Thèse Présentée à l'université Lille 1
En vue de l'obtention du grade de docteur de l'université Lille 1.

Distances matricielles dans la théorie des fonctions de croyance pour l'analyse et caractérisation des interactions entre les sources d'informations.

Par Mehena LOUDAHI

Soutenue le 01/12/2014, devant le jury composé de :

Didier DUBOIS	Président	Directeur de recherche CNRS	(61ème)	IRIT, Toulouse
Thierry DENŒUX	Rapporteur	Professeur des Universités	(61ème)	Heudiasyc, UTC
Arnaud MARTIN	Rapporteur	Professeur des Universités	(27ème)	IRISA, UR1
Sébastien DESTERCKE	Examinateur	Chercheur CNRS	(61ème)	Heudiasyc, UTC.
John Klein	Examinateur	Maitre de Conférences	(61ème)	Lagis, Lille1.
Jean-Marc Vannobel	Examinateur	Maitre de Conférences	(61ème)	Lagis, Lille1.
Olivier Colot	Examinateur	Professeur des Universités	(61ème)	Lagis, Lille1.

À la mémoire de :

Ma grand-mère,

Mon meilleur ami BELAID.

Remerciements

Le fruit de l'aventure dans des domaines scientifiques pointus et inexplorés ne peut être que surprises. Bonnes ou mauvaises, elles nous font grandir et nous apprennent à relever les défis et accomplir nos devoirs pour atteindre nos objectifs. Lorsque l'imagination et la créativité sont impliquées, la persévérance permet d'atteindre le recul nécessaire pour apporter quelques contributions novatrices et répondre aux problématiques posées au cours de l'aventure.

Une thèse de Doctorat n'est pas qu'un aboutissement. C'est aussi un parcourt, une expérience de vie assez riche en connaissances, en émotions et en souvenirs. C'est aussi un projet dont l'aboutissement dépasse de loin les termes du contrat doctoral.

Ce travail a été réalisé au sein de l'équipe "Signal et Image" du Laboratoire d'Automatique, Génie Informatique et Signal (Lagis UMR CNRS 8219) grâce à un financement ministériel. Ce financement fait l'objet d'un contrat doctoral entre 2011 et 2014 avec l'université Lille 1.

Je tiens à remercier les membres de mon jury d'avoir accepté d'évaluer mon travail et apporter des remarques et conseils constructifs ouvrant de nouvelles perspectives intéressantes à mes projets.

La thèse n'aurait toutefois pas pu s'être réalisée sans le soutien, la patience, les conseils et la confiance de mon directeur de thèse ; le professeur Olivier Colot, qui a cru en moi et m'a toujours bien conseillé, ainsi que mes co-encadrants le maître de conférences Jean-Marc Vannobel et le maître de conférences John Klein, pour tous leurs conseils et le temps qu'ils ont consacré afin de bien m'orienter durant la réalisation de mon projet de thèse.

Mes remerciements s'adressent particulièrement à John Klein, mon co-encadrant de thèse, pour sa présence et son dévouement ainsi que pour toute l'aide et la collaboration qu'il a généreusement apportées à mon travail.

Mes remerciement seront incomplets s'ils ne s'adressaient pas à ma famille qui m'a tant soutenu moralement et qui a toujours cru en moi et m'a encouragé pour aller de l'avant.

Mehena.

Résumé

Dans la théorie des fonctions de croyance, les distances sont des outils utiles pour l'approximation des fonctions de masse ou le clustering. Des approches efficaces sont proposées dans la littérature, spécialement les métriques complètes qui tiennent compte des interactions entre les éléments focaux. Dans cette thèse, un autre aspect particulier est examiné : la capacité de détecter une information commune en rapport à deux différents états de connaissance. Cette exigence, tout comme les deux autres mentionnées précédemment, sont formalisés sous forme de propriétés mathématiques. Afin de développer de nouvelles distances entre fonctions de croyance satisfaisant ces propriétés, des distances basées sur des normes entre les matrices de spécialisation dempsteriennes sont étudiées dans un premier temps. En effet, en utilisant les matrices de spécialisation dempsteriennes comme des représentations des fonctions de croyance, une distance est alors obtenue en calculant la norme de la différence entre ces matrices. N'importe quelle norme matricielle peut ainsi être utilisée pour définir une métrique complète. Il est prouvé que la distance du type L^1 basée sur les matrices de spécialisation dempsterienne réussit à satisfaire toutes les propriétés recherchées. Des liens intéressant et sans précédent entre la règle de combinaison conjonctive et cette distance sont démontrés.
En guise de généralisation de la nouvelle famille de distances, nous montrons aussi que d'autre matrices de croyance peuvent être utilisées pour évaluer une distance entre fonctions de croyance. Ces matrices sont les matrices d'α-spécialisation et d'α-généralisation qui sont étroitement liées aux règles de combinaison α-jonctives. Nous prouvons aussi que la distance basée sur la norme L^1 est consistante avec sa règle de combinaison α-jonctive correspondante. Par contre, la propriété structurelle n'est satisfaite que si $\alpha \in \left[\frac{1}{2}, 1\right]$ ou $\alpha = 0$. De plus, les α-jonctions sont des règles combinaisons dépendantes de la méta-connaissance [44]. Cette méta-connaissance inhérente aux α-jonctions est liée à la véracité des sources d'informations. Par conséquent, le comportement des distances entre fonctions de croyance est aussi analysé dans des situations diverses faisant intervenir une méta-connaissance incertaine ou partielle. Nous prouvons alors que, quand les deux sources mentent de la même façon, les distances basées sur la norme du type L^k ou d'opérateur 1 sont invariantes aux permutations des masses dues au mensonge des sources d'informations si $\alpha = 1$.
Plusieurs expériences de calcul des distances entre les fonctions de masse sont proposés et interprétés afin de permettre la compréhension des avantages et des limitations des nouvelles distances en comparaison aux autres distances existantes.

Abstract

Distances in evidence theory are useful tools for belief function approximation or clustering. Efficient approaches are found in the literature, especially full metrics taking focal element interactions into account. In this thesis, another aspect is investigated : the ability to detect common evidence pertaining to two different states of beliefs. This requirement, as well as the previously mentioned ones, are formalized through mathematical properties. To find a belief function distance satisfying the desired properties, matrix norms based distances between Dempsterian specialization matrices are investigated at first. Indeed, using specialization matrices as a representation of bodies of evidence, an evidential distance can be obtained by computing the norm of the difference of these matrices. Any matrix norm can be thus used to define a full metric. It is proved that the L^1 Dempsterian matrix distance succeeds to fulfil all requirements. Interesting and unprecedented ties between the conjunctive combination rule and this distance are demonstrated.

As a generalization of the newly introduced evidential distance familly, we also show that other matrices can be used to obtain new evidential distances. These matrices are the α-specialization and α-generalization matrices which are closely related to the α-junctive combination rules. We prove that any L^1 norm based distance thus defined turns out to be consistent with its corresponding α-junction. However, the structural property is satisfied if $\alpha \in \left[\frac{1}{2}, 1\right]$ or $\alpha = 0$. Furthermore, α-junctions are meta-data dependent combination rules [44]. The meta-data involved in α-junctions deals with the truthfulness of information sources. Consequently, the behavior of evidential distances is also analyzed in several situations involving uncertain or partial meta-knowledge about information source truthfulness. We then prove that, when two sources lie in the same way, distances relying on entry-wise norms or on the 1-operator norm are invariant to mass permutations pertaining to the lie if $\alpha = 1$.

Several mass function distance experiments are proposed and interpreted thereby allowing to understand advantages and limitations of the newly introduced distances as compared to existing ones.

Table des matières

Liste des tableaux

Table des figures

Notations mathématiques

Fonctions de croyance

$m,\ m_j\ m_j^\alpha$: fonction de masse qui peut être indicée indiquant par exemple la source j ou la fiabilité α.
$\overline{m}$: négation de la fonction de masse m.
$m(A)$: masse de croyance attribuée au sous-ensemble A.
m_A : fonction de masse catégorique totalement engagée sur A.
m_A^θ :fonction de masse à support simple.
S_j : source d'informations ou expert.
$\odot$: opérateur de combinaison dont on peur spécifier :
$\textcircled{\cap}$: combinaison conjonctive,
$\textcircled{\cup}$: combinaison disjonctive,
$\textcircled{\sqcup}$: combinaison disjonctive exclusive, sa fonction de combinaison duale est $\textcircled{\sqcap}$,
$\textcircled{\wedge}$: règle prudente de Denœux.
$\oplus$: règle de combinaison de Dempster.
$m_{12}, m_{1\odot 2}$: résultat d'une combinaison de deux fonctions de masse m_1 et m_2.
$m[B]$: conditionnement d'une fonction de masse m par le sous-ensemble B.
$\sqsubseteq_x$: relation d'ordre entre fonctions de croyance ; $m_1 \sqsubseteq_x m_2$ signifie que m_1 est plus engagée que m_2 au sens de l'argument x.
$Bel,\ Pl,\ q,\ b$: fonctions de croyance définies dans 2^Ω. Elle sont nommées dans l'ordre : crédibilité, plausibilité, communalité et implicabilité.
$BetP$: Probabilité pignistique.
κ : coefficient de conflit de Dempster.
a_n : auto-conflit résultant de la combinaison de n fois la même fonction de masse.
m^Ω : fonction de masse définie dans le cadre de discernement Ω.
$m^{\Omega\uparrow\Theta}$: extension vide de la fonction de croyance m de Ω vers Θ.

$m^{\Theta\downarrow\Omega}$: restriction de la fonction de croyance m de Θ vers Ω.
ε_Ω : espace vectoriel des fonctions de croyance.
$CP(m_1, m_2)$: coefficient de conflit entre m_1 et m_2.
$\textcircled{\cap}^\alpha$: α-conjonction.
$\textcircled{\cup}^\alpha$: α-disjonction.
g : fonction d'α-communalité.
h : fonction d'α-implicabilité.

Ensembles

$|\ .\ |$: cardinalité d'un ensemble.
$\Omega,\ \Theta$: cadre de discernement qui contient touts les hypothèses possibles à un problème
$\mathcal{M}^\Omega$: ensemble des fonctions de masse définies dans Ω.
$A,\ B,\ C,\ X,\ Y,\ A_i$: sous ensembles d'un cadre de discernement, ils représentent les élément focaux d'une fonction de masse et peuvent être indicés par i qui est l'ordre binaire.
$\overline{A}$: complémentaire du sous-ensemble A dans un cadre de discernement.
$\omega_i,\ \theta_i$: singleton du cadre de discernement Ω, respectivement Θ.
2^Ω : ensemble de toutes les disjonctions possibles de Ω (contient aussi l'ensemble vide $\emptyset$).
$\mathcal{F}_m$: noyau contenant les élément focaux de la fonction de masse m.
$S_x(m)$: ensemble des fonctions de masse plus riches que m au sens de l'ordre partiel $\sqsubseteq_x$.
$A\Delta B$: différence symétrique entre les ensembles A et B.

Fonctions diverses

$\langle m_1, m_2\rangle$: produit scalaire entre les fonctions de masse m_1 et m_2.
$d(m_1, m_2)$: distance entre les fonctions de masse m_1 et m_2.
d_{indice} : distance entre fonctions de croyance, l'*indice* indique le nom de la distance. On en cite par exemple :
d_J : distance de Jousselme.

d_D : distance de Diaz.
d_T : distance de Tessem.
d_{ZD} : distance de Zouhal et Denœux.
d_F : distance issue de la théorie des ensembles flous.
d_{IFi} : distance issue des ensembles flous intuitionnistes, l'indice $i \in \{1, 2, 3\}$.
$\mu,\ \nu$: fonctions d'appartenance et de non-appartenance floues.
d_{set} : distance d'ensemble.
1_{Cond} : fonction indicatrice dépendant de la condition $Cond$.
$\|.\|_W$: norme de type *Minkowski* pondérée par la matrice $\mathbf{W}$.
$d_W^{(p)}$: distance de *Minkowski* pondérée par la matrice $\mathbf{W}$ et $p \geq 1$ un entier non nul.
$\|.\|_k$: norme de type L^k.
$\|.\|_{opk}$: norme matriciel induite d'opérateur-k.
d_k : distance matricielle de type L^k entre fonctions de croyance.
d_{opk} : distance matricielle d'opérateur-k entre fonctions de croyance.
ρ : coefficient de normalisation des distances matricielles.
$O(.)$: ordre de complexité d'un calcul.
$d^{\cap}$ $d^{\cap,\alpha}$: distance d'α-spécialisation.
$d^{\cup}$ $d^{\cup,\alpha}$: distance d'α-généralisation.

Matrices

$Diag(\mathbf{v})$: matrice diagonale construite à partir du vecteur $\mathbf{v}$.
$\mathbf{M}$: matrice d'incidence utile pour transformer une fonction de masse en fonction de plausibilité.
$\mathbf{B}$: matrice d'implication utile pour transformer une fonction de masse en fonction d'implicabilité.
$\mathbf{M}_\alpha$: matrice d'$\alpha-$incidence.
$\mathbf{B}_\alpha$: matrice d'$\alpha-$implication.
$\mathbf{S}_i$: matrice de spécialisation dempsterienne d'indice de la fonction de masse m_i.
$\mathbf{Q}$: matrice de communalité sachant que $\mathbf{Q} = Diag(\mathbf{q})$.
$\mathbf{S}^\alpha,\ \mathbf{Q}^\alpha$: matrice issue d'une fonction de masse m affaiblie par un coefficient d'affaiblissement α.
$\mathbf{K},\ \mathbf{K}_i$: matrice de croyance relativement à une α-jonction indéfinie.
$\mathbf{K}^{\cap},\ \mathbf{K}^{\cup}$: matrice d'α-spécialisation, $\alpha-$généralisation.

Introduction générale

Souvent, on sous-entend par **distance évidentielle** tout critère destiné à être utilisé comme mesure de performance dans les algorithmes basés sur la théorie des fonctions de croyance (TFC). Dans un cadre formel, le terme "distance" est utilisé pour désigner un opérateur de mesure ayant des propriétés précises. Il est donc plus judicieux d'utiliser la terminologie **indice de dissimilarité** pour désigner la notion intuitive de différence entre fonctions de croyance. Au long de cette thèse nous nous appliquons à respecter au mieux ces différentes distinctions terminologiques.

Position du problème

La TFC, autrement appelée théorie de Dempster-Shafer [45], est un cadre formel largement utilisé dans les processus d'aide à la décision et de fusion d'informations. Elle permet de modéliser et de traiter des informations hétérogènes à la fois imprécises et/ou incertaines contrairement aux autres formalismes de traitement d'informations qu'elle généralise, à savoir, la théorie des probabilités, la théorie des possibilités et la théorie des ensembles de *Cantor*. Les approches s'appuyant sur la TFC traitent des données recueillies permettant de définir des fonctions d'ensemble décrivant l'état des connaissances d'une source d'informations sur un problème donné. Les ensembles de connaissances recueillis par chaque source individuelle sont appelés corps de preuves.
En cas de multiples sources d'informations, plusieurs approches ont mis en évidence la nécessité de comparer les éléments de preuve à l'aide d'un moyen pertinent. Par exemple, Zouhal et Denœux [59] utilisent un classifieur basé sur plus proches voisins où les paramètres intervenant dans leur approche sont optimisés en minimisant une dissimilarité évidentielle. Fixen et Mahler [14] introduisent une pseudo-métrique pour évaluer les performances de leur classifieur dont les sorties sont des fonctions de croyance. Dans [13] et [53], les données d'un capteur sont représentées comme des fonctions de croyance et la fiabilité du capteur est évaluée à l'aide de distances. Une difficulté, commune lorsque l'on travaille dans le cadre TFC, est l'approximation d'un élément de preuves pour obtenir un autre avec les propriétés désirées ou pour réduire la charge de calcul. Un moyen pratique pour effectuer ces approximations est de recourir à des distances comme proposé dans [1, 55, 11] ou [4].
De récentes recherches sur les indices dissimilarité entre fonctions de croyance ont été développées par la communauté de la TFC au cours des dernières années. Un état de l'art élaboré est proposé par Jousselme et Maupin dans [21]. Ce travail a permis de poser un certain nombre de concepts et de clarifier le comportement et les corrélations entre les différents indicateurs présents dans la littérature. Néanmoins, la définition de dissimilarités entre fonctions de croyance demeure encore un problème ouvert. Aucune définition formelle des propriétés, autres que celles d'une métrique, que devrait satisfaire une dissimilarité entre fonctions de croyance n'est encore établie. La structure en

treillis des informations traitées dans le cadre de la TFC, qui induit une forme d'interactions entre les éléments focaux, et l'influence du degré d'ignorance sont des contraintes principales qui doivent être prises en compte par une distance évidentielle. Définir une distance pertinente entre fonctions de croyance est une tâche difficile par le fait que deux aspects doivent être évalués conjointement dans une seule valeur, à savoir, l'incertitude (à la manière de l'entropie dans un cadre probabiliste) et l'imprécision (à la manière de la non-spécificité dans la théorie des possibilités).
D'autre part, l'agrégation des informations dans le cadre de la TFC est principalement basée sur deux opérateurs de fusion : l'opérateur conjonctif dans le cas où les sources sont fiables et distinctes et l'opérateur disjonctif dans le cas où l'on sait qu'une des sources au moins est fiable. Le processus de fusion permet d'apporter des informations supplémentaires à des éléments de preuve, ce qui a inévitablement un impact sur les distances évidentielles entre ces éléments après combinaison. Les interactions entre ces distances et les opérateurs de combinaison demeurent encore inexplorées dans la littérature.

Contribution

Tout d'abord, nous avons identifié trois principes auxquels il apparaît justifié qu'une distance évidentielle se conforme :

- (i) : deux fonctions de masse sont complètement similaires si et seulement si elles représentent des éléments de preuves identiques au regard de la solution au problème donné.
- (ii) : une distance évidentielle doit respecter les propriétés structurelles introduites dans [21] dans le but de prendre en compte les dépendances entre les ensembles de solution.
- (iii) : deux fonctions de masse qui soutiennent des croyances communes devraient être plus proches que des fonctions de masse qui s'engagent sur des hypothèses disjointes.

Avant de procéder à la définition de nouvelles distances évidentielles, nous nous sommes focalisés sur la formalisation de ces principes sous forme de propriétés mathématiques. La première est une propriété de métrique, impliquant donc que les mesures de dissimilarité que nous définissons doivent être des métriques complètes. Pour le second principe, une propriété dite **propriété structurelle** est proposée. Enfin, la définition mathématique apportée au principe des croyances communes est appelée **propriété de consistance** avec une règle de combinaison.

Distances de spécialisation

Vu l'intérêt des matrices de spécialisation dans un processus de combinaison conjonctive, celles-ci constituent des éléments essentiels dans la représentation des connaissances fournies par les sources d'informations. En effet, une matrice de spécialisation est en correspondance bijective avec la fonction de croyance qui lui correspond et contient tous les conditionnements possibles de celle-ci. L'idée est donc d'exploiter ces matrices dans le calcul d'une distance entre les fonctions de croyance.
Nous avons défini dans [32] une famille de distances issue des normes matricielles entre les matrices de spécialisation représentant les fonctions de croyance. L'étude de cette famille de métriques est détaillée dans le chapitre 3. Les aspects de consistance avec la règle conjonctive sont considérés puisque ces matrices interviennent dans le calcul de la fusion conjonctive. Les différents conditionnements contenus dans ces matrices tiennent compte de la structure en treillis de l'espace des fonctions de croyance.
Un algorithme de calcul rapide [31] est développé afin d'améliorer le temps de calcul de la classe des distances L^p qui est une sous-famille de la famille des distances de spécialisation introduites dans cette thèse. Initialement, ces distances sont calculées avec une complexité de $O\left(N^3\right)$ [1]. Grâce à cet algorithme nous pouvons réduire cette complexité à $O\left(N^{1.58}\right)$ et gagner en temps de calcul. Cet algorithme permet à ces distances d'être très compétitives en terme de temps de calcul en plus de leurs propriétés mathématiques et structurelles très attractives.

Généralisation aux $\alpha-$jonctions

Les $\alpha-$jonctions sont une famille de règles de combinaison linéaires généralisant les opérateurs marginaux qui sont la combinaison conjonctive et la combinaison disjonctive [51]. Tout d'abord, nous remarquons que la manipulation de ces règles dans l'espace vectoriel des fonctions de croyance se fait à travers des opérateurs linéaires faisant intervenir des matrices dépendant d'un paramètre α et dites **matrices de croyance**. Ces matrices sont en effet les matrices d'$\alpha-$spécialisation et d'$\alpha-$généralisation qui sont étroitement liées aux règles de combinaison $\alpha-$jonctives.
Comme nous l'avons vu dans [23], il est tout à fait possible d'utiliser les matrices d'$\alpha-$spécialisation et d'$\alpha-$généralisation pour définir une nouvelle famille de distances évidentielles basées sur les normes matricielles et qui généralise les distances de spécialisation. En plus des aspects structurels, les aspects de consistance des distances basées sur les matrices $\alpha-$jonctives avec les règles de combinaison qui leur sont relatives sont aussi explorés.
Les $\alpha-$jonctions sont interprétées comme des règles de combinaison qui tiennent compte des méta-informations liées à la véracité des sources d'informations comparées. L'in-

1. N est un entier tel que $N = 2^n$ avec n le nombre de solutions possibles au problème

fluence de ces méta-informations est aussi abordée dans des cas impliquant une méta-connaissance incertaine ou partielle.

Organisation

A la suite de cette introduction, notre travail est organisé comme suit :

- Le chapitre 1 est dédié aux concepts fondamentaux de la théorie des fonctions de croyance. Dans ce chapitre, une section est dédiée aux représentations matricielles des fonctions de croyance ; une attention particulière est consacrée aux matrices de spécialisation dempstériennes.
- Les notions de dissimilarité et de distances évidentielles sont abordées dans le chapitre 2. Dans ce chapitre, les indices de dissimilarité les plus connus entre fonctions de croyance sont présentés et répertoriés en quatre catégories.
- Une nouvelle famille de distances evidentielles basée sur les matrices de spécialisation est définie dans le chapitre 3. Les propriétés de cette famille de distances y sont aussi étudiées.
- Dans le chapitre 4, nous présentons une généralisation de la nouvelle famille de distances matricielles aux matrices issues des $\alpha-$jonctions. En plus des propriétés de ces distances, un autre aspect lié à leur nature en lien avec les méta-informations est abordé dans ce chapitre.

Nous finissons ce rapport avec une conclusion générale et quelques pistes pour des perspectives futures du travail.

Chapitre 1

La théorie des fonctions de croyance

Sommaire

La théorie des fonctions de croyance (TFC) offre un cadre formel pour la gestion des informations imparfaites. C'est un cadre qui modélise efficacement les données provenant de sources d'informations incertaines et/ou imprécises. La théorie des fonctions de croyance est un cadre général incluant la théorie des probabilités ainsi que la théorie des ensembles flous. Cette théorie est initiée en 1967 par Dempster puis développée dans les travaux de Shafer en 1976 [45] sous le nom de théorie de Dempster-Shafer.

Parmi les domaines applicatifs privilégiés de la TFC, on distingue la fusion d'informations qui est une étape cruciale dans le processus de prise de décision. La fusion d'informations permet en effet l'agrégation de connaissances incertaines et/ou imprécises dans l'objectif d'atteindre des niveaux informationnels plus élevés. Traditionnellement, l'outil de combinaison de base développé dans le cadre de cette théorie est la règle de combinaison orthogonale de Dempster. Son utilisation impose le respect de la contrainte d'indépendance des sources d'informations à combiner. Néanmoins, lors d'une combinaison, un conflit apparaît fréquemment. Ce dernier est principalement dû à des mesures aberrantes liées aux sources d'informations ainsi qu'à leur fiabilité. Une autre cause possible du conflit peut être aussi liée à l'approximation dans la modélisation utilisée pour traduire les données en fonctions de croyance. Comme nous le verrons au chapitre 2, cette notion de conflit peut être vue comme un indice de dissimilarité entre fonctions de croyance.

1.1 Concepts fondamentaux

1.1.1 Les fonctions de croyance

Le principe fondamental de la théorie des fonctions de croyance est inspiré de la théorie des probabilités supérieures et inférieures. Il est donc question de modéliser la croyance en un évènement en prenant en compte l'incertitude et l'imprécision de l'information acquise. Cette modélisation est basée sur des fonctions dites **fonctions de croyance** définies sur des sous-ensembles d'un ensemble global et qui prennent des valeurs dans l'intervalle $[0, 1]$.
La modélisation par fonctions de croyance repose principalement sur la définition préalable d'un ensemble appelé **cadre de discernement** contenant toutes les hypothèses singletons solutions potentielles du problème. Il est généralement noté $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ et contient un nombre fini n d'hypothèses exhaustives et exclusives qui sont les solutions au problème posé. Un exemple typique est celui où Ω est le domaine fini des valeurs que peut prendre un paramètre ω_0 inconnu. La condition d'exhaustivité du cadre de discernement est aussi appelée, dans le modèle de Shafer, "condition du monde fermé"

(*closed world*) impliquant la non-existence d'hypothèses solution au problème en dehors de l'ensemble Ω. Si cette condition n'est pas observée, on parlera alors de l'hypothèse du monde ouvert (*open world*) [48]. Il est toujours possible de garder l'hypothèse d'exhaustivité du cadre de discernement, en introduisant explicitement dans Ω un élément supplémentaire qui représente toute autre solution méconnue dans le cadre de discernement initial.
Bien que les hypothèses soient exclusives, tous les états de croyance, susceptibles d'être imprécis, ne sont pas représentables uniquement par les hypothèses singletons contenus dans Ω. On définit alors l'ensemble 2^Ω qui contient toutes les disjonctions possibles de Ω (ainsi que l'ensemble vide $\emptyset$) qui permettent de traduire des connaissances imprécises. L'ensemble 2^Ω est en effet appelé **ensemble des parties de** Ω. Il contient $N = 2^n$ éléments et il est défini comme suit :

$$2^\Omega = \{A | A \subseteq \Omega\}. \tag{1.1}$$

L'appellation **fonctions de croyance** est un terme générique qui fait référence à toutes les fonctions utilisées pour modéliser l'information dans la théorie de Dempster-Shafer. Ces fonctions sont les fonctions de masse m, les fonctions de crédibilité Bel et les fonctions de plausibilité Pl. Elles sont en correspondance bijective les unes par rapport aux autres et elles modélisent toutes le même état d'informations.
D'autres fonctions sont aussi définies dans la TFC. Il s'agit de la fonction de communalité q et de la fonction d'implicabilité b. Ces deux fonctions n'ont pas d'interprétation concrète dans la modélisation basée sur la TFC contrairement aux autres fonctions de croyance. Elles sont souvent utiles à des fins calculatoires et algorithmiques.

1.1.1.1 Fonctions de masse

L'élément fondamental de la modélisation dans la théorie des fonctions de croyance est la fonction de masse. Celle-ci est définie comme suit :

Définition 1. *Une* ***fonction de masse*** m_j *associée à une source d'informations* $\mathcal{S}_j$ *prend des valeurs dans l'intervalle* $[0, 1]$. *Elle est définie comme suit :*

$$m_j : 2^\Omega \longrightarrow [0, 1],$$

$$A \longrightarrow m_j(A),$$

sachant que :

$$\sum_{A \subseteq \Omega} m_j(A) = 1. \tag{1.2}$$

La masse $m_j(A)$ représente la partie du degré de croyance exactement placé dans l'hypothèse A. Cette croyance ne peut pas être affectée aux différents sous-ensembles

de A compte tenu de l'état actuel de la connaissance. Ceci devient possible sous réserve d'un nouvel apport de connaissances supplémentaires justifiant la spécification.

Définition 2. *Etant donné une fonction de masse m, les définitions suivantes sont essentielles dans la théorie des fonctions de croyance :*

- ***Éléments focaux et noyau***
  Les sous-ensembles A dont la masse est non nulle sont appelés ***éléments focaux*** *de la fonction de masse m. L'ensemble : $\mathcal{F}_m = \{A \subseteq \Omega / m(A) > 0\}$ est appelé* ***noyau*** *de la fonction de masse m.*
- ***Fonction de masse normale***
  Une fonction de masse est dite ***normale*** *si l'ensemble vide noté $\emptyset$ n'est pas un élément focal, $m(\emptyset) = 0$. Sinon elle est dite sous-normale. L'obtention d'une fonction de masse normale ${}_*m$ à partir d'une fonction de masse sous-normale m se fait par le processus de normalisation suivant :*

$$\begin{cases} {}_*m(A) = \frac{m(A)}{1-m(\emptyset)}, \ \forall A \subseteq \Omega, \ A \neq \emptyset, \\ {}_*m(\emptyset) = 0. \end{cases} \tag{1.3}$$

  Si $m(\emptyset) = 1$, cette transformation est simplement impossible. La condition $m(\emptyset) = 0$ est initialement imposée dans le modèle de Shafer mais elle est levée dans le modèle des connaissances transférables introduit par Smets [48].
- ***Fonction de masse dogmatique***
  Une fonction de masse est dite ***dogmatique*** *si Ω n'est pas un élément focal.*
- ***Fonction de masse simple***
  Une fonction de masse est dite ***simple*** *si elle a au plus deux éléments focaux incluant Ω. Elle est, par ailleurs, non dogmatique. Celle-ci est notée m_A^θ et elle est définie $\forall\ \theta \in [0,1]$ comme suit :*

$$m_A^\theta(X) = \begin{cases} 1-\theta & si\ X = A \subset \Omega, \\ \theta & si\ X = \Omega, \\ 0 & sinon. \end{cases} \tag{1.4}$$

- ***Fonction de masse catégorique***
  Une fonction de masse est dite ***catégorique*** *si elle possède un seul élément focal. Ceci implique que $|\mathcal{F}_m| = 1$ avec $|\mathcal{F}_m|$ le cardinal de $\mathcal{F}_m$. Autrement dit :*
  $\exists A \subseteq \Omega$ tel que $m(A) = 1$ et $m(B) = 0\ \forall B \neq A$. Cette fonction est notée m_A.
  Une fonction de masse catégorique est dite certaine car $m(A) = 1$ mais elle n'est précise que si A est un singleton, i.e. $|A| = 1$.
- ***Fonction de masse vide***

La fonction de masse vide m_Ω est une fonction qui admet pour unique élément focal le cadre de discernement Ω. Une fonction de masse vide correspond à l'ignorance totale impliquant que la source est certaine et que la solution se trouve dans le cadre de discernement Ω mais totalement incapable de la désigner avec précision.

- ***Fonction de masse bayésienne***
  Une fonction de masse est dite bayésienne si tous ses éléments focaux sont des singletons ; $\forall A \in \mathcal{F}_m, |A| = 1$. Une fonction de masse bayésienne peut alors être assimilée à une distribution de probabilités.
- ***Négation d'une fonction de masse***
  Selon Dubois et Prade, la négation d'une fonction de masse m est définie dans un cadre de discernement par [52] :

$$\overline{m}(A) = m(\overline{A}), \quad \forall A \subseteq \Omega. \tag{1.5}$$

où $\overline{m}$ est la négation de m et l'ensemble $\overline{A}$ est le complémentaire de A dans Ω.

1.1.1.2 Fonctions de crédibilité, de plausibilité, de communalité et d'implicabilité

Dans la théorie des fonctions de croyance, le même état de connaissance que celui modélisé par la fonction de masse m peut aussi être représenté par d'autres fonctions associées et équivalentes. Celles-ci sont la fonction de **crédibilité** Bel, la fonction de **plausibilité** Pl, la fonction de **communalité** q ainsi que la fonction **d'implicabilité** b. Elles sont respectivement définies par les équations suivantes $\forall A \subseteq \Omega$:

$$Bel(A) = \sum_{B \subseteq A, B \neq \emptyset} m(B), \tag{1.6}$$

$$Pl(A) = \sum_{\substack{B \in 2^\Omega \\ B \cap A \neq \emptyset}} m(B), \tag{1.7}$$

$$q(A) = \sum_{\substack{B \in 2^\Omega \\ B \supseteq A}} m(B), \tag{1.8}$$

$$b(A) = \sum_{\substack{B \in 2^\Omega \\ B \subseteq A}} m(B). \tag{1.9}$$

La quantité $Bel(A)$ représente la quantité d'informations exactement considérée dans l'hypothèse A. Elle intègre l'ensemble des croyances apportées aux éléments de cette disjonction. Nous pouvons facilement vérifier dans l'équation (1.6) que $Bel(\emptyset) = 0$ et que $Bel(\Omega) = 1$.

$Pl(A)$ peut être interprétée comme le degré de non-croyance en l'hypothèse complémentaire de A. La plausibilité est alors reliée à la crédibilité via la formule :

$$Pl(A) = 1 - m(\emptyset) - Bel(\overline{A}), \forall A \neq \emptyset$$

où $\overline{A}$ est le complémentaire de A dans Ω.
La fonction de plausibilité vérifie aussi les deux propriétés suivantes : $Pl(\emptyset) = 0$ et $Pl(\Omega) = 1 - m(\emptyset)$.
Les fonctions de communalité q et d'implicabilité b sont essentiellement utilisées pour simplifier les calculs engendrés par la combinaison des croyances. La fonction de communalité vérifie les propriétés : $q(\emptyset) = 1$ et $q(\Omega) = m(\Omega)$.

On remarque aussi que la communalité de la fonction de masse $\overline{m}$ est :

$$\begin{aligned} \overline{q}(A) &= \sum_{B \supseteq A} \overline{m}(B), \\ &= \sum_{B \supseteq A} m(\overline{B}), \\ &= \sum_{\overline{B} \subseteq \overline{A}} m(\overline{B}), \\ \Rightarrow \overline{q}(A) &= b(\overline{A}). \end{aligned}$$

De la même manière, on peut obtenir :

$$\overline{b}(A) = q(\overline{A}).$$

Il est essentiel de noter qu'une fonction de masse et ses différentes représentations associées, à savoir la crédibilité, la plausibilité, la communalité ainsi que l'implicabilité, sont toutes des fonctions en correspondance bijective les unes par rapport aux autres. Il existe donc une transformation unique pour passer de l'une de ces fonctions d'ensemble vers l'autre. Cette transformation est dite transformée de *Möbius*. Soient deux fonctions d'ensemble ϑ et $\varphi : 2^{\Omega} \longrightarrow [0, 1]$, telles que :

$$\vartheta(A) = \sum_{B \subseteq A} \varphi(B). \tag{1.10}$$

φ est la transformée de *Möbius* de ϑ. On a alors :

$$\varphi(A) = \mathfrak{M}_{\{\vartheta\}}(A) = \sum_{B \subseteq A} (-1)^{|A|-|B|} \vartheta(B). \tag{1.11}$$

La fonction ϑ est appelée transformée de *Möbius* inverse de la fonction φ.

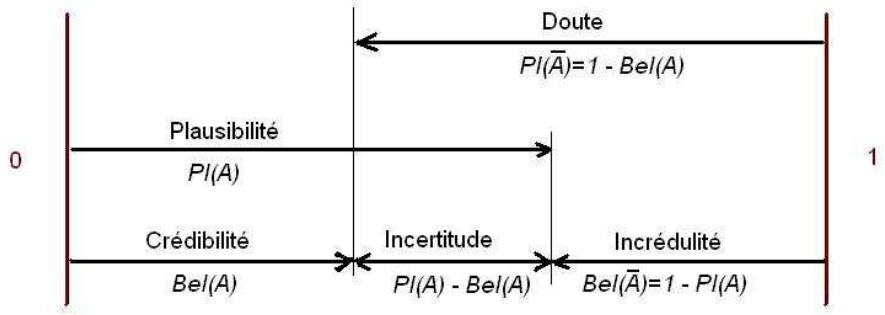

FIGURE 1.1 – Intervalle de croyance d'un sous-ensemble A

1.1.2 Interprétation des fonctions de plausibilité et de crédibilité

Etant donné deux fonctions Bel et Pl représentant un même état de connaissance d'une source d'informations, pour deux sous-ensembles $A, B \subseteq \Omega$, les relations suivantes sont vérifiées dans le modèle du monde fermé ($m(\emptyset) = 0$) :

$$Bel(A \cup B) \geq Bel(A) + Bel(B) - Bel(A \cap B). \tag{1.12}$$

$$Pl(A \cup B) \leq Pl(A) + Pl(B) - Pl(A \cup B). \tag{1.13}$$

$$Bel(A) + Bel(\overline{A}) \leq 1. \tag{1.14}$$

$$Pl(A) + Pl(\overline{A}) \geq 1. \tag{1.15}$$

$$0 \leq Bel(A) \leq Pl(A) \leq 1. \tag{1.16}$$

$$Bel(A) + Pl(\overline{A}) = Bel(\Omega) = Pl(\Omega) = 1. \tag{1.17}$$

Notons que les deux equations (1.12) et (1.13) correspondent respectivement aux deux propriétés de sur-additivité de la fonction Bel et de sous-additivité de la fonction Pl. L'intervalle d'incertitude relatif à une proposition A est borné par les quantités $Bel(A)$ et $Pl(A)$. Cet intervalle $[Bel(A), Pl(A)]$ est interprété comme un encadrement de la probabilité réelle $P(A)$ portée sur l'hypothèse A. Ainsi, nous pouvons écrire :

$$Bel(A) \leq P(A) \leq Pl(A).$$

Dans le cas ou le modèle est défini sans imprécision, nous pouvons noter que $Bel(A) = Pl(A) = P(A)$.
Il existe aussi dans l'intervalle de croyance d'autres fonctions symétriques à la crédibilité et à la plausibilité [25]. Celles-ci sont **l'incrédulité** qui est la crédibilité du contraire et **le doute** qui représente la plausibilité du contraire. Toutes les relations entre ces différentes mesures de croyance sont résumées dans la figure (1.1) [25].

1.1.3 Les combinaisons conjonctives et disjonctives des croyances

Dans un processus de prise de décision à l'aide de la TFC, la combinaison des fonctions de de masse est une étape cruciale. Cette opération permet l'agrégation des connaissances initiales fournies par diverses sources d'informations afin d'atteindre un niveau de connaissance élevé représenté par une seule fonction de masse synthétisant les données recueillies. Celle-ci tient compte, en effet, de toute l'information fournie par les différentes sources d'informations.
La majorité des règles de combinaison présentées dans la littérature utilisent soit l'opérateur conjonctif, soit l'opérateur disjonctif ou bien une combinaison particulière de ces deux derniers. En effet, étant donné deux fonctions de masse m_1 et m_2, définies sur Ω et issues de deux sources distinctes, un opérateur de combinaison fusionne les deux fonctions de masse précédentes en une nouvelle fonction de masse m_{12} traduisant le nouvel état de connaissance après avoir pris en considération les deux sources d'informations. La fusion est conjonctive si m_{12} est plus informative que les deux fonctions de masse initiales. Pour ce choix de combinaison, il est supposé que les deux sources d'informations soient fiables. Dans le cas où au moins une des sources est fiable, et que l'on ignore laquelle, on applique une stratégie de combinaison dite prudente utilisant l'opérateur disjonctif. Cette opération fournit alors une nouvelle masse combinée m_{12} moins informative que les différentes fonctions de masse combinées. Le choix de la stratégie de combinaison disjonctive implique une perte de spécificité.
Soient $m_{\textcircled{\cap}}$ et $m_{\textcircled{\cup}}$ les fonctions de masse obtenues après combinaison conjonctive et disjonctive d'un ensemble de fonctions $\{m_j\}_{j=1}^{M}$, On a :

– Pour la combinaison conjonctive

$$m_{\textcircled{\cap}}(A) = \sum_{\bigcap_{j=1}^{M} B_j = A} \prod_{j=1}^{M} m_j(B_j), \tag{1.18}$$

– Pour la combinaison disjonctive

$$m_{\textcircled{\cup}}(A) = \sum_{\bigcup_{j=1}^{M} B_j = A} \prod_{j=1}^{M} m_j(B_j). \tag{1.19}$$

On note $\textcircled{\cap}$ la règle de combinaison conjonctive et $\textcircled{\cup}$ la règle de combinaison disjonctive.

1.1.4 Le conditionnement

Dans la théorie des fonctions de croyance, le conditionnement d'une fonction de masse m par un sous-ensemble $B \subseteq \Omega$ est équivalent à la combinaison conjonctive de cette fonction de masse par la fonction catégorique $m_B(B) = 1$ engagée sur B. Le résultat de ce conditionnement est noté $m[B] = m \textcircled{\cap} m_B$. L'action de conditionnement quantifie le degré de croyance dans le cas où l'hypothèse B est retenue. La règle de conditionnement peut donc être calculée comme suit :

$$m[B](A) = \begin{cases} \sum\limits_{B \cap C = A} m(C) & \text{si } A \subseteq B, \\ 0 & \text{sinon.} \end{cases} \tag{1.20}$$

La masse de croyance initialement attribuée à $C \subseteq \Omega$ sera transférée vers le sous-ensemble $A = C \cap B$ après avoir retenu l'hypothèse B. Le conditionnement transfère donc toute la croyance initialement distribuée sur le cadre de discernement Ω vers les éléments du sous-ensemble $B \subseteq \Omega$. C'est cette idée du transfert qui est à la base du modèle de croyances transférables. Notons alors que la règle de combinaison conjonctive généralise le conditionnement. Son expression simplifiée est donnée par :

$$m_{1\textcircled{\cap}2}(A) = \sum_{B \subseteq \Omega} m_1[B](A) m_2(B), \quad \forall A \subseteq \Omega. \tag{1.21}$$

1.1.5 Le degré de Conflit

Lors de la combinaison conjonctive d'informations, un conflit apparaît fréquemment lorsque plusieurs sources d'informations imparfaites sont mises en jeu. Ce dernier est différent de zéro si les sources combinées ne s'accordent pas dans la description d'un même évènement en soutenant des hypothèses antagonistes. Dans un monde fermé, nous supposons connaître toutes les solutions au problème. Si alors deux sources sont en conflit, il est à déduire qu'au moins l'une d'entre elles est en erreur. Le degré de conflit est défini dans ce cas par la masse transférée à l'ensemble vide ($m(\emptyset)$) après une combinaison conjonctive. Cette masse est aussi appelée l'indice de conflit de Dempster et est donné par la formule :

$$\kappa = \sum_{\bigcap_{j=1}^{M} B_j = \emptyset} \prod_{j=1}^{M} m_j(B_j). \tag{1.22}$$

1.1.6 L'auto-conflit

La notion d'auto-conflit est introduite pour la première fois en 2006 par Osswald et Martin dans [38]. Étant donné que l'opérateur conjonctif n'est pas idempotent, la

combinaison de deux experts fournissant deux fonctions de masse identiques ne produit pas, en général, une même fonction de masse. Cette combinaison peut produire en outre une masse non nulle attribuée à l'ensemble vide bien que les deux masses combinées soient identiques. La notion d'auto-conflit permet alors de définir et de calculer le degré du conflit intrinsèque d'une fonction de masse [34, 33]. L'auto-conflit d'ordre n pour un expert est donné par la formule :

$$a_n = (\textcircled{\cap}_{i=1}^{n} m)(\emptyset) = \sum_{\bigcap_{i=1}^{n} B_i = \emptyset} \prod_{i=1}^{n} m(B_i). \tag{1.23}$$

La propriété principale de l'auto-conflit est qu'il augmente en fonction de l'ordre n. Autrement dit, au fur et à mesure que l'expert est combiné davantage avec lui même, l'auto-conflit augmente et tend davantage vers 1. Ainsi on peut écrire :

$$\begin{cases} 0 \leq a_n \leq 1 \quad \forall n \in \mathbb{N}^*, \\ a_n \leq a_{n+1}. \end{cases} \tag{1.24}$$

Exemple 1. *Étant donné une fonction de masse m définie sur $\Omega = \{\omega_1, \omega_2\}$:*

$$\mathbf{m} = 0.75\mathbf{m}_{\{\omega_1\}} + 0.2\mathbf{m}_{\{\omega_2\}} + 0.05\mathbf{m}_{\Omega}.$$

L'évolution de l'auto-conflit si l'on combine n fois la fonction m avec elle même est illustrée dans la figure 1.2.

1.1.7 Décomposition canonique conjonctive

Shafer [45] a défini une fonction de masse séparable comme le résultat de la combinaison conjonctive de fonctions de masse simples. Nous pouvons, en effet, définir une fonction de masse séparable comme suit :

$$m = \textcircled{\cap}_{\emptyset \neq A \subsetneq \Omega} m_A^{w(A)}. \tag{1.25}$$

sachant que $m_A^{w(A)}$ est une fonction de masse simple et que $w(A) \in [0, 1] \ \ \forall A \subsetneq \Omega$. Cette représentation est appelée par Shafer **décomposition canonique conjonctive**. Elle est unique si la fonction de masse m est non dogmatique.
Suite aux travaux de Shafer, Smets [50] a apporté une généralisation en 1995. Il a en effet, prouvé que toute fonction de croyance non-dogmatique peut être décomposée de façon unique en combinaison conjonctive de fonctions de masse simples généralisées. Une fonction de masse simple généralisée autorise des poids supérieurs à 1, alors : $w(A) \in [0, \infty[, \ \ \forall A \subsetneq \Omega$. La fonction des poids w peut être alors calculée à partir de la formule :

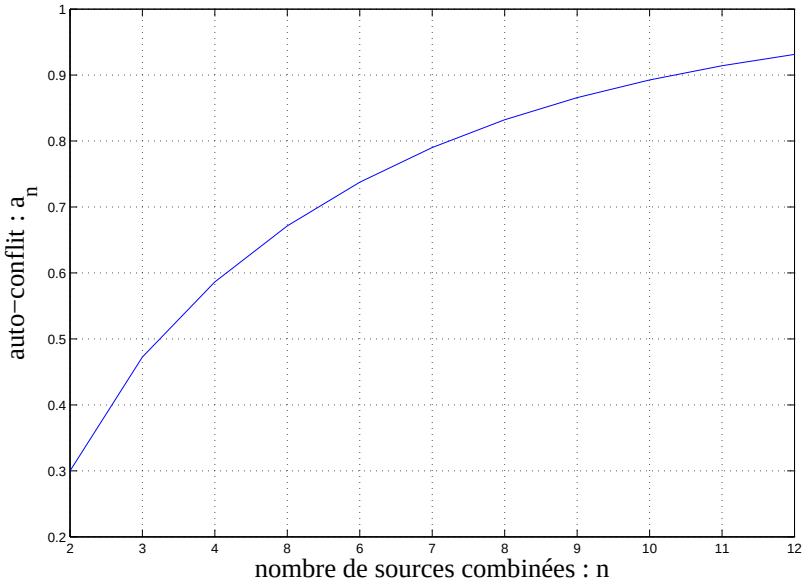

FIGURE 1.2 – Évolution de l'auto-conflit en fonction du nombre de fois que m est combinée avec elle même – Exemple 1.

$$w(A) = \prod_{B \supseteq A} q(B)^{(-1)^{|B|-|A|+1}}. \tag{1.26}$$

La décomposition canonique constitue une caractéristique assez importante pour les fonctions de masse non-dogmatiques. En effet, la combinaison conjonctive est alors équivalente à :

$$\begin{aligned} m_1 \textcircled{\cap} m_2 &= \left(\mathop{\textcircled{\cap}}_{A \subsetneq \Omega} m^{w_1(A)} \right) \textcircled{\cap} \left(\mathop{\textcircled{\cap}}_{A \subsetneq \Omega} m^{w_2(A)} \right), &(1.27)\\ &= \mathop{\textcircled{\cap}}_{A \subsetneq \Omega} m^{w_1(A) w_2(A)}. &(1.28) \end{aligned}$$

Nous pouvons aussi écrire la propriété précédente comme : $w_{1 \cap 2} = w_1 w_2$.

1.1.8 Relations d'ordre partiel entre fonctions de croyance

Parmi un ensemble de fonctions de masse candidates et compatibles avec les données collectées, la fonction de masse la plus appropriée est la moins informative. Ce choix est justifié par *le principe d'engagement minimal*. Ce principe joue un rôle comparable à celui du principe de maximum d'entropie dans la théorie des probabilités. Son application

dans la théorie des fonctions de croyance devient possible grâce à un certain nombre d'ordres partiels permettant la comparaison informationnelle des fonctions de masse. Quatre ordres partiels, généralisant l'inclusion ensembliste, sont proposés par Yager [56], Dubois et Prade [9] et Denœux [6]. Ces ordres partiels sont définis comme suit :

- $m_1 \sqsubseteq_{pl} m_2$ si $pl_1(A) \leq pl_2(A), \forall A \subseteq \Omega$,
- $m_1 \sqsubseteq_{q} m_2$ si $q_1(A) \leq q_2(A), \forall A \subseteq \Omega$,
- $m_1 \sqsubseteq_{w} m_2$ si $w_1(A) \leq w_2(A), \forall A \subset \Omega$,
- $m_1 \sqsubseteq_{s} m_2$ si il existe une matrice carrée S dont les termes généraux $S(A,B)$, $A, B \subseteq \Omega$ vérifient :

$$\begin{cases} \sum_{A \subseteq \Omega} S(A,B) = 1, & \forall B \subseteq \Omega, \\ S(A,B) > 0 \Rightarrow A \subseteq B, & \forall A, B \subseteq \Omega. \end{cases}$$

avec :

$$m_1(A) = \sum_{B \subseteq \Omega} S(A,B) m_2(B), \ \forall A \subseteq \Omega.$$

La quantité $S(A,B)$ est vue comme la proportion de la masse $m_2(B)$ transférée vers A. La fonction de masse m_1 est vue comme une **spécialisation** de m_2 à travers la matrice $S = [S(A,B)]$ qui est dite **matrice de spécialisation**.

