UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI MATEMATICA
PURA ED APPLICATA
DOTTORATO DI RICERCA IN MATEMATICA
XVI CICLO

Generating Functions and Finite Parameters Reductions in Field Theory

coordinatore: *Ch.mo Prof.* Bruno Chiarellotto

supervisore: *Ch.mo Prof.* Franco Cardin

dottorando: Alberto Lovison

31 ottobre 2004

Introduction

In this Ph.D. thesis we want to collect some results on different topics which share as a main feature the search of critical points of the so-called *generating functions*[1], and the allied subject of the reduction from infinite to finite parameters.

We give an overview of the main topics.

Asymptotic optics

The global vanishing trend

Generating functions are found as phase functions in oscillatory integrals, in the study of asymptotics in optics. Finding the critical points it is possible to reduce the integral to a finite sum, according to the well known *stationary phase method* (SPM). We set ourselves in the Gevrey class of functions, in order to obtain sharper estimates of the asymptotical behavior. We proved that if the amplitude $a(x)$ and the phase $\varphi(x)$ belong to the Gevrey class of functions G^s, $s > 1$, if $a(x)$ is compactly supported and $\varphi(x)$ has no critical points in $\operatorname{supp} a$, then the oscillatory integral satisfy the estimate:

$$I(k) := \int a(x)e^{ik\varphi(x)}dx = A\sqrt{k}e^{-\sigma k^{\frac{1}{s}}}\left[1 + O\left(\frac{1}{k}\right)\right], \qquad \text{as } k \to +\infty$$

(see chapter 1 and also [CL04]).

Loss of Gevrey regularity

Actually, the above result can be improved, including in the analysis Morse critical points, and even degenerate critical points.

In fact, when one has Morse critical points and one takes a G^s symbol for amplitude, the oscillatory integral gives, thorough SPM, a new Gevrey

[1]often referred to as generating families, or as phases, *e.g.* in wave phenomena.

symbol with a loss of Gevrey regularity from $s > 1$ to $2s - 1 > s$. This situation is analyzed in chapter 2 (see also a forthcoming paper [CGL04]). For sake of completeness, a proof for the Morse lemma in Gevrey classes is reported in section 2.3.

Reduction in field theory

General theory

A finite parameters reduction appears in the study of semilinear Dirichlet boundary value problems outlined in chapter 3. Providing some appropriate conditions which we will illustrate below, these equations, which are posed in fit functional spaces, *e.g.* in H_0^1, find a globally equivalent formulation in finite dimensional spaces, *i.e.* in $\mathbb{R}^m$. When these problems admit a variational principle $dJ = 0$, the reduction applies as well to the functional J, so the search for solutions is brought back to the determination of critical points of an m–variables function. A first illustration of this reduction technique in partial differential equations was performed in [Car03], as we essentially report in chapter 3). The exact finite parameters reduction stands on the fundamental works of Amann, Conley and Zehnder (see [AZ80, CZ86]).

Variational and topological techniques for existence

The most part of the results we met in literature is centered on existence questions (see for instance [Nir81, Ben95]). The most widely employed techniques are of topological type, *e.g.* degree theory and local inversion theorems, while finer variational techniques, as Morse theory or Lusternik-Schnirelmann theory, take place usually whenever some restrictions in the statement are provided.

Very sharp investigations can be done near bifurcation points, by means of the Liapunov-Schmidt reduction, see for instance [AP93]. The reduction techniques presented in this thesis, have, at least in principle, several affinities with the Liapunov-Schmidt procedure.

On the other hand, the Liapunov-Schmidt reduction can be applied only in a neighborhood of a bifurcation point, while the present techniques, though providing more restrictive assumptions, rewards by giving a global generating function, which critical points are all and exactly the solutions of the Dirichlet problem.

In chapter 4 we expose in detail a simple[2] existence result, stated in section 4.1.1, which has been carried over by means of Lusternik-Schnirelmann theory acting on the theory of Generating Functions Quadratic at Infinity (FGQI[3]). This result seems the first application of the FGQI to PDE problems. FGQI were developed, in a finite dimensional setting, mainly by Chaperon, Sikorav and Viterbo in a context of symplectic topology. See sections 4.3 and 4.4 and the original papers [Vit90, Cha84, Cha91, Sik86, Sik87, Thé99].

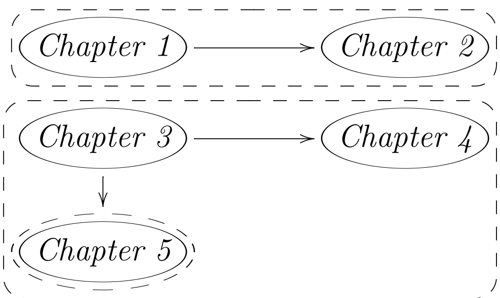
Numerical applications

Besides the above existence applications, this particular constructive finite dimensional reduction theory suggests us further more useful applications to algorithms of numerical approximation to solutions.

Following this indication we developed a very simple model in $\mathbb{R}$ employing finite elements in order to test the applicability of the theory on a physical problem (see chapter 5 and a forthcoming paper [CPL04]). Sample tests performed on this model agree with the theoretical estimates, though we haven't still evaluated the competitiveness with other more standard algorithms. Nevertheless, applying Peano-Picard and Newton-Raphson procedures, we found two nontrivial solutions for the nonlinear Dirichlet problem taken as a model.

Interdependence and redundancy

Chapters 1 and 2 are completely independent from the rest of this thesis. Chapters 4 and 5 employ the techniques exposed in detail in chapter 3. However, chapter 5 can be read alone, because in its introductive part the core results of the reduction technique of chapter 3 are repeated in short.



[2]Indeed this fact may also be reached straightforwardly by means of a standard Leray-Schauder's topological degree argument. We wish to thank a referee for this remark.

[3]Fonctions Génératrices Quadratiques à l'Infini

Contents

Chapter 1

Lack of critical phase points and exponentially faint illumination

The stationary phase method *(S.P.M.)* states that in the computation of oscillatory integrals, the contributions of non stationary points of the phase are smaller than any power n of $1/k$, for $k \to \infty$. Unfortunately, *S.P.M.* says nothing about the possible growth in the constants in the estimates with respect to the powers n. A quantitative estimate of oscillatory integrals with amplitude and phase in the Gevrey classes of functions shows that these contributions are asymptotically negligible, like $\exp(-ak^b)$, $a, b > 0$. An example in Optics is given.

1.1 Introduction

An *oscillatory integral* is an integral of the form

$$I(k) := \int_{u\in\Omega} a(u)e^{-ik\varphi(u)}du, \qquad \Omega \subseteq \mathbb{R}^d, \tag{1.1}$$

where a and φ are C^∞ real functions, called respectively amplitude and phase, and k is a (large) parameter. They are typically employed to represent solutions for linear PDE's depending on a real parameter, e.g. the Schrödinger equation or the Helmholtz equation.

A well known feature is the tight dependence on the values of a near the critical points of φ. More precisely, if in the domain of integration Ω there are no degenerate critical points for φ, the *stationary phase method*[1] applies, i.e.

$$I(k) \simeq \left(\frac{2\pi}{k}\right)^{\frac{d}{2}} \sum_{u_0 : \nabla\varphi(u_0)=0} a(u_0) \exp\{ik\varphi(u_0)\} \frac{e^{i\frac{\pi}{4}\operatorname{sgn}(\nabla^2\varphi(u_0))}}{\sqrt{\det \nabla^2\varphi(u_0)}}. \tag{1.2}$$

In particular, a very standard argument, states that the contributions to I coming from a compact subset $K \subset \Omega$ where there are no stationary points, tends to zero faster than every positive power of $\frac{1}{k}$, as $k \to \infty$.

This "superpolynomial estimate" can be written as

$$|I(k)| \leqslant A_n k^{-n}, \qquad \forall n \in \mathbb{N}, \tag{1.3}$$

(and will be revisited in Theorem 1). The coefficients A_n are comparable to the size of the n-th order derivatives of the amplitude function a. Unfortunately, for C^∞ compactly supported functions there is no *a priori* upper bound to the growth of these derivatives .

If the amplitude were analytic, it would be possible to estimate A_n by $n!$. In that case, choosing an optimal value of n depending on k and applying Stirling's formula, as shown in detail in the next section, one would find the expected exponential estimate for I:

$$|I(k)| \leqslant n! k^{-n}, \forall n \in \mathbb{N}, \qquad \Rightarrow \qquad |I(k)| \leqslant Ae^{-k}. \tag{1.4}$$

Unfortunately, analyticity is a requirement which cannot be satisfied by a non trivial compactly supported function. On the other hand, in order to have simpler integrals to manage, we choose to consider compactly supported amplitudes, as done, *e.g.*, in [AGZV88], although it would be possible to discuss the Stationary Phase Principle for non compactly supported amplitudes.

A reasonable way out can be found by turning to an intermediate class of functions placed between the spaces of the C^∞ functions and the analytic functions: the *Gevrey spaces*.

Functions in the Gevrey space $G^s(\Omega), s \geqslant 1$ satisfy the inequality:

$$|\partial^\alpha f(u)| \leqslant C^{|\alpha|+1}(\alpha!)^s \tag{1.5}$$

for every $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$, for every u in a compact subset $K \Subset \Omega$, and a suitable constant C, depending only on K.

[1] This short-wave approximation is usually referred in physics as the WKB method, and it seems that it was first worked out by F. Carlini [Car17] –we learned it in [Arn00]–, and later used by Kelvin, Stokes and many others in the 19th century.

Obviously, $G^1(\Omega)$ is the class of the analytic functions $\mathcal{A}(\Omega)$. It is also clear that there exist C^∞ functions which are not Gevrey for any s. To summarize, one has

$$\mathcal{A}(\Omega) \subsetneqq G^s(\Omega) \subsetneqq C^\infty(\Omega), \qquad s \in \mathbb{R}, s > 1. \tag{1.6}$$

Let $G^s_0(\Omega)$ denote the space of complactly supported "s-Gevrey" functions in Ω. Gevrey spaces $G^s_0(\Omega)$, with $s > 1$, are dense in $C^\infty_0(\Omega)$, in $L^p_0(\Omega)$, etc., essentially because partitions of unity can be constructed. See [Rod93] for an exhaustive treatise.

When a belongs to $G^s_0(\Omega)$, and φ to $G^s(\Omega)$, we obtain the exponential estimate:

$$|I(k)| \leqslant A\sqrt{k}\; e^{-\sigma k^{\frac{1}{s}}}, \qquad k \geqslant 1. \tag{1.7}$$

This result, allowing φ to belong to a Gevrey class of functions, improves a previous one by Todor Gramchev for analytical phases φ, see [Gra87].

1.2 A Gevrey exponentially decreasing estimate

The first theorem is standard; the second is the core of the main result, which we state as the third theorem. From now on we will assume Ω bounded.

Theorem 1. *Let $a \in C^\infty_0(\Omega)$, $\varphi \in C^\infty(\Omega)$, $\frac{\partial\varphi}{\partial u} \neq 0$, $\forall u \in \operatorname{supp} a$. Then*

$$|I(k)| \leqslant A_n k^{-n}, \qquad \forall n \in \mathbb{N}, \qquad k \geqslant 1. \tag{1.8}$$

Theorem 2. *Let $a \in G^s_0(\Omega)$, $\varphi = u_1$, then, for suitable constants A and σ,*

$$|I(k)| \leqslant A\sqrt{k}\; e^{-\sigma k^{\frac{1}{s}}}, \qquad k \geqslant 1. \tag{1.9}$$

This assertion is also true when φ is an arbitrary Gevrey function, in particular analytic, *i.e.* when $\varphi \in G^s(\Omega)$, $s \geqslant 1$.

Theorem 3. *Let $a \in G^s_0(\Omega)$ and φ be analytic or belonging to $G^s(\Omega)$, with $\frac{\partial\varphi}{\partial u} \neq 0$, $\forall u \in \operatorname{supp} a$. Then, for suitable constants A and σ,*

$$|I(k)| \leqslant A\sqrt{k}\; e^{-\sigma k^{\frac{1}{s}}}, \qquad k \geqslant 1. \tag{1.10}$$

Remark 1. In theorems (2) and (3) occurs a $\sqrt{k}$, coming from the Stirling formula, which could be absorbed by the exponential decay, by little increment of the constant σ. We keep on writing it since A and σ depend explicitly, as shown below, on the Gevrey constants C and s in (1.5), for the amplitude a and the phase φ.

Proof. Proof of Theorem 1 Following [AGZV88], we claim that there is no loss of generality if we assume $\varphi(u_1,\dots,u_d)=u_1$.

First consider an open covering $\{U_l\}$ of $\operatorname{supp} a$, such that in each U_l there is at least one non vanishing partial derivative of φ. Then, by means of a subordinate partition of unity ($\sum_{\iota\in\Upsilon}\theta_\iota=1$, $\operatorname{supp}\theta_\iota\subseteq U_l$, for some l), we can decompose the original integral as follows:

$$\int_{\operatorname{supp} a} ae^{ik\varphi}=\sum_{\iota\in\Upsilon}\int a\theta_\iota e^{ik\varphi}=\sum_{\iota\in\Upsilon}\int_{\operatorname{supp} a_\iota} a_\iota e^{ik\varphi}. \tag{1.11}$$

Since $\operatorname{supp} a$ is compact by hypothesis, Υ can be chosen finite. We now show that each of these integrals can be transformed into an integral of the desired form by means of a suitable change of variables.

By reordering variables and changing the orientation if needed, one can obtain $\frac{\partial\varphi}{\partial u_1}>0$. The map

$$u=(u_1,\dots,u_d)\overset{\eta}{\longmapsto} v=(\varphi(u),u_2,\dots,u_d), \tag{1.12}$$

is clearly a globally invertible diffeomorphism in every open convex subset of U. Then the inverse map $\xi:=\eta^{-1}$ fits our purpose. Indeed:

$$\begin{aligned} I(k)&=\int_\Omega a(u)e^{ik\varphi(u)}du=\int_{\tilde\Omega} a(\xi(v))e^{ik\varphi(\xi(v))}\left|J_\xi(v)\right|dv=\\ &=\int_{\tilde\Omega} a(\xi(v))\left|\frac{\partial\varphi}{\partial u_1}(\xi(v))\right|^{-1}e^{ikv_1}dv=\int_{\tilde\Omega}\tilde a(v)e^{ikv_1}dv. \end{aligned} \tag{1.13}$$

Now we prove the main fact. For every fixed $(d-1)$-tuple $(u_2,\dots,u_d)$, let

$$u_{1m}\ (\text{resp. } u_{1M})\ :=\inf\ (\text{resp. }\sup)\ \left\{u_1\,\middle|\,(u_1,u_2,\dots,u_d)\in\Omega\right\}. \tag{1.14}$$

Hence,

$$\begin{aligned} I(k)&=\int a(u)e^{iku_1}du=\int\int_{u_{1m}}^{u_{1M}} a(u)\frac{1}{ik}\frac{\partial}{\partial u_1}\left(e^{iku_1}\right)du_1du_2\dots du_d=\\ &=\int\left\{\left[a(u)\frac{1}{ik}e^{iku_1}\right]_{u_{1m}}^{u_{1M}}-\frac{1}{ik}\int_{u_{1m}}^{u_{1M}}\frac{\partial a(u)}{\partial u_1}e^{iku_1}du_1\right\}du_2\dots du_d=\\ &=-\frac{1}{ik}\int_\Omega\frac{\partial a(u)}{\partial u_1}e^{iku_1}du. \end{aligned}$$

Performing n times this procedure and taking absolute values gives:

$$\begin{aligned} |I(k)|&=\left(\frac{1}{k}\right)^n\left|\int_\Omega\left(\frac{\partial^n a}{\partial u_1^n}\right)e^{iku_1}du\right|\leqslant\\ &\leqslant\left(\frac{1}{k}\right)^n\int_\Omega\left|\left(\frac{\partial^n a}{\partial u_1^n}\right)\right|du=:A_nk^{-n}, \end{aligned} \tag{1.15}$$

as desired. □

Proof. Proof of Theorem 2 The hypotheses of this statement are exactly the hypotheses of the reduced case of the preceding theorem, apart from the stronger hypotheses on the amplitude. A straightforward substitution of (1.5) in the inequality of (1.15) gives

$$A_n \leqslant \text{meas}(\Omega)c^{n+1}(n!)^s, \tag{1.16}$$

with $c = C(\overline{\Omega})$, so for the integral we can write, setting $B = \text{meas}(\Omega)c$,

$$|I(k)| \leqslant Bk^{-n}c^n(n!)^s = B\left(\frac{c}{k}^{\frac{n}{s}} n!\right)^s, \qquad \forall n \in \mathbb{N} \tag{1.17}$$

which, by *Stirling's formula*:

$$n! = n^n e^{-n}\sqrt{2\pi n} e^{\frac{\theta(n)}{12n}}, \quad \text{where } 0 < \theta(n) < 1, \tag{1.18}$$

becomes

$$\begin{aligned} &= \quad B\left(\left(\tfrac{c}{k}\right)^{\frac{n}{s}} n^n e^{-n}\sqrt{2\pi n}\ e^{\frac{\theta(n)}{12n}}\right)^s, \\ &= \quad B\left(\left(\left(\tfrac{c}{k}\right)^{\frac{1}{s}} n\right)^n e^{-n}\sqrt{2\pi n}\ e^{\frac{\theta(n)}{12n}}\right)^s. \end{aligned}$$

Note that $e^{\frac{\theta(n)}{12n}}$ is surely smaller than 2 for all $n \geqslant 1$, and choose n in order to bound the quantity in the inner parentheses by 1, i.e. set $n^\star := n^\star(k) =$ (integer part of) $\left[\left(\frac{k}{c}\right)^{\frac{1}{s}}\right]$. Hence

$$|I| \quad \leqslant 2Be^{-s\left(\frac{k}{c}\right)^{\frac{1}{s}}}\sqrt{(2\pi)^s\frac{k}{c}} \quad = \quad A\sqrt{k}e^{-\sigma k^{\frac{1}{s}}}, \tag{1.19}$$

where

$$\begin{aligned} A &= 2B\sqrt{\tfrac{(2\pi)^s}{c}}, \\ \sigma &= sc^{-\frac{1}{s}}. \end{aligned}$$

□

Proof. Proof of Theorem 3 In order to reduce this statement to Theorem 2, we need to prove that the amplitude obtained performing a change of variables as in (1.13):

$$\tilde{a}(v) = a(\xi(v))\left|\frac{\partial\varphi}{\partial u_1}(\xi(v))\right|^{-1}, \tag{1.20}$$

is a Gevrey function as long as $a(u)$ and $\varphi(u)$ are. Hence we need to recall the following facts: if $\varphi \in G^s(\Omega)$,

(i) $\xi := \eta^{-1}$, where $\eta(u) = (\varphi(u), u_2, \dots, u_d)$ as in (1.12), is a G^s diffeomorphism.

(ii) $1/\left(\frac{\partial\varphi}{\partial u_1}(u)\right) \in G^s(\Omega)$.

(iii) The composition of a Gevrey function with a Gevrey diffeomorphism is still a Gevrey function of the same order s.

These are all known facts, a proof of which can be found in [LM70], where these results are proved in more general spaces of functions. Alternatively, one can find a proof which is specifically suitable for our purpose in [Gra02].
□

1.3 An application to optics

Consider the propagation of the light in the plane $\mathbb{R}^2$ due to a compact and regular emitting surface Σ (*i.e.* a closed simple curve), or, equivalently, consider a light emission for which such a Σ is a surface of constant phase. For monochromatic propagation of wave number k, the evolution is governed by the 'Helmholtz equation', which entails for the phase function to satisfy an Hamilton-Jacobi type equation: the *eikonal* equation. For isotropic homogeneous media (constant refractive index), the eikonal equation takes the form:

$$|\nabla\varphi(x)|^2 = 1. \tag{1.21}$$

A complete integral for this equation is given by

$$\varphi(x,\theta) := x \cdot n(\theta), \tag{1.22}$$

where $n : \mathbb{S}^1 \to \mathbb{R}^2$, $\theta \mapsto (\cos(\theta), \sin(\theta))$. Suppose the initial datum be σ given on Σ, which we think as parametrized by $\mathbb{S}^1 \ni \chi \mapsto x(\chi) \in \mathbb{R}^2$. It is well known that in the general case no global classical solution exists. However, the modern geometric Hamilton-Jacobi theory always admits globally defined solutions.[2] The solution of the Geometrical Cauchy Problem is a Lagrangian submanifold $\Lambda \subset T^\star\mathbb{R}^2$, globally generated by the Morse Family, (*i.e.* generating function) $\tilde\varphi$:

$$\tilde\varphi(x;\theta,\chi) := (x - x(\chi)) \cdot n(\theta) + \sigma(\chi),$$

$$\Lambda = \left\{ (x,p) : p = \frac{\partial\tilde\varphi}{\partial x},\ d_u\tilde\varphi = 0, i.e. \left\{ \begin{array}{l} \frac{\partial\tilde\varphi}{\partial\theta} = 0, \\ \frac{\partial\tilde\varphi}{\partial\chi} = 0. \end{array} \right. \right\}$$

[2]See e.g. [Arn00, Car02] and the bibliography quoted therein.

In such a case, the auxiliary variables $u = (\theta, \chi)$ belong to $\mathbb{S}^1 \times \mathbb{S}^1 = \mathbb{T}^2$. The amplitude of light observed in a point x of the plane is then given by an oscillatory integral, extended over the parameter space $\mathbb{T}^2$:

$$I(x;k) = \left(\frac{2\pi}{k}\right)^2 \int_{\mathbb{T}^2} a(x;\chi,\theta) e^{ik\tilde{\varphi}(x;\theta,\chi)} d\theta d\chi. \tag{1.23}$$

Here the amplitude function $a(x;\theta,\chi)$ satisfies the 'transport equations' because $I(x;k)$ is a solution of the Helmholtz equation.

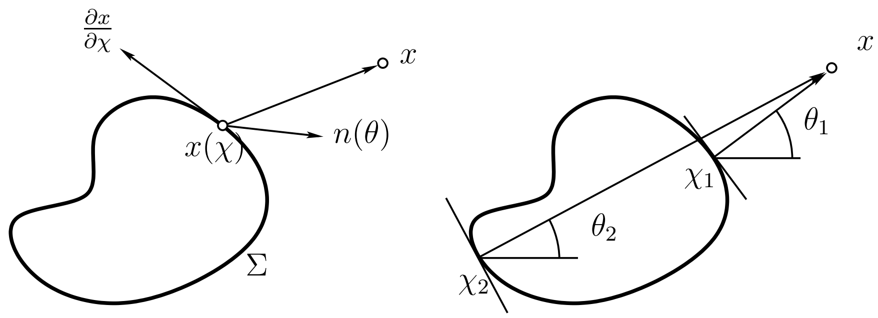


Figure 1.1: Light emission from a surface Σ

Stationary Phase states that the only relevant contributions in the asymptotic expression of $I(x;k)$ come from arbitrarily small neighbourhoods of the u-critical points of $\tilde{\varphi}$. They are given by

$$d_u\tilde{\varphi} = 0 : \begin{cases} \frac{\partial \tilde{\varphi}}{\partial \theta} = 0, \\ \frac{\partial \tilde{\varphi}}{\partial \chi} = 0. \end{cases} \Leftrightarrow \begin{cases} -n(\theta) \cdot \frac{\partial x}{\partial \chi}(\chi) = 0, & \text{(a)} \\ \frac{\partial n}{\partial \theta}(\theta) \cdot (x - x(\chi)) = 0. & \text{(b)} \end{cases} \tag{1.24}$$

From (a) one obtains that n must be $\perp$ to the emitting surface Σ, from (b) one obtains that $x - x(\chi)$ must be parallel to n. In other words, $u_0 = (\theta_0, \chi_0)$ is a critical point for $\tilde{\varphi}$ if and only if χ_0 is a critical point for the distance function between x and Σ and $n(\theta)$ is parallel to the connecting vector (see Figure 1.1).

