On Dimensional Reduction Techniques in Signal Processing and Applications in EMG Analysis

DIPLOMARBEIT
(Überarbeitete Version)

zum

Diplom-Mathematiker

des

Department Mathematik, Universität Hamburg

von

Mijail Guillemard

Juni 2008

Betreuer: Prof. Dr. Armin Iske
Zweite Gutachter: Prof. Dr. Vicente Cortés

On Dimensional Reduction Techniques in Signal Processing and Applications in EMG Analysis

Mijail Guillemard

We present an overview of background concepts related to recent developments in signal analysis and dimensional reduction techniques, with a particular focus on new aspects of frame theory and reproducing kernels Hilbert spaces. Frame theory is an important generalization of wavelet and short term Fourier transforms which extends in a significant way the standard concept of a basis in a vector space. On the one hand, novel characterizations relating reproducing kernels with frame analysis have opened new perspectives in approximation theory and radial basis functions. On the other hand, the field of manifold learning and dimensional reduction has provided novel approaches based on geometrical concepts for applications in data analysis. A recurrent aspect in some of these subjects is the adequate usage of selected eigenvectors and eigenvalues in order to obtain geometrical information of an underlying manifold. This property is used as a motivation to shortly mention elementary aspects of spectral theory and Gelfand transforms. We illustrate these interactions as a means for a better understanding of the theoretical and algorithmic aspects of these topics. An illustrative toy example in Electromyogram (EMG) signal analysis, combining dimensional reduction with wavelet transforms, is finally discussed.

Contents

Introduction

In the last decades, a fast development in different areas of numerical and functional analysis has delivered new tools and perspectives in pure and applied mathematics. Frame theory, as a generalization of Fourier and time-frequency transforms, allows a new range of possibilities in adaptive signal analysis. Recent developments in approximation theory, reproducing kernel Hilbert spaces, and radial basis functions, allow to efficiently interpolate and approximate functions in high dimensional spaces. The emergent areas of manifold learning and dimensional reduction techniques, together with the related setup of kernel methods, offer efficient tools for classification and analysis of high dimensional data. Despite the seemingly different environments and objectives of these topics, significant connections can be observed in these developments. Some particular common points could be traced back in the framework of spectral theory and Banach algebras, a machinery that explains essential aspects of the far-reaching concept of spectrum. A better comprehension of these conceptual connections could be used as a means to synthesize, improve and identify differences, as well as develop new strategies that combine the strength of each subject. The objective of this report is to present an overview of elementary concepts necessary for understanding these developments, together with an illustrative toy-application example in Electromyogram signal analysis.

Harmonic Analysis. Harmonic Analysis is a fundamental framework for analysis and synthesis of functions based on the concept of Fourier transform. Over the last two centuries, this concept was an important source of ideas, examples, and open problems in many fields, including analysis, topology, and differential equations. Nowadays, many extensions and generalizations of this framework are playing an important role in new theoretical and application programs. Time-frequency techniques are natural extensions, where the two classical examples are the wavelet and the short term Fourier transforms. The success of these extensions are now, once again, leading to novel analysis strategies, open problems, and new research directions.

Frames. Frame theory was originated with the work of Duffin and Schaeffer in 1952 as a part of their investigations on non-harmonic Fourier series. The emergence of wavelet theory in the late 1980s, and its success in application problems, reactivated the theoretical and practical investigation of frame theory. A *frame* in a separable Hilbert space $\mathcal{H}$ is a collection of vectors $F = \{f_i\}_{i\in I} \subset \mathcal{H}$, such that there exist two constants $0 < A \leq B$, with $A\|f\|^2 \leq \sum_{i\in I} |\langle f, f_i\rangle|^2 \leq B\|f\|^2$, for any $f \in \mathcal{H}$. This definition generalizes simultaneously the concept of a basis and time-frequency transform, allowing a greater flexibility and keeping the useful property of exact reconstruction. The strategy behind this generalization is to design adaptable building blocks, while maintaining the standard signal processing philosophy of analysis, filtering and synthesis. The flexibility of the frame definition introduces a range of nontrivial examples and unexplored phenomena.

Adaptive Frames. A natural question is how to use the flexibility of this definition for constructing frames that fits a particular application. This topic, denominated *adaptive frames*, is currently an active research area, both in theoretical and engineering domains. The engineering goal is to implement algorithms for constructing frames adapted to a given experimental data set. The corresponding theoretical goal is to guide the algorithm design by understanding the set of frames with predefined properties. One particular question in this context is to analyze the set of all possible

frames $F = \{f_i\}_{i\in I} \subset \mathbb{R}^n$ that lie in (or are close to) a given manifold $\mathcal{M} \subset \mathbb{R}^n$. Initial results in this direction have been discovered recently, revealing a range of very rich phenomena. In [22], it has been proved that the space of spherical tight frames $\mathcal{F}^{\mathbb{R}}_{k,n}$, all of whose vectors $\{f_i\}_{i=1}^k$ lie on the unit sphere of $\mathbb{R}^n$, is a real analytic manifold when k and n are relatively prime. The starting point for addressing this problem is to consider the orbit space $\mathcal{G}^{\mathbb{R}}_{k,n} = \mathcal{F}^{\mathbb{R}}_{k,n}/\mathcal{O}^{\mathbb{R}}_n$ for the action of the group of $n \times n$ orthogonal matrices $\mathcal{O}^{\mathbb{R}}_n$. Even for frames in the low dimensional space $\mathbb{R}^2$, the structures are already surprisingly rich: $\mathcal{G}^{\mathbb{R}}_{4,2}$ is a graph with twelve vertices and twenty-four edges, and $\mathcal{G}^{\mathbb{R}}_{5,2}$ is the orientable surface of genus 25. It is to be expected that more complex constraints, adapted to specific applications, would require a more detailed algorithmic and theoretical analysis, as well as new research directions in adaptive frames.

RKHS. The development of frame theory, and the understanding of its role in other areas of mathematics, is still in its early stages of development. Recent investigations have uncovered a significant role of frame theory in the area of reproducing kernel Hilbert spaces (RKHS) [48, 51]. For an arbitrary set X, a *reproducing kernel Hilbert space* $\mathcal{H}$ is defined as the set of complex valued functions on X, with the additional reproducing property $f(x) = \langle f, K_x\rangle$, for all $f \in \mathcal{H}$. The unique function defined as $K(x, y) = K_x(y)$ is the kernel of the Hilbert space $\mathcal{H}$. This concept has been a classical ingredient in diverse fields of analysis such as partial differential equations and integral operators. A very important property of a reproducing kernel is an expansion of the form $K(x, y) = \sum_{i\in I} \psi_i(x)\overline{\psi_i(y)}$, with $\{\psi_i\}_{i\in I}$ an orthonormal basis of $\mathcal{H}$. The novel particularity described in [48, 51], is that this expansion can be generalized to a more flexible situation where the set $\{\psi_i\}_{i\in I}$ is actually a frame. The implications of this new property are current research problems, and it is to be expected that the interaction with the adaptive frame domain will open new research questions.

RBF. The concept of Reproducing Kernel Hilbert Spaces is also an important tool in modern approximation theory. Radial basis functions (RBF) are used to build flexible and efficient interpolation procedures by using an adequate reproducing kernel Hilbert space as an approximation tool. Given a point set $X = \{x_1, \dots, x_m\} \subset \mathbb{R}^n$, that we assume to be sampled from a manifold $\mathcal{M}$ (that is $X \subset \mathcal{M}$), we want to approximate an unknown function $f : \mathcal{M} \to \mathbb{R}$, using the sample values $\{f(x_i)\}_{i=1}^m$. We consider $\mathcal{H}_K(\mathcal{M})$, a reproducing kernel Hilbert space of functions on $\mathcal{M}$, with kernel $K : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$. The main requirement for constructing efficient interpolation procedures is the *positive (semi)definiteness* of the kernel, meaning that the matrix $(K(x_i, x_j))_{1\le i,j\le m}$ is positive (semi)definite for any $X = \{x_1, \dots, x_m\} \subset \mathcal{M}$. With this property we can solve a system of linear equations in order to find an interpolant $s_f \in \mathcal{H}_K(\mathcal{M})$, with $s_f(x_i) = f(x_i), i = 1, \dots, m$. If the kernel can be written as $K(x, y) = \phi(\|x - y\|)$, for a scalar function $\phi : \mathbb{R}_+ \to \mathbb{R}$, we say that ϕ is a *radial basis function*. In our context, the natural questions are the implications for the RBF theory in the recent discovered frames characterizations, both in the adaptive frames, and reproducing kernel Hilbert space settings.

Dimensional Reduction and Manifold learning. In the context of the study of a finite scattered set $X = \{x_1, \dots, x_m\} \subset \mathbb{R}^n$, an emergent area of analysis are *dimensional reduction* and *manifold learning* techniques. The primary objective in *dimensionality reduction* is to construct a set $Y = \{y_1, \dots, y_m\} \subset \mathbb{R}^d$ in a low dimensional space (i.e. $d \le n$) in a way that certain characteristics of the dataset X are conserved. Among different options one example is to build Y in a way that the distances between the points

in X are conserved. Namely, we search for a configuration Y with $\|y_i - y_j\| \approx \|x_i - x_j\|$ for all $i, j \in \{1, \ldots, m\}$. In the case of *manifold learning*, the new hypothesis is that $X \subset \mathcal{M}$, namely, X is sampled from $\mathcal{M}$, a p-dimensional smooth compact submanifold of $\mathbb{R}^n$. As in dimensionality reduction, the objective is also to construct a low dimensional representation $Y = \{y_1, \ldots, y_m\} \subset \mathbb{R}^d, d \leq n$ that conserves some characteristics of the dataset X, but now, the geometrical environment introduced by $\mathcal{M}$ will play a crucial role in the algorithm design. Due to the Whitney embedding theorem (which states that any connected smooth p-dimensional manifold can be smoothly embedded in $\mathbb{R}^{2p+1}$ [18]) we require some conditions for the dimensions in this formulation, namely, $2p + 1 \leq d \leq n$. Our problem can also be formulated as the search for an adequate embedding E of the p-dimensional submanifold $\mathcal{M} \subset \mathbb{R}^n$ in $\mathbb{R}^d$, with $E : \mathcal{M} \subset \mathbb{R}^n \to \Omega \subset \mathbb{R}^d$, $X \subset \mathcal{M}$, $Y \subset \Omega$, Ω a p-dimensional submanifold, and $2p + 1 \leq d \leq n$. An important additional topic in this field are density conditions on the finite data set X with respect to $\mathcal{M}$, which need to be analyzed in order to guarantee a meaningful usage of the geometry of $\mathcal{M}$ [47].

Decreasing the amount of information in $X \subset \mathcal{M}$, and working with the low dimensional parameter space $Y \subset \Omega$ is crucial for many applications as in classification techniques, regression methods, data storage, reliable prediction, etc. Many tools in this domain have been derived on a statistical context, but with the exponential increase in computer power, new alternative strategies have been proposed, with a more geometrical and analytical background. An important topic of this framework are techniques based on kernel methods, which have been successfully used in order to exploit the similarities between elements in the dataset X. Many of these modern developments have their roots in fundamental concepts of linear algebra, with a main example being the *singular value decomposition*. The geometrical role of the eigenvectors of the matrix [1] XX^t, has been a key element in classical dimensional reduction methods such as *principal component analysis*, *multidimensional scaling*, and several modern extensions.

Spectral theory. The significant role of the eigenproblem formulation in some standard algorithms in dimensionality reduction, together with the crucial aspect of the concept of spectrum in Fourier analysis is a motivation to better understand these subjects in a unified environment. A branch of functional analysis with a remarkable power of synthesis and generalization is the framework of spectral theory, Gelfand transforms, and C^*-algebras, a machinery that unifies and explains multiple instances of the concept of spectrum. Fourier analysis and linear algebra are two important subjects where different perspectives of the notion of spectrum are fundamental. The spectral theory and Banach algebras developed by Gelfand and his collaborators have explained and synthesized, in a far-reaching design, diverse mathematical phenomena. A primal example of a C^*-algebra is given by the space of continuous complex-valued functions $C(\tau)$ on a compact Hausdorff space τ. A fundamental result of Gelfand theory reveals that the algebra $C(\tau)$ carries all necessary information for the reconstruction of τ, the key ingredient is the concept of spectrum of the algebra $C(\tau)$, denoted by $\sigma(C(\tau))$. This environment is not only useful in a conceptual level, but it has been successfully used to resolve complex problems of numerical analysis [4, 29]. A question might be how to understand the recent developments in kernel methods, and manifold learning in the light of spectral theory.

[1] Here, we abuse the notation and use the symbol X also for denoting the matrix $X = (x_{ij})_{n \times m}$, with $x_i = (x_{1i}, \ldots, x_{ni})^t$, for all $i = 1, \ldots, m$.

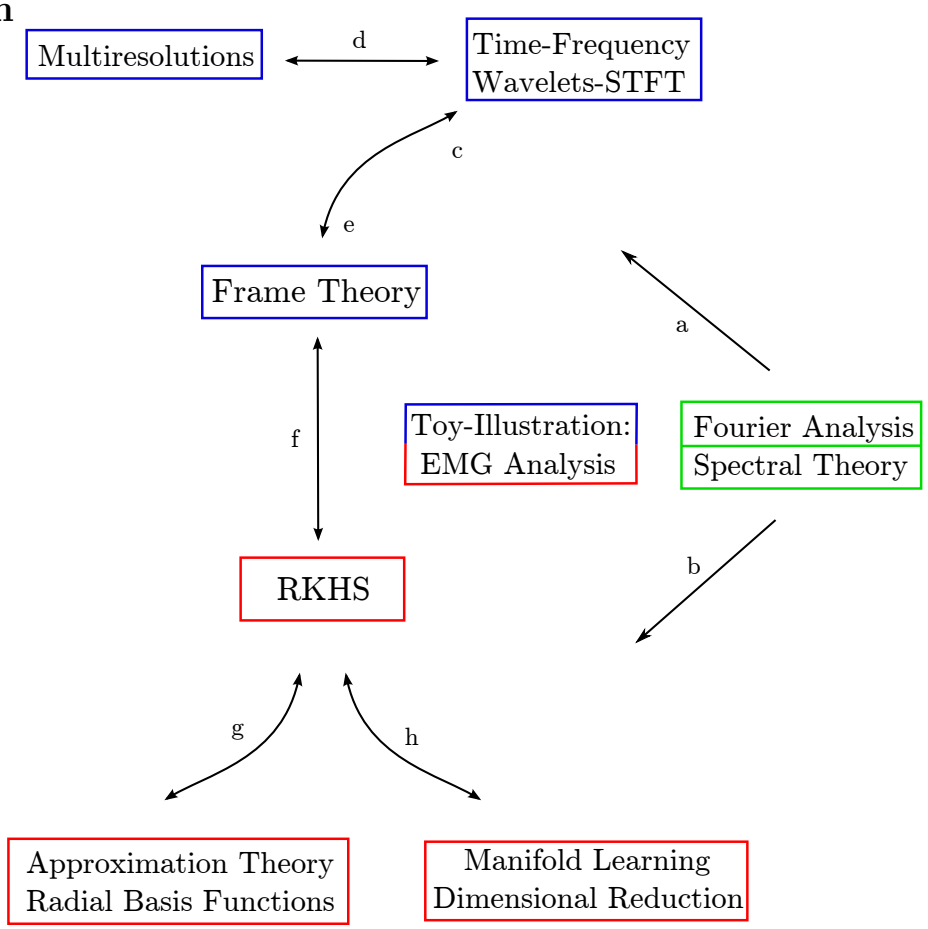
EMG analysis application. In order to present an application example illustrating the usage of these frameworks, we describe a toy problem in Electromyogram (EMG) signal analysis. The field of biological signal analysis has been increasingly demanding refined mathematical algorithms for supporting the study of living bodies. An electromyogram signal is an electrical measurement combining multiple *action potentials* propagating along motor neural cells. The analysis of these signals allows to obtain significant physiological information about the underlying muscle and nerve interactions. Standard applications of EMG analysis are diagnosis of diseases affecting the nervous and muscular system [38, 42], but additional uses have been developed in prosthetic and virtual devise interfacing [37], as well as athletic performance analysis [42]. An important task in this area is to construct adequate mathematical models of the underlying physiological mechanism. The understanding of the physical and biological processes behind the concept of action potential are actually active research topics. The standard explanation, the *Hodgkin-Huxley model*, is based on ionic exchanges on the cell membrane, but a recently proposed alternative model is based on the concept of *soliton* [30, 31], an important research topic in nonlinear PDE and wave propagation phenomena. The role of these models for algorithm design could be significant when constructing benchmark signals for evaluating and comparing new analysis algorithms. In the recent years, an important trend in electromyography has been the extensive usage of Fourier and wavelet transforms [26, 35, 38, 42]. In this report we experiment with a simple and illustrative framework that combines dimensional reduction techniques with standard signal transforms. A main goal is to describe, using a simplified environment, a concrete signal analysis problem that can be addressed with the concepts discussed in this report. The usage of nonlinear dimensional reduction techniques in application oriented problems in biology is an increasingly active research area (e.g. [12, 57]).

The structure of this report is as follows. In Chapter 1, we review background ideas in Fourier analysis and Spectral theory, stressing in particular the crucial concept of spectrum, not only in the harmonic analysis framework, but also its main role in linear algebra. In Chapter 2, we present a brief introduction to classical time-frequency transforms, focusing on the short term Fourier analysis, wavelet operators, and the important generalization of frame theory. Chapter 3 presents an introduction to reproducing kernel Hilbert spaces, and some applications in approximation theory, with particular emphasis in the recently discovered relations with frame analysis. Closely related concepts are discussed in Chapter 4, with the special usage of eigenvectors and eigenvalues in manifold learning and dimensional reduction techniques. Finally, the last chapter discusses the illustrative application of EMG signal analysis based on a combination of dimensional reduction techniques with Fourier and wavelet transforms.

Acknowledgments

I owe a great debt of gratitude to my advisor Prof. Armin Iske at the University of Hamburg for his continuous encouragement and positive support during the preparation of this report. A special thanks goes to Prof. Udo Zölzer for his constant support with my work at the Helmut Schmidt University. I would like also to express my gratitude to Prof. Vicente Cortés for his help and interest in this project.

Main Plan



Link	Description	Section
a	Fourier analysis: the conceptual and computational framework	1.1
b	Role of spectral theory in Fourier analysis and manifold learning	1.3
c	STFT and wavelets: two examples of time-frequency analysis	2.1
d	Multiresolutions structures of some wavelet functions	2.1
e	Frame theory generalizing bases and wavelet transforms	2.2
f	Frame theory and connections to RKHS	3.3
g	Role of RKHS in approximation theory	3.2
h	Role of kernels in dimensional reduction and manifold learning	4.3

The central topic of this plan is to present some perspectives concerning the interaction between frame theory and reproducing kernel Hilbert spaces (RKHS). In order to understand the central role of Fourier analysis, the language of group representations is required. A closely related framework is spectral theory and Gelfand transforms, which presents in a unified way background concepts in Fourier analysis and Linear Algebra. Kernel methods, under the scope of spectral theory, are important components in manifold learning and dimensional reduction techniques. We illustrate this set of interactions as a means for a better comprehension of the theoretical, practical, and algorithmic aspects of these topics. A simplified framework in Electromyogram (EMG) analysis is used as an illustrative example combining wavelet transforms and dimensionality reduction.

Notations

Fourier and Wavelet Analysis:

Notation	Description	Page
$\mathbb{T}$	Torus, Circle Group $\mathbb{T} := \{z \in \mathbb{C} : \lvert z\rvert = 1\} \cong \mathbb{R}/\mathbb{Z}$	15
$\hat{G}$	Dual Group $\hat{G} = \{\gamma : G \to \mathbb{T} : \gamma \text{ homomorphism}\}$	12
$C_0(G)$	Continuous Functions vanishing at infinity	13
$\lVert f\rVert_p$	p-norm $\lVert f\rVert_p := (\int \lvert f\rvert^p \, d\mu)^{1/p}\}$.	13
$L^p(G,\mu)$	Integrable functions $L^p(G,\mu) := \{f : G \to \mathbb{C}, \lVert f\rVert_p < \infty\}$.	13
$\mathcal{F}$	Fourier Transform	13
$\mathcal{A}$	Banach algebra	19
$M_n(\mathbb{C})$	C^*-algebra of $n \times n$ matrices over $\mathbb{C}$	19
$C(X)$	C^*-algebra of continuous complex-valued functions on X	19
$\sigma(\mathcal{A})$	Spectrum of an algebra	20
Γ	Gelfand transform $\Gamma : \mathcal{A} \to C(\sigma(\mathcal{A}))\quad \Gamma(x)(h) = \hat{x}(h) := h(x)$	20

Time Frequency Analysis and Frames:

Notation	Description	Page
$\mathcal{G}_g$	Short Term Fourier Transform	23
W_ψ	Wavelet Transform	25
ψ	Admissible wavelet	25
$\{V_i\}_{i\in\mathbb{Z}}$	Multiresolution	27
F	Frames $F = \{f_i\}_{i\in I} \subset \mathcal{H}$ with $A\lVert f\rVert^2 \leq \sum_{i\in I} \lvert\langle f, f_i\rangle\rvert^2 \leq B\lVert f\rVert^2$	28

Reproducing Kernels and Approximation Theory:

Notation	Description	Page
K	Kernel $K : E \times E \to \mathbb{R}$	29
Φ	Feature map $K(x,y) = \langle\Phi(x), \Phi(y)\rangle$	29
$\mathcal{H}_K$	Reproducing Kernel Hilbert Space (RKHS) with kernel K	29
$\mathcal{H}_\Phi$	RKHS with Feature space Φ	29
L_x	Point evaluation functional $L_x(f) = f(x)$	31
L_K	Integral Operator with kernel K	34

Dimensional Reduction:

Notation	Description	Page
PCD X	Point Cloud Data $X = \{x_k\}_{k=1}^m$	41
PCA	Principal Component Analysis	43
MDS	Multidimensional Scaling	46
ISOMAP	Isometric Feature Mapping	48

Application Examples, EMG experiments:

Notation	**Description**	**Page**
$u_k(t)$	EMG signal: sum of motor unit action potentials: $\sum_{j=1}^{p} V_{t_{ac}(k)}^{p_{jk}}(t)$	57,61
$V_{t_{ac}}^{p}(t)$	motor unit action potential (MUAP)	57,61
$g(t)$	polarization spike	60
t_{ac}	time constant of the afterpotential	62
q	number of MUAP in one EMG	62
$\mathcal{A}$	Parameterization map $\mathcal{A} : \Omega \to \Omega'$	58, 70
$\mathcal{T}$	Signal Transformation	58, 70
$\mathcal{P}$	Dimensional Reduction map	58, 70

1 Fourier transform and spectral concepts

Harmonic Analysis is an important cornerstone in modern mathematical techniques for decomposition, reconstruction and transformation of functions based the concept of Fourier transform. Historically this machinery has played an important role in many topics such as linear algebra, spectral theory, functional analysis, group theory, probability, and differential equations, among others. A notable amount of applications has been developed in the last two centuries, ranged over multiple areas such as signal processing, astronomy, differential equations, probability, chemistry, etc [10, 52]. A fundamental idea is that, given a function f defined in some space G, we build a two way mechanism that translates the information in f into another space related to the original domain G. In the case of the Fourier transform, the crucial aspect of such a scheme is the *duality* concept, described for this situation in a convenient context of group theory, an environment rich enough to explain most instances of Fourier transforms, while describing only the essential properties required for an harmonic analysis program.

One of the main problems with the classical framework in harmonic analysis is the difficulty when extracting time-dependent changes out of frequency information. Multiple replacements of the basic concept in Fourier analysis have been devised over the last decades, but despite its limitations, the principal conceptual features of harmonic analysis have a crucial role in defining, understanding and improving these new extensions. One example of these extensions is the Short Term Fourier Transform, which combines the time information with the Fourier analysis, by adequately splitting a signal in smaller chunks of information. The second extension is the wavelet transform that allows more flexible analysis procedures, and leads to important concepts such as multiresolution analysis and frame theory.

A very important domain, closely related to the harmonic analysis scheme, is the modern view of spectral theory under the framework of Gelfand transforms, Banach algebras and group representations, which allow to understand in a unified way core aspects of the concept of spectrum in operator theory and Fourier transforms. These subjects are not only crucial for an adequate understanding of the conceptual possibilities and extensions of harmonic analysis, but they are increasingly playing an important role in very concrete and applied subjects such as engineering aspects in robotics, image analysis, tomography, [10], design of efficient numerical signal processing algorithms [24, 25], etc.

1.1 Fourier Analysis

A basic strategy in signal processing is to construct an adequate transformation in order to extract special features of a signal. In the case of periodic functions, the mapping from the time to the frequency domain can be used for detecting harmonic content. A similar mechanism is a time-frequency transform, which allows to conveniently identify time-variable frequency features. An important characteristic of these transformations is the exact reconstruction procedures, which can be rephrased as isometry properties. Once an adequate isometric transformation is selected, a complete machinery for analysis, filtering, and synthesis can be constructed. This isomorphic transformation introduces the soul of this procedure: a duality mechanism that allows to construct two different perspectives of the same function.

There are different settings for developing this program, which mostly depend on the domain of the function to analyze, but all of them share a common background that represents a vital principle of harmonic analysis. The natural setting for this common background is to consider the domain of the functions as *locally compact groups*, namely, topological groups where the identity has a compact neighborhood. This environment is rich enough to explain the many instances of Fourier transforms, while representing only the essential properties of an harmonic analysis program. This accurate understanding of the Fourier transform started with the work of H. Weyl and L. Pontryagin in 1925, and has been since then an important motor for many branches in analysis [19]. The starting point is then to consider a locally compact abelian group together with the duality concept.

Definition 1.1 (Dual group)**.** The *dual* of a group G is defined by the set of *characters*

$$\hat{G} := \{\gamma : G \to \mathbb{T},\ \gamma \text{ homomorphism}\}.$$

That is, the set of homomorphisms between the group G, and the multiplicative abelian group of unit complex numbers $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\} \cong \mathbb{R}/\mathbb{Z}$, which is a subgroup of the multiplicative abelian group of all nonzero complex numbers $\mathbb{C}^\times := (\mathbb{C} - \{0\}, \times)$. If G is a topological group, the term *character* will mean *continuous character*.

The characters are elementary sort of functions that are used as analysis building blocks. A classical interpretation is to consider the group G as a *time space*, and the dual group $\hat{G}$ as the corresponding *frequency space*. In this framework, the characters are constant frequency functions, whose structure depends on the domain G. The commutativity condition of the group plays an important role: the case of locally compact non-abelian groups belongs to the more elaborate area of non-commutative harmonic analysis. Here we discuss only the simpler cases of locally compact abelian groups and compact groups.

Remark 1.1 (Topology in $\hat{G}$)**.** In order to work properly with the dual $\hat{G}$, an adequate topology needs to be constructed. This is particularly important when constructing a meaningful definition 1.2 of the Fourier transform. The strategy is to consider the topology generated by the uniform convergence. Namely, the topology generated by the uniform norm: $\|f\|_\infty = \sup\{|f(x)| : x \in \text{domain of } f\}$.

Remark 1.2 (Characters in group representations)**.** The concept of *character* belongs to the important realm of group representations which studies the structure of a group by mapping its properties to the richer environment of a vector space $\mathcal{H}$. A group representation generalizes the concept of character by considering homomorphisms of the group into a vector space of operators in some Hilbert space: $\rho : G \to \mathcal{U}(\mathcal{H})$. These definitions are not only crucial from a conceptual point of view, but they are also increasingly playing a main role in explaining and developing new efficient numerical algorithms in the context of signal processing [24, 25], or dealing with engineering aspects of robotics or image analysis [10].

The dual group $\hat{G}$ contains the characters of the group G, a question that naturally follows is to search for the characters of the dual group $\hat{\hat{G}}$. The answer of this question is an important characteristic of locally compact abelian groups, namely, the Pontryagin duality principle, which represents a landmark in Fourier analysis.

Theorem 1.3 (Pontryagin duality [36]). *There is a canonical isomorphism between a locally compact abelian group and its dual:*

$$f : G \to \hat{\hat{G}}, \quad f(g)(\gamma) \coloneqq \gamma(g), \quad g \in G,\ \gamma \in \hat{G}.$$

The main idea of this duality procedure is to identify in a natural way the characters of the dual group $\hat{G}$, as elements of the group G, indicating the crucial "reflexive" property of locally compact abelian groups [27, Sec.4.3].

In order to define integration procedures in a locally compact group, a proper measure needs to be constructed. A remarkable fact discovered by Alfred Haar in 1932 is that every locally compact group has an inherent measure invariant under the action of the group:

Theorem 1.4 (Haar Measure [27]). *For every locally compact group G there exists up to a positive multiplicative constant, a unique left (resp. right) measure μ with $\mu(aS) = \mu(S)$ (resp. $\mu(Sa) = \mu(S)$), for each $a \in G$, and $S \subset G$ a Borel subset. This measure will be called left (resp. right) Haar measure.*

Remark 1.5 (L^p spaces). With a convenient measure for any locally compact group in hand it is now possible to construct classical Banach spaces of integrable functions: $L^p(G, \mu) \coloneqq \{f : G \to \mathbb{C}, \|f\|_p \coloneqq (\int |f|^p \, d\mu)^{1/p} < \infty\}$. A natural consequence of the Haar measure are the invariance properties of the integral: $\int_G f(x) d\mu(x) = \int_G f(gx) \, d\mu(x)$.

The Fourier transform can now be defined as a linear operator that maps a function in the group into a function in the dual group. The first natural setting relates the space of integrable functions $L^1(G)$ with the space of continuous functions vanishing at infinity on the dual group, denoted by $C_0(G)$:

Definition 1.2 (Fourier transform in a locally compact group [27]). The *Fourier transform* is defined as a linear map

$$\mathcal{F} : L^1(G) \to C_0(\hat{G}) \qquad \mathcal{F}(f)(\gamma) = \langle \gamma, f \rangle = \int_G f(x) \overline{\gamma(x)} \, d\mu(x).$$

We denote $\mathcal{F}$ as the Fourier transform operator, but we use the notation $\hat{f}$ for the Fourier transform of $f \in L^1(G)$.

This map between the space of integrable functions $L^1(G)$ and the space $C_0(G)$, has a natural interpretation in the context of spectral theory and Banach algebras: we will see its potential in section 1.3. One of the following natural steps is to bring this transformation into a richer structural environment, namely a Hilbert space. As the space $L^1(G)$ has only a Banach structure, we move the map $\mathcal{F}$ to the Hilbert space setting of $L^2(G)$ using a classical analysis strategy [2, Sec.5.18]: by restricting the map $\mathcal{F}$ to $L^1(G) \cap L^2(G)$ it is possible to extend this transformation to $L^2(G)$. The key observation is to notice that $C_0(G) \subset L^1(G) \cap L^2(G)$ is a dense set in $L^2(G)$. We obtain the following crucial result of Fourier transforms and L^2 theory, namely, the Plancherel theorem, which states that $\mathcal{F}$ is in actually an isometry of $L^2(G)$:

Theorem 1.6 (The Plancherel Theorem [20, 27])**.** *The Fourier transform $\mathcal{F}$, extends to an isometry of $L^2(G)$:*

$$\int_{\hat{G}} |\hat{f}(\gamma)|^2 \, d\hat{\mu}(\gamma) = \int_G |f(x)|^2 d\mu(x).$$

Two basic operational procedures of the Fourier transform are related to the shift in the group translations. For the case of the time frequency interpretation these translations can be rephrased as a time-shift and frequency modulation.

Proposition 1.7 (Group translations)**.** *By defining the group translation by a factor $y \in G$ as $S_y : L^2(G, d\mu) \to L^2(G, d\mu)$, $S_y(f)(x) := f(x - y)$, and the corresponding modulation in the dual group $E_y : L^2(\hat{G}, d\mu) \to L^2(\hat{G}, d\mu)$, $E_y(r)(\gamma) := \overline{\gamma(y)} r(\gamma)$, we have the following relation:*

$$\mathcal{F} S_y = E_y \mathcal{F}.$$

Proof. A direct computation shows that:

$$\begin{aligned}
\mathcal{F}(S_y f)(\gamma) &= \langle \gamma, S_y f \rangle \\
&= \int_G f(x - y) \overline{\gamma(x)} \, d\mu(x) \\
&= \int_G f(x) \overline{\gamma(x + y)} \, d\mu(x) = \overline{\gamma(y)} \int_G f(x) \overline{\gamma(x)} \, d\mu(x) \\
&= \overline{\gamma(y)} \mathcal{F}(f)(\gamma) = E_y \mathcal{F}(f)(\gamma).
\end{aligned}$$

□

Remark 1.8. For the standard case of the Fourier transform of periodic functions in $\mathbb{R}$ we have the well known relations of translation in time and modulation in frequency: $\mathcal{F}(f(. - a))(n) = e^{ian} \hat{f}(n)$ and $\mathcal{F}(e^{ik \cdot} f(.))(n) = \hat{f}(n - k)$.