Interprétation : Si on admet que $m_1 \sqsubseteq_x m_2$ ($x \in \{pl, q, w, s\}$), on peut alors dire que m_1 est plus engagée que m_2 [6, 40, 43]. Ce constat signifie que l'état des croyances représentées par m_1 est le résultat d'un apport d'éléments de preuves plus précis et/ou certains que m_2. En conséquence, m_1 possède un contenu informationnel plus important que m_2. A partir de la notion de spécialisation, une nouvelle relation d'ordre partielle plus forte en découle. Il s'agit de **la spécialisation dempsterienne** [22]. Ainsi m_1 est dite **spécialisation dempsterienne** de m_2 ssi il existe une fonction de masse m telle que $m_1 = m \textcircled{\cap} m_2$. Cette relation d'ordre est notée $m_1 \sqsubseteq_d m_2$. Cette relation implique donc l'existence d'une **matrice de spécialisation dempsterienne** qui est une fonction de m. Dans [6], il est montré que la relation d'ordre $\sqsubseteq_w$ est plus forte que toutes les autres relations citées ci-dessus. On a en effet :

$$m_1 \sqsubseteq_w m_2 \Rightarrow m_1 \sqsubseteq_d m_2 \Rightarrow m_1 \sqsubseteq_s m_2 \Rightarrow \begin{cases} m_1 \sqsubseteq_{pl} m_2 \\ m_1 \sqsubseteq_{q} m_2 \end{cases}. \tag{1.29}$$

1.1.9 Autres règles de combinaison dans la TFC

L'importance de la combinaison de fonctions de masse a déjà été énoncé en section 1.1.3. Outre la règle conjonctive et la règle disjonctive énoncées précédemment, il existe de nombreuses autres règles développées dans la littérature.
Dans cette section, nous nous intéressons aux différentes stratégies de combinaison des croyances définies dans la TFC. Sans être exhaustifs, nous abordons les règles de com-

binaison les plus utilisées dans la littérature.

1.1.9.1 Propriétés fondamentales des règles de combinaison

La combinaison d'informations est une étape cruciale dans l'agrégation des croyances. Elle permet l'obtention d'éléments d'évidence plus informatifs que les éléments combinés. Les différentes règles de combinaison présentent des propriétés algébriques qui caractérisent le comportement de chacune d'entre-elles vis-à-vis des résultats obtenus.

Commutativité : Étant donné deux fonctions de masse m_1 et m_2, l'opérateur $\odot$ est dit commutatif si :

$$m_1 \odot m_2 = m_2 \odot m_1.$$

Associativité : L'opérateur $\odot$ est associatif ssi pour trois fonctions de masse m_1, m_2 et m_3 nous pouvons écrire :

$$(m_1 \odot m_2) \odot m_3 = m_1 \odot (m_2 \odot m_3).$$

Quasi-Associativité : L'opérateur $\odot$ peut être décomposé en plusieurs opérations élémentaires associatives.

Idempotence : L'opérateur $\odot$ est dit idempotent ssi pour une fonction de masse m nous pouvons écrire :

$$m \odot m = m.$$

Ce résultat implique qu'un élément d'évidence n'est pris en compte sous cette loi qu'une seule fois.

1.1.9.2 Règle de combinaison de Dempster

Également appelée somme orthogonale, la règle de combinaison de Dempster [5] est la version normalisée de la règle conjonctive. Elle vérifie donc les propriétés d'associativité et de commutativité. Dans le cas de plusieurs sources notées S_j $(j = 1 \dots M)$, la fonction de masse m_D, résultat de la combinaison des fonctions de masse relatives à ces sources, s'écrit comme suit :

$$\begin{cases} m_D(\emptyset) = 0, \\ m_D(A) = \frac{1}{1-\kappa} \sum\limits_{\bigcap_{j=1}^{M} B_j = A} \prod\limits_{j=1}^{M} m_j(B_j) \quad \text{si } A \neq \emptyset. \end{cases} \tag{1.30}$$

Notons que la règle de combinaison de Dempster n'est pas définie dans le cas où le

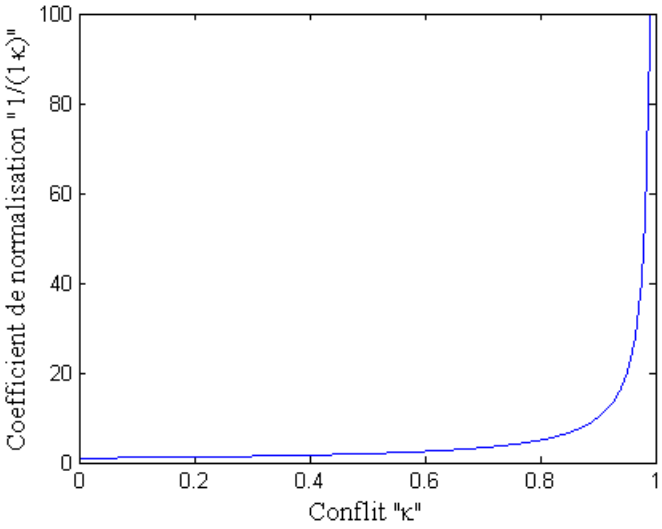

FIGURE 1.3 – Sensibilité de la règle de combinaison de Dempster à la variation du degré de conflit κ.

degré de conflit $\kappa = 1$. L'élément neutre de cette règle est l'ignorance totale (m_Ω). On note également que pour toute fonction catégorique sur un singleton $m_{\{\omega_i\}}$, la combinaison avec n'importe quelle autre fonction donnera $m_D = m_{\{\omega_i\}}$ à condition d'avoir toujours l'hypothèse $\kappa < 1$. La combinaison de Dempster propose une gestion du conflit en redistribuant celui-ci de façon équivalente et normalisée sur les différents éléments focaux. Les fortes variations du coefficient de normalisation $\frac{1}{1-\kappa}$ rendent la règle de combinaison de Dempster extrêmement sensible à la variation du degré de conflit [25, 26, 27]. La sensibilité du coefficient de normalisation de la règle de Dempster au degré de conflit κ est illustrée dans la figure 1.3

1.1.9.3 La règle de combinaison de Yager

Un autre point de vue de la gestion du conflit lors de la combinaison de sources d'informations imparfaites est apporté par Yager [57]. Ce dernier suppose qu'au moins une des sources combinées est fiable. Sans savoir laquelle, Yager répartit le conflit sur

l'ignorance totale. La règle de combinaison de Yager est alors :

$$m_{Yag}(A) = \begin{cases} \sum\limits_{\substack{\bigcap_{j=1}^{M} B_j = A \\ A \neq \emptyset}} \prod\limits_{j=1}^{M} m_j(B_j) & \text{si } A \neq \Omega \ , \\ 1 - \sum\limits_{A \in 2^{\Omega}/\{\Omega\}} \sum\limits_{\substack{\bigcap_{j=1}^{M} B_j = A \\ A \neq \emptyset}} \prod\limits_{j=1}^{M} m_j(B_j) & \text{sinon.} \end{cases} \tag{1.31}$$

1.1.9.4 La règle de Dubois et Prade

Cette règle propose une gestion plus fine du conflit en répartissant chaque composante partielle du conflit sur les ignorances partielles impliquant ce dernier. La règle de combinaison de Dubois et Prade [10] est alors définie pour tout $A \neq \emptyset$ de la façon suivante :

$$m_{DP}(A) = \sum_{\bigcap_{j=1}^{M} B_j = A} \prod_{j=1}^{M} m_j(B_j) + \sum_{\substack{\bigcup_{j=1}^{M} B_j = A \\ \bigcap_{j=1}^{M} B_j = \emptyset}} \prod_{j=1}^{M} m_j(B_j). \tag{1.32}$$

1.1.9.5 Règle de combinaison robuste de Florea (RCR)

Florea *et al.* [16] énoncent un formalisme générique de règles de combinaison dans le cadre de la théorie des fonctions de croyance basé sur la combinaison mixte et pondérée de la somme d'une combinaison conjonctive et d'une combinaison disjonctive. Cette règle de combinaison est reprise dans [15] où une étude spécifique est dédiée à deux lois de combinaison particulières, en l'occurrence la RCR symétrique et la RCR logarithmique. Dans un cadre général, cette famille de règles se définit comme suit : Étant donné deux fonctions de masse m_1 et m_2 définies dans Ω, la règle de combinaison robuste (RCR) est énoncée par la formule suivante :

$$m_{RCR}(A) = \begin{cases} \alpha(\kappa) m_{\textcircled{\cup}}(A) + \beta(\kappa) m_{\textcircled{\cap}}(A) & \text{si } A \neq \emptyset, \\ 0 & \text{sinon.} \end{cases} \tag{1.33}$$

où $\alpha(\kappa)$ et $\beta(\kappa)$ sont des fonctions qui dépendent du degré de conflit κ.

Du fait que m_{RCR} est une fonction de masse, $\alpha(\kappa)$ et $\beta(\kappa)$ doivent être choisies de sorte à ce que m_{RCR} puisse alors satisfaire la condition :

$$\begin{aligned} & \sum_{A\subseteq\Omega} m_{RCR}(A) = 1, \\ \Leftrightarrow \quad & \alpha(\kappa) \sum_{A\subseteq\Omega} m_{\textcircled{\cup}}(A) + \beta(\kappa) \sum_{A\subseteq\Omega} m_{\textcircled{\cap}}(A) = 1, \\ \Leftrightarrow \quad & \alpha(\kappa) + (1-\kappa)\beta(\kappa) = 1. \end{aligned} \tag{1.34}$$

Florea *et al.* adoptent une interprétation similaire à celle de Dubois et Prade concernant la redistribution du conflit. La formulation générale de la RCR répond alors aux conditions suivantes :

- α est une fonction croissante sachant que : $\alpha(0) = 0$ et $\alpha(1) = 1$,
- β est une fonction décroissante sachant que : $\beta(0) = 1$ et $\beta(1) = 0$,
- $1 - \alpha(\kappa) = (1-\kappa)\beta(\kappa),\ \forall\kappa \in [0,1]$.

1.1.9.6 Règle de Dezert et Smarandache (PCR)

Dezert et Smarandache [46] ont proposé une liste de méthodes de redistribution du conflit après combinaison. Ils s'intéressent à la redistribution de ce dernier sur les éléments focaux responsables de son apparition. Le principe de la PCR (Proportional Conflicting Redistribution) vise à transférer le conflit d'une façon proportionnelle vers les sous-ensembles non-vides responsables de son apparition.
Plusieurs versions d'opérateurs de combinaison sont synthétisées par Dezert et Smarandache pour redistribuer le conflit. La règle la plus aboutie est la PCR6. Elle est donnée pour deux sources d'informations par :

$$m_{PCR5}(A) = m_{\textcircled{\cap}}(A) + \sum_{\substack{B\subseteq\Omega \\ A\cap B=\emptyset}} \left(\frac{m_1(A)^2 m_2(B)}{m_1(A) + m_2(B)} + \frac{m_2(A)^2 m_1(B)}{m_2(A) + m_1(B)} \right). \tag{1.35}$$

L'extension de cette règle pour plusieurs sources d'informations est reprise par A. Martin et C. Osswald dans [38], [35], [33] puis généralisée sous l'appellation de PCR6. Celle-ci est donnée par la formule suivante :

$$m_{PCR6}(A) = m_{\textcircled{\cap}}(A) + \sum_{i=1}^{M} m_i(A)^2 \sum_{\bigcap_{k=1}^{M-1} B_{\sigma_i(k)} \cap A = \emptyset} \left(\frac{\prod_{j=1}^{M-1} m_{\sigma_i(j)}(B_{\sigma_i(j)})}{m_i(A) + \sum_{j=1}^{M-1} m_{\sigma_i(j)}(B_{\sigma_i(j)})} \right), \tag{1.36}$$

où σ_i prends des valeurs de 1 à M selon i :

$$\begin{cases} \sigma_i(j) = j & \text{si } j < i, \\ \sigma_i(j) = j+1 & \text{si } j \geq i. \end{cases} \tag{1.37}$$

Cette règle de combinaison assure une gestion plus fine de la redistribution du conflit puisqu'elle le redistribue proportionnellement sur les sous-ensembles responsables de son apparition. Néanmoins, elle est une règle de combinaison non-linéaire et aussi gourmande en terme de temps de calcul.

1.1.9.7 Règle prudente de Denœux

Quand un agent reçoit deux fonctions de masse m_1 et m_2 issues de deux sources fiables, la fonction de masse combinée m_{12} représente son nouvel état de connaissance. Elle doit être plus informative que m_1 et m_2 :

$$m_{12} \in S_x(m_1) \cap S_x(m_2). \tag{1.38}$$

$S_x(m)$ est l'ensemble des fonctions de masse plus riches que m au sens de l'ordre partiel $\sqsubseteq_x$ avec $x \in \{pl, q, s, w\}$.

L'application du principe d'engagement minimal permet en effet de choisir dans $S_x(m_1) \cap S_x(m_2)$ la fonction de masse la moins riche au sens de $\sqsubseteq_x$ si elle existe.
Soient m_1 et m_2 deux fonctions de masse non dogmatiques. L'élément le moins engagé dans $S_w(m_1) \cap S_w(m_2)$ existe et est unique. On le note $m_{1\textcircled{\wedge}2}$ et sa fonction de poids conjonctifs $w_{1\textcircled{\wedge}2}$ est telle que :

$$w_{12}(A) = w_1(A) \wedge w_2(A) \ ; \ \forall A \subsetneq \Omega. \tag{1.39}$$

où $\wedge$ désigne l'opérateur minimum.

La fonction $m_{1\textcircled{\wedge}2}$ est le résultat de la combinaison des fonctions de masse m_1 et m_2 en utilisant la règle prudente de Denœux $\textcircled{\wedge}$. D'après la décomposition canonique conjonctive, on a :

$$m_{1\textcircled{\wedge}2} = \bigcirc\!\!\!\!\!\!\cap_{A \subsetneq \Omega} m^{w_1(A) \wedge w_2(A)}. \tag{1.40}$$

Il est à noter que cette règle s'applique seulement aux fonctions de masse non dogmatiques.

1.1.9.8 Règle de combinaison hardie de Denœux

Dans le cas de la règle prudente, il est supposé que toutes les sources sont fiables. Nous avons aussi vu que cette règle étend l'intersection ensembliste. Si nous supposons une autre situation où une seule source d'informations est fiable, le comportement le plus adapté pour la combinaison est le comportement disjonctif. Dans ce cas, Denœux a montré dans [6] qu'il existe une relation d'ordre partielle basée sur les fonctions de masse séparables complémentaires notée $\sqsubseteq_\nu$ permettant de généraliser la disjonction ensembliste.

La fonction $m_{1\textcircled{\vee}2}$ est le résultat de la combinaison des fonctions de masse m_1 et m_2 en utilisant la règle hardie de Denœux $\textcircled{\vee}$:

$$m_{1\textcircled{\vee}2} = \bigcirc\!\!\!\!\!\!\cup_{A\neq\emptyset} m^{\nu_1(A)\wedge\nu_2(A)}. \tag{1.41}$$

Il est à noter que cette règle s'applique seulement aux fonctions de masse non normalisée ($m(\emptyset) \neq 0$).

1.1.9.9 Comparaison synthétique des règles de combinaison

Le Tableau 1.1 récapitule les différentes propriétés algébriques des lois de combinaison vues précédemment.

	Commutativité	Associativité	Quasi-associativité	Idempotence
Conjonctive	**o**	**o**		
Disjonctive	**o**	**o**		
Dempster	**o**	**o**		
Smets	**o**	**o**		
Yager	**o**		**o**	
Dubois et Prade	**o**		**o**	
RCR (Florea)	**o**		**o**	
PCR	**o**		**o**	
Denœux	**o**	**o**		**o**

TABLE 1.1 – Propriétés principales des différentes règles de combinaison

1.1.10 La prise de décision

Dans la théorie de la décision, le principe d'inférence bayésienne est un des plus utilisés. Cette approche consiste à élire l'action pour laquelle l'espérance du coût est la plus faible. Pour ce faire, il est tout d'abord impératif de définir un ensemble d'actions $\mathcal{A} = \{a_1, a_2, ..., a_n\}$ et une fonction coût $\lambda : \mathcal{A} \times \Omega \rightarrow \mathbb{R}$ sachant que la valeur $\lambda(a_i|\omega_j)$

représente le coût du choix de l'action a_i si la vérité est dans l'hypothèse ω_j. Le risque conditionnel de décider a_i relatif à une mesure de probabilité P est défini comme suit :

$$R_P(a_i) = \sum_{\omega_j \in \Omega} \lambda(a_i|\omega_j) P(\omega_j). \tag{1.42}$$

La règle de décision bayésienne consiste donc à choisir l'action a_i^* qui minimise le risque conditionnel :

$$a_i^* = arg \min_{a_i \in \mathcal{A}} \Big(R_P(a_i) \Big). \tag{1.43}$$

A l'issue du processus de combinaison, si m est une fonction synthétisant l'ensemble des connaissances, il est aisé de calculer des fonctions de crédibilité et de plausibilité à partir de cette fonction de masse. En effet, elles représentent la croyance minimale et la croyance maximale encadrant la probabilité effective de chaque hypothèse. Le problème qui se pose est donc formulé comme suit : *Afin d'utiliser la règle de décision de Bayes, comment peut-on transformer préalablement une fonction de masse m en une distribution de probabilité ?*

Dans [49], Smets définit la transformation pignistique pour transformer les fonctions de masse en distributions de probabilités. Il s'est basé sur le principe de la raison insuffisante, autrement dit, en l'absence d'informations justifiant le choix d'une hypothèse plutôt qu'une autre, on suppose qu'elles sont équiprobables. La transformation pignistique redistribue alors la masse $m(A)$ sur tous les éléments de A de façon équitable. Cette transformation permet d'obtenir une distribution de probabilité particulière dite probabilité pignistique et définie par :

$$BetP(\omega_i) = \sum_{A \subseteq \Omega, \omega_i \in A} \frac{1}{|A|} \frac{m(A)}{1 - m(\emptyset)}. \tag{1.44}$$

Nous remarquons que, contrairement à m, $BetP$ est une mesure additive, alors $\forall A,\ B \in 2^\Omega$; si $A \cap B \neq \emptyset,\ BetP(A \cup B) = BetP(A) + BetP(B)$. Cette transformation permet d'associer une distribution de probabilité pignistique à une seule fonction de croyance. Inversement, il peut exister une infinité de fonctions de croyance associées à une probabilité pignistique. Le passage d'une fonction de masse vers une distribution de probabilité pignistique engendre en effet une perte d'informations.

Dans le cas où la fonction coût ne prend des valeurs que dans $\{0, 1\}$, la minimisation du risque conditionnel au sein du processus d'inférence bayésien donné par la formule (1.42) revient à choisir l'hypothèse de plus grande probabilité pignistique.

D'autres règles de décision basées sur le maximum de crédibilité ou sur le maximum de plausibilité peuvent aussi être utilisées dans le cadre de la théorie des fonctions

de croyance. Celles-ci sélectionnent respectivement l'hypothèse la plus crédible et l'hypothèse la plus plausible [25, 33, 40]. Il est aussi possible d'employer d'autres risques conditionnels basés sur le principe de la définition des enveloppes supérieures et inférieures qui cernent la vraie probabilité de chaque hypothèse. En effet, deux risques inférieur et supérieur sont donc respectivement définis par les formules suivantes :

$$R_*(a_i) = \sum_{A \subseteq \Omega} m(A) \min_{\omega_j \in A} \lambda(a_i|\omega_j), \tag{1.45}$$

$$R^*(a_i) = \sum_{A \subseteq \Omega} m(A) \max_{\omega_j \in A} \lambda(a_i|\omega_j). \tag{1.46}$$

La minimisation des deux équations précédentes conduit à la définition de deux stratégies appelées respectivement **stratégie de décision optimiste** et **stratégie de décision pessimiste**. Les deux risques définis dans les équations (1.45) et (1.46) encadrent le risque pignistique :

$$R_*(a_i) \leq R_{Bet}(a_i) \leq R^*(a_i). \tag{1.47}$$

Notons que dans le cas où le coût est dans $\{0, 1\}$, la minimisation du risque inférieur conduit à une prise de décision **optimiste** basée sur le maximum de crédibilité, tandis que la minimisation du risque supérieur aboutit à une prise décision **pessimiste** basée sur le maximum de plausibilité.

1.2 Autres opérations sur les fonctions de croyance

1.2.1 Affaiblissement

Dans un processus de fusion d'informations, un doute peut être émis sur la fiabilité des informations recueillies par une source d'informations et représentées par la fonction de masse m. Afin de prendre en compte ce doute dans le processus de modélisation, une correction peut s'imposer sur cette même fonction de masse m si l'on connait le degré de fiabilité de la source d'informations en question. Une des solutions proposées pour répondre à cette question est **l'affaiblissement**.

L'affaiblissement d'une fonction de masse m est l'opération qui déplace une partie de la croyance de chaque sous-ensemble $A \subseteq \Omega$ vers un autre sous-ensemble $B \subseteq \Omega$ tel que $A \subseteq B$. La notion d'affaiblissement a pour objectif de pondérer les sources d'informations selon leur fiabilité afin de maîtriser leur impact dans le processus de combinaison.

La formulation la plus simple de l'affaiblissement est définie par Shafer [45] en 1976. Selon lui, il s'agit de déplacer proportionnellement une part de la croyance affectée à chaque sous-ensemble vers l'ensemble Ω qui caractérise l'ignorance totale. Ainsi on diminue l'engagement de la source d'informations en question et donc son impact dans

le processus de fusion (si celui-ci est de nature conjonctive).
Étant donné une fonction de masse m modélisant la connaissance d'une source S à un instant donné et un coefficient $\alpha \in [0,1]$ qui représente le taux de d'affaiblissement de la source S, l'opération d'**affaiblissement** est alors définie par la formule :

$$m^{\alpha}(A) = \begin{cases} (1-\alpha)m(A) & \text{si } A \neq \Omega, \\ \alpha + (1-\alpha)m(\Omega) & \text{sinon.} \end{cases} \tag{1.48}$$

Nous pouvons facilement constater que si $\alpha = 0$, aucune modification n'est apportée sur la fonction $m = m^0$. Une confiance totale est alors attribuée à la source S. Dans le cas contraire, si $\alpha = 1$, la fiabilité de l'information fournie pas S est totalement mise en cause. Aucune confiance n'est alors attribuée à la source S et l'information qu'elle fournit est totalement ignorée dans le cas d'une combinaison conjonctive en transformant la fonction m en fonction de masse vide, encore appelée ignorance totale. Ainsi, nous écrivons $m^1 = m_\Omega$. Nous remarquons aussi qu'une fonction de masse catégorique m_A affaiblie devient une fonction de masse à support simple m_A^α.

1.2.2 Grossissement et raffinement

Dans la théorie des fonctions de croyance, le degré de granularité d'un cadre de discernement Ω, tout comme la taille de celui-ci est souvent l'objet d'une convention. En effet, une hypothèse $\omega \in \Omega$ peut éventuellement regrouper plusieurs partitions d'un cadre de discernement Θ plus fin mais compatible avec Ω. Il est donc utile d'établir entre ces deux référentiels Ω et Θ des relations régies par les deux opérations de **raffinement** et de **grossissement**.

Soient Ω et Θ deux cadres de discernement à éléments finis tels que $|\Omega| < |\Theta|$. Une fonction $f : 2^\Omega \to 2^\Theta$ est appelée **raffinement** si elle vérifie les conditions suivantes :

- $\forall \omega \in \Omega,\ f(\{\omega\}) \neq \emptyset$,
- l'ensemble $\{f(\{\omega\}),\ \omega \in \Omega\} \subseteq 2^\Theta$ est une partition de Θ,
- pour tout sous-ensemble B de Ω, la relation suivante est vérifiée :

$$f(B) = \bigcup_{\omega \in B} f(\{\omega\}).$$

L'ensemble Θ est appelé **raffinement** de Ω et inversement Ω est dit **grossissement** de Θ. Soit, par exemple, un cadre de discernement $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Celui-ci est le

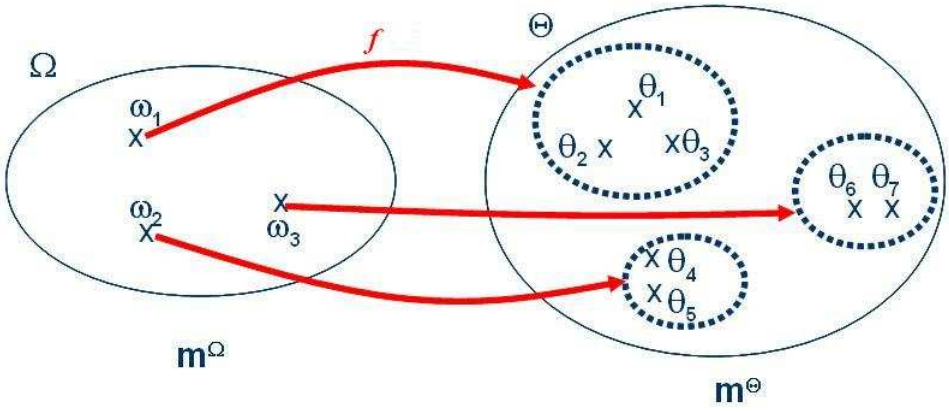

FIGURE 1.4 – Raffinement Θ du cadre de discernement Ω

grossissement de $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7\}$ défini par le raffinement :

$$\begin{cases} f(\{\omega_1\}) = \{\theta_1, \theta_2, \theta_3\}, \\ f(\{\omega_2\}) = \{\theta_4, \theta_5\}, \\ f(\{\omega_1\}) = \{\theta_6, \theta_7\}. \end{cases}$$

La situation expliquée dans l'exemple précédent est résumé dans la figure (1.4).

Étant donné un état de connaissance défini dans le cadre de discernement Ω par la fonction de massé notée m^Ω, le raffinement f défini entre les ensembles Ω et Θ est étendu aux fonctions de masse en transférant toute masse $m^\Omega(B \subseteq \Omega)$ vers l'ensemble $A = f(B)$ avec $A \subseteq \Theta$. Ce transfert est justifié par le principe de la raison insuffisante. Cette opération est appelée **extension vide** et définie par la formule :

$$m^{\Omega\uparrow\Theta}(A) = \begin{cases} m^\Omega(B) \ \text{ si } A = f(B) \text{ et } B \subseteq \Omega, \\ 0 \ \text{ sinon.} \end{cases} \tag{1.49}$$

L'opération inverse du raffinement consiste à transformer l'information contenue en m^Θ vers le cadre de discernement Ω. Cette opération est appelée **restriction** et elle est définie par :

$$m^{\Theta\downarrow\Omega}(A) = \sum_{\substack{B\subseteq\Theta \\ \overline{f}(B)=A}} m^\Theta(B), \quad \forall A \subseteq \Omega, \tag{1.50}$$

avec $\overline{f}$ une fonction de $2^\Theta \to 2^\Omega$ telle que $\forall B \subseteq \Theta$, $\overline{f}(B) = \{\omega \in \Omega | f(\{\omega\}) \cap B \neq \emptyset\}$. La fonction $\overline{f}$ est appelée réduction externe.

La définition de la fonction f est souvent liée au contexte applicatif. Elle permet le passage d'un niveau de hiérarchie à un autre via les équations (1.49) et (1.50). Toutefois, elle entraine néanmoins une complexité algorithmique du fait de la variation de la taille du cadre de discernement.

1.3 Fonctions de croyance et calcul matriciel

En 2001, Jousselme a utilisé pour ses travaux concernant les distances dans la théorie des fonctions de croyance [19, 21] une certaine représentation géométrique qui considère les fonctions de masse comme les composantes d'un espace vectoriel dont les vecteurs de base sont ordonnés selon un ordre dit **binaire** ou **naturel**. Cette représentation est ensuite développée par Cuzzolin [3, 4].

La notation matricielle est très utile dans la théorie des fonctions de croyance. Elle permet, à la fois, de mieux mettre en évidence la linéarité dont jouissent la majorité des opérations de base, mais aussi de simplifier très considérablement l'arithmétique des fonctions de croyance. Cette notion est introduite dans le cadre du modèle de croyances transférables en 2002 par Smets [52]. Il propose alors une analyse suffisamment élaborée pour appliquer le calcul matriciel aux fonctions de croyance. Cette analyse est reprise dans les travaux de Pichon [40].

1.3.1 Fonction de masse et ordre naturel

Pour des fins algorithmiques, une fonction de masse m (ainsi que les autre formes dérivées de représentation d'informations bel, pl, b et q) définie dans un cadre de discernement Ω est souvent vue comme un vecteur colonne stochastique de dimension $N = 2^{|\Omega|}$. Les éléments de ce dernier peuvent, en effet, être ordonnés de façon quelconque mais un ordre en particulier rend les calculs algorithmiques facilement accessibles. Cet ordre est appelé **Ordre Naturel** ou **Ordre Binaire**.

Dans un cadre de discernement ordonné de cardinalité $n = |\Omega|$, on attribue une représentation binaire sur n bits à chaque sous-ensemble $A \subseteq \Omega$. Les hypothèses $\omega_i \in A$ sont alors codées par la valeur 1 dans la représentation binaire. En effet, le i^{me} élément du vecteur $\mathbf{m}$ est associé au sous-ensemble A de Ω correspondant à la représentation binaire de la valeur $i-1$. Supposons, à titre d'exemple, un cadre de discernement $\Omega = \{\omega_1, \omega_2, \omega_3\}$. Le tableau 1.2 illustre l'ordre et la représentation binaire des éléments du vecteur $\mathbf{m}$. Le huitième élément correspond au sous-ensemble $\Omega = \{\omega_1, \omega_2, \omega_3\}$ car la représentation binaire de $8 - 1 = 7$ est 111. La représentation binaire 010 concerne le sous-ensemble $\{\omega_2\} \subset \Omega$ qui correspond au troisième élément du vecteur m puisque $2^1 + 1 = 3$.

TABLE 1.2 – Ordre des éléments du vecteur m quand $\Omega = \{\omega_1, \omega_2, \omega_3\}$

Position	$\omega_3\omega_2\omega_1$	A	m
1	000	$\emptyset$	$m(\emptyset)$
2	001	$\{\omega_1\}$	$m(\{\omega_1\})$
3	010	$\{\omega_2\}$	$m(\{\omega_2\})$
4	011	$\{\omega_1, \omega_2\}$	$m(\{\omega_1, \omega_2\})$
5	100	$\{\omega_3\}$	$m(\{\omega_3\})$
6	101	$\{\omega_1, \omega_3\}$	$m(\{\omega_1, \omega_3\})$
7	110	$\{\omega_2, \omega_3\}$	$m(\{\omega_2, \omega_3\})$
8	111	$\{\omega_1, \omega_2, \omega_3\}$	$m(\{\omega_1, \omega_2, \omega_3\})$

Nous adoptons pour la suite du travail les notations matricielles suivantes :

- Les vecteurs sont des vecteurs colonnes et les matrices sont carrées. La longueur des vecteurs est par défaut égale à $N = 2^{|\Omega|}$ et les matrices sont de taille $N \times N$.
- Les vecteurs et les matrices sont notés en gras. Une matrice peut être notée par l'écriture $\mathbf{M} = [M_{i,j}]$, mais aussi $\mathbf{M} = [M(A, B)]$, $\forall A, B \in \Omega$. Les indices des lignes et des colonnes i et j sont ceux correspondant aux sous-ensembles B_i et B_j de Ω ordonnés selon l'ordre binaire. Par exemple, si $B_i = \{\omega_1\}$ et $B_j = \{\omega_1, \omega_2\}$, on a $M(B_i, B_j) = M_{2,4}$.
- $\mathbf{1}$ est la matrice dont tous les éléments valent 1.
- $\mathbf{I}$ est la matrice identité dont tous les éléments sont nuls exceptés ceux de la diagonale principale qui sont égaux à 1.
- $\mathbf{J}$ est la matrice carrée dont tous les éléments sont nuls à l'exception des éléments de la deuxième diagonale qui sont identiques et dont la valeur est 1. Les éléments de la matrice $\mathbf{J}$ sont donc :

$$J(i,j) = \begin{cases} 1 & \text{si } \; i + j - 1 = N, \\ 0 & \text{sinon.} \end{cases} \tag{1.51}$$

  La propriété principale de la matrice $\mathbf{J}$ est que celle-ci inverse les lignes d'une matrice $\mathbf{M}$ quand le produit matriciel $\mathbf{J}.\mathbf{M}$ est réalisé, mais elle inverse l'ordre des colonnes si cette fois-ci c'est le produit $\mathbf{M}.\mathbf{J}$ qui est réalisé.

Les vecteurs fonctions de croyance sont alors notés en gras. Par exemple, on peut écrire pour les négations des fonctions de masse :

$$\overline{\mathbf{m}} = \mathbf{J}.\mathbf{m}, \quad \overline{\mathbf{q}} = \mathbf{J}.\mathbf{b}, \quad \overline{\mathbf{b}} = \mathbf{J}.\mathbf{q}.$$

1.3.2 La transformation des fonctions de masse

Les différentes transformations possibles entre les différentes représentations d'une fonction de croyance peuvent être exprimées en utilisant la notation matricielle.

Étant donné une fonction de masse m, la transformation de cette dernière en fonction de communalité est donnée par la relation :

$$q(A) = \sum_{\Omega \supseteq B \supseteq A} m(B).$$

Celle-ci peut être écrite comme suit :

$$q(A) = \sum_{\Omega \supseteq B} M(A,B).m(B),$$

avec

$$M(A,B) = \begin{cases} 1 & si A \subseteq B, \\ 0 & sinon. \end{cases}$$

Les éléments $M(A,B)$ sont les composantes d'une matrice $\mathbf{M}$ appelée **matrice d'incidence**. $\mathbf{M}$ est une matrice triangulaire supérieure. Elle permet d'avoir :

$$\mathbf{q} = \mathbf{M}.\mathbf{m},$$

$$\mathbf{m} = \mathbf{M^{-1}}.\mathbf{q}.$$

La **matrice de communalité** est définie par la matrice $\mathbf{Q}$ dont les éléments sont :

$$Q_{ij} = \begin{cases} q(A_i) & si \quad i = j, \\ 0 & sinon. \end{cases}$$

La représentation matricielle nous permet de réécrire les fonctions Bel et Pl comme :

$$\mathbf{Bel} = {}^t\mathbf{M}.\mathbf{m}, \ \mathbf{Pl} = \mathbf{N}.\mathbf{m}. \tag{1.52}$$

où ${}^t\mathbf{M}$ désigne la transposée de la matrice d'incidence $\mathbf{M}$. La matrice $\mathbf{N}$ est quant à elle définie par :

$$N(A,B) = \begin{cases} 1 & \text{si } A \cap B \neq \emptyset, \\ 0 & \text{sinon.} \end{cases} \tag{1.53}$$

Pour l'exemple du cadre de discernement $\Omega = \{\omega_1, \omega_2, \omega_3\}$, la matrice $\mathbf{M}$ est donnée dans le tableau 1.3.

Nous pouvons facilement constater que la matrice $\mathbf{M}$ est une matrice triangulaire supérieure et qu'elle est construite à partir du bloc de construction :

$$\mathbf{M^1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}.$$

TABLE 1.3 – Matrice d'incidence $\mathbf{M}$ quand $\Omega = \{\omega_1, \omega_2, \omega_3\}$

	$\emptyset$	$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_1, \omega_2\}$	$\{\omega_3\}$	$\{\omega_1, \omega_3\}$	$\{\omega_2, \omega_3\}$	$\{\omega_1, \omega_2, \omega_3\}$
$\emptyset$	1	1	1	1	1	1	1	1
$\{\omega_1\}$	0	1	0	1	0	1	0	1
$\{\omega_2\}$	0	0	1	1	0	0	1	1
$\{\omega_1, \omega_2\}$	0	0	0	1	0	0	0	1
$\{\omega_3\}$	0	0	0	0	1	1	1	1
$\{\omega_1, \omega_3\}$	0	0	0	0	0	1	0	1
$\{\omega_2, \omega_3\}$	0	0	0	0	0	0	1	1
$\{\omega_1, \omega_2, \omega_3\}$	0	0	0	0	0	0	0	1

La matrice $\mathbf{M^1}$ représente la matrice d'incidence $\mathbf{M}$ quand $|\Omega| = 1$. L'obtention de la matrice d'incidence pour un cadre de discernement à i éléments peut, en effet, être effectuée à partir de la matrice d'incidence d'un ensemble à $i-1$ éléments. Cette construction est possible grâce au produit de Kronecker du bloc $\mathbf{M^1}$ par la matrice obtenue pour l'ensemble à $i-1$ éléments[52] :

$$\mathbf{M^i} = \mathbf{Kron}\left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \mathbf{M^{i-1}}\right). \tag{1.54}$$

La transformation de la fonction de masse m en fonction d'implicabilité b est possible grâce à la matrice $\mathbf{B}$. On peut ainsi écrire $\mathbf{b} = \mathbf{B}.\mathbf{m}$ [52]. Cette matrice peut être calculée à partir de la matrice $\mathbf{M}$; nous avons alors $\mathbf{B} = \mathbf{J}.\mathbf{M}.\mathbf{J}$ [52]. Elle peut également s'obtenir, comme pour la matrice $\mathbf{M}$, en utilisant le produit de Kronecker. Le bloc de construction est cette fois-ci :

$$\mathbf{B^1} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Il en découle donc :

$$\mathbf{B^i} = \mathbf{Kron}\left(\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \mathbf{B^{i-1}}\right). \tag{1.55}$$

1.3.3 Matrices de spécialisation dempsteriennes

1.3.3.1 Généralités

Le concept de spécialisation est fondé sur la redistribution de la masse de croyance après l'ajout de nouveaux éléments d'informations. En effet, l'ajout de nouvelles connaissances modifie la connaissance initiale du problème et produit une fonction de masse au

moins autant engagée que la fonction initiale [22, 37]. Il est alors possible de stocker le transfert de masse dans une matrice appelée matrice de spécialisation.

Définition 3. *Une matrice de spécialisation* **S** *est une matrice stochastique [52] suivant les colonnes et dont les éléments vérifient* $S(A,B) = 0 \ \forall \ A \not\subseteq \quad B$.

Soit une fonction de masse initiale m_0 qui attribue au sous-ensemble B la valeur $m_0(B)$. Si $S(A,B) \in [0,1]$ est la proportion de la masse $m_0(B)$ transférée vers un sous-ensemble $A \subseteq B$. Après spécialisation, la nouvelle fonction de masse résultante m_1 est telle que $\forall A \subseteq \Omega$:

$$m_1(A) = \sum_{B \subseteq \Omega} S(A,B).m_0(B) \tag{1.56}$$

Dans le souci de conserver la totalité de la croyance $m_0(B)$ après le transfert, les deux conditions suivantes sont observées :

$$\begin{cases} \sum\limits_{A \subseteq B} S(A,B) = 1, \quad \forall B \subseteq \Omega \Leftrightarrow \mathbf{1}.\mathbf{S} = \mathbf{1}, \\ S(A,B) > 0 \ \Rightarrow \ A \subseteq B. \end{cases}$$

On peut facilement constater que $\lambda = 1$ est une valeur propre de la matrice de spécialisation **S** et que cette même matrice est une matrice triangulaire supérieure.
Si par exemple une fonction de masse m_1 est une spécialisation de m_0. Elle peut s'écrire sous une forme matricielle comme suit :

$$\mathbf{m}_1 = \mathbf{S}.\mathbf{m}_0.$$

Dans un processus de combinaison conjonctive, le transfert de masse est aussi considéré à travers la règle de conditionnement de Dempster. En effet, la quantité $(m \textcircled{\cap} m_B)(A)$ est une mesure des transferts de masse à partir d'un état de croyance défini par la fonction de masse m si l'on apprend que l'hypothèse $B \subset \Omega$ est vraie. La quantité $(m \textcircled{\cap} m_B)(A)$ est la proportion maximale de masse qui doit être transférée du sous-ensemble B vers $A \subset B$ si une source d'informations qui croit en B est à combiner avec m. Selon ce transfert, une matrice de spécialisation peut être définie. Celle-ci est la matrice de spécialisation dempsterienne.

Définition 4. *La matrice de spécialisation dempsterienne* **S** *décrivant la fonction de masse m définie dans un cadre de discernement Ω est une matrice carrée dont les éléments sont donnés par :*

$$S(A,B) = (m \textcircled{\cap} m_B)(A), \quad \forall A, B \subseteq \Omega. \tag{1.57}$$

Selon la définition de la règle de conditionnement dempsterienne, nous concluons que chaque fonction de masse définit une matrice de spécialisation dempsterienne unique et inversement. Cette correspondance bijective sera intuitivement constatée dans le théorème 2 ci-après. La matrice de spécialisation dempsterienne d'une fonction de masse ne décrit pas uniquement l'état de la connaissance à un instant donné mais également tous les états futurs qui peuvent être atteints à partir de celle-ci dans un processus de combinaisons conjonctives puisqu'elle contient tous les conditionnements possibles de la fonction de masse qui lui correspond.
Dans le cas de fonctions de masse catégoriques, la matrice de spécialisation dempsterienne est définie par le lemme suivant :

Lemme 1. *Étant donné* $\mathbf{S}_X$ *la matrice de spécialisation dempsterienne de la fonction de masse catégorique* m_X *définie dans le cadre de discernement* Ω*, les éléments de cette matrice sont définis comme suit :*

$$S_X(A,B) = \begin{cases} 1 & si\ X \cap B = A, \\ 0 & sinon. \end{cases} \tag{1.58}$$

Le lemme précédent est énoncé par Klawonn et Smets dans [22]. Nous le reprenons dans notre travail et nous y apportons une démonstration légèrement plus explicite :

Démonstration. A partir de la définition des matrices de spécialisation dempsteriennes, nous avons $S_X(A,B) = (m_X \textcircled{\cap} m_B)(A)$. Il résulte dans la théorie des fonctions de croyance que $m_X \textcircled{\cap} m_B = m_{X \cap B}$ parce que la théorie des fonctions de croyance généralise la théorie des ensemble de Cantor. Selon la définition d'une fonction de masse catégorique $m_{X \cap B}$, le résultat donné dans l'équation (1.58) est implicitement obtenu. □

1.3.3.2 Fusion conjonctive et matrices de spécialisation dempsteriennes

La loi de combinaison conjonctive, sans doute la règle de combinaison la plus utilisée dans les processus de fusion d'informations, peut être exprimée sous une forme matricielle. En effet, la révision conjonctive d'une fonction de masse m_1 par une autre fonction m_2 est assurée par une matrice de spécialisation dite dempsterienne [22] notée $\mathbf{S}_2$. Cette matrice est vue comme une fonction de la fonction de masse m_2.

$$\begin{aligned} m_1 \textcircled{\cap} m_2(A) &= \sum_{B \cap C = A} m_1(B).m_2(C), \\ &= \sum_{B \subseteq \Omega} \Big(\sum_{B \cap C = A} m_2(C) \Big).m_1(B), \\ &= \sum_{B \subseteq \Omega} (m \textcircled{\cap} m_B)(A).m_1(B), \\ &= \sum_{B \subseteq \Omega} S_2(A,B).m_1(B). \end{aligned}$$

La définition 4 nous permet de constater que les éléments $S_2(A,B)$ ne dépendent que de la fonction de masse m_2.
La représentation matricielle de la règle de combinaison conjonctive est alors :

$$\mathbf{m}_1 \textcircled{\cap} \mathbf{m}_2 = \mathbf{S}_2.\mathbf{m}_1 = \mathbf{S}_1.\mathbf{m}_2. \tag{1.59}$$

La règle de combinaison conjonctive doit satisfaire les propriétés de commutativité et d'associativité. Sa définition matricielle (équation (1.59)) met en exergue son aspect linéaire. Il est de plus aisé de prouver que la famille des matrices de spécialisation dempsteriennes est associative et commutative par rapport au produit matriciel.
Nous donnons ci-après des résultats établis par Smets [52] accompagnés de preuves concises.

Théorème 1. *Soient* $\mathbf{S}_1$ *et* $\mathbf{S}_2$ *deux matrices de spécialisation dempsteriennes issues de deux fonctions de masse* $\mathbf{m}_1$ *et* $\mathbf{m}_2$ *respectivement et soit* $\mathbf{S}_{12}$ *la matrice la spécialisation dempsterienne de* $m_{12} = m_1 \textcircled{\cap} m_2$*. On peut alors écrire :*

$$\mathbf{S}_{12} = \mathbf{S}_1.\mathbf{S}_2 = \mathbf{S}_2.\mathbf{S}_1.$$

Démonstration. Soient trois fonctions de masse $\mathbf{m}_1$, $\mathbf{m}_2$ et $\mathbf{m}_3$. Le résultat de leur combinaison conjonctive est donné par :

$$\begin{aligned} \mathbf{m}_{1 \textcircled{\cap} 2 \textcircled{\cap} 3} &= \mathbf{S}_1.\mathbf{m}_{2 \textcircled{\cap} 3}, \\ &= \mathbf{S}_1.\mathbf{S}_2.\mathbf{m}_3. \end{aligned}$$

Or nous savons que la combinaison conjonctive est commutative et associative. Ceci permet d'écrire :

$$\begin{aligned}\mathbf{m}_{1\textcircled{\cap}2\textcircled{\cap}3} &= \mathbf{m}_{2\textcircled{\cap}1\textcircled{\cap}3},\\ &= \mathbf{S}_2.\mathbf{S}_1.\mathbf{m}_3.\end{aligned}$$

Comme ce résultat est vrai pour toute fonction m_3, nous pouvons écrire donc que la matrice de spécialisation dempsterienne de la fonction de masse m_{12} est :

$$\mathbf{S}_{12} = \mathbf{S}_1.\mathbf{S}_2 = \mathbf{S}_2.\mathbf{S}_1.$$

□

Théorème 2. *Soit* $\mathbf{S}_1$ *une matrice de spécialisation dempsterienne et* $\mathbf{Q}_1$ *une matrice de communalité d'une fonction de masse* m_1*. Si* $\mathbf{M}$ *est la matrice d'incidence issue de ce cadre de discernement, on écrit alors :*

$$\mathbf{S}_1 = \mathbf{M}^{-1} * \mathbf{Q}_1 * \mathbf{M}. \tag{1.60}$$

Démonstration. soit $m_{12} = m_1 \textcircled{\cap} m_2$. On peut alors calculer le vecteur de communalité $\mathbf{q}_{12}$ correspondant à $\mathbf{m}_{12}$ comme suit :

$$\begin{aligned}\mathbf{q}_{12} &= \mathbf{M}.\mathbf{m}_{12},\\ &= \mathbf{M}.\mathbf{S}_1.\mathbf{m}_2,\\ &= \mathbf{M}.\mathbf{S}_1.(\mathbf{M}^{-1}.\mathbf{q}_2).\end{aligned}$$

Or, on a aussi par ailleurs :

$$\begin{aligned}\mathbf{q}_{12} &= Diag(\mathbf{q}_1).\mathbf{q}_2,\\ &= \mathbf{Q}_1.\mathbf{q}_2.\end{aligned}$$

Comme ces deux résultats sont vrais pour toute fonction de masse m_2, on a alors

$$\mathbf{S}_1 = \mathbf{M}^{-1} * \mathbf{Q}_1 * \mathbf{M}.$$

□

1.3.3.3 Matrice de spécialisation dempsterienne et affaiblissement

Étant donné une fonction de masse m définie dans un cadre de discernement Ω et $\alpha \in [0,1]$ un coefficient d'affaiblissement. Le vecteur fonction de masse résultant de l'affaiblissement de m par le coefficient α est :

$$\mathbf{m}^\alpha = (1-\alpha).\mathbf{m} + \alpha.\mathbf{m}_\Omega, \tag{1.61}$$

où $\mathbf{m}_\Omega$ est l'ignorance totale définie sur Ω.
Soit alors $\mathbf{q}^\alpha$ la communalité de la fonction de masse affaiblie. Celle-ci se calcule comme suit :

$$\mathbf{q}^\alpha = \mathbf{M}.\mathbf{m}^\alpha = \mathbf{M}.\Big((1-\alpha).\mathbf{m} + \alpha.\mathbf{m}_\Omega\Big).$$

on en déduit donc :

$$\mathbf{q}^\alpha = (1-\alpha).\mathbf{q} + \alpha.\mathbf{q}_\Omega, \tag{1.62}$$

et aussi :

$$\mathbf{Q}^\alpha = (1-\alpha).\mathbf{Q} + \alpha.\mathbf{Q}_\Omega. \tag{1.63}$$

Sachant que la matrice de communalité de l'ignorance totale est égale à la matrice identité de rang $2^{|\Omega|}$, on écrit alors $\mathbf{Q}_\Omega = \mathbf{I}$.

La matrice de spécialisation dempsterienne de la fonction de masse affaiblie $\mathbf{m}^\alpha$ est obtenue par :

$$\begin{aligned}
\mathbf{S}^\alpha &= \mathbf{M}^{-1}.\Big(Diag(\mathbf{M}.\mathbf{m}^\alpha)\Big).\mathbf{M}, \\
&= \mathbf{M}^{-1}.\Big(Diag(\mathbf{q}^\alpha)\Big).\mathbf{M}, \\
&= \mathbf{M}^{-1}.\mathbf{Q}^\alpha.\mathbf{M}, \\
&= \mathbf{M}^{-1}.\Big((1-\alpha).\mathbf{Q} + \alpha.\mathbf{Q}_\Omega\Big).\mathbf{M}, \\
&= (1-\alpha).\mathbf{M}^{-1}.\mathbf{Q}.\mathbf{M} + \alpha.\mathbf{M}^{-1}\mathbf{Q}_\Omega.\mathbf{M}.
\end{aligned}$$

Ce résultat nous permet d'écrire la relation suivante :

$$\mathbf{S}^\alpha = (1-\alpha).\mathbf{S} + \alpha.\mathbf{S}_\Omega, \tag{1.64}$$

où $\mathbf{S}_\Omega = \mathbf{I}$ est la matrice de spécialisation de l'ignorance totale définie sur Ω.