It is self-explanatory that this is a wave translation of the Fermat Principle of geometrical optics.

In order to exhibit an application of our result, fix an observation point x in the plane, and consider a Gevrey amplitude function $a(x;\theta,\chi)$, possibly obtained by means of a Gevrey partition of unity, which vanishes in a neighbourhood V of the critical points $u_\iota = (\theta_\iota, \chi_\iota)$ of the phase. (See $V = V_1 \cup V_2$ in Figure 1.2.)

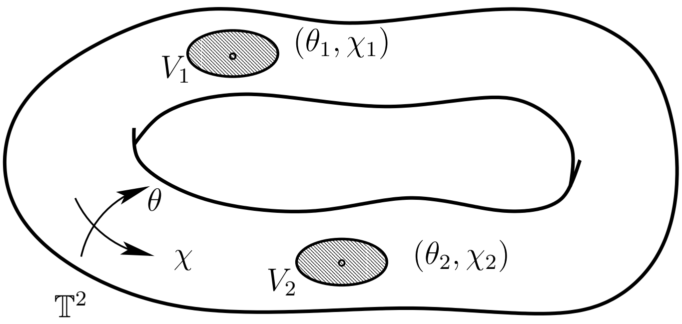


Figure 1.2: The domain of integration of the "reduced integral" of the example.

Thus, our result asserts that the oscillatory integral so obtained, is exponentially vanishing as k tends to infinity, as is the illumination at x, which is represented by this reduced integral.

Chapter 2

Loss of Gevrey regularity for asymptotic optics

In this chapter we will investigate the asymptotic behavior of oscillatory integrals from the Gevrey point of view. We will give formal asymptotic expansions and study the Gevrey character of oscillatory integrals, in comparison with the Gevrey character of their amplitudes. We will deduce a formula for the loss of Gevrey regularity when the phase functions are in the Morse class.

2.1 Oscillatory integrals and Gevrey character

Our aim is to give an asymptotic analysis from the Gevrey point of view of oscillatory integrals of the type

$$I_a^\varphi(\omega,\lambda) := \int_{\mathbb{R}^n} e^{i\lambda\varphi(x,\omega)} a(x,\omega,\lambda)\,dx \tag{2.1}$$

for $\lambda \to +\infty$, where $\omega \in \mathbb{R}^\ell$, the phase function φ is a real–valued smooth function in $\Omega \subset \mathbb{R}^n \times \mathbb{R}^\ell$, while the amplitude a is a complex–valued smooth function in $\mathbb{R}^n \times [1,+\infty)$, and also a symbol of order m (see the definitions below).

We will concentrate on the simpler case

$$I_p^\varphi(\lambda) := \int_{\mathbb{R}^n} e^{i\lambda\varphi(x)} p(x,\lambda)\,dx \tag{2.2}$$

where no external parameters ω appear, and consider Morse non degeneracy for φ, *i.e.* we require that every critical point x_0, $d\varphi(x_0) = 0$, to be non degenerate, *i.e.* $\det d^2\varphi(x_0) \neq 0$.

The first motivation for this investigation is to extend the results of chapter 1 to a first situation where critical points occur. In fact, the problems in optics require a rather precise analysis of integrals of the type of (2.2). Indeed, the Huygens principle in wave optics, states that the illumination in a point $\omega \in \mathbb{R}^3$ suscitated by a light wave which is completely determined, say, on a surface Σ, can be computed as a superposition of elementary contributions coming from every infinitesimal portion of Σ. This superposition of infinitesimal waves can be expressed precisely by means of an oscillatory integral.

Moreover, for $\omega \in \mathbb{R}^3$ fairly distant from the caustic and λ sufficiently large, the short wave approximation holds, *i.e.* the oscillatory integral can be substituted by a finite sum of contributions which can be plainly interpreted as rays. This fact is proved by means of the stationary phase method, and in a single time makes descend geometrical optics from wave optics and offers a support for the Fermat principle.

We set ourselves in the Gevrey class of functions in order to obtain finer estimates than which obtainable in the C^∞ class and also to allow ourselves to employ compactly supported amplitudes, which cannot be done in the analytical class, as explained in chapter 1.

We will also make use of refined scales of anisotropic Gevrey spaces splitting (separating) the Gevrey regularity with respect to the variables x and ω, and the large parameter λ.

Given $\rho, \sigma, \theta \geqslant 1$ we define the following spaces of formal symbols:

1. Let $F^\theta_m[1, +\infty)$ be the set of all formal series

$$\sum_{j=0}^{\infty} \kappa_{m-j}\lambda^{m-j} \in F_m[1, +\infty) \tag{2.3}$$

   such that there exists a constant C such that

$$|\kappa_{m-j}| \leqslant C^{j+1} (j!)^\theta, \qquad \forall j \in \mathbb{N} \tag{2.4}$$

2. Let $F^{\sigma,\theta}_m(\mathbb{R}^n \times [1, +\infty))$ be the set of all

$$\sum_{j=0}^{\infty} p_{m-j}(x)\lambda^{m-j} \in F_m(\mathbb{R}^n \times [1, +\infty)), \tag{2.5}$$

   such that, for every compact subset $K \subset \mathbb{R}^n$, there exists a constant C_K, such that

$$|\partial^\alpha p_{m-j}(x)| \leqslant C_K^{|\alpha|+j+1} (\alpha!)^\sigma (j!)^\theta, \qquad \forall \alpha \in \mathbb{N}^n, \forall x \in K. \tag{2.6}$$

3. Let $F_m^{\rho,\sigma,\theta}(\mathbb{R}^n \times [1,+\infty))$ be the set of all

$$\sum_{j=0}^{\infty} p_{m-j}(x,\omega)\lambda^{m-j} \in F_m(\mathbb{R}^n \times [1,+\infty)), \tag{2.7}$$

such that, for every compact subset $K \subset \mathbb{R}^n$, there exists a constant C_K, such that

$$\left|\partial_x^\alpha \partial_\omega^\beta p_{m-j}(x,\omega)\right| \leqslant C_K^{|\alpha|+\beta|+j+1} (\alpha!)^\sigma (\beta!)^\rho (j!)^\theta, \qquad \forall \alpha \in \mathbb{N}^n, \forall x \in K. \tag{2.8}$$

We observe that setting $\varepsilon = \lambda^{-1}$ we obtain that θ coincides with the Gevrey index for the formal Gevrey power series as in [Ram85] and in [Miy93].

We emphasize that there are essentially two major issues related to the study of the asymptotics of oscillatory integrals $I(\lambda)$, $\lambda \gg 1$ as above. The first goal is to derive a formal asymptotic expansion

$$I(\lambda) \sim \sum_{j=0}^{\infty} I_j \lambda^{\mu - j\nu}, \qquad I_j \in \mathbb{C},\ \lambda \gg 1 \tag{2.9}$$

for some $\mu \in \mathbb{R}$, $\nu > 0$, and secondly, to study the Gevrey character of the formal series above.

Secondly, one deals with the remainder, namely we investigate, roughly speaking, the rate of decay for $\lambda \to +\infty$ of

$$R_N(\lambda) := I(\lambda) - \sum_{j=0}^{N} I_j \lambda^{\mu - j\nu}, \qquad N \in \mathbb{N},\ \lambda \gg 1 \tag{2.10}$$

Again the Gevrey classes are a natural framework for estimating the type of decay of $R_N(\lambda)$.

2.2 Morse critical points for the phase

The main result of this chapter is on the loss

$$p \in F^{\sigma,\theta} \qquad \Rightarrow \qquad I \in F^{\max\{2\sigma-1,\theta\}}$$

of formal Gevrey regularity for the stationary phase method (cf. [Gra87], see also [GP95]) when the phase function is nonanalytic.

Note that when p is non analytic, *i.e.* $\sigma > 1$, then $2\sigma - 1 > \sigma$, so the Gevrey character of I is strictly greater (*worse*). Otherwise, when p is analytic, there is no degeneracy, because $2\sigma - 1 = 2 \cdot 1 - 1 = 1 = \sigma$.

We decouple the influence of the formal Gevrey character on the well known formal series appearing in SPM.

THEOREM 4. *Let* $\varphi(x) \in G^\sigma(\mathbb{R}^n, \mathbb{R})$ *and* $p(x,\lambda) \sim \sum p_j(x)\lambda^{-j} \in F_m^{\sigma,\theta}(\mathbb{R}^n, \mathbb{R})$. *Suppose there is a unique critical point* x_0 *for* φ *and suppose Morse non degeneracy,* i.e. $\exists! x_0, \nabla\varphi(x_0) = 0,$ *and* $\nabla^2\varphi(x_0) \neq 0$. *Then*

$$I(\lambda) = \int e^{i\lambda\varphi(x)} p(x,\lambda)dx = e^{i\lambda\varphi(x_0)} q(\lambda), \tag{2.11}$$

$$\textit{where} \quad q(\lambda) \sim \sum_{k=0}^{\infty} q_{-\frac{n}{2}+m-k} \lambda^{-\frac{n}{2}+m-k} \in F_{m-n/2}^{\max\{2\sigma-1,\theta\}}([1,+\infty)) \tag{2.12}$$

where, as it is well known from the complete asymptotic expansion for the stationary phase method

$$\begin{aligned} q_{-\frac{n}{2}+m-k} \quad &= \quad \frac{e^{i(\frac{\pi}{4})\operatorname{sgn} Q}}{\det Q} \sum_{j+s=k} \frac{(2i)^{-s}}{s!} \left\langle Q^{-1}\frac{\partial}{\partial y}, \frac{\partial}{\partial y} \right\rangle^s \times \\ &\qquad \times \left(p_{m-j}(\kappa(y),\lambda)\left|\det \kappa'(y)\right|\right)\Big|_{y=0} \end{aligned} \tag{2.13}$$

with $x = \kappa(y)$ *being the change of the variables transforming the phase function* $\varphi(x)$ *into* $\langle Qy, y\rangle$

Proof. By a generalization of the Morse lemma in G^σ classes (see Appendix 2.3 for the proof) there exists an appropriate G^σ change of variables $x = \kappa(y) \in G^\sigma$, with respect to which the phase function $\kappa^*\varphi(y) = \varphi(\kappa(y))$ becomes a quadratic form $\langle Qy, y\rangle$.

The new amplitude $\tilde{p}(y,\lambda)$ is defined by

$$\tilde{p}(y,\lambda) = p(\kappa(y),\lambda)\left|\det \kappa'(y)\right|.$$

Thus we can write,

$$q(\lambda) = \int e^{i\lambda\frac{\langle Qy,y\rangle}{2}} \tilde{p}(y,\lambda)dx,$$

where $\tilde{p}$ is still a formal $G^{\sigma,\theta}$ symbol.

According to the well known formula,

$$q_{-\frac{n}{2}+m-k} = \frac{e^{i(\frac{\pi}{4})\operatorname{sgn} Q}}{\det Q} \sum_{j+s=k} \frac{(2i)^{-s}}{s!} \left\langle Q^{-1}\frac{\partial}{\partial y}, \frac{\partial}{\partial y} \right\rangle^s \tilde{p}_{m-j}(y)\Big|_{y=0}.$$

The degree of derivation to which $\tilde{p}_{m-j}(y)$ is subject is $2s$, thus, by the inequality (2.4),

$$\begin{aligned}\left|q_{m-\frac{n}{2}-k}\right| \leqslant A \sum_{j+s=k} \frac{1}{2^s}\frac{1}{s!}\left|\partial^{2s}\tilde{p}_{m-j}(0)\right| \leqslant \\ \leqslant A \sum_{j+s=k} \frac{1}{2^s}\frac{1}{s!} C^{2s+j+1}\left((2s)!\right)^{\sigma}(j!)^{\theta} \approx \\ \approx Ak \max_{j+s=k} \frac{1}{2^s}\frac{1}{s!} C^{2s+j+1}(s)!^{2\sigma}(j!)^{\theta} \approx \\ \approx \left(\widetilde{C}\right)^{2k+1}(k!)^{\max\{2\sigma-1,\theta\}}, \quad (2.14)\end{aligned}$$

and the expression above yields the end of the proof. □

Remark 2. One can introduce Banach spaces of formal $G^{\sigma,\theta}$ symbols with norms of the type

$$\|p\|_{\sigma,\theta;T} := \sum_{\alpha\in\mathbb{Z}^n_+, j\in\mathbb{Z}_+} \frac{T^{|\alpha|+j}}{(\alpha!)^{\sigma}(j!)^{\theta}} \sup_{x\in K}|\partial^{\alpha}p_{m-j}(x)| \quad (2.15)$$

and, after precise combinatorial estimates via the Stirling formula, consider SPM with fixed phase as a linear operator acting between two Banach spaces with formal symbols. Such results, a part of the theoretical value in itself might be useful in future investigations. Motivations for such approach are based on results for singular PDE in Gevrey spaces, divergent Gevrey formal power series in Dynamical Systems.

We point out that if $\sigma=\theta$ we recover the loss of $\sigma-1$ Gevrey regularity studied in [Gra87, Cap03].

2.3 Gevrey anisotropy for the amplitude

Let Q be a symmetric non–degenerate real matrix in $\mathbb{R}^{2n}$ with signature type (n,n). The next theorem shows that we may reduce the loss of Gevrey smoothness if we impose additional regularity in anisotropic Gevrey spaces $G^{\sigma,\rho}$. Let $\sigma,\rho\geqslant 1$. We define $G^{\sigma,\rho;Q}(\mathbb{R}^{2n})$ as the set of all $g\in C^{\infty}(\mathbb{R}^{2n})$ such that there exist $S\in GL(2n;\mathbb{R})$ satisfying

$$S\circ Q\circ S^{-1} = \begin{pmatrix} 0 & I_n \\ I_n & 0 \end{pmatrix} \quad (2.16)$$

and the function $S^*g(z) = g(Sz)$ satisfies the following σ, ρ anisotropic Gevrey estimates: for every $K \subset\subset \mathbb{R}^{2n}$ there exists $C > 0$ such that

$$\sup_{(x,y)\in K} |\partial_x^\alpha \partial_y^\beta S^* g(x,y)| \leqslant C^{|\alpha|+|\beta|+1}(\alpha!)^\sigma(\beta!)^\rho, \qquad \alpha, \beta \in \mathbb{Z}_+^n \tag{2.17}$$

We note that $G^{\sigma,\rho;Q}(\mathbb{R}^{2n}) \subset G^{\max\{\sigma,\rho\}}(\mathbb{R}^{2n})$.

Setting $f(x,y) := S^*g(x,y)$, let us estimate its multi–derivative with respect to $\gamma = (\alpha, \beta) \in \mathbb{N}^n \times \mathbb{N}^n$,

$$\begin{aligned} \left|\left(\frac{\partial}{\partial(x,y)}\right)^\gamma f\right| = \left|\left(\frac{\partial}{\partial x}\right)^\alpha \left(\frac{\partial}{\partial y}\right)^\beta f\right| &\leqslant C^{|\alpha|+|\beta|+1}(\alpha!)^\sigma(\beta!)^\rho \leqslant \\ &\leqslant C^{|\alpha|+|\beta|+1}(\alpha!\beta!)^{\max\{\sigma,\rho\}} = C^{|\gamma|+1}(\gamma!)^{\max\{\sigma,\rho\}}, \end{aligned}$$

because, plainly, $\gamma! = \alpha!\beta!$ and $|\gamma| = |\alpha| + |\beta|$. So f is in $G^{\max\{\rho,\sigma\}}$, and as a result also g is, being S an analytical (linear!) diffeomorphism.

The next result shows that the $G^{\sigma,\rho;Q}(\mathbb{R}^{2n})$ anisotropic regularity of the amplitude leads to an improvement of the Gevrey index of the asymptotic expansion.

THEOREM 5. *Let Q be a $2n\times 2n$ non–degenerate real symmetric matrix with signature type (n,n). Let*

$$a \in G_0^{\sigma,\rho;Q}(\mathbb{R}^{2n}). \tag{2.18}$$

Then

$$I(\lambda) = \int e^{i\lambda\langle Qz,z\rangle} a(z)dz \sim \sum_{k=0}^{\infty} a_{-n-k}\lambda^{-n-k} \in F_{-n}^{\sigma+\rho-1}(\mathbb{R}^{2n}). \tag{2.19}$$

Proof. The proof follows from the fact that

$$\langle Qz, z\rangle = xy, \qquad z = (x,y) \in \mathbb{R}^{2n},$$

the explicit formulas for the asymptotic expansion and the assumption $a \in G_0^{\sigma,\rho;Q}(\mathbb{R}^{2n})$. Indeed, the form of Q implies that

$$\left\langle Q^{-1}\frac{\partial}{\partial z}, \frac{\partial}{\partial z}\right\rangle = \sum_{j=1}^{n} \frac{\partial^2}{\partial x_j \partial y_j}. \tag{2.20}$$

Next, in view of (2.18), we get

$$a_{-n-k} = \frac{e^{i(\frac{\pi}{4})n}}{(-1)^n}\frac{(2i)^{-k}}{k!}\left(\sum_{j=1}^{n}\frac{\partial^2}{\partial x_j \partial y_j}\right)^k S^*a(x,y)\Big|_{(x,y)=(0,0)}.$$

which allows us to estimate in the following way

$$\begin{aligned}|a_{-n-k}| \leqslant C_0 \frac{1}{2^k}\frac{1}{k!}\left|\partial_x^k\partial_y^k S^* a(x,y)\right|\Big|_{(x,y)=(0,0)} \leqslant \\ \leqslant C_0 \frac{1}{2^k}\frac{1}{k!} C^{2k+1}(k!)^{\sigma}(k!)^{\rho} \approx \\ \approx C_0 \tilde{C}^{k+1}(k)!^{\sigma+\rho-1}, \quad (2.21)\end{aligned}$$

as desired.

□

Appendix: The Morse lemma for Gevrey functions

Lemma 1 (Morse lemma). *Let $f(x) : \mathbb{R}^n \longrightarrow \mathbb{R}$ be a C^∞ function such that*

$$\begin{cases} f(0) = 0, \\ \nabla f(0) = 0, \\ \nabla^2 f(0) \equiv A \in M(n \times n, \mathbb{R}), \qquad \det A \neq 0. \end{cases} \tag{2.22}$$

Then there exists local coordinates $y = y(x)$ in a neighborhood of 0 with respect to which

$$f(x(y)) = \frac{1}{2}\langle Ay, y\rangle .$$

Proof. We can rewrite f as a quadratic form with non constant coefficients applying twice the Hadamard's lemma:

$$\varphi(x) = \frac{1}{2}\langle B(x)x, x\rangle = \sum_{j,k=1,\dots,n} \frac{1}{2} b_{jk}(x)x_j x_k, \tag{2.23}$$

where

$$b_{jk} = 2\int_0^1 (1-t)\frac{\partial^2 f}{\partial x_j \partial x_k}(tx)dt. \tag{2.24}$$

Let us consider the function

$$\begin{aligned} B : \mathbb{R}^n &\longrightarrow M(n\times n, \mathbb{R}), \\ x &\longmapsto (b_{jk})(x), \end{aligned}$$

for which

$$B(0) = A = \left(\frac{\partial^2 f}{\partial x_j \partial x_k}\right).$$

We are looking for a change of coordinates $y = y(x) = R(x)x$ with respect to which the matrix $\langle B(x(y))x(y), x(y)\rangle \equiv \langle Ay, y\rangle$ for all y in a neighborhood of 0, *i.e.*

$$\begin{aligned} y = R(x)x, \qquad R(x)^T AR(x) = B(x), \\ R(0) = \mathbb{I}, \qquad R(x) \in M(n \times n, \mathbb{R}). \end{aligned}$$

Indeed, with respect such coordinates,

$$\begin{aligned} f(x) = f(x) = \langle B(x)x, x\rangle = \left\langle R(x)^T AR(x)x, x\right\rangle = \\ = \langle AR(x)x, R(x)x\rangle = \langle Ay, y\rangle . \end{aligned}$$

Now, the existence of such an $R(x)$ is assured if the map

$$\mathrm{Sym}(n \times n, \mathbb{R}) \longrightarrow \mathrm{Sym}(n \times n, \mathbb{R}), \tag{2.25}$$

$$R \longmapsto R^T AR, \tag{2.26}$$

is an isomorphism, and we will prove its invertibility by the surjectivity of its differential in $x = 0, R = \mathbb{I}$. Differentiating (2.25) then, we obtain

$$\begin{aligned} d[R^T AR](S)\Big|_{\substack{x=0\\R=\mathbb{I}}} = \\ = \frac{d}{d\lambda}\left((R + \lambda S)^T A(R + \lambda S)\right)\Big|_{\substack{\lambda=0\\x=0\\R=\mathbb{I}}} = S^T A + AS, \end{aligned} \tag{2.27}$$

which clearly is surjective, for every symmetric matrix C is image of $S = \frac{A^{-1}C}{2}$. □

Remark 3. Moreover, the invertible map (2.25) is polynomial in the entries of R, thus its inverse, say $F(B(x)) = F(R^T(x)AR(x)) = R(x)$ is analytic in the entries of B. We obtain then $y(x) = R(x)x$ is a G^σ change of coordinates whenever the Morse function $f(x)$, and by consequence $B(x)$, as apparent from (2.24), is G^σ.

Chapter 3

Reduction for PDE's

In Section 1 we introduce the main problem and the reduction technique. The main requirement on the nonlinear function F is the Lipschitz property. Employing spectral decomposition, we reduce a boundary value problem in a function space to an "algebraic" equation in $\mathbb{R}^m$. In Section 2, the reduction is successively applied to the eventual variational principle, bringing back the search for solutions to the determination of critical points on a n-variables function.

3.1 Semilinear Dirichlet problem

Let $\Omega \subseteq \mathbb{R}^n$ be a Stokes' domain, and consider the following space,

$$H := H_0^1(\Omega, \mathbb{R}^k),$$

i.e. the Sobolev space obtained from $C_0^\infty(\Omega, \mathbb{R}^k)$ by completion with respect to one of the following equivalent norms $\|u\|_{L^2} + \|\nabla u\|_{L^2}$, $\|\nabla u\|_{L^2}$ or also $(-Lu, u)_{L^2}$, where L is an arbitrary elliptic operator.

