Università di Padova – Dipartimento di Matematica Pura ed Applicata

Scuole di Dottorato in Matematica Pura e Matematica Computazionale

Seminario Dottorato 2008/09

Preface

This document offers a large overview of the nine months' schedule of Seminario Dottorato 2008/09. Our "Seminario Dottorato", which is held in a stable framework since three years, is a double-aimed activity. At one hand, the speakers (usually young Ph.D. students or post-docs) 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.
At the end of this year's activity, we are happy to remark once more that this philosophy has been generally understood by the speakers, who also nicely agreed to write down these notes to leave a concrete footstep of their participation: we thank them all warmly.

Padova, 25 June 2009

Corrado Marastoni, Tiziano Vargiolu, Matteo Dalla Riva

Abstracts (from Seminario Dottorato's web page)

Wednesday 15 October 2008

The Grothendieck fundamental groups: a basic introduction

Nicola MAZZARI (Ph.D. in Pure Math., Dip. Mat. "F. Enriques", Milano)

We give a basic introduction to the Grothendieck fundamental group. In topology there are (at least) two ways to define the fundamental group $\pi_1(S, s)$ of a topological space S. Namely we can view it as the set of loops based on a point s up to homotopy, or as the group of automorphism of the universal cover. Only this second approach can be made algebraic and allows to define the Grothendieck (or etale) fundamental group $\pi_1^{\text{et}}(S, s)$ of a scheme S with respect to a geometric point s. We will consider only affine schemes and we don't assume the reader familiar with algebraic geometry.

Wednesday 29 October 2008

Information Flow on Trees: the Reconstruction Problem and the Purity of the Free Gibbs Measure

Marco FORMENTIN (Ph.D. in Applied Math., Dip. Mat.)

The Reconstruction Problem on a tree can be stated as follows. We send a signal from the root to the boundary, making a prescribed error at every edge of the tree. Suppose you know what happens at distance N from the origin of the tree. What can you say about the original signal sent from the root when N goes to infinity? This problem, concerning the flow of information on trees is equivalent to the purity of the free Gibbs Measure for the Ising/Potts models on a tree. Purity can be regarded as a special kind of phase transition. We review this equivalence and give old and new thresholds for the transition to purity for the free Potts Gibbs measure on regular trees.

Wednesday 12 November 2008

Synchronization and homomorphisms

Pablo SPIGA (Ph.D. in Pure Math., Dip. Mat.)

An automaton is a machine which can be in any of a set of internal states which cannot be directly observed. A synchronizing automaton is an automaton admitting a sequence of transitions which take the automaton from any state into a known state. In this talk we present some recent connections between synchronizing automatons, permutation groups and graph homomorphisms. All relevant definitions would be given during the talk.

Wednesday 26 November 2008

On matrices with the Edmonds-Johnson property

Alberto DEL PIA (Ph.D. in Applied Math., Dip. Mat.)

Integer programming is the problem of optimizing a linear function over the integral points in a polyhedron P, expressed as a system of linear inequalities. It is known that it is equivalent to optimizing such linear function over the polyhedron P_I, that is the convex hull of the integral points in P. One of the main problems in mathematical programming is to developed tools to get P_I, and one of such methods is the Chvatal-Gomory procedure. This procedure, starting from P, gives a sequence of smaller polyhedra, P', P'', ..., that converges to P_I in a finite number of iterations. The number of iterations needed to get P_I gives an order of complexity to the problems. We survey some old and new results of classes of problems in which $P' = P_I$.

Wednesday 10 December 2008

Some problems from Geometric Measure Theory: Plateau, Bernstein, Dido

Davide VITTONE (researcher in Pure Math., Dip. Mat.)

The aim of Geometric Measure Theory (GMT) is to approach geometric problems by means of measure-theoretical tools. After a brief presentation of some GMT definitions of "surface measure", we will summarize the history and main results about three classical questions: minimal surfaces (or Plateau problem), the Bernstein problem and the isoperimetric (or Dido) problem.

Wednesday 14 January 2009

(De)-Localization of some (1+1)-dimensional models

Martin BORECKI (Ph.D. in Applied Math., Technische Univ., Berlin)

We consider a (1+1)-dimensional model, i.e. a directed model for a linear chain. The chain is randomly distributed in space and undergoes an interaction with the environment and itself. Thus, it can be seen as a random polymer and we want to study its spatial distribution as a function of its length and its interaction parameters. The self-interaction consists of a Gradient and Laplacian mixture type, whereas the interaction with the environment is reduced to a delta-pinning, i.e. the chain gets a reward each time it touches the x-axis. We discuss the localization behaviour of the model, which displays remarkable differences (phase transitions) as the parameters of the interaction vary. Furthermore we consider what changes, if we additionally introduce an impermeable wall. Motivation, introduction and explanations will hopefully make the talk accessible to a large audience.

Wednesday 28 January 2009

An introduction to p-adic analysis

Valentina DI PROIETTO (Ph.D. in Pure Math., Dip. Mat.)

The p-adic numbers were discovered by Hensel at the end of nineteenth century and in the last century they came to a central role in number theory. In the first part of this talk we shall give the definition of the field of p-adic numbers and present some results in elementary p-adic analysis always comparing with results on classical analysis over the real numbers; in the second part we shall describe an example which explains a basic case of the result proven in our Ph.D thesis.

Wednesday 11 February 2009

Solving Mixed Integer Programs with Gomory Cutting Planes

Arrigo ZANETTE (Ph.D. in Applied Math., Dip. Mat.)

Gomory cutting planes were first introduced by Gomory in 1958 to solve Integer and Mixed Integer ('60) Programs (MIP). However they were soon abandoned in favor of enumeration tecniques, until, in 1996, they were revisited by Balas et al., becoming a fundamental tool for commercial MIP solver. Despite their long history and relative success, the lack of understading on their practical behaviour makes Gomory Cutting Planes an interesting research topic. In particular it is clear that they might perform much better than current implementations have managed to do, but nobody has found the right way of using them yet. In the talk we will review Gomory cutting planes, their typical usage in commercial MIP solvers and recent research findings that might eventually lead to a new performance breakthrough.

Wednesday 25 February 2009

An invitation to Frobenius manifolds

Luca Philippe MERTENS (Ph.D. in Pure Math., SISSA Trieste)

Frobenius manifolds are geometric structures encoding the dispersionless limit of a bihamiltonian integrable hierarchy. They were introduced by B. Dubrovin to study the remarkable connection between 1+1 integrable systems, 2D topological field theories and Gromov-Witten invariants of symplectic manifolds discovered by E. Witten and M. Kontsevich for the case of the KdV hierarchy. The first part of the talk will be an introduction to key ideas and definitions. We will define an appropriate class of integrable hierarchies and we will show how one can associate a Frobenius manifold to them, and vice versa. We will present in detail the example of the Toda Hierarchy, encoding the Gromov-Witten Invariants of the complex projective line. The second part of the talk will focus on recent developments of the theory related to 2+1 integrable systems. This class

of hierarchies naturally lead to the notion of infinite dimensional Frobenius manifold. We will present the case of the 2D Toda hierarchy, highlighting main differences with respect to the finite dimensional case and pointing out future applications of the theory.

Wednesday 11 March 2009

The Dynamics of a Spin-Flip System by an Example: the Curie-Weiss Model

Francesca COLLET (Ph.D. in Applied Math., Dip. Mat.)

A spin system is a system composed by N sites at which is associated randomly a +1 or -1 value, called spin. Each spin is influenced by all the others in the same way and this makes it flip with a certain probability. The dynamics just mentioned are completely described by the time evolution of the Magnetization (the sum of all the spin values divided by N), so it is sufficient to study the behavior of this last quantity. As N grows to infinity, its limiting dynamics are deterministic (driven by ODE) and exhibit a phase transition: multiple equilibrium solutions arise depending on the value of a parameter, which is the inverse of the temperature. The Curie-Weiss model is a basic example of it. After having recalled some notions of Probability, we try to explain in a simple and intuitive way, avoiding the most part of the technicalities, how the Curie-Weiss model evolves in time at different temperatures.

Wednesday 25 March 2009

Computing with Affine Algebraic Groups

Andrea PAVAN (Ph.D. in Pure Math., Dip. Mat.)

How can one solve the Rubik's Cube? The question turns out to be equivalent to a problem about groups, whose solution is provided by Computational Group Theory. More generally, CGT is concerned with designing and analyzing algorithms to compute information about groups which can be described by a finite amount of data. Examples include finite permutation groups, finitely presented groups, finitely generated matrix groups and polycyclic groups, which have been at the center of the subject since the beginning of the last century. On the contrary, very little work has been done on affine algebraic groups. These are, roughly speaking, groups whose elements are solutions to some system of polynomial equations in finitely many indeterminates. Although their structure is well understood, they have been rarely studied from a computational point of view. Two pioneers in the field are Grunewald and Segal, who developed the basis for many useful algorithms. In the first part of the talk we will give an introductory overview of both Computational Group Theory and the theory of affine algebraic groups. Then we will describe the work of Grunewald and Segal, as well as some improvements of their methods.

Wednesday 8 April 2009

Constraint Programming Techniques for Mixed Integer Linear Programs

Domenico SALVAGNIN (Ph.D. in Applied Math., Dip. Mat.)

Two paradigms in the field of optimization have reached a high degree of sophistication from the point of view of both theory and implementation: Constraint Programmming (CP) and Mixed Integer Programming (MIP). The CP and MIP paradigms have strengths and weaknesses that complement each other: thus an integration of the two has the potential to yield important benefits. In this talk I will provide a brief introduction of the two paradigms and present two cases of application of CP techniques, namely nogoods and propagation, to enhance MIP resolution algorithms, namely dominance detection and primal heuristics.

Thursday 23 April 2009

A primer in Arakelov geometry

Vincent MAILLOT (professor at CNRS - Paris VI)

1. In the first part of my talk, I'll introduce the basic notions and problems involved in the early developpements of Arakelov geometry.
2. In the second part, I'll give a more formal and systematic introduction to the subject. Time permitting, I'll explain the statement of the arithmetic Riemann-Roch theorem and I'll give a recent application to number theory.

Wednesday 13 May 2009

A global approach to multiobjective optimization

Alberto LOVISON (Univ. Padova, Dip. Mat.)

In real life situations, there are usually more than one objective to deal with in order to design a successful project. For instance, a good car should be fast, while being low consuming as possible, should protect occupants while keeping compact external dimensions, should be comfortable and maximize, or minimize, at the same time, many other performance indicators. Multiobjective optimization defines the mathematical framework for dealing with such problems. In this talk we propose a gentle introduction to this subject, strictly related to the standard single objective approach based on the study of critical points. We will recall the concept of Pareto critical set, introduced by Stephen Smale, and illustrate an effective algorithm for the global search of these critical sets. This topological approach allows the definition of a Morse theory for vector functions. On the other hand, severe restrictions derive from the curse of dimensionality and from the existence of structurally unstable singularities in higher dimensions.

Wednesday 20 May 2009

Introduction to Moduli Spaces

Ernesto MISTRETTA (Univ. Padova, Dip. Mat.)

We will explain the meaning of moduli spaces as spaces parametrizing geometrical objects, giving some well known examaples as Grassmannians and Projective Spaces. We will focus our attention on the moduli space of triangles, constructing it and elucidating problems of symmetries and monodromy, that appear in more sophisticated cases. Time permitting we will illustrate open problems and recent progress in the theory of moduli spaces of curves.

Wednesday 3 June 2009

Social Interactions and heterogeneous agent models. Applications to Economics and Finance

Marco TOLOTTI (Università di Venezia)

Relying on my work in the field of contagion models, based on interacting particle systems, I will discuss some open issues concerning the applicability of complex systems in Economics and Finance. I will present some applications of a class of Markov models that are in line with recent research in Economic Theory. In particular I will highlight the importance of modeling social interactions, bounded rationality, heterogeneous agents and random utilities. [Keywords: heterogeneous agent models, intensity-based models, mean field interactions, non reversible Markov processes, phase transition, random utilities, social interactions, stochastic population processes, strategic complementarities.]

Wednesday 17 June 2009

Finite p-groups

Eleonora CRESTANI (Ph.D. in Pure Math., Dip. Mat.)

A p-group is a group in which every element has order a power of p (where p is a prime). The first part of the seminar is an introduction to this area, and I will give same examples that, in particular, show why p-groups play such an important role in the theory of finite groups. In the second part of the seminar, some typical problems that arise in finite p-groups theory are presented, with a particular attention to the ones that I studied during my Ph.D. .

A note on Grothendieck fundamental group

Nicola Mazzari (*)

Abstract. We give a basic introduction to the Grothendieck fundamental group. We consider only affine schemes and we don't assume the reader familiar with algebraic geometry.

Introduction

Very roughly speaking we can say that algebraic/arithmetic geometry is the study of systems of polynomial equations. Let k be a commutative ring with unit and $f_1, ..., f_m \in k[T_1, ..., T_n]$ be polynomials of n unknowns with coefficient in k. We are interested to study the "solutions" of the system

$$\Sigma = \begin{cases} f_1(T_1, ..., T_n) & = 0 \\ \quad \vdots & \quad \vdots \\ f_m(T_1, ..., T_n) & = 0 \end{cases}$$

with values in a k-algebra R, i.e. $Z_\Sigma(R) := \{a = (a_1, ..., a_n) \in R^n \mid f_i(a_1, ..., a_m) = 0 \ \forall \ i\}$. Two well known particular cases are:

(a) if $R = k$ is a field and $n = m = 1$ we are in the setting of Galois theory.

(b) if $R = k$ is a field and the f_i are linear we are doing linear algebra.

Another important example is the case $k = \mathbb{C}$, then the set $Z_\Sigma(\mathbb{C})$ can be viewed as a closed sub-set of $\mathbb{C}^n$ with respect to the standard topology. Hence we can use topological methods in order to classify these zero-sets.

In topology there are (at least) two ways to define the fundamental group $\pi_1(S, s)$ of a topological space S. Namely we can view it as the set of loops based on a point s up to homotopy, or as the group of automorphism of the universal cover. This second approach can be made algebraic (i.e. it works for any ring k), as we are going to explain later, and allows to define the Grothendieck (or étale) fundamental group π_1^{et} of a system Σ (or more generally any noetherian connected scheme). This is a pro-finite group and it is

(*)Ph.D. course, Università di Milano, Dip. Mat. "F. Enriques". E-mail: nclmzzr@gmail.com. Seminar held on 15 October 2008.

isomorphic to the pro-finite completion of $\pi_1(Z_\Sigma(\mathbb{C}), s)$ when $k = \mathbb{C}$. This follows by some important differences between the topological and the algebraic setting.

In fact in the algebraic setting one cannot use the topological definition of a cover. This lead to the notion of étale morphism which is purely algebraic and corresponds to local isomorphisms when we are dealing with complex schemes.
There is also another problem to overcome. The universal étale cover rarely exists. The solution is to consider finite approximations. It follows that the étale fundamental group is a projective limit of finite groups.

1 Affine schemes

Let $f_1, ..., f_m \in k[T_1, ..., T_n]$ and consider the system of equations

$$\Sigma = \begin{cases} f_1(T_1, ..., T_n) & = 0 \\ \vdots & \vdots \\ f_m(T_1, ..., T_n) & = 0 \end{cases}$$

Let $Z_\Sigma(R) := \{a = (a_1, ..., a_n) \in R^n \mid f_i(a_1, ..., a_m) = 0 \ \forall \ i\}$ be the set of solutions of the system with values in R.

In order to have a better understanding of $Z_\Sigma(R)$ for any R we will use commutative algebra and study the ring $A = k[T_1, ..., T_n]/I$ associated to the system Σ, where $I = (f_1, ..., f_m)$. This is justified by the following

$$\begin{aligned} \mathrm{Hom}_{\mathsf{Alg}_k}(A, R) &= \mathrm{Hom}_{\mathsf{Alg}_k}\left(\frac{k[T_1, ..., T_n]}{(f_1, ..., f_m)}, R\right) \\ &= \{\phi : k[T_1, ..., T_n] \to R \mid \phi(f_i) = 0 \ \forall \ i\} \\ &= \{(a_1 = \phi(T_1), ..., a_n = \phi(T_n)) \mid f_i(a_1, ..., a_n) = 0 \ \forall \ i\} \\ &= Z_\Sigma(R) \ . \end{aligned}$$

Note that given a ring morphism $f : A \to B$ we have a natural transformation $f^* : \mathrm{Hom}_{\mathsf{Alg}_k}(B, -) \to \mathrm{Hom}_{\mathsf{Alg}_k}(A, -)$. Explicitly for any ring R we have $f^*_R : \mathrm{Hom}_{\mathsf{Alg}_k}(B, R) \to \mathrm{Hom}_{\mathsf{Alg}_k}(A, R)$, $f^*_R(g) = g \circ f$.

Examples 1.1 (i) Let k be a field. Then for any k-algebra R, $\mathrm{Hom}_{\mathsf{Alg}_k}(k, R) = \{pt\}$. In fact there is only one k-linear morphism sending $1 \mapsto 1$.

(ii) Let $A = k[T_1, ..., T_n]$ and R be an k-algebra. Then $\mathrm{Hom}(k[T_1, ..., T_n], R) = R^n$. Hence this ring corresponds to the n-dimensional affine space over k.

(iii) Let $A = k[T, T^{-1}]$ be the ring of Laurent polynomials. Then $\mathrm{Hom}_{\mathsf{Alg}_k}(A, R) = R^*$ is the set of invertible elements in R. This can be thought as the affine line (over k) without the origin.

(iv) Let A be a $\mathbb{C}$-algebra of the following type $\mathbb{C}[T_1, ..., T_n]/(f_1, ..., f_m)$ then the set of $\mathbb{C}$-points (or solutions over $\mathbb{C}$) $\mathrm{Hom}_{\mathsf{Alg}_k}(A, \mathbb{C})$ is naturally a closed subset of $\mathbb{C}^n$ w.r.t. the standard topology.

2 The topological fundamental group

Fix a (connected) topological space S. We assume that all the topological spaces in this section are path-connected and locally simply connected. These assumptions on S imply the existence of the universal cover.

Definition 2.1 *We say that a continuos and surjective map $\phi : X \to S$ is a* cover *of S (or a* covering space*) if for any point $s \in S$ there exists U, path-connected open neighborhood of s, such that the restriction*

$$\phi|_V : V \longrightarrow U$$

is an homeomorphism for any V path-connected component of $\phi^{-1}(U) \subset X$.

Examples 2.2 (i) Let $S = \mathbb{C}^* = \mathbb{C} \setminus 0$. Then the exponential map $\exp : \mathbb{C} \to \mathbb{C}^*$, $\exp(x) := \sum_n x^n/n!$, is a cover of $\mathbb{C}^*$. Note that $\#\exp^{-1}(s) = \infty$ for any $s \in \mathbb{C}^*$.
(ii) Let $S = \mathbb{C}^*$ and fix $n \in \mathbb{Z}$. Then we can define the following cover of S $\mu_n : \mathbb{C}^* \to \mathbb{C}^*$, $\phi(x) = x^n$. This cover is finite of degree n: i.e. $\#\mu_n^{-1}(s) = n$ for any $s \in \mathbb{C}^*$.

We define the category Cov_S of covers of S in the following way: an object of Cov_S is a cover $\phi : X \to S$; a morphism f from $\phi : X \to S$ to $\psi : Y \to S$ is a continuous map $f : X \to Y$ such that $\psi \circ f = \phi$.

Remark 2.3 Fix a cover $\phi : X \to S$ and a point $s \in S$.
(i) According to the previous definition the group $\mathrm{Aut}_{\mathsf{Cov}_S}(\phi)$ is the set of homeomorphisms $f : X \to X$ such that $\phi \circ f = \phi$. This group acts on the fiber $\phi^{-1}(s)$ by

$$\mathrm{Aut}_{\mathsf{Cov}_S}(\phi) \times \phi^{-1}(s) \to \phi^{-1}(s) \quad (f, x) \mapsto f(x)$$

(ii) The previous action is transitive if and only if $\phi_*(\pi_1(X, x)) \subset \pi_1(S, s)$ is a normal sub-group. In this case ϕ is called a Galois cover.
(iii) Also the fundamental group $\pi_1(S, s)$ acts on $\phi^{-1}(s)$ by lifting paths

$$\pi_1(S, s) \times \phi^{-1}(s) \to \phi^{-1}(s) \quad ([\gamma], x) \mapsto \gamma_x(x)$$

where γ_x is a lifting of γ starting from x.
The stabilizer of $x \in \phi^{-1}(s)$ is isomorphic to $\phi_*(\pi_1(X, x))$.

Theorem 2.4 *The functor $F : \mathsf{Cov}_S \to \mathsf{Set}$, $F(X \xrightarrow{\phi} S) := \phi^{-1}(s)$ is representable: i.e. there exists a (universal) cover $\phi^u : X^u \to S$ such that*

$$\mathrm{Hom}_{\mathsf{Cov}_S}(\phi^u, -) \cong F(-)$$

Moreover F induces an equivalence between the category Cov_S and the category of $\pi_1(S, s)$-sets.

Remark 2.5 (i) The universal cover can be characterized by the fact that X is simply connected. The existence of such a cover depends on the topological assumptions given at the beginning of the section.

(ii) An immediate corollary of the theorem is the canonical isomorphism

$$\pi_1(S, s) \cong \mathrm{Aut}_{\mathsf{Cov}(S)}(\phi^u)$$

hence the topological fundamental group can be defined via covers.

Example 2.6 In the case $S = \mathbb{C}^*$ the exponential map is the universal cover. An automorphism of the cover $\exp : \mathbb{C} \to \mathbb{C}^*$ is of the form $x \mapsto x + 2\pi i n$ for some $n \in \mathbb{Z}$. Hence $\pi_1(\mathbb{C}^*, s_0) \cong 2\pi i\mathbb{Z}$.

3 Grothendieck fundamental group

3.1 Étale covers

Definition 3.1 *Let A be a ring (noetherian, commutative, with unit). A ring morphism $f : A \to B$ is étale if it is of finite type, flat and unramified. If it also finite we call it an étale cover.*

This means that an étale cover of A is an A-algebra B which is a finitely generated A-module and for any $\mathfrak{m} \subset A$ maximal ideal the fiber $B \otimes A/\mathfrak{m}$ is a separable algebra over the field $A/\mathfrak{m}$. Recall that a separable algebra over a field k is isomorphic to a direct sum of finite separable extensions of k.
On the geometric side we have a map of functors $\phi = f^* : \mathrm{Hom}_{\mathsf{Alg}}(B, -) \to \mathrm{Hom}_{\mathsf{Alg}}(A, -)$ such that for any element (or k-point) $s \in \mathrm{Hom}_{\mathsf{Alg}}(A, k)$ the fiber $\phi^{-1}(s) \subset \mathrm{Hom}_{\mathsf{Alg}}(B, k)$ is a finite set of distinct points as in the topological case.

The basic example of an étale cover is the following.

Example 3.2 Let $A = k[T]$ where k is an algebraically closed field. Consider $B = k[T, S]/(f)$ and $\phi : A \to B$ defined by $\phi(T) = T \mod (f)$. Then ϕ is an étale cover of A if and only if for any $(a, b) \in \mathrm{Hom}_{\mathsf{Alg}}(B, k)$ we have $(\partial f/\partial S)(a, b) \neq 0$. In particular it is easy to check that we have an étale cover for $f = S - T^2$ (this is in fact an isomorphism), while for $f = S^2 - T$ there is a pathological point in $(0, 0)$.

A key point for the construction of the étale fundamental group is the existence of Galois covers. Let A be a ring and $f : A \to B$ an étale cover, it is a Galois cover there exists a finite group G acting (on the left) faithfully on B such that

(a) The rings of G-invariants $B^G = \{g \in G | gb = b\}$ is isomorphic to A.

(b) $f : A \to B$ is the canonical inclusion $B^G \to B$.

In this case the group of automorphism of the cover is G. If we start with G acting on B than the canonical map $B^G \to B$ is an étale cover if all the inertia groups of the action are trivial.

Example 3.3 Let $G = \mathbb{Z}/n\mathbb{Z}$ and $B = k[T, T^{-1}, S]/(S^n - T)$. Consider the following action: $[m] \cdot T = T$, $[m] \cdot S = S^m$, for $m \in \mathbb{Z}$. Then it easy to check that $B^G = k[T, T^{-1}]$ and that the canonical map $\iota : k[T, T^{-1}] \to k[T, T^{-1}, S]/(S^n - T)$ is étale. Moreover note that there is another way to write the same cover, namely

$$\mu_n : k[T, T^{-1}] \to k[T, T^{-1}] \quad \mu_n(T) = T^n$$

in fact it is easy to check that there is a ring ismorphism $\theta : k[T, T^{-1}, S]/(S^n - T) \to k[T, T^{-1}]$ such that $\theta \circ \iota = \mu_n$.

3.2 The main result

Let A be a ring (e.g. $A = k[T_1, ..., T_n]/I$) and fix $s \in \mathrm{Hom}_{\mathsf{Alg}}(A, \Omega)$ where Ω is an algebraically closed field. For any étale $f : A \to B$ cover we can consider the set of points of $\mathrm{Hom}_{\mathsf{Alg}}(B, \Omega)$ lying over s, i.e.

$$F(B) := \{t : B \to \Omega \mid t \circ f = s\}$$

This association induces a functor from the category of étale covers of A to the category of sets. Now we are ready to state the main result of these notes.

Theorem 3.4 *There exists a projective limit of Galois cover $(f_\alpha : A \to B_\alpha)$ such that for any étale cover $f : A \to B$*

$$F(B) \cong \operatorname*{colim}_\alpha \mathrm{Hom}_{\mathsf{Alg}_A}(B, B_\alpha)$$

(i.e. F is a pro-representable functor). Moreover we can define the pro-finite group $\pi_1^{\mathrm{et}}(A, s) := \lim_\alpha \mathrm{Aut}(B_\alpha) = \lim_\alpha F(B_\alpha)$ and F induces an equivalence of categories

$$F : \{\textit{étale covers of } A\} \to \{\pi_1^{\mathrm{et}}(A, s)\textit{-sets}\} \ .$$

Proof. See [sga1, Exp. V, §5 and 7]. □

Example 3.5 We already know that the (topological) fundamental group of $\mathbb{C}^*$ is isomorphic to $\mathbb{Z}$. In the algebraic setting $\mathbb{C}^*$ corresponds to the ring $A = k[T, T^{-1}]$. The following morphism

$$(-)^n : k[T, T^{-1}] \to B_n = k[T, T^{-1}] \quad T \mapsto T^n$$

is an étale cover, moreover it is a Galois cover with group $\mathbb{Z}/n\mathbb{Z}$. Hence if we take the projective limit $\lim_n \mathrm{Aut}(B_n) = \lim_n \mathbb{Z}/n\mathbb{Z} = \widehat{\mathbb{Z}}$ we obtain the pro-finite completion of $\mathbb{Z}$.

Example 3.6 Let $A = \mathbb{R}$, then it is easy to check that $\mathbb{R} \to \mathbb{C}$ is an étale cover with group $\mathbb{Z}/2\mathbb{Z}$ (the action is the complex conjugation). Moreover this the universal étale cover of $\mathbb{R}$ and $\pi_1^{\mathrm{et}}(\mathbb{R}, \mathbb{C}) = \mathrm{Aut}(\mathbb{R} \to \mathbb{C}) = \mathbb{Z}/2\mathbb{Z}$. We remark that this is a very special

situation. In fact we can hope to find the universal étale cover only in case π_1^{et} is a finite group.

All this theory can be generalized to the case of locally noetherian schemes. Unfortunately we can compute the étale fundamental group only in few cases, even for affine schemes. As final remarks we mention the following important results

(a) (Comparison with the topological fundamental group) Let A be a $\mathbb{C}$-algebra of finite type (i.e. $A = \mathbb{C}[T_1, ..., T_n]/I$) and let S be the set of $\mathbb{C}$-points $\text{Hom}_{\mathsf{Alg}}(A, \mathbb{C})$ endowed with the standard topology. Then there is a canonical isomorphism of pro-finite groups

$$\widehat{\pi}_1(S, s) \cong \pi_1^{\text{et}}(A, s)$$

where $\widehat{\pi}_1(S, s)$ is the completion of the (topological) fundamental group $\pi_1(S, s)$ w.r.t. the topology of finite index sub-groups. (See [sga1, XII.5.2]).

(b) (Galois theory) If k is a field and $\bar{k}$ is its algebraic closure, then $\pi_1^{\text{et}}(k, \bar{k}) = \text{Gal}(\bar{k}/k)$ is the absolute Galois group of k.

(c) If A is an algebra over a field k, then there is an exact sequence

$$1 \to \pi_1^{\text{et}}(A \otimes_k \bar{k}, \bar{k}) \to \pi_1^{\text{et}}(A, \bar{k}) \to \text{Gal}(\bar{k}/k) \to 1$$

Thus in general the étale fundamental group carries both a geometric and an arithmetic information.

References

[sga1] *Revêtements étales et groupe fondamental*. Séminaire de Géométrie Algébrique du Bois Marie 1960–1961 (SGA 1), Dirigé par Alexandre Grothendieck. Augmenté de deux exposés de M. Raynaud, Lecture Notes in Mathematics, Vol. 224.

Information Flow on Trees: the Reconstruction Problem and the Purity of the Free Gibbs Measure

Marco Formentin (*)

Abstract. The Reconstruction Problem on a tree can be stated as follows. We send a signal from the root to the boundary, making a prescribed error at every edge of the tree. Suppose you know what happens at distance N from the origin of the tree. What can you say about the original signal sent from the root when N goes to infinity? This problem, concerning the flow of information on trees is equivalent to the purity of the free Gibbs Measure for the Ising/Potts models on a tree. Purity can be regarded as a special kind of phase transition. We review this equivalence and give old and new thresholds for the transition to purity for the free Gibbs measure on regular trees.

1 Introduction

Graphs occur in many real-world communication networks. A graph (or a network) is a set of points (called vertices) with links (called edges) between them; Internet with web-pages and links between web-pages is one of the most notable examples: the vertices of the graph are the web-pages and a directed edge between two of them means you can click from one to another.
Other examples are the business relationships between companies, the networks of physical interactions between proteins and the relation predator-prey between species within an ecosystem; these are only few among all possible examples [1, 2, 4].

Many of such networks share the same properties. Many are small worlds, in the sense that most vertices are separated by a relatively short chain of edges. Moreover, many networks are "scale" free, i.e. the number of edges per vertex (i.e. the degree of a vertex) obeys a power-law, which means that the number of vertices with degree k is proportional to an inverse power of k.

(*)Dottorato di ricerca in Matematica Pura, Dipartimento di Matematica Pura ed Applicata, Università di Padova. E-mail: formen@math.unipd.it . Seminar held on 29 October 2008.

Processes on networks are common, too. For some real-word examples one could consider the spread of an opinion (or a disease) among a social network, or the flow of a signal on an information network. The structure of the network affects its performance. For instance, the topology of social networks affects the spread of information and disease.
A mathematical model of this kind of phenomena cannot be deterministic because there is something we cannot control and that we have to assume to be random. First, real networks are very large and they can grow too. As a result, their complete description is impossible, and researchers, both in the applications and in mathematics, have turned to their local description: how many vertices do they have, and by which rules are vertices connected to one another? These rules are usually local and probabilistic, which leads us to consider *Random Graphs*; this means that the number of vertices, the number of edges per vertex are not fixed and could be random variables. The graph could also grow in size following a random law adding vertices and edges. Turning back to the spread of a disease and to the flow of an information on a network, this corresponds to uncertainty about the number of people one knows, new persons one could meet, the numbers of channels transmitting the same information and so on. From this point of view, a Random Graph, with given distribution, is a natural starting point to study this kind of phenomena. Moreover, one also has to consider that contact between persons does not mean contagion for sure: there will be a probability that it happens. In the same way, for the information flow, one has to include in the model the possibility of a wrong transmission of the signal. This leads us to study *Stochastic Processes on Random Graphs*.

In the recent years a great interest grew in such models in applied sciences and they were used in Physics, Economy, Computer Sciences and Biology to describe a great variety of phenomena [1, 2, 3, 4]. This interest still continues nowadays giving to theoretical sciences much work to do and open problems. These brief notes focus on one of them concerning the flow of information on trees[(*)]: *The Reconstruction Problem*. Even if inspired by applications it has connection with mathematical and theoretical Physics questions like *Phase Transition for the Ising and Potts models on a graph* and in particular with the concept of *Purity of the Gibbs Measure*, that can be regarded as a special kind of phase transition.
Actually, interacting stochastic processes on trees and lattices often differ in a fundamental way: where a lattice model has a single transition point (a critical value for a parameter of the model) the corresponding model on a tree might possess multiple transition points. A main example of an interacting model is the usual ferromagnetic Ising model [5]. Here the interesting property which gives rise to a new transition is the purity (sometimes called extremality) of the free boundary condition Gibbs measure that as we will see is equivalent to the non solvability of the Recostruction Problem: an issue about information flow on trees.

In the following, we are going to describe these aspects in some details. First, we describe the Ising Model on a deterministic tree (i.e. the degree of a vertex is fixed) and its ferromagnetic phase transition; then we pass to define the Reconstruction Problem, for a $+/-$ signal sent from the root to the boundary of the tree. This problem is equivalent

[(*)] A tree is a graph without loops: there is no path going from a vertex to itself.

to the extremality of the free boundary Gibbs measure for the Ising model. We review quickly some known results for the Ising model on trees to go to the same problem for the Potts model where the possible signal sent from the origin can assume q values (i.e. $1, 2, \ldots, q$) instead of only two ($+$ or $-$) as in the former case. Not need to say that what can be said for the Potts model encloses, when $q = 2$, the results for the Ising model both on random and deterministic trees. For the Potts model the exact threshold for the extremality transition is not known.

The original results present in these notes are due to a joint work with C. Külske [15].

2 Ising Model on trees

In the next, the situation is as follows: we have a tree T, where we have chosen a special point, the root, that we denote by 0 (see Figure 1). The set of the vertices is called V, while G is the one of the edges. On a tree there is a unique chain of edges γ from one vertex v to another vertex w. This induces a natural notion of distance as the number of edges in γ. The set of vertices at distance n from the root is the level n of the tree. We indicates with T^N the sub-tree with just N levels and with ∂T^N the level N: i.e. the boundary of the sub-tree.

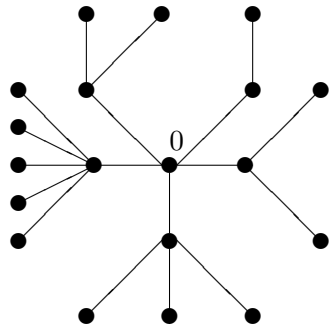


Figure 1. A tree rooted at 0.

The *Ising model* is obtained by putting at every vertex v of T^N a random variable $\eta(v) \in \{+1, -1\}$ (also called spin), and assigning to every configuration $\eta \in \{+1, -1\}^{\#(V)}$ the probability:

$$\mathbb{P}_N^\beta = \frac{1}{Z} \prod_{w:v\to w} e^{\beta\eta(v)\eta(w)} \tag{1}$$

where Z is a normalization factor.
The product runs over all the vertices, $v \to w$ means all the couples (v, w) where v is at distance one from w and w is a child of v meaning that it has a greater distance to the origin. We can look at the parameter β as the strength of the interactions between vertices at distance one. Coming the Ising model from Physics, β is often interpreted as the inverse of the temperature.
In (1) no boundary condition is specified, thus this is called the free Gibbs measure. When dealing with the ferromagnetic transition we are interested in $\mathbb{P}_N^{\beta,+}$, where + means that we set to +1 all the spins in ∂T^N. More precisely, we are interested in the limit:

$$\lim_{N\to\infty} \mathbb{P}_N^{\beta,+}(\eta(0) = +1). \tag{2}$$

We want to know if (2) is greater than $\frac{1}{2}$ for a certain range of the parameter β. Or, in other words, if for some values of β the plus-boundary condition persists to have influence at the root even when its distance from the root grows and N tends to infinity. For a regular tree where the degree k is the same for every vertex one has the following:

Theorem 2.1 *The inequality*

$$\lim_{N\to\infty} \mathbb{P}_N^{\beta,+}(\eta(0) = +1) > \frac{1}{2},$$

holds only if $k \tanh \beta > 1$.

For a proof see [5, 13].

So, ferromagnetic order on a tree is characterized by the fact that a plus-boundary condition at the leaves of a finite tree of depth N persists to have influence to the origin when N tends to infinity. From an heuristics point of view we can say that, if β is sufficiently large, there is a transport of information from the boundary of the tree to the origin even when N goes to infinity. Actually, there is a way to make precise this information theoretic interpretation of the Ising model: i.e. to show that non-solvabilty of the reconstruction problem is equivalent to the extremality of the free Gibbs measure. In the next section we show this equivalence.

3 Reconstruction Problem and extremality of the free Gibbs measure

The *Reconstruction Problem on binary channels* can be stated as follows. We send a signal from the root to the boundary, making a prescribed error at every edge of the tree. In this way one obtains a Markov chain indexed by the tree. That is, on the tree T we construct the following Markov process: to each edge e we associate a random variable $X(e)$ with

$$\mathbb{P}(X(e) = 1) = \epsilon = 1 - \mathbb{P}(X(e) = -1).$$

All the variables $X(e)$ are independent. The value of the spin $\eta(v)$ at the vertex v will be:

$$\eta(v) = \eta(0) \prod_{e\in\gamma} X(e),$$

where γ is the unique path going from the root 0 to the vertex v. While the initial value of $\eta(0)$ is chosen at random uniformly.
Suppose you know the values of the spins at distance N from the origin of the tree. What can you say about the spin at the root of the tree? Which is the probability of guessing the original value of the spin at 0 knowing that the configuration at the boundary ∂T^N is ξ, when N goes to infinity? These questions define the Reconstruction Problem.
In this problem the quantity of interest is(*)

$$\Delta_N(T,\epsilon) = \mathbb{E}\left(\left|\mathbb{P}^M(\eta(0)=1|\partial T^N=\xi) - \mathbb{P}^M(\eta(0)=-1|\partial T^N=\xi)\right|\right),$$

because it can be regarded as the difference between the probabilities of a correct and incorrect reconstruction.

Being at least $1/2$, the probability of a correct reconstruction can be written as $\frac{1+\Delta_N}{2}$. One wants to investigate if there exists a critical value ϵ_c such that:

$$\lim_{N\to\infty} \Delta_N(T,\epsilon) = 0 \text{ if } \epsilon > \epsilon_c.$$

For $\epsilon > \epsilon_c$ the problem is said to be non-solvable and solvable otherwise.

Now, it happens that if you choose

$$\frac{\epsilon}{1-\epsilon} = \exp(-2\beta).$$

on an infinite tree, the law of the Markov process defined before is the limit of the Gibbs measure

$$\mathbb{P}_N^\beta(\eta) = \frac{1}{Z}\prod_{\omega:v\to\omega} e^{\beta\eta(v)\eta(\omega)},$$

as the size of the tree grows to infinity. Then, $\lim_{N\to\infty}\Delta_N(T,\epsilon)=0$ means that

$$\lim_{N\to\infty}\mathbb{E}\left|\mathbb{P}_N^\beta(\eta(0)=1|\partial T^N=\xi) - \mathbb{P}_N^\beta(\eta(0)=-1|\partial T^N=\xi)\right| = 0,$$

that is equivalent to purity for the limiting Gibbs measure. Notice that here $\mathbb{P}_N^\beta(\eta(0)=\cdot|\partial T^N=\xi)$ is a random variable with respect to the free Gibbs measure on the boundary condition $\partial T^N=\xi$.

In this way one states that non-solvability of reconstruction is equivalent to the purity of the free Gibbs measure [10, 12], which is to say that there can be no transport of information along the tree between root and boundary, for typical signals.

In particular, for a regular tree of degree k, it holds the following theorem [5, 6, 7, 13].

