Université Sidi Mohamed Ben Abdellah
Faculté des Sciences Dhar El Mahraz- Fès
Centre d'Etudes Doctorales
"Sciences et Technologies"

Formation Doctorale : Mathématique et Applications

Discipline :

Spécialité : Equations aux Dérivées Partielles

Laboratoire : Analyse Mathématique et Applications

THESE DE DOCTORAT

Présentée par

BRAHMI Ibtissam

Stability of Eigenvalues in Partial Differential Equations – Deepness Analysis In Prototypes Based Unsupervised Learning

Soutenue le devant le jury composé de :

Pr. Aomar anane	***Ecole nationale des Sciences Appliquées, Oujda***	***Président***
Pr. Rosanna Verde	***Università degli Studi della Campania Luigi Vanvitelli , Naples***	***Rapporteur***
Pr. Youssef Akdim	***Faculté des Sciences Dhar El Mahraz, Fès***	***Rapporteur***
Pr. Said Raghay	***Faculyé des Sciences et Techniques, Marrakech***	***Rapporteur***
Pr. Souad Wardi	***Faculté des Sciences Dhar El Mahraz, Fès***	***Examinateur***
Pr. Rajae Aboulaich	***Ecole Mohammadia des Ingénieurs, Rabat***	**Examinateur**
Pr. Ahmed Youssfi	***Faculté des Sciences Dhar El Mahraz, Fès***	**Examinateur**
Pr. Abdelfettah Touzani	***Faculté des Sciences Dhar El Mahraz, Fès***	**Directeur de thèse**
Pr. Younès Bennani	***Université Paris 13, Sorbonne Paris Cité, Paris***	**Codirecteur de thèse**

Année universitaire : 2018-2019

Résumé

Le travail de recherche présenté dans cette thèse concerne deux parties distinctes, la première a été élaborée au sein du Laboratoire d'Analyse Mathématique et Applications de la Faculté des Sciences Dhar El Mahraz et concerne l'analyse de stabilité des valeurs propres suivant les variations de l'exposant d'intégrabilité dans les équations aux dérivées partielles à exposant variable. Afin d'apporter des éléments de réponse pour cette problématique, nous proposons différentes contributions. Une première contribution consiste à montrer la stabilité de la valeur propre à gauche qui est en général une question très délicate vu le manque des injections fonctionnelles adéquates, nous avons contourné à cette problématique en investissant dans la géométrie du domaine. Les deux autres contributions consistent à l'étude de stabilité des valeurs propres associées aux opérateurs du p(.)-Laplacien et p(.)-Biharmonique, respectivement. Nous avons également exploité la notion de Gamma-convergence pour nous aider à montrer les différentes convergences variationnelles. La deuxième partie a été réalisée au sein du Laboratoire d'Informatique Paris Nord et concerne l'analyse de profondeur de certains algorithmes à bases de prototypes dans l'apprentissage non supervisé. Dans un premier temps, nous avons montré mathématiquement que la Factorisation Matricielle Nonnégative Multicouches est une architecture profonde, on s'est reposé sur l'inverse généralisé de Penrose pour accomplir la démonstration puis on a validé expérimentalement en se agissant sur l'ordre de représentation. Dans un deuxième temps, nous avons conçu une nouvelle architecture profonde multicouches basée sur une version probabiliste des cartes auto-organisatrices, on a utilisé l'indice de Jensen-Shanon et l'indice de Silouhette pour valider notre nouvelle approche.

Mot clés: Equations aux Dérivées Partielles, Valeurs Propres, Stabilité, Factorisation Matricielle Nonnegative, Cartes Auto Organisatrices, Apprentissage Profond.

Abstract

The research work presented in this thesis concerns two distinct parts, the first was elaborated within the Laboratory of Mathematical Analysis and Applications of the Faculty of Sciences Dhar El Mahraz and concerns the analysis of stability of variational eigenvalues according to the variations of the exponent of integrability in partial differential equations with variable exponent. In order to provide elements of response to this problem, we propose various contributions. A first contribution is to show the stability of the eigenvalue on the left which is in general a very delicate issue given the lack of adequate functional injections, we have circumvented this problem by investing on the geometry of the domain. The other two contributions. The other two contributions consist on the study of the stability of eigenvalues associated with p(.)-Laplacien and p(.)-Biharmonique operators, respectively. We also exploited the notion of Gamma-convergence to help us show the different variational convergences. The second part was carried out within the Paris Nord Computer Laboratory and concerns the in-depth analysis of certain prototypes based on models in unsupervised learning, At first, we showed mathematically that the Multi-layer Nonnegative Matrix Factorisation is a deep architecture, we relied on the generalized inverse of Penrose to complete the demonstration and then we validated experimentally by acting on the order of representation. In a second step, we designed a new deep built-up architecture based on a probabilistic version of the self-organizing maps, we used the Jensen-indexShanon and the Silouhette index to validate our new approach.

Keywords: Partial Differential Equations, Eigenvalues, Stability, Nonnegative Matrix Factorization, Self Organizing Maps, Deep Learning.

List of Publications

International Journals:

Brahmi, I., Elkhalil, A., Touzani, A, Stability Of Eigenvalues Of A Nonlinear Variational Problem Involving p(X)Growth Condition With Varying p, Italian Journal of Pure and Applied Mathematics, 42, 2019, 449-457.

Brahmi, I., Elkhalil, A., Touzani, A. Stability of Eigenvalues of a Fourth Order Partial Differential Equation. International Journal of Mathematical Analysis, 13(3) 2019, 123-133.

Brahmi I., Cabanes G. Bennani Y. and Basarab M. Learning Useful Representations Through a Stacked Self Organizing Maps, Proceeding of International Joint Conference on Neural Networks (IJCNN), 2018, 1-8.

Brahmi I., Cabanes G. Bennani Y. and Touzani A. Deepness Analysis of Learned Factors in Multilayer NMF, Accepted in Australian Journal of Intelligent Information Processing Systems.

International Conferences:

Brahmi, I., Cabanes, G., Bennani Y.: Learning useful Representations Through Stacked Self-Organizing Maps. International Joint Conference on Neural Networks (IJCNN '18), Rio de Janeiro, Brazil.

Brahmi, I., Elkhalil, A., Touzani, A. : Stability of the mth variational eigenvalue for the p(x) Laplacian with varying p, 7ème édition du colloque bisannuel "Apprentissage Artificiel Fouille de Données" (AAFD) et Les 23èmes Rencontres annuelles de la Société Francophone de Classification (SFC) (AAFD SFC 2016)

Brahmi, I., Elkhalil, A., Touzani, A.: Continuity of The First Eigenvalue of The P(X) Laplacian Problem, 5ème congrès international de la Société Marocaine des Mathématiques Appliquées Février 2017.

Brahmi, I., Elkhalil, A., Touzani, A.: Continuity of The First Eigenvalue of The P(X)-Biharmonic Problem. Les quatrièmes journées d'analyse non linéaire, Oujda, 09-10 Décembre 2016

Contents

Part I

Stability Analysis of Associated Eigenvalues to Partial Differential Equations

Introduction

Our understanding of the phenomena of real world and our technology today are largely based on Partial Differential Equations, which will be abbreviated PDE in the following. Thanks to the modeling of these phenomena through PDE that we could understand the role of this or that parameter, and especially get forecasts sometimes extremely accurate. The mathematical study of PDEs has also given us learned to show a little modesty: we discovered the impossibility to predict in the medium term certain phenomena governed by nonlinear PDEs - think of the now famous butterfly effect: a small variation in initial conditions can in very long time lead to very large variations. On the other hand, we also learned to "hear the shape of a drum": we have mathematically demonstrated that the frequencies emitted by a drum during the vibration of the membrane, a phenomenon described by a PDE, allow to reconstitute perfectly the shape of the drum.

One important thing we need to keep in mind about PDEs is that we don't usually matter about getting their solutions explicitly! What mathematics can do on the other hand, is to say if one or more solutions exist, and sometimes very accurately describe certain properties of these solutions.

Indeed, after a qualitative description of physical phenomena by specialists (engineers, chemists, biologists...), we're on modeling, from the collected data, we have all the ingredients to write our mathematical model, we suppose in our case that the concrete problem is described by a PDE.

We move after that to analyze mathematically the eventual existing solution, discuss its uniqueness, continuity and other properties.

We end by the numerical stage, at first we transform our continuous problem to a discrete one, the notions of continuity in time and space are replaced by discrete ones, that the problem can be turned to a finite dimensional problem, to could write the numerical scheme, choice the numerical method and do numerical analysis to evaluate how well we've done, how much the discrete problem is close to the continuous one, and end by coding the method, implement it and analyze the results.

Our contributions in this thesis concern the mathmeatical analysis phase, where we study the continuity of solutions for a class of PDEs.

When did the PDEs appear? They were probably formulated for the first time at the birth of the rational mechanics during 17^{th} century (Newton, Leibniz ...). Then the "catalog" of the PDE was enriched as science develops and, in particular, physics.
If we must remember only few names, we must mention at first Euler, then Navier and Stokes, for the equations of fluid mechanics, those of Fourier for the equation of heat, Maxwell for those Electromagnetism, Schrödinger and Heisenberg for the equations of the quantum mechanics, and of course Einstein for the PDEs of theory of relativity.
However, the systematic study of PDEs is much more recent, and only during the 20^{th} century that mathematicians began to develop the necessary arsenal. A step forward was made by L. Schwartz when he introduced the distribution theory (around the 1950s), and at least comparable progress is to L. Hörmander for the development of pseudo-differential calculus (in the early 1970s). It is certainly good to keep in mind that PDEs studies remain a very important area of active research at the present century. Moreover, these studies don't only have a resonance in the applied science, but also play a very important role in the current development of mathematics itself, both in geometry and analysis.

The study of eigenvalues problems, our main concern in this part of thesis, finds its roots in the method of separation of variables.
Let's consider a class C of functions $u(t_1, ..., t_M, x_1, ..., x_N)$ with a couple of variables (x_i, t_i), we suppose that each function of the class C satisfies:

(a) A function in C is defined however $(t_1, ..., t_M)$ belongs to a set T and $(x_1, ..., x_N)$ belongs to a set X.

(b) There exists a class C_T of functions $v(t_1, ..., t_M)$ defined on T and a class $C_X$of functions $w(x_1, ..., x_N)$ defined on X such that if $v \in C_T$ and $w \in C_X$ then $vw \in C$.

(c) Each function u of C is a linear combination of products $v_i w_i$ with $v_i \in C_T$ and $w_i \in C_X$. (We say that C is tensorial product of C_T and C_X).

Let's suppose that we have two linear operators M and N defined on C, that could be factorized in the preceding manner, there exists two operators R and S defined on C_T and two operators A

and B defined on C_X such that for $v \in C_T$ and $w \in C_X$

$$M[vw] = R[v]B[w],$$

$$N[vw] = S[v]A[w].$$

We can write

$$M = RB, \ et \ \ N = SA.$$

We're looking for particular solutions of the equation

$$RBu + SAu = 0. \tag{1}$$

Assume that we can find a constant λ and an element $w \in C_X$ satisfying

$$Aw = \lambda Bw. \tag{2}$$

So the product $u = vw$ is a solution of (1) if and only if

$$RBwv + SAwv = 0.$$

$$RBwv + \lambda SBwv = 0.$$

In other words

$$Rv + \lambda Sv = 0. \tag{3}$$

Thus the problem (1) is decomposed on two problems defined on two smallest subspaces.
A value λ for which (2) have a non zero solution w is called **eigenvalue** and w is called the corresponding **eigenfunction**.

In physics, the eigenvalues are generally related to vibrations. The strings of violins, drums, bridges can still vibrate at a certain frequency. For more illustration, let's consider a simple example which is reduced to an eigenvalue problem. Let's suppose that we have a homogeneous chord of length l fixed at the ends and not undergoing any external force. Let the origin be one of the extremities and the X axis is directed along the string. A function $u(x;t)$ that describes small oscillations of the string is a solution of the homogeneous differential equation

$$\frac{\partial^2 u}{\partial t^2} - a^2 \frac{\partial^2 u}{\partial x^2} = 0 \tag{4}$$

With Dirichlet conditions $u(0;t) = 0; \quad u(l;t) = 0$ and $a \neq 0$.

The movement of the string is also defined by initial conditions. Suppose the solution can be written in the form $u(x,t) = X(x)T(t)$, where $T(t)$ is called oscillation law and it describes the character of movement of string individual points and $X(x)$ describes the string shape in different moments.

It's clear that $X(x)$ should satisfy boundary conditions $X(0) = 0 = X(l)$.

To obtain equations satisfied by $X(x)$ and $T(t)$, we replace the decomposition of $u(x,t)$ in equation (4), we obtain

$$X(x)T''(t) = a^2T(t)X''(x)$$

By separating the variables we obtain

$$\frac{T''(t)}{a^2T(t)} = \frac{X''(x)}{X(x)}$$

Since the left hand term depends only on t and the right hand term depends only on t, the two sides are equal to the same constant. We note the constant $-\lambda$ ($\lambda \neq 0$)

$$\frac{T''(t)}{a^2T(t)} = \frac{X''(x)}{X(x)} = -\lambda$$

So

$$T'' + \lambda a^2 T = 0, \; X'' + \lambda X = 0.$$

Therefore we have a simple eigenvalue problem. It's easy to see that the constant λ can take only the values

$$\lambda_n = \frac{\pi^2 n^2}{l^2} (n = 1, 2, 3, ...).$$

For the string with fixed ends, the waves shape is described by

$$X_n(x) = c \; sin\left(\frac{n\pi x}{l}\right), \;\; c = const.$$

To find functions $T_n(t)$ we replace λ_n with its values in T equation:

$$T_n'' + \frac{a^2\pi^2}{l^2} n^2 T_n = 0$$

The general solution of this equation has the form

$$T_n(t) = B_n sin\left(\frac{n\pi a}{l}\right) t + C_n cos\left(\frac{n\pi a}{l}\right) t = A_n sin\left(\frac{n\pi a}{l}\right) t + \varphi_n),$$

where B_n, C_n, A_n and φ_n are arbitrary constants.

The general solution is presented as

$$u(x;t) = (\frac{1}{\pi an} Esin\left(\frac{n\pi a}{l}\right) t + D\left(\frac{n\pi a}{l}\right) t) sin\left(\frac{n\pi x}{l}\right)$$

Eigenvalues and eigenfunctions often have physical interpretations: in the preceding example, eigenvalues λ_n define the frequency harmonic oscillations of the string and eigenfunctions X_n define the amplitudes of these oscillations.

Function spaces, in particular L^p spaces, play a central role in many questions in analysis. The special importance of L^p spaces may be said to derive from the fact that they offer a partial but useful generalization of the fundamental L^2 space of square integrable functions. In order of logical simplicity, the space L^1 comes first since it occurs already in the description of functions integrable in the Lebesgue sense. Connected to it via duality is the L^∞ space of bounded functions, whose supremum norm carries over from the more familiar space of continuous functions. Of independent interest is the L^2 space, whose origins are tied up with basic issues in Fourier analysis. The intermediate L^p spaces are in this sense an artifice, although of a most inspired and fortuitous kind.

The current high level of interest in such spaces stems largely from the fact that they provide a natural setting for the study both of partial differential equations with coefficients having variable rates of growth and regularity of solutions of variational problems involving integrals in which the integrand satisfies non standard growth conditions.
Conspicuous among such investigations in the work of [30] involving the modeling of electrorheological fluids. As might be expected, developments involving these spaces that are driven by the needs of such concrete problems have been accompanied by a good deal of pure theory designed to explore how close is the parallel between them and the classical spaces with constant exponent p. A notable use of continuity of eigenvalues is in bifurcation theory, exactly when we need to show that the first eigenvalue is a bifurcation point of the associated partial differential equation. For the general functional framework, let's consider Lebesgue spaces $L_p(\Omega)$ and Sobolev spaces

modeled on them: the spaces of variable integrability mentioned in this thesis appear in natural way when the constant p is replaced by a function on the underlying Ω, Our aim in this thesis is to present some stability results of eigenvalue problems associated with uniform variations of the variable exponent.
In this first part of this thesis we combine the two preceeding notions by treating partial differential equations involving eigenvalues in variable exponent spaces framework:

- In the first chapter, we present the general functional framework in which our problems are defined, we start by introducing the set of variable exponents, the geometric properties of our spacial domain, which is an important issue since the solutions behave differently depending on those properties, we then go to define the spaces with variable exponent, present the functional analytic properties, from Sobolov embeddings, to compact embeddings, without forgetting the Poincaré's inequality. We then introduce the topic of De Giorgi's $\Gamma-$convergence, this particular type of variational convergence that plays a central role for its compactness properties and for the large number of results concerning $\Gamma-$Limits of integral functions.

  We end this chapter by introducing a well established formula for defining a non-decreasing unbounded sequence of critical values of a convex $p-$homogeneous functional F, defined on some Banach space X, along the one-codimensional manifold $M = G^{-1}(\{1\})$, where G is another convex and $p-$homogeneous functional on X.

- In the second chapter, we study the stability of the variational eigenvalues with respect to uniform variations of the exponent associated to the problem:

$$(\mathcal{V}_{p(.)}) \quad \begin{cases} -\mathrm{div}\left(p(x)\left|\frac{\nabla u}{K(u)}\right|^{p(x)-2}\frac{\nabla u}{K(u)}\right) = \lambda S(u)p(x)\left|\frac{u}{k(u)}\right|^{p(x)-2}\frac{u}{k(u)} \\ u \in W_0^{1,p(.)}(\Omega), \end{cases},$$

  where $S(u) := \dfrac{\int_\Omega |\frac{\nabla u}{K(u)}|^{p(x)}dx}{\int_\Omega |\frac{u}{k(u)}|^{p(x)}dx}$, $K(u) := \|\nabla u\|_{p(.)}$ and $k(u) := \|u\|_{p(.)}$. Here $\|.\|_{p(.)}$ is the Luxemburg norm defined in Chapter 1.

  The above problem is in fact the Euler-Lagrange equation corresponding to the minimum of the Rayleigh quotient derived from norms, say: $R(u) = \frac{\|\nabla u\|_{p(.)}}{\|u\|_{p(.)}}$ Along this thesis,

we define the eigenvalues following Lusternik–Schnirelmann theory, and then study their behaviour when $p(.)$ is varied.

- Chapter 3 is concerned with the study of the stability of the first eigenvalue associated to the second order eigenvalue problem defined by the $p(.)-$Laplacian operator

$$\begin{cases} -\Delta_{p(x)}(u) = \lambda |u|^{p(x)-2}u & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases}$$

The associated variational problem is:

$$\int_\Omega |\nabla u|^{p(x)-2}\nabla u \nabla v \, dx = \lambda \int_\Omega |u|^{p(x)-2}uv \, dx \; \forall v \in W_0^{1,p(x)}(\Omega) \tag{5}$$

Which is the Euler-Lagrange associated to the minimization of: $\frac{\int_\Omega |\nabla u(x)| \, dx}{\int_\Omega |u(x)| \, dx}$

- The last chapter focuses on the study of the dependence on $p(.)$ of the first eigenvalue associated to the fourth order eigenvalue problem defined by the $p(.)-$Biharmonic operator

$$\begin{cases} \operatorname{div}(|\Delta u|^{p(x)-2}\Delta u) = \lambda |u|^{p(x)-2}u \;\; in \; \Omega & ; \\ u = \Delta u = 0 \qquad\qquad\qquad\qquad on \; \partial\Omega & , \end{cases} \tag{6}$$

The associated variational problem is:

$$\int_\Omega |\Delta u|^{p(x)-2}\Delta u \Delta v \, dx = \lambda \int_\Omega |u|^{p(x)-2}uv \, dx, \tag{7}$$

$$\forall v \in X := W_0^{1,p(x)}(\Omega) \cap W^{2,p(x)}(\Omega)$$

Which is the Euler-Lagrange associated to the minimization of: $\frac{\int_\Omega |\Delta u(x)| \, dx}{\int_\Omega |u(x)| \, dx}$

We close The part by a conclusion and somes perspectives.

CHAPTER 1

A Global Framework for Functional Spaces and Variational Convergence

Overview

In recent years, the nonlinear problems with variable exponential growth is a new research field that drew the interest of many mathematical researcher. The principal interest of these problems come mainly from their applications, a few examples to cite are image processing (remove noise, edge detection and image restoration) or the modelisation of the movement for electrorheological fluids. Other domain are in wide investigation of analytical results (finance, medicine, chemistry, quantum mechanics...).
This chapter is devoted to introduce some interesting definitions and properties, which are essential to define the problems studied in this thesis and to understand the used technics to prove the main results.

1.1 Variable Exponent Lebesgue and Sobolev Spaces

1.1.1 Variable Exponent Lebesgue Space

In this chapter we provide basic information about modular and spaces Muselak-Orlicz spaces which provide the framework for a variety of different function spaces, including classical Lebesgue and Orlicz spaces and variable exponent Lebesgue spaces.
For the investigation of Lebesgue spaces it's enough to stay in the framework of Banach spaces. In particular, the space and its topology is described in terms of a norm. However, in the context of Variable exponent spaces this is not the best way. Instead it's better to start with the so called modular which then includes a norm.
Let X be a linear space over $\mathbb{K}$ ($\mathbb{R}$ or $\mathbb{C}$). A map $\rho : X \to [0, \infty]$ is called a **convex modular** on X if

1. $\rho(x) = 0$ if and only if $x = 0$;
2. $\rho(\lambda x) = \rho(x)$ for all $x \in X$ and all $\lambda \in \mathbb{K}$ with $|\lambda| = 1$.
3. ρ is convex: $\rho(\alpha x + \beta y) \leq \alpha\rho(x) + \beta\rho(y)$ for all $x, y \in X$ and all $\alpha, \beta \geq 0$ with $\alpha + \beta = 1$.

Note that 1. and 3. imply that for each $x \in X$, the map $\lambda \mapsto \rho(\lambda x)$ is non-decreasing on $[0, \infty)$; also,

$$\begin{aligned} \rho(\lambda x) &= \rho(|\lambda|x) \leq |\lambda|\rho(x) \quad \text{if } \lambda \leq 1, \\ \rho(\lambda x) &= \rho(|\lambda|x) \geq |\lambda|\rho(x) \quad \text{if } \lambda \geq 1. \end{aligned} \tag{1.1}$$

A convex modular ρ is said to be (left-,right-) continuous if, for every $x \in X$, the map $\lambda \mapsto \rho(\lambda x)$ is (left-, right-) continuous on $[0, \infty)$ (on $(0, \infty)$ for left-continuity).

Given any normed linear space $(X,\ \|.\|)$, it is clear that $\|.\|$ is a continuous convex modular on X. In the reverse direction, let X be a linear space and consider the subset X_ρ defined by

$$X_\rho = \{x \in X\ :\ \rho(\lambda x) < \infty \quad \text{for some } \lambda > 0\},$$

Where ρ is a convex modular on X. This subset is called a modular space.

Proposition 1.1.1 *Let ρ be a convex modular on a linear space X and define*

$$\|x\|_\rho := \inf\{\lambda > 0 :\ \rho(x/\lambda) \leq 1\}\ (x \in X_\rho).$$

Then:

1. $(X_\rho, \|.\|_\rho)$ *is a normed linear space*
2. *if* $\rho(x) \leq 1$, *then* $\|x\|_\rho \leq 1$
3. *if ρ is left continuous, then $\|x\|_\rho \leq 1$ if and only if $\rho(x) \leq 1$. If ρ is continuous, then $\|x\|_\rho < 1$ if and only if $\rho(w) < 1$; and $\|x\|_\rho = 1$ if and only if $\rho(x) = 1$.*

The norm $\|.\|_\rho$ is called the **Luxemburg norm** on X_ρ.

