



**HAL**  
open science

# Bayesian Networks and Particle Filters for Joint Adaptive Equalization and Decoding

Pascal Cheung-Mon-Chan

► **To cite this version:**

Pascal Cheung-Mon-Chan. Bayesian Networks and Particle Filters for Joint Adaptive Equalization and Decoding. domain\_other. Télécom ParisTech, 2003. English. NNT: . pastel-00000732

**HAL Id: pastel-00000732**

**<https://pastel.hal.science/pastel-00000732>**

Submitted on 6 Sep 2004

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Thèse

présentée par

**Pascal Cheung-Mon-Chan**

pour obtenir le grade de docteur  
de l'École Nationale Supérieure des Télécommunications

Spécialité : Signal et Images

---

## **Réseaux bayésiens et filtres particuliers pour l'égalisation adaptative et le décodage conjoints**

---

Soutenue le 18 décembre 2003 devant le jury composé de

Cédric Demeure	
François Desbouvries	Rapporteur
Eric Moulines	
Jean-Yves Tourneret	Rapporteur
Sandrine Vaton	



*À ma famille.*



# Remerciements

*Je tiens tout d'abord à exprimer ma sincère et profonde reconnaissance à Eric Moulines, Professeur à l'Ecole Nationale Supérieure des Télécommunications, pour la grande confiance qu'il a toujours manifestée à mon égard au cours de cette thèse. J'ai beaucoup apprécié sa grande rigueur et sa remarquable culture scientifique et technique. Je n'oublierai jamais l'aide extrêmement précieuse qu'il m'a apportée pendant et surtout après la réalisation de ce travail.*

*Je suis reconnaissant envers Thalès Communications pour avoir bien voulu co-financer ce travail. Je remercie vivement Messieurs Cédric Demeure, Chef du service Traitement du Signal et des Images, et Dominique Mérel, Chef du laboratoire Radio Communications, pour m'avoir fourni les moyens dont j'avais besoin pour mener à bien mon travail. Je salue Didier Pirez, qui était chargé du suivi de cette thèse pour Thalès Communications. J'ai beaucoup apprécié la grande autonomie qu'il m'a accordée au cours de ce travail. Parmi tous les ingénieurs du service TSI, je tiens à remercier tout particulièrement Nicolas Marguinaud, Dominique Ragot et Cédric Dietzi pour leur aide sincère et désintéressée. Je remercie également Béatrice Col, Gwenaél Guilmin, Frédéric Lafaye, François Sirven, Yves-Marie Quémeneur et Christine Uhl pour leur attitude particulièrement amicale et bienveillante, grâce à laquelle cette thèse a pu être réalisée dans une atmosphère de confiance et de respect mutuel, propice à l'échange d'idées.*

*Je souhaiterais ensuite remercier toutes les personnes qui ont bien voulu prendre part au Jury de cette thèse. Je remercie très vivement Messieurs François Desbouvries, Professeur à l'Institut National des Télécommunications, et Jean-Yves Tourneret, Professeur à l'Institut National Polytechnique de Toulouse, pour leurs conseils et pour l'intérêt qu'ils ont porté à ma thèse en acceptant la charge de Rapporteur. Je remercie également Sandrine Vaton, Maître de Conférence à l'Ecole Nationale Supérieure des Télécommunications de Bretagne, pour sa lecture particulièrement minutieuse de mon manuscrit et pour ses remarques extrêmement pertinentes.*

*Je tiens à saluer tous les membres permanents et les thésards du département Traitement du Signal et des Images de l'ENST, et tout particulièrement Maurice Charbit, Olivier Cappé, Lisa Perros-Meilhac, Gersende Fort, Anahid Safavi, Emmanuele Grosicki, Serguey Burykh et Laurence Zelmar, grâce à qui mon séjour au département TSI de l'ENST a été pour moi une expérience agréable et enrichissante.*

*Enfin, je voudrais remercier ma famille, en particulier mon frère Jean-Yves, qui m'a toujours encouragé et soutenu.*



# Sommaire

<b>Introduction</b>	<b>9</b>
Etat de l'art . . . . .	9
Orientation et contenu de la thèse . . . . .	10
Plan de la thèse . . . . .	10
<b>1 L'indépendance conditionnelle</b>	<b>15</b>
1.1 Introduction . . . . .	15
1.2 L'indépendance conditionnelle entre sous-tribus . . . . .	15
1.3 L'indépendance conditionnelle entre sous-ensembles d'éléments aléatoires . . . . .	20
1.4 Conclusion . . . . .	21
<b>2 Les réseaux bayésiens</b>	<b>23</b>
2.1 Introduction . . . . .	23
2.2 Définition d'un réseau bayésien . . . . .	23
2.3 Construction d'un réseau bayésien . . . . .	27
2.4 Conclusion . . . . .	30
<b>3 L'algorithme de propagation de croyance</b>	<b>31</b>
3.1 Introduction . . . . .	31
3.2 Rappels et notations . . . . .	31
3.3 L'algorithme de propagation de croyance exacte . . . . .	33
3.4 L'algorithme de propagation de croyance approchée . . . . .	45
3.5 Conclusion . . . . .	53
<b>4 Application des réseaux bayésiens à l'égalisation</b>	<b>55</b>
4.1 Introduction . . . . .	55
4.2 L'égalisation d'un canal sélectif en fréquence . . . . .	55
4.3 L'égalisation adaptative d'un canal évanescent sélectif en fréquence . . . . .	60
4.4 L'égalisation adaptative d'un canal évanescent non-sélectif en fréquence . . . . .	68
4.5 Conclusion . . . . .	73
<b>5 L'algorithme de filtrage particulière à échantillonnage global</b>	<b>75</b>
5.1 Introduction . . . . .	75
5.2 L'algorithme de filtrage particulière à échantillonnage global . . . . .	76
5.3 Convergence de l'erreur quadratique moyenne . . . . .	79
5.4 Variantes de l'algorithme de filtrage particulière à échantillonnage global . . . . .	82



5.5	Conclusion . . . . .	84
<b>6</b>	<b>Application aux modèles dynamiques à processus indicateur latent</b>	<b>85</b>
6.1	Introduction . . . . .	85
6.2	Les modèles dynamiques à processus indicateur latent . . . . .	85
6.3	Le filtre particulaire à échantillonnage global immédiat . . . . .	90
6.4	Le filtre particulaire à échantillonnage global différé . . . . .	93
6.5	Conclusion . . . . .	99
<b>7</b>	<b>Application du filtrage particulaire à l'égalisation</b>	<b>101</b>
7.1	Introduction . . . . .	101
7.2	L'égalisation d'un canal sélectif en fréquence . . . . .	101
7.3	L'égalisation adaptative d'un canal évanescent sélectif en fréquence . . . . .	104
7.4	L'égalisation adaptative d'un canal évanescent non-sélectif en fréquence . . . . .	106
7.5	Simulations numériques . . . . .	108
7.6	Conclusion . . . . .	114
<b>8</b>	<b>L'égalisation adaptative et le décodage conjoints</b>	<b>117</b>
8.1	Introduction . . . . .	117
8.2	Modèle du système . . . . .	117
8.3	Réseau bayésien . . . . .	119
8.4	Application de l'algorithme de propagation de croyance approchée . . . . .	121
8.5	Principe de l'algorithme de turbo-filtrage particulaire . . . . .	125
8.6	Application de l'algorithme de turbo-filtrage particulaire . . . . .	127
8.7	Conclusion . . . . .	135
	<b>Conclusion</b>	<b>137</b>
<b>A</b>	<b>Quelques notions fondamentales de la théorie des graphes</b>	<b>139</b>
<b>B</b>	<b>Démonstration de la proposition 2.1</b>	<b>141</b>
<b>C</b>	<b>Résultats utiles</b>	<b>147</b>
<b>D</b>	<b>L'algorithme de propagation de croyance conditionnelle</b>	<b>153</b>
	<b>Bibliographie</b>	<b>163</b>

# Introduction

Dans un système de communication numérique, les canaux de transmission utilisés par un émetteur pour transmettre un message à un récepteur introduisent souvent des distorsions qui peuvent provoquer des erreurs de transmission.

L'objectif de cette thèse est de contribuer au développement de récepteurs plus robustes à ces distorsions en proposant de nouveaux algorithmes d'estimation du message transmis qui soient capables d'exploiter aussi efficacement que possible l'ensemble de l'information *a priori* dont dispose le récepteur : code de canal, entrelacement, modulation, type de distorsions introduites par le canal et présence de séquences de référence dans le message émis.

## Etat de l'art

Afin de comprendre l'originalité de l'approche que nous avons suivie, nous allons tout d'abord examiner brièvement la façon dont ces différentes informations sont traditionnellement prises en compte dans la conception des récepteurs actuels.

Jusqu'à la fin des années 1980, la plupart des récepteurs prenaient en compte ces différentes informations à l'aide de plusieurs étapes successives [BDI86]. Les distorsions du canal étaient d'abord corrigées par un égaliseur (le plus souvent linéaire) dont on adaptait les caractéristiques au canal à l'aide des séquences de référence. Ensuite, des décisions (éventuellement pondérées) étaient prises à partir du signal égalisé. Enfin, après désentrelacement, ces décisions étaient transmises à un décodeur qui reconstituait le message transmis. Aucune information n'était transmise par le décodeur à l'égaliseur dans cette approche – qui est d'ailleurs toujours utilisée aujourd'hui dans certains récepteurs fabriqués par Thalès Communications.

A la fin des années 1980, Eyuboglu [Eyu88] a proposé d'améliorer l'égalisation à l'aide des décisions prises par le décodeur et Zhou *et al.* [ZPL90] ont proposé d'effectuer l'égalisation et le décodage à l'aide de plusieurs itérations. Cependant, les performances des récepteurs obtenus restaient très sous-optimales.

En 1993, Berrou *et al.* [BGT93] ont proposé un nouveau type de code particulièrement performant appelé *turbo-code*, et dont la règle de décodage itérative, appelée *turbo-décodage*, a été reprise en 1995 par Douillard *et al.* [DJB<sup>+</sup>95] pour effectuer conjointement l'égalisation et le décodage à l'aide d'un processus itératif appelé *turbo-égalisation*. Depuis 1995, plusieurs variantes du récepteur présenté dans [DJB<sup>+</sup>95] ont été publiées. Ces variantes diffèrent entre elles essentiellement par le type d'égaliseur ou de décodeur utilisé, le but étant d'appliquer le principe de la turbo-égalisation à des canaux ou des codes différents de ceux initialement étudiés dans [DJB<sup>+</sup>95], ou d'obtenir un compromis différent entre la complexité et la performance du récepteur. Une comparaison plus détaillée de ces différentes variantes se trouve dans [TKS02].

## Orientation et contenu de la thèse

Après la publication de [BGT93] en 1993, de nombreux travaux ont été réalisés pour analyser l'algorithme de turbo-décodage. Ainsi, en 1995, Wiberg *et al.* [WLK95, Wib96] ont montré que l'algorithme de turbo-décodage peut être considéré comme une application particulière d'un algorithme plus général, qui opère sur des graphes de Tanner [Tan81], et que Wiberg *et al.* ont appelé algorithme *somme-produit*. En 1998, McELiece *et al.* [MMC98], ainsi que Frey *et al.* [FM98], ont montré que l'algorithme de turbo-décodage peut également être considéré comme une application particulière de l'algorithme de *propagation de croyance*, qui a été inventé par Pearl [Pea82, Pea86, Pea88] dans le domaine de l'intelligence artificielle, et qui opère en propageant des *messages* entre les sommets d'un *réseau bayésien*. Kschischang *et al.* [KF98] ont alors introduit les *graphes factoriels*, qui sont une généralisation des graphes de Tanner, et ont montré que l'algorithme de propagation de croyance de Pearl est équivalent à l'algorithme somme-produit appliqué au graphe factoriel correspondant à un réseau bayésien donné.

L'approche que nous avons suivie découle directement de ces différents travaux. Nous avons utilisé les réseaux bayésiens<sup>1</sup> pour représenter l'ensemble de l'information *a priori* dont dispose le récepteur et nous avons appliqué l'algorithme de propagation de croyance au réseau bayésien obtenu pour estimer les bits de données transmis. En outre, pour réduire la complexité de l'algorithme obtenu, nous avons utilisé un filtre particulière pour calculer de façon approchée certains des messages de l'algorithme de propagation de croyance.

Comparée à l'approche traditionnelle qui est fondée sur l'algorithme de turbo-décodage, notre approche représente un changement radical de paradigme. Plutôt que de considérer que l'on a une succession de traitements distincts (l'estimation du canal, l'égalisation, le désentrelacement et le décodage) que l'on essaie de combiner avec plus ou moins de succès, on part d'un seul et unique problème (l'estimation des bits de données transmis) que l'on résout en appliquant l'algorithme de propagation de croyance, éventuellement à l'aide d'approximations maîtrisées, à un réseau bayésien qui représente l'intégralité de l'information *a priori* dont dispose le récepteur.

Les avantages de cette approche unifiée sont multiples. Tout d'abord, cette approche est systématique : tous les calculs effectués correspondent à l'application de l'algorithme de propagation de croyance ; aucune justification subtile n'est nécessaire pour prouver leur validité. Ensuite, cette approche permet d'effectuer de façon méthodique les compromis entre performance et complexité : on part d'un algorithme optimal (ou quasiment optimal) et on en réduit progressivement la complexité en remplaçant certains calculs par des calculs approchés. Enfin, à long terme, cette approche nous permettra de bénéficier directement des travaux réalisés afin d'analyser l'algorithme de propagation de croyance et de l'améliorer [YFW01, YFW].

## Plan de la thèse

Le premier chapitre est consacré à la notion d'indépendance conditionnelle. Cette notion est l'une des notions centrales sur laquelle repose cette thèse : les relations d'indépendance

---

<sup>1</sup>Nous avons préféré utiliser les réseaux bayésiens plutôt que les graphes factoriels car les réseaux bayésiens sont plus compacts que les graphes factoriels et sont très faciles à comprendre intuitivement. Cependant, les deux formalismes sont équivalents et nous aurions obtenu les mêmes résultats avec des graphes factoriels.

conditionnelles permettent de simplifier le calcul de certaines lois conditionnelles et les différents réseaux bayésiens que nous serons amenés à construire auront pour fonction première de représenter les relations d'indépendance conditionnelle existant dans le système considéré. La définition utilisée par [Pea88] n'étant valable que lorsque certaines hypothèses sont vérifiées, nous sommes partis de la définition de l'indépendance conditionnelle entre sous-tribus [Nev70] et nous en étudions les propriétés dans un cadre très général.

Le deuxième chapitre est consacré à la définition et à la construction des réseaux bayésiens. Nous rappelons la définition de la  $d$ -séparation entre sommets donnée par [Pea88] et nous montrons qu'il est possible de représenter les relations d'indépendance conditionnelle telle que nous l'avons définie au premier chapitre à l'aide des relations de  $d$ -séparation. Nous sommes ainsi en mesure de généraliser à des éléments aléatoires quelconques la notion de réseau bayésien introduite par [Pea88].

Le troisième chapitre décrit l'algorithme de propagation de croyance. Nous définissons le concept de message envoyé par un sommet à un autre en utilisant la notion de transition entre espaces mesurables et nous donnons les équations de propagation de croyance en utilisant le théorème de Bayes généralisé (cf § C.1).

Le quatrième chapitre montre l'application des réseaux bayésiens au problème de l'égalisation. Nous nous intéressons aux trois situations suivantes : égalisation d'un canal sélectif en fréquence de réponse impulsionnelle connue, égalisation adaptative d'un canal évanescent sélectif en fréquence et égalisation adaptative d'un canal évanescent non-sélectif en fréquence. Nous construisons le réseau bayésien correspondant à chacune de ces situations et nous appliquons l'algorithme de propagation de croyance au réseau bayésien obtenu. Dans le cas de l'égalisation d'un canal sélectif en fréquence de réponse impulsionnelle connue, nous retrouvons l'algorithme BCJR [BCJR74], dont la complexité est exponentielle par rapport à la longueur de la réponse impulsionnelle du canal. Dans les deux autres cas, nous obtenons un algorithme dont la complexité est exponentielle par rapport au temps.

Le cinquième chapitre présente l'algorithme de filtrage particulaire à échantillonnage global. Ce nouvel algorithme permet de calculer de façon approchée une suite de mesures de probabilité définies sur une suite d'espaces mesurables discrets dont la taille croît de façon exponentielle. Nous démontrons que l'erreur quadratique moyenne de l'algorithme converge vers 0 lorsque le nombre de particules tend vers l'infini. Nous présentons ensuite brièvement les différentes variantes possibles de l'algorithme en fonction de la stratégie d'échantillonnage utilisée.

Le sixième chapitre montre l'application de l'algorithme de filtrage particulaire à échantillonnage global aux modèles dynamiques à processus indicateur latent. Nous définissons les notions d'opérateurs de prédiction de l'état, de prédiction de l'observation et de correction de l'état du modèle. Selon la loi *a posteriori* que l'on souhaite approcher, l'application de l'algorithme de filtrage particulaire à échantillonnage global à un modèle dynamique à processus indicateur latent permet d'obtenir deux algorithmes d'estimation : l'algorithme de filtrage particulaire à échantillonnage global immédiat et l'algorithme de filtrage particulaire à échantillonnage global différé. Nous décrivons en détail ces deux algorithmes à l'aide des opérateurs de prédiction et de correction définis précédemment.

Le septième chapitre montre l'application de l'algorithme de filtrage particulaire à échantillonnage global au problème de l'égalisation. Comme précédemment, nous nous intéressons à nouveau aux trois situations suivantes : égalisation d'un canal sélectif en fréquence de réponse impulsionnelle connue, égalisation adaptative d'un canal évanescent sélectif en fréquence et

égalisation adaptative d'un canal évanescent non-sélectif en fréquence. Nous constatons alors que le réseau bayésien correspondant à chacune de ces situations possède une structure de modèle dynamique à processus indicateur latent. Nous calculons ensuite explicitement les opérateurs de prédiction et de correction correspondant à chacun de ces modèles dynamiques à processus indicateur latent. Nous pouvons alors déterminer la complexité des algorithmes d'égalisation que l'on obtient en appliquant les algorithmes de filtrage particulaire à chacune de ces situations. Dans le cas de l'égalisation d'un canal sélectif en fréquence de réponse impulsionnelle connue, les algorithmes obtenus ont une complexité linéaire par rapport à la longueur de la réponse impulsionnelle du canal. Dans tous les cas, les algorithmes obtenus ont une complexité linéaire par rapport au temps. Enfin, nous étudions numériquement les performances de l'algorithme de filtrage particulaire à échantillonnage global appliqué à l'égalisation adaptative d'un canal évanescent non-sélectif en fréquence. Pour la situation considérée et en fonction du nombre de particules utilisées, nous constatons que les performances de l'algorithme proposé sont supérieures ou égales à celles de l'algorithme décrit dans [CWL00] avec les réglages préconisés dans cet article.

Le huitième chapitre montre l'application de l'ensemble des techniques étudiées précédemment au problème de l'égalisation adaptative et du décodage conjoints. A titre d'exemple, nous considérons dans cette partie de la thèse le cas d'un canal évanescent non-sélectif en fréquence. Il va sans dire que l'algorithme présenté dans ce chapitre est aussi applicable aux autres canaux étudiés précédemment. Nous commençons par construire le réseau bayésien correspondant à ce problème et nous appliquons l'algorithme de propagation de croyance à ce réseau. En utilisant les résultats du chapitre 4, il apparaît immédiatement que le calcul de certains messages exigerait un nombre d'opérations augmentant exponentiellement avec le temps. Nous remplaçons donc le calcul exact de ces messages par un calcul approché qui est effectué en appliquant l'algorithme de filtrage particulaire étudié précédemment à un réseau bayésien virtuel possédant une structure de modèle dynamique à processus indicateur latent. Nous obtenons ainsi un algorithme itératif, combinant filtrage particulaire et propagation de croyance, que nous avons appelé *algorithme de turbo-filtrage particulaire*.

Nous avons représenté les relations d'interdépendance entre les principaux domaines abordés au cours de cette thèse à l'aide du diagramme de la figure 1.

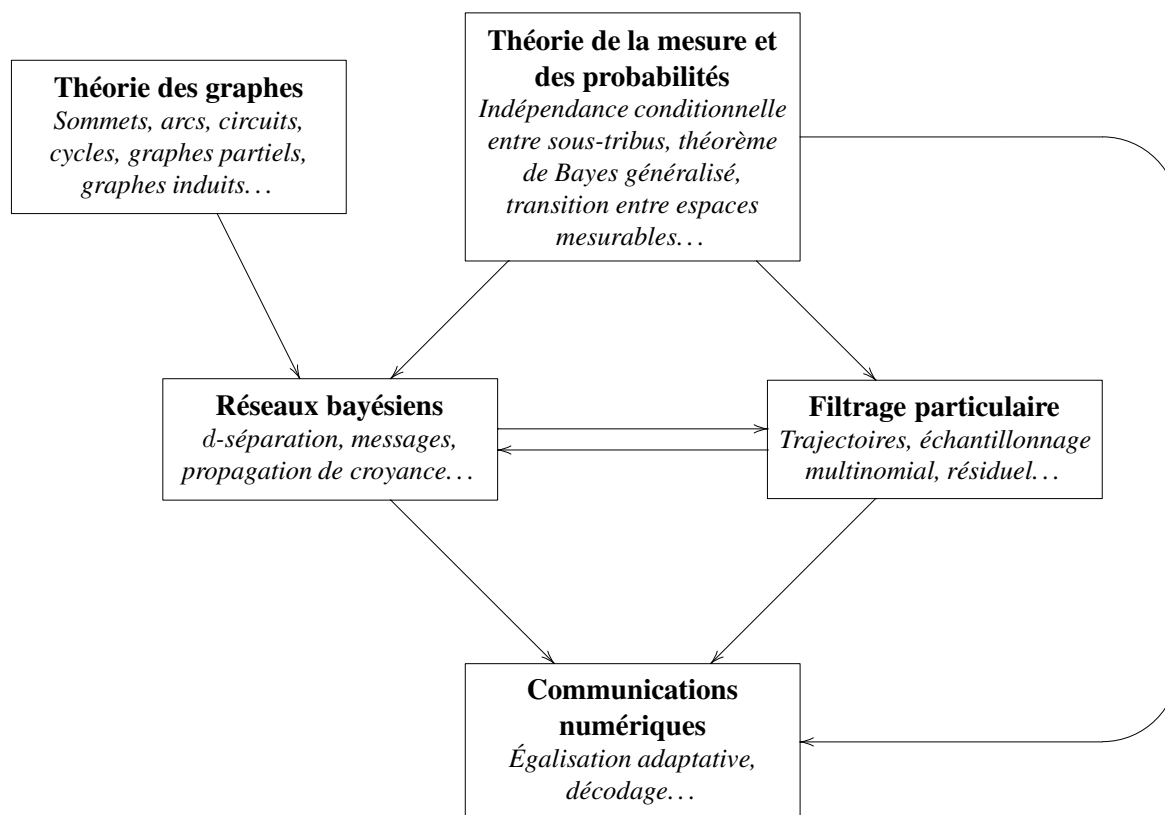


FIG. 1 – Interdépendance des domaines abordés dans ce document.

Ce diagramme montre que pour construire et utiliser les réseaux bayésiens, nous avons dû faire appel à des concepts issus à la fois de la théorie des graphes et de la théorie de la mesure. D'autre part, nous avons défini puis étudié l'algorithme de filtrage particulaire à échantillonnage global en utilisant des concepts et des résultats issus de la théorie de la mesure et des probabilités. Nous avons également utilisé les réseaux bayésiens pour formuler plus facilement l'algorithme de filtrage particulaire à échantillonnage global (cf la figure 5.2) et pour mettre en évidence les relations d'indépendance conditionnelles utilisées lors des calculs effectués pendant la mise en œuvre de cet algorithme (cf les §§ 6.3.2 et 6.4.2). Réciproquement, le filtrage particulaire nous a permis de calculer de façon approximative certains messages dont le calcul exact aurait exigé un nombre d'opérations trop élevé. Enfin, nous avons appliqué ces différentes techniques pour résoudre le problème de l'égalisation adaptative et du décodage conjoints.



# Chapitre 1

## L'indépendance conditionnelle

### 1.1 Introduction

La notion d'indépendance conditionnelle est l'une des notions centrales sur laquelle repose cette thèse. En effet, les relations d'indépendance conditionnelles permettent de simplifier le calcul de certaines lois conditionnelles et les différents réseaux bayésiens que nous serons amenés à construire auront pour fonction première de représenter les relations d'indépendance conditionnelle existant dans le système considéré. Le but de ce chapitre est de définir la notion d'indépendance conditionnelle et d'en étudier les propriétés.

La définition utilisée par [Pea88] n'étant valable que lorsque certaines hypothèses sont vérifiées, nous partirons de la définition de l'indépendance conditionnelle entre sous-tribus [Nev70] qui présente le grand avantage d'être valable dans tout espace de probabilité. Nous démontrons ensuite une propriété permettant de caractériser l'indépendance de deux sous-tribus conditionnellement à une autre sous-tribu. C'est notamment grâce à cette propriété caractéristique que les équations de propagation de croyance seront démontrées au chapitre 3. Nous établirons ensuite plusieurs autres propriétés vérifiées par la relation ternaire d'indépendance conditionnelle entre sous-tribus. Enfin, nous introduirons la notion d'indépendance conditionnelle entre ensembles d'éléments aléatoires et nous transporterons par homomorphisme les propriétés établies précédemment. Nous obtiendrons ainsi un ensemble de propriétés qui seront utilisées au chapitre 2 pour la construction des réseaux bayésiens.

### 1.2 L'indépendance conditionnelle entre sous-tribus

#### Définition 1.1

Soient  $(\Omega, \mathcal{F}, P)$  un espace de probabilité et  $\mathcal{G}, \mathcal{G}_1, \dots, \mathcal{G}_n$  des sous-tribus de  $\mathcal{F}$ . On dit que les sous-tribus  $\mathcal{G}_1, \dots, \mathcal{G}_n$  sont  $P$ -indépendantes conditionnellement à la sous-tribu  $\mathcal{G}$  si et seulement si on a, pour toutes variables aléatoires positives  $X_1, \dots, X_n$  respectivement  $\mathcal{G}_1$ -mesurable,  $\dots$ ,  $\mathcal{G}_n$ -mesurable,

$$\mathbb{E}^P(X_1 \cdots X_n | \mathcal{G}) = \mathbb{E}^P(X_1 | \mathcal{G}) \cdots \mathbb{E}^P(X_n | \mathcal{G}) \quad (1.1)$$

Dans le cas particulier où  $n = 2$ , on note  $I_P(\mathcal{G}_1, \mathcal{G}, \mathcal{G}_2)$  la  $P$ -indépendance des sous-tribus  $\mathcal{G}_1$  et  $\mathcal{G}_2$  conditionnellement à la sous-tribu  $\mathcal{G}$ .



**Remarque 1.1**

Dans la suite, lorsqu'il n'y a pas d'ambiguïté, on parlera simplement d'indépendance conditionnelle au lieu de  $P$ -indépendance conditionnelle.

**Proposition 1.1**

Soient  $(\Omega, \mathcal{F}, P)$  un espace de probabilité et  $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$  des sous-tribus de  $\mathcal{F}$ . Les sous-tribus  $\mathcal{G}_1$  et  $\mathcal{G}_2$  sont  $P$ -indépendantes conditionnellement à la sous-tribu  $\mathcal{G}_3$  si et seulement si on a, pour toute variable aléatoire positive  $X_1$   $\mathcal{G}_1$ -mesurable

$$\mathbb{E}^P(X_1 | \mathcal{G}_3 \vee \mathcal{G}_2) = \mathbb{E}^P(X_1 | \mathcal{G}_3) \quad (1.2)$$

où  $\mathcal{G}_3 \vee \mathcal{G}_2$  désigne la tribu engendrée par  $\mathcal{G}_3 \cup \mathcal{G}_2$ .

DÉMONSTRATION :

Supposons que les sous-tribus  $\mathcal{G}_1$  et  $\mathcal{G}_2$  sont  $P$ -indépendantes conditionnellement à la tribu  $\mathcal{G}_3$ . Soit alors une variable aléatoire positive  $X_1$   $\mathcal{G}_1$ -mesurable. Comme  $\mathcal{G}_3$  est inclus dans  $\mathcal{G}_3 \vee \mathcal{G}_2$ , la variable aléatoire positive  $\mathbb{E}^P(X_1 | \mathcal{G}_3)$ , qui est  $\mathcal{G}_3$ -mesurable, est donc également  $(\mathcal{G}_3 \vee \mathcal{G}_2)$ -mesurable. Afin de prouver l'équation 1.2, il nous suffit donc de montrer que pour tout  $A \in \mathcal{G}_3 \vee \mathcal{G}_2$ , on a

$$\mathbb{E}^P(\mathbb{1}_A \mathbb{E}^P(X_1 | \mathcal{G}_3)) = \mathbb{E}^P(\mathbb{1}_A X_1) \quad (1.3)$$

Pour cela, on remarque tout d'abord que  $\mathcal{G}_3 \vee \mathcal{G}_2$  est la tribu engendrée par l'ensemble  $\Gamma$  des éléments de  $\mathcal{F}$  constitués par l'intersection des éléments de  $\mathcal{G}_3$  et  $\mathcal{G}_2$ <sup>1</sup> i.e.  $\mathcal{G}_3 \vee \mathcal{G}_2 = \sigma(\Gamma)$  avec

$$\Gamma = \{B \in \mathcal{F} | B = A_3 \cap A_2, A_3 \in \mathcal{G}_3, A_2 \in \mathcal{G}_2\} \quad (1.4)$$

De plus, en utilisant la stabilité de  $\mathcal{G}_3$  et  $\mathcal{G}_2$  par intersection dénombrable (et donc finie), on voit facilement que  $\Gamma$  est stable par intersection finie. Par conséquent, le  $d$ -système<sup>2</sup> engendré par  $\Gamma$  est la tribu engendrée par  $\Gamma$ , à savoir  $\mathcal{G}_3 \vee \mathcal{G}_2$ . Autrement dit, on a

$$d(\Gamma) = \sigma(\Gamma) \quad (1.5)$$

$$= \mathcal{G}_3 \vee \mathcal{G}_2 \quad (1.6)$$

Considérons maintenant l'ensemble  $\mathcal{C}$  des éléments de  $d(\Gamma)$  vérifiant l'équation 1.3 i.e.

$$\mathcal{C} = \{B \in d(\Gamma) \mid \mathbb{E}^P(\mathbb{1}_B \mathbb{E}^P(X_1 | \mathcal{G}_3)) = \mathbb{E}^P(\mathbb{1}_B X_1)\} \quad (1.7)$$

Par construction,  $\mathcal{C}$  est inclus dans  $d(\Gamma)$ . Nous allons montrer que l'on a également  $d(\Gamma) \subset \mathcal{C}$  :

1. Montrons que  $\Gamma$  est inclus dans  $\mathcal{C}$ . Soit  $B$  un élément de  $\Gamma$ . Par définition de  $\Gamma$ ,  $B$  est de la forme  $B = A_3 \cap A_2$  avec  $A_3 \in \mathcal{G}_3$  et  $A_2 \in \mathcal{G}_2$ , et  $B$  vérifie donc

$$\mathbb{E}^P(\mathbb{1}_B \mathbb{E}^P(X_1 | \mathcal{G}_3)) = \mathbb{E}^P(\mathbb{1}_{A_2} \mathbb{1}_{A_3} \mathbb{E}^P(X_1 | \mathcal{G}_3)) \quad (1.8)$$

$$= \mathbb{E}^P(\mathbb{E}^P(\mathbb{1}_{A_2} \mathbb{1}_{A_3} \mathbb{E}^P(X_1 | \mathcal{G}_3) | \mathcal{G}_3)) \quad (1.9)$$

La variable aléatoire  $\mathbb{1}_{A_3} \mathbb{E}^P(X_1 | \mathcal{G}_3)$  étant  $\mathcal{G}_3$ -mesurable, on en déduit

$$\mathbb{E}^P(\mathbb{1}_B \mathbb{E}^P(X_1 | \mathcal{G}_3)) = \mathbb{E}^P(\mathbb{1}_{A_3} \mathbb{E}^P(X_1 | \mathcal{G}_3) \mathbb{E}^P(\mathbb{1}_{A_2} | \mathcal{G}_3)) \quad (1.10)$$

<sup>1</sup>En effet, comme  $\Omega$  appartient à  $\mathcal{G}_3$  et  $\mathcal{G}_2$ , on voit facilement que les sous-tribus  $\mathcal{G}_3$  et  $\mathcal{G}_2$  sont incluses dans  $\Gamma$  et donc dans  $\sigma(\Gamma)$ , ce qui implique  $\mathcal{G}_3 \vee \mathcal{G}_2 \subset \sigma(\Gamma)$ . De plus, comme  $\mathcal{G}_3 \vee \mathcal{G}_2$  est stable par intersection et contient  $\mathcal{G}_3$  et  $\mathcal{G}_2$ , l'ensemble  $\Gamma$  est inclus dans  $\mathcal{G}_3 \vee \mathcal{G}_2$ , ce qui implique  $\sigma(\Gamma) \subset \mathcal{G}_3 \vee \mathcal{G}_2$ .

<sup>2</sup>i.e. un ensemble de parties de  $\Omega$  contenant l'élément  $\Omega$  et stable par différence et par limite croissante. Certains auteurs appellent également un  $d$ -système un système de Dynkin.

Les sous-tribus  $\mathcal{G}_1$  et  $\mathcal{G}_2$  étant, par hypothèse,  $P$ -indépendantes conditionnellement à la tribu  $\mathcal{G}_3$ , on en déduit

$$\mathbb{E}^P(\mathbf{1}_B \mathbb{E}^P(X_1 | \mathcal{G}_3)) = \mathbb{E}^P(\mathbf{1}_{A_3} \mathbb{E}^P(X_1 \mathbf{1}_{A_2} | \mathcal{G}_3)) \quad (1.11)$$

La variable aléatoire  $\mathbf{1}_{A_3}$  étant  $\mathcal{G}_3$ -mesurable, on en déduit

$$\mathbb{E}^P(\mathbf{1}_B \mathbb{E}^P(X_1 | \mathcal{G}_3)) = \mathbb{E}^P(\mathbb{E}^P(\mathbf{1}_{A_3} X_1 \mathbf{1}_{A_2} | \mathcal{G}_3)) \quad (1.12)$$

$$= \mathbb{E}^P(\mathbf{1}_{A_3} X_1 \mathbf{1}_{A_2}) \quad (1.13)$$

$$= \mathbb{E}^P(\mathbf{1}_B X_1) \quad (1.14)$$

ce qui montre que  $B$  appartient à  $\mathcal{C}$  qui contient donc  $\Gamma$ .

2. Montrons que  $\mathcal{C}$  est un  $d$ -système :

(a) comme  $\Omega$  appartient à  $\Gamma$  qui est inclus dans  $\mathcal{C}$ ,  $\Omega$  appartient à  $\mathcal{C}$ .

(b) soit  $A$  et  $B$  des éléments de  $\mathcal{C}$  tels que  $A \subset B$ . Soit  $C = B \setminus A$ . L'élément  $C$  vérifie

$$\mathbb{E}^P(\mathbf{1}_C \mathbb{E}^P(X_1 | \mathcal{G}_3)) = \mathbb{E}^P((\mathbf{1}_B - \mathbf{1}_A) \mathbb{E}^P(X_1 | \mathcal{G}_3)) \quad (1.15)$$

$$= \mathbb{E}^P(\mathbf{1}_B \mathbb{E}^P(X_1 | \mathcal{G}_3)) - \mathbb{E}^P(\mathbf{1}_A \mathbb{E}^P(X_1 | \mathcal{G}_3)) \quad (1.16)$$

$$= \mathbb{E}^P(\mathbf{1}_B X_1) - \mathbb{E}^P(\mathbf{1}_A X_1) \quad (1.17)$$

$$= \mathbb{E}^P((\mathbf{1}_B - \mathbf{1}_A) X_1) \quad (1.18)$$

$$= \mathbb{E}^P(\mathbf{1}_C X_1) \quad (1.19)$$

ce qui montre que  $C$  appartient à  $\mathcal{C}$  qui est donc stable par différence.

(c) soit  $(A_n)$  une suite croissante d'éléments de  $\mathcal{C}$ . Soit  $A = \bigcup_n A_n$ . Les variables aléatoires  $\mathbf{1}_{A_n} \mathbb{E}^P(X_1 | \mathcal{G}_3)$  étant positives et vérifiant  $\mathbf{1}_{A_n} \mathbb{E}^P(X_1 | \mathcal{G}_3) \uparrow \mathbf{1}_A \mathbb{E}^P(X_1 | \mathcal{G}_3)$ , on a

$$\mathbb{E}^P(\mathbf{1}_A \mathbb{E}^P(X_1 | \mathcal{G}_3)) = \lim_n \mathbb{E}^P(\mathbf{1}_{A_n} \mathbb{E}^P(X_1 | \mathcal{G}_3)) \quad (1.20)$$

Comme les  $A_n$  sont des éléments de  $\mathcal{C}$ , on en déduit

$$\mathbb{E}^P(\mathbf{1}_A \mathbb{E}^P(X_1 | \mathcal{G}_3)) = \lim_n \mathbb{E}^P(\mathbf{1}_{A_n} X_1) \quad (1.21)$$

Les variables aléatoires  $\mathbf{1}_{A_n} X_1$  étant positives et vérifiant  $\mathbf{1}_{A_n} X_1 \uparrow \mathbf{1}_A X_1$ , on en déduit

$$\mathbb{E}^P(\mathbf{1}_A \mathbb{E}^P(X_1 | \mathcal{G}_3)) = \mathbb{E}^P(\mathbf{1}_A X_1) \quad (1.22)$$

ce qui montre que  $A$  appartient à  $\mathcal{C}$  qui est donc stable par limite croissante.

Comme  $\mathcal{C}$  est un  $d$ -système contenant  $\Gamma$ , on a donc  $d(\Gamma) \subset \mathcal{C}$  et donc, comme on avait déjà  $\mathcal{C} \subset d(\Gamma)$ , on a

$$\mathcal{C} = d(\Gamma) \quad (1.23)$$

$$= \mathcal{G}_3 \vee \mathcal{G}_2 \quad (1.24)$$

ce qui montre que l'équation 1.3 est vérifiée par tous les éléments de  $\mathcal{G}_3 \vee \mathcal{G}_2$ . On en déduit immédiatement que pour toute variable aléatoire  $X_1$   $\mathcal{G}_1$ -mesurable positive, on a

$$\mathbb{E}^P(X_1 | \mathcal{G}_3 \vee \mathcal{G}_2) = \mathbb{E}^P(X_1 | \mathcal{G}_3) \quad (1.25)$$

Réciproquement, supposons que les sous-tribus  $\mathcal{G}_1$ ,  $\mathcal{G}_2$  et  $\mathcal{G}_3$  sont telles que, pour toute variable aléatoire  $X_1$   $\mathcal{G}_1$ -mesurable positive, on a

$$\mathbb{E}^P(X_1|\mathcal{G}_3 \vee \mathcal{G}_2) = \mathbb{E}^P(X_1|\mathcal{G}_3) \quad (1.26)$$

Montrons alors que toutes variables aléatoires positives  $X_1$  et  $X_2$  respectivement  $\mathcal{G}_1$ -mesurable et  $\mathcal{G}_2$ -mesurable vérifient

$$\mathbb{E}^P(X_1 X_2|\mathcal{G}_3) = \mathbb{E}^P(X_1|\mathcal{G}_3)\mathbb{E}^P(X_2|\mathcal{G}_3) \quad (1.27)$$

Par définition de l'espérance conditionnelle d'une variable aléatoire positive, on a pour tout événement  $A_3 \in \mathcal{G}_3$

$$\mathbb{E}^P(\mathbb{1}_{A_3}\mathbb{E}^P(X_1 X_2|\mathcal{G}_3)) = \mathbb{E}^P(\mathbb{1}_{A_3} X_1 X_2) \quad (1.28)$$

$$= \mathbb{E}^P(\mathbb{E}^P(\mathbb{1}_{A_3} X_1 X_2|\mathcal{G}_3 \vee \mathcal{G}_2)) \quad (1.29)$$

La variable aléatoire  $\mathbb{1}_{A_3} X_2$  étant  $(\mathcal{G}_3 \vee \mathcal{G}_2)$ -mesurable, on en déduit

$$\mathbb{E}^P(\mathbb{1}_{A_3}\mathbb{E}^P(X_1 X_2|\mathcal{G}_3)) = \mathbb{E}^P(\mathbb{1}_{A_3} X_2 \mathbb{E}^P(X_1|\mathcal{G}_3 \vee \mathcal{G}_2)) \quad (1.30)$$

En utilisant l'équation 1.26, on en déduit

$$\mathbb{E}^P(\mathbb{1}_{A_3}\mathbb{E}^P(X_1 X_2|\mathcal{G}_3)) = \mathbb{E}^P(\mathbb{1}_{A_3} X_2 \mathbb{E}^P(X_1|\mathcal{G}_3)) \quad (1.31)$$

$$= \mathbb{E}^P(\mathbb{E}^P(\mathbb{1}_{A_3} X_2 \mathbb{E}^P(X_1|\mathcal{G}_3)|\mathcal{G}_3)) \quad (1.32)$$

La variable aléatoire  $\mathbb{1}_{A_3}\mathbb{E}^P(X_1|\mathcal{G}_3)$  étant  $\mathcal{G}_3$ -mesurable, on en déduit

$$\mathbb{E}^P(\mathbb{1}_{A_3}\mathbb{E}^P(X_1 X_2|\mathcal{G}_3)) = \mathbb{E}^P(\mathbb{1}_{A_3}\mathbb{E}^P(X_1|\mathcal{G}_3)\mathbb{E}^P(X_2|\mathcal{G}_3)) \quad (1.33)$$

On en déduit

$$\mathbb{E}^P(X_1 X_2|\mathcal{G}_3) = \mathbb{E}^P(X_1|\mathcal{G}_3)\mathbb{E}^P(X_2|\mathcal{G}_3) \quad (1.34)$$

Les sous-tribus  $\mathcal{G}_1$  et  $\mathcal{G}_2$  sont donc  $P$ -indépendantes conditionnellement à la tribu  $\mathcal{G}_3$ .  $\square$

### Proposition 1.2

Soit  $(\Omega, \mathcal{F}, P)$  un espace de probabilité. La relation ternaire  $I_P$  de  $P$ -indépendance conditionnelle entre sous-tribus de  $\mathcal{F}$ , que l'on note  $\mathcal{S}(\mathcal{F})$ , vérifie les propriétés suivantes :

#### 1. Symétrie :

$$\forall (\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3) \in \mathcal{S}(\mathcal{F})^3 \quad I_P(\mathcal{G}_1, \mathcal{G}_3, \mathcal{G}_2) \Rightarrow I_P(\mathcal{G}_2, \mathcal{G}_3, \mathcal{G}_1) \quad (1.35)$$

#### 2. Décomposition :

$$\forall (\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \mathcal{G}_4) \in \mathcal{S}(\mathcal{F})^4 \quad I_P(\mathcal{G}_1, \mathcal{G}_3, \mathcal{G}_2 \vee \mathcal{G}_4) \Rightarrow I_P(\mathcal{G}_1, \mathcal{G}_3, \mathcal{G}_2) \wedge I_P(\mathcal{G}_1, \mathcal{G}_3, \mathcal{G}_4) \quad (1.36)$$

#### 3. Union faible :

$$\forall (\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \mathcal{G}_4) \in \mathcal{S}(\mathcal{F})^4 \quad I_P(\mathcal{G}_1, \mathcal{G}_3, \mathcal{G}_2 \vee \mathcal{G}_4) \Rightarrow I_P(\mathcal{G}_1, \mathcal{G}_3 \vee \mathcal{G}_2, \mathcal{G}_4) \quad (1.37)$$

#### 4. Contraction :

$$\forall (\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \mathcal{G}_4) \in \mathcal{S}(\mathcal{F})^4 \quad I_P(\mathcal{G}_1, \mathcal{G}_3 \vee \mathcal{G}_2, \mathcal{G}_4) \wedge I_P(\mathcal{G}_1, \mathcal{G}_3, \mathcal{G}_2) \Rightarrow I_P(\mathcal{G}_1, \mathcal{G}_3, \mathcal{G}_2 \vee \mathcal{G}_4) \quad (1.38)$$

DÉMONSTRATION :

##### 1. Symétrie :

Cette propriété est évidente lorsque l'on considère l'équation 1.1 qui intervient dans la définition 1.1 de la  $P$ -indépendance conditionnelle.

##### 2. Décomposition :

Supposons que l'on a  $I_P(\mathcal{G}_1, \mathcal{G}_3, \mathcal{G}_2 \vee \mathcal{G}_4)$ . Considérons alors des variables aléatoires positives  $X_1$  et  $X_2$  respectivement  $\mathcal{G}_1$ -mesurable et  $\mathcal{G}_2$ -mesurable. Comme la sous-tribu  $\mathcal{G}_2$  est incluse dans la sous-tribu  $\mathcal{G}_2 \vee \mathcal{G}_4$ , la variable aléatoire  $X_2$  est également  $(\mathcal{G}_2 \vee \mathcal{G}_4)$ -mesurable et vérifie donc

$$\mathbb{E}^P(X_1 X_2 | \mathcal{G}_3) = \mathbb{E}^P(X_1 | \mathcal{G}_3) \mathbb{E}^P(X_2 | \mathcal{G}_3) \quad (1.39)$$

ce qui montre que l'on a  $I_P(\mathcal{G}_1, \mathcal{G}_3, \mathcal{G}_2)$ .

##### 3. Union faible :

Supposons que l'on a  $I_P(\mathcal{G}_1, \mathcal{G}_3, \mathcal{G}_2 \vee \mathcal{G}_4)$ . Considérons alors une variable aléatoire positive  $X_1$   $\mathcal{G}_1$ -mesurable. D'après la proposition 1.1, la variable aléatoire  $\mathbb{E}^P(X_1 | \mathcal{G}_3 \vee \mathcal{G}_2 \vee \mathcal{G}_4)$  vérifie<sup>3</sup>

$$\mathbb{E}^P(X_1 | \mathcal{G}_3 \vee \mathcal{G}_2 \vee \mathcal{G}_4) = \mathbb{E}^P(X_1 | \mathcal{G}_3) \quad (1.40)$$

Par conséquent, la variable aléatoire  $\mathbb{E}^P(X_1 | \mathcal{G}_3 \vee \mathcal{G}_2 \vee \mathcal{G}_4)$  est  $\mathcal{G}_3$ -mesurable et donc, *a fortiori*,  $(\mathcal{G}_3 \vee \mathcal{G}_2)$ -mesurable. De plus elle vérifie pour tout événement  $A$  de  $\mathcal{G}_3 \vee \mathcal{G}_2$  ( qui est inclus dans  $\mathcal{G}_3 \vee \mathcal{G}_2 \vee \mathcal{G}_4$  )

$$\mathbb{E}^P(\mathbf{1}_A \mathbb{E}^P(X_1 | \mathcal{G}_3 \vee \mathcal{G}_2 \vee \mathcal{G}_4)) = \mathbb{E}^P(\mathbf{1}_A X_1) \quad (1.41)$$

On en déduit

$$\mathbb{E}^P(X_1 | \mathcal{G}_3 \vee \mathcal{G}_2 \vee \mathcal{G}_4) = \mathbb{E}^P(X_1 | \mathcal{G}_3 \vee \mathcal{G}_2) \quad (1.42)$$

ce qui prouve, d'après la proposition 1.1, que l'on a  $I_P(\mathcal{G}_1, \mathcal{G}_3 \vee \mathcal{G}_2, \mathcal{G}_4)$ .

##### 4. Contraction :

Supposons que l'on a  $I_P(\mathcal{G}_1, \mathcal{G}_3 \vee \mathcal{G}_2, \mathcal{G}_4)$  et  $I_P(\mathcal{G}_1, \mathcal{G}_3, \mathcal{G}_2)$ . Considérons alors une variable aléatoire positive  $X_1$   $\mathcal{G}_1$ -mesurable. D'après la proposition 1.1, la variable aléatoire  $\mathbb{E}^P(X_1 | \mathcal{G}_3 \vee \mathcal{G}_2 \vee \mathcal{G}_4)$  vérifie

$$\mathbb{E}^P(X_1 | \mathcal{G}_3 \vee \mathcal{G}_2 \vee \mathcal{G}_4) = \mathbb{E}^P(X_1 | \mathcal{G}_3 \vee \mathcal{G}_2) \quad (1.43)$$

ainsi que

$$\mathbb{E}^P(X_1 | \mathcal{G}_3 \vee \mathcal{G}_2) = \mathbb{E}^P(X_1 | \mathcal{G}_3) \quad (1.44)$$

ce qui implique

$$\mathbb{E}^P(X_1 | \mathcal{G}_3 \vee \mathcal{G}_2 \vee \mathcal{G}_4) = \mathbb{E}^P(X_1 | \mathcal{G}_3) \quad (1.45)$$

ce qui prouve, d'après la proposition 1.1, que l'on a  $I_P(\mathcal{G}_1, \mathcal{G}_3, \mathcal{G}_2 \vee \mathcal{G}_4)$ .

□

---

<sup>3</sup>On notera que la relation  $\vee$  est associative, ce qui rend légitime la notation  $\mathcal{G}_3 \vee \mathcal{G}_2 \vee \mathcal{G}_4$ .

### 1.3 L'indépendance conditionnelle entre sous-ensembles d'éléments aléatoires

#### Définition 1.2

Soient  $(\Omega, \mathcal{F}, P)$  un espace de probabilité et  $U$  un ensemble fini d'éléments aléatoires de  $(\Omega, \mathcal{F})$  dans des espaces mesurables quelconques. Soient  $X, X_1, \dots, X_n$  des sous-ensembles de  $U$ . On dit que les sous-ensembles  $X_1, \dots, X_n$  sont  $P$ -indépendants conditionnellement au sous-ensemble  $X$  si et seulement si les sous-tribus  $\sigma(X_1), \dots, \sigma(X_n)$  sont  $P$ -indépendantes conditionnellement à la tribu  $\sigma(X)$ , avec la convention suivante :

$$\sigma(\emptyset) = \mathcal{F}_* \quad (1.46)$$

$$\forall Y \in \mathcal{P}(U) \setminus \{\emptyset\} \quad \sigma(Y) = \bigvee_{Y \in Y} \sigma(Y) \quad (1.47)$$

où  $\mathcal{F}_* = \{\emptyset, \Omega\}$  est la sous-tribu triviale de  $\mathcal{F}$ ,  $\sigma(Y)$  la sous-tribu de  $\mathcal{F}$  engendrée par l'élément aléatoire  $Y$  et  $\mathcal{P}(U)$  l'ensemble des parties de  $U$ . Dans le cas particulier où  $n = 2$ , on note  $I_P(X_1, X, X_2)$  la  $P$ -indépendance des sous-ensembles  $X_1$  et  $X_2$  conditionnellement au sous-ensemble  $X$ .

#### Proposition 1.3

Soient  $(\Omega, \mathcal{F}, P)$  un espace de probabilité et  $U$  un ensemble fini d'éléments aléatoires de  $(\Omega, \mathcal{F})$  dans des espaces mesurables quelconques. La relation ternaire  $I_P$  de  $P$ -indépendance conditionnelle entre parties de  $U$ , que l'on note  $\mathcal{P}(U)$ , vérifie les propriétés suivantes :

##### 1. Symétrie :

$$\forall (X_1, X_2, X_3) \in \mathcal{P}(U)^3 \quad I_P(X_1, X_3, X_2) \Rightarrow I_P(X_2, X_3, X_1) \quad (1.48)$$

##### 2. Décomposition :

$$\forall (X_1, X_2, X_3, X_4) \in \mathcal{P}(U)^4 \quad I_P(X_1, X_3, X_2 \cup X_4) \Rightarrow I_P(X_1, X_3, X_2) \wedge I_P(X_1, X_3, X_4) \quad (1.49)$$

##### 3. Union faible :

$$\forall (X_1, X_2, X_3, X_4) \in \mathcal{P}(U)^4 \quad I_P(X_1, X_3, X_2 \cup X_4) \Rightarrow I_P(X_1, X_3 \cup X_2, X_4) \quad (1.50)$$

##### 4. Contraction :

$$\forall (X_1, X_2, X_3, X_4) \in \mathcal{P}(U)^4 \quad I_P(X_1, X_3 \cup X_2, X_4) \wedge I_P(X_1, X_3, X_2) \Rightarrow I_P(X_1, X_3, X_2 \cup X_4) \quad (1.51)$$

DÉMONSTRATION :

Il suffit de remarquer que l'on a la relation suivante

$$\forall (X, Y) \in \mathcal{P}(U)^2 \quad \sigma(X \cup Y) = \sigma(X) \vee \sigma(Y) \quad (1.52)$$

et d'appliquer la proposition 1.2. □

## 1.4 Conclusion

Dans ce chapitre, nous avons défini la notion d'indépendance conditionnelle entre sous-tribus et nous avons démontré une propriété permettant de caractériser l'indépendance de deux sous-tribus conditionnellement à une autre sous-tribu. Nous avons également défini la notion d'indépendance conditionnelle entre ensembles d'éléments aléatoires et nous avons montré que la relation ternaire d'indépendance entre ensembles d'éléments aléatoires vérifie les propriétés de symétrie, de décomposition, d'union faible et de contraction.

Ces dernières propriétés avaient déjà été énoncées par [Pea88] dans le cas particulier de variables aléatoires discrètes, mais on notera qu'ici, à la différence de [Pea88], nous sommes parvenus à démontrer ces propriétés pour des éléments aléatoires *absolument quelconques*. En particulier, nous n'avons même pas eu besoin de supposer que la loi conjointe des éléments aléatoires considérés possède une densité par rapport à une quelconque mesure de domination ou que les variables aléatoires considérées prennent leurs valeurs dans des espaces séparables. Les différentes définitions et propriétés présentées dans ce premier chapitre sont donc très importantes car elles nous permettront dans la suite d'utiliser en toute rigueur les outils conçus initialement par [Pea88] dans un cadre beaucoup plus restreint que le nôtre. Ainsi, dans le prochain chapitre, nous utiliserons la définition de l'indépendance conditionnelle que nous venons de voir pour définir la notion de réseau bayésien dans un cadre très général, et nous utiliserons les propriétés que nous venons de démontrer pour construire les réseaux bayésiens dans ce cadre très général.



# Chapitre 2

## Les réseaux bayésiens

### 2.1 Introduction

Dans le chapitre précédent, nous avons défini la notion d'indépendance conditionnelle et nous en avons étudié les propriétés. Afin d'exploiter aisément ces propriétés, il peut être très utile de représenter graphiquement les relations d'indépendance conditionnelle existant dans un système. Aussi, dans ce chapitre, nous allons définir une famille de graphes appelés *réseaux bayésiens* qui permettent de représenter les relations d'indépendance conditionnelle et nous allons donner un moyen de construire ces graphes.

Ce chapitre faisant appel à de nombreuses notions issues de la théorie des graphes, nous conseillons au lecteur qui ne connaîtrait pas cette théorie de lire au préalable l'annexe A où nous avons rassemblé les concepts issus de la théorie des graphes utilisés dans ce document, ou, mieux encore, de consulter les premiers chapitres de [GM95].

### 2.2 Définition d'un réseau bayésien

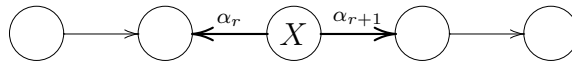
Le but de cette partie est de définir la notion de réseau bayésien. Pour cela, nous allons tout d'abord introduire la notion de *d-séparation* à l'aide des définitions 2.1, 2.2 et 2.3. Nous introduirons ensuite les notions de *modèle de dépendance* (définition 2.4) et de graphe constituant un *I-map minimal* pour un modèle de dépendance donné (définition 2.5). Enfin nous définirons un *réseau bayésien* comme étant un graphe constituant un *I-map minimal* pour le modèle de dépendance qui correspond à la relation d'indépendance conditionnelle définie au chapitre précédent (définition 2.6).

#### Définition 2.1

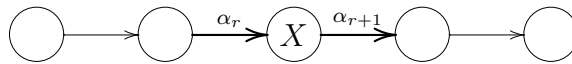
Soit  $G = (U, A)$  un graphe orienté sans circuit. Soit  $X$  un sommet du graphe orienté  $G$ . Soit  $(\alpha_1, \dots, \alpha_q)$  une chaîne du graphe orienté  $G$ .

On dit que la chaîne  $(\alpha_1, \dots, \alpha_q)$  **diverge** au sommet  $X$  si et seulement si  $q > 1$  et s'il existe  $r \in \{1, \dots, q-1\}$  tel que  $X$  est l'extrémité initiale des arcs  $\alpha_r$  et  $\alpha_{r+1}$ .

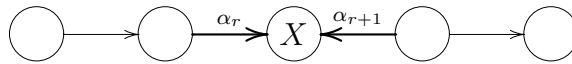


FIG. 2.1 – Exemple de chaîne divergeant en un sommet  $X$ .

On dit que la chaîne  $(\alpha_1, \dots, \alpha_q)$  est **en série** au sommet  $X$  si et seulement si  $q > 1$  et s'il existe  $r \in \{1, \dots, q - 1\}$  tel que  $X$  est soit l'extrémité terminale de  $\alpha_r$  et l'extrémité initiale de  $\alpha_{r+1}$ , soit l'extrémité initiale de  $\alpha_r$  et l'extrémité terminale de  $\alpha_{r+1}$ .

FIG. 2.2 – Exemple de chaîne en série en un sommet  $X$ .

On dit que la chaîne  $(\alpha_1, \dots, \alpha_q)$  **converge** au sommet  $X$  si et seulement si  $q > 1$  et s'il existe  $r \in \{1, \dots, q - 1\}$  tel que  $X$  est l'extrémité terminale des arcs  $\alpha_r$  et  $\alpha_{r+1}$ .

FIG. 2.3 – Exemple de chaîne convergeant en un sommet  $X$ .

### Définition 2.2

Soit  $G = (U, A)$  un graphe orienté sans circuit. Soit  $c$  une chaîne du graphe orienté  $G$ . Soit  $Z$  un sous-ensemble de  $U$ . On dit que la chaîne  $c$  est **bloquée** (ou **d-séparée**) par le sous-ensemble  $Z$  si et seulement si la chaîne  $c$  vérifie au moins l'une des conditions suivantes

1. la chaîne  $c$  diverge en un sommet de  $Z$
2. la chaîne  $c$  est en série en un sommet de  $Z$
3. la chaîne  $c$  converge en un sommet qui n'appartient pas à  $Z$  et dont aucun descendant propre dans le graphe  $G$  n'appartient à  $Z$ .

Dans le cas contraire, on dit que la chaîne  $c$  est **active** par rapport au sous-ensemble  $Z$ .

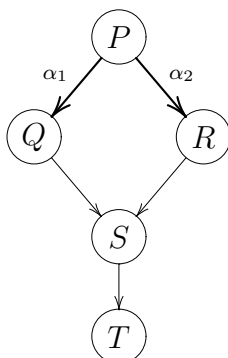


FIG. 2.4 – Exemple de chaîne bloquée par un ensemble. La chaîne  $(\alpha_1, \alpha_2)$  est bloquée par le sous-ensemble  $Z = \{P\}$ . En effet, la chaîne  $(\alpha_1, \alpha_2)$  diverge au sommet  $P$  qui appartient au sous-ensemble  $Z = \{P\}$ .

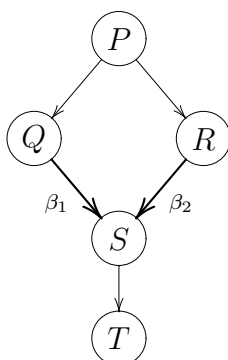


FIG. 2.5 – Autre exemple de chaîne bloquée par un ensemble. La chaîne  $(\beta_1, \beta_2)$  est bloquée par le sous-ensemble  $Z = \{P\}$ . En effet, la chaîne  $(\beta_1, \beta_2)$  converge au sommet  $S$  qui n'appartient pas au sous-ensemble  $Z = \{P\}$  et dont aucun descendant propre n'appartient au sous-ensemble  $Z = \{P\}$ .

### Définition 2.3

Soit  $G = (U, A)$  un graphe orienté sans circuit. Soient  $X_1, X_2, X_3$  des sous-ensembles de  $U$ . On dit que les sous-ensembles  $X_1$  et  $X_2$  sont **d-séparés** par le sous-ensemble  $X_3$  dans le graphe  $G$  si et seulement si les sous-ensembles  $X_1, X_2, X_3$  sont deux à deux disjoints et toutes les chaînes du graphe  $G$  joignant un sommet de  $X_1$  à un sommet de  $X_2$  sont bloquées par le sous-ensemble  $X_3$ . On note  $\langle X_1 | X_3 | X_2 \rangle_G$  la  $d$ -séparation des sous-ensembles  $X_1$  et  $X_2$  par le sous-ensemble  $X_3$  dans le graphe  $G$ .

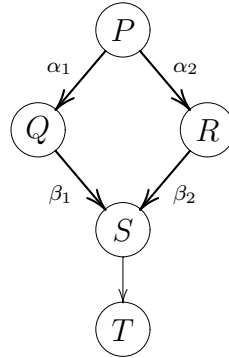


FIG. 2.6 – Exemple d'ensembles  $d$ -séparés. Les sous-ensembles  $\mathbf{X}_1 = \{Q\}$  et  $\mathbf{X}_2 = \{R\}$  sont  $d$ -séparés par le sous-ensemble  $\mathbf{X}_3 = \{P\}$ . En effet, les chaînes  $(\alpha_1, \alpha_2)$  et  $(\beta_1, \beta_2)$  sont bloquées par le sous-ensemble  $\mathbf{X}_3 = \{P\}$ .

#### Définition 2.4

Soit  $U$  un ensemble fini et  $I$  une relation ternaire sur les parties de  $U$ . On appelle **modèle de dépendance** sur  $U$  le couple  $(U, I)$ .

#### Définition 2.5

Soient  $U$  un ensemble fini,  $M = (U, I)$  un modèle de dépendance sur  $U$  et  $G = (U, A)$  un graphe orienté sans circuit sur  $U$ . On dit que le graphe orienté  $G$  est un **I-map** du modèle de dépendance  $M$  si et seulement si on a

$$\forall (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) \in \mathcal{P}(U)^3 \quad \langle \mathbf{X}_1 | \mathbf{X}_3 | \mathbf{X}_2 \rangle_G \Rightarrow I(\mathbf{X}_1, \mathbf{X}_3, \mathbf{X}_2) \quad (2.1)$$

On dit que le graphe orienté  $G$  est un **I-map minimal** du modèle de dépendance  $M$  si et seulement si  $G$  est un I-map du modèle de dépendance  $M$  et tout graphe partiel de  $G$  distinct de  $G$  n'est pas un I-map du modèle de dépendance  $M$ .

#### Définition 2.6

Soient  $(\Omega, \mathcal{F}, P)$  un espace de probabilité,  $U$  un ensemble fini d'éléments aléatoires de  $(\Omega, \mathcal{F})$  dans des espaces mesurables quelconques et  $G = (U, A)$  un graphe orienté sans circuit sur  $U$ . On dit que le graphe orienté  $G$  est un **réseau bayésien** sur  $U$  pour la mesure de probabilité  $P$  si et seulement si le graphe orienté  $G$  est un I-map minimal du modèle de dépendance  $(U, I_P)$ , où  $I_P$  est la relation de  $P$ -indépendance conditionnelle entre les parties de  $U$ .

Il ressort des définitions 2.5 et 2.6 que dans un réseau bayésien toute relation de  $d$ -séparation entre sous-ensembles de sommets implique une relation de  $P$ -indépendance conditionnelle entre les ensembles d'éléments aléatoires correspondant. Un réseau bayésien permet donc de représenter à l'aide d'une structure de graphe les relations d'indépendance conditionnelle entre ensembles d'éléments aléatoires.

## 2.3 Construction d'un réseau bayésien

Le but de cette partie est de déterminer un moyen de construire un réseau bayésien lorsque l'on dispose d'un ensemble d'éléments aléatoires vérifiant un certain nombre de relations d'indépendance conditionnelle. Pour cela, nous allons tout d'abord introduire les notions de *frontière markovienne* et de *graphoïde*. Nous aurons alors à notre disposition une proposition qui nous permettra de construire un  $I$ -map minimal à l'aide de frontières markoviennes. Nous pourrions ensuite appliquer directement cette proposition pour construire un réseau bayésien.

### Définition 2.7

Soient  $M = (U, I)$  un modèle de dépendance,  $B$  et  $V$  des parties de  $U$  et  $X$  un élément de  $V$ . On dit que le sous-ensemble  $B$  est une **couverture markovienne**<sup>1</sup> de l'élément  $X$  par rapport au sous-ensemble  $V$  pour le modèle de dépendance  $M$  si et seulement si  $B$  est inclus dans  $V \setminus \{X\}$  et vérifie

$$I(\{X\}, B, V \setminus (B \cup \{X\})) \quad (2.2)$$

On dit que le sous-ensemble  $B$  est une **frontière markovienne**<sup>2</sup> de l'élément  $X$  par rapport au sous-ensemble  $V$  pour le modèle de dépendance  $M$  si et seulement si le sous-ensemble  $B$  est une couverture markovienne de  $X$  par rapport à  $V$  pour le modèle de dépendance  $M$  et aucun sous-ensemble strict de  $B$  n'est une couverture markovienne de  $X$  par rapport à  $V$  pour le modèle de dépendance  $M$ .

Intuitivement, une frontière markovienne  $B$  d'un élément  $X$  par rapport à un ensemble  $V$  « protège » l'élément  $X$  de l'influence des autres éléments de  $V$  et aucun sous-ensemble strict de  $B$  ne possède cette propriété.

### Définition 2.8

Soit  $M = (U, I)$  un modèle de dépendance. On dit que le modèle de dépendance  $M$  est un **graphoïde** si et seulement si la relation ternaire  $I$  vérifie les propriétés suivantes

#### 1. Symétrie :

$$\forall (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3) \in \mathcal{P}(U)^3 \quad I(\mathbf{X}_1, \mathbf{X}_3, \mathbf{X}_2) \Rightarrow I(\mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_1) \quad (2.3)$$

#### 2. Décomposition :

$$\forall (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4) \in \mathcal{P}(U)^4 \quad I(\mathbf{X}_1, \mathbf{X}_3, \mathbf{X}_2 \cup \mathbf{X}_4) \Rightarrow I(\mathbf{X}_1, \mathbf{X}_3, \mathbf{X}_2) \wedge I(\mathbf{X}_1, \mathbf{X}_3, \mathbf{X}_4) \quad (2.4)$$

#### 3. Union faible :

$$\forall (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4) \in \mathcal{P}(U)^4 \quad I(\mathbf{X}_1, \mathbf{X}_3, \mathbf{X}_2 \cup \mathbf{X}_4) \Rightarrow I(\mathbf{X}_1, \mathbf{X}_3 \cup \mathbf{X}_2, \mathbf{X}_4) \quad (2.5)$$

#### 4. Contraction :

$$\forall (\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3, \mathbf{X}_4) \in \mathcal{P}(U)^4 \quad I(\mathbf{X}_1, \mathbf{X}_3 \cup \mathbf{X}_2, \mathbf{X}_4) \wedge I(\mathbf{X}_1, \mathbf{X}_3, \mathbf{X}_2) \Rightarrow I(\mathbf{X}_1, \mathbf{X}_3, \mathbf{X}_2 \cup \mathbf{X}_4) \quad (2.6)$$

<sup>1</sup>En anglais : *Markov blanket*.

<sup>2</sup>En anglais : *Markov boundary*.

**Proposition 2.1**

Soient  $M = (\mathbf{U}, I)$  un graphoïde,  $d = (X_1, \dots, X_n)$  un arrangement des éléments de  $\mathbf{U}$  et  $G = (\mathbf{U}, \mathbf{A})$  un graphe orienté sans circuit tel que pour tout  $i \in \{1, \dots, n\}$ , l'ensemble  $\text{Pa}(X_i)_G$  des parents de  $X_i$  dans  $G$  est une frontière markovienne de  $\{X_i\}$  par rapport à  $\{X_1, \dots, X_i\}$  pour le graphoïde  $M$ . Alors le graphe  $G$  est un  $I$ -map minimal du graphoïde  $M$ .

La démonstration de la proposition 2.1 a été donnée dans [VP88]. Comme cette proposition joue un rôle central dans la construction des réseaux bayésiens, nous en avons détaillé la démonstration à l'annexe B, page 141.

**Proposition 2.2**

Soient  $(\Omega, \mathcal{F}, P)$  un espace de probabilité,  $\mathbf{U}$  un ensemble fini d'éléments aléatoires de  $(\Omega, \mathcal{F})$  dans des espaces mesurables quelconques,  $d = (X_1, \dots, X_n)$  un arrangement des éléments de  $\mathbf{U}$  et  $b = (\mathbf{B}_1, \dots, \mathbf{B}_n)$  une suite de parties de  $\mathbf{U}$  tels que pour tout  $i \in \{1, \dots, n\}$ ,  $\mathbf{B}_i$  est une frontière markovienne de  $X_i$  par rapport à l'ensemble  $\{X_1, \dots, X_i\}$  pour le modèle de dépendance  $M = (\mathbf{U}, I_P)$ , où  $I_P$  est la relation de  $P$ -indépendance conditionnelle entre les parties de  $\mathbf{U}$ . Alors le graphe orienté  $G = (\mathbf{U}, \mathbf{A})$  vérifiant

$$\forall i \in \{1, \dots, n\} \quad \text{Pa}(\{X_i\})_G = \mathbf{B}_i \quad (2.7)$$

est un réseau bayésien sur  $\mathbf{U}$  pour la mesure de probabilité  $P$ . On dit alors que le réseau bayésien  $G$  est **engendré** par la **liste causale**  $L = (d, b)$ .

DÉMONSTRATION :

D'après la définition 2.8 et la proposition 1.3, le modèle de dépendance  $M = (\mathbf{U}, I_P)$  est un graphoïde. De plus, pour tout  $i \in \{1, \dots, n\}$ , l'ensemble  $\text{Pa}(X_i)_G$  des parents de  $X_i$  dans  $G$  est une frontière markovienne de  $\{X_i\}$  par rapport à  $\{X_1, \dots, X_i\}$  pour le graphoïde  $M = (\mathbf{U}, I_P)$ . D'après la proposition 2.1, le graphe  $G$  est donc un  $I$ -map minimal du graphoïde  $M$ , c'est-à-dire, d'après la définition 2.6, un réseau bayésien sur  $\mathbf{U}$  pour la mesure de probabilité  $P$ .  $\square$

Nous allons illustrer l'utilisation de la proposition 2.2 pour la construction des réseaux bayésiens à l'aide d'un exemple.

**Exemple 2.1**

Soient  $(\Omega, \mathcal{F}, P)$  un espace de probabilité,  $(E, \mathcal{E})$  un espace mesurable,  $t_0$  un entier strictement positif et  $S_0, U_0, Y_0, V_0, S_1, U_1, Y_1, V_1, \dots, S_{t_0}, U_{t_0}, Y_{t_0}, V_{t_0}$  des éléments aléatoires de  $(\Omega, \mathcal{F})$  dans  $(E, \mathcal{E})$  vérifiant, pour tout  $t \in \{0, 1, \dots, t_0\}$ ,

$$S_t = \begin{cases} f_t(S_{t-1}, U_t) & \text{si } t > 0 \\ f_0(U_0) & \text{si } t = 0 \end{cases} \quad (2.8)$$

$$Y_t = g_t(S_t, V_t) \quad (2.9)$$

où  $f_0$  est une application mesurable de  $(E, \mathcal{E})$  dans  $(E, \mathcal{E})$  et  $f_t$  et  $g_t$  sont des applications mesurables de  $(E, \mathcal{E}) \otimes (E, \mathcal{E})$  dans  $(E, \mathcal{E})$ . On suppose également que les éléments aléatoires  $U_0, V_0, U_1, V_1, \dots, U_{t_0}, V_{t_0}$  sont  $P$ -indépendants, que  $S_0$  n'est pas  $P$ -presque sûrement constant, que  $S_t$  n'est pas  $P$ -indépendant de  $S_{t-1}$  (resp.  $U_t$ ) conditionnellement à  $U_t$  (resp.  $S_{t-1}$ ) pour  $t \in \{1, \dots, t_0\}$  et que  $Y_t$  n'est pas  $P$ -indépendant de  $S_t$  (resp.  $V_t$ ) conditionnellement à  $V_t$  (resp.  $S_t$ ) pour  $t \in \{0, 1, \dots, t_0\}$ .

Afin d'obtenir un réseau bayésien sur  $U = \{S_0, U_0, Y_0, V_0, S_1, U_1, Y_1, V_1, \dots, S_{t_0}, U_{t_0}, Y_{t_0}, V_{t_0}\}$  pour la mesure de probabilité  $P$ , on pose

$$d = (U_0, S_0, V_0, Y_0, U_1, S_1, V_1, Y_1, \dots, U_{t_0}, S_{t_0}, V_{t_0}, Y_{t_0}) \quad (2.10)$$

$$b = (\mathbf{B}_{0,0}, \mathbf{B}_{0,1}, \mathbf{B}_{0,2}, \mathbf{B}_{0,3}, \mathbf{B}_{1,0}, \mathbf{B}_{1,1}, \mathbf{B}_{1,2}, \mathbf{B}_{1,3}, \dots, \mathbf{B}_{t_0,0}, \mathbf{B}_{t_0,1}, \mathbf{B}_{t_0,2}, \mathbf{B}_{t_0,3}) \quad (2.11)$$

avec  $\mathbf{B}_{0,0} = \emptyset$ ,  $\mathbf{B}_{0,1} = \{U_0\}$ ,  $\mathbf{B}_{0,2} = \emptyset$ ,  $\mathbf{B}_{0,3} = \{S_0, V_0\}$  et, pour tout  $t \in \{1, 2, \dots, t_0\}$ ,  $\mathbf{B}_{t,0} = \emptyset$ ,  $\mathbf{B}_{t,1} = \{S_{t-1}, U_t\}$ ,  $\mathbf{B}_{t,2} = \emptyset$ ,  $\mathbf{B}_{t,3} = \{S_t, V_t\}$ . On montre alors aisément que le couple  $L = (d, b)$  vérifie les hypothèses de la proposition 2.2, et on en déduit que le graphe représenté à la figure 2.7 est un réseau bayésien sur  $U$  pour la mesure de probabilité  $P$ .