Theorem 3.1 *The limiting Gibbs measure on a regular tree of degree k is pure, i.e.*

$$\lim_{N\to\infty}\mathbb{E}\left|\mathbb{P}_N^\beta(\eta(0)=1|\partial T^N=\xi) - \mathbb{P}_N^\beta(\eta(0)=-1|\partial T^N=\xi)\right| = 0 \tag{3}$$

(*) Here M stands for Markov. This the probability coming from the Markov process constructed before.

if $k\tanh^2(\beta) < 1$.(*)

A proof of the latter fact is contained in [6]. A beautiful alternative proof of the extremality for $k\tanh^2(\beta) < 1$ for regular trees was given by Ioffe [7]. The method used there in was elegant, but very much dependent on the nature of the Ising model's spin variable: i.e. on the fact that it can assume only two values. Here we give a proof that could be generalized to the Potts model. This is a simplified version of the one of Pementle and Peres.

4 Proof of Theorem 3.1

The method of the proof is by controlling the recursions from the outside to the inside of a tree of the log-likelihood ratios:

$$x_v^N := \log\left(\frac{\mathbb{P}_N^\beta(\eta(v)=+)}{\mathbb{P}_N^\beta(\eta(v)=-)}\right). \tag{4}$$

We prove that:

$$\lim_{N\to\infty}\mathbb{E}(|x_0^N|) = 0, \tag{5}$$

because this condition is equivalent to (3). Following [13] we prove (5) with the help of another quantity. We define:

$$Q_v^{N+}(x_v^N) = \int x_v^N dQ_v^{N+}(\xi) \tag{6}$$

where Q_v^{N+} is the probability for the boundary configuration ξ of T_v^N knowing that $\eta(v) = +1$:

$$Q_v^{N+}(\xi) = \mathbb{P}_N^\beta\left(\partial T_v^N = \xi | \eta(v) = +1\right). \tag{7}$$

In the same way one defines:

$$Q_v^{N-}(\xi) = \mathbb{P}_N^\beta\left(\partial T_v^N = \xi | \eta(v) = -1\right). \tag{8}$$

We need the following Lemmas:

Lemma 4.1 *For* x_v^N *one has the iteration:*

$$x_v^N = \sum_{v\to w} g(x_w^N), \tag{9}$$

(*)The Theorem is true with equality too. The bound is sharp. Here we give this version to have a proof with all the main ideas, but simpler.

with

$$g(x) = \log\left[\frac{\cosh\left(\frac{x}{2}\right) + \theta\sinh\left(\frac{x}{2}\right)}{\cosh\left(\frac{x}{2}\right) - \theta\sinh\left(\frac{x}{2}\right)}\right] \tag{10}$$

Lemma 4.2 *The projection of $Q_v^{N+}(\xi)$ onto the boundary condition of T_w^N is*

$$Q_v^{N+} = \frac{1+\theta}{2}Q_w^{N+} + \frac{1-\theta}{2}Q_w^{N-} \tag{11}$$

Lemma 4.3 *For every function f odd,*

$$\int f\left(x_v^N\right) dQ_v^{N+}(\xi) = \int f\left(|x_v^N|\right)\tanh\left(\frac{|x_v^N|}{2}\right) d\mathbb{P}_N^\beta(\xi). \tag{12}$$

Moreover, for the symmetry of the model with respect to the change of $+1$ with -1, one has :

$$-\int f\left(x_v^N\right) dQ_v^{N+}(\xi) = \int f\left(x_v^N\right) dQ_v^{N-}(\xi). \tag{13}$$

Notice that for Lemma 4.3 to prove (5) one could prove that

$$\lim_{N\to\infty} Q_0^{N+}(x_0^N) = 0$$

as we do in following, using the Banach-Cacciopoli's fixed point lemma. We use first Lemma 4.1 and Lemma 4.2 to compute:

$$\begin{aligned} Q_v^{N+}(x_v^N) &= Q_v^{N+}\left(\sum_\omega g(x_\omega^N)\right) = \\ &= \sum_\omega \left(\frac{1+\theta}{2}Q_\omega^{N+} + \frac{1-\theta}{2}Q_\omega^{N-}\right) g(x_\omega^N) = \\ &= \theta\sum_\omega Q_\omega^{N+} g(x_\omega^N) = \dots \end{aligned}$$

Now, suppose you are on a regular tree of degree k. In this case, because of the symmetry, we can say that $Q_\omega^{N+} g_\omega(x_\omega^N)$ are all equal even if rooted on different ω.

$$\dots = k\theta Q_\omega^{N+} g(x_\omega^N).$$

Now we take the Taylor expansion of $g(x_\omega)$. The function $g(x)$ is odd and concave for $x > 0$ thus, this for the Lemma 4.1 implies

$$Q_w^{N+}(g(x_\omega^N)) < Q_w^{N+}(\theta x_\omega^N) \tag{14}$$

and so

$$Q_v^{N+}(x_v) < k\theta^2 Q_\omega^{N+}(x_\omega). \tag{15}$$

Now, if $k\theta^2 < 1$ the Banach-Cacciopoli's fixed point lemma can be applied. This concludes the proof.

5 Extremality for the Potts Model

In this model the value of a spin runs over the set $\{1, 2, \dots, q\}$ and the free Gibbs measure is

$$\mathbb{P}_N^\beta = \frac{1}{Z} \prod_{v \to w} e^{2\beta\delta_{(\eta(v),\eta(w))}}. \tag{16}$$

Here $\delta_{(\cdot,\cdot)}$ is the Kronecker's delta. The extremality for this model
(i.e. $\lim_{N\to\infty} \mathbb{E}|\mathbb{P}_N^\beta(\eta(0) = i) - \mathbb{P}_N^\beta(\eta(0) = j)| = 0, \ \forall \ i, j = 1, \dots, q$) is equivalent to the solvability of the Reconstruction Problem with a q-ary symmetric channel. Here, the probability of changing the signal passing from v to w is

$$M(\eta(v) = i, \eta(w) = i) = 1 - (q-1)\lambda, \tag{17}$$

$$M(\eta(v) = i, \eta(w) = j) = \lambda \tag{18}$$

with $\lambda = \frac{1}{e^{2\beta}+q-1}$.

The best bound, valid for every k and every q, which has been previously given appears in [11]. Actually, in this situation, they prove the following:

Theorem 5.1 *Consider a tree T with degree k. For*

$$k\frac{2\theta^2}{q-(q-2)\theta} < 1,$$

with $\theta = \tanh\beta$, the free boundary condition Gibbs measure is extremal.

This condition, while being sharp at $q = 2$ (Ising model), is not sharp for $q > 2$.

In [15] we obtain a better threshold. We provide an explicit temperature interval below the ferromagnetic transition temperature for which this measure is extremal, improving older bounds of Mossel and Peres. In information theoretic language extremality of the Gibbs measure corresponds to non-reconstructability for symmetric q-ary channels. The bounds are optimal for the Ising model and appear to be close to what is conjectured to be the true values up to a factor of 0.0150 in the case $q = 3$ and 0.0365 for $q = 4$.

In this situation our main result is the following. Write

$$P = \{(p_i)_{i=1,\dots,q}, p_i \geq 0 \ \forall i, \ \sum_{i=1}^{q} p_i = 1\}$$

for the simplex of Potts probability vectors, then

Theorem 5.2 *The free boundary condition Gibbs measure is extremal when the condition* $k\frac{2\theta}{q-(q-2)\theta}\bar{c}(\beta,q) < 1$ *is satisfied. Here,*

$$\bar{c}(\beta,q) := \sup_{p\in P} \frac{\sum_{i=1}^{q}(qp_i-1)\log(1+(e^{2\beta}-1)p_i)}{\sum_{i=1}^{q}(qp_i-1)\log qp_i} \tag{19}$$

It appears that the supremum over P is achieved at the symmetric point $\frac{1}{q}(1,1,\dots,1)$ *only* in the Ising model $q = 2$. We recover the result of [11] (Theorem 5.2) from our bounds when we use the estimate $\bar{c}(\beta,q) \leq \theta$. This estimate we see indeed numerically. Moreover, numerically $\bar{c}(\beta,q)$ seems to decrease monotonically in q at fixed β. As a remark, we say that the theorem holds true a.s. with respect to the distribution of the degree at every vertex, also for Galton-Watson trees if the degree k is substituted with the mean degree of the random tree.

Let us compare with the recent literature. In their paper [16] Montanari and Mezard make the following conjecture:

Conjecture 5.3 (Mézard and Montanari 2006) *Consider the Potts model with q symbols on a d-ary tree and let $\lambda_2 = \frac{e^{2\beta}-1}{e^{2\beta}+q-1} = \frac{2\theta}{q-(q-2)\theta}$, with $\theta = \tanh(\beta)$, then if $q \leq 4$ and $k < k_{max}$, there is reconstruction if and only if $k\lambda_2^2 > 1$ (Kesten-Stigum bound). When $q \geq 5$ the Kesten-Stigum bound can never be sharp.*

This conjecture is based on extensive numerical simulations of the random recursion. Moreover, the restriction on k comes from the limitation on the values of k they can treat numerically and they actually think that $k_{max} = +\infty$.
Let us compare our bound with this conjecture. We obtain numerically $\bar{c}(\beta,q) = \frac{e^{2\beta}-1}{q-1+e^{2\beta}}(1+\epsilon(q))$ with $\epsilon(3) = 0.0150$ and $\epsilon(4) = 0.0365$. If we specialize to a binary tree, and take advantage of the possible temperature dependence of ϵ we obtain $\beta_c := \sup\{\beta, 2\frac{2\theta}{3-\theta}\bar{c}(\beta,3) < 1\} = 1.0434$ for $q = 3$ and $\beta_c := \sup\{\beta, 2\frac{2\theta}{4-2\theta}\bar{c}(\beta,4) < 1\} = 1.1555$ in the case $q = 4$.

After completion of the first draft of our work Sly's preprint [17] appeared where he proves the following.

Theorem 5.4 (Sly 2008) *When $q \leq 3$, and $k > k_{min}$, then Kesten-Stigum bound is sharp, while the Kesten-Stigum bound is never sharp when $q \geq 5$.*

His method uses large degrees to justify quadratic expansions by means of Central Limit Theorem (CLT) approximation and makes no statements for small degrees where our

estimates apply. In this region our estimates turn out to be, as of today, the closet to the Montanari-Mezard's conjecture. For $q \leq 4$ and d small, the problem of finding a sharp bound is still open. Moreover no sharp bound is known when $q \geq 5$; in this case there are numerical estimates that our rigorous thresholds fit very well.

References

[1] S. H. Strogatz, *Exploring complex networks.* Nature, **410** (2001).

[2] A-L. Barabasi, Z. N. Otvai, *Network biology: understanding the cell's functional organization.* Nature reviews, Genetics, **5** (2004).

[3] R. Albert, A-L. Barabasi, *Statistical Mechanics of complex networks.* Review of Modern Physics, **74** (2002).

[4] M. E. J. Newman, *The structure and function of complex networks.* SIAM reviews, **45**, no. 2, 167–256.

[5] R. J. Baxter, "Exactly solved models in statistical mechanics". London, Academic Press, 1982.

[6] P. M. Bleher, J. Ruiz, V.A. Zagrebnov, *On the purity of limiting Gibbs state for the Ising model on the Bethe lattice.* J. Stat. Phys. **79** (1995) 473–482.

[7] D. Ioffe, *A note on the extremality of the disordered state for the Ising model on the Bethe lattice.* Lett. Math. Phys **37** (1996), 137–143.

[8] D. Ioffe, *Extremality of the disordered state for the Ising model on general trees.* Trees (Versailles), 3–14, Progr. Probab. **40**, Birkhäuser, Basel (1996).

[9] R. Lyons, *Phase transitions on nonamenable graphs.* In: Probabilistic techniques in equilibrium and nonequilibrium statistical physics. J. Math. Phys. **41** no. 3 (2000), 1099–1126.

[10] S. Janson, E. Mossel, *Robust reconstruction on trees is determined by the second eigenvalue.* Ann. Probab. **32** no. 3B (2004), 2630–2649.

[11] E. Mossel, Y.Peres, *Information flow on trees.* The Annals of Applied Probability **13** No. 3 (2003), 817844.

[12] E. Mossel, *Reconstruction on trees: Beating the second eigenvalue.* Ann. Appl. Prob. **11** (2001), 285-300.

[13] R. Pemantle, Y. Peres, *The critical Ising model on trees, concave recursions and nonlinear capacity.* Preprint, arXiv:math/0503137v2 [math.PR] (2006).

[14] A. Gerschelfeld, A. Montanari, *Reconstruction Problem for models on random graphs.* 48th Annual Symposium on Foundations of Computer Science, Providence, RI (2007).

[15] M. Formentin, C. Külske, *On the Purity of the free boundary condition Potts measure on random trees.* To be published in Stochastic Processes and their Applications, available at arXiv:0810.0677, (2009).

[16] M. Mezard, A. Montanari, *Reconstruction on trees and spin glass transition.* J. Stat. Phys. **124** no. 6, 1317–1350 (2006).

[17] A. Sly, *Reconstruction of symmetric Potts models.* Preprint, arXiv:0811.1208 [math.PR] (2008).

Permutation groups and automata

Pablo Spiga [(*)]

1 Synchronization and automata

Let G be a *permutation group* on a finite set Ω, i.e. a subgroup of the symmetric group on Ω. A *partition* π of Ω is trivial if either it has a single part, or all its parts are singletons. A *section* or transversal of a partition π is a subset Δ of Ω such that Δ contains exactly one point from every part of π.

We recall a standard definition in permutation group theory.

Definition 1 *The permutation group G on Ω is* imprimitive *if there is a non-trivial partition π of Ω which is G-invariant, i.e. preserved by G. The permutation group is* primitive *otherwise.*

Definition 2 *The permutation group G on Ω is* non-synchronizing *if there is a non-trivial partition π of Ω and a subset Δ of Ω such that Δg is a section of π for any $g \in G$. The group G is* synchronizing *otherwise.*

The definition of synchronization for permutation groups is motivated by an application to the Černý conjecture for permutation groups, which we now describe.

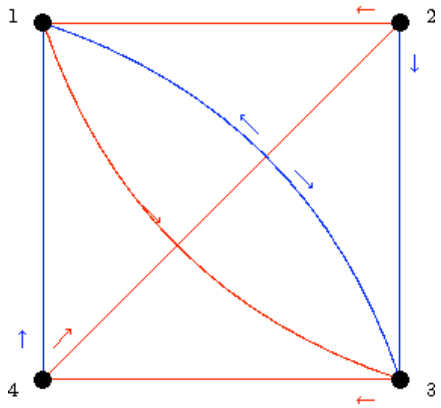
A finite deterministic *automaton* is a black box with a number of coloured buttons on the outside. The automaton can be in any of a set of internal states. Each button corresponds to a *transition* between states of the automaton. We can force the automaton to perform any desired sequence of transitions.

An automaton is usually represented as an edge-coloured directed graph, where the vertices of the digraph are the states of the automaton and the coloured edges are the transitions. Note that since we are requiring that the automaton can perform any desired sequence of transitions, there exists a unique edge of each colour leaving each vertex. In particular the digraph has constant out-degree.

A *reset word* for an automaton is a sequence of transitions which take the automaton from any state to a known state, in other words the composition of the corresponding

[(*)]Dottorato di ricerca in Matematica Pura, Dipartimento di Matematica Pura ed Applicata, Università di Padova. E-mail: spiga@math.unipd.it. Seminar held on 12 November 2008.

transitions is a constant mapping. An automaton is said to be *synchronizing* if it has a reset word.



In the automaton in the picture you can check that (Blue, Red, Blue, Blue) is a reset word which takes you to vertex 3 no matter where you start.

One of the oldest conjectures in automata theory is the *Černý conjecture*:

Conjecture 1 *If an automaton with n states is synchronizing, then the automaton has a reset word of length at most $(n-1)^2$.*

The definition of synchronizing permutation group is determined by a group theoretic approach to Černý conjecture for some classes of synchronizing automata. Note that an automaton in which every transition is a permutation is clearly non-synchronizing.

The original definition of synchronizing permutation group is the following:

Definition 3 *A permutation group G on Ω is synchronizing if, for any mapping $f : \Omega \to \Omega$ which is not a permutation, the semigroup generated by G and f contains a constant map.*

It turns out that Definition 2 and Definition 3 are equivalent.

Proposition 1 *A permutation group G is synchronizing with respect to Definition 2 if and only if G is synchronizing with respect to Definition 3.*

2 Separating and Synchronizing groups

Let Ω be a finite set of size n and G a permutation group on Ω. We will assume that G is transitive, i.e. for any $\alpha, \beta \in \Omega$ there exists $g \in G$ such that $\alpha^g = \beta$. Let R, S be subsets of Ω of size r and s respectively. The sets R, S will be said to be *inseparable* if $R^g \cap S^h \neq \emptyset$ for all $g, h \in G$. This condition is clearly equivalent to $R^g \cap S \neq \emptyset$ for all $g \in G$. By [4], if R, S are inseparable then $n \leq rs$. We focus on the extreme case where $n = rs$.

Proposition 2 *Suppose that R, S are inseparable subsets of size r, s for the permutation group G on Ω and $rs = n = |\Omega|$. Then $|R^g \cap S^h| = 1$ for all $g, h \in G$.*

According to the previous proposition we give the following definition.

Definition 4 *Let G be a transitive permutation on Ω. The group G is said to be* non-separating *if there exists a pair of inseparable sets R, S with $|R| = r > 1, |S| = s > 1$ and $rs = n = |\Omega|$. The group G is said* separating *otherwise.*

Proposition 3 *If G is separating, then G is synchronizing.*

Proof. Assume G is non-synchronizing as witnessed by the non-trivial partition π of Ω and the section S. Since G is transitive, it is easy to check that π is a uniform partition, i.e. all the parts in π have the same size. Let R be a part of π. Now, R, S are inseparable sets for G contradicting that G is separating. □

We point out that checking whether a group is separating is easier than checking whether a group is synchronizing. In fact, the definition of separating group involves just subsets of the set Ω, but the definition of synchronizing group requires to deal with partitions of the set Ω and their sections. In particular, an extensive study of synchronizing groups could be done first determining the class of separating groups and then "measuring" the difference between the synchronizing/separating definitions.

Proposition 4 *If G is synchronizing, then G is primitive.*

Proof. Assume G imprimitive and let π be a G-invariant partition of Ω. Now, let R be a part of π and S any section of π. Now, since π is G-invariant we have that Sg is a section for $\pi g = \pi$, for any $g \in G$. Thus G is non-synchronizing, a contradiction. □

The list of primitive permutation groups with $|\Omega| \leq 2000$ is known. So an obvious first step in classifying synchronizing/separating permutation groups is checking these "small" permutation groups for these properties. As we are going to see in Section 3, the checking is not trivial.

We point out that we only have a tiny handful of permutation groups which are synchronizing but non-separating. Actually we have only five examples in which three come from natural constructions in finite geometry and two are "sporadic". So, we do not understand at the time of this writing the difference between separating and synchronizing groups.

3 Graph theoretic invariants

In this section we mention that we can determine whether a group is separating/synchronizing through some graph theoretic invariants.

Let X be a graph (undirected). We denote by $\omega(X)$ the maximal size of a complete subgraph of X, i.e. a subset of vertices which are pairwise adjacent. We denote by $\alpha(X)$ the maximal size of a null subgraph of X, i.e. a subset of vertices which are pairwise non-adjacent (a clique in the complement of X). Finally, we denote by $\chi(X)$ the chromatic number of X, i.e. the minimum number of colours that are required to colour the vertices of X so that adjacent vertices have different colours.

We note that $\omega(X) \leq \chi(X)$, distinct vertices in a complete subgraph of X get distinct colours in a colouring of X.

In the next proposition we show that the synchronizing/separating property can be detected using the invariants $\omega(X), \alpha(X)$ and $\chi(X)$ for the graphs X admitting G as a group of automorphisms.

Proposition 5 *Let G be a transitive permutation group on Ω and $n = |\Omega|$.*

(*i*) *The group G is non-synchronizing if and only if there is a non-trivial G-invariant graph X for which $\omega(X) = \chi(X)$.*

(*ii*) *The group G is non-separating if and only if there is a non-trivial G-invariant graph X such that $\omega(X)\alpha(X) = n$.*

4 Some examples: classical groups on polar spaces

Let V be a vector space over a finite field $GF(q)$ on which there is a non-degenerate alternating bilinear, Hermitian, or quadratic form. The *classical group* is the group of isometries of this form, i.e. the isomorphisms of V leaving the form invariant.

The *polar space* associated with the form is the Buekenhout geometry whose points are the 1-dimensional subspaces of V and whose r-dimensional flats are the $r+1$-dimensional subspaces of V on which the form vanishes. In particular the lines, planes of the polar space are the $2, 3$-dimensional subspaces of V on which the form vanishes.

Assume that the dimension of a maximal subspace of V on which the form vanishes is greater than 2, i.e. the Witt index is greater than 1.

In this case, a classical group acting on the points of the polar space is a rank 3 group. In other words G has only two G-invariant graphs: the *collinearity graph* and its complement. We recall that the collinearity graph of the polar space is the graph where the vertices are the points of the polar space and two points are adjacent if they lie on a line of the polar space.

The aim of this section is applying Proposition 5 for a classical group acting on its polar space. In fact in this case we have only two graphs to study.

We recall that an *ovoid* is a set of points which meets every maximal subspace in a single point. In particular any two points of an ovoid are non-collinear. So an ovoid is a coclique in the collinearity graph. Also, since an ovoid meets every maximal subspace in a single point, we have that an ovoid is a coclique of maximum possible size in the collinearity graph.

A *spread* is a partition of the point set of the polar space into maximal subspaces. In particular it is a colouring of the complement of the collinearity graph.

Summing up, from Proposition 5 we get the following proposition.

Proposition 6 *Let G be a classical group acting on the points of its polar space. Then G is non-synchronizing if and only if there exists an ovoid and a spread, or there exists a partition of the polar space into ovoids.*

Deciding which polar spaces have ovoids and spreads has been studied by finite geometers for many years (the problem goes back to Segre), and the problem is far from a complete solution. In particular, these examples show that in general it is very hard understanding whether a permutation group is synchronizing.

References

[1] João Araújo, *A group theoretical approach to synchronizing automata and the Černý problem.* Preprint, Lisbon, January 2006.

[2] F. Arnold and B. Steinberg, *Synchronizing groups and automata.* Theoretical Computer Science 359 (2006), 101–110.

[3] P. J. Cameron, "Projective and Polar Spaces". QMW Maths Notes 13, QMW, London, 1991.

[4] J. D. Dixon and B. Mortimer, "Permutation groups". Springer-Verlag, New York 1996.

[5] A. Trahtman, *The road coloring problem.* Israel J. Math., to appear.

On matrices with the Edmonds-Johnson property

Alberto Del Pia (*)

Abstract. Integer programming is the problem of optimizing a linear function over the integral points in a polyhedron P, expressed as a system of linear inequalities. It is known that it is equivalent to optimizing such linear function over the polyhedron P_I, that is the convex hull of the integral points in P. One of the main problems in mathematical programming is to developed tools to get P_I, and one of such methods is the Chvatal-Gomory procedure. This procedure, starting from P, gives a sequence of smaller polyhedra, $P', P'', \ldots$, that converges to P_I in a finite number of iterations. The number of iterations needed to get P_I gives an order of complexity to the problems. We survey some old and new results of classes of problems in which $P' = P_I$, presented in the PhD Thesis of A. Del Pia [3].

1 Mathematical programming

A *linear programming problem* is a problem of the form

$$\begin{array}{ll} \max & cx \\ \text{s.t.} & x \in P, \end{array}$$

where $P \subseteq \mathbb{R}^n$ is a *polyhedron*, i.e. P is defined by a system of linear inequalities: $P = \{x : Ax \leq b\}$. A fundamental result states that any linear programming problem can be solved efficiently.

An *integer (linear) programming problem* is obtained from a linear programming problem by adding the constraint that all the variables must take an integer value. Thus an integer programming problem appears in the form

$$\begin{array}{ll} \max & cx \\ \text{s.t.} & x \in P \\ & x \in \mathbb{Z}^n. \end{array}$$

No algorithm is known to solve a general integer programming problem efficiently, and many researchers believe that no such algorithm exists.

(*)Dottorato di ricerca in Matematica Computazionale, Dipartimento di Matematica Pura ed Applicata, Università di Padova. E-mail: delpia@math.unipd.it. Seminar held on 26 November 2008.

For every polyhedron P, we denote with P_I the convex hull of the integral points in P. Then the integer programming problem over P is equivalent to the problem

$$\begin{array}{ll} \max & cx \\ \text{s.t.} & x \in P_I. \end{array} \tag{1}$$

We say that a polyhedron is *rational* if it is defined by a rational system of linear inequalities, and from now on we will work only with rational systems and polyhedrons. A theorem by Meyer [10] states that for any rational polyhedron P, the set P_I is also a polyhedron. Hence problem (1) is a linear programming problem.

In general it is very hard to find P_I starting from P. Several automatic procedures to find inequalities generating P_I have been introduced in literature. One of the simpler is the Chvátal closure, and we introduce it in the next section.

2 The Edmonds-Johnson property

If $P = \{x : Ax \leq b\}$, then each rational valid inequality for P is of the form

$$(\lambda A)x \leq \lambda b, \qquad \lambda \geq 0$$

where we can assume $\lambda A \in \mathbb{Z}^n$. Thus each inequality of the form

$$(\lambda A)x \leq \lfloor \lambda b \rfloor, \qquad \lambda \geq 0,\ \lambda A \in \mathbb{Z}^n$$

is valid for P_I. Such inequalities are called *Chvátal(-Gomory) inequalities*, and the vector λ is called *Chvátal-Gomory multiplier*.

For every polyhedron $P = \{x : Ax \leq b\}$, the *first (Chvátal) closure* of P is defined as

$$P' = \{x : (\lambda A)x \leq \lfloor \lambda b \rfloor, \forall \lambda \geq 0, \lambda A \in \mathbb{Z}^n\}.$$

A basic result, due to Chvátal [1], is that for every polyhedron P, the set P' is also a polyhedron.

Clearly we have that:

$$P \supseteq P' \supseteq P'' \supseteq \cdots \supseteq P_I.$$

A fundamental theorem states that for each polyhedron P, there exists a number t such that $P^{(t)} = P_I$, where $P^{(0)} := P, P^{(t+1)} := P^{(t)'}$. Thus we say that the *Chvátal rank* of a polyhedron P is the smallest number t such that $P^{(t)} = P_I$.

The *strong Chvátal rank* of a rational matrix A is the smallest number t such that the polyhedron defined by the system $b \leq Ax \leq c$, $l \leq x \leq u$ has Chvátal rank at most t for all integral vectors b, c, l, u. Matrices with strong Chvátal rank at most 1 are said to have the *Edmonds-Johnson property*, since it was shown by Edmonds and Johnson [7] that any integral matrix $A = (\alpha_{ij})$ such that $\sum_i |\alpha_{ij}| \leq 2$ for each column index j has Chvátal rank at most 1.

While the class of integral matrices with strong Chvátal rank 0 is well understood, since it is the class of totally unimodular matrices, no general characterization is known for integral matrices with the Edmonds-Johnson property.

Another class of matrices known to have the Edmonds-Johnson property is the class of the edge-node incidence matrices of bidirected graphs with no odd-K_4 minors (Gerards and Schrijver [9].

All matrices in the above classes are *totally half-modular*, that is, they are integral matrices such that for any nonsingular square submatrix B, $2B^{-1}$ is integral.

The class of totally half-modular matrices with the Edmonds-Johnson property is closed under the following operations:

(i) deleting or permuting rows or columns, or multiplying them by -1;

(ii) dividing by 2 an even row;

(iii) *pivoting* on a 1 entry, i.e. replacing matrix $\begin{pmatrix} 1 & g \\ f & D \end{pmatrix}$ by the matrix $\begin{pmatrix} -1 & g \\ f & D - fg \end{pmatrix}$, where f is a column vector, and g is a row vector.

We say that a matrix B is a *minor* of A if it arises from A by a series operations (i)-(iii).

Gerards and Schrijver [8] conjectured that a totally half-modular matrix has the Edmonds-Johnson property, if and only if it has no minor equal to $A_4 = \begin{pmatrix} 2 & 1 & 1 & 0 \\ 2 & 1 & 0 & 1 \\ 2 & 0 & 1 & 1 \end{pmatrix}$ or $A_3 = \begin{pmatrix} 1 & 2 & 0 \\ 1 & 2 & 2 \\ 1 & 0 & 2 \end{pmatrix}$. As far as we know, this question is wide open.

The two cited classes of matrices with the Edmonds-Johnson property are particular cases of this conjecture. In fact, the result of Gerards and Schrijver implies that an integral matrix $A = (\alpha_{ij})$ satisfying $\sum_j |\alpha_{ij}| \leq 2$ for each row index i, has the Edmonds-Johnson property if and only if it does not contain the matrix A_4 as a minor.

3 Our contribution

Our contribution consists into two classes of matrices with the Edmonds-Johnson property. Both results are particular cases of Gerards and Schrijver's conjecture.

3.1 A class from vertex covering problems

In Chapter 3 we study systems of the from

$$\begin{aligned} b \leq Mx \leq c \\ l \leq x \leq u, \end{aligned} \tag{2}$$

for integral vectors b, c, l, u, where M is obtained from a totally unimodular matrix with two nonzero elements per row by multiplying by 2 some of its columns.

The case where M is obtained from the transpose of the incidence matrix of a bipartite graph by multiplying by 2 some of the columns, has been studied by Conforti et al. in [2].

In this case, they derived an explicit characterization of the inequalities defining the integer hull, and showed that the problem of maximizing a linear function cx, with c integral, over the integer hull of (2) can be solved in strongly polynomial time.

We give an explicit description of a totally dual integral system that describes the integer hull of the polyhedron P defined by (2). Since the inequalities of such totally dual integral system are Chvátal inequalities for P, this implies that the matrix M has the Edmonds-Johnson property. We also derive a strongly polynomial time algorithm to find an integral optimal dual solution for the problem of maximizing a linear function with integer coefficients over the totally dual integral system describing the integer hull of (2). The results in Chapter 3 are joint work with G. Zambelli [5].

3.2 A class from bidirected graphs

A *bidirected graph* is a triple $G = (V, E, \sigma)$, where (V, E) is an undirected graph (possibly with loops and parallel edges) and σ is a *signing* of (V, E), i.e. a map that assigns to each $e \in E$ and $v \in e$ a *sign* $\sigma_{v,e} \in \{+1, -1\}$. For convenience, we define $\sigma_{v,e} := 0$ if $v \notin e$. We denote with $V(G)$ the nodes of G, and with $E(G)$ the edges of G. The *odd edges* of G are the edges with the same sign in its endnodes. A *cycle* in G is a closed path of length at least two. A cycle C in G is *even* if the number of odd edges in it is even, and is *odd* otherwise.

The *sign matrix* of a bidirected graph G is the $V \times E$ matrix $\Sigma(G) = (\sigma_{v,e})$. Given a bidirected graph G and a subset F of its edges, we denote by $A(G, F)$ the matrix obtained from $\Sigma(G)$ by multiplying by 2 the columns corresponding to the edges in F.

A matrix $A(G, F)$ is totally half-modular if and only if (G, F) satisfies the following.

Cycles condition: *no odd cycle of G contains edges in F.*

In Chapter 3 we characterize the pairs (G, F) that satisfy the cycles condition for which $A(G, F)$ has the Edmonds-Johnson property.

Let $G = (V, E, \sigma)$ be a bidirected graph and $F \subseteq E$. Given a node $v \in V$, the signing σ' obtained from σ by setting $\sigma'_{v,e} = -\sigma_{v,e}$ for all edges e incident with v is said to be obtained by *switching signs on the node v*.

Given $e = vw \in E$, the signing σ' obtained from σ by setting $\sigma'_{v,e} = -\sigma_{v,e}$, $\sigma'_{w,e} = -\sigma_{w,e}$, is said to be obtained by *switching signs on the edge e*.

Given a node $v \in V$, the pair (G', F') obtained from (G, F) by *deleting node v* is defined as follows. $V(G') = V \setminus \{v\}$, $E(G')$ contains all edges of E not incident to v and a loop on w for each edge $vw \in E$ with $v \neq w$. The signing on the edges of G' coincides with σ on $G \setminus v$, while $F' = F \cap E(G')$.

Given an edge $e \in E$, (G', F') is obtained from (G, F) by *deleting edge e* if $G' = (V, E \setminus \{e\}, \sigma')$ and $F' = F \setminus \{e\}$, where σ' coincides with σ on $E \setminus \{e\}$.

Let $e = vw \in E$ such that e is not a loop and $\sigma_{v,e} \neq \sigma_{w,e}$. We say that (G', F') is obtained from (G, F) by *contracting edge e* if G' is the bidirected graph obtained by replacing the nodes v, w with one new node $r \notin V$, by deleting all the edges vw with $\sigma_{v,vw} \neq \sigma_{w,vw}$, by replacing each edge vw with $\sigma_{v,vw} = \sigma_{w,vw}$ with a loop in r with sign

$\sigma_{v,vw}$, by replacing each edge $tv, t \neq w$ or $tw, t \neq v$, with an edge tr in $E(G')$, and by letting the signing in G' coincide with σ on $E(G')$. Let F' be the union of F and the set of the loops in r corresponding to edges vw in G with the same sign in their endnodes.

Given a pair (G, F), we call a pair (G', F') a *minor* of (G, F) if it is obtained by the latter through some of the following operations:

(1) switching signs on a node or on an edge of G;

(2) deleting a node or an edge in (G, F);

(3) contracting an edge $vw \in E(G) \setminus (F \cup L(G))$;

(4) contracting an edge $vw \in F \setminus L(G)$ such that v is incident only to edges in F.

Operation (1) corresponds to multiplying by -1 either the row of $A(G, F)$ corresponding to v, or the column of $A(G, F)$ corresponding to e. Operation (2) corresponds to deleting either the row of $A(G, F)$ corresponding to v, or the column of $A(G, F)$ corresponding to e. Operation (3) corresponds to pivoting the element in position (v, vw) of $A(G, F)$, and removing the row corresponding to v and the column corresponding to vw. Operation (4) corresponds to dividing by 2 the row of $A(G, F)$ corresponding to v, pivoting the element in position (v, vw), and removing the row corresponding to v and the column corresponding to vw.

Thus the class of pairs (G, F) that satisfy the cycles condition and such that $A(G, F)$ has the Edmonds-Johnson property is closed under taking minors.

Let G_4 be the pair (G, F) as in Figure 1.

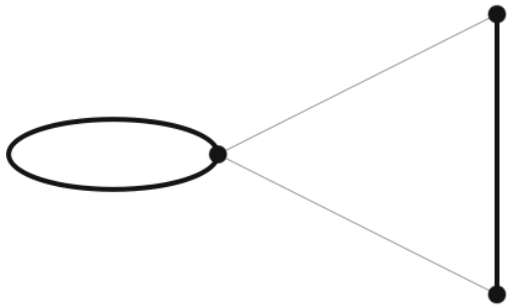


Figure 1. G_4: the boldfaced edges represent the edges in F.

The following is the main result of Chapter 3.

Theorem 1 *Given a pair (G, F) that satisfies the cycles condition, $A(G, F)$ has the Edmonds-Johnson property if and only if (G, F) does not contain G_4 as a minor.*

The theorem of Edmonds and Johnson discussed above is the special of Theorem 1 where F is a subset of the loops of G. Since the matrix A_3 is a minor of $A(G_4)$, Theorem 1 implies that Gerards and Schrijver's conjecture is true for totally half-modular matrices

obtained from $\{0, \pm 1\}$-matrices with at most two nonzero entries per column by multiplying by 2 some of the columns. In fact, the only minor that we need to exclude is A_3, because A_4 never appears.

We also show that, for each pair (G, F) that satisfies the cycles condition and that does not contain G_4 as a minor, one can minimize in polynomial time any linear function over the integer hull of $b \leq A(G, F)x \leq c$, $l \leq x \leq u$, for all integral vectors b, c, l, u. The results in Chapter 3 are joint work with A. Musitelli and G. Zambelli [4]. A partial result was shown by Del Pia and Zambelli [6].

References

[1] V. Chvatal, *Edmonds polytopes and a hierarchy of combinatorial problems.* Discrete Mathematics **4** (1973), 305-337.

[2] M. Conforti, A.M.H. Gerards, and G. Zambelli, *Mixed-integer vertex covers on bipartite graphs.* In *Integer Programming and Combinatorial Optimization* (M. Fischetti and D.P. Williamson eds.), proceedings of IPCO 2007, LNCS Vol. 4513, Springer (2007), 324-336.

[3] A. Del Pia, On matrices with the Edmonds-Johnson property, Ph.D. Thesis, Universitá di Padova, Padova, 2008.

[4] A. Del Pia, A. Musitelli, and G. Zambelli, *A class of matrices with the Edmonds-Johnson property.* Manuscript, 2008.

[5] A. Del Pia and G. Zambelli, *A class of matrices arising from bidirected graphs: total dual integrality and strong Chvátal rank.* Submitted, 2007.

[6] A. Del Pia and G. Zambelli, *A class of matrices with Strong Chvátal Rank 1.* Submitted, 2008.

[7] J. Edmonds and E.L. Johnson, *Matching, Euler tours and the Chinese postman.* Mathematical Programming **5** (1973), 88-124.

[8] A.M.H. Gerards, personal communication, 2007.

[9] A.M.H. Gerards and A. Schrijver, *Matrices with the Edmonds-Johnson property.* Combinatorica **6** (1986), 365-379.

[10] R.R. Meyer, *On the existence of optimal solutions to integer and mixed-integer programming problems.* Mathematical Programming **7** (1974), 223-235.

Some problems from Geometric Measure Theory: Plateau, Bernstein, Dido

Davide Vittone (*)

Abstract. The aim of Geometric Measure Theory (GMT) is to approach geometric problems by means of measure-theoretical tools. After a brief presentation of some GMT definitions of "surface measure", we will summarize the history and main results about three classical questions: minimal surfaces (or Plateau problem), the Bernstein problem and the isoperimetric (or Dido) problem.

1 Geometric Measure Theory

The aim of Geometric Measure Theory (GMT) is to attack geometric problems (e.g. minimal surfaces or the isoperimetric problems) by means of measure theoretical tools. The first works in this direction trace back to Besicovitch (around 1930), even if earlier elements can be found in papers by Hausdorff, von Koch, Cantor and others. Its development was influenced by mathematicians of the utmost importance as De Giorgi, Fleming, Almgren, Federer, but this brief list is surely incomplete.

In this talk we will focus our attention on three classic problems which have been solved through GMT tools: namely, the Plateau problem about minimal surfaces, the Berstein problem and Dido's isoperimetric problem. Before treating them we need however to introduce some specific notion from GMT; in particular, the nature of our problems requires to clarify what we mean by "surface measure".

Let S be a subset of $\mathbb{R}^n$ and let $m \in \mathbb{N}$ and $\delta > 0$ be fixed. We define

$$\mathcal{H}^m_\delta(S) := \inf\left\{ \omega_m \sum_{i=1}^{\infty} \left(\frac{diam(S_i)}{2}\right)^m : S \subset \bigcup_{i=1}^{\infty} S_i,\ diam(S_i) \leqslant \delta \right\},$$

where ω_m is the measure of the m-dimensional unit ball. Note that if $\delta_1 < \delta_2$ one has $\mathcal{H}^m_{\delta_1}(S) \geqslant \mathcal{H}^m_{\delta_2}(S)$ and so the *m-dimensional Hausdorff measure* of S

$$\mathcal{H}^m(S) := \lim_{\delta \to 0^+} \mathcal{H}^m_\delta(S)$$

(*)Dipartimento di Matematica Pura ed Applicata, Università di Padova, via Trieste 63, I-35122 Padova (Italy). E-mail: vittone@math.unipd.it . Seminar held on 10 December 2008.

is well defined. Since it will be useful in the sequel we notice that, more than the geometry of $\mathbb{R}^n$, the definition of Hausdorff measures involves the notion of *diameter* of sets: therefore, Hausdorff measures can be defined in general metric spaces.
Hausdorff measures are a good generalization of the notion of surface measure: indeed, if S is a regular m-dimensional surface then $\mathcal{H}^m(S) = \sigma_m(S)$, where σ_m is the m-dimensional surface measure associated with S. Moreover, one has $\mathcal{H}^n = \mathcal{L}^n$.

