

UNIVERSIDAD NACIONAL DE CÓRDOBA
FACULTAD DE MATEMÁTICA, ASTRONOMÍA Y FÍSICA

SERIE “B”
TRABAJOS DE MATEMÁTICA

N° 62/2012

IV Congreso Latinoamericano de Matemáticos

XXIV Encuentro de Estudiantes de Matemática
Unión Matemática Argentina

Notas de Cursos

6 al 10 de agosto de 2012, Córdoba, Argentina.

Carolina Araujo – Miguel Campercholi – Marilina Carena
Camilo Jadur, Jorge Yazlle – Uriel Kaufmann
Gabriela Ovando – Claudia Sagastizábal
Andrea Solotar, Mariano Suárez-Alvarez



Editores: Jorge G. Adrover – Gastón A. García

CIUDAD UNIVERSITARIA – (5000) CÓRDOBA
REPÚBLICA ARGENTINA

Prefacio

Los Congresos Latinoamericanos de Matemáticos (CLAM) son los encuentros matemáticos latinoamericanos de mayor envergadura e importancia en la región: incluyen sesiones de conferencias y comunicaciones en la mayoría de las áreas, junto con una serie de conferencias plenarios dictadas por distinguidos especialistas internacionales y cursos para estudiantes de posgrado y jóvenes investigadores.

Organizado cada cuatro años en alguno de los países de las sociedades miembros de la Unión Matemática de América Latina y el Caribe (UMALCA), los tres primeros congresos se realizaron en Brasil, México y Chile en ese orden. La Unión Matemática Argentina (UMA) obtuvo el honor de organizar el IV CLAM, del 6 al 10 de agosto de 2012.

Como adhesión a este significativo evento, la UMA decidió además modificar la fecha de su Reunión Anual, tradicionalmente a fines de septiembre, y realizarla sólo por 2012 en agosto, más precisamente del 5 al 8, para permitir una amplia participación de la comunidad argentina en ambos encuentros. Una sustancial parte de la Reunión Anual es constituida por la rica oferta de diversos cursos para Profesores y Estudiantes. Muchos de los Profesores a cargo de los cursos correspondientes a esta Reunión han redactado notas que sirven de apoyo a los mismos. En este volumen ponemos a disposición de los asistentes al curso y de potenciales interesados las notas de seis cursos del Encuentro de Estudiantes, a saber:

- “Dinámica Simbólica”, por Camilo Jadur y Jorge Yazlle.
- “Métricas en álgebras de Lie dos pasos nilpotentes”, por Gabriela Ovando.
- “Introducción a los reticulados”, por Miguel Campercholi.
- “Ecuaciones en derivadas parciales utilizando análisis funcional”, por Uriel Kaufmann.
- “Teorema del punto fijo en espacios métricos: una aplicación a fractales”, por Marilina Carena.
- “Anillos y módulos semisimples”, por Andrea Solotar y Mariano Suárez-Alvarez.

Además se incluyen las notas de dos de los cursos del CLAM:

- “Energy Problems: modeling and optimization methods”, por Claudia Sagastizábal.
- “Introduction to the Minimal Model Program in algebraic geometry”, por Carolina Araujo.

Agradecemos a los autores de las diversas notas por su desinteresado esfuerzo; a las instituciones que han permitido la realización de las Reuniones mencionadas y la edición de este volumen; y finalmente a Emilio Lauret por su dedicación en la coordinación de la reunión de todo el material.

Nicolás Andruskiewitsch

Córdoba, 10 de julio de 2012

Contenidos

IV Congreso Latinoamericano de Matemáticos

- *Introduction to the Minimal Model Program in algebraic geometry*
Carolina Araujo 1
- *Energy Problems: modeling and optimization methods*
Claudia Sagastizábal 17

XXIV Encuentro de Estudiantes de Matemática

- *Introducción a los reticulados*
Miguel Campercholi 53
- *Teorema del punto fijo en espacios métricos: una aplicación a fractales*
Marilina Carena 63
- *Dinámica Simbólica*
Camilo Jadur y Jorge Yazlle 81
- *Ecuaciones en derivadas parciales utilizando análisis funcional*
Uriel Kaufmann 117
- *Métricas en álgebras de Lie dos pasos nilpotentes*
Gabriela Ovando 125
- *Anillos y módulos semisimples*
Andrea Solotar y Mariano Suárez-Alvarez 137

INTRODUCTION TO THE MINIMAL MODEL PROGRAM IN ALGEBRAIC GEOMETRY

CAROLINA ARAUJO

ABSTRACT. These are lecture notes for the lectures given at the IV CONGRESO LATINOAMERICANO DE MATEMÁTICOS, Córdoba, Argentina, 2012.

CONTENTS

| | |
|---|----|
| Introduction | 1 |
| 1. The classification problem in Algebraic Geometry | 2 |
| 2. Classification of projective surfaces | 3 |
| 2.1. Intersection theory on surfaces | 3 |
| 2.2. Birational geometry of surfaces | 4 |
| 2.3. The role of the canonical class | 5 |
| 2.4. Birational invariants | 8 |
| 2.5. Kodaira dimension | 8 |
| 3. The MMP in higher dimensions | 9 |
| 3.1. Intersection product and spaces of curves and divisors | 9 |
| 3.2. The first theorems of the MMP | 10 |
| 3.3. Flips | 13 |
| 4. MMP with scaling | 14 |
| Suggested reading | 16 |
| References | 16 |

INTRODUCTION

The aim of these lectures is to introduce graduate students working on algebraic geometry and related fields to the main ideas of the so called Minimal Model Program, giving an overview of the subject. We have chosen to keep the lectures as elementary as possible, and hope that it is accessible to students who have completed a one-year basic course on algebraic geometry. In particular, we expect the students to be familiar with [7], whose notation we will follow throughout the text.

Given the limited time, we will not be able to go into the details of the amazing techniques developed in the context of the Minimal Model Program. Great part of these notes are devoted to reviewing the classification of surfaces, and translating it into modern language, while introducing some of the concepts that are key to the Minimal Model Program. We will spend very little time discussing the more recent and very exciting advances in the subject (specially those in [3]). We hope with these lectures to motivate the beginners to venture into more advanced material on the subject. We will suggest some further reading at the end of the notes.

2010 *Mathematics Subject Classification.* 14E30.

The author was partially supported by CNPq and Faperj Research Fellowships.

Throughout these lectures all varieties are assumed to be irreducible, reduced and defined over the field \mathbb{C} of complex numbers.

Curves and surfaces are always assumed to be irreducible, reduced and projective.

Recall that the Picard group of a projective variety X is the group $\text{Pic}(X)$ of invertible sheaves on X modulo isomorphism. Equivalently, $\text{Pic}(X)$ is the quotient of the group of Cartier divisors on X modulo linear equivalence. By abuse of notation, we often identify a Cartier divisor D on X with its class $[D]$ in $\text{Pic}(X)$.

We denote by Ω_X^1 the sheaf of Kähler differentials of X , and by $\omega_X = \wedge^{\dim X} \Omega_X^1$ its canonical sheaf. We denote by $K_X \in \text{Div}(X)$ any divisor on X such that $\omega_X \cong \mathcal{O}_X(K_X)$, and call it a *canonical divisor* or the *canonical class* of X .

1. THE CLASSIFICATION PROBLEM IN ALGEBRAIC GEOMETRY

We are interested in the following classical problem.

To classify projective varieties up to birational equivalence.

What do we mean by this? Here are some of our goals.

- We want to distinguish varieties by means of *invariants*.
- We want to pick distinguished representatives on each birational class. In some sense, these representatives should be the “simplest” varieties in their class.
- Given a projective variety, we want to understand the birational transformations needed to bring it to a distinguished representative of its class.

First of all, we recall the Hironaka’s famous Resolution of Singularities Theorem: any projective variety X admits a resolution of singularities, i.e., there exists a smooth projective variety \tilde{X} and a birational morphism $f : \tilde{X} \rightarrow X$. So we can always assume we start with a smooth variety, even though, as we shall see, it is unavoidable to work with (mildly) singular varieties in order to achieve our classification goals.

1.1 (Classification of curves). A smooth projective curve is nothing but a compact Riemann surface. Two smooth projective curves are birationally equivalent if and only if they are isomorphic. So there is a unique smooth projective model in each birational class of projective curves. We define *genus* $g(X)$ of a smooth projective curve X as

$$g(X) = h^0(X, \omega_X).$$

This numerical invariant allows us to completely solve the classification problem for curves.

- $g(X) = 0$ if and only if $X \cong \mathbb{P}^1$.
- $g(X) = 1$ if and only if X is an elliptic curve, and there is a 1-dimensional family of those, parametrized by \mathbb{C} (via the j -invariant).
- For each $g \geq 2$, there is an algebraic variety M_g of dimension $3g - 3$ parametrizing smooth projective curves of genus g .

For surfaces the situation is not as simple. Given a smooth projective surface S , we can consider the blowup \tilde{S} of S at a point $P \in S$. This is a smooth projective surface birationally equivalent but not isomorphic to S . It is easy to argue that S is “simpler” than \tilde{S} . It turns out that any smooth projective surface can be obtained from a distinguished representative of its class by a sequence of blowups. Such distinguished representatives are classically called *minimal surfaces*. We will revise and summarize this theory in Section 2, explaining how the 3 goals pointed out in the beginning of the section are achieved in this case.

While the classification of surfaces was established by the Italian school by the beginning of the 20th century, the first developments on the classification problem in higher dimensions started to take place in the beginning of the 1980's with important ideas from Mori and Reid, among others. With the contributions of many algebraic geometers, such as Kawamata, Kollár, Shokurov, just to mention a few, a powerful theory of classification of projective varieties was then developed. This is called the *Minimal Model Program* (MMP for short). The program was fully established for 3-folds by Mori in [11], yielding him the Fields Medal in 1990. We will give an overview of this program in Section 3.

Only part of the 3-fold theory could be carried out to higher dimensions, and new ideas and techniques were required. A major achievement was obtained recently by Birkar, Cascini, Hacon, and McKernan in [3]. We will address this briefly in Section 4.

2. CLASSIFICATION OF PROJECTIVE SURFACES

In this section we review the birational classification of complex projective surfaces. We start by recalling the intersection theory on surfaces. Then we state some of the classical results of Castelnuovo and Enriques. We refer to [7, Chapter V] and [2] for details and proofs.

At the end of the section, we rephrase these results from a modern perspective. This reinterpretation suggests generalizations to higher dimensions, which will be explored in the forthcoming sections.

2.1. Intersection theory on surfaces. Let S be a smooth surface.

Theorem 2.1 (Intersection form on surfaces). *There exists a unique symmetric bilinear form*

$$\cdot : \text{Div}(S) \times \text{Div}(S) \rightarrow \mathbb{Z}$$

satisfying the following conditions.

- (1) *Given $D, D' \in \text{Div}(S)$, the intersection number $D \cdot D'$ depends only on the linear equivalence classes of D and D' .*
- (2) *If C and D are curves on S meeting transversely, then $C \cdot D = \#(C \cap D)$.*

Definition 2.2. Two divisors $D, D' \in \text{Div}(S)$ are said to be *numerically equivalent* if $D \cdot C = D' \cdot C$ for every curve $C \subset S$. In this case we write $D \equiv D'$. We write $\text{Num}(S)$ for the quotient group $\text{Div}(S)/\equiv$. By the Theorem of the base of Néron-Severi, $\text{Num}(S)$ is a finitely generated abelian group. Its rank is called the *Picard number* of S , and is denoted by $\rho(S)$.

Later on it will be important to consider also the $\rho(S)$ -dimensional \mathbb{R} -vector space $N^1(S) := \text{Num}(S) \otimes_{\mathbb{Z}} \mathbb{R}$. The intersection form on S induces a nondegenerate symmetric bilinear form $\cdot : N^1(S) \times N^1(S) \rightarrow \mathbb{R}$.

Example 2.3. $S = \mathbb{P}^2$. In this case $\text{Pic}(S) = \mathbb{Z} \cdot [H]$, where $H \subset \mathbb{P}^2$ is a hyperplane section. The intersection form on S is given by $H^2 = 1$.

Example 2.4 (Hirzebruch surfaces). Let $n \in \mathbb{Z}$ be a non-negative integer, and consider the Hirzebruch surface $\mathbb{F}_n = \mathbb{P}(\mathcal{O}_{\mathbb{P}^1} \oplus \mathcal{O}_{\mathbb{P}^1}(n))$, with structure morphism $\pi : \mathbb{F}_n \rightarrow \mathbb{P}^1$. There is a section $\sigma \subset \mathbb{F}_n$ of π such that $\sigma^2 = -n$. If $n \geq 1$, then such section is unique. Moreover, if $\sigma' \subset \mathbb{F}_n$ is a section of π different from σ , then $\sigma'^2 \geq n$. We denote by $F \cong \mathbb{P}^1$ a fiber of π . We have

$$\text{Pic}(\mathbb{F}_n) = \mathbb{Z} \cdot [F] \oplus \mathbb{Z} \cdot [\sigma],$$

and the intersection form on \mathbb{F}_n is given by

- $F^2 = 0$;
- $F \cdot \sigma = 1$; and
- $\sigma^2 = -n$.

For $n = 0$, $\mathbb{F}_0 \cong \mathbb{P}^1 \times \mathbb{P}^1$. For $n = 1$, \mathbb{F}_1 is isomorphic to the blowup of \mathbb{P}^2 at one point, and under this isomorphism σ corresponds to the exceptional divisor. For $n \geq 2$, \mathbb{F}_n admits the following geometric realization. Let $\nu_n : \mathbb{P}^1 \rightarrow \mathbb{P}^n$ be the n^{th} Veronese embedding of \mathbb{P}^1 , given by $(s : t) \mapsto (s^n : s^{n-1}t : \cdots : st^{n-1} : t^n)$. Let $Z \subset \mathbb{P}^{n+1}$ be the cone over $\nu_n(\mathbb{P}^1)$ with vertex P . Then \mathbb{F}_n is isomorphic to the blowup of Z at the point P , and under this isomorphism σ corresponds to the exceptional divisor.

Example 2.5 (Blowups). Let S be a smooth surface and $P \in S$ a point. Let $\pi : \tilde{S} \rightarrow S$ be the blowup of S at P . We denote by $E = \pi^{-1}(P) \cong \mathbb{P}^1$ the exceptional divisor of π .

We have

$$\text{Pic}(\tilde{S}) = \pi^* \text{Pic}(S) \oplus \mathbb{Z} \cdot [E],$$

and the intersection form on \tilde{S} is given by

- $\pi^* D \cdot \pi^* D' = D \cdot D'$ for every $D, D' \in \text{Div}(S)$;
- $\pi^* D \cdot E = 0$ for every $D \in \text{Div}(S)$; and
- $E^2 = -1$.

Similarly, $N^1(\tilde{S}) = \pi^* N^1(S) \oplus \mathbb{R} \cdot [E]$, and thus $\rho(\tilde{S}) = \rho(S) + 1$. In this sense S is simpler than \tilde{S} .

The following concept is very important.

Definition 2.6. Let S be a smooth surface. We say that a divisor $D \in \text{Div}(S)$ is *nef* if $D \cdot C \geq 0$ for every curve $C \subset S$.

Examples 2.7.

- (1) Ample divisors are nef. We will see in Theorem 3.4 that nef divisors can be characterized as limits of ample divisors.
- (2) Let $\pi : S \rightarrow X$ be a morphism into a projective variety X , and let H be an ample Cartier divisor on X . Then $\pi^* H$ is a nef divisor on S . It is ample if and only if π is finite onto its image.

2.2. Birational geometry of surfaces. We now come the problem of determining the “simplest model” in each birational class of surfaces.

Definition 2.8. A smooth surface S is called a *minimal surface* if the following condition holds. If $\pi : S \rightarrow S'$ be a birational morphism onto another smooth surface, then π is an isomorphism.

Given a smooth surface S , how to determine whether it is a minimal surface? We recall that the structure of birational morphisms between smooth projective surfaces is well understood.

Theorem 2.9 (Factorization of birational morphisms of surfaces). *Let $\pi : S \rightarrow S'$ be a birational morphism between smooth surfaces. Then π is the composition of a finite number of blowups.*

So, to show that S is a minimal surface, one must show that S is not the blowup of any smooth surface. How do we check this condition?

Definition 2.10. A curve C on a smooth surface S is said to be a *-1-curve* if $C \cong \mathbb{P}^1$ and $C^2 = -1$.

As we saw in Example 2.5, if S is the blowup of a smooth surface, then it contains a -1 -curve, namely the exceptional divisor of the blowup. The next theorem says that the converse is true.

Theorem 2.11 (Castelnuovo’s contractibility theorem). *Let S be a smooth surface, and $C \subset S$ a -1 -curve. Then there exists a smooth projective surface S' and a point $P \in S'$ such that S is isomorphic to the blowup of S' at P , and under this isomorphism C corresponds to the exceptional divisor.*

In the situation of Theorem 2.11, we say that S' is obtained from S by contracting the -1 -curve C . Now we can state the classical MMP for surfaces.

2.12 (MMP for surfaces - classical version).

- (1) Start with a smooth projective surface S .
- (2) Ask: Does S contain a -1 -curve? If not, stop! S is a minimal surface. If yes, pick one such curve C and go to (3).
- (3) By Castelnuovo’s contractibility theorem, there is a blowup $f : S \rightarrow S'$ for which C is the exceptional divisor. Go back to (1) with S replaced with S' .

This process must stop after a finite number of steps because the Picard number, which is a positive integer, drops by one every time we contract a -1 -curve.

The MMP for surfaces provides a first step in the classification of surfaces: it tells us how to obtain a minimal surface in the birational class of any given surface. Then we ask the following natural questions:

- (1) Is the minimal surface in a given birational class unique?
- (2) Can we classify minimal surfaces in terms of some numerical invariants?

These questions were also classically answered by the Italian school. We will give the answers by the end of this section. At this point, we anticipate that the answer to question (1) depends on the birational class of the given surface S . More precisely, it depends on the behavior of the canonical class K_S . In fact, as we shall see shortly, the whole MMP for surfaces may be reformulated in terms of numerical properties of the canonical class.

2.3. The role of the canonical class. As described in 2.12, it is not at all clear how to generalize the MMP for surfaces to higher dimensions. More precisely, how to generalize the question “Does S contain a -1 -curve?” to higher dimensions? Our next goal is to rephrase the MMP for surfaces in such a way that it makes sense in arbitrary dimension. The key role will be played by the canonical class K_S of a smooth surface S and its numerical properties.

We start by recalling a very useful result.

Theorem 2.13 (Adjunction formula for surfaces). *Let $C \subset S$ be a curve, and $p_a(C) = h^1(C, \mathcal{O}_C)$ the arithmetic genus of C . Then*

$$2p_a(C) - 2 = (K_S + C) \cdot C.$$

Remark 2.14. Let $C \subset S$ be a curve. Then $p_a(C) = 0$ if and only if $C \cong \mathbb{P}^1$.

Examples 2.15.

- (1) $S = \mathbb{P}^2$. (Notation as in Example 2.3.) The canonical divisor is $K_{\mathbb{P}^2} = -3H$.
- (2) $S = \mathbb{F}_n$. (Notation as in Example 2.4.) The adjunction formula applied to F and σ yields:

$$K_{\mathbb{F}_n} = -2E - (2 + n) \cdot F.$$

- (3) Let $\pi : \tilde{S} \rightarrow S$ be the blowup of a smooth surface. (Notation as in Example 2.5.) The canonical divisor of \tilde{S} is given by $K_{\tilde{S}} = \pi^*K_S + E$.

Now we start rephrasing the classical MMP for surfaces in terms of numerical properties of the canonical class. The first step is to give a numerical characterization of -1 -curves.

Exercise 2.16. Let S be a smooth surface, and $C \subset S$ a curve. Show that

$$C \text{ is a } -1\text{-curve} \iff K_S \cdot C < 0 \text{ and } C^2 < 0.$$

It follows from Exercise 2.16, that if K_S is nef (see Definition 2.6), then S is necessarily a minimal surface. The converse is not true, as we shall see in Exercise 2.18.

Definition 2.17. A smooth surface is said to be a *scroll* if there exists a surjective morphism $\pi : S \rightarrow B$ onto a smooth curve B whose fibers are all isomorphic to \mathbb{P}^1 . In this case, it can be shown that there exists a rank 2 vector bundle E on B such that $S \cong \mathbb{P}(E)$. Moreover, if E' is another rank 2 vector bundle on B , then $\mathbb{P}(E) \cong \mathbb{P}(E')$ if and only if there is a line bundle L on B such that $E' \cong E \otimes L$. In particular, since every vector bundle on \mathbb{P}^1 decomposes as a direct sum of line bundles, rational scrolls are precisely the Hirzebruch surfaces.

A surface birationally equivalent to a scroll is called a *ruled surface*. A surface birationally equivalent to \mathbb{P}^2 is called a *rational surface*.

Exercise 2.18.

- (1) Show that \mathbb{P}^2 and scrolls are minimal surfaces, except for \mathbb{F}_1 .
- (2) Verify that if $S \cong \mathbb{P}^2$ or S is a scroll, then K_S is not nef.
- (3) Let $\pi : S \rightarrow B$ and $\pi' : S' \rightarrow B'$ be scrolls. Show that S and S' are birationally equivalent if and only if $B \cong B'$.

Conversely, it can be shown that the only minimal surfaces whose canonical classes are not nef are \mathbb{P}^2 and scrolls. (See Theorem 2.26 for a more precise statement.)

To distinguish between minimal surfaces with K_S nef and not nef, we introduce the following concept, which will generalize to higher dimensions.

Definition 2.19. We say that a surface S is a *minimal model* if K_S is nef.

Exercise 2.20. Let S and S' be birationally equivalent surfaces. Suppose that S and S' are minimal models. Show that $S \cong S'$.

(Hint: use the following structure theorem for birational maps between smooth surfaces: If $\varphi : S \dashrightarrow S'$ is a birational map between smooth surfaces, then there exist compositions of blowups $f : \tilde{S} \rightarrow S$ and $g : \tilde{S}' \rightarrow S'$, and isomorphism $\psi : \tilde{S} \rightarrow \tilde{S}'$ such that $\varphi = g \circ \psi \circ f^{-1}$.)

Remark 2.21. At this point one may ask whether a minimal model may be birational equivalent to a scroll. The answer is no. At the end of this section we will introduce birational invariants that can be used to distinguish between these two types of surfaces.

Next we introduce the *Mori cone* of a surface.

Definition 2.22. Let S be a smooth surface. The *Mori cone* of S is the closed convex cone $\overline{NE}(S) \subset N^1(S)$ generated by classes of curves $C \subset S$.

Definition 2.23. An *extremal face* F of a cone $N \subset \mathbb{R}^n$ is a subcone of N satisfying:

$$u, v \in N \text{ and } u + v \in F \implies u, v \in F.$$

A 1-dimensional extremal face of N is called an *extremal ray*.

Let $D : \mathbb{R}^n \rightarrow \mathbb{R}$ be a linear function. We write $N_{D \geq 0}$ for $\{z \in N \mid D(z) \geq 0\}$, and similarly for $N_{D=0}$, $N_{D \leq 0}$, etc. An extremal face $F \subset N$ such that $F \setminus \{0\} \subset N_{D < 0}$ is called a D -negative extremal face. If $F \subset N_{D=0}$, then we say that F is supported on D .

Exercise 2.24.

- (1) Let S be a smooth surface and $P \in S$ a point. Let $\pi : \tilde{S} \rightarrow S$ be the blowup of S at P . We denote by $E = \pi^{-1}(P) \cong \mathbb{P}^1$ the exceptional divisor of π . Show that $[E] \in N^1(\tilde{S})$ generates an extremal ray of $\overline{NE}(\tilde{S})$.
- (2) Let $n \in \mathbb{Z}$ be a non-negative integer, and consider the Hirzebruch surface $\mathbb{F}_n = \mathbb{P}(\mathcal{O}_{\mathbb{P}^1} \oplus \mathcal{O}_{\mathbb{P}^1}(n))$, with structure morphism $\pi : \mathbb{F}_n \rightarrow \mathbb{P}^1$. Denote by σ a section of π such that $\sigma^2 = -n$, and by F a fiber of π . Recall that $\{[\sigma], [F]\}$ is a basis for $N^1(S)$. Since $\overline{NE}(S)$ is a closed convex cone in a 2-dimensional vector space, it must have exactly 2 extremal rays. Show that these are generated by $[\sigma]$ and $[F]$.
(Hint: for $n \geq 1$, consider the structure morphism $\pi : \mathbb{F}_n \rightarrow \mathbb{P}^1$ and the blowup $f : \mathbb{F}_n \rightarrow Z$ onto the cone over $\nu_n(\mathbb{P}^1)$ described in Example 2.4.)

In general, the cone $\overline{NE}(S)$ may be “round”, and an extremal ray $R \subset \overline{NE}(S)$ may not be generated by the class of a curve. (See [9, Example 1.23].)

We observe that in Exercise 2.24 every extremal ray $R \subset \overline{NE}(S)$ is generated by the class of a curve, and moreover there exists a morphism $\varphi : S \rightarrow Y$ with the following property. For any curve $C \subset S$, $\varphi(C)$ is a point if and only if $[C] \in R$. This motivates the following definition.

Definition 2.25. Let S be a smooth surface, and F an extremal face of the Mori cone $\overline{NE}(S)$. A *contraction of F* is a morphism with connected fibers $\varphi_F : S \rightarrow Y$ onto a normal projective variety Y satisfying following property. For any curve $C \subset S$, $\varphi_F(C)$ is a point if and only if $[C] \in F$.

If the contraction of an extremal face of $\overline{NE}(S)$ exists, then it is unique by Stein factorization. The next theorem asserts that if R is a K_S -negative extremal ray, then the contraction of R always exists. It also gives a complete description of the contraction in this case.

Theorem 2.26. *Let S be a smooth surface, and R a K_S -negative extremal ray of the cone $\overline{NE}(S)$. Then $R = \mathbb{R}_{\geq 0}[C]$ for some rational curve $C \subset S$ (with $K_S \cdot C < 0$). Moreover, the contraction φ_R of R exists, and is one of the following:*

- (1) *If $C^2 < 0$, then $\varphi_R : S \rightarrow S'$ is the blowup of a smooth surface S' at one point, and C is the exceptional divisor.*
- (2) *If $C^2 = 0$, then $\varphi_R : S \rightarrow B$ realizes S as a scroll over a smooth curve B , and C is a fiber of φ_R .*
- (3) *If $C^2 > 0$, then $S \cong \mathbb{P}^2$, and $\varphi_R : \mathbb{P}^2 \rightarrow pt$.*

Definition 2.27. We call the morphisms of type (2) and (3) above *Mori fiber spaces*.

We are now ready to rephrase the classical MMP described in 2.12 in modern language.

2.28 (MMP for surfaces - modern version).

- (1) Start with a smooth projective surface S .
- (2) Ask: Is K_S nef? If yes, stop! S is a minimal model. If not, pick a K_S -negative extremal ray R of the cone $\overline{NE}(S)$ and go to (3).
- (3) Let $\varphi_R : S \rightarrow Y$ be the contraction of R . Ask: Is $\dim Y < 2$? If yes, stop! $\varphi_R : S \rightarrow Y$ is a Mori fiber space. If not, φ_R is the blowup of a smooth surface. Go back to (1) with S replaced with Y .

2.4. Birational invariants. We introduce some birational invariants for surfaces.

Definition 2.29. Let S be a smooth surface.

- (1) The *genus* of S is $p_g(S) := h^0(S, \omega_S)$.
- (2) More generally the *plurigenera* of S are $P_n(S) := h^0(S, \omega_S^{\otimes n})$, where n is a positive integer.
- (3) The *irregularity* of S is $q(S) := h^0(S, \Omega_S^1) = h^1(S, \mathcal{O}_S)$. (The last equality follows from Hodge duality.)

Theorem 2.30. *The quantities p_g , P_n and q are birational invariants for smooth surfaces.*

Proof. Let S and S' be smooth surfaces, and $\varphi : S \dashrightarrow S'$ a birational map. Then there is a finite subset $\Delta \subset S$ such that $\varphi|_{S \setminus \Delta} : S \setminus \Delta \rightarrow S'$ is a morphism. Given a 2-form $\omega \in H^0(S', \omega_{S'})$, we get a form $\varphi^*\omega \in H^0(S \setminus \Delta, \omega_{S \setminus \Delta})$. We may view $\varphi^*\omega$ as a meromorphic form on S with poles along Δ . Since Δ has codimension ≥ 2 in S , $\varphi^*\omega$ extends to a 2-form $\overline{\varphi^*\omega} \in H^0(S, \omega_S)$. This yields an inclusion $H^0(S', \omega_{S'}) \subset H^0(S, \omega_S)$. The same argument gives the reverse inclusion. Hence $p_g(S) = p_g(S')$.

The proof of birational invariance of P_n and q is analogous. \square

Exercise 2.31. Compute the birational invariants p_g , P_n and q for rational and ruled surfaces. (Hint: choose a suitable birational model.)

It turns out that the birational invariants p_g , P_n and q may be used to characterize rational and ruled surfaces. This is the content of the next result.

Theorem 2.32 (Numerical characterization of rational and ruled surfaces). *Let S be a smooth surface.*

- (1) (Castelnuovo) S is rational $\iff q(S) = 0$ and $P_n(S) = 0 \forall n \geq 1 \iff q(S) = 0$ and $P_2(S) = 0$.
- (2) (Enriques) S is ruled $\iff P_n(S) = 0 \forall n \geq 1 \iff P_{12}(S) = 0$.

Next we define *Kodaira dimension*. We give the definition for arbitrary smooth projective varieties, and then we specialize to the surface case.

2.5. Kodaira dimension. Let X be a smooth projective variety of dimension $n \geq 1$. Let $D \in \text{Div}(X)$ be a divisor and suppose that $H^0(X, \mathcal{O}_X(D)) \neq 0$. Pick a basis $\{s_0, \dots, s_k\}$ for $H^0(X, \mathcal{O}_X(D)) \cong \mathbb{C}^{k+1}$, and consider the rational map $\varphi_{|D|} : X \dashrightarrow \mathbb{P}^k$ that sends a point x at which not all the s_i 's vanish to the point $(s_0(x) : \dots : s_k(x)) \in \mathbb{P}^k$. We have $0 \leq \dim(\varphi_{|D|}(X)) \leq n$.

Definition 2.33. Let X be a smooth projective variety of dimension $n \geq 1$, and $D \in \text{Div}(X)$. Define the semigroup of D to be $\mathbb{N}(D) = \{m \geq 0 \mid H^0(X, \mathcal{O}_X(mD)) \neq 0\}$. The *Iitaka dimension* of D is defined to be

$$\kappa(D) = \begin{cases} -\infty, & \text{if } \mathbb{N}(D) = \{0\} \\ \max\{\dim(\varphi_{|mD|}(X)) \mid m \in \mathbb{N}(D)\}, & \text{if } \mathbb{N}(D) \neq \{0\} \end{cases}$$

Note that $\kappa(D) \in \{-\infty, 0, 1, \dots, n\}$. It can be shown that there exist positive constants c_1 and c_2 , depending on D , such that, for $m \in \mathbb{N}(D)$ sufficiently large, we have

$$c_1 \cdot m^{\kappa(D)} \leq h^0(X, \mathcal{O}_X(mD)) \leq c_2 \cdot m^{\kappa(D)}.$$

We say that the divisor D is *big* if $\kappa(D) = n$.

The *Kodaira dimension* of X is defined to be $\kappa(X) := \kappa(K_X)$.

Exercise 2.34. Show that the Kodaira dimension is a birational invariant for smooth projective varieties.

Examples 2.35.

- (1) Curves. If $\dim(X) = 1$, then $\kappa(X) \in \{-\infty, 0, 1\}$.
 - $g(X) = 0 \iff X \cong \mathbb{P}^1 \iff -K_X$ is ample $\iff \kappa(X) = -\infty$.
 - $g(X) = 1 \iff -K_X = 0 \iff \kappa(X) = 0$.
 - $g(X) \geq 2 \iff K_X$ is ample $\iff \kappa(X) = 1$.
- (2) Hypersurfaces. Let $X = X_d \subset \mathbb{P}^{n+1}$ be a smooth hypersurface of degree d . It follows from the adjunction formula that $K_X = (-n - 2 + d) \cdot H$, where H is the class of a hyperplane in \mathbb{P}^n .
 - $d < n + 2 \iff -K_X$ is ample $\iff \kappa(X) = -\infty$.
 - $d = n + 2 \iff K_X = 0 \iff \kappa(X) = 0$.
 - $d > n + 2 \iff K_X$ is ample $\iff \kappa(X) = n$.

Exercise 2.36. Let X and Y be smooth projective varieties, and suppose that $\kappa(X) = 0$. Show that $\kappa(X \times Y) = \kappa(Y)$.

Conclude that, for each positive integer n , and each $\kappa \in \{-\infty, 0, 1, \dots, n\}$, there exists a smooth projective variety X of dimension n and Kodaira dimension $\kappa(X) = \kappa$.

Definition 2.37. We say that a smooth projective variety X is of *general type* if $\kappa(X) = \dim(X)$.

2.38 (Enriques' classification of minimal surfaces). Let S be a smooth surface. Then $\kappa(S) \in \{-\infty, 0, 1, 2\}$. It follows from Theorem 2.32(2) that $\kappa(S) = -\infty$ if and only if S is a ruled surface.

On the other hand, if $\kappa(S) \geq 0$, then the MMP for S as described in 2.28 ends necessarily with a minimal model S_{min} . Moreover, by Exercise 2.20, S_{min} is unique up to isomorphism.

Minimal models S of surfaces can be divided into the following classes, according to the values of their birational invariants p_g , P_n , q and κ :

- (1) $\kappa(S) = 0$. There are 4 classes.
 - (a) $p_g(S) = q(S) = 0$. These are called Enriques' surfaces.
 - (b) $p_g(S) = 0$ and $q(S) = 1$. These are called bielliptic surfaces.
 - (c) $p_g(S) = 1$ and $q(S) = 0$. These are called K3 surfaces.
 - (d) $p_g(S) = 1$ and $q(S) = 2$. These are abelian surfaces.
- (2) $\kappa(S) = 1$. Such surfaces admit a fibration $f : S \rightarrow B$ onto a smooth curve whose generic fiber is an elliptic curve.
- (3) $\kappa(S) = 2$. Most surfaces lie in this class. These are called surfaces of general type.

3. THE MMP IN HIGHER DIMENSIONS

Now we want to extend the MMP for surfaces, as described in 2.28, to higher dimensions. Our first task is to introduce the intersection product, spaces of curves and divisors, which are different spaces in dimension bigger than 2, and special cones on them. The reference for most of this section is [9].

3.1. Intersection product and spaces of curves and divisors. Throughout this subsection let X be a smooth projective variety.

Definition 3.1. Consider the free abelian group $Z_1(X)$ generated by curves on X . We have an intersection product:

$$\cdot : \text{Pic}(X) \times Z_1(X) \rightarrow \mathbb{Z}$$

with the property that, if $D \in \text{Pic}(X)$ and $C \subset X$ is a curve, with normalization $n : \tilde{C} \rightarrow C$, then $D \cdot C$ equals the degree of the invertible sheaf $n^*(\mathcal{O}_X(D)|_C)$.

Two elements $D, D' \in \text{Pic}(X)$ are said to be *numerically equivalent* if $D \cdot \alpha = D' \cdot \alpha$ for every $\alpha \in Z_1(X)$. In this case we write $D \equiv D'$. We write $\text{Num}(X)$ for the quotient group $\text{Pic}(X)/\equiv$. By the Theorem of the base of Néron-Severi, $\text{Num}(X)$ is a finitely generated abelian group. Its rank is called the *Picard number* of X , and is denoted by $\rho(X)$. We define the $\rho(X)$ -dimensional \mathbb{R} -vector space $N^1(X) := \text{Num}(X) \otimes_{\mathbb{Z}} \mathbb{R}$.

Similarly, two cycles $\alpha, \alpha' \in Z_1(X)$ are said to be *numerically equivalent* if $D \cdot \alpha = D \cdot \alpha'$ for every $D \in \text{Pic}(X)$. In this case we write $\alpha \equiv \alpha'$. We define the $\rho(X)$ -dimensional \mathbb{R} -vector space $N_1(X) := (Z_1(X) \otimes_{\mathbb{Z}} \mathbb{R})/\equiv$.

The intersection product on X induces a perfect pairing $\cdot : N^1(X) \times N_1(X) \rightarrow \mathbb{R}$, making $N^1(X)$ and $N_1(X)$ dual vector spaces.

Next we introduce the *Mori cone* and the cone of nef divisors.

Definition 3.2. The *Mori cone* of X is the closed convex cone $\overline{NE}(X) \subset N_1(X)$ generated by classes of curves $C \subset X$.

We say that a divisor $D \in \text{Div}(X)$ is *nef* if $D \cdot C > 0$ for every curve $C \subset X$. This is equivalent to saying that $D \cdot \alpha \geq 0$ for every $\alpha \in \overline{NE}(X)$. So the dual cone of $\overline{NE}(X)$ under the intersection product is the closed convex cone $\text{Nef}(X) \subset N^1(X)$ generated by nef divisors. We call it the *nef cone* of X .

Remark 3.3. Similarly, one can define the *cone of pseudo-effective divisors* of X as the closed convex cone $\text{Pseff}(X) \subset N^1(X)$ generated by classes of effective divisors. It was proved in [4] that the dual cone of $\text{Pseff}(X) \subset N^1(X)$ under the intersection product is the closed convex cone in $N_1(X)$ generated by classes of the so called *strongly movable curves*. Strongly movable curves are images in X of curves obtained as complete intersections of suitable very ample divisors on birational modifications of X .

It is a formidable fact that many geometric properties of divisors depend only on their numerical class. The following are two important manifestations of this phenomenon.

Theorem 3.4. *Let $D \in \text{Div}(X)$ be a divisor.*

- (1) (Kleiman's ampleness criterion.) D is ample $\iff D \cdot \ell > 0 \ \forall \ell \in \overline{NE}(X) \setminus \{0\}$.
This is equivalent to saying that the class of D lies in the interior of $\text{Nef}(X)$.
- (2) (Kodaira's lemma.) D is big \iff the class of D lies in the interior of $\text{Pseff}(X)$.

3.2. The first theorems of the MMP. We want to run the program described in 2.28 starting with a smooth projective variety X of arbitrary dimension. We start by asking whether K_X is nef. If K_X is nef, then we stop and say that X is a minimal model. If K_X is not nef, then we may pick a K_X -negative extremal ray R of the Mori cone $\overline{NE}(X)$. As we shall see in Theorem 3.11 below, $R = \mathbb{R}_{\geq 0}[C]$ for some rational curve $C \subset X$, and the contraction of R (as in Definition 2.25) exists. Let us denote it by $\varphi_R : X \rightarrow Y$. If $\dim Y < \dim X$, then, as before, we stop and call $\varphi_R : X \rightarrow Y$ a Mori fiber space. If $\dim Y = \dim X$, then φ_R is a birational morphism, and we would like to replace X with Y and go back to the original question. Here we face a problem that did not appear in the surface case: the variety Y may be singular. This situation is illustrated in Example 3.6 below.

Exercise 3.5. Let $Y \subset \mathbb{P}^N$ be a smooth projective variety, and $C(Y) \subset \mathbb{P}^{N+1}$ the cone over Y with vertex P . Let X be the blowup of $C(Y)$ at the point P . Show that X is a smooth projective variety.

Example 3.6. Let $Y \subset \mathbb{P}^5$ be the Veronese embedding of \mathbb{P}^2 , and $C(Y) \subset \mathbb{P}^6$ the cone over Y with vertex P . One can check that the canonical divisor $K_{C(Y)}$ is not Cartier, while $2K_{C(Y)}$ is Cartier. Let $\pi : X \rightarrow C(Y)$ be the blowup of $C(Y)$ at the point P , and denote by $E \cong \mathbb{P}^2$ the exceptional divisor. By Exercise 3.5, X is a smooth projective 3-fold. One can check that $\mathcal{O}_X(E)|_E \cong \mathcal{O}_{\mathbb{P}^2}(-2)$. There is a morphism $p : X \rightarrow \mathbb{P}^2$, with fibers isomorphic to \mathbb{P}^1 , which resolves the indeterminacy of the projection $C(Y) \dashrightarrow Y \cong \mathbb{P}^2$ from the point P . One can show that $\text{Pic}(X) = \mathbb{Z} \cdot [p^*\mathcal{O}_{\mathbb{P}^2}(1)] \oplus \mathbb{Z} \cdot [E]$, and in $\text{Pic}(X)$ we have

$$2K_X = \pi^*2K_Y + E.$$

In $\text{Pic}(X) \otimes_{\mathbb{Z}} \mathbb{Q}$ we have

$$K_X = \pi^*K_Y + \frac{1}{2}E.$$

As for the space of 1-cycles, $N_1(X) = \mathbb{R} \cdot f \oplus \mathbb{R} \cdot \ell$, where f denotes the class of a fiber of p , and ℓ denotes the class of a curve on E corresponding to a line in \mathbb{P}^2 under the isomorphism $E \cong \mathbb{P}^2$. Note that f and ℓ are classes of curves contained on the fibers of the morphisms p and π , respectively. Hence f and ℓ generate extremal rays of the Mori cone $\overline{NE}(X) \subset N_1(X)$. The intersection product of curves and divisors on X gives: $K_X \cdot f = -1$ and $K_X \cdot \ell = -2$. Hence both f and ℓ generate K_X -negative extremal rays of $\overline{NE}(X)$. The contraction of the ray $\mathbb{R}_{\geq 0}f$ is the morphism $p : X \rightarrow \mathbb{P}^2$, while the contraction of the ray $\mathbb{R}_{\geq 0}\ell$ is the blowup $\pi : X \rightarrow C(Y)$. The latter is a birational morphism onto a *singular* variety.

This simple example brings a point that was understood since the beginning of the development of the MMP for higher dimensional varieties: singularities are unavoidable, and we must learn how to deal with them. A whole theory of singularities was developed in the context of the MMP. In these lectures we will only consider a small portion of it. Namely, we will define the smallest class of singularities \mathcal{S} that unavoidably appear when running the MMP starting with smooth projective varieties, and such that the steps of the MMP are still valid for projective varieties with singularities in the class \mathcal{S} .

Recall that we start the MMP by asking if K_X is nef. For this question to make sense, it is necessary that the divisor K_X is \mathbb{Q} -Cartier, i.e., some nonzero multiple of it is Cartier. In these lectures we will require something stronger, namely, that X is \mathbb{Q} -factorial.

Definition 3.7. Let X be an arbitrary projective variety. A \mathbb{Q} -divisor on X is a \mathbb{Q} -linear combination of prime Weil divisors on X . A \mathbb{Q} -divisor D on X is said to be \mathbb{Q} -Cartier if some nonzero multiple of D is a Cartier divisor. Two \mathbb{Q} -divisors D and D' on X are said to be \mathbb{Q} -linearly equivalent if there exists an integer $m > 0$ such that both mD and mD' are Cartier and $mD \sim mD'$. In this case we write $D \sim_{\mathbb{Q}} D'$.

We say that X is \mathbb{Q} -factorial if every \mathbb{Q} -divisor on X is \mathbb{Q} -Cartier.

Remark 3.8. The vector spaces $N^1(X)$ and $N_1(X)$, their intersection product, the cones of curves and divisors introduced in the beginning of this section, and the Kodaira dimension may all be defined more generally for \mathbb{Q} -factorial projective varieties. We leave this easy task to the reader.

If we start with a \mathbb{Q} -factorial projective variety X , then we can ask whether K_X is nef. If the answer is no, then we pick a K_X -negative extremal ray R of the Mori cone $\overline{NE}(X)$,

and we wish to consider the contraction of R . Now we encounter another problem. The Contraction Theorem that we need here is not valid for arbitrary \mathbb{Q} -factorial projective varieties. So we must consider a more restrictive class of possibly singular varieties. The following definition is not intuitive, but it is the right one in our context.

Definition 3.9. Let X be a normal projective variety, and suppose that K_X is \mathbb{Q} -Cartier. Let $f : \tilde{X} \rightarrow X$ be a log resolution of X . This means that \tilde{X} is a smooth projective variety, f is a birational morphism whose exceptional locus is the union of prime divisors E_i 's, and the divisor $\sum E_i$ has simple normal crossing support. There are uniquely defined rational numbers $a(E_i)$'s such that

$$K_{\tilde{X}} \sim_{\mathbb{Q}} f^* K_X + \sum_{E_i} a(E_i) E_i.$$

The $a(E_i)$'s do not depend on the log resolution, but only on the valuations associated to the E_i 's.

We say that X is *terminal* if, for some log resolution $f : \tilde{X} \rightarrow X$, $a(E_i) > 1$ for every f -exceptional prime divisor E_i . If this condition holds for some log resolution of X , then it holds for every log resolution of X .

Now we can state the first theorems of the MMP, which hold for the class of \mathbb{Q} -factorial terminal projective varieties.

Theorem 3.10 (Cone Theorem). *Let X be a \mathbb{Q} -factorial terminal projective variety. There is a countable set $\Gamma \subset \overline{NE}_1(X)$ of classes of rational curves $C \subset X$ with $0 < -K_X \cdot C \leq 2 \dim(X)$ such that*

- (1) *for any ample divisor A on X , there are finitely many classes $[C_1], \dots, [C_r]$ in Γ such that*

$$\overline{NE}_1(X) = \overline{NE}_1(X)_{(K_X+A) \geq 0} + \sum_{i=1}^r \mathbb{R}_{\geq 0}[C_i], \text{ and}$$

- (2) $\overline{NE}_1(X) = \overline{NE}_1(X)_{K_X \geq 0} + \sum_{[C] \in \Gamma} \mathbb{R}_{\geq 0}[C]$.

Theorem 3.11 (Contraction Theorem). *Let X be a \mathbb{Q} -factorial terminal projective variety. Let F be a K_X -negative extremal face of the Mori cone $\overline{NE}(X)$. Then there exists a unique morphism $\varphi_F : X \rightarrow Y$ onto a normal projective variety such that $(\varphi_F)_* \mathcal{O}_X = \mathcal{O}_Y$, and, for any curve $C \subset X$, $\varphi_F(C)$ is a point if and only if $[C] \in F$.*

Definition 3.12. Under the assumptions and notation of Theorem 3.11, we say that $\varphi_F : X \rightarrow Y$ is the *contraction of F* .

3.13 (Properties of contractions of K_X -negative extremal rays). Let X be a \mathbb{Q} -factorial terminal projective variety. Let R be a K_X -negative extremal ray of the cone $\overline{NE}(X)$, and $\varphi_R : X \rightarrow Y$ the contraction of R . There is an exact sequence

$$(3.1) \quad 0 \rightarrow \text{Pic}(Y) \xrightarrow{f^*} \text{Pic}(X) \rightarrow \mathbb{Z},$$

where the last map is given by intersection with a curve $C \subset X$ such that $R = \mathbb{R}_{\geq 0}[C]$. In particular, $\rho(X) = \rho(Y) + 1$. The *exceptional locus* of φ_R is the locus of X consisting of points at which φ_R fails to be a local isomorphism. One of the following situations occurs.

- (1) $\dim(Y) < \dim(X)$. We call such φ_R a *Mori fiber space*.
- (2) The morphism φ_R is birational and the exceptional locus of φ_R consists of a prime divisor on X . In this case, Y is a \mathbb{Q} -factorial terminal projective variety. We call such φ_R a *divisorial contraction*.

- (3) The morphism φ_R is birational and the exceptional locus of φ_R has codimension at least 2 in X . We call such φ_R a *small contraction*.

Definition 3.14. A \mathbb{Q} -factorial terminal projective variety X is called a *minimal model* if K_X is nef.

Let us resume our description of the MMP. We start with a smooth (or more, generally \mathbb{Q} -factorial terminal) projective variety X , and ask whether K_X is nef. If K_X is nef, then X is a minimal model and we stop. If K_X is not nef, then we pick a K_X -negative extremal ray R of the Mori cone $\overline{NE}(X)$, and consider its contraction $\varphi_R : X \rightarrow Y$. According to the description given in 3.13, there are 3 possibilities.

- (1) If $\varphi_R : X \rightarrow Y$ a Mori fiber space, then we stop.
- (2) If $\varphi_R : X \rightarrow Y$ is a divisorial contraction, then Y is \mathbb{Q} -factorial and terminal, and we go back to the original question with X replaced with Y . In this case $\rho(Y) = \rho(X) - 1$.
- (3) If $\varphi_R : X \rightarrow Y$ is a small contraction, then we are in trouble for the following reason.

Claim 3.1. *Under the assumptions of 3.13(3), K_Y is not \mathbb{Q} -Cartier.*

Proof. Let $C \subset X$ be a rational curve such that $R = \mathbb{R}_{\geq 0}[C]$. Suppose that K_Y is \mathbb{Q} -Cartier, and consider the \mathbb{Q} -divisor $\varphi_R^* K_Y$ on X . Since $\varphi_R(C)$ is a point, we have $\varphi_R^* K_Y \cdot C = 0$.

On the other hand, $\varphi_R^* K_Y$ coincides with K_X in the open subset of X where φ_R is an isomorphism. Since the exceptional locus of φ_R has codimension at least 2 in X , we must have $\varphi_R^* K_Y = K_X$ on X . However, by assumption, R is a K_X -negative extremal ray, and thus $K_X \cdot C < 0$, yielding a contradiction. \square

Since K_Y is not \mathbb{Q} -Cartier, in case (3) we cannot hope to continue running the MMP with X replaced with Y . The idea then is to do something different. Instead of contracting the ray R and replacing X with Y , we will perform a *flip* $\psi : X \dashrightarrow X^+$, and go back to the original question with X replaced with X^+ . We will explain the notion of flip in the next subsection.

3.3. Flips. We now come to a fundamental concept from the MMP.

Definition 3.15 (Flip). Let X be a \mathbb{Q} -factorial terminal projective variety, and $f = \varphi_R : X \rightarrow Y$ a small contraction associated to a K_X -negative extremal ray $R \subset \overline{NE}(X)$.

A *flip of f* is a commutative diagram

$$\begin{array}{ccc} X & \xrightarrow{\psi} & X^+, \\ & \searrow f & \swarrow f^+ \\ & & Y \end{array}$$

where $\psi : X \dashrightarrow X^+$ is a birational map and $f^+ : X^+ \rightarrow Y$ is a birational morphism satisfying the following conditions.

- (1) K_{X^+} is \mathbb{Q} -Cartier.
- (2) The exceptional locus of f^+ has codimension at least 2 in X^+ .
- (3) $K_{X^+} \cdot C > 0$ for every curve $C \subset X^+$ contracted by f^+ .

We refer to [9, Example 2.7] for an example of flip.

It is a difficult task to prove the existence of flips. In dimension 3, it was proved by Mori in [11]. In dimension 4, it was proved by Shokurov in [13]. In [6], Hacon and McKernan proved that flips exist in dimension n provided the existence of minimal models

in dimension $n - 1$. Using this inductive scheme, existence of flips in any dimension was finally proved in [3]. Given the existence of flips, it is not so difficult to prove that it satisfies the following properties.

3.16 (Properties of flips). Let the notation be as in Definition 3.15.

- (1) The flip of f is unique up to isomorphism. In fact, the existence of $f^+ : X^+ \rightarrow Y$ is equivalent to the finite generation of the \mathcal{O}_Y -algebra $\bigoplus_{m \in \mathbb{Z}_{\geq 0}} f_* \mathcal{O}_X(\lfloor mK_X \rfloor)$. Moreover, $f^+ : X^+ \rightarrow Y$ is precisely $\text{Proj}_Y \left(\bigoplus_{m \in \mathbb{Z}_{\geq 0}} f_* \mathcal{O}_X(\lfloor mK_X \rfloor) \right) \rightarrow Y$.
- (2) X^+ is a \mathbb{Q} -factorial terminal projective variety.

Notice moreover that $\psi : X \dashrightarrow X^+$ is an isomorphism in codimension 1. Hence, since both X and X^+ are \mathbb{Q} -factorial, we have $\rho(X^+) = \rho(X)$.

Now we can finally describe the MMP in arbitrary dimension.

3.17 (MMP in arbitrary dimension).

- (1) Start with a \mathbb{Q} -factorial terminal projective variety X .
- (2) Ask: Is K_X nef? If yes, stop! X is a minimal model. If not, pick a K_X -negative extremal ray R of the cone $\overline{NE}(X)$ and go to (3).
- (3) Let $\varphi_R : S \rightarrow Y$ be the contraction of R . There are 3 possibilities.
 - (a) If $\varphi_R : X \rightarrow Y$ a Mori fiber space, then we stop.
 - (b) If $\varphi_R : X \rightarrow Y$ is a divisorial contraction, then Y is \mathbb{Q} -factorial and terminal. Go back to (1) with X replaced with Y .
 - (c) If $\varphi_R : X \rightarrow Y$ is a small contraction, then consider the flip $\psi : X \dashrightarrow X^+$ of φ_R . Then X^+ is \mathbb{Q} -factorial and terminal. Go back to (1) with X replaced with X^+ .

In order to conclude the program, one must show that this process eventually stops. Every time we perform a divisorial contraction $X \rightarrow Y$, the Picard number drops by one, $\rho(Y) = \rho(X) - 1$. However, in the case of a flip $X \dashrightarrow X^+$, we have $\rho(X^+) = \rho(X)$. Therefore this process can only admit a finite number of divisorial contraction, while we have the following question:

Does there exist an infinite sequence of flips?

Termination of flips in dimension 3 was proved in [12]. However, to this date the answer to the question above is not known in arbitrary dimension. So the MMP as described in 3.17 has not been established in higher dimensions. However, in certain cases, a special instance of the MMP, called *MMP with scaling* was proved to terminate in any dimension in [3]. This is the subject of the next section.

4. MMP WITH SCALING

As we mentioned at the end of the previous section, if we start with a \mathbb{Q} -factorial terminal projective variety X , and run the MMP as described in 3.17, it is not clear that the process terminates. There is however a variation of this program, called the *MMP with scaling*, in which we start with an ample divisor H on X and, instead of choosing an arbitrary extremal ray at each step of the MMP, we use the divisor H to narrow (and sometimes decide) our choice of extremal ray. Here is how it works. At the first step, if K_X is not nef, then, instead of choosing an arbitrary K_X -negative extremal ray of $\overline{NE}(X)$, we proceed as follows. Since H is ample, $\overline{NE}(X) \setminus \{0\}$ is contained in the half-space $\{H > 0\}$. We move the hyperplane $\{K_X = 0\}$ in $N_1(X)$ toward $\{H = 0\}$

until it supports an extremal face F of $\overline{NE}(X)$, and then we choose an extremal ray contained in this face. More precisely, we define

$$\lambda = \inf \left\{ t \geq 0 \mid [K_X + tH] \in \text{Nef}(X) \right\},$$

and choose an extremal ray of $\overline{NE}(X)$ supported on $K_X + \lambda H$. (We invite the reader to draw a picture.) This is necessarily a K_X -negative extremal ray. Then we continue as in the ordinary MMP. If $\psi : X \dashrightarrow Y$ is a birational step in the MMP (i.e., either a divisorial contraction or a flip), then we replace X with Y and H with ψ_*H . The divisor ψ_*H is no longer ample. Nevertheless, the procedure just described can be repeated for Y and ψ_*H .

Definition 4.1. Let X be a \mathbb{Q} -factorial terminal projective variety, and H a \mathbb{Q} -divisor on X . Suppose that $K_X + \lambda H$ is nef for some $\lambda \geq 0$. (This holds for instance if H is ample.) We define the *nef threshold of H* by

$$\lambda(X, H) = \inf \left\{ \lambda \geq 0 \mid [K_X + \lambda H] \in \text{Nef}(X) \right\}.$$

The *Rationality Theorem* asserts that $\lambda(X, H) \in \mathbb{Q}$.

Now we describe the MMP with scaling in more detail. We start with a \mathbb{Q} -factorial terminal projective variety X , and an ample divisor H on X . We will define inductively (possibly finite) sequences of \mathbb{Q} -factorial terminal projective varieties X_i 's, together with \mathbb{Q} -divisors H_i 's on them such that $K_{X_i} + \lambda H_i$ is nef for some $\lambda \geq 0$. For each i , $\psi_i : X_i \dashrightarrow X_{i+1}$ will be either a divisorial contraction or a flip from the ordinary MMP, and $H_{i+1} = (\psi_i)_*H_i$.

Step 0. We set $X_0 = X$, and $H_0 = H$. We move to Step 1 with $n = 0$.

Step 1. Suppose we have constructed X_n and H_n . Set $\lambda_n = \inf \left\{ \lambda \geq 0 \mid [K_{X_n} + \lambda H_n] \in \text{Nef}(X_n) \right\}$. We move to Step 2.

Step 2. We ask whether K_{X_n} is nef (or, equivalently, if $\lambda_n = 0$).

If K_{X_n} is nef, then we stop and the sequence $\{X_i\}$ ends with the minimal model X_n .

If K_{X_n} is not nef, then there exists at least one K_{X_n} -negative extremal ray $R \subset \overline{NE}(X_n)$ such that $(K_{X_n} + \lambda_n H_n) \cdot R = 0$. We choose one such extremal ray R_n , and let $\varphi_n : X_n \rightarrow Y$ be the contraction of R_n . We move to Step 3.

Step 3. We check which of the three possibilities described in 3.13 occurs.

(1) If $\varphi_n : X_n \rightarrow Y$ is a Mori fiber space, then we stop and the sequence $\{X_i\}$ ends with X_n .

(2) If $\varphi_n : X_n \rightarrow Y$ is a divisorial contraction, then we set $X_{n+1} = Y$ and $H_{n+1} = (\varphi_n)_*H_n$. Since $(K_{X_n} + \lambda_n H_n) \cdot R_n = 0$, by (3.1),

$$K_{X_n} + \lambda_n H_n \sim_{\mathbb{Q}} (\varphi_n)^*(K_{X_{n+1}} + \lambda_n H_{n+1}).$$

Since $K_{X_n} + \lambda_n H_n$ is nef, this implies that $K_{X_{n+1}} + \lambda_n H_{n+1}$ is also nef. We go back to Step 1 replacing n with $n + 1$.

(3) If $f = \varphi_n : X_n \rightarrow Y$ is a small contraction, and $\psi : X_n \dashrightarrow X_n^+$ is the associated flip, then we set $X_{n+1} = X_n^+$, and $H_{n+1} = \psi_*H_n$. Consider the flip diagram:

$$\begin{array}{ccc} X_n & \xrightarrow{\psi} & X_{n+1} \\ & \searrow f & \swarrow f^+ \\ & & Y \end{array}$$

Since $(K_{X_n} + \lambda_n H_n) \cdot R_n = 0$, by (3.1), there exists a \mathbb{Q} -Cartier \mathbb{Q} -divisor D_Y on Y such that $K_{X_n} + \lambda_n H_n \sim_{\mathbb{Q}} f^* D_Y$. Then $K_{X_{n+1}} + \lambda_n H_{n+1} \sim_{\mathbb{Q}} (f^+)^* D_Y$. By hypothesis $K_{X_n} + \lambda_n H_n$ is nef. Thus D_Y is nef and so is $K_{X_{n+1}} + \lambda_n H_{n+1}$. We go back to Step 1 replacing n with $n + 1$.

In [3], the MMP with scaling was proved to terminate in the following two important cases:

- (1) X is of general type (this is equivalent to saying that K_X is big, i.e., K_X lies in the interior of $\text{Pseff}(X)$). In this case, the MMP with scaling ends with a minimal model.
- (2) X is uniruled (by [4] this is equivalent to saying that $K_X \notin \text{Pseff}(X)$). In this case, the MMP with scaling ends with a Mori fiber space.

SUGGESTED READING

The reader interested in a more detailed and rigorous introduction to the MMP and its techniques is referred to [9]. The texts [8] and [10] also provide a good introduction. All of these cover the “classical” MMP.

There are many notes available in the web discussing the more recent results from [3], including the MMP with scaling. In addition to [3] itself, the reader may consult the expository paper [5]. Those interested in the MMP with scaling for uniruled varieties exclusively may also look at [1].

REFERENCES

- [1] C. Araujo, *The cone of pseudo-effective divisors of log varieties after Batyrev*, Math. Z. **264**, no. 1, 179–193 (2010).
- [2] A. Beauville, *Complex Algebraic Surfaces*, London Mathematical Society Student Texts **No. 34** (1983).
- [3] C. Birkar, P. Cascini, C. Hacon, and J. McKernan, *Existence of minimal models for varieties of log general type*, J. Amer. Math. Soc., **No. 23(2)** 405–468 (2010).
- [4] S. Boucksom, J.-P. Demailly, M. Paun, and T. Peternell, *The pseudo-effective cone of a compact Kähler manifold and varieties of negative Kodaira dimension*, pre-print math.AG/0405285 (2004).
- [5] S. Druel, *Existence de modèles minimaux pour les variétés de type général*, Exposé 982, Séminaire Bourbaki, 2007/08, Astérisque **326**, 1–38 (2009).
- [6] C. Hacon, J. McKernan, *On the existence of flips*, preprint arXiv:math/0507597v1 [math.AG] (2005).
- [7] R. Hartshorne, *Algebraic Geometry*, Springer-Verlag, Graduate Texts in Mathematics **No. 52** (1977).
- [8] Y. Kawamata, K. Matsuda and K. Matsuki *Introduction to the minimal model problem*, in Algebraic geometry, Sendai, 1985, Adv. Stud. Pure Math., **vol. 10**, North-Holland, Amsterdam, 283–360 (1987).
- [9] J. Kollár and S. Mori, *Birational geometry of algebraic varieties*, Cambridge Tracts in Mathematics **vol. 134**, Cambridge University Press, Cambridge, (1998).
- [10] K. Matsuki, *Introduction to the Mori program*, Universitext, Springer-Verlag, New York, (2002).
- [11] S. Mori, *Flip theorem and the existence of minimal models for 3-folds*, J. Amer. Math. Soc. **1**, 117–253 (1988).
- [12] V. V. Shokurov, *A nonvanishing theorem*, Izv. Akad. Nauk SSSR Ser. Mat. **49**, no. 3, 635 – 651 (1985).
- [13] V. V. Shokurov, *Prelimiting flips*, Tr. Mat. Inst. Steklova **240**, Biratsion. Geom. Linein. Sist. Konechno Porozhdennye Algebr, 82219 (2003).

IMPA, ESTRADA DONA CASTORINA 110, RIO DE JANEIRO, RJ 22460-320, BRAZIL
E-mail address: caraujo@impa.br

ENERGY PROBLEMS: MODELING AND OPTIMIZATION METHODS

PROBLEMAS DE ENERGÍA: MODELIZACIÓN Y MÉTODOS DE OPTIMIZACIÓN

CLAUDIA SAGASTIZÁBAL

ABSTRACT. Los sistemas modernos de energía eléctrica proporcionan múltiples desafíos para el área de Optimización. La penetración cada vez mayor de fuentes de energía renovable introduce incertidumbre en problemas tradicionalmente modelizados de forma determinista. La liberalización del sector eléctrico trajo como consecuencia la necesidad de diseñar mercados apropiados para comercializar la energía. Estos mercados deben garantizar inversiones y al mismo tiempo ser capaces de reflejar adecuadamente interacciones estratégicas entre los diversos agentes del sector. En todos estos problemas es fundamental protegerse contra riesgos y volatilidades de precio o de demanda. Del punto de vista de la optimización, el hecho de representar la incertidumbre, el riesgo, y/o la competencia de diferentes empresas en un mercado de energía aumenta considerablemente el número de variables y restricciones del problema. Por esta razón, en general se debe hallar un buen compromiso entre la representatividad de los modelos y la resolución numérica de los problemas correspondientes: mientras más detallados sean los modelos, más difícil será la resolución del problema de optimización correspondiente. En este minicurso explicaremos diferentes técnicas de modelización y también como resolver problemas de optimización que presentan una estructura especial, como separabilidad. Analizaremos algunas variantes de métodos basados en relajación Lagrangeana, en la técnica llamada de Dantzig-Wolfe, y en la descomposición de Benders. A lo largo de la exposición usaremos como hilo conductor ejemplos de casos reales, para analizar el valor práctico de los métodos en términos de su eficacia para la resolución de los problemas considerados.

Observación: este texto fue redactado a partir de diversos trabajos de la autora citados en las referencias bibliográficas. Como tal, está escrito en inglés y no constituye un trabajo original, sino más bien una compilación de resultados anteriores, organizada para fines didácticos.

CONTENTS

| | |
|--|----|
| Introduction | 18 |
| 1. Energy Problems. An Overview | 19 |
| 2. Optimal Power Management | 20 |
| 2.1. Variables and Constraints | 21 |
| 2.2. Power Plant Modeling | 22 |
| 3. Lagrangian Relaxation Decomposition | 26 |
| 3.1. Direct Relaxation of Coupling Constraints | 26 |
| 3.2. Uncoupling via Variable Duplication | 27 |
| 4. Solving the Dual Problem | 28 |

2010 *Mathematics Subject Classification.* 65K10, 49J53, 49M05.

<http://w3.impa.br/~sagastiz/>. Research partially supported by Grants CNPq 303840/2011-0, AFOSR FA9550-08-1-0370, NSF DMS 0707205, as well as by PRONEX-Optimization and FAPERJ.

| | |
|--|----|
| 4.1. Inexact Bundle Methods | 29 |
| 4.2. Handling Inexact Oracle Information | 30 |
| 4.3. An Inexact Bundle Optimization Solver | 32 |
| 5. Back to the Primal Problem | 34 |
| 6. Multistage Decision Problems | 35 |
| 6.1. A Simplified Model for Hydro-Power Management under Uncertainty | 35 |
| 6.2. Decomposing Two-stage Problems | 37 |
| 6.3. Decomposing Multistage Problems | 38 |
| 6.4. Risk Averse SDDP Variants | 39 |
| 7. Variational Inequalities Decomposition | 40 |
| 7.1. Generalized Nash Games for Electricity Markets | 41 |
| 7.2. Finding Variational Equilibria | 42 |
| References | 44 |
| Appendix A. Convergence Analysis of Inexact Bundle Methods | 47 |
| A.1. Infinite Serious Steps | 47 |
| A.2. Finite Serious Steps with Infinite Noise Attenuation | 49 |
| A.3. Finite Serious Steps with Finite Noise Attenuation | 49 |

INTRODUCTION

Since the early days of Operations Research (soon after the Second World War), energy problems have provided a fertile field for application of the models and methods of Optimization. Optimization techniques contributed significantly to the successful resolution of several problems and provided the foundations for sound decision support. These activities, driven by challenging and complex application problems requiring the development of new algorithms, motivate in turn fundamental research. Such constant intertwining had a great positive impact in both fields.

In the past decades, to certain extent in a worldwide trend, the energy sector has undergone a restructuration that brought new interesting issues to be addressed by optimization. After a shift towards a full liberalization of the electricity market, a sequel of black-outs and the increasing concern on sustainable growth made the sector focus on the research of fast, permanent, and reliable solutions that involve some level of regulation as well as long-term planning programs. Decisions to be made by different agents (utilities, operators, regulatory bodies, governments) must take into account several conflicting objectives, of technical, economic, financial, or environmental nature, and this at various stages of decision making, ranging from the real time to the strategic government levels.

The great diversity of problems arising in the energy sector and the corresponding diversity of perspectives for assessing these problems are simply too broad to be covered here. Instead, we shall focus on some prototypical examples that give a good flavor of the challenges this type of problems poses from the optimization point of view. In all the examples, there is uncertainty that needs to be properly dealt with, as well as many variables (usually mixed-integer) and constraints (often nonlinear).

This work, which essentially gathers material from different sources in the bibliographical references, focuses on optimal power management and equilibrium problems. Section 1 starts with a general description of the energy sector and its related optimization problems. Section 2 focus on how to model power plants in the short term.

In Section 3 we explain how to decompose large-scale or complex problems using Lagrangian duality and obtain a nondifferentiable dual problem, easier to solve. The solution of the dual problem by inexact variants of bundle methods is addressed in Section 4 (a general convergence analysis for such methods can be found in the appendix). A short description of the so-called primal-recovery techniques is given in Section 5. Section 6 reviews two- and multi-stage stochastic programming problems as well as their decomposition, in the framework of long term power management problems. The final Section 7 is devoted to decomposition of variational inequalities, a useful setting to understand competition in an electricity market modeled as a generalized Nash Game.

Unless specified otherwise, throughout we use the Euclidean inner product $\langle x, p \rangle = x^\top p$ for two column vectors x and p , and denote the induced Euclidean norm by $|\cdot|$.

1. ENERGY PROBLEMS. AN OVERVIEW

An Independent System Operator (ISO) is an organization responsible for coordinating the operation of an electrical power system in an efficient and reliable manner, preferably seeking the consumers welfare. Among other tasks, the ISO is in charge of the dispatch, deciding how much electricity goes through each transmission line to bring power from a given generating company to a given consumption center. The ISO also manages the pricing of electricity, in a manner that depends on the degree of liberalization of the power system the ISO coordinates.

The ISO as well as other agents in the energy sector use optimization methods to make appropriate decisions on many fundamental issues, such as:

Generation and dispatch problems: how to operate -daily, weekly, or yearly- a mix of power plants or an individual plant, taking into account specific technological characteristics.

Transmission problems: how to deliver power through the electric network in a reliable manner.

Expansion problems: how to decide which plants are to be built in a five, ten, or thirty years future, as well as their location and network connections.

Competitive market problems: how to understand competitive interaction between several generating firms seeking to maximize profit.

Even if these problems are of different nature, they all involve maximal profits or minimal costs of generation, of transmission, of investment, of unsupplied energy. The representation of the power system configuration depends on the chosen time horizon (short, medium, long term). For example, transmission constraints can be represented by box constraints, corresponding to simple exchanges between subregions, or by linear or nonlinear inequalities, corresponding respectively to DC and AC power flows. Similarly, some parameters of the problem may have a deterministic or a stochastic nature, depending on the considered time horizon. Such is the case of water streamflows, of the evolution of oil, gas, and energy prices, and of demand of energy. In general, problems with short time horizon model in detail technological constraints, while for long time horizons the emphasis is put on a detailed model of uncertainty.

For all these problems, it is crucial to take into account specific features of electricity:

- power flows in the transmission network by following Kirchoff's laws of conservation of current and voltage in electrical circuits; and
- electricity is not storable, yet supply needs to meet the demand, which varies unwieldy.

The first item above has an important impact in transmission planning. In most network design problems, in telecommunication or transportation for instance, the routing of commodities can be decided by a manager, or fixed by a rule minimizing some utility (congestion, travel time, travel costs). By contrast, if the transmitted commodity is electricity, the network operator cannot choose through which edges the electric power is routed. As a result, when expanding the network capacity, adding new edges may make the network nonoperational, due to the presence of power flows that are unfeasible regarding Kirchoff's laws. Moreover, it may very well happen that cutting off an edge *increases* the network capacity. This specificity, known as transmission switching in the area, has been used for long time by network operators to temporarily change the topology of the system and increase transfer capacity or improve voltage profiles. For expansion problems, by contrast, transmission switching was only recently introduced, because its modeling increases significantly the number of 0-1 variables in the underlying mixed-integer bilinear program. We refer to [FOF08], [MPS10], [KSK10] for more details.

The second specific feature in the list above is not less important and has a double impact for generation problems. First, since demand cannot be known in advance and it takes time to generate power and distribute it through the network, a *spinning reserve* of extra generating capacity needs to be available to the system operator within a short interval of time. Such reserve needs to be enough but not too large, to keep costs low. The presence of intermittent sources of power, for example produced by wind farms that randomly change the generation levels of the system, make the determination of a “proper” amount of reserve extremely difficult. Second, the decision of which plant to dispatch is mainly based on generation costs. For thermal plants, such cost is related to burning some fuel (fossil, nuclear). But for hydro-plants, “fuel” is the stored water, which costs nothing. So, in principle, it seems cheaper to generate hydraulic electricity. But for power generation purposes, water reservoirs are nothing but electricity storages; since future streamflows are uncertain, so is the future availability of hydro-power. Thus, having depleted reservoirs in the future may be very costly and can even result in a black-out. It is then of paramount importance to devise sound mechanisms to price water as a source of electric power. Most of the models follow [PP91], defining the cost of water as its future cost of substitution, given by an estimate of the value function of a large multistage stochastic linear program. Often such problems are simply intractable; for the Brazilian power system, for example, uncertainty is represented by a tree with 120 time steps and 20^{119} scenarios. Only a combination of decomposition and sampling methods is applicable in this case; see Section 6.

2. OPTIMAL POWER MANAGEMENT

The optimal management of many power units is a challenging problem, from both economical and mathematical points of view. Because of interconnections between different electrical networks (of cities in a same country, or more generally, of countries in a same region), power management activities can be split into two classes, of *generation* and *distribution*, relatively independent of each other. Indeed, when solving generation problems in an interconnected system, electricity can be brought from far away if there is need, or, reciprocally, an excess of generation in some plant can be sent somewhere else. This flexibility makes it possible to use efficiently power plants with strict generation rules but low costs, in combination with more expensive plants that can produce electricity “as soon as required”. In a manner similarly, it is possible to exchange energy

with countries under different climatic conditions or with shifted load peaks, to reduce generation costs. However, the involved transmission costs and network losses should be kept in mind.

Given an electric generation mix of power plants (hydraulic, thermal -fuel or nuclear based-, eolian or other renewable), the aim is to minimize generation costs subject to operating constraints of generation units and other external constraints, like network flow capacities or satisfaction of demand. There are many different problems fitting such a large framework. In particular, the time horizon chosen for the scheduling highly determines the specificity of problems. Short, middle and long term decisions have their own peculiarities that need to be reflected in the modeling. Originally, short term problems were mostly modeled in a deterministic framework, see [BR92], [CR99], [BDMS03]. The modern trend, induced by the increasing penetration of wind power in the mix, is to employ stochastic models, to take into account wind uncertainty. For longer terms, uncertainty needs to be represented in the model, not to have poor solutions. Quoting from [BLRS01], *during the French winter, demand has uncertainties reaching up to several thousands of MW. When comparing this value to typical peak loads (70000 MW), we see that for the modeling to yield any significant values, it must explicitly incorporate the stochastic nature of the problem.*

2.1. Variables and Constraints. The following problem constitutes a good example of a realistic model. It is a short term problem, with detailed generation rules for operation of each plant and of each unit inside a plant. The solution of such problems, called of *unit commitment* gives a scheduling with directions for turning on/off all the units and how much to generate at each of them, for every 30 minutes.

Consider a set I of power plants whose generation is to be optimized in a manner that demand is satisfied at minimal cost. The decision horizon $\{1, \dots, T\}$, for example $T = 48$ half-hours; and for simplicity demand is supposed to be known. Let p_i^t denote the energy produced by the power plant $i \in I$ during the period t .

Hydraulic valleys are composed by plants and reservoirs, and each thermal plant may have several generators, or *units*. Letting I be the total number of units in the system, $p^t := (p_1^t, p_2^t, \dots, p_I^t)$ stands for the generation vector of whole mix at each time t ; likewise, $p_i := (p_i^1, p_i^2, \dots, p_i^T)$ will be the generation vector of unit i for the whole horizon of time.

Sometimes we will distinguish thermal and hydraulic components and write p_{t} and p_{h} .

A generic formulation for the optimal power management problem is the following:

$$(2.1) \quad \begin{cases} \min_p C(p) \\ p \in S \cap D, \end{cases}$$

where p is the generation, C is the operating cost function and S, D represent, respectively, static and dynamic constraints.

