



HAL
open science

Modèle d'interaction et performances du traitement du signal multimodal

Saloua Chlaily

► **To cite this version:**

Saloua Chlaily. Modèle d'interaction et performances du traitement du signal multimodal. Traitement du signal et de l'image [eess.SP]. Université Grenoble Alpes, 2018. Français. NNT : 2018GREAT026 . tel-01880756

HAL Id: tel-01880756

<https://theses.hal.science/tel-01880756>

Submitted on 25 Sep 2018

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

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

UNIVERSITÉ GRENOBLE ALPES

THÈSE

pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Signal, Image, Parole, Télécoms**

Arrêté ministériel : 25 mai 2016

Présentée par

Saloua CHLAILY

Thèse dirigée par **Christian JUTTEN** et
codirigée par **Pierre-Olivier AMBLARD** et **Olivier MICHEL**
préparée au sein du laboratoire **Grenoble Image Parole Signal
Automatique (GIPSA-lab)**
dans l'école doctorale **Electronique, Electrotechnique,
Automatique, Traitement du Signal (EEATS)**

Modèle d'interaction et performances du traitement du signal multimodal Interaction model and performance of multimodal signal processing

Thèse soutenue publiquement le **4 Avril 2018**,
devant le jury composé de :

Guillaume GINOLHAC

Professeur, Polytech Annecy-Chambéry, Président

Inbar FIJALKOW

Professeure, ENSEA, Université Cergy-Pontoise, Rapportrice

Jean-François BERCHER

Professeur, ESIEE, Université Paris-Est, Rapporteur

Pierre BORGNAT

Directeur de recherche au CNRS, ENS Lyon, Examineur

Christian JUTTEN

Professeur, GIPSA-lab, Communauté UGA, Directeur de thèse

Pierre-Olivier AMBLARD

Directeur de recherche au CNRS, GIPSA-lab, Co-directeur de thèse

Olivier MICHEL

Professeur, GIPSA-lab, Grenoble INP, Co-directeur de thèse



UNIVERSITÉ DE GRENOBLE ALPES
ÉCOLE DOCTORALE EEATS

Electronique, Electrotechnique, Automatique, Traitement du Signal

THÈSE

pour obtenir le titre de

docteur en sciences

de l'Université de Grenoble Alpes

Mention : SIGNAL, IMAGE, PAROLE, TÉLÉCOMS

par

Saloua CHLAILY

**Modèle d'interaction et performances du traitement du signal
multimodal**

Thèse dirigée par Christian JUTTEN, Pierre-Olivier AMBLARD et Olivier MICHEL
préparée au laboratoire Grenoble Image Parole Signal Automatique (GIPSA-lab)

Jury :

<i>Président :</i>	Guillaume GINOLHAC	- Polytech Annecy-Chambéry, France
<i>Rapporteurs :</i>	Inbar FIJALKOW	- ENSEA, Université Cergy-Pontoise, France
	Jean-François BERCHER	- ESIEE, Université Paris-Est, France
<i>Examineur :</i>	Pierre BORGNAT	- CNRS, ENS, Lyon
<i>Directeur :</i>	Christian JUTTEN	- Université Grenoble Alpes, GIPSA-lab, France
<i>Co-directeur :</i>	Pierre-Olivier AMBLARD	- CNRS, GIPSA-lab, France
<i>Invité :</i>	Olivier MICHEL	- Grenoble INP, GIPSA-lab, France

Remerciement

C'était qu'hier que j'ai commencé ma thèse, et ces trois années et demi se sont rapidement écoulées. Durant lesquelles j'ai bien grandi dans un bon environnement entourée par de bons gens. Je termine ma thèse par un petit mot pour les remercier tous.

J'étais très chanceuse d'avoir travaillé avec trois scientifiques éminents qui m'ont beaucoup appris. Christian, Pierre-Olivier et Olivier, je ne pourrai jamais vous remercier assez pour ce que vous avais fait pour moi. Merci pour votre temps, efforts et bienveillance. Chirstian, merci de ta relecture minutieuse de ma thèse qui a donné une telle qualité.

Je tiens à remercier les rapporteurs d'avoir accepté de relire ma thèse. Je remercie vivement tous les membres de Jury pour leurs remarques constructives et l'intérêt qu'ils ont porté à mon travail.

Je tiens à remercier Cyrille Siclet de m'avoir confié l'enseignement de certaines de ses classes. Je le remercie aussi pour sa disponibilité, ses conseils et son accompagnement pour la réussite de cette tâche.

Lors de ma thèse, j'ai collaboré avec deux post-docs qui sont à présents des maitres de conférences Chenfang Ren et Ebadollah roonizi. Je les remercie pour la richesse de nos discussions.

Je remercie toute GIPSA, en particulier les équipes CICS et VIBS dont j'ai fait partie ainsi que les ressources humaines pour leurs aides en particulier Martine et Lucie. Je remercie le conseil européen d'avoir financé mes travaux de recherche dans le cadre du projet CHESS.

L'un des précieux cadeaux que la vie peut nous offrir sont les amis. Cette thèse ma permis de rencontrer des gens d'or que je remercie pour leurs amours : Dora la souriante, Jitendra le grand coeur, Karina la gracieuse et sérieuse, Akerke la persévérante, Hardik le gentil toujours occupé, Fateme, Tien et Julien... Merci pour ces moments merveilleux qu'on a passé ensemble. Je remercie Dana et Francesca pour nos profondes conversations et pour leurs bons conseils.

Vient le temps de remercier les gens qui étaient bien longtemps à mes côtés avant et durant la thèse. Je commence par mes meilleures amies, que ni les années ni la distance n'ont pu affaiblir notre amitié : ma thérapeute Rjiwa, ma raison et mon compagnon dans les bêtises Malika.

Je tiens à remercier mes sources de joie et d'amour : ma soeur Btissam, mes deux frères Anouar et Bader.

A mes parents, la lumière de ma vie, un simple merci ne suffit pas, pas même des millions.

Table de matières

Table des figures	vii
Notations	xi
Introduction	1
I Modèle adéquat	5
1 Mesures d'information et d'estimation	7
2 Modalités avec un couplage fort	19
3 Modalités avec des couplages faibles	53
II Modèle inadéquat	73
4 Mesures d'information et d'estimation	77
5 Modèle inadéquat : Modalités avec un couplage fort	109
6 Modèle inadéquat : Modalités avec des couplages faibles	127
Conclusions et perspectives	147
Annexes	151
A Modèle adéquat : Calcul d'estimation	153
B Modèle adéquat : Calcul d'Information	161
C Modèle inadéquat : Démonstration du théorème 4.5	165
D Modèle inadéquat : Calcul d'estimation	173
E Calcul matriciel	179
Bibliographie	181

Table de matières détaillées

Table des figures	vii
Notations	xi
Introduction	1
I Modèle adéquat	5
1 Mesures d'information et d'estimation	7
1.1 Mesures d'information	8
1.1.1 Entropie	8
1.1.2 Entropie différentielle	9
1.1.3 Entropie relative	10
1.1.4 Information mutuelle	10
1.1.5 Information mutuelle conjointe	12
1.2 Mesures d'estimation	12
1.2.1 Estimateur optimal au sens de l'erreur quadratique moyenne	12
1.2.2 Information de Fisher	14
1.3 Information versus estimation	15
1.3.1 Identité de De Bruijn	15
1.3.2 Information de Fisher et divergence de Kullback-Leibler	16
1.3.3 Information mutuelle et erreur quadratique moyenne	16
1.4 Conclusion	17
2 Modalités avec un couplage fort	19
2.1 Introduction	19
2.2 Deux modalités	20
2.2.1 Modèle	20
2.2.2 Calculs d'estimation et d'information	21
2.2.3 Une ou deux modalités ?	22
2.2.4 Effet de la corrélation des bruits et des rapports signal à bruit	25

2.3	Trois modalités	31
2.3.1	Modèle	32
2.3.2	Valeurs éligibles des corrélations	33
2.3.3	Calculs d'estimation et d'information	36
2.3.4	Une ou trois modalités ?	37
2.3.5	Deux modalités parmi trois	39
2.3.6	Deux ou trois modalités ?	46
2.3.7	Effet des corrélations des bruits	48
2.4	Conclusion	51
3	Modalités avec des couplages faibles	53
3.1	Introduction	53
3.2	Modèle	54
3.3	Calculs d'estimation et d'information	55
3.4	Analyse séparée ou conjointe ?	56
3.5	Interaction des modalités	57
3.5.1	Modalités de même qualité ($\alpha_2 = \alpha_1$)	58
3.5.2	Modalités de différentes qualités ($\alpha_2 < \alpha_1$)	58
3.6	Effet des corrélations et des rapports signal à bruit	61
3.6.1	Deux modalités de même qualité	62
3.6.2	Deux modalités de qualités différentes	65
3.6.3	Deux modalités de qualités très différentes	68
3.7	Conclusion	70
II	Modèle inadéquat	73
4	Mesures d'information et d'estimation	77
4.1	Mesures d'information	78
4.1.1	Entropie inadéquate et entropie différentielle inadéquate	78
4.1.2	Information mutuelle inadéquate	81
4.2	Mesures d'estimation	83
4.2.1	Estimateur optimal au sens de l'erreur quadratique moyenne	83
4.3	Information versus estimation	84
4.3.1	Canal gaussien	85
4.3.2	Modèle gaussien	93
4.3.3	Exemple scalaire	98
4.4	Conclusion	106
5	Modèle inadéquat : Modalités avec un couplage fort	109
5.1	Modèle	110
5.2	Mesures d'estimation	111
5.3	Liens des bruits ignorés	113
5.3.1	Modèle inadéquat <i>versus</i> modèle adéquat	114

5.3.2	Modèle inadéquat <i>versus</i> une seule modalité	119
5.4	Inadéquation quelconque des liens entre les bruits	121
5.4.1	Modèle inadéquat <i>versus</i> modèle adéquat	121
5.4.2	Modèle inadéquat <i>versus</i> une seule modalité	123
5.5	Conclusion	124
6	Modèle inadéquat : Modalités avec des couplages faibles	127
6.1	Introduction	127
6.2	Modèle	128
6.3	Calculs d'estimation	129
6.4	Inadéquation du bruit	131
6.4.1	Corrélation des bruits ignorée	131
6.4.2	Inadéquation quelconque du bruit	138
6.5	Inadéquation du signal	141
6.6	Double inadéquation	145
6.7	Conclusion	146
	Conclusions et perspectives	147
	Annexes	151
A	Modèle adéquat : Calcul d'estimation	153
A.1	Modèle gaussien général	153
A.2	Une modalité scalaire	155
A.3	Deux modalités scalaires faiblement couplées	155
A.4	Deux modalités scalaires fortement couplées	157
A.5	Trois modalités scalaires	157
B	Modèle adéquat : Calcul d'Information	161
B.1	Modèle gaussien général	161
B.2	Une modalité scalaire	162
B.3	Deux modalités scalaires faiblement couplées	162
B.4	Deux modalités scalaires fortement couplées	162
B.5	Trois modalités scalaires	163
C	Modèle inadéquat : Démonstration du théorème 4.5	165
C.1	Lemmes	165
C.2	Démonstration du théorème 4.5	167
C.3	Permutation d'intégrale et de dérivée	168
C.3.1	La permutation en (C.4)	169
C.3.2	La permutation en (C.8)	169
C.3.3	La permutation en (C.6) et (C.12)	170
C.3.4	La permutation en (C.11)	170

D	Modèle inadéquat : Calcul d'estimation	173
D.1	Modèle gaussien général	173
D.2	Deux modalités scalaires faiblement couplées	174
D.3	Deux modalités scalaires fortement couplées	178
E	Calcul matriciel	179
E.1	Déterminant	179
E.2	Inverse	179
E.3	Gradient	179
	Bibliographie	181

Table des figures

Partie I : Modèle adéquat	5
Chapitre 1 : Mesures d'information et d'estimation	7
1.1 Estimateur optimal par projection	12
1.2 Canal gaussien scalaire	16
1.3 Canal gaussien vectoriel	17
Chapitre 2 : Des modalités avec un couplage fort	19
2.1 Deux modalités, avec un couplage fort, vues comme des canaux de communication	20
2.2 Une ou deux modalités de même qualité ?	23
2.3 Une ou deux modalités de qualités différentes ?	25
2.4 Redondance de deux modalités de même qualité	27
2.5 Redondance de deux modalités de différentes qualités	28
2.6 Effet d'une modalité de très faible qualité	30
2.7 Effet des qualités de modalités	31
2.8 Trois modalités avec un couplage fort vues comme des canaux de communication	32
2.9 Valeurs éligibles des corrélations en deux dimensions	34
2.10 Contours des valeurs admissibles des corrélations des bruits en trois dimensions	35
2.11 Une ou trois modalités ?	39
2.12 Deux modalités parmi trois dans le cas de bruits similairement liés	41
2.13 (x_1, x_2) ou (x_1, x_3) ? Intervalle des corrélations	44
2.14 (x_1, x_2) ou (x_1, x_3) ? Information mutuelle	44
2.15 (x_1, x_2) ou (x_1, x_3) ? Intervalle de corrélation	45
2.16 (x_1, x_2) ou (x_2, x_3) ? Information mutuelle	45
2.17 (x_1, x_3) ou (x_2, x_3) ? Intervalle de corrélation	46
2.18 (x_1, x_3) ou (x_2, x_3) ? Information mutuelle	46
2.19 Deux ou trois modalités ?	48
2.20 Redondance de trois modalités	49
2.21 Effet des qualités des modalités sur la redondance	50

Chapitre 3 : Des modalités avec des couplages faibles	53
3.1 Deux modalités avec des couplages faibles vues comme des canaux de communication	54
3.2 Interaction de deux modalités complémentaires	59
3.3 Deux modalités de même qualités - Intervalle de redondance	63
3.4 Deux modalités de même qualités - Information mutuelle	64
3.5 Deux modalités de différentes qualités - Intervalle de redondance	66
3.6 Deux modalités de différentes qualités - Information mutuelle	67
3.7 Deux modalités de qualités très différentes - Intervalle de redondance	68
3.8 Deux modalités de qualités très différentes - Information mutuelle	69
Partie II : Modèle inadéquat	73
Chapitre 4 : Mesures d'information et d'estimation	77
4.1 Estimateur inadéquat par projection	83
4.2 Canal gaussien linéaire	85
4.3 Canal gaussien linéaire sous inadéquation de l'entrée	86
4.4 Canal gaussien linéaire sous inadéquation du bruit	88
4.5 Canal gaussien linéaire sous inadéquation du signal d'entrée et du bruit	91
4.6 Canal gaussien scalaire	99
4.7 D_{kl} et l'excédent de l'EQM sous inadéquation du signal d'entrée	100
4.8 Information mutuelle et l'EQM sous inadéquation du signal d'entrée	101
4.9 D_{kl} et l'excédent de l'EQM sous inadéquation du bruit	103
4.10 Information mutuelle et l'EQM sous inadéquation du bruit	104
4.11 D_{kl} et l'excédent de l'EQM sous double inadéquation	105
4.12 Information mutuelle et l'EQM sous double inadéquation	106
Chapitre 5 : Modalités avec un couplage fort	109
5.1 Inadéquation sur le lien entre les bruits de deux modalités avec un couplage fort	110
5.2 Corrélation des bruits ignorée - Modèle inadéquat versus modèle anticipé	114
5.3 Corrélation des bruits ignorée - Modèle inadéquat versus modèle exact (1)	115
5.4 Corrélation des bruits ignorée - Modèle inadéquat versus modèle exact (2)	117
5.5 Corrélation des bruits ignorée - Modèle inadéquat versus modèle exact (3)	118
5.6 Corrélation des bruits ignorée - L'excédent de l'EQM	119
5.7 Corrélation des bruits ignorée - Modèle inadéquat versus une modalité	120
5.8 Inadéquation quelconque du bruit - Modèle inadéquat versus modèle exact	122
5.9 Inadéquation quelconque du bruit - Modèle inadéquat versus une modalité	124
Chapitre 6 : Des modalités avec des couplages faibles	127
6.1 Inadéquation sur les liens entre deux modalités avec des couplages faibles	128
6.2 Modèle inadéquat versus modèle exact - corrélation des bruits ignorée	132
6.3 Excédent de l'EQM - corrélation des bruits ignorée	134

6.4	Modèle inadéquat versus une modalité (1) - corrélation des bruits ignorée	135
6.5	Modèle inadéquat versus une modalité (2) - corrélation des bruits ignorée	135
6.6	les EQMs de l'estimation de s_2 en fonction de x_2 seule et x_1 seule.	136
6.7	Modèle inadéquat versus une modalité (3) - corrélation des bruits ignorée	137
6.8	Excédent de l'EQM - inadéquation quelconque du bruit	139
6.9	Modèle inadéquat versus une modalité (1) - Inadéquation quelconque du bruit	140
6.10	Modèle inadéquat versus une modalité (2) - Inadéquation quelconque du bruit	140
6.11	Excédent de l'EQM - inadéquation du signal utile (1)	142
6.12	Excédent de l'EQM - inadéquation du signal utile (2)	143
6.13	Modèle inadéquat versus une modalité (1) - Inadéquation du signal utile	144
6.14	Modèle inadéquat versus une modalité (2) - Inadéquation du signal utile	144
6.15	Excédent de l'EQM - double inadéquation	146

Notations

Dans un but de clarté, nous introduisons ici les notations adoptées tout au long de ce manuscrit.

Variables

z	Variable aléatoire scalaire et réalisation
\mathbf{z}	Vecteur aléatoire et réalisation
\mathbf{Z}	Matrice

Opérateurs

$ \cdot $	Déterminant
$Tr\{\cdot\}$	Trace
\cdot^T	Transposé
∇	Gradient

Fonctions

$\text{sgn}(\cdot)$	fonction signe	$\text{sgn}(x) = \begin{cases} -1 & \text{si } x < 0, \\ 0 & \text{si } x = 0, \\ 1 & \text{si } x > 0. \end{cases}$
---------------------	----------------	--

Canal de communication

\mathbf{s} ou s	signal d'intérêt
\mathbf{x}_i ou x_i	Observation, Modalité i
\mathbf{n}_i ou n_i	Bruit additif de la modalité i
\mathbf{H}_i ou α_i	Réponse du $i^{\text{ème}}$ canal de communication

Théorie d'information

$h_{p,p}(\cdot)$	Entropie et Entropie différentielle
$h_{p,q}(\cdot)$	Entropie inadéquate et Entropie différentielle inadéquate
$I_{p,p}(\cdot;\cdot)$	Information mutuelle
$I_{p,q}(\cdot;\cdot)$	Information mutuelle inadéquate
$D_{KL}(\cdot \cdot)$	Divergence de Kullback-Leibler

Théorie d'estimation

$\hat{\mathbf{s}}_p(\cdot), \hat{s}_p(\cdot)$ ou $\hat{s}_{i_p}(\cdot)$	Estimateur optimal de \mathbf{s} , s ou s_i
$\hat{\mathbf{s}}_q(\cdot), \hat{s}_q(\cdot)$ ou $\hat{s}_{i_q}(\cdot)$	Estimateur inadéquat de \mathbf{s} , s ou s_i
$\mathbf{M}_{p,p}(\cdot \cdot)$	Erreur quadratique moyenne d'estimation
$\mathbf{M}_{p,q}(\cdot \cdot)$	Erreur quadratique moyenne d'estimation inadéquate

Statistiques

$p(\cdot)$	Distribution des données dite adéquate
$q(\cdot)$	Distribution du modèle dite inadéquate
$\mathbb{E}_p[\cdot]$	Espérance par rapport à la vraie loi $p(\cdot)$
$\mathbb{E}_q[\cdot]$	Espérance par rapport à la loi inadéquate $q(\cdot)$
ρ_n	Coefficient de corrélation exact des bruits dans le cas de deux modalités
ρ_{ij}	Coefficient de corrélation exact des bruits n_i et n_j dans le cas de trois modalités
$\hat{\rho}_n$	Coefficient de corrélation supposé des bruits
ρ_s	Coefficient de corrélation exact des signaux s_1 et s_2
$\hat{\rho}_s$	Coefficient de corrélation supposé des signaux s_1 et s_2

Intervalles

I_{red}	Intervalle de redondance introduite sur le signal utile par la corrélation des bruits entre x_1 et x_2 .
$I_{une}(s x)$	Intervalle sur lequel une seule modalité x est plus favorable à l'estimation de s que deux modalités.

Modèles avec des signaux fortement couplés (h pour hard coupling)

\mathcal{M}_h^2	Modèle adéquat de deux modalités.
\mathcal{M}_{ho}^2	Modèle adéquat de deux modalités avec des bruits non liés.
$\widetilde{\mathcal{M}}_h^2$	Modèle inadéquat de deux modalités.
\mathcal{M}_h^3	Modèle adéquat de trois modalités.
\mathcal{M}_h^3	Modèle adéquat de trois modalités avec des bruits non liés.

Modèles avec des signaux faiblement couplés (s pour soft coupling)

\mathcal{M}_s^2	Modèle adéquat de deux modalités.
\mathcal{M}_{so}^2	Modèle adéquat de deux modalités avec des bruits non liés.
$\widetilde{\mathcal{M}}_s^2$	Modèle inadéquat de deux modalités.

Introduction

Les phénomènes physiques étudiés en traitement du signal sont souvent très complexes à appréhender en utilisant un seul système de mesure. Grâce aux avancées technologiques, nous disposons de différents types de systèmes de mesures qui utilisent différents principes physiques et permettent ainsi de capter différentes caractéristiques d'un phénomène. L'exploitation des données issues de différents types de systèmes de mesures est connue sous le nom de *multimodalité*. En général, on postule que la complémentarité et la redondance des données multimodales offrent une richesse et fiabilité en information qui conduisent à une meilleure compréhension du phénomène étudié.

L'idée clé de la multimodalité est de permettre aux modalités d'interagir et de s'informer mutuellement. Ce qui est possible grâce à une analyse simultanée (conjointe) des modalités qui exploite les relations entre les données. La multimodalité trouve application dans divers domaines tels que l'audiovisuel, l'imagerie médicale et la télédétection, etc.

Audiovisuel. L'audio et la vidéo sont deux modalités complémentaires qui sont fusionnées dans un grand nombre d'applications. A titre d'exemple, en application à la reconnaissance de la parole et l'interaction homme-machine, quoique le signal audio détienne toute l'information utile, il est plus sensible aux bruits que le signal vidéo. En plus de la robustesse de la vidéo, il s'est avéré que l'information visuelle est très importante à la compréhension de la parole. En effet, la perception humaine exploite en complément de la parole, le mouvement des lèvres, l'expression faciale et la gestuelle [1].

Imagerie médicale. Il existe une pléthore de modalités qui communiquent des informations structurelles ou fonctionnelles sur le cerveau humain. Nous citons à titre d'exemple l'électroencéphalographie (EEG) et l'imagerie par résonance magnétique fonctionnelle (IRMf). Ces deux modalités mesurent l'activité cérébrale en utilisant différents principes physiques. L'EEG mesure le champ électrique induit par l'activité cérébrale alors que l'IRMf détecte le changement de magnétisation entre les niveaux d'oxygène dû à cette même activité. En fusionnant ces deux modalités, complémentaires en terme de résolutions spatiale et temporelle, on espère atteindre une haute résolution spatio-temporelle [2].

Téledétection. On peut extraire différentes types d'informations sur l'environnement en utilisant différents systèmes de mesures. L'information sur la structure et l'élévation sont fournies par le SAR ("Synthetic Aperture Radar") et l'imagerie optique, le LIDAR ("light detection and ranging") fournit de l'information sur la structure et l'imagerie hyperspectrale et multispectrale sont informatives sur la composition des matériaux. Par conséquent, la fusion de ces modalités peut fournir une vue plus complète de la scène observée [3].

Malgré les gains très avantageux et attirants que nous promet la multimodalité, nous trouvons dans la littérature des exemples qui mettent en évidence que la fusion des données ne conduit pas toujours à de meilleures performances. McGurk et McDonald ont montré que la combinaison d'un son "ba" et un mouvement des lèvres "ga" peut générer une perception différente qui est "da". Cet effet est connu comme l'effet McGurk [4]. Un autre exemple en téledétection, pour une compétition annuelle sur la fusion des données, des images optiques et SAR ont été utilisées afin de détecter les changements après inondation dans une scène. Les résultats de la compétition montrent que, pour les méthodes non supervisées, l'image optique seule fournit des meilleurs résultats que l'image SAR. A l'opposé, les images optique et SAR combinées ont conduit à de moins bonnes performances que l'image optique seule [5].

Ces deux exemples montrent qu'il existe des cas où la multimodalité n'est pas bénéfique. Alors, quelles sont les conditions pour atteindre le bénéfice maximal de la multimodalité? Quoiqu'il paraisse normal de mettre en cause la performance de la multimodalité, peu de travaux se sont intéressés à cette question. En effet, la combinaison des données multimodales introduit plusieurs nouveaux défis en raison de l'hétérogénéité des données et de la complexité des systèmes physiques étudiés, etc [6], et la plupart des travaux se sont concentrés sur ces défis. Cependant, étant donné que la performance attendue de la multimodalité n'est pas toujours au rendez-vous, la quantification préalable du gain de la multimodalité est un défi aussi important que la combinaison des modalités.

Parmi, les travaux qui étudient les conditions pour une performance optimale de la multimodalité nous citons [7] et [8]. Dans [7], Bourrier et *al.* étudient le mode d'échantillonnage "optimal" dans le cas de deux modalités. Ils considèrent le problème d'estimation d'une variable latente continue à partir de deux observations bruitées avec des fréquences d'échantillonnage et des variances de bruit différentes. En raison de la difficulté d'établir des règles générales pour atteindre une meilleure performance, les auteurs proposent d'exploiter un modèle stochastique.

Fijalkow et *al.* considèrent dans [8] l'estimation paramétrique à partir de mesures multimodales, éventuellement non-linéaires. Afin, de contourner la complexité de l'analyse conjointe des données multimodales, les auteurs proposent une combinaison linéaire des estimateurs obtenus séparément de chaque modalité. Cependant, cette combinaison dépend du vecteur paramètre inconnu. Les auteurs proposent alors une combinaison sous-optimale et déterminent les conditions pour atteindre une meilleure performance qu'une seule modalité.

Le traitement conjoint des modalités repose sur l'interaction qui permet aux modalités de s'informer mutuellement. Par conséquent, nous croyons que le premier pas pour une combinaison

optimale est de comprendre comment les modalités interagissent. L'interaction des modalités est possible grâce aux liens entre les données. Ces liens reflètent la similarité et l'hétérogénéité d'information entre les modalités. Ils peuvent être modélisés par des dépendances statistiques, par exemple la corrélation ou la covariance, nous parlons dans ce cas de couplages faibles (*soft*) ; ou par des quantités liées ou communes, nous parlons dans ce cas de couplages forts (*hard*) [6]. Dans cette thèse nous proposons d'étudier l'effet des liens sur la multimodalité.

Puisqu'il s'agit du même système physique, les mesures multimodales peuvent être liées à travers le signal d'intérêt mais aussi à travers les bruits. A la différence de [7] et [8], nous supposons aussi une dépendance entre les bruits. Nous examinons, ainsi, l'effet des liens entre les signaux d'intérêt et des liens entre les bruits sur l'interaction des modalités.

Compte tenu de la complexité des systèmes physiques, objets des analyses multimodales, nous supposons aussi que la modélisation des liens est sujette à des erreurs. Ainsi, cette étude consiste en deux parties : la première partie suppose la bonne modélisation des liens, nous disons dans ce cas que le modèle est *adéquat* ; la deuxième partie suppose que les liens sont imparfaitement modélisés, nous disons dans ce cas que le modèle est *inadéquat*.

Dans notre étude nous considérons des modèles généraux et simples. Généraux, pour tirer des conclusions non dépendantes d'une application spécifique. Simples, afin de faciliter les interprétations des résultats et permettre une meilleure compréhension de l'interaction. Les outils que nous avons choisis pour notre étude sont l'information mutuelle et l'erreur quadratique moyenne.

Cette thèse est organisée comme suit :

La partie I est consacrée aux modèles adéquats où les liens entre les modalités sont parfaitement connus.

- Le chapitre 1 passe en revue les mesures fondamentales de la théorie de l'information et de la théorie de l'estimation, ainsi que quelques relations établies entre ces deux théories.
- Le chapitre 2 couvre l'interaction des modalités dont les signaux d'intérêt sont fortement couplés.
- Le chapitre 3 porte sur l'interaction des modalités complémentaires, *i.e* dont les signaux d'intérêt sont faiblement couplés.

La partie II concerne le modèle inadéquat où les liens entre les modalités sont modélisés inexactement.

- Le chapitre 4 adapte les définitions des mesures d'information et d'estimation du chapitre 1 aux modèles inadéquats et représente les nouvelles relations établies entre ces mesures.
- Le chapitre 5 met en exergue l'interaction des modalités sous inadéquation des liens entre les bruits.
- Le chapitre 6 concerne l'interaction des modalités sous une double inadéquation : inadéquation des liens entre les bruits et inadéquation des liens entre les signaux d'intérêt.

Contribution de thèse

Les contributions de cette thèse sont :

1. L'analyse détaillée, à travers des exemples de deux et trois modalités, des différentes situations de l'analyse multimodale (chapitres 2 et 3).
2. L'extension de certaines relations établies, dans un canal gaussien, entre la théorie de l'information et la théorie de l'estimation en considérant les erreurs sur les densités de probabilité du bruit et du signal utile (section 4.3 du chapitre 4).
3. L'analyse de l'impact, sur l'analyse multimodale, des fausses hypothèses sur les liens entre modalités (chapitres 5 et 6).

Partie I

Modèle adéquat

Mesures d'information et d'estimation

Sommaire

1.1	Mesures d'information	8
1.1.1	Entropie	8
1.1.2	Entropie différentielle	9
1.1.3	Entropie relative	10
1.1.4	Information mutuelle	10
1.1.5	Information mutuelle conjointe	12
1.2	Mesures d'estimation	12
1.2.1	Estimateur optimal au sens de l'erreur quadratique moyenne	12
1.2.2	Information de Fisher	14
1.3	Information versus estimation	15
1.3.1	Identité de De Bruijn	15
1.3.2	Information de Fisher et divergence de Kullback-Leibler	16
1.3.3	Information mutuelle et erreur quadratique moyenne	16
1.4	Conclusion	17

Dans notre étude de l'interaction entre les modalités, nous utilisons l'information mutuelle (théorie de l'information) et l'erreur quadratique moyenne (théorie de l'estimation) comme mesures de performance. L'information mutuelle quantifie la dépendance entre le signal d'intérêt et les mesures des différentes modalités. Elle caractérise donc le canal de transmission associé au système de mesure et permet de quantifier l'information maximale pouvant être obtenue sur le signal d'intérêt, c'est-à-dire le bénéfice maximal de la multimodalité. L'erreur quadratique moyenne évalue la précision d'estimation du signal d'intérêt à partir des modalités. Elle définit alors le bénéfice de la multimodalité sur l'estimation.

Dans ce chapitre nous rappelons la définition de ces deux mesures, mais pas uniquement. Nous présentons l'entropie et la divergence de Kullback-Leibler, qui sont liées à l'information mutuelle et constituent ensemble les mesures fondamentales de la théorie d'information. Nous introduisons tout d'abord les mesures d'information discrètes telles qu'elles ont été introduites par Shannon, afin de souligner les différences avec le cas continu, notamment pour l'interprétation

de l'entropie différentielle. Parmi les mesures de la théorie d'estimation, nous présentons l'erreur quadratique moyenne et l'information de Fisher.

La théorie de l'estimation s'intéresse à la manière d'estimer un signal et comment évaluer la précision de cette estimation. La théorie de l'information est concernée, entre autres, par la manière d'envoyer de l'information ainsi que son estimation au niveau du récepteur. Ces deux théories distinctes se rejoignent alors dans l'élimination du bruit. Il n'est donc pas surprenant qu'il existe des relations simples qui les relient. Nous présentons, à la fin de ce chapitre, quelques relations établies entre ces deux théories dans le cadre d'un canal gaussien.

Aux notations classiques des mesures d'estimation et d'informations nous ajoutons l'indice $\cdot_{p,q}$. Le premier indice, p , désigne la vraie densité des données. Le deuxième indice, q , fait référence à la densité supposée des données. L'indice $\cdot_{p,p}$ signifie donc que $q = p$ c'est-à-dire que la densité supposée correspond à la vraie densité. Nous disons dans ce cas que le modèle supposé est adéquat. Cette nouvelle notation permet de faire la différence entre les mesures classiques et les mesures similaires dans le cadre du modèle inadéquat pour lequel $q \neq p$.

1.1 Mesures d'information

La théorie de l'information a été introduite en 1948 par Claude Shannon dans son fameux article "A mathematical theory of communication" [9]. La raison d'être de cette théorie est de résoudre les problématiques de communication telles que la compression et la transmission des données. Depuis, les applications de cette théorie se sont étendues dans d'autres domaines : les statistiques, le traitement de signal, l'informatique, la cryptographie, l'économie, etc. Nous présentons dans ce qui suit les mesures fondamentales de cette théorie : l'entropie, l'information mutuelle et la divergence de Kullback-Leibler.