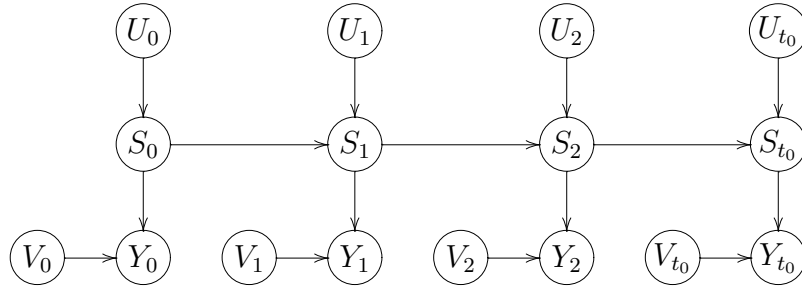


FIG. 2.7 – Réseau bayésien sur  $U$  pour la mesure de probabilité  $P$  engendré par la liste causale  $L = (d, b)$ . On a représenté le cas  $t_0 = 3$ .

Si on ne s'intéresse qu'aux éléments aléatoires  $S_t, Y_t$  pour  $t \in \{0, 1, \dots, t_0\}$ , on peut obtenir un réseau bayésien sur  $U' = \{S_0, Y_0, S_1, Y_1, \dots, S_{t_0}, Y_{t_0}\}$  pour la mesure de probabilité  $P$  en posant

$$d' = (S_0, Y_0, S_1, Y_1, \dots, S_{t_0}, Y_{t_0}) \quad (2.12)$$

$$b' = (\mathbf{B}'_{0,0}, \mathbf{B}'_{0,1}, \mathbf{B}'_{1,0}, \mathbf{B}'_{1,1}, \dots, \mathbf{B}'_{t_0,0}, \mathbf{B}'_{t_0,1}) \quad (2.13)$$

avec  $\mathbf{B}'_{0,0} = \emptyset$ ,  $\mathbf{B}'_{0,1} = \{S_0\}$  et, pour tout  $t \in \{1, 2, \dots, t_0\}$ ,  $\mathbf{B}'_{t,0} = \{S_{t-1}\}$  et  $\mathbf{B}'_{t,1} = \{S_t\}$ . En utilisant le réseau bayésien de la figure 2.7, on montre alors aisément que le couple  $L' = (d', b')$  vérifie les hypothèses de la proposition 2.2, et on en déduit que le graphe représenté à la figure 2.8 est un réseau bayésien sur  $U'$  pour la mesure de probabilité  $P$ .

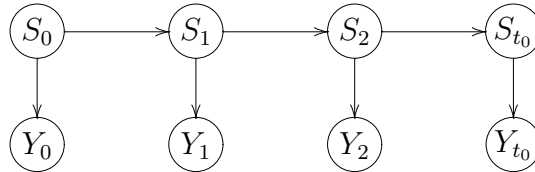


FIG. 2.8 – Réseau bayésien sur  $U'$  pour la mesure de probabilité  $P$  engendré par la liste causale  $L' = (d', b')$ . On a représenté le cas  $t_0 = 3$ .

## 2.4 Conclusion

Dans ce chapitre, nous avons présenté la notion de  $d$ -séparation entre ensembles de sommets d'un graphe orienté sans circuit et nous avons défini un réseau bayésien comme étant un graphe orienté sans circuit dont les sommets sont des éléments aléatoires et dont les arcs permettent de représenter à l'aide de relations de  $d$ -séparation les relations d'indépendance conditionnelle telle que définie dans le chapitre précédent. Nous avons ensuite présenté la notion de frontière markovienne et nous avons montré, en utilisant les propriétés de la relation d'indépendance conditionnelle démontrées au chapitre précédent, qu'il est possible de construire un réseau bayésien à partir d'une suite particulière de frontières markoviennes.

La définition des réseaux bayésiens à partir de la notion de  $d$ -séparation et leur construction à partir de suites de frontières markoviennes ont été énoncées par [Pea88] dans le cas particulier de variables aléatoires discrètes. Grâce à la définition de l'indépendance conditionnelle que nous avons donnée au chapitre précédent et aux propriétés de l'indépendance conditionnelle que nous avons démontrées dans ce même chapitre, nous sommes parvenus ici à généraliser la définition et la construction des réseaux bayésiens au cas d'éléments aléatoires *absolument quelconques*. Ainsi, nous pouvons maintenant utiliser en toute rigueur les réseaux bayésiens pour représenter et manipuler les relations d'indépendance conditionnelle existant par exemple entre des mesures aléatoires ou – autre exemple très utile en pratique – entre des variables aléatoires dont la loi conjointe n'admet pas de densité par rapport à la mesure de Lebesgue.

Après avoir défini les réseaux bayésiens et étudié leur construction dans le cadre très général qui est le nôtre, nous pouvons maintenant les appliquer à l'estimation statistique comme nous allons le voir dans le prochain chapitre.

# Chapitre 3

## L'algorithme de propagation de croyance

### 3.1 Introduction

Nous avons vu dans le chapitre précédent que les relations d'indépendance conditionnelle peuvent être représentées à l'aide de graphes appelés réseaux bayésiens. Dans ce chapitre, nous allons nous intéresser à un algorithme d'estimation statistique appelé *algorithme de propagation de croyance* qui s'applique à un réseau bayésien et qui exploite les relations d'indépendance conditionnelle représentées par ce réseau bayésien pour calculer – de façon exacte ou approchée – la loi *a posteriori* d'un sommet de ce réseau bayésien.

Nous allons tout d'abord rappeler plusieurs concepts dont nous aurons besoin pour formuler l'algorithme de propagation de croyance. Nous introduirons également plusieurs conventions de notation qui seront utilisées dans la suite. Nous montrerons ensuite que, dans le cas d'un réseau bayésien sans cycle, il est possible de calculer exactement la loi d'un sommet du réseau bayésien conditionnellement à un ensemble d'autres sommets, appelé *ensemble des observations*, à l'aide d'un ensemble d'équations locales appelées *équations de propagation de croyance* et nous appellerons *algorithme de propagation de croyance exacte* l'algorithme résultant. Enfin, dans le cas d'un réseau bayésien cyclique, nous utiliserons les équations de propagation de croyance pour calculer de façon itérative une approximation de la loi d'un sommet du réseau conditionnellement à un ensemble d'observations et nous appellerons *algorithme de propagation de croyance approchée* l'algorithme résultant. Comme les algorithmes de propagation de croyance exacte et de propagation de croyance approchée utilisent tous les deux les équations de propagation de croyance, il est d'usage de les considérer comme deux variantes d'un même algorithme que l'on conviendra d'appeler *algorithme de propagation de croyance*.

### 3.2 Rappels et notations

Dans cette partie, nous allons rappeler plusieurs concepts dont nous aurons besoin ultérieurement et introduire quelques conventions de notation spécifiques à ce document.

Nous allons commencer par rappeler les notions de transition et de transition finie entre espaces mesurables. Ces concepts, qui sont très utilisés dans l'étude des chaînes de Markov, nous permettront non seulement de formuler de façon très générale l'algorithme de propagation de croyance dans ce chapitre mais aussi d'énoncer de façon particulièrement concise l'algorithme



de filtrage particulière à échantillonnage global au chapitre 5.

### Définition 3.1

Soient  $(E_1, \mathcal{E}_1)$  et  $(E_2, \mathcal{E}_2)$  des espaces mesurables. On appelle **transition finie** (resp. **transition**) de l'espace mesurable  $(E_1, \mathcal{E}_1)$  dans l'espace mesurable  $(E_2, \mathcal{E}_2)$  toute fonction  $\Pi$  de  $E_1 \times \mathcal{E}_2$  dans  $\mathbb{R}_+$ ,  $(x_1, B_2) \mapsto \Pi(x_1; B_2)$ , vérifiant les propriétés suivantes :

1. pour tout  $x_1 \in E_1$ ,  $\Pi(x_1; \cdot)$  est une mesure finie (resp. une mesure de probabilité) sur l'espace mesurable  $(E_2, \mathcal{E}_2)$
2. pour tout événement  $B_2 \in \mathcal{E}_2$ ,  $\Pi(\cdot; B_2)$  est une variable aléatoire sur  $(E_1, \mathcal{E}_1)$ .

On note  $\Pi : (E_1, \mathcal{E}_1) \prec (E_2, \mathcal{E}_2)$  une transition finie  $\Pi$  de l'espace mesurable  $(E_1, \mathcal{E}_1)$  dans l'espace mesurable  $(E_2, \mathcal{E}_2)$ . On dit que la transition finie  $\Pi$  est une **transition bornée** si et seulement si il existe  $M \geq 0$  tel que pour tout  $x_1 \in E_1$ , on a

$$\Pi(x_1; E_2) \leq M \quad (3.1)$$

Après ce rappel, nous pouvons maintenant introduire plusieurs conventions relatives aux transitions.

### Convention 3.1

Soit  $I \subset \mathbb{N}$ . Soit  $\left( (E_i, \mathcal{E}_i) \right)_{i \in I}$  une suite d'espaces mesurables. Soient  $A, B$  et  $C$  des sous-ensembles finis et non-vides de  $I$  tels que  $A \cap C = \emptyset$ .

Soit  $F_A$  une mesure de probabilité sur l'espace mesurable  $\bigotimes_{i \in A} (E_i, \mathcal{E}_i)$ .

Soit  $\Pi_C^B$  une transition bornée de l'espace mesurable  $\bigotimes_{i \in B} (E_i, \mathcal{E}_i)$  dans l'espace mesurable  $\bigotimes_{i \in C} (E_i, \mathcal{E}_i)$ .

On pose  $R = B \setminus A$  et  $S = A \cup C$ . Alors,

1. si  $R$  est non vide, on pose, pour tout  $x = (x_i)_{i \in R} \in \prod_{i \in R} E_i$  et pour tout  $\Gamma \in \bigotimes_{i \in S} \mathcal{E}_i$ ,

$$F_A \otimes \Pi_C^B(x; \Gamma) \stackrel{\text{déf}}{=} \int \left[ \int \mathbf{1}_\Gamma((x_i)_{i \in S}) \Pi_C^B((x_i)_{i \in B}; d(x_i)_{i \in C}) \right] F_A(d(x_i)_{i \in A})$$

On vérifie alors que  $F_A \otimes \Pi_C^B$  est une transition finie de l'espace mesurable  $\bigotimes_{i \in R} (E_i, \mathcal{E}_i)$  dans l'espace mesurable  $\bigotimes_{i \in S} (E_i, \mathcal{E}_i)$ .

2. si  $R$  est vide, on pose, pour tout  $\Gamma \in \bigotimes_{i \in S} \mathcal{E}_i$ ,

$$F_A \otimes \Pi_C^B(\Gamma) \stackrel{\text{déf}}{=} \int \left[ \int \mathbf{1}_\Gamma((x_i)_{i \in S}) \Pi_C^B((x_i)_{i \in B}; d(x_i)_{i \in C}) \right] F_A(d(x_i)_{i \in A})$$

On vérifie alors que  $F_A \otimes \Pi_C^B$  est une mesure finie sur l'espace mesurable  $\bigotimes_{i \in S} (E_i, \mathcal{E}_i)$ .

### Convention 3.2

Soit  $(\Omega, \mathcal{F}, P)$  un espace de probabilité. Soient  $(E_1, \mathcal{E}_1)$ ,  $(E_2, \mathcal{E}_2)$  et  $(E_3, \mathcal{E}_3)$  des espaces mesurables. Soit  $X_2$  un élément aléatoire de  $(\Omega, \mathcal{F})$  dans  $(E_2, \mathcal{E}_2)$ . Soit une transition finie  $\Pi_{2,3}^1 : (E_1, \mathcal{E}_1) \prec (E_2, \mathcal{E}_2) \otimes (E_3, \mathcal{E}_3)$ . Alors, on définit la transition finie  $\sum_{X_2} \Pi_{2,3}^1 :$

$(E_1, \mathcal{E}_1) \prec (E_3, \mathcal{E}_3)$  en posant, pour tout  $x_1 \in E_1$  et pour tout  $B_3 \in \mathcal{E}_3$

$$\left( \sum_{X_2} \Pi_{2,3}^1 \right) (x_1; B_3) \stackrel{\text{déf}}{=} \Pi_{2,3}^1(x_1; E_2 \times B_3)$$

**Convention 3.3**

Soit  $(\Omega, \mathcal{F}, P)$  un espace de probabilité. Soient  $(E_1, \mathcal{E}_1)$  et  $(E_2, \mathcal{E}_2)$  des espaces mesurables. Soit une transition finie  $\Pi_2^1 : (E_1, \mathcal{E}_1) \prec (E_2, \mathcal{E}_2)$ . Soit  $g_2$  une fonction mesurable positive bornée sur  $(E_2, \mathcal{E}_2)$ . Alors, on définit la transition finie  $g_2.\Pi_2^1 : (E_1, \mathcal{E}_1) \prec (E_2, \mathcal{E}_2)$  en posant, pour tout  $x_1 \in \mathbb{E}_1$

$$(g_2.\Pi_2^1)(x_1; \cdot) \stackrel{\text{déf}}{=} g_2.\Pi_2^1(x_1; \cdot)$$

**Convention 3.4**

Soit  $(\Omega, \mathcal{F}, P)$  un espace de probabilité. Soit  $(E, \mathcal{E})$  un espace mesurable. Soit  $X$  un élément aléatoire de  $(\Omega, \mathcal{F})$  dans  $(E, \mathcal{E})$ . On convient de noter  $F_X = X(P)$  la loi de  $X$ .

**Convention 3.5**

Soit  $(\Omega, \mathcal{F}, P)$  un espace de probabilité. Soient  $(E_1, \mathcal{E}_1)$  et  $(E_2, \mathcal{E}_2)$  deux espaces mesurables. Soient  $X_1$  un élément aléatoire de  $(\Omega, \mathcal{F})$  dans  $(E_1, \mathcal{E}_1)$  et  $X_2$  un élément aléatoire de  $(\Omega, \mathcal{F})$  dans  $(E_2, \mathcal{E}_2)$ . On convient de noter  $\Pi_{X_2}^{X_1}$  une transition de  $(E_1, \mathcal{E}_1)$  dans  $(E_2, \mathcal{E}_2)$  telle que pour tout  $B_2 \in \mathcal{E}_2$ ,  $\Pi_{X_2}^{X_1}(X_1; B_2)$  est une version de  $P(X_2 \in B_2 | X_1)$ .

Dans la suite, nous supposons toujours qu'une telle transition existe. On sait que c'est notamment le cas lorsque  $E_1$  et  $E_2$  sont des espaces métriques complets séparables (autrement dit polonais) et  $\mathcal{E}_1$  et  $\mathcal{E}_2$  sont leurs tribus boréliennes. Par commodité de langage, nous nous référons alors à « la » transition  $\Pi_{X_2}^{X_1}$ , bien qu'en général il n'y ait pas unicité d'une telle transition.

Afin de formuler les équations de propagation de croyance, nous aurons également besoin de l'opérateur de Bayes qui intervient dans l'énoncé du théorème de Bayes généralisé (cf § C.1).

**Définition 3.2**

Soit  $(E, \mathcal{E})$  un espace mesurable. On appelle **opérateur de Bayes** et on note  $\Psi_{\text{Bayes}}$  l'opérateur qui, à toute fonction mesurable positive  $g$  sur  $(E, \mathcal{E})$  et à toute mesure  $\mu$  sur  $(E, \mathcal{E})$ , associe la mesure

$$\Psi_{\text{Bayes}}(g, \mu) = \begin{cases} \frac{g}{\mu(g)} \cdot \mu & \text{si } \mu(g) \neq 0 \\ \mu & \text{sinon.} \end{cases} \quad (3.2)$$

Enfin, nous introduisons une convention de notation qui nous sera utile pour formuler l'algorithme de propagation de croyance approchée.

**Convention 3.6**

Soit  $G = (U, \mathbf{A})$  un réseau bayésien. Soit  $c = (X, Y)$  un couple de sommets de  $G$ . On convient de noter  $\bar{c}$  le couple  $(Y, X)$ .

### 3.3 L'algorithme de propagation de croyance exacte

Nous allons maintenant nous intéresser au calcul de la loi d'un sommet d'un réseau bayésien conditionnellement à un ensemble d'autres sommets, appelé *ensemble des observations*. Nous allons montrer que dans le cas d'un réseau bayésien sans cycle, il est possible de calculer exactement cette loi *a posteriori* à l'aide d'une série de calculs locaux. Pour cela, nous allons

tout d'abord introduire la notion de *message* envoyé par un sommet à un autre sommet le long d'un arc. Nous montrerons ensuite que la loi *a posteriori* d'un sommet n'appartenant pas aux observations peut être calculée à l'aide des messages envoyés à ce sommet par ses parents et ses enfants. Enfin nous montrerons que ces messages peuvent à leur tour être calculés à partir d'autres messages grâce à un ensemble d'équations locales appelées *équations de propagation de croyance*. Nous appellerons alors *algorithme de propagation de croyance exacte* l'algorithme ainsi obtenu.

### 3.3.1 Messages

#### Définition 3.3

Soit  $(\Omega, \mathcal{F}, P)$  un espace de probabilité. Soit  $\mathcal{L}$  une partie finie de  $\mathbb{N}$ . Pour tout  $l \in \mathcal{L}$ , soit  $(E_l, \mathcal{E}_l)$  un espace mesurable et  $X_l$  un élément aléatoire de  $(\Omega, \mathcal{F})$  dans  $(E_l, \mathcal{E}_l)$ . On pose  $U = \{X_l\}_{l \in \mathcal{L}}$ . Soit  $G = (U, \mathbf{A})$  un réseau bayésien sur  $U$  pour la mesure de probabilité  $P$ . Soit  $Z$  une partie non-vide de  $U$  telle que les éléments de  $Z$  n'ont pas d'enfants dans  $G$ . Pour tout  $l \in \mathcal{L}$  tel que  $X_l \in Z$ , soit  $\nu_l$  une mesure positive et  $\sigma$ -finie sur  $(E_l, \mathcal{E}_l)$ . On suppose que la mesure  $\nu_l$  domine la loi de  $X_l$ . Alors, pour tout arc  $(X, Y)$  de  $G$ ,

1. on note  $Z_{XY}^+$  l'ensemble des éléments de  $Z$  qui sont joints à  $Y$  par une chaîne  $(\alpha_1, \dots, \alpha_q)$  telle que  $\alpha_q = (X, Y)$  et  $Z_{XY}^+$  le vecteur formé par les éléments de  $Z_{XY}^+$  rangés par indice croissant.
  - (a) Si  $Z_{XY}^+$  est non vide, on convient d'appeler **message de X à Y** et de noter  $\mu_{XY}^+$  la transition  $\Pi_X^{Z_{XY}^+}$ .
  - (b) Si  $Z_{XY}^+$  est vide, on convient d'appeler **message de X à Y** et de noter  $\mu_{XY}^+$  la loi  $F_X$ .
2. on note  $Z_{XY}^-$  l'ensemble des éléments de  $Z$  qui sont joints à  $X$  par une chaîne  $(\alpha_1, \dots, \alpha_q)$  telle que  $\alpha_q = (X, Y)$ ,  $Z_{XY}^-$  le vecteur formé par les éléments de  $Z_{XY}^-$  rangés par indice croissant et  $(\mathfrak{X}, \mathcal{X})$  l'espace image de  $X$ .
  - (a) Si  $Z_{XY}^-$  est non vide, on note  $(\mathfrak{Z}_{XY}^-, \mathcal{Z}_{XY}^-)$  l'espace image du vecteur  $Z_{XY}^-$  et  $\nu_{XY}^-$  la mesure produit sur  $(\mathfrak{Z}_{XY}^-, \mathcal{Z}_{XY}^-)$  définie par

$$\nu_{XY}^- = \bigotimes_{\substack{l \in \mathcal{L} \\ X_l \in Z_{XY}^-}} \nu_l \quad (3.3)$$

On suppose que, pour tout  $x \in \mathfrak{X}$ , la mesure  $\Pi_{Z_{XY}^-}^X(x; \cdot)$  admet une densité notée  $f_{Z_{XY}^-}^X(x; \cdot)$  par rapport à la mesure  $\nu_{XY}^-$  et que l'application  $f_{Z_{XY}^-}^X$  est une application mesurable de  $(\mathfrak{X}, \mathcal{X}) \otimes (\mathfrak{Z}_{XY}^-, \mathcal{Z}_{XY}^-)$  dans  $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ . On convient alors d'appeler **message de Y à X** et de noter  $\mu_{YX}^-$  l'application  $f_{Z_{XY}^-}^X$ .

- (b) Si  $Z_{XY}^-$  est vide, on convient d'appeler **message de Y à X** et de noter  $\mu_{YX}^-$  l'application mesurable constante de  $(\mathfrak{X}, \mathcal{X})$  dans  $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$  valant 1.

Enfin, on convient d'appeler  $Z$  l'ensemble des **observations**,  $Z_{XY}^+$  l'ensemble des **observations situées en amont de Y et reliées à Y via X** et  $Z_{XY}^-$  l'ensemble des **observations situées en aval de X et reliées à X via Y**.

### 3.3.2 Calcul de la loi *a posteriori* d'un sommet.

#### Proposition 3.1

On suppose que le réseau bayésien  $G$  ne possède pas de cycle et on considère un sommet  $X$  de  $G$  n'appartenant pas à  $\mathcal{Z}$ . On note  $U_1, \dots, U_r$  les parents de  $X$  et  $Y_1, \dots, Y_s$  les enfants de  $X$ . Alors, la loi *a posteriori* de  $X$  étant donnée les observations  $\mathcal{Z}$  vérifie

$$\forall z \in \mathfrak{Z} \quad \Pi_X^{\mathcal{Z}}(z; \cdot) = \Psi_{\text{Bayes}} \left( \prod_j \mu_{Y_j X}^-(\cdot; z_{XY_j}^-), \dots \right. \\ \left. \dots \sum_{(U_1, \dots, U_r)} \left[ \bigotimes_i \mu_{U_i X}^+(z_{U_i X}^+; \cdot) \right] \otimes \Pi_X^{(U_1, \dots, U_r)} \right) \quad (3.4)$$

avec les conventions suivantes

- $z_{XY_j}^-$  désigne le projeté de  $z$  sur l'espace  $\mathfrak{Z}_{XY_j}^-$
- $z_{U_i X}^+$  désigne le projeté de  $z$  sur l'espace  $\mathfrak{Z}_{U_i X}^+$
- lorsque l'ensemble  $\mathcal{Z}_{XY_j}^-$  est vide, le terme  $\mu_{Y_j X}^-(\cdot; z_{XY_j}^-)$  doit être remplacé par le terme  $\mu_{Y_j X}^-$  dans l'expression 3.4
- lorsque l'ensemble  $\mathcal{Z}_{U_i X}^+$  est vide, le terme  $\mu_{U_i X}^+(z_{U_i X}^+; \cdot)$  doit être remplacé par le terme  $\mu_{U_i X}^+$  dans l'expression 3.4
- lorsque  $X$  n'a pas de parent, l'expression 3.4 devient

$$\Pi_X^{\mathcal{Z}}(z; \cdot) = \Psi_{\text{Bayes}} \left( \prod_j \mu_{Y_j X}^-(\cdot; z_{XY_j}^-), F_X \right) \quad (3.5)$$

- lorsque  $X$  n'a pas d'enfants, l'expression 3.4 devient

$$\Pi_X^{\mathcal{Z}}(z; \cdot) = \sum_{(U_1, \dots, U_r)} \left[ \bigotimes_i \mu_{U_i X}^+(z_{U_i X}^+; \cdot) \right] \otimes \Pi_X^{(U_1, \dots, U_r)} \quad (3.6)$$

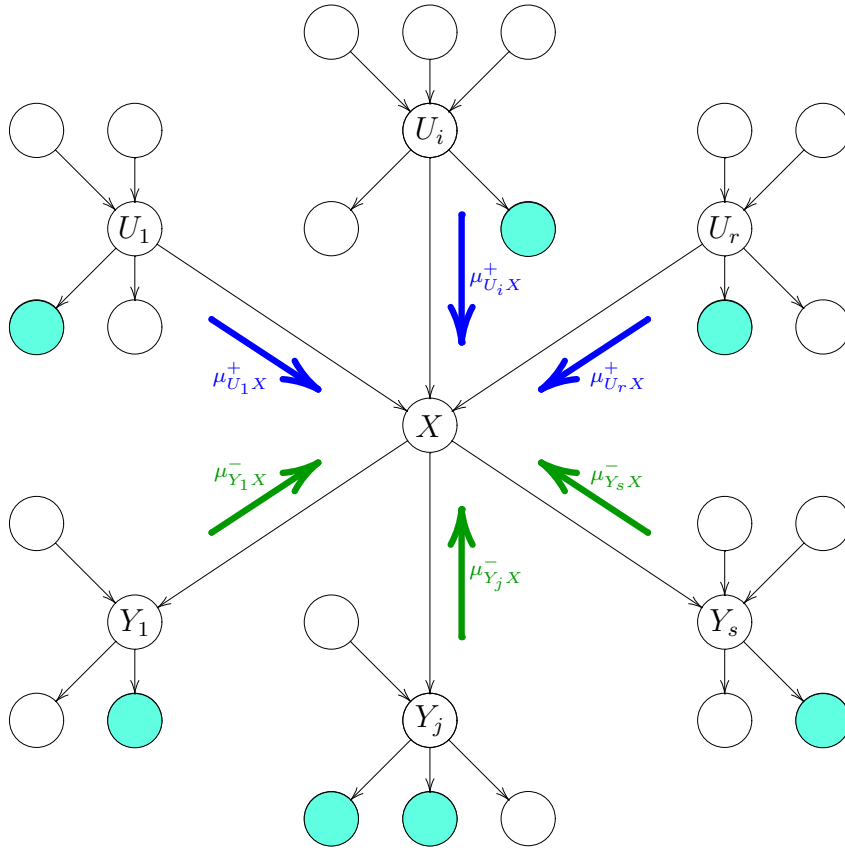


FIG. 3.1 – Calcul de  $\Pi_X^Z$ . Les sommets appartenant à l'ensemble des observations ont été représentés en vert.

DÉMONSTRATION :

Dans cette démonstration, on supposera que  $X$  possède au moins un parent et un enfant et que les ensembles  $Z_{XY_j}^-$  et  $Z_{U_iX}^+$  sont non-vides. On vérifiera aisément que le traitement de ces différents cas particuliers ne pose aucune difficulté.

Considérons l'ensemble  $Z_X^+ = \bigcup_i Z_{U_iX}^+$  des observations situées en amont de  $X$ . On remarque que pour tout  $i \in \{1, \dots, r\}$  on a

$$I_P(\{U_i\}, Z_{U_iX}^+, Z_X^+ \setminus Z_{U_iX}^+) \quad (3.7)$$

car tous les chemins joignant  $U_i$  à un sommet de  $Z_X^+ \setminus Z_{U_iX}^+$  convergent en  $X$  qui n'appartient pas à  $Z_{U_iX}^+$  et dont aucun descendant n'appartient à  $Z_{U_iX}^+$ . En utilisant la proposition 1.1, on en déduit

$$\Pi_{U_i}^{Z_{U_iX}^+}(z_{U_iX}^+; \cdot) = \Pi_{U_i}^{Z_X^+}(z_X^+; \cdot) \quad (3.8)$$

où  $Z_X^+$  désigne le vecteur formé par les éléments de  $Z_X^+$  rangés par indice croissant et  $z_X^+$  est le projeté de  $z$  sur l'espace image de  $Z_X^+$ . On remarque alors que l'on a également

$$I_P(\{U_1\}, Z_X^+, \{U_2\}) \quad (3.9)$$

car le chemin joignant  $U_1$  à  $U_2$  converge en  $X$  qui n'appartient pas à  $Z_X^+$  et dont aucun descendant n'appartient à  $Z_X^+$ . En utilisant la proposition C.4, on en déduit

$$\Pi_{U_1}^{Z_X^+}(z_X^+; \cdot) \otimes \Pi_{U_2}^{Z_X^+}(z_X^+; \cdot) = \Pi_{(U_1, U_2)}^{Z_X^+}(z_X^+; \cdot) \quad (3.10)$$

En remarquant ensuite que l'on a

$$I_P(\{U_1, U_2\}, Z_X^+, \{U_3\}) \quad (3.11)$$

on en déduit

$$\Pi_{U_1}^{Z_X^+}(z_X^+; \cdot) \otimes \Pi_{U_2}^{Z_X^+}(z_X^+; \cdot) \otimes \Pi_{U_3}^{Z_X^+}(z_X^+; \cdot) = \Pi_{(U_1, U_2)}^{Z_X^+}(z_X^+; \cdot) \otimes \Pi_{U_3}^{Z_X^+}(z_X^+; \cdot) \quad (3.12)$$

$$= \Pi_{(U_1, U_2, U_3)}^{Z_X^+}(z_X^+; \cdot) \quad (3.13)$$

En poursuivant ce raisonnement, on obtient finalement la formule suivante

$$\bigotimes_i \Pi_{U_i}^{Z_X^+}(z_X^+; \cdot) = \Pi_{(U_1, \dots, U_r)}^{Z_X^+}(z_X^+; \cdot) \quad (3.14)$$

On remarque également que l'on a

$$I_P(\{X\}, \{U_1, U_2, \dots, U_r\}, Z_X^+) \quad (3.15)$$

car tout chemin joignant un sommet de  $Z_X^+$  à  $X$  diverge ou est en série en un sommet de  $\{U_1, U_2, \dots, U_r\}$ . En utilisant la proposition 1.1, on en déduit

$$\Pi_X^{(U_1, \dots, U_r)}(u_1, \dots, u_r; \cdot) = \Pi_X^{(Z_X^+, U_1, \dots, U_r)}(z_X^+, u_1, \dots, u_r; \cdot) \quad (3.16)$$

Les équations 3.8, 3.14 et 3.16 impliquent

$$\begin{aligned} \sum_{(U_1, \dots, U_r)} \left[ \bigotimes_i \mu_{U_i X}^+(z_{U_i X}^+; \cdot) \right] \otimes \Pi_X^{(U_1, \dots, U_r)} &= \sum_{(U_1, \dots, U_r)} \Pi_{(U_1, \dots, U_r)}^{Z_X^+}(z_X^+; \cdot) \otimes \dots \\ &\dots \otimes \Pi_X^{(Z_X^+, U_1, \dots, U_r)}(z_X^+, \cdot; \cdot) \end{aligned} \quad (3.17)$$

En utilisant la proposition C.3, on en déduit

$$\sum_{(U_1, \dots, U_r)} \left[ \bigotimes_i \mu_{U_i X}^+(z_{U_i X}^+; \cdot) \right] \otimes \Pi_X^{(U_1, \dots, U_r)} = \sum_{(U_1, \dots, U_r)} \Pi_{(U_1, \dots, U_r, X)}^{Z_X^+}(z_X^+; \cdot) \quad (3.18)$$

$$= \Pi_X^{Z_X^+}(z_X^+; \cdot) \quad (3.19)$$

D'autre part, on remarque que l'on a

$$I_P(Z_{XY_1}^-, \{X\}, Z_{XY_2}^-) \quad (3.20)$$

car tout chemin joignant un sommet de  $Z_{XY_1}^-$  à un sommet de  $Z_{XY_2}^-$  diverge en  $X$ . En utilisant la proposition C.4, on en déduit, pour tout  $x \in \mathfrak{X}$

$$\Pi_{Z_{XY_1}^-}^X(x; \cdot) \otimes \Pi_{Z_{XY_2}^-}^X(x; \cdot) = \Pi_{(Z_{XY_1}^-, Z_{XY_2}^-)}^X(x; \cdot) \quad (3.21)$$

En remarquant ensuite que l'on a

$$I_P(\mathbf{Z}_{XY_1}^- \cup \mathbf{Z}_{XY_2}^-, \{X\}, \mathbf{Z}_{XY_3}^-) \quad (3.22)$$

on en déduit

$$\Pi_{\mathbf{Z}_{XY_1}^-}^X(x; \cdot) \otimes \Pi_{\mathbf{Z}_{XY_2}^-}^X(x; \cdot) \otimes \Pi_{\mathbf{Z}_{XY_3}^-}^X(x; \cdot) = \Pi_{(\mathbf{Z}_{XY_1}^-, \mathbf{Z}_{XY_2}^-)}^X(x; \cdot) \otimes \Pi_{\mathbf{Z}_{XY_3}^-}^X(x; \cdot) \quad (3.23)$$

$$= \Pi_{(\mathbf{Z}_{XY_1}^-, \mathbf{Z}_{XY_2}^-, \mathbf{Z}_{XY_3}^-)}^X(x; \cdot) \quad (3.24)$$

En poursuivant ce raisonnement, on obtient finalement la formule suivante

$$\bigotimes_j \Pi_{\mathbf{Z}_{XY_j}^-}^X(x; \cdot) = \Pi_{\mathbf{Z}_X^-}^X(x; \cdot) \quad (3.25)$$

avec

$$\mathbf{Z}_X^- \stackrel{\text{déf}}{=} (\mathbf{Z}_{XY_1}^-, \mathbf{Z}_{XY_2}^-, \dots, \mathbf{Z}_{XY_s}^-) \quad (3.26)$$

En posant  $\mathbf{Z}_X^- = \bigcup_j \mathbf{Z}_{XY_j}^-$ , on remarque que l'on a

$$I_P(\mathbf{Z}_X^+, \{X\}, \mathbf{Z}_X^-) \quad (3.27)$$

car tout chemin joignant un sommet de  $\mathbf{Z}_X^+$  à un sommet de  $\mathbf{Z}_X^-$  est en série ou diverge en  $X$ . En utilisant la proposition 1.1, on en déduit

$$\bigotimes_j \Pi_{\mathbf{Z}_{XY_j}^-}^X(x; \cdot) = \Pi_{\mathbf{Z}_X^-}^{(\mathbf{Z}_X^+, X)}(z_X^+, x; \cdot) \quad (3.28)$$

Or, pour tout  $j \in \{1, 2, \dots, s\}$ , la mesure  $\Pi_{\mathbf{Z}_{XY_j}^-}^X(x; \cdot)$  vérifie

$$\Pi_{\mathbf{Z}_{XY_j}^-}^X(x; \cdot) = f_{\mathbf{Z}_{XY_j}^-}^X(x; \cdot) \cdot \nu_{\mathbf{Z}_{XY_j}^-} \quad (3.29)$$

Par conséquent, les équations 3.28 et 3.29 impliquent

$$\Pi_{\mathbf{Z}_X^-}^{(\mathbf{Z}_X^+, X)}(z_X^+, x; \cdot) = f_{\mathbf{Z}_X^-}^{(\mathbf{Z}_X^+, X)}(z_X^+, x; \cdot) \cdot \nu_{\mathbf{Z}_X^-} \quad (3.30)$$

avec, pour tout  $z_X^{-'} = (z_{XY_1}^{-'}, z_{XY_2}^{-'}, \dots, z_{XY_s}^{-'})$ ,

$$f_{\mathbf{Z}_X^-}^{(\mathbf{Z}_X^+, X)}(z_X^+, x; z_X^{-'}) \stackrel{\text{déf}}{=} \prod_j f_{\mathbf{Z}_{XY_j}^-}^X(x; z_{XY_j}^{-'}) \quad (3.31)$$

$$= \prod_j \mu_{Y_j X}^-(x; z_{XY_j}^{-'}) \quad (3.32)$$

et

$$\nu_{\mathbf{Z}_X^-} \stackrel{\text{déf}}{=} \bigotimes_j \nu_{\mathbf{Z}_{XY_j}^-} \quad (3.33)$$

Les équations 3.19 et 3.32 impliquent alors

$$\begin{aligned} \Psi_{\text{Bayes}} \left( \prod_j \mu_{Y_j X}^-(\cdot; z_{XY_j}^-), \sum_{(U_1, \dots, U_r)} \left[ \bigotimes_i \mu_{U_i X}^+(z_{U_i X}^+; \cdot) \right] \otimes \Pi_X^{(U_1, \dots, U_r)} \right) = \dots \\ \dots \Psi_{\text{Bayes}} \left( f_{Z_X^-}^{(Z_X^+, X)}(z_X^+, \cdot; z_X^-), \Pi_X^{Z_X^+}(z_X^+; \cdot) \right) \end{aligned} \quad (3.34)$$

La mesure  $\nu_X^-$  étant positive et  $\sigma$ -finie, on peut utiliser la proposition C.2 pour calculer le membre de droite de l'équation 3.34. On obtient

$$\Psi_{\text{Bayes}} \left( \prod_j \mu_{Y_j X}^-(\cdot; z_{XY_j}^-), \sum_{(U_1, \dots, U_r)} \left[ \bigotimes_i \mu_{U_i X}^+(z_{U_i X}^+; \cdot) \right] \otimes \Pi_X^{(U_1, \dots, U_r)} \right) = \dots \quad (3.35)$$

$$\begin{aligned} \dots \Pi_X^{(Z_X^+, Z_X^-)}(z_X^+, z_X^-; \cdot) \\ = \Pi_X^Z(z; \cdot) \end{aligned} \quad (3.36)$$

ce qui est l'égalité à démontrer.  $\square$

### 3.3.3 Calcul du message envoyé par un parent à un enfant.

#### Proposition 3.2

On suppose que le réseau bayésien  $G$  ne possède pas de cycle et on considère un sommet  $X$  de  $G$  n'appartenant pas à  $\mathcal{Z}$ . On note  $U_1, \dots, U_r$  les parents de  $X$  et  $Y_1, \dots, Y_s$  les enfants de  $X$ . Alors, pour tout  $j_0 \in \{1, 2, \dots, s\}$  le message  $\mu_{XY_{j_0}}^+$  envoyé par  $X$  à  $Y_{j_0}$  vérifie

$$\begin{aligned} \forall z_{XY_{j_0}}^+ \in \mathfrak{Z}_{XY_{j_0}}^+ \quad \mu_{XY_{j_0}}^+(z_{XY_{j_0}}^+; \cdot) = \Psi_{\text{Bayes}} \left( \prod_{j \neq j_0} \mu_{Y_j X}^-(\cdot; z_{XY_j}^-), \dots \\ \dots \sum_{(U_1, \dots, U_r)} \left[ \bigotimes_i \mu_{U_i X}^+(z_{U_i X}^+; \cdot) \right] \otimes \Pi_X^{(U_1, \dots, U_r)} \right) \end{aligned} \quad (3.37)$$

avec les conventions suivantes :

- $z_{XY_j}^-$  désigne le projeté de  $z_{XY_{j_0}}^+$  sur l'espace  $\mathfrak{Z}_{XY_j}^-$
- $z_{U_i X}^+$  désigne le projeté de  $z_{XY_{j_0}}^+$  sur l'espace  $\mathfrak{Z}_{U_i X}^+$
- lorsque l'ensemble  $\mathcal{Z}_{XY_j}^-$  est vide, le terme  $\mu_{Y_j X}^-(\cdot; z_{XY_j}^-)$  doit être remplacé par le terme  $\mu_{Y_j X}^-$  dans l'expression 3.37
- lorsque l'ensemble  $\mathcal{Z}_{U_i X}^+$  est vide, le terme  $\mu_{U_i X}^+(z_{U_i X}^+; \cdot)$  doit être remplacé par le terme  $\mu_{U_i X}^+$  dans l'expression 3.37
- lorsque  $X$  a  $Y_{j_0}$  comme unique enfant et possède au moins un parent, l'expression 3.37 devient

$$\mu_{XY_{j_0}}^+(z_{XY_{j_0}}^+; \cdot) = \sum_{(U_1, \dots, U_r)} \left[ \bigotimes_i \mu_{U_i X}^+(z_{U_i X}^+; \cdot) \right] \otimes \Pi_X^{(U_1, \dots, U_r)} \quad (3.38)$$

- lorsque  $X$  n'a pas de parent et possède au moins un enfant autre que  $Y_{j_0}$ , l'expression 3.37 devient

$$\mu_{XY_{j_0}}^+(z_{XY_{j_0}}^+; \cdot) = \Psi_{\text{Bayes}} \left( \prod_{j \neq j_0} \mu_{Y_j X}^-(\cdot; z_{XY_j}^-), F_X \right) \quad (3.39)$$



– lorsque  $X$  n'a ni parent et ni d'autre enfant que  $Y_{j_0}$ , l'expression 3.37 devient

$$\mu_{XY_{j_0}}^+ = F_X \quad (3.40)$$

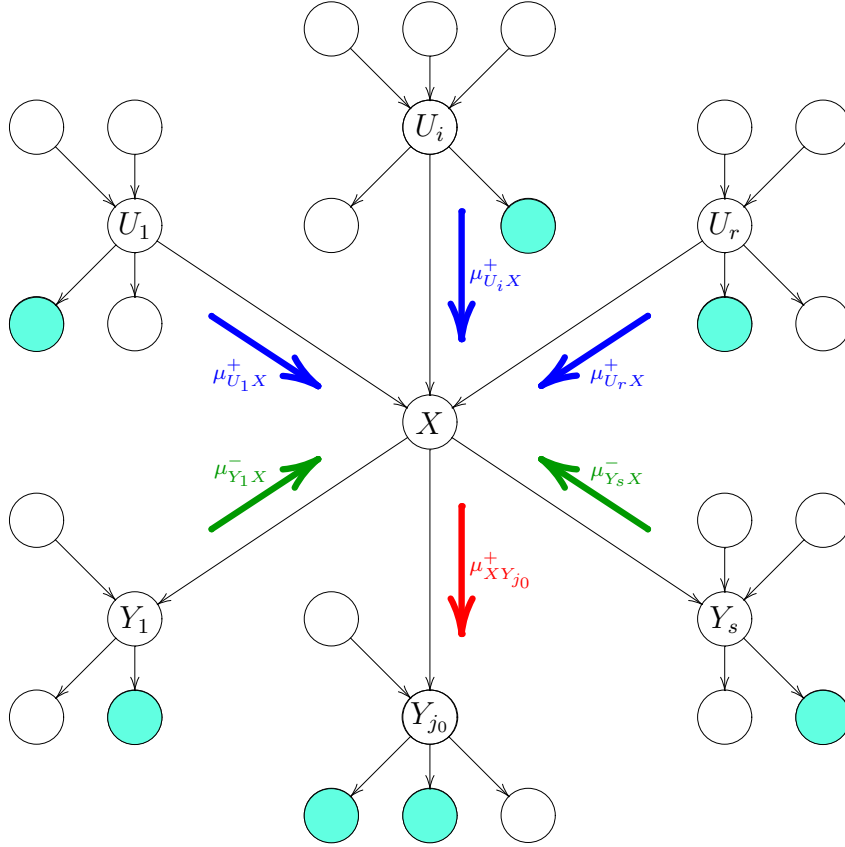


FIG. 3.2 – Calcul de  $\mu_{XY_{j_0}}^+$ .

DÉMONSTRATION :

Dans cette démonstration, on supposera que  $X$  possède au moins un parent et un enfant autre que  $Y_{j_0}$ , et que les ensembles  $Z_{XY_j}^-$  et  $Z_{U_i X}^+$  sont non-vides. On vérifiera aisément que le traitement de ces différents cas particuliers ne pose aucune difficulté.

En procédant comme dans la démonstration de la proposition 3.1, on montre que l'on a

$$\sum_{(U_1, \dots, U_r)} \left[ \bigotimes_i \mu_{U_i X}^+(z_{U_i X}^+; \cdot) \right] \otimes \Pi_X^{(U_1, \dots, U_r)} = \Pi_X^{Z_X^+}(z_X^+; \cdot) \quad (3.41)$$

et

$$\Pi_{(Z_{XY_j}^-)_{j \neq j_0}}^{(Z_X^+, X)}(z_X^+, x; \cdot) = f_{(Z_{XY_j}^-)_{j \neq j_0}}^{(Z_X^+, X)}(z_X^+, x; \cdot) \cdot \bigotimes_{j \neq j_0} \nu_{XY_j}^- \quad (3.42)$$

avec, pour tout  $(z_{XY_j}^{-'})_{j \neq j_0}$ ,

$$f_{(Z_{XY_j}^+, X)}(z_X^+, x; (z_{XY_j}^{-'})_{j \neq j_0}) = \prod_{j \neq j_0} \mu_{Y_j X}^-(x; z_{XY_j}^{-'}) \quad (3.43)$$

Les équations 3.41 et 3.43 impliquent alors

$$\begin{aligned} \Psi_{\text{Bayes}} \left( \prod_{j \neq j_0} \mu_{Y_j X}^-(\cdot; z_{XY_j}^-), \sum_{(U_1, \dots, U_r)} [\otimes_i \mu_{U_i X}^+(z_{U_i X}^+; \cdot)] \otimes \Pi_X^{(U_1, \dots, U_r)} \right) = \dots \\ \dots \Psi_{\text{Bayes}} \left( f_{(Z_{XY_j}^+, X)}(z_X^+, \cdot; (z_{XY_j}^-)_{j \neq j_0}), \Pi_X^{Z_X^+}(z_X^+; \cdot) \right) \end{aligned} \quad (3.44)$$

La mesure  $\otimes_{j \neq j_0} \nu_{XY_j}^-$  étant positive et  $\sigma$ -finie, on peut utiliser la proposition C.2 pour calculer le membre de droite de l'équation 3.44. On obtient

$$\Psi_{\text{Bayes}} \left( \prod_{j \neq j_0} \mu_{Y_j X}^-(\cdot; z_{XY_j}^-), \sum_{(U_1, \dots, U_r)} [\otimes_i \mu_{U_i X}^+(z_{U_i X}^+; \cdot)] \otimes \Pi_X^{(U_1, \dots, U_r)} \right) = \dots \quad (3.45)$$

$$\begin{aligned} \dots \Pi_X^{(Z_X^+, (Z_{XY_j}^-)_{j \neq j_0})} (z_X^+, (z_{XY_j}^-)_{j \neq j_0}; \cdot) \\ = \Pi_X^{Z_{XY_{j_0}^+}}(z_{XY_{j_0}^+}; \cdot) \end{aligned} \quad (3.46)$$

$$= \mu_{XY_{j_0}^+}^+(z_{XY_{j_0}^+}; \cdot) \quad (3.47)$$

ce qui est l'égalité à démontrer.  $\square$

### 3.3.4 Calcul du message envoyé par un enfant à un parent.

#### Proposition 3.3

On suppose que le réseau bayésien  $G$  ne possède pas de cycle et on considère un sommet  $X$  de  $G$  n'appartenant pas à  $\mathcal{Z}$ . On note  $U_1, \dots, U_r$  les parents de  $X$  et  $Y_1, \dots, Y_s$  les enfants de  $X$ . Alors, pour tout  $i_0 \in \{1, 2, \dots, r\}$ , si  $X$  n'appartient pas à l'ensemble des observations  $\mathcal{Z}$ , le message  $\mu_{XU_{i_0}}^-$  envoyé par  $X$  à  $U_{i_0}$  vérifie

$$\begin{aligned} \forall z_{U_{i_0} X}^- \in \mathfrak{Z}_{U_{i_0} X}^- \quad \mu_{XU_{i_0}}^-(\cdot; z_{U_{i_0} X}^-) = \alpha \sum_X \left\{ \left[ \prod_j \mu_{Y_j X}^-(\cdot; z_{XY_j}^-) \right] \times \dots \right. \\ \left. \dots \times \sum_{(U_i)_{i \neq i_0}} [\otimes_{i \neq i_0} \mu_{U_i X}^+(z_{U_i X}^+; \cdot)] \otimes \Pi_X^{(U_1, \dots, U_r)} \right\} \end{aligned} \quad (3.48)$$

avec

$$\alpha = \prod_{i \neq i_0} f_{Z_{U_i X}^+}(z_{U_i X}^+) \quad (3.49)$$

où  $f_{Z_{U_i X}^+}$  désigne la densité de la loi  $F_{Z_{U_i X}^+}$  du vecteur  $Z_{U_i X}^+$  par rapport à la mesure  $\nu_{U_i X}^+$  définie par

$$\nu_{U_i X}^+ = \otimes_{\substack{l \in \mathcal{L} \\ X_l \in \mathcal{Z}_{U_i X}^+}} \nu_l \quad (3.50)$$

et avec les conventions suivantes :

- $z_{XY_j}^-$  désigne le projeté de  $z_{U_{i_0}X}^-$  sur l'espace  $\mathfrak{Z}_{XY_j}^-$
- $z_{U_{i_0}X}^+$  désigne le projeté de  $z_{U_{i_0}X}^-$  sur l'espace  $\mathfrak{Z}_{U_{i_0}X}^+$
- lorsque l'ensemble  $\mathfrak{Z}_{XY_j}^-$  est vide, le terme  $\mu_{Y_jX}^-(\cdot; z_{XY_j}^-)$  doit être remplacé par le terme  $\mu_{Y_jX}^-$  dans l'expression 3.48
- lorsque l'ensemble  $\mathfrak{Z}_{U_{i_0}X}^+$  est vide, le terme  $\mu_{U_{i_0}X}^+(z_{U_{i_0}X}^+; \cdot)$  doit être remplacé par le terme  $\mu_{U_{i_0}X}^+$  dans l'expression 3.48, et le terme  $f_{Z_{U_{i_0}X}^+}(z_{U_{i_0}X}^+)$  vaut 1 dans l'expression 3.49
- lorsque  $X$  n'a pas d'enfants et possède au moins un parent autre que  $U_{i_0}$ , l'expression 3.48 devient

$$\mu_{XU_{i_0}}^-(\cdot; z_{U_{i_0}X}^-) = \alpha \quad (3.51)$$

- lorsque  $X$  a  $U_{i_0}$  comme unique parent et possède au moins un enfant, l'expression 3.48 devient

$$\mu_{XU_{i_0}}^-(\cdot; z_{U_{i_0}X}^-) = \sum_X \left\{ \left[ \prod_j \mu_{Y_jX}^-(\cdot; z_{XY_j}^-) \right] \cdot \Pi_X^{U_{i_0}} \right\} \quad (3.52)$$

- lorsque  $X$  n'a ni enfant ni d'autre parent que  $U_{i_0}$ , l'expression 3.48 devient

$$\mu_{XU_{i_0}}^- = 1 \quad (3.53)$$

Dans le cas où  $X$  appartient à  $\mathfrak{Z}$  et possède au moins un parent autre que  $U_{i_0}$ , le message  $\mu_{XU_{i_0}}^-$  vérifie<sup>1</sup>

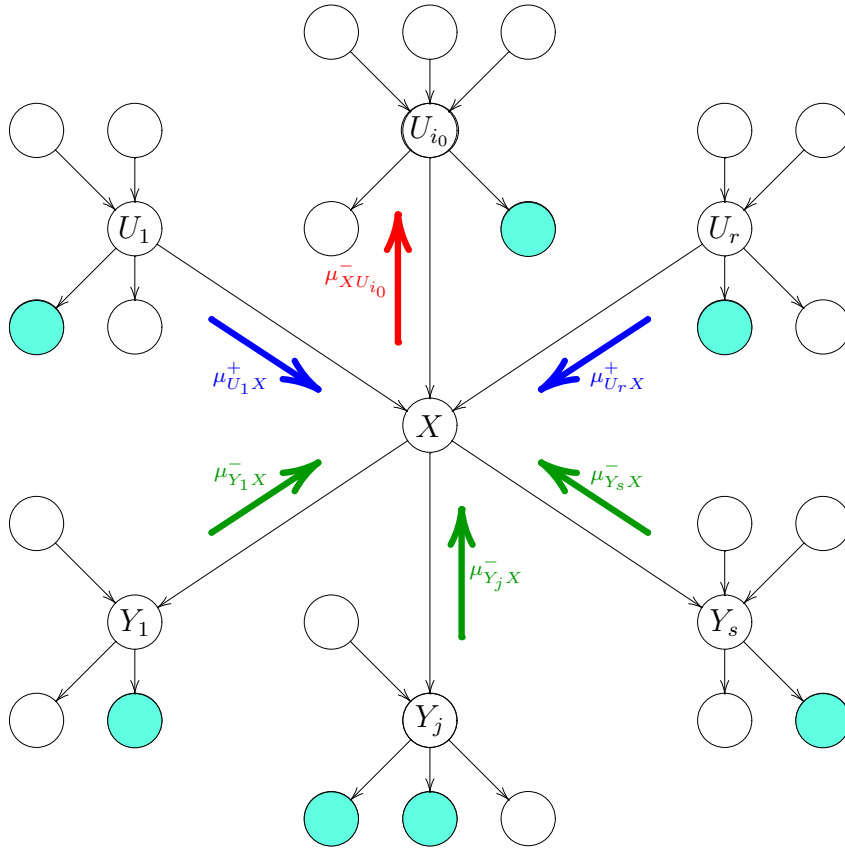
$$\mu_{XU_{i_0}}^-(\cdot; z_{U_{i_0}X}^-) = \alpha \frac{d \left[ \sum_{(U_i)_{i \neq i_0}} \left[ \bigotimes_{i \neq i_0} \mu_{U_iX}^+(z_{U_iX}^+; \cdot) \right] \otimes \Pi_X^{(U_1, \dots, U_r)} \right]}{d\nu_l}(x) \quad (3.54)$$

où  $x$  désigne le projeté de  $z_{U_{i_0}X}^-$  sur l'espace image de  $X$  et  $\nu_l$  est la mesure de domination sur l'espace image de  $X$ . Enfin, dans le cas où  $X$  appartient à  $\mathfrak{Z}$  et n'a pas d'autre parent que  $U_{i_0}$ , le message  $\mu_{XU_{i_0}}^-$  vérifie

$$\mu_{XU_{i_0}}^- = f_X^{U_{i_0}} \quad (3.55)$$

---

<sup>1</sup>On notera que lorsque  $X$  appartient à  $\mathfrak{Z}$ ,  $X$  n'a pas d'enfant car on a supposé que les éléments de  $\mathfrak{Z}$  n'ont pas d'enfant.


 FIG. 3.3 – Calcul de  $\mu_{X U_{i_0}}^-$ .

DÉMONSTRATION :

Dans cette démonstration, on supposera que  $X$  n'appartient pas à  $\mathcal{Z}$ , que  $X$  possède au moins un enfant et un parent autre que  $U_{i_0}$ , et que les ensembles  $\mathcal{Z}_{X Y_j}^-$  et  $\mathcal{Z}_{U_i X}^+$  sont non-vides. On vérifiera aisément que le traitement de ces différents cas particuliers ne pose aucune difficulté.

On convient de noter  $(\mathcal{U}_{i_0}, \mathcal{U}_{i_0})$  l'espace image de  $U_{i_0}$ . Pour tout  $u_{i_0} \in \mathcal{U}_{i_0}$  et pour tout  $z_{U_{i_0} X}^- \in \mathfrak{Z}_{U_{i_0} X}^-$ , on pose

$$f(u_{i_0}; z_{U_{i_0} X}^-) \stackrel{\text{déf}}{=} \alpha \sum_X \left\{ \left[ \prod_j \mu_{Y_j X}^-(\cdot; z_{X Y_j}^-) \right] \sum_{(U_i)_{i \neq i_0}} \left[ \bigotimes_{i \neq i_0} \mu_{U_i X}^+(z_{U_i X}^+; \cdot) \right] \otimes \Pi_X^{(U_1, \dots, U_r)} \right\} (u_{i_0}) \quad (3.56)$$

$$= \alpha \sum_X \left\{ \left[ \prod_j f_{Z_{X Y_j}^-}^X(\cdot; z_{X Y_j}^-) \right] \sum_{(U_i)_{i \neq i_0}} \left[ \bigotimes_{i \neq i_0} \Pi_{U_i}^{Z_{U_i X}^+}(z_{U_i X}^+; \cdot) \right] \otimes \Pi_X^{(U_1, \dots, U_r)} \right\} (u_{i_0}) \quad (3.57)$$

Alors, pour tout  $B \in \mathcal{Z}_{U_{i_0}^- X}$ , on a

$$\begin{aligned} \int_B f(u_{i_0}; z_{U_{i_0}^- X}^-) \nu_{U_{i_0}^- X}^-(dz_{U_{i_0}^- X}^-) &= \int \mathbf{1}_B(z_{U_{i_0}^- X}^-) \left[ \prod_j f_{Z_{XY_j}^-}^X(x; z_{XY_j}^-) \right] \times \dots \\ &\dots \times \Pi_X^{(U_1, \dots, U_r)}(u_1, \dots, u_r; dx) \left[ \prod_{i \neq i_0} f_{Z_{U_i X}^+}(z_{U_i X}^+) \Pi_{U_i}^{Z_{U_i X}^+}(z_{U_i X}^+; du_i) \right] \nu_{U_{i_0}^- X}^-(dz_{U_{i_0}^- X}^-) \end{aligned} \quad (3.58)$$

On définit les vecteurs  $V$ ,  $Z_{VX}^+$  et  $Z_X^-$  par

$$V \stackrel{\text{déf}}{=} (U_1, U_2, \dots, U_{i_0-1}, U_{i_0+1}, \dots, U_r) \quad (3.59)$$

$$Z_{VX}^+ \stackrel{\text{déf}}{=} (Z_{U_1 X}^+, Z_{U_2 X}^+, \dots, Z_{U_{i_0-1} X}^+, Z_{U_{i_0+1} X}^+, \dots, Z_{U_r X}^+) \quad (3.60)$$

$$Z_X^- \stackrel{\text{déf}}{=} (Z_{XY_1}^-, Z_{XY_2}^-, \dots, Z_{XY_s}^-) \quad (3.61)$$

Avec ces notations, on a alors

$$\int_B f(u_{i_0}; z_{U_{i_0}^- X}^-) \nu_{U_{i_0}^- X}^-(dz_{U_{i_0}^- X}^-) = \int \mathbf{1}_B(z_{U_{i_0}^- X}^-) f_{Z_X^-}^X(x; z_X^-) \Pi_X^{(U_{i_0}, V)}(u_{i_0}, v; dx) \times \dots \quad (3.62)$$

$$\begin{aligned} &\dots \times \Pi_{V^+ X}^{Z_{VX}^+}(z_{VX}^+; dv) F_{Z_{VX}^+}(dz_{VX}^+) \nu_X^-(dz_X^-) \\ &= \int \mathbf{1}_B(z_{U_{i_0}^- X}^-) \Pi_{Z_X^-}^X(x; dz_X^-) \Pi_X^{(U_{i_0}, V)}(u_{i_0}, v; dx) \times \dots \end{aligned} \quad (3.63)$$

$$\begin{aligned} &\dots \times \Pi_{V^+ X}^{Z_{VX}^+}(z_{VX}^+; dv) F_{Z_{VX}^+}(dz_{VX}^+) \\ &= \int \mathbf{1}_B(z_{U_{i_0}^- X}^-) \Pi_{Z_X^-}^X(x; dz_X^-) \Pi_X^{(U_{i_0}, V)}(u_{i_0}, v; dx) \times \dots \quad (3.64) \\ &\dots \times F_{(V, Z_{VX}^+)}(d(v, z_{VX}^+)) \end{aligned}$$

Or, on a

$$\Pi_X^{(U_{i_0}, V)}(u_{i_0}, v; \cdot) = \Pi_X^{(U_{i_0}, V, Z_{VX}^+)}(u_{i_0}, v, z_{VX}^+; \cdot) \quad (3.65)$$

et

$$F_{(V, Z_{VX}^+)} = \Pi_{(V, Z_{VX}^+)}^{U_{i_0}}(u_{i_0}; \cdot) \quad (3.66)$$

En reportant les égalités 3.65 et 3.66 dans l'équation 3.64 et en utilisant la proposition C.3, on obtient

$$\int_B f(u_{i_0}; z_{U_{i_0}^- X}^-) \nu_{U_{i_0}^- X}^-(dz_{U_{i_0}^- X}^-) = \int \mathbf{1}_B(z_{U_{i_0}^- X}^-) \Pi_{Z_X^-}^X(x; dz_X^-) \Pi_{(X, V, Z_{VX}^+)}^{U_{i_0}}(u_{i_0}; d(x, v, z_{VX}^+)) \quad (3.67)$$

$$= \int \mathbf{1}_B(z_{U_{i_0}^- X}^-) \Pi_{Z_X^-}^X(x; dz_X^-) \Pi_{(X, Z_{VX}^+)}^{U_{i_0}}(u_{i_0}; d(x, z_{VX}^+)) \quad (3.68)$$

Or, on a

$$\Pi_{Z_X^-}^X(x; \cdot) = \Pi_{Z_X^-}^{(U_{i_0}, X, Z_{VX}^+)}(u_{i_0}, x, z_{VX}^+; \cdot) \quad (3.69)$$

En reportant l'égalité 3.69 dans l'équation 3.68 et en utilisant la proposition C.3, on obtient

$$\int_B f(u_{i_0}; z_{\bar{U}_{i_0}X}^-) \nu_{\bar{U}_{i_0}X}^-(dz_{\bar{U}_{i_0}X}^-) = \int \mathbb{1}_B(z_{\bar{U}_{i_0}X}^-) \Pi_{(Z_{\bar{X}}, X, Z_{\bar{V}X}^+)}^{U_{i_0}}(u_{i_0}; d(z_{\bar{X}}^-, x, z_{\bar{V}X}^+)) \quad (3.70)$$

$$= \int \mathbb{1}_B(z_{\bar{U}_{i_0}X}^-) \Pi_{(Z_{\bar{X}}, Z_{\bar{V}X}^+)}^{U_{i_0}}(u_{i_0}; d(z_{\bar{X}}^-, z_{\bar{V}X}^+)) \quad (3.71)$$

$$= \Pi_{Z_{\bar{U}_{i_0}X}^-}^{U_{i_0}}(u_{i_0}; B) \quad (3.72)$$

Par conséquent, pour tout  $u_{i_0}$ , l'application  $f(u_{i_0}; \cdot)$  est la densité de la mesure  $\Pi_{Z_{\bar{U}_{i_0}X}^-}^{U_{i_0}}(u_{i_0}; \cdot)$  par rapport à la mesure  $\nu_{\bar{U}_{i_0}X}^-$ , ce qui prouve l'égalité 3.48.  $\square$

### 3.3.5 Définition de l'algorithme de propagation de croyance exacte

En considérant les propositions 3.2 et 3.3, on remarque que, dans un réseau bayésien sans cycle, il est possible de calculer tous les messages reçus par un sommet quelconque (n'appartenant pas à l'ensemble des observations) à partir des messages envoyés par les sommets de degré 1 du réseau bayésien<sup>2</sup> en appliquant de façon récursive ces deux propositions. Comme la proposition 3.1 permet de calculer la loi *a posteriori* de ce sommet à partir des messages reçus par ce sommet, on voit donc que l'ensemble des propositions 3.1, 3.2 et 3.3 permet de calculer la loi *a posteriori* de ce sommet à partir des messages envoyés par les sommets de degré 1. On convient alors d'appeler *algorithme de propagation de croyance exacte* l'algorithme ainsi obtenu et *équations de propagation de croyance* les équations 3.4, 3.37 et 3.48 qui sont utilisées par cet algorithme.

## 3.4 L'algorithme de propagation de croyance approchée

Dans le cas d'un réseau bayésien cyclique, il n'est plus possible de calculer exactement la loi *a posteriori* d'un sommet à l'aide de l'algorithme de propagation de croyance exacte étudié précédemment. Cependant, il a été constaté empiriquement [FM98, MMC98] que l'on peut obtenir dans certains cas une « bonne » approximation de cette loi *a posteriori* en appliquant itérativement les équations de propagation de croyance que nous avons vues précédemment. Dans cette partie, nous allons décrire en détail l'algorithme que l'on obtient en procédant ainsi et que nous conviendrons d'appeler *algorithme de propagation de croyance approchée*<sup>3</sup>. Pour cela, nous allons tout d'abord étendre la notion de message introduite au § 3.3.1 : les messages utilisés par l'algorithme de propagation de croyance approchée perdent la signification qu'ils avaient au § 3.3.1 et deviennent maintenant de simples intermédiaires de calcul. Nous décrirons ensuite l'application des équations de propagation de croyance pour calculer les messages échangés à

<sup>2</sup>Les sommets de degré 1 du réseau bayésien sont les sommets qui soit ont un unique parent et pas d'enfant, soit ont un unique enfant et pas de parent (cf la définition A.3).

<sup>3</sup>Dans la littérature, cet algorithme est également appelé *algorithme de propagation de croyance cyclique* (*loopy belief propagation algorithm*) [TJ02, Wei00]. Dans cette thèse, nous préférons utiliser le terme *algorithme de propagation de croyance approchée* pour souligner le fait que cet algorithme ne permet de calculer qu'une approximation de la loi *a posteriori* de chaque sommet.

chaque itération et obtenir une estimation de la loi *a posteriori* des sommets n'appartenant pas à l'ensemble des observations.

### 3.4.1 Messages

#### Définition 3.4

Soit  $(\Omega, \mathcal{F}, P)$  un espace de probabilité. Soit  $\mathcal{L}$  une partie finie de  $\mathbb{N}$ . Pour tout  $l \in \mathcal{L}$ , soit  $(E_l, \mathcal{E}_l)$  un espace mesurable et  $X_l$  un élément aléatoire de  $(\Omega, \mathcal{F})$  dans  $(E_l, \mathcal{E}_l)$ . On pose  $U = \{X_l\}_{l \in \mathcal{L}}$ . Soit  $G = (U, \mathbf{A})$  un réseau bayésien sur  $U$  pour la mesure de probabilité  $P$ . Soit  $Z$  une partie non-vide de  $U$  telle que les éléments de  $Z$  n'ont pas d'enfants dans  $G$ . Pour tout arc  $(X, Y)$  de  $G$ ,

1. on note  $(\mathfrak{X}, \mathcal{X})$  l'espace image de  $X$  et  $Z_{XY}^+$  l'ensemble des éléments de  $Z$  qui sont joints à  $Y$  par une chaîne  $(\alpha_1, \dots, \alpha_q)$  telle que  $\alpha_q = (X, Y)$ .
  - (a) Si  $Z_{XY}^+$  est non vide, on note  $(\mathfrak{Z}_{XY}^+, \mathcal{Z}_{XY}^+)$  le produit des espaces mesurables images des éléments de  $Z_{XY}^+$  et on convient alors d'appeler **message de X à Y** et de noter  $\mu_{XY}^+$  toute transition de l'espace mesurable  $(\mathfrak{Z}_{XY}^+, \mathcal{Z}_{XY}^+)$  dans l'espace mesurable  $(\mathfrak{X}, \mathcal{X})$ .
  - (b) Si  $Z_{XY}^+$  est vide, on convient d'appeler **message de X à Y** et de noter  $\mu_{XY}^+$  toute mesure de probabilité sur l'espace mesurable  $(\mathfrak{X}, \mathcal{X})$ .
2. on note  $Z_{XY}^-$  l'ensemble des éléments de  $Z$  qui sont joints à  $X$  par une chaîne  $(\alpha_1, \dots, \alpha_q)$  telle que  $\alpha_q = (X, Y)$ .
  - (a) Si  $Z_{XY}^-$  est non vide, on note  $(\mathfrak{Z}_{XY}^-, \mathcal{Z}_{XY}^-)$  le produit des espaces mesurables images des éléments de  $Z_{XY}^-$  et on convient alors d'appeler **message de Y à X** et de noter  $\mu_{YX}^-$  toute application mesurable de  $(\mathfrak{X}, \mathcal{X}) \otimes (\mathfrak{Z}_{XY}^-, \mathcal{Z}_{XY}^-)$  dans  $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ .
  - (b) Si  $Z_{XY}^-$  est vide, on convient d'appeler **message de Y à X** et de noter  $\mu_{YX}^-$  l'application mesurable constante de  $(\mathfrak{X}, \mathcal{X})$  dans  $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$  valant 1.

Comme précédemment, on convient d'appeler  $Z$  l'ensemble des **observations**,  $Z_{XY}^+$  l'ensemble des **observations situées en amont de Y et reliées à Y via X** et  $Z_{XY}^-$  l'ensemble des **observations situées en aval de X et reliées à X via Y**.

### 3.4.2 Description de l'algorithme

#### 3.4.2.1 Cadre de l'algorithme

On se place sous les hypothèses de la définition 3.4. On suppose également donnés, pour tout arc  $(X, Y)$  de  $G$ , un message initial  $\mu_{XY}^{+(0)}$  ainsi qu'un message initial  $\mu_{YX}^{-(0)}$  choisis de la façon suivante :

- dans le cas où  $X$  n'a pas de parents et possède uniquement  $Y$  comme enfant, on pose  $\mu_{XY}^{+(0)} = F_X$
- dans le cas où  $Y$  n'a pas d'enfant et possède uniquement  $X$  comme parent,
  - si  $Y \in Z$ , on suppose que, pour tout  $x \in \mathfrak{X}$ , la mesure  $\Pi_Y^X(x; \cdot)$  possède une densité  $f_Y^X(x; \cdot)$  par rapport à une mesure de domination  $\nu_l$  positive et  $\sigma$ -finie sur l'espace image de  $Y$  et que la fonction  $f_Y^X$  est mesurable, et on pose  $\mu_{YX}^{-(0)} = f_Y^X$
  - si  $Y \notin Z$ , on pose  $\mu_{YX}^{-(0)} = 1$

- dans tous les autres cas, les messages initiaux  $\mu_{XY}^{+(0)}$  et  $\mu_{YX}^{-(0)}$  sont choisis de façon arbitraire.

Enfin, on suppose donnée une liste  $\mathcal{L} = (c_1, c_2, \dots, c_p)$  de couples de sommets du réseau bayésien  $G$  telle que, pour tout  $k \in \{1, \dots, p\}$ ,

1. le couple  $c_k$  ou le couple  $\bar{c}_k$  est un arc du réseau bayésien  $G$ ,
2. lorsque  $c_k$  est un arc du réseau bayésien  $G$ , le premier sommet de  $c_k$  possède au moins un parent ou possède plus d'un enfant,
3. lorsque  $\bar{c}_k$  est un arc du réseau bayésien  $G$ , le premier sommet de  $c_k$  possède au moins un enfant ou possède plus d'un parent,
4. aucun couple n'apparaît plus d'une fois sur la liste  $\mathcal{L}$ . Autrement dit,

$$\forall k, k' \quad c_k = c_{k'} \Rightarrow k = k' \quad (3.73)$$

### Remarque 3.1

La liste  $\mathcal{L}$  servira au § 3.4.2.2 à spécifier l'ordre dans lequel les différents messages doivent être mis à jour à chaque itération : on calculera d'abord le message envoyé par le premier sommet du couple  $c_1$  au deuxième sommet du couple  $c_1$ , puis le message envoyé par le premier sommet du couple  $c_2$  au deuxième sommet du couple  $c_2$  et ainsi de suite. La condition 1 garantit que les  $c_k$  correspondent bien à des arcs de  $G$  car un sommet ne peut envoyer de message qu'à un parent ou à un enfant. Les conditions 2 et 3 permettent d'éviter le calcul à chaque itération des messages envoyés par les sommets de degré 1 au reste du réseau : ces messages restent constants d'une itération à l'autre. Enfin la condition 4 nous permet de simplifier la formulation de l'algorithme de propagation de croyance approchée. Il est possible de formuler l'algorithme de propagation de croyance approchée sans cette dernière hypothèse en utilisant des notations plus complexes.