Indeed, as scalar product we will consider

$$\begin{aligned} \langle \cdot, \cdot \rangle : H \times H &\longrightarrow \mathbb{R}, \\ (u, v) &\longmapsto \langle u, v \rangle := \int_\Omega -Lu \cdot v dx, \end{aligned} \tag{3.1}$$

thus the norm here considered will be $\|u\| := \langle u, u \rangle$.

Note that when $L = \Delta$, $\langle u, v \rangle := \int_\Omega \nabla u \cdot \nabla v dx$.

Consider now the following semilinear Dirichlet problem:

$$\begin{cases} N(u) := -Lu - \lambda u - \varepsilon F(u) = 0, & \text{in } \Omega, \\ u\Big|_{\partial\Omega} = 0, & \text{on } \partial\Omega, \end{cases} \tag{3.2}$$

where,

(i) L is a linear elliptic operator,

(ii) λ is a (positive) real number,

(iii) $F : H \to H$ is a nonlinear operator,

(iv) $\varepsilon \in \mathbb{R}$ is a (small) perturbative parameter.

3.1.1 Green operator of $-L$

The eigenfunctions associated to the eigenvalues of the elliptic operator L form a basis for the space H,

$$\begin{aligned} -L\hat{u}_j &= \lambda_j \hat{u}_j, \\ 0 &= \lambda_0 < \lambda_1 \leqslant \lambda_2 \leqslant \dots, \\ \|\hat{u}_j\| &= 1. \end{aligned}$$

Note that passing to different norms only requires a rescaling of the $\hat{u}_j$.

For every $v \in H$ we can write

$$v = \sum_{j=1}^{\infty} \langle v, \hat{u}_j \rangle \hat{u}_j = \sum_{j=1}^{\infty} v_j \hat{u}_j.$$

Sometimes it would be useful to identify $v \in H$ with the sequence $\{v_j\}_{j=1}^{\infty}$ of its coefficients in the eigenvectors representation.

We are able now to define the Green operator of L, $g : H \to H$, $g = (-L)^{-1}$,

$$gv = g\left(\sum_{j=1}^{\infty} v_j \hat{u}_j\right) = \sum_{j=1}^{\infty} \frac{v_j}{\lambda_j} \hat{u}_j.$$

It is clear that

$$-Lgv = g(-L)v = v, \qquad \forall v \in H.$$

3.1.2 Cut-off decomposition

For every fixed $m \in \mathbb{N}$, we can consider the following decomposition of H:

$$\begin{gathered} H = \mathbb{P}_m H \oplus \mathbb{Q}_m H, \\ v = \mathbb{P}_m v + \mathbb{Q}_m v = \sum_{j=1}^{m} v_j \hat{u}_j + \sum_{j=m+1}^{\infty} v_j \hat{u}_j. \end{gathered}$$

We will briefly write

$$v = \mu + \eta = \sum_{j=1}^{m} \mu_j \hat{u}_j + \sum_{j=m+1}^{\infty} \eta_j \hat{u}_j$$

we will refer to $\mu \in \mathbb{P}_m H$ as to the *finite* head of v, whereas to $\eta \in \mathbb{Q}_m H$ as to the *infinite* tail of v.

3.1.3 Splitting and reduction of the problem

If we substitute the relation $u = gv$ the Dirichlet problem 3.2, we obtain

$$\begin{aligned} v - \lambda g v &= \varepsilon F(gv), \\ v &= (\mathbb{I} - \lambda g)^{-1} \varepsilon F(gv), \end{aligned}$$

which solutions, in other words, are just the fixed points of the map:

$$\begin{aligned} H &\longrightarrow H, \\ v &\longmapsto (\mathbb{I} - \lambda g)^{-1} F(gv). \end{aligned}$$

Note that we can write easily the map $(\mathbb{I} - \lambda g)^{-1}$ as

$$(\mathbb{I} - \lambda g)^{-1} v = (\mathbb{I} - \lambda g)^{-1} \left(\{v_j\}_{j=1}^{\infty} \right) = \left\{ \frac{\lambda_j}{\lambda_j - \lambda} v_j \right\}_{j=1}^{\infty},$$

which of course is well defined when λ is different from every λ_j.

The application of the cut-off decomposition splits the original problem into a finite and an infinite part:

$$v = \mathbb{P}_m v + \mathbb{Q}_m v = \mu + \eta, \tag{3.3}$$

$$\mu = \mathbb{P}_m (\mathbb{I} - \lambda g)^{-1} \varepsilon F(g(\mu + \eta)), \qquad \text{“finite”} \tag{3.4}$$

$$\eta = \mathbb{Q}_m (\mathbb{I} - \lambda g)^{-1} \varepsilon F(g(\mu + \eta)). \qquad \text{“infinite”} \tag{3.5}$$

These equations can also be equivalently rewritten as follows,

$$\mathbb{P}_m N(g(\mu + \eta)) = 0, \qquad \text{“finite”} \tag{3.6}$$

$$\mathbb{Q}_m N(g(\mu + \eta)) = 0. \qquad \text{“infinite”} \tag{3.7}$$

Maybe, providing some appropriate hypotheses on F, it would be possible to uniquely solve for η with respect to μ in the infinite part of the equation. In fact, we can prove that

Proposition 1. *If $F : H \to H$ is Lipschitz, for every fixed $\varepsilon > 0$, there exists $m \in \mathbb{N}$, such that*

$$\eta \longmapsto (\mathbb{I} - \lambda g)^{-1} \varepsilon \mathbb{Q}_m F(g(\mu + \eta)), \tag{3.8}$$

is a contraction, for every fixed $\mu \in \mathbb{P}_m H$.

Proof. We can divide the proof into several steps,

1^{st} **step** Trivially, $\eta \mapsto g(\mu + \eta)$ is a Lipschitz map with constant $\frac{1}{\lambda_{m+1}}$.

2^{nd} **step** $(\mathbb{I} - \lambda g)^{-1}$ is bounded, indeed

$$\left\| (\mathbb{I} - \lambda g)^{-1} v \right\| = \left\| \left(\frac{\lambda_j}{\lambda_j - \lambda} v_j \right) \right\| \leqslant \sup_{j \in N} \left| \frac{\lambda_j}{\lambda_j - \lambda} \right| \|v\|,$$

and clearly $\sup \left| \frac{\lambda_j}{\lambda_j - \lambda} \right| := C_1 < +\infty$, because the sequence $\frac{\lambda_j}{\lambda_j - \lambda} \to 1$.

3^{rd} **step** by assumption, F is Lipschitz, so let us denote by C_2 the constant such that

$$\|F(u_1) - F(u_2)\| \leqslant C_2 \|u_1 - u_2\|.$$

Finally the map (3.8) is Lipschitz:

$$\left\| (\mathbb{I} - \lambda g)^{-1} \varepsilon \mathbb{Q}_m F(g(\mu + \eta_1)) - (\mathbb{I} - \lambda g)^{-1} \varepsilon \mathbb{Q}_m F(g(\mu + \eta_2)) \right\| \leqslant$$
$$\leqslant \frac{\varepsilon C_1 C_2}{\lambda_{m+1}} \|\eta_1 - \eta_2\|.$$

For every $\varepsilon > 0$ there exists a sufficiently large m, such that the Lipschitz constant $\frac{\varepsilon C_1 C_2}{\lambda_{m+1}}$ results in something smaller than 1, *i.e.* the map (3.8) is contractive as claimed. □

Let us denote by $\tilde{\eta}(m, \mu)$, or simply $\tilde{\eta}(\mu)$ if there is no ambiguity, the unique fixed point of the map (3.8). Substituting it into equation (3.4), we obtain

$$\mu = (\mathbb{I} - \lambda g)^{-1} \varepsilon \mathbb{P}_m F(g(\mu + \tilde{\eta}(m, \mu))), \tag{3.9}$$

which indeterminate is merely $\mu = (\mu_1, \ldots, \mu_m) \in \mathbb{R}^m$. The problem has reached a completely finite (say, algebraic) formulation, though passing through a contraction in the infinite dimensional space $\mathbb{Q}_m H$.

To every solution $\tilde{\mu}$ of this last equation corresponds a solution of the original Dirichlet Problem (3.2) by means of the formula:

$$\tilde{u} = g(\tilde{\mu} + \tilde{\eta}(m, \tilde{\mu})),$$

and clearly viceversa,

$$\tilde{\mu} = \mathbb{P}_m(-L)\tilde{u},$$

to every solution of (3.2) corresponds a solution of (3.9).

3.2 Qualitative analysis of the solution set

3.2.1 Increasing the number of the parameters

Fix $\varepsilon > 0$ and let $m_\varepsilon \in \mathbb{N}$ be the smallest integer such that $\frac{\varepsilon C_1 C_2}{\lambda_{m_\varepsilon+1}} < 1$.

We want to put in evidence that every $m > m_\varepsilon$ produces a reduced problem $(3.9)_m$ which possesses a smaller contractive constant, but the corresponding generated solutions are precisely whose of $(3.9)_{m_\varepsilon}$.

In some more detail, we want to show that the following problems are equivalent

(a) Find $\hat{\mu} \in \mathbb{R}^{m_\varepsilon}$, $\hat{\mu}$ solves $(3.9)_{m_\varepsilon}$

(b) Find $\check{\mu} \in \mathbb{R}^{m}$, $\check{\mu}$ solves $(3.9)_{m}$

(c) Find $u \in H$, u solves (3.2)

We will prove $(a) \Rightarrow (c) \Rightarrow (b)$. Indeed, if $\hat{\mu}$ solves (a), then $u := g(\hat{\mu} + \tilde{\eta}(m_\varepsilon, \hat{\mu}))$ solves (c), and as a result we get $\check{\mu} = \mathbb{P}_m(-L)u$ and $\check{\eta} = \mathbb{Q}_m(-L)u$. It is clear that $\check{\eta} = \tilde{\eta}(m, \check{\mu})$, because $\check{\eta}$ is a fixed point for $(3.8)_m$, with $\mu = \check{\mu}$, being u a fixed point (a solution) for the *functional* equation (c). The actual role of the contractiveness of $(3.8)_m$ is to guarantee the *uniqueness* of the fixed point, and as a consequence the good definition of the head-to-tail map $\mu \mapsto \tilde{\eta}(m, \mu)$. The same argument works also for the converse $(b) \Rightarrow (c) \Rightarrow (a)$, and finally we get

$$\begin{array}{ccccccccc}
\hat{\mu}_1 & \dots & \hat{\mu}_{m_\varepsilon} & \tilde{\eta}_1(m_\varepsilon,\hat{\mu}) & \tilde{\eta}_2(m_\varepsilon,\hat{\mu}) & \dots & \tilde{\eta}_{m-m_\varepsilon}(m_\varepsilon,\hat{\mu}) & \tilde{\eta}_{m-m_\varepsilon+1}(m_\varepsilon,\hat{\mu}) & \dots \\
|| & \dots & || & || & || & \dots & || & || & \dots \\
\check{\mu}_1 & \dots & \check{\mu}_{m_\varepsilon} & \check{\mu}_{m_\varepsilon+1} & \check{\mu}_{m_\varepsilon+2} & \dots & \check{\mu}_m & \tilde{\eta}_1(m,\check{\mu}) & \dots
\end{array}$$

The entry-to-entry equality of the preceding sequences may seem not completely clear at first sight. In particular, the interdependence between m_ε–head and m–tail, and between m–head and m_ε–tail, are not immediate. However, the entry–to–entry equality becomes apparent thinking to the functional solution u generated separately by each one of the reductions. In this way, the sequences can be considered simply as the projections on the eigenspectrum, and the difference between m and m_ε becomes only a matter of labelling.

$$\begin{pmatrix} \hat{\mu} \\ \tilde{\eta}(m_\varepsilon,\hat{\mu}) \end{pmatrix} = \begin{pmatrix} \hat{\mu} \\ \hat{\eta} \end{pmatrix} \leftarrow u \rightarrow \begin{pmatrix} \check{\mu} \\ \check{\eta} \end{pmatrix} = \begin{pmatrix} \check{\mu} \\ \tilde{\eta}(m,\check{\mu}) \end{pmatrix}$$

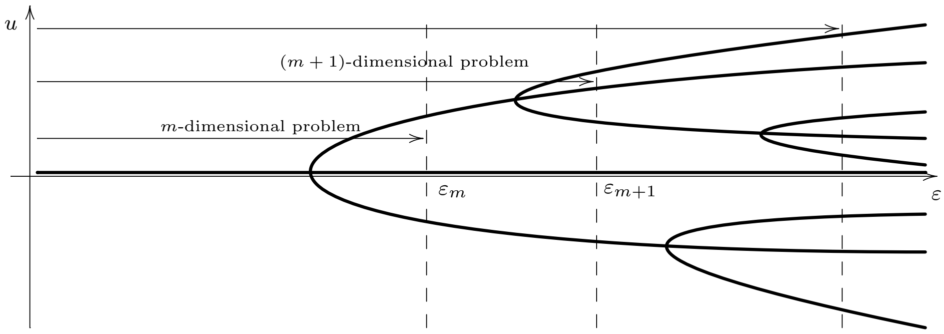
$$\begin{array}{ccccccccc}
\hat{\mu}_1 & \dots & \hat{\mu}_{m_\varepsilon} & \tilde{\eta}_1(m_\varepsilon,\hat{\mu}) & \dots & \tilde{\eta}_{m-m_\varepsilon}(m_\varepsilon,\hat{\mu}) & \tilde{\eta}_{m-m_\varepsilon+1}(m_\varepsilon,\hat{\mu}) & \dots \\
\uparrow & \dots & \uparrow & \uparrow & \dots & \uparrow & \uparrow & \dots \\
\langle u,\hat{u}_1\rangle & \dots & \langle u,\hat{u}_{m_\varepsilon}\rangle & \langle u,\hat{u}_{m_\varepsilon+1}\rangle & \dots & \langle u,\hat{u}_m\rangle & \langle u,\hat{u}_{m+1}\rangle & \dots \\
\downarrow & \dots & \downarrow & \downarrow & \dots & \downarrow & \downarrow & \dots \\
\check{\mu}_1 & \dots & \check{\mu}_{m_\varepsilon} & \check{\mu}_{m_\varepsilon+1} & \dots & \check{\mu}_m & \tilde{\eta}_1(m,\check{\mu}) & \dots
\end{array}$$

In spite of the equivalence of the solution sets of successive finite parameters reductions, it could be convenient to speed up the convergence of $\tilde{\eta}(m,\mu)$ increasing m. On the other hand, the cost of this advantage is larger number of variables to manage with, so an optimal choice of the dimension of the reduction must take into account these questions.

3.2.2 Hierarchically nested structure of the set of solutions

Because of the nonlinearity of F, the questions about existence, uniqueness and multiplicity of the solutions of (3.2) are far from trivial.

One could however naturally think that the complexity of the solution set could be somewhat simplified after the reductions described in this chapter. In fact, all the complexity of the solution set of (3.2) is preserved, but somehow pressed into a suitable finite dimensional space. The dimension of this space depends on the Lipschitz constant of F, *i.e.* on the magnitude of $\varepsilon|F'|$. More precisely, for every fixed $m \in \mathbb{N}$, there is a complete interval $[0, \varepsilon_m[$, for which the problem (3.2) can be generated by m eigenfunctions. If ε exceeds this threshold ε_m, it is necessary to increase the number of eigenfunctions to $m+1$, which keeps working until the next threshold $\varepsilon_{m+1} > \varepsilon_m$, and so on. This way, the solution set of the family of Dirichlet problems $(3.2)_\varepsilon$ can be seen as a *hierarchically nested system* of discrete problems, spanned by the growing parameter ε.



3.2.3 Looking for bifurcations

We do know very little about the global structure of the set of the solutions, we can only study the local structure of the reduced equation (3.9) for $\varepsilon \simeq 0$ and $\mu \simeq 0$. Let us rewrite equation (3.9) as

$$\mathcal{G}(\mu, \varepsilon) = 0. \tag{3.10}$$

Clearly, $\mathcal{G}(0,0) = 0$ for certainly $\mu = 0$ is a solution if $\varepsilon = 0$. Moreover,

$$\frac{\partial \mathcal{G}}{\partial \mu}(0,0) = \mathbb{I}_{\mu}, \tag{3.11}$$

so, by the Implicit Function Theorem (IFT), there exists a (possibly short) unique branch of solutions

$$]-\bar{\varepsilon}, \bar{\varepsilon}[\ni \varepsilon \longmapsto \mu(\varepsilon) \in \mathbb{R}^m, \qquad \mathcal{G}(\mu(\varepsilon), \varepsilon) \equiv 0, \tag{3.12}$$

passing thorough the origin $(0,0)$.

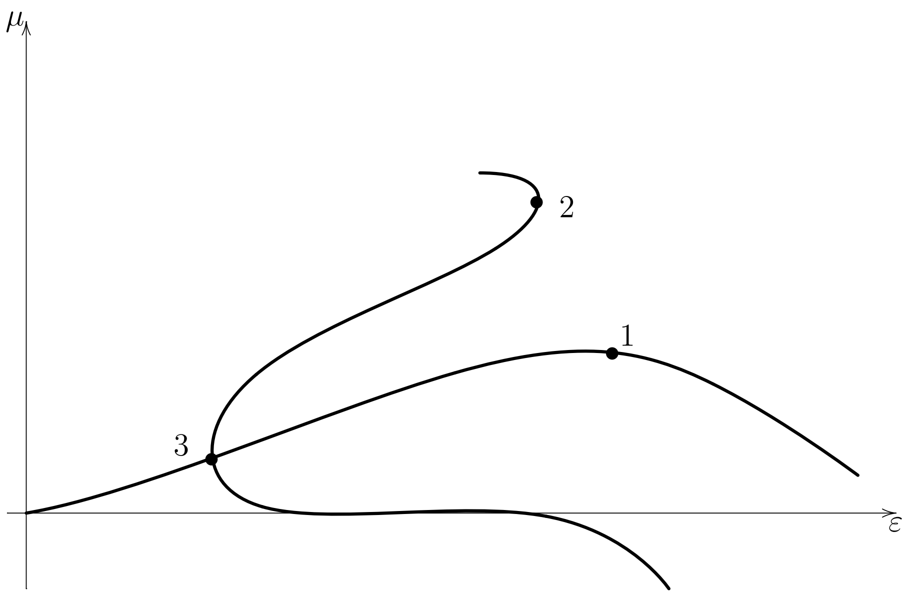
This branch keeps growing as long as $\det\left(\frac{\partial \mathcal{G}}{\partial \mu}\right) \neq 0$, otherwise, as soon as $\det\left(\frac{\partial \mathcal{G}}{\partial \mu}\right)\Big|_{\bar{\mu},\bar{\varepsilon}} = 0$, application of IFT fails, and everything could happen:

1. the branch keeps growing, even if the equation degenerates,

2. the branch turns back, towards smaller values of ε, or interrupts,

3. a new branch of solutions *bifurcates* from $(\mu(\bar{\varepsilon}), \bar{\varepsilon})$.

These situations are, only suggestively, represented here:



Existence questions like these will be treated more in detail in next chapter.

3.3 Infinite dimensional and reduced finite dimensional variational formulation.

If we assume the Gateaux derivative $N'(u)$ of the non linear operator considered (3.2) to be *symmetric* with respect to the L^2-scalar product, *i.e.* ,

$$(u, v) := \int_\Omega uvdx,$$
$$(N'(u)h, k) = (N'(u)k, h)\,, \qquad \forall h, k \in H,$$

the Volterra-Vainberg theorem admits us to write a variational principle,

$$\begin{aligned} J : H &\longrightarrow \mathbb{R}, \\ u \longmapsto J(u) &:= \int_{t=0}^{t=1} (N(tu), u)\, dt, \end{aligned} \tag{3.13}$$

which is equivalent to the original Dirichlet Problem, More precisely,

THEOREM 6. *Every critical point of J is a solution of (3.2), and viceversa,* i.e. ,

$$dJ(u)h = 0 \quad \forall h \in H \qquad \Leftrightarrow \qquad \begin{cases} N(u) = 0, \\ u\Big|_{\partial\Omega} = 0. \end{cases} \tag{3.14}$$

Proof.

$$dJ[u] \cdot h = \frac{d}{d\lambda} J[u + \lambda h]\Big|_{\lambda=0} = \int_0^1 (N'(tu)th, u) + (N(tu), h)\, dt =$$
$$= \int_0^1 (N'(tu)u, th) + (N(tu), h)\, dt = \int_0^1 \frac{d}{dt}(N(tu), th)\, dt = (N(u), h)\,.$$

So $dJ[u] = 0$ if and only if $N(u) = 0$ as claimed. □

3.3.1 Reduced variational principle

The finite parameters reduction of the preceeding section can be applied to the functional $J(u)$ to obtain a finite parameters variational principle, which is equivalent to the infinite dimensional one, *i.e.* every critical point of J is also a critical point for

$$\begin{aligned} W : \mathbb{R}^m &\longrightarrow \mathbb{R}, \\ W(\mu) &:= J(g(\mu + \tilde{\eta}(m, \mu))). \end{aligned} \tag{3.15}$$

We will show how this works proving the equivalence between the finite parameters variational principle $W(\mu)$ and the original Dirichlet problem (3.2).

THEOREM 7. *To every critical point $\tilde{\mu}$ of $W(\mu)$ corresponds a solution $\tilde{u} = g(\mu + \tilde{\eta}(\mu))$ of the problem (3.2) and viceversa to every solution $\tilde{u}$ of (3.2) corresponds a critical point $\tilde{\mu} = \mathbb{P}_m(-Lu)$ of the variational principle $W(\mu)$.*

Proof. Almost directly,

$$
\begin{aligned}
dW(\mu) &= dJ[u] \cdot \frac{d}{d\mu}\left(g(\mu + \tilde{\eta}(\mu))\right) d\mu = \\
&= \left(N(u)\Big|_{u=g(\mu+\tilde{\eta}(\mu))}, g(d\mu) + g(\tilde{\eta}'(\mu) d\mu) \right) = \\
&= \left(\mathbb{P}_m N(u) + \underbrace{\mathbb{Q}_m N(u)}_{=0 \text{ by } (3.7)}\Big|_{u=g(\mu+\tilde{\eta}(\mu))}, \underbrace{g(d\mu)}_{\in \mathbb{P}_m H} + \underbrace{g(\tilde{\eta}'(\mu) d\mu)}_{\in \mathbb{Q}_m H} \right) = \\
&= \left(\mathbb{P}_m N(g(\mu + \tilde{\eta}(\mu))), g(d\mu) \right).
\end{aligned}
$$

Being g acting as a diagonal isomorphism of $\mathbb{R}^m$ into itself, we can conclude that,

$$dW(\mu) = 0 \qquad \Leftrightarrow \qquad \mathbb{P}_m N(g(\mu + \tilde{\eta}(\mu))) = 0,$$

i.e. the variational principle is equivalent to the Dirichlet problem. □

We used in the last proof the fact that $\mathbb{P}_m H$ and $\mathbb{Q}_m H$ are still orthogonal also with respect to the L^2 scalar product, indeed:

LEMMA 2. $(\hat{u}_i, \hat{u}_j) = 0$ *whenever* $i \neq j$.