Static constraints are organized in two different subsets: $S = S_1 \cap S_2$. The first category, that is S_1 , gathers *coupling* constraints which relate generation of different units (and/or plants), like satisfaction of demand, network security, spinning reserve, at each time step. They are generally represented by a large-scale linear system. For example, if d^t is the requirement of demand at time t , then satisfaction of demand (also

called *load* constraint) is stated as follows:

$$(2.2) \quad \sum_{i \in I} p_i^t = d^t, \quad \text{for } t = 1, \dots, T.$$

For some models, demand is uncertain and depends on some random variable ξ . In this case, $d^t(\xi)$ is also random, and both the static feasible sets and the generation variables vary with uncertainty: $S = S(\xi)$ and $p = p(\xi)$.

Similar affine constraints can be written to include shortages and energy interchanges between subsystems of the mix. Transmission constraints (bounding the capacity of each arc in the network), and security constraints (ensuring the satisfaction of the demand even after the outage of one line) are also static constraints of type S_1 . In the DC flow approximation model these constraints are linear, but, as opposed to demand or reserve, they are often numerous, because the network has several hundreds of nodes. With security constraints, there will be several hundred thousand linear constraints at each time step. A possibility to avoid its explicit consideration is to introduce additional (linear) constraints in S_1 when necessary, along iterations.

The set S_1 also includes constraints linking units in the same plant that are required at each time step. For example, spinning reserve constraints, which guarantee that there is enough power started up to quickly face most random events (unit outage, errors in rain forecasts leading to underestimation of water inflows, etc).

Although some constraints, like reserve constraints, are naturally inequalities, we formulate them in the form of equality constraints, by introducing slack variables. These variables can be interpreted as the generation of a fictitious unit, say p_0^t , whose operating cost is zero. In (2.5) below we give the explicit form for the thermal spinning reserve constraint.

The second category S_2 represents static *non-coupling* constraints, like bounds on generation levels, that are required at each time step and for each unit.

Finally, the set of dynamic constraints $D = D_t \times D_h$ includes all the operating rules for each power plant, see § 2.2 below. Typically, in this set there is a coupling of variables of the same unit/plant along time steps.

In general, describing the dynamic sets is not a simple task, because of their disparity: thermal, nuclear and hydro plants have very different technological characteristics. Neither their operation costs nor their constraints are alike.

2.2. Power Plant Modeling. We give here some more details on the particular modeling of hydraulic and thermal plants. Note that in (2.1) it is assumed that each power plant i can be described straightforwardly in terms of its generation p_i . Actually, realistic models use state and control variables, denoted respectively by y and u , to represent p_i as a function of y and u . At first sight, this modification comes just to a change of variables in (2.1). However, it is important to realize that, although (2.1) is formally the same, the introduction of new variables can modify the structure of some constraints, that may no longer be linear, neither convex.

2.2.1. Hydraulic plants. The hydro-generation p_h depends on the discharge of water flowing through turbines (denoted below by Q), on the water-head and sometimes also on the spillage (denoted below by S). These values define state and control variables for the system, denoted respectively by y_h and u_h .

More precisely, a hydro-valley is a set of interconnected plants and reservoirs with natural inflows in each reservoir, see Fig. 1.

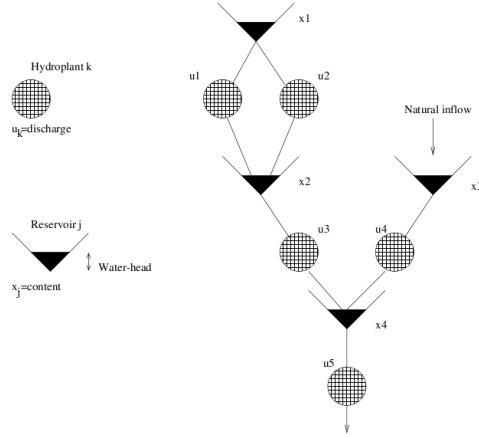


FIGURE 1. A hydro-valley

For each reservoir $i \leq I_h$, we let $j \in \text{UP}_i$ be the reservoirs immediately upstream (i.e., there is no other plant or reservoir in-between upstream the valley). The state variable $y_{h,i}$ has components $y_{h,i}^t$ for each time step $t \leq T$. Each component $y_{h,i}^t$ is the vector

$$y_{h,i}^t := \begin{pmatrix} V_i^t \\ E_i^t \end{pmatrix},$$

where V_i^t is the volume of water stored in the reservoir at the end of the time step and E_i^t is the evaporation during the time step. Evaporation depends on the current level of reservoirs and known rates of evaporation. Sometimes the model needs to consider the travel time of water between two hydro plants along the same valley. In this case, the state $y_{h,i}^t$ also includes upstream discharges: $Q_j^\tau + S_j^\tau$ for $\tau \leq t$ and $j \in \text{UP}_i$.

The control variable $u_{h,i}$ has components $u_{h,i}^t$ for each time step $t \leq T$. Each component $u_{h,i}^t$ describes volumes to be discharged (at the reservoir itself and at all the reservoirs immediately upstream), either by flowing through turbines (Q_i^t, Q_j^t) or by being spilled (S_i^t, S_j^t):

$$u_{h,i}^t := \begin{pmatrix} Q_i^t \\ S_i^t \\ (Q_j^t)_{j \in \text{UP}_i} \\ (S_j^t)_{j \in \text{UP}_i} \end{pmatrix}.$$

We see that to represent $p_{h,i} := p(y_{h,i}, u_{h,i})$ we use $T(2 + 2 + 2\#\text{UP}_i)$ variables.

The modeling of run-of-river plants is simpler, since their generation is a portion of the streamflow, immediately transformed into power, and subtracted from the demand d^t in (2.2). We refer to [GS12b] for more details.

Dynamic constraints are given by the stream-flow balance equations, establishing the physical coupling of reservoirs in the hydro valley:

$$(2.3) \quad V_i^t + Q_i^t - \sum_{j \in \text{UP}_i} Q_j^t + S_i^t - \sum_{j \in \text{UP}_i} S_j^t = V_i^{t-1} + A_i^t - E_i^t,$$

where A_i^t is the natural water inflow during the time step, a deterministic forecast, usually derived from a dynamic stochastic recourse model used for the medium term model, as the one in Section 6.

Hence,

$$D_{h,i} := \{p_{h,i} = p(y_{h,i}, u_{h,i}) : y_{h,i}^t \text{ and } u_{h,i}^t \text{ satisfy (2.3) for all } t < T\}$$

and $D_h := \prod_{i \leq I_h} D_{h,i}$. Since (2.3) is a linear equation relating $y_{h,i}^{t+1}$, $y_{h,i}^t$ and $u_{h,i}^t$, the inclusion $p_h \in D_h$ is a linear system, possibly large.

For hydraulic plants the operating cost function $C_h(p_h^t)$ is the so-called *future value* of water, i.e., the cost for time steps $\tau > t$ that would result from not having enough water in the reservoirs to satisfy the load d^τ . This substitution cost is given indirectly by the extra thermal generation and/or shortages of power that might have to be faced in the future, due to the lack of water. In some formulations, where the hydro-generation is not an important portion of the mix, it is simply modeled with a quadratic function of the water contents. For those systems predominantly hydraulic, it is estimated by using a cutting plane approximation built with information from the medium term model:

$$C_h(p_h^t) \approx \max_{k \leq K} \{ \langle \pi_t^k, y_h^t \rangle + \delta_t^k \},$$

where K is the number of cutting planes and, for each $k \leq K$, π_t^k is a vector of appropriate dimensions and δ_t^k is a scalar. Essentially, vectors π measure the sensitivity of C_h to variations on the state variables, typically the volume of water stored in the reservoirs. A distinctive feature of the future value function is that it is *not decomposable*, i.e., it cannot be expressed as a sum of individual costs over the mix; [Sag12].

An important family of static constraints coupling units of the same plant is given by the *generation* function for the reservoirs. As we mentioned, each plant generation p_i depends on the discharge of water flowing through turbines and on the water-head. In turn, water-head depends on volumes upstream and downstream the reservoir. Also, for some plants generation varies with spillage. For this reason, the non convex generation function is first convexified and this convex hull is then approximated by a piecewise linear function obtained using a Taylor expansion on the control variables (specifically, on Q_i^t and S_i^t):

$$S_{h,1} \supset \left\{ p_h = p(y_h, u) : p_i^t = \min_{k \leq K_i} \{ P_y^k \top y_{h,i}^t + P_{u_i}^k \top u_{h,i}^t + d_i^k \} \right. \\ \left. \text{for all } i \leq I_h \text{ and } t \leq T \right\},$$

where for each $i \leq I_h$, K_i is the number of pieces defining the linear function and, for each $k \leq K_i$, P_y^k , $P_{u_i}^k$ are vectors of appropriate dimensions and d_i^k is a scalar. Note that, via the volumes V_i^t , there is still a coupling along times due to (2.3). In our formulation, we eliminate this coupling by using *average* volumes in the generation function. This averaging, justified by the short time horizon, makes it possible to include the generation function constraints in the static set S_1 .

2.2.2. Thermal plants. For hydraulic plants both state and control variables are continuous. This is no longer the case for thermal plants, since on/off constraints require the use of integer variables. Thermal dynamic constraints are modeled by a graph whose nodes are associated with an integer operating mode $u_{t,i}^t$, a generation level p_i^t , and a cost C_t .

A thermal unit can only be operated after completing start-up phases and following some fixed output curves for the generation during shutdown or start-up times. There are also maintenance periods where the unit must be off line. Besides, not any range of variation is allowed for the generation.

Therefore, for thermal units we have that $p_i^t = p(p_i^t, u_{t,i}^t)$; in particular, when the unit is on line, $u_{t,i}^t > 0$ and the generation level satisfies $p_i^t \in [p_i, \bar{p}_i]$.

Suppose that I_t plants compose the thermal mix, and each plant j has K_j units indexed in the set J_j . In order to model the plant operation by means of a graph for each unit i we define the set of nodes as $V_{G_i} := V_i \setminus M_i$, where $V_i := \{(t, u_{\tau,i}^t) \in \{1, \dots, T\} \times \{0, 1\}\}$ and $M_i := \{(t, 1) \in \{1, \dots, T\} \times \{0, 1\} : m_i^t = 1\}$. The binary variables $u_{\tau,i}^t$ and m_i^t are associated, respectively, to the operating and maintenance modes. At time t , we set $u_{\tau,i}^t = 0$ if the i^{th} unit is off line and, likewise, $m_i^t = 1$ if the i^{th} unit is off for maintenance.

The set of edges E_{G_i} consists of elements $(v, w) \in V_{G_i} \times V_{G_i}$ satisfying

$$(2.4) \quad \begin{cases} v_t = w_t + 1 & \text{if } v_u = w_u \\ v_t = w_t + t_i^{up} & \text{if } v_u = w_u - 1 \\ v_t = w_t + t_i^{down} & \text{if } v_u = w_u + 1, \end{cases}$$

where t_i^{up} (resp. t_i^{down}) is the minimum time required to start up (resp. to shut down) unit i .

Finally, having the induced graph $G_i = (V_{G_i}, E_{G_i})$, we denote by G_i those binaries sequences representing connected paths between a node with $t = 1$ and another with $t = T$.

Altogether, dynamic constraints are given by

$$D_{\tau,i} := \{p_{\tau,i} = p(p_i, u_{\tau,i}) \in \mathfrak{R}^T : u_{\tau,i} \in G_i\} \quad \text{and} \quad D_{\tau} := \prod_{i \in J_j, j \leq I_{\tau}} D_{\tau,i}.$$

Depending on the regulatory rules, spinning reserve may be required to be available at each plant, or just in the whole system. In the former case, let R_j^t denote the amount of reserve required for the j^{th} plant at time t . Then each unit i in the plant has $r_i(p_i^t) := \bar{p}_i - p_i^t$ power available for reserve, so we have that

$$(2.5) \quad S_{\tau,1} := \left\{ p_{\tau}^t = p(p^t, u_{\tau}^t) : \sum_{i \in J_j} r_i(p_i^t) \geq R_j^t, \quad \text{for all } j \leq I_{\tau} \right\},$$

a linear expression, coupling all the units in the same plant. By contrast, maximum and minimum generation constraints

$$S_{\tau,2} := \left\{ p_{\tau} = p(p, u) : \right. \\ \left. u_{\tau,i}^t p_i^t \leq p_i^t \leq u_{\tau,i}^t \bar{p}_i \text{ for all } i \in J_j, j \leq I_{\tau} \text{ and } t \leq T \right\},$$

are neither coupling time steps nor units.

Note that, as opposed to hydraulic plants, there is no *generation* function for thermal plants: the state variable p gives in a straightforward manner the generation level p_{τ} . In addition, the cost function $C_{\tau}(p_{\tau}^t)$ is the sum over the thermal units of individual operating costs, generally given by quadratic or piecewise linear functions.

From the optimization point of view, problem (2.1) is large-scale, mixed-integer and non convex, with many coupling constraints, both along time steps and along different units/plants. Its solution calls for special techniques of decomposition, that can guarantee high precision. As often in energy problems, accuracy in the numerical results is of paramount importance. The reason is that some of the obtained results (marginal prices) are used as indicators of energy prices.

For deterministic generation problems we refer to [LPRS96], [CR99], [BDMS03]. Stochastic formulations have been considered in [TBL96], [NR00], [GNRS00] [BLRS01]; see also [NSW05], [PS06].

3. LAGRANGIAN RELAXATION DECOMPOSITION

The extreme complexity and diversity of the various elements defining a power system make the problems in the area hard to solve by a direct approach, and decomposition is therefore a must. Notwithstanding, decomposition should be done in a manner ensuring that the original problem is solved with relatively high accuracy. This is a common requirement in generation problems, because some of the optimal Lagrangian multipliers therein contribute to set electricity prices. Therefore, having, say three, instead of four significant digits may have a huge economic impact.

The real difficulty of the energy problems presented so far is, more than their scale, the underlying heterogeneity: a direct application of packages of nonlinear or combinatorial methods will simply fail, or give inaccurate results.

This is why it is suitable to transform this problem. Consider for example the unit-commitment problem (2.1) and suppose, for simplicity, that there are no spinning reserve constraints $S_{\mathbf{t},1}$ and $S_{\mathbf{h},1}$: the only static constraint is demand satisfaction (2.2):

$$\left\{ \begin{array}{l} \min_{p_{\mathbf{t}}, p_{\mathbf{h}}} C_{\mathbf{t}}(p_{\mathbf{t}}) + C_{\mathbf{h}}(p_{\mathbf{h}}) \\ p_{\mathbf{t}} \in D_{\mathbf{t}} \cap S_{\mathbf{t},2} \\ p_{\mathbf{h}} \in D_{\mathbf{h}} \cap S_{\mathbf{h},2} \\ (p_{\mathbf{t}}, p_{\mathbf{h}}) \in S_1 \iff \sum_{i \in I_{\mathbf{t}}} p_{\mathbf{t}}^t + \sum_{i \in I_{\mathbf{h}}} p_{\mathbf{h}}^t = d^t, t \leq T \end{array} \right.$$

The key is to observe that, if coupling constraints were not present, each difficulty (integer variables, nonlinear functions, individual plants) could be treated separately. More precisely, the coupling between the nonlinear functions with 0-1 variables in $D_{\mathbf{t}}$ and the large linear problems on continuous variables in S_1 should be eliminated.

Lagrangian relaxation is a convenient tool to bring on separability. There are several choices, depending on the problem structure, as shown below.

3.1. Direct Relaxation of Coupling Constraints. Suppose we relax the constraint of demand satisfaction with a multiplier $x = (x^1, \dots, x^t, \dots, x^T) \in \mathfrak{R}^T$. The corresponding Lagrangian

$$L(p, x) = C(p) + \sum_{t=1}^T x^t \left(\sum_{i=1}^I p_i^t - d^t \right)$$

is separable:

$$L(p, x) = - \sum_{t=1}^T x^t d^t + L_{\mathbf{t}}(p_{\mathbf{t}}, x) + L_{\mathbf{h}}(p_{\mathbf{h}}, x)$$

for

$$L_{\mathbf{t}}(p_{\mathbf{t}}, x) = C_{\mathbf{t}}(p_{\mathbf{t}}) + \langle x, p_{\mathbf{t}} \rangle \quad \text{and similarly for } L_{\mathbf{h}}.$$

The corresponding dual function inherits separability

$$\theta(x) = - \sum_{t=1}^T x^t d^t + \theta_{\mathbf{t}}(x) + \theta_{\mathbf{h}}(x)$$

for

$$\theta_{\nu}(x) = \min_{p_{\nu} \in D_{\nu} \cap S_{\nu,2}} C_{\nu}(p_{\nu}) + \langle x, p_{\nu} \rangle$$

where $\nu \in \{\mathbf{t}, \mathbf{h}\}$.

In this setting, a “subproblem” refers to the optimization problem above, which depends on the given x . The negative of each dual function is convex and nonsmooth

at those points x with more than one minimizer in the corresponding subproblem. More precisely, letting $\text{conv } P$ denote the convex hull of a set P ,

$$\partial(-\theta_\nu)(x) = \text{conv} \left\{ -p_\nu(x) : \text{for } p_\nu(x) \in D_\nu \cap S_{\nu,2} \text{ such that } \theta_\nu(x) = L_\nu(p_\nu(x), x) \right\}.$$

Therefore, evaluating each dual function and getting one subgradient involves the same computational effort, of finding a minimizer of the subproblem. This feature makes the dual problem suitable for *oracle* based methods, that is, algorithms designed on the knowledge, at each iterate, of the function value and of *one* subgradient (more details are given in Sec. 4 below).

3.2. Uncoupling via Variable Duplication. To achieve decomposition, we can introduce the artificial variable q_t duplicating p_t . We obtain the following equivalent problem:

$$(3.1) \quad \begin{cases} \min_{p_t, p_h, q_t} C_t(p_t) + C_h(p_h) \\ q_t \in D_t \cap S_{t,2} \\ p_h \in D_h \cap S_{h,2} \\ (p_t, p_h) \in S_1 \\ p_t = q_t, \end{cases}$$

where all the coupling is concentrated in the last constraint. In other words, without this constraint, it would be possible to solve separately two problems, on variables p_t , p_h , and q_t , respectively. Once again, relaxing the last constraint induces separability. This is because, for a multiplier vector $x \in \mathfrak{R}^{T I_t}$, the Lagrangian can be written in the form

$$L(p_t, p_h, q_t, x) = L_p(p_t, p_h, x) + L_q(q_t, x).$$

The *variable splitting* decomposition was introduced in [GK87] under the name of Lagrangean decomposition; see also [Gui03]. The approach yields purely thermal or purely hydro subproblems, corresponding to evaluating $\theta_i(x_i)$, along the lines in Section 3.1. Subproblems are typically network optimization problems, dealing with demand satisfaction and transmission or security constraints, as in [LPRS96], [DSM07].

Decomposition consists in solving, instead of (2.1), the convex dual problem

$$(3.2) \quad \min_{x \in \mathfrak{R}^T} (-\theta)(x),$$

by some nonsmooth method, which at iteration k updates the dual iterate x^k using the information of past oracle calls. As shown in Figure 2 In this *price-decomposition* scheme, the evaluation of the i -th dual function defines the i -th *subproblem*, while one iteration to solve (3.2) defines the *master program*.

With price decomposition, each subproblem involves variables corresponding to the i -th unit only. In the optimal power management setting, there will be purely thermal, or purely hydro subproblems. For unit-commitment scheduling, for instance, thermal subproblems are mixed 0-1 problems, often solved by dynamic programming, [MDS08], [FG06]. Hydro subproblems are generally large scale programs, sometimes involving 0-1 variables and/or nonlinear relations, [BDLM08], [FdS06].

Incidentally, an optimum \bar{x} of the dual problem (3.2) is nothing but a marginal cost, pricing the extra cost of satisfying an additional unit of demand. This *shadow price* is used by the ISO to establish price cap regulations in a competitive environment, or to define electricity prices in a centralized system.

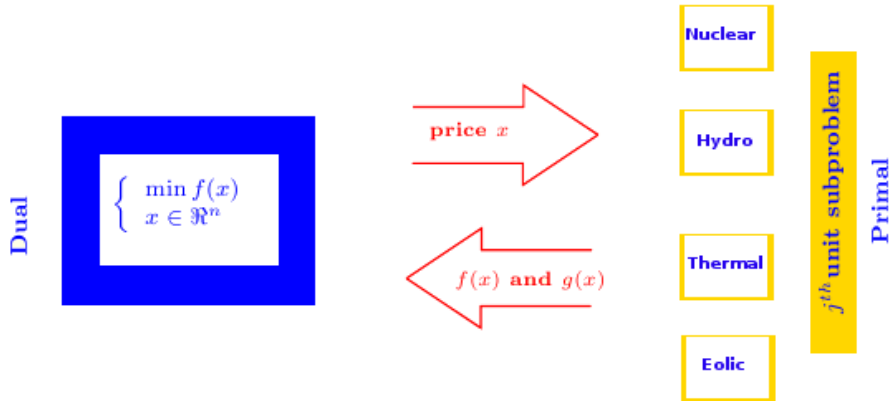


FIGURE 2. Price decomposition with exact subproblem solution

4. SOLVING THE DUAL PROBLEM

For simplicity, we now let $f = -\theta$ denote the objective function in (3.2). In our separable framework, this is a convex nonsmooth function whose evaluation at any given point x^k involves solving all the relevant subproblems. Once $f(x^k)$ is computed, the subproblems solution give a subgradient $g^k \in \partial f(x^k)$ for free:

$$g^k = d - \sum_{i=1}^I p_i(x^k) \quad \in \mathbb{R}^T \quad \text{for Section 3.1}$$

$$g^k = -p_t(x^k) + q_t(x^k) \quad \in \mathbb{R}^{T \times I_t} \quad \text{for Section 3.2.}$$

Note in passing that in both cases, the subgradient is the negative of the relaxed constraint.

When adopting a dual approach as in our setting, to successfully solve the original problem (2.1) amounts in particular to finding primal points that are feasible and satisfy the relaxed constraint. For the nonsmooth master program this means ensuring that, at least asymptotically, $g^k \rightarrow 0$. However, in nonsmooth optimization this is too strong of a requirement, as shown by the scalar function $f(x) = |x|$: only by hitting exactly its minimum we can hope to get a zero subgradient. Nonsmooth methods can at best provide *convex combinations*, using certain simplicial multipliers α_k , such that $\sum_k \alpha_k g^k \rightarrow 0$. This is because, from the primal point of view, any nonsmooth method asymptotically solves the *bi-dual* of the primal problem, which is a convexified variant of (2.1), see [FK00], [Lem01], [AW09].

In nonsmooth optimization, the three main classes of oracle based algorithms are the subgradient, cutting-plane, and stabilized cutting-plane methods, [BGLS06, Part II]. Regarding implementation, subgradients methods are the simplest one to code, while the last class, that includes bundle methods, is the more involved one. The increasing effort of implementation pays off when the problem needs to be solved with high precision.

Since accuracy of the dual solution is a concern in energy optimization, bundle methods appear as one of the best choices for the master program. Indeed, for a battery of unit-commitment problems with nine configurations of the Brazilian power system, the comparison in [BDMS03] shows numerically that subgradient methods struggle to achieve high accuracy for this type of problems.

A thorough description of nonsmooth algorithms, including bundle methods, can be found in [HUL93, vol. II] and [BGLS06, Ch. 10]. We will focus here on more recent bundle variants, capable of handling *inaccurate* function and subgradient evaluations, introduced in [Kiw06].

The reason is that having a nonsmooth method able to deal with inexact information makes it possible to accelerate calculations, for instance by exiting subproblems before optimality, or even by *skipping* some subproblem solution. The procedure is schematically represented by Figure 3.

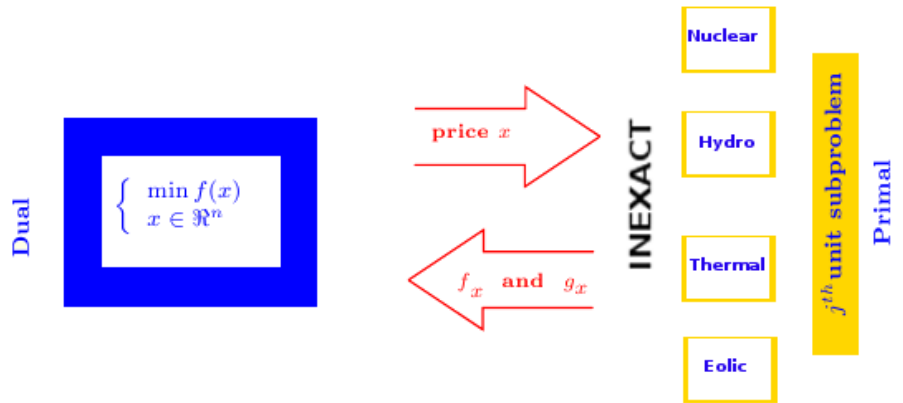


FIGURE 3. Price decomposition with inexact subproblem solution

The saving induced by inexact calculations can be significant. For a real-life French case, for a tolerance corresponding to a deviation in demand satisfaction of 20MW r (a negligible amount, if compared to the average power load of about 50.000MW), the (inexact) incremental bundle method [ES10] reaches the same precision as the proximal approach [LS97], but uses 25% less CPU-time.

4.1. Inexact Bundle Methods. We are interested in solving problem

$$\min\{f(x) : x \in \mathbb{R}^n\}$$

where f is a nonsmooth convex function, whose information is provided through an inexact oracle, like the one in Figure 3. Accordingly, instead of exact function and subgradient values, only *estimates* are available: $f_x/g_x \approx f(x)/g(x)$. The algorithm is based on

$$\begin{aligned} k & \text{ an iteration counter,} \\ \{x^j\}_{j \in J^k} & \text{ a set of points indexed by } J^k, \\ \hat{x}^k & \text{ the algorithmic center at iteration } k, \end{aligned}$$

that is, a point yielding a small enough function value, with respect to past iterations.

Since we are in a setting where subproblems are not solved exactly (or not solved at all), the function estimates f_{x^j} are denoted by

$$(4.1) \quad \begin{aligned} f^j & = f(x^j) - \varepsilon_f^j \quad \text{where } \varepsilon_f^j \text{ is an unknown error, and} \\ \hat{f}^k & = f(\hat{x}^k) - \hat{\varepsilon}_f^k \quad \text{where } \hat{\varepsilon}_f^k \text{ is an unknown error.} \end{aligned}$$

In these relations, the error sign is not specified, so that the true function value can be either overestimated or underestimated. Yet, errors are assumed to be bounded

$$(4.2) \quad |\varepsilon_f^j| \leq \bar{\varepsilon}^f \text{ for all } j,$$

noting that, since the algorithmic center is a past iterate, the bound above also holds for $\hat{\varepsilon}_f^k$, for all k . As for the subgradient estimates, $g^j = g_{x^j}$, for the moment we do not set any condition on their calculation, preferring to leave this point to Remark 4.1 below.

The *bundle information*

$$\mathbb{B}^k = \{\hat{x}^k, \hat{f}^k, \hat{g}^k\} \cup \{(x^j, f^j, g^j) : j \in J^k\},$$

is used to define a convex model for the function, denoted by \mathbb{M}^k . One possibility is to use the cutting-plane model, defined as the piecewise maximum of linearizations of f , that is, $\mathbb{M}^k(y) = \max\{f^j + \langle g^j, y - x^j \rangle : j \in \mathbb{B}^k\}$, and plotted in blue in Figure 4.

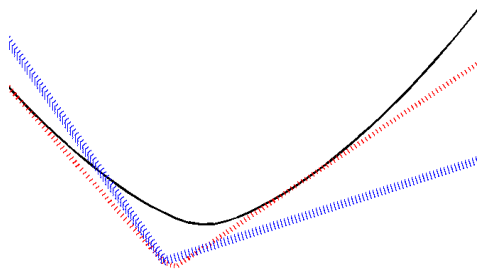


FIGURE 4. Exact cutting-plane model in red, inexact model in blue, for the function graphed in black

The choice of the model function is important for the defining the next iterate, since

$$x^{k+1} := \arg \min_y \{\mathbb{M}(y) + \frac{1}{2}\mu_k |y - \hat{x}^k|^2\}$$

for a given prox-parameter $\mu_k > 0$. Since this problem needs to be solved at every iteration, it must be reasonably easy. When using a cutting-plane model, finding a new iterate amounts to solving a quadratic programming problem.

By the optimality conditions of the problem above,

$$(4.3) \quad x^{k+1} = \hat{x}^k - \mu^k G^k \quad \text{where} \quad G^k \in \partial \mathbb{M}^k(x^{k+1}).$$

Once the new iterate is known, we define the *inexact aggregate linearization*

$$(4.4) \quad \mathbb{M}^k(y) := \mathbb{M}^k(x^{k+1}) + \langle G^k, y - x^{k+1} \rangle.$$

Clearly, because $G^k \in \partial \mathbb{M}^k(x^{k+1})$,

$$(4.5) \quad \mathbb{M}^k(y) \leq \mathbb{M}^k(y) \quad \text{for all } y \in \mathfrak{R}^n.$$

Finally, we consider the error and the optimality measures defined by

$$(4.6) \quad \mathbb{E}^k = \hat{f}^k - \mathbb{M}^k(\hat{x}^k) \quad \text{and} \quad V_k = \max \{ \mathbb{E}^k + \langle G^k, \hat{x}^k \rangle, |G^k| \}.$$

4.2. Handling Inexact Oracle Information. Since the function/subgradient information is inaccurate, the bundle method should check if the oracle noise becomes cumbersome and needs to be attenuated (think for instance of an extreme case, when the model is so polluted that it lies above the graph of the function: there is no chance to get a reasonable solution to the optimization problem).

A possibility to detect when noise introduced by the inexact evaluations became “too large” is the function value at the algorithmic center is *below* the minimum model value

(by convexity, this would never happen with exact oracle values). More precisely, when the noise measurement quantity defined below is negative:

$$\hat{f}^k - \left(\mathbb{M}^k(x^{k+1}) + \frac{1}{2}\mu_k|x^{k+1} - \hat{x}^k|^2 \right) < 0.$$

When the relation above holds, the algorithm maintains the model and the center, and reduces the prox-parameter. The new iterate yields a smaller noise measurement quantity, thus *attenuating* the noise induced by the inexact bundle information. For numerical convenience, we consider here an alternative mechanism that checks asymptotic satisfaction of the inequality above, based on a *relative* criterion. More precisely, noise is too large if

$$(4.7) \quad \text{NoNeg}_k := \frac{\hat{f}^k - \left(\mathbb{M}^k(x^{k+1}) + \frac{1}{2}\mu_k|x^{k+1} - \hat{x}^k|^2 \right)}{\frac{1}{2}\mu_k|x^{k+1} - \hat{x}^k|^2} < -\beta_k$$

for a parameter $\beta_k \in [0, 1]$.

Only when noise is declared acceptable, that is when $\text{NoNeg}_k \geq -\beta_k$, the algorithm examines if the new iterate is good enough to become the next center by checking if

$$(4.8) \quad f^{k+1} \leq \hat{f}^k - m\delta^k$$

for a certain nonnegative *predicted decrease*. When this relation holds, the iteration is declared *serious*, because it provided a new algorithmic center: $\hat{x}^{k+1} = x^{k+1}$. Otherwise, the iteration is declared *null*.

Usual definitions for the decrease are $\delta^k = \hat{f}^k - \mathbb{M}^k(x^{k+1})$, or $\delta^k = \hat{f}^k - \mathbb{M}^k(x^{k+1}) - \frac{1}{2}\mu_k|x^{k+1} - \hat{x}^k|^2$. We consider a slightly more general variant, and let

$$(4.9) \quad \delta^k := \hat{f}^k - \mathbb{M}^k(x^{k+1}) - \frac{1}{2}\alpha_k\mu_k|x^{k+1} - \hat{x}^k|^2 \quad \text{for a parameter } \alpha_k \in [0, 1].$$

With this definition, since the numerator in (4.7) equals $\delta^k - \frac{1}{2}(1 - \alpha_k)\mu_k|x^{k+1} - \hat{x}^k|^2$, we see that a noise attenuation step amounts to checking satisfaction of the inequality

$$(4.10) \quad \delta^k < \frac{1}{2} \left(1 - (\alpha_k + \beta_k) \right) \mu_k |x^{k+1} - \hat{x}^k|^2.$$

As explained below, this relation makes it possible to control the relation between noise attenuation and descent, a flexibility that can help the numerical performance of the algorithm. Generally, to progress towards a solution, it is preferable for the algorithm to:

- make more serious iterations, because serious iterates converge to a solution, and
- have few noise attenuation steps. Noise attenuation steps are undesirable to occur often, because they prevent the algorithm from having “true” iterates, for which to check the descent condition.

By (4.8), more serious iterations are achieved by taking larger α_k 's, while a larger β_k reduces the left hand side term in (4.10), making less likely noise attenuation. Since the choice of parameters α_k, β_k should ensure that the nominal decrease in (4.8) is nonnegative, in view of (4.10), the relation $(1 - \alpha_k - \beta_k) > 0$ should hold. Accordingly, we require that

$$(4.11) \quad \text{AB}_{\min} \leq \alpha_k + \beta_k \leq 1$$

for some positive constant AB_{\min} .

4.3. An Inexact Bundle Optimization Solver. An oracle computing approximate f values as in (4.1) as well as subgradient estimates is assumed to be given.

Step 0 (Input and Initialization) Select a initial starting point \hat{x}^0 a stopping tolerance $\text{TOL} \geq 0$, an Armijo-like parameter $m \in (0, 1)$. Initialize the iteration counter $k = 0$, the bundle index set $J^0 := \{0\}$, and the first candidate point $x^0 := \hat{x}^0$. Compute f^0 as well as g^0 . Choose the starting prox-parameter $\mu_0 > 0$, parameters $\alpha_0, \beta_0 \in [0, 1]$ satisfying (4.11).

Step 1 (Model Generation and QP Subproblem) Having the current algorithmic center \hat{x}^k , the bundle \mathbb{B}^k , and the prox-parameter μ_k , define the convex model function \mathbb{M}_k and compute $x^{k+1} = \arg \min \{ \mathbb{M}_k(y) + \frac{1}{2} \mu_k |y - \hat{x}^k|^2 : y \in \mathcal{R}^n \}$.

Define the predicted decrease δ^{k+1} as in (4.9), as well as the aggregate error and optimality certificate in (4.6).

Step 2 (Noise attenuation test)

If condition (4.10) is true, noise is too large: decrease the prox-parameter as in (4.12b) below; maintain the center and the bundle:

$$\left(\hat{x}^{k+1}, \mathbb{B}^{k+1} \right) = \left(\hat{x}^k, \mathbb{B}^k \right);$$

choose parameters $\alpha_{k+1}, \beta_{k+1}$ satisfying (4.11), and loop to Step 1.

Otherwise, if (4.10) does not hold, proceed to Step 3.

Step 3 (Stopping Test and New Oracle Information) Call the oracle to obtain f^{k+1}, g^{k+1} . If $V^{k+1} \leq \text{TOL}$ then stop.

Step 4 (Serious step test) Check the descent condition (4.8). If this condition is true, declare a serious iteration and set $\hat{x}^{k+1} = x^{k+1}$. Otherwise, declare a null step, maintained the center: $\hat{x}^{k+1} = \hat{x}^k$.

Step 5 (Bundle Management and updates) Choose a new prox-center μ_{k+1} satisfying (4.12a) or (4.12c) below, if the iteration was declared serious or null, respectively.

In all cases choose parameters $\alpha_{k+1}, \beta_{k+1}$ satisfying (4.11). Define the new bundle \mathbb{B}^{k+1} , for example by appending to the index set the last iterate information: $J^{k+1} = J^k \cup \{k+1\}$. Increase k by 1 and loop to Step 1. \square

Both in Step 2 and Step 5 there is some freedom in the choice of the new bundle \mathbb{B}^{k+1} . When noise is excessive, as in Step 2, the conservative choice of keeping the same cutting-planes models for f seems reasonable. As for Step 5, alternative choices for managing the bundle are possible as long as conditions (A.11) and (A.12) in the Appendix are satisfied. We refer to [BGLS06, Ch. 10] for more details.

The prox-parameter update depends on positive constants μ_{\max} and Δ , as follows:

$$(4.12a) \quad \mu_{k+1} \leq \mu_{\max} < +\infty \quad \text{if iteration } k \text{ was declared serious}$$

$$(4.12b) \quad \mu_{k+1} \leq \mu_k - \Delta < \mu_k \quad \text{if iteration } k \text{ resulted in noise attenuation}$$

$$(4.12c) \quad \mu_{k+1} \in [\mu_k, \mu_{\max}] \quad \text{if iteration } k \text{ was declared null, but } \mu^{k+1} \text{ increases} \\ \text{only if no noise was attenuated after generating } \hat{x}^k.$$

Finally, for all the bundle choices, the key is to ensure satisfaction

$$(4.13) \quad \mathbb{M}^k(y) \leq f(y) + \bar{\varepsilon}_g \text{ for all } y \in \mathcal{R}^n.$$

for some constant $\bar{\varepsilon}_g \geq 0$. Indeed, under this assumption, we show below that V^k is an *optimality certificate*.

Remark 4.1. *In our decomposition setting, the inexact oracle can be defined by stopping the subproblems solution before optimality is reached. For illustration purposes, consider Section 3.1 and suppose $x^j \in \mathfrak{R}^T$ is given. The nonsmooth objective function is $f(x) = \sum_{t=1}^T x^t d^t - \theta_{\mathbf{t}}(x) - \theta_{\mathbf{h}}(x)$, for*

$$\theta_{\nu}(x) = \min_{p_{\nu} \in D_{\nu} \cap S_{\nu,2}} C_{\nu}(p_{\nu}) + \langle x, p_{\nu} \rangle$$

where $\nu \in \{\mathbf{t}, \mathbf{h}\}$. The exact oracle values are given by minimizers $p_{\nu}(x^j)$ such that $\theta_{\nu}(x^j) = C_{\nu}(p_{\nu}(x^j)) + \langle x^j, p_{\nu}(x^j) \rangle$ and, hence,

$$f(x^j) = \sum_{t=1}^T x^{jt} d^t - C_{\mathbf{t}}(p_{\mathbf{t}}(x^j)) - \langle x^j, p_{\mathbf{t}}(x^j) \rangle - C_{\mathbf{h}}(p_{\mathbf{h}}(x^j)) - \langle x^j, p_{\mathbf{h}}(x^j) \rangle$$

and

$$g(x^j) = d - \sum_{i \in I_{\mathbf{t}}} p_{\mathbf{t}}(x^j) - \sum_{i \in I_{\mathbf{h}}} p_{\mathbf{h}}(x^j).$$

Any $p_{\nu}^j \in D_{\nu} \cap S_{\nu,2}$ can be used to build an inexact oracle, by letting

$$f^j = \sum_{t=1}^T x^{jt} d^t - C_{\mathbf{t}}(p_{\mathbf{t}}^j) - \langle x^j, p_{\mathbf{t}}^j \rangle - C_{\mathbf{h}}(p_{\mathbf{h}}^j) - \langle x^j, p_{\mathbf{h}}^j \rangle$$

and

$$g^j = d - \sum_{i \in I_{\mathbf{t}}} p_{\mathbf{t}}^j - \sum_{i \in I_{\mathbf{h}}} p_{\mathbf{h}}^j.$$

By definition of the dual functions, the corresponding linearizations satisfy the inequality $f^j + \langle g^j, y - x^j \rangle \leq f(y)$ for any $y \in \mathfrak{R}^n$. As a result, (4.13) holds for the cutting-plane model with $\bar{\varepsilon}_g = 0$.

Lemma 4.2. *If the algorithm builds models satisfying (4.13) the following holds:*

$$\hat{f}^k \leq \mathbf{M}^k(y) + V^k(1 + |y|) \quad \text{for all } y \in \mathfrak{R}^n.$$

As a result, if (4.13) holds, then

$$(4.14) \quad f(\hat{x}^k) \leq f(y) + \bar{\varepsilon} + V^k(1 + |y|) \quad \text{for all } y \in \mathfrak{R}^n.$$

Proof. Since the function \mathbf{M}^k from (4.4) is affine and G^k is its gradient, using the error definition in (4.6) and the Cauchy-Schwarz inequality we see that

$$\begin{aligned} \mathbf{M}^k(y) &= \mathbf{M}^k(\hat{x}^k) + \langle G^k, y - \hat{x}^k \rangle \\ &= \hat{f}^k - \mathbf{E}^k + \langle G^k, y - \hat{x}^k \rangle \\ &\geq \hat{f}^k - \left(\mathbf{E}^k + \langle G^k, \hat{x}^k \rangle + |G^k| |y| \right). \end{aligned}$$

The desired inequality follows from the definition of V^k in (4.6). Furthermore, using (4.1), (4.2), and (4.5),

$$\begin{aligned} f(\hat{x}^k) &= \hat{f}^k + \hat{\varepsilon}_f^k \leq \hat{f}^k + \bar{\varepsilon}^f \\ &\leq \mathbf{M}^k(y) + V^k(1 + |y|) + \bar{\varepsilon}^f \\ &\leq \mathbf{M}^k(y) + V^k(1 + |y|) + \bar{\varepsilon}^f, \end{aligned}$$

and (4.14) follows from (4.13). \square

The asymptotic analysis for inexact bundle methods is rather involved and technical, the full proofs can be found in the appendix. Basically, convergence of the algorithm

means that, at least for a subsequence, $V^k \rightarrow 0$. As a result, (4.14) gives that a cluster point of the serious subsequence is an $\bar{\varepsilon}$ solution to (3.2).

We now review how to estimate actual solutions of (2.1) by means the output of the bundle algorithm.

5. BACK TO THE PRIMAL PROBLEM

When adopting a dual approach as in our setting, to successfully solve the original problem (2.1) amounts in particular to finding primal points that are feasible and satisfy the relaxed constraint. For the nonsmooth master program this means ensuring that, at least asymptotically, $g^k \rightarrow 0$. However, in nonsmooth optimization this is too strong of a requirement, as shown by the scalar function $f(x) = |x|$: only by hitting exactly its minimum we can hope to get a zero subgradient. Nonsmooth methods can at best provide *convex combinations*, using certain simplicial multipliers α_j^k , such that $\sum_k \alpha_j^k g^j \rightarrow 0$. This is because, from the primal point of view, any nonsmooth method asymptotically solves the *bi-dual* of the primal problem, which is a convexified variant of (2.1), see [FK00], [Lem01], [AW09].

When the primal problem is convex to the extent that its KKT conditions are sufficient for optimality, solving the dual problem (3.2) is equivalent to solving the primal (2.1). Otherwise, the corresponding optimal values do not coincide and there is a positive duality gap. Having a small duality gap is a fundamental concern in power management problems (and one of the main reasons why Lagrangian Relaxation is so successful in the area). The geometrical study in [LR01] shows how different relaxation schemes impact on the size of duality gaps, with application to unit-commitment problems.

In general, the dual solution gives a primal point solving the bi-dual of (2.1). Following [DGL05], we refer to such primal point as a *pseudo* primal solution.

Pseudo solutions can be built by any nonsmooth algorithm, by making convex combinations of primal points; in [FK00] and [Lem01] there is a study for bundle methods, while [AW09] considers subgradient methods. Different approaches yield different convex multipliers, denoted by α in what follows. In bundle methods, the components of α correspond to the Lagrange multipliers in the following problem

$$(5.1) \quad \min_{r \in \mathfrak{R}, x} \left\{ r + \frac{1}{2} \mu |x - \hat{x}|^2 : r \geq f^j + \langle g^j, x - x^j \rangle \text{ for all } j \in \mathbb{B}^k \right\}.$$

Since this is just a rewriting of the bundle QP problem in Step 2 of the algorithm, the value of $\alpha \in \mathfrak{R}^{|\mathbb{B}|}$ is available at each iteration. The corresponding convex combination of the bundle subgradients (which is shown to asymptotically vanish for the serious subsequence) gives the pseudo primal point, that we denote with a “hat” superscript: (5.2)

$$\begin{aligned} \sum_{j \in \mathbb{B}^k} \alpha_j^k g^j &= d - \sum_{j \in \mathbb{B}^k} \alpha_j^k \sum_{i=1}^I p_i(x^j) & \text{so } \hat{p}^j &= \sum_{j \in \mathbb{B}^k} \alpha_j^k p(x^j) & \text{for Section 3.1} \\ \sum_{j \in \mathbb{B}^k} \alpha_j^k g^j &= - \sum_{j \in \mathbb{B}^k} \alpha_j^k p(x^j) + \sum_{j \in \mathbb{B}^k} \alpha_j^k q(x^j) & \text{so } \begin{cases} \hat{p}^j = \sum_{j \in \mathbb{B}^k} \alpha_j^k p(x^j) \\ \hat{q}^j = \sum_{j \in \mathbb{B}^k} \alpha_j^k q(x^j) \end{cases} & \text{for Section 3.2.} \end{aligned}$$

Under reasonable assumptions, see e.g. [FK00] or [Lem01], the final pseudo primal point, built with the convex multipliers corresponding to a solution of (3.2), solves the

bi-dual of problem (2.1). When (2.1) is not sufficiently convex, (2.1) and its bi-dual differ, so the final pseudo primal point will often fail to be feasible. For the dualization in Section 3.1, if the dynamic set involves 0-1 variables, the components of the pseudo point may just belong to the interval $[0, 1]$. Likewise for Section 3.2, the pseudo point \hat{p} is built with points satisfying the dynamic constraints only and, hence, it may not satisfy the static constraints. Reciprocally for \hat{q} , for which it may happen that $\hat{q}_i \notin D_i$ for some units i .

Therefore, after solving (3.2) it is necessary to proceed to a second phase to recover primal feasibility. In some cases, for instance for unit-commitment scheduling, the sequence of pseudo primal points can be used as a guide to quickly produce primal feasible points, as in [FG06]. Heuristic techniques are popular for dealing with the combinatorial nature of such problems. These are essentially convexification procedures that try to generate primal solutions even for discrete or non convex feasible sets. In such a setting, not much can be shown theoretically regarding the ability of the approaches in actually producing merely feasible primal points. Even less can be said in terms of primal optimality. We refer to [Sag12] for more details.

6. MULTISTAGE DECISION PROBLEMS

Some power problems are naturally sequential, in the sense that decisions are taken in a consecutive manner, over different times. When expanding the system capacity, for instance, it is natural to decide on the generation of each plant only after having decided which plants (new ones and/or old ones with added capacity) define the expanded system. For such two-stage processes, decomposition can be done in two levels. The water pricing mechanism mentioned in the final paragraph in Section 1 (and used to define the operating hydro cost in Section 2) is an example of a multistage decision process that covers several years.

The generation of hydro-power is a prototypical example of decisions taken along different time steps, as uncertainty realizes. The following simple model illustrates well key issues to be explored by the decomposition alternatives in this section.

6.1. A Simplified Model for Hydro-Power Management under Uncertainty.

For mid and long-term problems, a hydro-power dynamic set D_i may refer to a power plant i , to a hydro-valley with several cascaded hydro-plants, or to an equivalent reservoir aggregating all the hydro-plants in a geographical region. The corresponding generation p_i has a dynamic evolution depending on two variables: a state variable v_i^t and a control variable u_i^t . The state variable refers to the reservoir volume at the end of the time step; the control variable is the turbined outflow generating the power (there is no spillage).

At each time step, the reservoir volume decreases as a consequence of the turbined outflow (there is no evaporation) and increases by contributions of rain and upstream rivers. Since the natural water streamflow arriving at time t to the reservoir is uncertain and may depend on the past, we let $\xi^{[t]}$ denote the process history until time t . Hydro-generation is expressed as a function relating water to energy, depending not only on the state and control variables but also on uncertainty: $p^t = p(v^t, u^t, \xi^{[t]})$ for some function $p(\cdot)$.

The process ξ can be multivariate (with as many components as hydro-reservoirs compose the power mix), has sample space Ω , probability distribution \mathbf{P} , and generates the sigma-algebra $\mathcal{F}^t = \sigma(\xi^\tau, \tau \leq t)$. Since the realization $\tilde{\xi}^t$ becomes known at the

beginning of time step t , for some $q \geq 2$, the state and control variables (i.e., v^t and u^t) as well as the hydro-generation (p^t) are functions in $L_q(\Omega, \mathcal{F}^t, \mathbf{P})$.

For hydro-plants, the relation $p_i \in D_i$ in fact constrains the state and control variables; for example, the transition equation for the reservoir volume, or *water balance* constraint:

$$v^t = v^{t-1} - u^t + \xi^t,$$

which, by its physical nature, needs to be satisfied almost surely (a.s.). This relation the stochastic counterpart of the deterministic streamflow equations (2.3).

Another important constraint refers to *minzones*, or low critical values for the storage levels that the ISO is often concerned with:

$$v^t \geq v^{t,crit}.$$

The fact that a reservoir level drops below its minzone alerts the ISO to trigger preventive measures, set to avoid a lack of energy in the future. Accordingly, this constraint can be satisfied with high probability, not necessarily equal to one.

In what follows, given a past history of realizations $\tilde{\xi}^{[t]}$, we will consider the abstract dynamic constraints

$$(6.1) \quad A^t p^t + B^t p^{t-1} \geq b^t(\tilde{\xi}^{[t]}) \text{ for } t = 1, \dots, T,$$

with right-hand side uncertainty only for some vector functions $b^t(\cdot)$. The matrices A^t, B^t are deterministic and have suitable dimensions. We adopt this setting, said with *fixed recourse*, for simplicity and because it covers the hydro-power management problem considered here: when the generation function $p(\cdot)$ is affine, the water balance and minzone constraints can be written in the form above. In (6.1) the first time stage involves no uncertainty, because the right-hand side vector $b^1(\xi^1) = b^1$ is deterministic: all realizations $\tilde{\xi}$ of uncertainty have the same first component. Also, p^0 is known and depends on the initial volumes of the reservoirs: $p^0 = p(v^0)$.

The constraints above seem different in nature from the ones considered so far, separating dynamic and static constraints. Nevertheless, the abstract feasible set in (2.1) is covered by our stage-wise writing. The relation with the dynamic set D in (2.1) is straightforward. As for the static set S therein, it is not difficult to write static relations (such as demand satisfaction) in the form (6.1), by properly setting to zero some rows in the matrices B^t , thus eliminating the dependence on p^{t-1} for some constraints.

Costs could be convex functions; we focus here in linear programs, without loss of generality. Moreover, in this stage-wise setting, it is also convenient to write such costs (associated to thermal generation and deficit) as a sum over time steps. Therefore, knowing an uncertainty realization $\tilde{\xi}$, a particular instance of (2.1) has the form:

$$\left\{ \begin{array}{l} \min_p \quad \sum_{t=1}^T \langle C^t, p^t \rangle \\ s.t. \quad \text{for } t = 1, \dots, T : \\ \quad A^t p^t + B^t p^{t-1} \geq b^t(\tilde{\xi}^{[t]}). \end{array} \right.$$

This type of problem, however, needs to be “solved” for *any realization* $\tilde{\xi}$, giving a meaning in particular to constraint satisfaction: either by considering only the worst-case scenario; or by requiring satisfaction with probability one or with high probability; or by satisfying some stochastic dominance order. In general, since decision vectors are in fact functions, unless the stochastic process has a finite number of realizations, we will likely be dealing with an infinite dimensional problem.

In order to stay in a finite-dimensional setting (and thus take advantage of nonlinear programming techniques), one possibility is to use the Sample Average Approximation (SAA). Other approaches, formulating the problem in a robust, chance-constrained, or risk-averse manner are possible too.

The SAA approach discretizes the random variable ξ into a sample with N equiprobable scenarios $\{\xi(1), \dots, \xi(N)\}$, each one with length T ; see [KSHdM02]. The resulting optimization problem has N copies of the constraints above and, in a risk-neutral setting, uses as objective function the average cost over the sample.

6.2. Decomposing Two-stage Problems. We start with decomposition methods applicable when decisions are taken in two steps. Such is often the case for energy problems dealing with capacity expansion [BS88], generation planning [TBL96], [NR02], [CS97], environmental control [MS07], and finance [BS11], [FMR11].

Suppose that we set $T = 2$, and recall the first-stage vector b^1 is deterministic. With a slight abuse of notation, and because uncertainty arises only in the second stage, for the SAA we just consider the second components of the N scenarios: instead of $\tilde{\xi}$, a realization is denoted below by $\xi(i)$, for $i = 1, \dots, N$.

For a given initial state p^0 , the decision process solves the first-stage problem

$$(6.2) \quad \begin{cases} \min_{p^1} & \langle C^1, p^1 \rangle + \mathbf{E}[Q(p^1; \xi)] \\ \text{s.t.} & p^1 \in P^1 = \{q^1 : A^1 q^1 + B^1 p^0 \geq b^1\}. \end{cases}$$

The second term in the objective is the expected *recourse function*, defined over the optimal values of the second-stage problem

$$(6.3) \quad Q(p^1; \xi) = \begin{cases} \min_{p^2} & \langle C^2, p^2 \rangle \\ \text{s.t.} & p^2 \in P^2(p^1, \xi) = \{q^2 : A^2 q^2 + B^2 p^1 \geq b^2(\xi)\} \quad a.s., \end{cases}$$

where the inequality is to be satisfied almost surely, for each considered realization of the uncertainty ξ .

For any fixed p^1 , the SAA uses $\xi \in \{\xi(i), i = 1, \dots, N\}$, the N samples of the random data, to define the feasible sets $P^2(p^1, \xi(i))$ and approximate the expectation by averaging the expected values $Q(p^1; \xi(i))$.

Given p^1 and $\xi(i)$, the value $Q(p^1; \xi(i))$ is finite and can be computed either by solving (6.3) or its dual

$$(6.4) \quad \begin{cases} \max_{\lambda \geq 0} & \langle \lambda, b^2(\xi(i)) - B^2 p^1 \rangle \\ \text{s.t.} & A^{2\top} \lambda = C^2, \end{cases}$$

if both the primal and dual feasible sets are nonempty. As explained in [SDR09, Ch. 2], in this case the function $Q(\cdot; \xi(i))$ is convex, polyhedral, and subdifferentiable at p^1 , with

$$\partial Q(p^1; \xi(i)) = -B^{2\top} \lambda(p^1; \xi(i)) \quad \text{where} \quad \lambda(p^1; \xi(i)) \text{ solves (6.4).}$$

As with the Lagrangian Decomposition, the evaluation of the nonsmooth function and the computation of one subgradient entail the same computational effort, involving the solution of some optimization problem. In this Benders-like decomposition, the optimization problem (6.2) can be solved by any nonsmooth minimization algorithm. The expected recourse function is then approximated by some model \mathbb{M} , incorporating linearizations defined with function and subgradient information computed along iterations. For instance, the well-known L-shaped method [VW69] applies an algorithm of cutting planes; which is slow to converge, at least if an accurate solution is required. If

a bundle method was to be used instead, the fact that now (6.2) is constrained should be taken into account (the convex dual problems in Section 3 are unconstrained). Since the feasible set P^1 in (6.2) is a polyhedron, the corresponding affine constraints can be added to the bundle QP (5.1) without much difficulty.

It is shown in [SDR09, Ch. 5.8.2] that, under mild conditions, the accuracy in solving an SAA of (6.3) is of order $O(1/N^2)$. Therefore, to ensure an accurate solution, it is important to take N sufficiently large, using as many scenarios as possible. But a large number of scenarios involves the solution of many linear programs computing $Q(p^1; \xi(i))$ for $i = 1, \dots, N$. Such calculations can considerably slow down the optimization process of (6.2), because numerous linear programs need to be solved, and this, for each first-stage value p^1 computed along iterations.

Alternatively, for some scenarios one can either solve (6.3) or (6.4) only approximately, or just skip the solution of the linear program, replacing the missing information by some reasonable value. From the bundle optimization point of view, this amounts to having inexact linearizations for the recourse function, “polluted” with noise, that needs to be handled by a noise attenuation step.

The inexact bundle method in [OSS11] uses several different inexact approximations of (6.3) for two-stage stochastic linear programs. For the battery of examples considered therein, the best variant is the one that solves exactly (6.3) for a reduced subset of scenarios

$$\left\{ \xi(j_i) : i = 1, \dots, n \right\} \quad \text{with } n \ll N.$$

Having the corresponding dual solutions $\lambda(p^1; \xi(j_i))$, the optimal values of each unsolved problem (6.3) is approximated by the expression

$$\langle \lambda(p^1; \xi(j_i)), b^2(\xi) - B^2 p^1 \rangle.$$

In view of the equivalence (6.3)-(6.4), the scenario j_i chosen for the approximation above is the one for which the vectors $b^2(\xi) - B^2 p^1$ and $b^2(\xi(j_i)) - B^2 p^1$ are the closest. Because in (6.4) the feasible set is not uncertain but deterministic, the value above is a lower estimate for $Q(p^1; \xi(j_i))$ and the vector $-B^{2\top} \lambda(p^1; \xi(j_i))$ is a sound proxy subgradient.

The benchmark in [OSS11] shows that taking $n = 0.2N$ is sufficient for the inexact bundle method to obtain solutions with the same accuracy as the L-shaped method, but four times faster.

For related works we refer to [ZPR00], [HS94], [Yak94], [FS07]; see also [OS12].

6.3. Decomposing Multistage Problems. When $T > 2$, the accuracy estimate $O(1/N^2)$ still holds. Only that now $N = \prod_{i=2}^T N^t$, where N^t is the size of discretization chosen for ξ^t , the t -th component of uncertainty. In order to prevent N from growing exponentially, it is important to maintain controlled the number of scenarios, for example by employing some scenario selection technique [HR11]; see also [OSP⁺10] for an application in two-stage generation planning. For some problems, it may also be possible to use multi-scaling, regrouping several time steps into a single decision stage, as in [SYG06].

The multistage variant of the bi-level problem (6.2)-(6.3) has a multilevel formulation, nested over the time steps $t = 1, 2, \dots, T$. At each decision level t , uncertain data depends on the history of the stochastic process so far, denoted by $\xi^{[t]}$ (rainfall in a given month t depends on the pluvial regime of some past months, for instance). This feature is represented by a right hand side function $b^t(\cdot)$ from (6.1) that is affine on uncertainty, a particular setting general enough to cover seasonal time series modeled

as periodic autoregressive processes (like rains), as well as other forecasting models. Interstage dependent processes make it necessary to keep track of the relevant history of uncertainty along the solution procedure: the state is extended from p^t to $(p^t, \xi^{[t]})$. This feature, culprit of the well-known dimensionality curse in dynamic programming, notably complicates the numerical solution of the multistage stochastic program under consideration.

The goal is to find a *policy* or operation strategy that keeps the power system controlled in a sufficiently optimal way, both in terms of cost and minzone constraints, along the planning period. Similarly to (6.2)-(6.3), but for T decision levels, a nested decomposition defines the $(t + 1)$ -th recourse function $Q^{t+1}\left((p^t, \xi^{[t]}; \xi^{t+1})\right)$, to be approximated by cutting planes (regarding the water pricing mechanism mentioned in Section 1, the value of water is nothing but the final cutting-plane approximation for the first recourse function). The t -th optimization problem is akin to (6.2), replacing subindices “1” by t and “2” by $t + 1$. Also, the expected recourse function is now conditioned to the known history $\xi^{[t]}$.

Once more, a process discretization gives a sample of equiprobable scenarios, and the SAA problem requires satisfaction of constraints over the sample. Only that now the number of scenarios is huge, making a full parsing of the scenario tree computationally impossible. The algorithm of stochastic dual dynamic programming (SDDP), introduced in [PP91], addresses this issue by *sampling* paths over the scenario tree. Even though the sampling smartly overcomes the dimensionality curse to some extent, for long time horizons SDDP computational effort still remains significant (even after parallelization of the so-called forward pass).

We refer to [Sag12] and references therein for details.

6.4. Risk Averse SDDP Variants. The formulation above defines a problem like (6.2) at the t -th stage. These problems involve the expected value of the recourse function, conditioned to the history of realizations. With such a risk neutral approach, the ISO cannot hedge against extreme droughts, likely to make the reservoirs drop below the important minzone constraints. This concern has been addressed only recently, in two distinct ways, that could be thought of as having a *monetary* and *physical* focus, respectively.

6.4.1. Monetary Approaches. The first class, considered in [Sha11] and [PdM10], basically replaces the expectation by a *risk functional*, defined by a convex combination of the expectation and the Conditional Value at Risk (CVaR) of the recourse function. By [RU02], for a confidence level ε and a random variable Q representing a cost, its CVaR of level $\varepsilon \in (0, 1)$ is given by

$$CVaR_\varepsilon[Q] = \min_{z \in \mathbb{R}} \left(z + \frac{1}{\varepsilon} \mathbf{E}[\max(Q - z, 0)] \right).$$

Since the operation $\max(\cdot, 0)$ is not linear, a new state variable needs to be added, to keep track of the Value at Risk, given by the minimizer z above. Then a sampling method along the lines of SDDP is applied to a problem like where the expected recourse function is replaced in the objective by the risk functional

$$(1 - \kappa) \mathbf{E} \left[Q^{t+1} \left((p^t, \xi^{[t]}, z^t); \xi^{t+1} \right) | \tilde{\xi}^{[t]} \right] + \kappa CVaR_\varepsilon \left[Q^{t+1} \left((p^t, \xi^{[t]}, z^t); \xi^{t+1} \right) | \tilde{\xi}^{[t]} \right],$$

for $\kappa \in [0, 1]$: the closer to 1 the convex parameter is, the higher will be the risk aversion.

As discussed in [Sha11, Sec. 5], the risk averse formulation lacks an implementable stopping criterion. Both [PdM10] and [ST11] stop the procedure after having done the same number of iterations than those required for the risk neutral SDDP to trigger its stopping test.

6.4.2. *Physical Approaches.* The risk functional tries to find a compromise between minimizing the average cost and keeping controlled large increases in the recourse function, measured by its CVaR. The tuning on how such compromise is found is done by the parameter κ , which can vary along time steps, for instance increasing risk aversion as stages are further away in the future.

For the water pricing problem, the goal is to ensure satisfaction of the minzone constraints, preferably controlling that violations are not too large, if they occur. In this setting, the risk functional can be seen as a tool converting into a cost the future impact of such violations, thus justifying our naming “monetary” for the approach. The weight of that monetary impact depends on the parameter κ , which has no intuitive physical meaning with respect to the minzone constraints. As observed in [PdM10], the choice of such parameter is delicate, because it may force to operate the system in a manner resulting into a deficit of energy, not to violate the minzones. This is certainly not the best ISO behavior: minzones are meant to trigger some alert, *before* blacking out the power system.

The works [GS12a], [GS12b] consider an alternative approach, that handles uncertainty along a rolling horizon by means of individual chance constraints. Probabilistic constraints give a physical interpretation to risk, which in the water resource application is hydraulic, and not of financial nature. For the hydro-power management application considered in [GS12a] and [GS12b], mimicking the Brazilian power system, the rolling horizon policy compares favorably with the policy obtained by SDDP, involving computational times that are seven times inferior. Like SDDP, the policy built by this approach is implementable, non anticipative, and time consistent, only that computed with much less computational effort. Nevertheless, rather than a closed-loop feedback like SDDP, the rolling horizon approach can be seen as an adaptive open-loop controller, without feedback.

At this point, a word of caution is in order. The choice of a particular technique for handling uncertainty is clearly problem-dependent. To certain extent, such choice is also “school” dependent. The *SAA-school* advocates for the use of scenario trees; but there is another important group of researchers who prefer to formulate the problem in a continuous setting and approximate the probability distribution in some manner. There is always a trade-off to be found, for neither approach is free from weaknesses and/or criticism (a similar phenomenon appears in PDE-optimization and the approaches “First discretize, then optimize” versus “First optimize, then discretize”, [HT10]). Factors such as availability of sound statistical tools for approximating the distribution when compared to sampling; or the eventual use of the numerical solution found (strategic, like electricity prices, or just indicative, like future mean deficits) certainly play an important role in the decision.

7. VARIATIONAL INEQUALITIES DECOMPOSITION

Due to the presence of relatively few companies generating power in a given region, electricity markets are naturally set in an oligopolistic competition framework. Variational inequality models are useful to model strategic interactions in multi-settlement

power markets. Specifically, generating companies usually manage their power by optimizing a portfolio of alternative selling opportunities, given in the form of long-term contracts and immediate sales. Long-term contracts are negotiated in the *forward market*, for the company to deliver at given times, and for the contracted price, fixed amounts of energy during a period of five years or more. Immediate sales are settled in the *spot market*. The interest of this real-time market is that usually spot prices are higher than contracted ones. Nevertheless, since neither the price nor the quantity to be sold in the spot market is certain, contracts give to companies a sure and fixed income, thus hedging the risk of the spot market sales.

In centralized environments, there is no spot market and the ISO defines the price based on the marginal cost or shadow price. The situation is quite different in a deregulated market where, to coordinate the pricing, the ISO needs to analyze the competition mechanism of agents. We will focus here on game-theoretic formulations, although other approaches such as bi-level programming and mathematical programs with equilibrium constraints are also possible. References on the topic are numerous, we refer to the bibliography in [Sag12] for a thorough list.

7.1. Generalized Nash Games for Electricity Markets. Like the generation problems in Sections 3 and 6, game models can vary with the considered horizon (short/long term), the modeling of uncertainty and risk hedging, or the level of detail of technological features of power plants. Most importantly, models differ on the role of the ISO in the game: the ISO can be seen as a leader followed by the agents, or simply as another player, in equal terms with respect to the generating companies. We adopt this last model, said with *bounded rationality*, in which players ignore strategies of the adversaries and all decisions are taken simultaneously.

For convenience, the ISO is considered player number 0 while agents owning one or more generating companies are players with numbers $i = 1, \dots, N$. Accordingly, the i -th player owns n_i power plants, with generation denoted by $p_i \in \mathfrak{R}^{n_i}$ for $i = 1, \dots, N$. As in (2.1), each power plant generation is constrained by its own technology. Such constraints are abstractly denoted here by $p_i \in P_i$ for a suitable set P_i . Since generating companies aim at maximizing revenue, the i -th player problem is

$$\begin{cases} \max_{p_i} & r_i(\pi, p_i) \\ \text{s.t.} & p_i \in P_i, \end{cases} \quad \text{where} \quad r_i(\pi, p_i) = \pi \mathbf{1}_i^\top p_i - C_i(p_i)$$

defines the agent's profit. Specifically, $\mathbf{1}_i$ is an n_i -dimensional vector with all components equal to 1, p_i the generation offered to the market, C_i a convex cost function, and $\pi \in \mathfrak{R}$ is the unit price. Such price, that the ISO will pay to generators for electricity, is an unknown scalar. In our model, the price is a smooth concave function of the total generation offered for sale, $\sum_{i=1}^N \mathbf{1}_i^\top p_i$. For example, it can be a concave quadratic, defined by exogenous parameters that are market-dependent.

The purpose of the game is to find an equilibrium allowing generators to maximize profit and, at the same time, supplying the energy demand in a manner that is satisfactory for consumers and the society.

The ISO behavior is also driven by maximization of a suitable profit function, representing social welfare. Unsupplied demand entails the payment of a (high) penalty fee pen , fixed by the regulatory body. Also, the corresponding deficit of energy, denoted by p_0 in what follows, cannot be larger than a constant $\overline{\text{def}}$. We once more let p_{-i} denote the vector obtained after eliminating the sub-vector p_i from $p = (p_0, p_1, \dots, p_N)$. With

this notation, the total generation is obtained by summing the vector p_{-0} over all its components.

Suppose for simplicity demand of energy is exogenous and denoted by $d \in \mathfrak{R}$. Then the ISO seeks a solution to the problem

$$\left\{ \begin{array}{l} \max_{p_0} r_0(p_0, p_{-0}) \\ s.t. \quad p_0 \in P_0 = [0, \overline{\text{def}}] \\ \sum_{i=1}^N \mathbf{1}_i^\top p_i + p_0 = d, \end{array} \right. \quad \text{where } r_0(p_0, p_{-0}) = \pi \sum_{i=1}^N \mathbf{1}_i^\top p_i - C_i(p_i) - \text{pen} p_0.$$

Instead of generating costs, social welfare could include *wheeling revenues*, or charges paid to the network operator for transmission of the generated power. The justification here is that discouraging transmission expenses amounts to compete for using a, possibly congested, network of electric power lines. In this case, Kirchoff's laws for the network enter the ISO optimization problem.

To cast the ISO problem as one more player in the game, we let $\mathbf{1}_0 = 1 \in \mathfrak{R}$, $n_0 = 1$:

$$(7.1) \quad \text{for } i = 0, \dots, N \quad \left\{ \begin{array}{l} \max_{p_i} r_i(p_i, p_{-i}) \\ s.t. \quad p_i \in Q_i(p_{-i}), \end{array} \right. \\ \text{where} \quad Q_i(p_{-i}) = P_i \cap \left\{ q_i \in \mathfrak{R}^{n_i} : \mathbf{1}_i^\top q_i + \sum_{j \neq 0}^N \mathbf{1}_j^\top p_j = d \right\}.$$

Due to the coupling of the $N + 1$ feasible sets, the equilibrium conditions for (7.1) are given by a quasi-variational inequality, usually very difficult to solve. Finding a generalized Nash equilibrium of the game above is equivalent to solving a quasi-variational inequality problem, [FK10]. Under favorable circumstances, it is possible to find some, but not all, of the Nash points by solving a variational inequality instead (an easier problem). Those particular Nash points, called *variational equilibria*, have useful economical interpretation; see [FFP07]. For our game, the specific form of the feasible sets $Q_i(p_{-i})$, with a shared constraint, makes it possible to find variational equilibria, as explained below.

Let $n = \sum_{i=0}^N n_i$ and let $\nabla_i r_i(p)$ denote the n_i -dimensional vector obtained by differentiating the function r_i with respect to the components in the sub-vector p_i . Then, the feasible set

$$(7.2) \quad \mathbb{P} = \prod_{i=0}^N P_i \cap Q \quad \text{with} \quad Q = \left\{ q = (q_0, q_1, \dots, q_N) \in \mathfrak{R}^n : \sum_{i=0}^N \mathbf{1}_i^\top q_i = d \right\}$$

and the operator $F : \mathfrak{R}^n \rightarrow \mathfrak{R}^n$

$$(7.3) \quad F(p) = \left(\nabla_0 r_0(p), \nabla_1 r_1(p), \dots, \nabla_N r_N(p) \right),$$

not necessarily monotone, define the variational inequality

$$(7.4) \quad \text{finding } \bar{p} \in \mathbb{P} \text{ such that } \langle F(\bar{p}), p - \bar{p} \rangle \geq 0 \text{ for all } p \in \mathbb{P}.$$

A solution to this variational inequality gives a variational equilibrium for game (7.1).

7.2. Finding Variational Equilibria. When the ISO coordinates the pricing of a region with many power plants and agents selling energy, the game (7.1) becomes large and complex. The number of variables and constraints further increases in the presence of uncertainty, even more so if risk is a concern. In this situation, the variational

inequality (7.2)-(7.4) is potentially difficult to solve. However, the feasible set has a specific structure, that is amenable to decomposition: without the shared constraint of demand, the feasible set \mathbb{P} from (7.2) would decompose along players. The operator (7.3) is not separable as a sum, but it is possible to induce separability by a Jacobi-like approximation. We now describe a decomposition approach from [LSS11], that is applicable not only to non-monotone single-valued continuous operators like (7.3), but also to set-valued operators, as long as they are maximal monotone. Also, the feasible set \mathbb{P} can be the intersection of two closed sets, defined by affine equality and convex inequality constraints.

For linear programming problems, each iteration of the classical Dantzig-Wolfe decomposition is divided into two phases, dealing respectively with a master program and several subproblems solution. Basically, the master program solves the original problem over an easier feasible set. For (7.4), this means taking points in the set Q from (7.2) only using convex combinations of certain past iterates. Subproblems, instead, maximize a relaxed form of the individual profits, over each set P_i .

The method in [LSS11] follows the pattern above, with the additional feature of handling inexact subproblem solution in a rather comprehensive framework, which is shown to be convergent under mild assumptions. For illustration, we give here a combination of the Jacobi-like techniques with a Josephy-Newton approximation for the particular variational inequality problem (7.2)-(7.4).

Specifically, at the k -th iteration, feasible points $\{\mathbf{s}^0, \dots, \mathbf{s}^k\} \subset \mathbb{P}$ are available. Instead of (7.2)-(7.4), the *master problem* defines the set

$$\mathbb{Q}^k = \left\{ q \in \mathfrak{R}^n : q \in \text{conv}\{\mathbf{s}^0, \dots, \mathbf{s}^k\} \text{ and } \sum_{i=0}^N \mathbf{1}_i^\top q_i = d \right\}$$

and solves the variational inequality

$$\text{find } \mathbf{m}^{k+1} \in \mathbb{Q}^k \text{ such that } \langle F(\mathbf{m}^{k+1}), p - \mathbf{m}^{k+1} \rangle \geq 0 \text{ for all } p \in \mathbb{Q}^k.$$

The optimal multiplier corresponding to the (shared) demand constraint, $\lambda^{k+1} \in \mathfrak{R}$, defines a relaxed term to be incorporated in the subproblems, so that the corresponding feasible sets become separable. In view of the simple shared constraint in (7.2), such relaxed term has the form $\lambda^{k+1}(1, \dots, 1)^\top \in \mathfrak{R}^n$.

As mentioned, to induce subproblems separability, a Jacobi approach for (7.3) gives the operators

$$\nabla_i r_i(p_i, \mathbf{m}_{-i}^{k+1}), \quad \text{on sub-variables } p_i \text{ only.}$$

Afterwards, a Newton linearization around \mathbf{m}_i^{k+1} yields the approximation

$$F_i^{k+1} : \mathfrak{R}^{n_i} \rightarrow \mathfrak{R}^{n_i} \quad \text{defined by} \quad F_i^{k+1}(p_i) = \nabla_i r_i(\mathbf{m}^{k+1}) + \nabla_i^2 r_i(\mathbf{m}^{k+1})(p_i - \mathbf{m}_i^{k+1}).$$

The i -th *subproblem* adds to this approximation the relaxation term above and a regularization term, using a positive semidefinite matrix M_i^{k+1} of order n_i , to

$$\begin{aligned} &\text{find } \mathbf{s}_i^{k+1} \in P_i \text{ such that} \\ &\langle F_i^{k+1}(\mathbf{s}_i^{k+1}) + \lambda^{k+1} \mathbf{1}_i + M_i^{k+1}(\mathbf{s}_i^{k+1} - \mathbf{m}_i^{k+1}), p_i - \mathbf{s}_i^{k+1} \rangle \geq 0 \text{ for all } p_i \in P_i. \end{aligned}$$

In this expression, the role of matrices M_i^{k+1} is to ensure subproblems resolvability. Therefore, such matrices are null if the approximation F_i^{k+1} is already strongly monotone.

The procedure continues until satisfaction of certain stopping test, which essentially ensures that \mathbf{m}^k and \mathbf{s}^k are close enough, thus yielding a good estimate for a solution \bar{p} ; see [LSS11, Sec. 2.4]. The numerical experience in this work uses the well-known

PATH solver [FM99] on a battery of 60 test-cases, with up to $n = 10.000$ power plants divided among $N = 5$ players and the ISO, testing different solution procedures. The benchmark shows that, even with an extremely simplified modeling of the power system, large instances become intractable with a direct solution method, and can only be solved by decomposition. The benchmark also puts in evidence the good properties of the Jacobi-Newton variant above regarding scalability: as n grows, the computational effort grows very slowly, at a linear rate.

The Dantzig-Wolfe decomposition was applied for the first time to variational inequalities in [FC05]. In this work, the authors consider single-valued strictly monotone operators and assume that subproblems are solvable. The proposal was extended in [CF10] to handle more general situations, including Jacobi approximations (but no Newtonian linearizations). Related decomposition schemes have been considered in [GZE09], and [GF10]. A different class of methods for distributing calculations (as in the Jacobi approach), that combines projections with Tikhonov or proximal regularizations, has been considered in [KSH11].

