Università di Padova – Dipartimento di Matematica

Scuole di Dottorato in Matematica Pura e Matematica Computazionale

Seminario Dottorato 2012/13

Preface

This document offers a large overview of the nine months' schedule of Seminario Dottorato 2012/13. Our "Seminario Dottorato" (Graduate Seminar) is a double-aimed activity. At one hand, the speakers (usually Ph.D. students or post-docs, but sometimes also senior researchers) are invited to think how to communicate their own researches to a public of mathematically well-educated but not specialist people, by preserving both understandability and the flavour of a research report. At the same time, people in the audience enjoy a rare opportunity to get an accessible but also precise idea of what's going on in some mathematical research area that they might not know very well.
Let us take this opportunity to warmly thank the speakers once again, in particular for their nice agreement to write down these notes to leave a concrete footstep of their participation. We are also grateful to the collegues who helped us, through their advices and suggestions, in building an interesting and culturally complete program.

Padova, June 21st, 2013

Corrado Marastoni, Tiziano Vargiolu

Abstracts (from Seminario Dottorato's web page)

24 October 2012

Recollements from tilting modules

Alice PAVARIN (Padova, Dip. Mat., dott.)

In this talk I will expose some special and computable examples of recollements and tilting theory. I will start by presenting the concept of modules over a ring, complexes of modules and derived category of a ring. Then I will give examples of finitely generated tilting modules over a path algebra, that is, over a finite dimensional algebra generated by the paths of a finite, acyclic quiver. In the final part I will provide a glimpse over the general theory of recollements through the example of the endomorphism ring of the finitely generated tilting module $\mathbb{Q}+(\mathbb{Q}/\mathbb{Z})$ over the ring of integers $\mathbb{Z}$.

---

7 November 2012

FX derivatives and jump models

Giulio MIGLIETTA (Padova, Dip. Mat., dott.)

After a quick overview of the needed probabilistic concepts, I present the building blocks of modern mathematical finance with some emphasis on the foreign-exchange (FX) market. The anomalies recently experienced by the EURCHF (Euro - Swiss Franc) currency pair furnish the motivation for some simple extensions of the mainstream models used in the industry, namely models in which asset prices can exhibit discontinuities. I then analyze the effect of model choice (including jump models) on some contracts which are very liquid in the FX world, namely one-touch options and variance swaps.

---

21 November 2012

An Introduction to the Minimal Model Program

Gloria DELLA NOCE (Pavia, Dip. Mat., dott.)

The problem of reaching a birational classification of (smooth) complex projective varieties is classical in algebraic geometry. In dimension one, the problem comes out to be trivial and, in dimension 2, it was completely solved at the beginning of the XX century. We have to wait until the 1980's for an approach, due to S. Mori, to the 3-dimensional case, which is far more complicated and requires new tools. This is the beginning of the Minimal Model Program, which is nowadays

one of the most active areas of algebraic geometry. In this talk, after recalling some basic notions of algebraic geometry, I will introduce some of the tools of the Minimal Model Program, illustrating them through many examples.

---

5 December 2012

Geometric Surface Processing for Computer Graphics

Marco RUCCI (Padova, Dip. Mat., dott.)

The application of mathematical models based on Partial Differential Equations to image processing and computer graphics problems has been extremely successful over the past 20 years. A popular method to approach geometric processing and evolving surfaces has consisted of a Lagrangian setup where a surface is explicitly represented as a piecewise-linear mesh. In particular, geometric surface flows have been extensively used in mesh processing. In this talk we investigate curvature-driven surface evolutions and propose differential models and numerical solutions to Computer Graphics problems such as smoothing, remeshing, reconstruction, multiresolution surface representation and deformation.

---

19 December 2012

Geometric Quantization and Coherent States: a link between classical and quantum mechanics

Daniele FONTANARI (Padova, Dip. Mat., dott.)

A physical system can be mathematically described either in the quantum or in the classical framework, the latter being an accurate approximation only when the magnitude of the physical quantities that characterize the system is large compared to the value of the Planck constant h. It is usually accepted that the "right" classical description of a system should emerge from the quantum one as a "limit for h close to 0". The inverse problem is also of great interest: given a classical description of a system, is it possible to find a suitable quantum one? Such a procedure, when properly defined, is called quantization. In this talk I will give an introduction of the basic concepts and interpretations of (non-relativistic) Hamiltonian and quantum mechanics, then I will proceed with an overview of the apparatus of geometrical quantization. Finally, if time allows, I will show that coherent states naturally arise in the framework of geometric quantization and can be interpreted as quantum states that best approximate the classical ones. The talk will be self-contained and no prior knowledge on classical mechanics, quantum mechanics or geometric quantization is required.

---

16 January 2013

Power laws, possible explanations for their ubiquity

Alberto LOVISON (Padova, Dip. Mat.)

One of the emerging and most studied features of complex systems is the frequent occurrence of power law statistics, also referred to as scaling, or scale free statistics. Power laws have been firstly detected by Vilfredo Pareto in 1897 in the distribution of wealth among the italian population, whose key feature is the famous 80/20 law. Afterwards, power laws have been discovered in countless situations for diverse natural and artificial phenomena. Although during the two last decades several scientists strived for finding a universal and physically convincing explanation for the origin of power laws, this is still considered an open problem. In this talk we will give a brief survey on this topic and describe why multiplicative processes with a reflecting barrier could be a plausible explanation.

---

30 January 2013

Probabilistic Zeta function

Dung DUONG (Padova, Dip. Mat., dott.)

In this talk, I will introduce about the area of the probabilistic zeta function of finitely generated groups. I will start with some basic definitions, some examples, some basic properties and end up with some open problems.

---

6 February 2013

Zhanar TASPAGANBETOVA

Zhanar TASPAGANBETOVA (Padova, Dip. Mat. and Astana)

Weighted Hardy type inequalities restricted to the cones of monotone functions and sequences have been extensively studied in the last decades, especially in view of their applications in the estimation of maximal functions, in the theory of interpolation of operators and in the embedding theory of function spaces. In this talk, we introduce a Hardy type inequality on the cone of non-negative and non-increasing sequences. We give the statement and motivation of the problem. We describe the development and current status of the theory of Hardy type inequalities on the cones of monotone functions and sequences. Moreover, we also present some open problems.

---

13 February 2013

Portfolio optimization in defaultable markets under incomplete information

Giorgia CALLEGARO (Padova, Dip. Mat.)

We consider the problem of maximization of expected utility from terminal wealth in a market model that is driven by a possibly not fully observable factor process and that takes explicitly into account the possibility of default for the individual assets, as well as contagion (direct and information induced) among them. It is a multinomial model in discrete time that allows for an explicit solution. We discuss the solution within our defaultable and partial information setup, in particular we study its robustness. Numerical results are derived in the case of a log-utility function, and they can be analogously obtained for a power utility function.

---

27 February 2013

Ideas in finite group theory

Martino GARONZI (Padova, Dip. Mat., dott.)

In this talk I will present the basic ideas of finite group theory, which was invented by Evariste Galois to study polynomials. I will show how finite groups and polynomials interact. I will talk about the properties of the Symmetric group, which is one of the foundamental objects of the theory. I will illustrate the main elementary results of finite group theory (the theorems of Lagrange, Cauchy, Cayley and Sylow) with many examples and I will outline how problems are usually dealt with. I will introduce abelian, nilpotent, solvable and non-solvable groups with many examples. Then, I will present the specific problem I dealt with in my Ph.D thesis, the covering problem. Many examples will be given. If there is time, I will use not more than ten minutes to state some results I obtained in my thesis.

---

20 March 2013

Regularization techniques for linear inverse ill-posed problem

Silvia GAZZOLA (Padova, Dip. Mat., dott.)

Inverse problems are ubiquitous in many areas of science and engineering: they are typically modeled by Fredholm integral equations of the first kind and the available data are commonly affected by errors. Once discretized they give rise to ill-conditioned linear systems, often of huge dimensions: regularization consists in replacing the original system by a nearby problem with better numerical properties, in order to find a meaningful approximation of the exact solution. During this talk we will focus on problems regarding the restoration of images corrupted by blur and noise. We will review some standard regularization methods, both direct and iterative, and

we will introduce the most recent class of the Arnoldi-Tikhonov methods. The results of many numerical experiments will be shown, so to compare the different approaches and to contribute validating the newly-proposed strategies.

---

10 April 2013

Interpolation properties of Morrey-type spaces and their application

Diana DARBAYEVA K. (Padova, Dip. Mat. and Astana)

It is well known that in the theory of partial differential equations, alongside with weighted Lebesque spaces, Morrey-type spaces and their generalizations also play an important role. Our purpose is the introduction of some generalized Morrey-type spaces, which include classical Morrey spaces, and study their properties. We give examples. Moreover, in this talk we describe the interpolation theory of linear operators and consider some applications. We present a Marcinkiewicz-type interpolation theorem for generalized Morrey-type spaces. This theorem is then applied to obtain a Young-O'Neil-type inequality for the convolution operator in the generalized Morrey-type spaces, in particular in Morrey spaces. Moreover, we also present some open problems.

---

24 April 2013

Almost integrable Hamiltonian systems and the approach of the Perturbation Theory

Gabriella SCHIRINZI (Padova, Dip. Mat., dott.)

A Hamiltonian dynamical system is a system whose dynamic can be described by the Hamilton's equations. When we study such a system we try to find an explicit expression for the solutions of the equations, but this is not always possible, that is a Hamiltonian system is not always "integrable". Many of the most important physical systems are not integrable but their dynamic can be described by a Hamiltonian which is a small perturbation of an integrable one. This kind of functions, called "almost integrable Hamiltonians", are studied by the Hamiltonian Perturbation Theory. After recalling basic notions about Lagrangian and Hamiltonian mechanics, I will introduce almost integrable systems and give an introduction to the approach of the Perturbation Theory, stating the main results.

---

22 May 2013

A class of derivative-free nonmonotone algorithms for unconstrained optimization

Francesco RINALDI (Padova, Dip. Mat.)

Derivative-free methods represent a widely-used tool for solving problems where first order information is unavailable, unreliable, or impractical to obtain (for instance when the objective function is expensive to evaluate or somewhat noisy). In this talk, we first provide some basics on derivative-free optimization. Then, we present a class of derivative-free unconstrained minimization algorithms employing nonmonotone inexact linesearch techniques along a set of suitable search directions. In particular, we define globally convergent nonmonotone versions of some well-known derivative-free methods and we describe an algorithm combining coordinate rotations with approximate simplex gradients.

---

29 May 2013

Elliptic curves, an introduction and beyond

René SCHEIDER (Regensburg)

The talk provides a very elementary introduction to elliptic curves and their geometry, with the final aim of understanding the analytic geometry of the universal family with level N structure. As a broad audience shall be addressed, only few and basic prerequisites are required. In more detail, the plan is to cover the following topics: We explain how elliptic curves over a general field are defined and introduce their group law as well as the Weil Pairing. We then outline how these algebraic data express analytically if we work over the field of complex numbers. After this we are prepared to construct the universal elliptic curve with level N structure as a complex manifold.

---

19 June 2013

An introduction to Ramification Theory for Number Fields

Sophie MARQUES (Padova, Dip. Mat., and Bordeaux)

The question of prime decomposition in a finite extension of fields motivates classical ramification theory for field extensions. We propose to give a very little introduction to this deep subject which can also be extended to arithmetic geometry. After recalling some basic facts around finite field extensions, we will explain how to do arithmetic in number fields. This will permit us to define the ramification for number field and to give some criteria permitting to decide which primes are ramified or not. Finally, we will study the particular case of quadratic extensions in order to see how we can apply this theory.

Recollements from infinitely generated tilting modules

Alice Pavarin (*)

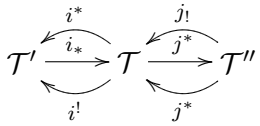
Abstract. The main aim of this talk is to introduce titling theory and in particular the equivalences induced by tilting modules at the level of derived categories of rings. This equivalences can be expressed by diagrams of functors between categories called recollements.

1 Introduction

Tilting theory owes its origin to Bernstein, Gelfand and Ponomarev (see [8]) who invented reflection functors (reformulated, some years later, by Auslander, Platzeck and Reiten in [2]). Tilting theory was born in the same philosophy as "Morita theory of equivalence", to simplify the study of the module category of an algebra A, by replacing A with another simpler algebra B. Indeed, if P is a projective generator of A-Mod, then $A\text{-Mod} \simeq \text{End}_A(P)\text{-Mod}$. Tilting modules can be viewed as the generalization of progenerators (finitely generated projective modules in the category of finitely generated modules A-mod). The difference between tilting theory and Morita theory is that, given a tilting module T over a finite dimensional algebra A and indicated with B its endomorphism algebra, the functors $\text{Hom}_A(T,-)$ and $T \otimes_B -$ do not provide and equivalence between A-mod and B-mod, but just between two pairs of subcategories (the torsion pairs $(\text{Ker}(\text{Hom}_A({}_AT,-), \text{Ker}(\text{Ext}^1_A({}_AT,-))$ and $(\text{Ker}(T\otimes_B -), \text{Ker}(\text{Tor}^B_1(T_B,-)))$. The main aim of this seminar is to present "short exact sequences" of categories induced by infinitely generated tilting modules. The study of infinitely generated tilting modules started for different reasons. In particular they simplify the study of some classes of finitely generated modules and they are involved also in some homological conjecture.
In the 80's works by several authors showed that a natural setting to interpret equivalences induced by classical tilting modules was that of derived categories. The first result in this direction was proved by Happel: if T is an n-tilting module over A, with endomorphism ring B, the pair $(G,H) := (T \overset{\mathbb{L}}{\otimes}_B -, \mathbb{R}\text{Hom}_A({}_AT,-))$ is no more an equivalence. Bazzoni,

(*)Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy; E-mail: alicepavarin@gmail.com. Seminar held on October 24th, 2012.

Mantese e Tonolo proved that H induces an equivalence with the quotient between $\mathcal{D}(A)$ and $\mathcal{D}(B)$ modulo the kernel of G. The equivalence proved in [6] can be expressed by a diagram of functors called recollement. A recollement of triangulated categories is a diagram



where the six functors involved are the derived version of Grothendieck's functors. In particular, they are paired in two adjoint triples, i_* is fully faithful and $\mathcal{T}''$ is equivalent to a Verdier quotient of $\mathcal{T}$ via j^* so that the straight arrows can be interpreted as an exact sequence of triangulated categories. The notion of recollements was introduced by Beilinson-Bernstein-Deligne [7] in a geometric context, where stratifications of varieties induce recollements of derived categories of constructible sheaves.

In this survey we will give some preliminary notions about categories, functors and module categories. We will present a simple example of a tilting module over finite dimensional path-algebra and we will briefly expose the construction of the derived category of a ring. Here we will show explicitly the result of [6] and an example of infinitely generated tilting module will be given.

2 Preliminaries

Definition 2.1 A category $\mathbb{C}$ consists of the following:

(a) a class of objects indicated by $\mathrm{Ob}(\mathbb{C})$.

(b) For each pair of objects X and Y, a set denoted by $\mathrm{Hom}_{\mathbb{C}}(X, Y)$ whose elements will be called *morphisms* or *maps* from $\mathcal{X}$ to $\mathcal{Y}$.

(c) For every triple A, B, C of objects, an associative composition law

$$\mathrm{Hom}_{\mathbb{C}}(A, B) \times \mathrm{Hom}_{\mathbb{C}}(B, C) \longrightarrow \mathrm{Hom}_{\mathbb{C}}(A, C);$$

the composite of the pair (f, g) will be denoted by gf.

(d) For every object A, a morphisms $\mathrm{Id}_A \in \mathrm{Hom}_{\mathbb{C}}(A, A)$, called the identity on A, such that, for each object B and for every morphisms $f \in \mathrm{Hom}_{\mathbb{C}}(A, B)$ and $g \in \mathrm{Hom}_{\mathbb{C}}(B, A)$, we have

$$f\mathrm{Id}_A = f \quad \text{and} \quad \mathrm{Id}_A g = g.$$

Given a category $\mathbb{C}$ and a morphism in it $f : M \to N$, f is called isomorphism if there exists a morphism $g : N \to M$ in $\mathbb{C}$, such that $fg = Id_N$ and $gf = Id_M$.

Example 1 It easy to verify that the following are examples of categories.

- Let us denote by Set the category where the objects are sets, and for each set X and Y, $\mathrm{Hom}_{Set}(X, Y)$ is the class of all maps between X and Y.
- Let us fix a camp k and denote with Vec_k the category where the objects are k-vector spaces and, for all X and Y in $\mathrm{Ob}(\mathrm{Vec}_k)$, $\mathrm{Hom}_{\mathrm{Vec}_k}(X, Y)$ is the class of all the k-linear maps between X and Y.
- Let us indicate with Top the category where the objects are topological spaces and morphisms are continuous maps.
- Let denote with Ab the category of abelian groups, where morphisms are group homomorphisms.

Given two categories $\mathbb{C}$ and $\mathcal{D}$, it is possible to define a "homomorphism" between them. We will call it "functor" between $\mathbb{C}$ and $\mathcal{D}$.

Definition 2.2 A functor F from a category $\mathbb{C}$ to a category $\mathcal{D}$ consists of the following:

(a) a map between $\mathrm{Ob}(\mathbb{C}) \longrightarrow \mathrm{Ob}(\mathcal{D})$. The image of A in $\mathrm{Ob}(\mathbb{C})$ via F is written $F(A)$.

(b) for every pair of objects M, N in $\mathbb{C}$, a map

$$\mathrm{Hom}_{\mathbb{C}}(M, N) \longrightarrow \mathrm{Hom}_{\mathcal{D}}(F(M), F(N))$$

such that $F(gf) = F(g)F(f)$ for each composable morphisms f and g, and $F(\mathrm{Id}_A) = \mathrm{Id}_{F(A)}$ for every object A in $\mathbb{C}$.

Let us present now two important examples of functors, that will be useful later on.

Example 2 Let V be a vector space over a field k.

(a) Let us define the tensor functor

$$V \otimes_k - : \mathrm{Vec}_k \longrightarrow \mathrm{Vec}_k$$

such that, for each vector space W and for each k-linear morphism $f : W \to U$,

$$W \mapsto V \otimes_k W, f \mapsto V \otimes_k f.$$

(b) Let us define the Hom functor as $\mathrm{Hom}_k(V, -) : \mathrm{Vec}_k \longrightarrow \mathrm{Ab}$ such that, for each vector space M and for each k-linear morphism $g : M \to N$

$$M \mapsto \mathrm{Hom}_k(M, V), g \mapsto \mathrm{Hom}_k(M, g).$$

Definition 2.3 A functor $F : \mathbb{C} \to \mathcal{D}$ is an equivalence if:

(a) for each object M in $\mathcal{D}$ there exists an object $N \in \mathbb{C}$ such that $F(N)$ is isomorphic to M.

(b) for each $A, B \in \mathbb{C}$, $\mathrm{Hom}_{\mathbb{C}}(A, B) \simeq \mathrm{Hom}_{\mathcal{D}}(F(A), F(B))$.

Let us note that if $F : \mathbb{C} \to \mathcal{D}$ is an equivalence, there exists a functor $G : \mathcal{D} \to \mathbb{C}$ such that $FG(D) \simeq D$, for all $D \in \mathcal{D}$, and $GF(C) \simeq C$ for all $C \in \mathbb{C}$. Moreover, let us suppose that $\mathbb{C}$ and $\mathcal{D}$ are additive categories (that is the the Hom spaces are abelian groups) then, $F(M) = 0$ if and only if $M = 0$, that is $\mathrm{Ker}(F) = \{0\}$.

3 The module category and tilting modules

Let us now introduce one of the most studied categories in algebra: the module category over a ring. Given a ring R, a module over R is a generalization of a vector space over a field.

Definition 3.1 Let R be a ring. A left R-module M is an abelian group such that there is a map, called *action*:

$$R \times M \longrightarrow M : \ (r, m) \mapsto r \cdot m$$

such that:

(a) for all $m, n \in Mnr \in R$, $r \cdot (m + n) = r \cdot m + r \cdot n$.

(b) For all $r, s \in R, m \in M$, $(rs) \cdot m = r \cdot (s \cdot m)$.

(c) For all $m \in M$, $1_R \cdot m = m$.

In the same way we can define the right R-modules and the R-bimodules, where the structure of left and right R-modules must be compatible.
To introduce the category of modules, we need to define what is a morphism between modules.

Definition 3.2 Let R be a ring and M and N be left R-modules. A morphism $f : M \to N$ is said to be R-linear if it is a morphism of abelian groups and, for all $r \in R, m \in M$, $f(r \cdot m) = r \cdot f(m)$.

Definition 3.3 Let R be a ring. We denote by R-Mod the category of left R-modules, where objects are the left R-modules and the morphisms are the R-linear morphisms. In the same way it can be defined the category of right R-module Mod-R.

Let us now introduce some "particular kinds" of modules that will be involved in the results presented in this survey.

Definition 3.4 A projective left module over a ring R, is a left module P, such that, for every surjection $f : M \to N$ in R-Mod and for every map $g : P \to N$, there is a map $h : P \to M$, such that $fh = g$.

Let us point out that a projective module P is direct summand of $R^{(I)}$ for some set I.

Definition 3.5 A short exact sequence in R-Mod, is a sequence of morphisms

$$0 \longrightarrow M \xrightarrow{f} N \xrightarrow{g} V \longrightarrow 0$$

with $M, N, V, f, g \in R$-Mod such that $\mathrm{Ker}(g) = \mathrm{Im}(f)$ and g is surjective and f is injective. We can regard M as a submodule of N and V as the quotient N/M. A *long exact sequence* is a sequence of B-modules and of morphisms of B-modules

$$0 \to M_1 \xrightarrow{f_1} M_2 \xrightarrow{f_2} .... \xrightarrow{f_n} M_n \to 0$$

such that, for all $0 \leq i \leq n$, there is a short exact sequence

$$\mathrm{Im}(f_i) = \mathrm{Ker}(f_{i+1}).$$

Let us now introduce tilting modules.

Definition 3.6 Let R be a ring and $n > 1$ an integer. An n-tilting module is a finitely generated R-module T such that:

- thre is a long exact sequence

  $$0 \to P_n \to P_{n-1} \to ... \to P_0 \to T$$

  with $P_n, ..., P_0$ finitely generated projective B-modules. The sequence above is called the *projective resolution* of T_B.

- There is a long exact sequence

  (1) $$0 \to \mathrm{Hom}_B(T, T) \to \mathrm{Hom}_B(P_0, T) \to ... \to \mathrm{Hom}_B(P_{n-1}, T) \to 0$$

  (that can be expressed saying that $\mathrm{Ext}^i(T, T) = 0$ for all $i > 0$). Usually the sequence above is given replacing T with $T^{(I)}$ for some set I, but this condition can be seen as a consequence of (1), since T is finitely generated.

- There exist modules T_0, T_1, ..., T_n that are direct summands of copies of T and a long exact sequence:

  $$0 \to R \to T_0 \to \cdots \to T_n \to 0.$$

Some of the simplest examples of tilting modules come from algebras of finite representation-type. In the following we will show an example of tilting module of projective dimension one. For the concept of path algebra and representations over it we refer to [3].

Example 3 Let us set B the path algebra of the quiver: $\underset{1}{\circ} \xleftarrow{a} \underset{2}{\circ} \xleftarrow{b} \underset{3}{\circ}$. Let us consider the module $T_B = 111 \oplus 100 \oplus 001$. Let us verify that T_B is a tilting module. Indeed, there is a short exact sequence

$$0 \to 011 \to 111 \oplus 100 \oplus 111 \to 001 \oplus 100 \oplus 111 \to 0$$

that is exactly the projective resolution of T_B. Moreover the sequence

$$0 \to \mathrm{Hom}_B(001 \oplus 100 \oplus 111, 010 \oplus 011) \to$$

$$\mathrm{Hom}_B(011, 001 \oplus 100 \oplus 111) \to \mathrm{Hom}_B(111 \oplus 100 \oplus 111, 001 \oplus 100 \oplus 111) \to 0$$

is exact. Finally, there is a short exact sequence

$$0 \to 100 \oplus 110 \oplus 111 \to 100 \oplus 111 \oplus 111 \to 001 \to 0,$$

where $100 \oplus 111 \oplus 111$ and 001 are direct summands of copies of T_B.

4 The derived category

As said in the introduction, the natural setting to study equivalences induced by tilting modules is the derived category. We will show briefly the main steps for the construction of the derived category and we will give just the idea of what a derived functor is.

Definition 4.1 A complex of left R-modules is an infinite sequence of left R-modules

$$M := ... \to M_n \overset{d_n^M}{\to} M_{n+1} \overset{d_{n+1}^M}{\to} M_{n+2} \to ....$$

such that the maps d_n^M are R-linear morphisms for each n and $d_{n+1}d_n = 0$ for all integer $n \in \mathbb{Z}$.
We define the n^{th} *homology* of M as the quotient

$$H^n(M) := \mathrm{Ker}(d_n^M)/\mathrm{Im}(d_{n-1}^M).$$

To introduce the category $\mathbb{C}(R)$ of complexes over R we need to define morphisms between complexes.

Definition 4.2 Let M, N two complexes of R-modules with differentials $d^M := (d_n^M)_{n\in\mathbb{Z}}$ and $d^N := (d_n^N)_{n\in\mathbb{Z}}$ respectively. A morphism of R-complexes or a *chain map* between M and N is a family of R-linear morphisms $f := (f_n)_{n\in\mathbb{Z}}$, with $f_n : M_n \to N_n$ such that $f_n d_{n-1}^M = d_{n-1}^N f_{n-1}$.

The category of complexes of R-modules $\mathbb{C}(R)$ is the category whose objects are the complexes over R and the morphisms are the chain maps.
Let us note that, for each integer n, the n^{th}-homology of M is still an R-module. Then every chain map induces a morphism between the sequences of the homologies ($\overline{f} := (\overline{f}_n)_{n\in\mathbb{Z}}$, with $\overline{f}_n : H^n(M) \to H^n(N)$). In particular we say that a chain map $f : M \to N$ is a *quasi-isomorphism* if $\overline{f}$ is an isomorphism. A complex P such that $H^n(P) = 0$ for every integer n is called *acyclic*. In particular an acyclic module is quasi isomorphic to the zero complex.

Definition 4.3 Let us consider the following equivalence relation in the sets of chain maps: we say that two chain maps f, g between the complexes M and N are homotopically

equivalent if there exists a family of R-linear maps $(h_n)_{n\in\mathbb{Z}}$ with $h_n : M_n \to N_{n-1}$ such that $g_n - f_n = d^N_{n-1}h_n - h_{n+1}d^M_n$ for every integer n.

To introduce the derived category we need an intermediate step, that is the homotopic category.

Definition 4.4 The *homotopic category* of the module category over a ring R, $\mathcal{H}(R)$, is the category where objects are complexes of R modules and morphisms are equivalence classes of chain maps via the homotopically equivalence. In particular a morphism in $\mathcal{H}(R)$ is zero if it is homotopic to the zero map. An object in $\mathcal{H}(R)$ is zero if its identity map is zero in $\mathcal{H}(R)$.

It is important to note that the class Σ of quasi-isomorphisms in $\mathcal{H}(B)$ is closed under composition and it has other "good properties" that made Σ a so-called multiplicative set. Then it is possible to consider the "localization" of $\mathcal{H}(B)$ at Σ. The byproduct of this localization is a "Verdier quotient" of $\mathcal{H}(B)$ and it is the so called derived category of B.

Definition 4.5 The derived category of a ring R, $\mathcal{D}(R)$, is the Verdier quotient of the homotopic category category $\mathcal{H}(R)$ via the class of quasi isomorphism Σ. In particular, in $\mathcal{D}(R)$ quasi isomorphism become invertible, and acyclic objects are isomorphic to zero.

Remark 1 Let us point out that if a functor $F : A\text{-Mod} \longrightarrow B\text{-Mod}$ (with A and B rings) has some "good properties", then it can be lift to a *derived functor* between derived categories. An example of such a functors are the tensor functor and the Hom functor.

5 Equivalences and recollements of derived categories

Let A be a ring. If M is a left A-module, then it is a right $\mathrm{End}_A(M)$-module. In particular the functors

$$\mathrm{Hom}_A(M,-) : A\text{-Mod} \to \mathrm{End}_A(M)\text{-Mod} \text{ and } M \otimes_{\mathrm{End}_A(M)} - : \mathrm{End}_A(M)\text{-Mod} \to A\text{-Mod}$$

are well define and it is possible to construct the respectively derived functors: $\mathbb{R}\mathrm{Hom}_A({}_AM,-)$ and $M \overset{\mathbb{L}}{\underset{\mathrm{End}_A(M)}{\otimes}} -$.

Theorem 5.1 *Let A be a finite dimensional algebra and ${}_AT$ a finitely generated tilting module over A. Set $B := \mathrm{End}_A({}_AT)$, then there is an equivalence:*

$$\mathcal{D}(A) \underset{\mathbb{R}\mathrm{Hom}_A({}_AT,-)}{\overset{T_B \overset{\mathbb{L}}{\underset{B}{\otimes}} -}{\rightleftarrows}} \mathcal{D}(B) \tag{0.1}$$

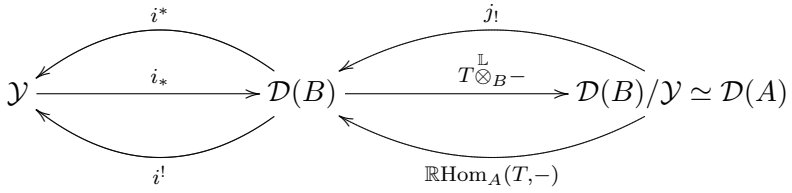
Remark 2 If we consider an infinitely generated titling module over a ring A wit endomorphism ring B, then

$$\mathcal{Y} := \mathrm{Ker}(T_B \overset{\mathbb{L}}{\otimes}_B -) \neq 0.$$

In particular $\mathcal{Y}$ is a subcategory of $\mathcal{D}(B)$ and there is an equivalence

$$\mathcal{D}(B)/\mathcal{Y} \simeq \mathcal{D}(A).$$

This result was proved by Bazzoni, Mantese and Tonolo in [6]. We can express this equivalence via a *recollement*, that is a diagram:



Recollements, as said in the introduction, can be seen as short exact sequences of categories, so we can see the first term as a subcategory of the central term and the third term as a quotient of the first two.

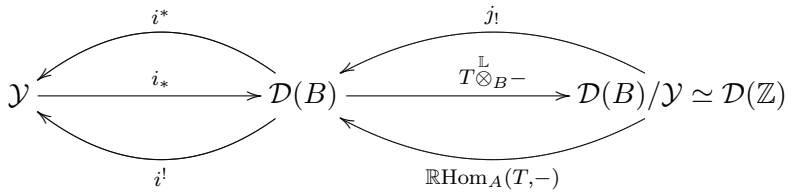
Example 4 A typical example of infinitely generated tilting module over the ring $\mathbb{Z}$ is given by

$$_{\mathbb{Z}}T = \mathbb{Q} \oplus \mathbb{Q}/\mathbb{Z}.$$

Its endomorphism ring is the matrix

$$B = \begin{pmatrix} \mathrm{End}_{\mathbb{Z}}(\mathbb{Q}/\mathbb{Z}) & 0 \\ \mathrm{Hom}_{\mathbb{Z}}(\mathbb{Q}, \mathbb{Q}/\mathbb{Z}) & \mathbb{Q} \end{pmatrix}$$

Then there is a recollement



References

[1] F. Borceux, "Handbook of Categorical Algebra: Basic Category Theory, Vol. 1". Cambridge University Press, 2008.

[2] M. Auslander, M.I. Platzeck, I. Reiten, *Coxeter functors without diagrams.* Trans. Amer. Math. Soc. 250 (1979), 1–46.

[3] I. Assem, D. Simson and A. Skowroński, "Elements of the representation theory of associative algebras. Vol. 1. Techniques of representation theory". London Mathematical Society Student Texts 65. Cambridge University Press, Cambridge, 2006.

[4] S. Bazzoni, *Equivalences induced by infinitely generated tilting modules.* Proc. Amer. Math. Soc. 138/2 (2009), 533–544.

[5] S. Brenner and M. C. R. Butler, *Generalizations of the Bernstein-Gel'fand-Ponomarev reflection functors.* In "Representation theory II", Proc. Second Internat. Conf., Carleton Univ., Ottawa, Ont., 1979, volume 832 of Lecture Notes in Math., Springer, Berlin (1980), 103–169.

[6] S. Bazzoni, F. Mantese and A. Tonolo, *Derived equivalence induced by infinitely generated n-tilting modules.* Proc. Amer. Math. Soc. 139/12 (2011) 4225–4234.

[7] A.A. Beilinson, J. Bernstein and P. Deligne, "Faisceaux pervers". Astérisque S.M.F. 100, 1982.

[8] I.N. Bernstein, I. M. Gel'fand and V. A. Ponomarev, *Coxeter functors, and Gabriel's theorem.* Uspehi Mat. Nauk 28 (1973), 19–33.

[9] E. Cline, B. Parshall and L. Scott, *Derived categories and Morita theory.* J. Algebra 104/2 (1986), 397–409.

[10] D. Happel, *On the derived category of a finite-dimensional algebra.* Comment. Math. Helv. 62/3 (1987), 339–389.

[11] Y. Miyashita, *Tilting modules of finite projective dimension.* Math. Z. 193/1 (1986), 113–146.

[12] D. Yang, *Recollements from generalized tilting.* Proc. Amer. Math. Soc. 140/1 (2012), 83–91.

Jump processes and FX derivatives

Giulio Miglietta [(*)]

Abstract. We will give a brief non-specialized introduction to Levy processes and how to construct them in a simple case starting from a Poisson random measure. Then we informally introduce some concepts in mathematical finance such as contingent claims and lognormal implied volatility and hint at the simple models in continuous-time (both with continuous and discontinuous sample paths). We conclude by an analysis of one-touch option prices under different underlying stochastic processes.

1 Levy Processes

1.1 Definition and some examples

A probability space $(\Omega, \mathscr{H}, \mathbb{P})$ is simply a measure space where the measure $\mathbb{P}$ is normalized to have total mass of one. If $(E, \mathscr{E})$ is a measurable space, an E-valued random variable is a measurable mapping from $(\Omega, \mathscr{H})$ to $(E, \mathscr{E})$. A stochastic process is a collection of random variables indexed by some arbitrary set, say $(X_t)_{t\in\mathbb{T}}$. Even if this is by no means the only case, in the following we will just be interested in real valued stochastic processes and the index set $\mathbb{T}$ will always be $\mathbb{R}_+$ which we will think of as the time.

Being a collection of random variables a (real) stochastic process can be thought as a (real) function of the two variables (ω, t). Almost inevitably when defining classes of processes we end up making requirements about $X_\cdot(\omega)$ (the so-called sample path) which have to hold for every ω and requirements about $X_t(\cdot)$ which have to hold for every t.

An notable class of processes is that of Levy processes, where we make a topological requirement on the sample paths and some independence and stationarity requirements on the increments. More specifically we have the following:

Definition 1 A Levy process is a stochastic process $(X_t)_{t>0}$ taking values in $\mathbb{R}$ with the following properties:

- $X_0 = 0$ and $X_\cdot$ is right-continuous with finite left-limit a.s.
- $\forall t$ and $h > 0$, $X_{t+h} - X_t$ is independent of $\mathcal{F}_t^X$ and has the same law as X_h.

[(*)] Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy; E-mail: giulio.miglietta@math.unipd.it. Seminar held on November 7th, 2012.

We now give some crucial examples of Levy processes by specifying the law of the increments and making some additional requirement on the topological properties of the sample paths. The first example is the fundamental Wiener process, which we define as follows:

Definition 2 A Wiener process is a Levy process X such that:

- $X_\cdot$ is continuous a.s.
- $\forall t$, X_t has the Gaussian distribution with zero mean and variance t.

The second fundamental example is the Poisson process:

Definition 3 A Poisson process with rate $\lambda \in \mathbb{R}_+$ is a Levy process X such that:

- $X_\cdot$ is counting function a.s.
- $\forall t$, X_t has the Poisson distribution with intensity λt.

It turns out that both the above definitions contains some redundancies. Roughly speaking, the conditions on the sample paths we imposed leave no choice for the law of the increments and vice-versa, but we will not dig any deeper into this topic.

In the following we will develop the machinery of Poisson random measures and then we will see how to use it to build (all kinds of) Levy processes.

1.2 Poisson Random Measures

In probability we can "randomize" any kind of object we want. In particular a real-valued random variable is the "randomization" of a real number. The "randomization" of a measure is the follwing:

Definition 4 Given a probability space $(\Omega, \mathscr{H}, \mathbb{P})$ and an arbitrary measurable space $(E, \mathscr{E})$, we say $M : \Omega \times \mathscr{E} \longrightarrow [0, \infty]$ is a random measure on $(E, \mathscr{E})$ if

- $M(\omega, \cdot)$ is a measure $\forall \omega \in \Omega$
- $M(\cdot, A)$ is $\mathscr{E}$-measurable $\forall A \in \mathscr{E}$

It is easy to show that if M is a random measure on $(E, \mathscr{E})$, then $\mu : A \longmapsto E(M(A))$ is a measure on E which is called the mean measure of M (under $\mathbb{P}$).

Also, to every $f \in \mathscr{E}_+$ we can associate a function $Mf : \Omega \longmapsto \mathbb{R}_+$ defined as

$$Mf : \omega \longmapsto \int_E M(\omega, dx) f(x) \tag{1}$$

which is readily seen to be $\mathscr{H}$-measurable. If we think of M as a collection of random variables $(M(\cdot, A))_{A \in \mathscr{E}}$ indexed by $\mathscr{E}$ we can define in a natural way the concept of law of a random measure. An extremely useful concept is that of Laplace functionals:

Definition 5 Let M be a random measure on $(E, \mathscr{E})$. We define the Laplace functional of M as the map $f \longmapsto E(e^{-Mf})$ from $\mathscr{E}_+$ into $\mathbb{R}_+$.

All familiar concepts for positive measures (finite, σ-finite, Σ-finite, diffuse i.e. with no atoms, purely atomic) get extended naturally to random measures if they hold $\mathbb{P}$-as. A key role in the sequel will be played by the following class of random measures:

Definition 6 Let M be a random measure on $(E, \mathscr{E})$. It is said to be additive if $M(A_1), M(A_2), \ldots, M(A_n)$ are independent for every choice of disjoint sets $A_1, A_2, \ldots, A_n$ in $\mathscr{E}$.