#### 3.4.2.2 Passage de l'itération $n$ à l'itération $n + 1$

Soit  $n \in \mathbb{N}$ . On suppose que, pour tout arc  $X \rightarrow Y$  de  $G$ , les messages  $\mu_{XY}^{+(n)}$  et  $\mu_{YX}^{-(n)}$  ont été définis. Pour tout  $k \in \{1, \dots, p\}$ ,

- si  $c_k$  est un arc du réseau bayésien  $G$ , on note  $X$  le premier sommet du couple  $c_k$  ;  $U_1, \dots, U_r$  les parents de  $X$  ;  $Y_1, \dots, Y_s$  les enfants de  $X$  et  $Y_{j_0}$  le deuxième sommet du couple  $c_k$ . On définit alors le message  $\mu_{XY_{j_0}}^{+(n+1)}$  par

$$\begin{aligned} \forall z_{XY_{j_0}}^+ \in \mathfrak{Z}_{XY_{j_0}}^+ \quad \mu_{XY_{j_0}}^{+(n+1)}(z_{XY_{j_0}}^+; \cdot) = \Psi_{\text{Bayes}} \left( \prod_{j \neq j_0} \mu_{Y_j X}^-(\cdot; z_{XY_j}^-), \dots \right. \\ \left. \dots \sum_{(U_1, \dots, U_r)} \left[ \bigotimes_i \mu_{U_i X}^+(z_{U_i X}^+; \cdot) \right] \otimes \Pi_X^{(U_1, \dots, U_r)} \right) \quad (3.74) \end{aligned}$$

avec

$$\mu_{Y_j X}^- = \begin{cases} \mu_{Y_j X}^{-(n+1)} & \text{si } \exists k' \in \{1, \dots, k-1\} \quad c_{k'} = (Y_j, X) \\ \mu_{Y_j X}^{-(n)} & \text{sinon} \end{cases} \quad (3.75)$$

$$\mu_{U_i X}^+ = \begin{cases} \mu_{U_i X}^{+(n+1)} & \text{si } \exists k' \in \{1, \dots, k-1\} \quad c_{k'} = (U_i, X) \\ \mu_{U_i X}^{+(n)} & \text{sinon} \end{cases} \quad (3.76)$$



et avec les conventions suivantes :

- $z_{XY_j}^-$  désigne le projeté de  $z_{XY_{j_0}}^+$  sur l'espace  $\mathfrak{Z}_{XY_j}^-$
- $z_{U_iX}^+$  désigne le projeté de  $z_{XY_{j_0}}^+$  sur l'espace  $\mathfrak{Z}_{U_iX}^+$
- lorsque l'ensemble  $\mathfrak{Z}_{XY_j}^-$  est vide, le terme  $\mu_{Y_jX}^-(\cdot; z_{XY_j}^-)$  doit être remplacé par le terme  $\mu_{Y_jX}^-$  dans l'expression 3.74
- lorsque l'ensemble  $\mathfrak{Z}_{U_iX}^+$  est vide, le terme  $\mu_{U_iX}^+(z_{U_iX}^+; \cdot)$  doit être remplacé par le terme  $\mu_{U_iX}^+$  dans l'expression 3.74
- lorsque  $X$  a  $Y_{j_0}$  comme unique enfant, l'expression 3.74 devient

$$\mu_{XY_{j_0}}^{+(n+1)}(z_{XY_{j_0}}^+; \cdot) = \sum_{(U_1, \dots, U_r)} [\bigotimes_i \mu_{U_iX}^+(z_{U_iX}^+; \cdot)] \otimes \Pi_X^{(U_1, \dots, U_r)} \quad (3.77)$$

- lorsque  $X$  n'a pas de parent, l'expression 3.74 devient

$$\mu_{XY_{j_0}}^{+(n+1)}(z_{XY_{j_0}}^+; \cdot) = \Psi_{\text{Bayes}}\left(\prod_{j \neq j_0} \mu_{Y_jX}^-(\cdot; z_{XY_j}^-), F_X\right) \quad (3.78)$$

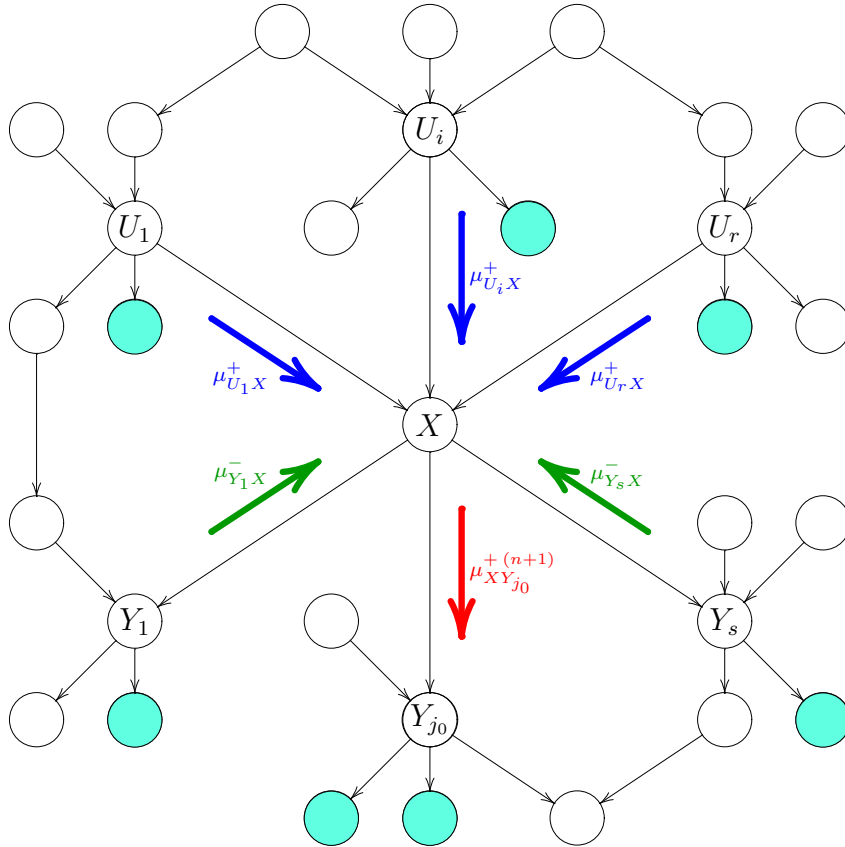


FIG. 3.4 – Calcul de  $\mu_{XY_{j_0}}^{+(n+1)}$ .

- si  $\bar{c}_k$  est un arc du réseau bayésien  $G$ , on note  $X$  le premier sommet du couple  $c_k$  ;  $U_1, \dots, U_r$  les parents de  $X$  ;  $Y_1, \dots, Y_s$  les enfants de  $X$  et  $U_{i_0}$  le deuxième sommet

du couple  $c_k$ . Si  $X \notin \mathbf{Z}$ , on définit alors le message  $\mu_{XU_{i_0}}^{-(n+1)}$  par

$$\forall z_{U_{i_0}X}^- \in \mathfrak{Z}_{U_{i_0}X}^- \quad \mu_{XU_{i_0}}^{-(n+1)}(\cdot; z_{U_{i_0}X}^-) = \sum_X \left[ \prod_j \mu_{Y_jX}^-(\cdot; z_{XY_j}^-) \times \dots \right. \\ \left. \dots \times \sum_{(U_i)_{i \neq i_0}} \left[ \bigotimes_{i \neq i_0} \mu_{U_iX}^+(z_{U_iX}^+; \cdot) \right] \otimes \Pi_X^{(U_1, \dots, U_r)} \right] \quad (3.79)$$

avec

$$\mu_{Y_jX}^- = \begin{cases} \mu_{Y_jX}^{-(n+1)} & \text{si } \exists k' \in \{1, \dots, k-1\} \quad c_{k'} = (Y_j, X) \\ \mu_{Y_jX}^{-(n)} & \text{sinon} \end{cases} \quad (3.80)$$

$$\mu_{U_iX}^+ = \begin{cases} \mu_{U_iX}^{+(n+1)} & \text{si } \exists k' \in \{1, \dots, k-1\} \quad c_{k'} = (U_i, X) \\ \mu_{U_iX}^{+(n)} & \text{sinon} \end{cases} \quad (3.81)$$

et avec les conventions suivantes :

- $z_{XY_j}^-$  désigne le projeté de  $z_{U_{i_0}X}^-$  sur l'espace  $\mathfrak{Z}_{XY_j}^-$
- $z_{U_iX}^+$  désigne le projeté de  $z_{U_{i_0}X}^-$  sur l'espace  $\mathfrak{Z}_{U_iX}^+$
- lorsque l'ensemble  $\mathbf{Z}_{XY_j}^-$  est vide, le terme  $\mu_{Y_jX}^-(\cdot; z_{XY_j}^-)$  doit être remplacé par le terme  $\mu_{Y_jX}^-$  dans l'expression 3.79
- lorsque l'ensemble  $\mathbf{Z}_{U_iX}^+$  est vide, le terme  $\mu_{U_iX}^+(z_{U_iX}^+; \cdot)$  doit être remplacé par le terme  $\mu_{U_iX}^+$  dans l'expression 3.79
- lorsque  $X$  n'a pas d'enfants, l'expression 3.79 devient

$$\mu_{XU_{i_0}}^{-(n+1)}(\cdot; z_{U_{i_0}X}^-) = 1 \quad (3.82)$$

- lorsque  $X$  a  $U_{i_0}$  comme unique parent, l'expression 3.79 devient

$$\mu_{XU_{i_0}}^{-(n+1)}(\cdot; z_{U_{i_0}X}^-) = \sum_X \left[ \prod_j \mu_{Y_jX}^-(\cdot; z_{XY_j}^-) \cdot \Pi_X^{U_{i_0}} \right] \quad (3.83)$$

Dans le cas où  $X$  appartient à l'ensemble des observations  $\mathbf{Z}$ , on définit le message  $\mu_{XU_{i_0}}^{-(n+1)}$  par

$$\mu_{XU_{i_0}}^-(\cdot; z_{U_{i_0}X}^-) = \frac{d \left[ \sum_{(U_i)_{i \neq i_0}} \left[ \bigotimes_{i \neq i_0} \mu_{U_iX}^+(z_{U_iX}^+; \cdot) \right] \otimes \Pi_X^{(U_1, \dots, U_r)} \right]}{d\nu_l}(x) \quad (3.84)$$

où  $x$  désigne le projeté de  $z_{U_{i_0}X}^-$  sur l'espace image de  $X$  et  $\nu_l$  est la mesure de domination sur l'espace image de  $X$ .

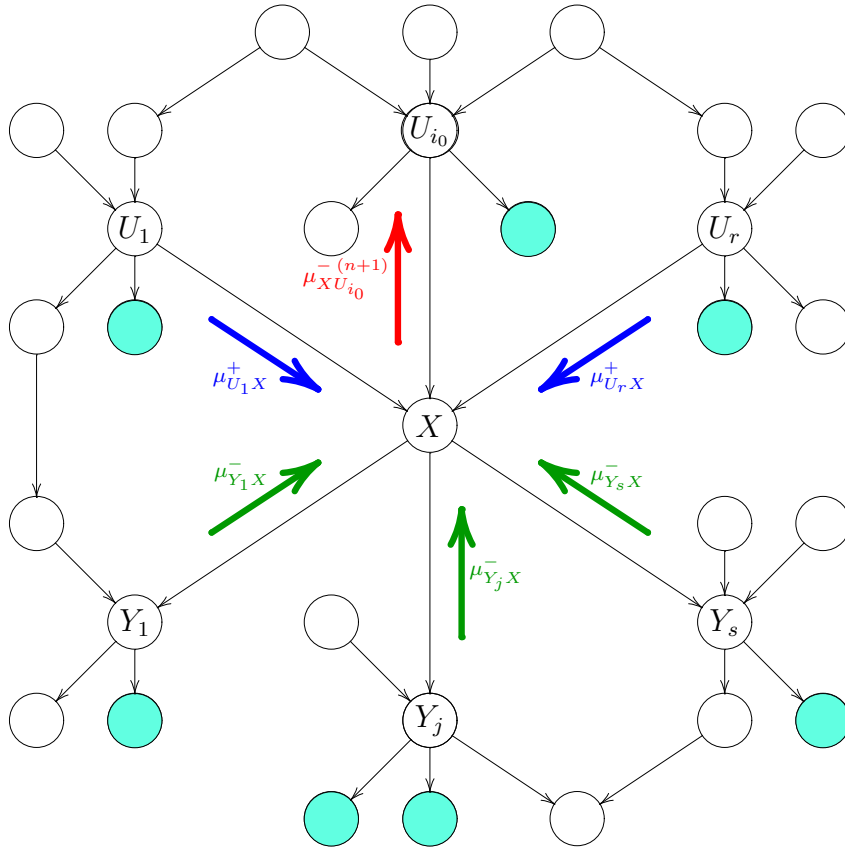


FIG. 3.5 – Calcul de  $\mu_{XU_{i_0}}^{-(n+1)}$ .

D'autre part, pour tout arc  $X \rightarrow Y$  de  $G$  tel que le couple  $(X, Y)$  n'apparaît pas sur la liste  $\mathcal{L}$ , on pose

$$\mu_{XY}^{+(n+1)} = \mu_{XY}^{+(0)} \quad (3.85)$$

et, pour tout arc  $X \rightarrow Y$  de  $G$  tel que le couple  $(Y, X)$  n'apparaît pas sur la liste  $\mathcal{L}$ , on pose

$$\mu_{YX}^{-(n+1)} = \mu_{YX}^{-(0)} \quad (3.86)$$

### Remarque 3.2

Au § 3.4.2.1, nous n'avons pas donné d'indication sur la façon de choisir la liste  $\mathcal{L}$  qui spécifie l'ordre dans lequel les messages sont mis à jour à chaque itération. Dans la littérature, on trouve plusieurs ordonnancements possibles. Ainsi, Kschischang et Frey [KF98] décrivent deux types d'ordonnement appelés respectivement *two-way schedule* et *flooding schedule*. Dans le premier cas, chaque sommet ne met à jour le message destiné à l'un de ses parents ou enfants que lorsqu'il a reçu un nouveau message de la part de tous ses autres parents et enfants. Dans le deuxième cas, chaque sommet met à jour le message destiné à l'un de ses parents ou enfants dès qu'il a reçu un nouveau message de la part de l'un de ses autres parents ou enfants.

Dans ce dernier cas, on peut être amené à calculer plusieurs fois au cours d'une même itération le message échangé entre deux sommets : la liste  $\mathcal{L}$  ne vérifie plus la condition 4 p. 47 et on doit alors utiliser des notations plus complexes pour formuler l'algorithme de propagation de croyance approchée.

### 3.4.2.3 Estimation des lois *a posteriori* à l'itération $n$ .

On suppose que tous les messages correspondant à l'itération  $n$  ont été calculés. Soit  $X$  un sommet de  $G$  n'appartenant pas à  $\mathcal{Z}$ . On note  $U_1, \dots, U_r$  les parents de  $X$  et  $Y_1, \dots, Y_s$  les enfants de  $X$ . L'estimation à l'itération  $n$  de la loi *a posteriori* de  $X$  étant donnée les observations  $\mathcal{Z}$  est donnée par

$$\forall z \in \mathfrak{Z} \quad \widehat{\Pi}_X^{\mathcal{Z}(n)}(z; \cdot) = \Psi_{\text{Bayes}} \left( \prod_j \mu_{Y_j X}^-(\cdot; z_{XY_j}^-), \dots \right. \\ \left. \dots \sum_{(U_1, \dots, U_r)} \left[ \bigotimes_i \mu_{U_i X}^+(z_{U_i X}^+; \cdot) \right] \otimes \Pi_X^{(U_1, \dots, U_r)} \right) \quad (3.87)$$

avec les conventions suivantes

- $z_{XY_j}^-$  désigne le projeté de  $z$  sur l'espace  $\mathfrak{Z}_{XY_j}^-$
- $z_{U_i X}^+$  désigne le projeté de  $z$  sur l'espace  $\mathfrak{Z}_{U_i X}^+$
- lorsque l'ensemble  $\mathcal{Z}_{XY_j}^-$  est vide, le terme  $\mu_{Y_j X}^-(\cdot; z_{XY_j}^-)$  doit être remplacé par le terme  $\mu_{Y_j X}^-(n)$  dans l'expression 3.87
- lorsque l'ensemble  $\mathcal{Z}_{U_i X}^+$  est vide, le terme  $\mu_{U_i X}^+(z_{U_i X}^+; \cdot)$  doit être remplacé par le terme  $\mu_{U_i X}^+(n)$  dans l'expression 3.87
- lorsque  $X$  n'a pas de parent, l'expression 3.87 devient

$$\widehat{\Pi}_X^{\mathcal{Z}(n)}(z; \cdot) = \Psi_{\text{Bayes}} \left( \prod_j \mu_{Y_j X}^-(\cdot; z_{XY_j}^-), F_X \right) \quad (3.88)$$

- lorsque  $X$  n'a pas d'enfants, l'expression 3.87 devient

$$\widehat{\Pi}_X^{\mathcal{Z}(n)}(z; \cdot) = \sum_{(U_1, \dots, U_r)} \left[ \bigotimes_i \mu_{U_i X}^+(z_{U_i X}^+; \cdot) \right] \otimes \Pi_X^{(U_1, \dots, U_r)} \quad (3.89)$$

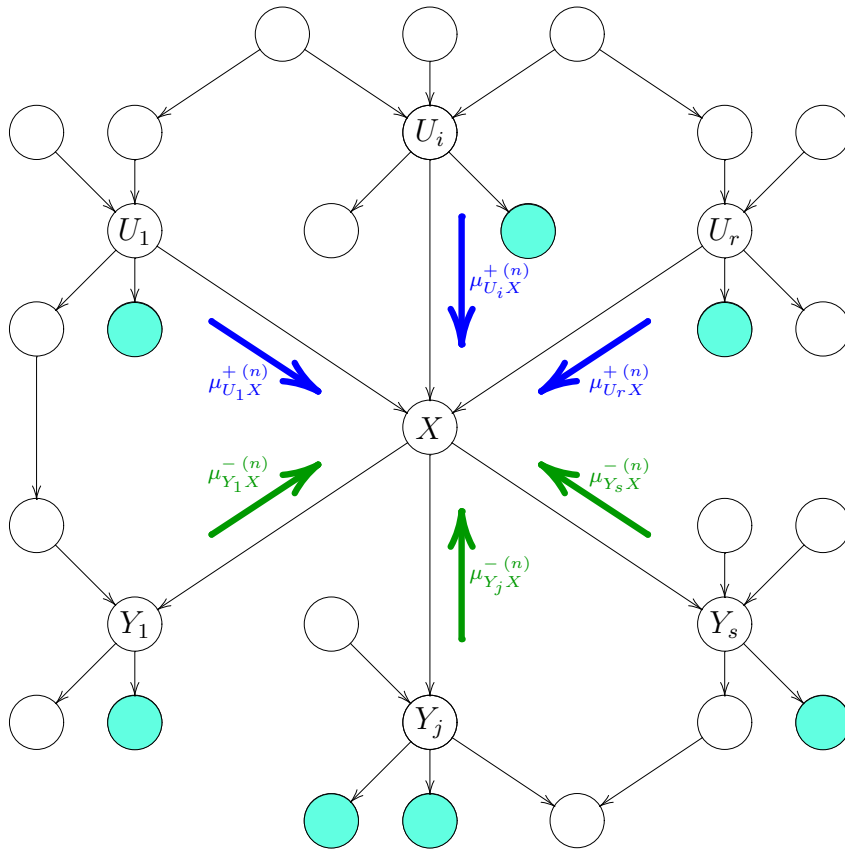


FIG. 3.6 – Calcul de  $\widehat{\Pi}_X^Z^{(n)}$ .

#### 3.4.2.4 Commentaires sur la convergence de l'algorithme

Comme on a pu le voir, l'algorithme de propagation de croyance approchée est un algorithme itératif. En notant  $\vec{\mu}^{(n)}$  les messages à l'itération  $n$ , on constate que les messages à l'itération  $n + 1$  sont calculés à partir des messages à l'itération  $n$ , autrement dit on a

$$\vec{\mu}^{(n+1)} = \Phi(\vec{\mu}^{(n)}) \quad (3.90)$$

où l'opérateur  $\Phi$  correspond aux calculs décrits au § 3.4.2.2. Lorsque l'algorithme de propagation de croyance converge, les messages obtenus correspondent donc à un point fixe de l'opérateur  $\Phi$ .

Lorsque le réseau bayésien  $G$  possède un unique cycle, les messages échangés entre les sommets de ce cycle peuvent être calculés à l'aide d'un opérateur linéaire  $\tilde{\Phi}$ . Les points fixes de l'algorithme correspondent alors aux vecteurs propres de  $\tilde{\Phi}$  associés à une valeur propre strictement positive et appliquer l'algorithme de propagation de croyance au réseau bayésien  $G$  revient en fait à rechercher un vecteur propre de l'opérateur  $\tilde{\Phi}$  par la méthode de la puissance [Wei00].

Lorsque le réseau bayésien  $G$  possède plusieurs cycles, il n'est plus possible de se ramener à l'étude d'un opérateur linéaire. La convergence de l'algorithme a cependant été étudiée dans la littérature dans un certain nombre de cas particuliers importants : variables aléatoires

gaussiennes [WF01], topologie particulière correspondant à celle des turbo-codes [Ric00]. . . La convergence de l'algorithme dans le cas général a également été étudiée à l'aide d'outils issus de la théorie des mesures de Gibbs [TJ02, YFW01, YFW].

### 3.5 Conclusion

Dans ce chapitre, nous avons présenté les algorithmes de propagation de croyance exacte et de propagation de croyance approchée. L'algorithme de propagation de croyance exacte s'applique à un réseau bayésien sans cycle. Il permet de calculer la loi *a posteriori* d'un sommet à l'aide de messages qui sont propagés entre les sommets du réseau bayésien grâce aux équations de propagation de croyance. L'algorithme de propagation de croyance approchée s'applique à un réseau bayésien cyclique. Il permet de calculer de façon itérative une approximation de la loi *a posteriori* d'un sommet en utilisant également les équations de propagation de croyance. Comme ces deux algorithmes utilisent tous les deux les équations de propagation de croyance, il est d'usage de les considérer comme deux variantes d'un unique algorithme : l'algorithme de propagation de croyance.

On notera que notre présentation de l'algorithme de propagation de croyance est différente de la présentation habituelle de cet algorithme dans la littérature : dans [Pea82, Pea86, Pea88, MMC98], les messages échangés entre les sommets sont définis comme étant des vecteurs de coefficients et on y appelle équations de propagation de croyance les équations qui opèrent sur ces coefficients. Nous avons choisi ici une présentation plus générale et abstraite de cet algorithme. Dans notre approche, les messages correspondent à des transitions entre espaces mesurables et les équations de propagation de croyance opèrent sur ces transitions en utilisant notamment l'opérateur de Bayes. Cette approche est plus adaptée à la définition très générale des réseaux bayésiens que nous avons donnée dans le chapitre précédent car elle n'exige pas que la loi *a posteriori* des sommets du réseau ait une forme particulière. Naturellement, dans le cas de réseaux bayésiens dont tous les sommets sont des variables aléatoires discrètes, les équations de propagation de croyance telle que nous les avons définies coïncident avec les équations figurant dans la littérature. Dans d'autres cas, nos équations de propagation de croyance peuvent conduire à des calculs insolubles analytiquement ou possédant une complexité trop élevée en pratique, mais il est alors possible de recourir à des méthodes d'approximation numérique pour remplacer le calcul exact des messages par des calculs approchés.

A l'issue de ce chapitre, nous disposons maintenant d'un ensemble de techniques que nous pouvons appliquer au problème de l'égalisation. C'est l'objet du prochain chapitre.



# Chapitre 4

## Application des réseaux bayésiens à l'égalisation

### 4.1 Introduction

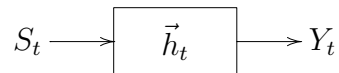
Dans les chapitres précédents, nous avons vu qu'il est possible de calculer la loi *a posteriori* d'un élément aléatoire en appliquant l'algorithme de propagation de croyance à un réseau bayésien représentant les relations d'indépendance conditionnelle du système étudié.

Nous allons maintenant appliquer les techniques que nous venons d'étudier au problème de l'égalisation. Nous allons nous intéresser à trois situations. Nous considérerons tout d'abord l'égalisation d'un canal sélectif en fréquence dont la réponse impulsionnelle est supposée connue. Nous examinerons ensuite l'égalisation adaptative d'un canal évanescent sélectif en fréquence dont la réponse impulsionnelle doit être estimée conjointement. Enfin nous étudierons l'égalisation adaptative d'un canal évanescent non-sélectif en fréquence dont la réponse impulsionnelle doit être estimée conjointement. Pour chacune de ces situations, nous construirons le réseau bayésien correspondant au système étudié, nous déterminerons les transitions entre parents et enfants dans le réseau bayésien obtenu et nous appliquerons l'algorithme de propagation de croyance à ce réseau bayésien.

### 4.2 L'égalisation d'un canal sélectif en fréquence

#### 4.2.1 Modèle du système

On suppose que des symboles complexes  $S_t$  appartenant à un alphabet  $\mathcal{A}$  sont transmis à travers un canal de réponse impulsionnelle complexe  $\vec{h}_t$  supposée connue dont on observe la sortie complexe  $Y_t$ .



Toutes les grandeurs sont échantillonnées au rythme symbole. On suppose que la sortie complexe  $Y_t$  du canal vérifie, pour tout  $t \in \mathbb{N}$ ,

$$Y_t = \vec{h}_t^T \cdot \vec{L}_t + V_t \quad (4.1)$$



où  $V_t$  suit la loi gaussienne  $\mathcal{N}_C(0, \mathcal{N}_0)$  et le vecteur  $\vec{L}_t$  est défini par

$$\vec{L}_t = \begin{cases} \begin{bmatrix} S_t & S_{t-1} & \dots & S_{t-L} \end{bmatrix}^T & \text{si } t \geq L \\ \begin{bmatrix} S_t & S_{t-1} & \dots & S_0 & 0 & \dots & 0 \end{bmatrix}^T & \text{sinon} \end{cases} \quad (4.2)$$

Enfin, on suppose que les variables aléatoires  $S_0, V_0, S_1, V_1, \dots, S_t, V_t$  sont indépendantes et que  $S_t$  suit une loi  $F_{S_t}$  sur l'alphabet  $\mathcal{A}$ . Pour tout  $t_0 \in \mathbb{N}$ , on recherche l'estimateur du maximum *a posteriori* de  $S_{t_0}$  étant données les observations  $Y_0, Y_1, \dots, Y_{t_0}$ .

## 4.2.2 Réseau bayésien

En procédant comme indiqué au § 2.3, on obtient le réseau bayésien représenté à la figure 4.1.

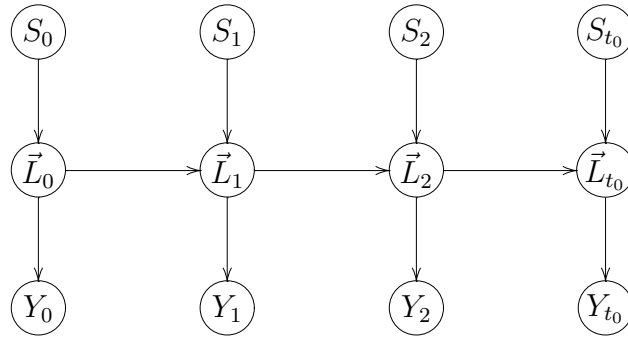


FIG. 4.1 – Réseau bayésien correspondant au problème de l'égalisation d'un canal sélectif en fréquence de réponse impulsionnelle connue. On a représenté le cas  $t_0 = 3$ .

## 4.2.3 Transitions entre parents et enfants

D'après l'équation 4.2, la transition  $\Pi_{\vec{L}_0}^{S_0}$  peut être définie par

$$\Pi_{\vec{L}_0}^{S_0}(s_0; \cdot) = \delta_{s_0 \vec{C}} \quad (4.3)$$

où  $\vec{C}$  est la matrice de taille  $(L+1) \times 1$  suivante

$$\vec{C} = [1 \ 0 \ 0 \ \dots \ 0]^T, \quad (4.4)$$

et, pour tout  $t > 0$ , la transition  $\Pi_{\vec{L}_t}^{(S_t, \vec{L}_{t-1})}$  peut être définie par

$$\Pi_{\vec{L}_t}^{(S_t, \vec{L}_{t-1})}(s_t, \vec{l}_{t-1}; \cdot) = \delta_{s_t \vec{C} + M \vec{l}_{t-1}} \quad (4.5)$$

où  $M$  est la matrice de taille  $(L + 1) \times (L + 1)$  suivante

$$M = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \quad (4.6)$$

D'après l'équation 4.1, pour tout  $t \in \mathbb{N}$ , la transition  $\Pi_{Y_t}^{\vec{L}_t}$  peut être définie par

$$\Pi_{Y_t}^{\vec{L}_t}(\vec{l}_t; \cdot) = \mathcal{N}_C(\vec{h}_t^T \cdot \vec{l}_t, \mathcal{N}_0) \quad (4.7)$$

$$= g(\vec{h}_t^T \cdot \vec{l}_t, \mathcal{N}_0) \cdot \lambda \quad (4.8)$$

où  $\lambda$  désigne la mesure de Lebesgue sur  $\mathbb{C}$  et  $g(\vec{h}_t^T \cdot \vec{l}_t, \mathcal{N}_0)$  désigne la densité de  $\mathcal{N}_C(\vec{h}_t^T \cdot \vec{l}_t, \mathcal{N}_0)$  par rapport à  $\lambda$ .

#### 4.2.4 Application de l'algorithme de propagation de croyance

Comme le réseaux bayésien représenté à la figure 4.1 ne possède pas de cycle, on peut appliquer l'algorithme de propagation de croyance exacte pour déterminer l'estimateur du maximum *a posteriori* de  $S_{t_0}$  étant données les observations  $Y_0, Y_1, \dots, Y_{t_0}$ . Les messages nécessaires au calcul de la loi *a posteriori* de  $S_{t_0}$  ont été représentés à la figure 4.2.

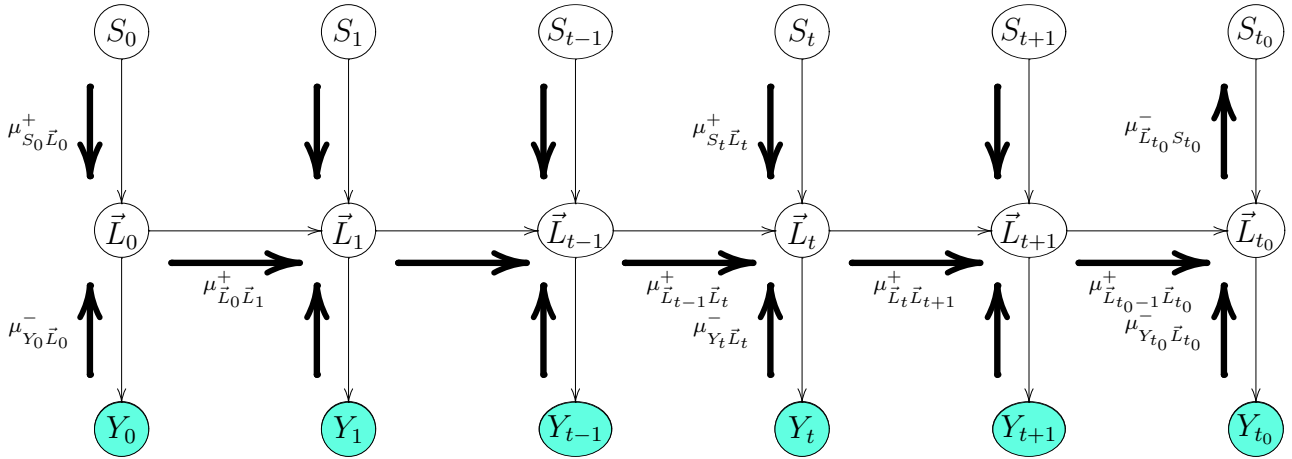


FIG. 4.2 – Application de l'algorithme de propagation de croyance au réseau bayésien correspondant au problème de l'égalisation d'un canal sélectif en fréquence de réponse impulsionnelle connue. Les variables aléatoires observées sont représentées en vert.

Pour tout  $t > 0$ , le message envoyé par  $\vec{L}_t$  à  $\vec{L}_{t+1}$  vérifie

$$\mu_{\vec{L}_t \vec{L}_{t+1}}^+(y_{0:t}; \cdot) = \Psi_{\text{Bayes}} \left( \mu_{Y_t \vec{L}_t}^-(\cdot; y_t), \sum_{\substack{S_t \\ \vec{L}_{t-1}}} (F_{S_t} \otimes \mu_{\vec{L}_{t-1} \vec{L}_t}^+(y_{0:t-1}; \cdot)) \otimes \Pi_{\vec{L}_t}^{(S_t, \vec{L}_{t-1})} \right) \quad (4.9)$$

En utilisant les équations 4.5 et 4.8, on en déduit, pour tout  $\vec{l}_t = [s_t \ s_{t-1} \ \dots \ s_{t-L}]^T$ ,

$$\begin{aligned} \mu_{\vec{L}_t \vec{L}_{t+1}}^+ (y_{0:t}; \{\vec{l}_t\}) &\propto g(\vec{h}_t^T \cdot \vec{l}_t, \mathcal{N}_0)(y_t) \times \dots \\ &\dots \times F_{S_t}(s_t) \sum_{s_{t-L-1}} \mu_{\vec{L}_{t-1} \vec{L}_t}^+ \left( y_{0:t-1}; \left\{ [s_{t-1} \ s_{t-2} \ \dots \ s_{t-L-1}]^T \right\} \right) \end{aligned} \quad (4.10)$$

En posant, pour tout  $t \geq 0$ ,

$$\alpha_t(s_t, s_{t-1}, \dots, s_{t-L+1}) \stackrel{\text{déf}}{=} \sum_{s_{t-L}} \mu_{\vec{L}_t \vec{L}_{t+1}}^+ \left( y_{0:t}; \left\{ [s_t \ s_{t-1} \ \dots \ s_{t-L}]^T \right\} \right) \quad (4.11)$$

la relation 4.10 implique, pour tout  $t > 0$ ,

$$\begin{aligned} \alpha_t(s_t, s_{t-1}, \dots, s_{t-L+1}) &\propto F_{S_t}(s_t) \sum_{s_{t-L}} g(\vec{h}_t^T \cdot [s_t \ s_{t-1} \ \dots \ s_{t-L}]^T, \mathcal{N}_0)(y_t) \times \dots \\ &\dots \times \alpha_{t-1}(s_{t-1}, s_{t-2}, \dots, s_{t-L}) \end{aligned} \quad (4.12)$$

L'initialisation de la récursion 4.12 se fait en remarquant que le message envoyé par  $\vec{L}_0$  à  $\vec{L}_1$  vérifie

$$\mu_{\vec{L}_0 \vec{L}_1}^+ (y_0; \cdot) = \Psi_{\text{Bayes}} \left( \mu_{Y_0 \vec{L}_0}^- (\cdot; y_0), \sum_{S_0} F_{S_0} \otimes \Pi_{\vec{L}_0}^{S_0} \right) \quad (4.13)$$

En utilisant les équations 4.3, 4.8 et 4.11, on en déduit

$$\begin{aligned} \alpha_0(s_0, s_{-1}, \dots, s_{-L+1}) &\propto F_{S_0}(s_0) \sum_{s_{-L}} g(\vec{h}_0^T \cdot [s_0 \ s_{-1} \ \dots \ s_{-L}]^T, \mathcal{N}_0)(y_0) \times \dots \\ &\dots \times \alpha_{-1}(s_{-1}, s_{-2}, \dots, s_{-L}) \end{aligned} \quad (4.14)$$

avec

$$\alpha_{-1}(s_{-1}, s_{-2}, \dots, s_{-L}) \stackrel{\text{déf}}{=} \begin{cases} 1 & \text{si } s_{-1} = s_{-2} = \dots = s_{-L} = 0 \\ 0 & \text{sinon} \end{cases} \quad (4.15)$$

Le message envoyé par  $\vec{L}_{t_0}$  à  $S_{t_0}$  vérifie

$$\mu_{\vec{L}_{t_0} S_{t_0}}^- (\cdot; y_{0:t_0}) = \sum_{\vec{L}_{t_0}} \mu_{Y_{t_0} \vec{L}_{t_0}}^- (\cdot; y_{t_0}) \sum_{\vec{L}_{t_0-1}} \mu_{\vec{L}_{t_0-1} \vec{L}_{t_0}}^+ (y_{0:t_0-1}; \cdot) \otimes \Pi_{\vec{L}_{t_0}}^{(S_{t_0}, \vec{L}_{t_0-1})} \quad (4.16)$$

En utilisant les équations 4.5 et 4.8, on en déduit pour tout  $s_{t_0}$

$$\begin{aligned} \mu_{\vec{L}_{t_0} S_{t_0}}^- (s_{t_0}; y_{0:t_0}) &= \sum_{(s_t)_{t_0-L-1 \leq t < t_0}} g(\vec{h}_{t_0}^T \cdot [s_{t_0} \ s_{t_0-1} \ \dots \ s_{t_0-L}]^T, \mathcal{N}_0)(y_{t_0}) \dots \\ &\dots \mu_{\vec{L}_{t_0-1} \vec{L}_{t_0}}^+ \left( y_{0:t_0-1}; \left\{ [s_{t_0-1} \ s_{t_0-2} \ \dots \ s_{t_0-L-1}]^T \right\} \right) \end{aligned} \quad (4.17)$$

On obtient alors la loi *a posteriori* de  $S_{t_0}$

$$\Pi_{S_{t_0}}^{Y_{0:t_0}}(y_{0:t_0}; \cdot) = \Psi_{\text{Bayes}}\left(\mu_{\vec{L}_{t_0} S_{t_0}}^-(\cdot; y_{0:t_0}), F_{S_{t_0}}\right) \quad (4.18)$$

En utilisant l'équation 4.17, on en déduit

$$\begin{aligned} \Pi_{S_{t_0}}^{Y_{0:t_0}}(y_{0:t_0}; \{s_{t_0}\}) &\propto F_{S_{t_0}}(s_{t_0}) \sum_{(s_t)_{t_0-L-1 \leq t < t_0}} g(\vec{h}_{t_0}^T [s_{t_0} \ s_{t_0-1} \ \dots \ s_{t_0-L}]^T, \mathcal{N}_0)(y_{t_0}) \dots \\ &\dots \mu_{\vec{L}_{t_0-1} \vec{L}_{t_0}}^+(y_{0:t_0-1}; \{[s_{t_0-1} \ s_{t_0-2} \ \dots \ s_{t_0-L-1}]^T\}) \end{aligned} \quad (4.19)$$

Avec la notation 4.11, l'équation 4.19 peut s'écrire

$$\begin{aligned} \Pi_{S_{t_0}}^{Y_{0:t_0}}(y_{0:t_0}; \{s_{t_0}\}) &\propto F_{S_{t_0}}(s_{t_0}) \sum_{(s_t)_{t_0-L \leq t < t_0}} g(\vec{h}_{t_0}^T [s_{t_0} \ s_{t_0-1} \ \dots \ s_{t_0-L}]^T, \mathcal{N}_0)(y_{t_0}) \dots \\ &\dots \alpha_{t_0-1}(s_{t_0-1}, s_{t_0-2}, \dots, s_{t_0-L}) \end{aligned} \quad (4.20)$$

En utilisant l'équation 4.12, on en déduit

$$\Pi_{S_{t_0}}^{Y_{0:t_0}}(y_{0:t_0}; \{s_{t_0}\}) \propto \sum_{(s_t)_{t_0-L+1 \leq t \leq t_0-1}} \alpha_{t_0}(s_{t_0}, s_{t_0-1}, \dots, s_{t_0-L+1}) \quad (4.21)$$

On obtient alors l'estimateur du maximum *a posteriori* de  $S_{t_0}$  étant données les observations  $Y_0, Y_1, \dots, Y_{t_0}$  en maximisant l'expression 4.21 par rapport à  $s_{t_0}$ .

### 4.2.5 Conclusion

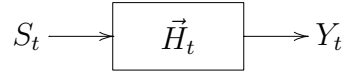
On remarque que les équations 4.12 et 4.21 sont identiques à celles que l'on obtiendrait en appliquant l'algorithme BCJR [BCJR74] pour déterminer la loi *a posteriori* de  $S_{t_0}$  étant données les observations  $Y_0, Y_1, \dots, Y_{t_0}$ . Ce résultat est cohérent avec les observations de Wiberg [Wib96] et de McEliece *et al.* [MMC98] : Wiberg a remarqué que l'on retrouve l'algorithme BCJR en appliquant l'algorithme somme-produit à un graphe de Tanner de structure équivalente à celle du réseau bayésien de la figure 4.1 et McEliece *et al.* ont fait une remarque similaire en utilisant l'algorithme de propagation de croyance pour décoder un code de convolution du type *tail-biting*.

D'autre part, la complexité de cet algorithme est  $O(t_0 \text{card}(\mathcal{A})^{L+1})$ . En pratique, cet algorithme ne peut donc être utilisé qu'avec des alphabets de faible taille et des canaux de réponse impulsionnelle courte.

## 4.3 L'égalisation adaptative d'un canal évanescent sélectif en fréquence

### 4.3.1 Modèle du système

On suppose que des symboles complexes  $S_t$  appartenant à un alphabet  $\mathcal{A}$  sont transmis à travers un canal de réponse impulsionnelle  $\vec{H}_t$  dont on observe la sortie complexe  $Y_t$ .



Toutes les grandeurs sont supposées échantillonnées au rythme symbole. On suppose que la sortie complexe  $Y_t$  du canal vérifie, pour tout  $t \in \mathbb{N}$ ,

$$Y_t = \vec{H}_t^T \cdot \vec{L}_t + V_t \quad (4.22)$$

où  $V_t$  suit la loi gaussienne  $\mathcal{N}_C(0, \mathcal{N}_0)$ , le vecteur  $\vec{L}_t$  est défini par

$$\vec{L}_t = \begin{cases} \left[ S_t & S_{t-1} & \dots & S_{t-L} \right]^T & \text{si } t \geq L \\ \left[ S_t & S_{t-1} & \dots & S_0 & 0 & \dots & 0 \right]^T & \text{sinon} \end{cases} \quad (4.23)$$

et le vecteur  $\vec{H}_t$  évolue suivant un processus autorégressif vectoriel d'ordre 1 défini par

$$\vec{H}_t = \begin{cases} \Lambda \vec{H}_{t-1} + \vec{W}_t & \text{si } t > 0 \\ \vec{\eta}_0 + \vec{W}_0 & \text{si } t = 0 \end{cases} \quad (4.24)$$

où  $\vec{W}_t$  suit la loi gaussienne  $\mathcal{N}_C(\vec{0}, R_G)$ ,  $\Lambda$  est une matrice complexe supposée connue et  $\vec{\eta}_0$  un vecteur complexe supposé connu. Enfin, on suppose que les variables aléatoires  $S_0, \vec{W}_0, V_0, S_1, \vec{W}_1, V_1, \dots, S_t, \vec{W}_t, V_t$  sont indépendantes et que  $S_t$  suit une loi  $F_{S_t}$  sur l'alphabet  $\mathcal{A}$ . Pour tout  $t_0 \in \mathbb{N}$ , on recherche l'estimateur du maximum *a posteriori* de  $S_{t_0}$  étant données les observations  $Y_0, Y_1, \dots, Y_{t_0}$ .

### 4.3.2 Réseau bayésien

En procédant comme indiqué au § 2.3, on obtient le réseau bayésien représenté à la figure 4.3.

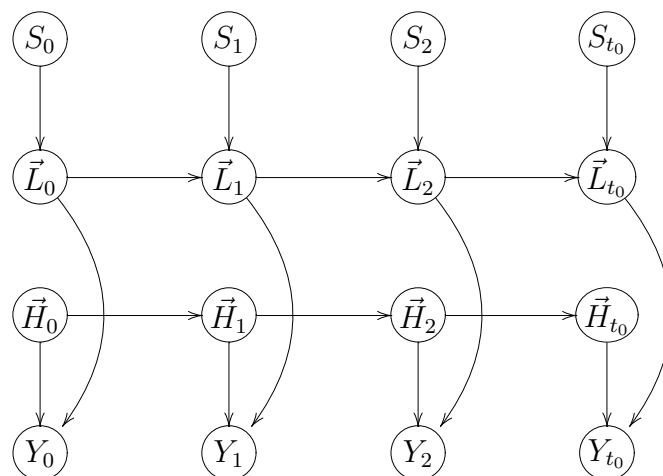


FIG. 4.3 – Réseau bayésien correspondant au problème de l'égalisation adaptative d'un canal sélectif en fréquence dont la réponse impulsionnelle évolue suivant un processus autorégressif vectoriel d'ordre 1. On a représenté le cas  $t_0 = 3$ .

Pour supprimer les cycles, on regroupe les variables  $\vec{L}_t$  et  $\vec{H}_t$  en introduisant la variable

$$\vec{B}_t = (\vec{L}_t, \vec{H}_t) \quad (4.25)$$

On obtient alors le réseau bayésien représenté à la figure 4.4.

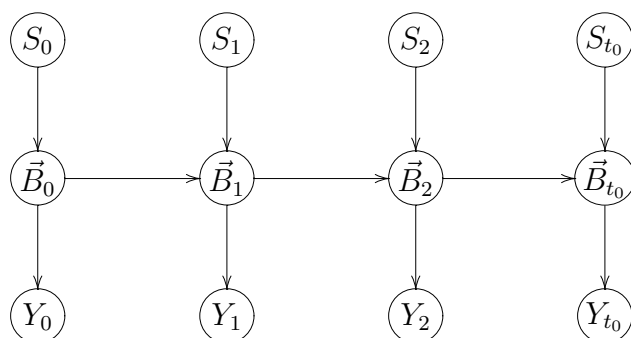


FIG. 4.4 – Réseau bayésien correspondant au problème de l'égalisation d'un canal sélectif en fréquence dont la réponse impulsionnelle évolue suivant un processus autorégressif après regroupement des variables  $\vec{L}_t$  et  $\vec{H}_t$ .

### 4.3.3 Transitions entre parents et enfants

D'après les équations 4.23 et 4.24, la transition  $\Pi_{\vec{B}_0}^{S_0}$  peut être définie par

$$\Pi_{\vec{B}_0}^{S_0}(s_0; \cdot) = \delta_{s_0 \vec{C}} \otimes \mathcal{N}_C(\vec{\eta}_0, R_G) \quad (4.26)$$

où  $\vec{C}$  est la matrice de taille  $(L+1) \times 1$  suivante

$$\vec{C} = [1 \ 0 \ 0 \ \dots \ 0]^T, \quad (4.27)$$

et, pour tout  $t > 0$ , la transition  $\Pi_{\vec{B}_t}^{(\vec{B}_{t-1}, S_t)}$  peut être définie par, pour tout  $\vec{b}_{t-1} = (\vec{l}_{t-1}, \vec{h}_{t-1})$  et pour tout  $s_t$ ,

$$\Pi_{\vec{B}_t}^{(\vec{B}_{t-1}, S_t)}(\vec{b}_{t-1}, s_t; \cdot) = \delta_{s_t \vec{C} + M \vec{l}_{t-1}} \otimes \mathcal{N}_C(\Lambda \vec{h}_{t-1}, R_G) \quad (4.28)$$

où  $M$  est la matrice de taille  $(L+1) \times (L+1)$  suivante

$$M = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \quad (4.29)$$

D'après l'équation 4.22, pour tout  $t \in \mathbb{N}$ , la transition  $\Pi_{\vec{Y}_t}^{\vec{B}_t}$  peut être définie par, pour tout  $\vec{b}_t = (\vec{l}_t, \vec{h}_t)$ ,

$$\Pi_{\vec{Y}_t}^{\vec{B}_t}(\vec{b}_t; \cdot) = \mathcal{N}_C(\vec{h}_t^T \cdot \vec{l}_t, \mathcal{N}_0) \quad (4.30)$$

$$= g(\vec{h}_t^T \cdot \vec{l}_t, \mathcal{N}_0) \cdot \lambda \quad (4.31)$$

où  $\lambda$  désigne la mesure de Lebesgue sur  $\mathbb{C}$  et  $g(\vec{h}_t^T \cdot \vec{l}_t, \mathcal{N}_0)$  désigne la densité de  $\mathcal{N}_C(\vec{h}_t^T \cdot \vec{l}_t, \mathcal{N}_0)$  par rapport à  $\lambda$ .

### 4.3.4 Application de l'algorithme de propagation de croyance

Comme le réseaux bayésien représenté à la figure 4.4 ne possède pas de cycle, on peut appliquer l'algorithme de propagation de croyance exacte pour déterminer l'estimateur du maximum *a posteriori* de  $S_{t_0}$  étant données les observations  $Y_0, Y_1, \dots, Y_{t_0}$ . Les messages nécessaires au calcul de la loi *a posteriori* de  $S_{t_0}$  ont été représentés à la figure 4.5.

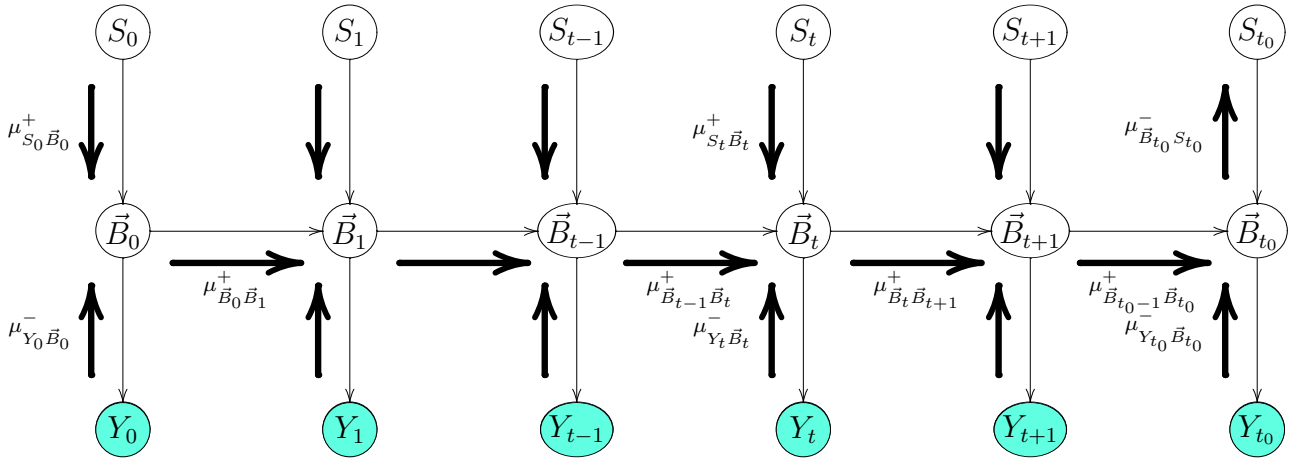


FIG. 4.5 – Application de l'algorithme de propagation de croyance au réseau bayésien correspondant au problème de l'égalisation d'un canal sélectif en fréquence dont la réponse impulsionnelle évolue suivant un processus autorégressif. Les variables aléatoires observées sont représentées en vert.

Afin de simplifier les notations, on convient dans la suite d'identifier l'ensemble des matrices de taille  $(L + 1) \times 1$  et à coefficients dans  $\mathbb{C}$  (resp.  $\mathcal{A}$ ) à l'ensemble  $\mathbb{C}^{L+1}$  (resp.  $\mathcal{A}^{L+1}$ ).

#### 4.3.4.1 Calcul du message envoyé par $\vec{B}_0$ à $\vec{B}_1$

Le message envoyé par  $\vec{B}_0$  à  $\vec{B}_1$  vérifie

$$\mu_{\vec{B}_0 \vec{B}_1}^+(y_0; \cdot) = \Psi_{\text{Bayes}} \left( \mu_{Y_0 \vec{B}_0}^-(\cdot; y_0), \sum_{S_0} F_{S_0} \otimes \Pi_{\vec{B}_0}^{S_0} \right) \quad (4.32)$$

Pour tout  $\Gamma \in \mathcal{P}(\mathcal{A}^{L+1}) \otimes \mathcal{B}(\mathbb{C}^{L+1})$ , on a donc

$$\mu_{\vec{B}_0 \vec{B}_1}^+(y_0; \Gamma) \propto \int \mathbf{1}_{\Gamma}(\vec{b}_0) \mu_{Y_0 \vec{B}_0}^-(\vec{b}_0; y_0) \Pi_{\vec{B}_0}^{S_0}(s_0; d\vec{b}_0) F_{S_0}(ds_0) \quad (4.33)$$

$$\propto \int \mathbf{1}_{\Gamma}(\vec{l}_0, \vec{h}_0) g(\vec{h}_0^T \cdot \vec{l}_0, \mathcal{N}_0)(y_0) \mathcal{N}_{\mathcal{C}}(\vec{\eta}_0, R_G)(d\vec{h}_0) \delta_{s_0 \vec{C}}(d\vec{l}_0) F_{S_0}(ds_0) \quad (4.34)$$

En utilisant la relation C.32, on en déduit

$$\begin{aligned} \mu_{\vec{B}_0 \vec{B}_1}^+(y_0; \Gamma) \propto & \int \mathbf{1}_{\Gamma}(\vec{l}_0, \vec{h}_0) [\mathcal{N}_0 + \vec{l}_0^T R_G \vec{l}_0^*]^{-1} \exp \left\{ \vec{m}_0(y_0, \vec{l}_0)^\dagger \Sigma_0(\vec{l}_0)^{-1} \vec{m}_0(y_0, \vec{l}_0) \right\} \dots \\ & \dots \mathcal{N}_{\mathcal{C}} \left( \vec{m}_0(y_0, \vec{l}_0), \Sigma_0(\vec{l}_0) \right) (d\vec{h}_0) \delta_{s_0 \vec{C}}(d\vec{l}_0) F_{S_0}(ds_0) \end{aligned} \quad (4.35)$$

avec

$$\vec{m}_0(y_0, \vec{l}_0) = \vec{\eta}_0 + [y_0 - \vec{l}_0^T \vec{\eta}_0] K_0(\vec{l}_0) \quad (4.36)$$

$$\Sigma_0(\vec{l}_0) = [I_{L+1} - K_0(\vec{l}_0) \vec{l}_0^T] R_G \quad (4.37)$$

$$K_0(\vec{l}_0) = [\mathcal{N}_0 + \vec{l}_0^T R_G \vec{l}_0^*]^{-1} R_G \vec{l}_0^* \quad (4.38)$$



Par conséquent, pour tout  $y_0 \in \mathbb{C}$  et pour tout  $\Gamma' \in \mathcal{P}(\mathcal{A}^{L+1})$ , on a

$$\Pi_{\vec{L}_0}^{Y_0}(y_0; \Gamma') = \Pi_{(\vec{L}_0, \vec{H}_0)}^{Y_0}(y_0; \Gamma' \times \mathbb{C}^{L+1}) \quad (4.39)$$

$$= \mu_{\vec{B}_0 \vec{B}_1}^+(y_0; \Gamma' \times \mathbb{C}^{L+1}) \quad (4.40)$$

$$\begin{aligned} &\propto \int \mathbb{1}_{\Gamma'}(\vec{l}_0) [\mathcal{N}_0 + \vec{l}_0^T R_G \vec{l}_0^*]^{-1} \times \dots \\ &\dots \times \exp\left\{ \vec{m}_0(y_0, \vec{l}_0)^\dagger \Sigma_0(\vec{l}_0)^{-1} \vec{m}_0(y_0, \vec{l}_0) \right\} \delta_{s_0 \vec{C}}(d\vec{l}_0) F_{S_0}(ds_0) \end{aligned} \quad (4.41)$$

ce qui implique, pour tout  $y_0 \in \mathbb{C}$  et pour tout  $\vec{l}_0 \in \mathcal{A}^{L+1}$ ,

$$\begin{aligned} \Pi_{\vec{L}_0}^{Y_0}(y_0; \{\vec{l}_0\}) &\propto \dots \\ &\dots [\mathcal{N}_0 + \vec{l}_0^T R_G \vec{l}_0^*]^{-1} \exp\left\{ \vec{m}_0(y_0, \vec{l}_0)^\dagger \Sigma_0(\vec{l}_0)^{-1} \vec{m}_0(y_0, \vec{l}_0) \right\} \sum_{s_0} \delta_{s_0 \vec{C}, \vec{l}_0} F_{S_0}(\{s_0\}) \end{aligned} \quad (4.42)$$

où  $\delta_{s_0 \vec{C}, \vec{l}_0}$  vaut 1 si  $s_0 \vec{C} = \vec{l}_0$ , et 0 sinon.

D'autre part, en appliquant l'algorithme de propagation de croyance conditionnelle (cf. l'annexe D) au réseau bayésien de la figure 4.6, on montre aisément que

$$\Pi_{\vec{H}_0}^{(Y_0, \vec{L}_0)}(y_0, \vec{l}_0; \cdot) = \mathcal{N}_C\left(\vec{m}_0(y_0, \vec{l}_0), \Sigma_0(\vec{l}_0)\right) \quad (4.43)$$

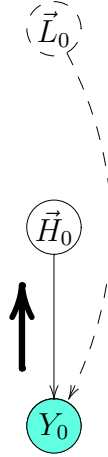


FIG. 4.6 – Calcul de la transition  $\Pi_{\vec{H}_0}^{(Y_0, \vec{L}_0)}$  par propagation de croyance conditionnelle. On prend comme observations  $Z = \{Y_0\}$  et comme ensemble de conditionnement  $C = \{\vec{L}_0\}$ .

En reportant les équations 4.41 et 4.43 dans l'équation 4.35, on obtient alors l'expression suivante pour le message envoyé par  $\vec{B}_0$  à  $\vec{B}_1$

$$\mu_{\vec{B}_0 \vec{B}_1}^+(y_0; \cdot) = \Pi_{\vec{L}_0}^{Y_0}(y_0; \cdot) \otimes \Pi_{\vec{H}_0}^{(Y_0, \vec{L}_0)}(y_0, \cdot; \cdot) \quad (4.44)$$

où les termes  $\Pi_{\vec{L}_0}^{Y_0}$  et  $\Pi_{\vec{H}_0}^{(Y_0, \vec{L}_0)}$  peuvent être calculées à l'aide des équations 4.42, 4.36, 4.37, 4.38 et 4.43.

4.3.4.2 Calcul par récurrence du message envoyé par  $\vec{B}_t$  à  $\vec{B}_{t+1}$ 

Soit  $t \geq 1$ . Supposons que le message envoyé par  $\vec{B}_{t-1}$  à  $\vec{B}_t$  vérifie, pour tout  $y_{0:t-1} \in \mathbb{C}^t$ ,

$$\mu_{\vec{B}_{t-1}\vec{B}_t}^+(y_{0:t-1}; \cdot) = \begin{cases} \Pi_{\vec{L}_0}^{Y_0}(y_0; \cdot) \otimes \Pi_{\vec{H}_0}^{(Y_0, \vec{L}_0)}(y_0, \cdot; \cdot) & \text{si } t = 1 \\ \sum_{\vec{L}_{0:t-2}} \Pi_{\vec{L}_{0:t-1}}^{Y_{0:t-1}}(y_{0:t-1}; \cdot) \otimes \Pi_{\vec{H}_{t-1}}^{(Y_{0:t-1}, \vec{L}_{0:t-1})}(y_{0:t-1}, \cdot; \cdot) & \text{si } t \geq 2 \end{cases} \quad (4.45)$$

avec, pour tout  $y_{0:t-1} \in \mathbb{C}^t$  et pour tout  $\vec{l}_{0:t-1} \in \mathcal{A}^{(L+1) \times t}$ ,

$$\Pi_{\vec{H}_{t-1}}^{(Y_{0:t-1}, \vec{L}_{0:t-1})}(y_{0:t-1}, \vec{l}_{0:t-1}; \cdot) = \mathcal{N}_C(\vec{m}_{t-1}(y_{0:t-1}, \vec{l}_{0:t-1}), \Sigma_{t-1}(\vec{l}_{0:t-1})) \quad (4.46)$$

Le message envoyé par  $\vec{B}_t$  à  $\vec{B}_{t+1}$  vérifie

$$\mu_{\vec{B}_t\vec{B}_{t+1}}^+(y_{0:t}; \cdot) = \Psi_{\text{Bayes}}\left(\mu_{Y_t\vec{B}_t}^-(\cdot; y_t), \sum_{\substack{S_t \\ \vec{B}_{t-1}}} (\mu_{\vec{B}_{t-1}\vec{B}_t}^+(y_{0:t-1}; \cdot) \otimes F_{S_t}) \otimes \Pi_{\vec{B}_t}^{(\vec{B}_{t-1}, S_t)}\right) \quad (4.47)$$

Pour tout  $\Gamma \in \mathcal{P}(\mathcal{A}^{L+1}) \otimes \mathcal{B}(\mathbb{C}^{L+1})$ , on a donc

$$\begin{aligned} \mu_{\vec{B}_t\vec{B}_{t+1}}^+(y_{0:t}; \Gamma) &\propto \int \mathbf{1}_\Gamma(\vec{b}_t) \mu_{Y_t\vec{B}_t}^-(\vec{b}_t; y_t) \dots \\ &\dots \Pi_{\vec{B}_t}^{(\vec{B}_{t-1}, S_t)}(\vec{b}_{t-1}, s_t; d\vec{b}_t) \mu_{\vec{B}_{t-1}\vec{B}_t}^+(y_{0:t-1}; d\vec{b}_{t-1}) F_{S_t}(ds_t) \end{aligned} \quad (4.48)$$

Dans la suite, on supposera  $t \geq 2$ , le traitement du cas  $t = 1$  étant similaire au traitement du cas général. Les équations 4.28, 4.31, 4.45 et 4.48 impliquent alors

$$\begin{aligned} \mu_{\vec{B}_t\vec{B}_{t+1}}^+(y_{0:t}; \Gamma) &\propto \int \mathbf{1}_\Gamma(\vec{l}_t, \vec{h}_t) g(\vec{h}_t^T \vec{l}_t, \mathcal{N}_0)(y_t) \mathcal{N}_C(\Lambda \vec{h}_{t-1}, R_G)(d\vec{h}_t) \mathcal{N}_C(\vec{m}_{t-1}(y_{0:t-1}, \vec{l}_{0:t-1}), \dots \\ &\dots \Sigma_{t-1}(\vec{l}_{0:t-1})) (d\vec{h}_{t-1}) \delta_{s_t \vec{C} + M \vec{l}_{t-1}}(d\vec{l}_t) \Pi_{\vec{L}_{0:t-1}}^{Y_{0:t-1}}(y_{0:t-1}; d\vec{l}_{0:t-1}) F_{S_t}(ds_t) \end{aligned} \quad (4.49)$$

En utilisant la relation C.36, on en déduit

$$\begin{aligned} \mu_{\vec{B}_t\vec{B}_{t+1}}^+(y_{0:t}; \Gamma) &\propto \int \mathbf{1}_\Gamma(\vec{l}_t, \vec{h}_t) g(\vec{h}_t^T \vec{l}_t, \mathcal{N}_0)(y_t) \mathcal{N}_C(\vec{m}_{t|t-1}(y_{0:t-1}, \vec{l}_{0:t-1}), \dots \\ &\dots \Sigma_{t|t-1}(\vec{l}_{0:t-1})) (d\vec{h}_t) \delta_{s_t \vec{C} + M \vec{l}_{t-1}}(d\vec{l}_t) \Pi_{\vec{L}_{0:t-1}}^{Y_{0:t-1}}(y_{0:t-1}; d\vec{l}_{0:t-1}) F_{S_t}(ds_t) \end{aligned} \quad (4.50)$$

avec

$$\vec{m}_{t|t-1}(y_{0:t-1}, \vec{l}_{0:t-1}) = \Lambda \vec{m}_{t-1}(y_{0:t-1}, \vec{l}_{0:t-1}) \quad (4.51)$$

$$\Sigma_{t|t-1}(\vec{l}_{0:t-1}) = \Lambda \Sigma_{t-1}(\vec{l}_{0:t-1}) \Lambda^\dagger + R_G \quad (4.52)$$

En utilisant la relation C.32, on en déduit

$$\begin{aligned} \mu_{\vec{B}_t\vec{B}_{t+1}}^+(y_{0:t}; \Gamma) &\propto \int \mathbf{1}_\Gamma(\vec{l}_t, \vec{h}_t) \left[ \mathcal{N}_0 + \vec{l}_t^T \Sigma_{t|t-1}(\vec{l}_{0:t-1}) \vec{l}_t^* \right]^{-1} \dots \\ &\dots \exp \left\{ -\vec{m}_{t|t-1}(y_{0:t-1}, \vec{l}_{0:t-1})^\dagger \Sigma_{t|t-1}(\vec{l}_{0:t-1})^{-1} \vec{m}_{t|t-1}(y_{0:t-1}, \vec{l}_{0:t-1}) + \dots \right. \\ &\dots \left. + \vec{m}_t(y_{0:t}, \vec{l}_{0:t})^\dagger \Sigma_t(\vec{l}_{0:t})^{-1} \vec{m}_t(y_{0:t}, \vec{l}_{0:t}) \right\} \mathcal{N}_C(\vec{m}_t(y_{0:t}, \vec{l}_{0:t}), \Sigma_t(\vec{l}_{0:t})) (d\vec{h}_t) \dots \\ &\dots \delta_{s_t \vec{C} + M \vec{l}_{t-1}}(d\vec{l}_t) \Pi_{\vec{L}_{0:t-1}}^{Y_{0:t-1}}(y_{0:t-1}; d\vec{l}_{0:t-1}) F_{S_t}(ds_t) \end{aligned} \quad (4.53)$$

avec

$$\vec{m}_t(y_{0:t}, \vec{l}_{0:t}) = \vec{m}_{t|t-1}(y_{0:t-1}, \vec{l}_{0:t-1}) + [y_t - \vec{l}_t^T \vec{m}_{t|t-1}(y_{0:t-1}, \vec{l}_{0:t-1})] K_t(\vec{l}_{0:t}) \quad (4.54)$$

$$\Sigma_t(\vec{l}_{0:t}) = [I_{L+1} - K_t(\vec{l}_{0:t}) \vec{l}_t^T] \Sigma_{t|t-1}(\vec{l}_{0:t-1}) \quad (4.55)$$

$$K_t(\vec{l}_{0:t}) = [\mathcal{N}_0 + \vec{l}_t^T \Sigma_{t|t-1}(\vec{l}_{0:t-1}) \vec{l}_t^*]^{-1} \Sigma_{t|t-1}(\vec{l}_{0:t-1}) \vec{l}_t^* \quad (4.56)$$

Or, en utilisant les relations d'indépendance conditionnelle représentées par le réseau bayésien de la figure 4.3, on montre que

$$\begin{aligned} \Pi_{\vec{L}_{0:t}}^{Y_{0:t}}(y_{0:t}; \cdot) &= \sum_{\vec{H}_{t-1:t}} \Psi_{\text{Bayes}} \left( f_{Y_t}^{(\vec{L}_t, \vec{H}_t)}(\cdot; y_t), \dots \right. \\ &\quad \left. \dots \sum_{S_t} [\Pi_{\vec{L}_{0:t-1}}^{Y_{0:t-1}}(y_{0:t-1}; \cdot) \otimes \Pi_{\vec{H}_{t-1}}^{(Y_{0:t-1}, \vec{L}_{0:t-1})}(y_{0:t-1}, \cdot; \cdot) \otimes F_{S_t}] \otimes \Pi_{(\vec{L}_t, \vec{H}_t)}^{(\vec{L}_{t-1}, \vec{H}_{t-1}, S_t)} \right) \end{aligned} \quad (4.57)$$

avec  $f_{Y_t}^{(\vec{L}_t, \vec{H}_t)}(\vec{l}_t, \vec{h}_t; \cdot) = g(\vec{h}_t^T \vec{l}_t, \mathcal{N}_0)$ . Après des calculs similaires aux calculs précédents, on en déduit, pour tout  $\Gamma' \in \mathcal{P}(\mathcal{A}^{(L+1) \times (t+1)})$ ,

$$\begin{aligned} \Pi_{\vec{L}_{0:t}}^{Y_{0:t}}(y_{0:t}; \Gamma') &\propto \int \mathbf{1}_{\Gamma'}(\vec{l}_{0:t}) \left[ \mathcal{N}_0 + \vec{l}_t^T \Sigma_{t|t-1}(\vec{l}_{0:t-1}) \vec{l}_t^* \right]^{-1} \dots \\ &\quad \dots \exp \left\{ -\vec{m}_{t|t-1}(y_{0:t-1}, \vec{l}_{0:t-1})^\dagger \Sigma_{t|t-1}(\vec{l}_{0:t-1})^{-1} \vec{m}_{t|t-1}(y_{0:t-1}, \vec{l}_{0:t-1}) + \vec{m}_t(y_{0:t}, \vec{l}_{0:t})^\dagger \times \dots \right. \\ &\quad \left. \dots \times \Sigma_t(\vec{l}_{0:t})^{-1} \vec{m}_t(y_{0:t}, \vec{l}_{0:t}) \right\} \delta_{s_t \vec{C} + M \vec{l}_{t-1}, \vec{l}_t} \Pi_{\vec{L}_{0:t-1}}^{Y_{0:t-1}}(y_{0:t-1}; d\vec{l}_{0:t-1}) F_{S_t}(ds_t) \end{aligned} \quad (4.58)$$

ce qui implique, pour tout  $y_{0:t} \in \mathbb{C}^{t+1}$  et pour tout  $\vec{l}_{0:t} \in \mathcal{A}^{(L+1) \times (t+1)}$ ,

$$\begin{aligned} \Pi_{\vec{L}_{0:t}}^{Y_{0:t}}(y_{0:t}; \{\vec{l}_{0:t}\}) &\propto \left[ \mathcal{N}_0 + \vec{l}_t^T \Sigma_{t|t-1}(\vec{l}_{0:t-1}) \vec{l}_t^* \right]^{-1} \dots \\ &\quad \dots \exp \left\{ -\vec{m}_{t|t-1}(y_{0:t-1}, \vec{l}_{0:t-1})^\dagger \Sigma_{t|t-1}(\vec{l}_{0:t-1})^{-1} \vec{m}_{t|t-1}(y_{0:t-1}, \vec{l}_{0:t-1}) + \vec{m}_t(y_{0:t}, \vec{l}_{0:t})^\dagger \times \dots \right. \\ &\quad \left. \dots \times \Sigma_t(\vec{l}_{0:t})^{-1} \vec{m}_t(y_{0:t}, \vec{l}_{0:t}) \right\} \sum_{s_t} \delta_{s_t \vec{C} + M \vec{l}_{t-1}, \vec{l}_t} \Pi_{\vec{L}_{0:t-1}}^{Y_{0:t-1}}(y_{0:t-1}; \{\vec{l}_{0:t-1}\}) F_{S_t}(\{s_t\}) \end{aligned} \quad (4.59)$$

D'autre part, en appliquant l'algorithme de propagation de croyance conditionnelle au réseau bayésien de la figure 4.7, on montre que

$$\Pi_{\vec{H}_t}^{(Y_{0:t}, \vec{L}_{0:t})}(y_{0:t}, \vec{l}_{0:t}; \cdot) = \mathcal{N}_C \left( \vec{m}_t(y_{0:t}, \vec{l}_{0:t}), \Sigma_t(\vec{l}_{0:t}) \right) \quad (4.60)$$

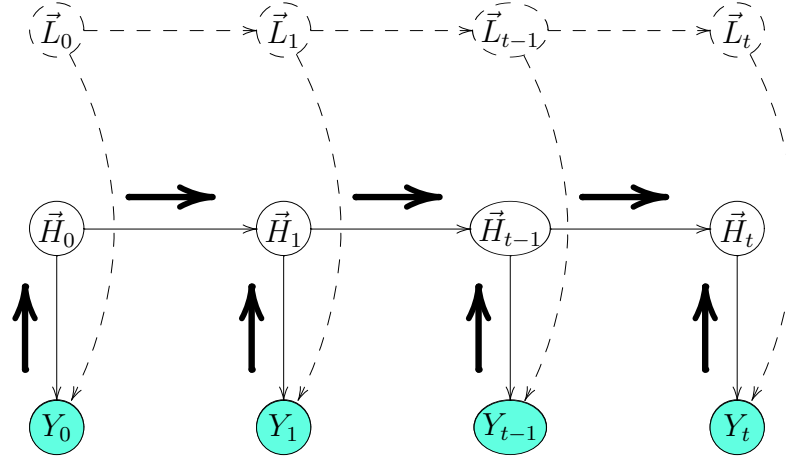


FIG. 4.7 – Calcul de la transition  $\Pi_{\vec{H}_t}^{(Y_{0:t}, \vec{L}_{0:t})}$  par propagation de croyance conditionnelle. On prend comme observations  $\mathbf{Z} = \{Y_0, Y_1, \dots, Y_t\}$  et comme ensemble de conditionnement  $\mathbf{C} = \{\vec{L}_0, \vec{L}_1, \dots, \vec{L}_t\}$ .

En reportant les équations 4.58 et 4.60 dans l'équation 4.53, on obtient alors l'expression suivante pour le message envoyé par  $\vec{B}_t$  à  $\vec{B}_{t+1}$

$$\mu_{\vec{B}_t \vec{B}_{t+1}}^+(y_{0:t}; \cdot) = \sum_{\vec{L}_{0:t-1}} \Pi_{\vec{L}_{0:t}}^{Y_{0:t}}(y_{0:t}; \cdot) \otimes \Pi_{\vec{H}_t}^{(Y_{0:t}, \vec{L}_{0:t})}(y_{0:t}, \cdot; \cdot) \quad (4.61)$$

où les transitions  $\Pi_{\vec{L}_{0:t}}^{Y_{0:t}}$  et  $\Pi_{\vec{H}_t}^{(Y_{0:t}, \vec{L}_{0:t})}$  peuvent être calculées par récurrence à l'aide des équations 4.51, 4.52, 4.54, 4.55, 4.56 et 4.59.