Proof. Clearly,

$$(\hat{u}_i, \hat{u}_j) = \frac{1}{\lambda_i}(-L\hat{u}_i, \hat{u}_j) = \frac{1}{\lambda_i}\langle \hat{u}_i, \hat{u}_j \rangle = 0.$$

□

This is the reason of the choices of the scalar product $\langle \cdot, \cdot \rangle$ to perform the spectral decomposition and the L^2 scalar product $(\cdot, \cdot)$ for the variational principle.

3.4 The fixed point map regularity

Because of the regularity of F required in order to consider the variational principle $J(u)$, the Lipschitz property can also be rewritten as

$$\|F'\| := \sup_{u \in H} \sup_{\|h\|=1} \|F'(u)h\| = C < +\infty.$$

From the regularity of F and by implicit function theorem we will prove the following:

PROPOSITION 2. *The fixed point map $\tilde{\eta}(\mu)$ is continuously differentiable.*

Proof. The contractive map

$$\eta \mapsto \mathcal{M}(\mu, \eta) := \mathbb{Q}_m F(g(\mu + \eta)), \tag{3.16}$$

is continuously differentiable with respect to η, and moreover, by the growing to infinity sequence of eigenvalues,

$$\left| \frac{\partial \mathcal{M}(\mu, \eta)}{\partial \eta} \right| \leqslant |F'| \frac{1}{\lambda_{m+1}} = \frac{C}{\lambda_{m+1}} < 1. \tag{3.17}$$

The differentiability of $\tilde{\eta}(\mu)$ can now be deduced by the implicit function theorem. Indeed, being $\tilde{\eta}$ the (unique) solution of

$$\mathcal{G}(\mu, \eta) := \mathcal{M}(\mu, \eta) - \eta = 0,$$

it suffices to show invertibility of the η-derivative of $\mathcal{G}$: $D_\eta \mathcal{M} - \mathbb{I}_\eta$.

Considering the identity,

$$(D_\eta \mathcal{M} - \mathbb{I}_\eta) \left(\mathbb{I}_\eta + D_\eta \mathcal{M} + \cdots + (D_\eta \mathcal{M})^k \right) = (D_\eta \mathcal{M})^{k+1} - \mathbb{I}_\eta,$$

and noting that

$$\lim_{k \to +\infty} \left\| (D_\eta \mathcal{M})^k \right\| = \lim_{k \to +\infty} \| D_\eta \mathcal{M} \|^k = 0,$$

we (formally) get

$$(D_\eta \mathcal{M} - \mathbb{I}_\eta)^{-1} = - \sum_{k=0}^{+\infty} (D_\eta \mathcal{M})^k .$$

The summation converges because of (3.17), indeed,

$$\left\| \sum_{k=0}^{+\infty} (D_\eta \mathcal{M})^k \right\| \leqslant \sum_{k=0}^{+\infty} \| D_\eta \mathcal{M} \|^k \leqslant \sum_{k=0}^{+\infty} \left(\frac{C}{\lambda_{m+1}} \right)^k = \frac{1}{1 - \frac{C}{\lambda_{m+1}}} < +\infty.$$

□

Chapter 4

Existence and multiplicity results. An overview

In this chapter we give an application of the techniques developed in the preceding chapter to existence problems. In Section 2, we put a brief survey on Lusternik-Schnirelmann theory, which will be employed in our existence theorem. In Section 3, we will explore quasi-quadratic functions, with particular care on their cohomological features, in relation with existence of critical points. In Section 4, we will prove that if F is C^1 and compactly supported, the reduced variational principle is a quasi quadratic function, so it possesses at least one critical point.

4.1 Introduction

In this section we will present a brief collection of results of existence and multiplicity of solutions for semilinear Dirichlet boundary value problems. We have mainly followed [Amb92] . We added also a result found by means of our reduction techniques.[1]

Let us consider the Dirichlet boundary value problem:

$$\begin{cases} -Lu = p(u), & \text{in } \Omega \subseteq \mathbb{R}^N, \\ u = 0, & \text{on } \partial\Omega, \end{cases} \tag{4.1}$$

where $p : \mathbb{R} \to \mathbb{R}$. The associated non linear operator on function spaces:

$$\begin{aligned} \tilde{p} : H &\longrightarrow H, \\ u &\longmapsto p \circ u, \end{aligned}$$

[1]See remarks in the introduction.

is a so-called *Nemitski* operator (*i.e.* a composition operator).

The first condition to be asked on p is requiring regularity

(p0). *$p \in C(\mathbb{R})$ is locally Hölder continuous.*

A second condition concerns the behavior at infinity,

(p1). $|p(s)| \leqslant a\,|s| + b, \qquad a, b > 0.$

1^{st} Existence Result. *If $p0$ and $p1$ hold, with $a < \lambda_1$, then there must be at least one solution for (4.1).*

This result is obtained exploiting the coercivity of the variational principle, which yields a global minimum. One can improve this result employing finer variational techniques, providing a finer knowledge about the behavior of p at 0 and at ∞.

(p2). $\limsup_{|s|\to+\infty} \frac{p(s)}{s} \leqslant a < \lambda_1$, *and* $p(s) = \lambda s - sh(s)$, *with* $h(0) = 0$.

2^{nd} Existence Result. $p0 + p2 \Rightarrow$

a) $\lambda > \lambda_1 \Rightarrow \exists$ *2 nontrivial solutions,*

b) $\lambda > \lambda_2 \Rightarrow \exists$ *3 nontrivial solutions,*

c) $\lambda > \lambda_k + p$ *odd* $\Rightarrow \exists 2k$ *nontrivial solutions.*

Lots of other results about asymptotical linearity at infinity (*e.g.* different asymptotes for $\pm\infty$, resonance $\lambda = \lambda_k$, additional dependence on $x \in \Omega$, etc...) can be found in [Amb92, AP93, Ben95, LMZ03, GMM01]

One can also consider polynomial growth at infinity,

(p3). *1.* $p \in C(\mathbb{R})$ *Hölder continuous and differentiable at 0.*

2. $\int_0^u p(t)dt \leqslant \theta u p(u), \qquad \exists r > 0, \exists \theta \in (0, \frac{1}{2}), \forall u \geqslant r.$

3. $|p(u)| \leqslant a_1 + a_2\,|u|^l, \qquad 1 < l < 2^\star - 1 := \frac{N+2}{N-2}.$

3^{rd} Existence Result. $p3 + p(0) = 0 \Rightarrow$ *(4.1) admits a nontrivial solution.*

This result is obtained via Mountain Pass Theorem or Linking Theorems.

When p is odd, this last result can be greatly improved. In fact it can be proved that

4^{th} Existence Result. *If p satisfy $p3$ and is odd, then (4.1) has infinitely many pairs of solutions.*

It is also possible to prove the existence of an arbitrarily large number of solutions when p is a (suitably small) perturbation of an odd function.

There are constitutive reasons imposing great care in the case when the growth at infinity reaches the Sobolev critical exponent $2^\star = \frac{2N}{N-2}$. In fact,

Non Existence Result. *If Ω is starshaped with respect to the origin, and $q \geqslant \frac{N+2}{N-2}$, then the boundary value problem,*

$$\begin{cases} -\Delta u = |u|^{q-1} u, \\ u = 0, \end{cases}$$

possesses only the trivial solution.

When p is non homogeneous, or Ω is not starshaped, there can be non-trivial solutions also with $q \geqslant 2^\star - 1$. Some difficulties must be overcome anyway, in particular the lacking of compactness of the immersion of H_0^1 into $L^{2^\star}$.

For a more complete discussion on these existence topics, the reader is referred to [Amb92, AP93, Ben95, LMZ03] and to the bibliography therein.

4.1.1 An existence result

As a byproduct of our reduction techniques we obtain the following simple existence result. It could be seen as an improvement of the first existence result of this section. We still do not know if there could be physically meaningful applications.

Theorem 8. *Let $p(s) = \lambda s + h(s)$, where $h : \mathbb{R} \to \mathbb{R}$. If h, its derivative h' and one of its primitives $\bar{h}$ are bounded, and $\lambda \neq \lambda_k, \forall k \in \mathbb{N}$, then (4.1) admits at least one solution.*

Example. *An example of such an h could be* $\sin(x)$ *for instance, or an arbitrary $h \in C_0^\infty(\mathbb{R})$, or $h \in C^1(\mathbb{R}) \cap L^1(\mathbb{R})$ such that* $\sup|h'| < +\infty$.

Here follows the idea of the proof. We gain the existence of a solution finding a critical point for the *energy functional*,

$$E(u) = \int_0^1 \langle N(tu), u \rangle \, dt,$$

In fact we employed the finite parameters functional $\tilde{E} : \mathbb{R}^m \to \mathbb{R}$, obtained via the techniques of chapter 3.

The main part of the work (see section 4.4) was to prove that $\tilde{E}$ is *quasi-quadratic*, *i.e.* it differs finitely from a non-degenerate quadratic form $\langle Qx, x\rangle$ in C^1 norm. (for a discussion on quasi-quadratic functions see section 4.3)

Proving then that the sublevel sets

$$Q^c = \left\{ x \middle| \ \langle Qx, x\rangle \leqslant c \right\}, \qquad \tilde{E}^c = \left\{ x \middle| \ \tilde{E}(x) \leqslant c \right\},$$

are diffeomorphic provided $c \gg 0$ is suitably large, we could apply the Lusternik-Schnirelmann theory (see section 4.2) and make correspond a critical point of $\tilde{E}$ to the critical point (0, of course) of Q. Note that this critical point is not a minimum in general.

Let us remark that this seems the first application in PDE of the theory of generating functions quadratic at infinity, originally developed by Chaperon, Sikorav and Viterbo studying Hamiltonian dynamics. (See [Cha84, Cha91, Sik86, Sik87, Vit90, Thé99])

4.2 Lusternik-Schnirelmann theory

We give an introduction to the Lusternik-Schnirelmann theory as we learned mainly by Claude Viterbo. The theorem proved here is the main tool employed in the construction of our existence result.

4.2.1 Motivation

Problem. *Let*

$$f : M \longrightarrow \mathbb{R}, \qquad f \in C^2.$$

Assume M compact. What can we say about the number of critical points of f?

$$\# \{x \in M : df(x) = 0\}?$$

Idea: Clearly there could not be an upper bound, because the local structure of f can be arbitrarily complicated. Maybe there are topological reasons from which we could argue a lower bound.

4.2.2 Construction of critical levels

Let us recall some notions on homology and cohomology. Cohomology on a differentiable manifold M can be defined as follows

$$H^k(M) := \frac{\{\text{closed } k-\text{forms on } \ M\}}{\{\text{exact } k-\text{forms on } \ M\}},$$

while, if $N \subseteq M$ is a submanifold (also with boundary), the relative cohomology of M on N is defined as,

$$H^k(M, N) := \{\alpha \text{ closed } k-\text{form on } \ M, \text{ exact on } \ N\}.$$

Idea: It is possible to associate a critical level $\gamma(\alpha, f)$ to every cohomology class, $\alpha \neq 0$, *i.e.* ,

$$\exists x, \quad f(x) = \gamma, \qquad df(x) = 0.$$

(There could be more than one critical point on a critical level.)

Thorough Poincaré duality, this claim can be rephrased in terms of homology classes. If $[\alpha] \in H_k(M)$, $[\alpha] \neq 0$, with representatives

$$\begin{aligned} j_\alpha : \alpha &\longrightarrow M, \\ \xi &\mapsto j_\alpha(\xi), \end{aligned}$$

the critical values are determined as:

$$\lambda_{[\alpha]} := \min_{\alpha \in [\alpha]} \max_{\xi \in \alpha} f \circ j_\alpha(\xi).$$

As an application, consider the situation represented in Figure 4.2.2, *i.e.* $M \cong \mathbb{T}^2 \subset \mathbb{R}^3$, while $f : M \to \mathbb{R}$ be a suitable "height" function. Consider then the singular k-cicles $N^k \longrightarrow M$.

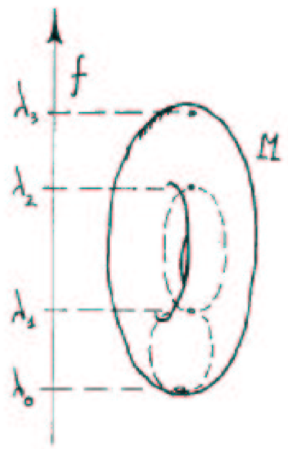


Figure 4.1: k-cycles on the torus

Starting from the 0-cycles on M, (*i.e.* the points), and letting them fall on the surface (in order to find the minimum of the height), one clearly finds the global minimum of f on M.

Looking at 1-cycles, *i.e.* $\mathbb{S}^1 \overset{j}{\hookrightarrow} M$, and taking the minimum height reached in each homotopy class, one finds the two saddles.

Lastly, considering 2-cycles, one finds the unique non degenerate class is the embedding of M into itself. Plainly, the minimum in the homotopy class of M of the maximum of f on M, keeps to be the maximum of f on M.

Let us now consider the problem from the cohomology viewpoint.

DEFINITION 4.2.1. *Let*

$$M^\lambda := \{x \in M : f(x) \leqslant \lambda\}.$$

denote the sublevel set.

Given any class $\alpha \in H^\star(M)$, $\alpha \neq 0$, it remains defined a critical level,

$$\gamma(\alpha, f) := \inf\left\{\lambda : \alpha\Big|_{M^\lambda} \neq 0 \quad \text{in} \quad H^\star(M^\lambda)\right\}. \tag{4.2}$$

$c = \gamma(\alpha, f)$ is a critical level because for every $\varepsilon > 0$, $H^\star(M^{c+\varepsilon}, M^{c-\varepsilon}) \neq 0$. Indeed, if c did not contain any critical point, then, by compactness, the same would be in $f^{-1}([c-\varepsilon, c+\varepsilon])$. It would be possible to deform $M^{c+\varepsilon}$ in $M^{c-\varepsilon}$, by means of the flow of the vector field

$$X(x) := -\frac{\nabla f(x)}{|\nabla f(x)|^2}, \qquad \text{per} \quad x \in M^{c+\varepsilon} \setminus M^{c-\varepsilon},$$

which could be regularly extended to all M to the zero vector field outside of a neighborhood of $M^{c+\varepsilon} \setminus M^{c-\varepsilon}$. One can easily check that if φ_t is the flow of $X(x)$, then

$$f(\varphi_t(x)) = f(x) + t.$$

By means of this deformation, α and $(\varphi_{2\varepsilon})^\star\alpha$ are homotopically equivalent, *i.e.* are in the same class, thus it cannot be $\alpha\Big|_{M^{c+\varepsilon}} \neq 0$ and $\alpha\Big|_{M^{c-\varepsilon}} = 0$.

Note that different cohomology classes could correspond to the same critical level.

The core result of this section is the following.

Theorem 9 (Lusternik-Schnirelmann). *Let $\beta \in H^\star(X) \setminus H^0(X)$. Then,*

1. $$\gamma(\alpha \wedge \beta, f) \geqslant \gamma(\alpha, f) \tag{$\star$}$$

2. *If ($\star$) is an equality, then the common critical level contains infinite many critical points.*

2'. *If ($\star$) is an equality, and if we denote by*

$$K_c = \{x : df(x) = 0, f(x) = c\}, \tag{4.3}$$

where $c = \gamma(\alpha \wedge \beta, f) = \gamma(\alpha, f)$, then

$$\beta \neq 0 \qquad \textit{in} \quad H^\star(K_c).$$

Proof. 1. Trivial, because every time $\alpha = 0$ in $H^\star(M^\lambda)$, surely $\alpha \wedge \beta = 0$ in $H^\star(M^\lambda)$.

2. Plainly 2' implies 2, because the unique nontrivial cohomology on a finite set of points is which of 0 degree. But β is at least a 1-forma. Then let us prove

2'. In our hypotheses we have $\alpha = 0$ in $H^\star(M^{c-\varepsilon})$ and $\alpha \wedge \beta \neq 0$ in $H^\star(M^{c+\varepsilon})$. In the situation represented in Figure 2, clearly $M^{c-\varepsilon}$ ed $M^{c+\varepsilon}$ cannot be diffeomorphic, moreover, K_c cannot be constituted by a unique point. Nevertheless, for a clearer visualization of the situation, we consider a section of M transversal to K_c.

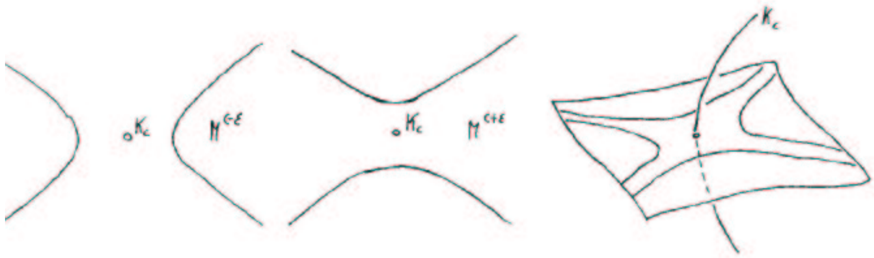


Figure 4.2: Sub-level sets

Let us assume by absurd the existence of a neighborhood U of K_c where $\beta = 0$.

We will then show the existence of a neighborhood W of K_c contained in U, and a retraction

$$M^{c+\varepsilon} \overset{\varphi}{\hookrightarrow} M^{c-\varepsilon} \cup W.$$

In such a case, the forms $\alpha \wedge \beta$ on $M^{c+\varepsilon}$ and on $M^{c-\varepsilon} \cup W$ are cohomologous, thus, being $\alpha = 0$ on $M^{c-\varepsilon}$ and $\beta = 0$ on W, we could argue $\alpha \wedge \beta = 0$ on $M^{c+\varepsilon}$, which is a contradiction.

Construction of W and of φ. The diffeomorphism φ will be given by the flow φ_s of the vector field $Y(x) := -\nabla f(x)$,[2] applied for a suitable

[2]In this case we clearly cannot employ $-\frac{\nabla f}{|\nabla f|^2}$, because we have to traverse a critical level, but we already suspected this, because such a diffeomorphism cannot exist, being different the cohomology groups between the two sublevel sets.

time. We also cannot keep U as it is, because we need something stable under φ, or something entering $M^{c-\varepsilon}$.

For this purpose we will take a subset of U and we will saturate it with the trajectories of the flow until we reach the boundaries of $M^{c+\varepsilon}$ and $M^{c-\varepsilon}$.

Let V open such that

$$K_c \subsetneqq V \subsetneqq \overline{V} \subsetneqq U.$$

We can assume without loss of generality $\overline{V}$ compact, being K_c compact.[3] Thus there exists $\delta := \operatorname{dist}(\overline{V}, \complement U) > 0$. Moreover, by compactness of M,[4] if ε is sufficiently small, there exists a $\eta > 0$ such that

$$|df(x)| > \eta, \qquad \forall x \in M^{c+\varepsilon} \setminus M^{c-\varepsilon}, \quad x \notin V.$$

As already observed, the distance between V and the boundary of U is non zero. Let us show that the trajectories connecting V to the boundary of $M^{c-\varepsilon}$, are shorter. Indeed,

$$\begin{aligned}\int_{s_0}^{s_1} |Y(\varphi_s(x))| ds &= -\int_{s_0}^{s_1} \frac{df(\varphi_s(x)) \cdot (-\nabla f(\varphi_s(x)))}{|df(\varphi_s(x))|} ds \leqslant \\ &\leqslant \frac{1}{\eta} \int_{s_1}^{s_0} df(\varphi_s(x)) \cdot \frac{d}{ds} \varphi_s(x) ds = \frac{1}{\eta} \left[f(\varphi_{s_0}(x)) - f(\varphi_{s_1}(x)) \right] \leqslant \\ &\leqslant \frac{c + \varepsilon - (c - \varepsilon)}{\eta} = \frac{2\varepsilon}{\eta}.\end{aligned}$$

Taking $\varepsilon < \frac{\eta}{2}\delta$, this length is less than δ. Thus $W := V \cup \{\text{trajectories}\} \subset U$, as wanted. (See Figure 2.)

[3] When M is non compact, one easily verifies that Palais-Smale condition implies compactness of K_c.

[4] or again by Palais-Smale

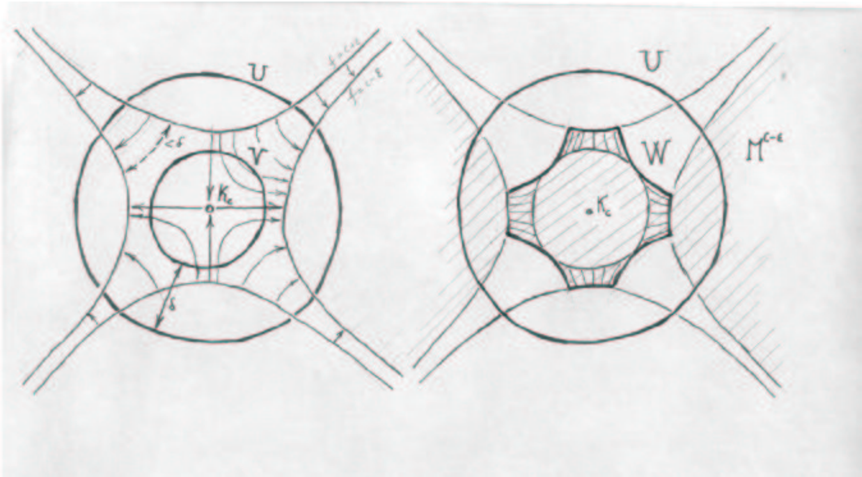


Figure 4.3: V "saturated"

Applying the retraction φ to $M^{c+\varepsilon}$. Let us give a look more in detail to what happens to $M^{c+\varepsilon}$ when applied the flow φ_s for a suitable time.

(a) $x \in W$. Two cases can occur: $f(\varphi_s(x)) \to c$, otherwise $f(\varphi_s(x))$ keeps on decreasing. In the first case the trajectory of x remains forever confined in W, because is moving towards a critical point. In the second case soon or later it has to become $f(\varphi_s(x)) < c-\varepsilon$. Thus $\varphi_s(x) \in M^{c-\varepsilon} \cup W$, for all $s > 0$.

(b) $x \notin W$. The trajectory of x will never meet V, so, as long as it remains in $M^{c+\varepsilon} \setminus M^{c-\varepsilon}$, it has to be $|df(\varphi_s(x))| > \eta$. If we take then $s_0 = \frac{2\varepsilon}{\eta^2}$,

$$
\begin{aligned}
f(\varphi_{s_0}(x)) - f(x) &= \int_0^{s_0} \frac{d}{ds}\left(f(\varphi_s(x))\right) ds = \\
&= -\int_0^{s_0} df(\varphi_s(x)) \cdot \nabla f(\varphi_s(x)) ds = -\int_0^{s_0} |df||\nabla f| ds \leqslant \\
&\leqslant -\int_0^{s_0} \eta \cdot \eta ds = -\eta^2 s_0 = -2\varepsilon,
\end{aligned}
$$

and by consequence,

$$f(\varphi_{s_0}(x)) \leqslant f(x) - 2\varepsilon \leqslant c + \varepsilon - 2\varepsilon = c - \varepsilon \quad \Rightarrow \quad \varphi_{s_0}(x) \in M^{c-\varepsilon},$$

as desired.