1.3.3.4 Matrices de spécialisation dempsteriennes des fonctions de masse à support simple

Dans la TFC, une fonction de masse à support simple $m = m_X^w$ n'est autre qu'une fonction catégorique $m = m_X$ affaiblie avec le coefficient w. La proposition suivante justifie ce cas particulier et énonce un résultat qui nous est utile pour des démonstrations ultérieures.

Proposition 1. *Supposons $m = m_X^w$ une fonction de masse à support simple définie dans un cadre de discernement Ω engagée sur le sous-ensemble $X \subsetneq \Omega$. Si sa matrice de spécialisation dempsterienne est notée par $\mathbf{S}$, alors :*

$$\mathbf{S} = \bar{w}\mathbf{S}_X + w\mathbf{I}, \tag{1.65}$$

où $\bar{w} = 1 - w$, $\mathbf{S}_X$ est la matrice de spécialisation dempsterienne de m_X, la fonction de croyance catégorique engagée sur X, et $\mathbf{I}$ la matrice identité.

Démonstration. Selon la définition des matrices de spécialisation :

$$S(A, B) = m[B](A),$$

avec $m[B](.)$ la fonction de croyance à support simple m après conditionnement par l'ensemble B. L'utilisation de la formule du conditionnement nous permet d'écrire :

$$S(A, B) = \sum_{\substack{C \subseteq \Omega \\ C \cap B = A}} m(C).$$

Puisque m possède uniquement deux éléments focaux, *i.e.* X et Ω, on a :

$$\begin{aligned} S(A, B) &= 1_{X \cap B = A} m(X) + 1_{\Omega \cap B = A} m(\Omega), \\ &= \bar{w} 1_{X \cap B = A} + w 1_{B = A}, \end{aligned} \tag{1.66}$$

où 1_{Cond} est la fonction indicatrice telle que :

$$1_{Cond} = \begin{cases} 1 & \text{si } Cond \text{ est vraie,} \\ 0 & \text{sinon.} \end{cases}$$

Notons $\mathbf{S}_X$ la matrice de spécialisation de m_X. D'après le lemme 1 nous pouvont écrire :

$$\mathbf{S} = \bar{w}\mathbf{S}_X + w\mathbf{I}.$$

□

1.4 Mesures d'ambiguïté dans la théorie des fonctions de croyance

Dans la théorie de Dempster-Shafer, il coexiste deux types d'ambiguïté, à savoir la discorde et la non-spécificité. Des investigations sont menées par plusieurs auteurs afin de définir une mesure **d'ambiguïté généralisée** dans la théorie des fonctions de croyance. Une telle mesure doit satisfaire les cinq axiomes fondamentaux introduits par Klir et Wierman en 1999 [24].

Dans la théorie classique des ensembles, le seul type de mesure d'ambiguïté défini est la non-spécificité d'un sous-ensemble. La mesure standard de non-spécificité définie dans ce cadre est la mesure d'ambiguïté de Hartley. Etant donné un sous-ensemble A d'un certain cadre de discernement Ω, la mesure de Hartley est définie comme suit :

$$H(A) = \log_2(|A|). \tag{1.67}$$

Dans la théorie des probabilités, ce type de mesure n'est plus utile puisque les distributions de probabilités sont uniquement engagées sur des singletons. Un autre type d'ambiguïté est alors introduit dans l'optique de quantifier la discorde des distributions de probabilités. Cette mesure est connue dans la littérature sous le nom d'**entropie de Shannon**. Elle mesure la distribution de l'incertitude sur l'information délivrée par une source d'informations. L'entropie de Shannon est en effet maximale si toutes les hypothèses sont équiprobables. Elle est définie pour une distribution de probabilité $P = \{p_x : x \in \Omega\}$ par la relation suivante :

$$S(P) = -\sum_{x\in\Omega} p_x \log_2 p_x. \tag{1.68}$$

Les deux mesures d'ambiguïté précédemment définies peuvent servir de base pour définir des mesures d'ambiguïté plus générales dans le cadre de la théorie des fonctions de croyance.

Jousselme *et al.* [18] ont tenté de définir une mesure d'ambiguïté totale à l'aide de la transformation pignistique ($BetP$). Cette mesure d'ambiguïté totale est donnée par la formule :

$$AT(m) = \sum_{x\in\Omega} BetP(x) \log_2(BetP(x)). \tag{1.69}$$

1.4.1 Mesure de la non-spécificité

La non-spécificité dans le cadre de la théorie des fonctions de croyance donne une information sur l'imprécision d'une fonction de masse. Elle est introduite par Dubois et Prade sous le nom de **mesure de Hartley généralisée**.

Etant donné une fonction de masse m définie dans un cadre de discernement Ω. La non-spécificité de la fonction m est définie par la formule :

$$N(m) = \sum_{\substack{A \subseteq \Omega \\ A \neq \emptyset}} m(A) \log_2(|A|). \tag{1.70}$$

$|A|$ désigne le cardinal du sous-ensemble A.

La mesure de Hartley dans la théorie des fonctions de croyance $N(m)$ correspond à la somme pondérée des non-spécificités $H(A)$. Les pondérations sont données par les valeurs des degrés de croyance alloués par la fonction de masse m à chacun des sous-ensembles de Ω. Nous pouvons facilement déduire que la mesure de non-spécificité prend ses valeurs dans l'intervalle $[0, \log_2(|\Omega|)]$.

Afin de pouvoir faire une comparaison de la non-spécificité des fonctions de masse dans un cadre absolu, Smarandache *et al.* [47] ont introduit une normalisation dans la formule (1.70). Ils aboutissent ainsi au résultat suivant :

$$N_S(m) = \sum_{A \subseteq \Omega} m(A) \frac{\log_2(|A|)}{\log_2(|\Omega|)}. \tag{1.71}$$

La fonction $N_S(m)$ tient ainsi ses valeurs dans l'intervalle $[0, 1]$ indépendamment de la taille du cadre de discernement.

1.4.2 La discorde

Plusieurs types de fonctions sont définies afin d'étendre la notion d'entropie de Shannon au cadre de la théorie des fonctions de croyance. A présent, il n'existe pas encore de mesure définitive permettant de quantifier de façon irréfutable la discorde d'une fonction de masse. Par définition, la discorde est la fonction qui mesure le niveau de conflit présent dans l'information modélisée en degrés croyance. Si par exemple, il y a une information qui donne une raison de placer une croyance dans une hypothèse A et qu'en même temps une autre information permet de placer une croyance sur une des autres hypothèses A' antagoniste $A \cap A' = \emptyset$, alors cette information introduit une ambiguïté issue de la contradiction des preuves. C'est ce type d'ambiguïté que mesure la discorde.

La formulation la plus adaptée pour mesurer cette dispersion est directement inspirée de la mesure d'entropie de Shannon :

$$S(m) = -\sum_{A \subseteq \Omega} m(A) \log_2(m(A)). \tag{1.72}$$

avec $m(A) > 0$. D'autres critères de discorde sont aussi introduits dans la littérature [47]. Ils sont basés sur les fonctions de plausibilité, de crédibilité et de probabilité pignistique. Les différents critères d'entropie de Shannon déduits sont respectivement donnés par les relations suivantes :

$$E(m) = -\sum_{A \subseteq \Omega} m(A) \log_2(Pl(A)), \tag{1.73}$$

$$C(m) = -\sum_{A \subseteq \Omega} m(A) \log_2(Bel(A)), \tag{1.74}$$

$$D(m) = -\sum_{A \subseteq \Omega} m(A) \log_2(BetP(A)). \tag{1.75}$$

Etant donné que $Bel(A) \leq BetP(A) \leq Pl(A) \quad \forall A \in 2^{\Omega}$, il est alors possible de déduire que :

$$E(m) \leq D(m) \leq C(m). \tag{1.76}$$

Dans le but d'effectuer une comparaison de plusieurs fonctions de masse, m_j avec $j \in \{1, 2, ..., M\}$, on peut se contenter de choisir un seul de ces différents critères. Le choix du critère approprié est souvent assujetti aux besoins de l'application.

1.5 Conclusion

Dans ce chapitre, les concepts fondamentaux de la théorie des fonctions de croyance ont été présentés. Cette théorie offre un cadre formel pour le traitement et la fusion des informations imprécises et/ou incertaines. C'est un cadre qui généralise à la fois la théorie des probabilités ainsi que la théorie des possibilités. En effet, les fonctions de croyance offrent un outil efficace pour la modélisation de l'incomplétude des informations, la prise en compte de la méconnaissance et de l'imprécision. La notion de fiabilité des informations est aussi prise en compte dans le processus de modélisation. Grâce à cette souplesse dans la gestion de l'information, la TFC est une des théories les plus utilisées dans les processus de fusion d'informations et d'aide à la décision.

Pour la suite de notre travail, nous nous intéresserons à la notion de distance entre les fonctions de croyance. Notamment, dans le prochain chapitre, nous présenterons les différents indices de dissimilarité et distances définis dans dans la littérature. Notre étude est loin d'être exhaustive mais nous nous intéresserons aux indices les plus connus dans la TFC.

Chapitre 2

Distances dans la théorie des fonctions de croyance

Sommaire

2.1 Introduction

D'un point de vue pratique, un calcul de distance dans la TFC a pour but de satisfaire plusieurs objectifs, à savoir l'optimisation et l'évaluation algorithmique pour la classification ou l'approximation de fonctions de masse d'une part, la définition du désaccord entre les fonctions de masse comparées d'une autre part. Dans la littérature, plusieurs approches ont mis en évidence la nécessité de comparer les fonctions de masse à l'aide de distances ou d'indices de dissimilarité pertinents. On peut citer, entre autres, la correction des fonctions de masse avant l'étape de fusion grâce à un processus d'affaiblissement dont les coefficients d'affaiblissement sont obtenus à l'aide d'un calcul de distance.

Au cours des dernières années, de plus en plus de travaux relatifs à la notion de distance ou de dissimilarité entre fonctions de croyance ont vu le jour. Ils s'appuient essentiellement sur une interprétation géométrique des fonctions de croyance. La distance la plus connue dans ce domaine est sans doute la distance de Jousselme [19]. Celle-ci tient compte de la quantification de la similarité entre les éléments focaux grâce aux coefficients de similarité de Jaccard. Jousselme *et al.* [20] ont proposé en 2010 une étude résumant certaines propriétés de la distance issue de l'interprétation géométrique de la théorie des fonctions de croyance. Puis en 2012, ils publient dans [21] un état de l'art très complet sur les distances dans la théorie des fonctions de croyance.

Dans la TFC, la majorité des dissimilarités définies dans la littérature sont fondées sur un produit scalaire, soit en l'exploitant directement, soit à travers une métrique. la famille de métriques la plus utilisée est la famille dite de *Minkowski*. Dans les sections ci-dessous, quatre grandes catégories de dissimilarités entre les fonctions de masse coexistent dans la théorie de Dempster-Shafer. Ces quatre catégories sont :

- les indices directs calculés dans l'espace des fonctions de croyance,
- les indices indirects évalués après transformation des fonctions de masse vers un autre espace de représentation,
- les indices mixtes qui sont des distances composites pouvant être construites à partir de deux ou plusieurs indices élémentaires,
- les indices bidimensionnels qui utilisent conjointement un indice de conflit et une mesure de distance élémentaire afin de définir un vecteur bidimensionnel.

Ces quatre catégories sont répertoriées selon la façon dont les fonctions de masse sont prises en compte dans l'évaluation de la distance qui les sépare ainsi que selon leur nature et leur contenu informationnel.

Ce chapitre est organisé comme suit : la section 2.2 présente les notions générales des distances et de dissimilarité dans la TFC. Les indices de dissimilarité directs sont présentés dans la section 2.3. Les indices de dissimilarité indirects sont traités dans la section 2.4. La section 2.5 est consacrée aux indices de dissimilarité mixtes. Enfin, les

indices de dissimilarité bidimensionnels sont présentés dans la section 2.6.

2.2 Généralités sur les distances évidentielles

D'une manière générale, nous désignons par le terme **distance évidentielle** une distance qui sert à quantifier la différence entre différents éléments de l'espace des fonctions de croyance. Le présent chapitre est dédié aux indices de dissimilarité qui recouvrent les distances entre fonctions de croyance. Néanmoins, la présente section est consacrée aux notions fondamentales relatives aux dissimilarités entre fonctions de croyance.

2.2.1 Interprétation géométrique de la TFC

L'interprétation géométrique de la TFC est une approche qui permet de considérer l'espace des fonctions de croyance comme un espace vectoriel dont les vecteurs de base sont générés par fonctions de masse catégoriques.
Soit $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$ un cadre de discernement contenant n hypothèses exclusives et exhaustives et soit ε_Ω, un espace vectoriel de dimension $N = 2^n$ tel que $\{m_A\}_{A\subseteq\Omega}$ est une base de ε_Ω. Une fonction de masse $\mathbf{m}$ est un vecteur dans l'espace ε_Ω. Chacune des composantes $m(A) \in [0,1],\ A \subseteq \Omega$, constitue la coordonnée du vecteur $\mathbf{m}$ dans la dimension $\mathbf{m_A}$ de l'espace ε_Ω. Ainsi, nous pouvons écrire :

$$\mathbf{m} = \sum_{A\subseteq\Omega} m(A).\mathbf{m_A}. \tag{2.1}$$

On appelle produit scalaire sur l'espace ε_Ω la forme bilinéaire linéaire $\langle ., . \rangle$ satisfaisant les propriétés suivantes :

- (1) Symétrie : $\langle \mathbf{m}_1, \mathbf{m}_2 \rangle = \langle \mathbf{m}_2, \mathbf{m}_1 \rangle$,
- (2) Linéarité : $\langle a.\mathbf{m}_1 + b.\mathbf{m}_2, \mathbf{m}_3 \rangle = a.\langle \mathbf{m}_1, \mathbf{m}_3 \rangle + b.\langle \mathbf{m}_2, \mathbf{m}_3 \rangle$,
- (3) Positivité : $\langle \mathbf{m}, \mathbf{m} \rangle \geq 0 \ \ \forall \mathbf{m} \in \varepsilon_\Omega$, l'égalité est valable pour $\mathbf{m} = 0$,

pour tout $\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3 \in \varepsilon_\Omega$ et $a, b \in \mathbb{R}$.

Une formulation générale du produit scalaire est donnée par l'équation :

$$\langle \mathbf{m}_1, \mathbf{m}_2 \rangle_W = {}^t\mathbf{m}_1.\mathbf{W}.\mathbf{m}_2, \tag{2.2}$$

où $\mathbf{W}$ est une matrice de pondération définie positive.

2.2.2 Distance entre fonctions de masse

En mathématiques, une distance est une application qui formalise l'idée intuitive de distance, c'est-à-dire la longueur du chemin qui sépare deux points dans un espace

donné. Dans l'espace vectoriel ε_Ω, une distance entre fonctions de croyance est définie comme suit :

Définition 5. *Dans l'espace vectoriel ε_Ω, une fonction $d : \varepsilon_\Omega^2 \longrightarrow \mathbb{R}$ est une métrique "complète" si elle satisfait les axiomes suivants :*

- *(1) Non-négativité : $d(m_1, m_2) \geq 0$,*
- *(2) Symétrie : $d(m_1, m_2) = d(m_2, m_1)$,*
- *(3) Réflexivité : $d(m, m) = 0$,*
- *(4) Séparabilité : $d(m_1, m_2) = 0 \Rightarrow m_1 = m_2$,*
- *(5) Inégalité triangulaire : $d(m_1, m_2) \leq d(m_1, m_3) + d(m_2, m_3)$,*

pour tout $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3) \in \varepsilon_\Omega^3$.

L'espace ε_Ω étant muni d'un produit scalaire est un espace normé noté $(\varepsilon_\Omega, \|.\|_W)$. Une distance peut alors être définie à partir de la norme $\|.\|_W$ comme suit :

$$d(m_1, m_2) = \|\mathbf{m}_1 - \mathbf{m}_2\|_W . \tag{2.3}$$

Si la norme du vecteur de masse $\mathbf{m}$ est naturellement définie par $\|\mathbf{m}\|_W = \sqrt{\langle \mathbf{m}, \mathbf{m} \rangle_W}$, une distance entre les éléments de l'espace ε_Ω est donc donnée par :

$$d(m_1, m_2) = \|\mathbf{m}_1 - \mathbf{m}_2\|_W = \sqrt{{}^t(\mathbf{m}_1 - \mathbf{m}_2) W (\mathbf{m}_1 - \mathbf{m}_2)}. \tag{2.4}$$

Si la fonction d ne satisfait pas tous les axiomes précédents, elle n'est alors plus une métrique *"complète"*. Le tableau 2.1 résume les différentes natures de métriques en fonction des axiomes satisfaits.

TABLE 2.1 – Propriétés des différents types de métriques

	Métrique	Semi-métrique	Quasi-métrique	Pseudo-métrique
(1) Non-négativité	x	x	x	x
(2) Symétrie	x	x		x
(3) Réflexivité	x	x	x	x
(4) Séparabilité	x	x	x	
(5) Inégalité triangulaire	x		x	x

2.2.3 Indices de dissimilarité entre fonctions de croyance

Dans la théorie des fonctions de croyance, un indice de dissimilarité peut être issu d'une distance entre les éléments représentant les fonctions de masse en concurrence. Étant donné une fonction $f(x)$ monotone croissante et une distance $d(m_1, m_2)$ normalisée entre deux fonctions de masse $d(m_1, m_2) \in [0, 1]$ tel que :

$$f(0) \leq f\big(d(m_1, m_2)\big) \leq f(1),$$

$$\Rightarrow 0 \leq \frac{f\big(d(m_1, m_2)\big) - f(0)}{f(1) - f(0)} \leq 1. \tag{2.5}$$

La normalisation de la fonction $f(x)$ issue de la formule (2.5) est exploitée pour définir tout un panel d'indices de dissimilarité entre fonctions de masse. Celle-ci peut être définie comme suit :

$$D(m_1, m_2) = \frac{f\big(d(m_1, m_2)\big) - f(0)}{f(1) - f(0)}. \tag{2.6}$$

Comparativement aux distances dans la TFC, un indice de dissimilarité D est une fonction satisfaisant les propriétés suivantes :

- $0 \leq D(m_1, m_2) \leq 1$.
- $D(m_1, m_2) = 0 \Leftrightarrow m_1 = m_2$.
- $D(m_1, m_2) = D(m_2, m_1)$.

Plusieurs choix pour la fonction f définissant D sont possibles, à noter par exemple :

- $f(x) = 1 - exp(-x)$:

$$\Rightarrow D(m_1, m_2) = \frac{1 - exp(-d(m_1, m_2))}{1 - exp(-1)},$$

- $f(x) = \frac{x}{1+x}$:

$$\Rightarrow D(m_1, m_2) = \frac{2.d(m_1, m_2)}{1 + d(m_1, m_2)}.$$

La souplesse dans le choix de la fonction $f(x)$ permet à ce genre d'indices de s'adapter aux besoins particuliers d'un certain cadre applicatif.

Pour la suite de notre travail, nous faisons le choix le plus simple $f(x) = x$. Ce qui revient à dire qu'une distance peut de toute évidence quantifier la dissimilarité entre fonctions de croyance.

2.2.4 Autres propriétés pour les distances évidentielles

Outre que les axiomes et propriétés métriques énoncés dans la définition 5, les propriétés de normalité et les propriétés structurelles sont très souhaitables pour une distance entre fonctions de croyance.

2.2.4.1 Normalité

Lorsqu'une distance est utilisée pour mesurer la similarité/dissimilarité entre fonctions de croyance, celle ci doit être normalisée afin d'avoir une lecture plus directe de la dissimilarité entre les élément comparés. En effet, si la distance après normalisation vaut 1, ce résultat est simplement interprétable par le fait que les fonctions de masse

comparées sont complètement dissimilaires. Bien souvent, la valeur d'une distance non normalisée dépend de la taille du cadre de discernement. Par ailleurs, la normalisation ramène la valeur de la distance dans l'intervalle $[0, 1]$ et permet une interprétation absolue et indépendante des spécificités liées au cadre de discernement.

2.2.4.2 Propriétés structurelles

Une distance entre fonctions de croyance doit satisfaire des propriétés spécifiques à la nature des fonctions de croyance. La structure en treillis de l'espace des fonctions de croyance doit être alors prise en compte dans le calcul de la distance entre fonctions de masse. Dans [21], Jousselme et *al.* ont partagé les propriétés structurelles en trois catégories :

- propriétés structurelles fortes où l'interaction entre les éléments focaux est prise en compte dans le calcul de la distance entre fonctions de masse. Cette interaction peut être par exemple une relation d'ordre partiel,
- propriété structurelle faible qui fait intervenir la cardinalité des éléments focaux dans le calcul de la distance,
- la dissimilarité structurelle est satisfaite si les ensembles des éléments focaux des fonctions de masse comparées sont mises en jeu dans le calcul de la distance. Cette propriété est la plus faible parmi celles énoncées dans [21].

2.3 Indices de dissimilarité directs

Dans les deux sous-sections suivantes, nous allons répartir les indices directs en deux sous-catégories. La première sous-section se consacrera aux indices exploitant directement un produit scalaire entre les vecteurs de fonctions de masse. La deuxième sera dédiée à la sous-catégorie d'indices issus de la famille des distances de *Minkowski*. Ces distances sont des métriques complètes. Ces deux sous-catégories d'indices ont le même fondement mathématique, à savoir la définition d'un produit scalaire mais elles se différencient par leur propriétés mathématiques.

2.3.1 Indices dérivés d'un produit scalaire

La mesure de désaccord la plus évidente dans la théorie des fonctions de croyance, et qui est sans doute la première mesure quantifiant l'interaction entre deux fonctions de masse, est le **degré de conflit** défini par Dempster.
Dans l'espace ε_Ω, le **degré de conflit** de Dempster peut être vu comme un indice issu du produit scalaire tel que :

$$\kappa(m_1, m_2) = m_{\textcircled{\cap}}(\emptyset) = {}^t\mathbf{m}_1.(\mathbf{1} - \mathbf{N}).\mathbf{m}_2, \tag{2.7}$$

où **1** est la matrice *"unité"* de dimension appropriée. La matrice **N** est définie dans l'équation (1.53).
Dans la littérature, on s'accorde à dire que l'indice κ est inadéquat pour décrire une dissimilarité entre les fonctions de croyance. Dans [28] et [34] par exemple, il est souligné que si le conflit interne d'une fonction de masse m_1 n'est pas nul, on aurait alors une valeur non nulle de $\kappa(m_1, m_1) \neq 0$. Cet indice n'est donc pas une métrique complète puisqu'il ne satisfait pas la propriété de réflexivité. En contrepartie, cet indice satisfait les propriétés structurelles énoncées par Jousselme dans [21]. Les interactions entre les éléments focaux du cadre de discernement sont donc pris en compte dans l'évaluation de κ par le fait que la matrice d'intersection des éléments focaux **N** intervient dans son calcul.

Shafer [45] propose un indice de désaccord (*weight of conflict*) d_κ basé sur le degré de conflit de Dempster :

$$d_\kappa(m_1, m_2) = -\log(1 - m_{\textcircled{\cap}}(\emptyset)). \tag{2.8}$$

Contrairement aux autres distances introduites dans la TFC qui prennent leur valeurs dans $[0, 1]$, cet indice possède la particularité de varier dans l'intervalle $[0, +\infty[$. Il possède les mêmes limitations que κ concernant la propriété de réflexivité. En outre, cet indice n'est pas défini dans le cas où $m_{\textcircled{\cap}}(\emptyset) = 1$.

Un autre procédé possible pour quantifier la différence entre les fonctions de masse consiste à utiliser le **cosinus** issu d'un produit scalaire comme indice de dissimilarité. Ce dernier est défini par :

$$cos^d_W(m_1, m_2) = 1 - \frac{\langle \mathbf{m}_1, \mathbf{m}_2 \rangle_W}{(\|\mathbf{m}_1\|_W).(\|\mathbf{m}_2\|_W)}. \tag{2.9}$$

La distance de *Bhattacharyya*, aussi connue comme le coefficient de *fidélité* dans la théorie des probabilités, peut aussi être induite d'un produit scalaire. Cette distance est basée sur la racine carrée de la distribution de probabilité. Son extension aux fonctions de croyance est définie par Ristic et Smets en proposant la distance de *Hellinger*, d_B est donnée par l'équation :

$$d_B(m_1, m_2) = \left(1 - \langle \sqrt{\mathbf{m}_1}, \sqrt{\mathbf{m}_2} \rangle_{\mathbf{I}}\right)^{\frac{1}{2}}, \tag{2.10}$$

où $\sqrt{\mathbf{m}_1}$ est le vecteur dont les éléments sont les racines carrées des éléments de $\mathbf{m}_1$.

2.3.2 Indices issus de la famille des distances de *Minkowski*

La famille des distances de *Minkowski* entre deux fonctions de masse est définie par la forme générique suivante :

$$d_W^{(p)}(m_1, m_2) = ({}^t[(\mathbf{Um}_1 - \mathbf{Um}_2)^{p/2}]\,[(\mathbf{Um}_1 - \mathbf{Um}_2)^{p/2}])^{1/p}, \tag{2.11}$$

où $\mathbf{W} = {}^t\mathbf{U}\,\mathbf{U}$, $p > 1$ un entier non nul, et chaque vecteur $\mathbf{v}^p,\ p \in \mathbb{R}$ est le vecteur dont les composantes sont v_i^p. En choisissant $\mathbf{U} = \mathbf{M}$, par exemple, on obtient la distance de *Minkowski* entre les vecteurs de croyance $\mathbf{Bel_1}$ et $\mathbf{Bel_2}$.
La structure de chaque fonction de masse doit être prise en compte dans la mesure de dissimilarité structurelle exprimée par la matrice $\mathbf{W}$. La distance de *Minkowski* est une métrique généralisée qui inclut plusieurs cas particulier connus de métriques. Théoriquement, une infinité d'indices découlent en faisant varier le paramètre p de l'équation(2.11). Nous citons dans ce qui suit certains cas particuliers :

2.3.2.1 Distances de *Chebychev*

Quand $p \longrightarrow \infty$, on définit alors la classe L_∞ des distances. Une distance appartenant à cette classe est appelée distance de *Chebychev*. Elle est obtenue en évaluant la limite de l'équation (2.11) quand p tend vers ∞. La forme générique de cette famille de distances est définie par :

$$d_W^{(\infty)}(m_1, m_2) = \max_{A \subseteq \Omega} \Big(|{}^t(\mathbf{Um}_1)\mathbf{m_A} - {}^t(\mathbf{Um}_2)\mathbf{m_A}| \Big), \tag{2.12}$$

où $\mathbf{m_A}$ est le vecteur de base de l'espace ε_Ω correspondant au sous-ensemble A.

2.3.2.2 Distances de *Manhattan*

La classe des distances dite de *Manhattan* est obtenue quand cette fois-ci p vaut 1. Cette famille est aussi appelée classe L^1 des distances. Elle est définie par :

$$d_{Bel}^{(1)}(m_1, m_2) = \sum_{A \subseteq \Omega} \Big| {}^t(\mathbf{Um}_1)\mathbf{m_A} - {}^t(\mathbf{Um}_2)\mathbf{m_A} \Big|. \tag{2.13}$$

2.3.2.3 Distances *euclidiennes*

La classe des distances dites *euclidiennes* ($p = 2$) est donnée par l'équation :

$$d_W^{(2)}(m_1, m_2) = \sqrt{\frac{1}{\eta}\,{}^t(\mathbf{m}_1 - \mathbf{m}_2)\,\mathbf{W}\,(\mathbf{m}_1 - \mathbf{m}_2)}, \tag{2.14}$$

où la matrice $\mathbf{W}$ est définie positive et η un coefficient de normalisation.
Le produit scalaire euclidien simple ($\mathbf{W} = \mathbf{I}$, matrice identité) ne fournit pas une mé-

trique très appropriée au formalisme des fonctions de croyance puisqu'il ne tient pas compte des interactions entre les éléments focaux. Une distance plus appropriée à la TFC doit prendre compte d'une part des désaccords entre les sous-ensembles du cadre de discernement Ω, et d'autre part de la distribution de chacune des fonctions de masse comparées sur ses éléments focaux.

En 2001, Jousselme *et al.* [19] ont introduit une distance entre deux fonctions de masse qui satisfait les propriétés structurelles en plus des propriétés métriques. Celle-ci est calculée à l'aide d'un produit scalaire basé sur la matrice de Jaccard $\mathbf{D_J}$ qui permet de tenir compte des différences de cardinalité des éléments focaux mis en jeu lors du calcul de distance. La matrice $\mathbf{D_J}$ contenant les différents indices de Jaccard est définie sur l'ensemble 2^Ω. La distance de Jousselme est alors donnée par :

$$d_J(m_1, m_2) = \sqrt{\frac{1}{2}({}^t\mathbf{m}_1 - \mathbf{m}_2)\, \mathbf{D_J}\, (\mathbf{m}_1 - \mathbf{m}_2)}, \tag{2.15}$$

avec :

$$D_J(A, B) = \begin{cases} 1 \quad \text{si } A = B = \emptyset, \\ \frac{|A \cap B|}{|A \cup B|} \quad \text{sinon.} \end{cases} \tag{2.16}$$

Le facteur $\frac{1}{2}$ garantit le fait que $0 \leq d_J(m_1, m_2) \leq 1$.

Dans [53], Sunberg et Rogers ont défini en 2013 une nouvelle métrique pour décrire le conflit entre les fonctions de masse. Cette métrique est spécialement développée pour mesurer le conflit entre des ensembles ordonnés. Elle utilise une distance de *Hausdorff* pour quantifier la distance entre les éléments focaux. Cette nouvelle métrique notée d_H est définie par :

$$d_H(m_1, m_2) = \sqrt{\frac{1}{2}\, {}^t(\mathbf{m}_1 - \mathbf{m}_2)\, \mathbf{D_H}\, (\mathbf{m}_1 - \mathbf{m}_2)}, \tag{2.17}$$

avec :

$$D_H(A, B) = \frac{1}{1 + KH(A, B)}, \tag{2.18}$$

où $K > 0$ est un paramètre de réglage défini par l'utilisateur. Pour des raisons de simplicité, le choix de $K = 1$ est retenu. $H(A, B)$ est la distance de *Hausdorff* entre les sous-ensembles A et B de Ω sachant que :

$$H(A, B) = \max\{\sup_{a \subseteq A} \inf_{b \subseteq B} d(a, b)\ ,\ \sup_{b \subseteq A} \inf_{a \subseteq B} d(a, b)\},$$

où $d(a, b)$ est la distance entre deux éléments des sous-ensembles. Elle peut être définie par toute métrique définie sur $\Omega \times \Omega$.

Diaz *et al.* [8] ont défini une distance qui pénalise de plus en plus les différences

entre les petites cardinalités comparativement aux cardinalités les plus importantes. La proximité par rapport à l'ignorance totale est ainsi un facteur primordial dans le calcul de la distance de Diaz. Ils proposent l'utilisation d'une fonction de similarité modifiée entre les éléments focaux. Cette fonction dépend de deux paramètres : la matrice de similarité S qui tient compte des dépendances des vecteurs de base de l'espace ε_Ω et le coefficient R qui décrit un seuil de pénalisation des cardinalités par rapport à l'ignorance totale. Diaz *et al.* ont proposé dans [8] différentes matrices de similarité S qui peuvent être utilisées dans le calcul de la distance d_D :

$$d_D(m_1, m_2) = \sqrt{\frac{1}{2}\ {}^t(\mathbf{m}_1 - \mathbf{m}_2)\ \mathbf{F}(\mathbf{S}, R)\ (\mathbf{m}_1 - \mathbf{m}_2)}, \tag{2.19}$$

où $\mathbf{S}$ est la matrice de similarité, R le coefficient qui évalue la proximité de l'ignorance totale et $\mathbf{F}$ une fonction croissante monotone dépendant de $\mathbf{S}$ et R.

D'autres auteurs ont choisi d'autres matrices de similarité pour définir des distances dans l'espace euclidien. Parmi ces matrices, on peut citer par exemple la matrice d'incidence $\mathbf{M}$ ou la matrice de *Dice* $\mathbf{D}$. Dans [21], des expériences montrent que les résultats obtenus par les distances définies dans l'espace euclidien sont en majorité très corrélés avec ceux engendrés par d_J ou d_D. Le tableau 2.2 résume certains des indices directs les plus connus définis dans l'espace euclidien.

TABLE 2.2 – Distances *euclidiennes* définies dans la TFC

Distance	Symbole	W
Fixen et Mahler	d_P	$P(A,B) = \frac{N.\lvert A\cap B\rvert}{\lvert A\rvert.\lvert B\rvert}$
Jousselme	d_J	$D_J(A,B) = \frac{\lvert A\cap B\rvert}{\lvert A\cup B\rvert}$
Dice	d_D	$D(A,B) = \frac{2.\lvert A\cap B\rvert}{\lvert A\rvert+\lvert B\rvert}$
Cuzzolin	d_C	$\mathbf{M}.{}^{\mathbf{t}}\mathbf{M}$

Les indices de dissimilarité directs définis dans le cadre de la théorie des fonctions de croyance utilisent l'information présente dans le vecteur de masse $\mathbf{m}$. Ils sont susceptibles de satisfaire certaines propriétés mathématiques assez importantes en plus des propriétés structurelles spécifiques au cadre de la théorie de Dempster-Shafer. La famille des distances de *Minkowski*, par exemple, permet d'assurer à la fois les propriétés d'une métrique complète ainsi que les propriétés structurelles. Cet avantage dépend d'un bon choix de la matrice $\mathbf{W}$ dans l'équation (2.11) qui doit être définie positive et tenir compte des similarités entre les éléments focaux du cadre de discernement Ω.

2.4 Indices de dissimilarité indirects

Dans cette catégorie d'indices, les fonctions de masse sont généralement transformées en probabilités subjectives ou en mesures floues. Une fois cette transformation assurée, la distance entre fonctions de masse est alors calculée dans le nouvel espace de représentation. La principale motivation de cette transformation est que, tant dans l'espace des probabilités que dans celui des mesures floues, des distances sont bien établies et justifiées dans la littérature. Néanmoins, une telle transformation engendre forcément des pertes d'informations du fait que la dimension du nouvel espace est inférieure à la dimension de l'espace des fonctions de croyance ε_Ω. Ce constat est attendu du fait que la théorie des fonctions de croyance généralise à la fois la théorie des probabilités ainsi que celle des possibilités [9].

2.4.1 Dissimilarités basées sur les probabilités

Le passage au domaine probabiliste se fait principalement via la transformation pignistique. Une distance est alors calculée entre les probabilités pignistiques issues des fonctions de masse transformées. On peut citer à titre d'exemple la distance de Tessem [54] et la distance de Zouhal et Denœux [59]. Ces distances sont toutes les deux des distances de type *Minkowski*. La distance de Tessem correspond à la norme L^∞. Celle-ci est définie par :

$$d_T(m_1, m_2) = \max_{A \subseteq \Omega} \{|BetP_1(A) - BetP_2(A)|\}. \tag{2.20}$$

De même, la distance de Zouhal et Denœux est une distance *euclidienne* et définie comme suit :

$$d_{ZD}(m_1, m_2) = \Big(\frac{1}{2} \sum_{\omega_i \in \Omega} (BetP_1(\omega_i) - BetP_2(\omega_i))^2\Big)^{\frac{1}{2}}. \tag{2.21}$$

2.4.2 Dissimilarités basées sur les fonctions d'appartenance floues

A l'instar des probabilités, la théorie des ensembles flous permet elle aussi de représenter des connaissances sous forme de mesures incertaines. En 2011, Han *et al.* [17] ont défini des indices de dissimilarité indirects basés sur la transformation des fonctions de masse en grandeurs floues. Ils ont pour cela utilisé deux méthodes. La première de ces méthodes consiste à transformer les fonctions de masse en fonctions d'appartenance floues, la seconde à les transformer en ensembles flous intuitionnistes.
Une fonction de masse est transformée en fonction d'appartenance floue en utilisant les plausibilités ou les crédibilités des singletons $\{\omega_i\}$ du cadre de discernement Ω. L'utilisation des fonctions de plausibilité nous permet de transformer un vecteur de masse $\mathbf{m}$

en le vecteur μ suivant :

$$\boldsymbol{\mu} = {}^t[\mu(\omega_1) \;\; \mu(\omega_2) \;\; \cdots \;\; \mu(\omega_n)], \tag{2.22}$$

où deux choix sont possibles : $\mu(\omega_i) = Pl(\{\omega_i\})$ ou-bien $\mu(\omega_i) = Bel(\{\omega_i\})$. L'indice de dissimilarité qui en découle est alors :

$$d_F(m_1, m_2) = 1 - \frac{\sum_{i=1}^{n}(\mu_1(\omega_i) \wedge \mu_2(\omega_i))}{\sum_{i=1}^{n}(\mu_1(\omega_i) \vee \mu_2(\omega_i))}. \tag{2.23}$$

L'opérateur $\wedge$ représente la conjonction (min) et $\vee$ la disjonction (max).
Dans le cas particulier où aucun singleton $\{\omega_i\}$ n'est un élément focal, il est alors nécessaire d'utiliser la plausibilité dans la transformation floue afin d'éviter des divisions par zéro.

2.4.3 Dissimilarités basées sur les ensembles flous intuitionnistes

Les fonctions d'appartenance et de non-appartenance à un ensemble flou intuitionniste d'une fonction de masse m sont définies par :

$$\begin{cases} \mu(\omega_i) = Bel(\{\omega_i\}), \\ \nu(\omega_i) = 1 - Pl(\{\omega_i\}). \end{cases}$$

Le degré d'hésitation de l'élément ω_i dans l'ensemble flou est défini quant à lui comme :

$$\pi(\omega_i) = 1 - \mu(\omega_i) - \nu(\omega_i).$$

Trois indices de dissimilarité sont alors définis dans cette optique. Ils sont exprimés par :

$$\begin{aligned} d_{IF1}(m_1, m_2) &= \frac{1}{2n}\sum_{i=1}^{n}\Big(|\mu_1(\omega_i) - \mu_2(\omega_i)| \\ &\quad + |\nu_1(\omega_i) - \nu_2(\omega_i)| + |\pi_1(\omega_i) - \pi_2(\omega_i)|\Big). \end{aligned} \tag{2.24}$$

$$d_{IF2}(m_1, m_2) = \frac{1}{\sqrt[p]{n}}\sqrt[p]{\sum_{i=1}^{n}\Big(\varphi_1(\omega_i) - \varphi_2(\omega_i)\Big)^p}, \tag{2.25}$$

$$d_{IF3}(m_1, m_2) = \frac{1}{\sqrt[p]{n}}\sqrt[p]{\sum_{i=1}^{n}\Big(\varphi^{\mu}_{1,2}(\omega_i) + \varphi^{\nu}_{1,2}(\omega_i)\Big)^p}, \tag{2.26}$$

où : $\varphi_k(\omega_i) = \frac{\mu_k(\omega_i)+1-\nu_k(\omega_i)}{2}$, $\varphi^{\mu}_{1,2}(\omega_i) = \frac{|\mu_1(\omega_i)-\mu_2(\omega_i)|}{2}$ et $\varphi^{\nu}_{1,2}(\omega_i) = |\frac{1-\nu_1(\omega_i)}{2} - \frac{1-\nu_2(\omega_i)}{2}|$.

Ces trois indices de dissimilarité sont bornés dans l'intervalle $[0, 1]$. Il est évident que

$d_{IF1} = d_{IF2} = d_{IF3} = 0$ quand $m_1 = m_2$.

Les distances indirectes sont, d'un point de vue mathématique, généralement des pseudo-métriques. Ceci est dû à la perte d'informations occasionnée lors de la transformation probabiliste ou floue des fonctions de masse. Sauf pour des cadres applicatifs dédiés, les indices de dissimilarités indirects sont par conséquent peu utilisés comparativement aux indices directs de la famille de *Minkowski* qui offrent une discrimination plus appréciée des fonctions de masse et des propriétés mathématiques plus générales.

2.5 Indices de dissimilarité mixtes

Les indices de dissimilarité mixtes sont des indices définis comme des fonctions d'au moins deux autres indices de dissimilarité élémentaires. Ce choix de stratégie est gourmand en terme de temps de calcul. Néanmoins, son utilisation est sensée apporter à la mesure de dissimilarité un maximum d'avantages contenus dans les indices élémentaires. Des travaux récents ont vu le jour mettant en exergue l'utilisation de deux indices différents afin de mieux discriminer les différences entre les fonctions de masse.

En 2010, Z. Liu *et al.* [29, 30] s'inspirent de la définition quantitative du conflit énoncée par W. Liu en 2006 [28] pour définir un indice de dissimilarité mixte entre les fonctions de masse. Les auteurs justifient leur définition en argumentant que la distance n'est sensible qu'à un seul aspect de toute la différence possible entre deux fonctions de masse comparées. Ils concluent alors que la dissimilarité entre les fonctions de masse doit être impérativement caractérisée par une mesure de distance et une mesure de conflit. Dans le but d'éviter les imperfections du degré de conflit $\kappa = m(\emptyset)$, ils introduisent le **coefficient du conflit** défini comme suit :

$$CP(m_1, m_2) = \begin{cases} 0 & \text{si } X_{max}^{\mathbf{m}_i} = X_{max}^{\mathbf{m}_i} \\ \max\limits_{x\in\Omega} \left(BetP_1(x)\right) \max\limits_{x\in\Omega} \left(BetP_2(x)\right) & \text{sinon.} \end{cases} \tag{2.27}$$

où $X_{max}^{\mathbf{m_i}} = \underset{x\in\Omega}{argmax} \left(BetP_i(x)\right)$.

Conjointement au coefficient du conflit, Z. Liu *et al.* proposent la distance de *Minkowski* définie dans l'espace des probabilités pignistiques pour définir un indice de dissimilarité mixte :

$$d_{BetP}^{(t)}(m_1, m_2) = \left(\frac{1}{2} \sum_{\omega_i\in\Omega} |BetP_1(\omega_i) - BetP_2(\omega_i)|^t\right)^{\frac{1}{t}}. \tag{2.28}$$

La dissimilarité mixte introduite par Z. Liu *et al.* satisfait les propriétés de commutativité et de monotonicité puisque elle est définie en utilisant la T-conorme de *Hamacher*. Elle

est donnée par la formule :

$$d_{Ham}(m_1, m_2) = \frac{d^{(t)}_{BetP}(m_1, m_2) + CP(m_1, m_2)}{1 + d^{(t)}_{BetP}(m_1, m_2).CP(m_1, m_2)}. \tag{2.29}$$

En 2013, Yang *et al.* [58] ont introduit un nouvel indice de dissimilarité mixte basé sur la distance de Jousselme[19] et le degré de conflit de Dempster. Cet indice est défini de la manière suivante :

$$d_Y(m_1, m_2) = \begin{cases} 0 & \text{si } d_J(m_1, m_2) = 0, \\ \left[\frac{1}{2}\left(\left(d_J(m_1, m_2)\right)^2 + \left(\kappa_{12}\right)^2\right)\right]^{\frac{1}{2}} & \text{sinon,} \end{cases} \tag{2.30}$$

où κ_{12} désigne le degré de conflit de Dempster entre m_1 et m_2.
Cet indice de dissimilarité défini dans l'équation (2.30) satisfait les propriétés mathématiques de non négativité, de symétrie et il est borné entre 0 et 1. Ces propriétés sont démontrées dans [58].

Cette même année, Mo *et al.* [36] utilisent conjointement les deux distances définies par Jousselme [19] et Sunberg [53] dans le but de définir une distance mixte qui généralise ces deux dernières tirant profit des avantages de chacune d'entre elles. Cette nouvelle distance est, elle aussi, une distance euclidienne appartenant à la famille de *Minkowski*. Elle est définie comme suit :

$$d_M(m_1, m_2) = \sqrt{\frac{1}{2}\ {}^t(\mathbf{m_1} - \mathbf{m_2})\mathbf{D_M}(\mathbf{m_1} - \mathbf{m_2})}, \tag{2.31}$$

où $\mathbf{D_M} = \alpha\mathbf{D_J} + (\mathbf{1} - \alpha)\mathbf{D_H}$ avec α un coefficient positif compris entre 0 et 1.
Il est à noter que, quand $\alpha = 1$, cette distance est égale à d_J définie dans l'équation (2.15) et quand $\alpha = 0$ celle-ci est alors égale à d_H définie dans l'équation (2.17).

Les indices de dissimilarité mixtes se justifient par la richesse de l'information qu'ils contiennent puisqu'ils sont construits à partir de deux indices de dissimilarité complémentaires différents. Néanmoins, le choix des indices qui les composent reste une question ouverte et dépend des objectifs applicatifs ciblés.

2.6 Indices de dissimilarité bidimensionnels

Dans [28], après avoir analysé le conflit de Dempster ainsi que la distance de Tessem, W. Liu a défini un indice de dissimilarité bidimensionnel du conflit. Selon Liu, ni le conflit de Dempster ($\kappa = m(\emptyset)$) ni la distance ne peuvent définir séparément et de façon complète le désaccord que présentent les fonctions de masse. L'auteur affirme qu'il est raisonnable de conclure que les sources comparées sont en conflit uniquement quand ces

deux indices sont conjointement élevés. Selon Jousselme [21], ce principe bidimensionnel peut être étendu en utilisant n'importe quel couple de dissimilarités entre fonctions de croyance. Par exemple, l'indice de dissimilarité peut se baser d'une part sur une mesure angulaire (issue du produit scalaire) liée aux aspects du conflit et d'autre part sur une distance qui informe sur les aspects métriques entre les vecteurs de fonctions de masse. La formulation générale de ces indices bidimensionnels est :

$$d_{V,W}(m_1, m_2) = \Big(\langle \mathbf{m}_1, \mathbf{m}_2 \rangle_V \ , \ d_W^{(P)}(\mathbf{m}_1, \mathbf{m}_2) \Big). \tag{2.32}$$

Une possible différence entre la matrice $\mathbf{V}$ de base du produit scalaire $\langle ., . \rangle_V$ et la matrice $\mathbf{W}$ de $d_W^{(P)}$ est susceptible d'accroitre la complémentarité des deux indices utilisés. L'apport de chacun d'entre eux est exploité pour avoir une interprétation plus adéquate du conflit entre les sources d'informations dans la théorie des fonctions de croyance.

2.7 Conclusion

Nous avons présenté dans ce chapitre un résumé des indices de dissimilarité les plus connus dans la TFC. Majoritairement, ces indices se basent sur la définition d'un produit scalaire en l'exploitant soit directement soit via la définition d'une distance. Le choix de la matrice de pondération dans la définition du produit scalaire permet de tenir compte des interactions entre les éléments focaux des fonctions de masse comparées. Selon la façon dont les fonctions de masse sont prises en compte dans le calcul de la dissimilarité, quatre grandes catégories sont définies :

- les indices directs qui se basent principalement sur la représentation géométrique de la TFC. Ces indices comparent directement les fonctions de masse dans l'espace vectoriel des fonctions de croyance.
- les indices indirects qui transforment d'abord les fonctions de masse vers un nouvel espace vectoriel dans lequel sont ensuite calculées les dissimilarités.
- les indices mixtes qui sont des fonctions d'au moins deux indices complémentaires différents choisis parmi les indices directs ou indirects.
- les indices bidimensionnels qui à leur tour utilisent deux indices élémentaires dont l'un est un produit scalaire et l'autre est une distance.

Contrairement aux indices de dissimilarité indirects qui sont en général des pseudo-métriques, et aux indices directs exploitant un produit scalaire, les distances de *Minkowski* offrent le plus de propriétés mathématiques. Elles sont en général des métriques complètes.

D'un point de vue mathématique, les indices bidimensionnels qui regroupent une mesure de conflit et une mesure de distance ne sont pas des grandeurs scalaires contrairement

aux indices mixtes. Leur structure vectorielle à deux dimensions rend leur utilisation et leur interprétation plus difficile face aux indices scalaires qui quant à eux se montrent plus intuitifs à interpréter.
Les indices de dissimilarité mixtes ou bidimensionnels tiennent compte de deux aspects différents du désaccord entre fonctions de croyance. Néanmoins ils nécessitent un temps de calcul plus important car ils dépendent inévitablement de deux indices élémentaires.

Dans le prochain chapitre, nous présenterons une nouvelle famille de distances entre les fonctions de croyance capables de tenir compte de l'existence de croyances communes entre les fonctions de masse comparées. Cette nouvelle famille est basée sur la norme matricielle des matrices de spécialisation dempsteriennes.