#### 4.3.4.3 Calcul du message envoyé par $\vec{B}_{t_0}$ à $S_{t_0}$

Le message envoyé par  $\vec{B}_{t_0}$  à  $S_{t_0}$  vérifie

$$\mu_{\vec{B}_{t_0} S_{t_0}}^-(\cdot; y_{0:t_0}) = \sum_{\vec{B}_{t_0}} \mu_{Y_{t_0} \vec{B}_{t_0}}^-(\cdot; y_{t_0}) \sum_{\vec{B}_{t_0-1}} \mu_{\vec{B}_{t_0-1} \vec{B}_{t_0}}^+(y_{0:t_0-1}; \cdot) \otimes \Pi_{\vec{B}_{t_0}}^{(S_{t_0}, \vec{B}_{t_0-1})} \quad (4.62)$$

Pour tout  $s_0 \in \mathcal{A}$ , on a donc

$$\begin{aligned} \mu_{\vec{B}_{t_0} S_{t_0}}^-(s_0; y_{0:t_0}) &= \dots \\ &\dots \int \mu_{Y_{t_0} \vec{B}_{t_0}}^-(\vec{b}_{t_0}; y_{t_0}) \Pi_{\vec{B}_{t_0}}^{(S_{t_0}, \vec{B}_{t_0-1})}(s_{t_0}, \vec{b}_{t_0-1}; d\vec{b}_{t_0}) \mu_{\vec{B}_{t_0-1} \vec{B}_{t_0}}^+(y_{0:t_0-1}; d\vec{b}_{t_0-1}) \end{aligned} \quad (4.63)$$

En utilisant les équations 4.28, 4.31, 4.60 et 4.61, on en déduit

$$\begin{aligned} \mu_{\vec{B}_{t_0} S_{t_0}}^-(s_0; y_{0:t_0}) &= \int g(\vec{h}_{t_0}^T \cdot \vec{l}_{t_0}, \mathcal{N}_0)(y_{t_0}) \mathcal{N}_C(\Lambda \vec{h}_{t_0-1}, R_G)(d\vec{h}_{t_0}) \mathcal{N}_C(\vec{m}_{t_0-1}(y_{0:t_0-1}, \dots \\ &\dots \vec{l}_{0:t_0-1}), \Sigma_{t_0-1}(\vec{l}_{0:t_0-1})) (d\vec{h}_{t_0-1}) \delta_{s_{t_0} \vec{C} + M \vec{l}_{t_0-1}}(d\vec{l}_{t_0}) \Pi_{\vec{L}_{0:t_0-1}}^{Y_{0:t_0-1}}(y_{0:t_0-1}; d\vec{l}_{0:t_0-1}) \end{aligned} \quad (4.64)$$

En effectuant des calculs similaires aux calculs précédents, on en déduit

$$\begin{aligned} \mu_{\vec{B}_{t_0} S_{t_0}}^-(s_0; y_{0:t_0}) &\propto \int \left[ \mathcal{N}_0 + \vec{l}_t^T \Sigma_{t_0|t_0-1} (\vec{l}_{0:t_0-1}) \vec{l}_{t_0}^* \right]^{-1} \exp \left\{ -\vec{m}_{t_0|t_0-1} (y_{0:t_0-1}, \vec{l}_{0:t_0-1})^\dagger \times \dots \right. \\ &\quad \dots \times \Sigma_{t_0|t_0-1} (\vec{l}_{0:t_0-1})^{-1} \vec{m}_{t_0|t_0-1} (y_{0:t_0-1}, \vec{l}_{0:t_0-1}) + \vec{m}_{t_0} (y_{0:t_0}, \vec{l}_{0:t_0})^\dagger \times \dots \\ &\quad \left. \dots \times \Sigma_{t_0} (\vec{l}_{0:t_0})^{-1} \vec{m}_{t_0} (y_{0:t_0}, \vec{l}_{0:t_0}) \right\} \delta_{s_{t_0} \vec{C} + M \vec{l}_{t_0-1}} (d \vec{l}_{t_0}) \Pi_{\vec{L}_{0:t_0-1}}^{Y_{0:t_0-1}} (y_{0:t_0-1}; d \vec{l}_{0:t_0-1}) \quad (4.65) \end{aligned}$$

#### 4.3.4.4 Calcul de la loi *a posteriori* de $S_{t_0}$

On obtient alors la loi *a posteriori* de  $S_{t_0}$

$$\Pi_{S_{t_0}}^{Y_{0:t_0}} (y_{0:t_0}; \cdot) = \Psi_{\text{Bayes}} \left( \mu_{\vec{B}_{t_0} S_{t_0}}^-(\cdot; y_{0:t_0}), F_{S_{t_0}} \right) \quad (4.66)$$

#### 4.3.4.5 Complexité de l'algorithme

Considérons le calcul du message envoyé par  $\vec{B}_t$  à  $\vec{B}_{t+1}$  à l'aide de l'équation 4.61. L'équation 4.59 montre que le cardinal du support de la mesure  $\Pi_{\vec{L}_{0:t}}^{Y_{0:t}} (y_{0:t}; \cdot)$  vaut  $\text{card}(\mathcal{A})^{t+1}$ . Comme, pour tout  $\vec{l}_{0:t}$  appartenant au support de la mesure  $\Pi_{\vec{L}_{0:t}}^{Y_{0:t}} (y_{0:t}; \cdot)$ , le calcul de la mesure  $\Pi_{\vec{H}_t}^{(Y_{0:t}, \vec{L}_{0:t})} (y_{0:t}, \vec{l}_{0:t}; \cdot)$  à l'aide des équations 4.51, 4.52, 4.54, 4.55 et 4.56 nécessite un nombre d'opérations en  $O(tL^2)$ , le calcul du message envoyé par  $\vec{B}_t$  à  $\vec{B}_{t+1}$  nécessite un nombre d'opérations en  $O(tL^2 \text{card}(\mathcal{A})^{t+1})$ . Ce calcul étant prédominant, la complexité de l'algorithme est  $O(t_0 L^2 \text{card}(\mathcal{A})^{t_0+1})$ .

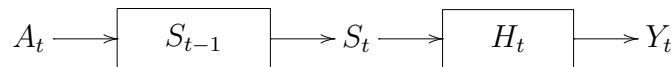
### 4.3.5 Conclusion

L'algorithme que nous avons obtenu en appliquant l'algorithme de propagation de croyance exact est optimal, mais sa complexité est exponentielle par rapport au temps, ce qui rend son implémentation impossible en pratique. Des approximations seront donc nécessaires afin d'obtenir un compromis plus satisfaisant entre les performances et la complexité.

## 4.4 L'égalisation adaptative d'un canal évanescent non-sélectif en fréquence

### 4.4.1 Modèle du système

On suppose que des symboles complexes  $A_t$  appartenant à un alphabet  $\mathcal{A}$  stable pour la multiplication sont codés différentiellement en des symboles  $S_t \in \mathcal{A}$  qui sont transmis à travers un canal non-sélectif en fréquence dont on observe la sortie complexe  $Y_t$ .



Toutes les grandeurs sont supposées échantillonnées au rythme symbole. On suppose que la sortie complexe  $Y_t$  du canal vérifie, pour tout  $t \in \mathbb{N}$ ,

$$Y_t = H_t S_t + V_t \quad (4.67)$$

où  $V_t$  suit la loi gaussienne  $\mathcal{N}_C(0, \mathcal{N}_0)$ , le symbole  $S_t$  est défini par

$$S_t = \begin{cases} A_t S_{t-1} & \text{si } t > 0 \\ A_0 & \text{si } t = 0 \end{cases} \quad (4.68)$$

et le coefficient  $H_t$  est défini par

$$H_t = \vec{\Theta}^T \vec{X}_t \quad (4.69)$$

$$\vec{\Theta} = [\theta_0 \ \theta_1 \ \dots \ \theta_L]^T \quad (4.70)$$

$$\vec{X}_t = \begin{cases} \Lambda \vec{X}_{t-1} + \vec{W}_t & \text{si } t > 0 \\ \vec{\eta}_0 + \vec{W}_0 & \text{si } t = 0 \end{cases} \quad (4.71)$$

où  $\vec{W}_t$  suit la loi gaussienne  $\mathcal{N}_C(\vec{0}, R_G)$ ,  $\vec{\eta}_0$  et  $\vec{\Theta}$  sont des vecteurs complexes supposés connus et  $\Lambda$  une matrice complexe également supposée connue et de la forme suivante

$$\Lambda = \begin{bmatrix} -\phi_1 & -\phi_2 & \dots & -\phi_L & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \quad (4.72)$$

On suppose également que la matrice de covariance  $R_G$  du bruit  $\vec{W}_t$  est de la forme suivante

$$R_G = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \quad (4.73)$$

Enfin, on suppose que les variables aléatoires  $A_0, \vec{W}_0, V_0, A_1, \vec{W}_1, V_1, \dots, A_t, \vec{W}_t, V_t$  sont indépendantes et que  $A_t$  suit une loi  $F_{A_t}$  sur l'alphabet  $\mathcal{A}$ . Pour tout  $t_0 \in \mathbb{N}$ , on recherche l'estimateur du maximum *a posteriori* de  $A_{t_0}$  étant données les observations  $Y_0, Y_1, \dots, Y_{t_0}$ .

#### Remarque 4.1

On vérifie aisément que les équations 4.69 à 4.73 impliquent que le coefficient  $H_t$  évolue selon un processus ARMA( $L, L$ ) défini par

$$H_t + \phi_1 H_{t-1} + \dots + \phi_L H_{t-L} = \theta_0 \nu_t + \theta_1 \nu_{t-1} + \dots + \theta_L \nu_{t-L} \quad (4.74)$$

avec  $\nu_t \sim \mathcal{N}_C(0, 1)$ . Le processus  $(H_t)$  est donc obtenu à l'aide d'un filtre de Butterworth dont la fréquence de coupure correspond à la fréquence Doppler du canal normalisée par rapport au débit symbole.

### 4.4.2 Réseau bayésien

En procédant comme indiqué au § 2.3, on obtient le réseau bayésien représenté à la figure 4.8.

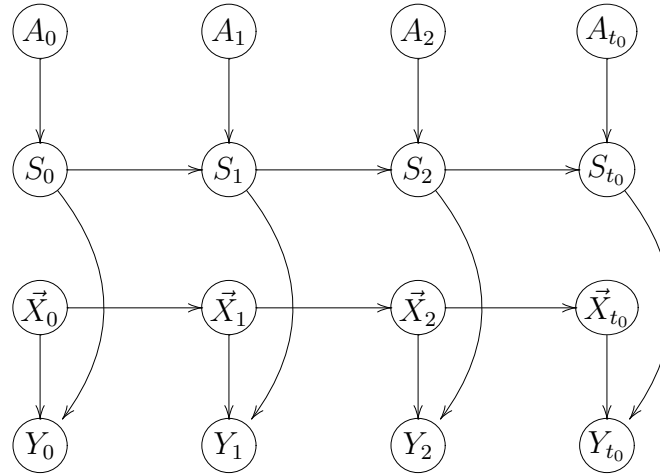


FIG. 4.8 – Réseau bayésien correspondant au problème de l'égalisation adaptative d'un canal évanescent non-sélectif en fréquence avec un codage différentiel. On a représenté le cas  $t_0 = 3$ .

Pour supprimer les cycles, on regroupe les variables  $S_t$  et  $X_t$  en introduisant la variable

$$\vec{B}_t = (S_t, X_t) \quad (4.75)$$

On obtient alors le réseau bayésien représenté à la figure 4.9.

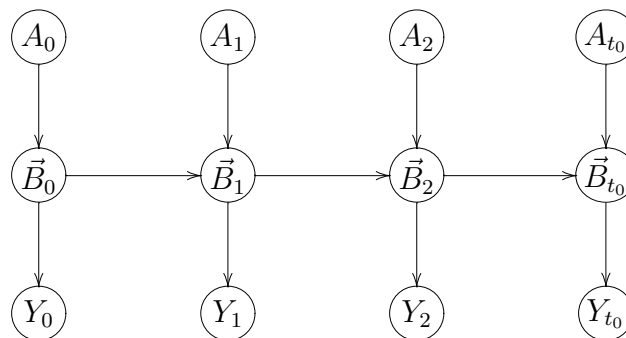


FIG. 4.9 – Réseau bayésien correspondant au problème de l'égalisation adaptative d'un canal évanescent non-sélectif en fréquence avec un codage différentiel. Les variables  $S_t$  et  $X_t$  ont été regroupées.

### 4.4.3 Transitions entre parents et enfants

D'après les équations 4.68 et 4.71, la transition  $\Pi_{\vec{B}_0}^{A_0}$  peut être définie par

$$\Pi_{\vec{B}_0}^{A_0}(a_0; \cdot) = \delta_{a_0} \otimes \mathcal{N}_C(\vec{\eta}_0, R_G) \quad (4.76)$$

et, pour tout  $t > 0$ , la transition  $\Pi_{\vec{B}_t}^{(\vec{B}_{t-1}, A_t)}$  peut être définie par, pour tout  $\vec{b}_{t-1} = (s_{t-1}, \vec{x}_{t-1})$  et pour tout  $a_t$ ,

$$\Pi_{\vec{B}_t}^{(\vec{B}_{t-1}, A_t)}(\vec{b}_{t-1}, a_t; \cdot) = \delta_{a_t s_{t-1}} \otimes \mathcal{N}_C(\Lambda \vec{x}_{t-1}, R_G) \quad (4.77)$$

D'après l'équation 4.67, pour tout  $t \in \mathbb{N}$ , la transition  $\Pi_{Y_t}^{\vec{B}_t}$  peut être définie par, pour tout  $\vec{b}_t = (s_t, \vec{x}_t)$ ,

$$\Pi_{Y_t}^{\vec{B}_t}(\vec{b}_t; \cdot) = \mathcal{N}_C(s_t \vec{x}_t^T \cdot \vec{\Theta}, \mathcal{N}_0) \quad (4.78)$$

$$= g(s_t \vec{x}_t^T \cdot \vec{\Theta}, \mathcal{N}_0) \cdot \lambda \quad (4.79)$$

où  $\lambda$  désigne la mesure de Lebesgue sur  $\mathbb{C}$  et  $g(s_t \vec{x}_t^T \cdot \vec{\Theta}, \mathcal{N}_0)$  désigne la densité de  $\mathcal{N}_C(s_t \vec{x}_t^T \cdot \vec{\Theta}, \mathcal{N}_0)$  par rapport à  $\lambda$ .

### 4.4.4 Application de l'algorithme de propagation de croyance

Comme le réseaux bayésien représenté à la figure 4.9 ne possède pas de cycle, on peut appliquer l'algorithme de propagation de croyance exacte pour déterminer l'estimateur du maximum *a posteriori* de  $A_{t_0}$  étant données les observations  $Y_0, Y_1, \dots, Y_{t_0}$ . Les messages nécessaires au calcul de la loi *a posteriori* de  $A_{t_0}$  ont été représentés à la figure 4.10.

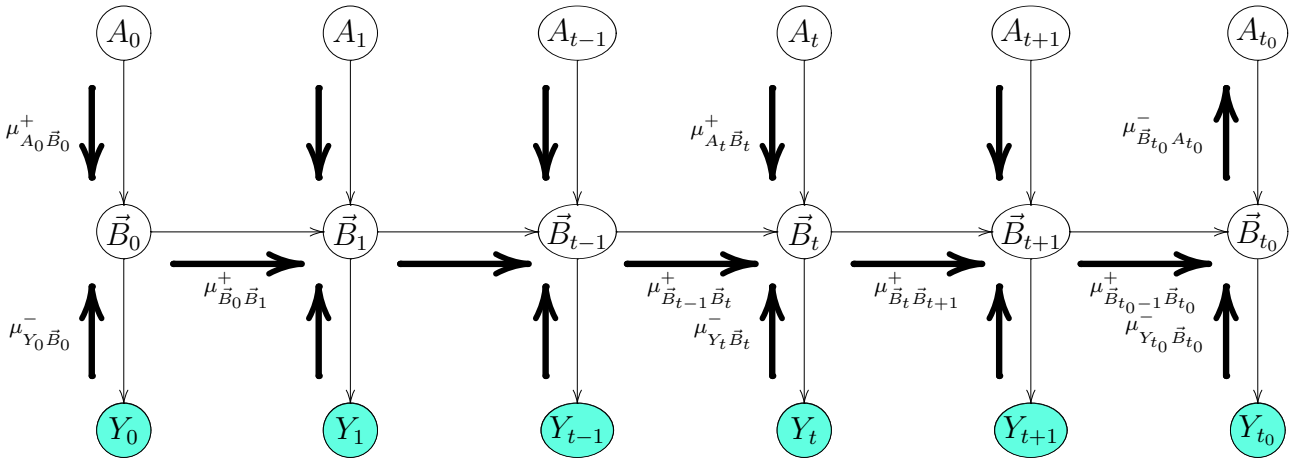


FIG. 4.10 – Application de l'algorithme de propagation de croyance au réseau bayésien correspondant au problème de l'égalisation adaptative d'un canal évanescent non-sélectif en fréquence avec un codage différentiel. Les variables aléatoires observées sont représentées en vert.

En procédant comme au § 4.3.4, on peut montrer que les messages nécessaires au calcul de la loi *a posteriori* de  $A_{t_0}$  étant données les observations  $Y_0, Y_1, \dots, Y_{t_0}$  peuvent s'écrire sous la forme indiquée ci-dessous.



#### 4.4.4.1 Message envoyé par $\vec{B}_0$ à $\vec{B}_1$

Le message envoyé par  $\vec{B}_0$  à  $\vec{B}_1$  vérifie

$$\mu_{\vec{B}_0\vec{B}_1}^+(y_0; \cdot) = \Pi_{S_0}^{Y_0}(y_0; \cdot) \otimes \Pi_{\vec{X}_0}^{(Y_0, S_0)}(y_0, \cdot; \cdot) \quad (4.80)$$

avec, pour tout  $y_0 \in \mathbb{C}$  et pour tout  $s_0 \in \mathcal{A}$ ,

$$\begin{aligned} \Pi_{S_0}^{Y_0}(y_0; \{s_0\}) &\propto [\mathcal{N}_0 + \vec{\Theta}^T R_G \vec{\Theta}^*]^{-1} \times \dots \\ &\dots \times \exp\left\{ \vec{m}_0(y_0, s_0)^\dagger \Sigma_0(s_0)^{-1} \vec{m}_0(y_0, s_0) \right\} F_{A_0}(\{s_0\}) \end{aligned} \quad (4.81)$$

$$\Pi_{\vec{X}_0}^{(Y_0, S_0)}(y_0, s_0; \cdot) = \mathcal{N}_C\left(\vec{m}_0(y_0, s_0), \Sigma_0(s_0)\right) \quad (4.82)$$

$$\vec{m}_0(y_0, s_0) = \vec{\eta}_0 + [y_0 - s_0 \vec{\Theta}^T \vec{\eta}_0] K_0(s_0) \quad (4.83)$$

$$\Sigma_0(s_0) = [I_{L+1} - K_0(s_0) s_0 \vec{\Theta}^T] R_G \quad (4.84)$$

$$K_0(s_0) = [\mathcal{N}_0 + \vec{\Theta}^T R_G \vec{\Theta}^*]^{-1} s_0^* R_G \vec{\Theta}^* \quad (4.85)$$

#### 4.4.4.2 Message envoyé par $\vec{B}_t$ à $\vec{B}_{t+1}$

Le message envoyé par  $\vec{B}_t$  à  $\vec{B}_{t+1}$  vérifie

$$\mu_{\vec{B}_t\vec{B}_{t+1}}^+(y_{0:t}; \cdot) = \sum_{S_{0:t-1}} \Pi_{S_{0:t}}^{Y_{0:t}}(y_{0:t}; \cdot) \otimes \Pi_{\vec{X}_t}^{(Y_{0:t}, S_{0:t})}(y_{0:t}, \cdot; \cdot) \quad (4.86)$$

avec, pour tout  $y_{0:t} \in \mathbb{C}^{t+1}$  et pour tout  $s_{0:t} \in \mathcal{A}^{t+1}$ ,

$$\begin{aligned} \Pi_{S_{0:t}}^{Y_{0:t}}(y_{0:t}; \{s_{0:t}\}) &\propto [\mathcal{N}_0 + \vec{\Theta}^T \Sigma_{t|t-1}(s_{0:t-1}) \vec{\Theta}^*]^{-1} \dots \\ &\dots \exp\left\{ -\vec{m}_{t|t-1}(y_{0:t-1}, s_{0:t-1})^\dagger \Sigma_{t|t-1}(s_{0:t-1})^{-1} \vec{m}_{t|t-1}(y_{0:t-1}, s_{0:t-1}) + \vec{m}_t(y_{0:t}, s_{0:t})^\dagger \times \dots \right. \\ &\quad \left. \dots \times \Sigma_t(s_{0:t})^{-1} \vec{m}_t(y_{0:t}, s_{0:t}) \right\} \sum_{a_t} \delta_{a_t s_{t-1}, s_t} \Pi_{S_{0:t-1}}^{Y_{0:t-1}}(y_{0:t-1}; \{s_{0:t-1}\}) F_{A_t}(\{a_t\}) \end{aligned} \quad (4.87)$$

$$\Pi_{\vec{X}_t}^{(Y_{0:t}, S_{0:t})}(y_{0:t}, s_{0:t}; \cdot) = \mathcal{N}_C\left(\vec{m}_t(y_{0:t}, s_{0:t}), \Sigma_t(s_{0:t})\right) \quad (4.88)$$

$$\vec{m}_{t|t-1}(y_{0:t-1}, s_{0:t-1}) = \Lambda \vec{m}_{t-1}(y_{0:t-1}, s_{0:t-1}) \quad (4.89)$$

$$\Sigma_{t|t-1}(s_{0:t-1}) = \Lambda \Sigma_{t-1}(s_{0:t-1}) \Lambda^\dagger + R_G \quad (4.90)$$

$$\begin{aligned} \vec{m}_t(y_{0:t}, s_{0:t}) &= \vec{m}_{t|t-1}(y_{0:t-1}, s_{0:t-1}) + \dots \\ &\quad \dots + [y_t - s_t \vec{\Theta}^T \vec{m}_{t|t-1}(y_{0:t-1}, s_{0:t-1})] K_t(s_{0:t}) \end{aligned} \quad (4.91)$$

$$\Sigma_t(s_{0:t}) = [I_{L+1} - s_t K_t(s_{0:t}) \vec{\Theta}^T] \Sigma_{t|t-1}(s_{0:t-1}) \quad (4.92)$$

$$K_t(s_{0:t}) = [\mathcal{N}_0 + \vec{\Theta}^T \Sigma_{t|t-1}(s_{0:t-1}) \vec{\Theta}^*]^{-1} s_t^* \Sigma_{t|t-1}(s_{0:t-1}) \vec{\Theta}^* \quad (4.93)$$

#### 4.4.4.3 Message envoyé par $\vec{B}_{t_0}$ à $A_{t_0}$

Le message envoyé par  $\vec{B}_{t_0}$  à  $A_{t_0}$  vérifie

$$\begin{aligned} \mu_{\vec{B}_{t_0} A_{t_0}}^-(a_0; y_{0:t_0}) &\propto \int \left[ \mathcal{N}_0 + \vec{\Theta}^T \Sigma_{t_0|t_0-1}(s_{0:t_0-1}) \vec{\Theta}^* \right]^{-1} \exp \left\{ -\vec{m}_{t_0|t_0-1}(y_{0:t_0-1}, s_{0:t_0-1})^\dagger \times \dots \right. \\ &\quad \dots \times \Sigma_{t_0|t_0-1}(s_{0:t_0-1})^{-1} \vec{m}_{t_0|t_0-1}(y_{0:t_0-1}, s_{0:t_0-1}) + \vec{m}_{t_0}(y_{0:t_0}, s_{0:t_0})^\dagger \times \dots \\ &\quad \left. \dots \times \Sigma_{t_0}(s_{0:t_0})^{-1} \vec{m}_{t_0}(y_{0:t_0}, s_{0:t_0}) \right\} \delta_{a_{t_0} s_{t_0-1}}(ds_{t_0}) \Pi_{S_{0:t_0-1}}^{Y_{0:t_0-1}}(y_{0:t_0-1}; ds_{0:t_0-1}) \end{aligned} \quad (4.94)$$

#### 4.4.4.4 Loi *a posteriori* de $A_{t_0}$

On obtient alors la loi *a posteriori* de  $A_{t_0}$

$$\Pi_{A_{t_0}}^{Y_{0:t_0}}(y_{0:t_0}; \cdot) = \Psi_{\text{Bayes}} \left( \mu_{\vec{B}_{t_0} A_{t_0}}^-(\cdot; y_{0:t_0}), F_{A_{t_0}} \right) \quad (4.95)$$

#### 4.4.4.5 Complexité de l'algorithme

En procédant comme au § 4.3.4.5, on montre que la complexité de cet algorithme est  $O(t_0 L^2 \text{card}(\mathcal{A})^{t_0+1})$ .

### 4.4.5 Conclusion

Comme dans le cas de l'égalisation adaptative d'un canal sélectif en fréquence, l'algorithme que nous avons obtenu en appliquant l'algorithme de propagation de croyance exact possède une complexité exponentielle par rapport au temps, ce qui rend son implémentation impossible en pratique. Dans ce cas également, des approximations seront nécessaires afin d'obtenir un compromis plus satisfaisant entre les performances et la complexité.

## 4.5 Conclusion

Dans ce chapitre, nous avons étudié l'application des réseaux bayésiens au problème de l'égalisation. Nous nous sommes intéressés aux trois situations suivantes : égalisation d'un canal sélectif en fréquence de réponse impulsionnelle connue, égalisation adaptative d'un canal évanescent sélectif en fréquence et égalisation adaptative d'un canal évanescent non-sélectif en fréquence. Nous avons construit le réseau bayésien correspondant à chacune des ces situations et nous avons appliqué l'algorithme de propagation de croyance exacte au réseau bayésien obtenu. Dans le cas de l'égalisation d'un canal sélectif en fréquence de réponse impulsionnelle connue, nous avons retrouvé l'algorithme BCJR, dont la complexité est exponentielle par rapport à la longueur de la réponse impulsionnelle du canal. Dans les deux autres cas, nous avons obtenu un algorithme dont la complexité est exponentielle par rapport au temps.

On voit donc qu'en utilisant l'algorithme de propagation de croyance pour égaliser un canal de réponse impulsionnelle longue ou pour effectuer conjointement l'estimation de canal et l'égalisation, on obtient des algorithmes qui sont optimaux mais inexploitable en raison de leur trop grande complexité. Il est alors nécessaire de recourir à des méthodes d'approximation numérique afin d'obtenir un compromis plus satisfaisant entre les performances et la complexité.

Aussi, dans le prochain chapitre, nous allons présenter une technique permettant de calculer de façon approchée une suite de mesures de probabilité. Nous verrons dans les chapitres suivants que cette technique permet d'obtenir un bon compromis entre les performances et la complexité lorsqu'on l'applique à l'égalisation. Nous verrons également qu'il est possible de combiner cette technique d'approximation avec l'algorithme de propagation de croyance pour effectuer conjointement l'égalisation adaptative et le décodage.

# Chapitre 5

## L'algorithme de filtrage particulaire à échantillonnage global

### 5.1 Introduction

Dans le chapitre précédent, nous avons vu que, dans certaines situations, l'application de l'algorithme de propagation de croyance peut conduire à des calculs dont la complexité est trop élevée en pratique. Aussi, dans ce chapitre, nous allons introduire un nouvel algorithme d'approximation numérique appelé *algorithme de filtrage particulaire à échantillonnage global* qui nous permettra d'obtenir un compromis plus satisfaisant entre les performances et la complexité.

L'algorithme que nous allons présenter fait partie de la famille des algorithmes séquentiels de Monte Carlo (*Sequential Monte Carlo algorithms* ou SMC) qui ont été introduits dans les années 1950 pour simuler les chaînes de polymère de grande longueur [HM54, RR55]. Pendant plusieurs décennies, le développement des méthodes SMC a été freiné par l'absence de moyens informatiques suffisants pour leur mise en œuvre. Au cours des années 1990, les progrès rapides de l'informatique ont provoqué un regain d'intérêt pour les méthodes SMC qui ont commencé à être utilisées dans des domaines aussi variés que le traitement du signal radar [GSS93], l'estimation bayésienne classique [KLW94], la génétique [ICK94], les communications numériques [LC95], l'analyse d'image [IB96], le diagnostic médical [BBGL97], la robotique [FBT99], la modélisation en finance [SP99] et le traitement de la parole [VADG02].

Les algorithmes SMC – que l'on appelle également *algorithmes de filtrage particulaire* – permettent d'approcher une suite de mesures de probabilité à l'aide de sommes finies de mesures de Dirac centrées en des points correspondant à des « particules ». A chaque itération, une particule peut donner naissance à des descendants ou disparaître, selon sa capacité à représenter la mesure de probabilité que l'on veut approcher. La principale différence entre chaque grande catégorie d'algorithme SMC réside dans la façon dont on fait évoluer le système de particules d'une itération à l'autre. En particulier, l'algorithme de filtrage particulaire à échantillonnage global que nous allons proposer se distingue des algorithmes SMC existants par sa façon novatrice de gérer l'évolution du système de particules d'une itération à l'autre.

Dans ce chapitre, nous présenterons tout d'abord le principe de l'algorithme de filtrage particulaire à échantillonnage global et nous expliquerons l'originalité de notre approche. Nous montrerons ensuite que l'erreur quadratique moyenne de l'algorithme converge vers 0 lorsque

le nombre de particules tend vers l'infini. Enfin, nous présenterons brièvement les différentes variantes possibles de l'algorithme en fonction de la stratégie d'échantillonnage utilisée.

## 5.2 L'algorithme de filtrage particulaire à échantillonnage global

### 5.2.1 Hypothèses

Soit  $\mathcal{A}$  un ensemble fini<sup>1</sup>. On considère une suite  $(F_t)_{t \in \mathbb{N}}$  de mesures de probabilité telles que

1. pour tout  $t$ ,  $F_t$  est une mesure de probabilité sur l'espace mesurable  $(\mathcal{A}, \mathcal{P}(\mathcal{A}))^{\otimes(t+1)}$
2. pour tout  $t \geq 1$ , il existe une transition finie  $Q_t : (\mathcal{A}, \mathcal{P}(\mathcal{A}))^{\otimes t} \prec (\mathcal{A}, \mathcal{P}(\mathcal{A}))$  telle que

$$F_t = F_{t-1} \otimes Q_t \quad (5.1)$$

### 5.2.2 Description de l'algorithme

#### 5.2.2.1 Initialisation de l'algorithme

Pour tout  $j \in \{1, \dots, N\}$ , on définit la variable aléatoire  $\Lambda^{(j,0)}$  à valeurs dans  $\mathcal{A}$  et correspondant à la  $j$ -ème trajectoire à l'instant initial par

$$\Lambda^{(j,0)} \sim F_0 \quad (5.2)$$

et par la condition, pour  $j > 1$ , que  $\Lambda^{(j,0)}$  est indépendant de  $\Lambda^{(1,0)}, \dots, \Lambda^{(j-1,0)}$ .

On définit la mesure de probabilité aléatoire  $F_0^N$  sur  $(\mathcal{A}, \mathcal{P}(\mathcal{A}))$  par

$$F_0^N \stackrel{\text{déf}}{=} \frac{1}{N} \sum_{j=1}^N \delta_{\Lambda^{(j,0)}} \quad (5.3)$$

#### 5.2.2.2 Passage de l'instant $t - 1$ à l'instant $t$

Soit  $t \geq 1$ . Pour tout  $j \in \{1, \dots, N\}$ , on suppose définie la variable aléatoire  $\Lambda^{(j,t-1)} = (\Lambda_0^{(j,t-1)}, \Lambda_1^{(j,t-1)}, \dots, \Lambda_{t-1}^{(j,t-1)})$  à valeurs dans  $\mathcal{A}^t$  et correspondant à la  $j$ -ème trajectoire à l'instant  $t - 1$ .

On définit la mesure de probabilité aléatoire  $F_{t-1}^N$  sur  $(\mathcal{A}, \mathcal{P}(\mathcal{A}))^{\otimes t}$  par

$$F_{t-1}^N \stackrel{\text{déf}}{=} \frac{1}{N} \sum_{j=1}^N \delta_{\Lambda^{(j,t-1)}} \quad (5.4)$$

---

<sup>1</sup>Nous supposons ici que l'ensemble  $\mathcal{A}$  est fini afin de simplifier la présentation du principe de notre algorithme. Cependant, cet algorithme peut être utilisé avec des espaces quelconques (cf la remarque 5.1, p. 78).

et on définit la mesure de probabilité aléatoire  $\tilde{F}_t^N$  sur  $(\mathcal{A}, \mathcal{P}(\mathcal{A}))^{\otimes(t+1)}$  par

$$\tilde{F}_t^N \stackrel{\text{déf}}{=} \alpha_t^N F_{t-1}^N \otimes Q_t \quad (5.5)$$

avec

$$\alpha_t^N \stackrel{\text{déf}}{=} \left\{ [F_{t-1}^N \otimes Q_t](\mathcal{A}^{t+1}) \right\}^{-1} \quad (5.6)$$

$$= \left\{ \frac{1}{N} \sum_{j=1}^N Q_t(\Lambda^{(j,t-1)}, \mathcal{A}) \right\}^{-1} \quad (5.7)$$

Pour tout  $j \in \{1, \dots, N\}$ , on définit la variable aléatoire  $\Lambda^{(j,t)} = (\Lambda_0^{(j,t)}, \Lambda_1^{(j,t)}, \dots, \Lambda_t^{(j,t)})$  à valeurs dans  $\mathcal{A}^{t+1}$  et correspondant à la  $j$ -ème trajectoire à l'instant  $t$  par

$$\Lambda^{(j,t)} \sim \tilde{F}_t^N \quad \text{conditionnellement à } \Lambda^{(1,t-1)}, \dots, \Lambda^{(N,t-1)} \quad (5.8)$$

et par la condition que  $\Lambda^{(j,t)}$  doit vérifier les relations d'indépendance conditionnelle représentées par le réseau bayésien de la figure 5.1. On notera en particulier que, pour  $j > 1$ ,  $\Lambda^{(j,t)}$  est indépendant de  $\Lambda^{(1,t)}, \dots, \Lambda^{(j-1,t)}$  conditionnellement à  $\Lambda^{(1,t-1)}, \dots, \Lambda^{(N,t-1)}$ .

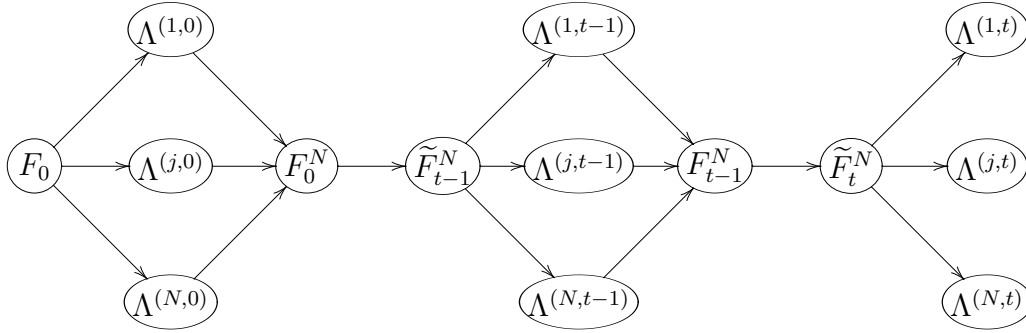


FIG. 5.1 – Réseau bayésien représentant les relations d'indépendance conditionnelle que doivent vérifier les  $\Lambda^{(j,t)}$ .

Enfin, on définit la mesure de probabilité aléatoire  $F_t^N$  sur  $(\mathcal{A}, \mathcal{P}(\mathcal{A}))^{\otimes(t+1)}$  par

$$F_t^N \stackrel{\text{déf}}{=} \frac{1}{N} \sum_{j=1}^N \delta_{\Lambda^{(j,t)}} \quad (5.9)$$

On convient d'appeler *algorithme de filtrage particulière à échantillonnage global* l'algorithme ci-dessus qui, à partir de la mesure initiale  $F_0$  et des transitions finies  $Q_t$ , permet d'obtenir une mesure  $F_t^N$  approchant la mesure  $F_t$ .

Le réseau bayésien correspondant aux différents éléments aléatoires utilisés par l'algorithme de filtrage particulière à échantillonnage global a été représenté à la figure 5.2.

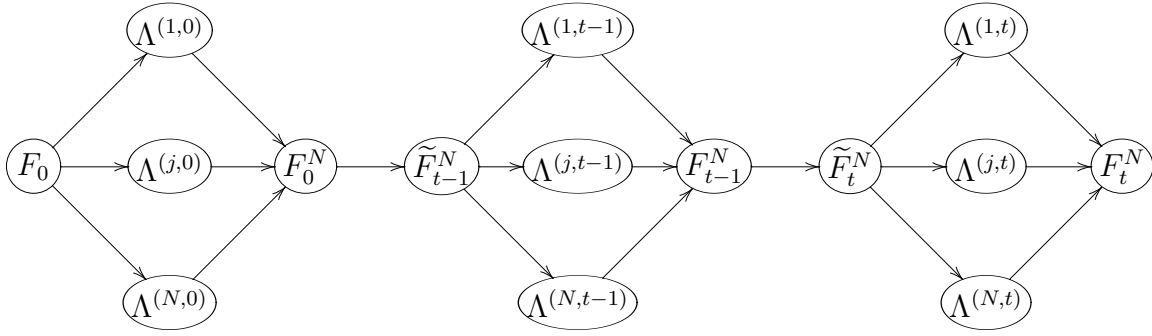


FIG. 5.2 – Réseau bayésien correspondant aux différents éléments aléatoires utilisés par l'algorithme de filtrage particulaire à échantillonnage global.

### Remarque 5.1

Comme nous avons supposé ici que l'ensemble  $\mathcal{A}$  est fini, la simulation des variables aléatoires  $\Lambda^{(j,t)}$  par tirage aléatoire est toujours possible lorsque l'on dispose de la mesure de probabilité  $\tilde{F}_t^N$ . Dans le cas où l'ensemble  $\mathcal{A}$  est quelconque, il peut arriver que l'on ne puisse pas générer aisément des échantillons de la mesure de probabilité  $\tilde{F}_t^N$ . Dans ce cas, on peut remplacer les transitions finies  $Q_t$  par des transitions finies  $Q'_t$  telles que

1. la mesure de probabilité  $F_t$  est absolument continue par rapport à la mesure  $F_{t-1} \otimes Q'_t$  ; autrement dit, la mesure  $F_t$  peut s'écrire sous la forme suivante

$$F_t = g_t \cdot F_{t-1} \otimes Q'_t \quad (5.10)$$

où  $g_t$  désigne la dérivée de Radon-Nikodym de  $F_t$  par rapport à  $F_{t-1} \otimes Q'_t$  ;

2. pour tout  $\lambda \in \mathcal{A}^t$ , il est possible de générer des échantillons de la mesure de probabilité  $\alpha(\lambda) Q'_t(\lambda; \cdot)$  avec  $\alpha(\lambda) = 1/Q'_t(\lambda; \mathcal{A})$ .

Si on dispose d'une approximation  $F_{t-1}^N = \sum_{j=1}^N w^{(j,t-1)} \delta_{\Lambda^{(j,t-1)}}$  de  $F_{t-1}$ , on peut alors obtenir une approximation  $F_t^N$  de  $F_t$  de la façon suivante :

- on définit la mesure de probabilité aléatoire  $\tilde{F}_t^N$  par

$$\tilde{F}_t^N \stackrel{\text{déf}}{=} \alpha_t^N F_{t-1}^N \otimes Q'_t \quad (5.11)$$

avec

$$\alpha_t^N \stackrel{\text{déf}}{=} \left\{ [F_{t-1}^N \otimes Q'_t](\mathcal{A}^{t+1}) \right\}^{-1} \quad (5.12)$$

$$= \left\{ \sum_{j=1}^N w^{(j,t-1)} Q'_t(\Lambda^{(j,t-1)}, \mathcal{A}) \right\}^{-1} \quad (5.13)$$

- pour tout  $j \in \{1, \dots, N\}$ , on définit la variable aléatoire  $\Lambda^{(j,t)} = (\Lambda_0^{(j,t)}, \Lambda_1^{(j,t)}, \dots, \Lambda_t^{(j,t)})$  à valeurs dans  $\mathcal{A}^{t+1}$  et correspondant à la  $j$ -ème trajectoire à l'instant  $t$  par

$$\Lambda^{(j,t)} \sim \tilde{F}_t^N \quad \text{conditionnellement à } \Lambda^{(1,t-1)}, \dots, \Lambda^{(N,t-1)} \quad (5.14)$$

et par la condition que  $\Lambda^{(j,t)}$  doit vérifier les relations d'indépendance conditionnelle représentées par le réseau bayésien de la figure 5.1.

– on définit la mesure de probabilité aléatoire  $F_t^N$  par

$$F_t^N \stackrel{\text{déf}}{=} \sum_{j=1}^N w^{(j,t)} \delta_{\Lambda^{(j,t)}} \quad (5.15)$$

avec

$$w^{(j,t)} \stackrel{\text{déf}}{=} \frac{g_t(\Lambda^{(j,t)})}{\sum_{k=1}^N g_t(\Lambda^{(k,t)})} \quad (5.16)$$

Comme la mesure  $\tilde{F}_t^N$  peut s'écrire de la façon suivante

$$\tilde{F}_t^N = \alpha_t^N \sum_{j=1}^N w^{(j,t-1)} \delta_{\Lambda^{(j,t-1)}} \otimes Q'_t \quad (5.17)$$

$$= \sum_{j=1}^N \tilde{w}^{(j,t-1)} \delta_{\Lambda^{(j,t-1)}} \otimes (\alpha^{(j,t-1)} Q'_t) \quad (5.18)$$

avec  $\alpha^{(j,t-1)} = 1/Q'_t(\Lambda^{(j,t-1)}; \mathcal{A})$  et  $\tilde{w}^{(j,t-1)} = \alpha_t^N w^{(j,t-1)} / \alpha^{(j,t-1)}$ , on peut générer la variable aléatoire  $\Lambda^{(j,t)} = (\Lambda_0^{(j,t)}, \Lambda_1^{(j,t)}, \dots, \Lambda_t^{(j,t)})$  en générant tout d'abord  $\Lambda_{0:t-1}^{(j,t)}$  selon la loi  $\sum_{i=1}^N \tilde{w}^{(i,t-1)} \delta_{\Lambda^{(i,t-1)}}$ , puis en générant  $\Lambda_t^{(j,t)}$  selon la loi  $\alpha(\Lambda_{0:t-1}^{(j,t)}) Q'_t(\Lambda_{0:t-1}^{(j,t)}; \cdot)$ .

### 5.2.3 Originalité de l'algorithme

Classiquement, un algorithme de filtrage particulière comme l'algorithme *Sequential Importance Sampling and Resampling* (SISR) décrit par exemple dans [LCL01, AMGC02, DGA00] comporte deux étapes. Au cours de la première étape, dite étape d'échantillonnage pondéré (*importance sampling step*), chaque trajectoire est prolongée à l'aide d'une fonction de proposition et est affectée d'un poids. Au cours de la deuxième étape, dite étape de rééchantillonnage (*resampling step*), la taille effective de l'échantillon  $N_{\text{eff}}$  est calculée et comparée au nombre  $N$  de particules ; lorsque le rapport  $N_{\text{eff}}/N$  est inférieur à un seuil arbitraire, les trajectoires prolongées sont rééchantillonnées en prenant en compte les poids calculés à l'étape d'échantillonnage.

L'algorithme que nous proposons ici utilise un principe différent. A chaque itération, l'algorithme proposé considère la population formée par toutes les prolongations possibles de toutes les trajectoires et sélectionne les nouvelles trajectoires en échantillonnant cette population. Nous obtenons ainsi un algorithme qui est plus simple à mettre en œuvre que l'algorithme SISR et dont les performances se sont avérées supérieures ou égales à celles de l'algorithme SISR au cours des simulations numériques que nous avons effectuées et qui seront présentées plus loin dans le chapitre 7.

## 5.3 Convergence de l'erreur quadratique moyenne

### Théorème 5.1

Pour tout  $t \in \mathbb{N}$ , il existe  $c_t \in \mathbb{R}^+$  tel que, pour tout  $N \in \mathbb{N}^*$  et pour toute fonction  $\varphi$  de  $\mathcal{A}^{t+1}$  dans  $\mathbb{R}$ , on a

$$\mathbb{E}^P \left\{ [F_t^N \varphi - F_t \varphi]^2 \right\} \leq c_t \frac{\|\varphi\|^2}{N} \quad (5.19)$$



avec  $\|\varphi\| = \max_{\lambda \in \mathcal{A}^{t+1}} \varphi(\lambda)$ .

DÉMONSTRATION :

La démonstration se fait par récurrence sur  $t$ .

Montrons que la proposition est vraie pour  $t = 0$ . Soient  $N \in \mathbb{N}^*$  et  $\varphi$  une fonction de  $\mathcal{A}$  dans  $\mathbb{R}$ . Calculons tout d'abord  $\mathbb{E}^P \{F_0^N \varphi\}$

$$\mathbb{E}^P \{F_0^N \varphi\} = \mathbb{E}^P \left\{ \frac{1}{N} \sum_{j=1}^N \varphi(\Lambda^{(j,0)}) \right\} \quad (5.20)$$

$$= \frac{1}{N} \sum_{j=1}^N \mathbb{E}^P \{ \varphi(\Lambda^{(j,0)}) \} \quad (5.21)$$

$$= \frac{1}{N} \sum_{j=1}^N F_0 \varphi \quad (5.22)$$

$$= F_0 \varphi \quad (5.23)$$

Calculons également  $\mathbb{E}^P \{[F_0^N \varphi]^2\}$

$$\mathbb{E}^P \{[F_0^N \varphi]^2\} = \mathbb{E}^P \left\{ \frac{1}{N^2} \sum_{\substack{1 \leq j \leq N \\ 1 \leq j' \leq N}} \varphi(\Lambda^{(j,0)}) \varphi(\Lambda^{(j',0)}) \right\} \quad (5.24)$$

$$= \frac{1}{N^2} \sum_{j=1}^N \mathbb{E}^P \{ \varphi^2(\Lambda^{(j,0)}) \} + \frac{1}{N^2} \sum_{\substack{1 \leq j \leq N \\ 1 \leq j' \leq N \\ j \neq j'}} \mathbb{E}^P \{ \varphi(\Lambda^{(j,0)}) \varphi(\Lambda^{(j',0)}) \} \quad (5.25)$$

$$= \frac{1}{N^2} \sum_{j=1}^N F_0[\varphi^2] + \frac{1}{N^2} \sum_{\substack{1 \leq j \leq N \\ 1 \leq j' \leq N \\ j \neq j'}} [F_0 \varphi]^2 \quad (5.26)$$

$$= \frac{1}{N} F_0[\varphi^2] + \left(1 - \frac{1}{N}\right) [F_0 \varphi]^2 \quad (5.27)$$

On en déduit

$$\mathbb{E}^P \{[F_0^N \varphi - F_0 \varphi]^2\} = \mathbb{E}^P \{[F_0^N \varphi - \mathbb{E}^P \{F_0^N \varphi\}]^2\} \quad (5.28)$$

$$= \mathbb{E}^P \{[F_0^N \varphi]^2\} - [\mathbb{E}^P \{F_0^N \varphi\}]^2 \quad (5.29)$$

$$= \frac{1}{N} F_0[\varphi^2] + \left(1 - \frac{1}{N}\right) [F_0 \varphi]^2 - [F_0 \varphi]^2 \quad (5.30)$$

$$= \frac{1}{N} (F_0[\varphi^2] - [F_0 \varphi]^2) \quad (5.31)$$

$$\leq \frac{1}{N} F_0[\varphi^2] \quad (5.32)$$

$$\leq \frac{1}{N} \|\varphi\|^2 \quad (5.33)$$

ce qui montre, en prenant  $c_0 = 1$ , que la proposition est vraie pour  $t = 0$ .

Supposons maintenant que la proposition est vraie à l'ordre  $t - 1$ . Soient  $N \in \mathbb{N}^*$  et  $\varphi$  une fonction de  $\mathcal{A}^{t+1}$  dans  $\mathbb{R}$ . L'inégalité de Minkowski s'écrit

$$\mathbb{E}^P \left\{ [F_t^N \varphi - F_t \varphi]^2 \right\}^{1/2} \leq \mathbb{E}^P \left\{ [F_t^N \varphi - \tilde{F}_t^N \varphi]^2 \right\}^{1/2} + \mathbb{E}^P \left\{ [\tilde{F}_t^N \varphi - F_t \varphi]^2 \right\}^{1/2} \quad (5.34)$$

Majorons tout d'abord le terme  $\mathbb{E}^P \left\{ [F_t^N \varphi - \tilde{F}_t^N \varphi]^2 \right\}^{1/2}$ . Soit  $\mathcal{G}_{t-1}$  la tribu engendrée par  $\Lambda^{(1,t-1)}, \Lambda^{(2,t-1)}, \dots, \Lambda^{(N,t-1)}$ . On démontre, en procédant comme précédemment, que l'on a

$$\mathbb{E}^P \left\{ [F_t^N \varphi - \tilde{F}_t^N \varphi]^2 \middle| \mathcal{G}_{t-1} \right\} = \frac{1}{N} (\tilde{F}_t^N [\varphi^2] - [F_t^N \varphi]^2) \quad (5.35)$$

$$\leq \frac{1}{N} \tilde{F}_t^N [\varphi^2] \quad (5.36)$$

$$\leq \frac{1}{N} \|\varphi\|^2 \quad (5.37)$$

et on en déduit

$$\mathbb{E}^P \left\{ [F_t^N \varphi - \tilde{F}_t^N \varphi]^2 \right\}^{1/2} = \mathbb{E}^P \left\{ \mathbb{E}^P \left\{ [F_t^N \varphi - \tilde{F}_t^N \varphi]^2 \middle| \mathcal{G}_{t-1} \right\} \right\}^{1/2} \quad (5.38)$$

$$\leq \frac{1}{\sqrt{N}} \|\varphi\| \quad (5.39)$$

Majorons maintenant le terme  $\mathbb{E}^P \left\{ [\tilde{F}_t^N \varphi - F_t \varphi]^2 \right\}^{1/2}$ . En utilisant les équations 5.1 et 5.5, on obtient

$$\mathbb{E}^P \left\{ [\tilde{F}_t^N \varphi - F_t \varphi]^2 \right\}^{1/2} = \mathbb{E}^P \left\{ [(\alpha_t^N \cdot F_{t-1}^N \otimes Q_t) \varphi - (F_{t-1} \otimes Q_t) \varphi]^2 \right\}^{1/2} \quad (5.40)$$

En utilisant l'inégalité de Minkowski, l'équation 5.40 implique

$$\begin{aligned} \mathbb{E}^P \left\{ [\tilde{F}_t^N \varphi - F_t \varphi]^2 \right\}^{1/2} &\leq \mathbb{E}^P \left\{ [\alpha_t^N (F_{t-1}^N \otimes Q_t) \varphi - (F_{t-1}^N \otimes Q_t) \varphi]^2 \right\}^{1/2} + \dots \\ &\dots + \mathbb{E}^P \left\{ [(F_{t-1}^N \otimes Q_t) \varphi - (F_{t-1} \otimes Q_t) \varphi]^2 \right\}^{1/2} \end{aligned} \quad (5.41)$$

Majorons le terme  $\mathbb{E}^P \left\{ [\alpha_t^N (F_{t-1}^N \otimes Q_t) \varphi - (F_{t-1}^N \otimes Q_t) \varphi]^2 \right\}^{1/2}$ . On remarque que l'on a

$$\left| \alpha_t^N (F_{t-1}^N \otimes Q_t) \varphi - (F_{t-1}^N \otimes Q_t) \varphi \right| = \left| \alpha_t^N (F_{t-1}^N \otimes Q_t) \varphi \right| \left| 1 - 1/\alpha_t^N \right| \quad (5.42)$$

$$= \left| \tilde{F}_t^N \varphi \right| \cdot \left| F_t(\mathcal{A}^{t+1}) - [F_{t-1}^N \otimes Q_t](\mathcal{A}^{t+1}) \right| \quad (5.43)$$

$$= \left| \tilde{F}_t^N \varphi \right| \cdot \left| [F_{t-1} \otimes Q_t](\mathcal{A}^{t+1}) - [F_{t-1}^N \otimes Q_t](\mathcal{A}^{t+1}) \right| \quad (5.44)$$

$$\leq \|\varphi\| \cdot \left| [F_{t-1} \otimes Q_t](\mathcal{A}^{t+1}) - [F_{t-1}^N \otimes Q_t](\mathcal{A}^{t+1}) \right| \quad (5.45)$$

$$\leq \|\varphi\| \cdot \left| F_{t-1} \psi - F_{t-1}^N \psi \right| \quad (5.46)$$

où l'on a posé, pour tout  $\lambda \in \mathcal{A}^t$ ,

$$\psi(\lambda) \stackrel{\text{déf}}{=} Q_t(\lambda, \mathcal{A}) \quad (5.47)$$

On en déduit

$$\mathbb{E}^P \left\{ [\alpha_t^N (F_{t-1}^N \otimes Q_t) \varphi - (F_{t-1}^N \otimes Q_t) \varphi]^2 \right\}^{1/2} \leq \|\varphi\| \mathbb{E}^P \left\{ [F_{t-1} \psi - F_{t-1}^N \psi]^2 \right\}^{1/2} \quad (5.48)$$

L'hypothèse de récurrence étant supposée vraie à l'ordre  $t-1$ , il existe donc  $c_{t-1} \in \mathbb{R}^+$  tel que

$$\mathbb{E}^P \left\{ [\alpha_t^N (F_{t-1}^N \otimes Q_t) \varphi - (F_{t-1}^N \otimes Q_t) \varphi]^2 \right\}^{1/2} \leq \sqrt{\frac{c_{t-1}}{N}} \|\psi\| \|\varphi\| \quad (5.49)$$

avec  $\|\psi\| = \sup_{\lambda \in \mathcal{A}^t} Q_t(\lambda, \mathcal{A})$ .

Majorons le terme  $\mathbb{E}^P \left\{ [(F_{t-1}^N \otimes Q_t) \varphi - (F_{t-1} \otimes Q_t) \varphi]^2 \right\}^{1/2}$ . Ce terme s'écrit

$$\mathbb{E}^P \left\{ [(F_{t-1}^N \otimes Q_t) \varphi - (F_{t-1} \otimes Q_t) \varphi]^2 \right\}^{1/2} = \mathbb{E}^P \left\{ [F_{t-1}^N \psi' - F_{t-1} \psi']^2 \right\}^{1/2} \quad (5.50)$$

où l'on a posé, pour tout  $\lambda \in \mathcal{A}^t$ ,

$$\psi'(\lambda) \stackrel{\text{déf}}{=} \int \varphi(\lambda, \lambda_t) Q_t(\lambda, d\lambda_t) \quad (5.51)$$

En utilisant l'hypothèse de récurrence à l'ordre  $t-1$ , on en déduit

$$\mathbb{E}^P \left\{ [(F_{t-1}^N \otimes Q_t) \varphi - (F_{t-1} \otimes Q_t) \varphi]^2 \right\}^{1/2} \leq \sqrt{\frac{c_{t-1}}{N}} \|\psi'\| \quad (5.52)$$

$$\leq \sqrt{\frac{c_{t-1}}{N}} \|\psi\| \|\varphi\| \quad (5.53)$$

Les inégalités 5.34, 5.39, 5.41, 5.49 et 5.53 impliquent donc

$$\mathbb{E}^P \left\{ [F_t^N \varphi - F_t \varphi]^2 \right\}^{1/2} \leq \frac{1 + 2\sqrt{c_{t-1}} \|\psi\|}{\sqrt{N}} \|\varphi\| \quad (5.54)$$

et donc

$$\mathbb{E}^P \left\{ [F_t^N \varphi - F_t \varphi]^2 \right\} \leq c_t \frac{\|\varphi\|^2}{N} \quad (5.55)$$

avec  $c_t = (1 + 2\sqrt{c_{t-1}} \|\psi\|)^2$  ce qui montre que la proposition est vraie à l'ordre  $t$ .  $\square$

## 5.4 Variantes de l'algorithme de filtrage particulaire à échantillonnage global

Afin de simplifier la présentation de l'algorithme de filtrage particulaire à échantillonnage global, nous avons supposé que la mesure  $F_t^N$  était déduite de la mesure  $\tilde{F}_t^N$  à l'aide de la stratégie d'échantillonnage caractérisée par l'équation 5.8. Dans la littérature consacrée au filtrage particulaire, cette stratégie d'échantillonnage est appelée *stratégie d'échantillonnage multinomiale*. D'autres stratégies d'échantillonnage existent dans la littérature. A titre d'exemple, nous présenterons ci-dessous la stratégie d'échantillonnage résiduelle qui a été originalement proposée par Liu et Chen [LC98] afin de réduire la variance introduite par l'opération d'échantillonnage.

### 5.4.1 Opérateur d'échantillonnage

Avant de présenter la stratégie d'échantillonnage résiduelle, nous allons tout d'abord définir plus précisément la notion d'opérateur d'échantillonnage.

#### Définition 5.1

Soient  $(\Omega, \mathcal{F}, P)$  un espace de probabilité,  $(E, \mathcal{E})$  un espace mesurable et  $k$  et  $n$  des entiers strictement positifs. On appelle **opérateur d'échantillonnage** de paramètre  $(k, n)$  sur l'espace mesurable  $(E, \mathcal{E})$  tout opérateur  $\Xi$  qui, à toute mesure de probabilité  $\mu$  sur  $(E, \mathcal{E})$  de la forme  $\mu = \sum_{i=1}^k w_i \mathbb{1}_{\Lambda_i}$  avec  $w_i \in \mathbb{R}^+$  et  $\Lambda_i \in E$ , associe une mesure de probabilité aléatoire  $\Xi(\mu)$  de la forme suivante

$$\Xi(\mu) = \frac{1}{n} \sum_{i=1}^k N_i \mathbb{1}_{\Lambda_i} \quad (5.56)$$

où les  $N_i$  sont des variables aléatoires sur  $(\Omega, \mathcal{F}, P)$  et à valeurs dans  $\{0, 1, \dots, n\}$  telles que

$$\frac{1}{n} \sum_{i=1}^k N_i = 1 \quad (5.57)$$

Lorsque les  $N_i$  vérifient également la condition suivante

$$\forall i \in \{1, \dots, k\} \quad \mathbb{E}^P(N_i) = n w_i \quad (5.58)$$

on dit que l'opérateur d'échantillonnage  $\Xi$  est **non-biaisé**.

### 5.4.2 Opérateur d'échantillonnage multinomial

#### Définition 5.2

On appelle **opérateur d'échantillonnage multinomial** de paramètre  $(k, n)$  sur l'espace mesurable  $(E, \mathcal{E})$  l'opérateur  $\Xi_M$  qui, à toute mesure de probabilité  $\mu$  sur  $(E, \mathcal{E})$  de la forme  $\mu = \sum_{i=1}^k w_i \mathbb{1}_{\Lambda_i}$  avec  $w_i \in \mathbb{R}^+$  et  $\Lambda_i \in E$ , associe la mesure de probabilité aléatoire  $\Xi_M(\mu)$  vérifiant

$$\Xi_M(\mu) = \frac{1}{n} \sum_{i=1}^k N_i \mathbb{1}_{\Lambda_i} \quad (5.59)$$

où  $(N_1, \dots, N_k)$  est une variable aléatoire multinomiale d'ordre  $n$  et de paramètre  $(w_1, \dots, w_k)$ .

On notera que l'opérateur d'échantillonnage multinomial est non-biaisé. On remarquera également que la mesure de probabilité aléatoire  $F_t^N$  définie à l'équation 5.9 et liée à la mesure de probabilité aléatoire  $\tilde{F}_t^N$  par l'équation 5.8, vérifie

$$F_t^N = \Xi_M^N(\tilde{F}_t^N) \quad (5.60)$$

où  $\Xi_M^N$  désigne l'opérateur d'échantillonnage multinomial de paramètre  $(N \text{ card}(\mathcal{A}), N)$  sur l'espace mesurable  $(\mathcal{A}, \mathcal{P}(\mathcal{A}))^{\otimes(t+1)}$ .

### 5.4.3 Opérateur d'échantillonnage résiduel

#### Définition 5.3

On appelle **opérateur d'échantillonnage résiduel** de paramètre  $(k, n)$  sur l'espace mesurable  $(E, \mathcal{E})$  l'opérateur  $\Xi_R$  qui, à toute mesure de probabilité  $\mu$  sur  $(E, \mathcal{E})$  de la forme  $\mu = \sum_{i=1}^k w_i \mathbb{1}_{\Lambda_i}$  avec  $w_i \in \mathbb{R}^+$  et  $\Lambda_i \in E$ , associe la mesure de probabilité aléatoire  $\Xi_R(\mu)$  vérifiant

$$\Xi_R(\mu) = \frac{1}{n} \sum_{i=1}^k N_i \mathbb{1}_{\Lambda_i} \quad (5.61)$$

$$N_i = \lfloor nw_i \rfloor + R_i \quad (5.62)$$

où  $(R_1, \dots, R_k)$  est une variable aléatoire multinomiale d'ordre  $\tilde{n}$  et de paramètre  $(\tilde{w}_1, \dots, \tilde{w}_k)$  avec

$$\tilde{n} = n - \sum_{i=1}^k \lfloor nw_i \rfloor \quad (5.63)$$

$$\tilde{w}_i = \frac{1}{\tilde{n}}(nw_i - \lfloor nw_i \rfloor) \quad (5.64)$$

On vérifiera aisément que l'opérateur d'échantillonnage résiduel est non-biaisé. En outre, on montre facilement que le théorème 5.1 reste valide lorsque l'on remplace les équations 5.8 et 5.9 par

$$F_t^N \stackrel{\text{déf}}{=} \Xi_R^N(\tilde{F}_t^N) \quad (5.65)$$

où  $\Xi_R^N$  désigne l'opérateur d'échantillonnage résiduel de paramètre  $(N \text{ card}(\mathcal{A}), N)$  sur l'espace mesurable  $(\mathcal{A}, \mathcal{P}(\mathcal{A}))^{\otimes(t+1)}$ . Autrement dit, la convergence de l'erreur quadratique moyenne du filtre particulaire à échantillonnage global est toujours assurée lorsque l'on utilise une stratégie d'échantillonnage résiduelle à la place d'une stratégie d'échantillonnage multinomiale.

## 5.5 Conclusion

Dans ce chapitre, nous avons présenté l'algorithme de filtrage particulaire à échantillonnage global. Nous avons ensuite montré que l'erreur quadratique moyenne de l'algorithme converge vers 0 lorsque le nombre de particules tend vers l'infini. Enfin, nous avons défini la notion d'opérateur d'échantillonnage et nous avons présenté brièvement les différentes variantes possibles de l'algorithme de filtrage particulaire à échantillonnage global en fonction de l'opérateur d'échantillonnage utilisé.

Dans le chapitre suivant, nous allons étudier l'application de cet algorithme à l'estimation statistique dans le cas d'un modèle dynamique à processus indicateur latent.

# Chapitre 6

## Application aux modèles dynamiques à processus indicateur latent

### 6.1 Introduction

Dans le chapitre précédent, nous avons présenté de façon très générale l'algorithme de filtrage particulière à échantillonnage global. Nous avons vu que cet algorithme permet de calculer de façon approchée une suite de mesures de probabilité. Dans ce chapitre, nous allons étudier l'application de cet algorithme lorsque la suite de mesures de probabilité que l'on souhaite approcher correspond à une suite de lois *a posteriori* possédant une structure particulière.

Pour cela, nous allons tout d'abord introduire le concept de *modèle dynamique à processus indicateur latent*. Ce concept nous sera très utile dans la suite car il nous permettra de traiter à l'aide d'un formalisme commun les différentes situations que nous avons rencontrées au chapitre 4. Nous définirons ensuite les notions d'opérateurs de *prédiction de l'état*, de *prédiction de l'observation* et de *correction de l'état* d'un modèle dynamique à processus indicateur latent. Nous pourrons alors appliquer l'algorithme de filtrage particulière à échantillonnage global aux modèles dynamiques à processus indicateur latent. Selon la loi *a posteriori* que l'on souhaite approcher, nous obtiendrons ainsi deux algorithmes d'estimation, que nous appellerons respectivement *filtre particulière à échantillonnage global immédiat* et *filtre particulière à échantillonnage global différé*, et que nous décrirons en détail à l'aide des opérateurs de prédiction et de correction définis précédemment.

### 6.2 Les modèles dynamiques à processus indicateur latent

#### 6.2.1 Définition d'un modèle dynamique à processus indicateur latent

On considère un espace de probabilité  $(\Omega, \mathcal{F}, P)$  et les processus stochastiques  $(\Lambda_t)_{t \in \mathbb{N}}$ ,  $(X_t)_{t \in \mathbb{N}}$  et  $(Y_t)_{t \in \mathbb{N}}$  sur  $(\Omega, \mathcal{F})$ , vérifiant, pour tout  $t \in \mathbb{N}$ , les relations d'indépendance conditionnelle représentées par le réseau bayésien de la figure 6.1.

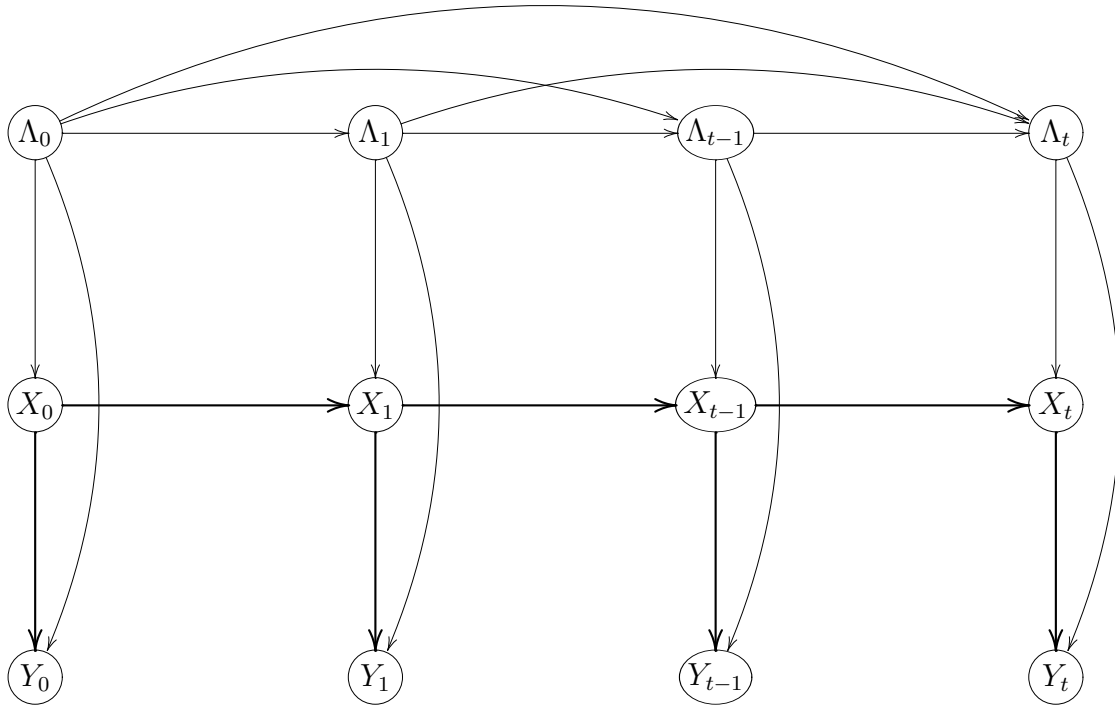


FIG. 6.1 – Réseau bayésien sur  $\{\Lambda_0, X_0, Y_0, \dots, \Lambda_t, X_t, Y_t\}$  pour la mesure de probabilité  $P$ .

On suppose également que les  $\Lambda_t$  prennent leurs valeurs dans un ensemble  $\mathcal{A}$  avec  $\text{card}(\mathcal{A}) < \infty$ .

On dit alors que le processus stochastique  $(\Lambda_t)_{t \in \mathbb{N}}$  est un **processus indicateur latent** et que les processus stochastiques  $(\Lambda_t)_{t \in \mathbb{N}}$ ,  $(X_t)_{t \in \mathbb{N}}$  et  $(Y_t)_{t \in \mathbb{N}}$  constituent un **modèle dynamique à processus indicateur latent**.

## 6.2.2 Opérateurs agissant sur un modèle dynamique à processus indicateur latent

Grâce aux relations d'indépendance conditionnelle qui existent au sein d'un modèle dynamique à processus indicateur latent, nous pouvons définir plusieurs opérateurs qui nous permettront de calculer les lois conditionnelles dont nous aurons besoin pour appliquer l'algorithme de filtrage particulaire à échantillonnage global à un modèle dynamique à processus indicateur latent.

### 6.2.2.1 Opérateur de prédiction de l'état

On peut déterminer la transition  $\Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t-1})}$  à partir de la transition  $\Pi_{X_{t-1}}^{(\Lambda_{0:t-1}, Y_{0:t-1})}$  en appliquant l'algorithme de propagation de croyance conditionnelle (cf. l'annexe D) au réseau bayésien de la figure 6.1 en prenant comme observations  $\{Y_0, Y_1, \dots, Y_{t-1}\}$  et comme ensemble de conditionnement  $\{\Lambda_0, \Lambda_1, \dots, \Lambda_t\}$  (cf figure 6.2).

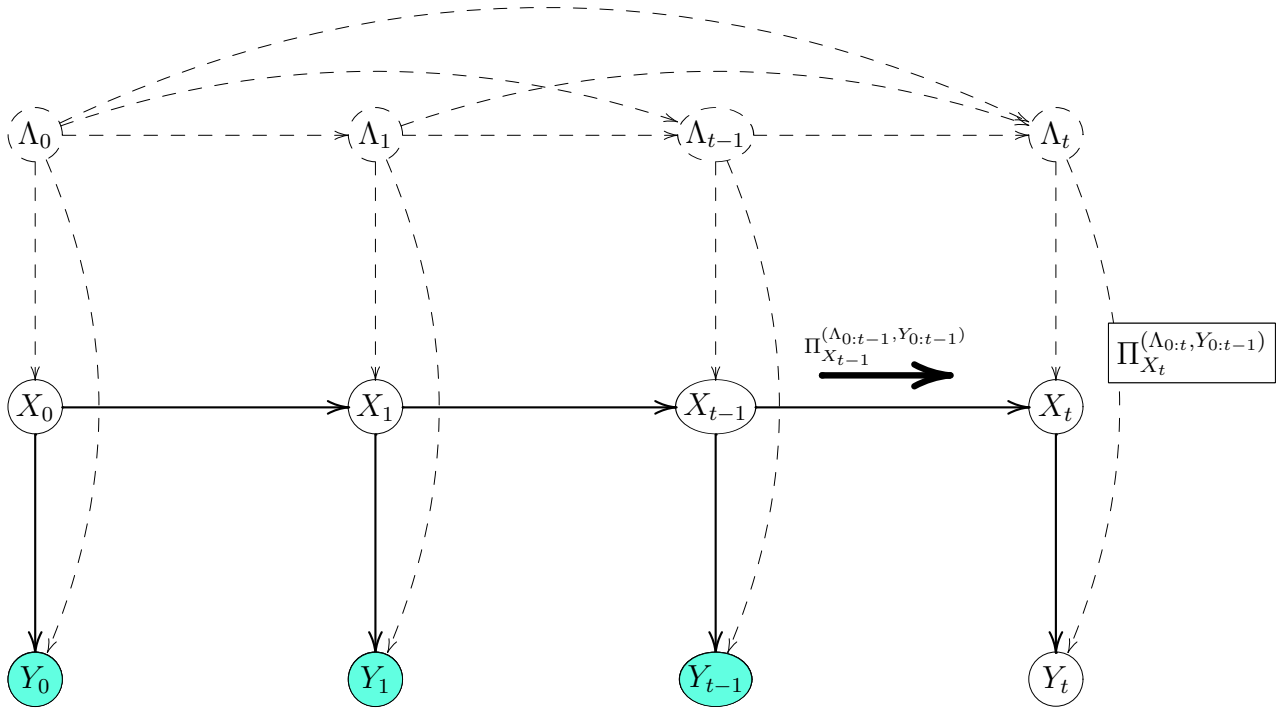


FIG. 6.2 – Calcul de la transition  $\Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t-1})}$  à partir de la transition  $\Pi_{X_{t-1}}^{(\Lambda_{0:t-1}, Y_{0:t-1})}$  par propagation de croyance conditionnelle, les observations étant  $\{Y_0, Y_1, \dots, Y_{t-1}\}$  et l'ensemble de conditionnement  $\{\Lambda_0, \Lambda_1, \dots, \Lambda_t\}$ .

On obtient alors l'équation suivante

$$\Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\lambda_{0:t}, y_{0:t-1}; \cdot) = \sum_{X_{t-1}} \Pi_{X_{t-1}}^{(\Lambda_{0:t-1}, Y_{0:t-1})}(\lambda_{0:t-1}, y_{0:t-1}; \cdot) \otimes \Pi_{X_t}^{(\Lambda_t, X_{t-1})}(\lambda_t, \cdot; \cdot) \quad (6.1)$$

Il est donc possible de définir un opérateur  $\Psi_{t|t-1}^{\lambda_t}$  permettant d'obtenir le terme  $\Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\lambda_{0:t}, y_{0:t-1}; \cdot)$  à partir du terme  $\Pi_{X_{t-1}}^{(\Lambda_{0:t-1}, Y_{0:t-1})}(\lambda_{0:t-1}, y_{0:t-1}; \cdot)$ . Cet opérateur est défini par la relation suivante

$$\Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\lambda_{0:t}, y_{0:t-1}; \cdot) = \Psi_{t|t-1}^{\lambda_t} \left( \Pi_{X_{t-1}}^{(\Lambda_{0:t-1}, Y_{0:t-1})}(\lambda_{0:t-1}, y_{0:t-1}; \cdot) \right) \quad (6.2)$$

On convient d'appeler  $\Psi_{t|t-1}^{\lambda_t}$  l'**opérateur de prédiction de l'état** du système à l'instant  $t$  conditionnellement aux observations à l'instant  $t-1$  et au processus indicateur latent jusqu'à l'instant  $t$ .

### 6.2.2.2 Opérateur de prédiction de l'observation

On peut déterminer la transition  $\Pi_{Y_t}^{(\Lambda_{0:t}, Y_{0:t-1})}$  à partir de la transition  $\Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t-1})}$  en appliquant l'algorithme de propagation de croyance conditionnelle au réseau bayésien de la figure 6.1 en prenant comme observations  $\{Y_0, Y_1, \dots, Y_{t-1}\}$  et comme ensemble de conditionnement  $\{\Lambda_0, \Lambda_1, \dots, \Lambda_t\}$  (cf figure 6.3).



