

INF431

**Algorithmes et Programmation:
du séquentiel au distribué**

École polytechnique

Partie I

François Morain & Jean-Marc Steyaert

Table des matières

I	Programmation	13
1	Programmer avec Java	15
1.1	Modularité	15
1.1.1	Motivation	15
1.1.2	Modules	16
1.1.3	Contrôle d'accès	17
1.1.4	Types	17
1.2	Programmation par objets	18
1.2.1	Motivations	18
1.2.2	L'héritage par l'exemple	18
1.2.3	Propriétés de l'héritage	20
1.2.4	Contrôle d'accès et sous-classes	22
1.2.5	Un peu de typage	22
1.3	Classes abstraites, interfaces	26
1.3.1	Classe abstraite	26
1.3.2	Interfaces	28
1.4	Les génériques de Java	28
1.5	Procédural ou Objets?	29
2	Complexité et structures de données	31
2.1	Complexité élémentaire	31
2.1.1	Introduction	31
2.1.2	Calculs élémentaires de complexité	32
2.2	Structures de données	33
2.2.1	Représentation des ensembles	33
2.2.2	Piles, files, tas	34
II	Graphes	35
3	Propriétés élémentaires des graphes	37
3.1	Introduction	37
3.2	Définitions	37

3.2.1	Graphes non orientés	39
3.2.2	Sous-graphes, etc.	39
3.2.3	Degré	40
3.2.4	Quelques graphes non orientés classiques	42
3.3	Représentations en machine	42
3.3.1	Représentation par une matrice	43
3.3.2	Représentation par un tableau de listes	43
4	Accessibilité	45
4.1	Problématique	45
4.2	Définitions ; propriétés fondamentales	45
4.3	Fermeture transitive	46
4.3.1	Première solution par produit de matrice	47
4.3.2	L'algorithme de Roy et Warshall	48
4.4	Plus court chemin	50
4.5	Arbres	51
5	Parcours	55
5.1	Le cas non orienté	55
5.1.1	Définitions et propriétés générales	55
5.1.2	Arbre couvrant	57
5.1.3	Parcours en largeur d'abord	57
5.1.4	Parcours en profondeur d'abord (<i>depth-first search</i>)	63
5.1.5	Construction de l'arbre couvrant	64
5.2	Parcours dans le cas orienté	67
5.2.1	Propriétés générales	67
5.2.2	Propriétés du parcours en profondeur d'abord	68
5.2.3	Graphes sans cycles	72
5.2.4	Tri topologique	72
6	Topologie	75
6.1	Forte connexité	75
6.1.1	Définition, exemples	75
6.1.2	Le retour de Trémaux	76
6.1.3	Points d'attache	78
6.1.4	Application à la forte connexité	78
6.1.5	L'algorithme	79
6.1.6	L'exemple détaillé	81
6.2	Connexité dans les graphes non orientés	83
6.2.1	Énumération des composantes connexes	83
6.2.2	Points d'articulation	83
6.3	Planarité	87
6.3.1	Faces	88
6.3.2	Graphe dual	88

6.3.3	Relation d'Euler	89
6.3.4	Un critère théorique de planarité	92
6.3.5	Vers un algorithme de test de planarité	93
6.3.6	Une idée d'algorithme	94
6.4	Les programmes en Java	96
6.4.1	Composantes fortement connexes	96
6.4.2	Composantes connexes	97
6.4.3	Points d'articulation	98
7	Euler et Hamilton	101
7.1	Euler : fondateur de la théorie des graphes	101
7.2	Hamilton (1805–1865)	102
8	Introduction à l'optimisation combinatoire	105
8.1	L'algorithme de Dijkstra	105
8.1.1	Théorie	105
8.1.2	Implantations	108
8.2	Arbre couvrant de poids minimal	113
8.2.1	L'algorithme de Prim	114
8.2.2	L'algorithme de Kruskal	115
8.2.3	Ébauche de preuve	117
8.2.4	Prim ou Kruskal?	117
8.3	Les n reines	117
8.3.1	Prélude : les n tours	118
8.3.2	Des reines sur un échiquier	118
8.4	Les ordinateurs jouent aux échecs	120
8.4.1	Principes des programmes de jeu	120
8.4.2	Algorithme minimax	121
8.4.3	Principe de l'élagage $\alpha\beta$	121
8.4.4	Retour aux échecs	122
III	Langages	125
9	Langages rationnels	127
9.1	Recherche de motifs dans un texte	127
9.1.1	Un algorithme naïf	128
9.1.2	L'algorithme KMP	129
9.1.3	Multimotifs	132
9.2	Automates finis	134
9.2.1	Définitions	134
9.2.2	Langage reconnu par un automate fini	136
9.2.3	Combinaisons d'automates	138
9.3	Expressions régulières et langages rationnels	141

9.3.1	Langages rationnels : définitions	141
9.3.2	La réciproque : le théorème de Kleene	143
9.4	Déterminisme contre Non-Déterminisme	144
9.4.1	Simulation du non-déterminisme	145
9.4.2	Suppression des ε -transitions	145
9.4.3	Suppression du non-déterminisme	146
9.5	Complexité et limites	147
9.5.1	Déterminisation	147
9.5.2	Automate minimal	148
9.6	Langages non rationnels : le lemme d'itération	150
9.7	Appendice	151
10	Parenthésages, Langages algébriques	155
10.1	Exemples de structures parenthésées	155
10.1.1	Notations	155
10.2	Grammaires hors-contexte	156
10.2.1	Automates, langages et grammaires	157
10.2.2	Terminologie	158
10.2.3	Ambiguïté	160
10.3	Propriétés des langages algébriques	160
10.4	Grammaires de Greibach et automates à pile	163
10.4.1	Forme normale de Greibach	163
10.4.2	Automates à pile	166
10.5	Analyse par programmation dynamique	169
10.5.1	Forme normale de Chomsky	169
10.5.2	Reconnaissance déterministe des langages algébriques	170
10.6	Algébricité	174
10.6.1	Le lemme d'itération des langages algébriques	174
10.6.2	Langage générateur	176
11	Analyse syntaxique et autres langages	181
11.1	Éléments d'analyse syntaxique	181
11.1.1	Analyse descendante	182
11.1.2	Analyse ascendante	183
11.2	Au-delà des algébriques	188
11.2.1	L-systèmes	188
11.2.2	Autres classes	191
12	Analyses lexicale et syntaxique par l'exemple	193
12.1	Motivations	193
12.2	Construction de l'analyseur lexical	195
12.2.1	Utilisation des diagrammes de transition	196
12.2.2	Le programme en Java	196
12.2.3	Pour aller plus loin	198

12.3	Analyse syntaxique	199
12.3.1	Arbre de syntaxe abstraite	199
12.3.2	Construction des ASA	200
IV	Annexes	205
13	Compléments de Java	207
13.1	Paquetages et espace des noms	207
13.1.1	Définition et construction	207
13.1.2	Contrôle d'accès	208
13.1.3	Utilisation des .jar (java archive)	209
14	Quelques classes prédéfinies de Java	211
14.1	La classe Collection	211
14.1.1	La classe Vector	212
14.1.2	La classe LinkedList	212
14.1.3	La classe Hashtable	212
14.1.4	La classe TreeSet	213
15	La classe Graphe	215
15.1	Les choix	215
15.2	grapheX comme exemple de package	216
15.2.1	Comment ça marche ?	216
15.2.2	Faire un .jar	216
15.3	Sommets	217
15.4	La classe Arc	218
15.5	La classe abstraite Graphe et deux implantations	220
15.5.1	Définitions	220
15.5.2	Numérotation	220
15.5.3	Implantation par matrice	222
15.5.4	Implantation avec des listes	225

Présentation

Les buts du cours

Dans une première partie, ce cours se veut la continuation de l'apprentissage de base de la programmation, auquel s'ajoute la description d'algorithmes sur les graphes. Dans un deuxième temps, on s'intéresse à la présentation de la théorie des réseaux, ainsi que des protocoles.

De nombreux ouvrages existent déjà sur la théorie des graphes, comme par exemple le polycopié de R. Cori et J.-J. Lévy [4]. Il est difficile d'être novateur sur un tel sujet. Nous avons toutefois privilégié une approche de la programmation utilisant les bibliothèques déjà programmées en Java, bénéficiant des types génériques. Par exemple, nous utiliserons les possibilités de simplification d'exposition des algorithmes en utilisant des itérateurs sur les sommets d'un graphe, plutôt que de rester à un codage classique des sommets par des entiers. De la même façon, une classe abstraite de graphes nous permettra de nous concentrer sur les algorithmes plutôt que sur l'implantation réelle, que nous reléguerons dans les implantations par matrice ou tableau de liste. Sans vouloir à tout crin programmer de façon outrancière, le code résultat nous semble à la fois plus lisible et plus proche du pseudocode dans lequel les algorithmes seront décrits.

Remerciements

J.-J. Lévy nous a permis d'utiliser son poly pour la promotion X2004, ce dont nous le remercions chaleureusement.

P. Chassignet a prêté son concours à l'élaboration des classes de Graphe utilisées pour le poly. Il a également répondu à de multiples questions de tous ordres. Merci à P. Baptiste pour les discussions sur l'optimisation combinatoire. Merci aussi à J. Cervelle pour avoir répondu à des questions pointues sur l'utilisation de Java, version IIIe millénaire, ainsi que pour la relecture. Que les premiers lecteurs soient également remerciés : F. Pottier pour sa méticulosité, N. Sendrier, A. Cohen.

Polycopié du cours INF431, Version 1.2b [12 janvier 2007].

Notations

Le cardinal d'un ensemble E sera noté $|E|$.

Algorithme, pseudocode et programme

Quand on commence à programmer, on étudie et on implante des algorithmes simples, dont l'écriture coule de source. Arrivés à ce niveau de votre apprentissage, nous allons voir des algorithmes de plus en plus complexe, et se pose le problème de leur description indépendante d'un langage particulier et en négligeant dans un premier temps les détails d'implantation qui pourraient avoir tendance à compliquer inutilement la vue d'ensemble de l'algorithme. Pour résumer, nous décrirons un algorithme de manière synthétique et générique, à l'aide d'un langage intermédiaire, le *pseudocode*. La correction de l'algorithme proviendra souvent d'un théorème, l'étude de sa complexité esquissée à partir du pseudocode, et dépendant souvent de l'implantation choisie.

Quel pseudocode et quelles propriétés demander ? Nous modifions le pseudocode présenté de façon très classique dans [5, p. 17-18]. La syntaxe est très proche de langages de type Pascal, C, ou Java, et les principes du langage sont habituels en Java (cf. les objets).

1. Le symbole `//` indique que le reste de la ligne est un commentaire.
2. Les instructions seront terminées par un `;`.
3. L'indentation indique une structure de bloc ; on se passera, sauf cas particulier, d'indicateurs de type **début** ou **fin** pour alléger la lecture.
4. Les affectations entre type simple s'écriront `x <- 3` ; les tests d'égalité `==`. On utilisera les opérateurs de post et pré-incrémentation de Java à l'occasion.
5. On utilisera des boucles **tantque**, **pour**, **répéter**. Les variables d'itération continueront d'exister à la sortie des itérations.
6. Les variables sont locales aux procédures, nous n'emploierons pas de variable globale, sauf indication explicite.
7. `A[i]` désigne le i -ème élément du tableau `A`, et `A[2..j]` représente le sous-tableau de `A` d'indices compris entre 2 et j . Les tableaux seront numérotés à partir de 0. Les tableaux bi-dimensionnels seront notés `A[i][j]`.
8. Les données composées seront traitées comme des objets, constitués d'*attributs* ou *champs*. Comme en Java, un objet sera vu comme un pointeur sur les données. Si `f` est un champ de l'objet `x`, on notera `x.f`. La longueur d'un tableau `A` sera `longueur(A)`. Si `x` et `y` sont deux objets, faire `x <- y` fait que `x` et `y` pointent sur le même objet. Si un objet ne contient rien, il contient `NIL`. Dans le cas où on voudra recopier les objets champ par champ, on écrira `x := y`.
9. Les paramètres sont passés à une procédure *par valeur*. Les objets sont passés comme en Java.

10. Comme en Java, le *et* et le *ou* logiques sont *court-circuitants* (on dit aussi *paresseux*) : quand on évalue x ET y , on évalue x , et seulement s'il est vrai on évalue y .
11. Une liste sera écrite $L = (x_0, x_1, \dots, x_n)$, on écrira la concaténation $L \# z = (x_0, x_1, \dots, x_n, z)$ ou $z \# L^- = (z, x_0, x_1, \dots, x_n)$. Le premier élément d'une liste sera récupéré comme $tête(L) = x_0$; écrire

```
y <- tête(L);
```

fera que y contiendra x_0 et il restera (x_1, \dots, x_n) dans L , ce que l'on pourra noter $reste(L)$ si besoin est, de même que les éléments sauf le dernier seront désignés par $début(L)$. La queue de L sera le dernier élément, $queue(L) = x_n$. On aura donc formellement :

```
L = tête(L) # reste(L) = début(L) # queue(L)
```

12. Un ensemble sera noté $\{a, b, \dots\}$. Notons que des implantations spécifiques pourront être données quand nécessaire.
13. Si l'on veut afficher quelque chose à l'écran, on utilisera l'instruction *écrire*.

Donnons un exemple simple : recherchons le minimum d'un tableau. Notons que le type du tableau n'est pas explicité, et qu'on suppose implicitement qu'il existe un ordre sur les éléments du tableau.

```
rechercheMinimum(A)
  imin <- 0;
  pour i <- 1 à longueur(A)-1 faire
    si A[i] < A[imin] alors
      // on a trouvé un minimum local
      imin <- i;
  retourner A[imin];
```

Il sera facile de traduire ce pseudocode en Java, quel que soit le type de A .

Première partie
Programmation

Chapitre 1

Programmer avec Java

Ce chapitre ne se propose pas comme une description exhaustive de Java, mais plutôt comme un guide à l'utilisation de certains traits avancés, comme l'héritage et les classes abstraites. Il ne saurait se substituer à un cours de langages de programmation, comme par exemple celui de Majeure 1. On trouvera des éléments complémentaires dans [8]. On pourra consulter également les tutoriels de Java sur la page de Sun¹. Notre optique est d'être capable d'utiliser ces notions quand nous aborderons les graphes.

Plus philosophiquement, les langages modernes tournent autour de cette question : comment écrire au kilomètre des programmes corrects, lisibles et efficaces ? Une des façons de répondre est d'utiliser des classes prédéfinies, efficaces, standardisées, et nous en reparlerons au chapitre 14. Nous nous intéresserons plutôt ici aux mécanismes qui permettent la réutilisation et le partage de ces ressources.

1.1 Modularité

1.1.1 Motivation

C'est là un concept fondamental de la programmation. Dans un programme, on cherche toujours où on doit mettre quelles données. Pour s'y retrouver, il est beaucoup plus simple de pouvoir contrôler quelles données sont accessibles à qui (fonctions, classes, programmes, etc.). Dans certains langages, on a accès à essentiellement deux niveaux : une variable peut être *locale* à une fonction (ou plus généralement à un *bloc* de code) et n'est pas vue de l'extérieur de la fonction ; ou bien elle est *globale*, ce qui fait qu'elle est accessible à toutes les fonctions à la fois. Ces deux niveaux se révèlent rapidement insuffisants.

On est donc amené à introduire des contrôles plus stricts. On peut rassembler dans une même bibliothèque (appelée généralement *module* dans ce contexte), toutes les fonctions opérant sur des objets communs. L'idée derrière les modules est bien évidemment celui de la réutilisation et de la composition de petits modules pour en faire des plus gros, ce qui conduit à de gros programmes, etc.

¹<http://java.sun.com/docs/books/tutorial/index.html>

1.1.2 Modules

Très généralement, un module se compose de deux parties :

- l'*interface*² accessible depuis l'extérieur du module ;
- l'*implémentation* accessible de l'intérieur du module.

Un des avantages de cette dichotomie est de permettre d'utiliser le principe d'*encapsulation*. On définit les fonctionnalités du module dans la partie interface (données partagées, types, fonctions), car c'est ce qui intéresse l'utilisateur, et on en réalise une implantation dans la partie implémentation. Prenons l'exemple classique de la file d'attente. On a besoin de savoir faire trois opérations : construction, ajouter, supprimer. Mais on peut implanter les files d'au moins deux façons : par un tableau géré de façon circulaire, ou bien par une liste. L'extérieur n'a pas besoin de savoir quelle représentation est utilisée, et parfois, on peut la changer sous ses pieds.

Dans certains langages (comme Pascal), un module contient une interface et une implantation, et on ne peut avoir plusieurs implantations correspondant à la même interface. En Java, l'utilisation des classes abstraites permet de résoudre ce problème (cf. section 1.3).

Reprenons le cas d'une file d'attente. On peut écrire :

```
public class FIFO{
    private int debut, fin;
    private boolean pleine, vide;
    private int[] contenu;

    public FIFO(int n){
        debut = 0; fin = 0;
        pleine = n == 0; vide = true;
        contenu = new int[n];
    }
    public static void ajouter(int x, FIFO f){
        if(f.pleine) throw new Error ("File Pleine.");
        f.contenu[f.fin] = x;
        f.fin = (f.fin + 1) % f.contenu.length;
        f.vide = false; f.pleine = f.fin == f.debut;
    }
    public static int supprimer(FIFO f){
        if(f.vide)
            throw new Error ("File Vide.");
        int res = f.contenu[f.debut];
        f.debut = (f.debut + 1) % f.contenu.length;
        f.vide = f.fin == f.debut; f.pleine = false;
        return res;
    }
}
```

²Attention : en Java, interface est un mot clef pour désigner autre chose, cf. section 1.3.

On rend tous les champs privés (voir section 1.1.3) de sorte que l'utilisateur ne puisse utiliser directement l'implantation (ce qui rendrait son code sensible à tout changement de la définition de FIFO). Par contre, ce qui est public, c'est le constructeur, ainsi que les fonctions d'accès.

On pourrait également écrire une autre classe avec la même interface (et les mêmes signatures de fonction), mais cette fois ci utilisant une liste chaînée en interne :

```
public class FIFO{
    private Liste debut, fin;
    public FIFO(int n){ debut = null; fin = null; }
    public static void ajouter(int x, FIFO f){
        if(f.fin == null)
            f.debut = f.fin = new Liste (x);
        else{
            f.fin.suivant = new Liste(x);
            f.fin = f.fin.suivant;
        }
    }
    public static int supprimer(FIFO f){
        if(f.debut == null) throw new Error ("File Vide.");
        else{
            int res = f.debut.val;
            if(f.debut == f.fin)
                f.debut = f.fin = null;
            else f.debut = f.debut.suivant;
            return res;
        }
    }
}
```

1.1.3 Contrôle d'accès

Pour les membres d'une classe, il y a 3 types d'accès :

- **public** pour permettre l'accès depuis toutes les classes.
- **private** pour restreindre l'accès aux seules expressions ou fonctions de la classe courante.
- par défaut pour autoriser l'accès depuis toutes les classes du même paquetage (cf. chapitre 13).

1.1.4 Types

De même, on peut élargir les types des données pour fabriquer un type adapté à un cas spécifique, des types peuvent être locaux ou globaux, etc. En Java, types et modules se confondent dans la notion de classe, qui est à la fois type utilisateur et module contenant les fonctions associés à ce type.

1.2 Programmation par objets

1.2.1 Motivations

Les premiers langages informatiques relèvent d'une programmation dite *procédurale* où on applique des fonctions (ou procédures) à des données. La programmation dirigée par les données (ou programmation par objet) est un paradigme plus récent. Pour simplifier, la première vision du monde est celle de calculs appliqués à des paramètres d'entrée (donc une approche très fonctionnelle), alors que la seconde considère les objets comme la quantité importante, et qu'un objet peut posséder ses propres fonctions de traitement.

En Java, un objet est une instance d'une classe. Il a un état (la valeur de ses champs de données) et un ensemble de méthodes attachées. Les données statiques (*static*) sont partagées par tous les objets de la classe.

Quels avantages a-t-on à programmer objet ? Même si la modularité n'impose pas la programmation objet, celle-ci bénéficie énormément de cette possibilité, tant les deux ont le même objectif de rassembler en un même endroit (du programme) les données et fonctions applicables à un objet. Un autre avantage est celui de l'héritage : on peut définir des classes, qu'on va étendre dans un but plus spécifique, et donc aboutir à une programmation incrémentale, qui nous rapproche de notre objectif de l'écriture de gros programmes. C'est le cas de l'AWT (*Abstract Windowing Toolkit*) qui permet de faire des dessins en Java.

1.2.2 L'héritage par l'exemple

Nous allons considérer l'exemple ultra-classique des points et des points colorés. On commence par une classe `Point` qui code un point du plan par ses coordonnées `x` et `y` :

```
class Point{
    double x, y;
}
```

On décide de créer un nouveau type qui aura la particularité d'être un point avec une couleur :

```
class PointC extends Point{
    final static int JAUNE = 0, ROUGE = 1;
    int c;
    PointC(double x0, double y0, int col){
        x = x0; y = y0; c = col; // [1]
    }
    public static void main(String[] args){
        PointC pc = new PointC(3, 4, JAUNE);
    }
}
```

La classe `PointC` est une *sous-classe* de `Point`. Elle *hérite* des champs et des méthodes définis dans sa classe père, ce qui explique la syntaxe de la ligne [1] : à la création d'un `PointC`, on a déjà à sa disposition les deux champs `x` et `y` hérités de la classe `Point`.

Complétons l'exemple :

```

class Point{
    double x, y;
    static int NP = 1;
    static void afficherAbscisse(Point p){
        System.out.println("P: "+p.x);
    }
}
class PointC extends Point{
    final static int JAUNE = 0, ROUGE = 1;
    int c;
    PointC(double x0, double y0, int col){
        x = x0; y = y0; c = col;
    }
    public static void main(String[] args){
        PointC pc = new PointC(3, 4, JAUNE);
        afficherAbscisse(pc);
        System.out.println("NP="+NP);
    }
}

```

Il y a conversion *implicite* de `PointC` en `Point`, dans l'appel de la méthode `afficherAbscisse` de la classe `Point`. D'autre part, la variable statique `NP` est elle aussi accessible depuis `PointC`.

Dans la sous-classe, on ajoute ou on redéfinit des champs ou méthodes : on dit qu'on a *spécialisé* (en anglais, c'est l'*overriding*) ces champs ou méthodes. Par exemple, dans le code :

```

class Point{
    ...
    static void afficherAbscisse(Point p){
        System.out.println("Point: "+p.x);
    }
}
class PointC extends Point{
    ...
    static void afficherAbscisse(Point p){
        System.out.println("PointC: "+p.x);
    }
    public static void main(String[] args){

```

```

    PointC pc = new PointC(3, 4, JAUNE);
    afficherAbscisse(pc);
    Point.afficherAbscisse(pc);
}
}

```

on a spécialisé dans `PointC` la méthode `afficherAbscisse`. On peut également spécialiser des méthodes d'objet :

```

class Point{
    ...
    void afficher(){
        System.out.println("p: "+x);
    }
}
class PointC extends Point{
    ...
    void afficher(){
        System.out.println("pc: "+x);
    }
    public static void main(String[] args){
        PointC pc = new PointC(3, 4, JAUNE);
        Point p = new Point();
        p.x = 1;

        pc.afficher();
        p.afficher();
    }
}

```

1.2.3 Propriétés de l'héritage

En Java, l'héritage est dit *simple*³, c'est-à-dire qu'une classe peut être sous-classe d'une autre classe et d'une seule. Par contre, une classe peut avoir plusieurs sous-classes. Les diagrammes d'héritage sont donc des arbres. Cela facilite grandement le choix de la méthode à appliquer à un objet, selon le principe suivant : on choisit *statiquement* la méthode la plus spécialisée, c'est-à-dire appartenant à la plus petite sous-classe connue contenant l'objet.

La classe père est accessible par le mot clef **super** (un peu comme (`ClassePere`) **this**); on ne peut l'utiliser que dans des méthodes d'objets : **super.f** est la méthode `f` de la classe parente; **super()** (avec d'éventuels arguments) est le constructeur (appelé par défaut) de la classe parente. Par exemple :

³contrairement à C++, Smalltalk ou Ocaml, dans lesquels l'héritage est *multiple*.

```

class PointC extends Point{
    final static int JAUNE = 0, ROUGE = 1;
    int c;
    PointC(double x0, double y0, int col){
        super(); // facultatif
        c = col;
    }
    public static void main(String[] args){
        PointC pc = new PointC(3, 4, JAUNE);
    }
}

```

Dans le cas d'un constructeur explicite pour la classe père, on est obligé d'appeler **super** :

```

class Point{
    double x, y;
    Point(double x0, double y0){
        x = x0; y = y0;
    }
}
class PointC extends Point{
    final static int JAUNE = 0, ROUGE = 1;
    int c;
    PointC(double x0, double y0, int col){
        super(x0, y0); // nécessaire
        c = col;
    }
    public static void main(String[] args){
        PointC pc = new PointC(3, 4, JAUNE);
    }
}

```

Exemple. En Java, on peut masquer des champs. Par exemple, le code suivant est (malheureusement ?) licite :

```

class PointC extends Point{
    final static int JAUNE = 0, ROUGE = 1;
    int x;
    int c;
    ...
}

```

Le nouveau x masque l'ancien, qui est toujours accessible par **super**.

1.2.4 Contrôle d'accès et sous-classes

Les champs ou méthodes d'une classe peuvent être :

- **public** pour permettre l'accès depuis toutes les classes ;
- **private** pour restreindre l'accès aux seules expressions ou fonctions de la classe courante ;
- par défaut pour autoriser l'accès depuis toutes les classes du même paquetage ;
- **protected** pour restreindre l'accès aux seules expressions ou fonctions de sous-classes de la classe courante.
- Un champ **final** ne peut être redéfini (on peut donc optimiser sa compilation).
- une classe **final** ne peut être spécialisée.
- une méthode peut être déclarée **final** (pas besoin de déclarer toute la classe **final**).

Les méthodes redéfinies doivent fournir au moins les mêmes droits d'accès que les méthodes originellement définies.

1.2.5 Un peu de typage

Théorie

Rappelons qu'un type est un ensemble de valeurs partageant une même propriété. En Java, il y a des types primitifs : **int**, **char**, etc ; une classe est un type. On a déjà rencontré des classes prédéfinies comme la classe `String`.

Comment le compilateur vérifie-t-il le typage ? Il s'agit d'une vérification statique (à la compilation) contrairement au typage dynamique (à l'exécution). L'intérêt du typage statique est de permettre la détection d'erreurs le plus tôt possible, mais aussi d'optimiser le code, et d'assurer une certaine sécurité (pas d'atteinte à la mémoire par exemple). Le typage dynamique peut être dans certains cas plus précis (prendre l'exemple du cas de **if(a) return 1; else return false;**).

Il n'est pas question ici de faire un cours de théorie des types. Nous renvoyons au cours Principes des Langages de Programmation de Majeure 1, ainsi qu'aux livres de Benjamin Pierce ou de Abadi-Cardelli.

Et en Java

En Java, chaque objet possède un certain type lors de sa création. L'opérateur **instanceof** teste l'appartenance d'un objet à une classe. Ainsi :

```
if(p instanceof PointC)
    System.out.println ("couleur = " + p.c);
```

Donc l'expression `(PointC)p` équivaut à :

```
if(p instanceof PointC)
    p
```

```

else
  throw new ClassCastException();

```

On parle de sous-classe comme on parle de sous-type. On note $x : t$ pour dire que x est de type t et $t <: t'$ signifie $x : t \Rightarrow x : t'$ pour tout x . On dit que t est un *sous-type* de t' . En Java, les types suivants sont des sous-types les uns des autres, et il y a conversion explicite si nécessaire :

```

byte <: short <: int <: long
float <: double
char <: int

```

Une classe est un type; une sous-classe un sous-type. On propage la notation : $C <: C'$ si C est une sous-classe de C' .

Ex. `PointC <: Point`.

On a déjà dit qu'en Java, l'héritage était simple. Il existe ainsi un arbre qui contient toute la hiérarchie des classes, dont la racine est la classe `Object` :

$$\forall C \quad C <: \text{Object}$$

(C est une classe ou un tableau quelconque).

Les méthodes de `Object` sont `clone`, `finalize`, `getClass`, `hashCode`, `notify`, `notifyAll`, `wait`, `equals`, `toString`. Seules `equals`, `toString`, `finalize` peuvent être redéfinies.

Comme conséquence de ce qui précède, *tous* les objets contiennent ces méthodes.

Comment se raccrochent les types primitifs dans ce tableau ? On convertit les scalaires `int`, `float`, `char`, ... en objets avec un « conteneur » (*autoboxing*) :

```

int x = 3;
Objet xobj = new Integer(x);
int y = xobj.intValue();

```

Quelques exemples extrêmes

Le code qui suit est valide, mais est (souvent) déconseillé⁴ :

```

class A{
  int x;

  A(){ x = 1; }
}
class B extends A{
  B(){ x = 2; }
  public static void main(String[] args){

```

⁴Cette section peut être sautée en première lecture.

```

    B b = new B();
    A a = new B();
    System.out.println(b.x);
    System.out.println(a.x);
}

```

Le programme affiche :

```

2
2

```

Le *type créé* de `a` est `B`, mais son *type apparent* est `A`. La conversion `A a = new B()` marche puisque `B` est sous-classe de `A`.

Cette propriété permet d'écrire :

```

class Point{
    double x, y;
    static void translation(Point p,
                           double dx, double dy){
        p.x += dx; p.y += dy;
    }
}
class PointC extends Point{
    final static int JAUNE = 0, ROUGE = 1;
    int c;
    PointC(double x0, double y0, int col){
        x = x0; y = y0; c = col;
    }
    public static void main(String[] args){
        PointC pc = new PointC(3, 4, JAUNE);
        translation(pc, 1, 2);
    }
}

```

Par contre, le code suivant ne peut marcher :

```

class Point{
    ...
    static Point translation(Point p,
                            double dx, double dy){
        Point pt = new Point();
        pt.x = p.x + dx; pt.y = p.y + dy;
        return pt;
    }
}

```

```

class PointC extends Point{
    ...
    public static void main(String[] args){
        PointC pc = new PointC(3, 4, JAUNE);
        pc = (PointC)translation(pc, 1, 2);
    }
}

```

Il n'y a pas de conversion père \rightarrow fils. La compilation marche (avec le cast), mais pas l'exécution (ClassCastException).

Héritage et surcharge

La surcharge est déterminée à la *compilation*. On parle de liaison *retardée* : l'héritage permet d'appeler une méthode de la même classe sans savoir exactement quelle sous-classe va l'instancier.

Pour illustrer cette notion, on demande quelle est la valeur affichée par le programme suivant ?

```

class C{
    void f(){
        g();
    }
    void g(){
        System.out.println(1);
    }
}
class D extends C{
    void g(){
        System.out.println(2);
    }
    public static void main(String[] args){
        D x = new D();
        x.f();
    }
}

```

La réponse est 2. En effet, que se passe-t-il ? Le compilateur ne se plaint pas, car à la création, x étant de type $D <: C$, il possède bien une méthode f . À l'exécution, on utilise la règle suivant laquelle on applique la méthode de la plus petite classe possible contenant x , ici la classe D , qui a bien une méthode g , qui affiche 2!

Notons que ce genre de propriété est difficile à utiliser quand on débute. Il n'accroît pas la lisibilité du code non plus, et donc nous engageons le programmeur à ne l'utiliser que quand la situation l'exige, et elle peut très bien ne jamais se produire...

Exercice 1.2.1 Quel est le résultat produit par le programme suivant quand $T, T' \in \{A, B\}$ et $U, U' \in \{A, B\}$ avec $T' <: T$ et $U' <: U$?

```

class A{
    public static void main(String[] args){
        T x = new T'(); U y = new U'();
        System.out.println (x.f(y));
    }

    int f(A y) { return 1; }
}

class B extends A{
    int f(B y) { return 2; }
}

```

1.3 Classes abstraites, interfaces

1.3.1 Classe abstraite

On veut permettre l'implantation de deux interfaces de modules de deux façons différentes. On sépare (comme dans le cas de la FIFO), la partie spécification, de la partie implantation. On pourra ainsi écrire (en utilisant une syntaxe objet) :

```

public abstract class FIFO{
    public abstract void ajouter(int x);
    public abstract int supprimer();
}

```

dont on aura deux implantations :

```

public class FIFOTAB extends FIFO{
    private int debut, fin;
    private boolean pleine, vide;
    private int[] contenu;

    public FIFO(int n){
        debut = 0; fin = 0;
        pleine = n == 0; vide = true;
        contenu = new int[n];
    }
    public void ajouter(int x){
        if(pleine)
            throw new Error ("File Pleine.");
        contenu[fin] = x;
    }
}

```

```

    fin = (fin + 1) % contenu.length;
    vide = false; pleine = fin == debut;
}
public static int supprimer(){
    if(vide)
        throw new Error ("File Vide.");
    int res = contenu[debut];
    debut = (debut + 1) % contenu.length;
    vide = fin == debut; pleine = false;
    return res;
}
}

```

et

```

public class FIFOLIB extends FIFO{
    private Liste debut, fin;
    public FIFO(int n){ debut = null; fin = null; }
    public void ajouter(int x){
        if(fin == null)
            debut = fin = new Liste(x);
        else{
            fin.suivant = new Liste(x);
            fin = fin.suivant;
        }
    }
    public int supprimer(){
        if(debut == null) throw new Error ("File Vide.");
        else{
            int res = debut.val;
            if(debut == fin)
                debut = fin = null;
            else debut = debut.suivant;
            return res;
        }
    }
}

```

qu'on pourra utiliser sans problème dans :

```

public class TestFIFO{
    public static void main(String[] args){
        FIFOTAB f = new FIFOTAB(10);
        f.ajouter(1);
        FIFOLISTE g = new FIFOLISTE(10);
        g.ajouter(1);
    }
}

```

```

        int x = g.supprimer();
    }
}

```

Énumérons quelques propriétés des classes abstraites :

- Une classe abstraite contient des champs indéfinis.
- On ne peut pas créer d’instances de classes abstraites.
- Permet d’utiliser des types disjonctifs (comparer avec les types somme de Caml, les variantes de Pascal; les union de C).

1.3.2 Interfaces

Parfois, on a besoin d’une classe plus rigide qu’une classe abstraite, qui ne spécifie pas complètement un module. On utilise alors une interface, qui est une classe abstraite avec les propriétés suivantes :

- dont tous les champs sont indéfinis. Ses champs de données sont constants ; ses méthodes sont abstraites et publiques.
- Une interface peut *spécialiser* une autre interface.
- Une classe peut *implémenter* une ou plusieurs interfaces.
- Il n’y a pas de fonctions statiques, ou champs de données modifiables dans un interface.
- Les interfaces sont une bonne manière d’exiger la présence de certains champs dans une classe.

Donnons un exemple simple :

```

public interface Couleur{
    final int ROUGE = 0, JAUNE = 1;
}
interface PointTournant{
    void static rotation(Point p, double theta);
}
interface PointMobile{
    void static translation(Point p,
                            double dx, double dy);
}
class Point implements PointTournant, PointMobile{
    ...
}

```

1.4 Les génériques de Java

Java possède depuis peu une notion de polymorphisme. On peut avoir des variables de type générique `<E>` :

```

public class LinkedList<E> {
    void add(int i, E x) { ... }
    E get(int i) { ... }
    E getFirst() { ... }
    ListIterator<E> listIterator(int i) { ... }
}
public interface ListIterator<E>
    extends Iterator<E>
    boolean hasNext();
    E next();
}

```

ce qui rend extrêmement pratique la réutilisation de code, ce dont nous ne nous privons pas dans le reste du cours. Par exemple, une liste d'entiers sera simplement utilisée de la façon suivante :

```

LinkedList<Integer> L = new LinkedList<Integer>();
L.add(1);

```

Signalons un inconvénient (expliqué dans la documentation de Sun). On ne peut écrire `E x = new E();` ou `E[] t = new E[10];` pour des raisons de typage. Cela nous conduira à quelques contorsions plus loin, mais ces classes sont très faciles à utiliser malgré tout.

1.5 Procédural ou Objets ?

On peut se poser diverses questions pour choisir : préfère-t-on contrôler les fonctions ou les données ? Les petits programmes peuvent se faire indifféremment dans les deux styles. Si le programme devient gros (disons au-delà de 10,000 lignes), la programmation objet a des avantages certains. Entrent alors en ligne de compte des considérations comme la stratégie de modification ou d'évolution du code, et du côté incrémental que cela peut représenter, ce qui est très prisé dans l'industrie.

La programmation objet est quasi impérative quand on utilise des classes prédéfinies déjà en syntaxe objet (AWT).

Gardons la tête froide. En dernière analyse, c'est une affaire de goût. . . . En tout cas, arrivés à ce point du cours, nous avons toutes les clefs pour commencer à naviguer dans les classes prédéfinies de Java, et pour programmer les graphes de manière agréable.

Chapitre 2

Complexité et structures de données

Le but de ce chapitre est de rappeler quelques notions simples de complexité. Pour illustrer les concepts, nous rappellerons les structures de données élémentaires vues dans les cours précédents, en insistant sur leur complexité en terme d'accès, insertion ou suppression. Nous indiquerons également sous quelle forme elles sont disponibles en Java dans l'annexe 14.

2.1 Complexité élémentaire

2.1.1 Introduction

La recherche en algorithmique porte sur l'élaboration d'algorithmes résolvant un problème donné. La comparaison des différentes méthodes proposées repose généralement sur le temps de calcul des différentes approches. On peut se pencher sur l'analyse exacte de l'algorithme, son coût asymptotique, son coût moyen ou dans le cas le pire, ou bien s'attacher à sa facilité d'implantation (c'est là plus subjectif). Le coût peut être un coût en temps aussi bien qu'en espace. Si $T(n)$ est le coût d'un algorithme opérant sur des entrées de taille n , on appelle coût amorti la quantité $T(n)/n$. Nous verrons que beaucoup d'algorithmes sur les graphes ont un coût amorti $O(1)$.

Être capable d'estimer le coût théorique d'un algorithme est important quand on l'implante, ce qui permet de vérifier qu'il évolue bien en fonction de ce que prédit la théorie.

Trouver des paramètres pertinents pour analyser un algorithme n'est pas toujours simple et demande souvent un peu de métier. Dans les exemples que nous rencontrerons, nous justifierons le choix du ou des paramètres.

Un sujet encore plus poussé est celui du temps minimum pour résoudre un problème donné ; la recherche de bornes inférieures est très difficile et possible seulement dans un petit nombre de cas. On peut aussi essayer de classifier les problèmes en fonction de leur difficulté, mais ce sujet sera abordé en Majeure 2.

2.1.2 Calculs élémentaires de complexité

Ce que l'utilisateur veut généralement est le temps que va prendre son ordinateur pour résoudre un problème donné. Évidemment, il entre en ligne de compte les performances de la machine. On les abstrait généralement en les cachant dans les constantes d'implantation. On reste à un niveau théorique en remplaçant le temps par le nombre d'opérations élémentaires effectuées (affectations, comparaisons, +, *, etc.). Il suffirait de mesurer le temps pris par celles-ci sur une machine donnée pour pouvoir extrapoler le temps pris sur l'ordinateur cible.

Énumérons quelques règles courantes :

1. La complexité d'une affectation, opération de lecture ou d'écriture est généralement de $O(1)$. C'est surtout vrai pour les types primitifs.
2. Le coût est considéré être additif : le coût de deux instructions indépendantes est $T(P; Q) = T(P) + T(Q)$. Le coût d'une itération est plus généralement : $T(\text{for}(i = 0; i < n; i++) P(i);) = \sum_{i=0}^{n-1} T(P(i))$.
3. Le coût d'une instruction conditionnelle est le maximum des coûts des deux branches (si le calcul de la condition est négligeable). On peut faire mieux si on connaît la probabilité de passer dans l'une ou l'autre branche pour évaluer le coût moyen.

Souvent, on compare le comportement asymptotique de deux algorithmes. Rappelons que si f, g sont deux fonctions à valeur ≥ 0 , on écrit $f = O(g)$ si et seulement si le rapport f/g est borné à l'infini :

$$\exists n_0, \exists K, \forall n \geq n_0, 0 \leq f(n) \leq Kg(n).$$

De telles comparaisons asymptotiques conduisent généralement à isoler les coûts dominants dans l'algorithme.

Donnons un exemple, celui de la multiplication de matrices. Soient A et B deux matrices $n \times n$ à coefficients entiers et $C = AB$. Rappelons que :

$$\forall i, j, c_{i,j} = \sum_{k=0}^{n-1} a_{i,k} b_{k,j}.$$

Le code Java correspondant est :

```

static int[][] mult(int[][] A, int[][] B){
    int n = A.length;
    int[][] C = new int[n][n];

    for(int i = 0; i < n; i++)
        for(int j = 0; j < n; j++)
            for(int k = 0; k < n; k++)
                C[i][j] += A[i][k] * B[k][j];
    return C;
}

```

Il est clair que le programme fait trois boucles **for** imbriquées, donc le coût est $O(n^3)$ multiplications, et $O(n^3)$ additions. Il est courant de supposer qu'une multiplication est plus coûteuse qu'une addition, donc on se contentera de considérer que la multiplication de matrices coûte $O(n^3)$.

Pour la petite histoire, Strassen a montré qu'il existe un algorithme demande $O(n^{\log_2 7})$ multiplications. Il est classique de conjecturer que l'algorithme optimal a une complexité $O(n^\omega)$ avec $\omega \geq 2$. Le record actuel est toujours 2.38 (D. Coppersmith et S. Winograd). Les optimistes pensent que $\omega = 2$ est la bonne valeur (il y a des progrès récents liés à la théorie des groupes, qui iraient dans ce sens¹). C'est un domaine de recherche actif de chercher à faire diminuer ω . Notons que beaucoup d'algorithmes, notamment de calcul formel, peuvent se ramener à la multiplication de matrices, ce qui en fait une primitive incontournable.

Citons d'autres exemples en vrac. Les meilleurs algorithmes de tri de n objets par comparateur ont une complexité en $O(n \log n)$: c'est le cas de heap sort, merge sort, quicksort. Trouver un élément dans une liste de taille n coûte $O(n)$ opérations. Recherche un élément dans un tableau trié coûte $O(\log n)$ (par dichotomie). Énumérer tous les éléments de $\{0, 1\}^n$ prend $O(2^n)$, ce qui fournit un exemple simple de problème difficile, dont le temps de calcul est exponentiel dans la taille de l'input.

2.2 Structures de données

Lors des cours précédents, des structures aussi simples que tableaux, listes et arbres ont été présentées, et des résultats prouvés sur leurs performances en temps d'accès. Nous allons prendre un point de vue dual. Soit E un ensemble de n objets (que nous supposons tous distincts). Quelles sont les représentations possibles de E et leurs différentes propriétés.

Dans la suite du cours, nous utiliserons ces structures de base pour représenter les graphes, mais également pour enrichir les graphes avec des propriétés particulières. Le choix d'une bonne structure d'accompagnement sera crucial dans la performance de certains algorithmes (Dijkstra par exemple).

2.2.1 Représentation des ensembles

On suppose que l'on veut stocker ou supprimer un élément, et tester si un élément appartient à E . Si on représente E par un tableau de taille n , ajouter un élément est facile, si on garde l'indice de la dernière case occupée. Retrouver (ou supprimer) un élément coûtera $O(n)$. Si un ordre sur les éléments est connu, on peut maintenir le tableau trié, ce qui nous donnera un coût de recherche ou suppression en $O(\log n)$. Il n'en reste pas moins que la gestion des places vides peut être compliqué.

Si on utilise une liste, les coûts d'insertion et suppression seront également en $O(n)$, mais la place sera toujours calculée au plus juste. Un ordre sur les éléments pourra aider un peu, au prix de la complication du code.

¹<http://www.cs.caltech.edu/~umans/papers/CKSU05.pdf>

Utiliser un arbre est possible si un ordre est connu. Cela garantira un coût d'insertion/suppression/recherche en $O(\log n)$, en équilibrant les arbres (AVL) ou par d'autres techniques (arbres bicolores, etc.). Les programmes résultant peuvent être compliqués.

La technique la plus efficace est celle du hachage. Elle permet d'avoir un coût d'insertion/suppression/recherche en $O(1)$, ce qui est optimal comme on s'en convaincra. L'idée consiste à associer de manière unique un entier à chaque objet, ce qui permettra de le repérer facilement dans un tableau. La difficulté est de trouver une bonne fonction de hachage. Une technique standard consiste à associer à un objet une représentation sous forme d'une chaîne de bits, puis à appliquer une fonction de transformation pour en tirer un petit entier.

2.2.2 Piles, files, tas

Les piles et les files permettent des opérations rapides sur des cas particuliers d'ensemble. Une pile permet de stocker le dernier élément arrivé avec un coût $O(1)$ et de le retirer avec le même coût. La file permet de stocker le dernier arrivé en $O(1)$ et de retirer le premier arrivé en $O(1)$ également.

Une file de priorité permet de stocker les éléments de E de sorte que le minimum soit trouvable en $O(1)$ et les coûts de mise à jour (insertion ou suppression) en $O(\log n)$. Les implantations les plus courantes utilisent un *tas*.

Deuxième partie

Graphes

Chapitre 3

Propriétés élémentaires des graphes

3.1 Introduction

Une partie importante de tout système d'informations consiste à modéliser les données ou les propriétés du système. Si les données sont indépendantes, on peut les représenter par un ensemble, qui sera implanté par exemple dans un tableau ou une liste. Si un lien hiérarchique simple est mis en évidence, on peut stocker les informations sous forme d'arbre (hiérarchie par exemple). Que faire quand les liens sont plus complexes, comme par exemple quand on doit représenter la carte d'un réseau routier, un circuit électronique ? On utilise alors souvent des graphes, qui permettent de coder de telles informations. Par exemple, on peut imaginer gérer une carte du réseau ferré en stockant les gares de France et les différentes distances entre ces gares.

Une fois que l'on dispose d'une telle représentation, on travaille généralement sur son abstraction en tant que graphe, ce qui permet d'utiliser de nombreux algorithmes résolvant des problèmes précis, comme par exemple trouver un chemin minimal en temps (ou en ressources) entre deux gares.

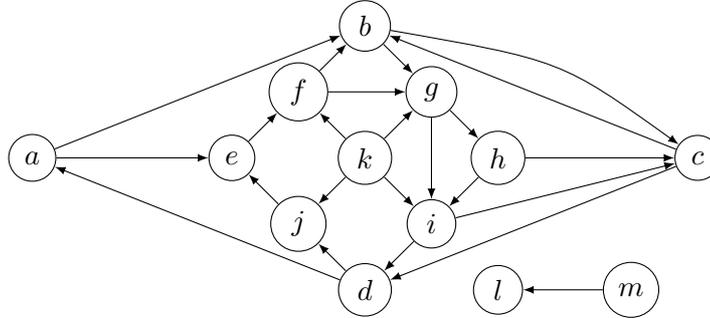
Citons encore comme exemples les graphes de dépendance entre fichiers sources de façon à optimiser la compilation, le diagramme d'héritage des classes dans un langage typé, mais aussi la représentation de tâches à accomplir dans une usine avec contraintes de précédence, etc.

3.2 Définitions

Un *graphe* $\mathcal{G} = (\mathcal{S}, \mathcal{A})$ est donné par un ensemble \mathcal{S} de *sommets* et un ensemble \mathcal{A} d'*arcs*, $\mathcal{A} \subset \mathcal{S} \times \mathcal{S}$.

On visualise en général un graphe par un dessin dans le plan. Les sommets sont représentés par des points, les arcs par des flèches. Le graphe \mathcal{G}_1 de la figure 3.1 a pour ensemble de sommets $\mathcal{S} = \{a, b, c, d, e, f, g, h, i, j, k, l, m\}$, et ensemble d'arêtes

$\mathcal{A} = \{(a, b), (b, c), (c, b), (c, d), (d, a), (a, e), (e, f), (f, b), (f, g), (b, g), (g, h), (h, c), (h, i), (g, i), (i, c), (i, d), (d, j), (j, e), (k, f), (k, g), (k, i), (k, j), (m, l)\}$.

FIG. 3.1 – Le graphe \mathcal{G}_1 .

Il est important de noter qu'il n'existe pas de dessin canonique d'un graphe. Par exemple, les deux dessins de la figure 3.2 correspondent au même graphe.

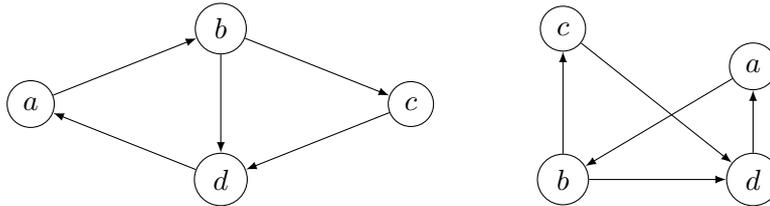
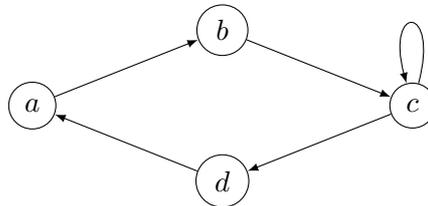


FIG. 3.2 – Deux dessins d'un même graphe.

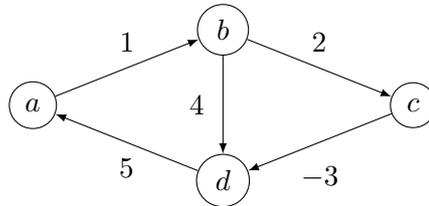
Un arc $\alpha = (a, b)$ est *orienté* de a vers b ; α a pour *origine* a et *destination* b ; l'arc α est *incident* à a ainsi qu'à b . On dira également que a est un *prédécesseur* de b , et b un *successeur* de a ; a et b seront également dits *adjacents*. L'arc (x, x) est une *boucle*.

Un graphe est *simple* s'il ne comporte pas de boucles. Le graphe de la figure 3.1 est simple. Le graphe \mathcal{G}_2 de la figure 3.3 ne l'est pas.

FIG. 3.3 – Le graphe \mathcal{G}_2 .

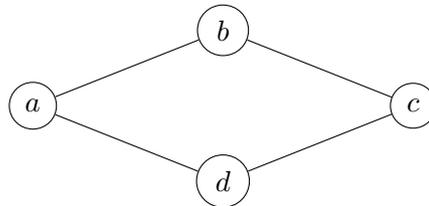
On peut enrichir un graphe en considérant des quantités associées à chaque arc, ou à chaque sommet. Penser par exemple à la carte de circulation d'une ville, avec ses sens

interdits et les longueurs des rues. On parlera d'*arcs valués*, c'est-à-dire qu'on utilisera une fonction $\nu : \mathcal{A} \rightarrow \mathbb{N}$ qui donnera le poids ou la valeur associée à un arc. C'est le cas du graphe \mathcal{G}_3 de la figure 3.4 dans lequel on a ajouté des poids sur chaque arc.

FIG. 3.4 – Le graphe valué \mathcal{G}_3 .

3.2.1 Graphes non orientés

Les graphes les plus généraux que nous considérerons sont *a priori orientés*, ce qui veut dire qu'un arc indique un sens de lecture ou de dessin. Autrement dit, si l'arc (a, b) existe, cela n'implique pas l'existence de l'arc (b, a) dans le graphe. Il peut arriver que pour tout arc (a, b) , son inverse (b, a) existe également. On dit que le graphe \mathcal{G} est *non orienté*. Dans ce cas, on utilise le terme *arête* au lieu d'arc et on note parfois $\{a, b\}$ l'arête reliant a à b ¹. On dessine alors des traits au lieu de flèches, comme dans l'exemple du graphe \mathcal{G}_4 de la figure 3.5.

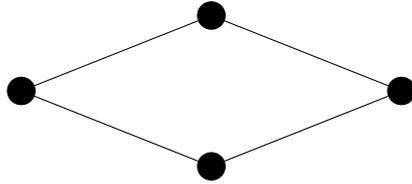
FIG. 3.5 – Le graphe \mathcal{G}_4 .

3.2.2 Sous-graphes, etc.

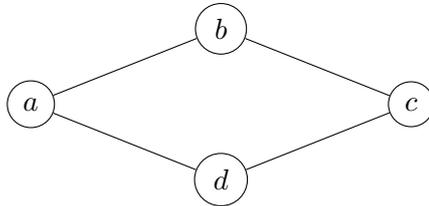
Il peut aussi arriver que seule la forme du graphe importe. On remplace alors les sommets par des cercles. Le graphe de la figure 3.5 devient alors simplement celui de la figure 3.6. De nombreux graphes deviennent isomorphes si on anonymise leurs nœuds.

De façon similaire, on peut avoir à considérer le graphe \mathcal{G} "sans ses flèches", c'est-à-dire considérer sa forme non orientée. On enlève alors les boucles et les flèches. La

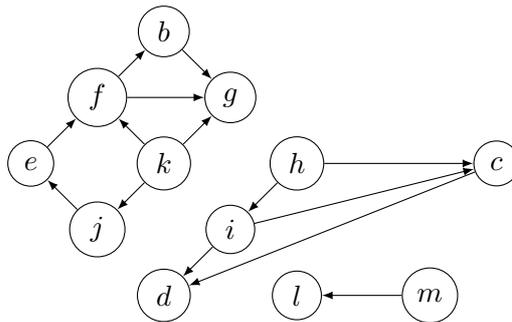
¹Nous considérerons des paires au sens propre, c'est-à-dire qu'un graphe non orienté ne possède pas de boucles.

FIG. 3.6 – Le graphe abstrait correspondant à \mathcal{G}_4 .

version non orientée (ou graphe non orienté *sous-jacent*) du graphe \mathcal{G}_2 est donnée à la figure 3.7.

FIG. 3.7 – Le graphe non-orienté sous-jacent au graphe \mathcal{G}_2 .

Si $\mathcal{G} = (\mathcal{S}, \mathcal{A})$ est un graphe et \mathcal{S}' un sous-ensemble de \mathcal{S} , on note $\mathcal{A}|\mathcal{S}'$ l'ensemble des arêtes de \mathcal{A} qui ont leurs deux extrémités dans \mathcal{S}' . Le graphe $\mathcal{G}' = \mathcal{G}|\mathcal{S}' = (\mathcal{S}', \mathcal{A}|\mathcal{S}')$ est appelé le *graphe induit* de \mathcal{G} par \mathcal{S}' . Tout graphe $(\mathcal{S}', \mathcal{A}')$ avec $\mathcal{S}' \subset \mathcal{S}$ et $\mathcal{A}' \subset \mathcal{A}$ est appelé *sous-graphe* de \mathcal{G} ; si $\mathcal{S}' = \mathcal{S}$, on parle de *graphe couvrant*. Pour le graphe de la figure 3.1, on trouve dans les figures 3.8, 3.9, 3.10 des exemples de chaque catégorie.

FIG. 3.8 – Un sous-graphe du graphe \mathcal{G}_1 .

3.2.3 Degré

Définition 3.2.1 Le degré d'un sommet s dans un graphe non orienté est le nombre d'arêtes incidentes à s .

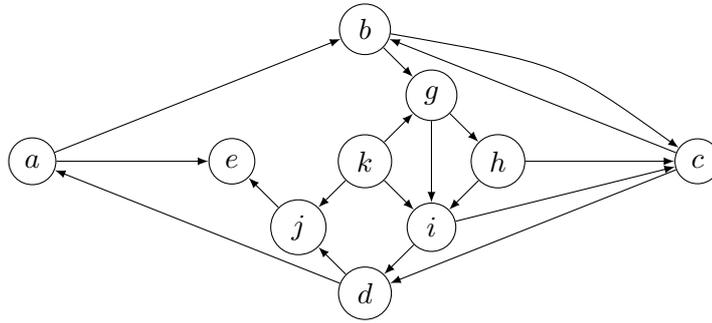
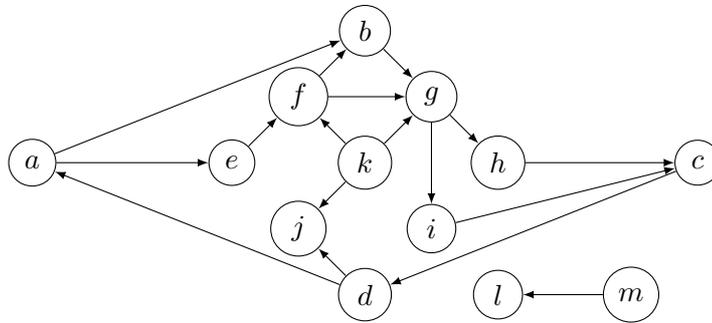
FIG. 3.9 – Un graphe induit de \mathcal{G}_1 .

FIG. 3.10 – Un graphe couvrant.