Chapitre 3

Distances de spécialisation dempsterienne

Sommaire

3.1 Introduction

3.1.1 Motivation

Une distance est utilisée dans la théorie des fonctions de croyance pour décrire la différence entre deux sources d'informations distinctes, ou plus exactement, entre deux jeux de données imparfaites recueillies par ces sources. Plusieurs indices sont définis dans la théorie de Dempster-Shafer afin de bien mettre en évidence les éventuelles dissimilarités entre les fonctions de croyance comparées qui représentent les données collectées sous forme synthétique.
Les indices obtenus ont un sens quand les interactions structurelles entre les sous-ensembles de Ω, telles que les relations d'inclusion, sont sont prises en compte. Par exemple, la distance euclidienne calculée directement entre deux fonctions de masse donne des résultats parfois contre-intuitifs car elle traite indifféremment tout ensemble focal. Cette exigence est une conclusion issue des travaux de Jousselme [21] où les justifications de l'intérêt des propriétés des métriques complètes sont aussi fournies.
Dans notre travail, nous nous intéressons à l'analyse d'un autre aspect inexploré des dissimilarités entre les fonctions de masse : **le pouvoir discriminant de la croyance commune**. En effet, des différences significatives sont observées entre les dissimilarités de fonctions de masse quand ces fonctions partagent des connaissances communes et quand celle-ci ne partagent aucune croyance avec les autres.
Il n'est pas évident de donner une définition d'un tel pouvoir discriminant des distances entre les fonctions de masse. La première étape vers cette définition est de savoir ce que **croyance commune** signifie. La règle de combinaison conjonctive vise à conserver une croyance commune tout en rejetant des croyances contradictoires. Une manière possible de caractériser une croyance commune est donc de comprendre quelles combinaisons conjonctives ont conduit aux deux états de croyances qui doivent être comparés. Supposons que trois fonctions de masse m_1, m_2 et m_3 soient définies dans un cadre de discernement Ω. De façon intuitive, nous remarquons que les fonctions de masse m_{13} et m_{23}, résultats de combinaisons conjonctives, partagent une croyance commune fournie par m_3. Une dissimilarité entre fonctions de croyance d devrait être capable de détecter cette situation dans le sens où l'on doit obtenir le résultat suivant : $d(m_{13}, m_{23}) < d(m_1, m_2)$. Analysons à présent un exemple élaboré dans le but de clarifier cette définition :

Exemple 2. *Étant donné trois fonctions de masse m_1, m_2 et m_3 définies dans le cadre de discernement $\Omega = \{\omega_1, \omega_2\}$ comme le montre le tableau 3.1.*

Les résultats obtenus en utilisant quelques uns des indices de dissimilarité les plus connus dans la littérature sont donnés dans le tableau 3.2.

Pour rappel, ces différents indices de dissimilarité sont définis dans le chapitre 2.

TABLE 3.1 – Differentes masses de croyance traités dans l'example 2

sous-ensemble	$\emptyset$	$\{\omega_1\}$	$\{\omega_2\}$	Ω
$m_1 = m^{0.2}_{\{\omega_1\}}$	0	0.8	0	0.2
$m_2 = m^{0.8}_{\{\omega_2\}}$	0	0	0.2	0.8
$m_3 = m_{\{\omega_2\}}$	0	0	1	0
$m_{13} = m_1 \textcircled{\cap} m_3$	0.8	0	0.2	0
$m_{23} = m_2 \textcircled{\cap} m_3$	0	0	1	0

TABLE 3.2 – Indices de dissimilarité entre m_1, m_2 et entre m_{13}, m_{23} - Exemple 2

Distances	$d(m_1, m_2)$	$d(m_{13}, m_{23})$
d_J	0.5831	0.8
d_T	1	0
d_{ZD}	1	0
d_{IF1}	0.8	0.4
d_{IF2}	0.5	0.4
d_{IF3}	0.5	0.4

Relativement à la définition de la croyance commune que nous avons établie précédemment, nous constatons que la distance de Jousselme est moins discriminante que les autres dissimilarités étudiées dans cet exemple. En regardant les résultats du tableau 3.1 concernant la combinaison conjonctive quand m_3 est combinée avec m_1 et m_2 respectivement, il apparaît que la masse de croyance est réattribuée uniquement au sous-ensemble $\{\omega_2\}$ et à l'ensemble vide. Par conséquent, l'hypothèse ω_1 ainsi que la part d'ignorantce totale sont éliminées par le processus de combinaison. C'est pourquoi nous devrions intuitivement avoir $d(m_{13}, m_{23})$ inférieure à $d(m_1, m_2)$. Il est important de souligner que le comportement de la distance de Jousselme n'est pas incorrect mais critiquable. Premièrement, la discrimination en fonction de la croyance commune n'est pas nécessaire dans tous les contextes applicatifs. Deuxièmement, d'autres critères décrivant l'information commune peuvent être définis et la distance de Jousselme pourrait très bien avoir un comportement satisfaisant à l'égard de ces nouveaux critères.
Concernant d'autres capacités discriminantes, tous les indices indirects sont critiquables par le fait qu'ils ne satisfont pas la propriété de définition d'une métrique. Une valeur nulle est par conséquent obtenue malgré le fait que les fonctions de masse comparées soient différentes. Cette situation est observée dans le tableau 3.2 pour les distances de Tessem d_T et de Zouhal d_{ZD} parce que m_{13} et m_{23} conduisent à la même décision au vu de leurs probabilités pignistiques.
Par conséquent, les remarques précédentes tendent à montrer que la recherche d'une distance entre fonctions de masse avec un pouvoir discriminant accru à l'égard des croyances communes est justifié et reste une tâche difficile à réaliser.
Dans ce qui suit, le problème de la conception de distances entre fonctions de masse

est formalisé par la définition de propriétés mathématiques. De nouvelles distances sont ensuite introduites dans la section 3.2 et l'évaluation de leurs pouvoirs discriminants comparée aux approches existantes est présentée dans la section 3.4.

3.1.2 Formulation du problème

Avant de définir de nouvelles distances entre fonctions de croyance, nous devons clairement établir les propriétés que ces distances entre fonctions de croyance doivent respecter. Ci-dessous, nous présentons trois principes que nous considérons d'une importance primordiale :

- (i) : deux fonctions de masse sont complètement similaires si elles représentent des états de connaissance identiques au regard de la solution au problème.
- (ii) : une distance entre fonctions de masse doit prendre en compte les dépendances entre les éléments de 2^{Ω} (même principe que la propriété structurelle forte de Jousselme [21]).
- (iii) : deux fonctions de masse qui soutiennent des croyances communes devraient être plus proches que des fonctions de masse engagées sur des hypothèses disjointes.

Les trois conditions précédentes sont traduites en propriétés mathématiques principales pour la conception d'une nouvelle distance. Le premier principe est assuré si la propriété de **définition** des métriques est vérifiée, ce qui signifie que les distances évidentielles doivent être des métriques complètes. Concernant le second principe, il n'existe pas de traduction formelle en propriété mathématique. Beaucoup de propriétés pourraient être considérées afin d'assurer cet objectif. Nous proposons d'utiliser la définition suivante à cet effet :

Définition 6. *Soit d_{set} une distance d'ensemble et d une distance entre fonctions de croyance. d est dite* ***structurée*** *relativement à d_{set} si, $\forall A, B, C, D \subseteq \Omega$, on a :*

$$d_{set}(A, B) \leq d_{set}(C, D) \Leftrightarrow d(m_A, m_B) \leq d(m_C, m_D). \tag{3.1}$$

Quand l'inégalité est stricte, la distance d est dite ***strictement structurée****.*

Dans le cadre la théorie des fonctions de croyance, l'intérêt de la propriété structurelle dépend du choix de la distance d_{set} qui est primordial. Par exemple, il est justifié de choisir pour d_{set} les distances de *Jaccard* et de *Hamming* car ces distances d'ensembles tiennent compte de l'interaction entre les éléments focaux de Ω.

Cette définition se justifie par le fait que la théorie des fonctions de croyance généralise la théorie des ensembles de *Cantor*. Par conséquent, une distance entre fonctions

de croyance se doit de généraliser la distance entre les ensembles.

La troisième condition trouve sa justification dans les connexions qui existent entre la dissimilarité entre fonctions de croyance et la règle de combinaison conjonctive :

Définition 7. *Étant donné d, une distance entre fonctions de masse dans un cadre de discernement Ω, d est dite* ***consistante conjonctivement*** *si pour toutes fonctions de masse m_1, m_2 et m_3 définies dans Ω, on a :*

$$d\left(m_1 \textcircled{\cap} m_3, m_2 \textcircled{\cap} m_3\right) \leq d\left(m_1, m_2\right). \tag{3.2}$$

La propriété précédente peut être interprétée comme suit : si la même croyance est incorporée dans deux fonctions de masse, ces deux dernières doivent être plus proches après combinaison qu'avant.
Étant donné que les matrices de spécialisation dempsteriennes sont étroitement liées à la règle de combinaison conjonctive, la recherche de nouvelles dissimilarités entre fonctions de croyance utilisant ces matrices s'avère une idée d'une importance primordiale. Cette idée sera développée au long des sections suivantes.

3.2 Nouvelles distances basées sur les matrices de spécialisation dempsteriennes

3.2.1 Rappel sur les matrices de spécialisation dempsteriennes

On rappelle que le concept de spécialisation est fondé sur la redistribution de la masse de croyance après l'ajout de nouveaux éléments d'informations. En effet, l'ajout de nouvelles connaissances modifie la connaissance initiale du problème et produit une fonction de masse au moins autant engagée que la fonction initiale [22, 37]. Il est alors possible de stocker le transfert de masse dans une matrice appelée matrice de spécialisation.
Si par exemple une fonction de masse m_1 est une spécialisation de m_0. Elle peut s'écrire sous une forme matricielle comme suit :

$$\mathbf{m_1} = \mathbf{S}.\mathbf{m_0},$$

Dans un processus de combinaison conjonctive, le transfert de masse est aussi considéré à travers la règle de conditionnement de Dempster. La quantité $(m \textcircled{\cap} m_B)(A)$ est la proportion maximale de masse qui doit être transférée du sous-ensemble B vers $A \subset B$ si une source d'informations qui croit en B est à combiner avec m. Selon ce transfert, la matrice de spécialisation dempsterienne $\mathbf{S}$ issue de la fonction de masse m est unique et

ses éléments sont donnés par :

$$S(A,B) = (m_{\textcircled{\cap}} m_B)(A), \quad \forall A, B \subset \Omega. \tag{3.3}$$

Il est aussi à rappeler que la matrice de spécialisation dempsterienne d'une fonction de masse ne décrit pas uniquement l'état de la connaissance à un instant donné mais également tous les états futurs qui peuvent être atteints à partir de celle-ci au travers de tous les conditionnements possibles de la fonction de masse qui lui correspond.

3.2.2 Nouvelles distances dans la théorie des fonctions de croyance

Une nouvelle famille de distances entre fonctions de croyance est introduite dans ce travail. Cette nouvelle famille est obtenue en comparant les fonctions de masse à travers leurs matrices de spécialisation dempsteriennes puisqu'il existe une correspondance bijective entre chaque fonction de croyance et sa matrice de spécialisation dempsterienne (regarder la section 1.3.3). En outre, lorsqu'une combinaison conjonctive est utilisée pour fusionner les informations issues de sources diverses, la façon dont une fonction de masse interagit avec les autres est décrite dans sa matrice de spécialisation dempsterienne qui donne une information non seulement sur l'état actuel de la connaissance mais aussi sur tous les états futurs. Notre idée est donc de comparer les fonctions de masse en utilisant leurs matrices de spécialisation dempsteriennes.

3.2.2.1 Généralités sur les normes matricielles

La norme $\mathcal{N}(x)$ est une application sur un espace métrique qui permet d'introduire une distance entre les éléments de l'espace en question. Dans un espace normé, une distance est définie par la norme de la différence entre deux éléments de cet espace :

$$d(x,y) = \mathcal{N}(x-y). \tag{3.4}$$

Notons $\mathcal{M}_N$ l'ensemble des matrices carrées réelles de dimension $N \times N$ doté des propriétés algébriques d'un espace vectoriel. Par conséquent, les distances matricielles ne sont pas très différentes des distance vectorielles. A la suite de ce travail, nous nous intéressons aux distances induites par des normes matricielles. Notons aussi que l'ensemble des matrices de spécialisation dempsteriennes est clairement un sous-ensemble de $\mathcal{M}_N$. Rappelons à présent la définition d'une norme matricielle :

Définition 8. *Dans le K-espace vectoriel $\mathcal{M}_N$ des matrices carrées de dimension $N \times N$, une norme matricielle $\|.\|$ est une application définie de $\mathcal{M}_N \longrightarrow \mathbb{R}_+$ satisfaisant les conditions :*

1. $\|\mathbf{A}\| = \mathbf{0} \Leftrightarrow \mathbf{A} = \mathbf{0}$,
2. $\|\lambda\mathbf{A}\| = |\lambda|.\|\mathbf{A}\|$ *pour tout* $\lambda \in K$,
3. $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$,

Une norme matricielle est dite sous-multiplicative si, en plus, elle satisfait la propriété suivante :

– $\|\mathbf{A}.\mathbf{B}\| \leq \|\mathbf{A}\| . \|\mathbf{B}\|$.

Intéressons nous dans notre travail aux deux familles les plus connues de normes matricielles : les normes d'opérateur (normes subordonnées) et les normes. On sous-entend par norme vectorielle d'une matrice la norme de vecteur calculée sur le vecteur colonne issue de la matrice de dimension $N \times N$ transformée et dont tous les éléments sont stockés dans un vecteur colonne de dimension $(N \times N) \times 1$.

Les normes d'opérateur k, notées $\|.\|_{opk}$, sont définies pour chaque matrice $\mathbf{A} \in \mathcal{M}_N$ s'il existe une norme vectorielle $\|\mathbf{x}\|_k$, $\mathbf{x} \in \mathbb{R}^N$ telle que :

$$\|\mathbf{A}\|_{opk} = \max_{\mathbf{x}\in\mathbb{R}^N, \mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_k}{\|\mathbf{x}\|_k} = \max_{\mathbf{x}\in\mathbb{R}^N, \|\mathbf{x}\|_k=1} \|\mathbf{A}\mathbf{x}\|_k \,, \tag{3.5}$$

avec $\|\mathbf{x}\|_k$ la norme L^k vectorielle classique du vecteur $\mathbf{x}$.
Nous pouvons aisément constater que de chaque norme vectorielle k découle une norme matricielle subordonnée. Trois cas particuliers sont considérés dans la suite de notre travail, à savoir : la norme d'opérateur-1, la norme d'opérateur-2 et la norme d'opérateur-∞. Elles sont définies par les formules suivantes :

- norme d'opérateur-1 :

$$\|\mathbf{A}\|_{op1} = \max_{1<j<n} \sum_{1<i<n} |\mathbf{A}_{ij}|. \tag{3.6}$$

  L'application de cette norme à n'importe quelle matrice de spécialisation dempsterienne $\mathbf{S}$ donne $\|\mathbf{S}\|_{op1} = 1$.
- norme d'opérateur-2 :

$$\|\mathbf{A}\|_{op2} = \sqrt{Spec(\mathbf{A}^* \times \mathbf{A})}. \tag{3.7}$$

  où $Spec(\mathbf{A}^* \times \mathbf{A})$ est le rayon spectral de $(\mathbf{A}^* \times \mathbf{A})$ sachant que $\mathbf{A}^*$ est la matrice adjointe de la matrice $\mathbf{A}$. Autrement dit $\|\mathbf{A}\|_{op2}$ est la valeur propre maximale de A.

- norme d'opérateur-∞ :

$$\|\mathbf{A}\|_{op\infty} = \max_{1<i<n} \sum_{1<j<n} |\mathbf{A}_{ij}|. \tag{3.8}$$

Les normes vectorielles L^k appliquées à des matrices sont notées par $\|.\|_k$ et définies comme suit :

$$\|\mathbf{A}\|_k = \left(\sum_{1\leq i,j\leq 2^{|\Omega|}} |\mathbf{A}_{ij}|^k \right)^{\frac{1}{k}}. \tag{3.9}$$

On fait aisément la distinction entre une norme L^k appliquée à un vecteur ou une matrice par le symbole mathématique associé qui est en minuscule ou en majuscule.

3.2.2.2 Distances basés sur les matrices de spécialisation

La famille des distances basées sur les matrices de spécialisation dans la TFC est une nouvelle famille de distances que nous définissons entre fonctions de masse. Cette nouvelle famille calcule la distance entre fonctions de masse sur la base d'une norme matricielle entre les matrices de spécialisation des fonctions de masse concernées.

Définition 9. *Soient deux fonctions de masse m_1 et m_2 définies dans un cadre de discernement Ω et de matrices de spécialisation dempsteriennes respectives $\mathbf{S}_1$ et $\mathbf{S}_2$. Nous appelons* ***distance de spécialisation*** *d_x entre m_1 et m_2 la distance définie à partir d'une norme matricielle donnée $\|.\|_x$, à savoir :*

$$d_x(m_1, m_2) = \frac{1}{\rho} \|\mathbf{S}_1 - \mathbf{S}_2\|_x, \tag{3.10}$$

où $\rho = \max_{m_i,m_j} \|\mathbf{S}_i - \mathbf{S}_j\|_x$ est un coefficient de normalisation.

Pour les distances de spécialisation matricielles de type L^k, le coefficient de normalisation est donné par :

$$\rho = \left(2\left(2^n - 1\right)\right)^{\frac{1}{k}}. \tag{3.11}$$

Démonstration. La preuve de ce résultat est données dans l'annexe A.1. □

Concernant les distances de spécialisation basées sur les normes subordonnées, des résultats de normalisation partiels sont regroupés dans le tableau 3.3.

TABLE 3.3 – Coefficient de normalisation des distances de spécialisation subordonnées (d'opérateur).

Distances	ρ
d_{Sop1}	2
d_{Sop2}	$\sqrt{2^n}$
$d_{Sop\infty}$	$2^n - 1$

3.2.3 Propriétés principales des distances de spécialisation

Après avoir défini une nouvelle famille de distances dans la théorie des fonctions de croyance, il est à présent impératif de déterminer leurs principales propriétés. Ces propriétés sont nécessaires pour pouvoir statuer sur le comportement que ces nouvelles distances établies vont adopter dans des situations pratiques.

3.2.3.1 Propriétés métriques

Proposition 2. *Toutes les distances de spécialisation sont des métriques complètes normalisées.*

Démonstration. Le fait que les distances de spécialisation soient définies à partir de normes matricielles implique que ces distances sont des métriques complètes. Concernant les distances L^k, l'équation (3.11) montre qu'elles sont bien normalisées. Selon l'équation (3.5), il est clair que les normes d'opérateur sont des sommes finies d'éléments de matrices finis. Ceci implique que $\forall k$, $d_{opk}(m_1, m_2) < \infty$ et par conséquent, il est toujours possible de trouver un facteur de normalisation. □

3.2.3.2 Propriété structurelle

Intéressons nous à présent à l'analyse de la propriété structurelle des distances de spécialisation entre fonctions de masse. On a la proposition suivante :

Proposition 3. *$\forall k < \infty$, toute distance de spécialisation L^k est une distance structurée.*

Démonstration. Pour la démonstration de ce résultat, consulter l'annexe A.2. □

Dans la démonstration de ce résultat, il est clair que les distances de spécialisation L^k généralisent la distance d'ensembles de Hamming. Par contre, il n'existe pas de distances de spécialisation d'opérateur structurée [1].

Pour la distance d_{op1}, ce résultat est évident puisqu'elle est aussi définie par la formule (3.12) donnée dans le lemme suivant :

1. Ces distances ne généralisent ni la distance de Hamming, ni la distance de Jaccard

Lemme 2. *La distance de spécialisation d'opérateur-1 d_{op1} est égale à la distance de Manhattan (L^1) entre les vecteurs des fonctions de masse :*

$$d_{op1}(m_1, m_2) = \frac{1}{2} \|\mathbf{m}_1 - \mathbf{m}_2\|_1 . \tag{3.12}$$

Démonstration. La démonstration de ce résultat est dans l'annexe A.3. □

A la vue de ce résultat, on voit bien que d_{op1} n'a aucune chance de prendre en compte les interactions entre éléments focaux.

3.2.3.3 Propriété de consistance conjonctive

Concernant la propriété de consistance conjonctive, trois résultats sont disponibles. Nous les énonçons dans les deux propositions 4,5 et 6 ci-dessous :

Proposition 4. *La distance de spécialisation d'opérateur-1 d_{op1} est consistante avec la règle conjonctive.*

Démonstration. Supposons m_1, m_2 et m_3 trois fonctions de masse définies sur Ω. Leurs matrices de spécialisation dempsteriennes respectives sont notées $\mathbf{S}_1$, $\mathbf{S}_2$ et $\mathbf{S}_3$. La norme d'opérateur-1 possède la propriété sous-multiplicative, c'est à dire que quelles que soient les matrices $\mathbf{A}$ et $\mathbf{B}$, on a :

$$\|\mathbf{AB}\|_{op1} \leq \|\mathbf{A}\|_{op1} . \|\mathbf{B}\|_{op1} . \tag{3.13}$$

On peut alors écrire :

$$\begin{aligned} \|(\mathbf{S}_1 - \mathbf{S}_2)\,\mathbf{S}_3\|_{op1} &\leq \|\mathbf{S}_1 - \mathbf{S}_2\|_{op1} . \|\mathbf{S}_3\|_{op1} , \\ \Leftrightarrow d_{op1}(m_1 \textcircled{\cap} m_3, m_2 \textcircled{\cap} m_3) &\leq d_{op1}(m_1, m_2) . \|\mathbf{S}_3\|_{op1} . \end{aligned}$$

Par définition de cette norme et des matrices de spécialisation, on a déjà vu que $\|\mathbf{S}_3\|_{op1} = 1$.
Par conséquent, la distance d_{op1} est conjonctivement consistante, et l'on peut alors écrire :

$$d_{op1}(m_1 \textcircled{\cap} m_3, m_2 \textcircled{\cap} m_3) \leq d_{op1}(m_1, m_2) .$$

□

Proposition 5. *La distance L^1 de spécialisation d_1 est consistante avec la règle conjonctive.*

Démonstration. Supposons m_1, m_2 et m_3 trois fonctions de masse définies dans Ω. Le lemme 2 implique :

$$\begin{aligned} \|(\mathbf{S}_1 - \mathbf{S}_2)\,\mathbf{S}_3\|_{op1} &= \|(\mathbf{m}_1 - \mathbf{m}_2)\,\textcircled{\cap}\mathbf{m}_3\|_1\,, \\ \text{et } \|(\mathbf{S}_1 - \mathbf{S}_2)\|_{op1} &= \|(\mathbf{m}_1 - \mathbf{m}_2)\|_1\,. \end{aligned}$$

Selon la proposition 4, nous pouvons établir que :

$$\|(\mathbf{m}_1 - \mathbf{m}_2)\,\textcircled{\cap}\mathbf{m}_3\|_1 \leq \|(\mathbf{m}_1 - \mathbf{m}_2)\|_1\,. \tag{3.14}$$

Par ailleurs, la norme 1 peut aussi être calculée en utilisant les vecteurs colonnes comme suit :

$$\|(\mathbf{S}_1 - \mathbf{S}_2)\,\mathbf{S}_3\|_1 = \sum_{A\subseteq\Omega} \|(\mathbf{m}_1\textcircled{\cap}\mathbf{m}_A - \mathbf{m}_2\textcircled{\cap}\mathbf{m}_A)\,\textcircled{\cap}\mathbf{m}_3\|_1\,.$$

Utilisons à présent l'équation (3.14) pour obtenir :

$$\begin{aligned} \|(\mathbf{S}_1 - \mathbf{S}_2)\,\mathbf{S}_3\|_1 &= \sum_{A\subseteq\Omega} \|(\mathbf{m}_1\textcircled{\cap}\mathbf{m}_A - \mathbf{m}_2\textcircled{\cap}\mathbf{m}_A)\,\textcircled{\cap}\mathbf{m}_3\|_1\,, \\ \|(\mathbf{S}_1 - \mathbf{S}_2)\,\mathbf{S}_3\|_1 &\leq \sum_{A\subseteq\Omega} \|(\mathbf{m}_1\textcircled{\cap}\mathbf{m}_A - \mathbf{m}_2\textcircled{\cap}\mathbf{m}_A)\|_1\,, \\ \|(\mathbf{S}_1 - \mathbf{S}_2)\,\mathbf{S}_3\|_1 &\leq \|(\mathbf{S}_1 - \mathbf{S}_2)\|_1\,, \\ \Leftrightarrow d_1\,(m_1\textcircled{\cap}m_3, m_2\textcircled{\cap}m_3) &\leq d_1\,(m_1, m_2)\,. \end{aligned}$$

□

Proposition 6. *La distance L^∞ de spécialisation $d_{S\infty}$ est consistante avec la règle conjonctive.*

Démonstration. Selon la propriété de sous-multiplicativité des normes matricielles, nous écrivons

$$\|(\mathbf{S}_1 - \mathbf{S}_2)\,\mathbf{S}_3\|_\infty \leq \|(\mathbf{S}_1 - \mathbf{S}_2)\|_\infty \|\mathbf{S}_3\|_\infty\,,$$

or nous savons que : $\|\mathbf{S}_3\|_\infty \leq 1$, car $\|\mathbf{S}_3\|_\infty = \max_{\{m[B](A)\}}$,d'où :

$$\begin{aligned} \|(\mathbf{S}_1 - \mathbf{S}_2)\,\mathbf{S}_3\|_\infty &\leq \|(\mathbf{S}_1 - \mathbf{S}_2)\|_\infty\,, \\ \Leftrightarrow d_{S\infty}\,(m_1\textcircled{\cap}m_3, m_2\textcircled{\cap}m_3) &\leq d_{S\infty}\,(m_1, m_2)\,. \end{aligned}$$

□

3.2.3.4 Rapport de croyance des distances de spécialisation L^k

Outre que les propriétés métriques, structurelle et de consistance conjonctive, les distances L^k possède d'autres propriétés pertinentes. Dans cette sous-section, le lien entre le rapport de croyance des fonctions de masse à support simple engagées sur un même sous ensemble et les distances L^k est illustré. Il est énoncé dans la proposition ci-après.

Proposition 7. *Supposons m_1 une fonction de masse définie dans Ω. Si $m_2 = m_A^{w_2}$ et $m_3 = m_A^{1-a+aw_2}$ sont deux fonctions de masse à support simple définies dans Ω et engagées sur le sous-ensemble A, la relation suivante est alors vérifiée pour les distances de spécialisation du type L^k notées d_k :*

$$\frac{d_k\left(m_1, m_1 \textcircled{\cap} m_2\right)}{d_k\left(m_1, m_1 \textcircled{\cap} m_3\right)} = \frac{m_2\left(A\right)}{m_3\left(A\right)} = a, \tag{3.15}$$

où $a \in [0,1]$.

Démonstration. Soit $\mathbf{S}_i$ la matrice de spécialisation dempsterienne de la fonction de masse à support simple m_i. Il est clair que si $\mathbf{S}_{12}$ désigne la matrice de spécialisation dempsterienne de la fonction de masse $m_1 \textcircled{\cap} m_2$, on peut alors écrire $\mathbf{S}_{12} = \mathbf{S}_1.\mathbf{S}_2$. Par conséquent, on aboutit au résultat suivant :

$$\begin{aligned} \mathbf{S}_1 - \mathbf{S}_{12} &= \mathbf{S}_1 - \mathbf{S}_1\mathbf{S}_2, \\ \mathbf{S}_1 - \mathbf{S}_{12} &= \mathbf{S}_1\left(\mathbf{I} - \mathbf{S}_2\right), \end{aligned} \tag{3.16}$$

avec $\mathbf{I}$ la matrice identité.
La proposition 1 implique que si m_2 est une fonction de masse à support simple, alors $\mathbf{S}_2 = \bar{w}_2\mathbf{S}_A + w_2\mathbf{I}$ avec $\mathbf{S}_A$ est la matrice de spécialisation dempsterienne de la fonction de croyance catégorique engagée sur A.
L'utilisation de ce résultat dans l'équation (3.16) implique :

$$\begin{aligned} \mathbf{S}_1 - \mathbf{S}_{12} &= \mathbf{S}_1\left(\mathbf{I} - \bar{w}_2\mathbf{S}_{A_2} - w_2\mathbf{I}\right), \\ \mathbf{S}_1 - \mathbf{S}_{12} &= \mathbf{S}_1\left(\bar{w}_2\mathbf{I} - \bar{w}_2\mathbf{S}_{A_2}\right), \\ \mathbf{S}_1 - \mathbf{S}_{12} &= \bar{w}_2\left(\mathbf{S}_1 - \mathbf{S}_1\mathbf{S}_{A_2}\right). \end{aligned} \tag{3.17}$$

Une conséquence immédiate de l'équation (3.17) est :

$$d\left(m_1, m_1 \textcircled{\cap} m_2\right) = \bar{w}_2 d\left(m_1, m_1 \textcircled{\cap} m_A\right). \tag{3.18}$$

Ce résultat peut aussi être appliqué à la fonction de croyance m_3 comme suit :

$$\begin{aligned} d_k\left(m_1, m_1 \textcircled{\cap} m_3\right) &= m_3\left(A\right) d_k\left(m_1, m_1 \textcircled{\cap} m_A\right), \\ d_k\left(m_1, m_1 \textcircled{\cap} m_3\right) &= \left(1-\left(1-a+aw_2\right)\right) d_{Sk}\left(m_1, m_1 \textcircled{\cap} m_A\right), \\ d_k\left(m_1, m_1 \textcircled{\cap} m_3\right) &= a\bar{w}_2 d_k\left(m_1, m_1 \textcircled{\cap} m_A\right). \end{aligned} \tag{3.19}$$

Finalement, la division de chaque terme de l'équation (3.19) par ceux de (3.18) donne :

$$\frac{d_k\left(m_1, m_1 \textcircled{\cap} m_2\right)}{d_k\left(m_1, m_1 \textcircled{\cap} m_3\right)} = \frac{m_2\left(A\right)}{m_3\left(A\right)} = a.$$

□

3.2.3.5 Interprétation de la valeur maximale des distances d_k

La signification de la valeur maximale d'une distance entre fonctions de croyance est une notion pertinente puisqu'elle permet de caractériser la distance en définissant la dissimilarité totale. Pour notre cas, la valeur maximale de la distance d_k n'est atteinte que dans le cas où les fonctions comparées sont en négation l'une par rapport à l'autre. Cette propriété est énoncée dans la proposition suivante :

Proposition 8. *Soit un cadre de discernement Ω et un élément $A \in 2^{\Omega}$. La distance d_k entre deux fonctions de masse m_1 et m_2 admet :*

$$d_k\left(m_1, m_2\right) = 1 \Leftrightarrow m_1 = m_A \text{ et } m_2 = m_{\overline{A}}. \tag{3.20}$$

Démonstration. La démonstration de cette propriété est donnée dans l'annexe A.4. □

La démonstration de la propriété précédente nous permet d'affirmer que la mesure de distance $d_k(m_1, m_2)$ est maximale et égale à 1 uniquement quand la fonction de masse m_1 est la négation de m_2 et que toutes deux sont des fonctions de masse catégoriques : $\exists A \subseteq \Omega : \overline{m}_1(A) = m_2(\overline{A}) = 1$.
La distance de Jousselme ainsi que la distance de Tessem sont égales à 1 quand les deux fonctions de masse comparées sont catégoriques et complètement conflictuelles : $m_1(A_1) = m_2(A_2) = 1$ et $A_1 \cap A_2 = \emptyset$. La spécificité des éléments focaux comparés n'est pas prise en compte dans ce cas.

3.2.3.6 Distance de spécialisation d_k et affaiblissement

Nous présentons dans cette sous-section le lien existant entre les distances d_k et la notion de l'affaiblissement des fonctions de croyance. Deux résultats sont présentés, le

premier concerne la distance entre deux fonctions de masse affaiblies avec un même coefficient et le deuxième se consacre à la distance entre une fonction de masse quelconque et une autre affaiblie.

Proposition 9. *Étant donné deux fonctions de masse m_1 et m_2 affaiblies avec un même coefficient d'affaiblissement $\alpha \in [0,1]$, la relation suivante est alors vraie pour toute distance d_k :*

$$d_k(m_1^\alpha, m_2^\alpha) = (1-\alpha).d_k(m_1, m_2) \leq d_k(m_1, m_2).$$

Démonstration. Par définition, la distance d_k des deux fonctions de masse affaiblies correspond à :

$$\begin{aligned} d_k(m_1^\alpha, m_2^\alpha) &= \frac{1}{\rho} \left\| \mathbf{S}_1^\alpha - \mathbf{S}_2^\alpha \right\|_k , \\ &= \frac{1}{\rho} \left\| \Big((1-\alpha)\mathbf{S}_1 + \alpha\mathbf{I}\Big) - \Big((1-\alpha)\mathbf{S}_2 + \alpha\mathbf{I}\Big) \right\|_k , \\ &= \frac{1}{\rho} \left\| (1-\alpha)\Big(\mathbf{S}_1 - \mathbf{S}_2\Big) \right\|_k , \\ &= (1-\alpha) d_k(m_1, m_2). \end{aligned}$$

De plus, nous savons que $\alpha \in [0,1]$. Ainsi, nous aboutissons au résultat :

$$d_k(m_1^\alpha, m_2^\alpha) \leq d_k(m_1, m_2). \tag{3.21}$$

□

Proposition 10. *Étant donné un cadre de discernement Ω, deux fonctions de masse définies dans Ω et un coefficient d'affaiblissement $\alpha \in [0,1]$, alors :*

1. $d_k({}^\alpha m_1, m_2) \leq (1-\alpha).d_k(m_1, m_2) + \alpha.d_k(m_2, m_\Omega)$,
2. $(1-\alpha).d_k(m_1, m_2) - \alpha.d_k(m_2, m_\Omega) \leq d_k({}^\alpha m_1, m_2)$.

Démonstration. La démonstration de la proposition 10 est donnée dans l'annexe A.5. □

L'interprétation graphique de la propriété précédente est illustrée dans la figure (3.1).

La figure (3.1) visualise les résultats énoncé dans la proposition 10. Comme nous pouvons le voir sur cette figure, la distance entre une fonction de masse affaiblie avec un coefficient α et une autre définit une forme d'inégalité triangulaire en impliquant la fonction d'ignorance totale qui est le résultat d'un affaiblissement complet. Cette inégalité triangulaire est pondérée par les coefficients α et $1-\alpha$.

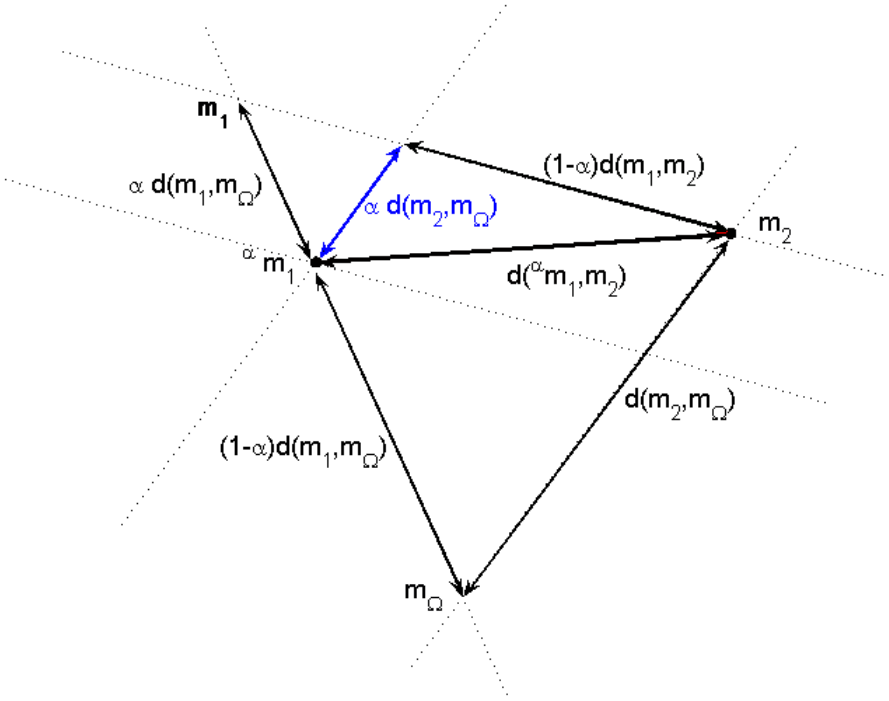

FIGURE 3.1 – Interprétation géométrique de la propriété 10

3.3 Calcul rapide des métriques de spécialisation L^k

Le calcul classique des distances de spécialisation L^k se fait à travers l'équation :

$$d_k(m_1, m_2) = \frac{1}{\rho} \left\| \mathbf{M}^{-1} \mathrm{Diag}\left(\mathbf{M}\left(\mathbf{m}_1 - \mathbf{m}_2\right)\right) \mathbf{M} \right\|_k . \tag{3.22}$$

Comme nous pouvons le remarquer dans (3.22), le calcul matriciel des distances de spécialisation d_k se fait à travers un produit matriciel dont la complexité est de l'ordre de $O(N^3)$. Une telle complexité de calcul, qui est plus grande que celles d'autres distances entre fonctions de masse, peut s'avérer prohibitive pour beaucoup d'applications. Par conséquent, de nouvelles méthodes de calcul sont élaborées dans cette section pour les distances de spécialisation de type L^k. Dans un premier temps, des résultats préliminaires sont donnés pour des cas particuliers de fonctions de masse. Puis, dans la dernière partie de cette section, nous élaborons un algorithme pour le calcul dans le cas général.

3.3.1 Distances d_k entre des fonctions de masse catégoriques

Une façon rapide de calculer la distance de spécialisation basée sur la norme L^k entre des masses de croyances catégoriques est introduite dans [32]. En effet, il est prouvé dans cet article qu'il existe un isomorphisme d'ordre entre la distance ensembliste de Hamming et les distances d_k restreintes aux fonctions de masse catégoriques. Précisément, nous avons :

$$d_k\left(m_A, m_B\right) = \left(\frac{N - 2^{n-|A\Delta B|}}{N-1}\right)^{\frac{1}{k}}, \tag{3.23}$$

où Δ est la différence symétrique entre deux ensembles. Le cardinal de la différence symétrique est égal à la distance de Hamming entre deux ensembles.
L'intérêt principal de l'équation (3.23) est de permettre le calcul des distances d_k entre fonctions de masse catégoriques avec une complexité en $O(1)$. Cette équation montre aussi qu'il existe un isomorphisme d'ordre entre la distance de Hamming et la distance de spécialisation d_k ce qui implique que cette distance tient compte des propriétés structurelles entre fonctions de masse.

3.3.2 Distances entre une fonction de masse catégorique et une autre fonction de masse quelconque

Dans cette sous-section, un cas plus large est exploré : il s'agit du calcul de la distance d_k entre une fonction de masse catégorique et une fonction de masse quelconque. Dans la proposition 11, nous introduisons un résultat uniquement valable pour la distance de spécialisation d_1 basée sur la norme matricielle L^1 :

Proposition 11. *Soit m une fonction de masse définie dans un cadre Ω. Soit aussi $A \subseteq \Omega$ et m_A une fonction de masse catégorique dans Ω. La matrice de spécialisation de m est notée $\mathbf{S}$ et celle de m_A est $\mathbf{S}_A$. Nous avons alors :*

$$d_1\left(m, m_A\right) = \frac{N - \|\mathbf{S} \circ \mathbf{S_A}\|_1}{N-1}, \tag{3.24}$$

$$= \frac{N - tr\left(\mathbf{S}.\,{}^t\mathbf{S}_A\right)}{N-1}, \tag{3.25}$$

avec $\circ$ le produit matriciel d'Hadamard, tr la trace matricielle et ${}^t\mathbf{S}_A$ la matrice transposée de la matrice $\mathbf{S}_A$.

Démonstration. Par définition de la norme L^1, nous avons :

$$\|\mathbf{S} - \mathbf{S}_A\|_1 = \sum_{X,Y \subseteq \Omega} |S(X,Y) - S_A(X,Y)|.$$

Il est connu que $S_A(X,Y) = 1$ si $A \cap Y = X$ et $S_A(X,Y) = 0$ sinon. Ceci implique :

$$\begin{aligned}
\|\mathbf{S} - \mathbf{S}_A\|_1 &= \sum_{\substack{X,Y \subseteq \Omega \\ X = A \cap Y}} (1 - S(X,Y)) + \sum_{\substack{X,Y \subseteq \Omega \\ X \neq A \cap Y}} S(X,Y), \\
&= \sum_{\substack{X,Y \subseteq \Omega \\ X = A \cap Y}} 1 + \sum_{\substack{X,Y \subseteq \Omega \\ X \neq A \cap Y}} S(X,Y) - \sum_{\substack{X,Y \subseteq \Omega \\ X = A \cap Y}} S(X,Y), \\
&= \|\mathbf{S}_A\|_1 + \sum_{X,Y \subseteq \Omega} S(X,Y) - 2 \sum_{\substack{X,Y \subseteq \Omega \\ X = A \cap Y}} S(X,Y), \\
&= \|\mathbf{S}_A\|_1 + \|\mathbf{S}\| \, 1 - 2 \, \|\mathbf{S} \circ \mathbf{S}_A\|_1 .
\end{aligned}$$

Puisque la norme L^1 de n'importe quelle matrice de spécialisation est égale à N, l'équation (3.24) est retrouvée. L'équation (3.25) est obtenue à partir de l'équation (3.24) en utilisant un résultat algébrique classique. □

En terme de temps de calcul, l'équation (3.24) devrait être préférée. En effet, les matrices de spécialisation ont $3^n = N^{\frac{\log(3)}{log(2)}}$ éléments non-nuls. Le produit de Hadamard peut quant à lui être restreint à un produit point par point de ces éléments non-nuls. La complexité de l'équation (3.24) est égale à $O\left(N^{\frac{\log(3)}{log(2)}}\right) \approx O\left(N^{1.58}\right)$.

3.3.3 Distances d_k entre fonctions de masse quelconques

Nous nous intéressons à présent au problème de calcul de la distances d_k dans le cas général. A cet effet, un algorithme de calcul rapide d'une matrice de spécialisation est introduit. Cet algorithme est inspiré du résultat suivant :

Proposition 12. *Soit m une fonction de masse définie dans un cadre de discernement Ω. Soit aussi $X \subseteq Y \subseteq \Omega$ et $z \notin Y$. Le résultat suivant est alors vérifié :*

$$m[Y](X) = m[Y \cup \{z\}](X) + m[Y \cup \{z\}](X \cup \{z\}). \tag{3.26}$$

Démonstration. La démonstration de la proposition 12 est illustrée dans l'annexe A.6. □

La proposition 12 est particulièrement intéressante pour le calcul des matrices de spécialisation. Comme nous le savons, tout élément de la matrice de spécialisation est issu d'un conditionnement de la masse m sachant un ensemble $Y \subseteq \Omega$. En effet, $S(X,Y) = m[Y](X)$. Cette proposition montre par conséquent que chaque élément de la matrice de spécialisation d'une fonction de masse m peut être obtenu en additionnant deux autres éléments provenant d'une colonne située à droite de cet élément et des lignes inférieures. Comme nous le savons aussi, la dernière colonne de la matrice $\mathbf{S}$ est égale à $\mathbf{m}$. Cette

matrice peut en effet être construite de façon incrémentale en commençant par la colonne d'indice $N-1$ jusqu'à la première colonne. Dans chaque colonne, nous commençons par l'élément le plus bas en remontant vers les éléments de la partie supérieure de la colonne en question. Cette procédure est donnée dans l'algorithme 1.

L'exemple suivant illustre le mécanisme de calcul rapide de la matrice de spécialisation

Algorithme 1 Calcul rapide d'une matrice de spécialisation **S**

entrées : m, N, $\mathbf{M}$.
$\mathbf{S} \leftarrow \mathbf{0}$, la matrice nulle.
pour $X \subseteq \Omega$ **faire**
  $S(X, \Omega) \leftarrow m(X)$
fin pour
pour $Y \subsetneq \Omega$ (suivant l'ordre binaire décroissant) **faire**
  pour $X \subseteq Y$ (suivant l'ordre binaire décroissant) **faire**
    si $M(X, Y) > 0$ **alors**
      choisir $z \in \bar{Y}$.
      $S(X, Y) \leftarrow S(X, Y \cup \{z\}) + S(X \cup \{z\}, Y \cup \{z\})$.
    fin si
  fin pour
fin pour
retourner **S**.
Fin

S développé dans l'algorithme 1.

Exemple 3. *Soit une fonction de masse m définie dans $\Omega = \{\omega_1, \omega_2, \omega_3\}$ sachant que :*

$$\begin{aligned} m(\{\omega_1\}) &= 0.4, \\ m(\{\omega_1, \omega_2\}) &= 0.2, \\ m(\{\omega_3\}) &= 0.3, \\ m(\Omega) &= 0.1\ . \end{aligned}$$

La matrice de spécialisation dempsterienne de la fonction de masse m est :

$$\mathbf{S} = \begin{bmatrix} 1 & 0.3 & 0.7 & 0.3 & 0.6 & 0 & 0 & 0 \\ 0 & 0.7 & 0 & 0.4 & 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 0.3 & 0 & 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0.3 & 0 & 0 & 0 & 0.2 \\ 0 & 0 & 0 & 0 & 0.4 & 0.3 & 0.3 & 0.3 \\ 0 & 0 & 0 & 0 & 0 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1 \end{bmatrix}$$

Comme nous le remarquons, la dernière colonne de la matrice $\mathbf{S}$ est égale au vecteur fonction de masse $\mathbf{m}$. Pour calculer les autres éléments de cette matrice, on procède comme illustré dans l'algorithme précédent. En voici quelque exemples :

$$\begin{aligned} S(\{\omega_1\},\{\omega_1,\omega_2\}) &= S(\{\omega_1\},\{\omega_1,\omega_2\}\cup\{\omega_3\}) + S(\{\omega_1\}\cup\{\omega_3\},\{\omega_1,\omega_2\}\cup\{\omega_3\}), \\ &= S(\{\omega_1\},\{\omega_1,\omega_2,\omega_3\}) + S(\{\omega_1,\omega_3\},\{\omega_1,\omega_2,\omega_3\}), \\ &= 0.4+0, \\ &= 0.4. \end{aligned}$$

$$\begin{aligned} S(\{\omega_1\},\{\omega_1\}) &= S(\{\omega_1\},\{\omega_1,\omega_2\}) + S(\{\omega_1,\omega_2\},\{\omega_1,\omega_2\}), \\ &= 0.4+0.3, \\ &= 0.7. \end{aligned}$$

Pour le calcul de $S(\{\omega_1\},\{\omega_1\})$, le même résultat peut aussi être obtenu en incluant le singleton ω_3 au lieu de ω_2, alors :

$$\begin{aligned} S(\{\omega_1\},\{\omega_1\}) &= S(\{\omega_1\},\{\omega_1,\omega_3\}) + S(\{\omega_1,\omega_3\},\{\omega_1,\omega_3\}), \\ &= 0.6+0.1, \\ &= 0.7. \end{aligned}$$

L'algorithme 1 de calcul rapide de la matrice de spécialisation peut être utilisé directement avec l'entrée $m_1 - m_2$ au lieu de m dans l'objectif d'obtenir la différence matricielle $\mathbf{S_1} - \mathbf{S_2}$. Cet algorithme permet aussi de calculer récursivement la distance d_k en la mettant à jour à chaque fois qu'un nouveau terme $S_1(X,Y) - S_2(X,Y)$ est obtenu.
Étant donné la définition de la matrice $\mathbf{M}$ donnée dans le chapitre 1, nous constatons que l'algorithme 1 présente 3^n boucles. De même que pour d_1 donnée dans l'équation (3.24), la complexité de calcul de la distance d_k entre des fonctions de masse quelconques est en effet $O\left(N^{1.58}\right)$. La majorité des métriques[2] énoncées dans l'état de l'art de la théorie des fonctions de croyance, telles que la distance de Jousselme, ont quant à elles recours à un produit entre une matrice de taille $N \times N$ et un vecteur de la fonction de masse de taille N. Leur complexité est alors $O\left(N^2\right)$ (en utilisant une programmation naïve). Par conséquent, nous avons bien réussi à rendre les distances de spécialisation basées sur les normes matricielles L^k au moins autant attractives en terme de temps de

2. Perry et Stephanou [39] ont défini une métrique complète avec une complexité égale à $O(N)$ mais elle échoue à saisir les aspects structurels de la comparaison des fonctions de masse (voir [21]).

calcul que les autres métrique définies dans la TFC.
La figure 3.2 illustre le temps de calcul de la distance d_k en utilisant l'algorithme 1 comparé au temps de calcul obtenu en utilisant l'équation (3.22). Ces résultats sont obtenus en utilisant un ordinateur portable doté d'un CPU Intel® centrino2 2.53 GHz et d'un environnement de programmation GNU Octave©. On peut constater que le rapport en log des temps de calcul est linéaire en fonction de n ce qui est conforme au fait que la complexité est passée de $O\left(N^3\right)$ à $O\left(N^{1.58}\right)$.

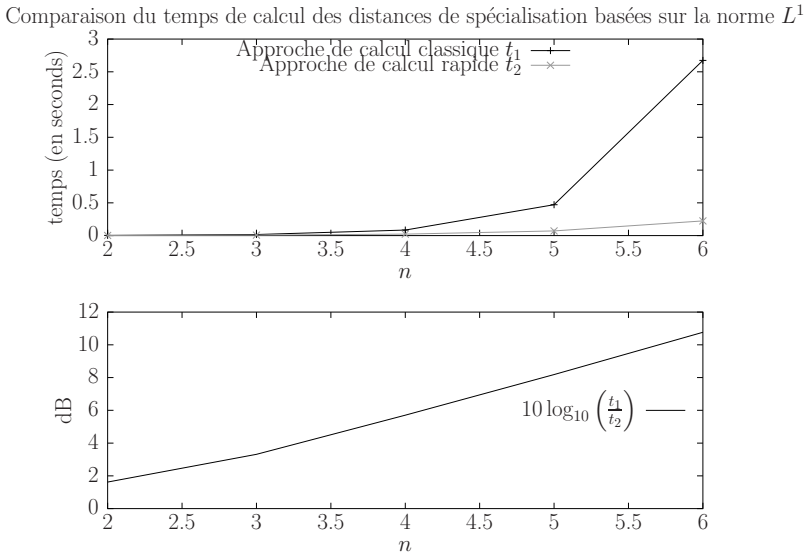

FIGURE 3.2 – Comparaison du temps de calcul de l'approche classique (équation (3.22)) et l'approche rapide (algorithme 1).