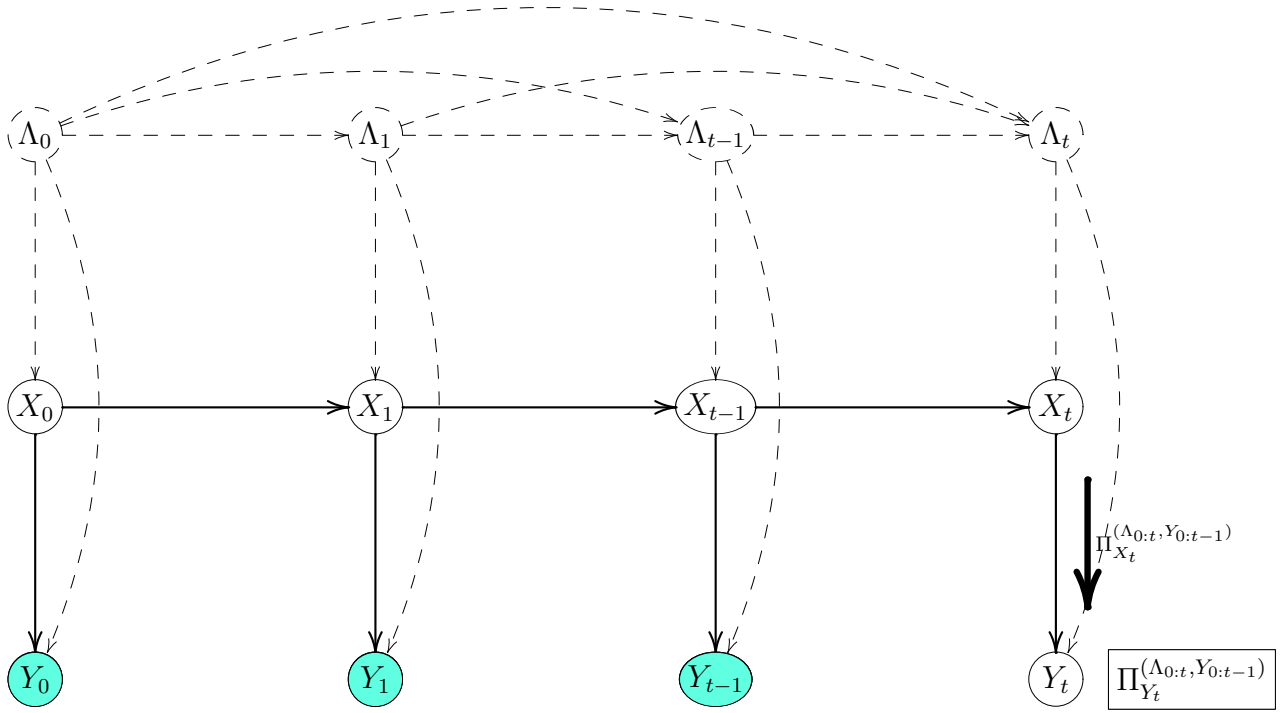


FIG. 6.3 – Calcul de la transition  $\Pi_{Y_t}^{(\Lambda_{0:t}, Y_{0:t-1})}$  à partir de la transition  $\Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t-1})}$  par propagation de croyance conditionnelle, les observations étant  $\{Y_0, Y_1, \dots, Y_{t-1}\}$  et l'ensemble de conditionnement  $\{\Lambda_0, \Lambda_1, \dots, \Lambda_t\}$ .

On obtient alors l'équation suivante

$$\Pi_{Y_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\lambda_{0:t}, y_{0:t-1}; \cdot) = \sum_{X_t} \Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\lambda_{0:t}, y_{0:t-1}; \cdot) \otimes \Pi_{Y_t}^{(\Lambda_t, X_t)}(\lambda_t, \cdot; \cdot) \quad (6.3)$$

Il est donc possible de définir un opérateur  $\Phi_t^{\lambda_t}$  permettant d'obtenir le terme  $\Pi_{Y_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\lambda_{0:t}, y_{0:t-1}; \cdot)$  à partir du terme  $\Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\lambda_{0:t}, y_{0:t-1}; \cdot)$ . Cet opérateur est défini par la relation suivante

$$\Pi_{Y_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\lambda_{0:t}, y_{0:t-1}; \cdot) = \Phi_t^{\lambda_t} \left( \Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\lambda_{0:t}, y_{0:t-1}; \cdot) \right) \quad (6.4)$$

On convient d'appeler  $\Phi_t^{\lambda_t}$  l'**opérateur de prédiction de l'observation** à l'instant  $t$  conditionnellement aux observations à l'instant  $t-1$  et au processus indicateur latent jusqu'à l'instant  $t$ .

### 6.2.2.3 Opérateur de correction de l'état

On peut déterminer la transition  $\Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t})}$  à partir de la transition  $\Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t-1})}$  et de la transition  $\Pi_{Y_t}^{(\Lambda_t, X_t)}$  en appliquant l'algorithme de propagation de croyance conditionnelle au réseau bayésien de la figure 6.1 en prenant comme observations  $\{Y_0, Y_1, \dots, Y_t\}$  et comme ensemble de conditionnement  $\{\Lambda_0, \Lambda_1, \dots, \Lambda_t\}$  (cf figure 6.4).

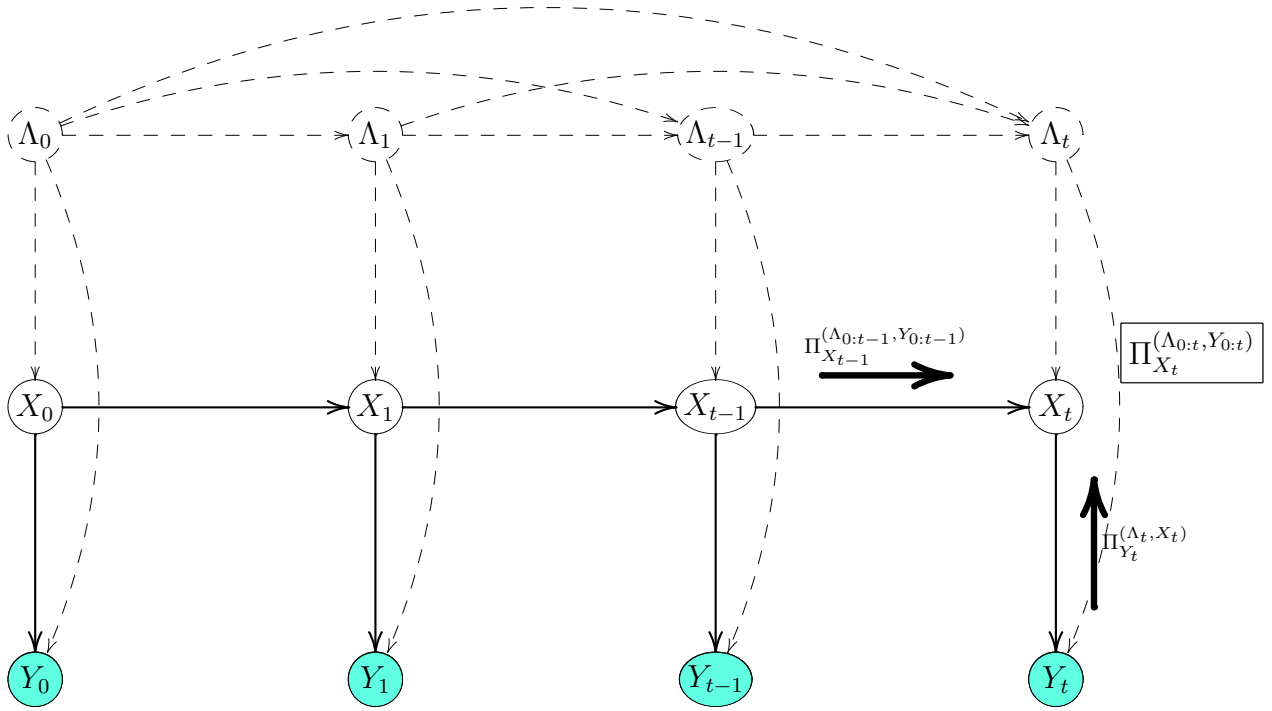


FIG. 6.4 – Calcul de la transition  $\Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t})}$  à partir de la transition  $\Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t-1})}$  et de la transition  $\Pi_{Y_t}^{(\Lambda_t, X_t)}$  par propagation de croyance conditionnelle, les observations étant  $\{Y_0, Y_1, \dots, Y_t\}$  et l'ensemble de conditionnement  $\{\Lambda_0, \Lambda_1, \dots, \Lambda_t\}$ .

On obtient alors l'équation suivante

$$\Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t})}(\lambda_{0:t}, y_{0:t}; \cdot) = \Psi_{\text{Bayes}}(f_{Y_t}^{(\Lambda_t, X_t)}(\lambda_t, \cdot; y_t), \Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\lambda_{0:t}, y_{0:t-1}; \cdot)) \quad (6.5)$$

où  $\Psi_{\text{Bayes}}$  désigne l'opérateur de Bayes (cf la convention 3.2). Il est donc possible de définir un opérateur  $\Psi_t^{(\lambda_t, y_t)}$  permettant d'obtenir le terme  $\Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t})}(\lambda_{0:t}, y_{0:t}; \cdot)$  à partir du terme  $\Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\lambda_{0:t}, y_{0:t-1}; \cdot)$ . Cet opérateur est défini par la relation suivante

$$\Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t})}(\lambda_{0:t}, y_{0:t}; \cdot) = \Psi_t^{(\lambda_t, y_t)}\left(\Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\lambda_{0:t}, y_{0:t-1}; \cdot)\right) \quad (6.6)$$

On convient d'appeler  $\Psi_t^{(\lambda_t, y_t)}$  l'**opérateur de correction de l'état** du système à l'instant  $t$  conditionnellement aux observations à l'instant  $t$  et au processus indicateur latent jusqu'à l'instant  $t$ .

Après avoir défini ces différents opérateurs, nous pouvons maintenant appliquer l'algorithme de filtrage particulière à échantillonnage global au modèle dynamique à processus latent décrit au § 6.2.1.

## 6.3 Le filtre particulaire à échantillonnage global immédiat

### 6.3.1 Principe de l'algorithme

Cet algorithme consiste à appliquer l'algorithme de filtrage particulaire à échantillonnage global afin d'approcher, pour toute suite d'observations  $(y_i)_{i \in \mathbb{N}}$  et pour tout  $t \in \mathbb{N}$ , la mesure

$$F_t = \Pi_{\Lambda_{0:t}}^{Y_{0:t}}(y_{0:t}; \cdot) \quad (6.7)$$

On utilise ensuite la loi approchée  $F_t^N$  obtenue pour estimer  $\Lambda_{t-\delta}$  (avec  $\delta \in \{0, \dots, t\}$ ).

### 6.3.2 Calcul de la transition finie $Q_t$

Afin d'appliquer l'algorithme de filtrage particulaire à échantillonnage global, il est nécessaire de déterminer, pour tout  $t \geq 1$ , une transition finie  $Q_t : (\mathcal{A}, \mathcal{P}(\mathcal{A}))^{\otimes t} \prec (\mathcal{A}, \mathcal{P}(\mathcal{A}))$  telle que

$$F_t = F_{t-1} \otimes Q_t \quad (6.8)$$

Pour tout  $t \in \mathbb{N}$ , on convient de noter  $f_t$  la densité de  $F_t$  par rapport à la mesure de comptage et pour tout  $t \geq 1$  et pour tout  $\lambda \in \mathcal{A}^t$ ,  $q_t(\lambda; \cdot)$  la densité de  $Q_t(\lambda; \cdot)$  par rapport à la mesure de comptage.

Avec ces conventions, on a, pour tout  $t \geq 1$  et pour tout  $\lambda_{0:t} \in \mathcal{A}^{t+1}$ ,

$$f_t(\lambda_{0:t}) = f_{\Lambda_{0:t}}^{Y_{0:t}}(y_{0:t}; \lambda_{0:t}) \quad (6.9)$$

$$\propto f_{Y_{0:t}}^{\Lambda_{0:t}}(\lambda_{0:t}; y_{0:t}) f_{\Lambda_{0:t}}(\lambda_{0:t}) \quad (6.10)$$

Le terme  $f_{Y_{0:t}}^{\Lambda_{0:t}}(\lambda_{0:t}, y_{0:t})$  de l'équation 6.10 peut se décomposer ainsi

$$f_{Y_{0:t}}^{\Lambda_{0:t}}(\lambda_{0:t}, y_{0:t}) = f_{Y_{0:t-1}}^{\Lambda_{0:t}}(\lambda_{0:t}; y_{0:t-1}) f_{Y_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\lambda_{0:t}, y_{0:t-1}; y_t) \quad (6.11)$$

Or, dans le réseau bayésien de la figure 6.1, toute chaîne joignant  $Y_{0:t-1}$  à  $\Lambda_t$  est bloquée par  $\Lambda_{0:t-1}$ . Par conséquent, l'élément aléatoire  $Y_{0:t-1}$  est indépendant de l'élément aléatoire  $\Lambda_t$  conditionnellement à  $\Lambda_{0:t-1}$ . On en déduit

$$f_{Y_{0:t}}^{\Lambda_{0:t}}(\lambda_{0:t}, y_{0:t}) = f_{Y_{0:t-1}}^{\Lambda_{0:t-1}}(\lambda_{0:t-1}; y_{0:t-1}) f_{Y_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\lambda_{0:t}, y_{0:t-1}; y_t) \quad (6.12)$$

D'autre part, le terme  $f_{\Lambda_{0:t}}(\lambda_{0:t})$  de l'équation 6.10 peut se décomposer ainsi

$$f_{\Lambda_{0:t}}(\lambda_{0:t}) = f_{\Lambda_{0:t-1}}(\lambda_{0:t-1}) f_{\Lambda_t}^{\Lambda_{0:t-1}}(\lambda_{0:t-1}; \lambda_t) \quad (6.13)$$

Par conséquent, les équations 6.10, 6.12 et 6.13 impliquent

$$\begin{aligned} f_t(\lambda_{0:t}) &\propto f_{Y_{0:t-1}}^{\Lambda_{0:t-1}}(\lambda_{0:t-1}; y_{0:t-1}) f_{Y_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\lambda_{0:t}, y_{0:t-1}; y_t) \times \dots \\ &\dots \times f_{\Lambda_{0:t-1}}(\lambda_{0:t-1}) f_{\Lambda_t}^{\Lambda_{0:t-1}}(\lambda_{0:t-1}; \lambda_t) \end{aligned} \quad (6.14)$$

Or, d'après l'équation 6.8, on a

$$f_t(\lambda_{0:t}) = f_{t-1}(\lambda_{0:t-1}) q_t(\lambda_{0:t-1}; \lambda_t) \quad (6.15)$$

avec, d'après l'équation 6.10, appliquée à l'instant  $t - 1$

$$f_{t-1}(\lambda_{0:t-1}) \propto f_{Y_{0:t-1}}^{\Lambda_{0:t-1}}(\lambda_{0:t-1}; y_{0:t-1}) f_{\Lambda_{0:t-1}}(\lambda_{0:t-1}) \quad (6.16)$$

En comparant les équations 6.14, 6.15 et 6.16, on obtient

$$q_t(\lambda_{0:t-1}; \lambda_t) \propto f_{Y_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\lambda_{0:t}, y_{0:t-1}; y_t) f_{\Lambda_t}^{\Lambda_{0:t-1}}(\lambda_{0:t-1}; \lambda_t) \quad (6.17)$$

### 6.3.3 Description de l'algorithme

#### 6.3.3.1 Notations

On convient tout d'abord de poser pour tout  $t \in \mathbb{N}$  et pour tout  $\lambda_{0:t} \in \mathcal{A}^{t+1}$

$$\gamma_t(\lambda_{0:t}) \stackrel{\text{déf}}{=} \begin{cases} f_{Y_0}^{\Lambda_0}(\lambda_0; y_0) f_{\Lambda_0}(\lambda_0) & \text{si } t = 0 \\ f_{Y_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\lambda_{0:t}, y_{0:t-1}; y_t) f_{\Lambda_t}^{\Lambda_{0:t-1}}(\lambda_{0:t-1}; \lambda_t) & \text{si } t > 0 \end{cases} \quad (6.18)$$

On remarque alors que, avec ces notations, l'expression 6.17 peut s'écrire

$$q_t(\lambda_{0:t-1}; \lambda_t) \propto \gamma_t(\lambda_{0:t}) \quad (6.19)$$

#### 6.3.3.2 Passage de l'instant $t - 1$ à l'instant $t$

Soit  $t \geq 1$ . Pour tout  $j \in \{1, \dots, N\}$ , on suppose que

- on a défini la trajectoire  $\Lambda^{(j,t-1)} = (\Lambda_0^{(j,t-1)}, \dots, \Lambda_{t-1}^{(j,t-1)})$
- on a déterminé la mesure de probabilité aléatoire

$$\begin{cases} \prod_{X_{t-1}}^{(\Lambda_{0:t-1}, Y_{0:t-2})}(\Lambda^{(j,t-1)}, y_{0:t-2}; \cdot) & \text{si } t > 1 \\ \prod_{X_0}^{\Lambda_0}(\Lambda^{(j,0)}; \cdot) & \text{si } t = 1 \end{cases} \quad (6.20)$$

alors

1. Pour tout  $j \in \{1, \dots, N\}$  et pour tout  $\lambda_t \in \mathcal{A}$ , on calcule

$$\begin{aligned} \prod_{X_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\Lambda^{(j,t-1)}, \lambda_t, y_{0:t-1}; \cdot) &= \dots \\ \dots &\begin{cases} \Psi_{t|t-1}^{\lambda_t} \circ \Psi_{t-1}^{(\Lambda_{t-1}^{(j,t-1)}, Y_{t-1})} \left( \prod_{X_{t-1}}^{(\Lambda_{0:t-1}, Y_{0:t-2})}(\Lambda^{(j,t-1)}, y_{0:t-2}; \cdot) \right) & \text{si } t > 1 \\ \Psi_{1|0}^{\lambda_1} \circ \Psi_0^{(\Lambda_0^{(j,0)}, Y_0)} \left( \prod_{X_0}^{\Lambda_0}(\Lambda^{(j,0)}; \cdot) \right) & \text{si } t = 1 \end{cases} \end{aligned} \quad (6.21)$$

$$\prod_{Y_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\Lambda^{(j,t-1)}, \lambda_t, y_{0:t-1}; \cdot) = \Phi_t^{\lambda_t} \left( \prod_{X_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\Lambda^{(j,t-1)}, \lambda_t, y_{0:t-1}; \cdot) \right) \quad (6.22)$$

$$\gamma_t(\Lambda^{(j,t-1)}, \lambda_t) = f_{Y_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\Lambda^{(j,t-1)}, \lambda_t, y_{0:t-1}; y_t) f_{\Lambda_t}^{\Lambda_{0:t-1}}(\Lambda^{(j,t-1)}; \lambda_t) \quad (6.23)$$

2. On calcule le coefficient de normalisation

$$\rho_t = \sum_j \sum_{\lambda_t} \gamma_t(\Lambda^{(j,t-1)}, \lambda_t) \quad (6.24)$$

3. On définit la mesure de probabilité aléatoire  $\tilde{F}_t^N$  par

$$\tilde{F}_t^N = \sum_{j=1}^N \sum_{\lambda_t \in \mathcal{A}} \tilde{f}_t^N(\Lambda^{(j,t-1)}, \lambda_t) \delta_{(\Lambda^{(j,t-1)}, \lambda_t)} \quad (6.25)$$

avec

$$\tilde{f}_t^N(\Lambda^{(j,t-1)}, \lambda_t) = \rho_t^{-1} \gamma_t(\Lambda^{(j,t-1)}, \lambda_t) \quad (6.26)$$

4. Pour tout  $j \in \{1, \dots, N\}$ , on définit la variable aléatoire  $\Lambda^{(j,t)} = (\Lambda_0^{(j,t)}, \Lambda_1^{(j,t)}, \dots, \Lambda_t^{(j,t)})$  correspondant à la  $j$ -ème trajectoire à l'instant  $t$  par

$$\Lambda^{(j,t)} \sim \tilde{F}_t^N \quad \text{conditionnellement à } \Lambda^{(1,t-1)}, \dots, \Lambda^{(N,t-1)} \quad (6.27)$$

et par la condition que  $\Lambda^{(j,t)}$  doit vérifier les relations d'indépendance conditionnelle représentées par le réseau bayésien de la figure 6.5.

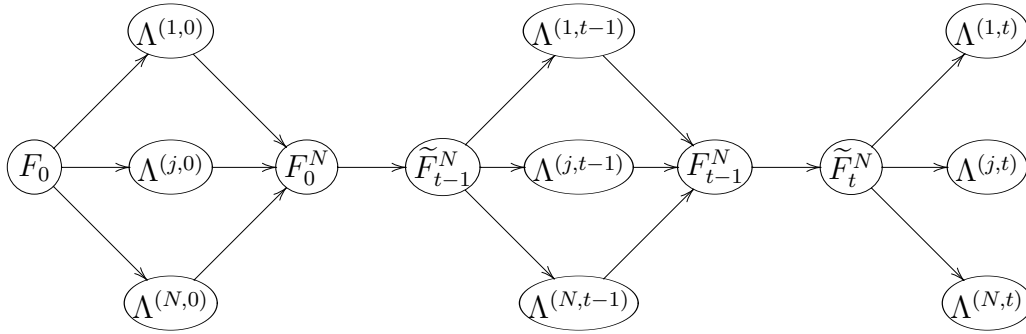


FIG. 6.5 – Réseau bayésien représentant les relations d'indépendance conditionnelle que doivent vérifier les  $\Lambda^{(j,t)}$ .

On notera que l'on a déjà calculé (cf l'équation 6.21) la mesure de probabilité  $\Pi_{X_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\Lambda^{(j,t)}, y_{0:t-1}; \cdot)$  associée à  $\Lambda^{(j,t)}$ .

### 6.3.3.3 Initialisation de l'algorithme

1. Pour tout  $\lambda_0 \in \mathcal{A}$ , on calcule

$$\Pi_{Y_0}^{\Lambda_0}(\lambda_0; \cdot) = \Phi_0^{\lambda_0} \left( \Pi_{X_0}^{\Lambda_0}(\lambda_0; \cdot) \right) \quad (6.28)$$

$$\gamma_0(\lambda_0) = f_{Y_0}^{\Lambda_0}(\lambda_0; y_0) f_{\Lambda_0}(\lambda_0) \quad (6.29)$$

2. On calcule le coefficient de normalisation

$$\rho_0 = \sum_{\lambda_0} \gamma_0(\lambda_0) \quad (6.30)$$

3. On obtient la mesure de probabilité  $F_0$  en utilisant l'équation suivante

$$F_0 = \sum_{\lambda_0 \in \mathcal{A}} f_0(\lambda_0) \delta_{\lambda_0} \quad (6.31)$$

avec

$$f_0(\lambda_0) = \rho_0^{-1} \gamma_0(\lambda_0) \quad (6.32)$$

4. Pour tout  $j \in \{1, \dots, N\}$ , on définit la variable aléatoire  $\Lambda^{(j,0)}$  correspondant à la  $j$ -ème trajectoire à l'instant  $t = 0$  par

$$\Lambda^{(j,0)} \sim F_0 \quad (6.33)$$

et lorsque  $j > 1$ , par la condition que  $\Lambda^{(j,0)}$  est indépendant de  $\Lambda^{(1,0)}, \dots, \Lambda^{(j-1,0)}$ .

#### 6.3.3.4 Estimation de $\Lambda_{t-\delta}$

Pour tout  $t \in \mathbb{N}$  et pour tout  $\delta \in \{0, \dots, t\}$ , on peut estimer  $\Lambda_{t-\delta}$  à l'aide de la mesure  $F_t^N$  qui est une approximation de la mesure  $F_t = \Pi_{\Lambda_{0:t}}^{Y_{0:t}}(y_{0:t}; \cdot)$ . Cependant, dans le cas  $t > 0$ , il est préférable d'utiliser la mesure  $\tilde{F}_t^N$  qui est une meilleure approximation de  $F_t$ . On obtient alors une approximation  $\hat{\Pi}_{\Lambda_{t-\delta}}^{Y_{0:t}}(y_{0:t}; \cdot)$  de la loi *a posteriori* de  $\Lambda_{t-\delta}$  vérifiant, pour tout  $\lambda_{t-\delta} \in \mathcal{A}$ ,

$$\hat{\Pi}_{\Lambda_{t-\delta}}^{Y_{0:t}}(y_{0:t}; \{\lambda_{t-\delta}\}) = \sum_{(\lambda_{t'})_{t' \neq t-\delta}} \tilde{F}_t^N(\{\lambda_{0:t}\}) \quad (6.34)$$

En notant  $(\tilde{\Lambda}^{(k,t)})_k$  les  $N \text{ card}(\mathcal{A})$  trajectoires correspondant à toutes les prolongations possibles à l'instant  $t$  des  $N$  trajectoires  $\Lambda^{(j,t-1)}$  à l'instant  $t - 1$ , l'équation 6.34 devient

$$\hat{\Pi}_{\Lambda_{t-\delta}}^{Y_{0:t}}(y_{0:t}; \{\lambda_{t-\delta}\}) = \sum_k \tilde{f}_t^N(\tilde{\Lambda}^{(k,t)}) \delta_{\tilde{\Lambda}_{t-\delta}^{(k,t)}, \lambda_{t-\delta}} \quad (6.35)$$

On obtient alors une estimation de  $\Lambda_{t-\delta}$  étant données les observations  $Y_{0:t} = y_{0:t}$  en maximisant l'expression 6.35 par rapport à  $\lambda_{t-\delta}$ .

## 6.4 Le filtre particulière à échantillonnage global différé

### 6.4.1 Principe de l'algorithme

Cet algorithme consiste à appliquer l'algorithme de filtrage particulière à échantillonnage global afin d'approcher, pour toute suite d'observations  $(y_i)_{i \in \mathbb{N}}$ , pour tout instant  $t \in \mathbb{N}$  et pour tout délai  $\Delta \in \mathbb{N}^*$ , la mesure

$$F_t = \Pi_{\Lambda_{0:t}}^{Y_{0:t+\Delta}}(y_{0:t+\Delta}; \cdot) \quad (6.36)$$

On utilise ensuite la loi approchée  $F_t^N$  obtenue pour estimer  $\Lambda_{t-\delta}$  (avec  $\delta \in \{0, \dots, t\}$ ).

### 6.4.2 Calcul de la transition finie $Q_t$

Afin d'appliquer l'algorithme de filtrage particulière à échantillonnage global, il est nécessaire de déterminer, pour tout  $t \geq 1$ , une transition finie  $Q_t : (\mathcal{A}, \mathcal{P}(\mathcal{A}))^{\otimes t} \prec (\mathcal{A}, \mathcal{P}(\mathcal{A}))$  telle que

$$F_t = F_{t-1} \otimes Q_t \quad (6.37)$$

Pour tout  $t \in \mathbb{N}$ , on convient de noter  $f_t$  la densité de  $F_t$  par rapport à la mesure de comptage et pour tout  $t \geq 1$  et pour tout  $\lambda \in \mathcal{A}^t$ ,  $q_t(\lambda; \cdot)$  la densité de  $Q_t(\lambda; \cdot)$  par rapport à la mesure de comptage.

Avec ces conventions, on a, pour tout  $t \geq 1$ , pour tout  $\Delta \geq 1$  et pour tout  $\lambda_{0:t} \in \mathcal{A}^{t+1}$ ,

$$f_t(\lambda_{0:t}) = f_{\Lambda_{0:t}}^{Y_{0:t+\Delta}}(y_{0:t+\Delta}; \lambda_{0:t}) \quad (6.38)$$

$$= \sum_{\lambda_{t+1:t+\Delta}} f_{\Lambda_{0:t+\Delta}}^{Y_{0:t+\Delta}}(y_{0:t+\Delta}; \lambda_{0:t+\Delta}) \quad (6.39)$$

$$\propto \sum_{\lambda_{t+1:t+\Delta}} f_{Y_{0:t+\Delta}}^{\Lambda_{0:t+\Delta}}(\lambda_{0:t+\Delta}; y_{0:t+\Delta}) f_{\Lambda_{0:t+\Delta}}(\lambda_{0:t+\Delta}) \quad (6.40)$$

Le terme  $f_{Y_{0:t+\Delta}}^{\Lambda_{0:t+\Delta}}(\lambda_{0:t+\Delta}; y_{0:t+\Delta})$  de l'équation 6.40 peut se décomposer ainsi

$$\begin{aligned} f_{Y_{0:t+\Delta}}^{\Lambda_{0:t+\Delta}}(\lambda_{0:t+\Delta}; y_{0:t+\Delta}) &= f_{Y_{0:t+\Delta-1}}^{\Lambda_{0:t+\Delta}}(\lambda_{0:t+\Delta}; y_{0:t+\Delta-1}) \times \dots \\ &\dots \times f_{Y_{t+\Delta}}^{(\Lambda_{0:t+\Delta}, Y_{0:t+\Delta-1})}(\lambda_{0:t+\Delta}, y_{0:t+\Delta-1}; y_{t+\Delta}) \end{aligned} \quad (6.41)$$

Or, dans le réseau bayésien de la figure 6.1, toute chaîne joignant  $Y_{0:t+\Delta-1}$  à  $\Lambda_{t+\Delta}$  est bloquée par  $\Lambda_{0:t+\Delta-1}$ . Par conséquent, l'élément aléatoire  $Y_{0:t+\Delta-1}$  est indépendant de l'élément aléatoire  $\Lambda_{t+\Delta}$  conditionnellement à  $\Lambda_{0:t+\Delta-1}$ . On en déduit

$$\begin{aligned} f_{Y_{0:t+\Delta}}^{\Lambda_{0:t+\Delta}}(\lambda_{0:t+\Delta}; y_{0:t+\Delta}) &= f_{Y_{0:t+\Delta-1}}^{\Lambda_{0:t+\Delta-1}}(\lambda_{0:t+\Delta-1}; y_{0:t+\Delta-1}) \times \dots \\ &\dots \times f_{Y_{t+\Delta}}^{(\Lambda_{0:t+\Delta}, Y_{0:t+\Delta-1})}(\lambda_{0:t+\Delta}, y_{0:t+\Delta-1}; y_{t+\Delta}) \end{aligned} \quad (6.42)$$

D'autre part, le terme  $f_{\Lambda_{0:t+\Delta}}(\lambda_{0:t+\Delta})$  de l'équation 6.40 peut se décomposer ainsi

$$f_{\Lambda_{0:t+\Delta}}(\lambda_{0:t+\Delta}) = f_{\Lambda_{0:t+\Delta-1}}(\lambda_{0:t+\Delta-1}) f_{\Lambda_{t+\Delta}}^{\Lambda_{0:t+\Delta-1}}(\lambda_{0:t+\Delta-1}; \lambda_{t+\Delta}) \quad (6.43)$$

Par conséquent, les équations 6.40, 6.42 et 6.43 impliquent

$$\begin{aligned} f_t(\lambda_{0:t}) &\propto \sum_{\lambda_{t+1:t+\Delta}} f_{Y_{0:t+\Delta-1}}^{\Lambda_{0:t+\Delta-1}}(\lambda_{0:t+\Delta-1}; y_{0:t+\Delta-1}) \times \dots \\ &\dots \times f_{Y_{t+\Delta}}^{(\Lambda_{0:t+\Delta}, Y_{0:t+\Delta-1})}(\lambda_{0:t+\Delta}, y_{0:t+\Delta-1}; y_{t+\Delta}) \times \dots \\ &\dots \times f_{\Lambda_{0:t+\Delta-1}}(\lambda_{0:t+\Delta-1}) \cdot f_{\Lambda_{t+\Delta}}^{\Lambda_{0:t+\Delta-1}}(\lambda_{0:t+\Delta-1}; \lambda_{t+\Delta}) \end{aligned} \quad (6.44)$$

Or le produit du premier et du troisième terme du membre de droite de l'équation 6.44 vérifie

$$f_{Y_{0:t+\Delta-1}}^{\Lambda_{0:t+\Delta-1}}(\lambda_{0:t+\Delta-1}; y_{0:t+\Delta-1}) f_{\Lambda_{0:t+\Delta-1}}(\lambda_{0:t+\Delta-1}) \propto f_{\Lambda_{0:t+\Delta-1}}^{Y_{0:t+\Delta-1}}(y_{0:t+\Delta-1}; \lambda_{0:t+\Delta-1}) \quad (6.45)$$

$$\begin{aligned} &\propto f_{\Lambda_{0:t-1}}^{Y_{0:t+\Delta-1}}(y_{0:t+\Delta-1}; \lambda_{0:t-1}) \times \dots \\ &\dots \times f_{\Lambda_{t+\Delta-1}}^{(Y_{0:t+\Delta-1}, \Lambda_{0:t-1})}(y_{0:t+\Delta-1}, \lambda_{0:t-1}; \lambda_{t:t+\Delta-1}) \end{aligned} \quad (6.46)$$

Le deuxième terme du membre de droite de l'équation 6.46 vérifie, d'après le théorème de Bayes

$$\begin{aligned}
f_{\Lambda_{t:t+\Delta-1}}^{(Y_{0:t+\Delta-1}, \Lambda_{0:t-1})}(y_{0:t+\Delta-1}, \lambda_{0:t-1}; \lambda_{t:t+\Delta-1}) &= \sigma_t(y_{0:t-1}, \lambda_{0:t-1}) \times \dots \\
&\dots \times f_{Y_{t:t+\Delta-1}}^{(Y_{0:t-1}, \Lambda_{0:t+\Delta-1})}(y_{0:t-1}, \lambda_{0:t+\Delta-1}; y_{t:t+\Delta-1}) \times \dots \\
&\dots \times f_{\Lambda_{t:t+\Delta-1}}^{(Y_{0:t-1}, \Lambda_{0:t-1})}(y_{0:t-1}, \lambda_{0:t-1}; \lambda_{t:t+\Delta-1})
\end{aligned} \tag{6.47}$$

où le coefficient de normalisation  $\sigma_t(y_{0:t-1}, \lambda_{0:t-1})$  est défini par

$$\begin{aligned}
\sigma_t(y_{0:t-1}, \lambda_{0:t-1}) &= \left[ \sum_{\lambda'_{t:t+\Delta-1}} f_{Y_{t:t+\Delta-1}}^{(Y_{0:t-1}, \Lambda_{0:t+\Delta-1})}(y_{0:t-1}, \lambda_{0:t-1}, \lambda'_{t:t+\Delta-1}; y_{t:t+\Delta-1}) \times \dots \right. \\
&\dots \left. \times f_{\Lambda_{t:t+\Delta-1}}^{(Y_{0:t-1}, \Lambda_{0:t-1})}(y_{0:t-1}, \lambda_{0:t-1}; \lambda'_{t:t+\Delta-1}) \right]^{-1}
\end{aligned} \tag{6.48}$$

Le dernier terme du membre de droite de l'équation 6.47 peut se simplifier en remarquant que, dans le réseau bayésien de la figure 6.1, toute chaîne joignant  $Y_{0:t-1}$  à  $\Lambda_{t:t+\Delta-1}$  est bloquée par  $\Lambda_{0:t-1}$ . Par conséquent,  $\Lambda_{t:t+\Delta-1}$  est indépendant de  $Y_{0:t-1}$  conditionnellement à  $\Lambda_{0:t-1}$  et le dernier terme du membre de droite de l'équation 6.47 vérifie donc

$$f_{\Lambda_{t:t+\Delta-1}}^{(Y_{0:t-1}, \Lambda_{0:t-1})}(y_{0:t-1}, \lambda_{0:t-1}; \lambda_{t:t+\Delta-1}) = f_{\Lambda_{t:t+\Delta-1}}^{\Lambda_{0:t-1}}(\lambda_{0:t-1}; \lambda_{t:t+\Delta-1}) \tag{6.49}$$

Les équations 6.44, 6.46, 6.47 et 6.49 impliquent donc

$$\begin{aligned}
f_t(\lambda_{0:t}) &\propto f_{\Lambda_{0:t-1}}^{Y_{0:t+\Delta-1}}(y_{0:t+\Delta-1}; \lambda_{0:t-1}) \sigma_t(y_{0:t-1}, \lambda_{0:t-1}) \times \dots \\
&\dots \times \sum_{\lambda_{t+1:t+\Delta}} f_{Y_{t:t+\Delta-1}}^{(Y_{0:t-1}, \Lambda_{0:t+\Delta-1})}(y_{0:t-1}, \lambda_{0:t+\Delta-1}; y_{t:t+\Delta-1}) f_{\Lambda_{t:t+\Delta-1}}^{\Lambda_{0:t-1}}(\lambda_{0:t-1}; \lambda_{t:t+\Delta-1}) \times \dots \\
&\dots \times f_{Y_{t+\Delta}}^{(\Lambda_{0:t+\Delta}, Y_{0:t+\Delta-1})}(\lambda_{0:t+\Delta}, y_{0:t+\Delta-1}; y_{t+\Delta}) f_{\Lambda_{t+\Delta}}^{\Lambda_{0:t+\Delta-1}}(\lambda_{0:t+\Delta-1}; \lambda_{t+\Delta})
\end{aligned} \tag{6.50}$$

$$\begin{aligned}
&\propto f_{\Lambda_{0:t-1}}^{Y_{0:t+\Delta-1}}(y_{0:t+\Delta-1}; \lambda_{0:t-1}) \sigma_t(y_{0:t-1}, \lambda_{0:t-1}) \times \dots \\
&\dots \times \sum_{\lambda_{t+1:t+\Delta}} \prod_{\tau=t}^{t+\Delta} f_{Y_\tau}^{(\Lambda_{0:\tau}, Y_{0:\tau-1})}(\lambda_{0:\tau}, y_{0:\tau-1}; y_\tau) f_{\Lambda_\tau}^{\Lambda_{0:\tau-1}}(\lambda_{0:\tau-1}; \lambda_\tau)
\end{aligned} \tag{6.51}$$

On en déduit

$$q_t(\lambda_{0:t-1}; \lambda_t) \propto \frac{\sum_{\lambda_{t+1:t+\Delta}} \prod_{\tau=t}^{t+\Delta} f_{Y_\tau}^{(\Lambda_{0:\tau}, Y_{0:\tau-1})}(\lambda_{0:\tau}, y_{0:\tau-1}; y_\tau) f_{\Lambda_\tau}^{\Lambda_{0:\tau-1}}(\lambda_{0:\tau-1}; \lambda_\tau)}{\sum_{\lambda'_{t:t+\Delta-1}} \prod_{\tau=t}^{t+\Delta-1} f_{Y_\tau}^{(\Lambda_{0:\tau}, Y_{0:\tau-1})}(\lambda_{0:t-1}, \lambda'_{t:\tau}, y_{0:\tau-1}; y_\tau) f_{\Lambda_\tau}^{\Lambda_{0:\tau-1}}(\lambda_{0:t-1}, \lambda'_{t:\tau-1}; \lambda'_\tau)} \tag{6.52}$$



### 6.4.3 Description de l'algorithme

#### 6.4.3.1 Notations

On convient tout d'abord de poser pour tout  $t \in \mathbb{N}$  et pour tout  $\lambda_{0:t} \in \mathcal{A}^{t+1}$

$$\gamma_t(\lambda_{0:t}) \stackrel{\text{déf}}{=} \begin{cases} f_{Y_0}^{\Lambda_0}(\lambda_0; y_0) f_{\Lambda_0}(\lambda_0) & \text{si } t = 0 \\ f_{Y_t}^{(\Lambda_{0:t}, Y_{0:t-1})}(\lambda_{0:t}, y_{0:t-1}; y_t) f_{\Lambda_t}^{\Lambda_{0:t-1}}(\lambda_{0:t-1}; \lambda_t) & \text{si } t > 0 \end{cases} \quad (6.53)$$

ainsi que,

$$D_t(\lambda_{0:t}) \stackrel{\text{déf}}{=} \sum_{\lambda_{t+1:t+\Delta}} \prod_{\tau=t+1}^{t+\Delta} \gamma_\tau(\lambda_{0:\tau}) \quad (6.54)$$

On remarque alors que, avec ces notations, l'expression 6.52 peut s'écrire

$$q_t(\lambda_{0:t-1}; \lambda_t) \propto \gamma_t(\lambda_{0:t}) \frac{D_t(\lambda_{0:t})}{D_{t-1}(\lambda_{0:t-1})} \quad (6.55)$$

#### 6.4.3.2 Passage de l'instant $t - 1$ à l'instant $t$

Soit  $t \geq 1$ . Pour tout  $j \in \{1, \dots, N\}$ , on suppose que

- on a défini la trajectoire  $\Lambda^{(j,t-1)} = (\Lambda_0^{(j,t-1)}, \dots, \Lambda_{t-1}^{(j,t-1)})$
- on a calculé  $D_{t-1}(\Lambda^{(j,t-1)})$
- pour tout  $\tau \in \{t, t+1, \dots, t+\Delta-1\}$  et pour tout  $\lambda_{t:\tau} \in \mathcal{A}^{\tau-t+1}$ , on a calculé  $\gamma_\tau(\Lambda^{(j,t-1)}, \lambda_{t:\tau})$
- pour tout  $\lambda_{t:t+\Delta-1} \in \mathcal{A}^\Delta$ , on a déterminé la mesure de probabilité aléatoire

$$\prod_{X_{t+\Delta-1}}^{(\Lambda_{0:t+\Delta-1}, Y_{0:t+\Delta-2})}(\Lambda^{(j,t-1)}, \lambda_{t:t+\Delta-1}, y_{0:t+\Delta-2}; \cdot) \quad (6.56)$$

alors

1. Pour tout  $j \in \{1, \dots, N\}$  et pour tout  $\lambda_{t:t+\Delta} \in \mathcal{A}^{\Delta+1}$ , on calcule

$$\begin{aligned} & \prod_{X_{t+\Delta}}^{(\Lambda_{0:t+\Delta}, Y_{0:t+\Delta-1})}(\Lambda^{(j,t-1)}, \lambda_{t:t+\Delta}, y_{0:t+\Delta-1}; \cdot) = \dots \\ & \Psi_{t+\Delta|t+\Delta-1}^{\lambda_{t+\Delta}} \circ \Psi_{t+\Delta-1}^{(\lambda_{t+\Delta-1}, Y_{t+\Delta-1})} \left( \prod_{X_{t+\Delta-1}}^{(\Lambda_{0:t+\Delta-1}, Y_{0:t+\Delta-2})}(\Lambda^{(j,t-1)}, \lambda_{t:t+\Delta-1}, y_{0:t+\Delta-2}; \cdot) \right) \end{aligned} \quad (6.57)$$

$$\begin{aligned} & \prod_{Y_{t+\Delta}}^{(\Lambda_{0:t+\Delta}, Y_{0:t+\Delta-1})}(\Lambda^{(j,t-1)}, \lambda_{t:t+\Delta}, y_{0:t+\Delta-1}; \cdot) = \dots \\ & \Phi_{t+\Delta}^{\lambda_{t+\Delta}} \left( \prod_{X_{t+\Delta}}^{(\Lambda_{0:t+\Delta}, Y_{0:t+\Delta-1})}(\Lambda^{(j,t-1)}, \lambda_{t:t+\Delta}, y_{0:t+\Delta-1}; \cdot) \right) \end{aligned} \quad (6.58)$$

$$\begin{aligned} \gamma_{t+\Delta}(\Lambda^{(j,t-1)}, \lambda_{t:t+\Delta}) &= f_{Y_{t+\Delta}}^{(\Lambda_{0:t+\Delta}, Y_{0:t+\Delta-1})}(\Lambda^{(j,t-1)}, \lambda_{t:t+\Delta}, y_{0:t+\Delta-1}; y_{t+\Delta}) \times \dots \\ & \dots \times f_{\Lambda_{t+\Delta}}^{\Lambda_{0:t+\Delta-1}}(\Lambda^{(j,t-1)}, \lambda_{t:t+\Delta-1}; \lambda_{t+\Delta}) \end{aligned} \quad (6.59)$$

2. Pour tout  $j \in \{1, \dots, N\}$  et pour tout  $\lambda_t \in \mathcal{A}$ , on calcule

$$D_t(\Lambda^{(j,t-1)}, \lambda_t) = \sum_{\lambda_{t+1:t+\Delta}} \prod_{\tau=t+1}^{t+\Delta} \gamma_\tau(\Lambda^{(j,t-1)}, \lambda_{t:\tau}) \quad (6.60)$$

$$\begin{aligned} &= \sum_{\lambda_{t+1}} \gamma_{t+1}(\Lambda^{(j,t-1)}, \lambda_{t:t+1}) \sum_{\lambda_{t+2}} \gamma_{t+2}(\Lambda^{(j,t-1)}, \lambda_{t:t+2}) \dots \\ &\dots \sum_{\lambda_{t+\Delta-1}} \gamma_{t+\Delta-1}(\Lambda^{(j,t-1)}, \lambda_{t:t+\Delta-1}) \sum_{\lambda_{t+\Delta}} \gamma_{t+\Delta}(\Lambda^{(j,t-1)}, \lambda_{t:t+\Delta}) \end{aligned} \quad (6.61)$$

3. On calcule le coefficient de normalisation

$$\rho_t = \sum_j \sum_{\lambda_t} \gamma_t(\Lambda^{(j,t-1)}, \lambda_t) \frac{D_t(\Lambda^{(j,t-1)}, \lambda_t)}{D_{t-1}(\Lambda^{(j,t-1)})} \quad (6.62)$$

4. On définit la mesure de probabilité aléatoire  $\tilde{F}_t^N$  par

$$\tilde{F}_t^N = \sum_{j=1}^N \sum_{\lambda_t \in \mathcal{A}} \tilde{f}_t^N(\Lambda^{(j,t-1)}, \lambda_t) \delta_{(\Lambda^{(j,t-1)}, \lambda_t)} \quad (6.63)$$

avec

$$\tilde{f}_t^N(\Lambda^{(j,t-1)}, \lambda_t) = \rho_t^{-1} \gamma_t(\Lambda^{(j,t-1)}, \lambda_t) \frac{D_t(\Lambda^{(j,t-1)}, \lambda_t)}{D_{t-1}(\Lambda^{(j,t-1)})} \quad (6.64)$$

5. Pour tout  $j \in \{1, \dots, N\}$ , on définit la variable aléatoire  $\Lambda^{(j,t)} = (\Lambda_0^{(j,t)}, \Lambda_1^{(j,t)}, \dots, \Lambda_t^{(j,t)})$  correspondant à la  $j$ -ème trajectoire à l'instant  $t$  par

$$\Lambda^{(j,t)} \sim \tilde{F}_t^N \quad \text{conditionnellement à } \Lambda^{(1,t-1)}, \dots, \Lambda^{(N,t-1)} \quad (6.65)$$

et par la condition que  $\Lambda^{(j,t)}$  doit vérifier les relations d'indépendance conditionnelle représentées par le réseau bayésien de la figure 6.6.

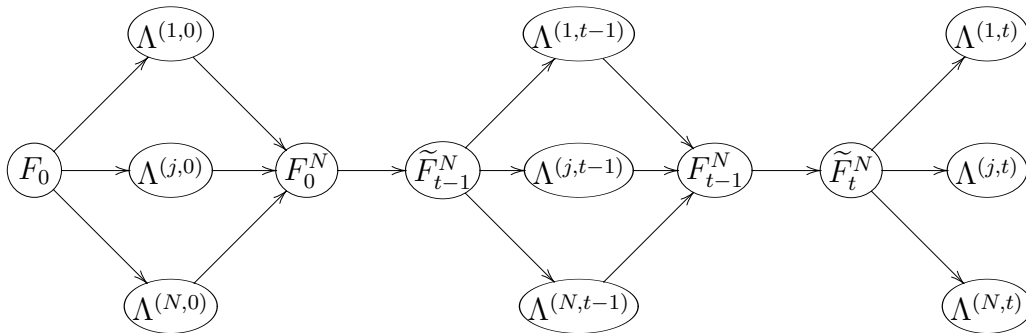


FIG. 6.6 – Réseau bayésien représentant les relations d'indépendance conditionnelle que doivent vérifier les  $\Lambda^{(j,t)}$ .

On notera que l'on a déjà calculé (cf les équations 6.57, 6.59 et 6.61) les grandeurs suivantes associées à  $\Lambda^{(j,t)}$  :

- $D_t(\Lambda^{(j,t)})$ ;
- $\gamma_\tau(\Lambda^{(j,t)}, \lambda_{t+1:\tau})$  pour tout  $\tau \in \{t+1, t+2, \dots, t+\Delta\}$  et pour tout  $\lambda_{t+1:\tau} \in \mathcal{A}^{\tau-t}$ ;
- $\Pi_{X_{t+\Delta}}^{(\Lambda_{0:t+\Delta}, Y_{0:t+\Delta-1})}(\Lambda^{(j,t)}, \lambda_{t+1:t+\Delta}, y_{0:t+\Delta-1}; \cdot)$  pour tout  $\lambda_{t+1:t+\Delta} \in \mathcal{A}^\Delta$ .

### 6.4.3.3 Initialisation de l'algorithme

1. Pour tout  $\lambda_0 \in \mathcal{A}$ , on calcule

$$\Pi_{Y_0}^{\Lambda_0}(\lambda_0; \cdot) = \Phi_0^{\lambda_0} \left( \Pi_{X_0}^{\Lambda_0}(\lambda_0; \cdot) \right) \quad (6.66)$$

$$\gamma_0(\lambda_0) = f_{Y_0}^{\Lambda_0}(\lambda_0; y_0) f_{\Lambda_0}(\lambda_0) \quad (6.67)$$

2. Pour tout  $\tau \in \{1, \dots, \Delta\}$  et pour tout  $\lambda_{0:\tau} \in \mathcal{A}^{\tau+1}$ , on calcule

$$\begin{aligned} \Pi_{X_\tau}^{(\Lambda_{0:\tau}, Y_{0:\tau-1})}(\lambda_{0:\tau}, y_{0:\tau-1}; \cdot) = \dots \\ \dots \begin{cases} \Psi_{1|0}^{\lambda_1} \circ \Psi_0^{(\lambda_0, Y_0)} \left( \Pi_{X_0}^{\Lambda_0}(\lambda_0; \cdot) \right) & \text{si } \tau = 1 \\ \Psi_{\tau|\tau-1}^{\lambda_\tau} \circ \Psi_{\tau-1}^{(\lambda_{\tau-1}, Y_{\tau-1})} \left( \Pi_{X_{\tau-1}}^{(\Lambda_{0:\tau-1}, Y_{0:\tau-2})}(\lambda_{0:\tau-1}, y_{0:\tau-2}; \cdot) \right) & \text{si } \tau > 1 \end{cases} \end{aligned} \quad (6.68)$$

$$\Pi_{Y_\tau}^{(\Lambda_{0:\tau}, Y_{0:\tau-1})}(\lambda_{0:\tau}, y_{0:\tau-1}; \cdot) = \Phi_\tau^{\lambda_\tau} \left( \Pi_{X_\tau}^{(\Lambda_{0:\tau}, Y_{0:\tau-1})}(\lambda_{0:\tau}, y_{0:\tau-1}; \cdot) \right) \quad (6.69)$$

$$\gamma_\tau(\lambda_{0:\tau}) = f_{Y_\tau}^{(\Lambda_{0:\tau}, Y_{0:\tau-1})}(\lambda_{0:\tau}, y_{0:\tau-1}; Y_\tau) f_{\Lambda_\tau}^{\Lambda_{0:\tau-1}}(\lambda_{0:\tau-1}; \lambda_\tau) \quad (6.70)$$

3. Pour tout  $\lambda_0 \in \mathcal{A}$ , on calcule

$$D_0(\lambda_0) = \sum_{\lambda_{1:\Delta}} \prod_{\tau=1}^{\Delta} \gamma_\tau(\lambda_{0:\tau}) \quad (6.71)$$

$$= \sum_{\lambda_1} \gamma_1(\lambda_{0:1}) \sum_{\lambda_2} \gamma_2(\lambda_{0:2}) \dots \sum_{\lambda_{\Delta-1}} \gamma_{\Delta-1}(\lambda_{0:\Delta-1}) \sum_{\lambda_\Delta} \gamma_\Delta(\lambda_{0:\Delta}) \quad (6.72)$$

4. On calcule le coefficient de normalisation

$$\rho_0 = \sum_{\lambda_0} \gamma_0(\lambda_0) D_0(\lambda_0) \quad (6.73)$$

5. On obtient la mesure de probabilité  $F_0$  en utilisant l'équation suivante

$$F_0 = \sum_{\lambda_0 \in \mathcal{A}} f_0(\lambda_0) \delta_{\lambda_0} \quad (6.74)$$

avec

$$f_0(\lambda_0) = \rho_0^{-1} \gamma_0(\lambda_0) D_0(\lambda_0) \quad (6.75)$$

6. Pour tout  $j \in \{1, \dots, N\}$ , on définit la variable aléatoire  $\Lambda^{(j,0)}$  correspondant à la  $j$ -ème trajectoire à l'instant  $t = 0$  par

$$\Lambda^{(j,0)} \sim F_0 \quad (6.76)$$

et lorsque  $j > 1$ , par la condition que  $\Lambda^{(j,0)}$  est indépendant de  $\Lambda^{(1,0)}, \dots, \Lambda^{(j-1,0)}$ . On notera que l'on a déjà calculé (cf les équations 6.68, 6.70 et 6.72) les grandeurs suivantes associées à  $\Lambda^{(j,0)}$  :

- $D_0(\Lambda^{(j,0)})$  ;
- $\gamma_\tau(\Lambda^{(j,0)}, \lambda_{1:\tau})$  pour tout  $\tau \in \{1, 2, \dots, \Delta\}$  et pour tout  $\lambda_{1:\tau} \in \mathcal{A}^{\tau-1}$  ;
- $\Pi_{X_\Delta}^{(\Lambda_{0:\Delta}, Y_{0:\Delta-1})}(\Lambda^{(j,0)}, \lambda_{1:\Delta}, y_{0:\Delta-1}; \cdot)$  pour tout  $\lambda_{1:\Delta} \in \mathcal{A}^\Delta$ .

#### 6.4.3.4 Estimation de $\Lambda_{t-\delta}$

Pour tout  $t \in \mathbb{N}$  et pour tout  $\delta \in \{0, \dots, t\}$ , on peut estimer  $\Lambda_{t-\delta}$  à l'aide de la mesure  $F_t^N$  qui est une approximation de la mesure  $F_t = \Pi_{\Lambda_{0:t}}^{Y_{0:t+\Delta}}(y_{0:t+\Delta}; \cdot)$ . Cependant, dans le cas  $t > 0$ , il est préférable d'utiliser la mesure  $\tilde{F}_t^N$  qui est une meilleure approximation de  $F_t$ . On obtient alors une approximation  $\hat{\Pi}_{\Lambda_{t-\delta}}^{Y_{0:t+\Delta}}(y_{0:t+\Delta}; \cdot)$  de la loi *a posteriori* de  $\Lambda_{t-\delta}$  vérifiant, pour tout  $\lambda_{t-\delta} \in \mathcal{A}$ ,

$$\hat{\Pi}_{\Lambda_{t-\delta}}^{Y_{0:t+\Delta}}(y_{0:t+\Delta}; \{\lambda_{t-\delta}\}) = \sum_{(\lambda_{t'})_{t' \neq t-\delta}} \tilde{F}_t^N(\{\lambda_{0:t}\}) \quad (6.77)$$

En notant  $(\tilde{\Lambda}^{(k,t)})_k$  les  $N$  card( $\mathcal{A}$ ) trajectoires correspondant à toutes les prolongations possibles à l'instant  $t$  des  $N$  trajectoires  $\Lambda^{(j,t-1)}$  à l'instant  $t-1$ , l'équation 6.77 devient

$$\hat{\Pi}_{\Lambda_{t-\delta}}^{Y_{0:t+\Delta}}(y_{0:t+\Delta}; \{\lambda_{t-\delta}\}) = \sum_k \tilde{f}_t^N(\tilde{\Lambda}^{(k,t)}) \delta_{\tilde{\Lambda}_{t-\delta}^{(k,t)}, \lambda_{t-\delta}} \quad (6.78)$$

On obtient alors une estimation de  $\Lambda_{t-\delta}$  étant données les observations  $Y_{0:t+\Delta} = y_{0:t+\Delta}$  en maximisant l'expression 6.78 par rapport à  $\lambda_{t-\delta}$ .

## 6.5 Conclusion

Dans ce chapitre, nous avons introduit le concept de modèle dynamique à processus indicateur latent. Nous avons ensuite étudié l'application de l'algorithme de filtrage particulière à échantillonnage global à un modèle dynamique à processus indicateur latent. Selon la loi *a posteriori* que l'on souhaite approcher, nous avons obtenu deux algorithmes d'estimation : le filtre particulière à échantillonnage global immédiat et le filtre particulière à échantillonnage global différé.

Dans le prochain chapitre, nous allons examiner l'application de ces deux algorithmes au problème de l'égalisation.



# Chapitre 7

## Application du filtrage particulaire à l'égalisation

### 7.1 Introduction

Dans le chapitre 4, nous avons vu qu'en utilisant l'algorithme de propagation de croyance pour égaliser un canal de réponse impulsionnelle longue ou pour effectuer conjointement l'estimation de canal et l'égalisation, nous obtenons des algorithmes qui sont optimaux mais inexploitable en raison de leur trop grande complexité. Aussi, dans le chapitre 5, nous avons introduit l'algorithme de filtrage particulaire à échantillonnage global, qui permet de calculer de façon approchée une suite de mesures de probabilité, et dans le chapitre 6, nous avons appliqué cet algorithme à un modèle dynamique à processus indicateur latent. Nous avons ainsi obtenu deux algorithmes de filtrage particulaire – le filtre particulaire à échantillonnage global immédiat et le filtre particulaire à échantillonnage global différé – que nous avons décrits en détail à l'aide des opérateurs de prédiction et de correction du modèle dynamique à processus indicateur latent.

Dans ce chapitre, nous allons étudier l'application de ces deux algorithmes à l'égalisation. Comme dans le chapitre 4, nous nous intéresserons à nouveau aux trois situations suivantes : égalisation d'un canal sélectif en fréquence de réponse impulsionnelle connue, égalisation adaptative d'un canal évanescent sélectif en fréquence et égalisation adaptative d'un canal évanescent non-sélectif en fréquence. Nous constaterons alors que le réseau bayésien correspondant à chacune des ces situations possède une structure de modèle dynamique à processus indicateur latent. Nous calculerons ensuite explicitement les opérateurs de prédiction et de correction correspondant et nous pourrons alors déterminer la complexité des algorithmes d'égalisation obtenus en appliquant les algorithmes de filtrage particulaire à chacune de ces situations. Enfin, nous étudierons numériquement les performances de ces algorithmes dans le cas de l'égalisation adaptative d'un canal évanescent non-sélectif en fréquence.

### 7.2 L'égalisation d'un canal sélectif en fréquence

Dans cette partie, nous allons étudier l'application du filtrage particulaire à échantillonnage global à l'égalisation d'un canal sélectif en fréquence de réponse impulsionnelle connue. A

cette fin, nous considérerons à nouveau le système qui a été décrit au § 4.2.1, p. 55 ainsi que le réseau bayésien correspondant qui a été représenté à la figure 4.1, p. 56.

## 7.2.1 Applicabilité des algorithmes de filtrage particulière

Le réseau bayésien de la figure 4.1 possède bien une structure permettant l'application des algorithmes de filtrage particulière en prenant comme processus indicateur latent le processus  $(S_t)_{t \in \mathbb{N}}$ . De plus, les opérateurs suivants admettent une forme explicite :

1. **opérateur de prédiction de l'état** : l'opérateur  $\Psi_{t|t-1}^{s_t}$  de prédiction de l'état du système à l'instant  $t$  conditionnellement aux observations à l'instant  $t-1$  et au processus indicateur latent jusqu'à l'instant  $t$  est défini par

$$\Pi_{\vec{L}_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot) = \Psi_{t|t-1}^{s_t} \left( \Pi_{\vec{L}_{t-1}}^{(S_{0:t-1}, Y_{0:t-1})}(s_{0:t-1}, y_{0:t-1}; \cdot) \right) \quad (7.1)$$

En posant pour tout  $s_{0:t} \in \mathcal{A}^{t+1}$

$$\vec{\varphi}_t(s_{0:t}) = \begin{cases} \begin{bmatrix} s_t & s_{t-1} & \dots & s_{t-L} \end{bmatrix}^T & \text{si } t \geq L \\ \begin{bmatrix} s_t & s_{t-1} & \dots & s_0 & 0 & \dots & 0 \end{bmatrix}^T & \text{sinon} \end{cases} \quad (7.2)$$

on montre aisément que les termes  $\Pi_{\vec{L}_{t-1}}^{(S_{0:t-1}, Y_{0:t-1})}(s_{0:t-1}, y_{0:t-1}; \cdot)$  et  $\Pi_{\vec{L}_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot)$  sont de la forme suivante

$$\Pi_{\vec{L}_{t-1}}^{(S_{0:t-1}, Y_{0:t-1})}(s_{0:t-1}, y_{0:t-1}; \cdot) = \delta_{\vec{\varphi}_{t-1}(s_{0:t-1})} \quad (7.3)$$

$$\Pi_{\vec{L}_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot) = \delta_{\vec{\varphi}_t(s_{0:t})} \quad (7.4)$$

avec

$$\vec{\varphi}_t(s_{0:t}) = s_t \vec{C} + M \vec{\varphi}_{t-1}(s_{0:t-1}) \quad (7.5)$$

Afin d'appliquer l'opérateur  $\Psi_{t|t-1}^{s_t}$ , il suffit donc de calculer le vecteur  $\vec{\varphi}_t(s_{0:t})$  à partir du vecteur  $\vec{\varphi}_{t-1}(s_{0:t-1})$  à l'aide de l'équation 7.5.

2. **opérateur de prédiction de l'observation** : l'opérateur  $\Phi_t$  de prédiction de l'observation à l'instant  $t$  conditionnellement aux observations à l'instant  $t-1$  et au processus indicateur latent jusqu'à l'instant  $t$  est défini par

$$\Pi_{Y_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot) = \Phi_t \left( \Pi_{\vec{L}_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot) \right) \quad (7.6)$$

Or, on a vu que le terme  $\Pi_{\vec{L}_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot)$  est de la forme suivante

$$\Pi_{\vec{L}_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot) = \delta_{\vec{\varphi}_t(s_{0:t})} \quad (7.7)$$

et on montre aisément que le terme  $\Pi_{Y_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot)$  est de la forme suivante

$$\Pi_{Y_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot) = \mathcal{N}_C(\vec{h}_t^T \cdot \vec{\varphi}_t(s_{0:t}), \mathcal{N}_0) \quad (7.8)$$

Afin d'appliquer l'opérateur  $\Phi_t$ , il suffit donc de calculer la moyenne  $\vec{h}_t^T \cdot \vec{\varphi}_t(s_{0:t})$  de  $\Pi_{Y_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot)$  à partir de  $\vec{\varphi}_t(s_{0:t})$ .

3. **opérateur de correction de l'état** : l'opérateur  $\Psi_t^{y_t}$  de correction de l'état du système à l'instant  $t$  conditionnellement aux observations à l'instant  $t$  et au processus indicateur latent jusqu'à l'instant  $t$  est défini par

$$\Pi_{\tilde{L}_t}^{(S_{0:t}, Y_{0:t})}(s_{0:t}, y_{0:t}; \cdot) = \Psi_t^{y_t} \left( \Pi_{\tilde{L}_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot) \right) \quad (7.9)$$

Or, on a vu que le terme  $\Pi_{\tilde{L}_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot)$  est de la forme suivante

$$\Pi_{\tilde{L}_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot) = \delta_{\tilde{\varphi}_t(s_{0:t})} \quad (7.10)$$

et on montre aisément que le terme  $\Pi_{\tilde{L}_t}^{(S_{0:t}, Y_{0:t})}(s_{0:t}, y_{0:t}; \cdot)$  est de la forme suivante

$$\Pi_{\tilde{L}_t}^{(S_{0:t}, Y_{0:t})}(s_{0:t}, y_{0:t}; \cdot) = \delta_{\tilde{\varphi}_t(s_{0:t})} \quad (7.11)$$

Dans ce cas particulier, l'opérateur de correction  $\Psi_t^{y_t}$  est donc réduit à l'identité.

Par conséquent, le filtre particulière à échantillonnage global immédiat décrit au § 6.3.3 et le filtre particulière à échantillonnage global différé décrit au § 6.4.3 peuvent être appliqués au système défini au § 4.2.1 en utilisant les formes explicites des opérateurs de prédiction et de correction que nous venons de donner.

## 7.2.2 Complexité des algorithmes d'égalisation obtenus

### 7.2.2.1 Filtre particulière à échantillonnage global immédiat

Pour tout  $t \in \{1, 2, \dots, t_0\}$ , le calcul de la mesure  $\tilde{F}_t^N$  à partir de la mesure  $F_{t-1}^N$  nécessite un nombre d'opérations en  $O(NL \text{card}(\mathcal{A}))$ . Le calcul de  $F_t^N$  à partir de la mesure  $\tilde{F}_t^N$  par échantillonnage multinomial ou résiduel nécessite un nombre d'opérations en  $O(N \text{card}(\mathcal{A}))$  [DGK01]. Par conséquent, la complexité de l'algorithme de particulière à échantillonnage global immédiat est  $O(t_0 NL \text{card}(\mathcal{A}))$ .

### 7.2.2.2 Filtre particulière à échantillonnage global différé

Pour tout  $t \in \{1, 2, \dots, t_0\}$ , le calcul de la mesure  $\tilde{F}_t^N$  à partir de la mesure  $F_{t-1}^N$  nécessite un nombre d'opérations en  $O(NL \text{card}(\mathcal{A})^{\Delta+1})$ . Comme dans le cas du filtre particulière à échantillonnage global immédiat, le calcul de  $F_t^N$  à partir de la mesure  $\tilde{F}_t^N$  par échantillonnage multinomial ou résiduel nécessite un nombre d'opérations en  $O(N \text{card}(\mathcal{A}))$ . Par conséquent, la complexité de l'algorithme de particulière à échantillonnage global différé est  $O(t_0 NL \text{card}(\mathcal{A})^{\Delta+1})$ .

## 7.2.3 Conclusion

Alors que la complexité de l'algorithme optimal obtenu au § 4.2.4 en appliquant l'algorithme de propagation de croyance était exponentielle par rapport à la longueur  $L$  de la réponse impulsionnelle du canal, les techniques de filtrage particulière nous ont permis d'obtenir des algorithmes dont la complexité est linéaire par rapport à  $L$ . Cette réduction de complexité a été obtenue en effectuant un compromis entre les performances et la complexité, compromis que l'on peut ajuster à l'aide du nombre de particules  $N$  (cf. § 5.3).



## 7.3 L'égalisation adaptative d'un canal évanescent sélectif en fréquence

Dans cette partie, nous allons étudier l'application du filtrage particulière à échantillonnage global à l'égalisation adaptative d'un canal évanescent sélectif en fréquence. A cette fin, nous considérerons à nouveau le système qui a été décrit au § 4.3.1, p. 60 ainsi que le réseau bayésien correspondant qui a été représenté à la figure 4.4, p. 61.

### 7.3.1 Applicabilité des algorithmes de filtrage particulière

Le réseau bayésien de la figure 4.4 possède bien une structure permettant l'application des algorithmes de filtrage particulière en prenant comme processus indicateur latent le processus  $(S_t)_{t \in \mathbb{N}}$ . De plus, les opérateurs suivants admettent une forme explicite :

1. **opérateur de prédiction de l'état** : l'opérateur  $\Psi_{t|t-1}^{s_t}$  de prédiction de l'état du système à l'instant  $t$  conditionnellement aux observations à l'instant  $t-1$  et au processus indicateur latent jusqu'à l'instant  $t$  est défini par

$$\Pi_{\vec{B}_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot) = \Psi_{t|t-1}^{s_t} \left( \Pi_{\vec{B}_{t-1}}^{(S_{0:t-1}, Y_{0:t-1})}(s_{0:t-1}, y_{0:t-1}; \cdot) \right) \quad (7.12)$$

En posant pour tout  $s_{0:t} \in \mathcal{A}^{t+1}$

$$\vec{\varphi}_t(s_{0:t}) = \begin{cases} \begin{bmatrix} s_t & s_{t-1} & \dots & s_{t-L} \end{bmatrix}^T & \text{si } t \geq L \\ \begin{bmatrix} s_t & s_{t-1} & \dots & s_0 & 0 & \dots & 0 \end{bmatrix}^T & \text{sinon} \end{cases} \quad (7.13)$$

on montre aisément que les termes  $\Pi_{\vec{B}_{t-1}}^{(S_{0:t-1}, Y_{0:t-1})}(s_{0:t-1}, y_{0:t-1}; \cdot)$  et  $\Pi_{\vec{B}_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot)$  sont de la forme suivante

$$\Pi_{\vec{B}_{t-1}}^{(S_{0:t-1}, Y_{0:t-1})}(s_{0:t-1}, y_{0:t-1}; \cdot) = \delta_{\vec{\varphi}_{t-1}(s_{0:t-1})} \otimes \dots \quad (7.14)$$

$$\dots \mathcal{N}(\vec{m}_{t-1}(s_{0:t-1}, y_{0:t-1}), \Sigma_{t-1}(s_{0:t-1}))$$

$$\Pi_{\vec{B}_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot) = \delta_{\vec{\varphi}_t(s_{0:t})} \otimes \dots \quad (7.15)$$

$$\dots \mathcal{N}(\vec{m}_{t|t-1}(s_{0:t-1}, y_{0:t-1}), \Sigma_{t|t-1}(s_{0:t-1}))$$

avec

$$\vec{m}_{t|t-1}(s_{0:t-1}, y_{0:t-1}) = \Lambda \vec{m}_{t-1}(s_{0:t-1}, y_{0:t-1}) \quad (7.16)$$

$$\Sigma_{t|t-1}(s_{0:t-1}) = \Lambda \Sigma_{t-1}(s_{0:t-1}) \Lambda^\dagger + R_G \quad (7.17)$$

$$\vec{\varphi}_t(s_{0:t}) = s_t \vec{C} + M \vec{\varphi}_{t-1}(s_{0:t-1}) \quad (7.18)$$

Afin d'appliquer l'opérateur  $\Psi_{t|t-1}^{s_t}$ , il suffit donc de calculer le vecteur  $\vec{\varphi}_t(s_{0:t})$  à partir du vecteur  $\vec{\varphi}_{t-1}(s_{0:t-1})$  à l'aide de l'équation 7.18, ainsi que les matrices  $(\vec{m}_{t|t-1}, \Sigma_{t|t-1})$  à partir des matrices  $(\vec{m}_{t-1}, \Sigma_{t-1})$  à l'aide des équations 7.16 et 7.17.

2. **opérateur de prédiction de l'observation** : l'opérateur  $\Phi_t$  de prédiction de l'observation à l'instant  $t$  conditionnellement aux observations à l'instant  $t-1$  et au processus indicateur latent jusqu'à l'instant  $t$  est défini par

$$\Pi_{Y_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot) = \Phi_t \left( \Pi_{\vec{B}_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot) \right) \quad (7.19)$$

Or, on a vu que le terme  $\Pi_{\vec{B}_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot)$  est de la forme suivante

$$\begin{aligned} \Pi_{\vec{B}_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot) &= \delta_{\vec{\varphi}_t(s_{0:t})} \otimes \dots \\ &\dots \mathcal{N}_C(\vec{m}_{t|t-1}(s_{0:t-1}, y_{0:t-1}), \Sigma_{t|t-1}(s_{0:t-1})) \end{aligned} \quad (7.20)$$

et on montre aisément que le terme  $\Pi_{Y_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot)$  est de la forme suivante

$$\Pi_{Y_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot) = \mathcal{N}_C \left( \nu_t(s_{0:t}, y_{0:t-1}), \Gamma_t(s_{0:t}) \right) \quad (7.21)$$

avec

$$\nu_t(s_{0:t}, y_{0:t-1}) = \vec{\varphi}_t(s_{0:t})^T \vec{m}_{t|t-1}(s_{0:t-1}, y_{0:t-1}) \quad (7.22)$$

$$\Gamma_t(s_{0:t}) = \vec{\varphi}_t(s_{0:t})^T \Sigma_{t|t-1}(s_{0:t-1}) \vec{\varphi}_t^*(s_{0:t}) + \mathcal{N}_0 \quad (7.23)$$

Afin d'appliquer l'opérateur  $\Phi_t$ , il suffit donc de calculer  $(\nu_t, \Gamma_t)$  à partir de  $(\vec{m}_{t|t-1}, \Sigma_{t|t-1}, \vec{\varphi}_t)$  à l'aide des équations 7.22 et 7.23.

3. **opérateur de correction de l'état** : l'opérateur  $\Psi_t^{y_t}$  de correction de l'état du système à l'instant  $t$  conditionnellement aux observations à l'instant  $t$  et au processus indicateur latent jusqu'à l'instant  $t$  est défini par

$$\Pi_{\vec{B}_t}^{(S_{0:t}, Y_{0:t})}(s_{0:t}, y_{0:t}; \cdot) = \Psi_t^{y_t} \left( \Pi_{\vec{B}_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot) \right) \quad (7.24)$$

Or, on a vu que le terme  $\Pi_{\vec{B}_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot)$  est de la forme suivante

$$\begin{aligned} \Pi_{\vec{B}_t}^{(S_{0:t}, Y_{0:t-1})}(s_{0:t}, y_{0:t-1}; \cdot) &= \delta_{\vec{\varphi}_t(s_{0:t})} \otimes \dots \\ &\dots \mathcal{N}_C(\vec{m}_{t|t-1}(s_{0:t-1}, y_{0:t-1}), \Sigma_{t|t-1}(s_{0:t-1})) \end{aligned} \quad (7.25)$$

et on montre aisément que le terme  $\Pi_{\vec{B}_t}^{(S_{0:t}, Y_{0:t})}(s_{0:t}, y_{0:t}; \cdot)$  est de la forme suivante

$$\Pi_{\vec{B}_t}^{(S_{0:t}, Y_{0:t})}(s_{0:t}, y_{0:t}; \cdot) = \delta_{\vec{\varphi}_t(s_{0:t})} \otimes \mathcal{N}_C(\vec{m}_t(s_{0:t}, y_{0:t}), \Sigma_t(s_{0:t})) \quad (7.26)$$

avec

$$\begin{aligned} \vec{m}_t(s_{0:t}, y_{0:t}) &= \vec{m}_{t|t-1}(s_{0:t-1}, y_{0:t-1}) + \dots \\ &\dots + [y_t - \vec{\varphi}_t(s_{0:t})^T \vec{m}_{t|t-1}(s_{0:t-1}, y_{0:t-1})] \vec{K}_t(s_{0:t}) \end{aligned} \quad (7.27)$$

$$\Sigma_t(s_{0:t}) = [I_{L+1} - \vec{K}_t(s_{0:t}) \vec{\varphi}_t(s_{0:t})^T] \Sigma_{t|t-1}(s_{0:t-1}) \quad (7.28)$$

$$\vec{K}_t(s_{0:t}) = [\mathcal{N}_0 + \vec{\varphi}_t(s_{0:t})^T \Sigma_{t|t-1}(s_{0:t-1}) \vec{\varphi}_t^*(s_{0:t})]^{-1} \Sigma_{t|t-1}(s_{0:t-1}) \vec{\varphi}_t^*(s_{0:t}) \quad (7.29)$$

Afin d'appliquer l'opérateur  $\Psi_t^{y_t}$ , il suffit donc de calculer  $(\vec{m}_t, \Sigma_t)$  à partir de  $(\vec{m}_{t|t-1}, \Sigma_{t|t-1}, \vec{\varphi}_t)$  à l'aide des équations 7.27, 7.28 et 7.29.