With this machinery in hand, it is now possible to understand all particular examples of Fourier analysis. The case of the trigonometric series was the first historical example considered by Joseph Fourier in the context of numerical solutions of the heat equation. For this situation, the natural working environment are functions defined on a compact support. The classical prototype is the multiplicative group of units $G = \mathbb{T}$, which can be used to analyze periodic functions. The corresponding dual group turns out to be the set of integers $\hat{\mathbb{T}} = \mathbb{Z}$. The characters for this group are identified with the trigonometric functions $\gamma_n(x) = e^{2\pi i n x}$, and by using directly the main construction, 1.2, we obtain the classical case of Fourier series [36]:

Definition 1.3 (Fourier series in $\mathbb{T}$)**.** The Fourier transform of a square integrable function on the torus $\mathbb{T}$, considered as the multiplicative group of complex numbers with absolute value 1, is defined as:

$$\mathcal{F} : L^2(\mathbb{T}) \to \ell^2(\mathbb{Z}) \qquad \mathcal{F}(f)(n) = c_n = \int_0^1 f(x) e^{-2\pi i n x} \, dx.$$

A similar situation arises when considering functions defined in a finite set. In order to construct a Fourier decomposition, a basic requirement is to have a group structure, which can be fulfilled by working with $G = \mathbb{Z}_p$, the group of integers modulo p. For this special situation, the dual group turns out to the isomorphic to the original group: $\hat{\mathbb{Z}}_p = \mathbb{Z}_p$. The characters are described with the pth roots of unity by $\gamma_n(x) = e^{2\pi i n x/p}, x, n \in \mathbb{Z}_p$. Again, using the main construction, 1.2, we obtain the well known formulas for the Finite Fourier Transform [1]:

Definition 1.4 (Finite Fourier Transform in $\mathbb{Z}_p$)**.** The Fourier transform of a function in the group of integers module p is defined as:

$$\mathcal{F} : L^2(\mathbb{Z}_p) \to L^2(\mathbb{Z}_p) \qquad \mathcal{F}(f)(n) = c_n = \sum_{k=0}^{p-1} f(x) e^{-2\pi i n x/p}$$

Remark 1.9. The general form of a character is $\gamma(x) = e^{if(x)}, x \in G$, and $f : G \to \mathbb{R}$. For the case $G = (\mathbb{R}, +)$, the homomorphism property $\gamma(xy) = \gamma(x)\gamma(y)$, can then be rephrased as a Cauchy's functional equation $f(x+y) = f(x)+f(y)$. The general handling of these equations is a non-trivial problem, but for our case of a continuous character we have classical result stating a parameterization $\hat{(\mathbb{R}, +)} = \{\gamma_m(x) = e^{imx}, m \in \mathbb{R}\} = (\mathbb{R}, +)$.

The third well known case of analysis corresponds to the functions defined on the real line. For this situation the working group and its dual turn out to be isomorphic $G = \mathbb{R}$, and $\hat{\mathbb{R}} = \mathbb{R}$. The corresponding characters are defined by $\gamma_\lambda(x) = e^{-2\pi i \lambda x}$. Using the main construction 1.2 we obtain the well known formulas for the Fourier Integral [36]:

Definition 1.5. The Fourier transform of a square integrable function in the real line is defined as:

$$\mathcal{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R}) \qquad \mathcal{F}(f)(\lambda) = \hat{f}(\lambda) = \int_{\mathbb{R}} f(x) e^{-2\pi i \lambda x}\, dx.$$

Convergence Issues

The questions raised by Joseph Fourier by the end of the 18th century for the expansions of functions as trigonometric series introduced a new point of view. Most of the functions used at that time where described with Taylor series, their good analytical behavior was well known, and integration and differentiation properties where particularly easy to handle. In contrast, the consideration of trigonometric functions as building blocks introduced many pathological and wild phenomena. In 1872, Weierstrass used trigonometric series to produce a classical example of a continuous and non differentiable function [2]:

$$f(x) \coloneqq \sum_{n=0}^{\infty} b^n \cos(a^n x), \quad 0 < b < 1, \quad ab > 1, \quad a \in \mathbb{N}.$$

To understand these new constructions, a precise definition of a function had to be worked out. These problems played an important role in the 19th century with the creation and formalization of new areas, such as analysis, topology, set theory, and

integration. The topic of Fourier series laid the ground of what is known today as classical harmonic Analysis. The exploration of this new arena in the 19th century, contrasted the well behaved constructions of analytic functions and Taylor series in the 18th century.

A main aspect of the theory of Fourier series are convergence problems. Given a function in some space, it can be far from trivial the question of whether its Fourier series converges to the function. These convergence problems have survived for a period of two centuries. The first positive behavior was discovered by Dirichlet in 1829, where he proved that the Fourier series of any piecewise smooth function is pointwise convergent to the function.

Theorem 1.10 (1829 Dirichlet)**.** *The Fourier series of a piecewise smooth function is pointwise convergent to the function.*

After this result, the accepted conjecture was that the Fourier series of any continuous function is pointwise convergent to the function. But this turned out to be wrong, and in 1909 Du Bois-Reymond presented a counterexample of a continuous function whose Fourier series diverge in a dense set.

Theorem 1.11 (1909 Du Bois-Reymond)**.** *There is a continuous function whose Fourier series diverge in a dense set.*

A second negative result came from Kolmogorov in 1926, where he proved that there is a function in L^1 whose Fourier series diverges almost everywhere.

Theorem 1.12 (1926 Kolmogorov)**.** *There is a function in L^1 whose Fourier series diverges almost everywhere.*

The following accepted conjecture of this period was then the existence of continuous functions whose Fourier series diverged everywhere. But once more, this intuition was wrong, and in 1966 Carleson proved in a famous result (and one of the most complex proofs in analysis) that the Fourier series of a function in L^2 converges almost everywhere [2, 20].

Theorem 1.13 (1966 Carleson)**.** *The Fourier series of a function in L^2 converges almost everywhere.*

1.2 Efficient algorithms and group representations

The 20th century presented a major development in the theoretical and practical aspects of harmonic analysis. A main technological milestone with a major economical and scientific impact was the successful implementation of fast algorithms for the Fourier transform. A main drawback of a direct application of the discrete Fourier transform in digital computers is the increasing complexity and computational time of the Definition 1.4. An efficient algorithm for resolving this problem, the Fast Fourier transform (FFT), was proposed by Cooley and Tukey in their celebrated paper of 1965 [13]. It was later realized that this idea was reinvented several times in the last 200 years. The first appearance dated back to Gauss, who used it for the interpolation of asteroidal orbits [52].

The importance of the Cooley and Tukey's papers was the timing in combination with the development of new computational resources in the 1960s. Two main aspects

where responsible for the immediate success of this publication: on the one hand the commercialization of transistor-based computers, and on the other hand the development of analog to digital converters. The result was a unprecedent combination of availability of raw data in many scientific domains, with the possibility of analysis with an efficient algorithm.

Since the rediscovery of the FFT in the 1960s, several reinterpretations of this algorithm have been devised leading to an important framework based on group representations that explains and extends the numerical efficient strategies used in the FFT [23–25]. This framework belongs to an important scheme that uses the language of group theory for a correct understanding of the efficiency of the finite or Discrete Fourier transform (DFT) [1]. Remember that the DFT is a linear map $F_n : \mathbb{C}^n \to \mathbb{C}^n$ that can be represented as a particular Vandermonde matrix, which are generally defined as $m \times n$ matrices containing a geometric progression on each row, that is $V_{ij} = \alpha_i^{j-1}$ for a given sequence $\{\alpha_i\}_{i=1}^m$ with $j = 1, \ldots, n$. The particular case of a DFT includes the powers of the nth root of unity as entries: by denoting $x, y \in \mathbb{C}^n$ we have

$$y = F_n x, \quad y_j = \sum_{k=0}^{n-1} \omega_n^{jk} x_k, \quad \omega_n = e^{-2\pi i/n}.$$

$$F_n = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ 1 & \omega_n & \ldots & \omega_n^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \omega_n^{n-1} & \ldots & \omega_n^{(n-1)(n-1)} \end{pmatrix}$$

The standard observation that the DFT of a vector requires $\mathrm{O}(n^2)$ operations, while the FFT reduces the execution speed to $\mathrm{O}(n\log(n))$, can be illustrated with the following factorization of the Fourier matrix F_4 [44], where we replace zeros by dots in order to visually emphasize the reduction of number of operations, and depict the implementation advantages of this factorization.

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix} = \begin{pmatrix} 1 & \cdot & 1 & \cdot \\ \cdot & 1 & \cdot & 1 \\ 1 & \cdot & -1 & \cdot \\ \cdot & 1 & \cdot & -1 \end{pmatrix} \begin{pmatrix} 1 & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot & i \end{pmatrix} \begin{pmatrix} 1 & 1 & \cdot & \cdot \\ 1 & -1 & \cdot & \cdot \\ \cdot & \cdot & 1 & 1 \\ \cdot & \cdot & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 & \cdot \\ \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 \end{pmatrix}$$

This elementary example is a basic observation of an important framework that uses the theory of group representations for designing fast signal processing numerical algorithms [23–25]. The principal idea is related to the notion of *symmetry* of a matrix M, a concept based on a pair of group representations used to create a framework for identifying redundancies in the matrix M. Constructing efficient factorizations of a matrix M follows as a consequence of the properties of these representations and the underlying group.

More precisely, we consider finite and complex representations (of *degree* n) of a group G defined as homomorphisms $\phi : G \to \mathrm{GL}_n(\mathbb{C})$ into the group $\mathrm{GL}_n(\mathbb{C})$ of invertible $n \times n$ matrices over the complex numbers. The first step when studying such objects is to decompose them into "irreducible" ones, which for this case means that a given

representation cannot be described as *direct sums*, namely, representations of the form $\phi \oplus \psi : \phi(g) \oplus \psi(g) = \begin{pmatrix} \phi(g) & 0 \\ 0 & \psi(g) \end{pmatrix}$.

The concept of *symmetry* of a matrix $M \in \mathbb{C}^{m \times n}$ is defined as a pair (ϕ_1, ϕ_2) of representations of a group G such that $\phi_2(g)M = M\phi_1(g)$, for all $g \in G$. In other words, the following commutative diagram holds for any $g \in G$:

$$\begin{array}{ccc} \mathrm{GL}_n(\mathbb{C}) & \xrightarrow{\phi_1(g)} & \mathrm{GL}_n(\mathbb{C}) \\ \Big\downarrow{\scriptstyle M} & & \Big\downarrow{\scriptstyle M} \\ \mathrm{GL}_m(\mathbb{C}) & \xrightarrow{\phi_2(g)} & \mathrm{GL}_m(\mathbb{C}) \end{array}$$

As a shortcut, we use the notation $\phi_1 \xrightarrow{M} \phi_2$ for expressing the symmetry (ϕ_1, ϕ_2) of M. Another way to describe this property is to say that M actually belongs to the *intertwining space* of the representations ϕ_1, and ϕ_2, namely, M belongs to the space:

$$\mathrm{Int}(\phi_1, \phi_2) := \{M \in \mathbb{C}^{\deg(\phi_2) \times \deg(\phi_1)} \mid \phi_2(g)M = M\phi_1(g),\ g \in G\}.$$

In order to consider meaningful representations we do not handle arbitrary symmetries, but instead we specialize our definition to particular types of representations including *permutation*, *monomial*, *irreducible*, etc. We say that ϕ is a permutation (resp. monomial) representation if all $\phi(g), g \in G$ are permutation matrices (resp. monomial matrices). A monomial matrix is a generalization of a permutation matrix which has exactly one non-zero entry in every row and column. It turns out that these types of special representations allows to detect useful patterns in the matrix M.

Now, the procedure for constructing efficient numerical algorithms, described as special factorizations of the matrix M, is based on the decomposition of the representations in a symmetry (ϕ_1, ϕ_2) of the matrix M. First, we decompose the representations ϕ_1 and ϕ_2 with the matrices A_1 and A_2 by considering the representations $\rho_i = \phi_i^{A_i}$, with $\phi_i^{A_i}(g) := A_i^{-1}\phi_i(g)A_i$ for $i = 1, 2$. This procedure can be represented with the following commutative diagram:

$$\begin{array}{ccc} \phi_1 & \xrightarrow{M} & \phi_2 \\ \Big\downarrow{\scriptstyle A_1} & & \Big\downarrow{\scriptstyle A_2} \\ \rho_1 & \xrightarrow{D} & \rho_2 \end{array}$$

In order to obtain a factorization of M which can be used for implementing efficient numerical algorithms (as illustrated in the previous factorization of the Fourier matrix F_4) we consider the matrix $D := A_1^{-1}\, M\, A_2$, which turns out to have a critical property of being *sparse*, and we obtain the desired efficient decomposition

$$M = A_1\, D\, A_2^{-1}.$$

This scheme can be further generalized when considering general fields $\mathbb{F}$ instead of our original case $\mathbb{C}$, and the resulting framework allows to answer the question of finding efficient methods for evaluating Mx, for $x \in \mathbb{F}^n$, $M \in \mathbb{F}^{m \times n}$ with as few arithmetic operations in the base field $\mathbb{F}$ as possible.

1.3 Elements of Spectral Theory

The Fourier Analysis generalization based on locally compact groups allows to better understand, unify and extend the analysis and synthesis philosophy devised in classical harmonic analysis. An intimately related framework is spectral theory and Gelfand transforms, a machinery that unifies and explains the multiple instances of the concept of spectrum. Fourier analysis and linear algebra are two fundamental domains where different perspectives of the notion of spectrum are deeply integrated. The multiple application contexts of these two domains indicates the far-reaching influence of spectral theory, but its strategy goes far beyond this unification procedure, it also allows to identify the same important mathematical structures in different working environments. One of the primal examples of spectral and Gelfand theory is the recovery of a topological Hausdorff space using the ring of continuous complex-valued functions defined on it. This framework underlines a deep interaction between geometrical and algebraic structures providing a main motor in the modern developments of differential geometry and algebraic topology. But in the last two decades this conceptual framework has begun to infiltrate more numerical and computational domains. In [4], for instance, the analysis of numerical problems related to Toeplitz matrices has been analyzed with C^*-algebra methods. In this section we present very elementary ideas of spectral theory, and refer to the main references for a comprehensive introduction [19, 27, 32, 39]. We follow closely the presentation in [27], extracting only elementary relevant results for understanding our context.

Definition 1.6 (Banach Algebra)**.** A *vector space* is a group $(\mathcal{A}, +)$ with a scalar multiplication, $\mathbb{F} \times \mathcal{A} \to \mathcal{A}$, over the field $\mathbb{F}\,(= \mathbb{R}, \mathbb{C})$ compatible with the group operation. An *associative algebra* is a vector space $\mathcal{A}$ with a multiplication (bilinear map $\mathcal{A} \times \mathcal{A} \to \mathcal{A}$) which is distributive and associative. If the multiplication operator has an *identity*, we have an *unital* algebra, which can be rephrased as a ring $(\mathcal{A}, +, \times)$ with a scalar product in $\mathbb{F}$. A *Banach algebra* is an normed algebra $\mathcal{A}$ that is a Banach space in which the multiplication and the norm relates as $\|xy\| \leq \|x\|\,\|y\|$, for all $x, y \in \mathcal{A}$.

Definition 1.7 (C^*-algebra)**.** A **-algebra* is an algebra with an *involution*, namely, an automorphism $^* : \mathcal{A} \to \mathcal{A}$, such that $(x+y)^* = x^* + y^*$, $(\lambda x)^* = \bar{\lambda} x^*$, $(xy)^* = y^* x^*$, and $x^{**} = x$. A *C^*-algebra* is Banach *-algebra such that $\|xx^*\| = \|x\|^2$, for all $x \in \mathcal{A}$.

Remark 1.14 (Examples of C^*-algebras)**.** The following list enumerates some important examples of Banach algebras, indicating the rich diversity and breath of this theory.

1. $M_n(\mathbb{C})$. The set $M_n(\mathbb{C})$ of all $n \times n$ matrices over $\mathbb{C}$ is a C^*-algebra with the usual addition and multiplication of matrices. The involution is given by the complex transpose $(T^*)_{ij} = \overline{T_{ji}}$, and the operator norm $\|T\|_{\text{op}} = \min\{c : \|Tx\| \leq c\|x\| \,\forall x \in \mathbb{C}^n\}$ is used to construct a Banach space.

2. $C(X)$. If X is a compact Hausdorff space, the space $C(X)$ of continuous complex-valued functions on X is a unital Banach algebra. The addition and multiplication are the usual pointwise operations. The involution is given by the pointwise complex conjugation, and we use the uniform (supremum) norm $\|f\|_\infty = \sup\{|f(x)| : x \in \text{domain of } f\}$.

3. $C_0(X)$. If X is a locally compact Hausdorff, the space of continuous functions vanishing at infinity $C_0(X)$ is a nonunital C^*-algebra. The addition, multiplication, involution, and norm are defined as in the previous example. Notice that the characteristic function, 1_X, is not in $C_0(X)$.

4. $L(\mathcal{H})$. For a Hilbert space $\mathcal{H}$, the set of bounded linear operators $L(\mathcal{H})$ is C^*-algebra with the pointwise addition, and the composition, and adjoint, as the multiplication and involution respectively. The norm is also given by the operator norm: $\|T\|_{\text{op}} = \sup_{\|x\|\leq 1}\|T(x)\|$.

5. $l^1(\mathbb{Z})$. The set of complex-valued sequences $a = (a_n)_{-\infty}^{\infty}$ such that $\|a\| = \sum_{-\infty}^{\infty} |a_n| < \infty$ is denoted by $l^1(\mathbb{Z})$, and is a unital $*$-algebra with the pointwise addition, and the convolution $a * b = c$, with $c_n = \sum_{-\infty}^{\infty} a_k b_{n-k}$. The involution is given by $(a^*)_n = \overline{a_{-n}}$. This algebra is not a C^*-algebra.

6. $L^1(\mathbb{R})$. The space $L^1(\mathbb{R})$ of integrable functions is a nonunital Banach $*$-algebra with the multiplication given by the convolution $(f * g)(x) = \int f(y)g(x-y)\,dy$, and the involution defined as $f^*(x) = \overline{f(-x)}$.

Definition 1.8 (Spectrum of an algebra)**.** Suppose $\mathcal{A}$ is a commutative unital Banach algebra. The spectrum of $\mathcal{A}$ is given by the set of all nonzero multiplicative functionals:

$$\sigma(\mathcal{A}) = \{\gamma : \mathcal{A} \to \mathbb{C}, \gamma \text{ nonzero homomorphism}\}$$

An important aspect of the the spectrum of an algebra is that as $\sigma(\mathcal{A})$ lies in the set of bounded linear functionals $\mathcal{A}^*$, the weak-* topology can be used to make $\sigma(\mathcal{A})$ a topological space. Additionally, with the help of the map $h \in \sigma(\mathcal{A}) \to \ker(h)$, we can identify $\sigma(\mathcal{A})$ with the set of maximum ideals of $\mathcal{A}$, providing an identification between topological and algebraic information related to $\mathcal{A}$.

The motivation for calling $\sigma(\mathcal{A})$ the spectrum is based on the fact that, in general, we have

$$\text{range}(\hat{x}) = \sigma(x),$$

and for particular cases (e.g. when $\mathcal{A}$ is generated by x_0 and e), $\hat{x}$ as a function in $\sigma(\mathcal{A})$ is an homomorphism from $\sigma(\mathcal{A})$ to $\sigma(x_0)$.

Definition 1.9 (Gelfand transform)**.** Suppose $\mathcal{A}$ is a commutative unital Banach algebra. The Gelfand transform is the homomorphism defined by

$$\Gamma : \mathcal{A} \to C(\sigma(\mathcal{A})) \quad \Gamma(x)(h) = \hat{x}(h) \coloneqq h(x) \quad \forall x \in \mathcal{A},\, h \in \sigma(\mathcal{A}).$$

One of the main characteristics of the Gelfand transform is the possibility to identify an arbitrary commutative unital C^*-algebra with a C^*-algebra of continuous functions defined over the spectrum, which is the celebrated Gelfand Naimark theorem:

Theorem 1.15 (Gelfand Naimark theorem: Gelfand transform of commutative unital C^*-algebras)**.** *For any commutative unital C^*-algebra $\mathcal{A}$, the Gelfand transform is an isometric *-isomorphism from $\mathcal{A}$ to $C(\sigma(\mathcal{A}))$.*

Spectral theorem

The classical spectral theorem of linear algebra, together with the standard concepts of spectrum, eigenvalue and eigenvectors can be generalized and embedded in the Gelfand theory scheme. Remember that the finite dimensional spectral theorem says that if T is a self-adjoint operator on $\mathcal{H}$, then there is an orthonormal basis of $\mathcal{H}$ consisting of eigenvectors of T with the following decomposition:

$$T = \sum_{\lambda \in \Sigma} \lambda P_\lambda,$$

where Σ is the spectrum of T, and P_λ is the orthogonal projection onto the eigenspace for λ. The extension of this property to a general scheme using Gelfand transforms has several variations, one particular situation considers an arbitrary C^*-subalgebra $\mathcal{A}$ of $\mathcal{L}(\mathcal{H})$ instead of a single operator T. We have in this scheme the following three basic ingredients:

(1) A Hilbert space $\mathcal{H}$.

(2) A commutative C^*-subalgebra $\mathcal{A}$ of $\mathcal{L}(\mathcal{H})$ containing the identity.

(3) $\Sigma := \sigma(\mathcal{A})$ the spectrum of $\mathcal{A}$.

Another crucial ingredient in this framework is the Gelfand transform $\hat{T}$ for any $T \in \mathcal{A}$, and the inverse operation $T_f \in \mathcal{A}$, well defined for any $f \in C(\Sigma)$ thanks to the Gelfand-Naimark Theorem 1.15.

An additional important component is the concept of *projection valued measure*, which extends the standard measure definition by mapping an element of a σ-algebra $\mathcal{A}$ of a measurable space $(\Omega, \mathcal{A})$ to a bounded linear operator in $\mathcal{L}(\mathcal{H})$. A projection valued measure $P : \mathcal{A} \to \mathcal{L}(\mathcal{H})$ fulfills an extended version of the properties of a standard measure.

(1) $P(E)$ is an orthogonal projection for any $E \in \mathcal{A}$.

(2) $P(\emptyset) = 0$ and $P(\mathcal{A}) = I$.

(3) $P(E \cap F) = P(E)P(F)$, for $E, F \in \mathcal{A}$.

(4) $P(\cup E_i) = \sum P(E_i)$ for any disjoint sequence $E_1, E_2, \ldots$, where the sum converges in the strong operator topology.

We can now extend the finite dimensional spectral theorem by considering the previous ingredients in one single package:

Theorem 1.16 (Spectral Theorem)**.** *Let $\mathcal{A}$ be a commutative C^* subalgebra of $\mathcal{L}(\mathcal{H})$ containing the identity, and let Σ be its spectrum. Then there is a unique projection valued measure P on Σ such that for all $T \in \mathcal{A}$ we have:*

$$T = \int \hat{T}\, dP.$$

Fourier analysis and Spectral theory

The Fourier analysis methods described in the previous sections, together with the standard concept of spectrum used in Fourier transforms, is another crucial structure that can be described and explained in the context of the Gelfand theory. In order to explain this remark, we can interpret the internal machinery in the harmonic analysis of a locally compact group G as an interplay between two frameworks: the study of representations of the group G, and the analysis of the Banach algebra $L^1(G)$.

On the one hand, one of the objectives of group representation theory is to *analyze* arbitrary representations of G, where a crucial example is the analysis of the regular representations, that is, homomorphisms based on left and right translations of the group. For instance, the *left regular representation* is defined as

$$\pi_L : G \to \mathcal{L}(L^1(G)),$$

with $(\pi_L(x)f)(y) = L_x f(y) := f(x^{-1}y)$, for $x, y \in G$, and $f \in L^1(G)$. In this situation, the term *analysis* means an explicit description of these representations in terms of basic building blocks, which are, for this environment, the *irreducible representations*, namely, representations that have no trivial invariant subspaces. If G is abelian, every irreducible representation is one dimensional, and a direct link with the set of characters $\hat{G}$ is established. In general, there are three particular situations that have to be considered when analyzing the structure of $\hat{G}$: the case of G being a locally compact abelian group, compact group, and locally compact nonabelian group.

On the other hand, we want to analyze complex-valued functions in the group G, belonging to the Banach algebra $L^1(G)$, that is we are dealing with the Gelfand theory domain. Now, these two issues (the general group representation problematic, and the analysis of $L^1(G)$) are intimately related via the concept of the Fourier transform, which turns out to be an instance of the Gelfand transform. More precisely, we can identify the dual group $\hat{G}$ with $\sigma(L^1(G))$, the spectrum of the algebra $L^1(G)$, by considering, for any character $\xi \in \hat{G}$, the functional

$$f \to \overline{\xi}(f) := \int \overline{\xi(x)} f(x)\, dx.$$

We can now obtain, as explained in Definition 1.9, an explicit description of the Gelfand transform on the Banach algebra $L^1(G)$:

$$\Gamma : L^1(G) \to C_0(\hat{G}) \qquad \Gamma(f)(\xi) := \overline{\xi}(f).$$

But, as explained in Definition 1.2, this functional rephrases the concept of the Fourier transform:

$$\mathcal{F} : L^1(G) \to C_0(\hat{G}) \qquad \mathcal{F}(f)(\xi) := \overline{\xi}(f).$$

We obtain therefore an important connection between the task in group representation theory of analyzing the irreducible representations (described for the case of abelian groups as the character group $\hat{G}$), and the fundamental objective of Gelfand theory when analyzing the Banach algebra $L^1(G)$. The key synchronizing link turns out to be the identification of the dual group $\hat{G}$ with the spectrum $\sigma(L^1(G))$, and the Fourier transform $\mathcal{F}$ with the Gelfand transform Γ.

2 Time-Frequency analysis and Frames

2.1 Continuous Transforms

A main problem in classical harmonic analysis is how to obtain the time evolution of a frequency component in a signal, using only the information provided by the Fourier Transform. In the first half of the 20th century, D. Gabor and other researchers proposed a scheme known today as the Gabor transform, whose first step consist of partitioning the signal in short consecutive and overlapped sections, using a Gauss function as a transformation window. In the second step of this procedure, the sequence of short sections are analyzed with the Fourier transform, in such a way that a the time evolution of the frequency components can be obtained. This procedure can be generalized by using other types of windows functions resulting in the well known *Short term Fourier Transform* (STFT).

The previous scheme is a one side of a two way road in which the first direction is the *analysis* of a function using the STFT as a decomposition tool, and the second direction is the *synthesis* of a function implemented using the STFT as a composition tool.

Due to the Heisenberg uncertainty principle, a tiling of the time-frequency space is inherent in each STFT transformation, with the particular characteristic of being uniform over all frequencies. In many applications this is not a desirable feature, as high frequency information requires a small resolution in the time domain while, inversely, a large window in required in order to correctly detect low frequency information. In the 1980s a remarkable collaboration between mathematicians, physicists, and engineers, improved the STFT scheme by formalizing the wavelet theory. One key idea of this new procedure is to design more adequate tilings of the time-frequency plane by using special basis functions called wavelets. Many important components of this framework where already known in the literature, but the main concepts where spread out over diverse subfields of classical harmonic analysis, group representations, and filterbank theory. The understanding of these ideas in the unified framework of wavelet theory allowed a deeper comprehension of time-frequency analysis, and led to further developments such as multiresolution structures and frame theory [50].

Short Term Fourier Transform

The Fourier transform provides information about the frequency content of a function, but an important question in signal processing is how to obtain the time evolution of the main frequency components of a signal. Standard procedures include the modulus of the Fourier transform, also known as *spectra*, as a main analysis tool, where a local maxima indicates the presence of an harmonic element. The crucial problem is that this procedure ignores the time evolution of this frequency local maxima. A similar problem happens when analyzing signals with different waveforms but similar frequency content, in which a direct consideration of the spectra information cannot be used to differentiate the signals. One solution strategy for revealing the relation between time and frequency content, is the framework proposed by A. Gabor and other researchers in the first half of the 20th century. The main idea is to use a window function for partitioning the signal before the application of the Fourier transform [11]:

Definition 2.1 (Short Term Fourier Transform (STFT) *analysis*)**.** The Short Term Fourier Transform in $L^2(\mathbb{R})$ is defined as a map $\mathcal{G} : L^2(\mathbb{R}) \to L^2(\mathbb{R} \times \mathbb{R})$, with

$$\mathcal{G}_g f(b,\omega) \coloneqq \langle f, g_{b,\omega} \rangle = \int_{\mathbb{R}} f(t)\overline{g_{b,\omega}(t)}\, dt, \quad g_{b,\omega}(t) \coloneqq g(t-b)e^{i\omega t}.$$

In this framework several possibilities are available for the window function g. For the case of a Gaussian window $g_\alpha(x) = 1/(2\sqrt{\pi\alpha})e^{-x^2/4\alpha}$ the corresponding map $\mathcal{G}_{g_\alpha}$ is known as *Gabor Transform*. Other window functions with different time-frequency behavior are available, such as the Hann, Hamming, Triangular, Kaiser, Blackman, etc. Many applications traditionally based on the Fourier transform could be improved by using the additional structure provided by the STFT scheme. In the last decades of the 20th century this framework opened several new areas of theory and applications under the name of *time-frequency analysis*.

The previous step is considered as an *analysis* procedure, where the map $\mathcal{G}_g f$ delivers a time-frequency view of the original function f. The second crucial component of this framework is a *synthesis* procedure, which corresponds to the possibility of reconstructing the original function f, using the analysis performed with $\mathcal{G}_g f$. The heart of the synthesis step is an isometry property of the STFT, which allows in a further step to construct an explicit reconstruction function.

Theorem 2.1 (Isometry of the STFT [11])**.** *If $g \in L^2(\mathbb{R})$, with $\|g\| = 1$, $tg(t) \in L^2(\mathbb{R})$ and $\omega\hat{g}(\omega) \in L^2(\mathbb{R})$, then*

$$\begin{aligned} 2\pi\langle f_1, f_2 \rangle &= \langle \mathcal{G}_g f_1, \mathcal{G}_g f_2 \rangle \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \langle f_1, g_{b,\omega}\rangle\overline{\langle f_2, g_{b,\omega}\rangle}\, db\, d\omega \qquad \forall f_1, f_2 \in L^2(\mathbb{R}). \end{aligned}$$

The reconstruction formula can now be directly obtained from the isometric property, providing the two way mechanism that translates information between f and $\mathcal{G}_g f$:

Corollary 2.2 (Reconstruction with STFT *synthesis*)**.** *Using the hypothesis of the Theorem 2.1, we have for any $t \in \mathbb{R}$ where f is continuous:*

$$f(t) = \frac{1}{2\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} \mathcal{G}_g f(b,\omega) g_{b,\omega}(t)\, db\, d\omega.$$

Wavelet Transforms

The STFT proved to be a very important tool for many problems where simultaneous time-frequency information is required. But this technique has some limitations, as its time-frequency tiling has a fixed resolution for all frequencies, providing an inconvenient scheme that handles high and low frequencies with the same fixed analysis tool. To resolve this problem, a new range of improvements came from a notable collaboration between mathematicians, physicists, and engineers in the 1980s, constructing the framework known today as wavelet theory. The new procedure extends the previous STFT construct by using a different strategy for analyzing time-frequency information. The core idea is to use a special type of window function, called *wavelet*, that is localized both in the time and the frequency domain and which, additionally, allows to replace the frequency modulation by a scaling operation. This new property introduces the possibility of constructing more adequate tiling of time-frequency plane [11, 33]:

Definition 2.2 (Wavelet Transform *analysis*)**.** The wavelet transform in $L^2(\mathbb{R})$ is defined as a map $W : L^2(\mathbb{R}) \to L^2(\mathbb{R} \times \mathbb{R}^*)$ with

$$W_\psi f(b,a) := \langle f, \psi_{b,a} \rangle = \int_{\mathbb{R}} f(t)\overline{\psi_{b,a}(t)}\,|a|^{-1/2}\,dt, \quad \psi_{b,a}(t) := \psi\left(\frac{t-b}{a}\right),$$

where we require $\psi \in L^2(\mathbb{R})$ to be an "admissible" function, that is:

$$C_\psi := \int_{\mathbb{R}} \frac{|\hat{\psi}(\omega)|^2}{|\omega|}\,d\omega < \infty.$$

In the same spirit as in the STFT, the crucial component that pairs the analysis procedure with the synthesis mechanism is the isometry property of the wavelet transform.