A different notion of "surface measure" is given by the *perimeter measure* of a set, that we are now going to introduce. Following an intuition by R. Caccioppoli, the idea is to think of (hyper-)surfaces as boundaries of sets; accordingly, more than *surfaces* and their measures we will focus our attention on *sets* and their "boundary measure".

Definition 1 *Given $E \subset \mathbb{R}^n$ measurable and $\Omega \subset \mathbb{R}^n$ open, the* perimeter *of E in Ω is*

$$P_E(\Omega) := \sup\left\{ \int_E \mathrm{div} g : g \in \mathbf{C}^\infty_c(\Omega, \mathbb{R}^n), |g| \leqslant 1 \right\}.$$

It is not difficult to show that P_E can be equivalently defined as the total variation $|\nabla \chi_E|$ of the distributional gradient of the characteristic function χ_E of E. Therefore, E has finite perimeter if and only if the distributional derivative $\nabla \chi_E$ is a vector measure with finite total variation. If E is a set with smooth boundary the divergence theorem gives

$$P_E(\Omega) = \sup_g \int_E \mathrm{div} g = \sup_g \int_{\partial E \cap \Omega} \langle g, \nu_{\partial E} \rangle \, d\sigma_{n-1} = \mathcal{H}^{n-1}(\partial E \cap \Omega)$$

and so $P_E = \mathcal{H}^{n-1} \llcorner \partial E$. The latter formula, however, cannot hold in general: a counterexample is given by the set $E = \mathbb{Q}^n \subset \mathbb{R}^n$, for which $P_E = 0$ while $\partial E = \mathbb{R}^n$.

We would like however to possess a representation formula of the type $P_E = \mathcal{H}^{n-1} \llcorner \mathcal{F}E$ for a suitable "boundary" $\mathcal{F}E$ of E. A smart attempt is that of considering a "measure theoretic" boundary $\partial_M E$ [(*)], defined as the set of those $x \in \mathbb{R}^n$ such that

$$0 < \mathcal{L}^n(E \cap B(x,r)) < \omega_n r^n \quad \text{for any } r > 0\,.$$

Note that $\partial_M E$ is stable under modifications of E on $\mathcal{L}^n$-negligible sets, exactly as the perimeter of E; moreover, the perimeter P_E is concentrated on $\partial_M E$. Nevertheless, $\partial_M E$ is not the good notion of "boundary" we were looking for. Consider in fact an enumeration $\{q_i\}_i$ of $\mathbb{Q}^n$ and the set $E := \bigcup_i B(q_i, 2^{-i})$. It is possible to prove that E has finite perimeter while $\mathcal{L}^n(\partial_M E) > 0$; the latter inequality is sufficient to conclude that $\mathcal{H}^{n-1}(\partial_M E) = \infty$.

According to E. De Giorgi's terminology we define the *reduced boundary* $\mathcal{F}E$ of E as the set of those points $x \in \partial_M E$ such that

$$\text{the limit } \nu_E(x) := \lim_{r \to 0} \frac{\nabla \chi_E(B(x,r))}{|\nabla \chi_E|(B(x,r))} \text{ exists and is a unit vector.}$$

[(*)] Actually, in the literature one usually refers to the measure theoretic boundary as the set

$$\partial^* E := \left\{ x \in \mathbb{R}^n : \liminf_{r \downarrow 0} \frac{\mathcal{L}^n(E \cap B(x,r))}{\omega_n r^n} > 0 \text{ and } \limsup_{r \downarrow 0} \frac{\mathcal{L}^n(B(x,r) \setminus E)}{\omega_n r^n} < 1 \right\}$$

which is different from our $\partial_M E$.

Example 2 If E is a square in $\mathbb{R}^2$, then $\partial E = \partial_M E$ is constituted by its sides and the vertices. The reduced boundary $\mathcal{F}E$ is instead given by the sides without the vertices, since at these points the previous limit ν_E is a vector, parallel to a diagonal of the square, of length $1/\sqrt{2}$.

The vector $\nu_E(x)$ is a sort of inward normal to the set E (see also the following Theorem 3, where a generalized Gauss-Green formula is stated): in some sense, the reduced boundary is the set of points where a tangent plane is defined. The following fundamental result was proved by De Giorgi in 1955.

Theorem 3 *If E has finite perimeter, then $P_E = \mathcal{H}^{n-1} \llcorner \mathcal{F}E$ and*

$$\int_E \mathrm{div} g = -\int_{\mathcal{F}E} \langle g, \nu_E \rangle \, d\mathcal{H}^{n-1} \qquad \forall g \in \mathbf{C}_c^\infty(\mathbb{R}^n, \mathbb{R}^n) \, .$$

Moreover, $\overline{\mathcal{F}E} = \partial E$ and $\mathcal{F}E$ is rectifiable*, i.e.*

$$\mathcal{F}E \ \subset \ \mathcal{N} \cup \bigcup_{i=1}^{\infty} S_i$$

where S_i are $\mathbf{C}^1$ hypersurfaces and $\mathcal{H}^{n-1}(\mathcal{N}) = 0$.

Remark 4 After possibly modifying E on a negligible set, it is always possible to suppose that $\partial E = \partial_M E$. We agree henceforth that such an assumption holds.

2 Minimal surfaces

The classical problem proposed by the phisicist A. F. Plateau is the following: given a closed curve $\gamma \subset \mathbb{R}^3$, does there exist the surface Σ with minimal area among those whose boundary is γ? Has Σ to be regular?
Again in the spirit of considering surfaces as boundaries of sets, our version of the Plateau problem will be the one considered by De Giorgi: more than minimal surfaces, our target will be constituted by perimeter minimizing sets. We say that E is *perimeter minimizing* in an open set $\Omega \subset \mathbb{R}^n$ if

$$P_E(\Omega) \leqslant P_F(\Omega) \qquad \text{for any } F \text{ with } E \Delta F \Subset \Omega.$$

The fundamental result about regularity of perimeter minimizing sets was obtained by De Giorgi:

Theorem 5 *Let E be perimeter minimizing. Then the reduced boundary $\mathcal{F}E$ is analytic. Moreover $\mathcal{F}E$ is open in ∂E and $\mathcal{H}^{n-1}(\partial E \setminus \mathcal{F}E) = 0$.*

Therefore, the regularity of the reduced boundary is the best possible; the core of De Giorgi's proof was to show that $\mathcal{F}E$ is of class $\mathbf{C}^{1,\alpha}$, since a very well established PDE theory guaranteed the step from $\mathbf{C}^{1,\alpha}$ to analytic regularity. De Giorgi's result, however, did not exclude the possibility of singular (e.g. corner-like) points, i.e. points belonging

to $\partial E \setminus \mathcal{F}E$. Besides the $\mathcal{H}^{n-1}$-negligibility of this *singular set*, De Giorgi believed that it has to be empty. In order to prove this, his strategy was the following. Suppose that the singular set is not empty, w.l.o.g. $0 \in \partial E \setminus \mathcal{F}E$, and consider the dilated sets $rE, r > 0$. Then a certain sequence of dilated sets $r_j E$ has to "converge" to a singular minimal cone $C \subset \mathbb{R}^n$. Precisely:

- as $r_j \to +\infty$, the associated characteristic functions $\chi_{r_j E}$ converge in $L^1_{loc}(\mathbb{R}^n)$ to χ_C;
- C is a cone ($C = rC \ \forall r > 0$), is perimeter minimizing (because an "L^1-limit" of perimeter minimizing set is perimeter minimizing) and has a singularity at 0 (i.e. C is not a halfspace: the fact that $0 \notin \mathcal{F}E$ prevents the existence of a "tangent plane").

Therefore, the nonexistence of singular minimal cones in $\mathbb{R}^n$ would imply the analytic regularity of minimal boundaries in $\mathbb{R}^n$. W. Fleming (1961) and F. Almgren J. (1965) proved, respectively, that no singular minimal cone exists in $\mathbb{R}^3$ and $\mathbb{R}^4$; J. Simons in 1967 extended the same result to $\mathbb{R}^n$, $n \leqslant 7$ by making use of a powerful differential geometry argument. As a corollary, perimeter minimizing sets in $\mathbb{R}^n$, $n \leqslant 7$, have analytic boundaries.
Simons' argument stopped in front of the so called *Simons' cone*

$$\mathcal{S} := \{(x, y) \in \mathbb{R}^8 \equiv \mathbb{R}^4 \times \mathbb{R}^4 : |x| < |y|\}.$$

In 1969, E. Bombieri, E. De Giorgi and E. Giusti were able to prove that $\mathcal{S}$ is perimeter minimizing, thus showing that $\partial E \setminus \mathcal{F}E$ can actually be nonempty. A year later, H. Federer completed the picture by proving that the dimension of the singular set cannot exceed $n - 8$: this result is clearly sharp because the Simons' cone in $\mathbb{R}^8$ has a nonempty singular set (constituted by the origin) of dimension 0.

3 The Bernstein problem

Before introducing the *Bernstein problem* we need to recall some preliminary fact. For a given regular function $u : \mathcal{D} \subset \mathbb{R}^n \to \mathbb{R}$, the area of its graph is given by

$$\mathscr{A}(u) := \mathcal{H}^n\big(\mathrm{gr}(u)\big) = \int_{\mathcal{D}} \sqrt{1 + |\nabla u|^2}\, d\mathcal{L}^n .$$

If u minimizes $\mathscr{A}$ (for some boundary datum), then u solves the *minimal surface equation*

$$\mathrm{div} \frac{\nabla u}{\sqrt{1 + |\nabla u|^2}} = 0 \quad \text{in } \mathcal{D}, \tag{3.1}$$

which is the Euler equation for the functional $\mathscr{A}$. It is immediate to see that affine functions on $\mathbb{R}^n$ are global solutions of (3.1). In 1915, S. Bernstein proved that for $n = 2$ also the converse is true: any global solution $u : \mathbb{R}^2 \to \mathbb{R}$ of (3.1) is affine. More "modern" proofs of the same fact have subsequently been found: we recall in particular the one by

J. C. C. Nitsche (1957) based on complex analysis. However, none of these proofs can be extended to higher dimensional cases in order to solve the so called Bernstein problem:

Do there exist global solutions $u : \mathbb{R}^n \to \mathbb{R}$ of (3.1) which are not affine functions?

A first partial answer was given by J. Moser (1961) who proved, as a corollary of his Harnack's inequality, that if $u : \mathbb{R}^n \to \mathbb{R}$ solves (3.1) and $|\nabla u| \leqslant C$, then u is affine. Later on, in 1964 (before his work with Bombieri and Giusti) De Giorgi was able to show that, if there exists a non-affine solution to the Bernstein problem in $\mathbb{R}^n$, then there exists a singular minimal cone in $\mathbb{R}^n$. As a consequence, there exists no global non-affine solution of (3.1) in $\mathbb{R}^3$. A year later, the already mentioned result by Almgren on the nonexistence of singular minimal cones in $\mathbb{R}^4$ allowed to extend the same result to $\mathbb{R}^4$. Similarly, after the 1967 result by Simons the nonexistence of global non-affine solution of (3.1) was proved in $\mathbb{R}^n,\ n \leqslant 7$.
De Giorgi knew how to construct a counterexample to the Bernstein problem in $\mathbb{R}^n$ provided a singular minimal cone was given in $\mathbb{R}^n$; we now know that Simons' cone is minimal but, at the time, De Giorgi believed it was not. It was Bombieri who convinced him that $\mathcal{S}$ had to be minimal: the result was the celebrated paper [1], were together with Giusti they were able to prove, at once, the minimality of $\mathcal{S}$, the existence of singular minimal surfaces and the existence of non-affine solutions to the Bernstein problem in $\mathbb{R}^n$ for any $n \geqslant 8$.

4 The isoperimetric problem

In Virgil's Aeneid it is told that Queen Dido, after arriving in Northern Africa, asked to Lybia's king Jarba for some land to settle in. The king's sarcastic reply was a promise of as much land as an ox leather could contain. Dido then cut the leather into a long thin strip, with which she surrounded the region where Carthago was founded.
The mathematical problem lying behind the myth is clear: which is the largest amount of area that can be enclosed by a curve of fixed length? Equivalently, one has the isoperimetric problem: which is the figure minimizing the perimeter among those with fixed volume? It is well known that the solution is provided by balls, but it took a long time to achieve a complete and rigorous proof of this.

In 1838 J. Steiner gave a first incomplete proof by making use of what was to become the key tool in the approach to the isoperimetric problem: the so called *Steiner symmetrization*, which we are now going to describe.
Let $\Pi \subset \mathbb{R}^n$ be an hyperplane and, for any $x \in \Pi$, let l_x be the line passing through x and orthogonal to Π. Given a measurable set $E \subset \mathbb{R}^n$ and $x \in \Pi$ define $d_E(x) := \mathcal{L}^1(E \cap l_x)$. The Steiner symmetrization of E (with respect to Π) is provided by the set $E_s \subset \mathbb{R}^n$ such that, for any $x \in \Pi$, $E_s \cap l_x$ is a segment of center x, orthogonal to Π, of length $d_E(x)$. By Fubini's theorem, this construction preserves the volume bound, i.e. $\mathcal{L}^n(E_s) = \mathcal{L}^n(E)$; the not immediate part is to prove that the perimeter of the symmetrized set is not greater than the one of the original set. Once this is know, it is relatively not difficult (by performing the symmetrization "sufficiently many times") to show that isoperimetric sets are given by balls.

It was only in 1935 that L. A. Lusternik gave a complete answer to the isoperimetric problem, proving that Steiner symmetrization does not increase the *Minkowski content*. The latter is another possible definition for the "boundary measure" of a set: we are however not going into details, just recalling that it coincide with the perimeter measure for smooth sets. Going back to our isoperimetric problem (the one concerning the minimization of the *perimeter* measure), a complete solution was given in 1958 by De Giorgi, who showed the key inequality $P_{E_s} \leqslant P_E$ and so the isoperimetric property of balls.

The isoperimetric problem was the motivation for a lot of beautiful mathematical intuitions. We would like to recall the solution, for the planar case, given in 1901 by A. Hurwitz, who made use of a nice argument of Fourier analysis. The seek for alternative proofs is motivated also by the importance of isoperimetric type problems in more general settings: for example, the nice solution to the Dido problem given by F. Hélein (1994) with the use of *calibrations* can be extended to show that balls are isoperimetric also in the sphere S^2 and the hyperbolic plane H^2.

5 Geometric Measure Theory in the Heisenberg group $\mathbb{H}^1$

In the last few years, a systematic attempt (which traces back to De Giorgi himself) to develop Geometric Measure Theory in the general framework of metric spaces has become the object of many studies. One of the most fruitful setting in this sense is provided by the *Heisenberg group*, a particular metric space whose structure has been of inspiration for many other more general investigations.
More precisely, the (first) Heisenberg group $\mathbb{H}^1$ is the noncommutative Lie group diffeomorphic to $\mathbb{R}^3$ whose Lie algebra admits a basis X, Y, T with $[X, Y] = T$ and $[X, T] = [Y, T] = 0$. The vector fields X and Y have to be considered as a sort of "privileged" directions; directions in $\text{span}\{X, Y\}$ are called *horizontal*.
Given $P, Q \in \mathbb{H}^1$, their *Carnot-Carathodory* distance is defined by

$$d_c(P, Q) := \inf \left\{ \ell(\gamma) : \gamma \text{ joins } P, Q \text{ and } \gamma' \in \text{span}\{X, Y\} \text{ a.e.} \right\}$$

where the length $\ell(\gamma)$ is computed in such a way that X, Y are orthonormal. The Hausdorff dimension of $\mathbb{H}^1$ is 4.
The Heisenberg group is a length space: namely, geodesics always exist but are (possibly) not unique, not even locally. More precisely, for any $s \neq 0$ the points P and $Q := \exp(sT)(P)$ are connected by a continuous family of geodesics. A geometric characterization of such a family is the following: if one identifies $\mathbb{H}^1$ with $\mathbb{R}^3$ by means of exponential coordinates $(x, y, t) = \exp(xX + yY + tT)(e)$ (e being the group identity), then any geodesic between P and Q is a spiral from P to Q. Any other geodesic can be obtained by rotating one of these spirals around (the line parallel to) the t-axis passing through P and Q, and any rotated spiral is a geodesic. In this way we obtain a rotationally invariant figure that we agree to denote by $\mathcal{I}_{P,s}$: we will meet it later.
Let us now introduce the $\mathbb{H}$-perimeter measure. Given $f : \mathbb{H}^1 \to \mathbb{R}$ we denote its distributional horizontal gradient by $\nabla_{\mathbb{H}} f := (Xf, Yf)$; we say that a continuous function f is of class $\mathbf{C}^1_{\mathbb{H}}$ if $\nabla_{\mathbb{H}} f$ is represented by a continuous function. Given a measurable set

$E \subset \mathbb{H}^1$, the $\mathbb{H}$-*perimeter measure* $P_E^{\mathbb{H}}$ of E is defined as the total variation of $\nabla_{\mathbb{H}}\chi_E$, or equivalently by

$$P_E^{\mathbb{H}}(\Omega) := \sup\left\{\int_E \operatorname{div}_{\mathbb{H}} g : g \in \mathbf{C}_c^\infty(\Omega, \mathbb{R}^2), |g| \leqslant 1\right\}$$

where $\operatorname{div}_{\mathbb{H}} g := Xg_1 + Yg_2$. If E has smooth boundary, then $P_E^{\mathbb{H}} = \mathcal{H}^3_{\mathbb{H}} \llcorner \partial E$, where $\mathcal{H}^m_{\mathbb{H}}$ is the m-dimensional Hausdorff measure associated with d_c; in some sense, we have to think of the dimension $3 = 4 - 1$ as the one associated with objects of codimension 1. Sets with locally finite Euclidean perimeter have locally finite $\mathbb{H}$-perimeter, while the converse is not true: sets with finite $\mathbb{H}$-perimeter have typically a fractal boundary.
In analogy with the classical case, for sets with finite $\mathbb{H}$-perimeter it is possible to define the $\mathbb{H}$-*reduced boundary* $\mathcal{F}_{\mathbb{H}}E$; the basic result in this sense is provided by the following rectifiability theorem proved by B. Franchi, R. Serapioni and F. Serra Cassano in 2001:

Theorem 6 *If $E \subset \mathbb{H}^1$ has finite $\mathbb{H}$-perimeter, then $P_E^{\mathbb{H}} = \mathcal{H}^3_{\mathbb{H}} \llcorner \mathcal{F}_{\mathbb{H}}E$ and $\mathcal{F}_{\mathbb{H}}E$ is $\mathbb{H}$-rectifiable, i.e.*

$$\mathcal{F}_{\mathbb{H}}E \ \subset\ \mathcal{N} \cup \bigcup_{i=1}^{\infty} S_i$$

where S_i are $\mathbf{C}^1_{\mathbb{H}}$ hypersurfaces and $\mathcal{H}^3_{\mathbb{H}}(\mathcal{N}) = 0$.

By considering $\mathbb{H}$-perimeter instead of classical perimeter it is possible to define $\mathbb{H}$-perimeter minimizing sets and to consider the associated problem of the regularity of their boundaries: very few is known in this sense. There are examples of $\mathbb{H}$-perimeter minimizing sets whose reduced boundary is not of class $\mathbf{C}^\infty$. What lacks is essentially the implication

$$\mathcal{F}_{\mathbb{H}}E \text{ of class } \mathbf{C}^{1,\alpha}_{\mathbb{H}} \quad \Longrightarrow \quad \mathcal{F}_{\mathbb{H}}E \text{ of class } \mathbf{C}^\infty.$$

It is not known whether $\mathcal{F}_{\mathbb{H}}E$ has to be at least of class $\mathbf{C}^{1,\alpha}_{\mathbb{H}}$.
Regarding the problem of the singular set, there are also examples of $\mathbb{H}$-perimeter minimizing sets E for which $\partial E \setminus \mathcal{F}_{\mathbb{H}}E$ is not empty. In the known examples $\mathcal{H}^1_{\mathbb{H}}(\partial E \setminus \mathcal{F}_{\mathbb{H}}E) \neq 0$; on the other hand it is possible to show that $\mathcal{H}^3_{\mathbb{H}}(\partial E \setminus \mathcal{F}_{\mathbb{H}}E) = 0$. An open problem is to give sharp estimates on the size of $\partial E \setminus \mathcal{F}_{\mathbb{H}}E$.

A Bernstein type problem has been formulated in $\mathbb{H}^1$ too; since we do not intend to enter into excessive technicalities, we will present it very briefly. Let us consider the subgroups $\mathbb{V} := \exp(X) \equiv \mathbb{R}$ and $\mathbb{W} := \exp(\operatorname{span}\{Y, T\})$, which is also normal and maximal. $\mathbb{H}^1$ is the semidirect product $\mathbb{W} \ltimes \mathbb{V}$. The *intrinsic graph* of a function $\phi : \mathbb{W} \to \mathbb{V}$ is

$$\operatorname{gr}\phi := \{P \cdot \phi(P) : P \in \mathbb{W}\}.$$

There is an integral formula for computing the measure $\mathcal{H}^3_{\mathbb{H}}(\operatorname{gr}\phi)$ of the graph of ϕ, and correspondingly there is a *minimal surface equation* ($\mathbb{H}$MSE) for intrinsic graphs. If $\operatorname{gr}\phi$ is a maximal subgroup (or cosets), then ϕ is a global solution of ($\mathbb{H}$MSE): these are in some sense analogous to affine Euclidean functions parametrizing hyperplanes (i.e. cosets of maximal subgroups). The Bernstein problem in $\mathbb{H}^1$ admits a positive answer: there exist global solutions of ($\mathbb{H}$MSE) which are not "affine". We would like to mention that

there exist also higher dimensional Heisenberg groups $\mathbb{H}^n \equiv \mathbb{R}^{2n+1}$ where the Bernstein problem could be formulated: "non-affine" solutions exist if $n \geqslant 5$, while the problem is open for $n = 2, 3, 4$.

Finally, the isoperimetric problem can be formulated also in $\mathbb{H}^1$, and is probably the most important open problem in this setting. Which is the set with smallest $\mathbb{H}$-perimeter among those with fixed volume? The most "naive" conjecture fails: balls in the Carnot-Carathodory metric are not isoperimetric. Isoperimetric sets have been conjectured by P. Pansu in 1983 and are provided by the already mentioned sets $\mathcal{I}_{P,s}$. However, a complete solution has not been achieved yet and only some partial results are known:

- M. Ritorè - C. Rosales (2005): if the isoperimetric set had boundary of class $\mathbf{C}^2$, then it would be $\mathcal{I}_{P,s}$;
- R. Monti - M. Rickly (2006): if the isoperimetric set were convex, then it would be $\mathcal{I}_{P,s}$;
- R. Monti (2008): if the isoperimetric set were axially symmetric around the t-axis, then it would be $\mathcal{I}_{P,s}$.

References. A good introductory reference to Geometric Measure Theory is [7]; for a more comprehensive study we address the reader to [5]. The monograph [6] is a very clear reference for minimal surfaces and the Bernstein problem; the same holds for [2, 4] regarding the isoperimetric problem. The books [3] and [8] can be used for an introduction to the Heisenberg group and related geometric problems.

References

[1] E. Bombieri, E. De Giorgi & E. Giusti, *Minimal Cones and the Bernstein Problem.* Inv. Math. 7 (1969), 243–268.

[2] Y. D. Burago & V. A. Zalgaller, "Geometric inequalities". Springer-Verlag, 1988.

[3] L. Capogna, D. Danielli, S. D. Pauls & J. T. Tyson, "An Introduction to the Heisenberg Group and the Sub-Riemannian Isoperimetric Problem". Birkhäuser, 2007.

[4] I. Chavel, "Isoperimetric inequalities. Differential geometric and analytic perspectives". Cambridge Tracts in Mathematics, 145. Cambridge University Press, Cambridge, 2001.

[5] H. Federer, "Geometric Measure Theory". Springer, 1969.

[6] E. Giusti, "Minimal surfaces and functions of bounded variation". Birkhäuser, 1984.

[7] F. Morgan, "Geometric measure theory. A beginner's guide (fourth edition)". Academic Press Inc., 2008.

[8] D. Vittone, Submanifolds in Carnot groups, Tesi di Perfezionamento, Scuola Normale Superiore, Pisa (2008).

(De)-Localization of some (1+1)-dimensional models

Martin Borecki (*)

Abstract. We consider a (1+1)-dimensional model, i.e. a directed model for a linear chain which is described by its configurations $\{(n, \varphi_n)\}_{0 \leq n \leq N}$. The chain is randomly distributed in space and undergoes an interaction with the environment and itself. Thus, it can be seen as a so called random polymer and we want to study its spatial distribution as a function of its length and its interaction parameters. The self-interaction consists of a Gradient and Laplacian mixture type, whereas the interaction with the environment will be reduced to a δ-pinning, i.e. the chain gets a reward $\varepsilon \geq 0$ each time it touches the x-axis. We discuss the localization behavior of the model which is substantially different, depending on the parameters α, β and $\varepsilon \geq 0$. Furthermore we consider what changes, if we additionally introduce an impermeable wall.

1 Motivation and introduction

We start our motivation from the point of view of polymers. These are well known objects to chemists, biologists and physicists. Typical examples are : plastic, rubber or soap and some more complicated biopolymers like: cellulose, filaments or DNA. In such a variety of different objects there is a big interest of investigating the interaction behavior. One could imagine for example the bundling of two (or more) polymers, which in the case of filaments would be important for the stability of the cytoskeleton of living cells. Another interaction of this kind is the phenomenon of denaturation of DNA. A second kind of interaction is the adsorption of a polymer onto a planar substrate, for instance a polymer attracted by a membrane.

In order to develop a mathematical model for polymers we simplify the point of view and think of a chain-like molecule built up from monomers which are connected by strong covalent bonds. By monomers we mean in our case only one kind of molecules. Furthermore two essential probabilistic properties will be needed to obtain a model for the so called random polymer. Namely, random distribution in space and interaction of the

(*)Ph.D. school "IRTG Stochastic Models of Complex Processes Berlin-Zurich" - Technische Universität Berlin, Institut für Mathematik, MA 7-4, Strasse des 17 Juni 136, 10623 Berlin (Germany). E-mail: borecki@math.TU-Berlin.DE . Seminar held on 14 January 2009.

polymer with itself and the environment.

2 The model

We consider the following pinning and wetting model, which is given by the spatial distribution of the chain $\varphi = \{\varphi_i\}_{1 \le i \le N-1}$:

$$\mathbb{P}^{(+)}_{\varepsilon,N}(d\varphi_1, ..., d\varphi_{N-1}) \ := \frac{\exp(-\mathcal{H}_N(\varphi))}{\mathcal{Z}^p_{\varepsilon,N}} \prod_{i=1}^{N-1} (d\varphi_i^{(+)} + \varepsilon\delta_0(d\varphi_i)) \ ,$$

where

- $\varepsilon \ge 0 \rightarrow$ pinning parameter
- $\delta_0(.)$ the Dirac mass at zero
- $d\varphi_i^{(+)}$ the Lebesgue measure on $\mathbb{R}$ or $\mathbb{R}_+$
- $\mathcal{Z}^{(+)}_{\varepsilon,N}$ the normalization constant (partition function)
- $\mathcal{H}_N$ Hamiltonian responsible for self-interaction .

The interaction with the environment is reduced to a δ-pinning at the x-axis. Therefore the chain is rewarded by touching this defect-line. The self-interaction for $\alpha, \beta > 0$ is chosen as follows

$$\begin{aligned} \mathcal{H}_N(\varphi_1, ..., \varphi_{N-1}) : &= \mathcal{H}_{[-1,N+1]}(\varphi_{-1}, ..., \varphi_{N+1}) \\ &= \frac{\alpha}{2} \sum_{i=1}^{N+1} (\nabla\varphi_i)^2 + \frac{\beta}{2} \sum_{i=0}^{N} (\Delta\varphi_i)^2 \end{aligned}$$

Here we take zero boundary conditions $\varphi_{-1} = \varphi_0 = \varphi_N = \varphi_{N+1} = 0$ and ∇, Δ denote of course the discrete gradient and Laplacian.

We remark that this model undergoes two opposite effects, the entropy and the energy represented by the self-interaction, the wall and the pinning. Therefore we ask

Q1: Is the energy large enough to pin the chain at the x-axis ?

Q2: Is there a critical energy (phase transition) ?

It is known that the model behaves dramatically different for the two extreme cases $\alpha = 0$ and $\beta = 0$. The natural question arises, if there are some critical values for α or β where the model changes its behavior. For this we need to clarify how the behavior can be studied in dependence of the parameters $\alpha, \beta, \varepsilon$ and N.

3 Localization-delocalization

To capture the competing behavior of the fluctuations of the free model and the interaction with the environment as $N \to \infty$ we introduce the following term.

Definition 1 (Free energy)

$$F(\varepsilon) := \lim_{N\to\infty} \frac{1}{N} \log\left(\frac{\mathcal{Z}_{\varepsilon,N}}{\mathcal{Z}_{0,N}}\right) \quad , \ \varepsilon \geq 0$$

It is well defined by a super-additivity argument and has the obvious, but important property :

$$\mathcal{Z}_{\varepsilon,N} \geq \mathcal{Z}_{0,N} \qquad \Longrightarrow \qquad F(\varepsilon) \geq F(0) = 0 \ .$$

One could ask whether there exists an ε such that F is positive. Indeed this is true and in this case it is possible to show that the average paths touch the defect line a positive fraction of time. Whereas, if F equals 0, then there are only $o(N)$ contacts. With that we can define

Definition 2 (Delocalization-localization) *The polymer measure $\mathbb{P}_{\varepsilon,N}$ is called*

- *delocalized, if* $\varepsilon \in \mathcal{D} := \{\varepsilon \geq 0 \,|\, F(\varepsilon) = 0\}$
- *localized, if* $\varepsilon \in \mathcal{L} := \{\varepsilon \geq 0 \,|\, F(\varepsilon) > 0\}$
- $\varepsilon_c := \sup_{\varepsilon\in\mathcal{L}} \varepsilon$ *(critical point)* .

4 Main results

We were able to prove a localization result which shows the "very strong" impact of the ∇-interaction.

Theorem 3 (Localization for the pinning case) *For every $\alpha, \beta > 0$ the model $\mathbb{P}^p_{\varepsilon,N}$ exhibits a trivial phase transition $\varepsilon_c = 0$, i.e.*

$$\mathcal{D} = \{0\} \quad \text{and} \quad \mathcal{L} = (0, \infty) \ .$$

Furthermore, on $\mathcal{L}$ the free energy is real analytic and

$$F^p(\varepsilon) \sim \log \varepsilon \quad , \varepsilon \to \infty \ .$$

We want to state also a conjecture in the wetting case.

Theorem 4 (Conjecture on wetting) *For every $\alpha, \beta > 0$ there exists $\varepsilon_c^+ > 0$, such that the model $\mathbb{P}^+_{\varepsilon,N}$ reveals the following delocalization-localization behavior:*

$$\mathcal{D} = [0, \varepsilon_c^+] \quad \text{and} \quad \mathcal{L} = (\varepsilon_c^+, \infty) \ .$$

We see that this model behaves like the ∇-model, which was investigated by [1], [3], [6], [7]. Also even a big impact of the Laplacian part, $\beta \gg 0$, has no influence in the change on localization behavior. We don't want to give the proof here, but it uses renewal technics from [4]. Instead we turn to an interesting extension of our model.

5 Extension to higher dimensions

We would like to extend the results from the $(1+1)$-dimensional case to the $(1+d)$-dimensional one. Depending on a pinning subspace of $\mathbb{R}^d$, the problem can be reduced to that of $(1+1)$-dimension, due to the choice of quadratic potential.

The following model is a generalization of our $(1+1)$-dimensional model for a linear chain with pinning and wetting. On the one hand the generalization is motivated by [2], where they introduce a pinning measure ν and certain pinning subspaces. So let's consider the chain, given by the polymer measure on $(\mathbb{R}^d)^{N-1}$:

$$\mathrm{P}^{(+)}_{\varepsilon,N,d,m}(d\varphi_1,...,d\varphi_{N-1}) \ := \frac{\exp(-\mathcal{H}_N(\varphi))}{\mathcal{Z}^{(+)}_{\varepsilon,N,d,m}} \prod_{i=1}^{N-1} \left(\varepsilon\nu(d\varphi_i^{(+)}) + d\varphi_i^{(+)}\right) \ ,$$

with

- $\varepsilon \geq 0 \rightsquigarrow$ the pinning parameter
- $+$ denotes the model with an additional wall, i.e. $\varphi_n \in \mathbb{R}^{d-1} \times \mathbb{R}_+$ for all n
- $d\varphi_i^{(+)}$ the Lebesgue measure on $\mathbb{R}^d$ or $\mathbb{R}^{d-1} \times \mathbb{R}_+$ respectively
- $\mathcal{Z}^{(+)}_{\varepsilon,N,d}$ the normalization constant (partition function)
- the Hamiltonian
$$H_N(\varphi) = \frac{\alpha}{2} \sum_{i=1}^{N+1} ||\nabla\varphi_i||_2^2 + \frac{\beta}{2} \sum_{i=0}^{N} ||\Delta\varphi_i||_2^2$$
with zero boundary conditions $\varphi_{-1} = \varphi_0 = \varphi_N = \varphi_{N+1} = 0$.
- Let M be an m-dimensional subspace of $\mathbb{R}^d$ for $0 \leq m \leq d-1$ (for the model with the wall we take $M \subset \partial(\mathbb{R}^{d-1} \times \mathbb{R}_+)$)
- the pinning measure on $\mathbb{R}^d$ is $\nu(dy) := dy^{(1)}\delta_0(dy^{(2)})$, where $dy^{(1)}$ is the Lebesgue measure on M and the second measure is the Dirac mass at 0 on $M^\perp$

Now with no loss of generality we take $r \in \{1, ..., d\}$ and

$$M = \{x \in \mathbb{R}^d \mid x_d = x_{d-1} = ... = x_{d-(r-1)} = 0\} \ .$$

We set $m := \dim(M) = d - r$ and it can be proven that

$$\mathcal{Z}^{(+)}_{\varepsilon,N,d,m} = (\mathcal{Z}_{0,N,1,0})^m \mathcal{Z}^{(+)}_{\varepsilon,N,d-m,0} \ .$$

It is an important observation to prove the following result:

Theorem 5 (Localization in higher dimensions)

- *The model exhibits a non-trivial phase transition for pinning and wetting, if* $r \geq 3$ *and*
- *trivial phase transition for pinning, if* $1 \leq r \leq 2$.

References

[1] E. Bolthausen, *Localization-Delocalization Phenomena for Random Interfaces.* ICM 2002, Vol. III, 1-3 (2002).

[2] E. Bolthausen, T. Funaki, T.Otobe. Probab. Theory Relat. Fields **143** (2009), pp. 441-480.

[3] F. Caravenna, G. Giacomin and L. Zambotti, *Sharp Asymptotic Behavior for Wetting Models in (1+1)-dimension.* Elect. J. Probab. **11** (2006), pp. 345-362.

[4] F. Caravenna, J.-D. Deuschel, *Pinning and Wetting Transition for (1+1)-dimensional Fields with Laplacian Interaction.* preprint (2007).

[5] F. Caravenna, J.-D. Deuschel, *Scaling Limits of (1+1)-dimensional Pinning Models with Laplacian Interaction.* preprint (2008).

[6] J.-D. Deuschel, G. Giacomin and L. Zambotti, *Scaling Limits of Equilibrium Wettin Models in (1+1)-dimension.* Probab. Theory Rel. Fields **132** (2005), pp. 471-500.

[7] Y. Isozaki and N. Yoshida, *Weakly Pinned Random Walk on the Wall: Pathwise Descriptions of the Phase Transitions.* Stoch. Proc. Appl. **96** (2001), pp. 261-284.

[8] M. Zerner, *Quelques propriétés spectrales des opérateurs positifs.* J. Funct. Anal. **72** (1987), pp. 381-417.

An introduction to p-adic analysis

Valentina Di Proietto (*)

Abstract. The p-adic numbers were discovered by Hensel at the end of nineteenth century and in the last century they came to a central role in number theory. In the first part we shall give the definition of the field of p-adic numbers and present some results in elementary p-adic analysis always comparing with results on classical analysis over the real numbers; in the second part we shall describe an example which explains a basic case of the result proven in our Ph.D thesis.

1 Motivation

Let us fix a prime number p. The field of p-adic numbers, that it is denoted by $\mathbb{Q}_p$, is a field which contains $\mathbb{Q}$ in which it is relatively simple to find roots of polynomials.
One of the typical problems in number theory is the following: taken a polynomial

$$F(X_1, \dots, X_n) \in \mathbb{Q}[X_1, \dots, X_n],$$

has F roots in $\mathbb{Q}$?
The hope is what is called local-global principle:
The existence of a rational root of F can be established by studying the existence of a root of F in $\mathbb{Q}_p \ \forall\ p$ and in $\mathbb{R}$, the field of real numbers. In some particular cases this hope becomes a theorem; what follows is called theorem of Hasse-Minkowski.

Theorem 1 *If $F(X_1, \dots, X_n)$ is an homogeneous polynomial of degree 2, then it has a non trivial root in $\mathbb{Q}$ if and only if it has a non trivial root in $\mathbb{Q}_p$, $\forall\ p$ and in $\mathbb{R}$.*

The local-global principle is false for a general $F(X_1, \dots, X_n)$; one example is the following:

$$F(X, Y, Z) = 3X^3 + 4Y^3 + 5Z^3$$

has a non trivial root in the field of real numbers and in all the p-adic fields, but it has no non trivial root in $\mathbb{Q}$ [Se].

(*)Dottorato di ricerca in Matematica Pura, Dipartimento di Matematica Pura ed Applicata, Università di Padova. E-mail: proietto@math.unipd.it. Seminar held on 28 January 2009.

2 Absolute values on fields

Definition 1 *Let K be a field, an absolute value is a map $|\ |: K \to \mathbb{R}_{\geq 0}$ such that*

1) $|x| = 0 \Leftrightarrow x = 0$,

2) $|xy| = |x||y|$,

3) $|x + y| \leq |x| + |y|$.

The absolute value is said to be non archimedean if it satisfies

3') $|x + y| \leq \max\{|x|, |y|\}$.

Obviously an absolute value which verifies 3') verifies 3), but the opposite is false. We are interested in absolute values over the field $\mathbb{Q}$; let us see some examples.

Example 1 We can define an absolute value taking

$$|x| = 1 \ \ \forall x \neq 0$$
$$|0| = 0.$$

This absolute value is non archimedean, it is called the trivial absolute value and we will avoid this case in what follows.

Example 2 We now examine the classical euclidean absolute value, denoted by $|\ |_\infty$ and defined as

$$|x|_\infty = x \ \ \forall x \geq 0$$
$$|x|_\infty = -x \ \ \forall x < 0.$$

It is archimedean, which means that $|\ |_\infty$ verifies 1), 2), 3), but it does not verifies 3').

Now we describe the main example which we are interested in.

Example 3 Let us fix p a prime number and ε a real number such that $0 < \varepsilon < 1$, we define the p-adic absolute value which we denote by $|\ |_p$ in the way that follows: if $x \in \mathbb{Q}$ and $x \neq 0$, than we can write

$$x = p^\alpha \frac{a}{b} \text{ with } \alpha \in \mathbb{Z},\ p \nmid a\ p \nmid b$$

and we define

$$|x|_p = \varepsilon^\alpha$$

and

$$|0|_p = 0.$$

This is the p-adic absolute value and one can easily verify that it is non archimedean.

An absolute value $|\ |$ on a field K induces a metric: the distance between two points x and y in K is defined in the following way:

$$d(x, y) := |x - y|.$$

Two absolute values on K are equivalent if they induce the same topology on K. There is a famous theorem, due to Ostrowski, which classifies the absolute values over the field of rational numbers.