Par conséquent, le filtre particulière à échantillonnage global immédiat décrit au § 6.3.3 et le filtre particulière à échantillonnage global différé décrit au § 6.4.3 peuvent être appliqués au système défini au § 4.3.1 en utilisant les formes explicites des opérateurs de prédiction et de correction que nous venons de donner.

### 7.3.2 Complexité de l'algorithme

En procédant comme au § 7.2.2, et en remarquant que les calculs correspondant aux opérateurs  $\Psi_{t|t-1}^{s_t}$ ,  $\Phi_t$  et  $\Psi_t^{y_t}$  exigent maintenant un nombre d'opérations en  $O(L^2)$ , on montre que, dans ce cas, la complexité de l'algorithme de filtrage particulière à échantillonnage global immédiat est  $O(t_0NL^2 \text{card}(\mathcal{A}))$  et celle de l'algorithme de filtrage particulière à échantillonnage global différé  $O(t_0NL^2 \text{card}(\mathcal{A})^{\Delta+1})$ .

### 7.3.3 Conclusion

Alors que la complexité de l'algorithme optimal obtenu au § 4.3.4 en appliquant l'algorithme de propagation de croyance était exponentielle par rapport au temps, les techniques de filtrage particulière nous ont permis d'obtenir des algorithmes dont la complexité est linéaire par rapport au temps. Comme nous l'avons déjà remarqué au § 7.2.3, cette réduction de complexité a été obtenue en effectuant un compromis entre les performances et la complexité, compromis que l'on peut ajuster à l'aide du nombre de particules  $N$ .

## 7.4 L'égalisation adaptative d'un canal évanescent non-sélectif en fréquence

Dans cette partie, nous allons étudier l'application du filtrage particulière à échantillonnage global à l'égalisation adaptative d'un canal évanescent non-sélectif en fréquence. A cette fin, nous considérerons à nouveau le système qui a été décrit au § 4.4.1, p. 68 ainsi que le réseau bayésien correspondant qui a été représenté à la figure 4.9, p. 70.

### 7.4.1 Applicabilité des algorithmes de filtrage particulière

Le réseau bayésien de la figure 4.9 possède bien une structure permettant l'application des algorithmes de filtrage particulière en prenant comme processus indicateur latent le processus  $(A_t)_{t \in \mathbb{N}}$ . De plus, les opérateurs suivants admettent une forme explicite :

1. **opérateur de prédiction de l'état** : l'opérateur  $\Psi_{t|t-1}^{a_t}$  de prédiction de l'état du système à l'instant  $t$  conditionnellement aux observations à l'instant  $t-1$  et au processus indicateur latent jusqu'à l'instant  $t$  est défini par

$$\Pi_{\vec{B}_t}^{(A_{0:t}, Y_{0:t-1})}(a_{0:t}, y_{0:t-1}; \cdot) = \Psi_{t|t-1}^{a_t} \left( \Pi_{\vec{B}_{t-1}}^{(A_{0:t-1}, Y_{0:t-1})}(a_{0:t-1}, y_{0:t-1}; \cdot) \right) \quad (7.30)$$

En posant pour tout  $a_{0:t} \in \mathcal{A}^{t+1}$

$$\varphi_t(a_{0:t}) = a_0 a_1 \dots a_t \quad (7.31)$$

on montre aisément que les termes  $\Pi_{\vec{B}_{t-1}}^{(A_{0:t-1}, Y_{0:t-1})}(a_{0:t-1}, y_{0:t-1}; \cdot)$  et  $\Pi_{\vec{B}_t}^{(A_{0:t}, Y_{0:t-1})}(a_{0:t}, y_{0:t-1}; \cdot)$  sont de la forme suivante

$$\begin{aligned} \Pi_{\vec{B}_{t-1}}^{(A_{0:t-1}, Y_{0:t-1})}(a_{0:t-1}, y_{0:t-1}; \cdot) &= \delta_{\varphi_{t-1}(a_{0:t-1})} \otimes \dots \\ &\dots \mathcal{N}_C(\vec{m}_{t-1}(a_{0:t-1}, y_{0:t-1}), \Sigma_{t-1}(a_{0:t-1})) \end{aligned} \quad (7.32)$$

$$\begin{aligned} \Pi_{\vec{B}_t}^{(A_{0:t}, Y_{0:t-1})}(a_{0:t}, y_{0:t-1}; \cdot) &= \delta_{\varphi_t(a_{0:t})} \otimes \dots \\ &\dots \mathcal{N}_C(\vec{m}_{t|t-1}(a_{0:t-1}, y_{0:t-1}), \Sigma_{t|t-1}(a_{0:t-1})) \end{aligned} \quad (7.33)$$

avec

$$\vec{m}_{t|t-1}(a_{0:t-1}, y_{0:t-1}) = \Lambda \vec{m}_{t-1}(a_{0:t-1}, y_{0:t-1}) \quad (7.34)$$

$$\Sigma_{t|t-1}(a_{0:t-1}) = \Lambda \Sigma_{t-1}(a_{0:t-1}) \Lambda^\dagger + R_G \quad (7.35)$$

$$\varphi_t(a_{0:t}) = a_t \cdot \varphi_{t-1}(a_{0:t-1}) \quad (7.36)$$

Afin d'appliquer l'opérateur  $\Psi_{t|t-1}^{a_t}$ , il suffit donc de calculer le terme  $\varphi_t(a_{0:t})$  à partir du terme  $\varphi_{t-1}(a_{0:t-1})$  à l'aide de l'équation 7.36, ainsi que les matrices  $(\vec{m}_{t|t-1}, \Sigma_{t|t-1})$  à partir des matrices  $(\vec{m}_{t-1}, \Sigma_{t-1})$  à l'aide des équations 7.34 et 7.35.

2. **opérateur de prédiction de l'observation** : l'opérateur  $\Phi_t$  de prédiction de l'observation à l'instant  $t$  conditionnellement aux observations à l'instant  $t-1$  et au processus indicateur latent jusqu'à l'instant  $t$  est défini par

$$\Pi_{Y_t}^{(A_{0:t}, Y_{0:t-1})}(a_{0:t}, y_{0:t-1}; \cdot) = \Phi_t \left( \Pi_{\vec{B}_t}^{(A_{0:t}, Y_{0:t-1})}(a_{0:t}, y_{0:t-1}; \cdot) \right) \quad (7.37)$$

Or, on a vu que le terme  $\Pi_{\vec{B}_t}^{(A_{0:t}, Y_{0:t-1})}(a_{0:t}, y_{0:t-1}; \cdot)$  est de la forme suivante

$$\begin{aligned} \Pi_{\vec{B}_t}^{(A_{0:t}, Y_{0:t-1})}(a_{0:t}, y_{0:t-1}; \cdot) &= \delta_{\varphi_t(a_{0:t})} \otimes \dots \\ &\dots \mathcal{N}_C(\vec{m}_{t|t-1}(a_{0:t-1}, y_{0:t-1}), \Sigma_{t|t-1}(a_{0:t-1})) \end{aligned} \quad (7.38)$$

et on montre aisément que le terme  $\Pi_{Y_t}^{(A_{0:t}, Y_{0:t-1})}(a_{0:t}, y_{0:t-1}; \cdot)$  est de la forme suivante

$$\Pi_{Y_t}^{(A_{0:t}, Y_{0:t-1})}(a_{0:t}, y_{0:t-1}; \cdot) = \mathcal{N}_C \left( \nu_t(a_{0:t}, y_{0:t-1}), \Gamma_t(a_{0:t}) \right) \quad (7.39)$$

avec

$$\nu_t(a_{0:t}, y_{0:t-1}) = \varphi_t(a_{0:t}) \vec{\Theta}^T \vec{m}_{t|t-1}(a_{0:t-1}, y_{0:t-1}) \quad (7.40)$$

$$\Gamma_t(a_{0:t}) = \vec{\Theta}^T \Sigma_{t|t-1}(a_{0:t-1}) \vec{\Theta}^* + \mathcal{N}_0 \quad (7.41)$$

Afin d'appliquer l'opérateur  $\Phi_t$ , il suffit donc de calculer  $(\nu_t, \Gamma_t)$  à partir de  $(\vec{m}_{t|t-1}, \Sigma_{t|t-1}, \varphi_t)$  à l'aide des équations 7.40 et 7.41.

3. **opérateur de correction de l'état** : l'opérateur  $\Psi_t^{y_t}$  de correction de l'état du système à l'instant  $t$  conditionnellement aux observations à l'instant  $t$  et au processus indicateur latent jusqu'à l'instant  $t$  est défini par

$$\Pi_{\vec{B}_t}^{(A_{0:t}, Y_{0:t})}(a_{0:t}, y_{0:t}; \cdot) = \Psi_t^{y_t} \left( \Pi_{\vec{B}_t}^{(A_{0:t}, Y_{0:t-1})}(a_{0:t}, y_{0:t-1}; \cdot) \right) \quad (7.42)$$

Or, on a vu que le terme  $\Pi_{\vec{B}_t}^{(A_{0:t}, Y_{0:t-1})}(a_{0:t}, y_{0:t-1}; \cdot)$  est de la forme suivante

$$\begin{aligned} \Pi_{\vec{B}_t}^{(A_{0:t}, Y_{0:t-1})}(a_{0:t}, y_{0:t-1}; \cdot) &= \delta_{\varphi_t(a_{0:t})} \otimes \dots \\ &\dots \mathcal{N}_C(\vec{m}_{t|t-1}(a_{0:t-1}, y_{0:t-1}), \Sigma_{t|t-1}(a_{0:t-1})) \end{aligned} \quad (7.43)$$

et on montre aisément que le terme  $\Pi_{\vec{B}_t}^{(A_{0:t}, Y_{0:t})}(a_{0:t}, y_{0:t}; \cdot)$  est de la forme suivante

$$\Pi_{\vec{B}_t}^{(A_{0:t}, Y_{0:t})}(a_{0:t}, y_{0:t}; \cdot) = \delta_{\varphi_t(a_{0:t})} \otimes \mathcal{N}_C(\vec{m}_t(a_{0:t}, y_{0:t}), \Sigma_t(a_{0:t})) \quad (7.44)$$

avec

$$\begin{aligned} \vec{m}_t(a_{0:t}, y_{0:t}) &= \vec{m}_{t|t-1}(a_{0:t-1}, y_{0:t-1}) + \dots \\ &\dots + [y_t - \varphi_t(a_{0:t}) \vec{\Theta}^T \vec{m}_{t|t-1}(a_{0:t-1}, y_{0:t-1})] \vec{K}_t(a_{0:t}) \end{aligned} \quad (7.45)$$

$$\Sigma_t(a_{0:t}) = [I_{L+1} - \varphi_t(a_{0:t}) \vec{K}_t(a_{0:t}) \vec{\Theta}^T] \Sigma_{t|t-1}(a_{0:t-1}) \quad (7.46)$$

$$\vec{K}_t(a_{0:t}) = [\mathcal{N}_0 + \vec{\Theta}^T \Sigma_{t|t-1}(a_{0:t-1}) \vec{\Theta}^*]^{-1} \varphi_t^*(a_{0:t}) \Sigma_{t|t-1}(a_{0:t-1}) \vec{\Theta}^* \quad (7.47)$$

Afin d'appliquer l'opérateur  $\Psi_t^{y_t}$ , il suffit donc de calculer  $(\vec{m}_t, \Sigma_t)$  à partir de  $(\vec{m}_{t|t-1}, \Sigma_{t|t-1}, \varphi_t)$  à l'aide des équations 7.45, 7.46 et 7.47.

Par conséquent, le filtre particulière à échantillonnage global immédiat décrit au § 6.3.3 et le filtre particulière à échantillonnage global différé décrit au § 6.4.3 peuvent être appliqués au système défini au § 4.4.1 en utilisant les formes explicites des opérateurs de prédiction et de correction que nous venons de donner.

## 7.4.2 Complexité de l'algorithme

En procédant comme au § 7.2.2, et en remarquant que les calculs correspondant aux opérateurs  $\Psi_{t|t-1}^{a_t}$ ,  $\Phi_t$  et  $\Psi_t^{y_t}$  exigent un nombre d'opérations en  $O(L^2)$ , on montre que, dans ce cas, la complexité de l'algorithme de filtrage particulière à échantillonnage global immédiat est  $O(t_0 N L^2 \text{card}(\mathcal{A}))$  et celle de l'algorithme de filtrage particulière à échantillonnage global différé  $O(t_0 N L^2 \text{card}(\mathcal{A})^{\Delta+1})$ .

## 7.4.3 Conclusion

Alors que la complexité de l'algorithme optimal obtenu au § 4.4.4 en appliquant l'algorithme de propagation de croyance était exponentielle par rapport au temps, les techniques de filtrage particulière nous ont permis à nouveau d'obtenir des algorithmes sous-optimaux dont la complexité est linéaire par rapport au temps.

Dans la prochaine partie, nous allons étudier numériquement les performances de ces algorithmes.

## 7.5 Simulations numériques

Nous avons étudié numériquement les performances de l'algorithme de filtrage particulière à échantillonnage global appliqué à l'égalisation adaptative d'un canal évanescent non-sélectif

en fréquence. Pour permettre la comparaison avec des travaux publiés antérieurement, nous avons considéré l'exemple étudié dans [CWL00, § VIII]. Dans cet exemple, le canal évanescent est modélisé par un processus  $(H_t)$  ARMA(3, 3) vérifiant

$$\begin{aligned} H_t - 2,374\ 09H_{t-1} + 1,929\ 36H_{t-2} - 0,532\ 08H_{t-3} \\ = 10^{-2}(0,894\ 09\nu_t + 2,682\ 27\nu_{t-1} + 2,682\ 27\nu_{t-2} + 0,894\ 09\nu_{t-3}) \end{aligned} \quad (7.48)$$

avec  $\nu_t \sim \mathcal{N}_C(0, 1)$ . Le processus  $(H_t)$  est donc obtenu à l'aide d'un filtre de Butterworth de fréquence de coupure 0,05, ce qui correspond à un canal à évanouissement rapide. On suppose également que l'on utilise une modulation par déplacement de phase à deux états, précédé par un codage différentiel. Autrement dit, les symboles de modulation  $S_t$  appartiennent à l'alphabet  $\mathcal{A} = \{-1, +1\}$  et sont obtenus par la relation  $S_t = A_t S_{t-1}$  à partir d'une suite de bits  $A_t \in \{-1, +1\}$ , indépendants et identiquement distribués selon une loi de Bernoulli de paramètre 1/2.

Les performances du récepteur à échantillonnage global ont été comparées à celles des récepteurs suivants :

1. *récepteur à canal connu* : on suppose que le récepteur connaît les véritables coefficients  $H_t$  et on utilise la règle de détection suivante :  $\hat{A}_t = \hat{S}_t \hat{S}_{t-1}$  avec  $\hat{S}_t = \text{sign}(\Re\{H_t^* Y_t\})$  ;
2. *récepteur aidé par un génie* : on suppose qu'un « génie » permet au récepteur d'observer  $\tilde{Y}_t = H_t + \tilde{V}_t$  avec  $\tilde{V}_t \sim \mathcal{N}_C(0, \mathcal{N}_0)$ . On utilise  $\tilde{Y}_t$  pour calculer par filtrage de Kalman une estimation  $\hat{H}_t$  du coefficient  $H_t$  et on utilise la règle de détection suivante :  $\hat{A}_t = \hat{S}_t \hat{S}_{t-1}$  avec  $\hat{S}_t = \text{sign}(\Re\{\hat{H}_t^* Y_t\})$  ;
3. *récepteur à détection différentielle* : ce récepteur n'essaie pas d'estimer les coefficients du canal. Les bits  $A_t$  sont estimés en utilisant la règle de détection différentielle incohérente :  $\hat{A}_t = \text{sign}(\Re\{Y_t^* Y_{t-1}\})$  ;
4. *récepteur à mélange de filtres de Kalman (MFK)* : le filtre particulière décrit dans [CWL00, § IV & V] est utilisé pour estimer les bits  $A_t$ . Le récepteur MFK utilise l'algorithme SISR afin de tirer des échantillons dans l'espace du processus indicateur et utilise un filtre de Kalman pour chaque trajectoire afin de calculer la loi de proposition et le poids d'importance de chaque trajectoire. Un rééchantillonnage a lieu lorsque le rapport entre la taille effective de l'échantillon [CWL00, Eq. (45)] et la taille réelle  $N$  de l'échantillon est inférieur à un seuil  $\beta$ . La méthode des poids différés décrite dans [CWL00, § V.A] est utilisée pour obtenir une estimation de  $A_t$  avec un retard  $\delta$ .

Dans toutes les simulations présentées ci-dessous, nous avons exclusivement utilisé des filtres particuliers à échantillonnage immédiat car dans les différentes situations considérées, l'utilisation d'un filtre particulier à échantillonnage différé n'apportait pas d'amélioration significative. Ceci est dû au fait que nous n'avons considéré, dans cette partie de la thèse, que des systèmes ne comportant pas de codage de canal.

Les figures 7.1 et 7.5 montrent la courbe de taux d'erreur binaire (TEB) de chaque récepteur en fonction du rapport signal sur bruit (RSB) pour les retards d'estimation respectifs  $\delta = 0$  et  $\delta = 1$ . Le RSB est défini comme étant le rapport  $\text{var}(H_t)/\text{var}(V_t)$  et le TEB a été obtenu en moyennant le taux d'erreur sur  $10^6$  symboles. Les 50 premiers symboles n'ont pas été pris en compte dans cette moyenne. Nous avons également représenté sur ces figures les courbes de taux d'erreur pour le récepteur à canal connu, le récepteur aidé par un génie, le récepteur à détection différentielle et le récepteur MFK. Dans le cas du récepteur MFK, nous avons considéré

les retards d'estimation  $\delta = 0$  et  $\delta = 1$  ainsi que les seuils de rééchantillonnage  $\beta = 0,1$  et  $\beta = 1$  (rééchantillonnage systématique). Les courbes des figures 7.1 et 7.5 ont été obtenues avec un récepteur à échantillonnage global et un récepteur MFK utilisant chacun 50 particules. Ces courbes montrent qu'avec 50 particules il n'existe pas de différence significative de performance entre le récepteur proposé et un récepteur MFK utilisant le même retard d'estimation. On note également qu'avec 50 particules, les performances du récepteur MFK ne changent pas de façon significative lorsque l'on modifie le seuil de rééchantillonnage  $\beta$  du récepteur MFK de  $\beta = 0,1$  à  $\beta = 1$ . Enfin, comme noté par [CWL00], on remarque que les performances des récepteurs à filtre particulaire s'améliorent de façon significative lorsque l'on augmente le retard d'estimation de  $\delta = 0$  à  $\delta = 1$ . Les performances du récepteur à échantillonnage global et du récepteur MFK sont alors très proches de la borne inférieure correspondant au récepteur aidé par un génie.

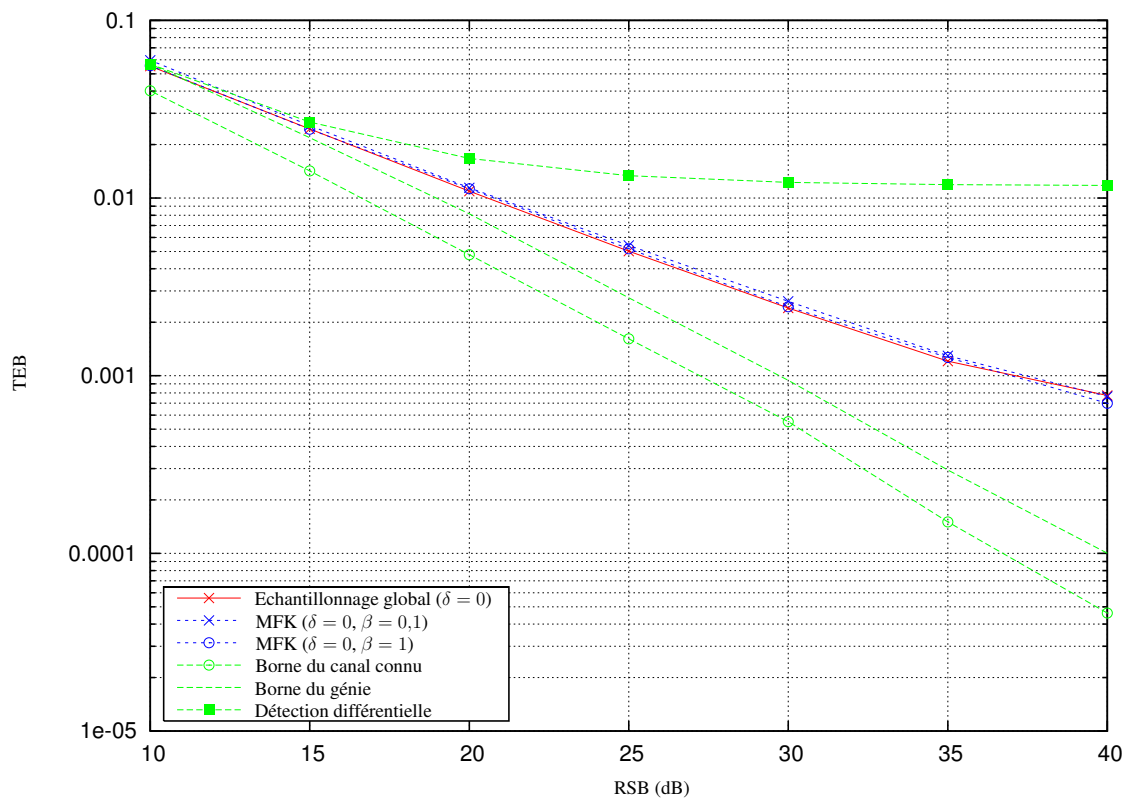


FIG. 7.1 – Courbe de taux d'erreur binaire du récepteur à échantillonnage global en fonction du rapport signal sur bruit pour un retard d'estimation  $\delta = 0$  et un nombre de particules  $N = 50$ . On a également représenté sur cette figure la courbe de taux d'erreur binaire du récepteur MFK avec  $\delta = 0$ ,  $N = 50$  et successivement  $\beta = 0,1$  et  $\beta = 1$ ; la courbe de taux d'erreur binaire du récepteur à canal connu; la courbe de taux d'erreur binaire du récepteur aidé par un génie et la courbe de taux d'erreur binaire du récepteur à détection différentielle.

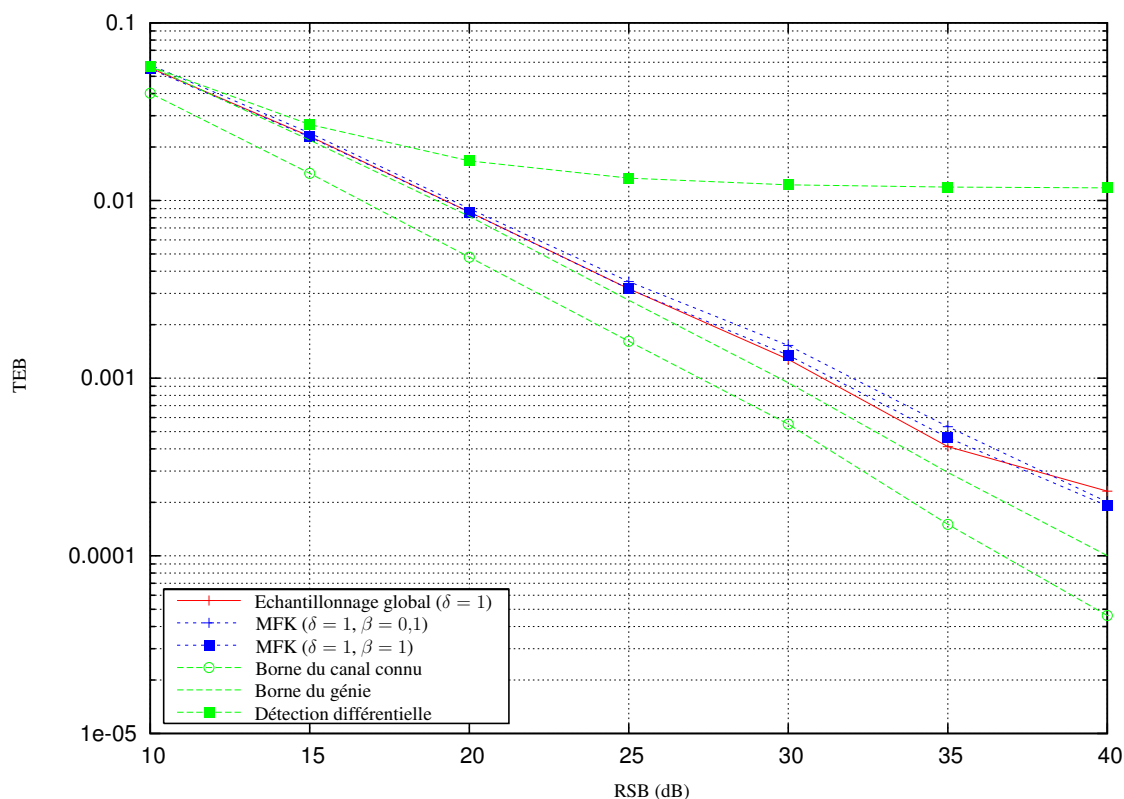


FIG. 7.2 – Courbe de taux d’erreur binaire du récepteur à échantillonnage global en fonction du rapport signal sur bruit pour un retard d’estimation  $\delta = 1$  et un nombre de particules  $N = 50$ . On a également représenté sur cette figure la courbe de taux d’erreur binaire du récepteur MFK avec  $\delta = 1$ ,  $N = 50$  et successivement  $\beta = 0,1$  et  $\beta = 1$  ; la courbe de taux d’erreur binaire du récepteur à canal connu ; la courbe de taux d’erreur binaire du récepteur aidé par un génie et la courbe de taux d’erreur binaire du récepteur à détection différentielle.

La figure 7.3 montre la courbe de taux d’erreur binaire du récepteur à échantillonnage global en fonction du nombre de particules pour un rapport signal sur bruit de 20 dB et un retard d’estimation  $\delta = 1$ . On a également représenté sur cette figure la courbe de taux d’erreur binaire du récepteur MFK avec, successivement, un seuil de rééchantillonnage  $\beta = 0,1$  et un seuil de rééchantillonnage  $\beta = 1$ . En examinant cette figure, on constate que lorsque le nombre de particules décroît de  $N = 50$  à  $N = 10$ , le taux d’erreur binaire du récepteur MFK utilisant un seuil de rééchantillonnage  $\beta = 0,1$  augmente de 67 % alors que le taux d’erreur binaire du récepteur à échantillonnage global n’augmente que de 11 % seulement. La figure 7.3 montre également que, dans la situation étudiée, la courbe de taux d’erreur binaire du récepteur à échantillonnage global est identique à celle d’un récepteur MFK utilisant un seuil de rééchantillonnage  $\beta = 1$  (rééchantillonnage systématique). Ce résultat suggère que, contrairement à ce qui est habituellement affirmé dans la littérature [CWL00, DGA00], un rééchantillonnage systématique des particules semble, pour des raisons qui restent encore obscures sur le plan théorique, être plus robuste qu’un rééchantillonnage non-systématique lorsque l’on réduit le nombre de particules afin de satisfaire les contraintes d’implémentation requises par les applications en temps-réel.



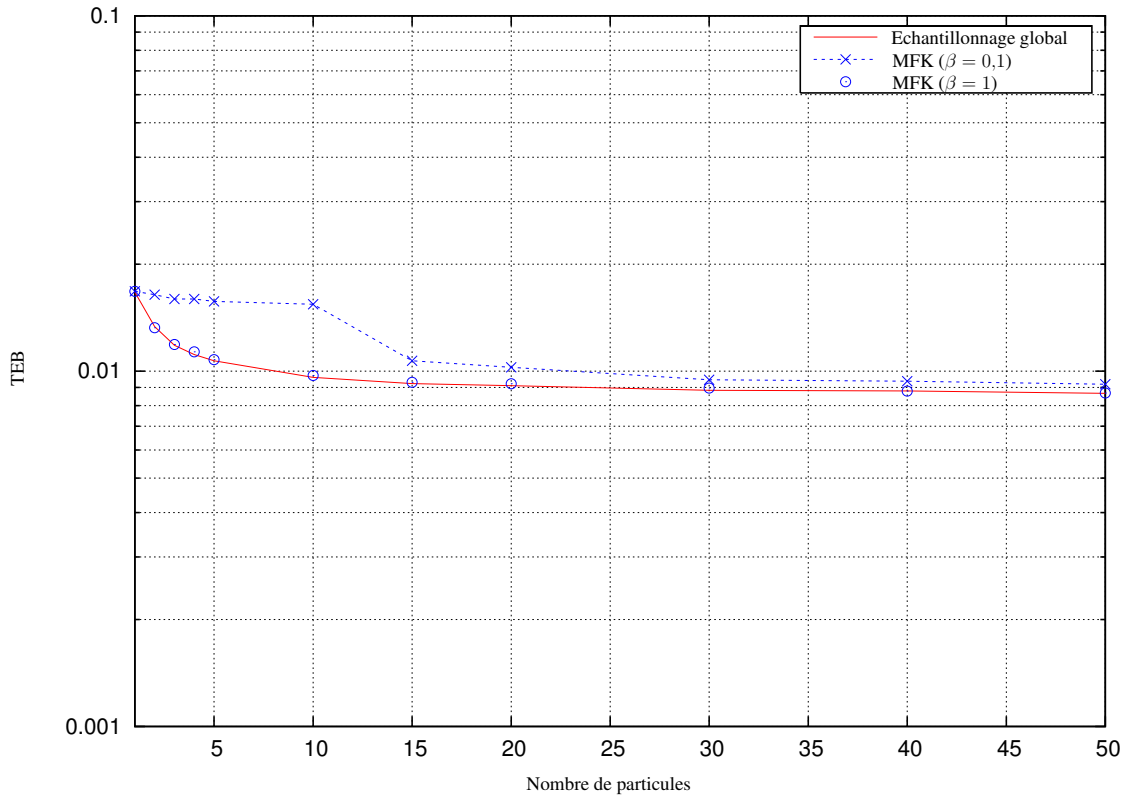


FIG. 7.3 – Courbe de taux d'erreur binaire du récepteur à échantillonnage global en fonction du nombre de particules pour un rapport signal sur bruit de 20 dB et un retard d'estimation  $\delta = 1$ . On a également représenté sur cette figure la courbe de taux d'erreur binaire du récepteur MFK avec  $\delta = 1$  et successivement  $\beta = 0,1$  et  $\beta = 1$ .

Les figures 7.4 et 7.5 montrent la courbe de taux d'erreur binaire de chaque récepteur en fonction du rapport signal sur bruit pour les retards d'estimation respectifs  $\delta = 0$  et  $\delta = 1$  lorsque le récepteur à échantillonnage global et le récepteur MFK utilisent 5 particules. En examinant ces deux figures, on constate qu'avec 5 particules il existe une différence significative de performance entre le récepteur proposé et le récepteur MFK utilisant le même retard d'estimation et un seuil de rééchantillonnage  $\beta = 0,1$ . Cette différence reste significative même pour des valeurs de rapport signal sur bruit proches de 10 dB. Ces deux figures montrent également que, dans la situation étudiée, la courbe de taux d'erreur binaire du récepteur à échantillonnage global est identique à celle d'un récepteur MFK utilisant le même retard d'estimation et un seuil de rééchantillonnage  $\beta = 1$ .

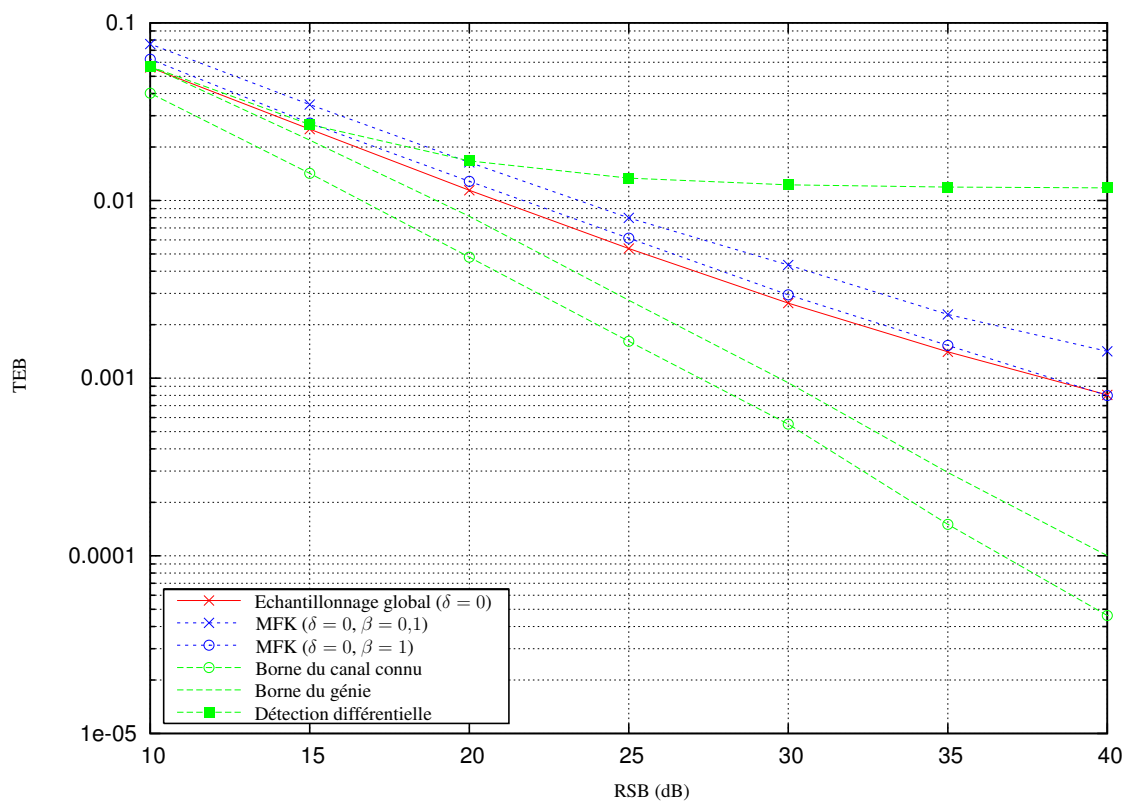


FIG. 7.4 – Courbe de taux d’erreur binaire du récepteur à échantillonnage global en fonction du rapport signal sur bruit pour un retard d’estimation  $\delta = 0$  et un nombre de particules  $N = 5$ . On a également représenté sur cette figure la courbe de taux d’erreur binaire du récepteur MFK avec  $\delta = 0$ ,  $N = 5$  et successivement  $\beta = 0,1$  et  $\beta = 1$  ; la courbe de taux d’erreur binaire du récepteur à canal connu ; la courbe de taux d’erreur binaire du récepteur aidé par un génie et la courbe de taux d’erreur binaire du récepteur à détection différentielle.

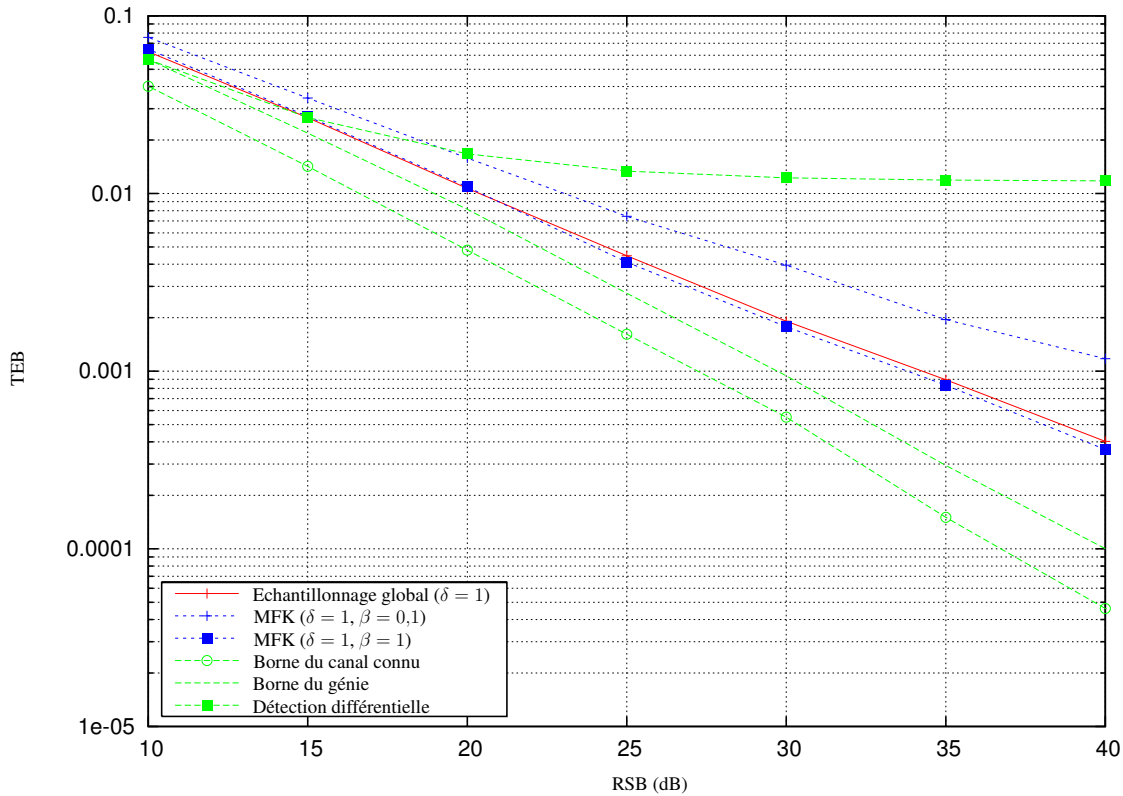


FIG. 7.5 – Courbe de taux d’erreur binaire du récepteur à échantillonnage global en fonction du rapport signal sur bruit pour un retard d’estimation  $\delta = 1$  et un nombre de particules  $N = 5$ . On a également représenté sur cette figure la courbe de taux d’erreur binaire du récepteur MFK avec  $\delta = 1$ ,  $N = 5$  et successivement  $\beta = 0,1$  et  $\beta = 1$  ; la courbe de taux d’erreur binaire du récepteur à canal connu ; la courbe de taux d’erreur binaire du récepteur aidé par un génie et la courbe de taux d’erreur binaire du récepteur à détection différentielle.

## 7.6 Conclusion

Dans ce chapitre, nous avons appliqué l’algorithme de filtrage particulière à échantillonnage global au problème de l’égalisation. Comme au chapitre 4, nous nous sommes intéressés aux trois situations suivantes : égalisation d’un canal sélectif en fréquence de réponse impulsionnelle connue, égalisation adaptative d’un canal évanescent sélectif en fréquence et égalisation adaptative d’un canal évanescent non-sélectif en fréquence. Grâce à la notion de modèle dynamique à processus indicateur latent introduite au chapitre 6, nous avons pu traiter ces différentes situations à l’aide d’un formalisme commun. En effet, pour appliquer les filtres particulières que nous avons présentés au chapitre 6, il nous a suffi de remarquer que le réseau bayésien correspondant à chacune de ces situations possède une structure de modèle dynamique à processus indicateur latent, et de calculer ensuite explicitement les opérateurs de prédiction et de correction correspondant. Nous avons ainsi obtenu des algorithmes dont la complexité est linéaire par rapport au temps. De plus, dans le cas de l’égalisation d’un canal sélectif en fréquence de réponse impulsionnelle connue, cette complexité est également linéaire par rapport à la longueur de la réponse impulsionnelle du canal. Enfin, nous avons étudié numériquement les performances de

l'algorithme de filtrage particulaire à échantillonnage global appliqué à l'égalisation adaptative d'un canal évanescent non-sélectif en fréquence. Même avec un faible nombre de particules, nous obtenons des performances proches de l'optimal. En outre, pour la situation considérée et en fonction du nombre de particules utilisées, nous constatons que les performances de l'algorithme proposé sont supérieures ou égales à celles de l'algorithme décrit dans [CWL00] avec les réglages préconisés dans cet article.

L'algorithme de filtrage particulaire à échantillonnage global permet donc d'obtenir un excellent compromis entre les performances et la complexité lorsqu'on l'applique à l'égalisation. Dans le prochain chapitre, nous verrons qu'il est également possible de combiner cette technique d'approximation avec l'algorithme de propagation de croyance pour effectuer conjointement l'égalisation adaptative et le décodage.



# Chapitre 8

## L'égalisation adaptative et le décodage conjoints

### 8.1 Introduction

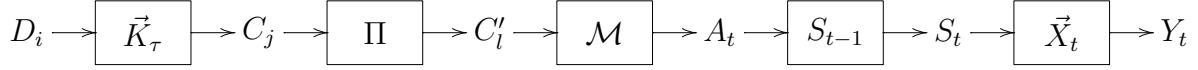
Dans le chapitre précédent, le filtrage particulière à échantillonnage global nous a permis d'obtenir des algorithmes d'égalisation réalisant un excellent compromis entre les performances et la complexité. Cependant, ces algorithmes ne prennent pas en compte le codage de canal et l'entrelacement utilisés par l'émetteur. Aussi, dans ce chapitre, nous allons mettre en œuvre l'ensemble des techniques étudiées précédemment afin de concevoir un algorithme d'égalisation adaptative et de décodage conjoints qui permettra au récepteur de prendre également en compte le code de canal et l'entrelacement utilisés par l'émetteur.

A titre d'exemple, nous considérerons dans ce chapitre le cas d'un canal évanescent non-sélectif en fréquence. On notera cependant que l'algorithme présenté dans ce chapitre est aussi applicable aux autres canaux étudiés précédemment. Après avoir décrit le modèle de système étudié, nous commencerons par construire le réseau bayésien correspondant et nous lui appliquerons l'algorithme de propagation de croyance. En utilisant les résultats du chapitre 4, nous constaterons alors que le calcul de certains des messages utilisés par l'algorithme de propagation de croyance exigerait un nombre d'opérations augmentant exponentiellement avec le temps. Aussi, nous remplacerons le calcul exact de ces messages par un calcul approché effectué en appliquant l'algorithme de filtrage particulière étudié précédemment. Nous obtiendrons ainsi un algorithme itératif, combinant filtrage particulière et propagation de croyance, que nous appellerons *algorithme de turbo-filtrage particulière* et dont nous décrirons en détail l'application au système étudié.

### 8.2 Modèle du système

On suppose que l'on veut transmettre des bits de données  $D_i$ . Ces bits de données sont transformés en des bits codés  $C_j$  par un codeur convolutif de taux  $k_0/n_0$ , de longueur de contrainte  $k_0\nu_0$  avec  $\nu_0 > 1$ , et dont l'état interne est noté  $\vec{K}_\tau$ . Les bits codés  $C_j$  sont entrelacés par un entrelaceur  $\Pi$  en des bits entrelacés  $C'_l = C_{\Pi^{-1}(l)}$  qui sont ensuite transformés en des symboles  $A_t$  complexes, de module unitaire, à l'aide d'un modulateur par déplacement de phase à

$M$  états (MDP-M). Les symboles de modulation  $A_t$  sont codés différentiellement en des symboles  $S_t$  qui sont transmis à travers un canal non sélectif en fréquence dont on observe la sortie complexe  $Y_t$ .



On suppose que, pour tout  $i \in \mathbb{N}$ , les bits de donnée  $D_0, D_1, \dots, D_i$  sont indépendants – sans être nécessairement identiquement distribués<sup>1</sup>. On note  $(\vec{D}_\tau)_{\tau \in \mathbb{N}}$  la suite de vecteurs telle que

$$\vec{D}_\tau \stackrel{\text{déf}}{=} [D_{\tau,1} \ D_{\tau,2} \ \dots \ D_{\tau,k_0}]^T \quad (8.1)$$

$$(D_0, D_1, D_2, \dots) = (D_{0,1}, D_{0,2}, \dots, D_{0,k_0}, D_{1,1}, D_{1,2}, \dots, D_{1,k_0}, \dots) \quad (8.2)$$

Pour tout  $\tau \in \mathbb{N}$ , on définit l'état  $\vec{K}_\tau$  du codeur convolutif par

$$\vec{K}_\tau \stackrel{\text{déf}}{=} \begin{cases} \left[ \vec{D}_\tau^T \ \vec{D}_{\tau-1}^T \ \dots \ \vec{D}_{\tau-\nu_0+1}^T \right]^T & \text{si } \tau \geq \nu_0 - 1 \\ \left[ \vec{D}_\tau^T \ \vec{D}_{\tau-1}^T \ \dots \ \vec{D}_0^T \ 0 \dots 0 \right]^T & \text{sinon} \end{cases} \quad (8.3)$$

On note  $(\vec{C}_\tau)_{\tau \in \mathbb{N}}$  la suite de vecteurs telle que

$$\vec{C}_\tau \stackrel{\text{déf}}{=} [C_{\tau,1} \ C_{\tau,2} \ \dots \ C_{\tau,n_0}]^T \quad (8.4)$$

$$(C_0, C_1, C_2, \dots) = (C_{0,1}, C_{0,2}, \dots, C_{0,n_0}, C_{1,1}, C_{1,2}, \dots, C_{1,n_0}, \dots) \quad (8.5)$$

On suppose qu'il existe une fonction vectorielle  $\vec{\psi} = (\psi_1, \dots, \psi_{n_0})$  telle que

$$\vec{C}_\tau = \vec{\psi}(\vec{K}_\tau) \quad (8.6)$$

On suppose que l'entrelaceur  $\Pi$  est un entrelaceur en blocs de longueur  $p$ . On suppose que le modulateur est caractérisé par une bijection  $\mathcal{M}$  qui à  $q_0$  bits associe un symbole appartenant à un alphabet  $\mathcal{A} \subset \mathbb{C}$ , stable pour la multiplication et de cardinal  $M = 2^{q_0}$ . Autrement dit, on a  $\mathcal{M} : \{0, 1\}^{q_0} \rightarrow \mathcal{A}$ ,  $\vec{b} = (b_1, \dots, b_{q_0}) \mapsto a = \mathcal{M}(\vec{b})$ . On suppose également que la longueur  $p$  de l'entrelaceur est un multiple de  $q_0$  et de  $n_0$ . On note  $(\vec{C}'_t)_{t \in \mathbb{N}}$  la suite de vecteurs telle que

$$\vec{C}'_t \stackrel{\text{déf}}{=} [C'_{t,1} \ C'_{t,2} \ \dots \ C'_{t,q_0}]^T \quad (8.7)$$

$$(C'_0, C'_1, C'_2, \dots) = (C'_{0,1}, C'_{0,2}, \dots, C'_{0,q_0}, C'_{1,1}, C'_{1,2}, \dots, C'_{1,q_0}, \dots) \quad (8.8)$$

Les symboles  $S_t$  sont reliés aux symboles  $A_t$  par l'équation suivante

$$S_t = \begin{cases} A_t S_{t-1} & \text{si } t > 0 \\ A_0 & \text{si } t = 0 \end{cases} \quad (8.9)$$

La sortie  $Y_t$  du canal vérifie, pour tout  $t \in \mathbb{N}$ ,

$$Y_t = H_t S_t + V_t \quad (8.10)$$

<sup>1</sup>On pourra ainsi prendre une mesure de Dirac comme a loi *a priori* d'un bit appartenant à une séquence de référence et une loi uniforme pour les autres bits.

où  $V_t$  suit la loi gaussienne  $\mathcal{N}_C(0, \mathcal{N}_0)$  et le coefficient  $H_t$  est défini par

$$H_t = \vec{\Theta}^T \vec{X}_t \quad (8.11)$$

$$\vec{\Theta} = [\theta_0 \ \theta_1 \ \dots \ \theta_L]^T \quad (8.12)$$

$$\vec{X}_t = \begin{cases} \Lambda \vec{X}_{t-1} + \vec{W}_t & \text{si } t > 0 \\ \vec{\eta}_0 + \vec{W}_0 & \text{si } t = 0 \end{cases} \quad (8.13)$$

où  $\vec{W}_t$  suit la loi gaussienne  $\mathcal{N}_C(\vec{0}, R_G)$ , les variables aléatoires  $V_0, \vec{W}_0, V_1, \vec{W}_1, \dots, V_t, \vec{W}_t$  sont indépendantes,  $\vec{\eta}_0$  et  $\vec{\Theta}$  sont des vecteurs complexes supposés connus et  $\Lambda$  une matrice complexe également supposée connue et de la forme suivante

$$\Lambda = \begin{bmatrix} -\phi_1 & -\phi_2 & \dots & -\phi_L & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \quad (8.14)$$

On suppose également que la matrice de covariance  $R_G$  du bruit  $\vec{W}_t$  est de la forme suivante

$$R_G = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \quad (8.15)$$

Enfin, on suppose que les processus  $(D_i)_{i \in \mathbb{N}}$ ,  $(V_t)_{t \in \mathbb{N}}$  et  $(\vec{W}_t)_{t \in \mathbb{N}}$  sont indépendants.

### 8.3 Réseau bayésien

Le réseau bayésien correspondant aux processus  $(D_i)_{i \in \mathbb{N}}$ ,  $(\vec{K}_\tau)_{\tau \in \mathbb{N}}$ ,  $(C_j)_{j \in \mathbb{N}}$ ,  $(C'_l)_{l \in \mathbb{N}}$ ,  $(A_t)_{t \in \mathbb{N}}$ ,  $(S_t)_{t \in \mathbb{N}}$ ,  $(\vec{X}_t)_{t \in \mathbb{N}}$  et  $(Y_t)_{t \in \mathbb{N}}$  a été représenté à la figure 8.1 pour  $t \in \{0, 1, 2\}$ , dans le cas particulier d'un code de convolution de taux  $2/3$ , d'un entrelaceur en bloc de taille  $p = 6$  et d'un modulateur à  $M = 2^2 = 4$  états.



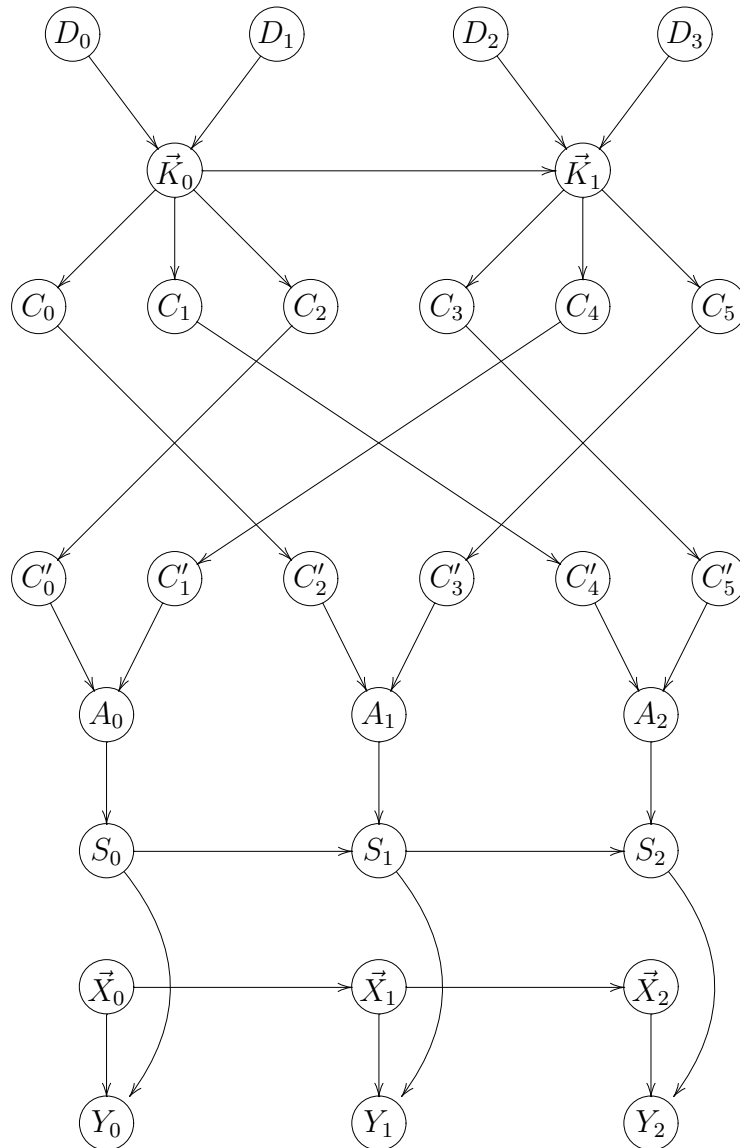


FIG. 8.1 – Réseau bayésien correspondant à un code de convolution de taux  $2/3$ , un entrelaceur en bloc de taille  $p = 6$ , un modulateur à  $M = 2^2 = 4$  états, un codeur différentiel et un canal évanescent non-sélectif en fréquence.

On peut regrouper les variables  $S_t$  et  $\vec{X}_t$ , qui correspondent respectivement à l'état du codeur différentiel et à l'état du canal, en introduisant la variable d'état conjointe

$$\vec{B}_t \stackrel{\text{déf}}{=} (S_t, \vec{X}_t) \quad (8.16)$$

On obtient alors le réseau bayésien de la figure 8.2.

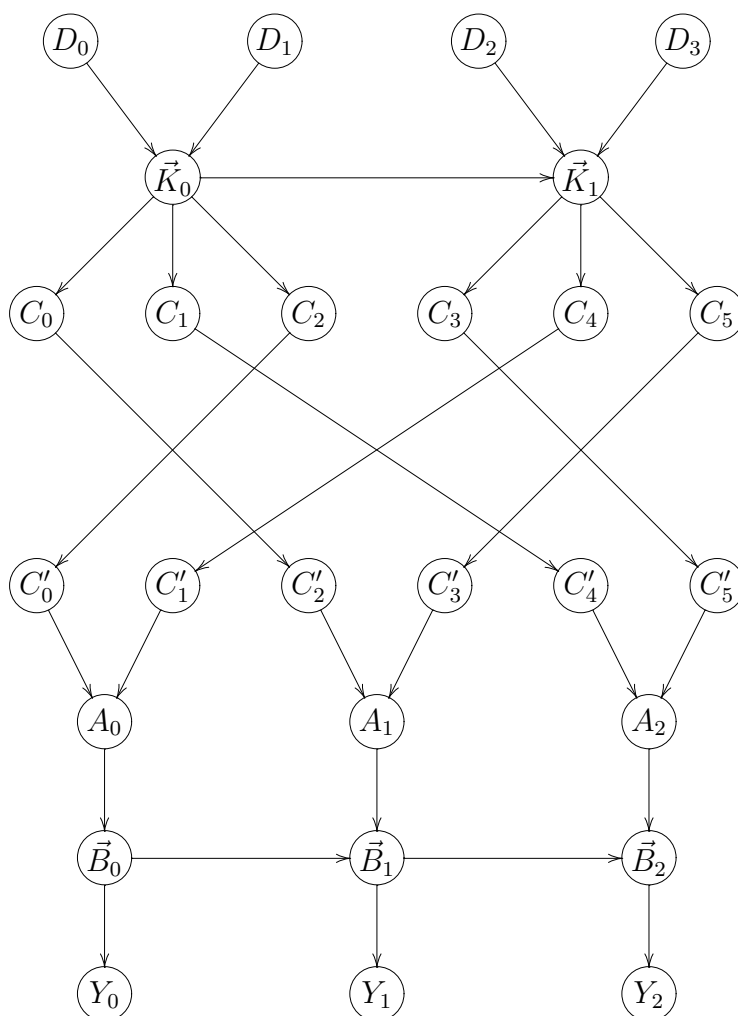


FIG. 8.2 – Réseau bayésien correspondant à un code de convolution de taux 2/3, un entrelaceur en bloc de taille  $p = 6$ , un modulateur à  $M = 2^2 = 4$  états, un codeur différentiel et un canal évanescent non-sélectif en fréquence. Les variables d'état du codeur différentiel et du canal ont été regroupées.

## 8.4 Application de l'algorithme de propagation de croyance approchée

### 8.4.1 Séparation des données en blocs

Afin d'appliquer l'algorithme de propagation de croyance approchée, les données sont séparées en blocs successifs qui correspondent chacun à  $m$  bits de données,  $m$  étant un multiple de  $p \frac{k_0}{n_0}$  i.e. le produit de la taille de l'entrelaceur et du taux du code de convolution<sup>2</sup>. On peut

<sup>2</sup>En pratique, la taille des blocs de données peut être limitée par la latence maximale d'estimation permise par l'application et la quantité maximale de mémoire disponible.

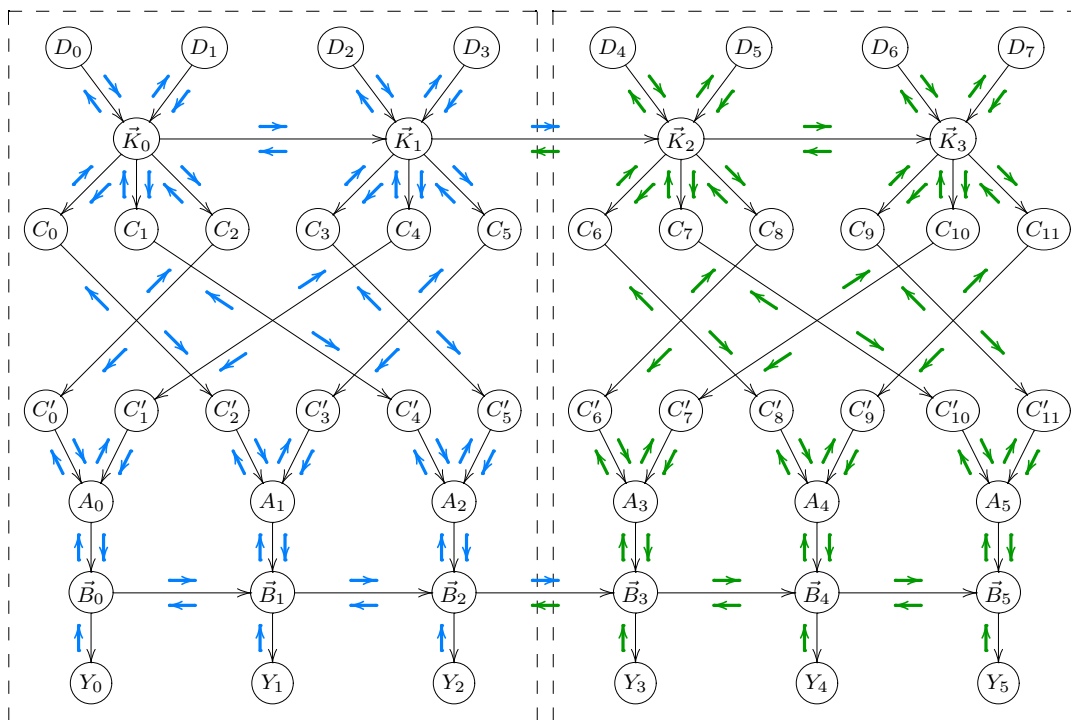


FIG. 8.3 – Réseau bayésien correspondant à un code de convolution de taux  $2/3$ , un entrelaceur en bloc de taille  $p = 6$ , un modulateur à  $M = 2^2 = 4$  états, un codeur différentiel et un canal évanescent non-sélectif en fréquence. Les variables d'état du codeur différentiel et du canal ont été regroupées. Les variables correspondant à deux blocs de bits de données de taille  $m = 4$  ont été représentées. Les messages obtenus à l'issue des itérations sur le premier bloc de données ont été représentés en bleu. Les autres messages, qui correspondent aux itérations sur le deuxième bloc de données, ont été représentés en vert.

alors appliquer l'algorithme de propagation de croyance en calculant de façon itérative les messages correspondant à un bloc de données, puis en propageant les messages issus de ce bloc de données vers le bloc suivant (cf Figure 8.3). On notera que la propagation des messages issus d'un bloc de données vers les blocs précédents est possible, mais n'est nécessaire que si l'on veut prendre compte les observations correspondant à ce bloc dans l'estimation des bits de données correspondant aux blocs précédents. Dans la suite, nous supposons que ce n'est pas le cas et nous ne propagerons donc pas les messages issus d'un bloc de données vers les blocs précédents.

#### 8.4.2 Calcul itératif des messages correspondant à un bloc de données

L'ordre dans lequel les messages correspondant à un bloc de données sont mis à jour à chaque itération est indiqué à la figure 8.4. A chaque itération sur un bloc de données, on calcule

1. le message envoyé par  $\vec{B}_t$  à  $\vec{B}_{t+1}$  pour  $t$  croissant
2. le message envoyé par  $\vec{B}_t$  à  $\vec{B}_{t-1}$  pour  $t$  décroissant

3. le message envoyé par  $\vec{B}_t$  à  $A_t$  pour  $t$  croissant
4. le message envoyé par  $A_t$  à  $C'_{t,q}$  pour  $t$  croissant et  $q$  croissant
5. le message envoyé par  $C'_l$  à  $C_{\Pi^{-1}(l)}$  pour  $l$  croissant
6. le message envoyé par  $C_{\tau,n}$  à  $\vec{K}_\tau$  pour  $\tau$  croissant et  $n$  croissant
7. le message envoyé par  $\vec{K}_\tau$  à  $\vec{K}_{\tau+1}$  pour  $\tau$  croissant
8. le message envoyé par  $\vec{K}_\tau$  à  $\vec{K}_{\tau-1}$  pour  $\tau$  décroissant
9. le message envoyé par  $\vec{K}_\tau$  à  $D_{\tau,k}$  pour  $\tau$  croissant et  $k$  croissant
10. le message envoyé par  $\vec{K}_\tau$  à  $C_{\tau,n}$  pour  $\tau$  croissant et  $n$  croissant
11. le message envoyé par  $C_j$  à  $C'_{\Pi(j)}$  pour  $j$  croissant
12. le message envoyé par  $C'_{t,q}$  à  $A_t$  pour  $t$  croissant et  $q$  croissant
13. le message envoyé par  $A_t$  à  $\vec{B}_t$  pour  $t$  croissant.

### Remarque 8.1

Comme nous l'avons indiqué au § 3.4, p. 45, l'ordre dans lequel les messages sont calculés est arbitraire. L'ordre que nous avons choisi ici est similaire à celui utilisé dans l'algorithme de turbo-décodage. Cependant, d'autres choix étaient *a priori* possibles [MV98b, MV98a, KF98].

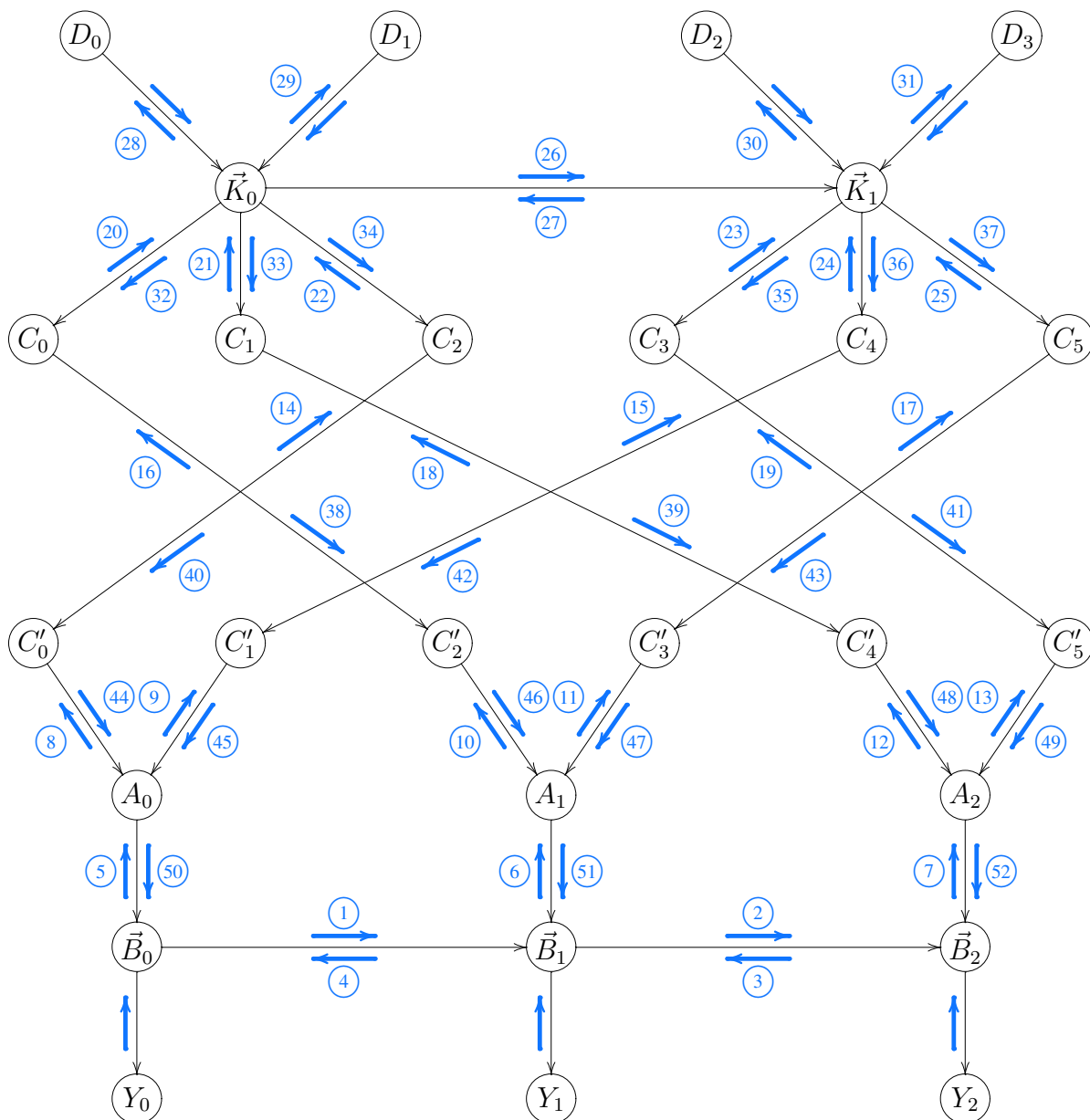


FIG. 8.4 – Calcul itératif des messages correspondant à un bloc de données.

### 8.4.3 Limitation de l'algorithme de propagation de croyance approchée

Comme nous l'avons vu au chapitre 4, le calcul du message envoyé par une variable  $\vec{B}_t$  à la variable d'état suivante  $\vec{B}_{t+1}$  exige un nombre d'opérations augmentant exponentiellement avec  $t$ . En pratique, il n'est donc pas possible d'appliquer directement l'algorithme que nous venons de décrire. Des approximations numériques supplémentaires sont indispensables. Nous allons voir dans la partie suivante que ces approximations peuvent être effectuées à l'aide des techniques de filtrage particulière que nous avons présentées au chapitre 5.

## 8.5 Principe de l'algorithme de turbo-filtrage particulière

L'idée est d'appliquer l'algorithme de propagation de croyance approchée en remplaçant le calcul exact d'une partie des messages échangés par un calcul approché effectué par filtrage particulière.

L'algorithme que nous proposons agit sur des blocs de données correspondant à  $m$  bits de données,  $m$  étant un multiple de  $p \frac{k_0}{n_0}$ , qui est le produit de la taille de l'entrelaceur et du taux du code de convolution. Dans la suite, nous nous intéresserons donc à l'estimation des bits de données dont l'indice est compris entre  $i_1$  et  $i_2$ , avec  $i_2 = i_1 + m - 1$  et  $i_1 \equiv 0 \pmod{m}$ .

Soit  $G$  le réseau bayésien correspondant aux variables aléatoires  $(D_i)_{0 \leq i \leq i_2}$ ,  $(\vec{K}_\tau)_{0 \leq \tau \leq \tau_2}$ ,  $(C_j)_{0 \leq j \leq j_2}$ ,  $(C'_l)_{0 \leq l \leq l_2}$ ,  $(A_t)_{0 \leq t \leq t_2}$ ,  $(\vec{B}_t)_{0 \leq t \leq t_2}$  et  $(Y_t)_{0 \leq t \leq t_2}$  avec

$$\tau_1 = \frac{i_1}{k_0} \qquad \tau_2 = \tau_1 + \frac{m}{k_0} - 1 \qquad (8.17)$$

$$j_1 = \frac{i_1 n_0}{k_0} \qquad j_2 = j_1 + \frac{m n_0}{k_0} - 1 \qquad (8.18)$$

$$l_1 = j_1 \qquad l_2 = j_2 \qquad (8.19)$$

$$t_1 = \frac{i_1 n_0}{q_0 k_0} \qquad t_2 = t_1 + \frac{m n_0}{q_0 k_0} - 1 \qquad (8.20)$$

Soit  $G'$  le sous-graphe de  $G$  induit par les sommets  $(A_t)_{0 \leq t \leq t_2}$ ,  $(\vec{B}_t)_{0 \leq t \leq t_2}$  et  $(Y_t)_{0 \leq t \leq t_2}$ . Considérons le calcul des messages  $(\mu_{\vec{B}_t A_t}^-)_{t_1 \leq t \leq t_2}$  envoyés par les  $\vec{B}_t$  aux  $A_t$ , à partir des messages  $(\mu_{A_t \vec{B}_t}^+)_{0 \leq t \leq t_2}$  envoyés par les  $A_t$  aux  $\vec{B}_t$ . On remarque alors que le calcul de ces messages est identique à celui que l'on ferait en appliquant l'algorithme de propagation de croyance exacte à un réseau bayésien virtuel  $\tilde{G}$  tel que

1.  $\tilde{G}$  a pour sommets des variables aléatoires notées  $(\tilde{A}_t)_{0 \leq t \leq t_2}$ ,  $(\tilde{\vec{B}}_t)_{0 \leq t \leq t_2}$  et  $(\tilde{Y}_t)_{0 \leq t \leq t_2}$
2. les arcs de  $\tilde{G}$  sont identiques<sup>3</sup> à ceux de  $G'$ .
3. les transitions entre les parents et les enfants de  $\tilde{G}$  sont identiques aux transitions entre les sommets correspondants de  $G'$ .
4. les variables aléatoires  $\tilde{A}_t$  avec  $0 \leq t \leq t_2$  sont indépendantes et ont pour loi *a priori*  $F_{\tilde{A}_t} = \mu_{A_t \vec{B}_t}^+(y_{0:t_2}; \cdot)$ .

A titre d'exemple, nous avons représenté à la figure 8.5 le réseau bayésien virtuel  $\tilde{G}$  associé au réseau bayésien  $G$  de la figure 8.3.

<sup>3</sup>Autrement dit, pour tout couple de sommets  $(X, Y)$  de  $G'$ ,  $(\tilde{X}, \tilde{Y})$  est un arc de  $\tilde{G}$  si et seulement si  $(X, Y)$  est un arc de  $G'$ .

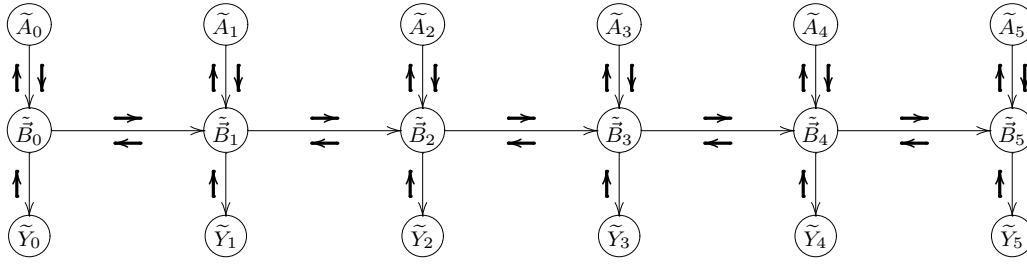


FIG. 8.5 – Réseau bayésien virtuel  $\tilde{G}$  associé au réseau bayésien  $G$  de la figure 8.3. Le calcul des messages  $(\mu_{\tilde{B}_t \tilde{A}_t}^-)_{t_1 \leq t \leq t_2}$  à partir des messages  $(\mu_{\tilde{A}_t \tilde{B}_t}^+)_{0 \leq t \leq t_2}$  dans le réseau  $G$  est identique à celui que l'on ferait en appliquant l'algorithme de propagation de croyance exacte au réseau bayésien virtuel  $\tilde{G}$ .

Par conséquent, pour tout  $y_{0:t_2}$ , le message  $\mu_{\tilde{B}_t \tilde{A}_t}^-(\cdot; y_{0:t_2})$  peut être considéré comme la vraisemblance  $f_{\tilde{Y}_{0:t_2}}^{\tilde{A}_t}(\cdot; y_{0:t_2})$  de  $\tilde{A}_t$  lorsque  $\tilde{Y}_{0:t_2} = y_{0:t_2}$ .