Any convex, continuous, strictly increasing function $\phi : [0, \infty) \to [0, \infty)$ such that $\phi(0) = 0$ and $\lim_{t\to\infty} \phi(t) = \infty$ is called an **Orlicz function**. Examples of such functions are provided by the maps with values $t^p (1 \leq p < \infty)$, $t \log(1 + t),\ \exp(t)$ at $t \in [0, \infty)$

Let Ω be a measurable subset of $\mathbb{R}^n$, $n \geq 2$, and let $\mathcal{M}(\Omega)$ be the set of all Lebesgue-measurable, scalar-valued functions on Ω. If ϕ is an Orlicz function, the function ρ defined by

$$\rho(f) = \int_\Omega \phi(|f(x)|)dx\ (f \in \mathcal{M}(\Omega))$$

is plainly a convex modular on $\mathcal{M}(\Omega)$.

The corresponding modular space, denoted by $L_\phi(\Omega)$ is called an Orlicz space:

$$L_\phi(\Omega) \;=\; \{f \in \mathcal{M}(\Omega) : \rho(\lambda f) < \infty \; for \; some \; \lambda > 0\};$$

The Luxemburg norm on $L_\phi(\Omega)$ is denoted by $\|.\|_{L_\phi(\Omega|)}$ and is given by

$$\|f\|_{L_\phi(\Omega)} \;=\; \inf\{\lambda > 0 : \int_\Omega \phi(|f(x)|)dx \leq 1\}.$$

A norm on $L_\phi(\Omega)$ equivalent to $\|.\|_{L_\phi(\Omega)}$ is given by means of the so-called **complementary function** (to ϕ). This is the function $\phi_* : [0,\infty) \to [0,\infty)$ defined by

$$\phi_*(s) = \sup\{st - \phi(t) : \; t \geq 0\};$$

it is assumed that ϕ has a right derivative that is zero at 0 and tends to infinity at infinity.

Then ϕ_* is a non-degenerate Orlicz function, corresponding to which the Orlicz norm $\|.\|_*$ on $L_\phi(\Omega)$ is defined by

$$\|f\|^*_{L_\phi(\Omega)} \;=\; \sup\left\{\left|\int_\Omega f(t)g(t)dt\right| \; : g \in L_{\phi_*}(\Omega), \int_\Omega \phi_*(|g(x)|)dx \leq 1\right\};$$

moreover,

$$\|f\|_{L_\phi(\Omega)} \leq \|f\|^*_{L_\phi(\Omega)} \leq 2\|f\|_{L_\phi(\Omega)}$$

Hölder's inequality holds in the form

$$\int_\Omega |f(t)g(t)|dt \leq \|f\|^*_{L_\phi(\Omega)}\|g\|_{L_{\phi_*}(\Omega)}$$

For all $f \in L_\phi(\Omega)$ and $g \in L_{\phi_*(\Omega)}$. The Orlicz spaces just introduced are particular cases of Orlicz-Musielak spaces, which we now define. Let $M : \; \Omega \times [0,\infty) \to [0,\infty)$ be such that $M(.,t)$ is measurable for every $t \geq$, and for almost all $x \in \Omega$, the function $M(x,.)$ is an Orlicz function. The Orlicz-Musielak space $L_M(\Omega)$ is defined by

$$L_M = \left\{f \in \mathcal{M}(\Omega) : \; \int_\Omega M(x,\lambda|f(x)|)dx < \infty \; for \; some \; \lambda > 0\right\};$$

endowed with the norm

$$\|f\|_{L_M(\Omega)} := \inf\left\{\lambda > 0 : \int_\Omega M(x, \lambda|f(x)|)dx \leq 1\right\},$$

it is a Banach space (see [23]).

It's now concenient to introduce the notion of a Banach space. A linear space $X \in \mathcal{M}(\Omega)$ is called a Banach function space if there is a map $\|.\|_X : \mathcal{M}(\Omega) \to [0, \infty)$ with the properties of a norm and such that

1. $f \in X$ if and only if $\|f\|_X < \infty$
2. $\|f\|_X = \||f|\|_X$ for all $f \in \mathcal{M}(\Omega)$;
3. if $0 \leq f_k \uparrow f$, then $\|f_k\|_X \uparrow \|f\|_X$;
4. if $E \subset \Omega$ and $|E| < \infty$, then $\|\chi_E\|_X < \infty$
5. if $E \subset \Omega$ and $|E| < \infty$, there is a constant $c(E)$ such that for all $f \in X$,

$$\int_E |f(x)|dx \leq c(E)\|f\|_X$$

Every Banach function space X is a Banach space when endowed with the norm $\|.\|_X$ (see [4]). It is a routine to check that the classical Lebesgue and Sobolev spaces are Banach function spaces. While every Orlicz space $L_\phi(\Omega)$ is a rearrangement-invariant (see [1] and [19]), this is not the case, in general, for the Orlicz-Musielak spaces $L_M(\Omega)$.

Given a Banach function space X, the set

$$X' := \left\{f \in \mathcal{M}(\Omega) : \int_\Omega |f(x)g(x)|dx < \infty \; for\; all\; g \in X\right\},$$

furnished with the norm

$$\|f\|'_X := \sup_{g \in S_X} |f(x)g(x)|\, dx,$$

where S_X is the unit sphere is a Banach function space called the **associate space** of X. Hölder's inequality holds in the form

$$\int_\Omega |f(x)g(x)|\, dx \leq \|f\|_X \|g\|_{X'}$$

for all $f \in X$ and $g \in X'$. Every Banach function space X coincides with its associate space X'', and $\|f\|_X = \|f\|_{X''}$ for all $f \in X$. In general, the associate space of X is (canonically isomorphic to) a closed subspace of the dual X^*. It is natural to investigate the circumstances under which X' coincides with X^*. To do this, we say that $f \in X$ has **absolutely continuous norm** if, for every decreasing sequence $G_{k\in\mathbb{N}}$ of subsets of Ω with $|G_k| \to 0$, we have $\|f\chi_{G_k}\|_X \to 0$; if every $f \in X$ has this property, X is said to have absolutely continuous norm.

We now turn to the particular type of Banach function space with which this thesis is concerned. Let Ω be a bounded set of $\mathbb{R}^N$, $N \geq 2$, we denote by $\mathcal{P}(\Omega)$ the class of variable exponents $p : \overline{\Omega} \to \mathbb{R}^+$ continuous and satisfying

$$1 \leq p^- := \inf_{x\in\overline{\Omega}} p(x) \leq p^+ := \sup_{x\in\overline{\Omega}} p(x) < \infty.$$

The conjugate function p' is defined by

$$p'(x) = \begin{cases} \frac{p(x)}{p(x)-1} & \text{if } 1 < p(x) < \infty, \\ 1 & \text{if } p(x) = 1. \end{cases}$$

It is easy to check that the function $\rho_{p(.),\Omega} : \mathcal{M}(\Omega) \to [0,\infty)$ defined by

$$\rho_{p(.),\Omega}(u) = \int_\Omega |u(x)|^{p(x)} dx$$

is a left-continuous convex modular on $\mathcal{M}(\Omega)$, if there is no confusion, we will write $\rho_{p(.)}(u)$ instead of $\rho_{p(.),\Omega}(u)$. The corresponding modular space is denoted by $L^{p(.)}(\Omega)$ and is called a **generalized Lebesgue space** or a **Lebesgue space with variable exponent**. Thus

$$L^{p(.)}(\Omega) = \{u \in \mathcal{M}(\Omega) : \ \rho_{p(.)}(u) < \infty\}$$

Note that $L^{p(.)}(\Omega)$ is an Orlicz-Musielak space with defining function M given by

$$M(x,t) = t^{p(x)}$$

Proposition 1.1.2 *Let $p \in \mathcal{P}(\Omega)$. Then $\|u\|_{p(.)} \leq 1$ if and only of $\rho_{p(.)}(u) \leq 1$; moreover if $\|u\|_{p(.)} \leq 1$, then $\rho_{p(.)}(u) \leq \|u\|_{p(.)} \leq 1$.*

Proposition 1.1.3 *Let $p \in \mathcal{P}(\Omega)$. Then for all $u \in L^{p(.)}(\Omega)$ we have*

$$\min\{\|u\|_{p(.)}^{p_-}, \|u\|_{p(.)}^{p^+}\} \leq \rho_{p(.)}(u) \leq \max\{\|u\|_{p(.)}^{p_-}, \|u\|_{p(.)}^{p^+}\}$$

Hölder's inequality is valid in the following form.

Theorem 1.1.4 *If $p \in \mathcal{P}(\Omega)$, then for all $u \in L^{p(.)}(\Omega)$ and $g \in L^{p'(.)}(\Omega)$*

$$\int_\Omega |f(x)g(x)|dx \leq r_p \|f\|_p \|g\|_{p'}$$

where r_p is a constant that depends on Ω, p^- and p^+.

Under the condition $p \in \mathcal{P}(\Omega)$, $L^{p(.)}(\Omega)$ is a reflexive and uniformly convex Banach function space.

1.1.2 Sobolev Spaces Variable Exponent

This section is intended to provide a vrief summary of the main properties of Soboloev spaces with variable integrability. In addition to be a measurable subset of $\mathbb{R}^N$, in this section Ω is a bounded domain with a Lipschitz boundary.
As is customary, $\nabla u = \left(\frac{\partial}{\partial x_j} u\right)_{1 \leq j \leq N}$ stands for the (weak) gradient of $u \in L^{p(.)}(\Omega)$.

Definition 1.1.5 *The Sobolev space $W^{1,p(.)}(\Omega)$ associated to the variable exponent p is defined as*

$$W^{1,p(.)}(\Omega) = \{u \in L^{p(.)}(\Omega) : |\nabla u| \in L^{p(.)}(\Omega)\},$$

endowed with the norm

$$\|u\|_{W^{1,p(.)}(\Omega)} = \|u\|_{1,p(.)} = \|u\|_{p(.)} + \|\nabla u\|_{p(.)}$$

To define the closure of the space of infinitely differentiable functions with compact support in Ω over $W^{1,p(.)}(\Omega)$, denoted $\overline{C_c^\infty(\Omega)}^{W^{1,p(.)}(\Omega)}$, the exponent $p(.)$ should satisfy the $\log-$Hölder continuity condition: $|p(x) - p(y)| \leq -\frac{L}{\log(|x-y|)}$ for some $L > 0$ and for all $x, y \in \Omega$ with

$0 < |x - y| < 1/2$.

We will then denote by $\mathcal{P}^{\log}(\Omega)$ the set of $p(.) \in \mathcal{P}(\Omega)$ that are $\log -$Hölder continuous.

If $p(.) \in \mathcal{P}^{\log}(\Omega)$ then we can define $W_0^{1,p(.)}(\Omega)$ as the closure of $\overline{C_c^\infty(\Omega)}$ over $W^{1,p(.)}(\Omega)$ and the critical exponent by $p^*(x) = \frac{Np(x)}{N-p(x)}$ for $p(x) < N$.

Proposition 1.1.6 *(see [12, 15])*

(i) *Assuming* $1 < p_- \leq p_+ < \infty$, *the spaces* $W^{1,p(x)}(\Omega)$ *and* $W_0^{1,p(x)}(\Omega)$ *are separable and reflexive Banach spaces.*

(ii) *If* $q(\cdot) \in C_+(\bar{\Omega})$ *and* $q(x) < p^*(x)$ *for any* $x \in \Omega$, *then the embedding* $W_0^{1,p(x)}(\Omega) \hookrightarrow\hookrightarrow L^{q(x)}(\Omega)$ *is continuous and compact.*

(iii) *Poincaré inequality : there exists a constant* $C > 0$, *such that*

$$\|u\|_{p(x)} \leq C \|\nabla u\|_{p(x)} \qquad \forall u \in W_0^{1,p(x)}(\Omega).$$

(vi) *Sobolev inequality : there exists an other constant* $C > 0$, *such that*

$$\|u\|_{p*(x)} \leq C \|\nabla u\|_{p(x)} \qquad \forall u \in W_0^{1,p(x)}(\Omega).$$

Remark 1.1.7 *By (iii) of the Proposition 1.1.6, we deduce that* $\|\nabla u\|_{p(x)}$ *and* $\|u\|_{1,p(x)}$ *are equivalent norms in* $W_0^{1,p(x)}(\Omega)$.

The next result concerns an estimation of the Luxemburg norm of the characteristic function that we will need afterward.

Corollary 1.1.8 *Let* $p \in \mathcal{P}^{log}(\mathbb{R}^N)$. *Then*

$$\|\chi_Q\|_{p(.)} \approx \begin{cases} |Q|^{\frac{1}{p(x)}}, & \text{if } |Q| \leq 2^N \text{ and } x \in Q; \\ |Q|^{\frac{1}{p_\infty}}, & \text{if} |Q| \geq 1. \end{cases}$$

for every cube (or ball) $Q \subset \mathbb{R}^N$; p_∞ *is defined by* $\frac{1}{p_\infty} := \lim_{|x| \to \infty} \frac{1}{p(x)}$.

1.2 $\Gamma-$Convergence

Given a real valued functional F on a set X, one of the main problems of the calculus of variations is to find the minimum value

$$m_X(F) \;=\; \inf_{x\in X} F(x),$$

together with the set of all minimum points

$$M_X(F) \;=\; \{x \in X \;:\; F(x) \;=\; m_X(F)\}$$

The aim of $\Gamma-$Convergence is to study the dependence of $m_X(F)$ and $M_X(F)$ and on the data of the problem, i.e, on F and X, in particular when F or X undergo severe perturbations.
In particular, if we assume that X is fixed and we consider a sequence (F_h) of perturbations of F, which converges to F in a very strong way, then, in general, we can prove by elementary arguments that the minimum values of the functionals F_h converge to the minimum value of F. For instance, if (F_h) converges to F uniformly on X, then $(m_X(F_h))$ converges to $m_X(F)$.
These elementary results however are not really suitable for many applications to Physics and Engineering, characterizd by perturbations of minimum problems for integral functionals of the form

$$F(u) \;=\; \int_\Omega f(x, Du(x))\, dx, \tag{1.2}$$

where Ω is a bounded open subset of $\mathbb{R}^N$, $f : \Omega \times \mathbb{R}^N \to [0, \infty[$ is a function satisfying suitable structure conditions, and $Du :\; \Omega \to\; \mathbb{R}^N$ denotes the gradient of the unknown function
$u :\; \Omega \to\; \mathbb{R}$

Definition 1.2.1 *Let X be a metric space. A sequence F_h of functionals $(F_h) : X \to \overline{\mathbb{R}}$ is said to $\Gamma(X)$ converges to $F : X \to \overline{\mathbb{R}}$ and we write $\Gamma(X) - \lim_{h\to\infty} F_h = F$ if the following hold*

i/ (liminf inequality) For every $u \in X$ and $(u_h) \subset X$ such that $u_h \to u$ in X we have

$$F(u) \le \liminf_{h\to\infty} F_h(u_h)$$

ii/ (limsup inequality) For every $u \in X$ there exists a sequence $(u_h) \subset X$ (called recovery sequence) such that $u_h \to u$ in X and

$$F(u) \geq \limsup_{h\to\infty} F_h(u_h)$$

The condition ii/ is characterized by the following (see [20]).

Proposition 1.2.2 *Let X be a topological space that satisfies the 1^{st} axiom of countability and assume that (u_h) is a sequence such that $u_h \to u$ in X as $h \to \infty$, $\limsup_{h\to\infty} F(u_h) \leq F(u)$ and such that for every $m \in \mathbb{N}$ there exists a sequence $\{u_{m,h}\}_h$, $u_{m,h} \to u_m$ as $h \to \infty$ with $\limsup_{h\to\infty} F_h(u_{m,h}) \leq F(u_m)$, then there exists a recovering sequence of u in the sense of ii/ of the preceding definition.*

Definition 1.2.3 *We say that a sequence (F_h) is equicoercive (on X) if for every $t \in \mathbb{R}$ there exists a closed countably compact subset K_t of X such that $\{F_h \leq t\} \subseteq K_t$ for all $h \in \mathbb{N}$.*

Proposition 1.2.4 ([27],Proposition 7.7) *The sequence (F_h) is equicoercive if and only if there exists a lower semi-continuous coercive function $\psi : X \to \overline{\mathbb{R}}$ such that $F_h \geq \psi$ on X for all $h \in \mathbb{N}$.*

Theorem 1.2.5 ([27], Theorem 7.8) *Suppose that (F_h) is equicoercive on X, then the $\underbrace{\Gamma - \limsup_{h\to\infty} F_h}_{F'}$ and $\underbrace{\Gamma - \liminf_{h\to\infty} F_h}_{F''}$ are coercive and*

$$\min_{x\in X} F'(x) = \limsup_{h\to\infty} \inf_{x\in X} F_h(x)$$

If in addition (F_h) $\Gamma-$converges to a function F in X then F is coercive and

$$\min_{x\in X} F(x) = \lim_{h\to\infty} \inf_{x\in X} F_h(x).$$

1.3 Basic Preliminaries on Nonlinear Eigenvalues

If X is a normed vector space, X^* will denote the strong dual space, consisting of all the linear functionals on X that are continuous with respect to the topology induced on X by the norm and $\langle ., . \rangle$ will denote the duality pairing between X and X^*. If Y is another normed space, then $\mathcal{L}(X, Y)$ will stand for the space of all continuous linear mappings from X to Y .

A topological space M is said to be a C^k Banach manifold modelled on the Banach space X if there exist a set I, an open covering $\{U_i\}_{i \in I}$ of M, a family of closed vector subspaces X_i of X and a collection of mappings $\varphi_i : U_i \rightarrow X_i$ which are homeomorphisms with their images, such that $\varphi_i(U_i \cap U_j)$ is open in X and $\varphi_j \circ \varphi_i^{-1}$ induces a C^k diffeomorphism of $\varphi_i(U_i \cap U_j)$ onto $\varphi_j(U_i \cap U_j)$.
When it happens that all the X_i's are one-codimensional subspaces of X, M is said to be a one-codimensional Banach manifold. Since in this Thesis we aim to adress some issues regarding real eigenvalues, which are nothing but critical levels of functionals along one-codimensional manifolds, we restrict ourselves to this case.
Let $G : X \rightarrow \mathbb{R}$ be a C^1 functional, such that the topological subspace M of X defined by

$$M = \{u \in X \ : \ G(u) = 1\}, \tag{1.3}$$

consists of regular points for G, that is $X_u = ker G'(u) \neq X$, for all $u \in M$.
We call the tangent space to M at its point u the vector space

$$T_u M = \{\varphi \in X \ : \ \langle G'(u), \varphi \rangle = 0\},$$

consisting of all tangent vectors to M at u.
If $F : X \rightarrow \mathbb{R}$ is a C^1 functional, then its restriction to M is also C^1, its differential at a point $u \in M$ being nothing but the restriction $F'(u)_{|T_u M}$ of the differential $F'(u)$. Thus, a number c is a **critical value** of F **along** M if $F(u) = c$ and there exist a point $u \in M$ such that

$$\langle F'(u), \varphi \rangle = 0, \ for \ all \ \varphi \in T_u M, \tag{1.4}$$

and if this happens u is called a **critical point** of F along M corresponding to the critical value λ. By Lagrange multipliers' rule, a point $u \in M$ is a critical point of F along M is such that

$$F'(u) = \lambda G'(u), \tag{1.5}$$

in X^* for some real number λ.

Definition 1.3.1 *(Nonlinear eigenvalues). Let F, G be C^1 even and positively homogeneous functionals of degree $p \geq 1$ on the Banach space X, and $M = G^{-1}(\{1\})$. A real number λ is said to be an **eigenvalue** of the pair (F, G) if there exists $u \in X \setminus \{0\}$ such that*

$$\langle F'(u), v\rangle = \lambda\langle G'(u), v\rangle, \tag{1.6}$$

*holds for all $v \in X$. If this is the case, then u is called an **eigenvector** corresponding to λ. Note that eigenvectors and eigenvalues of the pair (F, G) are precisely given by the critical points and critical values F along M.*

Note that eigenvectors and eigenvalues of the pair (F, G) are precisely given by the critical points and critical values F along M. To see that, note that (1.4) holds for all eigenvectors $u \in M$ corresponding to the eigenvalue λ. Conversely, if $u \in M$ is a constrained critical point associated with the critical value c, then equation (1.6) holds with $\lambda = c$. Indeed, there has to be λ such that (1.5) holds, and by plugging $u = v$ in, one gets

$$\lambda \;=\; \lambda\langle G'(u), u\rangle = \langle F'(u), u\rangle = F(u) = c.$$

In the follow we discuss the existence of eigenvalues for a pair (F, G) of C^1 functionals which are even and positively homogeneous of degree $p > 1$. A mapping f is said to be homoegeneous of degree p if $f(\alpha v) = \alpha^p f(v)$.
Min-max formulae of the type

$$\lambda_n = \inf_f \max_\omega \frac{F(f_\omega)}{G(f_\omega)}$$

play a role. The maximum is taken among all unit vectors ω in $\mathbb{R}^N$, whereas f ranges over all odd and continuous mappings $\omega \mapsto f_\omega$ from the unit sphere $\mathbb{S}^{N-1}$ of $\mathbb{R}^N$ into $M = G^{-1}(\{1\})$. A mapping f from $\mathbb{S}^{N-1}$ to M is said to be odd if $f_{-\omega} = -f_\omega$, for all $\omega \in \mathbb{S}^{N-1}$. This is a well established method for producing eigenvalues of the pair (F, G). The procedure hardly would deserve a comment. Yet, for sake of completeness we discuss a proof of this existence result in next section, nonetheless.

CHAPTER 2

Stability Of Eigenvalues Associated To Rayleigh Quotient Generated By Norm of Gradient Under Variations Of The Exponent

Overview

This chapter is devoted to answer the stability question of the m^{th} variational eigenvalue $\lambda^m_{p(.)}$ of the Euler-Lagrange equation corresponding to the minimization of the Rayleigh ratio $\frac{\|\nabla u\|_{p(.)}}{\|u\|_{p(.)}}$ among all $u \in W^{1,p(x)}_0 \setminus \{0\}$ when varying $p(.)$, we accomplish this by investigating an important property concerning the geometry of the variable exponent Sobolev spaces.
This work makes subject of paper [16].

2.1 Related works and main results

The generalized Orlicz-Lebesgue spaces $L^{p(.)}$ and the corresponding generalized Orlicz-Sobolev spaces $W^{k,p(.)}$ appeared in the literature for the first time in a 1931 paper of Orlicz [24]. Nearly half a century later these spaces attracted more and more attention. One of the driving forces for the rapid development of this theory has been the model of electrorheological smart fluids introduced by Rajagopal and Rúžička in [31] and [30]. In this context, it is important to study the dependence of these problems governed by the $p(.)$-Laplacian with respect to the rheological behavior describing by the function $p(x)$ which depends also on the location x itself.

Let Ω be a bounded set of $\mathbb{R}^N$, $N \geq 2$, we denote by $\mathcal{P}(\Omega)$ the class of variable exponents $p : \overline{\Omega} \to \mathbb{R}^+$ continuous and satisfying

$$1 \leq p^- := \inf_{x \in \overline{\Omega}} p(x) \leq p^+ := \sup_{x \in \overline{\Omega}} p(x) < \infty.$$

In general smooth functions are not dense in $W^{1,p(.)}(\Omega)$. For that we need to add the so-called vanishing log-Hölder continuity assumption on the exponent $p(.)$.

We say that $p \in \mathcal{P}^{\log}(\Omega)$ if $p \in \mathcal{P}(\Omega)$ and is $\log-$Hölder continuous, namely,
$|p(x) - p(y)| \leq -\frac{L}{\log(|x-y|)}$ for some $L > 0$ and for all $x, y \in \Omega$ with $0 < |x - y| < 1/2$.