Theorem 2.3 (Isometry of the Wavelet transform)**.** *If ψ is a basic wavelet, we have:*

$$\begin{aligned} C_\psi \langle f_2, f_2 \rangle &= \langle W_\psi f_1, W_\psi f_2 \rangle \\ &= \int_{\mathbb{R}^*} \int_{\mathbb{R}} (W_\psi f_1)(b,a)\overline{(W_\psi f_2)(b,a)}\,\frac{da}{a^2}\,db. \qquad \forall f_1, f_2 \in L^2(\mathbb{R}). \end{aligned}$$

As in the case of the STFT, we can derive a reconstruction formula for the wavelet transform, constructing a new translation mechanism between a function f and its transform $W_\psi f$:

Corollary 2.4 (Reconstruction with wavelet transforms *Synthesis*)**.** *With the hypothesis of the Theorem 2.3, we have for any $t \in \mathbb{R}$ where f is continuous:*

$$f(t) = \frac{1}{C_\psi} \int_{\mathbb{R}^*} \int_{\mathbb{R}} W_\psi f(b,a)\psi_{b,a}(t)\,\frac{da}{a^2}\,db.$$

Remark 2.5 (Admissibility condition)**.** We notice that the admissibility condition is clearly necessary in the reconstruction formula of the wavelet transform. Additionally, this requirement provides an intuitive explanation for the term "wavelet", meaning a "small wave". We remark that $C_\psi < \infty$ implies $\hat{\psi}(0) = 0$, which in turn implies the following property, intuitively related to the idea of a "small wave" localized over time:

$$\int_{\mathbb{R}} \psi(t)\,dt = 0.$$

A significant aspect of this framework is that several of the original ideas where scattered concepts already present in diverse fields as classical harmonic analysis, group representations and filterbank structures. For instance, the way the modulation and the scaling operators act in the STFT and wavelet transforms is reminiscent of the action of two important examples of locally compact non abelian groups used in physics: the Weyl-Heisenberg and Affine group. This remark opens a very important link with the field of group representations, a fact that was presented in the seminal paper of Grossmann, Morlet, and Paul in 1985 [28]. Similarly, a well known area in the engineering literature, filterbank theory, had some initial components of the discrete version of the wavelet transform, that where further developed with the the concept of multiresolution analysis. One of the very first examples of wavelets are the Haar functions, designed by A. Haar

in his PhD thesis in 1909. An impressive result of his work is that the Haar function generates a basis of $L^2(\mathbb{R})$ for which the expansion of any function converges uniformly. A phenomenon that contrast completely with the seemingly better behaved Fourier basis [50]. This result of classical harmonic analysis played an important role in the initial development of wavelet theory. But the unification work of this new framework placed in a coherent structure many aspects of these, in appearance, different areas of mathematics, engineering, and physics. This unification opens at the same time further directions for theoretical and application research.

A Global View

An important concept behind the STFT and Wavelet transforms lies in the way a locally compact group G acts in a Hilbert space of functions $\mathcal{H}$. This action is an irreducible and unitary group representation, $\pi : G \to \mathcal{U}(\mathcal{H})$, that fulfills square integrable conditions. It is important to notice that the group G plays a different role than the one used in the Fourier transform definition, here G acts in the space of functions, and in the case of the Fourier transform the group G is the domain of the functions. With this representation the so called voice transform is constructed by setting $V_\psi : \mathcal{H} \to L^2(G)$, with $V_\psi(f)(x) := \langle f, \pi(x)(\psi)\rangle$. The STFT and Wavelet transforms are typical examples where ψ corresponds to a window function for the former and to a wavelet for the latter. These transforms represent an interplay between $\mathcal{H}$ and $L^2(G)$, that allows to analyze the function f, by porting its information to a setting defined by G. Another way to rephrase this phenomenon is that the transformation V_ψ "unfolds" data present in f, using G as an analysis environment. In the case of the STFT transform, the Weyl-Heisenberg group represents the time-frequency background to which information from f is translated. In the case of the wavelet transform, the affine group is responsible for providing a time-scale representation of the function. In these situations, a fundamental objective is to understand the components of f, using G and V_ψ as new observation tools.

The first crucial concept upon which the voice transform results useful are the orthogonality relations. This means that for the square integrable representation π, there is a unique self-adjoint positive operator C on $\mathcal{H}$, such that for $f_1, f_2 \in \mathcal{H}$ and $\psi_1, \psi_2 \in \operatorname{dom} C$, we have:

$$\int_G \langle f_1, \pi(x)\psi_1\rangle \, \langle \pi(x)\psi_2, f_2\rangle d\mu(x) = \langle f_1, f_2\rangle \, \langle C\psi_1, C\psi_2\rangle.$$

These relationships imply that the transform V_ψ is an isometry from $\mathcal{H}$ into $L^2(G)$. Another way to express this idea is that the energy of f is conserved by V_ψ. It will then be possible, by restricting to the range of the transform, to build inversion formulas.

The second crucial concept of the voice transform is the characterization of its range. In this case it turns out that $V_\psi(\mathcal{H}) \subseteq L^2(G)$ is a reproducing kernel Hilbert space, meaning in particular that a reproducing formula is valid: $V_\psi(f) = V_\psi(f) * V_\psi(\psi),\ \forall f \in \mathcal{H}$. This reproducing formula can be rephrased by saying that a function $F \in L^2(G)$ belongs to $V_\psi(\mathcal{H})$, if and only if $F = F * V_\psi(\psi)$.

2.2 Discrete Transforms

The discretization of the STFT and wavelet transforms is a field with many aspects that are today very active research topics. The objective is, as in the case of the continuous

environment, to find conditions that ensure adequate analysis and synthesis algorithms. The fact that now we handle a discrete environment introduces additional requirements to be checked for a correct analysis/synthesis functionality. Given a discrete set of functions $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$ to be used in a analysis framework, a crucial aspect to consider is the correct construction of the corresponding set of functions in the synthesis framework, usually denominated the *dual* set $\{\psi^{j,k}\}_{j,k\in\mathbb{Z}}$. In general, given a set of analysis functions $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$, it is sometimes not immediately clear how to construct the corresponding dual $\{\psi^{j,k}\}_{j,k\in\mathbb{Z}}$. The general problem can then be described as the search for adequate functions $\{\psi_{j,k}\}_{j,k\in\mathbb{Z}}$, with a corresponding dual set $\{\psi^{j,k}\}_{j,k\in\mathbb{Z}}$ that can be used for constructing adequate analysis and synthesis procedures:

$$\begin{array}{cc} \text{Analysis} & \text{Synthesis} \\ W_\psi f(j,k) \coloneqq \langle f, \psi_{j,k}\rangle & f = \displaystyle\sum_{j,k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi^{j,k}. \end{array} \tag{2.1}$$

Two aspects that we will now shortly describe are the very active topics of frame theory and Multiresolution analysis, presenting two particular and crucial examples of the previous analysis/synthesis scheme.

Multiresolutions

After the initial understanding of continuous wavelet transforms in the middle of the 1980s, Y. Meyer and S. Mallat proposed an efficient algorithm for constructing wavelet functions. This algorithm consist of a decomposition of $L^2(\mathbb{R})$ in a nested sequence of spaces generated by the *scaling function* or *father wavelet* ϕ. It turns out that a multiresolution is a very natural structure that can be conveniently identified in many applications. Once a scaling function ϕ generating a multiresolution is found, the next step is to construct the wavelet ψ, sometimes denominated in this context as the *mother wavelet*. The crucial aspect is the usage of the relations between ϕ and ψ, explicitly described by the useful multiresolution structure.

Definition 2.3 (Multiresolution)**.** A multiresolution of $L^2(\mathbb{R})$ is defined as a sequence of subspaces of $L^2(\mathbb{R})$, denoted by $\{V_i\}_{i\in\mathbb{Z}}$ with the following properties:

1. $V_i \subset V_j,\ \forall i,j \in \mathbb{Z}, i < j.$
2. $f \in V_j \Rightarrow f(t) \in V_{j+1}.$
3. $f \in V_j \Rightarrow f(t + k/2^j) \in V_j,\ \forall j,k \in \mathbb{Z}.$
4. $\overline{\bigcup_{j\in\mathbb{Z}} V_j} = L^2(\mathbb{R}).$
5. $\bigcap_{j\in\mathbb{Z}} V_j = \{0\}.$
6. $V_0 = \overline{\text{span}\{\phi(x-k), k \in \mathbb{Z}\}}.$

This machinery was used by I. Daubechies in her celebrated discovery of compact supported wavelets [15], developing an important step in multiresolution analysis which today has grown into a theory with many interactions to diverse fields such as multiscale analysis, radial basis functions, diffusion wavelets, image processing, etc.

Frame Theory

The standard concept of a basis is of fundamental importance for analysis and synthesis of arbitrary functions in a given vector space $\mathcal{H}$. Given a basis $\{b_i\}_{i\in\mathbb{Z}}$ of $\mathcal{H}$, the analysis procedure maps an arbitrary element of $\mathcal{H}$ into a set of coefficients in a more controllable environment such as a sequence space. The synthesis algorithm is based on a reconstruction method that uses the vectors $\{b_i\}_{i\in\mathbb{Z}}$ defined in the analysis procedure. In each case, we use the same elements of the basis $\{b_i\}_{i\in\mathbb{Z}}$ as fundamental building blocks. Frame theory takes one step further and, in a same spirit as in the analysis/synthesis framework of Equations 2.1, it generalizes the concept of a basis while introducing additional flexibility for selecting the building blocks. One fundamental difference, is that a elements of a frame do not need to be linearly independent

Definition 2.4 (Frame of a Hilbert space)**.** A *Frame* is defined as a sequence of vectors $\{f_i\}_{i\in\mathbb{Z}}$ in $\mathcal{H}$ if there exist numbers $A, B > 0$ (the *lower* and *upper* frame bounds) such that for all $f \in \mathcal{H}$ we have:

$$A\|f\|^2 \leq \sum_i |\langle f, f_i\rangle|^2 \leq B\|f\|^2.$$

Remark 2.6 (Frame Operators)**.** It is sometimes convenient to rephrase this definition using the language of operators in order to better understand, and usefully related frame theory to basic aspects in operator theory. For a finite frame $\{f_i\}_{i=1}^r$, the first operator that a frame produces is the *analysis operator*, defined as

$$T : \mathcal{H} \to \mathbb{C}^r, \quad (T(f))_i := \langle f, f_i\rangle.$$

The definition 2.4, can be rephrased as that in order $\{f_i\}_{i=1}^r$ to be a Frame, we require T to be injective. A second fundamental operator used is the adjoint operator of T, which plays the role of *synthesis operator*:

$$T^* : \mathbb{C}^r \to \mathcal{H}, \quad T^*((c_i)_{i=1}^r) := \sum_{i=1}^r c_i f_i.$$

The composition of these maps is the *Frame operator*, which naturally links the original Definition 2.4, and the analysis/synthesis framework in Equations 2.1:

$$G = T^*T : \mathcal{H} \to \mathcal{H}, \quad G(f) := \sum_{i=1}^r \langle f, f_i\rangle f_i.$$

The first basic result of this scheme relates the operator G to the frame bounds with the property that $A = \sigma_{\min}(G)$, and $B = \sigma_{\max}(G)$. Namely, the smallest eigenvalue of the operator G is the lower frame bound, and the largest eigenvalue is the upper frame bound. The second result illustrates frame theory as a crucial example of the analysis/synthesis framework in the Equations 2.1:

$$f = GG^{-1}f = G^{-1}\Big(\sum_{i=1}^r \langle f, f_i\rangle f_i\Big) = \sum_{i=1}^r \langle f, f_i\rangle G^{-1}(f_i), \quad \forall f \in \mathcal{H}.$$

In this case, the set of vectors $\{G^{-1}f_i\}_{i=1}^r$, also denoted as $\{\overline{f_i}\}_{i=1}^r$, correspond to the dual frame of $\{f_i\}_{i=1}^r$, exactly in the same spirit as in the Equations 2.1.

3 Reproducing Kernels and Approximation Theory

In the last decades, the concept of kernel, and its related notion of reproducing kernel Hilbert space (RKHS), has played an increasingly important role in a broad range of applied and theoretical fields. Interpolation and approximation theory, signal sampling techniques, solution spaces of PDEs, characterization of the range of integral operators, and modern dimensional reduction and machine learning methods, are examples of important topics where the concept of kernel is of fundamental importance. Some interesting recent developments have revealed a significant interaction of frame theory with function kernels, adding further perspectives to the applicability of frame analysis. In this chapter, we present a short overview of the concept of kernels, reproducing kernel Hilbert spaces, and some selected applications. We focus primarily on interpolation and approximation methods, and the new interactions with frame theory. For the main ideas of these concepts we follow closely [43, 45, 60].

3.1 Reproducing Kernel Hilbert Spaces (RKHS)

Despite the multiple and diverse contexts in which the concept of RKHS appears, the main principles can be cast in one unified framework. As a starting point, there are two important, closely related, concepts: a kernel, and an underlying RKHS. Even though there is a close relationship between them, it is important to distinguish the differences of these points of view. One of the main motivations for defining a kernel is to analyze arbitrary unstructured sets by mapping its elements to a set with some useful structure: in the current case, the target will be a Hilbert space.

Definition 3.1 (Kernel, Feature map, Feature space [60])**.** Given a nonempty set E, a *kernel* is a function $K : E \times E \to \mathbb{R}$, such that a Hilbert space $\mathcal{H}$, and a map $\Phi : E \to \mathcal{H}$, exists with:

$$K(x, y) = \langle \Phi(x), \Phi(y) \rangle_{\mathcal{H}}, \quad \forall x, y \in E. \tag{3.1}$$

The map Φ, and the space $\mathcal{H}$, are denominated *feature map* and *feature space*.

The informal idea behind the concept of the kernel K is to measure and analyze the similarity between the elements of E, a set without any predefined structure, by using the scalar product of the Hilbert space $\mathcal{H}$. Notice that if we use as the scalar field $\mathbb{C}$ instead of $\mathbb{R}$, we have to take care of defining $K(x, y) = \langle \Phi(y), \Phi(x) \rangle$ due to the sesquilinearity of the scalar product in $\mathbb{C}$. There are no special constraints on the feature space $\mathcal{H}$, but as we will see in Proposition 3.16, the interesting candidates are essentially equivalent, and the prototypical examples will be given by reproducing kernel Hilbert spaces.

Definition 3.2 (Reproducing Kernel Hilbert Space)**.** The Hilbert space $\mathcal{H}_K$ of real functions defined on a nonempty set E is a *reproducing kernel Hilbert space* (RKHS), if there exist a map, the *reproducible kernel*, $K : E \times E \to \mathbb{R}$ with:

1. For $K_x : E \to \mathbb{R}$, $K_x(y) \coloneqq K(x, y)$, $y \in E$, we have $K_x \in \mathcal{H}_K$, $\forall x \in E$.
2. Reproducing property:

$$f(x) = \langle f, K_x \rangle, \quad \forall x \in E, \quad \forall f \in \mathcal{H}_K. \tag{3.2}$$

We have, as in the previous definition, an arbitrary nonempty set E as a starting point, but the focus now is on the particular type of Hilbert space $\mathcal{H}_K$, and the set of functions $\{K_x\}_{x\in E}$ used to generate the reproducing property (the crucial characteristic for the applications of this framework). The work that follows is to analyze the relation of these definitions by constructing adequate feature maps Φ, and presenting specific examples of RKHS with a given kernel K.

Remark 3.1 (Symmetric and positive semi-definite properties [43])**.** The reproducing property allows to immediate obtain several basic aspects of reproducing kernels:

1. $K(x,x) \geq 0$, for any $x \in E$.
2. $\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j K(x_i,x_j) \geq 0$, for any $\{x_i\}_{i=1}^{n} \subset E$, $\{\lambda_i\}_{i=1}^{n} \subset \mathbb{C}$.
3. $K(x,y) = \overline{K(y,x)}$, for any $x, y \in E$.

Defining $K_y(x) := K(x,y), x \in \mathcal{H}_K$, and using the equation 3.2, we obtain $K_y(x) = \langle K_y, K_x\rangle$ for all $x \in E$. When setting $x = y$, we obtain

$$K(y,y) = \langle K_y, K_y\rangle = \|K_y\|^2_{\mathcal{H}_K} \geq 0.$$

In a similarly spirit, if we select n points $\{x_i\}_{i=1}^{n} \subset E$, and n complex numbers $\{\lambda_i\}_{i=1}^{n} \subset \mathbb{C}$, when using the relation $K(x_i,x_j) = K_{x_i}(x_j) = \langle K_{x_i}, K_{x_j}\rangle$, we obtain the positive-semidefinite property:

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j K(x_i,x_j) = \sum_{i=1}^{n}\sum_{j=1}^{n} \lambda_i\lambda_j \langle K_{x_i}, K_{x_j}\rangle = \langle \sum_{i=1}^{n} \lambda_i K_{x_i}, \sum_{j=1}^{n} \lambda_j K_{x_j}\rangle \geq 0.$$

The symmetric relation follows directly from $K(x,y) = K_x(y) = \langle K_x, K_y\rangle = \overline{\langle K_y, K_x\rangle} = \overline{K_y(x)} = \overline{K(y,x)}$.

Remark 3.2 (RKHS Prototype I)**.** Given a symmetric positive-definite kernel K, there is a prototypical example of RKHS that can be constructed by generating a vector space with the functions $K_x : E \times E \to \mathbb{R}$, $K_x(y) = K(x,y)$:

$$\mathcal{H}_K := \overline{\text{span}\{K_x : x \in X\}}. \tag{3.3}$$

The scalar product is given by $\langle K_x, K_y\rangle_{\mathcal{H}_K} := K(x,y)$, and the feature map is given by

$$\Phi_K : E \to \mathcal{H}_K, \quad \Phi_K(x) := K_x \tag{3.4}$$

We will see alternative ways of constructing prototypical RKHS with a main result based on the Mercer's theorem 3.14.

One important characteristic of a Hilbert space, equivalent to the reproducing property is the continuity of the *point evaluation functionals* (or *dirac functionals*), namely, that given $x \in E$, the map $f \to f(x)$ is continuous for all $f \in \mathcal{H}$. This fact is a straightforward consequence of the Riesz representation theorem. Remember than in any Hilbert space $\mathcal{H}$, the map $L_g : \mathcal{H} \to \mathbb{R}$, $L_g(f) = \langle f, g\rangle, f \in \mathcal{H}$ is a linear and bounded (continuous) functional for any $g \in \mathcal{H}$. Conversely, the Riesz representation theorem, a fundamental property of Hilbert spaces [17], specifies that for any linear and bounded functional $L : \mathcal{H} \to \mathbb{R}$, there exist a unique vector $g \in \mathcal{H}$ with $L(f) = \langle f, g\rangle$, for any $f \in \mathcal{H}$.

Theorem 3.3 (RKHS and continuous point evaluation functionals)**.** *Let $\mathcal{H}$ be a Hilbert space of real functions defined in a nonempty set E, and let $L_x : \mathcal{H} \to \mathbb{R}$, with*

$$L_x(f) \coloneqq f(x), \tag{3.5}$$

the point evaluation functionals. The linear maps L_x are continuous for any $x \in E$, if and only if $\mathcal{H}$ has a reproducing property, $f(x) = \langle f, K_x \rangle$, for a set $\{K_x\}_{x \in E} \subset \mathcal{H}$, and any $f \in \mathcal{H}$, $x \in E$.

Proof. If the functional L_x is continuous, with the Riesz representation theorem, we have a vector $K_x \in \mathcal{H}$ with the reproducing property $L_x(f) = f(x) = \langle f, K_x \rangle$, for any $x \in E$, and $f \in \mathcal{H}$. Conversely, with the reproducing property, we can construct a bounded linear functional $L_x(f) = \langle f, K_x \rangle$ due to the continuity of the scalar product: that is, the point evaluation functionals are continuous. □

Remark 3.4 (Pointwise, uniform, strong and weak convergence)**.** Another quite particular property of reproducing kernel Hilbert spaces is the fact that the strong convergence implies the pointwise convergence. Remember that for any nonempty set A, and any metric space (M, d), a sequence of mappings $f_n : A \to M$ converges *pointwise (or simply)* to $f : A \to M$, if $f_n(x)$ converges to $f(x)$, for any $x \in A$, namely, $\lim_{n\to\infty} d(f_n(x), f(x)) = 0$. The convergence is called *uniform* if $\lim_{n\to\infty} \sup_{x\in E}(d(f_n(x), f(x))) = 0$. It is clear that uniform convergence implies pointwise convergence, but a classical example, $f_n(x) = x^n, x \in [0, 1]$, shows that the opposite is, in general, not true.

We recall two other important notions of convergence: the strong (or norm) convergence, and the weak convergence. In order to set these definitions we require a normed, (or Banach) space $\mathcal{B}$. The sequence x_n *converges weakly* to x, if $f(x_n)$ converges to $f(x)$ (i.e. $\lim_{n\to\infty} f(x_n) = f(x)$ in $\mathbb{C}$), for every bounded linear operator f in $\mathcal{B}$. If $\mathcal{B}$ is a Hilbert space, with the Riesz representation theorem we can write $\langle x_n, y \rangle \to \langle x, y \rangle$ (in $\mathbb{C}$), for any $y \in \mathcal{B}$. The sequence x_n *converges in norm (or strong)*, if $\|x_n - x\|_{\mathcal{B}}$ converges to 0. Due to the Cauchy Schwarz inequality, it is not difficult to see that strong convergence implies weak convergence (see [43],p18). But the converse is, in general, not true: take for instance any complete orthogonal system $\{\phi_n\}_{n\in\mathbb{N}}$ of a Hilbert space $\mathcal{H}$. The sequence ϕ_n converges to zero weakly (as $\sum_{n\in\mathbb{N}} |\langle \phi_n, f \rangle_{\mathcal{H}}| < \infty, \forall f \in \mathcal{H}$), but it is clear that $\|\phi_n\|_{\mathcal{H}}$ must not converge to 0.

Now, notice that, in general, it is not true that the strong convergence implies pointwise convergence. A standard example is a L^p space, where we can have pointwise divergence in any set of measure 0, without affecting the norm convergence. The interesting issue is that (due to the reproducing property) in a RKHS this counterintuitive fact actually holds.

Proposition 3.5 ([43])**.** *In a RKHS, strong convergence implies pointwise convergence.*

Proof. If for a reproducing kernel Hilbert space $\mathcal{H}$, each point evaluation functional is continuous, we have $|f(x)| \leq \|f\|_{\mathcal{H}} M_x$, and therefore $|f_n(x) - f(x)| = |(f_n - f)(x)| \leq \|f_n - f\|_{\mathcal{H}} M_x$. As the metric in the Hilbert space $\mathcal{H}$ is given by $d(f, g) := \|f - g\|_{\mathcal{H}}$, strong convergence implies pointwise convergence. □

The following elementary, but important, properties concerning the uniqueness of reproducing kernels, characterization of linear subspaces of RKHS, and orthogonal projections on RKHS are straightforward applications of the reproducing property.

Proposition 3.6. *The reproducing kernel K in a RKHS $\mathcal{H}_K$ is unique.*

Proof. Let K' another reproducing kernel of $\mathcal{H}_K$. We set $K_x(y) \coloneqq K(x,y)$, and $K'_x(y) \coloneqq K'(x,y)$. With the reproducing property we obtain: $\|K_x - K'_x\|^2 = \langle K_x - K'_x, K_x - K'_x\rangle = \langle K_x - K'_x, K_x\rangle - \langle K_x - K'_x, K_x\rangle = K_x - K'_x - K_x + K'_x = 0$. □

Proposition 3.7 (Linear subspaces of a RKHS)**.** *Any linear subspace $\mathcal{H}$ of a RKHS $\mathcal{H}_K$ is also a RKHS.*

Proof. Let the point evaluation functionals be $L_x(f) = f(x),\ f \in \mathcal{H}_K,\ x \in E$, and let $L_x|_{\mathcal{H}}$ be their restrictions to $\mathcal{H}$. Using the Proposition 3.3, and the continuity of the $L_x|_{\mathcal{H}}$, we can guarantee a RKHS structure for $\mathcal{H}$. □

Proposition 3.8 (RKHS as a subspace)**.** *If a RKHS $\mathcal{H}_K$ is a linear subspace of a Hilbert space $\mathcal{H}$, the orthogonal projection in $\mathcal{H}_k$ is given by*

$$P_{\mathcal{H}_K}(f)(x) = \langle f, K_x\rangle, \quad f \in \mathcal{H}. \tag{3.6}$$

Proof. For any $f \in \mathcal{H}$, we can write $f = f' + g$, with $f' \in \mathcal{H}_K$, and $g \in \mathcal{H}_K^{\perp}$. We have then $P_{\mathcal{H}_K}(f)(x) = \langle f', K_x\rangle + \langle g, K_x\rangle$. As $K_x \in \mathcal{H}_K$, we obtain $P_{\mathcal{H}_K}(f)(x) = \langle f, K_x\rangle$. □

We will now explore a close and important relation between the reproducing kernel K and a basis of the space $\mathcal{H}_K$ (Theorem 3.11). This relation can be conveniently used for checking whether a given Hilbert space has a reproducing kernel (Remark 3.13). We will see now that in the case of a RKHS, the kernel can be expanded as a product of the elements of the basis, provided that the Hilbert space is topologically separable. Even though many spaces that we are working with are separable, it is important to be able to recognize some important counterexamples.

Remark 3.9 (Separable spaces)**.** A topological space (X, τ) is *separable* if there is a countable dense subset D, namely, $\overline{D} = X,\ D \subset X$. A classical result indicates that a Hilbert space $\mathcal{H}$ is separable if and only if it has a countable orthonormal basis. Indeed, let $\{\psi_i\}_{i\in I} \subset \mathcal{H}$ be a orthonormal set, that is, $\langle \psi_i, \psi_j\rangle = \delta_{ij},\ i,j \in I$, with δ_{ij} the Kronecker delta. We have for any $i, j \in I, i \neq j$

$$\|\psi_i - \psi_j\|^2 = \langle \psi_i - \psi_j, \psi_i - \psi_j\rangle = \langle \psi_i, \psi_i\rangle - \langle \psi_i, \psi_j\rangle - \langle \psi_j, \psi_i\rangle + \langle \psi_j, \psi_j\rangle = 2.$$

If $\mathcal{H}$ is separable with a countable dense set D, and I is not countable, we have a contradiction using the density condition of D. The argument is to consider an injective map from I to D by selecting for every $\phi_i, i \in I$ an element in D. We have then a countable identification which contradicts the hypothesis of a non countable I. Conversely, we can use the countable property of the base field ($\mathbb{C}$ or $\mathbb{R}$) of $\mathcal{H}$ in order to construct a countable dense set given a countable orthonormal basis.

Remark 3.10 (Counterexamples of separable spaces)**.** As a topological concept, the separability of a Banach space depends on the norm that is used. For instance, a classical example of a nonseparable Banach space is the set of bounded operators in a Hilbert space, $B(\mathcal{H})$, with the norm topology, namely, the topology induced by the operator norm $\|T\|_{\text{op}} \coloneqq \sup_{\|x\|_{\mathcal{H}}\leq 1}\|T(x)\|_{\mathcal{H}}$, $T \in B(\mathcal{H})$. Even though this useful topology is a standard one for the vector space $B(\mathcal{H})$, it turns out to be too fine to allow the construction of

countable dense sets. Another standard example of a non-separable topological set can be described when considering sequences of complex numbers $(c_n)_{n=1}^{\infty}, c_n \in \mathbb{C}$. We recall the space $l^p(\mathbb{C}) := \{c = (c_n)_{n=1}^{\infty}, \|c\|_p := (\sum_{n\in\mathbb{Z}} |c_n|^p)^{1/p} < \infty\}$, for $1 \leq p < \infty$, and for $p = \infty$, $l^{\infty}(\mathbb{C}) := \{c = (c_n)_{n=1}^{\infty}, (c_n)$ is a bounded sequence, $\|c\|_{\infty} := \sup_{n\in\mathbb{N}} c_n < \infty\}$. An important subset of l^{∞} is the bounded sequences converging to zero with the $\| \|_{\infty}$ norm, $c_0 := \{c = (c_n), \lim_{n\to\infty} c_n = 0, \|c\|_{\infty} < \infty\}$. Among these examples, the space of bounded sequences, l^{∞}, is the only case of a non-separable space.

If a RKHS $\mathcal{H}_K$ is separable, a main property can now be presented, that not only provides an important characterization of $\mathcal{H}_K$, but also allows to easily identify simple examples and counter-examples of RKHS.

Theorem 3.11 (Basis and the Kernel decomposition in separable RKHS)**.** *Let $\mathcal{H}_K$ be a reproducing kernel Hilbert space. The following characterization of the reproducing kernel holds:*

$$K(x, y) = \sum_{i=1}^{\infty} \psi_i(x)\overline{\psi_i(y)}, \ \forall x, y \in E, \tag{3.7}$$

with $\{\psi_i\}_{i\in\mathbb{N}}$, a countable orthonormal system of $\mathcal{H}_K$, if and only if $\mathcal{H}_K$ is separable.

Proof. If the kernel decomposition 3.7 holds, we can define a function $K_x : E \to \mathbb{R}$ with $K_x := \sum_{i=1}^{\infty} \psi_i(x)\psi_j$ for any $x \in E$, we have then $K_x(y) = K(x, y)$, and with the reproducing property 3.2 we obtain:

$$f(x) = \langle f, K_x\rangle = \sum_{i=1}^{\infty} \langle f, \psi_i(x)\psi_i\rangle = \sum_{i=1}^{\infty} \langle f, \psi_i(x)\rangle\psi_i,$$

for any $f \in \mathcal{H}_K$ and $x \in E$, and therefore the orthonormal system $\{\psi_i\}_{i\in\mathbb{N}}$ is a countable orthonormal basis, that is, $\mathcal{H}_K$ is separable. Conversely, let $\mathcal{H}_K$ be a separable RKHS, with K its reproducing kernel. We can define $K_y(x) := K(x, y)$, with $K_y \in \mathcal{H}_K$, for all $y \in E$. As $\mathcal{H}_K$ is separable, there exist a countable orthonormal basis $\{\psi_i\}_{i\in\mathbb{N}}$ with $K_y = \sum_{i=1}^{\infty} c_i(y)\psi_i$. The coefficients $c_i(y)$ can be computed as scalar products $c_i(y) = \langle K_y, \psi_i\rangle = \overline{\psi_i(y)}$, we obtain then the decomposition 3.7. □

Remark 3.12 (Kernel decomposition is Basis independent)**.** Notice that the only requirement for the set $\{\psi_i\}_{i\in\mathbb{N}}$ in the proof of the decomposition 3.7, is to be a countable orthonormal system. For another countable orthonormal system $\{\phi_i\}_{i\in\mathbb{N}}$, we obtain, under the same hypothesis, the decomposition $K(x, y) = \sum_{i=1}^{\infty} \phi_i(x)\overline{\phi_i(y)}$.

Remark 3.13 (Simple examples and counterexamples of RKHS [62])**.** With the machinery that has been just developed, we can now present elementary examples and counterexamples of RKHS. It is easy to check that the space l^2 is actually a RKHS. Using the expression $|c_n| \leq 1, \forall c \in l^2, \|c\|_{l^2} = 1$, we can verify that the point evaluation functional $L_n(c) := c_n$ for any $c := (c_n)_{n=1}^{\infty} \in l^2$ and $n \in \mathbb{N}$, is continuous with operator norm $\|L_n\|_{op} = 1$. The kernel for this case turns out to be $K(n, m) = \delta_{nm}$.

Conversely, it is not difficult to verify that the important space of square integrable functions $L^2([-\pi, \pi])$, is not a RKHS. A countable orthonormal system $\{\psi_l\}_{l\in\mathbb{N}}$ exist, the

classical option is given by $\psi_l(t) = \sin(tl)/\sqrt{\pi}$, and the space $L^2([-\pi, \pi])$ is therefore separable. We can check, for instance, that the expression

$$K(t, s) \coloneqq \frac{1}{\pi} \sum_{l=1}^{\infty} \sin(tl) \sin(sl),$$

is not always convergent (take for instance $t = s = \pi/2$), rejecting the possibility of a reproducing property. Alternatively, we can analyze the point evaluation functionals, and check that $L_x(f) := f(x), f \in L^2([-\pi, \pi]), x \in [-\pi, \pi]$ is not continuous (for a fixed x, the expression $|f(x)|$ for $\|f\|_2 = 1$ is unbounded). We will see in the following paragraphs more elaborate examples of RKHS.

Mercer's theorem, feature maps and feature spaces

We are now prepared to address an important result in RKHS theory, the Mercer's theorem, which allows to construct a concrete example of a feature map and a feature space, providing a main link between the concepts of kernel, and RKHS (definitions 3.1 and 3.2). Up to know we have worked in a very general setting (a nonempty set E without any particular structure), which is an important requirement when considering the multiple and diverse applications of kernel methods. Now, for technical convenience when setting the framework of the Mercer's theorem, we assume a measurable space (E, μ), with $E \subset \mathbb{R}^n$ and μ a Borel measure. The following step is to construct our main tool, which is an integral operator L_K defined on the space of square integrable functions $L^2_\mu(E) \coloneqq \{f : E \to \mathbb{C}, \int_E |f(x)|^2 \, d\mu(x) < \infty\}$.

The main strategy we will follow, and the intuitive idea behind these new requirements, is to transport the information we have in the rarefied environment given by (E, μ) and K, into the richer structural setting of the linear space $L^2_\mu(E)$ and the linear operator L_K. Once we ensure that the spectral theorem machinery can be applied to the operator L_K, we obtain a useful decomposition for K, which allows to construct a prototypical example of RKHS $\mathcal{H}_K$.

Theorem 3.14 (Mercer's Theorem [45])**.** *Let K be a continuous, symmetric, positive-semidefinite kernel defined in a measurable space (E, μ), with $E \subset \mathbb{R}^n$, closed, and μ a Borel measure. We assume that $\iint_{E^2} K(x, y)^2 \, d\mu(x) \, d\mu(y) < \infty$. Let the integral operator $L_K : L^2(E) \to L^2(E)$ with*

$$L_K f(x) \coloneqq \int_E K(x, y) f(y) \, d\mu(y). \tag{3.8}$$

Then we have the following decomposition

$$K(x, y) = \sum_{i=1}^{\infty} \lambda_i \psi_i(x) \psi_i(y), \tag{3.9}$$

where $\{\lambda_i\}_{i=1}^{\infty}$ and $\{\psi_i\}_{i=1}^{\infty}$ are the eigenvalues and eigenvectors of the operator L_K.