Proposition 3.2.1 (hand-shaking lemma) Dans tout graphe \mathcal{G} non orienté, on a $\sum_s \deg(s) = 2|\mathcal{A}|$.

Démonstration. Chaque arête contribue pour deux sommets dans la somme. \square

Corollaire 3.2.1 Soit \mathcal{G} un graphe non orienté. Le nombre de sommets de degré impair est pair.

Démonstration. On peut récrire la somme précédente comme :

$$\sum_s \deg(s) = \sum_{s, \deg(s) \text{ impair}} \deg(s) + \sum_{s, \deg(s) \text{ pair}} \deg(s).$$

Les deux sommes extrêmes sont paires, donc celle du milieu également. \square

Dans le cas des graphes orientés, on peut définir de la même façon le *degré entrant* (le nombre d'arcs arrivant sur un sommet), et le *degré sortant* (le nombre d'arcs partant d'un sommet).

3.2.4 Quelques graphes non orientés classiques

Remarquons qu'une liste ou un arbre forment des exemples de graphes.

Le graphe *complet* à n sommets est le graphe dans lequel chaque sommet a $n - 1$ voisins, voir la figure 3.11. Le graphe *biparti complet* $K_{i,j}$ contient $i + j$ sommets répartis en deux sous-ensembles, l'un contenant i sommets, l'autre j sommets. On relie ensuite chacun des i sommets à tous les j autres. On a dessiné dans la figure 3.11 le graphe $K_{3,3}$.

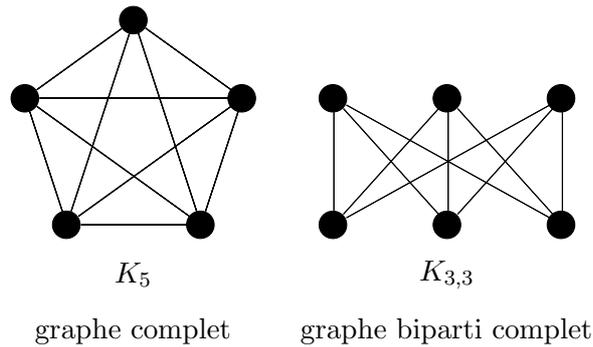


FIG. 3.11 – Graphes classiques.

3.3 Représentations en machine

Les dessins que nous avons faits dans la section précédente sont commodes pour faire des exemples et des raisonnements à la main, mais ils ne sont pas utilisables en machine. Dessiner dans le plan un graphe donné est en soi une tâche algorithmique et pratique relativement ardue. Donc nous devons trouver une autre représentation plus commode.

Le problème se réduit à représenter \mathcal{S} et $\mathcal{A} \subset \mathcal{S} \times \mathcal{S}$ en machine. Sans perte de généralité, on peut se ramener au cas où $\mathcal{S} = \{0, 1, \dots, n - 1\}$ si $n = |\mathcal{S}|$. Dans la pratique, il sera toujours possible de coder les éléments du graphe par un tel ensemble d'entiers, quitte à utiliser du hachage par exemple (voir les polys des cours précédents ou [5]). Passons au problème de la représentation de \mathcal{A} . Avec notre définition de graphe, il est clair que le cardinal de \mathcal{A} est borné par n^2 . Si $m = |\mathcal{A}|$ est proche de n^2 , on parlera de graphe *dense*. Si $m = o(n^2)$, on parlera de graphe *creux*. Il est très fréquent en pratique d'avoir à traiter des graphes creux (éléments finis, carte routière, graphe d'Internet, etc.). Les algorithmes que nous étudierons auront une complexité qui pourra bien souvent être décrite en terme de n et m , ce qui est toujours intéressant quand m

est loin de sa valeur maximale théorique $O(n^2)$.

3.3.1 Représentation par une matrice

C'est la façon la plus simple de représenter l'ensemble \mathcal{A} . On se donne donc une matrice \mathcal{M} de taille $n \times n$. Une telle matrice est appelée *matrice d'adjacence* du graphe. Il existe essentiellement deux choix pour le type de la matrice.

Le premier choix consiste à prendre une matrice de booléens tels que $\mathcal{M}_{i,j}$ est vrai si et seulement si $(i, j) \in \mathcal{A}$. Pour le graphe \mathcal{G}_2 de la figure 3.3, on numérote les sommets a, b, c, d selon la correspondance

$$\begin{array}{c|c|c|c} a & b & c & d \\ \hline 0 & 1 & 2 & 3 \end{array}$$

puis on construit la matrice booléenne (où nous avons mis T pour vrai et omis les valeurs fausses) :

i, j	0	1	2	3
0		T		
1			T	
2			T	T
3	T			

Le deuxième choix conduit à utiliser un tableau d'entiers (ou de flottants) quand les arcs de \mathcal{G} sont valués. On pose alors $\mathcal{M}_{i,j} = \nu((i, j))$ et on utilise une valeur par défaut quand l'arc (i, j) n'existe pas (par exemple une valeur plus grande que toutes les valeurs de ν). Pour le graphe \mathcal{G}_3 de la figure 3.4, on trouve :

i, j	0	1	2	3
0		1		
1			2	4
2				-3
3	5			

Dans le cas où \mathcal{G} est non orienté, la matrice \mathcal{M} est une matrice symétrique.

3.3.2 Représentation par un tableau de listes

Le stockage à l'aide de matrices prend une place proportionnelle à n^2 . Quand $m = |\mathcal{A}|$ est petit par rapport à n^2 , cela représente une perte de place parfois importante (ainsi qu'une perte de temps). D'où l'idée d'utiliser une représentation *creuse*, par exemple sous la forme d'un tableau de listes L . La liste L_i pour $i \in \mathcal{S}$, contiendra la liste des successeurs de i dans le graphe, c'est-à-dire

$$L_i = (j \in \mathcal{S}, (i, j) \in \mathcal{A}).$$

En utilisant une représentation "graphique" des listes, l'exemple du graphe \mathcal{G}_2 pourra conduire à :

$L[0] = (1)$

$L[1] = (2)$

$L[2] = (2, 3)$

$L[3] = (0)$

Il est important de noter, encore une fois, qu'une telle représentation n'est nullement canonique, car on peut décider de ranger les voisins dans n'importe quel ordre. De manière plus générale, on peut tout à fait représenter notre graphe sous forme creuse comme un tableau d'ensembles, les ensembles pouvant avoir une représentation et des propriétés variées.

La place mémoire nécessitée par une telle représentation est de $|S|$ ensembles. Dans le cas de listes chaînées, chaque ensemble demande une place $O(|L[i]|)$ (par exemple le contenu d'une cellule et d'un pointeur sur la suivante). En tout, on aura un stockage en $O(n + m)$, donc linéaire. En fonction des problèmes, on pourra trouver des représentations parfois plus adaptées.

Chapitre 4

Accessibilité

4.1 Problématique

Soit $\mathcal{G} = (\mathcal{S}, \mathcal{A})$ un graphe orienté. L'un des problèmes élémentaires que l'on peut résoudre est celui de l'accessibilité d'un sommet à partir d'un autre. Étant donné une carte ferroviaire, peut-on aller de Paris à Strasbourg? Combien de sommets intermédiaires devons-nous visiter au minimum pour se rendre de l'un à l'autre? La difficulté de ce genre d'exercice réside en ce que l'on dispose de l'information sous forme indirecte, c'est-à-dire qu'à partir de renseignements locaux (qui est voisin de qui), on doit déterminer un résultat global : existe-t-il un chemin qui m'amène d'un sommet à un autre?

4.2 Définitions ; propriétés fondamentales

Un *chemin* (s_0, s_1, \dots, s_p) dans \mathcal{G} est une suite finie non vide de sommets telle que $(s_k, s_{k+1}) \in \mathcal{A}$. La *longueur* du chemin est p , c'est le nombre d'arcs empruntés par le chemin. Notons que les arcs ne sont pas nécessairement distincts. Par exemple, la suite (c, a, b, c, e) est un chemin dans le graphe \mathcal{G}_1 de la figure 4.1.

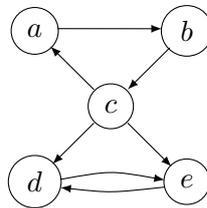


FIG. 4.1 – Le graphe \mathcal{G}_1 .

Un sommet y est un *descendant* (resp. *ascendant*) de x s'il existe un chemin de x vers y (resp. de y vers x). C'est un descendant (resp. ascendant) *propre* s'il existe un

chemin de longueur non nulle. Dans le graphe \mathcal{G}_1 , b est descendant propre de a , mais également de lui-même (par le chemin (b, c, a, b)).

Un chemin est *simple* si les arcs (s_{i-1}, s_i) pour $i = 1, \dots, p$ sont deux à deux distincts ; il est *élémentaire* si les sommets sont deux à deux distincts. Le chemin est un *circuit* si $p \geq 1$ et $s_p = s_0$; il est *élémentaire* si (s_0, \dots, s_{p-1}) est élémentaire.

Dans le graphe \mathcal{G}_1 , le chemin (a, b, c) est élémentaire ; le chemin (a, b, c, a) est un circuit élémentaire, mais pas un chemin élémentaire.

Dans le cas où \mathcal{G} n'est pas orienté, on définit une *chaîne* comme étant une suite finie de sommets (s_0, s_1, \dots, s_p) telle que deux sommets consécutifs soient extrémités d'une arête de \mathcal{A} , les arêtes étant distinctes deux à deux. Une chaîne est *élémentaire* si les sommets sont deux à deux distincts ; c'est un *cycle* si $s_0 = s_p$. Un cycle est *élémentaire* si $p \geq 1$ et (s_0, \dots, s_{p-1}) est une chaîne élémentaire.

Le résultat suivant est facile, mais mérite d'être cité.

Lemme 4.2.1 *L'ensemble des chemins d'un graphe est infini si et seulement si le graphe possède des circuits.*

Le lemme de König permet de se restreindre dans tous les cas aux chemins élémentaires. On dit que c' est un *chemin extrait* de c si c' est chemin, que c et c' ont même origine et extrémité et la suite des arcs de c' est une sous-suite de la suite des arcs de c . Pour le graphe \mathcal{G}_1 , le chemin (a, b, c) est un chemin extrait de (a, b, c, a, b, c) .

Lemme 4.2.2 (König) *De tout chemin, on peut extraire un chemin élémentaire.*

Démonstration. On effectue une récurrence sur la longueur p d'un chemin. Si $p = 0$, le chemin est élémentaire. Soit $c = (s_0, s_1, \dots, s_p)$ un chemin de longueur $p > 0$. S'il n'est pas élémentaire, il existe deux sommets égaux $s_i = s_j$ avec $0 \leq i < j \leq p$. Le chemin $c' = (s_0, \dots, s_i, s_{j+1}, \dots, s_p)$ est strictement plus petit que c et on peut en extraire un chemin élémentaire par hypothèse de récurrence, qui sera lui-même extrait de c . \square

4.3 Fermeture transitive

Cette section répond au problème déjà indiqué : existe-il un chemin entre deux sommets donnés de \mathcal{G} ? On obtient la réponse simultanément pour *toutes* les paires simultanément. Si la réponse est demandé pour un petit nombre de couples, un parcours sera sans doute plus efficace (voir chapitre 5). Par convention, un sommet sera toujours supposé accessible de lui-même.

Définition 4.3.1 *On appelle fermeture transitive du graphe \mathcal{G} le graphe $\mathcal{G}^* = (\mathcal{S}, \mathcal{A}^*)$ avec $(s, t) \in \mathcal{A}^*$ si et seulement s'il existe un chemin de s à t dans \mathcal{G} .*

Pour ce genre de problème où nous devons calculer un résultat global sur le graphe, il est logique de supposer que l'on va obtenir un graphe \mathcal{G}^* dense, donc plutôt favoriser des algorithmes basés sur la matrice d'adjacence (booléenne ou entière) associée au graphe \mathcal{G} . Pour simplifier l'exposition, nous allons supposer (dans cette section) que $\mathcal{S} = \{0, 1, 2, \dots, n-1\}$.

4.3.1 Première solution par produit de matrice

On va construire de proche en proche une suite de graphes $\mathcal{G}^{(r)}$, $1 \leq r \leq n$ de la façon suivante : $\mathcal{G}^{(1)}$ est le graphe initial ; pour $r \geq 2$, $\mathcal{G}^{(r)} = (\mathcal{S}, \mathcal{A}^{(r)})$ sera le graphe défini par $(i, j) \in \mathcal{A}^{(r)}$ si et seulement s'il existe un chemin de i à j de longueur exactement r . On notera $\mathcal{M}^{(r)}$ la matrice d'adjacence associée à $\mathcal{G}^{(r)}$.

S'il existe un chemin de longueur $r > 1$ entre i et j , il s'écrit (i, k, \dots, j) avec k voisin de i et (k, \dots, j) un chemin de longueur $r - 1$. La formule booléenne s'écrit alors :

$$\mathcal{M}_{i,j}^{(r)} = \bigvee_{k=0}^{n-1} \mathcal{M}_{i,k}^{(1)} \wedge \mathcal{M}_{k,j}^{(r-1)}.$$

On reconnaît là la formule du produit des matrices $\mathcal{M}^{(1)}$ et $\mathcal{M}^{(r-1)}$ où on a remplacé l'addition par le OU logique, et la multiplication par le ET logique. Plus généralement, si $(i, j) \in \mathcal{G}^*$, c'est qu'il existe un chemin de longueur $< n$ entre les deux (par le lemme de König).

Par la suite, on utilisera la syntaxe de l'addition et de la multiplication usuelles au lieu de OU et ET. Cela permet d'écrire :

$$\mathcal{M}^* = I_n + \mathcal{M}^{(1)} + \mathcal{M}^{(2)} + \dots + \mathcal{M}^{(n-1)}. \quad (4.1)$$

La matrice identité d'ordre n , notée I_n , a été ajoutée pour tenir compte de l'accessibilité de tout sommet avec lui-même.

Si le produit de deux matrices se fait en temps $O(n^\omega)$ (où ω est un exposant compris entre 2 et 3 – voir section 2.1.2), on voit que le calcul naïf nécessite $O(n^{1+\omega})$ opérations, soit généralement $O(n^4)$ en pratique.

Remarquons qu'on peut récrire la formule (4.1) sous la forme (en booléens) :

$$\mathcal{M}^* = (I + \mathcal{M})^{n-1}$$

ce qui fait qu'on peut calculer la puissance n -ième de la matrice en $O((\log n)n^\omega)$ par une méthode d'exponentiation binaire.

Reprenons l'exemple du graphe \mathcal{G}_1 , dont la matrice d'adjacence est :

$$\mathcal{M} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

On trouve :

$$\mathcal{M}^{(2)} = \mathcal{M}^{(1)} \cdot \mathcal{M}^{(1)} = \begin{pmatrix} & & & 1 & \\ 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{pmatrix}, \quad \mathcal{M}^{(3)} = \mathcal{M}^{(1)} \cdot \mathcal{M}^{(2)} = \begin{pmatrix} & & & & 1 \\ & & & & & 1 & 1 \\ & & & & & & 1 & 1 \\ & & & & & & & 1 & 1 \\ & & & & & & & & 1 \\ & & & & & & & & & 1 \end{pmatrix},$$

$$\mathcal{M}^{(4)} = \begin{pmatrix} & 1 & & 1 & 1 \\ & & 1 & 1 & 1 \\ 1 & & & 1 & 1 \\ & & & 1 & \\ & & & & 1 \end{pmatrix}.$$

Faisant la somme booléenne, on trouve :

$$\mathcal{M}^* = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ & & & 1 & 1 \\ & & & 1 & 1 \end{pmatrix}.$$

On peut également compter le nombre de chemins entre deux sommets x et y , en considérant la matrice d'adjacence entière.

Proposition 4.3.1 *Le nombre de chemins de longueur ℓ entre i et j est $\mathcal{M}_{i,j}^{(\ell)}$, \mathcal{M} étant considérée comme une matrice entière.*

4.3.2 L'algorithme de Roy et Warshall

Dans l'algorithme précédent, nous avons raisonné sur la longueur des chemins. Une autre idée consiste à faire augmenter le nombre de sommets possibles dans un chemin. On va construire là aussi une suite de graphes auxiliaires $\mathcal{G}^{(r)} = (\mathcal{S}, \mathcal{A}^{(r)})$, mais de la façon suivante : (i, j) est dans $\mathcal{A}^{(r-1)}$ s'il existe un chemin direct de i à j et donc ne passant par aucun sommet intermédiaire, c'est-à-dire le graphe de matrice $I_n + \mathcal{M}$ (chaque sommet étant accessible depuis lui-même). Un arc (i, j) est dans $\mathcal{A}^{(0)}$ s'il existe un chemin de i à j passant au plus par le sommet 0. Plus généralement, $(i, j) \in \mathcal{A}^{(r)}$ s'il existe un chemin de i à j ne passant par aucun sommet intermédiaire d'indice $> r$. Un tel chemin existe dans deux cas : ou bien il existait déjà un chemin de i à j ne passant pas par un sommet d'indice $> r - 1$, ou bien on peut aller de i à r dans les mêmes conditions, puis de r à j . En booléens :

$$\mathcal{M}_{i,j}^{(r)} = \mathcal{M}_{i,j}^{(r-1)} + \mathcal{M}_{i,r}^{(r-1)} \mathcal{M}_{r,j}^{(r-1)}.$$

Donnons le pseudocode pour l'algorithme :

```
RoyWarshall(M)

// M est la matrice d'adjacence booléenne n x n d'un graphe G

1. N := M; P := identite(n, n);

2. pour r <- 0 à n-1 faire
```

```

pour i <- 0 à n-1 faire
  pour j <-0 à n-1 faire
    P[i][j] <- N[i][j] OU (N[i][r] ET N[r][j]);
  N := P; // copie de P dans N
3. retourner N; // matrice d'adjacence de G*

```

La complexité de cet algorithme est clairement en $O(n^3)$. On peut remarquer que la matrice P est inutile et qu'on peut effectuer les calculs *en place*, c'est-à-dire directement dans N. Le code correspondant est alors :

```

RoyWarshall(M)

// M est la matrice d'adjacence booléenne n x n d'un graphe G

1. N := M; pour i <- 0 à n-1 faire N[i][i] <- true;

2. pour r <- 0 à n-1 faire
  pour i <- 0 à n-1 faire
    pour j <-0 à n-1 faire
      N[i][j] <- N[i][j] OU (N[i][r] ET N[r][j]);

3. retourner N; // matrice d'adjacence de G*

```

La justification de l'algorithme précédent dépend de la proposition suivante.

Proposition 4.3.2 *Pout tout r, i, j dans \mathcal{S} , les deux propriétés suivantes sont vraies :*

- (1) $(i, r) \in \mathcal{A}^{(r-1)} \iff (i, r) \in \mathcal{A}^{(r)}$;
- (2) $(r, j) \in \mathcal{A}^{(r-1)} \iff (r, j) \in \mathcal{A}^{(r)}$.

Démonstration. Traitons le point (1), le point (2) se traitant de façon similaire.

Dire que $(i, r) \in \mathcal{A}^{(r-1)}$ veut dire qu'il existe un chemin de i à r ne passant par aucun sommet intermédiaire d'indice $> r-1$; *a fortiori*, il ne passe pas par un sommet d'indice $> r$, d'où le sens direct.

Si $(i, r) \in \mathcal{A}^{(r)}$, cela veut dire qu'il existe un chemin de i à r ne passant par aucun sommet intermédiaire d'indice $> r$; en considérant un chemin élémentaire entre i et r , on voit qu'il ne peut passer par aucun sommet d'indice $> r-1$, à part le sommet r lui-même, donc $(i, r) \in \mathcal{A}^{(r-1)}$. \square

Donnons maintenant une implantation dans notre classe Graphe (voir l'annexe 15 pour la définition de la classe abstraite et des implantations) :

```

static Graphe RoyWarshall(Graphe G) {
  Graphe F = G.copie();

```

```

for(Sommet s : G.sommets())
    F.ajouterArc(s, s);
for(Sommet r : G.sommets()){
    for(Sommet s : G.sommets())
        for(Sommet t : G.sommets())
            if(!F.existeArc(s, t)
                && (F.existeArc(s, r)
                    && F.existeArc(r, t)))
                F.ajouterArc(s, t);
    }
return F;
}

```

4.4 Plus court chemin

Cette fois, on cherche à calculer pour chaque paire de sommets (i, j) le plus court chemin qui les joint, quand un chemin existe. Nous verrons plus loin comment répondre à la même question quand on cherche la distance entre un sommet donné et plusieurs sommets.

De façon générale, on peut se donner une fonction distance $d : \mathcal{A} \rightarrow \mathbb{R}_+$ vérifiant :

$$d((i, j)) = \begin{cases} > 0 & \text{si } (i, j) \in \mathcal{A}, \\ 0 & \text{si } i = j, \\ \infty & \text{sinon.} \end{cases}$$

De façon analogue au premier algorithme de recherche d'accessibilité, on peut établir une récurrence sur le nombre de voisins visités. On définit $D_{i,j}^{(k)}$ comme étant la longueur du plus court chemin de i à j ayant au plus k arcs. On a pour $k \geq 1$:

$$D_{i,j}^{(k)} = \min \left(D_{i,j}^{(k-1)}, \min_{0 \leq \ell < n} (D_{i,\ell}^{(k-1)} + D_{\ell,j}^{(k-1)}) \right),$$

avec $D_{i,j}^{(0)} = d((i, j))$. On en déduit un algorithme en $O(n^4)$, ainsi qu'un algorithme en $O((\log n)n^\omega)$ en utilisant de la multiplication rapide de matrices (dans une algèbre particulière).

L'algorithme de Floyd fonctionne comme l'algorithme d'accessibilité de Roy-Warshall. On considère $F_{i,j}^{(r)}$ comme étant la longueur du plus court chemin de i à j ne passant par aucun sommet intermédiaire d'indice $> r$. On a :

$$F_{i,j}^{(r)} = \min \left(F_{i,j}^{(r-1)}, F_{i,r}^{(r-1)} + F_{r,j}^{(r-1)} \right).$$

Le coût de l'algorithme est en $O(n^3)$.

Exercice 4.4.1 Écrire un programme qui implante cette méthode et qui stocke le plus court chemin, en mémorisant l'indice r fournissant le chemin minimum de i à j .

4.5 Arbres

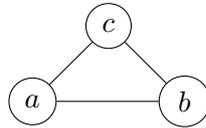
Nous allons revoir les propriétés des arbres, déjà introduits dans les cours précédents, comme cas particulier de graphes. On se place ici encore dans le cas des graphes simples (sans boucles), non orientés.

Lemme 4.5.1 *Soit \mathcal{G} un graphe simple non orienté. Si \mathcal{G} a un cycle, alors ce cycle est de longueur ≥ 3 .*

Démonstration. Comme \mathcal{G} est simple, on n'a pas droit au cycle



mais seulement à :



Le but de cette section est de donner une description géométrique des arbres :

Proposition 4.5.1 *Soit \mathcal{G} un graphe simple à n sommets. Les propriétés suivantes sont équivalentes et caractérisent le fait que \mathcal{G} est un arbre.*

- (i) \mathcal{G} est un graphe connexe sans cycle ;
- (ii) \mathcal{G} est un graphe connexe ayant $n - 1$ arêtes ;
- (iii) \mathcal{G} est un graphe sans cycle possédant $n - 1$ arêtes ;
- (iv) \mathcal{G} est un graphe dans lequel deux sommets quelconques sont liés par une seule chaîne ;
- (v) \mathcal{G} est un graphe connexe dont la connexité disparaît dès qu'on enlève une arête quelconque ;
- (vi) \mathcal{G} est un graphe sans cycle tel que l'ajout d'une arête quelconque crée un cycle et un seul.

Lemme 4.5.2 *Un graphe connexe à n sommets possède au moins $n - 1$ arêtes.*

Démonstration. On utilise une récurrence sur n . La propriété est vraie pour $n = 1$. Supposons $n \geq 2$ et soit $\mathcal{G} = (\mathcal{S}, \mathcal{A})$ un graphe connexe à n sommets. Soit u un sommet de \mathcal{G} et considérons le graphe \mathcal{G}_u induit de \mathcal{G} par $\mathcal{S} - \{u\}$. Soient C_1, C_2, \dots, C_p les composantes connexes de \mathcal{G}_u . Pour tout $k \in \{1, \dots, p\}$, il existe au moins une arête liant u à un sommet de C_k (sinon \mathcal{G} ne serait pas connexe). Soit $\mathcal{G}^{(k)}$ le sous-graphe induit par C_k . Par hypothèse de récurrence, les nombre d'arêtes et de sommets de $\mathcal{G}^{(k)}$ satisfont $m_k \geq n_k - 1$. On en déduit :

$$m \geq p + \left(\sum_{k=1}^p m_k \right) \geq \sum_{k=1}^p n_k = n - 1.$$

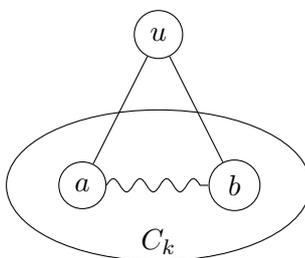
d'où le résultat. \square

Lemme 4.5.3 *Un graphe à n sommets ayant au moins n arêtes possède un cycle.*

Démonstration. On utilise là encore une récurrence. La propriété est vraie pour $n \leq 3$ (on ne suppose pas \mathcal{G} simple). Soit $n \geq 4$ et $\mathcal{G} = (\mathcal{S}, \mathcal{A})$ un graphe à n sommets et m arêtes, avec $m \geq n$. On raisonne par l'absurde en supposant \mathcal{G} sans cycle. Soit $u \in \mathcal{S}$ et \mathcal{G}_u le graphe induit par $\mathcal{S} - \{u\}$ comme dans le lemme précédent, et C_1, C_2, \dots, C_p les composantes connexes de \mathcal{G}_u , ainsi que $\mathcal{G}^{(k)}$ le sous-graphe induit par C_k . Chaque $\mathcal{G}^{(k)}$ est sans cycle, donc $m_k \leq n_k - 1$. Le nombre d'arêtes incidentes à u est

$$m - \sum_{k=1}^p m_k \geq m + p - \sum_{k=1}^p n_k \geq p + 1.$$

Il existe donc un k tel que u soit adjacent à deux sommets distincts de a et b de C_k .



Comme $\mathcal{G}^{(k)}$ est connexe, il existe une chaîne de a à b dans $\mathcal{G}^{(k)}$ et donc un cycle dans \mathcal{G} passant par u, a et b . Contradiction. \square

Démonstration de la proposition : on démontre (i) \Rightarrow (ii) \Rightarrow (iii) \Rightarrow (iv) \Rightarrow (v) \Rightarrow (vi) \Rightarrow (i).

(i) \Rightarrow (ii) : soit \mathcal{G} un graphe connexe sans cycle. D'après le lemme 4.5.2, on a $m \geq n - 1$. D'après le lemme 4.5.3, $m \leq n - 1$, d'où l'égalité.

(ii) \Rightarrow (iii) : si \mathcal{G} connexe et $n - 1$ arêtes avait un cycle, alors la suppression d'une arête du cycle entraînerait l'existence d'un graphe connexe à n sommets et $n - 2$ arêtes.

(iii) \Rightarrow (iv) : si \mathcal{G} est sans cycle et non connexe, chaque composante connexe C_k pour $k = 1..p$ est telle que $m_k = n_k - 1$. D'où $m = n - p$ et $p \geq 2$, ce qui est absurde. S'il existe deux chaînes entre deux sommets, il existe un cycle, contradiction.

(iv) \Rightarrow (v) : un graphe vérifiant (iv) est forcément connexe. Si \mathcal{G} restait connexe après disparition de l'arête $\{a, b\}$, il existerait une chaîne reliant a à b , ce qui contredirait l'unicité.

(v) \Rightarrow (vi) : un tel graphe est nécessairement sans cycle, sinon la disparition d'une arête ne suffirait pas à le déconnecter. Si on ajoute une arête $\{a, b\}$, on crée un cycle, puisqu'il existe déjà une chaîne entre a et b ; et ce cycle passerait par $\{a, b\}$. Si l'ajout créait deux cycles, c'est qu'il existait déjà deux chaînes.

(vi) \Rightarrow (i) : si l'ajout de $\{a, b\}$ crée un cycle, G est connexe, car sinon un pont entre deux composantes connexes ne créerait pas de cycle. \square

Exercices

1. [Centre d'un graphe orienté] Soit s un sommet du graphe $\mathcal{G} = (\mathcal{S}, \mathcal{A})$. L'*excentricité* de s est définie par :

$$\max_{t \in \mathcal{S}} \{\text{longueur du plus court chemin entre } s \text{ et } t\}.$$

Un *centre* du graphe est un sommet d'excentricité minimale. Donner un algorithme de recherche du centre d'un graphe.

Chapitre 5

Parcours

Dans ce chapitre, nous présentons des algorithmes fondamentaux permettant de découvrir des propriétés des graphes, orientés ou non. Une connaissance globale d'un graphe est généralement difficile (par exemple une représentation dense prendrait trop de place en mémoire), mais les algorithmes de parcours permettent de visiter les structures et d'en déduire des informations comme la connexité et la forte connexité, que nous verrons au chapitre 6. Les parcours permettent de résoudre de multiples problèmes, comme l'accessibilité de deux sommets, la recherche de cycles ou de circuits, ou encore trouver des solutions à des problèmes d'ordonnement.

5.1 Le cas non orienté

5.1.1 Définitions et propriétés générales

Nous commençons par supposer que le graphe \mathcal{G} est non orienté et connexe.

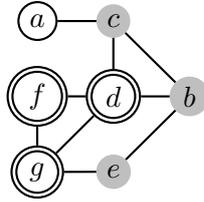
Définition 5.1.1 *Un parcours de $\mathcal{G} = (\mathcal{S}, \mathcal{A})$ à partir du sommet s est une liste de sommets L telle que :*

- le premier sommet de L est s ;
- chaque sommet de \mathcal{S} apparaît une fois et une seule dans L ;
- tout sommet de la liste (sauf le premier) est adjacent dans \mathcal{G} à au moins un sommet placé avant lui dans la liste.

À partir du sommet s , on explore une partie des sommets du graphe, les voisins de s . De proche en proche, nous sommes conduits à choisir les sommets suivants dans la bordure du graphe.

Définition 5.1.2 *La bordure $\mathcal{B}(T)$ de $T \subset \mathcal{S}$ est l'ensemble des sommets de $\mathcal{S} - T$ adjacents à T .*

Considérons par exemple le graphe \mathcal{G}_1 de la figure 5.1. La bordure de $\{d, f, g\}$ est $\{b, c, e\}$.

FIG. 5.1 – Le graphe \mathcal{G}_1 .

Définition 5.1.3 Le support $\sigma(L)$ de L est l'ensemble des sommets contenus dans L .

Par abus de notation, on notera souvent $\mathcal{B}(L) = \mathcal{B}(\sigma(L))$.

Une fois ces notations introduites, on peut esquisser un algorithme générique (non déterministe) de parcours d'un graphe \mathcal{G} .

```

parcoursGenerique(G, s)
1. L ← (s);
2. tantque B(L) est non vide faire
   2.1 choisir v dans B(L);
   2.2 L ← L # v;
  
```

À l'étape 2.1, on choisit un sommet, le choix dépendant des algorithmes. On parle souvent d'*exploration* du sommet.

Tout parcours induit une (re)numérotation des sommets de \mathcal{G} , par ordre d'apparition dans le parcours. Insistons sur le fait qu'il n'existe pas de parcours canonique d'un graphe, tout comme il n'y a pas de représentation canonique. On peut même esquisser un algorithme de numérotation associée au parcours :

```

parcoursGenerique(G, s)
1. numéro ← tableau de taille n = |S|; num ← 0;
2. L ← (s);
   numéro[s] ← num; num++;
3. tantque B(L) est non vide faire
   3.1 choisir v dans B(L);
   3.2 L ← L # v;
   3.3 numéro[v] ← num; num++;
  
```

Dans les programmes précédents, la bordure de L évolue de manière dynamique.

Proposition 5.1.1 Chaque sommet de \mathcal{G} est choisi une fois et une seule dans un parcours.

Il est difficile de parler de complexité générique, car nous n'avons pas spécifié comment gérer la bordure. Disons qu'au mieux, la complexité sera $O(n + m)$.

5.1.2 Arbre couvrant

Proposition 5.1.2 Soit $L = (x_0, \dots, x_{n-1})$ un parcours de \mathcal{G} . Pour tout $k, 1 \leq k < n$, soit x'_k un sommet de (x_0, \dots, x_{k-1}) adjacent à x_k . Le sous-graphe induit par les arêtes $(x_k, x'_k), 1 \leq k < n$, est un arbre, appelé arbre couvrant de G (relatif à L).

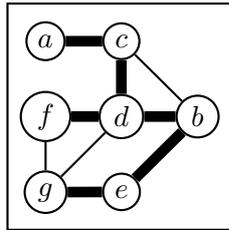
Définition 5.1.4 Les arêtes (x_k, x'_k) sont appelées arêtes de liaison.

Démonstration. On note G_p le graphe induit par $k \in \{0, \dots, p\}$ et on raisonne par récurrence sur p .

$p = 1$: seule arête (x_0, x_1) , donc $x'_1 = x_0$.

$p > 1$: G_{p-1} connexe $\Rightarrow G_p$ connexe car on a rajouté une arête avec un sommet dans G_{p-1} ; G_p a $p - 1$ arêtes et p sommets, n'a pas de cycle, donc est un arbre (en utilisant la proposition 4.5.1). \square

Ex. Un arbre couvrant du graphe \mathcal{G}_1 relatif à $L = (b, d, e, g, f, c, a)$:



Que faire quand \mathcal{G} n'est pas connexe ? On généralise à une *forêt* de recouvrement, en construisant un arbre couvrant par composante connexe.

5.1.3 Parcours en largeur d'abord

On peut programmer un parcours quelconque, en tirant au sort un sommet dans la bordure. Parmi tous les parcours possibles, deux sont très souvent employés, le *parcours en largeur d'abord*, et le *parcours en profondeur d'abord*.

L'idée de base du parcours en largeur d'abord (*breadth-first search*) consiste à parcourir les sommets d'un graphe à partir d'un sommet s donné en procédant par cercles concentriques.

Sur l'arbre binaire de la figure 5.2, en partant de la racine de l'arbre, cela revient à numéroter les sommets comme indiqué.

Dans le cas de l'arbre ci-dessus, on explore a et on stocke ses deux voisins b, c . On repart des voisins de b , à savoir d et e ; puis on prend les voisins de c , à savoir f et g .

Esquissons l'algorithme :

```
BFS(G, s)
```

1. explorer les voisins de s ;
2. explorer les voisins des voisins, etc.

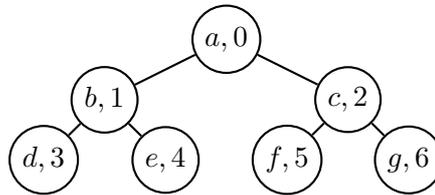


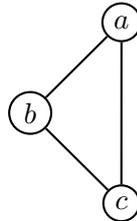
FIG. 5.2 – Parcours en largeur d’abord d’un arbre binaire.

Le point 1 est facile à réaliser, il suffit de parcourir la liste des voisins de s , qui sont les plus proches du sommet de départ. Pour aller plus loin, on va mettre les sommets dans une file d’attente, et les explorer un par un. Cela donne :

```

dfs(G, s)
1. F ← (s);
2. L ← (s);
3. tantque F n’est pas vide faire
   t ← tête(F);
   L ← L # t;
   pour u voisin de t faire
     F ← F # u;
  
```

Il est facile de voir que l’algorithme esquissé ci-dessus va boucler indéfiniment sur le triangle suivant :



Pour ne pas boucler, on va gérer l’état d’un sommet. Au début de l’algorithme, tous les sommets sont dans l’état *inexploré*. Cet état va évoluer au cours du temps. À chaque fois que l’on défile un sommet, son état devient *exploré*. À chaque fois qu’on ajoute un sommet à la file, son état devient *encours*, de sorte qu’un même sommet ne soit pas rajouté de nombreuses fois lors de l’exploration. Le programme devient alors :

```

dfs(G, s)
0. pourtout sommet t faire t.etat ← inexploré;
1. F ← (s); s.etat ← encours;
2. tantque F n’est pas vide faire
   t ← tête(F);
   pour u voisin de t faire
     F ← F # u;
  
```

```

si u.etat == inexploré alors
    u.etat <- encours;
    F <- F # u;
t.etat <- exploré;

```

Dans un parcours en largeur d'abord, le prochain sommet exploré est le plus ancien sommet de type encours.

Proposition 5.1.3 *La complexité de bfs est $O(|\mathcal{S}| + |\mathcal{A}|)$.*

Démonstration : notons d'abord que chaque sommet est ajouté dans la file au plus une fois, s'il n'a pas déjà été vu. Le nombre de passage dans la boucle 2. est donc au plus n . Récupérer la tête de F coûte $O(1)$ opérations. Pour chacun des n_t voisins de t , on teste l'état de la marque (coût en $O(1)$), et si le test est bon, on modifie l'état du sommet (coût encore en $O(1)$), puis on ajoute u à la file (encore $O(1)$). Par suite, le coût est borné par :

$$\sum_{t \in \mathcal{S}} \left(O(1) + \sum_{u \text{ voisin de } t} O(1) \right) = O(n) + O\left(\sum_{t \in \mathcal{S}} n_t\right).$$

En utilisant le fait que $\sum_{t \in \mathcal{S}} n_t = |\mathcal{A}| = m$, on en déduit le résultat. \square

Arrivés à ce point, la traduction en Java ne pose pas de problème majeur. On implante une file à l'aide de `LinkedList` et on utilise les méthodes idoines. Notons l'utilisation du marqueur présent dans la classe `Sommet`.

```

final static int inexplorable = 0, explore = 1, encours = 2;
void bfs(Sommet s){
    LinkedList<Sommet> F = new LinkedList<Sommet>();

    for(Sommet t : sommets())
        t.modifierMarque(inexplorable);
    s.modifierMarque(encours);
    F.addLast(s);
    while(! F.isEmpty()){
        Sommet t = F.removeFirst();
        for(Arc a : voisins(t)){
            Sommet u = a.d;
            if(u.valeurMarque() == inexplorable){
                u.modifierMarque(encours);
                F.addLast(u);
            }
        }
        t.modifierMarque(explore);
    }
}

```

On peut également se passer de la valeur `encours` dans beaucoup de cas, mais cela simplifie l'exposition.

Exercice 5.1.1 Montrer qu'on peut se contenter d'utiliser deux états pour un sommet, au lieu de trois.

Montrons sur un exemple le fonctionnement de l'algorithme pour le graphe dont les listes de voisins sont :

a: (g, e, b, f)

b: (c, a, d)

c: (b, e)

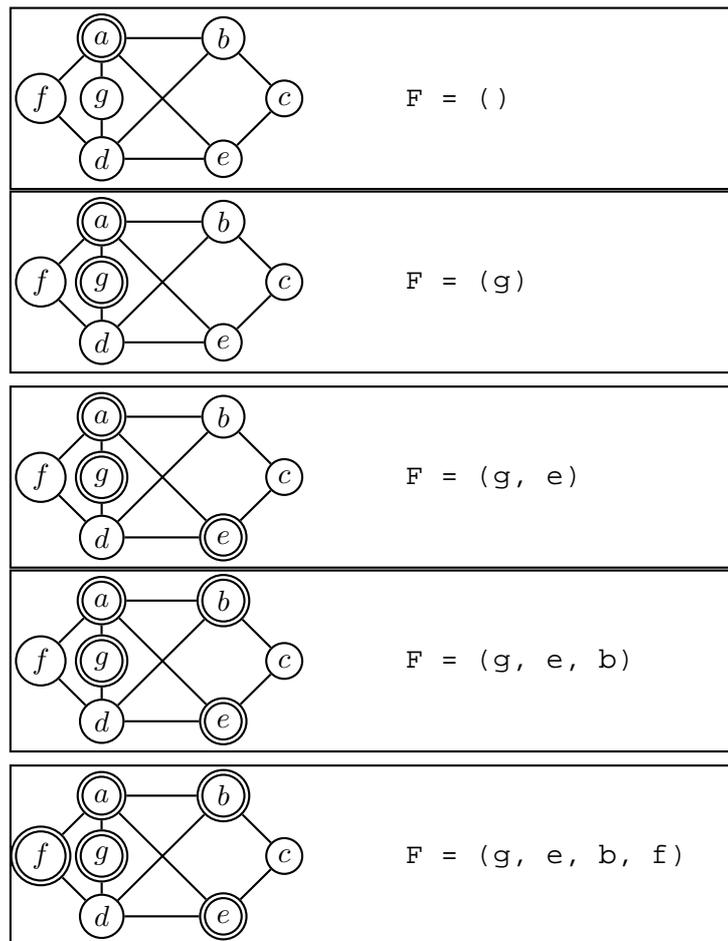
d: (b, f, e)

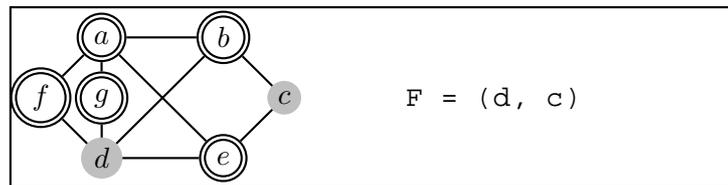
e: (c, d, a)

f: (a, d)

g: (a, d)

On indique dans les figures qui suivent l'état du graphe et l'état de la file au même moment :





En modifiant le programme pour numéroter les sommets dans l'ordre du parcours, on aurait, en utilisant une table de hachage pour les numéros, et en affichant les numéros au moment où ils sont calculés :

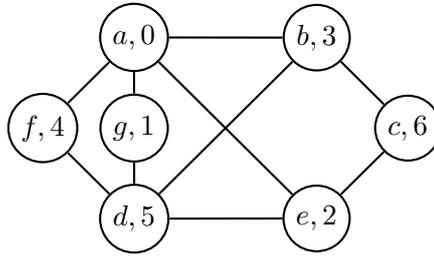
```

void bfsDistance(Sommet s){
    LinkedList<Sommet> F = new LinkedList<Sommet>();
    Hashtable<Sommet,Integer> numero =
        new Hashtable<Sommet,Integer>();
    int num = 0;

    for(Sommet t : sommets()){
        t.modifierMarque(inexplore);
        numero.put(t, 0);
    }
    s.modifierMarque(encours);
    F.addLast(s);
    numero.put(s, num++);
    while(! F.isEmpty()){
        Sommet t = F.removeFirst();
        for(Arc a : voisins(t)){
            Sommet u = a.d;
            if(u.valeurMarque() == inexplore){
                u.modifierMarque(encours);
                F.addLast(u);
                numero.put(u, num++);
                System.out.println("num["+u+"]="+numero.get(u));
            }
        }
        t.modifierMarque(explore);
    }
}

```

On trouve alors les numéros suivant :



Donnons un exemple d'application, celui où on cherche les distances entre s et les autres sommets. On va modifier le programme de sorte qu'il retourne un tableau des distances, ici une table de hachage indexée par les sommets.

```

Hashtable<Sommet,Integer> bfsDistance(Sommet s){
    LinkedList<Sommet> F = new LinkedList<Sommet>();
    Hashtable<Sommet,Integer> distance =
        new Hashtable<Sommet,Integer>();

    for(Sommet t : sommets()){
        t.modifierMarque(inexplore);
        distance.put(t, 0);
    }
    s.modifierMarque(encours);
    F.addLast(s);
    while(! F.isEmpty()){ // [1]
        Sommet t = F.removeFirst();
        for(Arc a : voisins(t)){
            Sommet u = a.d;
            if(u.valeurMarque() == inexplore){
                u.modifierMarque(encours);
                F.addLast(u);
                distance.put(u, distance.get(t) + 1);
            }
        }
        t.modifierMarque(explore);
    }
    return distance;
}

```

Pour le graphe présenté ci-dessus, on trouve :

a	b	c	d	e	f	g
0	1	2	2	1	1	1

Proposition 5.1.4 *L'algorithme bfsDistance calcule les distances minimales entre le sommet s et les autres sommets du graphe (qui sont dans sa composante connexe).*

Démonstration. Raisonnons par récurrence. À chaque entrée dans la boucle [1], les distances des sommets contenus dans la file F à s sont connues et exactes. C'est vrai à l'initialisation, puis quand on vient de visiter les voisins de s . À l'étape k , on ne rajoute dans la file que des sommets non encore visités, et qui sont donc les suivants à considérer, qui sont donc à une distance 1 de plus que le sommet t en cours d'exploration. \square

Nous verrons au chapitre 8 un autre algorithme pour résoudre le même problème, avec cette fois des distances positives quelconques pour chaque arc (algorithme de Dijkstra).

Exercice 5.1.2 Modifier l'algorithme bfs d'exploration pour détecter la présence de cycles dans un graphe.

5.1.4 Parcours en profondeur d'abord (*depth-first search*)

Dans ce cas-là, le prochain sommet visité est le plus récent sommet dont l'état a été déclaré encours. Cela revient à remplacer la file d'attente par une pile. Avec notre implémentation Java, il suffit de remplacer les `addLast` par des `addFirst` pour réaliser une recherche en profondeur :

```

void dfs(Sommet s){
    LinkedList<Sommet> F = new LinkedList<Sommet>();

    for(Sommet t : sommets())
        t.modifierMarque(inexplore);
    s.modifierMarque(encours);
    F.addFirst(s); // changement
    while(! F.isEmpty()){
        Sommet t = F.removeFirst();
        for(Arc a : voisins(t)){
            Sommet u = a.d;
            if(u.valeurMarque() == inexplore){
                u.modifierMarque(encours);
                F.addFirst(u); // changement
            }
        }
        t.modifierMarque(explore);
    }
}

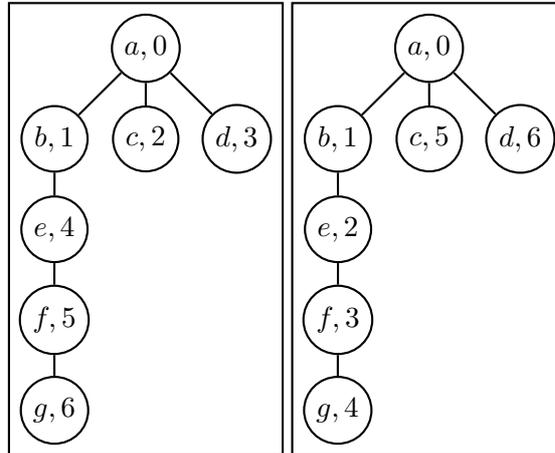
```

Le résultat suivant n'est que la copie du résultat sur bfs.

Proposition 5.1.5 *L'algorithme dfs a complexité $O(|S| + |A|)$.*

Démonstration. les opérations de pile ont le même coût que les opérations de file. \square

Comparons sur un graphe très simple les différences entre parcours en largeur et parcours en profondeur. À partir d'un sommet, le parcours en largeur va au plus près, alors que dans le parcours en profondeur, on va au plus loin :



Exercice 5.1.3 À quoi correspond une dfs sur un arbre ?

Souvent, un programme itératif utilisant une pile peut s'écrire facilement de manière récursive. C'est le cas ici pour la dfs, ce qui amène à écrire un code beaucoup plus compact.

```

void dfsRec(Sommet s){
    if(s.valeurMarque() == inexple){
        s.modifierMarque(encours);
        for(Arc a : voisins(s))
            dfsRec(a.d);
        s.modifierMarque(explre);
    }
}

```

Dans cette écriture, le rôle des états *exploré* et *en cours* est plus clairement identifiable. Un sommet est *en cours* tant que tous ses voisins et leurs descendants ne sont pas explorés.

Regardons sur un exemple (voir figure 5.3) le comportement de l'algorithme. Un sommet *encours* est doublement cerclé, un sommet *exploré* est grisé.

5.1.5 Construction de l'arbre couvrant

Nous allons plutôt construire la forêt de recouvrement, de façon à traiter le cas des graphes non connexes. Une forêt sera un ensemble d'arbres n -aires correspondant au parcours. Nous avons des arbres pour lesquels chaque nœud a *a priori* un nombre quelconque d'enfants. La classe proposée restera simple, chaque nœud de l'arbre étant une liste chaînée d'arbres :

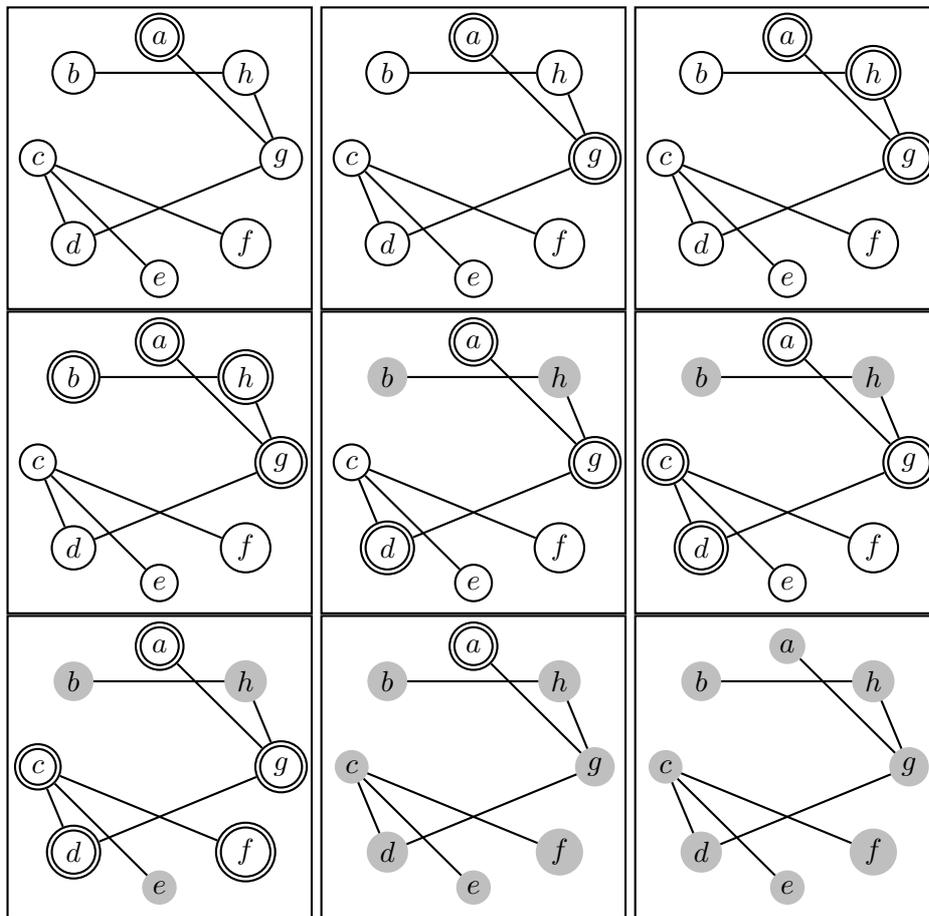


FIG. 5.3 – Exemple de déroulement de la dfs.

```

import java.io.*;
import java.util.*;

class Arbre{
    Sommet racine;
    LinkedList<Arbre> fils;

    Arbre(Sommet r){
        racine = r;
        fils = new LinkedList<Arbre>();
    }

    void ajouterFils(Arbre A){

```

```

        fils.addLast(A);
    }
}

```

La fonction ci-dessous fabrique un arbre couvrant à partir du sommet s , en ajoutant un sous-arbre par fils de s (on a modifié le programme pour ne pas retourner des arbres **null**) :

```

// on retourne un arbre couvrant de racine s
Arbre dfsRec(Sommet s){
    s.modifierMarque(encours);
    Arbre A = new Arbre(s);
    for(Arc a : voisins(s)){
        Sommet t = a.d;
        if(t.valeurMarque() == inexplorer)
            A.ajouterFils(dfsRec(t));
    }
    s.modifierMarque(explore);
    return A;
}

```

Pour stocker la forêt, on convient que l'arbre possède une racine spéciale, que nous appellerons ALPHA.

```

void dfsCouvrant(){
    Sommet ALPHA = new Sommet("ALPHA", 0);
    Arbre F = new Arbre(ALPHA);

    for(Sommet s : sommets())
        s.modifierMarque(inexplorer);
    for(Sommet s : sommets())
        if(s.valeurMarque() == inexplorer)
            F.ajouterFils(dfsRec(s));
}

```

On trouve à la figure 5.4 l'arbre couvrant correspondant au parcours de la figure 5.3.

Définition 5.1.5 Une arborescence est un arbre dont on a distingué un sommet, la racine.

Définition 5.1.6 Nous appellerons arborescence de Trémaux un arbre couvrant obtenu lors d'une dfs.

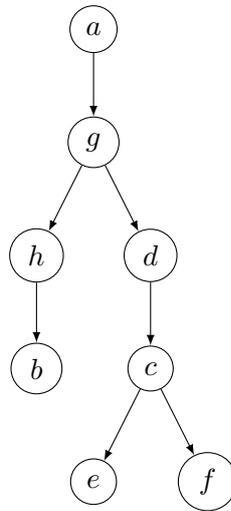


FIG. 5.4 – Arbre couvrant pour le parcours de la figure 5.3.

5.2 Parcours dans le cas orienté

5.2.1 Propriétés générales

On considère un graphe \mathcal{G} orienté sans boucle pour simplifier les définitions.

Les notions de bordure et parcours se généralisent sans problème au cas orienté.

Définition 5.2.1 La bordure $\mathcal{B}(T)$ d'une partie $T \subset \mathcal{S}$ est le sous-ensemble des sommets de $\mathcal{S} - T$ qui sont les extrémités d'un arc dont l'origine est dans T .

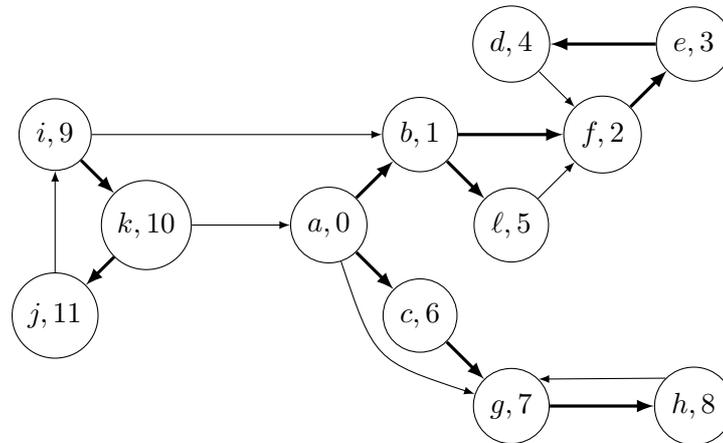
Définition 5.2.2 Un parcours de \mathcal{G} est une liste L de sommets de G telle que :

- chaque sommet de \mathcal{S} apparaît une fois et une seule dans L ;
- chaque sommet de L (sauf le premier) appartient à la bordure du sous-ensemble des sommets placés avant lui dans la liste, si toutefois celle-ci est non vide.

Les propriétés des parcours dans le cas non orienté se généralisent au cas orienté. On remplace les arêtes de liaison par les *arcs de liaison* (voir définition 5.1.4). On vérifie facilement :

Proposition 5.2.1 Les arcs de liaison d'un parcours constituent une forêt couvrante de \mathcal{G} .

Pour le parcours $L_3 = (a, b, f, e, d, \ell, c, g, h, i, k, j)$ du graphe \mathcal{G}_3 de la figure 5.5, on trouve deux arborescences indiquées en traits épais. Ces arborescences sont encore appelées arborescences de Trémaux.

FIG. 5.5 – Le graphe \mathcal{G}_3 .

5.2.2 Propriétés du parcours en profondeur d'abord

La dfs se programme comme dans le cas non orienté sans difficulté.

Exercice 5.2.1 Quelle structure de données peut-on utiliser pour répondre facilement à la question : y est-il un descendant du sommet s , dans le cas où s est fixé ?

Le rang

On peut lire de nombreuses propriétés du graphe sur un parcours L en profondeur d'abord, comme le parcours L_3 associé au graphe \mathcal{G}_3 de la figure 5.5. Notons $\text{rang}(x)$ le rang d'un sommet x dans le parcours L . Dans l'exemple, $\text{rang}(a) = 0$, $\text{rang}(b) = 1$, etc.

Il est facile de calculer le rang des sommets, au moment où on effectue le parcours en profondeur d'abord :

```

dfsRang(G)
1. rang <- tableau de taille n;
2. pourtout sommet s faire s.etat <- inexploré;
3. rg <- 0;
4. tantqu'il reste un sommet s non exploré faire
5.     rg <- dfsRec(rang, rg, s);

dfsRec(rang, rg, s)
// s est en cours
10. s.etat <- encours;
    rang[s] <- rg;
    rg <- rg + 1;
    pour t voisin de s faire
11.     si t.etat == inexploré alors

```

```

        rg <- dfsRec(rang, rg, t);
        s.etat <- exploré;
12. retourner rg.