We have proved that

$$\varphi_{s_0}(M^{c+\varepsilon}) \subseteq M^{c-\varepsilon} \cup W.$$

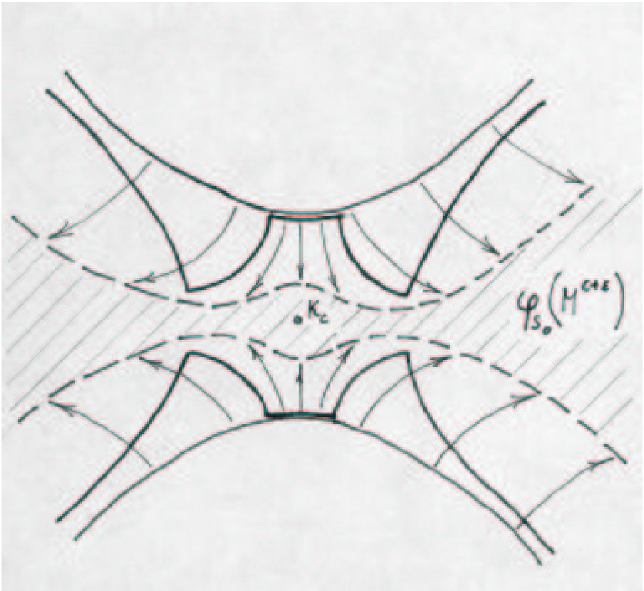


Figure 4.4: Applying the flow φ to $M^{c+\varepsilon}$

□

Corollary 1. *If we set*

$$cl(M) := \max\left\{k : \exists \alpha_1, \dots, \alpha_k \in H^\star(M) \setminus H^0(M)\right\},$$

then f possesses at least $cl(M)+1$ critical points.

Proof. Plainly, if 1 is the identity on $H^\star(M,\emptyset) = H^\star(M)$,

$$\gamma(1,f) \leqslant \gamma(\alpha_1,f) \leqslant \gamma(\alpha_1 \wedge \alpha_2, f) \leqslant \dots \leqslant \gamma(\alpha_1 \wedge \dots \wedge \alpha_k, f).$$

If the level sets are not all distinguished, there have to be infinite critical points. Otherwise, if the critical levels never coincide, there will be at least one critical point at each level, so they are at least $k+1$, as claimed. □

Example. *Let us consider the k-dimensional torus $\mathbb{T}^k$. Every f on the torus has always at least $k+1$ critical points. Indeed, if we consider the k angle functions on $\mathbb{T}^k$, they generate k distinguished forms $d\theta_1, \dots, d\theta_k$, with*

$$d\theta_1 \wedge \dots \wedge d\theta_k \neq 0.$$

Note. *The theorem works also in the non compact case, provided a suitable condition guaranteeing the existence of the flow of the gradient of* f*, in order to obtain the existence of a neighborhood of* K_c*, outside which* $|df| > \eta > 0$*.*

Thus it is requested that f *satisfy the* Palais-Smale condition, i.e.

$$\begin{aligned} &\forall \{n \mapsto x_n\}, \quad \textit{such that} \quad df(x_n) \to 0, \quad f(x_n) \quad \textit{bounded}, \\ &\qquad\qquad \textit{then} \quad \exists \{k \mapsto x_{n_k}\} \quad \textit{converging}. \end{aligned} \tag{P-S}$$

In such condition, K_c *results compact, and the existence of a suitable* η *is guaranteed. In next section, we will have to manage* $S, Q : \mathbb{R}^n \to \mathbb{R}$*, and to look for their critical points. Clearly, we will not be in the compact case, but it will be simple to check P-S condition when* Q *is a quadratic form and* S *is* C^1 *near to* Q*.*

4.2.3 Comparison between cup-length and category

Current techniques employed to find solutions in PDE are either topological, as local inversion theorems and degree theory, or variational, as Morse and Lusternik-Schnirelmann (L-S) theories.

Very often, L-S theory is introduced and developed around the topological notion of *category*. Here, we prefer to introduce this theory from the cohomological point of view, as we learned it by Claude Viterbo.

We thought it could be useful to show the equivalence between these two viewpoints.

Definition 4.2.2. *Let M be a orientable n-manifold. The* category *(*$\mathrm{cat}(M)$*) of M is the minimum number p such that there exist $A_1, \dots, A_p \subseteq M$, A_i closed and cotractible in M,*

$$\bigcup_{i=1}^{p} A_i = M.$$

It is usually non difficult to give upper bounds to the category. It seems harder to give lower bounds. We will expose a lower bound derived from cohomology.

Definition 4.2.3. *Let M be a orientable n-manifold. The* cohomological length (cup-length)($cl(M)$) *of M is the maximum number k for which there exist $\alpha_1, \dots, \alpha_k \in H^\star(M;\mathbb{Z}) \setminus H^0(M;\mathbb{Z})$ such that*

$$\alpha_1 \wedge \cdots \wedge \alpha_k \neq 0.$$

Theorem 10.

$$\mathrm{cat}(M) \geqslant cl(M) + 1.$$

Proof. Let $D : H^k(M;\mathbb{Z}) \longrightarrow H_{n-k}(M;\mathbb{Z})$ the Poincaré duality. If $\alpha, \beta \in H^\star(M;\mathbb{Z})$ are cocycles, and $\alpha \wedge \beta$ is the product in $H^\star(M;\mathbb{Z})$, holds

$$D(\alpha \wedge \beta) = D(\alpha) \cap D(\beta).$$

In other words, if γ_1, γ_2 are cycles corresponding respectively to α, β, the cycle $\gamma_1 \cap \gamma_2$ is the intersection of the manifolds, considered as generic.

Let k be the cup-length of M, *i.e.* there exist $\alpha_1, \dots, \alpha_k$, $\alpha_1 \wedge \cdots \wedge \alpha_k \neq 0$. Denote by $\gamma_i := D(\alpha_i)$, $i = 1, \dots, k$. So, by Poincaré, $D(\alpha_1 \wedge \cdots \wedge \alpha_k) = \gamma_1 \cap \cdots \cap \gamma_k = \gamma$ is a non zero cycle.

If it were $\mathrm{cat}(M) \leqslant k$, there would be $A_1, \dots, A_s$ closed contractible such that $M = \cup_{i=1}^{s} A_i$.

We can assume without loss of generality $s = k$. Let correspond γ_i to A_i (simply formally). Being A_i contractible, there is an immersion $H_l(M;\mathbb{Z}) \hookrightarrow H_l(M, A_i;\mathbb{Z})$, $\forall l > 0$. So there must be $\tilde{\gamma}_i$ homologous to γ_i with $\tilde{\gamma}_i \subseteq M \setminus A_i$. ($\gamma_i$ can be pulled away from A_i.)

Now,

$$\bigcap_{i=1}^{k} \tilde{\gamma}_i \cong \bigcap_{i=1}^{k} \gamma_i = \gamma \neq 0,$$

$$\gamma \cong \bigcap_{i=1}^{k} \tilde{\gamma}_i \subseteq \bigcap_{i=1}^{k} (M \setminus A_i) = M \setminus \left(\bigcup_{i=1}^{k} A_i \right) = \emptyset.$$

So we found $\gamma = 0$, which is a contradiction. □

Note. *For a complete overview on the modern applications of Lusternik-Schnirelmann theory see [CLOT03].*

4.3 *Quasi-quadratic* functions and their properties

At first, there were the Generating Functions Quadratic at Infinity or $FGQI$[5], which has been defined and developed by Marc Chaperon, Jean-Claude Sikorav and Claude Viterbo, in their works on hamiltonian dynamical systems. ([Cha84, Cha91, Sik86, Sik87, Thé99, Vit90]).

In particular, $FGQI$ are differentiable functions, $F : \mathbb{R}^n \longrightarrow \mathbb{R}$, such that $F(x) = \langle Qx, x\rangle$, for all $x \in \complement K$, for a prescribed $K \Subset \mathbb{R}^n$ and a non degenerate quadratic form Q.

For our purposes, the main features of $FGQI$ are the Palais-Smale property and, a consequence, the equivalence at infinity of the cohomological groups of F and Q, *i.e.*

$$H^\star(F^{-c}, F^c) \cong H^\star(Q^{-c}, Q^c),$$

for a sufficiently large $c > 0$.

On the other hand, Claude Viterbo and David Théret defined also *Generating Functions Quasi-Quadratic at Infinity*, $FGQQI$ or FGQ^2I, which will be extensively treated and exploited in this section and in the following one. For this class of functions still hold the Palais-Smale property and the cohomological groups equivalence. In fact, as proven in [Thé99], there is a complete equivalence between $FGQI$ and FGQ^2I, *i.e.* everything that could be done in a class can be done also in the other, because the sub-level sets $S^c := \{x : S(x) < c\}$ of a FGQQI function S are diffeomorphic to the sub-level sets of a suitable $FGQI$ function F.

We will exploit the topological properties of these functions in our existence theorem. It seems the first time FGQQI are applied in PDE theory.

Firstly we give the basic definition.

Definition 4.3.1. *We say that a function $S : \mathbb{R}^n \to \mathbb{R}$ is* quasi-quadratic *if there exists a non-degenerate quadratic form $\langle Au, u\rangle$ and a constant $K > 0$ such that*

$$\|S(u) - \langle Au, u\rangle\|_{C^1} = \|S(u) - \langle Au, u\rangle\| + \|S'(u) - 2Au\| \leqslant K$$

We recall the Palais-Smale Condition for a function $f : X \to \mathbb{R}$

Palais-Smale Condition: *Every sequence $\{x_i\}$ such that $|f(x_i)|$ being bounded and $|f'(x_i)| \to 0$, admits a converging subsequence.*

[5] in french, *fonctions génératrices quadratiques à l'infiní*

This condition is used when the domain of f is not compact, and for some reasons is needed the compactness of the critical levels. $FGQI$ and FGQ^2I will be proved to fulfill P-S, while the following elementary examples are not

Example. *1.* $e^{-\frac{x^2}{2}}$ *is not P-S. Consider the sequence* $n \mapsto x_n = n$.

2. sinx is not P-S. Consider $n \mapsto x_n = \frac{\pi}{2} + 2n\pi$.

Let us turn back to our quasi-quadratic functions.

PROPOSITION 3. *All the critical points of S are in a compact neighborhood of the origin.*

$$\frac{\partial S}{\partial \mu}(\bar{\mu}) = 0 \qquad \Rightarrow \qquad |A\bar{\mu}| \leqslant K \qquad \Rightarrow \qquad \bar{\mu} \in B\left(0, \frac{K}{\min \operatorname{Spec} A}\right).$$

A direct consequence is

PROPOSITION 4. *A quasi-quadratic function is Palais-Smale.*

Proof. Consider a P-S sequence $\{i \mapsto x_i\}$, *i.e.* $S(x_i)$ bounded, $S'(x_i) \to 0$. For i sufficiently large we have $|S'(x_i)| < K$, and by quasi-quadraticity, $|2Ax_i| < 2K$, then $x_i \in B\left(0, \frac{K}{\min \operatorname{Spec} A}\right)$. From a certain i onwards, the sequence is confined in a compact set, then there exists a converging subsequence. □

The main result of this section is the topological equivalence at infinity of the sublevel sets of the quasi-quadratic function and the ones of the quadratic form.

THEOREM 11. *For c sufficiently large,*

$$S^c \overset{\text{diffeo}}{\cong} Q^c.$$

Proof. Indeed,

$$\cdots \subsetneq S^c \subsetneq Q^{c+K} \subsetneq S^{c+2K} \subsetneq Q^{c+3K} \subsetneq \cdots$$

and certainly if c is large enough to overtake every critical value of S,

$$S^c \cong S^{c+K} \cong \cdots \cong S^{c+M} \qquad \text{and} \qquad Q^c \cong Q^{c+K} \cong \cdots \cong Q^{c+M}.$$

So we need only to prove

$$\boxed{S^c \cong Q^{c+M}}$$

for some $M > 0$.

It would be sufficient to find a differentiable function f such that $f^c = S^c$ e $f^{c+M} = Q^{c+M}$. For this purpose we need a differentiable Urysohn function $0 \leqslant \varphi \leqslant 1$ such that

$$\varphi(x) = \begin{cases} 0 \text{ se } S(x) \leqslant c, \\ 1 \text{ se } c + M \leqslant Q(x). \end{cases}$$

In that case we could set

$$f = (1 - \varphi)S + \varphi Q,$$

nevertheless we must exclude f to have intermediate critical points between c and $c + M$.

If we try to set up an estimate we find

$$\begin{aligned} |f'| &= \big|[(1-\varphi)S' + \varphi Q'] - [\varphi'(Q - S)]\big| \geqslant \\ &\geqslant \underbrace{|(1-\varphi)S' + \varphi Q'|}_{\cong |Q'| \pm K} + \underbrace{|\varphi'|}_{?} \underbrace{|Q - S|}_{\pm K}. \end{aligned}$$

This derivative is certainly non vanishing if $|\varphi'| \ll \frac{|Q'|}{K}$.

In what follows we will exibit a suitable Urysohn function φ and give also a lower bound for $|Q'|$ growing with c. □

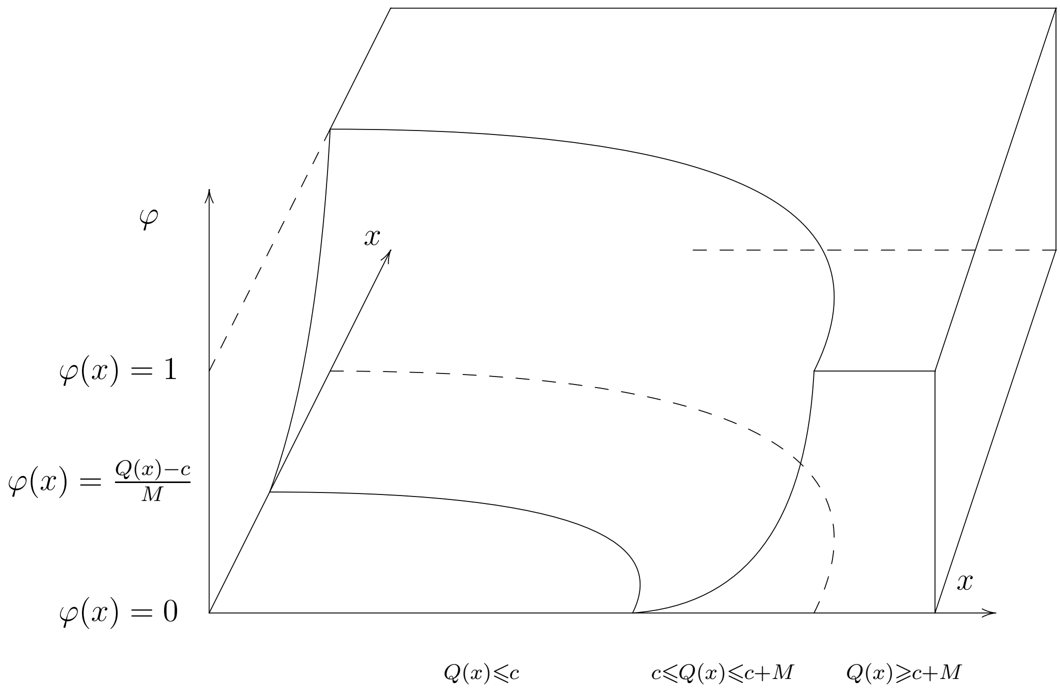
Being $S^c \subsetneqq Q^{c+K}$, it is sufficient that φ be 0 on Q^{c+K} and 1 on the complement of Q^{c+M}. Supposing M a lot larger than K, we will limit us to prove the existence of φ for Q^c and Q^{c+M}.

We will consider first a continuous version of φ, and then deform it to a differentiable one.

Lemma 3. *There exists a continuous Urysohn function φ which is 0 on S^c and 1 on the complement of Q^{c+M}.*

Proof. If we look at Q itself we see that it grows from c to $c + M$ in the desired sets, then it suffices to subtract c and divide by M to obtain the continuous φ:

$$\varphi(x) = \begin{cases} 0 & \text{if } Q(x) \leqslant c, \\ \frac{Q(x)-c}{M} & \text{if } c \leqslant Q(x) \leqslant c + M, \\ 1 & \text{if } c + M \leqslant Q(x). \end{cases}$$



Clearly, where the derivative exists, if we set $M \gg K$, we have $|\varphi'| = \frac{|Q'(x)|}{M} \ll \frac{|Q'(x)|}{K}$.

$\square$

In the next lemma we build up the regularization of this φ, maintaining the desired properties.

LEMMA 4. *Let $Q(x)$ a differentiable function without critical values between c and $c+M$. Then for every $\varepsilon > 0$ there always exists a differentiable function φ such that*

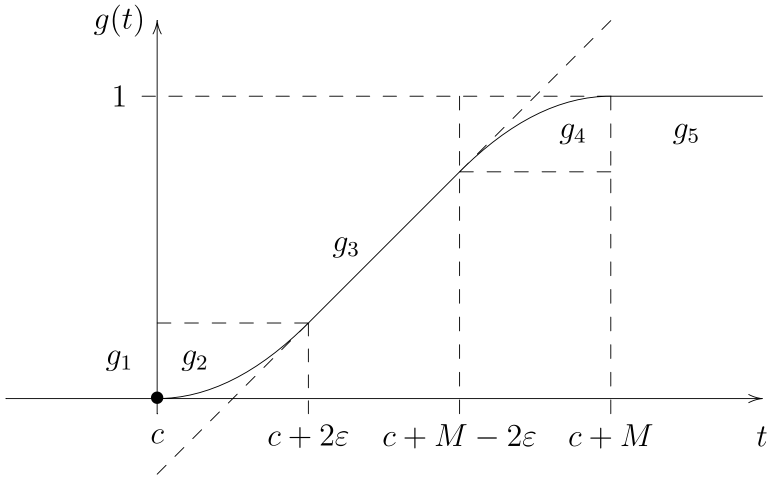
$$\varphi(x) = \begin{cases} 0 & \text{if } Q(x) \leqslant c, \\ 1 & \text{if } c + M \leqslant Q(x), \end{cases}$$

and furthermore that $|\varphi'(x)| \leqslant \frac{|Q'(x)|}{M-\varepsilon}$ for every x.

Proof. The idea consists in growing from 0 to 1 moving trough the integral paths of the vector field $\nabla Q(x)$. Taking thus a function $\varphi(x) = g(Q(x))$ does not alter the trajectories of the vector field: $\nabla(g \circ Q) = g' \cdot \nabla Q$.

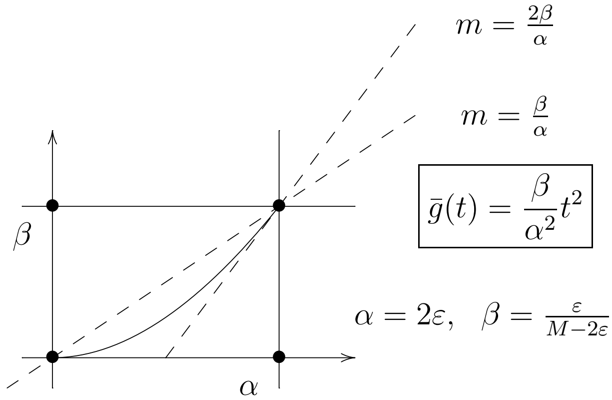
We only need to modulate the behavior of g near the angular points of the continuous version of φ seen before. We proceed as follows:

$$g(t) = \begin{cases} g_1 \equiv 0 & \text{if } \ t \leqslant c, \\ g_3 \equiv \frac{t-c-\varepsilon}{M-2\varepsilon} & \text{if } c + 2\varepsilon < t < c + M - 2\varepsilon, \\ g_5 \equiv 1 & \text{if } c + M \leqslant t. \end{cases}$$



It remains to define g_2 and g_4 in order to glue smootly the branches of g above. The conditions to be imposed to g_2 are the following

$$\begin{cases} g_2(c) = 0, \\ g_2(c+2\varepsilon) = \frac{\varepsilon}{M-2\varepsilon}, \\ g_2'(c) = 0, \\ g_2'(c+2\varepsilon) = \frac{1}{M-2\varepsilon}. \end{cases}$$



Trying to impose these conditions to a polynomial of degree two, we find

$$g_2(t) := \frac{1}{4\varepsilon(M-2\varepsilon)}(t-c)^2$$

Analogously it is easy to find

$$g_4 := c + M - \frac{1}{4\varepsilon(M-2\varepsilon)}(t-c-M)^2$$

and then, collecting all the pieces,

$$\varphi(x) = g(Q(x)) \qquad \Rightarrow \qquad \varphi'(x) = g'(Q(x)) \cdot Q'(x)$$

for which holds

$$\varphi'(x) = \begin{cases} g_1' \cdot Q' \equiv 0 & \text{se } \; Q(x) \leqslant c, \\ g_2' \cdot Q' = \frac{Q-c}{2\varepsilon(M-2\varepsilon)} \cdot Q' & \text{se } c < Q(x) < c + 2\varepsilon, \\ g_3' \cdot Q' = \frac{Q'}{M-2\varepsilon} & \text{se } c + 2\varepsilon < Q(x) < c + M - 2\varepsilon, \\ g_4' \cdot Q' = \frac{c+M-Q}{2\varepsilon(M-2\varepsilon)} \cdot Q' & \text{se } c + M - 2\varepsilon < Q(x) < c + M, \\ g_5' \cdot Q' \equiv 0 & \text{se } c + M \leqslant Q(x). \end{cases}$$

thus

$$|\varphi'(x)| = |g'(Q(x))| \cdot |Q'(x)| \leqslant \frac{1}{(M-2\varepsilon)} \cdot |Q'(x)|$$

as wanted. □

For sake of completeness, we show that $|Q'(x)|$ becomes arbitrarily large, for sufficiently large c.

In fact, if the quadratic form is undefined, $|Q'|$ is not bounded from above in a level set $Q = c$. (e.g. if $Q(x_1, x_2) = -x_1^2 + x_2^2 = c$, then $\forall M > 0, Q(M, \sqrt{c + M^2}) = c$ and $|Q'| \geqslant M$).

On the other hand, $|Q'|$ assumes its minimum in a compact subset of the level set (if $|x| > M$, then $|Q'(x)| > \min_i |\lambda_i| \cdot M$). It is easy to verify that the critical values are assumed on the eigenvectors of Q, thus the minimum of $|Q'|$ is reached by an eigenvector of the smallest in magnitude eigenvalue, say $\min |Q'| = |Q'(w_k)| = |\lambda_k w_k|$. But if $Q(w_k) = \lambda_k |v_k|^2 = c$, then $\min |Q'| = \sqrt{c \cdot \lambda_k}$, *i.e.* the minimum value of the gradient of Q is proportional to the square root of c.