The study of the eigenvalue problem associated to the minimization of the Rayleigh ratio $R(u) := \frac{K(u)}{k(u)}$, where $K(u) := \|\nabla u\|_{p(.)}$ and $k(u) := \|u\|_{p(.)}$, has attracted quite a lot of authors. For instance, in [13] the authors derived the associated Euler-Lagrange equation and stated it to

be

$$\left(\mathcal{V}_{p(.)}\right) \begin{cases} -\text{div}\left(p(x)\left|\frac{\nabla u}{K(u)}\right|^{p(x)-2}\frac{\nabla u}{K(u)}\right) = \lambda S(u)p(x)\left|\frac{u}{k(u)}\right|^{p(x)-2}\frac{u}{k(u)} \quad , \\ u \in W_0^{1,p(.)}(\Omega), \end{cases}$$

where $S(u) := \dfrac{\int_\Omega |\frac{\nabla u}{K(u)}|^{p(x)}dx}{\int_\Omega |\frac{u}{k(u)}|^{p(x)}dx}$. Here $\|.\|_{p(.)}$ is the Luxemburg norm defined in Chapter 1. Inspired from the work of [26], the authors in [29] showed the existence of an infinite non-decreasing sequence of non negative variational eigenvalues $\lambda^{(m)}_{p(x)}$ associated to $\left(\mathcal{V}_{p(.)}\right)$ defined by

$$\lambda^{(m)}_{p(x)} := \inf_{K\in\mathcal{W}^{(m)}_{p(x)}}\sup_{u\in K}\|\nabla u\|_{p(x)} \ \forall m \geq 1$$

where $\mathcal{W}^{(m)}_{p(x)}$ is the set of symmetric compact subsets of

$$\{u \in W_0^{1,p(x)}(\Omega) \ : \ \|u\|_{p(x)} = 1\}$$

such that $\gamma(K) \geq m$, $\gamma(K)$ stands for the Krasnosel'skii genus of K and $\|.\|_{p(.)}$ stands for the Luxemburg norm that will be defined in the next section

For $m = 1$, in 2014 the authors of [5] showed that the first eigenvalue is characterized to be the infimum value of Rayleigh quotient under the Luxemburg norm and that is continuous with respect to the exponent $p(.)$.

One year later, For $m \geq 1$ by using the results of [8], F. Colasuonno and M. Squassina [7] achieved to prove that the map $\Lambda_m(p(.)) := \lambda^{(m)}_{p(x)}$, $m \geq 1$ is right continuous. Let us mention that when p is not constant, contrary to constant exponent situation, the first eigenvalue can be zero. e can consider two different problems concerning eigenfunctions or modular eigenfunctions (with zero eigenvalues). In the present work, we are interested to the eigenfunctions, however the papers [22, 24, 25, 28] are devoted to modular eigenfunctions.

The purpose of this chapter is to study the stability phenomenon of the mth variational eigenvalue $\lambda^{(m)}_{p(.)}$ associated to problem $\left(\mathcal{V}_{p(.)}\right)$ under variations of the exponent $p(.)$.

More precisely, when the sequence $p_h(.)$ is increasing to $p(.)$, does $\lambda^{(m)}_{p_h(.)}$ converges to $\lambda^{(m)}_{p(.)}$?

As far as it's known there are no existing works studying the stability of the variational spectrum of the above problem when we have a convergence from the left of the variable exponents .

In the constant exponent case, the study of stability phenomenon has started at the end of the 1980s for eigenvalue problems associated to the $p-$Laplacian operator when Lindqvist [25]

showed that $\lim_{s\to p^+}\lambda_s=\lambda_p$ when $\lim_{s\to p^-}\lambda_s=\lambda_p$ is obtained provided that $\lim_{s\to p^-}\int_\Omega|\nabla u_s-\nabla u_p|^s\,dx=0$, whereas the inverse

$$\left(\lim_{s\to p^-}\lambda_s=\lambda_p\right)\Rightarrow\left(\lim_{s\to p^-}\int_\Omega|\nabla u_s-\nabla u_p|^s\,dx=0\right)$$

was declared as an open problem for general domains. The stability problem for more general operators was treated by [10].

The results shown in [6] were of big importance since they initiate two important works: (i) [28] where the author used a result of [6] to show the stability with respect to p of higher eigenvalues associated to the $p-$Laplacian by using a $\Gamma-$convergence argument, and (ii) [9] where the authors answered a question raised in [6] concerning the relation between the $\Gamma-converegence$ of some functionals f_h and that of the related functionals $\mathcal{F}_h^{(m)}:=\sup_K f_h$ (under some conditions on f_h and f, the limit of f_h). In fact, the "inverse" declared open by by Lindqvist [26] has been closed in [9]. In particular they showed the following result:

Corollary 2.1.1 ([8]) *Assume that*

$$f(u)=\left(\Gamma-\lim_{h\to\infty}\right)(u)\ for\ all\ u\in X$$

and that, for every strictly increasing sequence (h_n) *in* $\mathbb{N}$ *and every sequence* (u_n) *in* $X\setminus\{0\}$ *such that*

$$\sup_{n\in\mathbb{N}}f_{h_n}(u_n)<+\infty,$$

there exists a subsequence (u_{n_j}) *converging to some* u *in* X *with*

$$\lim_{j\to\infty}g_{h_{n_j}}(u_{n_j})=g(u).$$

Then, for every $m\geq 1$*, the sequence* $(\mathcal{F}_h^{(m)})$ *is asymptotically equicoercive*[1] *and*

$$\mathcal{F}^{(m)}(K)=\left(\Gamma-\lim_{h\to\infty}\mathcal{F}_h^{(m)}\right)(K)\ for\ all\ K\in\mathcal{K},$$

[1]We say that the sequence (F_h) is asymptotically equicoercive (on X), if for every strictly increasing sequence (h_n) in $\mathbb{N}$ and every sequence (u_n) in X satisfying $\sup_{n\in\mathbb{N}}F_{h_n}<+\infty$ there exists a subsequence (u_{n_j}) converging in X.

$$\inf_{k\in\mathcal{K}}\mathcal{F}^{(m)}(K)=\lim_{h\to\infty}\left(\inf_{k\in\mathcal{K}}\mathcal{F}_h^{(m)}(K)\right),$$

$$\inf_{k\in\mathcal{K}_s^{(m)}}\sup_K f=\lim_{h\to\infty}\left(\inf_{k\in\mathcal{K}_{s,h}^{(m)}}\sup_K f_h\right).$$

This latter kind of variational convergence ($\Gamma-convergence$) has became, after its introduction by De Giorgi, the commonly recognized notion of convergence for variational problems since it's designed to express the convergence of minimum problems.
Another work of [8] consists on giving an affirmative answer to the open problem posed in [25]. In fact, they gave several characterizations of the fact that $\lim_{s\to p^-}\lambda_s^{(1)}=\lambda_p^{(1)}$ and they proved the stability of the m^{th}eigenvalue associated to the $p-Laplacian$.

Our aim in this chapter is to prove the continuity of the map

$$\Lambda_m:\ (C(\Omega),\|.\|_\infty)\to\mathbb{R}$$

$$p(.)\mapsto\Lambda_m(p(.)):=\lambda_{p(.)}^{(m)}$$

with respect to variations of $p(.)$ that is, if $(p_h),p\ \in\mathcal{P}^{\log}(\Omega)$ and the sequence $p_h(.)$ is increasing to $p(.)$ then $\lim_{h\to\infty}\Lambda_m(p_h(.))=\Lambda_m(p(.))$.
In the next section we will prove an important functional property concerning Sobolev spaces with variable exponent refining, in terms of weakness of hypothesis, the one obtained in [21], this property will play a crucial role to obtain the stability. In fact we will show that if Ω satisfies the segment property then

$$W_0^{1,1}(\Omega)\cap W^{1,p(.)}(\Omega)=W_0^{1,p(.)}(\Omega).$$

Recall that an open subset Ω of $\mathbb{R}^N$ is said to have the segment property if, given any $x\in\partial\Omega$, there exists an open set G_x in $\mathbb{R}^N$ with $x\in G_x$ and y_x of $\mathbb{R}^N\setminus\{0\}$ such that, if $z\in\overline{\Omega}\cap G_x$ and $t\in]0,1[$, then $z+ty_x\in\Omega$.

This property allows us by a translation to push the support of a function u in Ω.

The third part will be dedicated to prove the main result of this chapter; we will proceed as follows: let $\underline{\mathcal{W}}_{p(x)}^{(m)}$ be the set of symmetric compact subsets of

$\{u \in W_0^{1,1}(\Omega) \cap W^{1,p(x)}(\Omega) \ : \ \|u\|_{p(x)} = 1\}$ such that $\gamma(K) \geq m$ and define

$$\underline{\Lambda}_m(p(.)) := \underline{\lambda}_{p(x)}^{(m)} := \inf_{K \in \underline{\mathcal{W}}_{p(x)}^{(m)}} \sup_{u \in K} \|\nabla u\|_{p(x)} \ \forall m \geq 1$$

we first show that

$$\lim_{h \to \infty} \underline{\Lambda}_m(p_h(.)) = \lim_{h \to \infty} \Lambda_m(p_h(.)) = \underline{\Lambda}_m(p(.))$$

then, using an important result concerning the $\Gamma-$convergence and the fact that

$$W_0^{1,1}(\Omega) \cap W^{1,p(.)}(\Omega) = W_0^{1,p(.)}(\Omega)$$

we will be able to prove that $\Lambda_m(p(.)) = \underline{\Lambda}_m(p(.))$ and then the continuity.

Remark 2.1.2 *Let E be a real Banach space, and denote by $\Sigma(E)$ the set of all compact subsets of E which do not contain 0 and are symmetric with respect to 0.*
The concept of genus was first introduced in 1952 as an alternative of the Ljusternik-Schnirelmann category in the minmax approach of critical points.
This genus measures the 'size' of some subsets of E which are symmetric with respect to the origin.
We define the genus $\gamma(A)$ of a non-empty set $A \in \Sigma(E)$ to be the smallest integer $n \in \mathbb{N}^*$ for which there exists a nontrivial and odd mapping $g \in C(A, \mathbb{R}^N \setminus \{0\})$. We write $\gamma(A) = +\infty$ if no such mapping exits, and $\gamma(\emptyset) = 0$.

2.2 The space $W^{1,p(.)}(\Omega) \cap W_0^{1,1}(\Omega)$

Often elements of variable exponent Sobolev space $W_0^{1,p(.)}(\Omega$ as well as when p is constant, are expressed to describe "$u_{|_{\partial\Omega}} = 0$". This fact is reasonably true only if $\partial\Omega$ is sufficiently smooth. To avoid this approach, we need to extend u to the whole space $\mathbb{R}^N$. Strictly speaking, for u belonging to $W_0^{1,1}(\Omega)$ will ensures that its extension, by zero outside of Ω, denoted $\widetilde{u}$ is in $W_0^{1,1}(\mathbb{R}^N)$ and belonging also to $W_0^{1,p(.)}(\Omega)$ will guarantee that $\widetilde{u}$ is in $W_0^{1,p(.)}(\mathbb{R}^N)$.

The purpose of this section is to prove the following theorem

Theorem 2.2.1 *Let Ω be a bounded domain satisfying the segment property and $p \in \mathcal{P}^{\log}(\Omega)$ satisfying $1 < p^- \leq p^+ < \infty$, we have*

$$W_0^{1,1}(\Omega) \cap W^{1,p(.)}(\Omega) = W_0^{1,p(.)}(\Omega).$$

The proof is based on fine properties of Sobolev functions. By definition, Sobolev functions are defined only up to Lebesgue measure zero and thus it is not always clear how to use their point-wise properties. We define a representative, called $quasicontinuous$, and exploit it to achieve the zero boundary value of u. We recall in the sequel some notions concerning $p(.) - Sobolev\ capacities$ and then some basic results of [21] (without proof) that we will need to prove Theorem 2.2.1.

Throughout next, we will use the notation and terminology of [21].

Definition 2.2.2 *Let $p \in \mathcal{P}(\mathbb{R}^N)$ satisfy $p(x) \in [1, \infty)$ for almost every x. For $E \subset \mathbb{R}^N$ we denote*

$$S_{p(.)}(E) := \left\{ u \in W^{1,p(.)}(\mathbb{R}^N)\ u \geq 1 \text{ in an open set containing } E \text{ and } u \geq 0 \right\}.$$

The Sobolev $p(.)$-capacity of E is defined by

$$C_{p(.)}(E) := \inf_{u \in S_{p(.)}(E)} \int_{\mathbb{R}^N} |u|^{p(x)} + |\nabla u|^{p(x)}\, dx$$

In case $S_{p(.)}(E) = \emptyset$ we set $C_{p(.)}(E) = \infty$.

Definition 2.2.3 *A claim is said to hold $p(.) - quasieverywhere$ if it holds everywhere except in a set of Sobolev $p(.) - capacity$ zero.*

A function $u : \Omega \to \mathbb{R}$ is said to be $p(.) - quasicontinuous$ if for every $\epsilon > 0$ there exists an open set U with $C_{p(.)}(U) < \epsilon$ such that u restricted to $\Omega \setminus U$ is continuous.

Theorem 2.2.4 *Let $p \in \mathcal{P}(\mathbb{R}^N)$ be such that $1 < p^- \leq p^+ < \infty$ and let $u \in W^{1,p(.)}(\mathbb{R}^N)$. Then there exists a set $E \subset \mathbb{R}^N$ of zero Sobolev $p(.) - capacity$ such that*

$$u^*(x) := \lim_{r \to 0} \fint_{B(x,r)} u(y)\, dy$$

exists for every $x \in \mathbb{R}^N \setminus E$.

The function u^* is the $p(.) - quasicontinuous$ representative of u.

Lemma 2.2.5 *Suppose that $p \in \mathcal{P}^{\log}(\mathbb{R}^N)$ satisfies $1 < p^- \leq p^+ \leq \infty$ and let $u \in W^{1,p(.)}(\mathbb{R}^N)$. Then*

$$C_{p(.)}\left(\left\{x \in \mathbb{R}^N : \limsup_{r\to 0} r^{p(x)} \fint_{B(x,r)} |\nabla u(y)|^{p(y)}\, dy > 0\right\}\right) = 0.$$

Theorem 2.2.6 *Let $p \in \mathcal{P}(\mathbb{R}^N)$ satisfy $1 < p^- \leq p^+ \leq \infty$. If the function $u \in W^{1,p(.)}(\mathbb{R}^N)$ is $p(.) - quasicontinuous$ and zero $p(.) - quasieverywhere$ in $\mathbb{R}^N \setminus \Omega$, then $u \in W_0^{1,p(.)}(\Omega)$.*

Theorem 2.2.7 *Let $p \in \mathcal{P}(\mathbb{R}^N)$ with $1 < p^- \leq p^+ < \infty$ and let u and v be $p(.)-quasicontinuous$ functions in $\mathbb{R}^N$. Suppose that $U \subset \mathbb{R}^N$ is open*

(a) if $u = v$ almost everywhere in U, then $u = v$ $p(.) - quasieverywhere$ in U.

(b) if $u \leq v$ almost everywhere in U, then $u \leq v$ $p(.) - quasieverywhere$ in U.

Remark 2.2.8 *The above results were extended recently to the case $p^- = 1$ by [14] except Theorem 2.2.6.*

Proof of Theorem 2.2.1. First, $W^{1,p(.)}(\Omega) \cap W_0^{1,1}(\Omega)$ is a closed linear subspace containing $W_0^{1,p(.)}(\Omega)$. The first inclusion of the theorem follows.

To prove the second inclusion, we start by extending p outside of Ω to the whole space $\mathbb{R}^N$ according to the result of [21, Prop.4.1.7] which permits to preserve the same infimum and log-Hölder continuity of $p(.)$. We suppose that this extension still denoted by p.

Let $u \in W^{1,p(.)}(\Omega) \cap W_0^{1,1}(\Omega)$ and $\widetilde{u}$ be its extension by zero outside of Ω to the whole space $\mathbb{R}^N$. Thus $\widetilde{u} \in W_0^{1,1}(\mathbb{R}^N)$ and

$$\int_{\mathbb{R}^N} \widetilde{u}\frac{\partial \phi}{\partial x_i} dx = -\int_{\Omega} \phi \frac{\partial u}{\partial x_i} dx,$$

for any test function $\phi \in C_c^{\infty}(\mathbb{R}^N)$. We deduce that $\widetilde{u} \in W^{1,p(.)}(\mathbb{R}^N)$. Now are ready to apply 2.2.6. For that it sufficient to prove that $\widetilde{u}$ satisfies the two assumptions of 2.2.7. Indeed, Let $\widetilde{u}$ denote the $p(.) - quasicontinuous$ representative of u (Theorem 2.2.4.)

$$\widetilde{u}(x) = \lim_{r\to 0^+} \fint_{B(x,r)} u(y)\, dy$$

Since u is zero almost everywhere in $\mathbb{R}^N \setminus \overline{\Omega}$, we find by Theorem 2.2.7 that $\widetilde{u}$ is zero $p(.)-quasieverywhere$ in $\mathbb{R}^N \setminus \overline{\Omega}$. Then if we show that $\lim\limits_{r\to 0} \fint_{B(x,r)} u\, dy = 0$ for $quasievery\ x \in \partial\Omega$ we can immediately conclude our result using Theorem 2.2.6.

Let $x \in \partial\Omega$, since $u \in W_0^{1,1}(\Omega)$ and Ω satisfies the segment property $u \in W_0^{1,1}(B(x,r))$ so that we have by Poincaré's inequality for a ball (see [11, Section 5.8.1])

$$\|u - \fint_{B(x,r)} u\, dy\|_{L^1(B(x,r))} \leq cr\|\nabla u\|_{L^1(B(x,r))}$$

That is,

$$\int_{B(x,r)} |u - \fint_{B(x,r)} u\, dy|\, dz \leq cr \int_{B(x,r)} |\nabla u(y)|\, dy.$$

That is,

$$\int_{B(x,r)} |u(y)|\, dy \leq cr \int_{B(x,r)} |\nabla u(y)|\, dy\ + \int_{B(x,r)} |u(y)|\, dy.$$

Then there is a constant $C > 0$ such that

$$\int_{B(x,r)} |u(y)|\, dy \leq R\, \|u\|_{1,1,B(x,r)} \leq\ CR|u|_{1,1,B(x,r)},$$

where $R = \max(cr, 1)$. Hence

$$\int_{B(x,r)} |u(y)|\, dy \leq CR\ \int_{B(x,r)} |\nabla u|\, dy.$$

By Hölder's inequality log-Hölder continuity and Corollary 1.1.8, we obtain

$$\fint_{B(x,r)} |u(y)|\, dy\ \leq\ CRr^{-n}\|1\|_{L^{p'(.)}(B(x,r))}\ \|\nabla u\|_{L^{p(.)}(B(x,r))}$$

$$\fint_{B(x,r)} |u(y)|\, dy\ \leq\ CRr^{-n+\frac{n}{p'(x)}}\|\nabla u\|_{L^{p(.)}(B(x,r))}$$

$$= \frac{CR}{r}\|r^{\frac{p(x)-n}{p(x)}}|\nabla u|\|_{L^{p(.)}(B(x,r))}$$

Since p is bounded, it is enough to show that

$$\frac{CR}{r}\int_{B(x,r)} r^{\frac{p(x)-n}{p(x)}\cdot p(y)}|\nabla u|^{p(y)}\, dy\ \leq\ CRr^{p(x)-1}\fint_{B(x,r)} |\nabla u|^{p(y)}\, dy \underset{r\to 0}{\longrightarrow} 0$$

for $quasievery\ x \in \partial\Omega$. By Lemma 2.2.5, we infer that $r^{p(x)}\fint_{B(x,r)} |\nabla u|^{p(y)}\, dy \underset{r\to 0}{\longrightarrow} 0$ and then $\lim\limits_{r\to 0}\fint_{B(x,r)} u\, dy = 0$ for almost every $x \in \partial\Omega$. The proof of the theorem is completely archived.

2.3 The Stability Result

Let $p(.) \in \mathcal{P}^{\log}(\Omega)$. Under the segment property $\underline{\lambda}^{(m)}_{p(.)} = \lambda^{(m)}_{p(.)}$ for any $m \geq 1$.
We define $g_{p(.)},\ \mathcal{E}_{p(.)}(u)\ and\ \underline{\mathcal{E}}_{p(.)}(u) :\ L^1(\Omega) \to [0, +\infty[$ as

$$g_{p(.)}(u) := \begin{cases} \|u\|_{p(x)}, & \text{if } u \in L^{p(x)}(\Omega); \\ 0, & \text{otherwise.} \end{cases}$$

$$\mathcal{E}_{p(.)}(u) := \begin{cases} \|\nabla u\|_{p(x)}, & \text{if } u \in W_0^{1,p(.)}(\Omega); \\ +\infty, & \text{otherwise.} \end{cases}$$

and

$$\underline{\mathcal{E}}_{p(.)}(u) := \begin{cases} \|\nabla u\|_{p(x)}, & \text{if } u \in W_0^{1,1}(\Omega) \cap W^{1,p(.)}(\Omega); \\ +\infty, & \text{otherwise.} \end{cases}$$

Before dressing our first result we give an elementary definition that we will work with in the next.

Definition 2.3.1 *Given a sequence of exponents $(p_h(.)) \subset \mathcal{P}^{\log}(\Omega)$, we say that $(p_h(.))$ is increasing (resp. decreasing) to $p(.) \in \mathcal{P}^{\log}(\Omega)$ if and only if*

$$p_h \text{ converges uniformly to } p \text{ in } \Omega$$

$$\text{and } \ p_h(x) \leq p(x) \text{ (resp. } p_h(x) \geq p(x)) \text{ for all } h \in \mathbb{N} \text{ and } x \in \Omega.$$

Recall that p_h converges uniformly to p if for all $\epsilon > 0\ \exists N_\epsilon \in \mathbb{N}$ such that
$|p_h(x) - p(x)| < \epsilon\ \forall x \in \Omega;\ h \geq N_\epsilon$.

Theorem 2.3.2 *For any domain Ω satisfying the segment property and for every sequence $(p_h(.)) \subset \mathcal{P}^{\log}(\Omega)$ increasing to $p(.) \in \mathcal{P}^{\log}(\Omega)$, it holds*

$$\Gamma - \lim_{h\to\infty} \underline{\mathcal{E}}_{p_h(.)} = \Gamma - \lim_{h\to\infty} \mathcal{E}_{p_h(.)} = \underline{\mathcal{E}}_{p(.)}$$

Proof 2.3.3 *If $\underline{\mathcal{E}}_{p_h(.)}(u) = +\infty$ then nothing to prove, else $\underline{\mathcal{E}}_{p_h(.)}(u) = \|u\|_{p_h(x)}$. Via the segment property $u \in W^{1,p(.)}(\Omega)$ and thanks to Theorem 3.5.7 of [21], we get*

$$\underline{\mathcal{E}}_{p_h(.)}(u) \to \underline{\mathcal{E}}_{p(.)}(u)$$

uniformly as $h \to \infty$.
Since $\underline{\mathcal{E}}_{p(.)}$ and $\underline{\mathcal{E}}_{p_h(.)}$ are lower semi-continuous, we deduce then from Remark 5.3 of [27] that $\Gamma - \lim_{h\to\infty} \underline{\mathcal{E}}_{p_h(.)} = \underline{\mathcal{E}}_{p(.)}$.
By the same argument we obtain $\Gamma - \lim_{h\to\infty} \mathcal{E}_{p_h(.)} = \underline{\mathcal{E}}_{p(.)}$

For the sake of completeness we recall some results from [7] that we will use later.