REFERENCES

- [AW09] Kurt Anstreicher and Laurence Wolsey, *Two “well-known” properties of subgradient optimization*, *Mathematical Programming* **120** (2009), 213–220, 10.1007/s10107-007-0148-y.
- [BDLM08] A. Borghetti, C. D’Ambrosio, A. Lodi, and S. Martello, *An MILP Approach for Short-Term Hydro Scheduling and Unit Commitment With Head-Dependent Reservoir*, *Power Systems, IEEE Transactions on* **23** (2008), no. 3, 1115–1124.
- [BDMS03] A. Belloni, A. Diniz, M.E. Maceira, and C. Sagastizábal, *Bundle relaxation and primal recovery in unit commitment problems. the brazilian case*, *Annals of Operations Research* **120** (2003), 21–44, 10.1023/A:1023314026477.
- [BGLS06] J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, and C. Sagastizábal, *Numerical optimization. theoretical and practical aspects*, Universitext, Springer-Verlag, Berlin, 2006, Second edition, xiv+490 pp.
- [BLRS01] L. Bacaud, C. Lemaréchal, A. Renaud, and C. Sagastizábal, *Bundle methods in stochastic optimal power management: a disaggregated approach using preconditioners*, *Computational Optimization and Applications* **20** (2001), no. 3, 227–244.
- [BR92] J. Batut and A. Renaud, *Daily generation scheduling with transmission constraints: a new class of algorithms*, *IEEE Transactions on Power Systems* **7** (1992), 982–989.
- [BS88] D. Bienstock and J. Shapiro, *Optimizing Resource Acquisition Decisions by Stochastic Programming*, *Manag. Science* **34** (1988), no. 7, 215–229.
- [BS11] S. Bruno and C. Sagastizábal, *Optimization of real asset portfolio using a coherent risk measure: Application to oil and energy industries*, *Optimization and Engineering* **12** (2011), 257–275, 10.1007/s11081-010-9127-x.
- [CF10] William Chung and J. David Fuller, *Subproblem Approximation in Dantzig-Wolfe Decomposition of Variational Inequality Models with an Application to a Multicommodity Economic Equilibrium Model*, *Operations Research* **58** (2010), 1318–1327.
- [CR99] A. J. Conejo and N. Jiménez Redondo, *Short-term hydro-thermal coordination by Lagrangian relaxation: solution of the dual problem*, *IEEE Transactions on Power Systems* **14** (1999), no. 1, 89–95.
- [CS97] Claus C. Caroe and Rudiger Schultz, *Dual decomposition in stochastic integer programming*, *Operations Research Letters* **24** (1997), 37–45.
- [DGL05] L. Dubost, R. Gonzalez, and C. Lemaréchal, *A primal-proximal heuristic applied to the French unit-commitment problem*, *Mathematical Programming* **104** (2005), 129–151, 10.1007/s10107-005-0593-4.
- [DSM07] A. Diniz, C. Sagastizábal, and M.E. Maceira, *Assessment of Lagrangian Relaxation with Variable Splitting for Hydrothermal Scheduling*, *Power Engineering Society General Meeting, 2007. IEEE*, june 2007, pp. 1–8.
- [ES10] G. Emiel and C. Sagastizábal, *Incremental-like bundle methods with application to energy planning*, *Computational Optimization and Applications* **46** (2010), 305–332.

- [FC05] J. David Fuller and William Chung, *Dantzig-Wolfe Decomposition of Variational Inequalities*, *Comput. Econ.* **25** (2005), 303–326.
- [FdS06] E.C. Finardi and E.L. da Silva, *Solving the hydro unit commitment problem via dual decomposition and sequential quadratic programming*, *Power Systems, IEEE Transactions on* **21** (2006), no. 2, 835 – 844.
- [FFP07] Francisco Facchinei, Andreas Fischer, and Veronica Piccialli, *On generalized Nash games and variational inequalities*, *Operations Research Letters* **35** (2007), no. 2, 159–164.
- [FG06] Antonio Frangioni and Claudio Gentile, *Solving nonlinear single-unit commitment problems with ramping constraints*, *Operations Research* **54** (2006), no. 4, 767 – 775.
- [FK00] Stefan Feltenmark and Krzysztof C. Kiwiel, *Dual Applications of Proximal Bundle Methods, Including Lagrangian Relaxation of Nonconvex Problems*, *SIAM Journal on Optimization* (2000), 697–721.
- [FK10] Francisco Facchinei and Christian Kanzow, *Generalized Nash Equilibrium Problems*, *Annals OR* **175** (2010), no. 1, 177–211.
- [FM99] Michael C. Ferris and Todd S. Munson, *Interfaces to PATH 3.0: Design, Implementation and Usage*, *Computational Optimization and Applications* **12** (1999), 207–227, 10.1023/A:1008636318275.
- [FMR11] Csaba Fábíán, Gautam Mitra, and Diana Roman, *Processing second-order stochastic dominance models using cutting-plane representations*, *Mathematical Programming* **130** (2011), 33–57, 10.1007/s10107-009-0326-1.
- [FOF08] E.B. Fisher, R.P. O’Neill, and M.C. Ferris, *Optimal transmission switching*, *Power Systems, IEEE Transactions on* **23** (2008), no. 3, 1346 –1355.
- [FS07] Csaba I. Fábíán and Zoltán Szóke, *Solving two-stage stochastic programming problems with level decomposition*, *Comput. Manag. Sci.* **4** (2007), no. 4, 313–353. MR 2351219 (2008f:90063)
- [GF10] S. Gabriel and J. Fuller, *A benders decomposition method for solving stochastic complementarity problems with an application in energy*, *Computational Economics* **35** (2010), no. 4, 301–329.
- [GK87] M. Guignard and S. Kim, *Lagrangian decomposition: a model yielding stronger Lagrangian bounds*, *Mathematical Programming* **39** (1987), no. 2, 215–228.
- [GNRS00] Ralf Gollmer, Matthias Nowak, Werner Romisch, and Rudiger Schultz, *Unit commitment in power generation a basic model and some extensions*, *Annals of Operations Research* **96** (2000), 167–189, 10.1023/A:1018947401538.
- [GS12a] V. Guigues and C. Sagastizábal, *Risk averse feasible policies for large-scale multistage stochastic linear programs*, *Math. Program.* (2012), Accepted for publication.
- [GS12b] ———, *The value of rolling horizon policies for risk-averse hydro-thermal planning*, *European Journal of Operations Research* **217** (2012), 219–240.
- [Gui03] M. Guignard, *Lagrangian decomposition*, *TOP* **11** (2003), no. 2, 151–228.
- [GZE09] Steven A. Gabriel, Jifang Zhuang, and Ruud Egging, *Solving stochastic complementarity problems in energy market modeling using scenario reduction*, *European Journal of Operational Research* **197** (2009), no. 3, 1028 – 1040.
- [HR11] Holger Heitsch and Werner Römisch, *Stability and scenario trees for multistage stochastic programs*, *Stochastic Programming (Gerd Infanger, ed.)*, *International Series in Operations Research & Management Science*, vol. 150, Springer New York, 2011, 10.1007/978-1-4419-1642-6_7, pp. 139–164.
- [HS94] Julia L. Higle and Suvrajeet Sen, *Finite master programs in regularized stochastic decomposition*, *Math. Programming* **67** (1994), no. 2, Ser. A, 143–168.
- [HT10] Michael Hinze and Fredi Tröltzsch, *Discrete concepts versus error analysis in pde-constrained optimization*, *GAMM-Mitteilungen* **33** (2010), no. 2, 148–162.
- [HUL93] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms*, *Grund. der math. Wiss.*, no. 305-306, Springer-Verlag, 1993, (two volumes).
- [Kiw06] K.C. Kiwiel, *A proximal bundle method with approximate subgradient linearizations*, *SIAM J. Optim.* **16** (2006), 1007–1023.
- [KSH11] A. Kannan, U.V. Shanbhag, and Harrison H.M., *Strategic behavior in power markets under uncertainty*, *Energy Systems* **2** (2011), no. 2, 115–141.

- [KSHdM02] Anton J. Kleywegt, Alexander Shapiro, and Tito Homem-de Mello, *The sample average approximation method for stochastic discrete optimization*, SIAM J. Optim. **12** (2001/02), no. 2, 479–502. MR 1885572 (2002j:90048)
- [KSK10] A. Khodaei, M. Shahidehpour, and S. Kamalinia, *Transmission switching in expansion planning*, Power Systems, IEEE Transactions on **25** (2010), no. 3, 1722–1733.
- [Lem01] C. Lemaréchal, *Lagrangian relaxation*, Computational Combinatorial Optimization (M. Jünger and D. Naddef, eds.), Springer Verlag, 2001, pp. 112–156.
- [LPRS96] C. Lemaréchal, F. Pellegrino, A. Renaud, and C. Sagastizábal, *Bundle methods applied to the unit-commitment problem*, System Modelling and Optimization (J. Doležal and J. Fidler, eds.), Chapman and Hall, 1996, pp. 395–402.
- [LR01] C. Lemaréchal and A. Renaud, *A geometric study of duality gaps, with applications*, Mathematical Programming **90** (2001), 399–427, 10.1007/PL00011429.
- [LS97] C. Lemaréchal and C. Sagastizábal, *Variable metric bundle methods: from conceptual to implementable forms*, Math. Program., Ser. A **76** (1997), 393–410.
- [LSS11] J. P. Luna, C. Sagastizábal, and M.V. Solodov, *A Class of Dantzig-Wolfe Type Decomposition Methods for Variational Inequality Problems*, Tech. report, Optimization Online, 2011.
- [MDS08] M. G. Martinez, A. Diniz, and C. Sagastizábal, *A comparative study of two forward dynamic programming techniques for solving local thermal unit commitment problem*, Proceedings of the 16th Power Systems Computation Conference, 2008, pp. 1–7.
- [MPS10] L. Moulin, M. Poss, and C. Sagastizábal, *Transmission expansion planning with re-design*, Energy Systems **1** (2010), 113–139.
- [MS07] R. Marcato and C. Sagastizábal, *Introducing Environmental Constraints in Generation Expansion Problems*, Numer. Linear Algebra Appl. **14** (2007), 351–368.
- [NR00] Matthias P. Nowak and Werner Romisch, *Stochastic Lagrangian Relaxation Applied to Power Scheduling in a Hydro-Thermal System under Uncertainty*, Annals of Operations Research **100** (2000), 251–272, 10.1023/A:1019248506301.
- [NR02] Robert Nurnberg and Werner Romisch, *A two-stage planning model for power scheduling in a hydro-thermal system under uncertainty*, Optimization and Engineering **3** (2002), 355–378, 10.1023/A:1021531823935.
- [NSW05] Matthias P. Nowak, Rudiger Schultz, and Markus Westphalen, *A stochastic integer programming model for incorporating day-ahead trading of electricity into hydro-thermal unit commitment*, Optimization and Engineering **6** (2005), 163–176, 10.1007/s11081-005-6794-0.
- [OS12] W. L. Oliveira and C. Sagastizábal, *Level Bundle Methods for Oracles with On-Demand Accuracy*, Submitted. Available at http://www.optimization-online.org/DB_HTML/2012/03/3390.html, 2012.
- [OSP⁺10] W. L. Oliveira, C. Sagastizábal, D.D.J. Penna, M.E.P. Maceira, and J.M. Damazio, *Optimal scenario tree reduction for stochastic streamflows in power generation planning problems*, Optimization Methods and Software **25** (2010), 917–936.
- [OSS11] W. L. Oliveira, C. Sagastizábal, and S. Scheimberg, *Inexact Bundle Methods for Two-Stage Stochastic Programming*, SIAM Journal on Optimization **21** (2011), 517–544.
- [PdM10] A. Philpott and V. de Matos, *Dynamic sampling algorithms for multi-stage stochastic programs with risk aversion*, Tech. report, Optimization Online, 2010.
- [PP91] M. Pereira and L. Pinto, *Multi-stage stochastic optimization applied to energy planning*, Math. Program. **52** (1991), no. 2, 359–375.
- [PS06] Andy Philpott and Rudiger Schultz, *Unit commitment in electricity pool markets*, Mathematical Programming **108** (2006), 313–337, 10.1007/s10107-006-0713-9.
- [RU02] R.T. Rockafellar and S. Uryasev, *Conditional Value-at-Risk for General Loss Distributions*, Journal of Banking and Finance **26** (2002), no. 7, 1443–1471.
- [Sag12] C. Sagastizábal, *Divide to conquer: decomposition methods for energy optimization*, Math. Program. B (2012), Accepted for publication. Special issue 21st ISMP, Berlin, 10.1007/s10107-012-0570-7.
- [SDR09] Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński, *Lectures on stochastic programming: Modeling and theory*, MPS-SIAM Series on Optimization, 2009.
- [Sha11] Alexander Shapiro, *Analysis of stochastic dual dynamic programming method*, European Journal of Operational Research **209** (2011), no. 1, 63–72.

- [ST11] Alexander Shapiro and Wajdi Tekaya, *Report for technical cooperation between Georgia Institute of Technology and ONS - Operador Nacional do Sistema Eléctrico*, Report 2: Risk Averse Approach, 2011.
- [SYG06] Suvrajeet Sen, Lihua Yu, and Talat Genc, *A stochastic programming approach to power portfolio optimization*, Operations Research **54** (2006), no. 1, 55–72.
- [TBL96] S. Takriti, J.R. Birge, and E. Long, *A stochastic model for the unit commitment problem*, IEEE Transactions on Power Systems **11** (1996), 1497–1508.
- [VW69] R. Van Slyke and R.J-B. Wets, *L-shaped linear programs with applications to control and stochastic programming*, SIAM Journal on Applied Mathematics **17** (1969), 638–663.
- [Yak94] D.S. Yakowitz, *A regularized stochastic decomposition algorithm for two-stage stochastic linear programs*, Computational Optimization and Applications **3** (1994), no. 1, 59–81.
- [ZPR00] Golbon Zakeri, Andrew B. Philpott, and David M. Ryan, *Inexact cuts in Benders decomposition*, SIAM J. Optim. **10** (2000), no. 3, 643–657 (electronic).

APPENDIX A. CONVERGENCE ANALYSIS OF INEXACT BUNDLE METHODS

We start by listing some consequences of various definitions in Section 4, obtained after some simple algebra. First, note that (4.4) written with $y = \hat{x}^k$, together with (4.3), the definition of \mathbf{E}^k in (4.6), and (4.9), yield the relation

$$(A.1) \quad \mathbf{E}^k = \delta^k - \frac{2 - \alpha_k}{2} \mu_k |x^{k+1} - \hat{x}^k|^2 \quad \text{and, therefore,} \quad \mathbf{E}^k \leq \delta^k \text{ at all iterations.}$$

Second, by adding δ^k to both members of the identity, we see that

$$(A.2) \quad \text{inequality (4.10) holds} \iff \delta^k + \mathbf{E}^k < -\frac{\alpha_k + 2\beta_k}{2} \mu_k |x^{k+1} - \hat{x}^k|^2$$

$$(A.3) \quad \iff \mathbf{E}^k < -\frac{1 + \beta_k}{2} \mu_k |x^{k+1} - \hat{x}^k|^2$$

where the last equivalence follows from (4.10) and the left hand side identity in (A.1).

Finally,

$$(A.4) \quad \text{if (4.10) does not hold, then } \delta^k \geq \max \left\{ \begin{array}{l} \frac{1}{2} (1 - (\alpha_k + \beta_k)) \mu_k |x^{k+1} - \hat{x}^k|^2 \\ -\mathbf{E}^k + \frac{\alpha_k + 2\beta_k}{2} \mu_k |x^{k+1} - \hat{x}^k|^2 \\ \mathbf{E}^k \end{array} \right\}.$$

The algorithm in Section 4.3 has individual counters for iterations declared serious, or of noise attenuation, denoted respectively by \mathbf{SA}^k and \mathbf{NA}^k . Suppose the algorithm loops forever, so that $k \rightarrow \infty$. Then

- (1) there are infinitely many serious iterates ($\mathbf{SA}^k \rightarrow \infty$),
- (2) the center is unchanged after a finite number of iterations ($\mathbf{SA}^k = \widehat{\mathbf{SA}}$ for all k large), and there are an infinite number of noise attenuation steps ($\mathbf{NA}^k \rightarrow \infty$),
or
- (3) the center is unchanged after a finite number of iterations ($\mathbf{SA}^k = \widehat{\mathbf{SA}}$ for all k large), there are a finite number of noise attenuation steps ($\mathbf{NA}^k = \widehat{\mathbf{NA}}$ for all k large).

We consider convergence in each of these cases now.

A.1. Infinite Serious Steps.

Theorem A.1 (Infinitely many serious iterates). *Suppose $\mathbf{SA}^k \rightarrow \infty$ and let K_s denote the indices of iterations that give a serious step (i.e., $\mathbf{SA}^k = k$). Then either $\inf_{x \in \mathbb{R}^n} f(x) = -\infty$, or $\delta^k \rightarrow 0$ as $K_s \ni k \rightarrow \infty$. Suppose, in addition, that (4.11),*

(4.12a) hold. Then, when $\delta^k \rightarrow 0$ as $K_s \ni k \rightarrow \infty$, $\liminf_{K_s \ni k \rightarrow \infty} V_k = 0$. Moreover, if the sequence $\{\hat{x}^k\}_{K_s}$ is bounded, then

$$\lim_{K_s \ni k \rightarrow \infty} V_k = 0.$$

Proof. At each serious step $k \in K_s$ we have

$$(A.5) \quad \hat{f}^{k+1} \leq \hat{f}^k - m\delta^k \quad \text{and} \quad \delta^k \geq 0,$$

where the second inequality follows from (4.11), recalling that a serious step can only take place when no noise attenuation occurs. Thus, the non increasing sequence $\{\hat{f}^k\}_{K_s}$ is either unbounded below or converges.

In the first case, by equation (4.1), the sequence $\{f(\hat{x}^k) - \hat{\varepsilon}_f^k\}_{K_s}$ is unbounded below and since $\{\hat{\varepsilon}_f^k\}$ is bounded, it follows that $\{f(\hat{x}^k)\} \rightarrow -\infty$ as $K_s \ni k \rightarrow \infty$.

In the case of convergent $\{\hat{f}^k\}_{K_s}$, the inequality (A.5) implies that $\delta^k \rightarrow 0$ as $K_s \ni k \rightarrow \infty$. Since there is no noise attenuation, test (4.10) fails and (4.3) yields that

$$\delta^k \geq \frac{1}{2} \left(1 - (\alpha_k + \beta_k)\right) \mu_k |x^{k+1} - \hat{x}^k|^2 = \frac{1}{2\mu_k} \left(1 - (\alpha_k + \beta_k)\right) |G^k|^2.$$

Together with assumptions (4.11) and (4.12a), we see that

$$(A.6) \quad \text{both } \frac{1}{2\mu_k} |G^k|^2 \text{ and } G^k \rightarrow 0 \quad \text{as } K_s \ni k \rightarrow \infty.$$

Moreover, by (A.1), $\delta^k \geq \mathbf{E}^k$, while (A.2) implies that $\delta^k \geq -\mathbf{E}^k - \frac{\alpha_k + 2\beta_k}{2} \mu_k |x^{k+1} - \hat{x}^k|^2 = -\mathbf{E}^k - \frac{\alpha_k + 2\beta_k}{2\mu_k} |G^k|^2$. It follows that $\lim \delta^k \geq \lim |\mathbf{E}^k|$ and, hence,

$$(A.7) \quad \mathbf{E}^k \rightarrow 0 \quad \text{as } K_s \ni k \rightarrow \infty.$$

If $\{\hat{x}^k\}_{K_s}$ is bounded, the limits in (A.6) and (A.7) and the definition of V_k in (4.6) show that $\lim_{K_s \ni k \rightarrow \infty} V_k = 0$, as stated.

If $\{\hat{x}^k\}_{K_s}$ is unbounded, both (A.6) and (A.7) hold and to show that $\liminf_{K_s \ni k \rightarrow \infty} V_k = 0$ the relation

$$(A.8) \quad \liminf_{K_s \ni k \rightarrow \infty} \langle G^k, \hat{x}^k \rangle \leq 0$$

should hold. To show this inequality, we start by taking squares in (4.3) to obtain $|\hat{x}^{k+1}|^2 - |\hat{x}^k|^2 = \frac{1}{2} |G^k|^2 - 2 \frac{1}{\mu_k} \langle G^k, \hat{x}^k \rangle$. Summing over $K_s \ni i \leq k$, we have that

$$(A.9) \quad \begin{aligned} |\hat{x}^{k+1}|^2 - |\hat{x}^1|^2 &= \sum_{K_s \ni i=1}^k |\hat{x}^{i+1}|^2 - |\hat{x}^i|^2 \\ &= \sum_{K_s \ni i=1}^k \frac{1}{\mu_i} \left(\frac{1}{\mu_i} |G^i|^2 - 2 \langle G^i, \hat{x}^i \rangle \right). \end{aligned}$$

If there were to exist some $\gamma < 0$ such that $\frac{1}{\mu_i} |G^i|^2 - 2 \langle G^i, \hat{x}^i \rangle \leq \gamma$ for all $i \in K_s$, then (4.12a) would imply that the right-hand side of the equality (A.9) tends to $-\infty$ as $k \rightarrow \infty$, while the left-hand side is bounded below by $-|\hat{x}^1|^2$. This contradiction implies that

$$\limsup_{K_s \ni k \rightarrow \infty} \left(\frac{1}{\mu_k} |G^k|^2 - 2 \langle G^k, \hat{x}^k \rangle \right) \geq 0.$$

Since the first term tends to zero, by (A.6), the claim in (A.8) holds. \square

In the next two cases, a finite number of serious steps occurs. That is, after a finite number of iterations the algorithmic center remains unchanged, so that there exists \hat{k} and \hat{x} such that $\hat{x}^k = \hat{x}$ for all $k > \hat{k}$. We require the additional assumptions on the model \mathbb{M}_k :

$$(A.10) \quad \begin{aligned} & \text{for all non-serious iterations } k, \\ & \mathbb{M}^k(\hat{x}) \leq f(\hat{x}) + \hat{\varepsilon}_g \text{ for a constant } \hat{\varepsilon}_g \geq 0 \end{aligned}$$

while, for all null iterations k ,

$$(A.11) \quad \mathbb{M}^{k+1}(y) \geq \mathbb{M}^k(y) \text{ and}$$

$$(A.12) \quad \mathbb{M}^{k+1}(y) \geq f^{k+1} + \langle g^{k+1}, y - x^{k+1} \rangle$$

with the sequence $\{g^k\}$ bounded if $\{x^k\}$ is bounded.

A straightforward consequence of (A.10) is that, by (4.1), the error defined in (4.6) is bounded below:

$$(A.13) \quad \text{if (A.10) holds, then } \mathbf{E}^k \geq -\bar{\varepsilon}_f - \hat{\varepsilon}_g.$$

A.2. Finite Serious Steps with Infinite Noise Attenuation. We now consider the case of a finite number of serious steps with an infinite sequence of noise attenuation steps.

Theorem A.2 (Finitely many serious iterates with infinite noise attenuation steps). *Suppose that $\mathbf{SA}^k = \widehat{\mathbf{SA}}$ for all k sufficiently large, and suppose that $\mathbf{NA}^k \rightarrow \infty$. Let K_t denote those iterations when inaccuracy is detected. If (A.10) holds, then $\lim_{K_t \ni k \rightarrow \infty} V_k = 0$.*

Proof. First note that if $k \in K_t$, then by equation (A.3), $\mathbf{E}^k < 0$. Moreover, combining (A.3) and (4.3), we see that

$$|G^k|^2 = \frac{1}{\mu_k} |x^{k+1} - \hat{x}|^2 < -\frac{2\mu_k \mathbf{E}^k}{(1 + \beta_k)}.$$

As a result, by (4.6) and (A.13),

$$0 \leq V^k \leq \sqrt{-2\mu_k \mathbf{E}^k / (1 + \beta_k)} (1 + |\hat{x}|) \leq \sqrt{2\mu_k (\bar{\varepsilon}_f + \hat{\varepsilon}_g) / (1 + \beta_k)} (1 + |\hat{x}|).$$

After all serious steps have been done, $\mathbf{NA}^k \rightarrow \infty$ and, since the sequence $\{\mu_k\}_{k \in K_t}$ is decreasing, $\mu_k \rightarrow 0$ and $V^k \rightarrow 0$ as $K_t \ni k \rightarrow \infty$. \square

A.3. Finite Serious Steps with Finite Noise Attenuation. Before stating the next theorem, recall that \mathbf{NA}^k is non-decreasing during null steps. Therefore, if a finite number of serious steps occurs, then \mathbf{NA}^k is non-decreasing after a finite number of iterations. In this case, if \mathbf{NA}^k is bounded above, then \mathbf{NA}^k must reach its finite limit after a finite number of iterations. After this point, no noise attenuation occurs ((4.10) does not hold) and \mathbf{NA}^k will remain unchanged in all future iterations. In order to prove convergence in this case we use the following lemma.

At null steps, the prox-parameter is allowed to increase only if there was no noise attenuation step after generating the current algorithmic center \hat{x} . Otherwise, the prox-parameter remains unchanged.

Lemma A.3 (Convergence of the value function). *Suppose that a finite number of serious iterates and noise attenuation steps occur, so that for all $k \geq \hat{k}$ we have $\mathbf{SA}^k = \widehat{\mathbf{SA}}$, $\mathbf{NA}^k = \widehat{\mathbf{NA}}$, and $\delta^k \geq \frac{1}{2} \left(1 - (\alpha_k + \beta_k)\right) \mu_k |x^{k+1} - \hat{x}|^2$. Let*

$$\mathbf{OV}^k := \mathbb{M}^k(x^{k+1}) + \frac{1}{2} \mu_k |x^{k+1} - \hat{x}|^2$$

be the optimal value of the problem solved to define x^{k+1} . Suppose the model functions satisfy (A.10) and (A.11). Then the values \mathbf{OV}^k are increasing and bounded above and, therefore, the sequence $\{\mathbf{OV}^k\}$ is convergent.

Proof. First note that, as no noise attenuation occurs past \hat{k} , we have that $\{\mu_k\}$ is a non-increasing sequence for all $k \geq \hat{k}$.

Consider $k \geq \max\{\hat{k}, \bar{k}\}$. Keeping in mind (A.11) and that μ_k is non-increasing, we have that

$$\begin{aligned} \mathbf{OV}^{k+1} &= \mathbb{M}^{k+1}(x^{k+2}) + \frac{1}{2} \mu_{k+1} |x^{k+2} - \hat{x}|^2 \\ &= \min_y \left\{ \mathbb{M}^{k+1}(y) + \frac{1}{2} \mu_{k+1} |y - \hat{x}|^2 \right\} \\ &\geq \min_y \left\{ \mathbf{M}^k(y) + \frac{1}{2} \mu_{k+1} |y - \hat{x}|^2 \right\} \\ &\geq \min_y \left\{ \mathbf{M}^k(y) + \frac{1}{2} \mu_k |y - \hat{x}|^2 \right\} \\ &= \mathbf{OV}^k. \end{aligned}$$

Therefore $\{\mathbf{OV}^k\}$ is non-decreasing, as stated. To see that the sequence $\{\mathbf{OV}^k\}$ is bounded above, note that

$$\mathbf{OV}^k \leq \mathbb{M}^k(\hat{x}) + \frac{1}{2} \mu_k |\hat{x} - \hat{x}|^2 = \hat{f} - \mathbf{E}^k \leq f(\hat{x}) + \bar{\varepsilon}_f + \hat{\varepsilon}_g,$$

by (A.13). It then follows that $\{\mathbf{OV}^k\}$ converges. \square

We are now in position to prove our last convergence statement.

Theorem A.4 (Finitely many serious iterates and noise attenuation steps with infinitely many null steps).

Suppose that a finite number of serious iterates and noise attenuation steps occur, so that for all $k \geq \hat{k}$ we have $\mathbf{SA}^k = \widehat{\mathbf{SA}}$, $\mathbf{NA}^k = \widehat{\mathbf{NA}}$, and $\delta^k + \mathbf{E}^k \geq 0$. Let \hat{x} be the point found by the last serious step (i.e., $\hat{x}^k = \hat{x}$ for all $k \geq \hat{k}$).

If (A.10) and (A.11) hold, the sequence $\{x^k\}$ is bounded.

If, in addition, (4.11), (A.12), and (4.12c) hold, then $\delta^k \rightarrow 0$. As a result,

$$\lim_{k \rightarrow \infty} x^k = \hat{x}, \quad \lim_{k \rightarrow \infty} \mathbf{M}^k(x^{k+1}) = \hat{f}, \quad \text{and} \quad \lim_{k \rightarrow \infty} V_k = 0.$$

Proof. We again assume that k is large enough, so that $\hat{x}^k = \hat{x}$ is fixed, no noise attenuation occurs. Thus $\{\mu_k\}$ is a non-decreasing sequence bounded below by certain positive minimal value, μ_{\min} . Let

$$\mathbf{OV}^k := \mathbb{M}^k(x^{k+1}) + \frac{1}{2} \mu_k |x^{k+1} - \hat{x}|^2$$

be the optimal value of the problem defining the next iterate. By (4.3), for all y it holds that

$$2\langle G^k, y - x^{k+1} \rangle = 2\mu_k \langle \hat{x} - x^{k+1}, y - x^{k+1} \rangle = \mu_k |x^{k+1} - \hat{x}|^2 + \mu_k |y - x^{k+1}|^2 - \mu_k |y - \hat{x}|^2.$$

Since \mathbf{M}^k is affine and G^k is its gradient, this implies that

$$\begin{aligned} \mathbf{M}^k(y) &= \mathbf{M}^k(x^{k+1}) + \langle G^k, y - x^{k+1} \rangle \\ &= \mathbf{M}^k(x^{k+1}) + \frac{1}{2} \mu_k |x^{k+1} - \hat{x}|^2 + \frac{1}{2} \mu_k |y - x^{k+1}|^2 - \frac{1}{2} \mu_k |y - \hat{x}|^2 \\ &= \mathbf{OV}^k + \frac{1}{2} \mu_k |y - x^{k+1}|^2 - \frac{1}{2} \mu_k |y - \hat{x}|^2. \end{aligned}$$

Therefore

$$(A.14) \quad \mathbb{M}^k(y) + \frac{1}{2}\mu_k|y - \hat{x}|^2 = \text{OV}^k + \frac{1}{2}\mu_k|y - x^{k+1}|^2.$$

Evaluating at $y = \hat{x}$, we see that $\mathbb{M}^k(\hat{x}) = \text{OV}^k + \frac{1}{2}\mu_k|\hat{x} - x^{k+1}|^2$. Assumptions (A.10) and (4.5) imply that $\text{OV}^k + \frac{1}{2}\mu_k|\hat{x} - x^{k+1}|^2 \leq f(\hat{x}) + \hat{\varepsilon}_g$. The optimal values converge, by Lemma A.3, so the second term above is bounded. Moreover, since $\mu_k \geq \mu_{\min}$, the sequence $\{x^{k+1}\}$ is bounded, as stated.

Using assumption (A.11) in (A.14), we see that $\mathbb{M}^{k+1}(y) + \frac{1}{2}\mu_k|y - \hat{x}|^2 \geq \text{OV}^k + \frac{1}{2}\mu_k|y - x^{k+1}|^2$, and evaluating at $y = x^{k+2}$, this implies $\text{OV}^{k+1} \geq \text{OV}^k + \frac{1}{2}\mu_k|x^{k+2} - x^{k+1}|^2$. Since $\{\text{OV}^k\}$ converges and $\{\mu_k\}$ is non-decreasing, this demonstrates that $|x^{k+2} - x^{k+1}|^2 \rightarrow 0$.

The assumption (4.11) implies that $\delta_k \geq 0$ when (4.10) fails. The descent test also fails: $f^{k+1} > \hat{f}^k - m\delta^k$. Adding δ^k to both terms and using the definition in (4.9) we obtain that

$$\begin{aligned} 0 \leq (1 - m)\delta^k &= \delta^k + f^{k+1} - \hat{f}^k \\ &= f^{k+1} - \mathbb{M}^k(x^{k+1}) - \frac{1}{2}\alpha_k\mu_k|x^{k+1} - \hat{x}|^2 \\ &< f^{k+1} - \mathbb{M}^k(x^{k+1}) \\ &\leq -\langle g^{k+1}, x^{k+2} - x^{k+1} \rangle + \mathbb{M}^{k+1}(x^{k+2}) - \mathbb{M}^k(x^{k+1}), \end{aligned}$$

where the last inequality follows from adding $\pm\mathbb{M}^{k+1}(x^{k+2})$ and applying assumption (A.12). Since the sequence $\{x^k\}$ is bounded, so is $\{g^k\}$, say by a constant \bar{g} . Continuing and using that $\mu_{\max} \geq \mu_{k+1} \geq \mu_k$, we have that

$$\begin{aligned} 0 \leq (1 - m)\delta^k &\leq \bar{g}|x^{k+2} - x^{k+1}| + \mathbb{M}^{k+1}(x^{k+2}) - \mathbb{M}^k(x^{k+1}) \\ &= \bar{g}|x^{k+2} - x^{k+1}| + \text{OV}^{k+1} - \frac{1}{2}\mu_{k+1}|x^{k+2} - \hat{x}|^2 - \text{OV}^k + \frac{1}{2}\mu_k|x^{k+1} - \hat{x}|^2 \\ &\leq \bar{g}|x^{k+2} - x^{k+1}| + \text{OV}^{k+1} - \text{OV}^k + \frac{1}{2}\mu_{k+1}(|x^{k+1} - \hat{x}|^2 - |x^{k+2} - \hat{x}|^2) \\ &= \bar{g}|x^{k+2} - x^{k+1}| + \text{OV}^{k+1} - \text{OV}^k + \\ &\quad \frac{1}{2}\mu_{k+1}(2\langle x^{k+2} - x^{k+1}, \hat{x} - x^{k+1} \rangle - |x^{k+2} - x^{k+1}|^2) \\ &\leq \bar{g}|x^{k+2} - x^{k+1}| + \text{OV}^{k+1} - \text{OV}^k + \frac{1}{2}\mu_{\max}|x^{k+2} - x^{k+1}||\hat{x} - x^{k+1}|. \end{aligned}$$

Since the sequence $\{x^k\}$ is bounded, $|x^{k+2} - x^{k+1}| \rightarrow 0$, and $\{\text{OV}^k\}$ converges, by Lemma A.3. So the right hand side of the above tends to zero and, therefore, $\delta^k \rightarrow 0$.

By (A.4) and (4.11), $\frac{1}{2\mu_k}|G^k|^2 \rightarrow 0$, and (4.12c) implies that $|G^k| \rightarrow 0$. Therefore, by definition (4.3), $x^k \rightarrow \hat{x}$. Using once more (A.4), and the fact that $\frac{1}{2\mu_k}|G^k|^2 \rightarrow 0$ provides $\lim \delta^k \geq \lim |\mathbf{E}^k|$, so $\lim |\mathbf{E}^k| = 0$ and the results follow from the definitions in (4.6). \square

The interest of the results above is that neither convexity of the function f nor subgradient information of f is used along the proofs. Of course, if gradient information is available, even if inaccurate, it should be used to define the models. As long as the model functions satisfy (A.10) *at non serious iterations*, and both (A.11) and (A.12) *at null iterations*, the optimality measure V_k will tend to zero for some infinite subsequence. At serious iterations, one should just ensure boundedness of the generated sequence: *it does not matter what the model is*.

To give a meaning to $V_k \rightarrow 0$ in terms of the problem we want to solve, the model should satisfy (4.13). In this case, the optimality certificate (4.14) will ensure optimality of \hat{x} (or of a limit point of the serious sequence). Clearly, if the function is convex, building a model satisfying (4.13) is possible. In particular, as shown in Remark 4.1, in the setting in Section 3 it is possible to defined an inexact oracle such that (4.13) holds

with $\bar{\varepsilon} = 0$. Since (4.13) implies (A.10), for the decomposition setting, only (A.11) and (A.12) need to be checked. These inequalities are satisfied by applying standard rules for *bundle compression*, described in [BGLS06, Ch. 10].

ON LEAVE FROM INRIA, FRANCE, VISITING RESEARCHER AT IMPA, ESTRADA DONA CASTORINA 110, JARDIM BOTÂNICO, RIO DE JANEIRO RJ 22460-320, BRAZIL

E-mail address: `sagastiz@impa.br`

INTRODUCCIÓN A LOS RETICULADOS

MIGUEL CAMPERCHOLI

ÍNDICE

| | |
|--|----|
| Introducción | 53 |
| 1. Conjuntos parcialmente ordenados | 53 |
| 1.1. Diagramas de Hasse | 54 |
| 1.2. Elementos maximales, máximos y supremos | 54 |
| 1.3. Homomorfismos de posets | 54 |
| 2. Reticulados | 55 |
| 2.1. Homomorfismos de reticulados | 57 |
| 3. Reticulados distributivos | 57 |
| 4. Las leyes de \cap y \cup | 60 |
| Referencias | 62 |

INTRODUCCIÓN

Un *reticulado* es un poset en el cual todo subconjunto finito tiene supremo e ínfimo. Estas estructuras surgen naturalmente en diversos campos de la matemática, y han sido extensamente estudiadas a partir de principios del siglo XX. Si bien la definición evidentemente enmarca a los reticulados en la teoría del orden, existe una definición equivalente que es puramente algebraica. Esta naturaleza dual hace que su teoría sea muy rica, pudiendo estudiarlos mediante técnicas algebraicas y geométricas.

El lector interesado en explorar esta apasionante teoría más allá de la breve introducción en sus manos puede consultar, por ejemplo, los textos [1], [2] y [3].

1. CONJUNTOS PARCIALMENTE ORDENADOS

Definición 1.1. Sea P un conjunto no vacío. Un *orden parcial* de P es una relación binaria \leq que para todo $a, b, c \in P$ cumple:

1. $a \leq a$ (reflexiva)
2. $a \leq b$ y $b \leq c$ implican $a \leq c$ (transitiva)
3. $a \leq b$ y $b \leq a$ implican $a = b$ (antisimétrica).

Un *conjunto parcialmente ordenado* o *poset* es un par $\langle P, \leq \rangle$, donde P es un conjunto no vacío y \leq es un orden parcial.

Ejemplo 1.2.

1. Los órdenes usuales de \mathbb{N} , \mathbb{Z} , \mathbb{Q} y \mathbb{R} son órdenes parciales de sus respectivos conjuntos.
2. En \mathbb{N} la relación dada por $n \leq m$ si y solo si n divide a m es un orden parcial.
3. Sea X un conjunto cualquiera. La relación de inclusión entre subconjuntos de X es un orden parcial de $\mathcal{P}(X)$.

4. El orden alfabético es un orden parcial sobre el conjunto de las palabras en español.

Dado un orden parcial \leq de P escribiremos:

- $a < b$ si y solo si $a \leq b$ y $a \neq b$
- $a \prec b$ si y solo si $a < b$ y no hay c tal que $a < c < b$. En este caso diremos que b cubre a a .

1.1. Diagramas de Hasse. Un poset finito $\langle P, \leq \rangle$ se puede *dibujar* siguiendo las siguientes instrucciones:

1. asociar a cada $a \in P$ un punto p_a del plano
2. trazar un segmento uniendo los puntos p_a y p_b , cada vez que $a \prec b$
3. realizar lo indicado en los puntos anteriores de manera que:
 - a) si $a \prec b$ entonces p_a está por debajo de p_b
 - b) si un punto p_a ocurre en un segmento del diagrama entonces lo hace en alguno de sus extremos.

Un dibujo de $\langle P, \leq \rangle$ obtenido de esta forma es llamado un *diagrama de Hasse* de $\langle P, \leq \rangle$.

Notar que la relación de orden de un poset queda plasmada en su diagrama, ya que $a < b$ se da si y solo si hay una sucesión de segmentos ascendentes desde p_a hasta p_b .

1.2. Elementos maximales, máximos y supremos.

Definición 1.3. Sean $\langle P, \leq \rangle$ un poset, $S \subseteq P$ y $a \in P$. Diremos que:

1. a es *maximal* si no hay $b \in P$ tal que $a < b$
2. a es *máximo* de P si $b \leq a$, para todo $b \in P$
3. a es *cota superior* de S si $b \leq a$ para todo $b \in S$
4. a es *supremo* de S si a es la menor cota superior de S . Es decir, si
 - a) a es cota superior de S , y
 - b) $a \leq b$ para toda cota superior b de S .

En forma dual se definen los conceptos *minimal*, *mínimo*, *cota inferior* e *ínfimo*.

No todo poset $\langle P, \leq \rangle$ cuenta con elementos maximales o máximos. Además puede haber subconjuntos de P para los cuales no exista supremo o ínfimo. Por otra parte, todo $S \subseteq P$ puede tener a lo sumo un supremo y un ínfimo, que en caso de existir serán denotados por $\sup S$ e $\inf S$.

1.3. Homomorfismos de posets.

Definición 1.4. Sean $\langle P, \leq \rangle$ y $\langle P', \leq' \rangle$ posets. Una función $h : P \rightarrow P'$ es un *homomorfismo* de $\langle P, \leq \rangle$ en $\langle P', \leq' \rangle$ si para todo $x, y \in P$ se cumple que $x \leq y$ implica $h(x) \leq' h(y)$. Diremos que h es un *isomorfismo* $\langle P, \leq \rangle$ en $\langle P', \leq' \rangle$ si es biyectiva y h^{-1} es homomorfismo de $\langle P', \leq' \rangle$ en $\langle P, \leq \rangle$.

Un homomorfismo biyectivo no necesariamente es un isomorfismo.

Lema 1.5. Supongamos que $h : P \rightarrow P'$ es un isomorfismo de $\langle P, \leq \rangle$ en $\langle P', \leq' \rangle$. Sean $S \subseteq P$ y $a, b \in P$.

1. a es cota superior de S si y solo si $h(a)$ es cota superior de $h(S)$.
2. $\sup S$ existe si y solo si $\sup h(S)$ existe. En el caso de que existan vale que $h(\sup S) = \sup h(S)$.
3. a es maximal si y solo si $h(a)$ es maximal.

4. $a \prec b$ si y solo si $h(a) \prec' h(b)$.

Demostración. (1) Supongamos que a es cota superior de S y sea $y \in h(S)$. Sea $s \in S$ tal que $y = h(s)$. Como $s \leq a$, tenemos que $y = h(s) \leq' h(a)$, y hemos probado que $h(a)$ es cota superior de $h(S)$.

Supongamos ahora que $h(a)$ es cota superior de $h(S)$ y fijemos $s \in S$. Ya que $h(s) \leq h(a)$, tenemos que $s = h^{-1}(h(s)) \leq h^{-1}(h(a)) = a$.

(2) Probamos que si existe $\sup S$ entonces $h(\sup S)$ es el supremo de $h(S)$. Por (1) $h(\sup S)$ es cota superior de $h(S)$. Sea b una cota superior de $h(S)$. Como $h^{-1}(b)$ es cota superior de S se cumple que $\sup S \leq h^{-1}(b)$, y esto implica $h(\sup S) \leq b$. La otra dirección es igual de fácil.

Los puntos (3) y (4) son dejados al lector. \square

Notar que si dos posets finitos son isomorfos, entonces pueden representarse con el mismo diagrama de Hasse.

2. RETICULADOS

Definición 2.1. Un *reticulado* es un conjunto parcialmente ordenado $\langle L, \leq \rangle$ tal que para todo $a, b \in L$, existen $\sup\{a, b\}$ e $\inf\{a, b\}$.

En un reticulado tenemos dos operaciones binarias, \vee e \wedge , naturalmente definidas:

$$a \vee b = \sup\{a, b\}$$

$$a \wedge b = \inf\{a, b\}.$$

Recopilamos en el siguiente lema algunas propiedades características de éstas operaciones.

Lema 2.2. Sean $\langle L, \leq \rangle$ un reticulado y $a, b, c, d \in L$. Vale que:

1. $a \leq a \vee b$ y $a \wedge b \leq a$.
2. $a \vee a = a \wedge a = a$.
3. $a \vee b = b \vee a$ y $a \wedge b = b \wedge a$.
4. $a \leq b$ si y solo si $a \vee b = b$ si y solo si $a \wedge b = a$.
5. $a \vee (a \wedge b) = a$ y $a \wedge (a \vee b) = a$.
6. $(a \vee b) \vee c = a \vee (b \vee c)$ y $(a \wedge b) \wedge c = a \wedge (b \wedge c)$.
7. Si $a \leq c$ y $b \leq d$, entonces $a \vee b \leq c \vee d$ y $a \wedge b \leq c \wedge d$.
8. $(a \wedge b) \vee (a \wedge c) \leq a \wedge (b \vee c)$.

Demostración. Los puntos (1)–(4) son consecuencia directa de la definición.

(5) Como $(a \wedge b) \leq a$ por (4) tenemos $(a \wedge b) \vee a = a$; y por (3) $a \vee (a \wedge b) = a$.

(6) Observemos que $a \vee (b \vee c)$ es cota superior de $\{a, b, c\}$ pues:

$$a \leq a \vee (b \vee c)$$

$$b \leq (b \vee c) \leq a \vee (b \vee c)$$

$$c \leq (b \vee c) \leq a \vee (b \vee c).$$

Además, como $a \vee (b \vee c)$ es cota superior de $\{a, b\}$, tenemos que

$$a \vee b \leq a \vee (b \vee c),$$

por lo cual $a \vee (b \vee c)$ es cota superior del conjunto $\{a \vee b, c\}$, lo cual nos dice que

$$(a \vee b) \vee c \leq a \vee (b \vee c).$$

Análogamente se puede probar la otra desigualdad.

(7) y (8) quedan como ejercicios para el lector. \square

Lema 2.3. *Dados elementos $a_1, \dots, a_n \in L$ vale que:*

$$\begin{aligned} (\dots (a_1 \vee a_2) \vee \dots) \vee a_n &= \sup\{a_1, \dots, a_n\} \\ (\dots (a_1 \wedge a_2) \wedge \dots) \wedge a_n &= \inf\{a_1, \dots, a_n\}. \end{aligned}$$

Demostración. Haremos inducción en n . Claramente el resultado vale para $n = 1, 2$. Fijamos $a_1, \dots, a_{n+1} \in L$ y veremos que

$$s_{n+1} \stackrel{def}{=} (\dots (a_1 \vee a_2) \vee \dots) \vee a_{n+1} = \sup\{a_1, \dots, a_{n+1}\}.$$

De nuestra hipótesis se deduce que

$$(*) \quad s_n \stackrel{def}{=} (\dots (a_1 \vee a_2) \vee \dots) \vee a_n = \sup\{a_1, \dots, a_n\},$$

de lo que se sigue fácilmente que s_{n+1} es cota superior de $\{a_1, \dots, a_{n+1}\}$. Sea c otra cota superior. Ya que c es también cota superior del conjunto $\{a_1, \dots, a_n\}$, (*) implica que

$$s_n \leq c.$$

Como además sabemos que $a_{n+1} \leq c$, podemos concluir que

$$s_n \vee a_{n+1} \leq c,$$

es decir $s_{n+1} \leq c$. □

En un abierto desafío a las rigurosas costumbres matemáticas vamos a usar el nombre ‘reticulado’ también para otro tipo de estructura.

Definición 2.4. Un *reticulado* es una terna $\langle L, \vee, \wedge \rangle$, donde L es un conjunto no vacío cualquiera y \vee e \wedge son dos operaciones binarias sobre L que para todo $a, b \in L$ cumplen:

- $a \vee a = a \wedge a = a$ (idempotencia)
- $a \vee b = b \vee a$ y $a \wedge b = b \wedge a$ (conmutatividad)
- $(a \vee b) \vee c = a \vee (b \vee c)$ y $(a \wedge b) \wedge c = a \wedge (b \wedge c)$ (asociatividad)
- $a \vee (a \wedge b) = a$ y $a \wedge (a \vee b) = a$ (absorción).

Un vistazo al Lema 2.2 nos asegura que si $\langle L, \leq \rangle$ es un reticulado, y definimos \vee y \wedge como el supremo y el ínfimo de L respectivamente, entonces $\langle L, \vee, \wedge \rangle$ es un reticulado. Vemos entonces que, en cada reticulado $\langle L, \leq \rangle$ encontramos así un único (¿por qué?) reticulado de la nueva especie. Como veremos a continuación, todo reticulado $\langle L, \vee, \wedge \rangle$ puede ser encontrado de esta manera.

Teorema 2.5. *Sea $\langle L, \vee, \wedge \rangle$ un reticulado. La relación binaria definida por*

$$a \leq b \text{ si y solo si } a \vee b = b$$

es un orden parcial sobre L para el cual se cumple

$$\begin{aligned} \sup\{a, b\} &= a \vee b \\ \inf\{a, b\} &= a \wedge b. \end{aligned}$$

Demostración. Dejamos al lector verificar que \leq es reflexiva y antisimétrica. Probamos que \leq es transitiva. Supongamos que $a \leq b$ y $b \leq c$. Entonces

$$a \vee c = a \vee (b \vee c) = (a \vee b) \vee c = b \vee c = c,$$

por lo cual $a \leq c$. Veremos ahora que $\sup\{a, b\} = a \vee b$. Es claro que $a \vee b$ es una cota superior del conjunto $\{a, b\}$. Si $a, b \leq c$, entonces

$$(a \vee b) \vee c = a \vee (b \vee c) = a \vee c = c,$$

y así $a \vee b \leq c$. Es decir que $a \vee b$ es la menor cota superior.

Para comprobar que $\inf\{a, b\} = a \wedge b$ bastará con ver que para todo $u, v \in L$,

$$u \leq v \text{ si y solo si } u \wedge v = u,$$

y luego aplicar un razonamiento simétrico al del caso de \vee . Si $u \leq v$, es decir $u \vee v = v$, tenemos que $u \wedge v = u \wedge (u \vee v) = u$. Recíprocamente, si $u \wedge v = u$, entonces

$$\begin{aligned} u \vee v &= (u \wedge v) \vee v \\ &= v \vee (u \wedge v) \\ &= v \vee (v \wedge u) \\ &= v, \end{aligned}$$

lo cual equivale a que $u \leq v$. □

Se desprende de nuestro anterior teorema que un reticulado es esencialmente lo mismo que un reticulado. Y viceversa.

2.1. Homomorfismos de reticulados.

Definición 2.6. Sean $\langle L, \vee, \wedge \rangle$ y $\langle L', \vee', \wedge' \rangle$ reticulados. Una función $h : L \rightarrow L'$ es un *homomorfismo* de $\langle L, \vee, \wedge \rangle$ en $\langle L', \vee', \wedge' \rangle$ si para todo $a, b \in L$ vale que $h(a \vee b) = h(a) \vee' h(b)$ y $h(a \wedge b) = h(a) \wedge' h(b)$.

Si además h es biyectiva diremos que h es un *isomorfismo*.

Listamos a continuación algunas propiedades básicas.

Proposición 2.7. Sea h un homomorfismo de $\langle L, \vee, \wedge \rangle$ en $\langle L', \vee', \wedge' \rangle$.

1. Si h es un isomorfismo entonces h^{-1} es un homomorfismo.
2. h es un homomorfismo de posets de $\langle L, \leq \rangle$ en $\langle L', \leq' \rangle$. La recíproca no se cumple.
3. Si g es un isomorfismo de posets de $\langle L, \leq \rangle$ en $\langle L', \leq' \rangle$ entonces g es un isomorfismo de $\langle L, \vee, \wedge \rangle$ en $\langle L', \vee', \wedge' \rangle$.

Demostración. (1) y (2) quedan como ejercicios para el lector. El punto (3) es inmediato del Lema 1.5. □

3. RETICULADOS DISTRIBUTIVOS

Nos enfocaremos ahora en los reticulados en los cuales las operaciones \wedge y \vee distribuyen entre sí.

Lema 3.1. Sea $\langle L, \vee, \wedge \rangle$ un reticulado. Son equivalentes:

1. $\forall a, b, c \quad a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$.
2. $\forall a, b, c \quad a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$.

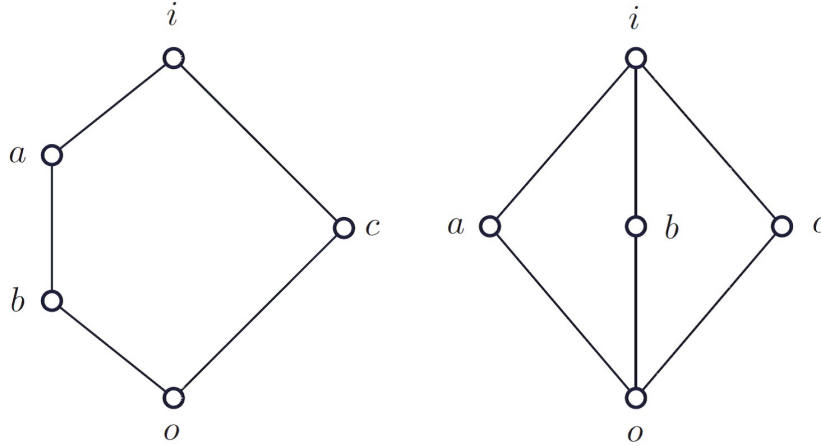
Demostración. Probamos (1) \Rightarrow (2), la otra implicación es análoga.

$$\begin{aligned} (a \vee b) \wedge (a \vee c) &= ((a \vee b) \wedge a) \vee ((a \vee b) \wedge c) \\ &= (a \vee (c \wedge (a \vee b))) \\ &= (a \vee ((c \wedge a) \vee (c \wedge b))) \\ &= a \vee (c \wedge b) = a \vee (b \wedge c). \end{aligned}$$

□

Un reticulado que cumple (1) y (2) es llamado *distributivo*.

Como lo muestra la siguiente figura, no todo reticulado es distributivo.



Por otra parte, si L es una familia de conjuntos cerrada bajo uniones e intersecciones finitas, entonces $\langle L, \cup, \cap \rangle$ obviamente satisface las leyes de distributividad. Esta clase de reticulados distributivos no sólo nos provee con una gran familia de ejemplos, sino que todo reticulado distributivo es esencialmente de esta forma. El resto de esta sección está dedicada a demostrar ese hecho.

Definición 3.2. Un *filtro* de $\langle L, \vee, \wedge \rangle$ es un subconjunto no vacío $F \subseteq L$ tal que:

1. $a, b \in F \Rightarrow a \wedge b \in F$ (cerrado bajo \wedge)
2. $a \in F$ y $a \leq b \Rightarrow b \in F$ (creciente).

Dado $S \subseteq L$, denotemos con $[S]$ al siguiente conjunto

$$\{a \in L : a \geq s_1 \wedge s_2 \wedge \cdots \wedge s_n, \text{ para algunos } s_1, s_2, \dots, s_n \in S, n \geq 1\}.$$

Lema 3.3. Si S es no vacío, entonces $[S]$ es un filtro. Más aún, $[S]$ es el menor filtro en contener a S .

Demostración. Ya que $S \subseteq [S]$, tenemos que $[S] \neq \emptyset$. Claramente $[S]$ cumple (2) de la Definición 3.2. Veamos que cumple (1). Si

$$\begin{aligned} a &\geq s_1 \wedge s_2 \wedge \cdots \wedge s_n \\ b &\geq t_1 \wedge t_2 \wedge \cdots \wedge t_m, \end{aligned}$$

con $s_1, s_2, \dots, s_n, t_1, \dots, t_m \in S$, entonces

$$a \wedge b \geq s_1 \wedge s_2 \wedge \cdots \wedge s_n \wedge t_1 \wedge t_2 \wedge \cdots \wedge t_m.$$

Por lo tanto $a \wedge b \in [S]$. □

Llamaremos a $[S]$ el *filtro generado por S* .

Sea $\langle P, \leq \rangle$ un poset. Un subconjunto $C \subseteq P$ es una *cadena* si para cada $a, b \in C$, se tiene que $a \leq b$ o $b \leq a$.

Lema 3.4 (Zorn). Si toda cadena del poset $\langle P, \leq \rangle$ tiene una cota superior, entonces hay un elemento maximal en $\langle P, \leq \rangle$.

Un filtro F de un reticulado $\langle L, \vee, \wedge \rangle$ será llamado *primo* cuando se cumplan:

- $F \neq L$
- $a \vee b \in F \Rightarrow a \in F$ o $b \in F$.

Teorema 3.5 (Filtro primo). Sea $\langle L, \vee, \wedge \rangle$ un reticulado distributivo y F un filtro. Supongamos $a \in L \setminus F$. Entonces hay un filtro primo P tal que $a \notin P$ y $F \subseteq P$.

Demostración. Sea

$$\mathcal{H} = \{H : H \text{ es un filtro, } a \notin H \text{ y } F \subseteq H\}.$$

Queremos aplicar el Lema de Zorn para obtener un elemento maximal del poset $\langle \mathcal{H}, \subseteq \rangle$, y para esto debemos ver que toda cadena tiene una cota superior. Fijemos entonces una cadena C de $\langle \mathcal{H}, \subseteq \rangle$. Si $C = \emptyset$, entonces cualquier elemento de \mathcal{H} es cota de C . Supongamos que C no es vacía, y sea G la unión de todos los filtros en C . Es claro que G es no vacío. Fijamos $b, b' \in G$, y sean $H, H' \in \mathcal{H}$ tales que $b \in H$ y $b' \in H'$. Como H y H' son comparables sabemos que $H \cup H'$ es un filtro. Por lo tanto $b \wedge b' \in H \cup H' \subseteq G$. Dejamos al lector el verificar que G es creciente.

Hemos visto que $G \in \mathcal{H}$, y es obvio que G es una cota superior de C . Por el lema de Zorn $\langle \mathcal{H}, \subseteq \rangle$ tiene un elemento maximal P . Nuestra prueba concluye mostrando que P es primo.

Razonaremos por el absurdo suponiendo que hay $p, q \in L \setminus P$ tales que $p \vee q \in P$. Al ser P maximal, tanto $[P \cup \{p\}]$ como $[P \cup \{q\}]$ no pueden ser elementos de \mathcal{H} . Esto nos dice que

$$a \in [P \cup \{p\}] \text{ y } a \in [P \cup \{q\}],$$

y así es que debe haber $p_1, \dots, p_n, q_1, \dots, q_m \in P$ que cumplen

$$a \geq p_1 \wedge \dots \wedge p_n \wedge p$$

$$a \geq q_1 \wedge \dots \wedge q_m \wedge q.$$

Si tomamos

$$s \stackrel{\text{def}}{=} p_1 \wedge \dots \wedge p_n \wedge q_1 \wedge \dots \wedge q_m,$$

se sigue que

$$a \geq s \wedge p$$

$$a \geq s \wedge q.$$

Pero entonces

$$a \geq (s \wedge p) \vee (s \wedge q) = s \wedge (p \vee q) \in P,$$

lo cual es una contradicción pues $a \notin P$. \square

Ahora sí estamos en condiciones de mostrar que todo reticulado distributivo es un reticulado de conjuntos.

Teorema 3.6. Sean $\langle L, \vee, \wedge \rangle$ un reticulado distributivo y X su conjunto de filtros primos. Para $a \in L$ definimos

$$\sigma(a) = \{P \in X : a \in P\}.$$

Entonces:

1. $\mathcal{L} = \{\sigma(a) : a \in L\}$ es cerrado bajo uniones e intersecciones finitas.
2. σ es un isomorfismo de $\langle L, \vee, \wedge \rangle$ en $\langle \mathcal{L}, \cup, \cap \rangle$.

Demostración. Ejercicio. \square

Concluimos esta sección con una consecuencia directa del Teorema 3.5 que necesitaremos más adelante.

Corolario 3.7. Sean $\langle L, \vee, \wedge \rangle$ un reticulado distributivo y $a, b \in L$ tales que $a \not\preceq b$. Entonces hay un filtro primo P tal que $a \notin P \ni b$.

Demostración. Tomar $F = \{x \in L : b \leq x\}$ y aplicar el Teorema 3.5. \square

4. LAS LEYES DE \cap Y \cup

Nos proponemos a continuación estudiar las familiares operaciones de unión e intersección entre conjuntos. Comencemos por listar algunas de sus propiedades conocidas:

- $A \cap A = A$
- $A \cup B = B \cup A$
- $A \cap (D \cap C) = (A \cap D) \cap C$
- $A \cup (E \cap X) = (A \cup E) \cap (A \cup X)$
- etcétera.

Por supuesto que las igualdades de esta lista valen para cualesquiera conjuntos A, B, C, \dots lo cual las hace mucho más interesantes que si describieran situaciones particulares. Observamos además que cada una de estas leyes tiene el mismo formato: es la igualdad entre dos expresiones formadas a partir de variables A, B, C, \dots (que representan conjuntos) conectadas por las operaciones \cap y \cup .

Veremos en esta sección que la teoría de reticulados distributivos permite dar un método muy simple para determinar si una igualdad de expresiones es una ‘ley’.

Comenzaremos por formalizar la noción de ‘expresión’. Sea

$$\text{Var} = \{X_1, X_2, \dots\}$$

un conjunto de numerable de variables. Definimos *expresión* mediante las siguientes reglas:

- $E_0 = \text{Var}$
- $E_{k+1} = E_k \cup \{(e_1 \cup e_2) : e_1, e_2 \in E_k\} \cup \{(e_1 \cap e_2) : e_1, e_2 \in E_k\}$
- $\text{Expresiones} = \bigcup_{k \geq 0} E_k$.

Sea e una expresión cuyas variables están todas en el conjunto $\{X_1, \dots, X_n\}$, y sean A_1, A_2, \dots, A_n conjuntos. Escribiremos $e(A_1, A_2, \dots, A_n)$ para denotar el conjunto que resulta de evaluar la expresión e en $X_1 = A_1, X_2 = A_2, \dots, X_n = A_n$.

Por ejemplo, si $e = (X_2 \cup X_4)$ entonces

$$e(\emptyset, \{1, 2\}, \{4\}, \{\pi\}, \mathbb{N}) = \{1, 2, \pi\}.$$

Sean e y e' expresiones. Diremos que la igualdad $e = e'$ es *válida* si para cualesquiera conjuntos A_1, \dots, A_n

$$(*) \quad e(A_1, \dots, A_n) = e'(A_1, \dots, A_n),$$

donde n es el menor natural tal que $\{X_1, \dots, X_n\}$ contiene todas las variables de e y e' .

Dado un conjunto S , diremos que la igualdad $e = e'$ es *válida para subconjuntos de S* si (*) se cumple siempre que $A_1, \dots, A_n \subseteq S$.

Como lo muestra el siguiente resultado los homomorfismos de reticulados preservan valuaciones de expresiones.

Lema 4.1. Sean \mathcal{F} y \mathcal{F}' familias de conjuntos cerradas bajo \cup e \cap , y sea h un homomorfismo de $\langle \mathcal{F}, \cup, \cap \rangle$ en $\langle \mathcal{F}', \cup, \cap \rangle$. Supongamos e es una expresión tal que todas sus variables están en $\{X_1, \dots, X_n\}$. Entonces para cualesquiera $A_1, \dots, A_n \in \mathcal{F}$ tenemos que

$$h(e(A_1, \dots, A_n)) = e(h(A_1), \dots, h(A_n)).$$

Demostración. Haremos inducción en k tal que $e \in E_k$. Si $e \in E_0 = \text{Var}$ el enunciado es obviamente cierto. Supongamos $e \in E_{k+1} \setminus E_k$ es tal que todas sus variables están en $\{X_1, \dots, X_n\}$, y fijamos $A_1, \dots, A_n \in \mathcal{F}$. Hay $e_1, e_2 \in E_k$ tal que $e = (e_1 \cup e_2)$ o

$e = (e_1 \cap e_2)$. Ya que los casos son simétricos supondremos que $e = (e_1 \cup e_2)$. Es claro que las variables de e_1 y e_2 están en $\{X_1, \dots, X_n\}$. Luego, por hipótesis inductiva, sabemos que

$$h(e_j(A_1, \dots, A_n)) = e_j(h(A_1), \dots, h(A_n)), \text{ para } j = 1, 2,$$

por lo cual

$$\begin{aligned} h(e(A_1, \dots, A_n)) &= h(e_1(A_1, \dots, A_n) \cup e_2(A_1, \dots, A_n)) \\ &= h(e_1(A_1, \dots, A_n)) \cup h(e_2(A_1, \dots, A_n)) \\ &= e_1(h(A_1), \dots, h(A_n)) \cup e_2(h(A_1), \dots, h(A_n)) \\ &= e(h(A_1), \dots, h(A_n)). \end{aligned}$$

□

Estamos en condiciones de caracterizar las igualdades válidas.

Teorema 4.2. *Son equivalentes:*

1. $e = e'$ es válida.
2. $e = e'$ es válida para subconjuntos del singulete $\{0\}$.

Demostración. (2) \Rightarrow (1) Razonaremos por el absurdo. Supongamos $e = e'$ es válida para subconjuntos de $\{0\}$, pero hay conjuntos A_1, \dots, A_n tales que

$$e(A_1, \dots, A_n) \neq e'(A_1, \dots, A_n).$$

Podemos suponer sin perder generalidad que $e(A_1, \dots, A_n) \not\subseteq e'(A_1, \dots, A_n)$. Sea $S = A_1 \cup \dots \cup A_n$. Observamos que $e(A_1, \dots, A_n)$ y $e'(A_1, \dots, A_n)$ son elementos de $\mathcal{P}(S)$. Por el Corolario 3.7 hay un filtro primo P del reticulado $\langle \mathcal{P}(S), \cup, \cap \rangle$ tal que

$$e(A_1, \dots, A_n) \notin P \ni e'(A_1, \dots, A_n).$$

Sea $h : \mathcal{P}(S) \rightarrow \{\emptyset, \{0\}\}$ definida por

$$h(A) = \begin{cases} \{0\} & \text{si } A \in P \\ \emptyset & \text{si } A \notin P. \end{cases}$$

No es difícil ver que h es homomorfismo de $\langle \mathcal{P}(S), \cup, \cap \rangle$ en $\langle \{\emptyset, \{0\}\}, \cup, \cap \rangle$. Luego, por el Lema 4.1 tenemos que

$$\begin{aligned} e(h(A_1), \dots, h(A_n)) &= h(e(A_1, \dots, A_n)) = \emptyset \\ e'(h(A_1), \dots, h(A_n)) &= h(e'(A_1, \dots, A_n)) = \{0\}. \end{aligned}$$

En particular,

$$e(h(A_1), \dots, h(A_n)) \neq e'(h(A_1), \dots, h(A_n)),$$

lo cual contradice el hecho de que $e = e'$ es válida para subconjuntos $\{0\}$. □

Es interesante observar que el teorema anterior transforma el problema de decidir si una igualdad es válida en una verificación finitaria (que podría ser realizada por una computadora).

REFERENCIAS

- [1] R. Balbes and P. Dwinger, *Distributive lattices*, University of Missouri Press, 1975.
- [2] B.A. Davey and H.A. Priestley, *Introduction to lattices and order*, Cambridge University Press, 2002.
- [3] G.A. Grätzer, *General lattice theory*, Birkhäuser Verlag, 2003.

FACULTAD DE MATEMÁTICA ASTRONOMÍA Y FÍSICA, UNIVERSIDAD NACIONAL DE CÓRDOBA,
ARGENTINA

E-mail address: `camper@famaf.unc.edu.ar`

TEOREMA DEL PUNTO FIJO EN ESPACIOS MÉTRICOS: UNA APLICACIÓN A FRACTALES

MARILINA CARENA

RESUMEN. Existen objetos que poseen la propiedad de ser *autosemejantes*, en el sentido que su estructura geométrica básica se repite en escalas distintas, por lo que el objeto total puede ser visto como una unión finita de copias de sí mismo en diferentes escalas. Mandelbrot propuso el término “fractal” para describir este tipo de objetos. Los fractales son objetos matemáticos de gran importancia por sus complejas propiedades. Una de las formas más conocidas para generar fractales es a través de un *sistema iterado de funciones* (SIF), que consiste en una familia de contracciones $\{\phi_1, \phi_2, \dots, \phi_M\}$ sobre un espacio métrico (X, d) , siendo $M \geq 2$. Si X es completo, la propiedad fundamental de un SIF es que determina un único conjunto compacto no vacío K que satisface $K = \bigcup_{i=1}^M \phi_i(K)$, por lo que K es llamado *atractor* o *conjunto invariante* por el SIF. Más aún, si \mathcal{K} denota el conjunto de todos los subconjuntos compactos no vacíos de X y T es la función definida sobre \mathcal{K} como

$$T(E) = \bigcup_{i=1}^M \phi_i(E),$$

entonces el atractor K puede obtenerse como límite en la métrica de Hausdorff d_H de la sucesión $T^n(E)$, donde E es un conjunto compacto no vacío cualquiera y T^n denota la composición de T consigo misma n veces. Esto fue probado por Hutchinson en [5] utilizando el teorema del punto fijo y la completitud del espacio (\mathcal{K}, d_H) , ya que la aplicación T así definida resulta contractiva, donde \mathcal{K} denota el conjunto de todos los subconjuntos compactos no vacíos de X .

El objetivo del curso es desarrollar los conceptos y resultados anteriormente mencionados, con el fin de obtener aproximaciones a fractales clásicos, como los conjuntos de Cantor, de Sierpinski y de Koch, diferentes a las usuales.

ÍNDICE

| | |
|-------------------------------------|----|
| Organización del curso | 64 |
| 1. Espacios métricos | 64 |
| 2. La distancia de Hausdorff | 67 |
| 3. Teorema del punto fijo de Banach | 70 |
| 4. Sistemas iterados de funciones | 72 |
| 5. Aproximaciones a fractales | 76 |
| Referencias | 80 |

2010 *Mathematics Subject Classification.* 54E50; 54E40.

Financiamiento: CONICET, CAI+D (UNL) y ANPCyT.

Agradecimientos: a Marcelo Actis y Gracia Salazar, por la lectura detallada de estas páginas, por sus correcciones y comentarios constructivos.

ORGANIZACIÓN DEL CURSO

En la Sección 1 se enunciarán algunos conceptos y resultados básicos en espacios métricos, que serán utilizados para el desarrollo del curso. Un ejemplo particular de métrica será presentado en la Sección 2: la distancia de Hausdorff entre conjuntos compactos. La Sección 3 será destinada a la prueba del teorema del punto fijo de Banach, el cual nos permitirá obtener aproximaciones a fractales clásicos, como los conjuntos de Cantor, Koch y Sierpinski, diferentes a las usuales. Estas aproximaciones serán presentadas en la Sección 5. Una herramienta fundamental para generar dichas aproximaciones serán los sistemas iterados de funciones (SIF), los cuales serán estudiados en la Sección 4, así como la aplicación contractiva inducida por un SIF dado.

1. ESPACIOS MÉTRICOS

En esta sección reuniremos algunos conceptos y resultados básicos de espacios métricos, necesarios para el desarrollo del curso. Un análisis más completo y detallado de este tema puede encontrarse en una gran cantidad de libros, como por ejemplo [6], [3], [1] y [4], entre otros. No es necesaria la lectura de esta sección para los alumnos familiarizados con espacio métricos.

Definición 1.1. Sea X un conjunto. Una **métrica** sobre X es una función no negativa y finita d definida sobre $X \times X$ tal que

- i) $d(x, x) = 0$ para todo $x \in X$;
- ii) $d(x, y) = 0$ implica $x = y$ (confiabilidad);
- iii) $d(x, y) = d(y, x)$ para todo $x, y \in X$ (simetría);
- iv) $d(x, y) \leq d(x, z) + d(z, y)$ para todo $x, y, z \in X$ (desigualdad triangular).

Si d satisface todas las propiedades anteriores excepto ii) entonces d es una **pseudométrica**. Decimos que el par (X, d) es un **espacio métrico** si d es una métrica sobre X .

Ejemplo 1.2.

- a) En el conjunto de los número reales \mathbb{R} podemos definir una métrica tomando el valor absoluto de la diferencia, es decir, $d(x, y) := |x - y|$. Las propiedades de métrica se deducen inmediatamente de las propiedades conocidas del valor absoluto. A esta métrica le llamaremos **distancia usual** en \mathbb{R} .
- b) Si (x_1, y_1) y (x_2, y_2) son puntos en \mathbb{R}^2 , las siguientes funciones son métricas en \mathbb{R}^2 :

$$\begin{aligned} d_1((x_1, y_1), (x_2, y_2)) &:= |x_1 - x_2| + |y_1 - y_2|; \\ d_2((x_1, y_1), (x_2, y_2)) &:= \sqrt{(|x_1 - x_2|^2 + |y_1 - y_2|^2)}; \\ d_\infty((x_1, y_1), (x_2, y_2)) &:= \max\{|x_1 - x_2|, |y_1 - y_2|\}. \end{aligned}$$

El hecho que estas tres funciones sean métricas en \mathbb{R}^2 es consecuencia de un resultado más general contenido en el Ejercicio 1.

- c) Sea $X = C[0, 1]$ el espacio de todas las funciones continuas en el intervalo $[0, 1]$. Para $f, g \in C[0, 1]$ se define

$$d(f, g) := \sup\{|f(x) - g(x)| : x \in [0, 1]\}.$$

Se puede verificar que d es una métrica sobre $C[0, 1]$.

- d) Sea X un conjunto no vacío. La **métrica discreta** sobre X se define como

$$d(x, y) := \begin{cases} 0 & \text{si } x = y; \\ 1 & \text{si } x \neq y. \end{cases}$$

Observación 1.3. Es importante señalar que si (X, d) es un espacio métrico, entonces todo subconjunto Y de X es a su vez un espacio métrico con la misma función d . Por ejemplo, el conjunto \mathbb{Q} de los números racionales es un espacio métrico con la distancia usual heredada de \mathbb{R} .

Los siguientes conceptos y resultados serán fundamentales en el desarrollo del curso, y en ellos (X, d) siempre denotará un espacio métrico.

Definición 1.4. Si $x \in X$ y $r > 0$, la **bola** (abierta) de centro x y radio r con respecto a la métrica d es

$$B_d(x, r) = \{y \in X : d(x, y) < r\}.$$

Definición 1.5. Un subconjunto E de X es **abierto** si para todo $x \in E$ existe $r > 0$ tal que $B_d(x, r) \subseteq E$, y es **cerrado** si su complemento $E^c = X \setminus E$ es abierto.

Definición 1.6. Sea $\{x_n\}$ una sucesión en X . Decimos que $\{x_n\}$ **converge** (con respecto a d) a un punto $x \in X$ si $\lim_{n \rightarrow \infty} d(x_n, x) = 0$. Lo denotaremos $x_n \rightarrow x$ o $\lim x_n = x$.

Definición 1.7. La sucesión $\{x_n\}$ es de **Cauchy** si $d(x_n, x_m) \rightarrow 0$ cuando $n, m \rightarrow \infty$. En otras palabras, si para todo $\varepsilon > 0$ existe un entero positivo N tal que

$$d(x_n, x_m) < \varepsilon \quad \text{para todo } n, m \geq N.$$

Definición 1.8. Decimos que el espacio métrico (X, d) es **completo** si toda sucesión de Cauchy en X es convergente. Un subconjunto E de X es llamado completo si toda sucesión de Cauchy en E converge a un elemento de E .

Puede probarse que \mathbb{R}^n equipado con la métrica usual

$$|x - y| = |(x_1, \dots, x_n) - (y_1, \dots, y_n)| = \left[\sum_{i=1}^n (x_i - y_i)^2 \right]^{1/2},$$

es completo (ver por ejemplo [6]).

El siguiente resultado relaciona espacios completos y cerrados, y su demostración puede encontrarse en [6] o [4], entre otros.

Proposición 1.9. *Sea (X, d) un espacio métrico. Entonces todo subconjunto completo de X es cerrado, y si X es completo entonces todo subconjunto cerrado de X es completo.*

Definición 1.10. Si $E \subseteq X$, la unión de todos los subconjuntos abiertos de E es llamado el **interior** de E , y lo denotaremos por E° . Además llamaremos **clausura** de E , y lo denotaremos con \overline{E} , a la intersección de todos los conjuntos cerrados que contienen a E .