```

On gère le tableau `rang` qui se remplit progressivement avec les rangs des sommets. On retourne systématiquement la prochaine valeur de rang à utiliser (variable `rg`).

En Java, le tableau `rang` pourra être une `Hashtable` :

```

void dfsRang(){
    Hashtable<Sommet,Integer> rang
        = new Hashtable<Sommet,Integer>();
    for(Sommet s : sommets())
        s.modifierMarque(inexplore);
    int rg = 0;
    for(Sommet s : sommets())
        if(s.valeurMarque() == inexplore)
            rg = dfsRec(rang, rg, s, s);
    for(Sommet s : sommets())
        System.out.print("rang["+s+"]="+rang.get(s));
}
int dfsRec(Hashtable<Sommet,Integer> rang, int rg,
            Sommet racine, Sommet s){
    s.modifierMarque(encours);
    rang.put(s, rg++);
    for(Arc a : voisins(s)){
        Sommet t = a.d;
        if(t.valeurMarque() == inexplore)
            rg = dfsRec(rang, rg, racine, t);
    }
    s.modifierMarque(explore);
    return rg;
}

```

Ex. On a indiqué dans la figure 5.5 les rangs des sommets pour le parcours L_3 .

Arborescence préfixe

Chaque passage dans la ligne 5 de l'algorithme 5.2.2 crée une nouvelle arborescence dont le sommet `s` sera la racine. On peut modifier les fonctions précédentes de façon à fabriquer de la même façon l'arborescence correspondante.

Les rangs des nœuds d'une arborescence de Trémaux forment un intervalle des entiers $[n, m]$. L'arborescence est *préfixe*, c'est-à-dire que pour tout sommet `s`, les rangs des descendants de `s` est un sous-intervalle de $[n, m]$ dont le plus petit élément est `rang(s)`.

Pour le parcours de la figure 5.5, on trouve que la première arborescence de Trémaux est celle de la figure 5.6.

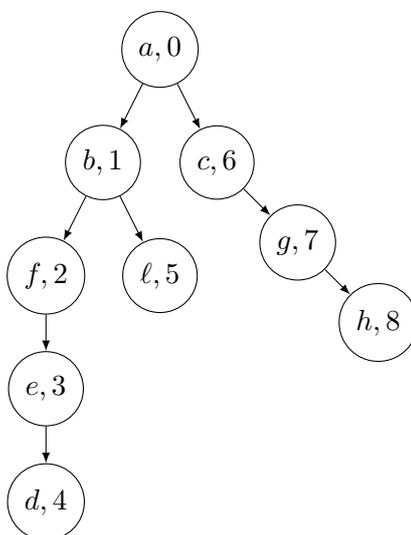


FIG. 5.6 – Première arborescence de Trémaux pour le parcours de la figure 5.5.

L'intervalle correspondant est $[0..8]$. Le premier sous-arbre de racine b correspond à $[1..5]$, subdivisé lui-même en $[2..4]$ (suivi de $[3..4]$ et $[4]$) et $[5]$; le second à $[6..8]$, puis $[7..8]$, puis $[8]$.

Les différents types d'arc

Soit \mathcal{F} la forêt couvrante induite de L . Rappelons que les arcs de \mathcal{F} sont les arcs de liaison. On peut distinguer trois autres types d'arcs dans \mathcal{G} .

Définition 5.2.3 *Un arc (s, t) est :*

- avant si s est un ascendant de t dans \mathcal{F} ;
- arrière si t est un ascendant de s dans \mathcal{F} ;
- transverse inter-arbre si s et t appartiennent à deux arborescences différentes, ou transverse intra-arbre si s et t ont un ancêtre commun z dans \mathcal{F} distinct de s et t . Si on n'a pas besoin de distinguer ces deux cas, on parlera d'arc transverse.

Dans l'exemple de la figure 5.5, l'arc (a, g) est avant ; (j, i) , (d, f) , (h, g) sont arrière ; (i, b) et (k, a) sont transverses inter-arbre (deux arborescences différentes) ; (l, f) est transverse intra-arbre (ancêtre commun).

La proposition suivante rassemble les propriétés de la fonction rang.

Proposition 5.2.2 (i) *Si t est un descendant (resp. ascendant) strict de s dans l'arborescence, alors $\text{rang}(s) < \text{rang}(t)$ (resp. $\text{rang}(s) > \text{rang}(t)$).*

(ii) *Si (s, t) est un arc arrière, alors $\text{rang}(t) < \text{rang}(s)$.*

(iii) Si (s, t) est un arc avant, alors $\text{rang}(s) < \text{rang}(t)$.

(iv) Le sommet t appartient à l'arborescence de racine s si et seulement si $\text{rang}(s) \leq \text{rang}(t) \leq \text{rang}(s) + \mathcal{D}(s)$ où $\mathcal{D}(s)$ est le nombre de descendants de s .

(v) Si (s, t) est un arc transverse, alors $\text{rang}(s) > \text{rang}(t)$. Autrement dit, s est visité après t .

Démonstration. Soient A_1, \dots, A_n les arborescences correspondant à L . Par définition des arborescences, il n'existe pas d'arc (u, v) avec $u \in A_i$ et $v \in A_j$ pour $i < j$.

(i-iv) On ne fait là que traduire les propriétés élémentaires de la fonction rang.

(v) Si (s, t) est tel que $s \in A_i$ et $t \in A_j$ avec $i \neq j$, on a nécessairement $i > j$ et donc $\text{rang}(s) > \text{rang}(t)$.

Dans le second cas, s et t sont dans la même arborescence et ont un ancêtre commun $z \notin \{s, t\}$. Si $\text{rang}(s) < \text{rang}(t)$, quand s est visité, t ne l'est pas encore et donc s sera un ascendant de t , ce qui est absurde, l'arc (s, t) serait soit de liaison, soit avant. \square

Montrons comment déduire de ces résultats un algorithme qui donne les types des arcs lors de la dfs. À l'étape 11, on déduit que (s, t) est un arc de liaison.

Si le sommet t a déjà été visité, il peut se passer différentes choses. Si t est dans l'état `encours`, cela veut dire que t est un ascendant de s , donc que (s, t) est un arc arrière.

Si t a déjà été exploré, son rang a été calculé. Soit r la racine de l'arborescence courante. Si $\text{rang}(t) < \text{rang}(r)$, c'est que t est dans une arborescence dont l'exploration est terminée, donc que (s, t) est transverse inter-arbre. Dans le cas contraire, t appartient à l'arborescence de racine r , contenant s ; t ne peut être un ascendant de s , sinon son état ne serait pas exploré, mais `encours`. Si $\text{rang}(s) < \text{rang}(t)$, t a été visité (et exploré) après s ; l'arc (s, t) est donc avant; si $\text{rang}(s) > \text{rang}(t)$, s est exploré après t , donc (s, t) est transverse intra-arbre.

Il suffit donc de garder la racine de l'arborescence courante pour avoir toutes les informations dont nous avons besoin. Nous donnons ci-dessous un programme qui écrit à l'écran les arcs visités ainsi que leur type :

Listing 5.1 – Trouver les types des arcs

```
dfsRec(rang, rg, racine, s)
// s est non exploré,
// racine est la racine de l'arborescence courante
10. s.etat <- encours;
    rang[s] <- rg;
    rg <- rg + 1;
    pour t voisin de s faire
11.     si t.etat == inexploré alors
        rg <- dfsRec(rang, rg, racine, t);
12.     sinon si t.etat == encours alors
        écrire "(s, t) arrière";
        sinon // t est déjà exploré
            si rang[t] < rang[racine] alors
```

```

        écrire "(s, t) transverse inter-arbre";
    sinon si rang[s] < rang[t] alors
        écrire "(s, t) avant";
    sinon // rang[s] > rang[t]
        écrire "(s, t) transverse intra-arbre";
    s.etat <- exploré;
13. retourner rg.

```

Exercice 5.2.2 Modifier l'algorithme 5.1 pour assigner à chaque sommet le numéro de l'arborescence dans lequel il est découvert.

Les propriétés que nous venons de démontrer, ainsi que les algorithmes décrits, formeront la base du calcul des composantes fortement connexes présenté au chapitre 6. Nous allons donner deux autres applications pour clore ce chapitre.

5.2.3 Graphes sans cycles

Proposition 5.2.3 Soit L un parcours en profondeur d'abord d'un graphe orienté \mathcal{G} . Le graphe \mathcal{G} est sans circuit si et seulement s'il n'existe pas d'arc arrière.

Démonstration. S'il existe un arc arrière, il referme un circuit, donc la condition est nécessaire.

Supposons que \mathcal{G} contienne un circuit C . Après avoir effectué la dfs correspondant à L , soit s le sommet de C de rang minimum. Soit t le prédécesseur de s sur le circuit ; t doit être un descendant de s dans une arborescence d'exploration ; donc (t, s) n'est pas transverse. Puisque $\text{rang}(t) > \text{rang}(s)$, ce ne peut être un arc de liaison. Donc (t, s) est un arc arrière. \square

Exercice 5.2.3 Montrer comment modifier l'algorithme précédent pour détecter si un graphe contient un circuit.

5.2.4 Tri topologique

Définition 5.2.4 Une liste topologique des sommets de $\mathcal{G} = (\mathcal{S}, \mathcal{A})$ est une permutation $(s_0, s_1, \dots, s_{n-1})$ des sommets de \mathcal{S} telle que pour tout arc (s_i, s_j) , on a $i < j$.

Si l'on imagine que les arcs désignent l'ordre de tâches à effectuer, autrement dit que $(s, t) \in \mathcal{A}$ si s doit être exécutée avant t , une liste topologique induit un ordre dans lequel quand chaque tâche est effectuée, alors toutes les précédentes dans le graphe l'ont été avant. La proposition qui suit montre qu'une telle liste existe s'il n'existe pas de tâches mutuellement dépendantes.

Proposition 5.2.4 Un graphe $\mathcal{G} = (\mathcal{S}, \mathcal{A})$ est sans circuit si et seulement s'il existe une liste topologique des sommets de \mathcal{G} .

Démonstration. Si \mathcal{G} possède une liste topologique, c'est qu'il existe une permutation $(s_0, s_1, \dots, s_{n-1})$ des sommets de \mathcal{S} telle que pour tout arc (s_i, s_j) , on a $i < j$. S'il existait un circuit $(s_{i_0}, s_{i_1}, \dots, s_{i_r}, s_{i_0})$, on aurait $i_0 < i_1 < \dots < i_r < i_0$, d'où une contradiction.

Démontrons la condition nécessaire par récurrence sur $n = |\mathcal{S}|$. La propriété est vraie pour $n = 1$. Supposons la vraie pour $n - 1$ avec $n \geq 2$. Soit \mathcal{G} un graphe sans circuit avec n sommets. Il existe au moins un sommet s dans descendants propres, car sinon nous aurions un circuit. Le sous-graphe \mathcal{G}' de \mathcal{G} induit par $\mathcal{S} - \{s\}$ a $n - 1$ sommets et ne possède pas de circuit. Appliquant l'hypothèse de récurrence, soit L' une liste topologique pour les sommets de \mathcal{G}' . La liste $L' \#(s)$ est une liste topologique pour les sommets de \mathcal{G} . \square

On peut modifier l'algorithme de dfs pour stocker l'ordre topologique des sommets d'un graphe sans circuit dans une liste L. Il suffit de rajouter à la liste L le sommet en fin d'exploration. Dans cet algorithme, nous n'avons en fait pas besoin du rang, ni de la racine :

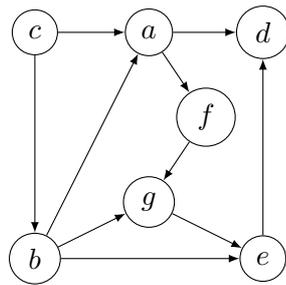
```

triTopologique(G)
1. pourtout sommet s faire s.etat <- inexploré;
2. L <- NIL;
3. tantqu'il reste un sommet s non exploré faire
   dfsTopo(L, s);

dfsTopo(L, s)
// s est non exploré
s.etat <- encours;
10. pour t voisin de s faire
    si t.etat == inexploré alors
        rg <- dfsTopo(L, t);
    sinon si t.etat == encours alors erreur;
s.etat <- exploré;
L <- s # L;

```

Pour le graphe \mathcal{G}_4 dessiné à la figure 5.7, on trouve que $L = (d, e, g, f, a, b, c)$.

FIG. 5.7 – Le graphe \mathcal{G}_4 .

Chapitre 6

Topologie

Nous présentons dans ce chapitre des algorithmes élaborés pour répondre à des questions liées à la topologie des graphes, comme la connexité, la forte connexité et la planarité. Décomposer un graphe en composantes connexes ou fortement connexes permet de réduire un problème sur un graphe à l'étude du problème sur les différents morceaux, ce qui permet souvent de gagner du temps. On en verra un exemple avec les tests de planarité. Les références utilisées pour ce chapitre sont [4, 3, 9, 2].

6.1 Forte connexité

On s'intéresse ici à un graphe \mathcal{G} orienté. Le concept de connexité n'est pas très pertinent dans ce cas-là, à moins qu'on ne s'intéresse aux propriétés du graphe sous-jacent.

6.1.1 Définition, exemples

Définition 6.1.1 *On dira qu'un graphe $\mathcal{G} = (\mathcal{S}, \mathcal{A})$ est fortement connexe si et seulement si pour tout couple $(s, t) \in \mathcal{A}$, il existe un chemin de s à t et un chemin de t à s .*

Le graphe \mathcal{G}_1 de la figure 6.1 est fortement connexe, le graphe \mathcal{G}_2 ne l'est pas.

Trouver les composantes fortement connexes d'un graphe est un problème important dans plusieurs domaines de l'informatique. Nous allons montrer comment un parcours en profondeur d'abord peut nous apporter des informations cruciales sur ces composantes.

Nous nous appuyerons sur l'exemple déjà donné au chapitre 5. Sur la figure 6.2, nous indiquons les composantes fortement connexes trouvées à la main :

Dans tout ce qui suit, on suppose qu'on a fixé un parcours L et que les sommets ont un rang calculé lors du parcours.

Notation : on notera $C(u)$ la composante fortement connexe contenant u .

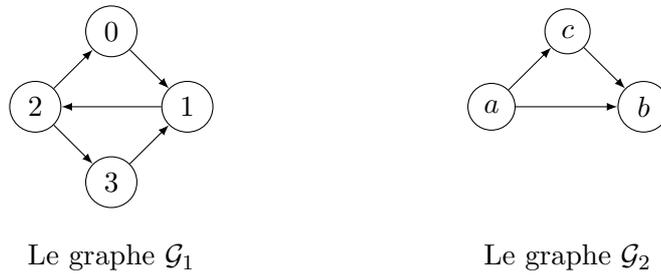
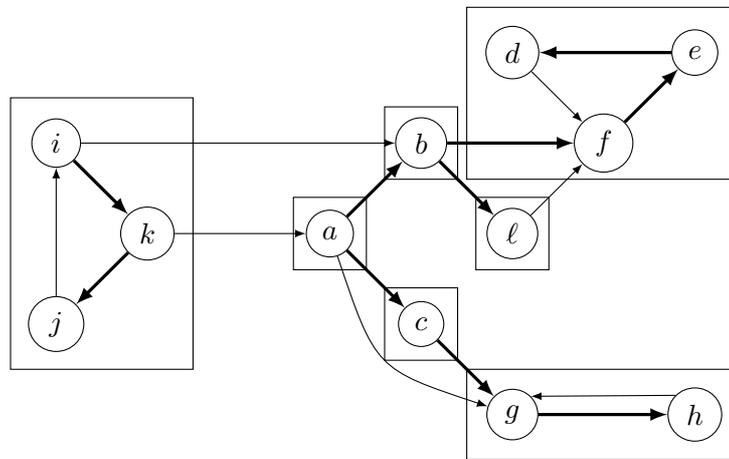


FIG. 6.1 – Exemples de graphe.

FIG. 6.2 – Le graphe \mathcal{G}_3 et ses composantes fortement connexes.

Un algorithme naïf consisterait à partir d'un sommet x et à établir la liste de ses descendants. Pour un tel sommet y , il suffirait de faire une dfs dans le graphe *renversé* de \mathcal{G} (c'est-à-dire inverser les flèches), à partir de x pour voir s'il existe un chemin vers y . Cet algorithme serait assez couteux, et nous allons donner un algorithme plus rapide, mais plus complexe et dû à Tarjan.

6.1.2 Le retour de Trémaux

Les sommets qui ne sont pas dans une arborescence de racine r ne sont pas dans $C(r)$, mais le contraire n'est pas vrai.

Définition 6.1.2 Une sous-arborescence (Y', T') de racine r' d'une arborescence (Y, T) de racine r est constituée par $Y' \subset Y$, $T' \subset T$, tel que (Y', T') forme une arborescence de racine r' .

On va montrer que la composante fortement connexe $C(u)$ est une sous-arborescence de (Y, T) . Plus précisément :

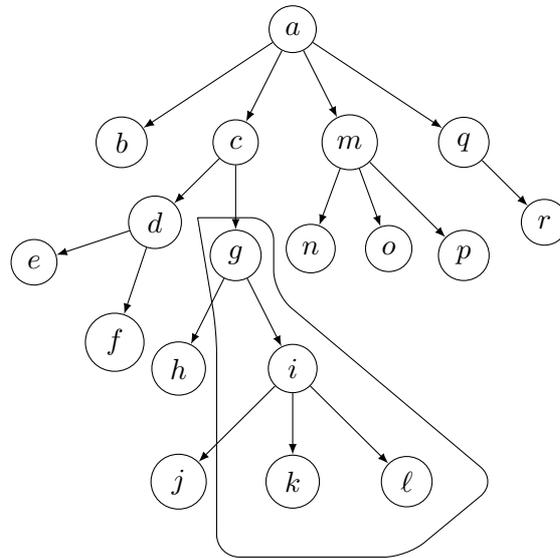


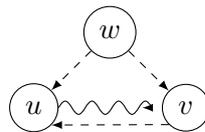
FIG. 6.3 – Exemple de sous-arborescence.

Proposition 6.1.1 Soit (Y, T) une arborescence de Trémaux de racine x . Soit u un sommet de Y et u_0 le sommet de plus petit rang dans $C(u)$. Alors :

- (i) pour tout $v \in C(u)$, tous les sommets du chemin de (Y, T) joignant u_0 à v sont dans $C(u)$;
- (ii) tous les éléments de $C(u)$ sont des descendants de u_0 dans (Y, T) .

Démonstration.

(i) Si u et v sont dans la même composante fortement connexe et w est sur un chemin de u à v , alors il existe un chemin entre u et w ainsi qu'entre w et u , donc $w \in C(u)$.



(ii) On va montrer que pour tout v dans $C(u)$, v est un descendant de u_0 dans (Y, T) . Supposons que ce ne soit pas vrai. Il existe un chemin c d'origine u_0 et extrémité v . Soit w le premier sommet du chemin qui n'est pas un descendant de u_0 , et w' le sommet précédent :

$$c = (u_0, \dots, w', w, \dots, v).$$

L'arc (w', w) n'est pas un arc de liaison (il n'est pas dans T), et n'est pas un arc avant (sinon on pourrait le remplacer par une suite d'arcs de T). Il est donc arrière ou transverse, et donc

$$\text{rang}(w) < \text{rang}(w').$$

D'après la propriété (i), w est dans $C(u)$ et par minimalité, on a $\text{rang}(u_0) < \text{rang}(w)$. Mais comme T est préfixe, cela veut justement dire que w est descendant de u_0 , contradiction. \square

6.1.3 Points d'attache

Définition 6.1.3 Soit $x \in \mathcal{G}$ racine de l'arborescence de Trémaux (Y, T) . Le point d'attache $at(y)$ de $y \in Y$ est le sommet de plus petit rang extrémité d'un chemin de \mathcal{G} , d'origine y et contenant au plus un arc (u, v) tel que $\text{rang}(u) > \text{rang}(v)$ (i.e., un arc de retour ou un arc transverse intra arbre). On suppose que le chemin vide d'origine et extrémité y est un tel chemin et donc $\text{rang}(at(y)) \leq \text{rang}(y)$. Le rang du point d'attache est appelé rang d'attache.

Un chemin de y à son point d'attache dans \mathcal{G} est soit vide (et $at(y) = y$), ou bien une suite d'arcs dans T suivis par un arc arrière ou un arc transverse. Les arcs avant ne servent à rien dans la recherche, ils peuvent être remplacés par les arcs de liaison (donc dans T). Les arcs de T partant de y conduisent à des sommets de rang supérieur à $\text{rang}(y)$.

Proposition 6.1.2 Le point d'attache $at(y)$ est le sommet de plus petit rang parmi les sommets suivants :

- y lui-même ;
- les points d'attache des fils de y dans (Y, T) ;
- les extrémités des arcs transverses (intra arbres) ou arrière d'origine y .

Nous donnons dans la figure 6.4 les rangs des sommets de \mathcal{G}_3 ainsi que leur rang d'attache (à l'extérieur des cercles).

6.1.4 Application à la forte connexité

Théorème 6.1.1 Si $u \in Y$ satisfait :

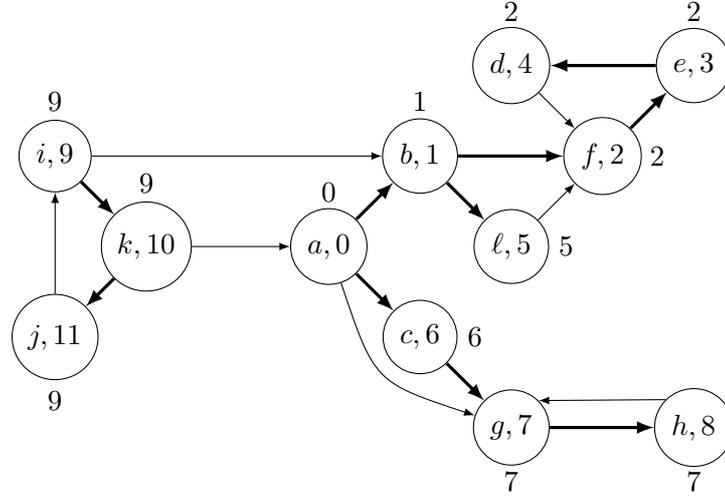
- (i) $u = at(u)$;
- (ii) Pour tout descendant v de u dans (Y, T) , on a $\text{rang}(at(v)) < \text{rang}(v)$.

Alors l'ensemble des descendants de u dans (Y, T) forme une composante fortement connexe de \mathcal{G} .

Le sommet u est appelé point d'entrée du parcours dans la composante fortement connexe.

Démonstration.

a) Soit v un descendant de u et montrons que $v \in C(u)$. Il est donc extrémité d'un chemin d'origine u . On veut montrer que u est extrémité d'un chemin d'origine v . Si ce n'est pas le cas, on peut toujours supposer que v est le sommet de plus petit rang pour lequel on ne puisse pas joindre u . Soit c_1 le chemin joignant v à $at(v)$. Soit c_2 un

FIG. 6.4 – Les rangs d'attache pour les sommets de \mathcal{G}_3 .

chemin de (Y, T) joignant u à v et $c_3 = c_1 \cup c_2 = (u, \dots, v, \dots, at(v))$. Par définition de $at(v)$, ce chemin contient au plus un arc de retour ou transverse :

$$\text{rang}(u) = \text{rang}(at(u)) \leq \text{rang}(at(v)) < v.$$

Comme (Y, T) est préfixe, on en déduit que $at(v)$ est un descendant de u . En utilisant l'hypothèse de minimalité sur v , on voit qu'il existe un chemin c_4 de $at(v)$ à u dans (Y, T) et par suite $c_2 \cup c_4$ est un chemin de v à u , contradiction.

b) Soit $w \in C(u)$. Montrons qu'il est un descendant de u , ce qui se fera à l'aide du résultat auxiliaire suivant :

Lemme 6.1.1 *Tout arc dont l'origine est dans $desc(u)$ a aussi son extrémité dans $desc(u)$.*

Démonstration du lemme :

Soit $(v_1, v_2) \in \mathcal{A}$ tel que $v_1 \in desc(u)$. Si $\text{rang}(v_2) > \text{rang}(v_1)$, v_2 est un descendant de v_1 qui est descendant de u , donc v_2 aussi. Si $\text{rang}(v_2) < \text{rang}(v_1)$, le chemin de u à v_2 contient exactement un arc arrière ou transverse, donc

$$\text{rang}(u) = \text{rang}(at(u)) \leq \text{rang}(v_2) < \text{rang}(v_1),$$

donc $v_2 \in desc(u)$ en utilisant une fois encore la préfixité de (Y, T) . \square

6.1.5 L'algorithme

L'algorithme consiste à faire un parcours en profondeur d'abord (dfs) de façon à calculer le rang et le rang d'attache dans le même temps, et d'en déduire les composantes

fortement connexes à la volée. On descend une arborescence jusqu'à trouver un point d'entrée. Une fois celui-ci trouvé, on en déduit la composante fortement connexe, qu'on retire du graphe, et on continue.

L'état d'un sommet va être modifié comme suit. Au départ, tous les sommets sont dans l'état `libre`. Quand on n'a pas encore trouvé la composante fortement connexe contenant le sommet `s`, son état est `empilé`, car nous allons stocker tous ces sommets dans la pile `pile`, qui est différente de la pile d'appels de la `dfs`. Quand on aura trouvé une composante fortement connexe et dépilé tous les sommets la constituant, ceux-ci seront dits `exclus`.

La fonction d'appel est :

```
cfc(G)
1. rang <- tableau de taille n;
   rat <- tableau de taille n; // numéro d'attache
2. pourtout sommet s faire s.etat <- libre;
3. rg <- 0;
4. pile <- NIL;
5. tantqu'il reste un sommet s libre faire
   rg <- dfsCfc(pile, rang, rg, rat, s);
```

On notera qu'on remplit des numéros d'attaches. Si on veut les points d'attache, il suffit de construire une table de sommets indexée par les rangs.

Tout se passe dans la fonction :

```
dfsCfc(pile, rang, rg, rat, s)
  s.etat <- empilé;
  rang[s] <- rg; rg <- rg+1;
  rat[s] <- rang[s]; // initialisation
  pile <- s # pile;
  pour t voisin de s faire
11.   si t.etat == libre alors
      // t est un descendant de s
      rg <- dfsCfc(pile, rang, rg, rat, t);
      rat[s] <- min(rat[s], rat[t]);
      sinon si rang[t] < rang[s] alors
      // (s, t) est arrièrè ou transverse
      si t.etat == empilé alors
      // t est dans C(s) car t descendant de s et
      // t est successeur de s
      rat[s] <- min(rat[s], rang[t]);
  si rat[s] == rang[s] alors
  // s est point d'entrèe, C(s) est dans la pile jusqu'à s
  répèter
    t <- tête(pile);
    t.etat <- exclus;
```

```

        écrire t;
        jusqu'à ce que t == s;
12. retourner rg.

```

Rappelons que faire `tête(pile)` modifie le premier élément de la pile. L'implantation en Java ne pose pas de problème. Voir la section 6.4.1.

6.1.6 L'exemple détaillé

On reprend l'exemple du graphe \mathcal{G}_3 , et on donne le déroulement de l'algorithme en détail, ce qui permet de mieux s'imprégner de son fonctionnement.

```

rang[a] <- 0;
rat[a] <- 0;
pile <- (a);
b est voisin de a
  rang[b] <- 1;
  rat[b] <- 1;
  pile = (b, a);
f est voisin de b
  rang[f] <- 2;
  rat[f] <- 2;
  pile = (f, b, a);
e est voisin de f
  rang[e] <- 3;
  rat[e] <- 3;
  pile = (e, f, b, a);
d est voisin de e
  rang[d] <- 4;
  rat[d] <- 4;
pile = (d, e, f, b, a);
      f est voisin de d, mais il est dans la pile et
      rang[f] < rang[d], donc rat[d] <- 2;
      rat[e] <- min(3, 2) = 2;
      rat[f] n'est pas modifié;
      rat[f] == rang[f] donc f est point d'entrée
      on dépile C(f) = {d, e, f};
rat[b] n'est pas modifié;
l est voisin de b
  rang[l] <- 5;
  rat[l] <- 5;
  pile = (l, b, a);
f est voisin de l, et il n'est plus dans la pile;
rat[l] == rang[l] donc l est point d'entrée
  on dépile C(l) = {l};

```

```

    rat[b] n'est pas modifié;
    rat[b] == rang[b] donc b est point d'entrée
        on dépile C(b) = {b};
c est voisin de a
    rang[c] <- 6;
    rat[c] <- 6;
    pile = (c, a);
g est voisin de c
    rang[g] <- 7;
    rat[g] <- 7;
    pile = (g, c, a);
h est voisin de g
    rang[h] <- 8;
    rat[h] <- 8;
    pile = (h, g, c, a);
    g est voisin de h, mais il est dans la pile et
        rang[g] < rang[h], donc rat[h] <- 7;
    rat[g] n'est pas modifié;
    rat[g] == rang[g] donc g est point d'entrée
        on dépile C(g) = {h, g};
    rat[c] n'est pas modifié;
    rat[c] == rang[c] donc c est point d'entrée
        on dépile C(c) = {c};
rat[a] n'est pas modifié;
g est voisin de a, mais il est exclus; // arc avant
rat[a] == rang[a] donc a est point d'entrée
    on dépile C(a) = {a};
// deuxième arborescence
rang[i] <- 9;
rat[i] <- 9;
pile = (i);
b est voisin de i mais il est exclus;
k est voisin de i
    rang[k] <- 10;
    rat[k] <- 10;
    pile = (k, i);
a est voisin de k mais il est exclus;
j est voisin de k
    rang[j] <- 11;
    rat[j] <- 11;
    pile = (j, k, i);
    i est voisin de j dans la pile
        rang[j] <- 9;
    rat[k] <- min(10, 9) = 9;

```

```

rat[i] n'est pas modifié;
rat[i] == rang[i] donc i est point d'entrée
  on dépile C(i) = {j, k, i};

```

6.2 Connexité dans les graphes non orientés

6.2.1 Énumération des composantes connexes

Soit \mathcal{G} un graphe non orienté. Montrons comment utiliser le parcours en largeur d'abord pour énumérer les composantes connexes du graphe. L'idée est de partir d'un sommet inexploré, puis de chercher tous ses voisins, et de proche en proche tous les sommets qui sont liés. Pour arriver au même but, un parcours en profondeur d'abord conviendrait également.

On peut par exemple afficher tous les membres d'une même composante. Le programme s'écrira :

```

composantesConnexes(G)
0. pourtout sommet t faire t.etat <- inexploré;
   nc <- 0;
1. pour s dans S
   si s.etat = inexploré alors
     nc <- nc+1; // numéro de composante
     afficher("Composante " + nc);
     bfs(G, s);

bfs(G, s)
1. F <- (s);
2. tantque F n'est pas vide faire
   t <- tête(F);
   t.etat <- exploré;
   afficher(t);
   pour u voisin de t faire
     si u.etat = inexploré alors
       u.etat <- encours;
       F <- F # u;

```

Il serait facile de stocker les sommets appartenant à la même composante connexe.

6.2.2 Points d'articulation

Théorie

On se donne un graphe connexe orienté \mathcal{G} . Dans certains problèmes, savoir que la connexité ne suffit pas.

Définition 6.2.1 *Un point d'articulation est un sommet dont la disparition supprime la connexité de \mathcal{G} .*

Définition 6.2.2 *Un graphe qui ne possède pas de point d'articulation est dit biconnexe.*

Définition 6.2.3 *Un sous-graphe maximal biconnexe est appelé bloc.*

On peut vouloir savoir s'il existe un sommet qui permet de séparer le graphe en deux (ou plusieurs) morceaux connexes. Sur une carte routière, un tel point sera névralgique pour le réseau et on devra donc surveiller qu'aucun accident ne le bloque. On verra à la section 6.3 que c'est la première opération à effectuer sur un graphe dont on veut tester la planarité, ce qui permettra de tester la planarité bloc par bloc.

On trouve à la figure 6.5 l'exemple du graphe \mathcal{G}_5 et ses points d'articulation indiqués par des cercles doubles. Les blocs sont également indiqués à droite du graphe.

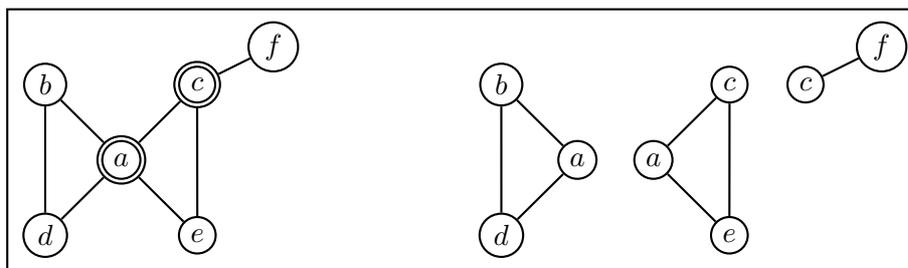


FIG. 6.5 – Le graphe \mathcal{G}_5 , ses points d'articulation et les blocs.

Une solution brutale consiste à enlever chaque sommet s à tour de rôle et à faire une bfs (ou dfs) pour tester la connexité du graphe restant. En déduire les blocs est un sous-produit de l'algorithme.

Quand le graphe est gros, c'est inefficace, et nous allons montrer qu'un algorithme très proche de celui cherchant les composantes fortement connexes permet de résoudre le problème avec une bonne complexité.

L'algorithme

Le sommet v est un point d'articulation si et seulement s'il existe deux sommets x et y tel que tout chemin de x à y passe par v . Dans ce cas là et seulement dans ce cas là, la disparition de v déconnecte \mathcal{G} .

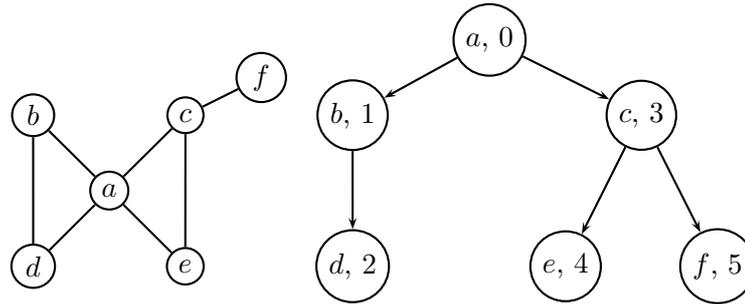
On va là encore utiliser le rang d'attache d'un sommet, comme déjà introduit dans le cas orienté.

Nous allons nous servir de l'exemple du graphe \mathcal{G}_5 pour illustrer notre propos. Conformément à l'ordre suivant pour les voisins des sommets :

a: (b, d, c, e)

b: (a, d)
 c: (a, e, f)
 d: (a, b)
 e: (a, c)
 f: (c)

on trouve :



Proposition 6.2.1 Soit T une arborescence de Trémaux associée à \mathcal{G} . Le sommet u est un point d'articulation si et seulement s'il existe deux sommets v et w tels que $(u, v) \in T$, $w \in T$, $w \neq u$, w n'est pas un descendant de v dans T , et $\text{rang}(at(v)) \geq \text{rang}(u)$.

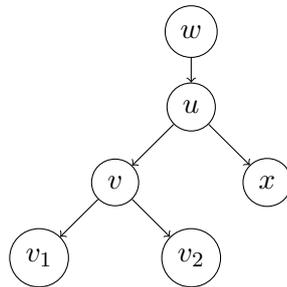
Exemple. (suite) On voit que dans notre exemple, a est point d'articulation (en utilisant $v = b$, $w = c$) ; c l'est également ($v = e$, $w = f$).

Démonstration. Supposons dans un premier temps que v et w existent. Comme $(u, v) \in T$ et $\text{rang}(at(v)) \geq \text{rang}(u)$, tout chemin d'origine v ne passant par u doit rester dans la sous-arborescence de racine v . Comme w n'est pas descendant de v dans T , un tel chemin ne peut contenir w . Les seuls chemins éventuels entre v et w doivent passer par u , donc u est point d'articulation.

Supposons maintenant que u est point d'articulation. Si u est la racine de T , il y a au moins deux arêtes qui partent de u ; s'il n'y en avait qu'une, alors il y aurait un chemin entre chaque paire de sommets de $\mathcal{S} - \{u\}$ ne contenant pas u . Soient (u, v) et (u, w) ces deux arêtes ; v et w conviennent.

Dans notre exemple, c'est le cas de a et des deux arêtes (a, b) et (a, c) dans T .

Si u n'est pas la racine de T , il a un ancêtre w . L'une des composantes biconnexes contenant u a tous ses sommets parmi les descendants de u dans T . Ces sommets sont tous des descendants d'un sommet v , avec (u, v) dans T . Ce seront les v et w demandés. Pourquoi est-ce vrai ? Tout d'abord, u a forcément un descendant, sinon ce ne serait pas un point d'articulation. Si la propriété était fautive, on aurait une situation comme :



Si x et v étaient dans la même composante bi-connexe, alors il existerait un chemin entre x et v ne passant par u , et donc x serait un descendant de v dans T . Contradiction. \square

Exemple. (suite) Dans notre exemple, pour le sommet c , on peut prendre $w = a$ et $v = f$.

L'algorithme que nous allons décrire ressemble beaucoup à celui qui cherche les composantes fortement connexes. Une différence importante est liée au traitement de la racine de l'arborescence, car nous devons compter le nombre de fils de la racine dans celle-ci et que donc on ne saura que la racine est point d'articulation que quand on aura terminé l'exploration.

```

pointsDArticulation(G)
1. rang <- tableau de taille n;
   rat <- tableau de taille n; // numéro d'attache
2. pourtout sommet s faire s.etat <- inexploré;
3. rg <- 0;
4. tantqu'il reste un sommet s inexploré faire
   // exploration composante connexe par composante connexe
   rg <- dfsPDARacine(rang, rg, rat, s);
  
```

Le code traitant la racine est relativement simple :

```

dfsPDARacine(pile, rang, rg, rat, s)
  s.etat <- encours;
  rang[s] <- rg; rg <- rg+1;
  rat[s] <- rang[s]; // initialisation
  nfils <- 0;
  pour t voisin de s faire
11.   si t.etat == inexploré alors
      // t est un descendant de s
      rg <- dfsPda(rang, rg, rat, s, t);
      nfils++;
   si nfils > 1 alors
      écrire s; // point d'articulation
   s.etat <- exploré;
12. retourner rg.
  
```

Le code pour un sommet non racine utilise le père du sommet courant. Un arc (s, p) avec p père de s n'est pas un arc intéressant dans le calcul du rang d'attache.

```

dfsPda(rang, rg, rat, pere, s)
  // pere est le père de s dans l'arborescence
  s.etat <- encours;
  rang[s] <- rg; rg <- rg+1;
  rat[s] <- rang[s]; // initialisation
  pour t voisin de s faire
11.   si t.etat == inexploré alors
      // t est un descendant de s
      rg <- dfsPda(rang, rg, rat, s, t);
      rat[s] <- min(rat[s], rat[t]);
      si rat[t] >= rang[s] alors
          écrire s; // point d'articulation
      sinon si (rang[t] < rang[s]) et (t ≠ pere) alors
          // (s, t) est arrière
          rat[s] <- min(rat[s], rang[t]);
  s.etat <- exploré;
12. retourner rg.

```

La recherche des blocs est une chose un peu plus délicate, car il faut définir la notion correcte de bloc (avec partage des sommets points d'articulation), et trouver une structure de données adéquate. Une variante particulièrement astucieuse du programme que nous venons de donner se trouve dans [6] et permet de trouver ces blocs.

Exercice 6.2.1 Exécuter l'algorithme précédent sur l'exemple de la figure 6.5.

6.3 Planarité

Tester la planarité d'un graphe est important dans plusieurs applications, comme le dessin de circuits imprimés, la réalisation de câblages. Nous allons donner quelques idées sur la façon dont on peut caractériser les graphes planaires et esquisser un algorithme de test. L'algorithme linéaire de Tarjan dépasse le cadre de ce cours et nous renvoyons à la littérature pour sa description. De nombreux résultats seront cités sans démonstration.

Dans cette section, \mathcal{G} sera un graphe *simple*, c'est-à-dire sans arêtes multiples et sans boucles.

Définition 6.3.1 *Un graphe est planaire si on peut le dessiner sans que les arcs se croisent.*

Les graphes \mathcal{G}_5 et K_4 de la figure 6.6 sont planaires (exercice), $K_{3,3}$ ne l'est pas. Dans ce domaine là également, il ne saurait y avoir de dessin canonique d'un graphe planaire, comme le montre l'exemple de K_4 .

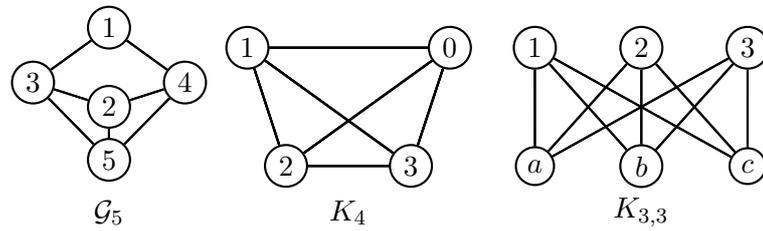


FIG. 6.6 – Des exemples de graphes planaires ou non.

6.3.1 Faces

Soit \mathcal{G} un *dessin* planaire d'un graphe planaire ayant n sommets et m arêtes. Les arêtes de \mathcal{G} découpent l'espace en *régions* ou *faces*; il existe une face spéciale, la *face infinie*.

Définition 6.3.2 Le degré d'une face est le nombre d'arêtes rencontrées dans un chemin qui suit la frontière de f .

Exemple. Le graphe de la figure 6.7 découpe l'espace en 4 faces, dont voici les degrés : $\deg(f_1) = 3$, $\deg(f_2) = 9$, $\deg(f_3) = 4$, $\deg(f_4) = 6$.

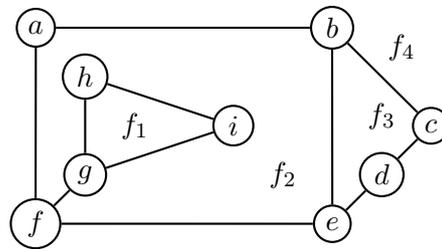


FIG. 6.7 – Un graphe planaire et les faces associées.

Exercice 6.3.1 Montrer que si \mathcal{G} est un arbre à n sommets, alors il n'y a qu'une face de degré $2n - 2$.

6.3.2 Graphe dual

Définition 6.3.3 Le graphe dual \mathcal{G}^* d'un graphe \mathcal{G} est obtenu à partir de \mathcal{G} en créant pour toute face de \mathcal{G} , un sommet de \mathcal{G}^* ; si deux faces sont contiguës, on crée une arête entre les deux sommets de \mathcal{G}^* .

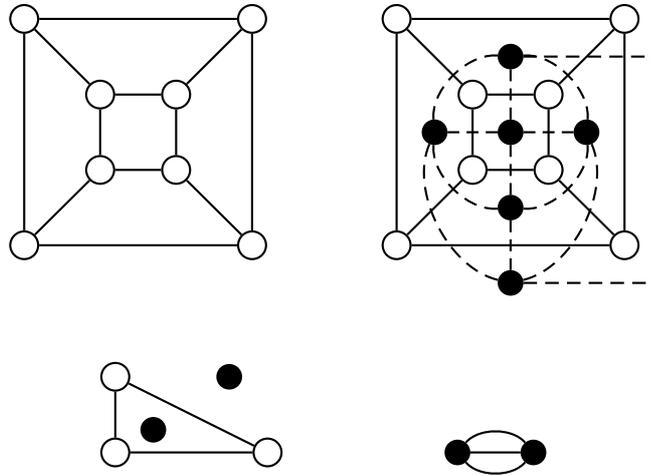


FIG. 6.8 – Graphes et leurs duals.

On donne à la figure 6.8 deux exemples de graphe avec son dual.

Le résultat suivant permet de relier les paramètres du graphe et de son dual.

Proposition 6.3.1 *Si \mathcal{G} a n sommets, m arêtes, f faces, alors \mathcal{G}^* a f sommets, m arêtes, n faces.*

On peut démontrer les deux théorèmes suivants :

Théorème 6.3.1 *Si \mathcal{G} est planaire alors \mathcal{G}^* est planaire.*

Définition 6.3.4 *On dit que deux graphes G et H sont isomorphes si et seulement si on passe de G à H par renumérotation des sommets.*

Théorème 6.3.2 *Le dual $(\mathcal{G}^*)^*$ est isomorphe à \mathcal{G} .*

Proposition 6.3.2 *Si \mathcal{G} est planaire, alors pour tout dessin planaire de \mathcal{G} , on a $\sum_f \deg(f) = 2m$.*

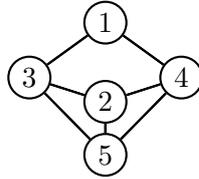
Démonstration. on applique le hand-shaking lemma (Lemme 3.2.1) au graphe dual. \square

6.3.3 Relation d'Euler

Théorème 6.3.3 *On note f le nombre de régions de \mathcal{G} . Alors*

$$f = m - n + 2.$$

Exemple. Pour le graphe



on trouve $n = 5$, $m = 7$, $f = 4$ et on a bien $4 = 7 - 5 + 2$.

Exercice 6.3.2 Montrer que le résultat est vrai pour les arbres.

Démonstration. On va raisonner par récurrence à m ou n fixé, avec l'hypothèse de récurrence :

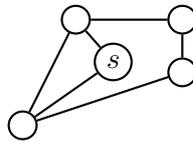
$H_{m,n}$: la relation d'Euler est vraie pour tous (i, j) , $i \leq m$ et $j \leq n$.

La propriété $H_{3,3}$ est vraie comme on peut s'en convaincre avec le dessin :



Effectuons une récurrence sur n , avec m fixé : soit \mathcal{G} un graphe connexe planaire avec m arêtes et $n + 1$ sommets. Plusieurs cas se présentent :

1. Le graphe \mathcal{G} possède un sommet intérieur s de degré $\nu \geq 1$

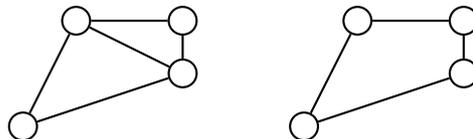


Si on enlève ce sommet et ces ν arcs ($\nu \geq 1$), le graphe restant \mathcal{G}' est toujours connexe planaire et

$$m' = m - \nu, n' = (n + 1) - 1, f' = f - (\nu - 1)$$

et donc : $f - m + (n + 1) = f' - m' + n' = 2$ par $H_{n,m}$.

2. Le graphe \mathcal{G} n'a pas de point intérieur. Supposons que \mathcal{G} contienne une corde. En l'enlevant, on crée un graphe \mathcal{G}' , qui est toujours connexe planaire et qui a une arête de moins, mais également une face de moins :



$$m' = m - 1, n' = n + 1, f' = f - 1.$$

On itère ce raisonnement tant qu'il reste une corde dans \mathcal{G} . On s'est donc ramené au cas où \mathcal{G} ne contient que des points sur le périmètre (et donc de chaque sommet partent

deux arêtes). Dans ce cas, on enlève un point quelconque en fusionnant les deux arêtes, ce qui donne \mathcal{G}' avec

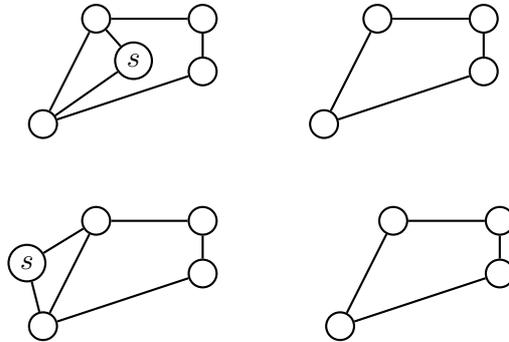
$$m' = m - 1, n' = n + 1 - 1 = n, f' = f$$

et on utilise $H_{n,m}$.

Effectuons maintenant une récurrence sur m , avec n fixé : soit \mathcal{G} un graphe planaire avec $m + 1$ arêtes et n sommets.

Soit s un sommet quelconque et $\nu \geq 1$ son degré. Le fait de l'enlever donne un nouveau graphe \mathcal{G}' qui est connexe planaire pour lequel :

$$m' = m - \nu, n' = n - 1, f' = f - (\nu - 1).$$



Ceci achève la démonstration. \square

Donnons quelques applications de la formule d'Euler.

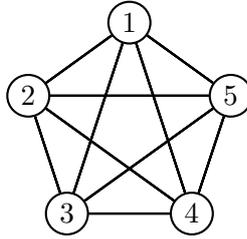
Corollaire 6.3.1 *Si \mathcal{G} est planaire avec $n \geq 3$, alors $m \leq 3n - 6$.*

Démonstration. Le degré de chaque face est au moins 3, d'où $3f \leq 2m$ et on utilise Euler. \square

Ce résultat est très important. Il montre en particulier qu'un graphe planaire ne peut pas avoir trop d'arêtes, et donc qu'en particulier $|\mathcal{A}| = O(|\mathcal{S}|)$, ce qui montre que le paramètre pertinent est n plutôt que m et laisse la place pour des algorithmes linéaires en le nombre de sommets. En outre, cela permet de faire des raisonnements évitant d'avoir à introduire le nombre de faces, donc de faire des dessins qui peuvent se révéler faux.

Corollaire 6.3.2 *Le graphe complet K_5 n'est pas planaire.*

Démonstration.



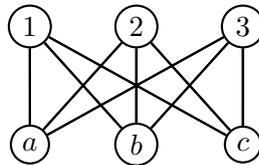
On trouve que $n = 5$, $m = 10$, mais $10 > 3 \times 5 - 6$. \square

Corollaire 6.3.3 *Si \mathcal{G} a $n \geq 3$ et n'a pas de triangle, alors $m \leq 2n - 4$.*

Démonstration. chaque face a degré au moins 4, donc $4f \leq 2m$ d'où le résultat. \square

Corollaire 6.3.4 *Le graphe $K_{3,3}$ n'est pas planaire.*

Démonstration.



On calcule $n = 6$, $m = 9$; le graphe n'a pas de triangle mais pourtant : $9 > (2 \times 6) - 4$. \square

6.3.4 Un critère théorique de planarité

Définition 6.3.5 *Deux graphes sont homéomorphes ssi on peut passer de l'un à l'autre par fusion ou scission d'arcs passant par un nœud de degré 2.*

Exemple. On part du graphe (a) de la figure 6.9 et supprime le sommet 1. On obtient le graphe (b) qui lui est homéomorphe.

L'un des théorèmes principaux du domaine est le suivant.

Théorème 6.3.4 (Kuratowski, 1930) *Un graphe \mathcal{G} est planaire si et seulement s'il ne contient aucun sous-graphe homéomorphe à K_5 ou $K_{3,3}$.*

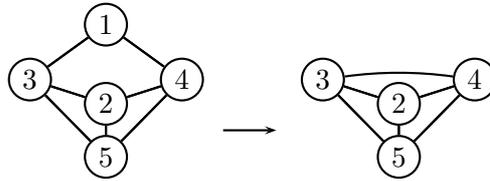
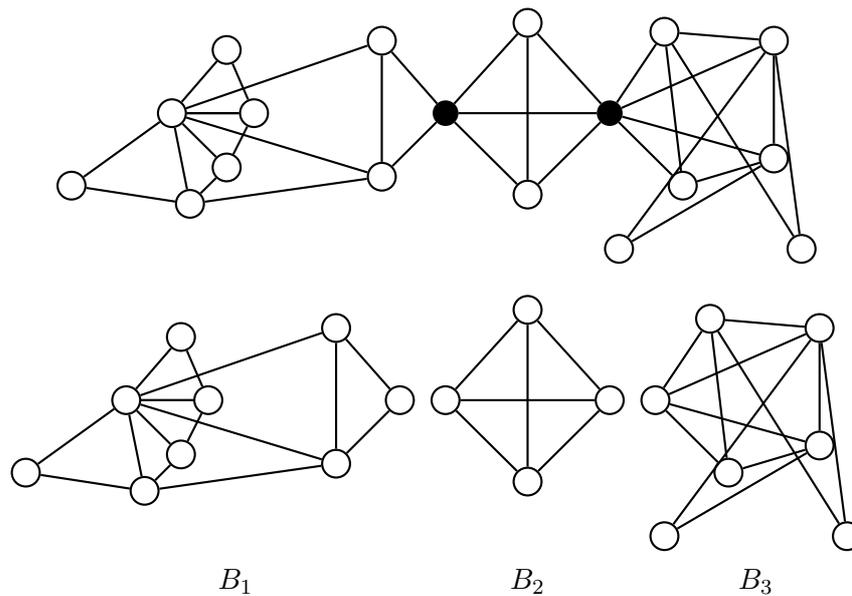


FIG. 6.9 – Deux graphes homéomorphes.

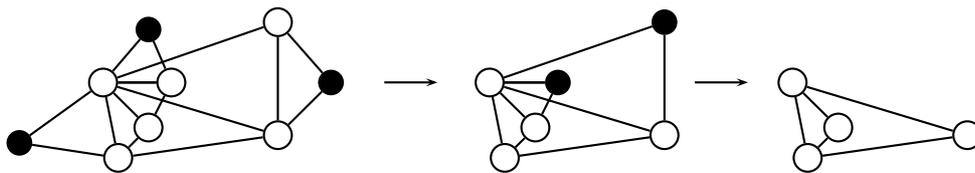
6.3.5 Vers un algorithme de test de planarité

Expliquons sur un exemple la manière de procéder pour se ramener à des cas plus simples. La première chose à faire pour déterminer si un graphe est planaire est d'abord de travailler composante connexe par composante connexe. Ensuite, on remarque qu'un graphe connexe est planaire si et seulement si chacun de ses blocs l'est. On les calcule donc par l'algorithme de la section 6.2.2.

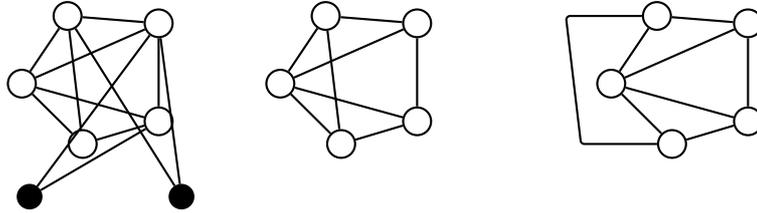
Considérons l'exemple du graphe suivant, que l'on coupe en trois blocs B_1 , B_2 , B_3 à partir des points d'articulation :



Le graphe B_2 est planaire par inspection (ou en remarquant que tout graphe avec $n < 5$ est planaire). Pour B_1 , on agit par homéomorphisme, en enlevant les sommets de degré 2 (les sommets en noir) tour à tour comme indiqué sur le dessin.



Il reste à traiter B_3 , pour lequel on procède par homéomorphisme, puis inspection :



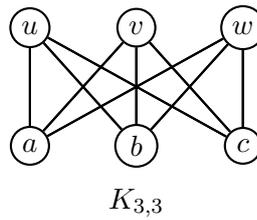
On peut aussi montrer que $m < 9$ implique \mathcal{G} planaire.

6.3.6 Une idée d'algorithme

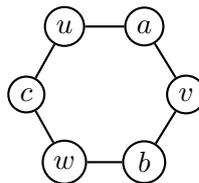
On suppose dans un premier temps que le graphe \mathcal{G} est hamiltonien : ou bien c'est un arbre (donc planaire), ou bien ses sommets peuvent être mis sur un polygone fermé (un cercle).

L'idée est de répartir les arêtes à l'extérieur ou à l'intérieur du cycle. Si on y arrive sans que les arêtes ne se coupent, \mathcal{G} est planaire, sinon, il ne l'est pas.

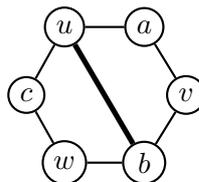
De nouveau, considérons l'exemple de $K_{3,3}$ et appliquons cette démarche.



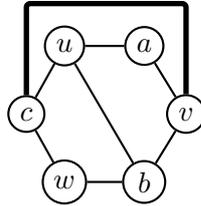
On sélectionne le circuit hamiltonien **uavbwc**. S'il existe un dessin planaire, alors ce circuit formera nécessairement un polygone :



Les arêtes non encore utilisées sont **ub**, **vc**, **wa**. On peut dessiner **ub** sans problème :

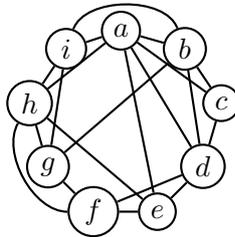


On peut dessiner l'arête **vc** en passant à l'extérieur du polygone :

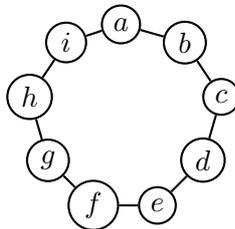


Il ne nous reste plus que l'arête **wa** à placer, mais on ne peut le faire sans couper les arêtes déjà présentes. On peut montrer que la conclusion reste valide en permutant ces trois arêtes entre elles, donc $K_{3,3}$ n'est pas planaire.