Theorem 2.3.4 ([7]) *Let $p, p_h \in \mathcal{P}^{log}(\Omega)$ such that $p_h \to p$ uniformly in Ω. Then, for every subsequence (p_{h_n}) and for every sequence $(u_n) \subset L^1(\Omega)$ verifying*

$$\sup_{n\in\mathbb{N}} \mathcal{E}_{p_{h_n}}(u_n) < \infty,$$

there exists a subsequence u_{n_j} such that, as $j \to \infty$,

$$u_{n_j} \to u \; in \; L^1(\Omega)$$

and

$$g_{p_{h_{n_j}}(x)}(u_{n_j}) \to g_{p(x)}(u).$$

Theorem 2.3.5 ([7]) *For every sequence $(p_h(.)) \subset \mathcal{P}^{\log}(\Omega)$ decreasing to $p(.) \in \mathcal{P}^{\log}(\Omega)$, it holds*

(a) $\Gamma - \lim_{h\to\infty} \underline{\mathcal{E}}_{p_h(.)} = \Gamma - \lim_{h\to\infty} \mathcal{E}_{p_h(.)} = \mathcal{E}_{p(.)}$

(b) $\Lambda_m(p_h(.)) \to \Lambda_m(p(.)) \; as \; h \to \infty$

Theorem 2.3.6 ([7]) *The following properties hold*

(a) $g_{p(x)}$ *is even positively homogenous of degree 1;*

(b) For every $b \in \mathbb{R}$ the restriction of $g_{p(x)}$ to $\{u \in L^1(\Omega) \ : \ \mathcal{E}_{p(x)} \leq b\}$ is continuous.

We establish next an important result concerning the left continuity of $\Lambda_m(p(.))$.

Theorem 2.3.7 *For every sequence $(p_h(.)) \subset \mathcal{P}^{\log}(\Omega)$ increasing to $p(.) \in \mathcal{P}^{\log}(\Omega)$, it holds $\lim_{h\to\infty} \Lambda_m(p_h(.)) = \Lambda_m(p(.))$*

Proof 2.3.8 *We need to show that the minmax values with respect to the $W^{1,p(x)}-$topology are equal to those with respect to the weaker topology $L^1(\Omega)$.*
Let us define $\mathcal{K}^{(m)}_{s,p(x)}$ to be the set of symmetric compact subsets K of $\{u \in L^1(\Omega) \ : \ \|u\|_{p(x)} = 1\}$ such that $\gamma(K) \geq m$, by Theorem $2.3.6 - (b)$ the hypothesis of [[9], Corollary 3.3] are fulfilled so that we have

$$\inf_{k\in\mathcal{K}^{(m)}_{s,p(x)}} \sup_{u\in K} \underline{\mathcal{E}}_{p(x)}(u) = \inf_{k\in\underline{\mathcal{W}}^{(m)}_{p(x)}} \sup_{u\in K} \underline{\mathcal{E}}_{p(x)}(u)$$

Thanks to Theorem 3.1 and Theorem 2.3.4, we deduce from Corollary 4.4 of [8] that

$$\inf_{K\in\mathcal{K}^{(m)}_{s,p(x)}} \sup_{u\in K} \mathcal{E}_{p(x)}(u) = \lim_{h\to\infty} \Big(\inf_{K\in\mathcal{K}^{(m)}_{s,p_h(x)}} \sup_{u\in K} \mathcal{E}_{p_h(x)}(u)\Big)$$

The proof ends by recalling that

$$\inf_{K\in\mathcal{K}^{(m)}_{s,p(x)}} \sup_{u\in K} \mathcal{E}_{p(x)}(u) = \inf_{K\in\mathcal{K}^{(m)}_{s,p(x)}} \sup_{u\in K} \underline{\mathcal{E}}_{p(x)}(u)$$

Corollary 2.3.9 *Let $p, (p_h(.)) \in \mathcal{P}^{\log}(\Omega)$ such that $p_h(.)$ increases to $p(.)$ and assume that Ω satisfies the segment property then Λ_m is continuous with respect to $p(.)$, that is,* $\lim_{h\to\infty} \Lambda_m(p_h(.)) = \Lambda_m(p(.))$

Proof 2.3.10 *The assertion follows from combining Theorem 2.3.7 and Theorem 2.3.5-(b).*

Remark 2.3.11 *Note that the stability of the eigenfunctions of the eigenvalue problem $\mathcal{V}_{p(.)}$ under perturbation of the exponent $p(.)$ was recently accomplished in [22].*

CHAPTER 3

Stability of Eigenvalues of a Second Order Partial Differential Equation

Overview

In this chapter we apply an elementary result of $\Gamma - \text{convergence}$ to show the dependence on $p(\cdot)$ of the first eigenvalue $\lambda_{p(x)}$ of the $p(x)-$Laplacian problem

$$\begin{cases} -\Delta_{p(x)}(u) = \lambda |u|^{p(x)-2}u & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega. \end{cases}$$

This result makes subject of paper [17].

3.1 Related Works and Main Result

The eigenvalue problem associated with the $p(x)-$Laplacian is defined by

$$\begin{cases} -\text{div}(|\nabla u|^{p(x)-2}\nabla u) = \lambda |u|^{p(x)-2}u & ; \\ u \in W_0^{1,p(x)}(\Omega) & . \end{cases} \tag{3.1}$$

where Ω is a bounded set of $\mathbb{R}^N$ with Lipschitzien boundary, $N \geq 2$ and

$p(\cdot) \in \mathcal{P}^{log}(\Omega)$, $\mathcal{P}^{log}(\Omega)$ defined as in Section 2.1.

A pair $(u, \lambda) \in W_0^{1,p(x)}(\Omega) \times \mathbb{R}$ is a weak solution of (3.1) provided that

$$\int_\Omega |\nabla u|^{p(x)-2}\nabla u \nabla v \, dx = \lambda \int_\Omega |u|^{p(x)-2}uv \, dx \; \forall v \in W_0^{1,p(x)}(\Omega) \tag{3.2}$$

u is called an eigenfunction and λ is called an eigenvalue.

(3.2) is the Euler-Lagrange equation associated to the minimization of the Rayleigh ratio

$$R(u) := \frac{\int_\Omega |\nabla u|^{p(x)} \, dx}{\int_\Omega |u|^{p(x)} \, dx}$$

This follows essentially from the next theorems:

Theorem 3.1.1 *The Fréchet derivative $F' \in L^{p'}(\Omega)$ $(1 < p^- \leq p^+ < \infty)$ of the modular*

$$F : L^{p(.)}(\Omega) \to [0, \infty),$$

defined by

$$F(u) = \int_\Omega |u(x)|^{p(x)} dx,$$

is given by

$$\langle v, F'_u \rangle \;=\; \int_\Omega p(x)|u(x)|^{p(x)-2}u(x)v(x)\,dx \;(v \in L^{p(.)}(\Omega)). \tag{3.3}$$

Results following form this theorem are:

Corollary 3.1.2 *Under the assumptions of Theorem 3.1.1, the Fréchet derivative of the modular:*

$$G:\; W_0^{1,p(.)}(\Omega) \to [0,\infty),$$

defined by

$$G(u) = \int_\Omega |\nabla u(x)|^{p(x)} dx,$$

is given by

$$\langle v, G'(u) \rangle \;=\; \int_\Omega p(x)|\nabla u(x)|^{p(x)-2}\nabla u(x).\nabla v(x)\,dx \;(v \in W_0^{1,p(.)}(\Omega)). \tag{3.4}$$

Theorem 3.1.3 *Under the assumptions of Theorem 3.1.1, the Fréchet derivative of the functional*

$$H:\; L^{p(.)}(\Omega) \to [0,\infty),$$

$$H(u) = \int_\Omega \frac{|u(x)|^{p(x)}}{p(x)} dx,$$

is given by

$$\langle v, H'(u) \rangle \;=\; \int_\Omega |\nabla u(x)|^{p(x)-2}u(x)v(x)dx \;(v \in L^{p(.)}(\Omega)). \tag{3.5}$$

Likewise, the Fréchet derivative of

$$I:\; W_0^{1,p(.)}(\Omega) \to [0,\infty),$$

defined by

$$I(u) = \int_\Omega \frac{|\nabla u(x)|^{p(x)}}{p(x)} dx,$$

is given by

$$\langle v, I'(u)\rangle \;=\; \int_{\Omega} |\nabla u(x)|^{p(x)-2}\nabla u(x)\nabla v(x)dx \;(v \in W_0^{1,p(.)}(\Omega)). \tag{3.6}$$

Coming back to problem (3.1), in [33], the authors showed that problem (3.1) has an infinite sequence of eigenvalues defined by

$$\lambda^{m,p(\cdot)}(u) = \inf_{K\in\mathcal{W}_{m,p(\cdot)}(\Omega)} \sup_{u\in K} \int_{\Omega} |\nabla u|^{p(x)}\, dx. \tag{3.7}$$

where $\mathcal{W}_{m,p(.)}(\Omega)$ is the set of symmetric and compact subsets of
$\{u \in W_0^{1,p(x)}(\Omega) :\ \int_\Omega |u|^{p(x)}\, dx = 1\}$ such that $i(K) \geq m$, and i denotes the Krasnoselskii's genus.[1] This result is obtained based on Lusternik-Schnirelmann theory, for more detail on this issue see [32].

When $m = 1$ we use the notation $\lambda^{p(\cdot)} = \lambda^{1,p(\cdot)}$.

The normalization condition $\int_\Omega |u|^{p(x)}\, dx = 1$ taken on u is necessary to avoid the fact that $\lambda^{m,p(\cdot)}(u) \equiv 0$. Indeed, the example introduced in [33] shows that this may happen even if $u \not\equiv 0$

The first eigenvalue (i.e. $m = 1$) is characterized to be

$$\lambda^{p(\cdot)} = \inf_{u\in\mathcal{M}} \int_{\Omega} |\nabla u|^{p(x)}\, dx\ ;\ \mathcal{M} = \{u \in W_0^{1,p(x)}(\Omega)/\ \int_{\Omega} |u|^{p(x)}\, dx = 1\} \text{ is a closed manifold.}$$

In the constant exponent case E. Parini showed in [28] the continuity with respect to p of the m^{th} variational eigenvalue in a sufficiently regular domain. More precisely the authors require the domain to be bounded with Lipschitz boundary.

A few years later, M. Degiovanni and M. Marzocchi showed in [9] the continuity of the m^{th} variational eigenvalue but requiring the domain to be only bounded open and connected subset of $\mathbb{R}^N$.

For other articles treating the stability see for example [10], [7], [25] and the references therein.

As far as we know the stability of eigenvalues associated to the $p(x)-$Laplacian problem have

[1] Let E be a real Banach space and $\Sigma(E)$ the set of all closed subsets of E which do not contain 0 and are symmetric with respect to 0. The Krasnoselkii's genus $i(A)$ of a non empty set $A \in \Sigma(A)$ is defined to be the smallest integer $m \in \mathbb{N}$ for which there exists an odd mapping $g \in C(A, \mathbb{R}^m \setminus \{0\})$. If no such mapping exists we write $i(A) = +\infty$.

not been treated before. Our main goal in this work is to show that the first eigenvalue associated to the $p(x)-$Laplacian problem $\lambda^{p(\cdot)}$ is continuous with respect to $p(\cdot)$, i.e. for all $h \in \mathbb{N}$

$$\lambda^{p_h(\cdot)} \longrightarrow \lambda^{p(\cdot)} \; as \; p_h(\cdot) \underset{h\to\infty}{\longrightarrow} p(\cdot) \; uniformly.$$

To achieve this goal we utilized an elementary result found in [27] concerning convergence of minimum values of a family of functionals F_h to the minimum value of a limit F.
Since it has been introduced in 1985 by De Giorgi, De Giorgi's $\Gamma-$convergence covers almost all other variational convergences in term of applications as it plays a central role for its compactness properties and for the large number of results concerning $\Gamma-$limits of integral functionals.
The most difficult issue in the study of the Rayleigh ratio with integrals is the non homogeneity of the integrals in marked contrast to the ratio with norms.

3.2 Proof of The Stability Result

First of all, let us present the functional notations adapted in this chapter.
$F: \; M \to [0, +\infty]$ defined by:

$$F(u) = \int_\Omega |\nabla u|^{p(x)} \, dx \tag{3.8}$$

and $F_h: \; M_h \to [0, +\infty]$ defined by:

$$F_h(u) = \int_\Omega |\nabla u|^{p_h(x)} \, dx \tag{3.9}$$

To show the continuity of the first eigenvalue we have to prove the following convergence

$$\inf_{u\in M_h} \int_\Omega |\nabla u|^{p_h(x)} \, dx \longrightarrow \inf_{u\in M} \int_\Omega |\nabla u|^{p(x)} \, dx \tag{3.10}$$

as $p_h(\cdot) \to p(\cdot)$ uniformly, where $M_h = \{u \in W_0^{1,p_h(\cdot)}(\Omega) / \; \int_\Omega |u|^{p_h(x)} \, dx = 1\}$ and $M = \{u \in W_0^{1,p(\cdot)}(\Omega) / \; \int_\Omega |u|^{p(x)} \, dx = 1\}$.
In view of Theorem 1.2.5, the convergence (4.5) will be attained if we show that the family F_h $\Gamma-$converges to F as $h \to \infty$ and F_h is equicoercive.
Our main concern in this section is to prove the following theorem

Theorem 3.2.1 *1. The family of functions (F_h) is equicoercive.*

2. Let $\Omega \subseteq \mathbb{R}^N$ be an open and bounded set with Lipschitzien boundary and let $(p_h) \subset \mathcal{P}^{\log}(\Omega)$ and $p \in \mathcal{P}^{\log}(\Omega)$ such that $p_h \to p$ uniformly in Ω. Then

$$\Gamma(L^{p(\cdot)})(\Omega) - \lim_{h\to\infty} F_h = F.$$

Proof.

1. For all $u \in X$ such that $\|u\|_X \geq 1$ we have $\frac{F(u)}{\|u\|_X} \geq \|u\|_X^{p_- -1}$. This implies

$$\Rightarrow \lim_{\|u\|_X\to\infty} \frac{F(u)}{\|u\|_X} = \infty.$$

So F is coercive and clearly lower semi-continuous so we deduce the equicoerciveness by simply taking $\Psi = F$ in Proposition 1.2.4.

2. `liminf inequality`

Let $u_h \to u$ in $L^{p(\cdot)}(\Omega)$. If $\liminf_{h\to\infty} F_h(u_h) = +\infty$ there is nothing to prove. Thus, we may assume, without loss of generality, that $u_h \in W^{1,p_h(\cdot)}(\Omega)$ and, up to a subsequence

$$\liminf_{h\to\infty} F_h(u_h) = \lim_{h\to\infty} F_h(u_h) < +\infty$$

Since $p_h \to p$ uniformly, $\forall \epsilon > 0$, $\exists N_\epsilon \in \mathbb{N}\ \forall h \geq N_\epsilon\ |p_h - p| < \epsilon$

$$\Rightarrow\ p(x) - \epsilon < p_h(x)\ \forall x \in \Omega\ \forall h \geq N_\epsilon \tag{3.11}$$

Thus, for all $h \geq N_\epsilon$ we have $W_0^{1,p_h(\cdot)}(\Omega) \subset W_0^{1,p(\cdot)-\epsilon}(\Omega)$ and

$$\|\nabla u_h\|_{p(\cdot)-\epsilon} \leq (1 + |\Omega|)\|\nabla u_h\|_{p_h(\cdot)} \leq D \tag{3.12}$$

where D is a constant not depending on ϵ.

Then(u_h) is a bounded sequence in $W_0^{1,p(\cdot)-\epsilon}(\Omega)$ which is reflexive, hence there exists a subsequence (u_h) such that $u_h \rightharpoonup u$ weakly in $W_0^{1,p(\cdot)-\epsilon}(\Omega)$ with $u \in W_0^{1,p(\cdot)-\epsilon}(\Omega)$ for all $\epsilon > 0$ sufficiently small.

We shall next show that the limit $u \in W_0^{1,p(\cdot)}(\Omega)$ for ϵ sufficiently small.

By the lower semicontinuity of the modular and inequality (4.2) we have

$$\begin{aligned}\int_\Omega |\nabla u|^{p(x)-\epsilon}\,dx &\le \liminf_{h\to\infty} \int_\Omega |\nabla u_h|^{p(x)-\epsilon}\,dx \\ &\le \liminf_{h\to\infty} \max\{\|\nabla u_h\|^{p^+-\epsilon}_{p(\cdot)-\epsilon}, \|\nabla u_h\|^{p^--\epsilon}_{p(\cdot)-\epsilon}\} \\ &\le \liminf_{h\to\infty} \max\left\{\left[(1+|\Omega|)|\nabla u_h|_{p_h(\cdot)}\right]^{p^+-\epsilon}\right. \\ &\qquad\qquad \left., \left[(1+|\Omega|)|\nabla u_h|_{p_h(\cdot)}\right]^{p^--\epsilon}\right\} \\ &\le \max\left\{D^{p^+-\epsilon}, D^{p^--\epsilon}\right\} < +\infty\end{aligned}$$

Thus

$$\sup_{\epsilon>0}\int_\Omega |\nabla u|^{p(x)-\epsilon}\,dx < +\infty$$

and by Fatou's lemma

$$|\nabla u|^{p(x)} = \liminf_{\epsilon\to 0} |\nabla u|^{p(x)-\epsilon} \in L^1(\Omega)$$

It follows that

$$u \in W_0^{1,p(\cdot)}(\Omega)$$

By Young's inequality and (3.11) we have

$$\begin{aligned}\int_\Omega |\nabla u_h|^{p(x)-\epsilon}\,dx &\le \int_\Omega \tfrac{p(x)-\epsilon}{p_h(x)}|\nabla u_h|^{p_h(x)}\,dx + \int_\Omega \tfrac{p_h(x)-p(x)+\epsilon}{p_h(x)}\,dx \\ &\le \int_\Omega |\nabla u_h|^{p_h(x)}\,dx + \epsilon\int_\Omega \tfrac{1}{p_h(x)}\,dx + \tfrac{1}{p_h^-}\|p_h - p\|_\infty\end{aligned}$$

In view of the weak convergence to u in $W_0^{1,p(\cdot)-\epsilon}(\Omega)$ and the uniform convergence of p_h to p we have

$$\int_\Omega |\nabla u|^{p(x)-\epsilon}\,dx \le \liminf_{h\to\infty}\int_\Omega |\nabla u_h|^{p_h(x)}\,dx + \epsilon\frac{|\Omega|}{p^-}$$

Since the inequality holds for each $\epsilon > 0$, letting $\epsilon \searrow 0$ and applying Fatou's lemma we obtain

$$\int_\Omega |\nabla u|^{p(x)}\,dx \le \liminf_{h\to\infty}\int_\Omega |\nabla u_h|^{p_h(x)}\,dx$$

`limsup inequality`

Suppose that $u \in W_0^{1,p(\cdot)}(\Omega)$ (if $u \notin W_0^{1,p(\cdot)}(\Omega)$ $F_h(u) = +\infty$ and there is nothing to prove).Since $p(\cdot)$ is $\log-$Hölder continuous $\overline{C_0^\infty(\Omega)} = W_0^{1,p(\cdot)}(\Omega)$ and then it exists a sequence $(u_n) \subset C_0^\infty(\Omega)$ such that

$$u_n \to u \in W_0^{1,p(\cdot)}(\Omega)$$

$$\text{i.e} \int_\Omega |\nabla u|^{p(x)}\, dx = \lim_{n\to+\infty} \int_\Omega |\nabla u_n|^{p(x)}\, dx \tag{3.13}$$

On the other hand, for each $m \in \mathbb{N}$

$$|u_m|^{p_h(x)} \longrightarrow |u_m|^{p(x)} \text{ as } h \to \infty \text{ for a.e } x \in \Omega$$

and by the uniform convergence of p_h to $p\ \forall \epsilon > 0$ we can find $N_\epsilon \in \mathbb{N}$ such that $p_h < p + \epsilon$ $\forall\ h \geq N_\epsilon$ and then

$$|u_m|^{p_h(x)} \leq |u_m|^{p(x)+\epsilon} + 1 \in L^1(\Omega)\ \forall\ x \in \Omega\ \forall\ h \geq N_\epsilon,$$

whence by the Dominated Convergence Theorem,

$$\lim_{n\to\infty} \int_\Omega |u_m|^{p_h(x)}\, dx = \int_\Omega |u_m|^{p(x)}\, dx\ \forall m \in \mathbb{N}.$$

The proof is completed by Proposition 1.2.2 .

Theorem 3.2.2 *The first eigenvalue associated to the $p(x)-$Laplacian $\lambda^{p(\cdot)}$ is continuous with respect to the exponent $p(\cdot)$.*

Proof.

By Theorem 4.2.1 the functional F defined in (4.4) is the $\Gamma-$limit of F_h and since the functionals F_h are coercive $\forall\ h \in \mathbb{N}$ by Theorem 1.2.5

$$\inf_{u\in M_h} F_h \longrightarrow \inf_{u\in M} F \text{ as } h \to \infty$$

Define $\lambda^{p_h(\cdot)} := \inf_{u\in M_h} F_h(u)$ and $\lambda^{p(\cdot)} := \inf_{u\in M} F(u)$ we conclude that

$$\lambda^{p_h(\cdot)} \longrightarrow \lambda^{p(\cdot)}$$

when $p_h(\cdot) \to p(\cdot)$ uniformly.

CHAPTER 4

Stability of Eigenvalues of a Fourth Order Partial Differential Equation

Overview

In this chapter of the first part we present the results concerning the stability of the first eigenvalue associated to the $p(.)$Biharmonic operator under variations of the index of integrability $p(.)$.

The current chapter makes subject of publication [18].

4.1 Related Works and Main Result

Let Ω be a bounded set of $\mathbb{R}^N$, $N \geq 2$. The eigenvalue problem associated with the $p(x)-$Biharmonic operator is given by

$$\begin{cases} \operatorname{div}(|\Delta u|^{p(x)-2}\Delta u) = \lambda |u|^{p(x)-2}u \ \ in\ \Omega & ; \\ u = \Delta u = 0 \qquad\qquad\qquad\qquad on\ \partial\Omega & , \end{cases} \tag{4.1}$$

where $p(\cdot) \in \mathcal{P}^{log}(\Omega)$ which is defined as in Section 2.1.

A pair $(u, \lambda) \in W_0^{1,p(x)}(\Omega) \cap W^{2,p(x)}(\Omega) \times \mathbb{R}$ is a weak solution of (4.1) provided that

$$\int_\Omega |\Delta u|^{p(x)-2}\Delta u \Delta v \, dx = \lambda \int_\Omega |u|^{p(x)-2}uv \, dx, \tag{4.2}$$

$$\forall v \in X := W_0^{1,p(x)}(\Omega) \cap W^{2,p(x)}(\Omega)$$

Notice that if $u \neq 0$, then u is called an eigenfunction associated to the eigenvalue λ.

(4.2) is the Euler-Lagrange equation associated to the minimization of the Rayleigh ratio

$$R(u) := \frac{\int_\Omega |\Delta u|^{p(x)} \, dx}{\int_\Omega |u|^{p(x)} \, dx}.$$

For the existence result, based on Ljusternik–Schnirelmann Ayoujil and El Amrouss showed in [2] the existence of an infinitely many solutions of problem (4.1) $\{(\pm u_{n,p(.)}, \lambda_{n,p(.)})\}$ where

$$\lambda_{n,p(.)} = \inf_{K \in \Sigma_{n,p(.)}} \sup_{u \in K} \int_\Omega |\Delta u_{n,p(x)}|^{p(x)} \, dx \ n = 1, 2, ...,$$

where $\Sigma_{n,p(.)}$ denotes the set of symmetric and compact subsets of $\{u \in X \ : \ \int_\Omega |u|^{p(x)} \, dx = 1\}$ such that $i(K) \geq m$ where i denotes the Krasnoselskii's genus.