1.1.1 Entropie

Soit X une variable aléatoire discrète, prenant les valeurs x_i avec la probabilité $p(x_i)$. L'entropie, notée $h_{p,p}(X)$, mesure l'incertitude moyenne de X avant mesure. Autrement dit la quantité d'information moyenne qu'apporte X . Elle est définie par l'espérance de l'incertitude $\log\left(\frac{1}{p(X)}\right)$:

$$h_{p,p}(X) \triangleq \mathbb{E}_p \left[\log \left(\frac{1}{p(X)} \right) \right] = \sum_i p(x_i) \log \left(\frac{1}{p(x_i)} \right). \quad (1.1)$$

Si le logarithme est en base 2, l'entropie est exprimée en *bits*. Dans ce travail, nous considérons le logarithme en base e où l'entropie est exprimée en *nats*. L'entropie est non-négative et est nulle si et seulement si X est certain ($\exists i$ tel que $p(x_i) = 1$). Elle est maximale si tous les événements sont équiprobables $p(x_i) = \frac{1}{k}$, où k est le nombre des valeurs possibles que peut prendre X . Pour toute fonction f inversible, $f(X)$ est dite équivalente à X et présente la même entropie que X , $h_{p,p}(f(X)) = h_{p,p}(X)$.

Soit Y une variable discrète qui dépend de X . Lorsque nous observons Y , nous obtenons de l'information sur X et son incertitude diminue. La quantité d'information sur X non présente

dans Y est quantifiée par l'entropie conditionnelle :

$$h_{p,p}(X|Y) \triangleq \mathbb{E}_p \left[\log \left(\frac{1}{p(X|Y)} \right) \right] = \sum_i \sum_j p(x_i, y_j) \log \left(\frac{1}{p(x_i|y_j)} \right), \quad (1.2)$$

où $p(x_i, y_j)$ est la probabilité conjointe de x_i et y_j , et $p(x_i|y_j) = \frac{p(x_i, y_j)}{p(y_j)}$ est la probabilité conditionnelle de x_i sachant y_j .

Puisque la connaissance de Y ne peut que réduire l'incertitude sur X alors :

$$h_{p,p}(X|Y) \leq h_{p,p}(X), \quad (1.3)$$

avec égalité si et seulement si X et Y sont indépendantes. L'incertitude moyenne sur X et Y est quantifiée par l'entropie conjointe $h_{p,p}(X, Y)$:

$$h_{p,p}(X, Y) \triangleq \mathbb{E}_p \left[\log \left(\frac{1}{p(X, Y)} \right) \right]. \quad (1.4)$$

En utilisant la relation de Bayes nous trouvons que l'entropie et l'entropie conditionnelle sont liées à l'entropie conjointe par les relations suivantes, dites règles de la chaîne :

$$\begin{aligned} h_{p,p}(X, Y) &= h_{p,p}(X) + h_{p,p}(Y|X), \\ &= h_{p,p}(Y) + h_{p,p}(X|Y). \end{aligned} \quad (1.5)$$

A partir de (1.3) et (1.5), on déduit que l'incertitude conjointe de deux variables est inférieure à la somme des incertitudes individuelles : $h_{p,p}(X, Y) \leq h_{p,p}(X) + h_{p,p}(Y)$ avec égalité si et seulement si les variables sont indépendantes.

1.1.2 Entropie différentielle

Soit \mathbf{x} un vecteur aléatoire de densité de probabilité $p(\mathbf{x})$. L'entropie de \mathbf{x} est dite différentielle et est notée $h_{p,p}(\mathbf{x})$. Elle est définie, par analogie au cas discret, par l'espérance de $\log \left(\frac{1}{p(\mathbf{x})} \right)$:

$$h_{p,p}(\mathbf{x}) \triangleq \mathbb{E}_p \left[\log \left(\frac{1}{p(\mathbf{x})} \right) \right] = \int p(\mathbf{x}) \log \left(\frac{1}{p(\mathbf{x})} \right) d\mathbf{x}. \quad (1.6)$$

Contrairement à l'entropie, l'entropie différentielle peut être négative. De plus, pour une fonction f inversible, les entropies différentielles de $f(\mathbf{x})$ et \mathbf{x} sont différentes ; en fait $h_{p,p}(f(\mathbf{x})) = h_{p,p}(\mathbf{x}) + \mathbb{E}_p[\log(|\mathbf{J}|)]$ où $\mathbf{J} = |\nabla_{\mathbf{x}} f(\mathbf{x})|$ est le jacobien de la fonction f . L'entropie différentielle ne peut donc être interprétée comme mesure d'incertitude ou d'information.

Exemple 1.1. L'entropie différentielle d'un vecteur gaussien de dimension k , $\mathbf{x} \sim p(\mathbf{x}) = \mathcal{N}(\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})$, est :

$$h_{p,p}(\mathbf{x}) = \frac{1}{2} \log \left((2\pi e)^k |\Sigma_{\mathbf{x}}| \right), \quad (1.7)$$

où $|\cdot|$ est l'opérateur déterminant. Nous remarquons que l'entropie de \mathbf{x} ne dépend pas de sa valeur moyenne, $\mu_{\mathbf{x}}$, mais uniquement de sa matrice de covariance $\Sigma_{\mathbf{x}}$. Si $|\Sigma_{\mathbf{x}}| < \frac{1}{(2\pi e)^k}$ alors $h_{p,p}(\mathbf{x}) < 0$.

Démonstration. Nous avons, $p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}_{\mathbf{x}}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^T \boldsymbol{\Sigma}_{\mathbf{x}}^{-1}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})\right)$, ainsi :

$$h_{p,p}(\mathbf{x}) = -\mathbb{E}_p [\log(p(\mathbf{x}))], \quad (1.8)$$

$$= \frac{1}{2} \mathbb{E}_p \left[\log\left((2\pi)^k |\boldsymbol{\Sigma}_{\mathbf{x}}|\right) \right] + \frac{1}{2} \mathbb{E}_p \left[\text{Tr} \left\{ (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^T \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \right\} \right], \quad (1.9)$$

$$= \frac{1}{2} \log\left((2\pi)^k |\boldsymbol{\Sigma}_{\mathbf{x}}|\right) + \frac{1}{2} \mathbb{E}_p \left[\text{Tr} \left\{ \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^T \right\} \right], \quad (1.10)$$

$$= \frac{1}{2} \log\left((2\pi)^k |\boldsymbol{\Sigma}_{\mathbf{x}}|\right) + \frac{1}{2} \text{Tr} \left\{ \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} \boldsymbol{\Sigma}_{\mathbf{x}} \right\}, \quad (1.11)$$

$$= \frac{1}{2} \log\left((2\pi)^k |\boldsymbol{\Sigma}_{\mathbf{x}}|\right) + \frac{k}{2}, = \frac{1}{2} \log\left((2\pi e)^k |\boldsymbol{\Sigma}_{\mathbf{x}}|\right). \quad (1.12)$$

où (1.9) s'ensuit étant donné que $(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^T \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})$ est une forme quadratique ainsi $(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^T \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) = \text{Tr} \left\{ (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^T \boldsymbol{\Sigma}_{\mathbf{x}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) \right\}$. (1.10) résulte de l'invariance de la trace à la permutation circulaire. (1.11) est obtenue par permutation de la trace et l'espérance. \square

Les entropies différentielles conditionnelle et conjointe sont définies similairement au cas discret :

$$h_{p,p}(\mathbf{x}|\mathbf{y}) \triangleq \mathbb{E}_p \left[\log \left(\frac{1}{p(\mathbf{x}|\mathbf{y})} \right) \right] = \int \int p(\mathbf{x}, \mathbf{y}) \log \left(\frac{1}{p(\mathbf{x}|\mathbf{y})} \right) d\mathbf{x}d\mathbf{y}, \quad (1.13)$$

$$h_{p,p}(\mathbf{x}, \mathbf{y}) \triangleq \mathbb{E}_p \left[\log \left(\frac{1}{p(\mathbf{x}, \mathbf{y})} \right) \right] = h_{p,p}(\mathbf{x}) + h_{p,p}(\mathbf{y}|\mathbf{x}), \quad (1.14)$$

où $p(\mathbf{x}, \mathbf{y})$ est la loi jointe de \mathbf{x} et \mathbf{y} , et $p(\mathbf{x}|\mathbf{y}) = \int p(\mathbf{x}, \mathbf{y}) d\mathbf{y}$ est la loi conditionnelle de \mathbf{x} sachant \mathbf{y} . Les inégalités présentées pour le cas discret s'appliquent aussi au cas continu, à savoir :

$$h_{p,p}(\mathbf{x}|\mathbf{y}) \leq h_{p,p}(\mathbf{x}), \quad (1.15)$$

$$h_{p,p}(\mathbf{x}, \mathbf{y}) \leq h_{p,p}(\mathbf{x}) + h_{p,p}(\mathbf{y}). \quad (1.16)$$

1.1.3 Entropie relative

L'entropie relative ou divergence de Kullback-Leibler a été introduite par Solomon Kullback et Richard Leibler en 1951 [10] et a été étudiée en détail dans [11]. Cette mesure quantifie la dissimilarité entre deux distributions de probabilité. Soient p et q deux lois de probabilité, la divergence de Kullback-Leibler est notée et définie par :

$$D_{KL}(p(\mathbf{x})||q(\mathbf{x})) \triangleq \mathbb{E}_p \left[\log \left(\frac{p(\mathbf{x})}{q(\mathbf{x})} \right) \right]. \quad (1.17)$$

La divergence de Kullback-Leibler n'est pas une distance puisqu'elle est asymétrique en p et q , $D_{KL}(p(\mathbf{x})||q(\mathbf{x})) \neq D_{KL}(q(\mathbf{x})||p(\mathbf{x}))$, et elle ne vérifie pas l'inégalité triangulaire. Elle est non-négative et est nulle si et seulement si $p = q$. Elle est finie si et seulement si le support de $p(\cdot)$ est inclus dans le support de $q(\cdot)$. C'est-à-dire que s'il existe \mathbf{x} tel que $p(\mathbf{x}) > 0$ tandis que $q(\mathbf{x}) = 0$, alors $D_{KL}(p(\mathbf{x})||q(\mathbf{x})) = +\infty$.

1.1.4 Information mutuelle

L'information mutuelle est une mesure de dépendance statistique ou d'information partagée entre deux vecteurs aléatoires. L'information mutuelle entre \mathbf{x} et \mathbf{y} est la divergence de Kullback-

Leibler entre la loi jointe, $p(\mathbf{x}, \mathbf{y})$, et le produit des marginales, $p(\mathbf{x})p(\mathbf{y})$:

$$I_{p,p}(\mathbf{x}; \mathbf{y}) = D_{KL}(p(\mathbf{x}, \mathbf{y}) || p(\mathbf{x})p(\mathbf{y})) = \mathbb{E}_p \left[\log \left(\frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} \right) \right]. \quad (1.18)$$

L'information mutuelle est non-négative (divergence de Kullback-Leibler) et peut être infinie. Elle est symétrique en \mathbf{x} et \mathbf{y} . $I_{p,p}(\mathbf{x}; \mathbf{y}) = 0$ si et seulement si $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y})$, c'est-à-dire si \mathbf{x} et \mathbf{y} sont indépendants. Elle peut être exprimée comme une différence d'entropies (entropies différentielles pour les vecteurs continus) :

$$I_{p,p}(\mathbf{x}; \mathbf{y}) = h_{p,p}(\mathbf{x}) - h_{p,p}(\mathbf{x}|\mathbf{y}) = h_{p,p}(\mathbf{y}) - h_{p,p}(\mathbf{y}|\mathbf{x}). \quad (1.19)$$

Ainsi, l'information mutuelle est aussi la diminution en incertitude sur \mathbf{x} lorsque \mathbf{y} est observé ou inversement. Maintenant si nous désirons quantifier l'information partagée entre \mathbf{x} et \mathbf{y} , et non présente dans un troisième vecteur aléatoire \mathbf{z} (c'est-à-dire qui n'est pas apportée par l'observation de \mathbf{z}), il faut utiliser l'information mutuelle conditionnelle. C'est encore une divergence de Kullback-Leibler entre la loi jointe conditionnelle à \mathbf{z} , $p(\mathbf{x}, \mathbf{y}|\mathbf{z})$, et le produit, $p(\mathbf{x}|\mathbf{z})p(\mathbf{y}|\mathbf{z})$:

$$I_{p,p}(\mathbf{x}; \mathbf{y}|\mathbf{z}) = D_{KL}(p(\mathbf{x}, \mathbf{y}|\mathbf{z}) || p(\mathbf{x}|\mathbf{z})p(\mathbf{y}|\mathbf{z})), \quad (1.20)$$

$$= h_{p,p}(\mathbf{x}|\mathbf{z}) - h_{p,p}(\mathbf{x}|\mathbf{y}, \mathbf{z}), \quad (1.21)$$

$$= h_{p,p}(\mathbf{y}|\mathbf{z}) - h_{p,p}(\mathbf{y}|\mathbf{x}, \mathbf{z}). \quad (1.22)$$

Si \mathbf{x} et \mathbf{y} sont indépendants conditionnellement à \mathbf{z} , *i.e* $p(\mathbf{x}, \mathbf{y}|\mathbf{z}) = p(\mathbf{x}|\mathbf{z})p(\mathbf{y}|\mathbf{z})$, alors $I_{p,p}(\mathbf{x}; \mathbf{y}|\mathbf{z}) = 0$. C'est-à-dire que indépendamment de \mathbf{z} , \mathbf{y} et \mathbf{x} ne partagent aucune information. Dans ce cas, \mathbf{x} , \mathbf{z} et \mathbf{y} forment dans ce sens une chaîne de Markov :

$$\mathbf{x} \longrightarrow \mathbf{z} \longrightarrow \mathbf{y}$$

Une chaîne de Markov est, en général, une suite de variables aléatoires qui représente l'état d'un système aléatoire. Cette chaîne est caractérisée par la propriété de Markov selon laquelle l'état futur, \mathbf{y} , est prédictible uniquement à partir de l'état présent \mathbf{z} . Autrement dit, l'état futur du système dépend du passé uniquement à travers son état présent, $p(\mathbf{y}|\mathbf{x}, \mathbf{z}) = p(\mathbf{y}|\mathbf{z})$. Ceci revient à dire d'un point de vue informationnel, que \mathbf{z} contient toute l'information qu'on peut tirer sur \mathbf{x} . Ainsi, $I_{p,p}(\mathbf{x}; \mathbf{z}) \geq I_{p,p}(\mathbf{x}; \mathbf{y})$. En fait :

$$I_{p,p}(\mathbf{x}; \mathbf{z}) = I_{p,p}(\mathbf{x}; \mathbf{y}) + I_{p,p}(\mathbf{x}; \mathbf{z}|\mathbf{y}). \quad (1.23)$$

On pourrait croire que la connaissance de \mathbf{z} diminuera l'information que partagent \mathbf{x} et \mathbf{y} , c'est-à-dire que $I_{p,p}(\mathbf{x}; \mathbf{y}|\mathbf{z}) \leq I_{p,p}(\mathbf{x}; \mathbf{y})$, alors que c'est pas toujours vrai. Il se peut que la dépendance entre deux variables aléatoires \mathbf{x} et \mathbf{y} , augmente par l'observation d'une troisième variable aléatoire \mathbf{z} . Par exemple, si \mathbf{x} et \mathbf{y} sont indépendants $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x})p(\mathbf{y})$ alors $I_{p,p}(\mathbf{x}; \mathbf{y}) = 0$. C'est-à-dire que \mathbf{x} et \mathbf{y} ne partagent aucune information. Alors que s'il existe \mathbf{z} dont \mathbf{x} et \mathbf{y} dépendent conditionnellement, $p(\mathbf{x}, \mathbf{y}|\mathbf{z}) \neq p(\mathbf{y}|\mathbf{z})p(\mathbf{x}|\mathbf{z})$ alors $I_{p,p}(\mathbf{x}; \mathbf{y}|\mathbf{z}) > 0$. C'est le cas par exemple si $\mathbf{z} = \mathbf{x} + \mathbf{y}$. Ainsi, dans ce cas, $I_{p,p}(\mathbf{x}; \mathbf{y}|\mathbf{z}) > I_{p,p}(\mathbf{x}; \mathbf{y})$.

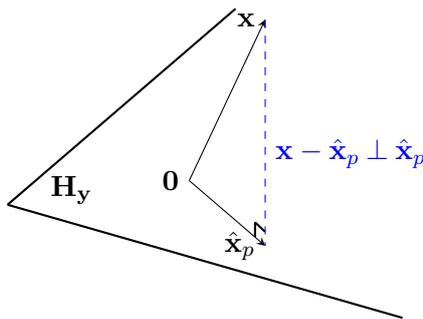


FIGURE 1.1 – Estimateur optimal par projection

1.1.5 Information mutuelle conjointe

Soient \mathbf{x} , \mathbf{y} et \mathbf{z} des vecteurs aléatoires de lois de probabilité $p(\mathbf{x})$, $p(\mathbf{y})$ et $p(\mathbf{z})$. L'information qu'apportent conjointement \mathbf{y} et \mathbf{z} sur \mathbf{x} est définie, de manière similaire à l'information mutuelle, comme la divergence de Kullback-Leibler entre la loi jointe $p(\mathbf{x}, \mathbf{y}, \mathbf{z})$ et le produit $p(\mathbf{x})p(\mathbf{y}, \mathbf{z})$:

$$I_{p,p}(\mathbf{x}; (\mathbf{y}, \mathbf{z})) = D_{KL}(p(\mathbf{x}, \mathbf{y}, \mathbf{z}) || p(\mathbf{x})p(\mathbf{y}, \mathbf{z})). \quad (1.24)$$

Cette quantité est symétrique en \mathbf{x} et (\mathbf{y}, \mathbf{z}) . Néanmoins, elle n'est pas symétrique en \mathbf{x} et \mathbf{y} ou en \mathbf{x} et \mathbf{z} . L'information mutuelle conjointe est liée à l'information mutuelle et l'information mutuelle conditionnelle par les relations :

$$\begin{aligned} I_{p,p}(\mathbf{x}; (\mathbf{y}, \mathbf{z})) &= I_{p,p}(\mathbf{x}; \mathbf{y}) + I_{p,p}(\mathbf{x}; \mathbf{z}|\mathbf{y}), \\ &= I_{p,p}(\mathbf{x}; \mathbf{z}) + I_{p,p}(\mathbf{x}; \mathbf{y}|\mathbf{z}). \end{aligned} \quad (1.25)$$

1.2 Mesures d'estimation

La théorie de l'estimation propose des méthodes pour estimer un paramètre ou un signal \mathbf{x} à partir d'une observation \mathbf{y} . L'estimation est basée sur la loi de probabilité de \mathbf{x} sachant \mathbf{y} , $p(\mathbf{x}|\mathbf{y})$, dite loi a posteriori. Cette loi peut être déterminée grâce à la relation de Bayes :

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})}{\int p(\mathbf{x})p(\mathbf{y}|\mathbf{x})d\mathbf{x}}. \quad (1.26)$$

La relation de Bayes relie la loi a posteriori, $p(\mathbf{x}|\mathbf{y})$, à $p(\mathbf{y}|\mathbf{x})$ et $p(\mathbf{x})$, où $p(\mathbf{x})$ est la loi a priori sur \mathbf{x} , qui résume toute l'information que nous avons sur le signal d'intérêt. Par conséquent, nous avons besoin de déterminer $p(\mathbf{y}|\mathbf{x})$ et $p(\mathbf{x})$ et ensuite appliquer la loi de Bayes afin de mettre à jour l'information que nous avons sur \mathbf{x} par l'information présente dans \mathbf{y} . Finalement, nous associons à l'estimateur de \mathbf{x} , noté $\hat{\mathbf{x}}_p$ et basé sur $p(\mathbf{y}|\mathbf{x})$, un critère (ou fonction de coût) mesurant la qualité de l'estimateur.

1.2.1 Estimateur optimal au sens de l'erreur quadratique moyenne

L'estimateur au sens de l'erreur quadratique moyenne (EQM) est l'un des estimateurs les plus répandus en raison de sa simplicité. Cet estimateur est obtenu par minimisation de la trace du

coût quadratique moyen :

$$\mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y}) = \mathbb{E}_p[(\mathbf{x} - \hat{\mathbf{x}}_p)(\mathbf{x} - \hat{\mathbf{x}}_p)^T]. \quad (1.27)$$

où :

$$\text{Tr} \{ \mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y}) \} = \mathbb{E}_p[(\mathbf{x} - \hat{\mathbf{x}}_p)^T (\mathbf{x} - \hat{\mathbf{x}}_p)] \quad (1.28)$$

Démonstration.

$$\text{Tr} \{ \mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y}) \} = \text{Tr} \{ \mathbb{E}_p[(\mathbf{x} - \hat{\mathbf{x}}_p)(\mathbf{x} - \hat{\mathbf{x}}_p)^T] \}, \quad (1.29)$$

$$= \mathbb{E}_p[\text{Tr} \{ (\mathbf{x} - \hat{\mathbf{x}}_p)(\mathbf{x} - \hat{\mathbf{x}}_p)^T \}], \quad (1.30)$$

$$= \mathbb{E}_p[\text{Tr} \{ (\mathbf{x} - \hat{\mathbf{x}}_p)^T (\mathbf{x} - \hat{\mathbf{x}}_p) \}], \quad (1.31)$$

$$= \mathbb{E}_p[(\mathbf{x} - \hat{\mathbf{x}}_p)^T (\mathbf{x} - \hat{\mathbf{x}}_p)]. \quad (1.32)$$

où (1.30) est obtenue par permutation de la trace et l'espérance puisque la trace est linéaire. (1.31) résulte de l'invariance de la trace à la permutation circulaire. Finalement, (1.32) s'ensuit car $(\mathbf{x} - \mu_{\mathbf{x}})^T \Sigma_{\mathbf{x}}^{-1} (\mathbf{x} - \mu_{\mathbf{x}})$ est une forme quadratique. \square

L'erreur d'estimation $\mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y})$ n'est pas une fonction des variables aléatoires \mathbf{x} et \mathbf{y} : la notation choisie signifie que c'est l'erreur d'estimation de \mathbf{x} en fonction de \mathbf{y} . L'estimateur au sens de l'erreur quadratique moyenne minimale, quant à lui, est une fonction de l'observation \mathbf{y} .

Soit $H_{\mathbf{y}}$ le sous-espace des observations muni du produit scalaire $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbb{E}[\mathbf{u}^T \mathbf{v}]$. Si $\mathbf{x} \in H_{\mathbf{y}}$, alors \mathbf{x} peut être déterminé par l'observation de \mathbf{y} , et une estimation parfaite avec une erreur d'estimation nulle est possible. Si $\mathbf{x} \notin H_{\mathbf{y}}$, alors la trace de l'EQM $\mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y})$ est minimisée par une projection orthogonale de \mathbf{x} sur $H_{\mathbf{y}}$ (voir figure 1.1). Par conséquent, $\mathbf{x} - \hat{\mathbf{x}}_p$ est orthogonale à $H_{\mathbf{y}}$, donc à tout estimateur basé sur \mathbf{y} . C'est le principe d'orthogonalité :

$$\mathbb{E}_p[(\mathbf{x} - \hat{\mathbf{x}}_p)f(\mathbf{y})^T] = 0 \quad \text{pour toute fonction } f(.). \quad (1.33)$$

Par conséquent $\mathbb{E}_p[(\mathbf{x} - \hat{\mathbf{x}}_p)\hat{\mathbf{x}}_p^T] = 0$ et l'erreur $\mathbf{M}_{p,p}$ peut s'écrire comme suit :

$$\mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y}) = \mathbb{E}_p[(\mathbf{x} - \hat{\mathbf{x}}_p)\mathbf{x}^T]. \quad (1.34)$$

Théorème 1.1: (Estimateur optimal [12, ch.2])

Soit $\mathbf{x} \sim p(\mathbf{x})$ le vecteur aléatoire que nous voulons estimer à partir de l'observation $\mathbf{y} \sim p(\mathbf{y})$. L'estimateur de \mathbf{x} en fonction de \mathbf{y} qui minimise la trace de l'EQM $\mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y})$, est la moyenne de la loi a posteriori $p(\mathbf{x}|\mathbf{y})$:

$$\underset{\hat{\mathbf{x}}_p}{\text{argmin}} \mathbb{E}_p[(\mathbf{x} - \hat{\mathbf{x}}_p)^T (\mathbf{x} - \hat{\mathbf{x}}_p)] = \mathbb{E}_p[\mathbf{x}|\mathbf{y}], \quad (1.35)$$

et l'erreur quadratique moyenne associée à cet estimateur est donnée par l'espérance par rapport à $p(\mathbf{y})$ de la matrice de covariance de \mathbf{x} sachant \mathbf{y} , $\Sigma_{\mathbf{x}|\mathbf{y}}$:

$$\mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y}) = \mathbb{E}_p[\Sigma_{\mathbf{x}|\mathbf{y}}] = \int \Sigma_{\mathbf{x}|\mathbf{y}} p(\mathbf{y}) d\mathbf{y}. \quad (1.36)$$

Démonstration. La démonstration de ce théorème peut être trouvée dans plusieurs livres d'estimation, notamment [12]. Néanmoins, nous la présentons ici.

Afin de déterminer l'estimateur qui minimise $\text{Tr}\{\mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y})\}$, nous fixons $\hat{\mathbf{x}}_p$ et nous mettons à zéro la dérivée de $\mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y})$ par rapport à $\hat{\mathbf{x}}_p$:

$$\begin{aligned} \frac{\partial}{\partial \hat{\mathbf{x}}_p} \text{Tr}\{\mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y})\} = 0 &\Rightarrow \frac{\partial}{\partial \hat{\mathbf{x}}_p} \int (\mathbf{x} - \hat{\mathbf{x}}_p)^T (\mathbf{x} - \hat{\mathbf{x}}_p) p(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} = 0 \\ &\Rightarrow \int -2(\mathbf{x} - \hat{\mathbf{x}}_p) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \int p(\mathbf{y}) d\mathbf{y} = 0 \\ &\Rightarrow \hat{\mathbf{x}}_p = \int \mathbf{x} p(\mathbf{x}|\mathbf{y}) d\mathbf{x} = \mathbb{E}_p[\mathbf{x}|\mathbf{y}] \end{aligned}$$

Calculons maintenant la dérivée seconde de $\text{Tr}\{\mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y})\}$,

$$\frac{\partial^2}{\partial \hat{\mathbf{x}}_p \hat{\mathbf{x}}_p^T} \text{Tr}\{\mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y})\} = \frac{\partial}{\partial \hat{\mathbf{x}}_p^T} \int -2(\mathbf{x} - \hat{\mathbf{x}}_p) p(\mathbf{x}|\mathbf{y}) d\mathbf{x} \int p(\mathbf{x}) d\mathbf{x} = 2 \quad (1.37)$$

Puisque la dérivée seconde de $\text{Tr}\{\mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y})\}$ est positive, $\mathbb{E}_p[\mathbf{x}|\mathbf{y}]$ effectivement minimise $\text{Tr}\{\mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y})\}$.

Démontrons maintenant que $\mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y}) = \mathbb{E}_p[\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}}]$. D'après le principe d'orthogonalité, $\mathbb{E}_p[\mathbf{x}^T \mathbb{E}_p[\mathbf{x}|\mathbf{y}]] = \mathbb{E}_p[\mathbb{E}_p[\mathbf{x}|\mathbf{y}]^T \mathbb{E}_p[\mathbf{x}|\mathbf{y}]]$, ainsi

$$\begin{aligned} \mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y}) &= \mathbb{E}_p[(\mathbf{x} - \mathbb{E}_p[\mathbf{x}|\mathbf{y}])\mathbf{x}^T] \\ &= \int \int \mathbf{x}\mathbf{x}^T p(\mathbf{x}|\mathbf{y}) d\mathbf{x} p(\mathbf{y}) d\mathbf{y} - \mathbb{E}_p[\mathbb{E}_p[\mathbf{x}|\mathbf{y}]\mathbb{E}_p[\mathbf{x}|\mathbf{y}]^T] \\ &= \int \left(\int \mathbf{x}\mathbf{x}^T p(\mathbf{x}|\mathbf{y}) d\mathbf{x} - \mathbb{E}_p[\mathbf{x}|\mathbf{y}]\mathbb{E}_p[\mathbf{x}|\mathbf{y}]^T \right) p(\mathbf{x}) d\mathbf{x} \\ &= \mathbb{E}_p[\boldsymbol{\Sigma}_{\mathbf{x}|\mathbf{y}}] \end{aligned}$$

□

1.2.2 Information de Fisher

1.2.2.1 Paramètre déterministe

L'information de Fisher est d'une importance primordiale pour l'estimation paramétrique, puisqu'elle définit la borne inférieure des variances des estimateurs non biaisés, connue comme la borne Cramér-Rao.

Soit \mathbf{x} un vecteur aléatoire paramétré par θ , $\mathbf{x} \sim p(\mathbf{x}; \theta)$. L'information de Fisher de \mathbf{x} est la matrice de covariance de la fonction score : $\nabla_{\theta} \log(p(\mathbf{x}; \theta))$. La moyenne de la fonction score vaut $\mathbf{0}$, $\mathbb{E}[\nabla_{\theta} \log(p(\mathbf{x}; \theta))] = \mathbf{0}$, [13] ainsi l'information de Fisher, $\mathbf{J}_{p,p}(\mathbf{x}; \theta)$, est égale à

$$\mathbf{J}_{p,p}(\mathbf{x}; \theta) = \mathbb{E}_p[\nabla_{\theta} \log(p(\mathbf{x}; \theta)) \nabla_{\theta} \log(p(\mathbf{x}; \theta))^T]. \quad (1.38)$$

1.2.2.2 Paramètre aléatoire

Pour un paramètre aléatoire θ , l'information de Fisher consiste en deux termes, l'information apportée par les observations et l'information a priori sur θ [12] :

$$\mathbf{J}_{p,p}(\mathbf{x}; \theta) = \mathbb{E}_p[\nabla_{\theta} \log(p(\mathbf{x}; \theta)) \nabla_{\theta} \log(p(\mathbf{x}; \theta))^T] + \mathbb{E}_p[\nabla_{\theta} \log(p(\theta)) \nabla_{\theta} \log(p(\theta))^T]. \quad (1.39)$$

1.2.2.3 Information de Fisher non paramétrique

L'information de Fisher pour un vecteur aléatoire \mathbf{x} est définie comme la variance de la fonction score, $\nabla_{\mathbf{x}} \log(p(\mathbf{x}))$:

$$\mathbf{J}_{p,p}(\mathbf{x}) = \mathbb{E}_p[\nabla_{\mathbf{x}} \log(p(\mathbf{x})) \nabla_{\mathbf{x}} \log(p(\mathbf{x}))^T]. \quad (1.40)$$

1.3 Information versus estimation

Plusieurs relations ont été établies entre les mesures d'information et les mesures d'estimation notamment dans les canaux gaussiens. La raison d'être de ces relations est le rapport de vraisemblance, sur lequel sont souvent basées les définitions de ces mesures [14]. Les mesures d'information sont définies comme des espérances des rapports de vraisemblance alors que les mesures d'estimation sont des optimisateurs de certains rapports de vraisemblance.

Des plus fameuses relations nous citons l'identité de De Bruijn [15] et l'I-MMSE [16]. L'identité de De Bruijn relie l'entropie à l'information de Fisher. L'I-MMSE relie l'information mutuelle à l'erreur quadratique moyenne. Ces relations trouvent application principalement dans les démonstrations dites réciproques du codage. Comme elles permettent d'étendre les résultats d'une théorie à l'autre.

1.3.1 Identité de De Bruijn

L'identité de De Bruijn est considérée comme la première relation établie entre la théorie de l'information et de l'estimation. Elle date de l'année 1959. La première démonstration de l'inégalité de la puissance entropique (Entropy power inequality) est basée sur cette identité, d'où sa renommée. L'inégalité de la puissance entropique donne la limite inférieure de la capacité d'un canal gaussien additif [15].

L'identité de De Bruijn, initialement établie dans le cas scalaire, relie l'entropie (paragraphe 1.1.2) et l'information de Fisher non paramétrique (paragraphe 1.2.2) dans un canal gaussien. Soit x une variable aléatoire quelconque de variance finie et soit y une variable gaussienne centrée réduite $y \sim \mathcal{N}(0, 1)$ indépendante de x , alors pour $t \geq 0$:

$$\frac{\partial}{\partial t} h_{p,p}(x + \sqrt{t}y) = \frac{1}{2} \mathbf{J}_{p,p}(x + \sqrt{t}y). \quad (1.41)$$

Dans la littérature, plusieurs généralisations de cette identité ont été proposées : pour des variables vectorielles [17], pour un bruit non centré-réduit [18, 19], pour un canal non-gaussien [20], *etc.* Dans le cas vectoriel, l'identité De Bruijn peut être exprimée comme suit [16] :

$$\nabla_{\mathbf{T}} h_{p,p}(\mathbf{x} + \mathbf{T}^{\frac{1}{2}} \mathbf{y}) = \mathbf{J}_{p,p}(\mathbf{x} + \mathbf{T}^{\frac{1}{2}} \mathbf{y}) \mathbf{T}^{\frac{1}{2}} \boldsymbol{\Sigma} \mathbf{T}^{-\frac{1}{2}}, \quad (1.42)$$

où $h_{p,p}(\cdot)$ est l'entropie (1.6) et $\mathbf{J}_{p,p}(\cdot)$ est l'information de Fisher non paramétrique (1.40). Avec, $\mathbf{T}^{\frac{1}{2}}$ une matrice inversible, \mathbf{x} un vecteur aléatoire et $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$ indépendant de \mathbf{x} .