Definition 7 Let μ be a measure on $(E, \mathscr{E})$. An additive random measure M is said to be Poisson with intensity μ if $M(A) \sim Poi_{\mu(A)}$ for every $A \in \mathscr{E}$.

It can be readily seen that is M is Poisson on $(E, \mathscr{E})$ with intensity μ, then

$$E(e^{-Mf}) = \exp\left(\int_E \mu(dx)(e^{-f(x)} - 1)\right) \tag{2}$$

and the converse holds as well. The existence of Poisson random measures with finite intensity, μ, is handled in the following constructive way:

Theorem 1 *Let K be a Poisson random variable with intensity $\mu(E) < \infty$ and, independent of it, let $(X_n)_{n\in\mathbb{N}}$ be an iid collection of random variables on $(E, \mathscr{E})$ with law $\frac{\mu(\cdot)}{\mu(E)}$. Then*

$$M := \sum_{n=0}^{K} \delta_{X_n} \tag{3}$$

is a Poisson measure with intensity μ

For a construction for an arbitrary Σ-finite intensity it is enough to repeat the previous exercise for every finite measure constituting the intensity. A crucial result for the following is a characterization of the finiteness of Mf in terms of the intensity.

Theorem 2 *Let M be a Poisson random measure on $(E, \mathscr{E})$, with intensity μ and let $f \in \mathscr{E}_+$. Then Mf is finite $\mathbb{P}$-as if and only if $\int_E \mu(dx) \min(|f(x)|, 1) < \infty$*

By exploiting the construction used to prove the existence of Poisson random measures, we could show that a Poisson measure with Σ-finite intensity is a counting measure if and only if its intensity is diffuse. Thus we could think of Poisson measures with diffuse intensity as random countable sets on $(E, \mathscr{E})$, which are often referred to as "point processes".

1.3 Levy Processes from Poisson random measures

Pure drifts (i.e. $X_t = bt$ for some real b) are obviously Levy processes and the same is true for Wiener processes as well. Since linear combinations of independent Levy processes is

easily shown to stay Levy, we now show how to build purely discontinuous Levy processes using Poisson random measures. Once we will have that, we will be able to build a large class of Levy processes just by summing up processes of these three fundamental examples. At the very end of this subsection we will hint at the fundamental result (Levy-Ito decomposition) that guarantees that *every* Levy process is the "limit" of processes of this kind.

Our goal here is to construct some stochastic processes with discontinuous sample paths from Poisson random measures, but since this will be carried out almost exclusively in a pathwise manner, we present first some deterministic results. In the following we will make use of the standard relationships between bounded variation functions on $[0,T]$ and real measures on $\mathcal{B}([0,T])$, with the following definitions.

Definition 8 Let $X : [0,T] \longrightarrow \mathbb{R}$ have bounded variation and be right-continuous. Denote by dX the real measure it induces on $\mathcal{B}([0,T])$

- we say it is a pure-jump function if dX is purely atomic;
- if, in addition, $Ato(dX)$ is finite, we say it is a step-function;
- if, furthermore, $dX\{t\} = 1$ for every $t \in Ato(dX)$, we say it is a counting function.

Now, given a counting measure on $[0,T] \times \mathbb{R}$, call it J, with the following properties

- $Ato(J) \bigcap (\{0\} \times \mathbb{R}) = \emptyset$
- $Ato(J) \bigcap ([0,T] \times \{0\}) = \emptyset$
- $\forall (t_1, x_1), (t_2, x_2) \in Ato(J)$, we have that $t_1 \neq t_2$
- $\forall t$, the map $(u,x) \mapsto \mathbb{I}_{[0,t]}(u)x \in L^1(J)$

then we can construct a pure-jump function, X, as follows

$$X : t \mapsto \int_{[0,t]\times\mathbb{R}} J(du,dx)x = \sum_{i\in I} \mathbb{I}_{[0,t]}(t_i)x_i \tag{4}$$

where we indexed the atoms of J by $I \ni i \mapsto (t_i, x_i)$. It is clear that X has a jump of size x_i at t_i, but there is not necessarily a way to order the t_i's increasingly. Also, note that $J([t_0,t_1] \times B)$ is the number of jumps of size in B that occur between t_0 and t_1. This reasoning can be also turned around, so that we can construct a counting measure on $[0,T] \times \mathbb{R}$ starting from a pure jump function, X, in the following way:

$$Ato(J) = \{(t,x) \in [0,T] \times \mathbb{R} : t \in disc(X) \text{ and } x = \Delta X_t\} \tag{5}$$

So far everything was analytic and deterministic. Now we turn in into stochastic by allowing the measure J to be a random measure. It is quite natural to take it Poisson, in order to have a certain degree of tractability of the resulting process.

It can be verified that if J is a Poisson random measure on $[0,T] \times \mathbb{R}$ with intensity $Leb \times \lambda$, where λ satisfies $\int_{\mathbb{R}} \lambda(dx) \min(|x|, 1) < \infty$ and $\lambda\{0\} = 0$ then X can be defined as

above. The most crucial requirement is the last integrability condition which guarantees that X_t is finite a.s. for every t thanks to the results on Poisson measures we gave in the previous subsection.

In this stochastic setting, the number of jumps of size in $B \subset \mathbb{R}$ which occur between t_0 and t_1 is $J([t_0, t_1] \times B)$ which has law $Poi_{(t_1-t_0)\lambda(B)}$. In particular, if $\lambda(B)$ is infinite, that random variable is infinite a.s.[(†)]. Thus a fundamental distinction has to be made between the cases of a finite or infinite measure λ. Infinite measures give rise to processes that exhibit an infinite number of jumps on bounded interval, which are often called infinite activity processes. Finite measures give rise to much simpler processes whose sample paths are step-functions which are called finite activity processes.

By exploiting the results on the Laplace functional of Poisson random measure, we can compute the characteristic function of $X_{t+h} - X_t$ as follows:

$$E(e^{iu(X_{t+h}-X_t)}) = exp\left(h \int \lambda(dx)(e^{iux} - 1)\right) \tag{6}$$

This shows that the increments of X are stationary, whereas additivity of J shows their independence: we thus showed that X is indeed a Levy process.

The crucial assumption that $\int_{\mathbb{R}} \lambda(dx) \min(|x|, 1) < \infty$ had to be made to ensure that the function $(u, x) \longmapsto \mathbb{I}_{[0,t]}(u)x$ is J integrable $\forall t$ almost surely. If this hypothesis is not met, the process might even be not well defined. It turns out, though, that if we replace $\min(|x|, 1)$ with $\min(|x|^2, 1)$ we can still somehow define the process X. This construction is quite technical and we will not deal with it in detail, suffices it to say that X is defined as the limit (in some appropriate sense) for $\epsilon \to 0$ of the sequence $(X^\epsilon)_\epsilon$, where X^ϵ is the process associated to the trace of J on $[0, \infty) \times (-\epsilon, \epsilon)^c$ (which can be shown to be Poisson with finite intensity).

Just for completeness, we close the picture by citing the deep theorem of Levy and Ito which asserts that *every* Levy process can be written as the sum of a pure drift, a Wiener process (multiplied by some constant) and a pure jump process of the same form as processes dealt with in this section.

1.4 Martingales

We briefly recall the notion of a martingale.

Definition 9 A real valued stochastic process X is said to be a martingale if X_t is integrable $\forall t$ and

$$X_t = \mathbb{E}[X_s || \mathcal{F}_t^X] \tag{7}$$

for every $t < s$.

There is a very direct and simple connection between Levy processes and martingales. Namely, if a Levy process X is integrable, then it is very easy to note that $(X_t - \mathbb{E}[X_t])_t$ is a martingale. Thus in particular, a Wiener process is a martingale and if X is a Poisson

[(†)] By convention a Poisson distribution with infinite intensity is a Dirac measure sitting at ∞

process with intensity λ then $(X_t - \lambda t)_t$ (the so called compensated Poisson process) is a martingale.

2 Basic concepts in mathematical finance

We now consider a financial market where a single asset is traded[‡]. We could think of this asset as a stock or an FX (foreign-exchange) rate, i.e. the price of a unit of some foreign currency. Given the fact that the evolution of the price process is random, it is natural to model it as a (continuous-time) stochastic process, call is $(S_t)_t$.

One of the main subjects of investigation in mathematical finance is the pricing of derivatives. A derivative (also known as a contingent claim) with maturity T on an asset is a contract which guarantees its holder to be paid some random amount of money at time T depending in some way (possibly quite complicated) upon the price fluctuations up to time T. Mathematically a derivative can be represented an $\mathcal{F}_T^S$-measurable random variable.

For example a contract that gives its holder the right but not the obligation to buy (resp., sell) the asset at time T at the pre-specified price K is called a call (resp., put) option with strike K. The associated random variable is of course $\max(S_T - K, 0)$ (resp., $\max(K - S_T, 0)$).

Any derivative can have some "barriers" embedded in it. Barriers are of two types, knock-in or knock-out. A derivative, call it X, with a knock-in (resp., knock-out) barrier at level B pays out only if the price process breaches (resp., never breaches) the level B before expiration. Mathematically the associated random variable is $X\mathbb{I}_{\{\tau_B<T\}}$ (resp., $X\mathbb{I}_{\{\tau_B>T\}}$). Here τ_B is the first hitting time of B by the price process S, i.e. $\tau_B = \inf\{t > 0 : S_t \leq B\}$ or $\inf\{t > 0 : S_t \geq B\}$ if the initial price was above the barrier B.

A one-touch (OT) option struck at level B, particularly common in the FX market, is a knock-in barrier option where the underlying claim X is equal to 1. In other words, a OT option pays 1 if and only if the price underlying asset ever touches the level B.

A fundamental concept in a financial market is that of admissible strategy. Its definition hinges upon the concept of stochastic integral with respect to whatever kind of class of processes S belongs to and this is beyond the scope of this note, so we just give an informal definition. A portfolio strategy is basically a stochastic process adapted to the filtration $\mathbb{F}^S$ which represents at each time how many stocks we are holding. We say that a strategy is admissible if all the changes in its value are due to fluctuations in the stock price and not to injections or withdrawal of money.

An arbitrage in a financial market is an admissible trading strategy such that its time 0 value is 0 whereas its time T value is non-negative and strictly positive with positive probability. Intuitively, an arbitrage opportunity is the possibility of making money out of nothing and it is natural that the guiding economic principle of model-building is to avoid the existence of such an opportunity. It is the so called first fundamental theorem of mathematical finance that gives sharp necessary and sufficient conditions for the absence

[‡]here and in the following, we are implicitly assuming the existence of a second asset whose price never fluctuates, think of it as a bank deposit where we can put all the money we do not invest in the other asset but that pays no interest rate.

of arbitrage opportunities, namely (under mild conditions on the process S) the existence of a probability measure $\mathbb{Q}$ equivalent[(§)] to the original one such that S is a $\mathbb{Q}$-martingale.

A claim with maturity T is said to be attainable if there exists an admissible trading strategy that replicates it, i.e. has a time T value equal to that of the claim. A financial market is said to be complete if every claim is attainable. The so-called second fundamental theorem states that an arbitrage-free market is complete if and only if the martingale measure is unique.

In the case of a arbitrage-free complete market, the price process of a claim X can be defined naturally as

$$\Pi_t(X) = E^{\mathbb{Q}}[X||\mathcal{F}_t] \tag{8}$$

This definition is very natural since the resulting process $\Pi(X)$ is a martingale which is equal to X at time T.

Thus, at least ideally, the specification of any financial model should in theory start from an arbitrary economically justifiable model and then an accurate no arbitrage analysis should be carried out, possibly by using the fundamental theorem given above. Actually it is often the case that a model assumes *a fortiori* the absence of arbitrage and makes the additional assumption that S is a martingale. The latter practice, widely used especially in term-structure modeling, is known as martingale modeling.

Let us conclude this section with some examples of possible models for an asset price.

A Wiener (normal model) or compensated Poisson processes of course would do, because of the relation between Levy processes and martingales we saw in the first section. An important drawback of these processes, though, is that they do not stay positive. This can be overcome by taking their exponential. In particular the single parameter (lognormal) model

$$S_t = e^{\sigma W_t - \frac{1}{2}\sigma^2 t} \tag{9}$$

(where W is a Wiener process) is commonly referred to as the Black-Scholes model. The analogous version for a purely discontinuous (two-parameter) approach[(¶)] is

$$S_t = e^{ln(1+J)N_t - \lambda J t} \tag{10}$$

where N is a Poisson process with intensity λ.

Both the above processes happen to be solutions of some stochastic differential equations. In particular the latter is the solution of

$$dS_t = S_t J(dN_t - \lambda dt) \tag{11}$$

[(§)] Two measures $\mathbb{P}$ and $\mathbb{Q}$ are said to be equivalent if $\mathbb{P}(A) = 0$ iff $\mathbb{Q}(A) = 0$ for every A in the underlying σ-algebra.

[(¶)] These two examples could be seen as particular cases of an "exp-Levy" model by using the characteristic exponent of the underlying Levy processes. This approach would shed some light on the intuitive meaning of the parameters.

so J can be seen as the percentage return on the stock when the process N jumps. The former lognormal model, on the other hand, can be seen as the solution of

$$dS_t = S_t \sigma dW_t \tag{12}$$

but here the dW_t integral cannot be defined as an ordinary Lebesgue-Stieltjes integral because it can be shown that W has infinite variation. One way to circumvent this problem has been found by Ito in the 40's and it is referred to as the Ito integral, with which we will not deal though. Natural generalizations of the above model have been proposed, namely letting the σ to be stochastic. In the so-called "local volatility" models σ depends on S_t only through a deterministic function, whereas in "stochastic volatility" models it is allowed to be a process on its own right.

Actually the purely lognormal (constant σ) model is so common that it is used as a benchmark in the following simple way. First note that if the underlying follows a lognormal model with volatility σ, the price of a call option (for fixed maturity T and strike K) is given by

$$c_{BS}(T, K; \sigma) := \mathbb{E}[max(S_T - K, 0)] \tag{13}$$

and this expression can be evaluated as a simple closed-form formula (the well-known Black-Scholes formula). At this point, for every model M for S we could define the lognormal implied volatility function σ_M^{imp} implicitly as the unique function such that

$$c_M(T, K) = c_{BS}(T, K; \sigma_M^{imp}(T, K)) \tag{14}$$

where of course $c_M(T, K)$ is the call price in the model M, which depends upon all the parameters used in M itself. It is straightforward to note that if M is a lognormal model with volatility σ^* then we would have a flat implied volatility function $\sigma_M^{imp}(T, K) = \sigma^*$ for every K and T.

3 Barrier and one-touch Options

One-touch (OT) options are path-dependent claims whose price depends in a non-trivial way upon the choice of the model.

The first fact behind the effect of different models on OT prices is the following: say model M^{neg} and model M^{pos} produce the same implied volatility at strike K, but implied volatility is downward sloping in M^{neg} and upward sloping in M^{pos}. Then a binary put with strike K will have a lower price in M^{neg} than in M^{pos}. The best way to see it is to think of the binary put as a limit of put spreads, i.e.

$$BinPut(K) \approx \frac{Put(K + \epsilon) - Put(K - \epsilon)}{2\epsilon} \tag{15}$$

and notice that M^{neg} gives a cheaper $Put(K + \epsilon)$ and a more expensive $Put(K - \epsilon)$.

The second thing to notice is that a OT down can be "approximately" replicated with two binary puts with the same strike. This is exactly true if the spot follows a normal process as:

$$dS_t = \sigma dW_t \tag{16}$$

The reasoning is simple. Buy two binaries and wait until the barrier is hit: if that never happens everything will expire worthless, otherwise at the hitting time the two binaries will be worth $\frac{1}{2}$ each, enough to handle the OT. In a lognormal model as

$$dS_t = S_t \sigma dW_t \tag{17}$$

an at-the-money (ATM) binary will not be worth exactly $\frac{1}{2}$ so we need a little adjustment which can be computed for example by using the symmetry formula

$$E[f(S_T)||\{S_0 = x\}] = E\left[\left(\frac{S_T}{x}\right) f\left(\frac{x^2}{S_T}\right)||\{S_0 = x\}\right] \tag{18}$$

Now it becomes clear that the key feature of the model that determines the price of OT options is the evolution of prices of ATM binaries, but in the light of what we saw before that is linked in a one-to-one fashion to the evolution of ATM volatility skew. A pure local volatility model, in which the smile tends to go in the opposite direction of the spot, will produce higher prices for OT than an stochastic volatility model, in which the smile typically tends to fluctuate with the spot.

References

[1] Carr, P., Ellis, K. and Gupta, V., *Static Hedging of Exotic Options*. Journal of Finance 53 (1998), 1165–1191.

[2] Cont, Rama, and Peter Tankov, "Financial Modelling With Jump Processes". Chapman and Hall/CRC Financial Mathematics Series, Boca Raton, 2004.

[3] Gatheral, Jim, "The Volatility Surface. A practitioner's guide". Wiley Finance, 2006.

[4] Merton, Robert C., *Option pricing when underlying stock returns are discontinuous*. Journal of Financial Economics, vol. 3/1-2 (1976), 125–144.

[5] Rogers, Chris, and David Williams, "Diffusions, Markov Processes and Martingales". Cambridge Mathematical Library, 2000.

Geometric Surface Processing for Computer Graphics

Marco Rucci (*)

1 Introduction

The application of mathematical models based on Partial Differential Equations (PDE) to image processing and computer graphics problems has been extremely successful over the past 20 years. In particular, geometric surface flows have been extensively used in mesh processing. While a large part of the image processing community solve the PDE models using an Eulerian methodology (typically, with level sets), Lagrangian representations of surfaces based on triangle meshes are most common in graphics. In this Lagrangian setting, discretization of continuous flows is usually achieved through the use of discrete differential operators or using finite element techniques.

We consider both discrete geometric flows, i.e., flows based on discrete analogous of continuous differential geometry quantities, and variational methods. Different approaches are based on the classical discretization of continuous models by finite volume and finite elements schemes.

Let $\mathcal{M}_0 = Image(X_0) := \{X_0(u), u \in [0,1] \times [0,1]\}$ be a compact, closed immersed orientable surface in $\mathbb{R}^3$ and X_0 be the corresponding parameter map. A geometric surface evolution consists of finding a family $\mathcal{M}(t) = Image(X(\cdot,t)), t \in [0,T), T > 0$ of smooth, closed, immersed orientable surfaces in $\mathbb{R}^3$ which evolve according to the flow equation (geometric flow)

$$\frac{\partial X}{\partial t} = -\beta \overrightarrow{N} + \alpha \overrightarrow{T},$$

where $\overrightarrow{N}$ is the unit normal vector to the surface, β is a velocity applied along the normal direction and α is the velocity in the tangent direction $\overrightarrow{T}$.

The family of manifolds $\mathcal{M}(t) \in \mathbb{R}^4$ moves along the normal direction driven by a normal velocity β which may be a function, for example, of the curvature and spatial position. The normal motion controls the geometry of the surface while the role of the

(*)Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy; E-mail: marco.rucci@gmail.com. Seminar held on December 5th, 2012.

tangential velocity is a sort of redistribution of the nodes which improves the accuracy of the surface representation.

In order to numerically approximate the PDEs on the evolving surface $\mathcal{M}(t)$, we define a discrete setting. The spatial approximation of $\mathcal{M}(t)$ is an evolving interpolated polyhedral mesh consisting of a union of faces whose vertices $X(t)$ lie on $\mathcal{M}(t)$, and $X(t)$ represents the parameterization of the surface itself. We can define, discretize and approximate differential evolutive PDE models based on local operators such as the Laplace-Beltrami, the intrinsic gradient and divergence.

2 Smoothing

We propose a solution to the fairing or smoothness problem. Such problem is formulated in terms of variational or energy based models in order to derive a nonlocal approach that performs smoothing by evolving the surface according to a fourth order Non Local Surface Diffusion Flow (NL-SDF) on $\mathcal{M}$. Results are summarized in [5].

2.1 The nonlocal variational fairing

For a surface parameterization X of $\mathcal{M}$ on a domain Ω, and a given vector field $f \in \mathbb{R}^{N_v \times 3}$, we consider the minimization of the following functional

$$\min_X \int_\Omega |\nabla_{w\mathcal{M}} X|^2 + \frac{\lambda}{2}(X - f)^2 d\omega, \tag{1}$$

where $\lambda > 0$ is a regularization parameter and $\nabla_{w\mathcal{M}}$ is a weighted gradient operator. The corresponding Euler-Lagrange descent flow can be written as

$$\frac{\partial X}{\partial t} = \int_\Omega (X(y) - X(x))W(x, y)d\omega + \lambda(f - X), \tag{2}$$

with $x, y \in \Omega$, (see [1] for a similar definition). Here $W(x, y)$ is the weight function, which satisfies $W(x, y) \geq 0$, and is symmetric $W(x, y) = W(y, x)$. The spatial discretization of (2) on the mesh M, is

$$\frac{\partial X_i}{\partial t} = \sum_{j \in N(i)} W_{ij}(X_j - X_i) + \lambda(f_i - X_i), \tag{3}$$

where X_i denotes the value of X at the ith vertex, $i = 1, \ldots, N_v$, and $N(i)$ is the set of 1-ring neighbor vertices of the ith vertex.

Let $f(x) := (f^1, f^2, f^3)(x)$ be a vector field on $\mathcal{M}$, $W(x, y)$ is the same for all vector components. Let $X(x) := (X^1, X^2, X^3)(x)$ be the coordinate function vector on M, where X^1 is the scalar function that defines the first coordinate of point $x \in \mathcal{M}$, and analogously for the second and the third coordinate scalar functions. Then the regularizing formulation (3) for each vector component $X^k, k = 1, 2, 3$, is

$$\frac{\partial X_i^k}{\partial t} = \sum_{j \in N(i)} W_{ij}(X_j^k - X_i^k) + \lambda(f_i^k - X_i^k), \tag{4}$$

by initializing each component k of X as $X^k|_{t=0} = f^k$.

If we let $W_{ij} = w_{ij}$, with w_{ij} defined by cotangent weight [2] then the regularized PDEs (4) can be interpreted as the spatial discretization on M of the well know *mean curvature flow* (MCF)

$$\frac{\partial X}{\partial t} = \triangle_{\mathcal{M}} X + \lambda(X_0 - X), \quad X|_{t=0} = X_0, \tag{5}$$

with initial surface X_0. The first term in (5) is the *regularization* term, while the second one is the *fidelity* term.

We propose the following *nonlocal weighted Laplace-Beltrami operator* on M,

$$L_w X_i = \sum_{j \in N(i)} (X_j - X_i) W_{ij} w_{ij}, \tag{6}$$

where w_{ij} are the laplacian discretization weights, while W_{ij} depends on a similarity measure between ith and jth vertex.

By initializing $X|_{t=0} = X_0$ and using the nonlocal operator (6), then (3) can be rewritten as

$$\frac{\partial X_i}{\partial t} = L_w X_i + \lambda(X_{0_i} - X_i). \tag{7}$$

Replacing X with the mean curvature normal vector $\overrightarrow{H}$ in (4), and considering a uniform discretization of the time interval $[0, T], T > 0$, with a temporal time step τ, then (7) can be fully discretized using, for example, an implicit scheme:

$$(I - \tau L_w)\overrightarrow{H}_i^{n+1} = \overrightarrow{H}_i^n, \quad \overrightarrow{H}|_{t=0} = \overrightarrow{H}^0, \tag{8}$$

where L_w is computed as given in (6), with initial condition $\overrightarrow{H}^0$ determined from X_0.

The two-step strategy first smooths the normal vectors allowing the mean curvature normals to diffuse on M, then the second step refits the parameterization X according to a given mean curvature distribution. The mean curvature smoothing (8) is "nonlocal". By this we mean that a "nonlocal" operator is used which includes weights that penalize the similarity between patches.

3 Remeshing

Remeshing refers to the redistribution of the sampling and connectivity of the geometry in order to satisfy mesh property requirements while maintaining surface features. We present an adaptive remeshing method that uses the mean curvature as an intrinsic measure of regularity. Results are reported in [3].

3.1 Adaptive Mesh Regularization

The AR method alternates equalization of edge lengths and vertex valence, which generate a new connectivity, with adaptive mesh regularization, which modifies the distribution of the vertices on the surface.

The mesh regularization method consists of a two-step PDE model. In the first step, the vertex area distribution function $A(X)$ defined on the mesh M with vertex set $X = \{X_i\}_{i=1}^{n_v}$, is diffused over the mesh, constrained by the mean curvature map. In the second step, the vertex position is tangentially relocated to obtain edges on element stars approximately of the same size, and all the vertex areas proportional to the surface features.

Let A_0 be the initial vertex area distribution function computed as the Voronoi area at each vertex on the mesh M, with vertex set X_0. Then in STEP 1, the vertex area distribution function $A(X)$ is diffused on $\mathcal{M}$ by solving

$$\frac{\partial A}{\partial t} = \triangle_{\mathcal{M}}^{w_H} A(X), \quad A(0) = A_0. \tag{9}$$

In (9) the operator $\triangle_{\mathcal{M}}^{w_H}$ is the *weighted Laplace-Beltrami operator* discretized on the mesh M by the matrix L_{w_H} with elements

$$L_{w_H}^{ij} = \frac{1}{\sum_{j\in N(i)} w_{ij}} \begin{cases} -\sum_{j\in N(i)} w_{ij} W_{ij} & i = j \\ +w_{ij} W_{ij} & i \neq j, j \in N(i) \\ 0 & otherwise \end{cases} \tag{10}$$

The weights W_{ij} prevent the area diffusion in high curvature regions. They depend on a similarity measure between the i^{th} and the j^{th} vertex, and are defined in terms of mean curvature values H on the mesh M. Increasing the number of time steps, the diffusion of (9) without weights converges to a constant area all over the entire mesh.

In STEP 2 of the AR algorithm the vertex position X is updated, taking into account the resulting $A(X)$ area distribution obtained in STEP 1, by solving the following constrained curvature diffusion equation

$$\frac{\partial X}{\partial t} = \nabla_{\mathcal{M}}^{w_A} \cdot (g(|H(X)|) \nabla_{\mathcal{M}}^{w_A} X), \quad X(0) = X_0, \tag{11}$$

where the function $g(\cdot)$, referred to as the diffusivity, is defined as

$$g(s) := \frac{1}{(1 + s^{\alpha})}, \tag{12}$$

where $\alpha > 0$ is a small positive constant value. The geometric evolution driven by (11) with high mean curvature values, that is, belonging to sharp creases and corners.

At each vertex X_i, linearizing (11) by evaluating $g(|H(X_i)|)$ with X_i from the previous time-step, the right-hand side of (11) reduces to

$$g(|H(X_i^{old})|) \triangle_{\mathcal{M}}^{w_A} X_i^{new}. \tag{13}$$

Finally, the displacement of the vertex X_i is in the tangent plane if we replace (11) with

$$\frac{\partial X_i}{\partial t} = (I - \overrightarrow{N}_i \overrightarrow{N}_i^T) g(|H(X_i)|) \triangle_{\mathcal{M}}^{w_A} X_i, \quad X(0) = X_0, \tag{14}$$

where $\overrightarrow{N}_i$ is the unit normal to the surface at X_i.

4 Reconstruction

We introduce a novel simple surface construction procedure based on functional optimization, which, for a given 3D curve network, automatically constructs a smooth surface preserving sharp features defined by the user. Results are summarized in [4].

The approach we follow for surface construction is based on the surface diffusion flow

$$\triangle_{\mathcal{M}} H = 0. \tag{15}$$

Equation (15) can be derived from minimizing the total curvature functional which leads to a minimal energy surface. The resulting surface has to satisfy both geometric constraints, given by a set $\overline{X_0}$ of points on the 3D curve network, and sharpness constraints associated at each given curve, while preserving the topology defined by the polyline-mesh.

Let X_0 be an initial surface which interpolates the set $\overline{X_0}$ of points on the 3D given curves and preserves the topology defined on the base-mesh, then we solve a global optimization problem by applying directly to the coordinate maps X the following fourth order flow

$$\frac{\partial X}{\partial t} = \triangle_{\mathcal{M}} \overrightarrow{H} + \lambda(X - \overline{X_0}), \quad X(0) = X_0, \tag{16}$$

where λ is a positive parameter which controls the effect of the data fidelity term that places positional constraints on all vertices of the 3D curves.

The construction method is based on a preliminary step where we construct a sufficiently Refined Mesh X_0, which includes $\overline{X_0}$, by tessellating each polygon in the Base Mesh following the same splitting rules of the generalized Catmull-Clark subdivision and iterating so that each n-sided face is subdivided into $16n$ quads. The newly introduced vertices move according to (16) towards a minimal energy surface which satisfies the given geometry and feature constraints.

References

[1] B. Dong, J. Ye, S. Osher, and I. Dinov, *Level Set Based Nonlocal Surface Restoration.* Multiscale Modeling and Simulation 7/2 (2008), 589–598.

[2] M. Meyer, M. Desbrun, P. Schröder, and A. H. Barr, *Discrete differential-geometry operators for triangulated 2-manifolds.* In "Visualization and Mathematics III", Springer Verlag (2003), 35–57.

[3] S. Morigi and M. Rucci, *Adaptive Tangential Remeshing.* In Proc. "Computer Graphics International", Ottawa, Canada (2011), 1–5.

[4] S. Morigi and M. Rucci, *Reconstructing surfaces from sketched 3D irregular curve networks.* In Proc. "Eighth Eurographics Symposium on Sketch-Based Interfaces and Modeling (SBIM)", T. Hammond and A. Nealen, editors, ACM Press (August 2011), 39–46.

[5] S. Morigi, M. Rucci, and F. Sgallari, *Nonlocal Surface Fairing.* In "LNCS 6667, SSVM 2011", A. M. Bruckstein, editor, Springer-Verlag Berlin Heidelberg (2011), 38–49.

Geometric quantization and coherent states: a link between classical and quantum mechanics

Daniele Fontanari (*)

Abstract. A physical system can be mathematically described either in the quantum or in the classical framework, the latter being an accurate approximation only when the magnitude of the physical quantities that characterize the system is large compared to the value of the Planck constant $\hbar$. It is usually accepted that the "right" classical description of a system should emerge from the quantum one as a limit (in some sense) for $\hbar$ close to 0 (semi classical limit). The inverse problem is also of great interest: given a classical description of a system is it possible to find a suitable quantum one? Such a procedure, when properly defined, is called quantization. When the classical system is described in the Hamiltonian formalism there are some natural postulates, the Dirac quantum conditions, that should hold for the quantized system. A procedure, among many others, that give an explicit solution is known as "geometric quantization". The importance of this method is because of the strong connection it maintains with the geometric structure of the underlying classical system.
I will give a review/introduction of the basic concepts and interpretations of (non-relativistic) Hamiltonian and quantum mechanics, in particular I will try to highlight the similarities between the the two descriptions with examples drawn from concrete systems. Then I will proceed with an overview of the apparatus of geometrical quantization, trying to analyze some fundamental results and drawbacks of this procedure. Finally I will show that coherent states naturally arise in the framework of geometric quantization (in the case of Kähler manifolds) and can be interpreted as quantum states that best approximate the classical ones and provide a means to uncover some aspects of the behaviour of the classical dynamics of a system in its quantum counterpart.

A physical system(†) can be described by a mathematical model which can be either described using two different formalisms: the classical or the quantum one.

Classical mechanics gives an accurate description of the system only when the *Planck constant* $\hbar \sim 1.05 \cdot 10^{-34} Js$ is negligible compared to the characteristic dynamical quantities of the system (such as energies or timescales).

While the classical and the quantum descriptions of a physical system are completely different, it is usually assumed that it should be possible to recover classical mechanics

(*) Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy; E-mail: danielefontanari@gmail.com . Seminar held on December 19th, 2012.

(†) In the following we will deal *exclusively* with non-relativistic systems.

from quantum mechanics in the so called semi classical limit (i.e. $\hbar \to 0$). It is thus of fundamental importance the problem of associating a classical system to the "right" quantum one or the converse of finding a suitable quantum system given its classical analogue. The latter goes under the name of *quantization*.

1 Mathematical foundations of classical mechanics

To understand the quantization procedure we must take into account the analogies and the differences between classical and quantum mechanics. We start by analyzing a simple classical system (a standard reference for classical mechanics is [1], while the theory of symplectic geometry is presented in [7], the motion of a massive particle immersed in a conservative force field. If $m > 0$ denotes the mass of the particle, $x \in \mathbb{R}^3$ its position and $V \in \mathcal{C}^\infty(\mathbb{R}^3)$ the force potential we have that the equation of motion for the particle are

$$m\ddot{x} = -\nabla V(x) \tag{1}$$

where ∇ denotes the gradient in $\mathbb{R}^3$ and $\ddot{x}$ the second derivative of x with respect to time. It is possible to rewrite equation (1) to obtain a first order equation defined in the variables $(x, p) \in \mathbb{R}^6$ (the phase space of the system):

$$\begin{aligned} \dot{x} &= \frac{p}{m} \\ \dot{p} &= -\nabla V(x) \end{aligned} \tag{2}$$

Defining the total energy of the system as a function (the *Hamiltonian* of the system) defined on the phase space:

$$H(x, p) = \frac{p^2}{2m} + V(x)$$

we can rewrite (2) as

$$\begin{aligned} \dot{x} &= \frac{\partial H(x,p)}{\partial p} \\ \dot{p} &= -\frac{\partial H(x,p)}{\partial x}. \end{aligned} \tag{3}$$

To generalize this situation we give the following:

Definition A *symplectic manifold* is a $2n$-dimensional $\mathcal{C}^\infty$ manifold $\mathcal{M}$ equipped with a differential 2-form ω, the *symplectic form*, such that

- ω is *closed* (i.e. $\mathrm{d}\omega = 0$).
- ω is *non-degenerate* (i.e. for $p \in \mathcal{M}$, $X \in T_p\mathcal{M}$ then $\iota_Y \iota_X \omega(p) = 0$ for every $Y \in T_p\mathcal{M}$ iff $X = 0$).

to every function $H \in \mathcal{C}^\infty(\mathcal{M})$ it is possible to associate its *Hamiltonian vector field* X_H defined by:

$$\iota_{X_H}\omega + \mathrm{d}H = 0.$$

X_H exists and it is uniquely defined by H thanks to the non degeneracy of ω.

In this context a physical system is described by the choice of a symplectic manifold $(\mathcal{M}, \omega)$ and a distinguished function, the *Hamiltonian*, $H \in \mathcal{C}^\infty(\mathcal{M})$ such that the flow $t \mapsto \Phi_H^t$ of X_H defines the dynamics of the system. In particular to every function $f \in \mathcal{C}^\infty(\mathcal{M})$ it is possible to associate[‡] its evolution at time $t \in \mathbb{R}$ $f_t \in \mathcal{C}^\infty(\mathcal{M}) = f \circ \Phi_H^t$. From the definition of X_H we have that

$$\frac{\mathrm{d}f_t}{\mathrm{d}t} = \iota_{X_H} \mathrm{d}f_t = \iota_{X_{f_t}} \iota_{X_H} \omega = \{f_t, H\} \tag{4}$$

where we introduced the *Poisson bracket* that to every pair of functions $f, g \in \mathcal{C}^\infty(\mathcal{M})$ associates a function $\{f, g\} := \iota_{X_f} \iota_{X_g} \omega \in \mathcal{C}^\infty(\mathcal{M})$.

2 Classical mechanics: Examples

The standard example of a symplectic manifold if given by $\mathcal{M} = \mathbb{R}^{2n}$ for some n with coordinates $(x_1, \dots, x_n, p_1, \dots, p_2)$ and $\omega = \sum_{i=1}^n \mathrm{d}p_i \wedge \mathrm{d}x_i$. In this case

$$\{f, g\} = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial x_i}.$$

In particular when $n = 3$ we recover (3).

A less trivial example of symplectic manifold is given by $\mathbb{S}^2$ equipped with its natural oriented area form as its symplectic form. In this case, if we identify $\mathbb{S}^2$ by the set of points in $\mathbb{R}^3$ of unitary norm then

$$\{f, g\}(x) = x \cdot (\nabla f \times \nabla g)$$

We can however consider local coordinates on $\mathbb{S}^2$, for instance defining $(X_S, Y_S) \in \mathbb{R}^2$ using stereographic coordinates by (here (x_1, x_2, x_3) is a point of unitary norm in $\mathbb{R}^3$):

$$X_S = \frac{x_1}{1 - x_3}$$
$$Y_S = \frac{x_2}{1 - x_3}$$

we have a local chart that cover $\mathbb{S}^2$ except for the north pole, while defining

$$X_N = \frac{x_1}{1 + x_3}$$
$$Y_N = -\frac{x_2}{1 + x_3}$$

we cover $\mathbb{S}^2$ but the south pole. Locally we have that

$$\omega = \frac{4}{(1 + X_S^2 + Y_S^2)^2} \mathrm{d}X_S \wedge \mathrm{d}Y_S = \frac{4}{(1 + X_N^2 + Y_N^2)^2} \mathrm{d}X_N \wedge \mathrm{d}Y_N.$$

[‡] for simplicity it is assumed that the flow of X_H is complete

It is also possible to consider $\mathbb{C}^n$ equipped with the symplectic structure inherited by the identification of $\mathbb{C}^n$ with $\mathbb{R}^{2n}$ (i.e the element $(z_1, \dots, z_n)$ of $\mathbb{C}^n$, with $z_i = x_i + \imath p_i$, is identified with $(x_1, \dots, x_n, p_1, \dots, p_n)$). Explicitly we have:

$$\omega = \frac{\imath}{2} \sum_{i=1}^{n} \mathrm{d}z_i \wedge \mathrm{d}\bar{z}_i.$$

Similarly, for $\mathbb{S}^2$, we can define complex coordinates $z_S, z_N \in \mathbb{C}$ by

$$z_S = X_S + \imath Y_S$$
$$z_N = X_N + \imath Y_N$$

we have that

$$\omega = \frac{2\imath}{(1 + z_S \bar{z}_S)^2} \mathrm{d}z_S \wedge \mathrm{d}\bar{z}_S = \frac{2\imath}{(1 + z_N \bar{z}_N)^2} \mathrm{d}z_N \wedge \mathrm{d}\bar{z}_N.$$

Moreover we see that the transition functions between the local charts are holomorphic, since $z_S = z_N^{-1}$, and so $\mathbb{S}^2$ is also a complex manifold (like, trivially, $\mathbb{C}^n$). This will play a central role in the following.