Observación 1.11. Notar que E es abierto si y sólo si $E = E^\circ$, y que E es cerrado si y sólo si $E = \overline{E}$.

Observación 1.12. Se deja como ejercicio ver que si E es un subconjunto de X y $x \in X$, entonces $x \in \overline{E}$ si y sólo si existe una sucesión $\{x_n\}$ en E que converge a x .

Definición 1.13. El **diámetro** de un subconjunto E de X como

$$\text{diam}E = \sup\{d(x, y) : x, y \in E\}.$$

Decimos que E es **acotado** si $\text{diam}E < \infty$.

Definición 1.14. Si $E \subseteq X$ y $\{V_\alpha\}_{\alpha \in A}$ es una familia de subconjuntos de X tales que $E \subseteq \bigcup_{\alpha \in A} V_\alpha$, entonces $\{V_\alpha\}_{\alpha \in A}$ es llamado un **cubrimiento** de E . Si el conjunto de índices A es finito, decimos que es un **cubrimiento finito** de E , y si cada V_α es abierto, decimos que la familia es un **cubrimiento abierto**.

Definición 1.15. Un conjunto que tiene la propiedad que de todo cubrimiento abierto se puede seleccionar un subcubrimiento finito se llama **conjunto compacto**.

El siguiente resultado relaciona algunos de los conceptos anteriores, y su prueba puede verse por ejemplo en [1] o [6].

Proposición 1.16. Sea (X, d) un espacio métrico. Si E es un subconjunto compacto de X , entonces E es cerrado y acotado.

El recíproco de la proposición anterior es falso en general, aunque es cierto en \mathbb{R}^n : todo subconjunto cerrado y acotado de \mathbb{R}^n es compacto. En el caso general es cierto siempre que (X, d) sea compacto.

Definición 1.17. Si (X, d_X) e (Y, d_Y) son dos espacios métricos, una función $f : X \rightarrow Y$ es llamada **continua en un punto** $x \in X$ si para todo $\varepsilon > 0$ existe $\delta > 0$ tal que $d_Y(f(x), f(z)) < \varepsilon$ siempre que $d_X(x, z) < \delta$ (o equivalentemente, $f(B_{d_X}(x, \delta)) \subseteq B_{d_Y}(f(x), \varepsilon)$). Decimos que f es **continua** si es continua en cada punto $x \in X$.

Se deja como ejercicio la prueba del siguiente resultado que relaciona compacidad y funciones continuas.

Teorema 1.18. Sean X e Y dos espacios métricos y sea $f : X \rightarrow Y$ es una función continua. Si K es un subconjunto compacto de X , entonces $f(K)$ es compacto en Y .

Ejercicios.

- Sean (X, d_X) , (Y, d_Y) dos espacios métricos. Probar que cada una de las siguientes funciones definen una métrica sobre $X \times Y$:
 - $d_1((x_1, y_1), (x_2, y_2)) = d_X(x_1, x_2) + d_Y(y_1, y_2)$.
 - $d_2((x_1, y_1), (x_2, y_2)) = (d_X(x_1, x_2)^2 + d_Y(y_1, y_2)^2)^{1/2}$.
 - $d_\infty((x_1, y_1), (x_2, y_2)) = \max\{d_X(x_1, x_2), d_Y(y_1, y_2)\}$.
- Probar que la distancia usual en \mathbb{R}^n es una métrica.
- Sea X un conjunto no vacío equipado con la métrica discreta. Probar que todo subconjunto de X es abierto y cerrado.
- Sea (X, d) un espacio métrico y sea Λ un conjunto de índices. Probar:
 - Si G_λ es abierto en X para todo $\lambda \in \Lambda$, entonces $G = \bigcup_{\lambda \in \Lambda} G_\lambda$ es abierto en X .
 - Si F_λ es cerrado en X para todo $\lambda \in \Lambda$, entonces $F = \bigcap_{\lambda \in \Lambda} F_\lambda$ es cerrado en X .
 - Si G_i es abierto en X para todo $i = 1, 2, \dots, n$, entonces $G = \bigcap_{i=1}^n G_i$ es abierto en X . ¿Qué ocurre con la intersección de una familia arbitraria de conjuntos abiertos?
 - Si F_i es cerrado en X para todo $i = 1, 2, \dots, n$, entonces $F = \bigcup_{i=1}^n F_i$ es cerrado en X . ¿Qué ocurre con la unión de una familia arbitraria de conjuntos cerrados?
- Sean (X, d) un espacio métrico, E un subconjunto de X y $x \in X$. Probar que $x \in \overline{E}$ si y sólo si existe una sucesión $\{x_n\}$ en E que converge a x .
- Demostrar que en todo espacio métrico (X, d) se tiene que $\overline{B_d(x, r)} \subseteq \{y : d(x, y) \leq r\}$. Dar un ejemplo para el cual la inclusión sea propia.

7. Sea (X, d) un espacio métrico y sea A un subconjunto no vacío de X . Dado un punto $x \in X$, la distancia de x a A se define como

$$d(x, A) = \inf\{d(x, a) : a \in A\}.$$

Mostrar que $d(x, A) = 0$ si y sólo si $x \in \bar{A}$.

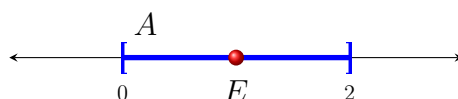
8. Sea (X, d) un espacio métrico y sean K_1, \dots, K_M una familia (finita) de conjuntos compactos. Probar que $K = \bigcup_{i=1}^M K_i$ es compacto. ¿Qué ocurre con la unión de una familia arbitraria de conjuntos compactos?
9. Sean X e Y dos espacios métricos y sea $f : X \rightarrow Y$ una función, y sea $f^{-1}(E) := \{x \in X : f(x) \in E\}$. Probar que las siguientes afirmaciones son equivalentes:
- f es continua.
 - $f^{-1}(G)$ es abierto en X para todo G abierto en Y .
 - $f^{-1}(F)$ es cerrado en X para todo F cerrado en Y .
10. Sea (X, d) un espacio métrico y sea $y \in X$ fijo. Probar que la función $f : X \rightarrow \mathbb{R}$ definida como $f(x) = d(x, y)$ es continua.
11. Sean (X, d_X) e (Y, d_Y) dos espacios métricos, y sea $f : X \rightarrow Y$ una función continua. Probar que si $\{x_n\}$ converge a x con respecto a d_X , entonces $\{f(x_n)\}$ converge a $f(x)$ con respecto a d_Y .
12. Probar el Teorema 1.18. [*Sugerencia:* utilizar la definición de compacidad.]

2. LA DISTANCIA DE HAUSDORFF

Dado un espacio métrico (X, d) la distancia usual entre dos subconjuntos no vacíos A y E de X , se define como

$$D(A, E) = \inf\{d(x, y) : x \in A, y \in E\} = \inf\{d(x, E) : x \in A\}.$$

Se deja como ejercicio probar que D no es una métrica (ni siquiera es una pseudo-métrica) sobre todos los subconjuntos no vacíos de X . A pesar de ello D define una función simétrica que resulta útil en muchos problemas en los que se necesita obtener una medida de la “distancia” entre dos conjuntos (es decir, cuando se quiere medir cuán lejos están uno del otro), pero no es efectiva cuando se necesita de comparar la “forma” entre los mismos. Para ilustrar esta idea, consideremos $X = \mathbb{R}$ equipado con la distancia usual, $A = [0, 2]$ y $E = \{1\}$.



El hecho que $E \subseteq A$ implica $D(A, E) = 0$. En palabras, A y E están a distancia nula. Sin embargo esto no nos brinda ninguna información sobre la “forma” de ambos conjuntos, ya que como se ve en el ejemplo, pueden ser muy distintos entre sí y aún así estar a distancia cero uno del otro.

El objetivo de esta sección es definir una distancia entre subconjuntos de un espacio métrico (X, d) que tenga en cuenta tanto la “distancia” entre ellos como su “forma”. Para ello necesitaremos la siguiente definición.

Definición 2.1. Sea (X, d) un espacio métrico. Si $E \subseteq X$ y $\varepsilon > 0$, el ε -engordado de E , denotado por $[E]_\varepsilon$, se define por

$$[E]_\varepsilon = \bigcup_{x \in E} B_d(x, \varepsilon) = \{x \in X : d(x, E) < \varepsilon\}.$$

Para ilustrar esta definición, notar que en el ejemplo anterior tenemos que $[E]_1 = (0, 2)$. Es decir, si “1-engordamos” el conjunto E (el cual está formado solamente por el punto $x = 1$) en \mathbb{R} con la distancia usual, obtenemos el intervalo abierto $(0, 2)$.

Ya estamos en condiciones de definir la la distancia que nos interesa.

Definición 2.2. Sean A y E subconjuntos de un espacio métrico (X, d) . La **distancia de Hausdorff** entre A y E se define como

$$d_H(A, E) = \inf\{\varepsilon > 0 : A \subseteq [E]_\varepsilon \text{ y } E \subseteq [A]_\varepsilon\}.$$

En palabras, la distancia de Hausdorff entre dos conjuntos es la “menor” cantidad posible que hay que engordar a ambos conjuntos para que cada uno esté contenido en el engordado del otro. Siguiendo con el ejemplo, se tiene que $d_H(A, E) = 1$. Esto nos da una noción de que ambos conjuntos no tienen una forma similar, pues nos dice que hay que engordar bastante (donde “bastante” significa una cantidad no nula) a al menos uno de ellos para que se parezca al otro. Sin embargo, notar que esta distancia no es confiable, puesto que por ejemplo si tomamos $F = (0, 2)$ y $A = [0, 2]$ como antes, entonces $d_H(A, F) = 0$ pero $A \neq F$. Por otro lado, no habría forma de incluir un conjunto en ningún engordado del conjunto vacío, pues por mucho que se lo engorde sigue siendo vacío. Además, si queremos que la distancia tome valores finitos, no podemos considerar conjuntos no acotados, pues la única forma de incluir un conjunto no acotado en el engordado de otro que sí lo sea, sería engordándolo a este último una cantidad infinita. Resumiendo, si queremos que d_H resulte una métrica, debemos restringir el dominio de los conjuntos: pediremos que sean no vacíos (para que esté bien definida), que sean cerrados (para que resulte confiable), y que sean acotados (para que tome valores finitos). Puesto que todo compacto es cerrado y acotado, será suficiente trabajar en la siguiente familia de conjuntos:

$$\mathcal{K} := \{K \subseteq X : K \neq \emptyset, K \text{ compacto}\}.$$

Teorema 2.3. d_H es una métrica sobre \mathcal{K} .

Observación 2.4. Notar que tanto \mathcal{K} como d_H dependen fuertemente de la métrica d considerada sobre X .

Observación 2.5. En realidad puede probarse que d_H sea una métrica sobre

$$\tilde{\mathcal{K}} := \{K \subseteq X : K \neq \emptyset, K \text{ es cerrado y acotado}\},$$

el cual contiene a \mathcal{K} (ver Proposición 1.16). Sin embargo, para los resultados de este curso nos bastará con trabajar con el conjunto \mathcal{K} . Por otra parte, para la construcción de los fractales clásicos se trabaja con $X = \mathbb{R}$ equipado con la distancia usual, y en este caso $\mathcal{K} = \tilde{\mathcal{K}}$ (pues K es compacto en \mathbb{R} si y sólo si es cerrado y acotado).

Dem. Teorema 2.3. De la definición se ve claramente que d_H es una función no negativa y que $d_H(A, A) = 0$. Para ver que es confiable, supongamos que $d_H(A, E) = 0$. Queremos ver que $A = E$. Por ser un ínfimo, la hipótesis $d_H(A, E) = 0$ implica que para cada $n \in \mathbb{N}$ se tiene que $A \subseteq [E]_{\frac{1}{n}}$ y $E \subseteq [A]_{\frac{1}{n}}$. Luego, si fijamos $x \in A$, para cada $n \in \mathbb{N}$ existe $x_n \in E$ tal que $d(x, x_n) < \frac{1}{n}$. Entonces $d(x_n, x)$ tiende a cero cuando n tiende a infinito. En otras palabras, x_n converge a x . Puesto que $\{x_n\} \subseteq E$ y E es cerrado, se tiene que $x \in E$. Hemos probado que $A \subseteq E$, y análogamente se prueba la otra inclusión.

De la definición se ve claramente que d_H es una función simétrica. Sólo resta ver que vale la desigualdad triangular. Para ver esto usaremos las siguientes propiedades, cuya prueba es sencilla y queda como ejercicio: si $A \subseteq E$ entonces $[A]_\varepsilon \subseteq [E]_\varepsilon$; además $[[A]_\varepsilon]_{\tilde{\varepsilon}} \subseteq [A]_{\varepsilon+\tilde{\varepsilon}}$ para todo $\varepsilon, \tilde{\varepsilon} > 0$.

Fijemos entonces A, B y E en \mathcal{K} . Notar que por ser la distancia de Hausdorff un ínfimo, para cada $\varepsilon > 0$ se tiene que $A \subseteq [B]_{d_H(A,B)+\varepsilon}$ y $B \subseteq [E]_{d_H(B,E)+\varepsilon}$. Aplicando las dos propiedades mencionadas obtenemos que

$$A \subseteq [B]_{d_H(A,B)+\varepsilon} \subseteq [[E]_{d_H(B,E)+\varepsilon}]_{d_H(A,B)+\varepsilon} \subseteq [E]_{d_H(A,B)+d_H(B,E)+2\varepsilon}.$$

De forma análoga se prueba que

$$E \subseteq [B]_{d_H(E,B)+\varepsilon} \subseteq [[A]_{d_H(B,A)+\varepsilon}]_{d_H(E,B)+\varepsilon} \subseteq [A]_{d_H(A,B)+d_H(B,E)+2\varepsilon}.$$

Por lo tanto, $d_H(A, E) \leq d_H(A, B) + d_H(B, E) + 2\varepsilon$. Puesto que esto vale para cada $\varepsilon > 0$, obtenemos que $d_H(A, E) \leq d_H(A, B) + d_H(B, E)$, como queríamos. Luego (\mathcal{K}, d_H) es un espacio métrico. \square

Si A y $\{A_n\}_{n=1}^\infty$ son conjuntos compactos no vacíos, decimos que la sucesión $\{A_n\}$ converge al conjunto A en el sentido de Hausdorff si $d_H(A_n, A) \rightarrow 0$ cuando $n \rightarrow \infty$, y lo denotamos por $A_n \xrightarrow{d_H} A$.

El espacio métrico (\mathcal{K}, d_H) es llamado el **espacio de Hausdorff inducido** por (X, d) , y como veremos luego, es el ambiente ideal para estudiar la geometría fractal. La propiedad más importante para los resultados de este curso está contenida en el siguiente teorema, el cual establece que el espacio de Hausdorff inducido por un espacio métrico preserva la completitud. Este hecho está enunciado en el siguiente teorema.

Teorema 2.6. *Si (X, d) es un espacio métrico completo, entonces (\mathcal{K}, d_H) también es completo.*

Idea de la prueba. Sea $\{A_n\}$ una sucesión de Cauchy en (\mathcal{K}, d_H) . Debemos probar que $\{A_n\}$ converge a algún elemento en \mathcal{K} . Debido al grado de dificultad de la prueba, sólo presentaremos al conjunto A que será el límite de la sucesión $\{A_n\}$. Dicho conjunto se define como

$$A := \{x \in X : \text{existe } \{x_n\} \text{ suc. de Cauchy en } X \text{ que converge a } x, \text{ con } x_n \in A_n\}.$$

Para completar la prueba restaría probar que A pertenece a \mathcal{K} y que $d_H(A_n, A) \rightarrow 0$ cuando $n \rightarrow \infty$. La prueba completa de este teorema puede encontrarse por ejemplo en [2]. \square

Como consecuencia directa del teorema anterior y del hecho que \mathbb{R}^n equipado con la distancia usual es un espacio métrico completo, tenemos el siguiente resultado.

Corolario 2.7. *El espacio de Hausdorff inducido por $(\mathbb{R}^n, |\cdot|)$ es un espacio métrico completo.*

Ejercicios.

1. Sea (X, d) un espacio métrico y sean $A, E \subseteq X$ no vacíos.

$$D(A, E) = \inf\{d(x, y) : x \in A, y \in E\}.$$

Probar que D no satisface la desigualdad triangular, por lo que no es una pseudo-métrica sobre los subconjuntos no vacíos de X .

2. Sea (X, d) un espacio métrico. Probar:

- a) si $A \subseteq E$ entonces $[A]_\varepsilon \subseteq [E]_\varepsilon$ para todo $\varepsilon > 0$;
 b) $[[A]_\varepsilon]_{\tilde{\varepsilon}} \subseteq [A]_{\varepsilon+\tilde{\varepsilon}}$ para todo $\varepsilon, \tilde{\varepsilon} > 0$ y todo $A \subseteq X$. Dar un ejemplo en el que la inclusión sea propia.
3. Sea $X = \mathbb{R}$, $d(x, y) = |x - y|$ y d_H la distancia de Hausdorff inducida. Hallar $d_H(A, E)$, siendo $A = [1, 3]$ y $E = [7, 15]$.
4. Sean A y E dos conjuntos en \mathcal{K} . Probar que

$$d_H(A, E) = \max\{\rho(A, E), \rho(E, A)\},$$

donde $\rho(A, E) = \sup_{x \in A} \{d(x, E)\}$ y $d(x, E) = \inf_{y \in E} \{d(x, y)\}$ es la distancia de x a E .

3. TEOREMA DEL PUNTO FIJO DE BANACH

Definición 3.1. Un punto fijo de una función f es un elemento x^* del dominio de f que satisface $f(x^*) = x^*$.

En palabras, x^* es un punto fijo de una función f si las gráficas de f y la función identidad $g(x) = x$ se intersecan en x^* .

Ejemplo 3.2. Sea $f(x) = x^2 - 6$. Encontrar los puntos fijos de f significa resolver la ecuación $x^2 - 6 = x$, o equivalentemente $x^2 - x - 6 = 0$. Aplicando la resolvente se obtiene que los puntos fijos de f son 3 y -2 (ver Figura 1).

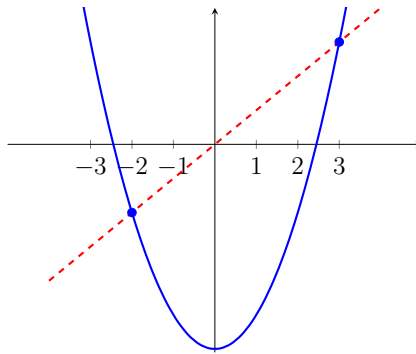


FIGURA 1. Puntos fijos de $f(x) = x^2 - 6$.

Es claro que no todas las funciones poseen un punto fijo. Sin embargo, existe una clase especial de funciones las cuales en un dominio adecuado poseen exactamente un punto fijo.

Definición 3.3. Sea (X, d) un espacio métrico. Una función $f : X \rightarrow X$ se dice que es una **aplicación contractiva** si existe un número positivo $0 \leq k < 1$ tal que

$$d(f(x), f(y)) \leq kd(x, y)$$

para todo $x, y \in X$. Si ocurre que $d(f(x), f(y)) = kd(x, y)$ para todo $x, y \in X$, decimos que f es una **similitud contractiva**.

Probaremos ahora el teorema que será la herramienta fundamental en el desarrollo de este trabajo, que es llamado Teorema del punto fijo de Banach o Teorema de la aplicación contractiva.

Teorema 3.4 (Teorema del punto fijo de Banach). *Sea (X, d) un espacio métrico completo. Si $f : X \rightarrow X$ es una aplicación contractiva, entonces f tiene un único punto fijo en X .*

Demostración. Comencemos probando la existencia de un punto fijo. Para ello fijemos $x_0 \in X$, y definamos la sucesión $x_n = f(x_{n-1}) = f^n(x_0)$ para $n \geq 1$, donde f^n denota la composición de f consigo misma n veces. Es decir, $x_1 = f(x_0)$, $x_2 = f(x_1) = (f \circ f)(x_0)$, y para cada natural n , se tiene

$$x_n = \underbrace{(f \circ f \circ \cdots \circ f)}_{n \text{ veces}}(x_0).$$

Veremos ahora que esta sucesión tiene un límite x^* en X , y que este límite es punto fijo de f . En efecto, de la definición de x_n y por ser f contractiva tenemos que

$$d(x_{n+1}, x_n) = d(f(x_n), f(x_{n-1})) \leq kd(x_n, x_{n-1}).$$

Reiterando razonamiento anterior obtenemos

$$d(x_{n+1}, x_n) \leq kd(x_n, x_{n-1}) \leq k^2d(x_{n-1}, x_{n-2}) \leq \cdots \leq k^nd(x_1, x_0).$$

Usaremos esto junto a la desigualdad triangular de d para probar que la sucesión $\{x_n\}$ es de Cauchy. Para ello fijemos dos enteros positivos m y n tales que $n > m$. Entonces

$$\begin{aligned} d(x_n, x_m) &\leq d(x_n, x_{n-1}) + d(x_{n-1}, x_{n-2}) + \cdots + d(x_{m+1}, x_m) \\ &\leq d(x_1, x_0)(k^{n-1} + k^{n-2} + \cdots + k^m) \\ &= d(x_1, x_0)k^m(k^{n-m-1} + \cdots + k^2 + k + 1) \\ &\leq d(x_1, x_0)k^m \sum_{i=1}^{\infty} k^i \\ &= k^m \frac{d(x_1, x_0)}{1 - k}. \end{aligned}$$

Esta última cantidad puede hacerse arbitrariamente pequeña eligiendo m suficientemente grande, puesto que $k < 1$. Es decir, $\{x_n\}$ es de Cauchy en (X, d) , el cual es completo. Luego existe el límite de $\{x_n\}$, al cual denotaremos x^* . Es decir, $d(x_n, x^*) \rightarrow 0$ cuando $n \rightarrow \infty$. Resta ver que $f(x^*) = x^*$. Para ello, notar que para todo $n \in \mathbb{N}$ se cumple que

$$d(x^*, f(x^*)) \leq d(x^*, x_n) + d(x_n, f(x^*)) \leq d(x^*, x_n) + kd(x_{n-1}, x^*).$$

Haciendo n tender a infinito obtenemos $d(x^*, f(x^*)) = 0$, lo cual implica $f(x^*) = x^*$ por ser d una métrica.

Para demostrar la unicidad no se necesita que el espacio sea completo, sino que sale del hecho de que f es contractiva. Es decir, una aplicación contractiva tiene a lo sumo un punto fijo. Para verlo, supongamos que x^* y \bar{x} son dos puntos fijos de f . Es decir, $f(x^*) = x^*$ y $f(\bar{x}) = \bar{x}$. Queremos ver que $x^* = \bar{x}$. En efecto,

$$d(f(x^*), f(\bar{x})) \leq kd(x^*, \bar{x}) = kd(f(x^*), f(\bar{x})).$$

Puesto que $k < 1$ esto sólo es posible si $d(f(x^*), f(\bar{x})) = 0$, lo cual implica $f(x^*) = f(\bar{x})$, por ser d una métrica. Puesto que x^* y \bar{x} son ambos puntos fijos de f , obtenemos $x^* = \bar{x}$, como queríamos ver. \square

Observación 3.5. Notar que la prueba del teorema proporciona un método constructivo para hallar el punto fijo: se fija un punto x_0 en X y se construye la sucesión $x_n = f(x_{n-1}) = f^n(x_0)$ para $n \geq 1$. El punto fijo x^* es el límite de esta sucesión. Es muy importante señalar que el punto fijo no depende del punto inicial, es decir, el punto de partida x_0 puede ser *cualquier* punto en X .

Observación 3.6. Otro hecho importante que se desprende de la demostración del teorema anterior, es que puede obtenerse una *cota para el error* en la aproximación al punto fijo x^* . En efecto, podemos fijar m y hacer tender n a infinito en la desigualdad

$$d(x_n, x_m) \leq k^m \frac{d(x_1, x_0)}{1 - k},$$

para obtener que

$$d(x^*, x_m) \leq k^m \frac{d(x_1, x_0)}{1 - k}$$

para todo m . Aquí hemos aplicado el hecho que $\lim_{n \rightarrow \infty} d(x_n, x_m) = d(x^*, x_m)$ (ver Ejercicios 10 y 11 de la Sección 1), y que x_n converge a x^* .

Ejercicios.

1. Sea (X, d) un espacio métrico y sea $f : X \rightarrow X$ una función contractiva. Probar que f es continua en X .
2. Sea $I = [a, b]$ un intervalo en \mathbb{R} y sea $f : I \rightarrow \mathbb{R}$ derivable. Demostrar que si la derivada f' satisface $|f'(x)| \leq K$ para todo $x \in I$, siendo $0 \leq K < 1$, entonces f es contractiva.

4. SISTEMAS ITERADOS DE FUNCIONES

Siguiendo la terminología introducida por Hutchinson en [5], presentaremos en esta sección el concepto de sistema iterado de funciones actuando sobre un espacio métrico.

Definición 4.1. Un **sistema iterado de funciones** (SIF) consiste en una familia de contracciones $\Phi = \{\phi_1, \dots, \phi_M\}$ sobre un espacio métrico (X, d) , donde $M \geq 2$. Esto significa que para cada $i = 1, \dots, M$ se tiene que $\phi_i : X \rightarrow X$ y existe una constante $k_i < 1$ tal que

$$(4.1) \quad d(\phi_i(x), \phi_i(y)) \leq k_i d(x, y)$$

para todo $x, y \in X$. Denotamos $k_{\max} = \max_{1 \leq i \leq M} k_i$.

Existen clases especiales de SIF, por ejemplo si en (4.1) se da la igualdad para todo i , se dice que el SIF está compuesto por similitudes contractivas. Exhibiremos ahora algunos ejemplos clásicos de SIF, los cuales utilizaremos luego para generar fractales clásicos.

Ejemplo 4.2 (Cantor). Sea $X = \mathbb{R}$ equipado con la distancia usual $d(x, y) = |x - y|$. Sea $\Phi = \{\phi_1, \phi_2\}$, donde

$$\phi_1(x) = \frac{1}{3}x, \quad \phi_2(x) = \frac{1}{3}x + \frac{2}{3}.$$

Entonces Φ es un SIF formado por 2 similitudes contractivas ($M = 2$) con el mismo coeficiente de contractividad $k = \frac{1}{3}$.

Para los siguientes ejemplos, consideremos $X = \mathbb{R}^2$ equipado con la distancia usual $d((x_1, y_1), (x_2, y_2)) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$.

Ejemplo 4.3 (Triángulo de Sierpinski). Sea $\Phi = \{\phi_1, \phi_2, \phi_3\}$, con

$$\phi_1(x, y) = \frac{1}{2}(x, y), \quad \phi_2(x, y) = \frac{1}{2}(x + 1, y), \quad \phi_3(x, y) = \frac{1}{2}\left(x + \frac{1}{2}, y + \frac{\sqrt{3}}{2}\right).$$

En este caso Φ es un SIF formado por 3 similitudes contractivas ($M = 3$) con el mismo coeficiente de contractividad $k = \frac{1}{2}$.

Ejemplo 4.4 (Carpeta de Sierpinski). Consideremos $\Phi = \{\phi_1, \dots, \phi_8\}$, con

$$\phi_1(x, y) = \frac{1}{3}(x, y), \quad \phi_2(x, y) = \frac{1}{3}(x + 1, y), \quad \phi_3(x, y) = \frac{1}{3}(x + 2, y),$$

$$\phi_4(x, y) = \frac{1}{3}(x, y + 1), \quad \phi_5(x, y) = \frac{1}{3}(x + 2, y + 1),$$

$$\phi_6(x, y) = \frac{1}{3}(x, y + 2), \quad \phi_7(x, y) = \frac{1}{3}(x + 1, y + 2), \quad \phi_8(x, y) = \frac{1}{3}(x + 2, y + 2).$$

Aquí tenemos $M = 8$ y el mismo factor de contractividad $k = \frac{1}{3}$ para cada función.

Ejemplo 4.5 (Curva de Koch). Sea $\Phi = \{\phi_1, \phi_2, \phi_3, \phi_4\}$, con

$$\phi_1(x, y) = \frac{1}{3}(x, y),$$

$$\phi_2(x, y) = \frac{1}{3}(x \cos(\pi/3) - y \sin(\pi/3) + 1, x \sin(\pi/3) + y \cos(\pi/3)),$$

$$\phi_3(x, y) = \frac{1}{3} \left(x \cos(\pi/3) - y \sin(\pi/3) + \frac{1}{2}, -x \sin(\pi/3) + y \cos(\pi/3) + \frac{\sqrt{3}}{2} \right),$$

$$\phi_4(x, y) = \frac{1}{3}(x + 2, y).$$

Ahora tenemos $M = 4$ y nuevamente el mismo factor de contractividad $k = \frac{1}{3}$ para cada función.

Notar que en los últimos 3 ejemplos, cada transformación $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ que compone el SIF tiene la forma

$$\phi(x, y) = A(x, y) + (x_0, y_0),$$

siendo A una matriz de 2×2 . Una función de este tipo es llamada **transformación afín** en el plano. Observar que se componen de una transformación lineal, realizada por la matriz A , seguida de una traslación producida por el vector (x_0, y_0) . Una transformación afín puede resultar o no contractiva, dependiendo de la matriz A , la cual puede producir escalamiento (dilataciones o contracciones), simetrías o rotaciones, entre otras cosas. Por ejemplo, la transformación ϕ_2 en el Ejemplo 4.5 produce una contracción de factor $1/3$ (homotecia con centro el origen y razón $1/3$), seguida de un giro con centro en el origen y ángulo $\pi/3$, y finalmente una traslación de vector $(1/3, 0)$. En forma matricial, nos queda:

$$\phi_2(x, y) = \begin{pmatrix} \cos(\pi/3) & -\sin(\pi/3) \\ \sin(\pi/3) & \cos(\pi/3) \end{pmatrix} \begin{pmatrix} 1/3 & 0 \\ 0 & 1/3 \end{pmatrix} (x, y) + (1/3, 0).$$

Luego, en este caso nos queda $A = \frac{1}{3} \begin{pmatrix} \cos(\pi/3) & -\sin(\pi/3) \\ \sin(\pi/3) & \cos(\pi/3) \end{pmatrix}$ y $(x_0, y_0) = (1/3, 0)$.

Se deja como ejercicio hallar la matriz A y el vector (x_0, y_0) en el resto de los ejemplos.

Puede probarse que una condición necesaria y suficiente para que una transformación afín resulte contractiva es que todos sus autovalores tengan módulo menor que la unidad.

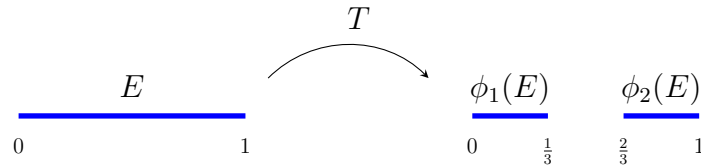
Dado un SIF $\Phi = \{\phi_1, \dots, \phi_M\}$, consideraremos la función T definida sobre \mathcal{K} como

$$T(E) = \bigcup_{i=1}^M \phi_i(E),$$

para cada E en \mathcal{K} . Esta función es llamada **transformación inducida por el SIF**.

Ejemplo 4.6. Sea T la transformación inducida por el SIF de Cantor dado en el Ejemplo 4.2, y consideremos $E = [0, 1]$. Entonces

$$T(E) = \phi_1(E) \cup \phi_2(E) = [0, 1/3] \cup [2/3, 1].$$



A los lectores familiarizados con el **conjunto ternario de Cantor**, lo anterior seguramente les recuerda al primer paso para la construcción del mismo. Repasemos su construcción geométrica. Se comienza con el intervalo real $[0, 1]$ y se elimina el segmento abierto correspondiente al tercio central del mismo, es decir, se quita el conjunto $(1/3, 2/3)$. Se procede recursivamente eliminando en cada paso el segmento abierto central de cada intervalo restante. Más precisamente, los pasos para la construcción del conjunto de Cantor son los siguientes:

1. Tomar el intervalo $[0, 1]$.
2. Dividir en 3 intervalos iguales y eliminar el intervalo abierto central.
3. Aplicar el paso anterior recursivamente a cada uno de los intervalos restantes.

El conjunto de Cantor se obtiene repitiendo el proceso anterior *infinitas* veces (ver Figura 2). Observar que cada paso de la construcción se compone de 2 copias del paso anterior, reducidas un tercio de su tamaño.

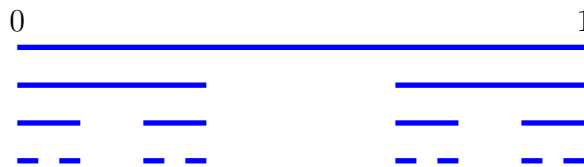


FIGURA 2. Primeras cuatro etapas de la construcción del conjunto de Cantor.

El conjunto de Cantor es un objeto matemático con muchas propiedades interesantes, como el hecho de tener medida cero, ser no numerable, ser cerrado (y por estar incluido en el $[0, 1]$ resulta además acotado, y por lo tanto compacto), entre otras. Pero la propiedad que nos interesa en este curso es la *autosimilaridad*. Un objeto **autosimilar** es aquel que puede ser visto como una unión finita de copias de sí mismo en diferentes escalas. Luego, si cambiamos la escala para ver con mayor detalle el objeto, cualquiera sea el grado de precisión elegido, nos encontramos con el mismo patrón que rige al todo. Mandelbrot propuso el término **fractal** para describir este tipo de objetos.

Volviendo al conjunto de Cantor, notar que puede verse como la unión de 2 copias de sí mismo reducidas a un tercio de su tamaño: una comenzando en el origen y la otra trasladada $2/3$ hacia la derecha. Más precisamente, llamando C al conjunto de Cantor tenemos que

$$(4.2) \quad C = \phi_1(C) \cup \phi_2(C),$$

siendo $\phi_1(x) = \frac{1}{3}x$ y $\phi_2(x) = \frac{1}{3}x + \frac{2}{3}$ las similitudes contractivas mencionadas en el Ejemplo 4.2. Reescribiendo (4.2) en términos de la transformación T inducida por el SIF $\Phi = \{\phi_1, \phi_2\}$, obtenemos $C = T(C)$. En palabras, *el conjunto de Cantor es un punto fijo para la transformación T inducida por el SIF.*

Esto nos conduce a plantearnos el problema en forma general: dado un SIF $\Phi = \{\phi_1, \dots, \phi_M\}$ y su correspondiente transformación inducida T , nos preguntamos:

- ¿existirá siempre un conjunto $K \in \mathcal{K}$ tal que $K = T(K)$? Es decir, ¿existirá un conjunto compacto no vacío K satisfaciendo $K = \bigcup_{i=1}^M \phi_i(K)$? Un conjunto que satisface esta igualdad es llamado **conjunto invariante por el SIF** (pues es un punto fijo de T) o **atractor** del SIF. En caso que cada una de las transformaciones que conforman el SIF sea una similitud contractiva, entonces el atractor es llamado también **conjunto autosimilar**.
- En caso que el atractor exista, ¿es único? ¿cómo se construye?

Las respuestas a estas preguntas se obtendrán utilizando el siguiente resultado, y serán enunciadas en la Sección 5.

Proposición 4.7. *Dado un sistema iterado de funciones $\Phi = \{\phi_1, \dots, \phi_M\}$ en un espacio métrico (X, d) , la transformación T inducida por Φ mapea \mathcal{K} en sí mismo y es una contracción de razón $k_{\max} = \max_{1 \leq i \leq M} k_i$.*

Demostración. Sea $E \in \mathcal{K}$. Para cada i fijo, ϕ_i es una función continua (por ser contractiva), por lo que $\phi_i(E)$ es un conjunto compacto. Luego $T(E)$ es compacto, por ser unión finita de conjuntos compactos. Por lo tanto $T : \mathcal{K} \rightarrow \mathcal{K}$.

Probaremos ahora que T es una aplicación contractiva en (\mathcal{K}, d_H) con coeficiente de contractividad $k = k_{\max}$, es decir, veremos que

$$d_H(T(E), T(F)) \leq k_{\max} d_H(E, F),$$

para todo $E, F \in \mathcal{K}$. En efecto, fijemos E y F en \mathcal{K} y, para cada $i = 1, \dots, M$ fijo, denotemos $E_i := \phi_i(E)$ y $F_i := \phi_i(F)$. Notar que si $E \subseteq [F]_\varepsilon$ y $F \subseteq [E]_\varepsilon$, entonces $E_i \subseteq [F_i]_{\varepsilon k_i}$ y $F_i \subseteq [E_i]_{\varepsilon k_i}$ para cada $i = 1, \dots, M$ (ver Ejercicio 4 de esta sección). Luego

$$T(E) = \bigcup_{i=1}^M E_i \subseteq \bigcup_{i=1}^M [F_i]_{\varepsilon k_i} \subseteq \left[\bigcup_{i=1}^M F_i \right]_{\varepsilon k_{\max}} = [T(F)]_{\varepsilon k_{\max}},$$

y

$$T(F) = \bigcup_{i=1}^M F_i \subseteq \bigcup_{i=1}^M [E_i]_{\varepsilon k_i} \subseteq \left[\bigcup_{i=1}^M E_i \right]_{\varepsilon k_{\max}} = [T(E)]_{\varepsilon k_{\max}}.$$

Por lo tanto

$$d_H(T(E), T(F)) \leq k_{\max} d_H(E, F).$$

□

Ejercicios.

1. Hallar la matriz A y el vector (x_0, y_0) en cada una de las transformaciones que aparecen en los Ejemplos 4.3, 4.4 y 4.5.
2. Probar que cada una de las funciones que componen los sistemas en los Ejemplos 4.3, 4.4 y 4.5, son transformaciones contractivas.

3. Sean $A = \{(0,0)\}$, B el segmento en \mathbb{R}^2 con vértices en $(0,0)$ y $(1,0)$, y C el cuadrado unitario en \mathbb{R}^2 . Aplicar a los conjuntos A , B y C cada una de las transformaciones que conforman los siguientes SIF. Graficar el resultado obtenido:
 - a) el SIF del Ejemplo 4.3;
 - b) el SIF del Ejemplo 4.4;
 - c) el SIF del Ejemplo 4.5.
4. Sea $\phi : X \rightarrow X$ una aplicación contractiva con factor de contracción k . Sean E y F dos subconjuntos de X tales que $E \subseteq [F]_\varepsilon$ para algún $\varepsilon > 0$. Probar que $\phi(E) \subseteq [\phi(F)]_{\varepsilon k}$.

5. APROXIMACIONES A FRACTALES

La Proposición 4.7, el teorema del punto fijo de Banach y el Teorema 2.6 serán herramientas fundamentales para responder las preguntas planteadas en la sección anterior. Al responderlas obtendremos además aproximaciones a fractales clásicos diferentes a las usuales, en el sentido de la distancia de Hausdorff. Enunciamos a continuación dichas respuestas, resumiendo a la vez los resultados contenidos en este curso que fueron necesarios para obtenerlas.

Dado un SIF $\Phi = \{\phi_1, \dots, \phi_M\}$ sobre un espacio métrico (X, d) , la transformación T inducida resulta una aplicación contractiva sobre \mathcal{K} con respecto a la métrica d_H . Si el espacio métrico (X, d) dado es completo, entonces (\mathcal{K}, d_H) también es un espacio métrico completo. Luego podemos aplicar el teorema del punto fijo de Banach para concluir que T tiene un único punto fijo K en \mathcal{K} . Es decir, existe un único subconjunto compacto y no vacío de X que satisface

$$K = \bigcup_{i=1}^M \phi_i(K).$$

Además, de la demostración del teorema del punto fijo se tiene que este conjunto K , llamado atractor del SIF, puede obtenerse como límite con respecto a la métrica d_H de la sucesión $E_n := T^n(E)$, donde T^n denota la composición de T consigo misma n veces, y E es *cualquier* conjunto compacto no vacío en X . De la definición de T , tenemos que

$$E_n = \bigcup_{i_1, \dots, i_n \in \{1, \dots, M\}} \phi_{i_1, i_2, \dots, i_n}(E),$$

donde $\phi_{i_1, i_2, \dots, i_n}(E)$ denota la composición $(\phi_{i_1} \circ \phi_{i_2} \circ \dots \circ \phi_{i_n})(E)$.

Ejemplo 5.1. Para ilustrar lo anterior, retomemos el conjunto de Cantor y lo visto en el Ejemplo 4.6. Allí teníamos $\Phi = \{\phi_1, \phi_2\}$ con $\phi_1(x) = x/3$ y $\phi_2(x) = x/3 + 2/3$ y $E = [0, 1]$. Vimos en dicho ejemplo la imagen de E mediante la transformación T inducida por Φ , es decir, obtuvimos $E_1 = T(E) = \phi_1(E) \cup \phi_2(E)$. Analicemos ahora E_2 y E_3 también.

$$\begin{aligned} E_2 &= T(T(E)) = T(E_1) = T(\phi_1(E) \cup \phi_2(E)) \\ &= \phi_1(\phi_1(E)) \cup \phi_1(\phi_2(E)) \cup \phi_2(\phi_1(E)) \cup \phi_2(\phi_2(E)) \\ &= \phi_1([0, 1/3]) \cup \phi_1([2/3, 1]) \cup \phi_2([0, 1/3]) \cup \phi_2([2/3, 1]) \\ &= [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]. \end{aligned}$$

Análogamente,

$$\begin{aligned}
 E_3 &= T(T^2(E)) = T(E_2) = \phi_1(E_2) \cup \phi_2(E_2) \\
 &= [0, 1/27] \cup [2/27, 1/9] \cup [2/27, 7/27] \cup [8/27, 1/3] \\
 &\cup [2/3, 19/27] \cup [20/27, 7/9] \cup [8/9, 25/27] \cup [26/27, 1].
 \end{aligned}$$

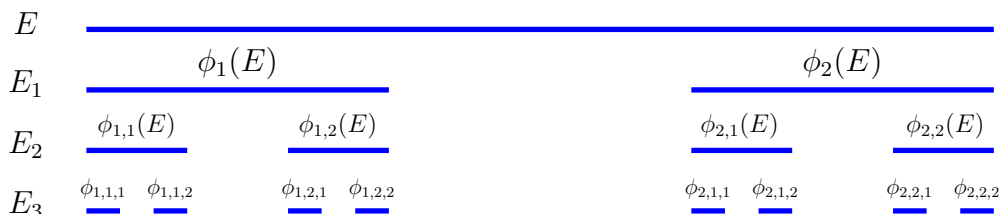


FIGURA 3. Aproximaciones al conjunto de Cantor, partiendo de $E = [0, 1]$.

Observar que de esta manera obtenemos las aproximaciones usuales al Conjunto de Cantor (ver Figura 3 y comparar con la Figura 2). Para resolver ciertos problemas en el fractal límite, a veces es posible resolver un problema equivalente en los conjuntos aproximantes, y tratar luego de extender las conclusiones obtenidas al conjunto límite. Para ello puede resultar de gran utilidad que los conjuntos aproximantes sean lo más simples posible, o que tengan una forma particular. Como veremos a continuación, esto puede obtenerse tomando como conjunto de partida un compacto distinto al intervalo $[0, 1]$. Por ejemplo, uno podría considerar que una sucesión de conjuntos aproximantes es “simple” si cada uno de sus elementos es un conjunto con una cantidad finita de puntos. Esto es lo que se obtiene partiendo de un conjunto con un solo punto, digamos $E = \{0\}$. Es fácil ver que cada elemento E_n de la sucesión aproximante generada a partir de E es un conjunto formado por 2^n puntos, correspondientes a los extremos derechos de los intervalos del nivel n de la construcción usual del conjunto de Cantor. A continuación exhibimos los tres primeros de ellos, los cuales se encuentran representados gráficamente en la Figura 4.

$$E_1 = \left\{ 0, \frac{2}{3} \right\}; \quad E_2 = \left\{ 0, \frac{2}{9}, \frac{2}{3}, \frac{8}{9} \right\}; \quad E_3 = \left\{ 0, \frac{2}{27}, \frac{2}{9}, \frac{8}{27}, \frac{2}{3}, \frac{20}{27}, \frac{8}{9}, \frac{26}{27} \right\}.$$

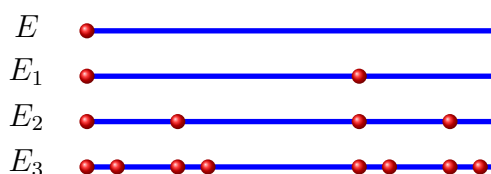


FIGURA 4. Aproximaciones al conjunto de Cantor, partiendo de $E = \{0\}$.

Puesto que nuestro espacio métrico $(\mathbb{R}, |\cdot|)$ es completo, hemos construido de esta manera una sucesión $\{E_n\}$ de conjuntos finitos que converge al conjunto de Cantor C con respecto a la distancia de Hausdorff. Es decir, $d_H(C, E_n)$ tiende a cero cuando n tiende a infinito. Esto implica que dado cualquier $\varepsilon > 0$ existe un número natural $N = N(\varepsilon)$ tal que $C \subseteq [E_N]_\varepsilon$. En palabras, existe un conjunto formado por una cantidad finita de puntos, tal que si lo “engordamos” un poco (poco= ε) contiene al

conjunto de Cantor. Notar que, puesto que nuestro contexto aquí es \mathbb{R} , el engordado de un punto es simplemente un intervalo con centro en él, es decir, $[\{x\}]_\varepsilon = (x - \varepsilon, x + \varepsilon)$. Es claro que mientras menor sea ε , mayor será el cardinal del conjunto finito al cual debemos engordar para contener al conjunto de Cantor. De hecho, de la Observación 3.6 se tiene que $d_H(C, E_n) \leq 3^{-n}$ (ver Ejercicio 1 en esta sección), por lo que para obtener $d_H(E_n, C) < \varepsilon$ será suficiente con tomar $n > \log_3(1/\varepsilon)$.

De esta manera, partiendo de diferentes conjuntos iniciales obtenemos diferentes sucesiones aproximantes, todas convergentes con respecto a la distancia de Hausdorff al mismo fractal límite. Para finalizar ilustraremos esto en otros dos fractales clásicos, comenzando por su construcción geométrica, la cual nos provee de las aproximaciones usuales a los mismos, y dando luego su construcción como límite de la sucesión generada por la transformación inducida por un SIF adecuado.

Ejemplo 5.2. La construcción geométrica del **triángulo de Sierpinski** se define de forma recursiva como sigue:

1. Se comienza con un triángulo equilátero E_0 , digamos de lado 1. Este paso se denomina “etapa cero” de la construcción.
2. Se conectan los puntos medios de cada lado, quedando así 4 triángulos equiláteros de lado $1/2$.
3. Se elimina el triángulo central. Lo que queda se denomina la “primera etapa” de la construcción y la denotamos E_1 .
4. Se repite el proceso anterior a cada uno de los triángulos restantes.

El triángulo de Sierpinski se obtiene repitiendo el proceso anterior infinitas veces. Observar que en cada paso el triángulo de Sierpinski se forma con tres copias del paso anterior, reducidas a la mitad de su tamaño.

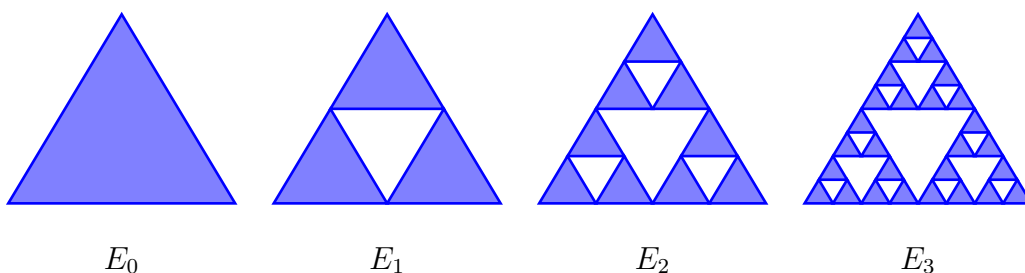


FIGURA 5. Primeras etapas de la construcción del triángulo de Sierpinski.

Notar que el proceso anterior puede reescribirse en términos del SIF Φ considerado en el Ejemplo 4.3. Más precisamente, para cada número natural n se tiene que $E_n = T^n(E_0)$, siendo T la transformación inducida por Φ , y E_0 el triángulo en el plano con vértices en $(0, 0)$, $(1, 0)$ y $(1/2, \sqrt{3}/2)$. Por lo visto anteriormente, podemos obtener diferentes aproximaciones al triángulo de Sierpinski tomando diferentes conjuntos de partida. Por ejemplo, si tomamos E_0 como el cuadrado unitario en \mathbb{R}^2 , obtenemos la siguiente sucesión de conjuntos aproximantes:

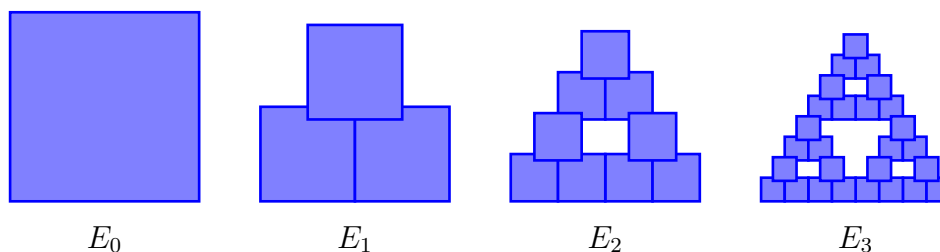


FIGURA 6. Aproximaciones al triángulo de Sierpinski, partiendo de $E_0 = [0, 1]^2$.

Esta aproximación puede considerarse *exterior*, en el sentido que cada elemento de la sucesión contiene al conjunto límite. En el Ejercicio 2 se considera una aproximación *interior*, puesto que cada elemento de la sucesión está contenido en el atractor.

Ejemplo 5.3. La construcción geométrica de la **alfombra (o carpeta) de Sierpinski** se define de forma recursiva de la siguiente forma:

1. Se comienza con un cuadrado.
2. El cuadrado se divide en 9 cuadrados iguales, y se elimina el cuadrado central.
3. El paso anterior vuelve a aplicarse recursivamente a cada uno de los cuadrados restantes.

La alfombra de Sierpinski es el límite de este proceso tras un número infinito de iteraciones. Luego de tres iteraciones se obtiene el conjunto ilustrado en la Figura 7. Al igual que antes, el proceso iterativo aplicado para obtener la carpeta de Sierpinski

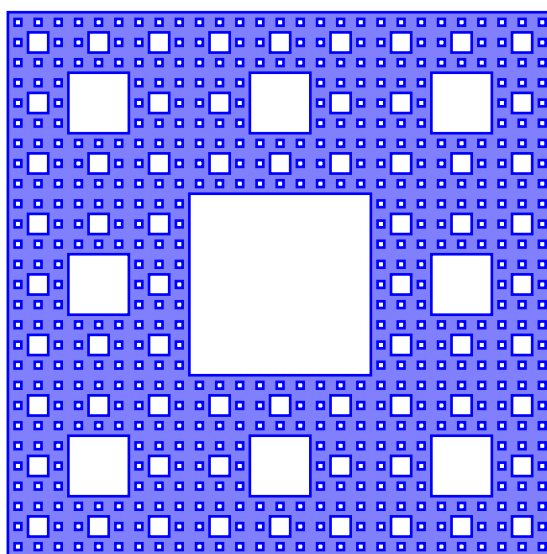


FIGURA 7. Carpeta de Sierpinski.

puede reescribirse en términos de la aplicación inducida por un SIF. En este caso, el sistema Φ es el presentado en el Ejemplo 4.4. Si T denota la transformación inducida por dicho sistema Φ , entonces en la Figura 7 se representa $C_3 = T^3(C_0)$, siendo $C_0 = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1, 0 \leq y \leq 1\}$ el cuadrado unitario en el plano. Podremos obtener diferentes aproximaciones a la carpeta de Sierpinski si partimos de un compacto distinto a C_0 (ver Ejercicio 3).

Ejercicios.

1. Sean T la transformación inducida por el SIF Φ del Ejemplo 4.2, $E_0 = \{0\}$ y $E_n = T^n(E_0)$. Probar que $d_H(C, E_n) \leq 3^{-n}$ para todo n .
2. Sea T la transformación inducida por el SIF Φ del Ejemplo 4.3. Graficar $T^n(E_0)$ para $n = 1, 2, 3$, siendo $E_0 = \{(0, 0)\}$. Hallar n tal que $d_H(E_n, S) < 10^{-2}$, siendo S el triángulo de Sierpinski y $E_n = T^n(E_0)$.
3. Sea T la transformación inducida por el SIF Φ del Ejemplo 4.4. Graficar $T^n(E_0)$ para $n = 1, 2, 3$, siendo
 - a) $E_0 = \{(0, 0)\}$;
 - b) $E_0 = \{(x, 0) : 0 \leq x \leq 1\}$;
 - c) E_0 el triángulo con vértices $(0, 0)$, $(1, 0)$ y $(1/2, 1)$.
 Estimar $d_H(S^*, E_3)$ en cada caso, siendo S^* la carpeta de Sierpinski.
4. Sea T la transformación inducida por el SIF Φ del Ejemplo 4.5, y sea $E_0 = \{(x, 0) : 0 \leq x \leq 1\}$. En la Figura 8 se ilustra $T^3(E_0)$. Graficar $T(E_0)$ y $T^2(E_0)$.

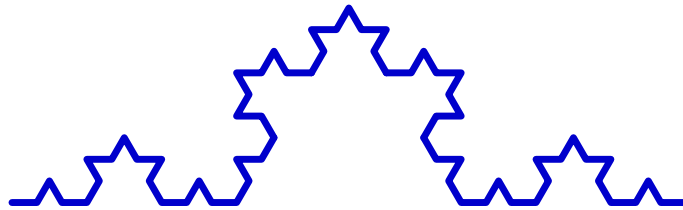


FIGURA 8. Curva de Koch

REFERENCIAS

- [1] George Bachman and Lawrence Narici. *Functional analysis*. Dover Publications Inc., Mineola, NY, 2000. Reprint of the 1966 original.
- [2] Michael F. Barnsley, *Fractals everywhere*. second ed., Academic Press Professional, Boston, MA, 1993, Revised with the assistance of and with a foreword by Hawley Rising, III.
- [3] Robert G. Bartle. *Introducción al análisis matemático*. Editorial Limusa, México, 1982. Translated from the English by Ma. Cristina Gutiérrez González.
- [4] Gerald B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons Inc., New York, 1999.
- [5] John E. Hutchinson. Fractals and self-similarity. *Indiana Univ. Math. J.*, 30(5):713–747, 1981.
- [6] Walter Rudin. *Principles of mathematical analysis*. McGraw-Hill Book Co., New York, third edition, 1976. International Series in Pure and Applied Mathematics.

INSTITUTO DE MATEMÁTICA APLICADA DEL LITORAL (CONICET-UNL), DEPARTAMENTO DE MATEMÁTICA (FHUC-UNL), SANTA FE, ARGENTINA.

E-mail address: mcarena@santafe-conicet.gov.ar

DINÁMICA SIMBÓLICA

CAMILO JADUR Y JORGE YAZLLE

RESUMEN. Un Sistema Dinámico es un par (X, T) donde X es un espacio métrico compacto y T una transformación continua de X en X . La Dinámica Topológica estudia las iteraciones de estas transformaciones. Los conceptos básicos en Dinámica Topológica son, entre otros, Minimalidad, Transitividad, Recurrencia, Estabilidad, Sensitividad, atractores y entropía topológica. La Dinámica Simbólica estudia los Sistemas Dinámicos cuyo espacio de configuraciones consiste de sucesiones infinitas de símbolos tomados de un alfabeto finito. Las principales clases de Sistemas Dinámicos Simbólicos son: Los Shift de Tipo Finito, los Shift Sóficos y Autómatas Celulares. Se dará una introducción a estos tópicos destacando los resultados principales.

ÍNDICE

| | |
|--|-----|
| 1. Espacios shift | 81 |
| 1.1. Definiciones básicas | 81 |
| 1.2. Full shifts | 82 |
| 1.3. Los Espacios Shift | 84 |
| 1.4. Lenguajes | 86 |
| 1.5. Códigos de ventana deslizante | 88 |
| 2. Shifts de tipo finito | 94 |
| 2.1. Restricciones de tipo finito | 94 |
| 2.2. Grafos y sus shifts | 97 |
| 2.3. Representación de shifts de tipo finito por medio de grafos | 103 |
| 3. Aspectos topológicos y dinámicos de los espacios shift | 105 |
| 3.1. Introducción | 105 |
| 3.2. Una métrica para $\mathcal{A}^{\mathbb{Z}}$ | 106 |
| 3.3. Sucesiones | 108 |
| 3.4. Cilindros. Conexidad | 110 |
| 3.5. Continuidad | 112 |
| 3.6. Sistemas dinámicos | 113 |
| Referencias | 116 |

1. ESPACIOS SHIFT

1.1. Definiciones básicas. Un **alfabeto** es un conjunto finito, cuyos elementos se denominan **símbolos**, o también **letras**. Un **bloque**, o **palabra**, sobre un alfabeto \mathcal{A} es una sucesión finita (posiblemente vacía) de letras de \mathcal{A} . Formalmente, una palabra u sobre \mathcal{A} es una función u del conjunto $\{x \in \mathbb{N} : 1 \leq x \leq n\}$ en \mathcal{A} (para algún entero $n \geq 0$). En el caso $n = 0$, se tiene la función vacía, o sucesión vacía, que llamaremos **palabra vacía** y denotaremos por ε . La **longitud** (o **largo**, o **tamaño**) de la palabra es n , es decir, la cantidad de términos de la sucesión, y se denota por $|u|$. La palabra

vacía tiene largo 0. Para una palabra no vacía u , el k -ésimo término de la sucesión se denota, como es costumbre, mediante u_k . Entonces, una palabra no vacía de largo n sobre \mathcal{A} es $u_1u_2\cdots u_n$, con $u_i \in \mathcal{A}$ para todo $i \in \{1, \dots, n\}$. Un bloque de longitud n se llama un **n -bloque**. Dos palabras u y v son **iguales** si tienen el mismo largo n y, para todo j entre 1 y n , es $u_j = v_j$.

El conjunto de todas las palabras de largo n sobre un alfabeto \mathcal{A} se designa por \mathcal{A}^n , y el conjunto de todas las palabras sobre \mathcal{A} se designa por \mathcal{A}^* . Es decir, $\mathcal{A}^* = \bigcup_{n=0}^{\infty} \mathcal{A}^n$. Notar que, para todo $n \geq 0$, \mathcal{A}^n es un conjunto finito, en tanto que \mathcal{A}^* es un conjunto que tiene una cantidad infinita de elementos.

Un alfabeto que a menudo usaremos es $\mathcal{A} = \{0, 1\}$, y las palabras que sobre él pueden formarse se llaman **sucesiones binarias**. Tenemos que, en este caso, es $\mathcal{A}^0 = \{\varepsilon\}$, $\mathcal{A}^1 = \{0, 1\}$, $\mathcal{A}^2 = \{00, 01, 10, 11\}$, $\mathcal{A}^3 = \{000, 001, 010, 011, 100, 101, 110, 111\}$, etc. Además, $\mathcal{A}^* = \{\varepsilon, 0, 1, 00, 01, 10, 11, \dots\}$.

Sea u un bloque. Un **subbloque** (o **subpalabra**) de u es una palabra w tal que existe $i \geq 1$ tal que $\forall k \in \mathbb{Z}, 1 \leq k \leq |w| \Rightarrow w_k = u_{i+k-1}$. De la definición, se desprende que la palabra vacía es subbloque de cualquier palabra, y que si $u = u_1 \cdots u_n$ es un n -bloque no vacío, sus subpalabras no vacías son de la forma $u_i \cdots u_j$, en donde i, j son enteros tales que $1 \leq i \leq j \leq n$. Si w es subbloque de u , escribimos $w \sqsubseteq u$, y decimos que w **ocurre en** u . Obviamente, $w \sqsubseteq u \Rightarrow |w| \leq |u|$.

Si u y v son dos bloques sobre \mathcal{A} , la **concatenación** de u y v es la nueva palabra que se obtiene al escribir los símbolos de u y, a continuación, los de v , sin ningún signo especial intermedio. La concatenación de u y v se escribe uv . Notar que, en general, $uv \neq vu$. Es directo ver que $|uv| = |u| + |v|$. De manera análoga, la concatenación de tres o más bloques es la palabra que se obtiene de escribir los respectivos símbolos consecutivamente. La concatenación de un bloque u consigo mismo una cantidad n de veces se designa por u^n . Formalmente:

$$u^n = \begin{cases} \varepsilon & \text{si } n = 0, \\ uu^{n-1} & \text{si } n > 0. \end{cases}$$

1.2. Full shifts. Dado un alfabeto \mathcal{A} , se denomina **full shift sobre el alfabeto \mathcal{A}** , o **full \mathcal{A} -shift**, al conjunto de todas las funciones de \mathbb{Z} en \mathcal{A} . El full shift sobre \mathcal{A} se denota por $\mathcal{A}^{\mathbb{Z}}$.

Una función de \mathbb{Z} en \mathcal{A} no es más que una sucesión bi-infinita de símbolos de \mathcal{A} , que puede escribirse así:

$$x = (x_n)_{n \in \mathbb{Z}} = \cdots x_{-2}x_{-1}x_0x_1x_2x_3 \cdots$$

Un ejemplo de tales sucesiones bi-infinitas para $\mathcal{A} = \{0, 1\}$ es la que tiene 0 en sus términos impares, y 1 en sus pares, es decir, $\cdots 010101010 \cdots$.

Para tener una idea de cuál es la “coordenada 0” de x , escribiremos un signo de puntuación a la izquierda de x_0 , así:

$$\cdots x_{-2}x_{-1}.x_0x_1x_2 \cdots$$

Los elementos de $\mathcal{A}^{\mathbb{Z}}$ reciben el nombre de **puntos**. Cada punto es, entonces, una sucesión bi-infinita, o tirilla bi-infinita, de símbolos. Dos puntos $x = (x_i)_{i \in \mathbb{Z}}$ e $y = (y_i)_{i \in \mathbb{Z}}$ son iguales si, y sólo si, $x_i = y_i$ para todo entero i .

Notar que $\mathcal{A}^{\mathbb{Z}}$ no contiene palabras, sino sucesiones bi-infinitas. \mathcal{A}^* y $\mathcal{A}^{\mathbb{Z}}$ no tienen ningún elemento en común.

Dado $x = \cdots x_{-2}x_{-1}x_0x_1x_2x_3 \cdots \in \mathcal{A}^{\mathbb{Z}}$, designamos por $x_{[i,j]}$ a la sucesión (finita) de los símbolos de x que van desde la coordenada i hasta la coordenada j , ambas

inclusive. Es decir, $x_{[i,j]} = x_i x_{i+1} \cdots x_{j-1} x_j$. Adoptamos la convención de que para $i > j$, $x_{[i,j]} = \varepsilon$ (formalmente, $x_{[i,j]}$ es la función u con dominio $\{n \in \mathbb{Z} : 1 \leq n \leq j - i + 1\}$ y codominio \mathcal{A} tal que $u_n = x_{i+n-1}$). Denotamos por $x_{[i,\infty)}$ a la sucesión infinita a derecha $x_i x_{i+1} \cdots$, mientras que $x_{(-\infty,i]}$ es la sucesión infinita a izquierda $\cdots x_{i-1} x_i$. Dados $u \in \mathcal{A}^*$ y $x \in \mathcal{A}^{\mathbb{Z}}$, decimos que u **ocurre en** x , o también que u **es palabra o bloque de** x (y lo escribimos $u \sqsubseteq x$) si existen enteros i, j tales que $x_{[i,j]} = u$. Notar que la palabra vacía ocurre en cualquier punto del full shift: $\forall x \in \mathcal{A}^{\mathbb{Z}}, x_{[1,0]} = \varepsilon$. El $(2k+1)$ -**bloque central** de x es $x_{[-k,k]}$.

Para un bloque no vacío $u = u_1 \cdots u_n \in \mathcal{A}^* - \{\varepsilon\}$, designamos por u^∞ al punto $x \in \mathcal{A}^{\mathbb{Z}}$ que resulta de la concatenación infinita (hacia ambos lados) de la palabra u consigo misma de modo que $\forall i \in \{0, \dots, n-1\}, x_i = u_{i+1}$:

$$u^\infty = \cdots u u u u u . u u u u u \cdots = \cdots u_1 \cdots u_n u_1 \cdots u_n . u_1 \cdots u_n u_1 \cdots u_n \cdots .$$

Obsérvese que la coordenada 0 es u_1 . Por ejemplo, si $u = 011$, entonces

$$u^\infty = \cdots 011011.011011011 \cdots ,$$

que no es el mismo que

$$\cdots 0110110.11011011 \cdots = (110)^\infty .$$

1.2.1. La transformación shift. Se puede pensar en el subíndice i de una sucesión $(x_i)_{i \in \mathbb{Z}}$ como indicador del tiempo, de modo que x_i representa el valor de la sucesión en el minuto i . El paso de una unidad de tiempo equivale entonces a desplazar (“shift”) cada coordenada de la sucesión un lugar hacia la izquierda. Esto define una transformación natural de $\mathcal{A}^{\mathbb{Z}}$ en $\mathcal{A}^{\mathbb{Z}}$, llamada **transformación shift** y denotada por σ , del siguiente modo:

$$\begin{aligned} \sigma & : \mathcal{A}^{\mathbb{Z}} \rightarrow \mathcal{A}^{\mathbb{Z}} \\ x & \mapsto y = \sigma(x) \text{ con } y_i = x_{i+1}. \end{aligned}$$

Por ejemplo, si $x = (00111)^\infty = \cdots 00111.0011100111 \cdots$, entonces

$$\sigma(x) = \cdots 001110.011100111 \cdots .$$

Conforme a notación standard en teoría de sistemas dinámicos, usaremos σx para $\sigma(x)$, y σx_i para $(\sigma(x))_i$.

σ es una función biyectiva. Su inversa, σ^{-1} , desplaza cada posición de una sucesión un lugar hacia la derecha. Si k es un entero positivo, σ^k designa a la composición de σ consigo misma k veces, y produce el efecto de desplazar todas las coordenadas de una sucesión k lugares a la izquierda ($\sigma^k x_i = x_{i+k}$), mientras que σ^{-k} es la composición de σ^{-1} consigo misma k veces, y mueve k lugares a la derecha ($\sigma^{-k} x_i = x_{i-k}$). σ^0 se define como la función identidad. Se puede resumir todo diciendo que $\sigma^k x_i = x_{i+k}$, sea k positivo o no. Nótese que, en consecuencia, $x_i = \sigma^k x_{i-k}$. Por ser composición de biyecciones, σ^k es una biyección de $\mathcal{A}^{\mathbb{Z}}$, para cualquier $k \in \mathbb{Z}$.

Un punto $x \in \mathcal{A}^{\mathbb{Z}}$ para el cual existe un $n \geq 1$ tal que $\sigma^n x = x$, se llama **punto periódico para** σ , y, en ese caso, n es un **período de** x . Un punto fijo para σ es de la forma a^∞ , para algún $a \in \mathcal{A}$. Si x es periódico de período n , también es periódico de período $2n, 3n, \dots$. El menor de los números positivos k tales que $\sigma^k x = x$ se llama **período mínimo de** x .

1.3. Los Espacios Shift.

Definición 1.1. Sea \mathcal{F} una colección de bloques sobre un alfabeto \mathcal{A} , es decir, $\mathcal{F} \subseteq \mathcal{A}^*$. Designamos por $\mathbb{X}_{\mathcal{F}}$ al subconjunto de $\mathcal{A}^{\mathbb{Z}}$ formado por todos aquellos puntos en los que **no ocurre** ningún bloque de \mathcal{F} . Es decir,

$$\begin{aligned}\mathbb{X}_{\mathcal{F}} &= \{x \in \mathcal{A}^{\mathbb{Z}} : \forall f \in \mathcal{F}, f \text{ no ocurre en } x\} \\ &= \{x \in \mathcal{A}^{\mathbb{Z}} : \forall i, j \in \mathbb{Z}, x_{[i,j]} \notin \mathcal{F}\}.\end{aligned}$$

Obsérvese que $x \notin \mathbb{X}_{\mathcal{F}} \Leftrightarrow \exists i, j \in \mathbb{Z} : x_{[i,j]} \in \mathcal{F}$.

A continuación, algunos ejemplos sobre $\mathcal{A} = \{0, 1\}$:

Ejemplo 1.2. Si $\mathcal{F}_1 = \emptyset$, es $\mathbb{X}_{\mathcal{F}_1} = \mathcal{A}^{\mathbb{Z}}$, pues trivialmente se tiene que $\forall x \in \mathcal{A}^{\mathbb{Z}}, \forall i, j \in \mathbb{Z}, x_{[i,j]} \notin \mathcal{F}_1$.

Ejemplo 1.3. Si $\mathcal{F}_2 = \{0, 1\}$, es $\mathbb{X}_{\mathcal{F}_2} = \emptyset$, pues $\forall x \in \mathcal{A}^{\mathbb{Z}}, x_{[0,0]} \in \mathcal{F}_2$.

Ejemplo 1.4. Si $\mathcal{F}_3 = \{\varepsilon\}$, es $\mathbb{X}_{\mathcal{F}_3} = \emptyset$, pues $\forall x \in \mathcal{A}^{\mathbb{Z}}, x_{[1,0]} = \varepsilon \in \mathcal{F}_3$.

Ejemplo 1.5. Si $\mathcal{F}_4 = \{0\}$, es $\mathbb{X}_{\mathcal{F}_4} = \{1^\infty\}$, pues cualquier punto x distinto de 1^∞ posee una coordenada i tal que $x_i = 0$, y entonces $x_{[i,i]} \in \mathcal{F}_4$.

Ejemplo 1.6. Si $\mathcal{F}_5 = \{00, 01\}$, es $\mathbb{X}_{\mathcal{F}_5} = \{1^\infty\}$ pues si un punto x tiene una coordenada $x_i = 0$, es $x_{i+1} = 0$ o $x_{i+1} = 1$, y en cualquier caso $x_{[i,i+1]} \in \mathcal{F}_5$.

Ejemplo 1.7. Si $\mathcal{F}_6 = \{11\}$, algunos puntos de $\mathbb{X}_{\mathcal{F}_6}$ son $(01)^\infty$, $(010)^\infty$, $(00100010)^\infty$, $\dots 0000.10000 \dots$, $\dots 00100000.00000 \dots$, y

$$\dots 00001000100101.01010010001000010000010 \dots$$

$\mathbb{X}_{\mathcal{F}_6}$ es precisamente el conjunto de puntos de $\mathcal{A}^{\mathbb{Z}}$ que no contienen dos 1 consecutivos. Este espacio shift es conocido como el **shift de la razón de oro**.

Ejemplo 1.8. Si $\mathcal{F}_7 = \{1^n : n \geq 2\}$, es $\mathbb{X}_{\mathcal{F}_7} = \mathbb{X}_{\mathcal{F}_6}$.

Ejemplo 1.9. Si $\mathcal{F}_8 = \mathcal{A}^*$, es $\mathbb{X}_{\mathcal{F}_8} = \emptyset$.

Ejemplo 1.10. Si $\mathcal{F}_9 = \{10^{2n+1}1 : n \in \mathbb{N}\}$, $\mathbb{X}_{\mathcal{F}_9}$ es precisamente el conjunto de puntos de $\mathcal{A}^{\mathbb{Z}}$ tales que entre dos ocurrencias consecutivas de 1 hay una cantidad par de 0 (es decir, aquellos puntos $x \in \mathcal{A}^{\mathbb{Z}}$ tales que $10^n 1 \sqsubseteq x \Rightarrow n$ es par). Este espacio shift es conocido como el **shift par**.

Proposición 1.11. Sean \mathcal{F} y \mathcal{F}' subconjuntos de \mathcal{A}^* . Entonces:

1. $\mathbb{X}_{\mathcal{F} \cup \mathcal{F}'} = \mathbb{X}_{\mathcal{F}} \cap \mathbb{X}_{\mathcal{F}'}$,
2. $\mathbb{X}_{\mathcal{F} \cap \mathcal{F}'} \supseteq \mathbb{X}_{\mathcal{F}} \cup \mathbb{X}_{\mathcal{F}'}$,
3. Si $\mathcal{F} \subseteq \mathcal{F}'$, entonces $\mathbb{X}_{\mathcal{F}} \supseteq \mathbb{X}_{\mathcal{F}'}$.

Dado $\mathcal{F} \subseteq \mathcal{A}^*$, de la definición de $\mathbb{X}_{\mathcal{F}}$ resulta claro que $\mathbb{X}_{\mathcal{F}} \subseteq \mathcal{A}^{\mathbb{Z}}$. Ahora bien, dado un subconjunto cualquiera X de $\mathcal{A}^{\mathbb{Z}}$, ¿existirá siempre un $\mathcal{F} \subseteq \mathcal{A}^*$ tal que $X = \mathbb{X}_{\mathcal{F}}$? Resulta que no siempre, y aquellos X para los que sí existe tal \mathcal{F} reciben un nombre especial y son objeto de nuestro estudio.

Definición 1.12. Un **espacio shift** sobre el alfabeto \mathcal{A} es un conjunto $X \subseteq \mathcal{A}^{\mathbb{Z}}$ tal que existe un $\mathcal{F} \subseteq \mathcal{A}^*$ tal que $X = \mathbb{X}_{\mathcal{F}}$.

Pensamos en \mathcal{F} como en un conjunto de **bloques prohibidos** para X .

Antes de ver ejemplos de subconjuntos de $\mathcal{A}^{\mathbb{Z}}$ que son o no espacios shift, veamos una propiedad importante que cumplen los conjuntos que sí lo son.

Proposición 1.13. Sean X un espacio shift, $x \in X$ y $k \in \mathbb{Z}$. Entonces $\sigma^k x$ también pertenece a X .

Corolario 1.14. Sea $X \subseteq \mathcal{A}^{\mathbb{Z}}$, y supongamos que existen $x \in X$ y $k \in \mathbb{Z}$ tales que $\sigma^k x \notin X$. Entonces X no es un espacio shift.

Corolario 1.15. Si X es un espacio shift y $k \in \mathbb{Z}$, entonces $\sigma^k(X) = X$.

La propiedad de ser $\sigma(X) = X$ se llama **invariancia por shift** o **shift invariancia**. Los corolarios anteriores dicen que todo espacio shift es shift invariante, o, igualmente, que un conjunto que no es shift invariante no puede ser un espacio shift. Por ejemplo, el conjunto unitario $X = \{(01)^\infty\}$ no es un espacio shift pues $\sigma(01)^\infty = (10)^\infty \notin X$. Sin embargo, la shift invariancia de un conjunto no garantiza que éste sea un espacio shift.

Ejemplo 1.16. Sea X el conjunto de todas las sucesiones bi-infinitas sobre $\{0, 1\}$ en las que hay exactamente una coordenada 1 y el resto son todas 0. Es decir, $X = \{(x_i)_{i \in \mathbb{Z}} : \exists k \in \mathbb{Z} : (x_k = 1 \wedge \forall i \neq k, x_i = 0)\}$. Si $x \in X$, σx también pertenece a X , es decir, X es invariante por shift. Sin embargo, veamos que X no es espacio shift. Para arribar a una contradicción, supongamos que lo fuera. Quiere decir que existiría un \mathcal{F} tal que $X = \mathbb{X}_{\mathcal{F}}$. Debe ocurrir que $\forall k \geq 1, 0^k \notin \mathcal{F}$ (pues $\forall k \geq 1, 0^k \sqsubseteq \dots 000000.1000000 \dots \in X$). Pero entonces 0^∞ pertenecería a $\mathbb{X}_{\mathcal{F}}$, con lo cual $0^\infty \in X$, que contradice la definición de X . La contradicción proviene de suponer la existencia del conjunto \mathcal{F} , es decir, de suponer que X es un espacio shift.