If we have a polynomial of degree two instead of the quadratic form, it is easy to reduce the problem to the one considered before by means of a shift of coordinates. Indeed, if

$$P(x) = \langle Ax, x \rangle + ax + b,$$

setting $x = y + v$ obtains

$$\begin{aligned} \tilde{P}(y) = P(y+v) &= \langle A(y+v), y+v \rangle + ay + av + b = \\ &= \langle Ay, y \rangle + 2\langle Av, y \rangle + ay + \langle Av, v \rangle + av + b = \langle Ay, y \rangle + cost. \end{aligned}$$

if $2\langle Av, y \rangle + ay = 0$, for all y. Thus it suffices that

$$v = -\frac{1}{2} A^{-1} a.$$

Clearly the geometry and the topology of the sub level sets is not affected by adding constants.

4.3.1 Existence of a critical point for S

For our purposes it is important to show the existence of at least one critical point for S.

(i) Trivially, Q has a Morse critical point (0!) with Morse index equal to $q = \operatorname{sgn} Q$.

(ii) It is not difficult to prove that

$$H^k(Q^c, Q^{-c}) = \begin{cases} \mathbb{R}, & \text{if } k = q, \\ 0, & \text{if } k \neq q. \end{cases} \tag{4.4}$$

(see, *e.g.* , [God69], page 188.)

(iii) Plainly, for $c \gg 0$ suitably large,

$$H^\star(S^c, S^{-c}) \cong H^\star(Q^c, Q^{-c}) \neq 0.$$

(iv) There must be at least a critical level for S between $-c$ and c, because, if not, it must be $S^c \overset{\text{diffeo}}{\cong} S^{-c}$, (by Moser's paths technique, see section 4.2). In such a case, $H^\star(S^c, S^{-c}) = 0$: a contradiction.

4.4 Quasi-quadratic reduced variational principle for an elliptic Dirichlet problem.

Consider the nonlinearly perturbed elliptic operator

$$N(u) = -Lu - \lambda u - \varepsilon F(u).$$

Consider $u \in H = H_0^1(\Omega)$ and the following scalar products

$$(u, v) := \int_\Omega uvdx, \qquad \langle u, v\rangle := (-Lu, v) = \int_\Omega (-Lu) \cdot vdx.$$

We recall the variational formulation (Volterra-Vainberg) associated to our problem (see section 3.3):

$$\begin{aligned} J(u) = & \int_0^1 (N(tu), u)\, dt = \\ & = -\left(\int_0^1 (Ltu, u)\, dt + \lambda \int_0^1 (tu, u)\, dt\right) - \int_0^1 (\varepsilon F(tu), u)\, dt = \\ & \qquad =: J^{L+\lambda}(u) + J^F(u). \end{aligned}$$

We have seen that u can be represented by means of the eigenfunctions of the elliptic operator L:

$$u = \sum_{j=1}^{\infty} \hat{u}_j \langle u, u_j\rangle = g(v) = \sum_{j=1}^{\infty} \frac{1}{\lambda_j} v_j \hat{u}_j = \sum_{j=1}^{m} \frac{1}{\lambda_j} \mu_j \hat{u}_j + \sum_{j=m+1}^{\infty} \frac{1}{\lambda_j} \eta_j \hat{u}_j.$$

When the Dirichlet problem collapses to a finite dimensional problem by means of the fixed point technique, η (and by consequence u) become a function of the first m parameters μ_j:

$$u(m, \mu) = g(v) = g(\mu + \tilde{\eta}(m, \mu))$$

A reduced functional can readily be obtained,

$$W(\mu) := J(u(m, \mu)),$$

and W is equivalent to J in the sense explained in section 3.3. The reduced functional splits analogously to J,

$$W(\mu) = J(u(m, \mu)) = J^{L+\lambda}(u(\mu)) + J^F(u(\mu)) = W^{L+\lambda}(\mu) + W^F(\mu),$$

and, furthermore, a quadratic form in the μ_j's can be isolated:

$$\begin{aligned} J^{L+\lambda}(u) = J^{L+\lambda}(gv) &= \sum_{j=1}^{\infty} \frac{1}{\lambda_j} v_j^2 - \lambda \left(\sum_{j=1}^{\infty} \frac{1}{\lambda_j^2} v_j^2 \right) = \\ &= \sum_{j=1}^{m} \frac{\mu_j^2}{\lambda_j} + \sum_{j=m+1}^{\infty} \frac{\tilde{\eta}_j^2(\mu)}{\lambda_j} - \lambda \left(\sum_{j=1}^{m} \frac{\mu_j^2}{\lambda_j^2} + \sum_{j=m+1}^{\infty} \frac{\tilde{\eta}_j^2(\mu)}{\lambda_j^2} \right) = \\ &= \sum_{j=1}^{m} \frac{\lambda_j - \lambda}{\lambda_j^2} \mu_j^2 + \sum_{j=m+1}^{\infty} \frac{\lambda_j - \lambda}{\lambda_j^2} \tilde{\eta}_j^2(\mu) =: W_q^{L+\lambda}(\mu) + W_{nq}^{L+\lambda}(\mu). \end{aligned}$$

In the general case, there is no reason for $W_{nq}^{L+\lambda}(\mu)$ and $W^F(\mu)$ to be quadratic.

In what follows, we will put in evidence that $W_q^{L+\lambda}(\mu)$ is a non degenerate quadratic form and, under some hypothesis on F we are going to discuss, we will obtain $W_{nq}^{L+\lambda}(\mu)$ and $W^F(\mu)$ to be bounded together with their first derivatives. In such a case, the reduced functional will result in a quasi quadratic function, and we will obtain the existence of at least one critical point, on the basis of the topological considerations exposed in section 4.3.1.

Hypothesis . *Assume that the nonlinear functional $F : H \to H$ is a Nemitski operator, $F(u) = f \circ u$, associated to $f : \mathbb{R} \longrightarrow \mathbb{R}$. We will assume f, its derivative f' and one of its primitives $\bar{f}$ being bounded,* i.e.

$$\exists K, C > 0, \forall s \in \mathbb{R}, \qquad |f(s)|, \left|\bar{f}(s)\right| \leqslant K, \qquad |f'| \leqslant C.$$

See example 4.1.1 for nontrivial functions in this class.

4.4.1 $W_{nq}^{L+\lambda}(\mu)$ and $W^F(\mu)$ are bounded

Lemma 5.

$$\boxed{\left|W^F(\mu)\right| \leqslant \varepsilon K \text{ meas}(\Omega).}$$

Proof. Whenever $u(x) \neq 0$,

$$\int_0^1 f(tu(x))dt = \frac{1}{u(x)} \int_0^{u(x)} f(\tau)\, d\tau,$$

thus, denoting by $\Omega' = \left\{ x \in \Omega \middle| \, u(x) \neq 0 \right\}$,

$$\left|W^F\right| = \left|\left(\int_0^1 \varepsilon F(tu)dt, u\right)\right| = \left|\int_{\Omega'} \varepsilon \left(\frac{1}{u(x)} \int_0^{u(x)} f(\tau)d\tau \cdot u\right) dx\right| =$$
$$= \varepsilon \left|\int_{\Omega'} \left(\int_0^{u(x)} f(\tau)d\tau\right) dx\right| = \varepsilon \left|\int_{\Omega'} \bar{f}(u(x))dx\right| \leqslant$$
$$\leqslant \varepsilon K \text{ meas}(\Omega).$$

$\square$

LEMMA 6.
$$\boxed{\left|W_{nq}^{L+\lambda}(\mu)\right| \leqslant \varepsilon^2 K^2 \frac{\lambda_{m+1}}{(\lambda_{m+1} - \lambda)^2}.}$$

Proof.
$$W_{nq}^{L+\lambda}(\mu) = \sum_{j=m+1}^{\infty} \frac{\lambda_j - \lambda}{\lambda_j^2} \tilde{\eta}_j(m, \mu)^2$$

Being λ fixed, while m can be chosen arbitrarily large, it is non restrictive to suppose $\lambda \ll \lambda_j$, $\forall j > m$.

$$\left|W_{nq}^{L+\lambda}(\mu)\right| \leqslant \frac{1}{\lambda_{m+1}} \sum_{j=m+1}^{\infty} \left|\underbrace{\frac{\lambda_j - \lambda}{\lambda_j}}_{<1}\right| |\tilde{\eta}_j(m, \mu)|^2 \leqslant \frac{1}{\lambda_{m+1}} \|\tilde{\eta}\|^2 .$$

On the other hand, being $\tilde{\eta}$ obtained by means of the fixed point technique, it satisfies

$$\tilde{\eta} = (\mathbb{I} - \lambda g)^{-1} \varepsilon \mathbb{Q}_m F g(\mu + \tilde{\eta}).$$

We prove that $\|(\mathbb{I} - \lambda g)^{-1}\| < \frac{\lambda_{m+1}}{\lambda_{m+1} - \lambda}$. Indeed,

$$(\mathbb{I} - \lambda g)x = \left\{ \left(\mathbb{I} - \frac{\lambda}{\lambda_j} x_j\right) \right\}_{j=m+1}^{\infty}, \qquad \forall x \in \mathbb{Q}_m H.,$$
$$(\mathbb{I} - \lambda g)^{-1}x = \left\{ \left(1 - \frac{\lambda}{\lambda_j}\right)^{-1} x_j \right\}_{j=m+1}^{\infty} = \left\{ \frac{\lambda_j}{\lambda_j - \lambda} x_j \right\}_{j=m+1}^{\infty},$$

and if we suppose $\lambda \ll \lambda_m \leqslant \lambda_j$,

$$\frac{\lambda_{m+1}}{\lambda_{m+1} - \lambda} \geqslant \frac{\lambda_j}{\lambda_j - \lambda} \xrightarrow[j \to \infty]{} 1.$$

Thus,

$$\|\tilde{\eta}\| \leqslant \frac{\lambda_{m+1}}{\lambda_{m+1}-\lambda}\varepsilon \left\|F(g(\mu+\tilde{\eta}))\right\| \leqslant \frac{\lambda_{m+1}}{\lambda_{m+1}-\lambda}\varepsilon K,$$

as claimed. □

4.4.2 The first derivatives of $W_{nq}^{L+\lambda}(\mu)$ and $W^F(\mu)$ are bounded

Now it is necessary to check if also the difference between the first derivatives of W and the ones of the quadratic form $W_q^{L+\lambda}$ is bounded by a constant.

First we look for an estimate for $\tilde{\eta}'(\mu)$ and $u'(\mu)$. From the fixed point equation defining $\tilde{\eta}$ we have:

$$\frac{\partial\tilde{\eta}}{\partial\mu_h} = (\mathbb{I}-\lambda g)^{-1}\varepsilon\mathbb{Q}_m F'(u(\mu))\cdot\left(g\left(\frac{\partial\mu}{\partial\mu_h}+\frac{\partial\tilde{\eta}}{\partial\mu_h}\right)\right), \qquad h=1,\dots,m.$$

$$\begin{aligned}\left\|\frac{\partial\tilde{\eta}}{\partial\mu_h}\right\| &= \left\|(\mathbb{I}-\lambda g)^{-1}\varepsilon\mathbb{Q}_m F'(u(\mu))\cdot\left(g\left(\frac{\partial\mu}{\partial\mu_h}+\frac{\partial\tilde{\eta}}{\partial\mu_h}\right)\right)\right\| \leqslant \\ &\leqslant \left\|(\mathbb{I}-\lambda g)^{-1}\right\|\varepsilon\left\|F'\right\|\cdot\left(\frac{1}{\lambda_h}+\frac{1}{\lambda_{m+1}}\left\|\frac{\partial\tilde{\eta}}{\partial\mu_h}\right\|\right) \leqslant \\ &\leqslant \varepsilon C\frac{\lambda_{m+1}}{\lambda_{m+1}-\lambda}\left(\frac{1}{\lambda_1}+\frac{1}{\lambda_{m+1}}\left\|\frac{\partial\tilde{\eta}}{\partial\mu_h}\right\|\right),\end{aligned}$$

thus,

$$\left\|\frac{\partial\tilde{\eta}}{\partial\mu_h}\right\|\left(1-\frac{\varepsilon C}{\lambda_{m+1}-\lambda}\right) \leqslant \frac{\varepsilon C}{\lambda_1}\frac{\lambda_{m+1}}{\lambda_{m+1}-\lambda},$$

$$\boxed{\|\tilde{\eta}'\| \leqslant \frac{\varepsilon C}{\lambda_1}\cdot\frac{\lambda_{m+1}}{\lambda_{m+1}-\lambda}\cdot\frac{1}{1-\varepsilon C\frac{1}{\lambda_{m+1}-\lambda}} = \frac{\varepsilon C}{\lambda_1}\cdot\frac{\lambda_{m+1}}{\lambda_{m+1}-\lambda-\varepsilon C}.}$$

Note that the rightmost fraction is well defined and positive if m is sufficiently large.

On the other hand, we have by consequence an estimate for u',

$$\begin{gathered} u(\mu) = g(\mu+\tilde{\eta}(\mu)) \Rightarrow u' = g(\mu'+\tilde{\eta}'), \\ \|u'\| = \|g(\mu'+\tilde{\eta}')\| \leqslant \|g\|\left(1+\|\tilde{\eta}'\|\right), \\ \boxed{\|u'\| \leqslant \frac{1}{\lambda_1}(1+\|\tilde{\eta}'\|)} \end{gathered}$$

Estimates for $\frac{\partial}{\partial\mu}W_{nq}^{L+\lambda}(\mu)$ and $\frac{\partial}{\partial\mu}W^F(\mu)$

Lemma 7. $\left|\frac{\partial}{\partial\mu}W_{nq}^{L+\lambda}(\mu)\right| \leqslant$ *constant.*

Proof. By definition we have

$$\frac{\partial W_{nq}^{L+\lambda}}{\partial\mu} = \frac{\partial}{\partial\mu}\left(\sum_{j=m+1}^{\infty}\frac{\lambda_j-\lambda}{\lambda_j^2}\tilde{\eta}_j^2(\mu)\right) = 2\sum_{j=m+1}^{\infty}\left(\frac{\lambda_j-\lambda}{\lambda_j}\frac{\partial\tilde{\eta}_j(\mu)}{\partial\mu}\right)\cdot\left(\frac{\tilde{\eta}_j(\mu)}{\lambda_j}\right),$$

then

$$\begin{aligned}\left\|\frac{\partial}{\partial\mu}W_{nq}^{L+\lambda}(\mu)\right\| &\leqslant 2\left\|\left\{\frac{\lambda_j-\lambda}{\lambda_j}\frac{\partial\tilde{\eta}_j}{\partial\mu}\right\}_{m+1}^{\infty}\right\|\cdot\frac{\|\tilde{\eta}\|}{\lambda_{m+1}} \leqslant \\ &\leqslant 2\frac{1}{\lambda_{m+1}}\sup_{j>m}\left|\frac{\lambda_j-\lambda}{\lambda_j}\right|\|\tilde{\eta}'\|\cdot\|\tilde{\eta}\| \leqslant \frac{2}{\lambda_{m+1}}\|\tilde{\eta}'\|\cdot\|\tilde{\eta}\|.\end{aligned}$$

□

Lemma 8. $\left|\frac{\partial}{\partial\mu}W^F(\mu)\right| \leqslant$ *constant.*

Proof. First let us recall that

$$\left(\int_0^1 F(tu)dt, u\right) = \int_\Omega \bar{f}(u(x))dx.$$

Next, it is easy to check,

$$\begin{aligned}\left|\frac{\partial}{\partial\mu}W^F(\mu)\right| &= \left|\frac{\partial}{\partial\mu}\left(\int_0^1 F(tu(\mu))dt, u\right)\right| = \\ &= \left|\frac{\partial}{\partial\mu}\int_\Omega \bar{f}(u(x))dx\right| = \left|\int_\Omega f(u(x))\frac{\partial u}{\partial\mu}(x)dx\right| \leqslant \\ &\leqslant \int_\Omega |f|\,\|u'\|\,dx \leqslant \|u'\|\,K\ \mathrm{meas}(\Omega).\end{aligned}$$

□

4.4.3 Existence theorem

The reduced functional $W(\mu)$ is *quasi*-quadratic in μ, *i.e.*

$$\begin{aligned}&\boxed{\|W(\mu)-\langle Q\mu,\mu\rangle\|_{C^1}} = \left\|W(\mu)-W_q^{L+\lambda}(\mu)\right\|_{C^1} = \\ &= \left\|W_{nq}^{L+\lambda}(\mu)+W^F(\mu)\right\|_{C^1} \leqslant \boxed{\left(\begin{smallmatrix}\text{Something}\\ \text{depending}\\ \text{only on}\end{smallmatrix}\right)(\varepsilon, C, m, K, \mathrm{meas}(\Omega), \lambda_1, \lambda_{m+1}),}\end{aligned}$$

where

$$Q := \begin{pmatrix} \frac{\lambda-\lambda_1}{\lambda_1^2} & & & \mathbb{O} \\ & \frac{\lambda-\lambda_2}{\lambda_2^2} & & \\ & & \ddots & \\ \mathbb{O} & & & \frac{\lambda-\lambda_m}{\lambda_m^2} \end{pmatrix}.$$

Chapter 5

Numerical application of the finite parameters reduction

In this chapter we present and discuss an algorithm for constructing solutions for a class of nonlinear PDEs, devised implementing the techniques exposed in chapter 3.
We will generate two nontrivial solutions for a nonlinear Dirichlet Problem, and check the finite difference machinery to match with the theoretical forecasts, either in the speed of convergence and in approaching to the solutions of the original problem.
The result seems satisfactory, as we will try to put in evidence.

5.1 Analytical setting

Our investigation takes place in $H := H_0^1(\Omega, \mathbb{R}^k)$, where we consider a non linear perturbation F of an elliptic operator L. For sake of brevity we will limit ourselves to the following simple Dirichlet boundary value problems of the form

$$\begin{cases} -Lu = F(u), & \text{in } \Omega, \\ u = 0, & \text{on } \partial\Omega. \end{cases} \tag{5.1}$$

The crucial hypothesis we ought to adopt is an uniform Lipschitz estimate of the nonlinear operator F. More precisely, we will assume $F : H \to H$ to be a Nemitski operator, *i.e.* $F(u) := f \circ u$, where $f : \mathbb{R} \to \mathbb{R}$ is Lipschitz,

$$|f(s_1) - f(s_2)| \leqslant C\,|s_1 - s_2|\,.$$

The core of the algorithm consists in the spectral decomposition of H w.r.t. the eigenspaces of $-L$, and in the exploitation of the Green operator $g =$

$(-L)^{-1}$, *i.e.* $g: H \to H$, $g \circ (-L) = -L \circ g = id_H$. The problem is translated through g and successively decomposed into a finite and an infinite part by means of a suitable cut-off.

Here is an outline of these steps, first the spectral decomposition and the Green operator,

$$-L\hat{u}_j = \lambda_j \hat{u}_j, \qquad \langle \hat{u}_i, \hat{u}_j \rangle = \delta_{ij}, \qquad 0 = \lambda_0 < \lambda_1 \leqslant \lambda_2 \leqslant \dots. \tag{5.2}$$

$$g(v) = g\left(\sum_{j=1}^{+\infty} v_j \hat{u}_j\right) = \sum_{j=1}^{+\infty} v_j \frac{1}{\lambda_j} \hat{u}_j, \tag{5.3}$$

thus the cut-off of the space H,

$$v = \sum_{j=1}^{+\infty} v_j \hat{u}_j = \sum_{j=1}^{m} v_j \hat{u}_j + \sum_{j=m+1}^{+\infty} v_j \hat{u}_j \in H, \tag{5.4}$$

$$v = \mathbb{P}_m v + \mathbb{Q}_m v = \mu + \eta, \qquad H = \mathbb{P}_m H \oplus \mathbb{Q}_m H. \tag{5.5}$$

Here the crucial starting point: we are going to search solutions of (5.1) represented by the form: $u = g(v)$, for suitable $v \in H$,

$$\begin{array}{lcl} -Lu & = & F(u), \\ -L\left(g\left(v\right)\right) & = & F\left(g\left(v\right)\right), \quad v = \mu + \eta, \\ \mu + \eta & = & F(g(\mu + \eta)), \end{array} \tag{5.6}$$

so the problem is splitted into

$$\begin{array}{lcl} \eta & = & \mathbb{Q}_m F(g(\mu + \eta)) \quad \text{(infinite part)} \\ \\ \mu & = & \mathbb{P}_m F(g(\mu + \eta)) \quad \text{(finite part)} \end{array} \tag{5.7}$$

Next, we will show that the infinite part of the equation, for suitable fixed cut-off m, is uniquely solved, for every fixed finite part $\mu \in \mathbb{P}_m H$. Indeed the map

$$\begin{array}{lcl} \mathbb{Q}_m H & \longrightarrow & \mathbb{Q}_m H \\ \eta & \longmapsto & \mathbb{Q}_m F(g(\mu + \eta)), \end{array} \tag{CTR}$$

is contractive, provided m being suitably large; by involving the Lipschitz constant C of F and recalling the monotone character of the spectral sequence $\{\lambda_j\}$:

$$\begin{aligned} &\|\mathbb{Q}_m F(g(\mu + \eta_1)) - \mathbb{Q}_m F(g(\mu + \eta_2))\| \leqslant \\ &\qquad \leqslant C\,\|g(\mu + \eta_1) - g(\mu + \eta_2)\| \leqslant C\frac{1}{\lambda_{m+1}}\,\|\eta_1 - \eta_2\|\,, \end{aligned}$$

so we can choose $m \in \mathbb{N}$ large enough to have $\frac{C}{\lambda_{m+1}} < 1$, and the *unique fixed point* $\tilde{\eta}(\mu)$ of this contraction solves the (infinite part) of the equation (5.2). It can be easily proved that the fixed point $\tilde{\eta}(\mu)$ inherits the regularity of F, being expressible as the implicit function of an equation involving F:

$$\mathcal{F}(\mu, \eta) = 0, \qquad \mathcal{F}(\mu, \eta) := \mu - \mathbb{Q}_m F(g(\mu + \eta)).$$

By substituting $\tilde{\eta}(\mu)$ into the finite equation (finite part), we get a *finite dimensional problem*:

$$\mu = \mathbb{P}_m F\left(g\left(\mu + \tilde{\eta}\left(\mu\right)\right)\right), \qquad \mu \in \mathbb{R}^m. \tag{5.8}$$

Although it is finite, in general we have not an *a priori* control about existence and uniqueness; more precisely, we could find no solutions, or many solutions, and possible bifurcation phenomena could happen for increasing (Lipschitz constant C of) F.

Finally, in correspondence to every solution μ^* of (5.8) we obtain a solution of the original nonlinear Dirichlet problem:

$$u = g\left(\mu^* + \tilde{\eta}\left(\mu^*\right)\right).$$

In the following sections we will exhibit a finite difference implementation of these ideas.

5.2 Numerical implementation

The previously outlined procedure can be implemented in a numerical framework by substituting the differential operator of the PDE with its discretized version. Using a finite elements approach, denoting by T_h a generic discretization of Ω, formed by n nodes and N subdivisions with characteristic length h, the discrete elliptic operator reduces to a symmetric positive-definite matrix $-L_h$. The numerical solution vector $u_h \in H_h := \mathbb{R}^n$, is given by the solution of the system of the nonlinear algebraic equations:

$$-L_h(u_h) = F_h(u_h), \tag{5.9}$$

where F_h is the discretization of the nonlinear function operator F.