3 Mathematical foundations of quantum mechanics

In quantum mechanics (see for instance [2, 9] for a complete reference on the subject. Here the theory will be only outlined) even the complete knowledge of the state of the system, unlike the classical case, does not allow the knowledge of the outcome of the measure of an observable (like the position or the velocity of a particle). We have instead that the possible outcomes are driven by a probability distribution.

In this case the system is defined by a (possibly infinite dimensional) separable complex Hilbert space $\mathcal{H}$. Every non zero element of $\mathcal{H}$ represents a state of the system (in the classical case states are represented by points of $\mathcal{M}$). Observables of the system (i.e. quantities that can be measured, their classical counterpart would be represented by functions on $\mathcal{M}$) are represented by self-adjoint operators on $\mathcal{H}$. If an observable A has discrete spectrum[(§)] then there exists a numerable orthonormal basis for $\mathcal{H}$, $(\psi_i)_{i \in \mathcal{I} \subseteq \mathbb{N}}$, such that every ψ_i is an eigenvector of A: i.e. $A\psi_i = \lambda_i \psi_i$ for some $\lambda_i \in \mathbb{R}$. If a generic state $\psi \in \mathcal{H}$ is expressed as:

$$\psi = \sum_{i \in \mathcal{I}} \psi(\lambda_i) \psi_i$$

for $\psi(\lambda_i) \in \mathbb{C}\ \forall i \in \mathcal{I}$, then $\|\psi\|^{-2} |\psi(\lambda_i)|^2$ is interpreted as the probability of obtaining λ_i as the outcome of a measure of the observable A when the system is in the state defined by ψ. An observable of particular importance is given by the Hamiltonian H. It measures the energy of the system and determines the time evolution by the *Schrödinger equation*.

[(§)] this is not always the case, however for simplicity we will not consider the case of observables with continuous or mixed spectrum.

If $\psi \in \mathcal{H}$ is a state, we define its evolution at the time $t \in \mathbb{R}$ by:

$$\begin{aligned} U(t) &:= e^{\frac{H}{\imath\hbar}t} \\ \psi_t &= U(t)\psi \end{aligned}$$

or equivalently:

$$\begin{aligned} \psi_0 &= \psi \\ \imath\hbar\frac{\mathrm{d}\psi_t}{\mathrm{d}t} &= H\psi_t. \end{aligned}$$

It is possible to give an equivalent characterization of the dynamics where the states are fixed and the observables are evolving (*Heisenberg picture*): if A is an observable, its evolution A_t is defined as:

$$A_t = U(t)^{-1}AU(t)$$

or equivalently:

$$\begin{aligned} A_0 &= A \\ \frac{\mathrm{d}A_t}{\mathrm{d}t} &= \frac{1}{\imath\hbar}[A_t, H] \end{aligned} \tag{5}$$

where $[A, B] := AB - BA$ is the commutator between operators.

4 Quantum mechanics: Examples

The quantum system modelling a particle of mass $m > 0$, in some potential $V : \mathbb{R}^3 \to \mathbb{R}$, is given by $\mathcal{H} = L^2(\mathbb{R}^3)$ on which observables $X_1, X_2, X_3, P_1, P_2, P_3, H$ are defined as

$$\begin{aligned} X_i\phi(x) &= x_i\phi(x) \\ P_i\phi(x) &= -\imath\hbar\frac{\partial\phi(x)}{\partial x^i} \\ H\psi(x) &= (\frac{P^2}{2m} + V(X))\psi(x) = -\frac{\hbar^2}{2m}\Delta\psi(x) + V(x)\psi(x) \end{aligned}$$

If we take H as the Hamiltonian of the system, the Heisenberg equations (5) can be rewritten as

$$\begin{aligned} \frac{\mathrm{d}X_i}{\mathrm{d}t} &= \frac{P_i}{m} \\ \frac{\mathrm{d}P_i}{\mathrm{d}t} &= -\frac{\partial V}{\partial x^i}(X). \end{aligned} \tag{6}$$

While neither the spectra of X_i or P_i are discrete, it is possible to identify the probability density relative to the (simultaneous) measurement of $X = (X_1, X_2, X_3)$ in the state ψ with the function $x \mapsto |\psi(x)|^2$. On the other hand if we define

$$\hat{\psi}(p) := (2\pi\hbar)^{-\frac{3}{2}} \int \psi(x) \exp\left(-\imath\hbar^{-1}p \cdot x\right) \mathrm{d}x$$

or, equivalently,

$$\psi(x) := (2\pi\hbar)^{-\frac{3}{2}} \int \hat{\psi}(p) \exp\left(\imath\hbar^{-1} p \cdot x\right) \mathrm{d}p$$

we have that

$$\widehat{P_i\psi}(p) = p_i \hat{\psi}(p).$$

It is known that this amounts in identifying the probability density relative to the (simultaneous) measurement of $P = (P_1, P_2, P_3)$ in the state ψ with the function $p \mapsto |\hat{\psi}(p)|^2$. We see however that $\hat{\psi}$ is basically the Fourier transform of ψ. It is well known that if we try to define a function ψ which is concentrated around a point $x \in \mathbb{R}^3$, then its Fourier transform $\hat{\psi}$ is spread on $\mathbb{R}^3$. If we identify X as the position of the particle and P as its linear momentum this means that it is not possible to determine with arbitrary accuracy both the position and the velocity of a particle (i.e. it is not possible to localize a state on the phase space). This is a consequence of the more general *Heisenberg indeterminacy principle*.

Another example is given by $\mathcal{H} = \mathbb{C}^n$. The Hermitian structure on $\mathcal{H}$ is in this case the standard one given by

$$v \cdot w = \sum_{i=1}^{n} \bar{v}_i w_i \qquad v, w \in \mathbb{C}^n.$$

It is possible to define three observables L_1, L_2, L_3 satisfying

$$[L_i, L_j] = \imath\hbar\epsilon_{ijk} L_k \tag{7}$$

It is possible to prove that (7) alone completely determines L_1, L_2, L_3 for any n. In concrete system we have that $\mathcal{H}$ appears as the quantum space (or at least the quantum space is a tensor product of Hilbert spaces and $\mathcal{H}$ appears as a factor) when we are dealing with internal rotational degrees of freedom. It is possible to see that $L_1^2 + L_2^2 + L_3^2$ is proportional to the identity. Thus we can interpret $L = (L_1, L_2, L_3)$ as point in $\mathbb{R}^3$ constrained on a sphere. Also in this case it is possible to see that, if ψ is an eigenstate of L_3 (i.e. the measure of L_3 on the state is deterministic), then there is a indeterminacy in the measure of L_1 and L_2. Again this can be interpreted as the impossibility to localize a state in the phase space.

5 Quantization

It is immediate to notice the similarity between (2) and (6), and between (4) and (5) (when $\{\cdot,\cdot\}$ is replaced by $(\imath\hbar)^{-1}[\cdot,\cdot]$ and the quantum observables by their "corresponding" classical ones). Motivated by this we investigate the possibility to extend this analogy for generic symplectic manifolds for appropriate quantum Hilbert spaces. To do this we resort to the so-called geometric quantization (good references for the geometric quantization procedure are [6, 10, 14]).

5.1 Prequantization

The first step in obtaining a quantum system from a classical one is the prequantization procedure. The $2n$-dimensional symplectic manifold $\mathcal{M}$ is equipped with a natural measure (Liouville measure) given by the $2n$-form ω^n. Let E be a complex vector fiber bundle over $\mathcal{M}$ with 1 dimensional fibers equipped with a Hermitian scalar product (such a E is usually called, in this context, a line bundle). If ψ_1 and ψ_2 are sections of E it is possible to define the function over $\mathcal{M}$ (ψ_1, ψ_2) by taking the inner product of ψ_1 and ψ_2 fiber wise. It is then possible to define $\langle \psi_1, \psi_2 \rangle \in \mathbb{C}$ as the integral of (ψ_1, ψ_2) over $\mathcal{M}$ under the Liouville measure. If now we define $\mathcal{H}'$ as the space of the sections ψ of E for which $\langle \psi, \psi \rangle$ is finite, we have that $(\mathcal{H}', \langle \cdot, \cdot \rangle)$ is an Hilbert space.

The fiber bundle E might also be equipped with a covariant derivative ∇ compatible with the Hermitian structure on the fibers of E. Given a vector field X tangent to $\mathcal{M}$ and a section ψ of E we have that $\nabla_X \psi$ is a section of E. It is a known fact that there exists a 2-form Ω (curvature form), such that for any pair of vector fields X, Y and any section ψ we have

$$\frac{\imath}{2}(\nabla_X \nabla_Y \psi - \nabla_Y \nabla_X \psi - \nabla_{[X,Y]}\psi) = \iota_X \iota_Y \Omega \psi.$$

Definition We say that a symplectic manifold $(\mathcal{M}, \omega)$ is quantizable if there exists a fiber bundle E equipped with a covariant derivative ∇ such that $\hbar\Omega = \omega$. In this case $\mathcal{H}'$, as defined above, is called the prequantum Hilbert space associated to (E, ∇).

In this case, given a smooth $\mathbb{R}$-valued function f on $\mathcal{M}$, we can define the self-adjoint operator $\mathcal{Q}(f)$ on $\mathcal{H}'$ as:

$$\mathcal{Q}(f)\psi(p) = -\imath\hbar\nabla_{X_f}\psi(p) + f(p)\psi(p).$$

The map $\mathcal{Q} : f \mapsto \mathcal{Q}(f)$ is called the *quatization map*.

Thanks to its definition, and to the condition on Ω, the quantization map satisfies the following properties:

- $\mathcal{Q}$ is linear: $\mathcal{Q}(\mu f + \nu g) = \mu\mathcal{Q}(f) + \nu\mathcal{Q}(g)$ for every f, g functions on $\mathcal{M}$ and $\mu, \nu \in \mathbb{R}$.
- If $1_{\mathcal{M}}$ is the function on $\mathcal{M}$ constantly equal to 1 and $\mathbb{I}_{\mathcal{H}'}$ is the identity operator on $\mathcal{H}'$, $\mathcal{Q}(1_{\mathcal{M}}) = \mathbb{I}_{\mathcal{H}'}$.
- For any pair of function f, g on $\mathcal{M}$, $\mathcal{Q}(\{f, g\}) = (\imath\hbar)^{-1}[\mathcal{Q}(f), \mathcal{Q}(g)]$.

These properties suggest that $\mathcal{H}'$ is the right space representing the quantum analogue of $\mathcal{M}$, and $\mathcal{Q}(f)$ the observable associated to f.

Not every symplectic manifold however admits a prequantum Hilbert space and a quantization map, and even if they are defined they might not be unique. However we have that

Proposition 1 [14] *If the* Weil integrality condition *for $\mathcal{M}$ holds, i.e. for every oriented 2-surface $\Sigma \subseteq \mathcal{M}$ we have that*

$$\frac{1}{2\pi\hbar}\int_\Sigma \omega \in \mathbb{Z},$$

then there exists at least a prequantization of $\mathcal{M}$.

Moreover we have that

Proposition 2 *If $\mathcal{M}$ is simply connected, then $\mathcal{M}$ does not admit more than one prequantization.*

This for instance implies that $\mathbb{R}^{2n}$ (equipped with its natural symplectic structure) admits a unique prequantization procedure ($\mathbb{R}^{2n}$ is simply connected and it can be shown that when ω is exact then the Weil integrality condition automatically holds). In this case it is possible to take E as the trivial bundle over $\mathbb{R}^{2n}$, its sections can be identified with complex valued functions on $\mathbb{R}^{2n}$ and $\mathcal{H}' = L^2(\mathbb{R}^{2n}; \omega^n)$. In this case we have that:

$$\mathcal{Q}(f)\psi(x,p) = -\imath\hbar\{\psi, f\}(x,p) + \left(f(x,p) - \sum_i p_i \frac{\partial f(x,p)}{\partial p_i}\right)\psi(x,p).$$

It is also possible to see that $\mathbb{S}^2$ admits a (unique) prequantization procedure if and only if the radius of $\mathbb{S}^2$ is proportional to $\hbar/2$. We will see later that this leads to the well known quantization of spin.

5.2 Geometric quantization: the real case

In general the problem of prequantization is that $\mathcal{H}'$ is, in some sense, too large. In particular it is possible to define states which are indefinitely localized in the phase space, which contradicts Heisenberg uncertainty principle. To solve this problem we should discard the localized states. To solve this problem we need to introduce the concept of polarization of a symplectic manifold.

Definition Let $\mathcal{M}$ be a symplectic manifold. A distribution $\mathcal{F}$ of $T\mathcal{M}$ is a real polarization if

- $\mathcal{F}$ is integrable.
- The leaves of the foliation defined by $\mathcal{F}$ are n-dimensional manifold.
- For every $x \in \mathcal{M}$ and for every $v, w \in \mathcal{F}_x = \mathcal{F} \cap T_x\mathcal{M}$ then $\iota_w\iota_v\omega(x) = 0$ (i.e. the leaves of $\mathcal{F}$ are *Lagrangian* submanifolds of $\mathcal{M}$).

We say that a smooth section ψ of the line bundle E over $\mathcal{M}$ is polarized (with respect to the polarization $\mathcal{F}$) if, for every $x \in \mathcal{M}$ and $X \in \mathcal{F}_x$, $\nabla_X\psi = 0$. Obviously the polarized sections defines a vector space. Chosen a polarization $\mathcal{F}$ of $\mathcal{M}$, with line bundle E admitting a prequantum space $\mathcal{H}'$, we consider, as our quantum Hilbert space $\mathcal{H}$, the subset of $\mathcal{H}'$ defined by polarized square-integrable sections. For example, when $\mathcal{M} = \mathbb{R}^{2n}$ a standard choice is the polarization spanned by $\frac{\partial}{\partial p_1}, \ldots, \frac{\partial}{\partial p_n}$. In this case a polarized section is a function ψ depending on the x coordinates alone and

$$\mathcal{Q}(f)\psi(x) = -\imath\hbar \sum_i \frac{\partial\psi(x)}{\partial x_i}\frac{\partial f(x,p)}{\partial p_i} + \left(f(x,p) - \sum_i p_i \frac{\partial f(x,p)}{\partial p_i}\right)\psi(x).$$

in particular we have that, if f is a function of the x variables alone,

$$\mathcal{Q}(f)\psi(x) = f(x)\psi(x)$$
$$\mathcal{Q}(p_i)\psi(x) = -\imath\hbar\frac{\partial\psi(x)}{\partial x_i}$$

as expected.

5.3 Geometric quantization: the Kähler case

The previous construction is still unsatisfactory. For instance it is easy to see that not every symplectic manifold $\mathcal{M}$ admits a real polarization, for example the sphere does not. Moreover $\mathcal{H}$ as defined above may be empty: if we consider the case of $\mathcal{M} = \mathbb{R}^{2n}$ it is immediate to check that non-zero functions depending only on the x variable are not integrable. To overcome these problem we introduce the concept of *complex polarization*:

Definition Let $\mathcal{M}$ be a symplectic manifold. A complex polarization $\mathcal{F}$ is a distribution on the complexified tangent space of $\mathcal{M}$, $T\mathcal{M}^{\mathbb{C}} = T\mathcal{M} \otimes \mathbb{C}$, such that

- $\mathcal{F}$ is integrable.
- At every point of $\mathcal{M}$, $\mathcal{F}$ has (complex) dimension n.
- For every $x \in \mathcal{M}$ and for every $v, w \in \mathcal{F}_x$ then $\iota_w\iota_v\omega(x) = 0$ (where the domain of ω is trivially extended by linearity to $T^{\mathbb{C}}\mathcal{M}$).

We say that a smooth section ψ of the line bundle E over $\mathcal{M}$ is polarized (with respect to the polarization $\mathcal{F}$) if, for every $x \in \mathcal{M}$ and $X \in \mathcal{F}_x$, $\nabla_X\psi = 0$ (here ∇ is trivially extended by linearity to $T^{\mathbb{C}}\mathcal{M}$). An interesting situation arises when $\mathcal{M}$ is a Kähler manifold:

Definition A symplectic manifold $\mathcal{M}$ is a Kähler manifold if it admits a complex structure compatible with the symplectic one:

$$\omega(\imath v, \imath w) = \omega(v, w).$$

In this case we have that

Proposition 3 *If $\mathcal{M}$ is a Kähler manifold then the distribution spanned by the vector fields $\frac{\partial}{\partial\bar{z}^i}$ defining the antiholomorpic tangent space of $\mathcal{M}$ is a complex distribution. Polarized sections are the ones whose local representatives are holomorphic functions.*

For instance both $\mathbb{C}^n$ and the sphere are Kähler manifolds. In all the practical cases $\mathcal{H}$ is non-trivial (i.e. there exist non-zero holomorphic sections which are square-integrable), for example in the case of the sphere of radius $\frac{\hbar n}{2}$, $\mathcal{H}$ can be identified with the space of polynomials of degree not greater that n, which is isomorphic to $\mathbb{C}^n$. In this case we have that if $x \in \mathbb{R}^3$ with $\|x\| = \frac{\hbar n}{2}$ parametrizes the sphere, then $\mathcal{Q}(x_i) = L_i$ as defined before.

5.4 BKS Quantization

Even in the Kähler case there is still a problem. Namely, if ψ is a polarized state and f is a function on $\mathcal{M}$, then $\mathcal{Q}(f)\psi$ might be not polarized. We should then restrict the analysis only to *polarized* observables (i.e. $\mathcal{Q}(f)\psi$ is polarized as long as ψ is polarized). It is known (from the celebrated Grönewold-van Hove theorem and its variants, see [4, 5, 13]) that there are not many observables that are polarized. To overcome this difficulty it is possible to modify the quantization procedure. Since $\mathcal{H}$ is (in the Kähler case) a closed subspace of $\mathcal{H}'$ it is possible to define the orthogonal projection π of $\mathcal{H}'$ onto $\mathcal{H}$. If we define

$$\mathcal{Q}_{\mathrm{BKS}}(f) = \pi \circ \mathcal{Q}(f)$$

or equivalently

$$\langle \psi', \mathcal{Q}_{\mathrm{BKS}}(f)\psi \rangle = \langle \psi', \mathcal{Q}(f)\psi \rangle$$

for every pair of polarized sections ψ, ψ', we have that $\mathcal{Q}_{\mathrm{BKS}}(f)\psi$ is polarized for every ψ and $\mathcal{Q}_{\mathrm{BKS}}(f) = \mathcal{Q}(f)$ when f is a polarized observable. The drawback of this method is that now we have the weaker condition on $\mathcal{Q}_{\mathrm{BKS}}$:

$$\mathcal{Q}_{\mathrm{BKS}}(\{f, g\}) = (\imath\hbar)^{-1}[\mathcal{Q}_{\mathrm{BKS}}(f), \mathcal{Q}_{\mathrm{BKS}}(g)] + O(\hbar)$$

however it can be shown that for any reasonable quantization procedure this is unavoidable.

6 Coherent states

Even if now we are able to associate to a classical system a quantum one, we still would like to be able to compare in a more direct way the dynamics in the two framework, in particular we would like to find the "right" quantum state $\psi_p \in \mathcal{H}$ which is the most localized in the point $p \in \mathcal{M}$ (clearly with respect to the limitations imposed by the Heisenberg uncertainty relations). We have that the linear map $\mathcal{H} \ni \psi \mapsto \psi(p) \in \mathbb{C}$ is continuous. Thanks to the Riesz representation theorem this means that there exists a state $\psi_p \in \mathcal{H}$, the coherent state centered in p (more on coherent states, whose presence is ubiquitous in physics, can be found in [3, 8], an interesting study on the relation between BKS-quantization and coherent states can be found in [11, 12], such that $\langle \psi_p, \psi \rangle = \psi(p)$. It is also possible to see ψ_p as the projection on $\mathcal{H}$, through π, of the Dirac delta distribution centered in p (it is possible to extend the domain of π to the space of distributions) and so ψ_p can be rightfully thought as the most localized state in p. It results also that ψ_p is, among the elements of $\mathcal{H}$ of the same norm, the state which attains the maximum absolute value in p.

Given a state ψ we can thus interpret $H_\psi(p) := |\psi(p)|^2 = |\langle \psi_p, \psi \rangle|^2$, the *Husimi* distribution of ψ, as the probability to find ψ localized in p and so, if we are interested in analyze the classical orbit of a point $p_0 \in \mathcal{M}$ from a quantum perspective, then we should study $H_{\psi_{p_0}(t)}(p)$.

While the theory that compares the quantum evolution of coherent states with its classical counterpart is largely to be developed, there are hints that suggest that there exists a strong link between the two behaviours. In particular from numerical simulations

it seems that localized chaotic structures on the phase space, typical of perturbed super-integrable systems, emerge in the quantum analogue. A result that help to understand the connection between the classical and quantum framework is the following:

Proposition 4 *Let $h \in \mathcal{C}^\infty(\mathcal{M})$ a classical Hamiltonian function and $\psi \in \mathcal{H}$ a quantum state. If we denote by H_{ψ_t} the Husimi distribution of the state ψ_t, the evolution of ψ at the time t under the quantum Hamiltonian $\mathcal{Q}_{BKS}(h)$, and by $H_\psi(t)$ the evolution of H_ψ at the time t under the classical flow of h, we have that:*

$$\frac{\mathrm{d}}{\mathrm{d}t}(H_\psi(t) - H_{\psi_t})_{|t=0} = O(\hbar)$$

This proposition implies that the wave function of a coherent state relative to a point p in $\mathcal{M}$ should evolve (at least for short timescales and when $\hbar$ can be ignored) along the classical orbit of p.

References

[1] Arnol'd, V.I., "Mathematical Methods of Classical Mechanics". 2nd edn., Springer, New York, 1989.

[2] Dirac P.A.M., "The Principles of Quantum Mechanics". 4th rev. edn., Oxford university press, Oxford, 1958.

[3] Gazeau J.P., "Coherent States in Quantum Physics". Wiley-VCH, Berlin, 2009.

[4] Gotay M.J., Grundling H., Hurst C.A., *A Groenewold-van Hove theorem for* $\mathbb{S}^2$. Trans. Am. Math. Soc. 348 (1996), 1579–1597.

[5] Groenewold, H.J., *On the Principles of Elementary Quantum Mechanics.* Physica 12 (1946), 405–460.

[6] Guillermin V. and Sternberg S., "Geometric asymptotics". American mathematical society, Providence, 1977.

[7] Libermann, P. and Marle, C.-M, "Symplectic Geometry and Analytical Mechanics". D. Reidel, Dordecht, 1987.

[8] Perelomov, A.M., "Generalized coherent states and their applications". Springer-Verlag, Berlin, 1987.

[9] Sakurai J.J., "Modern Quantum Mechanics". Rev. edn., Addison-Wesley, Reading, MA, 1994.

[10] Śniatycki J., "Geometric Quantization and Quantum Mechanics". Springer-Verlag, New York, 1908.

[11] Tuynman G.M., *Generalized Bergman kernels and geometric quantization.* J. Math. Phys. 28 (1987), 573–583.

[12] Tuynman G.M., *Quantization: Towards a comparison between methods.* J. Math. Phys. 28 (1987), 2829–2840.

[13] Van Hove L., *Sur le problème des relations entre les transformations unitaires de la méchanique quantique et les transformations canoniques de la mécanique classique.* Acad. Roy. Belgique Bull. Cl. Sci. 37 (1951), 610–620.

[14] Woodhouse, N.M.J., "Geometric Quantization". 2nd edn., Clarendon Press, Oxford, 1994.

The Probabilistic Zeta Function

Duong Hoang Dung [(*)]

Let G be a finite group and choose randomly k elements in G. One could ask whether those elements generate the group G, and what the probability $P(G,k)$ for this event is. The answer is easy and dates back to P. Hall in 1936 as the following:

Theorem 1 [9] *Let G be a finite group. The probability that k randomly chosen elements generate the group G is calculated by the following formula*

$$P(G,k) = \sum_{H \leq G} \frac{\mu_G(H)}{|G:H|^k}$$

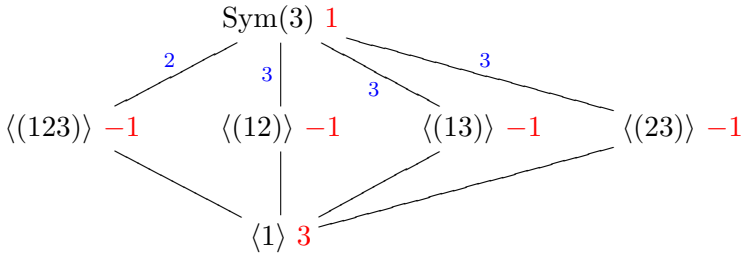
where $\mu_G(H)$ is the Möbius function defined over the subgroup lattice of G recursively by $\mu_G(G) = 1$ and $\mu_G(H) = -\sum_{H<K\leq G} \mu_G(K)$ if $H < G$.

Examples 2

- Let $G = C_p$ be the cyclic group of prime order p. Then

$$P(G,t) = 1 - \frac{1}{p^t}$$

- Let $G = \mathrm{Sym}(3)$ be the group of permutations on 3 letters. The subgroup lattice of G is described as the following:



$a_1 = \mu_G(\mathrm{Sym}(3)) = 1$
$a_2 = \mu_G(\langle(123)\rangle) = -1$

[(*)]Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy; E-mail: hduong@math.unipd.it . Seminar held on January 30th, 2013.

$a_3 = \mu_G(\langle(12)\rangle) + \mu_G(\langle(13)\rangle) + \mu_G(\langle(23)\rangle) = -3$
$a_6 = \mu_G(\langle 1\rangle) = 3$

$$P(\mathrm{Sym}(3), t) = 1 - \frac{1}{2^t} - \frac{3}{3^t} + \frac{3}{6^t} = \left(1 - \frac{1}{2^t}\right)\left(1 - \frac{3}{3^t}\right)$$

- $G = C_n$, the cyclic group of order n.

$$P(G, t) = \prod_{\substack{p|n \\ p \text{ prime}}} \left(1 - \frac{1}{p^t}\right)$$

- Let $G = C_p^k$ be the direct product of k copies of C_p, then

$$P(G, t) = \prod_{i=0}^{k-1} \left(1 - \frac{p^i}{p^t}\right)$$

- Let $G = \mathrm{Alt}(4)$ be the alternating group of degree 4, then

$$P(G, t) = \left(1 - \frac{2}{2^t}\right)\left(1 + \frac{2}{2^t}\right)\left(1 - \frac{1}{3^t}\right)$$

Proof. Let $\Phi(G, t)$ be the set of ordered t-tuples $(g_1, \cdots, g_t)$ such that $\langle g_1, \cdots, g_t\rangle = G$. Since every t-tuple generates a subgroup, we have

$$\sum_{H \leq G} \Phi(H, t) = |G|^t$$

By Möbius inversion formula, we obtain

$$\Phi(G, t) = \sum_{H \leq G} \mu_G(H)|H|^t$$

Dividing by $|G|^t$ in both sides gives us

$$P(G, t) = \frac{\Phi(G, t)}{|G|^t} = \sum_{H \leq G} \frac{\mu_G(H)|H|^t}{|G|^t} = \sum_{H \leq G} \frac{\mu_G(H)}{|G : H|^t}$$

□

We next would like to ask whether we have some analogue interpretations via subgroups and quotients. The following result dues to Gaschütz in [8]:

Theorem 3 *Let $N \trianglelefteq G$ be a normal subgroup of G. Then*

$$P(G, t) = P(G/N, t)P(G, N, t)$$

where $P(G, N, t)$ is the conditional probability that k randomly chosen elements generated G given that they together with N generate G.

One could expect that $P(G, N, t) = P(N, t)$. However, it is sometimes true in some cases.

Example 4 Let $G = \mathrm{Sym}(5)$ and $N = \mathrm{Alt}(5)$, then by [1], we have that

$$P(\mathrm{Sym}(5), t) = P(C_2, t)P(\mathrm{Alt}(5), t)$$

and so $P(\mathrm{Sym}(5), \mathrm{Alt}(5), t) = P(\mathrm{Alt}(5), t)$. However, if $G = \mathrm{Sym}(3)$ and $N = \mathrm{Alt}(3) = C_3$, we have that

$$P(\mathrm{Sym}(3), \mathrm{Alt}(3), t) = \left(1 - \frac{3}{3^t}\right) \neq P(C_3, t) = 1 - \frac{1}{3^t}$$

Remark 5 As we have seen above, notice that if $P(G, t)$ is irreducible, then it is easy to see that G is a simple group. However, the converse does not hold. The following example is also due to [1]:

$$P(\mathrm{PSL}(2, 7), \mathrm{t}) = \left(1 - \frac{2}{2^{\mathrm{t}}}\right)\left(1 + \frac{2}{2^{\mathrm{t}}} + \frac{4}{4^{\mathrm{t}}} - \frac{14}{7^{\mathrm{t}}} - \frac{28}{14^{\mathrm{t}}} + \frac{21}{21^{\mathrm{t}}} - \frac{28}{28^{\mathrm{t}}} + \frac{42}{42^{\mathrm{t}}}\right)$$

A question arises naturally is whether there exist examples of groups G such that $P_G(s)$ has a non-trivial factorization which does not come from normal subgroups, in particular, whether $P_G(s)$ is irreducible when G is a simple group. The answer is positive for all abelian simple groups, being $P_{\mathbb{Z}/p\mathbb{Z}}(s) = 1 - 1/p^s$. For nonabelian simple groups, some results were obtained my Damian, Lucchini and Morini in [4] as follows:

Theorem 6

(1) *For any prime $p \geq 5$, the polynomial $P_{\mathrm{Alt}(p)}(s)$ is irreducible.*

(2) *If $p = 2^t - 1$ and $t \equiv 3 \mod 4$ then $P_{\mathrm{PSL(2,p)}}(s)$ is reducible.*

These were extended by recent results of Patassini appeared in [13] and [15] in the following

Theorem 7

(1) *Assume that $k \geq 5$. If $k \leq 4.2 \cdot 10^{16}$ or $k \geq (e^{e^{15}} + 2)^3$, then $P_{\mathrm{Alt}(k)}(s)$ is irreducible. If we assume the Riemann Hypothesis, then $P_{\mathrm{Alt}(k)}(s)$ is always irreducible.*

(2) *Let S be a simple group of Lie type. Then $P_S(s)$ is reducible if and only if $S \cong A_1(p)$ for some Mersenne prime p such that $\log_2(p + 1) \equiv 3 \mod 4$.*

One then may ask which extra relationships between the combinatorial properties of $P(G, s)$ and the structural properties of G itself. There are many beautiful results have been being obtained. The following is about solvable groups:

Theorem 8 *Let G be a finite group. The following are equivalent:*

- *G is (pro)solvable.*
- *$P(G,s)$ can be written as the product of $(1-c_i/q_i^s)$ where $c_i \geq 0$ and q_i is a prime power.*
- *The sequence $\{a_n(G)\}$, where $a_n(G) = \sum_{|G:H|=n} \mu_G(H)$, is multiplicative, i.e., $a_{m.n}(G) = a_n(G)a_m(G)$ whenever m and n are coprime.*

Remark 9 Notice that if G is solvable, then $1/P(G,s)$ is a Dirichlet series with positive coefficients. It does not hold in general for non-solvable group.

Example 10 Let $G = \mathrm{Alt}(5)$ then

$$P(G,s) = 1 - \frac{5}{5^s} - \frac{6}{6^s} - \frac{10}{10^s} + \frac{20}{20^s} + \frac{60}{30^s} - \frac{60}{60^s}$$

while

$$1/P(G,s) = 1 + \frac{5}{5^s} + \frac{6}{6^s} + \frac{10}{10^s} - \frac{20}{20^s} + \cdots$$

Conjecture 11 *Let G be a finite group. Then G is solvable if and only if $1/P(G,s)$ is a Dirichlet series with positive coefficients.*

Another interesting connected problem is about the order complex associated to finite groups, as described as follows. Let $\mathcal{C}(G)$ be the poset consisting of proper right cosets Hx in the finite group G, ordered by inclusion $Hx < Ky$ if and only if $H \leq K$ and $Kx = Ky$. The simplicial complex is denoted by $\Delta(\mathcal{G})$ whose faces are the finite chains $H_1g_1 < H_2g_2 < \cdots < H_kg_k$ of elements of $\mathcal{C}(G)$. The reduced Euler characteristic of $\mathcal{C}(G)$ is defined by $\chi(\Delta(\mathcal{C}(G))) := \sum_n (-1)^n a_n = -1 + a_0 - a_1 + \cdots$, where a_n is the number of chains in $\mathcal{C}(G)$ of length n. Brown noticed in [2] that

Theorem 12 *Let G be a finite group, then*

$$\widetilde{\chi}(\Delta(\mathcal{C}(G))) = -P(G,-1)$$

As known that if $\widetilde{\chi}(\Delta(\mathcal{C}(G))) \neq 0$ then the simplicial complex $\Delta(\mathcal{C}(G))$ is not contractible. Brown then conjectured that

Conjecture 13 *Let G be a finite group. Then $P(G,-1) \neq 0$. That means the simplicial complexes of finite groups are non-contractible.*

He showed in [2] that the conjecture holds for solvable groups. Recently, Patassini showed that the conjecture holds for classical groups, $PSL(2,q)$, Suzuki and Ree groups (see [14, 16]). Other cases are still open. Notice that the conjecture can be verified for sporadic simple groups in GAP.

As we know, for finite groups, probability just means quotient. However, it is not the same for infinite groups. We need measures on groups. Luckily, they exist on compact groups, known there as Haar measure. More precisely, we will be considering profinite groups, i.e., inverse limits of finite groups. Let us first introduce about profinite groups and Haar measures on them.

A *topological group* is a group G which is also a topological space, such that the maps $g \mapsto g^{-1} : G \to G$ and $(g, h) \mapsto gh : G \times G \to G$ are both continuous. An easy example of topological groups are finite groups endowed with discrete topology. A *profinite group* is a compact Hausdorff topological group whose open subgroups forms a base for the neighborhoods of the identity. For such a group G, a subgroup is open if and only if it is closed and has finite index. And so the family of all open subgroups of G intersects in $\{1\}$. Moreover, a subset of G is open if and only if it is a union of cosets of open normal subgroups.

A second definition of profinite groups is based on the concept of an *inverse limit*. We recall briefly what it is. A *directed set* is a non-empty partially ordered set $(I, \leq)$ with the property that for every $i, j \in I$, there exists $k \in I$ such that $k \geq i$ and $k \geq j$. An *inverse system* of sets (or groups, rings or topological spaces) over I is a family of sets (or groups, etc.) $(G_i)_{i\in I}$ with maps (respectively homomorphisms, continous maps) $\phi_{ij} : G_i \to G_j$ whenever $i \geq j$, satisfying $\varphi_{ii} = Id_{G_i}$ and $\varphi_{jk} \circ \varphi_{ij} = \varphi_{ik}$ whenever $i \geq j \geq k$, where "$f \circ g$ means do g first, then f". The *inverse limit*

$$\varprojlim G_i = \varprojlim (G_i)_{i\in I}.$$

is the subset (or subgroup, etc.) of the Cartesian product $\prod_{i\in I} G_i$ consisting of all (g_i) such that $\varphi_{ij}(g_i) = g_j$ whenever $i \geq j$. Hence, if for each i, let π_i be the projection from $\varprojlim G_i$ to G_i, then for $i \geq j$, we have that $\varphi_{ij} \circ \pi_i = \pi_j$. The inverse limit must be universal in the sense that if there is an object Y together with projections $\lambda_i : Y \to G_i$ satisfying $\varphi_{ij} \circ \lambda_i = \lambda_j$ then there is a unique morphism $\phi : Y \to \varprojlim G_i$ such that $\pi_i \circ \phi = \lambda_i$ for each $i \in I$.

If each G_i is a finite group endowed with discrete topology and $\prod_{i\in I} G_i$ is given the product topology, then $\varprojlim G_i$ becomes a topological group, and this topological group is profinite.

If I is a family of normal subgroups of a given group G which is closed under taking intersection, we may order I by reverse inclusion, i.e., $N \geq M$ whenever $N \subseteq M$, and obtain an inverse system $(G/N)_{N\in I}$. And so the maps $\varphi_{N,M}$ are the natural epimorphisms $G/N \to G/M$ for $N \subseteq M$. We now come to the identicality of two definitions of profinite groups

Proposition 14 *If G is a profinite group then G is (topologically) isomorphic to $\varprojlim_{N \lhd_o G} (G/N)$. Conversely, the inverse limit of any inverse system of finite groups is a profinite group.*

Note that $\varprojlim (G/N)$, where N ranges over all normal subgroups of G of finite index, is called the *profinite completion* of G, denoted by $\widehat{G}$. If G is *residually finite*, i.e., the intersection of all above N's is trivial, then G is embedded into its profinite completion.

Equivalently, a profinite group is an inverse limit of finite groups (see [6, Proposition 1.3, p. 17]).

A typical example for a profinite group is $\mathbb{Z}_p$, the group of p-adic integers, where p is a fixed prime. We can express the expansion of elements in $\mathbb{Z}_p$ as

$$\mathbb{Z}_p = \varprojlim_n \mathbb{Z}/p^n\mathbb{Z} = \left\{ (x_n)_{n\geq 0} \in \prod_{n\geq 0} \mathbb{Z}/p^n\mathbb{Z} : \text{for all } n,\ x_{n+1} \equiv x_n \mod p^n \right\}.$$

We also define the ring $\widehat{\mathbb{Z}}$ to be the profinite completion of $\mathbb{Z}$. That is

$$\widehat{\mathbb{Z}} = \varprojlim_n \mathbb{Z}/n\mathbb{Z} = \left\{ (x_n)_{n\geq 1} \in \prod_{n=1}^{\infty} \mathbb{Z}/n\mathbb{Z} : \text{for all } n|m, a_m \equiv a_n \mod n \right\}.$$

It is also true that

$$\widehat{\mathbb{Z}} \cong \prod_p \mathbb{Z}_p.$$

Profinite groups take attentions of many number theorists since they in fact arise in number theory as Galois groups of (finite or infinite) Galois extensions of fields, with appropriate topology. Historically, this is the original motivation for the study of profinite groups, and Galois theory remains the main area of applications of results in profinite groups (see [17, Chapter 3]).

When talking about generators of a profinite group, we mean generators as a topological group, i.e., X generates G means that G is the smallest closed subgroup of G containing X. The closure of arbitrary subset X is $\overline{X} = \bigcap XN$, with N ranging over all open normal subgroups of G. If follows that X generates G if and only if each finite factor G/N is generated by XN/N. Thus G is generated, by d elements, say, if and only if each finite factor group G/N can be generated by d elements. So now, let G be a profinite group and μ the normalized Haar measure on G or on some direct power G^k. Now fix k and write

$$X(G,k) = \{(x_1, \cdots, x_k) \in G^k | \overline{\langle x_1, \cdots, x_k \rangle} = G\}$$

to denote the set of all k-tuples topologically generating the group G. We may therefore define

$$P(G,k) = \mu(X(G,k))$$

to be the *probability that a random k-tuple generates G*. Thus $0 \leq P(G,k) \leq 1$, and if $P(G,k) > 0$ then $d(G) \leq k$, where $d(G)$ is the minimal number of generators of G.