1.3.2 Information de Fisher et divergence de Kullback-Leibler

L'information de Fisher paramétrique est liée à la divergence de Kullback-Leibler par [11, p.26] :

$$\mathbf{J}_{p,p}(\mathbf{x}; \theta) = \nabla_{\hat{\theta}} D_{KL}(p(\mathbf{x}; \theta) || p(\mathbf{x}; \hat{\theta})) \Big|_{\hat{\theta}=\theta}, \quad (1.43)$$

où $\hat{\theta}$ est l'estimée de θ . Par un développement à l'ordre deux de la divergence, on montre que :

$$D_{KL}(p(\mathbf{x}; \theta) || p(\mathbf{x}; \hat{\theta})) = \frac{(\hat{\theta} - \theta)^2}{2} \mathbf{J}_{p,p}(\mathbf{x}; \theta) + o(\hat{\theta} - \theta)^2. \quad (1.44)$$

Ainsi la divergence de Kullback-Leibler est une parabole en fonction de θ , dont l'information de Fisher contrôle la courbure. A $D_{KL}(p(\mathbf{x}; \theta) || p(\mathbf{x}; \hat{\theta}))$ donnée, pour une grande information de Fisher, la courbe est plus piquée et l'estimée $\hat{\theta}$ est proche de la vraie valeur θ . Inversement, pour une faible information de Fisher, la courbe devient plate et l'estimée est très loin de la vraie valeur.

1.3.3 Information mutuelle et erreur quadratique moyenne

Une variante de l'identité de De Bruijn a été récemment établie par Guo *et al.* [16]. Elle est connue sous le nom I-MMSE, où le "I" désigne l'information mutuelle et "MMSE" représente l'erreur quadratique moyenne minimale. Cette relation énonce que le gradient de l'information mutuelle entre l'entrée et la sortie d'un canal gaussien est directement liée à l'erreur quadratique moyenne. Soit le canal gaussien scalaire de la figure 1.2, où $\alpha \in \mathbb{R}^{+*}$ représente le rapport signal à bruit et $s \in \mathbb{R}$ est une variable aléatoire quelconque indépendante de $n \sim \mathcal{N}(0, 1)$.

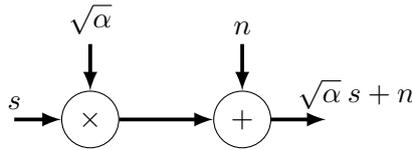


FIGURE 1.2 – Canal gaussien scalaire

La dérivée de l'information mutuelle $I_{p,p}(s; \sqrt{\alpha}s + n)$ entre l'entrée, s , et la sortie du canal, $\sqrt{\alpha}s + n$, par rapport à α est liée à l'erreur d'estimation $\mathbf{M}_{p,p}(x|\sqrt{\alpha}s + y)$ par la relation :

$$\frac{\partial}{\partial \alpha} I_{p,p}(s; \sqrt{\alpha}s + n) = \frac{1}{2} \mathbf{M}_{p,p}(s|\sqrt{\alpha}s + n). \quad (1.45)$$

La relation (1.45) montre que l'augmentation du rapport signal à bruit, α , induit une augmentation de l'information mutuelle. Le taux d'augmentation de l'information mutuelle est égale à la moitié de l'erreur quadratique moyenne minimale de l'estimateur optimal.

Cette relation est généralisée au cas vectoriel dans [21]. Soit le canal gaussien vectoriel :

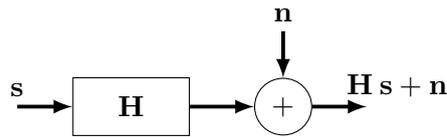


FIGURE 1.3 – Canal gaussien vectoriel

où $\mathbf{H} \in \mathbb{R}^{k \times l}$ est la matrice du canal, $\mathbf{s} \in \mathbb{R}^l$ est un vecteur aléatoire quelconque indépendant du bruit additif gaussien $\mathbf{n} \in \mathcal{N}(0, \Sigma_{\mathbf{n}})$. Palomar et Verdú ont démontré que :

$$\nabla_{\mathbf{H}} I_{p,p}(\mathbf{s}; \mathbf{H}\mathbf{s} + \mathbf{n}) = \Sigma_{\mathbf{n}}^{-1} \mathbf{H} \mathbf{M}_{p,p}(\mathbf{s} | \mathbf{H}\mathbf{s} + \mathbf{n}). \quad (1.46)$$

Cette généralisation donne la sensibilité de l'information mutuelle par rapport à la matrice du canal \mathbf{H} en fonction de l'erreur quadratique moyenne. Cette relation a été étendue à d'autres modèles et canaux : non-gaussiens [16], poissoniens [22–24], de Lévy [25], binomiales et binomiales négatifs [26–28].

La relation I-MMSE, offre des démonstrations élégantes et simples pour certaines démonstrations déjà existantes notamment pour l'inégalité de la puissance entropique. De plus, elle a permis d'établir des bornes pour certains canaux multi-utilisateurs [14].

1.4 Conclusion

Dans ce chapitre nous avons passé en revue les mesures fondamentales de la théorie de l'information. Notamment, l'entropie, l'information mutuelle et la divergence de Kullback-Leiber. A l'origine, ces mesures ont été définies pour des vecteurs aléatoires discrets. Au passage au cas continu, seule l'entropie perd certaines de ces propriétés. Parmi les mesures de la théorie de l'estimation, nous avons présenté l'erreur quadratique moyenne et l'information de Fisher. Finalement, la dernière section était consacrée à trois relations qui relient les mesures d'information aux mesures d'estimation.

Au moyen de l'information mutuelle et de l'erreur quadratique moyenne nous étudierons dans les prochains chapitres comment les modalités interagissent entre elles. Bien que ces deux mesures sont fortement liées dans les canaux gaussiens, l'information mutuelle est plus riche, puisqu'elle permet de détecter tout types de dépendances, mêmes non-linéaires, entre deux vecteurs aléatoires. De plus, l'information mutuelle dépend seulement des densités de probabilités. Ainsi, elle peut être utilisée dans différentes applications.

Modalités avec un couplage fort

Sommaire

2.1	Introduction	19
2.2	Deux modalités	20
2.2.1	Modèle	20
2.2.2	Calculs d'estimation et d'information	21
2.2.3	Une ou deux modalités ?	22
2.2.4	Effet de la corrélation des bruits et des rapports signal à bruit	25
2.3	Trois modalités	31
2.3.1	Modèle	32
2.3.2	Valeurs éligibles des corrélations	33
2.3.3	Calculs d'estimation et d'information	36
2.3.4	Une ou trois modalités ?	37
2.3.5	Deux modalités parmi trois	39
2.3.6	Deux ou trois modalités ?	46
2.3.7	Effet des corrélations des bruits	48
2.4	Conclusion	51

2.1 Introduction

La phrase d'Aristote "Le tout est plus grand que la somme de ses parties" traduit le principe de la multimodalité ou plutôt ce qui est espéré en utilisant des mesures multimodales. L'interprétation de cette citation est que l'analyse conjointe des données issues de différentes modalités (systèmes physiques) est meilleure que l'analyse séparée. Ce postulat s'appuie sur l'idée que l'analyse conjointe, contrairement à l'analyse séparée, permet l'interaction entre les données et révèle donc plus d'information. Cependant, en pratique cette amélioration attendue n'est pas toujours atteinte. Ainsi, les défis que soulève la multimodalité ne concernent pas seulement la conception des algorithmes de traitement conjoint des modalités, mais aussi la quantification préalable du bénéfice de la multimodalité et la détermination des conditions qui garantissent le gain maximal.

La première étape vers l'établissement des règles pour une meilleure performance de l'analyse conjointe des mesures multimodales, consiste à comprendre comment ces mesures interagissent. L'interaction entre les modalités n'est possible qu'en exploitant les liens entre les modalités. Des

liens peuvent exister entre les signaux d'intérêt mais aussi entre les bruits. Dans ce chapitre nous considérons un lien fort entre les signaux d'intérêt. C'est-à-dire que les signaux d'intérêt sont communs à toutes les modalités. Ce choix permet de mettre l'accent sur l'influence des liens entre les bruits qui sont souvent ignorés en pratique.

La multimodalité est largement utilisée dans différents domaines où les systèmes physiques sont complexes. Pour citer quelques exemples : télédétection, imagerie médicale, interaction homme machine, etc. Par conséquent, les méthodes multimodales dépendent de l'application visée. Pour tirer des conclusions qui ne dépendent pas de l'application et pour une meilleure compréhension nous considérons un modèle général très simple, qui a l'avantage d'être facile à interpréter et analyser.

Ce chapitre est découpé en trois parties. Nous considérons le cas de deux modalités dans la section 2.2 et celui de trois modalités dans la section 2.3. A partir des notions d'information et d'estimation, nous déterminerons le nombre optimal des modalités à considérer et nous étudions l'effet des liens entre les bruits sur l'interaction des modalités. La section 2.4 conclut ce chapitre.

2.2 Deux modalités

Dans cette section, nous considérons le cas de deux modalités. Après la présentation du modèle, nous détaillons les calculs d'estimation et d'information. Ces grandeurs permettront de déterminer les différentes situations de l'analyse multimodale selon les liens qui peuvent exister entre les bruits et les rapports signal à bruit des modalités.

2.2.1 Modèle

Considérons la figure 2.1, qui décrit schématiquement les mesures bruitées d'une variable $s \in \mathbb{R}$ à l'aide de deux modalités (systèmes de mesure) fournissant x_1 et x_2 .

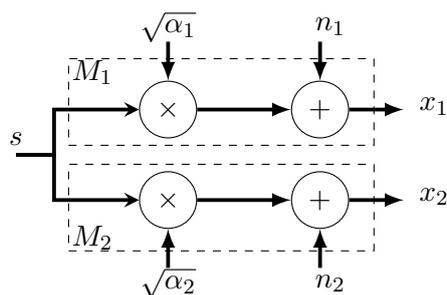


FIGURE 2.1 – Deux modalités, avec un couplage fort, vues comme des canaux de communication

Le modèle mathématique associé à cette figure, avec deux modalités et un couplage dur (hard) entre les signaux utiles, est noté \mathcal{M}_h^2 . Il est décrit par les équations :

$$\mathcal{M}_h^2 : \begin{cases} x_1 = \sqrt{\alpha_1}s + n_1 \\ x_2 = \sqrt{\alpha_2}s + n_2 \end{cases}, \quad (2.1)$$

où le “signal” d’entrée et d’intérêt $s \sim p(s)$ est la réalisation d’une variable aléatoire gaussienne de moyenne nulle et de variance unitaire, *i.e* $p(s) = \mathcal{N}(0, 1)$. Les bruits additifs $n_1 \sim p(n_1)$ et $n_2 \sim p(n_2)$ sont également des variables gaussiennes centrées réduites, *i.e* $\sim \mathcal{N}(0, 1)$. α_1 et α_2 sont les rapports signal à bruit des modalités x_1 et x_2 , respectivement. Sans restreindre la généralité, nous supposons que $\alpha_2 \leq \alpha_1$.

Nous supposons que la source s est indépendante des bruits n_1 et n_2 alors que les bruits sont corrélés. Nous notons par $\rho_n = \frac{\mathbb{E}_p[n_1 n_2]}{\mathbb{E}_p[n_1^2] \mathbb{E}_p[n_2^2]}$ le coefficient de corrélation des bruits. Il s’ensuit que, $x_1 \sim \mathcal{N}(0, 1 + \alpha_1)$ et $x_2 \sim \mathcal{N}(0, 1 + \alpha_2)$.

Liens entre les modalités :

- **Lien entre les signaux d’intérêt :** Un couplage fort est considéré entre les signaux d’intérêt représenté par la variable s , qui est commune aux deux modalités.
- **Lien entre les bruits :** Le couplage que nous considérons entre les bruits est plus faible. Il est représenté par le coefficient de corrélation, ρ_n , entre n_1 et n_2 .

Au moyen de l’erreur quadratique moyenne et de l’information mutuelle, que nous présentons ensuite, nous proposons de répondre aux questions suivantes :

1. Sous quelles conditions, deux modalités assurent-elles plus d’information qu’une seule modalité ?
2. Quels sont les liens des bruits qui favorisent l’interaction des modalités ?
3. Quel est l’impact des rapports signal à bruit des observations ? Une modalité de très faible signal à bruit est-elle sans intérêt ?

2.2.2 Calculs d’estimation et d’information

Estimation. Nous considérons le problème d’estimation de s en fonction d’une ou deux modalités. L’estimateur optimal au sens de l’erreur quadratique moyenne en fonction d’une seule modalité x_i , $i \in \{1, 2\}$, est par définition la moyenne de la loi a posteriori (théorème 1.2.1) :

$$\hat{s}_p(x_i) \triangleq \int s p(s|x_i) ds = \frac{\sqrt{\alpha_i}}{1 + \alpha_i} x_i. \quad (2.2)$$

L’erreur quadratique moyenne relative à cet estimateur est la variance de l’erreur $s - \hat{s}_p(x_i)$:

$$\mathbf{M}_{p,p}(s|x_i) \triangleq \mathbb{E} [(s - \hat{s}_p(x_i))^2] = \frac{1}{1 + \alpha_i}. \quad (2.3)$$

| *Démonstration.* Voir section A.2 de l’annexe A. □

Maintenant, si nous utilisons les deux modalités conjointement, *i.e* en exploitant leurs liens, l’estimateur optimal en fonction des deux modalités est :

$$\hat{s}_p(x_1, x_2) = \frac{(\sqrt{\alpha_1} - \sqrt{\alpha_2} \rho_n) x_1 + (\sqrt{\alpha_2} - \sqrt{\alpha_1} \rho_n) x_2}{(1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1 \alpha_2} + \rho_n)^2}. \quad (2.4)$$

L’erreur quadratique relative à $\hat{s}_p(x_1, x_2)$ est donnée par la variance de l’erreur $s - \hat{s}_p(x_1, x_2)$:

$$\mathbf{M}_{p,p}(s|x_1, x_2) \triangleq \mathbb{E} [(s - \hat{s}_p(x_1, x_2))^2] = \frac{1 - \rho_n^2}{(1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1 \alpha_2} + \rho_n)^2}. \quad (2.5)$$

| *Démonstration.* Voir section A.4 de l'annexe A. □

Information. Les deux modalités du modèle \mathcal{M}_h^2 peuvent aussi être vues comme deux canaux de communication où s est l'entrée et x_1 et x_2 sont des sorties (Figure 2.1). Nous pouvons donc considérer les mesures d'information. L'information qu'apporte une seule modalité x_i , $i \in \{1, 2\}$, sur s est donnée par l'information mutuelle entre s et x_i (paragraphe 1.1.4) :

$$I_{p,p}(s; x_i) = \frac{1}{2} \log(1 + \alpha_i). \quad (2.6)$$

| *Démonstration.* Voir section A.2 de l'annexe A. □

Alors que l'information qu'apportent conjointement x_1 et x_2 est donnée par l'information mutuelle conjointe entre s et (x_1, x_2) (paragraphe 1.1.5) :

$$I_{p,p}(s; (x_1, x_2)) = \frac{1}{2} \log \left(\frac{(1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1 \alpha_2} + \rho_n)^2}{1 - \rho_n^2} \right). \quad (2.7)$$

| *Démonstration.* Voir section B.4 de l'annexe B. □

Les expressions d'information et d'estimation, en (2.6) et (2.3), et en (2.7) et (2.5), sont liées par la relation $I_{p,p}(s; \cdot) = -\frac{1}{2} \log(\mathbf{M}_{p,p}(s|\cdot))$ qui est un cas particulier de la relation plus générale entre le gradient de l'information mutuelle et l'erreur quadratique moyenne dans un canal gaussien (paragraphe 1.3.3). L'analyse de ces grandeurs conduit alors à des résultats comparables, mais avec des interprétations selon deux points de vue différents.

2.2.3 Une ou deux modalités ?

Dans le cas multimodal, on ne se demande pas si une seule modalité est suffisante, puisqu'on postule généralement que le traitement conjoint des modalités est toujours plus avantageux. Nous envisageons cette question sur le modèle \mathcal{M}_h^2 : les deux modalités assurent-elle toujours plus d'information qu'une seule modalité ? Pour y répondre, nous comparons l'information communiquée conjointement par les deux modalités sur s , $I_{p,p}(s; (x_1, x_2))$, et l'information communiquée par une seule modalité $I_{p,p}(s; x_i)$, $i \in \{1, 2\}$. En s'appuyant sur la relation (1.25), nous pouvons écrire :

$$I_{p,p}(s; (x_1, x_2)) = I_{p,p}(s; x_1) + I_{p,p}(s; x_2|x_1) = I_{p,p}(s; x_2) + I_{p,p}(s; x_1|x_2). \quad (2.8)$$

Cette équation indique, par la non-négativité de l'information mutuelle $I_{p,p}(s; x_j|x_i) \geq 0$ ($i, j \in \{1, 2\}$ et $i \neq j$), que deux modalités présentent plus ou autant d'information qu'une seule modalité, $I_{p,p}(s; (x_1, x_2)) \geq I_{p,p}(s; x_i)$. Par conséquent, une seule modalité suffit pour tirer le maximum d'information possible sur s si et seulement si $I_{p,p}(s; x_j|x_i) = 0$. Nous disons alors que la modalité x_j est redondante sur s par rapport à x_i . Dans ce cas, x_j , x_i et s forment dans ce sens une chaîne de Markov (paragraphe 1.1.4), que nous pouvons schématiser ainsi :

$$x_j \longrightarrow x_i \longrightarrow s$$

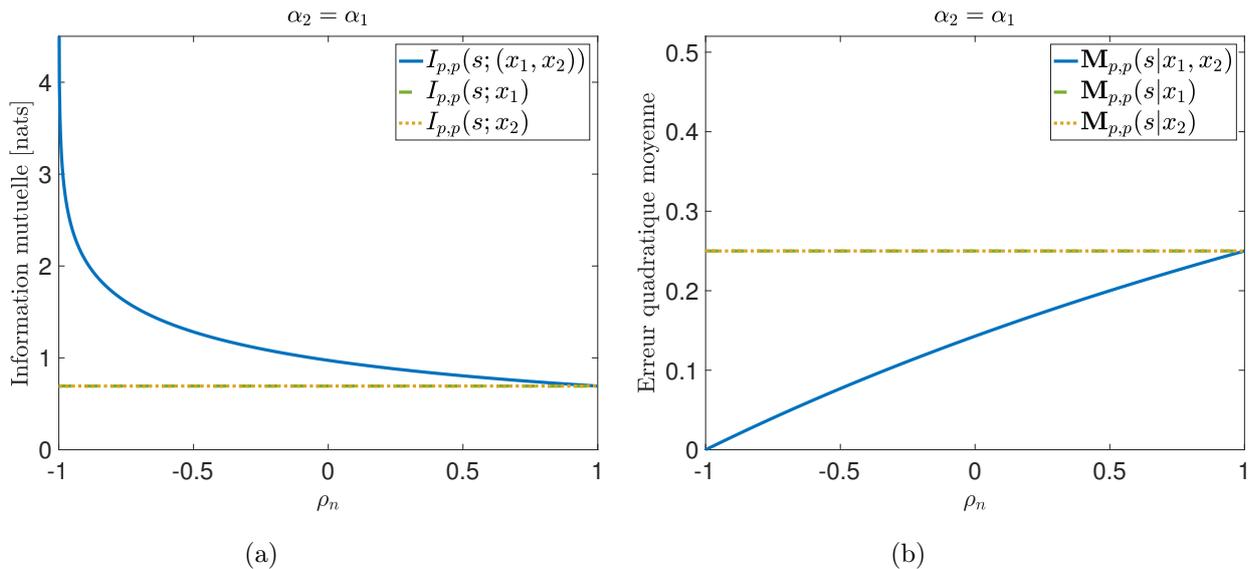


FIGURE 2.2 – Variations, en fonction de ρ_n pour $\alpha_1 = \alpha_2 = 3$, des informations mutuelles (a) et des erreurs quadratiques moyennes (b). L'information mutuelle est infinie pour $\rho_n = -1$ qui correspond à une erreur quadratique moyenne nulle.

Ce qui signifie que x_j n'apporte pas de nouveauté sur s comparé à x_i . Pour le modèle \mathcal{M}_h^2 , x_j est redondante à la modalité x_i sur s si $I_{p,p}(s; (x_1, x_2)) = I_{p,p}(s; x_i)$. A partir de (2.6) et (2.7) nous trouvons que ce cas correspond à $\rho_n = \frac{\sqrt{\alpha_j}}{\sqrt{\alpha_i}}$. Puisque $|\rho_n| \leq 1$, cette valeur de ρ_n n'est admissible que si $\alpha_j \leq \alpha_i$. C'est-à-dire que seule la modalité x_j peut être redondante, sur s , à la modalité x_i qui a le rapport signal à bruit le plus élevé.

$$I_{p,p}(s; (x_1, x_2)) = I_{p,p}(s; x_i) \Leftrightarrow \rho_n = \frac{\sqrt{\alpha_j}}{\sqrt{\alpha_i}} \quad \text{si } \alpha_j \leq \alpha_i. \quad (2.9)$$

Nous comparons, dans la suite, la performance des deux modalités à la performance d'une seule modalité pour $\alpha_2 = \alpha_1$, *i.e* modalités de même qualité, et $\alpha_2 < \alpha_1$, *i.e* modalités de différentes qualités.

2.2.3.1 Modalités de même qualité ($\alpha_2 = \alpha_1$)

Information. La figure 2.2a illustre les informations mutuelles en fonction de la corrélation des bruits ρ_n . Les informations communiquées individuellement par chaque modalité sur s sont représentées par deux lignes horizontales confondues : $I_{p,p}(s; x_i)$, $i \in \{1, 2\}$, ne dépend pas de ρ_n . Ces deux lignes en pointillés (.....) et en tirets (-.-) coïncident car $\alpha_1 = \alpha_2$: les deux modalités apportent en moyenne la même quantité d'information sur s , *i.e* $I_{p,p}(s; x_1) = I_{p,p}(s; x_2)$. La courbe pleine (—) représente l'information qu'apportent conjointement les deux modalités sur s :

$$I_{p,p}(s; (x_1, x_2)) \Big|_{\alpha_1=\alpha_2} = \frac{1}{2} \log \left(\frac{1 + 2\alpha_1 + \rho_n}{1 + \rho_n} \right). \quad (2.10)$$

Nous remarquons que $I_{p,p}(s; (x_1, x_2)) \Big|_{\alpha_1=\alpha_2}$ est au-dessus de $I_{p,p}(s; x_i) \Big|_{\alpha_1=\alpha_2}$ (droite horizontale). Ce qui signifie que conjointement les deux modalités communiquent plus d'information sur s

qu'une seule modalité, sauf si $\rho_n = 1$.

Pour $\rho_n = 1$, l'équation (2.10) devient égale à $I_{p,p}(s, x_i)$ et la courbe pleine est tangente à la ligne horizontale. Ce qui indique que les deux modalités sont totalement redondantes sur s . Le cas $\rho_n = 1$ correspond à la condition de redondance (2.9) pour $\alpha_2 = \alpha_1$.

Estimation. Les estimateurs optimaux en fonction d'une seule modalité sont différents, car bien que $\alpha_1 = \alpha_2$, $\hat{s}_p(x_1) \neq \hat{s}_p(x_2)$, puisque $x_1 \neq x_2$. Cependant ils ont la même précision $\mathbf{M}_{p,p}(s|x_1) = \mathbf{M}_{p,p}(s|x_2)$. La figure 2.2b représente la variation des erreurs quadratiques moyennes d'estimation, en fonction de ρ_n . $\mathbf{M}_{p,p}(s|x_1)$ et $\mathbf{M}_{p,p}(s|x_2)$ sont représentées par des droites horizontales puisqu'elles ne dépendent pas de ρ_n . La courbe pleine représente l'erreur d'estimation basée sur les deux modalités conjointement :

$$\mathbf{M}_{p,p}(s|x_1, x_2) \Big|_{\alpha_1=\alpha_2} = \frac{1 + \rho_n}{1 + 2\alpha_1 + \rho_n}. \quad (2.11)$$

Nous constatons que l'erreur d'estimation multimodale, $\mathbf{M}_{p,p}(s|x_1, x_2)$, est inférieure aux lignes horizontales correspondantes à $\mathbf{M}_{p,p}(s|x_1)$ et $\mathbf{M}_{p,p}(s|x_2)$. Ce qui signifie que l'estimateur :

$$\hat{s}_p(x_1, x_2) \Big|_{\alpha_1=\alpha_2} = \sqrt{\alpha_1} \frac{x_1 + x_2}{1 + 2\alpha_1 + \rho_n}, \quad (2.12)$$

est plus précis que les estimateurs $\hat{s}_p(x_1)$ et $\hat{s}_p(x_2)$.

Pour $\rho_n = 1$, les trois estimateurs, $\hat{s}_p(x_1, x_2) \Big|_{\alpha_1=\alpha_2}$, $\hat{s}_p(x_1)$ et $\hat{s}_p(x_2)$ ont la même erreur d'estimation et la courbe pleine est tangente aux lignes horizontales, $\mathbf{M}_{p,p}(s|x_1, x_2) = \mathbf{M}_{p,p}(s|x_1) = \mathbf{M}_{p,p}(s|x_2)$. Par conséquent, pour $\rho_n = 1$ une seule modalité x_1 ou x_2 suffit pour estimer s de façon optimale.

Pour $\alpha_1 = \alpha_2$, une seule modalité x_1 ou x_2 suffit à estimer s si et seulement si $\rho_n = 1$.

2.2.3.2 Modalités de différentes qualités ($\alpha_2 < \alpha_1$)

Information. La figure 2.3a représente, en fonction de ρ_n , les informations communiquées sur s par la modalité x_1 seule (.....), et par x_2 seule (-.-.-) et par les deux modalités conjointement (—). Nous remarquons que $I_{p,p}(s; x_1) > I_{p,p}(s; x_2)$ car, puisque $\alpha_1 > \alpha_2$, x_1 apporte plus d'information sur s que x_2 . De plus, nous constatons que $I_{p,p}(s; (x_1, x_2))$ est toujours strictement au-dessus de $I_{p,p}(s; x_2)$ et ne la croise jamais. C'est-à-dire que les deux modalités communiquent toujours plus d'information que la moins précise des modalités, x_2 . C'est parce que $\alpha_2 < \alpha_1$ que la condition de redondance (2.9) n'est pas vérifiée, $|\rho_n| \leq 1$ et $\frac{\sqrt{\alpha_1}}{\sqrt{\alpha_2}} > 1$. Toutefois, $I_{p,p}(s; (x_1, x_2))$ est tangente à $I_{p,p}(s; x_1)$ pour $\rho_n = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$:

$$I_{p,p}(s; (x_1, x_2)) \Big|_{\rho_n = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}} = I_{p,p}(s; x_1). \quad (2.13)$$

Dans ce cas la modalité x_2 est redondante sur s par rapport à x_1 . Ainsi, x_1 seule fournit l'information maximale qu'on peut obtenir sur s .

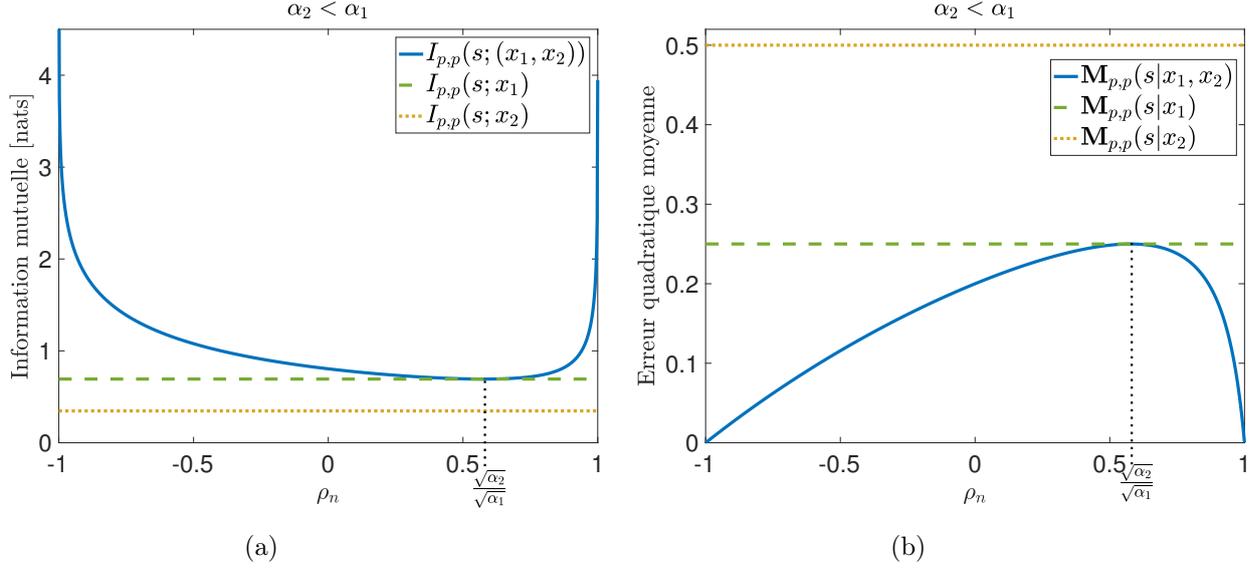


FIGURE 2.3 – Variations, en fonction de ρ_n pour $\alpha_2 = 1 < \alpha_1 = 3$, des informations mutuelles (a) et des erreurs quadratiques moyennes (b). L'information mutuelle est infinie pour $|\rho_n| = 1$ qui correspond à une erreur quadratique moyenne nulle.

Estimation. Dans la figure 2.3b, nous montrons les erreurs quadratiques moyennes d'estimation obtenues avec une seule modalité (lignes horizontales) et avec de deux modalités (—). Selon l'hypothèse $\alpha_2 < \alpha_1$, l'estimateur $\hat{s}_p(x_1)$ est plus précis que $\hat{s}_p(x_2)$ et présente donc moins d'erreur d'estimation. Effectivement, $\mathbf{M}_{p,p}(s|x_1) < \mathbf{M}_{p,p}(s|x_2)$. Nous constatons que l'erreur d'estimation multimodale $\mathbf{M}_{p,p}(s|x_1, x_2)$ est inférieure aux erreurs $\mathbf{M}_{p,p}(s|x_1)$ et $\mathbf{M}_{p,p}(s|x_2)$, et n'est tangente à $\mathbf{M}_{p,p}(s|x_1)$ que pour $\rho_n = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$. Ce qui signifie que l'estimateur optimal $\hat{s}_p(x_1, x_2)$ est strictement meilleur que l'estimateur $\hat{s}_p(x_2)$ et est meilleur que $\hat{s}_p(x_1)$ avec égalité en $\rho_n = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$. Pour $\rho_n = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$, le facteur de x_2 dans (2.4) s'annule et l'estimateur optimal multimodal, basé sur x_1 et x_2 , se réduit à l'estimateur optimal basé sur la seule modalité x_1 (2.2) :

$$\hat{s}_p(x_1, x_2) \Big|_{\rho_n = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}} = \hat{s}_p(x_1). \quad (2.14)$$

Pour cette valeur de ρ_n , la modalité x_1 suffit pour estimer s .

Pour $\alpha_2 < \alpha_1$, la seule modalité x_1 suffit à estimer s si et seulement si $\rho_n = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$.

2.2.4 Effet de la corrélation des bruits et des rapports signal à bruit

Dans cette sous-section, nous étudions l'influence de la qualité des modalités donc des mesures, c'est-à-dire du rapport signal à bruit, et l'effet du lien entre les bruits, c'est-à-dire de la corrélation des bruits ρ_n . Qu'est ce qui donne les meilleures performances : des modalités avec des qualités comparables ou des modalités avec des qualités disparates ? Une modalité x_i de très faible qualité (α_i petit) est-elle à négliger ? Nous avons montré que, pour une corrélation des bruits spécifique, les deux modalités peuvent être redondantes sur s . Cette cor-

relation entraîne-t-elle toujours une redondance ou peut-elle être bénéfique pour la multimodalité ?

Pour le modèle \mathcal{M}_h^2 , nous considérons trois cas de figures : $\alpha_2 = \alpha_1$, $\alpha_2 < \alpha_1$ et $\alpha_2 \ll \alpha_1$, afin de déterminer quel scénario augmente l'information et donne les meilleures estimations. Pour chaque scénario, nous examinons l'effet de la corrélation des bruits en comparant deux modèles : un premier modèle dont les deux modalités, x_1 et x_2 , ont des bruits corrélés, \mathcal{M}_h^2 , et un deuxième modèle dont les bruits des modalités sont non corrélés, noté \mathcal{M}_{ho}^2 . La décorrélation des bruits, $\rho_n = 0$, correspond à leur indépendance compte tenu de l'hypothèse de gaussianité. L'indépendance entre les bruits se traduit par l'indépendance des modalités conditionnellement à la source s , étant donné que les modalités restent liées par le signal d'intérêt. La comparaison des modèles \mathcal{M}_h^2 et \mathcal{M}_{ho}^2 permettra de déterminer si la corrélation des bruits est bénéfique ou néfaste à la multimodalité.