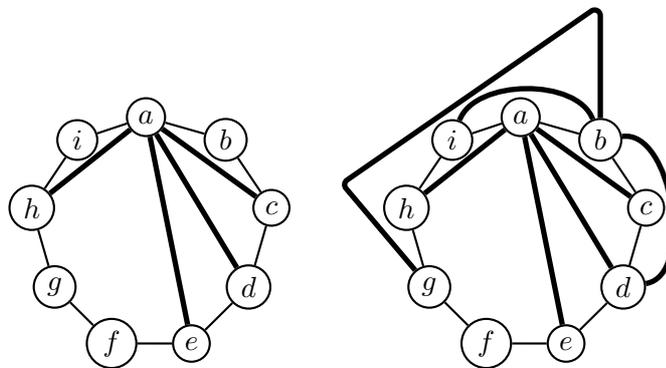
Donnons un deuxième exemple.



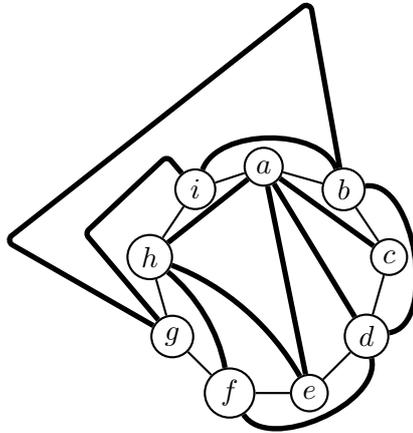
On choisit le cercle :



Il reste à placer les arêtes **ac, ad, ae, ah, bd, bg, bi, df, eh, fh, gi**. On place d'abord **ac, ad, ae, ah** à l'intérieur, puis **bd, bg, bi** à l'extérieur, conformément au dessin :



Finalement, on place **df**, **eh**, **fh**, **gi** :



et le graphe est donc bien planaire.

Il est clair que nous n'avons fait qu'esquisser l'algorithme. Nous renvoyons à l'algorithme de Tarjan qui fonctionne en temps linéaire en $O(|S|)$.

6.4 Les programmes en Java

6.4.1 Composantes fortement connexes

Nous ajoutons de nouvelles valeurs pour les marques.

```

final static int libre = 0, empile = 1, exclus = 2;

int dfsCfc(LinkedList<Sommet> pile,
            Hashtable<Sommet,Integer> rang, int rg,
            Hashtable<Sommet,Integer> rat, Sommet s){
    s.modifierMarque(empile);
    rang.put(s, rg++);
    int rats = rang.get(s);
    pile.addFirst(s);
    for(Arc a : voisins(s)){
        Sommet t = a.d;
        if(t.valeurMarque() == libre){
            rg = dfsCfc(pile, rang, rg, rat, t);
            rats = Math.min(rats, rat.get(t));
        }
        else if(rang.get(t) < rang.get(s)){
            if(t.valeurMarque() == empile)
                // t est dans C(s) car t descendant de s
                // et s -> t

```

```

        rats = Math.min(rats, rang.get(t));
    }
}
rat.put(s, rats);
if(rats == rang.get(s)){
    // s est un point d'entrée
    Sommet t;
    System.out.print("CFC={");
    do{
        t = pile.removeFirst();
        t.modifierMarque(exclus);
        System.out.print(" "+t);
    } while(! t.equals(s));
    System.out.println(" }");
}
return rg;
}

void cfc(){
    Hashtable<Sommet,Integer> rang =
        new Hashtable<Sommet,Integer>();
    Hashtable<Sommet,Integer> rat =
        new Hashtable<Sommet,Integer>();

    for(Sommet s : sommets())
        s.modifierMarque(libre);
    int rg = 0;
    LinkedList<Sommet> pile = new LinkedList<Sommet>();
    for(Sommet s : sommets())
        if(s.valeurMarque() == libre)
            rg = dfsCfc(pile, rang, rg, rat, s);
}

```

6.4.2 Composantes connexes

Ce qui peut donner en Java

```

void composantesConnexes(){
    for(Sommet s : sommets())
        s.modifierMarque(inexplore);
    int ncc = 0;
    for(Sommet s : sommets())
        if(s.valeurMarque() == inexplore){
            ncc++;
        }
}

```

```

        System.out.println("Composante "+ncc);
        bfs(s);
    }
}
void bfs(Sommet s){
    LinkedList<Sommet> f = new LinkedList<Sommet>();

    s.modifierMarque(encours);
    f.addLast(s);
    while(! f.isEmpty()){
        Sommet t = (Sommet)f.remove();
        t.modifierMarque(explore);
        System.out.println("J'explore "+t);
        for(Arc a : voisins(t)){
            Sommet u = a.d;
            if(u.valeurMarque() == inexplorer){
                u.modifierMarque(encours);
                f.addLast(u);
            }
        }
    }
}

```

6.4.3 Points d'articulation

```

// pere -> s
int dfsPda(Hashtable<Sommet,Integer> rang, int rg,
           Hashtable<Sommet,Integer> rat,
           Sommet pere, Sommet s){
    s.modifierMarque(encours);
    System.out.println("J'explore "+s);
    rang.put(s, rg++);
    int rats = rang.get(s);
    for(Arc a : voisins(s)){
        Sommet t = a.d;
        if(t.valeurMarque() == inexplorer){
            pile.addFirst(a);
            rg = dfsPda(rang, rg, rat, s, t);
            rats = Math.min(rats, rat.get(t));
            if(rat.get(t) >= rang.get(s)){
                System.out.print(s+" est point");
                System.out.println(" d'articulation");
            }
        }
    }
}

```

```

    }
    else if((rang.get(t) < rang.get(s))
            && !t.equals(pere)){
        System.out.print("(" + s + ", " + t + ") arrière");
        System.out.println(" ou transverse");
        rats = Math.min(rats, rang.get(t));
    }
}
rat.put(s, rats);
s.modifierMarque(explore);
return rg;
}

// s est la racine de l'arborescence
int dfsPDARacine(Hashtable<Sommet,Integer> rang, int rg,
                 Hashtable<Sommet,Integer> rat, Sommet s){
    s.modifierMarque(encours);
    rang.put(s, rg++);
    int rats = rang.get(s);
    int nfils = 0;
    for(Arc a : voisins(s)){
        Sommet t = a.d;
        if(t.valeurMarque() == inexplorer){
            rg = dfsPda(rang, rg, rat, s, t);
            nfils++;
        }
    }
    if(nfils > 1){
        System.out.print(s + " est racine");
        System.out.println(" ET point d'articulation");
    }
    rat.put(s, rats);
    System.out.print("rang[" + s + "]=" + rang.get(s));
    System.out.println(" rat[" + s + "]=" + rat.get(s));
    s.modifierMarque(explore);
    return rg;
}

void pointsDArticulation(){
    Hashtable<Sommet,Integer> rang =
        new Hashtable<Sommet,Integer>();
    Hashtable<Sommet,Integer> rat =
        new Hashtable<Sommet,Integer>();

```

```
for(Sommet s : sommets())
    s.modifierMarque(inexplore);
int rg = 0;
for(Sommet s : sommets())
    if(s.valeurMarque() == inexplore)
        // nouvelle composante connexe
        rg = dfsPDARacine(rang, rg, rat, s);
}
```

Chapitre 7

Euler et Hamilton

Dans ce chapitre, nous introduisons deux concepts utilisés couramment, ceux de graphe eulérien et graphe hamiltonien, notions qui prédatent l'utilisation des graphes en informatique, mais qui apparaissent comme les actes fondateurs de la théorie.

Les graphes considérés seront *a priori* non orientés et connexes, même si les notions s'étendent facilement au cas non orienté.

7.1 Euler : fondateur de la théorie des graphes

Le problème considéré à l'époque est celui des ponts de Koenigsberg, un plan d'une ville dont on se demandait s'il était possible de parcourir tous les ponts une fois et une seule au cours d'une promenade :



FIG. 7.1 – Ponts de Königsberg (Kaliningrad)

Définition 7.1.1 *On appelle cycle Eulérien un cycle qui passe par toutes les arêtes du graphe une fois et seule (mais peut-être plusieurs fois par le même sommet). Un graphe eulérien est un graphe possédant un cycle eulérien.*

Ex. Pour le graphe de la figure 7.2, **abcdefbgcegfa** ou **afgcedegbcefba** sont des cycles eulériens.

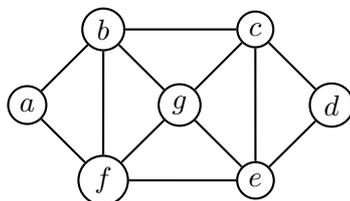


FIG. 7.2 – Le graphe \mathcal{G} .

Nous pouvons maintenant énoncer le résultat principal.

Théorème 7.1.1 (Euler, 1736) *Un graphe \mathcal{G} possède un cycle Eulérien si et seulement si chaque sommet est de degré pair.*

Corollaire 7.1.1 *Il n'existe pas de chemin passant une fois et une seule par les 7 ponts de Königsberg.*

Nous commençons par le résultat suivant :

Proposition 7.1.1 *Si tous les sommets sont de degré pair, alors \mathcal{G} peut être découpé en cycles n'ayant pas d'arête en commun.*

Démonstration. On part d'un sommet arbitraire u , et quand on arrive à un sommet, on sait qu'on peut ressortir par une autre arête. Comme \mathcal{G} est fini, on finit par rencontrer un sommet v qu'on a déjà rencontré, ce qui forme le cycle C_1 . Le graphe $H = \mathcal{G} - C_1$ possède la même propriété (composante connexe par composante connexe), etc. \square

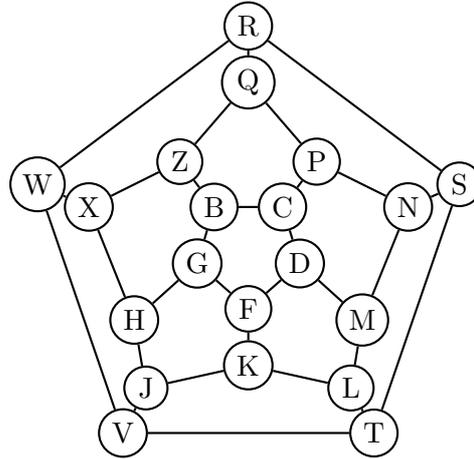
Démonstration du théorème d'Euler. (a) Si \mathcal{G} est eulérien, alors tout sommet est de degré pair. En effet, chaque sommet visité par le chemin apporte une contribution de 2 au degré du sommet.

(b) Montrons que si tout sommet est de degré pair, alors \mathcal{G} est eulérien. D'après la proposition, on peut découper \mathcal{G} en cycles disjoints C_1, C_2, \dots, C_k . On parcourt C_1 jusqu'à trouver un sommet de C_2 , que l'on parcourt. Si on n'a pas fini, on recommence le parcours de $C_1 \cup C_2$ jusqu'à trouver un sommet d'un cycle C_3 , etc. \square

7.2 Hamilton (1805–1865)

Définition 7.2.1 *Un cycle Hamiltonien est un cycle qui passe par tous les sommets une fois et seule. Un graphe hamiltonien est un graphe possédant un cycle hamiltonien.*

Ex. **abcdegfa** ou **afedcgba** sont des cycles hamiltoniens pour le graphe de la figure 7.2. Pour le graphe ci-dessous, on trouve que **BCPNMDFKLT SRQZXWVJHGB** et **BCPNMDFGHXWVJKLTSRQZB** sont deux cycles hamiltoniens.



Contrairement au cas eulérien, il n’y a pas de critère simple d’existence de cycle hamiltonien. On dispose de quelques résultats théoriques.

Proposition 7.2.1 *Le graphe K_n est hamiltonien, puisque tous les sommets sont reliés entre eux.*

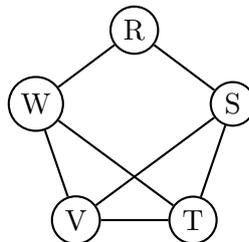
Intuitivement, il sera plus facile de trouver un chemin Hamiltonien si le graphe est dense, ce que tend à souligner le résultat qui suit.

Théorème 7.2.1 (Ore, 1960) *Si $n \geq 3$, et*

$$\deg(v) + \deg(w) \geq n$$

pour tout $(v, w) \notin \mathcal{A}$, alors \mathcal{G} est Hamiltonien.

Ex. Ce résultat permet de prouver que le graphe suivant est hamiltonien.



Exercices complémentaires

1. (Le graphe de de Bruijn) Le problème est le suivant : est-il possible de construire une suite circulaire de bits telle que tous les nombres de n chiffres binaires apparaissent par simple décalage d'un bit ?

a) On considère le cas $n = 3$. On construit le graphe dont les sommets sont les suites de 2 bits. Les arcs sortant du sommet s sont étiquetés par sb avec $b \in \{0, 1\}$. De même, les arcs entrant sont du type bs avec $b \in \{0, 1\}$. Dessiner le graphe correspondant.

Montrer qu'un circuit eulérien dans ce graphe donne la réponse, en sélectionnant la première lettre de chaque étiquette dans l'ordre du circuit, par exemple la suite circulaire 00010111.

b) Généraliser à n quelconque.

c) Généraliser à n chiffres sur un alphabet à m lettres.

Chapitre 8

Introduction à l'optimisation combinatoire

Les graphes permettent de modéliser de nombreux problèmes de nature combinatoire, comme trouver des chemins de coût minimaux, ou bien encore jouer à des jeux comme les échecs.

8.1 L'algorithme de Dijkstra

L'algorithme de Floyd permet de calculer les distances minimales entre tous les sommets avec une complexité en $O(n^3)$. Dans certains cas, on a besoin d'une information partielle, à savoir, la distance minimale entre deux sommets s et s' . Il se trouve qu'il est tout aussi facile de trouver les distances entre s et tous les autres sommets. C'est une approche dite *gloutonne*.

8.1.1 Théorie

On se donne un graphe valué avec

$$d(u, v) = \begin{cases} 0 & \text{si } u = v; \\ d(u, v) > 0 & \text{si } u \neq v; \\ +\infty & \text{si } (u, v) \notin \mathcal{A}. \end{cases}$$

Les principes de l'algorithme sont les suivants :

- L'algorithme construit un ensemble $T \subset S$ tel que pour tout $t \in T$, tout chemin minimal de s à t ne passe que par des éléments de T . L'ensemble T grossit incrémentalement de \emptyset à S .
- À chaque sommet $t \in S$, on associe une étiquette $L(t)$, initialisée à $d(s, t)$; à la fin de l'algorithme, cette valeur contient la valeur du plus petit chemin de s à t .

Donnons l'algorithme en pseudocode :

```
1. [Initialisation] pour tout  $t \neq s$  faire  $L(t) \leftarrow d(s, t)$ ;  
2.  $L(s) \leftarrow 0$ ;  $T \leftarrow \{s\}$ ;
```

```

3. tantque  $T \neq S$ 
4.   trouver  $v \notin T$  telque  $L(v) = \min \{L(t), t \notin T\}$ ;
5.    $T \leftarrow T \cup \{v\}$ ;
6.   pourtout  $t \notin T$  faire
7.      $L(t) \leftarrow \min (L(t), L(v) + d(v, t))$ ;
    
```

La validité de l'algorithme provient de la proposition suivante.

Proposition 8.1.1 À l'entrée dans la boucle 3, on a :

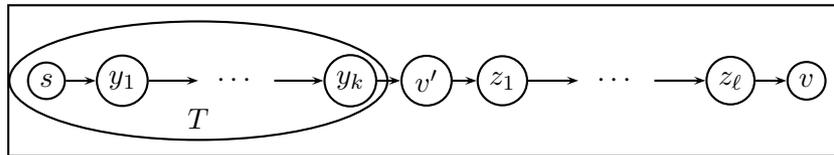
- (a) pour tout $t \in T$, $L(t)$ est la longueur d'un chemin minimal de s à t dans \mathcal{G} ;
- (b) pour tout $v \notin T$, $L(v)$ est la longueur d'un chemin minimal de s à v ne passant par aucun sommet de $S - T$ (en dehors de v).

Preuve. On procède par récurrence.

Pour $|T| = 1$, c'est-à-dire $T = \{s\}$: (a) et (b) sont vraies par construction (cf. ligne 2.).

Supposons donc $|T| > 1$. Juste avant la ligne 5, d'après l'hypothèse de récurrence, $L(v)$ est la distance d'un chemin minimal de s à v ne passant que par des sommets de T .

Raisonnons par l'absurde et supposons que quand on ajoute v , ce n'est pas la longueur d'un chemin minimal de s à v dans \mathcal{G} . Si c'est le cas, un plus court chemin de s à v doit contenir au moins un autre sommet qui ne soit pas dans T . Soit v' le premier que l'on rencontre en partant de s .



Soit c ce plus court chemin $c = (s, y_1, \dots, y_k, v', z_1, \dots, z_\ell, v)$; pour tout i , $y_i \in T$. Soit c' le chemin extrait $c' = (s, y_1, \dots, y_k, v')$. C'est un plus court chemin entre s et v' , car sinon, on aurait un chemin encore plus court de s à v . Par construction, la longueur δ du chemin c' vérifie $\delta < L(v)$.

Par hypothèse de récurrence sur (b), $\delta = L(v')$, et donc $L(v') < L(v)$, ce qui contredit la propriété à la ligne 5.

La propriété (a) est bien encore valide; (b) s'en déduit grâce aux lignes 6-7. \square

Ex. Considérons le graphe \mathcal{G}_1 de la figure 8.1. Nous allons calculer les distances minimales entre 0 et les autres sommets.

Au départ de l'algorithme, les valeurs de la fonction L sont données par :

t	0	1	2	3	4
$L(t)$	0	$+\infty$	3	1	6

L'ensemble T contient d'abord $\{0\}$.

On cherche alors le sommet v d'étiquette $L(v)$ minimale parmi les sommets non encore dans T . Ici, c'est le sommet 3, ce qui conduit au dessin

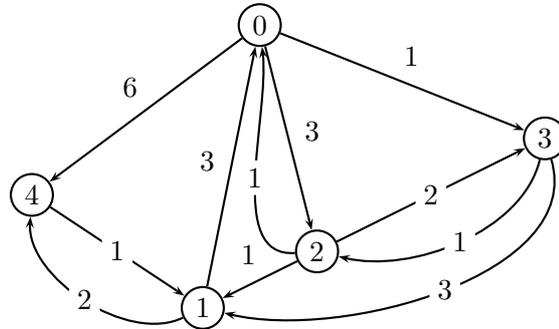


FIG. 8.1 – Le graphe \mathcal{G}_1 .

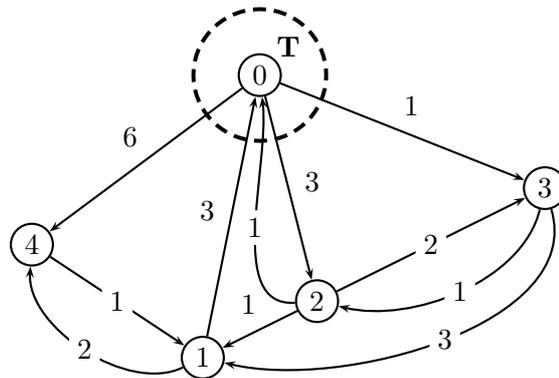


FIG. 8.2 – Étape 0.

La mise à jour de la fonction L donne :

t	1	2	3	4
$L(t)$	4	2	1	6

et on ajoute le sommet 2 dans T pour obtenir la figure 8.4.

La mise à jour de la fonction L donne :

t	1	2	3	4
$L(t)$	3	2	1	6

et on ajoute 1 à T . À l'étape 3, on trouve finalement

t	1	2	3	4
$L(t)$	3	2	1	5

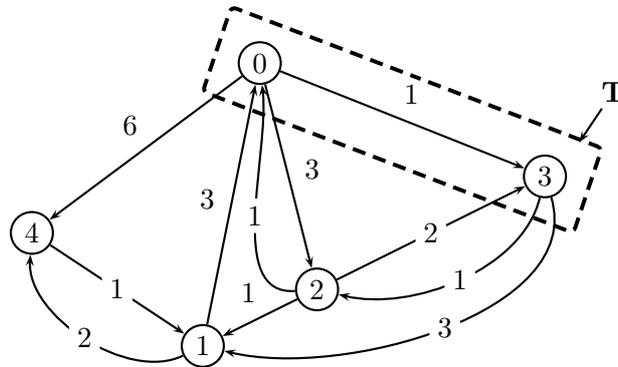


FIG. 8.3 – Étape 1.

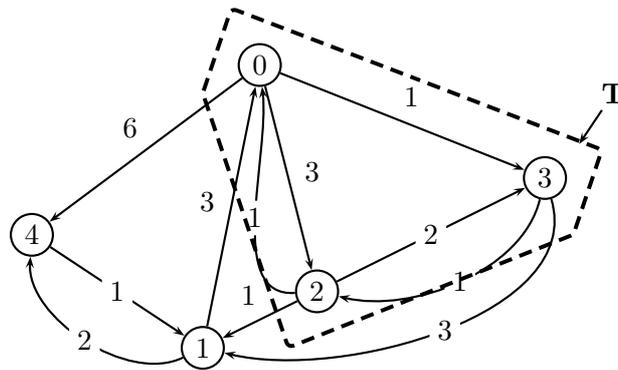


FIG. 8.4 – Étape 2.

8.1.2 Implantations

Cet algorithme est très intéressant à étudier et implanter, car le choix des structures de données a un impact crucial sur la complexité de l'algorithme. Notons tout de suite que l'algorithme va effectuer $n = |\mathcal{S}|$ étapes.

Préliminaires

La sortie de l'algorithme sera un tableau indexé par les sommets, donc une table de hachage indexée par les sommets.

Notons qu'il est plus facile de gérer l'ensemble $U = S - T$. Cela nous conduit à récrire le pseudocode sous la forme :

```

1. [Initialisation] pourtout  $t \neq s$  faire  $L(t) \leftarrow d(s, t)$ ;
2.  $L(s) \leftarrow 0$ ;  $U \leftarrow S - \{s\}$ ;
    
```

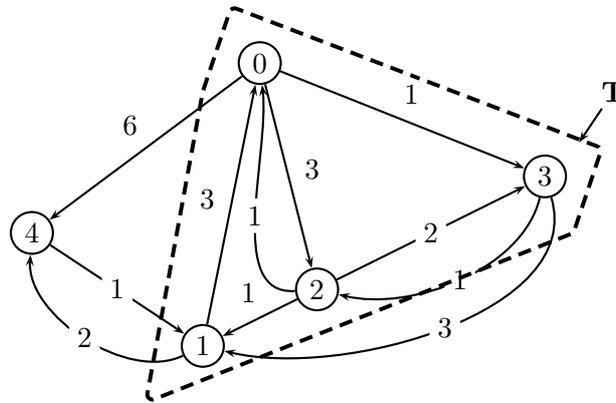


FIG. 8.5 – Étape 3.

```

3. tantque  $U \neq \emptyset$ 
4.   trouver  $v \in U$  telque  $L(v) = \min \{L(t), t \in U\}$ ;
5.    $U \leftarrow U - \{v\}$ ;
6.   pourtout  $t \in U$  faire
7.      $L(t) \leftarrow \min (L(t), L(v) + d(v, t))$ ;

```

Utilisation d'un ensemble

La première solution envisageable est de représenter U par un ensemble. L'étape 4 est la simple recherche du minimum dans un ensemble.

Quelle performance obtenons-nous? Trouver le minimum des $L(t)$ pour $t \in U$ coûte $O(|U|) = O(n)$. La mise à jour coûte $O(n_v)$ si n_v est le nombre de voisins de v , d'où un coût total :

$$\sum_{|T|=1}^n (O(n - |T|) + O(n_v)) = O(n^2) + O(|\mathcal{A}|) = O(n^2).$$

En Java, on peut par exemple utiliser un objet de `HashSet`, qui permet de récupérer un élément en temps constant. Le code correspondant est alors, en laissant les points de repère du pseudocode :

```

Hashtable<Sommet,Integer> Dijkstra(Sommet s){
    int INFINITY = 1000000; // [1]
    Hashtable<Sommet,Integer> L =
        new Hashtable<Sommet,Integer>();

    // 1. initialisation
    for(Sommet t : sommets())
        if(! t.equals(s))

```

```

        if(existeArc(s, t))
            L.put(t, valeurArc(s, t));
        else
            L.put(t, INFINITY);
// 2.
HashSet<Sommet> U
    = new HashSet<Sommet>(sommets()); // [2]
L.put(s, 0);
U.remove(s);
// 3.
while(! U.isEmpty()){
    // 4. recherche du minimum
    Sommet v = null;
    int Lv = INFINITY;
    for(Sommet t : U)
        if(L.get(t) < Lv){
            v = t;
            Lv = L.get(t);
        }
    // 5.
    U.remove(v);
    // 6.
    for(Arc a : voisins(v)){
        Sommet t = a.d;
        int tmp = Lv + a.valeur();
        if(tmp < L.get(t)) // [3]
            L.put(t, tmp);
    }
}
return L;
}

```

Apportons quelques précisions :

- 1 On a utilisé un majorant sur les valeurs des arcs. Le cas échéant, on peut déterminer un vrai majorant à la construction du graphe, pour un coût $O(|A|)$.
- 2 Il faut être prudent, car nous sommes amenés à enlever les sommets de U . Une écriture du type

```
Collection<Sommet> U = sommets();
```

aurait eu pour conséquence d'enlever les sommets de la table des sommets du graphe de départ, ce qui n'est pas une bonne idée en général.

- 3 On ne doit pas s'inquiéter du fait qu'on ne teste pas directement l'appartenance de t à U comme l'algorithme l'indique. En effet, si t est dans T , il n'y a aucun risque d'améliorer la valeur $L(t)$.

Utilisation d'une file de priorité

Nous décrivons dans cette section une amélioration de la complexité de la phase de recherche.

Quelle opération avons-nous à faire ? À chaque étape, on doit trouver l'élément v de $U = S - T$ qui minimise L . On peut utiliser une file de priorité qui permet de trouver le minimum d'un ensemble en temps $O(\log n)$, avec un coût de mise à jour en $O(\log n)$.

Le temps de calcul de notre algorithme serait alors :

$$\sum_{|T|=1}^n (O(\log n) + O(n_v \log n)) = O((|\mathcal{A}| + n) \log n),$$

ce qui est meilleur que $O(n^2)$ si $|\mathcal{A}| \ll n^2 / \log n$.

Donnons un programme Java implantant cette idée¹. Cela va nous permettre de donner des exemples non-triviaux de syntaxe et d'utilisation de Java.

En Java, la classe `TreeSet` permet d'implanter une file de priorité. Elle prend en paramètre une classe avec un comparateur (**extends** `Comparable<T>`) ou bien directement un comparateur à la création. C'est cette dernière syntaxe que nous allons donner. Nous choisissons d'utiliser le champ `marque` des sommets et donc une file de priorité de `Sommet`. Donnons la fonction d'initialisation de notre deuxième version de l'algorithme :

```

TreeSet<Sommet> Dijkstra2Init(Hashtable<Sommet,Integer> L,
                             Sommet s){
    int INFINITY = 1000000;
    TreeSet<Sommet> fp =
        new TreeSet<Sommet>(
            new Comparator<Sommet>() {
                public int compare(Sommet s1, Sommet s2) {
                    if(s1.valeurMarque()<s2.valeurMarque())
                        return -1;
                    else if(s1.valeurMarque()>s2.valeurMarque())
                        return 1;
                    if(s1.equals(s2))
                        return 0;
                    return -1;
                }
            }
        );
    for(Sommet u : sommets()){
        if(u.equals(s))
            L.put(u, 0);
    }
}

```

¹Le programme qui suit peut être omis dans une première phase de la lecture du poly.

```

        else if(existeArc(s, u)){
            int val = valeurArc(s, u);
            fp.add(new Sommet(u, val));
            L.put(u, val);
        }
        else{
            fp.add(new Sommet(u, INFINITY));
            L.put(u, INFINITY);
        }
    }
    return fp;
}

```

La boucle sur les sommets du graphe permet l'initialisation de la file de priorité, ainsi que de la table L qui a la même utilisation que dans la première version. Notez l'utilisation du constructeur `Sommet(Sommet s, int mm)`; ainsi que le rôle joué par `INFINITY` comme précédemment. Nous stockons deux fois la valeur $L(t)$, une fois dans le résultat final, une fois dans la file de priorité. Nous en avons besoin pour pouvoir enlever des couples $(t, L(t))$ de la file. L'initialisation telle qu'écrite a un coût $O(|S| + n_s) = O(n)$.

La fonction principale est alors :

```

Hashtable<Sommet,Integer> Dijkstra2(Sommet s){
    Hashtable<Sommet,Integer> L
        = new Hashtable<Sommet,Integer>();

    // 1. initialisation
    TreeSet<Sommet> U = Dijkstra2Init(L, s);
    // 3.
    while(! U.isEmpty()){
        // 4.
        Sommet v = U.first();
        U.remove(v);
        int Lv = v.valeurMarque();
        // 5.
        // 6.
        for(Arc a : voisins(v)){
            Sommet t = a.d;
            int tmp = Lv + a.valeur();
            if(tmp < L.get(t)){
                U.remove(new Sommet(t, L.get(t))); // [1]
                U.add(new Sommet(t, tmp));
                L.put(t, tmp);
            }
        }
    }
}

```

```

    }
    return L;
}

```

- 1 Pour mettre la file de priorité à jour, le plus simple est d'enlever le sommet, puis de mettre à la place un sommet avec sa nouvelle valeur.

On pourrait améliorer un peu les constantes d'implantation en étant moins brutal sur la mise à jour de U . Toutefois, ce n'est pas possible en restant générique. Il faudrait choisir une implantation par tas ou arbre et être capable d'aller chercher un nœud en plein milieu, pour accélérer la mise à jour. Nous laissons ce point comme sujet de méditation.

Pour aller plus loin

En utilisant un tas de Fibonacci, on peut faire décroître la complexité à $O(n \log n) + O(|A|)$. Nous renvoyons à la littérature pour cet algorithme qui dépasse le cadre actuel du cours.

8.2 Arbre couvrant de poids minimal

Imaginons un réseau de canaux de capacité connue devant irriguer un nombre donné de villes. Le problème est de minimiser la somme totale des capacités. On peut modéliser le problème par un graphe (non orienté) dont les sommets sont les villes et les arêtes les canaux, dont les capacités sont données. On notera $val(a)$ la valuation de l'arc a .

Considérons le graphe de la figure 8.6.

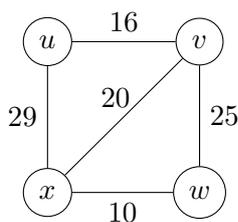


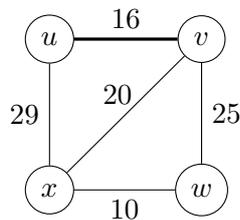
FIG. 8.6 –

Nous cherchons un sous-graphe connexe du graphe de départ, qui passe par tous les sommets du graphe, et de coût minimum. Si ce sous-graphe possédait un cycle, il suffirait de l'enlever pour diminuer son coût. Donc on cherche en fait un arbre couvrant de poids minimum. On note $\varpi(T)$ le poids de l'arbre, c'est-à-dire la somme des valeurs des arcs.

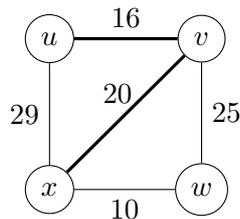
8.2.1 L'algorithme de Prim

1. Choisir un sommet initial s ;
2. **tantque** c'est possible
3. Trouver l'arête de poids minimal joignant un sommet déjà sélectionné **à** un sommet non encore sélectionné.

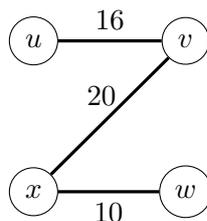
Reprenons encore le graphe donné en exemple, en commençant par exemple par le sommet u . On ajoute alors l'arête minimale joignant u , ici l'arête uv :



Puis nous passons en revue les arêtes contenant u et celles contenant v . L'arête de poids minimal est xv :



On cherche alors parmi les arêtes contenant $\{u, v, x\}$, ici wx :



Tous les sommets ayant été utilisés, l'algorithme s'arrête.

Comment peut-on implanter l'algorithme. Le plus efficace est de gérer une file de priorité d'arcs, classés par ordre croissant de valuation.

```

Prim(G = (S, A))
X <- S;
tantque X ≠ ∅ faire
  u <- élément suivant de X;
  X <- X - {u};
  
```

```

T[u] <- ∅; // arbre couvrant de racine u
F <- file de priorité avec les arêtes (u, x);
Y <- {u}; // ensemble des sommets de T[u]
tantque F ≠ ∅ faire
  a = (s, t) <- arête minimale dans F;
  si t ∈ Y alors
    // (s, t) ferme un cycle
  sinon
    Y <- Y ∪ {t};
    T[u] <- T[u] ∪ {(s, t)};
    pour v voisin de t
      ajouter (t, v) dans F;

```

Exercice 8.2.1 Implanter l'algorithme de Prim en Java.

Exercice 8.2.2 Montrer qu'il existe une variante de Prim fonctionnant en temps $O(m \log n)$.

8.2.2 L'algorithme de Kruskal

Il peut être décrit simplement comme suit :

```

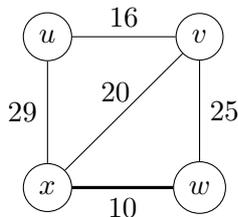
1. Trier les arêtes par ordre croissant pour obtenir
   B = (b0, b1, ..., bm);
2. T <- ∅;
3. pour i <- 0 à m-1 faire
   si bi ne crée pas de cycle
     ajouter bi à T;

```

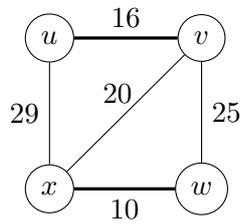
Revenons au cas du graphe de la figure 8.6. On trie les arêtes :

$$wx = 10, uv = 16, xv = 20, vw = 25, xu = 29.$$

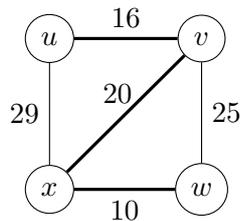
On commence par sélectionner l'arête wx , que nous dessinons en gras :



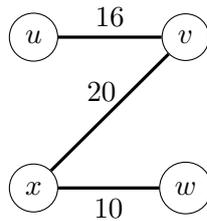
La deuxième arête est uv :



La troisième est xv , puisque nous n'avons toujours pas de cycle :



Nous ne pouvons pas utiliser vw , car cela créerait un cycle. Nous l'enlevons donc du graphe, ainsi que l'arête xu , qui elle aussi fermerait un cycle :



Le coût de l'étape 1 est bien sûr $O(m \log m)$ avec $m = |\mathcal{A}|$.

Comment implante la deuxième étape? On va créer une collection d'ensemble C indexée par les sommets. Si $t \in C[s]$, cela veut dire qu'il existe un chemin entre t et s dans la forêt en construction. Détaillons le code permettant d'insérer éventuellement une arête a . On suppose qu'à l'initialisation, on a mis tous les $C[s]$ à NIL.

```

insérer(C, a)
a = (s, t)
si C[s] == NIL et C[t] == NIL alors
  // l'arête (s, t) est nouvelle
  C[s] <- {s, t};
  C[t] <- C[s]; // partage
sinon si C[s] == NIL alors
  // t est déjà connu
  C[t] <- C[t] ∪ {s};
  C[s] <- C[t]; // partage
sinon si C[t] == NIL alors
  // s est déjà connu

```

```

C[s] <- C[s] ∪ {t};
C[t] <- C[s];
sinon // s et t sont déjà connus, on ne fait rien

```

Reprenons notre exemple. On commence par

$$C[w] = C[x] = \{w, x\}.$$

La deuxième arête est uv , ce qui crée

$$C[u] = C[v] = \{u, v\}.$$

Quand on veut insérer xv , on assiste à la fusion des composantes :

$$C[w] = C[x] = C[u] = C[v] = \{w, x, u, v\}.$$

Les deux dernières arêtes ne sont pas insérées, puisque les sommets sont déjà dans C .

Le lecteur attentif aura reconnu le principe de l'algorithme Union-Find. On sait que la complexité de traitement est $O(m\alpha(m))$ avec $\alpha(m)$ l'inverse de la fonction d'Ackermann $Ack(m, m)$, ce qui le rend quasi-linéaire.

Exercice 8.2.3 Implanter l'algorithme de Kruskal en Java.

8.2.3 Ébauche de preuve

Cette esquisse est valable pour les deux algorithmes.

Supposons que l'arbre construit T ne soit pas minimal et soit S un arbre couvrant de poids minimum $\varpi(S) < \varpi(T)$. Soit a une arête de poids minimum présente dans T mais pas dans S . Si on ajoute cette arête a à S , cela crée un cycle (car tous les sommets sont déjà présents dans S). Ce cycle contient une autre arête a' non présente dans T . Si on remplace a' par a dans S pour obtenir S' , cet arbre est encore un arbre couvrant. Par construction de T , le poids de a ne peut dépasser celui de a' , et donc celui de S' ne peut dépasser celui de S . De proche en proche, on peut donc passer de T à S , alors que le poids de T est plus grand que celui de S , contradiction.

8.2.4 Prim ou Kruskal ?

Dans sa variante rapide, Prim a coût $O(m \log n)$, Kruskal $O(m \log m)$. Comme $\log m = O(\log n)$, on est dans la même classe de complexité. Néanmoins, si on sait que $n \ll m$, alors Prim gagne un peu.

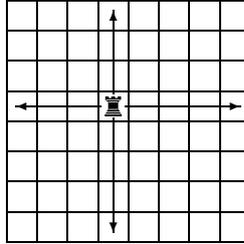
Si l'on regarde bien, Prim ne trouve l'arbre couvrant que pour la composante connexe du sommet de départ, alors que Kruskal construit tous les arbres en même temps.

8.3 Les n reines

Nous allons encore voir un algorithme de backtrack pour résoudre un problème combinatoire. Dans la suite, nous supposons que nous utilisons un échiquier $n \times n$.

8.3.1 Prélude : les n tours

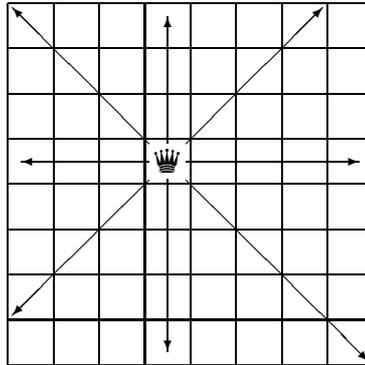
Rappelons quelques notions du jeu d'échecs. Une tour menace toute pièce adverse se trouvant dans la même ligne ou dans la même colonne.



On voit facilement qu'on peut mettre n tours sur l'échiquier sans que les tours ne s'attaquent. En fait, une solution correspond à une permutation de $1..n$, et on sait déjà faire. Le nombre de façons de placer n tours non attaquantes est donc $T(n) = n!$.

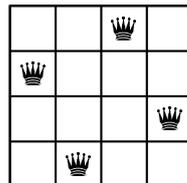
8.3.2 Des reines sur un échiquier

La reine se déplace dans toutes les directions et attaque toutes les pièces (adverses) se trouvant sur les même ligne ou colonne ou diagonales qu'elle.

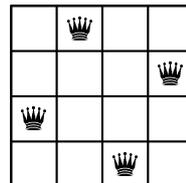


Une reine étant une tour avec un peu plus de pouvoir, il est clair que le nombre maximal de reines pouvant être sur l'échiquier sans s'attaquer est au plus n . On peut montrer que ce nombre est n pour $n = 1$ ou $n \geq 4$ ². Reste à calculer le nombre de solutions possibles, et c'est une tâche difficile, et non résolue.

Donnons les solutions pour $n = 4$:



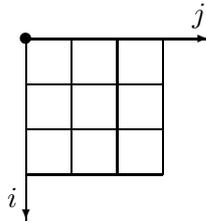
(2413)



(3142)

²Les petits cas peuvent se faire à la main, une preuve générale est plus délicate et elle est due à Ahrens, en 1921.

Expliquons comment résoudre le problème de façon algorithmique. On commence par chercher un codage d'une configuration. Une configuration admissible sera codée par la suite des positions d'une reine dans chaque colonne. On oriente l'échiquier comme suit :



Avec ces notations, on démontre :

Proposition 8.3.1 *La reine en position (i_1, j_1) attaque la reine en position (i_2, j_2) si et seulement si $i_1 = i_2$ ou $j_1 = j_2$ ou $i_1 - j_1 = i_2 - j_2$ ou $i_1 + j_1 = i_2 + j_2$.*

Démonstration : si elle sont sur la même diagonale nord-ouest/sud-est, $i_1 - j_1 = i_2 - j_2$; ou encore sur la même diagonale sud-ouest/nord-est, $i_1 + j_1 = i_2 + j_2$. \square

On va procéder comme pour les permutations : on suppose avoir construit une solution approchée dans $t[1..i_0[$ et on cherche à placer une reine dans la colonne i_0 . Il faut s'assurer que la nouvelle reine n'attaque personne sur sa ligne (c'est le rôle du tableau `utilise` comme pour les permutations), et personne dans aucune de ses diagonales (fonction `pasDeConflit`). Le code est le suivant :

```
// t[1..i0[ est déjà rempli
static void reinesAux(int[] t, int n,
                    boolean[] utilise, int i0){
    if(i0 > n)
        afficher(t);
    else{
        for(int v = 1; v <= n; v++){
            if(! utilise[v] && pasDeConflit(t, i0, v)){
                utilise[v] = true;
                t[i0] = v;
                reinesAux(t, n, utilise, i0+1);
                utilise[v] = false;
            }
        }
    }
}
```

La programmation de la fonction `pasDeConflit` découle de la proposition 8.3.1 :

```
// t[1..i0[ est déjà rempli
static boolean pasDeConflit(int[] t, int i0, int j){
```

```

int x1, y1, x2 = i0, y2 = j;

for(int i = 1; i < i0; i++){
    // on récupère les positions
    x1 = i;
    y1 = t[i];
    if((x1 == x2) // même colonne
        || (y1 == y2) // même ligne
        || ((x1-y1) == (x2-y2))
        || ((x1+y1) == (x2+y2)))
        return false;
    }
return true;
}

```

Notons qu'il est facile de modifier le code pour qu'il calcule le nombre de solutions. Terminons par un tableau des valeurs connues de $R(n)$:

n	$R(n)$	n	$R(n)$	n	$R(n)$	n	$R(n)$
4	2	9	352	14	365596	19	4968057848
5	10	10	724	15	2279184	20	39029188884
6	4	11	2680	16	14772512	21	314666222712
7	40	12	14200	17	95815104	22	2691008701644
8	92	13	73712	18	666090624	23	24233937684440

Vardi a conjecturé que $\log R(n)/(n \log n) \rightarrow \alpha > 0$ et peut-être que $\alpha = 1$. Rivin & Zabih ont d'ailleurs mis au point un algorithme de meilleur complexité pour résoudre le problème, avec un temps de calcul de $O(n^2 8^n)$.

8.4 Les ordinateurs jouent aux échecs

Nous ne saurions terminer un chapitre sur la recherche exhaustive sans évoquer un cas très médiatique, celui des ordinateurs jouant aux échecs.

8.4.1 Principes des programmes de jeu

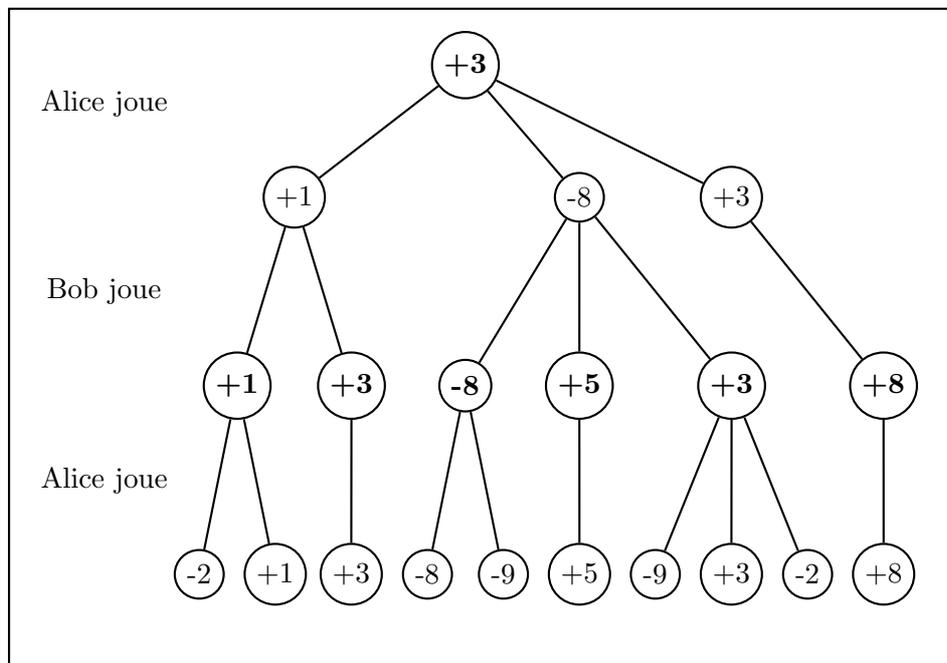
Deux approches ont été tentées pour battre les grands maîtres. La première, dans la lignée de Botvinnik, cherche à programmer l'ordinateur pour lui faire utiliser la démarche humaine. La seconde, et la plus fructueuse, c'est utiliser l'ordinateur dans ce qu'il sait faire le mieux, c'est-à-dire examiner de nombreuses données en un temps court.

Comment fonctionne un programme de jeu ? En règle général, à partir d'une position donnée, on énumère les coups valides et on crée la liste des nouvelles positions. On tente alors de déterminer quelle est la meilleure nouvelle position possible. On fait cela sur plusieurs tours, en parcourant un arbre de possibilités, et on cherche à garder le meilleur chemin obtenu.

Dans le meilleur des cas, l'ordinateur peut examiner tous les coups et il gagne à coup sûr. Dans le cas des échecs, le nombre de possibilités en début et milieu de partie est beaucoup trop grand. Aussi essaie-t-on de programmer la recherche la plus profonde possible.

8.4.2 Algorithme minimax

Dans cet algorithme, on imagine que le joueur A va prendre la meilleure décision possible compte tenu des coups qu'il peut jouer et des coups que son adversaire B peut jouer. À partir d'une position donnée au niveau 0, A joue, ouvrant un certain nombre de coups valides à B, ce qui conduit à d'autres coups, et ainsi de suite. Regardons ce qui se passe sur un exemple :

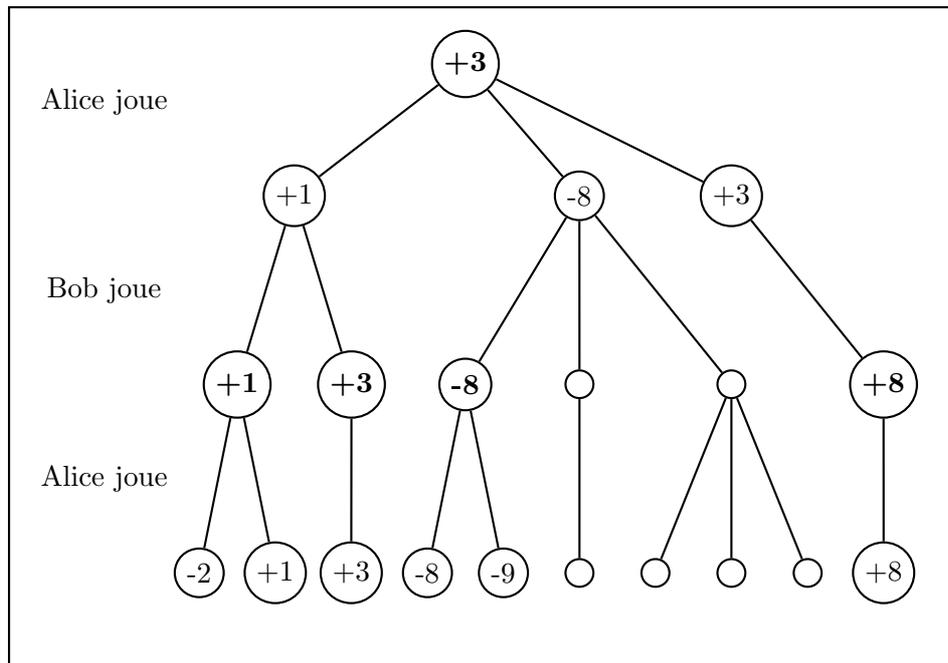


Au niveau le plus, A joue et cela conduit à des positions positives si la position est gagnante, négative sinon. Si A a le choix entre une position finale à +1 ou -2, elle va choisir +1, ce que l'on symbolise au cran précédent par +1 dans le nœud de l'arbre. A maximise ainsi l'évaluation de son coup. Au cran d'avant, B va au contraire minimiser le gain de A, puisqu'il veut gagner. C'est ainsi qu'entre un coup conduisant à +1 ou +3, il choisit +1. Parcourant ainsi toutes les branches de l'arbre, on voit que le meilleur coup pour A avec une profondeur de 4 conduit à une position +3.

Il est facile de voir que la complexité de cette approche est $O(b^{\text{prof}})$.

8.4.3 Principe de l'élagage $\alpha\beta$

Dans certains cas, on peut couper des branches de l'arbre, ce qui permet de gagner du temps dans l'exploration. Expliquons ce qui se passe sur un exemple :



8.4.4 Retour aux échecs

Codage d'une position

La première idée qui vient à l'esprit est d'utiliser une matrice 8×8 pour représenter un échiquier. On l'implante généralement sous la forme d'un entier de type long qui a 64 bits, un bit par case. On gère alors un ensemble de tels entiers, un par type de pièce par exemple.

On trouve dans la thèse de J. C. Weill un codage astucieux :

- les cases sont numérotées de 0 à 63;
- les pièces sont numérotées de 0 à 11 : pion blanc = 0, cavalier blanc = 1, ..., pion noir = 6, ..., roi noir = 11.

On stocke la position dans le vecteur de bits

$$(c_1, c_2, \dots, c_{768})$$

tel que $c_{64i+j+1} = 1$ ssi la pièce i est sur la case j .

Les positions sont stockées dans une table de hachage la plus grande possible qui permet de reconnaître une position déjà vue.

Fonction d'évaluation

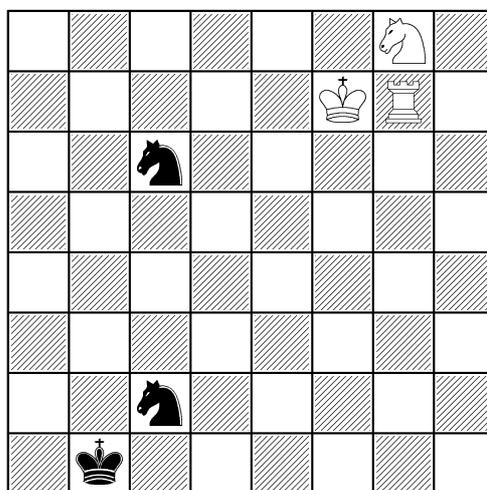
C'est un des secrets de tout bon programme d'échecs. L'idée de base est d'évaluer la force d'une position par une combinaison linéaire mettant en œuvre le poids d'une pièce (reine = 900, tour = 500, etc.). On complique alors généralement la fonction en fonction de stratégies (position forte du roi, etc.).

Bibliothèques de début et fin

Une façon d'accélérer la recherche est d'utiliser des bibliothèques d'ouvertures pour les débuts de partie, ainsi que des bibliothèques de fins de partie.

Dans ce dernier cas, on peut tenter, quand il ne reste que peu de pièces d'énumérer toutes les positions et de classer les perdantes, les gagnantes et les nulles. L'algorithme est appelé analyse rétrograde et a été décrit par Ken Thompson (l'un des créateurs d'Unix).

À titre d'exemple, la figure ci-dessous décrit une position à partir de laquelle il faut **243 coups** (contre la meilleure défense) à Blanc (qui joue) pour capturer une pièce sans danger, avant de gagner (Stiller, 1998).



Deep blue contre Kasparov (1997)

Le projet a démarré en 1989 par une équipe de chercheurs et techniciens : C. J. Tan, Murray Campbell (fonction d'évaluation), Feng-hsiung Hsu, A. Joseph Hoane, Jr., Jerry Brody, Joel Benjamin. Une machine spéciale a été fabriquée : elle contenait 32 nœuds avec des RS/6000 SP (chip P2SC) ; chaque nœud contenait 8 processeurs spécialisés pour les échecs, avec un système AIX. Le programme était écrit en C pour le IBM SP Parallel System (MPI). La machine était capable d'engendrer 200,000,000 positions par seconde (ou 60×10^9 en 3 minutes, le temps alloué). Deep blue a gagné 2 parties à 1 contre Kasparov³.

Deep Fritz contre Kramnik (2002)

C'est cette fois un ordinateur plus raisonnable qui affronte un humain : 8 processeurs à 2.4 GHz et 256 Mo, qui peuvent calculer 3 millions de coups à la seconde. Le programmeur F. Morsch a soigné la partie algorithmique. Kramnik ne fait que match nul (deux victoires chacun, quatre nulles), sans doute épuisé par la tension du match.

³www.research.ibm.com/deepblue

Conclusion

Peut-on déduire de ce qui précède que les ordinateurs sont plus intelligents que les humains ? Certes non, ils *calculent* plus rapidement sur certaines données, c'est tout. Pour la petite histoire, les joueurs d'échec peuvent s'adapter à l'ordinateur qui joue face à lui et trouver des positions qui le mettent en difficulté. Une manière de faire est de jouer systématiquement de façon à maintenir un grand nombre de possibilités à chaque étape.

Troisième partie

Langages

Chapitre 9

Langages rationnels

Une des fonctions principales des ordinateurs dans le monde est de rechercher des informations dans des bases de connaissance : courrier, fichiers de documentation, articles, livres, pages Web, etc. Dans les moteurs de recherche du Web, dans tous les systèmes d'interrogation bibliographiques, dans les traducteurs et compilateurs on retrouve cette fonction essentielle qui permet de retrouver rapidement les pages, les ouvrages, les passages du texte contenant un mot, un groupe de mots avec ou sans variante. Les logiciels les plus élaborés peuvent tenir compte de la phonétique comme dans l'annuaire électronique de France-Télécom. De fait, la théorie sous-jacente a été développée dans les années 1950, sous l'impulsion initiale de linguistes, de logiciens et d'informaticiens théoriciens pour donner la théorie des automates finis et des langages rationnels (et expressions régulières).

Par ailleurs, ces mêmes automates sont à la base de très nombreux automatismes développés par les électroniciens et autres concepteurs d'appareils électro-mécaniques. Les réseaux téléphoniques, les circuits de commande et de contrôle sont en fait des automates finis munis de senseurs et d'effecteurs. On remonte en cela à la tradition de Vaucanson et Jacquard (et bien d'autres). Les montres (mécaniques ou électroniques) sont de beaux exemples, ainsi que tous les appareils-photos et autres caméras ou lecteurs de CD.

Comprendre la puissance et les limites de ce modèle simple est donc important. On verra aussi apparaître le phénomène du non-déterminisme.

9.1 Recherche de motifs dans un texte

Cette section est dévolue à l'étude et à l'analyse d'algorithmes qui recherchent des motifs ou des ensembles de motifs simples. On commencera donc par la recherche d'une suite particulière de lettres (un *mot*) dans un texte, puis on compliquera en passant à un mot parmi un ensemble fini.

9.1.1 Un algorithme naïf

Considérons un ensemble fini A de lettres appelé *alphabet*. On appelle *mot* toute suite finie de lettres de A .

Définition 9.1.1 *L'ensemble des mots sur un alphabet A est noté A^* . Un mot $w \in A^*$ est défini par la suite de ses lettres : $w = w_1w_2\dots w_k$, où $k = |w|$ est la longueur, ou encore le nombre de lettres, de w .*

L'opérateur de concaténation est une loi interne binaire de A^ qui à deux mots u et v associe le mot $w = uv$, de longueur $|u| + |v|$, défini par*

$$w_i = \text{si } i \leq |u| \text{ alors } u_i \text{ sinon } v_{i-|u|}.$$

Il existe un mot unique de longueur 0, appelé mot vide et noté ε .

Notons, de plus, que la loi de concaténation n'est pas commutative, qu'elle est associative et que le mot vide est élément neutre.

Définition 9.1.2 *L'ensemble A^* muni de la concaténation a une structure de monoïde, dont on dit qu'il est libre car la décomposition d'un mot sur les lettres de l'alphabet est unique.*

Cette propriété de décomposition unique n'est pas vraie de tous les monoïdes. Elle est ici triviale, est caractéristique des ensembles de mots formant des codes (la décomposition est unique) et est à la base de la théorie des codes combinatoires.