Definition 15 A group G is called *positively finitely generated* (PFG) if $P(G,k) > 0$ for some choice of $k \in \mathbb{N}$.

Hence a PFG group is finitely generated. However, for a d-generated group G, it does not always hold that $P(G,d) > 0$. Kantor and Lubotzky showed in [10] that the group $G = \prod_{n\geq n_0} \mathrm{Alt}(n)^{n!/8}$ is 2-generated but $P(G,t) = 0$ for all $t > 0$.

Mann conjectured in [11] that for a PFG group G, the values $P(G,k)$ can be interpolated to an analytic function defined in some right half-plane of the complex plane. He then conjecture in [12] that if G is a PFG group, then the following series

$$P_G(s) := \sum_{H \leq_0 G} \frac{\mu_G(H)}{|G:H|^s}$$

converges absolutely in some right half-complex plane. Moreover, he noticed that, if the later conjecture holds, then $P_G(k) = P(G,k)$ for every $k \geq 1$.

So now we consider, generally, G to be a finitely generated profinite group. For each $n \in \mathbb{N}$, let

$$a_n := \sum_{|G:H|=n} \mu_G(H)$$

and rewrite the series $P_G(s)$ as the following

$$P_G(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}$$

Example 16 Let $G = \widehat{\mathbb{Z}}$ be the profinite completion of $\mathbb{Z}$. Then

$$P_{\widehat{\mathbb{Z}}}(k) = \begin{cases} 1/\zeta(k) & , \ k > 1 \\ 0 & , \ k = 1 \end{cases}$$

We do not know whether the series $P_G(s)$ converges, however, the inverse of $P_G(s)$ is called the *probabilistic zeta function* of G (Boston, Mann).

Hall noticed in [9] that if $\mu_G(H) \neq 0$ then H is an intersection of maximal subgroups. In this case, H contains the Frattini subgroup $\mathrm{Frat}(G)$ which is the intersection of all (closed) maximal subgroups of G. It implies in particular that $P_G(s) = P_{G/\mathrm{Frat}(G)}(s)$.

Example 17 Let G be a finitely generated pro-p-group. Then $G/\mathrm{Fratt}(G) \cong C_p^{d(G)}$, and so

$$P_G(s) = \prod_{i=0}^{d(G)-1} \left(1 - \frac{p^i}{p^s}\right)$$

Conjecture 18 *Let G be a finitely generated profinite group. Then $P_G(s)$ is rational if and only if $G/Frat(G)$ is a finite group.*

It was also shown in [5] by Detomi and Lucchini that the conjecture holds for prosolvable groups. For non-prosolvable groups, several results have been obtained in [5] and [7].

To a finitely generated profinite group G, we associated another Dirichlet series as the following:

$$\zeta_G(s) = \sum_{n=1}^{\infty} \frac{b_n}{n^s}$$

where, for each positive integer n, $b_n := b_n(G)$ is the number of open subgroups of index n in G. The series $\zeta_G(s)$ is called the *subgroup zeta function* of G.

Example 19 Let $G = \widehat{\mathbb{Z}}$, then

$$\zeta_G(s) = \zeta(s)$$

the Riemann zeta function.

So, in this case, $P_{\widehat{\mathbb{Z}}}(s)\zeta_{\widehat{\mathbb{Z}}}(s) = 1$. We say that a group G is ζ-reversible if $P_G(s)\zeta_G(s) = 1$.

One could ask the characterization of ζ-reversible groups. Notice that $P_G(s)\zeta_G(s) = 1$ is equivalent to the following:...

Corollary 20 *If $P_H(s) = P_G(s)$ for any open subgroup H of G then G is ζ-reversible.*

Conjecture 21 *Let G be a profinite group. Then G is ζ-reversible if and only if $P_G(s) = P_H(s)$ for every open subgroup H of G.*

If G is a finitely generated pro-p-group, then $G/\mathrm{Frat}(G) \cong C_p^{d(G)}$ and so $P_G(s) = P_{C_p^{d(G)}}(s)$. Hence, the Conjecture 21 is equivalent to the following

Conjecture 22 *Let G be a finitely generated pro-p-group. Then G is ζ-reversible if and only if $d(G) = d(H)$ for every open subgroup of G.*

The conjecture is still open. Some partial results have been obtained by Damian in Lucchini (see [3]).

References

[1] Nigel Boston, *A probabilistic generalization of the Riemann zeta functio.* Analytic Number Theory, vol. 1 (1996), 155–162.

[2] K.S. Brown, *The coset poset and probabilistic zeta function of a finite group.* J. Algebra 225 (2000), 989–1012.

[3] E. Damian and A. Lucchini, *Profinite groups whose the probabilistic zeta function concides with the subgroup zeta function.* Preprint (2009).

[4] E. Damian, A. Lucchini and F. Morini, *Some properties of the probabilistic zeta function on finite simple groups.* Pacific J. Math. 215/1 (2004), 3–14.

[5] E. Detomi and A. Lucchini, *Profinite groups with a rational probabilistic zeta function.* Journal of Group Theory 9/2 (2006), 203–217.

[6] J.D. Dixon, M.P.F. du Sautoy, A. Mann, and D. Segal, "Analytic pro-p groups". Volume 61 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, second edition, 1999.

[7] Duong Hoang Dung, "Profinite groups with a rational probabilistic zeta function". Ph.D. Thesis, Leiden Universiteit, 2013.

[8] W. Gaschütz, *Die Eulersche Funktion endlicher auflösbarer Gruppen (German).* Illinois J. Math. 3 (1959), 469–476.

[9] P. Hall, *The eulerian functions of a group.* Quart. J. Math. 7/1 (1936), 134–151.

[10] W.M. Kantor and A. Lubotzky, *The probability of generating a finite classical groups.* Geom. Ded. 36 (1990), 67–87.

[11] A. Mann, *Positively finitely generated groups.* Forum Math. 8 (1996), 429–459.

[12] A. Mann, *A probabilistic zeta function for arithmetic groups.* Internat. J. Algebra Comput. 15/5-6 (2005), 1053–1059.

[13] M. Patassini, *On the irreducibility of the Dirichlet polynomial of an alternating group.* Trans. Amer. Math. Soc., to appear.

[14] M. Patassini, *The probabilistic zeta function of $PSL(2,q)$, of the Suzuki groups ${}^2B_2(q)$ and of the Ree groups ${}^2G_2(q)$.* Pacific J. Math. 240/1 (2009), 185–200.

[15] M. Patassini, *On the irreducibility of the Dirichlet polynomial of a simple group of Lie type.* Israel J. Math. 185 (2011), 477–507.

[16] M. Patassini, *On the (non-)contractibility of the order complex of the coset poset of a classical group.* J. Algebra 343 (2011), 37–77.

[17] J.S. Wilson, "Profinite groups". Volume 19 of London Mathematical Society Monographs. New Series. The Clarendon Press Oxford University Press, New York, 1998.

Hardy type inequalities on the cone of monotone sequences

Zhanar Taspaganbetova (*)

Abstract. Weighted Hardy type inequalities restricted to the cones of monotone functions and sequences have been extensively studied in the last decades, especially in view of their applications in the estimation of maximal functions, in the theory of interpolation of operators and in the embedding theory of function spaces. In this talk, we introduce a Hardy type inequality on the cone of non-negative and non-increasing sequences. We give the statement and motivation of the problem. We describe the development and current status of the theory of Hardy type inequalities on the cones of monotone functions and sequences. Moreover, we also present some open problems.

1 Historical notes on the problem

The properties of the cone of monotone sequences of real numbers, of the cone of monotone functions, and several related extremum problems have an important significance in functional analysis, in the problems of the mathematical economics, of the theory of probability and statistics. The approximation numbers of operators, quantitative characteristics of the best approximations of functions, moment sequences of function are monotone sequences of numbers, which carry certain information. Many qualitative properties of this type can be expressed by functional relations of monotone sequences.

It is known that the properties of a class of functions or of a class of sequences of numbers can be obtained from the functional relations of their non-increasing rearrangements, which are monotone functions and monotone sequences, respectively. Therefore, the problem of establishing the various functional relationships on the cone of monotone sequences of numbers is an actual direction of mathematical analysis.

Hardy type inequalities on the cone of monotone functions and sequences have some applications in the investigation of boundedness of operators in Lorentz spaces and in the embedding theory in Lorentz spaces.

Hardy type inequalities on the cone of monotone functions and sequences have been intensively studied during the last two decades. In 1990 M. Ariño and B. Muckenhoupt

(*)The L. N. Gumilyov Eurasian National University, Munaitpasov st., 5, 010008, Astana, Kazakhstan; E-mail: zhanara.t.a@yandex.kz . Seminar held on February 6th, 2013.

[1] obtained a necessary and sufficient condition for the validity of the following inequality

$$\left(\int_0^\infty \left(\frac{1}{t}\int_0^t f(s)ds\right)^q u(t)dt\right)^{\frac{1}{q}} \le C\left(\int_0^\infty f^p(t)v(t)dt\right)^{\frac{1}{p}} \tag{1.1}$$

on the cone of non-negative and non-increasing functions f in the case $1 \le p = q < \infty$ and $u(t) = v(t)$. Previously, such problems were studied by D.W. Boyd [2] in 1967 and by S.G. Krein, Yu.I. Petunin, E.M. Semenov [3].

E. Sawyer [4] has extended the result of M. Ariño and B. Muckenhoupt to the case of different weights v and u and $1 < p, q < \infty$. Nowadays this result of E. Sawyer is known as the Sawyer duality principle. Moreover, E. Sawyer [4] has used this duality result to obtain necessary and sufficient conditions for which (1.1) holds for all non-negative and non-increasing functions f in the case $1 < p, q < \infty$. This result of E. Sawyer was extended by V.D. Stepanov [5] to the cases $0 < q < 1 < p < \infty$ and $0 < p \le q < \infty$, $0 < p < 1$. M.L. Goldman [6], G. Bennett and K.-G. Grosse-Erdmann [7] have characterized the weights u and v, for which inequality (1.1) holds for all non-negative and non-increasing functions f in the case $0 < q < p < 1$. The duality principle for the case $0 < p \le 1$ has been proved in [5], [8] and [9]. A simpler proof of the Sawyer duality principle has been obtained by V.D. Stepanov [5], M.J. Carro and J. Soria [8]. Some generalizations of Sawer duality formula were proved by D.E. Edmunds, R. Kerman, L. Pick [10] and A. Kamińska, M. Mastylo [11].

At the same time the investigation and generalization of the discrete Hardy inequality

$$\left(\sum_{i=1}^\infty u_i\left(\frac{1}{i}\sum_{j=1}^i f_j\right)^q\right)^{\frac{1}{q}} \le C\left(\sum_{i=1}^\infty v_i {f_i}^p\right)^{\frac{1}{p}} \tag{1.2}$$

on the cone of monotone sequences $f \ge 0$ was developed. Results on weighted Hardy inequalities on the cone of monotone sequences have been obtained by K.F. Andersen, H.P. Heinig [12], H.P. Heinig [13], L. Leindler [14], M.Sh. Braverman, V.D. Stepanov [15], J. Nemeth [16], F.P. Cass, W. Kratz [17], P.D. Johnson, R.N. Mohapatra, David Ross [18], R. Oinarov, S.Kh. Shalgynbaeva [19], G. Bennett, K.-G. Grosse-Erdmann [7], M.L. Goldman [6], [20], [21], S.Kh. Shalgynbaeva [22] and others.

In 1998 R. Oinarov, S.Kh. Shalgynbaeva [19] obtained an analogue of the Sawyer duality principle for the discrete case if $1 < p, q < \infty$. This result of R. Oinarov, S.Kh. Shalgynbaeva allows to reduce a Hardy type inequality on the cone of monotone sequences to a corresponding inequality on the cone of non-negative sequences from $l_{p,v}$. Moreover, R. Oinarov, S.Kh. Shalgynbaeva [19] have obtained criteria for the validity of inequality (1.2) for some other values of the parameters p and q.

In 2006 G. Bennett and K.-G. Grosse-Erdmann [7] obtained a complete characterization of the weights for which the Hardy inequality (1.2) holds on the cone of monotone sequences of different nature of the conditions of S.Kh. Shalgynbaeva.

M.L. Goldman in his papers has studied inequalities of the type (1.2) on the cone of monotone sequences and has applied the corresponding results to establish Hardy inequalities on the cone of quasi-monotone sequences, see e.g. [6], [20], [21].

Nowadays inequalities on the cone of monotone functions and sequences are still being intensively developed. This fact is confirmed by a great number of recent publications. For a history of Hardy type inequalities on the cones of monotone functions and sequences and for references to related results we refer to the book of A. Kufner, L. Maligranda and L.-E. Persson [23, Chapter 1], and to the PhD thesis of O. Popova [24].

2 Introduction

Let $1 < p, q < \infty$, $\frac{1}{p} + \frac{1}{p'} = 1$ and $u = \{u_i\}_{i=1}^{\infty},\ v = \{v_i\}_{i=1}^{\infty}$ be positive sequences of real numbers. Let $l_{p,v}$ be the space of sequences $f = \{f_i\}_{i=1}^{\infty}$ of real numbers such that

$$\|f\|_{p,v} := \left(\sum_{i=1}^{\infty} |v_i f_i|^p\right)^{\frac{1}{p}} < \infty, \qquad 1 < p < \infty.$$

We consider an inequality of the following form

$$\left(\sum_{i=1}^{\infty} u_i^q \left(\sum_{j=1}^{i} a_{i,j} f_j\right)^q\right)^{\frac{1}{q}} \le C \left(\sum_{i=1}^{\infty} v_i^p {f_i}^p\right)^{\frac{1}{p}} \tag{2.1}$$

on the cone of non-negative and non-increasing sequences $f = \{f_i\}_{i=1}^{\infty}$ of $l_{p,v}$, where C is a positive constant independent of f and $(a_{i,j})$ is a non-negative triangular matrix with entries $a_{i,j} \ge 0$ for $i \ge j \ge 1$ and $a_{i,j} = 0$ for $i < j$.

In [22] S.Kh. Shalgynbaeva has obtained necessary and sufficient conditions for the validity of (2.1) on the cone of monotone sequences for $1 < p \le q < \infty$ under the assumption that there exists $d \ge 1$ such that the inequalities

$$\frac{1}{d}(a_{i,k} + a_{k,j}) \le a_{i,j} \le d(a_{i,k} + a_{k,j}), \qquad i \ge k \ge j \ge 1 \tag{2.2}$$

hold.

Notation: If M and K are real valued functionals of sequences, then we understand that the symbol $M \ll K$ means that there exists $c > 0$ such that $M \le cK$, where c is a constant which may depend only on parameters such as p, q and r_n, but not on the sequences in the arguments of M and K. If $M \ll K \ll M$, then we write $M \approx K$.

3 Preliminaries and notation

In [25], the classes $\mathcal{O}_n^+$ and $\mathcal{O}_n^-$ of matrices $(a_{i,j})$ were defined for $n \ge 1$. Now we will give equivalent definitions of such classes.

We denote the elements of $\mathcal{O}_n^+$ or of $\mathcal{O}_n^-$ by the symbol $(a_{i,j}^{(n)})$.

We define the classes $\mathcal{O}_n^+$, $n \ge 0$ by induction. Let $(a_{i,j})$ be a matrix which is non-negative and non-decreasing in the first index for all $i \ge j \ge 1$. The class $\mathcal{O}_0^+$ consists of

the matrices of the type $a_{i,j}^{(0)} = \alpha_j$, $i \geq j \geq 1$. Let the classes $\mathcal{O}_\gamma^+$, $\gamma = 0, 1, \ldots, n-1$, $n \geq 1$ be defined. By definition, the matrix $(a_{i,j}) \equiv (a_{i,j}^{(n)})$ belongs to the class $\mathcal{O}_n^+$ if and only if there exist matrices $(a_{i,j}^{(\gamma)}) \in \mathcal{O}_\gamma^+$, $\gamma = 0, 1, \ldots, n-1$ and matrices $(b_{i,k}^{n,\gamma})$, $\gamma = 0, 1, \ldots, n$ such that

$$a_{i,j}^{(n)} \approx \sum_{\gamma=0}^{n} b_{i,k}^{n,\gamma} a_{k,j}^{(\gamma)} \tag{3.1}$$

for all $i \geq k \geq j \geq 1$, where $b_{i,k}^{n,n} \equiv 1$.

As above, we introduce the classes $\mathcal{O}_m^-$, $m \geq 0$. Let $(a_{i,j})$ be a matrix which is non-negative and non-increasing in the second index for all $i \geq j \geq 1$. By definition a matrix $(a_{i,j}) = (a_{i,j}^{(0)})$ belongs to the class $\mathcal{O}_0^-$ if and only if it has the form $a_{i,j}^{(0)} = \beta_i$ for all $i \geq j \geq 1$. Let the classes $\mathcal{O}_\gamma^-$, $\gamma = 0, 1, \ldots, m-1$, $m \geq 1$ be defined. A matrix $(a_{i,j}) = (a_{i,j}^{(m)})$ belongs to the class $\mathcal{O}_m^-$ if and only if there exist matrices $(a_{i,j}^{(\gamma)}) \in \mathcal{O}_\gamma^-$, $\gamma = 0, 1, \ldots, m-1$ and matrices $(d_{k,j}^{\gamma,m})$, $\gamma = 0, 1, \ldots, m$ such that

$$a_{i,j}^{(m)} \approx \sum_{\gamma=0}^{m} a_{i,k}^{(\gamma)} d_{k,j}^{\gamma,m} \tag{3.2}$$

for all $i \geq k \geq j \geq 1$, where $d_{k,j}^{m,m} \equiv 1$.

It is obvious that the classes $\mathcal{O}_1^\pm$ include the matrices, whose entries satisfy conditions (2.2). This implies that the classes $\mathcal{O}_n^+$, $n \geq 1$ and $\mathcal{O}_m^-$, $m \geq 1$ are wider than the classes of matrices which have been used in this connection before.

Such classes of operators include a lot of well-known classical operators such as the operator of multiple summation, Hölder's operator, Cesàro operator and others.

A continuous analogue of the classes $\mathcal{O}_n^+$ and $\mathcal{O}_n^-$, $n \geq 0$ has been studied by R. Oinarov in [26].

4 Main result

We define

$$V_k = \sum_{i=1}^{k} v_i^p, \quad A_{ik} = \sum_{j=1}^{k} a_{i,j}, \quad E_1 = \sup_{s \geq 1} V_s^{-\frac{1}{p}} \left(\sum_{i=1}^{s} A_{ii}^q u_i^q \right)^{\frac{1}{q}},$$

$$E_2 = \sup_{s \geq 1} \left(\sum_{k=1}^{s} \left(V_k^{-\frac{p'}{p}} - V_{k+1}^{-\frac{p'}{p}} \right) \left(\sum_{i=s}^{\infty} A_{ik}^q u_i^q \right)^{\frac{p'}{q}} \right)^{\frac{1}{p'}},$$

$$E_3 = \sup_{s \geq 1} \left(\sum_{k=s}^{\infty} u_k^q \left(\sum_{i=1}^{s} A_{ki}^{p'} \left(V_i^{-\frac{p'}{p}} - V_{i+1}^{-\frac{p'}{p}} \right) \right)^{\frac{q}{p'}} \right)^{\frac{1}{q}}.$$

Theorem 4.1 *Let $1 < p \le q < \infty$. Let the matrix $(a_{i,j})$ in (2.1) belong to the class $\mathcal{O}_m^+ \cup \mathcal{O}_m^-$, $m \ge 0$. Then the inequality (2.1) on the cone of non-negative and non-increasing sequences $f \in l_{p,v}$ holds if and only if at least one of the conditions $E_{12} = \max\{E_1,\ E_2\} < \infty$ and $E_{13} = \max\{E_1,\ E_3\} < \infty$ holds. Moreover, $E_{12} \approx E_{13} \approx C$, where C is the best constant in (2.1).*

Acknowledgement. The author thanks University Degli Studi di Padova for opportunity to have an internship, as well as expresses her gratitude to Professors Ryskul Oinarov and Massimo Lanza de Cristoforis for providing cooperative support and valuable assistance.

The paper was done under financial support by the Scientific Committee of RK MES, Grant No.1529/GF on priority area "Intellectual potential of the country".

References

[1] M. Ariño, B. Muckenhoupt, *Maximal functions on classical Lorentz spaces and Hardy's inequality with weights for non-increasing functions.* Trans. Amer. Math. Soc. 320 (1990), No. 2, pp. 727–735.

[2] D.W. Boyd, *The Hilbert transform on rearrangement-invariant spaces.* Canad. J. Math. 19 (1967), 599–616.

[3] S.G. Krein, Yu.I. Petunin, E.M. Semenov, "Interpolation of linear operators". Transl. Math. Monographs 54, Amer. Math. Soc., Providence, 1982.

[4] E. Sawyer, *Boundedness of classical operators on classical Lorentz spaces.* Studia Math. 96 (1990), No. 2, 145–158.

[5] V.D. Stepanov, *The weighted Hardy's inequality for non-increasing functions.* Trans. Amer. Math. Soc. 338/1 (1993), 173-186.

[6] M.L. Goldman, *Sharp estimates for the norms of Hardy-type operators on cones of quasimonotone functions.* Proc. Steklov Inst. Math. 232/1 (2001), 109–137.

[7] G. Bennett, K.-G. Grosse-Erdmann, *Weighted Hardy inequalities for decreasing sequences and functions.* Math. Ann. 334/3 (2006), 489–531.

[8] M.J. Carro, J. Soria, *Boundedness of some integral operators.* Canad. J. Math. 45/6 (1993), 1155–1166.

[9] H.P. Heinig, L. Maligranda, *Interpolation with weights in Orlicz spaces.* Boll. Un. Mat. Ital. B (7) 8/1 (1994), 37–55.

[10] D.E. Edmunds, R. Kerman, L. Pick, *Optimal Sobolev imbeddings involving rearrangement-invariant quasinorms.* J. Funct. Anal. 170 (2000), 307–355.

[11] A. Kamińska, M. Mastylo, *Duality and classical operators in function spaces.* Technical Report, Adam Mickiewicz University 112 (2001), 1–30.

[12] K.F. Andersen, H.P. Heinig, *Weighted norm inequalities for certain integral operators.* SIAM J. Math. 14 (1983), 834–844.

[13] H.P. Heinig, *Weighted norm inequalities for certain integral operator. II.* Proc. Amer. Math. Soc. 95 (1985), 387–395.

[14] L. Leindler, *Generalization of inequalities of Hardy and Littlewood.* Acta Sci. Math. 31 (1970), 279–285.

[15] M.Sh. Braverman and V.D. Stepanov, *On the discrete Hardy inequality.* Bull. London Math. Soc. 26/3 (1994), 283–287.

[16] J. Nemeth, *Generalizations of the Hardy-Littlewood inequality. II.* Acta Sci. Math. (Szeged) 35 (1994), 127–134.

[17] F.P. Cass, W. Kratz, *Nörlund and weighted mean matrices as bounded operators on l_p.* Rocky Mountain J. Math. 20 (1990), 59–74.

[18] P.D. Johnson, R.N. Mohapatra, David Ross, *Bounds for the operator norms of some Nörlund matrices.* Proc. Amer. Math. Soc. 124/2 (1996), 543–547.

[19] R. Oinarov, S.Kh. Shalgynbaeva, *Weighted Hardy inequalities on the cone of monotone sequences.* Izvestiya NAN RK, serial Phys.-Mat. 1 (1998), 33–42.

[20] M.L. Goldman, *Hardy type inequalities on the cone of quasimonotone functions.* Res. Rep. 98/31. Khabarovsk: Russ. Acad. Sci. Far-East Branch Comput. Center, 1998.

[21] M.L. Goldman, *Order-sharp estimates for Hardy-type operators on cones of quasimonotone functions.* Eurasian Math. J. 2/3 (2011), 143–146.

[22] S.Kh. Shalgynbaeva, *Weighted estimate for a class of matrices on the cone of monotone sequences.* Izvestiya NAN RK, serial Phys.-Mat. 5 (1998), 76–80.

[23] A. Kufner, L. Maligranda and L-E. Persson, "The Hardy Inequality. About its History and Some Related Results". Vydavatelský Servis, Plzeň, 2007.

[24] O. Popova, "Weighted Hardy-type inequalities on the cones of monotone and quasi-concave functions". PhD thesis, Luleå University of Technology, SE-971 87 Luleå, Sweden and Peoples' Friendship University of Russia, Moscow 117198, Russia, 2012.

[25] R. Oinarov, Zh. Taspaganbetova, *Criteria of boundedness and compactness of a class of matrix operators.* Journal of Inequalities and Applications 2012:53 (2012), 1–18.

[26] R. Oinarov, *Boundedness and compactness of integral operators of Volterra type.* Sibirskii Matematicheskii Zhurnal 48/5 (2007), 1100-1115.

Ideas in finite group theory

Martino Garonzi (*)

Abstract. In this note I present some of the main ideas of finite group theory, starting with examples of non-abelian groups (groups of matrices and groups of permutations), going to Galois theory, i.e. the way polynomials and groups interact, and finally simple groups, solvable groups and their role in understanding when the roots of a polynomial can be expressed by starting with the coefficients and performing sums, differences, products, divisions and root extractions. After that I will present my research topic with examples and some results.

When dealing with operations there are two possible notations, the additive notation and the multiplicative notation. In the additive notation the operation between two group elements a, b is denoted $a + b$ and the identity element is denoted 0. In the multiplicative notation the operation between two group elements a, b is denoted $a \cdot b$ or simply ab and the identity element is denoted 1. Usually the additive notation is reserved for the abelian case. I will mostly use the multiplicative notation. I will assume the reader to be familiar with the basic properties of groups and fields. The notation $H \leq G$ means that H is a subgroup of G, and the notation $H \trianglelefteq G$ means that H is a normal subgroup of G.

The given bibliography provides good reference books for the theory of (finite) groups.

I will start by recalling the isomorphism theorem.

Theorem 1 (Isomorphism Theorem) *Let $\varphi : G \to H$ be a group homomorphism and let $N := \ker(\varphi)$ be the kernel of φ, i.e. the set of elements $g \in G$ such that $\varphi(g) = 1$. Then $G/N \cong \varphi(G)$ via the canonical isomorphism $gN \mapsto \varphi(g)$.*

1 Some examples of groups

Usually abelian groups (commutative groups), i.e. groups in which any two elements a, b verify $ab = ba$ ("commute"), are familiar to every mathematician. Let us start with some examples of non-abelian groups.

(*)Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy; E-mail: mgaronzi@math.unipd.it . Seminar held on February 27th, 2013.

1.1 Matrices

Invertible matrices form a group. Let F be any field (for example $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\mathbb{Z}/p\mathbb{Z}$). I denote by $GL(n, F)$ the set of $n \times n$ invertible matrices with entries in the field F. The usual row-column multiplication gives $GL(n, F)$ the structure of a group, which is non-abelian if $n \geq 2$. It is usually called the "General Linear Group". It is non-abelian because, as it is well-known, the row-column multiplication is not a commutative operation. The set $F - \{0\} = F^*$ is a group with respect to multiplication, and it is abelian, isomorphic to $GL(1, F)$. Taking the determinant provides a (surjective!) group homomorphism

$$GL(n, F) \to F^*, \qquad A \mapsto \det(A)$$

(indeed $\det(AB) = \det(A)\det(B)$ by Binet's theorem) whose kernel, $SL(n, F) := \{A \in GL(n, F) \ : \ \det(A) = 1\}$, is called the "Special Linear Group". It is a normal subgroup of $GL(n, F)$ (being the kernel of a homomorphism) and using the isomorphism theorem we see that the quotient $GL(n, F)/SL(n, F)$ is isomorphic to F^*.

1.2 Permutations

I will denote by $\mathrm{Sym}(X)$ the set of bijections $X \to X$ (also called the permutations of X). The operation of usual composition of functions gives $\mathrm{Sym}(X)$ the structure of group. It is called the "Symmetric Group" of X. It is an easy exercise to show that if X, Y are equipotent sets then $\mathrm{Sym}(X)$ and $\mathrm{Sym}(Y)$ are isomorphic groups. If $X = \{1, \ldots, n\}$ I shall denote $\mathrm{Sym}(X)$ by $\mathrm{Sym}(n)$ or S_n. It is called the symmetric group of degree n. This group is non-abelian if and only if $n \geq 3$. An element of $\mathrm{Sym}(n)$ is called permutation of $\{1, \ldots, n\}$. The order of $\mathrm{Sym}(n)$ (its size as a set) is $n! = 1 \cdot 2 \cdots n$. I will use the standard cycle notation, which is best explained by means of examples:

$$(123)(4567): \qquad 1 \mapsto 2 \mapsto 3 \mapsto 1, \qquad 4 \mapsto 5 \mapsto 6 \mapsto 7 \mapsto 4.$$

$$(123 \cdots k): \qquad 1 \mapsto 2 \mapsto 3 \mapsto \cdots \mapsto k \mapsto 1 \qquad k\text{-cycle}.$$

Composition goes as follows:

$$(12)(234)(13) = (234), \qquad (143)(1352)(4312) = (13)(45).$$

Note that disjoint cycles always commute. The following calculation shows that $\mathrm{Sym}(n)$ is non-abelian for $n \geq 3$:

$$(12)(123) = (13), \qquad (123)(12) = (23).$$

Remark 1 Every permutation can be written uniquely (up to reordering) as **product of disjoint cycles**.

2-cycles are also called "transpositions". A permutation is called "even" (or "of sign 1") if it can be written as the product of an even number of transpositions, and "odd" (or "of sign -1") otherwise. For example $(12)(25)(13)(35)$ is even, $(13)(26)(43)$ is odd. The

identity of $\mathrm{Sym}(n)$ (the identity function $\{1,\dots,n\} \to \{1,\dots,n\}$) is considered to be the product of zero transpositions, hence an even permutation.

Remark 2 A product of disjoint cycles is an even permutation if and only if the number of cycles of even length is even.

For example $(123)(4567)$, $(12)(3456)(78)$ are odd, $(123)(45)(67)$, $(123)(4567)(89)$ are even.

Definition 1 The "cycle structure" of a permutation is the increasing sequence of the cycle lengths in the representation as a product of disjoint cycles. Cycles of length 1 are usually omitted.

So for example $(123)(4567)$, $(12)(3456)(78)$, $(123)(45)(67)$, $(123)(4567)(89)$ have cycle structure respectively $(3,4)$, $(2,2,4)$, $(2,2,3)$, $(2,3,4)$. Remark 2 implies that the cycle structure of a permutation determines its sign. Hence all elements of cycle structure $(3,4)$, $(2,2,4)$ are odd, and all elements of cycle structure $(2,2,3)$, $(2,3,4)$ are even.

Let us denote by C_2 the set $\{-1,1\}$ with the operation given by multiplication: $1\cdot 1 = (-1)\cdot(-1) = 1$ and $1\cdot(-1) = (-1)\cdot 1 = -1$. Then C_2 is a commutative group, it is a cyclic group (a group generated by one element) of order 2 (generated by -1), and it is isomorphic to $\mathrm{Sym}(2) = \{1,(12)\}$. Consider the map

$$\mathrm{sgn} : \mathrm{Sym}(n) \to \{-1,1\} = C_2, \qquad \sigma \mapsto \mathrm{sgn}(\sigma)$$

which sends any permutation to its sign (1 if it is even, -1 if it is odd). Then sgn is a (surjective!) group homomorphism whose kernel, $\mathrm{Alt}(n) = A_n := \{\sigma \in \mathrm{Sym}(n) : \mathrm{sgn}(\sigma) = 1\}$ is called the "Alternating Group" of degree n. It is a normal subgroup of $\mathrm{Sym}(n)$ (being the kernel of a homomorphism) and using the isomorphism theorem we see that the quotient $\mathrm{Sym}(n)/\mathrm{Alt}(n)$ is isomorphic to C_2, in particular $|\mathrm{Alt}(n)| = n!/2$.

For example

$$\mathrm{Alt}(4) = \{1,(12)(34),(13)(24),(14)(23),(123),(132),(124),(142),(234),(243),(134),(143)\}.$$

2 Galois Theory

2.1 The Galois group of a polynomial

Let us talk about the reason why groups were invented.

To each polynomial $f(X) \in \mathbb{Q}[X]$ without multiple roots can be attached a finite group G_f, called the **Galois group** of the polynomial (named after **Evariste Galois, 1811 - 1832**). It is the group defined as follows: if $a_1,\dots,a_n \in \mathbb{C}$ denote the distinct roots of $f(X)$ then

$$G_f = \mathrm{Aut}(\mathbb{Q}(a_1,\dots,a_n))$$

where $\mathbb{Q}(a_1,\dots,a_n)$ denotes the field generated by $a_1,\dots,a_n$, i.e. the intersection of the subfields of $\mathbb{C}$ containing $a_1,\dots,a_n$. That is, G_f is the group of field isomorphisms

$$\mathbb{Q}(a_1,\dots,a_n) \to \mathbb{Q}(a_1,\dots,a_n),$$

that is, the group of such maps which are bijective and respect identity elements, sums, and products. It goes without saying that the operation in G_f is again the usual composition of functions. Suppose that a is a root of $f(X) \in \mathbb{Q}[X]$ (i.e. $f(a) = 0$) and $g \in G_f$. We can consider the element $g(a) \in \mathbb{C}$. Since g is a ring homomorphism, it fixes every element of $\mathbb{Q}$ (this is easy to show starting from the identity $g(n) = g(1+\cdots+1) = g(1)+\cdots+g(1) = 1+\cdots+1 = n$ for $n \in \mathbb{N}$) and also $g(a)$ is a root of f, indeed writing $f(X) = \sum_i c_i X^i$ with c_i elements of $\mathbb{Q}$ we have

$$\begin{aligned} f(g(a)) &= \sum_i c_i g(a)^i = \sum_i g(c_i)g(a^i) = \sum_i g(c_i a^i) = \\ &= g(\sum_i c_i a^i) = g(f(a)) = g(0) = 0. \end{aligned}$$

This implies that the group G_f **permutes the roots of** f. In other words, G_f can be described (or better, "represented") as a subgroup of $\mathrm{Sym}(n)$. Indeed, the function

$$G_f \to \mathrm{Sym}(\{a_1, \ldots, a_n\})$$

which sends $g \in G_f$ to the permutation given by $a_i \mapsto g(a_i)$ (which, as we saw above, is well-defined) is an injective (!) group homomorphism. Injectivity follows from the fact that the only element of G_f which fixes all the roots is the identity.

For example the Galois group of $X^2 - 2$ is the automorphism group of the field $\mathbb{Q}(\sqrt{2})$, so it consists of two elements: the identity and the (unique!) field homomorphism $\tau : \mathbb{Q}(\sqrt{2}) \to \mathbb{Q}(\sqrt{2})$ which sends $\sqrt{2}$ to $-\sqrt{2}$. The Galois group of $X^2 - 2$ is cyclic of order 2. $G_f \cong \mathrm{Sym}(2) \cong C_2$.

2.2 Factorizations modulo prime numbers

I now want to show how polynomials and groups interact. The usual "reduction modulo n" that we have for integers can be performed also in polynomial rings. It is an interesting fact that a polynomial might be irreducible over $\mathbb{Z}$ but reducible modulo some prime p. For example $X^2 - 2$ is irreducible over $\mathbb{Z}$ (its roots are not integers) but

$$X^2 - 2 \equiv X^2 \mod 2, \qquad X^2 - 2 \equiv (X-3)(X-4) \mod 7.$$

Still there are some primes for which the given polynomial might remain irreducible, for example $X^2 - 2$ is irreducible modulo 3.

For the following theorem see [6], Lemmas 1 and 2.

Theorem 2 (Frobenius-Dedekind) *Let $f(X)$ be an irreducible polynomial of $\mathbb{Z}[X]$ of degree n. The following assertions are equivalent.*

- *There exists a prime p for which $f(X)$ mod p does not admit multiple irreducible factors (such prime is usually called "unramified") and the factorization pattern of $f(X)$ mod p is $(n_1, \ldots, n_t)$ (meaning that there are t irreducible factors of degrees $n_1, \ldots, n_t$).*

- *The Galois group of $f(X)$, seen as a (transitive) subgroup of Sym(n), contains an element of cycle structure $(n_1, \dots, n_t)$.*

This implies, for example, that an irreducible polynomial of degree n remains irreducible modulo some prime if and only if its Galois group, viewed as a subgroup of $\mathrm{Sym}(n)$, contains an n-cycle. Moreover, since every group contains the identity element, which has cycle structure $(1, \dots, 1)$, we deduce that given an irreducible polynomial there always exist primes p such that $P(X)$ splits into distinct linear factors modulo p (!). Such primes are infinitely many by Chebotarev's density theorem (cf. below).

There is a notion of "discriminant" valid for every polynomial. It is defined as $\prod_{i,j}^{n}(a_i - a_j)$ where $a_1, \dots, a_n$ are the roots of the polynomial. It is possible to show that the discriminant of a polynomial with integer coefficients is an integer. For small degrees it is reasonable to find a formula for the discriminant. For example,

- the discriminant of $aX^2 + bX + c$ is $b^2 - 4ac$,
- the discriminant of $X^3 + pX + q$ is $-4p^3 - 27q^2$.

A very useful property of the discriminant (which follows directly from its definition) is the following: given a polynomial $P(X)$, a prime number is ramified (i.e. $P(X)$ has multiple roots modulo that prime) if and only if it divides the discriminant of $P(X)$. In particular, there are only finitely many ramified primes.

In the case of cubics (polynomials of degree 3) the discriminant determines the Galois group: it is possible to show that an irreducible polynomial of degree 3 over $\mathbb{Q}$ has Galois group Sym(3) if its discriminant is a square in $\mathbb{Q}$, it has Galois group Alt(3) otherwise.

Let us consider the following examples:

- $X^3 + X^2 + X + 3$ (discriminant $-204 = -2^2 \cdot 3 \cdot 17$);
- $X^3 - 3X + 1$ (discriminant $81 = 3^4$).