Les mesures d'estimation et d'information relatives au deuxième modèle \mathcal{M}_{ho}^2 s'obtiennent à partir des mesures introduites à la section 2.2.2 en mettant ρ_n à 0. Par conséquent, l'estimateur optimal en fonction des deux modalités du modèle \mathcal{M}_{ho}^2 (c'est-à-dire avec bruits indépendants) et l'erreur associée sont :

$$\hat{s}_p(x_1, x_2) \Big|_{\rho_n=0} = \frac{\sqrt{\alpha_1}x_1 + \sqrt{\alpha_2}x_2}{1 + \alpha_1 + \alpha_2}, \quad (2.15)$$

$$\mathbf{M}_{p,p}(s|x_1, x_2) \Big|_{\rho_n=0} = \frac{1}{1 + \alpha_1 + \alpha_2}. \quad (2.16)$$

L'information qu'apportent conjointement les deux modalités sur s dans le cas du modèle \mathcal{M}_{ho}^2 est :

$$I_{p,p}(s; (x_1, x_2)) \Big|_{\rho_n=0} = \frac{1}{2} \log(1 + \alpha_1 + \alpha_2). \quad (2.17)$$

2.2.4.1 Modalités de même qualité ($\alpha_2 = \alpha_1$)

Nous commençons cette étude par le cas de deux modalités de qualités comparables, *i.e.* $\alpha_2 = \alpha_1$. Dans ce cas, les deux modalités apportent la même quantité d'information sur s : $I_{p,p}(s; x_2) = I_{p,p}(s; x_1)$. Attention, ceci ne veut pas dire qu'elles sont redondantes. En effet, x_1 et x_2 sont redondantes sur s si et seulement si $I_{p,p}(s; x_i|x_j) \Big|_{\alpha_1=\alpha_2} = 0$, c'est-à-dire si $\rho_n = 1$.

Information. La figure 2.4a illustre les informations mutuelles pour le modèle \mathcal{M}_h^2 (—) et pour le modèle \mathcal{M}_{ho}^2 (---). $I_{p,p}(s; (x_1, x_2)) \Big|_{\rho_n=0}$ ne dépend pas de ρ_n ($\rho_n = 0$), ainsi elle est représentée par une ligne horizontale.

Nous constatons que, $I_{p,p}(s; (x_1, x_2)) \leq I_{p,p}(s; (x_1, x_2)) \Big|_{\rho_n=0}$ pour $\rho_n \in [0, 1]$. Ainsi les corrélations positives des bruits introduisent de la redondance.

Réciproquement, pour $\rho_n \in [-1, 0]$, $I_{p,p}(s; (x_1, x_2)) \geq I_{p,p}(s; (x_1, x_2)) \Big|_{\rho_n=0}$. On en déduit que les corrélations négatives des bruits favorisent l'interaction et les modalités de \mathcal{M}_h^2 apportent plus d'information que celles du modèle \mathcal{M}_{ho}^2 . En particulier, l'information mutuelle est infinie pour $\rho_n = -1$. L'information infinie signifie que les deux modalités apportent toute l'information sur s .

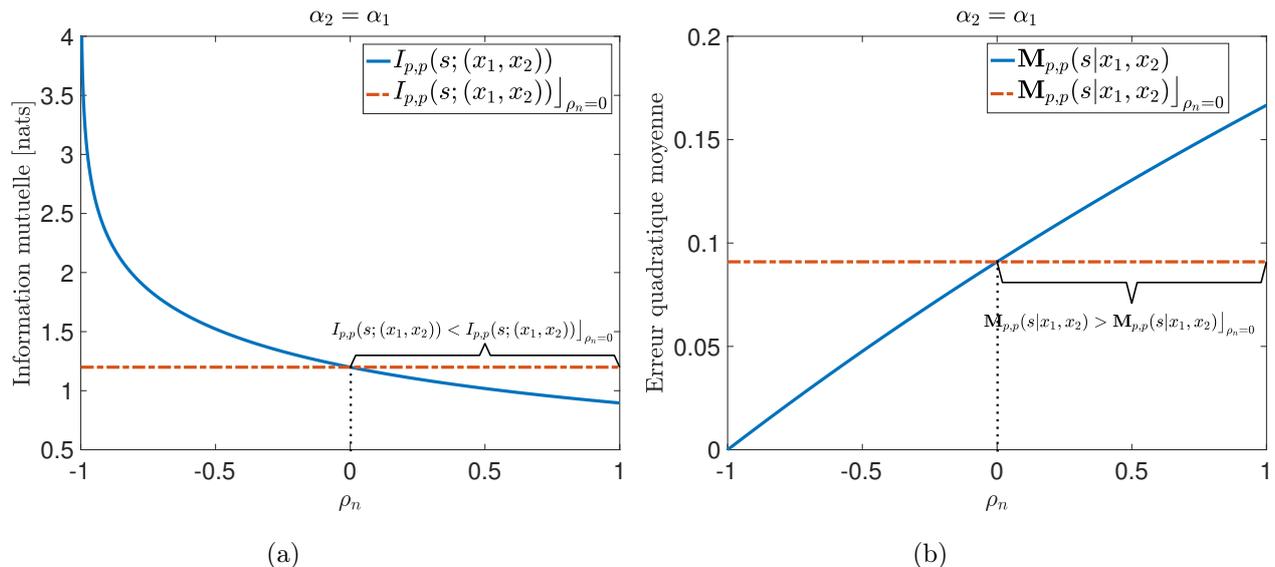


FIGURE 2.4 – Variations, en fonction de ρ_n pour $\alpha_2 = \alpha_1 = 5$, des informations mutuelles (a) et des erreurs quadratiques moyennes (b). L'information mutuelle est infinie pour $\rho_n = 1$ qui correspond à une erreur quadratique moyenne nulle.

Estimation. Nous représentons sur la figure 2.4b, les erreurs d'estimation multimodales $\mathbf{M}_{p,p}(s|x_1, x_2)$ (—) et $\mathbf{M}_{p,p}(s|x_1, x_2)|_{\rho_n=0}$ (---).

Pour $\rho_n \in [0, 1]$ l'estimateur optimal du modèle \mathcal{M}_h^2 est moins précis que l'estimateur optimal du modèle \mathcal{M}_{ho}^2 , car le bruit est plus mal compensé pour \mathcal{M}_h^2 que pour \mathcal{M}_{ho}^2 à cause de la redondance introduite par les corrélations positives, $\rho_n \geq 0$.

Réciproquement, pour les corrélations négatives $\mathbf{M}_{p,p}(s|x_1, x_2) \leq \mathbf{M}_{p,p}(s|x_1, x_2)|_{\rho_n=0}$, puisque la corrélation $\rho_n \leq 0$ permet de compenser le bruit. En particulier, le bruit est totalement supprimé pour $\rho_n = -1$ et l'erreur d'estimation est nulle : $\mathbf{M}_{p,p}(s|x_1, x_2)|_{\rho_n=-1, \alpha_1=\alpha_2} = 0$.

2.2.4.2 Modalités de différentes qualités ($\alpha_2 < \alpha_1$)

Considérons maintenant le cas $\alpha_2 < \alpha_1$. Par rapport au cas précédent, la modalité x_2 apporte moins d'information sur s que x_1 , $I_{p,p}(s; x_2) < I_{p,p}(s; x_1)$. Néanmoins, x_2 apporte une information plus fiable sur le bruit que x_1 .

Information. La figure 2.5a représente les informations mutuelles pour le modèle \mathcal{M}_h^2 (—) et pour le modèle \mathcal{M}_{ho}^2 (---) en fonction du paramètre ρ_n .

Nous remarquons que, sur une partie des corrélations positives, $I_{p,p}(s; (x_1, x_2)) \leq I_{p,p}(s; (x_1, x_2))|_{\rho_n=0}$. A partir des équations (2.7) et (2.17), nous trouvons que :

$$I_{p,p}(s; (x_1, x_2)) < I_{p,p}(s; (x_1, x_2))|_{\rho_n=0} \quad \text{si et seulement si} \quad \rho_n \in I_{red} = \left] 0, \frac{2\sqrt{\alpha_1\alpha_2}}{\alpha_1 + \alpha_2} \right[. \quad (2.18)$$

Nous concluons alors que la corrélation des bruits de l'intervalle I_{red} induit de la redondance et donc moins d'information. Une redondance totale entre les deux modalités de \mathcal{M}_h^2 est atteinte au point $\rho_n = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$ où $I_{p,p}(s; x_2|x_1) = 0$.

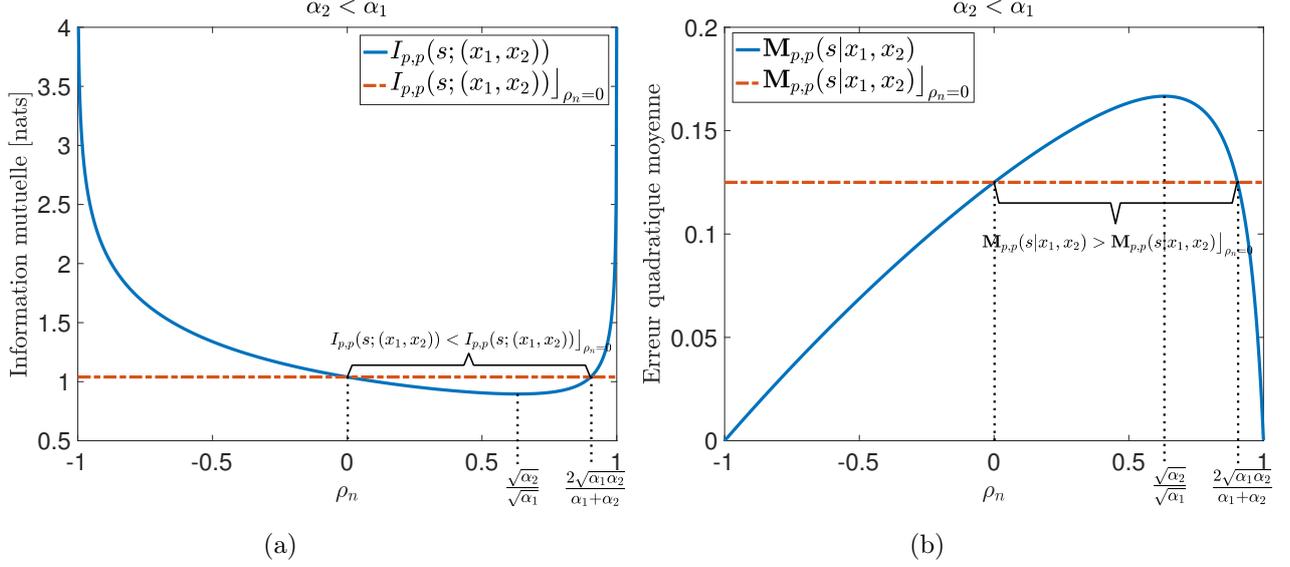


FIGURE 2.5 – Variations, en fonction de ρ_n pour $\alpha_2 = 2 < \alpha_1 = 5$, des informations mutuelles (a) et des erreurs quadratiques moyennes (b)

Au-delà de l'intervalle I_{red} , $I_{p,p}(s; (x_1, x_2)) \geq I_{p,p}(s; (x_1, x_2))|_{\rho_n=0}$, car la corrélation des bruits favorise l'interaction entre les modalités. Notamment, sur l'intervalle $\rho_n \in \left] \frac{2\sqrt{\alpha_1\alpha_2}}{\alpha_1+\alpha_2}, 1 \right]$ la corrélation positive n'induit plus de la redondance comme dans le cas $\alpha_1 = \alpha_2$. En particulier, l'information mutuelle est infinie pour $\rho_n = 1$ et $\rho_n = -1$.

Estimation. Les erreurs quadratiques moyennes pour le modèle \mathcal{M}_h^2 (—) et pour le modèle \mathcal{M}_{ho}^2 (---) sont illustrées par la figure 2.5b.

Sur l'intervalle I_{red} , $\mathbf{M}_{p,p}(s|x_1, x_2) \geq \mathbf{M}_{p,p}(s|x_1, x_2)|_{\rho_n=0}$. En raison de la redondance introduite par la corrélation des bruits, le bruit est mal compensé dans le cas du modèle \mathcal{M}_h^2 . L'erreur d'estimation est maximale pour $\rho_n = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$, puisque la modalité x_2 est totalement redondante sur s par rapport à x_1 (paragraphe 2.2.3).

Au-delà de I_{red} , nous constatons que $\mathbf{M}_{p,p}(s|x_1, x_2) \leq \mathbf{M}_{p,p}(s|x_1, x_2)|_{\rho_n=0}$. Ce qui signifie que le modèle \mathcal{M}_h^2 induit moins d'erreur d'estimation que le modèle \mathcal{M}_{ho}^2 , car la corrélation des bruits est exploitée pour réduire le bruit. Notamment, le bruit est complètement supprimé et le signal s est parfaitement estimé pour $\rho_n = 1$ et $\rho_n = -1$ où l'erreur quadratique moyenne d'estimation est nulle :

$$\hat{s}_p(x_1, x_2)|_{\rho_n=-1} = \frac{\sqrt{\alpha_1}x_1 + \sqrt{\alpha_2}x_2}{\sqrt{\alpha_1} + \sqrt{\alpha_2}}, \quad \mathbf{M}_{p,p}(s|x_1, x_2)|_{\rho_n=-1} = 0. \quad (2.19)$$

$$\hat{s}_p(x_1, x_2)|_{\rho_n=1, \alpha_1 \neq \alpha_2} = \frac{\sqrt{\alpha_1}x_1 - \sqrt{\alpha_2}x_2}{\sqrt{\alpha_1} - \sqrt{\alpha_2}}, \quad \mathbf{M}_{p,p}(s|x_1, x_2)|_{\rho_n=1, \alpha_1 \neq \alpha_2} = 0. \quad (2.20)$$

Notons que pour $\rho_n = 1$ et $\alpha_1 = \alpha_2$, l'estimateur (2.20) est non défini, puisque, comme nous avons vu auparavant, les deux modalités sont totalement redondantes sur s .

Les figures 2.5a et 2.5b montrent que les courbes pour $\rho_n \neq 0$ (—) et pour $\rho_n = 0$ (---) se croisent en deux points : $\rho_n = 0$ et $\rho_n = \frac{2\sqrt{\alpha_1\alpha_2}}{\alpha_1+\alpha_2}$. Pour $\rho_n = 0$, les deux modèles sont identiques,

$\mathcal{M}_h^2 = \mathcal{M}_{ho}^2$. Alors que pour $\rho_n = \frac{2\sqrt{\alpha_1\alpha_2}}{\alpha_1+\alpha_2}$, les deux modèles sont équivalents mais différents. Équivalents parce qu'ils induisent les mêmes erreurs d'estimation et fournissent la même quantité d'information sur s . Différents, parce qu'ils apportent différents éléments d'information sur s et que leurs estimateurs optimaux sont différents. En effet, pour $\rho_n = \frac{2\sqrt{\alpha_1\alpha_2}}{\alpha_1+\alpha_2}$, on a :

$$\hat{s}_p(x_1, x_2) \Big|_{\rho_n = \frac{2\sqrt{\alpha_1\alpha_2}}{\alpha_1+\alpha_2}} = \frac{\sqrt{\alpha_1}(\alpha_1 - \alpha_2)x_1 + \sqrt{\alpha_2}(\alpha_2 - \alpha_1)x_2}{(\alpha_1 + \alpha_2) \left((1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1\alpha_2} + \frac{2\sqrt{\alpha_1\alpha_2}}{\alpha_1+\alpha_2})^2 \right)}, \quad (2.21)$$

alors que pour $\rho_n = 0$,

$$\hat{s}_p(x_1, x_2) \Big|_{\rho_n=0} = \frac{\sqrt{\alpha_1}x_1 + \sqrt{\alpha_2}x_2}{1 + \alpha_1 + \alpha_2}. \quad (2.22)$$

Dans le cadre du modèle \mathcal{M}_h^2 , la redondance apparaît pour certaines corrélations positives si $\alpha_2 < \alpha_1$ et pour toutes les corrélations positives si $\alpha_2 = \alpha_1$. En revanche, quelles que soient les valeurs des rapports signal à bruit, les corrélations négatives des bruits assurent plus d'information que les corrélations positives ou nulles. Ainsi une corrélation négative entre les bruits est la condition qui améliore les performances de la multimodalité.

Les corrélations négatives des bruits favorisent toujours l'interaction des modalités.

2.2.4.3 Modalités de qualités très disparates ($\alpha_2 \ll \alpha_1$)

Nous considérons maintenant le cas de deux modalités de qualités très disparates $\alpha_2 \ll \alpha_1$. Dans le cas précédent $\alpha_2 < \alpha_1$, nous avons vu que x_2 apporte de l'information sur le bruit qui favorise l'interaction pour de fortes corrélations des bruits. Dans le cas $\alpha_2 \ll \alpha_1$, x_2 apporte une information négligeable sur s par rapport à x_1 , $I_{p,p}(s; x_2) \ll I_{p,p}(s; x_1)$, mais elle apporte réciproquement une information très fiable sur le bruit par rapport à x_1 . En effet, pour $\alpha_2 \rightarrow 0$ la modalité x_2 tend vers une référence du bruit n_2 . Par conséquent, Le modèle devient proche d'un modèle classique où x_2 devient une référence du bruit [29] :

$$\begin{cases} x_1 = \sqrt{\alpha_1}s + n_1 \\ x_2 = \sqrt{\alpha_2}s + n_2 \end{cases} \xrightarrow{\alpha_2 \rightarrow 0} n_2. \quad (2.23)$$

Pour des bruits non corrélés (modèle \mathcal{M}_{ho}^2), l'information sur le bruit qu'apporte x_2 n'a aucun intérêt. Dans ce cas seule la modalité x_1 est à utiliser pour estimer s :

$$I_{p,p}(s; (x_1, x_2)) \Big|_{\rho_n=0, \alpha_2 \ll \alpha_1} \approx I_{p,p}(s; x_1), \quad (2.24)$$

$$\hat{s}_p(s|x_1, x_2) \Big|_{\rho_n=0, \alpha_2 \ll \alpha_1} \approx \hat{s}_p(s|x_1). \quad (2.25)$$

Alors que pour des bruits corrélés (modèle \mathcal{M}_h^2), l'information qu'apporte x_2 sur n_2 peut être utilisée pour réduire le bruit en x_1 . Ainsi conjointement les deux modalités peuvent révéler plus d'information sur s .

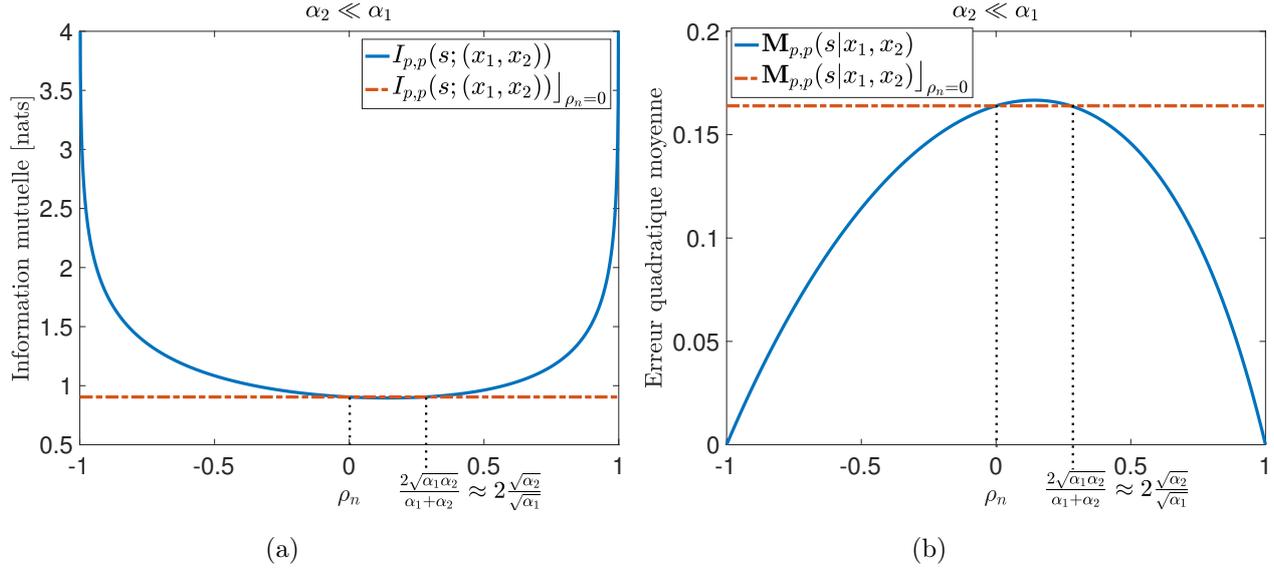


FIGURE 2.6 – Variations, en fonction de ρ_n pour $\alpha_2 = .1 \ll \alpha_1 = 5$, des informations mutuelles (a) et des erreurs quadratiques moyennes (b)

Information. Sur la figure 2.6a, nous représentons les informations mutuelles multimodales de \mathcal{M}_h^2 (—) et de \mathcal{M}_{ho}^2 (---). Nous remarquons que $I_{p,p}(s; (x_1, x_2))$ est supérieure à $I_{p,p}(s; (x_1, x_2))|_{\rho_n=0}$ presque tout le temps et que l'intervalle de redondance I_{red} tend vers 0. Par conséquent, conjointement les deux modalités de \mathcal{M}_h^2 communiquent souvent plus d'information que les modalités de \mathcal{M}_{ho}^2 car la corrélation des bruits favorise l'interaction.

Estimation. La figure 2.6b illustre les erreurs d'estimation multimodales $\mathbf{M}_{p,p}(s|x_1, x_2)$ (—) et $\mathbf{M}_{p,p}(s|x_1, x_2)|_{\rho_n=0}$ (---). Nous constatons que $\mathbf{M}_{p,p}(s|x_1, x_2)$ est souvent inférieure à $\mathbf{M}_{p,p}(s|x_1, x_2)|_{\rho_n=0}$, car l'information sur le bruit contenue dans x_2 est exploitée pour compenser le bruit en x_1 .

Une modalité de faible qualité communique de l'information fiable sur le bruit qui est utile à l'estimation multimodale du signal utile s .

2.2.4.4 $\alpha_2 = \alpha_1$ versus $\alpha_2 < \alpha_1$ versus $\alpha_2 \ll \alpha_1$:

Pour finir notre étude sur le gain de la multimodalité, nous comparons les trois scénarios $\alpha_2 = \alpha_1$, $\alpha_2 < \alpha_1$ et $\alpha_2 \ll \alpha_1$.

Information. La figure 2.7a représente les informations multimodales dans les cas : $\alpha_2 = \alpha_1$ (---), $\alpha_2 < \alpha_1$ (---) et $\alpha_2 \ll \alpha_1$ (—).

Nous constatons que sur l'intervalle $\rho_n \in [-1, 0.5 + \epsilon]$ avec $\epsilon \rightarrow 0$, $I_{p,p}(s; (x_1, x_2))|_{\alpha_2=\alpha_1}$ est supérieure à $I_{p,p}(s; (x_1, x_2))|_{\alpha_2 < \alpha_1}$ et $I_{p,p}(s; (x_1, x_2))|_{\alpha_2 \ll \alpha_1}$. Ainsi sur cet intervalle, les deux modalités de mêmes qualités communiquent le maximum d'information sur s .

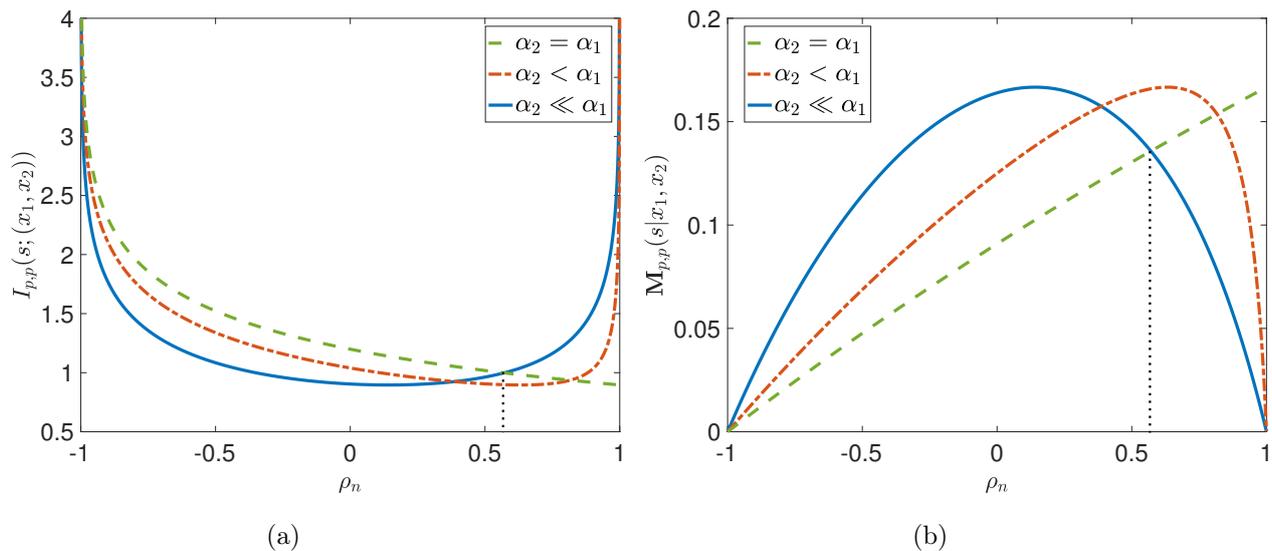


FIGURE 2.7 – Variations de l'information conjointe (a) et des erreurs quadratiques moyennes (b) en fonction de ρ_n pour différentes valeurs de α_2 où $\alpha_1 = 5$.

Réciproquement, sur l'intervalle $\rho_n \in [0.5 + \epsilon, 1]$ avec $\epsilon \rightarrow 0$, $I_{p,p}(s; (x_1, x_2)) \Big|_{\rho_n=0, \alpha_2 \ll \alpha_1}$ est supérieure aux deux autres informations multimodales. C'est-à-dire que, pour de fortes corrélations positives, les deux modalités de qualités très disparates sont plus informatives sur s .

Estimation. La figure 2.7b illustre les erreurs d'estimation multimodales dans les cas : $\alpha_2 = \alpha_1$ (---), $\alpha_2 < \alpha_1$ (-.-.) et $\alpha_2 \ll \alpha_1$ (—).

Nous constatons que $\mathbf{M}_{p,p}(s|x_1, x_2) \Big|_{\alpha_2 \ll \alpha_1}$ est inférieure à $\mathbf{M}_{p,p}(s|x_1, x_2) \Big|_{\alpha_2 = \alpha_1}$ et $\mathbf{M}_{p,p}(s|x_1, x_2) \Big|_{\alpha_2 < \alpha_1}$ pour $\rho_n \in [0.5 + \epsilon, 1]$ avec $\epsilon \rightarrow 0$. C'est-à-dire que nous tirons le maximum d'avantage de l'information sur le bruit qu'apporte x_2 pour $\alpha_2 \ll \alpha_1$, si les bruits sont fortement et positivement corrélés.

Réciproquement dans l'intervalle $[-1, 0.5 + \epsilon]$ où $\epsilon \rightarrow 0$, nous remarquons que $\mathbf{M}_{p,p}(s|x_1, x_2) \Big|_{\alpha_2 = \alpha_1}$ est en-dessous de $\mathbf{M}_{p,p}(s|x_1, x_2) \Big|_{\alpha_2 < \alpha_1}$ et $\mathbf{M}_{p,p}(s|x_1, x_2) \Big|_{\alpha_2 \ll \alpha_1}$. Nous ne pouvons pas conclure que les bruits sont mal compensés dans les autres cas comparé à $\alpha_2 = \alpha_1$, puisque la puissance du bruit est différente dans les trois scénarios. Ce qu'on peut dire c'est que l'estimateur optimal $\hat{s}_p(s|x_1, x_2) \Big|_{\alpha_2 = \alpha_1}$ est plus précis que $\hat{s}_p(s|x_1, x_2) \Big|_{\alpha_2 < \alpha_1}$ et $\hat{s}_p(s|x_1, x_2) \Big|_{\alpha_2 \ll \alpha_1}$. Ainsi, nous avons tout intérêt à avoir des modalités de mêmes qualités.

Pour une forte corrélation positive des bruits, deux modalités de qualités très disparates apportent plus d'information sur s . Alors que pour d'autres valeurs de corrélation, on a intérêt à avoir des modalités de même qualité.

2.3 Trois modalités

Dans cette section, nous considérons le cas de trois modalités. Nous examinons les mêmes problématiques que dans la section 2.2 : le nombre optimal des modalités, l'effet des liens

(corrélations des bruits) et l'effet de la qualité des mesures (rapports signal à bruit) sur l'interaction. Afin de mettre l'accent sur l'effet des corrélations des bruits, le signal d'intérêt est commun à toutes les modalités. Par conséquent, les modalités sont dites fortement couplées.

Nous avons montré que dans le cas de deux modalités du modèle \mathcal{M}_h^2 , le maximum de bénéfice de la multimodalité est assuré avec des bruits négativement corrélés. Ce résultat peut-il être étendu pour trois modalités ?

2.3.1 Modèle

Nous considérons le schéma de la figure 2.8, qui représente trois canaux linéaires de communication ou encore trois systèmes de mesures (modalités) d'une variable s fournissant des mesures bruitées x_1 , x_2 et x_3 .

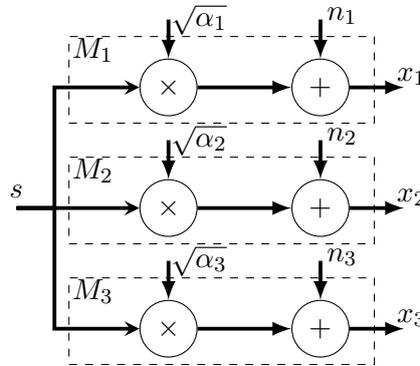


FIGURE 2.8 – Trois modalités avec un couplage fort vues comme des canaux de communication

Le modèle mathématique associé au schéma ci-dessus est :

$$\mathcal{M}_h^3 : \begin{cases} x_1 = \sqrt{\alpha_1}s + n_1 \\ x_2 = \sqrt{\alpha_2}s + n_2 \\ x_3 = \sqrt{\alpha_3}s + n_3 \end{cases}, \quad (2.26)$$

où $s \sim p(s)$ est le signal d'intérêt qui est une variable aléatoire gaussienne de moyenne nulle et de variance unitaire, *i.e* $p(s) = \mathcal{N}(0, 1)$. Les bruits additifs, n_1 , n_2 et n_3 , des trois canaux de communication sont des variables aléatoires gaussiennes de moyennes nulles et de variances unitaires, *i.e* $\sim \mathcal{N}(0, 1)$. α_1 , α_2 et α_3 sont, respectivement, les rapports signal à bruit des modalités x_1 , x_2 et x_3 . Sans perte de généralité, nous supposons que $\alpha_3 \leq \alpha_2 \leq \alpha_1$.

Nous supposons que la source s est indépendante des bruits alors que les bruits sont corrélés. Nous notons par $\rho_{12} = \frac{\mathbb{E}[n_1 n_2]}{\mathbb{E}[n_1^2] \mathbb{E}[n_2^2]}$, $\rho_{13} = \frac{\mathbb{E}[n_1 n_3]}{\mathbb{E}[n_1^2] \mathbb{E}[n_3^2]}$ et $\rho_{23} = \frac{\mathbb{E}[n_2 n_3]}{\mathbb{E}[n_2^2] \mathbb{E}[n_3^2]}$, les coefficients de corrélation de n_2 et n_1 , n_3 et n_1 et de n_3 et n_2 , respectivement. Ainsi, $x_1 \sim \mathcal{N}(0, 1 + \alpha_1)$, $x_2 \sim \mathcal{N}(0, 1 + \alpha_2)$ et $x_3 \sim \mathcal{N}(0, 1 + \alpha_3)$.

Soient $\mathbf{n} = [n_1, n_2, n_3]^T$ et $\mathbf{x} = [x_1, x_2, x_3]^T$. Nous notons par $\mathbf{\Sigma}$ et $\mathbf{\Omega}$ les matrices de covariances des bruits \mathbf{n} et des modalités \mathbf{x} , respectivement. Ainsi, selon les hypothèses précitées,

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}, \quad \Omega = \begin{pmatrix} 1 + \alpha_1 & \sqrt{\alpha_1 \alpha_2} + \rho_{12} & \sqrt{\alpha_1 \alpha_3} + \rho_{13} \\ \sqrt{\alpha_1 \alpha_2} + \rho_{12} & 1 + \alpha_2 & \sqrt{\alpha_2 \alpha_3} + \rho_{23} \\ \sqrt{\alpha_1 \alpha_3} + \rho_{13} & \sqrt{\alpha_2 \alpha_3} + \rho_{23} & 1 + \alpha_3 \end{pmatrix}. \quad (2.27)$$

Liens entre les modalités :

- **Lien entre les signaux d'intérêt :** Un couplage fort est considéré entre les signaux d'intérêt représenté par la variable s , qui est commune aux trois modalités.
- **Lien entre les bruits :** Le couplage que nous considérons entre les bruits est représenté par les coefficients de corrélation, ρ_{12} , entre n_1 et n_2 , ρ_{13} , entre n_1 et n_3 , et ρ_{23} , entre n_2 et n_3 .

Dans ce qui suit nous étudions les questions suivantes,

1. Sous quelles conditions, trois modalités assurent-elles plus d'information qu'une seule ou deux modalités ?
2. Parmi les trois modalités, quelles sont les deux modalités qui assurent plus d'information ?
3. Quelles sont les corrélations des bruits qui favorisent l'interaction ?