L'algorithme le plus simple de recherche de motif dans un texte essaie toutes les positions possibles du motif m en dessous du texte t . Comment tester qu'il existe une occurrence en position i ? Il suffit d'utiliser un indice j qui va servir à comparer $m[j]$ à $t[i+j]$ de proche en proche :

```
static boolean occurrence(String t, String m, int i){
    for(int j = 0; j < m.length(); j++)
        if(t.charAt(i+j) != m.charAt(j))
            return false;
    return true;
}
```

Nous utilisons cette primitive dans la fonction suivante, qui teste toutes les occurrences possibles :

```
static void naif(String t, String m){
    System.out.print("Occurrences en position :");
    for(int i = 0; i < t.length()-m.length(); i++)
        if(occurrence(t, m, i))
            System.out.print(" "+i+",");
    System.out.println("");
}
```

Cette fonction opère en un nombre de comparaisons qui dépend du motif et du texte. En effet, considérons le motif $m = aaaaaab$ et les textes $t_1 = aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaab$ et $t_2 = bbbbbbbbbbbbbbbbbbbbbbbbbb$; le nombre de comparaisons entre lettres du motif et du texte dans pour t_1 est équivalent à $|m|.|t_1|$ essentiellement alors qu'il n'est que de $|t_2|$ dans le deuxième cas. Même si on peut montrer qu'en moyenne le coût ne dépend pas de la taille du motif (pour un texte aléatoire), on aimerait avoir une méthode insensible à la nature du motif et du texte, indépendante de leurs régularités ou irrégularités.

9.1.2 L'algorithme KMP

On doit remarquer que le gros défaut de la méthode naïve est d'oublier ce qui a été lu et comparé du texte après chaque décalage. On pourrait pourtant économiser beaucoup de comparaisons si l'on pouvait mémoriser par exemple que t_1 ne contient que des a et donc qu'en cas de lecture de nombreux a consécutifs, il suffit de chercher le premier b . Plus précisément on peut construire, sur un exemple moins caricatural le dispositif suivant bien plus efficace.

Définition 9.1.3 Soit $w \in A^*$ un mot non vide ; pour toute décomposition de $w = f_1 f_2 \dots f_k$ telle que $f_i \neq \varepsilon$, on dira que f_i est un facteur de w ; f_1 sera appelé préfixe ou facteur gauche et f_k suffixe ou facteur droit.

Soient le motif $m = abaabbab$ et le texte $t = ababbabaabaabb\dots$. On peut alors commenter la recherche d'une occurrence de m dans t comme suit, les étoiles correspondant aux comparaisons couronnées de succès et les dièses aux comparaisons conclues par un échec :

```

t = a b a b b a b a a b a a b b . . .
m = a b a a b b a b
    * * * #                reprendre la comparaison après le
                           dernier a b
      a b a a b b a b
        #                reprendre la comparaison au début
          a b a a b b a b
            * * * * * #   reprendre après le dernier a b a
              a b a a b b a b
                * * * . . . etc.
    
```

- On aura compris :
- que lors de la première différence entre m et t , on a pu garder en mémoire le fait que leur dernières lettres lues du texte coïncident avec les deux premières du motif et donc,
 - qu'on peut poursuivre la lecture sur la troisième lettre du motif sans revenir en arrière sur le texte.

Il y a alors échec immédiat et comme aucun suffixe des 5 premières lettres du texte ne correspond à un préfixe du motif, on en déduit qu'on peut décaler entièrement le motif.

Pour le troisième échec on constate que "a b a" est le plus long suffixe de la portion de texte lue qui soit aussi préfixe du motif, et on décale donc le motif de 3 positions, etc.

On constate donc qu'il serait possible de ne pas revenir en arrière sur le texte. Le processus de reprise en cas de discordance entre le texte et le motif peut se représenter par le schéma suivant où il suffit de suivre les flèches en fonction de la lettre lue sur le texte. Un tel schéma est appelé *automate*. Chaque position numérotée est appelée *état* ; le dernier état (matérialisé par un \diamond) est appelé final et correspond à une occurrence du motif.

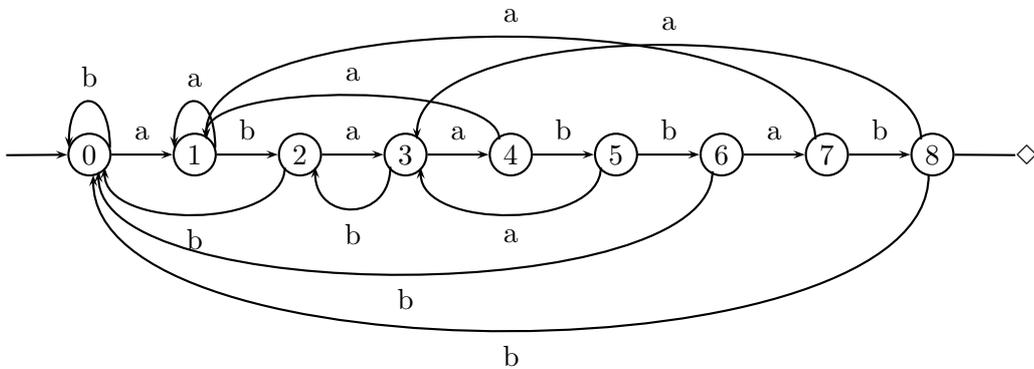


FIG. 9.1 – L'automate de reconnaissance du motif $m = abaabbab$

L'automate associé à un motif m , possède donc $|m| + 1$ états nommés $0, 1, 2, \dots, |m|$ et les *transitions* de l'automate (ses flèches) sont définies par la fonction $\tau : [0..|m|] \times A \rightarrow [0..|m|]$, telle que $\tau(i, \alpha)$ est le nombre de lettres du plus grand suffixe de $m[1..i]\alpha$ qui soit aussi préfixe de m .

On a donc :

$$\tau(i, \alpha) = \begin{cases} i + 1 & \text{si } \alpha = m[i + 1], \\ \max\{j \leq i \mid m[1..j] = m[i - j + 2..i]\alpha\} & \text{sinon.} \end{cases}$$

La construction de cette table τ prend un temps $O(|m|^3)$ si l'on procède de manière directe, mais une fois ce pré-traitement (on dit souvent en jargon *preprocessing*) la recherche du motif se fait en temps linéaire $O(|t|)$, puisqu'il suffit de parcourir le graphe de l'automate en lisant à chaque étape une lettre de t .

Cependant, il est possible de construire la table de l'automate en temps linéaire, à condition d'introduire la fonction auxiliaire $\delta : [0..|m|] \rightarrow [0..|m|]$, telle que $\delta(i)$ soit le nombre de lettres du plus grand préfixe de $m[1..i - 1]$ qui soit aussi suffixe de $m[2..i]$.

On a donc :

$$\delta(i) = \begin{cases} 0 & \text{si } i = 0, \\ \max\{j < i \mid m[1..j] = m[i - j + 1..i]\} & \text{sinon.} \end{cases}$$

On remarquera les différences subtiles dans le domaine de variation de la variable j . On a alors la propriété remarquable que τ et δ se calculent ensemble de façon très efficace.

Proposition 9.1.1 *Les fonctions τ et δ satisfont les récurrences croisées :*

$$\begin{aligned} \delta(i) &= \begin{array}{l} \text{si } i \leq 1 \text{ alors } 0 \\ \text{sinon } \tau(\delta(i-1), m[i]) \end{array} \\ \tau(i, \alpha) &= \begin{array}{l} \text{si } \alpha = m[i+1] \text{ alors } i+1 \\ \text{sinon si } i = 0 \text{ alors } 0 \\ \text{sinon } \tau(\delta(i), \alpha) \end{array} \end{aligned}$$

Preuve : (Esquisse)

Elle se fait par récurrence sur i .

Les choses sont évidentes pour $i = 0, 1$.

Dans le cas général, deux situations se présentent selon les valeurs respectives de $m[i+1]$ et $m[\delta(i)+1]$: on garde en mémoire que, par hypothèse de récurrence, $m[1..\delta(i)] = m[i-\delta(i)+1..i]$.

Dans le cas où $m[i+1] = m[\delta(i)+1] = \alpha$, alors m a la structure suivante lorsque les facteurs appelés f ne sont pas chevauchants : $m = f\alpha u f\alpha v f\alpha w$, où f est de longueur $\delta(i+1) - 1$, le premier α est donc en position $\delta(i+1)$, le second α est en position $\delta(i)+1$, et le troisième α est en position $i+1$. (Il est possible que les facteurs $f\alpha$ se chevauchent, mais ceci n'affecte pas le raisonnement.)

Il faut justifier qu'il n'existe pas de facteur gauche $g\alpha$ de m plus grand que $f\alpha$ qui soit aussi facteur droit de $m[1..i+1]$. Si tel était le cas, comme $m[1..\delta(i)]$ est aussi facteur droit de $m[1..\delta]$, il aurait aussi g comme facteur droit, ce qui est contraire à l'hypothèse de récurrence qui stipule que $\delta(i) = |f|$. En conséquence on a bien que $\tau(\delta(i), m[i+1]) = \delta(i+1)$ et satisfait donc les définitions de τ et δ .

Dans le cas où $m[i+1] \neq m[\delta(i)+1]$, m se décompose aussi selon un schéma semblable $m = f u f v f w$, les α étant simplement omis. Le raisonnement est exactement identique et le résultat aussi.

On vérifie de la même façon (mais c'est un peu une évidence) que la fonction δ se définit bien comme mentionné à partir de τ et de m .

Le calcul des fonctions τ et δ devient alors beaucoup plus rapide, puisque de quadratique en $|m|$, il devient linéaire, le théorème précédent permettant un calcul par récurrence linéaire.

Proposition 9.1.2 *Le calcul simultané des fonctions τ et δ , pour un motif m donné, se fait en temps et espace $O(|m|)$.*

On a donc le résultat (non trivial) de D. E. Knuth, J. H. Morris et V. R. Pratt, qui a des prolongements importants en matière de reconnaissance de mots à structure palindromique itérée.

Théorème 9.1.1 *La recherche de toutes les occurrences d'un motif m dans un texte t se fait en temps linéaire $O(|m| + |t|)$.*

Preuve : On commence par construire les tables τ et δ des fonctions τ et δ associées au motif m , puis on parcourt séquentiellement le texte t selon le programme suivant, dans lequel on dénote par l_{motif} et l_{texte} les longueurs respectives de m et t :

```

int etat = 0; // la position de départ de l'automate
for(int i = 0; i < T.length(); i++){
    if(M.charAt(etat+1) == T.charAt(i+1)){
        etat++;
        if(etat == M.length()){
            System.out.print("occurrence en :");
            System.out.println((i+1-M.length()));
        }
    }
    else
        etat = tau[delta[etat]][(int)T.charAt(i+1)];
}

```

Exercice 9.1.1 Compléter le programme précédent pour en faire un vrai programme Java.

Le coût de construction des tables étant $O(|m|)$ et celui d'exécution de la boucle de recherche étant trivialement $O(|t|)$, on en déduit l'assertion.

Remarques :

1. On peut ne pas précalculer les deux tables dans leur totalité, mais se contenter de les remplir au fur et à mesure des besoins en faisant un appel récursif si le résultat n'a pas été stocké dans la table auparavant. On économise ainsi du pré-traitement. Une telle technique de programmation est appelée *mémo-fonction*.
2. Lorsque l'alphabet est de grande cardinalité, par exemple ASCII 7 bits ou pire Unicode, la table τ sert à moins de 1%, voire moins de 10^{-4} . L'usage de mémo-fonctions est bien sûr une des réponses à cette situation ; on peut aussi convertir le problème en binaire et appliquer les méthodes précédentes dans le cas binaire sur des motif et texte un peu plus longs. Le lecteur trouvera matière à réflexion dans les exercices.

9.1.3 Multimotifs

Une première généralisation du problème est de rechercher un motif parmi plusieurs possibles. Peut-on gagner du temps en factorisant, pour ainsi dire, les recherches des divers motifs sur le même texte ?

La solution proposée est de nature semblable en compliquant légèrement la structure de l'automate du paragraphe précédent. Prenons un exemple avec pour ensemble de motifs $M = \{ababb, aaba, baa\}$.

On construit alors l'automate donné en figure 9.2, sur lequel on remarquera que le squelette a une forme arborescente fournie par l'arbre binaire (incomplet) des motifs

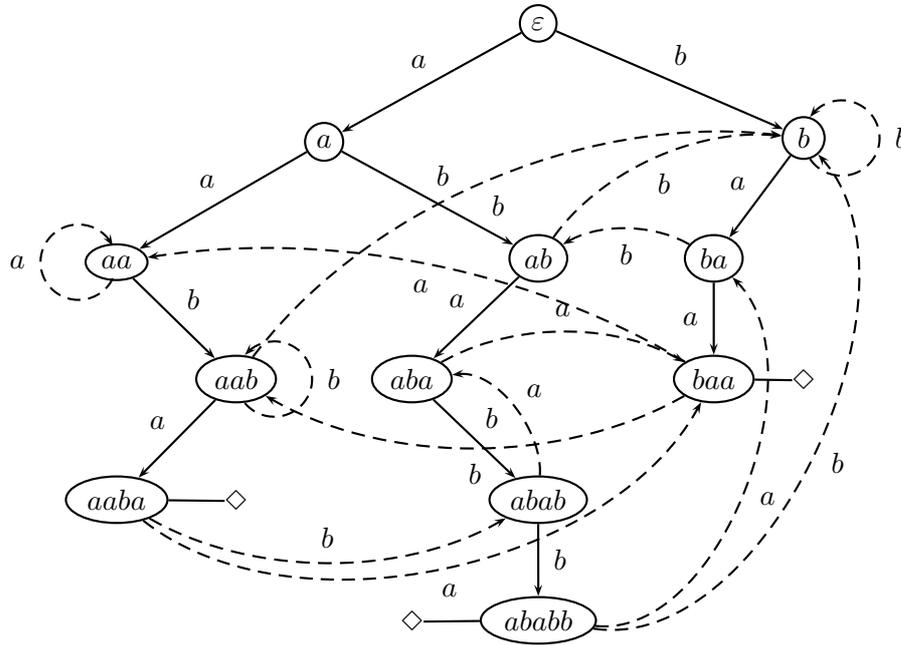


FIG. 9.2 – L’automate de reconnaissance du multimotif $M = \{ababb, aaba, baa\}$; le squelette de l’automate est marqué en arcs pleins alors que les autres actions possibles le sont en arcs pointillés ; les \diamond marquent les états finaux.

de l’ensemble M . Cet arbre est incomplet car il ne possède que trois branches, correspondant chacune à l’un des motifs ; de façon générale, il est construit sur l’ensemble des facteurs gauches des mots de M , encore appelés *préfixes* de M , et noté $FG(M)$. Les sommets de l’arbre (qui sont aussi les états de l’automate) sont précisément les éléments de $FG(M) = \{\varepsilon, a, b, aa, ab, ba, aab, aba, baa, aaba, abab, ababb\}$, où ε désigne le mot vide, préfixe de tout mot.

Les trois états $\{ababb, aaba, baa\}$ sont finaux car on n’y arrive que si l’on vient précisément de lire dans le texte un mot de M . Les autres arêtes sont calculées comme dans le cas précédent au moyen de la fonction $\tau(f, \alpha)$ dont la valeur est le plus grand suffixe de $f\alpha$ qui en soit aussi préfixe d’un mot de M :

$$\tau(f, \alpha) = \mu\{g \in FG(M) \mid \exists h, l \ f\alpha = gh = lg\},$$

définie sur $FG(M) \times \{a, b\}$ et où μ désigne l’opérateur *le plus long*. Remarquons que si $f\alpha \in FG(M)$, alors $g = f\alpha$.

Comme au paragraphe précédent on gagne du temps de calcul pour la fonction τ , en utilisant la fonction auxiliaire $\delta(f)$, pour $f \in FG(M)$, qui calcule le plus grand suffixe strict de f qui soit aussi dans $FG(M)$ et qui est définie par

$$\delta(f) = \mu\{g \in FG(M) \mid \exists h, l \neq \varepsilon \ f = gh = lg\}.$$

Cette fonction δ permet de définir τ et δ par les récurrences conjointes, où $f \in FG(M)$ et $\alpha \in \{a, b\}$:

$$\begin{aligned} \delta(\varepsilon) &= \varepsilon \\ \delta(f\alpha) &= \begin{array}{l} \text{si } |f| = \varepsilon \text{ alors } \varepsilon \\ \text{sinon } \tau(\delta(f), \alpha) \end{array} \\ \tau(f, \alpha) &= \begin{array}{l} \text{si } f\alpha \in FG(M) \text{ alors } f\alpha \\ \text{sinon si } f = \varepsilon \text{ alors } \varepsilon \\ \text{sinon } \tau(\delta(f), \alpha). \end{array} \end{aligned}$$

9.2 Automates finis

Nous présentons maintenant le formalisme général de cette classe de mécanisme de calcul simple. De fait, nous allons voir ces machines plus comme des machines à caractériser des familles de mots que comme des machines à calculer un résultat numérique ou symbolique. On dit que ces automates fonctionnent en *reconnaissance de langage*. Un des côtés les plus surprenants, et c'est un nouveau concept, est que ces machines sont *non-déterministes*, c'est-à-dire qu'elles sont susceptibles de plusieurs évolutions différentes dans le même environnement. Ce n'était pas le cas des deux exemples donnés dans la section précédente où le lecteur vérifiera qu'étant donné un texte le chemin suivi dans le graphe de l'automate est unique et pré-déterminé.

9.2.1 Définitions

Un automate fini agit sur les mots appartenant à l'ensemble A^* défini par l'alphabet fini A ; pour chaque mot, il permettra de déterminer s'il appartient ou non à l'ensemble caractérisé par l'automate, comme on le définit maintenant.

Définition 9.2.1 *Un automate fini A sur un alphabet fini A est défini par un quintuplet*

$$\mathcal{A} = \langle A, Q, q_0, F, \Delta \rangle,$$

où Q est l'ensemble fini des états, $q_0 \in Q$ est l'état initial, l'ensemble $F \subset Q$ est celui des états finaux et Δ , appelée table de transition est un ensemble fini

$$\Delta \subset Q \times (A \cup \{\varepsilon\}) \times Q.$$

Si la table de transition est une relation fonctionnelle, on dit que l'automate fini est déterministe.

Exemples :

1. L'automate de la figure 9.1 reconnaissant les mots contenant le motif *abaabbab* est défini comme suit :
 - alphabet : $A = \{a, b\}$;
 - états : $Q = \{0, 1, 2, 3, 4, 5, 6, 8\}$;

- état initial : 0 ;
- états finaux : $F = \{8\}$;
- table : $\Delta = \{(i, \alpha, j) \mid j = \tau(i, \alpha), 0 \leq i, j \leq 8, \alpha \in A\}$.

2. L'automate de la figure 9.2 reconnaissant les mots contenant l'un des motifs de l'ensemble $M = \{ababb, aaba, baa\}$ est défini comme suit :

- alphabet : $A = \{a, b\}$;
- états : $Q = FG(M) = \{\varepsilon, a, b, aa, ab, ba, aab, aba, baa, aaba, abab, ababb\}$;
- état initial : ε ;
- états finaux : $F = \{aaba, baa, ababb\}$;
- table : $\Delta = \{(f, \alpha, g) \mid g = \tau(f, \alpha), f, g \in Q, \alpha \in A\}$.

Il faut noter pour ces deux automates que Δ est en fait le graphe de la fonction τ et donc qu'étant donnés $f \in Q$ et $\alpha \in A$, il existe au plus un g tel que $(f, \alpha, g) \in \Delta$. L'automate est alors qualifié de *déterministe*, comme on le retrouvera plus bas.

3. Un automate pathologique : soit \mathcal{P}_7 défini comme suit :

- alphabet : $A = \{a, b\}$;
- états : $Q = \{0, 1, 2, 3, 4, 5, 6, 7\}$;
- état initial : 0 ;
- états finaux : $F = \{7\}$;
- table : $\Delta = \{(0, a, 0), (0, b, 0), (0, a, 1)\} \cup \{(i, a, i + 1), (i, b, i + 1) \mid 1 \leq i < 7\}$.

Graphiquement, il est sans doute plus facile de comprendre la structure de l'automate, comme on le voit sur la figure 9.3.

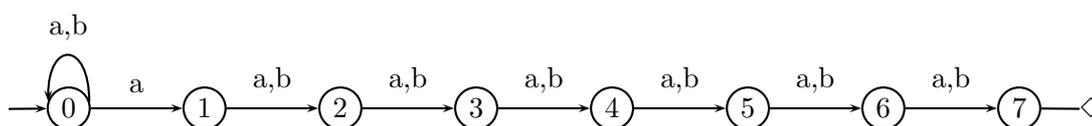


FIG. 9.3 – L'automate \mathcal{P}_7

Cet automate est non-déterministe car il existe deux façons de réagir lorsqu'on est dans l'état 0 et qu'on lit la lettre a ; une flèche va vers l'état 1 tandis que l'autre va vers 0. Par ailleurs cet automate est incomplet, car quand on arrive en 7, il n'y a pas moyen de poursuivre, ce qui se traduit par le fait qu'il faut tomber juste pour pouvoir accepter un mot. Bien sûr, on n'est pas autorisé à revenir en arrière et il faut donc deviner le bon moment pour se lancer dans la suite de transitions menant de 0 à 7...

Avec un peu d'entraînement, il est facile de voir que cet automate reconnaît un mot si et seulement si il contient un a en 7^e position avant sa fin...

9.2.2 Langage reconnu par un automate fini

On définit maintenant formellement l'ensemble des mots reconnus (ou acceptés) par un automate fini.

Définition 9.2.2 On appelle configuration d'un automate \mathcal{A} , la donnée d'un mot $w \in A^*$ sur son alphabet A et d'un état $q \in Q$. Une transition δ de l'automate s'applique à une configuration (w, q) si et seulement si une des deux conditions est satisfaite :

- $\delta = (q, \varepsilon, q')$ et alors l'automate peut passer dans la configuration (w, q') ;
- $\delta = (q, \alpha, q')$ et $w = \alpha v$ (commence donc par la lettre α) et alors l'automate peut passer dans la configuration (v, q') (en ayant donc lu la lettre α).

On notera $(w, q) \xrightarrow{\delta} (w, q')$ ou $(w, q) \xrightarrow{\delta} (v, q')$ l'application de la transition δ à la configuration (w, q) et le lecteur remarquera bien qu'il s'agit d'évolution possible, mais non forcée. La notion de calcul se définit comme suit.

Définition 9.2.3 On appelle calcul d'un automate \mathcal{A} toute suite de configurations c_0, c_1, \dots, c_k , telles que l'on peut passer de c_i à c_{i+1} par une transition de \mathcal{A} , pour $0 \leq i < k$.

Un calcul accepte un mot w si et seulement sa première configuration est $c_0 = (w, q_0)$ et sa dernière est $c_k = (\varepsilon, q)$, pour quelque $q \in F$ état final.

Remarquons donc qu'un calcul acceptant un mot de longueur n est de longueur au moins n , puisqu'il faut que l'automate lise toutes les lettres du mot. Sur l'exemple de la figure 9.3 le mot $f = \text{abbababbaaab}$ admet plusieurs suites de calcul :

- $C_1 : (\text{abbababbaaab}, 0) \rightarrow (\text{bbababbaaab}, 1) \rightarrow (\text{bababbaaab}, 2) \rightarrow (\text{ababbaaab}, 3) \rightarrow (\text{babbaaab}, 4) \rightarrow (\text{abbaaab}, 5) \rightarrow (\text{bbaaab}, 6) \rightarrow (\text{baaab}, 7)$ qui se bloque car il n'y a plus de transition applicable dans l'état 7 et le mot d'entrée n'a pas été lu entièrement ;
- $C_2 : (\text{abbababbaaab}, 0) \rightarrow (\text{bbababbaaab}, 0) \rightarrow (\text{bababbaaab}, 0) \rightarrow (\text{ababbaaab}, 0) \rightarrow (\text{babbaaab}, 0) \rightarrow (\text{abbaaab}, 0) \rightarrow (\text{bbaaab}, 1) \rightarrow (\text{baaab}, 2) \rightarrow (\text{aaab}, 3) \rightarrow (\text{aab}, 4) \rightarrow (\text{ab}, 5) \rightarrow (\text{b}, 6) \rightarrow (\varepsilon, 7)$ qui accepte donc le mot f ;
- $C_3 : (\text{abbababbaaab}, 0) \rightarrow (\text{bbababbaaab}, 0) \rightarrow (\text{bababbaaab}, 0) \rightarrow (\text{ababbaaab}, 0) \rightarrow (\text{babbaaab}, 0) \rightarrow (\text{abbaaab}, 0) \rightarrow (\text{bbaaab}, 0) \rightarrow (\text{baaab}, 0) \rightarrow (\text{aaab}, 1) \rightarrow (\text{aab}, 2) \rightarrow (\text{ab}, 3) \rightarrow (\text{b}, 4) \rightarrow (\varepsilon, 5)$ qui s'arrête dans un état non final.

On dira que le mot $f = \text{abbababbaaab}$ est reconnu par \mathcal{A} , car un des calculs au moins l'accepte. Ceci correspond donc au fait qu'il existe une manière correcte de l'analyser.

Définition 9.2.4 Un mot w est reconnu par un automate \mathcal{A} , s'il existe au moins un calcul de \mathcal{A} sur w qui l'accepte. On appelle langage reconnu par un automate \mathcal{A} , et on le note $\mathcal{L}(\mathcal{A})$, l'ensemble des mots reconnus par l'automate.

Si l'automate \mathcal{A} est déterministe, il existe un seul calcul possible par mot et par conséquent on détermine l'appartenance d'un mot quelconque w au langage $\mathcal{L}(\mathcal{A})$ en un temps $O(|w|)$. Si, au contraire, l'automate est non-déterministe, il faut trouver la

bonne analyse (si elle existe) qui permettra de décider positivement de l'appartenance de w au langage reconnu par \mathcal{A} .

Voici une autre façon de définir l'ensemble reconnu par un automate fini. Le lecteur pourra reconnaître des idées et concepts déjà vus en algorithmique des graphes.

Définissons pour chaque état $q \in Q$ le langage L_q comme étant l'ensemble des mots permettant d'atteindre l'état q à partir de l'état initial q_0 . Les transitions de l'automate du type (q, α, q') , pour $q, q' \in Q$ et $\alpha \in A \cup \{\varepsilon\}$, peuvent être interprétées sur le graphe de l'automate comme des arêtes entre états (q et q') étiquetées par la lettre α de A ou le mot vide $\varepsilon : q \xrightarrow{\alpha} q'$. Alors

$$L_q = \{w \mid \exists n \exists q_1, \dots, q_{n+1} = q \in Q, w_0, w_1, \dots, w_n \in A \cup \{\varepsilon\}, \\ q_0 \xrightarrow{w_0} q_1 \xrightarrow{w_1} \dots q_n \xrightarrow{w_n} q_{n+1} = q \text{ \& } w = w_0 w_1 \dots w_n\}$$

On doit noter ici le rôle que peuvent jouer les ε , la longueur de w n'étant pas forcément $n + 1$. Les L_q se définissent récursivement comme le montre la proposition suivante.

Proposition 9.2.1 *Les langages L_q , $q \in Q$, sont définis par*

$$L_q = K_q \cup \{w\alpha \mid \exists q' \in Q, w \in L_{q'} \text{ \& } (q', \alpha, q) \in \Delta\},$$

où $K_{q_0} = \{\varepsilon\}$ et $K_q = \emptyset$ pour $q \neq q_0$.

Preuve : Si w est un mot de $L_{q'}$ et si $(q', \alpha, q) \in \Delta$ est une transition de l'automate alors, il est évident que $w\alpha$ est un élément de L_q , d'après sa définition ; donc $L_q \subset \{w\alpha \mid \exists q' \in Q, \exists w \in L_{q'}, (q', \alpha, q) \in \Delta\}$, lorsque $q \neq q_0$, et $L_{q_0} \subset \{\varepsilon\} \cup \{w\alpha \mid \exists q' \in Q, \exists w \in L_{q'}, (q', \alpha, q_0) \in \Delta\}$, car $\varepsilon \in L_{q_0}$ par construction (et pour amorcer le dispositif).

Réciproquement, si w est un mot de L_q , c'est que lui-même ou son facteur gauche v défini par $w = v\alpha$, $\alpha \in A$, appartiennent à un $L_{q'}$ et vérifient la propriété ci-dessus. D'où la proposition.

Remarquons de plus que ces ensembles sont parfaitement bien définis par récurrence sur la longueur des chemins dans le graphe sous-jacent de l'automate, chemins que l'on peut par exemple exprimer au moyen des puissances de la matrice d'adjacence du graphe, déduite de la table de transitions. Il s'agit des plus petits ensembles satisfaisant la récurrence ; dans certains cas des sur-ensembles de ces ensembles peuvent aussi fournir une solution, mais nous ne la considérerons pas.

Sur ces bases, on pose alors la définition du langage reconnu par l'automate (équivalente à celle donnée informellement plus haut).

Définition 9.2.5 *Le langage reconnu par un automate \mathcal{A} est la réunion des langages*

$$\mathcal{L}(\mathcal{A}) = \cup_{q \in F} L_q.$$

Exemple : Pour l'automate des multi-motifs de la figure 9.2, l'ensemble $\mathcal{L}(\mathcal{A}) = L_{baa} \cup L_{aaba} \cup L_{ababb}$ est l'ensemble des mots *se terminant* par l'un des motifs de $M = \{baa, aaba, ababb\}$.

Comment faudrait-il le modifier pour qu'il reconnaisse l'ensemble de mots *contenant* l'un de ces motifs ?

Remarques :

1. Un mot est donc reconnu par un automate si et seulement s'il étiquette un chemin menant de l'état initial à un état final.
2. On peut aussi voir la définition des L_q comme un système d'équations ensemblistes, dont les L_q forment la plus petite solution vis-à-vis de l'inclusion. Cette idée sera exploitée au chapitre suivant.
3. Les automates sont a priori non-déterministes. Si la relation Δ est fonctionnelle, c'est-à-dire si $(q, \alpha, q'), (q, \alpha, q'') \in \Delta \Rightarrow q' = q''$, alors l'automate est appelé déterministe, comme noté plus haut : il n'y a alors pas d' ε -transitions.

Pour terminer ce paragraphe donnons deux propriétés élémentaires des langages reconnaissables par automates finis.

Propriétés élémentaires :

1. Tout ensemble fini est reconnu par un automate fini : il suffit de construire l'arbre des préfixes des mots du langage.
2. Le sens de lecture n'importe pas : si $L = \mathcal{L}(\mathcal{A})$ est reconnu par l'automate fini \mathcal{A} , alors il existe un automate \mathcal{B} qui reconnaît $L^R = \{w^R \mid w \in L\}$ ¹ La construction de \mathcal{B} consiste à renverser la table de transition de l'automate, comme on le verra en exercice.

9.2.3 Combinaisons d'automates

Appelons \mathcal{Aut} la famille des langages reconnus par des automates finis. Nous avons déjà vu deux des propriétés suivantes.

Proposition 9.2.2 *La classe \mathcal{Aut} des langages reconnus par automate fini contient les langages finis \mathcal{Fin} , et est close par l'opérateur miroir L^R . De plus elle est dénombrable.*

Nous étudions maintenant les opérations ensemblistes classiques. La classe est close par toutes les opérations classiques, mais nous ne traitons pour le moment la complémentation que dans un cas particulier.

Proposition 9.2.3 *La classe \mathcal{Aut} est close par union, concaténation, quasi-inverse, intersection, morphisme et morphisme inverse, complémentation.*

Preuve : Nous allons construire explicitement des automates réalisant ces opérations.

1. Union : Soient $\mathcal{A} = \langle \Sigma, Q^A, q_0^A, F^A, \Delta^A \rangle$ et $\mathcal{B} = \langle \Sigma, Q^B, q_0^B, F^B, \Delta^B \rangle$ deux automates reconnaissant les langages \mathcal{L}^A et \mathcal{L}^B respectivement. Pour simplifier, et sans aucune

¹On rappelle que la notation w^R désigne le mot w lu de droite à gauche de telle sorte que si $w = w_1w_2\dots w_{n-1}w_n$, on a $w^R = w_nw_{n-1}\dots w_2w_1$.

restriction, on peut supposer que les ensembles d'états Q^A et Q^B sont disjoints, ce qu'on peut toujours réaliser par renommage.

On construit alors un automate \mathcal{C} qui reconnaît $\mathcal{L}^C = \mathcal{L}^A \cup \mathcal{L}^B$ comme suit :

- alphabet Σ ;
- états $Q^C = \{q_0^C\} \cup Q^A \cup Q^B$;
- état initial q_0^C ;
- états finaux $F^C = F^A \cup F^B$;
- transitions $\Delta^C = \{(q_0^C, \varepsilon, q_0^A), (q_0^C, \varepsilon, q_0^B)\} \cup \Delta^A \cup \Delta^B$.

L'automate \mathcal{C} contient donc deux copies disjointes de \mathcal{A} et de \mathcal{B} ; de plus les deux ε -transitions supplémentaires permettent de se brancher à volonté sur l'une de ces copies et par conséquent de reconnaître précisément $\mathcal{L}^A \cup \mathcal{L}^B$.

2. Concaténation : On veut donc reconnaître le langage $\mathcal{L}^C = \mathcal{L}^A \cdot \mathcal{L}^B = \{fg \mid f \in \mathcal{L}^A \ \& \ g \in \mathcal{L}^B\}$, les automates \mathcal{A} et \mathcal{B} étant définis comme précédemment. L'automate \mathcal{C} qui reconnaît \mathcal{L}^C est défini comme suit :

- alphabet Σ ;
- états $Q^C = Q^A \cup Q^B$;
- état initial q_0^A ;
- états finaux F^B ;
- transitions $\Delta^C = \Delta^A \cup \Delta^B \cup \{(q_f, \varepsilon, q_0^B) \mid q_f \in F^A\}$.

Ainsi, une suite de calculs acceptant un mot w pour \mathcal{C} commence par reconnaître un mot de \mathcal{L}^A , puisqu'elle commence par l'état q_0^A et passe nécessairement par un état final de \mathcal{A} — seule façon de sauter vers un état final en passant dans \mathcal{B} en ayant, ce faisant, lu un facteur gauche u de w — puis lit le facteur droit restant v qui est précisément reconnu par \mathcal{B} puisque la suite de calculs commence après l' ε -transition $(q_f, \varepsilon, q_0^B)$ par l'état initial de \mathcal{B} et se termine sur un de ses états finaux.

La réciproque est triviale par construction.

3. Quasi-inverse : Cette opération ensembliste est très importante en informatique ainsi qu'en combinatoire des mots. C'est l'opération de fermeture des monoïdes définie en début de chapitre, mais on l'applique désormais à tout ensemble de mots (fini ou infini) : $\mathcal{X}^* = \{\varepsilon\} \cup \mathcal{X} \cup \mathcal{X}^2 \cup \dots \cup \mathcal{X}^k \cup \dots = \bigcup_{k \geq 0} \mathcal{X}^k$.²

Toujours avec les mêmes notations, construisons l'automate \mathcal{C} qui doit reconnaître le langage $\mathcal{L}^C = \mathcal{L}^{A^*}$:

- alphabet Σ ;
- états $Q^C = Q^A \cup \{q_0^C\}$;
- état initial q_0^C ;
- états finaux $\{q_0^C\}$;
- transitions $\Delta^C = \Delta^A \cup \{(q_0^C, \varepsilon, q_0^A)\} \cup \{(q_f, \varepsilon, q_0^C) \mid q_f \in F^A\}$.

L'introduction de l'état q_0^C , à la fois initial et final, permet de prendre en compte le mot vide ε qui appartient au langage \mathcal{L}^C indépendamment de son appartenance à \mathcal{L}^A ;

²Une remarque importante est que le monoïde \mathcal{X}^* ainsi défini n'a plus de raisons d'être libre, ce qui traduit le fait que la décomposition d'un mot de \mathcal{X}^* en facteurs appartenant à \mathcal{X} n'est pas forcément unique, un même mot pouvant avoir plusieurs scansions.

de plus cet état sert de point relai pour séparer chacun des facteurs d'un mot $w \in \mathcal{L}^{\mathcal{C}}$ qui doivent appartenir à $\mathcal{L}^{\mathcal{A}}$. Un mot est reconnu par l'automate \mathcal{C} si et seulement s'il est la concaténation de mots reconnus par \mathcal{A} .

4. Intersection : Soient donc $\mathcal{L}^{\mathcal{A}}$ et $\mathcal{L}^{\mathcal{B}}$ les langages reconnus par les automates \mathcal{A} et \mathcal{B} , définis comme précédemment. Construisons un automate \mathcal{C} qui va réaliser le produit de ces deux automates en permettant de rendre leur comportement synchrone et donc de les faire marcher en parallèle. Pour ce faire on introduit des ε -transitions qui permettent de boucler arbitrairement sur chaque état de \mathcal{A} et \mathcal{B} sans pour autant lire de lettre de la donnée, et même mieux, de pouvoir changer d'état sur l'un sans modifier l'état de l'autre.

Formellement, on ajoute donc respectivement aux ensembles $\Delta^{\mathcal{A}}$ et $\Delta^{\mathcal{B}}$ les transitions $\{(q, \varepsilon, q) \mid \forall q \in Q^{\mathcal{A}}\}$ et $\{(q, \varepsilon, q) \mid \forall q \in Q^{\mathcal{B}}\}$. Ce faisant on ne change en rien les langages $\mathcal{L}^{\mathcal{A}}$ et $\mathcal{L}^{\mathcal{B}}$ reconnus par ces automates modifiés que nous continuons d'appeler \mathcal{A} et \mathcal{B} , pour ne pas alourdir les notations.

On a alors formellement pour l'automate \mathcal{C} :

- alphabet Σ ;
- états $Q^{\mathcal{C}} = Q^{\mathcal{A}} \times Q^{\mathcal{B}}$;
- état initial $(q_0^{\mathcal{A}}, q_0^{\mathcal{B}})$;
- états finaux $F^{\mathcal{C}} = F^{\mathcal{A}} \times F^{\mathcal{B}}$;
- transitions $\Delta^{\mathcal{C}} = \{((q, r), \alpha, (q', r')) \mid \forall q, q', \alpha, r, r' : (q, \alpha, q') \in \Delta^{\mathcal{A}} \& (r, \alpha, r') \in \Delta^{\mathcal{B}}\}$.

On voit donc que si \mathcal{C} reconnaît un mot w , alors les deux projections du calcul sur les composantes \mathcal{A} et \mathcal{B} sont séparément des calculs valides de ces deux automates et donc $w \in \mathcal{L}^{\mathcal{A}} \cap \mathcal{L}^{\mathcal{B}}$.

Inversement, soit w appartenant à ces deux langages ; il est donc reconnu par deux séquences de calcul $\sigma^{\mathcal{A}}$ et $\sigma^{\mathcal{B}}$. Supposons qu'aucune des deux ne contienne d' ε -transition : elles ont donc la même longueur et lisent la même succession de lettres (celles de w) et, en conséquence, on peut les superposer et en faire le produit, transition par transition. De $q_0 \xrightarrow{w_0} q_1 \xrightarrow{w_1} \dots q_n \xrightarrow{w_n} q_{n+1}$ pour \mathcal{A} et $r_0 \xrightarrow{w_0} r_1 \xrightarrow{w_1} \dots r_n \xrightarrow{w_n} r_{n+1}$ pour \mathcal{B} , on obtient ainsi une séquence de calcul pour \mathcal{C} : $(q_0, r_0) \xrightarrow{w_0} (q_1, r_1) \xrightarrow{w_1} \dots (q_n, r_n) \xrightarrow{w_n} (q_{n+1}, r_{n+1})$, qui accepte w si et seulement si les deux séquences le faisaient.

Si l'une ou l'autre (ou les deux) des séquences de \mathcal{A} et \mathcal{B} contiennent des ε -transitions, il suffit (après avoir mis en correspondance les transitions lisant les lettres homologues de w) de placer en vis-à-vis dans l'autre séquence une ε -transition qui ne change pas l'état (ce sont précisément celles qui furent ajoutées en début de construction). Ainsi réalise-t-on l'automate produit qui reconnaît l'intersection des deux langages.

5. Morphisme et morphisme inverse : un morphisme h est une application d'un ensemble de mots \mathcal{A}^* dans un autre \mathcal{B}^* et qui est défini par sa restriction — $h : \mathcal{A} \rightarrow \mathcal{B}^*$ — à l'alphabet \mathcal{A} , puis étendu à tout \mathcal{A}^* par la règle $h(f.g) = h(f).h(g)$.

Montrons que si $\mathcal{L}^{\mathcal{A}}$ est reconnu par \mathcal{A} , il existe un automate \mathcal{C} qui reconnaît $H = h(\mathcal{L}^{\mathcal{A}})$. L'automate \mathcal{C} va reprendre la structure de l'automate \mathcal{A} en remplaçant chaque transition (q, α, q') , telle que $h(\alpha) = \beta_1\beta_2\dots\beta_k \neq \varepsilon$ par l'ensemble de transitions : $\{(q, \beta_1, r_1^{(q,\alpha,q')}), (r_1^{(q,\alpha,q')}, \beta_2, r_2^{(q,\alpha,q')}), \dots, (r_{k-1}^{(q,\alpha,q')}, \beta_k, q')\}$. Si le morphisme efface α ($h(\alpha) =$

ε), on remplace la transition par (q, ε, q') .

Il est facile de voir qu'en prenant des nouveaux états tous distincts, la structure de \mathcal{A} est maintenue dans \mathcal{C} et qu'un mot est accepté par \mathcal{C} si et seulement s'il est l'image d'un mot accepté par \mathcal{A} .

Pour le morphisme inverse (défini par $\mathcal{K} = h^{-1}(\mathcal{L}) = \{w \mid h(w) \in \mathcal{L}\}$), il faut construire un automate *universel*, capable de simuler tout automate fini donné par sa table, ce qui est la démarche symétrique de ce qui vient d'être fait. Voir les exercices pour des éléments de solution.

6. Complémentation (cas simplifié) :

Considérons le langage $\mathcal{L}^{\mathcal{A}}$ reconnu par \mathcal{A} et faisons sur cet automate les hypothèses supplémentaires suivantes (dont on montrera par la suite qu'elles peuvent toujours être satisfaites) :

- \mathcal{A} est déterministe et sans ε -transition ;
- \mathcal{A} est complet, en ce sens que, pour tout état $q \in Q^{\mathcal{A}}$ et toute lettre de l'alphabet $\alpha \in \Sigma$, il existe un état q' unique (car \mathcal{A} est déterministe) et une transition (q, α, q') correspondante dans la table $\Delta^{\mathcal{A}}$.

On a la propriété fondamentale suivante. Soit w un mot quelconque :

— puisque l'automate est complet, il existe au moins un chemin $q_0 \xrightarrow{w} q$ étiqueté par w ;

– puisque l'automate est complet et sans ε -transition, ce chemin est unique.

En conséquence, pour chaque mot w , il y a existence et unicité de l'état produit par la lecture de w à partir de l'état initial $q_0^{\mathcal{A}}$ et dont l'appartenance à $F^{\mathcal{A}}$ détermine l'appartenance de w à $\mathcal{L}^{\mathcal{A}}$. Soit \mathcal{C} l'automate obtenu, à partir de \mathcal{A} , en prenant $F^{\mathcal{C}} = Q^{\mathcal{A}} \setminus F^{\mathcal{A}}$; cet automate reconnaît exactement le complémentaire des mots acceptés par \mathcal{A} : $\mathcal{L}(\mathcal{C}) = Q \setminus \mathcal{L}(\mathcal{A})$.

On prouvera un peu plus tard la fermeture par complémentation en toute généralité.

9.3 Expressions régulières et langages rationnels

Les automates procurent un cadre effectif et calculatoire pour définir des classes de motifs, mais ils ne donnent pas nécessairement et directement une idée de la nature combinatoire de ces motifs. Une bien meilleure formulation est fournie par ce que l'on appelle expressions régulières dans la littérature anglo-saxonne (*regular expressions*) et langages rationnels dans la tradition française. Ces expressions régulières sont un des outils privilégiés de description de motifs utilisés dans les bases de données, les moteurs de recherche du Web et des systèmes documentaires. Ces concepts font donc partie de la culture de base de tout utilisateur éclairé et a fortiori de celle des concepteurs avancés.

9.3.1 Langages rationnels : définitions

La famille des langages rationnels se définit très simplement et de façon minimale comme le montre la définition ci-dessous. Cependant on verra qu'on peut étendre les

outils de description en augmentant ainsi la lisibilité de la définition sans pour autant sortir de la classe des langages rationnels.

Définition 9.3.1 Soit un alphabet fini A . On appelle famille des langages rationnels sur l'alphabet A la plus petite famille de sous-ensembles de A^* contenant les ensembles finis et close par union, produit de concaténation et quasi-inverse : on la note \mathcal{Rat} .

Cette définition concise a été reprise dans le monde de l'informatique sous le nom d'expressions régulières (*regular expressions* en anglais) et figure dans tous les langages de programmation de systèmes : Unix, DOS, VMS, etc.

L'appendice reprend une partie du manuel Linux qui définit les expressions régulières de la commande GNU-grep.

Nous allons considérer maintenant la version minimale, noyau du concept, qui définit les expressions régulières comme des expressions algébriques sur le monoïde des mots construits sur un alphabet A .

Définition 9.3.2 On appelle expression régulière — et on note \mathcal{Reg} leur ensemble — toute formule composée aux moyens des opérations suivantes, à partir d'un ensemble fini A appelé alphabet, des opérateurs $+$, \cdot , $*$ (les deux premiers sont binaires et le troisième unaire), ainsi que des parenthèses :

- tout élément de A est dans \mathcal{Reg} ;
- pour toute paire $e_1, e_2 \in \mathcal{Reg}$, alors $e_1 + e_2 \in \mathcal{Reg}$ et $e_1.e_2 \in \mathcal{Reg}$;
- pour toute $e \in \mathcal{Reg}$, alors $(e) \in \mathcal{Reg}$ et $(e)^* \in \mathcal{Reg}$;
- l'ensemble vide est noté ε .

Le lecteur notera que les parenthèses permettent de lever les ambiguïtés ; on utilisera aussi le fait que l'opération de quasi-inverse $*$ est prioritaire sur l'opération de concaténation \cdot qui est elle-même prioritaire sur l'union $+$. On supprimera généralement les parenthèses autour d'une lettre unique ainsi que les \cdot et on écrira par exemple a^*ba^* au lieu de $((a)^*) \cdot (b)((a)^*)$.

Chaque expression régulière e définit un sous-ensemble — noté $\mathcal{L}(e)$ et appelé langage associé à e — de l'ensemble A^* des mots sur l'alphabet A , comme explicité ci-dessous :

- $\mathcal{L}(\varepsilon) = \{\varepsilon\}$;
- $\mathcal{L}(a) = \{a\}$, pour toute lettre $a \in A$;
- $\mathcal{L}(e_1 + e_2) = \{w \mid w \in \mathcal{L}(e_1) \vee w \in \mathcal{L}(e_2)\}$;
- $\mathcal{L}(e_1.e_2) = \{w_1.w_2 \mid w_1 \in \mathcal{L}(e_1) \ \& \ w_2 \in \mathcal{L}(e_2)\}$;
- $\mathcal{L}(e^*) = \{\varepsilon\} \cup \mathcal{L}(e) \cup (\mathcal{L}(e))^2 \cup (\mathcal{L}(e))^3 \cup \dots$

On a donc de toute évidence $\mathcal{L}(\mathcal{Reg}) = \mathcal{Rat}$, puisqu'il ne s'agit que d'une question de formulation.

On a de plus et de façon élémentaire le premier pas de caractérisation calculatoire de cette définition ensembliste.

Proposition 9.3.1 Les langages définis par des expressions régulières sont reconnaissables par automates finis :

$$\mathcal{Rat} \subset \mathcal{Aut}.$$

Preuve : Elle reprend en grande partie les constructions de la proposition 2.5 et est basée sur une décomposition récursive de l'expression régulière.

1. Si la formule est une lettre, un mot ou encore un ensemble fini, on construit l'automate filiforme ou arborescent qui reconnaît la lettre, le mot ou l'ensemble fini, comme ceci fut fait dans la construction du squelette de l'automate KMP en première partie de chapitre.

2. Si la formule est de type $(e_1 + e_2)$ ou $(e_1.e_2)$ ou $(e_1)^*$, on utilise de façon récursive les automates \mathcal{A}_1 et \mathcal{A}_2 construits pour les expressions e_1 et e_2 et on leur applique les constructions de la proposition 2.5.

9.3.2 La réciproque : le théorème de Kleene

La propriété remarquable de cette section est que la réciproque est vraie, et donc que l'on dispose d'une double caractérisation pour ces ensembles de mots : ensembliste comme nous venons de le faire et calculatoire au moyen des automates finis. La preuve est due à S. C. Kleene et repose sur une méthode d'énumération nouvelle et très intéressante des chemins dans un graphe, que l'on peut exprimer comme suit : montrer que pour tout automate fini \mathcal{A} d'ensemble d'états Q , les langages³ L_q des mots qui étiqùent les chemins dans le graphe de l'automate sont tous rationnels.

De plus on sait produire algorithmiquement une expression régulière pour ces langages comme le montre la proposition suivante.

Proposition 9.3.2 *Étant donné un automate fini \mathcal{A} et un état q de \mathcal{A} , soit L_q l'ensemble des mots des mots permettant d'atteindre l'état q à partir de l'état initial q_0 . Alors $L_q \in \mathcal{Rat}$.*

L'algorithme ci-dessous fournit une expression régulière pour le langage L_q . Preuve :

0. Soit $\mathcal{A} = \langle A, Q, q_*, F, \Delta \rangle$ l'automate, dont on suppose seulement qu'il est sans ε -transitions ; il est donc non-déterministe a priori.

1. Ranger les états de Q dans un ordre quelconque : $Q = \{q_1, q_2, \dots, q_n\}$. L'état initial q_* est l'un de ces états mais son numéro n'a aucune importance.

Poser

$$\begin{aligned} Q_0 &= \emptyset \\ Q_1 &= \{q_1\} \\ Q_2 &= \{q_1, q_2\} \\ &\dots \\ Q_n &= \{q_1, q_2, \dots, q_n\} \end{aligned}$$

2. Initialisation :

Pour $1 \leq i, j \leq n$ faire

$$X_{i,j}^{Q_0} = \{\alpha \in A \mid (q_i, \alpha, q_j) \in \Delta\}.$$

³Revoir la définition formelle au § 2.2.2.

On remarquera que les $X_{i,j}^{Q_0}$ sont des ensembles finis que l'on représentera trivialement par l'expression régulière équivalente : $\{\alpha_1, \alpha_2, \dots, \alpha_k\} \equiv (\alpha_1 + \alpha_2 + \dots + \alpha_k)$.

3. Itération :

Pour $k = 1, n, 1$ faire Pour $1 \leq i, j \leq n$ faire

$$X_{i,j}^{Q_k} = X_{i,j}^{Q_{k-1}} + X_{i,k}^{Q_{k-1}} (X_{k,k}^{Q_{k-1}})^* X_{k,j}^{Q_{k-1}}.$$

Considérons les $X_{i,j}^{Q_1} = X_{i,j}^{Q_0} + X_{i,1}^{Q_0} (X_{1,1}^{Q_0})^* X_{1,j}^{Q_0}$; comme $X_{i,j}^{Q_0}$ est l'ensemble des lettres qui permettent de passer de q_i à q_j par une transition de l'automate, on en déduit que $X_{i,j}^{Q_1}$ est l'ensemble des mots qui amènent de q_i à q_1 , puis bouclent sur q_1 et vont en un coup de q_1 à q_j .

De même, on interprète $X_{i,j}^{Q_2} = X_{i,j}^{Q_1} + X_{i,2}^{Q_1} (X_{2,2}^{Q_1})^* X_{2,j}^{Q_1}$ comme étant l'ensemble des étiquettes des chemins qui vont de q_i à q_2 (premier passage) en passant par q_1 , puis bouclent sur q_2 en passant par q_1 et enfin vont de q_2 (dernier passage) à q_j en passant par q_1 . Remarquons que tout chemin allant de q_i à q_j en passant uniquement par q_1 et q_2 peut être ainsi décomposée et que $X_{i,j}^{Q_2}$ les contient donc tous au travers de l'expression régulière.

Et donc par récurrence, $X_{i,j}^{Q_k}$ contient tous les mots étiquetant des chemins allant de q_i à q_j en passant par les sommets de l'ensemble Q_k .

4. En conclusion, $X_{i,j}^{Q_n}$ contient tous les mots étiquetant des chemins allant de q_i à q_j en passant par les sommets de l'ensemble Q_n et par conséquent représente toutes les façons de passer de q_i à q_j . Comme cas particulier, on a donc pour tout état $q \in Q$,

$$L_q = X_{q^*,q}^{Q_n}.$$

En conséquence tout automate fini produit uniquement des langages rationnels, expressibles donc par expression régulière. D'où le théorème annoncé.

Théorème 9.3.1 [Kleene] *Les ensembles rationnels sont identiques aux ensembles reconnaissables par automate fini :*

$$\mathcal{A}ut = \mathcal{R}at.$$

Le nombre de phases de calcul de cet algorithme est $n = |Q|$, chaque phase comportant le calcul de n^2 expressions ; on pourrait donc en déduire que le temps de calcul est polynomial. Or il importe de noter que la taille de ces expressions peut quadrupler à chaque phase, ce qui fait que la longueur de l'expression pour $X_{i,j}^{Q_k}$ est $O(4^k \cdot |A|)$ et donc que si l'on procède par recopie le temps de calcul sera $O(n^2 4^n)$... De même, le simple fait d'écrire le résultat peut être exponentiel en $|Q|$.

9.4 Déterminisme contre Non-Déterminisme

La définition des automates finis comme reconnaisseurs non-déterministes est fondamentale du point de vue de leurs propriétés de composition et de pouvoir expressif.

Cependant, il faut bien reconnaître que du point de vue pratique de l'informaticien qui *calcule* le non-déterminisme est source de frustration et d'inefficacité. Nous allons développer plusieurs réponses à cette critique, visant à rendre déterministe le calcul du résultat attendu de l'automate non-déterministe.

9.4.1 Simulation du non-déterminisme

Considérons un automate fini $\mathcal{A} = \langle A, Q, q_*, F, \Delta \rangle$, dont on suppose seulement qu'il est sans ε -transitions, pour simplifier. Le mécanisme de l'automate fini est simulé au moyen d'une mémoire de taille Q qui contiendra à chaque pas de simulation le sous-ensemble des états accessibles depuis l'état initial. Plus précisément, si après avoir lu la k^{e} lettre de la donnée w , l'ensemble des états accessibles est

$$H_k = \{q \mid q_* \xrightarrow{w_1 \dots w_k} q\}$$

et si α est la lettre suivante de w , l'ensemble H_{k+1} est défini par :

$$H_{k+1} = \{q' \mid \exists q \in H_k \ \& \ (q, \alpha, q') \in \Delta\}.$$

Le programme qui effectue la simulation sur un mot w , vu comme une liste de lettres, a donc la forme suivante :

```

H <- {q*}; // état initial
répéter
  l <- tête(w); // la première lettre de w
  w <- queue(w); // qui est supprimée
  K <- {};
  pour q dans H faire
    pour (q, l, qq) dans Delta faire
      K <- K ∪ {qq};
    H := K; // recopie
jusqu'à ce que w == NIL;
si H contient un état de F alors
  accepter(w);

```

Ce programme opère en $O(|w| \cdot |\Delta|)$ opérations, et même un peu moins à condition de pouvoir faire une recherche rapide des transitions applicables à l'ensemble H courant.

Proposition 9.4.1 *La simulation d'un automate non-déterministe \mathcal{A} sur un mot de longueur n se fait en temps $O(n \cdot |\mathcal{A}|)$.*

9.4.2 Suppression des ε -transitions

Nous avons plusieurs fois fait l'hypothèse simplificatrice que l'automate ne comportait pas de transition du type (q, ε, q') . Nous montrons dans ce paragraphe comment on peut effectivement transformer tout automate fini \mathcal{A} en un automate fini équivalent et qui ne contient pas d' ε -transitions.

Le principe de l'opération est simple : soit un état q tel qu'il existe deux transitions, (q', α, q) pour un certain $\alpha \in A$ en amont et (q, ε, q'') en aval. On voit que l'on peut ajouter la transition (q', α, q'') à la table sans changer pour autant l'ensemble reconnu par \mathcal{A} . Si l'on répète cette opération pour toutes les suites d' ε -transitions issues de q et toutes les transitions amenant à q , on peut alors retirer la transition (q, ε, q'') . L'automate ainsi obtenu est équivalent à \mathcal{A} . D'où l'algorithme.