3.4 Comparaison des distances entre fonctions de croyance

Dans cette section nous effectuons une comparaison de différentes distances entre fonctions de croyance et fournissons des informations sur leur limites. Pour ce qui est des aspects théoriques, les propriétés des indices de dissimilarité entre fonctions de croyance sont résumés dans le tableau 3.4. Des cas d'étude sont ensuite proposés et focalisés sur les propriétés principales suivantes : définition, structure et consistance conjonctive. Les dissimilarités suivantes citées dans l'état de l'art sont retenues : d_J, d_T, d_{ZD}, d_{IF1}, d_{IF2} et d_{IF3}. Parmi les nouvelles distances définies dans ce chapitre, les mesures suivantes sont retenues : d_{op1}, d_{op2}, $d_{op\infty}$, d_1, d_2 et d_∞.

3.4.1 Récapitulatif des propriétés principales

Une liste des propriétés principales que satisfont les différentes dissimilarités étudiées est dressée dans le tableau 3.4 :

TABLE 3.4 – Résumé des propriétés principales satisfaites par les distances entre fonctions de masse.

	Pseudo-métriques					Métriques complètes						
						Distances classiques	Distances de spécialisation					
Dissimilarités	d_T	d_{ZD}	d_{IF1}	d_{IF2}	d_{IF3}	d_J	d_{op1}	d_{op2}	$d_{op\infty}$	d_1	d_2	d_∞
Normalité	×	×	×	×	×	×	×	×	×	×	×	×
Définition						×	×	×	×	×	×	×
structure						×				×	×	
Consistance conjonctive							×			×		×

L'analyse des résultats du tableau 3.4 montre clairement que, d'un point de vue théorique, la distance d_1 est plus performante que toutes les autres puisqu'elle est la seule distance possédant toutes les propriétés énoncées dans la section 3.2.3. Ce tableau montre aussi que la propriété la plus stricte est certainement la propriété de consistance conjonctive car elle n'est satisfaite que par trois des nouvelles distances parmi celles étudiées dans cette section.
Concernant la propriété structurelle, il existe des preuves formelles de la généralisation de la distance Jaccard par la distance de Jousselme. Par ailleurs, les distances de spécialisation du type L^k généralisent la distance d'ensemble de Hamming lorsque $k < \infty$.

3.4.2 Influence de la "définition" (*definiteness*)

Cette sous-section s'intéresse à l'influence de la propriété de *definiteness*. Notons qu'il est récemment prouvé formellement que d_J est définie [2]. L'exemple 2 montre clairement que les mesures d_Z et d_T sont des pseudo-distances et donc non définies. En effet des fonctions de masse isopignistiques[3] sont vues comme des fonctions de masse totalement identiques par ces deux dissimilarités dans le sens où elles produisent une même décision. Si les fonctions de masse ne sont plus traitées avant la décision, ce résultat est alors justifié. Toutefois, si les états de croyance peuvent encore évoluer grâce à des preuves supplémentaires, il est alors préférable de considérer ces deux fonctions de masse comme différentes.
Pour clarifier ce point de vue et inclure par la même occasion les mesures de dissimilarité intuitionnistes dans cette discussion, considérons l'exemple suivant :

3. Deux fonctions de masse sont isopignistiques si elles ont les mêmes distributions de probabilité pignistiques.

Exemple 4. *Supposons un cadre de discernement* $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ *et deux fonctions de masse* m_1 *et* m_2 *définies dans* Ω *comme suit :*

$$m_1(\{\omega_1, \omega_2\}) = m_1(\{\omega_3, \omega_4\}) = \frac{1}{2},$$
$$m_2(\{\omega_1, \omega_3\}) = m_2(\{\omega_2, \omega_4\}) = \frac{1}{2}.$$

Nous avons alors $d_{IF1}(m_1, m_2) = d_{IF2}(m_1, m_2) = d_{IF3}(m_1, m_2) = d_T(m_1, m_2) = d_Z(m_1, m_2) = 0$.

L'exemple 4 montre que des fonctions de masse engagées identiquement sur le même nombre de sur-ensembles d'hypothèses ω_i sont vus comme complètement similaires par les mesures intuitionnistes. Cette condition est plus robuste que la condition isopignistique. Encore une fois, si la prise de décision est immédiatement la prochaine étape dans le processus de calcul, ces résultats sont alors justifiés. En contre partie, supposons qu'un nouvel état de connaissance est collecté et que cette information est représentée par la fonction de masse $m_3 = m_{\{\omega_1,\omega_2\}}$. Quand on fusionne selon $\cap$ la fonction m_3 avec m_1 ainsi qu'avec m_2, il en résulte que les états de croyance évoluent de façons assez différentes :

- m_{13} est telle que $m_{13}(\{\omega_1, \omega_2\}) = m_{13}(\emptyset) = \frac{1}{2}$. Ceci sous-entend que la solution appartient à $\{\omega_1, \omega_2\}$ mais elle est complètement indéterminée dans ce sous-ensemble.
- m_{23} est telle que $m_{23}(\{\omega_1\}) = m_{23}(\{\omega_2\}) = \frac{1}{2}$. Une telle fonction de masse est une distribution de probabilité complètement engagée et présente les mêmes risques de choisir ω_1 ou ω_2 (équiprobabilité).

Par conséquent m_1 et m_2 devraient être vues comme des fonctions de masse dissimilaires.

3.4.3 Influence de l'interaction des éléments focaux

Cette sous-section s'intéresse à la propriété structurelle et l'intérêt pratique qu'elle suscite pour la comparaison des fonctions de masse. Pour mettre cette propriété en exergue, deux exemples sont étudiés. Dans ces exemples, des fonctions de masse catégoriques sont comparées. En effet, les fonctions de masse catégoriques sont les vecteurs de base de l'espace ε_Ω. Il est alors justifié de restreindre l'analyse à cette famille de fonctions de masse si l'on veut savoir comment les interactions entre éléments focaux sont prises en compte. Dans l'exemple 5, des fonctions de masse catégoriques et non conflictuelles sont mises en compétition, tandis que dans l'exemple 6 des fonctions de masse catégoriques et conflictuelles sont comparées. Notons que dans ces deux exemples, les courbes correspondant aux mesures d_{IF2} et d_{IF3} sont superposées. La même remarque est valable pour d_{op1} et d_∞.

Exemple 5. *Cet exemple est inspiré de celui présenté dans [19] et réutilisé dans [17]. Supposons deux fonctions de masse définies dans un cadre de discernement Ω de cardinalité $|\Omega| = 8$ tel que :*

$$\begin{aligned} m_1(\{\omega_1, \omega_2, \omega_3\}) &= 1, \\ m_2(A_t) &= 1. \end{aligned}$$

Le sous-ensemble A_t varie par inclusion successive d'un singleton à chaque étape de calcul allant de $\{\omega_1\}$ jusqu'à atteindre Ω à la dernière étape :

t : étape de calcul	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
élément focal A_t	$\{\omega_1\}$	$\{\omega_1, \omega_2\}$	$\{\omega_1, \omega_2, \omega_3\}$	$\{\omega_1, ..., \omega_4\}$	$\{\omega_1, ..., \omega_5\}$	$\{\omega_1, ..., \omega_6\}$	$\{\omega_1, ..., \omega_7\}$	Ω

Les résultats sont montrés dans la figure 3.3. Ces résultats incluent aussi les deux distances d'ensemble : la distance de Hamming et la distance de Jaccard.

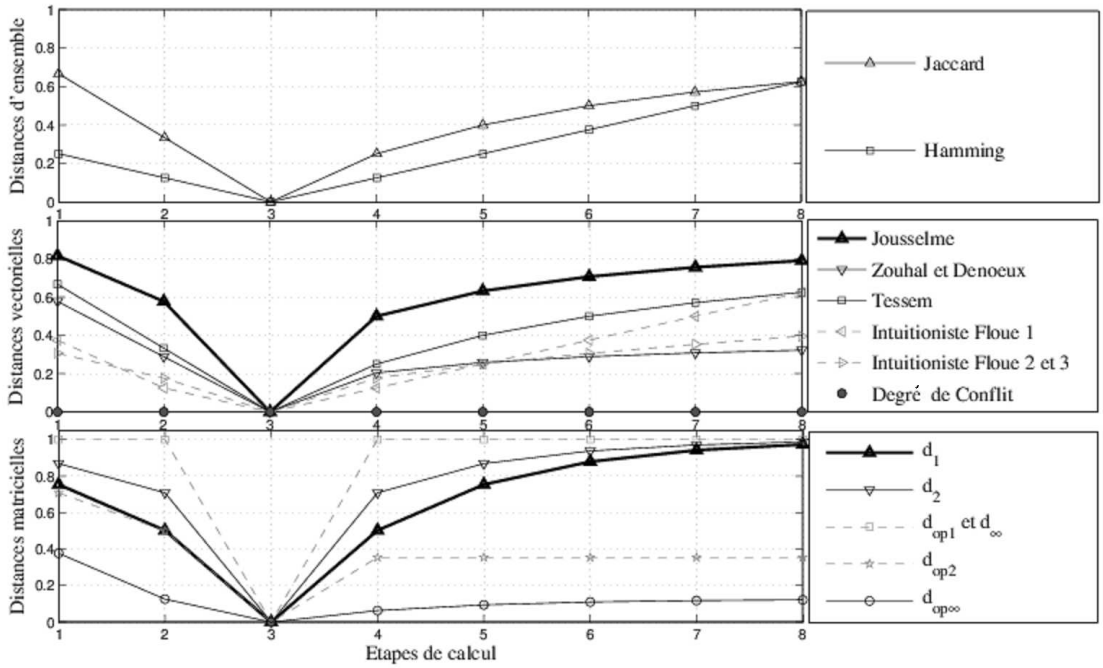

FIGURE 3.3 – Dissimilarités entre m_1 et m_2 - Exemple 5

Dans l'exemple 5, la majorité des mesures de dissimilarité présentent une monotonicité similaire des distances d'ensemble : elles décroissent de l'étape de calcul 1 à l'étape 3, où la valeur minimale est atteinte, et croissent de l'étape 3 à l'étape 8. Concernant l'étape 3, toutes les mesures sont nulles car $A_3 = \{\omega_1, \omega_2, \omega_3\}$. Notons que le conflit est nul tout au long de l'expérience. Ce résultat montre que ce critère n'est pas adapté à

la comparaison des fonctions de masse en toutes circonstances. A l'exception de d_{op1} et d_∞, nous pouvons aussi remarquer que les mesures étudiées prennent en compte la cardinalité des éléments focaux dans leur variation, notamment celle du sous-ensemble A_t. Ce résultat confirme que les interactions entre éléments focaux sont bien prises en compte par les mesures d_{ZD}, d_T, d_{IF1}, d_{IF2}, d_{IF3}, d_J, d_1, d_2, d_{op2} et $d_{op\infty}$.

Exemple 6. *Cet exemple est inspiré d'un exemple cité dans [19]. Supposons deux fonctions de masse définies dans un cadre de discernement de cardinalité* $|\Omega| = 8$ *sachant que :*

$$\begin{aligned} m_1(\{\omega_8\}) &= 1, \\ m_2(A_t) &= 1. \end{aligned}$$

Le sous-ensemble A_t *varie de* $\{\omega_1\}$ *jusqu'à* Ω *par inclusion successive d'un singleton, puis par omission successive d'un singleton de* $\{\Omega\}$ *jusqu'à atteindre* $\{\omega_8\}$ *à la dernière étape étape de calcul :*

t : étape de calcul	*1*	*2*	*...*	*7*	*8*	*9*	*...*	*14*	*15*	*16*
élément focal A_t	$\{\omega_1\}$	$\{\omega_1, \omega_2\}$	...	$\{\omega_1, ..., \omega_7\}$	Ω	$\{\omega_2, ..., \omega_8\}$	...	$\{\omega_6, \omega_7, \omega_8\}$	$\{\omega_7, \omega_8\}$	$\{\omega_8\}$

Les résultats obtenus sont montrés en figure 3.4. Ces résultats incluent aussi les deux distances d'ensemble : la distance de Hamming et la distance de Jaccard.

Dans l'exemple 6, le comportement attendu des mesures de dissimilarité étudié est :

- grandes valeurs de dissimilarité pour les étapes 1 à 7, car pour ces étapes $A_t \cap \{\omega_8\} = \emptyset$. A l'étape 7, A_7 est le complémentaire de l'ensemble $\{\omega_8\}$ dans Ω (dissimilarité totale).
- des valeurs de plus en plus petites et décroissantes de l'étape 8 à l'étape 15. Durant ces étapes, ω_8 est inclus dans A_t et $|A_t|$ décroit. Finalement à l'étape 15, nous remarquons que $m_1 = m_2$ (similarité complète).

Nous pouvons aisément voir que toutes les mesures sont décroissantes de façon monotone de l'étape 8 à l'étape 15 où une dissimilarité nulle est obtenue. En revanche lors des étapes 1 à 7, les dissimilarités peuvent être partagées en 4 catégories :

- les indices constants : d_J, d_T, d_∞, d_{op1} et $d_{op\infty}$. Ce comportement est expliqué par le fait que, de l'étape 1 à 7, les deux fonctions de masse comparées sont catégoriques et leurs éléments focaux sont disjoints. Il est par conséquent justifié d'interpréter ces fonctions de masse comme totalement dissimilaires.
- les indices décroissants : d_{ZD}. Ce comportement est interprété par le fait que la spécificité décroit de l'étape 1 à l'étape 7. En effet, A_7 contient plus d'ignorance que A_1.
- les indices croissants : d_1, d_2 et d_{IF1}. Ce comportement est expliqué par le fait qu'à l'étape 7, les deux fonctions de masse comparées sont catégoriques et leurs

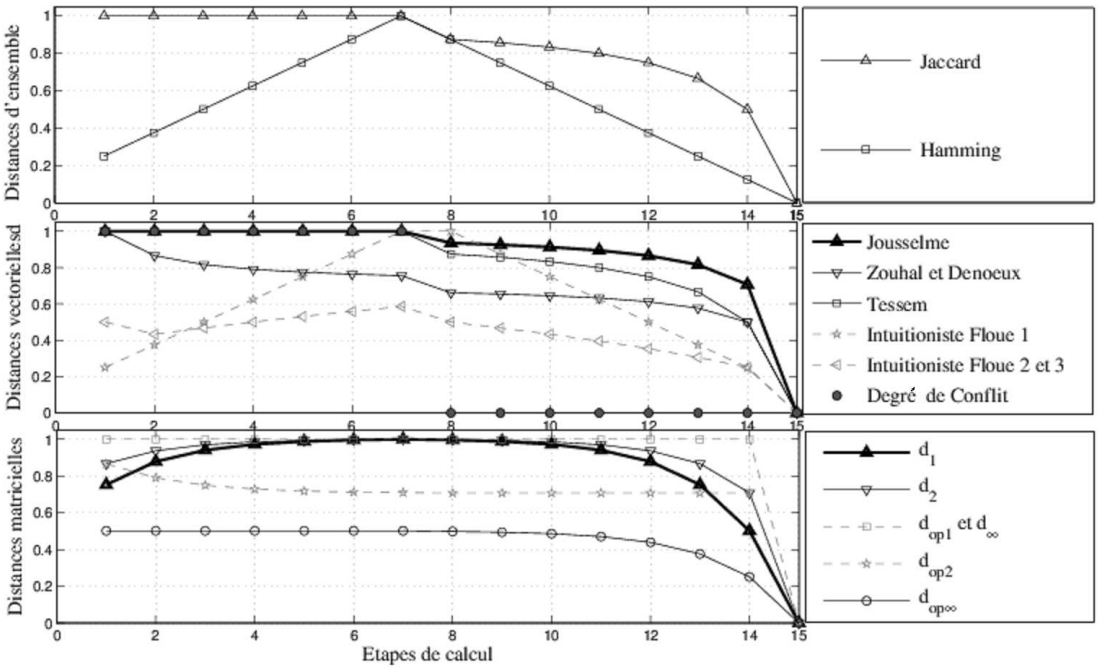

FIGURE 3.4 – Dissimilarités entre m_1 et m_2 - Exemple 6.

éléments focaux complémentaires dans Ω. Il est donc justifié que la dissimilarité totale ne soit atteinte qu'à l'étape 7. Une critique peut être soulevée concernant ces mesures puisqu'elle montre que $m_{\{\omega_8\}}$ est aussi plus proche de $m_{\{\omega_1\}}$ que de m_Ω. Par contre, $m_{\{\omega_8\}}$ et $m_{\{\omega_1\}}$ sont des fonctions de masse incompatibles.

– les indices non monotones : d_{IF2} et d_{IF3}. Ce comportement est probablement le moins satisfaisant de tous puisqu'il est très difficile d'interpréter les différentes variations de ces mesures.

3.4.4 Discrimination à l'égard de la connaissance commune

Dans cette sous-section, nous étudions le pouvoir discriminant des dissimilarités à l'égard d'informations communes. Revoyons dans un premier temps l'exemple 2 énoncé dans la motivation des développements de ce chapitre. Dans cet exemple, la distance entre m_1 et m_2 est comparée à la distance entre $m_1 \textcircled{\cap} m_3$ et $m_2 \textcircled{\cap} m_3$. Les résultats de cette expérience sont présentés dans le tableau 3.5 où des résultats additionnels concernant les distances de spécialisation nouvellement développées au cours de ce chapitre sont également illustrés.

Les distances d_J, d_2, d_{op2} et $d_{op\infty}$ peuvent être critiquées parce que deux fonctions de masse obtenues après combinaison avec un même état de connaissance contenu dans une troisième fonction de masse deviennent plus distantes qu'elles ne l'étaient avant la

TABLE 3.5 – Mesures de distance additionnelles entre m_1 et m_2 et entre m_{13} et m_{23} - Exemple 2.

Distances	$d(m_1, m_2)$	$d(m_{13}, m_{23})$
d_J	0.5831	0.8
d_T	1	0
d_{ZD}	1	0
d_{IF1}	0.8	0.4
d_{IF2}	0.5	0.4
d_{IF3}	0.5	0.4
d_1	0.6	0.5333
d_2	0.6324	0.6532
d_∞	0.1333	0.1333
d_{op1}	0.8	0.8
d_{op2}	0.5840	0.8
$d_{op\infty}$	0.3333	0.5333

combinaison. Le pouvoir discriminant de d_{op1} et de d_∞ semble être plus faible que celui des autres dissimilarités examinées parce que des distances égales sont observées.
Comme mentionné précédemment, les dissimilarités d_T, d_{ZD}, d_{IF1}, d_{IF2} et d_{IF3} ont un comportement satisfaisant dans cet exemple. Toutefois puisqu'elles ne satisfont pas la propriété de consistance conjonctive, d'autres exemples peuvent être développés pour prouver qu'elles ont un pouvoir discriminant plus faible que d_1 et d_{op1}.
Pour faire une comparaison plus fine, il est plus intéressant d'évaluer à quel degré les différentes mesures examinées respectent la propriété de consistance conjonctive. A cet effet, nous calculons le taux de compatibilité à la propriété 3.2.3.3. Supposons m_1, m_2 et m_3 trois fonctions de masse choisies aléatoirement. Si une distance d satisfait $d(m_1, m_2) \geq d(m_1 \textcircled{\cap} m_3, m_2 \textcircled{\cap} m_3)$ (inégalité (3.2)), alors un succès est attribué à celle-ci. Le taux de succès est alors obtenu en divisant le nombre de succès par le nombre total d'expériences aléatoires. Il est décidé de ne pas tirer uniformément les fonctions de masse dans le simplexe de l'espace ε_Ω. En effet, l'inégalité (3.2) est plus rarement vérifiée quand les fonctions de masse comparées sont séparables. En conséquence, les fonctions de masse aléatoires utilisées dans cette expérience sont séparables. Le tableau 3.6 donne les taux de succès des dissimilarités examinées.

Comme attendu, d_1, d_∞ et d_{op1} ont un taux de succès plein. La distance la plus discriminante et qui ne satisfait pas la propriété de consistance conjonctive est d_2.

3.4.5 Influence de la taille du cadre de discernement

Cette sous-section est dédiée à l'analyse du comportement des dissimilarités entre fonctions de masse en fonction des variations de la cardinalité du cadre de discernement. Cette analyse est basée sur l'exemple 7.

Exemple 7. *Cet exemple était proposé dans [17]. soit* $\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$ *un cadre*

TABLE 3.6 – Taux de succès de discrimination de l'information commune des dissimilarités examinées (10^4 réalisations).

Distances	Taux
d_J	86%
d_T	51.5%
d_{ZD}	57.9%
d_{IF1}	93.3%
d_{IF2}	82.3%
d_{IF3}	82.4%
d_1	100%
d_2	99.2%
d_∞	100%
d_{op1}	100%
d_{op2}	42%
$d_{op\infty}$	37%

de discernement. Trois fonctions de croyance sont définies dans Ω *comme suit :*

$m_1(\{\omega_i\}) = \frac{1}{n}, \quad \forall i \in \{1, \cdots, n\}$. *(fonction bayésienne),*
$m_2 = m_\Omega$ *(ignorance totale),*
$m_3 = m_{\{\omega_1\}}$ *(fonction catégorique).*

L'évolution des dissimilarités comparées dans cette étude en fonction de la taille du cadre de discernement n *est visualisée dans les figures 3.5 et 3.6.*

Selon la définition des fonctions de masse impliquées dans l'exemple 7, leur degré de spécificité et d'engagement sont tous différents. Le comportement attendu des dissimilarités entre fonctions de croyance est que la relation d'ordre entre les valeurs des distances devrait être préservé malgré les variations de n. Dans les figure 3.5 et 3.6, seules d_{IF2} et d_{op2} modifient l'ordre des distances. Une analyse des fonctions de croyance pourrait alors être différente après grossissement du cadre de discernement. Ce résultat est en effet insatisfaisant si une telle opération est désirée par l'utilisateur.
Bien que la discrimination des fonctions de croyance n'est pas l'objet d'étude de cette sous-section, nous remarquons pour la distance de Jousselme que les courbes correspondant à $d_J(m_1, m_2)$ et $d_J(m_1, m_3)$ sont superposées. Ce résultat signifie que, du point de vue de la distance de Jousselme, la fonction bayésienne est équidistante de la fonction catégorique et de l'ignorance totale. D'une part, il peut être justifié de préférer le fait que la fonction de masse bayésienne m_1 soit plus proche de la fonction catégorique m_3 que de l'ignorance totale m_2. En effet, m_1 et m_3 sont fortement engagées puisqu'elles sont des distributions de probabilités. A l'opposé, m_2 n'est pas du tout engagée puisqu'il s'agit d'une ignorance totale, ce qui s'explique par le fait que d_1 considère m_1 et m_3 comme des fonctions équidistantes de la fonction de masse m_2. D'autre part, il peut

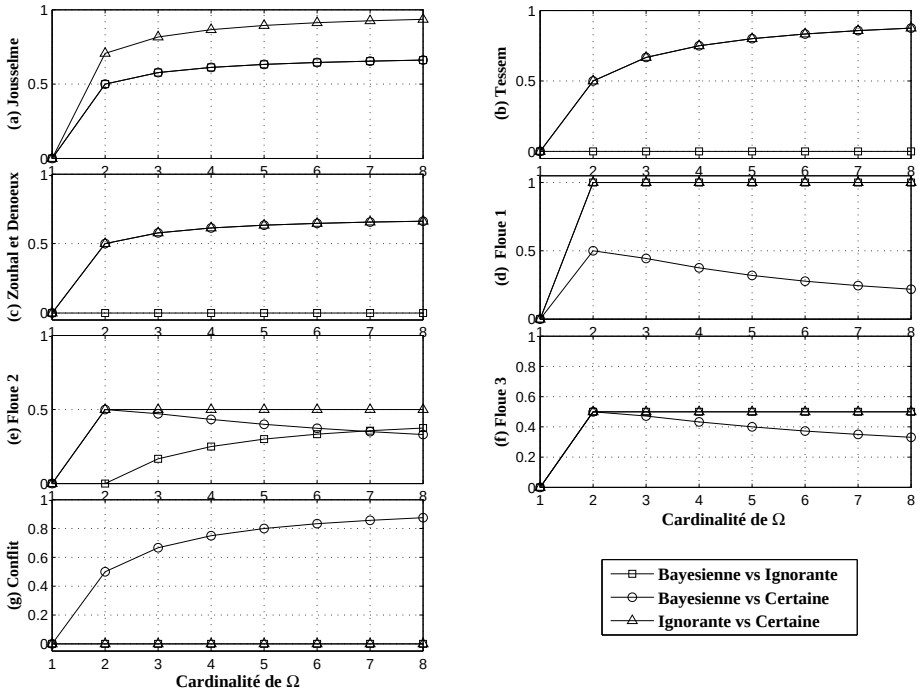

FIGURE 3.5 – dissimilarités entre m_1, m_2 et m_3 - Exemple 7 - Partie I.

aussi être justifié que m_1 et m_2 devraient être proches car elles sont isopignistiques. Le fait que m_1 et m_2 sont isopignistiques implique que leurs courbes sont constantes et nulles pour les deux mesures d_{ZD} et d_T. Cette dernière remarque illustre le fait que la spécificité n'est pas considérée de la même façon par toutes les dissimilarités étudiées et que le comportement souhaité dépend des besoins de l'utilisateur.

3.5 Conclusion

Dans ce chapitre, une nouvelle famille de distances entre fonctions de croyance a été développée. Étant donné que chaque fonction de masse est représentée de manière unique par sa matrice de spécialisation dempsterienne, la nouvelle famille de distances définie dans ce chapitre est basée sur les normes matricielles appliquées aux matrices de spécialisation. L'utilisation des matrices de spécialisation pour comparer les fonctions de croyance est justifiée par le fait qu'elles ne représentent pas uniquement l'état actuel de croyance mais aussi tous les conditionnements futurs potentiellement atteignables au sens de la combinaison conjonctive.
De plus, nous avons argumenté dans ce travail qu'une distance évidentielle se conforme à trois principes essentiels :

– une similarité totale n'est obtenue que si les états de croyance sont identiques,

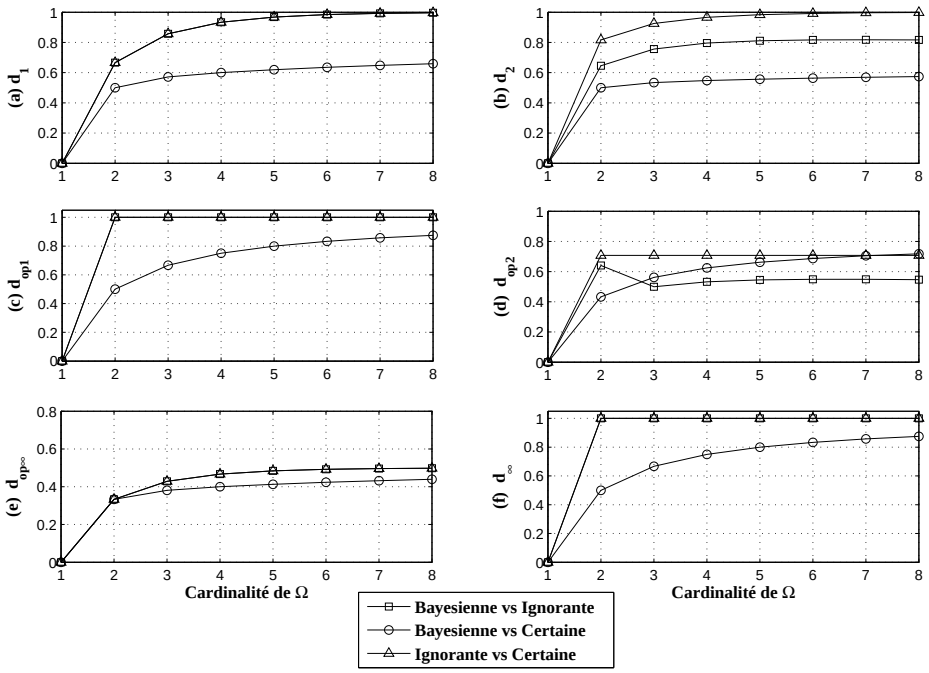

FIGURE 3.6 – dissimilarités entre m_1, m_2 et m_3 - Exemple 7 - Partie II (distances de spécialisation)

- les interactions entre les éléments focaux doivent être prises en compte,
- les états comprenant des connaissances communes sont plus proches que ceux ne partageant pas de connaissances.

Ces trois principes sont formalisés en propriétés mathématiques. Il est prouvé que seule la distance de spécialisation du type L^1 satisfait ces trois propriétés essentielles. La troisième propriété est appelée **consistance conjonctive**. Grâce à cette propriété, deux fonctions de masse incorporant un même état de connaissance à travers une combinaison conjonctive deviennent plus proches après combinaison. Par exemple, dans le cas de fonctions de masse séparables, cette propriété signifie que deux fonctions sont d'autant plus proches que les mêmes éléments entrent en jeu dans leur décomposition conjonctive.

En outre, plusieurs méthodes pour le calcul rapide des distances de spécialisation basées sur les normes matricielles L^k sont aussi introduites. Initialement, ces distances sont calculées avec une complexité de $O\left(N^3\right)$. Dans un cas général, nous avons élaboré un algorithme pour réduire cette complexité à $O\left(N^{1.58}\right)$ afin de gagner en temps de calcul. Dans le cas de fonctions de masse catégoriques, cette complexité est simplement de l'ordre de $O\left(1\right)$. L'utilisation de ces approches permettent à ces distances d'être très compétitives en terme de temps de calcul en plus de leurs propriétés mathématiques et

structurelles très attractives.

Ce chapitre contient aussi plusieurs cas d'étude qui illustrent les avantages et les limitations des nouvelles distances comparativement aux distances existantes. L'analyse des résultats de ces expériences soutient l'idée que le choix d'une distance dans le cadre de la théorie des fonctions de croyance devrait être fait selon l'application dédiée. En effet, aucune distance ne peut être considérée la plus performante en toutes les circonstances. L'adéquation entre les propriétés satisfaites par une distance et les objectifs ciblés par une application sont, en effet, les éléments essentiels dans le choix d'une distance appropriée entre fonctions de croyance.

Le prochain chapitre sera consacré à la généralisation des distances basées sur les matrices de spécialisation dempsterienne définies dans ce chapitre au cadre des $\alpha-$jonctions.

Chapitre 4

Généralisation des distances matricielles entre fonctions de croyance

Sommaire

4.1 Introduction

Dans un processus de fusion d'informations à l'aide de la TFC, la règle de combinaison la plus courante est, sans aucun doute, la règle de combinaison conjonctive. L'application de cette règle exige une confiance dans la véracité et l'indépendance des sources comparées. Quand la véracité de toutes les sources agrégées est incertaine, on suppose qu'au moins une d'entre elles est fiable sans savoir laquelle. Un comportement prudent peut alors s'imposer en appliquant la règle de combinaison disjonctive. Ces deux règles de combinaison sont les plus utilisées dans un processus de fusion d'informations basé sur la TFC. Ces règles de combinaison sont commutatives, associatives et admettent chacune un élément neutre unique. L'élément neutre de la combinaison conjonctive est la masse d'ignorance totale notée m_Ω, celui de la règle disjonctive est la masse totalement engagée sur l'ensemble vide notée $m_\emptyset$. On peut remarquer que cette dernière masse de croyance est la négation de la masse d'ignorance totale. Dubois et Prade [9] soulignent cette dualité et montrent que ces deux opérateurs de combinaison sont liés par les lois de *De Morgan* :

$$\begin{aligned} \overline{m_1 \textcircled{\cup} m_2} &= \overline{m_1} \textcircled{\cap} \overline{m_2}, \\ \overline{m_1 \textcircled{\cap} m_2} &= \overline{m_1} \textcircled{\cup} \overline{m_2}. \end{aligned} \tag{4.1}$$

En plus de la règle de combinaison conjonctive et disjonctive, deux autres opérateurs marginaux sont définis ; à savoir la règle de combinaison disjonctive exclusive notée $\underline{\cup}$ et sa négation $\underline{\cap}$. Étant donné deux fonctions de masse m_1 et m_2 définies dans Ω, ces deux dernières sont respectivement définies par les formules suivantes [51, 42] :

$$m_{1\underline{\cup}2} = \sum_{A=B\Delta C} m_1(B)m_2(C), \tag{4.2}$$

$$m_{1\underline{\cap}2} = \sum_{A=\underline{B\Delta C}} m_1(B)m_2(C), \tag{4.3}$$

La règle disjonctive exclusive $\underline{\cup}$ est utilisée dans le cas où nous supposons qu'une seule des sources d'informations exactement dit la vérité sans pouvoir déterminer exactement laquelle. Par ailleurs, la règle $\underline{\cap}$ correspond quant à elle au cas où l'on sait que la véracité de toutes les sources est soit approuvée ou contestée. Cette situation correspond au cas où toutes les sources disent la vérité ou aucune d'entre-elle ne dit la vérité. Comme les règles conjonctive et disjonctive, la règle disjonctive exclusive $\underline{\cup}$ et sa négation $\underline{\cap}$ sont aussi commutatives, associatives et possèdent respectivement un élément neutre unique : $m_\emptyset$ et m_Ω.

Dans [51], Smets fait remarquer que les combinaisons en utilisant les quatre règles

précédemment énoncées ne sont que des cas marginaux de l'agrégation des connaissances dans la TFC. Ces règles ne sont donc en réalité que des cas particuliers d'une famille de règles de combinaison appelées $\alpha-$jonctions et dont la propriété principale est la linéarité. Cette propriété se distingue plus explicitement dans la formulation matricielle de ces règles de combinaison. Étant donné que ces familles de règles de combinaison admettent une écriture matricielle qui généralise les quatre règles de combinaison marginales, une généralisation du calcul de distances entre fonctions de croyance basée sur le principe du chapitre 3 est donc possible.

4.2 Les $\alpha-$jonctions

4.2.1 Notions fondamentales

Les travaux de Smets [51] sur une nouvelle famille de règles de combinaisons linéaires, commutatives, associatives et qui admettent un seul élément neutre ont permis de définir ce que nous appelons aujourd'hui **les $\boldsymbol{\alpha}-$jonctions**. Cette famille de lois est donc définie et paramétrée par un coefficient α. Elle est basée soit sur un comportement conjonctif admettant m_Ω comme élément neutre ou sur un comportement disjonctif admettant l'élément neutre $m_\emptyset$. Cette famille recouvre à la fois la conjonction $\textcircled{\cap}$, la disjonction $\textcircled{\cup}$, la disjonction exclusive $\textcircled{\scriptsize\Delta}$ ainsi que sa négation $\textcircled{\scriptsize\overline{\cap}}$ qui correspondent aux valeurs extrêmes de $\alpha \in [0,1]$, d'où son appellation dédiée **les $\boldsymbol{\alpha}-$jonctions**.

Soient deux fonctions de masse m_1 et m_2 définies dans un cadre de discernement Ω. La combinaison m_{12} de ces deux dernières est, par définition, supposée être obtenu à l'aide d'un opérateur linéaire, commutatif et associatif.

$$\mathbf{m}_{12} = f\left(m_1, m_2\right). \tag{4.4}$$

Smets [51] a montré que les $\alpha-$jonctions forment l'unique famille d'opérateurs de combinaison possédant les propriétés principales suivantes :
$\forall m_1, m_2, m_3 \in \mathcal{M}^\Omega$

- Linéarité [1]
  $\forall \lambda \in [0,1], f\left(m, \lambda m_1 + (1-\lambda) m_2\right) = \lambda f\left(m, m_1\right) + (1-\lambda) f\left(m, m_2\right),$
- Commutativité :
  $f\left(m_1, m_2\right) = f\left(m_2, m_1\right),$
- Associativité :
  $f\left(f\left(m_1, m_2\right), m_3\right) = f\left(m_1, f\left(m_2, m_3\right)\right),$

1. L'opérateur est linéaire sur l'espace vectoriel engendré par les fonctions de masse catégoriques, mais la sortie de l'opérateur reste une fonction de masse uniquement en cas de combinaison convexe.

- Élément neutre :
  $\exists m_e \mid \forall \mathrm{m}, f(m, m_e) = m$,
- Anonymité : cette propriété implique qu'une permutation des éléments du cadre de discernement Ω n'affecte pas le résultat de la fusion en utilisant les $\alpha-$jonctions.
- Préservation du contexte : si pour un sous-ensemble A nous avons $Pl_1(A) = 0$ et $Pl_2(A) = 0$, alors $Pl_{12}(A) = 0$. Tout évènement impossible le reste après la combinaison.

Cet opérateur est assuré par la multiplication d'une matrice stochastique, induite par m_1 et notée $\mathbf{K}_{m_1}$, par le vecteur de masse $\mathbf{m}_2$. La combinaison peut alors s'écrire sous sa forme matricielle suivante :

$$\mathbf{m}_{12} = \mathbf{K}_{m_1}\mathbf{m}_2, \tag{4.5}$$

$$\text{où } \mathbf{K}_{m_1} = \sum_{A \subseteq \Omega} m_1(A).\mathbf{K}_A, \tag{4.6}$$

avec $\mathbf{K}_A$ la matrice de $\mathbf{m}_A$. Il est montré dans [51] que les matrices $\mathbf{K}_A$ ne dépendent pas de la fonction de masse m. Ce sont des matrices stochastiques de dimension $N \times N$ qui dépendent uniquement de l'élément neutre m_e et d'un paramètre α variant entre 0 et 1.
Pour la suite, nous convenons parfois de noter la matrice stochastique $\mathbf{K}_{m_1}$ par $\mathbf{K}_1$ pour alléger les écritures.

A l'issue de l'analyse des deux comportements principaux d'un processus de combinaison, en l'occurrence le comportement conjonctif et le comportement disjonctif, deux solutions seulement sont admissibles pour m_e. Dans le cas d'une combinaison conjonctive, l'élément neutre est $m_e = m_\Omega$, tandis que dans le cas d'une opération disjonctive, l'élément neutre est alors $m_e = m_\emptyset$. Le choix du comportement de la règle de combinaison impose en effet le choix de l'élément neutre. D'où l'existence de deux classes de règles de combinaison appartenant à la famille des $\alpha-$jonctions. Ces deux classes sont les $\alpha-$conjonctions et les $\alpha-$disjonctions. Dans ce qui suit nous discutons les particularités de chacune de ces classes ainsi que la définition des matrices $\mathbf{K}_A$ associées.

4.2.1.1 $\alpha-$conjonction

Dans ce cas l'élément neutre est $m_e = m_\Omega$. Les matrices $\mathbf{K}_A$ sont définies en fonction du paramètre $\alpha \in [0, 1]$ comme suit :

$$\mathbf{K}_\Omega = \mathbf{I},$$

$$K_A(X,Y) = \begin{cases} \alpha^{|\overline{A\Delta Y}|-|X|}.\overline{\alpha}^{|X|-|A\cap Y|} & \text{si } A\cap Y \subseteq X \subseteq \overline{A\Delta Y}, \\ 0 & \text{sinon.} \end{cases}$$

où $\overline{\alpha} = 1-\alpha$.

Étant donné deux fonctions de masse m_1 et m_2, l'$\alpha-$conjonction entre ces deux fonctions de masse est calculée pour une valeur constante de α. Elle est notée par :

$$m_{12} = m_1 \textcircled{\cap}^{\alpha} m_2. \tag{4.7}$$

Dans le cas particulier où m_1 et m_2 sont définies dans $\Omega = \{a,b\}$, la décomposition de la $\alpha-$conjonction entre ces deux fonctions de masse en fonction des matrices $\mathbf{K}_A$ est :

$$\begin{aligned} \mathbf{m}_1 \textcircled{\cap}^{\alpha} \mathbf{m}_2 &= \mathbf{K}_1 \mathbf{m}_2, \\ &= \Big(m_1(\emptyset)\mathbf{K}_\emptyset + m_1(a)\mathbf{K}_a + m_1(b)\mathbf{K}_b + m_1(\Omega)\mathbf{K}_\Omega\Big).\mathbf{m}_2. \end{aligned} \tag{4.8}$$

sachant que :

$$\mathbf{K}_\emptyset = \begin{bmatrix} \alpha^2 & \alpha & \alpha & 1 \\ \alpha\overline{\alpha} & 0 & \alpha & 0 \\ \alpha\overline{\alpha} & \overline{\alpha} & 0 & 0 \\ \overline{\alpha}^2 & 0 & 0 & 0 \end{bmatrix}, \quad \mathbf{K}_a = \begin{bmatrix} \alpha & 0 & 1 & 0 \\ 0 & \alpha & 0 & 1 \\ \overline{\alpha} & 0 & 0 & 0 \\ 0 & \overline{\alpha} & 0 & 0 \end{bmatrix}, \quad \mathbf{K}_b = \begin{bmatrix} \alpha & 1 & 0 & 0 \\ \overline{\alpha} & 0 & 0 & 0 \\ 0 & 0 & \alpha & 1 \\ 0 & 0 & \overline{\alpha} & 0 \end{bmatrix}.$$

Dans le cas marginal où $\alpha = 1$, la matrice $\mathbf{K}_1$ calculée en utilisant l'élément neutre $m_e = m_\Omega$ devient égale à la matrice de spécialisation dempsterienne, autrement dit $\mathbf{K}_1 = \mathbf{S}_1$. La combinaison a alors un comportement conjonctif pur. Ainsi on peut écrire $\mathbf{m}_1 \textcircled{\cap}^1 \mathbf{m}_2 = \mathbf{S}_1\mathbf{m}_2 = \mathbf{m}_1 \textcircled{\cap} \mathbf{m}_2$.
Le cas $\alpha = 0$ correspond à la règle $\underline{\textcircled{\cap}}$. Toute autre valeur de α traduit une loi de combinaison intermédiaire entre la règle conjonctive $\textcircled{\cap}$ et la règle $\underline{\textcircled{\cap}}$.

4.2.1.2 $\alpha-$disjonction

Tout comme les $\alpha-$conjonctions, la classe de lois de combinaison dites $\alpha-$disjonctives est définie en choisissant cette fois-ci l'élément neutre du comportement disjonctif qui est $m_e = m_\emptyset$. De même, les matrices $\mathbf{K}_A$ sont définies en fonction du paramètre $\alpha \in [0,1]$ comme suit :

$$\mathbf{K}_\emptyset = \mathbf{I},$$

$$K_A(X,Y) = \begin{cases} \alpha^{|X|-|A\Delta Y|}.\overline{\alpha}^{|A\cup Y|-|X|} & \text{si } A\Delta Y \subseteq X \subseteq A \cup Y, \\ 0 & \text{sinon.} \end{cases}$$

La règle de combinaison $\alpha-$disjonctive est notée $\textcircled{\cup}^{\alpha}$. Quand $\alpha = 1$, nous obtenons un comportement purement disjonctif. C'est à dire $m_1 \textcircled{\cup}^{1} m_2 = m_1 \textcircled{\cup} m_2$. En contrepartie, si $\alpha = 0$, le comportement de la règle de combinaison est celui d'un opérateur disjonctif exclusif $\underline{\textcircled{\cup}}$. Dans [52], Smets a montré que les règles $\alpha-$conjonctives et $\alpha-$disjonctives sont liées par les lois de *De Morgan*. On obtient alors :

$$\begin{aligned} \overline{m_1 \textcircled{\cup}^{\alpha} m_2} &= \overline{m_1} \textcircled{\cap}^{\alpha} \overline{m_2}, \\ \overline{m_1 \textcircled{\cap}^{\alpha} m_2} &= \overline{m_1} \textcircled{\cup}^{\alpha} \overline{m_2}. \end{aligned} \tag{4.9}$$

Dans le cas particulier où $\alpha = 1$, nous aboutissons à la dualité entre le comportement conjonctif pur et le comportement disjonctif pur énoncé dans l'équation (4.1).

4.2.2 Calcul des $\alpha-$jonctions

Quand les $\alpha-$jonctions sont apparues, une des limites à leur utilisation fut la complexité calculatoire accrue en comparaison des règles habituelles. En effet, le calcul des différentes matrice $\mathbf{K}_i$ régissant ces règles de combinaison n'est pas une tâche facile si l'on se réfère à leurs définitions. Dans [42], Pichon et Denœux ont développé des mécanismes de calcul simplifiés des combinaisons des croyances à l'aide des $\alpha-$jonctions. Ces mécanismes sont basées sur la généralisation des mécanismes utilisés dans le calcul des combinaisons purement conjonctives ou purement disjonctives.
Deux méthodes de calcul des $\alpha-$jonctions sont présentées dans les sous-sections suivantes. La première explique le calcul classique basée sur les principes d'$\alpha-$conditionnement et d'$\alpha-$généralisation, tandis que la deuxième méthode met en exergue les principes du calcul matriciel des $\alpha-$jonctions dans l'espace vectoriel des fonctions de masse.

4.2.2.1 Calcul classique d'une $\alpha-$jonction

L'opération de conditionnement est utilisée dans un processus de fusion d'informations conjonctif dès lors qu'une nouvelle connaissance du problème affirme avec certitude que la solution ω_0 est contenue dans un sous-ensemble $B \subset \Omega$. Cette nouvelle connaissance est modélisée dans le cadre de la TFC par $m(B) = 1$ et notée m_B. Le calcul du conditionnement d'une fonction de masse quelconque m_0 par le sous-ensemble B est effectué par la révision conjonctive de la masse m par m_B. On écrit alors $\mathbf{m}[B] = \mathbf{m} \textcircled{\cap} \mathbf{m_B}$. L'opération de conditionnement est réalisée à l'issue d'un transfert de toute connaissance

initialement définie dans Ω vers le sous ensemble $B \subset \Omega$. Dans [42], Pichon et Denœux s'inspirent de cette notion pour définir l'opérateur d'**$\alpha-$conditionnement**.

Définition 10. *L'$\alpha-$conditionnement [42] d'une fonction de masse m par un sous-ensemble $B \subset \Omega$ est égal à l'$\alpha-$conjonction de cette fonction de masse avec la masse catégorique focalisée en B notée m_B.*

Le résultat de l'opération d'$\alpha-$conditionnement d'une fonction de masse m par un sous-ensemble $B \subset \Omega$ est noté $m[B]^\alpha = m \textcircled{\cap}^\alpha m_B$. Cette opération est formulée dans [42] par l'expression suivante :

$$m[B]^\alpha(A) = \sum_{(X \cap B) \cup (\overline{X} \cap \overline{B} \cap C) = A} m(X) m_{[\alpha]}(C), \tag{4.10}$$

où $m_{[\alpha]}$ est la fonction de masse définie pour tout $A \subseteq \Omega$ par la formule :

$$m_{[\alpha]}(A) = \alpha^{|\overline{A}|}(1-\alpha)^{|A|}.$$

Comme nous le constatons, l'opération du $\alpha-$conditionnement n'est autre qu'un cas particulier de la règle $\alpha-$conjonctive dans lequel l'une des deux masses combinées est catégorique. La généralisation du calcul de la combinaison $\alpha-$conjonctive entre deux fonctions de masse quelconques m_1 et m_2 se base sur l'opération du $\alpha-$conditionnement. Elle est énoncée dans [42] par l'équation :

$$m_1 \textcircled{\cap}^\alpha m_2(A) = \sum_{B \subseteq \Omega} m_1[B]^\alpha(A) m_2(B),$$

ou encore

$$m_1 \textcircled{\cap}^\alpha m_2(A) = \sum_{(X \cap B) \cup (\overline{X} \cap \overline{B} \cap C) = A} m_1(X) m_2(B) \alpha^{|\overline{C}|}(1-\alpha)^{|C|}. \tag{4.11}$$

D'une façon analogue, une méthode similaire est développée pour élaborer un calcul simplifié des $\alpha-$disjonctions dans [42], La règle de combinaison $\alpha-$disjonctive $\textcircled{\cup}^\alpha$ est calculée cette fois-ci par la formule :

$$m_1 \textcircled{\cup}^\alpha m_2(A) = \sum_{(X \Delta B) \cup (X \cap B \cap C) = A} m_1(X) m_2(B) \alpha^{|C|}(1-\alpha)^{|\overline{C}|}. \tag{4.12}$$

où $X \Delta B$ désigne la différence symétrique entre les sous-ensembles X et B.