For $n = 1$ the first eigenvalue (denoted by $\lambda_{p(.)}$) is defined by

$$\lambda_{p(.)} := \inf_{u \in \mathcal{N}} \int_\Omega |\Delta u|^{p(x)}\, dx,$$

where the set

$$\mathcal{N} := \{u \in X / \int_\Omega |u|^{p(x)}\, dx = 1\}$$

is a closed manifold of class C^1. Here we index the eigenvalue by $p(.)$ to indicate the dependence on the exponent p. Throughout this chapter, the notion of the stability means exactly the continuity with respect the exponent $p(x)$.

For other articles studying the stability , the reader can refer to [10], [7], [25] and the references therein.

When seeking for papers treating the stability of eigenvalues for fourth order partial differential equations, we found just the paper of Benedikt [3], where the author dealt with the $p-$Biharmonic problem in the one dimensional case and showed the continue dependency on p of the eigenvalues; our result comes as a generalization of this work in the sense that we show the continuity with respect to $p(.)$ of the first eigenvalue of the $p(x)-$biharmonic problem in an open set $\Omega \subset \mathbb{R}^N$. Explicitly, show that the first eigenvalue associated to the $p(x)-$biharmonic problem $\lambda_{p(\cdot)}$ is continuous with respect to $p(\cdot)$, i.e., for all $h \in \mathbb{N}$

$$\lambda_{p_h(\cdot)} \longrightarrow \lambda_{p(\cdot)}\ as\ p_h(\cdot) \underset{h\to\infty}{\longrightarrow} p(\cdot)\ uniformly.$$

The Functional Framework

For any positive integer k we define the variable exponent Sobolev space $W^{k,p(\cdot)}(\Omega)$ by

$$W^{k,p(x)}(\Omega) := \{u \in L^{p(x)}(\Omega) :\ D^\alpha u \in L^{p(x)}(\Omega), |\alpha| \leq k\}.$$

Where D^α in the weak derivative of order α. Under the norm

$$\|u\|_{k,p(.)} := \sum_{|\alpha|\leq k} |D^\alpha u|_{p(x)}$$

$W^{k,p(\cdot)}(\Omega)$ is a Banach space. In fact, variable exponent Sobolev spaces are topological similar to classical Sobolev spaces: they are Banach spaces and they are reflexive and separable if and

only if $1 < p^- \leq p^+ < \infty$.

We endow X with the norm $\|.\|_X$ defined by

$$\|u\|_X = \inf \left\{ \mu > 0 : \int_\Omega \left(\left| \frac{u(x)}{\mu} \right|^{p(x)} + \left| \frac{\Delta u(x)}{\mu} \right|^{p(x)} \right) dx \leq 1 \right\}$$

Note that this norm is equivalent to $\|.\|_{2,p(.)}$ (see [2]). Endowed with this norm, X is a separable reflexive and Banach space.

We require $p(\cdot)$ to be in addition in $\mathcal{P}^{\log}(\Omega)$ to define $W_0^{1,p(x)}(\Omega)$ as the closure of $C_0^\infty(\Omega)$ (the space of indefinitely derivable functions with compact support) in $W^{1,p(x)}(\Omega)$ (see [21], Corollary 11.2.4).

4.2 The Stability Result

Define

$$M = \{u \in X \ / \int_\Omega |u|^{p(x)} \, dx = 1\}$$

and

$$M_h = \{u \in X_h := W_0^{1,p_h(x)}(\Omega) \cap W^{2,p_h(x)}(\Omega) \ / \int_\Omega |u|^{p_h(x)} \, dx = 1\}$$

Let $F : \ M \to [0, +\infty]$ defined by:

$$F(u) = \begin{cases} \int_\Omega |\Delta u|^{p(x)} \, dx & \text{if } u \in X; \\ +\infty & \text{otherwise.} \end{cases} \tag{4.3}$$

and $F_h : \ M_h \to [0, +\infty]$ defined by:

$$F_h(u) = \begin{cases} \int_\Omega |\Delta u|^{p_h(x)} \, dx & \text{if } u \in X_h; \\ +\infty & \text{otherwise.} \end{cases} \tag{4.4}$$

To show the stability of the first eigenvalue we have to prove the following convergence

$$\inf_{u \in M_h} \int_\Omega |\Delta u|^{p_h(x)} \, dx \longrightarrow \inf_{u \in M} \int_\Omega |\Delta u|^{p(x)} \, dx, \tag{4.5}$$

as $p_h(\cdot) \to p(\cdot)$ uniformly.

In view of Theorem 1.2.5, the convergence (4.5) will be attained if we show that the family F_h $\Gamma-$ converges to F, as $h \to \infty$ and F_h is equicoercive.

Our main concern in this section is to prove the following theorem.

Theorem 4.2.1 *Let $\Omega \subseteq \mathbb{R}^N$ be an open and bounded set and let $(p_h) \subset \mathcal{P}^{\log}(\Omega)$ and $p \in \mathcal{P}^{\log}(\Omega)$ such that $p_h \to p$ uniformly in Ω.Then*

1. *The family of functions (F_h) is equicoercive.*
2. $$\Gamma(L^{p(\cdot)})(\Omega) - \lim_{h\to\infty} F_h = F.$$

In the following we will prove the above assertions:

1. For all $u \in X$ such that $\|u\|_X \geq 1$ we have $\frac{F(u)}{\|u\|_X} \geq \|u\|_X^{p_- -1}$. This implies
$$\Rightarrow \lim_{\|u\|_X\to\infty} \frac{F(u)}{\|u\|_X} = \infty.$$
So F is coercive.

2. `liminf inequality`
Let $u_h \to u$ in $L^{p(\cdot)}(\Omega)$, if $lim\inf_{h\to\infty} F_h(u_h) = +\infty$ there is nothing to prove. Thus, we may assume without loss of generality, that $u_h \in X$.
Since $p_h \to p$ uniformly, $\forall \epsilon > 0\ \exists N_\epsilon \in \mathbb{N}\ \forall h \geq N_\epsilon$: $|p_h - p| < \epsilon$. Thus
$$.p(x) - \epsilon < p_h(x)\ \forall x \in \Omega\ \forall h \geq N_\epsilon. \tag{4.6}$$
Therefore for all $h \geq N_\epsilon$, $X_h \subset W^{2,p_h(.)}(\Omega) \subset W^{2,p(.)-\epsilon}(\Omega)$. Now by the embedding theorem we deduce that
$$\|u_h\|_{2,p(.)-\epsilon} \leq C\|u_h\|_X \leq C_1 \tag{4.7}$$
where C_1 is a constant not depending on ϵ.
Then (u_h) is a bounded sequence in $W^{2,p(\cdot)-\epsilon}(\Omega)$ which is reflexive. Hence there exists a subsequence (u_h) such that $u_h \rightharpoonup u$ weakly in $W^{2,p(\cdot)-\epsilon}(\Omega)$ with $u \in W^{2,p(\cdot)-\epsilon}(\Omega)$ for all $\epsilon > 0$ sufficiently small.
We shall next show that the limit $u \in X$.
We have $u \in W^{2,p(.)-\epsilon} \Rightarrow u \in W^{1,p(.)-\epsilon}(\Omega)$.
By the lower semicontinuity of the modular
$$\int_\Omega |\nabla u|^{p(x)-\epsilon}\, dx \leq \liminf_{h\to\infty} \int_\Omega |\nabla u_h|^{p(x)-\epsilon}\, dx.$$

On the other hands, by (4.6) $W_0^{1,p_h(.)}(\Omega) \subset W_0^{1,p(.)-\epsilon}(\Omega)$ and then

$$\|\nabla u_h\|_{p(.)-\epsilon} \leq (1+|\Omega|)\|\nabla u_h\|_{p_h(.)} \leq C_2.$$

So we obtain

$$\begin{aligned}\int_\Omega |\nabla u|^{p(x)-\epsilon}\,dx \leq \quad & \liminf_{h\to\infty} \max\{\|\nabla u_h\|_{p(.)-\epsilon}^{p^+-\epsilon}, \|\nabla u_h\|_{p(.)-\epsilon}^{p^--\epsilon}\} \\ & \leq \max\{C_2^{p^+-\epsilon}, C_2^{p^--\epsilon}\} < +\infty.\end{aligned}$$

Thus we get

$$\sup_{\epsilon>0}\int_\Omega |\nabla u|^{p(x)-\epsilon}\,dx < +\infty.$$

Now by Fatou's lemma

$$|\nabla u|^{p(x)} = \liminf_{\epsilon\to 0} |\nabla u|^{p(x)-\epsilon} \in L^1(\Omega)$$

it follows that $u \in W_0^{1,p(.)}(\Omega)$.

Still to show that $u \in W^{2,p(.)}(\Omega)$.

We have $u \in W^{2,p(.)-\epsilon}(\Omega)$, by the l.s.c. of the modular and inequality (4.7)

$$\begin{aligned}\int_\Omega |\Delta|^{p(x)-\epsilon}\,dx &\leq \liminf_{h\to\infty} \int_\Omega |\Delta u_h|^{p(x)-\epsilon}\,dx \\ &\leq \liminf_{h\to\infty} \int_\Omega \left(|\Delta u_h|^{p(x)-\epsilon} + |u_h|^{p(x)-\epsilon}\right)\,dx \\ &\leq \liminf_{h\to\infty} \max\{\|u_h\|_{2,p(.)}^{p^+-\epsilon}, \|u_h\|_{2,p(.)}^{p^--\epsilon}\} \\ &\leq \liminf_{h\to\infty} \max\{(C\|u_h\|_X)^{p^+-\epsilon}, (C\|u_h\|_X)^{p^--\epsilon}\} \\ &\leq \max\left\{C_1^{p^+-\epsilon}, C_1^{p^--\epsilon}\right\} < +\infty.\end{aligned}$$

Thus

$$\sup_{\epsilon>0}\int_\Omega |\Delta u|^{p(.)-\epsilon}\,dx \leq +\infty$$

and by Fatou's lemma $|\Delta u|^{p(x)-\epsilon} = \liminf_{\epsilon\to 0} |\Delta u|^{p(x)-\epsilon} \in L^1(\Omega)$.

Then we conclude that $u \in W^{2,p(.)}(\Omega)$. Which gives $u \in X$.

By Young's inequality and (4.5) we have

$$\begin{aligned}\int_\Omega |\Delta u_h|^{p(x)-\epsilon}\,dx &\leq \int_\Omega \tfrac{p(x)-\epsilon}{p_h(x)} |\Delta u_h|^{p_h(x)}\,dx + \int_\Omega \tfrac{p_h(x)-p(x)+\epsilon}{p_h(x)} \\ &\leq \int_\Omega |\Delta u_h|^{p_h(x)}\,dx + \epsilon \int_\Omega \tfrac{1}{p_h(x)}\,dx + \tfrac{1}{p_h^-}\|p_h - p\|_\infty.\end{aligned}$$

Since u_h converges weakly to u in $W^{2,p(.)-\epsilon}(\Omega)$ and $p_h(.)$ converges uniformly to $p(.)$ we have:

$$\int_\Omega |\Delta u|^{p(.)-\epsilon}\, dx \leq \liminf_{h\to\infty} \int_\Omega |\Delta u_h|^{p_h(x)}\, dx + \epsilon \frac{|\Omega|}{p^-}\ \forall \epsilon > 0.$$

Passing to the limit $\epsilon \to 0^+$ and taking into account Fatou's lemma again, we get

$$\int_\Omega |\Delta u|^{p(x)}\, dx \leq \liminf_{h\to\infty} \int_\Omega |\Delta| u^{p_h(x)}\, dx.$$

`limsup inequality`

Suppose that it exists $u \in X$, for all $(u_h) \subset X$ such that $u_h \to u$ in X

$$F(u) < \limsup_{h\to\infty} F_h(u_h).$$

Thus

$$\forall k \in \mathbb{N}\ \ \exists N \geq k \ such\ that\ F_N(u_N) > F(u).$$

That is,

$$\int_\Omega |\Delta u_N|^{p_N(x)}\, dx > \int_\Omega |\Delta u|^{p(x)}\, dx.$$

Therefore

$$\liminf_{N\to\infty} \int_\Omega |\Delta u|^{p_N(x)}\, dx > \liminf_{N\to\infty} \int_\Omega |\Delta u|^{p(x)}\, dx.$$

Then

$$\liminf_{N\to\infty} \int_\Omega |\Delta u|^{p_N(x)}\, dx > \int_\Omega |\Delta u|^{p(x)}\, dx.$$

This means that $\exists\, l \in \mathbb{N}$ such that $\forall m \in \mathbb{N}$ if $m \leq l$ then

$$\int_\Omega |\Delta u|^{p_m(x)} > \int_\Omega |\Delta u|^{p(x)}\, dx.$$

Which contradicts the uniform convergence of $p_h(.)$ (and then $p_m(.)$ as a subsequence) to $p(.)$. Finally, we conclude that for every $u \in X$ there exists a sequence $(u_h) \subset X$ such that $u_h \to u$ in X and

$$F(u) \geq \limsup_{h\to\infty} F_h(u_h).$$

Theorem 4.2.2 *The first eigenvalue associated to the $p(x)-$Biharmonic $\lambda_{p(\cdot)}$ is continuous with respect to the exponent $p(\cdot)$.*

Proof.

By Theorem 4.2.1 the functional F defined in (4.4) is the $\Gamma-$limit of F_h and since the functionals F_h are coercive for all $h \in \mathbb{N}$ by Theorem 1.2.5, we deduce that

$$\inf_{u\in M} F_h \longrightarrow \inf_{u\in M} F \text{ as } h \to \infty.$$

Now define

$$\lambda^{p_h(\cdot)} := \inf_{u\in\mathcal{M}} F(u) \quad \text{and} \quad \lambda^{p(\cdot)} := \inf_{u\in\mathcal{M}} F(u)$$

we conclude that

$$\lambda^{p_h(\cdot)} \longrightarrow \lambda^{p(\cdot)},$$

when $p_h(\cdot) \to p(\cdot)$ uniformly.

biblo

[1] E. Acerbi and G. Mingione. Regularity results for a class of functionals with non-standard growth. *Arch. Rational Mech. Anal.*, 156:121–140, 2001.

[2] A. Ayoujil and A.R. El Amrouss. On the spectrum of a fourth order elliptic equation with variable exponent. *Nonlinear Analysis: Theory, Methods Applications*, 71(10):4916 – 4926, 2009.

[3] Jiří Benedikt. Continuous dependence of eigenvalues of p-biharmonic problems on p. *Communications on Pure and Applied Analysis*, 12, 05 2013.

[4] C. Bennet and R. Sharpley. *Interpolation of operators*. Academic Press, New York, 1988.

[5] M. Bocea and M. Mihăilescu. On the continuity of the luxembourg norm of the gradient in $l^{p(.)}$ with respect to $p(.)$. *Proceedings of the American Mathematical Society*, 142(2):507–517, 2013.

[6] T. Champion and L. De Pascale. Asymptotic behaviour of nonlinear eigenvalue problems involving $p-$laplacian-type operators. *Proceedings of the Royal Society of Edinburgh: Section A Mathematics*, 137(06):1179–1195, 2007.

[7] F. Colasuonno and M. Squassina. Stability of eigenvalues for variable exponent problems. *Nonlinear Analysis: Theory, Methods & Applications*, 123:56–67, 2015.

[8] M. Degiovanni and M. Marzocchi. Limit of minmax values under γ convergence. *Electron. J. Differential Equations*, 2014:1–19, 2014.

[9] M. Degiovanni and M. Marzocchi. On the dependence on p of the variational eigenvalues of the $p-$laplace operator. *Potential Anal.*, 43(4):593–609, 2015.

[10] A. El Khalil, P. Lindqvist, and A. Touzani. On the stability of the first eigenvalue of $a_p u + \lambda g(x)|u|^{p-2}u = 0$ with varying p. *Rend. Mat. Appl.*, (7)24:321–336, 2004.

[11] L. Evans. *Partial Differential Equations*. American Mathematical Society, second ed., 2010.

[12] X. L. Fan and D. Zhao. On the generalised orlicz-sobolev space $w^{k,p(x)}(\omega)$. *J. Gansu Educ.*, 12(1):1–6, 1998.

[13] G. Franzina and P. Lindqvist. An eigenvalue problem with variable exponents. *Nonlinear Analysis: Theory, Methods and Applications*, 85:1–16, 2016.

[14] H. Hakkarainen and M. Nuortio. The variable exponent sobolev capacity and quasi-fine properties of sobolev functions in the case $p^- = 1$. *Journal of Mathematical Analysis and Applications*, 412(1):168–180, 2014.

[15] P. Harjulehto and P. Hästö. Sobolev inequalities for variable exponents attaining the values 1 and n. *Publ. Mat.*, 52(2):347–363, 2008.

[16] A. El Khalil I. Brahmi and A. Touzani. Stability of eigenvalues of a nonlinear variational problem involving $p(x)-$growth condition with varying p. *Accepted in Afrika Mathematika*.

[17] A. El Khalil I. Brahmi and A. Touzani. $\gamma-$convergence: an application to eigenvalue problems. *Italian Journal of Pure and Applied Mathematics*, 42:449–457, 2019.

[18] A. El Khalil I. Brahmi and A. Touzani. Stability of eigenvalues of a fourth order partial differential equation. *International Journal of Mathematical Analysis*, 13(3):123–133, 2019.

[19] E. De Pascale J. Appel and A. Vignoli. *Nonlinear Spectral Theory*. Walter De Gruyter, Berlin, 2014.

[20] J. Jost and X. Li-Jost. *Calculus of variations*. Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge.

[21] P. Hästö L. Diening, P. Harjulehto and M. Rúžička. *Lebesgue and Sobolev Spaces with Variable Exponents*. Lecture Notes in Mathematics - Springer-Verlag, Heidelberg, 2011.

[22] J. Lang and O. Méndez. Stability of the norm-eigenfunctions of the $p(.)-$laplacian. *Integral Equations and Operator Theory*, 1:1–13, 2016.

[23] J. Lindenstrauss and L. Tzafriri. *Classical Banach spaces I and II*. Springer-Verlag, Berlin-Heidelberg-New York.

[24] P. Lindqvist. Uber konjugierte exponentenfolgen. *Studia Math.*, 3:200–211, 1931.

[25] P. Lindqvist. Stability for the solutions of $\mathrm{div}(\nabla u|^{p-2}\nabla u)$= f with varying p. *Journal of Mathematical Analysis and Applications*, 127(1):93–10, 1987.

[26] P. Lindqvist. On nonlinear rayleigh quotients. *Potential Anal.*, 2(3):199–218, 1993.

[27] G. Dal Maso. *An introduction to* $\Gamma - converegnce$. Springer Science & Business Media, 1993.

[28] E. Parini. Continuity of the variational eigenvalues of the $p-$laplacian with respect to p. *Bulletin of the Australian Mathematical Society*, 83:376–381, 2011.

[29] K. Perera and M. Squassina. Asymptotic behavior of the eigenvalues of the $p(x)-$laplacian. *Manuscripta Math.*, 144:535–544, 2014.

[30] K. R. Rajagopal and M. Rúžička. On the modeling of electrorheological materials. *Mech. Research Comm.*, 23:401–407, 1996.

[31] K. R. Rajagopal and M. Rúžička. Mathematical modeling of electrorheological materials. *Cont. Mech. and Thermodynamics*, 13:59–78, 2001.

[32] A. Szulkin. Ljusternik–schnirelmann theory on c1-manifolds. *nn. Inst. H. Poincaré Anal. Non Linéaire*, 5:119–139, 1988.

[33] Q. Zhang X. Fan and D. Zhao. Eigenvalues of $p(x)-$laplacian dirichlet problem. *Journal of Mathematical Analysis and Applications*, 302(2):306–317, 2005.

Part II

Deepness Analysis in Prototypes based Unsupervised Learning

Introduction and General Context

One of the most interesting features of machine learning is that it lies on the boundary of several different academic disciplines, principally computer science, statistics, mathematics, and engineering. This has been a problem as well as an asset, since these groups have traditionally not talked to each other very much. To make it even worse, the areas where machine learning methods can be applied vary even more widely, from finance and business [43] to biology [36, 42, 70] and medicine [86, 88] to physics and chemistry [90] and beyond [34]. Around the world, computers capture and store terabytes of data every day. There are computers belonging to shops, banks, hospitals, scientific laboratories, and many more that are storing data incessantly. For example, banks are building up pictures of how people spend their money, hospitals are recording what treatments patients are on for which ailments (and how they respond to them). The challenge is to do something useful with this data: if the bank's computers can learn about spending patterns, can they detect credit card fraud quickly? If hospitals share information, then can treatments that don't work as well as expected be identified quickly? These are some of the questions that machine learning methods can be used to answer.

Science has also taken advantage of the ability of computers to store massive amounts of data. Biology has led the way, with the ability to measure gene expression in DNA micro-arrays producing immense data sets [50], along with protein transcription data and phylogenetic trees relating species to each other. However, other sciences have not been slow to follow. Astronomy [37] now uses digital telescopes, so that each night the world's observatories are storing incredibly high-resolution images of the night sky, around a terabyte per night. Equally, medical science stores the outcomes of medical tests from measurements as diverse as Magnetic Resonance Imaging (MRI) [96] scan and simple blood tests. The explosion in stored data is well known; the challenge is to do something useful with that data.

The size and complexity of these data sets means that humans are unable to extract useful information from them. Even the way that the data is stored works against us. Given a file full of numbers, our minds generally turn away from looking at them for long. Take some of the same data and plot it in a graph and we can do something, the graph is rather easier to look at and deal with. Unfortunately, our three-dimensional world doesn't let us do much with data in higher

dimensions. This is known as the curse of dimensionality. There are two things that we can do with this: reduce the number of dimensions (until our simple brains can deal with the problem) or use computers, which don't know that high-dimensional problems are difficult, and don't get bored with looking at a massive data files of numbers.

This is one reason why machine learning is becoming so popular. The problems of our human limitations go away if we can make computers do the dirty work for us.

Machine learning, then, is about making computers modify or adapt their actions so that these actions get more accurate, where accuracy is measured by how well the chosen action reflect the correct ones. It is only over the past decade or so that the inherent multi-disciplinarity of machine learning has been recognized. It merges ideas from neuroscience and biology, statistics, mathematics, and physics, to make computers learn.