2.3.2 Valeurs éligibles des corrélations

Les valeurs éligibles des corrélation, ρ_{12} , ρ_{13} , et ρ_{23} , sont telles que la matrice de covariance de $\mathbf{n} = [n_1, n_2, n_3]^T$, Σ , est semi-définie positive. Selon le critère de Sylvester [30], la matrice Σ est semi-définie positive si et seulement si les mineurs principaux sont non-négatifs. Les mineurs principaux sont les déterminants des sous-matrices carrées en haut à gauche de Σ ,

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}$$

Les déterminants de ces matrices sont les mineurs principaux de Σ

Par conséquent, Σ est semi-définie positive si et seulement si,

$$1 \geq 0, \quad \begin{vmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{vmatrix} \geq 0 \quad \text{et} \quad |\Sigma| = \begin{vmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{vmatrix} \geq 0. \quad (2.28)$$

Etant donné que $|\rho_{12}| \leq 1$, la deuxième condition est vérifiée. Ainsi les valeurs éligibles des corrélations sont déterminées par la troisième condition qui est :

$$|\Sigma| = 1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23} \geq 0. \quad (2.29)$$

Cette condition est symétrique en ρ_{12} , ρ_{13} et ρ_{23} . Pour une étude générale qui s'applique aux trois corrélations, nous considérons x , y et z tels que le triplet (x, y, z) représente toutes les combinaisons possibles des trois corrélations ρ_{12} , ρ_{13} et ρ_{23} :

$$(x, y, z) \in \{(\rho_{12}, \rho_{13}, \rho_{23}), (\rho_{12}, \rho_{23}, \rho_{13}), (\rho_{13}, \rho_{12}, \rho_{23}), (\rho_{13}, \rho_{23}, \rho_{12}), (\rho_{23}, \rho_{12}, \rho_{13}), (\rho_{23}, \rho_{13}, \rho_{12})\}.$$

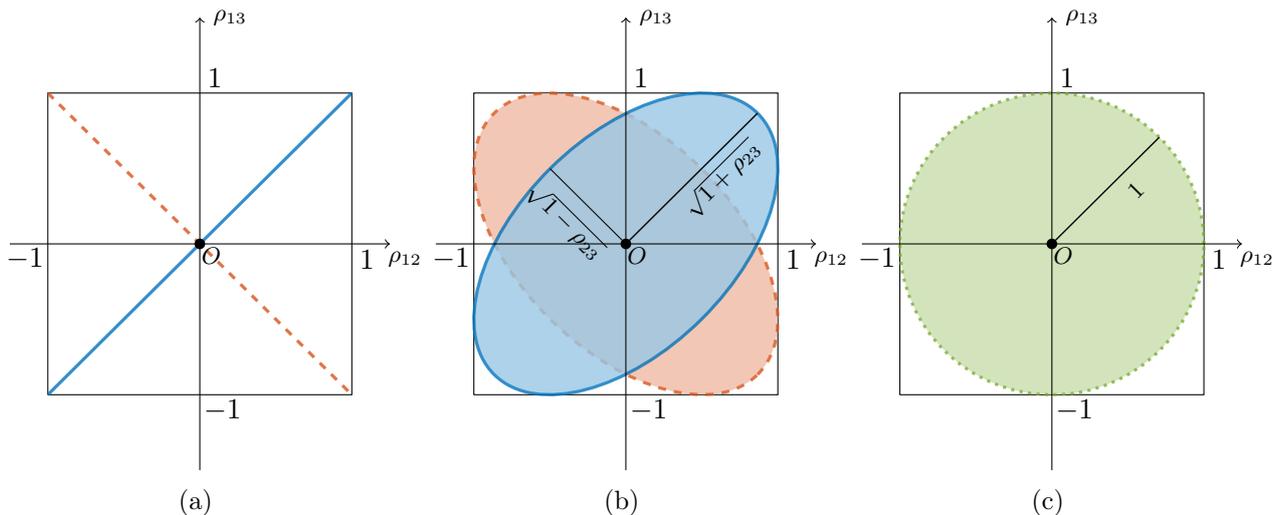


FIGURE 2.9 – Les valeurs éligibles de ρ_{12} et ρ_{13} sont présentées par des surfaces colorées pour ρ_{23} fixe : $|\rho_{23}| = 1$ (a), $|\rho_{23}| < 1$ (b) et $\rho_{23} = 0$ (c). Les régions bleues correspondent à $\rho_{23} > 0$ alors que les régions rouges (délimitées par des tirets) correspondent à $\rho_{23} < 0$.

Ainsi, nous pouvons écrire la condition (2.29) comme suit :

$$-1 + x^2 + y^2 + z^2 - 2xyz \leq 0. \quad (2.30)$$

Les valeurs des corrélations qui vérifient (2.30) sont, par conséquent, à l'intérieur d'une forme géométrique tridimensionnelle dont la surface extérieure est donnée par l'équation :

$$-1 + x^2 + y^2 + z^2 - 2xyz = 0. \quad (2.31)$$

En fixant deux corrélations, y et z , les valeurs éligibles de la troisième corrélation, x , sont à l'intérieur de l'intervalle :

$$x \in I_e = \left[yz - \sqrt{(1-y^2)(1-z^2)}, yz + \sqrt{(1-y^2)(1-z^2)} \right]. \quad (2.32)$$

Cet intervalle est la solution de l'inégalité de second degré (2.30) pour y et z fixes et x inconnu.

Représentation graphique en 2D. En général pour deux variables x et y , l'équation cartésienne d'une conique est donnée par :

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0, \quad (2.33)$$

où A, B, C, D, E et F sont des réels. Par analogie avec l'équation (2.31), $A = 1$, $B = -2z$, $C = 1$, $D = 0$, $E = 0$ et $F = z^2 - 1$. Par conséquent, en fixant un coefficient de corrélation, z , les deux autres corrélations, x et y , sont à l'intérieur d'une conique d'équation cartésienne :

$$-1 + x^2 + y^2 + z^2 - 2xyz = 0. \quad (2.34)$$

Le discriminant de cette conique est $\Delta = B^2 - 4AC = 4(z^2 - 1)$.

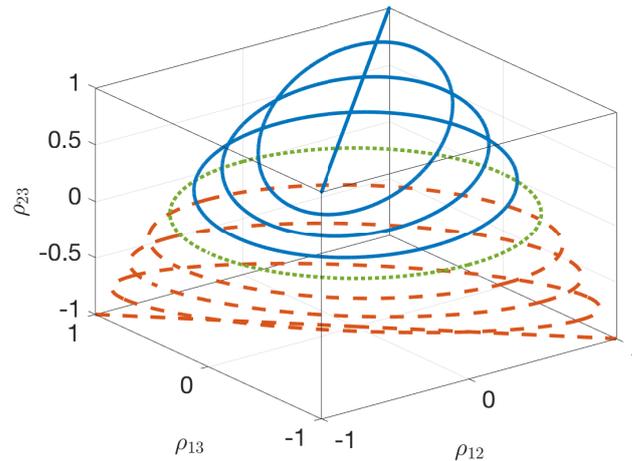
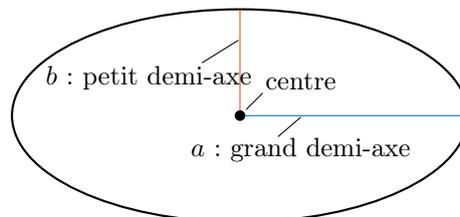


FIGURE 2.10 – Valeurs admissibles des corrélations des bruits représentées comme une superposition d’ellipses horizontales.

Si $z = \pm 1$, le discriminant est nul et $x = \pm y$ (Figure 2.9a). Autrement, *i.e.* $|z| < 1$, le discriminant est strictement négatif, $\Delta < 0$, et la conique est une ellipse d’équation canonique :

$$\frac{(x+y)^2}{2(1+z)} + \frac{(x-y)^2}{2(1-z)} = 1.$$

Le centre de cet ellipse est au point O de coordonnées $(x=0, y=0)$. Les deux axes de l’ellipse sont obtenus par une rotation des axes Ox et Oy d’un angle $\frac{\pi}{4}$ si $z > 0$, et d’un angle $-\frac{\pi}{4}$ si $z < 0$. La longueur du grand demi-axe est $a = \sqrt{1+|z|}$ et la longueur du petit demi-axe est $b = \sqrt{1-|z|}$ (Figure 2.9b).



Enfin, si $z = 0$, les demi-axes de la conique sont égaux, $a = b = 1$, et la conique est un cercle unitaire (Figure 2.9c).

Les segments $x = \pm y$ de la figure 2.9a peuvent être vus comme des ellipses dont le petit demi-axe est nul, $b = 0$, alors que la longueur du grand demi-axe a est, $a = \sqrt{2}$. Le cercle unitaire est également une ellipse dont le petit demi-axe est égal au grand demi-axe, $a = b$. Par conséquent, pour synthétiser, en fixant une corrélation les valeurs admissibles des deux autres sont toujours à l’intérieur d’une ellipse. Les longueurs des demi-axes de cette ellipse sont fonction de la corrélation fixe.

Représentation graphique en 3D. En variant la corrélation fixe, $z = \rho_{23}$, de -1 à 1 , les contours de la représentation 3D des corrélations admissibles sont une succession d’ellipses horizontales dont la direction et la longueur des axes dépendent de la valeur de z (figure 2.10) :

Pour $z \in [-1, 0[$, les demi-axes des ellipses sont obtenus en tournant les axes Ox et Oy , d'un angle $-\frac{\pi}{4}$. La longueur du petit-axe varie de 0 à 1 et la longueur du grand axe varie de $\sqrt{2}$ à 1 (ellipses en tirets de la figure 2.10).

Pour $z = 0$ l'ellipse est un cercle représenté en pointillés sur la figure 2.10.

Pour $z \in]0, 1]$, les demi-axes des ellipses sont obtenus en tournant les axes Ox et Oy , d'un angle de $\frac{\pi}{4}$. La longueur du petit-axe varie de 1 à 0 et la longueur du grand axe varie de 1 à $\sqrt{2}$ (Ellipses en ligne pleine de la figure 2.10).

Puisque ces résultats valent pour les trois corrélations, ainsi selon chaque axe ρ_{12} , ρ_{13} et ρ_{23} il y a une succession d'ellipses.

Les valeurs éligibles des corrélations des bruits sont telles que :

$$1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23} \geq 0 \quad (2.35)$$

2.3.3 Calculs d'estimation et d'information

Estimation. Nous nous intéressons au problème d'estimation de s en fonction des modalités. Les estimateurs optimaux de s au sens de l'erreur quadratique moyenne à partir d'une seule ou de deux modalités sont donnés en (2.2) et (2.4), respectivement, pour $i, j \in \{1, 2, 3\}$. En considérant les trois modalités, l'estimateur optimal s'écrit :

$$\hat{s}_p(x_1, x_2, x_3) = \frac{k_1 x_1 + k_2 x_2 + k_3 x_3}{|\mathbf{\Omega}|}, \quad (2.36)$$

avec :

$$k_1 = (1 + \alpha_3)(\sqrt{\alpha_1} - \sqrt{\alpha_2}\rho_{12}) + (\sqrt{\alpha_1\alpha_3} + \rho_{13})(\sqrt{\alpha_2}\rho_{23} - \sqrt{\alpha_3}) + (\sqrt{\alpha_2\alpha_3} + \rho_{23})(\sqrt{\alpha_3}\rho_{12} - \sqrt{\alpha_1}\rho_{23}),$$

$$k_2 = (1 + \alpha_3)(\sqrt{\alpha_2} - \sqrt{\alpha_1}\rho_{12}) + (\sqrt{\alpha_2\alpha_3} + \rho_{23})(\sqrt{\alpha_1}\rho_{13} - \sqrt{\alpha_3}) - (\sqrt{\alpha_1\alpha_3} + \rho_{13})(\sqrt{\alpha_2}\rho_{13} - \sqrt{\alpha_3}\rho_{12}),$$

$$k_3 = (1 + \alpha_2)(\sqrt{\alpha_3} - \sqrt{\alpha_1}\rho_{13}) + (\sqrt{\alpha_2\alpha_3} + \rho_{23})(\sqrt{\alpha_1}\rho_{12} - \sqrt{\alpha_2}) + (\sqrt{\alpha_1\alpha_2} + \rho_{12})(\sqrt{\alpha_2}\rho_{13} - \sqrt{\alpha_3}\rho_{12}).$$

$\mathbf{\Omega}$ est la matrice de covariance de (x_1, x_2, x_3) (2.27), ainsi son déterminant, noté $|\mathbf{\Omega}|$, est égal à :

$$\begin{aligned} |\mathbf{\Omega}| &= (1 + \alpha_1)(1 + \alpha_2)(1 + \alpha_3) + 2(\sqrt{\alpha_1\alpha_2} + \rho_{12})(\sqrt{\alpha_1\alpha_3} + \rho_{13})(\sqrt{\alpha_2\alpha_3} + \rho_{23}) \\ &\quad - (1 + \alpha_2)(\sqrt{\alpha_1\alpha_3} + \rho_{13})^2 - (1 + \alpha_3)(\sqrt{\alpha_1\alpha_2} + \rho_{12})^2 - (1 + \alpha_1)(\sqrt{\alpha_2\alpha_3} + \rho_{23})^2. \end{aligned} \quad (2.37)$$

L'erreur associée à l'estimateur multimodal $\hat{s}_p(x_1, x_2, x_3)$ est le rapport des déterminants des matrices $\mathbf{\Sigma}$ et $\mathbf{\Omega}$:

$$\mathbf{M}_{p,p}(s|x_1, x_2, x_3) = \frac{|\mathbf{\Sigma}|}{|\mathbf{\Omega}|}, \quad (2.38)$$

où $\mathbf{\Sigma}$ est la matrice de covariance des bruits (2.27), dont le déterminant est :

$$|\mathbf{\Sigma}| = 1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23} \quad (2.39)$$

Démonstration. Voir section A.5 de l'annexe A. □

Information. Les informations communiquées par une et deux modalités sur s sont fournies par les équations (2.6) et (2.7), respectivement, pour $i, j \in \{1, 2, 3\}$. Tandis que, l'information communiquée conjointement par les trois modalités est donnée par l'information mutuelle conjointe entre s et (x_1, x_2, x_3) :

$$I_{p,p}(s; (x_1, x_2, x_3)) = \frac{1}{2} \log \left(\frac{|\Omega|}{|\Sigma|} \right) = \frac{1}{2} \log \left(\mathbf{M}_{p,p}(s|x_1, x_2, x_3)^{-1} \right). \quad (2.40)$$

Démonstration. Voir section B.5 de l'annexe B. □

2.3.4 Une ou trois modalités ?

D'après l'hypothèse $\alpha_1 \geq \alpha_2 \geq \alpha_3$, la modalité x_1 donne une mesure plus précise sur s que x_2 qui, à son tour, est plus précise que x_3 . D'un point de vue information mutuelle, ceci se traduit par : $I_{p,p}(s; x_1) \geq I_{p,p}(s; x_2) \geq I_{p,p}(s; x_3)$. Ainsi, si l'on veut choisir une modalité parmi les trois pour estimer s , il s'agira de x_1 .

Maintenant, pour déterminer si les trois modalités apportent plus d'information sur s que x_1 seule, nous comparons les informations fournies sur s par les trois modalités, $I_{p,p}(s; (x_1, x_2, x_3))$, et l'information que x_1 seule apporte sur s , $I_{p,p}(s; x_1)$. L'information communiquée sur s par les trois modalités peut être décomposée, selon la relation (1.25), comme suit :

$$I_{p,p}(s; (x_1, x_2, x_3)) = I_{p,p}(s; (x_1, x_2)) + I_{p,p}(s; x_3|x_1, x_2) \quad (2.41)$$

$$= I_{p,p}(s; x_1) + I_{p,p}(s; x_2|x_1) + I_{p,p}(s; x_3|x_1, x_2) \quad (2.42)$$

L'équation (2.42) indique que l'information communiquée par les trois modalités sur s est égale à la somme des informations sur s présentes dans x_1 ($I_{p,p}(s; x_1)$), dans x_2 conditionnellement à x_1 ($I_{p,p}(s; x_2|x_1)$) et dans x_3 indépendamment du couple (x_1, x_2) ($I_{p,p}(s; x_3|x_1, x_2)$).

Etant donné que $I_{p,p}(s; x_2|x_1)$ et $I_{p,p}(s; x_3|x_1, x_2)$ sont non négatives alors $I_{p,p}(s; (x_1, x_2, x_3)) \geq I_{p,p}(s; x_1)$. Ainsi, les trois modalités présentent autant ou plus d'information sur s que x_1 , avec égalité si et seulement si $I_{p,p}(s; x_2|x_1) = 0$ et $I_{p,p}(s; x_3|x_1, x_2) = 0$. Puisque $I_{p,p}(s; x_2|x_1) = 0$, la dernière équation équivaut $I_{p,p}(s; x_3|x_1) = 0$, comme nous le montrons ci-dessous :

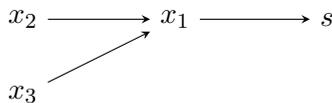
Démonstration. Etant donné que $I_{p,p}(s; x_3|x_1, x_2) = 0$ et $I_{p,p}(s; x_2|x_1) = 0$, alors $p(s, x_3|x_1, x_2) = p(s|x_1, x_2) p(x_3|x_1, x_2)$ et $p(s; x_2|x_1) = p(s|x_1) p(x_2|x_1)$ (paragraphe 1.1.4).

Ainsi :

$$\begin{aligned} p(s, x_3|x_1) &= \int p(s, x_3|x_1, x_2) p(x_2|x_1) dx_2, \\ &= \int p(s|x_1, x_2) p(x_3|x_1, x_2) p(x_2|x_1) dx_2, \\ &= \int p(s, x_2|x_1) p(x_3|x_1, x_2) dx_2, \\ &= p(s|x_1) \int p(x_2|x_1) p(x_3|x_1, x_2) dx_2, \\ &= p(s|x_1) p(x_3|x_1). \end{aligned}$$

Par conséquent, $I_{p,p}(s; x_3|x_1) = 0$. □

La modalité x_1 suffit, alors, à estimer s si et seulement si x_2 (resp. x_3), x_1 et s forment dans ce sens une chaîne de Markov (paragraphe 1.1.4) :



Nous disons dans ce cas que x_2 et x_3 sont redondantes par rapport à x_1 sur s .

Dans le cadre du modèle \mathcal{M}_h^3 , x_2 et x_3 sont redondantes à x_1 sur s si et seulement si $\rho_{12} = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$ et $\rho_{13} = \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_1}}$. Ces valeurs de corrélation sont les solutions des égalités : $I_{p,p}(s; x_2|x_1) = 0$ et $I_{p,p}(s; x_3|x_1) = 0$ (paragraphe 2.2.3).

Pour $\rho_{12} = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$ et $\rho_{13} = \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_1}}$, les facteurs de x_2 et x_3 de l'estimateur optimal $\hat{s}_p(x_1, x_2, x_3)$ (2.36) sont nuls, *i.e* $k_2 = 0$ et $k_3 = 0$. Ainsi, $\hat{s}_p(x_1, x_2, x_3) = \hat{s}_p(x_1)$ et $\mathbf{M}_{p,p}(s|x_1, x_2, x_3) = \mathbf{M}_{p,p}(s|x_1)$.

Dans ce cas de figure, l'intervalle I_e des valeurs éligibles de ρ_{23} est donné par (2.32) en posant $x = \rho_{23}$, $y = \rho_{12} = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$ et $z = \rho_{13} = \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_1}}$:

$$\rho_{23} \in I_e = \left[\frac{\sqrt{\alpha_2\alpha_3} - \sqrt{(\alpha_1 - \alpha_3)(\alpha_1 - \alpha_2)}}{\alpha_1}, \frac{\sqrt{\alpha_2\alpha_3} + \sqrt{(\alpha_1 - \alpha_3)(\alpha_1 - \alpha_2)}}{\alpha_1} \right]. \quad (2.43)$$

Cet intervalle se réduit en un point, $\rho_{23} = \frac{\sqrt{\alpha_2\alpha_3}}{\alpha_1}$, si $\alpha_2 = \alpha_1$ ou si $\alpha_3 = \alpha_1$. Et il se répartit sur $[-1, 1]$ si $\alpha_2 \ll \alpha_1$ et $\alpha_3 \ll \alpha_1$.

$$I_{p,p}(s; (x_1, x_2, x_3)) = I_{p,p}(s; x_1) \Leftrightarrow \rho_{12} = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}} \text{ et } \rho_{13} = \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_1}}. \quad (2.44)$$

Information. Nous considérons le cas de redondance entre x_3 et x_1 sur s , *i.e* $\rho_{13} = \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_1}}$. La figure 2.11a illustre, en fonction de la corrélation ρ_{12} , l'information multimodale $I_{p,p}(s; (x_1, x_2, x_3))$ pour différentes valeurs de ρ_{23} : $\rho_{23} = -0.7$ (---), $\rho_{23} = 0$ (—) et $\rho_{23} = 0.7$ (-.-) et $I_{p,p}(s; x_1)$ (.....).

Nous constatons que, pour chaque valeur de ρ_{23} , $I_{p,p}(s; (x_1, x_2, x_3))$ est représentée seulement sur une partie de l'intervalle $[-1, 1]$. Cette partie correspond aux valeurs éligibles de ρ_{12} pour ρ_{23} et ρ_{13} fixées (2.32).

Pour $\rho_{12} = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$, $I_{p,p}(s; (x_1, x_2, x_3))|_{\rho_{23}=0} = I_{p,p}(s; (x_1, x_2, x_3))|_{\rho_{23}=0.7} = I_{p,p}(s; x_1)$. C'est-à-dire que pour cette valeur de ρ_{12} , puisque $\rho_{13} = \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_1}}$, les trois modalités communiquent autant d'information que x_1 seule. Ainsi elles sont redondantes en information sur s par rapport à x_1 .

Toutefois, $I_{p,p}(s; (x_1, x_2, x_3))|_{\rho_{23}=-0.7} > I_{p,p}(s; x_1)$. C'est-à-dire que, pour $\rho_{23} = -0.7$, les trois modalités sont toujours plus informatives que x_1 seule. C'est parce que ρ_{12} ne peut pas être égale à $\frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$ (pas éligible) car $\rho_{23} = -0.7$ ne fait pas partie de l'intervalle (2.43). Nous pouvons

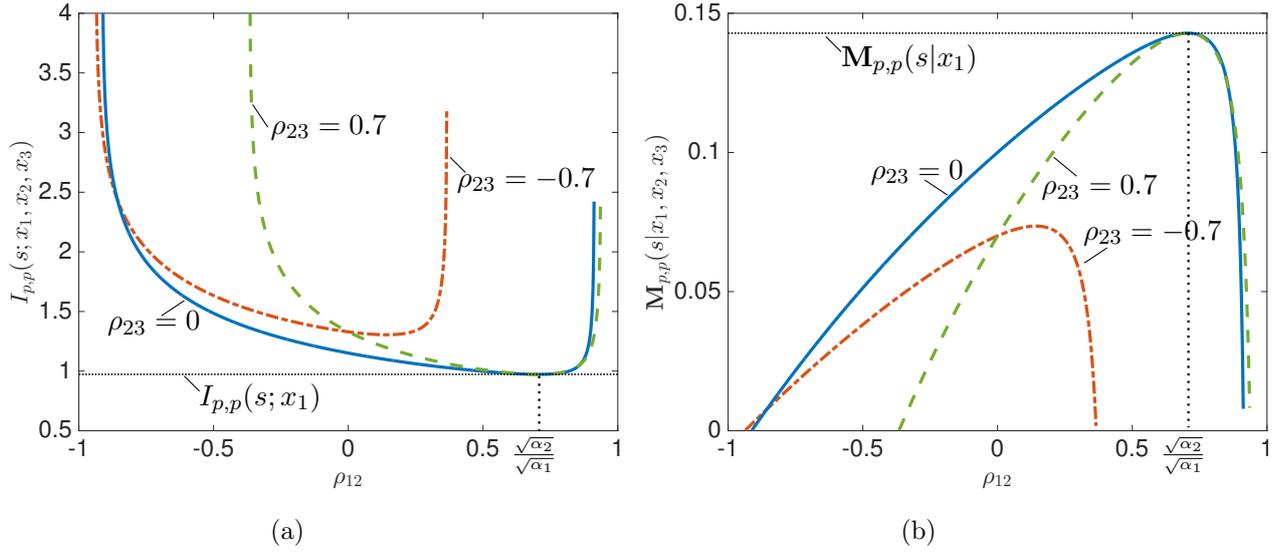


FIGURE 2.11 – Variations de $I_{p,p}(s; (x_1, x_2, x_3))$ (a) et $\mathbf{M}_{p,p}(s|x_1, x_2, x_3)$ (b) en fonction de ρ_{12} pour différentes valeurs de ρ_{23} , où $\rho_{13} = \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_1}}$, $\alpha_1 = 6$, $\alpha_2 = 3$ et $\alpha_3 = 1$.

alors conclure que, les trois modalités sont conjointement strictement plus informatives que la seule modalité x_1 pour toutes valeurs de ρ_{23} en dehors de l'intervalle I_e .

Estimation. Les mêmes remarques valent en termes des erreurs d'estimation qu'illustre la figure 2.11a.

Puisque $\rho_{23} = 0$ et $\rho_{23} = 0.7$ appartiennent à l'intervalle (2.43), la valeur $\rho_{12} = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$ est éligible. Ainsi, les erreurs d'estimation multimodales $\mathbf{M}_{p,p}(s|x_1, x_2, x_3) \Big|_{\rho_{23}=0}$ et $\mathbf{M}_{p,p}(s|x_1, x_2, x_3) \Big|_{\rho_{23}=0.7}$ sont tangentes à $\mathbf{M}_{p,p}(s|x_1)$ au point $\rho_{23} = \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_2}}$.

Réciproquement, pour $\rho_{23} = -0.7$, $\mathbf{M}_{p,p}(s|x_1, x_2, x_3) \Big|_{\rho_{23}=-0.7} > \mathbf{M}_{p,p}(s|x_1)$ car la valeur $\rho_{12} = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$ n'est pas éligible. C'est-à-dire que conjointement les trois modalités sont strictement plus précises que la seule modalité x_1 .

$$\text{Si } \rho_{23} \in I_e \text{ (2.43)} \Rightarrow I_{p,p}(s; (x_1, x_2, x_3)) > I_{p,p}(s; x_1).$$

2.3.5 Deux modalités parmi trois

Nous supposons toujours que nous disposons de trois modalités. En revanche, c'est seulement deux modalités que nous voudrions choisir afin d'estimer s . Lesquelles faut-il choisir ? En termes d'information mutuelle, cela revient à comparer les informations communiquées conjointement par les trois couples, (x_1, x_2) , (x_1, x_3) et (x_2, x_3) , sur s . A première vue, le choix évident est le couple (x_1, x_2) qui possède les rapports signal à bruit les plus élevés. Mais est-ce vraiment le cas ?

2.3.5.1 Mêmes liens

Les mesures d'information et d'estimation dans le cas du modèle \mathcal{M}_h^3 dépendent des corrélations des bruits et des rapports signal à bruit. Afin de mettre l'accent sur les rapports signal à bruit, nous supposons dans un premier temps que toutes les modalités sont similairement liées, *i.e* $\rho_{12} = \rho_{13} = \rho_{23} = \rho$.

Dans le cas où tous les bruits sont indépendants entre eux, *i.e* $\rho = 0$, l'information communiquée sur s par un couple (x_i, x_j) est fonction seulement des rapports signal à bruit :

$$I_{p,p}(s; (x_i, x_j)) = \frac{1}{2} \log(1 + \alpha_i + \alpha_j), \quad i, j \in \{1, 2, 3\}. \quad (2.45)$$

Ainsi, $I_{p,p}(s; (x_1, x_2)) \geq I_{p,p}(s; (x_1, x_3)) \geq I_{p,p}(s; (x_2, x_3))$ selon l'hypothèse $\alpha_1 \geq \alpha_2 \geq \alpha_3$.

Toutefois, pour des bruits corrélés, *i.e* $\rho \neq 0$, ce classement peut varier selon que cette corrélation induit ou non de la redondance. Dans le paragraphe 2.2.4, nous avons vu que certaines corrélations positives des bruits induisent de la redondance entre deux modalités. Cet intervalle de redondance (2.18), que nous avons noté I_{red} , s'étale sur toutes les valeurs positives de corrélation, *i.e* $I_{red} = [0, 1]$, dans le cas de deux modalités avec des rapports signal à bruit égaux. Lorsque les rapports signal à bruit deviennent plus disparates I_{red} se rétrécit, car l'information sur le bruit qu'offre la modalité de faible qualité permet de réduire le bruit.

Pour $i, j, k \in \{1, 2, 3\}$, en comparant $I_{p,p}(s; (x_i, x_j))$ et $I_{p,p}(s; (x_j, x_k))$ nous trouvons que, si $\alpha_k \leq \alpha_i$, (x_j, x_k) communique plus d'information sur s que (x_i, x_j) si et seulement si $\rho \in \left[\frac{\sqrt{\alpha_i} + \sqrt{\alpha_k}}{2\sqrt{\alpha_j}}, 1 \right]$. Cette condition ne peut être remplie que si $\frac{\sqrt{\alpha_i} + \sqrt{\alpha_k}}{2\sqrt{\alpha_j}} \leq 1$, puisque la valeur absolue du coefficient de corrélation est toujours inférieure à 1.

Si $\alpha_k \leq \alpha_i$, alors

$$I_{p,p}(s; (x_i, x_j)) \leq I_{p,p}(s; (x_j, x_k)) \Leftrightarrow \rho \in \left[\frac{\sqrt{\alpha_i} + \sqrt{\alpha_k}}{2\sqrt{\alpha_j}}, 1 \right]. \quad (\rho = \rho_{12} = \rho_{13} = \rho_{23}) \quad (2.46)$$

La figure 2.12a représente, en fonction de ρ , les informations qu'apporte chaque couple de modalités sur s : $I_{p,p}(s; (x_1, x_2))$ (—), $I_{p,p}(s; (x_1, x_3))$ (-.-.-) et $I_{p,p}(s; (x_2, x_3))$ (-.-.-). La figure 2.12b représente, les erreurs d'estimation multimodales : $\mathbf{M}_{p,p}(s|x_1, x_2)$ (—), $\mathbf{M}_{p,p}(s|x_1, x_3)$ (-.-.-) et $\mathbf{M}_{p,p}(s|x_2, x_3)$ (-.-.-).

- (x_1, x_2) ou (x_1, x_3) :

Nous commençons cette étude en comparant les informations fournies sur s par (x_1, x_2) , *i.e* $I_{p,p}(s; (x_1, x_2))$, et (x_1, x_3) , *i.e* $I_{p,p}(s; (x_1, x_3))$. La modalité x_1 étant commune aux deux couples, selon la relation (1.25), cette comparaison est équivalente à comparer les informations additionnelles qu'apportent x_2 conditionnellement à x_1 sur s , $I_{p,p}(s; x_2|x_1)$, et x_3 conditionnellement à x_1 sur s , $I_{p,p}(s; x_3|x_1)$,

$$I_{p,p}(s; (x_1, x_2)) - I_{p,p}(s; (x_1, x_3)) = I_{p,p}(s; x_1) + I_{p,p}(s; x_2|x_1) - (I_{p,p}(s; x_1) + I_{p,p}(s; x_3|x_1))$$

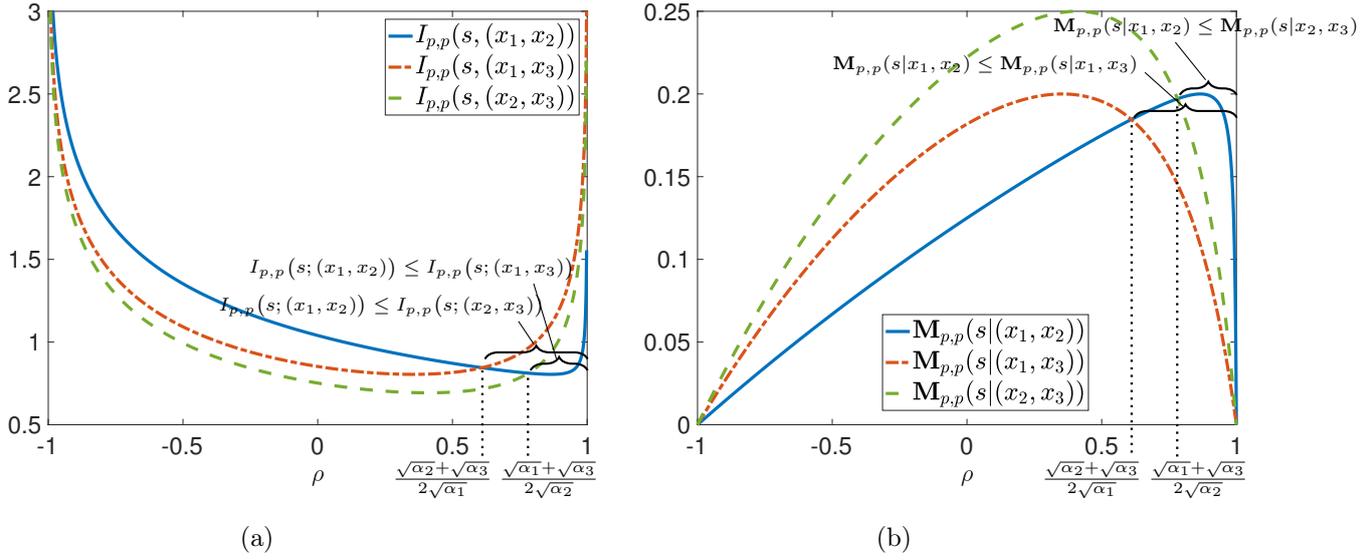


FIGURE 2.12 – Variations des informations mutuelles (a) et des erreurs quadratiques moyennes (b), dans le cas de bruits similairement liés, en fonction de $\rho = \rho_{12} = \rho_{13} = \rho_{23}$ pour $\alpha_1 = 4$, $\alpha_2 = 2$ et $\alpha_3 = 0.5$

$$= I_{p,p}(s; x_2|x_1) - I_{p,p}(s; x_3|x_1) \quad (2.47)$$

Information. Sur la figure 2.12a, nous remarquons que pour $\rho \in \left[-1, \frac{\sqrt{\alpha_2 + \sqrt{\alpha_3}}}{2\sqrt{\alpha_1}} \right]$, $I_{p,p}(s; (x_1, x_2)) \geq I_{p,p}(s; (x_1, x_3))$ car, puisque $\alpha_2 \geq \alpha_3$, x_2 apporte une information plus fiable sur s que x_3 , $I_{p,p}(s; x_3) \leq I_{p,p}(s; x_2)$.