The following tables contain the reductions of these polynomials modulo various prime numbers. Note that the factorization pattern $(1, 2)$ shows up in the first table but does not in the second. This is explained by Frobenius theorem: Sym(3) (the Galois group of $X^3 + X^2 + X + 3$) contains permutations of cycle structure $(1, 2)$ (2-cycles) but Alt(3) (the Galois group of $X^3 - 3X + 1$) does not! Indeed

$$\mathrm{Sym}(3) = \{1, (12), (13), (23), (123), (132)\}, \qquad \mathrm{Alt}(3) = \{1, (123), (132)\}.$$

Moreover, by a result known as "Chebotarev density theorem", the proportion of primes yielding a given factorization pattern equals the proportion of elements of the Galois group with the associated cycle structure. This is why, for example, the proportion of primes for which the second polynomial remains irreducible is about $2 : 1$: because in Alt(3) there are twice more 3-cycles than $(1, 1, 1)$-cycles.

p	X^3+X^2+X+3	p	X^3+X^2+X+3
2	$(X+1)^3$	53	$(X+43)(X^2+11X+5)$
3	$X(X+2)^2$	59	$(X+12)(X^2+48X+15)$
5	X^3+X^2+X+3	61	$(X+6)(X^2+56X+31)$
7	$(X+4)(X^2+4X+6)$	67	$(X+23)(X+52)(X+60)$
11	X^3+X^2+X+3	71	$(X+38)(X+52)(X+53)$
13	X^3+X^2+X+3	73	$(X+34)(X^2+40X+28)$
17	$(X+5)(X+15)^2$	79	$(X+74)(X^2+6X+31)$
19	X^3+X^2+X+3	83	$(X+45)(X^2+39X+72)$
23	X^3+X^2+X+3	89	$(X+32)(X^2+58X+14)$
29	$(X+11)(X+23)(X+25)$	97	$(X+59)(X^2+39X+28)$
31	$(X+15)(X^2+17X+25)$	101	$(X+75)(X^2+27X+97)$
37	$(X+25)(X^2+13X+9)$	103	X^3+X^2+X+3
41	X^3+X^2+X+3	107	X^3+X^2+X+3
43	X^3+X^2+X+3	113	X^3+X^2+X+3
47	$(X+31)(X^2+17X+38)$	127	X^3+X^2+X+3

p	X^3-3X+1	p	X^3-3X+1
2	X^3+X+1	53	$(X+18)(X+39)(X+49)$
3	$(X+1)^3$	59	$X^3+56X+1$
5	X^3+2X+1	61	$X^3+58X+1$
7	X^3+4X+1	67	$X^3+64X+1$
11	X^3+8X+1	71	$(X+16)(X+25)(X+30)$
13	$X^3+10X+1$	73	$(X+14)(X+25)(X+34)$
17	$(X+3)(X+4)(X+10)$	79	$X^3+76X+1$
19	$(X+10)(X+12)(X+16)$	83	$X^3+80X+1$
23	$X^3+20X+1$	89	$(X+12)(X+36)(X+41)$
29	$X^3+26X+1$	97	$X^3+94X+1$
31	$X^3+28X+1$	101	$X^3+98X+1$
37	$(X+14)(X+28)(X+32)$	103	$X^3+100X+1$
41	$X^3+38X+1$	107	$(X+7)(X+40)(X+60)$
43	$X^3+40X+1$	113	$X^3+110X+1$
47	$X^3+44X+1$	127	$(X+53)(X+87)(X+114)$

2.3 The Inverse Galois Problem

The most famous open problem in group theory is probably the Inverse Galois Problem.

Is it true that for any finite group G there exists a polynomial $f(X) \in \mathbb{Q}[X]$ with $G_f \cong G$?

This problem has been solved for abelian groups (even *solvable* groups, cf. below for the definition), but the answer in general is not known.

3 Cauchy, Lagrange, Sylow, Cayley

Let us list the important results of "elementary" finite group theory. Given a subset X of a group G I will denote by $\langle X \rangle$ the **subgroup generated** by X in G, i.e. the intersection of the subgroups of G containing X. I will rather write $\langle x_1, \ldots, x_n \rangle$ instead of $\langle \{x_1, \ldots, x_n\} \rangle$.

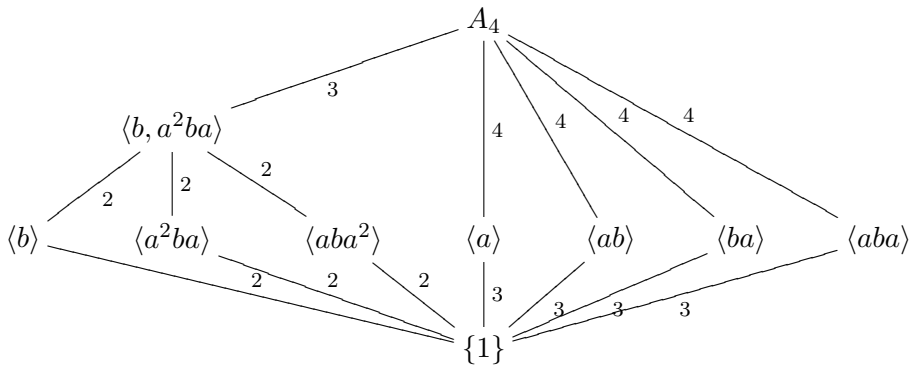
- the **"order" of an element** $g \in G$, denoted $o(g)$, is the smallest positive integer n such that the product of g with itself n times, $g^n = g \cdots g$ (n times), equals 1;
- the **"order" of a subgroup** $H \leq G$, denoted $|H|$, is its size.
- It turns out that $|\langle g \rangle| = o(g)$.

Theorem 3 (Lagrange (1736 - 1813)) *Let G be a finite group, and let $H \leq G$. Then $|H|$ divides $|G|$. The integer $|G|/|H| = |G : H|$ is called the "**index**" of H in G.*

Not every divisor of $|G|$ necessarily equals the size of a subgroup of G (cf. the example below), but...

Theorem 4 (Cauchy (1789 - 1857)) *Let G be a finite group, and let p be a prime dividing $|G|$. Then there exists $g \in G$ of order p.*

Consider the following example: the alternating group of degree 4. $A_4 = \langle a, b \rangle$ where $a = (123)$ and $b = (12)(34)$. $|A_4| = 4!/2 = 12 = 2^2 \cdot 3$. Here is its subgroup lattice (the labelling numbers denote the indeces):



$A_4 = \{1, (12)(34), (13)(24), (14)(23), (123), (132), (124), (142), (234), (243), (134), (143)\}$.

Note that although 6 divides 12, A_4 **has no subgroups of order** 6.

Although it is not true that there exist subgroups of G of order any given divisor of $|G|$, Cauchy's theorem implies that they exist if the given divisor is a prime. The next natural step is to ask what happens with prime-powers. Suppose $|G|$ is divisible by a prime-power p^k. Can we always find a subgroup $H \leq G$ with $|H| = p^k$? The answer is yes.

Theorem 5 (Sylow (1832 - 1918)) *Let G be a finite group and write $|G| = mp^n$ where p is a prime and m is not divisible by p.*

- *G contains a subgroup P of order p^n. P is called "Sylow p-subgroup" of G.*
- *G contains a subgroup of order p^k for every $0 \leq k \leq n$.*
- *If P, Q are two Sylow p-subgroups of G then they are conjugated: there exists $g \in G$ such that $g^{-1}Pg = Q$.*
- *The number of Sylow p-subgroups of G is congruent to 1 mod p.*
- *If H is a subgroup of G such that $|H|$ is a power of p then there exists a Sylow p-subgroup P of G such that $H \leq P$.*

Consider the following example. Let $F = \mathbb{Z}/5\mathbb{Z}$ and let

$$G := \{\begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \ : \ a, b, c \in F, \ a, c \neq 0\},$$

$$H := \{\begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \in G \ : \ b = 0\},$$

$$K := \{\begin{pmatrix} a & 0 \\ b & c \end{pmatrix} \in G \ : \ a = c = 1\}.$$

- G is a group (with respect to multiplication) of order $4^2 \cdot 5 = 2^4 \cdot 5$,
- $|H| = 4^2 = 2^4 \ \Rightarrow \ H$ is a Sylow 2-subgroup of G and
- $|K| = 5 \ \Rightarrow K$ is a Sylow 5-subgroup of G.

Finally, Cayley theorem says that every finite group can be found inside some symmetric group.

Theorem 6 (Cayley (1821 - 1895)) *Let G be a group. Then the map*

$$G \to Sym(G), \quad g \mapsto (x \mapsto gx)$$

is an injective homomorphism.
In particular, G is isomorphic with a subgroup of $Sym(G)$.

Corollary 1 *Let G be a finite group. There exists a positive integer n such that G is isomorphic with a subgroup of Sym(n).*

Cayley's theorem says that we may choose $n = |G|$. But sometimes we can choose a smaller n (cf. [7]).

For example, if G_f is the Galois group of the polynomial $f(X) \in \mathbb{Q}[X]$ with n distinct roots, then the permutation action of G_f on the n roots gives an injective homomorphism $G_f \to \mathrm{Sym}(n)$.

4 Simple groups, solvable groups

A group G is said to be "simple" if the only normal subgroups of G are $\{1\}$ and G. Since every subgroup of an abelian group is normal, it is easy to show that

Remark 3 (Abelian simple groups) Abelian simple groups are the cyclic groups of prime order,

$$C_p = \{g, g^2, \ldots, g^{p-1}, g^p = 1\} \cong (\mathbb{Z}/p\mathbb{Z}, +) = \{1, 2, \ldots, p-1, p = 0\}.$$

The following results provide infinite families of non-abelian simple groups.

Theorem 7 (Alternating Groups) *If $n \geq 5$ is an integer, Alt(n) is a non-abelian simple group.*

Theorem 8 (Projective Linear Groups) *Let F be a field, and let $GL(n, F)$ be the group of invertible matrices over F. Let $SL(n, F)$ be the subgroup of $GL(n, F)$ consisting of matrices of determinant 1. Let Z be the subgroup of $GL(n, F)$ consisting of scalar matrices. If $n \geq 2$ and $|F| \geq 4$, the quotient*

$$PSL(n, F) := SL(n, F)/Z \cap SL(n, F)$$

*(**projective linear group**) is a non-abelian simple group.*

Given a finite group G, we can costruct longest possible chains of subgroups of the form

$$\{1\} = G_0 \lhd G_1 \lhd G_2 \lhd \cdots \lhd G_k = G.$$

Maximality of k implies that the factor groups G_i/G_{i-1} are all simple groups. Such chain is called "**composition series**" and its factors G_i/G_{i-1} are called "**composition factors**".

Theorem 9 (Jordan-Holder) *Any two composition series of a given finite group have the same length and the same composition factors (up to reordering and isomorphism).*

For example the composition factors of the cyclic group C_n correspond to the prime divisors of n, counted with multiplicity. If $n = 60 = 2^2 \cdot 3 \cdot 5$,

$$1 \lhd \langle g^{30} \rangle (\cong C_2) \lhd \langle g^{15} \rangle (\cong C_4) \lhd \langle g^5 \rangle (\cong C_{12}) \lhd \langle g \rangle = C_{60}.$$

Definition 2 (Solvable groups) If the composition factors of the finite group G are all abelian (hence cyclic of prime order) then G is said to be **solvable**.

Evariste Galois proved that the zeros of a polynomial $f(X) \in \mathbb{Q}[X]$ can be expressed by starting from the elements of $\mathbb{Q}$ and performing sums, differences, products, divisions, and root extractions if and only if the Galois group G_f is solvable. In this case $f(X)$ is said to be "**solvable by radicals**". Let us give some examples.

The Galois group of $f(X) = X^4 - 4X + 2 \in \mathbb{Z}[X]$ is S_4, so $f(X)$ is solvable by radicals. Indeed, S_4 is solvable:

$$\{1\} \xrightarrow{C_2} \langle (12)(34) \rangle \xrightarrow{C_2} O_2(S_4) \xrightarrow{C_3} A_4 \xrightarrow{C_2} S_4$$

Arrows are inclusions. $O_2(S_4)$ denotes the intersection of the Sylow 2-subgroups of S_4: it is a normal subgroup of S_4 of order 4 isomorphic to the Klein group $C_2 \times C_2$. The composition factors of S_4 are C_2 (three times) and C_3. $|S_4| = 24 = 2^3 \cdot 3$. More generally, all polynomials of degree 2, 3, 4 are solvable by radicals. Indeed, all subgroups of Sym(4) are solvable. On the other hand, the symmetric group S_n is not solvable when $n \geq 5$:

$$\{1\} \xrightarrow{A_n} A_n \xrightarrow{C_2} S_n$$

The composition factors of S_n are A_n (not abelian) and C_2.

Let a, b, c be indeterminates over $\mathbb{Q}$. The roots of the polynomial $P(X) = aX^2 + bX + c$ are given by the well-known formula

$$x_{1,2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

It follows that $P(X)$ is solvable by radicals over $\mathbb{Q}(a, b, c)$. I want to consider now degrees larger than 2.

For the following discussion we refer the reader to [14, Theorem 4.15]. Let $a_0, \ldots, a_{n-1}$ be indeterminates over $\mathbb{Q}$. It is interesting to ask when the generic polynomial of degree n

$$P(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_1 X + a_0$$

is solvable by radicals over the field generated by its coefficients, $\mathbb{Q}(a_0, \ldots, a_{n-1})$. In other words, we ask when the roots of $P(X)$ can be expressed by starting from the coefficients $a_0, \ldots, a_{n-1}$ and performing sums, differences, products, divisions and root extractions. It turns out that $P(X)$ is irreducible in $\mathbb{Q}(a_0, \ldots, a_{n-1})[X]$, it has distinct roots (as does any irreducible polynomial in characteristic zero) and its Galois group over $\mathbb{Q}(a_0, \ldots, a_{n-1})$ is Sym(n). Since, as we have seen, Sym(n) is not a solvable group if $n \geq 5$, it follows that $P(X)$ is solvable by radicals if and only if $n \leq 4$.

As you might have noticed, above I used the expression "over the field generated by its coefficients". Let us clarify this. If F/K (to be read "F over K") is any **field extension**

(meaning that F and K are fields and F contains K) then the Galois group of F/K, denoted $\mathcal{G}(F/K)$, is defined to be the set of all field automorphisms g of F such that (*) $g(a) = a$ for every $a \in K$. Note that if $K = \mathbb{Q}$ then condition (*) is automathic. The inclusion $K \subseteq F$ gives F a canonical structure of K-vector space. The extension F/K is said to be "finite" if F has finite dimension as K-vector space. The "degree" of the finite field extension F/K, usually denoted $[F : K]$, is the dimension $\dim_K(F)$. So for example $\mathbb{C}/\mathbb{R}$ is a finite field extension of degree 2 with Galois group C_2 (its two elements are the identity and the *complex conjugation* $a + ib \mapsto a - ib$). A finite extension F/K is said to be a *Galois extension* if

$$\{a \in F \ : \ g(a) = a \ \forall g \in \mathcal{G}(F/K)\} = K.$$

It turns out that an extension F/K is a Galois extension if and only if $[F : K] = |\mathcal{G}(F/K)|$, that is, the degree equals the size of the Galois group. For example, whenever $f(X)$ is a polynomial in $\mathbb{Q}[X]$, with roots $a_1, \ldots, a_n \in \mathbb{C}$, the extension $\mathbb{Q}(a_1, \ldots, a_n)/\mathbb{Q}$ is a Galois extension. For example consider $f(X) = X^2 + 1 \in \mathbb{R}[X]$: since $\mathbb{C} = \mathbb{R}(i) = \mathbb{R}(i, -i)$, the extension $\mathbb{C}/\mathbb{R}$ is Galois.

Here is an example of an extension that is **not Galois**: $F/K = \mathbb{Q}(\sqrt[3]{2})/\mathbb{Q}$. Indeed, the only K-automorphism of F is the identity, $\mathrm{id}_F : F \to F$ ($\sqrt[3]{2}$ is the only root of $X^3 - 2$ that belongs to F !), so $|\mathcal{G}(F/K)| = 1$, while the degree $[F : K]$ is 3: a K-basis is given by 1, $\sqrt[3]{2}$, $\sqrt[3]{4}$. However, F is contained in a Galois extension of K: $\mathbb{Q}(a, b, c)/\mathbb{Q}$, where a, b, c are the three roots of $X^3 - 2$ in $\mathbb{C}$. It is an extension of degree 6 with Galois group isomorphic to Sym(3).

Here comes the main property of Galois extensions, which makes them very nice. If F/K is a finite Galois extension then the correspondences

$$H \mapsto \{a \in F \ : \ h(a) = a \ \forall h \in H\}$$

$$L \mapsto \{g \in \mathcal{G}(F/K) \ : \ g(a) = a \ \forall a \in L\}$$

provide inclusion-reversing bijections, inverses of each other, between the family of subgroups of $\mathcal{G}(F/K)$ and the family of fields L such that $K \subseteq L \subseteq F$ (*intermediate fields* of F/K). So, if you want to see how the intermediate field lattice of a Galois extension looks like just take the subgroup lattice of its Galois group and turn it upside-down (remember that inclusions are reversed).

I now spend some words on the classification of the finite simple groups. The starting point for the classification was the following beautiful result, which is known as the "odd order theorem". The proof is very long. Recently (September 2012) it was checked by the computer program Coq, essentially proving algorithmically that the proof is correct. This was achieved by a team led by Georges Gonthier (cf. http://ssr2.msr-inria.inria.fr/~ jenkins/current/progress.html).

Theorem 10 (Feit, Thompson, 1962-1963) *Any finite group of odd order is solvable.*

It is easy to show that this is equivalent to say that every finite non-abelian simple group has even order (just look at a composition series). In particular, by Cauchy Theorem, any finite non-abelian simple group contains involutions (elements of order 2). Finite simple groups have been classified using the centralizers of the involutions. The centralizer of an element $x \in G$ is the set of elements which commute with x, $C_G(x) := \{g \in G \ : \ gx = xg\}$.

Proposition 1 ([1], (45.4)) *Let G be a finite simple group and let t be an involution in G, $n := |C_G(t)|$. Then $|G| \leq (2n^2)!$.*

An immediate corollary is:

Theorem 11 (Brauer-Fowler) *Let H be a finite group. Then there exists at most a finite number of finite simple groups G with an involution t such that $C_G(t) \cong H$.*

This should clarify why the Feit-Thompson theorem is considered to be the starting point of the classification. The classification theorem is too long to be fully stated here, so I will give a short version of it. Groups of "Lie type" are particular groups of matrices over finite fields, as in the case of the projective special linear group $PSL(n, F)$.

Theorem 12 (Classification of the Finite Simple Groups) *Let S be a finite simple group. Then one of the following holds.*

- *$S \cong C_p$ for some prime p.*
- *$S \cong Alt(n)$ for some integer $n \geq 5$.*
- *S is a group of Lie type.*
- *S is one of 26 sporadic groups.*

5 Some more beautiful results

The following result is one of the few results which guarantee the existence of subgroups of some order.

Theorem 13 (Schur-Zassenhaus) *Let G be a finite group and let N be a normal subgroup of G. If $|N|$ and $|G : N|$ are coprime then G admits a subgroup of size $|G : N|$.*

The following result is a generalization of Sylow's Theorem in the case of solvable groups. If π is a set of prime numbers, a "Hall π-subgroup" of G is a subgroup H of G such that $|H|$ and $|G : H|$ are coprime and all prime divisors of $|H|$ belong to π.

Theorem 14 [Hall] *Let G be a finite solvable group, and let π a set of prime numbers. Then G admits Hall π-subgroups, and any two Hall π-subgroups of G are conjugated.*

Classically this is proved using the Schur-Zassenhaus theorem.

6 Covering finite groups

Now I will say something about my own research. From now on all considered groups will be assumed to be finite.

6.1 Sigma

A "cover" of a group G is a family $\mathcal{H}$ of proper subgroups of G such that $\bigcup_{H\in\mathcal{H}} H = G$. It is easy to see that a group admits covers if and only if it is noncyclic. This follows from the equality $\bigcup_{g\in G}\langle g\rangle = G$ and the fact that a proper subgroup cannot contain the elements g such that $\langle g\rangle = G$.

Define $\sigma(G)$, the "**covering number**" of G, to be the smallest size of a cover of G. This notion was introduced in [8]. A cover of G of size $\sigma(G)$ will be called "minimal cover". If G is cyclic, set $\sigma(G) = \infty$ with the convention that $n < \infty$ for every integer n. It is obvious, but worth remarking, that if $\mathcal{H}$ is any cover of G then $\sigma(G) \leq |\mathcal{H}|$. The following basic result shows that if $N \trianglelefteq G$ then

$$\sigma(G) \leq \sigma(G/N).$$

Theorem 15 (Correspondence theorem) *Let G be a group and let $N \trianglelefteq G$. The correspondences*

$$\varphi : H/N \mapsto \{g \in G \ : \ gN \in H/N\}, \qquad \psi : K \mapsto KN/N$$

provide canonical bijections, inverses of each other, between the family of subgroups of G/N and the family of subgroups of G containing N. Moreover, they both send normal subgroups to normal subgroups.

Indeed, if $\mathcal{H}$ is a cover of G/N then $\{\varphi(H) \ : \ H \in \mathcal{H}\}$ is a cover of G of size $|\mathcal{H}|$.

For example, if $n \neq 9$ is an odd integer larger than 1 then $\sigma(\mathrm{Sym}(n)) = 2^{n-1}$. A minimal cover of G is given by the following family:

$$(**)\{\mathrm{Alt}(n)\} \cup \{\mathrm{Sym}(a) \times \mathrm{Sym}(b) \ : \ 1 \leq a, b \leq n-1, \ a+b = n\}.$$

The subgroups of $\mathrm{Sym}(n)$ isomorphic to $\mathrm{Sym}(a) \times \mathrm{Sym}(b)$, for $a + b = n$, are obtained by considering partitions $\{1, \ldots, n\} = A \cup B$ with $A \cap B = \emptyset$, $|A| = a$ and $|B| = b$. Indeed, for such a partition, it turns out that

$$\{g \in \mathrm{Sym}(n) \ : \ g(x) \in A \ \forall x \in A\} \cong \mathrm{Sym}(a) \times \mathrm{Sym}(b).$$

Such subgroups of $\mathrm{Sym}(n)$ are called "maximal intransitive". The reason why family (**) is a cover of $\mathrm{Sym}(n)$ is that n being odd, the n-cycles are even permutations, hence they belong to $\mathrm{Alt}(n)$, and all the other permutations belong to some maximal intransitive subgroup (they have nontrivial orbits - just look at the cycle structure and group the cycles in two blocks). Family (**) is a cover also for $n = 9$ but it is still not known whether it is minimal or not in this case.

It is easy to show that a group cannot be written as the union of two proper subgroups. Therefore $\sigma(G) \geq 3$ always. What can be said about the groups G with $\sigma(G) = 3$?

Theorem 16 (Scorza, 1926) *Let G be a group. Then $\sigma(G) = 3$ if and only if G admits a normal subgroup N such that $G/N \cong C_2 \times C_2$.*

Note that the implication $\Leftarrow$ is easy: if $G/N \cong C_2 \times C_2$ then $\sigma(G) \leq \sigma(G/N) = \sigma(C_2 \times C_2) = 3$. On the other hand $\sigma(G) \geq 3$ (this is always true), so $\sigma(G) = 3$. The reason why $\sigma(C_2 \times C_2) = 3$ is that $C_2 \times C_2$ has only three nontrivial proper subgroups, and they have size 2. In general if p is any prime number then $C_p \times C_p$ has precisely $p + 1$ proper nontrivial subgroups, and they are all cyclic of order p (so they are both minimal and maximal subgroups). It follows that $\sigma(C_p \times C_p) = p + 1$. Indeed, any nontrivial element of $C_p \times C_p$ determines the proper subgroup in which it is contained: it is the subgroup which it generates. So in this case in order to cover the group all proper nontrivial subgroups have to be considered.

I now argue that whenever G is a noncyclic group and $|G|$ is a power of a prime p (i.e. G is a "p-group") we have $\sigma(G) = p + 1$.

Lemma 1 (The Minimal Index Lower Bound) *Let G be a non-cyclic group, and write $G = H_1 \cup \cdots \cup H_n$ as union of $n = \sigma(G)$ proper subgroups. Let $\beta_i := |G : H_i| := |G|/|H_i|$ for $i = 1, \ldots, n$. Then $\min\{\beta_1, \ldots, \beta_n\} < \sigma(G)$.*

Proof. We may assume that $\beta_1 \leq \cdots \leq \beta_n$. Since $1 \in H_1 \cap \ldots \cap H_n$ the union $H_1 \cup \ldots \cup H_n$ is not disjoint and hence

$$|G| < \sum_{i=1}^{n} |H_i| = |G| \sum_{i=1}^{n} \frac{1}{\beta_i} \leq \frac{|G|n}{\beta_1}.$$

Therefore $\beta_1 < n$. □

Let us apply this to the case $|G| = p^n$. The index of any subgroup of G is a divisor of $|G|$ (Lagrange Theorem) so if a subgroup of G is proper then its index is at least p. It follows that $p < \sigma(G)$ so $p + 1 \leq \sigma(G)$. We are left to prove that $\sigma(G) \leq p + 1$, and for this it is enough to find a normal subgroup N of G such that $G/N \cong C_p \times C_p$ (indeed $\sigma(G) \leq \sigma(G/N)$ and $\sigma(C_p \times C_p) = p + 1$). This follows from the following fact, which I will state without proof (the proof is a bit technical). Recall that a subgroup H of G is called "maximal" if it is not properly contained in a subgroup of G, and the "Frattini subgroup" of a group G is the intersection of the maximal subgroups of G, denoted $\Phi(G)$. It is a normal subgroup of G. Moreover, denote by $d(G)$ the least integer d such that there exist d elements $x_1, \ldots, x_d \in G$ with $\langle x_1, \cdots, x_d \rangle = G$.

Proposition 2 *Let G be a p-group and let $d = d(G)$. Then $G/\Phi(G) \cong {C_p}^d$.*

Now if the p-group G is not cyclic, i.e. if $d > 1$, then ${C_p}^d$ clearly admits a quotient isomorphic to $C_p \times C_p = {C_p}^2$, therefore $\sigma(G) \leq \sigma(G/\Phi(G)) = \sigma({C_p}^d) \leq \sigma(C_p \times C_p) = p+1$.

Using this we can deduce the value of $\sigma(G)$ whenever G is an abelian group. We first need a lemma.

Lemma 2 *If A, B are two groups of coprime order then $\sigma(A \times B) = \min\{\sigma(A), \sigma(B)\}$.*

Proof. Since $|A|$ and $|B|$ are coprime, the subgroups of $A \times B$ are of the form $H \times K$ with $H \leq A$ and $K \leq B$. With this in mind, the proof becomes technical. I will omit the details. □

Indeed, by the structure theorem of finite abelian groups (and the Chinese Remainder Theorem), any finite abelian group is a direct product of cyclic groups of prime power order, so

Proposition 3 *Let G be a noncyclic abelian group, and write $G = \prod_{i=1}^{k} C_{p_i^{n_i}}$. Then $\sigma(G) = p + 1$ where p is the smallest prime number such that there exist two distinct $i, j \in \{1, \dots, k\}$ with $p_i = p_j = p$.*

6.2 Direct products

Lemma 2 deals with direct products of groups A, B of coprime order. Let us give an example in which $A = B$. I will compute $\sigma(S \times S)$ when S is a nonabelian simple group. I will prove that $\sigma(S \times S) = \sigma(S)$. This will give me the opportunity to discuss more general facts.

Lemma 3 (Intersection argument) *Let K be a maximal subgroup of a group G and let $\mathcal{H}$ be a minimal cover of G. If $\sigma(G) < \sigma(K)$ then $K \in \mathcal{H}$. Equivalently, if $K \notin \mathcal{H}$ then $\sigma(K) \leq \sigma(G)$.*

Proof. We have

$$K = K \cap G = K \cap \bigcup_{H \in \mathcal{H}} H = \bigcup_{H \in \mathcal{H}} (K \cap H)$$

therefore, if $\sigma(G) < \sigma(K)$, this union cannot consist of proper subgroups of K, thus there exists $H \in \mathcal{H}$ such that $K \cap H = K$, i.e. $K \subseteq H$. Since K is maximal it follows that $K = H \in \mathcal{H}$. □

Corollary 2 *Let M be a maximal subgroup of G, not normal, such that $\sigma(G) < \sigma(M)$. Then $|G : M| < \sigma(G)$.*

Proof. Using standard arguments of group actions, it is possible to prove that the number of conjugates of $H \leq G$ equals the index in G of the "normalizer" $N_G(H) := \{g \in G \ : \ g^{-1}Hg = H\}$ of H in G. We always have $H \subseteq N_G(H)$ and $N_G(H) = G$ if and only if H is normal in G. Now, M being maximal and not normal in G, $N_G(M) = M$ therefore M has $|G : M|$ conjugates in G. Since $\sigma(G) < \sigma(M)$, they all belong to every minimal cover of G by the intersection argument. In particular $\sigma(G) \geq |G : M|$. Now, the $|G : M|$ conjugates of M cover less than $|G| = |M| \cdot |G : M|$ group elements (they all contain the identity element), so we get the strict inequality $\sigma(G) > |G : M|$. □

This is actually the main argument used in [9] and in my Ph.D thesis. Let me be more precise about this. Recall that if A, B are two subgroups of a group G then the product AB is defined as $AB := \{ab \ : \ a \in A, \ b \in B\}$. It turns out that $|AB| = |A| \cdot |B|/|A \cap B|$ (nice exercise). A **supplement** of the normal subgroup $N \trianglelefteq G$ is a subgroup $H \leq G$ such

that $HN = G$. A **complement** of N is a supplement H of N such that $H \cap N = \{1\}$. In this case we also say that H complements N in G. If H complements N then $|G| = |HN| = |H| \cdot |N|$, so $|G : N| = |H|$. The above corollary implies the following.

Proposition 4 (The Maximal Complement Argument) *Let N be a nonsolvable normal subgroup of the group G and suppose that there exists a maximal subgroup M of G that complements N. Then $\sigma(G) = \sigma(G/N)$.*

This is the argument that allows to produce results about the structure of σ-elementary groups (cf. the following subsection). The proof is a bit technical but I hope that by writing it down I will give some ideas about of the kind of arguments needed in this kind of analysis.

Proof. Thanks to nonsolvability of N we may assume that N does not contain nontrivial central elements of G (i.e. elements g lying in the center of G). Indeed if $g \in N$ and $g \in Z(G)$ we may consider the quotient $G/\langle g \rangle$ and proceed by induction on $|G|$. As a consequence of the classification of finite simple groups, N does not have fixed-point-free automorphisms (recall that an automorphism φ of N, i.e. a group isomorphism $N \to N$, is said to be fixed-point-free if $\varphi(x) \neq x$ whenever $x \in N$ and $x \neq 1$). It follows that the family $\{C_G(x) \ : \ 1 \neq x \in N\}$ covers G, where $C_G(x) = \{g \in G \ : \ gx = xg\}$. Indeed, if $g \in G$ then the map $N \to N$, $x \mapsto g^{-1}xg$ is an automorphism of N. Since N does not contain nontrivial central elements, $C_G(x) \neq G$ for every $1 \neq x \in N$. Therefore $\{C_G(x) \ : \ 1 \neq x \in N\}$ is a cover of G of size $|N| - 1$, so $\sigma(G) \leq |N| - 1$.

Now, M is not normal in G, otherwise $G \cong N \times M$ and maximality of M would imply that $|N| = |G : M| = p$ is a prime, contradicting the nonsolvability of N. Therefore Corollary 2 implies that $|N| = |G : M| < \sigma(G)$. This contradicts the fact that $\sigma(G) \leq |N| - 1$. □

Let us show how this implies that $\sigma(S \times S) = \sigma(S)$ whenever S is a nonabelian simple group. The following is a standard fact and a nice exercise.

Proposition 5 *Let G be a group. Then G is simple if and only if*

$$\Delta_G := \{(g, g) \ : \ g \in G\} < G \times G$$

is a maximal subgroup of $G \times G$.

It follows that Δ_S is a maximal subgroup of $S \times S$ that complements $S \times \{1\}$, and we may apply the Maximal Complement Argument. Actually in this case much less is needed: since $S \times S \to S$, $(x, y) \mapsto x$ is a surjective homomorphism with kernel $\{1\} \times S$, by the Isomorphism Theorem (Theorem 1) and the fact that S being noncyclic, it admits as a cover the family of its nontrivial cyclic subgroups, it follows that

$$\sigma(S \times S) \leq \sigma(S) \leq |S| - 1,$$

and now since $|S \times S : \Delta_S| = |S|$ Corollary 2 yields a contradiction.

A result I obtained in a joint work with A. Lucchini is a generalization of this fact to all direct products:

Theorem 17 (Lucchini A., G 2010 [11]) *Let $\mathcal{M}$ be a minimal cover of a direct product $G = H_1 \times H_2$ of two finite groups. Then one of the following holds:*

(a) $\mathcal{M} = \{X \times H_2 \mid X \in \mathcal{X}\}$ *where $\mathcal{X}$ is a minimal cover of H_1. In this case $\sigma(G) = \sigma(H_1)$.*

(b) $\mathcal{M} = \{H_1 \times X \mid X \in \mathcal{X}\}$ *where $\mathcal{X}$ is a minimal cover of H_2. In this case $\sigma(G) = \sigma(H_2)$.*

(c) There exist $N_1 \trianglelefteq H_1$, $N_2 \trianglelefteq H_2$ with $H_1/N_1 \cong H_2/N_2 \cong C_p$ and $\mathcal{M}$ consists of the maximal subgroups of $H_1 \times H_2$ containing $N_1 \times N_2$. In this case $\sigma(G) = p + 1$.

6.3 σ-elementary groups

Suppose we want to compute $\sigma(G)$ for a given group G. If there exists $N \trianglelefteq G$ such that $\sigma(G) = \sigma(G/N)$ then we may consider the group G/N instead of G. This gives a sort of reduction and leads to the following definition.

Definition 3 (σ-elementary groups) A group G is called σ-elementary if $\sigma(G) < \sigma(G/N)$ whenever $\{1\} \neq N \trianglelefteq G$. G is called n-elementary if G is σ-elementary and $\sigma(G) = n$.

This notion was introduced in [8] and thoroughly studied in [9] (there these groups are called σ-primitive).

For example:

- If G is any group then there exists a normal subgroup N of G such that G/N is σ-elementary and $\sigma(G) = \sigma(G/N)$ (just choose a proper normal subgroup N of G such that $\sigma(G) = \sigma(G/N)$ and proceed by induction on $|G|$).
- If p is any prime number, $C_p \times C_p$ is $(p + 1)$-elementary (the nontrivial proper quotients are all cyclic of size p).
- The only 3-elementary group is $C_2 \times C_2$ (Scorza's Theorem).
- 6.2 implies that if S is a nonabelian simple group then $S \times \cdots \times S = S^m$ is σ-elementary if and only if $m = 1$.
- If $n \geq 3$ is an integer and $n \neq 4$ then $\mathrm{Sym}(n)$ is σ-elementary: its only nontrivial proper quotient is C_2. $\mathrm{Sym}(4)$ is not σ-elementary: it admits $\mathrm{Sym}(3)$ as homomorphic image (quotient) and $\sigma(\mathrm{Sym}(4)) = \sigma(\mathrm{Sym}(3)) = 4$.
- If G/N is cyclic whenever $\{1\} \neq N \trianglelefteq G$ then G is σ-elementary. The converse is true for solvable groups but false in general. An example is $I \rtimes \mathrm{Alt}(p)$ where $I = \{(x_1, \ldots, x_p) \in \mathbb{F}_2^p : \sum_{i=1}^p x_i = 0\}$ and p is a prime not of the form $\frac{q^n-1}{q-1}$ with q a prime power, the action is the usual one on the p coordinates.

Let us list the known facts concerning σ-elementary groups. Recall that $\Phi(G)$, the Frattini subgroup of G, is the intersection of the maximal subgroups of G, $Z(G)$, the center of G, is the subgroup $\{g \in G \ : \ xg = gx \ \forall x \in G\}$, and G', the derived subgroup of G, is the intersection of the normal subgroups N of G such that G/N is abelian.

Proposition 6 *Let G be a σ-elementary group.*

- $\Phi(G) = \{1\}$.
- *If G is non-abelian then it has trivial center: $Z(G) = \{1\}$.*
- *If G is abelian then $G \cong C_p \times C_p$ for some prime p.*
- *Scorza's theorem: if $\sigma(G) = 3$ then $G \cong C_2 \times C_2$.*
- *Scorza's theorem revisited: if $\sigma(G) = p + 1$ with p the smallest prime divisor of $|G|$ then $G \cong C_p \times C_p$.*
- *Let n be a positive integer. There are only finitely many σ-elementary groups G with $\sigma(G) = n$.*
- *If $H_1 \times H_2$, a direct product of two non-trivial groups, is σ-elementary then $H_1 \cong H_2 \cong C_p$ for some prime p (this follows from Theorem 17).*
- *If G is σ-elementary, $\{1\} \neq N \trianglelefteq G$ and G/N is solvable then G/N is cyclic. In particular G/G' is cyclic.*

A result I obtained is the determination of all n-elementary groups with $n \leq 25$.

Theorem 18 (G 2009 [10]) *All σ-elementary groups G with $\sigma(G) \leq 25$ are known.*

$\sigma(G)$	G
3	$C_2 \times C_2$
4	$C_3 \times C_3, Sym(3)$
5	$Alt(4)$
6	$C_5 \times C_5, D_{10}, AGL(1,5)$
7	$\emptyset$
8	$C_7 \times C_7, D_{14}, 7:3, AGL(1,7)$
9	$AGL(1,8)$
10	$3^2:4, AGL(1,9), Alt(5)$
11	$\emptyset$
12	$C_{11} \times C_{11}, 11:5,$ $D_{22}, AGL(1,11)$
13	$Sym(6)$
14	$C_{13} \times C_{13}, D_{26}, 13:3,$ $13:4, 13:6, AGL(1,13)$

$\sigma(G)$	G
15	$SL(3,2)$
16	$Sym(5), Alt(6)$
17	$2^4:5, AGL(1,16)$
18	$C_{17} \times C_{17}, D_{34}, 17:4,$ $17:8, AGL(1,17)$
19	$\emptyset$
20	$C_{19} \times C_{19}, AGL(1,19),$ $D_{38}, 19:3, 19:6, 19:9$
21	$\emptyset$
22	$\emptyset$
23	M_{11}
24	$C_{23} \times C_{23}, D_{46},$ $23:11, AGL(1,23)$
25	$\emptyset$

Scorza's Theorem can be read off from the top left line of the above table. Also, we see that there are some numbers n such that $\sigma(G) \neq n$ for every group G (7, 11, 19, 21, 22, 25). The following is an open question: are there infinitely many such n?