Application of machine learning methods to large databases is called data mining [34,54,61,63]. The analogy is that large volume of earth and raw material is extracted from a mine, which when processed leads to a small amount of very precious material; similarly, in data mining, a large volume of data is processed to construct a simple model with valuable use, for example, having high predictive accuracy. Its application areas are abundant: In addition to retail, in finance banks analyze their past data to build models to use in credit applications, fraud detection, and the stock market. In manufacturing, learning models are used for optimization, control, and troubleshooting. In medicine, learning programs are used for medical diagnosis. In telecommunications, call patterns are analyzed for network optimization and maximizing the quality of service. In science, large amounts of data in physics, astronomy, and biology can only be analyzed fast enough by computers. The World Wide Web is huge; it is constantly growing, and searching for relevant information cannot be done manually. Example 1: A search engine (e.g., Google) receives hundreds of millions of queries every day. Each query can be viewed as a transaction where the user describes her or his information need. What novel and useful knowledge can a search engine learn from such a huge collection of queries collected from users over time? Interestingly, some patterns found in user search queries can disclose invaluable knowledge that cannot be obtained by reading individual data items alone. For example, Google's Flu Trends uses specific search terms as indicators of flu activity. It found a close relationship between the number of people who search for flu-related information and the number of people who actually have flu symptoms. A

pattern emerges when all of the search queries related to flu are aggregated. Using aggregated Google search data, Flu Trends can estimate flu activity up to two weeks faster than traditional systems can [59]. This example shows how data mining can turn a large collection of data into knowledge that can help meet a current global challenge. Example 2: A supermarket chain that has hundreds of stores all over a country selling thousands of goods to millions of customers. The point of sale terminals record the details of each transaction: date, customer identification code, goods bought and their amount, total money spent, and so forth. This typically amounts to gigabytes of data every day. What the supermarket chain wants is to be able to predict who are the likely customers for a product. Again, the algorithm for this is not evident; it changes in time and by geographic location. The stored data becomes useful only when it is analyzed and turned into information that we can make use of, for example, to make predictions. We do not know exactly which people are likely to buy this ice cream flavor, or the next book of this author, or see this new movie, or visit this city, or click this link. If we knew, we would not need any analysis of the data; we would just go ahead and write down the code. But because we do not, we can only collect data and hope to extract the answers to these and similar questions from data. We do believe that there is a process that explains the data we observe. Though we do not know the details of the process underlying the generation of data, for example customer behavior, we know that it is not completely random. People do not go to supermarkets and buy things at random. We may not be able to identify the process completely, but we believe we can construct a good and useful approximation. That approximation may not explain everything, but may still be able to account for some part of the data. We believe that though identifying the complete process may not be possible, we can still detect certain patterns or regularities. This is the niche of machine learning. Such patterns may help us understand the process, or we can use those patterns to make predictions: Assuming that the future, at least the near future, will not be much different from the past when the sample data was collected, the future predictions can also be expected to be right. Based on the available information and on the desired objectives, there are different types of Machine Learning Algorithms:

- Supervised learning (Classification): A training set of examples with the correct responses (targets) are provided and, based on this training set, the algorithm generalizes to respond

correctly to all possible inputs. This is also called learning from exemplars. [46, 58, 87]

- Unsupervised learning (Clustering): Correct responses are not provided, instead the algorithm tries to identify similarities between the inputs so that inputs that have something in common are categorized together. [38, 62, 69, 74].

- Reinforcement learning: This is somewhere between supervised and unsupervised learning. The algorithm gets told when the answer is wrong, but does not get told how to correct it. It has to explore and try out different possibilities until it works out how to get the answer right . Reinforcement learning is sometimes called learning with a critic because of this monitor that scores the answer, but does not suggest improvements. [75, 107, 112]

- Evolutionary learning: Biological evolution can be seen as a learning process: biological organisms adapt to improve their survival rates and chance of having offspring in their environment. [93, 113] The Unsupervised Learning (Clustering) is going to be the focus of this thesis. Data Visualization equally. So, we'll have a look at what it is, and the kinds of problems that can be solved using it.

In this second part of this thesis we propose a deepness analysis of levels of data representation, based on Self Organizing Maps and Nonnegative Matrix Factorization, we will pursue the following plan

- In the first chapter we give a general essay of methods of dimensionality reduction since they are an innovative and important tools in the fields of data analysis, data mining, and machine learning. We also give an overview on deep learning methods that aim at learning feature hierarchies with features from higher levels of the hierarchy formed by the composition of lower level features. We then focus on two particular methods: Nonnegative Matrix Factorization and Self Organizing Maps, two robust algorithms that are able to reduce the data dimensionality in a nonlinear way, and that are our main concern in deepness analysis through this part of thesis.

- The second chapter is devoted to the first contribution of the in the field of deep learning. Hierarchical structures have been known for decades for their exceptional properties that

make them ideal for the representation of data and were suggested as a particularly important method for the organization of concepts. The contribution proposed in this chapter, fills a gap Matrix Factorisation (NMF) hierarchical structures. The candidates proves that this model based on prototypes is a truly profound architecture. We provemathematically and experimentally that each layer depends on the previous layers. We then conclude that different layers of the multilayered nonnegative matrix factorisation are not only dependent but that the order of their construction is predominant. In in other words, the multilayer NMF model is indeed a method of reducing dimensionality and hierarchical classification. This involves learning how to several levels of representation, corresponding to different levels of abstractions.

- In the last chapter we explore an original strategy for the construction of deep networks, based on stacking layers of SOM (Self Organizing Maps) with finished weights. Its objective is to show that our approach, with enough hidden variables, is able to represent any “soft” distribution on visible variables, where "soft" means that the distribution contains no probability of 1 or 0. The proposed algorithm is tested on artificial and real data sets. The effect of added hidden layers for data structure representation is experimentally demonstrated.

CHAPTER 5

Unsupervised Learning, Deep Learning, Prototype Based Models and Dimensionality Reduction Algorithms

5.1 Unsupervised Learning

Human beings seem to be able to learn without explicit supervision. One aim of unsupervised learning is to mimic this aspect of human learning; hence, this type of learning tends to use methods that are more plausible from a biological perspective than those using error descent methods. For example, such algorithms involve local processing at each synapse and it is not necessary for global information to pass through the network. So an unsupervised network must self-organise with respect to its internal parameters, without external prompting, and to do so, it must react to some aspect of the input data. Typically this will be either redundancy in the input data or clusters in the data; i.e. there must be some structure in the data to which it can respond. One essential exploitation of unsupervised learning is that it's an important approach to exploratory analysis of unlabelled data, but remains a difficult problem. Without a priori knowledge on the structure of a database, clustering just detect automatically the presence of relevant subgroups (or clusters). A cluster can be defined as a set of similar data between them and not very similar with data belonging to another cluster (internal homogeneity and external separation). Clusters can also be described as regions of the data representation space containing a relatively high density of data points, separated by a relatively low density zone. The detection of these clusters plays an indispensable role in understanding varied phenomena described by a set of observations. Many classification methods have been proposed in the literature. The more conventional are hierarchical and biased methods. Although these methods have long been very popular, they are increasingly replaced using methods such as a self-organizing card or Self Organizing Map (SOM) [80, 84].

5.2 Hierarchical Structures and Deep Learning

Theoretical results suggest that in order to learn the kind of complicated functions that can represent high-level abstractions (e.g., in vision, language, and other AI-level tasks), one may need deep architectures. Deep architectures are composed of multiple levels of non-linear operations, such as in neural nets with many hidden layers or in complicated propositional formulae re-using many sub-formulae. Searching the parameter space of deep architectures is a difficult

task, but learning algorithms such as those for Deep Belief Networks have recently been proposed to tackle this problem with notable success, beating the state of-the-art in certain areas. This section discusses the motivations and principles regarding learning algorithms for deep architectures, in particular those exploiting as building blocks unsupervised learning of single-layer models such as Restricted Boltzmann Machines, used to construct deeper models such as Deep Belief Networks. This is addressed specially to emphasize the formal and right definition of deep architectures, and to justify why in particular we want to migrate from classical Multilayer Architectures to Deep Architectures, the main purpose in this thesis.

Depth of architecture refers to the number of levels of composition of non-linear operations in the function learned. Whereas most current learning algorithms correspond to shallow architectures ($1, 2$ or 3 levels), the mammal brain is organized in a deep architecture [102] with a given input percept represented at multiple levels of abstraction, each level corresponding to a different area of cortex. Humans often describe such concepts in hierarchical ways, with multiple levels of abstraction. The brain also appears to process information through multiple stages of transformation and representation. This is particularly clear in the primate visual system [102], with its sequence of processing stages: detection of edges, primitive shapes, and moving up to gradually more complex visual shapes. Inspired by the architectural depth of the brain, neural network researchers had wanted for decades to train deep multi-layer neural networks [40, 109], but no successful attempts were reported before 2006: researchers reported positive experimental results with typically two or three levels (i.e., one or two hidden layers), but training deeper networks consistently yielded poorer results. Something that can be considered a breakthrough happened in 2006: Hinton et al. at University of Toronto introduced Deep Belief Networks (DBNs) [68], with a learning algorithm that greedily trains one layer at a time, exploiting an unsupervised learning algorithm for each layer, a Restricted Boltzmann Machine (RBM) [56]. Shortly after, related algorithms based on auto-encoders were proposed [39, 97], apparently exploiting the same principle: guiding the training of intermediate levels of representation using unsupervised learning, which can be performed locally at each level.

5.3 Prototype based-models and Dimensionality Reduction Methods

In machine learning we distinguish parametric and non parametric models, the first category involves two steps:

1. Select a form of a learning function
2. Learn the coefficients of this function from the training data

They don't require a lot of training data and so show to be very fast when we want to learn from data.
When we have a lot of data and want to take advantage from, either by reducing the dimensionality, extracting features or seeking to construct a mapping function, we use non parametric methods, also called prototype based models.
Dimensionality reduction techniques can be categorized in two classes:

- Feature Selection and
- Feature Extraction

A feature selection method is a process that selects a subset of k original (and supposed relevant) features for spanning a reduced space that may better describe the phenomena of interest. Feature selection mechanisms reduce the computational costs, but a good trade-off between accuracy of the results and efficiency is needed. On the other hand, feature extraction methods try to capture hidden properties of data and discover the minimum number of uncorrelated or lowly correlated factors that can be used to better describe the phenomena of interest. It is accomplished by the creation of new features obtained as functions of the original data. Reduction of the computational complexity of data both in time (for elaboration) and in space (for storage) and the discovery of latent structure hidden in data, (meaningful structures and/or unexpected relationships among variables) are some of the advantages resulting from feature extraction methods.

Data projection or visualization, which facilitates the analysis of internal data set structures for the human expert, figures prominently among the applications of unsupervised learning methods. This can be achieved by data projection over more informative axes or by generating maps that represent the inner structure of data sets. Techniques such as Hebbian learning and Self-Organizing Maps (SOM) are probably the most widely used technique.
Topology-preserving maps consist of algorithms that visualize and interpret large high-dimensional data sets, which means that the topology preserving map is a useful tool for data mining by visual inspection. Typical applications are visualization of process states or financial results by representing instances of central dependency in the map data.
In the following we will point out two fundamental non parametric algorithms that are robust in dimensionality reduction and visualization and make subject of deepness analysis in this thesis

5.3.1 Nonnegative Matrix Factorization

Nonnegative Matrix Factorization (abbreviated in NMF) is a linear algebra basic modeling approach of dimensionality reduction adapted to matrices with high dimension [89]. By decomposing a data set into two matrices, a prototypes and a partition (Latent components) matrices, it allows retrieving sparse and easily interpretable features, which have made of it an efficient widely used tool in several issues, particularly in high dimensional data analysis. It also serves in clustering tasks, for topics recovery and in temporal segmentation. NMF was introduced first by Lee and Seung in [89] where they present an algorithm that learns parts of faces and semantic features of text. From then on NMF has found several potential scientific and engineering applications including:

- Image processing ([103] and [49]): representation of faces images, classification of images.
- Text processing ([114], [41] and [76]): classification of documents, extraction of semantic characters in texts.
- Source separation ([111] and [95]): separating voices in speech mixtures and singing voices in polyphonic mixtures.

- Economy ([53]): diversification of actions.
- Biology ([45] and [57]): clustering of genes implicated in cancer.

There exists several variants of NMF (see [44, 71, 94, 106] and the references therein). For example, symmetric NMF, semi-orthogonal NMF and three factor NMF were developed by imposing additional constraints to the resulting matrices in order to solve some gaps from which basic NMF suffers (e.g. imposing sparsity leads to have unique factors).

Basically, NMF seeks to estimate F and G knowing only X. This is generally achieved by minimizing a cost function $D(X||FG)$ and setting all negative elements of F and G to 0 or some small $\epsilon > 0$. Explicitly, the problem can be formulated in the following way:

$$\min_{F \in \mathbb{R}_+^{d \times k}, G \in \mathbb{R}_+^{k \times N}} D(X||FG) \tag{5.1}$$

Note that the estimation of the unknown matrices allows an unsupervised classification. The choice of the cost function depends on the probability distribution of the estimated components and on the data structure. The most used cost functions in the literature are the Frobenius norm:

$$D_F(X||FG) = \frac{1}{2}\|X - FG\|_F^2$$

and Kullback-Leiber norm, also referred as I-divergence:

$$D_{KL}(X||FG) = Xlog(\frac{X}{FG}) - X + FG$$

For more details about update rules associated to each of the above cost functions see [52].

The lack of convexity of the optimization problem (5.1) with respect to both F and G is an ill-conditioned problem that enhance the risk of converging to local minima of $D(X||FG)$. To alleviate this problem, Cichoki and Zdunek developed in [51] a multilayer approach where, the basic matrix F is replaced by a set of cascaded (factor) matrices. Besides many applications of Multilayer NMF in medicine and physics ([48], [35]), there exists many works that analyzed the behavior of different resulting factors versus the increasing number of layers, for example the authors in [98] found that the overall sparsity decreases within layers as we decompose more our

initial data matrix.

After obtaining prototypes and partition matrices F_1 and G_1 from initial data matrix X, they decompose again the resulting partition matrix F_1 via standard NMF to new prototypes and partition matrices F_2 and G_2. Multilayer NMF consists of repeating this process as many times as the defined number of wanted layers. Thus the model can be described as:

$$X = F_1 F_2 ... F_L G_L + E$$

where E is the error of approximation.

From now on and to simplify the writing we will neglect the error of approximation E in the equations below. In the first step, we perform the standard NMF decomposition:

$$X \cong F_1 G_1 \in \mathbb{R}^{d \times N}$$

In the second step, the results obtained from the first step are used as a new input data to NMF algorithm:

$$X \leftarrow G_1$$

we perform a similar factorization :

$$G_1 \cong F_2 G_2 \in \mathbb{R}^{k \times N}$$

After L steps (L is defined by the user), the multilayer NMF decomposition has the form:

$$X \cong F_1 F_2 ... F_L G_L$$

With $F = F_1 F_2 ... F_L$ and $G = G_L$.

5.3.2 Self Organizing Maps

5.3.2.1 Background and Notations

We recall the general principle of the now very popular self-organizing map (SOM) algorithm originally devised by Teuvo Kohonen in 1982 [77,78] and revisited in [79,81–83,85]. Kohonen's

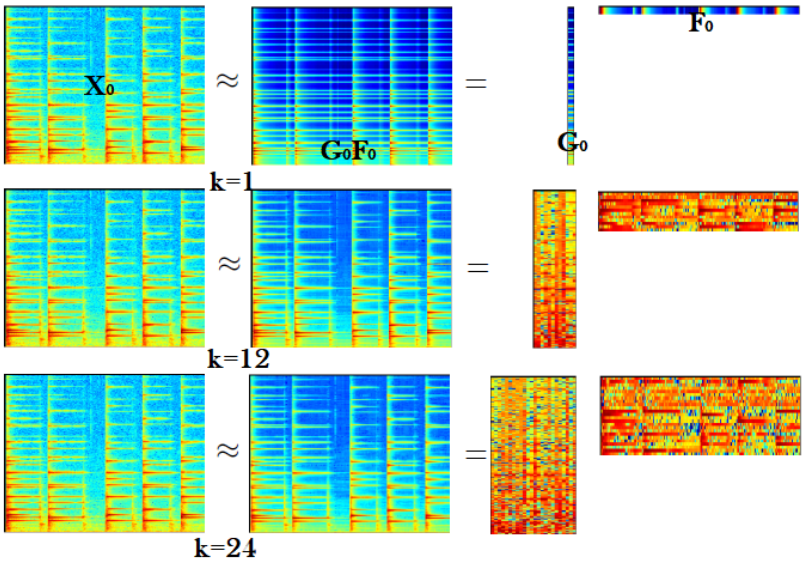


Figure 5.1: Standard Nonnegative Matrix Factorization for different orders of decomposition

Self-Organizing Map (SOM) [5] is a data analysis method that combines vector quantization with topology preservation. With each quantizer we associate a fixed location in a 'latent' space. The latent space is of much lower dimension (typically just two) than the data space. The SOM algorithm finds locations for the quantizers in the data space such that the summed squared distance from data to closest node is small and simultaneously topology is preserved. Topology preservation means that nearby components in latent space are also nearby in the data space. Topology preservation makes SOMs useful for data visualization and dimensionality reduction. It was presented as a model for the self-organization of neural connections. What immediately raised the interest of the scientific community was the *self-organization* property of such a simple algorithm to produce organization. The SOM map consider a finite number of cells located in a regular ordered grid G. We define the proximity on G as follows: for each pair of cells (c, r) on the map, $\delta(c, r)$ is the length of the unique shortest path linking cells c and r on the grid, i.e. the Manhattan distance.

For each cell c this distance defines a topological neighborhood whose influence is controlled by a positive, symmetrical, monotonically time dependent kernel function with rapid decay to

infinity as follows:

$$\mathcal{K}_t(c,r) = \frac{1}{\sigma_t}\exp\left(-\frac{\delta^2(c,r)}{\sigma_t^2}\right) \text{ with } \sigma_t = \sigma_0\left(\frac{\sigma_f}{\sigma_0}\right)^{\frac{t}{t_{max}}} \tag{5.2}$$

where σ_t is the temperature's function modeling the neighborhood's range, σ_0 and σ_f represent respectively the initial and the final temperature (e.g. $\sigma_0 = 2$ and $\sigma_f = 0.5$), t is the time instance and t_{max} - is the maximum allotted time. The size of this neighborhood is limited by the standard Gaussian deviation σ_t. The range σ_t is a function decreasing in time, so, the neighborhood function $\mathcal{K}_t(c,r)$ will have the same trend with a standard deviation decreasing in time. Therefore the value of $\mathcal{K}_t(c,r)$ decreases with increasing distance between c and r, and $\mathcal{K}_t(c,c) = 1$. Note that the neighborhood function can be constant or time dependent [77,78,81,83]. In this way the map G is endowed with a neighborhood system by considering the mutual influence of the two cells c and r defined by $\mathcal{K}_t(c,r)$.

We associate to each cell $c \in G$ the m-dimensional weight vector w_c. The network state at some fixed time is given by:

$$\mathcal{W} = (w_1,\ w_2,\ \ldots,\ w_{|G|}),\ \text{where}\ \mid G \mid\ \text{is the cardinality of } G.$$

Let $X = \{x^1, \ldots, x^N\} \subset \mathbb{R}^m$ be a finite set of inputs indexed by $n = 1, \ldots, N$. We define for a given state $\mathcal{W}$ and input x the **winning** unit c^* as the unit whose weight w_{c^*} is the closest to the input x.

The SOM algorithm usually unfolds in two phases: at first a large adaptation parameter is used so that the presented data affect several prototypes on large neighborhood area. In this second phase, the adaptation parameter decreases toward 0, and the neighborhood gets smaller until only a single unit is affected by each data. The SOM algorithm for some input x starts at $t = t_i$, ends at $t = t_f$ and is recursively defined by:

$$\begin{cases} c^* := \operatorname{argmin}_{c\in G}\{d(x, w_c)\} \\ w_c|_{new} = w_c|_{old} - \sigma_t\mathcal{K}_t(c, c^*)(w_c|_{old} - x), \forall c \in G. \end{cases} \tag{5.3}$$

Here $d(\cdot,\cdot)$ denotes the square of the euclidean distance in $\mathbb{R}^m$ and $\sigma_t \in \mathbb{R}$ is the learning rate or the adaptation gain parameter which is $]0,1[$-valued decreasing with time by construction.

Thus the network defines a map $\Phi_{\mathcal{W}} : x \longmapsto c^*(x)$, from X to G, and the goal of the learning algorithm is to converge to a network state such the $\Phi_{\mathcal{W}}$ map will be "topology preserving"in some sense. For a given state $\mathcal{W}$, let us denote $P_c(\mathcal{W})$ the set of the inputs such that c is the **winning** unit, that is $P_c(\mathcal{W}) = \Phi_{\mathcal{W}}^{-1}(c)$. Therefore the set of the classes $P_c(\mathcal{W})$ is the Euclidean Voronoi tessellation of the convex hull of X related to $\mathcal{W}$. The essential parameters are: (1) the dimension m of the input space, (2) the topology of the network and (3) the distribution probability of the inputs. Several variations of the original SOM algorithm have been proposed in the literature: (i) different divergence measures have been proposed or (ii) alternative learning algorithms for SOMs have been proposed [47, 110]. In this paper we show how to combine the benefits of self-organizing maps and hierarchical probabilistic models. We present a general learning algorithm, similar to Kohonen's original SOM algorithm, based on the standard EM learning algorithm using a slightly modified expectation-step.

5.3.2.2 Energy of SOM

We recall the energy formulation of SOM introduced in [65]. In what follows, let p_c^n be the probability that the input $x^n \in X$ is assigned to the cell c with the weight w_c. Therefore $\sum_{c\in G} p_c^n = 1$, and $p_c^n \geq 0$. When we assign the input x to the cell map c, there exists a confusion probability given by $\mathcal{K}_t(c, r)$, that the input x^n is instead represented by the the weight vector w_r associated to some cell r. To ensure the definition of the confusion probability, we assume that for all $c \in G$ we have $\sum_{r\in G} \mathcal{K}_t(c, r) = 1$. The goal of SOM is to find the probability assignment $\mathcal{P} = \{p_c^n; c \in G, n = 1, ..., N\}$ and the weights $\mathcal{W} = \{w_c; c \in G\}$ which minimize the following error function:

$$E(\mathcal{P}, \mathcal{W}) = \sum_{n=1}^{N} \sum_{c\in G} p_c^n \sum_{r\in G} \mathcal{K}_t(c, r) d(x^n, w_r), \tag{5.4}$$

where recall that d is the square of the euclidian distance in $\mathbb{R}^m$. For a probability assignment $\mathcal{P}$ and the inputs X we define the entropy term by:

$$H(\mathcal{P}) = \sum_{n=1}^{N} \sum_{c\in G} p_c^n \log\left(\frac{p_c^n}{q_c}\right), \text{ where } q_c = 1/ \mid G \mid . \tag{5.5}$$

With this definitions in mind, the " free energy functional" of the SOM is a weighted combination of (5.4) and (5.5) and is given by:

$$F(\mathcal{P}, \mathcal{W}) = \beta E(\mathcal{P}, \mathcal{W}) + H(\mathcal{P}), \tag{5.6}$$

the parameter $\beta > 0$ allows to tune the influence of the entropy term. Here the neighborhood function, implemented by the $\mathcal{W}$, is fixed, but the winner assignment is soft. Instead of selecting one 'winner' an unconstrained distribution over the components is used: the $\mathcal{P}$. The β is used for annealing: for very small β the entropy term, with only one global optimum, becomes dominant, whereas for large β the quantization error, with many local optima, becomes dominant. By gradually increasing β more and more structure is added to the objective function.

To learn this parameter, we start with $\beta = \frac{1}{2}$ and follow a descend gradient strategy. The gradient of the " free energy functional" with respect to β writes:

$$\nabla_{\beta} F(\mathcal{P}, \mathcal{W}) = E(\mathcal{P}, \mathcal{W}). \tag{5.7}$$

The learning rule for β for fixed $\mathcal{P}$ and $\mathcal{W}$ becomes:

$$\beta|_{new} = \beta|_{old} - \epsilon E(\mathcal{P}, \mathcal{W}), \tag{5.8}$$

where ϵ is the learning rate. The key idea in (5.6) is to optimize on a family of smoothed cost functions, the so-called free energy, which is parameterized by a parameter β. This parameter β effectively determines the amount of smoothing that is done on the original cost function. For low values of β the original cost function is smoothed to such a degree that only one global minimum remains, which is easily determined by gradient descent or in our case by an EM algorithm. For higher values of β more and more of the structure of the original cost function is reflected in the free energy. For $\beta \to \infty$, finally, the free energy assumes the exact form of the original cost function During the optimization process, it is observed, that the model vectors remain at the center of mass of the data vectors up to a certain value β^*. At that point the representation undergoes a transition and the model vectors split up in data space [60].