Proof Sketch. The first step in the strategy of the proof is to verify that L_K is a positive, self-adjoint compact operator, we can then construct an orthonormal basis $\{\psi_i\}_{i=1}^{\infty}$ of

$L^2(E)$, consisting of eigenvectors, with corresponding positive eigenvalues λ_i. This basis is then used for building the kernel expansion 3.9. In order to compute the adjoint of L_K, we use $\langle L_H f, g\rangle = \langle f, L_H^* g\rangle$, with:

$$\langle L_H f, g\rangle = \int_E \int_E K(x,y) f(y) \overline{g(x)}\, d\mu(x)\, d\mu(y) = \int_E f(z) \overline{L_H^* g(z)} d\mu(z),$$

by setting, $L_H^* g(z) := \int_E \overline{K(x,z)} g(x)\, d\mu(x) = \int_E K^*(z,x) g(x)\, d\mu(x)$, with $K^*(x,z) := \overline{K(z,x)}$. Using the symmetric property of K, we obtain $L_K = L_K^*$ (see also [36, p91]). With the spectral theorem we obtain the orthonormal basis $\{\psi_i\}_{i=1}^{\infty}$ and corresponding eigenvalues λ_i, used to construct the decomposition 3.9. □

Remark 3.15 (RKHS Prototype II [45])**.** Given a symmetric positive-semidefinite kernel K, the Mercer's theorem allows to construct another prototypical example of a RKHS. Based on the decomposition 3.9, we can construct a feature map $\Phi_\mu : E \to l^2(\mathbb{C})$, with $\Phi_\mu(x) = (\sqrt{\lambda_i}\psi_i(x))_{i=1}^{\infty}$. If the number of nonzero eigenvalues is $N < \infty$, we use the vector space $\mathbb{R}^N$ instead of l^2. As explained in [45], the eigenvectors $\{\psi_k\}$ and eigenvalues $\{\lambda_i\}$ depends on the measure μ, namely, by selecting a different measure, we obtain a different feature map Φ.

We will now verify that the different RKHS prototypes we have obtained so far, and any other that can be constructed, are essentially equivalent. The main idea is to construct an isometry between the first RKHS prototype we described in Remark 3.2, and an arbitrary feature space $\mathcal{H}$ with feature map $\Phi : E \to \mathcal{H}$, and $\langle \Phi(x), \Phi(y)\rangle = K(x,y)$.

Proposition 3.16 (Equivalence of Feature maps [45])**.** *Let E be a nonempty set, $K : E \times E \to \mathbb{R}$ a positive-semidefinite kernel, and Φ an arbitrary feature map with feature space $\mathcal{H}$, that is, $\Phi : E \to \mathcal{H}$, with $\langle \Phi(x), \Phi(y)\rangle_{\mathcal{H}} = K(x,y)$. We define the following vector spaces:*

$$\mathcal{H}_\Phi := \overline{span\{\Phi_x, x \in E\}}, \quad \mathcal{H}_K := \overline{span\{K_x, x \in E\}}, \tag{3.10}$$

with $\Phi_x := \Phi(x)$, and $K_x(y) := K(x,y)$, for any $x, y \in E$. Denoting by $\mathbb{R}^E$ the vector space of real-valued functions on E, we define the linear operator

$$L_\Phi : \mathcal{H} \to \mathbb{R}^E, \quad L_\Phi(v)(x) := \langle v, \Phi_x\rangle_{\mathcal{H}}. \tag{3.11}$$

Then, the restriction $L_\Phi|_{\mathcal{H}_\Phi}$ is an isometry (isometric isomorphism) from $\mathcal{H}_\Phi$ to $\mathcal{H}_K$.

Proof. If $L_\Phi(v) = 0$, we have $\langle v, \Phi_x\rangle = 0, \forall x \in E$. Then, $\ker L_\Phi = \mathcal{H}_\Phi^\perp$, and therefore $L_\Phi|_{\mathcal{H}_\Phi}$ is bijective. Notice now that $L_\Phi(\Phi_y)(x) = \langle \Phi_y, \Phi_x\rangle = K(x,y) = K_y(x)$ for any $x \in E$, therefore L_Φ maps the function Φ_y to K_y, which implies that $\text{span}\{\Phi_x, x \in E\}$ is isomorphic to $\text{span}\{K_x, x \in E\}$. The isometry property follows from $\langle \Phi_x, \Phi_y\rangle_{\mathcal{H}} = K(x,y) = \langle K_x, K_y\rangle_{\mathcal{H}_K}$. □

Remark 3.17 (Applications to Sampling theory [62])**.** With the general framework of RKHS we have just obtained we get a deeper understanding of the important field of sampling theory. Given a RKHS $\mathcal{H}_K$ with kernel $K : E \times E \to \mathbb{R}$, the main component for constructing a sampling procedure is an adequate selection of points $\{t_k\}_{k\in\mathbb{N}} \subset E$, such that $\{K_{t_k}\}_{k\in\mathbb{N}}$ is a complete orthogonal system of $\mathcal{H}_K$. In this context

we have $\|K_{t_k}\|^2 = \langle K_{t_k}, K_{t_k}\rangle = K(t_k, t_k)$, $f(t_k) = \langle f, K_{t_k}\rangle$, and therefore the sampling reconstruction formula is:

$$f(x) = \sum_{k\in\mathbb{N}} \langle f, K_{t_k}\rangle \frac{K_{t_k}(x)}{\sqrt{K(t_k,t_k)}} = \sum_{k\in\mathbb{N}} f(t_k) \frac{K_{t_k}(x)}{\sqrt{K(t_k,t_k)}}, \quad \forall f \in \mathcal{H}_K, \ x \in E. \tag{3.12}$$

We can use this scheme to obtain the well known case of Nyquist-Shannon sampling framework when using B_ω, the space of square integrable functions whose Fourier transform is supported in the interval $[-\omega, \omega]$ (also known as the space of bandlimited functions, or the Paley-Wiener space). This space turns out to be a RKHS with kernel:

$$K(x, y) = \frac{\sin(\omega(x-y))}{\omega(x-y)}. \tag{3.13}$$

Using the function $\operatorname{sinc}(x) := \sin(x)/x$, and the sequence $\{t_k := k\Delta, k \in \mathbb{Z}\}$, for a sampling step Δ, we obtain the corresponding well-known sampling formula:

$$f(x) = \frac{1}{2\omega} \sum_{k\in\mathbb{Z}} f(t_k) \operatorname{sinc}\left(\frac{\omega}{\pi}(x - t_k)\right), \quad \forall f \in B_\omega, \ x \in \mathbb{R}. \tag{3.14}$$

3.2 Elementary Properties in Approximation Theory

The field of approximation theory includes a combination of numerical, functional and linear analysis techniques in order to design computationally feasible methods for function reconstruction. A main problem consist in recovering a function $f : \Omega \to \mathbb{R}$, from a finite set of values $f(x_1), \ldots, f(x_n)$, with $X := \{x_1, \ldots x_n\}$, and $\Omega \subset \mathbb{R}^d$. In order to set an adequate environment for measuring the error behavior, execution speed, and quality of the approximation procedures, the first step is to identify the structure of the spaces where the function f lies. An important scenario in approximation theory is to consider a normed, or Banach space, and use the structural rich possibilities of linear spaces for building efficient approximation methods. But the fundamental definitions concerning the existence and unicity of approximation schemes can already be conveniently presented in the context of a metric space.

Definition 3.3 (Best approximation)**.** Let (M, d) be a metric space, and $U \subset M$. The *best approximation* of $f \in M$ in U is the element $u^\star \in U$ with

$$d(f, u^\star) = d(f, U) \quad \text{for} \quad d(f, U) \coloneqq \inf\{d(f, u), u \in M\} \tag{3.15}$$

The basic questions that have to be addressed are the existence, uniqueness, and construction of best approximations. An elementary property for a subset U to fulfill in order to guarantee the existence of a best approximation, is the compactness. These basic properties can be found in a standard text on Approximation theory (e.g. [58]).

Proposition 3.18 (Existence of best approximation for compact sets)**.** *If $U \subset M$ is a compact set in the metric space (M, d), then for every $f \in M$, there exist a best approximation element.*

Proof. When defining $d := \inf\{d(f,u), u \in M\}$, we can construct a sequence $\{u_k\}_{k\in\mathbb{N}} \subset U$ such that $d(f,u_k) \to d$, when $n \to \infty$. The main question whether there exist an optimum element in U is answered by the compact property of U, which ensures that for any sequence $\{u_k\}$, a limit point $u^\star \in U$ exist. In order to verify that $u^\star$ fulfills the Equation 3.15, we use the triangle inequality $d(f,u^\star) \leq d(f,u_k) + d(u_k,u^\star)$. We can assume that $d(u_k,u^\star) \to 0$, when $n \to \infty$, and therefore $d(f,u^\star) = d$. □

We will refine now the situation by selecting a normed linear space V instead of a metric space when searching for conditions for best approximations.

Proposition 3.19 (Existence of best approximation finite dimensional normed spaces)**.** *Let U a finite dimensional subspace of a normed space V. For every $f \in V$ there exist a best approximation in U.*

Proof. We can construct a compact subset $U_0 := \{u \in U, \|f-u\| \leq \|f-u_0\|\}$, for an arbitrary $u_0 \in U$. The previous Property 3.18, ensures then the existence of a best approximation. □

Up to know we have addressed the case of existence of best approximations. In order to guarantee uniqueness, an important property of the space V to fulfill is the strict convexity of the norm.

Remark 3.20 (Strictly convex norm)**.** A norm $\|\,\|$ in a vector space V is said to be *strictly convex* if $\|\alpha v + \beta w\| < \alpha\|v\| + \beta\|w\|$, for $\alpha+\beta = 1$, $\alpha,\beta \in [0,1]$. (We mean a *convex* norm when $\|\alpha v + \beta w\| \leq \alpha\|v\| + \beta\|w\|$). Due to the sublinearity every norm is convex, but in the case of L^p spaces, the $\|\|_\infty$ and $\|\|_1$ are the only cases that are not strictly convex.

Proposition 3.21 (Uniqueness of best approximation)**.** *If U if a subspace of a strictly convex normed space V, then every element $f \in V$, has at most one best approximation.*

Proof. Let $f \in V$, and assume that we have two different best approximation elements $u_1, u_2 \in U$, with $d := \|f-u_1\| = \|f-u_2\|$, we can use the strict convexity property to compute:

$$\|f - \frac{1}{2}(u_1+u_2)\| < \frac{1}{2}\|f-u_1\| + \frac{1}{2}\|f-u_2\| = d, \tag{3.16}$$

which contradicts the main assumption, and therefore at most one best approximation element exist. □

Now, we are going to refine even more our framework, and consider a pre-Hilbert space $\mathcal{H}$ as a main environment for characterizing best approximations.

Proposition 3.22 (Best approximations in pre-Hilbert spaces)**.** *Let $\mathcal{H}$ be pre-Hilbert space, and $U \subset \mathcal{H}$ a linear subspace. The element $u^\star \in U$ is a best approximation of an element $f \in \mathcal{H}$, if and only if:*

$$\langle f - u^\star, v\rangle = 0, \quad \forall v \in U. \tag{3.17}$$

Proof. Assuming that an element $u^\star \in U$ fulfills the orthogonality Equation 3.17, we have with the Pythagoras theorem, for any $u \in U$:

$$\|f-u\|^2 = \|(f-u^\star)+(u^\star-u)\| = \|f-u^\star\| + \|u^\star-u\| > \|f-u^\star\|. \tag{3.18}$$

In other words, we have ensured that $u^\star$ is a best approximation of f.

Now, assuming that the orthogonality Equation 3.17 does not hold for some $v \in U$, we can select a $\lambda := -\langle f-u^\star, v\rangle / \|v\|^2$, and compute:

$$\begin{aligned}\|f-u^\star+\lambda v\|^2 &= \|f-u^\star\| + 2\lambda\langle f-u^\star, v\rangle + \lambda^2\|v\|^2 \\ &< \|f-u^\star\|^2.\end{aligned}$$

In other words, $u^\star$ cannot be a best approximation. □

RKHS and approximation theory

With these elementary background ideas in approximation theory, we can now address an important goal: the possibility of constructing approximation schemes using an RKHS as a main tool. We follow closely the presentation in [56], and refer to [34, 54, 55], for further details.

Remark 3.23 (RKHS best approximation scheme)**.** The plan of the approximation procedure based on RKHS is as follows. The first component of this framework is a (unknown) function $f : \Omega \to \mathbb{R}$ to be reconstructed from a set of (known) values $f(x_1), \ldots, f(x_n)$, based on the sampling elements $X = \{x_1, \ldots, x_n\} \subset \Omega \subset \mathbb{R}^d$. The second component is a RKHS $\mathcal{H}_K$ with kernel $K : \Omega \times \Omega \to \mathbb{R}$, and the hypothesis that $f \in \mathcal{H}_K$. Now, the core idea is to construct a finite dimensional linear subspace of $\mathcal{H}_K$ that will be used as an approximation tool:

$$S_X := \text{span}\{K_x,\ x \in X\},\ \text{with } K_x : \Omega \to \mathbb{R},\ K_x(y) := K(x,y). \tag{3.19}$$

With these ingredients, we can directly apply the basic framework of approximation theory in order to construct efficient reconstruction procedures. From an application point of view, an important step is to carefully select a meaningful kernel K, with corresponding RKHS $\mathcal{H}_K$. We will now see two important properties of this scheme: the uniqueness and optimality of the best approximation procedures.

Theorem 3.24 (Approximation in RKHS [56])**.** *Let $\mathcal{H}_K$ be a RKHS, with kernel $K : \Omega \times \Omega \to \mathbb{R}$, and $\Omega \subset \mathbb{R}^d$. Let $f \in \mathcal{H}_K$ a function, and let the sampling values $f(x_1), \ldots, f(x_n)$ defined over the finite set $X := \{x_1, \ldots, x_n\} \subset \Omega$. Then there exist a best approximation element $f_X^* \in S_X := span\{K_x, x \in X\}$, where its coefficients can be found by solving the linear system:*

$$f_X^*(x_k) = \sum_{i=1}^{n} a_i^* K(x_i, x_k) = f(x_k), \quad 1 \le k \le n. \tag{3.20}$$

Proof. As $S_X \subset \mathcal{H}_K$ is a finite Hilbert space, with the Propositions 3.19 and 3.22, we ensure that there exist a best approximation $f_X^* \in S_X$ of $f \in \mathcal{H}_K$, with

$$\langle f - f_X^*, s\rangle = 0, \quad \forall s \in S_X. \tag{3.21}$$

We have, in particular, with the reproducing property (Equation 3.2), and the function $K_{x_k}(y) \coloneqq K(x_k, y)$:

$$0 = \langle f - f_X^*, K_{x_k} \rangle = f(x_k) - f_X^*(x_k), \quad 1 \leq k \leq n. \tag{3.22}$$

Being f_X^* an element of S_X, we can now obtain the Equation 3.20. □

Theorem 3.25 (Optimal approximation in RKHS [56])**.** *Under the conditions of the Theorem 3.24, we have, additionally, an optimality condition of the interpolant f^*, namely:*

$$\min_{g \in O_X(f)} \|g\|_K = \|f^*\|_K, \qquad O_X(f) \coloneqq \{g \in S_X,\ g|_X = f|_X\}.$$

Proof. For any $g \in O_X(f)$, we can write with the law of cosines:

$$\|g\|_K^2 = \|f_X^* + (g - f_X^*)\|_K^2 = \|f_X^*\|^2 + 2\langle f_X^*, g - f_X^* \rangle + \|g - f_X^*\|^2. \tag{3.23}$$

The scalar product in this equation can now be analyzed by considering the functions $K_{x_i}(y) \coloneqq K(x_i, y), y \in X$:

$$\begin{aligned} \langle f_X^*, g - f_X^* \rangle &= \langle \sum_{i=1}^n a_i^* K_{x_i}, g - f_X^* \rangle = \sum_{i=1}^n a_i^* \langle K_{x_i}, g - f_X^* \rangle \\ &= \sum_{i=1}^n a_i^* (g(x_i) - f_X^*(x_i)) \end{aligned}$$

As $f_X^*(x_i) = f(x_i)$, and $g(x_i) = f(x_i)$, for $i = 1, \ldots, n$, we have $\langle f_X^*, g - f_X^* \rangle = 0$. Therefore, using the Equation 3.23, we obtain $\|g\| \geq \|f_X^*\|$, for all $g \in O_X(f)$. □

3.3 Kernels and Frames

In this final section we reach or main objective: connecting the background ideas we have seen in the last two chapters, by presenting a conceptual interaction between frame theory and reproducing Kernel Hilbert spaces. These results have been reported in the literature over the last years [48, 51], and its development and further consequences are current active research topics.

One of the main objectives of this framework is to use the additional flexibility of a frame over a basis in a Hilbert space in order to extend the choice for construction of kernels and RKHS tailored to a given problem and application.

Theorem 3.26 (RKHS from a framable Hilbert space [51])**.** *Let $\mathcal{H}$ be a Hilbert space of functions over $\Omega \subset \mathbb{R}^d$, with a frame $\{\phi_i\}_{i \in I} \subset \mathcal{H}$. (I being a finite or infinite index set) Let $K_x(y) \coloneqq \sum_{i \in I} \overline{\phi_i}(y)\phi_i(x)$, (here $\{\overline{\phi_i}\}_{i \in I}$ is the dual frame of $\{\phi_i\}_{i \in I}$). If the following condition holds*

$$\|K_x\|_{\mathcal{H}} < \infty, \quad \forall x \in \Omega,$$

then $\mathcal{H}$ is a reproducing kernel Hilbert space.

Proof. The strategy of the proof consist of using the frame analysis property in order to verify the continuity condition of point evaluation functionals (Theorem 3.3), together with a straightforward application of the the Cauchy Schwarz inequality.

Due to the frame analysis property, we have:

$$f = \sum_{i \in I} \langle f, \overline{\phi_i} \rangle \phi_i, \quad \forall f \in \mathcal{H}. \tag{3.24}$$

The scalar product used here is defined in $\mathcal{H}$, and to make more precise this statement, we use the following seminorn in $\mathcal{H}$: $\|f\|_x = |f(x)|$ for $x \in \Omega$, which allows to rephrase the Equation 3.24 equivalently as a pointwise convergence $f(x) = \sum_{i \in I} \langle f, \overline{\phi_i} \rangle \phi_i(x)$, for all $x \in \Omega$. Now, this frame analysis property can also be written as:

$$f(x) = \langle f, \sum_{i \in I} \overline{\phi_i} \phi_i(x) \rangle, \quad \forall x \in \Omega, \tag{3.25}$$

and using the Cauchy Schwarz inequality, we obtain:

$$|f(x)| = |\langle f, \sum_{i \in I} \overline{\phi_i} \phi_i(x) \rangle| \leq \|f\| \ \ \| \sum_{i \in I} \overline{\phi_i} \phi_i(x) \|. \tag{3.26}$$

As the main hypothesis of this theorem is that $K_x = \sum_{i \in I} \overline{\phi_i} \phi_i(x)$ is bounded, we have a bounded linear operator $L_x(f) := f(x)$, which implies that $\mathcal{H}$ is a RKHS due to the Theorem 3.3. □

Theorem 3.27 (Reproducing Kernel of a framable RKHS [51])**.** *If $\mathcal{H}_K$ is a reproducing kernel Hilbert space of functions over a $\Omega \subset \mathbb{R}^d$, which contains a frame $\{\phi_i\}_{i \in I}$, then the reproducing kernel can be written as:*

$$K(x, y) = \sum_{i \in I} \overline{\phi_i}(x) \phi_i(y). \tag{3.27}$$

Proof. On the one hand, in a similar spirit as in the previous Theorem 3.26, any function in $\mathcal{H}_K$ can be written (thanks to the frame property) as:

$$f(x) = \langle f, \sum_{i \in I} \overline{\phi_i} \phi_i(x) \rangle, \quad \forall x \in \Omega, \quad \forall f \in \mathcal{H}_K. \tag{3.28}$$

On the other hand, since $\mathcal{H}_K$ is a reproducing kernel Hilbert space, we have a kernel K, with

$$f(x) = \langle f, K_x \rangle, \quad \forall x \in \Omega, \quad \forall f \in \mathcal{H}_K. \tag{3.29}$$

Now, due to the unicity of the reproducing kernel proved in Proposition 3.6, we obtain the kernel decomposition based on a frame $\{\phi_i\}_{i \in I}$ of Equation 3.27. □

4 Dimensional Reduction and Manifold Learning

In the last decades there has been an important demand of new analysis tools for dealing with the remarkable increase of raw data in many scientific fields. More efficient computer hardware and the extensive growth of storage capacity allows to easily record, transmit and store scientific data in diverse fields such as medicine, geography, social sciences, physics, chemistry and biology, to name just a few. At the same time, the exponential increase of processing power allows to design new approaches in data analysis, intertwining abstract mathematical tools with numerical techniques. In this context, a significant development in the last years has been observed under the name of *dimensionality reduction* and *manifold learning*, whose objective is to analyze a high dimensional dataset in order to extract a low dimensional parameterization. The main emphasis in these modern tools is on the usage of geometrical and analytical based approaches, which can be considered as a complementary strategy to more classical statistical oriented methods used in machine learning and data mining.

In the following sections we present elementary dimensional reduction techniques, including the well known Principal Component Analysis (PCA) method, and the alternative strategy of Multidimensional Scaling (MDS). We then mention modern frameworks for handling non-linear manifolds, focusing in particular on the recent Isomap algorithm, and its crucial usage of kernel techniques. This overview indicates a natural link with our previous chapter, while completing the necessary background for the following chapter, where an illustrative algorithm is presented combining signal processing techniques with dimensional reduction methods.

4.1 Basic Definitions

A standard way for representing experimental information is given by the concept of *Point Cloud Data*, defined as a finite set of vectors $X = \{x_k\}_{k=1}^m \subset \mathbb{R}^n$. As we will now see, the frameworks we present have a similar objectives, but contrary to the plain *dimensional reduction* scheme, the additional hypothesis in *manifold learning* allows to introduce more refined tools from differential geometry.

Remark 4.1 (Dimensional Reduction)**.** The *dimensional reduction* problem considers the case when much of the information described by X is redundant and can be discarded by constructing a low dimensional representation $Y = \{y_k\}_{k=1}^m \subset \mathbb{R}^d$, with $d \leq n$.

The main objective is to design a dataset Y in a way that certain desired characteristics of X are conserved. For instance, the strategy in *Multidimensional scaling* (MDS) is to build Y in a way that the distances between the points in X are conserved. Namely, we search for Y with $\|y_i - y_j\| \approx \|x_i - x_j\|$ for all $i, j \in \{1, \ldots, m\}$. Another example is the case when the dataset X lies in the vicinity of an hyperplane in $\mathbb{R}^n$: for this situation the strategy of *Principal Component Analysis* is to construct Y by projecting the set X in the hyperplane. The reduced dataset Y can then be used for analysis or classification purposes, allowing, for instance, to run more efficiently different numerical algorithms.

Remark 4.2 (Manifold Learning)**.** In *manifold learning*, the main hypothesis is that the elements x_i are vectors lying in (or are close to) a manifold $\mathcal{M}$. We consider the case $X \subset \mathcal{M}$, namely, X is sampled from $\mathcal{M}$, a p-dimensional smooth compact

submanifold of $\mathbb{R}^n$. The existence of this manifold is a reasonable hypothesis often fulfilled in many applications. As in dimensionality reduction, the objective is also to construct a low dimensional representation $Y = \{y_1, \ldots, y_m\} \subset \mathbb{R}^d, d \leq n$ that conserves some characteristics of the dataset X, but now, the geometrical environment introduced by $\mathcal{M}$ will play a crucial role in the algorithm design.

Due to the Whitney embedding theorem (which states that any connected smooth p-dimensional manifold can be smoothly embedded in $\mathbb{R}^{2p+1}$ [18]) we require some conditions for the dimensions in this formulation, namely, $2p + 1 \leq d \leq n$. Our problem can also be formulated as the search for an adequate embedding E of the p-dimensional submanifold $\mathcal{M} \subset \mathbb{R}^n$ in $\mathbb{R}^d$, with $E : \mathcal{M} \subset \mathbb{R}^n \to \Omega \subset \mathbb{R}^d$, $X \subset \mathcal{M}$, $Y \subset \Omega$, Ω a p-dimensional submanifold, and $2p + 1 \leq d \leq n$.

An important additional topic in this field are density conditions on the finite data set X with respect to $\mathcal{M}$, which need to be analyzed in order to guarantee a meaningful usage of the geometry of $\mathcal{M}$: recent results for ensuring the correct computation of the homology of a manifold from finite random sampled data have been presented in [47]. In a similar spirit, *persistent topology* is an algorithm developed for computing homological information of $\mathcal{M}$ using as input the finite sample $\{x_k\}_{k=1}^m$ [66].

4.2 Elementary techniques

Elementary techniques in the manifold learning and dimensional reduction framework have already been developed over the last century, with the *principal component analysis* as a core method with a long history of applications in multiple fields. A main characteristic of this algorithm is to express the dimensional reduction framework as an eigenproblem involving the singular value decomposition, and underlying the primal role of spectral concepts in this field. This procedure includes a direct construction of a linear projection using a selected set of eigenvectors. Another important dimensional reduction method, using an alternative strategy to the PCA technique, is Multidimensional scaling (MDS), which also uses as a core component special eigenvalues and eigenvectors. But the internal machinery in MDS is significantly different: contrary to the case of PCA, the MDS scheme constructs a low dimensional configuration without considering an explicit projection map.

Principal Component Analysis (PCA)

Principal Component Analysis (PCA) is a technique with multiple origins and different applications developed since the beginning of the last century [49]. Equivalent formulations arising from different backgrounds are the Karhunen-Loève, and the Hotelling transforms. The core component of this algorithm is one of the basic tools in the dimensional reduction framework, and its main input, as in any standard reduction algorithm, is a set of vectors $X = \{x_k\}_{k=1}^m \subset \mathbb{R}^n$, represented also (by abuse of notation) in matrix form as $X = (x_1, \ldots, x_m) \in \mathbb{R}^{n \times m}$. Two standard ways of formulating the PCA algorithm is as an optimization problem or in a linear algebra setting. The PCA transform can be described as a search for a hyperplane H minimizing the sum of the distances from the elements of $\{x_k\}_{k=1}^m$ to their projections in H. An equivalent formulation is the search for a hyperplane H maximizing the variance of the set of projections of the elements of $\{x_k\}_{k=1}^m$ in H. These formulations can be equivalently described in the context of the

singular value decomposition. A fourth possible equivalent context is the manipulation of the eigenvectors of the covariance matrix XX^t.

The Figure 1 represents the general idea in PCA transforms. Given the data set $\{x_k\}_{k=1}^m$, the task is to find an hyperplane in such a way that the projections $\{P(x_k)\}_{k=1}^m$ are distributed with maximum variance, or equivalently, the sum of the distances between the points and their projections, $\sum_{i=1}^m \|x_i - P(x_i)\|$, is minimum. The solution is given by the hyperplane specified by the eigenvectors of the matrix XX^t, which can also be described with the singular value decomposition of X.

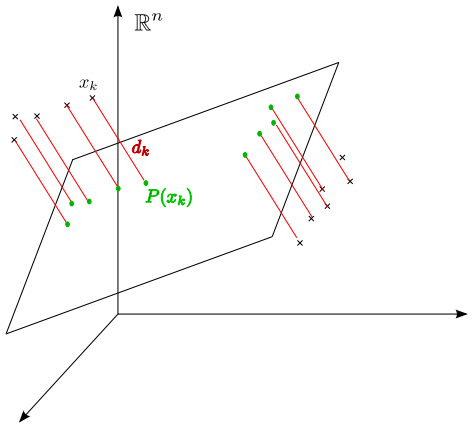


Figure 1: PCA main idea

Theorem 4.3 (Principal component analysis)**.** *Let $X = (x_1 \dots x_m) \in \mathbb{R}^{n\times m}$ be a matrix with vector entries representing a centered (or zero-mean) dataset (i.e. $\frac{1}{m}\sum_{k=1}^m x_k = 0$). The following characterizations for a projection $P : \mathbb{R}^n \to \mathbb{R}^n$, rank$(P) = p < n$, are equivalent:*

(1) err$(P, X) := \sum_{k=1}^m \|x_k - P(x_k)\|^2$ is minimum.

(2) var$(P(X)) := \sum_{k=1}^m \|P(x_k)\|^2$ maximum.

(3) The matrix representation of P is composed by the p eigenvectors corresponding to the p largest eigenvalues of the covariance matrix $XX^t \in \mathbb{R}^{n\times n}$.

Proof. The equivalence between *(1)* and *(2)* is a direct consequence of the orthogonality property of the linear projection P, and the Pythagoras theorem relating the norm of the vectors x_k, $P(x_k)$, and $x_k - P(x_k)$:

$$\|x_k\|^2 = \|x_k - P(x_k)\|^2 + \|P(x_k)\|^2, \quad k = 1, \dots, m. \tag{4.1}$$

We obtain, as a consequence, the following relation between err(P, X), and var$(P(X))$:

$$\text{err}(P, X) = \sum_{k=1}^m \|x_k\|^2 - \text{var}(P(X)). \tag{4.2}$$

With this expression, it is clear that a minimization of err(P, X) is equivalent to a maximization of var$(P(X))$. Now, in order to study the property *(3)*, we analyze in more

detail the expression $\text{var}(P(X)) = \sum_{k=1}^{m} \|P(x_k)\|^2$. By considering an orthonormal basis $\{v_i\}_{i=1}^{p} \subset \mathbb{R}^n$ of the Euclidean space $\mathbb{R}^p$, we have $P(x_k) = \sum_{i=1}^{p} \langle v_i, x_k \rangle v_i$, and therefore, due to the Pythagoras theorem, we obtain:

$$\sum_{k=1}^{m} \|P(x_k)\|^2 = \sum_{k=1}^{m} \sum_{i=1}^{p} |\langle v_i, x_k \rangle|^2. \tag{4.3}$$

By exchanging the sums, and recalling that we use the scalar product over the field $\mathbb{R}$, we can reformulate the term $\sum_{k=1}^{m} |\langle v_i, x_k \rangle|^2$ as

$$\sum_{k=1}^{m} |\langle v_i, x_k \rangle|^2 = \sum_{k=1}^{m} \langle v_i, x_k \rangle \langle x_k, v_i \rangle = \langle v_i, \sum_{k=1}^{m} x_k \langle x_k, v_i \rangle \rangle = \langle v_i, \sum_{k=1}^{m} x_k \sum_{j=1}^{n} x_{jk} v_{ji} \rangle. \tag{4.4}$$

Now, by considering the matrix $S := XX^t = \left(\sum_{k=1}^{m} x_{lk} x_{jk} \right)_{l,j=1}^{n}$, we have:

$$(Sv_i)_l = \sum_{j=1}^{n} \sum_{k=1}^{m} x_{lk} x_{jk} v_{ji} = \sum_{k=1}^{m} x_{lk} \sum_{j=1}^{n} x_{jk} v_{ji}. \tag{4.5}$$

With the Formulas 4.4 and 4.5, we obtain $\sum_{k=1}^{m} |\langle v_i, x_k \rangle|^2 = \langle v_i, Sv_i \rangle$, and therefore:

$$\text{var}(P(X)) = \sum_{k=1}^{m} \|P(x_k)\|^2 = \sum_{i=1}^{p} \langle v_i, Sv_i \rangle. \tag{4.6}$$

Finally, the maximization of $\sum_{i=1}^{p} \langle v_i, Sv_i \rangle$ occurs, due to the Lemma 4.4, when $\{v_i\}_{i=1}^{p}$ are the p eigenvectors, corresponding to the largest p eigenvalues of the matrix S. □

Lemma 4.4. *The function $W(v) := \langle v, Sv \rangle, v \in \mathbb{R}^n, \|v\| = 1$, with $S \in \mathbb{R}^{n \times n}$, a real symmetric matrix, is maximized when v is the eigenvector corresponding to the largest eigenvector of S.*

Proof. The symmetric property of the matrix S implies the existence of a set of eigenvalues $\{\lambda_i\}_{i=1}^{n}$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, and a corresponding orthonormal set of eigenvectors $\{e_i\}_{i=1}^{n}$. For any $v \in \mathbb{R}^n, \|v\| = 1$, we have $v = \sum_{i=1}^{n} \alpha_i e_i$, and $\sum_{i=1}^{n} \alpha_i^2 = 1$. Therefore, $W(v) = \langle v, Sv \rangle = \sum_{i=1}^{n} \langle \alpha_i e_i, \alpha_i \lambda_i e_i \rangle = \sum_{i=1}^{n} \alpha_i^2 \lambda_i$. It follows that $W(v) \leq \lambda_1$, which is an equality only for $v = e_1$. □

Remark 4.5 (SVD and PCA)**.** It is an important and easy observation the fact that the PCA transform is closely related to the Singular Value Decomposition (SVD) of X. Namely, for any $X \in \mathbb{R}^{n \times m}$, the SVD is a factorization $X = UDV^t$, with $D \in \mathbb{R}^{n \times m}$ a diagonal matrix with the *singular values*, and $U \in \mathbb{R}^{n \times n}$, $V \in \mathbb{R}^{m \times m}$ unitary matrices. We obtain the expression $XX^t = U(DD^t)U^t$, using the computation of the SVD of X, in order to obtain the eigendecomposition of the covariance matrix XX^t.