4.2.2.2 Calcul matriciel d'une $\alpha-$conjonction

Dans [51], Smets a montré que la matrice $\mathbf{M}_\alpha$ composée des vecteurs propres de la matrice de croyance $\mathbf{K}_m$, induite d'une masse notée m et définie dans Ω, intervient dans le calcul matriciel d'une $\alpha-$jonction. En effet, on distingue deux cas différents pour la matrice $\mathbf{K}_m$. Selon le choix d'une $\alpha-$conjonction cette matrice est notée $\mathbf{K}_m^{\cap}$ ou d'une $\alpha-$disjonction où elle est notée $\mathbf{K}_m^{\cup}$. Notons que la matrice $\mathbf{K}_m^{\cap}$ généralise, dans le cas $\alpha-$conjonctif, la matrice de spécialisation dempsterienne $\mathbf{S}_m$. La matrice $\mathbf{M}_\alpha$ n'est, en effet, qu'une généralisation de la matrice d'incidence $\mathbf{M}$ définie dans le chapitre 1, section 1.3.2. De façon analogue à la règle de combinaison conjonctive, cette matrice permet de généraliser la notion de communalité. La fonction d'$\alpha-$communalité g est alors définie dans l'expression :

$$\forall \mathbf{m} \in \varepsilon_\Omega,\ \mathbf{g} = \mathbf{M}_\alpha \mathbf{m}. \tag{4.13}$$

Smets [51] a aussi montré que, pour deux fonctions de masse quelconques m_1 et m_2, la règle $\alpha-$conjonctive se traduit pour tout $A \subseteq \Omega$ par :

$$g_{1\textcircled{\cap}^\alpha 2}(A) = g_1(A).g_2(A), \tag{4.14}$$

sachant que $\mathbf{g}_1 = \mathbf{M}_\alpha.\mathbf{m}_1$ et $\mathbf{g}_2 = \mathbf{M}_\alpha.\mathbf{m}_2$. L'expression (4.14) montre que la combinaison de deux fonctions de masse m_1 et m_2 en utilisant une $\alpha-$conjonction peut s'exécuter simplement par le produit point-à-point de leurs fonctions d'$\alpha-$communalité respectives g_1 et g_2.
La fonction d'$\alpha-$communalité résultante $g_{1\textcircled{\cap}^\alpha 2}$ peut être déduite de la masse résultante du processus d'$\alpha-$conjonction entre m_1 et m_2 :

$$\mathbf{g}_{1\textcircled{\cap}^\alpha 2} = \mathbf{M}_\alpha.\mathbf{m}_{1\textcircled{\cap}^\alpha 2}. \tag{4.15}$$

A partir des équations (4.14) et (4.15), nous obtenons le résultat suivant qui permet de calculer une $\alpha-$conjonction à l'aide des fonctions d'$\alpha-$communalité :

$$\mathbf{m}_{1\textcircled{\cap}^\alpha 2} = {\mathbf{M}_\alpha}^{-1}.Diag(\mathbf{g}_1).\mathbf{g}_2, \tag{4.16}$$

où $Diag(\mathbf{g}_1)$ est la matrice diagonale dont les éléments diagonaux sont les composantes du vecteur $\mathbf{g}_1$. A partir de l'équation (4.16), nous retrouvons facilement la relation qui lie la matrice $\mathbf{K}_{m_1}^{\cap}$ avec la matrice $\mathbf{M}_\alpha$:

$$\mathbf{K}_{m_1}^{\cap} = {\mathbf{M}_\alpha}^{-1}.Diag(\mathbf{M}_\alpha.\mathbf{m}_1).\mathbf{M}_\alpha. \tag{4.17}$$

Dans les travaux de Pichon et Denœux [42], un algorithme simple d'obtention de la matrice $\mathbf{M}_\alpha$ est présenté. Grâce à ce résultat, l'expression (4.17) devient particulièrement

intéressante puisqu'elle permet de simplifier le calcul de la matrice $\mathbf{K}^{\cap}_{m_1}$. Le calcul de la matrice $\mathbf{M}_\alpha$ est donné par un produit de *Kronecker* :

$$\begin{aligned} {\mathbf{M}_\alpha}^{i+1} &= \mathbf{Kron}\left(\begin{bmatrix} 1 & 1 \\ \alpha-1 & 1 \end{bmatrix}, {\mathbf{M}_\alpha}^{i}\right), \\ {\mathbf{M}_\alpha}^{1} &= 1, \end{aligned} \tag{4.18}$$

où ${\mathbf{M}_\alpha}^{i}$ correspond à la matrice $\mathbf{M}_\alpha$ dans le cas où $|\Omega| = i$. En examinant l'équation (4.18), nous remarquons facilement que, quand $\alpha = 1$, la matrice $\mathbf{M}_\alpha$ est égale à la matrice d'incidence $\mathbf{M}$. Dans ce cas, la fonction g associée à m se confond avec la fonction de communalité q.

4.2.2.3 Calcul matriciel d'une $\alpha-$disjonction

D'une façon analogue à la famille des $\alpha-$conjonctions, le calcul matriciel d'une $\alpha-$disjonction de deux fonctions de masse m_1 et m_2 est donné par :

$$\mathbf{m}_{1 \textcircled{\cup}^{\alpha} 2} = \mathbf{K}^{\cup}_{m_1}.\mathbf{m}_2, \tag{4.19}$$

où $\mathbf{K}^{\cup}_{m_1}$ est la matrice d'$\alpha-$généralisation de la fonction de masse m_1. Cette combinaison peut s'exprimer par un simple produit des deux fonctions d'$\alpha-$implicabilité h_1 et h_2 correspondant respectivement aux masses m_1 et m_2. Il en découle donc :

$$h_{1 \textcircled{\cup}^{\alpha} 2}(A) = h_1(A).h_2(A), \tag{4.20}$$

avec $A \subseteq \Omega$ et $\mathbf{h}_1 = \mathbf{B}_\alpha.\mathbf{m_1}$ sachant que la matrice $\mathbf{B}_\alpha$ est une matrice stochastique dont les éléments dépendent de α.
Comme pour le cas des $\alpha-$conjonctions, la relation matricielle entre la matrice $\mathbf{K}^{\cup}_{m_1}$ avec sa matrice de valeurs propres est donnée par :

$$\mathbf{K}^{\cup}_{m_1} = {\mathbf{B}_\alpha}^{-1}.Diag(\mathbf{B}_\alpha.\mathbf{m}_1).\mathbf{B}_\alpha. \tag{4.21}$$

Selon le théorème 12.1 dans [52], la relation entre les deux matrices d'$\alpha-$spécialisation $\mathbf{K}^{\cap}_{m_1}$ et d'$\alpha-$généralisation $\mathbf{K}^{\cup}_{m_1}$ est :

$$\mathbf{K}^{\cup}_{m_1} = \mathbf{J}\mathbf{K}^{\cap}_{m_1}\mathbf{J}, \tag{4.22}$$

A partir de l'équation (4.22), nous pouvons facilement démontrer que :

$$\mathbf{B}_\alpha = \mathbf{J}\mathbf{M}_\alpha\mathbf{J}. \tag{4.23}$$

L'application du résultat donné par l'équation (4.23) sur (4.18) nous permet de calculer

la matrice $\mathbf{B}_\alpha$ à l'aide d'un calcul itératif basé sur le produit de *Kronecker* :

$$\begin{aligned} {\mathbf{B}_\alpha}^{i+1} &= \mathbf{Kron}\left(\begin{bmatrix} 1 & \alpha-1 \\ 1 & 1 \end{bmatrix}, {\mathbf{B}_\alpha}^{i}\right), \\ {\mathbf{B}_\alpha}^{1} &= 1, \end{aligned} \tag{4.24}$$

où ${\mathbf{B}_\alpha}^{i}$ correspond à la matrice $\mathbf{B}_\alpha$ dans le cas où $|\Omega| = i$. Nous pouvons déduire que dans le cas où $\alpha = 1$ la fonction h associée à m se confond avec la fonction d'implicabilité b.

4.3 Distances basées sur les matrices $\alpha-$jonctives

La distance entre fonctions de croyance peut se définir à partir de la distance entre les matrices de croyance correspondantes puisqu'il existe une correspondance bijective entre les fonctions de masse et les matrices de croyance relatives à une $\alpha-$jonction donnée. En conséquent, la norme de la différence entre deux matrices de croyance peut permettre de définir une distance entre fonctions de croyance à l'instar des distances de spécialisation vues dans le chapitre 3.

Définition 11. *une* ***distance d'α-spécialisation*** $d^{\cap}$ *est une fonction définie en s'appuyant sur une norme matricielle* $\|.\|$ *et relative à une* $\alpha-$*conjonction telle que :*

$$\begin{aligned} d^{\cap} : \varepsilon_\Omega \times \varepsilon_\Omega &\longrightarrow \mathbb{R}_+ \\ m_1 \times m_2 &\rightarrow \frac{1}{\rho}\left\|\mathbf{K}^{\cap}_{1,\alpha} - \mathbf{K}^{\cap}_{2,\alpha}\right\|, \end{aligned} \tag{4.25}$$

où $\mathbf{K}^{\cap}_{i,\alpha}$ *est la matrice d'*α*-spécialisation correspondant à la masse* m_i *et* ρ *un coefficient de normalisation sachant que* $d^{\cap}(m_\emptyset, m_\Omega) = 1$.

Définition 12. *une* ***distance de α-généralisation*** $d^{\cup}$ *est une fonction définie en s'appuyant sur une norme matricielle* $\|.\|$ *et relative à une* $\alpha-$*disjonction telle que :*

$$\begin{aligned} d^{\cup} : \varepsilon_\Omega \times \varepsilon_\Omega &\longrightarrow \mathbb{R}_+ \\ m_1 \times m_2 &\rightarrow \frac{1}{\rho}\left\|\mathbf{K}^{\cup}_{1,\alpha} - \mathbf{K}^{\cup}_{2,\alpha}\right\|, \end{aligned} \tag{4.26}$$

où $\mathbf{K}^{\cup}_{i,\alpha}$ *est la matrice d'*α*-généralisation correspondant à la masse* m_i *et* ρ *un coefficient de normalisation sachant que* $d^{\cup}(m_\emptyset, m_\Omega) = 1$.

La famille des distances d'α-spécialisation est une extension de la famille des distances de spécialisation définie dans [32] qui correspond au cas où $\alpha = 1$. Parmi les normes matricielles existantes, nous nous intéressons dans ce chapitre, comme dans le

chapitre 3, aux normes les plus utilisées, à savoir : les normes **d'opérateur**, aussi connues comme normes induites, et les normes L^k de type *Minkowski*.

Nous remarquons facilement que la norme d'opérateur 1 de toute matrice de croyance est $\|\mathbf{K}\|_{op1} = 1$. Les normes matricielles dites vectorielles sont définies comme des normes L^k si les matrices sont vues comme des vecteurs dans l'espace $\mathcal{M}_N$. Les normes vectorielles et matricielles du type L^k sont toutes deux notées $\|.\|_k$. Elles se distinguent facilement puisque les vecteurs sont notés en minuscules tandis que les matrices sont notées en majuscules.

4.3.1 Propriétés des distances de d'$\alpha-$spécialisation et d'$\alpha-$généralisation

Comme pour les distances de spécialisation définies dans le chapitre 3, les distances matricielles entre les fonctions de croyance généralisées aux α-jonctions possèdent un certain nombre de propriétés. Il est intéressant d'étudier si ces nouvelles distances ont les mêmes propriétés que les distances de spécialisation. Cette étude est donnée ci-après.

4.3.1.1 Résultats préliminaires

Nous donnons tout d'abord quelques résultats préliminaires qui serviront dans les preuves d'adéquation des distances d'$\alpha-$spécialisation et d'$\alpha-$généralisation aux différentes propriétés traitées dans cette section.

Remarque 1. *Supposons m_1 et m_2 deux fonctions de masse définies dans Ω. $\mathbf{K}_{m_1}$ et $\mathbf{K}_{m_2}$ sont les matrices de croyance respectives de m_1 et m_2 relatives à une α-jonction notée $\odot^\alpha$. Notons $\mathbf{K}_{m_{12}}$ la matrice de croyance de $m_1 \odot^\alpha m_2$. Alors, nous avons d'après [52] :*

$$\mathbf{K}_{m_{12}} = \mathbf{K}_{m_1}.\mathbf{K}_{m_2}. \tag{4.27}$$

Démonstration. Supposons m_1, m_2 et m_3 trois fonctions de masse définies dans Ω et $\mathbf{K}_{m_1}$, $\mathbf{K}_{m_2}$ et $\mathbf{K}_{m_3}$ leurs matrices de croyance respectives relativement à une α-jonction notée $\odot^\alpha$. Puisque toutes les α-jonctions sont associatives, on peut écrire :

$$\begin{aligned}
(m_1 \odot^\alpha m_2) \odot^\alpha m_3 &= m_1 \odot^\alpha (m_2 \odot^\alpha m_3) \\
\Leftrightarrow \mathbf{K}_{m_{12}}.\mathbf{m}_3 &= \mathbf{K}_{m_1}.(\mathbf{m}_{2\odot^\alpha 3}) \\
\Leftrightarrow \mathbf{K}_{m_{12}}.\mathbf{m}_3 &= \mathbf{K}_{m_1}.(\mathbf{K}_{m_2}.\mathbf{m}_3).
\end{aligned}$$

L'équation précédente est vérifiée pour toute fonction de masse m_3, et donc : $\mathbf{K}_{m_{12}} = \mathbf{K}_{m_1}.\mathbf{K}_{m_2}$. □

Un premier résultat concernant les nouvelles familles de distances définies dans ce chapitre est la dualité entre les distances d'$\alpha-$spécialisation et les distances d'$\alpha-$généralisation

qui trouve ses origines dans les lois de *De Morgan* démontrées dans [52]. Cette dualité est formalisée dans la proposition suivante :

Proposition 13. *Supposons $\alpha \in [0,1]$. Soit $d^{\cap,\alpha}$ une distance d'$\alpha-$spécialisation relative à la règle α-conjonctive $\textcircled{\cap}^{\alpha}$. Soit $d^{\cup,\alpha}$ la distance d'$\alpha-$généralisation relative à la règle α-disjonctive $\textcircled{\cup}^{\alpha}$. Pour toutes fonctions de masse m_1 et m_2 dans un cadre Ω, nous avons :*

$$d^{\cap,\alpha}(m_1,m_2) = d^{\cup,\alpha}(\overline{m}_1,\overline{m}_2). \tag{4.28}$$

Démonstration. Soit $\mathbf{K}^{\cap}_{m_i}$ la matrice d'α-spécialisation et $\mathbf{K}^{\cup}_{m_i}$ la matrice d'α-généralisation de m_i avec $i \in \{1;2\}$. Selon le théorème 12.1 dans [52], nous avons $\mathbf{K}^{\cap}_{m_i} = \mathbf{J}\mathbf{K}^{\cup}_{m_i}\mathbf{J}$ avec $\mathbf{J}$ la matrice antidiagonale définie dans le chapitre 1. Il en résulte que :

$$\begin{aligned} d^{\cap,\alpha}(m_1,m_2) &= \frac{1}{\rho}\left\|\mathbf{K}^{\cap}_{m_1} - \mathbf{K}^{\cap}_{m_2}\right\|, \\ &= \frac{1}{\rho}\left\|\mathbf{J}\mathbf{K}^{\cup}_{\overline{m}_1}\mathbf{J} - \mathbf{J}\mathbf{K}^{\cup}_{\overline{m}_2}\mathbf{J}\right\|, \\ &= \frac{1}{\rho}\left\|\mathbf{J}\left(\mathbf{K}^{\cup}_{\overline{m}_1} - \mathbf{K}^{\cup}_{\overline{m}_2}\right)\mathbf{J}\right\|. \end{aligned}$$

Étant donné que $\mathbf{J}$ est une matrice de permutation et que n'importe quelle norme matricielle considérée dans ce manuscrit se calcule par un produit point-à-point, nous avons :

$$\left\|\mathbf{J}\left(\mathbf{K}^{\cup}_{\overline{m}_1} - \mathbf{K}^{\cup}_{\overline{m}_2}\right)\mathbf{J}\right\| = \left\|\mathbf{K}^{\cup}_{\overline{m}_1} - \mathbf{K}^{\cup}_{\overline{m}_2}\right\| = \rho d^{\cup,\alpha}(\overline{m}_1,\overline{m}_2).$$

□

La proposition 13 met en exergue le lien de dualité qui existe entre les distances d'α-spécialisation et les distances d'α-généralisation. Quand $\alpha \in \{0;1\}$, une situation bien plus particulière encore se présente à nous. Celle-ci est illustrée dans le lemme 3 suivant :

Lemme 3. *Soit $d^{\cap,\alpha}$ une distance d'α-spécialisation relative à la règle α-conjonctive $\textcircled{\cap}^{\alpha}$ basée sur une norme vectorielle ou d'opérateur. Soit $d^{\cup,\alpha}$ une distance d'α-généralisation relative à la règle α-disjonctive $\textcircled{\cup}^{\alpha}$ et issue de la même norme. Pour toutes fonctions de masse m_1 et m_2 définies dans Ω, on a :*

$$d^{\cap,0} = d^{\cup,0}, \tag{4.29}$$
$$d^{\cap,1} = d^{\cup,1}. \tag{4.30}$$

Démonstration. Soient $\mathbf{K}^{\cap}_{i,0}$ et $\mathbf{K}^{\cup}_{i,0}$ les matrices de 0-spécialisation et de 0-généralisation

de m_i avec $i \in \{1;2\}$. On montre que $\mathbf{JK}^{\cap}_{i,0} = \mathbf{K}^{\cup}_{i,0}$. Le même raisonnement que celui de la preuve de la proposition 13 donne $d^{\cap,0} = d^{\cup,0}$.
Soient $\mathbf{K}^{\cap}_i$ et $\mathbf{K}^{\cup}_i$ les matrices de spécialisation et de généralisation de m_i avec $i \in \{1;2\}$. On montre que pour tout A et $B \subseteq \Omega$ on a $K^{\cap}_i(A \cap B, B) = K^{\cup}_i\left(A \cup \overline{B}, \overline{B}\right)$. Autrement dit, les matrices $\mathbf{K}^{\cap}_i$ et $\mathbf{K}^{\cup}_i$ contiennent les mêmes éléments mais à des positions différentes. Aussi, un raisonnement similaire à celui de la preuve de la proposition 13 donne $d^{\cap,1} = d^{\cup,1}$. □

Ce résultat montre clairement que les distances d'α-spécialisation et les distances d'α-généralisation sont identiques pour les valeurs extrêmes de α. En outre, la proposition 13 nous permet d'anticiper le fait que si une distance d'α-spécialisation satisfait une propriété donnée, alors il en va probablement de même pour sa distance d'α-généralisation homologue.

4.3.1.2 Propriétés métriques

Proposition 14. *Toutes les distances d'α-spécialisation où d'α-généralisation sont des métriques complètes normalisées.*

Démonstration. Le fait que les distances d'α-spécialisation où d'α-spécialisation sont définies à partir de normes matricielles complètes implique que ces distances sont des métriques complètes normalisées.
Concernant la normalisation, la même explication que dans le cas des distances de spécialisation reste valable dans le cas présent. □

La section suivante est consacrée à la propriété de consistance des distances d'α-spécialisation et des distances d'α-généralisation, tandis que la propriété structurelle est traitée dans la section 4.3.1.3.

4.3.1.3 Propriétés structurelles des distances matricielles entre fonctions de croyance

Comme évoqué précédemment, les relations d'inclusion et d'intersection entre les éléments focaux d'une fonction de masse devraient induire des conséquences dans l'évaluation des distances. En effet, si $A \cap B \neq \emptyset$ tandis que $A \cap C = \emptyset$, nous en déduisons intuitivement que les vecteurs de base $\mathbf{m}_A$ et $\mathbf{m}_B$ sont plus proches que $\mathbf{m}_A$ et $\mathbf{m}_C$. Une distance entre fonctions de croyance qui tient compte de cette interaction est une distance qui respecte le principe structurel. Pour rappel, une distance structurée est définie dans la définition 6 du chapitre 3.
Parmi les familles de distances d'$\alpha-$spécialisation et d'$\alpha-$généralisation définies dans

ce travail, seule la distance d_1 basée sur la norme L^1 est structurée. Ce résultat est principalement basé sur le lemme suivant :

Lemme 4. *Supposons m_A et m_B deux fonctions de masse catégoriques portant sur A et B, deux sous-ensembles de Ω. Soit d_1 la distance L^1 relative à une α-jonction donnée. $\forall \alpha \in [0,1]$, le résultat suivant est vérifié pour d_1 :*

$$d_1\left(m_A, m_B\right) = \frac{2N}{\rho}\left(1 - \frac{\alpha^{\max\{|A\backslash B|;|B\backslash A|\}}}{2^{|A\Delta B|}}\right). \tag{4.31}$$

Démonstration. La démonstration du lemme 4 est donnée dans l'annexe A.7 □

A l'aide du lemme précédent, nous pouvons à présent introduire et prouver la propriété suivante :

Proposition 15. *Soit d_1 la distance basée sur la norme L^1 relativement à une α-jonction donnée. d_1 est structurée si $\alpha = 0$ ou $\frac{1}{2} \leq \alpha < 1$. Elle est strictement structurée si $\alpha = 1$.*

Démonstration. Considérons d'abord que d_1 est définie relativement à une règle α-conjonctive. A partir de la proposition 4, nous avons : $\forall \alpha \in [0,1]$:

$$\begin{aligned} d_1\left(m_A; m_B\right) &= \frac{2N}{\rho}\left(1 - \frac{\alpha^{\max\{|A\backslash B|;|B\backslash A|\}}}{2^{|A\Delta B|}}\right), \\ &\geq \frac{2N}{\rho}\left(1 - \frac{1^{\max\{|A\backslash B|;|B\backslash A|\}}}{2^{|A\Delta B|}}\right), \\ &\geq \frac{2N}{\rho}\left(1 - \left(\frac{1}{2}\right)^{|A\Delta B|}\right). \end{aligned} \tag{4.32}$$

Supposons à présent que $\alpha \geq \frac{1}{2}$, nous écrivons alors :

$$\begin{aligned} \alpha \geq \frac{1}{2} &\Leftrightarrow \left(\frac{\alpha}{2}\right)^{|A\Delta B|} \geq \left(\frac{1}{4}\right)^{|A\Delta B|}, \\ &\Leftrightarrow -\left(\frac{\alpha}{2}\right)^{|A\Delta B|} \leq -\left(\frac{1}{2}\right)^{|A\Delta B|+1}, \\ &\Leftrightarrow d_1\left(m_A; m_B\right) \leq \frac{2N}{\rho}\left(1 - \left(\frac{1}{2}\right)^{|A\Delta B|+1}\right). \end{aligned} \tag{4.33}$$

Soient A, B, C et D les quatre sous-ensembles de Ω sachant que $|A\Delta B| < |C\Delta D|$. Nous

avons :

$$\begin{aligned} |A\Delta B| < |C\Delta D| &\Leftrightarrow |A\Delta B| + 1 \leq |C\Delta D|, \\ &\Leftrightarrow \frac{2N}{\rho}\left(1 - \left(\frac{1}{2}\right)^{|A\Delta B|+1}\right) \leq \frac{2N}{\rho}\left(1 - \left(\frac{1}{2}\right)^{|C\Delta D|}\right), \\ &\Rightarrow d_1\left(m_A; m_B\right) \leq d_1\left(m_C; m_D\right), \end{aligned}$$

car ces distances se trouvent dans deux intervalles disjoints.

Réciproquement, supposons que nous avons $d_1\left(m_A; m_B\right) < d_1\left(m_C; m_D\right)$. A partir du raisonnement précédent, l'hypothèse $|A\Delta B| > |C\Delta D|$ induit une contradiction, d'où $|A\Delta B| \leq |C\Delta D|$. Finalement, puisque la cardinalité de la différence symétrique est la distance d'ensembles de Hamming, alors d_1 est structurée si $\frac{1}{2} \leq \alpha < 1$.
Pour le cas où $\alpha = 1$, la preuve que d_1 est strictement structurée est donnée dans l'annexe A.2. Pour la cas $\alpha = 0$, nous avons $d_1\left(m_A; m_B\right) = \frac{2N}{\rho} 1_{A \neq B}$. La distance adopte un comportement binaire. Elle est cependant structurée puisqu'elle est constante quand $A \neq B$ et elle est nulle quand $A = B$.
Des contre-exemples peuvent être trouvés pour montrer que d_1 n'est pas structurée si $0 < \alpha < \frac{1}{2}$.
Intéressons nous à présent au cas α-disjonctif. Pour la suite de cette démonstration, nous distinguons $d_{1,\cap}$ et $d_{1,\cup}$. A partir de la proposition 13 et étant donné que $d_{1,\cap}$ est structurée par rapport à la distance d'ensembles de Hamming, nous avons alors :

$$\begin{aligned} |A\Delta B| \leq |C\Delta D| &\Leftrightarrow d_{1,\cap}\left(m_A, m_B\right) \leq d_{1,\cap}\left(m_C, m_D\right), \\ &\Leftrightarrow d_{1,\cup}\left(\overline{m}_A, \overline{m}_B\right) \leq d_{1,\cup}\left(\overline{m}_C, \overline{m}_D\right), \\ &\Leftrightarrow d_{1,\cup}\left(m_{\overline{A}}, m_{\overline{B}}\right) \leq d_{1,\cup}\left(m_{\overline{C}}, m_{\overline{D}}\right). \end{aligned}$$

Puisque $|A\Delta B| \leq |C\Delta D| \Leftrightarrow |\overline{A}\Delta\overline{B}| \leq |\overline{C}\Delta\overline{D}|$, $d_{1,\cup}$ hérite de $d_{1,\cap}$ la propriété structurelle (strictement ou pas). □

4.3.1.4 Propriété de consistance avec les α-jonctions

On rappelle qu'une distance est consistante avec une règle de combinaison donnée si l'ajout d'une certaine information à deux fonctions de masse les rend plus proches. Dans cette section, nous donnons plusieurs résultats concernant la consistance des distances d'α-spécialisation et les distances d'α-généralisation. Le résultat suivant montre la consistance de ces distances dans le cas où la norme d'opérateur 1 est choisie :

Proposition 16. *Toute distance d'α-spécialisation ou d'α-généralisation définie en fonction de la norme d'opérateur-1 d_{op1} est consistante avec la règle d'α-jonction qui lui*

correspond.

Démonstration. Supposons m_1, m_2 et m_3 trois fonctions de masse définies dans Ω. $\mathbf{K}_1$, $\mathbf{K}_2$ et $\mathbf{K}_3$ sont leurs matrices de croyance relatives à une certaine α-jonction notée $\odot^\alpha$. La norme d'opérateur 1 possède la propriété de sous-multiplicativité. C'est à dire que pour toutes matrices $\mathbf{A}$ et $\mathbf{B}$, nous avons :

$$\|\mathbf{AB}\|_{op1} \leq \|\mathbf{A}\|_{op1} . \|\mathbf{B}\|_{op1} . \tag{4.34}$$

Nous pouvons écrire alors :

$$\begin{aligned} \|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{K}_3\|_{op1} &\leq \|\mathbf{K}_1 - \mathbf{K}_2\|_{op1} . \|\mathbf{K}_3\|_{op1} , \\ \|\mathbf{K}_1\mathbf{K}_3 - \mathbf{K}_2\mathbf{K}_3\|_{op1} &\leq d_{op1}(m_1, m_2) . \|\mathbf{K}_3\|_{op1} , \\ d_{op1}(m_1 \odot^\alpha m_3, m_2 \odot^\alpha m_3) &\leq d_{op1}(m_1, m_2) . \|\mathbf{K}_3\|_{op1} . \end{aligned}$$

Par définition, toutes les matrices de croyance sont stochastiques, d'où $\|\mathbf{K}_3\|_{op1} = 1$. Par conséquent, la distance d_{op1} est consistante avec la règle de combinaison $\odot^\alpha$, nous pouvons alors écrire :

$$d_{op1}(m_1 \odot^\alpha m_3, m_2 \odot^\alpha m_3) \leq d_{op1}(m_1, m_2) . \tag{4.35}$$

□

Un résultat similaire est obtenu quand les distances d'α-spécialisation ou d'α-généralisation sont basées sur la norme L^1. Afin de mettre en œuvre la preuve de ce résultat, nous introduisons tout d'abord la propriété suivante :

Proposition 17. *Soient m_1 et m_2 deux fonctions de masse définies dans Ω et deux ensembles A et B sachant que $A \subseteq B \subseteq \Omega$. Dans ces conditions, les propriétés suivantes sont vérifiées pour toute α-conjonction $\textcircled{\cap}^\alpha$ et toute α-disjonction $\textcircled{\cup}^\alpha$:*

$$\|\boldsymbol{m}_1 \textcircled{\cap}^\alpha \boldsymbol{m}_A - \boldsymbol{m}_2 \textcircled{\cap}^\alpha \boldsymbol{m}_A\|_1 \leq \|\boldsymbol{m}_1 \textcircled{\cap}^\alpha \boldsymbol{m}_B - \boldsymbol{m}_2 \textcircled{\cap}^\alpha \boldsymbol{m}_B\|_1 , \tag{4.36}$$

$$\|\boldsymbol{m}_1 \textcircled{\cup}^\alpha \boldsymbol{m}_A - \boldsymbol{m}_2 \textcircled{\cup}^\alpha \boldsymbol{m}_A\|_1 \geq \|\boldsymbol{m}_1 \textcircled{\cup}^\alpha \boldsymbol{m}_B - \boldsymbol{m}_2 \textcircled{\cup}^\alpha \boldsymbol{m}_B\|_1 . \tag{4.37}$$

Démonstration. La démonstration de la propriété 17 est donnée dans l'annexe A.8. □

Corollaire 1. *Soient m_1 et m_2 deux fonctions de masse définies dans Ω. $\mathbf{K}_1$ et $\mathbf{K}_2$ sont leurs matrices de croyance respectives relativement à une α-jonction particulière. Pour une α-jonction quelconque, nous avons :*

$$d_{op1}(m_1, m_2) = \frac{1}{\rho}\|\mathbf{K}_1 - \mathbf{K}_2\|_{op1} = \frac{1}{\rho}\|\mathbf{m}_1 - \mathbf{m}_2\|_1 . \tag{4.38}$$

Démonstration. Utilisons la définition suivante de la norme matricielle d'opérateur 1 $\mathbf{K}_1 - \mathbf{K}_2$:

$$\|\mathbf{K}_1 - \mathbf{K}_2\|_{op1} = \max_{B\subset\Omega} \sum_{A\subset\Omega} |K_1(A,B) - K_2(A,B)|. \tag{4.39}$$

Intéressons-nous à présent à la classe des α-conjonctions et à la classe des α-disjonctions séparément :

– Dans le cas d'une α-conjonction, l'équation (4.39) s'écrit :

$$\begin{aligned} \|\mathbf{K}_1 - \mathbf{K}_2\|_{op1} &= \max_{B\subset\Omega} \sum_{A\subset\Omega} |\left(m_1 \textcircled{\cap}^{\alpha} m_B\right)(A) - \left(m_2 \textcircled{\cap}^{\alpha} m_B\right)(A)|, \\ &= \max_{B\subset\Omega} \|(\mathbf{m}_1 \textcircled{\cap}^{\alpha} \mathbf{m}_B) - (\mathbf{m}_2 \textcircled{\cap}^{\alpha} \mathbf{m}_B)\|_1 . \end{aligned}$$

En utilisant la proposition 17, il est clair que la norme maximum est obtenue pour le plus grand ensemble B. C'est à dire que quand $B = \Omega$, donc :

$$\begin{aligned} \|\mathbf{K}_1 - \mathbf{K}_2\|_{op1} &= \|(\mathbf{m}_1 \textcircled{\cap}^{\alpha} \mathbf{m}_\Omega) - (\mathbf{m}_2 \textcircled{\cap}^{\alpha} \mathbf{m}_\Omega)\|_1 . \\ &= \|\mathbf{m}_1 - \mathbf{m}_2\|_1 . \end{aligned}$$

– Dans le cas des α-disjonctions, l'équation (4.39) s'écrit :

$$\begin{aligned} \|\mathbf{K}_1 - \mathbf{K}_2\|_{op1} &= \max_{B\subset\Omega} \sum_{A\subset\Omega} |\left(m_1 \textcircled{\cup}^{\alpha} m_B\right)(A) - \left(m_2 \textcircled{\cup}^{\alpha} m_B\right)(A)|, \\ &= \max_{B\subset\Omega} \|(\mathbf{m}_1 \textcircled{\cup}^{\alpha} \mathbf{m}_B) - (\mathbf{m}_2 \textcircled{\cup}^{\alpha} \mathbf{m}_B)\|_1 . \end{aligned}$$

En utilisant la proposition 17, il est clair que la norme maximum est obtenue pour le plus petit ensemble B. C'est à dire que quand $B = \emptyset$, donc :

$$\begin{aligned} \|\mathbf{K}_1 - \mathbf{K}_2\|_{op1} &= \|(\mathbf{m}_1 \textcircled{\cup}^{\alpha} \mathbf{m}_\emptyset) - (\mathbf{m}_2 \textcircled{\cup}^{\alpha} \mathbf{m}_\emptyset)\|_1 . \\ &= \|\mathbf{m}_1 - \mathbf{m}_2\|_1 . \end{aligned}$$

□

Ce résultat signifie que la distance d'opérateur 1 d_{op1} est égale à la distance de type L^1 entre des vecteurs fonctions de masse pour une valeur quelconque du coefficient α.

Les deux résultats préliminaires énoncés précédemment, nous permettent à présent d'obtenir la propriété de consistance des distances d'$\alpha-$spécialisation ou d'$\alpha-$généralisation basée sur la norme L^1 à l'aide de la proposition suivante :

Proposition 18. *Toute distance d'$\alpha-$spécialisation ou d'$\alpha-$généralisation basée sur la norme L^1, notée d_1, est consistante avec la règle d'α-jonction qui lui correspond.*

Démonstration. Supposons m_1, m_2 et m_3 trois fonctions de masse définies dans Ω. Supposons aussi que $\mathbf{K}_1$, $\mathbf{K}_2$ et $\mathbf{K}_3$ sont leurs matrices de croyance respectives relatives à une α-jonction notée $\odot^\alpha$. Le corollaire 1 donne :

$$\begin{aligned} \|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{K}_3\|_{op1} &= \|(\mathbf{m}_1 - \mathbf{m}_2) \odot^\alpha \mathbf{m}_3\|_1\,, \\ \text{et } \|(\mathbf{K}_1 - \mathbf{K}_2)\|_{op1} &= \|(\mathbf{m}_1 - \mathbf{m}_2)\|_1\,. \end{aligned}$$

Aussi, la proposition 16 stipule que $\|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{K}_3\|_{op1} \leq \|(\mathbf{K}_1 - \mathbf{K}_2)\|_{op1}$, on peut alors écrire :

$$\|(\mathbf{m}_1 - \mathbf{m}_2) \odot^\alpha \mathbf{m}_3\|_1 \leq \|(\mathbf{m}_1 - \mathbf{m}_2)\|_1\,. \tag{4.40}$$

De plus, la norme 1 peut aussi être exprimée en utilisant les vecteurs colonnes comme suit :

$$\|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{K}_3\|_1 = \sum_{A\subseteq\Omega} \|(\mathbf{m}_1 \odot^\alpha \mathbf{m}_A - \mathbf{m}_2 \odot^\alpha \mathbf{m}_A) \odot^\alpha \mathbf{m}_3\|_1\,.$$

L'équation (4.40) permet d'écrire :

$$\begin{aligned} \|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{K}_3\|_1 &= \sum_{A\subseteq\Omega} \|(\mathbf{m}_1 \odot^\alpha \mathbf{m}_A - \mathbf{m}_2 \odot^\alpha \mathbf{m}_A) \odot^\alpha \mathbf{m}_3\|_1\,, \\ \|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{K}_3\|_1 &\leq \sum_{A\subseteq\Omega} \|(\mathbf{m}_1 \odot^\alpha \mathbf{m}_A - \mathbf{m}_2 \odot^\alpha \mathbf{m}_A)\|_1\,, \\ \|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{K}_3\|_1 &\leq \|(\mathbf{K}_1 - \mathbf{K}_2)\|_1\,, \\ \Leftrightarrow d_1\left(m_1 \odot^\alpha m_3, m_2 \odot^\alpha m_3\right) &\leq d_1\left(m_1, m_2\right). \end{aligned}$$

□

Un autre résultat est obtenu cette fois-ci en observant les distances d'$\alpha-$spécialisation ou d'$\alpha-$généralisation basées sur la norme L^∞. Ce dernier est énoncé dans la proposition suivante :

Proposition 19. *Toute distance d'$\alpha-$spécialisation ou d'$\alpha-$généralisation basée sur la norme L^∞, notée d_∞, est consistante avec la règle d'α-jonction qui lui correspond.*

Démonstration. Soient m_1, m_2 et m_3 trois fonctions de masse définie dans Ω. $\mathbf{K}_1$, $\mathbf{K}_2$ et $\mathbf{K}_3$ sont les matrices de croyance respectives correspondant à m_1, m_2 et m_3 relativement à une α-jonction donnée notée $\odot^\alpha$. La norme L^∞ d'une matrice est égale au maximum des normes L^∞ des vecteurs colonne qui la composent.
Étant donné qu'un vecteur colonne de $\mathbf{K}_i$ peut s'écrire $m_{i|^\alpha B} = m_i \odot^\alpha m_B$ avec $B \subseteq \Omega$,

il existe un sous ensemble X tel que :

$$\begin{aligned}\|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{K}_3\|_\infty &= \left\|(\mathbf{K}_1 - \mathbf{K}_2)\mathbf{m}_{3|^\alpha X}\right\|_\infty, \\ &= \left\|(\mathbf{K}_1 - \mathbf{K}_2)\sum_{Y\subseteq\Omega}\mathbf{m}_{3|^\alpha X}(Y)\mathbf{m}_Y\right\|_\infty, \\ &\leq \sum_{Y\subseteq\Omega}\mathbf{m}_{3|^\alpha X}(Y)\,\|(\mathbf{K}_1 - \mathbf{K}_2)\mathbf{m}_Y\|_\infty.\end{aligned}$$

En outre, la norme L^∞ de $\mathbf{K}_1 - \mathbf{K}_2$ est égale au maximum des normes infinies de ses vecteurs colonnes :

$$\max_{Y\subseteq\Omega}\|(\mathbf{K}_1 - \mathbf{K}_2)\mathbf{m}_Y\|_\infty = \|\mathbf{K}_1 - \mathbf{K}_2\|_\infty.$$

Chaque terme de l'inégalité précédente est maximisé par $\|\mathbf{K}_1 - \mathbf{K}_2\|_\infty$ ce qui donne :

$$\begin{aligned}\|(\mathbf{K}_1 - \mathbf{K}_2)\,\mathbf{K}_3\|_\infty &\leq \sum_{Y\subseteq\Omega}\mathbf{m}_{3|^\alpha X}(Y)\,\|\mathbf{K}_1 - \mathbf{K}_2\|_\infty, \\ &\leq \|\mathbf{K}_1 - \mathbf{K}_2\|_\infty.\end{aligned}$$

Après normalisation, l'inégalité précédente donne :

$$d_\infty\left(m_1 \odot^\alpha m_3, m_2 \odot^\alpha m_3\right) \leq d_\infty\left(m_1, m_2\right).$$

□

4.3.2 Comparaison des distances entre fonctions de croyances

4.3.2.1 Tests basés sur les aspects structurels

Pour étudier la propriété structurelle liée à la structure en treillis de l'espace des fonctions de croyance, nous proposons de visualiser le comportement des différentes distances et dissimilarités évidentielles dans le cadre de l'exemple d'application suivant :

Exemple 8. *Cet exemple est inspiré d'un exemple présenté dans [19] et réutilisé dans [17, 31]. Supposons deux fonctions de masse définies dans la cadre de discernement* $\Omega = \{\omega_i\}_{i=1}^n$ *telles que :*

$$\begin{aligned}m_1 &= m_X, \\ m_2 &= m_{A_i},\end{aligned}$$

avec $X = \{\omega_1, \omega_2, \omega_3\}$ *et* $n = |\Omega| = 7$. *Le sous-ensemble* A_i *varie par inclusions successives d'un singleton à chaque étape de calcul allant de* $\emptyset$ *jusqu'à atteindre* Ω *à la*

dernière étape de calcul :

i : étape de calcul	*élément focal* A_i
0	$\emptyset$
1	$\{\omega_1\}$
2	$\{\omega_1, \omega_2\}$
3	$\{\omega_1, \omega_2\, \omega_3\} = X$
4	$\{\omega_1, ..., \omega_4\}$
...	...
7	Ω

Les résultats sont montrés dans la figure 4.1 pour les distances d'α-spécialisation, les distances évidentielles et les dissimilarités mentionnées dans le chapitre 2. Puisque cette expérience implique des fonctions de masse catégoriques, deux distances d'ensemble sont utilisées : la distance de *Hamming* d_{ham} et la distance de *Jaccard* d_{jac}. Notons qu'il est inutile d'inclure les distances d'α-généralisation dans cet exemple puisque, selon la proposition 13, nous savons qu'elles se comportent de façon similaire aux distances d'α-spécialisation dans un processus de comparaison des fonctions de masse catégoriques.

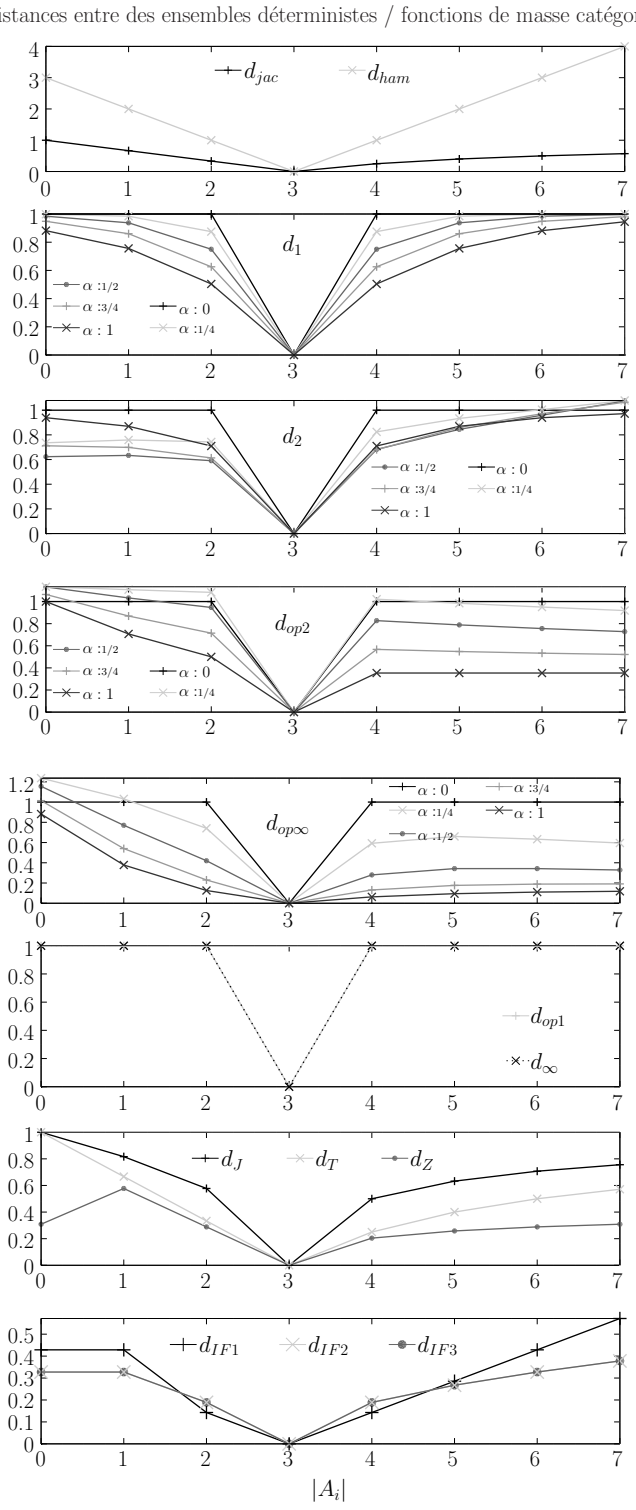

FIGURE 4.1 – Distances évidentielles et dissimilarités entre deux fonctions de masse catégoriques : $m_1 = m_X$ et $m_2 = m_{A_i}$ définies dans Ω, avec $|\Omega| = 7$ et $|X| = 3$. Le sous-ensemble A_i varie par inclusions successives de $\emptyset$ vars Ω et $A_3 = X$. Les distances matricielles impliquées dans cette expérience sont des distances d'α-spécialisation. Cinq valeurs de α sont consideées : $\left\{0; \frac{1}{4}; \frac{1}{2}; \frac{3}{4}; 1\right\}$.

Une distance ou une dissimilarité qui respecte la propriété structurelle devrait être décroissante quand $i \in \{0; ...; 3\}$ et croissante quand $i \in \{3; ...; 7\}$. A cet égard, d_2 (quand $\alpha = 0.25$), d_{op2} (quand $\alpha = 0.5$), $d_{op\infty}$ (quand $\alpha = 0.25$) et d_Z ne sont pas structurées. Par ailleurs beaucoup de distances non-structurées semblent avoir un comportement satisfaisant. Par exemple, on voit que la distance d_1 respecte le principe structurel même si $\alpha \in]0; \frac{1}{2}[$. Il est aussi à remarquer que les courbes de d_{op1} et d_∞ sont superposées quelque soit la valeur de α.

4.3.2.2 Tests de consistance des distances entre fonctions de masse

Dans cette partie, nous illustrons la consistance de toutes les distances relatives à la classe des règles α-conjonctives et α-disjonctives. Dans l'objectif de fournir ces résultats, il est nécessaire de générer aléatoirement des fonctions de masse. Pour des raisons similaires au chapitre 3, section 3.4.4, les fonctions de masse tirées aléatoirement sont ici des fonctions de masse simples. Au lieu de prélever uniformément les fonctions de masse, nous proposons d'utiliser le schéma suivant :

1. choisir aléatoirement un entier i entre 1 et $|2^\Omega| - 1$ avec des chances égales pour tous les nombres,
2. choisir aléatoirement un nombre réel x selon la loi uniforme $\mathcal{U}_{[0,1]}$,
3. allouer une masse x au sous-ensemble dont l'indice est i et une masse $1 - x$ à l'ensemble Ω.

Ce plan génère des fonctions de masse à support simple aléatoires.
Dans la figure 4.2, les taux de consistance pour les règles α-conjonctives et plusieurs distances entre fonctions de croyance sont montrés. Une itération de cette expérience consiste au choix de trois fonctions de masse à support simple et à vérifier si l'inégalité 4.35 est respectée. Les taux de consistance correspondent alors au nombre de fois que la propriété est vérifiée par rapport au nombre total des itérations. Pour cette expérience, $1e4$ itérations sont utilisées. La figure 4.3 montre le même résultat pour la classe des règles α-disjonctives.
Comme attendu, nous pouvons voir sur la figure 4.2 les distances d_1, d_∞ et d_{op1} sont totalement consistantes avec la règle $\alpha-$conjonctive tandis que les distances classiques présentent des taux de consistance inférieurs à 10%. Dans la figure 4.3, nous remarquons aussi que les distances d_1, d_∞ et d_{op1} sont totalement consistantes avec la règle $\alpha-$disjonctive. Les distances classiques, quant à elle, présentent un taux de consistance croissant en fonction de α. Ce taux atteint 100% quant $\alpha = 1$.
Le résultat obtenu pour les distances matricielles d_1, d_∞ et d_{op1} était attendu en vu des résultats démontrés dans la section 4.3.1.4.

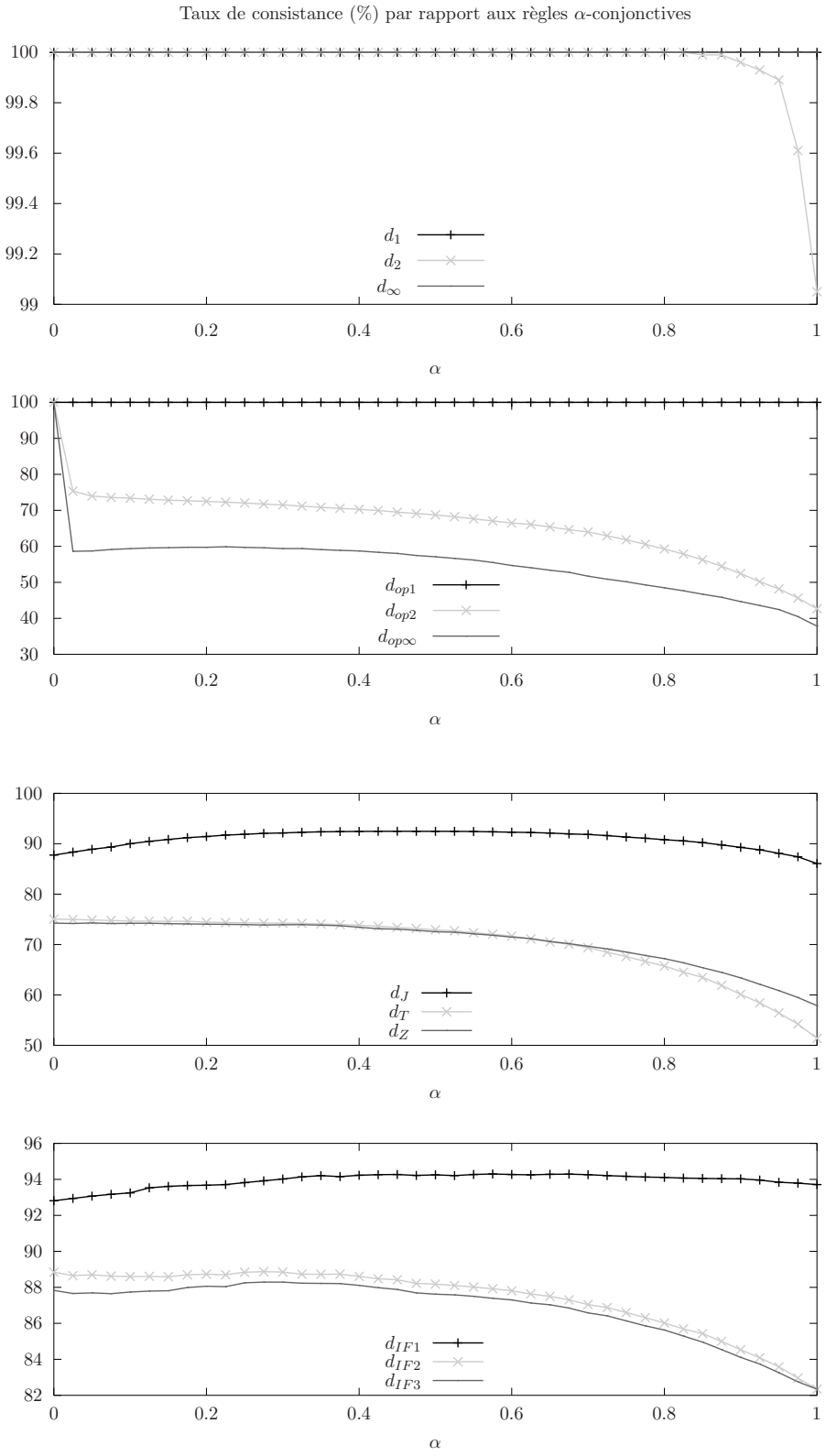

FIGURE 4.2 – Taux de consistance des distances entre fonctions de croyance avec les règles α-conjonctives en fonction du paramètre α.