CHAPTER 6

Multilayer Nonnegative Matrix Factorization is a Deep Architecture

6.1 Overview

A remarkable variant of NMF that improves NMF performances is the Multilayer NMF, introduced by Cichoki and Zdunek in [51]. The main idea in Multilayer NMF is that, after obtaining the prototypes and the partition matrices from an initial data matrix, it decomposes again the partition matrix via standard NMF to obtain an additional prototypes matrix and a new partition matrix. This process can be repeated as many times as the defined number of wanted layers. The output is a list of ordered prototypes matrices and one partition matrix. This family of NMF is now used in many data mining applications and thus remains a topic of ongoing interest. Despite the qualities of this approach, however, it is not clear if the obtained structure includes a proper hierarchy, in the sense that the different layers of prototypes matrices are not interchangeable.

Our objective in this chapter is to show the hierarchy between different layers in Multilayer NMF by using a mathematical approach. If we look in the literature, we find that most of the works done in that direction doesn't formalize enough the existing link, they just make some experiments to either learn the link or to assimilate the multilayer NMF process to an existing hierarchical structure and then conclude that it is self-hierarchical. However, such approach does not really ensure the hierarchy, as it lacks a mathematical formalization of the dependence between layers. In this work, the dependence is explicitly established, showing that Multilayer NMF is properly a deep process.

Our approach consists of expressing each prototypes matrix F_{i+1} corresponding to the layer $i+1$ in terms of the prototypes matrix F_i computed in the preceding step. Explicitly we infer a nonlinear function f that ensures the link between different prototypes matrices, i.e $F_{i+1} = f(F_i)$. The dependence being established, we then quantify this dependence between different layers and show that they present a hierarchical structure. We show this by proving that the order between characteristics is more important in the first layers than in the last layers. We split our resulting characteristics into three sets and compare the reconstruction error in the three sub-factorizations. We prove that an explicit order exists in the construction process of the feature matrices.

The remaining of the chapter is organized as follows. In section 2 we present some related works that treat the deepness of multilayer NMF in their point of view. Section 3 is dedicated to the formal definition of NMF and its derivatives. In section 4 we present our mathematical results concerning the existence of a mathematical function linking the layers. We drop up our experimental analysis in section 5. Finally, in section 6, we conclude and give some perspectives. The current chapter makes subject of publication [73].

6.2 Related works

In [104] and its extended version [105], Hyun Ah Song and Soo Young proposed a hierarchical feature extraction method to extract features on several stages by applying Multilayer NMF to a document data set. This yields to a concept learning hierarchy after combining relationship of features. They also extended non-smooth NMF into several layers for hierarchical learning. Experimentally this process showed a high performed reconstruction and classification.

In [108], the authors proposed a deep semi NMF which is a deep framework factorization that optimizes semi NMF factors with the aim of improvement of clustering performance. This approach is based on the learning of a hierarchical structure of features. The learning process is spread on different layers according to different attributes of data whose are automatically learned. The introduction of nonlinear functions between layers allowed extraction of features for each of the latent features.

In [55], the authors viewed Multilayer NMF as a deep architecture containing all elements of a deep neural network (pooling layer, activation function, backpropagation). According to the authors, this structure is able to produce a hierarchical classification of most type of data (audio, text, images). Following their approach, the hierarchy is ensured via the equivalence established between a standard deep neural network and what they named "Deep NMF".

Another approach, established by Le Roux, Hershey and Weninger in [64], investigates a model-based neural network architecture-like to unfolding NMF iterations and untie its basis parameters, they used back propagation for update rules and conclude that they infer a "Deep NMF model" just by stacking this iterations of unfolding NMF and consider them as a sequence

of layers in a deep neural network-like architecture.

In our point of view, the preceding works are not convincing enough to proof the actual "deepness" and hierarchy of the layered structure. It is unreasonable to justify the existence of a hierarchy by the existence of stacked layers. Rather, a deep hierarchical structure must be justified by the quality of the relationship linking those layers. To be more explicit, the term 'deep' was first associated to clustering and classification processes, Deep Neural Networks (DNN) being the most popular approach. If we dive in the way hidden layers are constructed, for example for the first hidden layer, we find out that each neuron is connected to a local receptive field built from the input data by learning its weights and then the receptive fields are moved up iteratively until they cover all the input data. This is what ensures the existence of deepness and hierarchy, since the inferior layers represents simple aspects of the original data (for example, borders in the case of images) while superior layers represent more complex aspects of data (e.g. patterns).

Our first motivation behind this work is to really prove hierarchy and deep structure of Multilayer NMF. The innovative strategy we follow is to start by establishing a function that ensures the formal link between different prototypes' matrices. We accomplish this using Moore Penrose inverse since we don't impose any regularity constraint on our matrices. After that, we construct an indexed ordered sequence of resulting prototypes' matrices, $A = \{F_1, F_2, ..., F_L\}$ and extract from it three recovering subsequences: $A_1 = \{F_1, F_2, ..., F_{L/3}\}$, $A_2 = \{F_{\frac{L}{3}+1}, ..., F_{2L/3}\}$ and $A_3 = \{F_{\frac{2L}{3}+1}, ..., F_L\}$. Say that A is a hierarchical structure turns to prove that the elements of A_1 play a greater role in representing the original data, because they come first in the process of factorization. This is done by comparing the error between different subsequences $A_i,\ i \in \{1, 2, 3\}$ and data matrix X.

6.3 Multilayer Nonnegative Matrix Factorization

Besides the useful visualizations provided by the approach, NMF is mostly used to reduce the dimensionality of the data. The main principle is to factorize the data matrix $X \in \mathbb{R}^{d\times N}$ into two matrices: $X = FG + E$, where $F \in \mathbb{R}_+^{d\times k}$ is the matrix of prototypes, $G \in \mathbb{R}_+^{k\times N}$

is the partition matrix (Cluster indicators) and E is the error of approximation. The order of decomposition k is chosen in such way that $(N + d)k << Nd$, which makes NMF a robust technique of dimensionality reduction.

In the next section, we present our mathematical result that explicit the relation between learned factors, we find the function f that links different F_i, says $F_{i+1} = f(F_i)$.

6.4 Existence of hierarchy between layers

The basis notions of NMF and its related derivatives being given, we infer the mathematical function that expresses each layer matrix in terms of the preceding layer matrix, this ensures that the matrices are indeed related mathematically. In order to find the dependence function f let us write for instance the two first lines of Multilayer NMF:

$$X \cong F_1 G_1 \tag{6.1}$$

$$G_1 \cong F_2 G_2 \tag{6.2}$$

By injecting the second equation into the first one we obtain

$$X \cong F_1(F_2 G_2)$$

Thus, to find the function f we should extract F_2 in terms of F_1. Since we don't have any information about the classical invertibility of the matrices (not necessarily square and even then non-singular) we will use the generalized inverse (or pseudoinverse) especially the Moore-Penrose inverse to get $F_2 = f(F_1)$.
Let us give a brief definition of what Moor-Penrose inverse is.

6.4.1 Moore-Penrose inverse

The advantage of the generalized inverse is that it covers a wide spectrum of matrices, even singular or rectangular and it fits with the classical inverse if the matrix is non-singular. An

important illustration of the generalized inverse that we exploit in our work is that it allows the solvability of linear systems of the form $Ax = b$ for any matrix A and any vectors x, b.

Definition. In 1955 Penrose showed that for any arbitrary matrix A there is a unique matrix X (called generalized inverse of A and often noted $A^\dagger$) satisfying the four equations

$$AXA = A$$

$$XAX = X$$

$$(AX)^* = AX$$

$$(XA)^* = XA$$

Because this unique generalized inverse was defined separately by Penrose (in a different manner) this generalized inverse is called Moore-Penrose inverse. An important task when manipulating generalized inverses is to find explicitly the expression of $A^\dagger$. A particular case where we can find an explicit formulation of $A^\dagger$ is when A of rank r has a full rank factorization $A = FG$, we then have:

1. $A^\dagger = G^*(F^*AG^*)^{-1}F^*$
2. $F^\dagger = (F^TF)^{-1}F^T$

From this last property, if $A \in \mathbb{R}^{m\times n}$ of full row rank m, $A = I_mA$ is a full rank factorization of A. Then

$$A^\dagger = A^T(AA^T)^{-1}$$

In our case and with the purpose to be more general and englobe all types of matrices we suppose that the matrix in question does not satisfy any type of regularity from the ones mentioned above. Now let's exploit this to find the features matrices dependance function.

6.4.2 Features matrices dependence function

Coming back to equation (6.1) and put $X = B$, $F_1 = D$ and $G_1 = C$, equation (6.1) is then equivalent to $B = DC$ and resolving the problem of finding f such that $F_2 = f(F_1)$ comes to find a way to take D to the left side.

We start by considering the system $Ax = b$, we express x in terms of A and b and then we will be able to find a general formula that express C in terms of B and D by applying the formula obtained from the linear system to columns of B. Note that a solution of (6.1) should be understood in the last squares sense. Following [**?**], we have:

$$x = A^{\dagger}b + (I - A^{\dagger}A)y \tag{6.3}$$

y an arbitrary vector, where the particular solution $A^{\dagger}b$ is the least squares solution LSS that has Minimal Euclidean Norm (abbreviated MNLSS).

Our equation is about the system:

$$B = DC \tag{6.4}$$

where X, B are matrices with same number of columns. For each column of B we apply (6.3) and conclude that the general solution of (6.4) is:

$$C = D^{\dagger}B + (I - D^{\dagger}D)Y \tag{6.5}$$

or

$$D = BC^{\dagger} + (I - CC^{\dagger})Y \tag{6.6}$$

Y an arbitrary matrix. Here again the particular solutions $D^{\dagger}B$ and $BC^{\dagger}$ are MNLSS for the appropriate matrix norm.

The equations of a multilayer NMF obtained at steps $i - 1$ and i respectively are:

$$G_{i-1} \cong F_i G_i \tag{6.7}$$

$$G_i \cong F_{i+1} G_{i+1} \tag{6.8}$$

We start by injecting equation (6.8) in equation (6.7):

$$G_{i-1} \cong F_i F_{i+1} G_{i+1}$$

we then apply (6.6) to obtain:

$$F_i F_{i+1} \cong G_{i-1} G_{i+1}^{\dagger} + (I - G_{i+1} G_{i+1}^{\dagger}) Y_1$$

Y_1 an arbitrary matrix.
By reapplying (6.5) and injecting equation (6.8) on this last equation we finally obtain:

$$F_{i+1} \cong F_i^{\dagger} \left(F_i G_i G_{i+1}^{\dagger} + (I - G_{i+1} G_{i+1}^{\dagger}) Y_1 \right) + (I - F_i^{\dagger} F_i) Y_2$$

Y_2 an arbitrary matrix.

This is to conclude that in fact each learned factor can be expressed in terms of the preceding one, that it depends strongly on the preceding learned factor through all layers of Multilayer NMF, in the next section we check experimentally that the order of F_i matters in the reconstruction layers of our input data.
In the next section we analyze experimentally the proved deepness between multilayer NMF layers, we do this by perturbing the order of features matrices and see if this affects the reconstruction.

6.5 Experimental Validation of the Hierarchical Dependence

Hierarchical ordered data structures and hierarchical feature learning show which extracted features develop at each step of a stacked process at each step, Multilayer Nonnegative Matrix Factorization (NMF) in our case, and how the aggregation of lower features leads to the construction of higher layer features.

To measure and gauge an eventual existing hierarchy between different layers of Multilayer NMF, and following our intuition that says the more representative and paramount layers (in the sense of similarity with entry data matrix' characteristics) are the first resulting matrices, i.e. that an order applies along and across the whole layers, we partitioned our resulting L layers on 3 sets, permutate alternatively matrices of each set and measure the error between the product of

the permutated matrices and the initial data matrix, we found that the error in the first set is more important than in the second set which is superior than the error in the last set, that is the first resulting matrices are closer to the data.

The first set consists of the first third of the F_is, say $A_1 = \{F_1, F_2, ..., F_{L/3}\}$ the second set is constituted from the second third $A_2 = \{F_{\frac{L}{3}+1}, ..., F_{2L/3}\}$ and the third of the last third of the resulting matrices $A_3 = \{F_{\frac{2L}{3}+1}, ..., F_L\}$.

To show that Multilayer NMF leads to a hierarchical structure not because of the existence of multiple layers, but due to an order established implicitly between them, a hefty task was to find a way to explicit this hierarchy and prove effectively that this order follows a dynamic through the L layers. To this end, we acted on the order of the layers of each of the previously defined set. Our intuition is the following: unlike Deep Neural Networks, Multilayer NMF generates features in a decreasing order in the terms of representability of the original data. So, to visualize this, we disrupt the order of G_is by permutating them. This implies an increasing shift in the difference between the original data matrix and the product of the permutated partitioned features matrices (this difference is called reconstruction error), when the percentage of permuted layers increases. The expression of this error is given for each of the three sets, respectively by: $\|X - permut_i(F_1, F_2, ..., F_{L/3})\|^2$, $\|X - permut_i(F_{\frac{L}{3}+1}, .., F_{2L/3})\|^2$, and $\|X - permut_i(F_{\frac{2L}{3}+1}, .., F_L)\|^2$, with $i \in \{5\%, 10\%, 15\%, 20\%\}$.

This experimental protocol has been applied on 100 set of 1000 artificial data generated from three Gaussian distribution in ten dimensions. For each dataset, we trained a 100-layer NMF, before applying the permutation protocol. Figure 6.1 shows the difference between X and a random permutation of respectively 5%, 10%, 15% and 20% of the $L/3$ first layers in A_1, A_2 and A_3, respectively. From the figure, we can see that perturbations in the first third of prototype matrices give rise to an error of 10, this drastically decreases to 0.20 when we act on permutating in the second third, and ends by being an error of 0.08 when we perturb the order of the last third of our prototype' matrices.

To visualize the relative importance of layers order in the hierarchy, we tested the impact of perturbations on a set of images (Figure 6.2). Each image is described as a matrix representing

Figure 6.1: Reconstruction error after the permutations in the first, second and last third of prototypes' matrices.

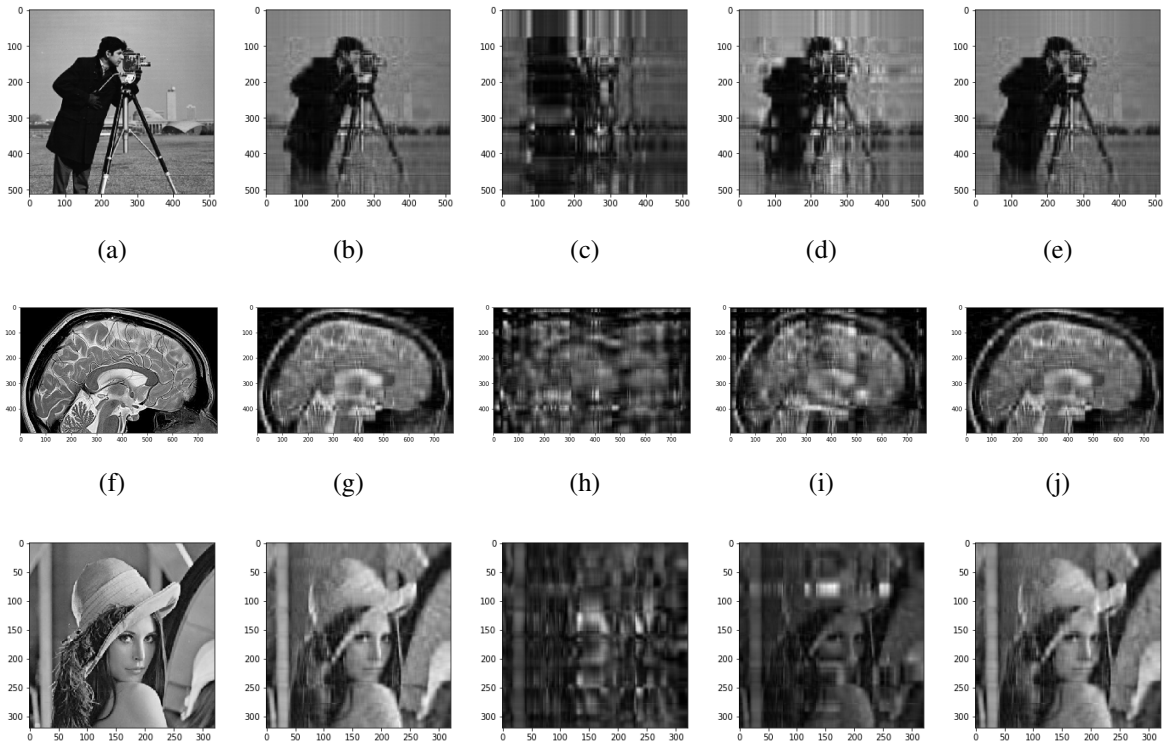
the gray levels. The matrix were decomposed with a Multilayer NMF with 50 layers, then reconstructed via the product of the features matrices. A random permutation of a pair of features matrix in the decomposition leads to an increase of the reconstruction error, as seen in the figure. In accordance to the results presented above, perturbations in the first layers of the decomposition have a higher effect than perturbations in the intermediate layers and a much higher effect than perturbations in the last layers.

The main conclusions we can infer from the obtained results are the following:

- The more we perturb the matrices' order the more the error increases.
- As the number of permuted layers increases, the effect of the perturbations is more and more important.
- The perturbations on the first third of matrices have an extra effect compared to perturbations on the two other thirds.

To summarize, the first ordered feature matrices resulting from the Multilayer NMF are the closest to the data characteristics. Since under perturbations of the order of the stacked matrices, those characteristics become less related to our original data, we can see in the first $L/3$ layers

that the error is huge compared to the others partitions. This is explained by the existence of an implicit hierarchy imbedded in the Multilayer NMF, allowing a deep learning framework for this type of approaches.



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

(k) (l) (m) (n) (o)

Figure 6.2: Visualization of the reconstruction error without and with perturbation of the layers order. The first column ((a), (f) and (k)) shows the original images, the second ((b), (g) , (l)) is the image reconstructed from a Multi-layer NMF. The last three columns present the reconstructed image obtained after perturbations on the first third of the layers ((c), (h), (m)), the second third ((d), (i), (n)) and the last ((e), (j), (o)).

CHAPTER 7

Stacked Deep Self Organizing Maps

Overview

In this part we explore an original strategy for building deep networks, based on stacking layers of Self Organizing Maps (SOM) with finite weights. We aim to show that our approach, with enough hidden variables, is capable to represents any "soft" distribution over the visible variables, where "soft" means that the distribution does not contain any probabilities of 1 or 0. The algorithm compute the model one layer at a time. The first layer receives the input observations and compute a probability of membership for each observation and each neuron. These probabilities become the input of the second layer, which compute a new set of probabilities, and so on. The number of neurons decrease in each layer after the first. The proposed algorithm is experimentally tested on artificial and real data-sets. The effect of the added hidden layers for the representation of data structure is experimentally demonstrated. The current chapter makes subject of publication [72].

7.1 Architecture

We consider a hierarchical system constructed by stacking a finite number of layers $(\mathcal{L})_{l=1}^{L}$, where L denotes the total number of levels as in [100, 101]. All the layers in our architecture have the same block structure $\mathcal{L} = (Input, Transform, Output)$.

The $Input$ block could be the observed data X for the first layer in the architecture or could be the output block of the previous layer in the architecture. The $Transform$ block part is basically a Maximum Likelihood (ML) learning algorithm. The basic principle of the ML method is that the event with the greatest probability is most likely to occur, so parameters should be chosen that maximize the probability of the observed sample data. The ML estimation possesses several attractive properties including: consistency, asymptotic normality, and efficiency when the sample size tends to infinity. The $Output$ block contains all the different vision on the $Input$ obtained through the $Transform$ block, each component in the $Output$ block represents a particular vision of the $Input$ data and is a new over-complete representation with redundant information of the $Input$. This idea was already used in [66, 67] to design learning algorithm for restricted Boltzmann machines (RBM) that contain many layers of hidden variables. As in the

case of RBMs data-dependent distributions are estimated using a variational approximation that tends to focus on a single mode. Nevertheless, several differences between the two approaches should be mentioned. In our case the SOM preserves the topology through the layers hierarchy and also we are able to process directly the continuous data.

7.1.1 Algorithm

The algorithm compute the model one layer at a time. The first layer receives the input observations and compute a probability of membership for each observation and each neuron. These probabilities (the output of the first level) become the input of the second layer, which compute a new set of probabilities, and so on. The number of neurons decrease in each layer after the first.

Starting from the second layer, input and output values are probabilities. Therefore, whereas in the first layer we use the euclidian distance to compute the similarities $d(x, w)$ between observations and prototypes, in the following layers the Kullback-Leibler divergence is more adapted, as both inputs and prototypes are a set of probabilities.

The full process is described in algorithm 1.

Algorithm 1 :Topological Stacking SOM

input : The observations X, the number of layers L, the number of neurons N_1 in the first layer

output: A membership probability $\mathcal{P}_L$ for each observation and each neuron in the last layer

for $l \leftarrow 1$ **to** L **do**

// Initialization:

if *l = 1* **then**

Select the input X

else

Select the input matrix: $X = \mathcal{P}_{l-1}$ Update the number of neurons: $N_l = N_{l-1} - 4\sqrt{N_{l-1}}$, the value is rounded and kept $N_l > 4$

end

Initialize the topology of the l^{th} map, the prototypes $\mathcal{W}^l$ and the value of β_l according to preceding section

// Main loop

while *the model is not stable* **do**

// Expectation phase

Compute $\mathcal{P}_l = \{p_c^n(\mathcal{W}^l); c = 1, ..., N_l, n = 1, ..., N\}$ for each observation n and each neuron c:

$$p_c^n(\mathcal{W}^l) = \frac{\sum_{r \in G} \mathcal{K}_t(c, r) e^{-\beta_l \cdot d(x^n, w_r^l)}}{\sum_{r', r \in G} \mathcal{K}_t(r', r) e^{-\beta_l \cdot d(x^n, w_r^l)}}.$$

with $d(x^n, w_r)$ the square of the euclidean distance if $l = 1$ and the Kullback-Leibler divergence otherwise.

// Minimization phase

Update $\mathcal{W}_l = \{w_c(\mathcal{P}^l); c = 1, ..., N_l\}$ for each neuron c:

$$w_c(\mathcal{P}^l) = \frac{\sum_{n=1}^{N_l} \sum_{r \in G} p_r^n \mathcal{K}_t(c, r) x^n}{\sum_{n=1}^{N_l} \sum_{r \in G} p_c^n \mathcal{K}_t(c, r)}$$

end

end

7.2 Experimental results of Stacked Self Organizing Maps

As in any machine learning problem, choosing the right single-number evaluation metric is a crucial step as it allows us to dive into experiments with clear direction for progress and hence quickly decide what is working best. Therefore, and because we are in a probabilistic concept, we assumed that inputs with similar probability distributions belong to the same cluster and consequently probability distributions of other inputs should be different and our metric should tell us roughly how true it is.
Taking that into consideration, we started with a metric based on the well-known Jensen Shannon divergence, representing the ratio of the mean divergence between each object i and the object in the same cluster C, by the mean divergence between i and the objects in other clusters. To test the quality of the proposed approach, we compared the model proposed by a mono-layered topological clustering the models obtained with our multi-layer approach. The main idea to test is that the multi-layer approach provides a better representation of the data structure into clusters. The representation of objects from the same cluster should be similar, whereas the representation of objects from different clusters should be different.