Independent Component Analysis (ICA)

Even though the spirit of the algorithm we describe in this section does not directly belongs to the dimensional reduction framework, its frequent use in combination with

PCA, makes it an important component in the field of data analysis and signal separation. For completeness, we present in this section a very short description of its core ideas, and relation to the dimensional reduction framework.

In the late 1980s the signal processing community presented an algorithm for signal separation known as Independent Component Analysis. The main idea uses the crucial concept of independent random variables, and represented a major milestone in blind source separation techniques. The importance of this algorithm was quickly recognized, and multiple applications were implemented in the area of audio analysis, telecommunications, biomedical signal processing, electrical engineering, and many others. Today, several generalizations and new applications of this framework are mainstream research topics.

The input of the ICA algorithm is a *Point cloud data*, defined as a finite sequence of vector values, written in matrix form as $X = (x_1 \dots x_m) \in \mathbb{R}^{n\times m}$. The objective is to find a sequence of source signals $S = (s_1 \dots s_m) \in \mathbb{R}^{n\times m}$, assuming a linear dependence between X and S. By denoting the *mixing matrix* as $W \in \mathbb{R}^{n\times n}$, this property can be expressed as:

$$X = WS \quad \text{with} \quad X = (x_1 \dots x_m), \quad \text{and} \quad S = (s_1 \dots s_m).$$

In this equation, the mixing matrix W and the source signals S are the unknown variables to be found with the ICA procedure. The second core assumption of the ICA algorithm is the statistical independence of the signals $\{s_i\}_{i=1}^n$. In order to resolve this problem, a general strategy can be described with the following measure for a set of random variables $Y = \{y_i\}_{i=1}^n$:

$$I(Y) = D(P_Y, \prod_i P_{Y_i}) \quad \text{with} \quad D(f,g) = \int_{\mathbb{R}} f(x) \log\left(\frac{f(x)}{g(x)}\right) dx.$$

The measure I allows to compute the degree of statistical independence by comparing the joint distribution P_Y, and the marginal distributions P_{Y_i}. The comparison function D, used in the measure I, is the *Kullback-Leibler distance*, also known as *relative entropy*. The function I allows to express the ICA algorithm as an optimization problem, where the solution space is the *General Linear Group*, defined as the set of $n \times n$ invertible matrices: $GL(n,\mathbb{R}) = \{A \in \mathbb{R}^{n\times n},\ \det(A) \neq 0\}$, with $f(A) := I(A^{-1}X)$, and $X = (x_1 \dots x_m) \in \mathbb{R}^{n\times m}$:

$$\min f(A) \qquad \text{with} \qquad A \in GL(n,\mathbb{R})$$

Multidimensional Scaling (MDS)

Another important tool in the dimensional reduction framework is a technique with a long tradition of over 80 years, and extensive applications, denominated Multidimensional scaling (MDS) [14]. The objective is similar as in the PCA scheme, but its internal strategy is significantly different, and the alternative formulations it includes will be of particular importance when designing important generalizations as we will see with the Isomap algorithm.

Given a dataset $X = (x_1, \dots, x_m) \in \mathbb{R}^{n\times m}$, the MDS algorithm looks for a configuration of points $Y = (y_1, \dots, y_m) \in \mathbb{R}^{p\times m}$, with $p \leq n$, in such a way that a given

relationship between the data elements $\{x_i\}_{i=1}^m$ is conserved. A prototypical example of such a relation is the distance function, but a different type of similarity measures might be considered. An important difference with the PCA framework is that the MDS does not find an explicit projection map, instead, an optimization procedure is implemented in a way that an optimal configuration in the low dimensional space is found. The problem of finding an explicit function mapping the elements $\{x_i\}_{i=1}^m$ to their low dimensional representation $\{y_i\}_{i=1}^m$, can be considered as a separate task, where particular interpolation and approximation techniques can play an important role.

Definition 4.1 (Distance Multidimensional Scaling (Distance MDS))**.** Let a matrix data $X \in \mathbb{R}^{n\times m}$ and its distance relationships d_{ij} be given as:

$$X = (x_1, \ldots, x_m) \in \mathbb{R}^{n\times m}, \qquad d_{ij} := \|x_i - x_j\|, \quad i, j = 1, \ldots, m. \tag{4.7}$$

The *multidimensional scaling* problem is defined as the search for a configuration of points $Y \in \mathbb{R}^{p\times m}$, with $p \leq n$, that minimizes the error function $\text{err}(Y, \{d_{ij}\}_{i,j=1}^m)$ with:

$$Y = (y_1, \ldots, y_m) \in \mathbb{R}^{p\times m}, \qquad \text{err}(Y, \{d_{ij}\}_{i,j=1}^m) := \sum_{i,j=1}^m (d_{ij} - \|y_i - y_j\|)^2. \tag{4.8}$$

In other words, the lower dimensional configuration of points $\{y_i\}_{i=1}^m \subset \mathbb{R}^p$ preserves the distance relationships of the higher dimensional dataset $\{x_i\}_{i=1}^m \subset \mathbb{R}^n$.

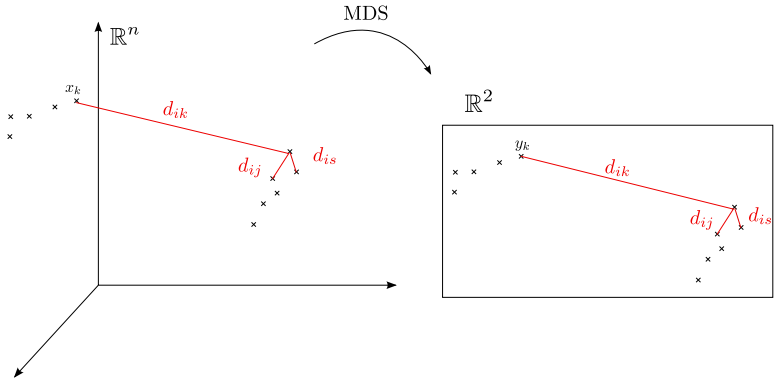


Figure 2: Distance MDS concept: preserve distance relations

Besides this original optimization problem, there are several important reinterpretations of the MDS that can be used for designing alternative algorithmic solutions. An important formulation translates the MDS scheme to an eigenproblem by relating the distance between two vectors with their scalar product. In order to differentiate these perspectives, the formulation that uses the distance and the optimization problem previously defined is sometimes denominated *distance MDS*. The eigenproblem formulation that follows is sometimes denominated *classical MDS*, and it has been used to design important generalizations, as we will see in the definition 4.10 of the continuum Isomap. We will now describe in more detail the eigenproblem formulation, which allows to design one possible solution strategy for the MDS framework.

Theorem 4.6 (MDS Eigenproblem formulation (Classical MDS) [65]). *Let $D = (d_{ij}^2)_{i,j=1}^m$, be the symmetric square distance matrix with $d_{ij} := d(x_i, x_j) = \|x_i - x_j\|$, and $X = (x_1, \ldots, x_m) \in \mathbb{R}^{n \times m}$, which we assumed to be centered (that is $\sum_{i=1}^m x_i = 0$).*

The matrix X can be recovered, up to an orthogonal transformation, by considering

$$Y = diag(\lambda_1^{1/2}, \ldots, \lambda_m^{1/2}) U^t,$$

where $\lambda_1 \geq \cdots \geq \lambda_m \geq 0$, and $U \in \mathbb{R}^{m \times n}$ are the eigenvalues and the corresponding eigenvectors of the matrix B with

$$B \coloneqq -\frac{1}{2} JDJ, \quad with \quad J := I - (1/m) ee^t, \quad e := (1, \ldots, 1)^t \in \mathbb{R}^m. \tag{4.9}$$

Proof. The core idea of this construction is to use a straightforward relation between the matrix of scalar products XX^t, and the given matrix of square distances D. This relation can in turn be used to compute X up to an orthogonal transformation.

The distance and the scalar product between two vectors x_i and x_j can be related as

$$d^2(x_i, x_j) = \|x_i - x_j\|^2 = \langle x_i, x_i \rangle - 2\langle x_i, x_i \rangle + \langle x_j, x_j \rangle. \tag{4.10}$$

Using $A := [\langle x_1, x_1 \rangle, \ldots, \langle x_m, x_m \rangle]^t$, the expression 4.10 can then be used to construct the matrix relation $D = Ae^t - 2XX^t + eA^t$. In order to recover the matrix of scalar products XX^t, we use the $m \times m$ square matrix J with the value $1 - 1/n$ in the main diagonal, and $-1/n$ elsewhere, that is: $J := I_m - (1/m)ee^t$. We notice that $Je = 0$ (as $1 - 1/m + (m-1)(-1/m) = 0$), and therefore we obtain a relation between the matrix of scalar products XX^t, and the given matrix of square distances D:

$$XX^t = -\frac{1}{2} JDJ.$$

The eigendecomposition of the matrix $B := -\frac{1}{2} JDJ = U \mathrm{diag}(\lambda_1, \ldots, \lambda_m) U^t$, allows to construct X, up to an orthogonal transformation $Y = \mathrm{diag}(\lambda_1^{1/2}, \ldots, \lambda_m^{1/2}) U^t$. □

Remark 4.7 (Entries of the matrix B 4.9 [14]). Each entry of the matrix B can be explicitly formulated with the following expressions, derived from the relation between the distance and the scalar product (formula 4.10), and the zero-mean property ($\sum_{i=1}^m x_i = 0$):

$$\frac{1}{m} \sum_{i=1}^m d_{ij}^2 = \frac{1}{m} \sum_{i=1}^m \langle x_i, x_i \rangle + \langle x_j, x_j \rangle,$$
$$\frac{1}{m} \sum_{j=1}^m d_{ij}^2 = \langle x_i, x_i \rangle + \frac{1}{m} \sum_{j=1}^m \langle x_j, x_j \rangle.$$

By combining these two relations we have:

$$\frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m d_{ij}^2 = \frac{2}{m} \sum_{i=1}^m \langle x_i, x_i \rangle.$$

We obtain, finally, an expression for the entries of the matrix B:

$$B_{ij} = \langle x_i, x_j \rangle = -\frac{1}{2} \left(d_{ij}^2 - \frac{1}{m} \sum_{i=1}^m d_{ij}^2 - \frac{1}{m} \sum_{j=1}^m d_{ij}^2 + \frac{1}{m^2} \sum_{i,j=1}^m d_{ij}^2 \right). \tag{4.11}$$

This formula will play an important role when generalizing important extensions of the MDS framework, as we will see in the following sections with the continuum Isomap concept.

4.3 Isomap and Kernel Techniques

There are several ways in which the fundamental techniques of PCA and MDS can be generalized for dealing with non-linear manifolds. An important strategy is the context of kernel techniques, which uses the fundamental tools of Reproducing Kernel Hilbert Spaces, and the crucial role of the Mercer theorem, represented in the prototypical framework of the kernel PCA algorithm [59]. In the recent years, significant developments have provided additional perspectives for dealing with the non-linear case. An important trigger was the publication, in the year 2000, of two reports proposing the Isomap and Local linear embedding (LLE) techniques. These reports have attracted a considerable and renewed interest for geometrical based methods in data analysis. Further results in the manifold learning and dimensional reduction setting have been devised using tools from differential geometry and algebraic topology [5, 47, 61, 66]. In the following list, adapted from [6], we present a very short snapshot of the developments in this field over the last years:

2000: Whitney Embedding based method [5]

2000: Isomap, LLE [53, 65]

2002: Laplacian Eigenmaps [3]

2003: Hessian Eigenmaps [21]

2004: Diffusion Maps, Local Tangent Space Alignment (LTSA) [64]

2005: Sample Logmaps [7]

2005: Riemannian Normal Coordinates [8]

In the following paragraphs we will describe the main components of the Isomap algorithm, a standard benchmark example for these algorithms denominated Roll Swiss dataset, and an important conceptual framework, the continuum Isomap, which clarifies fundamental properties of the Isomap algorithm.

Remark 4.8 (From MDS to Isomap)**.** When dealing with non-linear manifolds, a main problem with the MDS framework is the usage of the Euclidean distance in the formulation of the optimization problem. A typical example that illustrates this point is the Swiss roll dataset, defined as a finite sampling on surface parameterized by

$$f(u, v) = (u\cos(u), v, u\sin(u)), \quad u \in [3\pi/2, 9\pi/2],\ v \in [0, 1]. \tag{4.12}$$

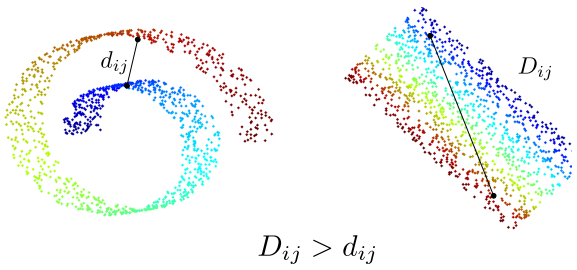


Figure 3: Isomap: Euclidean Distances vs Geodesic Distances

As can be seen in this example, using the Euclidean distance d_{ij} is not a meaningful strategy for describing the distance relationship between the points x_i and x_j, instead using the geodesic distance D_{ij}, represents a sensitive option adapted to the geometry of the dataset.

Remark 4.9 (Isomap Strategy)**.** In order to handle these kind of non linear manifolds, an elementary strategy, used in the Isomap algorithm, is to compute geodesic distances by considering the shortest path between groups of neighboring points. This procedure first identifies neighbor points using a k-nearest neighbors or ϵ radius criteria algorithms. Once the neighboring points are identified, the geodesic distance is computed by finding minimum paths between two given points using the neighborhood connections. Once the new geodesic distances for the given data set are obtained, the same procedure used in the MDS setting can be applied: by resolving an optimization problem, we construct a configuration of points Ω in a lower dimensional space that matches the distances in the original dataset $\mathcal{M}$.

1 **Neighborhood graph construction**: Define a graph where each vertex is a datapoint, and each edge connects two points if they fulfill an ϵ-radius or k-nearest neighbors criteria

2 **Geodesic Distance construction**: Compute the geodesic distance between each point using the previously constructed graph by finding the shortest paths between points.

3 **d-dimensional embedding**: Use the geodesic distance previously constructed in a MDS algorithm for computing a d-dimensional embedding.

Figure 4: Isomap Algorithm

The fact that now we use the geodesic distances allows to construct a configuration of points that represents a more accurate picture of the original dataset. As a comparison, we can see in the following picture the 2-dimensional parameterization of the Swiss roll dataset obtained with Isomap and PCA.

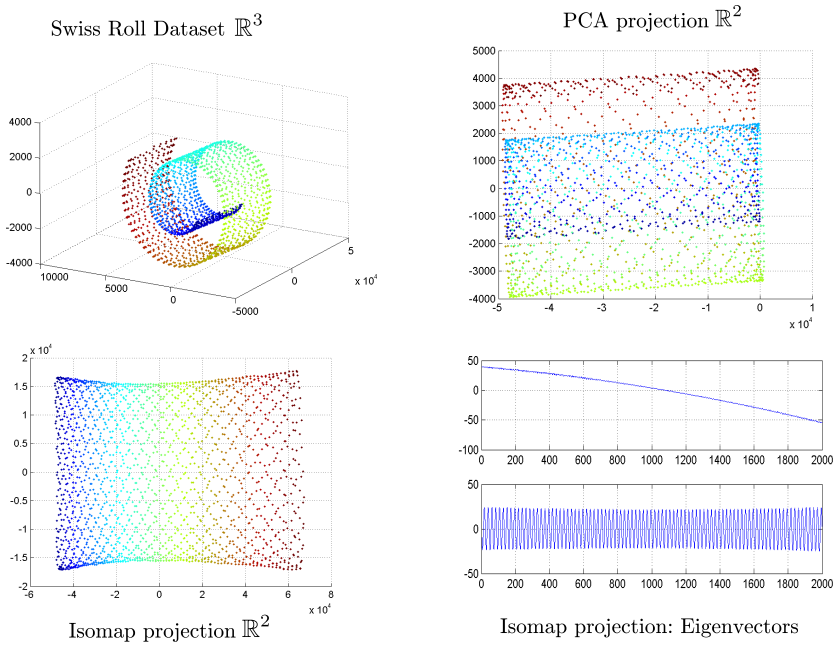


Figure 5: Isomap projection vs PCA projection

Continuum Isomap

The strategy proposed in the Isomap algorithm is an significant step in classification and data analysis of information lying on nonlinear manifolds, but fundamental questions concerning its functionality have not been clearly addressed in its original formulation. An important issue is that Isomap, originally designed as an algorithmic solution, does not include a well defined theoretical framework that could be used for a better understanding of its properties. In particular, no clear conditions where provided for Isomap that guarantee a correct reconstruction of the parameter space. For an accurate understanding of this algorithm, plain numerical experiments are unreliable verification procedures, and a better understanding of the underlying conceptual machinery is required. This problem has been recently addressed in a more conceptual framework denominated *continuum Isomap* [65], identifying some conditions on the manifold $\mathcal{M}$ for which the Isomap algorithm correctly constructs the low dimensional parameterization.

Remark 4.10 (Continuum Isomap formulation [65])**.** The idea of the *continuum Isomap* consist in generalizing the Isomap algorithm by constructing a kernel function K defined over the manifold $\mathcal{M}$ given the geodesic distances. The first step is the construction of a similarity kernel $K : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ generalizing the Equation 4.11:

$$K(x,y) = \frac{1}{2M}\int_{\mathcal{M}} (d^2(x,t) + d^2(t,y) - d^2(x,y))\,dt - \frac{1}{2M^2}\int_{\mathcal{M}}\int_{\mathcal{M}} d^2(t,s)\,dt\,ds. \quad (4.13)$$

Here, $M := \int_{\mathcal{M}} dt$. In a second step, following the original Isomap framework, we consider the eigenvectors ϕ_i of the integral operator with kernel K:

$$F_K(\phi)(x) \coloneqq \int_{\mathcal{M}} K(x,y)\phi(y)\,dy = \lambda\phi(x).$$

The main result of [65] is a theorem indicating necessary conditions for reconstructing the parameterization space Ω of a manifold $\mathcal{M}$, given only geodesic distance information. This result is a very particular and easy situation that implies the simplest structure in $\mathcal{M}$: as a manifold, its atlas consist of one single chart with Ω being its domain. Despite the simplicity of this situation, the result in [65] provides a controlable environment where we can garantee that the Isomap algorithm perfectly succeeds its objective. The key feature is that the d largest eigenvectors of the integral operator with kernel K allows to reconstruct the parameter space Ω:

Theorem 4.11 (Zha, Zhang, 2003)**.** *If $f : \Omega \subset \mathbb{R}^d \to \mathcal{M} \subset \mathbb{R}^n$ is an isometry with Ω open and convex, then the vector function θ build with the first d eigenvectors of F_K can be written as:*

$$\theta := (\phi_1, \ldots, \phi_d)^t = P(\tau - c),$$

with $\tau \in \Omega$, P is a constant orthogonal matrix, and c is a constant vector.

Consequences of the main theorem

An immediate application of this theorem allows to correctly interpret the usage of the Isomap algorithm in the following parameterization of the Swiss roll data set:

$$f(u, v) = (u\cos(u), v, u\sin(u))^t, \tag{4.14}$$

The Jacobian matrix of this transformation can be used to verify that f is not an isometry: $J_f(u,v)^t J_f(u,v) = \text{diag}(\sqrt{1+u^2}, 1)$. The initial assumption that for this particular Swiss roll data the Isomap algorithm constructs a perfect rectangular shape as a parameterized space must then be verified. The strategy proposed in [65] is to search for a suitable change of variable that will produce an isometry. The resulting map can then be analyzed with the Theorem 4.11, and the actual parameterization space can be revealed. Applied to our particular case, the change of variable transforms the map $f(u, v) = [u\cos(u), v, u\sin(u)]^t$ to an isometry, and the resulting vector function is

$$\theta = ((u\sqrt{1+u^2} + \text{arcsinh}(u))/2, v)^t.$$

In this expression, we found an explanation for the slightly distorted rectangular shape obtained empirically in Figure 5, as the term $u\sqrt{1+u^2} + \text{arcsinh}(u)$ produces a non linear curve revealing what really the Isomap algorithm computes:

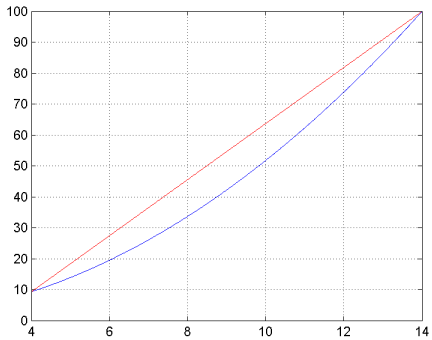


Figure 6: Isomap distortion: $u\sqrt{1+u^2} + \text{arcsinh}(u), u \in [3\pi/2, 9\pi/2]$, of the Swiss Roll from Equation 4.14 (see also Figure 5 and Reference [65])

5 Applications Examples

5.1 Action Potentials Background and EMG signals

Many living organisms are composed of a highly complex conglomeration of cells presenting a rich variety of interaction mechanisms, many of which are far from being understood, and represent very active research topics. One important example of cell interaction involves the propagation of electrical potentials along cell membranes. For instance, in the case of vertebrates, a voluntary muscular contraction includes an elaborate sequence of steps for the generation, transmission and translation of electrical discharges in neural cells. The analysis and mathematical signal interpretation of these electrical pulses, usually denominated *action potentials*, allows to obtain significant information about the underlying muscle and nerve interactions. The increasing demands in the field of biological signal analysis requires refined mathematical algorithms for supporting the physiological study of living tissues.

An important technique for measuring the electrical potentials generated by the muscular and neuronal activity are *Electromyogram (EMG)* signals. A standard EMG measurement procedure consist of inserting intramuscular needle electrodes in order to detect electrical potentials in a small localized muscular area. A less invasive method is denominated *surface EMG*, and consist of using skin surface electrodes for recording the electrical activity of wider muscular areas. The advantage of this second approach is the simplicity and comfortability of the measurement operation, but the drawback is the complexity of the recorded signal, which requires refined decomposition algorithms for extracting meaningful information. Standard applications of EMG analysis in medicine are diagnosis of diseases affecting the nervous and muscular system, but additional uses have been recently developed for prosthetic and virtual devise interfacing [63], as well as athletic performance analysis. The potential applications of EMG signals, and the complexity of the inherent physiological processes, are important stimulations for developing new mathematical and signal analysis techniques.

An important task in these developments is to construct accurate mathematical models of the action potential and its physiological reactions. In our context, an essential usage of these models is to design methods for evaluating and comparing new analysis algorithms. The accurate understanding of the generation and transmission of action potentials in neuronal cells is actually a very active research area. Standard descriptions are based on the *Hodgkin-Huxley model*, which explains an action potential as a purely electrical phenomenon based on the ionic exchange mechanisms in the cell membrane. But several experimental observations, in particular some temperature fluctuations along the cell membrane, cannot be explained with the Hodgkin-Huxley model. Recent alternative proposals [30, 31] have clarified these properties with the notion of a *soliton*, which can be roughly defined as a solitary wave pulse that conserves its energy while propagating in some medium. The concept of soliton is found in several physical wave propagation phenomena, ranging from nonlinear optics to fluid dynamics, and a precise definition can be specified in the very active research context of nonlinear partial differential equations. In the context of action potential mechanisms, the soliton model offers an alternative explanation based on thermodynamical properties of the cell membrane. From a signal processing point of view, the components of this new framework might be important to consider when designing refined models for studying EMG analysis algorithms.

Our purpose in this chapter is to illustrate a toy-example EMG analysis problem, together with a straightforward solution strategy that combines dimensional reduction methods with Fourier and wavelet transforms. An important incentive behind our plan is the problem of accurately detecting muscle fatigue factors from complex EMG measurements. The framework we present can be seen as an initial step towards this goal, addressing first a very simplified version of this problem with elementary conceptual and numerical settings. The first objective is to evaluate the feasibility of the strategy with easy controllable benchmark signals, before attempting to address more elaborate EMG models. In short, our toy-problem is to extract the inherent parameterization characterizing a set of EMG signals composed of a complex mixture of action potentials. Our main hypothesis is that the standard Fourier-based analysis techniques used in the EMG literature are not sensitive enough for some type of signals with complex spectral characteristics. The alternative strategy we analyze, combines dimensional reduction methods, (PCA and Isomap), with Fourier and wavelet transforms. Further work, in a theoretical, numerical, and EMG modeling context, should evaluate the applicability of these strategies in real EMG signals.

The plan of this chapter is as follows. We first present elementary information on cell physiology required for a general understanding of the action potential concept. For completeness, and as a trigger for further steps, we shortly mention the Hodgkin-Huxley mechanism, and the more recent soliton model. We then explain the framework of our EMG analysis problem, together with the simplified model for simulating action potentials. In the final section, we discuss the settings used in our numerical experiments, and different results illustrating the error behavior and algorithm quality.

Basic Motor System Physiology

In order to better understand the mathematical models describing the propagation of action potentials, some elementary background concepts in physiology are required. The high complexity of the motor system will only allow us to mention minimal aspects of the topics we require, including: neuron and muscle cell interactions, cell membrane biology, and models for action potentials propagation. The human motor system is a very elaborate mechanism combining an intricate communication between the central nervous system and the skeletal muscles. The concept of *motor unit* is the main building block, which combines a motor neuron to several muscle cells.

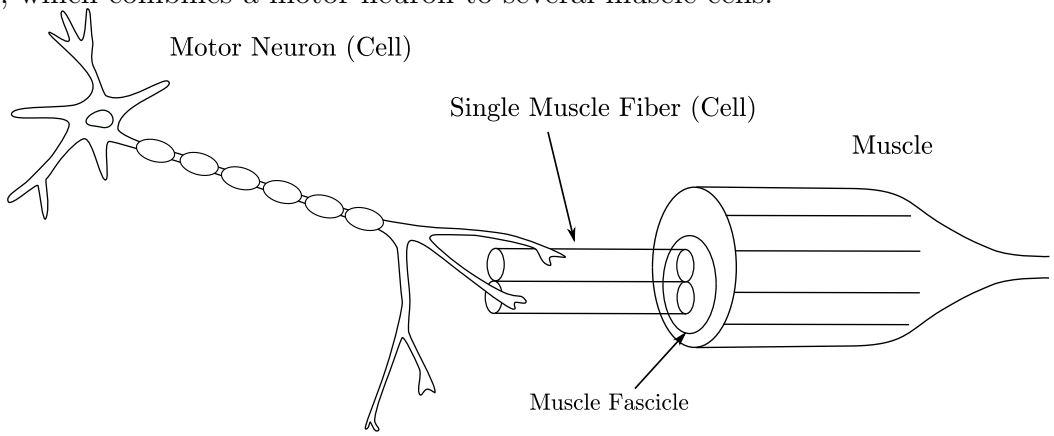


Figure 7: Motor Unit: motor neuron and muscle fibers

A motor neuron is highly specialized cell, dedicated to the processing and transmission of information to one or multiple muscle cells. The communication medium are the *nerve fibers* or *axons*, which are thin and long projections of the nerve cell, with a typical diameter of 1 micrometer, and an extension which can be of one meter or longer. Each axon can innervate several muscle cells, and the number of innervations depends greatly on the muscles type, required strength and precision movement. One of the strongest muscles in the human body is the mastication muscle, the masseter, with (approximately) 1020 motor neurons innervating each one 980 muscle fibers. The small muscles in the hand are specialized for refined movement: for instance, the Lumbricalis 1 is controlled by 98 motor neurons, each one innervating 110 muscle fibers:

Muscle	Number of motor units	Number of muscle fibers	Average number of muscle fibers
Arm Muscles			
Biceps brachii	774	580.000	750
Brachioradialis	330	130.000	390
Hand Muscles			
Lumbricalis 1	98	10.300	110
Opponens Pollicis	133	79.000	595
Mastication Muscles			
Masseter	1020	1.000.000	980
Temporalis	1150	1.500.000	13000

Figure 8: Approximate number of motor units in some human muscles [9, 46]

As most eukariotic cells found in multicellular organisms, an important component of a neuron is the *cell nucleus* containing genetic material in form of DNA molecules. There is a rich variety of neurons that can be classified according to different specializations, but in the case of motorneurons, a crucial characteristic are the long axons terminating in the *neuromuscular junction* that transmits information originated in the neuron cell's body to the muscular cells. A main transmission mechanism is the *action potential*, which can be defined as a signal that travels along the axon to the neuromuscular junction, where a complex set of processes is triggered, communicating with the muscle cells.

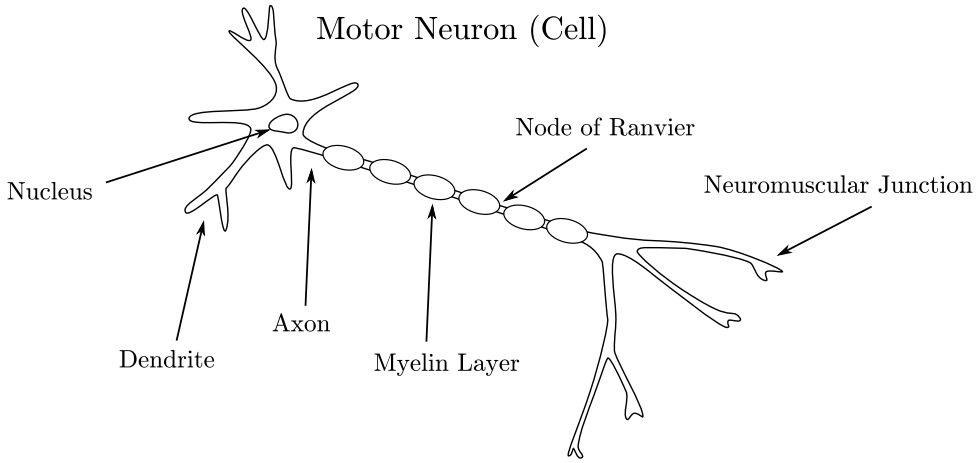


Figure 9: Neuron: central part and axon characteristics

A *skeletal muscle* is one type of muscle tissue specialized in voluntary muscle contractions, whose elaborate structure includes in particular a tight group of elongated *muscle cells*, also denominated *fibers*. There are several types of muscle fibers depending on particular characteristics as resistance to fatigue, force production, contraction time, size of motor neuron, etc. A crucial component of a muscle fiber are the *myofibrilis*, which are cylindrical units responsible for muscle contraction. Each myofibril of a skeletal muscle contains consecutive subunits called *sarcomeres*, which are protein complexes serving as the basic unit in the muscle contraction process.

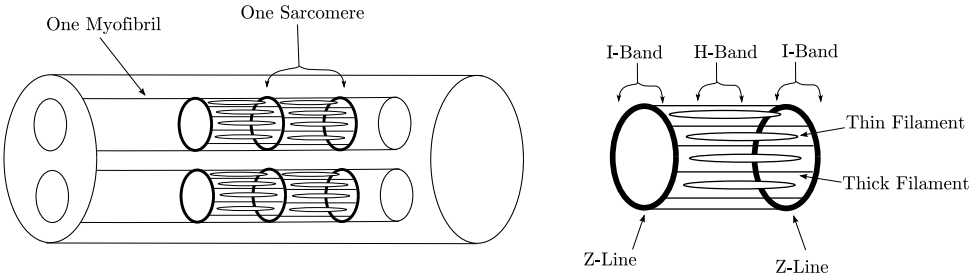


Figure 10: Muscle Fiber and its basic unit, the Sarcomere

Models for action potentials

The information transmission mechanism based on the action potential, occurs along the axon in the cell membrane, which is composed of a *phospholipid layer*, a structural unit containing a variety of molecules, such as proteins and lipids, and separating the extracellular from the intracellular environment.

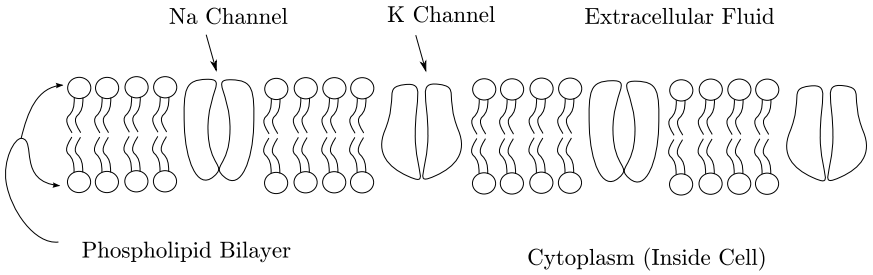


Figure 11: Cell Membrane: Phospholipid Bilayer and Protein channels

The Hodgkin-Huxley model, a major achievement in physiology of the second part of the 20th century, explains the action potential mechanism as an exchange of potassium and sodium ions through the ion channels in the phospholipid layer. The complex sequence of phases activating and deactivating the ion channels, results in the transmission of a signal down to the axon with a well defined bell shaped waveform that includes a slow repolarization phase (Figures 12 and 13). An alternative model recently proposed in [30, 31], clarifies some temperatures fluctuations occurring in the cell membrane that cannot be explained with the Hodgkin-Huxley model. The primary hypothesis of this model consist of a thermodynamical effect occurring when a soliton wave travels along the axon, producing a local change in density and thickness in the cell membrane.