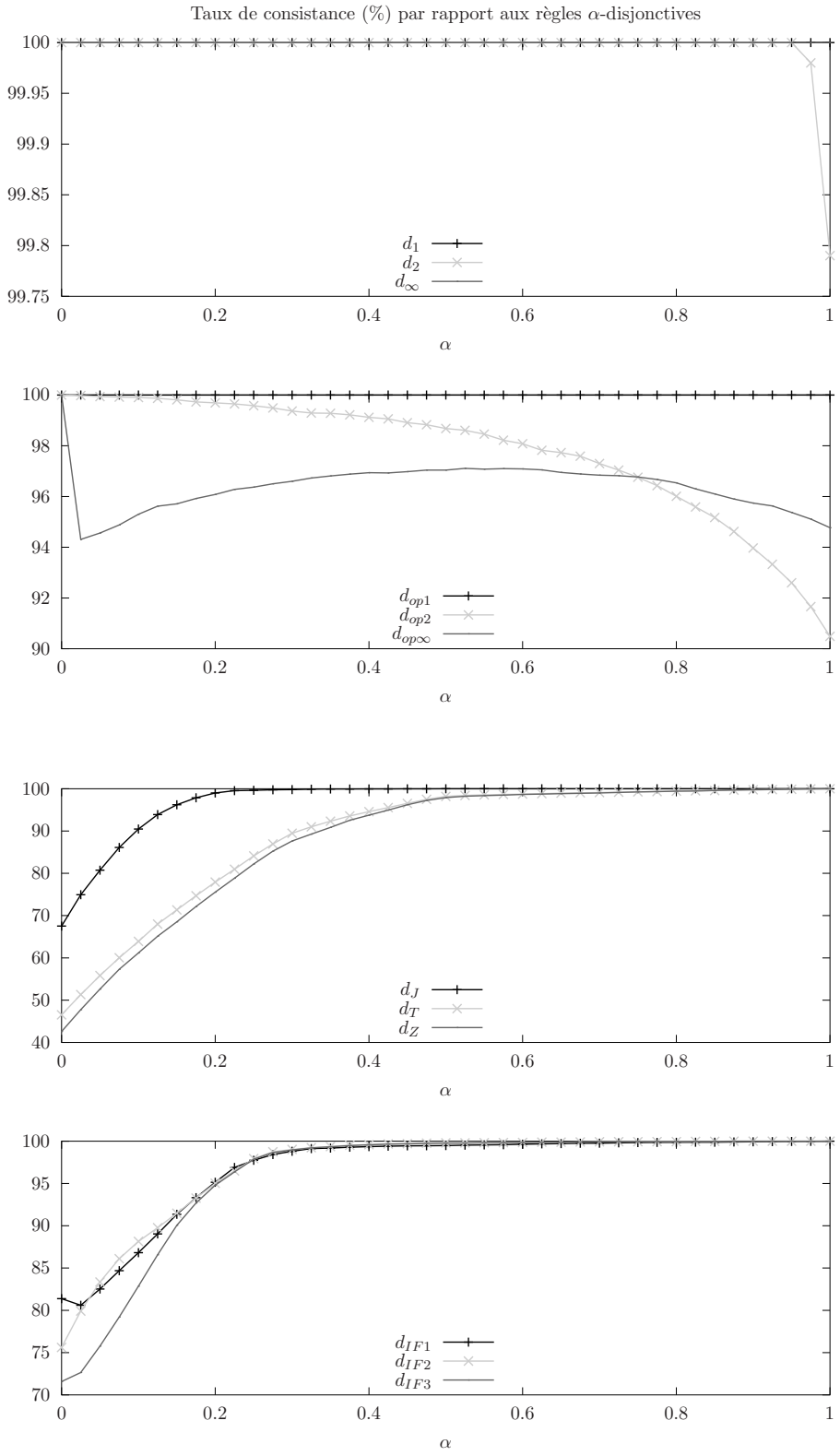

FIGURE 4.3 – Taux de consistance des distances entre fonctions de croyance avec les règles α-disjonctives en fonction du paramètre α.

4.4 Influence des méta-informations sur les distances entre fonctions de croyance

Les $\alpha-$jonctions correspondent à une forme particulière de connaissance sur la véracité des sources d'informations. Dans les applications où ce type de connaissance est disponible, les $\alpha-$jonctions peuvent alors trouver leur utilité. Étant donné $\omega = \omega_0$, il est montré dans [42] que le paramètre α impliqué dans le cas d'une $\alpha-$conjonction peut être interprété en terme de la plausibilité que la source d'informations mente puisque cette plausibilité est égale à $1 - \alpha$. Dans le cas $\alpha-$disjonctif, le paramètre α est égal à la plausibilité que la source dise la vérité. Plus de détails sur l'interprétation des α-jonctions et leur lien avec les méta-informations sont disponibles dans [42, 40]. Nous donnons quelques uns de ces éléments d'interprétation dans ce qui suit.

4.4.1 Méta-informations et α-junctions

L'interprétation des α-jonctions est étroitement liée aux connaissances disponibles sur la **véracité** des sources d'informations mises en jeu avant le processus de fusion. Ce type de connaissance est appelé **méta-information**. La véracité des sources est en effet un type particulier de méta-information. Dans un cadre pratique, la véracité peut être vue ou interprétée de plusieurs façons différentes mais, en tenant compte du contexte des $\alpha-$jonctions, nous adoptons la définition suivante :

*Une source d'informations $\mathcal{S}_i$ est dite **mensongère** si elle soutient le contraire de ce qu'elle considère vrai.*

Supposons que S_1 soutienne $\{\omega_0 \in A\}$ et S_2 soutienne $\{\omega_0 \in B\}$. En fonction de la définition de la véracité, différentes stratégies de combinaison sont intuitivement nécessaires selon différentes méta-informations :

- si les deux sources S_1 et S_2 sont véridiques, alors l'hypothèse à considérer est $A \cap B$,
- si S_1 est véridique tandis que S_2 ne l'est pas, alors l'hypothèse à considérer est $A \cap \overline{B}$,
- si S_2 est véridique tandis que S_1 ne l'est pas, alors l'hypothèse à considérer est $\overline{A} \cap B$,
- si les deux sources S_1 et S_2 sont mensongères, alors l'hypothèse à considérer est $\overline{A} \cap \overline{B}$.

Dans chacune des situations précédentes, une décision différente est retenue. Nous constatons en effet l'importance de prendre en considération les méta-informations dans un problème de fusion d'informations. En général, notre connaissance de la véracité des sources est imprécise et incertaine. Cette connaissance est alors exprimée et modélisée

par une fonction de masse dans un autre domaine noté $\mathcal{T}_i$. Dans [41], Pichon explique qu'un élément $t_A^i \in \mathcal{T}_i$ est compris comme le fait que les sources sont toutes véridiques quand elles s'engagent sur $\{\omega_0 \in A\}$ et elles sont toutes mensongères quand elles s'engagent sur l'hypothèse $\{\omega_0 \in \overline{A}\}$.
Quand on considère deux sources $(S_1; S_2)$, deux méta-évènements appartenant à $\mathcal{T}_1 \times \mathcal{T}_2$, Pichon prouve en outre que :

- pour les α-conjonctions, la méta-information mise en jeu est que chaque événement $\{$*soit les deux sources sont totalement véridiques ou commettent le même mensonge* $\overline{A}\} = \{(t_\Omega^1; t_\Omega^2); (t_A^1; t_A^2)\}$ a la probabilité $\alpha^{|A|}\overline{\alpha}^{|\overline{A}|}$.
- pour les α-disjonctions, la méta-information mise en jeu est que chaque événement $\{$*une source est complètement véridique tandis que l'autre ment au moins à propos de* $A\} = \bigcup_{X \subseteq \overline{C}} \{(t_\Omega^1; t_X^2); (t_X^1; t_\Omega^2)\}$ possède la probabilité $\alpha^{|\overline{A}|}\overline{\alpha}^{|A|}$.

En particulier, quand $\alpha = 1$, $A = \Omega$ est le seul choix donnant une mesure de probabilité non-nulle dans le cas conjonctif tandis que $A = \emptyset$ dans le cas disjonctif. Étant donné que t_Ω implique l'absence de tout mensonge, la méta-information est alors réduite à :

- pour la règle conjonctive, l'évènement $\{$*toutes les sources sont totalement véridiques*$\}$ a la probabilité 1.
- pour la règle disjonctive, l'évènement $\{$*au moins une des sources est totalement véridique*$\}$ a la probabilité 1.

Notons que les α-jonctions sont comprises comme des cas particuliers de processus de combinaison introduits dans [44] où un cadre général de la combinaison des fonctions de masse sous différentes méta-informations est formalisé.

Dans les sous-sections suivantes, nous mettons en lumière le comportement des distances d'α-spécialisation et d'α-généralisation dans le contexte d'un problème de fusion d'informations dépendant de méta-informations. Dans un premier temps, nous nous intéressons au cas où la méta-information est inconnue puis nous nous focalisons sur la situation où la méta-information est certaine. Nous nous limitons évidemment dans les deux cas aux méta-informations concernant les $\alpha-$jonctions.

4.4.2 Méta-information inconnue et distances entre fonctions de masse

Supposons le cas où l'on doit fusionner des sources d'informations incertaines et imprécises dans le cadre de la TFC. Supposons aussi que l'on a observé qu'un opérateur de combinaison $\alpha-$jonctif a du sens dans le contexte de l'application. Il est assez facile de vérifier en pratique que :

- la combinaison d'une fonction de masse avec une combinaison convexe de deux autres est identique à la combinaison convexe de la première opérande avec les deux autres (linéarité),

- l'ordre de traitement des sources d'informations n'affecte pas le résultat (commutativité et associativité),
- soit m_Ω ou $m_\emptyset$ n'impacte pas le résultat des combinaisons (élément neutre),
- la permutation des éléments de Ω n'impacte pas le résultat de la combinaison (anonymat),
- les hypothèses impossibles avant la combinaison le demeurent même après (préservation du contexte).

A l'exception de la linéarité, toutes les autres hypothèses sont très fréquemment vérifiées dans un processus de fusion d'informations. La linéarité est plus rarement observée mais elle est souvent souhaitable dans de nombreux cas.
Accepter une α-jonction comme règle de combinaison pour le problème en question implique l'existence d'une certaine méta-information sous-jacente mais cette méta-information n'est pas souvent disponible. En supposant qu'une comparaison de fonctions de masse doit être effectuée, il faut résoudre le problème de calcul de la distance entre fonctions de masse sous la méta-information inconnue sur l'état des sources d'informations. Une des réponses possibles à ce problème est de calculer toutes les distances d'α-spécialisation et d'α-généralisation et de retenir leur valeurs extrêmes. On obtient alors un intervalle comme valeur de la distance sachant que la valeur réelle de la distance appartient à cet intervalle. Cette imprécision sur la distance reflète notre méconnaissance de la méta-information. Une telle approche imprécise basée sur un intervalle n'est pas rare dans le cadre de la TFC. Destercke et Burger [7], par exemple, ont introduit une mesure du conflit entre fonctions de croyance variant dans un intervalle. Pour une meilleure évaluation de cette solution, considérons l'exemple suivant :

Exemple 9. *Supposons que Ω est un cadre de discernement et X un sous-ensemble de Ω. Soit m une fonction de masse dans Ω sachant que :*

$$\mathbf{m} = 0.3\mathbf{m}_X + 0.5\mathbf{m}_{\overline{X}} + 0.2\mathbf{m}_\Omega.$$

La figure 4.4 montre les distances d'α-spécialisation et d'α-généralisation basées sur la norme L^1 entre m et m_X quant $|\Omega| = 3$ et $|X| = 2$.

TABLE 4.1 – Distances entre m et m_X - Exemple 9

Distances	$d(m, m_X)$
d_J	0.57
d_T	0.57
d_Z	0.49
d_{IF1}	0.57
d_{IF2}	0.40
d_{IF3}	0.40

Distances	$d(m, m_X)$
d_1	$[0.59; 0.70]$
d_2	$[0.56; 0.62]$
d_∞	0.70
d_{op1}	0.70
d_{op2}	$[0.47; 0.70]$
$d_{op\infty}$	$[0.31; 0.70]$

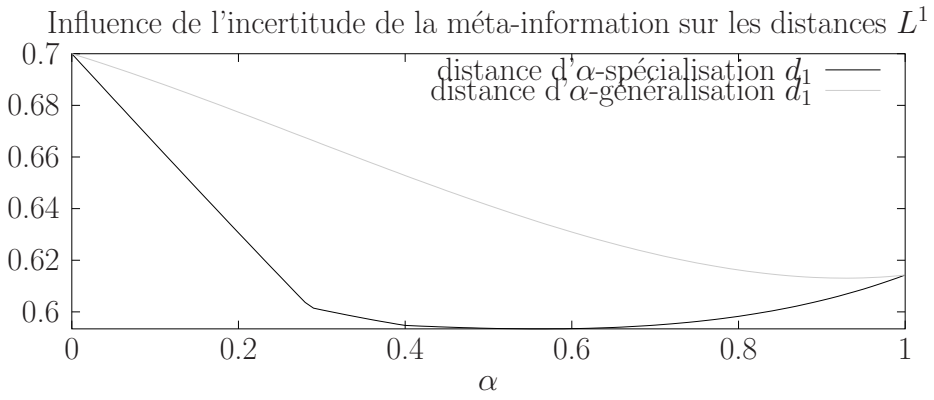

FIGURE 4.4 – Differentes distances d'α-spécialisation et d'α-généralisation basées sur la norme L^1 calculées entre deux fonctions de masse sachant que $m_1 = m_X$ et $m_2 = 0.3m_X + 0.5m_{\overline{X}} + 0.2m_\Omega$, avec $|\Omega| = 3$ et $|X| = 2$.

Comme imposé par le lemme 3, les deux courbes coïncident aux valeurs extrêmes de α. Par contre, les valeurs minimale et maximale des distances ne sont pas nécessairement atteintes en ces points. En outre, ces distances sont en général non convexes en fonction de α. Néanmoins, il est à remarquer qu'elles possèdent un comportement régulier et un minimum unique (dans cet exemple).
Dans l'exemple 9, le résultat de l'expérience donne : $d_1(m, m_X) \in [0.59; 0.7]$. Une question évidente se pose alors : Comment peut-on exploiter une distance donnée sous forme d'intervalle ? Cette question est fréquente dans le cadre des probabilités imprécises. Comme souvent, la distance minimale et la distance maximale peuvent être vues respectivement comme la plus risquée et la plus prudente des valeurs à considérer. Des solutions intermédiaires sont aussi possibles mais la détermination de l'utilité d'une distance définie dans un intervalle dépend étroitement de l'application dédiée.

D'autres estimations de la distance entre m et m_X sont regroupées dans le tableau 4.1. Deux remarques en découlent :

- concernant les distances basées sur les matrices de croyance, les plus précises (valeur unique) sont données par d_∞ et d_{op1}. Ce résultat ne devrait pas être extrapolé pour d_∞ parce que le calcul de cette distance avec d'autres masses peut donner des intervalles complètement différents. En ce qui concerne d_{op1}, cette distance ne peut guère être considérée comme dépendante de méta-informations parce que celle-ci est égale à la distance basée sur la norme L^1 entre les vecteurs de masse.
- bien que les distances classiques ne sont pas destinées à être utilisées dans un tel contexte, puisqu'elles ne sont pas développées pour cet objectif, les valeurs qu'elles fournissent sont incluses dans au moins un des intervalles produits par les distances basées sur des matrices.

4.4.3 Même état des sources et distances entre fonctions de masse

Examinons à présent la situation complémentaire où la méta-information est certaine et précise. C'est à dire les évènements $\{$*la source* S_1 *est dans l'état* $t_A\}$ et $\{$*la source* S_2 *est dans l'état* $t_B\}$ sont sûrs. Dans ces conditions, un mécanisme de correction peut être utilisé sur chaque fonction de masse afin de les rendre véridiques. En effet, supposons qu'une de ces sources, notée S_1, fournit une information qui supporte $\{\omega_0 \in Y\}$. Puisque S_1 est dans un état t_A de méta-information, le témoignage sera partitionné en deux parties : $\{\omega_0 \in Y \cap A\}$ et $\left\{\omega_0 \in Y \cap \overline{A}\right\}$. La première partie reste inchangée tandis qu'un mensonge effectué sur l'autre partie modifiant le témoignage en son complémentaire dans $\overline{A}$: $\left\{\omega_0 \in \overline{Y} \cap \overline{A}\right\}$. En conséquence, S_1 délivre une fonction de masse m_1 dont l'élément focal est $\overline{Y \Delta A} = (A \cap Y) \cup \left(\overline{Y} \cap \overline{A}\right)$. Si $m_{1|t_A}$ représente la fonction de masse que S_1 aurait du fournir, nous déduisons donc :

$$m_{1|t_A}(Y) \quad = \quad \sum_{\substack{X \subseteq \Omega \\ Y = \overline{A \Delta X}}} m_1(X) = m_1\left(\overline{A \Delta Y}\right), \tag{4.41}$$

parce que $X = \overline{A \Delta Y} \Leftrightarrow Y = \overline{A \Delta X}$.

Ce mécanisme de correction appartient à la famille des corrections basées sur le comportement introduites par Pichon *et al.* [44].

En utilisant ce mécanisme, on peut calculer la distance entre les fonctions de masse corrigées sans se préoccuper de la méta-information. Bien que ce mécanisme soit simple et facile à implémenter, il est toutefois peu probable que ces méta-informations soient disponibles en pratique.

Une situation plus réaliste est celle où toutes les sources mentent de la même façon sans connaitre la forme du mensonge. Par exemple, plusieurs systèmes réels possèdent des capteurs redondants qui pourraient mal fonctionner de la même façon quand ils sont soumis aux mêmes conditions expérimentales. Dans ce cas particulier, une comparaison

pertinente entre fonctions de masse est assurée par certaines distances de spécialisation malgré l'imprécision des méta-informations. En effet, nous avons le résultat suivant :

Proposition 20. *Soient m_1 et m_2 deux fonctions de masse définies dans Ω issues de deux sources d'informations S_1 et S_2, toutes deux dans un état t_A. Si d est une distance de spécialisation dépendant de la norme L^k ou de la norme d'opérateur 1, nous avons alors :*

$$d\left(m_1, m_2\right) = d\left(m_{1|t_A}, m_{2|t_A}\right), \tag{4.42}$$

avec $m_{1|t_A}$ et $m_{2|t_A}$ les fonctions de masse que devraient fournir les sources d'informations S_1 et S_2 si elles étaient totalement véridiques.

Démonstration. A partir de l'équation (4.41), une fonction de masse corrigée $m_{i|t_A}$ est égale à $m_i \circ \sigma$ avec σ une permutation telle que :

$$\begin{aligned} \sigma : 2^\Omega &\rightarrow 2^\Omega, \\ X &\rightarrow \overline{A \Delta X}. \end{aligned}$$

Ceci est suffisant pour conclure concernant la distance basée sur la norme d'opérateur 1 car cette norme est équivalente à une norme vectorielle dont le calcul est invariant aux permutations des éléments des fonctions de masse.
Pour les normes L^k, nous devons d'abord remarquer que $\forall A, B, X \subseteq \Omega$:

$$X \cap B = A \cap B \Leftrightarrow \sigma\left(X\right) \cap B = \sigma\left(A\right) \cap B.$$

Il s'en suit que

$$m_i[B]\left(A \cap B\right) = m_{i|t_A}[B]\left(\sigma\left(A\right) \cap B\right),$$

avec $m_{i|t_A}[B] = m_{i|t_A} \textcircled{\cap} m_B$. Notons par σ_B l'application définie par : $\sigma_B\left(A\right) = \sigma\left(A\right) \cap B$. On peut prouver que σ_B est une permutation dans 2^B.
Soient $\mathbf{S}_i$ et $\mathbf{S}_{i|t_A}$ les matrices de spécialisation correspondant aux masses m_i et $m_{i|t_A}$ respectivement. Rappelons que $\mathbf{m}_i[B]$ et $\mathbf{m}_{i|t_A}[B]$ sont les j^{ieme} vecteurs colonnes de $\mathbf{S}_i$ et $\mathbf{S}_{i|t_A}$, avec j l'entier correspondant à B au sens de l'ordre binaire. Il apparaît que chaque colonne de $\mathbf{S}_{i|t_A}$ est obtenue par la permutation des éléments non-nuls de la même colonne de la matrice $\mathbf{S}_i$.
Puisque le calcul de la norme L^k est invariant aux permutations des composantes des matrices comparées, nous avons :

$$\begin{aligned} \left\|\mathbf{S}_{1|t_A} - \mathbf{S}_{2|t_A}\right\|_k &= \left\|\mathbf{S}_1 - \mathbf{S}_2\right\|_k, \\ \Leftrightarrow d_k\left(m_1, m_2\right) &= d_k\left(m_{1|t_A}, m_{2|t_A}\right). \end{aligned}$$

□

La proposition 20 est intéressante dans le sens où, quand les sources sont corrompues de la même façon, la même valeur de distance est obtenue que si elles fournissaient les vraies informations. Cette propriété est illustrée dans l'exemple suivant :

Exemple 10. *Les mêmes fonctions de masse que dans l'exemple 9 sont utilisées :* $\mathbf{m}_1 = \mathbf{m}_X$ *et* $\mathbf{m}_2 = 0.3\mathbf{m}_X + 0.5\mathbf{m}_{\overline{X}} + 0.2\mathbf{m}_\Omega$. *Par contre, nous choisissons dans ce cas* $|\Omega| = 7$ *et* $|X| = 3$. *Supposons que* A_i *est un sous-ensemble variant par inclusions successives de* $\emptyset$ *à* Ω *et* $A_3 = X$.

La figure 4.5 montre les distances entre $m_{1|t_{A_i}}$ *et* $m_{2|t_{A_i}}$.

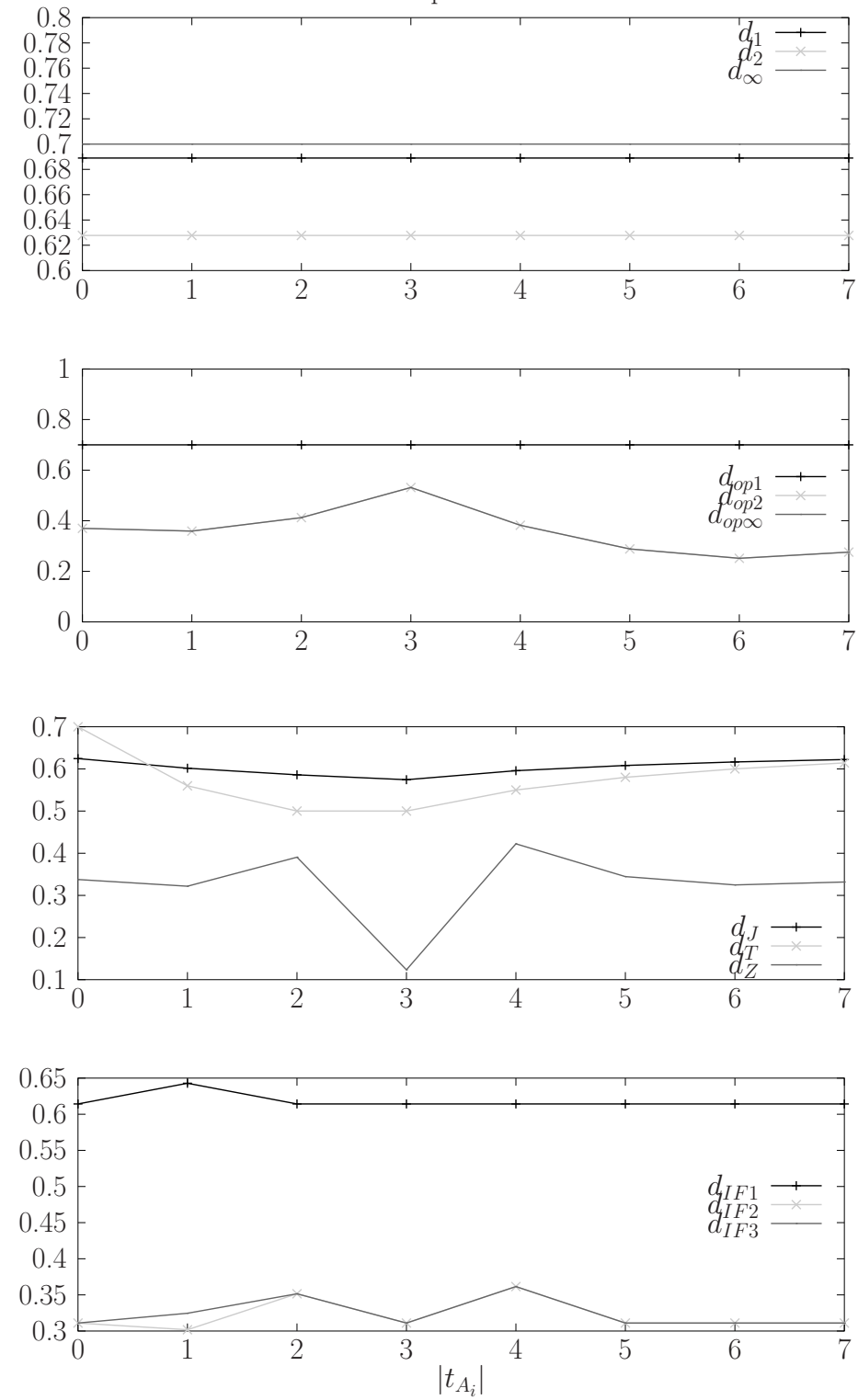

FIGURE 4.5 – Differentes distances entre des sources de même état de véracité calculées entre deux fonctions $m_{1|t_{A_i}}$ et $m_{2|t_{A_i}}$ sachant que $m_1 = m_X$ et $m_2 = 0.3m_X + 0.5m_{\overline{X}} + 0.2m_\Omega$, avec $|\Omega| = 7$ et $|X| = 3$. L'ensemble A_i varie par inclusions succéssives de $\emptyset$ jusqu'à Ω et $A_3 = X$.

Des écarts importants sont observés pour les distances classiques comme pour d_{op2} et $d_{op\infty}$ montrant que leurs valeurs sont peu pertinentes si la méta-connaissance ne peut pas être filtrée.

4.5 Conclusion

Dans la TFC, les $\alpha-$jonctions sont des familles d'opérateurs de combinaison linéaires, commutatifs et associatifs dont l'agrégation des connaissances est assurée grâce à des matrices stochastiques dites **matrices de croyance**. Elles généralisent des règles de combinaison familières telles que les règles de combinaison conjonctive et disjonctive.

La comparaison de deux fonctions de masse peut être effectuée par le calcul de la norme de la différence entre leurs matrices de croyances correspondantes. Cette comparaison est justifiée par le fait qu'il existe une correspondance bijective entre chaque fonction de masse et sa matrice de croyance induite. En examinant les différentes normes matricielles, des familles infinies de distances basées sur les matrices de croyance sont définies dans ce chapitre. Ces familles de distances généralisent la famille des distances de spécialisation proposée dans le chapitre 3.

Nous avons montré dans ce travail que certaines de ces familles de distances sont consistantes avec les $\alpha-$jonctions dans le sens où deux fonctions de masse sont plus proches entre elles après avoir été conjointement enrichies d'un apport d'informations identique via une $\alpha-$jonction. Ce résultat s'explique par le fait que les matrices de croyance utilisées dans les distances que nous proposons dans ce travail sont elles-mêmes impliquées dans la définition des $\alpha-$jonctions.

Étant donné que les fonctions de masse sont des fonctions d'ensembles et que leur domaine possède une structure de treillis, une distance entre fonctions de croyance devrait être capable de prendre en compte cette structure. A cet effet, nous avons aussi pu voir que la sous-famille des distances basées sur la norme L^1 généralise une distance d'ensemble si le paramètre α est nul ou appartient à l'intervalle $[0.5, 1]$. Notons que la généralisation d'une distance d'ensemble est une propriété contraignante. Certaines distances entre fonctions de croyance ne généralisent pas une distance d'ensemble mais tiennent compte de la structure des fonctions de masse d'une manière satisfaisante.

En outre, l'influence de la méta-information concernant la véracité des sources est étudiée dans ce chapitre. Nous avons pu conclure que les distances matricielles définies fournissent un intervalle de valeurs dans le cas où la méta-information est incertaine. Quand les sources d'informations sont corrompues de la même façon, les distances basées

sur les normes matricielles L^k où la norme d'opérateur 1 sont invariantes aux mensonges que peuvent générer les sources d'informations à condition de se restreindre à une distance de spécialisation ($\alpha = 1$).

Conclusion générale

Dans cette thèse, une nouvelle famille de distances entre fonctions de croyance a été proposée. Cette nouvelle famille est basée sur les matrices de croyances qui interviennent dans la représentation matricielle de la fusion d'informations. Étant donné que chaque fonction de masse est représentée de façon unique par une matrice de croyance relativement à une règle de combinaison donnée, la distance entre les fonctions de masse se calcule entre leurs matrices de croyance respectives.

En plus des propriétés métriques qu'offre cette nouvelle famille de distances basée sur les normes matricielles, deux autres propriétés principales sont formalisées dans cette thèse :

- la propriété structurelle. Celle-ci est respectée par une distance entre fonctions de croyance si elle généralise une distance d'ensembles,
- la propriété de consistance avec une règle de combinaison. La propriété de consistance avec une règle de combinaison implique, si elle est respectée, que deux fonctions de masse sont plus proches l'une par rapport à l'autre après avoir été enrichies d'un apport informationnel supplémentaire via l'application de la règle en question.

Dans un premier temps, notamment dans le chapitre 3, nous avons tout d'abord développé la nouvelle famille de distances basée sur les matrices de spécialisations dempsteriennes. Cette famille de distances respecte les propriétés métriques. En plus, la propriété structurelle est respectée par les distances de spécialisation du type L^k avec $k < \infty$ puisqu'elles généralisent la distance d'ensemble de *Hamming*. Par ailleurs, la propriété de consistance conjonctive est respectée par trois distances de spécialisation ; à savoir les distances basées sur les normes L^1, L^∞ et sur la norme d'opérateur-1. Cette dernière propriété est sans doute la propriété la plus rare dans la TFC puisqu'aucune autre distance définie dans la littérature ne la satisfait. Nous retenons que, parmi les distances étudiées, la distance de spécialisation du type L^1 est la seule distance qui respecte à la fois les propriétés métriques, structurelles ainsi que de consistance conjonctive.

Dans ce même chapitre, plusieurs méthodes pour le calcul rapide des distances de spécialisation du type L^k sont aussi introduites. Initialement, ces distances sont calculées avec une complexité en $O(N^3)$ avec $N = 2^{|\Omega|}$. Nous avons élaboré un algorithme pour réduire cette complexité à $O(N^{1.58})$ et gagner en temps de calcul. Dans le cas de fonctions de masse catégoriques, cette complexité est simplement de l'ordre de $O(1)$. L'utilisation de ces approches permet aux distances L^k d'être très compétitives en terme de temps de calcul en plus de leurs propriétés mathématiques très attractives.

Une généralisation de la famille des distances basées sur les matrices dempsteriennes est développée dans le chapitre 4. Cette généralisation s'appuie sur les matrices de

croyance engendrées par les règles de combinaison $\alpha-$jonctives. Elle est motivée par le principe de consistance avec les opérateurs de combinaison $\alpha-$jonctifs. En effet,les $\alpha-$jonctions sont des règles de combinaison qui généralisent la règle conjonctive ainsi que la règle disjonctive. Cette généralisation permet de préserver la propriété de consistance de chacune des distances d_{op1}, d_1 et d_∞ avec la règle $\alpha-$jonctive qui lui correspond. La sous-famille des distances du type L^1 généralise une distance d'ensemble uniquement si le paramètre α est nul ou appartient à l'intervalle $[0.5, 1]$. Il est à remarquer que certaines distances entre fonctions de croyance ne généralisent pas une distance d'ensemble mais tiennent compte tout de même de la structure des fonctions de masse d'une manière satisfaisante. C'est notamment le cas de la distance de type L^1 quand α appartient à $]0, 0.5[$.

Étant donné que l'emploi d'une $\alpha-$jonction peut se justifier par une méta-information concernant la véracité des différentes sources d'informations, l'influence de cette méta-information dans le calcul de distances entre fonctions de masse est étudiée dans cette thèse. Dans le cas où la méta-information est incertaine, nous avons proposé que les distances basées sur les matrices de croyance et généralisées aux $\alpha-$jonctions soient données sous forme d'intervalles. Il est aussi à retenir que si l'état des méta-informations des sources d'informations est le même, alors les distances de spécialisation du type L^1 et d'opérateur 1 sont invariantes aux éventuels mensonges que peuvent générer les sources d'informations. Ce résultat reste important si nous savons a priori que les sources d'informations mentent de la même façon. Quelle que soit la teneur du mensonge, la valeur proposée par la distance sera donc pertinente.
Si parmi toutes les distances proposées dans ce mémoire, une était à retenir, ce serait sans nul doute la distance de spécialisation du type L^1 puisqu'elle respecte le plus de propriétés que toutes ses concurrentes. Cette distance est :

- une métrique complète,
- structurée,
- consistante avec toute $\alpha-$jonction,
- invariante à un mensonge identique des sources d'information,
- calculable en un temps de calcul comparable aux approches classiques.

En perspectives, nous visons à approfondir l'étude sur l'impact des méta-informations sur les distances matricielles entre les fonctions de croyances. Nous réservons un intérêt particulier au cas de véracité et de mensonge identique qui est souvent le cas dans un système de capteurs évoluant dans un même environnement. Nous souhaitons aussi réaliser des applications par la suite sur un système de type VANETS à plusieurs véhicules échangeant des données dans un réseau ad-hoc [12]. Dans ce système, tous les véhicules sont équipés des mêmes capteurs qui sont susceptibles de défaillir de manière identique. Nous estimons en effet, qu'il peut être intéressant de comparer les éléments de preuves

fournis par chacun des véhicules pour identifier un comportement suspect ou singulier. Nous visons aussi à appliquer les nouvelles familles de distances dans des processus d'approximation de fonctions de masse par la minimisation d'une distance. Dans ce champs nous envisageons d'étudier les approximations de fonctions de masse quelconques par des fonctions de masse appartenant à une sous famille ayant des propriétés particulières telles que les fonctions les plus engagées qu'une certaine fonction par exemple, les fonctions consonantes ou les fonctions bayésiennes.

Annexes

En annexes de ce document sont regroupées diverses démonstrations de propositions et lemmes évoqués dans les chapitres 3 et 4.

A.1 Démonstration de la formule du coefficient de normalisation pour les distances de spécialisation L^k

Soient deux fonctions de masse m_1 et m_2, dont les matrices de spécialisation dempsteriennes respectives sont $\mathbf{S_1}$ et $\mathbf{S_2}$. Si $k < \infty$, la distance L^k entre les matrices de spécialisation s'écrit comme :

$$\|\mathbf{S_1} - \mathbf{S_2}\|_k = \left(\sum_{i=1}^{2^{|\Omega|}} \sum_{j=1}^{2^{|\Omega|}} |S_{1,ij} - S_{2,ij}|^k \right)^{\frac{1}{k}}.$$

La première colonne est toujours la même quelle que soit la matrice de spécialisation. Par conséquent, on obtient :

$$\begin{aligned} \|\mathbf{S_1} - \mathbf{S_2}\|_k &= \left(\sum_{i=1}^{2^{|\Omega|}} \sum_{j=2}^{2^{|\Omega|}} |S_{1,ij} - S_{2,ij}|^k \right)^{\frac{1}{k}}, \\ &\leq \left(\sum_{i=1}^{2^{|\Omega|}} \sum_{j=2}^{2^{|\Omega|}} |S_{1,ij}|^k + |S_{2,ij}|^k \right)^{\frac{1}{k}}. \end{aligned}$$

Étant donné que tous les éléments d'une matrice de spécialisation sont positifs et inférieurs ou égaux à 1, alors :

$$\|\mathbf{S_1} - \mathbf{S_2}\|_k \leq \left(\sum_{i=1}^{2^{|\Omega|}} \sum_{j=2}^{2^{|\Omega|}} S_{1,ij} + S_{2,ij} \right)^{\frac{1}{k}}$$

Sachant que la somme de chaque colonne d'une matrice de spécialisation est égale à 1, on obtient :

$$\begin{aligned} \|\mathbf{S_1} - \mathbf{S_2}\|_k &\leq \left(\sum_{j=2}^{2^{|\Omega|}} (1+1) \right)^{\frac{1}{k}} \\ &\leq \left(2\left(2^{|\Omega|} - 1\right) \right)^{\frac{1}{k}}. \end{aligned}$$

Finalement, procédons au calcul de la distance non normalisée entre le fonction d'ignorance totale m_Ω e t la fonction totalement conflictuelle $m_\emptyset$:

$$\|\mathbf{S_\Omega} - \mathbf{S_\emptyset}\|_k = \left(\sum_{i=1}^{2^{|\Omega|}} \sum_{j=1}^{2^{|\Omega|}} |\mathbf{I}_{ij} - \mathbf{Z}_{ij}|^k \right)^{\frac{1}{k}},$$

avec $\mathbf{Z}$ la matrice de spécialisation de $m_\emptyset$. cette matrice est définie comme suit $\mathbf{Z}_{ij} = 1$ si $i = 1$ et $\mathbf{Z}_{ij} = 0$ ailleurs. Ce qui donne : $\|\mathbf{S_\Omega} - \mathbf{S_\emptyset}\|_k = \left(2\left(2^{|\Omega|} - 1\right) \right)^{\frac{1}{k}}$. En conclusion, nous avons :

$$\max_{m_i, m_j} \|\mathbf{S_i} - \mathbf{S_j}\|_k = \left(2\left(2^n - 1\right)\right)^{\frac{1}{k}}. \tag{43}$$

En outre, si $k = \infty$, la distance L^∞ entre les matrices de spécialisation s'écrit :

$$\|\mathbf{S_1} - \mathbf{S_2}\| \infty = \max_{1 \leq i,j \leq 2^{|\Omega|}} |S_{1,ij} - S_{2,ij}|.$$

Il est donc clair que $\|\mathbf{S_1} - \mathbf{S_2}\| \infty \leq 1$. En plus, nous avons $\|\mathbf{S_\Omega} - \mathbf{S_\emptyset}\| \infty = 1$. Par conséquent, le résultat le résultat donné en équation 43 est vérifié même quand $k = \infty$.

A.2 Preuve que les distances de spécialisation L^k sont structurées (proposition 3)

Soit un cadre de discernement Ω. La distance d'ensemble de *Hamming* d_{ham} entre les sous-ensembles A_1 et A_2 est définie comme suit :

$$d_{ham}(A_1, A_2) = |A_1 \Delta A_2|, \tag{44}$$

avec Δ la différence symétrique. Dans cette démonstration, nous montrons que toute distance de spécialisation L^k généralise d_{ham}.

La distance de spécialisation L^k entre deux fonctions de masse catégoriques m_{A_1} et m_{A_2} définies dans Ω est donnée par :

$$\|\mathbf{S}_{A_1} - \mathbf{S}_{A_2}\|_k = \left(\sum_{A,B \subseteq \Omega} |S_{A_1}(A,B) - S_{A_2}(A,B)|^k \right)^{\frac{1}{k}}, \quad (45)$$

avec $\mathbf{S}_{A_1}$ et $\mathbf{S}_{A_2}$ les matrices de spécialisation dempsteriennes de m_{A_1} et m_{A_2} respectivement. Étant donné l'équation (1.58), alors :

$$\|\mathbf{S}_{A_1} - \mathbf{S}_{A_2}\|_k = \left(\sum_{A,B \subseteq \Omega} |1_{A_1 \cap B = A} - 1_{A_2 \cap B = A}|^k \right)^{\frac{1}{k}}, \quad (46)$$

avec 1_{Cond} la fonction indicatrice définie par :

$$1_{Cond} = \begin{cases} 1 & \text{si } Cond \text{ est vraie,} \\ 0 & \text{sinon.} \end{cases}$$

Les éléments de la matrice de spécialisation se compensent les uns par les autres dans le calcul de la distance à travers la différence absolue si $A_1 \cap B = A = A_2 \cap B$. C'est à dire quand $B \subseteq (A_1 \cap A_2) \cup (\Omega \setminus (A_1 \cup A_2)) = \Omega \setminus (A_1 \Delta A_2)$. On obtient alors :

$$\|\mathbf{S}_{A_1} - \mathbf{S}_{A_2}\|_k = \left(\sum_{\substack{A,B \subseteq \Omega \\ B \not\subseteq \Omega \setminus (A_1 \Delta A_2)}} 1^k_{A_1 \cap B = A} + \sum_{\substack{A,B \subseteq \Omega \\ B \not\subseteq \Omega \setminus (A_1 \Delta A_2)}} 1^k_{A_2 \cap B = A} \right)^{\frac{1}{k}}, \quad (47)$$

$$= \left(\sum_{\substack{A,B \subseteq \Omega \\ B \not\subseteq \Omega \setminus (A_1 \Delta A_2)}} 1_{A_1 \cap B = A} + \sum_{\substack{A,B \subseteq \Omega \\ B \not\subseteq \Omega \setminus (A_1 \Delta A_2)}} 1_{A_2 \cap B = A} \right)^{\frac{1}{k}}. \quad (48)$$

Dans chacune de ces sommes, on choisit B fixe et contenant A car $\forall i, A = B \cap A_i$ et ces intersections sont uniques. Le calcul de ces sommes est alors équivalent au calcul des choix possibles de B sachant que $B \not\subseteq \Omega \setminus (A_1 \Delta A_2)$. Alors :

$$\|\mathbf{S}_{A_1} - \mathbf{S}_{A_2}\|_k = \left(2 \sum_{\substack{A,B \subseteq \Omega \\ B \not\subseteq \Omega \setminus (A_1 \Delta A_2)}} 1 \right)^{\frac{1}{k}}. \quad (49)$$

B est choisi tel qu'il contient au moins un élément de $(A_1 \Delta A_2)$ et n'importe quels autres éléments ou non de $\Omega \setminus (A_1 \Delta A_2)$. Il existe $2^{|A_1 \Delta A_2|} - 1$ façons possibles de choisir un

élément dans $A_1 \Delta A_2$. Il existe $2^{|\Omega \setminus (A_1 \Delta A_2)|}$ façons de choisir un sous ensemble quelconque de $\Omega \setminus (A_1 \Delta A_2)$ incluant l'ensemble vide. Par conséquent, nous avons :

$$\begin{aligned} \|\mathbf{S}_{A_1} - \mathbf{S}_{A_2}\|_k &= \left(2\left(2^{|A_1 \Delta A_2|} - 1\right)\left(2^{|\Omega \setminus (A_1 \Delta A_2)|}\right)\right)^{\frac{1}{k}}, & (50) \\ &= \left(2\left(2^n - 2^{|\Omega \setminus (A_1 \Delta A_2)|}\right)\right)^{\frac{1}{k}}. & (51) \end{aligned}$$

Les distances de spécialisation L^k entre fonctions de masse catégoriques s'écrivent alors :

$$\begin{aligned} d_{Sk}\left(m_{A_1}, m_{A_2}\right) &= \left(\frac{2^n - 2^{|\Omega \setminus (A_1 \Delta A_2)|}}{2^n - 1}\right)^{\frac{1}{k}}, & (52) \\ &= \left(\frac{2^n - 2^{n - d_{ham}(A_1, A_2)}}{2^n - 1}\right)^{\frac{1}{k}}, & (53) \\ &= f \circ d_{ham}(A_1, A_2), & (54) \end{aligned}$$

avec f une fonction monotone croissante. La relation (3.1) est satisfaite.

A.3 Preuve que le distance de spécialisation d'opérateur 1 est égale à la distance de type L^k entre les vecteurs fonctions de masse (Lemme 2)

Soient m_1 et m_2 deux fonctions de masse définies dans Ω. Pour démontrer le Lemme 17, il est d'abord nécessaire de remarquer que la propriété suivante est vérifiée pour la norme vectorielle L^1 entre fonctions de masse :

$$\|(\mathbf{m}_1 - \mathbf{m}_2) \textcircled{\cap} \mathbf{m}_A\|_1 \leq \|\mathbf{m}_1 - \mathbf{m}_2\|_1, \quad (55)$$

avec m_A la fonction de masse catégorique engagée sur le sous-ensemble A. Prouvons d'abord ce résultat intermédiaire. Selon la définition de la norme vectorielle L^1, on écrit :

$$\|(\mathbf{m}_1 - \mathbf{m}_2) \textcircled{\cap} \mathbf{m}_A\|_1 = \sum_{B \subseteq \Omega} |m_1 \textcircled{\cap} m_A(B) - m_2 \textcircled{\cap} m_A(B)|.$$

La combinaison avec des fonctions de masse catégoriques m_A se résume à un conditionnement par rapport au sous-ensemble A. Selon la définition du conditionnement, nous obtenons :

$$\|(\mathbf{m}_1 - \mathbf{m}_2) \textcircled{\cap} \mathbf{m}_A\|_1 = \sum_{B \subseteq \Omega} | \sum_{\substack{C \subseteq \Omega \\ C \cap A = B}} m_1(C) - m_2(C) \ |.$$

Le conditionnement par rapport à A implique que les masses non inclues dans A deviennent nulles. Alors, on écrit :

$$\|(\mathbf{m}_1 - \mathbf{m}_2) \textcircled{\cap} \mathbf{m}_A\|_1 = \sum_{B \subseteq A} | \sum_{\substack{C \subseteq \Omega \\ C \cap \bar{A} = B}} m_1(C) - m_2(C) \, |.$$

En outre, le conditionnement implique aussi que C est un sur-ensemble de B et peut être écrit comme $C = B \cup D$ avec $D \subseteq \Omega \setminus A$. Ceci permet d'obtenir :

$$\|(\mathbf{m}_1 - \mathbf{m}_2) \textcircled{\cap} \mathbf{m}_A\|_1 = \sum_{B \subseteq A} | \sum_{D \subseteq \Omega \setminus A} m_1(B \cup D) - m_2(B \cup D) \, |.$$

En utilisant l'inigalité triangulaire de la norme vectorielle L^1, on obtient :

$$\|(\mathbf{m}_1 - \mathbf{m}_2) \textcircled{\cap} \mathbf{m}_A\|_1 \leq \sum_{B \subseteq A} \sum_{D \subseteq \Omega \setminus A} |m_1(B \cup D) - m_2(B \cup D)|.$$

Toutes les unions d'ensembles B et D sont incluses dans Ω. Puisqu'elles sont choisies dans des ensembles complémentaires A et $\Omega \setminus A$, leur union est unique. Par conséquent, l'ensemble des unions $B \cup D$ est inclus dans 2^Ω. alors :

$$\begin{aligned} \|(\mathbf{m}_1 - \mathbf{m}_2) \textcircled{\cap} \mathbf{m}_A\|_1 &\leq \sum_{E \subseteq \Omega} |m_1(E) - m_2(E)|, \\ &\leq \|(\mathbf{m}_1 - \mathbf{m}_2)\|_1 . \end{aligned}$$

L'étape suivante pour démontrer la proposition 17 consiste en une comparaison deux-à-deux des colonnes des matrices de spécialisation. Remarquons que chaque vecteur colonne d'une matrice de spécialisation $\mathbf{S}_1$ est la représentation vectorielle de la fonction de masse $m_1 \textcircled{\cap} m_A$. Par conséquent, la norme d'opérateur 1 s'exprime comme suit :

$$\|\mathbf{S_1} - \mathbf{S_2}\|_{op1} = \max_{A \subseteq \Omega} \|(\mathbf{m}_1 - \mathbf{m}_2) \textcircled{\cap} \mathbf{m}_A\|_1 .$$

Étant donné l'équation (55), le vecteur colonne avec la norme maximale est nécessairement $(m_1 - m_2) \textcircled{\cap} m_\Omega = (m_1 - m_2)$. On a alors :

$$\|\mathbf{S_1} - \mathbf{S_2}\|_{op1} = \|\mathbf{m}_1 - \mathbf{m}_2\|_1 .$$

A.4 Démonstration de l'interpretation de la valeur maximale des distances L^k (proposition 8)

Pouvons que :

$$d_k(m_1, m_2) = 1 \Leftrightarrow m_1 = m_X, m_2 = m_{\overline{X}}. \tag{56}$$

Les deux implications doivent être prouvées séparément :

- Supposons que m_1 et m_2 sont deux distributions de masse définies dans Ω sachant que $d_k(m_1, m_2) = 1$. D'après la définition de la distance d_k, nous pouvons obtenir :

$$\sum_{A,B\subseteq\Omega} |S_1(A,B) - S_2(A,B)|^k = 2(2^n - 1),$$

où n est la cardinalité de Ω. Sachant que pour toutes distributions de masse m_i, on a $S_i(\emptyset,\emptyset) = 1$, l'expression précédente s'écrit comme suit :

$$\sum_{\substack{B\subseteq\Omega \\ B\neq\emptyset}} \sum_{A\subseteq\Omega} |S_1(A,B) - S_2(A,B)|^k = 2(2^n - 1).$$

sachant que : $\sum_{\substack{B\subseteq\Omega \\ B\neq\emptyset}} 1 = (2^n - 1)$. Nous pouvons alors aboutir au résultat :

$$\sum_{\substack{B\subseteq\Omega \\ B\neq\emptyset}} \left(2 - \sum_{A\subseteq\Omega} |S_1(A,B) - S_2(A,B)|^k\right) = 0. \tag{57}$$

Par ailleurs, le développement de l'expression suivante nous permet d'écrire :

$$\sum_{A\subseteq\Omega} |S_1(A,B) - S_2(A,B)|^k \leq \sum_{A\subseteq\Omega} \Big(|S_1(A,B)|^k + |S_2(A,B)|^k\Big).$$

En plus, nous savons que tous les éléments d'une matrice de spécialisation sont positifs et inférieurs ou égaux à 1. Ce constat implique que pour tout A, B dans Ω, nous pouvons avoir $|S_i(A,B)|^k \leq S_i(A,B)$. Ce résultat nous permet d'écrire :

$$\sum_{A\subseteq\Omega} |S_1(A,B) - S_2(A,B)|^k \leq \sum_{A\subseteq\Omega} \Big(S_1(A,B) + S_2(A,B)\Big).$$

En outre, toute matrice de spécialisation S_i est définie de sorte que pour tout

$B \subseteq \Omega$, $\sum_{A\subseteq\Omega} S_i(A,B) = 1$. Ainsi nous obtenons le résultat suivant :

$$\begin{aligned} \sum_{A\subseteq\Omega} |S_1(A,B) - S_2(A,B)|^k &\leq 2, \\ 2 - \sum_{A\subseteq\Omega} |S_1(A,B) - S_2(A,B)|^k &\geq 0 \end{aligned} \tag{58}$$

En utilisant l'équation (58), l'expression (57) devient une somme de termes positifs. Et comme cette somme est égale a zéro, nous pouvons donc constater que tous ces termes sont nuls. Par conséquent, nous pouvons obtenir :

$$\forall B \subseteq \Omega \text{ et } B \neq \emptyset, \sum_{A\subseteq\Omega} |S_1(A,B) - S_2(A,B)|^k = 2. \tag{59}$$

Prouvons à présent que l'équation (59) implique que m_1 et m_2 sont catégoriques :

– Quand $k > 1$:

L'équation (59) peut s'écrire de la façon suivante en particulier quand $B = \Omega$:

$$\begin{aligned} 2 &= \sum_{A\subseteq\Omega} |S_1(A,\Omega) - S_2(A,\Omega)|^k, \\ 2 &\leq \sum_{A\subseteq\Omega} \left(|S_1(A,\Omega)|^k + |S_2(A,\Omega)|^k \right), \end{aligned}$$

$$2 \leq \sum_{A\subseteq\Omega} \left(|m_1(A)|^k + |m_2(A)|^k \right). \tag{60}$$

Posons à présent l'hypothèse suivante : m_1 n'est pas une distribution de masse catégorique. Cette hypothèse implique que $\forall A \subsetneq \Omega$, $m_1(A) < 1$ et par conséquent $\sum_{A\subseteq\Omega} |m_1(A)|^k < 1$ pour $k > 1$. Ainsi nous, obtenons :

$$\sum_{A\subseteq\Omega} \left(|m_1(A)|^k + |m_2(A)|^k \right) < 2, \tag{61}$$

Les deux équations (60) et (61) sont contradictoires. Ceci implique que, pour le cas où $k > 1$, la distribution de masse m_1 est catégorique et engagée sur l'ensemble X. Le même raisonnement est opéré pour la distribution de masse m_2, qui est catégorique et engagée sur l'ensemble Y.