6.4 A conjecture

Definition 4 (Minimal normal subgroups) A minimal normal subgroup of a group G is a non-trivial normal subgroup N of G which does not contain any non-trivial normal subgroup of G different from N.

Let us give some examples.

- If p is a prime, $C_p \times C_p$ has $p+1$ minimal normal subgroups.
- If S is a simple group, it is its unique minimal normal subgroup.
- If $n \geq 3$ is an integer and $n \neq 4$ then the unique minimal normal subgroup of $\mathrm{Sym}(n)$ is $\mathrm{Alt}(n)$.
- The unique minimal normal subgroup of $\mathrm{Sym}(4)$ is

$$V = \{1, (12)(34), (13)(24), (14)(23)\}.$$

- If $k \geq 1$ is an integer and S is a non-abelian simple group then the minimal normal subgroups of $S \times \cdots \times S = S^k$ are its k direct factors, $S \times \{1\} \times \cdots \times \{1\}, \ldots, \{1\} \times \cdots \times \{1\} \times S$.
- If F is a field with at least 4 elements and $n \geq 2$, the unique minimal normal subgroup of $PGL(n, F)$ is $PSL(n, F)$.

Given a finite group G denote by $mn(G)$ the **number of minimal normal subgroups** of G.

The known examples of σ-elementary groups either are abelian isomorphic to $C_p \times C_p$ or admit only one minimal normal subgroup. The main problem I dealt with in my Ph.D thesis is the following conjecture, still open.

Conjecture 1 (A. Lucchini, E. Detomi) *Let G be a non-abelian σ-elementary group. Then $mn(G) = 1$.*

If $mn(G) = 1$ we usually say that G is **monolithic**.

Here is what I can say when the covering number is "small". In the following result the **wreath product** $\mathrm{Alt}(5) \wr C_2$ is the semidirect product $(\mathrm{Alt}(5) \times \mathrm{Alt}(5)) \rtimes C_2$ where the action is given by the exchange of the two coordinates.

Theorem 19 *Let G be a non-abelian σ-elementary group such that $\sigma(G) \leq 56$. Then G is monolithic. Moreover, its minimal normal subgroup is either simple or abelian.*

Moreover $\sigma(Alt(5) \wr C_2) = 57$*,* $Alt(5) \wr C_2$ *is monolithic and its minimal normal subgroup is* $Alt(5) \times Alt(5)$*, not simple and not abelian.*

A subgroup H of a group G is said to be **subnormal** if there exists a chain $H \lhd H_2 \lhd \cdots \lhd H_n = G$. Subnormal subgroups are not necessarily normal, for example in Sym(4), $\langle (12)(34) \rangle \lhd V$ and $V \lhd \mathrm{Sym}(4)$ but $\langle (12)(34) \rangle$ is not normal in Sym(4). A **minimal subnormal subgroup** is a subnormal subgroup which does not properly contain nontrivial subnormal subgroups of G. Note that minimal subnormal subgroups are always simple groups. Here is another result I proved in my Ph.D thesis.

Theorem 20 *Let* G *be a non-abelian* σ*-elementary group, and suppose that every minimal subnormal subgroup of* G *is isomorphic to an alternating group* $Alt(n)$ *with* n *large enough and even. Then* G *is monolithic.*

References

[1] M. Aschbacher, "Finite Group Theory". Cambridge Studies in Adv. Math. 10, 2000.

[2] M. Isaacs, "Finite Group Theory". Graduate Studies in Mathematics 92, A.M.S., 2008.

[3] D. Gorenstein, "Finite Groups". Chelsea Publishing Co., New York, 1980.

[4] P.J. Cameron, "Permutation Groups". London Mathematical Society Student Texts, 45. Cambridge University Press, Cambridge, 1999.

[5] D.J.S. Robinson, "A Course in the Theory of Groups". Graduate Texts in Mathematics, 80. Springer-Verlag, New York, 1996.

[6] R. Brandl, *Integer polynomials that are reducible modulo all primes.* Amer. Math. Monthly 93 (1986), 286–288.

[7] D. Easdown, C. Praeger, *On Minimal Faithful Permutation Representations of Finite Groups.*

[8] J.H.E. Cohn, *On n-sum groups.* Math. Scand., 75/1 (1994), 44–58.

[9] E. Detomi, A. Lucchini, *On the Structure of Primitive n-Sum Groups.* CUBO A Mathematical Journal 10/03 (2008), 195–210.

[10] M. Garonzi, *Finite Groups that are the union of at most 25 proper subgroups.* Journal of Algebra and Its Applications 12/4 (2013) 1350002.

[11] M. Garonzi, A. Lucchini, *Direct products of groups as unions of proper subgroups.* Archiv der Mathematik, ISSN: 0003-889X.

[12] A. Maróti, M. Garonzi, *Covering certain wreath products with proper subgroups.* J. Group Theory 14/1 (2011), 103–125.

[13] M. Garonzi, *Covering certain monolithic groups with proper subgroups.* Communications in Algebra 41/2 (2013), 471–491.

[14] N. Jacobson, "Basic Algebra 1". W.H. Freeman and Company, second edition, 1985.

Regularization by means of Generalized Arnoldi-Tikhonov methods

Silvia Gazzola (*)

Abstract. Inverse problems are ubiquitous in many areas of science and engineering: they are typically modeled by Fredholm integral equations of the first kind with compact kernel and the available data are commonly affected by errors. Once discretized they give rise to ill-conditioned linear systems, often of huge dimensions: regularization consists in replacing the original system by a nearby problem with better numerical properties, in order to to find a meaningful approximation of the exact solution. We will review some standard regularization methods, both direct and iterative, and we will introduce the most recent class of the Arnoldi-Tikhonov methods; our focus will be on problems regarding the restoration of images corrupted by blur and noise. Some numerical experiments will be shown, so to compare the different approaches and to contribute validating the newly-proposed strategies.

1 Introduction

The concept of ill-posed problems goes back to J. Hadamard (1865-1963), who essentially defined a problem as ill-posed if the solution is not unique or if it is not a continuous function of the data. Ill-posed problems arise in the form of inverse problems in many areas of science and engineering, that is every time that one is interested in determining the initial structure of a physical system from its measured behavior. We are interested in the broad class of inverse problems that, in a continuous setting, can be mathematically formulated as Fredholm integral equations of the first kind with a square integrable kernel. They all can be written in generic form as

$$\int_0^1 K(s,t)\, f(t)dt = g(s), \quad 0 \le s \le 1, \tag{1}$$

where the right-hand side function g and the kernel K are known, while f is unknown. If we consider the singular value expansion of the kernel K, we notice that the singular values decay to zero (the smoother the kernel the faster the decay) and the corresponding singular functions show increasing oscillations. Discretizing the problem (1) leads to the

(*)Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy; E-mail: gazzola@math.unipd.it . Seminar held on March 20th, 2013.

linear system

$$Ax = b. \tag{2}$$

Let us examine the singular value decomposition (SVD) of the matrix $A \in \mathbb{R}^{n\times n}$ (for the sake of simplicity we are just considering square matrices), defined as

$$A = U\Sigma V^T, \tag{3}$$

where $U \in \mathbb{R}^{n\times n}$, $V \in \mathbb{R}^{n\times n}$ are orthogonal matrices (i.e. $U^TU = UU^T = I_n$ and $V^TV = VV^T = I_n$, where I_n is the identity matrix of order n), and $\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_n)$, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$. We typically observe that the singular values (the diagonal elements of Σ) rapidly decrease and cluster at zero, with no evident gap between them (cf. Figure 1, frame **(a)**), while the corresponding singular vectors (the columns of the matrices U and V) exhibit increasing oscillations. This implies that the 2-norm condition number of the matrix A, defined as the ratio σ_1/σ_n, is very high; indeed, the bad conditioning of the matrix A reproduces the ill-posedness of the continuous problem. One more feature to be considered is that the available right-hand-side b is affected by error: of course there is always some perturbation introduced during the discretization process, but often even the model (1) is corrupted by unavoidable measurements errors. Therefore, we could better specify (2) as

$$Ax_{ex} + e = b_{ex} + e = b, \tag{4}$$

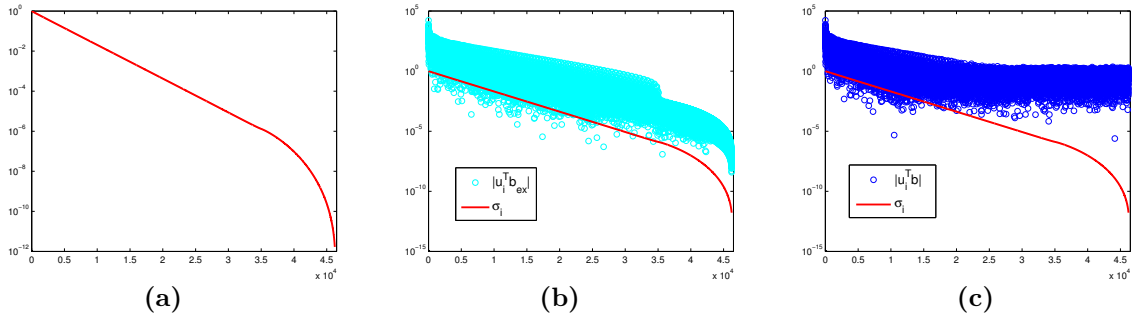
where e is an unknown error (or noise). Concerning the last formulation, our task is to recover an approximation of the solution x_{ex} starting from the available corrupted vector b.

A "simple"inversion of the matrix A, besides being often computationally unfeasible or extremely expensive, cannot give meaningful solutions because of the errors in b and the huge conditioning of A. Indeed, let us theoretically consider the expression of the solution with respect to the singular value decomposition (3)

$$x = A^{-1}b = \sum_{i=1}^{n} \frac{u_i^T b_{ex}}{\sigma_i} v_i \; + \; \sum_{i=1}^{n} \frac{u_i^T e}{\sigma_i} v_i. \tag{5}$$

If the right-hand-side only consists of the term b_{ex} (an extremely rare circumstance), the approximate solution can be straightforwardly recovered thanks to the so-called Discrete Picard Condition, which states that the Fourier coefficients $u_i^T b_{ex}$ decay as the singular values σ_i, on average (cf. Figure 1, frame **(b)**). Therefore, as far as only the first term in (5) is concerned, the solution x is well-behaved. However, as we consider the contributions given by the so-called inverted noise (the second term in (5)), the Fourier coefficients $u_i^T b$ level off after some index $\bar{i}$ (cf. Figure 1, frame **(c)**), hence the ratio $u_i^T b/\sigma_i$ is huge for $i > \bar{i}$ and the solution is dominated by highly-oscillating components (that is, by noise). We cite [4] as an excellent introduction to the solution of ill-posed problems and as a general reference for regularization methods.

The main application we are considering is image deblurring. We assume to know the operator K defining the blur and the available image g that is affected by blur and noise; K is often given as a PSF (Point Spread Function) that, depending on the circumstances, can be recovered experimentally or defined by a mathematical function (for instance by a Gaussian). The goal is to obtain a meaningful approximation of the exact and unknown image f. Mathematically, a grayscale image X is represented as a 2D array, whose integer entries define the intensity to be assigned to each pixel. In order to write a linear system like (2) we have to rearrange the involved images into 1D arrays (vectors) and we do this by stacking the columns of the 2D arrays; the matrix A is defined starting from the PSF and the so-called boundary conditions, i.e. the supposed behavior of the image outside the recorded scene. The ideally exact image we are employing to perform the tests in the following sections is shown in Figure 2, frame **(a)**: the size of this image is 215×215 pixels and the dimension of the associated blurring matrix A is 46225. The dimensions of the systems associated to image restoration problems are always huge. We corrupt the exact image applying a mild Gaussian blur and adding some white noise such that the ratio $\|e\|/\|b\|$ (noise level) is 10^{-2}. When we deal with medical and astronomical images, finding good reconstructions of the corrupted images in a quick and reliable way is particularly important. We refer to [5] for a clear and complete introduction to image deblurring.



(a) **(b)** **(c)**

Figure 1: **(a)** Decay of the singular values σ_i; **(b)** Discrete Picard Condition, only b_{ex} is involved; **(c)** Discrete Picard Condition, b affected by noise.

2 Regularization Methods

As said in the Introduction, if we wish to approximate the exact solution of the discrete ill-posed problem (2) by just inverting the matrix A, we actually obtain a vector dominated by noise. Therefore, in order to reconstruct a solution that resembles the behavior of the exact one, we have to replace the original system (2) by a nearby problem with better numerical properties. This process is called regularization. In the following we give a brief summary of the most well-known regularization techniques, introducing the class of the direct, iterative and hybrid methods. As we will see, every regularization method is not complete without specifying a rule for choosing the regularization parameter (that can be a different quantity, depending on the class of methods we are treating). A lot of techniques have been developed in order to set the regularization parameter; they can be applied,

with minor changes, to all the regularization methods described below. In Section 3.2 we are going to describe a sort of revisited discrepancy principle, that is a well-known parameter choice method that can be applied only if a fairly good approximation of the norm of the error e in (4) is known. Of course there are also methods that do not rely on this hypothesis, for example the L-curve criterion and the GCV method. We refer to [7] for an accurate description of these parameter choice strategies, along with some more recent ones.

2.1 Regularization by Direct Methods

The class of direct methods collects all the strategies that are essentially based on the SVD decomposition of the matrix A. Referring to expression (5), the regularized solution computed by a direct method can be written in this way

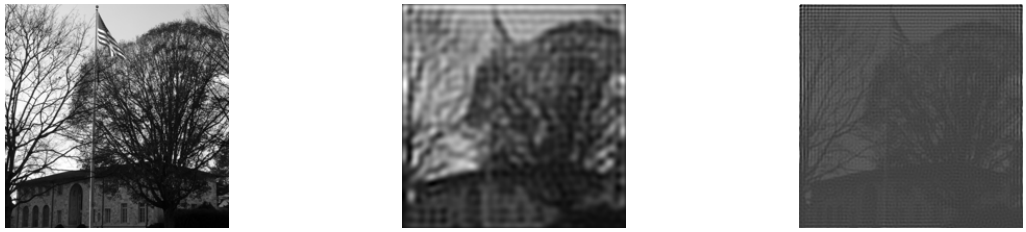
$$x_{reg} = \sum_{i=1}^{n} \phi_i \frac{u_i^T b}{\sigma_i} v_i, \tag{6}$$

i.e. the expression of the solution in the singular vectors basis is modified premultiplying each coefficient by a scalar ϕ_i, called filter factor. For this reason, direct methods are also referred to as spectral filtering methods; different filter factors define different direct methods. We consider the following ones:

- **Truncated Singular Value Decomposition** (**TSVD**): the regularized solution is defined by just taking the first $\bar{k}$ components in (5)

$$x_{TSVD} = \sum_{i=1}^{\bar{k}} \frac{u_i^T b}{\sigma_i} v_i \quad (\bar{k} < n), \quad \phi_i = \begin{cases} 1 & \text{if } i \leq \bar{k} \\ 0 & \text{if } i > \bar{k} \end{cases}. \tag{7}$$

  In this setting, the scalar $\bar{k}$ acts as a regularization parameter. If we choose $\bar{k}$ too small, i.e. if we take too few components in (5), the solution is over-regularized, i.e. we filter out too much and the solution appear to be too "smooth"(cf. Figure 2, frame **(c)**). On the contrary, if we choose $\bar{k}$ too big, i.e. if we take too many components in (5), the solution is under-regularized, i.e. we do not filter out enough and the solution is still dominated by noise (cf. Figure 2, frame **(c)**).



(a) **(b)** **(c)**

Figure 2: **(a)** exact image, **(b)** over-regularized solution by TSVD method, **(c)** under-regularized solution by TSVD method.

- **Tikhonov Regularization** (**Standard Form**) consists in substituting the system (2) with the following penalized minimization problem

$$x_{Tikh} = \min_{x\in\mathbb{R}^n} \{\|b - Ax\|^2 + \lambda\|x\|^2\}, \quad x_{Tikh} = \sum_{i=1}^{n} \overbrace{\frac{\sigma_i^2}{\sigma_i^2 + \lambda}}^{\phi_i} \frac{u_i^T b}{\sigma_i} v_i. \tag{8}$$

The objective function in (8) is the sum of two terms. The first one is a fit-to-data term: the smaller this contribution, the closer the regularized solution to the unregularized one; the second one is a regularization term: the smaller this norm, the smoother the regularized solution. The weight to be assigned to the second term is specified by the regularization parameter $\lambda > 0$: the smaller the λ, the more under-regularized the solution (cf. Figure 3, frame **(b)**), the larger the λ the more over-regularized the solution (cf. Figure 3, frame **(c)**). This can be understood also looking at the expressions for the filters factor ϕ_i and at the plot in Figure 3, frame **(a)**.

(a) **(b)** **(c)**

Figure 3: **(a)**: behavior of the Tikhonov filter factors versus the singular values of the matrix A, both in logarithmic scale, for the case $\lambda = 10^{-6}$ (blue line) and $\lambda = 5 \cdot 10^{-2}$ (red line); **(b)**: solution corresponding to the case $\lambda = 10^{-6}$ (under-regularization); **(c)**: solution corresponding to the case $\lambda = 5 \cdot 10^{-2}$ (over-regularization).

Problem (8) is equivalent to the Least-Squares problem

$$\min_{x\in\mathbb{R}^n} \left\| \begin{pmatrix} A \\ \sqrt{\lambda} I \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|^2 \tag{9}$$

and to the associated normal equations

$$(A^T A + \lambda I)x = A^T b. \tag{10}$$

As said at the beginning of this section, the SVD of the matrix A should be available in order to efficiently implement direct methods. This requirement can be computationally satisfied only by small to middle dimensional linear systems. When dealing with the huge

dimensional systems associated to image deblurring, we can only hope to compute the SVD of matrices A with special structures (we refer to [5] for a full justification of this statement).

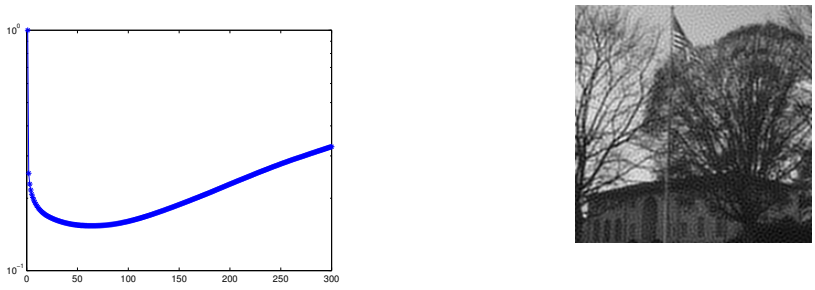
2.2 Regularization by Iterative Methods

The class of iterative regularization methods can be successfully applied to huge dimensional systems because, basically, just matrix-times-vector products or inversions of small-size auxiliary matrices are required at each iteration. Many different iterative methods have been applied with the aim of computing a regularized solution. Regularization is achieved by early termination of the iterations; the regularization properties of some iterative methods (like Landweber and CG methods, cf. [4]) have been studied theoretically. We are going to focus on iterative methods that compute, at each step, a solution belonging to a Krylov subspace. Starting from a matrix $\bar{A} \in \mathbb{R}^{n\times n}$ and a vector $\bar{b} \in \mathbb{R}^n$, Krylov subspaces are vectorial spaces of increasing dimension h defined as:

$$\mathcal{K}_h(\bar{A}, \bar{b}) = \text{span}\{\bar{b}, \bar{A}\bar{b}, \ldots, \bar{A}^{h-1}\bar{b}\}, \quad h \ll n. \tag{11}$$

Krylov subspace methods differ in the way the matrix $\bar{A}$ and the vector $\bar{b}$ are chosen in (11) and in the constraints imposed on the solution. For instance, let us consider the CGLS method: at step m the solution belongs to the space $\mathcal{K}_m(A^TA, A^Tb)$ and is determined so to minimize $\|r - r_k\|_2^2$, where $r = b - A\hat{x}$ ($\hat{x} = A^\dagger b$ is the pseudo-inverse solution of (2)) and $r_k = b - Ax_k$ is the residual at the kth iteration (cf. [1]).

The main problem when performing regularization by means of an iterative method is semiconvergence, i.e. at the beginning of the process the relative error keeps to decrease but, after some iterations, it starts to rise again (Figure 4, frame **(a)**). Therefore determining when exactly to stop the iterations is a very important issue in order to give a meaningful reconstruction, before the computed solution is dominated by noise. When considering iterative methods, the number of performed iterations is the regularization parameter: if we stop too early we have over-regularization, if we stop too late we have under-regularization.



(a) **(b)**

Figure 4: **(a)**: history of the relative errors associated to the CGLS method applied to the considered test problem, **(b)**: reconstructed solution at the 70th iteration.

2.3 Regularization by Hybrid Method

Hybrid methods incorporate an iterative and a direct approach to regularization: more precisely, a direct regularization method is applied to a system of increasingly bigger dimension. This approach has been originally introduced in [6], where the authors propose to perform a step of the Lanczos bidiagonalization algorithm at each iteration and regularize the projected system by Tikhonov method. Basically, the so-called Lanczos-hybrid method aims at *regularizing a projection*. The Arnoldi-Tikhonov method has been more recently introduced in [2], where the authors suggest to solve the Problem (8) by projecting it into the Krylov subspaces $\mathcal{K}_h(A, b)$, generated by performing a step of the Arnoldi algorithm at each iteration. Basically this method aims at *projecting the regularization*. Although in this section we are not exploring the details of hybrid methods, in the following we are going to focus exclusively on the Arnoldi-Tikhonov method, describing a generalization of it. With respect to the purely iterative approach, the semiconvergent behavior of the regularized solution is overcome. The main disadvantage of hybrid methods is that we have to set a new regularization parameter at each step and choose when to stop the iterations (i.e. the dimension of the Krylov subspace in which we want to find the regularized solution): however a meaningful solution can be typically computed after just a few iterations and so the computational cost of this methods is very low.

3 The Generalized Arnoldi-Tikhonov Method

Here we introduce a generalization of the Arnoldi-Tikhonov method that can be applied to the following generalized Tikhonov regularization method:

$$\min_{x\in\mathbb{R}^n} \left\{ \|b - Ax\|^2 + \lambda \|L(x - x_0)\|^2 \right\}. \tag{12}$$

With respect to (8), the above formulation allows to force proximity to an initial guess x_0 (if no initial guess is available we simply take $x_0 = 0$) and additional smoothness (if we take as L a finite-difference approximation of a derivative operator) into the regularized solution . This new strategy is called Generalized Arnoldi-Tikhonov (GAT) method; the ideas summarized in this section are fully explained in [3].

3.1 Formulation

Let us first consider the Krylov subspace $\mathcal{K}_m(A, r_0)$, defined as in (11). A basis for this Krylov subspace can be built using the Arnoldi algorithm [8], which leads to the associated decomposition

$$AV_m = V_{m+1}\bar{H}_m, \tag{13}$$

where $V_{m+1} = [v_1, ..., v_{m+1}] \in \mathbb{R}^{n\times(m+1)}$ has orthonormal columns that span the Krylov subspace $\mathcal{K}_{m+1}(A, r_0)$, and v_1 is defined as $r_0/\|r_0\|_2$. The matrix $\bar{H}_m \in \mathbb{R}^{(m+1)\times m}$ is an upper Hessenberg matrix. The GAT method searches for approximations x_m of the solution of Problem (12) belonging to $x_0 + \mathcal{K}_m(A, r_0)$. Therefore, replacing $x = x_0 + V_m y$

($y \in \mathbb{R}^m$) into (12), and using the properties of the matrices generated by the Arnoldi algorithm (13) yields the reduced minimization problem

$$y_m = \min_{y \in \mathbb{R}^m} \left\{ \left\| \bar{H}_m y - \|r_0\|_2 e_1 \right\|_2^2 + \lambda \left\| L V_m y \right\|_2^2 \right\}. \tag{14}$$

Similarly to what we did in Section 2.1, we consider the equivalent formulations

$$y_m = \min_{y \in \mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} L V_m \end{pmatrix} y - \begin{pmatrix} \|r_0\|_2 \, e_1 \\ 0 \end{pmatrix} \right\|_2^2 \tag{15}$$

and

$$(\bar{H}_m^T \bar{H}_m + \lambda V_m^T L^T L V_m) y_m = \bar{H}_m^T c \,. \tag{16}$$

At each step of the Arnoldi algorithm we solve the reduced-dimension least squares formulation (15), with a newly defined regularization parameter λ.

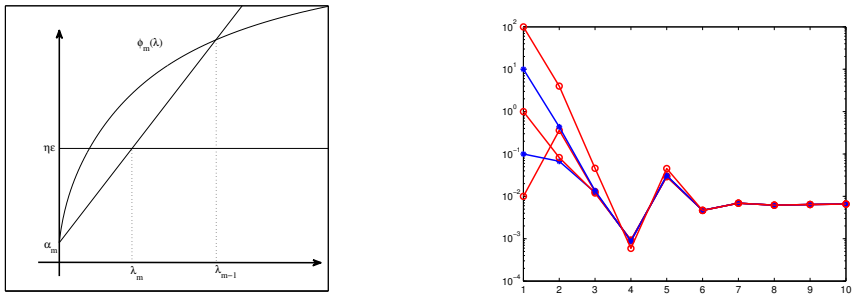
3.2 Parameter Choice Strategy

Assuming that the quantity $\varepsilon = \|e\|_2$ is known, a successful strategy to define λ as well as a stopping criterion (i.e. the dimension m of the Krylov subspace) is the discrepancy principle adapted to the iterative setting of the GAT method. Specifically, at each iteration we can define the discrepancy function

$$\phi_m(\lambda) = \|Ax_m - b\|_2 \, = \|\bar{H}_m y_m - c\|_2 \tag{17}$$

where $c = \|r_0\|_2 e_1 \in \mathbb{R}^{m+1}$. We say that the discrepancy principle is satisfied as soon as

$$\phi_m(\lambda) \leq \eta\varepsilon, \quad \text{where} \quad \eta \gtrapprox 1 \,. \tag{18}$$



(a) **(b)**

Figure 5: **(a)**: The discrepancy function $\phi_m(\lambda)$ along with its linear approximation and the horizontal line corresponding to $\|e\|$; **(b)**: values of λ versus the number of performed iterations, for different $\lambda_0 \in [10^{-2}, 10^2]$.

At each iteration we approximate the discrepancy function (17) by

$$\phi_m(\lambda) \simeq \alpha_m + \lambda \beta_m, \tag{19}$$

that is a linear function with respect to λ (cf. Figure 5, frame **(a)**), where

- $\alpha_m \in \mathbb{R}$ is simply obtained taking $\lambda = 0$; employing the formulation (16) we get

$$\alpha_m = \phi_m(0) = \left\| \bar{H}_m (\bar{H}_m^T \bar{H}_m)^{-1} \bar{H}_m^T c - c \right\|_2 . \tag{20}$$

- $\beta_m \in \mathbb{R}$ is defined by the ratio

$$\beta_m = \frac{\phi_m(\lambda_{m-1}) - \alpha_m}{\lambda_{m-1}} , \tag{21}$$

  where $\phi_m(\lambda_{m-1})$ is obtained by solving the m-dimensional Problem (15) using the parameter λ_{m-1}, which is defined at the previous step; λ_0 must be set by the user, the default value is $\lambda_0 = 1$ (the GAT method is very robust with respect to the choice of λ_0, cf. Figure 5, frame **(b)**).

To select λ_m for the next step of the generalized Arnoldi-Tikhonov algorithm we impose

$$\phi_m(\lambda_m) = \eta\varepsilon \tag{22}$$

and we again force the approximation

$$\phi_m(\lambda_m) = \alpha_m + \lambda_m \beta_m . \tag{23}$$

Substituting in (23) and using the condition (22), we obtain

$$\lambda_m = \left| \frac{\eta\varepsilon - \alpha_m}{\phi_m(\lambda_{m-1}) - \alpha_m} \right| \lambda_{m-1} . \tag{24}$$

The absolute value has been considered in order to avoid the negativity of the first values of λ_m. Numerically, formula (24) is very stable, in the sense that after the discrepancy principle is satisfied, λ_m is almost constant for growing values of m.

4 Numerical Experiments

We show the behavior of the Generalized Arnoldi-Tikhonov method when applied to the test problem we have been considering so far. We employ the regularization matrix L defined as

$$L = \begin{pmatrix} D_1 \otimes I \\ I \otimes D_1 \end{pmatrix}, \quad \text{where} \quad D_1 = \begin{pmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{pmatrix} .$$

The matrix D_1 written above is a scaled finite difference approximation of the one-dimensional first derivative.

For this example, the GAT method can deliver a solution after just four iterations of the Arnoldi algorithm, i.e. the solution belongs to the space $\mathcal{K}_4(A, b)$. Anyway, in order to assess the behavior of the method after the stopping criterion has been satisfied, we decide to proceed till the 20th step. In Figure 6 we show some meaningful quantities associated to the GAT method: we can appreciate that the values of relative errors, the discrepancy and the regularization parameter are very stable with respect to the number of iterations. The GAT scheme is computationally very efficient, since all the computations (including

the ones required to set the regularization parameter λ_m) are performed using projected quantities and therefore just involve small-dimensional matrices and vectors.

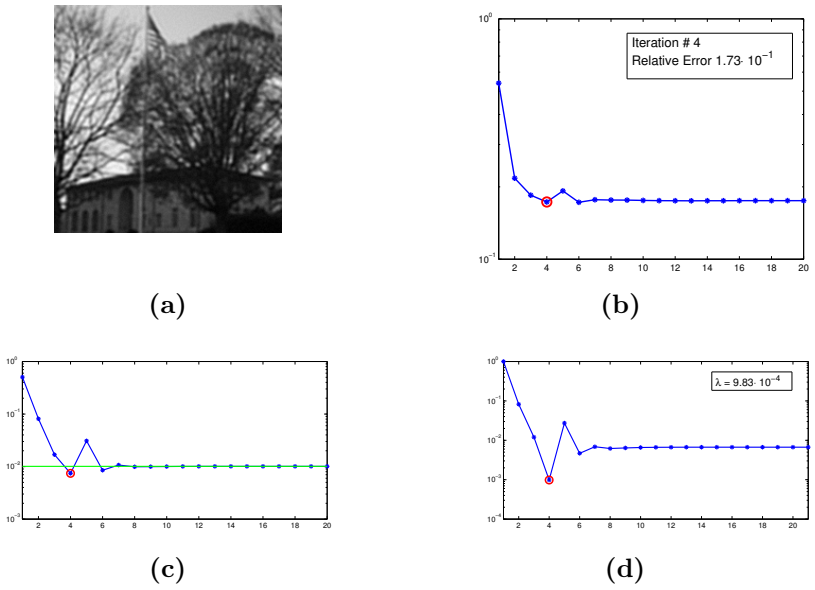


Figure 6: **(a)**: image restored employing the GAT method, **(b)**: history of the relative errors, **(c)**: values of the discrepancy function $\phi_m(\lambda_{m-1})$ versus the number of iterations, **(d)**: values of the regularization parameter λ_m.

References

[1] Å. Björck, "Numerical Methods for Least Squares Problems". SIAM, Philadelphia, 1996.

[2] D. Calvetti, S. Morigi, L. Reichel, F. Sgallari, *Tikhonov regularization and the L-curve for large discrete ill-posed problems.* J. Comput. Appl. Math. 123 (2000), 423–446.

[3] S. Gazzola, P. Novati, *Automatic parameter setting for Arnoldi-Tikhonov methods.* Submitted (2012).

[4] P.C. Hansen, "Rank-Deficient and Discrete Ill-Posed Problems. Numerical Aspects of Linear Inversion". SIAM, Philadelphia, 1998.

[5] P.C. Hansen, J.G. Nagy, D.P. O'Leary, "Deblurring Images. Matrices, Spectra and Filtering". SIAM, Philadelphia, 2006.

[6] D.P. O'Leary, J.A. Simmons, *A bidiagonalization-regularization procedure for large-scale discretizations of ill-posed problems.* SIAM J. Sci. Statist. Comput. 2 (1981), 474–489.

[7] L. Reichel, G. Rodriguez, *Old and new parameter choice rules for discrete ill-posed problems.* Numer. Algorithms, in press (2012).

[8] Y. Saad, "Iterative methods for Sparse Linear Systems". 2nd edition, SIAM, Philadelphia, 2003.

Interpolation properties of Morrey-type spaces and their application

Diana Darbayeva (*)

Abstract. It is well known that in the theory of partial differential equations, alongside with weighted Lebesque spaces, Morrey spaces and their generalizations also play an important role. Our purpose is the introduction of some generalized Morrey-type spaces, which include classical Morrey spaces, and study their properties. Moreover, we discuss the interpolation theory of linear operators and consider some applications. We prove a Marcinkiewicz-type interpolation theorem for generalized Morrey-type spaces. This theorem is then applied to obtain a Young-O'Neil-type inequality for the convolution operator in generalized Morrey-type spaces, in particular, in Morrey spaces.

1 Introduction

Fundamental interpolation theorems such as the Riesz's convexity theorem (1926), Thorin's complex version of Riesz's theorem (1939) and Marcinkiewicz's interpolation theorem (1939) are the basic models of the real interpolation method of Peetre and the complex interpolation method of Calderón. Both have found widespread application in functional analysis, approximation theory, PDE, calculus of variation, harmonic analysis. More details can be found in the books of Bergh-Löfström [1], Brudnyi-Krein-Semenov [2] and Triebel [19].

This is a note about interpolation properties of Morrey-type spaces. Classical Morrey spaces M_p^λ were studied as a consequence of the regular solutions to nonlinear elliptic equations and systems. For the properties and applications of M_p^λ we refer to [11, 15]. Interpolation properties of classical Morrey spaces and their generalizations were considered in many papers. Some results for classical Morrey spaces were obtained in Stampacchia [17], Campanato and Murthy [8], Peetre [15]. In particular, in [15] it is proved that

$$(M_p^{\lambda_0}, M_p^{\lambda_1})_{\theta,\infty} \subset M_p^\lambda\,,$$

where $\lambda = (1-\theta)\lambda_0 + \theta\lambda_1$. In Ruiz and Vega [16], Blasco, Ruiz and Vega [3] it is proved

(*)The L. N. Gumilyov Eurasian National University, Munaitpasov st., 5, 010008, Astana, Kazakhstan; E-mail: d.darbaeva@yandex.kz. Seminar held on April 10th, 2013.

that such inclusion is strict, i.e.

$$(M_p^{\lambda_0}, M_p^{\lambda_1})_{\theta,\infty} \neq M_p^{\lambda}\,.$$

The case of local Morrey-type spaces was considered in Burenkov and Nursultanov [7], where it was proved, that the interpolation spaces are again local Morrey-type spaces with the same integrability parameter. Further generalization of interpolation properties for general local Morrey-type spaces over a given measurable spaces has been discussed in [6].

However there is still interesting interpolation theorem involved with generalized Morrey-type spaces including classical Morrey spaces. Our goal is the introduction of some generalized class of Morrey spaces $M_{p,q}^{\alpha}$, which allow to prove a Marcinkiewicz-type interpolation theorem (Theorem 4 in Section 4).

The next step was to study the convolution operator

$$(Tf)(x) = (K * f)(x) = \int\limits_{\mathbb{R}^n} K(x-y)f(y)dy$$

in the spaces $M_{p,q}^{\alpha}$, where $K(\cdot)$ is a locally integrable function on $\mathbb{R}^n$.

For the convolution operator the generalizations of Young-O'Neil's inequality were obtained in Blozinski [4], Kerman [9], Kostyuchenko and Nursultanov [10], Stepanov [18], Nursultanov and Tikhonov [12, 13], and others. We also continue the study Young-O-Neil's inequality in the case of Morrey spaces. In particular, by applying the Marcinkiewicz-type interpolation theorem, we prove boundedness of the convolution operator for classical Morrey spaces (Theorem 6 in Section 5). There are many possibilities for application, because many problems in PDE are using boundedness property of that or other operator in Morrey spaces and their generalizations.

2 Classical interpolation theorems and applications

Let $\Omega \subset \mathbb{R}^n$ and let (Ω, μ) be a space with a positive measure μ. By $L_p(\Omega, d\mu)$ (or L_p) we denote the Lebesque space of all functions f μ-measurable on Ω for which

$$||f||_{L_p(\Omega,d\mu)} = \left(\int\limits_{\Omega} |f(x)|^p d\mu\right)^{\frac{1}{p}}.$$

Here $1 \leq p \leq \infty$ and

$$||f||_{L_\infty(\Omega,d\mu)} = \text{ess sup}_{\,x\in\Omega}|f(x)|$$

if $p = \infty$.

Let T be a linear operator acting from L_p into L_q, i.e. $T(\alpha f + \beta g) = \alpha T(f) + \beta T(g)$. If also T is a bounded operator or if the quantity

$$M = \sup_{f\neq 0} \frac{\|Tf\|_{L_q}}{\|f\|_{L_p}}$$

is finite, then we shall write

$$T : L_p \to L_q.$$

The number M is called the norm of the operator T.

The following classical interpolations theorems are well known.

Theorem 1 (Riesz-Thorin's interpolation theorem [1]) *Suppose that* $p_0 \neq p_1,\ q_0 \neq q_1$, *and*

$$T : L_{p_0}(U, d\mu) \to L_{q_0}(V, d\nu) \quad \textit{with the norm } M_0,$$
$$T : L_{p_1}(U, d\mu) \to L_{q_1}(V, d\nu) \quad \textit{with the norm } M_1.$$

Then

$$T : L_p(U, d\mu) \to L_q(V, d\nu)$$

with the norm $M \leq M_0^{1-\theta} M_1^{\theta}$, *where* $\theta \in (0, 1)$, *and*

$$\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1},\ \ \frac{1}{q} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}.$$

This interpolation theorem shows that weak conditions at the extreme points $(\frac{1}{p_0}, \frac{1}{q_0})$ and $(\frac{1}{p_1}, \frac{1}{q_1})$ still guarantee the boundedness of the operator for intermediate values.

The following theorem applies to the case of the Lorentz spaces. Let us recall the definition of the Lorentz space. Let f^* denote the non-increasing rearrangement of a function f. By $L_{p,q}$ we denote the Lorentz space of all functions f μ-measurable on Ω for which

$$||f||_{L_{p,q}} = \left(\int_0^{\infty} \left(t^{\frac{1}{p}} f^*(t)\right)^q \frac{dt}{t} \right)^{\frac{1}{q}},\ \ 1 \leq p < \infty,\ \ 1 \leq q \leq \infty$$

(with the usual supremum interpretation when $q = \infty$).