Réciproquement, sur l'intervalle $\left[\frac{\sqrt{\alpha_2 + \sqrt{\alpha_3}}}{2\sqrt{\alpha_1}}, 1 \right]$, $I_{p,p}(s; (x_1, x_2)) \leq I_{p,p}(s; (x_1, x_3))$. Parce que pour de fortes corrélations des bruits, alors que l'information qu'apporte x_2 induit de la redondance, l'information qu'apporte x_3 sur le bruit est bien exploitée pour réduire le bruit.

Estimation. Nous constatons sur la figure 2.12b que $\mathbf{M}_{p,p}(s|x_1, x_2) \leq \mathbf{M}_{p,p}(s|x_1, x_3)$ pour $\rho \in \left[-1, \frac{\sqrt{\alpha_2 + \sqrt{\alpha_3}}}{2\sqrt{\alpha_1}} \right]$. C'est-à-dire que l'estimateur $\hat{s}_p(x_1, x_2)$ est plus précis que l'estimateur $\hat{s}_p(x_1, x_3)$.

Réciproquement, pour $\rho \in \left[\frac{\sqrt{\alpha_2 + \sqrt{\alpha_3}}}{2\sqrt{\alpha_1}}, 1 \right]$, l'estimateur $\hat{s}_p(x_1, x_2)$ est moins précis que l'estimateur $\hat{s}_p(x_1, x_3)$ et $\mathbf{M}_{p,p}(s|x_1, x_2) \geq \mathbf{M}_{p,p}(s|x_1, x_3)$.

- (x_1, x_2) ou (x_2, x_3) :

A présent c'est x_2 qui est commune aux deux couples. Par conséquent, Nous comparons l'information qu'apporte x_1 sur s conditionnellement à x_2 , $I_{p,p}(s; x_1|x_2)$, et l'information qu'apporte x_3 sur s conditionnellement à x_2 , $I_{p,p}(s; x_3|x_2)$:

$$I_{p,p}(s; (x_1, x_2)) - I_{p,p}(s; (x_2, x_3)) = I_{p,p}(s; x_1|x_2) - I_{p,p}(s; x_3|x_2) \quad (2.48)$$

Cette comparaison est assez similaire au cas précédent.

Information. Sur la figure 2.12a, nous constatons que $I_{p,p}(s; (x_1, x_2)) \leq I_{p,p}(s; (x_2, x_3))$ sur l'intervalle $\rho \in \left[\frac{\sqrt{\alpha_1 + \sqrt{\alpha_3}}}{2\sqrt{\alpha_2}}, 1 \right]$. Cette condition n'est vérifiée que si $\sqrt{\alpha_3} \leq 2\sqrt{\alpha_2} - \sqrt{\alpha_1}$. En effet, la qualité de x_3 doit être assez faible pour que l'information qu'elle apporte sur le bruit soit assez précise et assurer plus d'information sur s avec x_2 . Cette information est bien exploitée pour de fortes corrélation des bruits.

Estimation. Nous constatons sur la figure 2.12b que l'estimateur $\hat{s}_p(x_1, x_2)$ est plus précis que l'estimateur $\hat{s}_p(x_1, x_3)$ sur l'intervalle $\rho \in \left[-1, \frac{\sqrt{\alpha_1 + \sqrt{\alpha_3}}}{2\sqrt{\alpha_2}} \right]$.

Réciproquement, pour $\rho \in \left[\frac{\sqrt{\alpha_2 + \sqrt{\alpha_3}}}{2\sqrt{\alpha_1}}, 1 \right]$, l'estimateur $\hat{s}_p(x_1, x_2)$ est moins précis que l'estimateur $\hat{s}_p(x_2, x_3)$.

- (x_1, x_3) ou (x_2, x_3) :

Notre choix de couple est déterminé par les informations additionnelles qu'apportent x_1 ou x_2 sur s conditionnellement à x_3 , puisque x_3 est commune aux deux couples :

$$I_{p,p}(s; (x_1, x_3)) - I_{p,p}(s; (x_2, x_3)) = I_{p,p}(s; x_1|x_2) - I_{p,p}(s; x_3|x_2) \geq 0 \quad (2.49)$$

Information. Nous constatons que $I_{p,p}(s; (x_1, x_3)) \geq I_{p,p}(s; (x_2, x_3))$ pour toutes les valeurs de corrélation. C'est-à-dire que le couple (x_1, x_3) communique toujours plus d'information que (x_2, x_3) sur s . Parce que, selon l'hypothèse $\alpha_1 \geq \alpha_2 \geq \alpha_3$, la condition $\rho \in \left[-1, \frac{\sqrt{\alpha_1 + \sqrt{\alpha_2}}}{2\sqrt{\alpha_3}} \right]$ n'est jamais remplie.

Estimation. Sur la 2.12b, $\forall \rho \in [-1, 1]$, $\mathbf{M}_{p,p}(s|x_1, x_3) \leq \mathbf{M}_{p,p}(s|x_2, x_3)$. C'est-à-dire que l'estimateur $\hat{s}_p(x_1, x_2)$ est toujours plus précis que l'estimateur $\hat{s}_p(x_1, x_3)$, car x_1 apporte l'information la plus fiable sur le signal s et x_3 en apporte sur le bruit.

- **Synthèse :**

Finalement, pour $|\rho| = 1$ les trois couples sont équivalents car ils apportent toute l'information sur s :

$$I_{p,p}(s; (x_1, x_2)) \Big|_{|\rho|=1} = I_{p,p}(s; (x_1, x_3)) \Big|_{|\rho|=1} = I_{p,p}(s; (x_2, x_3)) \Big|_{|\rho|=1} = +\infty \quad (2.50)$$

Pour $\rho < 0$, la corrélation des bruits favorise l'interaction des modalités quelle que soit la qualité des modalités. Puisque, x_1 et x_2 ont les rapports signal à bruit les plus élevés, ils sont conjointement plus informatifs sur s que (x_1, x_3) et (x_2, x_3) .

Réciproquement, les corrélations positives introduisent de la redondance entre les modalités. Dans ce cas, une modalité de faible qualité est plus intéressante car elle fournit moins d'information sur le signal d'intérêt, et ainsi elle introduit moins de redondance. Elle fournit, en revanche, plus d'information sur le bruit qui permet une meilleure compensation du bruit pour de fortes valeurs de ρ . Effectivement, sur l'intervalle $\left[\frac{\sqrt{\alpha_2 + \sqrt{\alpha_3}}}{2\sqrt{\alpha_1}}, 1 \right]$, le couple (x_1, x_3) présente

plus d'information sur s que (x_1, x_2) et (x_2, x_3) .

Parmi trois modalités avec des bruits similairement corrélés, le couple de deux modalités avec les rapports signal à bruit les plus élevés ne communique pas toujours le maximum d'information.

2.3.5.2 Liens différents

Nous avons vu précédemment que, pour des bruits similairement liés, bien que le couple (x_1, x_2) comprenne les modalités de meilleures qualités, il communique moins d'information que les couples (x_1, x_3) et (x_2, x_3) pour certaines corrélations positives. Nous considérons maintenant le cas général où les liens entre les modalités sont quelconques.

Pour $i, j, k \in \{1, 2, 3\}$, en comparant $I_{p,p}(s; (x_i, x_j))$ et $I_{p,p}(s; (x_j, x_k))$ nous trouvons que (x_i, x_j) apporte moins d'information sur s que (x_j, x_k) si et seulement si :

$$\rho_{ij} \in \left[\frac{\sqrt{\alpha_i \alpha_j} (1 - \rho_{jk}^2) - \delta}{(\sqrt{\alpha_k} - \sqrt{\alpha_j} \rho_{jk})^2 + \alpha_j (1 - \rho_{jk}^2)}, \frac{\sqrt{\alpha_i \alpha_j} (1 - \rho_{jk}^2) + \delta}{(\sqrt{\alpha_k} - \sqrt{\alpha_j} \rho_{jk})^2 + \alpha_j (1 - \rho_{jk}^2)} \right], \quad (2.51)$$

où $\delta = (\sqrt{\alpha_k} - \sqrt{\alpha_j} \rho_{jk}) \sqrt{(\sqrt{\alpha_k} - \sqrt{\alpha_j} \rho_{jk})^2 + (\alpha_j - \alpha_i)(1 - \rho_{jk}^2)} \in \mathbb{R}$.

- (x_1, x_2) ou (x_1, x_3) :

Nous comparons tout d'abord, les couples (x_1, x_2) et (x_1, x_3) . Dans le cas $\rho_{12} = \rho_{13} = \rho$, (x_1, x_3) assure plus d'information que (x_1, x_2) pour de fortes corrélations des bruits qui permettent la meilleure exploitation de l'information sur le bruit fournie par x_3 . Cependant pour $\rho_{12} \neq \rho_{13}$, à partir de (2.51), nous obtenons que (x_1, x_3) est plus informatif sur s que (x_1, x_2) , $I_{p,p}(s; (x_1, x_2)) \leq I_{p,p}(s; (x_1, x_3))$ si et seulement si :

$$\rho_{12} \in \left[\frac{\sqrt{\alpha_1 \alpha_2} (1 - \rho_{13}^2) - \delta}{(\sqrt{\alpha_3} - \sqrt{\alpha_1} \rho_{13})^2 + \alpha_1 (1 - \rho_{13}^2)}, \frac{\sqrt{\alpha_1 \alpha_2} (1 - \rho_{13}^2) + \delta}{(\sqrt{\alpha_3} - \sqrt{\alpha_1} \rho_{13})^2 + \alpha_1 (1 - \rho_{13}^2)} \right], \quad (2.52)$$

où $\delta = (\sqrt{\alpha_3} - \sqrt{\alpha_1} \rho_{13}) \sqrt{(\sqrt{\alpha_3} - \sqrt{\alpha_1} \rho_{13})^2 + (\alpha_1 - \alpha_2)(1 - \rho_{13}^2)}$. L'intervalle (2.52) dépend de la valeur de ρ_{13} .

La figure 2.13 illustre la variation de cet intervalle selon le paramètre ρ_{13} . Si $\rho_{13} = \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_1}}$, l'intervalle (2.52) est réduit en un point qui correspond à $\rho_{12} = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$. La longueur de l'intervalle augmente en s'éloignant de $\rho_{13} = \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_1}}$ et s'étale sur $[-1, 1]$ pour $|\rho_{13}| = 1$.

Information. La figure 2.14 illustre $I_{p,p}(s; (x_1, x_2))$ (.....) et $I_{p,p}(s; (x_1, x_3))$ pour différentes valeurs de ρ_{13} : $\rho_{13} = -0.7$ (---), $\rho_{13} = \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_1}}$ (—) et $\rho_{13} = 0.7$ (-.-).

Pour $\rho_{13} = \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_1}}$, x_3 n'apporte aucune information supplémentaire sur s comparé à x_1 . Ainsi, $I_{p,p}(s; (x_1, x_3)) = I_{p,p}(s; x_1)$. $I_{p,p}(s; x_1)$ est la valeur minimale de $I_{p,p}(s; (x_1, x_2))$ qui n'est atteinte que quand x_2 est totalement redondante sur s comparé à x_1 . C'est le cas si $\rho_{12} = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$. Nous

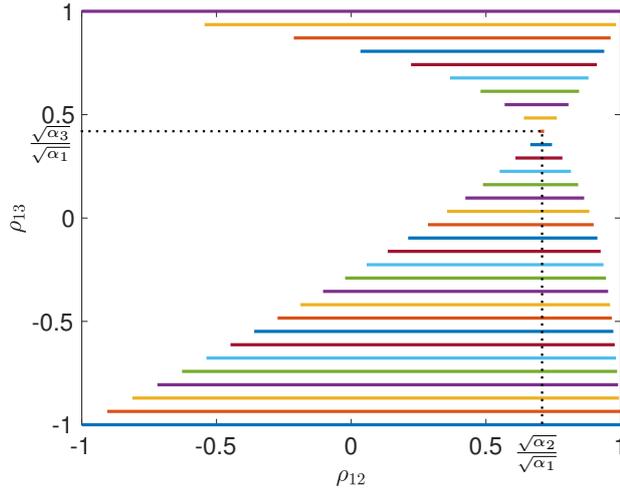


FIGURE 2.13 – Les valeurs de ρ_{12} telles que $I_{p,p}(s; (x_1, x_2)) \leq I_{p,p}(s; (x_1, x_3))$, selon les valeurs de ρ_{13} .

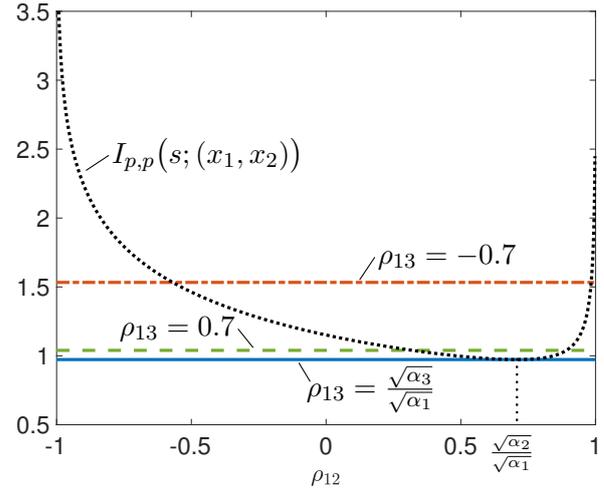


FIGURE 2.14 – $I_{p,p}(s; (x_1, x_2))$ (en pointillés) et $I_{p,p}(s; (x_1, x_3))$ (en droites horizontales) pour différentes valeurs de ρ_{13} , en fonction de ρ_{12} .

constatons que, effectivement $I_{p,p}(s; (x_1, x_2))$ est tangente à $I_{p,p}(s; (x_1, x_3))$ en $\rho_{12} = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$ pour $\rho_{13} = \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_1}}$.

Nous remarquons que pour $\rho_{13} = 0.7$ et $\rho_{13} = -0.7$, $I_{p,p}(s; (x_1, x_3))$ est supérieure à $I_{p,p}(s; (x_1, x_2))$ pour certaines valeurs de ρ_{12} . Parce que, pour $\rho_{13} \neq \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_1}}$, les modalités x_1 et x_3 deviennent moins redondantes et révèlent de l'information sur s qui peut surpasser $I_{p,p}(s; (x_1, x_2))$. Cette information est plus importante pour les corrélations négatives qui favorisent le plus l'interaction des modalités.

- (x_1, x_2) ou (x_2, x_3)

Considérons maintenant, les couples de modalités (x_1, x_2) et (x_2, x_3) . Le couple (x_2, x_3) présente plus d'information que (x_1, x_2) , i.e $I_{p,p}(s; (x_1, x_2)) \leq I_{p,p}(s; (x_2, x_3))$, si :

$$\rho_{12} \in \left[\frac{\sqrt{\alpha_1 \alpha_2} (1 - \rho_{23}^2) - \delta}{(\sqrt{\alpha_3} - \sqrt{\alpha_2} \rho_{23})^2 + \alpha_2 (1 - \rho_{23}^2)}, \frac{\sqrt{\alpha_1 \alpha_2} (1 - \rho_{23}^2) + \delta}{(\sqrt{\alpha_3} - \sqrt{\alpha_2} \rho_{23})^2 + \alpha_2 (1 - \rho_{23}^2)} \right]. \quad (2.53)$$

où $\delta = (\sqrt{\alpha_3} - \sqrt{\alpha_2} \rho_{23}) \sqrt{(\sqrt{\alpha_3} - \sqrt{\alpha_2} \rho_{23})^2 + (\alpha_2 - \alpha_1)(1 - \rho_{23}^2)}$. Cependant, δ est réel si et seulement si :

$$\rho_{23} \in \left[-1, \frac{\sqrt{\alpha_2 \alpha_3} - \sqrt{(\alpha_1 - \alpha_3)(\alpha_1 - \alpha_2)}}{\alpha_1} \right] \cup \left[\frac{\sqrt{\alpha_2 \alpha_3} + \sqrt{(\alpha_1 - \alpha_3)(\alpha_1 - \alpha_2)}}{\alpha_1}, 1 \right], \quad (2.54)$$

Ce qui signifie que (x_2, x_3) ne peut assurer plus d'information que (x_1, x_2) , que pour des corrélations très fortes des bruits. La figure 2.15, présente la variation de l'intervalle (2.53) selon les valeurs de ρ_{23} .

Information. Sur la figure 2.16, nous montrons $I_{p,p}(s; (x_1, x_2))$ (.....) en fonction de ρ_{12} . En outre, $I_{p,p}(s; (x_2, x_3))$ est illustrée, pour différentes valeurs de ρ_{23} : $\rho_{23} = -0.7$ (---), $\rho_{23} = \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_1}}$ (—) et $\rho_{23} = 0.7$ (-.-).

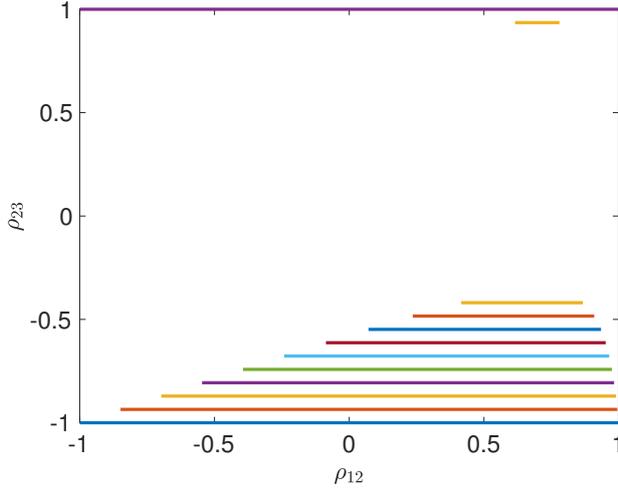


FIGURE 2.15 – Les valeurs de ρ_{12} telles que $I_{p,p}(s; (x_1, x_2)) \leq I_{p,p}(s; (x_2, x_3))$, selon les valeurs de ρ_{23} .

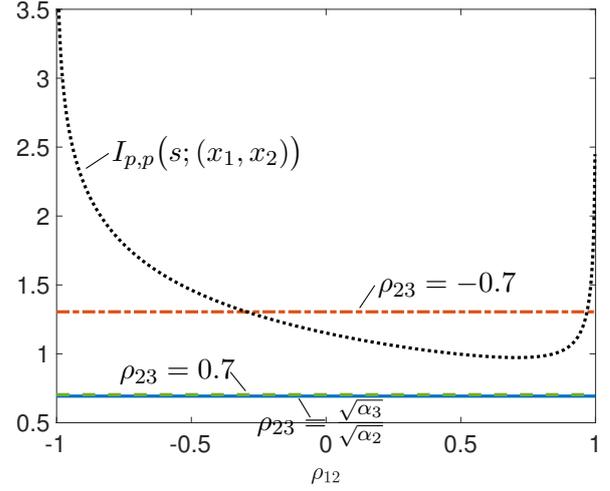


FIGURE 2.16 – $I_{p,p}(s; (x_1, x_2))$ (en pointillés) et $I_{p,p}(s; (x_2, x_3))$ (en droites horizontales) pour différentes valeurs de ρ_{23} , en fonction de ρ_{12} .

Si $\rho_{23} = \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_2}}$, x_3 est totalement redondante à x_2 sur s . Le couple (x_2, x_3) est alors équivalent à x_2 , $I_{p,p}(s; (x_2, x_3)) = I_{p,p}(s; x_2)$. Toutefois, l'information que présente x_2 sur s est toujours strictement inférieure à l'information que présente le couple (x_1, x_2) , ainsi $I_{p,p}(s; (x_1, x_2)) > I_{p,p}(s; (x_2, x_3)) \Big|_{\rho_{23} = \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_2}}}$.

Si $\rho_{23} = 0.7$, même si les bruits sont fortement corrélés et les modalités sont moins redondantes, l'information que (x_2, x_3) communique sur s est légèrement supérieure à $I_{p,p}(s; x_2)$. Ainsi, $I_{p,p}(s; (x_1, x_2)) > I_{p,p}(s; (x_2, x_3)) \Big|_{\rho_{23} = 0.7}$.

Réciproquement, pour $\rho_{23} = -0.7$, (x_2, x_3) devient plus informatif sur s et surpasse $I_{p,p}(s; (x_1, x_2))$ pour certaines valeurs de ρ_{12} .

Par conséquent, même si le couple (x_2, x_3) est le couple de très faible rapport signal à bruit, il peut assurer plus d'information que (x_1, x_2) pour des corrélations très fortes des bruits surtout pour les corrélations négatives.

- (x_1, x_3) ou (x_2, x_3)

Finalement, nous comparons les couples (x_1, x_3) ou (x_2, x_3) . A partir de (2.51) nous obtenons que (x_2, x_3) présente plus d'information que (x_1, x_3) si :

$$\rho_{13} \in \left[\frac{\sqrt{\alpha_1 \alpha_3} (1 - \rho_{23}^2) - \delta}{(\sqrt{\alpha_2} - \sqrt{\alpha_3} \rho_{23})^2 + \alpha_3 (1 - \rho_{23}^2)}, \frac{\sqrt{\alpha_1 \alpha_3} (1 - \rho_{23}^2) + \delta}{(\sqrt{\alpha_3} - \sqrt{\alpha_2} \rho_{23})^2 + \alpha_3 (1 - \rho_{23}^2)} \right], \quad (2.55)$$

où $\delta = (\sqrt{\alpha_2} - \sqrt{\alpha_3} \rho_{23}) \sqrt{(\sqrt{\alpha_2} - \sqrt{\alpha_3} \rho_{23})^2 + (\alpha_3 - \alpha_1)(1 - \rho_{23}^2)} \in \mathbb{R}$. Similairement au cas précédent, (x_2, x_3) peut communiquer plus d'information sur s que (x_1, x_3) uniquement pour de fortes corrélations des bruits, notamment si

$$\rho_{23} \in \left[-1, \frac{\sqrt{\alpha_2 \alpha_3} - \sqrt{(\alpha_2 - \alpha_3)(\alpha_1 - \alpha_2)}}{\alpha_1} \right] \cup \left[\frac{\sqrt{\alpha_2 \alpha_3} + \sqrt{(\alpha_2 - \alpha_3)(\alpha_1 - \alpha_2)}}{\alpha_1}, 1 \right]. \quad (2.56)$$

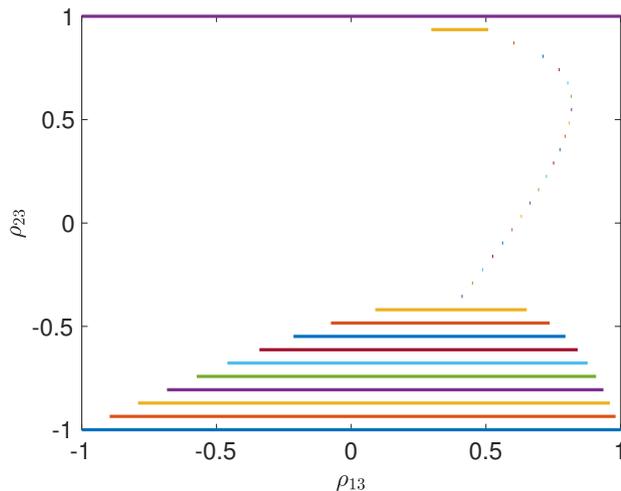


FIGURE 2.17 – Les valeurs de ρ_{12} telles que $I_{p,p}(s; (x_1, x_2)) \leq I_{p,p}(s; (x_2, x_3))$, selon les valeurs de ρ_{23} .

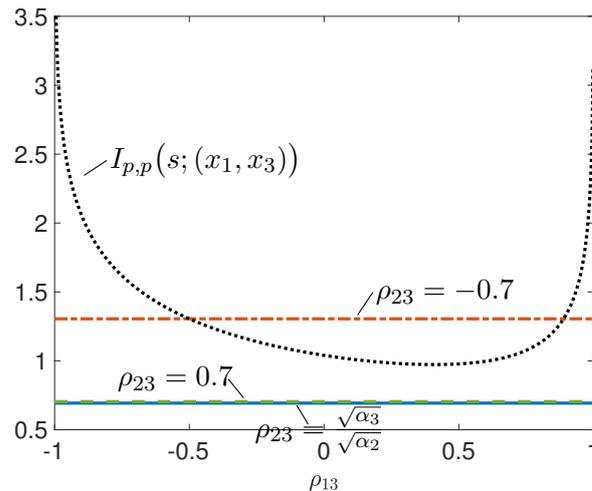


FIGURE 2.18 – $I_{p,p}(s; (x_1, x_2))$ (en pointillés) et $I_{p,p}(s; (x_2, x_3))$ (en droites horizontales) pour différentes valeurs de ρ_{23} , en fonction de ρ_{12} .

La figure 2.17, présente la variation de l'intervalle (2.55) selon les valeurs de ρ_{23} . Nous constatons que le couple (x_2, x_3) est susceptible de communiquer plus d'information sur s , quand les bruits sont négativement corrélés.

Information. Sur la figure 2.18, nous montrons $I_{p,p}(s; (x_1, x_3))$ (.....) en fonction de ρ_{13} . En outre, $I_{p,p}(s; (x_2, x_3))$ est illustrée, pour différentes valeurs de ρ_{23} : $\rho_{23} = -0.7$ (---), $\rho_{23} = \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_1}}$ (—) et $\rho_{23} = 0.7$ (-.-).

Nous tirons les mêmes conclusions que précédemment. Nous remarquons que, $I_{p,p}(s; (x_2, x_3)) < I_{p,p}(s; (x_1, x_3))$ pour $\rho_{23} = \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_1}}$ et $\rho_{23} = 0.7$. En fait, pour ces valeurs de corrélation, le couple (x_2, x_3) est moins informatif que la modalité x_1 seule. Au contraire, pour une corrélation négative ($\rho_{23} = -0.7$), la paire de modalités (x_2, x_3) peut apporter plus d'information que (x_1, x_3) , pour certaines valeurs de ρ_{13} .

Choisir deux modalités parmi trois est simple tant que les bruits sont indépendants ou similairement corrélés. Cependant, si les bruits sont arbitrairement corrélés, les conditions du choix optimal dépendent de plusieurs paramètres, à savoir les rapports signal à bruit et les corrélations des bruits, ce qui rend la tâche plus difficile. Cependant, ce qui est certain, c'est que nous devons privilégier les modalités avec des bruits négativement corrélés.

2.3.6 Deux ou trois modalités ?

Dans ce paragraphe nous examinons les conditions pour lesquelles deux modalités suffisent à tirer le maximum de profit de la multimodalité. Nous proposons de déterminer les cas où une troisième modalité x_j est totalement redondante aux deux autres modalités x_i et x_k , $i, j, k \in \{1, 2, 3\}$. A cette fin, nous comparons l'information conjointe qu'apportent les trois modalités, $I_{p,p}(s; (x_1, x_2, x_3))$, et l'information communiquée par les deux modalités x_i et x_j ,

$I_{p,p}(s; (x_i, x_j))$.

Nous pouvons décomposer $I_{p,p}(s; (x_1, x_2, x_3))$ par la règle de la chaîne en la somme de l'information conjointe que présentent deux modalités sur s , $I_{p,p}(s; (x_i, x_j))$, et l'information que présente la troisième modalité conditionnellement aux deux autres sur s , $I_{p,p}(s; x_k|x_i, x_j)$:

$$I_{p,p}(s; (x_1, x_2, x_3)) = I_{p,p}(s; (x_i, x_j)) + I_{p,p}(s; x_k|x_i, x_j) \quad (2.57)$$

Si $I_{p,p}(s; x_k|x_i, x_j) = 0$, alors x_k n'apporte pas de nouveauté sur s comparé à (x_i, x_j) . Nous disons dans ce cas que x_k est redondante sur s par rapport au couple (x_i, x_j) . C'est-à-dire que le triplet (x_1, x_2, x_3) est aussi précis à l'estimation de s que le couple (x_i, x_j) . Ainsi, x_i et x_j conjointement suffisent à estimer s .

Nous supposons que les deux modalités x_i et x_j sont conjointement plus informatives sur s qu'une seule modalité, $I_{p,p}(s; (x_i, x_j)) > I_{p,p}(s; x_i)$ et $I_{p,p}(s; (x_i, x_j)) > I_{p,p}(s; x_j)$. C'est-à-dire que x_j n'est pas redondante à x_i sur s , i.e $\rho_{ij} \neq \frac{\sqrt{\alpha_j}}{\sqrt{\alpha_i}}$, $\alpha_j \leq \alpha_i$. Dans le cadre du modèle \mathcal{M}_h^3 , en résolvant l'égalité $I_{p,p}(s; (x_1, x_2, x_3)) = I_{p,p}(s; (x_i, x_j))$, nous obtenons que la prise en compte de x_k est inutile si ρ_{ik} et ρ_{jk} sont liés comme suit :

$$\rho_{jk} = \frac{(\sqrt{\alpha_j}\rho_{ij} - \sqrt{\alpha_i})}{(\sqrt{\alpha_j} - \sqrt{\alpha_i}\rho_{ij})} \rho_{ik} + \frac{\sqrt{\alpha_k}(1 - \rho_{ij}^2)}{(\sqrt{\alpha_j} - \sqrt{\alpha_i}\rho_{ij})}. \quad (2.58)$$

Dans ce cas, les valeurs admissibles de ρ_{ik} sont dans l'intervalle :

$$\rho_{ik} \in \left[\frac{\sqrt{\alpha_i\alpha_k}(1 - \rho_{ij}^2) - (\sqrt{\alpha_i}\rho_{ij} - \sqrt{\alpha_j})\sqrt{(\sqrt{\alpha_i}\rho_{ij} - \sqrt{\alpha_j})^2 + (\alpha_i - \alpha_k)(1 - \rho_{ij}^2)}}{(\sqrt{\alpha_i}\rho_{ij} - \sqrt{\alpha_j})^2 + \alpha_i(1 - \rho_{ij}^2)}, \frac{\sqrt{\alpha_i\alpha_k}(1 - \rho_{ij}^2) + (\sqrt{\alpha_i}\rho_{ij} - \sqrt{\alpha_j})\sqrt{(\sqrt{\alpha_i}\rho_{ij} - \sqrt{\alpha_j})^2 + (\alpha_i - \alpha_k)(1 - \rho_{ij}^2)}}{(\sqrt{\alpha_i}\rho_{ij} - \sqrt{\alpha_j})^2 + \alpha_i(1 - \rho_{ij}^2)} \right]. \quad (2.59)$$

D'après (2.58), ρ_{jk} est une fonction affine de ρ_{ik} , $\rho_{jk} = c \rho_{ik} + d$, dont le coefficient directeur, $c = \frac{\sqrt{\alpha_j}\rho_{ij} - \sqrt{\alpha_i}}{\sqrt{\alpha_j} - \sqrt{\alpha_i}\rho_{ij}}$, et le coefficient à l'origine, $d = \frac{\sqrt{\alpha_k}(1 - \rho_{ij}^2)}{\sqrt{\alpha_j} - \sqrt{\alpha_i}\rho_{ij}}$, dépendent de ρ_{ij} . Le coefficient directeur est négatif pour $\rho_{ij} \geq \frac{\sqrt{\alpha_j}}{\sqrt{\alpha_i}}$. Réciproquement, $c \geq 0$ pour $\rho_{ij} \leq \frac{\sqrt{\alpha_j}}{\sqrt{\alpha_i}}$. Le coefficient à l'origine est nul pour $\rho_{ij} = \pm 1$.

Si $\rho_{ij} = \pm 1$, alors (2.58) se réduit à $\rho_{jk} = \pm \rho_{ik}$ et (2.59) devient $\rho_{ik} \in [-1, 1]$. La relation $\rho_{jk} = \pm \rho_{ik}$ représente les valeurs éligibles de ρ_{jk} et ρ_{ik} pour $\rho_{ij} = \pm 1$. Par conséquent, x_k est redondante à (x_i, x_j) pour toutes les valeurs admissibles de ρ_{jk} et ρ_{ik} si $\rho_{ij} = \pm 1$. Parce que, le couple (x_i, x_j) communique toute l'information sur s , $I_{p,p}(s; (x_i, x_j)) = +\infty$.

Si $\rho_{12} = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$ la modalité x_2 est redondante sur s par rapport à x_1 , $I_{p,p}(s; (x_1, x_2)) = I_{p,p}(s; x_1)$. Par conséquent, la modalité x_3 est bénéfique tant que $\rho_{13} \neq \frac{\sqrt{\alpha_3}}{\sqrt{\alpha_1}}$. C'est-à-dire, tant que x_3 n'est pas aussi redondante sur s par rapport à x_1 .