Il est donc possible, en principe, d'obtenir une approximation  $\hat{\mu}_{\tilde{A}_t \tilde{B}_t}^-(\cdot; y_{0:t_2})$  du message  $\mu_{\tilde{B}_t \tilde{A}_t}^-(\cdot; y_{0:t_2})$  en appliquant un filtre particulière au modèle dynamique à processus indicateur latent formé par les processus  $(\tilde{A}_t)_{0 \leq t \leq t_2}$ ,  $(\tilde{B}_t)_{0 \leq t \leq t_2}$  et  $(\tilde{Y}_t)_{0 \leq t \leq t_2}$ . En effet, un filtre particulière permet d'obtenir une approximation  $\hat{\Pi}_{\tilde{A}_{0:t_2}}^{\tilde{Y}_{0:t_2}}(y_{0:t_2}; \cdot)$  de la loi de  $\tilde{A}_{0:t_2}$  conditionnellement à  $\tilde{Y}_{0:t_2} = y_{0:t_2}$ . Pour tout  $t \in \{t_1, \dots, t_2\}$ , on en déduit une approximation  $\hat{\Pi}_{\tilde{A}_t}^{\tilde{Y}_{0:t_2}}(y_{0:t_2}; \cdot)$  de la loi de  $\tilde{A}_t$  conditionnellement à  $\tilde{Y}_{0:t_2} = y_{0:t_2}$

$$\hat{\Pi}_{\tilde{A}_t}^{\tilde{Y}_{0:t_2}}(y_{0:t_2}; \cdot) = \sum_{(\tilde{A}_{t'})_{t' \neq t}} \hat{\Pi}_{\tilde{A}_{0:t_2}}^{\tilde{Y}_{0:t_2}}(y_{0:t_2}; \cdot) \quad (8.21)$$

On en déduit

$$\hat{f}_{\tilde{Y}_{0:t_2}}^{\tilde{A}_t}(\cdot; y_{0:t_2}) \propto \frac{d\hat{\Pi}_{\tilde{A}_t}^{\tilde{Y}_{0:t_2}}(y_{0:t_2}; \cdot)}{dF_{\tilde{A}_t}} \quad (8.22)$$

$$\propto \frac{d\hat{\Pi}_{\tilde{A}_t}^{\tilde{Y}_{0:t_2}}(y_{0:t_2}; \cdot)}{d\mu_{\tilde{A}_t \tilde{B}_t}^+(y_{0:t_2}; \cdot)} \quad (8.23)$$

où, pour toute mesure signée  $\mu$  et pour toute mesure positive finie  $\nu$  telles que  $\mu \ll \nu$ , la notation  $\frac{d\mu}{d\nu}$  désigne la dérivée de Radon-Nikodym de  $\mu$  par rapport à  $\nu$ . On peut donc poser

$$\hat{\mu}_{\tilde{A}_t \tilde{B}_t}^-(\cdot, y_{0:t_2}) = \frac{d\hat{\Pi}_{\tilde{A}_t}^{\tilde{Y}_{0:t_2}}(y_{0:t_2}; \cdot)}{d\mu_{\tilde{A}_t \tilde{B}_t}^+(y_{0:t_2}; \cdot)} \quad (8.24)$$

Cependant, en pratique, pour des blocs de grande taille, l'approximation de  $\Pi_{\tilde{A}_t}^{\tilde{Y}_{0:t_2}}$  par  $\hat{\Pi}_{\tilde{A}_t}^{\tilde{Y}_{0:t_2}}$  peut être très mauvaise en raison de l'appauvrissement du filtre particulière lorsque l'instant  $t$

de l'estimation est trop éloigné de l'instant  $t_2$  de l'échantillonnage. Une meilleure estimation de  $\Pi_{\tilde{A}_t}^{\tilde{Y}_{0:t_2}}$  peut alors être  $\hat{\Pi}_{\tilde{A}_t}^{\tilde{Y}_{0:\min(t+\delta, t_2)}}$  où  $\delta$  est un retard fixe.

## 8.6 Application de l'algorithme de turbo-filtrage particulière

### 8.6.1 Traitement d'un bloc de données

Soit  $i_1 > 0$  tel que  $i_1 \equiv 0 \pmod{m}$ . On pose  $i_2 = i_1 + m - 1$  et on définit  $\tau_1, \tau_2, j_1, j_2, l_1, l_2, t_1$  et  $t_2$  à l'aide des équations 8.17, 8.18, 8.19 et 8.20. On note  $G$  le réseau bayésien correspondant aux variables aléatoires  $(D_i)_{0 \leq i \leq i_2}, (\vec{K}_\tau)_{0 \leq \tau \leq \tau_2}, (C_j)_{0 \leq j \leq j_2}, (C'_l)_{0 \leq l \leq l_2}, (A_t)_{0 \leq t \leq t_2}, (\vec{B}_t)_{0 \leq t \leq t_2}$  et  $(Y_t)_{0 \leq t \leq t_2}$  et  $G'$  le sous-graphe de  $G$  induit par les sommets  $(A_t)_{0 \leq t \leq t_2}, (\vec{B}_t)_{0 \leq t \leq t_2}$  et  $(Y_t)_{0 \leq t \leq t_2}$ .

On suppose que l'on a calculé les  $N$  trajectoires  $(A^{(j, t_1-1)})_{1 \leq j \leq N}$  du filtre particulière à l'instant  $t_1 - 1$  ainsi que le terme

$$\alpha_{\tau_1-1}(\vec{d}_{\tau_1-1}, \vec{d}_{\tau_1-2}, \dots, \vec{d}_{\tau_1-\nu_0+1}) \stackrel{\text{déf}}{=} \sum_{\vec{d}_{\tau_1-\nu_0}} \mu_{\vec{K}_{\tau_1-1} \vec{K}_{\tau_1}}^+ (y_{0:t_1-1}; \left\{ \left[ \vec{d}_{\tau_1-1}^T \quad \vec{d}_{\tau_1-2}^T \quad \dots \quad \vec{d}_{\tau_1-\nu_0}^T \right]^T \right\}) \quad (8.25)$$

correspondant au message  $\mu_{\vec{K}_{\tau_1-1} \vec{K}_{\tau_1}}^+$  envoyé par  $\vec{K}_{\tau_1-1}$  à  $\vec{K}_{\tau_1}$ . Pour tout  $t \in \{t_1, t_1 + 1, \dots, t_2\}$ , on pose

$$\mu_{A_t \vec{B}_t}^{+(0)} = \mathcal{U}_{\mathcal{A}} \quad (8.26)$$

où  $\mathcal{U}_{\mathcal{A}}$  désigne la loi uniforme sur l'alphabet  $\mathcal{A}$ . On note également  $\tilde{G}^{(0)}$  le réseau bayésien tel que

1.  $\tilde{G}^{(0)}$  a pour sommets des variables aléatoires notées  $(\tilde{A}_t^{(0)})_{0 \leq t \leq t_2}, (\tilde{B}_t^{(0)})_{0 \leq t \leq t_2}$  et  $(\tilde{Y}_t^{(0)})_{0 \leq t \leq t_2}$
2. les arcs de  $\tilde{G}^{(0)}$  sont identiques à ceux du sous-graphe  $G'$ .
3. les transitions entre les parents et les enfants de  $\tilde{G}^{(0)}$  sont identiques aux transitions entre les sommets correspondants de  $G'$ .
4. les variables aléatoires  $\tilde{A}_t^{(0)}$  avec  $0 \leq t \leq t_2$  sont indépendantes et ont pour loi *a priori*

$$F_{\tilde{A}_t^{(0)}} = \begin{cases} \mu_{A_t \vec{B}_t}^+ & \text{si } 0 \leq t \leq t_1 - 1 \\ \mu_{A_t \vec{B}_t}^{+(0)} = \mathcal{U}_{\mathcal{A}} & \text{si } t_1 \leq t \leq t_2 \end{cases} \quad (8.27)$$

où, pour tout  $t \in \{0, 1, \dots, t_1 - 1\}$ , la notation  $\mu_{A_t \vec{B}_t}^+$  désigne le message envoyé par  $A_t$  à  $\vec{B}_t$  lors de la dernière itération de l'algorithme sur le bloc de données correspondant à  $A_t$  et  $\vec{B}_t$ .



Pour tout  $t' \in \{t_1, \dots, t_2\}$ , on calcule une approximation  $\widehat{\Pi}_{\widetilde{A}_{0:t'}}^{\widetilde{Y}_{0:t'}}(y_{0:t'}; \cdot)$  de la loi de  $\widetilde{A}_{0:t'}^{(0)}$  conditionnellement à  $\widetilde{Y}_{0:t'}^{(0)} = y_{0:t'}$  en appliquant un filtre particulière au modèle dynamique à processus indicateur latent dont le réseau bayésien est  $\widetilde{G}^{(0)}$ .<sup>4</sup> On en déduit pour tout  $t \in \{t_1, \dots, t_2\}$  une approximation  $\widehat{\Pi}_{\widetilde{A}_t}^{\widetilde{Y}_{0:t}}(y_{0:t}; \cdot)$  de la loi de  $\widetilde{A}_t^{(0)}$  conditionnellement à  $\widetilde{Y}_{0:t}^{(0)} = y_{0:t}$

$$\widehat{\Pi}_{\widetilde{A}_t}^{\widetilde{Y}_{0:t}}(y_{0:t}; \cdot) = \sum_{(\widetilde{A}_{t'}^{(0)})_{t'' \neq t}} \widehat{\Pi}_{\widetilde{A}_{0:t'}}^{\widetilde{Y}_{0:t'}}(y_{0:t'}; \cdot) \quad (8.28)$$

avec  $t' = \min(t + \delta, t_2)$  et on pose

$$\mu_{\widetilde{B}_t A_t}^{-(1)}(\cdot; y_{0:t_2}) = \frac{d\widehat{\Pi}_{\widetilde{A}_t}^{\widetilde{Y}_{0:t}}(y_{0:t}; \cdot)}{d\mu_{A_t \widetilde{B}_t}^{+(0)}(y_{0:t_2}; \cdot)} \quad (8.29)$$

Pour tout  $q \in \{1, 2, \dots, q_0\}$ , on calcule le message  $\mu_{A_t C'_{t,q}}^{-(1)}$  envoyé par  $A_t$  à  $C'_{t,q}$

$$\mu_{A_t C'_{t,q}}^{-(1)}(\cdot; y_{0:t_2}) = \sum_{A_t} \mu_{\widetilde{B}_t A_t}^{-(1)}(\cdot; y_{0:t_2}) \sum_{(C'_{t,q'})_{q' \neq q}} \left[ \bigotimes_{q'' \neq q} \mu_{C'_{t,q''} A_t}^{+(0)}(y_{0:t_2}; \cdot) \right] \otimes \Pi_{A_t}^{(C'_{t,1}, \dots, C'_{t,q_0})} \quad (8.30)$$

avec

$$\Pi_{A_t}^{(C'_{t,1}, \dots, C'_{t,q_0})}(c'_{t,1}, \dots, c'_{t,q_0}; \cdot) = \delta_{\mathcal{M}(c'_{t,1}, \dots, c'_{t,q_0})} \quad (8.31)$$

On en déduit

$$\mu_{A_t C'_{t,q}}^{-(1)}(c'_{t,q}; y_{0:t_2}) = \sum_{(c'_{t,q'})_{q' \neq q}} \mu_{\widetilde{B}_t A_t}^{-(1)}(\mathcal{M}(c'_{t,1}, \dots, c'_{t,q_0}); y_{0:t_2}) \prod_{q'' \neq q} \mu_{C'_{t,q''} A_t}^{+(0)}(y_{0:t_2}; \{c'_{t,q''}\}) \quad (8.32)$$

Initialement, on peut prendre

$$\mu_{C'_{t,q''} A_t}^{+(0)}(y_{0:t_2}; \cdot) = \mathcal{U}_{\{0,1\}} \quad (8.33)$$

ce qui implique

$$\mu_{C'_{t,q''} A_t}^{+(0)}(y_{0:t_2}; \{c'_{t,q''}\}) = \frac{1}{2} \quad (8.34)$$

<sup>4</sup>On notera que les calculs nécessaires à l'application de ce filtre particulière ne concernent que le bloc de données considéré puisqu'il suffit de prolonger jusqu'à l'instant  $t' \in \{t_1, \dots, t_2\}$  les  $N$  trajectoires déjà obtenues à l'instant  $t_1 - 1$  pour obtenir une approximation de la loi de  $\widetilde{A}_{0:t'}^{(0)}$  conditionnellement à  $\widetilde{Y}_{0:t'}^{(0)} = y_{0:t'}$ .

Pour tout  $l \in \{l_1, l_1 + 1, \dots, l_2\}$ , on pose  $j = \Pi^{-1}(l)$  et on détermine le message envoyé par  $C'_l$  à  $C_j$  de la façon suivante

$$\mu_{C'_l C_j}^{-(1)}(\cdot; y_{0:t_2}) = \sum_{C'_l} \mu_{A_t C'_l}^{-(1)}(\cdot; y_{0:t_2}) \Pi_{C'_l}^{C_j} \quad (8.35)$$

avec

$$\Pi_{C'_l}^{C_j}(c_j; \cdot) = \delta_{c_j} \quad (8.36)$$

$$t = \left\lfloor \frac{l}{q_0} \right\rfloor \quad (8.37)$$

où, pour tout  $x \in \mathbb{R}$ ,  $\lfloor x \rfloor$  désigne la partie entière de  $x$ . On en déduit

$$\mu_{C'_l C_j}^{-(1)}(c_j; y_{0:t_2}) = \mu_{A_t C'_l}^{-(1)}(c_j; y_{0:t_2}) \quad (8.38)$$

De même, pour tout  $\tau \in \{\tau_1, \tau_1 + 1, \dots, \tau_2\}$  et pour tout  $n \in \{1, 2, \dots, n_0\}$ , on peut déterminer le message envoyé par  $C_{\tau,n}$  à  $\vec{K}_\tau$

$$\mu_{C_{\tau,n} \vec{K}_\tau}^{-(1)}(\cdot; y_{0:t_2}) = \sum_{C_{\tau,n}} \mu_{C'_{\Pi(j)} C_{\tau,n}}^{-(1)}(\cdot; y_{0:t_2}) \Pi_{C_{\tau,n}}^{\vec{K}_\tau} \quad (8.39)$$

avec

$$\Pi_{C_{\tau,n}}^{\vec{K}_\tau}(\vec{k}_\tau, \cdot) = \delta_{\psi_n(\vec{k}_\tau)} \quad (8.40)$$

$$j = \tau n_0 + n - 1 \quad (8.41)$$

On en déduit

$$\mu_{C_{\tau,n} \vec{K}_\tau}^{-(1)}(\vec{k}_\tau; y_{0:t_2}) = \mu_{C'_{\Pi(j)} C_{\tau,n}}^{-(1)}(\psi_n(\vec{k}_\tau); y_{0:t_2}) \quad (8.42)$$

Pour tout  $\tau \in \{\tau_1 + 1, \tau_1 + 2, \dots, \tau_2\}$ , on peut alors déterminer le message envoyé par  $\vec{K}_\tau$  à  $\vec{K}_{\tau-1}$

$$\mu_{\vec{K}_\tau \vec{K}_{\tau-1}}^{-(1)}(\cdot; y_{0:t_2}) = \sum_{\vec{K}_\tau} \mu_{\vec{K}_{\tau+1} \vec{K}_\tau}^{-(1)}(\cdot; y_{0:t_2}) \prod_n \mu_{C_{\tau,n} \vec{K}_\tau}^{-(1)}(\cdot; y_{0:t_2}) \sum_{\vec{D}_\tau} \left[ \bigotimes_k \mu_{D_{\tau,k} \vec{K}_\tau}^{+(0)} \right] \otimes \Pi_{\vec{K}_\tau}^{(\vec{K}_{\tau-1}, \vec{D}_\tau)} \quad (8.43)$$

avec la convention

$$\mu_{\vec{K}_{\tau_2+1} \vec{K}_{\tau_2}}^{-(1)} = 1, \quad (8.44)$$

avec

$$\mu_{D_{\tau,k}\vec{K}_{\tau}}^{+(0)} = F_{D_{\tau,k}} \quad (8.45)$$

et avec, pour tout  $\vec{k}_{\tau-1} = \left[ \vec{d}_{\tau-1}^T \quad \vec{d}_{\tau-2}^T \dots \quad \vec{d}_{\tau-\nu_0}^T \right]^T$  et pour tout  $\vec{d}_{\tau}$ ,

$$\Pi_{\vec{K}_{\tau}}^{(\vec{K}_{\tau-1}, \vec{D}_{\tau})}(\vec{k}_{\tau-1}, \vec{d}_{\tau}; \cdot) = \delta \left[ \vec{d}_{\tau}^T \quad \vec{d}_{\tau-1}^T \quad \dots \quad \vec{d}_{\tau-\nu_0+1}^T \right]^T \quad (8.46)$$

On en déduit

$$\begin{aligned} \mu_{\vec{K}_{\tau}\vec{K}_{\tau-1}}^{-(1)}(\vec{k}_{\tau-1}; y_{0:t_2}) &= \sum_{\vec{d}_{\tau}} \mu_{\vec{K}_{\tau+1}\vec{K}_{\tau}}^{-(1)} \left( \left[ \vec{d}_{\tau}^T \quad \vec{d}_{\tau-1}^T \quad \dots \quad \vec{d}_{\tau-\nu_0+1}^T \right]^T ; y_{0:t_2} \right) \times \dots \\ &\dots \times \left[ \prod_n \mu_{C_{\tau,n}\vec{K}_{\tau}}^{-(1)} \left( \left[ \vec{d}_{\tau}^T \quad \vec{d}_{\tau-1}^T \quad \dots \quad \vec{d}_{\tau-\nu_0+1}^T \right]^T ; y_{0:t_2} \right) \right] \left[ \prod_k F_{D_{\tau,k}}(\{d_{\tau,k}\}) \right] \end{aligned} \quad (8.47)$$

On remarque alors que l'expression 8.47 ne dépend pas de  $\vec{d}_{\tau-\nu_0}$ . On peut donc poser, pour tout  $\tau \in \{\tau_1, \tau_1 + 2, \dots, \tau_2\}$ ,

$$\beta_{\tau}^{(1)}(\vec{d}_{\tau}, \vec{d}_{\tau-1}, \dots, \vec{d}_{\tau-\nu_0+2}) \stackrel{\text{déf}}{=} \mu_{\vec{K}_{\tau+1}\vec{K}_{\tau}}^{-(1)} \left( \left[ \vec{d}_{\tau}^T \quad \vec{d}_{\tau-1}^T \quad \dots \quad \vec{d}_{\tau-\nu_0+1}^T \right]^T ; y_{0:t_2} \right) \quad (8.48)$$

La relation 8.47 s'écrit alors

$$\begin{aligned} \beta_{\tau-1}^{(1)}(\vec{d}_{\tau-1}, \vec{d}_{\tau-2}, \dots, \vec{d}_{\tau-\nu_0+1}) &= \sum_{\vec{d}_{\tau}} \beta_{\tau}^{(1)}(\vec{d}_{\tau}, \vec{d}_{\tau-1}, \dots, \vec{d}_{\tau-\nu_0+2}) \times \dots \\ &\dots \times \left[ \prod_n \mu_{C_{\tau,n}\vec{K}_{\tau}}^{-(1)} \left( \left[ \vec{d}_{\tau}^T \quad \vec{d}_{\tau-1}^T \quad \dots \quad \vec{d}_{\tau-\nu_0+1}^T \right]^T ; y_{0:t_2} \right) \right] \left[ \prod_k F_{D_{\tau,k}}(\{d_{\tau,k}\}) \right] \end{aligned} \quad (8.49)$$

avec  $\beta_{\tau_2}^{(1)} = 1$ .

Pour tout  $\tau \in \{\tau_1, \tau_1 + 1, \dots, \tau_2 - 1\}$ , on peut également déterminer le message envoyé par  $\vec{K}_{\tau}$  à  $\vec{K}_{\tau+1}$

$$\begin{aligned} \mu_{\vec{K}_{\tau}\vec{K}_{\tau+1}}^{+(1)}(y_{0:t_2}; \cdot) &= \Psi_{\text{Bayes}} \left( \prod_n \mu_{C_{\tau,n}\vec{K}_{\tau}}^{-(1)}(\cdot; y_{0:t_2}), \right. \\ &\left. \sum_{(\vec{K}_{\tau-1}, \vec{D}_{\tau})} \left[ \mu_{\vec{K}_{\tau-1}\vec{K}_{\tau}}^{+(1)}(y_{0:t_2}; \cdot) \otimes \left( \bigotimes_k \mu_{D_{\tau,k}\vec{K}_{\tau}}^{+(0)} \right) \right] \otimes \Pi_{\vec{K}_{\tau}}^{(\vec{K}_{\tau-1}, \vec{D}_{\tau})} \right) \end{aligned} \quad (8.50)$$

avec la convention  $\mu_{\vec{K}_{\tau_1-1}\vec{K}_{\tau_1}}^{+(1)} = \mu_{\vec{K}_{\tau_1-1}\vec{K}_{\tau_1}}^{+}$ . En utilisant les équations 8.45 et 8.46, on en déduit,

pour tout  $\vec{k}_\tau = \left[ \vec{d}_\tau^T \quad \vec{d}_{\tau-1}^T \cdots \vec{d}_{\tau-\nu_0+1}^T \right]^T$ ,

$$\begin{aligned} \mu_{\vec{K}_\tau \vec{K}_{\tau+1}}^{+(1)}(y_{0:t_2}; \{\vec{k}_\tau\}) &\propto \left[ \prod_n \mu_{C_{\tau,n} \vec{K}_\tau}^{-(1)}(\vec{k}_\tau; y_{0:t_2}) \right] \times \dots \\ &\dots \times \left[ \prod_k F_{D_{\tau,k}}(\{d_{\tau,k}\}) \right] \sum_{\vec{d}_{\tau-\nu_0}} \mu_{\vec{K}_{\tau-1} \vec{K}_\tau}^{+(1)}(y_{0:t_2}; \left\{ \left[ \vec{d}_{\tau-1}^T \quad \vec{d}_{\tau-2}^T \cdots \vec{d}_{\tau-\nu_0}^T \right]^T \right\}) \end{aligned} \quad (8.51)$$

En posant

$$\alpha_\tau^{(1)}(\vec{d}_\tau, \vec{d}_{\tau-1}, \dots, \vec{d}_{\tau-\nu_0+2}) \stackrel{\text{d\u00e9f}}{=} \sum_{\vec{d}_{\tau-\nu_0+1}} \mu_{\vec{K}_\tau \vec{K}_{\tau+1}}^{+(1)}(y_{0:t_2}; \left\{ \left[ \vec{d}_\tau^T \quad \vec{d}_{\tau-1}^T \cdots \vec{d}_{\tau-\nu_0+1}^T \right]^T \right\}) \quad (8.52)$$

la relation 8.51 implique

$$\begin{aligned} \alpha_\tau^{(1)}(\vec{d}_\tau, \vec{d}_{\tau-1}, \dots, \vec{d}_{\tau-\nu_0+2}) &\propto \dots \\ &\dots \left[ \prod_k F_{D_{\tau,k}}(\{d_{\tau,k}\}) \right] \sum_{\vec{d}_{\tau-\nu_0+1}} \left[ \prod_n \mu_{C_{\tau,n} \vec{K}_\tau}^{-(1)} \left( \left[ \vec{d}_\tau^T \quad \vec{d}_{\tau-1}^T \cdots \vec{d}_{\tau-\nu_0+1}^T \right]^T ; y_{0:t_2} \right) \right] \times \dots \\ &\dots \times \alpha_{\tau-1}^{(1)}(\vec{d}_{\tau-1}, \vec{d}_{\tau-2}, \dots, \vec{d}_{\tau-\nu_0+1}) \end{aligned} \quad (8.53)$$

Pour tout  $\tau \in \{\tau_1, \tau_1 + 1, \dots, \tau_2\}$  et pour tout  $k \in \{1, 2, \dots, k_0\}$ , on peut alors calculer le message envoy\u00e9 par  $\vec{K}_\tau$  \u00e0  $D_{\tau,k}$

$$\begin{aligned} \mu_{\vec{K}_\tau D_{\tau,k}}^{-(1)}(\cdot; y_{0:t_2}) &= \sum_{\vec{K}_\tau} \mu_{\vec{K}_{\tau+1} \vec{K}_\tau}^{-(1)}(\cdot; y_{0:t_2}) \prod_n \mu_{C_{\tau,n} \vec{K}_\tau}^{-(1)}(\cdot; y_{0:t_2}) \dots \\ &\dots \sum_{\vec{K}_{\tau-1}} \sum_{(D_{\tau,k'})_{k' \neq k}} \left[ \mu_{\vec{K}_{\tau-1} \vec{K}_\tau}^{+(1)}(y_{0:t_2}; \cdot) \otimes \left( \bigotimes_{k'' \neq k} \mu_{D_{\tau,k''} \vec{K}_\tau}^{+(0)} \right) \right] \otimes \prod_{\vec{K}_\tau}^{(\vec{K}_{\tau-1}, \vec{D}_\tau)} \end{aligned} \quad (8.54)$$

En utilisant la relation 8.46, on en d\u00e9duit, pour tout  $d_{\tau,k}$ ,

$$\begin{aligned} \mu_{\vec{K}_\tau D_{\tau,k}}^{-(1)}(d_{\tau,k}; y_{0:t_2}) &= \sum_{(d_{\tau,k'})_{k' \neq k}} \sum_{(\vec{d}_{\tau'})_{\tau-\nu_0 < \tau' < \tau}} \mu_{\vec{K}_{\tau+1} \vec{K}_\tau}^{-(1)} \left( \left[ \vec{d}_\tau^T \quad \vec{d}_{\tau-1}^T \cdots \vec{d}_{\tau-\nu_0+1}^T \right]^T ; y_{0:t_2} \right) \dots \\ &\dots \prod_n \mu_{C_{\tau,n} \vec{K}_\tau}^{-(1)} \left( \left[ \vec{d}_\tau^T \quad \vec{d}_{\tau-1}^T \cdots \vec{d}_{\tau-\nu_0+1}^T \right]^T ; y_{0:t_2} \right) \dots \\ &\dots \left[ \sum_{\vec{d}_{\tau-\nu_0}} \mu_{\vec{K}_{\tau-1} \vec{K}_\tau}^{+(1)}(y_{0:t_2}; \left\{ \left[ \vec{d}_{\tau-1}^T \quad \vec{d}_{\tau-2}^T \cdots \vec{d}_{\tau-\nu_0}^T \right]^T \right\}) \right] \left[ \prod_{k'' \neq k} F_{D_{\tau,k''}}(d_{\tau,k''}) \right] \end{aligned} \quad (8.55)$$

Pour tout  $\tau \in \{\tau_1, \tau_1 + 1, \dots, \tau_2\}$  et pour tout  $k \in \{1, 2, \dots, k_0\}$ , on obtient alors l'estimation suivante de la loi *a posteriori* de  $D_{\tau,k}$

$$\widehat{\prod_{D_{\tau,k}}^{Y_{0:t_2}}}_{(1)}(y_{0:t_2}; \cdot) = \Psi_{\text{Bayes}} \left( \mu_{\vec{K}_\tau D_{\tau,k}}^{-(1)}(\cdot; y_{0:t_2}), F_{D_{\tau,k}} \right) \quad (8.56)$$

En utilisant l'équation 8.55, on en déduit

$$\begin{aligned} \widehat{\Pi}_{D_{\tau,k}}^{Y_{0:t_2}}(1) (y_{0:t_2}; \{d_{\tau,k}\}) &\propto \sum_{(d_{\tau,k'})_{k' \neq k} (\vec{d}_{\tau'})_{\tau-\nu_0 < \tau' < \tau}} \sum \mu_{\vec{K}_{\tau+1}\vec{K}_{\tau}}^{- (1)} \left( \left[ \vec{d}_{\tau}^T \quad \vec{d}_{\tau-1}^T \cdots \quad \vec{d}_{\tau-\nu_0+1}^T \right]^T ; y_{0:t_2} \right) \cdots \\ &\cdots \prod_n \mu_{C_{\tau,n}\vec{K}_{\tau}}^{- (1)} \left( \left[ \vec{d}_{\tau}^T \quad \vec{d}_{\tau-1}^T \cdots \quad \vec{d}_{\tau-\nu_0+1}^T \right]^T ; y_{0:t_2} \right) \cdots \\ &\cdots \left[ \sum_{\vec{d}_{\tau-\nu_0}} \mu_{\vec{K}_{\tau-1}\vec{K}_{\tau}}^{+ (1)} \left( y_{0:t_2}; \left\{ \left[ \vec{d}_{\tau-1}^T \quad \vec{d}_{\tau-2}^T \cdots \quad \vec{d}_{\tau-\nu_0}^T \right]^T \right\} \right) \right] \left[ \prod_{k''} F_{D_{\tau,k''}}(d_{\tau,k''}) \right] \end{aligned} \quad (8.57)$$

Avec les notations 8.48 et 8.52, l'équation 8.57 peut s'écrire

$$\begin{aligned} \widehat{\Pi}_{D_{\tau,k}}^{Y_{0:t_2}}(1) (y_{0:t_2}; \{d_{\tau,k}\}) &\propto \sum_{(d_{\tau,k'})_{k' \neq k} (\vec{d}_{\tau'})_{\tau-\nu_0+2 \leq \tau' \leq \tau-1}} \sum \alpha_{\tau}^{(1)}(\vec{d}_{\tau}, \vec{d}_{\tau-1}, \dots, \vec{d}_{\tau-\nu_0+2}) \times \cdots \\ &\cdots \times \beta_{\tau}^{(1)}(\vec{d}_{\tau}, \vec{d}_{\tau-1}, \dots, \vec{d}_{\tau-\nu_0+2}) \end{aligned} \quad (8.58)$$

Pour tout  $\tau \in \{\tau_1, \tau_1+1, \dots, \tau_2\}$  et pour tout  $n \in \{1, 2, \dots, n_0\}$ , on peut également calculer le message envoyé par  $\vec{K}_{\tau}$  à  $C_{\tau,n}$

$$\begin{aligned} \mu_{\vec{K}_{\tau}C_{\tau,n}}^{+ (1)} (y_{0:t_2}; \cdot) &= \Psi_{\text{Bayes}} \left( \mu_{\vec{K}_{\tau+1}\vec{K}_{\tau}}^{- (1)} (\cdot; y_{0:t_2}) \prod_{n' \neq n} \mu_{C_{\tau,n'}\vec{K}_{\tau}}^{- (1)} (\cdot; y_{0:t_2}), \right. \\ &\left. \sum_{(\vec{K}_{\tau-1}, \vec{D}_{\tau})} \left[ \mu_{\vec{K}_{\tau-1}\vec{K}_{\tau}}^{+ (1)} (y_{0:t_2}; \cdot) \otimes \left( \bigotimes_k \mu_{D_{\tau,k}\vec{K}_{\tau}}^{+ (0)} \right) \otimes \Pi_{\vec{K}_{\tau}}^{(\vec{K}_{\tau-1}, \vec{D}_{\tau})} \right] \right) \end{aligned} \quad (8.59)$$

où l'on a également utilisé la convention 8.44. En utilisant les équations 8.45 et 8.46, on en déduit, pour tout  $\vec{k}_{\tau} = \left[ \vec{d}_{\tau}^T \quad \vec{d}_{\tau-1}^T \cdots \quad \vec{d}_{\tau-\nu_0+1}^T \right]^T$

$$\begin{aligned} \mu_{\vec{K}_{\tau}C_{\tau,n}}^{+ (1)} (y_{0:t_2}; \{\vec{k}_{\tau}\}) &\propto \mu_{\vec{K}_{\tau+1}\vec{K}_{\tau}}^{- (1)} (\vec{k}_{\tau}; y_{0:t_2}) \left[ \prod_{n' \neq n} \mu_{C_{\tau,n'}\vec{K}_{\tau}}^{- (1)} (\vec{k}_{\tau}; y_{0:t_2}) \right] \times \cdots \\ &\cdots \times \left[ \prod_k F_{D_{\tau,k}}(\{d_{\tau,k}\}) \right] \sum_{\vec{d}_{\tau-\nu_0}} \mu_{\vec{K}_{\tau-1}\vec{K}_{\tau}}^{+ (1)} \left( y_{0:t_2}; \left\{ \left[ \vec{d}_{\tau-1}^T \quad \vec{d}_{\tau-2}^T \cdots \quad \vec{d}_{\tau-\nu_0}^T \right]^T \right\} \right) \end{aligned} \quad (8.60)$$

Avec les notations 8.48 et 8.52, l'équation 8.60 peut s'écrire

$$\begin{aligned} \mu_{\vec{K}_{\tau}C_{\tau,n}}^{+ (1)} (y_{0:t_2}; \{\vec{k}_{\tau}\}) &\propto \beta_{\tau}^{(1)}(\vec{d}_{\tau}, \vec{d}_{\tau-1}, \dots, \vec{d}_{\tau-\nu_0+2}) \left[ \prod_{n' \neq n} \mu_{C_{\tau,n'}\vec{K}_{\tau}}^{- (1)} (\vec{k}_{\tau}; y_{0:t_2}) \right] \times \cdots \\ &\cdots \times \left[ \prod_k F_{D_{\tau,k}}(\{d_{\tau,k}\}) \right] \alpha_{\tau-1}^{(1)}(\vec{d}_{\tau-1}, \vec{d}_{\tau-2}, \dots, \vec{d}_{\tau-\nu_0+1}) \end{aligned} \quad (8.61)$$

Pour tout  $j \in \{j_1, j_1 + 1, \dots, j_2\}$ , on peut alors déterminer le message envoyé par  $C_j$  à  $C'_{\Pi(j)}$

$$\mu_{C_j C'_{\Pi(j)}}^{+(1)}(y_{0:t_2}; \cdot) = \sum_{\vec{k}_\tau} \mu_{\vec{k}_\tau C_j}^{+(1)}(y_{0:t_2}; \cdot) \Pi_{C_j}^{\vec{k}_\tau} \quad (8.62)$$

avec

$$\Pi_{C_j}^{\vec{k}_\tau}(\vec{k}_\tau; \cdot) = \delta_{\psi_n(\vec{k}_\tau)} \quad (8.63)$$

$$\tau = \left\lfloor \frac{j}{n_0} \right\rfloor \quad (8.64)$$

$$n = j - \tau n_0 + 1 \quad (8.65)$$

On en déduit

$$\mu_{C_j C'_{\Pi(j)}}^{+(1)}(y_{0:t_2}; \{c_j\}) = \sum_{\substack{\vec{k}_\tau \\ \psi_n(\vec{k}_\tau) = c_j}} \mu_{\vec{k}_\tau C_j}^{+(1)}(y_{0:t_2}; \{\vec{k}_\tau\}) \quad (8.66)$$

Pour tout  $t \in \{t_1, \dots, t_2\}$  et pour tout  $q \in \{1, 2, \dots, q_0\}$ , on peut calculer le message  $\mu_{C'_{t,q} A_t}^{+(1)}$  envoyé par  $C'_{t,q}$  à  $A_t$

$$\mu_{C'_{t,q} A_t}^{+(1)}(y_{0:t_2}; \cdot) = \sum_{C_j} \mu_{C_j C'_{t,q}}^{+(1)}(y_{0:t_2}; \cdot) \otimes \Pi_{C'_{t,q}}^{C_j} \quad (8.67)$$

avec

$$\Pi_{C'_{t,q}}^{C_j}(c_j; \cdot) = \delta_{c_j} \quad (8.68)$$

$$j = \Pi^{-1}(tq_0 + q - 1) \quad (8.69)$$

On en déduit

$$\mu_{C'_{t,q} A_t}^{+(1)} = \mu_{C_j C'_{t,q}}^{+(1)} \quad (8.70)$$

Pour tout  $t \in \{t_1, \dots, t_2\}$ , on peut calculer le message envoyé par  $A_t$  à  $\vec{B}_t$

$$\mu_{A_t \vec{B}_t}^{+(1)}(y_{0:t_2}; \cdot) = \sum_{(C'_{t,1}, \dots, C'_{t,q_0})} [\mu_{C'_{t,1} A_t}^{+(1)}(y_{0:t_2}; \cdot) \otimes \dots \otimes \mu_{C'_{t,q_0} A_t}^{+(1)}(y_{0:t_2}; \cdot)] \otimes \Pi_{A_t}^{(C'_{t,1}, \dots, C'_{t,q_0})} \quad (8.71)$$

En utilisant l'équation 8.31, on en déduit

$$\mu_{A_t \vec{B}_t}^{+(1)}(y_{0:t_2}; \{a_t\}) = \sum_{\substack{(c'_{t,1}, \dots, c'_{t,q_0}) \\ \mathcal{M}(c'_{t,1}, \dots, c'_{t,q_0}) = a_t}} \mu_{C'_{t,1} A_t}^{+(1)}(y_{0:t_2}; \{c'_{t,1}\}) \times \dots \times \mu_{C'_{t,q_0} A_t}^{+(1)}(y_{0:t_2}; \{c'_{t,q_0}\}) \quad (8.72)$$

On considère alors le réseau bayésien  $\tilde{G}^{(1)}$  tel que

1.  $\tilde{G}^{(1)}$  a pour sommets des variables aléatoires notées  $(\tilde{A}_t^{(1)})_{0 \leq t \leq t_2}$ ,  $(\tilde{B}_t^{(1)})_{0 \leq t \leq t_2}$  et  $(\tilde{Y}_t^{(1)})_{0 \leq t \leq t_2}$
2. les arcs de  $\tilde{G}^{(1)}$  sont identiques à ceux du sous-graphe  $G'$ .
3. les transitions entre les parents et les enfants de  $\tilde{G}^{(1)}$  sont identiques aux transitions entre les sommets correspondants de  $G'$ .
4. les variables aléatoires  $\tilde{A}_t^{(1)}$  avec  $0 \leq t \leq t_2$  sont indépendantes et ont pour loi a priori

$$F_{\tilde{A}_t^{(1)}} = \begin{cases} \mu_{A_t \tilde{B}_t}^+ & \text{si } 0 \leq t \leq t_1 - 1 \\ \mu_{A_t \tilde{B}_t}^{+(1)} & \text{si } t_1 \leq t \leq t_2 \end{cases} \quad (8.73)$$

Pour tout  $t' \in \{t_1, \dots, t_2\}$ , on calcule une approximation  $\hat{\Pi}_{\tilde{A}_{0:t'}}^{\tilde{Y}_{0:t'}^{(1)}}(y_{0:t'}; \cdot)$  de la loi de  $\tilde{A}_{0:t'}^{(1)}$  conditionnellement à  $\tilde{Y}_{0:t'}^{(1)} = y_{0:t'}$  en appliquant un filtre particulière au modèle dynamique à processus indicateur latent dont le réseau bayésien est  $\tilde{G}^{(1)}$ . On en déduit pour tout  $t \in \{t_1, \dots, t_2\}$  une approximation  $\hat{\Pi}_{\tilde{A}_t^{(1)}}^{\tilde{Y}_{0:t'}^{(1)}}(y_{0:t'}; \cdot)$  de la loi de  $\tilde{A}_t^{(1)}$  conditionnellement à  $\tilde{Y}_{0:t'}^{(1)} = y_{0:t'}$

$$\hat{\Pi}_{\tilde{A}_t^{(1)}}^{\tilde{Y}_{0:t'}^{(1)}}(y_{0:t'}; \cdot) = \sum_{(\tilde{A}_{t''}^{(1)})_{t'' \neq t}} \hat{\Pi}_{\tilde{A}_{0:t''}^{(1)}}^{\tilde{Y}_{0:t''}^{(1)}}(y_{0:t''}; \cdot) \quad (8.74)$$

avec  $t' = \min(t + \delta, t_2)$  et on pose

$$\mu_{\tilde{B}_t A_t}^{-(2)}(\cdot; y_{0:t_2}) = \frac{d\hat{\Pi}_{\tilde{A}_t^{(1)}}^{\tilde{Y}_{0:t'}^{(1)}}(y_{0:t'}; \cdot)}{d\mu_{A_t \tilde{B}_t}^{+(1)}(y_{0:t_2}; \cdot)} \quad (8.75)$$

On poursuit ensuite les itérations sur ce bloc de données jusqu'à ce que le nombre  $r$  d'itérations effectuées atteint un certain seuil  $r_{\max}$  ou que la variation de l'estimation de la loi *a posteriori* des  $D_{\tau, k}$  avec  $k \in \{1, 2, \dots, k_0\}$  d'une itération à l'autre reste en dessous d'un certain seuil.<sup>5</sup> On calcule alors à l'aide de la formule 8.53 le terme  $\alpha_{\tau_2}$  qui correspond à cette dernière itération et on le stocke pour le traitement du bloc de données suivant. De même, les  $N$  trajectoires  $(A^{(j, t_2)})_{1 \leq j \leq N}$  qui ont été calculées lors de cette dernière itération doivent être également stockées pour le traitement du bloc de données suivant.

<sup>5</sup>On peut par exemple prendre comme critère d'arrêt

$$\forall k \in \{1, 2, \dots, k_0\} \quad \left| \widehat{\Pi}_{D_{\tau, k}}^{Y_{0:t_2}}^{(r)}(y_{0:t_2}; \{0\}) - \widehat{\Pi}_{D_{\tau, k}}^{Y_{0:t_2}}^{(r-1)}(y_{0:t_2}; \{0\}) \right| \leq \epsilon_0 \quad (8.76)$$

où  $\epsilon_0$  est un seuil arbitraire.

### 8.6.2 Initialisation

L'initialisation du filtre particulaire a été décrite dans le chapitre consacré au filtrage particulaire.

L'initialisation de la récursion 8.53 se fait en remarquant que le message envoyé par  $\vec{K}_0$  à  $\vec{K}_1$  à l'itération  $r$  vérifie

$$\mu_{\vec{K}_0 \vec{K}_1}^{+(r)}(y_{0:t_2}; \cdot) = \Psi_{\text{Bayes}} \left( \prod_n \mu_{C_{0,n} \vec{K}_0}^{-(r)}(\cdot; y_{0:t_2}), \sum_{\vec{D}_0} \left[ \bigotimes_{k=1}^{k_0} F_{D_{0,k}} \right] \otimes \Pi_{\vec{K}_0}^{\vec{D}_0} \right) \quad (8.77)$$

avec  $t_2 = \frac{mn_0}{q_0 k_0} - 1$  et  $\Pi_{\vec{K}_0}^{\vec{D}_0}(\vec{d}_0; \cdot) = \delta_{\begin{bmatrix} \vec{d}_0 \\ 0 \end{bmatrix}}$ . On en déduit

$$\begin{aligned} \alpha_0^{(r)}(\vec{d}_0, \vec{d}_{-1}, \dots, \vec{d}_{-\nu_0+2}) &\propto \left[ \prod_k F_{D_{0,k}}(\{d_{0,k}\}) \right] \times \dots \\ \dots \times \sum_{\vec{d}_{-\nu_0+1}} &\left[ \prod_n \mu_{C_{0,n} \vec{K}_0}^{-(r)} \left( \begin{bmatrix} \vec{d}_0^T & \vec{d}_{-1}^T & \dots & \vec{d}_{-\nu_0+1}^T \end{bmatrix}^T; y_{0:t_2} \right) \right] \alpha_{-1}(\vec{d}_{-1}, \vec{d}_{-2}, \dots, \vec{d}_{-\nu_0+1}) \end{aligned} \quad (8.78)$$

avec

$$\alpha_{-1}(\vec{d}_{-1}, \vec{d}_{-2}, \dots, \vec{d}_{-\nu_0+1}) \stackrel{\text{déf}}{=} \begin{cases} 1 & \text{si } \vec{d}_{-1} = \vec{d}_{-2} = \dots = \vec{d}_{-\nu_0+1} = 0 \\ 0 & \text{sinon} \end{cases} \quad (8.79)$$

## 8.7 Conclusion

Dans ce dernier chapitre, nous avons montré comment l'ensemble des techniques étudiées précédemment peuvent être combinées pour concevoir de façon méthodique un récepteur de communication numérique.

L'approche que nous avons suivie comporte trois grandes étapes. Tout d'abord, nous représentons l'ensemble de l'information *a priori* dont dispose le récepteur à l'aide d'un réseau bayésien. Puis nous appliquons l'algorithme de propagation de croyance à ce réseau bayésien pour calculer itérativement une approximation de la loi *a posteriori* des bits transmis. Enfin, nous réduisons la complexité de l'algorithme obtenu à l'aide de techniques d'approximation numérique comme le filtrage particulaire.

A titre d'exemple, nous avons appliqué cette approche pour concevoir un algorithme d'égalisation et de décodage conjoints dans le cas d'un canal évanescent non-sélectif en fréquence. Grâce à cette approche, nous avons obtenu un algorithme d'égalisation et de décodage conjoints qui combine la propagation de croyance et le filtrage particulaire et qui, par construction, prend en compte l'ensemble de l'information *a priori* dont dispose le récepteur : code de canal, entrelacement, modulation, type de distorsions introduites par le canal et présence de séquences de référence dans le message émis.





# Conclusion

Au cours de cette thèse, nous avons étudié les réseaux bayésiens, le filtrage particulaire et leur application à l'égalisation adaptative et au décodage conjoints.

Nous avons généralisé la notion de réseau bayésien afin de pouvoir représenter de façon totalement rigoureuse les relations d'indépendance conditionnelle entre éléments aléatoires quelconques, et nous avons donné une formulation de l'algorithme de propagation de croyance adaptée à cette nouvelle définition en utilisant le théorème de Bayes généralisé.

Nous avons introduit un nouveau type de filtre particulaire que nous avons appelé *filtre particulaire à échantillonnage global*. Nous avons démontré que l'erreur quadratique moyenne de ce filtre converge vers 0 lorsque le nombre de particules tend vers l'infini. Nous avons constaté lors des simulations numériques que les performances de ce nouvel algorithme sont supérieures ou égales à celles de l'algorithme décrit dans [CWL00].

Nous avons ensuite combiné la propagation de croyance et le filtrage particulaire. Plus précisément, nous avons utilisé un filtre particulaire pour calculer de façon approchée certains messages de l'algorithme de propagation de croyance, dont le calcul exact aurait eu une complexité trop élevée. Nous avons appelé ce nouvel algorithme *algorithme de turbo-filtrage particulaire*.

Enfin, nous avons montré que ces différentes techniques permettent de concevoir un récepteur de communications numériques en utilisant une approche *unifiée et globale*. En effet, plutôt que de considérer que l'on a une succession de traitements distincts (l'estimation du canal, l'égalisation, le désentrelacement et le décodage) que l'on essaie de combiner, nous sommes partis d'un seul et unique problème (l'estimation des bits de données transmis) que nous avons résolu en appliquant de façon approchée l'algorithme de propagation de croyance à un réseau bayésien qui représente l'intégralité de l'information *a priori* dont dispose le récepteur.

Les suites que l'on peut donner à cette thèse sont multiples. On peut comparer systématiquement les différentes méthodes d'approximation de l'algorithme de propagation de croyance et rechercher des critères simples permettant de choisir le type d'approximation à adopter en fonction du compromis complexité-performance souhaité. On peut également étendre cette étude à des canaux non-linéaires ou à des systèmes multi-utilisateurs. Enfin, on peut examiner l'intérêt des algorithmes de propagation de croyance généralisée pour les réseaux bayésiens que nous avons étudiés au cours de cette thèse.



# Annexe A

## Quelques notions fondamentales de la théorie des graphes

### Définition A.1

Soit  $U$  un ensemble fini. Soit  $A$  un ensemble de couples d'éléments de  $U$ . On appelle

- **graphe orienté** sur  $U$  le couple  $G = (U, A)$
- **sommets** du graphe orienté  $G$  les éléments de  $U$
- **arcs** du graphe orienté  $G$  les éléments de  $A$
- **ordre** du graphe orienté  $G$  le cardinal de  $U$ .

Tout arc du graphe orienté  $G$  est donc de la forme  $(X, Y)$  avec  $X \in U$  et  $Y \in U$ . Le sommet  $X$  est appelé **extrémité initiale** de l'arc  $(X, Y)$  et le sommet  $Y$  **extrémité terminale** de l'arc  $(X, Y)$ . On convient de noter  $X \rightarrow Y$  l'arc  $(X, Y)$ .

On dit qu'un arc  $\alpha$  du graphe orienté  $G$  est une **boucle** si et seulement si l'extrémité initiale de  $\alpha$  est égale à son extrémité terminale.

### Définition A.2

Soient  $G$  un graphe orienté et  $X$  et  $Y$  deux sommets du graphe orienté  $G$ . On dit que le sommet  $X$  est un **parent** du sommet  $Y$  dans le graphe orienté  $G$  ou que le sommet  $Y$  est un **enfant** du sommet  $X$  dans le graphe orienté  $G$  si et seulement si le couple  $(X, Y)$  est un arc du graphe orienté  $G$ .

### Définition A.3

Soient  $G$  un graphe orienté et  $X$  un sommet du graphe orienté  $G$ . On appelle **degré** du sommet  $X$  le nombre d'arcs de  $G$  ayant  $X$  comme extrémité.

### Définition A.4

Soit  $G = (U, A)$  un graphe orienté. On appelle **sous-graphe** du graphe orienté  $G$  tout graphe orienté  $G' = (U', A')$  vérifiant

$$U' \subset U \quad (\text{A.1})$$

$$A' \subset (U' \times U') \cap A \quad (\text{A.2})$$

Lorsque  $A' = (U' \times U') \cap A$ , on dit que  $G'$  est le sous-graphe de  $G$  **induit** (on dit aussi **engendré**) par le sous-ensemble de sommets  $U'$ .

**Définition A.5**

Soit  $G = (U, A)$  un graphe orienté. On appelle **graphe partiel** du graphe orienté  $G$  tout graphe orienté  $G'$  de la forme  $G' = (U, A')$  avec  $A' \subset A$ .

**Définition A.6**

Soit  $G$  un graphe orienté. Soit  $q$  un entier strictement positif. On appelle **chaîne de longueur**  $q$  du graphe orienté  $G$  toute suite  $(\alpha_1, \dots, \alpha_q)$  d'arcs distincts deux à deux du graphe orienté  $G$  vérifiant l'une des conditions suivantes :

- $q$  est supérieur ou égal à 3 et, pour tout  $r \in \{2, \dots, q-1\}$ , l'arc  $\alpha_r$  possède une extrémité en commun avec l'arc  $\alpha_{r-1}$  et l'autre en commun avec l'arc  $\alpha_{r+1}$ ,
- $q$  est égal à 2 et l'arc  $\alpha_1$  possède une extrémité en commun avec l'arc  $\alpha_2$ ,
- $q$  est égal à 1.

Lorsqu'une chaîne  $(\alpha_1, \dots, \alpha_q)$  est de longueur supérieure ou égale à 2, l'extrémité  $X$  de  $\alpha_1$  qui n'est pas commune avec l'arc  $\alpha_2$  et l'extrémité  $Y$  de  $\alpha_q$  qui n'est pas commune avec l'arc  $\alpha_{q-1}$  sont appelées les **extrémités** de la chaîne  $(\alpha_1, \dots, \alpha_q)$  et on dit alors que la chaîne  $(\alpha_1, \dots, \alpha_q)$  **joint** les sommets  $X$  et  $Y$ . Lorsqu'une chaîne  $L$  est réduite à un unique arc  $\alpha_1$ , on appelle extrémités de la chaîne  $L$  les extrémités  $X$  et  $Y$  de l'arc  $\alpha_1$  et on dit que la chaîne  $L$  joint les sommets  $X$  et  $Y$ .

**Définition A.7**

Soit  $G$  un graphe orienté. Soit  $L$  une chaîne du graphe orienté  $G$ . On dit que  $L$  est un **cycle** du graphe orienté  $G$  si et seulement si les deux extrémités de la chaîne  $L$  sont identiques.

**Définition A.8**

Soit  $G$  un graphe orienté. On dit que  $G$  est **connexe** si et seulement si pour tout couple  $(X, Y)$  de sommets distincts de  $G$ , il existe une chaîne  $L$  du graphe orienté  $G$  qui joint les sommets  $X$  et  $Y$ .

**Définition A.9**

Soit  $G$  un graphe orienté. Soit  $(\alpha_1, \dots, \alpha_q)$  une chaîne du graphe orienté  $G$ . On dit que la chaîne  $(\alpha_1, \dots, \alpha_q)$  est un **chemin** du graphe orienté  $G$  si et seulement si pour tout  $r \in \{1, \dots, q-1\}$ , l'extrémité terminale de l'arc  $\alpha_r$  coïncide avec l'extrémité initiale de l'arc  $\alpha_{r+1}$ . L'extrémité initiale  $X$  (resp. terminale  $Y$ ) de l'arc  $\alpha_1$  (resp.  $\alpha_q$ ) est alors appelée **extrémité initiale** (resp. **terminale**) du chemin  $(\alpha_1, \dots, \alpha_q)$  et on dit que le chemin  $(\alpha_1, \dots, \alpha_q)$  **va** du sommet  $X$  au sommet  $Y$ .

**Définition A.10**

Soit  $G$  un graphe orienté. Soit  $L$  un chemin du graphe orienté  $G$ . On dit que  $L$  est un **circuit** du graphe orienté  $G$  si et seulement si les deux extrémités du chemin  $L$  sont identiques.

**Définition A.11**

Soit  $G$  un graphe orienté. Soient  $X$  et  $Y$  deux sommets du graphe orienté  $G$ . On dit que le sommet  $X$  est un **ancêtre propre** du sommet  $Y$  dans le graphe orienté  $G$ , ou que le sommet  $Y$  est un **descendant propre** du sommet  $X$  dans le graphe orienté  $G$  si et seulement si il existe un chemin du graphe orienté  $G$  qui va du sommet  $X$  au sommet  $Y$ .

# Annexe B

## Démonstration de la proposition 2.1

La démonstration de la proposition 2.1 comporte deux étapes. On démontre tout d'abord par récurrence que le graphe  $G$  est un  $I$ -map du graphoïde  $M$  (proposition B.1), puis on démontre par l'absurde que le graphe  $G$  est un  $I$ -map minimal (proposition B.2).

### Proposition B.1

Soient  $M = (\mathbf{U}, I)$  un graphoïde,  $d = (X_1, \dots, X_n)$  un arrangement des éléments de  $\mathbf{U}$  et  $G = (\mathbf{U}, \mathbf{A})$  un graphe orienté sans circuit tel que pour tout  $i \in \{1, \dots, n\}$ , l'ensemble  $\text{Pa}(X_i)_G$  des parents de  $X_i$  dans  $G$  est une couverture markovienne de  $\{X_i\}$  par rapport à  $\{X_1, \dots, X_i\}$  pour le graphoïde  $M$ . Alors le graphe  $G$  est un  $I$ -map du graphoïde  $M$ .

DÉMONSTRATION :

Démontrons cette propriété par récurrence sur le nombre  $n$  d'éléments du graphoïde.

Lorsque  $n = 1$ , l'ensemble  $\mathbf{U}$  est réduit au singleton  $\{X_1\}$  et l'ensemble des parties de  $\mathbf{U}$  est alors  $\mathcal{P}(\mathbf{U}) = \{\emptyset, \mathbf{U}\}$ . Pour démontrer que le graphe  $G$  est un  $I$ -map du graphoïde  $M$ , il suffit donc de montrer les implications suivantes :

$$\langle \emptyset | \emptyset | \emptyset \rangle_G \Rightarrow I(\emptyset, \emptyset, \emptyset) \quad (\text{B.1})$$

$$\langle \mathbf{U} | \emptyset | \emptyset \rangle_G \Rightarrow I(\mathbf{U}, \emptyset, \emptyset) \quad (\text{B.2})$$

$$\langle \emptyset | \mathbf{U} | \emptyset \rangle_G \Rightarrow I(\emptyset, \mathbf{U}, \emptyset) \quad (\text{B.3})$$

$$\langle \emptyset | \emptyset | \mathbf{U} \rangle_G \Rightarrow I(\emptyset, \emptyset, \mathbf{U}) \quad (\text{B.4})$$

Par hypothèse, l'ensemble  $\text{Pa}(X_1)_G$  des parents de  $X_1$  dans  $G$  est une couverture markovienne de  $\{X_1\}$  par rapport à  $\{X_1\}$ . Le graphe  $G$  étant supposé sans circuit, le sommet  $X_1$  ne peut être son propre parent et l'ensemble  $\text{Pa}(X_1)_G$  est donc vide. On en déduit

$$I(\{X_1\}, \emptyset, \{X_1\} \setminus (\emptyset \cup \{X_1\})) \quad (\text{B.5})$$

i.e.

$$I(\mathbf{U}, \emptyset, \emptyset) \quad (\text{B.6})$$

ce qui prouve l'implication B.2.

En utilisant la propriété de symétrie du graphoïde  $M$  (cf équation 2.3), on a

$$I(\mathbf{U}, \emptyset, \emptyset) \Rightarrow I(\emptyset, \emptyset, \mathbf{U}) \quad (\text{B.7})$$

ce qui prouve l'implication B.4.

En utilisant la propriété de décomposition du graphoïde  $M$  (cf équation 2.4), on a

$$I(\emptyset, \emptyset, U) \Rightarrow I(\emptyset, \emptyset, U \cup \emptyset) \quad (\text{B.8})$$

$$\Rightarrow I(\emptyset, \emptyset, U) \wedge I(\emptyset, \emptyset, \emptyset) \quad (\text{B.9})$$

ce qui prouve l'implication B.1.

En utilisant la propriété d'union faible du graphoïde  $M$  (cf équation 2.5), on a

$$I(\emptyset, \emptyset, U \cup \emptyset) \Rightarrow I(\emptyset, \emptyset \cup U, \emptyset) \quad (\text{B.10})$$

ce qui prouve l'implication B.3. La propriété à démontrer est donc vraie pour  $n = 1$ .

Supposons maintenant que la propriété est vraie pour  $n - 1$  éléments et montrons qu'elle est vraie pour  $n$  éléments,  $n$  étant strictement plus grand que 1. Considérons donc  $X, Y, Z \in \mathcal{P}(U)$  vérifiant  $\langle X | Z | Y \rangle_G$  et montrons que l'on a  $I(X, Z, Y)$  :

1. si  $X_n$  n'appartient à aucun des sous-ensembles  $X, Y$  et  $Z$ , alors les sous-ensembles  $X, Y$  et  $Z$  sont inclus dans  $U_{n-1} = \{X_1, \dots, X_{n-1}\}$ . En notant  $A_{n-1}$  l'ensemble des arcs de  $G$  dont aucune des extrémités n'est  $X_n$ , on voit facilement que le graphe orienté  $G_{n-1} = (U_{n-1}, A_{n-1})$  est un sous-graphe du graphe orienté  $G$ . Toute chaîne  $c$  de  $G_{n-1}$  joignant un sommet de  $X$  à un sommet de  $Y$  est donc une chaîne de  $G$ . Comme on a  $\langle X | Z | Y \rangle_G$ , la chaîne  $c$  est bloquée dans  $G$  (et donc dans  $G_{n-1}$  également) par le sous-ensemble  $Z$ . On en déduit  $\langle X | Z | Y \rangle_{G_{n-1}}$ . En appliquant l'hypothèse de récurrence au graphoïde  $M_{n-1} = (U_{n-1}, I_{n-1})$ , où  $I_{n-1}$  désigne la restriction de  $I$  à  $U_{n-1}$ ; à l'arrangement  $d_{n-1} = \{X_1, \dots, X_{n-1}\}$  et au graphe  $G_{n-1}$ , on en déduit  $I(X, Z, Y)$ .
2. si  $X_n$  appartient à  $X$ , alors par hypothèse on a

$$I(\{X_n\}, B, R) \quad (\text{B.11})$$

où  $B = \text{Pa}(X_n)_G$  désigne l'ensemble des parents de  $X_n$  dans  $G$  et  $R = U \setminus (\text{Pa}(X_n)_G \cup \{X_n\})$  désigne l'ensemble des éléments de  $U$  qui sont distincts de  $X_n$  et qui ne sont pas parents de  $X_n$  dans  $G$ .

Notons  $X' = X \setminus \{X_n\}$  l'ensemble des éléments de  $X$  distincts de  $X_n$  et partitionnons l'ensemble  $B$  en les sous-ensembles suivants

- $B_{X'} = X' \cap B$  l'ensemble des éléments de  $X'$  parents de  $X_n$  dans  $G$
- $B_Y = Y \cap B$  l'ensemble des éléments de  $Y$  parents de  $X_n$  dans  $G$
- $B_Z = Z \cap B$  l'ensemble des éléments de  $Z$  parents de  $X_n$  dans  $G$
- $B_0 = B \setminus (B_{X'} \cup B_Y \cup B_Z)$  l'ensemble des parents de  $X_n$  dans  $G$  qui n'appartiennent ni à  $X'$ , ni à  $Y$ , ni à  $Z$

De même, partitionnons l'ensemble  $R$  en les sous-ensembles suivants

- $R_{X'} = X' \cap R$  l'ensemble des éléments de  $X'$  qui ne sont pas parents de  $X_n$  dans  $G$
- $R_Y = Y \cap R$  l'ensemble des éléments de  $Y$  qui ne sont pas parents de  $X_n$  dans  $G$
- $R_Z = Z \cap R$  l'ensemble des éléments de  $Z$  qui ne sont pas parents de  $X_n$  dans  $G$
- $R_0 = R \setminus (R_{X'} \cup R_Y \cup R_Z)$  l'ensemble des éléments de  $U$  qui sont distincts de  $X_n$ , qui ne sont pas parents de  $X_n$  dans  $G$  et qui n'appartiennent ni à  $X'$ , ni à  $Y$ , ni à  $Z$

La relation B.11 peut alors s'écrire ainsi

$$I(\{X_n\}, B_{X'} \cup B_Y \cup B_Z \cup B_0, R_{X'} \cup R_Y \cup R_Z \cup R_0) \quad (\text{B.12})$$

En utilisant la propriété d'union faible du graphoïde  $M$  (cf équation 2.5), on en déduit

$$I(\{X_n\}, B_{X'} \cup R_{X'} \cup B_Y \cup B_Z \cup R_Z \cup B_0, R_Y \cup R_0) \quad (\text{B.13})$$

i.e.

$$I(\{X_n\}, X' \cup B_Y \cup Z \cup B_0, R_Y \cup R_0) \quad (\text{B.14})$$

Or, on a également supposé

$$\langle X | Z | Y \rangle_G \quad (\text{B.15})$$

i.e.

$$\langle X' \cup \{X_n\} | Z | Y \rangle_G \quad (\text{B.16})$$

Par conséquent,  $X_n$  ne peut avoir de parents appartenant à  $Y$ . Autrement dit, on a

$$B_Y = \emptyset \quad (\text{B.17})$$

$$R_Y = Y \quad (\text{B.18})$$

On déduit donc de la relation B.14 et des équations B.17 et B.18 la relation suivante

$$I(\{X_n\}, X' \cup Z \cup B_0, Y \cup R_0) \quad (\text{B.19})$$

En utilisant la propriété de décomposition du graphoïde  $M$  (cf équation 2.4), on en déduit

$$I(\{X_n\}, X' \cup Z \cup B_0, Y) \quad (\text{B.20})$$

Afin d'utiliser la propriété de contraction du graphoïde  $M$  (cf équation 2.6), montrons maintenant que l'on a la relation suivante

$$I(X' \cup B_0, Z, Y) \quad (\text{B.21})$$

Montrons tout d'abord en raisonnant par l'absurde que  $Z$   $d$ -sépare  $B_0$  de  $Y$  dans  $G$ . Supposons que  $Z$  ne  $d$ -sépare pas  $B_0$  de  $Y$  dans  $G$ . Il existe alors une chaîne  $c = (\alpha_1, \dots, \alpha_q)$  active par rapport à  $Z$  joignant un sommet  $Y$  de  $Y$  et un sommet  $B$  de  $B_0$ . Par définition de  $B_0$ , le sommet  $B$  est un parent de  $X_n$  qui n'appartient ni à  $X'$ , ni à  $Y$ , ni à  $Z$ . Il existe donc un arc  $\alpha$  de  $G$  dont l'extrémité initiale est  $B$  et l'extrémité terminale  $X_n$ . On considère alors la chaîne  $c' = (\alpha_1, \dots, \alpha_q, \alpha)$  qui joint les sommets  $Y$  et  $X_n$ . Comme la chaîne  $c = (\alpha_1, \dots, \alpha_q)$  est active par rapport à  $Z$  et comme la chaîne  $c'$  est en série ou diverge en  $B$  qui n'appartient pas à  $Z$ , la chaîne  $c'$  est active par rapport à  $Z$ , ce qui est absurde puisque d'après la relation B.16,  $Z$   $d$ -sépare  $X'$  de  $\{X_n\}$  dans  $G$ . On en déduit que  $Z$   $d$ -sépare  $B_0$  de  $Y$  dans  $G$ .

De plus, d'après la relation B.16,  $Z$   $d$ -sépare  $X'$  de  $Y$  dans  $G$ . On en déduit que  $Z$   $d$ -sépare  $X' \cup B_0$  de  $Y$  dans  $G$ . Autrement dit, on a

$$\langle X' \cup B_0 | Z | Y \rangle_G \quad (\text{B.22})$$

Comme  $X_n$  n'appartient ni à  $X' \cup B_0$ , ni à  $Z$ , ni à  $Y$ , on peut utiliser le résultat démontré précédemment dans le cas numéro 1 et déduire finalement de la relation B.22 la relation B.21.



En utilisant la propriété de contraction du graphoïde  $M$  (cf équation 2.6), on déduit des relations B.20 et B.21 la relation

$$I(\{X_n\} \cup X' \cup B_0, Z, Y) \quad (\text{B.23})$$

ainsi que, en utilisant la propriété de décomposition du graphoïde  $M$  (cf équation 2.4), la relation

$$I(\{X_n\} \cup X', Z, Y) \quad (\text{B.24})$$

qui est la relation à démontrer dans le cas numéro 2.

3. si  $X_n$  appartient à  $Y$ , on peut se ramener au cas précédent en utilisant la propriété de symétrie du graphoïde  $M$  et la propriété de symétrie de la relation de  $d$ -séparation.
4. si  $X_n$  appartient à  $Z$ , on note  $Z' = Z \setminus \{X_n\}$  l'ensemble des éléments de  $Z$  distincts de  $X_n$  et on a

$$\langle X \mid \{X_n\} \cup Z' \mid Y \rangle_G \quad (\text{B.25})$$

Toute chaîne  $c$  joignant un sommet de  $X$  et un sommet de  $Y$  est donc bloquée par  $\{X_n\} \cup Z'$ . Autrement dit, la chaîne  $c$  vérifie au moins l'une des conditions suivantes :

- (a) la chaîne  $c$  diverge en  $X_n$  ou en un sommet de  $Z'$
- (b) la chaîne  $c$  est en série en  $X_n$  ou en un sommet de  $Z'$
- (c) la chaîne  $c$  converge en un sommet qui n'appartient pas à  $\{X_n\} \cup Z'$ , et dont aucun descendant propre dans le graphe  $G$  n'appartient à  $\{X_n\} \cup Z'$ .

Or on a supposé que pour tout  $i \in \{1, \dots, n\}$ , l'ensemble  $\text{Pa}(X_i)_G$  des parents de  $X_i$  dans  $G$  est une couverture markovienne de  $\{X_i\}$  par rapport à  $\{X_1, \dots, X_i\}$ , ce qui implique que  $X_n$  ne fait pas partie des parents de  $X_i$ . Par conséquent,  $X_n$  n'a pas d'enfants dans  $G$  et n'est donc l'extrémité initiale d'aucun arc de  $G$ . La chaîne  $c$  ne peut donc diverger ou être en série en  $X_n$ . On en déduit immédiatement que la chaîne  $c$  est également bloquée par  $Z'$ . On a donc

$$\langle X \mid Z' \mid Y \rangle_G \quad (\text{B.26})$$

Montrons maintenant que les relations B.25 et B.26 impliquent

$$\langle X \mid Z' \mid \{X_n\} \rangle_G \vee \langle \{X_n\} \mid Z' \mid Y \rangle_G \quad (\text{B.27})$$

Supposons que la proposition B.27 est fautive. Alors  $X$  et  $\{X_n\}$  ne sont pas  $d$ -séparés par  $Z'$ , et  $\{X_n\}$  et  $Y$  ne sont pas  $d$ -séparés par  $Z'$ . Il existe donc une chaîne  $c = (\alpha_1, \dots, \alpha_q)$  active par rapport à  $Z'$  joignant un sommet  $X \in X$  et le sommet  $X_n$ , et une chaîne  $c' = (\alpha'_1, \dots, \alpha'_{q'})$  active par rapport à  $Z'$  joignant le sommet  $X_n$  et un sommet  $Y \in Y$ . On définit les indices  $i$  et  $i'$  de la façon suivante

$$i = \min \{ r \mid X_n \in \mathcal{E}(\alpha_r), 1 \leq r \leq q \} \quad (\text{B.28})$$

$$i' = \max \{ r \mid X_n \in \mathcal{E}(\alpha'_r), 1 \leq r \leq q' \} \quad (\text{B.29})$$

où la notation  $\mathcal{E}(\alpha)$  désigne l'ensemble des extrémités de l'arc  $\alpha$ . Considérons alors la chaîne  $c'' = (\alpha_1, \dots, \alpha_i, \alpha'_{i'}, \dots, \alpha'_{q'})$  qui joint les sommets  $X$  et  $Y$ . Le sommet  $X_n$  étant une extrémité commune des arcs  $\alpha_i$  et  $\alpha'_{i'}$ , la chaîne  $c''$  vérifie l'une des conditions suivantes :

- (a) la chaîne  $c''$  converge en  $X_n$  qui appartient à  $Z' \cup \{X_n\}$ . On vérifie alors facilement que la chaîne  $c''$  est active par rapport à  $Z' \cup \{X_n\}$ , ce qui contredit la proposition B.25.

(b) la chaîne  $c''$  diverge ou est en série en  $X_n$  qui n'appartient pas à  $Z'$ . On vérifie alors facilement que la chaîne  $c''$  est active par rapport à  $Z'$ , ce qui contredit la proposition B.26.

Par conséquent, la proposition B.27 est vraie.

On remarque maintenant que les propositions B.26 et B.27 impliquent la proposition suivante

$$\langle \mathbf{X} \mid \mathbf{Z}' \mid \mathbf{Y} \cup \{X_n\} \rangle_G \vee \langle \mathbf{X} \cup \{X_n\} \mid \mathbf{Z}' \mid \mathbf{Y} \rangle_G \quad (\text{B.30})$$

ce qui nous ramène donc au cas numéro 3 ou au cas numéro 2. On en déduit

$$I(\mathbf{X}, \mathbf{Z}', \mathbf{Y} \cup \{X_n\}) \vee I(\mathbf{X} \cup \{X_n\}, \mathbf{Z}', \mathbf{Y}) \quad (\text{B.31})$$

En utilisant la propriété d'union faible du graphoïde  $M$  (cf équation 2.5), on en déduit

$$I(\mathbf{X}, \mathbf{Z}' \cup \{X_n\}, \mathbf{Y}) \quad (\text{B.32})$$

qui est la relation à démontrer dans le cas numéro 4.

Par conséquent, la propriété à démontrer est vraie pour  $n$  éléments et donc, en raisonnant par récurrence, pour un nombre quelconque d'éléments.  $\square$

### Proposition B.2

Soient  $M = (\mathbf{U}, I)$  un graphoïde,  $d = (X_1, \dots, X_n)$  un arrangement des éléments de  $\mathbf{U}$  et  $G = (\mathbf{U}, \mathbf{A})$  un graphe orienté sans circuit tel que pour tout  $i \in \{1, \dots, n\}$ , l'ensemble  $\text{Pa}(X_i)_G$  des parents de  $X_i$  dans  $G$  est une frontière markovienne de  $\{X_i\}$  par rapport à  $\{X_1, \dots, X_i\}$  pour le graphoïde  $M$ . Alors le graphe  $G$  est un  $I$ -map minimal du graphoïde  $M$ .

DÉMONSTRATION :

D'après la proposition B.1, le graphe  $G$  est un  $I$ -map du graphoïde  $M$ . Supposons qu'il existe  $G'$  un graphe partiel de  $G$  qui est distinct de  $G$  et qui est un  $I$ -map du graphoïde  $M$ . Comme  $G'$  est distinct de  $G$  il existe un arc  $\alpha$  de  $G$  qui n'est pas un arc de  $G'$ . Soit  $X_k$  l'extrémité terminale de cet arc  $\alpha$ . On a  $k > 1$  car l'ensemble  $\text{Pa}(X_1)_G$  des parents de  $X_1$  dans  $G$  forme, par hypothèse, une frontière markovienne de  $X_1$  par rapport à  $\{X_1\}$  et est donc réduit à l'ensemble vide, alors que l'extrémité initiale de l'arc  $\alpha$  fait partie de l'ensemble  $\text{Pa}(X_k)_G$  des parents de  $X_k$  dans  $G$ , qui n'est donc pas vide.

Montrons que l'on a

$$\langle \{X_k\} \mid \text{Pa}(X_k)_{G'} \mid \{X_1, \dots, X_{k-1}\} \setminus \text{Pa}(X_k)_{G'} \rangle_{G'} \quad (\text{B.33})$$

S'il n'existe pas de chaîne de  $G'$  joignant  $X_k$  à un sommet  $X_i \in \{X_1, \dots, X_{k-1}\} \setminus \text{Pa}(X_k)_{G'}$ , alors la relation B.33 est vraie. Sinon, soit  $c = (\alpha_1, \dots, \alpha_q)$  une chaîne de  $G'$  joignant  $X_k$  et un sommet  $X_i \in \{X_1, \dots, X_{k-1}\} \setminus \text{Pa}(X_k)_{G'}$ .

Si  $X_k$  est l'extrémité terminale de l'arc  $\alpha_1$ , alors on a  $q > 1$  car  $X_i$  ne fait pas partie des parents de  $X_k$  dans  $G'$ . Par conséquent, en notant  $X_j$  l'extrémité initiale de l'arc  $\alpha_1$ , le chemin  $c$  diverge ou est en série en  $X_j$  qui est un parent de  $X_k$  dans  $G'$ . Le chemin  $c$  est donc bloqué par  $\text{Pa}(X_k)_{G'}$ .