Theorem 2 (General Marcinkiewicz's interpolation theorem [1]) *Suppose that*

$$T : L_{p_0,\tau_0}(U, d\mu) \to L_{q_0,s_0}(V, d\nu) \quad \textit{with the norm } M_0$$

and

$$T : L_{p_1,\tau_1}(U, d\mu) \to L_{q_1,s_1}(V, d\nu) \quad \textit{with the norm } M_1,$$

where $p_0 < p_1,\ q_0 < q_1$. *Let* $\frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1},\ \ \frac{1}{q} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}$. *Then*

$$T : L_{p,\tau}(U, d\mu) \to L_{q,\tau}(V, d\nu),\ \ 0 < \tau \leq \infty$$

with the norm $M \leq c M_0^{1-\theta} M_1^{\theta}$, *where* $\theta \in (0, 1)$. *In particular, if* $p \leq q$, *then*

$$T : L_p(U, d\mu) \to L_q(V, d\nu).$$

This interpolation result was appeared in the note of Marcinkiewicz in 1939, however the proof of Marcinkiewicz's theorem was published by Zigmund in 1956. Further many mathematicians investigated the generalizations of Marcinkiewicz's theorem for the abstract spaces. Now we consider simple application of such interpolation results.

Example 1 (Young's inequality) Let $U = V = \mathbb{R}^n$ and $d\mu = dx$ (Lebesque measure). Consider the convolution operator

$$K * f = \int_{\mathbb{R}^n} K(x-y)f(y)dy,$$

where $K \in L_r(\mathbb{R}^n),\ 1 \le r \le \infty$. Then by Hölder's inequality we have

$$\|K * f\|_{L_\infty} \le \|K\|_{L_r}\|f\|_{L_{r'}}\ ,$$

where $\frac{1}{r} + \frac{1}{r'} = 1$, and by generalized Minkowski's inequality

$$\|K * f\|_{L_r} \le \|K\|_{L_r}\|f\|_{L_1}.$$

So, we have boundedness

$$K * f : L_r \times L_{r'} \to L_\infty \quad \text{with the norm } \ M_0 \le \|K\|_{L_r},$$

$$K * f : L_r \times L_1 \to L_r \quad \text{with the norm } \ M_1 \le \|K\|_{L_r}.$$

By using complex interpolation (Theorem 1) we obtain boundedness

$$K * f : L_r \times L_p \to L_q,$$

where $\frac{1}{p} = \frac{1-\theta}{r'} + \frac{\theta}{1},\ \frac{1}{q} = \frac{1-\theta}{\infty} + \frac{\theta}{r}$, which gives $1 + \frac{1}{q} = \frac{1}{p} + \frac{1}{r}$, with the norm $M \le \|K\|_{L_r}$. Equivalently, we can formulate the Young's inequality: if $1 \le p, q, r \le \infty$ and $1 + \frac{1}{q} = \frac{1}{p} + \frac{1}{r}$, then

$$\|K * f\|_{L_q} \le \|K\|_{L_r}\|f\|_{L_p}.$$

By using real interpolation (Theorem 2) we obtain

$$\|K * f\|_{L_{q,\tau}} \le c\|K\|_{L_r}\|f\|_{L_{p,\tau}},$$

where $1 < p, q, r < \infty$ and $1 + \frac{1}{q} = \frac{1}{p} + \frac{1}{r}$.

Remark 1 If in the last inequality we only assume that $K \in L_{r,\infty}(\mathbb{R}^n)$, then this stronger result is due to O'Neil [14]. Most of the other classical inequalities were proved by using such interpolation results and were useful for numerous applications, fractional differential, embedding theorems etc.

3 Morrey-type spaces. Properties and examples.

We shall start from the definition of classical Morrey spaces.

Definition 1 (C. Morrey, 1938) Let $0 < p \leq \infty, 0 \leq \lambda \leq \frac{n}{p}$. We say that a function f belongs to the Morrey space $M_p^\lambda(\mathbb{R}^n)$ if $f \in L_p^{loc}(\mathbb{R}^n)$, and the following expression is finite

$$||f||_{M_p^\lambda} \equiv ||f||_{M_p^\lambda(\mathbb{R}^n)} = \sup_{x\in\mathbb{R}^n} \sup_{r>0} r^{-\lambda}||f||_{L_p(B(x,r))}.$$

Here $B(x,r)$ is the open ball of radius $r > 0$ with center at point $x \in \mathbb{R}^n$.

Here, compared with the original definition in [11], we write $r^{-\lambda}$ instead of $r^{-\frac{\lambda}{p}}$ and change the restriction on the parameter λ respectively. And also in [11] $p \in [1, \infty]$, but there is no problem if we extend it to $(0, \infty]$. Note also that the space M_p^λ is a Banach space for $1 \leq p \leq \infty$ and a quasi-Banach space for $0 < p < 1$.

Now we consider the example, which shows relation with the Lebesgue space.

Example 2

1) If $\lambda = 0,\ 0 < p \leq \infty$, then $M_p^\lambda(\mathbb{R}^n) = L_p(\mathbb{R}^n)$.

2) If $\lambda = \frac{n}{p},\ 0 < p < \infty$, then $M_p^{\frac{n}{p}}(\mathbb{R}^n) = L_\infty(\mathbb{R}^n)$.

3) If $\lambda < 0$ or $\lambda > \frac{n}{p}$, then $M_p^\lambda = \Theta$, where Θ is the set of all functions equivalent to 0 on $\mathbb{R}^n$.

Remark 2 Note that the space M_p^λ does not coincide with a Lebesque space if and only if $0 < p < \infty$ and $0 < \lambda < \frac{n}{p}$. Moreover, $L_\infty \cap L_p \subset M_p^\lambda$.

Next we introduce some generalized Morrey-type spaces.

Definition 2 Let $0 < p < \infty,\ 0 < q \leq \infty$ and $0 \leq \alpha \leq \frac{n}{p}$. For any $f \in L_p^{loc}(\mathbb{R}^n)$ we set for $q < \infty$,

$$||f||_{M_{p,q}^\alpha} = \left(\int_0^\infty \left(t^{-\alpha} \sup_{x\in\mathbb{R}^n} \|f\|_{L_p(B(x,t))} \right)^q \frac{dt}{t} \right)^{1/q}, \tag{1}$$

and for $q = \infty$,

$$||f||_{M_{p,\infty}^\alpha} = \sup_{x\in\mathbb{R}^n} \sup_{t>0} t^{-\alpha}\|f\|_{L_p(B(x,t))}.$$

We define the space $M_{p,q}^\alpha$ as the set of all functions $f \in L_p^{loc}(\mathbb{R}^n)$ such that $||f||_{M_{p,q}^\alpha} < \infty$. Taking this fact into account, we shall always assume that $0 < \alpha < \frac{n}{p}$ for $q < \infty$ and $0 \leq \alpha \leq \frac{n}{p}$ for $q = \infty$ when considering the spaces $M_{p,q}^\alpha$. Note that if $\alpha < 0$

or $\alpha = 0$ or $\alpha \geq \frac{n}{p}$ for $q < \infty$ and if $\alpha > \frac{n}{p}$ for $q = \infty$, then $M^{\alpha}_{p,q} = \Theta$. Note also that if $q = \infty$, then the space $M^{\alpha}_{p,q}$ coincides with the classical Morrey space, i.e.

$$M^{\alpha}_{p,\infty} = M^{\lambda}_{p}.$$

We also need a local variant of (1), which was defined in [5] as follows:

Definition 3 Let $0 < p < \infty$, $0 < q \leq \infty$ and $0 \leq \alpha \leq \frac{n}{p}$. The local Morrey-type spaces are defined as the space of all functions f Lebesque measurable on $\mathbb{R}^n$ with finite quasi-norm

$$||f||_{LM^{\alpha}_{p,q}} = \left(\int_0^{\infty} \left(t^{-\alpha} \|f\|_{L_p(B(0,t))} \right)^q \frac{dt}{t} \right)^{1/q}$$

if $0 < q < \infty$, and with the usual supremum interpretation when $q = \infty$.

Next we consider embedding properties of Morrey-type spaces.

Lemma 1

(i) *If* $1 \leq p_0 < p_1 < \infty$, *then*

$$M^{\alpha_1}_{p_1,q} \hookrightarrow M^{\alpha_0}_{p_0,q},$$

where $\alpha_0 = \alpha_1 - \frac{n(p_0 - p_1)}{p_0 p_1}$;

(ii) *If* $0 < q_0 < q_1 \leq \infty$, *then*

$$M^{\alpha}_{p,q_0} \hookrightarrow M^{\alpha}_{p,q_1}.$$

Remark 3 Lemma 1 also holds for local Morrey-type spaces, because the proof follows the same scheme.

Lemma 2 (Hölder's inequality for $M^{\alpha}_{p,q}$) *Let* $p, p_0, p_1 \in [1, \infty)$, $q, q_0, q_1 \in [1, \infty]$, $\frac{1}{p} = \frac{1}{p_0} + \frac{1}{p_1}$, $\frac{1}{q} = \frac{1}{q_0} + \frac{1}{q_1}$, $\alpha = \alpha_0 + \alpha_1$, *then* $fg \in M^{\alpha}_{p,q}$, *and*

$$||fg||_{M^{\alpha}_{p,q}} \leq ||f||_{M^{\alpha_0}_{p_0,q_0}} ||g||_{M^{\alpha_1}_{p_1,q_1}}$$

for all functions $f \in M^{\alpha_0}_{p_0,q_0}$ *and* $g \in M^{\alpha_1}_{p_1,q_1}$.

Consider the following examples for local Morrey-type spaces.

Example 3 Let $\beta \in \mathbb{R}$, $0 < p < \infty$, $0 < q \leq \infty$, $0 < \alpha < \frac{n}{p}$ if $q < \infty$, and $0 \leq \alpha \leq \frac{n}{p}$ if $q = \infty$. If $q < \infty$, then $|x|^{\beta} \notin LM^{\alpha}_{p,q}$ for all $\beta \in \mathbb{R}$. If $q = \infty$, then $|x|^{\beta} \in LM^{\alpha}_{p,\infty} \Leftrightarrow \beta = \alpha - \frac{n}{p}$.

Example 4 Let $\beta \in \mathbb{R}$, $0 < p < \infty$, $0 < q \leq \infty$, $0 < \alpha < \frac{n}{p}$ if $q < \infty$, and $0 \leq \alpha \leq \frac{n}{p}$ if $q = \infty$. Then $|x|^{\beta} \chi_{B(0,1)}(x) \in LM^{\alpha}_{p,q} \Leftrightarrow \beta > \alpha - \frac{n}{p}$ if $q < \infty$ and $\beta \geq \alpha - \frac{n}{p}$ if $q = \infty$.

Example 5 Let $\beta \in \mathbb{R},\ 0 < p < \infty,\ 0 < q \leq \infty,\ 0 < \alpha < \frac{n}{p}$ if $q < \infty$, and $0 \leq \alpha \leq \frac{n}{p}$ if $q = \infty$. Then $|x|^{\beta}(1 + |\ln |x||)^{\gamma}\chi_{B(0,1)}(x) \in LM^{\alpha}_{p,q} \Leftrightarrow \beta > \alpha - \frac{n}{p},\ \gamma \in R$ or $\beta = \alpha - \frac{n}{p},\ \gamma < -\frac{1}{q}$ for $q < \infty$ and $\gamma \leq 0$ for $q = \infty$. (Here $\chi_{B(0,1)}$ – the characteristic function of the ball $B(0,1)$.)

4 Interpolation properties of Morrey-type spaces

We shall now investigate the interpolation behavior of Morrey-type spaces, which was considered in many works in the 60-s. We start with

Theorem 3 (Stampacchia [17], Campanato-Murthy [8]) *Let T be a linear operator such that*

$$T : L_{P_0} \to M^{\lambda_0}_{p_0},$$

$$T : L_{P_1} \to M^{\lambda_1}_{p_1}.$$

Then

$$T : L_P \to M^{\lambda}_{p},$$

where $\frac{1}{P} = \frac{1-\theta}{P_0} + \frac{\theta}{P_1},\ \frac{1}{p} = \frac{1-\theta}{p_0} + \frac{\theta}{p_1}$ *and* $\lambda = (1-\theta)\lambda_0 + \theta\lambda_1,\ 0 \leq \theta \leq 1$.

Actually one can replace L_{P_0}, and L_{P1} by abstract space A_0 and A_1, and L_P by an abstract interpolation space $A = (A_0, A_1)_{\theta,\infty}$. In particular, in [15] it is proved that

$$(M^{\lambda_0}_p, M^{\lambda_1}_p)_{\theta,\infty} \subset M^{\lambda}_p\,,$$

where $\lambda = (1-\theta)\lambda_0 + \theta\lambda_1$. In Ruiz and Vega [16], Blasco, Ruiz and Vega [3] it is proved that such inclusion is strict, i.e.

$$(M^{\lambda_0}_p, M^{\lambda_1}_p)_{\theta,\infty} \neq M^{\lambda}_p\,.$$

However there is still interesting interpolation theorem involved with generalized Morrey-type spaces including classical Morrey spaces. We state our main result – the analog of Marcinkiewicz's interpolation theorem for $M^{\alpha}_{p,q}$.

Theorem 4 *Let $0 < \alpha_0 < \alpha_1 < \frac{n}{p},\ 0 < \beta_0 < \beta_1 < \frac{n}{q},\ 0 < p, q < \infty$, and let T be a linear operator. Suppose that the following inequalities hold for all $z \in \mathbb{R}^n$*

$$||Tf||_{LM^{\alpha_i}_{p,\infty,z}} \leq M_i ||f||_{LM^{\beta_i}_{q,1,z}}, \quad i = 0, 1.$$

Then

$$||Tf||_{M^{\alpha}_{p,\tau}} \leq c M_0^{1-\theta} M_1^{\theta} ||f||_{M^{\beta}_{q,\tau}},$$

in particular, if $\tau = \infty$, then

$$||Tf||_{M^{\lambda}_{p}} \leq c M_0^{1-\theta} M_1^{\theta} ||f||_{M^{\nu}_{q}},$$

where $\alpha = (1-\theta)\alpha_0 + \theta\alpha_1,\ \beta = (1-\theta)\beta_0 + \theta\beta_1,\ 0 < \tau \leq \infty,\ 0 < \theta < 1$, and c depends only on the parameters $p, \beta_0, \beta_1, \theta$.

Remark 4 Here we denote by $LM^{\alpha}_{p,q,z}$ the usual local Morrey-type spaces for a fixed $z \in \mathbb{R}^n$, in which case the behavior of the quasi-norm $\|f\|_{L_p(B(z,t))}$ is important only in a neighborhood of the point z. However, if f belongs to the generalized Morrey-type spaces $M^{\alpha}_{p,q}$, then the uniform in $z \in \mathbb{R}^n$ behavior of the quasi-norm $\|f\|_{L_p(B(z,t))}$ is assumed.

Application of Theorem 4 allows us to obtain different statements for the singular operators in Morrey spaces and their generalizations.

5 Some applications

• Interpolation technique is the instrument in the studying of integral operator in the space of integrable functions.

We present our result – Young-O'Neil-type inequality for generalized Morrey-type spaces. First of all we have proved boundedness of the convolution operator for local Morrey-type spaces.

Theorem 5 *Let $1 < p, q < \infty$ and $0 < \nu < \frac{n}{p}$ $0 < \lambda < \frac{n}{q}$. Let $1 + \frac{1}{q} = \frac{1}{p} + \frac{1}{r} + \frac{\lambda - \nu}{n}$. If $f \in LM^{\nu}_{p,\infty,z}$ and $K \in L_{r,\infty}(\mathbb{R}^n)$, then the convolution operator $K * f \in LM^{\lambda}_{q,\infty,z}$. Moreover, the following estimate holds for all $z \in \mathbb{R}^n$*

$$||K * f||_{LM^{\lambda}_{q,\infty,z}} \leq c||K||_{L_{r,\infty}(\mathbb{R}^n)}||f||_{LM^{\nu}_{p,\infty,z}},$$

where c independent of function K.

So, by applying Theorem 4 and Theorem 5, we prove Young-O'Neil-type inequality for the generalized Morrey-type spaces, in particular, for classical Morrey spaces.

Theorem 6 *Let $1 < p, q < \infty$, $0 < \tau \leq \infty$, $0 < \nu < \frac{n}{p}$, $0 < \lambda < \frac{n}{q}$, and $1 + \frac{1}{q} = \frac{1}{p} + \frac{1}{r} + \frac{\lambda - \nu}{n}$. If $f \in M^{\nu}_{p,\tau}$ and $K \in L_{r,\infty}(\mathbb{R}^n)$, then the convolution operator $K * f \in M^{\lambda}_{q,\tau}$, and the following estimate holds*

$$||K * f||_{M^{\nu}_{p,\tau} \to M^{\lambda}_{q,\tau}} \leq c||K||_{L_{r,\infty}(\mathbb{R}^n)},$$

in particular, if $\tau = \infty$, then

$$||K * f||_{M^{\nu}_{p} \to M^{\lambda}_{q}} \leq c||K||_{L_{r,\infty}(\mathbb{R}^n)},$$

where c independent of function K.

• Interpolation technique can also be used to obtain various limiting cases of the well-known "regularity theorems" of boundary value problems: Let u be a fundamental solution of an elliptic partial differential operator of $Au = f$, where $f \in L_{\infty}$, then all partial derivatives $D^m u$ of order m are locally in M^{λ}_{p}. (See [15].)

Open problems

• Extend the boundedness of the convolution operator up to obtaining the low estimates of the convolution in generalized Morrey-type spaces.
• The author intends to consider in the future the Fourier transform in Morrey-type spaces, namely to study relations between summability of Fourier coefficients and integrability of functions, which belong to the Morrey-type spaces. It is also useful to obtain some estimates and various integral operators on these spaces.

Acknowledgements. The author thanks the University of Padova for the opportunity to have an internship, and expresses her gratitude to Pier Domenico Lamberti for useful comments. This work was partially supported by the grants of the Ministry of education and science of the Republic of Kazakhstan (projects 1412/ГФ MOH PK and 0744/ГФ MOH PK).

References

[1] J. Bergh, J. Löfström, "Interpolation spaces. An introduction". Springer-Verlag, Berlin-Heidelberg-New York, 1976.

[2] Yu.A. Brudnyi, S.G. Krein, E.M. Semenov, *Interpolation of linear operators.* Itogi Nauki i Tekhn. Ser. Mat. Anal. 24 (1986), 3–163.

[3] O. Blasco, A. Ruiz, L. Vega, *Non interpolation in Morrey-Campanato and block spaces.* Ann. Scuola Norm. Super. Pisa 28/1 (1999), 31–40.

[4] A.P. Blozinski, *On a convolution theorem for $L_{p,q}$ spaces.* Trans. Amer. Math. Soc. 164 (1972), 255–265.

[5] V.I. Burenkov, H.V. Guliyev, V.S. Guliyev, *Necessary and sufficient conditions for boundedness of the fractional maximal operator in the local Morrey-type spaces.* Doklady Ross. Akad. Nauk. Matematika 409 (2006), 443–447.

[6] V.I. Burenkov, D.K. Darbayeva, E.D. Nursultanov, *Description of interpolation spaces for general local Morrey-type spaces.* Eurasian Mathematical Journal 4/1 (2013), 46–53.

[7] V.I. Burenkov, E.D. Nursultanov, *Description of interpolation spaces for local Morrey-type spaces.* Trudy Math. Inst. Steklov 269 (2010), 46–56.

[8] A. Campanato, M.K.V. Murthy, *Una generalizzazione del teorema di Riesz-Thorin* (Italian). Ann. Scuola Norm. Super. Pisa 19 (1965), 87–100.

[9] R.A. Kerman, *Convolution theorems with weights.* Trans. Amer. Math. Soc. 280/1 (1983), 207–219.

[10] A.G. Kostyuchenko, E.D. Nursultanov, *On integral operators in L_p-spaces.* Fundam. Prikl. Mat. 5/2 (1999), 475–491.

[11] C.B. Morrey, *On the solution of quasi-linear elliptic partial differential equations.* Trans. Amer. Math. Soc. 43 (1938), 126–166.

[12] E.D. Nursultanov, S. Tikhonov, *Convolution inequalities in Lorentz spaces.* J. Fourier Anal. Appl. 17 (2011), 486–505.

[13] E.D. Nursultanov, S. Tikhonov, *Net spaces and boundedness of integral operators.* Journal of Geometric Analysis 21/4 (2011), 950–981.

[14] R. O'Neil, *Convolution operators and $L(p,q)$ spaces.* Duke Math. J. 30 (1963), 129–142.

[15] J. Peetre, *On the theory of $\mathcal{L}_{p,\lambda}$ spaces.* J. Funct. Anal. 4 (1969), 71–87.

[16] A. Ruiz, L. Vega, *Corrigenda to "Unique continuation for Schrödinger operators" and a remark on interpolation on Morrey spaces.* Publ. Mat., Barc. 39 (1995), 405–411.

[17] G. Stampacchia, *$\mathcal{L}_{p,\lambda}$-spaces and interpolation.* Comm. Pure Appl. Math. 17 (1964), 293–306.

[18] V.D. Stepanov, "Some topics in the theory of integral convolution operators". Dalnauka. Vladivostok, 2000.

[19] H. Triebel, "Interpolation theory. Function spaces. Differential operators". VEB, Berlin, 1978.

Almost integrable Hamiltonian systems and the approach of the Perturbation theory

Gabriella Schirinzi (*)

1 Introduction

Almost integrable Hamiltonian systems are dynamical systems which can be described by the Hamiltonian formulation and are not integrable according to Liouville-Arnol'd theorem. Nevertheless, their Hamiltonian can be written as a small perturbation of an integrable one. One may easily conclude that since the Hamiltonian is very close to an integrable one, then the same holds also for the dynamics of such kind of systems. But unfortunately this is not true, hence it's necessary to find suitable techniques to study the behaviour of almost integrable systems. This is what the Hamiltonian Perturbation Theory does: it studies almost integrable systems with the purpose of controlling their evolution in time comparing the dynamics to the one of the correspondent integrable systems.
I will shortly recall some basical notions about Lagrangian and Hamiltonian formulations of dynamical systems; in particular for Hamiltonian systems I will state the celebrated Liouville- Arnold's theorem about integrable systems. Then I will introduce almost integrable systems and the main difficulties in studying their dynamics. Finally I will introduce the two main theorems of the Hamiltonian Perturbation theory, which can overcome such difficulties and prove some stability properties for almost integrable systems.

2 The Lagrangian formulation

In what follows we will denote by a dot the partial derivative with respect to the time.
Let us consider a system of particles with n degrees of freedom. Suppose all forces involved are conservative, and possible constraints are holonomic and time independent. As well

(*)Ph.D. course, Università di Padova, Dip. Matematica, via Trieste 63, I-35121 Padova, Italy; E-mail: schiri@math.unipd.it . Seminar held on April 24th, 2013.

known, the motion of such a system is described by the Newton's second law as well as by the Lagrangian formulation:

$$(1) \qquad \begin{cases} \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_1}\right) = \frac{\partial L}{\partial q_1} \\ \\ \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_2}\right) = \frac{\partial L}{\partial q_2} \\ \\ \vdots \\ \frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_n}\right) = \frac{\partial L}{\partial q_n} \end{cases}$$

This formulation consists of n differential equations of the second order in the n independent variables $q_1, \dots, q_n$, which are called *generalized variables*. The function $L(q, \dot{q})$ [(†)] is the *Lagrangian function* and is given by the difference between the total kinetic energy of the system $T(q, \dot{q})$ and the total potential energy $V(q)$: $L(q, \dot{q}) = T(q, \dot{q}) - V(q)$.

Let us recall some useful definitions:

Definition 2.1 (Configuration space) the n-dimensional space of the values of the generalized coordinates $(q_1, \dots, q_n)$

Definition 2.2 (Phase space) the $2n$-dimensional space of the values of the generalized coordinates and their velocities $(q_1, \dots, q_n, \dot{q}_1, \dots, \dot{q}_n)$

Definition 2.3 (Trajectories of the motion) the solutions $(q_1(t), q_2(t), \dots, q_n(t))$ of system (1) (which are curves in the configuration space)

When we have a dynamical system (1) and we are able to write explicitly the solutions, then we know everything about its dynamics: given an initial state of the system at a certain time, we can predict the behaviour of the motion at any moment. But unfortunately, not always we can calculate the solutions of (1). Nevertheless, there can exist some first integrals of the motion, which may help us simplify the equations in (1) and deduce useful informations about the dynamics. Let us recall what a first integral is:

Definition 2.4 (First integral of the motion) a function which has constant value along any trajectories of the motion (is conserved), that is of the form:

$$f(q_1(t), \dots, q_n(t), \dot{q}_1(t), \dots, \dot{q}_n(t)) = \text{constant}$$

Example 1 In conservative systems, the total energy $E = T + V$ is a first integral of the motion.

From a geometrical point of view, a first integral describes a surface in the phase space with the following property: if the motion starts on such a surface, then it is bound to

[(†)] For short we denote by q the whole set of variables $(q_1, q_2, \dots, q_n)$, and we do analogously for any other set of variables

move on it for all times. Forcing motions to stay on invariant surfaces, first integrals reduce the number of degrees of freedom of the system, hence they simplify the equations of the motion. In particular, when there is a sufficient number of first integrals, the equations of the motion become integrable. In fact, given a generic differential system (not necessarily a Lagrangian system) with $2n-$dimensional phase space, if it admits $2n-1$ independent first integrals, then it can be solved by quadratures.
There exists an alternative formulation of the motion of a dynamical system, still consisting in a differential system, for which it has been proved an integrability result based on a lower number of first integrals of the motion (but satisfying suitable conditions). This is the so called *Hamiltonian formulation of the motion.*

3 The Hamiltonian formulation

The Hamiltonian formulation of the motion of a dynamical system can be derived from the Lagrangian one, and consists of the following system of $2n$ differential equations of the first order:

$$\begin{cases} \dot{q}_i(t) = \frac{\partial H}{\partial p_i} \\ \\ \dot{p}_i(t) = -\frac{\partial H}{\partial q_i} \qquad i = 1, \dots, n \end{cases} \tag{2}$$

The equations depend on $2n$ independent variables $q_1, \dots, q_n, p_1, \dots, p_n$ called *canonical variables*, where the first n are the generalized variables, and the last n are the correspondent *conjugate momenta*, given by:

$$p_i := \frac{\partial L}{\partial \dot{q}_i} \tag{3}$$

The function $H(q,p)$ is the *Hamiltonian function* and can be obtained from the Lagrangian $L(q,\dot{q})$ by a change of coordinates, from the old coordinates $(q,\dot{q})$ to the new ones (q,p). Such a transformation is known under the name of Legendre transformation, and can be summarized as follows:

(a) suppose we can invert (3) with respect to $\dot{q}_i$: $\dot{q}_i = \phi_i(q,p)$

(b) define the *generalized energy*:

$$\widetilde{H}(q,\dot{q}) := \sum_{i=1}^{n} \frac{\partial L}{\partial \dot{q}_i}\dot{q}_i - L \tag{4}$$

(c) substitute (3) in (4) and we get the Hamiltonian:

$$H(q,p) = \widetilde{H}(q,\phi(q,p)) = \sum_{i=1}^{n} p_i\phi_i(q,p) - L(q,\phi(q,p))$$

We underline that under our hypothesis, the Langrangian function doesn't depend explicitly on time and, consequently, the same holds for the Hamiltonian. It is possible to prove that when the Hamiltonian doesn't depend explicitly on time, it is a first integral of its motion.
The Hamiltonian formulation turns out to be very useful when there are cyclic variables. Let us recall the definition of cyclic variable:

Definition 3.1 (Cyclic variable) a variable which doesn't appear explicitly in the Hamiltonian

When there is a cyclic variable, there is also a first integral of the motion, and the Hamilton's equations appear simpler. To understand why, let us consider an example.

Example 2 Suppose we have a Hamiltonian function such that all the variables q_i, $i = 1, \dots, n$, are cyclic: $H(q, p) = H(p)$. Then, from Hamilton's equations (2) we conclude that all the variables p_i, $i = 1, \dots, n$, are first integrals of the motion, in fact:

$$\dot{p_i}(t) = -\frac{\partial H}{\partial q_i} \equiv 0 \Longrightarrow p_i = \alpha_i \quad \forall i \quad \text{(constants)}$$

Consequently, Hamilton's equations are simply:

$$\begin{cases} \dot{q_i}(t) = \frac{\partial H}{\partial p_i} =: \omega_i(\alpha_i) \\ \\ \dot{p_i}(t) = 0 \end{cases} \Longrightarrow \begin{cases} q_i(t) = \omega_i(\alpha_i)t + \beta_i \\ \\ p_i(t) = \alpha_i \qquad i = 1, \dots, n \end{cases}$$

(β_i and α_i depend on initial conditions). In this case the equations of the motion can be easily integrated, and the trajectories of the system are linear motions lying on the invariant surfaces described by the first integrals $p_i = \alpha_i$, $i = 1, \dots, n$.

Now, one may wonder if a Hamiltonian system with $2n$-dimensional phase space, having n first integrals of the motion, does always admit a choice of canonical variables such that n of them are cyclic. This turns out to be true only under suitable conditions, as it is stated in the celebrated *Liouville-Arnol'd Theorem*, which characterizes integrable Hamiltonian systems. Before stating the theorem, let us recall a useful definition:

Definition 3.2 (Poisson Brackets) Given the functions $u(q, p)$ and $v(q, p)$, their Poisson Bracket is defined as follows:

$$\{u, v\} = \sum_{i=1}^{n} \Big(\frac{\partial u}{\partial q_i} \frac{\partial v}{\partial p_i} - \frac{\partial u}{\partial p_i} \frac{\partial v}{\partial q_i} \Big)$$

Moreover u and v are said to be in mutual involution if

$$\{u, v\} = 0$$

Theorem 3.3 (Liouville-Arnol'd Theorem) *Suppose:*

- *we have a Hamiltonian* $H(q,p) : M \to \mathbb{R}$ *and* n *first integrals of the motion* $f_i(q,p) : M \to \mathbb{R}$, $i = 1, \dots, n$, *where* $M \subseteq \mathbb{R}^{2n}$ *is a symplectic manifold*
- $f_1, \dots, f_n$ *are all in* ***mutual involution*** *each other*
- $f_1, \dots, f_n$ *are* ***functionally independent*** *on a level set* $N_c = \{(q,p) \in M : f_i(q,p) = c_i,\ i = 1, \dots, n\}$, $\quad c = (c_1, \dots, c_n)$, *and assume* N_c *is compact and connected*

$\Longrightarrow$ *then in a neighbourhood of* N_c *we can introduce the so called* <u>*action-angle variables* (I, φ)</u> *and in the new variables the Hamiltonian depends only on the actions:*

$$H'(I, \varphi) = H'(I) \tag{5}$$

The Hamiltonian (5) has the same form as the Hamiltonian in Example 2, hence the Hamilton's equations are:

$$\begin{cases} \dot{I}(t) = 0 \\ \\ \dot{\varphi}(t) = \omega(I), \quad \text{with } \omega(I) := \frac{\partial H'}{\partial I} \end{cases}$$

with trivial solutions $I(t) = I(0)$ and $\varphi(t) = \omega(I(0))t + \varphi(0)$. In particular, Liouville-Arnol'd theorem states that the neighbourhood of N_c is diffeomorphic to the n-dimensional torus $\mathbb{T}^n$, hence any motion of a system satisfying the hypotheses of the theorem, is linear on the invariant torus $I = I(0)$, with constant velocity $\omega(I(0))$.

Examples 3 Here are some examples of systems which are integrable according to Liouville-Arnol'd theorem:

- all Hamiltonian systems with one degree of freedom
- the 2-body problem
- some special systems (Euler rigid body, Lagrange spinning top...)

Then, thanks to Liouville-Arnol'd theorem, we have a complete characterization of the motions of those systems which are integrable according to this theorem. Nevertheless, most of the real physical systems are not integrable. In fact, the existence of a sufficient number of functionally independent first integrals and their mutual involution, are very restrictive hypotheses, which rarely are satisfied.

Example 4 The n-body problem is not integrable.

3.1 Almost integrable Hamiltonian systems

Even if they are not integrable according to Liouville-Arnol'd theorem, many of the real physical systems can be put in a form which is called *almost integrable*. It means they can be described by a Hamiltonian which appears as a small perturbation of an integrable one:

$$H(I,\varphi) = h(I) + \varepsilon f(I,\varphi) \tag{6}$$

The function $h(I)$ is an integrable Hamiltonian, in fact it depends only on the action variables; ε is a small parameter ($|\varepsilon| \ll 1$) which measures the intensity of the perturbation $f(I,\varphi)$, which is not integrable. The associated Hamiltonian system is

$$\begin{cases} \dot{\varphi}(t) = \frac{\partial H}{\partial I} &= \omega(I) + \varepsilon \frac{\partial f}{\partial I} \\ \dot{I}(t) = -\frac{\partial H}{\partial \varphi} &= -\varepsilon \frac{\partial f}{\partial \varphi} \end{cases} \tag{7}$$

where $\omega(I) := \frac{\partial h}{\partial I}$ is the frequency vector of the integrable Hamiltonian $h(I)$.

The fact that an almost integrable Hamiltonian function is close to an integrable one, doesn't mean that the same holds for the dynamics. In fact, orbits of an almost integrable system can get very far from the ones of the correspondent integrable system, and can be also very irregular. The **Hamiltonian Perturbation Theory** studies the behaviour of almost integrable Hamiltonian systems: in particular, it searches the longest possible time-scale on which action variables remain close to the ones of the correspondent integrable system (**stability**).

4 Hamiltonian Perturbation Theory

It is given an almost integrable Hamiltonian (6): we want to control the variation of the action variables such that they remain close to the ones of the integrable system (which are $I(0)$) for the longest possible time. First we observe from Hamilton's equations it follows a first, trivial, estimate for the actions. In fact, if we assume the perturbation $f(I,\varphi)$ is bound in norm by A, then from (7) we have the so called **a priori estimate**:

$$|I(t) - I(0)| \leq \varepsilon A|t| = \sqrt{\varepsilon} A \sqrt{\varepsilon}|t| \tag{8}$$

Estimate (8) implies that up to times of the order of $1/\sqrt{\varepsilon}$ the variation of the actions remains $\sqrt{\varepsilon}$-limited.
We would like to improve such an estimate, that is to extend the stability time as much as possible. But to do that we need to reduce the perturbation, increasing its order. In fact, suppose the order of the perturbation is ε^2; then, from Hamilton's equations (7) it follows:

$$|I(t) - I(0)| \leq \varepsilon^2 A|t| = \sqrt{\varepsilon} A \left(\varepsilon \sqrt{\varepsilon}|t|\right)$$

In this case the variation of the actions remains $\sqrt{\varepsilon}$-limited up to times of the order of $1/\varepsilon\sqrt{\varepsilon}$, which is a time longer than the one in the a priori estimate.

If we want to reduce the perturbation, we need to find a suitable change of variables, and if we are able to do it, it means we can perform a so called **perturbative step**.

4.1 The fundamental equation of the perturbation theory

There are two methods for changing variables in a Hamiltonian system:

- the **generating function** method
- the **Lie series** method

When we try to perform a perturbative step, both methods lead to the same differential equation in the unknown function $\chi(I, \varphi)$:

$$\boxed{\omega(I) \cdot \frac{\partial \chi}{\partial \varphi}(I, \varphi) = f(I, \varphi)} \tag{9}$$

This is the **fundamental equation of the perturbation theory**.

A perturbative step consists in searching a solution to equation (9). Suppose $f(I, \varphi)$ and $\chi(I, \varphi)$ can be expressed in Fourier series:

$$f(I, \varphi) = \sum_{k \in \mathbb{Z}^n} f_k(I) e^{ik \cdot \varphi} \qquad \chi(I, \varphi) = \sum_{k \in \mathbb{Z}^n} \chi_k(I) e^{ik \cdot \varphi}$$

then equation (9) implies:

$$\chi(I, \varphi) = -i \sum_{k \in \mathbb{Z}^n / \{0\}} \frac{f_k(I)}{\omega(I) \cdot k} e^{ik \cdot \varphi} \tag{10}$$

The function $\chi(I, \varphi)$ is well defined only for those actions in the phase space such that:

$$\omega(I) \cdot k \neq 0 \qquad \forall\, k \in \mathbb{Z}^n / \{0\}$$

that is actions corresponding to *non-resonant motions* (I correspond to a *resonant motion* if $\omega(I) \cdot k = 0$ for some $k \in \mathbb{Z}^n / \{0\}$). Actually, we should keep sufficiently far from resonant motions, hence actions should satisfy the stronger condition, for some positive constants γ, $\beta \in \mathbb{R}$:

$$|\omega(I) \cdot k| \geq \frac{\gamma}{|k|^\beta} \qquad \forall\, k \in \mathbb{Z}^n / \{0\} \tag{11}$$

known under the name of *diophantine condition*.

4.2 The Poincaré's difficulty (small denominators)

Poincaré proved that in general actions satisfying the diophantine condition (11) form in the phase space a dense set. In particular, this happens when the following two conditions hold at the same time:

(a) the components of $\omega(I)$ are *functionally independent*:

$$\det\Big|\frac{\partial\omega}{\partial I}\Big| = \det\Big|\frac{\partial^2 h}{\partial I^2}\Big| \neq 0$$

(b) the perturbation $f(I,\varphi)$ is *generic*, that means for each $k \in \mathbb{Z}^n$ there exists k' parallel to k such that $f_{k'}(I) \neq 0$.

Since the actions satisfying the diophantine condition form a dense set, we cannot ensure convergence for the series defining χ (10) in any open subset of the phase space, and consequently we cannot perform a perturbative step.

4.3 KAM Theorem and Nekhoroshev Theorem

There have been found some ways to overcome the Poincaré's difficulty, in particular two theorems, which are the most important in the Hamiltonian perturbation theory, can perform perturbative steps and then extend the stability time for the motions of almost integrable Hamiltonian systems with suitable small ε. Such theorems are the **KAM Theorem** [(‡)][1954] and the **Nekhoroshev Theorem** [1977]. They make some regularity assumptions on the Hamiltonian function, especially on its integrable part, and prove different stability properties for all, or almost all, initial data in the phase space.