Ejemplo 1.17. El conjunto vacío y el full shift $\{0, 1\}^{\mathbb{Z}}$ son espacios shift, según se mostró en los ejemplos 1.2 y 1.3. En general, cualquier full shift es un espacio shift. Los conjuntos $\mathbb{X}_{\mathcal{F}_4}$ y $\mathbb{X}_{\mathcal{F}_6}$ de los ejemplos 1.5 y 1.7 son también espacios shift. Los ejemplos 1.5 y 1.6 muestran que un mismo espacio shift puede ser descrito a través de diferentes colecciones de bloques prohibidos.

La restricción de σ a un espacio shift X será denotada por σ_X . En vistas del Corolario 1.15, σ_X es una biyección de X en X . Las definiciones anteriormente dadas para puntos fijos y periódicos de $\mathcal{A}^{\mathbb{Z}}$ para σ son igualmente aplicables a un espacio shift X y su respectiva σ_X .

El siguiente resultado será útil en lo sucesivo, pues afirma que el conjunto de bloques prohibidos para un espacio shift puede verse como constituido por palabras de longitud al menos N , donde N es cualquier entero positivo.

Proposición 1.18. Sean X un espacio shift y N un entero positivo. Entonces, existe $\mathcal{F} \subseteq \mathcal{A}^*$ tal que todo bloque en \mathcal{F} tiene longitud al menos N , y $X = \mathbb{X}_{\mathcal{F}}$.

Demostración. Siendo X un espacio shift, hay un $\mathcal{F}' \subseteq \mathcal{A}^*$ tal que $X = \mathbb{X}_{\mathcal{F}'}$. Hagamos

$$\mathcal{F}_1 = \{w \in \mathcal{A}^N : \exists u \in \mathcal{F}' : u \sqsubseteq w\}, \quad \mathcal{F}_2 = \{u \in \mathcal{F}' : |u| > N\},$$

$$\mathcal{F} = \mathcal{F}_1 \cup \mathcal{F}_2.$$

Es decir, \mathcal{F} consta de todas las palabras de \mathcal{F}' que tengan largo mayor que N , más todas aquellas palabras de largo N sobre el alfabeto que contengan una subpalabra que esté en \mathcal{F}' . Observar que $\mathcal{F}_1 \cap \mathcal{F}_2 = \emptyset$. Es directo ver que todo bloque en \mathcal{F} tiene largo

al menos N . Veamos que $X = \mathbb{X}_{\mathcal{F}}$:

$$\begin{aligned}
x \notin \mathbb{X}_{\mathcal{F}} &\Leftrightarrow \exists i, j \in \mathbb{Z} : x_{[i,j]} \in \mathcal{F} \\
&\Leftrightarrow \exists i, j \in \mathbb{Z} : x_{[i,j]} \in \mathcal{F}_1 \vee x_{[i,j]} \in \mathcal{F}_2 \\
&\Leftrightarrow \exists i, j \in \mathbb{Z} : (j - i + 1 = N \wedge \exists u \in \mathcal{F}' : u \sqsubseteq x_{[i,j]}) \\
&\quad \vee (j - i + 1 > N \wedge x_{[i,j]} \in \mathcal{F}') \\
&\Leftrightarrow \exists u \in \mathcal{F}' : u \sqsubseteq x \\
&\Leftrightarrow x \notin \mathbb{X}_{\mathcal{F}'} = X.
\end{aligned}$$

□

1.4. Lenguajes.

Definición 1.19. Un **lenguaje** sobre un alfabeto \mathcal{A} es cualquier subconjunto de \mathcal{A}^* .

Si X es un subconjunto de $\mathcal{A}^{\mathbb{Z}}$ (no necesariamente un espacio shift) y n un entero no negativo, denotamos por $\mathcal{B}_n(X)$ al conjunto de bloques de longitud n que ocurren en puntos de X , es decir,

$$\mathcal{B}_n(X) = \{u \in \mathcal{A}^n : \exists x \in X : u \sqsubseteq x\}.$$

Por ejemplo, si $X = \mathcal{A}^{\mathbb{Z}}$, entonces $\mathcal{B}_n(X)$ es precisamente \mathcal{A}^n . Si $X = \mathbb{X}_{\{11\}}$, es $\mathcal{B}_0(X) = \{\varepsilon\}$, $\mathcal{B}_1(X) = \{0, 1\}$, $\mathcal{B}_2(X) = \{00, 01, 10\}$, $\mathcal{B}_3(X) = \{000, 001, 010, 100, 101\}$.

Nótese que si $X = \mathbb{X}_{\mathcal{F}}$ y $f \in \mathcal{F}$, no puede haber un n tal que $f \in \mathcal{B}_n(X)$. Es decir, $\forall n \in \mathbb{N}, \mathcal{F} \cap \mathcal{B}_n(X) = \emptyset$. Sin embargo, bien puede ocurrir que un bloque de longitud n que no esté en \mathcal{F} tampoco esté en $\mathcal{B}_n(X)$.

Definición 1.20. Sea X un subconjunto de $\mathcal{A}^{\mathbb{Z}}$. El **lenguaje de X** , denotado por $\mathcal{B}(X)$, es el conjunto de todos los bloques que ocurren en puntos de X . En otras palabras,

$$\mathcal{B}(X) = \{u \in \mathcal{A}^* : \exists x \in X : u \sqsubseteq x\} = \bigcup_{n=0}^{\infty} \mathcal{B}_n(X).$$

Ejemplo 1.21. $\mathcal{B}(\mathbb{X}_{\{11\}}) = \{\varepsilon, 0, 1, 00, 01, 10, 000, 001, 010, 100, 101, \dots\}$.

Ejemplo 1.22. Sean $X_1 = \{(01)^\infty, (10)^\infty\}$ y $X_2 = \{(01)^\infty\}$; se tiene que

$$\mathcal{B}(X_1) = \{\varepsilon, 0, 1, 01, 10, 010, 101, 0101, 1010, \dots\} = \mathcal{B}(X_2)$$

a pesar de que $X_1 \neq X_2$; notar, sin embargo, que X_1 es un espacio shift, mientras que X_2 no lo es.

Así como \mathcal{F} se interpretaba como un conjunto de bloques prohibidos para $\mathbb{X}_{\mathcal{F}}$, $\mathcal{B}(\mathbb{X}_{\mathcal{F}})$ es el conjunto de bloques permitidos para $\mathbb{X}_{\mathcal{F}}$. Cuando un bloque está en el lenguaje de un espacio shift X , decimos de él que **ocurre en X** , o que **aparece en X** , o que **está en X** , o que **está permitido en X** .

Una consideración análoga a la de más arriba: si $X = \mathbb{X}_{\mathcal{F}}$, entonces $\mathcal{F} \cap \mathcal{B}(X) = \emptyset$. Sin embargo, bien puede ocurrir que un bloque que no esté en \mathcal{F} tampoco esté en $\mathcal{B}(X)$.

Si un bloque está en el lenguaje de un espacio shift, la definición nos dice que hay un punto del espacio en el cual, en determinada posición, encontramos ese bloque. Sin embargo, usando shift invariancia, podemos ver que el bloque aparece en algún punto del espacio en la coordenada que uno quiera.

No cualquier lenguaje (en el sentido de la definición 1.19) es el lenguaje de algún espacio shift. Por ejemplo, resulta directo ver que el lenguaje de un espacio shift no vacío es un conjunto infinito, de modo que un subconjunto finito no vacío de \mathcal{A}^* no puede ser el lenguaje de ningún espacio shift sobre \mathcal{A} . Buscamos caracterizar a aquellos lenguajes que son los lenguajes de los espacios shift.

Proposición 1.23. *Sean X un espacio shift y $\mathcal{B}(X)$ su lenguaje. Si $v \in \mathcal{B}(X)$, entonces:*

1. *cualquier subbloque de v pertenece también a $\mathcal{B}(X)$.*
2. *hay bloques u y w no vacíos en $\mathcal{B}(X)$ tales que $uvw \in \mathcal{B}(X)$.*

Demostración. Como $v \in \mathcal{B}(X)$, existen $i, j \in \mathbb{Z}$ y $x \in X$ tales que $v = x_{[i,j]}$. Si $v' \sqsubseteq v$, v' ocurre también en x , de modo que $v' \in \mathcal{B}(X)$. Además, haciendo $u = x_{i-1}$ y $w = x_{j+1}$, se tiene que u y w son bloques no vacíos que están en $\mathcal{B}(X)$ y que $uvw = x_{[i-1,j+1]} \sqsubseteq x$, por lo que $uvw \in \mathcal{B}(X)$. \square

La proposición anterior establece condiciones necesarias para los bloques del lenguaje de un espacio shift. Resulta que esas condiciones son también suficientes: cualquier lenguaje cuyas palabras satisfagan esas condiciones, es el lenguaje de algún espacio shift.

Proposición 1.24. *Sea \mathcal{L} un lenguaje con la propiedad de que para todo $v \in \mathcal{L}$: (a) cualquier subbloque de v está en \mathcal{L} , y (b) hay bloques no vacíos u y w en \mathcal{L} tales que uvw está en \mathcal{L} . Entonces \mathcal{L} es el lenguaje de algún espacio shift.*

Demostración. Consideremos el espacio shift $X = \mathbb{X}_{\mathcal{L}^c}$, y verifiquemos que $\mathcal{B}(X) = \mathcal{L}$. Lo haremos por doble inclusión:

Si $v \in \mathcal{B}(X)$, existe $x \in X$ tal que $v \sqsubseteq x$, de modo que $v \notin \mathcal{L}^c$, o sea que $v \in \mathcal{L}$.

Sea ahora v un bloque de longitud n en \mathcal{L} , digamos $v = x_0 \cdots x_{n-1}$. Por aplicación de la condición (b), hay bloques u y w , cuyos símbolos denotaremos, respectivamente, $x_{-j} \cdots x_{-1}$ y $x_n \cdots x_k$ tales que el bloque $v' = uvw = x_{-j} \cdots x_{-1} x_0 \cdots x_{n-1} x_n \cdots x_k$ está en \mathcal{L} . Aplicando nuevamente la condición (b) a v' hay bloques $u' = x_{-j'} \cdots x_{-j-1}$ y $w' = x_{k+1} \cdots x_{k'}$ tales que el bloque $x_{-j'} \cdots x_{-j-1} x_{-j} \cdots x_{-1} x_0 \cdots x_{n-1} x_n \cdots x_k x_{k+1} \cdots x_{k'}$ está en \mathcal{L} . Continuando de esta manera, tenemos definido x_i para todo entero i , de modo tal que el punto $x = (x_i)_{i \in \mathbb{Z}}$ cumple que todos sus bloques están en \mathcal{L} , y $v = x_{[0,n-1]}$. Luego, $x \in \mathbb{X}_{\mathcal{L}^c}$, ya que ningún bloque de x está en \mathcal{L}^c , y $v \sqsubseteq x$, por lo que $v \in \mathcal{B}(\mathbb{X}_{\mathcal{L}^c}) = \mathcal{B}(X)$. \square

De las dos proposiciones anteriores, vemos que un subconjunto de \mathcal{A}^* es el lenguaje de algún espacio shift si, y sólo si, todos sus bloques cumplen las propiedades (a) y (b) enunciadas en la Proposición 1.24.

Hay una correspondencia biunívoca entre el conjunto de todos los espacios shift y el conjunto de todos los lenguajes de espacios shift. Es decir, el lenguaje de un espacio shift determina completamente a ese espacio shift. Dicho de otro modo, no hay dos espacios shift distintos que tengan el mismo lenguaje, según veremos a continuación.

Proposición 1.25. *Sea $X \subseteq \mathcal{A}^{\mathbb{Z}}$. X es un espacio shift si, y sólo si, $X = \mathbb{X}_{\mathcal{B}(X)^c}$.*

Corolario 1.26. *Sean X_1 y X_2 espacios shift. Entonces, $X_1 = X_2 \Leftrightarrow \mathcal{B}(X_1) = \mathcal{B}(X_2)$.*

En este corolario, es esencial que X_1 y X_2 sean espacios shift (ver ejemplo 1.22).

Todos estos resultados muestran que si bien un mismo espacio shift X puede ser descrito a través de varios conjuntos prohibidos, hay un conjunto de bloques prohibidos

maximal, que es el complemento del lenguaje de ese espacio shift: cualquier \mathcal{F} tal que $X = \mathbb{X}_{\mathcal{F}}$ cumple con que $\mathcal{F} \subseteq \mathcal{B}(X)^c$, pues ya sabemos que $\mathcal{F} \cap \mathcal{B}(X) = \emptyset$.

Resumimos la correspondencia entre espacios shift y lenguajes de espacios shift a través de las siguientes “ecuaciones”:

$$\mathcal{L} = \mathcal{B}(\mathbb{X}_{\mathcal{L}^c}) \qquad X = \mathbb{X}_{\mathcal{B}(X)^c}.$$

La Proposición 1.25 tiene otro útil corolario, que dice que si un punto x de $\mathcal{A}^{\mathbb{Z}}$ tiene todos sus bloques en el lenguaje de un espacio shift X , entonces el punto x necesariamente pertenece a X .

Corolario 1.27. *Sea X un subconjunto de $\mathcal{A}^{\mathbb{Z}}$. X es un espacio shift si, y sólo si,*

$$(\forall x \in \mathcal{A}^{\mathbb{Z}}, (\forall i, j \in \mathbb{Z}, x_{[i,j]} \in \mathcal{B}(X)) \Rightarrow x \in X).$$

1.5. Códigos de ventana deslizante. En Teoría de Códigos, interesan las maneras que hay de traducir un mensaje (fuente) sobre un alfabeto \mathcal{A} en mensaje sobre otro alfabeto \mathcal{U} (posiblemente el mismo que el del mensaje fuente). Esto puede ser visto (sobre todo si el mensaje fuente es largo) como una manera de transformar puntos de un espacio shift en puntos de otro espacio shift, a través de una función. De las maneras en que esto se puede hacer, hay una clásica que explicamos seguidamente.

Supongamos dados dos enteros m y n , con $m + n \geq 0$, y supongamos también dada una función Φ que transforma bloques de tamaño $m + n + 1$ sobre \mathcal{A} en símbolos de \mathcal{U} . Entonces, una manera de transformar un $x \in \mathcal{A}^{\mathbb{Z}}$ en un $y \in \mathcal{U}^{\mathbb{Z}}$ es mirar el bloque que hay en una “ventana” de ancho $m + n + 1$ alrededor de la coordenada i -ésima de x y aplicarle a ese bloque la función Φ para obtener la correspondiente coordenada del transformado. Más concretamente, la i -ésima coordenada de y será $\Phi(x_{[i-m, i+n]})$. Para calcular luego la $(i + 1)$ -ésima coordenada de y , desplazamos toda la ventana una posición a la derecha, y repetimos. Llevando esto a cabo para todo $i \in \mathbb{Z}$, obtenemos el transformado de x . Esta operación de ver el contenido de la ventana, transformar usando Φ y deslizar, motiva que la función transformadora reciba el nombre de código de ventana deslizante, que abreviaremos CVD. Formalmente:

Definición 1.28. Sean $m, n \in \mathbb{Z}$ con $m + n \geq 0$, \mathcal{A} y \mathcal{U} alfabetos, X un espacio shift sobre \mathcal{A} , y Φ una función de $\mathcal{B}_{m+n+1}(X)$ en \mathcal{U} . El **código de ventana deslizante con memoria m y anticipación n inducido por Φ** es la transformación $\phi : X \rightarrow \mathcal{U}^{\mathbb{Z}}$ definida por $y = \phi(x)$ con $y_i = \Phi(x_{[i-m, i+n]})$.

Notar, en la definición de arriba, la distinción tipográfica entre ϕ (que transforma tirillas bi-infinitas) y Φ (que transforma bloques de tamaño $m + n + 1$). Decimos que ϕ es **inducida por Φ** , o que **Φ induce a ϕ** , y, para expresar este hecho, escribimos $\phi = \Phi_{\infty}^{[-m, n]}$. La función Φ se llama **función de bloques**, o **función inductora**, de ϕ . De los números m y n decimos que son, respectivamente, la **memoria** y la **anticipación** de ϕ (o también de Φ), por el hecho de que indican cuántas coordenadas “del pasado” y “del futuro” hay que conocer para obtener el transformado en la posición que corresponde “al presente”.

Ejemplo 1.29. $\mathcal{A} = \{0, 1\}, \mathcal{U} = \{A, B, C\}, m = 1, n = 2, X = \mathbb{X}_{\{110, 101\}}, \Phi : \mathcal{B}_4(X) \rightarrow \mathcal{U}$ dada por

$$\begin{array}{llll} \Phi(0000) = A, & \Phi(0001) = A, & \Phi(0010) = B, & \Phi(0011) = A, \\ \Phi(0100) = C, & \Phi(1000) = B, & \Phi(1001) = C, & \Phi(1111) = A. \end{array}$$

Entonces, si $\phi = \Phi_{\infty}^{[-m,n]}$ y $x = \cdots 001.000010010001 \cdots$, tenemos que

$$\phi(x) = \cdots BC.BAABCCBCBA \cdots$$

pues

$$\begin{aligned}\phi(x)_{-2} &= \Phi(x_{[-2-m,-2+n]}) = \Phi(x_{[-3,0]}) = \Phi(0010) = B, \\ \phi(x)_{-1} &= \Phi(x_{[-1-m,-1+n]}) = \Phi(x_{[-2,1]}) = \Phi(0100) = C, \\ \phi(x)_0 &= \Phi(x_{[0-m,0+n]}) = \Phi(x_{[-1,2]}) = \Phi(1000) = B, \\ \phi(x)_1 &= \Phi(x_{[1-m,1+n]}) = \Phi(x_{[0,2]}) = \Phi(0000) = A,\end{aligned}$$

y así.

Ejemplo 1.30. $\mathcal{A}, \mathcal{U}, X, x$ y Φ iguales que en el ejemplo anterior, excepto que $m = 0, n = 3$. En este caso, $\phi(x) = \cdots BCB.AABCCBCBA \cdots$.

Ejemplo 1.31. $\mathcal{A} = \{0, 1\} = \mathcal{U}, X = \mathcal{A}^{\mathbb{Z}}, m = 0 = n, \Phi : \mathcal{B}_1(X) \rightarrow \mathcal{U}$ dada por $\Phi(0) = 1 = \Phi(1)$, y $F = \Phi_{\infty}^{[-m,n]}$. Entonces, para todo $x \in \mathcal{A}^{\mathbb{Z}}, F(x) = 1^{\infty}$.

Ejemplo 1.32. \mathcal{A} cualquier alfabeto, X cualquier espacio shift sobre $\mathcal{A}, \mathcal{U} = \mathcal{A}, m = 0, n = 1, \Phi(ab) = b, \phi = \Phi_{\infty}^{[-m,n]}$. Entonces, $\phi = \sigma_X$, pues, para $x \in X, \phi x_i = \Phi(x_i x_{i+1}) = x_{i+1} = \sigma_X(x)_i$.

Ejemplo 1.33. \mathcal{A} cualquier alfabeto, X cualquier espacio shift sobre $\mathcal{A}, \mathcal{U} = \mathcal{A}, m = 1, n = 0, \Phi(ab) = a, \phi = \Phi_{\infty}^{[-m,n]}$. Entonces, $\phi = \sigma_X^{-1}$, pues, para $x \in X, \phi x_i = \Phi(x_{i-1} x_i) = x_{i-1} = \sigma_X^{-1}(x)_i$.

Al igual que con la transformación shift σ , si ϕ es un CVD, emplearemos (aunque no obligatoriamente) la notación ϕx para $\phi(x)$ y ϕx_i para $(\phi(x))_i$.

Los CVD para los cuales $m = 0 = n$ se llaman **códigos monobloque**, dado que actúan mirando bloques que tienen un solo símbolo (el que corresponde al “presente”).

Si un CVD ϕ con dominio X es tal que $\phi(X)$ es un subconjunto de algún espacio shift $Y \subseteq \mathcal{U}^{\mathbb{Z}}$, podemos mirar a ϕ como una función entre los espacios shift X e Y , y diremos entonces que el CVD es un código entre los espacios shift X e Y .

No cualquier función ϕ entre dos espacios shift X e Y es un CVD. Para que lo sea, debemos encontrar un m , un n y una $\Phi : \mathcal{B}_{m+n+1}(X) \rightarrow \mathcal{B}_1(Y)$ tal que $\forall x \in X, \forall i \in \mathbb{Z}, \phi x_i = \Phi(x_{[i-m, i+n]})$. Si no es posible el hallazgo de una función inductora, ϕ no es un CVD.

Ejemplo 1.34. $\mathcal{A} = \{0, 1\} = \mathcal{U}, X = \mathcal{A}^{\mathbb{Z}}, Y = \mathcal{U}^{\mathbb{Z}}, \phi : X \rightarrow Y$ definida por

$$\phi(x) = \begin{cases} 0^{\infty} & \text{si } x = 0^{\infty}, \\ 1^{\infty} & \text{si } x \neq 0^{\infty}. \end{cases}$$

Veamos que ϕ no es un CVD. Supongamos que lo fuera. Debe haber una función inductora Φ con memoria m y anticipación n . Sea $x = (x_i)_{i \in \mathbb{Z}}$ definido por $x_{n+1} = 1$, y $x_i = 0$ para todo $i \neq n+1$. Ya que es $(\phi 0^{\infty})_0 = 0$ (pues $\phi 0^{\infty} = 0^{\infty}$), debe ser $\Phi(0^{m+n+1}) = 0$. Pero entonces $\phi x_0 = \Phi(x_{[-m,n]}) = \Phi(0^{m+n+1}) = 0$. Contradicción, pues por la definición de ϕ se tiene que $\phi(x) = 1^{\infty}$, de donde $\phi x_0 = 1$.

A veces será conveniente “agrandar” el ancho de la ventana para la función inductora (sin modificar, obviamente, el CVD en cuestión). Para hacer esto, simplemente debemos ignorar las coordenadas añadidas a la ventana. Esto es lo que dice el siguiente resultado:

Dado un CVD con memoria m y anticipación n inducido por una transformación de bloques Φ (que transforma bloques de tamaño $m+n+1$ en símbolos), se puede

definir, de manera natural, una función Φ^* que transforma bloques de tamaño $m+n+k$ ($k \geq 1$), haciendo

$$\begin{aligned} \Phi^* (a_{-m}a_{-m+1} \cdots a_0a_1 \cdots a_{n-1}a_na_{n+1} \cdots a_{n+k-1}) \\ = \Phi (a_{-m} \cdots a_n) \Phi (a_{-m+1} \cdots a_{n+1}) \cdots \Phi (a_{k-m-1, n+k-1}). \end{aligned}$$

Si además convenimos en que Φ^* , aplicado a un bloque de tamaño menor que $m+n+1$ produce la palabra vacía, tenemos que, para $p \geq 0$, es

$$\Phi^* (a_1 \cdots a_p) = \begin{cases} \varepsilon & \text{si } p < m+n+1, \\ \Phi (a_1 \cdots a_{m+n+1}) \Phi^* (a_2 \cdots a_p) & \text{si } p \geq m+n+1. \end{cases}$$

Φ y Φ^* son funcionalmente tan parecidas que en general usaremos la misma notación Φ para referirnos a ambas.

Un hecho importante a tener en cuenta es que la composición de CVD es otro CVD.

Proposición 1.35. Sean X, Y y Z espacios shift, y supongamos que $\phi_1 : X \rightarrow Y$ y $\phi_2 : Y \rightarrow Z$ son códigos de ventana deslizante. Entonces $\phi_2 \circ \phi_1$ es un código de ventana deslizante desde X hacia Z .

Demostración. Supongamos ϕ_1 inducido por Φ_1 con memoria m_1 y anticipación n_1 , y ϕ_2 inducido por Φ_2 con memoria m_2 y anticipación n_2 . Sea $\phi : X \rightarrow Z$ definido por $\phi x = \phi_2 (\phi_1 x)$. Designemos por ϕ' al CVD con memoria m_1+m_2 y anticipación n_1+n_2 inducido por la función $\Phi : \mathcal{B}_{m_1+m_2+n_1+n_2+1}(X) \rightarrow \mathcal{B}_1(Z)$ definida por

$$\Phi (a_{-m_1-m_2} \cdots a_0 \cdots a_{n_1+n_2}) = \Phi_2 (\Phi_1 (a_{-m_1-m_2} \cdots a_0 \cdots a_{n_1+n_2}))$$

(nótese que Φ_1 corresponde en realidad a Φ_1^* como se definió antes). Para $x \in X$ y para cualquier $i \in \mathbb{Z}$ es

$$\begin{aligned} \phi' x_i &= \Phi (x_{[i-m_2-m_1, i+n_2+n_1]}) = \Phi_2 (\Phi_1 (x_{[i-m_2-m_1, i+n_2+n_1]})) \\ &= \Phi_2 ((\phi_1 x)_{[i-m_2, i+n_2]}) = \phi_2 (\phi_1 x)_i = \phi x_i \end{aligned}$$

de modo que $\phi' x = \phi x$. Luego, $\phi = \phi_2 \circ \phi_1$ es un CVD. \square

El enunciado anterior, junto a los ejemplos 1.32 y 1.33, permite ver que para cualquier espacio shift X y cualquier $k \geq 1$, σ_X^k y σ_X^{-k} son CVD.

Si $\phi : X \rightarrow Y$ es un CVD y $x \in X$, entonces calcular ϕ a la sucesión desplazada $\sigma_X(x)$ da lo mismo que desplazar la sucesión $\phi(x)$ usando σ_Y . Esta propiedad de conmutar con las σ es una de las claves para saber si una transformación es un CVD.

Proposición 1.36. Sean X e Y dos espacios shift, y k un entero. Si ϕ es un código de ventana deslizante, entonces $\phi \circ \sigma_X^k = \sigma_Y^k \circ \phi$, es decir, el siguiente diagrama conmuta:

$$\begin{array}{ccc} X & \xrightarrow{\sigma_X^k} & X \\ \downarrow \phi & & \phi \downarrow \\ Y & \xrightarrow{\sigma_Y^k} & Y \end{array}$$

Demostración. Supongamos que ϕ es inducido por Φ con memoria m y anticipación n . Debemos probar que $\forall x \in X, \phi (\sigma_X^k(x)) = \sigma_Y^k(\phi(x))$. Sea entonces $i \in \mathbb{Z}$. Se tiene que $\phi (\sigma_X^k(x))_i = \Phi ((\sigma_X^k(x))_{[i-m, i+n]}) = \Phi (x_{[i-m+k, i+n+k]}) = \Phi (x_{[(i+k)-m, (i+k)+n]}) = (\phi x)_{i+k} = \sigma_Y^k(\phi x)_i$. Como i es arbitrario, se tiene lo deseado. \square

La condición de conmutar con σ es necesaria, pero no suficiente, para que una transformación sea CVD. El Ejemplo 1.34 sirve para demostrarlo, pues ya hemos visto que la función allí estudiada no es un CVD, y sin embargo conmuta con σ : sea $x \in \{0, 1\}^{\mathbb{Z}}$; si $x = 0^\infty$, es $\sigma x = 0^\infty = \phi x$, por lo que $\sigma(\phi(x)) = 0^\infty = \phi(\sigma x)$; y si $x \neq 0^\infty$, es $\sigma x \neq 0^\infty$, por lo que $\phi x = \phi(\sigma x) = 1^\infty$, teniendo entonces que $\sigma(\phi x) = 1^\infty = \phi(\sigma x)$; resumiendo, $\forall x \in \{0, 1\}^{\mathbb{Z}}$, $\sigma(\phi x) = \phi(\sigma x)$.

Los siguientes resultados establecen condiciones necesarias y suficientes para que una transformación entre dos espacios shift sea un CVD.

Lema 1.37. *Sean X e Y dos espacios shift, y ϕ una función de X en Y tal que $\phi \circ \sigma_X = \sigma_Y \circ \phi$. Entonces $\phi \circ \sigma_X^{-1} = \sigma_Y^{-1} \circ \phi$.*

Demostración. Puesto que σ_X y σ_Y son biyecciones, se tiene que $\sigma_Y^{-1} \circ \sigma_Y$ y $\sigma_Y \circ \sigma_Y^{-1}$ son la identidad en Y , así como $\sigma_X^{-1} \circ \sigma_X$ y $\sigma_X \circ \sigma_X^{-1}$ son la identidad en X . De allí que $\phi \circ \sigma_X^{-1} = (\sigma_Y^{-1} \circ \sigma_Y) \circ (\phi \circ \sigma_X^{-1}) = \sigma_Y^{-1} \circ (\sigma_Y \circ \phi) \circ \sigma_X^{-1} = \sigma_Y^{-1} \circ (\phi \circ \sigma_X) \circ \sigma_X^{-1} = (\sigma_Y^{-1} \circ \phi) \circ (\sigma_X \circ \sigma_X^{-1}) = \sigma_Y^{-1} \circ \phi$, como se quería probar. \square

Lema 1.38. *Sean X e Y dos espacios shift, y ϕ una función de X en Y tal que $\phi \circ \sigma_X = \sigma_Y \circ \phi$. Entonces, para todo $k \in \mathbb{Z}$, $\phi \circ \sigma_X^k = \sigma_Y^k \circ \phi$.*

Demostración. Ya que σ_X^0 y σ_Y^0 son, respectivamente, la identidad en X y en Y , la proposición se cumple para $k = 0$. Lo que resta probar es, entonces, que $\forall k \geq 1$, $\phi \circ \sigma_X^k = \sigma_Y^k \circ \phi$ y $\phi \circ \sigma_X^{-k} = \sigma_Y^{-k} \circ \phi$. Lo haremos por inducción sobre k . Para $k = 1$, es cierto por hipótesis y por Lema 1.37. Suponiendo válido para k , veamos que vale para $k + 1$:

- $\phi \circ \sigma_X^{k+1} = (\phi \circ \sigma_X^k) \circ \sigma_X = (\sigma_Y^k \circ \phi) \circ \sigma_X = \sigma_Y^k \circ (\phi \circ \sigma_X) = \sigma_Y^k \circ (\sigma_Y \circ \phi) = \sigma_Y^{k+1} \circ \phi$
- $\phi \circ \sigma_X^{-k-1} = (\phi \circ \sigma_X^{-k}) \circ \sigma_X^{-1} = (\sigma_Y^{-k} \circ \phi) \circ \sigma_X^{-1} = \sigma_Y^{-k} \circ (\phi \circ \sigma_X^{-1}) = \sigma_Y^{-k} \circ (\sigma_Y^{-1} \circ \phi) = \sigma_Y^{-k-1} \circ \phi$

\square

Proposición 1.39. *Sean X e Y dos espacios shift, y ϕ una función de X en Y . ϕ es un código de ventana deslizante si, y sólo si, las dos condiciones siguientes se satisfacen simultáneamente:*

1. $\phi \circ \sigma_X = \sigma_Y \circ \phi$.
2. *Existen un entero $N \geq 0$ y una función $\Phi : \mathcal{B}_{2N+1}(X) \rightarrow \mathcal{B}_1(Y)$ tales que para todo $x \in X$, $(\phi x)_0 = \Phi(x_{[-N, N]})$.*

La segunda condición de la proposición anterior establece que existe $N \geq 0$ tal que $(\phi x)_0$ es función de lo que x contiene entre las coordenadas $-N$ y N . La negación de esta condición significa que para todo $N \geq 0$, existen $x, y \in X$ tales que $x_{[-N, N]} = y_{[-N, N]}$ pero $(\phi x)_0 \neq (\phi y)_0$.

1.5.1. Inmersiones, factores y conjugaciones. Para las siguientes definiciones, sean X e Y dos espacios shift, y ϕ una función de X en Y .

Definición 1.40. ϕ es una **inmersión de X en Y** si ϕ es un código de ventana deslizante inyectivo. Se dice que X **está inmerso en Y** si existe una inmersión de X en Y .

Definición 1.41. ϕ es un **código factor de X sobre Y** si ϕ es un código de ventana deslizante sobreyectivo. Se dice que Y **es un factor de X** si existe un código factor de X sobre Y .

Definición 1.42. ϕ es una **conjugación de X a Y** si ϕ es un código de ventana deslizante biyectivo. Se dice que X es **conjugado a Y** (simbolizado $X \sim Y$) si existe una conjugación de X a Y .

Por ejemplo, σ_X y σ_X^{-1} son conjugaciones de X a X ,

La definición establece que una conjugación es un CVD biyectivo (por lo cual, toda conjugación es también un código factor y una inmersión), y esto nos permite garantizar sólo que su inversa es una función biyectiva (pero, por lo visto hasta aquí, no necesariamente un CVD). Más adelante, demostraremos que si ϕ es conjugación, ϕ^{-1} es un CVD (y, en consecuencia, también es una conjugación).

Proposición 1.43. *La composición de dos conjugaciones es otra conjugación.*

Demostración. Se deduce directamente de la Proposición 1.35 y del hecho que la composición de funciones biyectivas es una función biyectiva. \square

Como consecuencia, σ_X^k y σ_X^{-k} son también conjugaciones de X a X , para cualquier espacio shift X y cualquier $k \geq 0$, pues σ_X y σ_X^{-1} lo son.

Veamos la acción de un CVD sobre los puntos periódicos de un espacio shift. El siguiente resultado muestra que las inmersiones (y en consecuencia las conjugaciones) preservan el mínimo período de un punto periódico.

Proposición 1.44. *Sean $\phi : X \rightarrow Y$ un código de ventana deslizante, $x \in X$ un punto periódico período n para σ_X , e $y = \phi x$. Entonces y tiene también período n para σ_Y , y el mínimo período de y divide al mínimo período de x . Si ϕ es además una inmersión, el período mínimo de y es igual al período mínimo de x .*

Demostración. Ya que $\sigma_X^n(x) = x$, tenemos que, por la propiedad de conmutar ϕ con σ , $\phi x = \phi(\sigma_X^n(x)) = \sigma_Y^n(\phi x)$, es decir, $y = \sigma_Y^n(y)$. Entonces y tiene período n para σ_Y . Si x tiene mínimo período n_0 , entonces y tiene período n_0 , y el mínimo período de y divide a n_0 . Si ϕ es inyectiva, tenemos que $\sigma_X^n(x) = x$ si, y sólo si, $\sigma_Y^n(\phi x) = \phi x$ (la ida está demostrada al principio, y para la vuelta: $\sigma_Y^n(\phi x) = \phi x \Rightarrow \phi(\sigma_X^n(x)) = \phi x \Rightarrow \sigma_X^n(x) = x$ por la inyectividad de ϕ). Entonces, x tiene período n si, y sólo si, y tiene período n . Dicho de otro modo, el conjunto de períodos para x es el mismo que el conjunto de períodos para y , de manera que el período mínimo de x es el mismo que el período mínimo de y . \square

De esta proposición se deduce que dos espacios shift que no tienen la misma cantidad de puntos periódicos de período n (cualquiera sea n) no pueden ser conjugados. Es decir, la cantidad de puntos de período n de un espacio shift es un **invariante por conjugación**: $\forall n > 0, X \sim Y \Rightarrow |\{x \in X : x \text{ tiene período } n\}| = |\{x \in Y : x \text{ tiene período } n\}|$.

1.5.2. Imagen de un espacio shift por un código de ventana deslizante. Hasta ahora, hemos visto que un CVD ϕ actúa transformando puntos de un espacio shift X en puntos de algún full shift $\mathcal{U}^{\mathbb{Z}}$. Cabe preguntarse cómo actúa “globalmente” ϕ , en el sentido de qué características tiene $\phi(X)$ como subconjunto de $\mathcal{U}^{\mathbb{Z}}$. Mostraremos ahora que $\phi(X)$ es un espacio shift sobre el alfabeto \mathcal{U} .

Lema 1.45. *Sea ϕ un código de ventana deslizante monobloque desde un espacio shift X hacia un full shift $\mathcal{U}^{\mathbb{Z}}$. Entonces, $\phi(X)$ es un espacio shift sobre el alfabeto \mathcal{U} .*

Demostración. Llamemos \mathcal{A} al alfabeto de X . Supongamos que ϕ está inducida por $\Phi : \mathcal{B}_1(X) \rightarrow \mathcal{U}$. Hagamos $\mathcal{L} = \{\Phi(w) : w \in \mathcal{B}(X)\}$, y verifiquemos que $\phi(X) = \mathbb{X}_{\mathcal{L}^c}$, con lo que quedará probado que $\phi(X)$ es un espacio shift.

1. Sea $y \in \phi(X)$. Existe $x \in X$ tal que $y = \phi x$. Sea u tal que $u \sqsubseteq y$. Entonces existen $i, j \in \mathbb{Z}$ tales que $u = y_{[i,j]} = (\phi x)_{[i,j]} = \Phi(x_{[i,j]})$, es decir, $u = \Phi(w)$ con $w \in \mathcal{B}(X)$. Luego, $u \in \mathcal{L}$, por lo que $u \notin \mathcal{L}^c$. Como u era un bloque arbitrario de y , vemos que $y \in \mathbb{X}_{\mathcal{L}^c}$. Entonces, $\phi(X) \subseteq \mathbb{X}_{\mathcal{L}^c}$.

2. Sea $y \in \mathbb{X}_{\mathcal{L}^c}$. Entonces todo bloque de y está en \mathcal{L} , de donde se deduce que para todo $n \geq 0$, existe $w \in \mathcal{B}(X)$ tal que $y_{[-n,n]} = \Phi(w)$. Por lo tanto, para todo $n \geq 0$, existe $x^{(n)} \in X$ tal que $x_{[-n,n]}^{(n)} = w$ y $(\phi x^{(n)})_{[-n,n]} = \Phi(x_{[-n,n]}^{(n)}) = y_{[-n,n]}$. Usaremos la sucesión $\{x^{(n)}\}_{n \in \mathbb{N}}$ para encontrar un punto $x \in X$ tal que $\phi x = y$, con lo que quedará probado que $y \in \phi(X)$, y, con ello, que $\mathbb{X}_{\mathcal{L}^c} \subseteq \phi(X)$. Como \mathcal{A} es finito, hay una cantidad infinita de n tales que $x^{(n)}$ tienen igual símbolo a_0 en la posición 0. Sea S_0 el conjunto de todos esos n . Como \mathcal{A}^3 es finito, hay un subconjunto infinito S_1 de S_0 y símbolos a_{-1} y a_1 tales que $\forall n \in S_1, x_{[-1,1]}^{(n)} = a_{-1}a_0a_1$. Continuando de esta manera, para cada $k \geq 1$ conseguimos un subconjunto infinito S_k de S_{k-1} y símbolos a_{-k} y a_k tales que $\forall n \in S_k, x_{[-k,k]}^{(n)} = a_{-k} \cdots a_0 \cdots a_k$. Definamos $x = (a_k)_{k \in \mathbb{Z}}$. Ocurrirá que $\forall k \geq 0, \forall n \in S_k, x_{[-k,k]}^{(n)} = x_{[-k,k]}$. Si $u = x_{[i,j]}$ para algunos enteros i y j , haciendo $k = \max\{|i|, |j|\}$ será $x_{[-k,k]} = x_{[-k,k]}^{(n)}$ para cualquier $n \in S_k$, por lo que existe $n \geq 0$ tal que $u \sqsubseteq x^{(n)} \in X$, es decir, $u \in \mathcal{B}(X)$; puesto que cualquier bloque de x está en el lenguaje del espacio shift X , se tiene que, por Corolario 1.27, $x \in X$. Sea ahora $k \geq 0$. Tomemos $n \in S_k$ tal que $n \geq k$ (tal elección es posible pues S_k es un conjunto infinito). Tenemos que

$$(\phi x)_{[-k,k]} = \Phi(x_{[-k,k]}) = \Phi(x_{[-k,k]}^{(n)}) = y_{[-k,k]}.$$

Es decir, $\forall k \geq 0, (\phi x)_{[-k,k]} = y_{[-k,k]}$. Entonces, $\phi x = y$. \square

Teorema 1.46. *Sea ϕ un código de ventana deslizante desde un espacio shift X hacia un full shift $\mathcal{U}^{\mathbb{Z}}$. Entonces, $\phi(X)$ es un espacio shift sobre el alfabeto \mathcal{U} .*

1.5.3. *Inversa de una conjugación.* Mostremos ahora que la inversa de una conjugación es siempre una conjugación.

Lema 1.47. *Sea ϕ una conjugación monobloque del espacio shift X al espacio shift Y . Entonces, ϕ^{-1} es una conjugación de Y a X .*

Demostración. Sean \mathcal{A} el alfabeto para X y \mathcal{U} el de Y . Supongamos que ϕ está inducida por $\Phi : \mathcal{B}_1(X) \rightarrow \mathcal{B}_1(Y)$. Hagamos $\psi = \phi^{-1}$. Debemos mostrar que ψ es un CVD biyectivo de Y a X . Por ser la inversa de una función biyectiva, ψ es también biyectiva, así que sólo resta verificar que es CVD, para lo que usaremos la Proposición 1.39.

1. Por ser ϕ un CVD, tenemos que $\phi \circ \sigma_X = \sigma_Y \circ \phi$, por lo que $\psi \circ \phi \circ \sigma_X \circ \psi = \psi \circ \sigma_Y \circ \phi \circ \psi$, y entonces $\sigma_X \circ \psi = \psi \circ \sigma_Y$. Por lo tanto, ψ cumple la condición de conmutar con σ .

2. Supongamos que no existe un $N \geq 0$ tal que para todo $y \in Y$, $(\psi y)_0$ dependa de $y_{[-N,N]}$. Esto quiere decir que para cada $n \geq 0$ existen $y^{(n)}$ e $\widehat{y}^{(n)}$ en Y tales que $y_{[-n,n]}^{(n)} = \widehat{y}_{[-n,n]}^{(n)}$ pero $\psi(y^{(n)})_0 \neq \psi(\widehat{y}^{(n)})_0$. Llamemos $x^{(n)} = \psi(y^{(n)})$ y $\widehat{x}^{(n)} = \psi(\widehat{y}^{(n)})$. Tenemos que, para cada $n \geq 0$, tanto $x^{(n)}$ como $\widehat{x}^{(n)}$ están en X , y además $x^{(n)} \neq \widehat{x}^{(n)}$ pues ambos difieren en la coordenada 0. Consideremos las dos sucesiones en X , $\{x^{(n)}\}_{n \in \mathbb{N}}$ y $\{\widehat{x}^{(n)}\}_{n \in \mathbb{N}}$. Como \mathcal{A} es finito, debe haber un símbolo $a_0 \in \mathcal{A}$, y un conjunto infinito S'_0 de enteros tal que $\forall n \in S'_0, (x^{(n)})_0 = a_0$; nótese que $\forall n \in S'_0, (\widehat{x}^{(n)})_0 \neq a_0$. Puesto que \mathcal{A} es finito, debe haber un símbolo $b_0 \neq a_0$ y un subconjunto infinito S'_0 de S'_0 tal que

$\forall n \in S_0, (\widehat{x}^{(n)})_0 = b_0$. Observar que, entonces, $\forall n \in S_0, (x^{(n)})_0 = a_0 \neq b_0 = (\widehat{x}^{(n)})_0$. Como la cantidad de bloques en \mathcal{A}^3 es finita, debe haber un subconjunto infinito S'_1 de S_0 y símbolos a_{-1} y a_1 tales que $\forall n \in S'_1, (x^{(n)})_{[-1,1]} = a_{-1}a_0a_1$, y consecuentemente debe haber un subconjunto infinito S_1 de S'_1 y símbolos b_{-1} y b_1 tales que $\forall n \in S_1, (\widehat{x}^{(n)})_0 = b_{-1}b_0b_1$. Notar que $\forall n \in S_1, (x^{(n)})_{[-1,1]} = a_{-1}a_0a_1 \neq b_{-1}b_0b_1 = (\widehat{x}^{(n)})_{[-1,1]}$. Continuando de esta manera, para cada $k \geq 1$ encontramos un subconjunto infinito S_k de S_{k-1} y símbolos a_{-k}, a_k, b_{-k} y b_k tales que $\forall n \in S_k, (x^{(n)})_{[-k,k]} = a_{-k} \cdots a_0 \cdots a_k \neq b_{-k} \cdots b_0 \cdots b_k = (\widehat{x}^{(n)})_{[-k,k]}$. Tomemos $x = (a_k)_{k \in \mathbb{Z}}$ y $\widehat{x} = (b_k)_{k \in \mathbb{Z}}$. Obsérvese que $\forall k \geq 0, \forall n \in S_k, x_{[-k,k]} = x_{[-k,k]}^{(n)}$ y $\widehat{x}_{[-k,k]} = \widehat{x}_{[-k,k]}^{(n)}$. Si $u = x_{[i,j]}$ para algunos enteros i y j , haciendo $k = \max\{|i|, |j|\}$ será $x_{[-k,k]} = x_{[-k,k]}^{(n)}$ para cualquier $n \in S_k$, por lo que existe $n \geq 0$ tal que $u \sqsubseteq x^{(n)} \in X$, es decir, $u \in \mathcal{B}(X)$; puesto que cualquier bloque de x está en el lenguaje del espacio shift X , se tiene que, por Corolario 1.27, $x \in X$. Igualmente, $\widehat{x} \in X$. Además, $x \neq \widehat{x}$ pues $x_0 = a_0 \neq b_0 = \widehat{x}_0$. Sea ahora $k \geq 0$. Tomemos $n \in S_k$ tal que $n \geq k$ (tal elección es posible pues S_k es un conjunto infinito). Tenemos que

$$\begin{aligned} (\phi x)_{[-k,k]} &= \Phi(x_{[-k,k]}) = \Phi(x_{[-k,k]}^{(n)}) = y_{[-k,k]}^{(n)} = \widehat{y}_{[-k,k]}^{(n)} \\ &= \Phi(\widehat{x}_{[-k,k]}^{(n)}) = \Phi(\widehat{x}_{[-k,k]}) = (\phi \widehat{x})_{[-k,k]}. \end{aligned}$$

Es decir, $\forall k \geq 0, (\phi x)_{[-k,k]} = (\phi \widehat{x})_{[-k,k]}$. De allí, tendríamos que $\phi x = \phi \widehat{x}$ pese a ser $x \neq \widehat{x}$. Contradicción a la inyectividad de ϕ . La contradicción proviene de suponer que no hay un $N \geq 0$ tal que ψy_0 es función de $y_{[-N,N]}$. \square

Teorema 1.48. *Sea ϕ una conjugación del espacio shift X al espacio shift Y . Entonces, ϕ^{-1} es una conjugación de Y a X .*

2. SHIFTS DE TIPO FINITO

Los espacios shift que pueden ser descriptos por un conjunto *finito* de bloques prohibidos se llaman *shifts de tipo finito* (abreviado STF). A pesar de ser los “más simples”, juegan un rol fundamental en ciertas áreas de la Matemática, tales como sistemas dinámicos. Su estudio también ha proporcionado soluciones a problemas prácticos de importancia, tales como encontrar esquemas de codificación eficientes para almacenar datos en discos de computadoras.

Una razón por la que los STF son tan útiles es que tienen una representación sencilla usando un grafo finito dirigido. Preguntas sobre el shift pueden, a menudo, reformularse como preguntas sobre la matriz de adyacencia del grafo. Resultados básicos del álgebra lineal nos ayudan a analizar esta matriz y encontrar respuestas.

En este capítulo, primero presentamos los STF, y damos algunos ejemplos típicos. Luego se explica las conexiones con los grafos dirigidos y las matrices, e introducimos la fundamental operación de desdoblamiento de estados. Concluimos con una breve descripción sobre el almacenamiento magnético, indicando por qué la shift de paso limitado (1.3) es la característica central de la mayoría de los métodos más usados para el almacenamiento de datos en discos rígidos de computadoras.

2.1. Restricciones de tipo finito. Primero definimos shifts de tipo finito, y luego explicamos el hecho de que tienen una propiedad (“markoviana”) de memoria finita.

Definición 2.1. Un **shift de tipo finito** (STF) es un espacio shift que puede describirse por un conjunto finito de bloques prohibidos, es decir, un espacio shift X para el cual existe algún conjunto finito \mathcal{F} de bloques tal que $X = \mathbb{X}_{\mathcal{F}}$.

Ya hemos tratado antes con varios STF.

Ejemplo 2.2. El full shift $\mathcal{A}^{\mathbb{Z}}$ es un STF: podemos tomar simplemente $\mathcal{F} = \emptyset$ pues nada está prohibido, de modo que $\mathcal{A}^{\mathbb{Z}} = \mathbb{X}_{\mathcal{F}}$.

Ejemplo 2.3. El shift de la razón de oro X (ejemplo 1.7) es un STF, pues podemos tomar $\mathcal{F} = \{11\}$, obteniendo $X = \mathbb{X}_{\mathcal{F}}$.

Ejemplo 2.4. Sea X el conjunto de todos los puntos de $\{e, f, g\}^{\mathbb{Z}}$ tales que e puede ser seguido sólo por e o por f , f puede ser seguido sólo por g , y g puede ser seguido sólo por e o por f . Entonces X es el espacio shift que resulta de tomar $\mathcal{F} = \{eg, fe, ff, gg\}$, y, en consecuencia, es un STF.

Notemos que un STF X podría también describirse mediante un conjunto infinito de bloques prohibidos. De hecho, la Proposición 1.25 muestra que ésto puede ocurrir siempre que X no sea el full shift, pues el complemento del lenguaje de X es infinito. La definición de STF sólo pide que exista *algún* \mathcal{F} adecuado.

Supongamos que $X \subseteq \mathcal{A}^{\mathbb{Z}}$ es STF, y que $X = \mathbb{X}_{\mathcal{F}}$ para una colección finita \mathcal{F} de bloques. Sea N la longitud del bloque más largo en \mathcal{F} . Si formamos la colección \mathcal{F}_N de todos los bloques de longitud N sobre \mathcal{A} que contienen algún subbloque en \mathcal{F} , entonces tendremos que $\mathbb{X}_{\mathcal{F}_N} = \mathbb{X}_{\mathcal{F}}$ (prop. 1.18) y los bloques de \mathcal{F}_N tienen todos la misma longitud N . Por ejemplo, si $\mathcal{A} = \{0, 1\}$ y $\mathcal{F} = \{11, 000\}$, entonces $\mathcal{F}_3 = \{110, 111, 011, 000\}$. A veces, será conveniente asumir que este procedimiento ya ha sido llevado a cabo, y que todos los bloques prohibidos tienen igual longitud.

Si todos los bloques en \mathcal{F} tienen longitud N , entonces $x \in \mathcal{A}^{\mathbb{Z}}$ está en $\mathbb{X}_{\mathcal{F}}$ exactamente cuando $x_{[i, i+N-1]} \notin \mathcal{F}$ para todo $i \in \mathbb{Z}$, o, equivalentemente, cuando $x_{[i, i+N-1]} \in \mathcal{B}_N(\mathbb{X}_{\mathcal{F}})$ para todo $i \in \mathbb{Z}$ (Supongamos que $\forall i \in \mathbb{Z}, x_{[i, i+N-1]} \notin \mathcal{F}$. Sea u un bloque de x . Si $|u| \neq N, u \notin \mathcal{F}$ pues \mathcal{F} tiene sólo bloques de tamaño N , y si $|u| = N, u \notin \mathcal{F}$ por la hipótesis; entonces ningún bloque de x está en \mathcal{F} , por lo que $x \in \mathbb{X}_{\mathcal{F}}$ y $\forall i \in \mathbb{Z}, x_{[i, i+N-1]} \in \mathcal{B}_N(\mathbb{X}_{\mathcal{F}})$. Recíprocamente, si un bloque de tamaño N está en $\mathcal{B}_N(\mathbb{X}_{\mathcal{F}})$, no puede estar en \mathcal{F}). Luego, para determinar si x está o no en $\mathbb{X}_{\mathcal{F}}$, sólo necesitamos “escanear” las coordenadas de x con una “ventana” de ancho N , y verificar que cada bloque visto a través de esta ventana está en la colección permitida $\mathcal{B}_N(X)$. Esta observación es, a menudo, útil cuando se debe decidir si un dado espacio shift es o no de tipo finito.

Ejemplo 2.5. El shift par X (ejemplo 1.10) no es un STF, pues, si lo fuera, existiría $N \geq 1$ y una colección de N -bloques tal que $X = \mathbb{X}_{\mathcal{F}}$. En dicha colección, no puede haber bloques de la forma 0^k , ni $0^k 10^l$ para ningún $k, l \geq 0$, ya que $0^\infty 10^\infty \in X$. Pero entonces $0^\infty 10^{2N+1} 10^\infty \in \mathbb{X}_{\mathcal{F}}$ pues ninguno de sus N -bloques está en \mathcal{F} . Sin embargo, ese punto no pertenece a X por la definición del shift par.

Hay también una noción de “memoria” para un STF:

Definición 2.6. Un shift de tipo finito **es de memoria** M si puede describirse mediante una colección de bloques prohibidos que tienen todos longitud $M + 1$.

Para motivar esta definición, supongamos que todos los bloques de \mathcal{F} tienen longitud $M + 1$. Sea $u = a_1 a_2 \dots a_n$ un bloque con longitud n mucho mayor que M . Supongamos que una máquina lee los símbolos de u uno por vez, de izquierda a derecha. Para que

esta máquina detecte si u contiene o no un bloque prohibido, sólo debe recordar los previos M símbolos leídos, es decir, necesita sólo M espacios de memoria.

Observación 2.7. Sea X un STF de memoria M , descrito a través de un conjunto \mathcal{F} de bloques de longitud $M + 1$.

1. Sea u un bloque de \mathcal{A}^* que no contiene ningún subbloque en \mathcal{F} . Esto no garantiza que $u \in \mathcal{B}(X)$. Por ejemplo: $\mathcal{F} = \{10, 11\}$, $u = 001$. u no contiene subbloques en \mathcal{F} , pero $u \notin \mathcal{B}(X)$ pues $X = \{0^\infty\}$.
2. Sea $\mathcal{F}' = \mathcal{A}^{M+1} - \mathcal{B}(\mathbb{X}_{\mathcal{F}})$. Entonces, $\mathcal{F} \subseteq \mathcal{F}'$, y $\mathbb{X}_{\mathcal{F}'} = \mathbb{X}_{\mathcal{F}}$. Para ver la primera aserción, si $f \in \mathcal{F}$, $|f| = M + 1$ y $f \notin \mathcal{B}(\mathbb{X}_{\mathcal{F}})$, luego $f \in \mathcal{F}'$. Para la segunda, ya que $\mathcal{F} \subseteq \mathcal{F}'$, se tiene que $\mathbb{X}_{\mathcal{F}'} \subseteq \mathbb{X}_{\mathcal{F}}$. Si $x \in \mathbb{X}_{\mathcal{F}}$ y u es un bloque de longitud $M + 1$ que ocurre en x , $u \in \mathcal{B}(\mathbb{X}_{\mathcal{F}})$, por lo que $u \notin \mathcal{F}'$. Luego x no contiene bloques en \mathcal{F}' , mostrando que $x \in \mathbb{X}_{\mathcal{F}'}$ y, consecuentemente, $\mathbb{X}_{\mathcal{F}} \subseteq \mathbb{X}_{\mathcal{F}'}$.

Notar que un STF de memoria M es también de memoria K para todo $K \geq M$. Un STF de memoria 0 es simplemente un full shift, pues prohibir bloques de tamaño 1 es meramente quitar letras del alfabeto. Se puede pensar en un STF de memoria 1 como en uno dado por una colección de símbolos junto con una descripción de cuáles símbolos pueden seguir a cuáles, como en el ejemplo 2.4. Posteriormente veremos muchos ejemplos de tales STF de memoria 1.

El lenguaje de un STF está caracterizado por la propiedad de que si dos palabras se solapan lo suficiente, pueden ser unidas a lo largo del solape para formar otra palabra en el lenguaje.

Teorema 2.8. *Sea X un espacio shift sobre el alfabeto \mathcal{A} , y $M \geq 0$. X es STF de memoria M si, y sólo si, $\forall u, v, w \in \mathcal{A}^*$, $uv \in \mathcal{B}(X) \wedge vw \in \mathcal{B}(X) \wedge |v| \geq M \Rightarrow uvw \in \mathcal{B}(X)$*

Demostración. Para la necesidad de la condición, sea \mathcal{F} una colección de bloques de tamaño $M + 1$ tal que $X = \mathbb{X}_{\mathcal{F}}$. Sean u, v y w en \mathcal{A}^* tales que uv y vw estén en $\mathcal{B}(X)$ con $|v| \geq M$. Existen x e y en X tales que $x_{[i, i+|uv|-1]} = uv$, $y_{[j, j+|vw|-1]} = vw$. Consideremos el punto $z = x_{(-\infty, i+|uv|-1]}y_{[j+|v|, \infty)} = \dots x_{i-2}x_{i-1}uvw y_{j+|vw|}y_{j+|vw|+1} \dots$. No puede contener un bloque en \mathcal{F} pues, al ser $|v| \geq M$, una ventana de ancho $M + 1$ no puede ver simultáneamente símbolos de u, v y w , por lo que todo bloque que esta ventana vea corresponde a un bloque en x o en y . Entonces $z \in X$ y $uvw \in \mathcal{B}(X)$.

Para la suficiencia, definamos $\mathcal{F} = \mathcal{A}^{M+1} - \mathcal{B}_{M+1}(X)$, y veamos que $X = \mathbb{X}_{\mathcal{F}}$. Si $x \in X$ e $i \in \mathbb{Z}$, $x_{[i, i+M]} \in \mathcal{B}_{M+1}(X)$, es decir, $\forall i, x_{[i, i+M]} \notin \mathcal{F}$. Luego $x \in \mathbb{X}_{\mathcal{F}}$, y se deduce que $X \subseteq \mathbb{X}_{\mathcal{F}}$. Ahora, sea $x \in \mathbb{X}_{\mathcal{F}}$. Mostremos, por inducción sobre n , que $\forall n \geq M, x_{[0, n]} \in \mathcal{B}(X)$. Para $n = M$ es cierto pues $x_{[0, M]} \notin \mathcal{F}$, por lo que $x_{[0, M]} \in \mathcal{B}_{M+1}(X)$, y, de allí, $x_{[0, M]} \in \mathcal{B}(X)$; esto muestra también que $\forall n \in \{0, \dots, M-1\}, x_{[0, n]} \in \mathcal{B}(X)$. Supongamos que $x_{[0, n]} \in \mathcal{B}(X)$. Hagamos $u = x_{[0, n-M]}$, $v = x_{[n-M+1, n]}$ y $w = x_{n+1}$. Observemos que $uv = x_{[0, n]} \in \mathcal{B}(X)$, $vw = x_{[n-M+1, n+1]} \notin \mathcal{F}$ y $|vw| = M + 1$, por lo que $vw \in \mathcal{B}_{M+1}(X) \subseteq \mathcal{B}(X)$. Además, $|v| = M$. Entonces $uvw \in \mathcal{B}(X)$, es decir, $x_{[0, n+1]} \in \mathcal{B}(X)$. Un razonamiento análogo en la otra dirección muestra que $\forall m, n \in \mathbb{Z}, x_{[m, n]} \in \mathcal{B}(X)$, es decir, cualquier bloque en x está en $\mathcal{B}(X)$, y, siendo X espacio shift, $x \in X$, por lo que $\mathbb{X}_{\mathcal{F}} \subseteq X$. \square

Ejemplo 2.9. El teorema anterior da otra forma de ver que el shift par X no es de tipo finito. Pues si lo fuera, sería de memoria M para algún $M \geq 1$. Ya que $10^{2M+1} \in \mathcal{B}(X)$ y que $0^{2M+1} \in \mathcal{B}(X)$, deberíamos tener que $10^{2M+1} \in \mathcal{B}(X)$, lo que violaría la definición de shift par.

Consideremos el CVD ϕ con memoria 0 y anticipación 1 que tiene por dominio el shift de la razón de oro X , inducido por Φ tal que $\Phi(00) = 1$, $\Phi(01) = 0 = \Phi(10)$. Ocurre que $\Phi(u) = 10^k 1$ si, y sólo si, existe $n \geq 0$ tal que $u = 0(01)^{2n} 00$, por lo que $k = 2n$. Luego, $\phi(X) \subseteq Y$, donde Y es el shift par. Además, si $y \in Y$, es posible construir $x \in X$ tal que $y = \phi(x)$ (teniendo en cuenta que si $y_{[i, i+2k+1]} = 10^{2k} 1$, entonces debe ser $x_{[i, i+2k+2]} = 0(01)^k 00$). Entonces, ϕ es un código factor del shift de la razón de oro en el shift par, siendo el primero de tipo finito y el segundo no. Es fácil construir también códigos de ventana deslizante sobreobjetivos desde espacios shift que no son de tipo finito hacia otros que sí lo son (por ejemplo, desde el shift par hacia $\{0^\infty\}$ por el CVD de memoria 0 y anticipación 0 inducido por $\Phi(0) = \Phi(1) = 0$). Sin embargo, las conjugaciones preservan el carácter de tipo finito.

Teorema 2.10. *Sea X un shift de tipo finito, e Y un espacio shift conjugado a X . Entonces Y es un shift de tipo finito.*

Demostración. Sean \mathcal{A} y \mathcal{U} los alfabetos para X e Y respectivamente. Supongamos que X es un STF de memoria M . Sea $\phi : X \rightarrow Y$ una conjugación inducida por Φ , y $\psi : Y \rightarrow X$ su inversa, inducida por Ψ . Podemos suponer que ϕ y ψ tienen ambas la misma memoria y anticipación l . Sabemos que $\psi \circ \phi$ es código de ventana deslizante con memoria y anticipación $2l$, y que una función inductora es $\Psi \circ \Phi : \mathcal{B}_{4l+1}(X) \rightarrow \mathcal{B}_1(X)$. Dicha composición de inductoras, aplicada a un bloque $a_{-2l} \dots a_0 \dots a_{2l}$ devuelve el carácter central a_0 ; como consecuencia de ésto, si s, t y w son bloques con $|s| = |t| = 2l$, $\Phi \circ \Psi(swt) = w$. Mostraremos que Y es un STF de memoria $M + 4l$, usando el teorema anterior. Sean $u, v, w \in \mathcal{U}^*$ tales que $uv \in \mathcal{B}(Y)$, $vw \in \mathcal{B}(Y)$, $|v| \geq M + 4l$, digamos $v = v_1 \dots v_k$ con $k \geq M + 4l$. Existen $s, t \in \mathcal{B}(Y) : swv \in \mathcal{B}(Y)$, $vwt \in \mathcal{B}(Y)$, $|s| = |t| = 2l$. Sea $u' = \Psi(suv_1 \dots v_{2l})$, $w' = \Psi(v_{k-2l+1} \dots v_k wt)$. Entonces $\Psi(suv) = u' \Psi(v) \in \mathcal{B}(X)$, $\Psi(vwt) = \Psi(v)w' \in \mathcal{B}(X)$. Como $|v| \geq M + 4$, será $|\Psi(v)| = |v| - 2l \geq M$. Luego $\Psi(suvw) = u' \Psi(v)w' \in \mathcal{B}(X)$. Por ello, $\Phi(\Psi(suvw)) \in \mathcal{B}(Y)$, es decir, $uvw \in \mathcal{B}(Y)$. \square

2.2. Grafos y sus shifts. Un método fundamental para construir shifts de tipo finito se basa en un grafo finito dirigido, considerando la colección de todos los caminos bi-infinitos (es decir, sucesiones de aristas *consecutivas*) en el grafo. En un sentido que precisaremos más adelante, cualquier shift de tipo finito puede ser recodificado en un shift de aristas. En esta sección, introducimos los shifts de aristas, y usamos la matriz de adyacencia del grafo para responder preguntas sobre su shift. Realmente, la razón por la que un shift de aristas puede ser entendido mucho mejor que un shift general es que podemos aplicar la potente maquinaria del álgebra lineal a su matriz de adyacencia.

Comenzamos con las definiciones básicas concernientes a la teoría de grafos.

Definición 2.11. Un **grafo finito dirigido** G es una cuaterna (V, Σ, i, t) donde:

V es un conjunto finito cuyos elementos llamaremos **vértices**, **nodos** o **estados**.

Σ es un conjunto finito cuyos elementos llamaremos **aristas** o **arcos**.

i es una función de Σ en V . Para una arista a , $i(a)$ se llama **vértice inicial de a** . También decimos que **la arista a arranca en el nodo $i(a)$** .

t es una función de Σ en V . Para una arista a , $t(a)$ se llama **vértice terminal de a** . También decimos que **la arista a termina en el nodo $t(a)$** .

Una manera cómoda de visualizar un grafo es mediante un dibujo donde los nodos se representan por círculos (con el nombre del nodo en su interior) y cada arista a mediante una flecha desde $i(a)$ hasta $t(a)$, con un rótulo que indica su nombre. La

figura 1 muestra un grafo con $V = \{I, J\}$ y ocho aristas, teniéndose que, por ejemplo, $i(e) = I$, $t(e) = J$, $i(f) = t(f) = i(g) = t(g) = J$, etc.

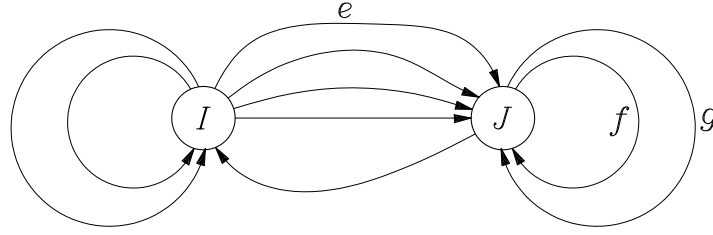


FIGURA 1.

Muchas veces, cuando hagamos referencia a distintos grafos en forma simultánea, emplearemos la notación $V(G)$, $\Sigma(G)$, i_G y t_G para indicar conjuntos de nodos, aristas y funciones inicial y terminal de un grafo G en particular.

Definición 2.12. Sea $G = (V, \Sigma, i, t)$ un grafo.

- Para nodos I, J de G , denotaremos por Σ_I al conjunto de todas las aristas de G que arrancan en I , es decir, $\Sigma_I = \{a \in \Sigma : i(a) = I\}$. Similarmente, Σ^J denotará al conjunto de todas las aristas de G que terminan en J , es decir, $\Sigma^J = \{a \in \Sigma : t(a) = J\}$. Reservamos el símbolo Σ_I^J para el conjunto de todas las aristas de G que empiezan en I y terminan en J , es decir $\Sigma_I^J = \{a \in \Sigma : i(a) = I \wedge t(a) = J\}$. Llamamos **grado de salida** de I a $|\Sigma_I|$, la cantidad de aristas que arrancan en I , y **grado de entrada** de J a $|\Sigma^J|$, la cantidad de aristas que llegan a J . La **matriz de adyacencia** A_G del grafo G se define como la matriz cuadrada de $|V| \times |V|$ indizada por V , cuyo elemento en la posición IJ es $A_{IJ} = |\Sigma_I^J|$. Es decir, A_{IJ} es la cantidad de aristas de I a J en G .
- Un **lazo** o **bucle** en G es una arista a tal que $i(a) = t(a)$.
- Un **camino** π en el grafo G es una sucesión finita de aristas $\pi = e_1 e_2 \dots e_k$ tal que $\forall i \in \{1, \dots, k-1\}, t(e_i) = i(e_{i+1})$. Definimos el **vértice inicial del camino** π como $i(\pi) = i(e_1)$, el vértice inicial de la primera arista del camino, y el **vértice final de** π como $t(\pi) = t(e_k)$, el vértice final de la última arista del camino. La **longitud** del camino π es la cantidad de aristas que contiene, denotada por $|\pi|$. Es decir, $|\pi| = k$. La sucesión vacía cumple también con la definición de camino, con longitud 0 y vértices inicial y final idénticos, pudiendo ser cualquiera de los nodos de G . Un **ciclo** en G es un camino π en G tal que $i(\pi) = t(\pi)$. Un **ciclo simple** en G es un ciclo $\pi = e_1 \dots e_k$ tal que los estados $i(e_1), \dots, i(e_k)$ son distintos.

Desde el punto de vista de la dinámica simbólica, interesa la *conectividad* de las aristas de un grafo, no los nombres de las mismas ni de los vértices del grafo. Por ello, grafos diferentes pueden en realidad representar un mismo objeto desde ese punto de vista.

Definición 2.13. Dados dos grafos G y H , un **homomorfismo** de G a H es un par de funciones (∂, Φ) con $\partial : V(G) \rightarrow V(H)$ y $\Phi : \Sigma(G) \rightarrow \Sigma(H)$ tales que $\forall e \in \Sigma(G), i_H(\Phi(e)) = \partial(i_G(e))$ y $t_H(\Phi(e)) = \partial(t_G(e))$. En caso de ser ambas funciones inyectivas, el homomorfismo se llama **inmersión** de G en H . Si ambas funciones son biyectivas, se llama **isomorfismo** entre G y H . Dos grafos se dicen **isomorfos** si hay un isomorfismo entre ellos, y en ese caso escribimos $G \cong H$. Un isomorfismo es, a los

finés prácticos, simplemente un cambio de nombre de nodos y aristas. La relación de ser isomorfos es una relación de equivalencia entre grafos.

Definición 2.14. Un **subgrafo** H del grafo G es un grafo tal que $V(H) \subseteq V(G)$, $\Sigma(H) \subseteq \Sigma(G)$, y las funciones i_H y t_H son, respectivamente, las restricciones de i_G y t_G a $\Sigma(H)$.

En términos de matrices de adyacencia, si H es subgrafo de G se tiene que $(A_H)_{I,J} \leq (A_G)_{I,J}$ para toda pareja de vértices $I, J \in V(H)$.

En cuanto a los isomorfismos, se tiene que dos grafos G y H son isomorfos si, y sólo si, existe una matriz de permutación P tal que $A_G = PA_HP^{-1}$. Recordemos que una matriz de permutación es una matriz cuadrada con entradas 0 o 1 que tiene exactamente un 1 en cada fila, y exactamente un 1 en cada columna. Es fácil demostrar si P es matriz de permutación, P también lo es, y $P^{-1} = P^T$. Si (∂, Φ) es un isomorfismo entre $G = (V, \Sigma, i_G, t_G)$ y $H = (W, \Sigma', i_H, t_H)$, se define P con filas indizadas según $V(G)$ y columnas indizadas según $V(H)$ mediante

$$P_{I,J} = \begin{cases} 1 & \text{si } \partial(I) = J, \\ 0 & \text{si } \partial(I) \neq J. \end{cases}$$

Resulta directo chequear que P es matriz de permutación. Además, teniendo en cuenta que $e \in \Sigma_I^J \iff \Phi(e) \in \Sigma_{\partial I}^{\partial J}$, a través de un cálculo rutinario se puede ver que $(A_G P)_{I,J} = (PA_H)_{I,J}$ para todos I, J , de modo que $A_G = PA_HP^{-1}$.

Recíprocamente, si G y H son grafos tales que admiten una matriz de permutación P tal que $A_G = PA_HP^{-1}$, considerando a las filas de P en el mismo orden que las de A_G y a las columnas de P en el mismo orden que las de A_H , se define $\partial I = J \iff P_{I,J} = 1$. De la condición $A_G = PA_HP^{-1}$, se deduce que $(A_G)_{I,J} = (A_H)_{\partial I, \partial J}$, con lo que los grafos G y H son isomorfos.

2.2.1. Shifts de aristas. A continuación definiremos el espacio shift generado por un grafo.

Definición 2.15. Sea G un grafo con matriz de adyacencia A . El **shift de aristas** de G , denotado por \mathbb{X}_G o \mathbb{X}_A , es el conjunto de todos los caminos bi-infinitos de G . Es decir:

$$\mathbb{X}_G = \{x = (x_k)_{k \in \mathbb{Z}} : \forall k \in \mathbb{Z}, x_k \in \Sigma \wedge t(x_k) = i(x_{k+1})\}.$$

Como puede verse, \mathbb{X}_G es un subconjunto del full shift $\Sigma^{\mathbb{Z}}$, que tiene como alfabeto el conjunto Σ de aristas del grafo.

Ejemplo 2.16. Sean G_1 y G_2 los grafos mostrados en la figura 2.

Entonces, \mathbb{X}_{G_1} es el full shift sobre el alfabeto $\{0, 1, \dots, r-1\}$, mientras que \mathbb{X}_{G_2} es el espacio shift $\mathbb{X}_{\{eg, fe, ff, gg\}}$ del ejemplo 2.4.

Justifiquemos que el uso del término “shift” para los shifts de aristas es adecuado.

Proposición 2.17. Si $G = (V, \Sigma, i, t)$ es un grafo, entonces \mathbb{X}_G es un espacio shift. Más aún, es un shift de tipo finito de memoria 1.

Demostración. Consideremos la colección finita

$$\mathcal{F} = \{ef \in \Sigma^2 : t(e) \neq i(f)\}$$

y mostremos que $\mathbb{X}_G = \mathbb{X}_{\mathcal{F}}$: Sea $x \in \mathbb{X}_G$, y $k \in \mathbb{Z}$. Debe ser $t(x_k) = i(x_{k+1})$ por la definición de \mathbb{X}_G . Luego $x_k x_{k+1} \notin \mathcal{F}$. Entonces, ningún bloque de x está en \mathcal{F} , y

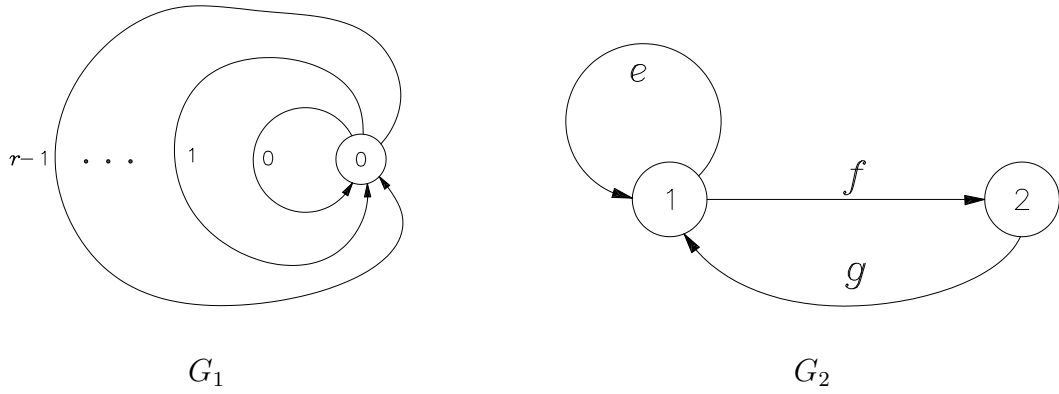


FIGURA 2.

$x \in \mathbb{X}_{\mathcal{F}}$. Recíprocamente, sea $x \in \mathbb{X}_{\mathcal{F}}$, y $k \in \mathbb{Z}$. $x_k x_{k+1} \notin \mathcal{F}$, por lo que $t(x_k) = i(x_{k+1})$. Como esto ocurre para todo $k \in \mathbb{Z}$, se ve que $x \in \mathbb{X}_G$.

Como \mathcal{F} tiene todos sus bloques de tamaño 2, \mathbb{X}_G es un shift de tipo finito de memoria 1. □

Es sencillo demostrar que si G y H son isomorfos, entonces \mathbb{X}_G y \mathbb{X}_H son conjugados: si (∂, Φ) es un isomorfismo entre G y H , Φ puede verse como la función inductora (biyectiva) de un código de ventana deslizante de memoria y anticipación 0, que, en vistas de la biyectividad de Φ , resultará ser una conjugación entre \mathbb{X}_G y \mathbb{X}_H .

También es fácil ver que si H es subgrafo de G , $\mathbb{X}_H \subseteq \mathbb{X}_G$.

Observemos el grafo de la figura 3. En él, la arista a no forma parte de ningún camino bi-infinito, pues desde $I = t(a)$ no arranca ninguna arista. Podemos pensar en el nodo I como en un “nodo muerto”. Una situación análoga ocurre con el nodo J , al cual no llega ninguna arista, de modo que las aristas b y c que arrancan en J tampoco pueden formar parte de un camino bi-infinito.