Élimination des ε -transitions :

1. Calculer pour chaque état q de l'automate la liste des états qui en sont accessibles par une suite d' ε -transitions : E_q .
2. Pour chaque transition (q', α, q) , $\alpha \neq \varepsilon$, ajouter les transitions (q', α, q'') pour tout $q'' \in E_q$.
3. Supprimer toutes les ε -transitions.

Le nouvel automate \mathcal{B} ainsi obtenu est équivalent à \mathcal{A} car pour tout mot w , si \mathcal{A} le reconnaît :

- par une suite de calcul sans ε -transitions, cette suite existe aussi dans \mathcal{B} ;
- par une suite de calcul avec ε -transitions, une suite existe aussi dans \mathcal{B} qui est obtenue en remplaçant les séquences maximales du type $q \xrightarrow{\alpha} q_1 \xrightarrow{\varepsilon} q_2 \xrightarrow{\varepsilon} \dots \xrightarrow{\varepsilon} q'$ dans \mathcal{A} par $q \xrightarrow{\alpha} q'$ dans \mathcal{B} .

Inversement, toute suite dans \mathcal{B} possède par construction son homologue avec ε -transitions dans \mathcal{A} .

9.4.3 Suppression du non-déterminisme

Reprenons le § 2.4.1 et exprimons les choses un peu différemment : l'ensemble des valeurs possibles pour la liste H est fini, puisqu'il correspond à l'ensemble des parties $\mathcal{P}(Q) = 2^Q$ de l'ensemble des états Q . L'algorithme consiste en une boucle qui lit séquentiellement les lettres de la donnée w et transforme la valeur de H en passant d'un élément de $\mathcal{P}(Q)$ à un autre. Ceci n'est finalement rien d'autre que le comportement d'un automate fini que nous définissons maintenant formellement.

Soit $\mathcal{A} = \langle A, Q, q_*, F, \Delta \rangle$ un automate fini sans ε -transitions. Soit $\hat{\mathcal{A}} = \langle A, \mathcal{P}(Q), \{q_*\}, \hat{F}, \hat{\Delta} \rangle$ l'automate fini défini par $\hat{F} = \{H \in \mathcal{P}(Q) \mid H \cap F \neq \emptyset\}$ pour ensemble d'états finaux et $\hat{\Delta} = \{(H, \alpha, H') \mid H \in \mathcal{P}(Q) \ \& \ H' = \{q' \mid q \in H \ \& \ (q, \alpha, q') \in \Delta\}\}$ pour table de transitions. L'automate $\hat{\mathcal{A}}$ est appelé *automate des parties* et il est évidemment déterministe par construction.

C'est alors une simple affaire de rephrasage de la preuve de correction de l'algorithme du § 2.4.1 que de prouver que les deux langages $\mathcal{L}(\mathcal{A})$ et $\mathcal{L}(\hat{\mathcal{A}})$ sont identiques.

Théorème 9.4.1 *Tout automate fini est équivalent à un automate fini déterministe : l'automate des parties.*

Ce résultat est remarquable d'un point de vue théorique. Il permet en particulier de démontrer en toute généralité la propriété de clôture de la famille \mathcal{Rat} par complémentation. Il faut cependant être conscient du fait que la taille de l'ensemble des parties $\mathcal{P}(Q)$ est $2^{|Q|}$ et donc devient rapidement prohibitive. Les Unixiens confirmés auront à cœur de comprendre les limitations imposées à la commande `grep`.

9.5 Complexité et limites

9.5.1 Déterminisation

Nous donnons dans ce paragraphe un exemple où le saut exponentiel de taille de l'automate déterministe par rapport à celle du non-déterministe est effectivement atteinte.

Considérons sur l'alphabet $A = \{a, b\}$ la famille des langages $P_n = (a + b)^* a (a + b)^{n-1}$, pour $n \geq 1$.

1. L'ensemble P_n est trivialement reconnu par un automate non-déterministe à $n + 1$ états, comme le montre la figure 9.3 dans le cas $n = 7$.

L'automate \mathcal{P}_n associé est défini comme suit :

- alphabet : $A = \{a, b\}$;
- états : $Q = \{0, 1, \dots, n\}$;
- état initial : 0 ;
- états finaux : $F = \{n\}$;
- table : $\Delta = \{(0, a, 0), (0, b, 0), (0, a, 1)\} \cup \{(i, a, i + 1), (i, b, i + 1) \mid 1 \leq i < n\}$.

2.1 Le langage P_1 est reconnu par un automate déterministe \mathcal{P}_1 à 2 états $Q_1 = \{\varepsilon, a\}$, d'état initial ε , d'état final $\{a\}$ et de table $\Delta_1 = \{(\varepsilon, a, a), (\varepsilon, b, \varepsilon), (a, a, a), (a, b, \varepsilon)\}$. Il est facile de voir que ces 2 états sont nécessaires car seuls les mots se terminant par a sont acceptés.

2.2 Le langage P_2 est reconnu par un automate déterministe \mathcal{P}_2 à 4 états $Q_2 = \{\varepsilon, a, aa, ab\}$, d'état initial ε , d'états finaux $\{aa, ab\}$ et de table $\Delta_2 = \{(\varepsilon, a, a), (\varepsilon, b, \varepsilon), (a, a, aa), (a, b, ab), (aa, a, aa), (aa, b, ab), (ab, a, a), (ab, b, \varepsilon)\}$. De nouveau ces 4 états sont nécessaires car chacun mémorise des situations distinctes pour les deux dernières lettres lues à savoir en régime permanent et dans l'ordre des états bb, ba, aa, ab , et qui ne peuvent pas être fusionnées sous peine de reconnaître des mots n'appartenant pas au langage.

2.3 Donnons comme ultime exemple l'automate déterministe \mathcal{P}_3 à 8 états qui reconnaît P_3 :

- $Q_3 = \{\varepsilon, a, aa, ab, aaa, aab, aba, abb\}$ est l'ensemble des états ;
- ε est l'état initial ;
- $F = \{aaa, aab, aba, abb\}$ est l'ensemble des états finaux ;
- les transitions sont données par $\Delta_3 = \{(\varepsilon, a, a), (\varepsilon, b, \varepsilon), (a, a, aa), (a, b, ab), (aa, a, aaa), (aa, b, aab), (ab, a, aba), (ab, b, abb), (aaa, a, aaa), (aaa, b, aab), (aab, a, aba), (aab, b, abb), (aba, a, aa), (aba, b, ab), (abb, a, a), - (abb, b, \varepsilon)\}$. Si l'on prend deux états quelconques, il est toujours possible de trouver un mot pour lesquels le résultat du calcul est différent et par conséquent tous les états sont nécessaires : par exemple avec les états aa et abb , le mot a conduit à un état final pour le premier et pas pour le second.

En règle générale, l'automate déterministe \mathcal{P}_n qui reconnaît P_n comporte 2^n états que l'on dénotera par les mots de longueur au plus n commençant par a plus le mot vide :

- $Q_n = \{\varepsilon, a, aa, ab, aaa, aab, aba, abb, \dots, a^n, a^{n-1}b, a^{n-2}ba, a^{n-2}bb, \dots, ab^{n-2}a, ab^{n-1}\}$;

- ε est l'état initial ;
- $F = \{a^n, a^{n-1}b, a^{n-2}ba, a^{n-2}bb, \dots, ab^{n-2}a, ab^{n-1}\}$ est l'ensemble des états finaux ;
- les transitions sont données par les règles suivantes exécutées dans cet ordre : $(\varepsilon, a, a) \in \Delta$ et $(\varepsilon, b, \varepsilon) \in \Delta$; si q est un état de longueur inférieure strictement à n , $(q, a, qa) \in \Delta$ et $(q, b, qb) \in \Delta$; si q est de longueur n et $q = aaf$ alors $(q, a, afa) \in \Delta$ et $(q, b, afb) \in \Delta$; si q est de longueur n et $q = ab^k f$, f ne commençant pas par b , alors $(q, a, fa) \in \Delta$ et $(q, b, fb) \in \Delta$.

Le lecteur reprendra les arguments ci-dessus pour se convaincre de la justesse de la proposition.

Proposition 9.5.1 *L'automate déterministe minimal équivalent à un automate non-déterministe de taille n peut avoir $\Omega(2^n)$ états.*

9.5.2 Automate minimal

Le dernier concept de cette section est qu'il existe une forme canonique des automates déterministes qui "contient" en un sens mathématiquement précis toutes les autres et qui est appelée "automate minimal". Dans toute la suite on suppose que les automates ont été préalablement nettoyés de leurs ε -transitions et que tous leurs états sont utiles, c'est-à-dire sont accessibles depuis l'état final et peuvent conduire à un état final.

Indiscernabilité

Oublions cependant les automates et considérons un langage $L \subset A^*$ quelconque. Ce langage permet de définir une relation d'équivalence sur les mots de A^* que l'on appelle la *L-indiscernabilité*.

Définition 9.5.1 *Soit $L \subset A^*$ un langage. On dit que deux mots x et y de A^* sont L-indiscernables si et seulement si*

$$\forall w \in A^* \quad xw \in L \Leftrightarrow yw \in L.$$

On note $x \equiv_L y$ la relation de L-indiscernabilité.

Il est évident que \equiv_L est une relation d'équivalence quel que soit le langage L .

Classes et automates finis

Considérons maintenant un automate fini *déterministe* \mathcal{A} , son ensemble d'états Q et notons L le langage qu'il reconnaît. Rappelons que $L_q = \{x \mid q_* \xrightarrow{x} q\}$ est l'ensemble des mots qui amènent de l'état initial dans l'état q .

Deux mots x et y d'un même L_q sont L-indiscernables, puisque pour tout mot w , l'automate agit de la même façon sur xw et yw , le résultat étant dans les deux cas celui du calcul de \mathcal{A} sur w à partir de l'état q .

De plus, soient deux états q et q' distincts et donc tels que $L_q \cap L_{q'} = \emptyset$ (puisque \mathcal{A}

est déterministe) ; alors s'il existe $x \in L_q$ et $x' \in L_{q'}$ tels que $x \equiv_L x'$, alors les deux ensembles tout entiers le sont : $L_q \equiv_L L_{q'}$, ce que nous écrirons plus simplement au niveau des états, $q \equiv_L q'$. Les deux états q et q' sont L -indiscernables.

Si deux états d'un automate \mathcal{A} sont $\mathcal{L}(\mathcal{A})$ -indiscernables, on peut donc simplifier cet automate en leur donnant le même nom et en modifiant la table de transitions en conséquence. Un critère simple d'indiscernabilité est le suivant, dont la preuve est évidente.

Lemme 9.5.1 *Soient deux états q et q' d'un automate fini déterministe \mathcal{A} qui vérifient la propriété : $\forall \alpha \in A, (q, \alpha, r) \in \Delta \Leftrightarrow (q', \alpha, r) \in \Delta$. Alors q et q' sont $\mathcal{L}(\mathcal{A})$ -indiscernables.*

On construit donc à partir de cette remarque un automate qui ne peut plus être réduit et qu'on appelle *automate minimal*. Cet algorithme construit l'automate minimal à partir des états finaux, puis par accréation des états indiscernables ; il sera détaillé en exercices. On montre maintenant que cet automate est unique car il ne dépend en fait que du langage et non de l'automate.

Langages rationnels

Tout langage rationnel $L \subset A^*$ étant reconnaissable par un automate fini détermine ce faisant une partition finie de A^* en classes de L -indiscernabilité : par exemple, celle déterminée par les états de l'automate fini.

Inversement, considérons un langage L dont les classes d'indiscernabilité

$$C_1, C_2, \dots, C_p \subset A^*$$

sont en nombre fini p . On a donc :

- $C_i \cap C_j = \emptyset$, pour tous $1 \leq i, j \leq p$;
- $\cup_{1 \leq i \leq p} C_i = A^*$;
- $\forall x, y \in C_i, x \equiv_L y$.

Alors L est union de certaines des classes C_i , puisque chaque classe est soit contenue dans L soit disjointe de lui (en prenant $w = \varepsilon$ dans la définition des classes). Par ailleurs pour chaque lettre $\alpha \in A$ et chaque classe C_i il existe un j unique tel que $C_i \alpha \subset C_j$, car $x \equiv_L y \Rightarrow x\alpha \equiv_L y\alpha$.

Considérons alors l'automate fini \mathcal{A} dont :

- les états sont les C_i ,
- l'état initial celui des C_i qui contient le mot vide ε ,
- les états finaux sont les $C_i \subset L$,
- la table est définie par les (i, α, j) tels que $C_i \alpha \subset C_j$.

Cet automate reconnaît à l'évidence le langage L et est fini. D'où le théorème.

Théorème 9.5.1 *Un langage $L \subset A^*$ est rationnel si et seulement s'il existe une partition finie de A^* en classes L -indiscernables.*

Automate minimal

Combinons maintenant les deux derniers paragraphes. Tout automate fini \mathcal{A} reconnaît un langage rationnel L dont l'ensemble des classes d'indiscernabilité est fini et correspond à la *partition la plus grossière* selon \equiv_L (celle qui possède le moins de classes). Les classes L_q de \mathcal{A} sont donc des sous-ensembles de ces classes d'après les raisonnements des paragraphes précédents et il est donc impossible de trouver un automate plus petit que celui construit sur les classes d'indiscernabilité. D'où le théorème.

Théorème 9.5.2 [Myhill, Nerode]

Tout langage rationnel admet un unique automate déterministe minimal : celui défini par les classes d'indiscernabilité du langage.

9.6 Langages non rationnels : le lemme d'itération

Ce théorème donne une indication sur la structure répétitive inévitable des mots des langages rationnels : on l'appelle encore *Lemme de l'étoile* en raison de la forme de son énoncé.

Théorème 9.6.1 *Soit $L \in \mathcal{Rat}$ un langage rationnel infini. Il existe un entier λ tel que, pour tout mot $w \in L$, $|w| \geq \lambda$, w peut se factoriser en $w = xyz$, $y \neq \varepsilon$ et tel que*

$$xy^*z \subset L.$$

En d'autres termes, le langage infini des mots obtenus en itérant un nombre arbitraire de fois le facteur y est inclus dans L ; ou encore, les mots assez longs de tout langage rationnel contiennent des formes de périodicités.

Preuve : Soit \mathcal{A} un automate (déterministe) reconnaissant L ; soit λ son nombre d'états et considérons pour un mot w du langage, de longueur $|w| \geq \lambda$, le calcul qui le reconnaît :

$$q_* \xrightarrow{w_1} q_1 \xrightarrow{w_2} q_2 \xrightarrow{w_3} \dots \xrightarrow{w_n} q_f.$$

Il y a strictement plus de λ occurrences d'états dans cette chaîne et donc deux au moins sont identiques, par exemple à q ; soit y le facteur de w qui les sépare :

$$q \xrightarrow{y} q.$$

On peut répéter cette séquence et obtenir une nouvelle séquence de calcul qui valide un mot plus long, de même qu'on peut la couper sans altérer la validité du calcul, de sorte que toutes les séquences du type

$$q_* \xrightarrow{x} q \left(\xrightarrow{y} q \right)^* \xrightarrow{z} q_f,$$

sont valides et reconnaissent précisément les mots du type xy^*z définis dans l'énoncé.

Ce résultat est un outil puissant pour montrer que certains langages ne sont pas rationnels. Nous donnons à titre d'exemple un corollaire qui sera repris dans le chapitre suivant.

Corollaire 9.6.1 *Le langage $P = \{a^n b^n \mid n \in \mathbb{N}\}$ n'est pas rationnel.*

Preuve : Soient $x, y, z \in \{a, b\}^*$, avec $y \neq \varepsilon$, tels que $xy^*z \subset P$. Alors y doit contenir autant de a que de b , les a précédant les b , $y = a^l b^l$; comme ce nombre l n'est pas nul, $yy = a^l b^l a^l b^l$, et donc le mot xy^2z n'est pas dans P . Contradiction.

9.7 Appendice

Nous reproduisons ci-dessous la définition des expressions régulières pour la commande `grep` de Linux, telle qu'elle apparaît dans le manuel de référence. Cette commande construit un automate à partir de l'expression régulière puis teste si le fichier proposé en second argument contient des facteurs qui satisfont cette expression. Il s'agit d'expressions étendues, mais qui restent équivalentes à des expressions usuelles.

REGULAR EXPRESSIONS

A regular expression is a pattern that describes a set of strings. Regular expressions are constructed analogously to arithmetic expressions, by using various operators to combine smaller expressions. Grep understands two different versions of regular expression syntax: "basic" and "extended." In GNU grep, there is no difference in available functionality using either syntax. In other implementations, basic regular expressions are less powerful. The following description applies to extended regular expressions; differences for basic regular expressions are summarized afterwards.

The fundamental building blocks are the regular expressions that match a single character. Most characters, including all letters and digits, are regular expressions that match themselves. Any metacharacter with special meaning may be quoted by preceding it with a backslash.

A list of characters enclosed by [and] matches any single character in that list; if the first character of the list is the caret ^ then it matches any character not in the list. For example, the regular expression [0123456789] matches any single digit. A range of ASCII characters may be specified by giving the first and last characters, separated by a hyphen. Finally, certain named classes of characters are predefined. Their names are self explanatory, and they are [:alnum:], [:alpha:], [:cntrl:], [:digit:], [:graph:], [:lower:], [:print:], [:punct:], [:space:], [:upper:], and [:xdigit:]. For example, [[:alnum:]] means [0-9A-Za-z], except the latter form is dependent upon the ASCII character encoding, whereas the former is portable. (Note that the brackets in these class names are part of the symbolic names, and must be included in addition to the brackets delimiting the bracket list.) Most metacharacters lose their special meaning inside lists. To include a literal] place it first in the list. Similarly, to include a literal ^ place it anywhere but first. Finally, to include a literal - place it last.

The period `.` matches any single character. The symbol `\w` is a synonym for `[[[:alnum:]]]` and `\W` is a synonym for `[^[:alnum:]]`.

The caret `^` and the dollar sign `$` are metacharacters that respectively match the empty string at the beginning and end of a line. The symbols `<` and `>` respectively match the empty string at the beginning and end of a word. The symbol `\b` matches the empty string at the edge of a word, and `\B` matches the empty string provided it's not at the edge of a word.

A regular expression may be followed by one of several repetition operators:

- ? The preceding item is optional and matched at most once.
- * The preceding item will be matched zero or more times.
- + The preceding item will be matched one or more times.
- {n} The preceding item is matched exactly n times.
- {n,} The preceding item is matched n or more times.
- {n,m} The preceding item is matched at least n times, but not more than m times.

Two regular expressions may be concatenated; the resulting regular expression matches any string formed by concatenating two substrings that respectively match the concatenated subexpressions.

Two regular expressions may be joined by the infix operator `|`; the resulting regular expression matches any string matching either subexpression.

Repetition takes precedence over concatenation, which in turn takes precedence over alternation. A whole subexpression may be enclosed in parentheses to override these precedence rules.

The backreference `\n`, where `n` is a single digit, matches the substring previously matched by the `n`th parenthesized subexpression of the regular expression.

In basic regular expressions the metacharacters `?`, `+`, `{`, `|`, `(`, and `)` lose their special meaning; instead use the backslashed versions `\?`, `\+`, `\{`, `\|`, `\(`, and `\)`.

Exercices

1. Décrire les langages engendrés par les expressions régulières suivantes :
 - $0(0+1)^*0$;
 - $((\varepsilon+0)1^*)^*$;
 - $(0+1)^*0(0+1)(0+1)$;
 - $0^*10^*10^*10^*$;
 - $(00+11)^*((01+10)(00+11)^*(01+10)(00+11)^*)^*$.
2. Écrire des expressions régulières pour les langages suivants :
 - tous les mots qui contiennent les 5 voyelles dans l'ordre ;
 - toutes les suites de chiffres sans qu'aucun ne soit répété ;
 - toutes les suites de chiffres sans qu'aucun ne soit répété consécutivement ;
 - les mots qui ne contiennent pas le facteur 011 ;
 - les mouvements du jeu d'échecs.
3. Construire les automates qui reconnaissent les langages :
 - $(a+b)^*$;
 - $(a^*+b^*)^*$;
 - $((\varepsilon+a)b^*)^*$;
 - $(a+b)^*abb(a+b)^*$.
 Montrer leurs calculs possibles sur le mot *ababbab*.
4. Construire des automates déterministes équivalents.
5. Les faire aussi petits que possible.
6. Montrer que $(a+b)^*$, $(a^*+b^*)^*$ et $((\varepsilon+a)b^*)^*$ engendrent le même langage.
7. Les langages suivants sont-ils rationnels ?
 - i) $\{w \in (a+b)^* \mid |w|_a = |w|_b\}$, en notant $|w|_a$ le nombre de a dans w ;
 - ii) $\{a^i b^j \mid i \neq j\}$;
 - iii) $\{a^i b^j \mid i = j \pmod k\}$, pour k fixé ;
 - iv) $\{c^k a^i b^j \mid i = j \pmod k\}$.
8. Dans un automate fini certains états peuvent être inutiles ; expliquer pourquoi. Peut-on les éliminer algorithmiquement ? Comment ?
9. Peut-on déterminer si le langage reconnu par un automate fini est non vide ? Estimer la complexité de mise en œuvre de la réponse.
10. Peut-on déterminer algorithmiquement si deux automates engendrent le même langage ?
11. Écrire complètement l'algorithme d'élimination des ε -transitions d'un automate.

Chapitre 10

Parenthésages, Langages algébriques

Le lemme de l'étoile montre clairement les limites des expressions régulières et des automates finis pour exprimer des propriétés structurelles à distance dans les mots. Un grand nombre de cas pratiques relèvent pourtant de ce type de contraintes : expressions arithmétiques et logiques, exécution des appels de fonctions dans les langages de programmation acceptant la récursivité, démontage et remontage d'assemblages complexes de pièces mécaniques, structures biologiques pour certains ARN et protéines, structures mathématiques en combinatoire et en physique statistique, etc. Développer un formalisme permettant de décrire et d'analyser ces structures emboîtées est donc une nécessité. N. Chomsky, le premier, a compris qu'il fallait dépasser le formalisme des expressions régulières pour exprimer les structures grammaticales des langues naturelles, en particulier celles des langues indo-européennes ; il a ainsi construit le premier modèle de "grammaires hors-contexte" (*contextfree grammars*) et en a étudié les propriétés mathématiques et informatiques avec M.-P. Schützenberger, à qui l'on doit de nombreux développements. La théorie s'est aussi imposée, peu après, dans la conception des compilateurs, et donc dans celle des langages de programmation. Elle est devenue un point de passage inévitable et est intégrée, sous des formes particulières, dans les langages et systèmes, avec par exemple le concept de *stream* et de *parser* de Caml et le module *Yacc* d'Unix. Ce chapitre est consacré aux bases conceptuelles ; le suivant sera une brève introduction à la construction d'analyseurs syntaxiques, comme on en trouve dans les compilateurs.

10.1 Exemples de structures parenthésées

10.1.1 Notations

La structure la plus évidente est celle des expressions arithmétiques, qu'elle soit classique, comme dans

$$(a + b)(a^2 - 2ab + b^2),$$

ou informatique comme dans

$$(a+b) * (a*a-2*a*b+b*b),$$

où le lecteur remarquera que l'on a déjà utilisé des conventions sur la " priorité " des opérateurs pour évaluer sans ambiguïté les sommes de produits.

Dans cette notation dite " infixe ", car les opérateurs sont placés entre les opérandes, on doit préciser (à défaut de tout parenthéser) que les multiplications et divisions sont exécutées *avant* les additions et soustractions elles-mêmes prioritaires sur l'opposé (– unaire), et que les opérations de même priorité sont exécutées de gauche à droite. Dans ces conditions le système est interprétable de façon unique. Dès les années 20, les logiciens, en particulier ceux de l'école polonaise avaient mis au point des systèmes de notation non ambigus et sans parenthèses : les formes polonaises préfixée ou postfixée encore appelées respectivement notations préfixe et postfixe. Ainsi pour l'expression qui nous sert d'exemple a-t-on en forme préfixée

$$**ab+ -*aa*ab*bb,$$

et en forme postfixée

$$ab+aa*ab*-bb**+.$$

Toutes ces représentations découlent en fait naturellement d'une seule et même notation qui voit et implante les expressions sous forme d'arbres.

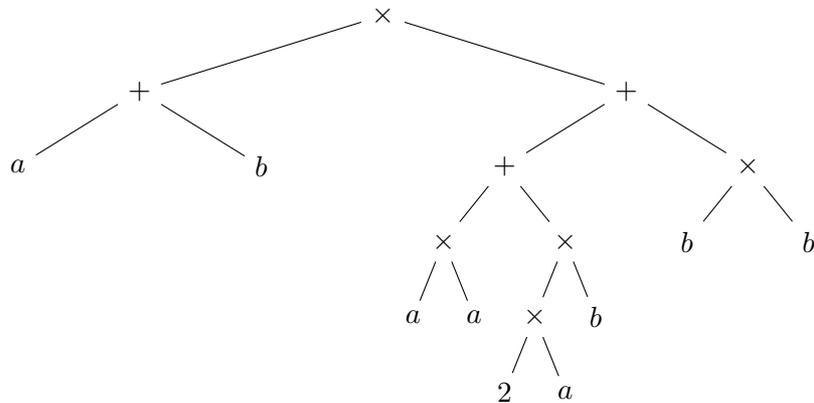


FIG. 10.1 – L'arbre pour l'expression $(a+b)(a^2-2ab+b^2)$.

10.2 Grammaires hors-contexte

Le problème est d'étendre le mécanisme des automates finis et des expressions régulières. On verra dans la section suivante comment on peut procéder au moyen

d'automates. Dans cette section on revient d'abord sur les langages rationnels associés à un automate et on dégage le concept de grammaire de réécriture qui va être étendu pour définir une classe plus importante de langages.

10.2.1 Automates, langages et grammaires

À partir d'un automate fini $\mathcal{A} = \langle A, Q, q_0, Q_f, \Delta \rangle$, on a vu au chapitre précédent qu'on définissait les langages L_q , pour chaque état $q \in Q$, constitués des mots qui étiquettent un chemin dans le graphe de l'automate amenant de l'état initial q_0 à l'état q . On a vu également que ces langages se définissent récursivement par un système de relations du type

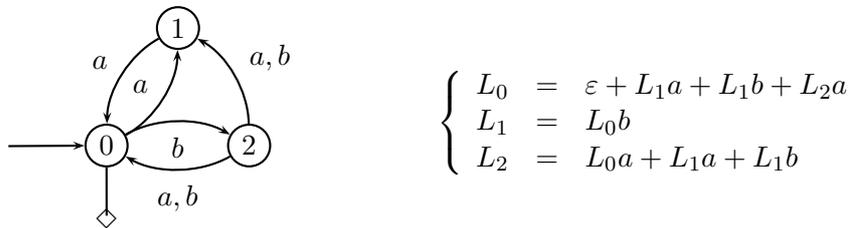
$$L_q = \{wa \mid a \in A \cup \{\varepsilon\} \ \& \ \exists q' \ w \in L_{q'} \ \& \ \langle q', a, q \rangle \in \Delta\},$$

ce qui s'écrit encore, en utilisant la concaténation ensembliste,

$$L_q = \bigcup_{\langle q', a, q \rangle \in \Delta} L_{q'}a,$$

en ajoutant pour l'état initial que $\varepsilon \in L_{q_0}$.

Par exemple, on aura le système suivant pour l'automate dessiné en vis-à-vis :



Comme on l'a fait dans les expressions régulières, on utilise le + pour noter l'union ensembliste, et la concaténation est notée par la simple juxtaposition.

Parcourir un chemin dans l'automate est équivalent à une réécriture depuis une "variable" L_q correspondant à un état final — 0 en l'occurrence. On notera que cette réécriture se fait en sens inverse du parcours du graphe de l'automate :

$$L_0 \rightarrow L_1a \rightarrow L_0ba \rightarrow L_2aba \rightarrow L_0aaba \rightarrow aaba.$$

Considérons maintenant le système plus compliqué suivant :

$$L = aLb + \varepsilon.$$

Une suite de réécritures conservant l'ordre des facteurs dans la concaténation donne typiquement :

$$L \rightarrow aLb \rightarrow aaLbb \rightarrow aaaLbbb \rightarrow aaabbb,$$

ce qui par récurrence immédiate nous permet de caractériser et d'engendrer le langage $\{a^n b^n \mid n \geq 0\}$ qui demeurerait inaccessible dans le cadre des expressions régulières.

De la même façon, le système de règles \mathcal{D} , défini par

$$S = (S)S + x,$$

permet de définir les mots correctement parenthésés.

Comme précédemment, il est judicieux de construire un arbre qui explicite les récritures successives, comme le montre la figure 10.2.

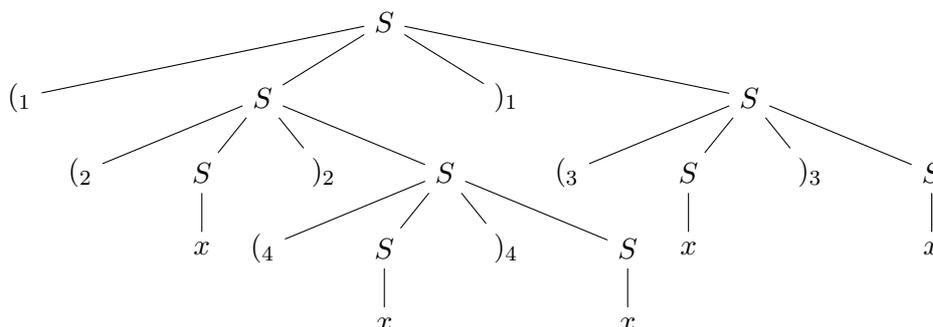


FIG. 10.2 – Un arbre de dérivation pour $(1(2x))2(3x)3x)1(4x)4x$: on a numéroté les parenthèses pour mettre en évidence les paires ouvrante-fermante homologues.

10.2.2 Terminologie

Définition 10.2.1 Une grammaire hors-contexte (context-free dans la terminologie de Chomsky) est un système de récriture \mathcal{G} défini par :

- V , un ensemble fini de variables (ou non-terminaux) ;
- T , un ensemble fini de terminaux (alphabet du langage) ;
- $\mathcal{R} \subset V \times (V \cup T)^*$, un ensemble fini de règles de récriture ;
- $S \in V$, un symbole distingué appelé axiome.

Dans l'exemple des expressions arithmétiques de la section 1, les variables sont l'ensemble $\{\text{Expression}, \text{Terme}, \text{Facteur}\}$, les terminaux sont les chiffres et les lettres, les règles sont données dans la figure 3.4, et l'axiome est la variable `Expression`. On prendra soin de noter que les variables d'une formule ne sont pas celles de la grammaire mais au contraire ses terminaux.

Définissons maintenant le langage engendré par une grammaire, en formalisant le concept de dérivation : on omettra de préciser dans toute la suite du chapitre que les grammaires sont hors-contexte.

Définition 10.2.2 Étant donné une grammaire \mathcal{G} et deux mots $w, w' \in (V \cup T)^*$, on dit que w se récrit en w' (ou que w' dérive en un coup de w), si et seulement si :

- w se décompose en $w = xvy$, $x, y \in (V \cup T)^*$ et $v \in V$;
- w' se décompose en $w' = xfy$, $x, y \in (V \cup T)^*$ (les mêmes) et $f \in (V \cup T)^*$;
- (v, f) est une règle de \mathcal{G} .

On note une réécriture $w \rightarrow \mathcal{G}w'$ ou encore $w \rightarrow w'$ lorsque la grammaire est implicite.

Une suite de réécritures

$$w \rightarrow w_1 \rightarrow w_2 \rightarrow \dots \rightarrow w_n = w'$$

est appelée dérivation et est notée $w \xrightarrow{*} w'$.

Si, de plus, le mot w' est tout entier sur l'alphabet terminal ($w' \in T^*$), on dit que la dérivation est terminale.

Un même mot peut posséder plusieurs dérivations ; ainsi le mot

$$({}_1(2x))_2({}_3x)_3x)_1(4x)_4x$$

est-il dérivable des deux façons suivantes où l'on a souligné la variable réécrite à chaque étape :

$$\begin{aligned} S \rightarrow (S)\underline{S} \rightarrow (S)(\underline{S})S \rightarrow (\underline{S})(x)S \rightarrow ((S)\underline{S})(x)S \rightarrow ((S)(\underline{S})S)(x)S \rightarrow ((S)(x)\underline{S})(x)S \\ \rightarrow ((S)(x)x)(x)\underline{S} \rightarrow ((\underline{S})(x)x)(x)x \rightarrow ((x)(x)x)(x)x, \text{ et} \end{aligned}$$

$$\begin{aligned} S \rightarrow (\underline{S})S \rightarrow ((\underline{S})S)S \rightarrow ((x)\underline{S})S \rightarrow ((x)(\underline{S})S)S \rightarrow ((x)(x)\underline{S})S \rightarrow ((x)(x)x)\underline{S} \\ \rightarrow ((x)(x)x)(\underline{S})S \rightarrow ((x)(x)x)(x)\underline{S} \rightarrow ((x)(x)x)(x)x. \end{aligned}$$

Les arbres de dérivation associés sont identiques ; dans la seconde dérivation, on a toujours effectué la réécriture la plus à gauche dans l'arbre. Et il est évident que cette suite de dérivations est toujours possible et est caractéristique d'un arbre.

Proposition 10.2.1 *Toute dérivation dans une grammaire hors-contexte est équivalente à une dérivation dans laquelle le symbole non-terminal le plus à gauche est systématiquement réécrit.*

On en “ déduit ” que les réécritures ne dépendent pas du contexte et que chaque sous-arbre de l'arbre de dérivation peut être calculé indépendamment des autres.

Définition 10.2.3 *On appelle langage engendré par une grammaire \mathcal{G} l'ensemble des mots sur l'alphabet terminal dérivables de l'axiome de la grammaire :*

$$\mathcal{L}(\mathcal{G}) = \{w \in T^* \mid S \xrightarrow{*} w\}.$$

Remarquons encore que les mots obtenus à chaque étape de la dérivation sont les lisières des arbres de dérivation successifs et que les mots du langage sont les lisières écrites sur l'alphabet terminal. On appelle lisière le mot obtenu en lisant les feuilles de l'arbre en ordre préfixe.

10.2.3 Ambiguïté

On a déjà fait remarquer que l'analyse d'une formule pouvait être multiple comme dans le cas des expressions arithmétiques dans lesquelles on ne tiendrait pas compte des priorités. Ce souhait d'avoir une propriété de " décodage unique " est évident si l'on veut associer comme en linguistique un sens unique (incarné par l'arbre syntaxique) à la phrase (le mot engendré par la grammaire), souhait encore plus impératif en programmation, où il faut qu'un programme soit exécuté de la même façon par toutes les machines. On souhaite donc construire des grammaires possédant la propriété de n'engendrer chaque mot du langage qu'une fois et une seule.

Définition 10.2.4 *Une grammaire G est dite non-ambiguë si et seulement si chaque mot du langage possède un arbre de dérivation unique.*

Par exemple, la grammaire déjà écrite pour le langage $\{a^n b^n \mid n \geq 0\}$ est non-ambiguë. On pourra facilement l'étendre en des grammaires non-ambiguës pour les langages $L_1 = \{a^n b^n c^m \mid n, m \geq 0\}$ et $L_2 = \{a^n b^m c^m \mid n, m \geq 0\}$. Il est cependant impossible de construire une grammaire non-ambiguë pour le langage union $L_1 \cup L_2$ qui est donc inhéremment ambigu.

La théorie de l'ambiguïté est complexe et dépasse largement le cadre de ce cours.

10.3 Propriétés des langages algébriques

Les langages engendrés par des grammaires hors-contexte satisfont des systèmes d'équations qui sont le pendant des systèmes algébriques en variables commutatives, de la même façon que les langages rationnels, associés aux expressions régulières satisfont des équations linéaires et sont le pendant des fractions rationnelles en variables non-commutatives. Nous les baptiserons donc langages algébriques selon la terminologie de M.-P. Schützenberger et noterons \mathcal{Alg} leur classe. Ils vérifient un certain nombre de propriétés de clôture ensemblistes.

Proposition 10.3.1 *La classe \mathcal{Alg} des langages algébriques est close par union, produit et quasi-inverse. Elle est close par homomorphisme et substitution algébrique.*

Preuve : Soient $L_1 = \mathcal{L}(\mathcal{G}_1)$ et $L_2 = \mathcal{L}(\mathcal{G}_2)$ les langages engendrés par les grammaires \mathcal{G}_1 et \mathcal{G}_2 d'axiomes respectifs S_1 et S_2 . Remarquons qu'on peut toujours supposer que les alphabets de variables V_1 et V_2 sont disjoints (en les renommant). On construit alors explicitement des grammaires pour $L_1 \cup L_2$, $L_1 L_2$ et L_1^* , à partir de $\mathcal{G}_1 = \langle V_1, T_1, \mathcal{R}_1, S_1 \rangle$ et de $\mathcal{G}_2 = \langle V_2, T_2, \mathcal{R}_2, S_2 \rangle$.

La nouvelle grammaire s'écrit alors sous la forme générique $\mathcal{G} = \langle V, T, \mathcal{R}, S \rangle$, où les ensembles V et \mathcal{R} sont définis dans chaque cas comme suit.

Union :

- $T = T_1 \cup T_2$;
- $V = \{S\} \cup V_1 \cup V_2$;
- $\mathcal{R} = \{S = S_1 + S_2\} \cup \mathcal{R}_1 \cup \mathcal{R}_2$.

Produit :

- $T = T_1 \cup T_2$;
- $V = \{S\} \cup V_1 \cup V_2$;
- $\mathcal{R} = \{S = S_1 S_2\} \cup \mathcal{R}_1 \cup \mathcal{R}_2$.

Quasi-inverse :

- $T = T_1$;
- $V = \{S\} \cup V_1$;
- $\mathcal{R} = \{S = S_1 S + \varepsilon\} \cup \mathcal{R}_1$.

Homomorphisme :

L'homomorphisme est une application h de T_1 dans T_2^* ; on l'étend à V_1 comme l'identité ($h(v) = v$ sur V_1) ;

- $T = T_2$;
- $V = V_1$;
- $\mathcal{R} = \{S = S_1\} \cup h(\mathcal{R}_1)$, chaque règle étant transformée par h qui laisse les variables inchangées.

Substitution algébrique :

La substitution σ est une application h de T_1 dans T_2^* (en supposant T_2 disjoint de T_1) telle que l'image de chaque lettre $x \in T_1$ est un langage algébrique L_x inclus dans T_2^* , de grammaire \mathcal{G}_x et d'axiome S_x ; étant donné un mot $w = x_1 x_2 \dots x_k$, son image par la substitution est le langage $\sigma(w) = \{w_1 w_2 \dots w_k \mid w_i \in L_{x_i}\}$; pour ce langage on a la grammaire avec :

- $T = T_2$;
- $V = V_1 \cup T_1 \cup_{x \in T_1} V_x$;
- $\mathcal{R} = \{S = S_1\} \cup \mathcal{R}_1 \cup_{x \in T_1} \mathcal{R}_x \cup_{x \in T_1} (x = S_x)$.

De façon assez immédiate les grammaires s'imbriquent les unes dans les autres, conduisant à l'effet cherché.

Une autre opération d'importance est l'intersection avec un langage rationnel, autrement dit, la superposition à la structure parenthétique de régularités décrites par expressions régulières. Le prochain théorème montre comment l'on peut prendre en compte les deux systèmes de contraintes en croisant une grammaire pour un langage algébrique avec un automate fini qui reconnaît le langage rationnel. Ce théorème montre aussi l'importance des arbres de dérivation dans ces processus.

Théorème 10.3.1 *La classe Alg des langages algébriques est close par intersection avec un langage rationnel.*

Preuve : Soit L un langage algébrique donné par sa grammaire $\mathcal{G} = \langle V, T, \mathcal{R}, S \rangle$ et soit K un langage rationnel reconnu par l'automate déterministe $\mathcal{A} = \langle T, Q, q_0, F, \Delta \rangle$.

Le langage $L \cap K \subset T^*$ est défini par la grammaire suivante \mathcal{G}' :

- $V' = \{S_0\} \cup \{(q, v, q') \mid q, q' \in Q, v \in V\}$;
- S_0 est l'axiome ;
- les règles sont de deux types \mathcal{R}_1 et \mathcal{R}_2 et expriment de façon récursive que l'ensemble des mots engendrés par la variable (q, v, q') est celui des mots $w \in T^*$ qui, simultanément, font passer l'automate \mathcal{A} de l'état q dans l'état q' .

Les variables de \mathcal{G}' susceptibles de produire des mots de $L \cap K$ doivent donc être du type (q_0, S, q_f) , $q_f \in F$ étant un état final de l'automate. Pour n'avoir qu'un seul axiome dans la grammaire nous utilisons S_0 et \mathcal{R}_1 est formé des règles

$$\mathcal{R}_1 = \{S_0 = (q_0, S, q) \mid q \in F\}.$$

L'ensemble \mathcal{R}_2 décrit récursivement le double processus de L et K . Pour chaque règle $v = f$ terminale, c.-à-d. où $f \in T^*$, \mathcal{R}_2 contient les règles du type

$$\{(q, v, q') = f \mid q \xrightarrow{f} q'\};$$

et pour chaque règle de la forme $v = f_0 v_1 f_1 v_2 \dots v_k f_k$, où $v_i \in V$ et $f_i \in T^*$, \mathcal{R}_2 contient les règles du type

$$\begin{aligned} \{(q, v, q') = f_0(q_1, v_1, q'_1) f_1(q_2, v_2, q'_2) \dots (q_k, v_k, q'_k) f_k \\ \mid \forall i, 1 \leq i < k, q'_i \xrightarrow{f_i} q_{i+1} \ \& \ q \xrightarrow{f_0} q_1 \ \& \ q'_k \xrightarrow{f_k} q'\}. \end{aligned}$$

Par récurrence sur la structure des dérivations on voit qu'à chaque étape de la construction de l'arbre d'un mot du langage L on propage la contrainte que ce mot soit aussi reconnu par l'automate de K , puisque l'on impose — ce que la grammaire \mathcal{G}' vérifie — que le mot engendré par L amène aussi l'automate dans un état terminal.

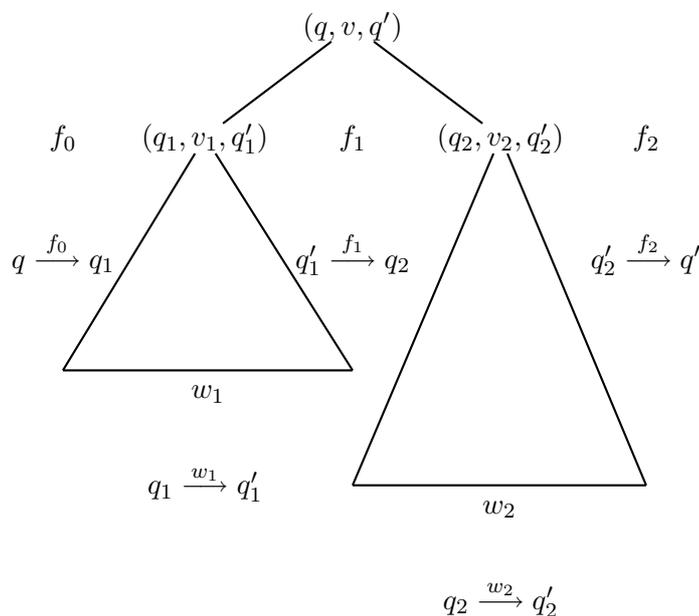


FIG. 10.3 – Illustration pour l'arbre de dérivation dans la grammaire avec intersection par un rationnel.

10.4 Grammaires de Greibach et automates à pile

On montre dans cette section que l'on peut mettre les grammaires dans une forme canonique qui permet, premièrement, d'éliminer les récritures récursives de variables dans le sous-arbre gauche, et deuxièmement de mettre en place un mécanisme de reconnaissance des mots du langage, un peu moins coûteux en mémoire et en temps que le processus d'exploration brut. Ce processus, appelé *automate à pile* reste cependant non-déterministe et il faudra bien comprendre qu'il est impossible de faire autrement. On s'éloigne ainsi de la situation rencontrée pour les langages rationnels et les automates finis.

10.4.1 Forme normale de Greibach

Les grammaires hors-contexte définies plus haut possèdent des règles de réécriture du type $v = w$, où $w \in V^*$; on peut donc "tourner en rond" dans les dérivations avant de commencer à produire des lettres de l'alphabet terminal. Ceci rend l'analyse prédictive des mots assez compliquée et on souhaiterait pouvoir se passer de telles productions.

Proposition 10.4.1 *Il existe un algorithme qui teste si le langage engendré par une grammaire hors-contexte est vide ou non.*

Preuve : Considérons un arbre de dérivation qui produise un mot w du langage L donc sur l'alphabet terminal, en supposant le langage non vide. Deux cas peuvent se produire : les branches issues de la racine (l'axiome) sont étiquetées par des variables toutes différentes, ou bien une variable, au moins, est répétée.

Considérons par exemple la grammaire $\mathcal{G} = \langle \{A, B, C, S\}, \{a, b, c\}, S, \{S = AC, A = aB + \varepsilon, B = Ab, C = Cc + \varepsilon\} \rangle$.

Un arbre de dérivation pour le mot $aabbc$ est donné ci-dessous, et on voit que la variable A apparaît répétée sur une branche alors que C ne l'est pas. Il est cependant possible en tronquant cet arbre de produire un mot terminal de la grammaire qui soit sans redondance comme le montre l'exemple.

On en arrive donc à la situation suivante :

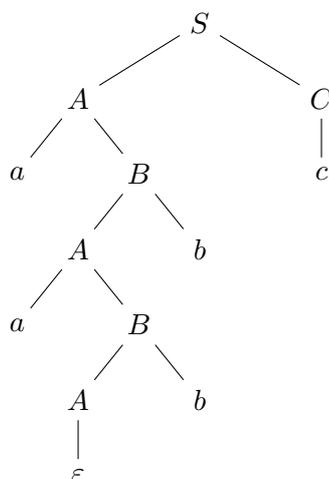
- si L est non vide, il existe un mot dont l'arbre de dérivation a une profondeur au plus $m = |V|$, nombre de variables de la grammaire puisque, sur chacune de ses branches, aucune feuille n'est répétée ;
- il suffit d'énumérer tous les arbres de dérivation possibles de profondeur au plus m pour décider de la vacuité de $\mathcal{L}(\mathcal{G})$.

Considérons maintenant les variables inutiles ; elles sont de deux types : celles qui ne produisent pas de mots sur T^* et celles qu'on ne peut pas atteindre depuis l'axiome.

Proposition 10.4.2 *Toute grammaire peut être transformée en une grammaire équivalente où chaque variable produit un mot terminal : $v \xrightarrow{*} f \in T^*$.*

Preuve : On procède en deux temps :

- pour chaque $v \in V$, tester si les mots engendrés à partir de v (pris pour axiome)



dans la grammaire \mathcal{G} forment un langage vide ;

— si tel est le cas, supprimer la variable v de la grammaire ainsi que toutes les productions qui y font référence.

Le processus ne modifie pas les mots du langage puisque si un mot dérivé contient une variable v non productive, alors il ne peut pas se dériver en un mot terminal. Par ailleurs les dérivations terminales ne sont pas touchées.

Proposition 10.4.3 *Pour tout langage algébrique L , on peut construire une grammaire hors-contexte telle que pour toute variable v il existe une dérivation terminale où elle intervient : $S \xrightarrow{*} fvg \xrightarrow{*} fwg$, où $v \in V$ et $f, g, w \in T^*$.*

Preuve : Dans la grammaire précédente, toutes les variables produisent un mot terminal, donc si $S \xrightarrow{*} fvg$, on en déduit que v satisfait la propriété. Il suffit donc de rechercher systématiquement les variables accessibles à partir de l'axiome par une suite de dérivations, ce qui se fait en considérant le graphe G dont les sommets sont les variables de \mathcal{G} et les arêtes sont telles que (v, v') existe si et seulement s'il existe une règle dans \mathcal{G} de la forme $v = fv'g$. On cherche alors dans le graphe G les sommets accessibles de S l'axiome, par exemple en construisant un arbre de Trémaux. Les sommets non accessibles sont retirés de la grammaire ainsi que les productions qui les contiennent.

Enfin il est possible de supprimer les simples récritures de variables du type $v = v'$ en les remplaçant par les membres droits des productions récrivant v' .

Proposition 10.4.4 *Pour chaque grammaire, on peut construire une grammaire équivalente dans laquelle il n'y a pas de règle du type $v = v'$.*

Les transformations précédentes permettent de simplifier notablement les grammaires et d'en éliminer les variables scories. Elles ne permettent pas de trouver facilement un arbre de dérivation potentiel pour un mot si celui-ci appartient au langage

engendré. Une étape importante, que nous retrouverons en compilation, est d'éliminer les appels récursifs gauches, ce qui revient sur la grammaire à faire que tous les membres droits de règle commencent par un symbole terminal. Commençons par deux lemmes.

Lemme 10.4.1 *Dans une grammaire hors-contexte où $v' = \gamma_1 + \gamma_2 + \dots + \gamma_k$ est un ensemble de règles, on peut remplacer les règles du type $v = \alpha v' \beta$ par $v = \alpha \gamma_1 \beta + \alpha \gamma_2 \beta + \dots + \alpha \gamma_k \beta$ sans changer le langage engendré.*

Preuve : On note qu'en procédant ainsi on ne fait que compresser deux récritures en une seule, et donc que le résultat est inchangé.

Le second lemme est plus intéressant car il va permettre de transformer des récurrences droites en des récurrences gauches. Il repose sur une propriété élémentaire des langages rationnels et des grammaires qui les engendrent.

Lemme 10.4.2 *Soit v une variable dont les récritures possibles sont données par les règles $v = v\alpha_1 + v\alpha_2 + \dots + v\alpha_k + \beta_1 + \beta_2 + \dots + \beta_l$, où les α_i et β_j sont dans $(V \cup T)^*$. Alors, on ne modifie pas le langage engendré en les remplaçant par les deux jeux de règles :*

$$\begin{aligned} v &= \beta_1 + \beta_2 + \dots + \beta_l + \beta_1 \hat{v} + \beta_2 \hat{v} + \dots + \beta_l \hat{v} \\ \hat{v} &= \alpha_1 + \alpha_2 + \dots + \alpha_k + \alpha_1 \hat{v} + \alpha_2 \hat{v} + \dots + \alpha_k \hat{v}. \end{aligned}$$

Preuve : Il suffit de vérifier que les deux jeux de règles engendrent également l'ensemble

$$(\beta_1 + \beta_2 + \dots + \beta_l)(\alpha_1 + \alpha_2 + \dots + \alpha_k)^*.$$

Pour ce faire, on se restreint à l'étude des dérivations gauches de la grammaire qui produisent, comme on le sait, l'ensemble des mots du langage.

On est maintenant en position de conclure et de prouver le théorème de Greibach.

Théorème 10.4.1 *Tout langage algébrique peut être engendré par une grammaire où toutes les règles sont du type*

$$v = aw, \text{ où } v \in V, a \in T, w \in V^*.$$

Preuve :

1. Commençons par mettre la grammaire sous la forme suivante

$$v = a_1 w_1 + a_2 w_2 + \dots + a_k w_k + w'_1 + w'_2 + \dots + w'_l, \text{ où } v \in V, a_i \in T, w_j, w'_j \in V^*.$$

Ceci est toujours possible en créant de nouvelles variables de telle sorte que si $v = a_1 h_1 a_2 h_2 \dots a_k h_k$, avec $a_i \in T$ et $h_j \in V^*$, alors on remplace cette règle par :

$$v = a_1 h_1 u_1, u_1 = a_2 h_2 u_2, \dots, u_{k-1} = a_k h_k.$$

Si le membre droit ne commence pas par un terminal, la même démarche conduit à la forme annoncée.

2. Les règles du type $v = aw$ sont maintenant conservées dans la grammaire finale.

3. Il faut maintenant traiter les autres pour éliminer les récurrences à gauche.

3.1 Soit V l'ensemble des variables. On va transformer les règles, au prix de l'ajout de quelques variables auxiliaires Z , pour faire en sorte que se dégage un ordre $V' = \{v_1, v_2, \dots, v_k, \hat{v}_1, \hat{v}_2, \dots, \hat{v}_k\} = V \cup \hat{V}$ tel que dans les productions de type $v = w$, w commence par une variable de rang supérieur strictement à celui de v .

Par un algorithme inspiré de celui du tri topologique, on peut trouver un début d'ordre, $v_1 < v_2 < \dots < v_i$ satisfaisant cette propriété, et si $i = k$ alors la première phase de la transformation est achevée.

Sinon, soit v_i la première variable telle qu'il existe $v_i = v_j w$, avec $j \leq i$. Alors comme v_j vérifie la propriété que toutes ses productions commencent par des $v_{j'}$, avec $j' > j$ ou des terminaux, on peut faire la substitution de ces membres droits dans ceux des règles attachées à v_i et on répète le processus jusqu'à ce que les variables de tête des productions de v_i soient toutes d'indice au moins i . La grammaire est équivalente par le lemme 10.4.1.

3.2 On est maintenant en présence de récurrences gauches du type $v = vw + v'w'$ (v' est d'indice supérieur à v ou est terminal) et qui sont donc passibles du lemme 10.4.2 et sont transformées en $v = v'w'\hat{v} + v'w'$ et $\hat{v} = w + w\hat{v}$.

Considérons alors la dernière variable v_k : comme il n'y a pas de v d'indice supérieur toutes ses productions commencent par un terminal. Par récurrence, les productions de v_{k-1} commencent soit par un terminal soit par v_k et par substitution (lemme 10.4.1) on les fait toutes commencer par un terminal. En itérant, toutes les productions des variables originelles commencent donc par un terminal.

Enfin, on remarque que les productions des variables additionnelles commencent toujours par un terminal ou une variable originelle ; par substitution encore, on se ramène à avoir toujours un terminal en tête de règle.

Ce qui prouve le théorème.

10.4.2 Automates à pile

Le théorème de forme normale de Greibach permet alors un passage facile à une famille d'automates reconnaissant les langages algébriques. (On pourrait aussi le faire directement).

Commençons par deux exemples.

Pour reconnaître si un mot w est dans le langage $M = \{fcf^R \mid f \in (a+b)^*\}$ et qui est évidemment algébrique, on peut procéder ainsi :

- lire le mot f jusqu'au c et le stocker dans une liste chaînée $\pi(f)$;
- poursuivre la lecture après le c et comparer chaque lettre lue de la donnée avec la lettre précédente de la liste $\pi(f)$;
- si une discordance apparaît dans le processus, le mot n'est pas dans M ; si, au contraire on épuise la donnée en même temps que l'on atteint le début de la liste π , le

Le calcul d'un automate à pile est défini par une suite de *configurations* compatible avec la table de transitions. On notera bien que ces automates ne sont pas déterministes *a priori*.

Une configuration d'un automate est un triplet (f, q, π) , où $f \in T^*$ est le suffixe de la donnée qui reste à lire, $q \in Q$ est l'état courant et $\pi \in P^*$ est le mot de pile : par convention l'extrémité accessible est la première lettre de π .

Pour en décrire le fonctionnement, il est commode d'introduire les fonctions `push` et `pop` sur la pile, et d'utiliser les notations Caml (`hd` et `tl`) pour la gestion des files :

— la fonction `pop` a pour effet de supprimer la lettre de sommet de pile (l'extrémité accessible de la pile) : $\text{pop}(\pi) = (\text{tl } \pi)$, mais on ne peut pas faire de `pop` sur une pile vide car il y a alors erreur :

— la fonction `push(w)` exécute d'abord un `pop`, puis empile le mot w , de telle sorte que $\text{push}(w, \pi) = w @ \text{pop}(\pi)$, à condition que la pile ne soit pas vide. Remarquons que $\text{pop} \equiv \text{push}(\varepsilon)$.

On passe de la configuration (f, q, π) à la configuration (f', q', π') à condition qu'il existe une règle dans la table Δ , $(q, \alpha, p) \rightarrow (q', w)$, où α peut être ε est un mot (éventuellement vide) sur l'alphabet de pile et telle que : $f = \alpha f'$ et $\pi' = \text{push}(w, \pi)$.

Un calcul sur le mot f commence par la configuration (f, q_0, p_0) .

Si la pile est vide en même temps que le mot de donnée est lu, l'automate s'arrête avec succès. Si aucune règle n'est applicable le calcul n'est pas valide. Un mot est accepté par l'automate si et seulement s'il existe une suite de configurations qui l'amène à la pile vide. Il faut remarquer qu'il n'y a pas unicité de la suite de calculs, que pour un même mot, il peut exister une ou même des suites qui l'acceptent et d'autres qui ne le font pas. Il s'agit ici encore d'une conséquence irréductible du non-déterminisme.