EMG Background concepts

Electromyogram (EMG) signals are important examples of physiological data obtained by measuring electrical potentials generated by muscle cells. A typical electromyogram measurement consist in inserting needle electrodes in order to measure electrical activity in a small, localized muscular area. But an important alternative procedure, the surface electromyogram, consist in recording electrical activity in the skin surface, covering a wider muscular area with multiple muscle fibers. The surface electromyogram has clear advantages for manipulation and recording procedures, but the drawback are the more complex recorded signals that require refined analysis procedures.

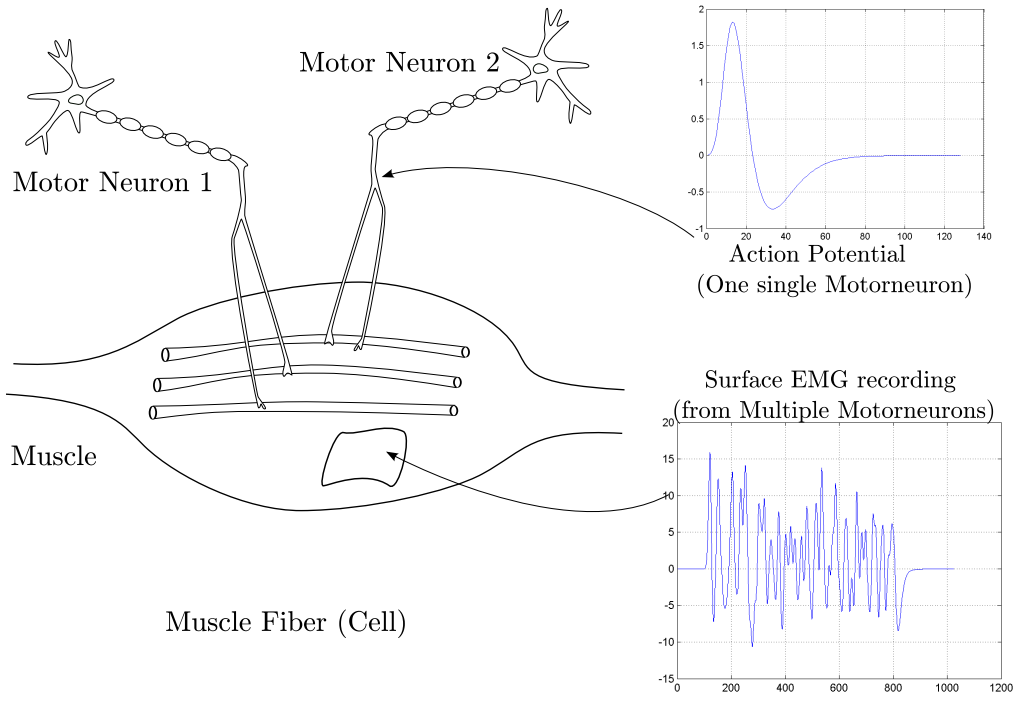


Figure 12: Action potential and Surface EMG

In Figure 12 we can compare a clear, well defined shape of an action potential originated from a single motor neuron, and the more complex combination of action potentials originated from multiple motor neurons, illustrating the surface electromyogram scheme.

One important application of these concepts is the analysis of EMG signals for the assessment of muscle fatigue. In the context of muscular activity (or any general biological process), the notion of fatigue is a very difficult concept to define, but there has been some important characterizations, developed over the last years, using the spectral behavior of EMG signals. It is a well known observation that the mean of the amplitude of the Fourier transform decreases during voluntary sustained contractions. This property offers a good example for the usage of several signal processing methods for extracting physiological information from EMG signals.

5.2 EMG Analysis Algorithms

With the elementary background information on motor unit potentials from the previous sections, we now address our particular EMG signal analysis problem. Given a set of EMG signals describing the muscular activity over a given period of time, our primary objective is to compute a function describing the fatigue evolution. The toy environment we present uses motor action potential models for setting fatigue parameters that can be used for constructing benchmark EMG signals. The philosophy we use is based on the fact that, as defined in our particular toy-environment, the fatigue evolution is an inherent parameterization of the set of EMG signals, which allows to consider its estimation as a low dimensionality reduction problem of the set of EMG signals. The challenge is to find an adequate transformation of these signals, such that a low dimensional representation can be used to reasonably estimate the fatigue evolution.

The synthetic and controlled framework we prepare can be seen as a very preliminary step for studying more realistic and complex situations of real life EMG signals. A typical real-life scenario would be to consider EMG signals recorded from a subject performing regular muscle contractions over a given period of time. One particular application of these experiments is to evaluate the athletic performance of highly trained subjects. A standard strategy is to analyze the time evolution of the EMG spectral characteristics for estimating the fatigue evolution. But in some situations, it can be the case that the high performance nature of the athlete does not allow to clearly identify the fatigue evolution with standard spectrum analysis tools. In order to address these situations, we evaluate the alternative analysis strategies using first a simplified and controlled environment of synthetic EMG signals.

We have therefore two main steps to perform. On the one hand, we construct a set of EMG benchmark synthetic signals where a given parameter can be use to simulate a very simplified and schematic version of the fatigue evolution. On the other hand, we test several strategies for estimating the fatigue evolution out of the EMG signal dataset. In this work, we place particular emphasis on techniques combining dimensional reduction techniques with Fourier and wavelet transforms.

Even though we use the problem of muscle fatigue estimation as a main motivation and application problem, our focus is primarily on illustrating a toy-example where the proposed techniques, combining dimensional reduction with standard signal transforms, could provide some meaningful information. In order to eventually apply these strategies in more realistic situations, further conceptual and experimental work is required. In the following subsections we present the result of this toy-example, illustrating the possibilities of this particular combination of methods.

Problem Setting

In order to describe in more concrete terms our hypothesis and settings, we define the basic elements required for defining the working environment. The main input of the algorithms we describe is a set of synthetic EMG signals where a given parameter can be used to implement a toy-example of the fatigue concept.

Definition 5.1 (EMG dataset)**.** We define our *EMG dataset* as a set of functions $U = \{u_k : I \to \mathbb{R}\}_{k\in J}$, with $I = \{1, \ldots, n\}$, and $J = \{1, \ldots, m\}$. The motivation for the the

index sets I, and J is to describe time parameters: I corresponds to the time evolution along one single EMG signal, and J describes the time evolution along the complete dataset of EMG signals.

Remark 5.1 (Main hypothesis: low dimensional representation as fatigue measure)**.** The main hypothesis of our setting is that an there exists an inherent low dimensional parameterization that we represent by a continuous function $\mathcal{A} : \Omega \subset \mathbb{R}^d \to \mathbb{R}^n$, with $d \leq n$, which simulates (in a simplified way) the fatigue concept. Given a dataset $U = \{u_k\}_{k=1}^m$, the main task is therefore to delete redundant information from this high dimensional set, in a way that only essential data characterizing the time evolution is kept, reconstructing the parameterization set Ω.

$$\Omega \subset \mathbb{R}^d \xrightarrow{\mathcal{A}} U = \{u_k\}_{k=1}^m \subset \mathbb{R}^n \tag{5.1}$$

The standard strategy is to construct the frequency spectrum of each EMG signal u_k, and compute the mean values in order to estimate the fatigue evolution. In this case we can consider the mean value as a low dimensional parameter characterizing the EMG dataset. But our second main hypothesis is that the EMG dataset presents a complex time evolution, in such a way that the mean values cannot provide sensible information for estimating the fatigue evolution. The numerical experiments we present in the following sections will illustrate this point.

Definition 5.2 (Fourier and Wavelet transforms of a EMG dataset)**.** We denote the Fourier and Wavelet transform of u_k by $\mathcal{F}(u_k)$ and $\mathcal{W}(u_k)$ respectively, and the corresponding sets $\mathcal{F}(U) := \{\mathcal{F}(u_k),\, k \in J\}$, and $\mathcal{W}(U) := \{\mathcal{W}(u_k), k \in J\}$, with $U := \{u_k, k \in J\}$.

Algorithm Strategy

The main strategy we will now explore is a straightforward application of *dimensionality reduction* or *manifold learning* techniques combined with wavelet and Fourier transforms for estimating the fatigue measure of a EMG dataset.

Remark 5.2 (General description of the Analysis Strategy)**.** Given a EMG dataset $U = \{u_k : I \to \mathbb{R}\}_{k=1}^m$, with $I = \{1, \ldots, n\}$, where each u_k represents one EMG signal, the strategy is to find an adequate signal transformation $\mathcal{T}$, and a (possibly nonlinear) map $\mathcal{P} : \mathbb{R}^n \to \mathbb{R}^d$ to a lower dimensional space (e.g. $d = 1, 2$, or 3) in a way that the resulting set $\mathcal{P}(\mathcal{T}(U))$ approximates the fatigue measure function. The map $\mathcal{P}$ depends on the type of method that is applied `M`, and the set S constructed out of the dataset U (i.e. $S = \mathcal{T}(U)$). We can use a diagram for depicting this procedure, where the main input is the dataset U, the unknown components are the low dimensional characterization Ω, and the embedding map $\mathcal{A}$.

$$\begin{array}{ccc} \Omega \subset \mathbb{R}^d & \xrightarrow{\mathcal{A}} & U = \{u_k\}_{k=1}^m \subset \mathbb{R}^n \\ & & \Big\downarrow \mathcal{T} \\ \Omega' \subset \mathbb{R}^d & \xleftarrow{\mathcal{P}} & S = \{\mathcal{T}(u_k)\}_{k=1}^m \subset \mathbb{R}^n \end{array} \tag{5.2}$$

We will see in the following paragraphs what kind of methods `M` do we use, and how to build exactly the set S, and the corresponding low dimensional representation Ω'.

Remark 5.3 (General problem description)**.** We can synthesize the previous observations by noting that the input of our main problem (Diagram 5.2) is the dataset of signals $U = \{u_k\}_{k=1}^m$, and the main hypothesis is existence of a continuous function $\mathcal{A} : \Omega \subset \mathbb{R}^d \to U \subset \mathbb{R}^n$ describing a low dimensional parameterization Ω of the dataset U. The principal problem is to find adequate signal transforms T, and dimensional reduction techniques $\mathcal{P}$, (with $S = \mathcal{T}(U)$) that can be used to approximate the parameterization space Ω.

Remark 5.4 (Method `M`, and the projection $\mathcal{P}$)**.** The questions to address now is how to construct the map $\mathcal{P}$, and what is exactly the input set S. The dimensional reduction and manifold learning machinery provides several possibilities for constructing maps $\mathcal{P}$. In the following experiments, we will focus on three basic options: the elementary PCA procedure, Multidimensional Scaling and the Isomap algorithm.

Remark 5.5 (Nonexplicit construction of the map $\mathcal{P}$)**.** It is important to notice that the map $\mathcal{P}$ is not always explicitly constructed as already indicated in the MDS or Isomap frameworks. In other words, our primary task will be to construct the set $\mathcal{P}(\mathcal{T}(U)) \subset \mathbb{R}^d$ out of the dataset $U \subset \mathbb{R}^n$, without necessarily computing an explicit construction of the map $\mathcal{P}$. This situation is the case, for instance, of the original formulation of the MDS algorithm, where an optimization procedure is described in order to find an adequate configuration of points in a low dimensional space 4.1. For these cases, an explicit computation of the map $\mathcal{P}$ can be considered as a separate problem, and modern approximation methods (e.g. [34]) are available for implementing efficient solution strategies.

Remark 5.6 (Construction of the set S from the dataset U)**.** In the map $\mathcal{P}$, we consider the input set S as constructed out of the the Fourier or wavelet transforms $\mathcal{F}(u_k)$, and $\mathcal{W}(u_k)$ of the EMG signals. The map can then be described as a transformation $\mathcal{P}(\mathcal{F}(u_k))$, or $\mathcal{P}(u_k)$). An important reference option to consider is the case when no particular transformation is applied to the EMG signals, and we apply directly the method `M` to the set $S = U$.

The idea behind the application of the Fourier or wavelet transforms is that the original dataset u_k is usually overloaded with information that does not directly contribute to the low dimensional characterization approximating the fatigue evolution. The objective of the Fourier and wavelet operators is to filter out some of this data, providing an important preprocessing step for the dimensional reduction algorithm.

Remark 5.7 (Mean and Median of the EMG power spectra)**.** Notice that the classical strategy of computing the mean values of the spectral frequencies of the signals u_k, is a particular case of this framework: the map for this case is $\mathcal{P} : U \to \mathbb{R}$ with $\mathcal{P}(u_k) = \|\mathcal{F}(u_k)\|_2$, and $U \coloneqq \{u_k, k \in T\}$.

Remark 5.8 (Algorithm strategy examples)**.** We will use the following dimensional reduction techniques `M`, where the goal is to approximate the parameterization set Ω: $\texttt{M} \in \{\text{Mean}, \text{Median}, \text{PCA}, \text{MDS}, \text{Isomap}\}$ and the set S will be constructed out of of the Fourier or Wavelet transforms (i.e. $\{\mathcal{M}, \mathcal{F}(\mathcal{M}), W(\mathcal{M})\}$.

The main problem to analyze is to find a good combination of maps $\mathcal{P}$ and $\mathcal{T}$, in a way that the resulting configuration of points $\Omega' = \mathcal{P}(\mathcal{T}(U))$ delivers an accurate approximation of the fatigue measure. We will test several dimensional reduction strategies

$\mathcal{P}$, and signal transformations $\mathcal{T}$, that depend on the complexity of the EMG dataset $U = \{u_k\}_{k=1}^m$.

The concept of dimensionality reduction involves a geometrical understanding of $\mathcal{W}(U)$, together with optimization procedures for finding a meaningful map $\mathcal{P}$. The key problem of this procedure can then be explained as how to correctly identify the geometry of the dataset $\mathcal{W}(U)$ in order to construct a meaningful parameterization space.

In the following subsections we will present several experiments for evaluating the extraction of a low dimensional parameterization of a EMG dataset based on dimensional reduction techniques combined with Fourier and wavelet transforms. There are two main steps in these experiments: preparation of the EMG dataset with a predefined fatigue measure function, and analysis of the EMG dataset with different algorithm strategies.

EMG Analysis Experiments: Dataset Construction

For constructing the EMG dataset, we use three steps which combines the concepts of action potential, a simplified fatigue measure parameter, and a schematic version of a EMG function:

1. Construction of a prototype action potential waveform.
2. Construction of a fatigue parameter evolution.
3. Construction of EMG dataset signals.

There are several important parameters of this dataset that will play an crucial role when selecting different analysis strategies. These include in particular the dimension d in which the parameterization space $\Omega \subset \mathbb{R}^d$ lives, and several values describing the complexity of the construction of the EMG signals (Remark 5.14).

Remark 5.9 (Construction of the action potential waveform)**.** In the construction of an action potential function, used as a basic building block of a EMG signal, we follow the model used in [40], which is described in further detail and used in a special application in [37,41]. This model of an action potential waveform is constructed from the derivative of the function V, which includes the fast depolarization represented as a narrow spike shape, and the slow repolarization step, also known as the exponential afterpotential.

$$V_{t_{ac}}(t) \coloneqq ag(t) + b(g * sr)(t), \quad sr(t) \coloneqq e^{-t/t_{ac}} 1_{[0,\infty[}(t). \tag{5.3}$$

The function sr is used to construct the slow repolarization component. As described in [40], the main parameters are a, the size of the spike, b, the size of the afterpotential, and t_{ac}, the time constant of the afterpotential. A possible option of the spike function g is based on the gamma function and two parameters, k and n, used to control the width and asymmetry of the spike:

$$g(t) \coloneqq \frac{k^{n+1}}{\Gamma(n+1)} t^n e^{-kt} 1_{[0,\infty[}(t). \tag{5.4}$$

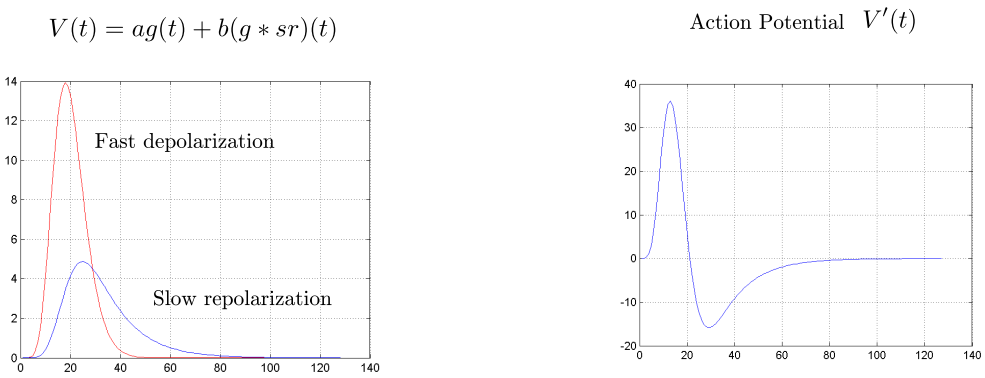


Figure 13: Construction of the action potential waveform

Remark 5.10 (EMG Signal Construction: MUAP)**.** In our toy-environment, we construct a simplified version of a EMG signal (which for our toy-framework will be also be identified as a Motor Unit Action potential MUAP) by adding several action potentials occurring at different time positions. In the following expression, our prototypical EMG signal is composed of a combination of q action potentials located at the positions $p_j,\ j = 1, \ldots, q$.

$$u(t) := \sum_{j=1}^{q} V_{t_{ac}}^{p_j}(t), \qquad V_{t_{ac}}^{p_j}(t) := V_{t_{ac}}(t - p_j). \tag{5.5}$$

In the following paragraphs, we will generalize this expression and we will construct a EMG dataset, by considering a set of u_k signals, where both t_{ac}, and p_j will depend on the parameter k.

Remark 5.11 (Simplified version of the fatigue parameter)**.** A real understanding of the muscle fatigue concept requires the assessment of multiple complex physiological phenomena. Our objective here is not to deal with this difficult problem, but instead to use this environment as a motivation for constructing toy-frameworks for testing alternative EMG signal analysis strategies. Our simplified framework is based on considerations of the effects of muscle fatigue on EMG signals properties discussed, for instance, in [42, Section 9.7.1], and [16]. The core idea is that the effects of the muscle fatigue on the power spectral characteristics of the EMG signal are, up to some extend, controlled by the shape of the waveform of the motor unit action potential. We use this information, and we oversimplify these properties, by considering the fatigue parameter to be controlled by t_{ac}, the time constant of the afterpotential. This allows to handle a simple toy framework that we can use for an initial testing and verification of our alternative analysis strategies.

Remark 5.12 (Construction of the EMG dataset)**.** In order to construct the EMG dataset we extend the definition of a EMG signal described in the Remark 5.10, by combining a set of action potentials in a sequence of signals parameterized by a given fatigue function. Specifically, based on the definition 5.1, we construct a sequence of signals $U = \{u_k\}_{k=1}^{m}$, with:

$$u_k(t) := \sum_{j=1}^{q} V_{t_{ac}(k)}^{p_{jk}}(t), \qquad V_{t_{ac}(k)}^{p_{jk}}(t) := V_{t_{ac}(k)}(t - p_{jk}). \tag{5.6}$$

In our toy-environment, the function $t_{ac}(k)$ describes the fatigue evolution that will play the role of unknown function. We can illustrate this construction of an EMG dataset with the following diagram, where we present a prototype motor unit action potential waveform, $V_{t_{ac}}$ parameterized by the time of the after potential t_{ac}, which changes over time with a given function $t_{ac}(k), k = 1, \ldots, m$. Here the parameter k, plays the role of another time variable, used also as an index for each entry of the EMG dataset. By adding q motor unit action potentials at different time positions, we obtain the dataset $U = \{u_k\}_{k=1}^{m}$.

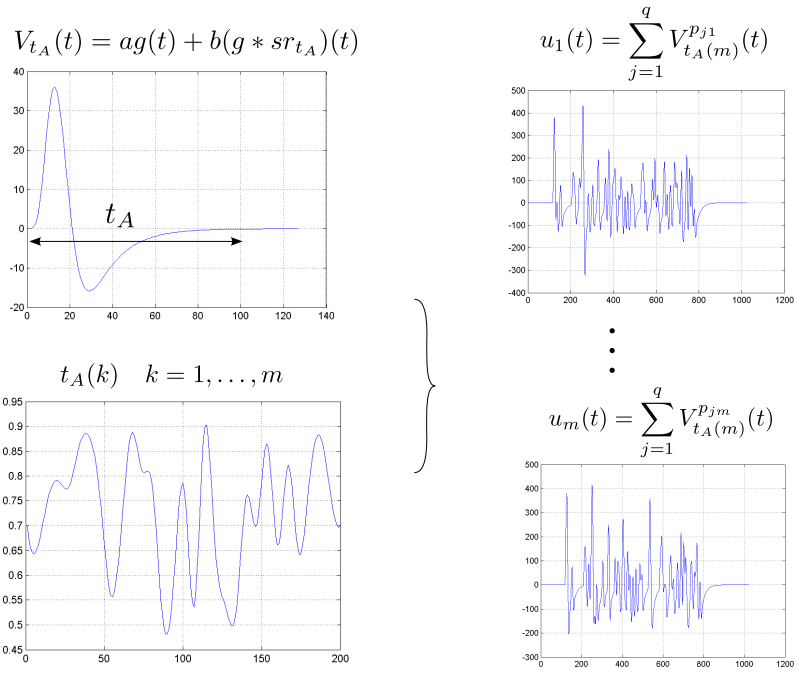


Figure 14: Construction of the EMG dataset

Remark 5.13 (Selection of the fatigue evolution)**.** In order to consider the approximation obtained with the different analysis strategies, we select a particular non trivial fatigue evolution depicted in the lower left graph of the Figure 14. This function has been constructed by interpolating random values over the range of the time span of the EMG dataset. The main objective of the algorithms strategies we present in this chapter is to approximate this function under special settings in the construction EMG dataset construction.

Remark 5.14 (Main Parameters and Functions in the EMG dataset)**.** Among the multiple settings of the EMG database, there are three main parameters playing a significant role in the analysis results:

q: number of action potentials used in a MUAP signal.

p_j: time shift behavior for the action potentials.

t_{ac}: range of fatigue parameterization values.

We will distinguish two principal situations depending on the behavior of the phase shift p_j: the case of constant phase shifts (i.e. $p_{jk} = p_j,\ k = 1, \ldots, m,\ j = 1, \ldots, q$), and the case of of phase shifts varying randomly over time (i.e. p_{jk} varies randomly with the variable k).

Remark 5.15 (Construction of the EMG Dataset: Extended version)**.** We extend the previous dataset construction to include a more refined version of EMG signals. We can rephrase and generalize the construction of the parameterization space by considering the function $t_{ac} : J \to \mathbb{R}^d$, such that $\Omega = t_{ac}(J)$. Each entries of this vector function $t_{ac}(k) = (t^1_{ac}(k), \ldots, t^d_{ac}(k))$, for $k = 1, \ldots, m$, (with $t^\alpha_{ac} : J \to \mathbb{R}$), is used to control a particular family of action potentials. We have therefore a more general description of the construction of the EMG dataset that depends on the dimensionality of the parameterization space $\Omega = t_{ac}(J)$:

$$u_k(t) = \sum_{\alpha=1}^{d} \sum_{j=1}^{q} V^\alpha_{t^\alpha_{ac}(k)}(t - p^\alpha_{jk}). \tag{5.7}$$

In this equation, we consider d types of action potentials $V^\alpha_{t^\alpha_{ac}}$, that are combined in a single EMG signal, u_k, that we use to construct our toy-example dataset. Each type α of action potential $V^\alpha_{t^\alpha_{ac}}$, has a particular set of formulas 5.3, 5.4, characterizing different types of spike behaviors. The objective of our toy-framework will be to find strategies for reconstructing the functions t^α_{ac} (e.g the parameterization space Ω), as described in the Diagram 5.2.

EMG Analysis Experiments: Results

We now apply the different analysis strategies discussed in the previous sections to the EMG dataset $U = \{u_k\}_{k=1}^m$. We will first address standard strategies involving the Fourier power spectrum, and elementary measures as the mean or median values. We will then consider more refined alternatives involving basic dimensional reduction strategies, including PCA, MDS and Isomap techniques. We finally analyze the combination of this methods with wavelet transforms.

Remark 5.16 (Case 1: Mean-Median, Power spectrum, constant phase shift)**.** The first test consist of applying standard strategies as the mean computation of the power spectrum of the dataset signals u_k. In Figure 15, we consider a dataset where each EMG signal u_k is composed of $q = 10$ synthetic action potential waveforms placed at some arbitrary, but constant, positions p_j. Namely, for each $k = 1, \ldots, m$ EMG signal (here with $m = 200$) we have $u_k(t) \coloneqq \sum_{j=1}^q V^{p_j}_{t_{ac}(k)}(t)$, with $V^{p_j}_{t_{ac}(k)}(t) \coloneqq V_{t_{ac}(k)}(t - p_j)$.

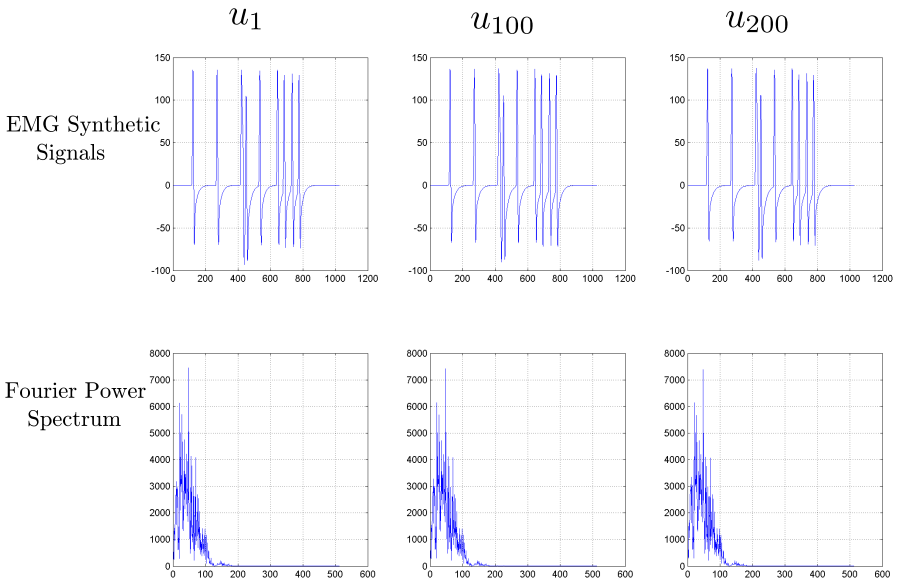


Figure 15: EMG synthetic signals examples and their Fourier spectra

As indicated in the Remark 5.14, there are two principal situations depending on the behavior of the fatigue function: a constant, and a random shift position variation. The case of the constant variation represents a very simple situation where the mean of the Fourier power spectra recovers effectively the fatigue parameterization. In Figure 16 we depict the function $t_{ac}(k)$, together with the evolution of the mean and median values of the Fourier power spectrum.

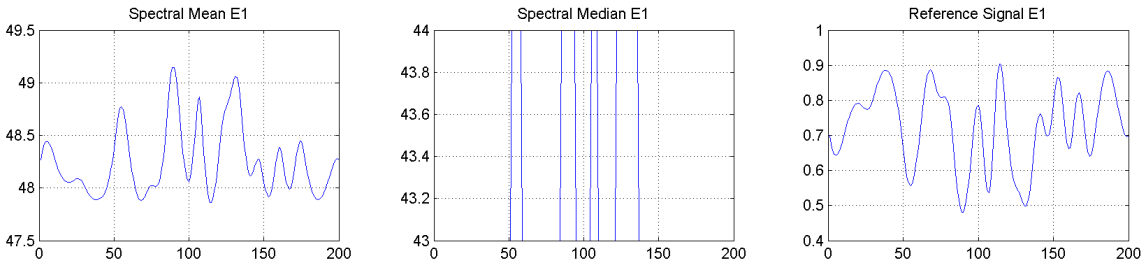


Figure 16: Mean, Median, and reference signal $t_{ac}(k)$

The mean values of the power spectral density allow to obtain an approximation of the unknown parameterization t_{ac}. Due to Heisenberg uncertainty principle, reflected in the inverse relation between the standard deviation of a function and its Fourier transform, we obtain a reflection of the graph of the power spectra mean function with respect to the parameterization t_{ac}. In the particular settings of this EMG dataset, the median value appears to be a very poor approximator of our target function t_{ac}.

Remark 5.17 (Case 2: Mean-Median, Power spectrum, random phase shifts)**.** We address now the more complex situation of random phase shifts. Namely, we consider the case where the phase varies randomly over time as described in equation 5.6. As depicted in Figure 17, this situation represents a more complex situation for the mean and median parameters which are far from providing an approximation for the function t_{ac}.

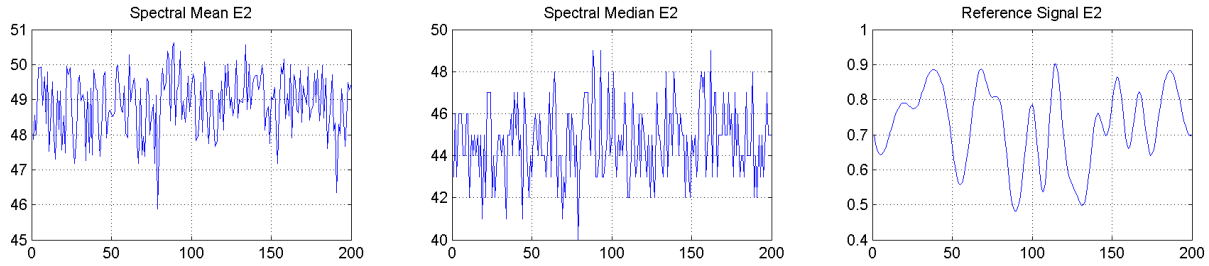


Figure 17: Mean, Median, and reference signal $t_{ac}(k)$: random phase shifts

Remark 5.18 (Dimensional reduction techniques vs standard power spectral mean values)**.** The mean or median values of the Fourier power spectrum used so far in the literature represent a characterization that depends only on the particular signal being analyzed. In our special context, it is more reasonable to consider a characterization that takes into account not just a single, but a group of signals in our dataset. This sort of characterization would be sensitive to the overall behavior of the EMG signals, and as such, would provide a more reliable approximation of the fatigue evolution. In the following steps we will illustrate this remark with additional experiments.

Remark 5.19 (Case 3: Power Spectrum, constant phase shifts, dimensional reduction)**.** The two previous cases indicate that, for this EMG dataset example, the standard techniques used for fatigue characterization are unreliable as soon as we have random phase shifts in the EMG signals. We now try a different strategy which consist on using alternative characterizations using dimensional reduction techniques. As indicated in the Remark 5.18, by considering the overall behavior of the EMG dataset, a dimensional reduction strategy might provide a better approximation of the fatigue evolution.

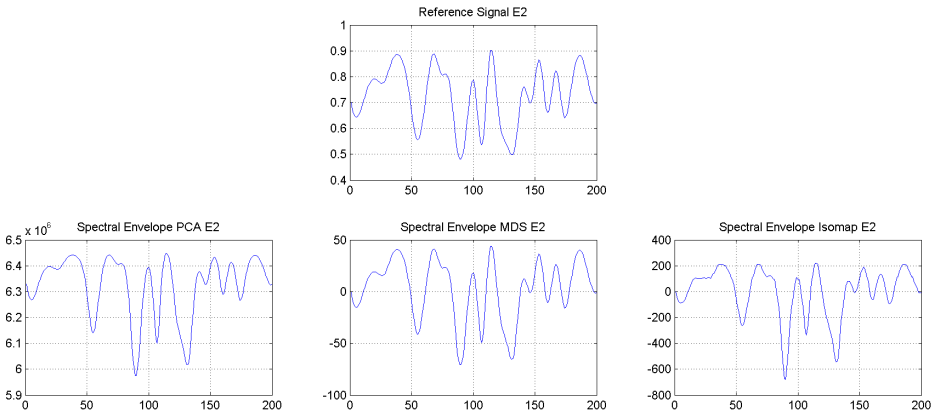


Figure 18: PCA, MDS and Isomap of Power Spectrum, constant phase shifts

The Figure 18 indicates that in the simplified situation of constant phase shifts we can approximate the fatigue parameterization t_{ac}, as it was in the case of the mean value characterization in Figure 16.

Remark 5.20 (Case 4: Power Spectrum, random phase shifts, dimensional reduction)**.** We introduce now a random phase shift p_j, and we check whether the Fourier Spectrum

would be sensitive enough for detecting the parameterization function. The result, indicated in Figure 19, illustrates that for these kind of examples, the combination of dimensional reduction techniques, with Fourier Spectrum appears to be problematic for approximating the fatigue parameterization t_{ac}.