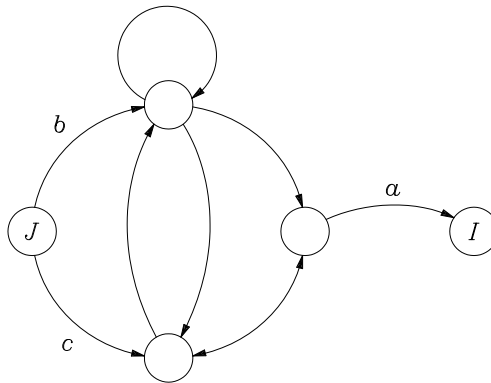


FIGURA 3.

Definición 2.18. Un nodo de un grafo se llama **nodo muerto** si a él no llega, o de él no sale, ninguna arista.

Los nodos muertos de un grafo G pueden eliminarse, conjuntamente con todas las aristas que a él llegan o salen, sin alterar el shift de aristas asociado. En el subgrafo que resulta de esta eliminación, pueden a su vez haber quedado nodos muertos. Sin embargo, repitiendo el proceso de eliminación sobre el subgrafo la cantidad de veces que sea necesaria hasta que no queden nodos muertos, se obtiene como resultado un

subgrafo del grafo original que contiene todas las aristas que forman parte de algún camino bi-infinito en G .

Ejemplo 2.19. La figura 4 muestra el resultado del proceso de eliminación de nodos muertos del grafo de la figura 3.

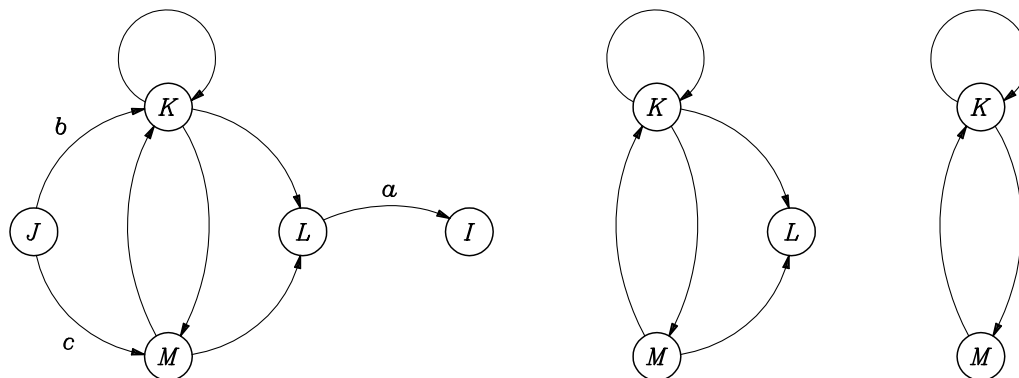


FIGURA 4.

Definición 2.20. Un grafo se llama **esencial** si no contiene nodos muertos.

Es fácil ver que si $u_1 \cdots u_k$ es un bloque de largo k que está en el lenguaje de \mathbb{X}_G , entonces u es un camino en G (si $u \sqsubseteq x \in \mathbb{X}_G$, existe $j \in \mathbb{Z}$ tal que $x_{[j+1, j+k]} = u_1 \cdots u_k$, de donde, para cualquier $n \in \{1, \dots, k-1\}$, es $t(u_n) = t(x_{j+n}) = i(x_{j+n+1}) = i(u_{n+1})$, por lo que $u_1 \cdots u_k$ es camino en G). Sin embargo, no es cierto que cualquier camino en un grafo G sea un bloque del lenguaje de \mathbb{X}_G . Por ejemplo, cualquier arista vinculada a un nodo muerto en un grafo no esencial es un camino de longitud 1 en G , pero no puede formar parte de un camino bi-infinito en G , por lo que no pertenece a $\mathcal{B}_1(\mathbb{X}_G)$.

Lema 2.21. Sea G un grafo esencial. Entonces, el conjunto de los caminos (finitos) en G es un precisamente el conjunto de bloques en el lenguaje de \mathbb{X}_G .

Demostración. Por nuestros comentarios anteriores, ya sabemos que los bloques en el lenguaje de \mathbb{X}_G son caminos en G . Probemos ahora que, bajo la hipótesis de que G es esencial, los caminos de G son bloques en el lenguaje de \mathbb{X}_G . Sea π un camino en G . Por ser G esencial, hay una arista a_1 que llega a $i(\pi)$ y una arista b_1 saliendo de $t(\pi)$. Inductivamente, para cada $k \geq 2$, hay una arista a_k llegando a $i(a_{k-1})$ y una arista b_{k-1} saliendo de $t(b_k)$. Entonces, $\cdots a_2 a_1 \pi b_1 b_2 \cdots$ es camino bi-infinito en G , y π ocurre en él. Luego, π está en $\mathcal{B}(\mathbb{X}_G)$. \square

Proposición 2.22. Si $G = (V(G), \Sigma(G), i_G, t_G)$ es un grafo, existe un único subgrafo H de G tal que H esencial y $\mathbb{X}_G = \mathbb{X}_H$.

Demostración. Sea $\Sigma(H) = \mathcal{B}_1(\mathbb{X}_G)$, es decir, el conjunto de todas las aristas de G que intervienen en algún camino bi-infinito en G . Hagamos $V(H) = \{i(a) : a \in \Sigma(H)\}$, el conjunto de nodos visitados por las aristas de H , y sean i_H y t_H las respectivas restricciones de i_G y t_G a $\Sigma(H)$. Entonces H es subgrafo de G , y, en consecuencia, cualquier camino bi-infinito en H es también camino bi-infinito en G , así que $\mathbb{X}_H \subseteq \mathbb{X}_G$.

Veamos que también $\mathbb{X}_G \subseteq \mathbb{X}_H$: si $x \in \mathbb{X}_G$, se tiene que para todo $k \in \mathbb{Z}$, x_k y x_{k+1} están ambos en $\mathcal{B}_1(\mathbb{X}_G)$, y además $t_G(x_k) = i_G(x_{k+1})$. Luego, para todo $k \in \mathbb{Z}$, es $x_k \in \Sigma(H)$ y además $t_H(x_k) = t_G(x_k) = i_G(x_{k+1}) = i_H(x_{k+1})$, de donde $x \in \mathbb{X}_H$.

Además, H es esencial, pues si $I \in V(H)$, debe haber una arista $a \in \Sigma(H)$ tal que $I = i_G(a)$. Como $a \in \mathcal{B}_1(\mathbb{X}_G)$, existe $x \in \mathbb{X}_G$ tal que $x_k = a$ para algún $k \in \mathbb{Z}$. Entonces $t_H(x_{k-1}) = t_G(x_{k-1}) = i_G(x_k) = i_H(x_k) = I$, mostrando que la arista x_{k-1} llega a I y la arista x_k sale desde I .

Para ver que H es único, consideremos cualquier subgrafo esencial H' de G tal que $\mathbb{X}_{H'} = \mathbb{X}_G$. Si hubiera una arista en H que no esté en H' o una en H' que no esté en H , sería $\mathcal{B}_1(\mathbb{X}_H) \neq \mathcal{B}_1(\mathbb{X}_{H'})$ (lema 2.21, considerando que H y H' son esenciales), por lo que $\mathcal{B}(\mathbb{X}_G) = \mathcal{B}(\mathbb{X}_H) \neq \mathcal{B}(\mathbb{X}_{H'})$, es decir, sería $\mathbb{X}_G \neq \mathbb{X}_{H'}$, y tendríamos una contradicción. Luego, H y H' tienen las mismas aristas, y siendo ambos esenciales, tienen también los mismos nodos, de donde concluimos que ambos son iguales. \square

Ya que el subgrafo H de la proposición anterior contiene la única parte de G usada para la dinámica simbólica, usualmente restringiremos nuestra atención a los grafos esenciales.

La matriz de adyacencia de un grafo puede usarse para obtener información sobre la cantidad de caminos de longitud $n \geq 0$ en el grafo, y, consecuentemente, si el grafo es esencial, para obtener la cantidad de bloques de tamaño n en $\mathcal{B}(\mathbb{X}_G)$, esto es, $|\mathcal{B}_n(\mathbb{X}_G)|$.

Proposición 2.23. *Sea G un grafo con matriz de adyacencia A , y sea $n \geq 0$. Entonces:*

1. *Para vértices I, J cualesquiera en G , el número de caminos de longitud n desde I hasta J en G es $(A^n)_{IJ}$, es decir, la posición IJ de la matriz A^n .*
2. *El número de ciclos de longitud n en G es $\text{tr}(A^n)$, la traza de A^n , y es igual al número de puntos periódicos de \mathbb{X}_G de período n .*

Demostración. 1. Probamos por inducción sobre n . Sean $I, J \in V(G)$. Si $I = J$, $(A^0)_{IJ} = 1$, y el único camino de I a J de longitud 0 es el camino vacío; si $I \neq J$, $(A^0)_{IJ} = 0$ y no hay camino en G de I a J de longitud 0. Luego, la proposición es verdadera para $n = 0$. Supongamos que la proposición es cierta para n . Un camino de I a J de longitud $n + 1$ consiste en un camino de longitud n de I a K seguido de una arista desde K hasta J , donde K es algún nodo. La cantidad de caminos de longitud $n + 1$ de I a J cuyo penúltimo vértice visitado es K es entonces la cantidad de caminos de longitud n de I a K multiplicada por la cantidad de aristas de K a J , que por hipótesis inductiva y definición de matriz de adyacencia es $(A^n)_{IK} A_{KJ}$. Para obtener todos los caminos desde I a J en G de longitud $n + 1$ hacemos variar a K en el conjunto de nodos de G y sumamos. Es decir, la cantidad de caminos de longitud $n + 1$ de I a J es

$$\sum_{K \in V(G)} (A^n)_{IK} A_{KJ}$$

que es precisamente la definición de $(A^{n+1})_{IJ}$, con lo que queda demostrada la primera parte.

2. La primera parte del enunciado se sigue del ítem anterior y de la definición de ciclo. Para la segunda parte, notar que si π es un ciclo de longitud n en G , entonces π^∞ es un punto periódico en \mathbb{X}_G de período n , mientras que si $x \in \mathbb{X}_G$ es un punto periódico de período n , entonces $x_{[0, n-1]}$ debe ser un ciclo de longitud n en G . Esto establece una correspondencia 1 a 1 entre los ciclos en G de longitud n y los puntos en \mathbb{X}_G de período n . \square

Recordemos que un espacio shift X se dice irreducible si para cualquier pareja de bloques $u, w \in \mathcal{B}(X)$, existe $v \in \mathcal{B}(X)$ tal que $uvw \in \mathcal{B}(X)$. Es fácil caracterizar a los grafos que dan lugar a shifts de aristas irreducibles.

Definición 2.24. Un grafo G se dice **irreducible** si para cualquier pareja I, J de vértices en G , existe un camino π en G tal que $i(\pi) = I$ y $t(\pi) = J$. Si G no es irreducible, se llama **reducible**.

La figura 5 muestra un ejemplo de grafo reducible, ya que no hay en él un camino desde I hasta J .

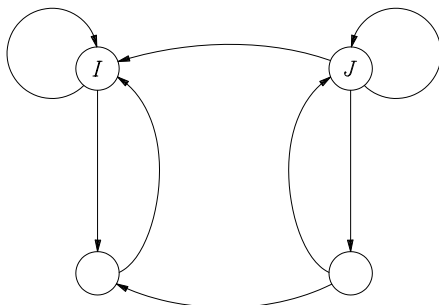


FIGURA 5.

2.3. Representación de shifts de tipo finito por medio de grafos. Hemos presentado, en la sección anterior, los shifts de aristas, que resultaron ser STF de memoria 1. Por lo tanto, un STF que no tenga memoria 1 no puede ser un shift de aristas. Más aún, el siguiente ejemplo muestra que hay STF de memoria 1 que no son shifts de aristas.

Ejemplo 2.25. No existe un grafo G tal que \mathbb{X}_G sea el shift de la razón de oro. Para verlo, supongamos que G cumple que $\mathbb{X}_G = \mathbb{X}_{\{11\}}$. Por prop. 2.22, podríamos suponer que G es esencial. Entonces debería contener exactamente dos aristas llamadas 0 y 1. La arista 1 debería iniciar en un nodo y terminar en otro distinto (de lo contrario, 1^∞ estaría en \mathbb{X}_G , lo que no puede ser) y entonces la arista 0 debería ir desde $t(1)$ hasta $i(1)$ (pues en otro caso G no sería esencial). Pero entonces $\mathbb{X}_G \neq \mathbb{X}_{\{11\}}$, contradicción que proviene de suponer la existencia de tal G .

A pesar de que los shifts de aristas aparecen como particulares shifts de tipo finito, mostraremos ahora que todo shift de tipo finito puede recodificarse a un shift de aristas.

Teorema 2.26. *Si X es un shift de tipo finito de memoria M , entonces $X^{[M+1]}$ es un shift de aristas. En consecuencia, todo shift de tipo finito es conjugado a un shift de aristas.*

Demostración. Sea X un STF de memoria M . Podemos suponer $M \geq 1$. Sea \mathcal{F} tal que $X = \mathbb{X}_{\mathcal{F}}$, siendo todos los bloques en \mathcal{F} de largo $M + 1$. Hagamos $V = \mathcal{B}_M(X)$ y $\Sigma = \mathcal{B}_{M+1}(X)$. Para $a = a_1 a_2 \cdots a_M a_{M+1} \in \Sigma$, definamos $i(a) = a_1 \cdots a_M$ y $t(a) = a_2 \cdots a_M a_{M+1}$. Puede verse que $i(a)$ y $t(a)$ están en V , ya que son subbloques de $a_1 \cdots a_M a_{M+1} \in \mathcal{B}(X)$. Definamos $G = (V, \Sigma, i, t)$. Notar que, dados vértices $I = a_1 \cdots a_M$ y $J = b_1 \cdots b_M$, hay una única arista de I a J si, y sólo si, $a_2 \cdots a_M = b_1 \cdots b_{M-1}$ y $a_1 \cdots a_M b_M \in \mathcal{B}_{M+1}(X)$. Es decir, hay arista de I a J si, y sólo si, I y J solapan progresivamente y el bloque obtenido de agregar a I el último símbolo de J está en $\mathcal{B}_{M+1}(X)$.

Verifiquemos ahora que $\mathbb{X}_G = X^{[M+1]}$.

Sea $\bar{x} \in \mathbb{X}_G$. Se tiene que, para todo $k \in \mathbb{Z}$, $\bar{x}_k \in \Sigma$ (es decir, \bar{x}_k es un bloque de tamaño $M + 1$ en el lenguaje de X), y $t(\bar{x}_k) = i(\bar{x}_{k+1})$. Pero si $\bar{x}_k = a_1 \cdots a_{M+1}$ y

$\bar{x}_{k+1} = b_1 \cdots b_{M+1}$, es $t(\bar{x}_k) = a_2 \cdots a_{M+1}$ e $i(\bar{x}_{k+1}) = b_1 \cdots b_M$. Se tiene entonces que \bar{x}_k y \bar{x}_{k+1} solapan progresivamente. Definamos entonces $x = \beta_{M+1}^{-1}(\bar{x})$. Para cualquier $k \in \mathbb{Z}$, $x_{[k, k+M]} = \bar{x}_k \in \mathcal{B}_{M+1}(X)$, por lo que x no tiene bloques en \mathcal{F} , es decir, $x \in X$, y entonces $\bar{x} = \beta_{M+1}(x)$ pertenece a $X^{[M+1]}$.

Ahora sea $\bar{x} \in X^{[M+1]}$. Sea x tal que $\beta_{M+1}(x) = \bar{x}$. Se tiene que, para cualquier $k \in \mathbb{Z}$, es $\bar{x}_k = x_k \cdots x_{k+M}$ y $\bar{x}_{k+1} = x_{k+1} \cdots x_{k+M+1}$, por lo que $\bar{x}_k \in \mathcal{B}_{M+1}(X) = \Sigma$ y $t(\bar{x}_k) = x_{k+1} \cdots x_{k+M} = i(\bar{x}_{k+1})$ de acuerdo a la definición de G . Entonces, $\bar{x} \in \mathbb{X}_G$.

La última afirmación del teorema se deduce del hecho de que $X \sim X^{[M+1]}$. \square

Ejemplo 2.27. Si $X = \mathbb{X}_{\{11\}}$, X es STF de memoria 1. Tenemos que $\mathcal{B}_1(X) = \{0, 1\} = V$ y que $\mathcal{B}_2(X) = \{00, 01, 10\} = \Sigma$. Definimos $i(00) = i(01) = 0$, $i(10) = 1$, $t(00) = t(10) = 0$ y $t(01) = 1$. El grafo (V, Σ, i, t) mostrado en la figura 6 es precisamente $X^{[2]}$.

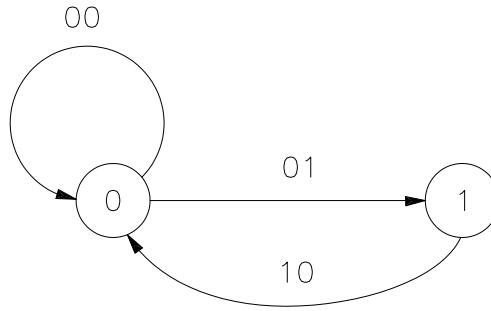


FIGURA 6.

Un camino bi-infinito en un grafo G determina una sucesión bi-infinita de vértices de ese grafo, a saber, los vértices que el camino bi-infinito atraviesa (pero a una sucesión bi-infinita de vértices puede corresponderle más de un camino bi-infinito, a menos que el grafo no posea *aristas en paralelo* entre dos nodos). Formalmente, sea $G = (V, \Sigma, i, t)$ un grafo, y sea $(x_k)_{k \in \mathbb{Z}}$ un punto de \mathbb{X}_G . Entonces, la sucesión bi-infinita de vértices que este camino bi-infinito determina es el punto $(i(x_k))_{k \in \mathbb{Z}}$, que pertenece al full shift $V^{\mathbb{Z}}$.

Para considerar estos recorridos bi-infinitos, introducimos la *matriz de incidencia* de un grafo.

Definición 2.28. Sea $G = (V, \Sigma, i, t)$ un grafo con matriz de adyacencia A_G . La **matriz de incidencia** de G es la matriz B , de tamaño $|V| \times |V|$ con filas y columnas indizadas por V , tal que, para vértices $m, n \in V$,

$$B_{mn} = \begin{cases} 1 & \text{si } (A_G)_{mn} > 0, \\ 0 & \text{si } (A_G)_{mn} = 0. \end{cases}$$

Es decir, la entrada (m, n) de la matriz de incidencia es 1 o 0, dependiendo de si hay o no arista en G desde m hacia n (en caso de haber, no importa cuántas hay).

Ahora consideraremos el conjunto de todas las sucesiones bi-infinitas de vértices que determinan los caminos bi-infinitos en G , y que resultará ser también un espacio shift.

Definición 2.29. Sea $G = (V, \Sigma, i, t)$ un grafo con matriz de incidencia B . El **shift de vértices** de G es el conjunto

$$\hat{\mathbb{X}}_G = \{(x_k)_{k \in \mathbb{Z}} \in V^{\mathbb{Z}} : \forall k \in \mathbb{Z}, B_{x_k x_{k+1}} = 1\}.$$

Proposición 2.30. *Los shifts de vértices son shifts de tipo finito de memoria 1.*

Demostración. Sea $G = (V, \Sigma, i, t)$ un grafo con matriz de incidencia B . Consideremos $\mathcal{F} = \{mn \in V^2 : B_{mn} = 0\}$, y veamos que $\hat{\mathbb{X}}_G = \mathbb{X}_{\mathcal{F}}$.

Sea $x \in \hat{\mathbb{X}}_G$, y consideremos $j, k \in \mathbb{Z}$. Si $k - j + 1 \neq 2$, $x_{[j,k]} \notin \mathcal{F}$ (pues \mathcal{F} contiene sólo palabras de largo 2). Y si $k - j + 1 = 2$, es $x_{[j,k]} = x_j x_{j+1}$, pero, por estar x en $\hat{\mathbb{X}}_G$, es $B_{x_j x_{j+1}} = 1$, por lo que $x_{[j,k]} \notin \mathcal{F}$. Entonces ningún bloque de x pertenece a \mathcal{F} , es decir, $x \in \mathbb{X}_{\mathcal{F}}$.

Recíprocamente, si $x \in \mathbb{X}_{\mathcal{F}}$, se tiene que, para todo $j \in \mathbb{Z}$, $x_{[j,j+1]} \notin \mathcal{F}$, de donde $B_{x_j x_{j+1}} = 1$, por lo que $x \in \hat{\mathbb{X}}_G$.

Ya que todos los bloques de \mathcal{F} tienen largo 2, se sigue que \hat{X}_G es un STF de memoria 1. \square

Proposición 2.31.

1. La familia de todos los shifts de tipo finito de memoria 1 es la familia de todos los shifts de vértices.
2. Cualquier shift de aristas es un shift de vértices (en un grafo diferente).
3. Si X es un shift de tipo finito de memoria M , $X^{[M]}$ es un shift de vértices. De hecho, existe un grafo G tal que $\hat{\mathbb{X}}_G = X^{[M]}$ y $\mathbb{X}_G = X^{[M+1]}$.

Demostración. 1. La proposición 2.30 muestra que cualquier shift de vértices es un SFT de memoria 1. Ahora, sea $\mathbb{X}_{\mathcal{F}}$ un espacio shift con todos los bloques en \mathcal{F} de largo 2. Definamos un grafo G como sigue: su conjunto de vértices es $V = \mathcal{B}_1(\mathbb{X}_G)$, y para $m, n \in V$, pongamos una única arista mn de m a n si, y sólo si, $mn \notin \mathcal{F}$. La matriz de incidencia de G tendrá entonces, en la posición (m, n) , un 1 si $mn \notin \mathcal{F}$, y un 0 si $mn \in \mathcal{F}$. Luego, un punto de $V^{\mathbb{Z}}$ está en $\hat{\mathbb{X}}_G$ precisamente cuando para todo $k \in \mathbb{Z}$, $B_{x_k x_{k+1}} = 1$, o, equivalentemente, cuando $x_k x_{k+1} \notin \mathcal{F}$, mostrando que $\hat{\mathbb{X}}_G = \mathbb{X}_{\mathcal{F}}$.

2. El enunciado es consecuencia directa del ítem anterior y del hecho de que todo shift de aristas es un STF de memoria 1.

3. Veamos que el grafo G construido en la prueba del teorema 2.26, aparte de satisfacer que $\mathbb{X}_G = X^{[M+1]}$, cumple que $\hat{\mathbb{X}}_G = X^{[M]}$.

Recordemos que el grafo G construido tiene conjunto de vértices $V = \mathcal{B}_M(X)$, y entre los vértices $I = a_1 \cdots a_M$ y $J = b_1 \cdots b_M$ hay arista si, y sólo si, $a_2 \cdots a_M = b_1 \cdots b_{M-1}$ y $a_1 \cdots a_M b_M \in \mathcal{B}_{M+1}(X)$. Es decir, llamando B a la matriz de incidencia, se tiene que $B_{IJ} = 1$ si, y sólo si, I y J solapan progresivamente y el bloque obtenido de agregar a I el último símbolo de J está en $\mathcal{B}_{M+1}(X)$.

Recordemos también que es $X = \mathbb{X}_{\mathcal{F}}$ para alguna colección \mathcal{F} de bloques de tamaño $M + 1$.

Sea $\bar{x} \in \hat{\mathbb{X}}_G$. Se tiene que, para todo $k \in \mathbb{Z}$, $\bar{x}_k \in V$, y $B_{\bar{x}_k \bar{x}_{k+1}} = 1$, de donde \bar{x}_k y \bar{x}_{k+1} solapan progresivamente. Definamos $x = \beta_M^{-1}(\bar{x})$. Se tiene que, para cualquier $k \in \mathbb{Z}$, $x_k \cdots x_{k+M} \in \mathcal{B}_{M+1}(X)$ (pues es el bloque que resulta de agregar a \bar{x}_k el último símbolo de \bar{x}_{k+1}), es decir, $x_k \cdots x_{k+M} \notin \mathcal{F}$, y entonces $x \in X$, así que $\bar{x} \in X^{[M]}$.

Ahora sea $\bar{x} \in X^{[M]}$. Sea x tal que $\beta_M(x) = \bar{x}$. Se tiene que, para cualquier $k \in \mathbb{Z}$, es $\bar{x}_k = x_k \cdots x_{k+M-1} \in \mathcal{B}_M(X)$, $\bar{x}_{k+1} = x_{k+1} \cdots x_{k+M} \in \mathcal{B}_M(X)$, y $x_k \cdots x_{k+M} \in \mathcal{B}_{M+1}(X)$, por lo que $B_{\bar{x}_k \bar{x}_{k+1}} = 1$. Luego, $\bar{x} \in \hat{\mathbb{X}}_G$. \square

3. ASPECTOS TOPOLÓGICOS Y DINÁMICOS DE LOS ESPACIOS SHIFT

3.1. Introducción. El objetivo de este capítulo es analizar a los espacios shift a través de sus características topológicas. Mediante la definición de una manera adecuada para medir distancias entre tirillas biinfinitas, podremos ver a la full shift como un

espacio métrico y estudiarla con las herramientas clásicas de la topología y el análisis. De ésto, resultará una caracterización de los espacios shift y de los códigos de ventana deslizante a través de propiedades topológicas. Esto forma parte de investigaciones más o menos recientes (alrededor de 1970) en el área de la Dinámica Simbólica, de modo que se trata de tópicos relativamente nuevos en Matemática, y en franco proceso de expansión.

No se pretende dar aquí un estudio detallado de la teoría general de espacios métricos, sino más bien enunciar las definiciones y propiedades básicas de los mismos con el único objetivo de aplicarlos al estudio de los espacios shift.

En el presente capítulo, primero definiremos una métrica para la full shift y veremos algunas propiedades básicas importantes de la misma, para luego estudiar la convergencia de sucesiones y arribar a la conclusión de que los espacios shift están caracterizados por ser subconjuntos compactos y shift-invariantes de la full shift. Luego analizaremos la continuidad de acuerdo a la métrica definida, encontrando que los códigos de ventana deslizante se caracterizan por ser las funciones entre espacios shift que son continuas y conmutan con la función shift.

3.2. Una métrica para $\mathcal{A}^{\mathbb{Z}}$.

Definición 3.1. Una **métrica** o **función distancia** para un conjunto X es una función $d : X \times X \rightarrow \mathbb{R}_0^+$ con las siguientes propiedades:

1. $d(x, y) = 0 \Leftrightarrow x = y$,
2. $\forall x, y \in X, d(x, y) = d(y, x)$ (axioma de simetría),
3. $\forall x, y, z \in X, d(x, z) \leq d(x, y) + d(y, z)$ (desigualdad triangular).

El par (X, d) se denomina **espacio métrico**.

Para llevar adelante el objetivo planteado en la introducción, debemos dotar a $\mathcal{A}^{\mathbb{Z}}$ de una métrica adecuada. Para nuestros propósitos, “adecuada” va a querer decir que dos puntos x e y de $\mathcal{A}^{\mathbb{Z}}$ estarán tanto más cerca cuanto más grande sea el bloque central en que coinciden. Así, por ejemplo, $\dots 00010,110101110000\dots$ y $\dots 10010,110101110010\dots$ deberán estar entre sí a una distancia menor que la distancia que haya entre los puntos $\dots 00010,1101110000\dots$ y $\dots 00011,11111111\dots$. Para ello, definiremos $d : \mathcal{A}^{\mathbb{Z}} \times \mathcal{A}^{\mathbb{Z}} \rightarrow \mathbb{R}_0^+$ mediante

$$d(x, y) = \begin{cases} 0 & \text{si } x = y, \\ 2^{-\min\{|k|:x_k \neq y_k\}} & \text{si } x \neq y. \end{cases}$$

En palabras, para encontrar la distancia entre dos puntos distintos x e y , hay que detectar la coordenada entera k (positiva o negativa) en la que x e y difieren más cercana a la coordenada 0, y evaluar $2^{-|k|}$.

Ejemplo 3.2. Encontremos la distancia entre los puntos

$$x = \dots 00010,110101110000\dots \qquad y = \dots 10010,110101110010\dots$$

A fin de visualizar mejor, dispongamos ambos puntos uno debajo del otro, y marquemos con \times las coordenadas en que $x_k \neq y_k$:

| | | | | | | | | | | | | | | | | | | | |
|-----|-----|----------|----|----|----------|----|---|---|---|---|---|---|---|---|---|---|----------|----|-----|
| | | -5 | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | |
| x | ... | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | ... |
| | | \times | | | \times | | | | | | | | | | | | \times | | |
| y | ... | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | ... |

La marca \times más cercana a la coordenada 0 corresponde a $k = -2$, de modo que $d(x, y) = 2^{-|-2|} = 1/4$.

Notemos que, de acuerdo a esta definición, el mayor valor que puede tomar d es 1, que corresponde al caso en que los dos puntos entre los que se está midiendo la distancia difieren en la coordenada 0.

Justifiquemos que la función d así definida es efectivamente una métrica para $\mathcal{A}^{\mathbb{Z}}$:

- Dados dos puntos x e y en $\mathcal{A}^{\mathbb{Z}}$, o bien son iguales o bien son distintos. En el primer caso, es $d(x, y) = 0 \in \mathbb{R}_0^+$. En el caso en que $x \neq y$, el conjunto $\{k : x_k \neq y_k\} \subseteq \mathbb{Z}$ es no vacío, de modo que el conjunto $\{|k| : x_k \neq y_k\} \subseteq \mathbb{N}$ tampoco es vacío, y entonces tiene un mínimo elemento N . Ya que la función exponencial 2^{-t} es estrictamente positiva para $t \in \mathbb{R}$, $d(x, y) = 2^{-N} \in \mathbb{R}_0^+$. Esto muestra también que $d(x, y) = 0 \Leftrightarrow x = y$.
- El cumplimiento del axioma de simetría es directo de ver.
- Para la desigualdad triangular, sean $x, y, z \in \mathcal{A}^{\mathbb{Z}}$ dos a dos distintos. Llamemos

$$N_{xy} = \min\{|k| : x_k \neq y_k\}, \quad N_{yz} = \min\{|k| : y_k \neq z_k\}, \quad N_{xz} = \min\{|k| : x_k \neq z_k\}.$$

Tomemos $N = \min\{N_{xy}, N_{yz}\}$.

- Si $N = 0$, es $y_0 \neq x_0$ o $y_0 \neq z_0$, de modo que $d(x, y) = 1$ o $d(y, z) = 1$, y por lo tanto es $d(x, y) + d(y, z) \geq 1 \geq d(x, z)$.
- Si $N > 0$, es $x_{[-N+1, N-1]} = z_{[-N+1, N-1]}$, pues si tomamos i tal que $|i| < N$, tenemos que $|i| < N_{xy}$ y $|i| < N_{yz}$, de donde $x_i = y_i$ e $y_i = z_i$, por lo que $x_i = z_i$. O sea que $N_{xz} \geq N$. Por lo tanto, observando que $2^{-N_{xz}} \leq 2^{-N}$ y que es $N = N_{xy}$ o $N = N_{yz}$, se tiene

$$d(x, z) = 2^{-N_{xz}} \leq 2^{-N} \leq 2^{-N_{xy}} + 2^{-N_{yz}} = d(x, y) + d(y, z).$$

Es decir, en cualquier caso se verifica la desigualdad triangular.

La función distancia puede restringirse de $\mathcal{A}^{\mathbb{Z}} \times \mathcal{A}^{\mathbb{Z}}$ a $X \times X$ para cualquier subconjunto X de la full shift, y el par (X, d) sigue siendo un espacio métrico.

La métrica recientemente definida recibe el nombre de **métrica usual para los espacios shift**, y es a la que haremos referencia por defecto siempre que hablemos de distancia en espacios shift, a menos que especifiquemos lo contrario.

Una condición necesaria y suficiente para que dos puntos distintos estén a distancia menor que un ε estrictamente positivo es que ambos coincidan en una ventana central de tamaño $2k + 1$, en donde $k = \lfloor \log \frac{1}{\varepsilon} \rfloor$ (a menos que especifiquemos otra cosa, log siempre significará logaritmo en base 2; además, para $r \in \mathbb{R}$, $\lfloor r \rfloor$ es la parte entera o piso de r , es decir, el mayor entero que no supera a r). Para probar la afirmación, notemos que

$$\begin{aligned} d(x, y) < \varepsilon &\Leftrightarrow 2^{-\min\{|k| : x_k \neq y_k\}} < \varepsilon \Leftrightarrow \min\{|k| : x_k \neq y_k\} > \log \frac{1}{\varepsilon} \\ &\Leftrightarrow \min\{|k| : x_k \neq y_k\} > \left\lfloor \log \frac{1}{\varepsilon} \right\rfloor \Leftrightarrow x_{[-\lfloor \log \frac{1}{\varepsilon} \rfloor, \lfloor \log \frac{1}{\varepsilon} \rfloor]} = y_{[-\lfloor \log \frac{1}{\varepsilon} \rfloor, \lfloor \log \frac{1}{\varepsilon} \rfloor]}. \end{aligned}$$

Juntando esto con el hecho de que dos puntos de la full shift que están a distancia 0 coinciden en el bloque central de tamaño $2k + 1$ cualquiera sea $k \in \mathbb{N}$, tenemos el siguiente resultado.

Proposición 3.3. *Sea $\varepsilon > 0$, y x e y dos puntos de $\mathcal{A}^{\mathbb{Z}}$. Se tiene que*

$$d(x, y) < \varepsilon \iff x_{[-\lfloor \log \frac{1}{\varepsilon} \rfloor, \lfloor \log \frac{1}{\varepsilon} \rfloor]} = y_{[-\lfloor \log \frac{1}{\varepsilon} \rfloor, \lfloor \log \frac{1}{\varepsilon} \rfloor]}.$$

En particular, para cualquier $K \in \mathbb{N}$, tenemos que $d(x, y) < 2^{-K} \Leftrightarrow x_{[-K, K]} = y_{[-K, K]}$.

Definición 3.4. Sea (X, d) un espacio métrico, $x \in X$ y $r \in \mathbb{R}^+$. El **entorno de radio r con centro en x** , denotado $B_r(x)$, es el conjunto de puntos del espacio cuya distancia a x es menor que r .

En virtud de la Proposición 3.3, el entorno de radio ε con centro en $x \in X \subseteq \mathcal{A}^{\mathbb{Z}}$, para la métrica usual, es

$$B_\varepsilon(x) = \left\{ y \in X : x_{[-\lfloor \log \frac{1}{\varepsilon} \rfloor, \lfloor \log \frac{1}{\varepsilon} \rfloor]} = y_{[-\lfloor \log \frac{1}{\varepsilon} \rfloor, \lfloor \log \frac{1}{\varepsilon} \rfloor]} \right\}.$$

Particularmente, para $K \in \mathbb{N}$, $B_{2^{-K}}(x) = \{y \in X : x_{[-K, K]} = y_{[-K, K]}\}$.

Definición 3.5. Dado un subconjunto E de un espacio métrico (X, d) , se dice que un punto $x \in E$ es un **punto interior a E** si existe $r > 0$ tal que $B_r(x) \subseteq E$. E es **abierto en X** si todos sus puntos son puntos interiores a E .

Como caso particular, los entornos de cualquier espacio métrico son conjuntos abiertos. Puede demostrarse que cualquier conjunto abierto de un espacio métrico es una unión de entornos, y viceversa. Además, la unión arbitraria de conjuntos abiertos es un conjunto abierto, y la intersección finita de conjuntos abiertos es también un conjunto abierto.

En el caso particular de la full shift, a partir de la definición de conjunto abierto y de la caracterización de los entornos, se deduce que un subconjunto X de $\mathcal{A}^{\mathbb{Z}}$ es abierto si, y sólo si, para cada $x \in X$, existe $K \in \mathbb{N}$ tal que cualquier punto $y \in \mathcal{A}^{\mathbb{Z}}$ que satisface $y_{[-K, K]} = x_{[-K, K]}$ pertenece también a X . Tal K dependerá, en general, de x .

Definición 3.6. Dado un subconjunto E de un espacio métrico (X, d) , se dice que un punto $x \in X$ es un **punto de acumulación de E** si todo entorno con centro en x tiene un punto de $E - \{x\}$. E es **cerrado** si contiene a todos sus puntos de acumulación. La **clausura de E** es la unión de E con el conjunto de sus puntos de acumulación.

Se tiene que E es cerrado si, y sólo si, su complemento es abierto. Además, E es cerrado si, y sólo si, la clausura de E coincide con E . Puede verse, usando las leyes de De Morgan, que la intersección arbitraria de conjuntos cerrados es un conjunto cerrado, así como la unión finita de conjuntos cerrados es también un conjunto cerrado.

3.3. Sucesiones. Una sucesión en un espacio métrico (X, d) es una función de \mathbb{N} en X . Denotaremos por x_n al n -ésimo término de la sucesión, y por $\{x_n\}_{n \in \mathbb{N}}$ a la sucesión completa. Cuando el espacio métrico en cuestión sea un espacio shift, denotaremos $x^{(n)}$ al n -ésimo término de la sucesión, pues seguiremos usando x_n para referirnos al símbolo que haya en la n -ésima coordenada de la tirilla biinfinita x .

Definición 3.7. Una sucesión $\{x_n\}_{n \in \mathbb{N}}$ en el espacio (X, d) se dice **convergente en X** si existe $p \in X$ tal que $\lim_{n \rightarrow \infty} d(x_n, p) = 0$. En este caso, escribimos $\lim_{n \rightarrow \infty} x_n = p$, o también $x_n \xrightarrow[n \rightarrow \infty]{} p$.

Obsérvese que $\lim_{n \rightarrow \infty} x_n = p$ si, y sólo si, $\forall \varepsilon > 0, \exists N \in \mathbb{N} : \forall n \geq N, d(x_n, p) < \varepsilon$.

Caracterizaremos ahora la convergencia de sucesiones en espacios shift. Básicamente, una sucesión $\{x^{(n)}\}_{n \in \mathbb{N}}$ en un espacio shift converge a $x \in \mathcal{A}^{\mathbb{Z}}$ cuando para cualquier K natural, por grande que sea, el bloque central de tamaño $2K + 1$ de x coincide con el bloque central de tamaño $2K + 1$ de $x^{(n)}$, para todo n suficientemente grande.

Proposición 3.8. Sea $\{x^{(n)}\}_{n \in \mathbb{N}}$ una sucesión en el espacio shift X con la métrica usual, y $x \in \mathcal{A}^{\mathbb{Z}}$. Se tiene que $\lim_{n \rightarrow \infty} x^{(n)} = x$ si, y sólo si, $\forall K \in \mathbb{N}, \exists N \in \mathbb{N} : \forall n \geq N, x_{[-K, K]}^{(n)} = x_{[-K, K]}$.

Demostración. Dado que la función exponencial a tasa menor que 1 tiende asintóticamente a 0, por la Proposición 3.3 tenemos que:

$$\begin{aligned} \lim_{n \rightarrow \infty} x^{(n)} = x &\iff \forall \varepsilon > 0, \exists N \in \mathbb{N} : \forall n \geq N, d(x^{(n)}, x) < \varepsilon \\ &\iff \forall K \in \mathbb{N}, \exists N \in \mathbb{N} : \forall n \geq N, d(x^{(n)}, x) < 2^{-K} \\ &\iff \forall K \in \mathbb{N}, \exists N \in \mathbb{N} : \forall n \geq N, x_{[-K, K]}^{(n)} = x_{[-K, K]}. \end{aligned}$$

□

Ejemplo 3.9. Estudiar la convergencia de la sucesión $\{x^{(n)}\}_{n \in \mathbb{N}}$ en $\{0, 1\}^{\mathbb{Z}}$ definida por

$$x_i^{(n)} = \begin{cases} 1 & \text{si } |i| = n, \\ 0 & \text{si } |i| \neq n. \end{cases}$$

Veamos el siguiente cuadro:

| | | | | | | | | | | | |
|-----------|-----|----|----|----|----|---|---|---|---|---|-----|
| | | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | |
| $x^{(0)}$ | ... | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | ... |
| $x^{(1)}$ | ... | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | ... |
| $x^{(2)}$ | ... | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | ... |
| $x^{(3)}$ | ... | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | ... |
| | | ⋮ | | | | | | | | | |

Rápidamente nos damos cuenta que la ventana central de tamaño $2K + 1$ de $x^{(n)}$ consta de puros 0, para todo $n \geq K + 1$. Luego, de acuerdo al criterio de convergencia en espacios shift, el límite de la sucesión es el punto 0^∞ .

Ejemplo 3.10. Estudiar la convergencia de la sucesión $\{x^{(n)}\}_{n \in \mathbb{N}}$ en $\{0, 1\}^{\mathbb{Z}}$ definida por

$$x^{(n)} = \begin{cases} (0001)^\infty & \text{si } n \text{ es par,} \\ (01)^\infty & \text{si } n \text{ es impar.} \end{cases}$$

Dispongamos los términos encolumnados, como antes:

| | | | | | | | | | | | |
|-----------|-----|----|----|----|----|---|---|---|---|---|-----|
| | | -4 | -3 | -2 | -1 | 0 | 1 | 2 | 3 | 4 | |
| $x^{(0)}$ | ... | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | ... |
| $x^{(1)}$ | ... | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | ... |
| $x^{(2)}$ | ... | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | ... |
| $x^{(3)}$ | ... | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | ... |
| | | ⋮ | | | | | | | | | |

Aquí vemos que la coordenada 1, por ejemplo, no se estabiliza nunca: para cualquier N natural y a en el alfabeto $\{0, 1\}$, existe $n > N$ tal que $x_1^{(n)} \neq a$. Entonces, esta sucesión no es convergente.

Definición 3.11. Sean $\{x_n\}_{n \in \mathbb{N}}$ y $\{z_k\}_{k \in \mathbb{N}}$ dos sucesiones en el espacio métrico (X, d) . Se dice que $\{z_k\}_{k \in \mathbb{N}}$ es **subsucesión de** $\{x_n\}_{n \in \mathbb{N}}$ si existe una sucesión de números naturales estrictamente creciente $\{n_k\}_{k \in \mathbb{N}}$ tal que para todo $k \in \mathbb{N}$, es $z_k = x_{n_k}$.

Definición 3.12. Sea (X, d) un espacio métrico. Un subconjunto K de X se dice **compacto** si toda sucesión en K posee subsucesión convergente en K .

Es bien sabido que un subconjunto K es compacto si, y sólo si, cualquier familia de abiertos que unidos cubren a K posee una subfamilia *finita* que también cubre a K . Además, todo subconjunto compacto de un espacio métrico es un conjunto cerrado, y si el espacio métrico es compacto, entonces la familia de los subconjuntos compactos del espacio es precisamente la de los subconjuntos cerrados.

El siguiente resultado muestra que cualquier espacio shift (en particular la full shift) es compacto para la métrica usual.

Proposición 3.13. *Sea X un espacio shift con la métrica usual. Entonces X es un subconjunto compacto de $\mathcal{A}^{\mathbb{Z}}$.*

Demostración. Sea $\{x^{(n)}\}_{n \in \mathbb{N}}$ una sucesión en X . Ya que \mathcal{A} es un conjunto finito, debe existir $a_0 \in \mathcal{A}$ y un subconjunto infinito B_0 de números naturales tal que para todo $n \in B_0$, $x_0^{(n)} = a_0$. Elijamos cualquier $n_0 \in B_0$. Como \mathcal{A}^3 es finito, deben existir a_{-1} y a_1 en \mathcal{A} , y un subconjunto infinito $B_1 \subseteq B_0$ tal que para todo $n \in B_1$, $x_{[-1,1]}^{(n)} = a_{-1}a_0a_1$. Elijamos $n_1 \in B_1$ tal que $n_1 > n_0$ (tal elección es posible, pues B_1 es un conjunto infinito de números naturales, y por tanto no acotado). Siguiendo de esta manera, para cada $k \in \mathbb{N}$, como \mathcal{A}^{2k+1} es finito, deben existir a_{-k} y a_k en \mathcal{A} , y un subconjunto infinito $B_k \subseteq B_{k-1}$ tal que para todo $n \in B_k$, $x_{[-k,k]}^{(n)} = a_{-k} \dots a_0 \dots a_k$. Elijamos $n_k \in B_k$ tal que $n_k > n_{k-1}$. De este modo, hemos definido (inductivamente) a_{-k} , a_k y n_k para todo $k \in \mathbb{N}$. Como $\{n_k\}_{k \in \mathbb{N}}$ es creciente, $\{x^{(n_k)}\}_{k \in \mathbb{N}}$ es subsucesión de $\{x^{(n)}\}_{n \in \mathbb{N}}$. Definamos $x = (a_k)_{k \in \mathbb{Z}}$. Dado que todo bloque de x ocurre en algún $x^{(n_k)} \in X$, tenemos que x pertenece al espacio shift X . Además, $\lim_{k \rightarrow \infty} x^{(n_k)} = x$, pues dado $K \in \mathbb{N}$, tomando $N = K$, tenemos que para todo $k \geq N$ es $x_{[-K,K]}^{(n_k)} = a_{-K}a_{-K+1} \dots a_0 \dots a_{K-1}a_K = x_{[-K,K]}$. Entonces X es compacto. \square

La sola compacidad no caracteriza completamente a los espacios shift: por ejemplo, el conjunto unitario $\{(01)^\infty\}$ es un subconjunto compacto de la full $\{0, 1\}$ -shift (como cualquier conjunto finito de cualquier espacio métrico) pero no es un espacio shift pues carece de la shift-invariancia. Precisamente, compacidad más shift-invariancia equivalen a tener carácter de espacio shift.

Teorema 3.14. *Un subconjunto de $\mathcal{A}^{\mathbb{Z}}$ es un espacio shift si, y sólo si, es compacto (para la métrica usual de los espacios shift) y shift-invariante.*

Demostración. La shift-invariancia de un espacio shift fue demostrada en el primer capítulo, y la compacidad en la Proposición 3.13.

Recíprocamente, supongamos que X es un subconjunto compacto y shift-invariante de $\mathcal{A}^{\mathbb{Z}}$. Por ser compacto, X es cerrado, y entonces X^c es abierto. De allí que para cada $y \in X^c$ existe $K(y) \in \mathbb{N}$ tal que $B_{2-K}(y)$ está contenido en X^c . Tomemos $\mathcal{F} = \{y_{[-K(y), K(y)]} : y \in X^c\}$, y verifiquemos que $X = \mathbb{X}_{\mathcal{F}}$. Si $x \in X^c$, entonces $x_{[-K(x), K(x)]} \in \mathcal{F}$, de donde $x \notin \mathbb{X}_{\mathcal{F}}$. Luego, $\mathbb{X}_{\mathcal{F}} \subseteq X$. Ahora, sea $x \in X$ y $u \sqsubseteq x$. Si u tiene longitud par, $u \notin \mathcal{F}$ pues \mathcal{F} consta sólo de palabras de largo impar. Si $|u| = 2n + 1$, por la shift-invariancia de X debe haber un $x' \in X$ tal que $x'_{[-n, n]} = u$. Si u perteneciese a \mathcal{F} , existiría $y \in X^c$ tal que $K(y) = n$ y además $y_{[-n, n]} = u$. Pero entonces, ya que $x'_{[-n, n]} = u = y_{[-K(y), K(y)]}$, tendríamos que x' pertenecería a $B_{2-K(y)}(y) \subseteq X^c$, vale decir, x' pertenecería a X^c , lo que sería una contradicción pues $x' \in X$. Luego, $u \notin \mathcal{F}$, y $X \subseteq \mathbb{X}_{\mathcal{F}}$. \square

3.4. Cilindros. Conexidad. $\mathcal{A}^{\mathbb{Z}}$ puede ser visto como el producto cartesiano infinito $\prod_{i \in \mathbb{Z}} \mathcal{A}$. Como en cualquier producto cartesiano, se define un **cilindro** como un

subconjunto C del producto para el cual un número finito de coordenadas están fijas. En el caso particular de los espacios shift, nos van a interesar los cilindros en que un número finito de coordenadas consecutivas están fijas. Formalmente:

Definición 3.15. Sea \mathcal{A} un alfabeto, X un espacio shift sobre \mathcal{A} , $u \in \mathcal{A}^*$ y $k \in \mathbb{Z}$. El **cilindro en X asociado al bloque u en la posición k** es el conjunto $C_k^X(u)$ definido por

$$C_k^X(u) = \{x \in X : x_{[k, k+|u|-1]} = u\}.$$

En palabras, en el espacio shift X , el cilindro asociado a u en la posición k es el conjunto de puntos del espacio X tales que el bloque u ocurre en la posición k . Obsérvese que para cualquier $n \in \mathbb{Z}$, es $\sigma_X^n(C_k^X(u)) = C_{k-n}^X(u)$. Además, $C_k^X(u) = \emptyset$ si, y sólo si, $u \notin \mathcal{B}(X)$.

Si x es un punto del espacio shift X y ε es un real positivo, entonces

$$\begin{aligned} B_\varepsilon(x) &= \left\{ y \in X : x_{[-\lfloor \log \frac{1}{\varepsilon} \rfloor, \lfloor \log \frac{1}{\varepsilon} \rfloor]} = y_{[-\lfloor \log \frac{1}{\varepsilon} \rfloor, \lfloor \log \frac{1}{\varepsilon} \rfloor]} \right\} \\ &= C_{-\lfloor \log \frac{1}{\varepsilon} \rfloor}^X \left(x_{[-\lfloor \log \frac{1}{\varepsilon} \rfloor, \lfloor \log \frac{1}{\varepsilon} \rfloor]} \right). \end{aligned}$$

En particular, para $k \in \mathbb{N}$, se tiene que

$$B_{2^{-k}}(x) = \{y \in X : y_{[-k, k]} = x_{[-k, k]}\} = C_{-k}^X(x_{[-k, k]}).$$

Como vemos, los entornos en un espacio shift son cilindros. La recíproca no es cierta en general: no todo cilindro es un entorno. Sin embargo, todo cilindro es un conjunto abierto. En efecto, sea $x \in C_k^X(u)$. Tomando $N = \max\{|k|, |k+|u|-1|\}$, se tiene que $-N \leq k \leq k+|u|-1 \leq N$, de modo que

$$\begin{aligned} B_{2^{-N}}(x) &= C_{-N}^X(x_{[-N, N]}) = \{y \in X : y_{[-N, N]} = x_{[-N, N]}\} \\ &\subseteq \{y \in X : y_{[k, k+|u|-1]} = x_{[k, k+|u|-1]}\} = C_k^X(u). \end{aligned}$$

Pero además los cilindros son cerrados, pues

$$\begin{aligned} (C_k^X(u))^c &= \{x \in X : x_{[k, k+|u|-1]} \neq u\} \\ &= \bigcup_{v \in \mathcal{A}^{|u|} - \{u\}} \{x \in X : x_{[k, k+|u|-1]} = v\} \\ &= \bigcup_{v \in \mathcal{A}^{|u|} - \{u\}} C_k^X(v). \end{aligned}$$

Es decir, el complemento de un cilindro es una unión de conjuntos abiertos, y por lo tanto es un conjunto abierto; luego, el cilindro es cerrado.

Entonces, los cilindros en un espacio shift son conjuntos a la vez abiertos y cerrados. Esto revela otra característica topológica de los espacios shift: son conjuntos **totalmente disconexos**. Recordamos que una forma de caracterizar a los conjuntos disconexos de un espacio métrico es como aquellos conjuntos que poseen un subconjunto propio no vacío que es a la vez abierto y cerrado. Y un conjunto totalmente disconexo es aquel para el cual todo subconjunto de más de un punto es disconexo. El conjunto clásico de Cantor en la recta real tiene esta misma estructura: se trata de una nube de puntos no aislados en la cual ningún par de puntos distintos pertenece a una misma pieza conexa.

3.5. Continuidad.

Definición 3.16. Sean (X, d_X) e (Y, d_Y) dos espacios métricos. Una función $\phi : X \rightarrow Y$ se dice **continua en** $x \in X$ si para toda sucesión $\{x_n\}_{n \in \mathbb{N}}$ en X tal que $\lim_{n \rightarrow \infty} x_n = x$, se tiene que $\lim_{n \rightarrow \infty} \phi(x_n) = \phi(x)$. Si $A \subseteq X$, ϕ se dice **continua en** A si ϕ es continua en cada $x \in A$. Si ϕ es continua en X , biyectiva y además su función inversa es también continua, ϕ se denomina un **homeomorfismo entre** X e Y , y los espacios X e Y se dicen **homeomorfos**.

Una función entre dos espacios métricos es continua en todo su dominio si, y sólo si, la preimagen de cualquier subconjunto abierto del codominio es un subconjunto abierto del dominio. La composición de funciones continuas es también una función continua.

La definición de continuidad que hemos dado corresponde en realidad a la definición de la *continuidad secuencial o sucesional*, y equivale a la clásica definición $\varepsilon - \delta$ de continuidad: ϕ es continua en x si $\forall \varepsilon > 0, \exists \delta > 0 : d_X(y, x) < \delta \Rightarrow d_Y(\phi(y), \phi(x)) < \varepsilon$. Pero la continuidad desde el punto de vista de las sucesiones es más fácil de manejar cuando el espacio métrico es un espacio shift, ya que tenemos completamente caracterizadas a las sucesiones que convergen en un espacio shift.

Proposición 3.17. *Todo código de ventana deslizante entre dos espacios shift es una función continua en todo su dominio.*

Demostración. Sean X e Y dos espacios shift y $\phi : X \rightarrow Y$ un CVD inducido por Φ con memoria y anticipación $L \geq 0$. Sean $x \in X$ y $\{x^{(n)}\}_{n \in \mathbb{N}}$ una sucesión en X tal que $\lim_{n \rightarrow \infty} x^{(n)} = x$. Debemos ver que $\forall K \in \mathbb{N}, \exists N \in \mathbb{N} : \forall n \geq N, (\phi x^{(n)})_{[-K, K]} = (\phi x)_{[-K, K]}$. Sea $K \in \mathbb{N}$. Por la hipótesis de convergencia de $\{x^{(n)}\}_{n \in \mathbb{N}}$ a x , tenemos que existe $N \in \mathbb{N}$ tal que $\forall n \geq N, x^{(n)}_{[-K-L, K+L]} = x_{[-K-L, K+L]}$. Veamos que para todo $n \geq N$ es $(\phi x^{(n)})_{[-K, K]} = (\phi x)_{[-K, K]}$: si i es un entero tal que $-K \leq i \leq K$, será $-K - L \leq i - L \leq i + L \leq K + L$, por lo que $x^{(n)}_{[i-L, i+L]} = x_{[i-L, i+L]}$, y entonces

$$(\phi x^{(n)})_i = \Phi \left(x^{(n)}_{[i-L, i+L]} \right) = \Phi \left(x_{[i-L, i+L]} \right) = (\phi x)_i.$$

Luego, $\lim_{n \rightarrow \infty} \phi(x^{(n)}) = \phi(x)$, de donde se deduce que ϕ es continua en x , y ya que x es arbitrario, ϕ es continua en X . \square

La sola continuidad de una función entre dos espacios shift no garantiza que sea un CVD. Hace falta algo más: la conmutabilidad de la función con las shifts de los respectivos espacios. Este resultado se obtuvo por Hedlund y otros en el año 1969. Antes de demostrarlo, enunciemos algunos resultados conocidos de la teoría general de espacios métricos, y cuyas demostraciones quedan de ejercicio.

Proposición 3.18. *Sean A y B dos subconjuntos compactos disjuntos no vacíos de un espacio métrico. Existe un número real estrictamente positivo δ tal que para todos $a \in A$ y $b \in B$, $d(a, b) \geq \delta$.*

Corolario 3.19. *Sea $\{A_k\}_{k=1}^n$ una familia finita de subconjuntos compactos no vacíos de un espacio métrico, dos a dos disjuntos. Existe $\delta > 0$ tal que si $x \in A_i$ e $y \in A_j$ con $i \neq j$, entonces $d(x, y) \geq \delta$.*

Proposición 3.20. *Sea ϕ una función continua entre los espacios métricos X e Y , con X compacto, y sea F un subconjunto compacto de Y . Entonces, $\phi^{-1}(F)$ es un subconjunto compacto de X .*

Teorema 3.21. (*Curtis, Lyndon, Hedlund*) Sean X e Y dos espacios shift, y ϕ una función de X en Y . ϕ es un código de ventana deslizante si, y sólo si, ϕ es una función continua tal que $\phi \circ \sigma_X = \sigma_Y \circ \phi$.

Demostración. Sabemos, de los primeros capítulos, que si ϕ es un CVD, conmuta con las σ , y la Proposición 3.17 muestra que ϕ es continua.

Para la vuelta, recordemos otra caracterización de los CVD: ϕ es un CVD si, y sólo si, conmuta con las σ y existe un $N \in \mathbb{N}$ tal que $\phi(x)_0$ es función de $x_{[-N,N]}$. El hecho de conmutar con σ lo tenemos por hipótesis, así que sólo resta probar la existencia de un tal N , para lo cual usaremos la hipótesis de la continuidad de ϕ . Sea $\mathcal{U} = \mathcal{B}_1(\phi(X))$, el conjunto de símbolos que aparecen en los transformados de los puntos de X . Obviamente, \mathcal{U} es un subconjunto del alfabeto de Y . Para cada $b \in \mathcal{U}$, el cilindro $C_0^Y(b)$ es un subconjunto cerrado de Y , y como Y es compacto, $C_0^Y(b)$ es compacto. Además, si b y c son elementos de \mathcal{U} con $b \neq c$, es $C_0^Y(b) \cap C_0^Y(c) = \emptyset$. Para cada $b \in \mathcal{U}$, hagamos $E_b = \phi^{-1}(C_0^Y(b))$. Por su definición y por la Proposición 3.20, cada E_b es un subconjunto compacto no vacío de X , y si $b \neq c$, $E_b \cap E_c = \emptyset$, pues si $x \in E_b$, es $\phi(x) \in C_0^Y(b)$, luego $\phi(x) \notin C_0^Y(c)$, y entonces $x \notin E_c$. Además, cada $x \in X$ pertenece a un único E_b . Así que la familia $\{E_b\}_{b \in \mathcal{U}}$ es una familia de conjuntos compactos que constituye una partición para X . Luego, por el Corolario 3.19, existe $\delta > 0$ tal que $d(x, y) \geq \delta$ toda vez que $x \in E_b$ e $y \in E_c$ con $b \neq c$. Tomemos $N \in \mathbb{N}$ suficientemente grande como para que $2^{-N} < \delta$. Si x e y son puntos de X tales que $x_{[-N,N]} = y_{[-N,N]}$, entonces $d(x, y) < 2^{-N} < \delta$, por lo que x e y deben estar en un mismo E_b . Esto significa que existe $b \in \mathcal{U}$ tal que $\phi(x)$ y $\phi(y)$ están ambos en $C_0^Y(b)$, o sea, $\phi(x)_0 = b = \phi(y)_0$. Pero esto quiere decir que la coordenada 0 del transformado por ϕ de $x \in X$ depende sólo de $x_{[-N,N]}$, como queríamos probar. \square

3.6. Sistemas dinámicos. Los sistemas físicos pueden usualmente describirse a través de una cantidad finita de mediciones. Por ejemplo, un péndulo oscilante constituye un sistema que queda totalmente caracterizado por el ángulo respecto de la vertical y por la velocidad angular. Un gas se describe a través de la posición y momentum de cada molécula. A medida que transcurre el tiempo, esos valores cambian. Si denotamos por M al conjunto de todos los valores posibles para un dado sistema, el estado del mismo en un instante dado corresponde a uno de esos estados. Imaginemos que filmamos el sistema y que luego observamos los cuadros que corresponden al instante $t = 0$, luego a $t = 1$ seg, $t = 2$ seg, etc. Cada cuadro depende del anterior, de acuerdo a la manera en que el sistema evolucione durante cada intervalo de un segundo. Si suponemos que las leyes que gobiernan el sistema no cambian con el tiempo, la dependencia de cada cuadro respecto del anterior se traduce en una función $T : M \rightarrow M$ que usualmente es continua. Así, si x_0 es el estado del sistema en $t = 0$, $T(x_0)$ describe el estado del sistema en $t = 1$, $T \circ T(x_0)$ en $t = 2$, y así. Por lo tanto, estudiar el comportamiento del sistema en el tiempo se reduce a estudiar las iteraciones de T , es decir, la composición de T consigo misma. Para $n \geq 0$, designaremos por T^n a la composición de T consigo misma n veces. Es decir, T^0 es la identidad, y, para $n > 0$, $T^n = T \circ T^{n-1}$.

Definición 3.22. Un **sistema dinámico** es un par (M, T) , en donde M es un espacio métrico compacto y T es una transformación continua en M . Si T es un homeomorfismo, el sistema dinámico se dice **inversible**.

Para nosotros, los sistemas dinámicos más interesantes serán aquellos en los que el conjunto M sea un espacio shift. En particular, si X es un espacio shift, el par (X, σ_X)

es un sistema dinámico que denominamos **sistema dinámico shift**. Si ϕ es un CVD de X en X , el par (X, ϕ) es también un sistema dinámico.

Definición 3.23. Para un sistema dinámico (M, T) , la **órbita** de un punto $x \in M$ es la sucesión $\{T^n x\}_{n \in \mathbb{N}}$. Si el sistema es inversible, la órbita de x es la sucesión $\{T^n x\}_{n \in \mathbb{Z}}$.

Definición 3.24. Para un sistema dinámico (M, T) , un **punto periódico** es un punto $x \in M$ tal que $T^n x = x$ para algún $n > 0$, y tal n se llama un **período de x** . Si x es un punto periódico, el menor de los $n > 0$ tales que $T^n x = x$ se llama **período mínimo de x** . Una órbita $\{T^n x\}$ es **periódica** si x es un punto periódico.

Un conjunto de problemas que interesan en relación a estos tópicos involucra el estudio del comportamiento de la órbita de puntos y de conjuntos en un dado sistema dinámico. Por ejemplo, si se tiene un sistema dinámico (M, T) en particular, puede interesar saber cómo se distribuyen las órbitas de los puntos de un dado subconjunto abierto U de M , o si el conjunto de puntos periódicos del espacio es denso.

Otro conjunto de problemas tiene que ver con problemas de clasificación de sistemas dinámicos: ¿Cuándo dos sistemas dinámicos aparentemente distintos son “el mismo sistema”?

Definición 3.25. Sean (M, T) y (N, S) dos sistemas dinámicos. Un **homomorfismo** θ de (M, T) a (N, S) es una función continua $\theta : M \rightarrow N$ tal que $\theta \circ T = S \circ \theta$. Los homomorfismos de un sistema dinámico a sí mismo se denominan **endomorfismos** del sistema dinámico.

Una simple inducción muestra que si θ es un homomorfismo de (M, T) a (N, S) , entonces, para cualquier entero positivo n , se tiene que $\theta \circ T^n = S^n \circ \theta$.

Notemos que, debido al Teorema de Curtis, Lyndon y Hedlund, los homomorfismos entre los sistemas dinámicos shift son precisamente los códigos de ventana deslizante.

Definición 3.26. Una **inmersión** es un homomorfismo inyectivo, un **factor** es un homomorfismo sobreyectivo, y una **conjugación topológica** es un homomorfismo biyectivo. Dos sistemas dinámicos se dicen **topológicamente conjugados** si existe una conjugación topológica entre ellos.

Las nociones de inmersión, factor y conjugación conducen a problemas de clasificación de sistemas dinámicos. Dos sistemas dinámicos conjugados son esencialmente el mismo, pero la cuestión de si dos dados sistemas dinámicos son conjugados, o uno factor del otro, o uno inmerso en el otro, puede ser muy difícil de resolver.

Para detectar si dos sistemas dinámicos *no* son conjugados, una estrategia es la siguiente: asociemos a cada sistema dinámico con alguna entidad (que puede ser un número, un conjunto, o lo que sea) de manera que dos sistemas conjugados tengan asociados la misma entidad. Luego, dados dos sistemas dinámicos en particular, obtengamos las entidades asociadas respectivas, y si éstas no coinciden sabremos que los sistemas no son conjugados (aunque puede ocurrir que no podamos deducir nada si vemos que ambas entidades coinciden). Dichas entidades reciben el nombre de **invariantes por conjugación**.

Una de tales entidades es el conjunto de puntos periódicos de período n , donde n es un natural fijo. Ya hemos visto, en el capítulo primero, la demostración para el caso en que los sistemas involucrados son sistemas dinámicos shift. Veamos el resultado en general.

Proposición 3.27. *Sea n un entero positivo. La cantidad de puntos periódicos de período n es un invariante por conjugación.*

Demostración. Sean (M, T) y (N, S) sistemas dinámicos conjugados, y $\theta : M \rightarrow N$ una conjugación. Denotemos por $p_n(T)$ al cardinal del conjunto de puntos periódicos de período n para T , y $p_n(S)$ al de los puntos periódicos para S . Para $x \in M$, se tiene que

$$T^n x = x \Leftrightarrow \theta(T^n x) = \theta(x) \Leftrightarrow S^n(\theta(x)) = \theta(x)$$

de modo que θ establece una biyección entre los puntos fijos de T^n y los de S^n . Por lo tanto, $p_n(T) = p_n(S)$. \square

Un argumento similar al anterior muestra que el número de puntos periódicos de período mínimo n es también invariante por conjugación.

Veamos ahora otro invariante por conjugación.

Definición 3.28. Un sistema dinámico (M, T) es **topológicamente transitivo** si para cualquier par de abiertos no vacíos U y V , existe $n > 0$ tal que $(T^n U) \cap V \neq \emptyset$.

Es decir, un sistema topológicamente transitivo mueve a cada conjunto abierto lo suficiente como para que llegue a “tocar” a cualquier otro conjunto abierto, en alguna iteración.

Proposición 3.29. Sea (M, T) un sistema topológicamente transitivo, y U y V dos conjuntos abiertos. Existe una sucesión $\{n_k\}_{k \in \mathbb{N}}$ de naturales estrictamente creciente tal que para todo $k \in \mathbb{N}$, es $(T^{n_k} U) \cap V \neq \emptyset$.

Demostración. Definiremos $\{n_k\}_{k \in \mathbb{N}}$ inductivamente. Por la transitividad topológica de (M, T) , hay un $n_1 \in \mathbb{N}$ tal que $(T^{n_1} U) \cap V \neq \emptyset$. Supongamos definidos $n_1 < n_2 < \dots < n_k$ tales que $(T^{n_i} U) \cap V \neq \emptyset$. Haciendo $V' = (T^{n_k})^{-1} V$, tenemos que V' es abierto pues T^{n_k} , como composición de la función continua T consigo misma n_k veces, es una función continua. Además, es $V' \neq \emptyset$, pues como $(T^{n_k} U) \cap V \neq \emptyset$, existe $x \in U$ tal que $T^{n_k} x$ pertenece a V . Por lo tanto, existe $m > 0$ tal que $(T^m U) \cap V' \neq \emptyset$, es decir, $(T^m U) \cap (T^{n_k})^{-1} V \neq \emptyset$. Luego $T^{n_k}((T^m U) \cap (T^{n_k})^{-1} V) \neq \emptyset$. Pero $T^{n_k}((T^m U) \cap (T^{n_k})^{-1} V) \subseteq (T^{n_k}(T^m U)) \cap (T^{n_k}((T^{n_k})^{-1} V)) \subseteq (T^{m+n_k} U) \cap V$. Por lo tanto, tomando $n_{k+1} = m + n_k > n_k$, tenemos que $(T^{n_{k+1}} U) \cap V \neq \emptyset$. \square

Proposición 3.30. La transitividad topológica es invariante por conjugación.

Demostración. Sean (M, T) y (N, S) sistemas dinámicos conjugados, con (M, T) topológicamente transitivo, y $\theta : M \rightarrow N$ una conjugación. Sean U y V abiertos en N . Denotemos $U' = \theta^{-1}(U)$ y $V' = \theta^{-1}(V)$. Como θ es continua, U' y V' son abiertos en M , por lo que existe $n > 0$ tal que $(T^n U') \cap V' \neq \emptyset$, es decir, $(T^n(\theta^{-1}(U))) \cap \theta^{-1}(V) \neq \emptyset$. Por lo tanto, siendo θ una biyección, tenemos que

$$\begin{aligned} \emptyset &\neq \theta((T^n(\theta^{-1}(U))) \cap \theta^{-1}(V)) \\ &= \theta(T^n(\theta^{-1}(U))) \cap \theta(\theta^{-1}(V)) \\ &= S^n(\theta(\theta^{-1}(U))) \cap V \\ &= (S^n U) \cap V. \end{aligned}$$

Entonces, (N, S) es también topológicamente transitivo. \square

Irreducibilidad: un espacio shift X es irreducible si para cualquier par de palabras u y w en el lenguaje, hay una palabra v también en $\mathcal{B}(X)$ tal que $uvw \in \mathcal{B}(X)$. Veremos ahora que en el contexto de los sistemas dinámicos shift, irreducibilidad equivale a transitividad topológica.

Proposición 3.31. *Un sistema dinámico shift es topológicamente transitivo si, y sólo si, el espacio shift es irreducible.*

Demostración. Supongamos que el sistema dinámico shift (X, σ_X) es topológicamente transitivo. Sean u y w en $\mathcal{B}(X)$. Hagamos $U = C_0^X(u)$ y $V = C_0^X(w)$. Ya que U y V son abiertos y no vacíos, por Proposición 3.29, existe $n > |u|$ tal que $(\sigma_X^n U) \cap V \neq \emptyset$, es decir, hay un $x \in X$ tal que $x \in \sigma_X^n(C_0^X(u)) = C_{-n}^X(u)$ y $x \in C_0^X(w)$. Es decir que $x_{[-n, -n+|u|-1]} = u$ y $x_{[0, |w|-1]} = w$. Como $n > |u|$, es $-n + |u| - 1 < 0$, de modo que podemos tomar $v = x_{[-n+|u|, -1]}$ para ver que $uvw \sqsubseteq x$, es decir, $uvw \in \mathcal{B}(X)$.

Recíprocamente, supongamos que X es un espacio shift irreducible, y sean U y V dos abiertos no vacíos en el espacio métrico X . Sean $x \in U$, $y \in V$, y tomemos $k, l \in \mathbb{N}$ tal que $B_{2^{-k}}(x) \subseteq U$ y $B_{2^{-l}}(y) \subseteq V$. Hagamos $u = x_{[-k, k]}$ y $w = y_{[-l, l]}$. Sabemos que $B_{2^{-k}}(x) = C_{-k}^X(u)$ y que $B_{2^{-l}}(y) = C_{-l}^X(w)$. Como X es irreducible, existe $v \in \mathcal{B}(X)$ tal que $uvw \in \mathcal{B}(X)$. Tomemos $z \in X$ tal que $z_{[-k, -k+|u|+|v|+|w|-1]} = uvw$. Tenemos que $z \in C_{-k}^X(u) \subseteq U$, y además, puesto que $|u| = 2k + 1$ y $|w| = 2l + 1$,

$$z \in C_{-k+|u|+|v|}^X(w) = \sigma_X^{k-|u|-|v|-l}(C_{-l}^X(w)) = \sigma_X^{-k-1-|v|-l}(C_{-l}^X(w)) \subseteq \sigma_X^{-k-1-|v|-l}(V).$$

Es decir, $z \in U$ y $\sigma_X^{k+l+|v|+1}(z) \in V$. Por lo tanto, tomando $n = k + l + |v| + 1 > 0$, tenemos que $(\sigma_X^n U) \cap V \neq \emptyset$. \square

REFERENCIAS

- [1] D. Lind and B. Marcus. *An Introduction to Symbolic Dynamics and Coding*.
- [2] B. Kitchens. *Symbolic Dynamics*.
- [3] Petr Kurka. *Topological and Symbolic Dynamics*.
- [4] M. Brin and G. Stuk. *Introduction to Dynamical Systems*.

FACULTAD DE CIENCIAS EXACTAS, UNIVERSIDAD NACIONAL DE SALTA
E-mail address: jadur@unsa.edu.ar

SEDE REGIONAL METÁN-ROSARIO DE LA FRONTERA, UNIVERSIDAD NACIONAL DE SALTA
E-mail address: yazlle@unsa.edu.ar

ECUACIONES EN DERIVADAS PARCIALES UTILIZANDO ANÁLISIS FUNCIONAL

URIEL KAUFMANN

RESUMEN. La idea de este curso es dar una introducción a los espacios de Sobolev y probar (de manera muy simple) en estos espacios un teorema de existencia y unicidad de soluciones (ver Teorema 2.4) para problemas del tipo

$$\begin{cases} -\Delta u = f & \text{en } \Omega \\ u = 0 & \text{en } \partial\Omega, \end{cases}$$

donde $\Omega \subset \mathbb{R}^n$ ($n \geq 1$) es un abierto acotado y f es una función dada. Mencionamos que las técnicas que desarrollaremos se pueden adaptar a problemas mucho más generales. Una muy buena exposición de estos temas (con un nivel accesible) se puede encontrar en los libros [1] y [2]. Al final de estas notas hay un breve apéndice con la notación utilizada.

ÍNDICE

| | |
|---|-----|
| 1. Introducción y resultados auxiliares | 117 |
| 2. Resultado principal | 121 |
| 3. Apéndice | 123 |
| Referencias | 124 |

1. INTRODUCCIÓN Y RESULTADOS AUXILIARES

Motivación de la definición de derivada débil. Sean $I := (a, b)$, $u \in \mathcal{C}^1(\bar{I})$ y $\phi \in \mathcal{C}_c^\infty(I)$. Integrando por partes resulta

$$\int_a^b u\phi' = u\phi|_a^b - \int_a^b u'\phi = - \int_a^b u'\phi$$

pues ϕ es de soporte compacto. Análogamente, si $\Omega \subset \mathbb{R}^n$ es un abierto acotado con frontera suave, $u \in \mathcal{C}^1(\bar{\Omega})$ y $\phi \in \mathcal{C}_c^\infty(\Omega)$, del teorema de Gauss-Green se tiene ($\nu = (\nu^1, \dots, \nu^n)$ vector normal exterior unitario a $\partial\Omega$)

$$(1.1) \quad \int_{\Omega} u\phi_{x_i} = \int_{\partial\Omega} u\phi\nu^i - \int_{\Omega} u_{x_i}\phi = - \int_{\Omega} u_{x_i}\phi$$

En general, si u no es diferenciable, u_{x_i} no tiene por qué existir. Sin embargo, (1.1) tiene sentido si $u \in L^1_{loc}(\Omega)$ y si en el miembro derecho ponemos una función $v \in L^1_{loc}(\Omega)$ en lugar de u_{x_i} .

Definición 1.1. Sean $\Omega \subseteq \mathbb{R}^n$ un abierto y $u, v \in L^1_{loc}(\Omega)$. Decimos que v es la *derivada débil respecto a x_i* de u si

$$\int_{\Omega} u\phi_{x_i} = - \int_{\Omega} v\phi \quad \text{para toda } \phi \in \mathcal{C}_c^\infty(\Omega),$$

y en tal caso decimos que $u_{x_i} = v$ en sentido débil.

Observación 1.2.

(i) La derivada débil de u , si existe, es única. En efecto, si existen $v, w \in L^1_{loc}(\Omega)$ tales que

$$-\int_{\Omega} v\phi = \int_{\Omega} u\phi_{x_i} = -\int_{\Omega} w\phi \quad \text{para toda } \phi \in \mathcal{C}_c^{\infty}(\Omega),$$

entonces $\int_{\Omega} (v-w)\phi = 0$ para toda $\phi \in \mathcal{C}_c^{\infty}(\Omega)$ y por tanto (ver por ejemplo, [1], Corollary 4.24) $v = w$ a.e. $x \in \Omega$.