Theorem 2 *Every non trivial absolute value on $\mathbb{Q}$ is equivalent to one of the absolute value $|\ |_p$, where p is a prime number, or to the euclidean absolute value $|\ |_\infty$.*

So examples 1, 2, 3 are the only possible examples of absolute values over $\mathbb{Q}$.

3 Properties of the non archimedean absolute value

In this section we will see some topological properties of a field K endowed with a non archimedean absolute value $|\ |$.

Lemma 1 *Let K be a field and $|\ |$ a non archimedean absolute value on K. If $x, y \in K$ are such that $|x| \neq |y|$, then*

$$|x + y| = \max\{|x|, |y|\}.$$

Proof. By hypothesis we know that $|x| \neq |y|$, so we can suppose that $|x| > |y|$. So by the non archimedean inequality we have that $|x + y| \leq |x|$. But on the other hand

$$|x| = |x + y - y| \leq \max\{|x + y|, |y|\} = |x + y|.$$

□

This simple lemma has a funny corollary.

Corollary 1 *In a non archimedean space all triangles are isosceles.*

Proof. Let us suppose that we have a triangle with vertices $x, y, z \in K$. So the lengths of the edges are $|x - y|, |y - z|, |x - z|$. Let us suppose that two of them are different ($|x - y| \neq |y - z|$), so

$$|x - z| = |x - y + y - z| = \max\{|x - y|, |y - z|\}.$$

□

Now we look at the balls. If $a \in K$ and $r \in \mathbb{R}_{\geq 0}$

$$\mathrm{B}(a, r^-) := \{x\ \in K|\ |x - a| < r\}$$

is called the open ball of center a and radius r and

$$\mathrm{B}(a, r^+) := \{x\ \in K|\ |x - a| \leq r\}$$

is called the closed ball of center a and radius r. Obviously open balls are open sets and closed balls are closed sets with respect to the topology induced by the absolute value, but in the case of a non archimedean space strange things happen.

Proposition 1 *If K is a non archimedean field, the following holds.*

(i) *If $b \in \mathrm{B}(a, r^-)$ (resp. $b \in \mathrm{B}(a, r^+)$), then $\mathrm{B}(a, r^-) = \mathrm{B}(b, r^-)$ (resp. $\mathrm{B}(a, r^+) = \mathrm{B}(b, r^+)$). In other words every point contained in an open (resp. closed) ball is the center of the open (resp. closed) ball.*

(ii) *$\mathrm{B}(a, r^-)$ is both open and closed.*

(iii) *If $r \neq 0$, then $\mathrm{B}(a, r^+)$ is both open and closed.*

(iv) *If a, $b \in K$, r,$s \in \mathbb{R}^*_{\geq 0}$ and $\mathrm{B}(a, r^-) \cap \mathrm{B}(b, s^-) \neq \emptyset$ (resp. $\mathrm{B}(a, r^+) \cap \mathrm{B}(b, s^+) \neq \emptyset$), then $\mathrm{B}(a, r^-) \subset \mathrm{B}(b, s^-)$ (resp. $\mathrm{B}(a, r^+) \subset \mathrm{B}(b, s^+)$)or $\mathrm{B}(b, s^-) \subset \mathrm{B}(a, r^-)$ (resp. $\mathrm{B}(b, s^+) \subset \mathrm{B}(a, r^+)$).*

Proof. Look at [Go, proposition 2.3.6]. □

4 Analytic properties of p-adic numbers

We still did not defined what a p-adic number is. The field of p-adic numbers is constructed in the same way as the field of real numbers. Let us recall the construction of $\mathbb{R}$. We start from $\mathbb{Q}$ and construct a field $\mathbb{R}$ which contains $\mathbb{Q}$ such that

- the absolute value $|\ |_\infty$ extends to $\mathbb{R}$,
- $\mathbb{R}$ is complete with respect to the metric given by the extension of the absolute value $|\ |_\infty$,
- $\mathbb{Q}$ is dense in $\mathbb{R}$.

The field $\mathbb{R}$ is called the completion of $\mathbb{Q}$ for the absolute value $|\ |_\infty$. It is a basic theorem in topology the existence and the uniqueness of the completion.

Definition 2 *The completion of $\mathbb{Q}$ with respect to the p-adic absolute value is called the field of p-adic numbers and denoted by $\mathbb{Q}_p$.*

There is a concrete description of p-adic numbers.

Theorem 3 *Any p-adic number α can be written in the form*

$$\alpha = \sum_{j=n}^{\infty} a_j p^j \tag{1}$$

where $a_j \in \mathbb{Z}$ and n is such that $|\alpha|_p = \varepsilon^n$. The integer coefficients are unique modulo p. If $0 \leq a_j \leq p-1$, then the expression in (1) is called the canonical expansion.

Proof. Look at [Ba, theorem 2.1]. □

In [Ba, pp. 36-37] one can find the algorithm to write a rational number in the form given by (1).

We want to do analysis over $\mathbb{Q}_p$; this can be vary complicated, in fact as we saw before the topology induced by the norm on is very strange. Moreover the following holds.

Proposition 2 *The field $\mathbb{Q}_p$ is totally disconnected.*

Proof. Let us take A a subset of $\mathbb{Q}_p$, which contains x and y. Then A is disconnected, since if we fix $\delta = \frac{1}{2}d(x,y)$, we can write A as the disjoint union of the sets A_1 and A_2, where $A_1 = \mathrm{B}(x,\delta) \cap A$ and $A_2 = A - A_1$. The set A_1 is open and closed in A and the same is true for A_2. □

Let us look now at Cauchy sequences. In $\mathbb{Q}_p$ there is a very simple criterion to established if a sequence is Cauchy or not.

Lemma 2 *The sequence $x_n \in \mathbb{Q}_p$ is Cauchy with respect to the p-adic absolute value $|\ |_p$ if and only if*

$$\lim_{n\to\infty} |x_{n+1} - x_n|_p = 0.$$

Proof. Let us suppose that $m = n + r$, then $|x_m - x_n|_p = |x_{n+r} - x_{n+r-1} + x_{n+r-1} + \cdots + x_{n+1} - x_n|_p \leq \max\{|x_{n+r} - x_{n+r-1}|, \ldots, |x_{n+1} - x_n|\}$. □

The analogous of lemma 2 over the real numbers with the euclidean absolute value is obviously false because the sequence $x_n = (-1)^n$ clearly satisfies that

$$\lim_{n\to\infty} |x_{n+1} - x_n|_\infty = 0,$$

but it is not a Cauchy sequence because it does not converge.
Lemma 2 has as consequence a surprising convergence criterion for series of p-adic numbers.

Proposition 3 *The series*

$$\sum_{n=0}^{\infty} a_n,$$

with $a_n \in \mathbb{Q}_p$, is convergent if and only if

$$\lim_{n\to\infty} |a_n|_p = 0.$$

Moreover if the series converges, then $|\sum_{n=0}^{\infty} a_n|_p \leq \max_n |a_n|_p$.

Proof. A series converges if the sequence of partial sum converges, but a_n is the difference between the n-th term of the sequence of partial sum and the $(n-1)$-th term of the sequence of partial sum. Being a_n infinitesimal we can conclude, thanks to lemma 2, that the series converges. □

As for lemma 2, the analogous of proposition 3 over the real numbers with the euclidean absolute value is false, because the series $\sum_n \frac{1}{n}$ does not converge with respect to the euclidean absolute value, but its n-th term is infinitesimal.
Let us look now at power series; as for series of p-adic numbers the convergent criterion is much more simple that in the case of real or complex numbers.

Proposition 4 *For a power series*

$$f(x) = \sum_{n=0}^{\infty} a_n x^n,$$

where $a_n \in \mathbb{Q}_p$*, we define*

$$\rho := \frac{1}{\limsup \sqrt[n]{|a_n|_p}}.$$

The following hold:

(i) *if* $\rho = 0$*,* $f(x)$ *converges only for* $x = 0$*,*

(ii) *if* $\rho = \infty$*,* $f(x)$ *converges* $\forall\ x \in \mathbb{Q}_p$*,*

(iii) *if* $0 < \rho < \infty$ *and* $\lim_{n\to\infty} |a_n|_p \rho^n = 0$*,* $f(x)$ *converges if and only if* $x \leq \rho$*,*

(iv) *if* $0 < \rho < \infty$ *and* $\lim_{n\to\infty} |a_n|_p \rho^n \neq 0$*,* $f(x)$ *converges if and only if* $x < \rho$*.*

Proof. Look at [Go, proposition 4.2.1]. □

In the following example we study the convergence of the exponential power series on $\mathbb{Q}_p$ and we will see that, contrary to what happens over the complex numbers, it has a very small radius of convergence.

Example 4 Let us consider the power series

$$\exp(x) := \sum_n \frac{x^n}{n!};$$

we want to see for which value of $x \in \mathbb{Q}_p$ it converges. Thanks to calculations which can be founded in [DGS, pag. 50], one can prove that

$$|n!|_p = \varepsilon^{\frac{n-S(n)}{p-1}}$$

if $n = a_0 + a_1 p + \cdots + a_l p^l$, with $0 \leq a_i \leq p-1$ and $S(n) = a_0 + a_1 + \cdots + a_l$.
Hence the $p-adic$ absolute value of the n-th term of the series is

$$\left|\frac{x^n}{n!}\right|_p = |x|_p^n \varepsilon^{-\frac{n-S(n)}{p-1}}.$$

Let us suppose that $|x|_p < \varepsilon^{\frac{1}{p-1}}$, then $|x|_p \leq \varepsilon^{\frac{1+c}{p-1}}$ for $c > 0$ and

$$\left|\frac{x^n}{n!}\right|_p \leq \varepsilon^{\frac{n(1+c)}{p-1}} \varepsilon^{-\frac{n-S(n)}{p-1}}.$$

Because $nc + S(n) \to \infty$ for $n \to \infty$,

$$\left|\frac{x^n}{n!}\right|_p \to 0$$

and $\exp(x)$ converges for $|x|_p < \varepsilon^{\frac{1}{p-1}}$.
If $|x|_p = \varepsilon^{\frac{1}{p-1}}$ for $n = p^m$, $S(n) = 1$, then

$$\left|\frac{x^n}{n!}\right|_p = \varepsilon^{\frac{n}{p-1}} \varepsilon^{-\frac{n-1}{p-1}} = \varepsilon^{\frac{1}{p-1}},$$

which does not converge to 0. So we can conclude that $\exp(x)$ converges if and only if $|x|_p < \frac{1}{p-1}$.

The fact that $\mathbb{Q}_p$ is disconnected is a problem for a good theory of holomorphic functions. The only possible approach seems to be via power series expansions. A reasonable definition for a locally analytic function is the following: a function $f : U \to \mathbb{Q}_p$, where U is an open set, is locally analytic if it admits a convergent power series expansion around every point of $x \in U$.
Let us choose $U = \mathrm{B}(0, 1^+)$ and as f the function defined in the following way:

$$f \equiv 0 \text{ on } \mathrm{B}(0, 1^-)$$
$$f \equiv 1 \text{ on } \Gamma(0, 1) = \{x \in \mathbb{Q}_p |\ |x|_p = 1\}.$$

Both $\mathrm{B}(0, 1^-)$ and $\Gamma(0, 1)$ are open and f fulfills the definition that we gave of locally analytic function, but we do not want to think to f as analytic. The solution to this problem is in what is called rigid analytic geometry studied by Tate, Bosch, Raynaud, Berkovich, Huber, and others. The general principle is that one consider power series which admit a convergent power series expansion globally. Thank to this theory one can construct over $\mathbb{Q}_p$ a theory of analytic varieties, which are called rigid analytic varieties, analogous to the theory of analytic varieties over the complex numbers.

5 Logarithmic extensions for differential equations

Before defining what a differential equation is in the p-adic world, we recall something about classical differential equations.
For more details on the contents of this subsection the interested reader can look at [Bl], [De], [Ke], [Ma], [Sh].
Let us consider the field $\mathbb{C}((t))$; a derivation is a map

$$d : \mathbb{C}((t)) \to \mathbb{C}((t))\frac{dt}{t},$$

where dt is a symbol, such that

$$d(f+g) = df + dg$$

$$d(fg) = fdg + gdf \text{ for } f, g \in \mathbb{C}((t))$$

and

$$d(c) = 0 \quad \forall c \in \mathbb{C}.$$

We will consider as d the map that acts on $f \in \mathbb{C}((t))$ on the following way:

$$d : f \mapsto t\frac{df}{dt}dt.$$

Associated to the map d there is a natural differential equation: we ask for which $Y \in \mathbb{C}((t))$ the following is verified

$$dY = 0 \Leftrightarrow t\frac{dY}{dt} = 0.$$

A generalization of this concept is what is called a module with connection; we fix (E, ∇) where E is a finite dimensional vector space over $C((t))$ and ∇ is a $\mathbb{C}$-linear map

$$\nabla : E \to E\frac{dt}{t}$$

which verifies

$$\nabla(v+w) = \nabla(v) + \nabla(w),$$

$$\nabla(fv) = f\nabla(v) + vdf \ \forall v, w \in E, \ f \in \mathbb{C}((t)).$$

As for the case of derivation we can associate to a couple (E, ∇) a differential equation. We fix a basis of E over $\mathbb{C}((t))$ in such a way that $E = \oplus_{i=1}^{n}\mathbb{C}((t))e_i$ and we suppose that the action of the connection ∇ on $e = (e_1, \ldots, e_n)$ is given by

$$\nabla(e) = eG(t)\frac{dt}{t},$$

where $G(t)$ is an $n \times n$ matrix with coefficients in $\mathbb{C}((t))$. The associated differential equation is given by the function $Y \in \mathbb{C}((t))$ such that

$$\nabla(Ye) = 0 \Leftrightarrow YeG(t)\frac{dt}{t} + edY = 0$$

$$\Leftrightarrow t\frac{dY}{dt} = -G(t)Y.$$

We are interested in understanding the conditions for which, given a module with connection (E, ∇) with E a vector space of dimension n over $\mathbb{C}((t))$, there exists a module with connection $(\tilde{E}, \tilde{\nabla})$ where $\tilde{E}$ is a $\mathbb{C}[[t]]$ free module of rank n and $\tilde{\nabla} : \tilde{E} \to \tilde{E}\frac{dt}{t}$ a connection that induces the given connection ∇ on E.
The answer in this case is a condition that it is call regular singularity. We fix a norm $|\ |$

which corresponds to the norm on $\mathbb{C}((t))$ induced by the variable t and we consider the following map:

$$\partial : E \to E\frac{dt}{t} \to E$$

which is the composite of the map given by the connection ∇ and the isomorphism that sends $\frac{dt}{t}$ in 1; we say that (E, ∇) is regular singular if

$$|\partial|_{sp} := \lim |\partial^n|^{\frac{1}{n}} = |d|_{sp} = 1.$$

For example the differential equation given by $t\frac{d}{dt}Y = 0$ is regular singular and the equation $t\frac{d}{dt}Y = \frac{1}{t}Y$ is not.
In a more general setting we can consider X a smooth variety over $\mathbb{C}$ and an open $U \subset X$ such that $X - U = D$ is a divisor with normal crossing. There is a notion of module with connection on varieties which generalizes the notion that we gave (see for example [De], [Ka]). We suppose to have a module with connection (E, ∇) defined on U and we ask when there exists a module with connection $(\tilde{E}, \tilde{\nabla})$ on X with logarithmic singularities along D such that it extends (E, ∇). The answer is a result by Deligne [De (II, proposition 5.4)] and deals with a generalization of the notion of regular singularity.
In the p-adic case instead of considering $\mathbb{C}((t))$ we look at the functions on what is called the annulus of radii $0 \leq \lambda < 1$ and 1, i.e. we consider the Laurent power series $\sum_n a_n t^n$ with a_n in $\mathbb{Q}_p$ such that $\lim |a_n|_p \rho^n = 0$ for every $\rho \in [\lambda, 1)$. If we denote by $A_{\lambda,1}$ the ring of the function with this convergent condition, we can define as before what is a module with connection on $A_{\lambda,1}$: it will be a couple (E, ∇) with E an $A_{\lambda,1}$ free module and $\nabla : E \to E\frac{dt}{t}$ with the same properties as before.
The question parallel to the one asked before is the following: given (E, ∇) a module with connection on $A_{\lambda,1}$ with $\lambda > 0$, is there a $(\tilde{E}, \tilde{\nabla})$ module with connection on $A_{0,1}$ which extends (E, ∇)?
Again in this case the answer deals with a p-adic analogous of the notion of regular singularity, a condition that is called Robba condition [ChMe]. In particular the result is that a module with connection on $A_{\lambda,1}$ which satisfies the Robba condition (and some other conditions that are called Non-Liouville) extends to a module with connection on $A_{0,1}$ ([BlCh], [ChMe]).
A part of the work done in our thesis deals with the problem of extending some special types of modules with connection defined on arbitrary dimensional rigid analytic varieties [DP].

References

[Ba] G. Bachman, "Introduction to p-adic numbers and valuation theory". Academic Press, New York-London, 1964.

[Bl] F. Baldassarri, *Towards an algebraic proof of Deligne's regularity criterion. An informal survey of open problems.* Milan Jou. Math. **73** (2005), 237-258.

[BlCh] F. Baldassarri, B. Chiarellotto, *Algebraic versus rigid cohomology with logarithmic coefficients.* In Barsotti Symposium in Algebraic Geometry, Academic Press (1994), 11-50.

[ChMe] G. Christol, Z. Mebkhout, *Équations différentielles p-adiques et coefficients p-adiques sur les courbes.* Astérisque **279** (2002), 345-410.

[De] P. Deligne, "Équations différentielles à point singuliers réguliers". Lecture Notes in Mathematics, Springer-Verlag, 1970.

[DP] V. Di Proietto, On p-adic differential equations on semistable varieties, Ph.D Thesis, Università Degli Studi di Padova, 2009.

[DGS] B. Dwork, G. Gerotto, F. J. Sullivan, "An introduction to G-functions". Annals of Mathematics Studies, Princeton University Press, 1994.

[Go] F. Q. Gouvêa, "p-adic numbers. An introduction". Universitext, Springer-Verlag, Berlin, 1993.

[Ka] N. M. Katz, *Nilpotent connection and the monodromy theorem: an application of a result of Turrittin.* Pub. Math. IHES, **39** (1970), 175-232.

[Ke] K. S. Kedlaya, *Semistable reduction for overconvergent F-isocrystals,I: Unipotence and logarithmic extensions.* Compositio Math., **143** (2007), 1164-1212.

[Ma] J. I. Manin, *Moduli Fuchsiani.* Ann. Sc. Norm. Sup. Pisa, (3), **19** (1965), 113-126.

[Se] E. S. Selmer, *The Diophantine equation* $ax^3 + by^3 + cz^3 = 0$. Acta Math. **85** (1951), 203–362.

[Sh] A. Shiho, *On logarithmic extension of overconvergent isocrystals.* arXiv:0806.4394v1, 2008.

Solving Mixed Integer Programs with Gomory Cutting Planes

Arrigo Zanette (*)

1 Introduction

Gomory cutting planes were first introduced by Gomory in 1958 to solve Integer and Mixed Integer ('60) Programs (MIP). However they were soon abandoned in favor of enumeration tecniques, until, in 1996, they were revisited by Balas et al., becoming a fundamental tool for commercial MIP solver. Despite their long history and relative success, the lack of understanding on their practical behaviour makes Gomory Cutting Planes an interesting research topic. In particular it is clear that they might perform much better than current implementations have managed to do, but nobody has found the right way of using them yet. In the present we will review Gomory cutting planes, their typical usage in commercial MIP solvers and recent research findings that might eventually lead to a new performance breakthrough.

2 The mathematical model

A *Mixed Integer linear Program* (MIP) is defined as follows:

$$
\begin{array}{rrl}
z_{MIP} = & \min & c^T x \\
& \text{subject to} & Ax \leq b \\
& & l \leq x \leq u \\
& & x \in \mathbb{R}^n \\
& & x_j \in \mathbb{Z} \quad \forall j \in I
\end{array}
\tag{1}
$$

where $c \in \mathbb{Q}^n$ is the cost vector, $A \in \mathbb{Q}^{m \times n}$ is the constraints matrix, $b \in \mathbb{Q}^m$ the right-hand side, $l \in (\mathbb{Q} \cup \{-\infty\})^m$ and $u \in (\mathbb{Q} \cup \{\infty\})^m$ the bounds on variables, typically treated apart from the rest of constraints. Variables constrained to integrality are indexed in $I \subseteq N = \{1, \ldots, n\}$. Related to the matrix notation in which it is written, constraints

(*)Dottorato di ricerca in Matematica Computazionale, Dipartimento di Matematica Pura ed Applicata, Università di Padova. E-mail: zanettea@math.unipd.it . Seminar held on 11 February 2009.

are also called *rows* and variables *columns*. Integer variables x_j with bounds $0 \leq x_j \leq 1$ are called *binary* variables and typically have a different semantic than other integer variables, called *general integer*. Indeed, binary variables are typically used to model the common situation in which yes-or-no decisions arise. Any vector x satisfying constraints in (1) is called *feasible* solution. We will denote by z^*_{MIP} the optimal integer solution of the MIP problem.

Geometrically speaking, constraints (1) define the set X obtained intersecting the polyhedron $P_{LP} = \{x \in \mathbb{R}^n_+ : Ax \leq b\}$ and $\mathbb{Z}^{|I|}$. The *integer hull* of P_{LP} is the convex hull of points in X, $\text{conv}(X)$. A fundamental result of Meyer [11] is that the integer hull of the rational polyhedron P_{LP} is again a rational polyhedron, P_{MIP}. Namely, $\text{conv}(X)$ is given by the intersection of a finite number of inequalities defining P_{MIP}. So, provided that the polyhedral description of $\text{conv}(X)$ is known, integrality constraints become redundant and the problem simple [13]. However, the polyhedral description is hard to find in general. This fact supports the theoretical importance of cutting plane algorithms (see Section 4), in which $\text{conv}(X)$ is iteratively constructed by adding valid inequalities for $\text{conv}(X)$ to P_{LP}.

3 Linear programming

Discarding the integrality constraints in (1), the MIP model becomes a Linear Programming (LP) model. LP can be solved efficiently, for example, by using the Simplex Method invented by Dantzig in 1947.

The simplex algorithm performs computations on the problem data arranged in a matrix, called *tableau*. The first row of the tableau, row 0, stores the coefficients of the objective function. Other rows store the constraints matrix A. The first column, column 0, is reserved for the right-hand-side values.

A *basis* B is a square invertible matrix obtained by selecting a subset of m columns (i.e., *basic* variables). The unique solution of the linear system obtained by setting to zero all non-basic variables is called *basic solution*.

The simplex algorithm iterates through basic solutions, by exchanging one column in the basis (*leaving variable*) with another outside the basis (*entering variable*): this operation is called *pivot*.

The choice of entering and leaving variables can be done by following different measures of improvement. Indeed, the algorithm comes into two main variants. On one hand, the *primal* simplex method visits only feasible solutions by iteratively improving the objective function value, until the optimum is eventually reached. On the other hand, the *dual* simplex method visits only more-than-optimal solutions until it eventually finds a feasible solution.

The twofold nature of the algorithms relies on the LP *strong duality*, that is, there exists a dual version of the original (primal) problem, defined in a dual space, which will attain the same optimal value as the primal problem—provided such optimum exists and is finite (refer to [2] for a treatment of linear programming duality).

When the measure of improvement fails the algorithm can cycle. This is due to either primal or dual degeneracy. In case of primal degeneracy, there are many different bases

corresponding to the same basic solution; in case of dual degeneracy, there are different basic solutions with the same cost. In both the situations, some countermeasure is required to avoid cycling. The first proposed solution was to use a lexicographic simplex algorithm [7]. Heuristic alternatives were then developed to address the computational burden of the lexicographic simplex. The reader is referred to [14] for more details on these techniques, and on the practical implementation of the simplex method as well.

4 Cutting planes

If the LP relaxation does not suffice to solve the MIP problem, one may try to strengthen the relaxation by adding new inequalities and solving the LP relaxation again. As mentioned in Section 2, theory [11] suggests that iterating this process one could eventually find the MIP optimum, since MIP is a polyhedron. This technique is called *cutting planes method*.

A *cutting plane* is an inequality valid for P_{MIP} but violated by at least one point of P_{LP}. A way of deriving a valid inequality for P_{MIP} is by aggregating rows in P_{LP} with some non-negative multipliers $u \in \mathbb{R}^n_+$:

$$u^T A x \leq u^T b$$

Since x is non-negative the inequality obtained by rounding down the coefficients of the left-hand-side

$$\lfloor u^T A \rfloor x \leq u^T b \tag{2}$$

is again a valid cut for P_{LP}. Now, since $x \in X$ is integer the strengthened cut:

$$\lfloor u^T A \rfloor x \leq \lfloor u^T b \rfloor \tag{3}$$

is a valid inequality for X. Cuts derived in this way are called Chvátal-Gomory cuts. Chvátal proved in [5] that taking successive *closures* of this simple kind of cuts, i.e., the (undominated) cuts derived from all the possible multipliers u in the separation phase, suffices to ensure convergence. However, it is not clear how to derive a violated cut, given a point x^*.

Gomory addressed that issue many years before, in his seminal work [8], showing that violated Chvátal-Gomory cuts can be read directly from the fractional rows of the current LP optimal tableau. In such a case, u appearing in (2) corresponds to the entries of a row of the optimal basis inverse. The Gomory procedure is the following. Starting from the Tableau row i

$$x_i + \sum_{j \in N \setminus B} a_{ij} x_j = a_{i0} \tag{4}$$

we can relax it to inequality $x_i + \sum_{j \in N \setminus B} a_{ij} x_j \leq a_{i0}$. Applying the Chvátal-Gomory rounding procedure we get the *Integer Gomory* cut:

$$x_i + \sum_{j \in N \setminus B} \lfloor a_{ij} \rfloor x_j \leq \lfloor a_{i0} \rfloor \tag{5}$$

Subtracting (5) from (4), and denoting by f_{ij} the fractional part of a_{ij} yields

$$\sum_{j \in N \setminus B} f_{ij} x \geq f_{i0}. \tag{6}$$

(6) is called *Fractional Gomory* cut (FGC).

Provided that the MIP model has an integer objective function, that is the vector c^T has only integer coefficients, Gomory gave a proof of convergence that relies on the lexicographic simplex method for getting rid of degeneracy in the LP relaxations. The lexicographic simplex objective is to minimize (or maximize) the solution vector formed by the objective function z, interpreted as the most important lexicographic variable, followed by the other variables. This is equivalent to solve the original LP amended with the following objective function:

$$\min M^n x_0 + M^{n-1} x_1 \ldots + M x_{n-1} + x_n \tag{7}$$

where M is a sufficiently large positive integer, and $x_0 = z$. Now let us suppose x_j is the first fractional variable, according to the lexicographic order. The key point of the first proof is that a Gomory cut derived from the row of the current lexicographically optimal tableau where x_j is basic, will increment the value of the objective function (7) by, at least, $M^{n-j}(\lceil x_j \rceil - x_j)$. Thus, assuming without loss of generality that (7) is bounded, the number of iterations of Gomory's algorithm is finite.

4.1 Fortune of Gomory cutting planes

After the first successes (see [4], for example) it was observed quite soon [9] that a pure cutting plane algorithm, suffers from an early *tailing off* phenomenon: large sequence of cuts are added without any significant improvement towards integrality. Due to this observation, pure cutting plane methods were abandoned in practice in favor of explicit enumerative techniques.

Nowadays, Gomory cutting planes are commonly used inside MIP solvers to strengthen the problem formulation. Their revival is due to Balas et al. [3], in which they report a breakthrough in the solution of binary MIP problems using rounds of GMI cuts embedded in a enumerative code. Since then, GMI cuts have become a common feature in all MIP solvers.

In addition to Gomory cutting planes, other cuts are among the arsenal of state-of-the-art MIP solvers. These cutting planes do not have a general applicability, but were developed by studying the polyhedral structure of specific problems such as knapsack, fixed-charge flow, vertex packing/covering. Solvers try to automatically identify these common substructures in general instance and to add the appropriate cuts. A comprehensive description can be found in [10].

5 Towards a computationally sound pure cutting plane algorithm

In [15] the authors addressed some important issues arising when designing a computationally sound pure cutting plane method.

Indeed, while modern MIP solver are based on the Branch&Cut [1], that mixes explicit enumeration and cutting planes, pure cutting plane algorithms, based on Gomory cuts alone, are typically not used in practice, due to their poor convergence properties.

In a sense, branching can be viewed as just a "symptomatic cure" to the well-known drawbacks of Gomory cuts—saturation, bad numerical behavior, etc. From the cutting plane point of view, however, the cure is even worse than the disease, in that it hides the real source of the troubles.

The purpose of work [15] is to try to come up with a viable pure cutting plane method (i.e., one that is not knocked out by numerical difficulties), even if on most problems it will not be competitive with the branch-and-bound based methods.

To this end, the authors review the theoretical properties of Gomory cutting planes. In particular Gomory used the lexicographic simplex algorithm to deal with degeneracy proving the converge of his cutting plane algorithm.

Actually, dual degeneracy is massive and unavoidable in Integer Programming. Moreover, common heuristics successfully used to cope with degeneracy solving a single LP are unable to overcome degeneracy solving an LP described by a set of dynamically separated constraints (cuts), as required by cutting plane algorithms. Only the lexicographic simplex algorithm, together with the right-paired Gomory cuts, is able to ensure a monotone, non-degenerate progression towards the final integer optimum.

However, unlike stated in the original paper [8], it turns out that the lexicographic simplex algorithm is not just a useful tool to simplify the proof of convergence, but it has strong practical implications too. Indeed, the effects of uncontrolled dual degeneracy are made clear by a detailed analisys on a small test problem. The authors show that iterating a wrong choice of the cut-generating basic solution can soon lead to very numerically unstable cuts.

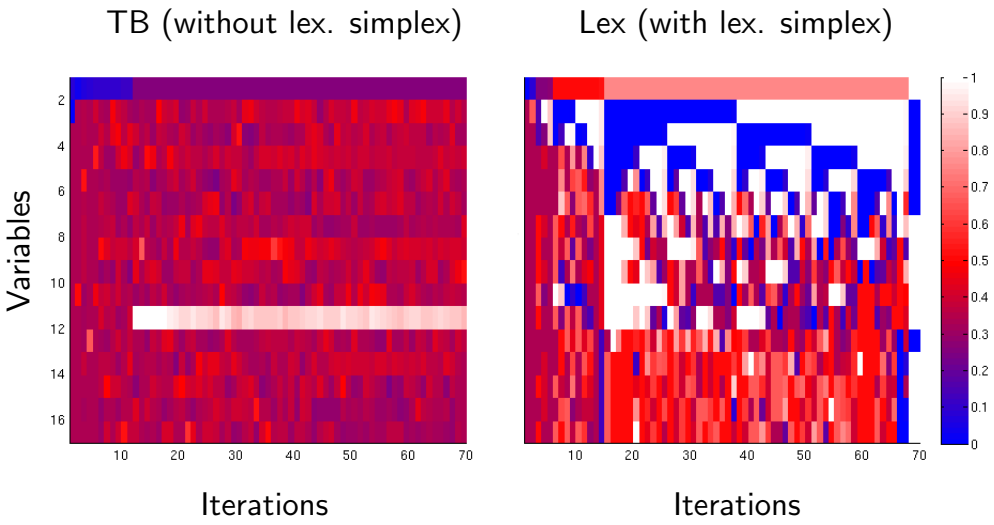


Figure 1. Fractional spectrography of the sequence of solutions provided by the Gomory cutting plane method with (Lex) and without (TB) lexicographic reoptimization (single-cut version on instance `stein15`).

Figure 1 gives a graphical representation of the fractional solutions visited during the optimization of the MIPLIB instance `stein15`. This kind of graphical representation has been introduced in [15] and called *fractional spectrography* because it depicts the spectrum of fractionalities along the execution of the algorithm. A fractional spectrography is a pseudo-color plot of the 3-dimensional data set formed by iterations, variables, and variable values. Iterations are represented in the horizontal axis, variables in the vertical axis, while variable values are represented in a color scale. Variables are listed in their lexicographic order: variable 0 is the objective function, variable 1 is the lexicographically first variable, and so on. For a problem with only binary variables, the white color encodes a 1 and the blue color a 0. Nonbinary variables as, for example, the objective function, are normalized into the interval $[0, 1]$. All red shadings between blue and white correspond to values between the lower and upper bounds. This coding helps giving a visualization of the degree of fractionality of a solution.

In Figure 1, the left chart shows the behavior of an implementation of the textbook Gomory method that generates one cut a time (as in the Gomory proof of convergence) but does not use the lexicographic method. It is clear that, very soon, the method starts visiting fractional solutions that are closer and closer one to each other, until the chart assumes an almost uniform red shading indicating almost no change in the fractional solution components.

To confirm their findings, the authors give an efficient implementation of the lexicographic simplex algorithm and compare the results obtained using a "textbook" implementation of the pure Gomory cutting plane algorithm, and our implementation, based of the lexicographic simplex. Their implementation turned out to outperform the textbook implementation, solving to optimality instances that have never been solved by pure cutting planes.

They finally extend our analisys to Mixed Integer Gomory cuts. They propose an adjustable approximate version of MIG cuts, that allow one to separate a set of cuts, ranging from Fractional Gomory cuts to MIG cuts, simply by setting a parameter appropriately. Preliminary results, show that a tighter approximation of MIG cuts leads to larger cut coefficients and, eventually, to numerical problems.

5.1 The enumerative nature of Gomory cutting planes

In [6] the work [15] is extended considering the enumerative nature of Gomory's method.

This interpretation was first introduced by Nourie and Venta [12], mainly as a theoretical tool to give an elegant proof of Gomory cutting plane algorithm convergence.

The search tree implicitly enumerated by Gomory's algorithm can be read from the right (Lex) part of Figure 1. Figure 2 gives a clearer pictorial intuition of this fact, overlaying part of the search tree on the corresponding fractional spectrography plot.

Unlike [12], in [6] the authors point out the practical importance of the enumerative interpretation: since Gomory's method is in fact cast in its enumerative framework, its performance can only be improved if one gets rid of the rigidity of the enumerative shell.

They therefore propose and computationally analyze new versions of the Gomory method, that borrow from native enumerative schemes some of their main features. In particular, they analyze cutting plane algorithms where the objective function is treated

implicitly, and/or the lexicographic order is redefined on the fly to mimic a sound branching strategy. Preliminary computational results seem to indicate that the new methods have some potentials.

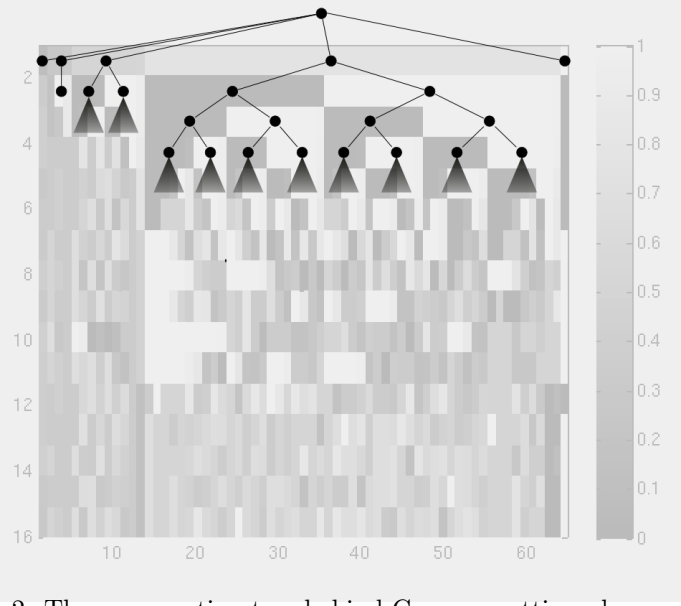


Figure 2. The enumeration tree behind Gomory cutting plane method.

Future research should try to put even more flexibility in pure cutting plane methods. In doing so, authors expect that a better understanding of the interaction between cuts and enumeration will be gained, with a positive follow-up for the MIP solvers side too.

References

[1] M. W. Padberg and G. Rinaldi, *A branch-and-cut algorithm for the resolution of large-scale symmetric traveling salesman problems.* SIAM Rev. 33 (1991), 60–100.

[2] A. Bachem and W. Kern, "Linear Programming Duality. An Introduction to Oriented Matroids". Number 074 in Universitext. Springer, 1992.

[3] E. Balas, S. Ceria, G. Cornujols, and N. Natra j., *Gomory cuts revisited.* Operations Research Letters 19 (1996), 1–9.

[4] E. M. L. Beale, *Survey of integer programming.* OR, 16(2) (jun 1965), 219–228.

[5] V. Chvátal, *Edmonds polytopes and a hierarchy of combinatorial problems.* Discrete Mathematics 4 (1973, 305–337.

[6] A. Zanette, E. Balas, M. Fischetti, *On the enumerative interpretation of Gomorys dual cutting plane method.* Technical report (2009).

[7] A. Orden, G. B. Dantzig and P. Wolfe, *The generalized simplex method for minimizing a linear form under linear inequality restraints.* Pacific J. Math. 5(2) (1955), 183–195.

[8] R. E. Gomory, *Outline of an algorithm for integer solutions to linear programs.* Bulletin of the American Society 64 (1958), 275–278.

[9] R. E. Gomory, *Early integer programming.* Operations Research 50 (Jan.-Feb. 2002), 78–81.

[10] H. Marchand, A. Martin, R. Weismantel, and L. Wolsey, *Cutting planes in integer and mixed integer programming.* Discrete Applied Mathematics, 123(1-3) (November 2002), 397-446.

[11] R. R. Meyer, *On the existence of optimal solutions to integer and mixed integer programming problems.* Mathematical Programming 7 (1974), 223–235.

[12] F.J. Nourie and E.R. Venta, *An upper bound on the number of cuts needed in Gomorys method of integer forms.* Operations Research Letters 1 (1982), 129–133.

[13] A. Schrijver, "Theory of linear and integer programming". John Wiley & Sons, Inc., New York, NY, USA, 1986.

[14] T. Terlaky, *Book review: Computational techniques of the simplex method.* Computational Optimization and Applications 26(2), 209–210.

[15] A. Zanette, M. Fischetti, and E. Balas, *Can pure cutting plane algorithms work?.* In IPCO (2008), 416–434.

An Invitation to Frobenius Manifolds

Luca Philippe Mertens (*)

Abstract. Frobenius manifold provide a geometric formulation of WDVV equations and their "hidden" integrable structure. They were introduced by B. Dubrovin to study the remarkable connection between integrable systems, 2D topological field theories and Gromov-Witten invariants of symplectic manifolds discovered by E. Witten and M. Kontsevich. The goal of these notes is to give an overview of the main ideas and definitions of Frobenius manifolds and integrable hierarchies, and to present some recent developments of the theory related to infinite dimensional Frobenius manifolds.

Sunto. Le varietà di Frobenius forniscono un' interpretazione geometrica delle equazioni WDVV e della loro struttura integrabile. Esse furono introdotte da B. Dubrovin per studiare la connessione tra sistemi integrabili, teorie di campo topologiche due dimensionali e invarianti di Gromov-Witten di varietà simplettiche scoperta da E. Witten e M. Kontsevich. Lo scopo di queste note è fornire un'introduzione alle principali idee e definizioni della teoria delle varietà di Frobenius e delle gerarchie integrabili, e di presentare alcuni recenti sviluppi della teoria riguardanti le varietà di Frobenius infinito dimensionali.

1 Introduction

Roughly speaking, Frobenius manifolds are manifolds endowed with a flat metric η and a Frobenius algebra structure on the tangent space. They were introduced by B. Dubrovin [9] to investigate the relationship between integrable systems and 2D topological field theories. The remarkable connection between these previously unrelated fields of mathematics and physics was firstly discovered by E. Witten and M. Kontsevich for the case of the KdV hierarchy [15].