Pour $i = 1, j = 2$ et $k = 3$, nous représentons sur la figure 2.19 les droites (2.58) pour différentes valeurs de ρ_{12} .

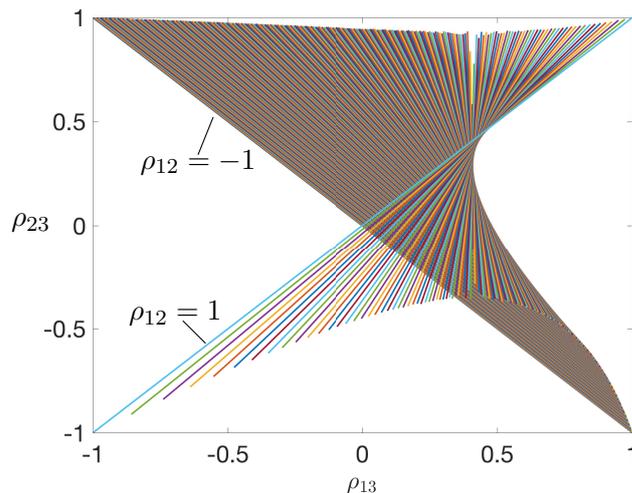


FIGURE 2.19 – Les conditions de redondance de x_3 à (x_1, x_2) , pour différentes valeurs de ρ_{12} .

2.3.7 Effet des corrélations des bruits

Les liens entre les modalités peuvent susciter l'interaction entre les données comme ils peuvent introduire de la redondance. En vue de bien profiter de la multimodalité, quels liens doivent exister entre les modalités ? Autrement dit, quelles sont les corrélations des bruit qui assurent le maximum d'information ? Pour y répondre, similairement au modèle \mathcal{M}_h^2 , nous comparons deux modèles : le modèle avec des bruits corrélés, \mathcal{M}_h^3 , et le modèle avec des bruits non corrélés et ainsi indépendants selon l'hypothèse de gaussianité, \mathcal{M}_{ho}^3 . Dans le cas de deux modalités, la redondance est possible pour des corrélations positives des bruits, notamment pour $\rho_{ij} \in \left] 0, \frac{2\sqrt{\alpha_i\alpha_j}}{\alpha_i+\alpha_j} \right[$. Comment peut-on généraliser ce résultat dans le cas de trois modalités ?

L'information qu'apportent les modalités sur s dans le cadre du modèle \mathcal{M}_{ho}^3 est obtenue à partir de (2.40) en mettons toutes les corrélations à zéro. Ainsi :

$$I_{p,p}(s; (x_1, x_2, x_3)) \Big|_{\rho_{12}=\rho_{13}=\rho_{23}=0} = \frac{1}{2} \log(1 + \alpha_1 + \alpha_2 + \alpha_3). \quad (2.60)$$

De même, l'estimateur optimal en fonction des trois modalités du modèle \mathcal{M}_{ho}^3 est obtenu à partir de (2.36) pour $\rho_{12} = \rho_{13} = \rho_{23} = 0$:

$$\hat{s}_p(x_1, x_2, x_3) \Big|_{\rho_{12}=\rho_{13}=\rho_{23}=0} = \frac{\sqrt{\alpha_1} x_1 + \sqrt{\alpha_2} x_2 + \sqrt{\alpha_3} x_3}{1 + \alpha_1 + \alpha_2 + \alpha_3}. \quad (2.61)$$

L'erreur d'estimation multimodale engendrée par le modèle \mathcal{M}_{ho}^3 , est :

$$\mathbf{M}_{p,p}(s|x_1, x_2, x_3) \Big|_{\rho_{12}=\rho_{13}=\rho_{23}=0} = \frac{1}{1 + \alpha_1 + \alpha_2 + \alpha_3}. \quad (2.62)$$

Lorsque les modalités du modèle \mathcal{M}_{ho}^3 communiquent plus d'information que les modalités du modèle \mathcal{M}_h^3 , nous concluons que la corrélation des bruits introduit de la redondance entre les modalités. Autrement, si $I_{p,p}(s; (x_1, x_2, x_3)) \geq I_{p,p}(s; (x_1, x_2, x_3)) \Big|_{\rho_{12}=\rho_{13}=\rho_{23}=0}$ alors la corrélation des bruits favorise l'interaction ce qui permet de révéler plus d'information sur s .

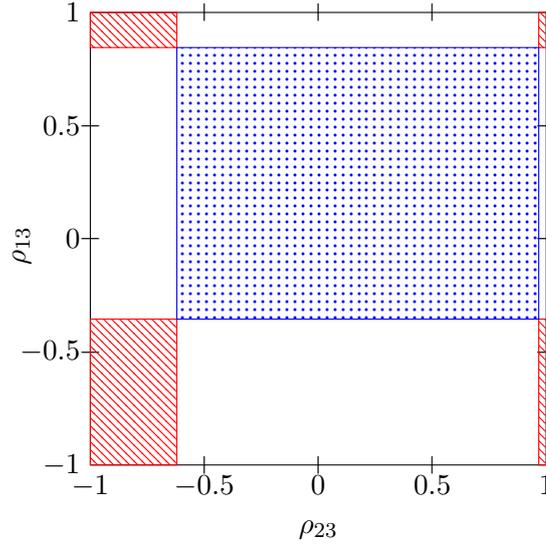


FIGURE 2.20 – Valeurs de ρ_{13} et ρ_{23} qui peuvent introduire de la redondance, où $\alpha_1 = 6$, $\alpha_2 = 3$ et $\alpha_3 = 1$.

A partir de (2.60) et (2.40), nous trouvons que $I_{p,p}(s; (x_1, x_2, x_3)) \leq I_{p,p}(s; (x_1, x_2, x_3)) \Big|_{\rho_{12}=\rho_{13}=\rho_{23}=0}$ si et seulement si :

$$\rho_{12} \in \left[\frac{\rho_{13}((\alpha_1 + \alpha_2 + \alpha_3)\rho_{23} - \sqrt{\alpha_2\alpha_3}) + \sqrt{\alpha_1}(\sqrt{\alpha_2} - \sqrt{\alpha_3}\rho_{23}) - \sqrt{\Delta}}{\alpha_1 + \alpha_2}, \frac{\rho_{13}((\alpha_1 + \alpha_2 + \alpha_3)\rho_{23} - \sqrt{\alpha_2\alpha_3}) + \sqrt{\alpha_1}(\sqrt{\alpha_2} - \sqrt{\alpha_3}\rho_{23}) + \sqrt{\Delta}}{\alpha_1 + \alpha_2} \right] \quad (2.63)$$

où $\Delta = ((\sqrt{\alpha_3}\rho_{23} - \sqrt{\alpha_2})^2 - (1 - \rho_{23}^2)(\alpha_1 + \alpha_3))((\sqrt{\alpha_3}\rho_{13} - \sqrt{\alpha_1})^2 - (1 - \rho_{13}^2)(\alpha_2 + \alpha_3))$. Même si cette condition est donnée pour ρ_{12} , elle dépend des autres corrélations ρ_{13} et ρ_{23} . Cette condition est valable pour $\Delta \geq 0$, c'est-à-dire si :

$$\left\{ \begin{array}{l} \rho_{23} \in \left[-1, \frac{\sqrt{\alpha_2\alpha_3} - \sqrt{(\alpha_1 + \alpha_2)(\alpha_1 + \alpha_3)}}{\alpha_1 + \alpha_2 + \alpha_3} \right] \cup \left[\frac{\sqrt{\alpha_2\alpha_3} + \sqrt{(\alpha_1 + \alpha_2)(\alpha_1 + \alpha_3)}}{\alpha_1 + \alpha_2 + \alpha_3}, 1 \right] \\ \rho_{13} \in \left[-1, \frac{\sqrt{\alpha_1\alpha_3} - \sqrt{(\alpha_1 + \alpha_2)(\alpha_2 + \alpha_3)}}{\alpha_1 + \alpha_2 + \alpha_3} \right] \cup \left[\frac{\sqrt{\alpha_1\alpha_3} + \sqrt{(\alpha_1 + \alpha_2)(\alpha_2 + \alpha_3)}}{\alpha_1 + \alpha_2 + \alpha_3}, 1 \right] \end{array} \right. , \quad (2.64)$$

ou si

$$\left\{ \begin{array}{l} \rho_{23} \in \left[\frac{\sqrt{\alpha_2\alpha_3} - \sqrt{(\alpha_1 + \alpha_2)(\alpha_1 + \alpha_3)}}{\alpha_1 + \alpha_2 + \alpha_3}, \frac{\sqrt{\alpha_2\alpha_3} + \sqrt{(\alpha_1 + \alpha_2)(\alpha_1 + \alpha_3)}}{\alpha_1 + \alpha_2 + \alpha_3} \right] \\ \rho_{13} \in \left[\frac{\sqrt{\alpha_1\alpha_3} - \sqrt{(\alpha_1 + \alpha_2)(\alpha_2 + \alpha_3)}}{\alpha_1 + \alpha_2 + \alpha_3}, \frac{\sqrt{\alpha_1\alpha_3} + \sqrt{(\alpha_1 + \alpha_2)(\alpha_2 + \alpha_3)}}{\alpha_1 + \alpha_2 + \alpha_3} \right] \end{array} \right. , \quad (2.65)$$

Les valeurs de ρ_{13} et ρ_{23} telles que $\Delta \geq 0$ sont représentées sur la figure 2.20 en surfaces hachurées et en pointillés. Cependant, les valeurs de l'intervalle (2.64), présentées par des surfaces hachurées sur la figure 2.20, ne vérifient pas la condition d'éligibilité des corrélations (2.29). Par conséquent, seules les valeurs de ρ_{13} et ρ_{23} de l'intervalle (2.65), présentées en pointillés, peuvent introduire de la redondance.

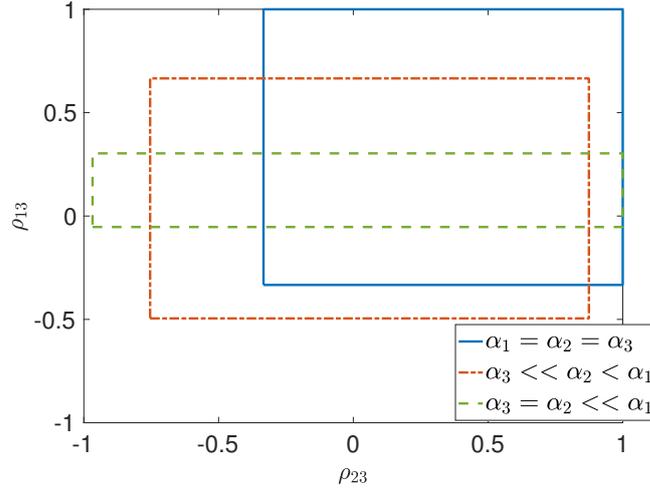


FIGURE 2.21 – Les carrés délimitent les valeurs de ρ_{13} et ρ_{23} qui peuvent introduire de la redondance entre les modalités, pour différentes valeurs des rapports signal à bruit :

$$\alpha_3 = \alpha_2 = \alpha_1, \alpha_3 \ll \alpha_2 < \alpha_1 \text{ et } \alpha_3 = \alpha_2 \ll \alpha_1.$$

Les carrés de la figure 2.21 délimitent les valeurs de ρ_{13} et ρ_{23} qui peuvent introduire de la redondance entre les modalités, c'est-à-dire les valeurs des conditions (2.65), pour différentes valeurs des rapports signal à bruit : $\alpha_3 = \alpha_2 = \alpha_1$ (—), $\alpha_3 \ll \alpha_2 < \alpha_1$ (---) et $\alpha_3 = \alpha_2 \ll \alpha_1$ (-.-).

- $\alpha_1 = \alpha_2 = \alpha_3$

Similairement au cas de deux modalités, toutes les corrélations positives du bruit engendrent de la redondance entre les modalités. Mais pas uniquement, certaines corrélations négatives rendent les modalités partiellement redondantes. Les valeurs de ρ_{23} et ρ_{13} qui peuvent introduire de la redondance entre les modalités sont obtenues à partir de (2.65) :

$$\rho_{23} \in \left[\frac{-2}{3}, 1 \right] \quad \text{et} \quad \rho_{13} \in \left[\frac{-2}{3}, 1 \right]. \quad (2.66)$$

- $\alpha_3 \ll \alpha_2 < \alpha_1$

Si une des trois modalités a un rapport signal à bruit négligeable par rapport aux deux autres modalités, alors la redondance a lieu pour les faibles valeurs de corrélation. Parce que pour de fortes corrélations l'information que x_3 apporte sur le bruit est exploitée. Si $\alpha_3 \rightarrow 0$, alors les conditions (2.65) deviennent :

$$\begin{cases} \rho_{23} \in \left[\frac{-\sqrt{(\alpha_1 + \alpha_2)\alpha_1}}{\alpha_1 + \alpha_2}, \frac{\sqrt{(\alpha_1 + \alpha_2)\alpha_1}}{\alpha_1 + \alpha_2} \right] \\ \rho_{13} \in \left[\frac{-\sqrt{(\alpha_1 + \alpha_2)\alpha_2}}{\alpha_1 + \alpha_2}, \frac{\sqrt{(\alpha_1 + \alpha_2)\alpha_2}}{\alpha_1 + \alpha_2} \right] \end{cases}, \quad (2.67)$$

Nous constatons que les corrélations négatives et positives deviennent équivalentes dans le sens où elles sont susceptibles de rendre les modalités redondantes.

- $\alpha_3 = \alpha_2 \ll \alpha_1$

Finalement, si deux modalités parmi trois ont de très faibles rapports signal à bruit alors les modalités deviennent moins redondantes et révèlent toujours plus d'information sur s . Si $\alpha_2 \rightarrow 0$ et $\alpha_3 \rightarrow 0$, alors les conditions (2.65) deviennent :

$$\rho_{13} \rightarrow 0 \quad \forall \rho_{23}. \quad (2.68)$$

Dans le cas de trois modalités, non seulement les corrélations positives induisent de la redondance mais aussi certaines corrélations négatives notamment les faibles valeurs. Par contre, lorsque les trois bruits sont positivement corrélés, le risque qu'ils induisent de la redondance est plus élevé.

Si nous excluons tous les cas de redondance qui peuvent y avoir lieu entre les trois modalités (sous-section 2.3.6), x_1 , x_2 et x_3 peuvent assurer toute l'information sur s si et seulement si $I_{p,p}(s; (x_1, x_2, x_3)) = +\infty$ et $\mathbf{M}_{p,p}(s|x_1, x_2, x_3) = 0$. C'est le cas si,

$$\rho_{12}^2 + \rho_{13}^2 - 2\rho_{12}\rho_{13}\rho_{23} + \rho_{23}^2 - 1 = 0 \quad (2.69)$$

Nous parlons, dans ce cas, d'estimation singulière puisque la matrice de covariance des bruits est singulière, *i.e* $|\Sigma| = 0$.

Dans le cas de trois modalités, les valeurs négatives de corrélation peuvent engendrer de la redondance.

2.4 Conclusion

Dans ce chapitre, nous avons étudié la question suivante : Comment les liens entre les modalités peuvent-ils améliorer ou diminuer la précision de l'estimation et le contenu informationnel des modalités ? Nous avons considéré deux modèles, à deux ou trois modalités, et nous avons étudié principalement le rôle de la corrélation du bruit qui représente le lien entre les modalités.

La première partie était consacrée à l'étude du modèle \mathcal{M}_h^2 à deux modalités. Cette étude a permis de souligner l'importance de la corrélation des bruits et l'intérêt des modalités de très faibles rapports signal à bruit. Ce qui réfute certaines idées reçues. En effet, la corrélation des bruits n'introduit pas toujours de la redondance. Ce lien favorise souvent l'interaction entre les modalités, notamment pour des valeurs négatives de corrélation. De plus, contrairement à ce que nous pouvons croire, prendre en compte une seconde modalité de faible qualité est parfois plus bénéfique qu'une seconde modalité de bonne qualité.

Dans la deuxième partie de ce chapitre, nous avons examiné l'interaction de trois modalités. Dans ce cas, les situations multimodales sont nombreuses et l'interprétation des résultats devient plus complexe. En raison du nombre de paramètres impliqués, il est difficile de tirer des conclusions générales et simples. Cependant, certains résultats contre-intuitifs

apparaissent encore, montrant l'importance du rapport signal à bruit et des corrélations des bruits.

Notre étude, dans ce chapitre, était limitée aux modalités dont les signaux d'intérêt sont fortement couplés, étant donné que le signal d'intérêt est commun à toutes les modalités. Dans le prochain chapitre, nous étendons ces résultats aux modalités avec des signaux faiblement couplés.

Modalités avec des couplages faibles

Sommaire

3.1	Introduction	53
3.2	Modèle	54
3.3	Calculs d'estimation et d'information	55
3.4	Analyse séparée ou conjointe ?	56
3.5	Interaction des modalités	57
3.5.1	Modalités de même qualité ($\alpha_2 = \alpha_1$)	58
3.5.2	Modalités de différentes qualités ($\alpha_2 < \alpha_1$)	58
3.6	Effet des corrélations et des rapports signal à bruit	61
3.6.1	Deux modalités de même qualité	62
3.6.2	Deux modalités de qualités différentes	65
3.6.3	Deux modalités de qualités très différentes	68
3.7	Conclusion	70

3.1 Introduction

Un phénomène peut se manifester de différentes façons qu'un seul appareil de mesures est incapable de discerner. Par conséquent, le recours à plusieurs modalités (systèmes de mesures) paraît judicieux afin de capter toutes ses diversités. Ce concept est connu sous le nom de multimodalité. Dans le sens où elles reflètent des caractéristiques différentes du même système physique, ces modalités sont alors qualifiées de complémentaires. Quoique complémentaires, elles sont liées car elles résultent du même système physique. La combinaison des données multimodales repose sur l'exploitation de ces liens pour permettre aux modalités de s'informer mutuellement, afin de révéler de nouvelles informations qu'elles ne peuvent pas fournir lorsqu'elles sont traitées séparément.

A titre d'exemple de modalités complémentaires nous pouvons citer le LIDAR et l'imagerie hyperspectrale, en télédétection, et l'EEG et l'IRMf, en imagerie médicale. Le LIDAR apporte une information très fiable sur l'élévation des objets alors que l'imagerie hyperspectrale différencie les matériaux. L'EEG et l'IRMf sont complémentaires en résolutions spatiale et temporelle.

On affirme généralement que la diversité et la complémentarité de ces données multimodales aboutissent à une meilleure connaissance du système. Mais qu'en est-il vraiment ? Quelles condi-

tions garantissent la combinaison des avantages de ces modalités et non de leurs inconvénients ? Pour y répondre nous devons comprendre comment ces modalités interagissent.

Nous considérons un modèle simple mais général, qui a l'intérêt de fournir des résultats analytiques faciles à interpréter et à analyser. Ce modèle consiste en deux modalités complémentaires. C'est-à-dire deux modalités qui observent deux parties différentes, mais corrélés, d'un signal d'intérêt. Ce modèle est donc plus général que celui du Chapitre 2 où les modalités partagent le même signal d'intérêt. Nous nous intéressons aux mêmes questions que précédemment à savoir le nombre optimal des modalités et l'effet des liens sur leur interaction. Ce chapitre considère l'effet du nouveau paramètre, qui est la corrélation des signaux, sur les résultats préalablement obtenus au chapitre 2.

Ce chapitre est organisé comme suit. Le modèle et les mesures d'information et d'estimation associées sont introduits dans les sections 3.2 et 3.3. L'influence des liens et rapports signal à bruit sur les performances attendues est ensuite développée dans les sections 3.4, 3.5 et 3.6. Nous concluons ce chapitre dans la section 3.7.

3.2 Modèle

Considérons le schéma suivant, qui décrit les mesures bruitées d'un vecteur $\mathbf{s} \in \mathbb{R}^2$ à l'aide de deux modalités complémentaires :

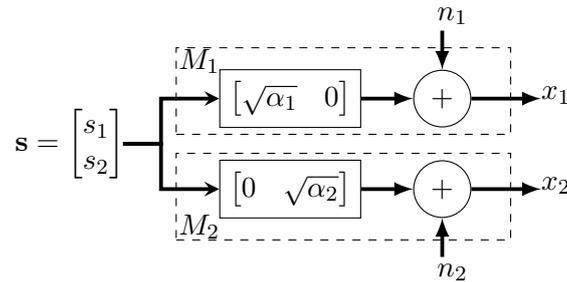


FIGURE 3.1 – Deux modalités avec des couplages faibles vues comme des canaux de communication

Le modèle mathématique associé à ce schéma est noté " \mathcal{M}_s^2 ", où l'exposant "2" et l'indice "s" dénotent le nombre de modalités et le couplage faible ("soft") entre les signaux utiles. le modèle \mathcal{M}_s^2 est décrit par les équations :

$$\mathcal{M}_s^2 : \begin{cases} x_1 = \sqrt{\alpha_1} s_1 + n_1 \\ x_2 = \sqrt{\alpha_2} s_2 + n_2 \end{cases}, \quad (3.1)$$

où s_1 , s_2 , n_1 et n_2 sont des variables gaussiennes. Sans perte de généralités nous supposons que ces variables sont de moyennes nulles et de variances unitaires, *i.e.* $\sim \mathcal{N}(0, 1)$. Nous supposons que les bruits sont indépendants de s_1 et s_2 . Les variances des signaux d'intérêt et des bruits étant unitaires, α_1 et α_2 représentent les rapports signal à bruit des modalités. Nous supposons que $\alpha_2 \leq \alpha_1$.

Le fait que les deux modalités observent le même système physique est modélisé par la corrélation entre s_1 et s_2 que nous notons $\rho_s = \frac{\mathbb{E}_p[s_1 s_2]}{\sqrt{\mathbb{E}_p[s_1^2] \mathbb{E}_p[s_2^2]}}$. Nous supposons aussi que les

bruits sont corrélés, étant donné que les systèmes d'acquisition peuvent être perturbés par des imperfections liées. Nous notons la corrélation entre les bruits $\rho_n = \frac{\mathbb{E}_p[n_1 n_2]}{\mathbb{E}_p[n_1^2] \mathbb{E}_p[n_2^2]}$. Par conséquent, les corrélations ρ_s et ρ_n constituent les liens entre les modalités x_1 et x_2 .

Liens entre les modalités :

- **Lien entre les signaux d'intérêt :** Un couplage faible considéré entre les signaux d'intérêt est représenté par le coefficient de corrélation, ρ_s , entre s_1 et s_2 .
- **Lien entre les bruits :** Le couplage que nous considérons entre les bruits est représenté par le coefficient de corrélation, ρ_n , entre n_1 et n_2 .

les modalités x_1 et x_2 étant complémentaires, il est évident que l'utilisation des deux est nécessaire pour avoir de l'information sur les deux parties de $\mathbf{s} = [s_1, s_2]$, à moins que $s_1 = \pm s_2$. C'est pourquoi, au lieu d'étudier le nombre optimal des modalités pour estimer \mathbf{s} , qui est évidemment deux, nous étudions le nombre optimal pour estimer s_i . Autrement dit, nous comparons l'analyse séparée et l'analyse conjointe. Dans le cas de l'analyse séparée, seule la modalité x_i est utilisée pour estimer s_i . Alors que, dans le cas de l'analyse conjointe les deux modalités sont utilisées simultanément pour estimer s_i en exploitant les liens ρ_s et ρ_n . Nous étudions ensuite l'effet des corrélations, ρ_s et ρ_n , et des rapports signal à bruit, α_1 et α_2 , sur le contenu informationnel des modalités et la précision d'estimation.

3.3 Calculs d'estimation et d'information

Estimation. L'estimateur optimal de s_i fonction d'une seule modalité x_i , $i \in \{1, 2\}$, est :

$$\hat{s}_{i_p}(x_i) \triangleq \int s_i p(s_i | x_i) ds_i = \frac{\sqrt{\alpha_i}}{1 + \alpha_i} x_i. \quad (3.2)$$

L'erreur quadratique moyenne relative à cet estimateur est la variance de l'erreur $s - \hat{s}_{i_p}(x_i)$:

$$\mathbf{M}_{p,p}(s_i | x_i) \triangleq \mathbb{E} [(s_i - \hat{s}_{i_p}(x_i))^2] = \frac{1}{1 + \alpha_i}. \quad (3.3)$$

| *Démonstration.* Voir section A.2 de l'annexe A. □

Maintenant si nous utilisons les deux modalités conjointement, *i.e* en exploitant les liens entre les modalités, l'estimateur optimal de s_i est :

$$\hat{s}_{i_p}(x_1, x_2) = \frac{k_{i1} x_1 + k_{i2} x_2}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n)^2}. \quad (3.4)$$

Avec :

$$k_{i1} = \alpha_j \sqrt{\alpha_i} (1 - \rho_s^2) + (\sqrt{\alpha_i} - \sqrt{\alpha_j} \rho_n \rho_s), \quad (3.5)$$

$$k_{i2} = \sqrt{\alpha_j} \rho_s - \sqrt{\alpha_i} \rho_n. \quad (3.6)$$

L'erreur quadratique relative à l'estimateur multimodal est donnée par la variance de l'erreur $s_i - \hat{s}_{i_p}(x_1, x_2)$:

$$\mathbf{M}_{p,p}(s_i | x_1, x_2) = \frac{\alpha_j (1 - \rho_s^2) + 1 - \rho_n^2}{(1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n)^2}. \quad (3.7)$$

| *Démonstration.* Voir section A.3 de l'annexe A. □

Information. L'information présente dans une seule modalité x_i sur s_i , $i \in \{1, 2\}$, est donnée par l'information mutuelle entre s_i et x_i :

$$I_{p,p}(s_i; x_i) = \frac{1}{2} \log(1 + \alpha_i). \quad (3.8)$$

| *Démonstration.* Voir section B.2 de l'annexe B. □

L'information apportée conjointement par x_1 et x_2 sur s_i est donnée par l'information mutuelle entre s_i et (x_1, x_2) :

$$I_{p,p}(s_i; (x_1, x_2)) = \frac{1}{2} \log(\mathbf{M}_{p,p}(s_i|x_1, x_2)^{-1}) = \frac{1}{2} \log\left(\frac{(1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1\alpha_2}\rho_s + \rho_n)^2}{\alpha_j(1 - \rho_s^2) + 1 - \rho_n^2}\right) \quad (3.9)$$

| *Démonstration.* Voir section B.3 de l'annexe B. □

Pour $\rho_s = 1$, le modèle \mathcal{M}_s^2 est équivalent en moyenne au modèle \mathcal{M}_h^2 étudié dans le chapitre 2.

$$\text{Si } \rho_s = 1 \text{ alors } \begin{cases} x_1 = \sqrt{\alpha_1} s_1 + n_1 \\ x_2 = \sqrt{\alpha_2} s_2 + n_2 \end{cases} \iff \begin{cases} x_1 = \sqrt{\alpha_1} s + n_1 \\ x_2 = \sqrt{\alpha_2} s + n_2 \end{cases} \quad \text{où } s \sim \mathcal{N}(0, 1). \quad (3.10)$$

L'information mutuelle et l'erreur quadratique moyenne étant liées par $I_{p,p}(s_i; \cdot) = \frac{1}{2} \log(\mathbf{M}_{p,p}(s_i|\cdot)^{-1})$, elles donnent des résultats comparables. Une information mutuelle minimale (resp. maximale) correspond à une erreur d'estimation maximale (resp. minimale). Notamment, une information mutuelle infinie est associée à une erreur d'estimation nulle. Dans ce chapitre ces deux mesures sont utilisées indifféremment.

3.4 Analyse séparée ou conjointe ?

L'analyse séparée, comme son nom l'indique traite les deux modalités séparément sans exploiter leurs liens mutuels. Par conséquent, dans le cadre du modèle \mathcal{M}_s^2 , seule la modalité x_i est utilisée pour estimer s_i , $i \in \{1, 2\}$. Tandis que l'analyse conjointe utilise les deux modalités simultanément, en tirant profit de leurs liens, pour estimer s_i . De ce fait, comparer les deux analyses est équivalent à comparer les informations mutuelles $I_{p,p}(s_i; x_i)$ et $I_{p,p}(s_i; (x_1, x_2))$. L'analyse séparée est équivalente à l'analyse conjointe si $I_{p,p}(s_i; (x_1, x_2)) = I_{p,p}(s_i; x_i), \forall i \in \{1, 2\}$. C'est-à-dire, si et seulement si $I_{p,p}(s_i; x_j|x_i) = 0$, puisque $I_{p,p}(s_i; (x_1, x_2)) = I_{p,p}(s_i; x_i) + I_{p,p}(s_i; x_j|x_i), \forall i \in \{1, 2\}$. C'est le cas si les deux modalités sont originellement non liées ou si elles sont redondantes.

Pour le modèle \mathcal{M}_s^2 , à partir des équations (3.9) et (3.8) nous trouvons que l'analyse séparée est équivalente à l'analyse conjointe, *i.e* $I_{p,p}(s_i; x_j|x_i) = 0$ pour $i \in \{1, 2\}$, si et seulement si $\sqrt{\alpha_1}\rho_s = \sqrt{\alpha_2}\rho_n$ et $\sqrt{\alpha_2}\rho_s = \sqrt{\alpha_1}\rho_n$. Pour ces valeurs de corrélations, le coefficient de x_j de l'estimateur multimodale est nul, ainsi $\hat{s}_{i_p}(x_1, x_2) = \hat{s}_{i_p}(x_i) \forall i \in \{1, 2\}$. Cette condition

correspond aux deux situations suivantes :

- $\rho_n = \rho_s = 0$:

Les deux modalités sont non liées et totalement indépendantes selon l'hypothèse de gaussianité. Par conséquent, la modalité x_j est non informative sur la partie non observée de \mathbf{s} , *i.e.* s_i . Dans ce cas la seule modalité x_i est à utiliser pour estimer s_i .

- $\alpha_1 = \alpha_2$ et $\rho_n = \rho_s$:

Pour $\alpha_1 = \alpha_2$ et $\rho_n = \rho_s$, les modalités sont redondantes l'une sachant l'autre sur la partie non observée $I_{p,p}(s_1; x_2|x_1) = I_{p,p}(s_2; x_1|x_2) = 0$.

L'analyse séparée est équivalente à l'analyse conjointe si $\rho_n = \rho_s = 0$ ou si $\alpha_1 = \alpha_2$ et $\rho_n = \rho_s$.

Dans ces situations particulières, l'analyse séparée est équivalente à l'analyse conjointe pour estimer s_1 et s_2 . Cependant, dans d'autres cas, il se peut que la seule modalité x_i suffise à estimer s_i , *i.e.* $\mathbf{M}_{p,p}(s_i|x_1, x_2) = \mathbf{M}_{p,p}(s_i|x_i)$, alors que pour estimer s_j les deux modalités soient plus précises, *i.e.* $\mathbf{M}_{p,p}(s_j|x_1, x_2) > \mathbf{M}_{p,p}(s_j|x_j)$. C'est-à-dire que x_j est redondante sur s_i par rapport à x_i , $I_{p,p}(s_i; x_j|x_i) = 0$, alors que x_i n'est pas redondante sur s_j par rapport à x_j , $I_{p,p}(s_j; x_i|x_j) > 0$. C'est le cas si $\sqrt{\alpha_j} \rho_s = \sqrt{\alpha_i} \rho_n$ et $\sqrt{\alpha_i} \rho_s \neq \sqrt{\alpha_j} \rho_n$.

La seule modalité x_i suffit à estimer s_i si et seulement si $\sqrt{\alpha_j} \rho_s = \sqrt{\alpha_i} \rho_n$.

Dans le cas particulier $|\rho_s| = 1$, les signaux utiles s_1 et s_2 sont équivalents. Selon l'hypothèse $\alpha_2 \leq \alpha_1$, si $\rho_n = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}} \rho_s$ la modalité x_2 est totalement redondante par rapport à x_1 sur l'estimation de s_1 et s_2 , $I_{p,p}(s_i; (x_1, x_2)) = I_{p,p}(s_i; x_1)$ (paragraphe 2.2.3). C'est-à-dire que la modalité x_1 suffit à estimer s_1 et s_2 . Par conséquent, le modèle \mathcal{M}_s^2 est équivalent à la seule modalité x_1 .

La seule modalité x_1 suffit à estimer s_1 et s_2 si et seulement si $|\rho_s| = 1$ et $\rho_n = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}} \rho_s$.

Hormis ces cas, l'analyse conjointe est plus informative sur s_1 et s_2 que l'analyse séparée, $I_{p,p}(s_i; (x_1, x_2)) > I_{p,p}(s_i; x_i)$.

3.5 Interaction des modalités

Dans cette section nous examinons la manière dont les modalités x_1 et x_2 interagissent. Nous étudions l'interaction des modalités en comparant les informations additionnelles $I_{p,p}(s_1; x_2|x_1)$ et $I_{p,p}(s_2; x_1|x_2)$. Cette comparaison permet de savoir sur quelle partie de \mathbf{s} l'interaction révèle

plus d'information.

Si $I_{p,p}(s_i; x_j | x_i) = 0$, alors l'interaction ne révèle aucune information additionnelle sur s_i . Ainsi la seule modalité x_i communique le maximum d'information possible sur s_i . Nous avons vu auparavant que ce cas correspond à la condition $\sqrt{\alpha_j} \rho_s = \sqrt{\alpha_i} \rho_n$ (section 3.4).