(ii) Es claro de (1.1) que si $u \in \mathcal{C}^1(\Omega)$ entonces u_{x_i} existe en sentido débil y coincide con la derivada usual. Recíprocamente, se puede mostrar que si u_{x_i} existe en sentido débil y además $u_{x_i} \in \mathcal{C}(\Omega)$ entonces la derivada usual respecto a x_i existe y coincide con u_{x_i} .

(iii) De forma análoga a lo anterior se pueden definir las derivadas débiles de cualquier orden. Más precisamente, si $\alpha := (\alpha_1, \dots, \alpha_n) \in \mathbb{N}_0^n$ y $|\alpha| := \sum \alpha_j$, decimos que v es la α -ésima derivada débil de u si

$$\int_{\Omega} u\phi_{x_1^{\alpha_1} \dots x_n^{\alpha_n}} = (-1)^{|\alpha|} \int_{\Omega} v\phi \quad \text{para toda } \phi \in \mathcal{C}_c^{\infty}(\Omega).$$

En el siguiente ejemplo vemos que la derivada débil efectivamente extiende la noción usual de derivada.

Ejemplo 1.3. Sean $\Omega = (0, 2)$ y

$$u(x) := \begin{cases} x & \text{si } x \in (0, 1], \\ 1 & \text{si } x \in [1, 2), \end{cases} \quad v(x) := \begin{cases} 1 & \text{si } x \in (0, 1], \\ 0 & \text{si } x \in (1, 2). \end{cases}$$

Veamos que $u' = v$ en sentido débil. Sea $\phi \in \mathcal{C}_c^{\infty}(\Omega)$ (en particular, $\phi(0) = \phi(2) = 0$). Integrando por partes obtenemos

$$\begin{aligned} \int_0^2 u\phi' &= \int_0^1 x\phi'(x) dx + \int_1^2 \phi' = x\phi(x) \Big|_0^1 - \int_0^1 \phi(x) dx + \phi(2) - \phi(1) \\ &= \phi(1) - \int_0^1 \phi(x) dx - \phi(1) = -\int_0^1 \phi(x) dx = -\int_0^2 v\phi \end{aligned}$$

que es lo que queríamos ver.

Ejemplo 1.4. Sean $\Omega = (0, 2)$ y u dada por

$$u(x) := \begin{cases} x & \text{si } x \in (0, 1], \\ 2 & \text{si } x \in (1, 2). \end{cases}$$

Veamos que u' no existe en sentido débil (obviamente tampoco en sentido usual). Supongamos por el absurdo que existe $v \in L^1_{loc}(\Omega)$ tal que

$$\int_{\Omega} u\phi' = -\int_{\Omega} v\phi \quad \text{para toda } \phi \in \mathcal{C}_c^{\infty}(\Omega).$$

Entonces

$$\begin{aligned} -\int_0^2 v\phi &= \int_0^2 u\phi' = \int_0^1 x\phi'(x) dx + \int_1^2 2\phi' \\ &= x\phi(x) \Big|_0^1 - \int_0^1 \phi(x) dx + 2(\phi(2) - \phi(1)) \\ &= \phi(1) - \int_0^1 \phi(x) dx - 2\phi(1) = -\phi(1) - \int_0^1 \phi. \end{aligned}$$

y por lo tanto

$$(1.2) \quad \phi(1) = \int_0^2 v\phi - \int_0^1 \phi.$$

Tomemos ahora una sucesión de funciones $\{\phi_j\}_{j \in \mathbb{N}} \subset \mathcal{C}_c^\infty(\Omega)$ satisfaciendo

- $0 \leq \phi_j(x) \leq 1$ para todo $x \in \Omega$,
- $\phi_j(1) = c \neq 0$ para todo $j \in \mathbb{N}$,
- $\lim_{j \rightarrow \infty} \phi_j(x) = 0$ si $x \neq 1$.

(Para ver cómo construir tal sucesión ver por ejemplo (3.1) en el apéndice al final.) Poniendo ahora ϕ_j en lugar de ϕ en (1.2) y utilizando el teorema de convergencia dominada (ver Teorema 3.1) se tiene

$$c = \lim_{j \rightarrow \infty} \phi_j(1) = \lim_{j \rightarrow \infty} \left[\int_0^2 v\phi_j - \int_0^1 \phi_j \right] = 0,$$

lo cual es una contradicción.

Definición 1.5. Sea $\Omega \subseteq \mathbb{R}^n$ un abierto. El *espacio de Sobolev* $W^{1,2}(\Omega)$ viene dado por

$$W^{1,2}(\Omega) := \left\{ u \in L^2(\Omega) : \begin{array}{l} u_{x_i} \text{ existe en sentido débil y} \\ u_{x_i} \in L^2(\Omega) \text{ para todo } 1 \leq i \leq n \end{array} \right\}.$$

Equipado con el producto escalar

$$\langle u, v \rangle_{W^{1,2}} := \langle u, v \rangle_{L^2} + \sum_{i=1}^n \langle u_{x_i}, v_{x_i} \rangle_{L^2} = \int_{\Omega} \left[uv + \sum_{i=1}^n u_{x_i} v_{x_i} \right] = \int_{\Omega} [uv + \langle \nabla u, \nabla v \rangle]$$

resulta ser un espacio de Hilbert reflexivo y separable (ver por ejemplo, [1], Proposition 9.1). La norma asociada es

$$\|u\|_{W^{1,2}} := \left[\|u\|_{L^2}^2 + \sum_{i=1}^n \|u_{x_i}\|_{L^2}^2 \right]^{\frac{1}{2}} = \left[\int_{\Omega} u^2 + \sum_{i=1}^n (u_{x_i})^2 \right]^{\frac{1}{2}} = \left[\int_{\Omega} u^2 + |\nabla u|^2 \right]^{\frac{1}{2}}.$$

Observación 1.6.

(i) Para $k \in \mathbb{N}$ y $1 \leq p \leq \infty$ se pueden definir análogamente los espacios de Sobolev $W^{k,p}$ como el conjunto de funciones $u \in L^p$ tales que todas sus derivadas parciales hasta el orden k existen en sentido débil y pertenecen a L^p . Estos espacios son de Banach y si $p = 2$ son Hilbert.

(ii) De la Observación 1.2 (i) se sigue que $\mathcal{C}_c^1(\Omega) \subset W^{1,2}(\Omega)$ y que si Ω es *acotado* se tiene $\mathcal{C}^1(\overline{\Omega}) \subset W^{1,2}(\Omega)$. Si Ω no es acotado entonces $\mathcal{C}^1(\overline{\Omega}) \not\subset W^{1,2}(\Omega)$ (¿por qué?) y (aunque Ω sea acotado) $\mathcal{C}^1(\Omega) \not\subset W^{1,2}(\Omega)$ (¿por qué?).

(iii) Toda función $u \in L^1_{loc}(\Omega)$ se puede identificar naturalmente con una distribución $T_u \in \mathcal{D}'(\Omega)$, y entonces existe $\frac{\partial}{\partial x_i} T_u \in \mathcal{D}'(\Omega)$. Usando el lenguaje de la *teoría de distribuciones* podemos decir que $W^{1,2}(\Omega)$ es el conjunto de funciones $u \in L^2(\Omega)$ tales que todas sus derivadas parciales (en el sentido de las distribuciones) pertenecen a $L^2(\Omega)$.

Una de las ventajas que tiene plantear problemas de ecuaciones diferenciales en los espacios de Sobolev $W^{k,p}(\Omega)$ es que en ellos hay “muchas más” funciones que en los espacios con derivadas usuales (y por lo tanto en algún sentido hay más posibilidades de encontrar soluciones), y además se puede permitir que los datos conocidos (por ejemplo, f en el problema del resumen) sean discontinuos y/o no acotados. Por otro lado, estos

problemas usualmente vienen asociados a condiciones en la frontera de Ω , y las funciones de los espacios de Sobolev están definidas salvo por conjuntos de medida cero, y $\partial\Omega$ tiene medida n -dimensional cero! Para arreglar este “detalle” se definen ciertos subespacios que mostramos a continuación.

Definición 1.7. Sea $\Omega \subseteq \mathbb{R}^n$ un abierto. El *espacio de Sobolev* $W_0^{1,2}(\Omega)$ viene dado por

$$(1.3) \quad W_0^{1,2}(\Omega) := \overline{\mathcal{C}_c^\infty(\Omega)}^{\|\cdot\|_{W^{1,2}}},$$

que resulta ser de Hilbert por ser un subespacio cerrado de un Hilbert.

Observación 1.8.

(i) Teniendo en cuenta la Observación 1.6 (i) podemos también definir los espacios $W_0^{k,p}$.

(ii) Si bien las funciones en $W_0^{1,2}$ están definidas salvo en conjuntos de medida nula, podemos decir que son “a grosso modo” aquellas funciones de $W^{1,2}$ que se anulan en la frontera. Más precisamente, se puede probar lo siguiente (ver [1], Theorem 9.17): Sea $u \in W^{1,2}(\Omega) \cap \mathcal{C}(\bar{\Omega})$. Entonces $u = 0$ en $\partial\Omega$ implica $u \in W_0^{1,2}(\Omega)$. Recíprocamente, si $\partial\Omega$ es suave, entonces $u \in W_0^{1,2}(\Omega)$ implica $u = 0$ en $\partial\Omega$.

Como las funciones de $W_0^{1,2}(\Omega)$ “se anulan” en $\partial\Omega$, uno no necesita la norma L^2 de estas funciones para controlar su convergencia en $W_0^{1,2}$, basta con la norma L^2 de sus derivadas. Esto saldrá como consecuencia del siguiente teorema. Antes de enunciarlo recordamos la noción de derivada direccional. Si una función v es diferenciable y ξ es un vector unitario en \mathbb{R}^n , entonces el límite

$$\lim_{h \rightarrow 0} \frac{v(x + h\xi) - v(x)}{h}$$

existe y se dice derivada en la dirección de ξ . Por ejemplo, $\partial_{x_1}v$ es la derivada en la dirección de $e_1 := (1, 0, \dots, 0)$. Para ver que el límite anterior efectivamente existe escribimos $F(t) := v(x + th\xi)$ y entonces del teorema del valor medio y la regla de la cadena resulta

$$v(x + h\xi) - v(x) = F(1) - F(0) = F'(\theta) = \langle \nabla v(x + \theta h\xi), h\xi \rangle.$$

Dividiendo ahora por h y haciendo $h \rightarrow 0$ tenemos

$$\lim_{h \rightarrow 0} \frac{v(x + h\xi) - v(x)}{h} = \langle \nabla v(x), \xi \rangle := \partial_\xi v(x).$$

Es natural entonces la siguiente

Definición 1.9. Dado ξ vector unitario en \mathbb{R}^n y $u \in W^{1,2}(\Omega)$ definimos $\partial_\xi u(x) := \langle \nabla u(x), \xi \rangle$, donde ∇u se entiende en sentido débil. Notar que $\partial_\xi u \in L^2(\Omega)$.

Teorema 1.10 (Desigualdad de Poincaré). *Sea ξ un vector unitario en \mathbb{R}^n , y sea $\Omega \subset \mathbb{R}^n$ un abierto acotado. Entonces existe $c > 0$ que sólo depende de Ω tal que*

$$\|u\|_{L^2} \leq c \|\partial_\xi u\|_{L^2} \quad \text{para toda } u \in W_0^{1,2}(\Omega).$$

Demostración. Como la medida de Lebesgue es invariante por rotaciones podemos elegir el sistema de coordenadas de manera que $\xi = e_1$. Tomamos primero $u \in \mathcal{C}_c^\infty(\Omega)$ y la extendemos por cero fuera de Ω . Como Ω es acotado existe $R > 0$ tal que $\Omega \subset B_R(0)$. Ahora, para $x = (x_1, \dots, x_n) \in \Omega$ tenemos

$$u(x) = u(x) - u(-R, x_2, \dots, x_n) = \int_{-R}^{x_1} \partial_{x_1} u(s, x_2, \dots, x_n) ds$$

y entonces elevando al cuadrado y usando la desigualdad de Hölder resulta

$$\begin{aligned} u^2(x) &= \left[\int_{-R}^{x_1} \partial_{x_1} u(s, x_2, \dots, x_n) ds \right]^2 \leq \left[\int_{-R}^{x_1} |\partial_{x_1} u(s, x_2, \dots, x_n)| ds \right]^2 \\ &\leq |x_1 + R| \int_{-R}^{x_1} [\partial_{x_1} u(s, x_2, \dots, x_n)]^2 ds \leq |x_1 + R| \int_{-R}^R [\partial_{x_1} u(s, x_2, \dots, x_n)]^2 ds. \end{aligned}$$

Integrando esta desigualdad en x_1 obtenemos

$$\begin{aligned} \int_{-R}^R u^2(x_1, \dots, x_n) dx_1 &\leq \frac{(x_1 + R)^2}{2} \Big|_{-R}^R \int_{-R}^R [\partial_{x_1} u(x_1, \dots, x_n)]^2 dx_1 \\ &= 2R^2 \int_{-R}^R [\partial_{x_1} u(x_1, \dots, x_n)]^2 dx_1 \end{aligned}$$

e integrando en las otras $n - 1$ direcciones queda

$$(1.4) \quad \|u\|_{L^2(\Omega)} \leq \sqrt{2}R \|\partial_{x_1} u\|_{L^2(\Omega)}$$

que es exactamente lo que queríamos (recordar que elegimos $\xi = e_1$). Si ahora $u \in W_0^{1,2}(\Omega)$ es arbitraria, tomamos $\phi_j \in C_c^\infty(\Omega)$ con $\phi_j \rightarrow u$ en $W^{1,2}$ (tales ϕ_j existen *por definición*, ver (1.3)). O sea, $\phi_j \rightarrow u$ en L^2 y $\partial_{x_i} \phi_j \rightarrow \partial_{x_i} u$ en L^2 para todo $1 \leq i \leq n$. En particular $\|\phi_j\|_{L^2(\Omega)} \rightarrow \|u\|_{L^2(\Omega)}$ y $\|\partial_{x_1} \phi_j\|_{L^2(\Omega)} \rightarrow \|\partial_{x_1} u\|_{L^2(\Omega)}$ (en efecto, de la desigualdad triangular se tiene $0 \leq \left| \|\phi_j\|_{L^2(\Omega)} - \|u\|_{L^2(\Omega)} \right| \leq \|\phi_j - u\|_{L^2(\Omega)}$ y análogamente para la derivada). Entonces utilizando (1.4) con ϕ_j en lugar de u y pasando al límite se termina la prueba. \square

Corolario 1.11 (Normas equivalentes). *Sea $\Omega \subset \mathbb{R}^n$ un abierto acotado. Entonces en $W_0^{1,2}(\Omega)$ son equivalentes las normas*

$$\|u\|_{W^{1,2}}, \quad \|\nabla u\|_{L^2}.$$

Demostración. Por un lado se tiene

$$\|u\|_{W^{1,2}}^2 = \int_{\Omega} u^2 + |\nabla u|^2 \geq \int_{\Omega} |\nabla u|^2 = \|\nabla u\|_{L^2}^2,$$

y por otra parte, para $\xi \in \mathbb{R}^n$ con $|\xi| = 1$, del teorema anterior resulta

$$(1.5) \quad \int_{\Omega} u^2 \leq c^2 \int_{\Omega} (\partial_{\xi} u)^2 = c^2 \int_{\Omega} \langle \nabla u, \xi \rangle^2 \leq c^2 \int_{\Omega} |\nabla u|^2$$

y entonces

$$\|u\|_{W^{1,2}}^2 = \int_{\Omega} u^2 + |\nabla u|^2 \leq (1 + c^2) \int_{\Omega} |\nabla u|^2 = (1 + c^2) \|\nabla u\|_{L^2}^2.$$

\square

2. RESULTADO PRINCIPAL

Daremos ahora la definición de *solución débil* del problema

$$(2.1) \quad \begin{cases} -\Delta u = f & \text{en } \Omega \\ u = 0 & \text{en } \partial\Omega. \end{cases}$$

Recordemos que u se dice *solución clásica* de (2.1) si $u \in C^2(\Omega) \cap C(\bar{\Omega})$ y (2.1) se cumple en sentido usual.

Motivación de la definición de solución débil. Supongamos que u es solución clásica de (2.1), y sea $\phi \in \mathcal{C}_c^\infty(\Omega)$. Como u_{x_i} tiene derivada usual continua que coincide con la derivada débil, multiplicando por ϕ la primera igualdad de (2.1), integrando sobre Ω y utilizando la definición de derivada débil (para u_{x_i}) tenemos

$$(2.2) \quad \int_{\Omega} f\phi = - \int_{\Omega} (\Delta u) \phi = - \int_{\Omega} \sum_i u_{x_i x_i} \phi = \int_{\Omega} \sum_i u_{x_i} \phi_{x_i} = \int_{\Omega} \langle \nabla u, \nabla \phi \rangle,$$

o sea

$$(2.3) \quad \int_{\Omega} \langle \nabla u, \nabla \phi \rangle = \int_{\Omega} f\phi \quad \text{para toda } \phi \in \mathcal{C}_c^\infty(\Omega).$$

Si u no es diferenciable ∇u no tiene por qué existir. Sin embargo, (2.3) tiene sentido si $u \in W^{1,2}(\Omega)$ y $f \in L^2(\Omega)$ (aún poniendo $\phi \in W_0^{1,2}$ en lugar de $\phi \in \mathcal{C}_c^\infty$). Como queremos reemplazar la segunda condición de (2.1) (ya que a priori si $u \in W^{1,2}(\Omega)$, u no está definida en $\partial\Omega$) pediremos que $u \in W_0^{1,2}(\Omega)$.

Definición 2.1. Sean $\Omega \subset \mathbb{R}^n$ un abierto y $f \in L^2(\Omega)$. Decimos que u es *solución débil* de (2.1) si

$$(2.4) \quad u \in W_0^{1,2}(\Omega), \quad \int_{\Omega} \langle \nabla u, \nabla v \rangle = \int_{\Omega} f v \quad \text{para toda } v \in W_0^{1,2}(\Omega).$$

Observación 2.2.

(i) Notar que f no necesariamente es ni continua ni acotada en Ω . Notar también que decimos que u es solución de una ecuación de orden dos a pesar que u sólo tiene derivadas (débiles) de orden uno!

(ii) Es fácil ver que es equivalente en la definición (2.4) poner v en $W_0^{1,2}$ o en \mathcal{C}_c^∞ .

(iii) Supongamos Ω es *acotado*. Si $u \in \mathcal{C}^2(\Omega) \cap \mathcal{C}^1(\overline{\Omega})$ es solución clásica, entonces de la Observación 1.6 (ii) tenemos que $u \in W^{1,2}(\Omega)$; y como $u = 0$, de la Observación 1.8 (ii) tenemos además que $u \in W_0^{1,2}(\Omega)$. Es claro además de las cuentas de arriba que se verifica (2.4), o sea u es solución débil. Es fácil encontrar ejemplos que muestran que esto no es cierto en general si Ω no es acotado. En efecto, tomemos por ejemplo $\Omega := \mathbb{R}^n - \overline{B_1(0)}$ y $u(x) = |x|^2 - 1$. Es claro que u es solución clásica del problema $-\Delta u = -2n$ en Ω y $u = 0$ en $\partial\Omega$. Sin embargo u no puede ser solución débil ya que $u \notin W_0^{1,2}(\Omega)$ (de hecho, $u \notin L^2(\Omega)$).

(iv) Supongamos que Ω tiene frontera suave. Sea u solución débil y además $u \in \mathcal{C}^2(\Omega) \cap \mathcal{C}(\overline{\Omega})$. De la Observación 1.8 (ii) es $u = 0$ en $\partial\Omega$, y además razonando como en (2.2) obtenemos

$$- \int_{\Omega} \phi \Delta u = \int_{\Omega} f \phi \quad \text{para toda } \phi \in \mathcal{C}_c^\infty(\Omega)$$

y entonces (ver por ejemplo, [1], Corollary 4.24) $-\Delta u = f$ a.e. $x \in \Omega$. De hecho, como $u \in \mathcal{C}^2(\Omega)$, la última igualdad vale en todo Ω (redefiniendo eventualmente f en un conjunto de medida cero), o sea u es solución clásica.

(v) Hay otra forma natural de definir solución de (2.1) involucrando las derivadas débiles: dada $f \in L^2(\Omega)$, decimos que $u \in W^{2,2}(\Omega) \cap W_0^{1,2}(\Omega)$ es *solución fuerte* de (2.1) si las derivadas débiles satisfacen $-\Delta u = f$ a.e. $x \in \Omega$. Notar que en este caso u tiene derivadas (débiles) hasta el orden dos. Esta definición se extiende de manera natural a una función $f \in L^p(\Omega)$ con $p \in (1, \infty)$.

Antes de enunciar y probar el teorema principal, recordamos el teorema de representación de Riesz (ver por ejemplo [2], Theorem 1.4).

Teorema 2.3. Sea H un espacio de Hilbert real y

$$H^* := \{\varphi^* : H \rightarrow \mathbb{R} : \varphi^* \text{ es lineal y continuo}\}$$

su dual. Entonces para cada $\varphi^* \in H^*$ existe un único $h \in H$ tal que

$$\varphi^*(v) = \langle h, v \rangle_H \quad \text{para todo } v \in H.$$

Más aún, $\|h\| = \sup_{0 \neq v \in H} \frac{\varphi^*(v)}{\|v\|}$.

El teorema de Riesz junto con la desigualdad de Poincaré (en realidad, su corolario) nos permiten probar un teorema de existencia y unicidad de manera muy simple.

Teorema 2.4 (Existencia, unicidad y dependencia continua). Sean $\Omega \subset \mathbb{R}^n$ un abierto acotado y $f \in L^2(\Omega)$. Entonces existe una única solución débil $u \in W_0^{1,2}(\Omega)$ de (2.1) y

$$(2.5) \quad \|u\|_{W_0^{1,2}} \leq c \|f\|_{L^2} \quad \text{con } c > 0 \text{ sólo dependiendo de } \Omega.$$

Demostración. Tomamos $H = W_0^{1,2}(\Omega)$, y por el Corolario 1.11 es posible elegir $\|u\|_{W_0^{1,2}} = \|\nabla u\|_{L^2}$, o sea $\langle u, v \rangle_{W_0^{1,2}} = \int_{\Omega} \langle \nabla u, \nabla v \rangle$. Definimos $\varphi^* : H \rightarrow \mathbb{R}$ por $\varphi^*(v) = \int_{\Omega} f v$. Es claro que φ^* es lineal y está bien definido (o sea, la integral es finita). Además, usando (1.5) es

$$(2.6) \quad |\varphi^*(v)| \leq \int_{\Omega} |f v| \leq \|f\|_{L^2} \|v\|_{L^2} \leq c \|f\|_{L^2} \|\nabla v\|_{L^2} = c \|f\|_{L^2} \|v\|_{W_0^{1,2}}$$

y por tanto φ^* es continuo y luego $\varphi^* \in H^*$. Pero entonces por el teorema de Riesz existe un único $u \in W_0^{1,2}(\Omega)$ tal que

$$\int_{\Omega} f v = \varphi^*(v) = \langle u, v \rangle_{W_0^{1,2}} = \int_{\Omega} \langle \nabla u, \nabla v \rangle \quad \text{para todo } v \in W_0^{1,2},$$

o sea, existe una única solución débil de (2.1). Para ver (2.5) ponemos $v = u$ en (2.4) y procediendo como en (2.6) resulta

$$\|u\|_{W_0^{1,2}}^2 = \int_{\Omega} \langle \nabla u, \nabla u \rangle = \int_{\Omega} f u \leq \|f\|_{L^2} \|u\|_{L^2} \leq c \|f\|_{L^2} \|u\|_{W_0^{1,2}}$$

y esto termina la prueba. □

3. APÉNDICE

Notación.

- Ω denota un abierto en \mathbb{R}^n , $\partial\Omega =$ frontera de Ω , $\overline{\Omega} = \Omega \cup \partial\Omega =$ clausura de Ω .
Escribimos $\Omega_1 \subset\subset \Omega_2$ si $\overline{\Omega_1} \subset \Omega_2$ y $\overline{\Omega_1}$ es compacto.
- $B_R(x_0) = \{x \in \mathbb{R}^n : |x - x_0| < R\}$.
- Si $u : \Omega \rightarrow \mathbb{R}$, escribimos $u(x) = u(x_1, \dots, x_n)$ para $x \in \Omega$.
- $\text{sop } u = \overline{\{x \in \Omega : u(x) \neq 0\}}$ = soporte de u .
- $u_{x_i}(x) = \partial_{x_i} u(x) = \lim_{h \rightarrow 0} \frac{u(x+he_i) - u(x)}{h}$ si es que el límite existe, y análogamente las derivadas de mayor orden.
- $\Delta u = \sum_{i=1}^n u_{x_i x_i} =$ laplaciano de u .
- $C(\Omega) = \{u : \Omega \rightarrow \mathbb{R} : u \text{ es continua en } \Omega\}$.
- $C^k(\Omega) = \{u : \Omega \rightarrow \mathbb{R} : u \text{ es continuamente diferenciable } k \text{ veces en } \Omega\}$.
- $C(\overline{\Omega}) = \{u \in C(\Omega) : u \text{ es uniformemente continua en } \Omega\}$.

- $\mathcal{C}^k(\overline{\Omega}) = \left\{ u \in \mathcal{C}^k(\Omega) : \begin{array}{l} \text{todas las derivadas parciales de todos los órdenes} \\ \text{(hasta } k \text{) son } \textit{uniformemente} \text{ continuas en } \Omega \end{array} \right\}$.

Es fácil ver que una función uniformemente continua en Ω se extiende (de manera única) continuamente a $\partial\Omega$. Por lo tanto podemos pensar que las funciones en $\mathcal{C}^k(\overline{\Omega})$ (y sus derivadas) están definidas en $\overline{\Omega}$.

- $\mathcal{C}^\infty(\Omega) = \bigcap_{k \in \mathbb{N}} \mathcal{C}^k(\Omega)$.
- $\mathcal{C}_c(\Omega)$, $\mathcal{C}_c^\infty(\Omega)$, etc. denotan las funciones en $\mathcal{C}(\Omega)$, $\mathcal{C}^\infty(\Omega)$, etc. que tienen *soporte compacto*.
- $L^p(\Omega) = \{u : \Omega \rightarrow \mathbb{R} : u \text{ es medible Lebesgue y } \|u\|_{L^p} < \infty\}$, con

$$\|u\|_{L^p} := \left(\int_{\Omega} u^p \right)^{\frac{1}{p}} \text{ si } p \in [1, \infty), \quad \|u\|_{L^\infty} := \sup \operatorname{ess}_{\Omega} |u|.$$

- $L_{loc}^p(\Omega) = \left\{ u : \Omega \rightarrow \mathbb{R} : u \in L^p(\tilde{\Omega}) \text{ para todo } \tilde{\Omega} \subset\subset \Omega \right\}$.

Teorema 3.1 (Convergencia Dominada de Lebesgue). *Sea $\{f_j\}$ una sucesión de funciones en $L^1(\Omega)$ satisfaciendo*

- (a) $f_j(x) \rightarrow f(x)$ a.e. $x \in \Omega$,
- (b) existe $g \in L^1(\Omega)$ tal que para todo j , $|f_j(x)| \leq g(x)$ a.e. $x \in \Omega$.

Entonces $f \in L^1(\Omega)$ y $\|f_j - f\|_{L^1(\Omega)} \rightarrow 0$.

Funciones en $\mathcal{C}_c^\infty(\Omega)$. Dado cualquier $\Omega \subseteq \mathbb{R}^n$ y $B_R(x_0) \subset \Omega$, es fácil construir funciones en $\mathcal{C}_c^\infty(\Omega)$ con soporte en $\overline{B_R(x_0)}$ comenzando a partir de una *única* función. En efecto, tomemos primero $f : \mathbb{R} \rightarrow \mathbb{R}$ dada por

$$f(x) := \begin{cases} e^{-1/x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

Se tiene $0 \leq f \leq 1$, y es fácil ver que $f \in \mathcal{C}^\infty(\mathbb{R})$ (de hecho, $f^{(j)}(x) = p(1/x) e^{-1/x}$ si $x > 0$ y $f^{(j)}(x) = 0$ en caso contrario, donde p es algún polinomio). Definimos ahora $\tilde{\rho} : \mathbb{R}^n \rightarrow \mathbb{R}$ por $\tilde{\rho}(x) := f(1 - |x|^2)$ y entonces $\tilde{\rho} \in \mathcal{C}^\infty(\mathbb{R}^n)$, $0 \leq \tilde{\rho} \leq 1$ y $\operatorname{sop} \tilde{\rho} = \overline{B_1(0)}$. Notar que $\tilde{\rho}$ es una función *radial* (o sea, sólo depende de $|x|$) y que

$$\tilde{\rho}(x) = \begin{cases} e^{\frac{1}{|x|^2-1}} & \text{si } |x| < 1, \\ 0 & \text{si } |x| \geq 1. \end{cases}$$

Finalmente, escribimos $\rho(x) := \tilde{\rho}\left(\frac{x-x_0}{R}\right)$ y resulta $\rho \in \mathcal{C}_c^\infty(\Omega)$ con $\operatorname{sop} \rho = \overline{B_R(x_0)}$.

Notar que para construir la sucesión de funciones ϕ_j del ejemplo 2 podemos tomar una sucesión ε_j con $\varepsilon_j > 0$ y $\varepsilon_j \rightarrow 0$ y definir

$$(3.1) \quad \phi_j(x) = \tilde{\rho}\left(\frac{x-1}{\varepsilon_j}\right), \quad x \in (0, 2).$$

REFERENCIAS

- [1] H. Brezis, *Functional analysis, Sobolev spaces and partial differential equations*. Universitext. Springer, New York (2011).
- [2] M. Chipot, *Elliptic equations: an introductory course*. Birkhäuser Advanced Texts, Birkhäuser Verlag, Basel (2009).

MÉTRICAS EN ÁLGEBRAS DE LIE DOS PASOS NILPOTENTES

GABRIELA P. OVANDO

RESUMEN. En estas notas veremos cómo ciertos conceptos geométricos pueden ser leídos a nivel algebraico. Trabajaremos con álgebras de Lie 2-pasos nilpotentes munidas de métricas definidas o indefinidas. Veremos que la signatura de la métrica influye en diversos aspectos. Trataremos dos casos extremos por un lado el de productos internos y por otro el de las métricas ad-invariantes.

ÍNDICE

| | |
|--|-----|
| Introducción | 125 |
| 1. Aplicaciones bilineales en espacios vectoriales | 126 |
| 2. Álgebras de Lie dos pasos nilpotentes | 128 |
| 3. Productos internos en \mathfrak{n} | 130 |
| 3.1. Elementos geométricos en \mathfrak{n} | 132 |
| 4. Métricas indefinidas en \mathfrak{n} | 133 |
| 4.1. Centro no degenerado | 133 |
| 4.2. Centro degenerado | 134 |
| Referencias | 136 |

INTRODUCCIÓN

Una de las motivaciones para estudiar métricas en álgebras de Lie 2-pasos nilpotentes proviene de la geometría. En efecto si suponemos que una cierta variedad provista de una métrica (definida o no) admite la acción de un grupo de isometrías 2-pasos nilpotente entonces esa variedad es llamada una nilvariedad y si además suponemos que el subgrupo de isotropía es trivial, entonces la variedad se identifica naturalmente con un grupo de Lie nilpotente.

Las álgebras de Lie 2-pasos nilpotentes constituyen la contraparte algebraica de los grupos de Lie 2-pasos nilpotentes y bajo ciertas hipótesis la geometría de una nilvariedad puede ser estudiada en muchos aspectos a través de su álgebra de Lie. Nuestro objetivo aquí, obviando el aspecto geométrico global, es ofrecer una muestra de cómo se traducen a nivel algebraico diversos conceptos y cómo se trabajan los mismos. Así empezaremos por repasar propiedades de aplicaciones bilineales en espacios vectoriales para luego trabajar el caso de formas bilineales en álgebras de Lie 2-pasos nilpotentes.

Nos interesa más en detalle trabajar dos situaciones bastante diferentes: por un lado el caso de álgebras de Lie provistas de un producto interno, lo cual correspondería a una nilvariedad provista de una métrica riemanniana invariante a izquierda y luego por otro lado las álgebras de Lie 2-pasos con métricas ad-invariantes, lo cual induce una métrica bi-invariante a nivel del respectivo grupo de Lie. En este último caso veremos

2010 *Mathematics Subject Classification.* 53B30, 22E25.

Parcialmente financiado por SCyT — Universidad Nacional de Rosario.

cómo la existencia de una tal métrica impone fuertes restricciones en la estructura del álgebra de Lie respectiva.

Para el tipo de problemas en geometría riemanniana remito por ejemplo al survey [1] y trabajos de diferentes autores que pueden revistarse en la bibliografía sobre nilvariedades munidas de métricas riemanniana invariantes (por ejemplo [5, 6] y sus referencias!). El caso pseudo-riemanniano recién comienza a explorarse en los últimos tiempos, ver las referencias en [8, 9, 10].

1. APLICACIONES BILINEALES EN ESPACIOS VECTORIALES

Sea V un espacio vectorial (real). Recordemos que una aplicación lineal de V en otro espacio vectorial (real) W es una aplicación $T : V \rightarrow W$ tal que

$$T(u + rv) = Tu + rTv \quad \text{para } u, v \in V, r \in \mathbb{R}.$$

La aplicación T también puede ser llamada un operador o una transformación lineal. Cuando W es el cuerpo de escalares de V , la transformación lineal T suele denominarse una funcional lineal.

Definición 1.1. Una *aplicación bilineal* B sobre V con valores en W es una aplicación $B : V \times V \rightarrow W$ la cual asigna a cada par (u, v) un elemento $B(u, v) \in W$ tal que para cada $u_0 \in V, v_0 \in V$ las aplicaciones

- $B(u_0, \cdot) : V \rightarrow W$ y
- $B(\cdot, v_0) : V \rightarrow W$

son lineales. Más aún, B se dice

- *simétrica* si $B(u, v) = B(v, u)$ para todos $u, v \in V$,
- *antisimétrica* si $B(u, v) = -B(v, u)$ para todos $u, v \in V$.

Cuando W coincide con el cuerpo de escalares de V , en este caso \mathbb{R} , suele decirse que B es una *forma bilineal* sobre V .

Ejemplo 1.2. El corchete de Lie en un álgebra de Lie es una aplicación bilineal antisimétrica con valores en la misma álgebra de Lie (veremos esto más adelante).

Sea V un espacio vectorial real de dimensión n y sea e_1, e_2, \dots, e_n una base de V . Supongamos que B es una forma bilineal sobre V con valores en W . Así para $u = \sum_{i=1}^n x_i e_i, v = \sum_{j=1}^n y_j e_j$ tenemos

$$B\left(\sum_{i=1}^n x_i e_i, \sum_{j=1}^n y_j e_j\right) = \sum_{i=1}^n \sum_{j=1}^n x_i y_j B(e_i, e_j),$$

lo cual indica que esencialmente B queda determinada si conocemos $B(e_i, e_j) \in W$.

En particular si B es una forma bilineal y armamos la matriz $A = (a_{ij})$ donde $a_{ij} = B(e_i, e_j)$ y tomamos las matrices de coordenadas de u y v en la base ordenada e_1, e_2, \dots, e_n entonces

$$(1.1) \quad B(u, v) = [u]^t [A] [v]$$

donde $[u]$ y $[v]$ denotan las matrices de coordenadas de u y v respectivamente. Recíprocamente si tomamos una matriz $n \times n$ A es fácil verificar que (1.1) define una forma bilineal B sobre V tal que $a_{ij} = B(e_i, e_j)$.

La matriz A se dice la matriz de B en la base ordenada e_1, e_2, \dots, e_n . Definimos además el *rango* de B como el rango de la matriz asociada a B y lo denotaremos como $\text{rango}(B)$.

Ejercicio 1.3. Pruebe que el rango de B no depende de la base, con lo cual la definición anterior es buena.

Proposición 1.4. Sea V un espacio vectorial de dimensión n y sea B una forma bilineal sobre V . Son equivalentes:

- $\text{rango}(B) = n$;
- para todo $u \neq 0$ existe $v \in V$ tal que $B(u, v) \neq 0$;
- para todo $v \neq 0$ existe $u \in V$ tal que $B(u, v) \neq 0$.

Demostración. Ejercicio. □

Definición 1.5. Una forma bilineal B sobre un espacio vectorial de dimensión n se dice *no degenerada* (o *no singular*) si satisface alguna de las condiciones equivalentes de la proposición anterior.

Estudiemos más en detalle las formas bilineales simétricas. Si B es una forma bilineal simétrica, la forma cuadrática asociada a B es la función q de V en \mathbb{R} definida por

$$q(v) = B(v, v)$$

y tenemos la denominada identidad de polarización:

$$B(u, v) = \frac{1}{4}q(u+v) - \frac{1}{4}q(u-v).$$

Ejemplo 1.6. Para $u, v \in \mathbb{R}^n$ con $u = \sum_{i=1}^n x_i e_i$, $v = \sum_{j=1}^n y_j e_j$ en la base canónica, el producto interno canónico denotado $\langle \cdot, \cdot \rangle$ está dado por $\langle u, v \rangle = \sum_{i=1}^n x_i y_i$. Este es un ejemplo de forma bilineal simétrica no degenerada. Más aún, su forma cuadrática asociada es

$$q(x_1, \dots, x_n) = x_1^2 + x_2^2 + \dots + x_n^2.$$

Dos vectores u, v en un espacio vectorial V munido de una forma bilineal simétrica B se dicen *ortogonales* si $B(u, v) = 0$ y la forma bilineal B se dice *definida positiva* si $B(v, v) > 0$ para todo $v \in V - \{0\}$. Notemos que esto implica que B es no degenerada.

Teorema 1.7. Sea V un espacio vectorial real de dimensión n y sea B una forma bilineal simétrica sobre V . Entonces existe una base de V en la cual B está representada por una matriz diagonal. Más aún, si el rango de B es r se puede elegir una base v_1, v_2, \dots, v_n tal que

$$B(v_i, v_i) = \pm 1 \quad i = 1, 2, \dots, r.$$

y el número de vectores de esa base para los cuales $B(v_i, v_i) = 1$ es independiente de la elección de la base.

Demostración. Ver pruebas por ejemplo en el libro [4]. □

La *signatura* de la forma bilineal simétrica no degenerada B es el par (k, s) donde k denota la cantidad de -1 que se obtiene en la matriz de B del teorema anterior, mientras que s es la cantidad de 1 .

De aquí en más una *métrica* en un espacio vectorial V será una forma bilineal simétrica y no degenerada sobre V .

Un *producto interno* en un espacio vectorial V de dimensión n es una forma bilineal simétrica definida positiva y por lo tanto de signatura $(0, n)$; una *métrica Lorentziana* será una métrica de signatura $(1, n-1)$ y una *métrica neutral* es una de signatura (m, m) con $m = \frac{1}{2}n$.

Dada una métrica $\langle \cdot, \cdot \rangle$ en un espacio vectorial de dimensión finita V , y dado un subespacio vectorial $W \subseteq V$, se define

$$W^\perp = \{x \in V \mid \langle x, w \rangle = 0 \text{ para todo } w \in W\}$$

y vale

$$(1.2) \quad \dim V = \dim W + \dim W^\perp,$$

en tanto que si $\langle \cdot, \cdot \rangle$ es definida positiva entonces tenemos $V = W \oplus W^\perp$ suma directa de espacios vectoriales.

Un subespacio vectorial W se dice

- *isotrópico* si $W \subseteq W^\perp$,
- *totalmente isotrópico* si $W = W^\perp$,
- *no degenerado* si $W \cap W^\perp = \{0\}$.

En el último caso $V = W \oplus W^\perp$ y por otro lado la existencia de W totalmente isotrópico por (1.2) implica que $\dim V = 2 \dim W$.

Proposición 1.8. *Sea V un espacio vectorial munido de una métrica y sea $W \subset V$ un subespacio totalmente isotrópico de V . Entonces existe un subespacio complementario U de W tal queda*

$$(1.3) \quad V = W \oplus U,$$

donde U puede ser elegido como un subespacio totalmente isotrópico o como un subespacio no degenerado tal que la restricción de la métrica a U resulte un producto interno en U .

Demostración. Ejercicio. □

Observación 1.9. Supongamos que V es un espacio vectorial de dimensión finita y que B_1 y B_2 son formas bilineales sobre V . Entonces fijada una base de V tendremos sus matrices asociadas A_1 y A_2 respectivamente. Así resulta

$$B_1(u, v) = [u]^t A_1 [v] \quad B_2(u, v) = [u]^t A_2 [v].$$

La pregunta es ¿qué relación podemos establecer entre B_1 y B_2 ? El punto es que si suponemos que B_1 es no degenerada entonces existe una T tal que $B_2(u, v) = B_1(u, Tv)$ con lo cual en términos de matrices tenemos

$$A_2 = A_1 C,$$

siendo C la matriz de T en la base elegida. Y si B_1 y B_2 son matrices simétricas la matriz de T resulta simétrica.

2. ÁLGEBRAS DE LIE DOS PASOS NILPOTENTES

Sea \mathfrak{n} un espacio vectorial real de dimensión finita. Sea $\mathfrak{z} \subset \mathfrak{n}$ un subespacio vectorial de \mathfrak{n} . Definiremos una operación binaria en \mathfrak{n} que denotaremos con un corchete $[\cdot, \cdot] : \mathfrak{n} \times \mathfrak{n} \rightarrow \mathfrak{n}$ tal que:

- $[\cdot, \cdot]$ es bilineal y antisimétrica
- $[x, y] \in \mathfrak{z}$ para todos $x, y \in \mathfrak{n}$
- $[x, z] = 0$ para todo $x \in \mathfrak{n}, z \in \mathfrak{z}$.

Definición 2.1. Un par $(\mathfrak{n}, [\cdot, \cdot])$ que satisface lo anterior se dice un *álgebra de Lie real 2-pasos nilpotente*.

La operación binaria definida en \mathfrak{n} se llama un *corchete de Lie en \mathfrak{n}* .

Observemos que esta operación binaria no es conmutativa, es asociativa puesto que

$$[u, [v, w]] = 0 \quad \text{para todos } u, v, w \in \mathfrak{n}$$

y con esto el corchete de Lie en \mathfrak{n} satisface la siguiente *identidad de Jacobi*

$$(2.1) \quad [u, [x, y]] + [x, [y, u]] + [y, [u, x]] = 0 \quad \text{para todos } x, y, z \in \mathfrak{n}.$$

Ejemplo 2.2. Una posibilidad es tomar $\mathfrak{z} = \{0\}$ en un espacio vectorial dado, lo cual es equivalente a $[x, y] = 0$ para todos x, y . En este caso el álgebra de Lie se dice *abeliana* y el corchete de Lie se dice trivial.

Notemos que bajo este corchete trivial, cualquier espacio vectorial puede ser considerado un álgebra de Lie. Algunos autores eligen no considerar a estas álgebras abelianas como ejemplos de álgebras de Lie 2-pasos nilpotentes.

Ejemplo 2.3. *El álgebra de Heisenberg.* Consideremos el conjunto de matrices 3×3 triangulares superiores estrictas

$$\begin{pmatrix} 0 & x & z \\ 0 & 0 & y \\ 0 & 0 & 0 \end{pmatrix}$$

Munido de la suma usual de matrices (sumar coordenada a coordenada) y el producto de una matriz por un escalar, este conjunto resulta un espacio vectorial real de dimensión tres. Una base de tal espacio vectorial está constituido por las matrices

$$X = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad Y = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \quad Z = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Definamos un corchete de Lie en este espacio vectorial por

$$(2.2) \quad [A, B] = AB - BA$$

donde AB denota el producto usual de matrices. Verifiquemos que con este producto en nuestro caso ocurre que

$$[u, v] \in \text{gen}\{Z\} \quad \text{y} \quad [u, Z] = 0 \quad \text{para todos } u, v.$$

Así este espacio vectorial de dimensión tres munido de este corchete es conocido como el álgebra de Heisenberg de dimensión tres que denotaremos \mathfrak{h}_3 .

El álgebra de Heisenberg de dimensión $2n + 1$ denotada \mathfrak{h}_{2n+1} puede ser construida con el corchete de Lie a partir de las matrices $(n + 2) \times (n + 2)$ de la forma

$$\begin{pmatrix} 0 & x_1 & x_2 & \dots & x_n & z \\ 0 & 0 & 0 & \dots & 0 & y_1 \\ 0 & 0 & 0 & \dots & 0 & y_2 \\ \vdots & \ddots & \vdots & \dots & 0 & y_{n-1} \\ 0 & 0 & 0 & \dots & 0 & y_n \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}$$

Notemos que por la bilinealidad de $[\cdot, \cdot]$ el corchete de Lie queda determinado por sus valores en una base de \mathfrak{n} . En efecto, supongamos que \mathfrak{n} es un espacio vectorial de dimensión n y sea e_1, e_2, \dots, e_n una base de \mathfrak{n} . Así cualquier vector en \mathfrak{n} se escribirá como

$$x = x_1e_1 + x_2e_2 + \dots + x_n e_n$$

donde x_i son las coordenadas de x . Luego si $y = \sum_{i=1}^n y_i e_i$ tendremos

$$[x, y] = \left[\sum_{i=1}^n x_i e_i, \sum_{j=1}^n y_j e_j \right] = \sum_{i=1}^n x_i \left[e_i, \sum_{j=1}^n y_j e_j \right] = \sum_{i=1}^n \sum_{j=1}^n x_i y_j [e_i, e_j],$$

o sea que conocidas las coordenadas de x e y para determinar el corchete $[x, y]$ necesitaríamos conocer los corchetes $[e_i, e_j]$. Así

$$[e_i, e_j] = \sum_{k=1}^n c_{ij}^k e_k$$

y los escalares c_{ij}^k son llamados los *coeficientes de estructura* de \mathfrak{n} .

Ejemplo 2.4. *Heisenberg cuaterniónico.* Sea $n \in \mathbb{N}$. Para cada i con $1 \leq i \leq n$ sea \mathbb{H}^i el espacio vectorial de dimensión cuatro con base X_i, Y_i, V_i, W_i . Sea $\mathfrak{v} = \bigoplus \mathbb{H}^i$. Tomemos \mathfrak{z} un espacio vectorial de dimensión tres con base Z_1, Z_2, Z_3 y sea $\mathfrak{n} = \mathfrak{z} \oplus \mathfrak{v}$. Munimos a \mathfrak{n} de un corchete como sigue

$$[z, u] = 0 \quad \text{para todo } z \in \mathfrak{z}$$

y haciendo de $[\cdot, \cdot]$ antisimétrico

$$\begin{aligned} [X_i, Y_i] &= Z_1, & [X_i, V_i] &= Z_2, & [X_i, W_i] &= Z_3, \\ [Y_i, V_i] &= Z_3, & [Y_i, W_i] &= -Z_2, & [V_i, W_i] &= Z_1. \end{aligned}$$

El álgebra de Lie resultante se llama el álgebra de Heisenberg cuaterniónica de dimensión $4n + 3$.

Definición 2.5. Un álgebra de Lie 2-pasos nilpotente \mathfrak{n} es no singular si $\text{ad} : \mathfrak{n} \rightarrow \mathfrak{z}$ es suryectiva para todo $x \in \mathfrak{n} - \mathfrak{z}$.

Ejercicio 2.6. Probar que \mathfrak{h}_{2n+1} es un espacio vectorial de dimensión $2n+1$ encontrando una base para dicho espacio vectorial. Además probar que el corchete de Lie de matrices en (2.2) hace de este espacio vectorial un álgebra de Lie 2-pasos nilpotente.

Ejercicio 2.7. Un álgebra de Lie real \mathfrak{g} es un espacio vectorial real junto con un corchete $[\cdot, \cdot] : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$ que es \mathbb{R} -bilineal, antisimétrico y satisface la identidad de Jacobi (2.1).

Pruebe que el conjunto de matrices reales $n \times n$ munido del corchete de matrices (2.2) es un álgebra de Lie real. Más generalmente cualquier álgebra asociativa A admite un corchete de Lie si se define $[a, b] := ab - ba$ para todos $a, b \in A$.

Ejercicio 2.8. Verifique que \mathfrak{z} es un ideal en \mathfrak{n} , esto es $[x, z] \in \mathfrak{z}$ para todo $x \in \mathfrak{n}, z \in \mathfrak{z}$.

Volvamos a las álgebras de Lie 2-pasos nilpotentes. El subespacio vectorial \mathfrak{z} es un ideal llamado el *centro* de \mathfrak{n} . Otro subespacio distinguido en una tal \mathfrak{n} es el denominado *conmutador* que es el subespacio de \mathfrak{n} generado por los corchetes de elementos de \mathfrak{n} :

$$C(\mathfrak{n}) := \text{gen}\{[x, y] \mid x, y \in \mathfrak{n}\}$$

Entonces para un álgebra de Lie 2-pasos nilpotente \mathfrak{n} ocurre que $C(\mathfrak{n}) \subseteq \mathfrak{z}$ y eligiendo un subespacio complementario $\tilde{\mathfrak{z}} \subseteq \mathfrak{z}$ tal que

$$\mathfrak{z} = C(\mathfrak{n}) \oplus \tilde{\mathfrak{z}}$$

suma directa de espacios vectoriales donde resulta que tanto $C(\mathfrak{n})$ como $\tilde{\mathfrak{z}}$ son ideales de \mathfrak{n} .

3. PRODUCTOS INTERNOS EN \mathfrak{n}

Sea \mathfrak{n} álgebra de Lie 2-pasos nilpotentes y supongamos que $\langle \cdot, \cdot \rangle$ es un producto interno en \mathfrak{n} . En este caso \mathfrak{n} admite una descomposición en suma directa de subespacios de la forma

$$\mathfrak{n} = \mathfrak{z} \oplus \mathfrak{v} \quad \text{con} \quad \mathfrak{v} = \mathfrak{z}^\perp.$$

Usando esta definición uno obtiene las siguientes relaciones

- a) $\langle j(z)x, j(z^*)x \rangle = \langle z, z^* \rangle |x|^2$ para todos $z, z^* \in \mathfrak{z}$, $x \in \mathfrak{v}$;
- b) $\langle j(z)x, j(z)y \rangle = |z|^2 \langle x, y \rangle$ para todos $z \in \mathfrak{z}$, $x, y \in \mathfrak{v}$;
- c) $|j(z)x| = |z||x|$ para todos $z \in \mathfrak{z}$, $x \in \mathfrak{v}$;
- d) $j(z) \circ j(z^*) + j(z^*) \circ j(z) = -2\langle z, z^* \rangle \text{Id}$ para todos $z, z^* \in \mathfrak{z}$.

Ejercicio 3.4. Pruebe las igualdades anteriores (por favor).

Lema 3.5. Sea \mathfrak{n} un álgebra de Lie 2-pasos nilpotente. Entonces \mathfrak{n} es no singular si y sólo si para cualquier producto interno $\langle \cdot, \cdot \rangle$ en \mathfrak{n} las aplicaciones $\{j(z) : z \in \mathfrak{z}\}$ son no singulares en $\mathfrak{v} = \mathfrak{z}^\perp$ para cualquier vector no nulo $z \in \mathfrak{z}$.

3.1. Elementos geométricos en \mathfrak{n} . En función de un producto interno definido en un álgebra de Lie 2-pasos nilpotente en esta parte vamos a dar algunos elementos geométricos. Empezaremos por definir la *conexión métrica* en \mathfrak{n} . Sea $\nabla : \mathfrak{n} \times \mathfrak{n} \rightarrow \mathfrak{n}$ la aplicación \mathbb{R} -bilineal dada por la fórmula

$$\nabla_x y = \frac{1}{2} ([x, y] - \text{ad}(x)^*(y) - \text{ad}(y)^*(x)).$$

No es difícil probar que

$$(3.3) \quad \nabla_x y = \begin{cases} \frac{1}{2}[x, y] & \text{para } x, y \in \mathfrak{v}, \\ \nabla_y x = -\frac{1}{2}j(y)x & \text{para } x \in \mathfrak{v}, y \in \mathfrak{z}, \\ 0 & \text{para } x, y \in \mathfrak{z}. \end{cases}$$

Ejercicio 3.6. Pruebe esto último.

Para x, y elementos en \mathfrak{n} el tensor de curvatura viene dado por

$$R(x, y) = [\nabla_x, \nabla_y] - \nabla_{[x, y]}.$$

Usando (3.3) uno obtiene

$$(3.4) \quad R(x, y)z = \begin{cases} \frac{1}{2}j([x, y])z - \frac{1}{4}j([y, z])x + \frac{1}{4}j([x, z])y & \text{para } x, y, z \in \mathfrak{v}, \\ \frac{1}{4}[j(z)x, y] - [x, j(z)y] & \text{para } x, y \in \mathfrak{v}, z \in \mathfrak{z}, \\ -\frac{1}{4}[x, j(y)z] & \text{para } x, z \in \mathfrak{v}, y \in \mathfrak{z}, \\ -\frac{1}{4}j(y)j(z)x & \text{para } x \in \mathfrak{v}, y, z \in \mathfrak{z}, \\ \frac{1}{4}[j(x), j(y)]z & \text{para } x, y \in \mathfrak{z}, z \in \mathfrak{v}, \\ 0 & \text{para } x, y, z \in \mathfrak{z}. \end{cases}$$

Ejercicio 3.7. Pruebe lo anterior.

Sea $\Pi \subseteq \mathfrak{n}$ un plano, esto es un subespacio vectorial de dimensión dos y sea Q el número real dado por

$$Q(x, y) = \langle x, x \rangle \langle y, y \rangle - \langle x, y \rangle^2.$$

La *curvatura seccional* de Π es el número real

$$K(x, y) := \langle R(x, y)y, x \rangle / Q(x, y),$$

el cual es independiente de la elección de la base.

Tomemos ahora una base ortonormal para Π , esto es un conjunto linealmente independiente $\{x, y\}$ tal que $\langle x, y \rangle = 0$ y $\langle x, x \rangle = 1 = \langle y, y \rangle$.

Por (3.4) se obtiene

$$(3.5) \quad K(x, y) = \begin{cases} -\frac{3}{4}\langle [x, y], [x, y] \rangle & \text{para } x, y \in \mathfrak{v}, \\ -\frac{1}{4}\langle j(y)x, j(y)x \rangle & \text{para } x \in \mathfrak{v}, y \in \mathfrak{z}, \\ 0 & \text{para } x, y \in \mathfrak{z}. \end{cases}$$

El tensor de Ricci viene dado por $Ric(x, y) = \text{traza}(z \rightarrow R(z, x)y)$, $z \in \mathfrak{n}$ para elementos arbitrarios $x, y \in \mathfrak{n}$.

Proposición 3.8. *Sea $\{z_i\}$ una base ortonormal de \mathfrak{z} y $\{v_j\}$ una base ortonormal de \mathfrak{v} . Vale*

$$Ric(x, y) = \begin{cases} 0 & \text{para } x \in \mathfrak{v}, y \in \mathfrak{z}, \\ \frac{1}{2} \sum_i \langle j(z_i)^2 x, y \rangle & \text{para } x, y \in \mathfrak{v}, \\ -\frac{1}{4} \sum_j \langle j(x)j(y)v_j, v_j \rangle & \text{para } x, y \in \mathfrak{z}. \end{cases}$$

Debido a simetrías del tensor de curvatura el tensor de Ricci resulta una aplicación bilinear simétrica en \mathfrak{n} y por lo tanto existe una aplicación lineal simétrica $T : \mathfrak{n} \rightarrow \mathfrak{n}$ tal que $Ric(x, y) = \langle Tx, y \rangle$ para todos $x, y \in \mathfrak{n}$. La aplicación T es llamada la *transformación de Ricci*. Sea $\{e_k\}$ una base ortonormal de \mathfrak{n} ; se tiene

$$Ric(x, y) = \sum_k \langle R(e_k, x)y, e_k \rangle = \langle -\sum_k R(e_k, x)e_k, y \rangle$$

lo que implica

$$(3.6) \quad T(x) = -\sum_k R(e_k, x)e_k.$$

Por lo expuesto en la Proposición 3.8 tenemos que \mathfrak{z} y \mathfrak{v} son subespacios T -invariantes y

$$T(x) = \begin{cases} \frac{1}{2} \sum_i j(z_i)^2 x & x \in \mathfrak{v}, \\ \frac{1}{4} \sum_j [v_j, j(x)v_j] & x \in \mathfrak{z}. \end{cases}$$

donde $\{z_i\}$ y $\{v_j\}$ son bases ortonormales de \mathfrak{z} y \mathfrak{v} respectivamente.

4. MÉTRICAS INDEFINIDAS EN \mathfrak{n}

Supongamos ahora que \mathfrak{n} está provista de una métrica $\langle \cdot, \cdot \rangle$. En este caso podremos considerar dos situaciones: que el centro sea degenerado o no.

4.1. Centro no degenerado. En este caso se procede de modo similar al caso de producto interno y empezamos por tomar

$$\mathfrak{n} = \mathfrak{z} \oplus \mathfrak{v} \quad \text{con } \mathfrak{v} = \mathfrak{z}^\perp.$$

También es posible definir una aplicación $j(z)$ como en (3.1) que resultará una aplicación antisimétrica o antiadjunta respecto de $\langle \cdot, \cdot \rangle_{\mathfrak{v}}$ pero tal que su matriz no será necesariamente antisimétrica. Varios resultados del caso definido positivo que vimos en la sección anterior pueden ser adaptados a este caso. Para más detalle sobre geometría pseudo-riemanniana ver [7].

Observación 4.1. Un ejercicio interesante usando los conceptos en [7] es escribir lo hecho en la sección anterior para esta situación. Se puede corroborar resultados mirando [9].

4.2. Centro degenerado. Aquí la situación se complica ya que no es tan fácil proponer una descomposición de \mathfrak{n} en una suma directa de su centro y un espacio complementario canónico o conveniente. El problema es que el centro \mathfrak{z} es degenerado y controlar \mathfrak{z} y \mathfrak{z}^\perp implica fijar otros elementos (ver por ejemplo [9] y referencias allí para diferentes soluciones a este problema de descomposición).

Trabajaremos aquí el caso de las denominadas métricas ad-invariantes.

Definición 4.2. Sea \mathfrak{n} un álgebra de Lie real. Una métrica $\langle \cdot, \cdot \rangle$ se dice ad-invariante si

$$(4.1) \quad \langle [x, y], z \rangle + \langle y, [x, z] \rangle = 0 \quad \text{para todos } x, y, z \in \mathfrak{n}.$$

Lema 4.3. Sea $(\mathfrak{n}, \langle \cdot, \cdot \rangle)$ un álgebra de Lie 2-pasos nilpotente equipada con una métrica ad-invariante.

1. Si \mathfrak{h} es un ideal de \mathfrak{n} entonces \mathfrak{h}^\perp es también un ideal en \mathfrak{n} .
2. $C^1(\mathfrak{n})^\perp = \mathfrak{z}(\mathfrak{n})$.

Observación 4.4. La segunda afirmación del lema anterior admite un enunciado más general para cualquier álgebra de Lie \mathfrak{g} y vale $C^r(\mathfrak{g})^\perp = C_r(\mathfrak{g})$ para todo r (ver por ejemplo [2]).

Así este lema y lo visto antes implican

$$(4.2) \quad \dim \mathfrak{n} = \dim \mathfrak{z}(\mathfrak{n}) + \dim C^1(\mathfrak{n}).$$

Ejemplo 4.5. Sea \mathfrak{n} un álgebra de Lie 2-pasos nilpotente equipada con una métrica ad-invariante. Asumamos que el centro coincide con el conmutador $\mathfrak{z}(\mathfrak{n}) = C^1(\mathfrak{n})$, entonces (4.2) dice que la métrica es neutral y $\dim \mathfrak{n} = 2 \dim \mathfrak{z}(\mathfrak{n})$. Notemos que ésta es una restricción algebraica muy fuerte y así por ejemplo el álgebra de Heisenberg \mathfrak{h}_n no admite una métrica ad-invariante.

Ejemplo 4.6. Ejemplos de álgebras de Lie nilpotentes satisfaciendo (4.2) se obtienen al considerar el producto semidirecto de \mathfrak{n} con su espacio dual via la representación coadjunta. Esto es tomamos

$$(4.3) \quad \mathfrak{g} := \mathfrak{n} \oplus \mathfrak{n}^* \quad \text{suma directa de espacios vectoriales}$$

y munimos a \mathfrak{g} del corchete de Lie dado por

$$[(x_1, \varphi_1), (x_2, \varphi_2)] = ([x_1, x_2], -\text{ad}(x_1)^*\varphi_2 + \text{ad}(x_2)^*\varphi_1)$$

donde $\text{ad}(x)^*\varphi(y) = \varphi([x, y])$ para $x, y \in \mathfrak{n}$, $\varphi \in \mathfrak{n}^*$. La métrica neutral en \mathfrak{g} está dada por

$$(4.4) \quad \langle (x_1, \varphi_1), (x_2, \varphi_2) \rangle = \varphi_1(x_2) + \varphi_2(x_1) \quad \text{para } x_1, x_2 \in \mathfrak{n}, \varphi_1, \varphi_2 \in \mathfrak{n}^*.$$

Entonces la métrica neutral en \mathfrak{g} resulta una métrica ad-invariante.

Ejercicio 4.7. Pruebe que \mathfrak{g} es un álgebra de Lie 2-pasos nilpotente si \mathfrak{n} lo es y que la métrica definida en el ejemplo anterior es ad-invariante y neutral.

Observación 4.8. La condición (4.2) no asegura la existencia de métricas ad-invariantes en álgebras de Lie 2-pasos nilpotentes (ver por ejemplo [8]).

Proposición 4.9. Sea \mathfrak{n} un álgebra de Lie 2-pasos nilpotente munida de una métrica ad-invariante $\langle \cdot, \cdot \rangle$. Sea \mathfrak{m} un subespacio complementario del conmutador en el centro, esto es

$$\mathfrak{z} = C^1(\mathfrak{n}) \oplus \mathfrak{m},$$

suma directa de espacios vectoriales. Entonces \mathfrak{m} es no degenerado.

Demostración. Sea $y \in \mathfrak{m}$ y supongamos que $\langle x, y \rangle = 0$ para todo $x \in \mathfrak{m}$. Como $y \in \mathfrak{z}(\mathfrak{n}) = C^1(\mathfrak{n})^\perp$ resulta que $\langle y, z \rangle = 0$ para todo $z \in \mathfrak{z}$. Así $z \in C^1(\mathfrak{n}) \cap \mathfrak{m} = \{0\}$ con lo cual $y = 0$, lo que demuestra que \mathfrak{m} es no degenerado. \square

Denotemos por $\tilde{\mathfrak{z}}$ un subespacio complementario de $C^1(\mathfrak{n})$ en el centro \mathfrak{z} . No es difícil ver que $\tilde{\mathfrak{z}}$ es un ideal de \mathfrak{n} . Entonces un álgebra de Lie 2-pasos nilpotente munida de una métrica ad-invariante se descompone como

$$\mathfrak{n} = \tilde{\mathfrak{z}} \oplus \tilde{\mathfrak{n}}$$

donde $\tilde{\mathfrak{n}}$ resulta ser un álgebra de Lie 2-pasos nilpotente cuyo centro coincide con su conmutador. Así la descomposición anterior es una suma directa de ideales, donde el conmutador de \mathfrak{n} coincide con el de $\tilde{\mathfrak{n}}$.

Se define el *corango* de \mathfrak{n} como el número $corango \mathfrak{n} = \dim \mathfrak{z} - \dim C^1(\mathfrak{n})$.

Ejercicio 4.10. Escribir el conmutador y el centro de \mathfrak{n} y $\tilde{\mathfrak{n}}$.

En virtud de lo anterior básicamente la métrica ad-invariante queda determinada a partir de un álgebra de Lie 2-pasos nilpotente en la cual el conmutador coincide con el centro, esto es de corango cero. En lo que sigue intentaremos dar un modelo similar a lo hecho para productos internos, esto es definir ciertas aplicaciones similares a las j .

El siguiente proceso provee una forma de construir álgebras de Lie 2-pasos nilpotente con una métrica ad-invariante. Sea \mathfrak{v} un espacio vectorial equipado con un producto interno $\langle \cdot, \cdot \rangle_{\mathfrak{v}}$ y sea $\rho : \mathfrak{v} \rightarrow \mathfrak{so}(\mathfrak{v})$ una aplicación lineal inyectiva satisfaciendo

$$(4.5) \quad \rho(u)v + \rho(v)u = 0 \quad \text{para todos } u, v \in \mathfrak{v}$$

o equivalentemente $\rho(x)x = 0$ para todo $x \in \mathfrak{v}$. Denotemos por

$$\mathfrak{n}(\mathfrak{v}, \rho) = \mathfrak{v}^* \oplus \mathfrak{v} \quad \text{suma directa de espacios vectoriales}$$

munido de la métrica neutral $\langle \cdot, \cdot \rangle$ ya dada en (4.4). Definamos un corchete de Lie $[\cdot, \cdot]$ en $\mathfrak{n}(\mathfrak{v}, \rho)$ con las condiciones

$$[\mathfrak{v}^*, \mathfrak{n}(\mathfrak{v}, \rho)] = 0 \quad \text{y} \quad [\mathfrak{v}, \mathfrak{v}] \subseteq \mathfrak{v}^*$$

y de modo que

$$\langle [u, v], w \rangle = \langle \rho(w)u, v \rangle_{\mathfrak{v}} \quad \text{para todos } u, v, w \in \mathfrak{v}.$$

Como $\rho(u) \in \mathfrak{so}(\mathfrak{v})$ para cualquier $v \in \mathfrak{v}$ resulta que $[\cdot, \cdot]$ es antisimétrico y (4.5) implica que $\text{ad}(x)$ son antisimétricas. Así $\mathfrak{n}(\mathfrak{v}, \rho)$ es un álgebra de Lie 2-pasos nilpotente de corango cero y la cual admite una métrica ad-invariante. Notemos que el corango nulo resulta del hecho que ρ es inyectiva.

Esta álgebra $\mathfrak{n}(\mathfrak{v}, \rho)$ se denomina el *cotangente modificado* de \mathfrak{v} . Finalmente podemos adjuntar por suma directa ortogonal un espacio vectorial \mathbb{R}^m munido de cualquier métrica para obtener un álgebra de Lie 2-pasos nilpotente de corango m . Esta es esencialmente la forma de las álgebras de Lie 2-pasos nilpotente munidas de métricas ad-invariantes según lo probado en [8].

Teorema 4.11. *Sea $(\mathfrak{n}, \langle \cdot, \cdot \rangle)$ un álgebra de Lie 2-pasos nilpotente de corango m munida de una métrica ad-invariante. Entonces $(\mathfrak{n}, \langle \cdot, \cdot \rangle)$ es isométrica isomorfa a una suma directa ortogonal del álgebra de Lie abeliana \mathbb{R}^m y el cotangente modificado de $\mathfrak{v} := C^1(\mathfrak{n})$.*

Aclaremos que un isomorfismo $\psi : \mathfrak{n}_1 \rightarrow \mathfrak{n}_2$ es un isomorfismo de espacios vectoriales tal que $\psi[x, y] = [\psi x, \psi y]$ para todos $x, y \in \mathfrak{n}_1$. Se dice isométrico si es una isometría respecto de métricas en \mathfrak{n}_1 y \mathfrak{n}_2 respectivamente.

Ejemplo 4.12. Sea \mathfrak{v} un espacio vectorial munido de un producto interno $\langle \cdot, \cdot \rangle$ y una base ortonormal $\mathcal{B} = \{v_1, v_2, \dots, v_p\}$. Sea $\rho : \mathfrak{v} \rightarrow \mathfrak{so}(\mathfrak{v})$ una aplicación lineal inyectiva satisfaciendo (4.5). Sea A^i la matriz de $\rho(v_i)$ en la base \mathcal{B} y denotemos por a_{jk}^i a las entradas de A^i . Entonces tenemos que cada matriz A^i es antisimétrica, con lo cual vale

$$(4.6) \quad a_{jk}^i = -a_{kj}^i \quad \text{para todos } i, j, k,$$

por otro lado la relación (4.5) aplicada a los $\rho(v_i)$ dice que

$$(4.7) \quad a_{jk}^i = -a_{ki}^j \quad \text{para todos } i, j, k,$$

lo cual prueba que $a(i, j, k)$ es trilineal y alternante como función de i, j, k . Así en dimensión tres, las matrices que obtenemos son de la forma

$$A^1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & a \\ 0 & -a & 0 \end{pmatrix} \quad A^2 = \begin{pmatrix} 0 & 0 & -a \\ 0 & 0 & 0 \\ a & 0 & 0 \end{pmatrix} \quad A^3 = \begin{pmatrix} 0 & a & 0 \\ -a & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

las cuales conforman un conjunto linealmente independiente para $a \neq 0$.

Ejercicio 4.13. Pensar cuatro matrices reales 4×4 que satisfagan (4.6) y (4.7) y probar que un tal conjunto no puede ser linealmente independiente con lo cual no hay álgebras de Lie 2-pasos nilpotentes de corango nulo en dimensión ocho que admitan métrica ad-invariante.

Observación 4.14. En [8] se puede ver que existen álgebras de Lie 2-pasos nilpotentes de corango nulo con métricas ad-invariantes en dimensiones mayores que ocho.

REFERENCIAS

- [1] J. Berndt, F. Tricceri, L. Vanhecke, *Generalized Heisenberg Groups and Damek-Ricci Harmonic Spaces*, Lecture Notes in Math., Vol. 1598, Springer-Verlag (1995).
- [2] V. del Barco, G. Ovando, *Free Nilpotent Lie Algebras Admitting Ad-Invariant Metrics* J. Algebra, **366**, 205–216 (2012).
- [3] P. Eberlein, *Geometry of 2-step nilpotent groups with a left-invariant metric*, Ann. scient. Éc. Norm. Sup. 4^e série, **27**, 611–660 (1994).
- [4] K. Hoffman, R. Kunze, *Álgebra lineal*, Prentice Hall (1973).
- [5] J. Lauret, *Homogeneous nilmanifolds of dimension 3 and 4*, Geometriae Dedicata **68**, 145–155 (1997).
- [6] J. Lauret, *Modified H-type groups and symmetric-like Riemannian spaces*, Diff. Geom. Appl., **10**, 121–143 (1999).
- [7] B. O'Neill, *Semi-Riemannian Geometry: with Applications to Relativity*, Academic Press, (1983).
- [8] G. Ovando, *Two-step nilpotent Lie algebras with ad-invariant metrics and a special kind of skew-symmetric maps*, J. Algebra Appl. **6** no. 6, 897–917 (2007).
- [9] G. Ovando, *Naturally reductive pseudo Riemannian 2-step nilpotent Lie groups* arXiv:0911.4067 (aparecerá en Houston J. Math.).
- [10] G. Ovando, *Examples of naturally reductive pseudo-Riemannian Lie groups*, arXiv:1104.4971 (2011).

CONICET — UNIVERSIDAD NACIONAL DE ROSARIO. DEPTO. DE MATEMÁTICA, ECEN — FCEIA, PELLEGRINI 250. 2000 ROSARIO

E-mail address: gabriela@fceia.unr.edu.ar

ANILLOS Y MÓDULOS SEMISIMPLES

ANDREA SOLOTAR – MARIANO SUÁREZ-ALVAREZ

RESUMEN. Este curso tratará del estudio de un tipo de anillos para los que hay una descripción completa de la categoría de módulos: los anillos semisimples.

Nos ocuparemos en particular del caso de representaciones de grupos finitos sobre k -espacios vectoriales, cuando la característica de k no divide al orden del grupo.

ÍNDICE

| | |
|---|-----|
| Introducción | 137 |
| 1. Módulos simples y semisimples | 137 |
| 1.1. Módulos simples | 137 |
| 1.2. Módulos semisimples | 139 |
| 1.3. El radical de un módulo | 142 |
| 2. Anillos semisimples | 144 |
| 2.1. Caracterización de anillos semisimples | 144 |
| 2.2. Teorema de Wedderburn | 146 |
| 2.3. El radical de un anillo | 148 |
| 2.4. Algebras de grupo | 151 |
| 2.5. Ejercicios | 153 |
| Referencias | 156 |

INTRODUCCIÓN

Los anillos semisimples tienen categorías de módulos muy sencillas: todos los módulos son proyectivos e inyectivos. Los anillos semisimples son hereditarios. En general, la dimensión global de un anillo A mide cuan lejos está A de ser semisimple.

Para un anillo conmutativo, ser semisimple es equivalente a ser un producto finito de cuerpos. En el caso no conmutativo, el teorema de Wedderburn dice que todo anillo semisimple es un producto finito de anillos de matrices sobre anillos de división.

El objetivo de este curso es caracterizar los anillos y los módulos semisimples y demostrar sus propiedades más importantes.

1. MÓDULOS SIMPLES Y SEMISIMPLES

1.1. Módulos simples. Sea A un anillo. Trabajaremos en general con A -módulos a izquierda.

Nos interesa describir propiedades de A que hagan que todo A -módulo sea proyectivo, o inyectivo, o libre.

Por ejemplo, para que todo A -módulo sea proyectivo se necesita que todo A -módulo sea sumando directo de un libre. En particular, todo ideal de A debe ser un sumando directo de A .

Definición 1.1. Decimos que un A -módulo S es *simple* si es no nulo y no posee submódulos propios no triviales.

Ejemplo 1.2. Si $A = k$ es un cuerpo o, más generalmente, un álgebra de división, entonces un A -módulo es simple sii tiene dimensión 1.

Ejemplo 1.3. Si $A = \mathbb{Z}$, entonces un A -módulo, es decir un grupo abeliano, es simple sii es finito de orden primo.

Es consecuencia directa de la definición que un módulo no nulo es simple si está generado por cualquiera de sus elementos no nulos.

Lema 1.4. *Sea A un anillo.*

- Si $\mathfrak{a} \triangleleft_l A$, entonces A/\mathfrak{a} es simple sii \mathfrak{a} es un ideal izquierdo maximal. En particular, existen siempre módulos simples.
- Si S es un módulo simple y $m \in S \setminus 0$, entonces $\text{ann}(m)$ es un ideal a izquierda maximal en A y $S \cong A/\text{ann}(m)$.

Demostración. La primera afirmación es clara. La existencia de módulo simples es consecuencia de ella y de la existencia de ideales a izquierda maximales. Si S es un módulo simple y $m \in S \setminus 0$, entonces el morfismo $a \in A \mapsto am \in S$ es sobreyectivo, así que $S \cong A/\text{ann}(m)$ y, por la primera parte, $\text{ann}(m)$ es un ideal a izquierda maximal. \square

La segunda afirmación de este lema implica que la colección de las clases de isomorfismo de A -módulos simples es un conjunto: en efecto, cada una de esas clases posee un representante que es un ideal a izquierda maximal de A y estos forman un conjunto.

Lema 1.5. *Sea $\phi : A \rightarrow B$ un morfismo de anillos sobreyectivo y sea S un B -módulo simple. Entonces el A -módulo $\phi^*(S)$ obtenido de S por restricción de escalares a lo largo de ϕ es simple.*

Demostración. Como ϕ es sobreyectivo, el grupo abeliano subyacente de un A -submódulo propio no trivial de $\phi^*(S)$ determina un B -submódulo propio no trivial de S . \square

Notemos que sin la condición de que el morfismo ϕ sea sobreyectivo la conclusión del lema no es necesariamente válida. Por ejemplo, si $\phi : \mathbb{Z} \rightarrow \mathbb{Q}$ es la inclusión, entonces el \mathbb{Z} -módulo $\phi^*(\mathbb{Q})$, obtenido del \mathbb{Q} -módulo simple \mathbb{Q} , no es simple.