The central notion of the theory is that of *potential* of the Frobenius manifold. This is a locally defined function F on the manifold from which one can recover the coefficients of the metric and the structure constants of the Frobenius algebra by taking the second and third order derivatives of F.

Once a system of flat coordinates $\{v^\alpha\}_{\alpha\in I}$ is chosen, one can define an infinite set of mutually commuting non linear PDEs of evolutionary type in the "time" variables $\{t^{\alpha,n}\}$ where $\alpha \in I$ and $n \in \mathbb{N}$. This integrable system is called the *principal hierarchy* of the

(*)SISSA - Via Beirut 2–4 - 34014 Trieste (Italy); e-mail: mertens@sissa.it - Seminar held on 25 February 2009.

Frobenius manifold. The potential of the Frobenius manifold is the restriction to the *small phase space* $\left\{t^{\alpha,0}\right\}_{\alpha\in I}$ of the logarithm of a *tau* function(*) of the principal hierarchy [6].

The key observation of Witten was that in the KdV case the logarithm of this τ function coincides with the generating function of the intersection numbers of the so called tautological classes in $\mathbf{H}^*(\bar{\mathcal{M}}_{g,n})$, the ϵ expansion corresponding to the genus expansion of the generating function. Frobenius Manifolds turned out to be a key tool in the study of sophisticated topological objects. For example, Frobenius Manifolds appear as quantum cohomologies of smooth projective varieties encoding the structure of genus zero Gromov-Witten invariants.

Summary. In the first two sections we recall the basic notions and definitions of Frobenius manifolds and integrable systems. In section 3 we describe the classical example of the Frobenius manifold associated to the Toda hierarchy. Finally, in the last section we present an example of infinite dimensional Frobenius manifold.

2 What is a Frobenius manifold?

Definition 2.1 *A* Frobenius structure $(\mathcal{M},\eta,\cdot,e,E)$ *of charge* d *on a manifold* $\mathcal{M}$ *is given by:*

- *A non degenerate symmetric* $(0,2)$ *tensor*(*) η. *We require* η *to be flat in the usual metric sense (vanishing of the curvature tensor).*
- *An associative, commutative algebra structure with unity* $(\mathbf{T}\mathcal{M},\cdot)$ *over the tangent space. Let* c *be the 3-tensor defined by* $c(x,y,z):=\eta(x,y\cdot z)$ *for* x,y,z *in* $\mathbf{T}\mathcal{M}$. *We require* c *to be symmetric in* x,y,z, *and* $\nabla_w c(x,y,z)$ *to be symmetric in* x,y,z,w.
- *A covariantly constant vector field* e, *which is the unity of the algebra structure.*
- *An affine vector field* E, *i.e.* $\nabla\nabla E=0$, *called the* Euler vector field. *The Euler vector field acts conformally on the structure coefficients of the product and the metric:*

$$\begin{aligned} [E,x\cdot y]-[E,x]\cdot y-x\cdot[E,y] &= x\cdot y \\ E<x,y>-<[E,x],y>-<x,[E,y]> &= (2-d)<x,y> \end{aligned}$$

The symmetry of c is equivalent to the requirement that η is *invariant* with respect to the product: $\eta(a\cdot b,c)=\eta(a,b\cdot c)$ for every $a,b,c\in\mathbf{T}\mathcal{M}$. For every $\mathbf{t}\in\mathcal{M}$ we have that $(\mathbf{T}_{\mathbf{t}}\mathcal{M},\cdot,\eta)$ is a *Frobenius algebra*: a finite dimensional associative unital algebra equipped with a symmetric non degenerate invariant bilinear form.

Let $\{t^\alpha\}_{\alpha\in I}$ be a set of local flat coordinates. The requirement that ∇c is a symmetric tensor implies that locally $c_{\alpha,\beta,\gamma}=\nabla_\alpha\nabla_\beta\nabla_\gamma F$. The function F is called the *potential*

(*) A τ function corresponds to a solution of the hierarchy depending on all the time variables $t^{\alpha,n}$.

(*) We will refer to it as a metric, although we don't require η to be positive definite.

of the Frobenius manifold. In the flat coordinates t^α, the *associativity equations* for the structure coefficients of the algebra give the following set of non linear PDEs for F:

$$\frac{\partial^3 F}{\partial t^\alpha \partial t^\beta \partial t^\lambda}\eta^{\lambda,\mu}\frac{\partial^3 F}{\partial t^\mu \partial t^\gamma \partial t^\delta} = \frac{\partial^3 F}{\partial t^\gamma \partial t^\beta \partial t^\lambda}\eta^{\lambda,\mu}\frac{\partial^3 F}{\partial t^\mu \partial t^\alpha \partial t^\delta} \qquad \alpha,\beta,\gamma,\delta \in I \tag{2.1}$$

These are the celebrated WDVV equations [9].

Theorem 2.2 *A Frobenius manifold structure is locally equivalent to the existence of a potential F satisfying the WDVV equations and the two properties:*

Normalization: $F_{1,\lambda,\mu} = \eta_{\lambda,\mu}$ *where* $\frac{\partial}{\partial v^1} = e$ *is a marked coordinate*

Conformal structure: $\mathcal{L}_E F = (3-d)F +$ *(quadratic terms)*

WDVV equations firstly appeared in the physics literature as a set of equations to describe the partition function $F(\mathbf{t})$ of a 2D TCFT (topological conformal field theory) together with its deformations preserving topological invariance [15]. Following the axiomatic formulation of a TFT (topological field theory) given by M.F. Atiyah [2], one can establish an equivalence of categories between 2D TFT and commutative Frobenius algebras [11]. For any TCFT one can construct a canonical moduli space of deformations parametrized by the *coupling constants* $\mathbf{t} = (t^\alpha, t^\beta, t^\gamma, \ldots)$ of the theory. The function $F(\mathbf{t})$ is the partition function of the theory. In this framework, one can think of a Frobenius manifold as a moduli space of 2D topological field theories. $\mathcal{M}$ is the space of parameters of the model, and for every fixed $\mathbf{t} \in \mathcal{M}$ the Frobenius algebra $(\mathbf{T_t}\mathcal{M}, \cdot, \eta)$ describes the corresponding perturbation of the theory.

3 What is an integrable hierarchy?

Quoting an introduction of N. Hitchin [10] on the subject:

> Integrable systems, what are they? The question [...] can be dismissed as Louis Amstrong is reputed to have done once when asked what jazz was - "If you gotta ask, you'll never know!"

Let $\mathcal{A} := \mathbb{C}\left[u^i, u^i_x, u^i_{xx}, \ldots\right]$ be the ring of differential polynomials in the u^i. This ring has a natural differential ∂_x, we define the formal integration as the cokernel of the map ∂_x. Following Dickey [5], we write an evolutionary PDE as $\partial_t u^i = P_i(u^i, u^i_x, u^i_x x, \ldots)$, where $P_i \in \mathcal{A}$ for every $i \in I$. We define an integrable hierarchy to be a system of compatible non linear PDEs of evolutionary type for a set of functions $\{u_i(x)\}_{i\in I}$ depending on a formal "space variable" x. Each equation of the hierarchy describes the evolution with respect to a formal "time variable" t. The fundamental requirement is that all these differential equations are compatible: $\partial_{t_1}\partial_{t_2}u_i - \partial_{t_2}\partial_{t_1}u_i = 0$ for every $i \in I$.

A τ *function* is a solution of an auxiliary differential equation, the Hirota equation, giving a solution of the whole integrable hierarchy. More precisely one can define the functions $u_i(x, \mathbf{t})$ in terms of logarithmic derivatives of the τ function, and these functions u_i will satisfy all the equations of the hierarchy.

Poisson structures

We define the *formal loop space* $\mathcal{L}(\mathcal{M})$ of $M = \{u^i\}$ to be the space of local functionals $\bar{P} := \int P dx$, where $P \in \mathcal{A}$. One may think of it as the space of formal maps from the unit circle to $\mathcal{M}$. Locally, a point of the loop space will be represented by a set of functions $u_i(x)$, where x is the independent variable on the unit circle.

Given a flat controvariant metric $g^{i,j}$ over $\mathcal{M}$ with Christoffel symbols $\Gamma_k^{i,j}$ one can define a Poisson bracket of Dubrovin-Novikov type over $\mathcal{L}(\mathcal{M})$:

$$\{\bar{P}, \bar{Q}\} := \int \frac{\delta P}{\delta u^i(x)} \{u^i(x), u^j(y)\} \frac{\delta Q}{\delta u^j(y)} dy \tag{3.2}$$

where: $\{u^i(x), u^j(y)\} = g^{i,j}\delta'(x-y) + \Gamma_k^{i,j} u^k \delta(x-y)$, while $\frac{\delta P}{\delta u^i(x)}$ is the (formal) variational derivative. The hierarchy possess a *Hamiltonian structure* when there exists a Poisson bracket $\{\quad,\quad\}$ such that each flow of the hierarchy can be written in the form $\partial_t u_i = \{\bar{H}, u_i\}$. The local functional $\bar{H}$ is called the *Hamiltonian* of the flow, while the differential polynomial H inducing it its *Hamiltonian density*. We are interested in integrable hierarchies possessing a bihamiltonian structure. A *bihamiltonian structure* is the data of two Poisson structures which are compatible, i.e. $\{\quad,\quad\}_\lambda := \{\quad,\quad\}_2 - \lambda\{\quad,\quad\}_1$ is a Poisson bracket for every $\lambda \in \mathbb{C}$.

Dispersionless limit

The *dispersionless limit* is a procedure one can perform to obtain an auxiliary system of quasilinear PDEs. This system is usually easier to study, and gives information on the behavior of the solutions of the original *dispersive* hierarchy. Roughly speaking, what one has to do is to replace $t \mapsto \epsilon t, x \mapsto \epsilon x$ and then take the limit $\epsilon \to 0$. Note that one can perform the limit procedure to all the machinery introduced so far (local functionals, Hamiltonian structures,...). The algebraic way to do this is to extend the above theory to the ring of formal power series $\mathcal{A}[[\epsilon]]$. The construction is to consider the ring of formal power series in ϵ with coefficients in the ring of differential polynomials [6].

From Frobenius Manifolds to Integrable Hierarchies

There is a canonical way to construct a bihamiltonian integrable hierarchy on the loop space of a Frobenius Manifold $\mathcal{L}(\mathcal{M})$:

- The bihamiltonian structure is given by two controvariant flat metrics. The first controvariant metric is given by dualizing η to the cotangent space. The second, called the *intersection form*, is defined by $g^{i,j} = E^k c_k^{i,j}$. One can prove that the two Poisson structures are compatible.
- To construct the Hamiltonian densities we have introduce the *deformed flat connection*. This is a controvariant connection ∇_λ depending on a complex parameter λ: $(\nabla_\lambda)_u v = \nabla_u v + \lambda u \cdot v$. A *horizontal section* is a function $f(u,\lambda) \in C^\infty(\mathcal{M} \times \mathbb{CP}^1)$ such that $\nabla_\lambda df = 0$. We take a basis of horizontal sections f^α and we consider their

power series expansion in λ, $f^\alpha = \sum_{k\geq 0} \theta^{\alpha,k}\lambda^k$. The family of functions $\left\{\theta^{\alpha,k}\right\}$ is the set Hamiltonian densities of the principal hierarchy.

All the Hamiltonians $H^{\alpha,k} := \oint \theta^{\alpha,k}dx$ mutually commute, i.e. their induced flows are compatible. This integrable system is called the *principal hierarchy*.

Remark 3.3 Note that the time variables of the principal hierarchy are labeled by two indices, an index α running over the set of flat coordinates and an integer index $k \geq 0$. This very simple observation will be crucial in the last section.

Under the additional assumption of *semi simplicity* of the Frobenius manifold, one can prove that the principal hierarchy is complete: every Hamiltonian commuting with all the Hamiltonians of the principal hierarchy can be written as a finite liner combination of them. Finally, under the further assumption of *quasi triviality* of the Frobenius Manifold one can reconstruct the full dispersive hierarchy starting from the dispersionless one [6].

4 A classical example

We consider the manifold $\mathbb{C}^2$ with potential, Euler vector field and identity given by:

$$F(u,v) = \frac{1}{2}v^2u + e^u \qquad E = v\partial_v + 2\partial_u \qquad e = \partial_v$$

This is a Frobenius manifold with $\eta(\partial_u, \partial_v) = 1$, 0 otherwise. The induced product is $\partial_u \cdot \partial_u = e^u\partial_u$, $\partial_v \cdot \partial_u = \partial_u$, $\partial_v \cdot \partial_v = \partial_v$.

The associated integrable hierarchy is the extended Toda hierarchy [8]. Introduce the Lax symbol: $\mathcal{L} = p + v + \frac{e^u}{p}$. Then:

$$\partial_{t^{u,n}}\mathcal{L} = \frac{1}{(n+1)!}\left\{(\mathcal{L}^n)_+, \mathcal{L}\right\} \tag{4.3}$$

$$\partial_{t^{u,n}}\mathcal{L} = \frac{2}{n!}\left\{(\mathcal{L}^n \ln(\mathcal{L} - c_n))_+, \mathcal{L}\right\} \tag{4.4}$$

where $(\sum_{k\in\mathbb{Z}} f_k p^k)_{\geq 0} = \sum_{k\geq 0} f_k p^k$ is the projection on the positive part of the symbol, $\{f,g\} := p\frac{df}{dp}\frac{dg}{dx} - p\frac{dg}{dp}\frac{df}{dx}$ is the canonical Poisson bracket, and $c_n = \sum_{k=1}^n \frac{1}{k}$.

Gromov-Witten Interpretation.

The logarithm of the τ function given by the topological solution is the generating function for the GW invariants of $\mathbb{CP}^1$. This function depends on two sets of times $\{t^{u,n}, t^{v,n}\}_{n\in\mathbb{N}}$. The restriction of τ to $t^{u,0}, t^{v,0}$ coincides with the potential. The flat coordinates u, v of the Frobenius manifold give a basis of the quantum cohomology $H^*(\mathbb{CP}^1)$, where $v = 1 \in H^0(\mathbb{CP}^1)$, $u = \omega \in H^2(\mathbb{CP}^1)$.

5 Infinite dimensional Frobenius manifolds

It is a well known fact in the field of integrable systems that the Toda hierarchy introduced in the previous section may be seen as a reduction of a larger integrable system called the 2D Toda hierarchy. We recall the Lax formulation of the dispersionless 2D Toda hierarchy:

Definition 5.4 *Let $\mathcal{L}(p), \bar{\mathcal{L}}(p)$ be two formal power series in p^{-1}, p defined by $\mathcal{L}(p) = \sum_{k=-1}^{\infty} u_{-k} p^{-k}$, $\bar{\mathcal{L}}(p) = \sum_{k=-1}^{\infty} \bar{u}_k p^k$ where $u_1 = 1, \bar{u}_{-1} \neq 0$. The dispersionless 2D Toda hierarchy is described by the following set of quasilinear PDEs for the functions $u_i(x), \bar{u}_i(x)$:*

$$\begin{aligned} \partial_{t_n} \mathcal{L} &= \{\mathcal{A}_n, \mathcal{L}\} \quad , \quad \partial_{\bar{t}_n} \mathcal{L} = \{\mathcal{B}_n, \mathcal{L}\} \\ \partial_{t_n} \bar{\mathcal{L}} &= \{\mathcal{A}_n, \bar{\mathcal{L}}\} \quad , \quad \partial_{\bar{t}_n} \bar{\mathcal{L}} = \{\mathcal{B}_n, \bar{\mathcal{L}}\} \end{aligned}$$

Where $\mathcal{A}_n(p) := (\mathcal{L}^n(p))_{<0}$ and $\mathcal{B}_n(p) := (\bar{\mathcal{L}}^n(p))_{\geq 0}$, while $\{f, g\} := p\frac{df}{dp}\frac{dg}{dx} - p\frac{dg}{dp}\frac{df}{dx}$ is the canonical Poisson bracket.

This hierarchy was introduced by K. Takasaki and T. Takebe [13]. Quoting Takasaki [14]:

> A very intriguing open problem is to interpret the unreduced dispersionless KP and Toda hierarchies as a kind of infinite dimensional Frobenius manifolds. This will require a drastic modification of the notion of Frobenius manifold

In the previous sections we have seen how to associate a bihamiltonian integrable system to a Frobenius manifold. Clearly we can proceed the other way round, and ask ourself if a given integrable system can be constructed starting from a Frobenius manifold. According to Remark 3.3, the existence of an (infinite dimensional) Frobenius manifold associated to the 2D Toda hierarchy would imply the existence of a bigger hierarchy extending the one given in Definition 5.4. This infinite dimensional Frobenius manifold was eventually constructed by G. Carlet, B. Dubrovin and the author [4].

Definition 5.5 *Let $\mathcal{M}_0$ be the set of functions on the closed unitary disk(*) that are holomorphic on the punctured disc and have a simple pole at 0. Let $\mathcal{M}_\infty$ be the set of functions on the the complementary disk (i.e. $|z| \geq 1$) that are holomorphic and with a simple pole at ∞.*

$$\mathcal{M}_{2DT} := \left\{ (\bar{\lambda}, \lambda) \in \mathcal{M}_0 \times \mathcal{M}_\infty \mid \lambda + \bar{\lambda} \text{ is injective on the unit circle } |z| = 1 \right\}$$

Theorem 5.6 *$\mathcal{M}_{2DT}$ has a Frobenius Manifold structure with potential F given by:*

$$\begin{aligned} F(\mathbf{t}(\mathbf{w}), u, v) =& \frac{1}{2} \oint_\Gamma \oint_\Gamma \mathrm{Li}_3 \left(\frac{w_1 e^{t(w_1)}}{w_2 e^{t(w_2)}} \right) dw_1\, dw_2 + \oint_\Gamma \left(\frac{e^u}{w e^{t(w)}} - w e^{t(w)} \right) dw \\ &+ \frac{(2v + t_{-1})}{4} \oint_\Gamma (t(w))^2\, dw + \frac{1}{2} v^2 u - e^u \end{aligned} \tag{5.5}$$

$$E = -\sum_{\alpha \in \mathbb{Z}} \alpha t^\alpha \partial_\alpha + v \partial_v + 2\partial_u \qquad e = \partial_v$$

(*) Holomorphic in this case means holomorphic in some open set containing the closed disk

Here for a given point $(\bar{\lambda}, \lambda) \in \mathcal{M}$ *we have that* $t(w) := \log\left(\frac{z(w)}{w}\right)$*, where* $z(w)$ *is the inverse function of the map* $w(z) = \bar{\lambda}(z) + \lambda(z)$*.* Γ *is the image of the unit circle under the map* $w(z)$*. The coefficients of the function* $\mathbf{t}(w) = \sum_{\alpha\in\mathbb{Z}} t^\alpha w^\alpha$ *are, together with* u *and* v*, a system of flat coordinates for* $\mathcal{M}_{2DT}$*.*

Theorem 5.7 *$\mathcal{M}_{2DT}$ is semisimple.*

Starting from the Frobenius manifolds we can construct the complete set of first integrals of 2D Toda. To each flat coordinate will correspond a hierarchy of flows. The first non trivial flow of each hierarchy is called the primary flow.

Theorem 5.8 *The primary flows of the principal hierarchy associated with $\mathcal{M}_{2DT}$ have the following form:*

$$\frac{\partial\lambda(z)}{\partial t^{\alpha,0}} = \frac{1}{\alpha+1}\left\{w^{\alpha+1}(z)_{<0}, \lambda(z)\right\}, \quad \frac{\partial\bar{\lambda}(z)}{\partial t^{\alpha,0}} = -\frac{1}{\alpha+1}\left\{w^{\alpha+1}(z)_{\geq 0}, \bar{\lambda}(z)\right\}$$
$$\alpha \in \mathbb{Z}, \quad \alpha \neq -1,$$

$$\frac{\partial\lambda(z)}{\partial t^{-1,0}} = \left\{\left(\log\frac{w(z)}{z}\right)_{<0} + \log z, \lambda(z)\right\}, \quad \frac{\partial\bar{\lambda}(z)}{\partial t^{-1,0}} = -\left\{\left(\log\frac{w(z)}{z}\right)_{\geq 0}, \bar{\lambda}(z)\right\}$$
$$\frac{\partial}{\partial t^{v,0}} = \frac{\partial}{\partial x}$$
$$\frac{\partial}{\partial t^{u,0}} = -\frac{\partial}{\partial \bar{t}_1}$$

All these flows are symmetries of the dispersionless limit of the 2D Toda hierarchy, hence they extend the classical dispersionless 2D Toda hierarchy of Takasaki and Takebe.

Note that these new flows are not well defined at a formal level, where one is faced with the problem of performing multiplications of symbols which are not defined in the formal case. In order to do this we have to abandon the formal setting and introduce the space of holomorphic symbols $\mathcal{M}_{2DT}$. The principal hierarchy induced by this manifold is an extension of the dispersionless 2D Toda hierarchy, endowed with the bihamiltonian structure given by G. Carlet in [3]. Some of the flows of this hierarchy already appeared, under the assumption of bounded Lax operators, in the papers of Adler and Van Moerbeke [1]. The Hamiltonian formulation of these flows and their bihamiltonian recursion relation is a new result which follows from the application of the Frobenius manifold theory.

References

[1] M. Adler, P. van Moerbeke, *String-orthogonal polynomials, string equations, and* 2*-Toda symmetries*. Comm. Pure Appl. Math. 50 (1997), no. 3, 241–290.

[2] M. F. Atiyah, *Topological quantum field theories*. Publ. Math. I.H.E.S. 68 (1988) 175.

[3] G. Carlet, *The Hamiltonian structures of the two-dimensional Toda lattice and R-matrices*. Lett. Math. Phys. **71** (2005) 209–226.

[4] G. Carlet, B. Dubrovin, L. Ph. Mertens, *Infinite Dimensional Frobenius Manifolds for* $2+1$ *Integrable Systems*. To appear on the ArXiv.

[5] L. A. Dickey, "Soliton Equations and Hamiltonian Systems". World Scientific, 1991 ISBN 9810236840.

[6] B. Dubrovin, *Normal forms of hierarchies of integrable PDEs, Frobenius manifolds and Gromov - Witten invariants*. ArXiv:math/0108160.

[7] B. Dubrovin, *Integrable systems in topological field theory*. Nucl. Phys. B379 (1992), 627–689.

[8] G. Carlet, B. Dubrovin, Y. Zhang, *The extended Toda hierarchy*. (English, Russian summary) Mosc. Math. J. 4 (2004), no. 2, 313–332, 534.

[9] B. Dubrovin, *Geometry and analytic theory of Frobenius manifolds*. Proceedings of ICM98, Vol. 2, 315-326.

[10] N. J. Hitchin, G. Segal, R. S. Ward, N. M. J. Woodhouse, "Integrable Systems: Twistors, Loop Groups, and Riemann Surfaces". Published by Oxford University Press, 1999; ISBN 0198504217, 9780198504214.

[11] J. Kock, "Frobenius Algebras and 2D Topological Quantum Field Theories". Cambridge University Press, 2004; ISBN 0521540313, 9780521540315.

[12] K. Ueno, K. Takasaki, *Toda lattice hierarchy*. In: *Group Representations and Systems of Differential Equations* (Tokyo, 1982), 1–95, *Adv. Stud. Pure Math.*, **4**, North-Holland, Amsterdam, 1984.

[13] K. Takasaki, T. Takebe, *Integrable hierarchies and dispersionless limit*. (English summary) Rev. Math. Phys. 7 (1995), no. 5, 743–808.

[14] P. Guha, K. Takasaki, *Dispersionless hierarchies, Hamilton-Jacobi theory and twistor correspondences*. (English summary) J. Geom. Phys. 25 (1998), no. 3-4, 326–340.

[15] E. Witten, *Two-dimensional gravity and intersection theory on moduli space*. Surv. in Diff. geom. 1 (1991), 243–310.

The Dynamics of a Spin-Flip System by an Example: the Curie-Weiss Model

Francesca Collet (*)

Abstract. A spin system is a system composed by N sites at which is associated randomly a $+1$ or -1 value, called spin. Each spin is influenced by all the others in the same way and this makes it flip with a certain probability. The dynamics just mentioned are completely described by the time evolution of the Magnetization (the sum of all the spin values divided by N), so it is sufficient to study the behavior of this last quantity. As N grows to infinity, its limiting dynamics are deterministic (driven by an ordinary differential equation) and exhibit a phase transition: multiple equilibrium solutions arise depending on the value of a parameter, which is the inverse of the temperature. The Curie-Weiss model is a basic example of it. After having recalled some notions of Probability, we try to explain, avoiding the most part of the technicalities, how the Curie-Weiss model evolves in time at different temperatures.
The few results we will state in these notes are due to a joint work with Paolo Dai Pra.

1 Preliminary Notions

The present short section reviews some notions of Probability directly useful in the sequel. The intent is to recall briefly the characterization of Continuous-Time Markov chains [1] and to define the weak convergence for stochastic processes with cadlag trajectories [5, 6].

Definition 1 *A stochastic process is an object of the form*

$$\underline{X} = (\Omega, \mathscr{A}, (\mathscr{A}_t)_{t\in\mathscr{T}}, (X_t)_{t\in\mathscr{T}}, P)$$

where

- $\mathscr{A}$ *is a* σ*-field of parts of* Ω*;*
- P *is a probability measure on* $(\Omega, \mathscr{A})$*;*
- $(\mathscr{A}_t)_{t\in\mathscr{T}}$ *is a filtration, that is a family of sub-*σ*-fields of* $\mathscr{A}$ *increasing in* t*:* $\mathscr{A}_s \subseteq \mathscr{A}_t$ *if* $s \leq t$*;*

(*)Dottorato di ricerca in Matematica Computazionale, Dipartimento di Matematica Pura ed Applicata, Università di Padova. E-mail: `fcollet@math.unipd.it`. Seminar held on 11 March 2009.

- $\mathscr{T} \subseteq \mathbb{R}^+$ *is the time set;*
- $(X_t)_{t\in\mathscr{T}}$ *is a family of random variables on* $(\Omega, \mathscr{A})$ *taking values in a measurable space* $(E, \mathscr{E})$ *and such that it is adapted to the filtration* $(\mathscr{A}_t)_{t\in\mathscr{T}}$ *(in other words, for every* t*,* X_t *is* $\mathscr{A}_t$*-measurable).*

We now proceed to give the basic definitions concerning Continuous-Time Markov chains. Let E be a countable set, called the state space, and let $\underline{X}$ be an E-valued stochastic process, indexed by $t \in \mathbb{R}^+$. The probability distribution of $\underline{X}$ consists of the data $P(X_{t_1} = x_1, \dots, X_{t_k} = x_k)$, for all $t_1, \dots, t_k \geq 0$ and all $x_1, \dots x_k \in E$.

Definition 2 *The* E*-valued stochastic process* $\underline{X}$ *is called a Continuous-Time Markov chain if for all* $x, y, x_1, \dots, x_k \in E$ *and for all* $0 \leq s_1 \leq \cdots \leq s_k \leq s \leq t$*,*

$$P(X_t = y | X_s = x, X_{s_k} = x_k, \dots, X_{s_1} = x_1) = P(X_t = y | X_s = x)\,.$$

If $P(X_t = y|X_s = x)$ depends on s, t only through $t - s$ we say that the chain is time-homogeneous.
In this case, let

$$S_t := \{P(X_t = y|X_0 = x_0)\}_{x_0,y\in E}\,.$$

Proposition 1 $(S_t)_{t\geq 0}$ *is a semigroup, i.e.* $S_0 = I$ *and* $S_{s+t} = S_s \circ S_t$*.*

The distribution of X_0 and the semigroup $(S_t)_{t\geq 0}$ identify the law of the process; in fact, for $0 < t_1 < \cdots < t_k$ and $x_0, x_1, \dots, x_k \in E$

$$\begin{aligned} P(X_{t_k} = x_k, X_{t_{k-1}} = x_{k-1}, \dots, X_0 = x_0) = \\ = (S_{t_k - t_{k-1}})_{x_{k-1},x_k} (S_{t_{k-1}-t_{k-2}})_{x_{k-2},x_{k-1}} \dots (S_{t_1})_{x_0,x_1} P(X_0 = x_0)\,, \end{aligned}$$

where $(S_{t_j - t_i})_{x_i,x_j} = P(X_{t_j} = x_j | X_{t_i} = x_i)$.
If $t \mapsto S_t$ is continuous, it can be shown that

$$\lim_{t\to 0} \frac{S_t - I}{t} = L \qquad \text{exists}$$

and

$$S_t = e^{tL}\,.$$

The operator L is the *Infinitesimal Generator* of the Markov chain and allows us to construct its transition semigroup.

Finally, to conclude this introductory part, we need the last concept: weak convergence.

Definition 3 *Let $(E, \mathscr{E})$ be a metric space provided with the Borel σ-field.*
A sequence of random variables $X_n : (\Omega_n, \mathscr{A}_n, P_n) \longrightarrow (E, \mathscr{E})$, $n \in \mathbb{N}$, converges in distribution to $X : (\Omega, \mathscr{A}, P) \longrightarrow (E, \mathscr{E})$ if the sequence of the laws of μ_{X_n} converges weakly to μ_X; that is, for every continuous and bounded $f : E \longrightarrow \mathbb{R}$,

$$\int_E f(x)\mu_{X_n}(dx) \xrightarrow{n\to+\infty} \int_E f(x)\mu_X(dx)\,.$$

The stochastic processes whose trajectories live in the Skorohod Space (the space of the functions that are continuous on the right and have limit from the left, endowed with the Skorohod topology) can be interpreted as random variables taking values in this space:

$$\begin{array}{rccc} \underline{X}: & (\Omega, \mathscr{A}, P) & \longrightarrow & (\text{SKOROHOD SPACE, BOREL } \sigma\text{-FIELD}) \\ & \omega & \longmapsto & [t \mapsto X_t(\omega)]\,. \end{array}$$

Hence, we can transpose the definition above to define the weak convergence for sequences of processes of this type.

2 The Curie-Weiss Model

We consider N sites and we associate with each of them a spin value, which is a random variable taking values in the set $\mathscr{S} = \{-1, +1\}$. At time $t = 0$, we construct a configuration $\underline{\sigma}(0) = (\sigma_j(0))_{j=1}^N \in \mathscr{S}^N$, choosing N independent and identically distributed spins, with common law λ. The initial configuration evolves in time depending on the interaction between the sites. We assume this interaction to be of mean-field type, meaning that all the sites interact with all the others in the same way. This hypothesis allows us to suppose that

- the interaction depends on the value of the magnetization of the system

$$m_N^{\underline{\sigma}}(t) = \frac{1}{N}\sum_{j=1}^N \sigma_j(t)\,;$$

- the rates of the transitions are of the form

$$\sigma_j \longrightarrow -\sigma_j \qquad \text{at rate} \qquad e^{-\beta\sigma_j m_N^{\underline{\sigma}}}\,,$$

  where β, positive parameter, can be interpreted as the inverse of the temperature.

To be more formal, the stochastic process $\underline{\sigma}(t) = (\sigma_j(t))_{j=1}^N$, with t belonging to a generic time interval $[0, T]$, where T is fixed, describes a N-spin system evolving as a Continuous-Time Markov chain on $\mathscr{S}^N$ (the space of the configurations), with Infinitesimal Generator L_N acting on functions $f : \mathscr{S}^N \longrightarrow \mathbb{R}$ as follows:

$$L_N f(\underline{\sigma}) = \sum_{j=1}^N e^{-\beta\sigma_j m_N^{\underline{\sigma}}} \nabla_j^{\underline{\sigma}} f(\underline{\sigma}), \tag{1}$$

where $\nabla_j^\sigma f(\underline{\sigma}) = f(\underline{\sigma}^j) - f(\underline{\sigma})$ and the k-th component of $\underline{\sigma}^j$, which has the meaning of a spin flip at site j, is

$$\sigma_k^j = \begin{cases} \sigma_k & \text{for} \quad k \neq j \\ -\sigma_k & \text{for} \quad k = j \end{cases} .$$

The expression (1) describes a system of mean-field coupled spins, subject to a dynamics constructed to favor the configurations where the most part of them takes the same value. The quantity $\sigma_j(t)$ represents the time evolution on $[0,T]$ of j-th spin value; it is the trajectory of the single j-th spin in time. The space of all these paths is $\mathcal{D}[0,T]$, which is the space of the right-continuous, piecewise-constant functions from $[0,T]$ to $\mathscr{S}$. We endow $\mathcal{D}[0,T]$ with the Skorohod topology, which provides a metric and a Borel σ-field.

We now describe the dynamics of the process (1), in the limit as $N \longrightarrow +\infty$, in a fixed time interval $[0,T]$ and later, the large time behavior of the limiting dynamics will be discussed.
The key quantity of this analysis is the stochastic process $m_N^\sigma(t)$, since it is a sufficient statistic for our model. In this context it means that its evolution is Markovian and then it can be described through an Infinitesimal Generator, deduced by (1). Therefore, the study of the dynamics of the whole model is completely determined by the dynamics of its magnetization and the problem is reduced to be one-dimensional.

Theorem 1 *For $t \in [0,T]$, T fixed, in the limit as $N \longrightarrow +\infty$, the magnetization $m_N^\sigma(t)$ converges weakly to m_t^σ, which is the solution of the following ordinary differential equation*

$$\dot{m}_t^\sigma = -2m_t^\sigma \cosh(\beta m_t^\sigma) + 2\sinh(\beta m_t^\sigma)\,. \tag{2}$$

The equation (2) drives the behavior of the system governed by generator (1) in the infinite volume limit. We are interested in the detection of the t-stationary solution(s) of this equation. We recall that to be t-stationary solution for (2) means to satisfy the equation

$$-2m_t^\sigma \cosh(\beta m_t^\sigma) + 2\sinh(\beta m_t^\sigma) = 0\,.$$

Lemma 1 *Any equilibrium solution of (2) is of the form $m_*^\sigma = \tanh(\beta m_*^\sigma)$.*

Depending on the values of the parameter β, the model exhibits a phase transition, which is the appearance of multiple stable equilibria. In fact,

Theorem 2 *Consider the equation (2).*

- *For $\beta \leq 1$, it has 0 as a unique equilibrium solution (see Figure 1(a)) and it is globally asymptotically stable, i.e. for every initial condition m_0^σ*

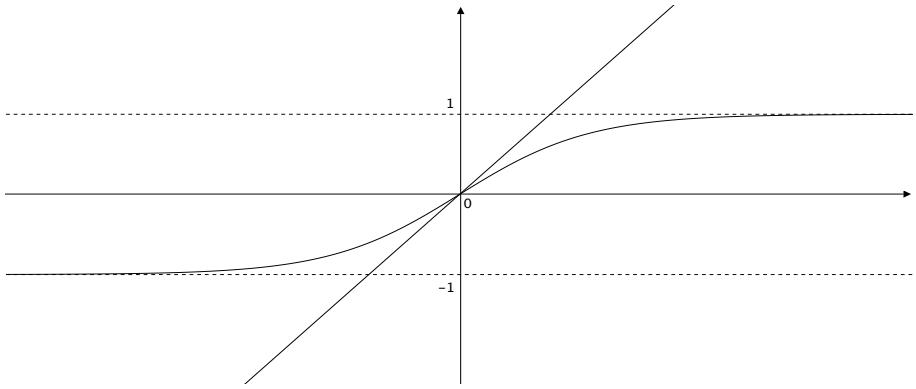
$$\lim_{t\to+\infty} m_t^\sigma = 0\,.$$

- *For $\beta > 1$, the point 0 is still an equilibrium and, moreover, two further equilibria arise:*

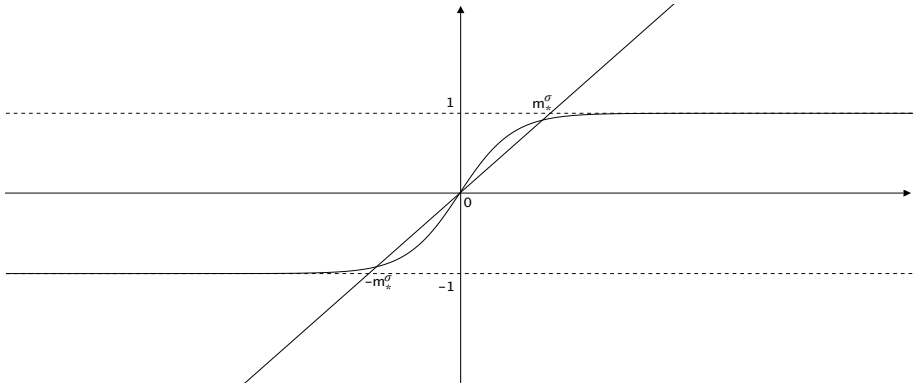
$$m_*^\sigma \qquad \text{and} \quad -m_*^\sigma\,,$$

where m_*^σ is the unique positive solution of $x = \tanh(\beta x)$ (see Figure 1(b)). In this case, the phase space $[-1,1]$ is bi-partitioned by the origin in two domains of attraction. Given an initial condition m_0^σ,

$$\lim_{t\to+\infty} m_t^\sigma = \begin{cases} m_*^\sigma & \text{if } m_0^\sigma \in (0,1] \\ -m_*^\sigma & \text{if } m_0^\sigma \in [-1,0) \\ 0 & \text{if } m_0^\sigma = 0\,. \end{cases}$$



(a) Case $\beta < 1$



(b) Case $\beta > 1$

Figure 1. Solution(s) of $m_*^\sigma = \tanh(\beta m_*^\sigma)$.

When the parameter β takes the values such that the system has exactly one stationary solution, we say the system to be in a subcritical regime; while, when more than one equilibrium appears, we are in a supercritical regime.
We want to study the fluctuations of the magnetization around its limiting dynamics.

We can capture different features of these fluctuations depending on whether or not the time is rescaled with N. If time is not rescaled and we consider the evolution in a time interval $[0, T]$, with T fixed, a Central Limit Theorem holds true for the magnetization for all regimes; in other words, the fluctuations of the magnetization converge to a Gaussian process.

Theorem 3 *For $t \in [0, T]$, with T fixed, the fluctuations of the magnetization*

$$z_N(t) := N^{1/2} \left(m_N^\sigma(t) - m_t^\sigma \right)$$

converge, in the sense of weak convergence of stochastic processes, to a limiting Gaussian Process $z(t)$, which is the unique solution of a linear diffusion equation.

Whenever time is rescaled in such a way T goes to infinity as N does, we may observe different behaviors.

- If $\beta > 1$, we expect to find a metastability-type phenomenon; in other words, the system may spend a very long time in a neighborhood of a stable equilibrium of the limiting dynamics and then switches to a neighborhood of another equilibrium.
- If $\beta < 1$, the Central Limit Theorem holds uniformly in time.
- If $\beta = 1$ (the parameters of the system are in the boundary between the subcritical and the supercritical regimes) the size of the Normal fluctuations must be further rescaled (in space and in time), because their size around the deterministic limit increases in time, and the limit can be non-Gaussian.

We focus on the dynamics of the critical fluctuations, i.e. the fluctuations of the magnetization in the infinite volume limit and at the critical point. We consider $\beta = 1$ and let us assume that the initial condition $\lambda^{\otimes N}$ is a product measure such that $m_0^\sigma = 0$ and so $m_t^\sigma = 0$ for every value of $t \geq 0$, since it is an equilibrium solution.

Theorem 4 *For $t \in [0, T]$, if we consider the critical fluctuation process*

$$x_N(t) := N^{1/4} m_N^\sigma(N^{1/2} t) \tag{3}$$

then, as $N \longrightarrow +\infty$, $x_N(t)$ converges, in the sense of weak convergence of stochastic processes, to a limiting Non-Gaussian process $x(t)$, which is the unique solution of the following stochastic differential equation:

$$\begin{cases} dx(t) &= -2x^3(t)dt + \sqrt{2}dB(t) \\ x(0) &= 0 \end{cases}$$

where $B(t)$ is a Standard Brownian Motion.

We are interested in the dynamics of the critical fluctuations in a slightly generalized system, a *disordered* mean-field system: the Random Curie-Weiss model. We aim at analyzing the effect of the disorder in the dynamics of critical fluctuations, as compared with the homogeneous case.