The corresponding symmetric eigenproblem can be written as:

$$-L_h u = \lambda u, \tag{5.10}$$

where

$$0 < \lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n,$$
$$u_1, u_2, \ldots, u_n,$$

are the real positive eigenvalues and the corresponding eigenvectors. Note that the eigenvalues and eigenvectors thus defined converge to the eigenvalues and eigenfunctions of the continuous problem (5.1) in the limit when $h \to 0$ and $n \to +\infty$ (see [GPP95]).

From (5.22), (5.23) and (5.24) we can observe what follows:

Remark 4. The first eigenvalues of the finite element problem (5.10) converge to fixed values when the nodal spacing of the discretization is made arbitrarily small, *i.e.* , when the dimension of $-L_h$ goes to infinity. Hence the leftmost eigenpairs of the eigenproblems converge to the same leftmost eigenspectrum;

Remark 5. The largest eigenvalues cannot converge and grow as n^2 when $h \to 0$, consistently with theory. This suggests that the highest frequencies are in essence inversely related to the magnitude of the discretization error, while the lowest frequencies represent the fundamental natural modes of the physical system described by the PDE (5.1).

In analogy to the continuous case (5.2), the discrete Green operator g_h of $-L_h$ can be easily written, w.r.t. the basis of H given by the eigenvectors of $-L_h$, as

$$g_h(u_k) = \left(-L_h^{-1}\right)(u_k) := \frac{1}{\lambda_k} u_k \qquad k = 1, \dots, n.$$

Since $-L_h$ is s.p.d., any vector $v \in H$ can be expressed as:

$$v = a_1 u_1 + \cdots + a_n u_n = Ua,$$

where U denotes the matrix whose columns are the eigenvectors u_k:

$$U := [u_1, \dots, u_n].$$

The Green operator applied to v gives:

$$g_h(v) = \frac{a_1}{\lambda_1} u_1 + \cdots + \frac{a_n}{\lambda_n} u_n = U\Lambda^{-1} a,$$

where Λ is the diagonal matrix of the eigenvalues (ordered accordingly to the corresponding eigenvectors).

The algorithm described in the previous section applies directly to the discretized problem, and proceeds as follows. For a given m, the vector space H is split into two subspaces $P_m H$ and $Q_m H$, where $P_m H \subseteq H$ is generated by the first m eigenvectors $u_1, \dots, u_m$, while $Q_m H \subseteq H$ is generated by $u_{m+1}, \dots, u_n$. Consequently, the projectors P_m and Q_m, which are the discrete counterparts of $\mathbb{P}_m$ and $\mathbb{Q}_m$ in (5.5), can be explicitly written by means of the two matrices:

$$V_1 := [u_1, \dots, u_m], \qquad V_2 := [u_{m+1}, \dots, u_n], \qquad [V_1, V_2] = U.$$

For every $v = \hat{\mu} + \hat{\eta} \in H$, we have:

$$\hat{\mu} := P_m v = V_1 V_1^T v = V_1 a', \qquad a' \in \mathbb{R}^m, \tag{5.11}$$
$$\hat{\eta} := Q_m v = V_2 V_2^T v = V_2 a'', \qquad a'' \in \mathbb{R}^{n-m}. \tag{5.12}$$

In summary, the discrete version of (5.6) becomes:

$$-L_h u = F_h(u), \tag{5.13}$$
$$-L_h\left(g_h(v)\right) = F_h\left(g_h(v)\right), \tag{5.14}$$
$$\hat{\mu} \oplus \hat{\eta} = F_h\left(g_h(\hat{\mu} + \hat{\eta})\right). \tag{5.15}$$

The numerical algorithm is thus formed by two finite dimensional fixed point iterations:

$$\hat{\eta} = Q_m F_h\left(g_h(\hat{\mu} + \hat{\eta})\right), \qquad \in \mathbb{R}^{n-m}, \tag{5.16}$$
$$\hat{\mu} = P_m F_h\left(g_h(\hat{\mu} + \hat{\eta})\right), \qquad \in \mathbb{R}^{m}, \tag{5.17}$$

with (5.16) satisfying:

$$\left\| Q_m F_h\left(g_h(\hat{\mu} + \hat{\eta}_1)\right) - Q_m F_h\left(g_h(\hat{\mu} + \hat{\eta}_2)\right) \right\| \leqslant \frac{C}{\lambda_{m+1}} \left\| \hat{\eta}_1 - \hat{\eta}_2 \right\|.$$

To prove the last assertion, we first note that F_h is Lipschitz whenever f is, *i.e.* , for every $u, \bar{u} \in H_h$,

$$\left\| F_h(u) - F_h(\bar{u}) \right\| = \left\| \begin{pmatrix} f(u_1) - f(\bar{u}_1) \\ \vdots \\ f(u_n) - f(\bar{u}_n) \end{pmatrix} \right\| =$$
$$= \left\| \begin{pmatrix} c_1(u_1 - \bar{u}_1) \\ \vdots \\ c_n(u_n - \bar{u}_n) \end{pmatrix} \right\| \leqslant C \left\| u - \bar{u} \right\|,$$

where C is the Lipschitz constant of f, while $c_1, \ldots, c_n$ are suitable positive numbers such that $c_j < C, \forall j$. The proof is completed by:

$$\left\| g_h(\hat{\mu} + \hat{\eta}_1) - g_h(\hat{\mu} + \hat{\eta}_2) \right\| = \left\| g_h(\hat{\eta}_1 - \hat{\eta}_2) \right\| \leqslant \frac{1}{\lambda_{m+1}} \left\| \hat{\eta}_1 - \hat{\eta}_2 \right\|.$$

The value of m is chosen so that the Lipschitz constant $\frac{C}{\lambda_{m+1}}$ becomes sufficiently smaller than 1 and guarantees a fast convergence of the fixed point iteration (5.16). The fixed point problem (5.17) can be solved either by Peano-Picard iterations or by a Newton-Raphson technique, but we do not have estimates on contractiveness.

Numerical performance does not allow for the full solution of the eigenproblem (5.10), which has been assumed so far. However, the ellipticity of $-L_h$ suggests that the discrete Green function tends to zero rapidly, according to the continuous case,but we are not aware of any useful theoretical estimate of this behavior. Thus in the following section we will try to evaluate experimentally, applying the proposed scheme to a sample problem, the minimum size of the eigensolution of (5.10) that gives a predefined accuracy.

5.2.1 Sample tests for the Peano-Picard procedure

Consider the one-dimensional Laplacian operator on the unit interval;

$$L := \Delta = \frac{\partial^2}{\partial x^2} \qquad \Omega := [0,1] \subseteq \mathbb{R}, \qquad \mathcal{H} := H_0^1([0,1],\mathbb{R}), \tag{5.18}$$

which, after the discretization with characteristic length $h = \frac{1}{n+1}$, and elimination from $-L_h$ of the two Dirichlet boundary conditions, will be represented by the well known tridiagonal matrix:

$$-L_h = \frac{1}{h^2}\begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & \ddots & \ddots & \\ & & \ddots & 2 & -1 \\ & & & -1 & 2 \end{pmatrix}. \tag{5.19}$$

which operates on the finite dimensional space $H := \mathbb{R}^n$. By very classical computations,the eigenpairs of (5.10) are found to be:

$$\lambda_k = 4(n+1)^2 \sin^2\left(\frac{k\pi}{2(n+1)}\right), \qquad k = 1,\dots,n, \tag{5.20}$$

$$u_k = \{u_{k,i}\} = \left\{\sqrt{\frac{2}{n+1}}\sin\left(\frac{k\pi}{n+1}i\right)\right\}, \qquad k = 1,\dots,n. \tag{5.21}$$

Since $\sin t \approx t$ as $t \to 0$, the leftmost eigenvalues converge to the quantities

$$\lambda_k(-L_h) \longrightarrow (\pi k)^2, \qquad k = 1,\dots,n, \tag{5.22}$$

while the eigenvectors behave as:

$$u_k \longrightarrow \sin(k\pi x), \qquad k = 1,\dots,n, \quad 0 \leqslant x \leqslant 1. \tag{5.23}$$

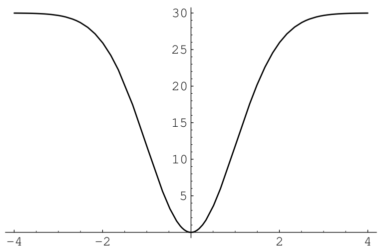


Figure 5.1: Graphic of $f(x) = 30\left(1 - e^{-\frac{x^2}{2}}\right)$

The largest eigenvalue is provided by (5.20) with $k = n$:

$$\lambda_n(-L_h) = \frac{4}{h^2}\sin^2\left(\frac{\pi}{2}\frac{n}{n+1}\right) \approx \frac{4}{h^2}. \tag{5.24}$$

Note that the eigenpairs obviously satisfy Remarks 4 and 5.

Our sample test considers the Nemitski operator F associated to the function

$$f : \mathbb{R} \to \mathbb{R}, \qquad f(x) = \alpha\left(1 - e^{-\frac{x^2}{2}}\right).$$

(See figure 5.1). Note that F is Lipschitz, with Lipschitz constant equal to

$$C = \sup|f'| = \alpha/\sqrt{e} \approx 18.1959(\alpha = 30).$$

Discretizing with $n = 80$ nodes, we can look for a suitable eigenvalue to perform the cutoff. The first candidates for contractive factors M are found to be

$$M = \frac{|f'|}{\lambda_1} \approx 1.84364, \qquad M = \frac{|f'|}{\lambda_2} \approx 0.460912, \qquad M = \frac{|f'|}{\lambda_3} \approx 0.204852, \qquad \ldots$$

We set $m = 2$, so that the contractive factor for the fixed point iteration given in (5.16) is 0.204852. The splitting is organized as follows:

$$H = \mathbb{R}^{80} = P_2H \oplus Q_2H,$$
$$v = \hat{\mu} + \hat{\eta} = a_1'u_1 + a_2'u_2 + a_1''u_3 + \cdots + a_{n-2}''u_n,$$

To show the convergence characteristics of the map (5.16), we perform several iterations by applying the Peano-Picard procedure. All the eigenpairs of (5.19) are employed in the calculations. The effects of using a much smaller number of eigenpairs ($k \ll n$) will be addressed in the next section.

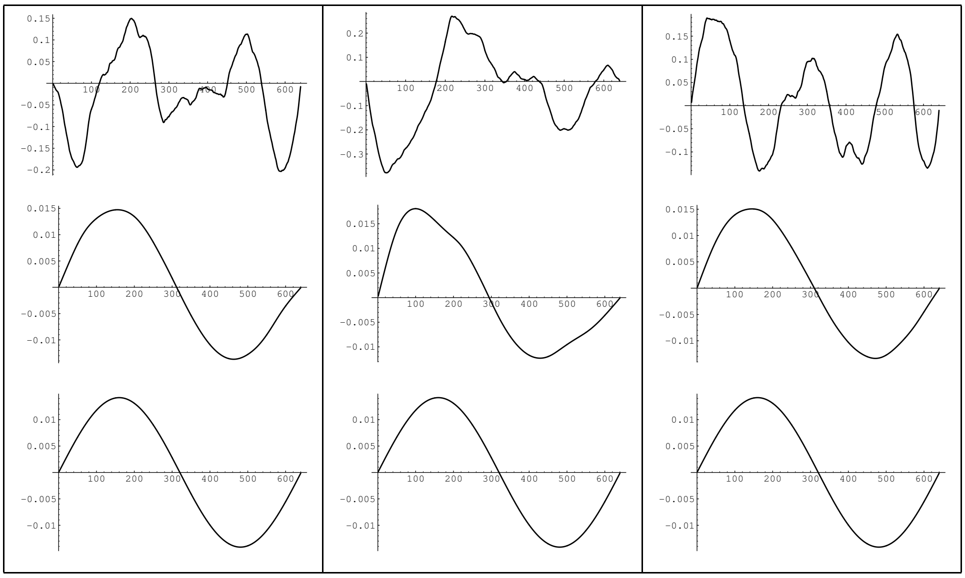


Figure 5.2: Convergence of the contractive map (5.16) for the same finite part $\mu^{(0)} = 10u_2$ and different randomly generated starting queues $\eta^{(0)}$

We fix as a first guess $\hat{\mu}^{(0)} = 500u_1 + 500u_2$ (u_1 and u_2 are the first two eigenvectors), and randomly generate $\hat{\eta}^{(0)}$. Denoting by $\hat{\eta}^{(j)}$ the j-th iterate of the contraction map, we expect the following estimate to be fulfilled,

$$\left\|\hat{\eta}^{(j+1)} - \hat{\eta}^{(j)}\right\| \leqslant 0.21 \left\|\hat{\eta}^{(j)} - \hat{\eta}^{(j-1)}\right\|,$$

The results are reported in Table 5.2.1 and illustrated in figure 5.2. After 20 iterations, the norm of the difference between two successive iterations, $\varepsilon_\eta^{(j)} = \left\|\hat{\eta}^{(j)} - \hat{\eta}^{(j-1)}\right\|$, becomes smaller than 10^{-15} with a contractive factor of approximately 0.16. Note that M is always smaller than the theoretical predictions and seems to stabilize around the 12^{th} iteration. After that, small oscillations appear due to round-off errors. Similar behavior is found when changing the initial guess $\mu^{(0)}$. The number of iterations changes slightly, while M always converges to 0.16. The solution of the full problem is obtained by solving

$$\hat{\mu} = P_m F_h \left(g_h \left(\hat{\mu} + \hat{\eta}(\hat{\mu})\right)\right). \tag{5.25}$$

Here we iterate by means again of the Peano-Picard procedure, though we do not possess any contraction result. Note that in this phase a more performing Newton-Raphson procedure could be employed. It is reported in the last section. Note that at each of the iterations of the $\hat{\mu}$–map we have to solve the $\hat{\eta}$–map. To ensure convergence of the latter we perform a fixed number of iterations equal to 20. This allows $\varepsilon_\eta^{(20)}$ to become always smaller than 10^{-15}.

j	$\varepsilon_\eta^{(j)}$	M	j	$\varepsilon_\eta^{(j)}$	M
1	8.21310×10^{1}	$\star$	11	6.33852×10^{-7}	0.169058
2	5.56597×10^{0}	0.067770	12	1.07158×10^{-7}	0.169058
3	9.43390×10^{-1}	0.169492	13	1.81159×10^{-8}	0.169059
4	1.60178×10^{-1}	0.169790	14	3.06254×10^{-9}	0.169053
5	2.71269×10^{-2}	0.169355	15	5.17750×10^{-10}	0.169059
6	4.58876×10^{-3}	0.169159	16	8.75862×10^{-11}	0.169167
7	7.75914×10^{-4}	0.169090	17	1.47106×10^{-11}	0.167956
8	1.31182×10^{-4}	0.169067	18	2.57688×10^{-12}	0.175171
9	2.21776×10^{-5}	0.169060	19	4.18135×10^{-13}	0.162264
10	3.74931×10^{-6}	0.169058	20	1.28513×10^{-13}	0.307348

Table 5.1: Convergence of the Peano-Picard iterations applied to (17) with $m = 2$, $\mu_0 = 500u_1 + 500u_2$ and using all the eigenspectrum of (20), $(n = 640, l = 640)$.

Table 5.2.1 reports the results of this problem. Note that numerically calculated contractive factor of this small scale $(m = 2)$ problem is rather small, achieving a value of about 0.22 (see table 5.2.1, 4^{th} column). As apparent from table 5.2.1, and from figure 5.3, the map converges spontaneously to

$$\hat{\mu} = 626.853u_1 - (7.67118 \times 10^{-14})u_2,$$

in 20 iterations. By means of the contraction map we can build the approximate solution of the discretized problem,

$$\bar{u} = g_h(\hat{\mu} + \hat{\eta}(\hat{\mu})),$$

and, by interpolation, construct a candidate solution for the analytical problem (5.1),

$$\mathcal{H} \ni \tilde{u}(x) := \text{interpolation}(\bar{u}).$$

Now we give an estimate of the goodness of the solution evaluating the residue function $E(x) := -\frac{\partial^2}{\partial x^2}\tilde{u}(x) - F(\tilde{u}(x))$. Next we evaluate the L^1 and the L^2 norm of the residue function. We check the theoretical convergence of the finite difference method does not degenerate as the number of subdivisions $n = 10, 20, \ldots, 320, 640$ grows. Theoretically, the norm of the error function should decrease proportionally to the square of the number of subdivisions. As apparent from table 5.2.1, doubling the subdivisions, the error seems asymptotically decrease by $\frac{1}{4}$, as expected. See also figure 5.4.

j	$\mu^{(j)}$	$\varepsilon_\mu^{(j)}$	M
1	$(230.589, -40.3486)$	6.7037×10^1	$\star$
2	$(283.236, -16.8716)$	5.7644×10^1	0.859881
3	$(369.349, -6.96820)$	8.6680×10^1	1.503710
4	$(482.325, -2.46623)$	1.1307×10^2	1.304410
5	$(573.516, -0.605729)$	9.1209×10^1	0.806688
6	$(612.605, -0.103476)$	3.9092×10^1	0.428595
7	$(623.499, -0.0150424)$	1.0895×10^1	0.278693
$\vdots$	$\vdots$	$\vdots$	$\vdots$
16	$(626.853, -2.61551 \times 10^{-10})$	1.7451×10^{-5}	0.225253
17	$(626.853, -3.56424 \times 10^{-11})$	3.9308×10^{-6}	0.225253
18	$(626.853, -4.93335 \times 10^{-12})$	8.8543×10^{-7}	0.225253
19	$(626.853, -5.30742 \times 10^{-13})$	1.9945×10^{-7}	0.225253
20	$(626.853, -7.67118 \times 10^{-14})$	4.4925×10^{-8}	0.225252

Table 5.2: Convergence behavior of the complete map starting with $m = 2$, $\mu^{(0)} = 100u_1 - 100u_2$ and using all the eigenspectrum of (20), $(n = 640, l = 640)$.

L^1–norm of the residue			L^2–norm of the residue		
n	n^{th}–residue	$\frac{n^{th}\text{–residue}}{2n^{th}\text{–residue}}$	n	n^{th}–residue	$\frac{n^{th}\text{–residue}}{2n^{th}\text{–residue}}$
10	9.07257×10^{-1}	$\star$	10	1.53448	$\star$
20	2.81267×10^{-1}	3.22561	20	7.778759×10^{-1}	1.97042
40	7.59899×10^{-2}	3.701369	40	2.47547×10^{-1}	3.14590
80	1.55879×10^{-2}	4.874926	80	5.12011×10^{-2}	4.83480
160	3.30348×10^{-3}	4.718633	160	9.88305×10^{-3}	5.18070
320	7.46062×10^{-4}	4.427889	320	1.94486×10^{-3}	5.08160
640	1.76359×10^{-4}	4.230356	640	4.02237×10^{-4}	4.83512

Table 5.3: Behavior of the residue of the approximated solution when the number of subdivisions n increases $(m = 2, l = n)$.

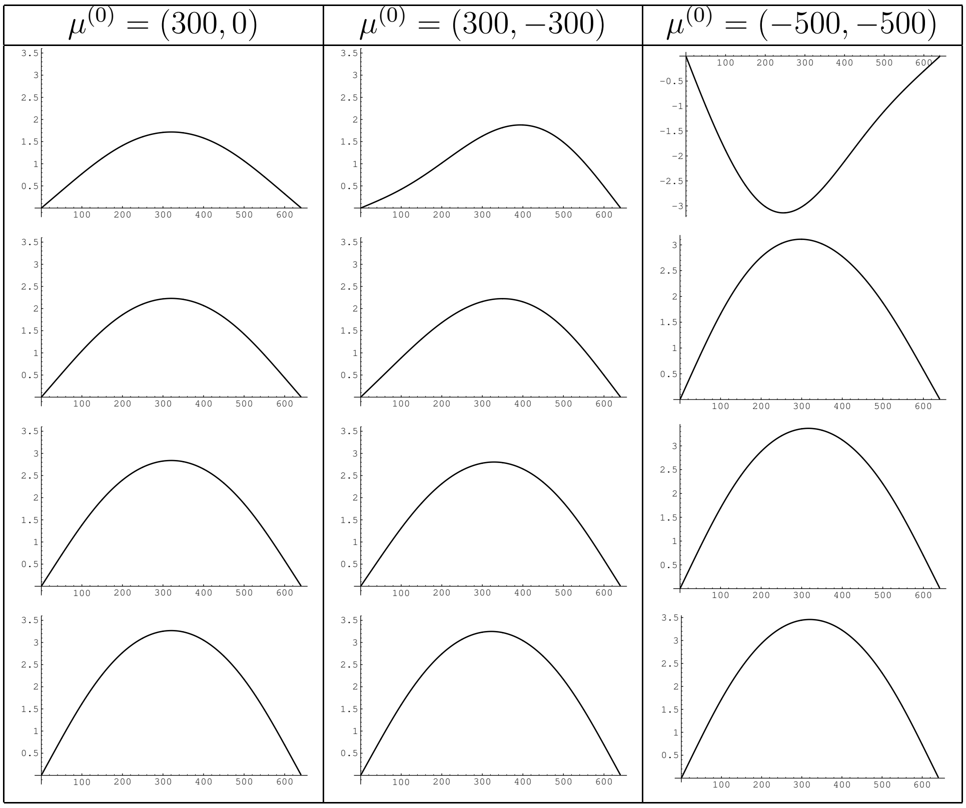


Figure 5.3: Peano-Picard iteration for the map (5.25) from different $\mu^{(0)}$ converging to the same solution $\hat{\mu} \approx (626.853, 0)$.

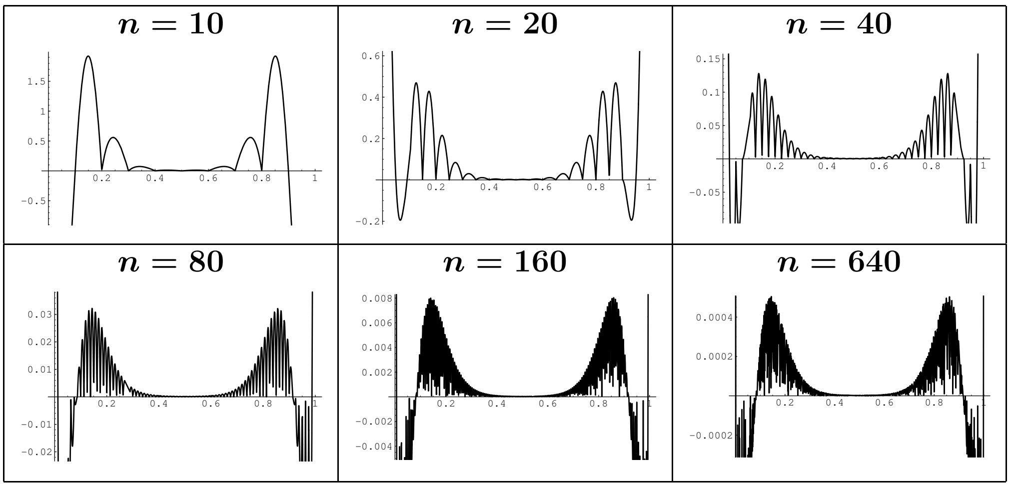


Figure 5.4: Graphic of the residue function $E(x) = -L\tilde{u}(x) - F(\tilde{u}(x))$.