Si  $X_k$  est l'extrémité initiale de l'arc  $\alpha_1$ , on définit l'ensemble d'indices  $S$  par

$$S = \{r \mid T(\alpha_r) \neq I(\alpha_{r+1}), 1 \leq r \leq q-1\} \quad (\text{B.34})$$

où  $T(\alpha_r)$  désigne l'extrémité terminale de l'arc  $\alpha_r$  et  $I(\alpha_{r+1})$  l'extrémité initiale de l'arc  $\alpha_{r+1}$ . L'ensemble  $S$  est non vide car sinon la chaîne  $c$  formerait un chemin qui irait du sommet  $X_k$  au sommet  $X_i$ , et la suite  $(j_0, j_1, \dots, j_q)$  formée par l'indice (dans l'arrangement  $d$ ) des sommets rencontrés en parcourant ce chemin serait strictement croissante. Comme  $j_0 = k$  et  $j_q = i$ , on aurait alors  $k < i$  ce qui est

incompatible avec le fait que  $X_i$  appartienne à  $\{X_1, \dots, X_{k-1}\} \setminus \text{Pa}(X_k)_{G'}$ . Par conséquent, l'ensemble  $S$  est non vide et on peut donc définir l'indice  $s$  par

$$s = \min S \tag{B.35}$$

La chaîne  $c$  converge en  $T(\alpha_s)$  et on montre facilement par un raisonnement similaire au précédent que ni le sommet  $T(\alpha_s)$  ni aucun de ses descendants propres dans  $G'$  n'appartient à  $\text{Pa}(X_k)_{G'}$ . Par conséquent, la relation B.33 est vraie. Comme le graphe  $G'$  est supposé être un  $I$ -map du graphoïde  $M$ , la relation B.33 implique donc

$$I(\{X_k\}, \text{Pa}(X_k)_{G'}, \{X_1, \dots, X_{k-1}\} \setminus \text{Pa}(X_k)_{G'}) \tag{B.36}$$

Or, l'arc  $\alpha$  étant un arc de  $G$  mais pas de  $G'$ , l'extrémité initiale de l'arc  $\alpha$  appartient à  $\text{Pa}(X_k)_G$  mais pas à  $\text{Pa}(X_k)_{G'}$ , qui est donc strictement inclus dans  $\text{Pa}(X_k)_G$ . Par conséquent, la relation B.36 implique que  $\text{Pa}(X_k)_G$  n'est pas une frontière markovienne de  $\{X_k\}$  par rapport à  $\{X_1, \dots, X_k\}$  pour le graphoïde  $M$ , ce qui est absurde.  $\square$

# Annexe C

## Résultats utiles

### C.1 Théorème de Bayes généralisé

#### Proposition C.1

Soit  $(E, \mathcal{E})$  un espace mesurable. Soient  $\mu$  et  $\lambda$  deux mesures positives  $\sigma$ -finies sur  $(E, \mathcal{E})$  et telles que  $\mu$  est absolument continue par rapport à  $\lambda$ . Soit  $\nu$  une mesure signée sur  $(E, \mathcal{E})$  et absolument continue par rapport à  $\mu$ . Alors la dérivée de Radon-Nikodym de la mesure  $\nu$  par rapport à la mesure  $\mu$  vérifie

$$\frac{d\nu}{d\mu} = \frac{d\nu}{d\lambda} \cdot \frac{d\lambda}{d\mu} \quad (\mu\text{-p.s.}) \quad (\text{C.1})$$

et l'ensemble des  $x \in E$  tels que  $\frac{d\mu}{d\lambda}(x) = 0$  est de  $\mu$ -mesure nulle.

DÉMONSTRATION :

Cf [Shi96], §II-7. □

#### Définition C.1

Soit  $(E, \mathcal{E})$  un espace mesurable. On appelle **opérateur de Bayes** et on note  $\Psi_{\text{Bayes}}$  l'opérateur qui, à toute fonction mesurable positive  $g$  sur  $(E, \mathcal{E})$  et à toute mesure  $\mu$  sur  $(E, \mathcal{E})$ , associe la mesure

$$\Psi_{\text{Bayes}}(g, \mu) = \begin{cases} \frac{g}{\mu(g)} \cdot \mu & \text{si } \mu(g) \neq 0 \\ \mu & \text{sinon.} \end{cases} \quad (\text{C.2})$$

#### Proposition C.2

Soit  $(\Omega, \mathcal{F}, P)$  un espace de probabilité. Soient  $X_0, X_1$  et  $X_2$  des éléments aléatoires de l'espace mesurable  $(\Omega, \mathcal{F})$  dans, respectivement, les espaces mesurables  $(E_0, \mathcal{E}_0)$ ,  $(E_1, \mathcal{E}_1)$  et  $(E_2, \mathcal{E}_2)$  tels qu'il existe une transition  $\Pi_{X_1}^{X_0} : (E_0, \mathcal{E}_0) \prec (E_1, \mathcal{E}_1)$ , une transition  $\Pi_{X_2}^{(X_0, X_1)} : (E_0, \mathcal{E}_0) \otimes (E_1, \mathcal{E}_1) \prec (E_2, \mathcal{E}_2)$ , une mesure  $\lambda_2$  positive et  $\sigma$ -finie sur l'espace mesurable  $(E_2, \mathcal{E}_2)$  et une fonction mesurable  $f_{X_2}^{(X_0, X_1)}$  de  $(E_0, \mathcal{E}_0) \otimes (E_1, \mathcal{E}_1) \otimes (E_2, \mathcal{E}_2)$  dans  $(\overline{\mathbb{R}}_+, \mathcal{B}(\overline{\mathbb{R}}_+))$  vérifiant les propriétés suivantes :

1. pour tout  $B_1 \in \mathcal{E}_1$ ,  $\Pi_{X_1}^{X_0}(X_0; B_1)$  est une version de  $P(X_1 \in B_1 | X_0)$

2. pour tout  $B_2 \in \mathcal{E}_2$ ,  $\Pi_{X_2}^{(X_0, X_1)}(X_0, X_1; B_2)$  est une version de  $P(X_2 \in B_2 | X_0, X_1)$
3. pour tout  $(x_0, x_1) \in E_0 \times E_1$ , on a

$$\Pi_{X_2}^{(X_0, X_1)}(x_0, x_1; \cdot) = f_{X_2}^{(X_0, X_1)}(x_0, x_1; \cdot) \cdot \lambda \quad (\text{C.3})$$

Alors, l'application  $\Pi_{X_1}^{(X_0, X_2)}$  définie par, pour tout  $(x_0, x_2) \in E_0 \times E_2$

$$\Pi_{X_1}^{(X_0, X_2)}(x_0, x_2; \cdot) \stackrel{\text{déf}}{=} \Psi_{\text{Bayes}}(f_{X_2}^{(X_0, X_1)}(x_0, \cdot; x_2), \Pi_{X_1}^{X_0}(x_0; \cdot)) \quad (\text{C.4})$$

est une transition de l'espace mesurable  $(E_0, \mathcal{E}_0) \otimes (E_2, \mathcal{E}_2)$  dans l'espace mesurable  $(E_1, \mathcal{E}_1)$  telle que pour tout  $B_1 \in \mathcal{E}_1$ ,  $\Pi_{X_1}^{X_0, X_2}(X_0, X_2; B_1)$  est une version de  $P(X_1 \in B_1 | X_0, X_2)$ .

DÉMONSTRATION :

Soit  $B_1 \in \mathcal{E}_1$ . On note  $F_{(X_0, X_2)}$  la loi de  $(X_0, X_2)$ . La probabilité conditionnelle  $P(X_1 \in B_1 | (X_0, X_2) = \cdot)$  vérifie l'égalité  $F_{(X_0, X_2)}$ -p.s. suivante

$$P(X_1 \in B_1 | (X_0, X_2) = \cdot) = \frac{dQ}{dF_{(X_0, X_2)}} \quad (\text{C.5})$$

où  $Q$  est la mesure sur  $(E_0, \mathcal{E}_0) \otimes (E_2, \mathcal{E}_2)$  définie par, pour tout  $B \in \mathcal{E}_0 \otimes \mathcal{E}_2$ ,

$$Q(B) \stackrel{\text{déf}}{=} \mathbb{E}^P \{ \mathbf{1}_{B_1}(X_1) \mathbf{1}_B(X_0, X_2) \} \quad (\text{C.6})$$

et  $dQ/dF_{(X_0, X_2)}$  désigne la dérivée de Radon-Nikodym de  $Q$  par rapport à  $F_{(X_0, X_2)}$ .

Or, en notant  $F_{(X_0, X_1)}$  la loi de  $(X_0, X_1)$  et  $F_{(X_0, X_1, X_2)}$  la loi de  $(X_0, X_1, X_2)$ , on remarque que la mesure  $Q$  vérifie, pour tout  $B \in \mathcal{E}_0 \otimes \mathcal{E}_2$ ,

$$Q(B) = \int \mathbf{1}_{B_1}(x_1) \mathbf{1}_B(x_0, x_2) F_{(X_0, X_1, X_2)}(dx_0, dx_1, dx_2) \quad (\text{C.7})$$

$$= \int \left\{ \int \mathbf{1}_{B_1}(x_1) \mathbf{1}_B(x_0, x_2) \Pi_{X_2}^{(X_0, X_1)}(x_0, x_1; dx_2) \right\} F_{(X_0, X_1)}(dx_0, dx_1) \quad (\text{C.8})$$

$$= \int \left\{ \int \mathbf{1}_{B_1}(x_1) \mathbf{1}_B(x_0, x_2) f_{X_2}^{(X_0, X_1)}(x_0, x_1; x_2) \lambda(dx_2) \right\} F_{(X_0, X_1)}(dx_0, dx_1) \quad (\text{C.9})$$

Comme la mesure  $\lambda$  est  $\sigma$ -finie et comme la mesure  $F_{(X_0, X_1)}$  est une mesure de probabilité, on peut appliquer le théorème de Fubini à l'équation C.9. On obtient alors

$$Q(B) = \int \left\{ \int \mathbf{1}_{B_1}(x_1) \mathbf{1}_B(x_0, x_2) f_{X_2}^{(X_0, X_1)}(x_0, x_1; x_2) F_{(X_0, X_1)}(dx_0, dx_1) \right\} \lambda(dx_2) \quad (\text{C.10})$$

$$= \int \left\{ \int \left[ \int \mathbf{1}_{B_1}(x_1) \mathbf{1}_B(x_0, x_2) f_{X_2}^{(X_0, X_1)}(x_0, x_1; x_2) \Pi_{X_1}^{X_0}(x_0; dx_1) \right] F_{X_0}(dx_0) \right\} \lambda(dx_2) \quad (\text{C.11})$$

où  $F_{X_0}$  désigne la loi de  $X_0$ . Comme la mesure  $\lambda$  est  $\sigma$ -finie et comme la mesure  $F_{X_0}$  est une mesure de probabilité, on peut appliquer le théorème de Fubini à l'équation C.11. On obtient alors

$$Q(B) = \int \left[ \mathbf{1}_B(x_0, x_2) \int \mathbf{1}_{B_1}(x_1) f_{X_2}^{(X_0, X_1)}(x_0, x_1; x_2) \Pi_{X_1}^{X_0}(x_0; dx_1) \right] F_{X_0} \otimes \lambda(dx_0, dx_2) \quad (\text{C.12})$$

La mesure  $Q$  est donc absolument continue par rapport à la mesure  $F_{X_0} \otimes \lambda$  et sa densité par rapport à la mesure  $F_{X_0} \otimes \lambda$  vérifie l'égalité  $F_{X_0} \otimes \lambda$ -p.s. suivante

$$\frac{dQ}{dF_{X_0} \otimes \lambda}(x_0, x_2) = \int \mathbf{1}_{B_1}(x_1) f_{X_2}^{(X_0, X_1)}(x_0, x_1; x_2) \Pi_{X_1}^{X_0}(x_0; dx_1) \quad (\text{C.13})$$

De même, on montre que la mesure  $F_{(X_0, X_2)}$  est absolument continue par rapport à la mesure  $F_{X_0} \otimes \lambda$  et sa densité par rapport à la mesure  $F_{X_0} \otimes \lambda$  vérifie l'égalité  $F_{X_0} \otimes \lambda$ -p.s. suivante

$$\frac{dF_{(X_0, X_2)}}{dF_{X_0} \otimes \lambda}(x_0, x_2) = \int f_{X_2}^{(X_0, X_1)}(x_0, x_1; x_2) \Pi_{X_1}^{X_0}(x_0; dx_1) \quad (\text{C.14})$$

Comme les mesures  $F_{(X_0, X_2)}$  et  $F_{X_0} \otimes \lambda$  sont positives et  $\sigma$ -finies, on peut appliquer la proposition C.1 aux mesures  $Q$ ,  $F_{(X_0, X_2)}$  et  $F_{X_0} \otimes \lambda$ . Les équations C.5, C.13 et C.14 impliquent alors les égalités  $F_{(X_0, X_2)}$ -p.s. suivantes

$$P(X_1 \in B_1 | (X_0, X_2) = (x_0, x_2)) = \frac{dQ}{dF_{X_0} \otimes \lambda} / \frac{dF_{(X_0, X_2)}}{dF_{X_0} \otimes \lambda}(x_0, x_2) \quad (\text{C.15})$$

$$= \frac{\int \mathbf{1}_{B_1}(x_1) f_{X_2}^{(X_0, X_1)}(x_0, x_1; x_2) \Pi_{X_1}^{X_0}(x_0; dx_1)}{\int f_{X_2}^{(X_0, X_1)}(x_0, x_1; x_2) \Pi_{X_1}^{X_0}(x_0; dx_1)} \quad (\text{C.16})$$

et l'ensemble des  $(x_0, x_2) \in E_0 \times E_2$  tels que  $\int f_{X_2}^{(X_0, X_1)}(x_0, x_1; x_2) \Pi_{X_1}^{X_0}(x_0; dx_1) = 0$  est de  $F_{(X_0, X_2)}$ -mesure nulle. Par conséquent, l'application  $\Pi_{X_1}^{X_0, X_2}(\cdot; B_1)$  est mesurable et la variable aléatoire  $\Pi_{X_1}^{X_0, X_2}(X_0, X_2; B_1)$  est une version de  $P(X_1 \in B_1 | X_0, X_2)$ . Comme, pour tout  $(x_0, x_2) \in E_0 \times E_2$ ,  $\Pi_{X_1}^{X_0, X_2}(x_0, x_2; \cdot)$  est une mesure de probabilité, l'application  $\Pi_{X_1}^{(X_0, X_2)}$  est donc bien une transition de  $(E_0, \mathcal{E}_0) \otimes (E_2, \mathcal{E}_2)$  dans  $(E_1, \mathcal{E}_1)$  telle que pour tout  $B_1 \in \mathcal{E}_1$ ,  $\Pi_{X_1}^{X_0, X_2}(X_0, X_2; B_1)$  est une version de  $P(X_1 \in B_1 | X_0, X_2)$ .  $\square$

## C.2 Propriétés des transitions

### Proposition C.3

Soit  $(\Omega, \mathcal{F}, P)$  un espace de probabilité. Soient  $X_0, X_1$  et  $X_2$  des éléments aléatoires de l'espace mesurable  $(\Omega, \mathcal{F})$  dans, respectivement, les espaces mesurables  $(E_0, \mathcal{E}_0)$ ,  $(E_1, \mathcal{E}_1)$  et  $(E_2, \mathcal{E}_2)$ . On suppose que

1. il existe une transition  $\Pi_{X_1}^{X_0} : (E_0, \mathcal{E}_0) \prec (E_1, \mathcal{E}_1)$  telle que, pour tout  $B_1 \in \mathcal{E}_1$ , la variable aléatoire  $\Pi_{X_1}^{X_0}(X_0, B_1)$  sur  $(\Omega, \mathcal{F})$  est une version de la probabilité conditionnelle  $P(X_1 \in B_1 | X_0)$
2. il existe une transition  $\Pi_{X_2}^{(X_0, X_1)} : (E_0 \times E_1, \mathcal{E}_0 \otimes \mathcal{E}_1) \prec (E_2, \mathcal{E}_2)$  telle que, pour tout  $B_2 \in \mathcal{E}_2$ , la variable aléatoire  $\Pi_{X_2}^{(X_0, X_1)}((X_0, X_1), B_2)$  sur  $(\Omega, \mathcal{F})$  est une version de la probabilité conditionnelle  $P(X_2 \in B_2 | X_0, X_1)$ .

Pour tout  $x_0 \in E_0$  et pour tout  $B \in \mathcal{E}_1 \otimes \mathcal{E}_2$ , on pose

$$\Pi(x_0, B) = \int \left\{ \int \mathbf{1}_B(x_1, x_2) \Pi_{X_2}^{(X_0, X_1)}((x_0, x_1), dx_2) \right\} \Pi_{X_1}^{X_0}(x_0, dx_1) \quad (\text{C.17})$$

Alors l'application  $\Pi$  ainsi définie est une transition de l'espace mesurable  $(E_0, \mathcal{E}_0)$  dans l'espace mesurable  $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$  telle que, pour tout  $B \in \mathcal{E}_1 \otimes \mathcal{E}_2$ , la variable aléatoire  $\Pi(X_0, B)$  sur  $(\Omega, \mathcal{F})$  est une version de la probabilité conditionnelle  $P((X_1, X_2) \in B | X_0)$ .

DÉMONSTRATION :

On montre aisément que  $\Pi$  est une transition de l'espace mesurable  $(E_0, \mathcal{E}_0)$  dans l'espace mesurable  $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ . D'autre part, pour tout  $B \in \mathcal{E}_1 \otimes \mathcal{E}_2$ , on a

$$\Pi(X_0, B) = \Pi_{X_1}^{X_0}(\Pi_{X_2}^{(X_0, X_1)} \mathbf{1}_B)(X_0) \quad (\text{C.18})$$

$$= \mathbb{E}^P [\Pi_{X_2}^{(X_0, X_1)} \mathbf{1}_B(X_0, X_1) | X_0] \quad (P\text{-p.s.}) \quad (\text{C.19})$$

$$= \mathbb{E}^P [\mathbb{E}^P [\mathbf{1}_B(X_1, X_2) | X_0, X_1] | X_0] \quad (P\text{-p.s.}) \quad (\text{C.20})$$

Comme  $\sigma(X_0) \subset \sigma(X_0, X_1)$ , on en déduit

$$\Pi(X_0, B) = \mathbb{E}^P [\mathbf{1}_B(X_1, X_2) | X_0] \quad (P\text{-p.s.}) \quad (\text{C.21})$$

ce qui montre que la variable aléatoire  $\Pi(X_0, B)$  sur  $(\Omega, \mathcal{F})$  est une version de la probabilité conditionnelle  $P((X_1, X_2) \in B | X_0)$ .  $\square$

#### Proposition C.4

Soit  $(\Omega, \mathcal{F}, P)$  un espace de probabilité. Soient  $X_0, X_1$  et  $X_2$  des éléments aléatoires de l'espace mesurable  $(\Omega, \mathcal{F})$  dans, respectivement, les espaces mesurables  $(E_0, \mathcal{E}_0)$ ,  $(E_1, \mathcal{E}_1)$  et  $(E_2, \mathcal{E}_2)$ . On suppose que

1. les éléments aléatoires  $X_1$  et  $X_2$  sont  $P$ -indépendants conditionnellement à  $X_0$  ; autrement dit, les éléments aléatoires  $X_0, X_1$  et  $X_2$  vérifient

$$I_P(\{X_1\}, \{X_0\}, \{X_2\}) \quad (\text{C.22})$$

2. il existe une transition  $\Pi_{X_1}^{X_0} : (E_0, \mathcal{E}_0) \prec (E_1, \mathcal{E}_1)$  telle que, pour tout  $B_1 \in \mathcal{E}_1$ , la variable aléatoire  $\Pi_{X_1}^{X_0}(X_0; B_1)$  sur  $(\Omega, \mathcal{F})$  est une version de la probabilité conditionnelle  $P(X_1 \in B_1 | X_0)$ .
3. il existe une transition  $\Pi_{X_2}^{X_0} : (E_0, \mathcal{E}_0) \prec (E_2, \mathcal{E}_2)$  telle que, pour tout  $B_2 \in \mathcal{E}_2$ , la variable aléatoire  $\Pi_{X_2}^{X_0}(X_0; B_2)$  sur  $(\Omega, \mathcal{F})$  est une version de la probabilité conditionnelle  $P(X_2 \in B_2 | X_0)$ .

Pour tout  $x_0 \in E_0$  et pour tout  $B \in \mathcal{E}_1 \otimes \mathcal{E}_2$ , on pose

$$\Pi(x_0; B) = \Pi_{X_1}^{X_0}(x_0, \cdot) \otimes \Pi_{X_2}^{X_0}(x_0, \cdot)(B) \quad (\text{C.23})$$

Alors l'application  $\Pi$  ainsi définie est une transition de l'espace mesurable  $(E_0, \mathcal{E}_0)$  dans l'espace mesurable  $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$  telle que, pour tout  $B \in \mathcal{E}_1 \otimes \mathcal{E}_2$ , la variable aléatoire  $\Pi(X_0, B)$  sur  $(\Omega, \mathcal{F})$  est une version de la probabilité conditionnelle  $P((X_1, X_2) \in B | X_0)$ .

DÉMONSTRATION :

On montre aisément que  $\Pi$  est une transition de l'espace mesurable  $(E_0, \mathcal{E}_0)$  dans l'espace mesurable  $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ . D'autre part, pour tout  $B \in \mathcal{E}_1 \otimes \mathcal{E}_2$ , on a

$$\Pi(X_0, B) = \Pi_{X_2}^{X_0}(\Pi' \mathbf{1}_B)(X_0) \quad (\text{C.24})$$

où  $\Pi'$  est la transition de l'espace mesurable  $(E_0 \times E_2, \mathcal{E}_0 \otimes \mathcal{E}_2)$  dans l'espace mesurable  $(E_1, \mathcal{E}_1)$  définie par

$$\forall (x_0, x_2) \in E_0 \times E_2 \quad \forall B_1 \in \mathcal{E}_1 \quad \Pi'(x_0, x_2; B_1) \stackrel{\text{déf}}{=} \Pi_{X_1}^{X_0}(x_0; B_1) \quad (\text{C.25})$$

La transition  $\Pi'$  vérifie, pour tout  $B_1 \in \mathcal{E}_1$ ,

$$\Pi'(X_0, X_2; B_1) = \Pi_{X_1}^{X_0}(X_0; B_1) \quad (\text{C.26})$$

$$= P(X_1 \in B_1 | X_0) \quad (P\text{-p.s.}) \quad (\text{C.27})$$

Comme les éléments aléatoires  $X_1$  et  $X_2$  sont  $P$ -indépendants conditionnellement à  $X_0$ , on en déduit, en utilisant la proposition 1.1,

$$\Pi'(X_0, X_2; B_1) = P(X_1 \in B_1 | X_0, X_2) \quad (P\text{-p.s.}) \quad (\text{C.28})$$

Comme, pour tout  $B_2 \in \mathcal{E}_2$ , la variable aléatoire  $\Pi_{X_2}^{X_0}(X_0, B_2)$  sur  $(\Omega, \mathcal{F})$  est une version de la probabilité conditionnelle  $P(X_2 \in B_2 | X_0)$ , l'équation C.24 implique

$$\Pi(X_0; B) = \mathbb{E}^P [\Pi' \mathbf{1}_B(X_0, X_2) | X_0] \quad (P\text{-p.s.}) \quad (\text{C.29})$$

$$= \mathbb{E}^P \left[ \mathbb{E}^P [\mathbf{1}_B(X_1, X_2) | X_0, X_2] \middle| X_0 \right] \quad (P\text{-p.s.}) \quad (\text{C.30})$$

Comme  $\sigma(X_0) \subset \sigma(X_0, X_2)$ , on en déduit

$$\Pi(X_0, B) = \mathbb{E}^P [\mathbf{1}_B(X_1, X_2) | X_0] \quad (P\text{-p.s.}) \quad (\text{C.31})$$

ce qui montre que la variable aléatoire  $\Pi(X_0, B)$  sur  $(\Omega, \mathcal{F})$  est une version de la probabilité conditionnelle  $P((X_1, X_2) \in B | X_0)$ .  $\square$

## C.3 Propriétés des lois gaussiennes

### Convention C.1

Soit  $d$  un entier strictement positif. Soit  $K \in \mathcal{M}_{d,d}(\mathbb{C})$  une matrice hermitienne définie positive. Soit  $m \in \mathcal{M}_{d,1}(\mathbb{C})$ . On convient de noter  $g(m, K)$  la densité de la loi gaussienne circulaire symétrique  $\mathcal{N}_C(m, K)$  par rapport à la mesure de Lebesgue.

### Proposition C.5

Soient  $d_1$  et  $d_2$  des entiers strictement positifs. Soient  $K_1 \in \mathcal{M}_{d_1,d_1}(\mathbb{C})$  et  $K_2 \in \mathcal{M}_{d_2,d_2}(\mathbb{C})$  des matrices hermitiennes définies positives. Soient  $m_1 \in \mathcal{M}_{d_1,1}(\mathbb{C})$  et  $A \in \mathcal{M}_{d_1,d_2}(\mathbb{C})$ . Alors, pour tout  $x \in \mathcal{M}_{d_1,1}(\mathbb{C})$  et pour tout  $y \in \mathcal{M}_{d_2,1}(\mathbb{C})$ , on a

$$g(m_1, K_1)(x) g(Ax, K_2)(y) = \frac{1}{\det(\pi[K_2 + AK_1A^\dagger])} e^{-(m_1^\dagger K_1^{-1} m_1 + y^\dagger K_2^{-1} y - m_3^\dagger K_3^{-1} m_3)} \times \dots \dots \times g(m_3, K_3)(x) \quad (\text{C.32})$$

avec

$$m_3 = m_1 + \gamma(y - Am_1) \quad (\text{C.33})$$

$$K_3 = (I_{d_1} - \gamma A) K_1 \quad (\text{C.34})$$

$$\gamma = K_1 A^\dagger (K_2 + AK_1 A^\dagger)^{-1} \quad (\text{C.35})$$

où  $I_{d_1}$  désigne la matrice identité de  $\mathcal{M}_{d_1,d_1}(\mathbb{C})$ .



**Proposition C.6**

Soient  $d_1$  et  $d_2$  des entiers strictement positifs. Soient  $K_1 \in \mathcal{M}_{d_1, d_1}(\mathbb{C})$  et  $K_2 \in \mathcal{M}_{d_2, d_2}(\mathbb{C})$  des matrices hermitiennes définies positives. Soient  $m_1 \in \mathcal{M}_{d_1, 1}(\mathbb{C})$  et  $A \in \mathcal{M}_{d_1, d_2}(\mathbb{C})$ . Alors, pour tout  $y \in \mathcal{M}_{d_2, 1}(\mathbb{C})$ , on a

$$\int g(Ax, K_2)(y) \mathcal{N}_{\mathbb{C}}(m_1, K_1)(dx) = g(m_3, K_3)(y) \quad (\text{C.36})$$

avec

$$m_3 = Am_1 \quad (\text{C.37})$$

$$K_3 = AK_1A^\dagger + K_2 \quad (\text{C.38})$$

# Annexe D

## L'algorithme de propagation de croyance conditionnelle

Dans cette annexe, nous donnons une version conditionnelle de l'algorithme de propagation de croyance exacte décrite au § 3.3.

### D.1 Messages

#### Définition D.1

Soit  $(\Omega, \mathcal{F}, P)$  un espace de probabilité. Soit  $\mathcal{L}$  une partie finie de  $\mathbb{N}$ . Pour tout  $l \in \mathcal{L}$ , soit  $(E_l, \mathcal{E}_l)$  un espace mesurable et  $X_l$  un élément aléatoire de  $(\Omega, \mathcal{F})$  dans  $(E_l, \mathcal{E}_l)$ . On pose  $\mathbf{U} = \{X_l\}_{l \in \mathcal{L}}$ . Soit  $G = (\mathbf{U}, \mathbf{A})$  un réseau bayésien sur  $\mathbf{U}$  pour la mesure de probabilité  $P$ . Soit  $\mathbf{Z}$  une partie non-vide de  $\mathbf{U}$  telle que les éléments de  $\mathbf{Z}$  n'ont pas d'enfants dans  $G$ . Pour tout  $l \in \mathcal{L}$  tel que  $X_l \in \mathbf{Z}$ , soit  $\nu_l$  une mesure positive et  $\sigma$ -finie sur  $(E_l, \mathcal{E}_l)$ . On suppose que la mesure  $\nu_l$  domine la loi de  $X_l$ . Soit  $\mathbf{C}$  une partie non-vide de  $\mathbf{U} \setminus \mathbf{Z}$  telle qu'aucun sommet de  $\mathbf{C}$  ne possède de parent dans  $\mathbf{U} \setminus \mathbf{C}$ . On note  $\mathbf{C}$  le vecteur formé par les éléments de  $\mathbf{C}$  rangés par indice croissant,  $(\mathfrak{C}, \mathcal{C})$  l'espace image du vecteur  $\mathbf{C}$  et  $G'$  le sous-graphe de  $G$  induit par  $\mathbf{U} \setminus \mathbf{C}$ . Alors, pour tout arc  $(X, Y)$  de  $G'$ ,

1. on note  $\mathbf{Z}_{XY}^+$  l'ensemble des éléments de  $\mathbf{Z}$  qui sont joints à  $Y$  par une chaîne  $(\alpha_1, \dots, \alpha_q)$  de  $G'$  telle que  $\alpha_q = (X, Y)$ ,  $\mathbf{Z}_{XY}^+$  le vecteur formé par les éléments de  $\mathbf{Z}_{XY}^+$  rangés par indice croissant,  $\mathbf{C}_{XY}^+$  l'ensemble des éléments de  $\mathbf{C}$  qui ont un enfant joint à  $Y$  par une chaîne  $(\alpha_1, \dots, \alpha_q)$  de  $G'$  telle que  $\alpha_q = (X, Y)$ , et  $\mathbf{C}_{XY}^+$  le vecteur formé par les éléments de  $\mathbf{C}_{XY}^+$  rangés par indice croissant. On convient d'appeler **message de X à Y** et de noter  $\mu_{XY}^+$ 
  - la transition  $\Pi_X^{(\mathbf{C}_{XY}^+, \mathbf{Z}_{XY}^+)}$  si  $\mathbf{C}_{XY}^+$  et  $\mathbf{Z}_{XY}^+$  sont non vides,
  - la transition  $\Pi_X^{\mathbf{C}_{XY}^+}$  si  $\mathbf{C}_{XY}^+$  est non vide et  $\mathbf{Z}_{XY}^+$  est vide,
  - la transition  $\Pi_X^{\mathbf{Z}_{XY}^+}$  si  $\mathbf{C}_{XY}^+$  est vide et  $\mathbf{Z}_{XY}^+$  est non vide,
  - la mesure  $F_X$  si  $\mathbf{C}_{XY}^+$  et  $\mathbf{Z}_{XY}^+$  sont vides.
2. on note  $\mathbf{Z}_{XY}^-$  l'ensemble des éléments de  $\mathbf{Z}$  qui sont joints à  $X$  par une chaîne  $(\alpha_1, \dots, \alpha_q)$  de  $G'$  telle que  $\alpha_q = (X, Y)$ ,  $\mathbf{Z}_{XY}^-$  le vecteur formé par les éléments de  $\mathbf{Z}_{XY}^-$  rangés par indice croissant,  $(\mathfrak{Z}_{XY}^-, \mathcal{Z}_{XY}^-)$  l'espace image du vecteur  $\mathbf{Z}_{XY}^-$  et  $\nu_{XY}^-$  la

mesure produit sur  $(\mathfrak{Z}_{XY}^-, \mathcal{Z}_{XY}^-)$  définie par

$$\nu_{XY}^- = \bigotimes_{\substack{l \in \mathcal{L} \\ X_l \in \mathcal{Z}_{XY}^-}} \nu_l \quad (\text{D.1})$$

On note également  $C_{XY}^-$  l'ensemble des éléments de  $\mathcal{C}$  qui ont un enfant joint à  $X$  par une chaîne  $(\alpha_1, \dots, \alpha_q)$  de  $G'$  telle que  $\alpha_q = (X, Y)$ ,  $C_{XY}^-$  le vecteur formé par les éléments de  $C_{XY}^-$  rangés par indice croissant,  $(\mathfrak{C}_{XY}^-, C_{XY}^-)$  l'espace image du vecteur  $C_{XY}^-$  et  $(\mathfrak{X}, \mathcal{X})$  l'espace image de  $X$ . Alors,

- si  $C_{XY}^-$  et  $\mathcal{Z}_{XY}^-$  sont non vides, on suppose que, pour tout  $(c_{XY}^-, x) \in \mathfrak{C}_{XY}^- \times \mathfrak{X}$ , la mesure  $\Pi_{\mathcal{Z}_{XY}^-}^{(C_{XY}^-, X)}(c_{XY}^-, x; \cdot)$  admet une densité notée  $f_{\mathcal{Z}_{XY}^-}^{(C_{XY}^-, X)}(c_{XY}^-, x; \cdot)$  par rapport à la mesure  $\nu_{XY}^-$  et que l'application  $f_{\mathcal{Z}_{XY}^-}^{(C_{XY}^-, X)}$  est une application mesurable de  $(\mathfrak{C}_{XY}^-, C_{XY}^-) \otimes (\mathfrak{X}, \mathcal{X}) \otimes (\mathfrak{Z}_{XY}^-, \mathcal{Z}_{XY}^-)$  dans  $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ . On convient alors d'appeler **message de Y à X** et de noter  $\mu_{YX}^-$  l'application  $f_{\mathcal{Z}_{XY}^-}^{(C_{XY}^-, X)}$ ,
- si  $C_{XY}^-$  est vide et  $\mathcal{Z}_{XY}^-$  non vide, on suppose que, pour tout  $x \in \mathfrak{X}$ , la mesure  $\Pi_{\mathcal{Z}_{XY}^-}^X(x; \cdot)$  admet une densité notée  $f_{\mathcal{Z}_{XY}^-}^X(x; \cdot)$  par rapport à la mesure  $\nu_{XY}^-$  et que l'application  $f_{\mathcal{Z}_{XY}^-}^X$  est une application mesurable de  $(\mathfrak{X}, \mathcal{X}) \otimes (\mathfrak{Z}_{XY}^-, \mathcal{Z}_{XY}^-)$  dans  $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$ . On convient alors d'appeler **message de Y à X** et de noter  $\mu_{YX}^-$  l'application  $f_{\mathcal{Z}_{XY}^-}^X$ ,
- Si  $\mathcal{Z}_{XY}^-$  est vide, on convient d'appeler **message de Y à X** et de noter  $\mu_{YX}^-$  l'application mesurable constante de  $(\mathfrak{X}, \mathcal{X})$  dans  $(\mathbb{R}^+, \mathcal{B}(\mathbb{R}^+))$  valant 1.

Enfin, on convient d'appeler  $\mathcal{C}$  l'ensemble de **conditionnement**,  $\mathcal{Z}$  l'ensemble des **observations**,  $\mathcal{Z}_{XY}^+$  l'ensemble des **observations situées en amont de Y dans G' et reliées à Y via X** et  $\mathcal{Z}_{XY}^-$  l'ensemble des **observations situées en aval de X dans G' et reliées à X via Y**.

## D.2 Calcul de la loi *a posteriori* conditionnelle d'un sommet

### Proposition D.1

On suppose que le réseau bayésien  $G'$  ne possède pas de cycle et on considère un sommet  $X$  de  $G'$  n'appartenant pas à  $\mathcal{Z}$ . On note  $U_1, \dots, U_r$  les parents de  $X$  dans  $G'$ ,  $Y_1, \dots, Y_s$  les enfants de  $X$  dans  $G'$ ,  $C_X$  les parents de  $X$  dans  $G$  n'appartenant pas à  $G'$ ,  $C_X$  le vecteur formé par les éléments de  $C_X$  rangés par indice croissant et  $(\mathfrak{C}_X, C_X)$  l'espace image de  $C_X$ . Alors, la loi a posteriori de  $X$  étant donnée les observations  $\mathcal{Z}$  et conditionnellement à  $\mathcal{C}$  vérifie

$$\begin{aligned} \forall (c, z) \in \mathfrak{C} \times \mathfrak{Z} \quad \Pi_X^{(C, Z)}(c, z; \cdot) &= \Psi_{\text{Bayes}} \left( \prod_j \mu_{Y_j X}^-(c_{XY_j}^-, \cdot; z_{XY_j}^-), \dots \right. \\ &\quad \left. \dots \sum_{(U_1, \dots, U_r)} \left[ \bigotimes_i \mu_{U_i X}^+(c_{U_i X}^+, z_{U_i X}^+; \cdot) \right] \otimes \Pi_X^{(C_X, U_1, \dots, U_r)}(c_X, \cdot; \cdot) \right) \quad (\text{D.2}) \end{aligned}$$

avec les conventions suivantes

- $c_{XY_j}^-$  désigne le projeté de  $c$  sur l'espace  $\mathfrak{C}_{XY_j}^-$
- $c_{U_i X}^+$  désigne le projeté de  $c$  sur l'espace  $\mathfrak{C}_{U_i X}^+$

- $z_{XY_j}^-$  désigne le projeté de  $z$  sur l'espace  $\mathfrak{Z}_{XY_j}^-$
- $z_{U_iX}^+$  désigne le projeté de  $z$  sur l'espace  $\mathfrak{Z}_{U_iX}^+$
- lorsque  $X$  n'a pas de parent dans  $G'$ , l'expression D.2 devient

$$\Pi_X^{(C,Z)}(c, z; \cdot) = \Psi_{\text{Bayes}} \left( \prod_j \mu_{Y_jX}^-(c_{XY_j}^-, \cdot; z_{XY_j}^-), \Pi_X^{C_X}(c_X; \cdot) \right) \quad (\text{D.3})$$

- lorsque  $X$  n'a pas d'enfants, l'expression D.2 devient

$$\Pi_X^{(C,Z)}(c, z; \cdot) = \sum_{(U_1, \dots, U_r)} \left[ \bigotimes_i \mu_{U_iX}^+(c_{U_iX}^+, z_{U_iX}^+; \cdot) \right] \otimes \Pi_X^{(C_X, U_1, \dots, U_r)}(c_X, \cdot; \cdot) \quad (\text{D.4})$$

- lorsque l'ensemble  $C_X$  est vide, le terme  $\Pi_X^{(C_X, U_1, \dots, U_r)}(c_X, \cdot; \cdot)$  doit être remplacé par le terme  $\Pi_X^{(U_1, \dots, U_r)}$  dans les expressions D.2 et D.4, et le terme  $\Pi_X^{C_X}(c_X; \cdot)$  doit être remplacé par le terme  $F_X$  dans l'expression D.3
- lorsque l'ensemble  $\mathfrak{Z}_{XY_j}^-$  est vide, le terme  $\mu_{Y_jX}^-(c_{XY_j}^-, \cdot; z_{XY_j}^-)$  doit être remplacé par le terme  $\mu_{Y_jX}^-$  dans les expressions D.2 et D.3
- lorsque l'ensemble  $C_{U_iX}^+$  est non vide et l'ensemble  $\mathfrak{Z}_{U_iX}^+$  est vide, le terme  $\mu_{U_iX}^+(c_{U_iX}^+, z_{U_iX}^+; \cdot)$  doit être remplacé par le terme  $\mu_{U_iX}^+(c_{U_iX}^+; \cdot)$  dans les expressions D.2 et D.4
- lorsque l'ensemble  $C_{U_iX}^+$  est vide et l'ensemble  $\mathfrak{Z}_{U_iX}^+$  est non vide, le terme  $\mu_{U_iX}^+(c_{U_iX}^+, z_{U_iX}^+; \cdot)$  doit être remplacé par le terme  $\mu_{U_iX}^+(z_{U_iX}^+; \cdot)$  dans les expressions D.2 et D.4
- lorsque les ensembles  $C_{U_iX}^+$  et  $\mathfrak{Z}_{U_iX}^+$  sont vides, le terme  $\mu_{U_iX}^+(c_{U_iX}^+, z_{U_iX}^+; \cdot)$  doit être remplacé par le terme  $\mu_{U_iX}^+$  dans les expressions D.2 et D.4

DÉMONSTRATION :

La démonstration de cette proposition est similaire à celle de la proposition 3.1. □

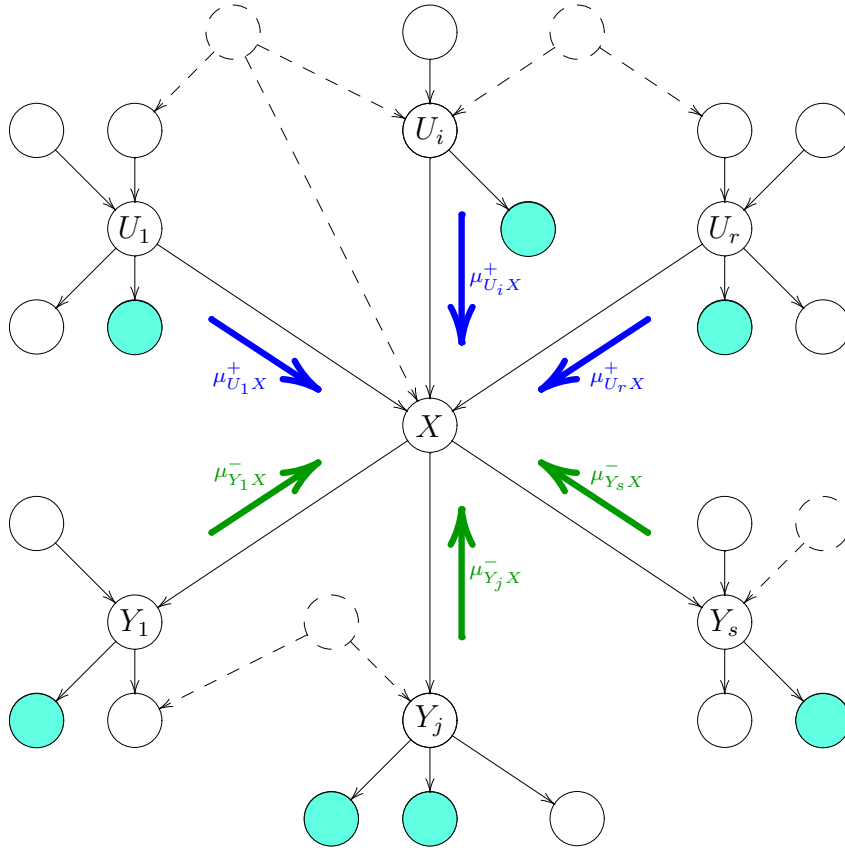


FIG. D.1 – Calcul de  $\Pi_X^{(C,Z)}$ . Les sommets de  $Z$  ont été représentés en vert. Les sommets de  $C$  et les arcs du graphe  $G$  qui n'appartiennent pas au graphe  $G'$  ont été représentés à l'aide de tirets.

### D.3 Calcul du message envoyé par un parent à un enfant

#### Proposition D.2

On suppose que le réseau bayésien  $G'$  ne possède pas de cycle et on considère un sommet  $X$  de  $G'$  n'appartenant pas à  $Z$ . On note  $U_1, \dots, U_r$  les parents de  $X$  dans  $G'$ ,  $Y_1, \dots, Y_s$  les enfants de  $X$  dans  $G'$ ,  $C_X$  les parents de  $X$  dans  $G$  n'appartenant pas à  $G'$ ,  $C_X$  le vecteur formé par les éléments de  $C_X$  rangés par indice croissant et  $(\mathfrak{C}_X, C_X)$  l'espace image de  $C_X$ . Alors, pour tout  $j_0 \in \{1, 2, \dots, s\}$  le message  $\mu_{XY_{j_0}}^+$  envoyé par  $X$  à  $Y_{j_0}$  vérifie

$$\begin{aligned} \forall (c_{XY_{j_0}}^+, z_{XY_{j_0}}^+) \in \mathfrak{C}_{XY_{j_0}}^+ \times \mathfrak{Z}_{XY_{j_0}}^+ \quad \mu_{XY_{j_0}}^+(c_{XY_{j_0}}^+, z_{XY_{j_0}}^+; \cdot) = \Psi_{\text{Bayes}} \left( \prod_{j \neq j_0} \mu_{Y_j X}^-(c_{XY_j}^-, \cdot; z_{XY_j}^-), \dots \right. \\ \left. \dots \sum_{(U_1, \dots, U_r)} [\bigotimes_i \mu_{U_i X}^+(c_{U_i X}^+, z_{U_i X}^+; \cdot)] \otimes \Pi_X^{(C_X, U_1, \dots, U_r)}(c_X, \cdot; \cdot) \right) \quad (\text{D.5}) \end{aligned}$$

avec les conventions suivantes :

- $c_{XY_j}^-$  désigne le projeté de  $c_{XY_{j_0}}^+$  sur l'espace  $\mathfrak{C}_{XY_j}^-$

- $c_{U_i X}^+$  désigne le projeté de  $c_{XY_{j_0}}^+$  sur l'espace  $\mathfrak{C}_{U_i X}^+$
- $z_{XY_j}^-$  désigne le projeté de  $z_{XY_{j_0}}^+$  sur l'espace  $\mathfrak{Z}_{XY_j}^-$
- $z_{U_i X}^+$  désigne le projeté de  $z_{XY_{j_0}}^+$  sur l'espace  $\mathfrak{Z}_{U_i X}^+$
- lorsque  $X$  a  $Y_{j_0}$  comme unique enfant et possède au moins un parent dans  $G'$ , l'expression 3.37 devient

$$\mu_{XY_{j_0}}^+(c_{XY_{j_0}}^+, z_{XY_{j_0}}^+; \cdot) = \sum_{(U_1, \dots, U_r)} \left[ \bigotimes_i \mu_{U_i X}^+(c_{U_i X}^+, z_{U_i X}^+; \cdot) \right] \otimes \Pi_X^{(C_X, U_1, \dots, U_r)}(c_X, \cdot; \cdot) \quad (D.6)$$

- lorsque  $X$  n'a pas de parent dans  $G'$  et possède au moins un enfant autre que  $Y_{j_0}$ , l'expression D.5 devient

$$\mu_{XY_{j_0}}^+(c_{XY_{j_0}}^+, z_{XY_{j_0}}^+; \cdot) = \Psi_{\text{Bayes}} \left( \prod_{j \neq j_0} \mu_{Y_j X}^-(c_{XY_j}^-, \cdot; z_{XY_j}^-), \Pi_X^{C_X}(c_X; \cdot) \right) \quad (D.7)$$

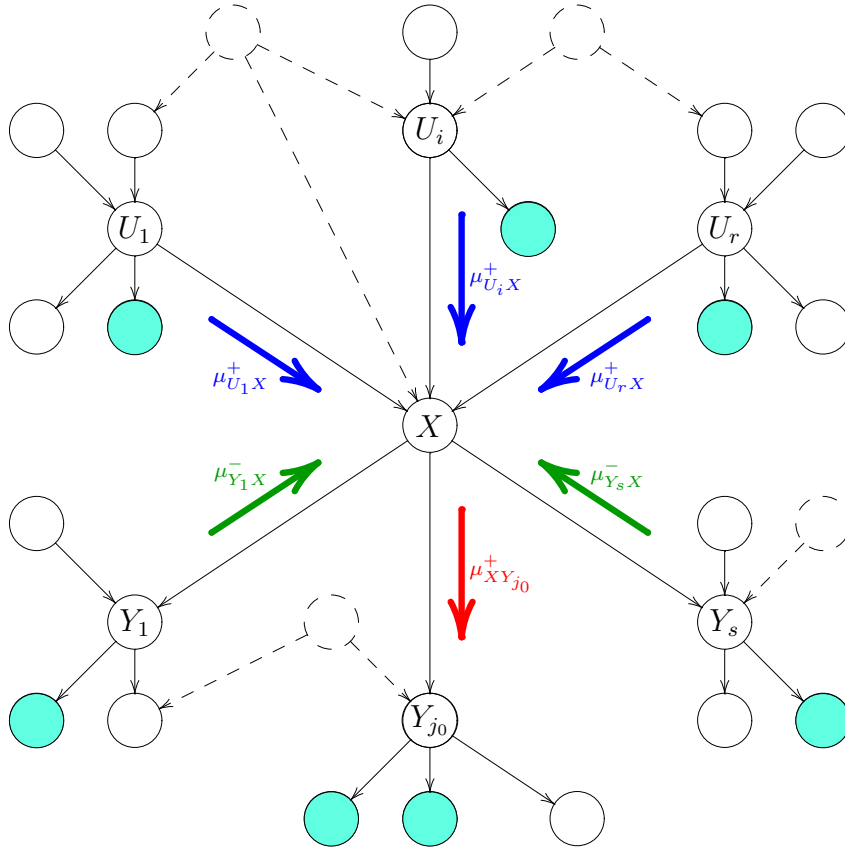
- lorsque  $X$  n'a ni parent dans  $G'$  ni d'autre enfant que  $Y_{j_0}$ , l'expression D.5 devient

$$\mu_{XY_{j_0}}^+ = \Pi_X^{C_X} \quad (D.8)$$

- lorsque l'ensemble  $C_X$  est vide, le terme  $\Pi_X^{(C_X, U_1, \dots, U_r)}(c_X, \cdot; \cdot)$  doit être remplacé par le terme  $\Pi_X^{(U_1, \dots, U_r)}$  dans les expressions D.5 et D.6, le terme  $\Pi_X^{C_X}(c_X; \cdot)$  doit être remplacé par le terme  $F_X$  dans l'expression D.7 et le terme  $\Pi_X^{C_X}$  doit être remplacé par le terme  $F_X$  dans l'expression D.8
- lorsque l'ensemble  $\mathfrak{Z}_{XY_j}^-$  est vide, le terme  $\mu_{Y_j X}^-(\cdot; z_{XY_j}^-)$  doit être remplacé par le terme  $\mu_{Y_j X}^-$  dans les expressions D.5 et D.7
- lorsque l'ensemble  $\mathfrak{C}_{U_i X}^+$  est non vide et l'ensemble  $\mathfrak{Z}_{U_i X}^+$  est vide, le terme  $\mu_{U_i X}^+(c_{U_i X}^+, z_{U_i X}^+; \cdot)$  doit être remplacé par le terme  $\mu_{U_i X}^+(c_{U_i X}^+; \cdot)$  dans les expressions D.5 et D.6
- lorsque l'ensemble  $\mathfrak{C}_{U_i X}^+$  est vide et l'ensemble  $\mathfrak{Z}_{U_i X}^+$  est non vide, le terme  $\mu_{U_i X}^+(c_{U_i X}^+, z_{U_i X}^+; \cdot)$  doit être remplacé par le terme  $\mu_{U_i X}^+(z_{U_i X}^+; \cdot)$  dans les expressions D.5 et D.6
- lorsque les ensembles  $\mathfrak{C}_{U_i X}^+$  et  $\mathfrak{Z}_{U_i X}^+$  sont vides, le terme  $\mu_{U_i X}^+(c_{U_i X}^+, z_{U_i X}^+; \cdot)$  doit être remplacé par le terme  $\mu_{U_i X}^+$  dans les expressions D.5 et D.6
- lorsque les ensembles  $\mathfrak{C}_{XY_{j_0}}^+$  ou  $\mathfrak{Z}_{XY_{j_0}}^+$  sont vides, le terme  $\mu_{XY_{j_0}}^+(c_{XY_{j_0}}^+, z_{XY_{j_0}}^+; \cdot)$  dans les expressions D.5, D.6 et D.7 doit être modifié de façon similaire.

DÉMONSTRATION :

La démonstration de cette proposition est similaire à celle de la proposition 3.2. □

FIG. D.2 – Calcul de  $\mu_{XY_{j_0}}^+$ .

## D.4 Calcul du message envoyé par un enfant à un parent

### Proposition D.3

On suppose que le réseau bayésien  $G'$  ne possède pas de cycle et on considère un sommet  $X$  de  $G'$  n'appartenant pas à  $\mathbf{Z}$ . On note  $U_1, \dots, U_r$  les parents de  $X$  dans  $G'$ ,  $Y_1, \dots, Y_s$  les enfants de  $X$  dans  $G'$ ,  $C_X$  les parents de  $X$  dans  $G$  n'appartenant pas à  $G'$ ,  $C_X$  le vecteur formé par les éléments de  $C_X$  rangés par indice croissant et  $(\mathfrak{C}_X, \mathcal{C}_X)$  l'espace image de  $C_X$ . Alors, pour tout  $i_0 \in \{1, 2, \dots, r\}$ , si  $X$  n'appartient pas à l'ensemble des observations  $\mathbf{Z}$ , le message  $\mu_{XU_{i_0}}^-$  envoyé par  $X$  à  $U_{i_0}$  vérifie

$$\forall (c_{U_{i_0}X}^-, z_{U_{i_0}X}^-) \in \mathfrak{C}_{U_{i_0}X}^- \times \mathfrak{Z}_{U_{i_0}X}^- \quad \mu_{XU_{i_0}}^-(c_{U_{i_0}X}^-, \cdot; z_{U_{i_0}X}^-) = \alpha \sum_X \left\{ \left[ \prod_j \mu_{Y_jX}^-(c_{XY_j}^-, \cdot; z_{XY_j}^-) \right] \times \dots \right. \\ \left. \dots \times \sum_{(U_i)_{i \neq i_0}} \left[ \bigotimes_{i \neq i_0} \mu_{U_iX}^+(c_{U_iX}^+, z_{U_iX}^+; \cdot) \right] \otimes \Pi_X^{(C_X, U_1, \dots, U_r)}(c_X, \cdot; \cdot) \right\} \quad (\text{D.9})$$

avec

$$\alpha = \prod_{i \neq i_0} f_{Z_{U_i X}^+}^{C_{U_i X}^+}(c_{U_i X}^+; z_{U_i X}^+) \quad (\text{D.10})$$

où  $f_{Z_{U_i X}^+}^{C_{U_i X}^+}(c_{U_i X}^+; \cdot)$  désigne la densité de la mesure  $\Pi_{Z_{U_i X}^+}^{C_{U_i X}^+}(c_{U_i X}^+; \cdot)$  par rapport à la mesure  $\nu_{U_i X}^+$  définie par

$$\nu_{U_i X}^+ = \bigotimes_{\substack{l \in \mathcal{L} \\ X_l \in \mathbf{Z}_{U_i X}^+}} \nu_l \quad (\text{D.11})$$

et avec les conventions suivantes :

- $c_{XY_j}^-$  désigne le projeté de  $c_{U_{i_0} X}^-$  sur l'espace  $\mathfrak{C}_{XY_j}^-$
- $c_{U_i X}^+$  désigne le projeté de  $c_{U_{i_0} X}^-$  sur l'espace  $\mathfrak{C}_{U_i X}^+$
- $z_{XY_j}^-$  désigne le projeté de  $z_{U_{i_0} X}^-$  sur l'espace  $\mathfrak{Z}_{XY_j}^-$
- $z_{U_i X}^+$  désigne le projeté de  $z_{U_{i_0} X}^-$  sur l'espace  $\mathfrak{Z}_{U_i X}^+$
- lorsque  $X$  n'a pas d'enfants et possède au moins un parent dans  $G'$  autre que  $U_{i_0}$ , l'expression D.9 devient

$$\mu_{XU_{i_0}}^-(c_{U_{i_0} X}^-, \cdot; z_{U_{i_0} X}^-) = \alpha \quad (\text{D.12})$$

- lorsque  $X$  a  $U_{i_0}$  comme unique parent dans  $G'$  et possède au moins un enfant, l'expression D.9 devient

$$\mu_{XU_{i_0}}^-(\cdot; z_{U_{i_0} X}^-) = \sum_X \left\{ \left[ \prod_j \mu_{Y_j X}^-(c_{XY_j}^-, \cdot; z_{XY_j}^-) \right] \cdot \Pi_X^{(C_X, U_{i_0})}(c_X, \cdot; \cdot) \right\} \quad (\text{D.13})$$

- lorsque l'ensemble  $\mathbf{Z}_{U_{i_0} X}^-$  est vide, l'expression D.9 devient

$$\mu_{XU_{i_0}}^- = 1 \quad (\text{D.14})$$

- lorsque l'ensemble  $\mathbf{C}_{U_{i_0} X}^-$  est vide et l'ensemble  $\mathbf{Z}_{U_{i_0} X}^-$  est non vide, le terme  $\mu_{XU_{i_0}}^-(c_{U_{i_0} X}^-, \cdot; z_{U_{i_0} X}^-)$  doit être remplacé par le terme  $\mu_{XU_{i_0}}^-(\cdot; z_{U_{i_0} X}^-)$  dans les expressions D.9 et D.13
- lorsque l'ensemble  $\mathbf{C}_X$  est vide, le terme  $\Pi_X^{(C_X, U_1, \dots, U_r)}(c_X, \cdot; \cdot)$  doit être remplacé par le terme  $\Pi_X^{(U_1, \dots, U_r)}$  dans l'expression D.9, et le terme  $\Pi_X^{(C_X, U_{i_0})}(c_X, \cdot; \cdot)$  doit être remplacé par le terme  $\Pi_X^{C_X}$  dans l'expression D.13
- lorsque l'ensemble  $\mathbf{C}_{XY_j}^-$  est vide et l'ensemble  $\mathbf{Z}_{XY_j}^-$  est non vide, le terme  $\mu_{Y_j X}^-(c_{XY_j}^-, \cdot; z_{XY_j}^-)$  doit être remplacé par le terme  $\mu_{Y_j X}^-(\cdot; z_{XY_j}^-)$  dans les expressions D.9 et D.13
- lorsque l'ensemble  $\mathbf{Z}_{XY_j}^-$  est vide, le terme  $\mu_{Y_j X}^-(c_{XY_j}^-, \cdot; z_{XY_j}^-)$  doit être remplacé par le terme  $\mu_{Y_j X}^-$  dans les expressions D.9 et D.13
- lorsque l'ensemble  $\mathbf{C}_{U_i X}^+$  est non vide et l'ensemble  $\mathbf{Z}_{U_i X}^+$  est vide, le terme  $\mu_{U_i X}^+(c_{U_i X}^+, z_{U_i X}^+; \cdot)$  doit être remplacé par le terme  $\mu_{U_i X}^+(c_{U_i X}^+; \cdot)$  dans l'expression D.9 et le terme  $f_{Z_{U_i X}^+}^{C_{U_i X}^+}(c_{U_i X}^+; z_{U_i X}^+)$  vaut 1 dans l'expression D.10



- lorsque l'ensemble  $C_{U_i X}^+$  est vide et l'ensemble  $Z_{U_i X}^+$  est non vide, le terme  $\mu_{U_i X}^+(c_{U_i X}^+, z_{U_i X}^+; \cdot)$  doit être remplacé par le terme  $\mu_{U_i X}^+(z_{U_i X}^+; \cdot)$  dans l'expression D.9 et le terme  $f_{Z_{U_i X}^+}^{C_{U_i X}^+}(c_{U_i X}^+; z_{U_i X}^+)$  doit être remplacé par le terme  $f_{Z_{U_i X}^+}(z_{U_i X}^+)$  dans l'expression D.10
- lorsque les ensembles  $C_{U_i X}^+$  et  $Z_{U_i X}^+$  sont vides, le terme  $\mu_{U_i X}^+(c_{U_i X}^+, z_{U_i X}^+; \cdot)$  doit être remplacé par le terme  $\mu_{U_i X}^+$  dans l'expression D.9 et le terme  $f_{Z_{U_i X}^+}^{C_{U_i X}^+}(c_{U_i X}^+; z_{U_i X}^+)$  vaut 1 dans l'expression D.10

Dans le cas où  $X$  appartient à  $Z$  et possède au moins un parent dans  $G'$  autre que  $U_{i_0}$ , le message  $\mu_{X U_{i_0}}^-$  vérifie<sup>1</sup>

$$\mu_{X U_{i_0}}^-(c_{U_{i_0} X}^-, \cdot; z_{U_{i_0} X}^-) = \dots \int \alpha \frac{d \left[ \sum_{(U_i)_{i \neq i_0}} \left[ \bigotimes_{i \neq i_0} \mu_{U_i X}^+(c_{U_i X}^+, z_{U_i X}^+; \cdot) \right] \otimes \Pi_X^{(C_X, U_1, \dots, U_r)}(c_X, \cdot; \cdot) \right]}{d\nu_l}(x) \quad (\text{D.15})$$

où  $x$  désigne le projeté de  $z_{U_{i_0} X}^-$  sur l'espace image de  $X$  et  $\nu_l$  est la mesure de domination sur l'espace image de  $X$ , et avec les mêmes conventions que précédemment lorsque les ensembles  $C_{U_{i_0} X}^-$ ,  $C_X$ ,  $C_{U_i X}^+$  ou  $Z_{U_i X}^+$  sont vides. Enfin, dans le cas où  $X$  appartient à  $Z$  et n'a pas d'autre parent dans  $G'$  que  $U_{i_0}$ , le message  $\mu_{X U_{i_0}}^-$  vérifie

$$\mu_{X U_{i_0}}^- = \begin{cases} f_X^{(C_X, U_{i_0})} & \text{si } C_X \neq \emptyset \\ f_X^{U_{i_0}} & \text{sinon} \end{cases} \quad (\text{D.16})$$

<sup>1</sup>On notera que lorsque  $X$  appartient à  $Z$ ,  $X$  n'a pas d'enfant car on a supposé que les éléments de  $Z$  n'ont pas d'enfant.

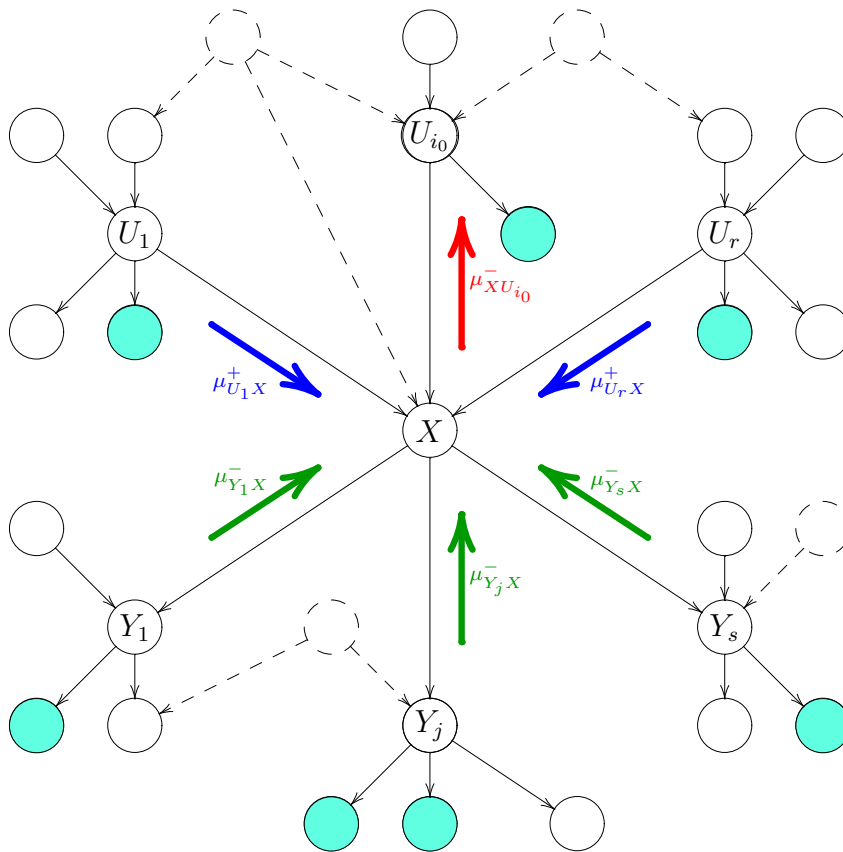


FIG. D.3 – Calcul de  $\mu_{XU_{i_0}}^-$ .

## D.5 Définition de l'algorithme de propagation de croyance conditionnelle

Lorsque le réseau bayésien  $G'$  est sans cycle, on peut calculer tous les messages envoyés à un sommet quelconque de  $G'$  (n'appartenant pas à l'ensemble des observations) à partir des messages envoyés par les sommets de degré 1<sup>2</sup> dans  $G'$  en appliquant de façon récursive les propositions D.2 et D.3. On peut alors calculer la loi a posteriori de ce sommet conditionnellement à  $C$  en appliquant la proposition D.1.

On convient d'appeler *algorithme de propagation de croyance conditionnelle* l'algorithme ci-dessus qui, à partir des messages envoyés par les sommets de degré 1 dans  $G'$ , permet de calculer la loi a posteriori d'un sommet quelconque de  $G'$  n'appartenant pas à l'ensemble des observations, conditionnellement à  $C$ .

<sup>2</sup>Les sommets de degré 1 dans  $G'$  sont les sommets qui, soit ont dans  $G'$  un unique parent et pas d'enfant, soit ont dans  $G'$  un unique enfant et pas de parent (cf la définition A.3).



# Bibliographie

- [AMGC02] M. ARULAMPALAM, S. MASKELL, N. GORDON & T. CLAPP – « A tutorial on particle filters for on line non-linear/non-gaussian bayesian tracking », *IEEE Trans. on Signal Processing* **50** (2002), p. 241–254. 79
- [BBGL97] C. BERZUINI, N. G. BEST, W. R. GILKS & C. LARIZZA – « Dynamic conditional independence models and Markov chain Monte Carlo methods », *Journal of the American Statistical Association* **92** (1997), p. 1403–1412. 75
- [BCJR74] L. BAHL, J. COCKE, F. JELINEK & J. RAVIV – « Optimal decoding of linear codes for minimizing symbol error rate », *IEEE Trans. Inform. Theory* **20** (1974), p. 284–287. 11, 59
- [BDI86] J. BIC, D. DUPONTEIL & J. IMBEAUX – *Éléments de communications numériques*, Dunod, 1986. 9
- [BGT93] C. BERROU, A. GLAVIEUX & P. THITIMAJSHIMA – « Near shannon limit error-correcting coding and decoding : Turbo-codes », *ICC 93* **2** (1993), p. 1064–1070. 9, 10
- [CWL00] R. CHEN, X. WANG & J. LIU – « Adaptive joint detection and decoding in flat-fading channels via mixture Kalman filtering », *IEEE Trans. Inform. Theory* **46** (2000), p. 2079–2094. 12, 109, 110, 111, 115, 137
- [DGA00] A. DOUCET, S. GODSILL & C. ANDRIEU – « On sequential Monte-Carlo sampling methods for bayesian filtering », *Statistics and Computing* **10** (2000), p. 197–208. 79, 111
- [DGK01] A. DOUCET, N. GORDON & V. KRISHNAMURTHY – « Particle filters for state estimation of jump Markov linear systems », *IEEE Trans. on Signal Processing* **49** (2001), p. 613–624. 103
- [DJB<sup>+</sup>95] C. DOUILLARD, M. JEZEQUEL, C. BERROU, A. PICART, P. DIDIER & A. GLAVIEUX – « Iterative correction of intersymbol interference : Turbo-equalization », *Europ. Trans. on Telecomm.* **6** (1995), p. 507–511. 9
- [Eyu88] M. EYUBOGLU – « Detection of coded modulation signals on linear, severely distorted channels using decision-feedback noise prediction with interleaving », *IEEE Trans. on Communications* **36** (1988), p. 401–409. 9
- [FBT99] D. FOX, W. BURGARD & S. THRUN – « Markov localization for mobile robots in dynamic environments », *Journal of Artificial Intelligence Research* **11** (1999), p. 391–427. 75

- [FM98] B. J. FREY & D. J. C. MACKAY – « A revolution : Belief propagation in graphs with cycles », in *Advances in Neural Information Processing Systems 10*, MIT Press, 1998. 10, 45
- [GM95] M. GONDRAN & M. MINOUX – *Graphes et algorithmes, 3<sup>e</sup> édition*, Éditions Eyrolles, 1995. 23
- [GSS93] N. GORDON, D. SALMOND & A. SMITH – « Novel approach to non-linear/non-gaussian bayesian state estimation », *Radar and Signal Processing, IEE-Proceedings-F* **140** (1993), p. 107–113. 75
- [HM54] J. HAMMERSLEY & K. MORTON – « Poor man’s Monte Carlo », *Journal of the Royal Statistical Society B* **16** (1954), p. 23–38. 75
- [IB96] M. ISARD & A. BLAKE – « Contour tracking by stochastic propagation of conditional density », in *European Conference on Computer Vision*, vol. 1, 1996, p. 343–356. 75
- [ICK94] M. IRWING, N. COX & A. KONG – « Sequential imputation for multilocus linkage analysis », *Proc. Natl. Acad. Sci. USA* **91** (1994), p. 11684–11688. 75
- [KF98] F. KSCHISCHANG & B. FREY – « Iterative decoding of compound codes by probability propagation in graphical models », *IEEE JSAC* **16** (1998), p. 219–230. 10, 50, 123
- [KLW94] A. KONG, J. S. LIU & W. WONG – « Sequential imputation method and bayesian missing data problems », *Journal of the American Statistical Association* **89** (1994), p. 278–288. 75
- [LC95] J. LIU & R. CHEN – « Blind deconvolution via sequential imputations », *Journal of the American Statistical Association* **90** (1995), p. 567–576. 75
- [LC98] — , « Sequential Monte-Carlo methods for dynamic systems », *Journal American Statistical Association* **93** (1998), p. 1032–1044. 82
- [LCL01] J. LIU, R. CHEN & T. LOGVINENKO – « A theoretical framework for sequential importance sampling and resampling », in *Sequential Monte Carlo Methods in Practice*, Springer-Verlag, New York, 2001, p. 225–245. 79
- [MMC98] R. J. MCELIECE, D. MACKAY & J.-F. CHENG – « Turbo decoding as an instance of Pearl’s belief propagation algorithm », *IEEE JSAC* **16** (1998), p. 140–152. 10, 45, 53, 59
- [MV98a] P. MESHKAT & J. D. VILLASENOR – « Generalized versions of turbo decoding in the framework of Bayesian networks and Pearl’s belief propagation algorithm », in *ICC’98*, vol. 1, 1998, p. 121–125. 123
- [MV98b] — , « New schedules for information processing in turbo decoding », in *ISIT’98*, 1998, p. 118. 123
- [Nev70] J. NEVEU – *Bases mathématiques du calcul des probabilités*, Masson, 1970. 11, 15
- [Pea82] J. PEARL – « Reverend Bayes on inference engines : a distributed hierarchical approach », in *Proc. AAAI National Conference on AI*, 1982, p. 133–136. 10, 53
- [Pea86] — , « Fusion, propagation, and structuring in belief networks », *Artificial Intelligence* **29** (1986), no. 3, p. 241–288. 10, 53

- [Pea88] —, *Probabilistic reasoning in intelligent systems*, Morgan-Kaufmann, 1988. 10, 11, 15, 21, 30, 53
- [Ric00] T. RICHARDSON – « The geometry of turbo-decoding dynamics », *IEEE Trans. Inform. Theory* **46** (2000), no. 1, p. 9–23. 53
- [RR55] M. ROSENBLUTH & A. ROSENBLUTH – « Monte Carlo calculation of the average extension of molecular chains », *Journal of Chemical Physics* **23** (1955), p. 356–359. 75
- [Shi96] A. N. SHIRYAEV – *Probability*, Springer-Verlag, 1996. 147
- [SP99] N. SHEPARD & M. K. PITT – « Filtering via simulation : auxiliary particle filter », *Journal of the American Statistical Association* **94** (1999), p. 590–599. 75
- [Tan81] R. M. TANNER – « A recursive approach to low complexity codes », *IEEE Trans. Inform. Theory* **27** (1981), no. 5, p. 533–547. 10
- [TJ02] S. TATIKONDA & M. I. JORDAN – « Loopy belief propagation and Gibbs measures », *Proc. Uncertainty in Artificial Intell.* **18** (2002), p. 493–50. 45, 53
- [TKS02] M. TUCHLER, R. KOETTER & A. SINGER – « Turbo equalization : principles and new results », *IEEE Trans. on Communications* **50** (2002), p. 754–767. 9
- [VADG02] J. VERMAAK, C. ANDRIEU, A. DOUCET & S. GODSILL – « Particle methods for Bayesian modeling and enhancement of speech signals », *IEEE Trans. on Speech and Audio Processing* **10** (2002), p. 173–185. 75
- [VP88] T. VERMA & J. PEARL – « Causal networks : Semantics and expressiveness », in *Proceedings, 4th Workshop on Uncertainty in Artificial Intelligence* (Minneapolis, MN), 1988, p. 352–359. 28
- [Wei00] Y. WEISS – « Correctness of local probability propagation in graphical models with loops », *Neural Computation* **12** (2000), no. 1, p. 1–41. 45, 52
- [WF01] Y. WEISS & W. T. FREEMAN – « Correctness of belief propagation in gaussian graphical models of arbitrary topology », *Neural Computation* **13** (2001), no. 10, p. 2173–2200. 53
- [Wib96] N. WIBERG – « Codes and decoding on general graphs », Thèse, Linköping University, Suède, 1996. 10, 59
- [WLK95] N. WIBERG, H.-A. LOELIGER & R. KÖTTER – « Codes and iterative decoding on general graphs », *Europ. Trans. Telecomm.* **6** (1995), p. 513–525. 10
- [YFW] J. S. YEDIDIA, W. T. FREEMAN & Y. WEISS – « Constructing free energy approximations and generalized belief propagation algorithms », Submitted to *IEEE Trans. Inform. Theory*. 10, 53
- [YFW01] —, « Generalized belief propagation », in *Advances in Neural Information Processing Systems 13* (T. K. Leen, T. G. Dietterich & V. Tresp, eds.), MIT Press, 2001, p. 689–695. 10, 53
- [ZPL90] K. ZHOU, J. PROAKIS & F. LING – « Decision-feedback equalization of time-dispersive channels with coded modulation », *IEEE Trans. on Communications* **38** (1990), p. 18–24. 9