3 The Random Curie-Weiss model

We consider the Curie-Weiss model and we add a site-dependent, i.i.d. random environment, acting as an inhomogeneity in the structure of the system. Using the same notation as before, we briefly sketch the main results regarding the Random Curie-Weiss model.

Let $\underline{\eta} = (\eta_j)_{j=1}^N$ be a sequence of i.i.d., symmetric, Bernoulli random variables, taking values in $\mathscr{S}$. That is, $P(\eta_j = -1) = P(\eta_j = +1) = \frac{1}{2}$, for any j.
For given $\underline{\eta}$, let $\underline{\sigma}(t) = (\sigma_j(t))_{j=1}^N$, with $t \in [0,T]$, be the N-spin system evolving as a Continuous-Time Markov chain on $\mathscr{S}^N$, with generator L_N, acting on functions $f : \mathscr{S}^N \longrightarrow \mathbb{R}$ as follows:

$$L_N f(\underline{\sigma}) = \sum_{j=1}^{N} e^{-\beta\sigma_j(m_N^{\sigma}+h\eta_j)}\, \nabla_j^{\sigma} f(\underline{\sigma}) \tag{4}$$

where $\nabla_j^{\sigma} f(\underline{\sigma}) = f(\underline{\sigma}^j) - f(\underline{\sigma})$ and the k-th component of $\underline{\sigma}^j$, which has the meaning of a spin flip at site j, is

$$\sigma_k^j = \begin{cases} \sigma_k & \text{for} \quad k \neq j \\ -\sigma_k & \text{for} \quad k = j \end{cases} .$$

Let β, positive parameter, be the inverse of the temperature and η_j be the direction of the local magnetic field (of intensity h) associated with the site j. The quantity $e^{-\beta\sigma_j(m_N^{\sigma}+h\eta_j)}$ represents the jump rate of the spins; the rate at which the transition $\sigma_j \longrightarrow -\sigma_j$ occurs for some j.
The expression (4) describes a system of mean-field coupled spins, which feel the attraction of all the others but also of the local field we associate with each site.

Remark 1 The realization of the magnetic field does not evolve in time; it is an initial datum and it is not changed by the dynamics.

The initial condition $\underline{\sigma}(0)$ is assumed to have product distribution $\lambda^{\otimes N}$, with λ probability measure on $\mathscr{S}$ (meaning that, at time $t = 0$, we are choosing independent, identically distributed spins, with common law λ). The quantity $\sigma_j(t)$ represents the time evolution on $[0,T]$ of j-th spin value; it is the trajectory of the single j-th spin in time. The space of all these paths is $\mathcal{D}[0,T]$, which is the space of the right-continuous, piecewise-constant functions from $[0,T]$ to $\mathscr{S}$. We endow $\mathcal{D}[0,T]$ with the Skorohod topology, which provides a metric and a Borel σ-field.

As in the previous case, we reduce this system to be finite dimensional. A two-dimensional "magnetization vector" is necessary to describe completely the time evolution of the system:

$$\begin{pmatrix} m_N^{\sigma}(t) \\ \\ m_N^{\sigma\eta}(t) \end{pmatrix} = \begin{pmatrix} \frac{1}{N}\sum_{j=1}^N \sigma_j(t) \\ \\ \frac{1}{N}\sum_{j=1}^N \eta_j\sigma_j(t) \end{pmatrix} .$$

The infinite volume limit ($N \longrightarrow +\infty$) for the dynamics of the system governed by (4) is given by the limiting dynamics $(m_t^\sigma, m_t^{\sigma\eta})$.

Theorem 5 *For $t \in [0,T]$, T fixed, in the limit as $N \longrightarrow +\infty$, the magnetization converges weakly to the solution of the following system of ordinary differential equations*

$$\begin{aligned} \dot{m}_t^\sigma &= -2\, m_t^\sigma \cosh(\beta h)\cosh(\beta m_t^\sigma) - 2\, m_t^{\sigma\eta}\sinh(\beta h)\sinh(\beta m_t^\sigma) + \\ &\quad +2\cosh(\beta h)\sinh(\beta m_t^\sigma) \\ \dot{m}_t^{\sigma\eta} &= -2\, m_t^\sigma \sinh(\beta h)\sinh(\beta m_t^\sigma) - 2\, m_t^{\sigma\eta}\cosh(\beta h)\cosh(\beta m_t^\sigma) + \\ &\quad +2\sinh(\beta h)\cosh(\beta m_t^\sigma)\,. \end{aligned}$$

Depending on the parameters β and h, the system exhibits a phase transition. For $\beta \leq \cosh^2(\beta h)$, $(0, \tanh(\beta h))$ is the unique stationary solution; while, for $\beta > \cosh^2(\beta h)$, $(0, \tanh(\beta h))$ is still a stationary solution and (at least) two further equilibria arise. We refer to [4] for a complete phase diagram.

We are going to consider the fluctuations of the magnetization around its limiting dynamics. With regard to the Normal fluctuations, it remains proved the following Theorem.

Theorem 6 *For $t \in [0,T]$, with T fixed, the fluctuations of the magnetization*

$$\begin{aligned} z_N(t) &:= N^{1/2}\left(m_N^\sigma(t) - m_t^\sigma\right) \\ w_N(t) &:= N^{1/2}\left(m_N^{\sigma\eta}(t) - m_t^{\sigma\eta}\right) \end{aligned}$$

converge, in the sense of weak convergence of stochastic processes, to a limiting Gaussian Process $(z(t), w(t))$, which is the unique solution of a linear diffusion equation.

We focus on the critical dynamics of the system. We construct the fluctuations in the threshold case, when $\beta = \cosh^2(\beta h)$, and we look at their long-time behavior. The size of the Normal fluctuations must be further rescaled (in space and in time), because their size around the deterministic limit increases in time. In this case we will still obtain Normal fluctuations.
We consider $\beta = \cosh^2(\beta h)$ and let us assume that the initial condition $\lambda^{\otimes N}$ is a product measure such that $m_0^\sigma = 0$, $m_0^{\sigma\eta} = \tanh(\beta h)$ and so $m_t^\sigma = 0$, $m_t^{\sigma\eta} = \tanh(\beta h)$ for every value of $t \geq 0$, since it is an equilibrium solution.

Theorem 7 *For $t \in [0,T]$, if we consider the two-dimensional critical fluctuation process*

$$\begin{aligned} x_N(t) &:= N^{1/4} m_N^\sigma(N^{1/4}t) \\ y_N(t) &:= N^{1/4}\left(m_N^{\sigma\,\eta}(N^{1/4}t) - \tanh(\beta h)\right) \end{aligned} \tag{5}$$

then, as $N \longrightarrow +\infty$, $y_N(t)$ *collapses and* $x_N(t)$ *converges, in the sense of weak convergence of stochastic processes, to a limiting Gaussian process*

$$x(t) = 2\,\mathscr{H}\,\sinh(\beta h)\,t\,,$$

where $\mathscr{H}$ *is a Standard Gaussian random variable.*

4 Conclusions

We point out the fact that the inhomogeneous critical fluctuation process exists in a shorter time-scale than the homogeneous one, in fact in (5) we can amplify the time only by a factor $N^{1/4}$, instead of the usual scale $N^{1/2}$, as in (3). The reason of this difference is the random field we have introduced, which generates the constant drift $\mathscr{H}$, which does not allow us to rescale the time more. Besides, the limit of disordered critical fluctuations is Gaussian, since solution of a deterministic equation with constant (but random) drift given by a Gaussian random variable; while, it is not when there is no added field.

References

[1] Pierre Brémaud, "Markov chains: Gibbs fields, Monte Carlo simulation, and queues". Springer-Verlag, New York,1999.

[2] Francesca Collet, The impact of disorder in the critical dynamics of mean-field models, PhD thesis, Department of Pure and Applied Mathematics, University of Padova, 2009.

[3] Francis Comets and Thomas Eisele, *Asymptotic dynamics, noncritical and critical fluctuations for a geometric long-range interacting model*. Comm. Math. Phys. 118 (1988), 531–567.

[4] Paolo Dai Pra and Frank den Hollander, "McKean-Vlasov limit for interacting random processes in random media". Technical report, Department of Mathematics, University or Nijmegen, 1995.

[5] Stewart N. Ethier and Thomas G. Kurtz, "Markov processes: characterization and convergence". John Wiley & Sons Inc., New York, 1986.

[6] Albert N. Shiryaev, "Probability". Springer-Verlag, New York, second edition, 1996.

Computing with Affine Algebraic Groups

Andrea Pavan (*)

Abstract. How can one solve the Rubik's Cube? The question turns out to be equivalent to a problem about groups, whose solution is provided by Computational Group Theory. More generally, CGT is concerned with designing and analyzing algorithms to compute information about groups which can be described by a finite amount of data. Examples include finite permutation groups, finitely presented groups, finitely generated matrix groups and polycyclic groups, which have been at the center of the subject since the beginning of the last century. On the contrary, very little work has been done on affine algebraic groups. These are, roughly speaking, groups whose elements are solutions to some system of polynomial equations in finitely many indeterminates. Although their structure is well understood, they have been rarely studied from a computational point of view. Two pioneers in the field are Grunewald and Segal, who developed the basis for many useful algorithms. In the first part of these notes we will give an introductory overview of both Computational Group Theory and the theory of affine algebraic groups. Then we will describe the work of Grunewald and Segal, as well as some improvements of their methods.

1 The Rubik's Cube

Imagine you have a Rubik's cube in its original configuration, like the one in Figure 1.

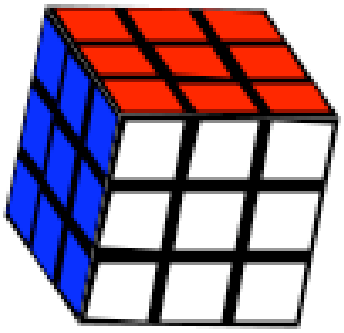


Figure 1: Rubik's cube in its original configuration.

You can begin playing with it applying any sequence of basic transformations, that

(*)Dottorato di ricerca in Matematica Pura, Dipartimento di Matematica Pura ed Applicata, Università di Padova. E-mail: pan@math.unipd.it. Seminar held on 25 March 2009.

it to say, 90-degrees rotations of a side in the direction you prefer. For example, you may rotate the upper side either clockwise or counterclockwise, and then the right side clockwise, or the front side counterclockwise... Eventually, you will reach a configuration like the one in Figure 2.

Figure 2: a shuffled Rubik's cube.

Now suppose you want to bring the cube back to its original configuration. Of course you can try to apply some randomly chosen basic transformations to the cube, and see what happens. In this way, though, it is likely that you will spend a very very long time before reaching your goal – if ever! Fortunately, a bit of math can help you. Let us fix a numbering for the facets of the cube (a facet is one of the $3 \times 3 = 9$ squares every side of the cube is divided in). For example, with respect to the original cube (Figure 1), we could assign the number 1 to the facet on the top-left corner of the front side. Note that this facet can alternatively be described as the white facet of the unique cubic containing also a red facet and a blue facet (a cubic is one of the $3 \times 3 \times 3 = 27$ little cubes Rubik's cube is composed of, and it contains either zero, one, two or three facets). Also, we could assign the number 2 to the facet on the right of facet 1 – that is to say, to the white facet of the cubic containing also a red facet, the number 3 to the facet on the right of facet 2, the number 4 to the facet below facet 1 – equivalently, to the white facet on the cubic containing also a blue facet, and so on. Of course, the last facet will have number $9 \times 6 = 54$. Then we can associate to the shuffled cube in Figure 2 a function sending a number m to the number of the facet whose position in the original cube is the same as the position of facet m in the shuffled cube. Of course, it is a permutation on the set $\{1, 2, \ldots, 54\}$, that is to say, an element of the symmetric group S_{54} on the set of the first 54 positive integers. Also, it is possible to compute it very explicitely. For example, recall that facet 4 was the white facet of the cubic containing a blue facet, too. Therefore looking at Figure 2 we soon realize that the position of facet 4 in the shuffled cube is the central facet of the upper row of the front side, that is to say, the position occupied by facet 2 in the original cube: thus 4 is sent to 2. In the very same way, we can associate an element of S_{54} to every shuffled cube, and it is quite evident that a shuffled cube is completely described by it; in particular, a shuffled cube is in fact the original cube if and only if the permutation is the identical one. Furthermore, it is possible to associate a permutation to every basic transformation, namely, the permutation associated to the shuffled cube

obtained applying the basic transformation to the original cube. In this way, if ω is the permutation of a basic transformation, that is to say, a 90-degrees rotation of a side in some direction, then ω^{-1} is the permutation associated to the 90-degrees rotation of the same side in the opposite direction; for this reason, the latter transformation is called the inverse of the former. To our purposes, the fundamental fact is the following: if π is the permutation of a shuffled cube, and ω is the permutation of a basic transformation, then applying the basic transformation to the shuffled cube we obtain a new shuffled cube whose permutation is $\omega \circ \pi$. A first consequence is that the subset of S_{54} consisting of the permutations associated to the shuffled cubes is precisely the subgroup of S_{54} generated – even as a monoid – by the permutations corresponding to the basic transformations. Secondly, and more importantly, it suggests a possible strategy for solving our problem: (1) find the permutation corresponding to the shuffled cube, (2) write it as a product

$$\omega_1 \omega_2 \dots \omega_m$$

in S_{54}, where the ω_i's are permutations corresponding to the basic transformations, and then (3) apply to the shuffled cube the inverses of the basic transformations corresponding to $\omega_1, \omega_2, \dots, \omega_m$. In this way we will end up with the cube in its original configuration.

2 Computational Group Theory

Step (2) above is a particular instance of: Let Ω be a finite set, G a subgroup of the symmetric group $\mathrm{S}(\Omega)$ on Ω generated by $g_1, \dots, g_m$, and let g be an element of G. Write g in the form

$$g_{i_1}^{e_1} \cdots g_{i_n}^{e_n}$$

where the g_{i_j}'s are chosen among the given generating set and the e_j's are ± 1. This is an example of an algorithmic problem for permutation groups. The main theoretical tool for solving it, as well as many others, is the Schreier-Sims algorithm, which computes a "good" set of generators for G. More precisely, let $\omega_1, \dots, \omega_l \in \Omega$, and let X be a finite set of generators for G. Then $\omega_1, \dots, \omega_l$ is called a basis for G if $\cap_{i=1}^{l} G_{\omega_i} = 1$ (here G_{ω_i} denotes the stabilizer of ω_i in G), and X is called a strong generating set for G (with respect to the basis $\omega_1, \dots, \omega_l$) if

$$X \cap \bigcap_{i=1}^{k} G_{\omega_i}$$

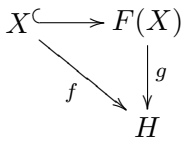
is a generating set for $\bigcap_{i=1}^{k} G_{\omega_i}$ for every integer k between 1 and l. Schreier-Sims algorithm, starting from any finite set of elements of $\mathrm{S}(\Omega)$, computes a basis and a strong generating set for the subgroup they generate. It is based on the well-known

Lemma 1 (Schreier) *Let G be a group generated by a subset S, H a subgroup of G and T a right transversal of H in G. For every $g \in G$, let us denote by $\overline{g}$ the unique element of T such that $Hg = H\overline{g}$. Then*

$$\{ts\overline{ts}^{-1} \mid s \in S, t \in T\}$$

generates H.

Algorithmic problems for finitely presented groups have been extensively studied, too. Recall that a free group (on a set X) is a group $F(X)$ containing X and satisfying the following universal property: for every function f from X to a group H, there exists a unique group morphism g from $F(X)$ to H such that



is commutative. Concretely, $F(X)$ is realized as a suitable quotient set of the set of words in the alphabet

$$X \cup X^{-1} = X \cup \{x^{-1} \mid x \in X\}$$

and the multiplication comes from concatenation of words. As an example, if $x, y, z \in X$, then the words $xyy^{-1}z$ and xz represent the same element of $F(X)$, and the product of the elements (represented by) xz and zyx^{-1} is the element (represented by) $xzzyx^{-1}$. Moreover, recall that a group is finitely presented group if it has the form

$$G = F(X)/N$$

where $F(X)$ is the free group on a finite set $X = \{x_1, \ldots, x_m\}$ and N is a normal subgroup of $F(X)$ with the additional property that there exist $r_1, \ldots, r_n \in F(X)$ such that N is the smallest normal subgroup of $F(X)$ containing them. If this is the case, we also say that

$$\langle x_1, \ldots, x_m \mid r_1, \ldots, r_n \rangle$$

is a finite presentation for G in the generators $x_1, \ldots, x_m$ and with relations $r_1, \ldots, r_n$. In 1911 Dehn asked if, fixed a group G given by means of a finite presentation, there exists an algorithm that, starting from an element $g \in G$ described as a word in the generators, decides whether $g = 1$ or not. This is the so-called "word problem" for finitely presented groups. In the '50 Novikov and Boone showed that, using the widely accepted notion of algorithm proposed by Turing, the answer in general is "no". Algorithmic problems that do not admit a solution are called undecidable, and they are very typical in the computational theory of finitely presented groups. For example, the conjugacy problem, that is to say, given x and y in G, to decide whether there exists $z \in G$ such that $zxz^{-1} = y$, and the isomorphism problem, which consists in deciding whether G is isomorphic to another finitely presented group H again given by means of a finite presentation, are both undecidable. As one of the few effective techniques available in dealing with finitely presented groups, we cite the Todd-Coxeter coset enumeration procedure. Given a finitely generated subgroup H of G whose index is known to be finite, it computes a right transveral for H is G as well as the image of the natural right action of G on the set of right cosets of H in G. However, the situation is much better if we restrict to polycyclic groups. These are solvable groups such that every subgroup is finitely generated. Not only they are finitely presented, but they even admit a finite presentation in some generators $g_1, \ldots, g_m$ such that the relations have the form (for suitable indexes i and j)

- $g_i^{e_i} = g_{i+1}^{a_{i+1}} \cdots g_m^{a_m}$,
- $g_i^{g_j} = g_{j+1}^{b_{i,j,j+1}} \cdots g_m^{b_{i,j,m}}$ and
- $g_i^{g_j^{-1}} = g_{j+1}^{c_{i,j,j+1}} \cdots g_m^{c_{i,j,m}}$,

where the e_i's, the a_i's, the $b_{i,j,k}$'s and the $c_{i,j,k}$'s are integers. Exploiting the existence a finite presentation of this kind, which is deemed polycyclic, it is possible to prove that most of the algorithmic problems for polycyclic group – even the word problem and the conjugacy problem – are decidable.

So far we described the problems which have traditionally been the main subject of investigation of Computational Group Theory. However, more recently attention was also focused on new problems that arise naturally when we consider some other classes of groups, such as matrix groups and black-box groups. The former are subgroups of some general linear group $\mathrm{GL}_n(F)$ over a finite field F. A theorem due to Aschbacher assures that they fit into nine (sometimes overlapping) classes, and the leading problem in this research area is to develop algorithms that, starting from a matrix group, find at least one of the classes it belongs to. The latter are groups whose elements are represented by words of bounded length on a finite alphabet, and in which multiplication is performed by an "oracle" or a "black-box". Otherwise stated, we can have access to the result of a multiplication, but we do not know the rules according to which it is computed. This model of computation is well-suited if we want to design algorithms that only depend on group-theoretical properties. However, forgetting about the nature of the group we are dealing with may have some drawbacks. As an example, suppose we want to compute the 100^{th} power of the cycle (123) of S_3. Obviously it has order 3, hence

$$(123)^{100} = (123).$$

On the contrary, in order to compute $(123)^{100}$ in the context of black-box groups we are forced to perform a computation like this:

$$(123)(123)\cdots(123) \quad (100 \text{ times!})$$

Another class of groups for which it is possible to develop an algorithmic theory will be introduced in a while.

3 The Theory of Affine Algebraic Groups

Let k be a field, and let $p_1, \ldots, p_n \in k[X_1, \ldots, X_m]$. Then we can consider the following system of polynomial equations (in finitely many indeterminates and equations, as we will always implicitely assume):

$$\begin{cases} p_1(X_1, \ldots, X_m) = 0 \\ \vdots \\ p_n(X_1, \ldots, X_m) = 0 \end{cases} \tag{1}$$

For any k-algebra R it makes sense to consider the solutions of (1) with coefficients in R: these are the m-uples $(a_1, \dots, a_m)$ of elements of R such that, when we substitute X_1 by a_1, X_2 by a_2, and so on, we have that all the equalities in (1) are satisfied. Moreover, if we put

$$A = k[X_1, \dots, X_m]/I$$

where I is the ideal generated by $p_1, \dots, p_n$, then the solutions of (1) with coefficients in R correspond to the morphisms of k-algebras from A to R, whose set is usually denoted by $\mathrm{Hom}(A, R)$, via the function that sends $f : A \to R$ to the m-uple

$$\big(f(X_1 + I), \dots, f(X_m + I)\big).$$

In a manner of speaking, the function $\mathrm{Hom}(A, \bullet)$ that to every k-algebra R associates the set $\mathrm{Hom}(A, R)$ represents up to the form the solutions of (1). The other way round, it follows from Hilbert's basis theorem that for every finitely generated k-algebra A the function $\mathrm{Hom}(A, \bullet)$ represents the solutions of a system of polynomial equations. Moreover, these properties are shared by other mathematical objects. To see this, note that $\mathrm{Hom}(A, \bullet)$ can be made into a functor – the functor represented by A – from the category of k-algebras to the category of sets sending every morphism $f : R \to S$ to the function

$$\sharp \circ f : \mathrm{Hom}(A, R) \to \mathrm{Hom}(A, S) \quad g \mapsto f \circ g.$$

Also, recall that if $\mathbf{X}$ is another functor from the category of k-algebras to the category of sets, then we say that $\mathbf{X}$ and $\mathrm{Hom}(A, \bullet)$ are isomorphic if it is possible to associate to every k-algebra R a bijection β_R between $\mathbf{X}(R)$ and $\mathrm{Hom}(A, R)$ in such a way that they are compatible as a whole, that is to say, for every morphism $f : R \to S$ of k-algebras the diagram

$$\begin{array}{ccc} \mathbf{X}(R) & \xleftarrow{\ \beta_R\ } & \mathrm{Hom}(A, R) \\ \Big\downarrow{\scriptstyle \mathbf{X}(f)} & & \Big\downarrow{\scriptstyle \sharp \circ f} \\ \mathbf{X}(S) & \xleftarrow{\ \beta_S\ } & \mathrm{Hom}(A, S) \end{array}$$

is commutative. If this is the case, since $\mathrm{Hom}(A, \bullet)$ represents the solutions of (1), so does $\mathbf{X}$. Summing up, functors from k-algebras to sets which are isomorphic to a functor represented by a finitely generated k-algebra are the appropriate mathematical formalization of the notion of solutions of a system of polynomial equations. For this reason, they deserve a name: affine algebraic sets over (the ground field) k. Furthermore, if $\mathbf{X}$ and $\mathrm{Hom}(A, \bullet)$ are isomorphic, we also say that A is the coordinate ring of $\mathbf{X}$. This definition makes sense by virtue of

Lemma 2 (Yoneda) *Suppose that the functors represented by the k-algebras A and B are isomorphic. Then $A \simeq B$, too.*

As a typical example of an affine algebraic set over k, let us consider the functor that to every R associates

$$V \otimes R$$

where V is a finite dimensional k-vector space. To see that it is actually an affine algebraic set, it is enough to show that as a coordinate ring we can choose $k[X_1, \ldots, X_m]$, where m is the dimension of V.

Now it comes the definition we were waiting for: an affine algebraic group over k is a functor from the category of k-algebras to the category of groups that, once we regard it as a functor to the category of sets, is also an affine algebraic set. There exists a natural notion of morphism between two affine algebraic groups $\mathbf{G}$ and $\mathbf{H}$. It is a family of group morphisms from $\mathbf{G}(R)$ to $\mathbf{H}(R)$ – one for every R – which are compatible as a whole (in the very same meaning as for an isomorphism of functors). Then it makes sense to consider the category of affine algebraic groups, and we can use category theory to introduce the concepts of monomorphism, epimorphism, isomorphism, subgroup, and so on. For example, a morphism of affine algebraic groups will be called monomorphism if it is so as an arrow in the category of affine algebraic groups. In this way most of the theory of abstract groups remains true in this new context, *mutatis mutandis*. In particular, a morphism from $\mathbf{G}$ to $\mathbf{H}$ is a monomorphism if and only if for every R the corresponding map $\mathbf{G}(R) \to \mathbf{H}(R)$ is a monomorphism, and the same is true for isomorphism. However, the maps $\mathbf{G}(R) \to \mathbf{H}(R)$ being epimorphic for every R is just a sufficient condition for the epimorphicity of $\mathbf{G} \to \mathbf{H}$. Also, if $\mathbf{N}$ is a normal subgroup of $\mathbf{G}$ then it is possible – though not easy – to show that there exists $\mathbf{Q}$ and a morphism π from $\mathbf{G}$ to $\mathbf{Q}$ such that

$$1 \to \mathbf{N} \hookrightarrow \mathbf{G} \xrightarrow{\pi} \mathbf{Q} \to 1$$

is exact: therefore quotients exist. Moreover, if $\mathbf{H}$ and $\mathbf{N}$ are subgroups of $\mathbf{G}$ and $\mathbf{N}$ normal, then

$$\frac{\mathbf{HN}}{\mathbf{N}} \simeq \frac{\mathbf{H}}{\mathbf{H} \cap \mathbf{N}},$$

that is to say, the third isomorphism theorem holds. Furthermore, for every integer n it is possible to consider the affine algebraic group GL_n sending every algebra R to $\mathrm{GL}_n(R)$. These groups, as well as their subgroups, are called (affine) algebraic matrix groups. It is a classical result that every group $\mathbf{G}$ is isomorphic to a subgroup of some GL_n. Hence algebraic matrix groups are very typical. When the ground field is $\mathbb{Q}$, then the function that to every group $\mathbf{G}$ associates $\mathbf{G}(\mathbb{C})$ induces a bijection from the set of subgroups of GL_n to the set of Zariski-closed subgroups of $\mathrm{GL}_n(\mathbb{C})$ defined over $\mathbb{Q}$. This fact has also an historical interest, since the defining property of the latter set is precisely the definition of affine algebraic group over $\mathbb{Q}$ that was given first.

4 Algorithms for Affine Algebraic Groups

Exploiting Yoneda's lemma, it is possible to show that the multiplication in an affine algebraic group $\mathbf{G}$ can be encoded by a morphism of algebras from the coordinate ring A of $\mathbf{G}$ to $A \otimes A$, which is called co-multiplication – and it could even be showed that A equipped with co-multiplication is a Hopf algebra. Therefore $\mathbf{G}$ can be described by a finite amount of data: namely, a finite family of polynomials $p_1, \ldots, p_n$ in $k[X_1, \ldots, X_m]$ such that

$$A = k[X_1, \ldots, X_m]/(p_1, \ldots, p_n)$$

is the coordinate ring of $\mathbf{G}$ and, for every X_i, the image of its equivalence class in A through co-multiplication. As another application of Yoneda's lemma, a morphism from $\mathbf{G}$ to another group $\mathbf{H}$ is encoded by a morphism from the coordinate ring B of $\mathbf{H}$ to A, which again can be described by a finite amount of data. This makes possible to develop an algorithmic theory for affine algebraic groups. Moreover, the class of affine algebraic groups is very different from the other classes of groups that have been traditionally considered in Computational Group Theory. To see this, it is enough to note that in general $\mathbf{G}(k)$ is neither finite nor even finitely generated, while permutation groups, finitely presented groups, matrix groups and black-box groups are. Nevertheless, very little work has been done on affine algebraic groups, and there are only few algorithmic problems for which a solution is already known. As an example, let us consider this: let $\mathbf{H}$ and $\mathbf{K}$ be subgroups of $\mathbf{G}$ with coordinate ring B and C, respectively, and let $f : A \to B$ and $g : A \to C$ be the maps corresponding to the inclusions. We want to understand whether $\mathbf{H} \leq \mathbf{K}$. It turns out that this is equivalent to

$$\ker g \leq \ker f.$$

Checking this condition is not a trivial problem, but fortunately there exist well-known techniques to solve it, and that rest upon the theory of Grobner basis. Another problem that has already been solved comes from the general theory of affine algebraic groups. More precisely, we know that for every finite dimensional vector space V there exists a functor Lie from the category of subgroups of GL_V to the category of sub-Lie-algebras of $\mathfrak{gl}_V$. When the ground field has characteristic 0, the functor has good properties. For example, it is injective and inclusion preserving. Also, starting from the coordinate ring of a group it is very easy to compute its Lie algebra. However for a number of applications it is useful to take the inverse route. This turns out to be a challenging problem, and a solution has been provided only recently by Willem de Graaf. Again by general theory it is well-known that every affine algebraic group admits a series like the one in Figure 3.

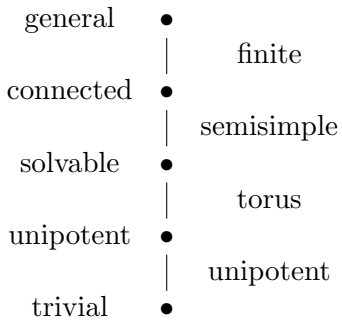


Figure 3: a series for a general affine algebraic group.

By this we mean that every group admits a normal subgroup which is connected and such that the quotient of the original group over it is finite, and further that such a subgroup has in turn a normal subgroup which is solvable and such that the quotient is semisimple, and so on. In this context, a group is connected if it is so as a topological space – with respect to a certain topology with which every affine algebraic group is naturally endowed.

Also, the concepts of finite, solvable, unipotent, semisimple and toric affine algebraic group are more or less straightforward generalizations of the homonymous notions that can be found in the theory of (matrix) groups (and will be partially explained in the following). In the '80, Grunewald and Segal (see also Figure 4)

Figure 4: from left to right, Dan Segal and Fritz Grunewald. Maybe some of you will remember Dan Segal since it held a talk in the Seminario Dottorato 2007/08.

presented an algorithm that, starting from any group $\mathbf{G}$, computes a series for $\mathbf{G}$ of the above kind. Later on, de Graaf developed another strategy for performing the same task and that turned out to be more efficient. As a last example, we will describe a problem that stems from a celebrated theorem due to Borel and Harish-Chandra. A representation of an affine algebraic group $\mathbf{G}$ on a finite dimensional vector space V associates to every algebra R a R-linear representation of $\mathbf{G}(R)$ on $R \otimes V$ in such a way that these representations satisfy the appropriate version of the usual compatibility condition. In 1962, Borel and Harsh-Chandra showed that, if the ground field is $\mathbb{Q}$ and the representation is faithful (*i.e.* for every R the representation of $\mathbf{G}(R)$ is faithful), then for every full-dimensional lattice L of V the "arithmetic" group

$$\mathbf{G}_L = \{g \in \mathbf{G}(\mathbb{Q}) \mid g.L = L\}$$

is finitely generated. Then it makes sense to ask whether there exists an algorithm for computing a finite set of generators for $\mathbf{G}_L$. In the '80, Grunewald and Segal proved that the problem is decidable. As a consequence, they showed that the isomorphism problem for finitely generated nilpotent groups is decidable, too.

An algorithm is called practical if its effective implementation on a computer, usually obtained with the aid of a computer algebra like GAP and MAGMA, runs in a reasonable amount of time on a large enough class of non-trivial inputs. Unfortunately, the algorithm presented by Grunewald and Segal is not. Moreover, the set of generators it computes is in general very (and unnecessarily) large. During my Ph.D. thesis I developed in collaboration with de Graaf two new algorithms that solve the same problem in two special cases and that do not suffer the same drawbacks of the algorithm of Grunewald and Segal. The first algorithm we presented deals with the case when $\mathbf{G}$ is unipotent. These are groups such that every representation on a non-zero vector space admits a non-zero fixed vector, that

is to say, a $0 \neq v \in V$ such that $g.1_R \otimes v = 1_R \otimes v$ for every algebra R and every $g \in \mathbf{G}(R)$. As an easy consequence, every representation admits a flag, that is a sequence

$$0 = V_0 < V_1 < \cdots < V_n = V$$

of subspaces of the underlying vector space V such the V_i/V_{i-1}'s consist of vectors that are fixed with respect to the natural action of $\mathbf{G}$ on V/V_{i-1}. In this case $\mathbf{G}_L$ is a torsion-free nilpotent group whose Hirsch length is equal to the dimension of $\mathbf{G}$. Our algorithm computes even a polycyclic sequence for $\mathbf{G}_L$ of the same length. Roughly speaking, the strategy is to find a flag for the representation, and then to use it in order to reduce the computation to a similar one, with the crucial difference that this time it involves a representation which admits a flag consisting of fewer subspaces. Also, a major role is played by the fact that, since we are considering unipotent groups, $\mathbf{G}$ and its Lie algebra are isomorphic as affine algebraic sets via a slight generalization of the usual exponential and logarithmic maps. Therefore we can easily find many points of $\mathbf{G}(\mathbb{Q})$. The algorithm has been implemented in GAP and, although some numerical tests as well as some theoretical considerations suggest that the running time is exponential on the dimension of the representation, it turned out that the algorithm is efficient enough to tackle non-trivial examples. See also Table 1.

index of the group	running time for the 1st family	running time for the 2nd family	running time for the 3rd family
6	0.7	0.4	1.8
7	3	3	11
8	24	16	95
9	204	133	963

Table 1: running time (in seconds, on a standard computer) of the algorithm for unipotent groups applied to some elements of three families of groups (of increasing dimension) indexed by natural numbers.

The second algorithm solves the problem in the special case in which $\mathbf{G}$ is a torus. This means that $\mathbf{G}$ is connected and that every representation of $\mathbf{G}_{\mathbb{C}}$ is diagonalizable, that is to say, the underlying vector space is the sum of its one-dimensional $\mathbf{G}_{\mathbb{C}}$-stable subspaces – also, here $\mathbf{G}_{\mathbb{C}}$ is the affine algebraic group over $\mathbb{C}$ that to every $\mathbb{C}$-algebra associates $\mathbf{G}(R)$. In this case, it is known that $\mathbf{G}_L$ is a finitely generated abelian group. We described an algorithm that computes a good finite set of generators for $\mathbf{G}_L$; by "good", of course, we mean a set of generators that resembles the structure of $\mathbf{G}_L$ as a direct sum of cyclic groups. The strategy was to find a semisimple, commutative and finite dimensional $\mathbb{Q}$-algebra A acting on V and with the property that $\mathbf{G}_L$ can be embedded in the stabilizer $A_L^\times$ of L with respect to the natural action the group of units $A^\times$ of A on L and, even more, that $\mathbf{G}_L$ is equal to the kernel of a group morphism from $A_L^\times$ to a direct product of multiplicative groups of number fields. With the aid of some well-known algorithms due to Pohst, Zassenhaus and Ge, we also showed how to make all these constructions and

computations effective. The algorithm was implemented in MAGMA, and some numerical tests show that it can be considered practical. See also Table 2.

index of the group	running time
10	1581
11	17.2
12	169
13	1581

Table 2: running time (in seconds, on a standard computer) of the algorithm for tori applied to some elements of a family of groups (of increasing dimension) indexed by natural numbers.

The next special case that we would like to face in the future is that of solvable groups. These are the natural generalization of solvable groups in the abstract case. Also, every unipotent group and every torus is solvable, and every solvable group admits a series like the one in Figure 5.

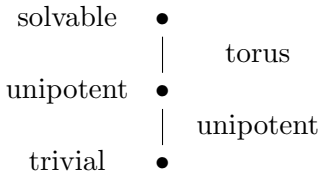


Figure 5: a series for a solvable affine algebraic group.

Also, it is well-known that

Theorem 1 (Lie-Kolchin) *Let* $\mathbf{G}$ *be a solvable and connected affine algebraic group over a field* k *of characteristic* 0*. For every representation of* $\mathbf{G}$ *on a finite dimensional vector space* V *there exists a chain*

$$0 = V_0 < V_1 < \cdots < V_n = V$$

of $\mathbf{G}$*-stable subspaces of* V *such that the image of the natural action of* $\mathbf{G}$ *on the* V_i/V_{i-1}*'s is a torus.*

These considerations seem to indicate that a sharp refinement of the techniques employed in the two special cases considered so far could lead to a practical solution of the problem for this more general class of groups.

Some bibliography

For a general introduction to Computational Group Theory, see:

- Holt, Derek F.; Eick, Bettina; O'Brien, Eamonn A. Handbook of computational group theory. Discrete Mathematics and its Applications (Boca Raton). *Chapman & Hall/CRC, Boca Raton, FL*, 2005.

Two classical introductory books to the theory of algebraic groups are:

- Borel, Armand. Linear algebraic groups. Second edition. Graduate Texts in Mathematics, 126. *Springer-Verlag, New York*, 1991.
- Waterhouse, William C. Introduction to affine group schemes. Graduate Texts in Mathematics, 66. *Springer-Verlag, New York-Berlin*, 1979.

Finally, some research papers of the algorithmic theory of affine algebraic groups:

- Grunewald, Fritz; Segal, Daniel. Some general algorithms. I. Arithmetic groups. *Ann. of Math.* (2) 112 (1980), no. 3, 531–583.
- de Graaf, Willem A. Constructing algebraic groups from their Lie algebras. *Journal of Symbolic Computation*, available online 13 February 2009.
- de Graaf, Willem A.; AP. Constructing arithmetic subgroups of unipotent groups. *Journal of Algebra*, to appear.

Constraint Programming Techniques for Mixed Integer Linear Programs

Domenico Salvagnin (*)

Abstract. Many decision problems in industry, logistics, and telecommunications can be viewed as satisfiability or optimization problems. Two paradigms have reached a high degree of sophistication from the point of view of both theory and implementation: Constraint Programming (CP) and Mixed Integer Programming (MIP). The CP and MIP paradigms have strengths and weaknesses that complement each other. On the one hand, CP, through the use of sophisticated propagation techniques, privileges primal inference. On the other hand, MIP, through the techniques of relaxation and strengthening through cutting planes, privileges dual inference.

In these notes we will deal with *optimization problems*, i.e., problems of the form

$$z = \min f(x) \tag{1}$$

$$\mathcal{C} \tag{2}$$

$$x \in D \tag{3}$$

Here $f(x)$ is a real valued *objective function* of the variable x to be minimized, D is the *domain* of x and $\mathcal{C}$ is a finite set of constraints that must be satisfied. Any $x \in D$ is called a *solution*; if it satisfies all constraints in $\mathcal{C}$, it is called a *feasible* solution. A feasible solution x^* is *optimal* if $f(x^*) \leq f(x)$ for all feasible solutions x.

A problem with no feasible solutions is called *infeasible*. On the other hand, if there is no lower bound on $f(x)$ over the feasible set, the problem is called *unbounded*. In the following we will assume that an optimization problem is either infeasible, unbounded or has finite optimum value(*).

Since the key to efficiently solve an optimization problem is to exploit its particular structure, optimization methods come in a great variety of forms. Different methods with different properties have been developed in different research communities, yet the key ingredients of optimization methods are quite common among them. In these notes we will briefly introduce two modeling and optimization paradigms, namely *Constraint Programming* and *Mixed Integer Linear Programming*, highlighting commonalities and differences, strengths and weaknesses of all of them. More details can be found in [25].

(*)Dottorato di ricerca in Matematica Computazionale, Dipartimento di Matematica Pura ed Applicata, Università di Padova. E-mail: salvagni@math.unipd.it. Seminar held on 8 April 2009.

(*)This is not always the case. For example we rule out problems like $\min e^{-x}$ subject to $x \geq 0$.

1 Constraint Programming

Constraint Programming (CP) is the study of computational systems based on constraints. It is an emergent paradigm to declarative model and effectively solve large, often combinatorial, optimization problems.

The basic concepts of constraint programming date back to the early studies in the sixties and seventies done by the *artificial intelligence* community [24, 23]. Further steps were achieved when it was noted [9, 13] that logic programming (and declarative programming in general) was just a particular kind of constraint programming (hence the development of *Constraint Logic Programming*). However, this does not mean that constraint programming is restricted to declarative languages (like CHiP [7], CLP($\mathcal{R}$) [13] or Prolog III [5], since constraint programming services can and are actually implemented also for imperative ones (like ILOG Solver [20] or Gecode [10].