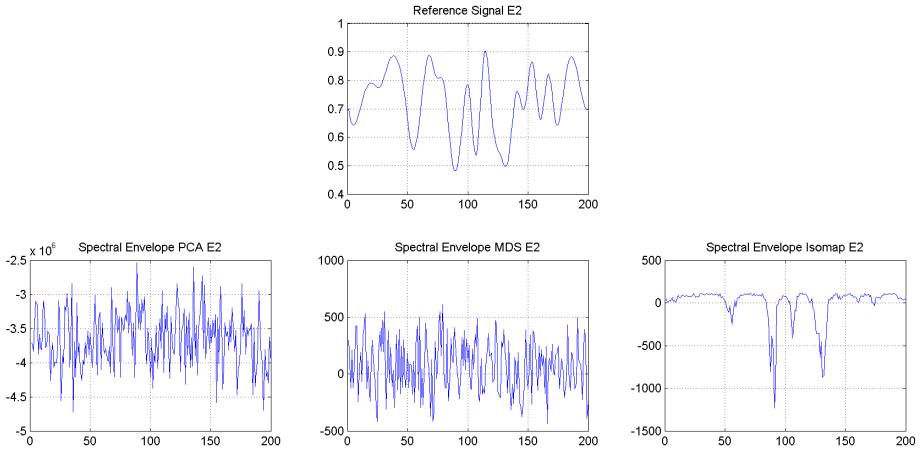


Figure 19: PCA, MDS and Isomap of Power Spectrum, random phase shifts

Remark 5.21 (Case 5: Plain Signals, constant phase shifts, dimensional reduction)**.** As an additional check, we test the results of dimensional reduction techniques applied to the plain signals, without considering any particular transformation. The case of constant phase shifts is presented in Figure 20, and, as expected, all dimensional reduction methods are approximating the fatigue parameterization without strong distortions.

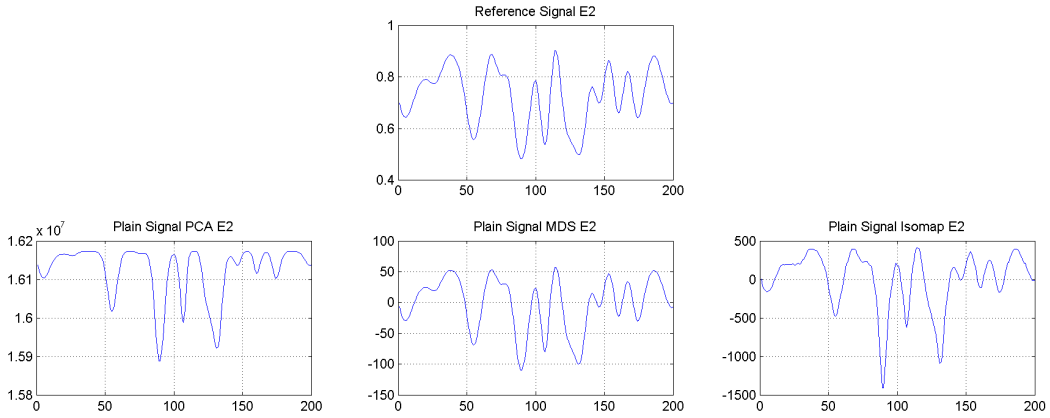


Figure 20: PCA, MDS and Isomap of Plain Signals Database, constant phase shifts

Remark 5.22 (Case 6: Plain Signals, random phase shifts, dimensional reduction)**.** The case of random phase shifts is more challenging for this combination of dimensional reduction with Plain EMG dataset signals. Isomap is the only method that appears to poorly approximate the fatigue parameterization, indicating that some transformation of the signals, as in the case of figure 19 is indeed required.

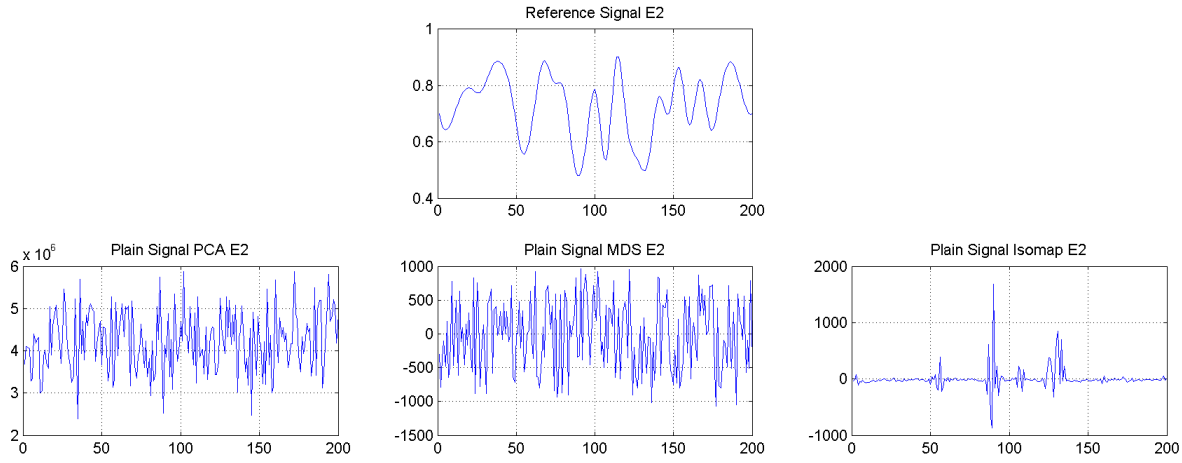


Figure 21: PCA, MDS and Isomap of Plain Signals Database, random phase shifts

Remark 5.23 (Case 7: Wavelet transformations with dimensional reduction)**.** In order to consider wavelet transforms with dimensional reduction techniques in our particular framework, we first experiment with the different possibilities provided by the multiresolution structure. The finite wavelet transform examples we apply converts a EMG signal represented as a vector of size $n = 1024$ into a vector of the same size, whose contents are computed according to a particular multiresolution scheme. For instance, if we select a multiresolution structure with 8 bands, the resulting scheme maps a vector u_k of size 1024 into a vector $\mathcal{W}(u_k)$ of the same size with 9 components:

$$V_4 \oplus W_4 \oplus W_8 \oplus W_{16} \oplus W_{32} \oplus W_{64} \oplus W_{128} \oplus W_{256} \oplus W_{512} = V_{1024} \tag{5.8}$$

Here we use a particular notation for the multiresolution structure: for each linear space in this decomposition we have $\dim(W_j) = \dim(V_j) = j$. We denote by $P_{W_j}(u_k) \in W_j$ (resp. $P_{V_j}(u_k) \in V_j$) the transformation mapping the vector u_k into the space W_j (resp. into the approximation space V_j).

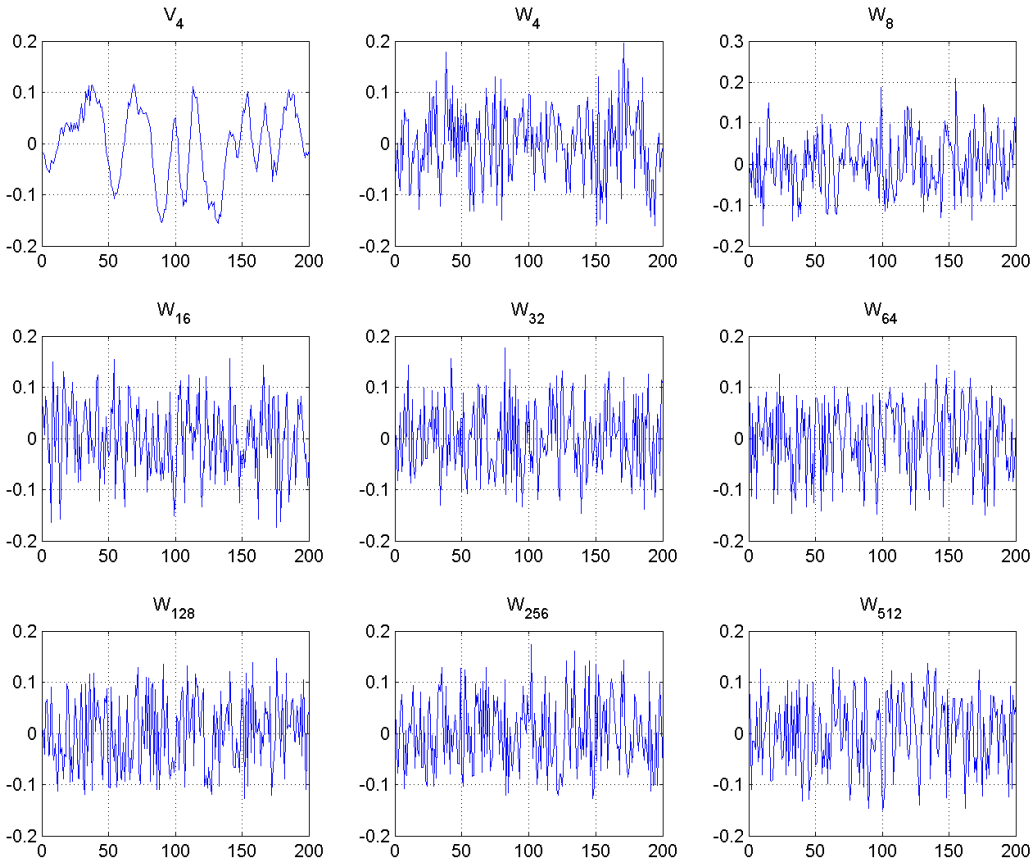


Figure 22: Dimensional Projections of the sets $\{P_{V_4}(u_k)\}_{k=1}^m$ and $\{P_{W_j}(u_k)\}_{k=1}^m$.

In Figure 22 we use a EMG dataset generated with random phase shifts, and we present the results of projecting the set of coefficients $\{P_{V_4}(u_k)\}_{k=1}^m$, and $\{P_{W_j}(u_k)\}_{k=1}^m$, (for $j = 4, 8, 16, \ldots, 512$) using Isomap as dimensional reduction method. As can be seen, the set $\{P_{V_4}(u_k)\}_{k=1}^m$ approximates up to some degree the parameterization function t_{ac}, but the coefficients obtained with the spaces W_j and their dimensional reduction versions

$\{P_{W_j}(u_k)\}_{k=1}^m$, are not as sensitive enough as the case of the approximation space V_4. In the following steps we will investigate in more detail the error behavior and further possibilities of this property.

As an additional information, in Figure 23 we depict a more precise view of the dimensional reduction projection of the signals in $\{P_{V_4}(u_k)\}_{k=1}^m \subset \mathbb{R}^4$, for the previous example. The upper right figure presents the eigenvector obtained with the PCD constructed out of the data depicted in the four signals with title `Band8Coef1`, up to `Band8Coef4`.

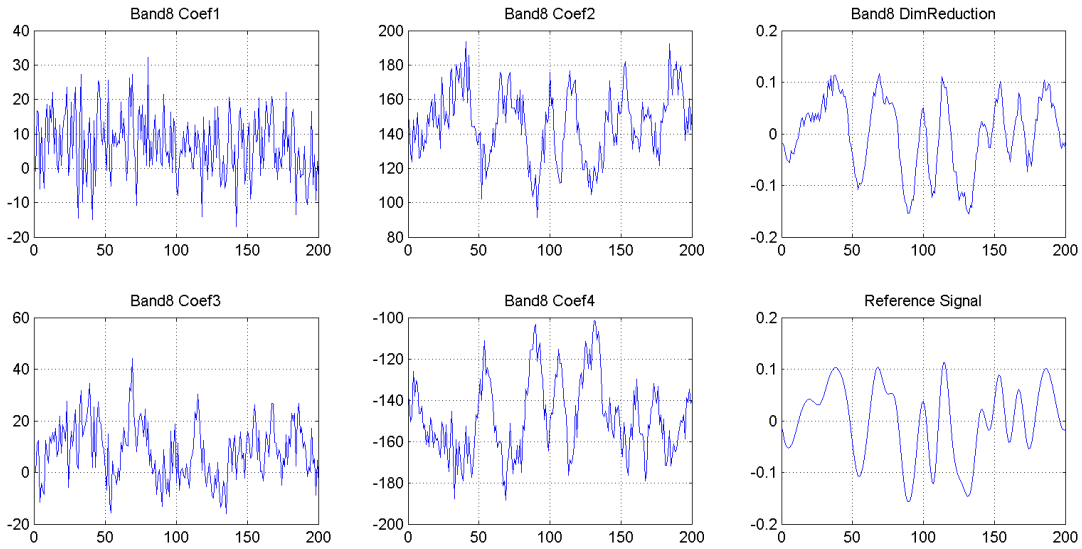


Figure 23: Four band 8 wavelet signals from V_4 and its Isomap projection

The task that follows is to gather these numerical results, and present a more detailed overview of some particular cases depending on the complexity of the database $U = \{u_k\}_{k=1}^m$. Our objective is to assemble some examples where the combination of wavelet transforms and dimensional reduction methods provide some improvement over direct implementations. The focus will be the error analysis of the strategy suggested by the results in the previous example from Figures 22 and 23.

EMG Analysis Experiments: Results Overview

We present the overview results of several examples illustrating different properties of the analysis strategies. As depicted in the following pages, the increasing complexity of the dataset examples requires more refined strategies, and the combination of wavelet transforms and dimensional reduction techniques appears to improve, for some situations, the approximation error. In the following remarks, we describe in more detail the strategy we analyze, and the corresponding error analysis. The presentation of the dataset examples includes a sequence of plots with the results of all implemented techniques, and subsequently, the error analysis of our proposed technique.

Remark 5.24 (Multiresolution and Dimensional Reduction strategy)**.** Based on the numerical experiments described in the previous remarks, the strategy we will now consider consist of using a multiresolution decomposition of each element of the dataset $U = \{u_k\}_{k=1}^m$, followed by a dimension reduction transform that will be used to approximate the parameterization function t_{ac}.

In order to generalize and formalize the procedure inspired by the experiments in Remark 5.23, we can rephrase the strategy using the projection $P_{V_{2^b}} : \mathbb{R}^m \to V_{2^b}$ in the approximation space V_{2^b} described in a multiresolution analysis. We recall that we use a special notation for expressing the multiresolution:

$$V_{2^b} \oplus W_{2^b} \oplus W_{2^{b+1}} \oplus W_{2^{b+2}} \oplus \cdots \oplus W_{2^{k-1}} = V_{2^k} \tag{5.9}$$

Here we have $\dim(V_{2^b}) = 2^b$, and in our particular numerical examples we use $k = 10$, and $b = 0, \ldots, 9$. We can now formulate the approximation scheme as a particular case of the Diagram 5.2, where $\mathcal{T} = P_{V_{2^b}}$, and the resulting approximation Ω_b, depends on a selected multiresolution level b.

$$\begin{array}{ccc} \Omega \subset \mathbb{R}^d & \xrightarrow{\mathcal{A}} & U = \{u_k\}_{k=1}^m \subset \mathbb{R}^n \\ & & \Big\downarrow {\scriptstyle P_{V_{2^b}}} \\ \Omega_b \subset \mathbb{R}^d & \xleftarrow{\mathcal{P}} & S_b = \{P_{V_{2^b}}(u_k)\}_{k=1}^m \subset \mathbb{R}^{2^b} \end{array} \tag{5.10}$$

As we will see in the following experiments, the main task is to investigate the levels b that provide an adequate approximation Ω_b of Ω, depending on the complexity of a particular dataset $U = \{u_k\}_{k=1}^m$.

Remark 5.25 (Error Analysis)**.** In order to consider the error produced by the approximation scheme described in Remark 5.24, we compute, for each level b, the difference between the approximation space Ω_b, and the original parameterization Ω. The procedure we use depends on our particular toy-framework, and the design of Ω. As described in Remark 5.15, we use in our numerical experiments the space Ω constructed as a curve in $\mathbb{R}^d$. Now, the error we define will compare each entry t_{ac}^α with a corresponding component of the dimensional reduction map $\mathcal{P}^\alpha$ (to be defined shortly):

$$\mathrm{err}_\alpha^{\mathcal{P}}(b) := \mathrm{err}_\alpha^{\mathcal{P}}(\Omega, \Omega_b) := \|\overline{\overline{t_{ac}^\alpha}} - \overline{\overline{(\mathcal{P}^\alpha(P_{V_{2^b}}(u_k)))_{k=1}^m}}\|_2. \tag{5.11}$$

Here, we use a special vector normalization (denoted as $\overline{\overline{t_{ac}^\alpha}}$), resulting in a zero-mean, unit-norm vector, defined as follows:

$$\overline{\overline{v}} := \frac{v - \overline{v}e}{\|v - \overline{v}e\|_2}, \quad \text{with} \quad \overline{v} := \frac{1}{m}\sum_{i=1}^{m} v_i, \quad v \in \mathbb{R}^m, \quad \text{and} \quad e := (1, \ldots, 1)^t \in \mathbb{R}^m.$$

In order to explain the $\mathcal{P}^\alpha$ component, we recall that with the image of the dimensional reduction map $\mathcal{P}$ we have a d-dimensional vector $\mathcal{P}(P_{V_{2^b}}(u_k)) \in \mathbb{R}^d$, corresponding to the d eigenvectors from the biggest d eigenvalues constructed in a Isomap or PCA reduction technique. By taking each entry of $\mathcal{P} = (\mathcal{P}^\alpha)_{\alpha=1}^d$, and considering the evolution over the m EMG signals u_k, we obtain a m dimensional vector $(\mathcal{P}^d(P_{V_d}(u_k)))_{k=1}^m$ that we can compare with the entries of the parameterization map t_{ac}^α. Notice that we still have to correctly identify the components of the vector $\mathcal{P}$ that correspond to the ones in t_{ac}, a problem that we resolve in our numerical experiments by trial and error.

In our particular experiments, we have $u_k \in \mathbb{R}^{1024}$, and therefore the options $b = 0, \ldots, 9$ can be considered. The resulting error analysis graph will contain several entries corresponding to the different multiresolution decompositions available for a vector of size 1024. For instance, in Figure 24 we have a prototypical error plot used in the experiments presented in the following sections (e.g. Figures 27, 30, 33, 36).

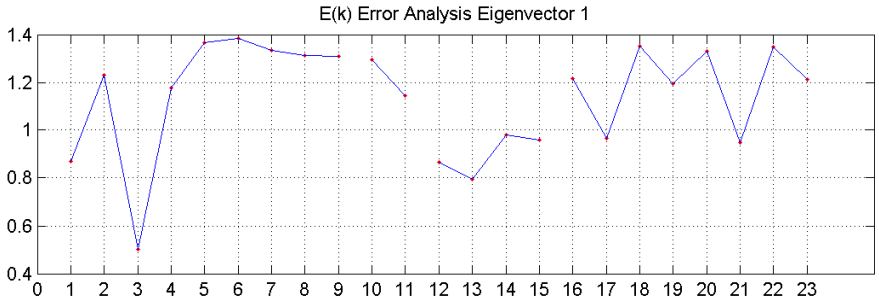


Figure 24: Example of an Error Analysis Plot (from Figure 36)

The first 9 entries of the graph correspond to the values $\text{err}_\alpha^{\mathcal{P}}(b), b = 0, \ldots, 8$ (denoted in the plot as $E(k), k = 1, \ldots, 9$). In order to visualize whether the dimensional reduction improves the approximation, we include also the error committed when considering directly the entries of the wavelet coefficients. Namely, as in the previous error construction, we use the entries of the vector $P_{V_{2^b}}^\alpha$ with $\alpha = 1, \ldots, 2^b$, for creating a m dimensional vector that can be compared with the original parameterization t_{ac}^α. We define the error components $E(k)$, for $k > 9$, with $E(k) = \text{err}_\alpha^P(b)$, $b = k - 1$, and:

$$\text{err}_\alpha^P(b) := \text{err}_\alpha^P(\Omega, \Omega_b) := \|\overline{\overline{t_{ac}^\alpha}} - \overline{\overline{(P_{V_{2^b}}^\alpha(u_k))_{k=1}^m}}\|_2. \tag{5.12}$$

Notice that, for instance, in the particular example of Figure 24, we have the first multiresolution structure V_{2^0}, represented by the corresponding error $E(1)$, and computed with the signal $(P_{V_{2^0}}^1(u_k))_{k=1}^m$. In this case, we have one single signal, and therefore no dimensional reduction is required (see the first line of Figure 25, and the corresponding plot in Figure 24).

Multiresolution Level V_{2^b}	**Dimensional Reduction Error**	**Multiresolution Error Start, . . . ,End**
$b = 0, \quad V_0$	$E(1)$	$- - -$
$b = 1, \quad V_2$	$E(2)$	$E(10), E(11)$
$b = 2, \quad V_4$	$E(3)$	$E(12), E(13), E(14), E(15)$
$b = 3, \quad V_8$	$E(4)$	$E(16), \ldots, E(23)$

Figure 25: Structure of Error Analysis plot in Figure 24

For the following multiresolution structure, V_{2^1}, we have the values $E(10)$, and $E(11)$, which corresponds to the error committed by $(P^1_{V_{2^1}}(u_k))_{k=1}^m$, and $(P^2_{V_{2^1}}(u_k))_{k=1}^m$, respectively. The error of the first eigenvector, $(\mathcal{P}^1(P_{V_{2^1}}(u_k)))_{k=1}^m$, of the dimensional reduction of these vectors is represented by $E(2)$. For this case, we don't observe a significant increase in the approximation quality when using dimensional reduction (see the second line of Figure 25, and the corresponding plot in Figure 24).

The following multiresolution structure, V_{2^2}, is used to compute the errors $E(3)$ for the dimensional reduction version, and $E(k), k = 12, \ldots, 15$, for non dimensional reduced wavelet coefficients. In this situation, we observe a significant improvement in the approximation quality when using dimensional reduction, illustrating the possibilities of the framework indicated in Diagrams 5.2 and 5.10.

Example 1: Constant phase shifts, One parameterization curve

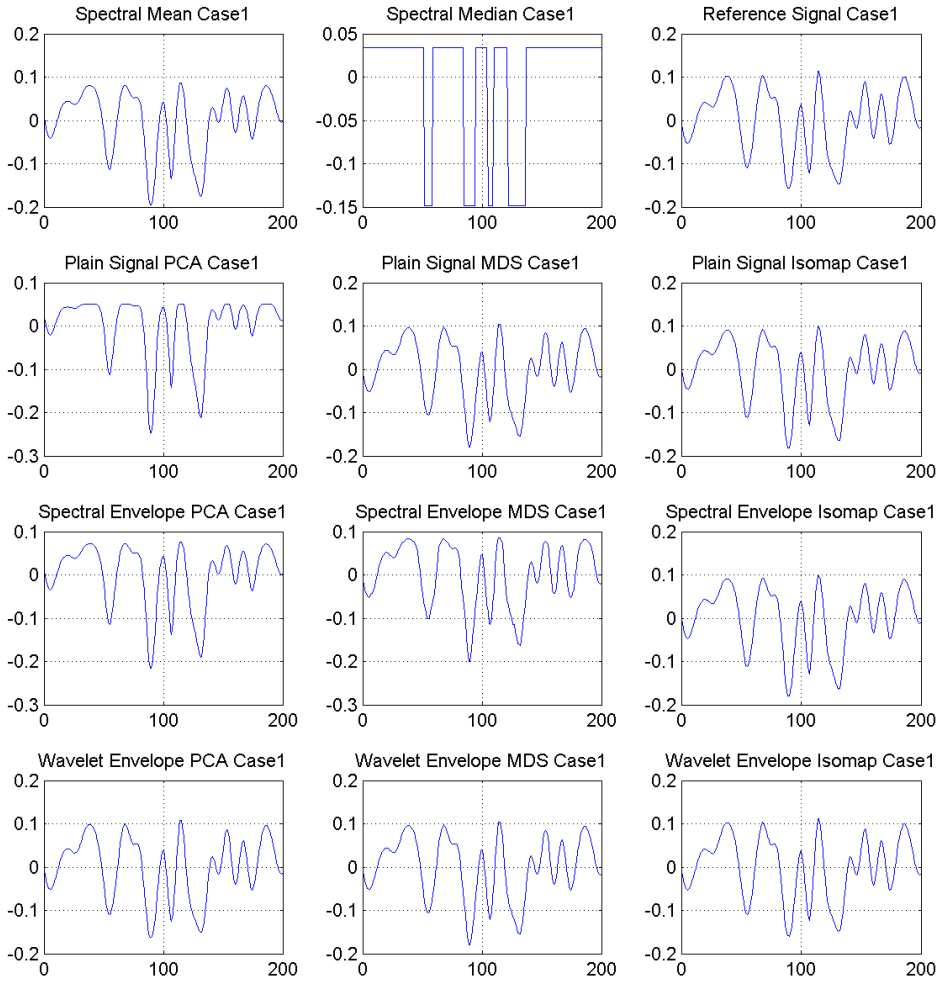


Figure 26: All Techniques: Example 1

In this example, we used one single parameterization curve for controlling the time evolution of the dataset. Constant phase shifts are used in the construction of the EMG signals. Example 1 parameters `rndFluct=0; reps=10; dim=1; (pos1,len1)=(100,700);`:

$$\boxed{u_k(t) := \sum_{j=1}^{q} V_{t_{ac}}(t - p_{jk})}, \quad q = 10, \quad p_{jk} = p_j, \quad 100 \leq p_j \leq 700, \quad u_k \in \mathbb{R}^{1024}. \quad (5.13)$$

We notice that most of the methods are capable of approximating, up to some degree, the reference parameterization signal, depicted in the right upper corder of this diagram. As we will see in the following examples, this stable behavior will not be the case as soon as the complexity of the signals is increased when random phase shifts are used.

Example 1: Constant phase shifts, `dim=1`, Error Analysis

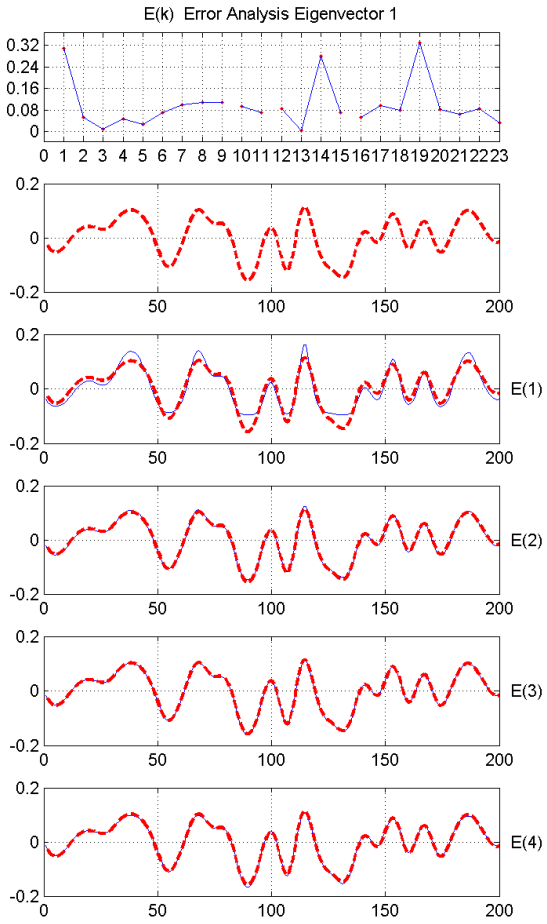


Figure 27: Error Analysis: Example 1

In Figure 27, the red dotted curve represents the reference signal, and the blue curves are the corresponding approximations. In this error analysis, we observe a local minima in $E(3)$ which corresponds to a multiresolution structure V_{2^2}. This value represents the first eigenvector of the dimensional reduction projection of the vectors used to compute the errors $E(k), k = 12, 13, 14, 15$.

$E(3)$	$E(12)$	$E(13)$	$E(14)$	$E(15)$
0.0087	0.0839	0.0041	0.2805	0.0699

Figure 28: Error values: Eigenvector 1, case $E(3)$

As can be seen $E(13)$ is actually smaller than $E(3)$, and therefore no significant improvement has been achieved with the dimensional reduction. However, we can argue that in a standard situation where no further information is available, the dimensional reduction procedure could help to identify a optimal vector among the corresponding values $E(k), k = 12, 13, 14, 15$.

Example 2: Random phase shifts, One parameterization curve

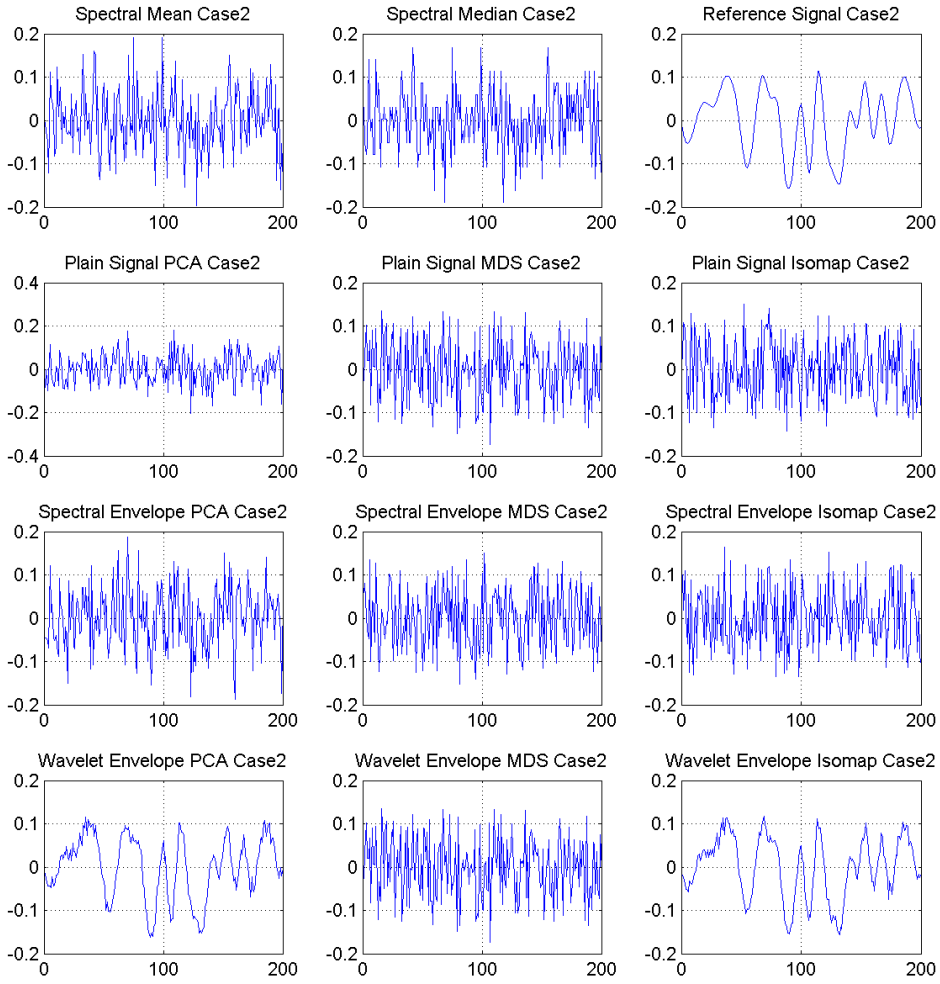


Figure 29: All Techniques: Example 2

In this example, we used one single parameterization curve for controlling the time evolution of the dataset, and random phase shifts varying between 0 and 70 samples for the construction of the EMG signals. Example 2 Parameters: `rndFluct=70; reps=80; dim=1; (pos1,len1)=(100,700);`

$$\boxed{u_k(t) := \sum_{j=1}^{q} V_{t_{ac}}(t - p_{jk}),} \quad q = 80, \quad p_{jk} = p_j + 70\text{rand}, \quad 100 \leq p_j \leq 700 \qquad (5.14)$$

We notice that only PCA and Isomap are capable of approximating, up to some degree, the reference parameterization signal, depicted in the right upper corder of this diagram. Due to the random phase shifts p_{jk}, all other methods are unable to identify correctly the main fatigue evolution t_{ac}.

Example 2: Random phase shifts, `dim=1`, Error Analysis

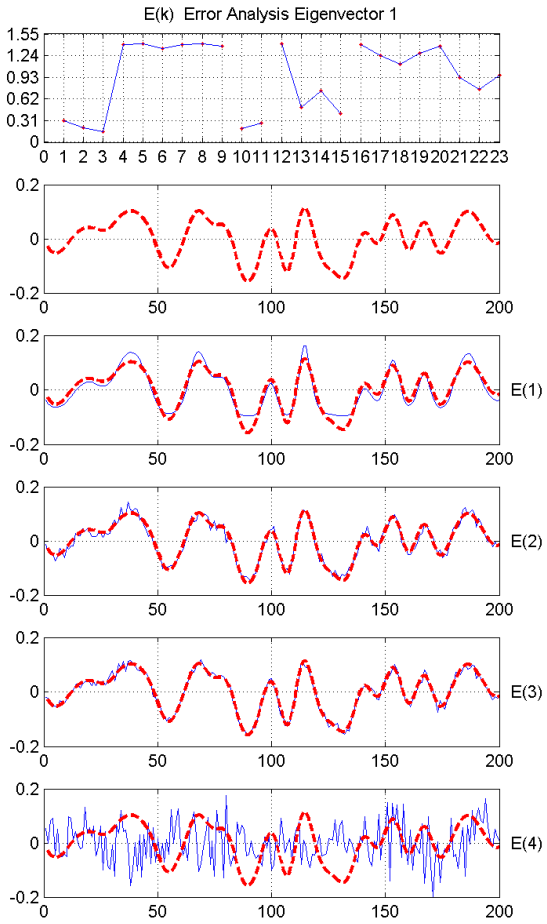


Figure 30: Error Analysis: Example 2

In Figure 30, the red dotted curve represents the reference signal, and the blue curves are the corresponding approximations. In this error analysis, we observe, once more, a local minima in $E(3)$ which corresponds to a multiresolution structure V_{2^2}. This value represents the first eigenvector of the dimensional reduction projection of the vectors used to compute the errors $E(k), k = 12, 13, 14, 15$.

$E(3)$	$E(12)$	$E(13)$	$E(14)$	$E(15)$
0.1518	1.4108	0.5046	0.7398	0.4119

Figure 31: Error values: Eigenvector 1, case $E(3)$

In this case, the value $E(3)$ constructed with a dimensional reduction map, represents a significant improvement with respect to all other vectors used to compute the errors $E(k), k = 12, 13, 14, 15$. This illustrates an example of the positive contribution of combining dimensional reduction techniques with wavelet transforms.