Improving the performances

As claimed in the previous section, we are not able in general to determine the complete eigensolution of the elliptic operator $-L_h$. Nevertheless, theoretical considerations suggest that a not so big number of eigenvectors could suffice to evaluate a good approximated solution of the b.v.p. We consider the solution $\bar{u}$ so far determined employing $l = n = 640$ eigenvectors as the "exact" solution, and we try to approximate it progressively reducing the number $l \ll n$ of eigenvectors involved to generate the solution.

First we test the contractiveness of the generator of the queue $\hat{\eta}$. We will calculate and follow the behavior of $\varepsilon_\eta^{(j)} := \left\|\hat{\eta}^{(j+1)} - \hat{\eta}^{(j)}\right\|$ and the contractive factor $M_j = \frac{\varepsilon_\eta^{(j)}}{\varepsilon_\eta^{(j-1)}}$. Recall that theoretically the contractive factor should be $M \leqslant 0.20485$. See the table 5.2.1

Next we try the Peano-Picard procedure to obtain successive approximate solutions, increasing the number $l < 640$ of eigenvectors involved, and check the eventual rate of convergence to the more accurate, but more expensive, solution $\tilde{u}$, obtained employing all eigenvectors ($l = n = 640$).

See tables 5.2.1.

5.2.2 Newton-Raphson procedure

The solutions $\hat{\mu}$ of the reduced equation (5.25)

$$\hat{\mu} = P_m F_h(g_h(\hat{\mu} + \tilde{\eta}(\hat{\mu}))),$$

are, in other words, the fixed points of the map

$$\mu \mapsto PP(\mu) := P_m F_h(g_h(\hat{\mu} + \tilde{\eta}(\hat{\mu}))).$$

The Peano-Picard procedure consists in the iterated application of $PP(\cdot)$ from a tentative starting point μ_0.

$$\mu_1 = PP(\mu_0), \dots, \mu_i = PP^i(\mu_0), \dots$$

If a limit is reached then a solution of the original problem is found. Alternatively, the solutions of (5.25) can also be considered as the zeros of the map:

$$\begin{aligned} NR : \mathbb{R}^m &\to \mathbb{R}^m, \\ \mu &\mapsto NR(\mu) := \mu - P_m F_h(g_h(\hat{\mu} + \tilde{\eta}(\hat{\mu}))), \end{aligned}$$

which could be sought by means of the Newton-Raphson procedure. Namely, we look for a limit in the sequence of the iterated of the map:

$$\mu \mapsto -(J_{NR}(\mu))^{-1} NR(\mu).$$

$l = 160$		
j	$\varepsilon_\eta^{(j)}$	M
1	7.84170×10^{1}	$\star$
2	3.99984×10^{0}	0.051007
3	2.89345×10^{-1}	0.072339
4	2.34318×10^{-2}	0.080982
5	2.18111×10^{-3}	0.093083
6	2.22913×10^{-4}	0.102202
7	2.37485×10^{-5}	0.106537
8	2.56871×10^{-6}	0.108163
9	2.79277×10^{-7}	0.108722
10	3.04167×10^{-8}	0.108912
11	3.31465×10^{-9}	0.108975
12	3.61567×10^{-10}	0.109081
13	3.94282×10^{-11}	0.109048
14	4.31287×10^{-12}	0.109385
15	3.79512×10^{-13}	0.111259

$l = 40$		
j	$\varepsilon_\eta^{(j)}$	M
1	6.99369×10^{1}	$\star$
2	3.90643×10^{0}	0.055857
3	2.76129×10^{-1}	0.070686
4	2.17559×10^{-2}	0.078789
5	1.98066×10^{-3}	0.091040
6	2.00151×10^{-4}	0.101052
7	2.12306×10^{-5}	0.106073
8	2.29291×10^{-6}	0.108000
9	2.49164×10^{-7}	0.108667
10	2.71326×10^{-8}	0.108894
11	2.95671×10^{-9}	0.108973
12	3.22190×10^{-10}	0.108969
13	3.51193×10^{-11}	0.109002
14	3.89529×10^{-12}	0.110916
15	4.26181×10^{-13}	0.109409

$l = 10$		
j	$\varepsilon_\eta^{(j)}$	M
1	6.63682×10^{1}	$\star$
2	4.00190×10^{0}	0.060298
3	2.79898×10^{-1}	0.069941
4	2.15908×10^{-2}	0.077138
5	1.93161×10^{-3}	0.089464
6	1.93531×10^{-4}	0.100192
7	2.04669×10^{-5}	0.105755
8	2.20834×10^{-6}	0.107898
9	2.39886×10^{-7}	0.108628
10	2.61154×10^{-8}	0.108866
11	2.84542×10^{-9}	0.108955
12	3.09856×10^{-10}	0.108896
13	3.37722×10^{-11}	0.108993
14	3.72282×10^{-12}	0.110233
15	3.92485×10^{-13}	0.105427

$l = 4$		
j	$\varepsilon_\eta^{(j)}$	M
1	5.69242×10^{1}	$\star$
2	3.07766×10^{0}	0.054066
3	1.81176×10^{-1}	0.058868
4	1.29883×10^{-2}	0.071689
5	1.13693×10^{-3}	0.087535
6	1.09579×10^{-4}	0.096382
7	1.08595×10^{-5}	0.099102
8	1.08213×10^{-6}	0.099648
9	1.07857×10^{-7}	0.099671
10	1.07438×10^{-8}	0.099611
11	1.06966×10^{-9}	0.099561
12	1.06517×10^{-10}	0.099580
13	1.06174×10^{-11}	0.099678
14	1.05226×10^{-12}	0.099107
15	1.35749×10^{-13}	0.129007

Table 5.4: Convergence of Peano-Picard iterations applied to (5.16) starting with $m = 2$, $\mu^{(0)} = 500u_1 + 500u_2$ and $(n = 640, l = 160, 40, 10, 4)$.

$\boldsymbol{l = 160}$			
j	$\mu^{(j)}$	$\mu^{(j)} - \bar{\mu}$	M_j
1	$(557.101, 115.658)$	135.063	$\star$
2	$(605.908, 21.1875)$	29.7917	0.220576
3	$(621.783, 3.16681)$	5.97733	0.200637
4	$(625.692, 0.443249)$	1.24227	0.207830
5	$(626.590, 0.0610481)$	0.269357	0.216826
6	$(626.793, 0.00837702)$	0.0597325	0.221759
7	$(626.839, 0.00114853)$	0.0133737	0.223893
8	$(626.849, 0.000157440)$	0.00300559	0.224738
9	$(626.851, 0.0000215808)$	0.000676438	0.225059
10	$(626.852, 2.95813454 \times 10^{-6})$	0.000152321	0.225180

$\boldsymbol{l = 40}$			
j	$\mu^{(j)}$	$\mu^{(j)} - \bar{\mu}$	M_j
1	$(557.101, 115.658)$	135.063	$\star$
2	$(605.908, 21.1875)$	29.7917	0.2205767
3	$(621.783, 3.16681)$	5.97733	0.2006370
4	$(625.692, 0.4432489)$	1.24227	0.2078307
5	$(626.590, 0.06104813)$	0.2693578	0.2168263
6	$(626.793, 0.008377023)$	0.05973271	0.2217596
7	$(626.839, 0.001148533)$	0.01337393	0.2238963
8	$(626.849, 0.0001574401)$	0.003005773	0.2247486
9	$(626.851, 0.00002158085)$	0.0006766204	0.2251069
10	$(626.852, 2.95813 \times 10^{-6})$	0.0001525030	0.2253893

$\boldsymbol{l = 4}$			
j	$\mu^{(j)}$	$\mu^{(j)} - \bar{\mu}$	M_j
1	$(557.082, 115.788)$	135.184	$\star$
2	$(605.986, 21.2413)$	29.7753	0.220257
3	$(621.865, 3.17323)$	5.91074	0.198510
4	$(625.757, 0.443697)$	1.18196	0.199969
5	$(626.646, 0.0610455)$	0.215283	0.182140
6	$(626.846, 0.00836796)$	0.0105572	0.0490387
7	$(626.890, 0.00114610)$	0.0384218	3.63938
8	$(626.901, 0.000156945)$	0.0484506	1.26101
9	$(626.903, 0.0000214909)$	0.0507005	1.04643
10	$(626.903, 2.94278 \times 10^{-6})$	0.0512045	1.00994

Table 5.5: Convergence of the Peano-Picard iterations applied to (5.25) with $m = 2$, $\mu_0 = 600u_1 + 0u_2$ and with $n = 640$ nodes and employing $l = 160, 40, 4$ eigenvectors.

The determination of the Jacobian

$$J_{NR}(\mu) = \left(\frac{\partial NR_i}{\partial \mu_j}(\mu) \right)_{i,j=1,\dots,m},$$

$$\begin{aligned} \frac{\partial NR_i}{\partial \mu_j}(\mu) = \delta_{ij} - \frac{\partial (F_h)_i}{\partial u_r} \cdot \left(\frac{1}{\lambda_r} \frac{\partial}{\partial \mu_j}(\mu + \tilde{\eta}(\mu)) \right) = \\ = \delta_{ij} - \frac{\partial (F_h)_i}{\partial u_j} \frac{1}{\lambda_j} - \sum_{r=3}^{k} \frac{\partial (F_h)_i}{\partial u_r} \cdot \frac{1}{\lambda_r} \frac{\partial \tilde{\eta}_r(\mu)}{\partial \mu_j}. \end{aligned} \quad (5.26)$$

is performed in a few steps.

First we calculate the Jacobian of $F_h(u)$ with respect to the eigenvectors coordinates, *i.e.* considering $u[u_1, \dots, u_k](x) = u_1 \hat{u}_1(x) + \dots + u_k \hat{u}_k(x)$. Thus,

$$\begin{aligned} \frac{\partial (F_h)_i}{\partial u_r}(u) = \frac{\partial}{\partial u_r} \left((F_h)_i (u_1 \hat{u}_1 + \dots + u_k \hat{u}_k) \right) = \\ = \frac{\partial}{\partial u_r} \left(\langle F_h(u_1 \hat{u}_1 + \dots + u_k \hat{u}_k), \hat{u}_i \rangle \right) = \\ = \frac{\partial}{\partial u_r} \sum_{l=1}^{n} f(u[u_1, \dots, u_k](x_l)) \hat{u}_i(x_l) = \\ = \sum_{l=1}^{n} \frac{df}{du}(u(x_l)) \frac{\partial u}{\partial u_r}(x_l) \hat{u}_i(x_l) = \\ = \sum_{l=1}^{n} f'(u(x_l)) \hat{u}_r(x_l) \hat{u}_i(x_l). \end{aligned} \quad (5.27)$$

Next, in order to determine the derivative of $\tilde{\eta}(\mu)$, we differentiate the defining equation,

$$\tilde{\eta}(\mu) = Q_m F_h(g_h(\mu + \tilde{\eta}(\mu)))$$

$$\frac{\partial \tilde{\eta}_s}{\partial \mu_j}(\mu) = \frac{\partial (F_h)_r}{\partial u_j} \frac{1}{\lambda_j} + \sum_{s=3}^{k} \frac{\partial (F_h)_r}{\partial u_s} \frac{1}{\lambda_r} \frac{\partial \tilde{\eta}_r}{\partial \mu_j}(\mu),$$

and solve the linear system:

$$\sum_{s=3}^{k} \left(\delta_{rs} - \frac{\partial (F_h)_r}{\partial u_s} \frac{1}{\lambda_r} \right) \frac{\partial \tilde{\eta}_r}{\partial \mu_j} = \frac{1}{\lambda_j} \frac{\partial (F_h)_r}{\partial u_j}.$$

$\boldsymbol{\mu_0 = (100, 0)}$			
j	μ_j	ε^j_μ	$\frac{\varepsilon^{j-1}_\mu}{(\varepsilon^j_\mu)^2}$
1	$(205.949, 1.66223 \times 10^{-15})$	105.95	$\star$
2	$(160.633, -1.08118 \times 10^{-15})$	45.316	0.00403693
3	$(159.585, -5.48825 \times 10^{-16})$	1.04871	0.000510684
4	$(159.582, 6.32719 \times 10^{-16})$	0.00283031	0.0025735
5	$(159.582, 2.17691 \times 10^{-15})$	2.10223×10^{-8}	0.00262429
L^2-Residue = 0.00103257			

$\boldsymbol{\mu_0 = (600, 0)}$			
j	μ_j	ε^j_μ	$\frac{\varepsilon^{j-1}_\mu}{(\varepsilon^j_\mu)^2}$
1	$(627.622, 1.59563 \times 10^{-15})$	27.6225	$\star$
2	$(626.853, -4.35704 \times 10^{-17})$	0.769375	0.00100835
3	$(626.852, 1.88864 \times 10^{-16})$	0.000522536	0.000882756
4	$(626.852, 1.73901 \times 10^{-15})$	2.41811×10^{-10}	0.000885613
L^2-Residue = 0.0205632			

Table 5.6: Solutions found starting from $\mu_0 = (100, 0)$ and $\mu_0 = (600, 0)$, with $n = 640$ subdivisions, employing $k = 32$ eigenvectors

Application of Newton-Raphson

Implementing the previous algorithm we obtained the same solution found by Peano-Picard, and also a new solution. (See figure 5.5). As apparent from the tables 5.6, the ratio $\frac{\varepsilon^j_\mu}{(\varepsilon^{j+1}_\mu)^2}$ converges, in agree with the theoretical prediction of Newton-Raphson algorithm. Furthermore, the new solution found $(159.582, 0)$, could not be reached by Peano-Picard. Indeed, tentative iterations of the previous procedure diverge from it, though slowly, as illustrated in table 5.2.2.

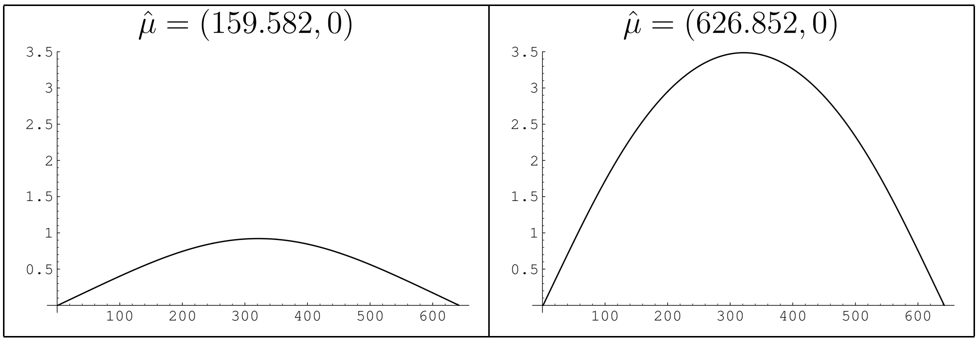


Figure 5.5: The two solutions for the map (5.25) obtained by Newton-Raphson procedure.

j	$\mu^{(j)}$	$\varepsilon_\mu^{(j)}$	M
1	$(159.582, -4.78656\,10^{-16})$	0.000136467	$\star$
2	$(159.582, -3.60152 \times 10^{-16})$	0.000230661	1.69024
3	$(159.581, 3.31963 \times 10^{-17})$	0.000389873	1.69024
$\vdots$	$\vdots$	$\vdots$	$\vdots$
15	$(159.064, 1.01548 \times 10^{-15})$	0.211715	1.68935
16	$(158.706, 3.33287 \times 10^{-16})$	0.357531	1.68874
17	$(158.103, 1.48207 \times 10^{-15})$	0.603402	1.68769
18	$(157.085, 1.93238 \times 10^{-15})$	1.01728	1.6859
19	$(155.373, 1.36685 \times 10^{-15})$	1.71191	1.68284
20	$(152.502, 8.66 \times 10^{-16})$	2.87176	1.67752
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Table 5.7: Divergence of Peano-Picard iterations applied to (5.16) near a solution found by means of Newton-Raphson $m = 2$, $\mu^{(0)} = 159.582u_1 + 0u_2$ and $(n = l = 640)$.

Bibliography

[AGZV88] V. I. Arnol′d, S. M. Guseĭn-Zade, and A. N. Varchenko. *Singularities of differentiable maps. Vol. II*, volume 83 of *Monographs in Mathematics*. Birkhäuser Boston Inc., Boston, MA, 1988. Monodromy and asymptotics of integrals, Translated from the Russian by Hugh Porteous, Translation revised by the authors and James Montaldi.

[Amb92] Antonio Ambrosetti. Critical points and nonlinear variational problems. *Mém. Soc. Math. France (N.S.)*, (49):139, 1992.

[AP93] Antonio Ambrosetti and Giovanni Prodi. *A primer of nonlinear analysis*, volume 34 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1993.

[Arn00] V. I. Arnold. Symplectic geometry and topology. *J. Math. Phys.*, 41(6):3307–3343, 2000.

[AZ80] Herbert Amann and Eduard Zehnder. Periodic solutions of asymptotically linear Hamiltonian systems. *Manuscripta Math.*, 32(1-2):149–189, 1980.

[Ben95] Vieri Benci. Introduzione alla teoria di morse e applicazioni. In *Equazioni Differenziali e Calcolo delle Variazioni (a cura di G. Buttazzo, A. Marino, M. K. V. Murthy)*, volume 39 of *Quaderni dell'Unione Matematica Italiana*, pages 11–109. Pitagora Editrice, Bologna, 1995.

[Cap03] M. Cappiello. Pseudodifferential parametrices of infinite order for SG-hyperbolic problems. *Rend. Sem. Mat. Univ. Politec. Torino*, 61(4):411–441 (2004), 2003.

[Car17] F Carlini. Ricerche sulla convergenza della serie che serve alla soluzione del problema di keplero. *Schumacher Astronomische Nachrichten*, 28,30, 1817.

[Car02] Franco Cardin. The global finite structure of generic envelope loci for Hamilton-Jacobi equations. *J. Math. Phys.*, 43(1):417–430, 2002.

[Car03] Franco Cardin. Global finite generating functions for field theory. In *Classical and quantum integrability (Warsaw, 2001)*, volume 59 of *Banach Center Publ.*, pages 133–142. Polish Acad. Sci., Warsaw, 2003.

[CGL04] Franco Cardin, Todor Gramtchev, and Alberto Lovison. Loss of Gevrey regularity for asymptotic optics. *preprint*, 2004.

[Cha84] Marc Chaperon. Une idée du type "géodésiques brisées" pour les systèmes hamiltoniens. *C. R. Acad. Sci. Paris Sér. I Math.*, 298(13):293–296, 1984.

[Cha91] Marc Chaperon. Lois de conservation et géométrie symplectique. *C. R. Acad. Sci. Paris Sér. I Math.*, 312(4):345–348, 1991.

[CL04] Franco Cardin and Alberto Lovison. Lack of critical points and exponentially faint illumination. *to appear in Meccanica*, 2004.

[CLOT03] Octav Cornea, Gregory Lupton, John Oprea, and Daniel Tanré. *Lusternik-Schnirelmann category*, volume 103 of *Mathematical Surveys and Monographs.* American Mathematical Society, Providence, RI, 2003.

[CPL04] Franco Cardin, Mario Putti, and Alberto Lovison. Implementation of finite parameters reductions. *submitted to Journal of Computational Physics*, 2004.

[CZ86] C. Conley and E. Zehnder. A global fixed point theorem for symplectic maps and subharmonic solutions of Hamiltonian equations on tori. In *Nonlinear functional analysis and its applications, Part 1 (Berkeley, Calif., 1983)*, volume 45 of *Proc. Sympos. Pure Math.*, pages 283–299. Amer. Math. Soc., Providence, RI, 1986.

[GMM01] M. Girardi, L. Mastroeni, and M. Matzeu. Existence and regularity results for non-negative solutions of some semilinear elliptic variational inequalities via mountain pass techniques. *Z. Anal. Anwendungen*, 20(4):845–857, 2001.

[God69] Claude Godbillon. *Géométrie différentielle et mécanique analytique.* Hermann, Paris, 1969.

[GP95] T. Gramchev and G. Popov. Nekhoroshev type estimates for billiard ball maps. *Ann. Inst. Fourier (Grenoble)*, 45(3):859–895, 1995.

[GPP95] Giuseppe Gambolati, Giorgio Pini, and Mario Putti. Nested iterations for symmetric eigenproblems. *SIAM J. Sci. Comput.*, 16(1):173–191, 1995.

[Gra87] Todor V. Gramchev. The stationary phase method in Gevrey classes and Fourier integral operators on ultradistributions. In *Partial differential equations (Warsaw, 1984)*, volume 19 of *Banach Center Publ.*, pages 101–112. PWN, Warsaw, 1987.

[Gra02] Todor Gramchev. Implicit function theorem in the gevrey classes. *preprint*, 2002.

[LM70] J.-L. Lions and E. Magenes. *Problèmes aux limites non homogènes et applications. Vol. 3.* Dunod, Paris, 1970. Travaux et Recherches Mathématiques, No. 20.

[LMZ03] Marcello Lucia, Paola Magrone, and Huan-Song Zhou. A Dirichlet problem with asymptotically linear and changing sign nonlinearity. *Rev. Mat. Complut.*, 16(2):465–481, 2003.

[Miy93] Masatake Miyake. An operator $L = aI - D_t^j D_x^{-j-\alpha} - D_t^{-j} D_x^{j+\alpha}$ and its nature in Gevrey functions. *Tsukuba J. Math.*, 17(1):85–98, 1993.

[Nir81] L. Nirenberg. Variational and topological methods in nonlinear problems. *Bull. Amer. Math. Soc. (N.S.)*, 4(3):267–302, 1981.

[Ram85] Jean-Pierre Ramis. Phénomène de Stokes et filtration Gevrey sur le groupe de Picard-Vessiot. *C. R. Acad. Sci. Paris Sér. I Math.*, 301(5):165–167, 1985.

[Rod93] Luigi Rodino. *Linear partial differential operators in Gevrey spaces.* World Scientific Publishing Co. Inc., River Edge, NJ, 1993.

[Sik86] Jean-Claude Sikorav. Sur les immersions lagrangiennes dans un fibré cotangent admettant une phase génératrice globale. *C. R. Acad. Sci. Paris Sér. I Math.*, 302(3):119–122, 1986.

[Sik87] Jean-Claude Sikorav. Problèmes d'intersections et de points fixes en géométrie hamiltonienne. *Comment. Math. Helv.*, 62(1):62–73, 1987.

[Thé99] David Théret. A complete proof of Viterbo's uniqueness theorem on generating functions. *Topology Appl.*, 96(3):249–266, 1999.

[Vit90] Claude Viterbo. Recent progress in periodic orbits of autonomous Hamiltonian systems and applications to symplectic geometry. In *Nonlinear functional analysis (Newark, NJ, 1987)*, volume 121 of *Lecture Notes in Pure and Appl. Math.*, pages 227–250. Dekker, New York, 1990.

Index