– Quand $k = 1$:

L'équation (59) implique $\forall B \subseteq \Omega$, $\mathcal{F}_{m_1[B]} \cap \mathcal{F}_{m_2[B]} = \emptyset$. C'est à dire que $m_1[B]$ et $m_2[B]$ ont toujours des éléments focaux différents deux à deux. On montre

alors que la seule situation possible est que $|\mathcal{F}_{m_1}| = |\mathcal{F}_{m_2}| = 1$.
Finalement, nous prouvons dans ce qui suit que les m_1 et m_2 sont engagées sur des sous-ensembles complémentaire dans Ω et ce $\forall k \geq 1$.
Posons à présent l'hypothèse supplémentaire suivante : $\exists Z \neq \emptyset$ tel que $Z = \Omega \setminus (X \cup Y)$. Nous pouvons alors appliquer la relation (59) avec $B = Z$. Dans ce cas particulier, nous obtenons :

$$\sum_{A \subseteq \Omega} |S_1(A, Z) - S_2(A, Z)|^k = 2. \tag{62}$$

Selon la définition des matrices de spécialisation, nous pouvons écrire : $S_i(A, Z) = m_i[Z](A)$ avec $m_i[Z]$ la fonction de masse m_i après conditionnement par Z. Or, puisque m_1 et m_2 sont catégoriques et que $X \cap Z = \emptyset$ et $Y \cap Z = \emptyset$, nous avons :

$$m_1[Z] = m_2[Z] = m_\emptyset,$$

la distribution de masse totalement conflictuelle. Ce résultat implique que :

$$\sum_{A \subseteq \Omega} |S_1(A, Z) - S_2(A, Z)|^k = 0,$$

qui est en contradiction avec l'équation(62). Par conséquent, la dernière hypothèse est fausse, impliquant que $X = \Omega \setminus Y$.

- Réciproquement, supposons à présent que m_1 est catégorique engagée sur le sous-ensemble X, m_2 est aussi catégorique engagée sur le sous-ensemble Y et que $Y = \overline{X}$. par définition de la distance L^k, nous avons :

$$d_k(m_1, m_2) = \left(\frac{1}{2(2^n - 1)} \sum_{A,B \subseteq \Omega} |S_1(A, B) - S_2(A, B)|^k \right)^{\frac{1}{k}},$$

où n est la cardinalité de Ω. Cette équation peut se réécrire comme suit :

$$d_k(m_1, m_2)^k = \frac{1}{2(2^n - 1)} \sum_{A,B \subseteq \Omega} |S_1(A, B) - S_2(A, B)|^k,$$

L'application de l'équation (1.66) de la proposition proposition (1) pour m_1 et m_2 simultanément nous donne le résultat suivant :

$$d_k(m_1, m_2)^k = \frac{1}{2(2^n - 1)} \sum_{A,B \subseteq \Omega} |1_{A_1 \cap B = A} - 1_{A_2 \cap B = A}|^k. \tag{63}$$

Or, nous savons que si $B = \emptyset$, le résultat suivant est vérifié :

$$1_{A_1 \cap \emptyset = \emptyset} - 1_{A_2 \cap \emptyset = \emptyset} = 0.$$

La formule (63) devient donc :

$$d_k\left(m_1, m_2\right)^k = \frac{1}{2\left(2^n - 1\right)} \sum_{\substack{A,B \subseteq \Omega \\ B \neq \emptyset}} \left|1_{A_1 \cap B = A} - 1_{A_2 \cap B = A}\right|^k. \tag{64}$$

Puisque $A_1 = \Omega \setminus A_2$, nous constatons alors que :

$$\forall B \subseteq \Omega \setminus \emptyset,\ A_1 \cap B = \emptyset \Rightarrow A_2 \cap B = B. \tag{65}$$

Par conséquent, nous pouvons donc écrire pour tout $B \neq \emptyset$, sous-ensemble de Ω :

$$\begin{aligned} 1_{A_1 \cap B = A} = 1 \quad &\Rightarrow \quad 1_{A_2 \cap B = A} = 0, \\ 1_{A_2 \cap B = A} = 1 \quad &\Rightarrow \quad 1_{A_1 \cap B = A} = 0. \end{aligned}$$

A présent, la formule (64) peut se traduire de la manière suivante :

$$d_k\left(m_1, m_2\right)^k = \frac{1}{2\left(2^n - 1\right)} \left(\sum_{\substack{A,B \subseteq \Omega \\ B \neq \emptyset}} 1_{A_1 \cap B = A} + \sum_{\substack{cA,B \subseteq \Omega \\ B \neq \emptyset}} 1_{A_2 \cap B = A} \right). \tag{66}$$

Le calcul des deux précédentes sommes se résume à un problème de comptage. Pour chacune des deux sommes, il y a $2^n - 1$ choix du sous-ensemble B. Le choix de B implique la définition de A. Par conséquent, nous obtenons :

$$\begin{aligned} d_k\left(m_1, m_2\right)^k &= \frac{1}{2\left(2^n - 1\right)} \left(2^n - 1 + 2^n - 1\right), \\ d_k\left(m_1, m_2\right)^k &= 1. \\ \Rightarrow d_k\left(m_1, m_2\right) &= 1. \end{aligned}$$

A.5 Démonstration de la proposition 10

1. Montrons que $d_k(m_1^\alpha, m_2) \leq (1-\alpha).d_k(m_1, m_2) + \alpha.d_k(m_2, m_\Omega)$:

$$\begin{aligned} d_k(m_1^\alpha, m_2) &= \frac{1}{\rho}\left\|\mathbf{S}_1^\alpha - \mathbf{S}_2\right\|_k \\ &= \frac{1}{\rho}\left\|\Big((1-\alpha)\mathbf{S}_1 + \alpha\mathbf{I}\Big) - \mathbf{S}_2\right\|_k \\ &= \frac{1}{\rho}\left\|(1-\alpha)\Big(\mathbf{S}_1 - \mathbf{S}_2\Big) + (\alpha)\Big(\mathbf{I} - \mathbf{S}_2\Big)\right\|_k . \end{aligned}$$

Le résultat précédent nous permet d'écrire par inégalité triangulaire :

$$d_k(m_1^\alpha, m_2) \leq (1-\alpha)d_k(m_1, m_2) + \alpha d_k(m_2, m_\Omega). \tag{67}$$

Notons que l'égalité est vérifiée pour les cas extrêmes. Autrement dit, les cas où $\alpha = 1$, $m_1 = m_2$ et/ou $m_2 = m_\Omega$.

(a) si $\alpha = 1$, alors $m_1^\alpha = m_\Omega$.
On obtient donc

$$d_k(m_1^1, m_2) = d_k(m_\Omega, m_2)$$

(b) Si $m_1 = m_2$:

$$\begin{aligned} d_k(m_1^\alpha, m_1) &= \frac{1}{\rho}\left\|\mathbf{S}_1^\alpha - \mathbf{S}_1\right\|_k \\ &= \frac{1}{\rho}\left\|\Big((1-\alpha)\mathbf{S}_1 + \alpha\mathbf{I}\Big) - \mathbf{S}_1\right\|_k \\ &= \frac{1}{\rho}\left\|\alpha\Big(\mathbf{I} - \mathbf{S}_1\Big)\right\|_k \\ \Rightarrow d_k(m_1^\alpha, m_1) &= \alpha d_k(m_1, m_\Omega). \end{aligned}$$

(c) Si $m_2 = m_\Omega$:

$$\begin{aligned} d_k(m_1^\alpha, m_\Omega) &= \frac{1}{\rho} \|\mathbf{S}_1^\alpha - \mathbf{I}\|_k \\ &= \frac{1}{\rho} \left\| \Big((1-\alpha)\mathbf{S}_1 + \alpha\mathbf{I} \Big) - \mathbf{I} \right\|_k \\ &= \frac{1}{\rho} \left\| (1-\alpha)\Big(\mathbf{S}_1 - \mathbf{I} \Big) \right\|_k \\ \Rightarrow d_k(m_1^\alpha, m_\Omega) &= (1-\alpha) d_k(m_1, m_\Omega). \end{aligned}$$

2. Montrons que $(1-\alpha).d_k(m_1, m_2) - \alpha.d_k(m_2, m_\Omega) \leq d_k(m_1^\alpha, m_2)$:
soit :

$$\begin{aligned} (1-\alpha) d_k(m_1, m_2) &= \frac{1-\alpha}{\rho} \|\mathbf{S}_1 - \mathbf{S}_2\|_k \\ &= \frac{1}{\rho} \left\| (1-\alpha)\Big(\mathbf{S}_1 - \mathbf{S}_2 \Big) \right\|_k \\ &= \frac{1}{\rho} \left\| (1-\alpha)\mathbf{S}_1 + \alpha\mathbf{I} - \mathbf{S}_2 - \alpha\Big(\mathbf{I} - \mathbf{S}_2 \Big) \right\|_k \\ &= \frac{1}{\rho} \left\| \Big(\mathbf{S}_1^\alpha - \mathbf{S}_2 \Big) + \alpha\Big(\mathbf{S}_2 - \mathbf{I} \Big) \right\|_k \\ &\leq d_k(m_1^\alpha, m_2) + \alpha d_k(m_2, m_\Omega). \end{aligned}$$

A l'issue de ce calcul nous obtenons le résultat suivant :

$$(1-\alpha) d_k(m_1, m_2) - \alpha d_k(m_2, m_\Omega) \leq d_k(m_1^\alpha, m_2) \tag{68}$$

A.6 Démonstration de la proposition 12

Par définition, le conditionnement de Dempster s'écrit :

$$\begin{aligned} m[Y](X) &= \sum_{\substack{A \subseteq \Omega \\ A \cap Y = X}} m(A) \\ &= \sum_{\substack{A \subseteq \Omega \\ A \cap Y = X, z \in A}} m(A) + \sum_{\substack{A \subseteq \Omega \\ A \cap Y = X, z \notin A}} m(A). \end{aligned} \tag{69}$$

Intéressons nous à présent au premier terme de l'équation (69). Si $z \in A$ alors :

$$\begin{aligned} A \cap (Y \cup \{z\}) &= (A \cap Y) \cup (A \cap \{z\}\ ), \\ &= A \cap Y \cup \{z\}. \end{aligned}$$

De plus, si $A \cap Y = X$, alors :

$$A \cap (Y \cup \{z\}) \quad = \quad X \cup \{z\}\,.$$

Réciproquement, supposons que : $X \cup \{z\} = A \cap (Y \cup \{z\}) = (A \cap Y) \cup (A \cap \{z\}\,)$. Puisque $z \notin Y$, nous en déduisons aussi que $z \notin A \cap Y$, d'où : $z \in A \cap \{z\}$. Par conséquent, ce résultat implique : $z \in A$. Nous avons aussi : $X \cup \{z\} = (A \cap Y) \cup \{z\}$. Puisque $z \notin Y$, nous obtenons alors : $(A \cap Y) \cup \{z\} \setminus \{z\} = A \cap Y$. En outre, puisque $z \notin Y$, alors $z \notin X$. Alors :

$$X \cup \{z\} \setminus \{z\} = X = A \cap Y.$$

A l'issue du raisonnement précédent, nous en déduisons que :

$$\sum_{\substack{A \subseteq \Omega \\ A \cap Y = X, z \in A}} m(A) \quad = \quad \sum_{\substack{A \subseteq \Omega \\ A \cap (Y \cup \{z\}) = X \cup \{z\}}} m(A) = m[Y \cup \{z\}]\,(X \cup \{z\})\,.$$

Intéressons nous à présent au terme restant dans l'équation (69). Supposons que $z \notin A$ et que $A \cap Y = X$. Nous pouvons ainsi écrire :

$$\begin{aligned} A \cap (Y \cup \{z\}) &= (A \cap Y) \cup (A \cap \{z\}\,)\,, \\ &= A \cap Y \cup \emptyset, \\ &= X. \end{aligned} \tag{70}$$

Réciproquement, supposons que : $X = A \cap (Y \cup \{z\}) = (A \cap Y) \cup (A \cap \{z\}\,)$. Puisque $z \notin Y$ alors $z \notin X$. Nous obtenons ainsi : $z \notin A \cap \{z\}$, ce qui implique $z \notin A$.
De ce fait nous aboutissons au résultat :

$$X = A \cap Y \cup \emptyset = A \cap Y.$$

A l'issue de ce raisonnement, nous en déduisons que :

$$\sum_{\substack{A \subseteq \Omega \\ A \cap Y = X, z \notin A}} m(A) \quad = \quad \sum_{\substack{A \subseteq \Omega \\ A \cap (Y \cup \{z\}) = X}} m(A) = m[Y \cup \{z\}]\,(X)\,.$$

A.7 Démonstration du lemme 4

Considérons dans un premier temps que d_1 est définie relativement à une règle α-conjonctive. Par définition de la distance d_1, nous écrivons :

$$d_1(m_A, m_B) = \frac{1}{\rho} \|\mathbf{K}_A - \mathbf{K}_B\|_1, \quad (71)$$

où $\mathbf{K}_A$ et $\mathbf{K}_B$ sont les matrices de croyance correspondant à m_A et m_B relativement à la règle $\textcircled{\cap}^\alpha$ et ρ un coefficient de normalisation. Selon les résultats énoncés dans [23], il est établi que l'élément $K_A(X,Y)$ de la matrice $\mathbf{K}_A$ est non nul si et seulement si $A \cap Y \subseteq X \subseteq \overline{A \Delta Y}$. Par conséquent, nous avons :

$$\begin{aligned} \|\mathbf{K}_A - \mathbf{K}_B\|_1 &= \sum_{X,Y} |K_A(X,Y) - K_B(X,Y)|, \\ &= \sum_{X,Y} |K_A(X,Y)\, 1_{A\cap Y \subseteq X \subseteq \overline{A\Delta Y}} - K_B(X,Y)\, 1_{B\cap Y \subseteq X \subseteq \overline{B\Delta Y}}|, \end{aligned}$$

avec 1_{cond} la fonction indicatrice telle que $1_{cond} = 1$ si la condition *cond* est vérifiée et $1_{cond} = 0$ sinon. Il y a une soustraction si toutes les conditions des fonctions indicatrices sont conjointement vérifiées. Analysons à présent cette situation :

- Si X est un sur-ensemble à la fois de $A \cap Y$ et de $B \cap Y$, alors X est un sur-ensemble de $(A \cup B) \cap Y = (A \cap Y) \cup (B \cap Y)$.
- Si X est un sous-ensemble à la fois de $\overline{A \Delta Y}$ et de $\overline{B \Delta Y}$, alors il est un sous-ensemble de $\overline{A \cup B \cup Y} \cup (A \cap B \cap Y) = \overline{A \Delta Y} \cap \overline{B \Delta Y}$.

Soit $I = A \cap B \cap Y$ et $U = A \cup B \cup Y$. On écrit alors :

$$\begin{aligned} \|\mathbf{K}_A - \mathbf{K}_B\|_1 &= \sum_{\substack{X,Y \\ (A\cup B)\cap Y \subseteq X \subseteq \overline{U} \cup I}} |K_A(X,Y) - K_B(X,Y)|, \\ &\quad + \sum_{\substack{X,Y \\ A\cap Y \subseteq X \subseteq \overline{A\Delta Y} \\ B\cap Y \nsubseteq X \text{ ou } X \nsubseteq \overline{B\Delta Y}}} K_A(X,Y) + \sum_{\substack{X,Y \\ B\cap Y \subseteq X \subseteq \overline{B\Delta Y} \\ A\cap Y \nsubseteq X \text{ ou } X \nsubseteq \overline{A\Delta Y}}} K_B(X,Y) \\ &= \|\mathbf{K}_A\|_1 + \|\mathbf{K}_B\|_1 + \\ &\quad \sum_{\substack{X,Y \\ (A\cup B)\cap Y \subseteq X \\ X \subseteq \overline{U} \cup I}} |K_A(X,Y) - K_B(X,Y)| - K_A(X,Y) - K_B(X,Y). \end{aligned}$$

Étant donné que la norme L^1 d'une distance matricielle quelconque entre fonctions de masse est égale à N et que les matrices de croyance sont positives, nous pouvons

aussi écrire :

$$\|\mathbf{K}_A - \mathbf{K}_B\|_1 \;=\; 2N-2 \sum_{\substack{X,Y \\ (A\cup B)\cap Y\subseteq X \\ X\subseteq \overline{U}\cup I}} \min\left\{K_A(X,Y); K_B(X,Y)\right\}. \tag{72}$$

Prouvons à présent que $(A\cup B)\cap Y \subseteq \overline{U}\cup I \Leftrightarrow Y\subseteq \overline{A\Delta B}$:

$$\begin{aligned}
(A\cup B)\cap Y \subseteq \overline{U}\cup I \;&\Leftrightarrow\; (A\cup B)\cap Y \subseteq \overline{A\cup B\cup Y}\cup (A\cap B\cap Y),\\
&\Leftrightarrow\; (A\cup B)\cap Y \subseteq \left(\overline{A\cup B}\cap \overline{Y}\right)\cup (A\cap B\cap Y),\\
&\Leftrightarrow\; (A\cup B)\cap Y \subseteq A\cap B\cap Y,\\
&\Leftrightarrow\; (A\cup B)\cap Y \setminus (A\cap B\cap Y)=\emptyset,\\
&\Leftrightarrow\; (A\Delta B)\cap Y=\emptyset,\\
&\Leftrightarrow\; Y\subseteq \overline{A\Delta B}.
\end{aligned}$$

L'équation (72) peut alors être réecrite comme suit :

$$\|\mathbf{K}_A - \mathbf{K}_B\|_1 \;=\; 2N-2 \sum_{\substack{Y \\ Y\subseteq \overline{A\Delta B}}} \sum_{\substack{X \\ (A\cup B)\cap Y\subseteq X \\ X\subseteq \overline{U}\cup I}} \min\left\{K_A(X,Y); K_B(X,Y)\right\}.$$

Étant donné que $Y\subseteq \overline{A\Delta B}$, nous avons $(A\cup B)\cap Y = A\cap B\cap Y$. par conséquent, il apparait que tous les sous-ensembles X sont l'union de I avec un certain sous-ensemble donné de $\overline{U}$. La seconde somme peut alors être réécrite comme suit :

$$\|\mathbf{K}_A - \mathbf{K}_B\|_1 \;=\; 2N-2 \sum_{\substack{Y \\ Y\subseteq \overline{A\Delta B}}} \sum_{\substack{X \\ X\subseteq \overline{U}}} K_{A\wedge B}(X\cup I, Y), \tag{73}$$

avec $\mathbf{K}_{A\wedge B}$ le minimum point à point des matrices $\mathbf{K}_A$ et $\mathbf{K}_B$. Remarquons que pour chaque sous-ensemble C sachant que $C\cap X=\emptyset$, nous avons :

$$\begin{aligned}
K_{A\wedge B}(X\cup C, Y) \;&=\; \min\left\{\frac{\alpha^{|\overline{A\Delta Y}|}}{\overline{\alpha}^{|A\cap Y|}}\left(\frac{\overline{\alpha}}{\alpha}\right)^{|X\cup C|}; \frac{\alpha^{|\overline{B\Delta Y}|}}{\overline{\alpha}^{|B\cap Y|}}\left(\frac{\overline{\alpha}}{\alpha}\right)^{|X\cup C|}\right\},\\
&=\; \min\left\{\frac{\alpha^{|\overline{A\Delta Y}|}}{\overline{\alpha}^{|A\cap Y|}}; \frac{\alpha^{|\overline{B\Delta Y}|}}{\overline{\alpha}^{|B\cap Y|}}\right\}\left(\frac{\overline{\alpha}}{\alpha}\right)^{|X\cup C|},\\
&=\; \min\left\{\frac{\alpha^{|\overline{A\Delta Y}|}}{\overline{\alpha}^{|A\cap Y|}}; \frac{\alpha^{|\overline{B\Delta Y}|}}{\overline{\alpha}^{|B\cap Y|}}\right\}\left(\frac{\overline{\alpha}}{\alpha}\right)^{|X|+|C|}.
\end{aligned}$$

La somme sur tous les sous-ensembles X peut alors être calculée explicitement :

$$\begin{aligned}
\sum_{\substack{X\\ X\subseteq\overline{U}}} K_{A\wedge B}\left(X\cup I,Y\right) &= \min\left\{\frac{\alpha^{|\overline{A\Delta Y}|}}{\overline{\alpha}^{|A\cap Y|}};\frac{\alpha^{|\overline{B\Delta Y}|}}{\overline{\alpha}^{|B\cap Y|}}\right\}\left(\frac{\overline{\alpha}}{\alpha}\right)^{|I|}\sum_{\substack{X\\ X\subseteq\overline{U}}}\left(\frac{\overline{\alpha}}{\alpha}\right)^{|X|},\\
&= \min\left\{\frac{\alpha^{|\overline{A\Delta Y}|}}{\overline{\alpha}^{|A\cap Y|}};\frac{\alpha^{|\overline{B\Delta Y}|}}{\overline{\alpha}^{|B\cap Y|}}\right\}\left(\frac{\overline{\alpha}}{\alpha}\right)^{|I|}\sum_{i=1}^{|\overline{U}|}\left(\begin{array}{c}|\overline{U}|\\ i\end{array}\right)\left(\frac{\overline{\alpha}}{\alpha}\right)^{i},\\
&= \min\left\{\frac{\alpha^{|\overline{A\Delta Y}|}}{\overline{\alpha}^{|A\cap Y|}};\frac{\alpha^{|\overline{B\Delta Y}|}}{\overline{\alpha}^{|B\cap Y|}}\right\}\left(\frac{\overline{\alpha}}{\alpha}\right)^{|I|}\left(\frac{\overline{\alpha}}{\alpha}+1\right)^{|\overline{U}|},\\
&= \min\left\{\frac{\alpha^{|\overline{A\Delta Y}|}}{\overline{\alpha}^{|A\cap Y|}};\frac{\alpha^{|\overline{B\Delta Y}|}}{\overline{\alpha}^{|B\cap Y|}}\right\}\left(\frac{\overline{\alpha}}{\alpha}\right)^{|I|}\alpha^{-|\overline{U}|}.
\end{aligned}$$

L'équation (73) peut alors être réécrite comme suit :

$$\left\|\mathbf{K}_A-\mathbf{K}_B\right\|_1 = 2N-2\sum_{\substack{Y\\ Y\subseteq\overline{A\Delta B}}}\min\left\{\frac{\alpha^{|\overline{A\Delta Y}|}}{\overline{\alpha}^{|A\cap Y|}};\frac{\alpha^{|\overline{B\Delta Y}|}}{\overline{\alpha}^{|B\cap Y|}}\right\}\left(\frac{\overline{\alpha}}{\alpha}\right)^{|I|}\alpha^{-|\overline{U}|}.$$

Puisque $Y\subseteq\overline{A\Delta B}$, Y est l'union de deux sous-ensembles disjoints $W\subseteq A\cap B$ et $Z\subseteq\overline{A\cup B}$. En particulier, nous avons :

- $|\overline{A\Delta Y}| = |\overline{A\Delta\left(W\cup Z\right)}| = |\overline{\left(A\cup Z\right)\setminus W}| = n-|A|-|Z|+|W|$ et également $|\overline{B\Delta Y}| = n-|B|-|Z|+|W|$,
- $|A\cap Y| = |A\cap\left(W\cup Z\right)| = |W|$ et également $|B\cap Y| = |W|$,
- $|I| = |W|$,
- $|\overline{U}| = |\overline{A\cup B\cup Z}| = n-|A\cup B|+|Z|$.

Cette décomposition de Y nous permet d'écrire :

$$\begin{aligned}
\min\left\{\frac{\alpha^{|\overline{A\Delta Y}|}}{\overline{\alpha}^{|A\cap Y|}};\frac{\alpha^{|\overline{B\Delta Y}|}}{\overline{\alpha}^{|B\cap Y|}}\right\} &= \min\left\{\frac{\alpha^{n-|A|-|Z|+|W|}}{\overline{\alpha}^{|W|}};\frac{\alpha^{n-|B|-|Z|+|W|}}{\overline{\alpha}^{|W|}}\right\},\\
&= \min\left\{\alpha^{-|A|};\alpha^{-|B|}\right\}\frac{\alpha^{n-|Z|+|W|}}{\overline{\alpha}^{|W|}},\\
&= \frac{\alpha^{n-|Z|+|W|-\min\{|A|;|B|\}}}{\overline{\alpha}^{|W|}}.
\end{aligned}$$

En reprenant le calcul de $\|\mathbf{K}_A - \mathbf{K}_B\|_1$, les simplifications des termes à la puissance de $|W|$ et $|Z|$ sont observées, ceci nous donne :

$$\begin{aligned}
\|\mathbf{K}_A - \mathbf{K}_B\|_1 &= 2N - 2 \sum_{\substack{W \\ W \subseteq A \cap B}} \sum_{\substack{Z \\ Z \subseteq \overline{A \cup B}}} \alpha^{|A \cup B| - \min\{|A|;|B|\}}, \\
&= 2N - 2\alpha^{|A \cup B| - \min\{|A|;|B|\}} 2^{|A \cap B|} 2^{|\overline{A \cup B}|}, \\
&= 2N \left(1 - \alpha^{|A \cup B| - \min\{|A|;|B|\}} 2^{|A \cap B| - |A \cup B|}\right), \\
&= 2N \left(1 - \frac{\alpha^{\max\{|A \setminus B|;|B \setminus A|\}}}{2^{|A \Delta B|}}\right).
\end{aligned}$$

Intéressons nous à présent au cas α-disjonctif. Pour le reste de cette démonstration, nous distinguons alors $d_{1,\cap}$ et $d_{1,\cup}$. A partir de la proposition 13, on écrit :

$$\begin{aligned}
d_{1,\cup}(m_A, m_B) &= d_{1,\cap}(\overline{m}_A, \overline{m}_B), \\
&= d_{1,\cap}(m_{\overline{A}}, m_{\overline{B}}), \\
&= \frac{2N}{\rho} \left(1 - \frac{\alpha^{\max\{|\overline{A} \setminus \overline{B}|;|\overline{B} \setminus \overline{A}|\}}}{2^{|\overline{A} \Delta \overline{B}|}}\right), \\
&= \frac{2N}{\rho} \left(1 - \frac{\alpha^{\max\{|B \setminus A|;|A \setminus B|\}}}{2^{|A \Delta B|}}\right), \\
&= d_{1,\cap}(m_A, m_B).
\end{aligned}$$

A.8 Démonstration de la proposition 17

Soit $\mathbf{K}_A$ la matrice de croyance d'une fonction de masse catégorique quelconque m_A relative à une α-jonction. Si $A = B$, les propriétés sont alors trivialement prouvées. Si $A \subsetneq B$, $\exists x \in B$ sachant que $x \notin A$. Intéressons-nous au cas de la classe des α-conjonctions et des α-disjonctions séparément :

– dans le cas d'une α-conjonction, on peut écrire :

$$\begin{aligned}
\|\mathbf{m}_1 \textcircled{\cap}^{\alpha} \mathbf{m}_A - \mathbf{m}_2 \textcircled{\cap}^{\alpha} \mathbf{m}_A\|_1 &= \|\mathbf{K}_A.\mathbf{m}_1 - \mathbf{K}_A.\mathbf{m}_2\|_1 \\
&= \|\mathbf{K}_A(\mathbf{m}_1 - \mathbf{m}_2)\|_1 \\
&= \left\| \prod_{x' \notin A} \mathbf{K}_{\overline{\{x'\}}} (\mathbf{m}_1 - \mathbf{m}_2) \right\|_1 \\
&= \left\| \mathbf{K}_{\overline{\{x\}}}.\mathbf{K}_{A \cup \{x\}} (\mathbf{m}_1 - \mathbf{m}_2) \right\|_1 \qquad (74)
\end{aligned}$$

La définition de la norme matricielle d'opérateur 1 d'une matrice $\mathbf{K}_{\overline{\{x\}}}$ s'écrit

comme suit :

$$\left\|\mathbf{K}_{\overline{\{x\}}}\right\|_{op1} = \max_{\mathbf{u}\in\mathbb{R}^{|2^{\Omega}|}} \frac{\left\|\mathbf{K}_{\overline{\{x\}}}\mathbf{u}\right\|_1}{\|\mathbf{u}\|_1}.$$

Comme $\mathbf{K}_{A\cup\{x\}}(\mathbf{m}_1-\mathbf{m}_2)$ est un vecteur particulier appartenant à $\mathbb{R}^{|2^{\Omega}|}$, on obtient :

$$\left\|\mathbf{K}_{\overline{\{x\}}}\right\|_{op1} \geq \frac{\left\|\mathbf{K}_{\overline{\{x\}}}.\mathbf{K}_{A\cup\{x\}}(\mathbf{m}_1-\mathbf{m}_2)\right\|_1}{\left\|\mathbf{K}_{A\cup\{x\}}(\mathbf{m}_1-\mathbf{m}_2)\right\|_1}. \quad (75)$$

L'utilisation de l'inégalité précédente dans l'équation (74) donne :

$$\|\mathbf{m}_1 \textcircled{\cap}^{\alpha}\mathbf{m}_A - \mathbf{m}_2 \textcircled{\cap}^{\alpha}\mathbf{m}_A\|_1 \leq \left\|\mathbf{K}_{\overline{\{x\}}}\right\|_{op1} \left\|\mathbf{K}_{A\cup\{x\}}(\mathbf{m}_1-\mathbf{m}_2)\right\|_1.$$

Puisque ${}^t\mathbf{K}_{\overline{\{x\}}}$ est une matrice stochastique, nous avons $\left\|\mathbf{K}_{\overline{\{x\}}}\right\|_{op1}=1$. Ceci nous permet d'écrire :

$$\|\mathbf{m}_1 \textcircled{\cap}^{\alpha}\mathbf{m}_A - \mathbf{m}_2 \textcircled{\cap}^{\alpha}\mathbf{m}_A\|_1 \leq \left\|\mathbf{m}_{A\cup\{x\}} \textcircled{\cap}^{\alpha}\mathbf{m}_1 - \mathbf{m}_{A\cup\{x\}} \textcircled{\cap}^{\alpha}\mathbf{m}_2\right\|_1. \quad (76)$$

Le même résultat peut être utilisé par inclusions successives des autres éléments x' appartenant à B et qui n'appartiennent pas à A, donc :

$$\|\mathbf{m}_1 \textcircled{\cap}^{\alpha}\mathbf{m}_A - \mathbf{m}_2 \textcircled{\cap}^{\alpha}\mathbf{m}_A\|_1 \leq \|\mathbf{m}_B \textcircled{\cap}^{\alpha}\mathbf{m}_1 - \mathbf{m}_B \textcircled{\cap}^{\alpha}\mathbf{m}_2\|_1. \quad (77)$$

– dans le cas d'une α-disjonction, on peut écrire :

$$\begin{aligned}\left\|\mathbf{m}_1 \textcircled{\cup}^{\alpha}\mathbf{m}_{A\cup\{x\}} - \mathbf{m}_2 \textcircled{\cup}^{\alpha}\mathbf{m}_{A\cup\{x\}}\right\|_1 &= \left\|\mathbf{K}_{A\cup\{x\}}.\mathbf{m}_1 - \mathbf{K}_{A\cup\{x\}}.\mathbf{m}_2\right\|_1 \\ &= \left\|\mathbf{K}_{A\cup\{x\}}(\mathbf{m}_1-\mathbf{m}_2)\right\|_1 \\ &= \left\|\prod_{x'\in A\cup\{x'\}}\mathbf{K}_{\{x'\}}(\mathbf{m}_1-\mathbf{m}_2)\right\|_1 \\ &= \left\|\mathbf{K}_{\{x\}}.\mathbf{K}_A(\mathbf{m}_1-\mathbf{m}_2)\right\|_1. \quad (78)\end{aligned}$$

La définition de la norme d'opérateur 1 pour une matrice $\mathbf{K}_{\{x\}}$ s'écrit comme suit :

$$\left\|\mathbf{K}_{\{x\}}\right\|_{op1} = \max_{\mathbf{u}\in\mathbb{R}^{|2^{\Omega}|}} \frac{\left\|\mathbf{K}_{\{x\}}\mathbf{u}\right\|_1}{\|\mathbf{u}\|_1}.$$

Comme $\mathbf{K}_A(\mathbf{m}_1 - \mathbf{m}_2)$ est un vecteur particulier appartenant à $\mathbb{R}^{|2^\Omega|}$, on obtient :

$$\left\|\mathbf{K}_{\{x\}}\right\|_{op1} \geq \frac{\left\|\mathbf{K}_{\{x\}}.\mathbf{K}_A(\mathbf{m}_1 - \mathbf{m}_2)\right\|_1}{\left\|\mathbf{K}_A(\mathbf{m}_1 - \mathbf{m}_2)\right\|_1}. \tag{79}$$

L'utilisation de l'inégalité précédente dans l'équation (78) donne :

$$\left\|\mathbf{m}_1 \textcircled{\cup}^\alpha \mathbf{m}_{A\cup\{x\}} - \mathbf{m}_2 \textcircled{\cup}^\alpha \mathbf{m}_{A\cup\{x\}}\right\|_1 \leq \left\|\mathbf{K}_{\{x\}}\right\|_{op1} \left\|\mathbf{K}_A(\mathbf{m}_1 - \mathbf{m}_2)\right\|_1.$$

Et puisque ${}^t\mathbf{K}_{\{x\}}$ est une matrice stochastique, nous avons $\left\|\mathbf{K}_{\{x\}}\right\|_{op1} = 1$. ceci nous permet d'écrire :

$$\left\|\mathbf{m}_1 \textcircled{\cup}^\alpha \mathbf{m}_{A\cup\{x\}} - \mathbf{m}_2 \textcircled{\cup}^\alpha \mathbf{m}_{A\cup\{x\}}\right\|_1 \leq \left\|\mathbf{m}_A \textcircled{\cup}^\alpha \mathbf{m}_1 - \mathbf{m}_A \textcircled{\cup}^\alpha \mathbf{m}_2\right\|_1. \tag{80}$$

Ce résultat peut être utilisé par inclusion successive des autres éléments x' appartenant à B et qui n'appartiennent pas à A, donc :

$$\left\|\mathbf{m}_1 \textcircled{\cup}^\alpha \mathbf{m}_A - \mathbf{m}_2 \textcircled{\cup}^\alpha \mathbf{m}_A\right\|_1 \geq \left\|\mathbf{m}_B \textcircled{\cup}^\alpha \mathbf{m}_1 - \mathbf{m}_B \textcircled{\cup}^\alpha \mathbf{m}_2\right\|_1. \tag{81}$$

Bibliographie

[1] Mathias Bauer : Approximation algorithms and decision making in the dempster-shafer theory of evidence : An empirical study. *International Journal of Approximate Reasoning*, 17:217 – 237, 1997.

[2] Mathieu Bouchard, Anne-Laure Jousselme et Pierre-Emmanuel Doré : A proof for the positive definiteness of the Jaccard index matrix. *International Journal of Approximate Reasoning*, 54(5):615 – 626, 2013.

[3] Fabio Cuzzolin : Geometry of Dempster's rule of combination. *IEEE Transactions on Systems, Man, and Cybernetics. Part B : Cybernetics*, 34(2):961 – 977, 2004.

[4] Fabio Cuzzolin : A geometric approach to the theory of evidence. *IEEE Transactions on Systems, Man and Cybernetics - Part C : Application And Reviews*, 38(4):522 – 534, 2008.

[5] Arthur Pentland Dempster : Upper and lower probabilities induced by a multiple valued mapping. *Annals of Mathematical Satistics*, 38:325 – 339, 1967.

[6] Thierry Denœux : Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artificial Intelligence*, 172:234 – 264, 2008.

[7] Sébastien Destercke et Thomas Burger : Toward an axiomatic definition of conflict between belief functions. *Cybernetics, IEEE Transactions on*, 43(2):585–596, April 2013.

[8] Javier. Diaz, Maria. Rifqi et Bernadette. Bouchon-Meunier : A similarity measure between basic belief assignments. *In Proceedings of the 9th International Conference on Information Fusion (FUSION 2006)*, pages 1–6, Florence, Italy, 2006.

[9] Didier Dubois et Henri Prade : A set-theoretic view of belief functions : logical operations and approximations by fuzzy sets. *International Journal of General Systems*, 12(3):193–226, 1986.

[10] Didier Dubois et Henri Prade : Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4(3):244–264, 1988.

[11] Didier Dubois et Henri Prade : Consonant approximations of belief functions. *International Journal of Approximate Reasoning*, 4:419 – 449, 1990.

[12] Nicole El Zoghby, Véronique Cherfaoui, Bertrand Ducourthial et Thierry Denoeux : Distributed Data fusion for detecting Sybil attacks in VANETs. *In Belief functions : theory and applications.*, volume AISC 164, pages 351–358, Compiègne, France, 2012.

[13] Zied Elouedi, Khaled Mellouli et Philippe Smets : Assessing sensor reliability for multisensor data fusion within the transferable belief model. *IEEE Transactions on Systems, Man, and Cybernetics. Part B : Cybernetics*, 34(1):782 – 787, 2004.

[14] Dale Fixen et Ronald Mahler : The modified Dempster-Shafer approach to classification. *IEEE Transactions on Systems, Man and Cybernetics. Part A : Systems and humans*, 27(1): 96 – 104, 1997.

[15] Mihai Cristian Florea et al. : Robuste combination rules for evidence theory. *Information Fusion*, 10:183–197, 2009.

[16] Mihai Cristian Florea, Jean Dezert, Pierre Valin, Florentin Smarandache et Anne-Laure Jouselme : Adaptative combination rule and and proportional conflict redistribution rule for information fusion. *In Proceedings of the COGnitive systems with Interactive Sensors (COGIS-6)*, volume abs/cs/0604042, pages 1–8, Paris, France, March 2006.

[17] Deqiang Han, Jean Dezert et Chongzhao Han ahd Yi Yang : New dissimilarity measure in evidence theory. *In Proceedings of the 14th International Conference on Information Fusion (FUSION 2011)*, pages 1 – 7, Chicago, IL. United States, 2011.

[18] Anne-Laure Jousselme, Chunsheng Liu Dominic Grenier et Eloi Bossé : Measuring ambiguity in the evidence theory. *IEEE Transactions on Systems, Man, and Cybernetics -Part A : Systems and Humans*, 36 No 5:890–903, 2006.

[19] Anne-Laure Jousselme, Dominic Grenier et Eloi Bossé : A new distance between two bodies of evidence. *Information Fusion*, 2:91–101, 2001.

[20] Anne-Laure Jousselme et Patrick Maupin : On some properties of distances in evidence theory. *In Proceedings of BELIEF. International workshop on the theory of belief functions*, pages 1 – 6, Brest, France, 2010.

[21] Anne-Laure Jousselme et Patrick Maupin : Distances in evidence theory : Comprehensive survey and generalizations. *International Journal of Approximate Reasoning*, 53:118 – 145, 2012.

[22] Frank Klawonn et Philippe Smets : The dynamic of belief in the transferable belief model and specialization-generalization matrices. *In* D. Dubois, M. P. Wellman, B. d'Ambrosio et Ph. Smets, éditeurs : *Proceedings of the eighth International Conference on Uncertainty in Artificial Intelligence*, volume 92, pages 130 –137, San Mateo, CA. USA, 1992. Morgan Kaufmann Publishers Inc.

[23] John Klein, Mehena Loudahi, Jean-Marc Vannobel et Olivier Colot : α-junctions of categorical mass functions. *In Belief Functions : Theory and Applications – Accepted*, volume XX de *Advances in Intelligent and Soft Computing*, pages 1 – 10. Springer Berlin Heidelberg, 2014.

[24] George J. Klir et Mark J. Wierman : *Uncertainty-Based Information : Elements of Generalized Information Theory,*. Springer, 1999.

[25] Eric Lefevre : *fusion adaptée d'informations conflictuelles dans le cadre de la théorie de l'évidence*. Thèse de doctorat, Institut National des Sciences Appliquées de Rouen, 2001.

[26] Eric Lefevre, Olivier Colot et Patrick Vannnoorenberghe. : Information et combinaison : les liaisons conflictuelles. *Traitement du Signal*, 18-n 3:161–177, 2001.

[27] Eric Lefevre, Olivier Colot et Patrick Vannnoorenberghe. : Belief function combination and conflict management. *Information Fusion*, 3:149–162, 2002.

[28] Weiru Liu : Analysing the degree of conflict among belief functions. *Artificial Intelligence*, 170(11):909–924, 2006.

[29] Zhun-Ga Liu, Jean Dezert et Quan Pan : A new measure of dissimilarity between two basic belief assignments. *In Information fusion (submitted)*, pages 1 – 11, 2010.

[30] Zhun-Ga Liu, Jean Dezert, Quan Pan et Grégoire Mercier : Combination of sources of evidence with different discounting factors based on a new dissimilarity measure. *Decision Support Systems*, 52:133 –141, 2011.

[31] Mehena Loudahi, John Klein, Jean-Marc Vannobel et Olivier Colot : Fast computation of l^p norm-based specialization distances between bodies of evidence. *In Belief Functions : Theory and Applications – Accepted*, volume XX de *Advances in Intelligent and Soft Computing*, pages 1 – 9. Springer Berlin Heidelberg, 2014.

[32] Mehena Loudahi, John Klein, Jean-Marc Vannobel et Olivier Colot : New distances between bodies of evidence based on dempsterian specialization matrices and their consistency with the conjunctive combination rule. *International Journal of Approximate Reasoning*, (Volume 55, Issue 5):1093–1112, 2014.

[33] Arnaud Martin : *Habilité à Diriger des Recherches : Modélisation et gestion du conflit dans la théorie des fonctions de croyances.* Thèse de doctorat, Université de Bretagne Occidentale, 2009.

[34] Arnaud Martin, Anne-Laure Jousselme et Christophe Osswald : Conflict measure for the discounting operation on belief functions. *In Proceedings of the 11th International Conference on Information Fusion (FUSION 2008)*, pages 1–8, Cologne, Germany, June 2008.

[35] Arnaud Martin et Christophe Osswald : Une nouvelle règle de combinaison répartissant le conflit - application en imagerie sonar et classification de cibles radar. *Traitement du Signal*, 24-n 2:71–82, 2007.

[36] Hongming Mo, Xiaoyan Su, Yong Hu et Yong Deng : A generalized evidence distance. *Computing Research Repository (CoRR)*, abs/1311.4056, 2013.

[37] Paul-André Monney : Dempster specialization matrices and the combination of belief functions. *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, Lecture Notes in Computer Science*, 2143:316–327, 2001.

[38] Christophe Osswald et Arnaud Martin : Understanding the large family of dempster-shafer theory's fusion operators - a decision-based measure. *In Proceedings of the 9th International Conference on Information Fusion (FUSION)*, pages 1–7, Florence. Italy, July 2006.

[39] Walter L. Perry et Harry E. Stephanou. : Belief function divergence as a classifier. *In Intelligent Control, 1991., Proceedings of the 1991 IEEE International Symposium on*, pages 280–285, Aug 1991.

[40] Frédéric Pichon : *Fonctions de Croyance : Décompositions Canoniques et Règles de combinaison.* Thèse de doctorat, Université de Technologie de Compiègne – UTC, 2009.

[41] Frédéric Pichon : On the α-conjunctions for combining belief functions. *In Belief Functions : Theory and Applications*, volume 164 de *Advances in Intelligent and Soft Computing*, pages 285–292. Springer Berlin Heidelberg, 2012.

[42] Frédéric Pichon et Thierry Denœux : Interpretation and computation of α-junctions for combining belief functions. *In 6th International Symposium on Imprecise Probability : Theories and Applications ISIPTA-09*, pages 1 – 10, Durham, United Kingdom, July 2009.

[43] Frédéric Pichon et Thierry Denœux : The unnormalized dempsterś rule of combination : A new justification from the least commitment principle and some extensions. *Journal of Automated Reasoning*, 45(1):61–87, 2010.

[44] Frédéric Pichon, Didier Dubois et Thierry Denœux : Relevance and truthfulness in information correction and fusion. *International Journal of Approximate Reasoning*, 53:159 – 175, 2012.

[45] Glenn Shafer : *A mathematical theory of evidence*. Princeton University Press, 1976.

[46] Florentin Smarandache et Jean Dezert : Proportional conflict redistribution rules for information fusion. *In Proceedings of the 8th International Conference on Information Fusion (FUSION)*, volume cs.AI/0408064, pages 1–41, Philadelphia. United States, 25-29 July 2005.

[47] Florentin Smarandache, Arnaud Martin et Christophe Osswald : Contradiction measures and specificity degrees of basic belief assignments. *In The International Conference on Information Fusion*, volume abs/1109.3700, pages 1–8, Chicago. United States, 2011.

[48] Philippe Smets : The combination of evidence in the transferable belief model. *IEEE Transactions on On Pattarn Analysis And Machine Intelligence.*, 11(5):447–458, 1990.

[49] Philippe Smets : Constructing the pignistic probability function in a context of uncertainty. *Uncertainty in Artificial Intelligence*, 5:29–39, 1990.

[50] Philippe Smets : The canonical decomposition of a weighted belief. *14th international joint conference on Artificial intelligence*, 2:1896–1901, 1995.

[51] Philippe Smets : The $\alpha-$junctions : Combination operators applicable to belief functions. *In* DovM. Gabbay, Rudolf Kruse, Andreas Nonnengart et HansJürgen Ohlbach, éditeurs : *Qualitative and Quantitative Practical Reasoning*, volume 1244 de *Lecture Notes in Computer Science*, pages 131–153. Springer Berlin Heidelberg, 1997.

[52] Philippe Smets : The application of the matrix calculus to belief functions. *International Journal of Approximate Reasoning*, 31:1–30, 2002.

[53] Zachary Sunberg et Jonathan Rogers : A belief function distance metric for orderable sets. *Information Fusion*, 14:361 – 373, 2013.

[54] Bjornar Tessem : Approximations for efficient computation in the theory of evidence. *Artificial Intelligence*, 61:315 – 329, 1993.

[55] Thomas Weiler et Ulrich Bodenhofer : Approximation of belief functions by minimizing Euclidean distances. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 11(06):170 –177, 2003.

[56] Ronald Robert Yager : The entailment principle for dempster-shafer granules. *International Journal of Intelligent Systems*, 1(4):247 – 262, 1986.

[57] Ronald Robert Yager : On the dempster-shafer framework and new combinaison rules. *Informations Sciences*, 41:93 – 137, 1987.

[58] Jianping YANG, Bing BAI et Xiaojun Jiang nd HONG-ZHONG HUANG : A novel method for measuring the dissimilarity degree between two pieces of evidence in dempster-shafer evidence théorie. *In International Confernce on Quality, Reliability, Risk, Maintenance, and Safety Engennering (QR2MSE)*, 2013.

[59] Lalla Mariem ZOUHAL et Tierry DENŒUX : An evidence-theoretic k-nn rule with parameter optimisation. *IEEE Transactions on Systems, Man and Cybernetics. Part C : Application and reviews*, 28(2):263 – 271, 1998.