Voici un exemple d'automate à pile :

$T = \{a, b\}, Q = \{0, 1, 2\}, P = \{b, c\}, q_0 = 0, p_0 = c$

$\Delta = \{(0, a, c) \rightarrow (0, c), (0, b, c) \rightarrow (0, bc), (0, b, b) \rightarrow (0, bb), (0, b, b) \rightarrow (1, bb) \text{ (on empile les } b)\}$

$(0, b, b) \rightarrow (1, bb) \text{ (on change de mode)}$

$(1, a, b) \rightarrow (1, \varepsilon) \text{ (on compte autant de } a \text{ que de } b)$

$(1, b, c) \rightarrow (2, c) \text{ (on a trouvé le motif)}$

$(2, a, c) \rightarrow (2, c), (2, b, c) \rightarrow (2, c), (2, b, c) \rightarrow (2, \varepsilon)\}$ (pour terminer le mot).

Cet automate reconnaît les mots de la forme $a^*ba^{i_1}ba^{i_2}b\dots ba^{i_k}b$ pour lesquels il existe un j tel que $j = i_j$. On voit bien que ce processus est fondamentalement non-déterministe, car il existe des mots acceptant plusieurs analyses correctes.

Théorème 10.4.2 *La class Alg est identique à celle des langages reconnaissables par automate à pile.*

Preuve :

Considérons un langage L engendré par une grammaire en forme de Greibach \mathcal{G} . Les règles de cette grammaire sont donc de la forme $v = av_1v_2\dots v_k$ ou $v = a$, pour $a \in T$ lettre de l'alphabet terminal et pour des v et $v_i \in V$ variables.

Construisons un automate à pile comme suit : T est son alphabet d'entrée, V est son alphabet de pile et S (l'axiome de \mathcal{G}) est son symbole de début de pile, Q et réduit à

un seul état qu'on ne note donc pas. La table de transition est formée des deux types de règles trivialement déduites de celles de \mathcal{G} :

— pour $v = a$ de \mathcal{G} , Δ contient $(-, a, v) \rightarrow (-, \varepsilon)$, ce qui revient à lire a tout en dépilant v ;

— pour $v = av_1v_2\dots v_k$ de \mathcal{G} , Δ contient $(-, a, v) \rightarrow (-, v_1v_2\dots v_k)$, ce qui revient à lire a tout en remplaçant v par $v_1v_2\dots v_k$ en sommet de pile, v_1 devenant le nouveau sommet.

Considérons alors une dérivation gauche d'un mot f dans \mathcal{G} . Les mots produits par les récritures successives sont de la forme hw , où $h \in T^*$ est un préfixe de f et $w \in V^*$ fournira le suffixe g de $f = hg$. Un calcul valide de l'automate défini ci-dessus sur le mot f est celui où les mots de pile de sont précisément les w successifs de la dérivation gauche et chaque lettre de f est lue au moment où elle est produite par une récriture du non-terminal gauche. Donc le mot est accepté par l'automate. Réciproquement tout calcul de l'automate acceptant un mot est immédiatement interprétable en terme de dérivation gauche de ce mot, par construction même.

Inversement, on construit pour tout automate à pile un automate équivalent sans états, ceux-ci étant mémorisés dans la pile. L'automate résultant est alors semblable à celui construit dans la partie directe ; on construit par le même mécanisme (mais en sens inverse) une grammaire hors-contexte dont il est évident qu'elle engendre le même langage.

D'où le théorème.

10.5 Analyse par programmation dynamique

Dans cette section, on montre l'existence d'un algorithme déterministe en temps polynomial pour décider si un mot est dans le langage engendré par une grammaire ou non.

10.5.1 Forme normale de Chomsky

Une autre normalisation des productions d'une grammaire hors-contexte est d'une grande utilité. Elle permet de rendre plus réguliers les arbres de dérivations en rendant essentiellement leurs sommets internes binaires. Les productions d'une telle grammaire sont de l'une des deux formes suivantes : $A = BC$ pour $A, B, C, \in V$ ou $A = a$ pour $a \in T$. On voit donc que les nœuds binaires de l'arbre correspondent à des récritures d'une variable en deux variables, alors que les nœuds unaires sont une récriture terminale produisant une lettre du mot engendré.

Proposition 10.5.1 *Pour toute grammaire hors-contexte, il existe une grammaire équivalente sous forme normale de Chomsky.*

Preuve : On peut supposer, sans restriction, que la grammaire \mathcal{G} est propre, c'est-à-dire que toutes les variables sont utiles, qu'il n'y a pas de règle du type $A = B$, pour $A, B \in V$, et qu'à l'exception de l'axiome aucune ne produit le mot vide. Toutes ces constructions ont été vues précédemment. On transforme alors les règles de \mathcal{G} comme

suit, en deux temps, en introduisant autant de nouvelles variables que nécessaire :

1. chaque règle de la forme $A = a$, $A \in V$ et $a \in T$, est conservée ;
2. chaque règle de la forme $A = w$, $A \in V$ et $w \in (V \cup T)^*$, est remplacée par $A = W_1W_2\dots W_{|w|}$, où les W_i soit sont $W_i = w_i$ si $w_i \in V$, soit sont de nouvelles variables si $w_i \in T$ auquel cas on ajoute la règle $W_i = w_i$;
3. les règles $A = W_1W_2\dots W_k$ ainsi créées sont ensuite normalisées pour $k > 2$ et transformées en l'ensemble de règles binaires : $A = W_1W'_2; W'_2 = W_2W'_3; \dots; W'_{k-1} = W_{k-1}W_k$.

Ainsi la nouvelle grammaire est en forme de Chomsky et il est évident qu'elle est strictement équivalente à \mathcal{G} . Remarquons de plus que sa taille est de l'ordre de 4 fois celle de \mathcal{G} .

10.5.2 Reconnaissance déterministe des langages algébriques

La mise en forme de Chomsky des grammaires hors-contexte permet de programmer très simplement l'analyse des mots du langage. En effet, définissons la fonction $Drv : V \times T^* \rightarrow \{0, 1\}$, où $\{0, 1\}$ sont les valeurs de vérité faux et vrai comme suit, en utilisant pour convention que les variables sont représentées par des capitales ($A, B, C, \dots \in V$) et que les mots sur l'alphabet terminal le sont par des minuscules ($f, g, h, \dots \in T^* \setminus \{\varepsilon\}$). Alors :

$$Drv(A, f) = \begin{array}{ll} & \{\text{pour } f \neq \varepsilon\} \\ \text{si } f \in T & \text{alors } (A = f) \in \mathcal{R}? \\ \text{sinon} & \exists g, h \neq \varepsilon \exists A, B, C \\ & \{ f = gh \ \& \ A = BC \in \mathcal{R} \ \& \ Drv(B, g) \ \& \ Drv(C, h) \}. \end{array}$$

Une telle définition se traduit immédiatement dans tout langage de programmation puisqu'il suffit de programmer les deux quantificateurs existentiels par deux boucles :
 — la première sur les césures de f en $f = gh$;
 — la seconde sur toutes les règles de \mathcal{R} dont le membre gauche est la variable A .
 Voici comment on peut l'exprimer en pseudocode, en utilisant les fonctions booléennes de test d'appartenance à l'alphabet (`estLettre`) et à l'ensemble \mathcal{R} des règles de la grammaire (`estRegle`) :

```
// R est une liste de listes [A B C] ou [A a]
derive(A, f)
// retourne true si et seulement si A produit f
si (longueur(f) == 1) ET estLettre(tête(f)) alors
    retourner estRegle([A, tête(f)]);
sinon
    retourner itererDecoupe(A, [tête(f)], queue(f));

itererDecoupe(A, g, f)
// recherche une césure correcte pour A->f
```

```

si f == [] alors
  retourner false;
sinon
  si itererRegle(A, R, g, f) alors
    retourner true;
  sinon
    retourner itererDecoupe(A, g # [tête(f)], queue(f));

itererRegle(A, R, g, h)
// recherche une règle correcte pour A -> gh
si R == [] alors
  retourner false;
sinon
  r <- tête(R); R1 <- queue(R);
  si A == r ET longueur(queue(r)) == 2 alors
    r = (r1, r2);
    si derive(r1, g) ET derive(r2, h) alors
      retourner true;
    sinon
      retourner itererRegle(A, R1, g, h);
  sinon
    retourner itererRegle(A, R1, g, h);

```

La programmation est donc simple en C, Caml ou Java, mais elle sera en général très inefficace car le nombre de façons de décomposer en facteurs analysables est essentiellement proportionnel au nombre d'arbres binaires à n sommets, ou encore de parenthésages à n paires de parenthèses, quantité appelée *nombre de Catalan*¹ et valant

$$C_n = \frac{1}{2n+1} \binom{2n}{n} \approx \frac{1}{\sqrt{\pi}} 4^n n^{-3/2}.$$

Le coût d'exécution de ce programme est donc exponentiel !

Il existe pourtant deux façons de faire tomber sa complexité dans le domaine du raisonnable polynomial. Elles reposent toutes deux sur la remarque que l'on répète un très grand nombre de fois les mêmes appels récursifs, et donc qu'il faut éviter de reprendre ces calculs soit en les mémorisant soit en les organisant mieux.

Mémo-fonctions

Cette méthode est utilisée pour accélérer l'accès aux valeurs déjà calculées d'une fonction ; elle est particulièrement efficace lorsque la définition est hautement récursive et permet, moyennant très peu de modifications au code d'en accélérer prodigieusement l'exécution. Prenons par exemple le cas de la fonction qui calcule les nombres du triangle

¹C'est un équivalent de ces nombres dans le cas ternaire que l'on a vu apparaître au chapitre 1 lors du dénombrement des fonctions booléennes.

de Pascal selon leur définition récursive standard : $\binom{n}{0} = 1$, $\binom{n}{p} = 0$ si $p > n$, $\binom{n+1}{p} = \binom{n}{p-1} + \binom{n}{p}$.

Utilisons un tableau memo de longueur suffisante pour stocker les $\frac{(N+1)(N+2)}{2}$ valeurs de la fonction jusqu'au rang N . On initialise ce tableau à 0 et on rangera la valeur de $\binom{n}{p}$ en position $\frac{(n+1)(n+2)}{2} + p$, commençant ainsi à l'adresse 1..

Le programme initial pour la fonction comb,

```

comb(n, p)
si n == 0 alors
  retourner 1;
sinon
  si p > n alors
    retourner 0;
  sinon
    retourner comb(n-1, p-1) + comb(n-1, p);

```

est changé en

```

memo <- tableau de taille 10000;

mcomb(n, p)
si n == 0 alors
  retourner 1;
sinon
  si p > n alors
    retourner 0;
  sinon
    vm <- memo[(n+1)*(n+2)/2+p];
    si vm != 0 alors
      retourner vm;
    sinon
      vm <- mcomb(n-1, p-1) + mcomb(n-1, p);
      memo[(n+1)*(n+2)/2+p] <- vm;
    retourner vm;

```

Nous avons utilisé ici le fait qu'il existait une bijection simple entre les éléments du triangle de Pascal et le vecteur memo. Si tel n'est pas le cas, on peut avoir recours à une table de hachage qui permet de ranger un ensemble difficilement ordonnable et de retrouver facilement les éléments déjà rangés, la structure du nouveau programme restant la même.

Dans le cas de la reconnaissance d'un langage algébrique par la fonction `derive`, on doit donc construire une table à double entrée, fonction de l'ensemble des non-terminaux de la grammaire V et de l'ensemble des facteurs du mot f qui sont en nombre $|f|^2$. Chaque appel de la fonction `derive` sur un mot de longueur p demande donc une lecture de la table des productions de la grammaire et la recherche des valeurs

de la fonction pour les p décompositions possibles de la donnée. En conséquence le coût de cette implantation de l'algorithme est $O(|\mathcal{G}| \cdot |f|^3)$, ce qui est polynomial et donc acceptable.

La programmation dynamique : Cocke, Kasami, Younger.

On peut aussi programmer cet algorithme de recherche d'un arbre de dérivation de façon entièrement tabulaire et complètement itérative, en utilisant un des paradigmes classiques de programmation, appelé *programmation dynamique*, très en vogue en mathématiques appliquées. Cela consiste essentiellement à démonter complètement le mécanisme des appels récursifs, puis à reconstruire le calcul à partir des appels terminaux (sur des mots de 1 lettre), puis en remontant aux facteurs de 2 lettres, puis 3 et ainsi de suite jusqu'à atteindre le mot f de départ de longueur n . On utilise alors un tableau T à deux dimensions de taille $n \times n$, dont les éléments sont des listes de non-terminaux de la grammaire \mathcal{G} : chaque liste est composée des non-terminaux qui peuvent engendrer un facteur bien déterminé du mot, selon le schéma suivant.

La première ligne contient les variables qui engendrent les lettres du mot f , en utilisant donc les règles du type $A \rightarrow a$:

$$T[1, i] = \{A \mid A \rightarrow f_i \in \mathcal{R}\}.$$

Dans la seconde ligne on assemble les informations de la première, pour déterminer comment on peut former les digrammes (deux lettres consécutives) de f , selon la règle :

$$\begin{aligned} T[2, i] &= \{A \mid A \rightarrow f_i f_{i+1} \in \mathcal{R}\} \\ &= \{A \mid \exists A \rightarrow BC \in \mathcal{R} \ \& \ B \in T[1, i] \ \& \ C \in T[1, i+1]\}. \end{aligned}$$

Avant de donner la règle générale, considérons la troisième ligne où se détermine la génération des trigrammes (trois lettres consécutives) de f , à partir des deux précédentes, selon la règle :

$$\begin{aligned} T[3, i] &= \{A \mid A \rightarrow f_i f_{i+1} f_{i+2} \in \mathcal{R}\} \\ &= \{A \mid \exists A \rightarrow BC \in \mathcal{R} \ \& \\ &\quad ((B \in T[1, i] \ \& \ C \in T[2, i+1]) \vee ((B \in T[2, i] \ \& \ C \in T[1, i+2])))\}. \end{aligned}$$

De façon générale, la k -ième ligne contient les non-terminaux possibles pour les facteurs de longueur k du mot f et se calcule comme suit à partir des lignes précédentes (on a évidemment $k + i \leq n + 1$) :

$$\begin{aligned} T[k, i] &= \{A \mid A \rightarrow f_i f_{i+1} \dots f_{i+k-1} \in \mathcal{R}\} \\ &= \{A \mid \exists A \rightarrow BC \in \mathcal{R}, \ \exists l < k, \ B \in T[l, i] \ \& \ C \in T[k-l, i+l-1]\}. \end{aligned}$$

Le processus est itéré jusqu'à la n -ième ligne du tableau dans laquelle on ne remplit que la case $T[n, 1]$. Par construction ce processus n'est que la mise en forme itérative de la méthode récursive et en conséquence, le mot f appartient au langage $\mathcal{L}(\mathcal{G})$ si et seulement si l'axiome S de la grammaire appartient à l'ensemble $T[n, 1]$.

Le tableau T comporte $\frac{n(n+1)}{2}$ cases effectivement utilisées, et le temps de calcul pour une cellule de la k -ième ligne est $O(|\mathcal{G}|.k)$; en conséquence le temps total est cubique en $|f|$.

Théorème 10.5.1 *Les deux algorithmes ci-dessus déterminent l'appartenance d'un mot de longueur n à un langage algébrique en temps $O(n^3)$.*

Borne inférieure

Dans la pratique les algorithmes de coût cubique sont encore trop complexes ; on verra dans la suite du chapitre qu'en imposant des restrictions à la grammaire, on peut souvent faire l'analyse en temps linéaire. Dans le cas général, le meilleur résultat connu est dû à L. Valiant qui a montré que l'on pouvait ramener le problème à celui de la résolution de systèmes linéaires et donc à la multiplication matricielle, problème connu depuis V. Strassen pour être moins que cubique.

Voici un point récent sur la question par Steven Finch :

“ Define the exponent of matrix multiplication ω as the infimum of all real numbers τ such that multiplication of $n \times n$ matrices may be achieved with $O(n^\tau)$ multiplications. Clearly $\omega \leq 3$ and it can be proved that $\omega \geq 2$.

Strassen discovered a surprising base algorithm to compute the product of 2×2 matrices with only seven multiplications. The technique can be recursively extended to large matrices via a tensor product construction. In this case, the construction is very simple : large matrices are broken down recursively by partitioning the matrices into quarters, sixteenths, etc. This gives $\omega \leq \ln(7)/\ln(2) < 2.808$.

More sophisticated base algorithms and tensor product constructions permit further improvements. Many researchers have contributed to this problem, including Pan who found $\omega < 2.781$ and Strassen who found $\omega < 2.479\dots$

Coppersmith & Winograd presented a new method, based on a combinatorial theorem of Salem & Spencer, which gives dense sets of integers containing no three terms in arithmetic progression. They consequently obtained $\omega < 2.376$, which is the best known upper bound today.

Is $\omega = 2$? Bürgisser called this the central problem of algebraic complexity theory. ”

10.6 Algébricité

De la même façon que l'on a caractérisé les langages rationnels par leur propriété de finitude de leurs classes d'indiscernabilité, on caractérise les langages algébriques par leur système de parenthésages. Nous allons d'abord démontrer un analogue du lemme d'itération, puis nous établirons que le langage des parenthèses est le prototype de toute la famille \mathcal{Alg} .

10.6.1 Le lemme d'itération des langages algébriques

La proposition 3.3 traitait de la finitude des langages algébriques ; réinterprétons-la sous un autre angle. Supposons donc qu'une variable A se répète sur une branche

d'un arbre de dérivation, comme illustré par la figure 10.5, et que cette même variable produise également un mot y non vide sur l'alphabet terminal. Alors on peut couper le sous-arbre intermédiaire ou bien le recopier à volonté, obtenant de nouveaux arbres de dérivation valides comme le montre la figure 10.6.

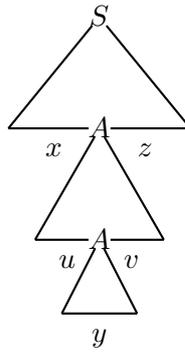


FIG. 10.5 – La répétition d'une variable sur une branche.

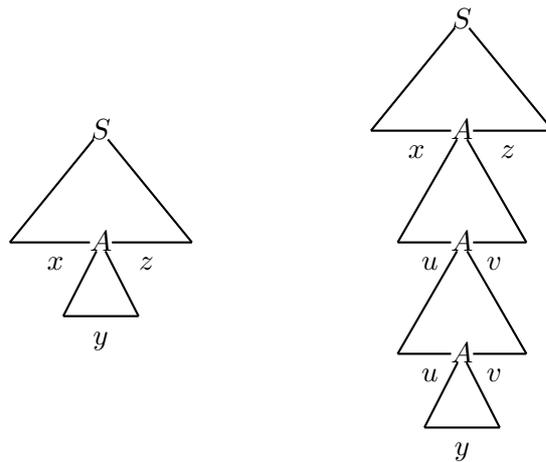


FIG. 10.6 – Itérations possibles d'une variable sur une branche : 0 fois à gauche et 2 fois à droite.

Le lemme d'itération pour les langages algébriques s'énonce donc comme suit.

Théorème 10.6.1 [Bar-Hillel, Perles, Shamir]

Soit L un langage algébrique infini ; il existe un entier λ tel que pour tout mot $w \in L$ assez grand ($|w| \geq \lambda$), le mot w se décompose en

$$w = xuyvz, \quad |uv| \geq 1,$$

et le langage L admet pour sous-ensemble

$$\{xu^n yv^n z \mid n \geq 0\} \subset L.$$

Preuve : Si L est infini, les arbres de dérivation sont de profondeur non bornée (car le degré de chaque sommet est borné) ; les branches sont de longueur non bornée et quand celle-ci dépasse le nombre de variables, l'une d'entre elles au moins apparaît deux fois. De plus, on peut choisir une grammaire sans ε -productions et débarrassée des productions du type $A \rightarrow B$. La situation est alors celle de la figure 10.5 et le sous-arbre de racine A produisant le mot uAv , $uv \neq \varepsilon$, peut être itéré arbitrairement :

$$A \xrightarrow{*} uAv \xrightarrow{*} uuAvv \xrightarrow{*} uuuAvvv \xrightarrow{*} u^n Av^n \xrightarrow{*} \dots,$$

laquelle dérivation combinée avec $S \xrightarrow{*} xAy$ et $A \xrightarrow{*} y$ conduit à l'énoncé.

On notera que seule la contrainte $uv \neq \varepsilon$ est imposable ; en effet, le langage peut parfaitement être rationnel et alors l'un des deux u ou v peut être vide. Enfin la valeur de λ peut être fixée à $\kappa^{|V|}$, où κ est la longueur de la plus grande production.

On en déduit alors que certains langages ne peuvent pas être décrits par des grammaires hors-contexte. C'est le cas de tous les langages qui comportent des systèmes de parenthèses croisés ou encore dont les contraintes correspondent à des nœuds.

Proposition 10.6.1 *Le langage $\{a^n b^n c^n \mid n \geq 0\}$ n'est pas algébrique.*

Preuve : S'il l'était, pour n assez grand on aurait, par le lemme d'itération, une décomposition du type $a^n b^n c^n = xuyvz$ et une contradiction car en itérant u et v , on ne peut pas maintenir la forme globale des mots, comme on l'a vu pour les langages rationnels.

On en déduit donc en particulier que l'intersection de deux langages algébriques ne l'est pas forcément.

10.6.2 Langage générateur

Pour terminer, nous allons montrer que tout langage algébrique peut s'obtenir à partir d'un seul et même langage appelé *langage générateur* par application des deux opérations de morphisme, et intersection avec un langage rationnel.

Définition 10.6.1 *Soit $X_n = \{(,)_i \mid 1 \leq i \leq n\}$ un alphabet à $2n$ symboles. On appelle langage de Dyck D_n^* l'ensemble des mots bien parenthésés sur cet alphabet et qui est défini par la grammaire d'axiome S :*

$$S = \sum_{1 \leq i \leq n} (,)_i S + \varepsilon.$$

Étant donné un langage $L \subset T^*$ dont la grammaire $\mathcal{G} = \langle V, T, S, \Delta \rangle$ est sous forme normale de Chomsky, nous allons montrer qu'il existe un homomorphisme h et un langage rationnel K tels que, pour un certain n ,

$$L = h(D_n \cap K).$$

En fait nous allons utiliser un ensemble de parenthèses P indicées de façon plus complexe (et surtout plus parlante), et nous considérerons le langage D_P des mots bien parenthésés sur P mais il sera toujours possible de les renuméroter sur \mathbb{N} . Il est défini par les trois règles suivantes.

1. Les lettres de l'alphabet terminal T seront représentées par la paire de parenthèses $(a)_a$, pour tout $a \in T$.
2. Chaque usage d'une règle terminale du type $A \rightarrow a$ sera rendu par une paire de parenthèses $\{A\}_A^a$.
3. Chaque usage d'une règle terminale du type $A \rightarrow BC$ sera rendu par un mot de parenthèses $\{A\}_A^{BC} (\frac{BC}{A})_A^{BC} [\frac{BC}{A}]_A^{BC} \}_A^{BC}$.

Considérons le langage \hat{L} défini par la grammaire $\hat{\mathcal{G}} = \langle V \cup T, P, S, \hat{\Delta} \rangle$ dont les productions sont définies par :

1. pour tout $a \in T$, $a \rightarrow (a)_a$;
2. pour tout $A \rightarrow a \in \Delta$, $A \rightarrow \{A\}_A^a$;
3. pour tout $A \rightarrow BC \in \Delta$, $A \rightarrow \{A\}_A^{BC} (\frac{BC}{A})_A^{BC} [\frac{BC}{A}]_A^{BC} \}_A^{BC}$.

Associons alors à toute dérivation d'un mot w dans L , le mot de parenthèses obtenues en appliquant les règles homologues de $\hat{\mathcal{G}}$; le mot obtenu \hat{w} contient à l'évidence toute la dérivation de w et permet de le retrouver en effaçant toutes les parenthèses autres que les (a) . Cette propriété se vérifie par récurrence sur les dérivations ; donc

$$L = h(\hat{L}),$$

où h est ce morphisme effaceur. De plus, on peut faire un certain nombre de déductions de proximité.

R1 : Tout mot de \hat{L} dérivé par exemple d'une règle $A \rightarrow a \in \Delta$, est forcément du type $\{A\}_A^a (a)_a \}_A^a$, contraintes qui sont rationnelles car locales.

R2 : Tout mot de \hat{L} dérivé par exemple d'une règle $A \rightarrow BC \in \Delta$, commence par $\{A\}_A^{BC} (\frac{BC}{A})_A^{BC}$, se termine par $[\frac{BC}{A}]_A^{BC} \}_A^{BC}$ et doit contenir un digramme $)_A^{BC} [\frac{BC}{A}$ à l'exclusion de toute autre combinaison.

R3 : Dans tout mot du langage \hat{L} :

- une lettre de type $(\frac{BC}{A})_A^{BC}$ doit être suivie d'une lettre de type $\{x\}_B^x$ où x est une lettre de T ou un digramme de V^2 ;
- une lettre de type $[\frac{BC}{A}]_A^{BC}$ doit être suivie d'une lettre de type $\{x\}_C^x$ où x est une lettre de T ou un digramme de V^2 ;
- une lettre de type $)_A^{BC}$ doit être précédée d'une lettre de type $\{x\}_B^x$ où x est une lettre de T ou un digramme de V^2 ;
- une lettre de type $[\frac{BC}{A}]_A^{BC}$ doit être précédée d'une lettre de type $\{x\}_C^x$ où x est une lettre de T ou un digramme de V^2 .

R4 : Les mots du langage \hat{L} comencent par une lettre du type $\{S\}_S^{BC}$ ou $\{S\}_S^a$ et se terminent par une $\}_S^{BC}$ ou $\}_S^a$.

Toutes ces contraintes sont exprimables par un automate fini qui définit donc sur l'alphabet P un langage K et l'on a $\hat{L} \subset D_P \cap K$.

Réciproquement, soit un mot \hat{w} de $K_P = D_P \cap K$; il est donc mot de parenthèses, et de plus il vérifie les contraintes de K . Par récurrence sur ses parenthèses, on vérifie qu'il est analysable selon la grammaire $\hat{\mathcal{G}}$. En effet :

– tout mot de K_P contenant une des quatre lettres de la séquence $\{ {}_A^a (a)_a \}_A^a$ contient forcément cette séquence qui dans $\hat{\mathcal{G}}$ est produite par une règle $A \rightarrow a \in \Delta$, par les règles R1 et R2 ;

– tout mot contenant une des six lettres de la séquence $\{ {}_A^{BC} ({}_A^{BC})_A^{BC} [{}_A^{BC}]_A^{BC} \}_A^{BC}$ contient toutes les autres et dans cet ordre du fait de son appartenance à D_P et à K par la règle R3, et si, par hypothèse de récurrence ses facteurs w_B et w_C placés entre les parenthèses homologues $({}_A^{BC})_A^{BC}$ et $[{}_A^{BC}]_A^{BC}$ respectivement sont engendrés (nécessairement à partir des variables B et C) par la grammaire $\hat{\mathcal{G}}$, alors le mot $w = \{ {}_A^{BC} ({}_A^{BC} w_B)_A^{BC} [{}_A^{BC} w_C]_A^{BC} \}_A^{BC}$ l'est également à partir de A en raison de la construction de la grammaire ;

– enfin la règle R4 impose que le mot soit issu de l'axiome de la grammaire.

En conséquence,

$$\hat{L} = D_P \cap K$$

et par application du morphisme effaceur h , on en déduit le théorème.

Théorème 10.6.2 *Étant donné un langage algébrique L , il existe un langage de Dyck D , un langage rationnel K et un morphisme h tels que :*

$$L = h(D \cap K).$$

Exercices

1. Décrire des automates à pile pour les langages suivants :

a) $\{w c w^R \mid w \in \{a, b\}^*\}$;

b) $\{w w^R \mid w \in \{a, b\}^*\}$;

c) $\{w \mid |w|_a = |w|_b\}$;

d) $\{a^n b^m \mid n \leq m \leq 2n\}$;

e) le langage engendré par la grammaire

$$\mathcal{G} = \langle \{S, A\}, \{a, b\}, S, \{S = aAA, A = bS + aS + a\} \rangle.$$

2. Donner une grammaire pour le langage reconnu par l'automate s'arrêtant sur pile vide et défini par

$$M = \langle \{q_0, q_1\}, \{0, 1\}, \{Z_0, X\}, q_0, Z_0, \delta \rangle,$$

où δ consiste en les actions :

$$\begin{array}{ll} \delta(q_0, 1, Z_0) = \{(q_0, XZ_0)\} & \delta(q_0, \varepsilon, Z_0) = \{(q_0, \varepsilon)\} \\ \delta(q_0, 1, X) = \{(q_0, XX)\} & \delta(q_1, 1, X) = \{(q_1, \varepsilon)\} \\ \delta(q_0, 0, X) = \{(q_1, X)\} & \delta(q_1, 0, Z_0) = \{(q_0, Z_0)\} \end{array}$$

3. Montrer que l'on peut définir l'acceptation d'un automate à pile par un ensemble d'états finaux sans modifier la puissance de calcul du modèle.

4. Quels sont les automates déterministes dans l'exercice 1 ci-dessus ?

5. Montrer que, pour tout automate à pile, on peut lui trouver un équivalent à un seul état.

6. Appliquer l'algorithme d'élimination des récurrences gauches à la grammaire formée des règles

$$A_1 = A_2 A_2 + 0, A_2 = A_1 A_2 + 1.$$

7. Donner une méthode pour calculer le nombre de dérivations possibles pour les mots de longueur n dans une grammaire en forme de Chomsky. Peut-on en tirer une information en ce qui concerne l'ambiguïté de la grammaire ? Appliquer cette méthode au langage $\{a^i b^j c^k \mid (i = j) \vee (j = k)\}$.

8. Montrer que le langage $K_n = \{a^{2^k} \mid k \leq n\}$ est rationnel et construire une expression régulière ou un automate. Exprimer cet ensemble au moyen d'une grammaire algébrique. Quel est le rapport des tailles de description ?

Montrer que le langage $K_* = \{a^{2^k} \mid k \geq 0\}$ n'est pas algébrique.

9. Que peut-on dire des langages

$$L_{15} = \{ab^{i_1} ab^{i_2} a \dots ab^{i_{15}} \mid \exists l \ i_l = l\}$$

et

$$L_* = \{ab^{i_1} ab^{i_2} a \dots ab^{i_k} \mid k \geq 1 \ \& \ \exists l \leq k \ i_l = l\}?$$

Chapitre 11

Analyse syntaxique et autres langages

La théorie des langages de parenthèses est d'une grande importance par ses applications dans l'analyse des langues naturelles et dans celle des langages de programmation. Dans le cas des langues naturelles, elle a été à l'origine d'un changement radical dans la méthodologie, mais nous ne développerons pas ce point. Pour ce qui est de l'analyse syntaxique des programmes, premier point obligé de la compilation et de l'interprétation, cette théorie, construite en même temps que l'on concevait les premiers langages de programmation comme Fortran, Lisp et Cobol, a permis de se sortir de l'approche pragmatique du groupe Fortran pour autoriser le développement d'outils systématiques et puissants.

Nous aborderons ensuite quelques prolongements sur d'autres classes de langages aux caractéristiques bien différentes et dont l'utilité n'est pas moindre.

11.1 Éléments d'analyse syntaxique

L'analyse syntaxique vise à fournir pour une phrase, un texte, un programme une traduction abstraite qui permette de lui donner une forme dénuée de toute ambiguïté en vue de la traduire dans un autre langage ou un univers sémantique bien défini.

On a vu au chapitre précédent que la reconnaissance des mots d'un langage algébrique se fait en temps $O(n^3)$ par une méthode de programmation dynamique. Les améliorations basées sur la multiplication matricielle rapide ne sont pas compétitives dans la pratique en raison de la trop grande complexité de mise en place de ces algorithmes récursifs.

On recherche des sous-classes de grammaires qui permettent de réaliser cette opération en temps linéaire et qui produisent également un arbre d'analyse syntaxique comme résultat : c'est cet arbre qui définira une première forme de sémantique pour le texte analysé, qu'il soit un programme ou un texte littéraire.

11.1.1 Analyse descendante

Considérons la grammaire \mathcal{G}_1 d'axiome S dont les règles sont les suivantes :

$$\begin{aligned} S &= aB + bCc + c, \\ B &= bB + aCC, \\ C &= aSS + c. \end{aligned}$$

Le mot $w = aaacc$ appartient au langage engendré et ne possède qu'un seul arbre de dérivation correspondant à la dérivation gauche : $S \rightarrow aB \rightarrow aaCC \rightarrow aaaSSC \rightarrow aaacSC \rightarrow aaaccC \rightarrow aaacc$.

En effet, on constate que puisque w commence par a , seule la règle $S \rightarrow aB$ a pu être utilisée, puis que seule $B \rightarrow aCC$ a pu fournir le second a de w , et ainsi de suite car la grammaire \mathcal{G}_1 a la propriété que chaque paire (α, X) , pour $\alpha \in A$ et $X \in V$, détermine de façon unique la production à appliquer et donc conduit à une analyse *déterministe et linéaire* des mots du langage. En fait c'est aussi l'automate à pile trivialement associé à la grammaire qui est déterministe et que l'algorithme suggéré ci-dessus ne fait que simuler. Cette situation peut se généraliser comme le montrent les exemples suivants.

Soit la grammaire \mathcal{G}_2 d'axiome S dont les règles sont les suivantes :

$$\begin{aligned} S &= aBBB + bC, \\ B &= bB + aC, \\ C &= SS + c. \end{aligned}$$

Cette grammaire n'a plus la bonne propriété de \mathcal{G}_1 , cependant on constate que si l'on développe le premier S de la production $C = SS$ en $C = aBBBS + bCS$, on se ramène au cas précédent et donc on sait analyser les mots du langage en temps linéaire.

La grammaire \mathcal{G}_3 d'axiome S et règles

$$\begin{aligned} S &= aBB + bC, \\ B &= CB + aC, \\ C &= abS + c \end{aligned}$$

nous met dans une situation encore différente : la simple réécriture de $B = CB$ en $B = abSB + cB$ ne permet pas de lever l'indétermination en raison de l'existence de la production $B = aC$. Toutefois si l'on développe encore ce C , on arrive à remplacer les deux productions initiales de B par $B = abSB + cB + aabS + ac$, et alors on peut savoir quelle production appliquer quand on doit dériver un B produisant un a en regardant la lettre suivante ! Cette anticipation (*look-ahead* en anglais) conduit à un algorithme d'analyse du même style et linéaire.

En toute généralité cette méthode s'appuie sur une propriété des dérivations gauches que nous formalisons maintenant.

Définition 11.1.1 Soit FG_k la fonction définie sur A^* pour $k \in \mathbb{N}$ par

$$FG_k(f) = \begin{cases} f & \text{si } |f| \leq k, \\ g & \text{tel que } f = gh \text{ \& } |g| = k \text{ sinon.} \end{cases}$$

Une grammaire est $LL(k)$ si et seulement si pour toute variable v admettant deux dérivations telles que $v \xrightarrow{*} f_1w_1$ et $v \xrightarrow{*} f_2w_2$, avec $f_1, f_2 \in T^k$ et $w_1, w_2 \in (T \cup V)^*$ et de surcroît

$$FG_k(f_1) \neq FG_k(f_2),$$

alors nécessairement les deux dérivations divergent dès la première réécriture :

$$\begin{array}{l} v \rightarrow \hat{v}_1 \xrightarrow{*} f_1w_1 \\ v \rightarrow \hat{v}_2 \xrightarrow{*} f_2w_2, \text{ avec } \hat{v}_1 \neq \hat{v}_2. \end{array}$$

Cette définition est clairement la formalisation des remarques informelles faites en début de paragraphe. On en déduit immédiatement que la méthode de reconnaissance esquissée ci-dessus se généralise au cas des grammaires $LL(k)$. L'automate à pile va prendre de l'avance et stocker dans son mécanisme d'états finis les k derniers symboles lus, ce qui lui permet de prendre la bonne décision en matière de dérivation... Il faut bien sûr préparer ces décisions, ce qui se fait en construisant des tables pour l'automate à pile, qui du même coup est déterministe. On renvoie à l'une des éditions du Dragon [1] pour les astuces d'implantation.

Proposition 11.1.1 *Tout langage engendré par une grammaire $LL(k)$ est reconnaissable par un automate à pile déterministe en temps linéaire.*

Pour terminer ce paragraphe, notons que l'on peut tester la propriété d'être une grammaire $LL(k)$ en temps $O(|\mathcal{G}|^k)$, ce qui est en gros le coût de construction de l'automate (ou des tables du programme). Enfin le fait pour un langage d'être engendré par une grammaire $LL(k)$ entraîne sa non-ambiguïté.

11.1.2 Analyse ascendante

Tous les langages ne sont pas forcément susceptibles de posséder une grammaire $LL(k)$: ainsi les langages produisant des systèmes de parenthèses comme celui donné par la grammaire \mathcal{G}_4 ci-dessous d'axiome S

$$\begin{array}{l} S = [S] + R, \\ R = E \equiv E, \\ E = [E \otimes E] + a + b, \end{array}$$

pour lequel il est impossible de décider, quel que soit le k , si les parenthèses initiales relèvent de S ou de E ...

La levée des ambiguïtés ne peut se faire que si l'on reconstruit différemment l'arbre syntaxique sous-jacent, en l'occurrence à partir de la droite (et non plus de gauche) ce qui est possible si la grammaire a certaines caractéristiques.

Nous allons considérer dans ce paragraphe des dérivations *droites* uniquement. On fera aussi la supposition que toutes les variables sont utiles, en ce sens qu'elles contribuent à former des mots du langage : on sait qu'il est facile de purifier la grammaire pour se ramener à ce cas.

Définition 11.1.2 *Une dérivation est appelée droite si et seulement si à chaque étape la réécriture porte sur le non-terminal le plus à droite.*

Dans ce paragraphe nous utiliserons les conventions typographiques suivantes : tout mot $w \in (V \cup T)^*$ est scandé en $w = \alpha A f$, où $f \in T^*$ est son plus long facteur droit terminal, $A \in V$ est la variable la plus à droite sur laquelle portera la réécriture et $\alpha \in (V \cup T)^*$ est le contexte gauche.

La réécriture (droite) de A selon la règle $A = \beta$, pour $\beta \in (V \cup T)^*$ donnera alors :

$$\alpha A f = w \rightarrow_d w' = \alpha \beta f.$$

Toute la question est de déterminer dans quelles conditions la connaissance d'information sur β permet d'en déduire que c'est précisément la règle $A = \beta$ qui avait été appliquée. Pour ce faire nous introduisons le concept fondamental de toute la reconnaissance montante (ou droite) : le contexte gauche d'une variable.

Définition 11.1.3 *Soit \mathcal{G} une grammaire. Pour toute variable $A \in V$ on appelle contexte gauche de la variable l'ensemble des facteurs gauches qu'elle définit en étant en position droite dans une dérivation droite :*

$$CG(A) = \{\alpha \mid S \xrightarrow{*}_d \alpha A f, \alpha \in (V \cup T)^*, f \in T^*\}.$$

Par exemple pour la grammaire de règles $S = aAb$ et $A = aAcAb + c$ et d'axiome S , on a :

- $CG(S) = \{\varepsilon\}$;
- $CG(A) = a(aAc)^*(a + \varepsilon)$.

Le lecteur est invité à construire les contextes gauches des variables de la grammaire \mathcal{G}_4 ci-dessus.

On a la propriété remarquable qui conditionne toutes les constructions à venir.

Proposition 11.1.2 *Pour toute grammaire \mathcal{G} et toute variable A utile, le contexte gauche de A est un langage rationnel : $CG(A) \in \mathcal{Rat}$.*

Preuve : Les règles de \mathcal{G} où A apparaît s'écrivent de façon générique sous la forme $B_i = \gamma_i A \delta_i$ avec $\alpha_i, \delta_i \in (V \cup T)^*$: on doit noter que cette scansion est différente de celle de la convention du début du paragraphe et qu'il y a autant de telles décompositions, pour une règle donnée, qu'il y a d'occurrences de la lettre A dans le membre droit de la règle. Les langages de contextes gauches vérifient alors le système d'équations

$$CG(A) = CG(B_1)\gamma_1 + CG(B_2)\gamma_2 + \dots + \varepsilon(A),$$

où $\varepsilon(A) = \text{si } (A = S) \text{ alors } \{\varepsilon\} \text{ sinon } \emptyset$.

Un tel système d'équations n'est autre qu'une grammaire linéaire droite, ou encore le système d'équations associé aux langages d'étiquettes des états d'un automate fini. Ces langages sont donc tous rationnels et nous avons construit implicitement un automate qui les reconnaît.

Définition 11.1.4 Soit $A = w$ une règle de la grammaire \mathcal{G} ; on appelle *contexte gauche de la règle $A = w$* le langage

$$CG(A = w) = CG(A)w.$$

Ces langages sont donc tous rationnels et effectivement calculables à partir de la grammaire.

Nous en arrivons à la notion cruciale qui va permettre de retrouver en fonction de ce qui a été lu sur le texte et de ce que la machine a mémorisé quelle production a été employée dans le processus de réécriture droite par lequel on tente d'analyser le texte. Pour ceci on étend la notation concernant les facteurs gauches.

Notation 11.1.1 Soit L un langage. On note $FG(L)$ l'ensemble des facteurs gauches de L :

$$FG(L) = \{f \mid \exists g \, fg \in L\}.$$

Définition 11.1.5 Une grammaire est dite $LR(0)$ si et seulement si pour toute paire de productions $A = w$ et $A' = w'$ on a la propriété :

$$CG(A = w) \cap FG(CG(A' = w')) = \emptyset.$$

En termes intuitifs, cette propriété permet de garantir que l'on sait caractériser les productions par la seule connaissance des facteurs gauches des mots de dérivation droite dans l'arbre syntaxique, et partant que l'on pourra le reconstruire du bas vers le haut et de droite à gauche. C'est le sens de la dénomination LR qui vient de l'anglais *Left-Right*.

Proposition 11.1.3 Il existe un algorithme pour déterminer si une grammaire est $LR(0)$.

Preuve : Les deux langages de contextes gauches sont rationnels, ainsi que le langage des facteurs gauches. Les rationnels sont clos par intersection et l'on sait déterminer si un rationnel est vide ou non.

Passons maintenant à l'algorithme de reconnaissance des mots du langage engendré par la grammaire $LR(0)$.

Chaque langage de contexte gauche de règle $CG(A = w)$ est reconnu par un automate $\mathcal{A}_{A=w}$: nous noterons, pour chaque règle R_i , l'automate correspondant par \mathcal{A}_{R_i} . On construit alors l'automate produit, pour toutes les règles de la grammaire,

$$\mathcal{A}_{\mathcal{G}} = \mathcal{A}_{R_1} \times \mathcal{A}_{R_2} \times \dots \times \mathcal{A}_{R_p}.$$

Cette construction est identique à celle de l'automate reconnaissant l'intersection de deux langages rationnels : les états de \mathcal{A}_G sont des p -uplets formés sur les états des automates \mathcal{A}_{R_i} à ceci près que l'on se soucie peu des états finaux de \mathcal{A}_G . Une autre façon de voir cet automate produit est de dire que l'on effectue les calculs séparés de chacun des p automates en parallèle.

L'alphabet de \mathcal{A}_G est $V \cup T$, son ensemble d'états $Q_{R_1} \times Q_{R_2} \times \dots \times Q_{R_p}$ et son état initial le p -uplet des états initiaux de chaque automate. La table de transitions s'obtient comme le produit des tables de chaque automate.

Proposition 11.1.4 *À chaque étape du calcul de \mathcal{A}_G sur un mot w , au plus une des composantes de l'état courant est état final de l'automate correspondant \mathcal{A}_{R_i} .*

Preuve : La propriété de la grammaire d'être $LR(0)$ est précisément que $CG(R_i) \cap FG(CG(R_j)) = \emptyset$, pour $i \neq j$, ce qui entraîne l'impossibilité pour deux automates d'être simultanément dans un état final. Il est parfaitement possible qu'aucun automate ne soit dans un état final – et c'est même ce qui se passe le plus souvent.

Considérons donc une dérivation droite pour un mot f de $\mathcal{L}(\mathcal{G})$, dans laquelle nous sélectionnons un sous-arbre particulier correspondant à la dérivation (droite) d'un non-terminal particulier :

$$S \xrightarrow{*}_d \alpha A w \rightarrow_d \alpha \beta w \xrightarrow{*}_d f,$$

où $\alpha, \beta \in (V \cup T)^*$, $w \in T^*$ et, de plus, $f = w'w$ avec $\alpha\beta \xrightarrow{*}_d w' \in T^*$.

Si l'on donne le mot $\alpha\beta$ à lire à l'automate \mathcal{A}_G , celui doit indiquer que $\alpha\beta \in CG(A = \beta)$, et donc que, en remontant la dérivation pour la reconstruire, $\alpha\beta$ devrait être réduit à αA .

Supposons, pour simplifier, que ce même A ait lui-même été produit par une règle de la forme $B = \gamma A x$ où $x \in T^*$; alors w admet nécessairement x pour facteur gauche, et l'automate \mathcal{A}_G doit reconnaître que le mot $\alpha A x$ est dans l'ensemble $CG(B = \gamma A x)$ et que donc $\delta B y \rightarrow \alpha A w$, avec $\alpha = \delta \gamma$ et $w = x y$. On peut donc réduire le mot courant $\alpha A w$ à $\delta B y$ et repartir avec les lettres de y pour poursuivre la réduction.

Dans le cas général, le x précédent contient des variables que le mécanisme aura su réduire en partant de la variable la plus à droite, selon un processus récursif que nous explicitons maintenant et qui se présente comme un automate à pile déterministe.

L'automate à pile :

- alphabet $X = V \cup T$;
- alphabet de pile $X \times Q_{\mathcal{A}}$, soit un ensemble de $(p + 1)$ -uplets composés d'une lettre de X et d'un état de \mathcal{A}_G .

Ses règles de fonctionnement sont données par le programme suivant qui s'appuie sur le fait qu'à tout instant la pile vérifie la propriété que si $\pi \in X^*$ est le mot de pile lu sur sa première composante alors le p -uplet de sommet de pile q_π est celui des états des p automates \mathcal{A}_{R_i} de contextes gauches des règles après lecture de ce mot π . De plus ceci est vrai à tous les niveaux de la pile, la preuve se faisant par récurrence sur le fonctionnement.

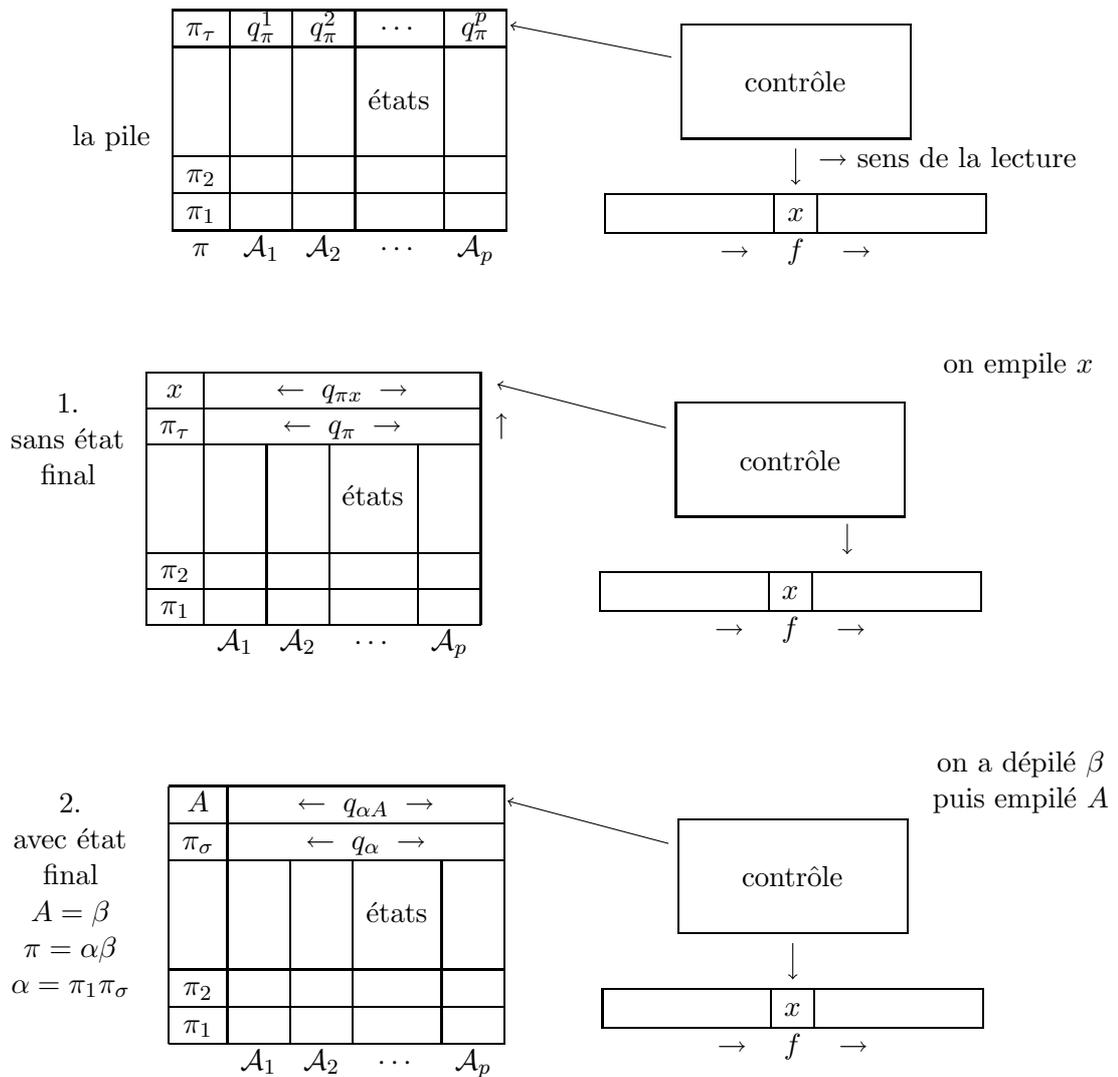


FIG. 11.1 – Une représentation de la pile et du fonctionnement de l'automate de reconnaissance $LR(0)$

Le fonctionnement est régi par les deux cas :

- si q_π n'a pas de composante qui soit état final d'un des \mathcal{A}_{R_i} , alors lire la lettre suivante x de la donnée et empiler $(x, q_{\pi x})$, où $q_{\pi x}$ est le p -uplet d'états obtenus en faisant lire à chaque automate la lettre x ;
- si q_π a une composante qui soit état final de \mathcal{A}_{R_i} (par exemple et on sait que ce i est unique), correspondant à la règle $A = \beta$, alors nécessairement β est suffixe de $\pi = \alpha\beta$ et l'automate dépile toute la partie correspondant à β ne laissant que le fond de pile avec α en première colonne ; puis il empile $(A, q_{\pi A})$.

Ainsi, à chaque instant où une composante du sommet de pile est un état final, ce mécanisme permet de garantir que le mot de pile est facteur gauche d'une dérivation droite légale et que toute dérivation droite légale conduira à avoir dans la pile le seul niveau (S, q_S) nécessaire et suffisant pour garantir l'appartenance du mot au langage. Toute autre situation conduit au blocage de l'automate et par conséquent au rejet de la donnée.

Ces grammaires sont une des catégories les plus utilisées en compilation ; on leur donne souvent un peu plus de souplesse en autorisant un certain *look-ahead* ce qui donne les grammaires *LALR* dont les tables sont beaucoup plus petites et dont le fleuron est le fameux YACC d'Unix (*Yet Another Compiler Compiler*) qui permet de fabriquer des compilateurs à partir d'une grammaire *LALR(1)*.

11.2 Au-delà des algébriques

On a montré que le langage $\{a^n b^n c^n \mid n \geq 0\}$ n'est pas algébrique en utilisant le lemme d'itération pour ces langages. Or il est de nombreux cas où des phénomènes de croissance multiple et contrôlée s'observent et donc où le cadre des grammaires algébriques est trop restrictif. Le premier à avoir fait ce genre de remarque est Sir D'Arcy Wentworth Thompson (1860-1948) qui fit le premier travail de mathématicien sur la croissance des formes en biologie (1917). Cette approche plutôt analytique fut reprise par Aristid Lindenmeyer qui développa une théorie des L-systèmes pour simuler la croissance des structures biologiques (1968)¹.

Nous donnons ici un bref aperçu de ces systèmes de réécriture.

11.2.1 L-systèmes

Dans les modèles hors-contextes des chapitres précédents, toutes les réécritures se font indépendamment les unes des autres. Lindenmeyer a eu l'idée d'imposer une réécriture de *toutes* les lettres à chaque étape de dérivation : ainsi avec la règle élémentaire $A = AA$, on fabrique comme ensemble de mots dérivés de l'axiome A , la famille $\{A^{2^k} \mid k \geq 0\}$ dont on vérifie assez facilement qu'elle n'est pas algébrique. Nous montrons ici à titre d'exemples trois classes couramment utilisées.

¹voir par exemple la page <http://www.xs4all.nl/~ljlapre/index.html>.

Les 0L-systèmes

Commençons par le plus simple des L -systèmes.

Définition 11.2.1 *Un 0L-système est défini par la donnée d'une grammaire $\mathcal{G} = \langle \Sigma, w, \mathcal{R} \rangle$, où Σ est un alphabet fini, $w \in \Sigma^+$ est l'axiome et \mathcal{R} est l'ensemble des règles — sous-ensemble fini de $\Sigma \times \Sigma^*$.*

À notre habitude chaque règle sera écrite sous la forme $A = f$. Notons que nous n'avons pas de distinction a priori entre lettres terminales et non-terminales. Un tel système peut être non-déterministe en ce sens qu'une même lettre peut se récrire de plusieurs manières différentes.

Un mot $u \in \Sigma^*$ se dérive en $v \in \Sigma^*$ — ce qu'on notera à l'usuel $u \rightarrow v$ — si et seulement si $u = u_1 u_2 \dots u_n$, $u_i \in \Sigma$, et il existe des règles $u_i = v_i$ dans l'ensemble \mathcal{R} telles que $v = v_1 v_2 \dots v_n$.

Le langage engendré par un tel système est l'ensemble des mots dérivables en un nombre fini de coups de l'axiome :

$$\mathcal{L}(\mathcal{G}) = \{f \mid w \xrightarrow{*} f\}.$$

Un bon exemple de 0L-système — 0 signifie “sans contexte” et L est évidemment un rappel de Lindenmeyer — est donné par :

$$\mathcal{F} = \langle \{\diamond, \circ\}, \circ, \{\circ = \diamond, \diamond = \circ\circ\} \rangle .$$

Les premiers mots de $\mathcal{L}(\mathcal{F})$ sont assez facilement vus pour être dans l'ordre :

$$\begin{aligned} f_0 & : \circ \\ f_1 & : \diamond \\ f_2 & : \circ \diamond \\ f_3 & : \diamond \circ \diamond \\ f_4 & : \circ \diamond \circ \diamond \\ f_5 & : \diamond \circ \diamond \circ \diamond \circ \diamond \\ & \text{etc.} \end{aligned}$$

Le lecteur montrera sans peine que $f_{n+2} = f_n f_{n+1}$ et que les longueurs de ces mots sont les nombres de Fibonacci.

Les 0L-systèmes sont trop contraints par les mécanismes de croissance en parallèle pour jouir des propriétés de fermeture prouvées pour les langages rationnels ou algébriques. On peut trouver des contre-exemples pour prouver que la famille des langages engendrés par les 0L-systèmes *n'est close par aucune* des opérations d'union, concaténation, intersection avec les rationnels, homomorphisme, etc.

Cette famille ne contient même pas les ensembles finis, comme par exemple $\{a, aaa\}$!

Les $E0L$ -systèmes

Pour étendre convenablement la famille des langages — d'où le E — il faut réintroduire un peu de souplesse dans la synchronisation et pour ce faire permettre des récritures non productives ultimement de façon à pouvoir mieux sélectionner. On introduit donc en plus des lettres de l'alphabet, des variables supplémentaires qui seront destinées à disparaître dans les mots sélectionnés : en d'autres termes on coupe l'ensemble produit par un $0L$ -système par un langage rationnel du type Σ^* .