Constraint programming deals with *Constraint Satisfaction Problems*. A constraint satisfaction problem (CSP) is defined as a triple $(x, D, \mathcal{C})$, where

- $x = \{x_1, \ldots, x_n\}$ denotes a finite set of *variables*
- $D = D_1 \times \ldots \times D_n$ represents the *domain* of the variables
- $\mathcal{C}$ denotes a finite set of constraints restricting the values that the variables can simultaneously take.

Note that both domains and constraints can be of arbitrary type. A CSP where all domains are finite, although not necessarily numeric, is called a *finite domain* CSP, CSP(FD) for short.

We may want to find one solution, all solutions, or prove that no solution exists. If we are also given an objective function $f : D \to \mathbb{R}$ mapping each complete solution to a real value we may also want to find an optimal (or at least good) feasible solution—in this case we speak of constraint optimization problems.

From the complexity point view, since CSPs include satisfiability problems as special case, we have that CSP is $\mathcal{NP}$-hard.

1.1 CP Solvers

The most common algorithm to solve CSPs is *backtracking search*. This technique consists in extending a (initially empty) partial solution assigning a value from its domain to an uninstantiated variable (this is called *labeling*). If every variable is assigned a value without violating any constraint, then we have found a solution. Otherwise we need to backtrack (undo) one or more assignments and pick different values. This scheme can be organized as a search tree; every node (subproblem) corresponds to a partial assignment and is derived by its parent with the addition of a variable assignment[(*)].

This simple scheme is sufficient to solve CSP, or at least CSP(FD), but it is hardly a practical scheme by itself. The key element for solving CSPs in practice is the use of

[(*)] In practice it is possible and worthwhile to generalize the concept of variable assignment to an arbitrary split of the subproblem space.

constraint propagation. Constraint propagation's task is to analyze the set of constraints and the domain of the current subproblem, and to infer additional constraints and domain reductions. Constraints discovered this way are always *implied* by the original constraint set $\mathcal{C}$, and thus unnecessary to correctly model the problem; yet they are helpful in reducing the amount of search that needs to be done to solve the problem, because they make implicit information explicit. If constraint propagation is used to infer only domain reductions it is called *domain propagation*.

If a constraint optimization problem is to be solved, a constraint

$$f(x) < f(\tilde{x})$$

is added to the set of constraints whenever a new *incumbent* $\tilde{x}$—the best solution available—is found. The search terminates when the feasible set becomes empty.

2 Mixed Integer Linear Programming

A *mixed integer linear program* is defined as follows

$$z = \min c^T x \tag{4}$$

$$Ax \leq b \tag{5}$$

$$l \leq x \leq u \tag{6}$$

$$x \in \mathbb{R}^n \tag{7}$$

$$x_j \in \mathbb{Z}\ \forall j \in I \tag{8}$$

where $A \in \mathbb{Q}^{m \times n}$, $c \in \mathbb{Q}^n$, $b \in \mathbb{Q}^m$, $l \in (\mathbb{Q} \cup \{-\infty\})^m$, $u \in (\mathbb{Q} \cup \{\infty\})^m$ and $I \subseteq N = \{1, \ldots, n\}$. Here $c^T x$ is the *objective function*, $Ax \leq b$ are the linear constraints, l and u are simple *lower* and *upper bounds* on the problem variables x, and I is the subset of indices denoting the variables required to be integer. Integer variables with bounds $0 \leq x_j \leq 1$ are called *binary* variables and play a special role in MIP modeling/solving. We will denote the set of binary variables indices with $B \subseteq I$. Variables not required to be integer are called *continuous* variables.

There are several special cases of the above general model, namely:

- *linear programs* (LP) if $I = \emptyset$
- *integer programs* (IP) if $I = N$
- *binary programs* (BP) if $B = I = N$

Note that a MIP is just a special case, albeit a very important one, of CP, where all constraints and the objective function are required to be linear and we have only integer or real-valued domains. Despite these limitations, MIPs proved to be very effective in modeling and solving both theoretical and practical optimization problems.

Since SAT is a special case of binary program, BP, IP and MIP are $\mathcal{NP}$-hard. Only LPs were shown to be polynomially solvable [14], yet the algorithm used to solve them in practice, the *simplex* [6] algorithm, is non-polynomial.

2.1 MIP solvers

State-of-the-art MIP solvers are based, like CSP and SAT solvers, on a backtracking scheme, namely the *branch-and-bound* (B&B) scheme. This technique optimizes the search exploiting *relaxations*.

The B&B algorithm consists of the following steps:

- At the beginning of the search, we initialize a queue $\mathcal{Q}$ of unprocessed subproblems with the original problem P and set the value z^{best} of the current incumbent to $+\infty$.
- We choose a subproblem Q from $\mathcal{Q}$. If the queue is empty we are done.
- We solve the LP relaxation R of Q, i.e., we solve a relaxed version of Q where we have dropped all integrality constraints. If the relaxation is feasible we denote with x^* and z^* the relaxed solution and its value, respectively.
- If $z^* \geq z^{best}$ than we can discard the current subproblem immediately, because it cannot contain a feasible solution better than the current incumbent (*bounding*).
- If x^* is feasible for the original problem, i.e., x^* satisfies all integrality requirements, then we have found a new incumbent and we can update x^{best} and z^{best}. Otherwise, we split the current subproblem Q into smaller parts and we add them to $\mathcal{Q}$. This is usually done by choosing a variable x_j that has a fractional value x_j^* in the relaxation and by imposing the disjunction $x_j \leq \lfloor x_j^* \rfloor \vee x_j \geq \lceil x_j^* \rceil$ (we say that we have *branched* on variable x_j).

The effectiveness of the B&B scheme depends on the convergence rate of primal (z^{best}) and dual (the smallest relaxation value z^* of unprocessed nodes) bounds. These bounds are affected by many factors:

- The node selection strategy used to choose the next subproblem Q to extract from $\mathcal{Q}$. There are two main possible strategies to visit the search tree. One is *best-first* search, where the node with best dual bound (lowest z^*) is chosen, while the other is *depth-first* search, where the deepest unprocessed node is chosen. The first strategy tries to move the dual bound as quickly as possible, thus minimizing the number of nodes of the tree, but in doing so it is quite slow at finding new incumbents and is more memory demanding. On the other hand, depth-first has lower memory requirements and a greater node throughput. It tends to find primal solutions earlier in the search; however, a bad branching decision can make the dual bound move very slowly. State-of-the-art MIP solvers use a hybrid strategy combining the characteristics of both methods [1].
- The branching strategy, i.e., how we choose the variable x_j to branch on. We would like to choose the variable that leads to the highest dual bound change, but this may be very difficult to estimate. What is done in practice is to select a manageable list of "candidate" variables, and estimate the change of LP bound that would occur branching on them. Many methods have been studied to get a good estimate in a reasonable amount of time, among them *strong branching*, *pseudocost branching*, and *reliability branching* [1].

- How fast we find an incumbent solution providing a good primal bound. Just waiting to find good feasible solutions at B&B nodes by solving LP relaxations has proven not to be effective, so MIP solvers resort to *primal heuristics* [4]. These heuristics are used both to find a first feasible solution to the model, with techniques such as rounding, partial enumeration and diving or to improve existing ones, usually through local search algorithms.

- How tight is the LP relaxation of the subproblems. There are essentially two techniques to improve the LP relaxation of a model, namely *preprocessing* and *cutting planes*. Preprocessing can alter quite significantly the formulation of a problem with techniques such as fixing, aggregation, substitution, redundancy detection, coefficient reduction, and bound strengthening, without changing the set of feasible (in case of primal reductions) or optimal (in case of dual reductions) solutions [22]. Cutting planes are linear constraints that are satisfied by all feasible solutions but not by every solution of the LP relaxation, hence the strengthening of the LP relaxation. Cutting planes can be generated "a priori" or "on the fly" in order to cut a fractional vertex of the LP relaxation (*separation*). Cutting planes implemented in MIP solvers include both generic cuts valid for any MIP problem, like Gomory mixed-integer and mixed-integer rounding cuts [15, 17] and strong polyhedral cuts studied for particular linear constraints. The combination of cutting planes and B&B was introduced successfully in [19, 18] under the name *Branch & Cut* (B&C) and is the algorithm of choice implemented in state-of-the-art MIP solvers.

3 Integration

Despite their differences, the two aforementioned paradigms all share the same key ingredients, namely *search*, *inference* and *relaxation*.

Search is the enumeration of problem *restrictions*, each of which is obtained by adding constraints to the original problem P. The solution space of a given optimization problem is almost invariably exponentially large and highly irregular, and a natural strategy is to split this search space into smallest pieces, to be solved separately, and to pick the best solution found. As we have seen, a very common search strategy is *branching search*, which recursively splits the feasible set of a given problem until the resulting subproblem is easy to solve or proven infeasible. As we have seen, branching search—combined with appropriate inference and relaxation techniques—is at the basis of the most efficient strategies for solving CP problems (backtracking algorithm) and MIP problems (branch-and-bound or branch-and-cut algorithms). However, branching search is not the only option: another general approach is that of *constraint-directed search*, in which once a restriction is solved a constraint is generated that excludes that restriction and possibly others that are proven to be no better (such constraint is named *nogood*). The search continues until the nogoods collected in the process exclude all the search space. Moreover, if the nogoods are written to impose bounds on the optimal value instead of excluding solutions (in this case they are named *nogood bounds*), then we have that the nogood set can be seen as a problem relaxation. SAT solving algorithms such as *dynamic backtracking* [11], as well as decom-

position techniques for MIPs, such as Benders' decompositions [3] and its generalizations, can be seen as cases of constraint-directed search.

Inference is the act of revealing implicit constraints from the existing ones, in order to reduce the search space. The connection between inference and optimization is fundamental; indeed, any optimization problem can in principle be expressed as an inference problem, namely the problem of deriving from the constraints the best lower bound on the objective function valid for the solution space. Inference is used extensively in all three paradigms we have seen, although in different ways. The rich modeling language of the constraint programming paradigm allows one to fully exploit the structure of the constraints, by means of sophisticated and powerful constraint propagation algorithms. These structures are often lost when translating these constraints (if possible) into linear inequalities. The unrestricted nature of CP constraints is also its biggest weakness, because it prevents inferences for the problem as a whole. On the other hand, MIP solver have to deal only with linear inequalities, for which global inference methods, such as the Farkas' lemma and Gomory cutting planes separators, are known.

Relaxation means replacing the solution space of our problem with a larger, but more easily searchable, one. Relaxations provide information to guide (through relaxed solutions) and accelerate (through bounds) the search phase. Linear relaxations are the key ingredient of B&B and B&C algorithms for MIPs. The existence of a good (and fast) relaxation is one of the biggest advantages of the MIP paradigm over CP. Due to the very generic nature of CP constraints, a relaxation of the problem is often unavailable to CP solvers. As such, CP solvers can perform much weaker global inferences and dual reductions.

In the last years, several researchers in the CP and MIP communities have investigated the possibility to integrate the methodologies proper to these paradigms [12, 16]. Such an integration has the potential to yield important benefits, and the literature on this topic is recent but growing. From the modelling/programming point of view, many CP languages include nowadays Operation Research techniques for the propagation of global constraints [16]. Global constraints enable the use of local relaxations of structured constraints, providing precious information for cost-based filtering [8]. Some languages enable the use of hybridization techniques, where different solvers work on distinct parts of the optimization process. Several methods combine CP with linear or Lagrangian relaxation, and maintain a unique model that is used by distinct solvers that cooperate through common data structures, such as the domains of the variables and their extremes, while others combine MIP and CP through generalizations of Benders decomposition. More recently, the Constraint Integer Programming [1] and search-infer-and-relax [12] paradigms proposed a higher level of integration between CP and MIP. Both paradigms are currently implemented in experimental codes, namely SCIP and SIMPL, respectively.

References

[1] Tobias Achterberg, Constraint Integer Programming, PhD thesis, Technische Universität Berlin, 2007.

[2] Tobias Achterberg and Timo Berthold, *Improving the feasibility pump.* Discrete Optimization 4 (2007), 77–86.

[3] J. F. Benders, *Partitioning procedures for solving mixed-variables programming problems.* Numerische Mathematik 4 (December 1962), 238–252.

[4] Timo Berthold, Primal Heuristics for Mixed Integer Programs, Master's thesis, Technische Universität Berlin, 2006.

[5] A. Colmerauer, "An introduction to Prolog III". Draft, 1987.

[6] G. B. Dantzig, "Linear Programming and Extensions". Princeton University Press, Princeton, New Jersey, 1963.

[7] Mehmet Dincbas, Pascal Van Hentenryck, Helmut Simonis, Abderrahmane Aggoun, T. Graf, and F. Berthier, *The constraint logic programming language CHIP.* In *FGCS*, Tokyo, Japan (1988), pages 693–702.

[8] Filippo Focacci, Andrea Lodi, and Michela Milano, *Cost-based domain filtering.* In *Principles and Practice of Constraint Programming*, Springer-Verlag (1999), 189–203.

[9] Hervé Gallaire, *Logic programming: Further developments.* In *SLP* (1985), 88–96.

[10] Gecode Team, "Gecode: Generic constraint development environment, 2008". Available from `http://www.gecode.org`.

[11] Matthew L. Ginsberg, James M. Crawford, and David W. Etherington, "Dynamic backtracking". Technical report, 1996.

[12] J. N. Hooker, "Integrated Methods for Optimization". Springer, 2006.

[13] Joxan Jaffar and Jean-Louis Lassez, *Constraint logic programming.* In *POPL* (1987), 111–119.

[14] L. G. Khachian, *A polynomial algorithm in linear programming.* Soviet Mathematics Doklady 20(1) (1979), 191–194.

[15] H. Marchand and L. A. Wolsey, *Aggregation and mixed integer rounding to solve MIPs.* Operations Research 49(3) (2001), 363–371.

[16] M. Milano, "Constraint and Integer Programming: Toward a Unified Methodology". Kluwer Academic Publishers, 2003.

[17] G. Nemhauser and L. A. Wolsey, *A recursive procedure to generate all cuts for 0-1 mixed integer programs.* Mathematical Programming 46 (1990), 379–390.

[18] M. Padberg and G. Rinaldi, *Optimization of a 532-city symmetric traveling salesman problem by branch and cut.* Operations Research Letters 6 (1987),1–8.

[19] M. W. Padberg and Rinaldi G., *A branch-and-cut algorithm for the resolution of large-scale symmetric traveling salesman problems.* SIAM Review 33 (1991).

[20] Jean-Francois Puget, *A C++ implementation of CLP.* In *Proceedings of the Second Singapore International Conference on Intelligent Systems*, Singapore, 1994.

[21] Francesca Rossi, Peter van Beek, and Toby Walsh, editors, "The Handbook of Constraint Programming". Elsevier, 2006.

[22] M. W. P. Savelsbergh, *Preprocessing and probing for mixed integer programming problems.* ORSA Journal on Computing 6 (1994), 445–454.

[23] Ivan E. Sutherland, *Sketchpad: a man-machine graphical communication system.* In *DAC '64: Proceedings of the SHARE design automation workshop*, New York, NY, USA, 1964. ACM.

[24] David Waltz, *Understanding line drawings of scenes with shadows.* In *The Psychology of Computer Vision*, McGraw-Hill, 1975.

[25] Domenico Salvagnin, Constraint Programming Techniques for Mixed Integer Linear Programs, PhD thesis, University of Padova, 2009.

A Global Framework for Multiobjective Optimization

Alberto Lovison (*)

Abstract. In real life situations, there are usually more than one objective to deal with in order to design successful projects. For instance, a good car should be fast while keeping fuel consumption as low as possible, should protect occupants maintaining compact external dimensions, et cetera. Multiobjective optimization defines the mathematical framework for dealing with such problems. In this talk we propose a gentle introduction to this subject recalling the notion of Pareto critical set, introduced by Stephen Smale, and illustrate an effective global search algorithm. We discuss advantages and drawbacks of this method, mentioning open issues involving the curse of dimensionality, Morse theory and singularity theory.

Introduction

We introduce multiobjective optimization having in mind mainly engineering design situations, like designing aircrafts, cars, electronic devices, and so on. In these situations there are usually many design parameters (input variables controllable by the designes) but also many fitness or performance criteria, each one defined by one or more performance indicators. Rarely, one can define an *overall* performance indicator, but even in this cases it is very difficult to make explicit its dependence on design parameters.

Originated from the work of Vilfredo Pareto in economics, multiobjective optimization is a discipline stating that in several situations, driving the optimization process following single performance indicators could not be the best strategy, but rather suggests to try to improve several indicators at the same time, producing different solutions, and at a later stage to make the final choice.

An Olympic analogy

In this section we introduce Pareto optima as a possible solution to the problem of defining the winners of athletic competitions. We ask if it could be possible to define in some sense an "absolute gold medal", i.e., an athlete somehow superior to everyone. Maybe we should

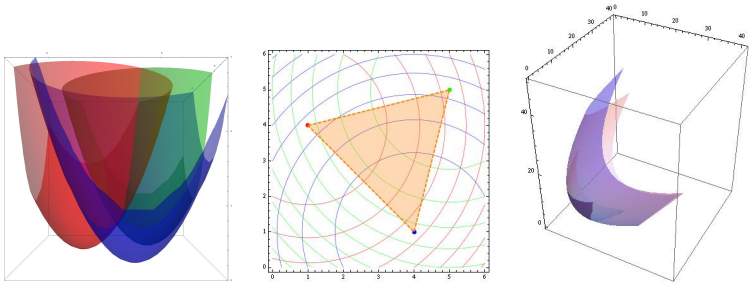
(*)Dipartimento di Matematica Pura ed Applicata, Università di Padova, via Trieste 63, I-35122 Padova (Italy). E-mail: lovison@math.unipd.it. Seminar held on 13 May 2009.

ask him to be able to win every kind of athletic competition, or at least a large part of them. Apart from very special cases, it is very unlikely for a single athlete to win in more than a competition. We could then try to define which competition is the most important, but such an approach will certainly involve subjective or arbitrary choices and a mathematical discussion of the problem becomes clumsy if not impossible to handle. Thus, at a first stage, our simplified mathematical model should elect a different equivalent gold medal for each different competition.

Next, we consider only three competitions (e.g., running, swimming, cycling) assuming that each function will associate a performance to each athlete in terms of a real number. The winner will be the athlete realizing the smallest (or larger) performance. We then assume that athletes are infinite and could be parametrized by means of few (say two) real continuous variables. Our caricature of the Olympic games consists then in a vector function:

$$u : W \subseteq \mathbb{R}^2 \longrightarrow \mathbb{R}^3, \qquad \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \longmapsto \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} (x_1, x_2).$$

A further simplification states that the u_i's are positive definite quadratic functions, as represented in Figure 1(a).



(a) (b) (c)

Figure 1. Simplified version of the Olympic games. (a) Each competition is represented by a positive definite quadratic functional. (b) Optima in Pareto sense span the full triangle having the optima of the single functionals as vertices. (c) Mapping of the domain on the performances spaces. The solid surface is the image of the Pareto optima, i.e., the Pareto front.

It seems clear that only in pathological cases the three optima coincide or are aligned, usually we should find three optima in general position, i.e., three champions. The existence of an *Atleta di Taranto* is a zero probability event. From an unbiased point of view each winner deserves an equivalent medal. A question naturally arises: what should be said about athletes running faster than the swimming champion and at the same time swimming faster than the running champion? Maybe it is possible to tailor a suitable *triathlon* fitting to such a person, i.e., a suitable (linear convex) combination of the three contests:

$$u(x, y)_{[\alpha,\beta,\gamma]} = \alpha u_1(x, y) + \beta u_2(x, y) + \gamma u_3(x, y), \qquad \alpha, \beta, \gamma \geq 0.$$

Every possible choice for α, β, γ potentially elects its own winner. This does not mean that *everybody* is the champion of a suitable contest. Elementary considerations lead to the conclusion that the set of champions is not the whole $\mathbb{R}^2$, but rather the full triangle having the individual optima as vertices. The set of (weak) optima is called *Pareto efficient set*, or *Pareto optima*, while the set of optimal vector values is the mapping of the optima in $\mathbb{R}^3$ and is called *Pareto front*, or *Pareto frontier*.

Geometry and topology of Pareto sets

In the previous example Pareto sets show nontrivial geometrical structure and regularity, i.e, they resemble surfaces with piece–wise smooth boundaries. Is it the same in the general case? If so, can we define such regularity more precisely? Are these properties stable, i.e., are these sets homotopically or isotopically equivalent under small perturbation of the functionals? In fact, Smale, de Melo and Wan [S73, S75, dM76a, dM76b, W75] showed that under mild transversality conditions on u, the set of Pareto optima is a *stratified set* in the sense of Thom, or an $(m-1)$*–manifold with corners*, in the sense of Cerf.

From these results descend important consequences on numerical methods, i.e., the possibility of discuss in some sense the convergence of algorithms to portions of surfaces. On the other hand, heuristic approaches to Pareto optimization usually offer scattered sets of points as optima and cannot recognize connected components of Pareto optimal sets. Possibly, substructures can be highlighted via heuristic approaches like clusterization, but they require an almost exhaustive exploration, involve fine tuning of parameters or arbitrary choices.

The Smale framework for Pareto optimization

We will try to define a *discretization* of the Pareto efficient set in terms of *simplicial complexes* (i.e., a mesh). A more ambitious aim is to open the road to a numerical translation of Smale's theory, defining a vector analogue of Newton methods for scalar optimization, i.e., following gradient descent directions, choosing the step length according to curvature information, and so on. Furthermore, the application of vector versions of topological methods like Morse theory could be applied to prove existence or multiplicity results for physical or industrial applications. We start from the classical definitions by V. Pareto [P97].

Definition 1 (Pareto dominance). *x dominates $\bar{x}$ if $u_1(x) \geq u_1(\bar{x}), \ldots, u_n(x) \geq u_n(\bar{x})$ and exists j such that $u_j(x) > u_j(\bar{x})$.*

It is easy to verify that Pareto dominance is a pre–order.

Definition 2 (Pareto optima) *x_p is Pareto optimum (*max*) if $\nexists x \in D$ such that x is Pareto–dominated by x_p*

Pareto maxima cannot be dominated by other points, while Pareto minima cannot dominate other points. These definition applies straightforwardly to heuristic strategies, i.e., it is immediate to extract the set of optima from a finite set of points by direct

application of Pareto's definitions. Defining Newton methods obviously require first and second derivatives, i.e., differential translations for these definitions.

In the '70, stimulated by Nobel prize for economics Gerard Debreu, Stephen Smale applied differential topology and global analysis to economics, and defined a theory of critical points for vector functions. In particular, he introduced the notion of *Pareto critical set* θ, having in mind necessary and sufficient conditions for maxima and minima for scalar functions. Quoting Smale [S75]:

> *"We study the local and global nature of θ, as one uses freshman calculus to study the maximum of a single function."*

According to Wan [W75], critical points for scalar functions are the points which satisfy the *necessary conditions* for local optimality. The necessary condition for scalar functions is the vanishing of all derivatives. The differential for vector functions is not supposed to vanish in any point in general, even when the input space is a compact manifold. Otherwise, usually there exists a nontrivial set where the rank of the jacobian is non maximal, the *singular set* Σ. Along the singular set the gradients of the functional are linearly dependent. We observe, recalling Pareto conditions, that in optimal configurations there should not exist any direction along which all functionals improve at the same time. So linear dependence of gradients should be integrated by a positivity condition on the coefficients of the linear combination giving zero.

More precisely, we will consider an n–dimensional compact manifold W as parameters space and assume all functions smooth. The differential necessary condition we stated above reads as follows. Consider the open cone in the tangent space of T_xW: $C_x := Du(x)^{-1}(Pos)$, where Pos is the positive orthant in the output space: $Pos := \left\{ y \in \mathbb{R}^m \middle| \; y_i > 0, \forall i \right\}$.

Definition 3 (Pareto critical set).

$$\theta = \{x \in W | C_x \text{ is empty}\}$$

Example 1 Let $W = \mathbb{S}^2 \hookrightarrow \mathbb{R}^3$ and let $u : \mathbb{S}^2 \to \mathbb{R}^2$ the projection along the vertical direction $u(x_1, x_2, x_3) := (x_1, x_2)^T$. It is easy to verify that the singular set Σ is the equator of $\mathbb{S}^2$, while θ is composed by the two red branches. See Figure 2.

It is straightforward to prove the following:

Proposition 1 (First order proposition). *$x \in \theta$ if and only if:*

(a) *The $Du_i(x)$ do not belong to a unique open half space in $T_x^\star(W)$.*

(b) *$\exists \lambda_i \geq 0$, $i = 1, \ldots, m$, not all zero such that $\sum \lambda_i Du_i(x) = 0$.*

The positivity of the coefficients of the linear combination implies that it is impossible to (infinitesimally) improve one of the functionals without making decrease some of the others.

We need now a criterion for distinguishing the nature of the critical points. A notion of stability is introduced to this aim. The curve $(a,b) \ni t \mapsto \phi(t) \in X$ is said *admissible*, if $\frac{d}{dt}u_i(\phi(t)) = \nabla u_i(\phi(t)) \cdot \phi'(t) > 0$, $\forall t \in (a,b), \forall i = 1,\dots,m$.

Definition 4 (Stable Pareto critical points). *A Pareto critical point x is said* stable, *$x \in \theta_s$, if, given a neighborhood V_x of x in W, exists a neighborhood U_x of x in V_x, such that every admissible curve $\phi : [a,b) \to W$, with $\phi(a) \in U_x$ satisfies* $\mathrm{Image}(\phi) \subset V_x$.

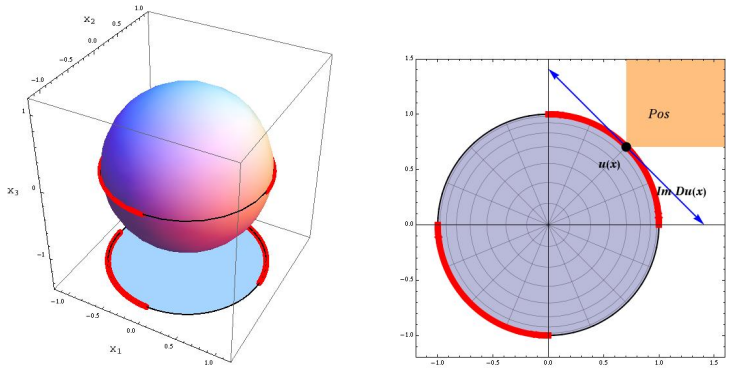


Figure 2. Pareto critical set θ (in red) for the projection along the first two coordinates of a 2–sphere. In black the singular set Σ.

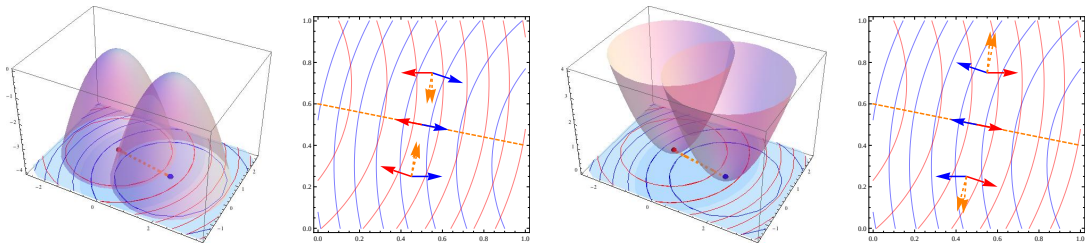


Figure 3. Examples of (a) stable and (b) unstable Pareto critical sets. Admissible curves and Pareto critical sets are drawn in orange.

Stability and instability is illustrated for elementary cases in Figure 3. It is easy to prove that local Pareto optima coincides with stable Pareto critical points. The generalized Hessian can be defined intrinsically along the Pareto critical set, requiring the *rank assumption* $\mathrm{rk}\,(Du) = m - 1$, and restricting the second differential to the kernel of the tangent map.

Definition 5 (Generalized Hessian). *Let $x \in \theta, x \notin \partial\theta$ satisfy rank assumption, then $\mathbb{R}^m / ImDu(x)$ has dimension one and has canonically defined positive ray. The bilinear*

form

$$H_x : \ker Du(x) \times \ker Du(x) \longrightarrow \mathbb{R}^m / ImDu(x)$$
$$(v, w) \longmapsto \langle D^2u(x)v, w \rangle$$

is called 2nd intrinsic derivative *of Porteous and Mather.*

The index and nullity of H_x are well defined. Pareto stability can be formulated in terms of the Hessian.

Proposition 2 (Second order proposition). *Let* $x \in \theta, x \notin \partial\theta$ *and* $\text{corank}Du(x) = 1$. *Then,*

(a) *If the generalized Hessian* H_x *is negative definite then* $x \in \theta_s$.

(b) *Let* $\lambda_i \geq 0$ *as in the 1st order proposition; then:*

$$H_x = \sum \lambda_i D^2 u_i(x), \qquad (on\ \ker Du(x))$$

A global algorithm for Pareto sets

We propose now a numerical translation of *"1st order proposition"*, i.e., a strategy for tracing the Pareto critical set in the simplest case: $W = [0,1]^2$, $m = 2$, $u : [0,1]^2 \to \mathbb{R}^2$. The singular set Σ is the set where the determinant of the Jacobian is zero, while the Pareto critical set is the subset of Σ where gradients oppose each other, i.e., $Du_1(x) \cdot Du_2(x) < 0$. The algorithm simply extends a standard strategy for tracing contour plots:

Algorithm 1 Global search strategy for Pareto sets

1: Consider a tessellation of the domain $[0,1]^2$.
2: Compute derivatives of u, $\det J_x$ and $Du_1(x) \cdot Du_2(x)$ for every node.
3: Consider the cells where $\det J_x$ changes sign. Find Σ by inverse linear interpolation.
4: Among Σ cells mark Pareto cells θ as which where $Du_1(x) \cdot Du_2(x) < 0$.
5: Refine the global tessellation or the Pareto cells.

The workings and the convergence of the algorithm depends on the stability of the mapping u, which is guaranteed for almost all functions in the simple 2–dimensional case we considered.

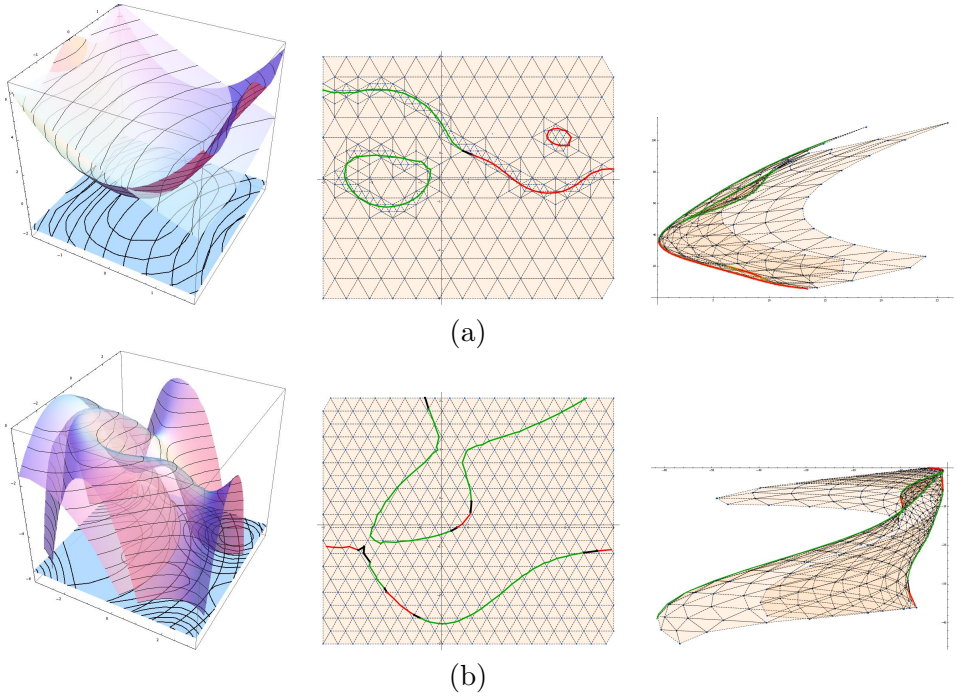
In Figure 4 are plotted the outcomes of the application of the algorithm described above on the following vector functions.

Example 2 Let $u_1 = x_1^2 + x_2^2 + 4\left(e^{-(x_1+2)^2 - x_2^2} + e^{-(x_1-2)^2 - x_2^2}\right), u_2 = (x_1-6)^2 + (x_2+0.5)^2$. The lack of convexity for u_1 produces a small local critical set (see Figure 4(a)).

Example 3 The *Poloni function* is a combination of trigonometric functions and a quadratic one:

$$\begin{aligned} u_1 = &- 1 - (0.873 - 2.\cos(x_1) - 1.5\cos(x_2) + 0.5\sin(x_1) + \sin(x_2))^2 \\ &- (2.748 - \cos(x_1) - 0.5\cos(x_2) + 1.5\sin(x_1) + 2\sin(x_2))^2, \\ u_2 = &- (3 + x_1)^2 - (1 + x_2)^2. \end{aligned}$$

The algorithm reveals five distinct branches (see Figure 4(b)).



(a)

(b)

Figure 4. Applications of the global algorithm for Pareto critical sets. (a) Example 2. (b) Example 3.

Perspectives

The implementation can be extended straightforwardly to higher input dimensions, however, usual restrictions due to the curse of dimensionality have to be considered, i.e., the complexity grows dramatically with input dimension making unrealistic the implementation of global search algorithms for more than seven input variables.

Serious problems are encountered also for higher output dimension, indeed the volume of m–dimensional manifolds with boundary concentrates in an ϵ neighborhood of their boundary, if m large. Being global Pareto fronts included in the boundary of the image of the vector function, we observe that a large fraction of the observations lie on the Pareto front. As a result, very refined meshes, i.e., enormous datasets, are necessary for discriminating optima from non optimal configurations.

Other severe limitation descend from singularity theory. De Melo [dM76a] proved that there exists an open and dense subset of $C^\infty(W, \mathbb{R}^3)$ whose mappings are θ–stable. Roughly speaking, "close" mappings determine "close" and homotopical Pareto critical sets. As a result, θ–stability guarantees some kind of topological convergence for the algorithm described above, e.g., the number of branches composing θ is determined by relatively coarse meshes. Moreover, the piecewise approximations are homotopically equivalent to θ from a certain stage on. However, θ–stability cannot hold for structurally unstable singularities, which occur for non–zero measure subsets of $C^\infty(W, \mathbb{R}^m)$ when the number of functions m is larger than 6 (see [A68]).

Finally, Smale [S73] delineates a generalization of Morse theory for vector functions. Wan [W75] gives a detailed proof for Morse inequalities in the case of two functions, proving that under generic conditions θ splits into a finite number of smooth arcs along which the index is constant. The minimum number of those arcs is related to the Betti numbers of the manifold W by means of the Morse theory. It is possible that this and other topological methods could be applied to vector functions leading to existence and multiplicity results for physical or industrial problems.

References

[A68] V. I. Arnol'd, "Singularities of Smooth Mappings". Russ. Math. Surv., 1968.

[dM76a] W. de Melo, *Stability and optimization of several functions.* Topology (1976).

[dM76b] W. de Melo, *On the Structure of the Pareto Set of Generic Mappings.* Bol. Soc. Bras. (1976).

[P97] V. Pareto, "Cours d'économie politique". 1897.

[S73] S. Smale, *Global analysis and economics I: Pareto optimum and a generalization of Morse theory.* Dynamical Systems, New York (1973).

[S75] S. Smale, *Optimizing Several Functions.* Manifolds, Tokyo (1975).

[W75] Y. H. Wan, *Morse Theory for Two Functions.* Topology (1975).

A Basic Introduction to Moduli Spaces

Ernesto C. Mistretta [(*)]

Abstract. We will explain the meaning of moduli spaces as spaces parametrizing geometrical objects, giving some well known examples such as Projective Spaces. We will focus our attention on the moduli space of triangles, constructing it and elucidating problems of symmetries and monodromy, that appear in more sophisticated cases.

Introduction

Our starting point in the Theory of Moduli Spaces is the investigation of the meaning of "classifying objects". By objects we mean the "geometrical objects" of a given Category, *i.e.* where isomorphisms are given.

By Classification we mean a procedure that consists in different steps:

- 1^{st} step:
  (a) finding/describing/constructing one (and only one) representative for each isomorphism class of the given objects;
  (b) given an object, being able to tell which is the corresponding representative, and to construct an isomorphism with that representative.
- 2^{nd} step:
  Give a geometrical structure to the set X of isomorphism classes, reflecting how "we move" among these objects in the following sense: a family of objects $\{Y_b\}_{b\in B}$ parametrized by B should correspond bijectively to the map $b \mapsto [Y_b]$ from B to X, this map respecting the geometrical structure of X (*i.e.* being continuous, or algebraic, or differentiable, etc.).

In this notes, addressed to Ph.D. students in all disciplines of mathematics, we will try to illustrate how to proceed with these 2 steps with some particularly simple examples: Platonic solids, rotations of the plane, projective spaces, and we will illustrate how to construct the moduli space of triangles, trying to shed light on the meaning of the difficulties arising in these simple constructions, that are to be found in all coarse moduli spaces.

[(*)]Dipartimento di Matematica Pura ed Applicata, Università di Padova, via Trieste 63, I-35122 Padova (Italy). E-mail: ernesto@math.unipd.it. Seminar held on 20 May 2009.

We will not introduce the concept of stacks or orbifolds, even though the examples treated lead naturally to that construction, and the main example, the moduli space of triangles, is developed from an idea of Behrend, Conrad, *et al.* to illustrate the concept of stacks (cf. [BCE$^+$]).

The notions required are the basic notions of a Topology course.

1 Some (not so) simple (as they seem) examples

In order to illustrate the procedure described above, we will start by giving some well known examples:

- platonic solids;
- rotations of the plane.

1.1 Platonic Solids

The classification of Platonic Solids is known since the ancient times: there are 5 classes of isomorphisms of regular solids. Let us see what it does mean in the sense of the classification problem.

We can define the family of Platonic Solids as the family of convex polyhedrons $P \subset \mathbb{R}^3$ which are regular, in the sense that the faces of a Platonic solid are congruent regular polygons, with the same number of faces meeting at each vertex. Then isomorphisms are bijective maps $P \to P'$ obtained as the restriction to P a composition of homothethies translations and orthogonal morphisms of $\mathbb{R}^3$.

In this sense the set of isomorphism classes is a set of 5 elements, distinguished by the number of faces:

As the number of faces is a finite discrete set, there is no way to "pass continuously" from a solid to another, but considering properly the moduli space of Platonic Solid is not as simple as considering a discrete space with five elements.... We will see why, when describing the moduli space of triangles.

1.2 Rotations of the Plane

Another well known set of geometrical object is the set of rotations of the plane about the origin.

Once we fix a base for the plane, we can identify this set with a group of matrices and we know that these are 2×2 matrices of the form

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

such that ${}^tA \cdot A = \mathrm{Id}$ and $\det(A) = 1$, and that these must have the form

$$A = \begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix}$$

for some $\vartheta \in [0, 2\pi]$. We can identify then the set of rotations of the plane with $S^1 = \mathbb{R}/2\pi\mathbb{Z}$.

We define a *family of rotations over a topological space* T, to be a family

$$\left\{ A_t = \begin{pmatrix} a(t) & b(t) \\ c(t) & d(t) \end{pmatrix} \mid A_t \cdot A_t = \mathrm{Id}\ ,\ \det(A_t) = 1 \right\}_{t \in T}$$

such that the function $t \mapsto (a(t), b(t), c(t), d(t))$ be a continuous map $T \to \mathbb{R}^4$.