As the output of our model is a matrix of probability associated to each input, the measure of similarity between objects' representation must be adapted. We chose here to use the Jensen-Shannon divergence which is a measure based on the Kullback-Leibler relative entropy. The Kullback-Leibler measure between two distributions of probability P and Q is defined as:

$$D_{KL}(P||Q) = \sum_{w} P(w) \frac{P(w)}{Q(w)}. \tag{7.1}$$

This measure is not symmetric, therefore the Jensen-Shannon divergence is often preferred :

$$D_{JS}(P||Q) = \frac{1}{2} D_{KL}(P||M) + \frac{1}{2} D_{KL}(Q||M) \tag{7.2}$$

with M the average of the two distributions:

$$M = \frac{1}{2} P + Q. \tag{7.3}$$

The quality index chosen for the experiment is based on the Jensen-Shannon divergence. It is the ratio of the mean divergence between each object i and the objects in the same cluster C, by the mean divergence between i and the objects in other clusters :

$$I_{JS} = \sum_{i} \frac{\sum_{j \in C, j \neq i} D_{JS}(\mathcal{P}_L(i)||\mathcal{P}_L(j))}{\sum_{k \notin C} D_{JS}(\mathcal{P}_L(i)||\mathcal{P}_L(k))}. \tag{7.4}$$

The smaller the value of this index, the better the representation of the data structure.
Executing this metric was computationally expensive, so we decided that running time should be taken into consideration and we need to find a metric that trades off the performance of our model and the similarity concept cited above.
So we searched for lesser computational formulas, and we came across the generalized Jensen Shannon divergence that uses the Shannon entropy [92] and the derived version that uses the centroid of all probability distributions [92]. executing these formulas made us realize that having results between -1 and 1 would be easier to interpret.
Here, the silhouette index [99] seemed to be a good choice as it fulfilled this need and considered to have enough credibility to judge the effectiveness of a model like ours: The highest the value of this index, the better the representation of the data structure. More merely, a close value to 1 means that inputs lie well within their cluster and a close value to -1 means wrong assignment.
And because we are in a probabilistic concept, we decided to slightly adjust the silhouette score formula: computing the Euclidean distance in the first layer and the testing distance measures which are used to quantify the similarity between two probability distributions in the rest of the layers.

We tested the algorithm on a set of artificial and real data-sets. The artificial data-sets "Art 1", "Art 10" and "Art 50" are generated from a Gaussian Mixture distribution with 1000 vectors in 2, 10 and 50 dimensions distributed in 6 clusters. The real data-set are from the UCI machine Learning repository [91]. We chose data-sets with different number of objects and different dimensions. The data-sets are given in Table 7.1.

We chose in these experiment a map of 10×10 neurons for the first layer, with the number of rows and column decreased by one in each subsequent layer (10×10 for the first, then 9×9 for the second and so on), with a minimal size of 5×5. The models have been trained with 5 layers.

Table 7.1: Data-sets description

Data-sets	Samples total	Dimensionality
Art 1	1000	2
Art 10	1000	10
Art 50	1000	50
Iris	150	4
Diabete	442	10
Wine	178	13
Boston	506	13
Digits	1797	64
Cancer	569	30
Housing	20640	8

Results

The maps architecture is without a doubt a major parameter as it has a tremendous influence on the final output.Let's think of that as any deep learning model: having an insufficient number

Table 7.2: Quality index ($I_{JS} \times 10^2$) for each data-set with respect to the number of layers.

Data-sets	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
Art 1	0.82	0.23	0.08	0.05	0.04
Art 10	0.28	0.05	4×10^{-3}	2×10^{-3}	1×10^{-3}
Art 50	0.11	0.02	3×10^{-3}	1×10^{-3}	7×10^{-4}
Iris	3.04	0.66	0.32	0.06	0.06
Diabete	3.19	1.69	1.04	0.49	0.36
Wine	2.40	0.58	0.58	0.63	0.59
Boston	0.20	0.21	0.21	0.22	0.21
Digits	1.94	0.76	0.58	0.46	0.37
Cancer	0.32	0.25	0.22	0.21	0.23
Housing	0.12	0.03	0.02	4×10^{-3}	3×10^{-3}

of neurons(or layers) will lead to what is called under-fitting meaning that your model will have poor performance. Whereas, having too many neurons(or layers) will make you over-fit your data. The comparison of quality of the objects' representation for different number of layers are presented in Table 7.2 and Figure 7.1. We tested models with different number of layers, from one to five, and computed the I_{JS} index values for each model and each data-set.

The effect of the multi-layered approach is clear. The values of I_{JS} decrease when the number of layers increase, meaning that the representation of the data structure improves with the number of layers: the output profiles of objects from different clusters are more clearly differentiated, whereas the representation of objects from the same cluster are more similar. The multi-layered model is therefore more efficient to produce a concise description of the similarities between objects. As visualized in Figure 7.1, the effect is usually stronger when we increase the number of layers of models with a low number of layers. The strongest effect is the increase of one layer to two. Further increase have different effect depending on the data set, for example data-set such as "Art 1", "Iris" or "Digits" benefit from an increase of the number of layer up to five layers. However, for some data-set we observe no further improvement after a few number of layers (e.g. "Wine" or "Diabetes"). This seems to indicate the existence of an optimal number of layers, above which the added computation time is not compensated with an improvement of the output quality. This optimum is data-set dependent, however it doesn't seem to be related to either the number of objects or the dimensionality of the data-set.

Another interesting perspective is that by increasing the number of self-organizing layers, we notice a convergence after a certain number of layers: The model stabilizes. It either cycles between a sequence of values or remains fixed.

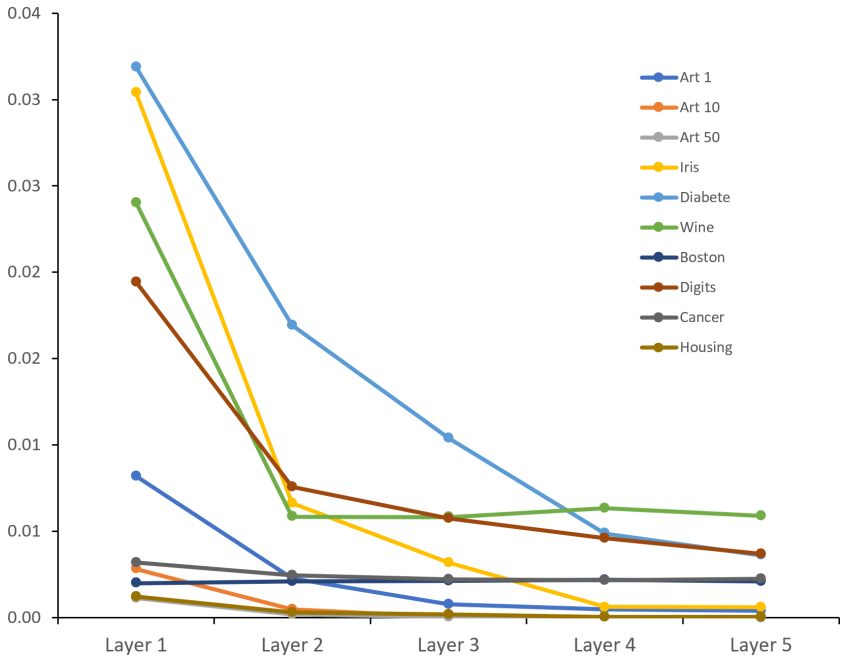
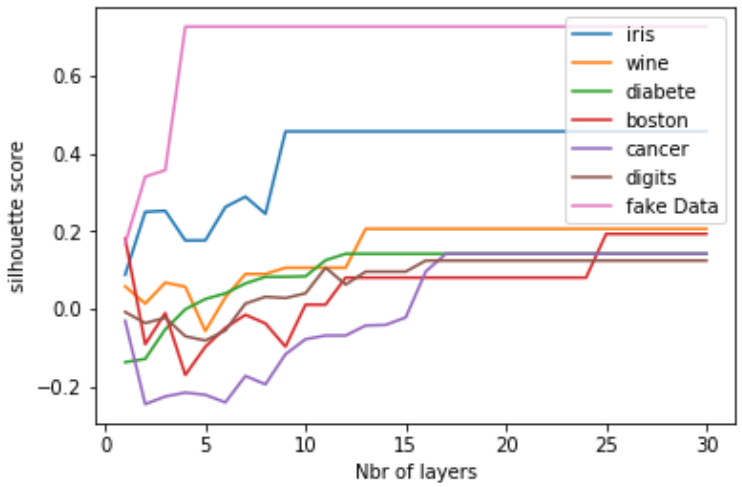


Figure 7.1: Visualization of I_{JS} for each data-set with respect to the number of layers.



We note that in the very first hidden layers, the model tends to behave strangely. Nevertheless, those layers are necessary for the model to learn data features and to output better accurate mappings then the first layer and therefore better representation of the topological properties of input data.

We also behold that when the dataset size is small, the model tends to converge more quickly (layer 9 in iris and 13 in wine) than other dataset (layer 16 in digits).

biblo2

[34] In Ian H. Witten, Eibe Frank, and Mark A. Hall, editors, *Data Mining: Practical Machine Learning Tools and Techniques (Third Edition)*, The Morgan Kaufmann Series in Data Management Systems, pages 587 – 605. Morgan Kaufmann, Boston, third edition edition, 2011.

[35] Deng ao Li, Jing ang Lv, Ju-Min Zhao, and Jin Zhang. An effective way of j wave separation based on multilayer nmf. In *Comp. Math. Methods in Medicine*, 2014.

[36] Pierre Baldi and Søren Brunak. *Bioinformatics: The Machine Learning Approach*. 01 2001.

[37] Nicholas Ball and Robert Brunner. Data mining and machine learning in astronomy. *International Journal of Modern Physics D*, 19, 06 2009.

[38] H. B. Barlow. Unsupervised learning. chapter Unsupervised Learning, pages 1–17. Bradford Company, Scituate, MA, USA, 1999.

[39] Y. Bengio, Pascal Lamblin, D. Popovici, Hugo Larochelle, and U. Montreal. Greedy layer-wise training of deep networks. volume 19, 01 2007.

[40] Yoshua Bengio and Yann Lecun. *Scaling learning algorithms towards AI*. MIT Press, 2007.

[41] Michael W. Berry and Murray Browne. Email surveillance using non-negative matrix factorization. *Comput. Math. Organ. Theory*, 11(3):249–264, October 2005.

[42] Harsh Bhasin and Surbhi Bhatia. Application of genetic algorithms in machine learning. 05 2012.

[43] Indranil Bose and Radha K. Mahapatra. Business data mining - a machine learning perspective. *Information Management*, 39(3):211 – 225, 2001.

[44] . Boutsidisa and E. Gallopoulosb. Svd based initialization : A head start for nonnegative-matrix factorization. 2007.

[45] Jean-Philippe Brunet, Pablo Tamayo, Todd Golub, and Jill Mesirov. Metagenes and molecular pattern discovery using matrix factorization. *Proceedings of the National Academy of Sciences of the United States of America*, 101:4164–9, 04 2004.

[46] Rich Caruana and Alexandru Niculescu-Mizil. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, pages 161–168, New York, NY, USA, 2006. ACM.

[47] Vikas Chaudhary, R.S. Bhatia, and Anil K. Ahlawat. A novel self-organizing map (som) learning algorithm with nearest and farthest neurons. *Alexandria Engineering Journal*, 53(4):827–831, 2014.

[48] Lei Chen, Shengbo Chen, and Xulin Guo. Multilayer nmf for blind unmixing of hyperspectral imagery with additional constraints. *Photogrammetric Engineering Remote Sensing*, 83:307–316, 04 2017.

[49] Lele Cheng, Jinjun Wang, Yihong Gong, and Qiqi Hou. Robust deep auto-encoder for occluded face recognition. In *Proceedings of the 23rd ACM International Conference on Multimedia*, MM '15, pages 1099–1102, New York, NY, USA, 2015. ACM.

[50] Sung-Bae Cho and Hong-Hee Won. Machine learning in dna microarray analysis for cancer classification. In *APBC*, 2003.

[51] Andrzej Cichocki and Rafal Zdunek. Zdunek, r.: Multilayer nonnegative matrix factorization. electronics letters 42, 947-948. *Electronics Letters*, 42:947– 948, 02 2006.

[52] Andrzej Cichocki, Rafal Zdunek, Anh Huy Phan, and S. Amari. Nonnegative matrix and tensor factorizations - applications to exploratory multi-way data analysis and blind source separation. 2009.

[53] Slim Essid and Cédric Févotte. Smooth nonnegative matrix factorization for unsupervised audiovisual document structuring. *IEEE Transactions on Multimedia*, 15:415–425, 2013.

[54] Usama M. Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, and Ramasamy Uthurusamy, editors. *Advances in Knowledge Discovery and Data Mining*. American Association for Artificial Intelligence, Menlo Park, CA, USA, 1996.

[55] Jennifer Flenner and Blake Hunter. A deep non-negative matrix factorization neural network. 2017.

[56] Yoav Freund and David Haussler. Unsupervised learning of distributions on binary vectors using two layer networks. In *Proceedings of the 4th International Conference on Neural Information Processing Systems*, NIPS'91, pages 912–919, San Francisco, CA, USA, 1991. Morgan Kaufmann Publishers Inc.

[57] Yuan Gao and George Church. Improving molecular cancer class discovery through sparse non-negative matrix factorization. *Bioinformatics*, 21(21):3970–3975, November 2005.

[58] Zoubin Ghahramani and Michael I. Jordan. Supervised learning from incomplete data via an em approach. In *Proceedings of the 6th International Conference on Neural Information Processing Systems*, NIPS'93, pages 120–127, San Francisco, CA, USA, 1993. Morgan Kaufmann Publishers Inc.

[59] Jeremy Ginsberg, Matthew Mohebbi, Rajan Patel, Lynnette Brammer, Mark Smolinski, and Larry Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457:1012–4, 12 2008.

[60] Thore Graepel, Matthias Burger, and Klaus Obermayer. Self-organizing maps: Generalizations and new optimization techniques. *Neurocomputing*, 21(1–3):173–190, 1998.

[61] Jiawei Han, Micheline Kamber, and Jian Pei. 1 - introduction. In Jiawei Han, Micheline Kamber, and Jian Pei, editors, *Data Mining (Third Edition)*, The Morgan Kaufmann Series in Data Management Systems, pages 1 – 38. Morgan Kaufmann, Boston, third edition edition, 2012.

[62] John A. Hartigan. *Clustering Algorithms*. John Wiley & Sons, Inc., New York, NY, USA, 99th edition, 1975.

[63] Trevor Hastie, Robert Tibshirani, Jerome Friedman, and James Franklin. The elements of statistical learning: Data mining, inference, and prediction. *Math. Intell.*, 27:83–85, 11 2004.

[64] John Hershey, Jonathan Le Roux, and Felix Weninger. Deep unfolding: Model-based inspiration of novel deep architectures. 09 2014.

[65] T. Heskes and B. Kappen. Error potentials for self-organization. In *Proceedings of Joint FUZZ-IEEEâ™93 and ICNNâ™93 [IJCNN93]*, volume III, pages 1219–1223, San Francisco, California, 1993.

[66] G.E Hinton, S. Osindero, and Y.W. Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18:1527–1554, 2006.

[67] G.E Hinton and R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[68] Geoffrey E. Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural Comput.*, 18(7):1527–1554, July 2006.

[69] Thomas Hofmann. Unsupervised learning by probabilistic latent semantic analysis. *Mach. Learn.*, 42(1-2):177–196, January 2001.

[70] John H. Holland. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence*. MIT Press, Cambridge, MA, USA, 1992.

[71] Patrik O. Hoyer. Non-negative matrix factorization with sparseness constraints. *J. Mach. Learn. Res.*, 5:1457–1469, December 2004.

[72] G. Cabanes I. Brahmi and Y. Bennani. Learning useful representations through a stacked self organizing maps. *Proceeding of International Joint Conference on Neural Networks*, pages 1–8, 2018.

[73] G. Cabanes Y. Bennani I. Brahmi and A. Touzani. A learning algorithm for Boltzmann machines. *Accepted in Australian Journal of Intelligent Information Processing Systems*, 2019.

[74] Anil K. Jain and Richard C. Dubes. *Algorithms for Clustering Data*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988.

[75] Leslie Pack Kaelbling, Michael L. Littman, and Andrew W. Moore. Reinforcement learning: A survey. *J. Artif. Int. Res.*, 4(1):237–285, May 1996.

[76] Jingu Kim and Haesun Park. Sparse nonnegative matrix factorization for clustering. 2008.

[77] T. Kohonen. Analysis of a simple self-organizing process. *Biol. Cyb.*, 44:135–140, 1982.

[78] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biol. Cyb.*, 43:59–69, 1982.

[79] T. Kohonen. Self-organization and associative memory. *Springer*, 1984.

[80] T. Kohonen. *Self-organization and Associative Memory: 3rd Edition*. Springer-Verlag, Berlin, Heidelberg, 1989.

[81] T. Kohonen. Speech recognition based on topology preserving neural maps. *Neural Computation*, 1989.

[82] T. Kohonen. Self-organizing maps : optimization approaches. *Artificial neural networks*, II:981–990, 1991.

[83] T. Kohonen. Self-organizing maps. *Springer*, 30, 1995.

[84] T. Kohonen, M. R. Schroeder, and T. S. Huang, editors. *Self-Organizing Maps*. Springer-Verlag, Berlin, Heidelberg, 3rd edition, 2001.

[85] T. Kohonen and P. Somervuo. Self-organizing maps of symbol strings. *Neurocomputing*, 21(1):19–30, 1998.

[86] Igor Kononenko. Machine learning for medical diagnosis: history, state of the art and perspective. *Artificial Intelligence in Medicine*, 23(1):89 – 109, 2001.

[87] Sotiris Kotsiantis, I. Zaharakis, and P. Pintelas. Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review*, 26:159–190, 11 2006.

[88] Nada Lavrac. Machine learning for data mining in medicine. In *Proceedings of the Joint European Conference on Artificial Intelligence in Medicine and Medical Decision Making*, AIMDM '99, pages 47–64, Berlin, Heidelberg, 1999. Springer-Verlag.

[89] Daniel Lee and H. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788–91, 11 1999.

[90] Hongdong Li, Yizeng Liang, and Qingsong Xu. Support vector machines and its applications in chemistry. *Chemometrics and Intelligent Laboratory Systems*, 95(2):188 – 198, 2009.

[91] M. Lichman. UCI machine learning repository, 2013.

[92] J. Lin. Divergence measures based on the shannon entropy. *IEEE Trans. Inf. Theor.*, 37(1):145–151, September 2006.

[93] Yong Liu, Xin Yao, and Tetsuya Higuchi. Evolutionary ensembles with negative correlation learning. *Evolutionary Computation, IEEE Transactions on*, 4:380 – 387, 12 2000.

[94] Siwei Lyu and Xin Wang. On algorithms for sparse multi-factor nmf. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'13, pages 602–610, USA, 2013. Curran Associates Inc.

[95] Nasser Mohammadiha, Paris Smaragdis, and Arne Leijon. Supervised and unsupervised speech enhancement using nonnegative matrix factorization. *Audio, Speech, and Language Processing, IEEE Transactions on*, 21:2140–2151, 10 2013.

[96] Francisco Pereira, Tom M. Mitchell, and Matthew M Botvinick. Machine learning classifiers and fmri: A tutorial overview. *NeuroImage*, 45:s199–s209, 2009.

[97] Marc'Aurelio Ranzato, Christopher Poultney, Sumit Chopra, and Yann LeCun. Efficient learning of sparse representations with an energy-based model. In *Proceedings of the 19th International Conference on Neural Information Processing Systems*, NIPS'06, pages 1137–1144, Cambridge, MA, USA, 2006. MIT Press.

[98] I. Redko and Y. Bennani. Sparsity analysis of learned factors in multilayer nmf. In *2015 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7, July 2015.

[99] Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53 – 65, 1987.

[100] B. Sallans. Learning factored representations for partially observable Markov decision processes. In S. A. Solla, T. K. Leen, and K-R Müller, editors, *Advances in Neural Information Processing Systems*, volume 12, pages 1050–1056. The MIT Press, Cambridge, 2000.

[101] B. Sallans and G. E. Hinton. Reinforcement learning with factored states and actions. *Journal of Machine Learning Research*, 5:1063–1088, 2004.

[102] Thomas Serre, Gabriel Kreiman, Minjoon Kouh, Charles Cadieu, Ulf Knoblich, and Tomaso Poggio. A quantitative theory of immediate visual recognition. *Progress in brain research*, 165:33–56, 02 2007.

[103] J. Sivic, B.C. Russell, A.A. Efros, A. Zisserman, and W.T. Freeman. Discovering objects and their location in images. *IEEE International Conference on Computer Vision*, pages 370–377, 01 2005.

[104] Hyun Song and Soo-Young Lee. Hierarchical data representation model - multi-layer nmf. 01 2013.

[105] Hyun Song and Soo-Young Lee. Hierarchical representation using nmf. volume 8226, pages 466–473, 11 2013.

[106] Suvrit Sra. Nonnegative matrix approximation: Algorithms and applications. 2006.

[107] Richard S. Sutton and Andrew G. Barto. *Introduction to Reinforcement Learning*. MIT Press, Cambridge, MA, USA, 1st edition, 1998.

[108] G. Trigeorgis, K. Bousmalis, S. Zafeiriou, and B. W. Schuller. A deep matrix factorization method for learning attribute representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(3):417–429, March 2017.

[109] Paul E. Utgoff and David J. Stracuzzi. Many-layered learning. *Neural Comput.*, 14(10):2497–2529, October 2002.

[110] J.J. Verbeek, N. Vlassis, and B.J.A. Krose. Self-organizing mixture models. *Neurocomputing*, 63:99–123, 2005.

[111] Tuomas Virtanen. Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria. *Audio, Speech, and Language Processing, IEEE Transactions on*, 15:1066 – 1074, 04 2007.

[112] Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8:229–256, 1992.

[113] Ronald J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine Learning*, 8:229–256, 1992.

[114] Wei Xu, Xin Liu, and Yihong Gong. Document clustering based on non-negative matrix factorization. pages 267–273, 01 2003.

General Conclusion

Through this thesis we presented in the first part the stability analysis of variational eigenvalues associated to different partial differential equations,say the $p(.)-$Laplacian, $p(.)-$Biharmonic and the eigenvalues associated to the minimization of Rayleigh quotient generated by norms, we study its robustness with respect to the integrability index $p(.)$.

A possible perspective of those results could be either the study of the convergence properties of the first eigenfunctions, bifurcation results from the first eigenvalue, or the extension of the stability to the eigenvalues of higher order. In the second part, we start by studying the dependence between features matrices obtained via Multilayer Nonnegative Matrix Factorization. We proofed the dependence function linking those matrices. We also investigated the existence of a hierarchical structure by permutating the order of the layers independently on features matrices at three different level in the layers. We showed that permutations of feature matrices in the first third of the layers induce a more important error than permutations of feature matrices of the last third of the layers. This means that there exists a real hierarchy between the layers and that it is justified to use deep learning processes on Multilayer Nonnegative Matrix Factorizations. An important theoretical study of Multilayer NMF that we envisage to accomplish in the future is to determine the optimal number of layers, that is at which level of decomposition from the basis matrices should we stop the process? It is also worth noting that this study could be generalized to tensor factorization.

The second novelty we propose is a new formalism for a multi-layered Self-Organizing Map. The proposed approach allows a more condensed description of the data structure. The final clustering can be hard or soft, thanks to a probabilistic formalism. It's complexity is linear regarding the number of objects in the data-set, allowing real-life data analysis. The experiments show the quality of the obtained model in comparison to standard topological approaches. We have shown that it is possible to learn a cascade, densely-connected, network one layer at a time. We have assumed that at each level the log-likelihood could be maximized conditionally inde-

pendent form previous level. Moreover by using the neighborhood function we have designed a topology preserving network. This is equivalent to having an undirected model which can be learned efficiently using the maximization of the log-likelihood on each layer. Our approach can also be understood as constrained variational learning because of the penalty term.