Définition 11.2.2 *Un $E0L$ -système est défini par la donnée d'une grammaire $\mathcal{G} = \langle \Sigma \cup V, w, \mathcal{R} \rangle$, où Σ et V sont des alphabets finis, $w \in (V \cup \Sigma)^+$ est l'axiome et \mathcal{R} est l'ensemble des règles — sous-ensemble fini de $(V \cup \Sigma) \times (V \cup \Sigma)^*$. Les règles de production des mots dérivés de l'axiome w sont celles du paragraphe précédent, mais le langage engendré par ce système est restreint à Σ^* .*

Avec $V = \{S, A, B, C, \bar{A}, \bar{B}, \bar{C}, N\}$ comme variables, $\Sigma = \{a, b, c\}$ comme alphabet final, S comme axiome et

$$\begin{aligned} \mathcal{R} = \{ & S = ABC, A = A\bar{A} + a, B = B\bar{B} + b, C = C\bar{C} + c, \\ & \bar{A} = \bar{A} + a, \bar{B} = \bar{B} + b, \bar{C} = \bar{C} + c \\ & a = N, b = N, c = N, N = N\}, \end{aligned}$$

on construit comme seuls mots sur l'alphabet terminal $\{a, b, c\}$ les mots de la forme

$$\{a^n b^n c^n \mid n \geq 1\}.$$

En effet, si les A, B, C et les $\bar{A}, \bar{B}, \bar{C}$ ne se récrivent pas de manière homologue et synchrone, ils introduiront en passant dans l'état a, b, c respectivement des N ineffaçables...

On notera que ce langage n'est pas algébrique.

Contrairement à ce qui se passe pour les $0L$ -systèmes, la classe des langages définis par des $E0L$ -systèmes a des propriétés de clôture semblables à celles des langages algébriques qu'elle contient. Ce que nous énonçons sans démonstration, car elle est en tous points semblable à celle faite pour les langages algébriques.

Proposition 11.2.1 *La classe des langages définis par des $E0L$ -systèmes contient la famille des langages algébriques et est close par union, produit, quasi-inverse et intersection avec un rationnel.*

Le lecteur est invité à montrer que tout langage rationnel peut être engendré par un $E0L$ -système.

Les $ET0L$ -systèmes

La dernière généralisation présentée dans ce court exposé consiste à ranger les règles en plusieurs sous-ensembles, $\mathcal{R} = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \dots \cup \mathcal{R}_p$, appelés *tables* (d'où le T de l'acronyme) et à utiliser dans chaque récriture parallèle les règles de la même table :

la table varie évidemment au cours du temps. Le cas des *EOL*-système est un cas particulier à une seule table.

Considérons le système suivant dont l'alphabet terminal est $\Sigma = \{A, N, 1, (,)\}$, l'alphabet non-terminal est $V = \{S, R, T, F\}$, l'axiome est $w = S$ et les tables sont respectivement — en notant α un symbole a priori quelconque de l'un des deux alphabets :

$$\begin{aligned} \mathcal{R}_1 &= \begin{cases} S = (xAyAz)S + (xAyAz), & \text{pour } x, y, z \in \{T, NT, F, NF\} \\ & \text{et au moins un des } x, y, z \in \{T, NF\}, \\ \alpha = \alpha, & \text{pour } \alpha \neq S; \end{cases} \\ \mathcal{R}_2 &= \begin{cases} S = R, F = 1F, T = 1 + 1T, \\ \alpha = \alpha, \text{ pour } \alpha \neq S, T, F; \end{cases} \\ \mathcal{R}_3 &= \begin{cases} S = R, T = 1T, F = 1 + 1F, \\ \alpha = \alpha, \text{ pour } \alpha \neq S, T, F. \end{cases} \end{aligned}$$

Ce système produit par exemple le mot

$$(11AN1AN111)(AN11AN1111A111)(1A11AN1111),$$

que l'on va interpréter comme un codage de la formule booléenne

$$(x_2 + \bar{x}_1 + \bar{x}_3)(\bar{x}_2 + \bar{x}_4 + x_3)(x_1 + x_2 + \bar{x}_4)$$

dont on vérifie qu'elle est satisfaisable.

Avec cette interprétation des choses — ou plus exactement des mots engendrés par le système — on constate et prouve que le système engendre précisément les formules booléennes en FNC qui sont satisfaisables. On a soupçonné au premier chapitre que cet ensemble est difficile à caractériser algorithmiquement et nous étayerons cette idée dans le dernier chapitre.

Cette famille des langages *ETOL* est donc plus large que la précédente et vérifie les mêmes propriétés de clôture.

11.2.2 Autres classes

De fait il est possible de définir bien d'autres classes. Il existe des hiérarchies infinies de classes de langages reconnaissables par diverses adaptations ou extensions de grammaires, systèmes de réécriture, automates ou autres machines. On pourrait parler aussi de machines non-déterministes et on pourrait aussi évoquer les machines probabilistes.

Chapitre 12

Analyses lexicale et syntaxique par l'exemple

Une bonne référence pour ce qui suit reste le livre surnommé le dragon [1] (également traduit en français). On pourra également consulter [7]. On procédera plus par exemples qu'en théorisant le sujet.

12.1 Motivations

Un des problèmes fondamentaux de l'informatique est l'écriture de compilateurs. Être capable de récupérer de l'information dans un gros fichier de données est une autre problématique fréquente (penser à des résultats d'expériences). On a souvent à écrire de petits langages adaptés à des problèmes spécifiques, et transformer le programme en objets plus utilisables est primordial.

Le schéma type d'un compilateur classique se trouve à la figure 12.1.

Le programme source subit de multiples transformations avant que du code exécutable soit produit. Notez également le rôle non négligeable des interactions avec le gestionnaire d'erreurs, dont la présence est indispensable pour l'utilisateur.

Nous allons surtout insister sur le début de la chaîne, laissant la suite pour le cours de Compilation de Majeure 2.

L'analyse lexicale sert à isoler les lexèmes dans une chaîne de caractères.

Définition 12.1.1 *Un lexème (en anglais token) est une unité élémentaire importante d'une chaîne pour une phase ultérieure de calcul.*

Ex. De

```
public static void main(String[] args){
    int n = 2;
}
```

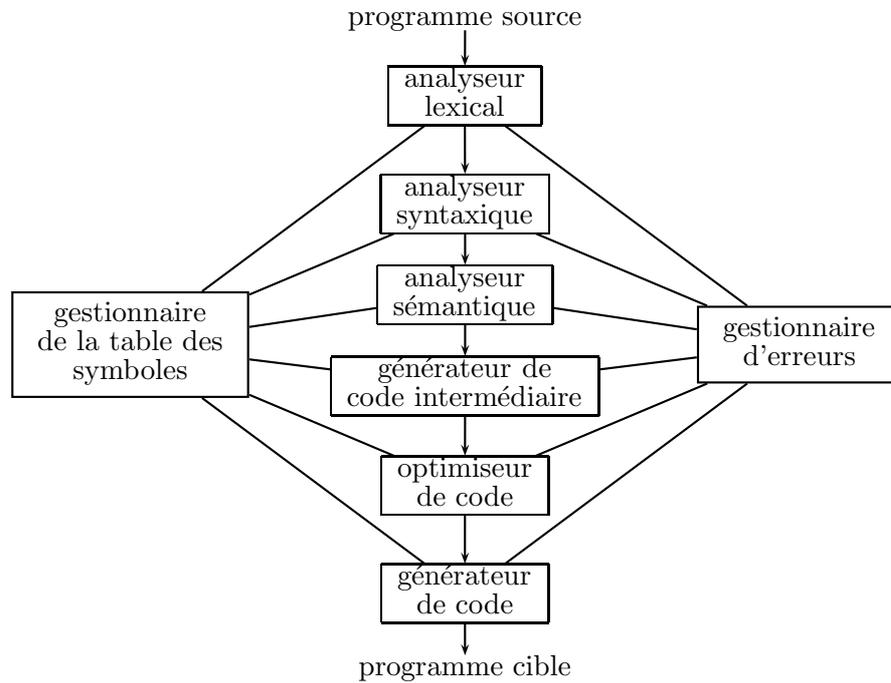


FIG. 12.1 – Schéma type d'un compilateur.

on récupère les mots clefs **public**, **static**, etc., des variables `args`, `n` et une constante `2`.

Plus généralement, on récupère un *identificateur*, une constante *numérique*, un *opérateur*, une *chaîne* de caractères, une constante *caractère*, etc. D'habitude, les espaces, tabulations, retours à la ligne, commentaires ne sont pas des lexèmes.

Donnons des exemples, avec comme source un morceau de programme Java, tout d'abord

```
" expression = 3 * x + 2 ; "
```

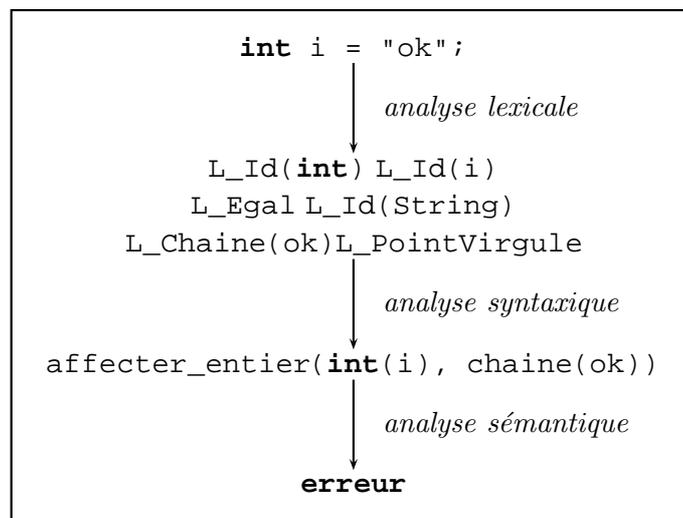
On obtient comme résultat la suite de lexèmes suivante

```
L_Id(expression) L_Egal L_Nombre(3) L_Mul L_Id(x)
L_Plus L_Nombre(2) L_PointVirgule L_EOF
```

On obtiendra le même résultat si la chaîne d'entrée est

```
" expression =
    3 * x
    + 2 ; "
```

Donnons un cas qui provoquera une erreur.



Intéressons-nous maintenant à un programme Java un peu plus vrai :

```

" class PremierProg{
  public static void main(String[] args){
    System.out.println(\"Bonjour!\");
  }
} "
  
```

On obtiendra alors :

```

L_Id(class) L_Id(PremierProg) L_AccG L_Id(public)
L_Id(static) L_Id(void) L_Id(main)
L_ParG L_Id(String) L_CroG L_CroD L_Id(args) L_ParD
L_AccG L_Id(System) L_Point L_Id(out) L_Point
L_Id(println) L_ParG L_Chaine(Bonjour!) L_ParD
L_PointVirgule L_AccD L_AccD L_EOF
  
```

12.2 Construction de l'analyseur lexical

Nous allons donner les principes de construction d'un analyseur lexical pour une version simplifiée de Java. On se donne trois lexèmes définis par des expressions régulières :

Identificateur = Lettre (Lettre | Chiffre)*

Entier = Chiffre Chiffre*

Operateur = + | - | * | /

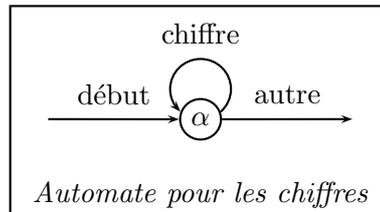
Lettre = a | b... | z | A | B... | Z

Chiffre = 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9

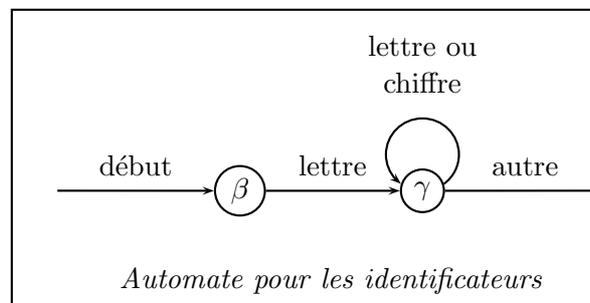
Blancs = (' ' | '\t' | '\r' | '\n')*

12.2.1 Utilisation des diagrammes de transition

Le principe de l'analyse est le suivant. On connecte entre eux des automates élémentaires qui savent reconnaître chaque lexème. Par exemple, un chiffre est reconnu par l'automate :

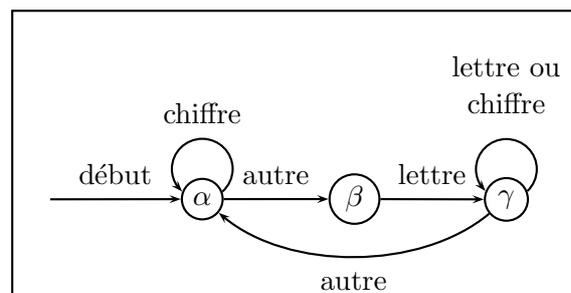


De même qu'un identificateur aura pour automate :



À chaque tour de boucle, on accumule les caractères lus. Si on sort avec une chaîne non nulle, on a trouvé un lexème et son type.

Il ne reste alors plus qu'à assembler les diagrammes de transition pour pouvoir reconnaître plus de lexèmes :



On a ainsi construit l'embryon du programme qui reconnaît un entier ou un identificateur. Une fois le schéma complété, il ne reste plus qu'à en déduire le programme correspondant.

12.2.2 Le programme en Java

On va représenter chacun des lexèmes par un objet contenant la nature du lexème (une constante entière), une valeur éventuelle (si on a affaire à une constante numérique), un nom (si nécessaire) :

```

class Lexeme{
    final static int L_Nombre = 0, L_Id = 1,
                  L_Plus = '+', L_Moins = '-',
                  L_Mul = '*', L_Div = '/',
                  L_ParG = '(', L_ParD = ')',
                  L_CroG = '[', L_CroD = ']',
                  L_EOF = -1;

    int nature;
    int valeur;
    String nom;

    Lexeme(int t, int i) { nature = t; valeur = i; }
    Lexeme(int t, String s) { nature = t; nom = s; }
    Lexeme(int t) { nature = t; }

    // aller chercher le lexème suivant
    static Lexeme suivant() { ... }

```

La principale difficulté d'un analyseur lexical tient à la gestion du caractère courant sur le tampon de lecture. Ici, on suppose que le caractère courant est une variable de classe. La méthode `avancer` est une des primitives les plus simples qui met à jour ce caractère courant. Dans la foulée, on écrit aussi une méthode qui "saute" les caractères inutiles (disons les espaces) :

```

static int c;    // caractère courant

static void avancer(){
    try{
        c = in.read();
    } catch (IOException e) { }
}

static void sauterLesBlancs(){
    while(Character.isWhitespace(c))
        avancer();
}

```

On peut alors écrire la méthode suivante :

```

static Lexeme suivant(){
    sauterLesBlancs();
    if(Character.isLetter(c))
        return new Lexeme(L_Id, identificateur());
    else if(Character.isDigit(c))
        return new Lexeme(L_Nombre, nombre());
}

```

```

else switch(c){
    case '+': case '-': case '*': case '/':
    case '(': case ')':
        char c1 = (char)c;
        avancer();
        return new Lexeme(c1);
    default:
        throw new Error ("Caractère illégal");
}
}

```

Le programme précédent n'est pas complet, car il ignore les problèmes liés à la fin de fichier. On laisse cela au lecteur.

Le traitement des deux automates pour les deux expressions régulières simples sont alors :

```

static String identificateur(){
    StringBuffer r;
    while(Character.isLetterOrDigit(c)){
        r = r.append(c);
        avancer();
    }
    return r;
}
static int nombre(){
    int r = 0;
    while(Character.isDigit(c)){
        r = r * 10 + c - '0';
        avancer();
    }
    return r;
}

```

12.2.3 Pour aller plus loin

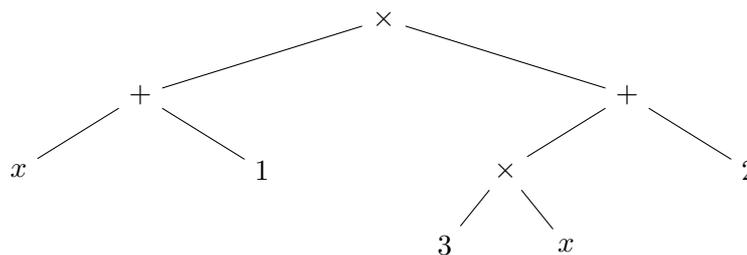
Quand il y a de nombreux mots-clé, on leur donne (comme pour les opérateurs) des numéros de lexème distincts. Chaque lexème identificateur est comparé à une *table des mots-clé*. En fonction de la taille, on peut se contenter d'une table séquentielle, ou bien d'une table de hachage, ou encore d'un arbre de recherche. Nous renvoyons au dragon pour plus de détails si besoin est.

Dans la vraie vie, il existe de nombreux outils permettant de simplifier l'analyse lexicale, comme `lex` dans le monde Unix. En Java, et après avoir écrit son premier lexeur à la main, il est recommandé d'utiliser la classe `StringTokenizer`.

12.3 Analyse syntaxique

12.3.1 Arbre de syntaxe abstraite

Le résultat de l'analyse syntaxique d'un texte (par exemple d'un programme ou d'une formule) sera souvent un *arbre de syntaxe abstraite* (ASA), c'est-à-dire une structure de données adaptée au traitement subséquent. Par exemple, à partir de l'expression arithmétique $(x + 1) * (3 * x + 2)$, on pourra construire l'arbre



puis se servir de cet arbre pour évaluer l'expression en $x = 10$, ou encore afficher la formule de façon agréable.

Donnons une implantation simple de ce concept, basé sur la grammaire simplifiée des expressions arithmétiques, et utilisant un type Terme qui est l'union des différents types possibles.

```

class Terme{

    final static int ADD = 0, SUB = 1, MUL = 2,
                  DIV = 3, MINUS = 4,
                  VAR = 5, CONST = 6;

    int nature;
    Terme a1, a2;
    String nom;
    int val;

    // des constructeurs pratiques
    Terme(int t, Terme a) {nature = t; a1 = a; }
    Terme(int t, Terme a, Terme b){
        nature = t; a1 = a; a2 = b;
    }
    Terme(String s) {nature = VAR; nom = s; }
    Terme(int v) {nature = CONST; val = v; }
}

```

Un terme correspondant à l'expression $(x + 1) * (3 * x + 2)$ pourra être créé par :

```

Terme t = new Terme(MUL,
    new Terme(ADD, new Terme("x"), new Terme(1)),

```

```

new Terme(ADD,
    new Terme(MUL, new Terme(3), new Terme("x")),
    new Terme(2));

```

Une telle expression pourra être évaluée dans un *environnement* e , c'est-à-dire d'une table des valeurs des variables, typiquement une table de couples (nom, valeur). Le code correspondant pourra être :

```

static int evaluer(Terme t, Environnement e){
    switch(t.nature){
        case ADD: return evaluer(t.a1, e) + evaluer(t.a2, e);
        case SUB: return evaluer(t.a1, e) - evaluer(t.a2, e);
        case MUL: return evaluer(t.a1, e) * evaluer(t.a2, e);
        case DIV: return evaluer(t.a1, e) / evaluer(t.a2, e);
        case CONST: return t.val;
        case VAR: return valeurDe(t.nom, e);
        default: throw new Error ("Erreur dans evaluation");
    }
}

static int valeurDe(String s, Environnement e) {
    if(e == null) throw new Error("Variable non définie");
    if(e.nom.equals(s)) return e.val;
    else return valeurDe(s, e.suivant);
}

```

Exercice 12.3.1 Écrire un code Java pour dériver formellement un Terme par rapport à une variable donnée.

Les arbres de syntaxe abstraite sont donc très utiles pour manipuler des programmes ou des formules. Il reste maintenant à savoir passer de l'expression proprement dite à l'ASA correspondant.

12.3.2 Construction des ASA

Le cas des structures parenthésées

C'est le cas le plus simple, comme déjà esquissé au chapitre 10. L'écriture $(x + 1) * (3 * x + 2)$ est habituelle, mais on pourrait aussi l'écrire (formes polonaises) :

- en notation préfixée : $* + x1 + *3x2$
- en notation postfixée : $x1 + 3x * 2 + *$

Ces deux formules sont très faciles à analyser pour construire l'ASA. Esquisons un programme Java qui construirait un ASA à partir d'une expression préfixée :

```

static Lexeme lc; // lexème courant

```

```

static void avancer() {lc = Lexeme.suivant(); }

static Terme lireExpr(){
  if(lc.nature == Lexeme.L_Mul){
    avancer(); Terme g = lireExpr();
    avancer(); Terme d = lireExpr();
    return new Terme(MUL, g, d);
  }
  else if(lc.nature == Lexeme.L_Plus){ ... }
  else if(lc.nature == Lexeme.L_Nombre){
    return new Terme(lc.valeur);
  }
  else if(lc.nature == Lexeme.L_Id){
    return new Terme(lc.nom);
  }
  else{ ... }
}

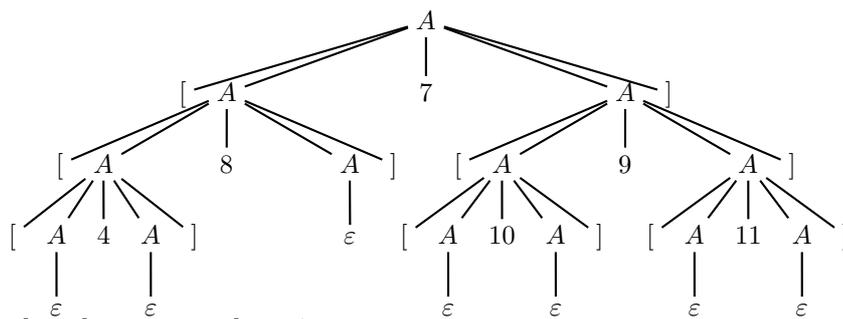
```

Un exemple d'analyse descendante

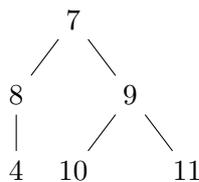
Contentons-nous de l'exemple d'une grammaire $LL(1)$, et considérons l'exemple de la représentation linéaire d'arbre binaire, dont la grammaire est définie par :

$$\begin{aligned} \Sigma &= \{[,], nb\}, V = \{A\} \\ A &\rightarrow [A nb A] \\ A &\rightarrow \end{aligned}$$

Par exemple, le mot $[[[4]8]7[[10]9[11]]]$ aura pour arbre de syntaxe concrète (ou *arbre syntaxique*) :



et arbre de syntaxe abstraite :



202 *CHAPITRE 12. ANALYSES LEXICALE ET SYNTAXIQUE PAR L'EXEMPLE*

Le principe de la méthode d'analyse descendante est de partir de l'axiome A en choisissant l'une des deux règles de production en fonction du premier lexème (terminal) de la chaîne d'entrée.

On commence par lire un crochet gauche

[[[4] 8] 7 [[10] 9 [11]]]
 ↑

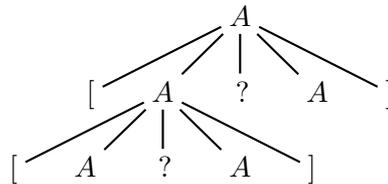
ce qui donne le début de l'arbre :



On lit alors le deuxième crochet gauche :

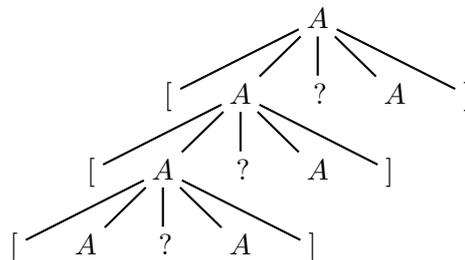
[[[4] 8] 7 [[10] 9 [11]]]
 ↑

ce qui déclenche :



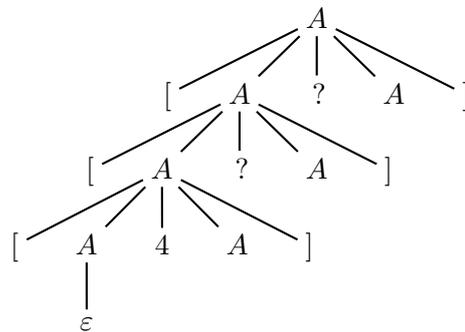
Le troisième crochet conduit à :

[[[4] 8] 7 [[10] 9 [11]]]
 ↑

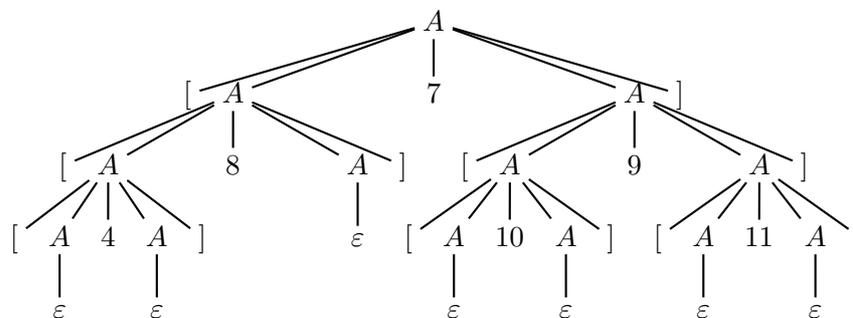


La lecture du 4 conduit à :

[[[4] 8] 7 [[10] 9 [11]]]
 ↑



et finalement, de proche en proche, on trouve :



Un programme Java qui implanterait cette idée est le suivant :

```

static Lexeme lc; // lexème courant
static void avancer() {lc = Lexeme.suivant(); }

static Arbre lireArbre(){
  if(lc.nature == Lexeme.L_CroG){
    avancer(); Arbre a = lireArbre();
    if(lc.nature == Lexeme.L_Nombre){
      int x = lc.val;
      avancer(); Arbre b = lireArbre();
      if(lc.nature == Lexeme.L_CroD){
        avancer(); return new Arbre(x, a, b);
      } else
        throw new Error("Syntaxe: \"\"]\" manquant.");
    } else
      throw new Error("Syntaxe: nombre manquant.");
    }
  }
  else return null;
}

```

Expressions générales

Si le programme précédent est particulièrement facile, c'est que la forme préfixée se prête remarquablement bien à la traduction. Pour le faire à partir de la forme postfixée, il faut utiliser une pile qui permet de mémoriser les sous-expressions construites et de les assembler au fur et à mesure que la lecture d'opérateurs conduira à le faire. Cette construction est laissée à titre d'exercice.

Dans le cas général, il faut revenir à la description des expressions et à leur décomposition hiérarchique. Sans chercher à expliciter plus le formalisme rappelons que les formes valides d'expressions Caml, C ou Java sont usuellement définies par des équations dont nous nous inspirons pour exprimer ce qu'est une expression arithmétique valide ainsi que son interprétation induite des règles de priorité énoncées en début de chapitre. Cette description est appelée forme de Backus-Naur en hommage à deux grands concepteurs de compilateurs, pour Fortran et Algol respectivement.

Quatrième partie

Annexes

Chapitre 13

Compléments de Java

13.1 Paquetages et espace des noms

13.1.1 Définition et construction

En Java, il existe déjà de nombreuses structures de données = objets (ou tableaux). Cela fait beaucoup de classes, de fichiers `.class`, et un fort risque de collisions dans l'espace des noms.

D'où l'idée de rassembler les classes en *paquetages* (en anglais, *package*). La syntaxe d'utilisation est la suivante :

```
package ma_lib1;           package ma_lib2;  
public class FIFO{...} public class FIFO{...}
```

Pour utiliser un paquetage pour l'utiliser, on doit *l'importer* :

```
import ma_lib1.FIFO;  
class Test{  
    public static void main(String[] args){  
        int n = Integer.parseInt(args[0]);  
        FIFO f = new FIFO(n);  
        FIFO.ajouter(1, f);  
    }  
}
```

L'instruction

```
package id1.id2...idk;
```

qualifie les noms des champs publics d'une unité de compilation, comme par exemple

- `ma_lib1.FIFO`, `ma_lib1.FIFO.ajouter`, `ma_lib1.FIFO.supprimer`
(premier paquetage)
- `ma_lib2.FIFO`, `ma_lib2.FIFO.ajouter`, `ma_lib2.FIFO.supprimer`
(deuxième paquetage)

On référence ces champs publics avec leurs *noms qualifiés* (classe `Test`) ou avec une *forme courte* si on importe la classe avant son utilisation

```
import id1.id2...idk.C;
```

où C est un nom de classe ou le symbole $*$ pour importer toutes les classes d'un paquetage.

L'emplacement des paquetages dépend de deux paramètres :

- le *nom* $id_1.id_2\dots id_k$ qui désigne l'emplacement $id_1/id_2/\dots/id_k$ dans l'arborescence des répertoires du système de fichiers.
- la *variable d'environnement* CLASSPATH qui donne une suite de racines à l'arborescence des paquetages.

La valeur de CLASSPATH est fixée par une commande Unix :

```
setenv CLASSPATH ".:$HOME/if431/DM1:/users/profs/info/chassignet/Jaxx'
```

(pour csh, tcsh) ou

```
export CLASSPATH=".:$HOME/if431/DM1:/users/profs/info/chassignet/Jaxx'
```

(pour sh, bash).

Un fichier `.java` démarre souvent comme suit :

```
package ma_lib1;
import java.util.*;
import java.io.*;
import java.awt.*;
```

Cette unité de compilation fera partie du paquetage `ma_lib1` et importe les paquetages :

- `java.util` qui contient des classes *standards* pour les tables, les piles, les tableaux de taille variable, etc.
- `java.io` qui contient les classes d'*entrées-sorties*.
- `java.awt` qui contient les classes *graphiques*.
- `java.math` qui contient les classes *mathématiques*.

On renvoie à la liste des paquetages dans la page web du cours. Si aucun paquetage précisé, l'unité de compilation fait partie du paquetage *anonyme* localisé dans le *répertoire courant*.

13.1.2 Contrôle d'accès

Pour les membres d'un paquetage, il y a deux types d'accès :

- une *classe publique* peut être accédée de l'extérieur de son paquetage ;
- une classe sans qualificatif n'est accessible que depuis son paquetage ;
- de l'extérieur du paquetage, double-protection : classe et champ doivent être publics ;
- **Remarque** : le *class loader* ne vérifie pas cette discipline, car il accède à la méthode `main` sans que sa classe ne soit publique !

- *une seule* classe publique C par unité de compilation
⇒ une seule classe publique par fichier $C.java$
(Certains compilateurs sont laxistes. Mais mieux vaut privilégier la portabilité)

13.1.3 Utilisation des `.jar` (java archive)

Un fichier `.jar` est une forme compressée des répertoires. Un `.jar` contient une arborescence de paquetages en un seul fichier (cf. la commande Unix `jar`). Les fichiers `.zip` sont aussi utilisables.

Chapitre 14

Quelques classes prédéfinies de Java

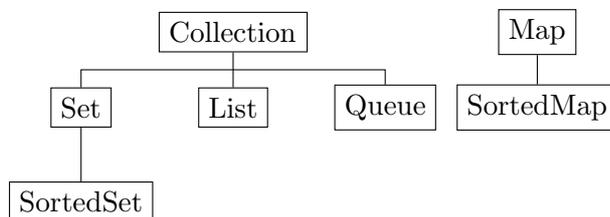
Nous présentons rapidement la classe des collections, ainsi que quelques-unes des classes que nous utiliserons au moment d'implanter les opérations et algorithmes sur les graphes.

Rappelons que le but de ces classes est de faciliter le partage et la réutilisation de code. Cela permet d'obtenir une certaine concision du code, ainsi qu'un prototypage plus rapide. Gardons cependant en tête que dans certains cas, il vaut mieux, après cela, récrire le code avec une structure de données ultra-spécifique si besoin est.

14.1 La classe Collection

En Java, la classe `Collection`¹ regroupe les structures de données les plus utilisées dans une forme standard. Elle incorpore plusieurs interfaces, associées à plusieurs implantations. De nombreux algorithmes de manipulation sont déjà implantés, ce qui facilite la réutilisation de code, donc simplifie l'écriture des programmes, et permet de les écrire dans une forme relativement standardisée.

Le cœur des collections est donné dans le diagramme suivant :



Remarquons que ces interfaces sont génériques, donc permettent de définir des Collections de type quelconque. Les collections permettent de gérer des ensembles d'éléments.

¹<http://java.sun.com/docs/books/tutorial/collections/index.html>

La classe `Set` s'intéresse aux ensembles dont les éléments ne sont pas dupliqués ; `SortedSet` stocke les éléments de manière ascendante. La `List` permet de stocker des éléments (éventuellements égaux) en ayant un certain contrôle sur l'endroit de stockage. Une `Queue` permet de modéliser une file ou une file de priorité.

Une `Map` associe une valeur à une clef. Ce sera la base des tables de hachage. `SortedMap` est la variante gardant un ordre sur les clefs.

Une collection peut être parcourue à l'aide de :

```
for(Object o : collection)
    System.out.println(o);
```

ou bien

```
for(Iterator<?> it = c.iterator(); it.hasNext(); )
    System.out.println(it.next());
```

Dans le cas où on veut pouvoir supprimer un élément, il faut utiliser l'itérateur explicite, comme dans :

```
static void filter(Collection<?> c) {
    for(Iterator<?> it = c.iterator(); it.hasNext(); )
        if(!cond(it.next()))
            it.remove();
}
```

Les autres opérations possibles sur les collections sont décrites dans la documentation de Java, ainsi que les fonctions sur les autres classes.

14.1.1 La classe `Vector`

Cette classe implante des tableaux de taille variable, c'est-à-dire que le tableau peut croître ou diminuer en fonction des accès. Il est conseillé dans certains cas d'utiliser `ArrayList` à la place de `Vector`, notamment dans le cas d'accès concurrents.

14.1.2 La classe `LinkedList`

Elle permet d'implanter les listes chaînées, les piles et les files, ainsi que les listes appelées *dequeue* (double-ended queues). Parmi les opérations intéressantes, notons `addFirst`, `removeFirst` (qui permettent de simuler une pile), ou bien de même `addLast`, `removeLast`, toutes opérations se faisant en temps constant.

14.1.3 La classe `Hashtable`

Cette classe permet d'implanter une table de hachage d'objets (non nuls), à condition que la classe des clefs implante les deux méthodes `equals` et `hashCode`. La table peut être paramétrée à la création en fonction d'une taille prévisible.

L'exemple typique d'utilisation est :

```
Hashtable nombres = new Hashtable();
nombres.put("un", new Integer(1));
nombres.put("deux", new Integer(2));
nombres.put("trois", new Integer(3));

Integer n = (Integer)nombres.get("deux");
if(n != null){
    System.out.println("deux = " + n);
}
```

La méthode `containsKey(Object key)` permet de tester si la clef existe dans la table. Toutes les opérations d'insertion, tests d'appartenance et suppression se font en $O(1)$.

14.1.4 La classe TreeSet

C'est la classe qui permet d'implanter les files de priorité sur des objets. Il faut que la classe en question soit une extension de type `Comparable` ou utilise un comparateur explicite.

Donnons un premier exemple, qui implante une classe paire d'entiers :

```
import java.io.*;
import java.util.*;

public class Paire implements Comparable{
    int i, j;

    Paire(int ii, int jj){
        i = ii; j = jj;
    }

    public int compareTo(Object o){
        Paire p = (Paire)o;
        if(i < p.i) return -1;
        if(i > p.i) return 1;
        if(j < p.j) return -1;
        if(j > p.j) return 1;
        return 0;
    }

    public boolean equals(Object o){
        return compareTo(o) == 0;
    }
}
```

```
public static void main(String[] args){
    TreeSet<Paire> t = new TreeSet<Paire>();

    for(int i = 1; i >= 0; i--){
        for(int j = 1; j >= 0; j--){
            t.add(new Paire(i, j));
        }
    }
    while(! t.isEmpty()){
        Paire p = t.first();
        t.remove(p);
        System.out.println(p.i+" "+p.j);
    }
}
```

L'exécution du programme donnera :

```
0 0
0 1
1 0
1 1
```

On aurait pu écrire également le programme suivant :

```
public static void main(String[] args){
    TreeSet<Paire> t =
        new TreeSet<Paire>(
            new Comparator<Paire>(){
                public int compare(Paire p1, Paire p2){
                    if(p1.i < p2.i) return -1;
                    if(p1.i > p2.i) return 1;
                    if(p1.j < p2.j) return -1;
                    if(p1.j > p2.j) return 1;
                    return 0;
                }
            }
        );

    for(int i = 1; i >= 0; i--){
        .....
    }
}
```

pour le même résultat.

Chapitre 15

La classe Graphe

15.1 Les choix

On fait souvent l'hypothèse que les sommets d'un graphe sont numérotés de 0 à $n - 1$, de façon à utiliser des implantations à base de matrices d'adjacence ou bien de pouvoir faire des calculs. Toutefois, de nombreux algorithmes sur les graphes sont écrits en manipulant des sommets abstraits. Il nous a paru plus proche d'une certaine réalité d'en faire autant.

Après réflexion, écrire une classe générique de graphe est une tâche quasi-impossible, car chaque algorithme qui opère sur les graphes utilisent des structures de données auxiliaires et on ne peut envisager tous les cas possibles. Néanmoins, une certaine abstraction est possible. Cela a conduit à définir une classe `Sommet` qui code un sommet, dont le nom est une chaîne, et qui contient un marqueur, qui servira par exemple à modéliser l'état d'un sommet lors d'un parcours (bfs ou dfs). Il serait facile de modifier ce type `Sommet` en fonction des applications. Pour aller au-delà, la classe `Arc` s'est imposée, un arc générique étant formé de deux sommets (origine et extrémité) et doté d'une valeur (un entier). On peut alors réaliser des listes de sommets, des piles d'arcs, etc.

On peut alors définir une classe abstraite `Graphe` qui nous permettra d'écrire tous les algorithmes du poly. Cette classe est accompagnée de deux implantations : `GrapheMatrice` (pour utiliser des matrices d'adjacence) et `GrapheListe` qui utilise des tableaux de listes de voisin. Tout a été écrit dans `Graphe` de façon à privilégier les parcours sur les voisins d'un sommet, sans savoir quelle représentation sera finalement prise pour un problème donné. Pour faciliter le traitement des sommets du graphe, on introduit une classe numérotation qui permet d'assurer un ordre sur les sommets par exemple. Pour en comprendre l'intérêt, on regardera le code de la fonction `deFichier` de la classe `GrapheListe`.

Remerciements : les classes mises au point ont bénéficié de l'aide irremplaçable de P. Chassignet. Elles ont été testées en même que le poly a été écrit. Elles sont disponibles sur la page web du cours.

15.2 grapheX comme exemple de package

Il est apparu naturel de regrouper tout cela dans un **package** au sens de Java, auquel le nom de **grapheX** a été donné. Cela permet de donner un exemple réel d'utilisation de package.

15.2.1 Comment ça marche ?

Après récupération sur le réseau¹, on exécute les commandes :

```
tar zxvf grapheX.tgz
```

Cela crée un répertoire `grapheX`, qui contient les fichiers sources :

```
Arbre.java   Graphe.java       GrapheMatrice.java  Sommet.java
Arc.java     GrapheListe.java  Numerotation.java
```

Notez la commande **package** `grapheX`; présente dans chacun des fichiers.

Au même niveau que `grapheX` se trouve un fichier de menu étendu qui montre comment utiliser les fonctions contenues dans le package (fichier `TestGraphe.java`). Ce fichier se compile par :

```
javac TestGraphe.java
```

et ira compiler tout seul dans le répertoire `grapheX` les différentes classes, les fichiers `.class` restant dans `grapheX`.

15.2.2 Faire un .jar

Si on le souhaite, on peut faire un `.jar`. Nous en profitons pour donner la syntaxe (en Unix), qui n'est pas si évidente que cela :

```
cd grapheX; javac *.java # il faut les .class!
jar -cf grapheX.jar grapheX
```

On peut alors compiler `TestGraphe.java` :

```
javac -cp grapheX.jar:. TestGraphe.java
```

Pour exécuter, il suffit de taper

```
java -cp grapheX.jar:. TestGraphe ../Data/prim1.in Prim
```

¹<http://www.enseignement.polytechnique.fr/informatique/INF431/>

15.3 Sommets

Un sommet d'un graphe sera supposé contenir comme information une chaîne de caractères. On a ajouté une marque, qui sera utilisée (à travers ses méthodes d'accès), lors des explorations, pour gérer l'état du sommet. On peut facilement afficher un sommet `s` à l'écran par

```
System.out.println(s);
```

Nous aurons besoin de piles ou de tables de hachage de sommets. Pour permettre une gestion agréable, nous avons spécialisé les méthodes `equals` et `hashCode`. Cela permettra d'écrire du code :

```
HashSet<Sommet> HS = new HashSet<Sommet>();
Sommet s = new Sommet("a", 0);
HS.put(s);
if(HS.contains(new Sommet("a", 0)))
    System.out.println(true);
```

et qu'il affiche `true`. Notons que la valeur de hachage dépendra du nom du sommet, pas de l'état du marqueur.

```
package grapheX;

/**
 * Classe de Sommets
 *
 * @author FMorain (morain@lix.polytechnique.fr)
 * @author PChassignet (chassignet@lix.polytechnique.fr)
 * @version 2006.11.27
 */
public class Sommet{
    String nom;
    private int marque;

    public Sommet(String nn, int mm){
        nom = nn;
        marque = mm;
    }

    public Sommet(Sommet s, int mm){
        nom = s.nom;
        marque = mm;
    }
}
```

```
public int valeurMarque(){
    return marque;
}

public void modifierMarque(int m){
    marque = m;
}

public boolean equals(Object o){
    return nom.equals(((Sommet)o).nom);
}

public int compareTo(Object o){
    Sommet s = (Sommet)o;
    return nom.compareTo(s.nom);
}

public String toString(){
    return ""+nom;
}

public int hashCode(){
    return nom.hashCode();
}
}
```

15.4 La classe Arc

Nous utilisons les mêmes principes que pour la classe Sommet en permettant la gestion facile de pile ou tables d'Arcs.

```
package grapheX;

import java.io.*;
import java.util.*;

/**
 * Classe d'arcs
 *
 * @author FMorain (morain@lix.polytechnique.fr)
 * @author PChassignet (chassignet@lix.polytechnique.fr)
 * @version 2006.11.27
 */
```

```
// L'arc o -> d avec valeur val
public class Arc{
    Sommet o, d;
    private int val;

    public Arc(Sommet o0, Sommet d0, int val0){
        this.o = o0;
        this.d = d0;
        this.val = val0;
    }

    public Arc(Arc a){
        this.o = a.o;
        this.d = a.d;
        this.val = a.val;
    }

    public int valeur(){
        return val;
    }

    public void modifierValeur(int vv){
        this.val = vv;
    }

    public boolean equals(Object aa){
        Arc a = (Arc)aa;
        return o.equals(a.o) && d.equals(a.d) && (val == a.val);
    }

    public String toString(){
        return "("+this.o+", "+this.d+")";
    }

    public int hashCode(){
        String str = "+"+this;
        return str.hashCode();
    }
}
```

Une remarque sur la méthode `hashCode`. On veut qu'un arc soit repérable par ses deux sommets, pas par sa valeur. On pourrait changer ce comportement si besoin est.

15.5 La classe abstraite Graphe et deux implantations

Le but de la classe `Graphe` est de permettre de manipuler facilement des graphes, comme on peut le lire dans les chapitres du poly concernant les graphes. Notons en passant que nous utiliserons systématiquement des itérateurs sur les sommets d'un graphe ou les voisins d'un sommet.

Le parti pris est celui d'une collection d'arcs, donc d'un graphe *a priori* orienté. Dans de rares cas, l'utilisation d'une classe plus spécifique d'arête au lieu d'arc pourrait être envisagée.

15.5.1 Définitions

Les quelques méthodes décrites suffisent à implanter tous les algorithmes décrits dans le cours.

```

package grapheX;

import java.io.*;
import java.util.*;

/**
 * Classe abstraite de graphes
 *
 * @author FMorain (morain@lix.polytechnique.fr)
 * @version 2006.11.27
 */
public abstract class Graphe{

    public abstract int taille();
    public abstract Graphe copie();

    public abstract void ajouterSommet(Sommet s);
    public abstract boolean existeArc(Sommet s, Sommet t);
    public abstract void ajouterArc(Sommet s, Sommet t, int val);
    public abstract int valeurArc(Sommet s, Sommet t);
    public abstract void enleverArc(Sommet s, Sommet t);

    public abstract Collection<Sommet> sommets();

```

15.5.2 Numérotation

Cette classe est utilisée dans les deux implantations de la classe abstraite, car nous avons souvent besoin (en interne) d'un ordre sur les sommets.

Il est important, en interne, d'avoir un moyen de numéroter les sommets lors de leur création. C'est le rôle de cette classe, qui sera utilisée dans les deux classes `GrapheMatrice` et `GrapheListe`. Pour que l'utilisateur ne soit pas tenté d'utiliser cette numérotation, tous les champs sont privés. On peut récupérer des itérateurs sur les sommets, au moyen de la méthode `elements()`.

```
package grapheX;

import java.io.*;
import java.util.*;

/**
 * Numérotation des graphes
 *
 * @author FMorain (morain@lix.polytechnique.fr)
 * @author PChassignet (chassignet@lix.polytechnique.fr)
 * @version 2006.11.22
 */
public class Numerotation{
    private int compteur;
    private Hashtable<Sommet,Integer> HSI;
    private Vector<Sommet> VS;

    public Numerotation(int n){
        compteur = -1;
        HSI = new Hashtable<Sommet,Integer>();
        VS = new Vector<Sommet>(n);
        VS.setSize(n);
    }

    public int taille(){
        return VS.size();
    }

    public boolean ajouterElement(Sommet s){
        if(!HSI.containsKey(s)){
            compteur++;
            HSI.put(s, compteur);
            VS.set(compteur, s);
            return true;
        }
        return false;
    }
}
```

```

    public int numero(Sommet s){
        return HSI.get(s);
    }

    public Sommet elementAt(int i){
        return VS.elementAt(i);
    }

    public Collection<Sommet> elements(){
        return VS;
    }
}

```

15.5.3 Implantation par matrice

C'est là une implantation proche des matrices d'adjacence, même si on utilise des vecteurs de vecteurs.

```

package grapheX;

import java.io.*;
import java.util.*;

/**
 * Graphes implantés dans des "matrices"
 *
 * @author FMorain (morain@lix.polytechnique.fr)
 * @version 2007.01.12
 */
public class GrapheMatrice extends Graphe{
    private Vector<Vector<Arc>> M;
    private Numerotation numerotation;

    public int taille(){
        return M.size();
    }

    public GrapheMatrice(int n){
        numerotation = new Numerotation(n);
        M = new Vector<Vector<Arc>>(n);
        M.setSize(n);
    }
}

```

```
public void ajouterSommet(Sommet s){
    if(numerotation.ajouterElement(s)){
        int n = taille();
        Vector<Arc> vs = new Vector<Arc>(n);
        vs.setSize(n);
        M.set(numerotation.numero(s), vs);
    }
}

public boolean existeArc(Sommet s, Sommet t){
    int si = numerotation.numero(s);
    int ti = numerotation.numero(t);
    return M.get(si).get(ti) != null;
}

private boolean existeArc(int i, int j){
    return M.get(i).get(j) != null;
}

public void ajouterArc(Sommet s, Sommet t, int val){
    ajouterSommet(s);
    ajouterSommet(t);
    int si = numerotation.numero(s);
    int ti = numerotation.numero(t);
    M.get(si).set(ti, new Arc(s, t, val));
}

public int valeurArc(Sommet s, Sommet t){
    int si = numerotation.numero(s);
    int ti = numerotation.numero(t);
    return M.get(si).get(ti).valeur();
}

private int valeurArc(int i, int j){
    return M.get(i).get(j).valeur();
}

public void enleverArc(Sommet s, Sommet t){
    int si = numerotation.numero(s);
    int ti = numerotation.numero(t);
    M.get(si).remove(ti);
}
```

```
public void modifierValeur(Sommet s, Sommet t, int val){
    int si = numerotation.numero(s);
    int ti = numerotation.numero(t);
    M.get(si).get(ti).modifierValeur(val);
}

public LinkedList<Arc> voisins(Sommet s){
    LinkedList<Arc> l = new LinkedList<Arc>();
    int si = numerotation.numero(s);

    for(int j = 0; j < taille(); j++)
        if(existeArc(si, j))
            l.addLast(M.get(si).get(j));
    return l;
}

public Collection<Sommet> sommets(){
    return numerotation.elements();
}

public GrapheMatrice copie(){
    int n = taille();
    GrapheMatrice G = new GrapheMatrice(n);
    for(int i = 0; i < n; i++)
        G.ajouterSommet(numerotation.elementAt(i));
    for(int i = 0; i < n; i++)
        for(int j = 0; j < n; j++)
            if(M.get(i).get(j) != null)
                G.ajouterArc(numerotation.elementAt(i),
                             numerotation.elementAt(j),
                             valeurArc(i, j));
    return G;
}

public static GrapheMatrice deMatrice(int[][] M){
    int n = M.length;
    GrapheMatrice G = new GrapheMatrice(n);

    for(int i = 0; i < n; i++)
        G.ajouterSommet(new Sommet(i+"", 0));
    for(int i = 0; i < n; i++){
        for(int j = 0; j < M[i].length; j++)
            if(M[i][j] == 1)
                G.ajouterArc(G.numerotation.elementAt(i),
```

```
        G.numerotation.elementAt(j),
        1);
    }
    return G;
}
}
```

15.5.4 Implantation avec des listes

```
package grapheX;

import java.io.*;
import java.util.*;

/**
 * Graphes implantés dans des "listes"
 *
 * @author FMorain (morain@lix.polytechnique.fr)
 * @version 2006.11.27
 */

public class GrapheListe extends Graphe{
    private Vector<LinkedList<Arc>> L;
    private Numerotation numerotation;

    public int taille(){
        return L.size();
    }

    public GrapheListe(int n){
        numerotation = new Numerotation(n);
        L = new Vector<LinkedList<Arc>>(n);
        L.setSize(n);
    }

    public void ajouterSommet(Sommet s){
        if(numerotation.ajouterElement(s))
            L.set(numerotation.numero(s), new LinkedList<Arc>());
    }

    public boolean existeArc(Sommet s, Sommet t){
        for(Arc a : L.get(numerotation.numero(s)))
```

```
        if((a.d).equals(t))
            return true;
        return false;
    }

    private boolean existeArc(int i, int j){
        Sommet t = numerotation.elementAt(j);
        for(Arc a : L.get(i))
            if(a.d.equals(t))
                return true;
        return false;
    }

    public void ajouterArc(Sommet s, Sommet t, int val){
        ajouterSommet(s);
        ajouterSommet(t);
        int si = numerotation.numero(s);
        L.get(si).addLast(new Arc(s, t, val));
    }

    public void ajouterArc(int i, int j, int val){
        L.get(i).addLast(new Arc(numerotation.elementAt(i),
                                numerotation.elementAt(j),
                                val));
    }

    public int valeurArc(Sommet s, Sommet t){
        for(Arc a : L.get(numerotation.numero(s)))
            if(a.d.equals(t))
                return a.valeur();
        return -1; // convention
    }

    public int valeurArc(int i, int j){
        Sommet t = numerotation.elementAt(j);
        for(Arc a : L.get(i))
            if(a.d.equals(t))
                return a.valeur();
        return -1; // convention
    }

    public void enleverArc(Sommet s, Sommet t){
        int si = numerotation.numero(s);
        Arc a = null;
```

```
        for(Arc aa : L.get(numerotation.numero(s)))
            if(aa.d.equals(t)){
                a = aa;
                break;
            }
        if(a != null)
            L.get(numerotation.numero(s)).remove(a);
    }

    public void modifierValeur(Sommet s, Sommet t, int val){
        for(Arc a : L.get(numerotation.numero(s)))
            if(a.d.equals(t)){
                a.modifierValeur(val);
                return;
            }
    }

    public LinkedList<Arc> voisins(Sommet s){
        return L.get(numerotation.numero(s));
    }

    public Collection<Sommet> sommets(){
        return numerotation.elements();
    }

    public GrapheListe copie(){
        int n = taille();
        GrapheListe G = new GrapheListe(n);
        for(int i = 0; i < n; i++)
            G.ajouterSommet(numerotation.elementAt(i));
        for(int i = 0; i < n; i++){
            // recopie dans le même ordre
            LinkedList<Arc> Li = G.L.get(i);
            for(Arc a : L.get(i))
                Li.addLast(a);
        }
        return G;
    }

    // retourne vrai si le caractère c est dans str
    private static boolean option(String str, char c){
        for(int i = 0; i < str.length(); i++)
            if(str.charAt(i) == c)
                return true;
    }
}
```

```

    return false;
}

public static GrapheListe deFichier(String nomfic){
    try{
        Scanner scan =
            new Scanner(
                new BufferedReader(new FileReader(nomfic)));
        System.out.println(scan.next());
        int n = scan.nextInt();
        GrapheListe G = new GrapheListe(n);
        String str = scan.next();
        boolean estValue = option(str, 'v');
        boolean estSym = option(str, 's');
        boolean avecCouples = option(str, 'c');

        System.out.println("n = "+n);
        for(int i = 0; i < n; i++){
            Sommet s = new Sommet(scan.next(), 0);
            G.ajouterSommet(s);
        }
        if(avecCouples){
            System.out.println("Avec couples");
            // on lit des lignes "i j dij" ou "i j"
            while(scan.hasNext()){
                Sommet s = new Sommet(scan.next(), 0);
                Sommet t = new Sommet(scan.next(), 0);
                int si = G.numerotation.numero(s);
                int ti = G.numerotation.numero(t);
                if(estValue)
                    G.ajouterArc(si, ti, (int)scan.nextInt());
                else
                    G.ajouterArc(si, ti, 1);
            }
        }
        else{
            // format "s ns t0 t1 ... t{ns-1}"
            // ou      "s ns t0 v0 t1 v1 ... t{ns-1} v{ns-1}"
            System.out.println("Avec listes, estvalue="+estValue);
            for(int r = 0; r < n; r++){
                Sommet s = new Sommet(scan.next(), 0);
                int si = G.numerotation.numero(s);
                int nj = (int)scan.nextInt();
                for(int k = 0; k < nj; k++){

```

```

        Sommet t = new Sommet(scan.next(), 0);
        int ti = G.numerotation.numero(t);
        if(estValue)
            G.ajouterArc(si, ti,
                        (int)scan.nextInt());
        else
            G.ajouterArc(si, ti, 1);
    }
}
System.out.println("G="+G);
if(estSym)
    // on doit symétriser G
    for(Sommet s : G.sommets())
        for(Sommet t : G.sommets())
            if(G.existeArc(s, t)
                && !G.existeArc(t, s))
                G.ajouterArc(s, t,
                            G.valeurArc(s, t));

    return G;
}
catch(Exception e) { }
return null;
}
}

```

La méthode `deFichier` donne un exemple d'utilisation des différentes primitives. Donnons un exemple de fichier d'entrée. Le fichier :

```

#prim1
4 vs
u v w x
u 2 v 16 x 29
v 3 u 16 x 20 w 25
w 2 v 25 x 10
x 3 u 29 v 20 w 10

```

La première ligne est un commentaire qui n'est pas traité par le programme. La seconde contient le nombre de sommets du graphe. Suivent des codes comme `v` pour graphe valué, `s` pour graphe symétrique (non orienté). La ligne suivante contient les noms des 4 sommets du graphe. Ils seront stockés dans cet ordre dans une table, ce qui fait que

```

for(Sommet s : sommets())
    System.out.println(s);

```

affichera

u
v
w
x

Suivent alors 4 lignes. Chacune commence par un nom de sommet, puis par le nombre de voisins k , puis k paires nom de sommet suivi de la valeur de l'arc. Ce fichier correspond au graphe de la figure 8.6.

Un graphe non valué est stocké sous une forme plus simple :

```
#pda1
6 s
a b c d e f
a 4 b d c f
b 2 a d
c 3 a e f
d 2 a b
e 2 a c
f 1 c
```

C'est le graphe représenté à la figure 6.5.

Bibliographie

- [1] A. V. Aho, R. Sethi, and J. D. Ullman. *Compilers : Principles, Techniques, and Tools*. Addison–Wesley, 1986. A été traduit en français.
- [2] J. M. Aldous and R. J. Wilson. *Graphs and applications – an introductory approach*. Springer, 2000.
- [3] D. Beauquier, J. Berstel, and P. Chrétienne. *Éléments d’algorithmique*. Masson, Paris, 1992.
- [4] R. Cori and J.-J. Lévy. *Algorithmes et Programmation*. Cours de l’École polytechnique, 1998.
- [5] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms*. MIT Press/McGraw-Hill, 1990.
- [6] A. Gibbons. *Algorithmic Graph Theory*. Cambridge University Press, 1988.
- [7] D. Kozen. *Automata and Computability*. Springer, 1997. <http://www.cs.cornell.edu/~kozen/>.
- [8] J.-J. Lévy. *Informatique fondamentale*. Cours de l’École polytechnique, 2004.
- [9] E. M. Reingold, J. Nievergelt, and N. Deo. *Combinatorial algorithms – theory and practice*. Prentice-Hall, 1977.