Example 3: Random phase shifts, Two parameterization curves

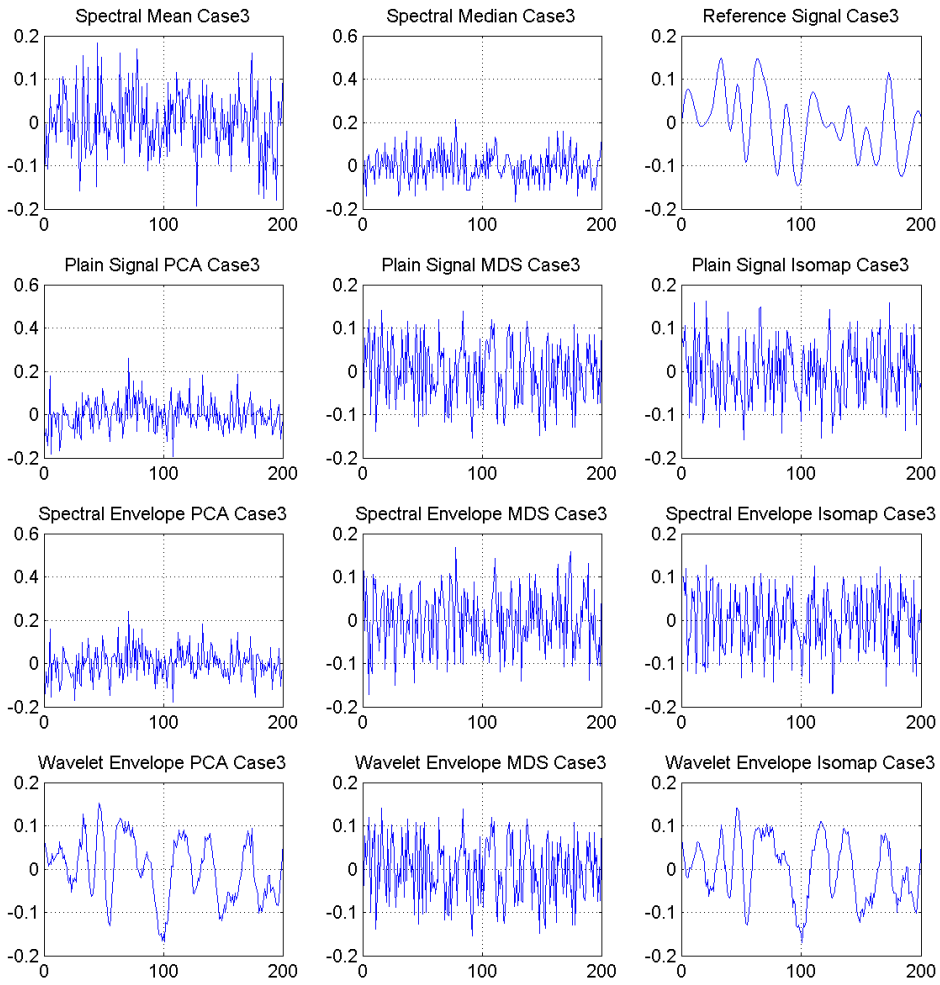


Figure 32: All Techniques: Example 3

In this example, we used two parameterization curves for controlling the time evolution of the dataset, and random phase shifts varying between 0 and 70 samples for the construction of the EMG signals. Example 3 Parameters: `rndFluct=70; reps=80; dim=2; (pos1,len1)=(100,500);`

$$\boxed{u_k(t) := \sum_{j=1}^{q} V^1_{t_{ac}}(t - p^1_{jk}) + \sum_{j=1}^{q} V^2_{t_{ac}}(t - p^2_{jk}),} \tag{5.15}$$

$$q = 80, \quad p^\alpha_{jk} = p^\alpha_j + 70\text{rand}, \quad \alpha = 1, 2 \tag{5.16}$$

$$100 \leq p^1_j \leq 500, \qquad 300 \leq p^2_j \leq 800. \tag{5.17}$$

Here, as before, we have as good approximation candidates, the dimensional reduction procedures obtained with PCA and Isomap. It is important to remember that these graphs are describing only on parameterization curve. In the following error analysis, we consider both components t^α_{ac}, $\alpha = 1, 2$.

Example 3: Random phase shifts, `dim=2`, Error Analysis

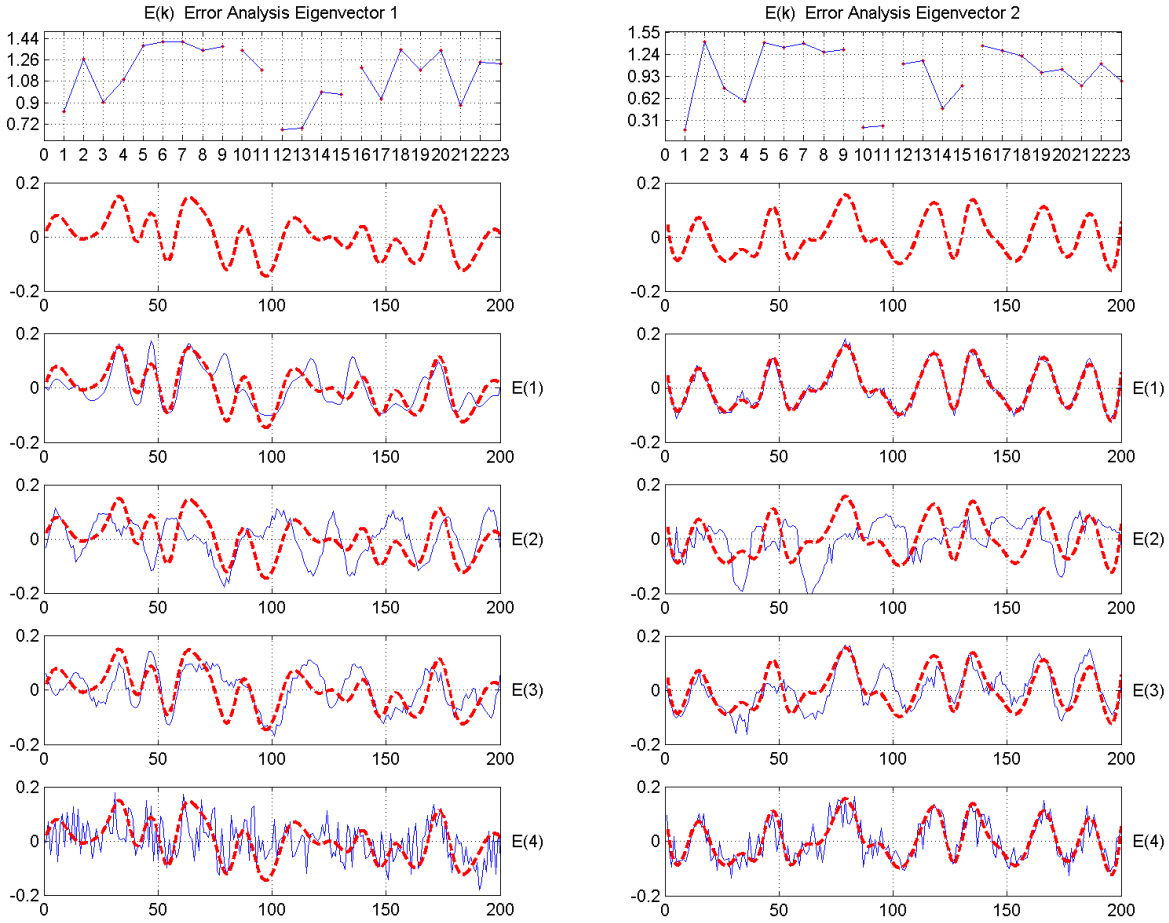


Figure 33: Error Analysis: Example 3

In Figure 33, the red dotted curve represents the reference signal, and the blue curves are the corresponding approximations. In this error analysis, we observe, contrary to the previous example, a significant distortion caused by the dimensional reduction in all cases of error components $E(2), E(3), E(4)$. This happens for both eigenvectors 1 and 2, suggesting that the best strategy for this case, is to consider the first multiresolution structure V_{2^0}. This indicates that further analysis is required to fully understand when, why, and how new improvements can be achieved with this combination of dimensional reduction and wavelet transforms.

$E(3)$	$E(12)$	$E(13)$	$E(14)$	$E(15)$
0.9028	0.6750	0.6842	0.9881	0.9700

Figure 34: Error values: Eigenvector 1, case $E(3)$

Example 4: Random phase shifts, Three parameterization curves

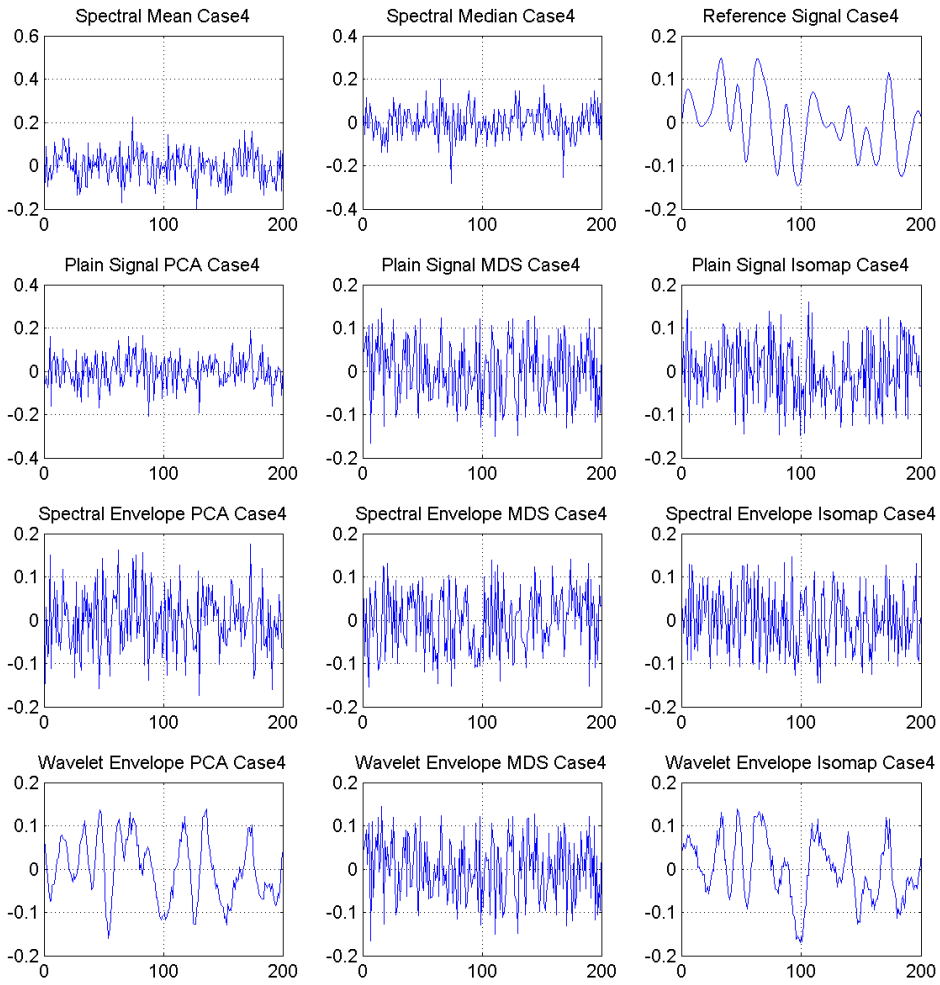


Figure 35: All Techniques: Example 4

In this example, we used three parameterization curves for controlling the time evolution of the dataset, and random phase shifts varying between 0 and 70 samples for the construction of the EMG signals. Example 4 Parameters: `rndFluct=70; reps=80; dim=3; (pos1,len1)=(100,500);`

$$\boxed{u_k(t) := \sum_{j=1}^{q} V^1_{t_{ac}}(t - p^1_{jk}) + \sum_{j=1}^{q} V^2_{t_{ac}}(t - p^2_{jk}) + \sum_{j=1}^{q} V^3_{t_{ac}}(t - p^3_{jk}),} \tag{5.18}$$

$$q = 80, \quad p^\alpha_{jk} = p^\alpha_j + 70\text{rand}, \quad \alpha = 1, 2, 3 \tag{5.19}$$

$$100 \le p^1_j \le 500, \quad 300 \le p^2_j \le 800 \quad 500 \le p^2_j \le 800. \tag{5.20}$$

Here, as before, we have as good approximation candidates, the dimensional reduction procedures obtained with PCA and Isomap. We recall that these graphs are describing only one parameterization curve. In the following error analysis plots, we consider all components t^α_{ac}, $\alpha = 1, 2, 3$.

Example 4: Random phase shifts, `dim=3`, Error Analysis

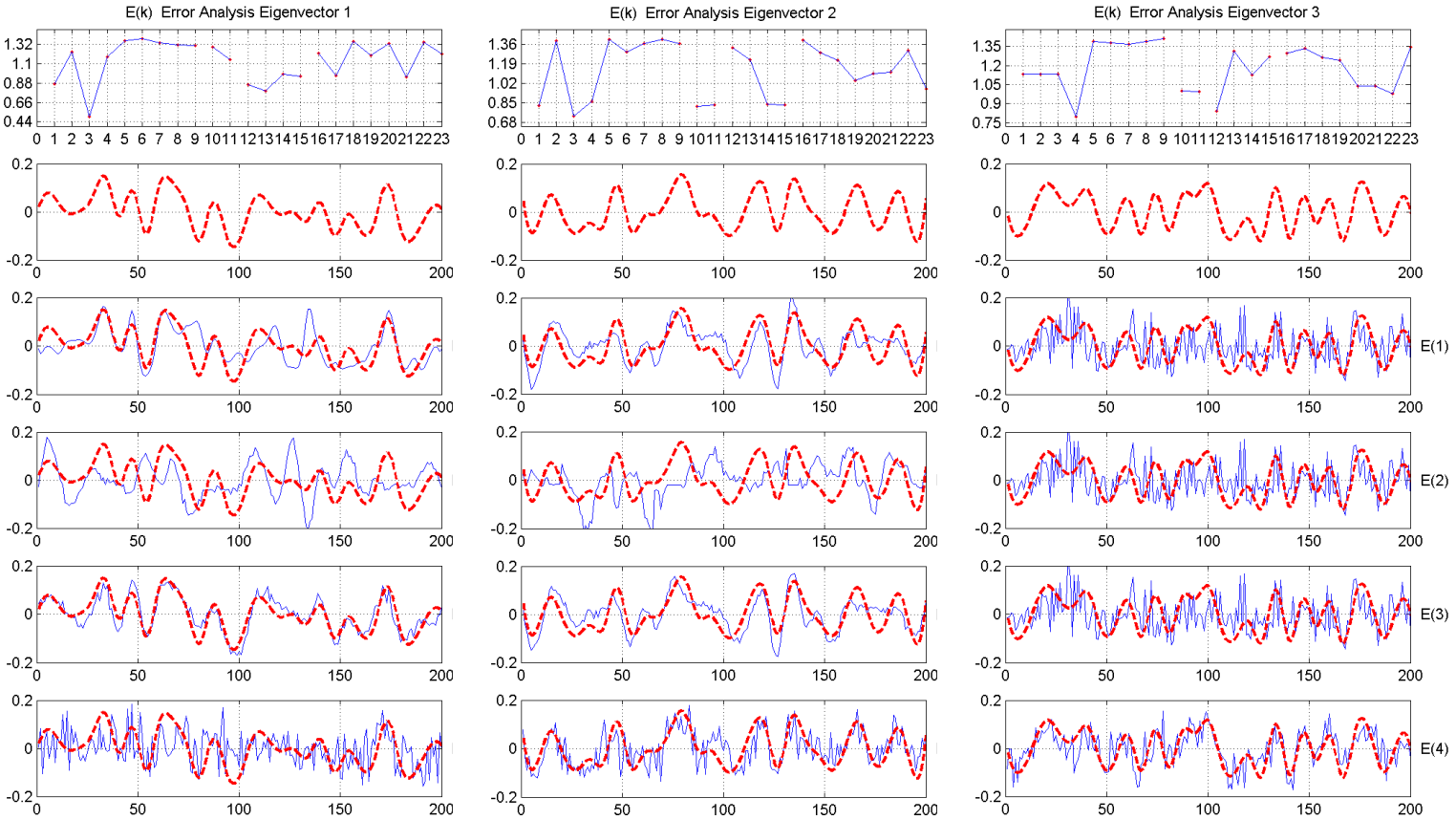


Figure 36: Error Analysis: Example 4

In Figure 36, the red dotted curve represents the reference signal, and the blue curves are the corresponding approximations. In this error analysis we observe, contrary to the previous Example 3, a significant improvement when using dimensional reduction techniques in the cases of the Eigenvector 1 and 2, and the error value $E(3)$. We have a similar improvement for the Eigenvector 3, corresponding to the parameterization t^3_{ac}, and the error value $E(4)$ with respect to the non dimensional reduced values $E(k), k = 16, \ldots, 23$.

$E(3)$	$E(12)$	$E(13)$	$E(14)$	$E(15)$
0.5011	0.8650	0.7934	0.9802	0.9597

Figure 37: Error values: Eigenvector 1, case $E(3)$

$E(3)$	$E(12)$	$E(13)$	$E(14)$	$E(15)$
0.7339	1.3302	1.2224	0.8382	0.8350

Figure 38: Error values: Eigenvector 2, case $E(3)$

$E(4)$	$E(16)$	$E(17)$	$E(18)$	$E(19)$	$E(20)$	$E(21)$	$E(22)$	$E(23)$
0.7991	1.2966	1.3356	1.2656	1.2430	1.0329	1.0388	0.9808	1.3447

Figure 39: Error values: Eigenvector 3, case $E(4)$

The behavior we see in this example illustrates the potential advantages we are attempting to understand of this combination of wavelet transforms and dimensional reduction techniques. Even though we have a mixture of examples indicating that a better comprehension and analysis is required in order to identify when and how this combination provides a significant approximation improvement, the tendency we have seen in some of this experiments, suggest that interesting perspectives.

Conclusions and Overview

The objective of the toy framework we have presented is to illustrate some perspectives combining dimensional reduction techniques with Fourier and wavelet transformations. The input of our problem is a set of signals $U = \{u_k\}_{k=1}^m \subset \mathbb{R}^n$, with the assumption that there is a low dimensional parameterization space $\Omega \subset \mathbb{R}^d$, with $d < n$, (i.e. there exist a continuous map $\mathcal{A} : \Omega \to U$ relating the low and high dimensional representations). In our particular example, the parameter space Ω is constructed as the image of a function $t_{ac} : I_m = \{1, \ldots, m\} \to \mathbb{R}^d$, with $\Omega = t_{ac}(I_m)$. The main problem is to estimate an approximation of the parameterization space Ω, given only the dataset U. The strategy we analyze is to construct an approximation $\Omega' = \mathcal{P}(\mathcal{T}(U))$, by combining a signal transformation $\mathcal{T}$ with a dimensional reduction method $\mathcal{P}$.

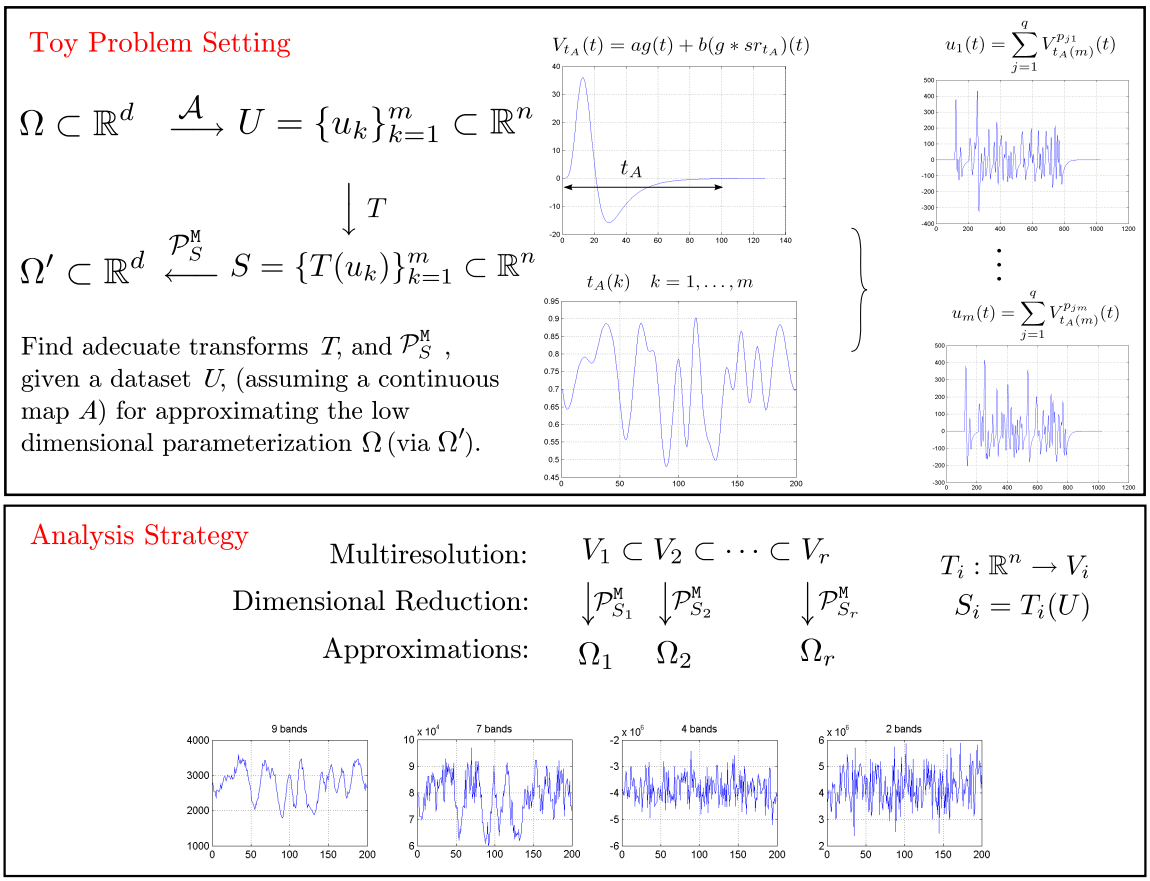


Figure 40: Overall structure of the toy-framework and some properties

The motivation behind this toy-framework is to study alternative analysis strategies for EMG signals, where one particular problem is to estimate a fatigue measure, represented in our examples as the parameterization function t_{ac}. The principal objective is to experiment with this combination of signal transformations and dimensional reduction methods. A main observation is that for a particular set of examples, a combination of multiresolution decomposition with dimensional reduction methods allows to improve the approximation of the parameter function t_{ac}. Further analysis is required for a better understanding of the properties of this framework, but the empirical results obtained with these examples might lead to a more robust and conceptual approximation scheme.

References

[1] M. An and R. Tolimieri. *Time-Frequency Representations*. Birkhäuser, 1997.

[2] G. Bachman, L. Narici, and E. Beckenstein. *Fourier and Wavelet Analysis*. Springer-Verlag. New York, Inc, 2000.

[3] M. Belkin and P. Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation, 2003.

[4] A. Böttcher. C^*-algebras in numerical analysis. *Irish Math. Soc. Bulletin*, 45:57–133, 2000.

[5] DS Broomhead and M. Kirby. A New Approach to Dimensionality Reduction: Theory and Algorithms. *SIAM Journal on Applied Mathematics*, 60:2114, 2000.

[6] A. Brun. *Manifold learning and representations for image analysis and visualization*. Linköping University, Department of Biomedical Engineering, February 2006.

[7] A. Brun, C. Westin, M. Herberthsson, and H. Knutsson. Sample logmaps: Intrinsic processing of empirical manifold data. *Proceedings of the (SSBA) Symposium on Image Analysis*, 1, 2006.

[8] A. Brun, C.-F. Westin, M. Herberthson, and H. Knutsson. Fast manifold learning based on riemannian normal coordinates. In *Proceedings of the SCIA;05*, pages 920–929, Joensuu, Finland, June 2005.

[9] F. Buchthal and H. Schmalbruch. Motor unit of mammalian muscle. *Physiological Reviews*, 60(1):90–142, 1980.

[10] G.S. Chirikjian and A.B. Kyatkin. *Engineering Applications of Noncommutative Harmonic Analysis: With Emphasis on Rotation and Motion Groups*. CRC Press, 2000.

[11] C. K. Chui. *An Introduction to Wavelets*. Academic Press, 1992.

[12] R. H. Clewley, J. M. Guckenheimer, and F. J. Valero-Cuevas. Estimating Effective Degrees of Freedom in Motor Systems. *IEEE Transactions on Biomedical Engineering*, 55(2):430–442, 2008.

[13] James W. Cooley and John W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19:297–301, 1965.

[14] T.F. Cox and M.A.A. Cox. *Multidimensional Scaling*. Chapman & Hall/CRC, 2001.

[15] I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial Mathematics, 1992.

[16] C.J. De Luca. The use of surface electromyography in biomechanics. *J Appl Biomech*, 13(2):135–63, 1997.

[17] J. Dieudonné. *Éléments D'Analyse. Tome 1*. Paris, Gauthier-Villars, 1971.

[18] J. Dieudonné. *Éléments D'Analyse. Tome 3.* Paris, Gauthier-Villars, 1974.

[19] J. Dieudonné. *Éléments D'Analyse. Tome 6.* Paris, Gauthier-Villars, 1975.

[20] J. Dieudonné. *Panorama des mathématiques pures, le choix bourbachique.* Gauthier-Villars, 1979.

[21] D.L. Donoho and C. Grimes. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences of the United States of America*, 100(10):5591, 2003.

[22] K. Dykema and N. Strawn. Manifold structure of spaces of spherical tight frames. `arXiv:math/0307367`, 2003.

[23] S. Egner. *Zur algorithmischen Zerlegungstheorie linearer Transformationen mit Symmetrie.* PhD thesis, Universität Karlsruhe, 1997.

[24] S. Egner, J. Johnson, D. Padua, M. Püschel, and J. Xiong. Automatic Derivation and Implementation of Signal Processing Algorithms t. *Urbana*, 51:61801, 2001.

[25] S. Egner and M. Puschel. Symmetry-based matrix factorization. *Journal of Symbolic Computation*, 37(2):157–186, 2004.

[26] K. Englehart, B. Hudgins, P.A. Parker, and M. Stevenson. Classification of the myoelectric signal using time-frequency based representations. *Medical Engineering and Physics*, 21(6-7):431–438, 1999.

[27] G. B. Folland. *A course in abstract harmonic analysis.* CRC Press, Boca Ratón, 1995.

[28] A. Grossmann, J. Morlet, and T. Paul. Transforms associated to square integrable group representations i: General results. *J. Math. Phys.*, 26(10), 1985.

[29] R. Hagen, S. Roch, and B. Silbermann. *C*-Algebras and Numerical Analysis.* Marcel Dekker Inc, 2000.

[30] T. Heimburg and A.D. Jackson. On soliton propagation in biomembranes and nerves. *Proceedings of the National Academy of Sciences*, 102(28):9790–9795, 2005.

[31] T. Heimburg and A.D. Jackson. On the action potential as a propagating density pulse and the role of anesthetics. `arXiv:physics/0610117`, 2006.

[32] E. Hewitt and K. Ross. *Abstract harmonic analysis I.* Springer Verlag, Berlin, 1963.

[33] M. Holschneider. *Wavelets An Analysis Tool.* Clarendon Press Oxford, 1995.

[34] A. Iske. *Multiresolution Methods in Scattered Data Modelling.* Springer, 2004.

[35] S. Karlsson, J. Yu, and M. Akay. Enhancement of spectral analysis of myoelectric signals duringstatic contractions using wavelet methods. *Biomedical Engineering, IEEE Transactions on*, 46(6):670–684, 1999.

[36] A. A. Kirillov and A. D. Gvishiani. *Theorems and Problems in Functional Analysis*. Springer-Verlag, 1982.

[37] Wheeler KR, MH Chang, and KH Knuth. Gesture-Based Control and EMG Decomposition. *IEEE Transactions on Systems, Man and Cybernetics Part C, Applications and Reviews*, 36(4):503–514, 2006.

[38] R.T. Lauer, C. Stackhouse, P.A. Shewokis, B.T. Smith, M. Orlin, and J.J. McCarthy. Assessment of wavelet analysis of gait in children with typical development and cerebral palsy. *Journal of Biomechanics*, 38(6):1351–1357, 2005.

[39] L. H. Loomis. *An Introduction to Abstract Harmonic Analysis*. Van Nostrand, Princeton, 1953.

[40] KC McGill and ZC Lateva. A model of the muscle-fiber intracellular action potentialwaveform, including the slow repolarization phase. *Biomedical Engineering, IEEE Transactions on*, 48(12):1480–1483, 2001.

[41] KC McGill and Z.C.S. Xiao. A model of the muscle action potential for describing the leadingedge, terminal wave, and slow afterwave. *Biomedical Engineering, IEEE Transactions on*, 48(12):1357–1365, 2001.

[42] R. Merletti and P. Parker. *Electromyography: Physiology, Engineering, and Noninvasive Applications*. John Wiley & Sons, 2004.

[43] H. Meschkowski. *Hilbertsche Räume mit Kernfunktion*. Springer Berlin, 1962.

[44] P. A. Milder, F. Franchetti, J. C. Hoe, and M. Püschel. Discrete Fourier transform compiler: From mathematical representation to efficient hardware. CSSI Technical Report CSSI-07-01, Carnegie Mellon University, 2007.

[45] H.Q. Minh, P. Niyogi, and Y. Yao. Mercer's Theorem, Feature Maps, and Smoothing. *Proceedings of the 19th Annual Conference on Learning Theory (COLT)*, 2006.

[46] M. Nikolic. *Detailed Analysis of Clinical Electromyography Signals*. PhD thesis, Institut for Kliniske Neurofag og Psykiatri Panum Instituttet, Denmark, 2002.

[47] P. Niyogi, S. Smale, and S. Weinberger. Finding the homology of submanifolds with high confidence from random samples. *to appear, Discrete and Computational Geometry*, 2006.

[48] R. Opfer. Tight frame expansions of multiscale reproducing kernels in sobolev spaces. *Applied and Computational Harmonic Analysis*, 20:357–374, 2006.

[49] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6):559–572, 1901.

[50] Ed. J. P. Pier. *Development of Mathematics 1950-2000*. Birkhäuser, 2000.

[51] A. Rakotomamonjy and S. Canu. Frames, Reproducing kernels, regularization and learning. *Journal of Machine Learning Research*, 06:1485–1515, September 2005.

[52] D. N. Rockmore. The FFT - an algorithm the whole family can use. *Computing in Science and Engineering*, 2(1):60–64, January/February 2000.

[53] S.T. Roweis and L.K. Saul. Nonlinear Dimensionality Reduction by Locally Linear Embedding, 2000.

[54] R. Schaback. Native Hilbert spaces for radial basis functions I. *International Series of Numerical Mathematics*, 132:255–282, 1999.

[55] R. Schaback. A unified theory of radial basis functions: Native Hilbert spaces for radial basis functions II. *Journal of Computational and Applied Mathematics*, 121(1-2):165–177, 2000.

[56] R. Schaback. Kernel-Based Meschless Methods (Lecture Notes). Institut für Numerische und Angewandte Mathematik (NAM), Georg-August-Universität Göttingen, 2007.

[57] I. Sergii and A.J. Douglas. Non-linear dimensionality reduction of signaling networks. *BMC Systems Biology* `http://www.biomedcentral.com/1752-0509/1/27`, 2007.

[58] A. Shadrin. Approximation Theory (Lecture Notes). DAMTP University of Cambridge, 2005.

[59] J. Shawe-Taylor and N. Cristianini. *Kernel Methods for Pattern Analysis*. Cambridge University Press, June 2004.

[60] I. Steinwart, D. Hush, and C. Scovel. An explicit description of the reproducing kernel Hilbert spaces of Gaussian RBF kernels. *Information Theory, IEEE Transactions on*, 52(10):4635–4643, 2006.

[61] J.B. Tenenbaum, V. de Silva, , and J.C. Langford. A Global Geometric Framework for Nonlinear Dimensionality Reduction. *Science*, 290(5500):2319–2323, 2000.

[62] M.D. van der Laan. *Signal Sampling Techniques for Data Acquisition in Process Control*. PhD thesis, University of Groningen, Netherlands, 1995.

[63] K.R. Wheeler, M.H. Chang, and K.H. Knuth. Gesture Based Control and EMG Decomposition. *IEEE Transactions on Systems, Man and Cybernetics*, 1(11):1, 2005.

[64] H. Zha and Z. Zhang. Principal Manifolds and Nonlinear Dimension Reduction via Local Tangent Space Alignment. *SIAM Journal of Scientific Computing*, 26(1):313–338, 2004.

[65] H. Zha and Z. Zhang. Continuum Isomap for manifold learnings. *Computational Statistics and Data Analysis*, 52(1):184–200, 2007.

[66] A. Zomorodian and G. Carlsson. Computing persistent homology. *Discrete Comput. Geom.*, 33(2):249–274, 2005.

Erklärung

Hiermit erkläre ich, dass die vorliegende Arbeit von mir selbständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel erstellt wurde.

Hamburg, Juni 2008