Then we see that for every topological space T we have a bijection

$$\begin{array}{ccc} \{\text{families of rotations over } T\} & \leftrightarrow & \{\text{continuous functions } T \to S^1\} \\ \left\{ A_t = \begin{pmatrix} \cos\varphi(t) & -\sin\varphi(t) \\ \sin\varphi(t) & \cos\varphi(t) \end{pmatrix} \right\}_{t \in T} & \leftarrow\!\shortmid & (\varphi \colon T \to \mathbb{R}/2\pi\mathbb{Z}) \end{array}$$

This means that S^1 is a *fine moduli space of rotations of the plane*. In the next section we will analyze in more details what this concept means, showing that projective spaces are fine moduli spaces of lines in a vector space.

2 Projective spaces

We know that real projective spaces

$$\mathbb{P}^n = \mathbb{P}(\mathbb{R}^{n+1}) = \{\ell \subset \mathbb{R}^{n+1} \mid \ell \text{ is a line through the origin}\}\ ,$$

and in general $\mathbb{P}(V)$ for a finite dimensional real vector space V, have their own topology, differential structure, algebraic structure, etc.

By construction this space parametrizes the lines in a given vector space. We want to see this space as a moduli space of lines in V, in the sense of the two steps described in the introduction.

The first step is obtained by construction. As for the second step, the geometry of the projective space is given by the classical constructions, for example considering the

quotient $(V - \{0\})/\mathbb{R}^*$. To understand how "we move" in this geometry, we have to describe what families of lines in V parametrized by a base B are. In order to do this we have to define vector bundles over a topological space B, which are basically families of vector spaces parametrized by B, and vector bundle morphisms.

Definition 2.1 *A* real vector bundle *of rank k over a topological space B is a continuous map of topological spaces $\pi\colon E \to B$ satisfying the following properties:*

(i) *$\forall p \in B$ the fibers $\pi^{-1}(p)$ are real vecotr spaces of dimension k;*

(ii) *there exists an open covering $\{U_i\}_{i\in I}$ of B, and homeomorphisms $\varphi_i\colon \pi^{-1}(U_i) \xrightarrow{\sim} U_i \times \mathbb{R}^k$ such that:*

(a) *the diagrams*

$$\begin{array}{ccc} U_i & \xrightarrow[\cong]{\varphi_i} & U_i \times \mathbb{R}^k \\ {\scriptstyle\pi}\downarrow & \swarrow_{pr_1} & \\ U_i & & \end{array}$$

commute for all i;

(b) *the morphisms $\varphi_{i|\pi^{-1}(p)}\colon \pi^{-1}(p) \xrightarrow{\sim} p \times \mathbb{R}^k$ are linear for all $i \in I$ and $p \in U_i$.*

Definition 2.2 *A* morphism *of vector bundles $\pi\colon E \to B$ and $\pi'\colon E' \to B$ over B is a continuous map $\varphi\colon E \to E'$ such that:*

(i) *$\pi = \pi' \circ \varphi$;*

(ii) *$\forall p \in B$ the map $\varphi_{|\pi^{-1}(p)}\colon \pi^{-1}(p) \to (\pi')^{-1}(p)$ is linear.*

A vector bundle morphism is called injective, surjective, or an isomorphism*, if the map φ is injective, surjective, or a homeomorphism.*

Example 2.3 Given any real vector space V of finite dimension k, the first projection $pr_1\colon B \times V \to B$ is a vector bundle on th topological space B, called *trivial vector bundle* with fiber V.

Example 2.4 Given a differentiable manifold M of dimension n, the tangent bundle $\mathcal{T}M$ is a vector bundle of dimension n on M. If the differentiable manifold M is a submanifold of $\mathbb{R}^N$, then the differential of the inclusion $f\colon M \subset \mathbb{R}^N$ gives an injective morphism of vector bundles $df\colon \mathcal{T}M \hookrightarrow \mathcal{T}\mathbb{R}^N = \mathbb{R}^N \times \mathbb{R}^N$.

Definition 2.5 *A* family of lines *over a topological space B is a rank* 1 *vector bundle over B. A* family of lines in a vector space V *over a topological space B is a rank* 1 *vector bundle $L \to B$ over B together with an injective morphism φ in the trivial bundle with fiber V on B, $\varphi\colon L \hookrightarrow B \times V$.*

Two families of lines in V over B, $\varphi\colon L \hookrightarrow B \times V$ and $\varphi'\colon L' \hookrightarrow B \times V$, are isomorphic if there exist an isomorphism of vector bundles $\psi\colon L \to L'$ such that $\varphi' \circ \psi = \varphi$.

We have a particular family of lines in V over $\mathbb{P}(V)$, called *tautological family* or *universal fmily*:

$$\mathcal{H} := \{(\ell, v) \in \mathbb{P}(V) \times V \mid v \in \ell\} \subset \mathbb{P}(V) \times V \ .$$

Given a continuous map $f\colon A \to B$ and a family of lines in V over B, $\pi\colon L \to B$, we can construct a family of lines in V over A, called the pull back f^*L of L to A, in the following way:

$$f^*L := \{(x, t) \in A \times L \mid f(x) = \pi(t) \in B\} \subset A \times V \ ,$$

where the map $f^*L \to A$ is the restriction of the first projection.

Given a family of lines $\pi\colon L \to B$ in V over B, we have a map:

$$\begin{array}{rccc} \Psi_\pi & : & B & \to & \mathbb{P}(V) \\ & & b & \mapsto & (\pi^{-1}(b) \subset V) \end{array}$$

After proving that this map is continuous, and that two isomorphic families induce the same map, one can prove that the family $\Psi_\pi^*\mathcal{H}$ over B is isomorphic to the family $\pi\colon L \to B$ of lines in V over B. This proves the following

Theorem 2.6 *For every topological space B there is a (natural) bijection*

$$\begin{array}{ccc} \{\textit{families of lines in } V \textit{ over } B\}/\sim & \leftrightarrow & \{\textit{continuous functions } B \to \mathbb{P}(V)\} \\ (\pi\colon L \to B) & \mapsto & \Psi_\pi \\ \Phi^*\mathcal{H} \subset B \times V & \leftarrow\!\shortmid & \Phi\colon B \to \mathbb{P}(V) \end{array}$$

that is, $\mathbb{P}(V)$ is a fine moduli space *for lines in V.*

Remark 2.7 The term *natural* used in the theorem means that for any couple of topological spaces A and B, and any continuous function $f\colon A \to B$, the following diagram is commutative:

$$\begin{array}{ccc} \{\text{families of lines in } V \text{ over } B\}/\sim & \xrightarrow{\sim} & \{\text{continuous functions } B \to \mathbb{P}(V)\} \\ \Big\downarrow f^* & & \Big\downarrow -\circ f \\ \{\text{families of lines in } V \text{ over } A\}/\sim & \xrightarrow{\sim} & \{\text{continuous functions } A \to \mathbb{P}(V)\} \end{array}$$

where f^* is the map that associate to a family $L \to B$ of lines in V over B the family f^*L of lines in V over A, and $- \circ f$ is the map that associate to a function $\varphi\colon B \to \mathbb{P}(V)$ the function $\varphi \circ f\colon A \to \mathbb{P}(V)$.

In the language of categories the last theorem shows that the functor from topological spaces to sets, that associates to any topological space B the set of isomorphism classes of family of lines in V over B, is *represented* by $\mathbb{P}(V)$ with its universal family. This means exactly that $\mathbb{P}(V)$ is a *fine moduli space* for lines in V.

3 Triangles

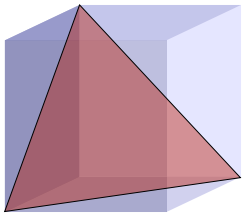
We want to proceed now with the classification illustrated in the introduction for our toy example, the construction of the moduli space of triangles.

Our spaces are plane triangles, and isomorphisms between triangles are bijective maps which are restriction to a triangle of a composition of translations homotheties and orthogonal endomprphisms of the plane.

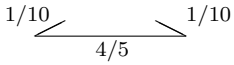
3.1 The Coarse Moduli Space of Triangles

As for the first step we want to construct a family of representative for each triangle. As we allow homotheties, we will consider only triangles whose perimeter is 1. Two such triangles are isomorphic if and only if they are congruent, and we know since middle school that 2 such triangles are congruent if and only if they have the 3 edges of the same length. We can describe all such triangles by the lengths of their 3 edges, say (a, b, c).

We have then a subset of $\mathbb{R}^3$, parametrizing all points (a, b, c) with $a + b + c = 1$, and $a, b, c > 0$, this is the interior of the standard 2-dimensional simplex $\Delta_2 \subset \mathbb{R}^3$:



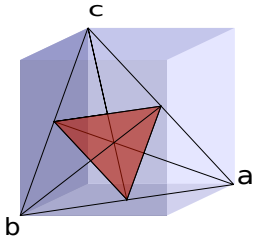
Obviously not every point $(a, b, c) \in \mathrm{Int}(\Delta_2)$ of this set correspond to the lenght of the edges of a triangle, *e.g.* $(4/5, 1/10, 1/10)$:



However if we restrict to the points satisfying the conditions "every edge is smaller than the sum of the other edges":

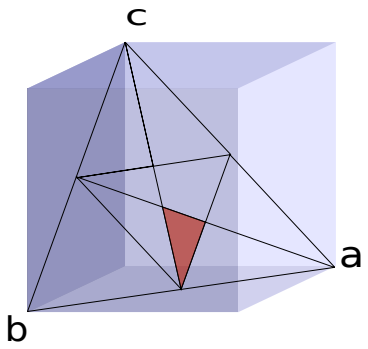
$$a + b > c \ , \ a + c > b \ , \ b + c > a$$

we see that every point of this set correspond to a triangle of perimeter 1, and to any such triangle correspond (at least) one point of this set:



In order to to fulfill the characterisics demanded on step 1 and have only one point for every equivalence class, we have to avoid the points that correspond to the same three edges with a different order, hence we can restrict to the points satisfying also:

$$a \geqslant b \geqslant c \ ,$$



In this way we constructed a set $\mathcal{T} \subset \mathbb{R}^3$ whose points are in one-to-one correspondence with the equivalence classes of triangles. This set being (a triangle!) in $\mathbb{R}^3$, it carries in a natural way a topology. As in the case of projective spaces, in order to give a meaning to this topology (when are two triangles near? what is a curve in this space?), we have to consider families of triangles.

3.2 Families of Triangles

We have to define families of triangles of perimeter 1 varying continuously over a topological space.

We have to define what metrics on vector bundles are:

Definition 3.1 *Let $\pi\colon E \to B$ be a vector bundle over B, then a* metric *on E is the data of scalar products $g_b\colon E_b \times E_b \to \mathbb{R}$ on the vector spaces $E_b := \pi^{-1}(b)$ for all $b \in B$, such that for any pair of sections of π, $s, t\colon U \to E$ on an open set $U \subseteq B$, the function $b \mapsto g_b(s(b), t(b))$ is a continuous map $U \to \mathbb{R}$. Then (E, g) is called a* metric vector bundle.

We want to define families $T \to B$ such that the fibers T_b are triangles of perimeter 1 varying in a continuous way:

Definition 3.2 *A family of triangles over a topological space B is a pair (E, T), where $\pi\colon E \to B$ is a metric vector bundle of rank 2, and $T \subset E$ is a subset such that:*

(i) *For every $p \in B$, $T \cap E_b$ is a set of 3 non collinear points, forming a triangle with perimeter 1;*

(ii) *$\pi_{|T}\colon T \to B$ is a topological covering space.*

We say that 2 families (E, T) and (E', T') of triangles over B are isomorphic if there exists a homeomorphism $\varphi\colon T \to T'$ such that:

(i) *if* $\pi\colon E \to B$ *and* $\pi'\colon E' \to B$ *are the 2 vector bundles, then* $\pi' \circ \varphi = \pi_{|T}$*;*

(ii) *the map* φ *induce affine isometries between* E_p *and* E'_p *for all* $p \in B$*.*

In particular an isomorphism of families of triangles (E, T) *and* (E', T') *over* B *induces an isomorphism of covering spaces of* $T \to B$ *and* $T' \to B$*.*

We call $T_p := T \cap E_p$ the triangle over the point $p \in B$. As E_p is a metric space for all $p \in B$, then we have the lengths of the three sides of the triangle and we call them $a(p) \geqslant b(p) \geqslant c(p)$. By construction the point $(a(p), b(p), c(p))$ belongs to $\mathcal{T}$.

Hence, we have a function

$$\begin{array}{rccc} \Upsilon_\pi & : & B & \to & \mathcal{T} \\ & & p & \mapsto & (a(p), b(p), c(p)) \end{array}$$

As the map $\pi_{|T}\colon T \to B$ is a topological covering space, choosing a trivialization $\{U_i\}_{i\in I}$ for $\pi_{|T}$ and for the vector bundle $\pi\colon E \to B$ at the same time, we can see locally the family of triangles as moving in one fixed plane V, *i.e.* as three continuous function $x, y, z\colon U_i \to V$. This proves that the function Υ_π is continuous.

3.3 Is there a Fine Moduli Space of Triangles?

We have shown that there exists a space $\mathcal{T}$ whose points are in one-to-one correspondence with isomorphism classes of triangles. And that we have a continuous map

$$\begin{array}{rccc} \Upsilon_\pi & : & B & \to & \mathcal{T} \\ & & p & \mapsto & (a(p), b(p), c(p)) \end{array}$$

for any family of triangles $\pi\colon T \subset E \to B$ over B.

After defining pull-backs of families, as in the case of the projective space, one can check that the map

$$\begin{array}{ccc} \{\text{families of triangles over } B\}/\sim & \longrightarrow & \{\text{continuous functions } B \to \mathcal{T}\} \\ (\pi\colon T \subset E \to B) & \longmapsto & \Upsilon_\pi \end{array}$$

is natural in the sense of Remark 2.7

Definition 3.3 *The fact that the space* $\mathcal{T}$ *verifies*

(i) *existence of such a* natural *map* $(\pi\colon T \subset E \to B) \mapsto \Upsilon_\pi$*,*

(ii) *a* bijective *correspondence between* $\mathcal{T}$ *and the isomorphism classes of triangles, or, more precisely, the bijectivity of the map above when the space* B *consists in 1 point,*

is denoted by saying that the space $\mathcal{T}$ *is a* coarse moduli space of triangles*.*

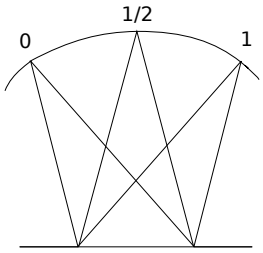
We may ask whether we have the same properties as in the case of projective spaces, *i.e.* if we constructed a *fine moduli space* for triangles. To verify whether this is the case, we have to see whether the map

$$\begin{array}{ccc} \{\text{families of triangles over } B\}/\sim & \longrightarrow & \{\text{continuous functions } B \to \mathcal{T}\} \\ (\pi\colon T \subset E \to B) & \longmapsto & \Upsilon_\pi \end{array}$$

is bijective.

We will see that this map is not injective in general, by the following example.

Example 3.4 Let us consider the following family of triangles over S^1: fix an ellipse such that the triangle formed by a point on the ellipse and the two focuses has perimeter 1. Fix a triangle on this ellipse and consider its symmetric triangle with respect to the axe of the ellipse passing between the 2 focuses as in the picture below:



then parametrize the family by moving the vertex on top along the ellipse, as to pass at time 0 by the first vertex, and at time 1 at its symmetric.

As we have isomorphic triangles at points 0 and 1, this is a family over S^1, let us call this family $p\colon \mathcal{P} \to S^1$.

Then consider another family obtained as before, but moving the vertex form 0 to $1/2$ as in figure, and then coming back to 0. Call this family $q\colon \mathcal{Q} \to S^1$. Note that over the point $1/2$ we have an isosceles triangle.

The two families coincide in the interval $[0, 1/2]$, while above the points t in the interval $[1/2, 1]$ the triangles $p^{-1}(t)$ and $q^{-1}(t)$ are isomorphic as they are symmetric with respect to the axe passing between the 2 focuses.

Then the maps $\Upsilon_p\colon S^1 \to \mathcal{T}$ and $\Upsilon_q\colon S^1 \to \mathcal{T}$ coincide, but the two families are not isomorphic, in fact the two coverings of S^1 are not isomorphic: the covering $q\colon \mathcal{Q} \to S^1$ is a trivial covering, as it can be observed following the orbit of the 3 vertices under the action of the fundamental group of S^1. While the covering $p\colon \mathcal{P} \to S^1$ is not trivial: the action of the fundamental group of S^1 exchanges the vertices on the base of the triangle.

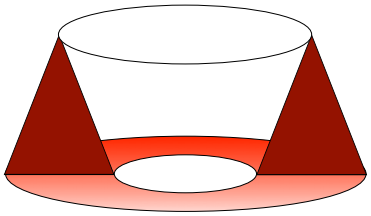
One can wonder whether this lack of injectivity is due to the construction of our space $\mathcal{T}$, and whether there might be another space $\mathcal{S}$ which is a fine moduli space for triangles, *i.e.* we could have another natural bijection

$$\begin{array}{ccc} \{\text{families of triangles over } B\}/\sim & \xrightarrow{\sim} & \{\text{continuous functions } B \to \mathcal{S}\} \\ (\pi\colon T \subset E \to B) & \longmapsto & \Gamma_\pi \end{array}$$

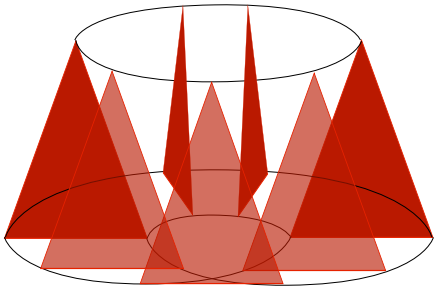
But this is not the case. In fact if we want any space $\mathcal{S}$ to parametrize triangles and families of triangles *naturally* (as in the sense of Remark 2.7), then for any family of triangles $p\colon T \to B$ the map *must* associate to $b \in B$ the point in $\mathcal{S}$ representing the triangle $p^{-1}(b)$. We can see this in another example of two non isomorphic families giving rise to the same map.

Example 3.5 We will consider two other families of triangles over S^1 giving rise to a constant map with value the point corresponding to an isosceles triangle T_0 with edges of legth (a, a, b).

The first one is the trivial family $r\colon S^1 \times T_0 \to S^1$:



The second one is a family $s : T \to S^1$ obtained as the first one on the interval $[0, 1/2]$, and then by flipping the isosceles triangle T_0 along its symmetry axe over the interval $[1/2, 1]$ as in the picture below:



These two families are non isomorphic, as the second one is not trivial. This can be seen in many ways, either observing that the covering $s : T \to S^1$ is non trivial as it exchanges the vertices at the base of the isosceles triangle, or remarking that the boundaries of the triangles describe 2 topological surfaces. This surface being clearly a torus in the case of the trivial family, and containing a Möbius strip (described by the base of the isosceles triangles) in the case of the second family, hence being non orientable (it is a Klein bottle indeed).

However the two families induce the same map, the constant map

$$S^1 \to \{(a, a, b)\} \subset \mathcal{T},$$

as the fibers are all isomorphic to the isosceles triangle T_0.

4 Conclusions

As it can be seen in the last example, the reason why the space $\mathcal{T}$ is a coarse moduli space, but not a fine one, is to be found in the existence of symmetries of triangles. In

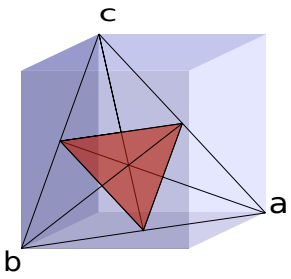
fact, whenever there is a triangle symmetry, we can define a family on S^1, having that symmetry as monodromy action of the corresponding 3-folded covering.

In fact we can construct a fine moduli space $\widetilde{\mathcal{T}}$ of "labelled triangles" that parametrizes triangles with an ordering of the edges (not necessarily by their length). Leaving to the reader to fill up the details, we see that this moduli space construct by parametrizing all the possible length of the 3 triangles edges (a, b, c) without imposing the conditions

$$a \geqslant b \geqslant c \ ,$$

as we did before:

$$\widetilde{\mathcal{T}} := \{(a, b, c) \in (0, 1)^3 \mid a + b > c \ , \ a + c > b \ , \ b + c > a\}$$



Then a (coarse) moduli space of triangles is obtained $\widetilde{\mathcal{T}}$ by forgetting the labeling of the edges, hence by quotienting $\widetilde{\mathcal{T}}$ by the action of $\mathcal{S}_3$. This action can be seen on the picture above as the action of the symmetries of the red triangle. The points which have a non-trivial stabilizer for this action are exactly the points corresponding to a triangle that has symmetries, an isosceles or equilateral triangle, and this points are the ones lying on the axes of the red triangle in the picture above.

The deep theory that provides the techniques to treat this problems is the theory of algebraic stacks (in the algebraic framework) or orbifolds (in the topological/differential setting). We do not pretend to discuss here any aspect of this theory, but we say that it allows to take into account in quotient spaces also the stabilizers for the action of certain groups, and not only the orbits for this action. So the moduli orbifold for triangles is to be considered this quotient orbifold noted $[\widetilde{\mathcal{T}}/\mathcal{S}_3]$, while the coarse moduli space is the topological quotient $[\widetilde{\mathcal{T}}/\mathcal{S}_3] \cong \mathcal{T}$.

In the same way, the moduli orbifold of platonic solids must take into account their symmetries. For example the moduli space orbifold parametrizing tetrahedrons will be the orbifold quotient $[\{*\}/\mathcal{S}_4]$ of a point for the (trivial) action of $\mathcal{S}_4$ on this point, while the coarse moduli space will be a single point. And so on for the other 4 solids (with different groups acting).

The problems that appeared in the examples treated here are to be found also in more complicated theories, such as the theory of moduli spaces of algebraic curves, where the concept of algebraic stack first appeared.

We refer to the work-in-progress book written by Kai Behrend, Brian Conrad, Dan Edidin, William Fulton, Barbara Fantechi, Lothar Göttsche and Andrew Kresch [BCE$^+$] for the interested reader (provided he has some notions of algebraic geometry).

References

[BCE+] Kai Behrend, Brian Conrad, Dan Edidin, William Fulton, Barbara Fantechi, Lothar Göttsche, and Andrew Kresch, "Algebraic stacks". Preprint available on `http://www.math.uzh.ch/index.php?ve_vo_det&key2=580&no_cache=1`.

Social interactions and heterogeneous agent models. Applications to economics and finance.

Marco Tolotti (*)

Abstract. Relying on my work in the field of contagion models, based on interacting particle systems, I will discuss some open issues concerning the applicability of complex systems in economics and finance. I will present some applications of a class of Markov models that are in line with recent research in economic theory. In particular I will highlight the importance of modeling social interactions, bounded rationality, heterogeneous agents and random utilities.

Sunto. Prendendo spunto dalla ricerca svolta nell'ambito dei modelli di contagio, basati su sistemi di particelle interagenti, discuterò una serie di questioni aperte sull'applicabilità di sistemi complessi alla teoria economica e alla finanza. In particolare presenterò qualche applicazione di una classe di modelli Markoviani che ben si inseriscono in un nascente filone di ricerca in ambito economico che vuole catturare aspetti non convenzionali quali interazioni sociali, razionalità limitata, eterogeneità degli agenti, casualità nei processi decisionali.

Keywords: intensity-based models, mean-field interactions, non-reversible Markov processes, random utility models, social interactions.

1 Introduction

Do *social interactions* and *agents' heterogeneity* really matter in economic theory? For a long period, the neoclassical (general equilibrium) approach with complete markets has provided a negative answer.

Social interactions and heterogeneity are only two of the aspects analyzed into the range of the so called *complex systems*. The debate concerning the applicability of complex systems in economics goes back to the eighties and the nineties and addresses some open issues left unsolved by the *neoclassical* approach based on the representative (perfectly rational) agent. Just to give the intuition, we could say that the neoclassical approach is

(*)Università Ca' Foscari di Venezia, Dipartimento di Matematica Applicata, Dorsoduro 3825/E - 30123 Venezia, Italy - tolotti@unive.it. Seminar held on 3 June 2009. This seminar is based on joint works with E. Barucci, P. Dai Pra, W. Runggaldier and E. Sartori.

based on a strong *rationality* paradigm that can be translated into the following modeling assumptions: perfect information shared by all the agents (no learning is needed), no space for errors in the decision process, no influence of the society on the personal beliefs.

On the opposite, the complex approach has been motivated by the desire to explain some empirical phenomena that can not be explained under the strict hypothesis of prefect rationality and efficiency of the markets. We are referring for instance to *i*) the statistical properties of financial time series (power laws with heavy tails in the stock returns, high temporal correlations, scaling laws); *ii*) path dependence and non-ergodicity in the pattern of the adoption of a new technology; *iii*) social behaviors as herding, peer pressure and conformity effects. A more detailed analysis of these issues can be found in [4, 5].

Here we focus on some of these aspects. In particular we shall describe a framework where *bounded rationality*, *social interactions* and *heterogeneity* can be introduced but maintaining tractability.

2 Random utility models

This is the case of the class of problems we intend to analyze: binary choices with interacting agents and random utility functions. The setting is the one posed in [1, 2].

The economy is made up of I agents facing a discrete binary choice problem: $\omega_i \in \{-1; +1\}$ for $i = 1, ..., I$.

Agents' utility function is made up of three components: private utility, social utility and an error term. For agent i we have that

$$u_i(\omega_i) = v(\omega_i) + J_i \omega_i \bar{m}_i^e + \epsilon(\omega_i). \tag{1}$$

$v(\omega_i)$ is any kind of private utility associated with the binary choice. Being ω a binary variable, we can assume w.l.o.g. $v(\omega) = h\omega + k$, for $h, k \in \mathbb{R}$. Notice that $h > 0$ makes the choice $\omega = 1$ preferable. $S(\omega_i, \bar{m}_i^e) = J_i \omega_i \bar{m}_i^e$ is the social component of the utility where $\bar{m}_i^e$ is the expectation from the point of view of agent i of the behavior of the others. Finally $\epsilon(\omega_i)$ is a random term whose distribution is extreme valued, i.e., $P(\epsilon(-1) - \epsilon(1) \leq x) = 1/(1 + e^{-\beta x})$, where $\beta > 0$ is a measure of the impact of the random component in the decision process: high values of β means that the deterministic part plays a relevant role in the maximization of the utility. Instead when β tends to zero the error term dominates and the choice between $\omega = 1$ or $\omega = -1$ is basically a coin tossing. The error component can be interpreted as a *bounded rationality* effect on the behavior of the agents.

We assume that the propensity to conformity of the agents (labeled by J_i for $i = 1, ..., N$) is not the same for all agents and is not constant. Note that $J_i > 0$ means that the agent is conformational (she tends to follow the social behavior). In what follows, we assume that $J \in \{-1; +1\}$; in this way we capture the presence of conformist and non conformist agents.

In this setting it can be shown that

$$P(\omega_i |\, \bar{m}_i^e\,) = \frac{e^{\beta\omega_i\left(h + J_i \bar{m}_i^e\right)}}{e^{\beta\omega_i\left(h + J_i \bar{m}_i^e\right)} + e^{-\beta\omega_i\left(h + J_i \bar{m}_i^e\right)}}. \tag{2}$$

In [2] it is proved that when agents are homogeneous ($J_i = J$ for all i) and share the same expectations on the choice of the others, i.e. when $\bar{m}_i^e = m$ for all i, then the equilibria of the system are described by the following fixed point argument:

$$m = \tanh(\beta h + \beta J m). \tag{3}$$

The main consequence of this fact is that the equilibria of the system can be one or three depending on the value of β and J. In particular, for high values of β and J there are multiple equilibria.

Some questions arise. Is it possible to characterize the equilibria of this system assuming heterogeneity and a dynamic updating of the beliefs of the agents? What is the relationship between the static equilibria provided by (3) and the steady states found via a dynamic approach? We try to give some insights in these directions.

2.1 Dynamic set up

We denote by $\omega_i(t) \in \{-1; 1\}$, where $i : 1, ..., I$ and $t \in [0, T]$, the choice of the $i - th$ agent at time t and by $\underline{\omega}(t) = (\omega_1(t), ..., \omega_I(t))$ the vector of the state variables (agents' decisions at time t).

We assume that agents update their decisions at random Poissonian times characterized by certain *intensities* or *rates* (the inverse of the average waiting times) that depend on the state of the economy at that time and on the information of the agent.

Inspired by [1], we proposed a dynamic version of equation (2):

$$\lambda_i(t) = \lim_{\tau \to 0} \frac{1}{\tau} P(\omega_i(t+\tau) \neq \omega_i(t) | \, \underline{\omega}(t) \,) = e^{-\beta \omega_i(t) \left(h + J_i(t) \frac{\sum_j \omega_j(t)}{I} \right)}, \tag{4}$$

where now all the state variables are indexed with time and where $\lambda_i(t)$ denotes the local rate of probability that agents i changes his choice between time t and t^+, given the state of the system at time t.

Having in mind applications where the propensity of the agents to conformity is endogenously varying, we define a similar dynamics for the J_i variables:

$$\mu_i(t) = \lim_{\tau \to 0} \frac{1}{\tau} P(J_i(t+\tau) \neq J_i(t) | \, \underline{\omega}(t) \,) = e^{-\gamma J_i(t) \frac{\sum_j \omega_j(t)}{I}}, \tag{5}$$

where γ is a parameter that quantify the dependence of J_i by the state of the world.

We notice that the $i-$th agent's decision depends on the system only through the *aggregate statistic*

$$s_I = \frac{1}{I} \sum_{j=1}^{I} \omega_j.$$

This variable is indeed an *empirical mean* of the system and incorporates only a partial (averaged) information on the state vector $\underline{\omega}$. This simplifying assumption is called *mean field assumption*: the interaction among different agents only depends on the value of s_I. Notice that higher values of s_I imply an higher probability for agent i to choose $\omega_i = 1$

(when $J_i = +1$). On the other hand, high values of s_I make also larger the probability that $J_i = +1$.

Intensities as (4) make the state variables evolve as a continuous-time Markov chain on $\{-1, 1\}^{2N}$ with the following infinitesimal generator:

$$\mathcal{L}_I f(\underline{\omega}, \underline{J}) = \sum_{i=1}^{I} e^{-\beta\omega_i(h+J_i s_I)} \left(f(\underline{\omega}^i, \underline{J}) - f(\underline{\omega}, \underline{J})\right) + \sum_{i=1}^{I} e^{-\gamma J_i s_I} \left(f(\underline{\omega}, \underline{J}^i) - f(\underline{\omega}, \underline{J})\right) \tag{6}$$

where $\underline{\omega}^i$ (resp. $\underline{J}^i$) denotes the vector $\underline{\omega}$ (resp. $\underline{J}$) where the $i-$th component has been switched:

$$\omega_j^i = \begin{cases} \omega_j & \text{for } j \neq i \\ -\omega_i & \text{for } j = i. \end{cases}$$

As argued in [1], heterogeneity leads to what is called *non reversibility* of the dynamical system. Reversibility is related to the shape of the generator (6) that describes the time evolution of the state variables of the system. It can be shown that when the system is reversible (homogeneous) there are standard techniques useful to describe the stationary distributions and hence the steady states (equilibria) of the system. In our case it is not possible to rely on such methodologies; in particular it has not yet been addressed in the literature whether more complex (non reversible) models might exhibit a behavior in line with the findings as in [1, 2]. One of our aims is to explore exactly this conjecture: is it possible to describe general Markov models that exhibit this kind of equilibria linked to the paper by Brock and Durlauf? We shall provide some insights in this directions showing in particular how to take advantage of the Markovian approach.

In order to characterize the equilibria we shall look at the *limiting behavior* of the system, i.e., the dynamics and the steady states of the $I \to \infty$ case. Our findings are in line with (3): for low values of β and γ (low interaction) there exists a unique (stationary) equilibrium; for high values there are instead three equilibria.

3 Results: Dynamics and equilibria

Inspired by [3], we now state a law of large number on a family of probability measures that will help us in determining the dynamics of the population behavior of the system. In what follows we shall denote with $(\omega_i[0,T], J_i[0,T])$ the trajectory on $[0,T]$ of the state indicators of the i-th agent. We also denote with $\mathcal{D}([0,T])$ the Skorohod space of (discontinuous) trajectories on $[0,T]$ endowed with the weak topology. With the notation $\mathcal{M}_1(X)$ we denote the space of probability measures on X.

Let $(\omega_i[0,T], J_i[0,T])_{i=1}^I \in \mathcal{D}([0,T])^{2I}$ denote a path of the system process in the time-interval $[0,T]$ for a generic $T > 0$. We define the so called *empirical measure* of the $I-$dimensional system as

$$\rho_I = \frac{1}{I} \sum_{i=1}^{I} \delta_{(\omega_i[0,T], J_i[0,T])}. \tag{7}$$

We may think of ρ_I as a (random) element of $\mathcal{M}_1(\mathcal{D}([0,T]) \times \mathcal{D}([0,T]))$, the space of probability measures on $\mathcal{D}([0,T]) \times \mathcal{D}([0,T])$ endowed with the weak convergence topology.

Let now q be any probability measure on $\{-1;1\}^2$. Define

$$m_q^\omega := \sum_{\omega,J=\pm 1} \omega\, q(\omega, J),$$

that can be interpreted as the average choice under q.

Theorem 3.1 *Suppose that the distribution at time $t = 0$ of the Markov process $(\underline{\omega}(t), \underline{J}(t))_{t\geq 0}$ with generator (6) is such that the random variables $(\omega_i(0), J_i(0))$, for $i = 1, \dots, I$, are independent and identically distributed with law λ. Then there exists a probability $Q^* \in \mathcal{M}_1(\mathcal{D}([0,T]) \times \mathcal{D}([0,T]))$ such that*

$$\rho_I \;\to\; Q^* \quad \textit{almost surely}$$

in the weak topology. Moreover, if $q_t \in \mathcal{M}_1(\{-1;1\}^2)$ denotes the marginal distribution of Q^* at time t, then q_t is the unique solution of the nonlinear (McKean-Vlasov) equation

$$\begin{cases} \frac{\partial q_t}{\partial t} = \mathcal{L} q_t, t \in [0,T] \\ q_0 = \lambda \end{cases} \tag{8}$$

where

$$\mathcal{L}q(\omega, J) = \nabla^\omega \left[e^{-\beta\omega(h + J m_q^\omega)} q(\omega, J) \right] + \nabla^J \left[e^{-\gamma J m_q^\omega} q(\omega, J) \right] \tag{9}$$

with $(\omega, J) \in \{-1, ; 1\}^2$ and where $\nabla^x f(x,y) = f(-x,y) - f(x,y)$. ∎

Equation (8) describes the dynamics of the system with generator (6) in the limit as $I \to +\infty$. In what follows we characterize the equilibrium points, or stationary (in t) solutions of equation (8), i.e. solutions of $\mathcal{L}q_t = 0$ and, more generally, the large time behavior of its solutions.

Lemma 3.2 *Let q_t be as defined in (8) and define the expectations:*

$$m_t^\omega := \sum_{\omega,J=\pm 1} \omega q_t(\omega, J), \quad m_t^J := \sum_{\omega,J=\pm 1} J q_t(\omega, J), \quad m_t^{\omega J} := \sum_{\omega,J=\pm 1} \omega J q_t(\omega, J). \tag{10}$$

Then equation (8) can be rewritten in the following form:

$$\begin{cases} \dot{m}_t^\omega &= 2S(\beta h)S(\beta m_t^\omega) m_t^{\omega J} - 2S(\beta h)C(\beta m_t^\omega) + \\ &\quad + 2C(\beta h)S(\beta m_t^\omega) m_t^J - 2C(\beta h)C(\beta m_t^\omega) m_t^\omega \\ \dot{m}_t^J &= 2S(\gamma m_t^\omega) - 2C(\gamma m_t^\omega) m_t^J \\ \dot{m}_t^{\omega J} &= 2[S(\beta h)S(\beta m_t^\omega) + S(\gamma m_t^\omega)] m_t^\omega - 2S(\beta h)C(\beta m_t^\omega) m_t^J \\ &\quad + 2C(\beta h)S(\beta m_t^\omega) - 2[C(\beta h)C(\beta m_t^\omega) + C(\gamma m_t^\omega)] m_t^{\omega J}. \end{cases} \tag{11}$$

where $C(x) = \cosh(x)$ and $S(x) = \sinh(x)$. ∎

A numerical inspection of these equations shows that, depending on the values of the parameters, we can have either one unique solution (for low values of β and γ) or three different solutions (for high values of the parameters).

To better understand the qualitative representation of the steady states, we analyze the simplified system where $h = 0$. In this case (11) reduces to the following:

$$\left\{ \begin{array}{lcl} \dot{m}_t^\omega & = & 2S(\beta m_t^\omega) m_t^J - 2C(\beta m_t^\omega) m_t^\omega \\ \dot{m}_t^J & = & 2S(\gamma m_t^\omega) - 2C(\gamma m_t^\omega) m_t^J \\ \dot{m}_t^{\omega J} & = & 2S(\gamma m_t^\omega) m_t^\omega + 2S(\beta m_t^\omega) - 2[C(\beta m_t^\omega) + C(\gamma m_t^\omega)] m_t^{\omega J}. \end{array} \right. \tag{12}$$

Here the dynamics of $(\dot{m}_t^\omega, \dot{m}_t^J)$ does not depend on $\dot{m}_t^{\omega J}$. This means that the differential system (12) is essentially two-dimensional: first one solves the two-dimensional system (on $[-1, 1]^2$)

$$(\dot{m}_t^\omega, \dot{m}_t^J) = V(m_t^\omega, m_t^J), \tag{13}$$

with $V(x, y) = (2\sinh(\beta x)y - 2\cosh(\beta x)x, 2\sinh(\gamma x) - 2\cosh(\gamma x)y)$, and then one solves the third equation in (12), which is linear in $m_t^{\omega J}$.

In particular the solutions of the system $V(x, y) = \underline{0}$ are

$$x = \tanh(\beta x)\tanh(\gamma x)\ ; \quad y = \tanh(\gamma x).$$

Notice that x plays exactly the role of m as in equation (3). Indeed, $x = \tanh(\beta x)\tanh(\gamma x)$ is a fixed point argument for which $x = 0$ is always solution. Moreover, for values of γ and β large enough there are three solutions: $(0, x_1, x_2)$ such that $0 < x_1 < x_2$. These findings are qualitatively in line with the equilibria found in [2] for the static and homogeneous model (see equation (3)).

4 Conclusions

We have presented a dynamic version of a random utility model. The main novelty of our analysis with respect to the literature is that we have introduced *agents' heterogeneity* in the social interaction attitude. Agents are heterogeneous in their degree of conformity/complementarity and this feature changes over time endogenously as a function of the behavior of the agents. We have shown how to introduce this source of complexity maintaining tractability.

A first application of these results to finance can be found in [3], it concerns the analysis of credit portfolio losses. We believe that this dynamic formulation can be fruitfully applied also to the study of complex credit derivatives and to the micro-foundation of prices in financial markets. In the context of social studies we are now applying it to the study of crimes rates and in the context of technology adoption patterns.

References

[1] Blume, L.E. and Durlauf, S., *Equilibrium concepts for social interaction models*. International Game Theory Review, 5 Issue: 3 (2003), 193-209.

[2] Brock, W. and Durlauf, S., *Discrete choice with social interactions*. Review of economic studies 68 (2001), 235-260.

[3] Dai Pra, P., Runggaldier, W.J., Sartori, E., Tolotti, M., *Large portfolio losses, a dynamic contagion model*. The Annals of Applied Probability, Vol. 19, No. 1 (2009), 347-394.

[4] Durlauf, S., *Complexity and empirical Economics*. The Economic Journal 115 (2005), F225-F243.

[5] Hommes, C.H., *Heterogeneous Agent Models in Economics and Finance*. Handbook of Computational Economics, in: Leigh Tesfatsion & Kenneth L. Judd (ed.), Handbook of Computational Economics, edition 1, volume 2, chapter 23, pages 1109-1186 Elsevier, 2006.