- **KAM Theorem** proves perpetual stability for only non-resonant motions. That is for almost all initial data in the phase space, except a set with small Lebesgue measure $O(\sqrt{\epsilon})$, orbits of an almost integrable system remain close to the ones of the correspondent integrable system forever.

  Application. This theorem has been used to prove stability in the circular restricted three body problem in the planar case.

- **Nekhoroshev Theorem** proves an exponentially long stability for all motions of an almost integrable system. The stability time is given by:

  $$T = t_0 e^{(\frac{\epsilon_0}{\epsilon})^b}$$

  with t_0, ϵ_0, b positive constants.

  Application. This theorem has been used to prove stability of some equilibrium points in the circular restricted three body problem in the spatial case.

(‡) after Kolmogorov, Arnol'd and Moser.

References

[1] Arnold, V.I., "Mathematical Methods of Classical Mechanics". Second Edition, Ed. Springer.

[2] Arnold, V.I., *Proof of a theorem of A.N. Kolmogorov on the preservation of conditionally periodic motions under a small perturbation of the Hamiltonian.* Uspehi, Mat. Nauk 18/5 (113), 13–40 (1963).

[3] G. Benettin, F. Fassò, M. Guzzo, *Nekhoroshev stability of L4 and L5 in the spatial restricted three body problem.* Regul. Chaotic Dyn. 3, 56–72 (1998).

[4] Goldstein H., Poole C., Safko J., "Classical Mechanic". Third Edition, 2002 Pearson Education, Inc., publishing as Addison Wesley.

[5] Kolmogorov, A.N., *On conservation of conditionally periodic motions for a small change in Hamilton's function.* Dokl. Akad. Nauk SSSR (N.S.) 98, 527–530 (1954).

[6] Moser, J., *New aspects in the theory of stability of Hamiltonian systems.* Comm. Pure Appl. Math 11, 81–114 (1958).

[7] Nekhoroshev, N.N., *An exponential estimate of the time of stability of nearly integrable Hamiltonian systems. I.* Usp. Mat. Nauk 32, No. 6, 5–66 (1977); Russ. Math. Surv. 32, 1–65 (1977).

[8] Nekhoroshev, N.N., *An exponential estimate of the time of stability of nearly integrable Hamiltonian systems. II.* Tr. Semin. Petrovsk. 5, 5–50 (1979); In: Oleinik, O.A. (ed.) Topics in Modern Mathematics, Petrovskii Semin., No. 5. New York: Consultant Bureau (1985).

A class of derivative-free nonmonotone algorithms for unconstrained optimization

Francesco Rinaldi [(*)]

Abstract. We first present some basic concepts related to derivative-free optimization. Then, we describe a class of algorithms that makes use of nonmonotone inexact linesearches along a set of search directions satisfying appropriate conditions.

Keywords. Derivative-free Methods, Nonmonotone Techniques, Unconstrained Optimization.

1 Introduction

Extensive useful information is contained in the derivatives of any function one wishes to minimize. Anyway, in many instances (at least some) derivatives are unavailable or unreliable. This situation frequently arises in real-world problems and requires the use of suitably developed approaches i.e. derivative-free approaches. In [1], a comprehensive study of derivative-free methods is given. ¿From now on, we focus on the following problem:

$$\min_{x \in \mathbb{R}^n} f(x),$$

and we assume that:

- $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function;
- first order information is not available.

In derivative-free monotone linesearch methods the key idea is that of finding a new point along a search direction that guarantees

- a sufficient decrease of the objective function;
- a sufficiently large movement along a given direction.

[(*)]Università di Padova, Dipartimento di Matematica, via Trieste 63, I-35121 Padova, Italy; E-mail: **rinaldi@math.unipd.it** . Seminar held on May 22nd, 2013.

The basic ingredients of such algorithms are: a suitably chosen set of search directions and an Armijo-type linesearch. Here, we first give some details about classic linesearch methods, then we briefly describe a class of algorithms that make use of nonmonotone linesearches [2].

2 Search directions

Without first order information, we cannot guarantee to have a descent direction. In other words, we cannot use a single search direction at each iteration (like e.g. in gradient-based methods). In practice, a suitably chosen set of search directions is considered. We give here some basic definitions:

Definition 1 A *positive span* of a set of vectors $\{v_1, \dots, v_r\}$ is the cone

$$\{y \in \mathbb{R}^n : \ y = \sum_{i=1}^{r} \gamma_i v_i, \ \gamma_i \geq 0, \ i = 1, \dots, r\}.$$

Definition 2 A *positive spanning set* in $\mathbb{R}^n$ is a set of vectors whose positive span is $\mathbb{R}^n$.

We report a classic result related to positive spanning sets (see e.g. [1]):

Proposition 3 *If $\{v_1, \dots, v_r\}$ is a positive spanning set, then for every nonzero vector $w \in \mathbb{R}^n$ there exists an index $i \in \{1, \dots, r\}$ for which $w^T v_i > 0$.*

So if $w = -\nabla f(x) \neq 0$, there exists a v_i :

$$-\nabla f(x)^T v_i > 0$$

and v_i is a descent direction.
A set $\{v_1, \dots, v_r\}$ is said to be *positively dependent* if one of its vectors is a positive combination of the others; otherwise, the set is *positively independent*.

Definition 4 A *positive basis* in $\mathbb{R}^n$ is a positively independent set whose positive span is $\mathbb{R}^n$.

Two examples of positive basis for $\mathbb{R}^n$ are:

- the coordinate directions and their negative counterparts, that is

$$\{s_1, \dots, s_{2n}\} = \{e_1, \dots, e_n, -e_1, \dots, -e_n\};$$

- the coordinate directions and the negative of their sum, that is

$$\{s_1, \dots, s_{n+1}\} = \{e_1, \dots, e_n, -\sum_{i=1}^{n} e_i\}.$$

3 Derivative-Free Linesearches and Global Convergence

In derivate-free methods an *Armijo Backtracking Rule* is usually considered. Given a search direction d_k, the goal is finding an α_k such that

$$f(x_k + \alpha_k d_k) \leq f(x_k) - \gamma(\alpha_k \|d_k\|)^2$$

with $\gamma > 0$. The algorithm starts with a given stepsize $\Delta_k > 0$ and stops after a finite number of iterations with a stepsize α_k. At the end of the procedure we can have:

- $\alpha_k \neq 0$: we get a new point along the search direction that guarantees sufficient decrease;
- $\alpha_k = 0$: we have a failure.

An expansion step is usually included in the Linesearch to guarantee that the chosen stepsize is sufficiently large (this enables to give an arbitrary value to Δ_k). The expansion starts every time Δ_k is accepted and Goldstein Conditions:

$$f(x_k + \alpha_k d_k) \leq f(x_k) - \gamma_1(\alpha_k \|d_k\|)^2$$

$$f(x_k + \alpha_k d_k) \geq f(x_k) - \gamma_2(\alpha_k \|d_k\|)^2$$

with $\gamma_1 < \gamma_2$, are not satisfied.

Nonmonotone linesearches can be very useful when dealing with an objective function that is noisy or with steep sided valleys. In this cases, the monotone reduction of the function corresponds to very small movement along the search direction. Nonmonotone acceptance rules can improve robustness and efficiency by imposing sufficient reduction with respect to a reference value W_k that satisfies the condition

$$f(x_k) \leq W_k \leq \max_{0 \leq j \leq \min(k,M)} [f(x_{k-j})], \tag{1}$$

for a given integer $M \geq 0$. Thus, we can define a *Nonmonotone Armijo Backtracking Rule* (see [3] for further details). In practice, given a search direction d_k, the goal is finding an α_k:

$$f(x_k + \alpha_k d_k) \leq W_k - \gamma(\alpha_k \|d_k\|)^2$$

with $\gamma > 0$. In the Armijo Search, the stepsize is reduced until either the condition of sufficient decrease is satisfied, and hence the sign of α_k is fixed, or the length of the tentative step $\alpha\|d_k\|$ becomes smaller than an adjustable bound ρ_k. In the latter case, the value α_k determined by the algorithm is set equal to zero and the last value of α is indicated by $\eta_k > 0$. We report below a scheme of the Nonmonotone Derivative-Free LineSearch (NDFLS) Algorithm.
The following result can be proved for the NDFLS Algorithm [3]:

Proposition 5 *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuously differentiable and assume that the level set $\mathcal{L}_0 = \{x \in \mathbb{R}^n : \ f(x) \leq f(x_0)\}$ is compact.*

(i) *Algorithm* NDFLS *determines, in a finite number of steps, a scalar* α_k *such that*

$$f(x_k + \alpha_k d_k) \leq W_k - \gamma(\alpha_k)^2 \|d_k\|^2. \tag{2}$$

(ii) *Let* $\{x_k\}$ *be a the sequence of points in* $\mathbb{R}^n$ *and let* K *be an infinite index set such that*

$$x_{k+1} = x_k + \alpha_k d_k, \quad \text{for all} \quad k \in K$$

where $d_k \in \mathbb{R}^n$, $d_k \neq 0$ *and* $\alpha_k \in \mathbb{R}$ *is determined by means of Algorithm* NDFLS. *Assume that* $\rho_k \to 0$ *for every infinite subsequence of* $\{x_k\}_K$ *such that* $\alpha_k = 0$. *Then, we have:*

$$\lim_{k\to\infty, k\in K} \frac{\nabla f(x_k)^T d_k}{\|d_k\|} = 0. \quad \square$$

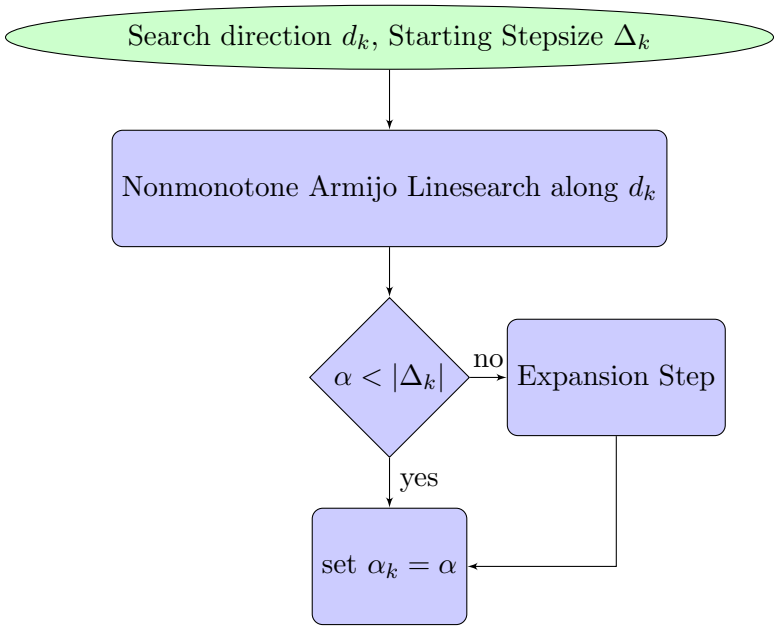


Figure 1: NDFLS Algorithm.

In order to guarantee the global convergence of a derivative free methods, we need to impose conditions on the search directions. In particular, the following assumption [2] needs to be satisfied:

Assumption 6 Let $\{d_k\}$ be the sequence of search directions used in the algorithm. There exists a value $N > 0$ and n integers $j(k, i)$, for $i = 1, \ldots, n$, such that

$$k \leq j(k, 1) \leq j(k, 2) \leq \cdots \leq j(k, n) \leq k + N$$

and the n sequences $\{p_k^i\}$, defined by

$$p_k^i = \frac{d_{j(k,i)}}{\|d_{j(k,i)}\|}, \quad i = 1, \ldots, n,$$

have the property that every limit point $(\bar{p}^1, \bar{p}^2, \ldots \bar{p}^n)$ of $\{(p_k^1, p_k^2, \ldots p_k^n)\}$ positively span $\mathbb{R}^n$. □

In [2], we extend to nonmonotone methods (that use the NDFLS Algorithm and a set of direction satisfying Assumption 6) global convergence conditions already established for derivative-free monotone linesearch-based methods.

4 A conceptual scheme of the algorithms

In this section, we give an informal outline of the class of methods proposed in [2]. All the algorithms generate an infinite sequence

$$x_{k+1} = x_k + \alpha_k d_k, \tag{3}$$

where $d_k \in \mathbb{R}^n$ is a search direction, $\alpha_k \in \mathbb{R}$ is a stepsize along d_k and $x_0 \in \mathbb{R}^n$ is a given point.

Each major step of the algorithms can be divided into three different phases:

(a) *Basic search.* Starting from the current point x_k, we choose a finite set of search directions and we perform a nonmonotone linesearch along each of them. During this phase, when needed, we can further store a set of tentative points $y^j \in \mathbb{R}^n$ computed along the directions and the corresponding function values $f(y^j)$, for $j = 0, 1, \ldots, q$.

(b) *Acceleration step.* Given the available information gathered during phase (a), that is $y^j, f(y^j)$ for $j = 0, 1, \ldots, q$, we determine a new search direction on the basis of some local model of f and perform a nonmonotone linesearch along this direction. The attempt is that of improving substantially the results of phase (a) by computing, for instance, an approximation to the steepest descent direction or by determining a suitable pattern on the basis of the previous steps.

(c) *Rotation of the search directions.* We perform a rotation (according to some given criterion) of the directions.

By using the results considered in the previous section it is possible to prove the global convergence of this class of derivative-free algorithms (see [2] for further details).

References

[1] A.R. Conn, K. Scheinberg, L.N. Vicente, "Introduction to Derivative-Free Optimization". MPS-SIAM Series on Optimization, Philadelphia, PA, 2008.

[2] L. Grippo and F. Rinaldi, *A class of derivative-free nonmonotone optimization algorithms employing coordinate rotations and gradient approximations*. Submitted (2013).

[3] L. Grippo and M. Sciandrone, *Nonmonotone derivative-free methods for nonlinear equations*. Computational Optimization and Applications 27 (2007), 297–328.

An introduction to Ramification theory for number fields

Sophie Marques (*)

Abstract. The question of prime decomposition in a finite extension of fields motivates classical ramification theory for field extensions. We propose to give a very little introduction in this deep subject which can be also extend to arithmetic geometry.
After recalling some basic facts around finite field extensions, we will explain how to do arithmetic in number fields. This will permit us to define the ramification for number field and to give some criterions permitting to decide which primes are ramified or not. Finally, we will study the particular case of quadratic extensions in order to see how we can apply this theory.

In mathematics, ramification is a geometric term used for 'branching out', in the way that the square root function, for complex numbers, can be seen to have two branches differing in sign. It is also used from the opposite perspective (branches coming together) as when a covering map degenerates at a point of a space, with some collapsing together of the fibers of the mapping. In algebraic number theory, roughly speaking, ramification means prime numbers factoring into some repeated prime factors. The purpose of this talk is to understand the meaning of ramification for number fields though very simple examples. The reader who is interested in this topic can find more details in [Sam67], [Lan94] and [Neu99].

1 Overview of finite extensions of fields

Suppose that L/K is a field extension (which means that L is a field and K is a subfield of L).

We call L/K to be **finite** if as a vector space over K, L is of finite dimension; the **degree of** L/K, denoted by $[L : K]$, is defined to be the vector space dimension of L over K. Given $\alpha_1, ..., \alpha_n \in L$, we denote by $K(\alpha_1, ..., \alpha_n)$ (resp. $K[\alpha_1, ..., \alpha_n]$) the smallest subfield (resp. subring) of L containing K and the elements $\alpha_1, ..., \alpha_n$.

(*) Algant, Université de Bordeaux, IMB, UMR 5251, F-33400 Talence, France and Università degli studi di Padova, Dipartimento di Matematica, via Trieste 63, I-35121 Padova, Italy; E-mail: sophie.marques@math.u-bordeaux1.fr . Seminar held on June 19th, 2013.

If L'/K is another extension, then a homomorphism $\sigma : L \to L'$ such that $\sigma(c) = c$ for all $c \in K$ is called a K-homomorphism of $L \to L'$. Note that a K-homomorphism is always injective and if $[L : K] = [L' : K]$, then it is surjective. Thus if $L = L'$, then such maps are called K-automorphisms of L. The set of all K-automorphisms of L is clearly a group where the group operation defined by composition of maps. This is called the **Galois group of** L/K and is denoted by $Gal(L/K)$.

An element $\alpha \in L$ is said to be **algebraic over** K if it satisfies a nonzero polynomial with coefficients in K. Suppose that $\alpha \in L$ is algebraic over K. Then a nonzero polynomial of least possible degree satisfied by α is clearly irreducible and, moreover, it is unique if we require it to be monic; this monic irreducible polynomial will be denoted by $Irr(\alpha, K)$, and called the **minimal polynomial of** α over K. The extension L/K is said to be algebraic if every $\alpha \in L$ is algebraic over K. If L/K is algebraic, then we call it **separable** if $Irr(\alpha, K)$ has distinct roots (in some extension of K) for every $\alpha \in L$, and we call it **normal** if $Irr(\alpha, K)$ has all its roots in L for every $\alpha \in L$. It may be noted that if L/K is algebraic, then it is normal if and only if any K-homomorphism of L into some extension L' of L maps L onto itself. We call L/K to be a **Galois extension** if it is finite, separable and normal.

We call the **trace map**, denoted by $Tr_{L/K}$ (respectively **norm**, denoted by $\mathcal{N}_{L/K}$) is a K-linear map of $L \to K$ sending $a \in L$ to $Tr_{L/K}(a)$, the trace of the K-linear transformation $L \to L$ mapping $x \in L$ to ax (respectively the determinant of this linear transformation). We recall that the **trace** of a linear transformation of some finite dimensional vector space is equal to the trace of the matrix associated to this linear application, for some fixed basis and only depends on the vector space. If L/K is Galois, for any $a \in L$,

$$Tr_{L/K}(a) = \sum_{\sigma \in Gal(L/K)} \sigma(a) \text{ and } \mathcal{N}_{L/K}(a) = \prod_{\sigma \in Gal(L/K)} \sigma(a).$$

If $K = \mathbb{Q}$ and L is a subfield of $\mathbb{C}$ such that $[L : \mathbb{Q}] = 2$, then it is called a **quadratic field**. In general, a subfield of $\mathbb{C}$ which is of finite degree over $\mathbb{Q}$ is known as an **algebraic number field** or simply, a **number field**. An algebraic element such that $Irr(\alpha, \mathbb{Q})$ belongs to $\mathbb{Z}[X]$ is called an **integer**. The set of the algebraic integers of K form a ring called the **ring of integers of** K, denoted by $\mathcal{O}_K$. Moreover, for any $a \in \mathcal{O}_K$, we can prove $\mathcal{N}_{L/K}(a)$ and $Tr_{L/K}(a)$ are in $\mathbb{Z}$.

2 How to do arithmetic in a number field?

Arithmetic (from the Greek word arithmos "number") is the oldest and most elementary branch of mathematics. It involves the study of quantity, especially as the result of operations that combine numbers. In common usage, it refers to the simpler properties when using the traditional operations of addition, subtraction, multiplication and division with smaller values of numbers.

In order to do arithmetic over a number field K, we need a equivalent of $\mathbb{Z}$. This equivalent could be given by the ring of integers $\mathcal{O}_K$. Indeed, the ring $\mathbb{Z}$ is the simplest possible ring of integers, we have $\mathcal{O}_{\mathbb{Q}} = \mathbb{Z}$.

Nevertheless, one essential property of $\mathbb{Z}$ is missing when we take the ring of integers of a number field: the unicity of the decomposition in irreducible factors. That is, that each element in $\mathbb{Z}$ can be decomposed uniquely as a product of prime numbers (which play the role of irreducible factors in $\mathbb{Z}$). For instance, for $K = \mathbb{Q}(\sqrt{-5})$, we have $\mathcal{O}_K = \mathbb{Z}[\sqrt{-5}]$ and in this ring, $(1+\sqrt{-5})(1-\sqrt{-5}) = 2 \times 3$. But, the algebraic integers have respectively as norm over $\mathbb{Q}$ 6, 6, 4 and 9, then irreducible since $\mathbb{Z}[\sqrt{-5}]$ has no elements with norm 2 or 3.

In order to have a good divisibility theory in K, we have to consider not the integers but objects which extend the divisibility of the algebraic integers and for which we have so the unicity of the decomposition as prime factors. One of the founder results of the algebraic number theory, anticipated by Kummer and proved by Dedekind, is that the ideals can play this role. In particular, the good equivalents are (non-zero) prime ideals of $\mathcal{O}_K$.

One natural question is then to know how the arithmetic of $\mathbb{Q}$ can pass to the one of the number field K. The question make sense, since $\mathbb{Q}$ is then a subring of $\mathcal{O}_K$ and to every integer p of $\mathbb{Z}$ is then naturally associated the ideal $p\mathcal{O}_K$ of $\mathcal{O}_K$. By unicity of the decomposition of ideals in prime factors, it is enough to look the prime ideal decomposition of the ideal $p\mathcal{O}_K$ in K when p is a prime integer. This leads to the notion of ramification.

3 Ramification of prime integers in number fields

Let $K/\mathbb{Q}$ be a number field of degree n.

3.1 Definitions

Let p be a non-zero prime integer of $\mathbb{Z}$. By unicity of the decomposition in prime ideal factors, we can write uniquely:

$$p\mathcal{O}_K = \mathfrak{P}_1^{e_1} ... \mathfrak{P}_r^{e_r}$$

where the $\mathfrak{P}_i$ are distinct non-zero prime ideals of $\mathcal{O}_K$ and the e_i are integers ≥ 1 called the **ramification index**.

We say that prime integer p is **ramified** in K if the ramification index $e_i > 1$ for some i moreover if $e_i \nmid p$ then p is said **tamely ramified**. Otherwise, we say that p is **unramified** in K. The extension $K/\mathbb{Q}$ is said to be **unramified** if every nonzero prime integer of $\mathbb{Z}$ is unramified in K.

We denote moreover f_i, for any i, the degree of the field extension $k(\mathfrak{P}_i)/\mathbb{F}_p$, where $k(\mathfrak{P}_i) = \mathcal{O}_K/\mathfrak{P}_i$ denotes the residual field of the ideal $\mathfrak{P}_i$ and $\mathbb{F}_p$ denotes the finite field

$\mathbb{Z}/p\mathbb{Z}$. We have then the classical property:

$$\sum_{i=1}^{r} e_i f_i = [K : \mathbb{Q}] = n$$

In particular, r can take the values from 1 to n. When $r = 1$, we say that $\mathfrak{p}$ is **not decomposed** and when $r = n$, we say that $\mathfrak{p}$ is **totally decomposed**. Moreover, when the extension $K/\mathbb{Q}$ is Galois then the ramification index e_i and the degree of the residue extensions f_i are all the same and the formula is of the form $ref = n$.

3.2 Discriminant

We call **integral basis of the number field** K a basis $\{\alpha_1, ..., \alpha_n\}$ of K as vector space over $\mathbb{Q}$ such that $\alpha_i \in \mathcal{O}_K$ for $1 \leq i \leq n$. The **absolute discriminant**, denoted by d_K, depending only of the field K is defined to be the determinant of the $n \times n$ matrix

$$(Tr_{K/\mathbb{Q}}(\alpha_i \alpha_j))_{1 \leq i,j \leq n}$$

where $\{\alpha_1, ..., \alpha_n\}$ is an integral basis of K.

We can prove that d_K is an element of $\mathbb{Z}$ and that the prime integers of $\mathbb{Z}$ ramified in K are exactly the ones dividing the discriminant.

3.3 Kummer's Theorem

Kummer's Theorem shows how the decomposition of extended prime ideals can be "read off" from the factorization of a polynomial, for a certain class of rings. It may be observed that the hypothesis of this theorem is satisfied in the case of quadratic. More precisely,

Theorem 3.1 *Let $K/\mathbb{Q}$ be a number field. Suppose that $\mathcal{O}_K = \mathbb{Z}[\alpha]$ for some $\alpha \in \mathcal{O}_K$ and $f(X) = Irr(\alpha, \mathbb{Q})$. Suppose*

$$\bar{f}(X) = \prod_{i=1}^{r} \bar{p_i}(X)^{e_i}$$

is the factorization of $\bar{f}(X)$ (the reduction of $f(X)$ modulo p) into powers of distinct monic irreducible polynomials in $\mathbb{F}_p[X]$. Let $p_i(X)$ be the monic polynomial in $\mathbb{Z}[X]$ whose reduction mod p is $\bar{p_i}(X)$. Then the prime in $\mathcal{O}_K$ lying over p are precisely $\mathfrak{P}_1, ..., \mathfrak{P}_g$, where $\mathfrak{P}_i = p\mathcal{O}_K + p_i(\alpha)\mathcal{O}_K$. Moreover,

$$p\mathcal{O}_K = \prod_{i=1}^{r} \mathfrak{P}_i^{e_i}$$

is the factorization of $p\mathcal{O}_K$ into powers of distinct primes in $\mathcal{O}_K$, the ramification index of $\mathfrak{P}_i$ over p is the above exponent e_i, and the residue degree f_i of $\mathfrak{P}_i$ over p is the degree of the irreducible factor $\bar{p_i}(X)$.

4 Example: study of quadratic Fields.

Let us illustrate the above definitions and results with the study of a particular case. Let $K/\mathbb{Q}$ be a quadratic extension.

4.1 Description of K

Suppose $\alpha \in K$ is any element such that $\alpha \notin \mathbb{Q}$. Then $[\mathbb{Q}(\alpha) : \mathbb{Q}]$ must be > 1 and it must divide $[K : \mathbb{Q}] = 2$. Therefore $K = \mathbb{Q}(\alpha)$ and α satisfies an irreducible quadratic, say $P(X) = X^2 + bX + c$, with coefficients in $\mathbb{Q}$. Using the well known formula for roots of quadratic polynomial, we can replace α by $\sqrt{m}$ so that $K = \mathbb{Q}(\sqrt{m})$, where m is some squarefree element of $\mathbb{Q}$ and $\sqrt{m}$ denotes an element of K whose square is m in particular an element of $\mathcal{O}_K$. With this, we can write $K = \{r + s\sqrt{m} : r, s \in \mathbb{Q}\}$. The other root, say β, of $P(X)$ must satisfy $\alpha + \beta = -b$, and hence it is also in K. So $K/\mathbb{Q}$ is normal. Also clearly $\beta \neq \alpha$ and so $K/\mathbb{Q}$ is separable as well. Thus a quadratic extension is always a Galois extension. The Galois group $Gal(K/\mathbb{Q})$ is a group of order 2, and the non-identity element in it is the automorphism of K which maps α to β and $Gal(K/\mathbb{Q}) = \{Id, \sigma\}$, where Id denotes the identity automorphism of K and σ is the $\mathbb{Q}$-automorphism defined by $\sigma(r + s\sqrt{m}) = r - s\sqrt{m}$.

4.2 Description of the ring of integers $\mathcal{O}_K$

Let K be a quadratic field and $\mathcal{O}_K$ be its ring of integers. As noted before, we have $K = \mathbb{Q}(\sqrt{m})$, where m is a squarefree integer. We now attempt to give a more concrete description of $\mathcal{O}_K$. First, note that $\mathbb{Z}[\sqrt{m}] = \{r + s\sqrt{m} : r, s \in \mathbb{Z}\} \subset \mathcal{O}_K$. Let $x = a + b\sqrt{m} \in \mathcal{O}_K$ for some $a, b \in \mathbb{Q}$. Then $Tr(x) = 2a \in \mathbb{Z}$ and $\mathcal{N}_{K/\mathbb{Q}}(x) = a^2 - mb^2 \in \mathbb{Z}$. Since m is squarefree and $a^2 - mb^2 \in \mathbb{Z}$, we see that $a \in \mathbb{Z}$ if and only if $b \in \mathbb{Z}$. Thus if $a \notin \mathbb{Z}$, then we can find an odd integer a_1 such that $2a = a_1$, and relatively prime integers b_1 and c_1 with $c_1 > 1$ such that $b = b_1/c_1$. Now

$$(a_1 = 2a \in \mathbb{Z} \text{ and } a^2 - mb^2 \in \mathbb{Z}) \Rightarrow (4|c_1^2 a_1^2 \text{ and } c_1^2|4mb_1^2) \Rightarrow c_1 = 2$$

Hence b_1 is odd and $a_1^2 - mb_1^2 \equiv 0(mod\ 4)$. Also a_1 is odd, and therefore, $m \equiv 1(mod\ 4)$.

It follows that if $m \not\equiv 1(mod 4)$, then $a, b \in \mathbb{Z}$, and so in this case,

$$\mathcal{O}_K = \{a + b\sqrt{m} : a, b \in \mathbb{Z}\}$$

and $\{1, \sqrt{m}\}$ is an integral basis.

In the case $m \equiv 1(mod\ 4)$, the preceding observations imply that

$$\mathcal{O}_K \subset \Big\{\frac{a_1 + b_1\sqrt{m}}{2} : a_1, b_1 \text{ are integers having the same parity, i.e., } a_1 \equiv b_1(mod\ 2)\Big\}$$

and, moreover, $(1 + \sqrt{m})/2 \in \mathcal{O}_K$ since it is a root of $X^2 - X - (m-1)/4$; therefore $\mathcal{O}_K = \mathbb{Z}[(1 + \sqrt{m})/2]$ and $\{1, (1 + \sqrt{m})/2\}$ is an integral basis.

4.3 Ramification of prime integers in quadratic fields

Let p be a rational prime. We are interested in the decomposition of the extended ideal $p\mathcal{O}_K$. The formula $\sum_{i=1}^{r} e_i f_i = n$ shows that r as well as e_i, f_i can only be 1 or 2, and that the situation has to be one of the following.

(a) $r = 2$, $e_1 = f_1 = e_2 = f_2 = 1$ so that $p\mathcal{O}_K = \mathfrak{P}_1\mathfrak{P}_2$ for some distinct primes $\mathfrak{P}_1, \mathfrak{P}_2$ of $\mathcal{O}_K$ with $\mathcal{O}_K/\mathfrak{P}_i = \mathbb{Z}/p\mathbb{Z}$. In this case, we say that p is a **decomposed (or split) prime**, or that p **decomposes (or splits) in** $\mathcal{O}_K$.

(b) $r = 1$, $e_1 = 2$, $f_1 = 1$ so that $p\mathcal{O}_K = \mathfrak{P}^2$ for some prime $\mathfrak{P}$ of $\mathcal{O}_K$ with $\mathcal{O}_K/\mathfrak{P} = \mathbb{Z}/p\mathbb{Z}$. In this case p is a **ramified prime**.

(c) $r = 1$, $e_1 = 1$, $f_1 = 2$ so that $p\mathcal{O}_K = \mathfrak{P}$ for some prime $\mathfrak{P}$ of $\mathcal{O}_K$ with $[\mathcal{O}_K/\mathfrak{P} : \mathbb{Z}/p\mathbb{Z}] = 2$. In this case, we say that p is an **inertial prime**.

4.3.1 Ramification of prime integers in $K = \mathbb{Q}(i)$.

Consider now the quadratic field $K = \mathbb{Q}(i)$, where i denotes a square root of -1. We know now that $\mathcal{O}_K$ is the ring $\mathbb{Z}[i]$ of Gaussian integers. This ring of integers is principal and the decomposition in prime factors is unique. So, the study of ramification in this field is not so hard. Indeed, let p be a prime integer if

(a) $p \equiv 1 (mod\ 4)$, then we know (by a classical result of Fermat) that p can be written as a sum of two squares. Thus there exist $a, b \in \mathbb{Z}$ such that $p = a^2+b^2 = (a+bi)(a-bi)$. It can be seen that $(a + bi)$ and $(a - bi)$ are distinct prime ideals in $\mathcal{O}_K$. Thus for the prime ideal $p\mathbb{Z}$, we have $r = 2$, $e_1 = e_2 = 1$ and (since $\sum e_i f_i = 2$) $f_1 = f_2 = 1$. Then p is decomposed.

(b) $p \equiv 3 (mod\ 4)$, it is not difficult to see that p generates a prime ideal in $\mathbb{Z}[i]$ and so for such a prime, we have $r = 1 = e_1$ and $f_1 = 2$. Then p is inertial.

(c) if $p = 2$, we have $2 = (1 + i)(1 - i)$. But $(1 + i)$ and $(1 - i)$ differ only by a unit (namely, $-i$) and thus they generate the same prime ideal. So 2 is a ramified prime and for it, we have $g = 1 = f_1$ and $e_1 = 2$. Then 2 is ramified

4.3.2 Ramification of prime integers in a general quadratic field.

Now let's figure out which primes precisely are ramified, inert or decomposed for a general quadratic field. As noted earlier, we have $K = \mathbb{Q}(\sqrt{m})$, for some uniquely determined squarefree integer m (with $m \neq 0, 1$). Let $\mathcal{O}_K$ be the ring of integers of K. We have also seen that

$$\mathcal{O}_K = \begin{cases} \mathbb{Z}[\sqrt{m}] & \text{if } m \equiv 2, 3 (mod\ 4) \\ \mathbb{Z}[1 + \sqrt{m})/2] & \text{if } m \equiv 1 (mod\ 4) \end{cases}$$

First we consider :

(a) **Case 1:** $m \not\equiv 1(mod\ 4)$, i.e., $m \equiv 2, 3(mod\ 4)$.

In this case, $\mathcal{O}_K = \mathbb{Z}[\sqrt{m}]$ and $f(X) = X^2 - m$ is the minimal polynomial of $\sqrt{m}$ over $\mathbb{Q}$. By Kummer's Theorem, the factorization of $p\mathcal{O}_K$ is determined by the factorization of $\bar{f}(X)$, the reduction of $f(X)$ modulo p.

If $p|m$ respectively $p = 2$, then $\bar{f}(X) = X^2$ respectively $(X-1)^2$, and hence $p\mathcal{O}_K = \mathfrak{P}^2$, with $\mathfrak{P} = (p, \sqrt{m})$ respectively $\mathfrak{P} = (p, 1 - \sqrt{m})$, and p is ramified.

If $p \nmid m$ and $p \neq 2$, then $\bar{f}(X)$ is either irreducible in $(\mathbb{Z}/p\mathbb{Z})[X]$ or has two distinct roots in $\mathbb{Z}/p\mathbb{Z}$. The latter is the case if and only if m is a square mod p, i.e., $m \equiv x^2(mod\ p)$ for some integer x. So we know which primes are decomposed and which are inertial. The result can be conveniently expressed using the Legendre symbol, which is defined thus.

$$\left(\frac{m}{p}\right) = \begin{cases} 1 & \text{if } p \nmid m \text{ and } m \text{ is a square mod } p \\ -1 & \text{if } p \nmid m \text{ and } m \text{ is not a square mod } p \\ 0 & \text{if } p|m. \end{cases}$$

What we have shown so far is that if $m \equiv 2, 3(mod\ 4)$, then

$$p \text{ is } \begin{cases} \text{decomposed} & \text{if } p \neq 2 \text{ and } \left(\frac{m}{p}\right) = 1 \\ \text{ramified} & \text{if } p = 2 \text{ or } \left(\frac{m}{p}\right) = 0 \\ \text{inertial} & \text{if } p \neq 2 \text{ and } \left(\frac{m}{p}\right) = -1. \end{cases}$$

(b) **Case 2:** $m \equiv 1(mod\ 4)$.

In this case, $\mathcal{O}_K = \mathbb{Z}[(1 + \sqrt{m})/2]$ and $f(X) = X^2 - X - (m-1)/4$ is the minimal polynomial of $(1 + \sqrt{m})/2$ over $\mathbb{Q}$.

If $p = 2$, then $\bar{f}(X)$ has a root mod p if and only if $(m-1)/4 \equiv 0(mod\ 2)$, i.e., $m \equiv 1(mod\ 8)$ (because $x^2 - x = x(x-1) \equiv 0(mod\ 2)$ for any $x \in \mathbb{Z}$), and in this case, each of the two distinct elements in $\mathbb{Z}/2\mathbb{Z}$ is a root of $\bar{f}(X)$, which implies that 2 is a decomposed prime.

If $p = 2$ and $m \neq 1(mod\ 8)$, then $\bar{f}(X)$ has to be irreducible in $(\mathbb{Z}/2\mathbb{Z})[X]$, and so 2 is an inertial prime.

Now assume that $p \neq 2$. Then the "roots" $(1 \pm \sqrt{m})/2$ of $X^2 - X - (m-1)/4$ will exist in $\mathbb{Z}/p\mathbb{Z}$ if and only if $\sqrt{m}$ exists in $\mathbb{Z}/p\mathbb{Z}$, or equivalently, m is a square mod p. Moreover, $\bar{f}(X)$ has multiple roots in $\mathbb{Z}/p\mathbb{Z}$ if and only if $p|m$. Thus, by Kummer's Theorem, we find that p is ramified if and only if $p|m$, and if $p \neq 2$ and $p \nmid m$, then p is decomposed or inertial according as m is or is not a square mod p.

So if $m \equiv 1(mod\ 4)$, then

$$p \text{ is } \begin{cases} \text{decomposed} & \text{if } p \equiv 2 \text{ and } m \equiv 1(mod\ 8) \text{ or if } p \neq 2 \text{ and } \left(\frac{m}{p}\right) = 1 \\ \text{ramified} & \text{if } p|m, \text{ i.e., } \left(\frac{m}{p}\right) = 0 \\ \text{inertial} & \text{if } p = 2 \text{ and } m \not\equiv 1(mod\ 8) \text{ or if } p \neq 2 \text{ and } \left(\frac{m}{p}\right) = -1 \end{cases}$$

4.4 Computing the discriminant d_K

We can now compute the discriminant of K as follows.

$$d_K = \begin{cases} Det \begin{pmatrix} 2 & 0 \\ 0 & 2m \end{pmatrix} & = 4m \quad if\ m \equiv 2, 3 (mod\ 4) \\ Det \begin{pmatrix} 2 & 1 \\ 1 & (1+m)/2 \end{pmatrix} & = m \quad if\ m \equiv 1 (mod\ 4) \end{cases}$$

It may be remarked that the integer d_K determines the quadratic field K completely. This prove that in this particular following case, the result mentioned before saying that p is a ramified prime in K if and only if $p|d_K$.

References

[Lan94] S. Lang, "Algebraic number theory". Graduate Texts in Mathematics, vol. 110. Springer-Verlag, New York, second edition, 1994.

[Neu99] J. Neukirch, "Algebraic number theory". Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 332. Springer-Verlag, Berlin, 1999. Translated from the 1992 German original and with a note by Norbert Schappacher, with a foreword by G. Harder.

[Sam67] P. Samuel, "Théorie algébrique des nombres". Hermann, Paris, 1967.