Lema 1.6. *Sean A y B anillos y sean $\pi_1 : A \times B \rightarrow A$ y $\pi_2 : A \times B \rightarrow B$ las proyecciones canónicas. Sean \mathcal{S}_A y \mathcal{S}_B conjuntos completos de representantes de las clases de isomorfismo de los A - y B -módulos simples, respectivamente. Entonces*

$$\mathcal{S} = \{\pi_1^*(S) : S \in \mathcal{S}_A\} \cup \{\pi_2^*(S) : S \in \mathcal{S}_B\}$$

es un conjunto completo de representantes de las clases de isomorfismo de los $A \times B$ -módulos simples.

Demostración. Sean $e_1 = (1_A, 0)$, $e_2 = (0, 1_B) \in A \times B$. Para cada $A \times B$ -módulo M , consideramos los subgrupos abelianos $M_1 = e_1 M$ y $M_2 = e_2 M$ de M . Definimos una acción de A sobre M_1 poniendo

$$a \cdot m = (a, 0)m, \quad \forall a \in A, m \in M_1.$$

Notemos que esto tiene sentido porque $(a, 0)m = e_1(a, 0)m \in M_1$ si $a \in A$ y $m \in M_1$. Es fácil ver que, dotado de esta acción, M_1 resulta un A -módulo. De manera similar hacemos de M_2 un B -módulo. Calculando directamente, se ve que la aplicación

$$m \in M \mapsto (e_1 m, e_2 m) \in \pi_1^*(M_1) \oplus \pi_2^*(M_2)$$

es un isomorfismo de $A \times B$ -módulos.

Sea S un $A \times B$ -módulo simple. Las observaciones recién hechas implican que hay un isomorfismo de $A \times B$ -módulos $S \cong \pi_1^*(S_1) \oplus \pi_2^*(S_2)$. Como S es simple, necesariamente o bien $S_1 = 0$ o bien $S_2 = 0$. Supongamos, por ejemplo, que es $S_2 = 0$. Para ver que S es isomorfo a un elemento de \mathcal{S} , entonces, alcanza con mostrar que S_1 es simple, pero esto es inmediato, ya que todo A -submódulo propio no trivial de S_1 es un $A \times B$ -submódulo propio no trivial de S .

Para terminar, tenemos que ver que los elementos de \mathcal{S} son no isomorfos dos a dos. Consideremos primero un par de A -módulos simples $S, T \in \mathcal{S}_A$ tales que existe un isomorfismo $f : \pi_1^*(S) \rightarrow \pi_1^*(T)$. Es claro que el isomorfismo de grupos abelianos $S \rightarrow T$ subyacente a f es A -lineal, así que es $S \cong T$ y, entonces, $S = T$. De la misma forma, si $S, T \in \mathcal{S}_B$ son B -módulos simples tales que $\pi_2^*(S) \cong \pi_2^*(T)$, entonces $S = T$.

Queda entonces solamente por considerar la posibilidad de que existan un A -módulo simple S y un B -módulo simple T tales que $\pi_1^*(S) \cong \pi_2^*(T)$. De hecho, esto no puede ocurrir porque $e_1\pi_1^*(S) = \pi_1^*(S) \neq 0$ y $e_1\pi_2^*(T) = 0$. \square

Proposición 1.7. (Schur, 1905) Sean S y S' dos A -módulos simples y M un A -módulo cualquiera.

- Si $f : S \rightarrow M$ es un morfismo no nulo, entonces f es inyectivo.
- Si $g : M \rightarrow S$ es un morfismo no nulo, entonces g es sobreyectivo.
- Todo morfismo no nulo $f : S \rightarrow S'$ es un isomorfismo.

En particular, $\text{End}_A(S)$ es un anillo de división.

Demostración. Las dos primeras afirmaciones siguen inmediatamente de la observación de que $\ker f$ e $\text{im } g$ son submódulos de S . La tercera es consecuencia de las dos primeras. \square

Proposición 1.8. Sea $r \in \mathbb{N}$ y sea $\{S_i : 1 \leq i \leq r\}$ un conjunto de A -módulos simples no isomorfos dos a dos. Si $1 \leq i \leq r$, sea $n_i \in \mathbb{N}$ y pongamos, si $1 \leq j \leq n_i$, $S_{i,j} = S_i$. Sea $M = \bigoplus_{i=1}^r \bigoplus_{j=1}^{n_i} S_{i,j}$. Entonces hay un isomorfismo de anillos

$$\text{End}_A(M) \cong M_{n_1}(D_1) \times \cdots \times M_{n_r}(D_r)$$

con $D_i = \text{End}_A(S_i)$ para cada $i \in \{1, \dots, r\}$.

Demostración. Esto es consecuencia inmediata de la descripción de los endomorfismos de M como matrices de morfismos $S_{i,j} \rightarrow S_{i',j'}$ y la tercera parte de 1.7. \square

1.2. Módulos semisimples. Si N y S son submódulos de un A -módulo M , se dice que S es un complemento de N si $S \cap N = 0$ y $S + N = M$. Es claro que se trata de una relación simétrica, que el complemento cuando existe no es único y que hay anillos A con A -módulos M con submódulos que no tienen complemento. Veremos en esta sección que la existencia de complemento para todo submódulo es una característica de los módulos semisimples.

Definición 1.9. Un A -módulo M es *semisimple* si M es suma de submódulos simples. El anillo A es *semisimple* si A , considerado como A -módulo a izquierda, es semisimple.

Ejemplos 1.10.

- Todo módulo simple es semisimple. En particular, el módulo $\{0\}$ es semisimple.
- Si k es un cuerpo, todo k -espacio vectorial es k -módulo semisimple.
- Si k es un cuerpo y $A = k \times \cdots \times k$ (n -factores), entonces A es un anillo semisimple.

- \mathbb{Z} no es un \mathbb{Z} -módulo semisimple pues los ideales de \mathbb{Z} no son sumandos directos de \mathbb{Z} .
- Un grupo abeliano G es simple si y sólo si $G \cong \mathbb{Z}_p$ para algún número primo p . G es semisimple si y sólo si

$$G \cong \bigoplus_{p \text{ primo}} \mathbb{Z}_p^{(I_p)}$$

para ciertos conjuntos I_p .

- Si k es un cuerpo, G es un grupo finito y la característica de k divide a $|G|$, entonces kG no es kG -módulo semisimple.

Lema 1.11. *Sea $M = \sum_{i \in I} S_i$ un A -módulo con S_i simple para cada $i \in I$. Si $N \subset M$ es un submódulo, entonces existe $J \subset I$ tal que $M = N \oplus \bigoplus_{i \in J} S_i$.*

Demostración. Sea $\mathcal{J} = \{J \subset I : \text{la suma } N + \sum_{i \in J} S_i \text{ es directa}\}$. Claramente $\emptyset \in \mathcal{J}$ así que $\mathcal{J} \neq \emptyset$. Si $C = \{J_\lambda\}_{\lambda \in \Lambda}$ es una cadena en \mathcal{J} , pongamos $J_C = \bigcup_{\lambda \in \Lambda} J_\lambda$. Veamos que $J_C \in \mathcal{J}$:

- En primer lugar, si $m \in N \cap \sum_{i \in J_C} S_i$ existe un subconjunto finito $I' \subset J_C$ tal que $m \in \sum_{i \in I'} S_i$ y, como C es una cadena, es posible encontrar $\lambda \in \Lambda$ con $I' \subset J_\lambda$, de manera que $m \in N \cap \sum_{i \in J_\lambda} S_i = 0$. Esto nos dice que $N \cap \sum_{i \in J_C} S_i = 0$.
- Por otro lado, si $i_0 \in J_C$ y $m \in S_{i_0} \cap (N + \sum_{i \in J_C \setminus \{i_0\}} S_i)$, existe un subconjunto finito $I' \subset J_C \setminus \{i_0\}$ tal que $m \in S_{i_0} \cap (N + \sum_{i \in I'} S_i)$. Si $\lambda \in \Lambda$ es tal que $I' \cup \{i_0\} \subset J_\lambda$, entonces $m \in S_{i_0} \cap (N + \sum_{i \in J_\lambda \setminus \{i_0\}} S_i) = 0$. Concluimos que $S_{i_0} \cap (N + \sum_{i \in J_C \setminus \{i_0\}} S_i) = 0$.

El lema de Zorn asegura, en estas condiciones, que \mathcal{J} posee un elemento maximal J . Pongamos $M' = N + \sum_{i \in J} S_i$. La elección de J implica, por supuesto, que esta suma es directa.

Para terminar, veamos que $M' = M$. Para hacerlo, y como $M = \sum_{i \in I} S_i$, basta mostrar que para cada $i \in I$ tenemos $S_i \subset M'$. Consideremos entonces $i_0 \in I$ y supongamos que $S_{i_0} \not\subset M'$. Como S_{i_0} es simple, debe ser entonces $S_{i_0} \cap M' = 0$ y, en particular, $J \cup \{i_0\} \in \mathcal{J}$. Esto contradice la elección de J , así que nuestra suposición debe ser falsa, esto es, debe ser $S_{i_0} \subset M'$. \square

Lema 1.12. *Sea M un A -módulo tal que todo submódulo de M es un sumando directo. Entonces todo submódulo de M posee un submódulo simple.*

Demostración. Basta probar que todo submódulo cíclico de M posee un submódulo simple. Consideremos entonces $m \in M \setminus 0$ y $Am \subset M$ el submódulo cíclico generado por m en M . Sea $\mathfrak{a} \triangleleft_l A$ un ideal a izquierda maximal tal que $\mathfrak{a} \supset \text{ann}(m)$. Entonces $\mathfrak{a}m$ es un submódulo maximal de Am y $Am/\mathfrak{a}m$ es simple.

Por hipótesis, existe $L \subset M$ tal que $M = \mathfrak{a}m \oplus L$. Entonces

$$Am = (\mathfrak{a}m \oplus L) \cap Am = \mathfrak{a}m \oplus (L \cap Am)$$

y vemos que $L \cap Am \cong Am/\mathfrak{a}m$ es un submódulo simple de Am . \square

Teorema 1.13. *Sea M un A -módulo. Las siguientes condiciones son equivalentes:*

1. M es semisimple;
2. M es suma directa de submódulos simples;
3. todo submódulo de M es un sumando directo.

Demostración. Para ver que la primera afirmación implica la segunda, basta tomar $N = 0$ en 1.11. Es inmediato que la segunda implica la primera, y es consecuencia directa de 1.11 que la primera implica la tercera.

Veamos, para terminar, que 3) implica 1). Sea M un A -módulo en el que todo submódulo es un sumando directo, sea M' la suma de todos los submódulos simples de M y supongamos, para llegar a una contradicción, que $M' \subsetneq M$. Por hipótesis, existe un submódulo $N \subset M$ no nulo tal que $M = M' \oplus N$ y, por 1.12, existe un submódulo $S \subset N$ simple. Como $S \cap M' = 0$, esto contradice la elección de M' . \square

Corolario 1.14. *Sea $M = \sum_{i \in I} S_i$ con S_i simple para cada $i \in I$ y sea $N \subset M$ un submódulo. Entonces existe $J \subset I$ tal que $N \cong \bigoplus_{i \in J} S_i$.*

En particular, todo submódulo de un módulo semisimple es semisimple.

Demostración. El teorema implica que N es un sumando directo de M , de manera que existe un submódulo $P \subset M$ tal que $M = N \oplus P$ y 1.11 nos da un conjunto $J \subset I$ tal que $M = \bigoplus_{i \in J} S_i \oplus P$. Luego $N \cong M/P \cong \bigoplus_{i \in J} S_i$. \square

Notemos que no es cierto, en las condiciones del corolario, que exista $J \subset I$ tal que $N = \bigoplus_{i \in J} S_i$. Por ejemplo, sea $A = k$ un cuerpo, $M = k^2$, $\{e_1, e_2\}$ la base canónica de M , $I = \{1, 2\}$, $S_i = \langle e_i \rangle$ si $i \in I$ y $N = \langle e_1 + e_2 \rangle$. Entonces S_i es simple para $i \in I$ y $M = \sum_{i \in I} S_i$ es semisimple, pero claramente N no es suma de una parte de $\{S_i : i \in I\}$.

Corolario 1.15. *Si*

$$0 \longrightarrow M' \xrightarrow{f} M \longrightarrow M'' \longrightarrow 0$$

es una sucesión exacta de A -módulos y M es semisimple, entonces la sucesión se parte y tanto M' como M'' son semisimples.

Demostración. El submódulo $f(M')$ de M es un sumando directo, así que la sucesión exacta se parte y $M \cong M' \oplus M''$. Luego M' y M'' son isomorfos a submódulos de M , que son semisimples en vista de 1.14. \square

Observación 1.16. Si M es un A -módulo con un submódulo semisimple N tal que además M/N es semisimple, no es cierto en general que M sea semisimple. Un (contra)ejemplo de esta situación es la extensión

$$0 \longrightarrow \mathbb{Z}_p \longrightarrow \mathbb{Z}_{p^2} \longrightarrow \mathbb{Z}_p \longrightarrow 0$$

Proposición 1.17. *Sea $M = \bigoplus_{i \in I} S_i$ un A -módulo semisimple con S_i simple para cada $i \in I$. Entonces M es artiniiano sii es n otheriano sii es finitamente generado sii I es finito.*

Demostraci3n. Es claro que si I es infinito, entonces M no es ni artiniiano, ni n otheriano, ni finitamente generado. Supongamos entonces que I es finito y hagamos inducci3n sobre $|I|$; notemos que si $|I| \leq 1$ entonces no hay nada que probar.

Ahora bien, si $i_0 \in I$, entonces hay una sucesi3n exacta corta

$$0 \longrightarrow \bigoplus_{i \in I \setminus i_0} S_i \longrightarrow M \longrightarrow S_{i_0} \longrightarrow 0$$

Aplicando la hip3tesis de inducci3n, vemos que $\bigoplus_{i \in I \setminus i_0} S_i$ es artiniiano y n otheriano. Como lo mismo vale para S_{i_0} , entonces M es artiniiano y n otheriano, como queramos. \square

Considerando el caso de los espacios vectoriales, es claro que la escritura de un módulo semisimple como suma directa de submódulos simples no es, en general, única. Tenemos, sin embargo, el siguiente resultado:

Proposición 1.18. *Sea M un A -módulo. Para cada A -módulo simple S , sea*

$$M_S = \sum_{f \in \text{hom}_A(S, M)} \text{im } f.$$

Entonces M_S es un A -submódulo de M y depende únicamente de la clase de isomorfismo de S , de manera que si c es la clase de isomorfismo de S , podemos escribir $M_c = M_S$. Además, si M es semisimple, existe un conjunto \mathcal{S}_M de clases de isomorfismo de A -módulos simples tal que $M_c \neq 0$ para todo $c \in \mathcal{S}_M$ y

$$M = \bigoplus_{c \in \mathcal{S}_M} M_c.$$

Demostración. Que para todo A -módulo simple M_S es un A -submódulo de M y que depende solamente de la clase de isomorfismo de S es claro. Supongamos entonces que M es semisimple y verifiquemos la última afirmación.

Sea $M = \bigoplus_{i \in I} S_i$ una descomposición de M como suma de A -módulos simples y sea \mathcal{S} el conjunto de las clases de isomorfismo de A -módulos simples.

Para cada $c \in \mathcal{S}$, consideremos un A -módulo simple S_c tal que $S_c \in c$ y pongamos $I_c = \{i \in I : S_i \cong S_c\}$. Es claro que si $\mathcal{S}_M = \{c \in \mathcal{S} : I_c \neq \emptyset\}$, obtenemos una partición $\{I_c : c \in \mathcal{S}_M\}$ de I . Además, si $M'_c = \bigoplus_{i \in I_c} S_i$ para cada $c \in \mathcal{S}_M$, es

$$M = \bigoplus_{c \in \mathcal{S}_M} M'_c.$$

Para terminar, mostraremos que $M_c = M'_c$ para cada $c \in \mathcal{S}_M$.

Sean $c, c' \in \mathcal{S}_M$, $f : S_c \rightarrow M$ un morfismo de A -módulos y supongamos que existe $m \in \text{im } f \cap \bigoplus_{i \in I_{c'}} S_i$ tal que $m \neq 0$. Entonces 1.7 implica que f es inyectivo, que $S_c \cong \text{im } f \subset \bigoplus_{i \in I_{c'}} S_i$ y entonces, como el módulo de la izquierda es semisimple, que existe $i \in I_{c'}$ tal que $S_c \cong S_i$. La elección de I_c implica entonces que $c = c'$. Esto nos dice que $M_c \subset M'_c$. Como la inclusión recíproca es evidente, esto termina la prueba de la proposición. \square

Si c es una clase de isomorfismo de A -módulos simples y M un A -módulo, el submódulo M_c de M construido en 1.18 es la *componente isotípica de M de tipo c* .

1.3. El radical de un módulo. Sea A un anillo. Recordemos que si M es un A -módulo a izquierda, entonces el *radical* $\text{rad } M$ de M es la intersección de sus submódulos maximales. Es claro que

$$(1.1) \quad \text{rad } M = \bigcap_{\substack{h: M \rightarrow S \\ S \text{ simple}}} \ker h.$$

Cuando M no posee submódulos maximales, es $\text{rad } M = M$.

Lema 1.19. *Sea A un anillo y M un A -módulo no nulo finitamente generado. Entonces $\text{rad } M$ es un submódulo propio de M .*

Demostración. Sea \mathcal{M} el conjunto de todos los submódulos propios de M y $\{x_1, \dots, x_n\}$ un subconjunto de M tal que $M = \sum_{i=1}^n Ax_i$. Claramente $\mathcal{M} \neq \emptyset$. Veamos que se trata de un conjunto inductivo.

Si $\{M_\lambda\}_{\lambda \in \Lambda} \subset \mathcal{M}$ es una cadena en \mathcal{M} , sea $N = \bigcup_{\lambda \in \Lambda} M_\lambda$. Como la unión es creciente, si fuese $N = M$, entonces existiría $\lambda \in \Lambda$ tal que $\{x_1, \dots, x_n\} \subset M_\lambda$, esto es, tal que $M = M_\lambda$, lo que es imposible. Luego $N \in \mathcal{M}$ y vemos que cada cadena en \mathcal{M} es acotada.

El lema de Zorn nos permite concluir, entonces, que existe un submódulo propio maximal $N \subsetneq M$. Como $\text{rad } M \subset N$, esto prueba el lema. \square

Si un módulo no es finitamente generado, puede coincidir con su radical. Por ejemplo, si $A = \mathbb{Z}$ y $M = \mathbb{Q}$, entonces $\text{rad } M = M$. En efecto, M no posee submódulos maximales: si $N \subset M$ es maximal, entonces M/N es un grupo abeliano simple y existe p primo tal que $M/N \cong \mathbb{Z}_p$. Pero entonces la proyección canónica es un morfismo no nulo $\mathbb{Q} \rightarrow \mathbb{Z}_p$, lo que es imposible.

Lema 1.20. *Sea A un anillo.*

- (a) *Si $f : M \rightarrow N$ un morfismo de A -módulos, entonces $f(\text{rad } M) \subset \text{rad } N$.*
- (b) *Si M es un A -módulo y $N \subset \text{rad } M$ es un submódulo, entonces $\text{rad}(M/N) = (\text{rad } M)/N$.*
- (c) *Si M es un A -módulo, $\text{rad}(M/\text{rad } M) = 0$.*

Demostración. (a) Sea $m \in \text{rad } M$ y sea $h : N \rightarrow S$ un morfismo de A -módulos con S simple. Entonces (1.1) implica que $h(f(m)) = 0$. Usando (1.1) otra vez, vemos que $f(m) \in \text{rad } N$.

(b) Si $\pi : M \rightarrow M/N$ es la proyección canónica, la primera parte nos dice que $(\text{rad } M)/N = \pi(\text{rad } M) \subset \text{rad}(M/N)$. Recíprocamente, supongamos que $m \in M \setminus \text{rad } M$. Entonces existe un morfismo $h : M \rightarrow S$ con S simple y $h(m) \neq 0$. Como $N \subset \text{rad } M \subset \ker h$, h induce un morfismo $\bar{h} : M/N \rightarrow S$ tal que $\bar{h} \circ \pi = h$. Es $\bar{h}(\pi(m)) \neq 0$ y vemos que $\pi(m) \notin \text{rad}(M/N)$.

(c) Esto sigue de tomar $N = \text{rad } M$ en (b). \square

La razón por la que estamos interesados en el radical es la siguiente caracterización de la semisimplicidad:

Proposición 1.21. *Sea A un anillo y M un A -módulo.*

- (a) *Si M es semisimple, entonces $\text{rad } M = 0$.*
- (b) *Las siguientes afirmaciones son equivalentes:*
 - (i) *M es artiniano y $\text{rad } M = 0$.*
 - (ii) *M es suma directa finita de submódulos simples.*

Demostración. Sea M un A -módulo semisimple y sea $M = \bigoplus_{i \in I} S_i$ una descomposición de M como suma directa de A -módulos simples. Si para cada $i \in I$ notamos $\pi_i : M \rightarrow S_i$ a la proyección canónica, entonces que $\text{rad } M \subset \bigcap_{i \in I} \ker \pi_i = 0$. Esto prueba (a). Veamos (b).

Probemos que (i) implica (ii): Sea \mathcal{M} el conjunto de los submódulos maximales de M y sea $\mathcal{I} = \{\bigcap_{N \in F} N : F \subset \mathcal{M} \text{ es finito}\}$. Como M es artiniano, \mathcal{I} posee un elemento minimal M_0 . Supongamos que $F \subset \mathcal{M}$ es una familia finita de submódulos maximales tal que $M_0 = \bigcap_{N \in F} N$. Si $m \in M_0 \setminus 0$, y como $m \notin \text{rad } M$, existe un submódulo $N \in \mathcal{M}$ tal que $m \notin N$. Pero entonces $M_0 \subsetneq M_0 \cap N \in \mathcal{I}$, lo que es imposible. Vemos así que debe ser $M_0 = 0$.

Esto implica que la aplicación $M \rightarrow \bigoplus_{N \in F} M/N$, que en cada componente es una proyección canónica, es inyectiva. Notemos que, como $F \subset \mathcal{M}$, el A -módulo

$\bigoplus_{N \in F} M/N$ es semisimple. Como M es isomorfo a un submódulo de esta suma directa, es él mismo semisimple. En vista de 1.17, M es isomorfo a una suma directa finita de submódulos simples.

Veamos finalmente que (ii) implica (i): Si $M = \bigoplus_{i \in I} S_i$ es una suma directa finita de submódulos simples, entonces $\text{rad } M = 0$ por la parte (a) de la proposición y es artiniiano en vista de 1.17. \square

2. ANILLOS SEMISIMPLES

2.1. Caracterización de anillos semisimples. Recordemos que un anillo A es *semisimple* si lo es como A -módulo a izquierda. Si bien esta definición puede parecer arbitraria, veremos en esta sección que no lo es, es decir que A es semisimple como A -módulo a izquierda sii lo es como A -módulo a derecha. Esto resulta del Teorema de Wedderburn.

Comenzaremos probando algunas propiedades de los módulos simples sobre anillos semisimples.

Lema 2.1. *Sea A un anillo semisimple. Si S es un A -módulo simple, entonces existe un ideal minimal $\mathfrak{a} \triangleleft_l A$ tal que $S \cong \mathfrak{a}$.*

Demostración. Si $m \in S \setminus 0$, entonces $S \cong A/\text{ann}(m)$ y $\text{ann}(m)$ es un ideal a izquierda maximal. Como A es semisimple, $\text{ann}(m)$ es un sumando directo de A y existe un ideal $\mathfrak{a} \triangleleft_l A$ tal que $A = \text{ann}(m) \oplus \mathfrak{a}$.

Observemos que \mathfrak{a} es un A -módulo semisimple. En efecto, supongamos que posee un submódulo propio no nulo $\mathfrak{b} \subset \mathfrak{a}$. Entonces \mathfrak{b} posee un complemento $\mathfrak{b}' \subset \mathfrak{a}$ tal que $\mathfrak{a} = \mathfrak{b} \oplus \mathfrak{b}'$ y $\text{ann}(m) \subsetneq \text{ann}(m) \oplus \mathfrak{b} \subsetneq A$. Esto contradice la maximalidad de $\text{ann}(m)$. Vemos así que \mathfrak{a} es un ideal a izquierda minimal.

Para terminar, notamos que $S \cong A/\text{ann}(m) = (\text{ann}(m) \oplus \mathfrak{a})/\text{ann}(m) \cong \mathfrak{a}$. \square

Proposición 2.2. *Sea A un anillo semisimple y sea \mathcal{S} el conjunto de las clases de isomorfismo de A -módulos simples. Para todo $c \in \mathcal{S}$, la componente isotópica $A_c \subset A$ correspondiente a c es un ideal bilátero no nulo.*

Demostración. Sea $c \in \mathcal{S}$ y sea S un A -módulo simple tal que $S \in c$. Recordemos de 1.18 que

$$A_c = \sum_{f \in \text{hom}_A(S, A)} \text{im } f.$$

Sea $x \in A_c$ y $b \in A$. Entonces existe $n \in \mathbb{N}$ y morfismos $f_1, \dots, f_n : S \rightarrow A$ tales que $x = \sum_{i=1}^n f_i(a)$, de donde $xb = \sum_{i=1}^n f_i(a)b = \sum_{i=1}^n (f_i b)(a) \in A_c$. Vemos así que A_c es un ideal bilátero.

Por otro lado, 2.1 implica que existe un ideal $\mathfrak{b} \triangleleft_l A$ tal que $S \cong \mathfrak{b}$. Claramente, esto nos dice que $\mathfrak{b} \subset A_c$ y, entonces, que $A_c \neq 0$. \square

Proposición 2.3. *Sean A y B dos anillos semisimples. Entonces $A \times B$ es semisimple.*

Demostración. Supongamos que $A = \sum_{i \in I} S_i$ y $B = \sum_{j \in J} T_j$ con S_i un A -módulo simple para cada $i \in I$ y T_j un B -módulo simple para cada $j \in J$, y consideremos las proyecciones canónicas $\pi_1 : A \times B \rightarrow A$ y $\pi_2 : A \times B \rightarrow B$. Es inmediato que

$$A \times B = \sum_{i \in I} \pi_1^*(S_i) + \sum_{j \in J} \pi_2^*(T_j),$$

de manera que $A \times B$ es semisimple porque $\pi_1^*(S_i)$ y $\pi_2^*(S_j)$ son simples cualesquiera sean $i \in I$ y $j \in J$, en vista de 1.5. \square

Un argumento inductivo a partir de 2.3 prueba, más generalmente, que un producto directo finito de anillos semisimples es semisimple.

Observación 2.4. Recordemos que si D es un anillo de división y $n \in \mathbb{N}$, entonces $M_n(D)$ es artiniano a izquierda y simple. La siguiente proposición nos dice que obtenemos de esta forma todos los anillos artinianos simples y que éstos son semisimples:

Proposición 2.5. *Un anillo A artiniano a izquierda y simple es semisimple y todos sus módulos simples son isomorfos. Además, si S es un A -módulo simple, entonces $D = \text{End}_A(S)^{\text{op}}$ es un anillo de división y existe $n \in \mathbb{N}$ tal que $A \cong M_n(D)$.*

Demostración. Sea A un anillo artiniano simple. Como es artiniano, existe un ideal minimal $\mathfrak{a} \triangleleft_l A$. Sea c la clase de isomorfismo de \mathfrak{a} y A_c la componente isotópica de A correspondiente a c . Como 2.2 nos dice que A_c es un ideal bilátero no nulo y estamos suponiendo que A es simple, debe ser $A = A_c$. En consecuencia, A_c es la única componente isotópica no nula y vemos que c es la única clase de isomorfismo de A -módulos simples.

Sea $X = \text{hom}_A(\mathfrak{a}, A)$ y consideremos el morfismo $\phi : \mathfrak{a}^{(X)} \rightarrow A$ tal que $\phi((a_f)_{f \in X}) = \sum_{f \in X} f(a_f)$. Es sobreyectivo: en efecto, $\text{im } \phi = A_c = A$. Como A es A -módulo proyectivo, concluimos que A es isomorfo a un sumando directo del módulo semisimple $\mathfrak{a}^{(X)}$. Usando 1.14, vemos entonces que existe $X' \subset X$ tal que $A \cong \mathfrak{a}^{(X')}$; en particular, A es semisimple.

Finalmente, como A es finitamente generado, debe ser X' finito y concluimos que existe $n \in \mathbb{N}$ tal que $A \cong \mathfrak{a}^n$. El lema de Schur 1.7 implica que $D = \text{End}_A(\mathfrak{a})$ es un anillo de división y entonces $A^{\text{op}} \cong \text{End}_A(A) \cong \text{End}_A(\mathfrak{a}^n) = M_n(D)$. Por supuesto, esto nos dice que $A \cong M_n(D)^{\text{op}} \cong M_n(D^{\text{op}})$. \square

Observación 2.6. Sea D un anillo de división, $n \in \mathbb{N}$ y sea $A = M_n(D)$. Sea $S = D^n$ el D -módulo de los vectores columna con coeficientes en D . Es claro que S es un A -módulo izquierdo con respecto a la multiplicación matricial. Es fácil ver, como en el caso de los espacios vectoriales, que se trata de un A -módulo simple. Obtenemos así un representante de la única clase de isomorfismo de A -módulos simples.

El siguiente teorema da una caracterización de los anillos semisimples en términos de sus módulos.

Teorema 2.7. *Sea A un anillo, son equivalentes:*

1. A es semisimple.
2. Todo A -módulo es semisimple.
3. Todo A -módulo libre es semisimple.
4. Todo A -módulo es proyectivo.
5. Toda extensión de A -módulos es trivial.
6. Todo A -módulo es inyectivo.
7. Todo ideal a izquierda de A es inyectivo.
8. Todo cociente de A es proyectivo.

Demostración. \blacksquare (1) implica (2). Todo A -módulo es suma de submódulos cíclicos, así que basta ver que todo A -módulo cíclico es semisimple. Si M es cíclico, $M \cong A/I$ para algún ideal a izquierda I . Como A es semisimple, el corolario 1.15 nos dice que M es semisimple.

\blacksquare (2) implica (3). Esto es inmediato.

\blacksquare (3) implica (4). Sea M un A -módulo. Existe un módulo libre L y un epimorfismo $p : L \rightarrow M$. El Teorema 1.13 nos dice que $\ker(p)$ es un sumando directo de L , así

que $L/\ker(p) \cong M$ también es isomorfo a un sumando directo de L . Esto muestra que M es proyectivo.

- (4) implica (5). Consideramos una extensión de A -módulos

$$0 \longrightarrow X \longrightarrow Y \longrightarrow Z \longrightarrow 0.$$

Como Z es proyectivo, esta sucesión se parte.

- (5) implica (6). Sea M un A -módulo. Como la hipótesis implica que toda sucesión exacta de la forma

$$0 \longrightarrow M \longrightarrow X \longrightarrow Y \longrightarrow 0$$

se parte, M es inyectivo.

- (6) implica (7). Esto es inmediato.
- (7) implica (8). Sea I un ideal. Por hipótesis, I es inyectivo, así que la sucesión exacta

$$0 \longrightarrow I \longrightarrow A \longrightarrow A/I \longrightarrow 0$$

se parte. Esto dice que A/I es isomorfo a un sumando directo de A . En particular, A/I es proyectivo.

- (8) implica (1). Por el Teorema 1.13 basta ver que todo ideal I de A es un sumando directo. Consideremos de nuevo la sucesión

$$0 \longrightarrow I \longrightarrow A \longrightarrow A/I \longrightarrow 0.$$

Como A/I es proyectivo, esta sucesión se parte. Luego I es un sumando directo. \square

2.2. Teorema de Wedderburn.

Teorema 2.8. (Wedderburn, 1908 [5]; Artin) *Sea A un anillo. Las siguientes afirmaciones son equivalentes:*

1. A es semisimple;
2. todo A -módulo es semisimple;
3. existen $r, n_1, \dots, n_r \in \mathbb{N}$ y anillos de división D_1, \dots, D_r tales que hay un isomorfismo de anillos $A \cong \mathbf{M}_{n_1}(D_1) \times \dots \times \mathbf{M}_{n_r}(D_r)$.

Demostración. Ya vimos que (1) y (2) son equivalentes.

Mostremos que (2) implica (3). La hipótesis nos dice que, en particular, ${}_A A$ es semisimple, así que 1.13 implica que $A \cong \bigoplus_{i \in I} S_i$ con S_i submódulo simple de A para cada $i \in I$. Como A es finitamente generado, 1.17 implica que I es finito. La afirmación (3) sigue entonces de los isomorfismos de anillos

$$A^{\text{op}} \cong \text{End}_A(A) \cong \text{End}_A\left(\bigoplus_{i \in I} S_i\right)$$

y de 1.8. Finalmente, la implicación restante es consecuencia inmediata de 2.3 y 2.4. \square

Podemos describir los parámetros que aparecen en la tercera afirmación del teorema de la siguiente manera:

Proposición 2.9. *Sea A un anillo semisimple. Entonces hay un número finito de clases de isomorfismo de A -módulos simples. Sea $\{S_1, \dots, S_r\}$ un conjunto de representantes dos a dos no isomorfos para estas clases de isomorfismo y, para cada $i \in \{1, \dots, r\}$, pongamos $D_i = \text{End}_A(S_i)$. Si $i \in \{1, \dots, r\}$, entonces $\text{hom}_A(S_i, A)$ es un D_i -módulo a derecha de dimensión $n_i = \dim_{D_i} \text{hom}_A(S_i, A)$ finita. Hay un isomorfismo de anillos $A \cong \mathbf{M}_{n_1}(D_1^{\text{op}}) \times \dots \times \mathbf{M}_{n_r}(D_r^{\text{op}})$.*

Demostración. Esto es consecuencia del teorema y del lema de Schur. \square

Todo lo que hemos hecho ha sido considerando módulos a izquierda, pero claramente podemos desarrollar una teoría simétrica con módulos a derecha. El siguiente corolario, sin embargo, justifica la asimetría de la definición 1.9:

Corolario 2.10. *Un anillo A es semisimple sii el A -módulo a derecha A es semisimple.*

Demostración. La tercera condición de 2.8 es evidentemente simétrica con respecto a la izquierda y la derecha. \square

Corolario 2.11. *Un anillo semisimple es artiniiano y notheriano.*

Demostración. Esto es consecuencia directa de 1.17. \square

Corolario 2.12. *Si A es un anillo semisimple, existen finitas clases de isomorfismo de A -mdulos simples.*

Demostración. En efecto, el nmero de clases de isomorfismo de A -mdulos simples es el nmero r que aparece en la tercera parte de 2.8. \square

Observacin 2.13. Sea k un anillo conmutativo y sea A una k -lgebra que es semisimple como anillo. Entonces los anillos de divisin que aparecen en la tercera afirmacin de 2.8 son k -lgebras y el isomorfismo all mencionado es un isomorfismo de k -lgebras.

Cuando A es un lgebra sobre un cuerpo k algebraicamente cerrado, podemos ser ms precisos en la tercera afirmacin de 2.8, ya que no hay k -lgebras de divisin de dimensin finita no triviales:

Lema 2.14. *Sea k un cuerpo algebraicamente cerrado. Si D es una k -lgebra de dimensin finita, entonces $D \cong k$.*

Demostracin. Sea D una k -lgebra de divisin y supongamos que existe $a \in D$ tal que el conjunto $\{1_D, a\}$ es linealmente independiente sobre k . Consideremos el morfismo de k -lgebras $f : p \in k[X] \mapsto p(a) \in D$. Como $\dim_k A < \infty$, es $\ker f \neq 0$ y existe $p \in k[X]$ mnico tal que $\ker f = (p)$. Ms an, como k es algebraicamente cerrado, existe $\lambda \in k$ y $q \in k[X]$ tal que $p = (X - \lambda)q$. Esto implica que $(a - \lambda 1_D)q(a) = 0$ en D : como $a \neq \lambda 1_D$, debe ser $q(a) = 0$, lo que contradice la eleccin de p , ya que $\deg q < \deg p$.

Vemos as que debe ser $\dim_k D = 1$, esto es, $D \cong k$. \square

Teniendo esto en cuenta, es claro que 2.8 implica la siguiente proposicin:

Proposicin 2.15. *Sea k un cuerpo algebraicamente cerrado y A una k -lgebra. Entonces A es semisimple sii existen $r, n_1, \dots, n_r \in \mathbb{N}$ tales que $A \cong M_{n_1}(k) \times \dots \times M_{n_r}(k)$.* \square

Recordemos, por otro lado, el siguiente teorema:

Teorema 2.16. (Wedderburn, 1905 [4]; Dickson, 1905 [2]) *Un anillo de divisin finito es un cuerpo.* \square

El teorema 2.8 implica, en vista de esta descripcin de los anillos de divisin finitos, la siguiente proposicin:

Proposicin 2.17. *Un anillo finito A es semisimple sii existen $r, n_1, \dots, n_r \in \mathbb{N}$ y cuerpos finitos k_1, \dots, k_r tales que $A \cong M_{n_1}(k_1) \times \dots \times M_{n_r}(k_r)$.* \square

Observación 2.18. Sean $r, n_1, \dots, n_r \in \mathbb{N}$ y D_1, \dots, D_r son anillos de división y consideremos el anillo $A = M_{n_1}(D_1) \times \dots \times M_{n_r}(D_r)$. Si $1 \leq i \leq r$, sea $\pi_i : A \rightarrow M_{n_i}(D_i)$ la proyección en el i -ésimo factor y sean, para cada i , $M_i = D_i^{n_i}$ el $M_{n_i}(D_i)$ -módulo simple construido en 2.6 y $S_i = \pi_i^*(M_i)$. Entonces $\{S_i : 1 \leq i \leq r\}$ es un conjunto completo de representantes de las clases de isomorfismo de A -módulos simples.

Notemos que si k es un cuerpo y A es una k -álgebra, entonces es claro que $\dim_k S_i = n_i \dim_k D_i$ para cada $i \in \{1, \dots, r\}$.

2.3. El radical de un anillo. El *radical de Jacobson* de un anillo A (o, simplemente, el *radical*) es el ideal a izquierda $J(A) = \text{rad } A$.

La segunda parte de la proposición 1.21 tiene como consecuencia inmediata el siguiente teorema:

Teorema 2.19. *Sea A un anillo artiniano. Entonces A es semisimple sii $J(A) = 0$. \square*

La siguiente proposición lista alguna de las propiedades más importantes del radical de un anillo:

Proposición 2.20. *Sea A un anillo y $J(A)$ su radical de Jacobson.*

- (a) $J(A)$ es un ideal bilátero.
- (b) $a \in J(A)$ sii para todo $x \in A$, $1 - xa$ es inversible a izquierda.
- (c) $J(A)$ es el único elemento maximal de $\{I \triangleleft A : \forall x \in I, 1 - x \in A^\times\}$.
- (d) $J(A) = \text{rad } {}_A A = \text{rad } A_A$.

Demostración. Si $b \in A$ y $f : a \in A \mapsto ab \in A$, entonces $f \in \text{hom}_A(A, A)$ y 1.20 implica que $f(J(A)) \subset J(A)$. Esto dice, precisamente, que el ideal a izquierda $J(A)$ también es un ideal a derecha y prueba (a).

Supongamos que $a, x \in A$ son tales que $1 - xa$ no tiene inverso a izquierda. Entonces el ideal $A(1 - xa)$ es propio y existe un ideal a izquierda maximal M tal que $1 - xa \in M$. Como $J(A) \subset M$, esto implica que $a \notin J(A)$: en efecto, si $a \in J(A)$, sería $1 = (1 - xa) + xa \in M + J(A) \subset M$. Esto muestra la necesidad de la condición en (b).

Para ver la suficiencia, consideremos $a \in A$ tal que $1 - xa$ es inversible a izquierdo para todo $x \in A$. Supongamos que $a \notin J(A)$, de manera que existe un ideal maximal $M \triangleleft_l A$ tal que $a \notin M$. Pero entonces $1 \in A = M + Aa$ y vemos que existe $x \in A$ tal que $y = 1 - xa \in M$. Esto es absurdo, porque por hipótesis y es inversible a izquierda.

Sea $\mathcal{I} = \{I \triangleleft A : \forall x \in I, 1 - x \in A^\times\}$. Para ver (c) tenemos que mostrar que $J(A) \in \mathcal{I}$ y que $J(A) \subset I$ para todo $I \in \mathcal{I}$.

Veamos que $J(A) \in \mathcal{I}$. Ya sabemos que $J(A) \triangleleft A$. Sea $x \in J(A)$. La parte (b) implica que $1 - x$ es inversible a izquierda, de manera que existe $z \in A$ tal que $z(1 - x) = 1$. Como $1 - z = -zx \in J(A)$, otra vez (b) nos dice que $z = 1 - (1 - z)$ es inversible a izquierda, esto es, que existe $w \in A$ tal que $wz = 1$. Como $wz = 1 = z(1 - x)$, debe ser $w = 1 - x$ y entonces $z = (1 - x)^{-1}$. Concluimos que $J(A) \in \mathcal{I}$, como queríamos.

Sea ahora $I \in \mathcal{I}$ y sea $a \in I$. Si $x \in A$, $ax \in I$ así que por hipótesis $1 - ax \in A^\times$. Usando (b) vemos que $a \in J(A)$. Así, $I \subset J(A)$.

Para terminar, notemos que (d) sigue inmediatamente del hecho de que la afirmación (c) es simétrica con respecto a la izquierda y la derecha. \square

Sabiendo que el radical es un ideal bilátero, el siguiente enunciado tiene sentido:

Proposición 2.21. *Si A es un anillo artiniano, entonces $A/J(A)$ es anillo semisimple.*

Demostración. La tercera parte de 1.20 implica que $\text{rad } {}_A(A/J(A)) = 0$, de manera que 1.21 nos permite concluir que $A/J(A)$ es semisimple como A -módulo.

Ahora bien, un subgrupo abeliano de $A/J(A)$ es un A -submódulo sii es un $A/J(A)$ -submódulo. Esto nos dice que $A/J(A)$ es semisimple también como $A/J(A)$ -módulo, esto es, que $A/J(A)$ es semisimple como anillo. \square

De hecho, dado un anillo A , el radical $J(A)$ es el menor ideal de A tal que $A/J(A)$ es semisimple. En efecto, si $I \triangleleft A$ es un ideal tal que A/I es semisimple, entonces $0 = \text{rad}_{A/I} A/I = \text{rad}_A A/I$, de manera que $J(A) = \text{rad}_A A \subset I$, como consecuencia de la segunda parte de 1.20.

Esta proposición tiene la siguiente consecuencia extremadamente útil:

Proposición 2.22. *Sea A un anillo artiniiano y M un A -módulo. Entonces $\text{rad } M = J(A)M$.*

Demostración. Pongamos $J = J(A)$. Si $m \in M$ y $f : a \in A \mapsto am \in M$, entonces 1.20(a) nos dice que $Jm = f(\text{rad } A) \subset \text{rad } M$. Como esto es cierto para todo $m \in M$, $JM \subset \text{rad } M$. Usando ahora 1.20(b), vemos que $(\text{rad } M)/JM = \text{rad}(M/JM)$. Pero M/JM es un A/J -módulo y A/J es un anillo semisimple, de modo que $\text{rad}_{A/J} M/JM = 0$ en vista de 1.21(a). Como un A -submódulo de M/JM es lo mismo que un A/J -submódulo, esto implica que $(\text{rad } M)/JM = \text{rad}_A M/JM = 0$ y, en definitiva, que $\text{rad } M = JM$. \square

Observación 2.23. En algunos textos, aparece la siguiente definición de anillo simple: *un anillo A se dice simple si es artiniiano y no tiene ideales biláteros propios.* Notamos que la condición de artiniiano es esencial si se desea que la definición de simple implique semisimple, como lo muestra el siguiente ejemplo:

Sea k un cuerpo de característica cero. El *álgebra de Weyl* $A_1(k)$ es la subálgebra de $\text{End}_k(k[X])$ generada por los endomorfismos $p, q \in \text{End}_k(k[X])$ tales que

$$q(f) = Xf \quad \text{y} \quad p(f) = \frac{d}{dX}f$$

para todo $f \in k[X]$.

Es fácil ver que $[p, q] = 1$ (verificarlo!). Usando esto, se puede ver que $\{p^i q^j : i, j \in \mathbb{N}_0\}$ es una base de $A_1(k)$ como k -espacio vectorial. Si $P \in A_1(k)$ se escribe de la forma $P = \sum_{i=0}^n f_i(q)p^i$, en donde cada f_i es un polinomio en q y $f_n \neq 0$, diremos que el *grado* de P es n y que el polinomio $f_n \in k[q]$ es el coeficiente principal de P .

Dejamos como ejercicio verificar que:

- Si $P \in A_1(k)$ es un elemento de grado n con coeficiente principal f_n , entonces $[P, q]$ es un elemento de grado $n - 1$ y su coeficiente principal es nf_n .
- Si $f \in k[q]$, entonces $[p, f] = f'$.

Usando estas dos afirmaciones, es fácil ver que $A_1(k)$ no tiene ideales biláteros propios. En efecto, supongamos que I es un ideal bilátero no nulo de $A_1(k)$.

Sea $P_0 \in I$ un elemento no nulo arbitrario y sea g el grado de P_0 . Definamos inductivamente $P_{i+1} = [P_i, q]$ para cada $i \in \mathbb{N}_0$. Entonces es fácil ver, usando la primera afirmación, que P_g es un elemento no nulo de $I \cap k[q]$. Escribamos $Q_0 = P_g$ y sea d el grado de Q_0 en $k[q]$. Definamos $Q_{i+1} = [p, Q_i]$ para cada $i \in \mathbb{N}_0$. Usando ahora la segunda afirmación, vemos que Q_d es un elemento no nulo de $I \cap k$. Como es inversible, vemos que $I = A_1(k)$.

Un elemento x de un anillo A es *nilpotente* si existe $n \in \mathbb{N}$ tal que $x^n = 0$. Un ideal $\mathfrak{a} \triangleleft A$ es *nil* si todos sus elementos son nilpotentes. Finalmente, un ideal $\mathfrak{a} \triangleleft A$ es *nilpotente* si existe $n \in \mathbb{N}$ tal que $\mathfrak{a}^n = 0$.

Proposición 2.24. *Sea A un anillo.*

(a) *Todo ideal nil está contenido en $J(A)$.*

Supongamos ahora que A es artiniiano a izquierda.

(b) *$J(A)$ es el ideal bilátero nilpotente más grande.*

(c) *$J(A)$ es el único ideal nil \mathfrak{a} de A tal que A/\mathfrak{a} es semisimple.*

Demostración. (a) Sea $\mathfrak{a} \triangleleft A$ un ideal nil. Si $x \in A$ y $n \in \mathbb{N}$ es tal que $x^n = 0$, entonces $(1-x) \sum_{i=0}^{n-1} x^i = 1$, de manera que $1-x$ es inversible. Usando la tercera parte de 2.20, vemos que $x \in J(A)$.

(b) Es claro que todo ideal nilpotente es nil, así que (a) implica que $J(A)$ contiene a todo ideal nilpotente. La afirmación (b) quedará probada, entonces, si mostramos que $J = J(A)$ es nilpotente.

La cadena de ideales $J \supset J^2 \supset J^3 \supset \dots$ debe estabilizarse, así que existe $n \in \mathbb{N}$ tal que $J^n = J \cdot J^n$. Supongamos que $J^n \neq 0$. Entonces el conjunto de ideales a izquierda $\mathcal{I} = \{\mathfrak{a} \triangleleft_l A : \mathfrak{a} = J\mathfrak{a}\}$ es no vacío. Como A es artiniiano, existe un elemento $\mathfrak{a} \in \mathcal{I}$ minimal. Además, como $\mathfrak{a} = J\mathfrak{a} = J^2\mathfrak{a} = \dots = J^n\mathfrak{a}$, existe $x \in \mathfrak{a}$ tal que $J^n x \neq 0$ y, entonces, $J \cdot J^n x = J^n x$. Así, $J^n x \in \mathcal{I}$. Como $J^n x \subset \mathfrak{a}$, la elección de \mathfrak{a} implica que $\mathfrak{a} = J^n x$ y vemos que \mathfrak{a} es finitamente generado. Pero entonces

$$\begin{aligned} \mathfrak{a} &\supsetneq \text{rad } \mathfrak{a} && \text{por 1.19} \\ &= J\mathfrak{a} && \text{por 2.22} \\ &= \mathfrak{a} \end{aligned}$$

Esto es imposible y debe ser, en consecuencia, $J^n = 0$.

Finalmente, sea $\mathfrak{a} \triangleleft A$ un ideal nil tal que A/\mathfrak{a} es semisimple. Por la parte (a), $\mathfrak{a} \subset J(A)$; por otro lado, como el ideal $J(A)/\mathfrak{a} \triangleleft A/\mathfrak{a}$ es nilpotente, la parte (b) nos dice que $J(A)/\mathfrak{a} \subset \text{rad}(A/\mathfrak{a})$. Como estamos suponiendo que A/\mathfrak{a} es semisimple, $\text{rad}(A/\mathfrak{a}) = 0$ y vemos que $J(A) = \mathfrak{a}$. Esto prueba (c). \square

La última parte de esta proposición puede ser usada frecuentemente para identificar el radical de un álgebra.

Por ejemplo, sea $Q = (Q_0, Q_1)$ un quiver finito, k un cuerpo y kQ la k -álgebra de caminos sobre Q . Sea $R \triangleleft kQ$ el ideal generado por los caminos de longitud 1 y sea $I \triangleleft kQ$ un ideal admisible, esto es, sea I un ideal tal que (i) $I \subset R^2$ y (ii) existe $n \in \mathbb{N}$ tal que $I \supset R^n$. Consideremos el álgebra $A = kQ/I$. Afirmamos que $\text{rad } A = R/I$.

Notemos que A es artiniiana, ya que $\dim_k A < \infty$ debido a la condición (ii). Ahora bien, el cociente $A/R = (kQ/I)/(R/I) \cong kQ/R \cong k^{|Q_0|}$ es isomorfo a un producto de $|Q_0|$ copias de k , así que es semisimple, y, por otro lado, todo elemento de R/I es nilpotente en A en vista de la condición (ii). Luego 2.24(c) implica que R/I es el radical de A .

Proposición 2.25. *Sea A un anillo y sea $J(A)$ su radical de Jacobson. Si S es un A -módulo simple, entonces $J(A)S = 0$ y S es, de manera natural, un $A/J(A)$ -módulo simple. Recíprocamente, si S es un $A/J(A)$ -módulo simple, entonces, vía restricción de escalares a lo largo de la proyección canónica $A \rightarrow A/J(A)$, S es un A -módulo simple.*

Demostración. Claramente alcanza con probar que si S es un A -módulo simple, entonces $J(A)S = 0$.

Sea $s \in S$. La aplicación $f : a \in A \mapsto as \in S$ es un morfismo de A -módulos no nulo, así que $J(A) \subset \ker s$. Esto significa que $J(A)s = 0$. En consecuencia $J(A)S = 0$, como queríamos ver. \square

Corolario 2.26. *Sea A un anillo. Hay un biyección entre las clases de isomorfismo de A -módulos simples y las clases de isomorfismo de $A/J(A)$ -módulos simples.*

En particular, si A es artiniiano, hay un número finito de clases de isomorfismo de A -módulos simples.

Demostración. La primera afirmación es consecuencia inmediata de 2.25. Para ver la segunda, basta observar que si A es artiniiano, 2.21 nos dice que $A/J(A)$ es semisimple y entonces 2.12 junto con la primera parte implican la finitud del conjunto de clases de isomorfismo de A -módulos simples. \square

2.4. Algebras de grupo. Veremos ahora un familia importante de ejemplos de álgebra semisimple.

Fijemos un grupo G y un cuerpo k . Notamos $\text{cl}(G)$ al conjunto de las clases de conjugación de G .

Recordemos que la k -álgebra de grupo kG es la k -álgebra que, como k -módulo, es el k -módulo libre con base G y en la que el producto es el único producto k -bilineal y asociativo que extiende al de G .

Una *representación de G (sobre k)* es un par (M, ρ) formado por un k -espacio vectorial M y un homomorfismo de grupos $\rho : G \rightarrow \text{GL}(M)$. En general, escribimos M en lugar de (M, ρ) , cuando esto no dé lugar a confusiones.

Si (M, ρ) y (M', ρ') son dos representaciones de G , un *morfismo de representaciones de G* $f : (M, \rho) \rightarrow (M', \rho')$ es un morfismo $f : M \rightarrow M'$ de k -espacios vectoriales tal que para todo $g \in G$ se tiene que $\rho'(g) \circ f = f \circ \rho(g)$.

Observación 2.27. Sea M un kG -módulo. Sobre el k -espacio vectorial M podemos construir una representación (M, ρ) de G definiendo $\rho : G \rightarrow \text{GL}(M)$ de manera que

$$\rho(g)(m) = gm, \quad \forall g \in G, m \in M.$$

Si $f : M \rightarrow M'$ es un morfismo de kG -módulos y (M, ρ) y (M', ρ') son las representaciones de G correspondientes, es claro que $f : (M, \rho) \rightarrow (M', \rho')$ es un morfismo de representaciones.

Recíprocamente, si (M, ρ) es una representación de G , podemos hacer de M un kG -módulo si definimos la acción $kG \times M \rightarrow M$ poniendo

$$x \cdot m = \sum_{g \in G} a_g \rho(g)(m), \quad \forall x = \sum_{g \in G} a_g g \in kG, m \in M.$$

Como antes, si $f : (M, \rho) \rightarrow (M', \rho')$ es un morfismo de representaciones de G , entonces la aplicación k -lineal $f : M \rightarrow M'$ es de hecho kG -lineal.

Vemos así que las nociones de kG -módulo y de representación de G son equivalentes.

La *representación trivial* de G es la representación (k, ρ) con $\rho : G \rightarrow \text{GL}(k)$ el homomorfismo trivial. El kG -módulo *trivial* k es el kG -módulo correspondiente a (k, ρ) .

Observación 2.28. Sea M un kG -módulo de dimensión 1 y sea $m \in M \setminus 0$. Si $g \in G$, entonces g es una unidad de kG y existe $\rho_M(g) \in k^\times$ tal que $gm = \rho_M(g)m$. Obtenemos así un morfismo de grupos $\rho_M : G \rightarrow k^\times$; de hecho, si identificamos a k^\times con $\text{GL}(M)$, (M, ρ_M) es la representación de G correspondiente al kG -módulo M . Como k^\times es un grupo abeliano, el subgrupo derivado G' de G está contenido en el núcleo de ρ_M , y ρ_M induce entonces un morfismo de grupos $\bar{\rho}_M : G/G' \rightarrow k^\times$.

El morfismo $\bar{\rho}_M$ no depende de la elección del elemento $m \in M \setminus 0$. Más aún, si M' es un kG -módulo isomorfo a M , es fácil verificar que $\bar{\rho}_{M'} = \bar{\rho}_M$. Así, $\bar{\rho}_M$ depende solamente de la clase de isomorfismo $[M]$ de M .

El resultado más importante sobre álgebras de grupos es el siguiente:

Teorema 2.29. (Teorema de Maschke) *Sea G un grupo finito y sea k un cuerpo en el que $|G|$ es inversible. Entonces $k[G]$ es un anillo semisimple.*

Demostración. Sea M un $k[G]$ -módulo y $S \subseteq M$ un submódulo. Mostremos que S es un sumando directo. Para eso, alcanza con mostrar que existe un morfismo de $k[G]$ -módulos $\phi : M \rightarrow S$ tal que $\phi|_S = \text{Id}_S$.

Como k es un cuerpo, existe ciertamente una transformación k -lineal $\pi : M \rightarrow S$ tal que $\pi|_S = \text{Id}_S$. Definamos $\phi : M \rightarrow S$ poniendo

$$\phi(m) = \frac{1}{|G|} \sum_{g \in G} g\pi(g^{-1}m)$$

para todo $m \in M$. Afirmamos que ϕ es $k[G]$ -lineal y que $\phi|_S = \text{Id}_S$.

Si $s \in S$, entonces $g^{-1}s \in S$ y $\pi(g^{-1}s) = g^{-1}s$, así que

$$\begin{aligned} \phi(s) &= \frac{1}{|G|} \sum_{g \in G} g\pi(g^{-1}s) = \frac{1}{|G|} \sum_{g \in G} gg^{-1}s \\ &= \frac{1}{|G|} \sum_{g \in G} s = \frac{|G|}{|G|}s = s. \end{aligned}$$

Esto dice que $\phi|_S = \text{Id}_S$. Por otro lado, si $h \in G$ y $m \in M$, es

$$\phi(hm) = \frac{1}{|G|} \sum_{g \in G} g\pi(g^{-1}hm);$$

Si ponemos $g' = hg$ en la suma, esto queda

$$\begin{aligned} &= \frac{1}{|G|} \sum_{g \in G} g\pi(g^{-1}hm) = \frac{1}{|G|} \sum_{g' \in G} g'\pi((g')^{-1}hm) \\ &= \frac{1}{|G|} \sum_{g' \in G} hg\pi(g^{-1}h^{-1}hm) = \frac{1}{|G|} h \left(\sum_{g' \in G} g\pi(g^{-1}m) \right) \\ &= h\phi(m). \end{aligned}$$

Vemos así que ϕ es un morfismo de $k[G]$ -módulos. □

Observación 2.30. Supongamos desde ahora que k es un cuerpo en el que $|G|$ es inversible. Sea $\{S_1, \dots, S_r\}$ un conjunto completo de representantes de las clases de isomorfismo de kG -módulos simples y para cada $i \in \{1, \dots, r\}$, sea $D_i = \text{End}_{kG}(S_i)$ y sea $n_i = \dim_{D_i} \text{hom}_{kG}(S_i, kG)$. Entonces, como en la sección anterior, hay un isomorfismo de anillos

$$(2.1) \quad kG \cong M_{n_1}(D_1) \times \cdots \times M_{n_r}(D_r).$$

Observemos que D_i es una k -álgebra de división de dimensión finita para cualquier $i \in \{1, \dots, r\}$.

Proposición 2.31. *Es $\sum_{i=1}^r n_i^2 \dim_k D_i = |G|$.*

Demostración. El resultado sigue de tomar dimensión sobre k en (2.1). □

Proposición 2.32. *Sea $i \in \{1, \dots, r\}$.*

- Si $\dim_k S_i = 1$, entonces $n_i = \dim_k D_i = 1$.
- Si k es algebraicamente cerrado y $n_i = 1$, entonces $\dim_k S_i = 1$.

Demostración. Las dos afirmaciones siguen de la igualdad $n_i \dim_k D_i = \dim_k S_i$ observada en 2.18 y del lema 2.14. \square

Recordemos que si R es un anillo -cuyo centro denotaremos $Z(R)$, dado $n \in \mathbb{N}$, entonces hay un isomorfismo $Z(M_n(R)) \cong Z(R)$; en efecto, la aplicación $r \in Z(R) \mapsto rl \in M_n(R)$ es un morfismo de anillos inyectivo que tiene a $Z(M_n(R))$ como imagen. Por otro lado, si R y S son anillos, entonces hay un isomorfismo evidente $Z(R \times S) \cong Z(R) \times Z(S)$.

Proposición 2.33. *Es $r \leq |\text{cl}(G)|$. Si k es algebraicamente cerrado, entonces vale la igualdad.*

Demostración. Aplicando a (2.1) las observaciones que preceden al enunciado, vemos que hay un isomorfismo

$$(2.2) \quad Z(kG) \cong Z(D_1) \times \cdots \times Z(D_r).$$

Para cada $c \in \text{cl}(G)$ pongamos $z_c = \sum_{g \in c} g \in kG$. Es fácil ver que el conjunto $\{z_c : c \in \text{cl}(G)\}$ es una base del k -espacio vectorial $Z(kG)$ así que, en particular, $\dim_k Z(kG) = |\text{cl}(G)|$. Por otro lado, cualquiera sea $i \in \{1, \dots, r\}$, es claro que $k1_{D_i} \subset Z(D_i)$, de manera que $\dim_k Z(D_i) \geq 1$. Teniendo esto en cuenta, el isomorfismo (2.2) da inmediatamente la desigualdad del enunciado.

Si k es algebraicamente cerrado, entonces $D_i \cong k$ para todo $i \in \{1, \dots, r\}$, así que en este caso es $\dim_k Z(D_i) = 1$. Vale entonces que $r = |\text{cl}(G)|$ en este caso. \square

Si k es algebraicamente cerrado, podemos explicitar el isomorfismo (2.1).

Proposición 2.34. *Sea k un cuerpo algebraicamente cerrado. Entonces si $\{S_i\}_{i=1}^r$ es un conjunto completo de representantes de clases de isomorfismo de kQ -módulos simples, hay un isomorfismo $\phi : kG \rightarrow \prod_{i=1}^r \text{End}_k(S_i)$ tal que, si para cada $i \in \{1, \dots, r\}$, $\pi_r : \prod_{i=1}^r \text{End}_k(S_i) \rightarrow \text{End}_k(S_i)$ es la proyección k -ésima, entonces*

$$\pi_i(\phi(g)) = \rho_{S_i}(g)$$

para cada $i \in \{1, \dots, r\}$ y $g \in G$.

2.5. Ejercicios.

1. Descomponer a $\mathbb{R}[\mathbb{Z}_2]$, $\mathbb{R}[\mathbb{Z}_3]$ y $\mathbb{C}[\mathbb{Z}_3]$ como producto de anillos de matrices sobre álgebras de división, como en el Teorema de Wedderburn. Sugerencia: encontrar módulos simples sobre los respectivos anillos. Antes de hacer cuentas, sabiendo que las únicas álgebras de dimensión finita sobre \mathbb{R} son \mathbb{R} , \mathbb{C} y \mathbb{H} . Cuales son las posibilidades?
2. Probar que si A es anillo semisimple y L es un ideal a izquierda de A entonces:
 - existe $e \in A$ idempotente tal que $L = Ae$.
 - A no tiene ideales a izquierda nilpotentes.
 - Si L es simple entonces el idempotente es primitivo, esto es, si $e = e_1 + e_2$ con $e_i^2 = e_i$ y $e_1 e_2 = 0 = e_2 e_1$, entonces alguno de los e_i es cero.
3. Sea k un cuerpo y $T_2(k)$ el conjunto de matrices triangulares superiores de 2×2 a coeficientes en k , que no es un anillo semisimple. Calcular $\text{rad}(T_2(k))$ y $T_2(k)/\text{rad}(T_2(k))$.
4. Sea k un cuerpo y $A = k \times k$ con el producto coordenada a coordenada. Mostrar que A es semisimple pero no simple. Quiénes son los idempotentes ortogonales que suman uno?
5. Para que $n \in \mathbb{N}$ es \mathbb{Z}_n un anillo semisimple? Para alguno que no sea semisimple, dar un ejemplo de módulo que no sea proyectivo.

6. Sea A un anillo semisimple y M un A -módulo. A partir del teorema de Wedderburn sabemos que $A \cong \prod_{i=1}^n M_{r_i}(D_i)$ donde cada $D_i = \text{End}_A(L_i)^{\text{op}}$ es el anillo de endomorfismos del ideal simple L_i , y r_i es la cantidad de veces que aparece L_i en A como sumando directo. A su vez, M se descompone en suma directa de submódulos simples, cada uno de ellos isomorfo a algún L_i (por qué?). Dar una condición necesaria y suficiente sobre la multiplicidad de cada L_i en M para decidir cuándo M es libre. Concluir que si A es semisimple, entonces A tiene noción de rango.
7. Sea $A = M_n(k)$ con k un cuerpo, ver que es un ejemplo de anillo con noción de rango pero que no existe ningún morfismo de anillos $A \rightarrow D$ con D un anillo de división.
8. Sea A un anillo y sea M un A -módulo simple. Entonces o bien M , considerado como grupo abeliano, es isomorfo a una suma directa de copias de \mathbb{Q} , o bien existe $p \in \mathbb{N}$ primo tal que M es, considerado como grupo abeliano, isomorfo a una suma directa de copias de \mathbb{Z}_p .
9. Sea A un anillo conmutativo y M y N dos A -módulos a izquierda y a derecha respectivamente. Probar que si M o N es semisimple, entonces $M \otimes_A N$ es semisimple.
10.
 - Si A es un anillo semisimple y $B \subset A$ es un subanillo, es B necesariamente semisimple?
 - Si A es un anillo semisimple e $I \triangleleft A$ es un ideal bilátero, probar que A/I es semisimple.
11.
 - Sean A y B anillos y $n, m \in \mathbb{N}$. Entonces $M_m(M_n(A)) \cong M_{mn}(A)$ y $M_n(A \times B) \cong M_n(A) \times M_n(B)$.
 - Si A es un anillo semisimple y $n \in \mathbb{N}$, entonces $M_n(A)$ es semisimple.
 - Sea A un anillo y sea $n \in \mathbb{N}$. Sea P el conjunto de vectores *fila* de n componentes en A y sea Q el conjunto de vectores *columna* de n componentes en A . Entonces P es un A - $M_n(A)$ -bimódulo y Q es un $M_n(A)$ - A -bimódulo con acciones de $M_n(A)$ inducidas por el producto matricial. Más aún, hay un isomorfismo $Q \otimes_A P \cong M_n(A)$ de $M_n(A)$ -bimódulos y un isomorfismo $P \otimes_{M_n(A)} Q \cong A$ de A -bimódulos.
Como consecuencia de esto, si M es un A -módulo a izquierda, entonces
$$P \otimes_{M_n(A)} (Q \otimes_A M) \cong M.$$
 - Si M es un A - B -bimódulo y N es un B -módulo izquierdo proyectivo, entonces $M \otimes_B N$ es un A -módulo proyectivo.
 - Sea A un anillo. Si existe $n \in \mathbb{N}$ tal que $M_n(A)$ es semisimple, entonces el anillo A mismo es semisimple.
12. Sea A un anillo, M un A -módulo finitamente generado. Si $B = \text{End}_A(M)$ y A es semisimple, entonces B es semisimple. Notemos que esto tiene como caso particular a la segunda parte del ejercicio 11, ya que si $M = A^n$, entonces $\text{End}_n(M) \cong M_n(A)$.
13.
 - Un anillo artiniiano a izquierda sin divisores de cero es un anillo de división.
 - Si A es un anillo sin divisores de cero tal que $M_n(A)$ es semisimple para algún $n \in \mathbb{N}$, entonces A es un anillo de división.

Si $n \in \mathbb{N}$, sea G_n un grupo cíclico de orden n y sea $g_n \in G_n$ un generador.

14. Sea k un cuerpo de característica cero. Si $kG_n \cong M_{n_1}(D_1) \times \cdots \times M_{n_r}(D_r)$ es la factorización de kG_n como k -álgebra dada por el teorema de Wedderburn, de

manera que $r \in \mathbb{N}$, $n_1, \dots, n_r \in \mathbb{N}$ y D_1, \dots, D_r son k -álgebras de división, entonces $n_1 = n_2 = \dots = n_r = 1$ y D_i es un cuerpo para cada $i \in \{1, \dots, r\}$.

En particular, hay exactamente r isoclases de kG_n -módulos simples y si S_1, \dots, S_n son representantes de estas clases, hay un isomorfismo de kG_n -módulos $kG_n \cong \bigoplus_{i=1}^r S_i$.

15. Sea k un cuerpo de característica cero. Sea M un kG_n -módulo simple y sea $a : m \in M \mapsto g_n m \in M$ la multiplicación por g_n . Entonces $a \in \text{End}_{kG_n}(M)$ porque kG_n es un anillo conmutativo. Sea $\mu \in k[X]$ el polinomio minimal de a sobre k . Muestre que μ es irreducible en $k[X]$. Además, si $k = \mathbb{Q}$, entonces μ tiene coeficientes enteros.

16. Sea $\Omega_n \subset \mathbb{C}^\times$ el subgrupo multiplicativo de \mathbb{C}^\times de las raíces n -ésimas de la unidad.

- La aplicación $\phi : \chi \in \text{hom}_{\text{Grp}}(G_n, \Omega_n) \mapsto \chi(g_1) \in \Omega_n$ es un isomorfismo de grupos abelianos. Esto implica que el conjunto $\hat{G}_n = \text{hom}_{\text{Grp}}(G_n, \Omega_n)$ tiene exactamente n elementos; llamémoslos χ_1, \dots, χ_n .
- Muestre que si $\chi, \rho \in \hat{G}_n$, entonces

$$\sum_{g \in G_n} \chi(g) \rho(g^{-1}) = \delta_{\chi, \rho}.$$

Sugerencia: Multiplique el miembro izquierdo de esta igualdad por $(1 - \chi(g_1) \rho(g_1^{-1}))$.

- Si $\chi \in \hat{G}_n$, sea $e_\chi = \frac{1}{n} \sum_{g \in G_n} \chi(g^{-1}) g \in \mathbb{C}G_n$. Entonces si $\chi, \rho \in \hat{G}_n$,

$$e_\chi^2 = e_\chi, \quad e_\chi e_\rho = 1, \quad \text{cuando } \chi \neq \rho, \quad \text{y} \quad \sum_{\chi \in \hat{G}_n} e_\chi = 1.$$

- Consideremos el anillo $A = \mathbb{C} \times \dots \times \mathbb{C}$ con n factores y sean $x_1, \dots, x_n \in A$ los elementos de la base canónica. Hay un isomorfismo de anillos $\phi : \mathbb{C}G_n \rightarrow A$ tal que $\phi(e_{\chi_i}) = x_i$ si $1 \leq i \leq n$. Describa representantes para cada isoclase de $\mathbb{C}G_n$ -módulos simples.
17. ▪ Sea p un número primo. Si $0 \leq k < l$, sea $\phi_{k,l} : \mathbb{Q}G_{p^l} \rightarrow \mathbb{Q}G_{p^k}$ el único morfismo de anillos tal que $\phi_{k,l}(g_{p^l}) = g_{p^k}$. Entonces $\ker \phi_{k,l} = \langle g_{p^l}^{p^k} - 1 \rangle$. Además, si $0 \leq r < k < l$, es $\phi_{r,l} = \phi_{r,k} \circ \phi_{k,l}$.
- Sea p un número primo y pongamos $\Phi_p = \sum_{i=0}^{p-1} X^i \in \mathbb{Z}[X]$. Entonces

$$X^{p^l} - 1 = (X - 1) \prod_{i=0}^{l-1} \Phi_p(X^{p^i})$$

y cada uno de los factores $\Phi_p(X^{p^i})$ con $0 \leq i < l$ es irreducible en $\mathbb{Q}[X]$.

- Sea p un número primo impar. Sea $l \geq 1$ y sea M un $\mathbb{Q}G_{p^l}$ -módulo simple. Si $\dim_{\mathbb{Q}} M < p^l - p^{l-1}$, entonces existe $k < l$ y un $\mathbb{Q}G_{p^k}$ -módulo simple N tal que $M \cong \phi_{k,l}^*(N)$.
- Sea p un número primo impar. Notemos M_0 al único $\mathbb{Q}G_1$ -módulo simple. Entonces, para todo $l \geq 1$ existe, a menos de isomorfismo, un único $\mathbb{Q}G_{p^l}$ -módulo simple M_l tal que

$$\dim_{\mathbb{Q}} M_l \geq p^l - p^{l-1}.$$

Además, se tiene que

- $\dim_{\mathbb{Q}} M_l = p^l - p^{l-1}$; y
- $\mathbb{Q}G_{p^l} \cong \bigoplus_{i=0}^{l-1} \phi_{i,l}^*(M_i) \oplus M_l$.

Sugerencia: Haga inducción con respecto a l .

- Enuncie y pruebe enunciados análogos a los dos últimos para $p = 2$.
- Sea $p \in \mathbb{N}$ primo, $l \geq 1$ y sea M_l un $\mathbb{Q}G_{p^l}$ -módulo simple de dimensión $p^l - p^{l-1}$. Entonces M_l posee una base con respecto a la cual la matriz de la aplicación $a : m \in M \mapsto g_{p^l} m \in M$ es la matriz compañera del polinomio $\Phi_p(X^{p^l})$.
- Sea $f \in \mathbb{Q}[X]$ un polinomio mónico irreducible. Sea $a \in M_n(\mathbb{Q})$ la matriz compañera de f . Entonces, si $\mathcal{C}(a) \subset M_n(\mathbb{Q})$ es el centralizador de a en $M_n(\mathbb{Q})$, hay un isomorfismo de anillos $\mathcal{C}(a) \cong \mathbb{Q}[X]/(f)$.
- Sea $p \in \mathbb{N}$ primo. Para cada $l \in \mathbb{N}$, sea $\zeta_l \in \mathbb{C}$ una raíz primitiva p^l -ésima de la unidad y sea $\mathbb{Q}(\zeta_l)$ el menor subcuerpo de \mathbb{C} que la contiene. Entonces hay un isomorfismo de álgebras

$$\mathbb{Q}G_{p^l} \cong \mathbb{Q} \times \mathbb{Q}(\zeta_1) \times \cdots \times \mathbb{Q}(\zeta_l).$$

- Supongamos que n es impar y que $n = p_1^{m_1} \cdots p_r^{m_r}$ es la factorización de n como producto de potencias de primos distintos. Entonces $G_n \cong G_{p_1^{m_1}} \times \cdots \times G_{p_r^{m_r}}$.
Si M es un $\mathbb{Q}G_n$ -módulo simple, entonces existen $l_1, \dots, l_r \in \mathbb{N}$ tales que $l_i \leq m_i$ si $i \in \{1, \dots, r\}$ y

$$M \cong M_{p_1, m_1, l_1} \boxtimes \cdots \boxtimes M_{p_r, m_r, l_r}.$$

Aquí $M_{p, m, l}$ es el único $\mathbb{Q}G_{p^m}$ -módulo simple de dimensión $p^l - p^{l-1}$.

18. Muestre que si $k \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$, entonces $kS_3 \cong k \times k \times M_2(k)$.
19. Encuentre la descomposición de Wedderburn para kD_4 con $k \in \{\mathbb{Q}, \mathbb{R}, \mathbb{C}\}$ si $D_4 = \langle s, t : s^2 = t^4 = 1, sts = t^{-1} \rangle$.
20. Sea $Q = \{\pm 1, \pm i, \pm j, \pm k\}$ el grupo de los cuaterniones unitarios. Muestre que

$$\mathbb{Q}Q \cong \mathbb{Q} \times \mathbb{Q} \times \mathbb{Q} \times \mathbb{Q} \times \mathbb{H}_{\mathbb{Q}},$$

$$\mathbb{R}Q \cong \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{R} \times \mathbb{H}_{\mathbb{R}},$$

$$\text{y } \mathbb{C}Q \cong \mathbb{C} \times \mathbb{C} \times \mathbb{C} \times \mathbb{C} \times M_2(\mathbb{C}).$$

Aquí $\mathbb{H}_{\mathbb{R}}$ es el anillo de los cuaterniones reales y $\mathbb{H}_{\mathbb{Q}}$ es el análogo definido sobre \mathbb{Q} .

REFERENCIAS

- [1] F. Anderson y K. Fuller, Rings and categories of modules, *Graduate Texts in Mathematics* No. 13, Springer-Verlag, 1992, 376 pp.
- [2] L. E. Dickson, *On finite algebras*, Gött. Nachr., 358–393 (1905).
- [3] H. Maschke, *Über den arithmetischen Charakter der Coefficienten der Substitutionen endlicher linearer Substitutionsgruppen*, Math. Ann. No. 50, 492–498 (1898).
- [4] J. H. Maclagan Wedderburn, *A theorem on finite algebras*, Trans. Amer. Math. Soc. No. 6 (3), 349–352 (1905).
- [5] J. H. Maclagan Wedderburn, *On hypercomplex numbers*, London M. S. Proc. (2) No. 6, 77–118 (1908).

DEPARTAMENTO DE MATEMÁTICA, FCEYN, INSTITUTO DE MATEMÁTICA LUIS SANTALÓ, IMAS-CONICET, UNIVERSIDAD DE BUENOS AIRES. 1428, BUENOS AIRES

E-mail address: asolotar@dm.uba.ar, mariano@dm.uba.ar