Si $I_{p,p}(s_1; x_2 | x_1) = I_{p,p}(s_2; x_1 | x_2)$, alors l'interaction des modalités permet de révéler autant d'information sur s_1 que sur s_2 . C'est-à-dire que l'interaction entre les modalités est équivalente pour l'estimation de s_1 et de s_2 . Attention, ceci ne veut pas dire que $I_{p,p}(s_1; (x_1, x_2)) = I_{p,p}(s_2; (x_1, x_2))$. En effet, les deux modalités communiquent conjointement la même quantité d'information sur s_1 et s_2 seulement si $\alpha_1 = \alpha_2$ ou si $\rho_s = \pm 1$.

Si $I_{p,p}(s_i; x_j | x_i) > I_{p,p}(s_j; x_i | x_j)$, alors l'interaction est plus bénéfique à l'estimation de s_i qu'à l'estimation de s_j .

A partir des équations (3.8) et (3.9), et selon la relation (1.25), nous obtenons que :

$$I_{p,p}(s_i; x_j | x_i) = I_{p,p}(s_i; (x_1, x_2)) - I_{p,p}(s_i; x_i) \quad (3.11)$$

$$= \frac{1}{2} \log \left(\frac{(1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n)^2}{(\alpha_j(1 - \rho_s^2) + 1 - \rho_n^2)(1 + \alpha_i)} \right). \quad (3.12)$$

3.5.1 Modalités de même qualité ($\alpha_2 = \alpha_1$)

Si $\alpha_2 = \alpha_1$, alors

$$I_{p,p}(s_1; x_2 | x_1) \Big|_{\alpha_1 = \alpha_2} = \frac{1}{2} \log \left(\frac{(1 + \alpha_1)^2 - (\alpha_1 \rho_s + \rho_n)^2}{(\alpha_1(1 - \rho_s^2) + 1 - \rho_n^2)(1 + \alpha_1)} \right), \quad (3.13)$$

$$= I_{p,p}(s_2; x_1 | x_2) \Big|_{\alpha_1 = \alpha_2}. \quad (3.14)$$

Ainsi, les modalités de même qualité s'influencent de la même façon.

Deux modalités de même qualité interagissent identiquement :

$$I_{p,p}(s_1; x_2 | x_1) = I_{p,p}(s_2; x_1 | x_2).$$

3.5.2 Modalités de différentes qualités ($\alpha_2 < \alpha_1$)

Maintenons nous considérons le cas $\alpha_2 < \alpha_1$. Nous comparons les informations $I_{p,p}(s_1; x_2 | x_1)$ et $I_{p,p}(s_2; x_1 | x_2)$ en vue de comprendre comment x_i et x_j interagissent. La figure 3.2 illustre les informations additionnelles $I_{p,p}(s_1; x_2 | x_1)$ (—) et $I_{p,p}(s_2; x_1 | x_2)$ (-.-), en fonction de ρ_n , pour différentes corrélations des signaux : $\rho_s = 0$ (figure 3.2a), $\rho_s = 0.5$ (figure 3.2b) et $\rho_s = 1$ (figure 3.2c).

- $\rho_s = 0$:

Ce cas est représenté par la figure 3.2a. Pour $\rho_s = 0$, s_1 et s_2 sont indépendants. Ainsi, les modalités x_1 et x_2 échangent uniquement de l'information sur les bruits. Selon l'hypothèse $\alpha_2 < \alpha_1$, la modalité x_2 communique une information plus précise sur le bruit que x_1 . C'est pourquoi, on a $I_{p,p}(s_1; x_2 | x_1) \geq I_{p,p}(s_2; x_1 | x_2)$ avec égalité si et seulement si $\rho_n = 0$.

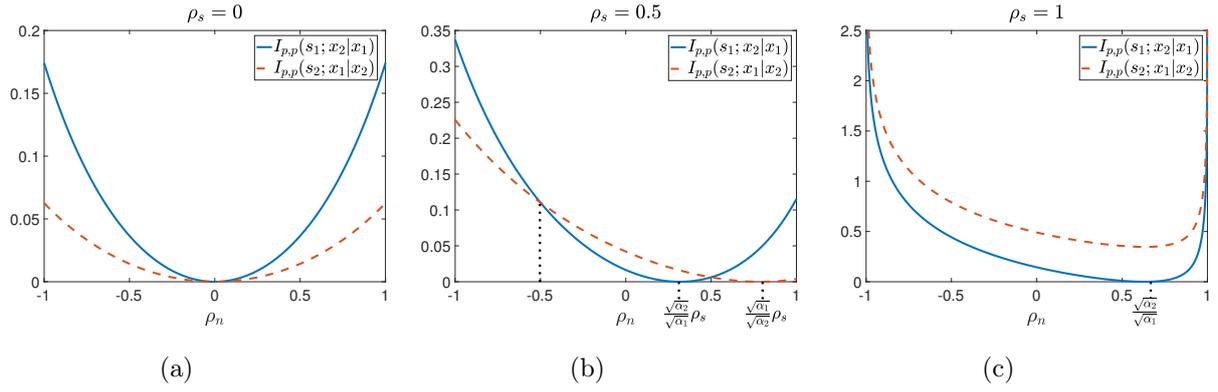


FIGURE 3.2 – Variations, en fonction de ρ_n , des informations mutuelles conjointes entre s_1 et (x_1, x_2) et entre s_2 et (x_1, x_2) . Pour différentes valeurs de ρ_s : $\rho_s = 0$ (a), $\rho_s = 0.5$ (b) et $\rho_s = 1$ (c). Avec $\alpha_1 = 5$ et $\alpha_2 = 2$. L'échelle de l'axe vertical est différente pour chaque figure en vue d'une meilleur visibilité.

Si, de plus, $\rho_n = 0$, les deux modalités sont totalement indépendantes et $I_{p,p}(s_1; x_2|x_1) = I_{p,p}(s_2; x_1|x_2) = 0$, *i.e* aucune interaction n'a lieu entre les modalités.

- $\rho_s = \frac{\sqrt{\alpha_i}}{\sqrt{\alpha_j}} \rho_n$:

L'information additionnelle que x_j apporte sur s_i est nulle, $I_{p,p}(s_i; x_j|x_i) = 0$, car x_j est redondante sur s_i par rapport à x_i . C'est-à-dire que l'information que x_j apporte sur s_i est déjà fournie par x_i . Dans ce cas, x_i suffit à estimer s_i . Cependant pour estimer s_j , les deux modalités sont plus informatives que la seule modalité x_j à moins que $\alpha_i = \alpha_j$. Si de plus $\alpha_i = \alpha_j$, l'analyse séparée est équivalente à l'analyse conjointe.

Nous constatons sur la figure 3.2a, que $I_{p,p}(s_1; x_2|x_1) = I_{p,p}(s_2; x_1|x_2)$ en $\rho_n = 0$. Ceci n'est pas dû à la redondance mais à la totale indépendance entre x_1 et x_2 .

La figure 3.2b montre que $I_{p,p}(s_1; x_2|x_1) = 0$ en $\rho_n = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}} \rho_s$ alors que $I_{p,p}(s_2; x_1|x_2) = 0$ en $\rho_n = \frac{\sqrt{\alpha_1}}{\sqrt{\alpha_2}} \rho_s$.

Nous remarquons sur la figure 3.2c que $I_{p,p}(s_1; x_2|x_1) = 0$ en $\rho_n = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$. Cependant, $I_{p,p}(s_2; x_1|x_2)$ est strictement positive, car la condition $\rho_n = \frac{\sqrt{\alpha_1}}{\sqrt{\alpha_2}}$ ne peut pas être vérifiée puisque $\frac{\sqrt{\alpha_1}}{\sqrt{\alpha_2}} > 1$ alors que la valeur absolue du coefficient de corrélation est toujours inférieure à 1.

- $|\rho_s| = |\rho_n|$:

Dans ce cas les bruits et les signaux sont similairement corrélés, ainsi les échanges d'information sur les signaux et sur les bruits sont égaux. C'est-à-dire que les deux modalités s'influencent de la même manière :

$$I_{p,p}(s_1; x_2|x_1) = \frac{1}{2} \log \left(\frac{(1 + \alpha_1)(1 + \alpha_2) - \rho_s^2 (\sqrt{\alpha_1 \alpha_2} + 1)^2}{(1 - \rho_s^2)(1 + \alpha_1)(1 + \alpha_2)} \right), \quad (3.15)$$

$$= I_{p,p}(s_2; x_1|x_2). \quad (3.16)$$

Pour le cas particulier $|\rho_s| = |\rho_n| = 1$, le modèle \mathcal{M}_s^2 est équivalent en moyenne au modèle suivant :

$$\begin{cases} x_1 = \sqrt{\alpha_1} s + n \\ x_2 = \text{sgn}(\rho_s) \sqrt{\alpha_2} s + \text{sgn}(\rho_n) n \end{cases}, \quad \text{où } s \text{ et } n \sim \mathcal{N}(0, 1). \quad (3.17)$$

Dans ce cas, $I_{p,p}(s_i; x_j | x_i) = +\infty$. C'est le seul cas où l'interaction entre les modalités permet de révéler toute l'information sur s_i . L'erreur d'estimation multimodale est nulle $\mathbf{M}_{p,p}(s_i | x_1, x_2) = 0$ et l'estimateur optimal associé est :

$$\hat{s}_{i_p}(x_1, x_2) = \frac{x_i - \text{sgn}(\rho_n) x_j}{\sqrt{\alpha_1} - \text{sgn}(\rho_s) \text{sgn}(\rho_n) \sqrt{\alpha_2}}. \quad (3.18)$$

Pour $\alpha_2 = \alpha_1$ et $\text{sgn}(\rho_s) = \text{sgn}(\rho_n)$ le dénominateur de (3.18) est nul et cet estimateur ne s'applique pas. En effet, les deux modalités sont redondantes et $\hat{s}_{i_p}(x_1, x_2) = \hat{s}_{i_p}(x_1) = \hat{s}_{i_p}(x_2)$ (section 3.4).

Sur la figure 3.2b, nous constatons que $I_{p,p}(s_1; x_2 | x_1)$ et $I_{p,p}(s_2; x_1 | x_2)$ se croisent pour $|\rho_n| = |\rho_s| = 0.5$. Similairement, sur la figure 3.2c, pour $|\rho_n| = |\rho_s| = 1$ on a $I_{p,p}(s_1; x_2 | x_1) = I_{p,p}(s_2; x_1 | x_2) = +\infty$.

- $|\rho_s| > |\rho_n|$:

Les signaux d'intérêt sont plus corrélés que les bruits, ainsi les modalités échangent plus d'information sur les signaux que sur les bruits. Puisque x_1 communique plus d'information sur le signal d'intérêt, selon l'hypothèse $\alpha_2 < \alpha_1$, alors $I_{p,p}(s_1; x_2 | x_1) < I_{p,p}(s_2; x_1 | x_2)$.

Sur la figure 3.2b, $I_{p,p}(s_1; x_2 | x_1) < I_{p,p}(s_2; x_1 | x_2)$ sur l'intervalle $[-0.5, 0.5]$ qui correspond à la condition $|\rho_s| > |\rho_n|$. Réciproquement, sur la figure 3.2c, $I_{p,p}(s_2; x_1 | x_2)$ est toujours supérieure à $I_{p,p}(s_1; x_2 | x_1)$ car $\rho_s = 1$.

- $|\rho_s| < |\rho_n|$:

Dans ce cas, les bruits sont plus corrélés que les signaux ainsi l'interaction entre les modalités concerne principalement les bruits. Par conséquent, $I_{p,p}(s_1; x_2 | x_1) > I_{p,p}(s_2; x_1 | x_2)$, car x_2 apporte plus d'information sur le bruit que x_1 .

Sur la figure 3.2b, $I_{p,p}(s_1; x_2 | x_1) > I_{p,p}(s_2; x_1 | x_2)$ pour $|\rho_s| = 0.5 < |\rho_n|$. C'est-à-dire pour des corrélations des bruits au delà de l'intervalle $[-0.5, 0.5]$.

Sur la figure 3.2c $I_{p,p}(s_2; x_1 | x_2)$ est toujours supérieure à $I_{p,p}(s_1; x_2 | x_1)$ car, puisque $|\rho_s| = 1$, les signaux sont toujours plus corrélés que les bruits, $|\rho_s| \geq |\rho_n|$.

En résumé, deux modalités, de même qualité ou dont les signaux et les bruits sont similairement corrélés, interagissent équitablement.

Autrement, si les bruits sont plus corrélés que les signaux, la modalité de faible qualité communique plus d'information lors de l'interaction. Réciproquement, si les signaux sont plus corrélés que les bruits alors la modalité de rapport signal à bruit le plus élevé communique plus d'information.

Finalement, nous pouvons conclure que la corrélation des signaux, ρ_s , a un effet "opposé" à la corrélation des bruits, ρ_n , sur l'interaction des modalités.

3.6 Effet des corrélations et des rapports signal à bruit

Maintenant, afin de déterminer les conditions pour tirer le maximum de profit des modalités, nous étudions l'effet de la corrélation des bruits, l'effet de la corrélation des signaux d'intérêt ainsi que l'impact des rapports signal à bruits sur le contenu informationnel des modalités. Cette étude s'inscrit dans la continuité de la sous-section 2.2.4 où seulement était considéré l'effet du lien entre les bruits, puisque ρ_s était égale à 1.

Nous examinons, pour une corrélation des signaux d'intérêt donnée ρ_s , l'impact de la corrélations des bruits dans trois scénarios des rapports signal à bruit : deux modalités de qualités comparables, $\alpha_2 = \alpha_1$, de qualités différentes, $\alpha_2 < \alpha_1$, et de qualités très différentes, $\alpha_2 \ll \alpha_1$.

Dans chaque scénario, afin de déterminer si la corrélation des bruits favorise ou non l'interaction des modalités, nous comparons deux modèles : un modèle avec des bruits corrélés, \mathcal{M}_s^2 , et un modèle avec des bruits non corrélés, \mathcal{M}_{so}^2 . Suivant l'hypothèse de gaussianité, des bruits non corrélés, *i.e.* $\rho_n = 0$, correspondent à des bruits indépendants.

La comparaison des modèles \mathcal{M}_{so}^2 et \mathcal{M}_s^2 ne permet pas seulement de déterminer les corrélations des bruits qui favorisent l'interaction. Elle permet aussi, si nous pouvons choisir entre différentes mesures, de préciser s'il est plus avantageux de considérer deux modalités avec des bruit corrélés, \mathcal{M}_s^2 , ou deux modalités avec des bruits indépendants, \mathcal{M}_{so}^2 .

L'estimateur optimal au sens de l'erreur quadratique moyenne et l'EQM associée, dans le cadre du modèle \mathcal{M}_{so}^2 , sont obtenus à partir de (3.4) et de (3.7), en mettant ρ_n à zéro :

$$\hat{s}_{i,p}(x_1, x_2) \Big|_{\rho_n=0} = \frac{(\alpha_j \sqrt{\alpha_i} (1 - \rho_s^2) + \sqrt{\alpha_i}) x_i + \sqrt{\alpha_j} \rho_s x_j}{(\alpha_1 + 1)(\alpha_2 + 1) - \alpha_1 \alpha_2 \rho_s^2}, \quad (3.19)$$

$$\mathbf{M}_{p,p}(s_i | x_1, x_2) \Big|_{\rho_n=0} = \frac{\alpha_j (1 - \rho_s^2) + 1}{(1 + \alpha_1)(1 + \alpha_2) - \alpha_1 \alpha_2 \rho_s^2}. \quad (3.20)$$

Similairement, l'information conjointe qu'apportent deux modalités avec des bruits indépendants sur s est obtenue à partir de (3.9), en mettant ρ_n à zéro :

$$I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_n=0} = \frac{1}{2} \log \left(\frac{(1 + \alpha_1)(1 + \alpha_2) - \alpha_1 \alpha_2 \rho_s^2}{\alpha_j (1 - \rho_s^2) + 1} \right). \quad (3.21)$$

Nous remarquons que l'erreur quadratique (3.20) et l'information conjointe (3.21) dépendent du carré de ρ_s . Par conséquent, pour des bruits indépendants, les corrélations des signaux de

même valeur absolue, $\pm\rho_s$, ont la même influence sur le contenu informationnel et la précision d'estimation des modalités :

$$I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_n=0, \rho_s} = I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_n=0, -\rho_s}, \quad (3.22)$$

$$\mathbf{M}_{p,p}(s_i|x_1, x_2) \Big|_{\rho_n=0, \rho_s} = \mathbf{M}_{p,p}(s_i|x_1, x_2) \Big|_{\rho_n=0, -\rho_s}. \quad (3.23)$$

Lorsque les modalités de \mathcal{M}_{so}^2 communiquent plus d'information que les modalités du modèle \mathcal{M}_s^2 , nous concluons que la corrélation des bruits entraîne de la redondance en information sur s_i . Réciproquement, quand les modalités \mathcal{M}_{so}^2 sont moins informatives que les modalités de \mathcal{M}_s^2 , alors la corrélation des bruits favorise l'interaction des modalités.

En résolvant l'inégalité $I_{p,p}(s_i; (x_1, x_2)) \leq I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_n=0}$ nous concluons que la corrélation des bruits est néfaste à la multimodalité si et seulement si $\rho_n \in I_{red}$, où I_{red} est l'intervalle de redondance partielle entre les modalités x_1 et x_2 sur s_i :

$$I_{red}(s_i|x_1, x_2) = [0, \rho_n^*] \quad \text{où} \quad \rho_n^* = \rho_s \frac{2\sqrt{\alpha_1\alpha_2}(1 + \alpha_j(1 - \rho_s^2))}{\alpha_1\alpha_2(1 - \rho_s^2) + \alpha_i + \alpha_j\rho_s^2}. \quad (3.24)$$

Pour $\rho_n = \rho_n^*$ les modalités de \mathcal{M}_{so}^2 communiquent autant d'information que les modalités de \mathcal{M}_s^2 car $I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_n=\rho_n^*} = I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_n=0}$. Cependant, les deux modèles dans ce cas n'apportent pas la même information sur s_i , car $\hat{s}_{i_p}(x_1, x_2) \Big|_{\rho_n=\rho_n^*} \neq \hat{s}_{i_p}(x_1, x_2) \Big|_{\rho_n=0}$.

D'après (3.24), le signe de ρ_n^* est égale au signe de ρ_s . Par conséquent, la redondance peut apparaître entre les modalités si la corrélation des signaux a le même signe que la corrélation des bruits. Ce résultat généralise le résultat du chapitre 2. Dans le cadre du modèle \mathcal{M}_h^2 du chapitre 2, ce sont les corrélations positives des bruits qui induisent de la redondance puisque $\rho_s = 1$.

La redondance a lieu pour des corrélations entre les signaux d'intérêt, ρ_s , et entre les bruits, ρ_n , de même signe.

3.6.1 Deux modalités de même qualité

Pour $\alpha_2 = \alpha_1$, les deux modalités ont la même précision, *i.e* $\mathbf{M}_{p,p}(s_1|x_1) = \mathbf{M}_{p,p}(s_2|x_2)$, et communiquent la même quantité d'information sur la partie observée de \mathbf{s} , *i.e* $I_{p,p}(s_1; x_1) = I_{p,p}(s_2; x_2)$. De plus conjointement, les deux modalités communiquent la même quantité d'information sur s_i , *i.e* $I_{p,p}(s_1; (x_1, x_2)) = I_{p,p}(s_2; (x_1, x_2))$.

Intervalle de redondance. Les figures 3.3a et 3.3b représentent la variation de l'intervalle de redondance introduite par ρ_n sur s_1 et sur s_2 , respectivement, $I_{red}(s_i|x_1, x_2)$ (3.24) :

$$I_{red}(s_i|x_1, x_2) \Big|_{\alpha_2=\alpha_1} = \left[0, \rho_n^* = \frac{2\rho_s(1 + \alpha_1(1 - \rho_s^2))}{\alpha_1(1 - \rho_s^2) + 1 + \rho_s^2} \right] \quad (3.25)$$

Les segments représentent la longueur de l'intervalle de redondance pour ρ_s donné. Les intervalles de redondance pour s_1 et s_2 sont égaux, *i.e* $I_{red}(s_1|x_1, x_2) = I_{red}(s_2|x_1, x_2)$ car $\alpha_2 = \alpha_1$.

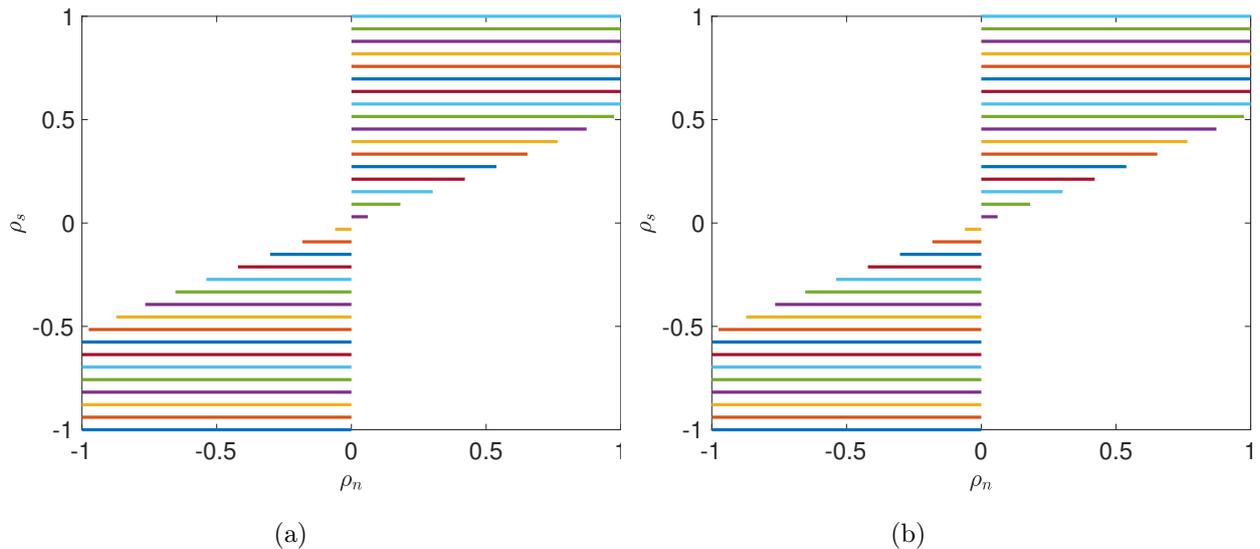


FIGURE 3.3 – Variation de l'intervalle de redondance introduite par la corrélation des bruits entre les modalités sur s_1 , $I_{red}(s_1|x_1, x_2)$ (a) et sur s_2 , $I_{red}(s_2|x_1, x_2)$ (b), où $\alpha_1 = \alpha_2 = 5$. Chaque segment présente, pour ρ_s donné, les valeurs de ρ_n qui introduisent de la redondance entre les modalités pour l'estimation de s_1 (a) et pour l'estimation de s_2 (b). Les figures (a) et (b) sont identiques car $I_{red}(s_1|x_1, x_2) = I_{red}(s_2|x_1, x_2)$, puisque $\alpha_2 = \alpha_1$.

Nous remarquons que pour des signaux d'intérêt faiblement corrélés, $\rho_s \rightarrow 0$, la redondance a lieu pour des bruits faiblement corrélés de même signe, $\rho_n^* \rightarrow 0$. L'intervalle de redondance s'élargit quand la corrélation des signaux augmente et ρ_n^* tend vers ± 1 pour de fortes corrélations de signaux. Pour interpréter ces résultats nous considérons trois cas de corrélations de signaux : $\rho_s = -1$, $\rho_s = 0$ et $\rho_s = 1$.

Information mutuelle. Les figures 3.4a et 3.4b illustrent les informations multimodales $I_{p,p}(s_1; (x_1, x_2))$ et $I_{p,p}(s_2; (x_1, x_2))$, respectivement, en fonction de ρ_n pour différentes valeurs de ρ_s . Les lignes horizontales correspondent aux informations multimodales du modèle $\mathcal{M}_{s_0}^2$, $\rho_n = 0$. Les autres courbes correspondent aux informations multimodales du modèle \mathcal{M}_s^2 . Les informations multimodales de $\mathcal{M}_{s_0}^2$ et \mathcal{M}_s^2 associées à une valeur identique de ρ_s sont représentées par des lignes de même style.

Nous constatons que pour ρ_s et ρ_n fixées, $I_{p,p}(s_1; (x_1, x_2)) = I_{p,p}(s_2; (x_1, x_2))$ car $\alpha_2 = \alpha_1$. Aussi, $I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=1, \rho_n=0} = I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=-1, \rho_n=0}$, car d'après (3.21) pour des bruits indépendants le signe de ρ_s n'a pas d'impact seule sa valeur absolue compte.

- $\rho_s = 0$:

La corrélation des signaux ρ_s traduit le taux de "similarité" des parties s_1 et s_2 . Pour $\rho_s = 0$, s_1 et s_2 sont indépendantes et donc totalement différentes.

Dans le cadre du modèle modèle \mathcal{M}_s^2 l'information qu'apportent conjointement x_1 et x_2 sur

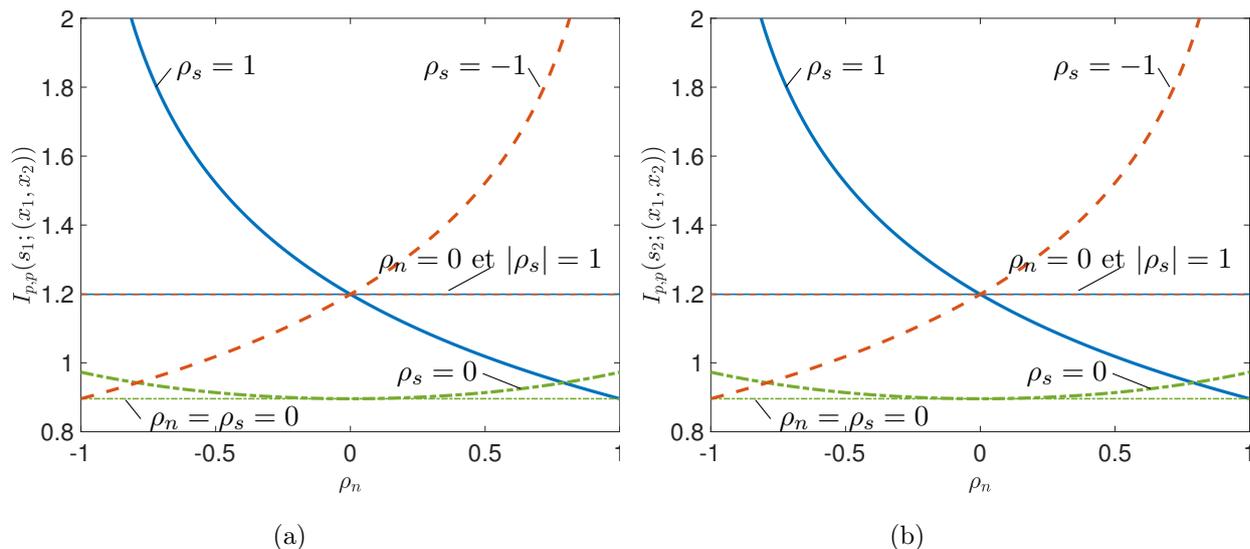


FIGURE 3.4 – Variations, en fonction de ρ_s , des informations mutuelles conjointes entre s_1 et (x_1, x_2) (a) et entre s_2 et (x_1, x_2) (b) où $\alpha_2 = \alpha_1 = 5$. Les figures (a) et (b) sont identiques car $I_{p,p}(s_1; (x_1, x_2)) = I_{p,p}(s_2; (x_1, x_2))$, puisque $\alpha_2 = \alpha_1$.

s_i est obtenue à partir de (3.9), en mettant ρ_s à zéro :

$$I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=0} = \frac{1}{2} \log \left(\frac{(1 + \alpha_1)^2 - \rho_n^2}{1 + \alpha_1 - \rho_n^2} \right) \geq I_{p,p}(s_i; x_i). \quad (3.26)$$

Tant que $\rho_n \neq 0$, x_1 et x_2 apportent conjointement strictement plus d'information sur s_i qu'une seule modalité. C'est-à-dire que même si les signaux d'intérêt sont indépendants les deux modalités interagissent tant que les bruits sont liés. La deuxième modalité x_j n'apporte pas d'information sur le signal s_i , mais elle en apporte sur le bruit.

Dans ce cas, la corrélation des bruits n'introduit jamais de redondance, $I_{red}(s_i|x_1, x_2) = \{0\}$. Par conséquent, les modalités de \mathcal{M}_s^2 sont plus informatives que les modalités de $\mathcal{M}_{s_0}^2$, $I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=0} \geq I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=0, \rho_n=0}$ avec égalité si et seulement si $\rho_n = 0$. Au contraire si les bruits sont indépendants, $\rho_n = 0$, aucune interaction ne peut avoir lieu en raison de la totale indépendance entre les modalités, et on a :

$$I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=0, \rho_n=0} = I_{p,p}(s_i; x_i). \quad (3.27)$$

Même si les signaux d'intérêts sont indépendants, les modalités interagissent toujours si la corrélation entre les bruits est non nulle, $\rho_n \neq 0$.

- $|\rho_s| = 1$:

L'information mutuelle entre s_i et les deux modalités pour $|\rho_s| = 1$ est, d'après (3.9) :

$$I_{p,p}(s_i; (x_1, x_2)) \Big|_{|\rho_s|=1, \alpha_2=\alpha_1} = \begin{cases} \frac{1}{2} \log \left(1 + \frac{2\alpha_i}{1-|\rho_n|} \right), & \text{si } \text{sgn}(\rho_s) \neq \text{sgn}(\rho_n) \\ \frac{1}{2} \log (1 + 2\alpha_i), & \text{si } \rho_n = 0 \\ \frac{1}{2} \log \left(1 + \frac{2\alpha_i}{1+|\rho_n|} \right), & \text{si } \text{sgn}(\rho_s) = \text{sgn}(\rho_n) \end{cases} \quad (3.28)$$

Lorsque $\text{sgn}(\rho_n) \neq \text{sgn}(\rho_s)$, les modalités de \mathcal{M}_{so}^2 sont plus informatives que les modalités de \mathcal{M}_{so} . C'est-à-dire que bien que s_1 et s_2 soient redondants ($|\rho_s| = 1$), nous arrivons à compenser le bruit sans affecter le signal d'intérêt ou abaisser sa qualité. Notamment, pour $\rho_n = \pm 1$, le bruit est complètement supprimé et l'erreur d'estimation multimodale $\mathbf{M}_{p,p}(s_i|x_1, x_2)$ est nulle.

Cependant, pour $\text{sgn}(\rho_n) = \text{sgn}(\rho_s)$, les modalités de \mathcal{M}_{so}^2 sont plus informatives que les modalités de \mathcal{M}_s^2 . C'est-à-dire que le bruit est mal compensé et les modalités deviennent partiellement redondantes. Notamment, pour $\rho_n = \pm 1$, les deux modalités sont totalement redondantes et ne sont conjointement pas plus précises qu'une seule modalité, $\hat{s}_{i_p}(x_1, x_2) = \hat{s}_{i_p}(x_1) = \hat{s}_{i_p}(x_2)$.

Nous constatons sur la figure 3.4 que :

- $I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=1} > I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=1, \rho_n=0}$ pour $\rho_s = 1$ et $\rho_n < 0$ (figure 3.4a —) et pour $\rho_s = -1$ et $\rho_n > 0$ (figure 3.4b ---). En particulier, pour $\rho_n = \pm 1$, $I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=1, \rho_n=-1} = +\infty$.
- $I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=1} < I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=1, \rho_n=0}$ pour $\rho_s = 1$ et $\rho_n > 0$ (figure 3.4a ---) et pour $\rho_s = -1$ et $\rho_n < 0$ (figure 3.4b —). Notamment, pour $\rho_n = \pm 1$, $I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=1} = I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=0, \rho_n=0}$, car les modalités sont totalement redondantes.

Si nous comparons les informations multimodales du modèle \mathcal{M}_{so}^2 (droites horizontales), nous constatons que $I_{p,p}(s_i; (x_1, x_2)) \Big|_{|\rho_s|=1, \rho_n=0} > I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=0, \rho_n=0}$. C'est-à-dire qu'en cas de bruits indépendants, la corrélation entre s_1 et s_2 révèle toujours plus d'information.

Maintenant, si nous comparons les informations multimodales du modèle \mathcal{M}_s^2 , nous constatons que pour de fortes corrélations des bruits, $I_{p,p}(s_i; (x_1, x_2)) \Big|_{|\rho_s|=1} > I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=0}$. C'est-à-dire qu'un modèle avec des signaux indépendants est plus intéressant.

3.6.2 Deux modalités de qualités différentes

Nous considérons maintenant le cas de deux modalités de différentes qualités, $\alpha_2 \leq \alpha_1$. Dans ce cas, la modalité x_1 est plus précise que x_2 , *i.e* $\mathbf{M}_{p,p}(s_1|x_1) \leq \mathbf{M}_{p,p}(s_2|x_2)$, ainsi l'information qu'elle communique sur s_1 est supérieure à l'information que x_2 communique sur s_2 , *i.e* $I_{p,p}(s_1; x_1) \geq I_{p,p}(s_2; x_2)$. De plus, conjointement, les deux modalités apportent plus d'information sur s_1 que sur s_2 , $I_{p,p}(s_1; (x_1, x_2)) \geq I_{p,p}(s_2; (x_1, x_2))$, avec égalité si et seulement si $|\rho_s| = 1$. Pour $|\rho_s| = 1$, les signaux d'intérêt s_1 et s_2 sont équivalents en moyenne.

Intervalle de redondance. Les figures 3.5a et 3.5b représentent la variation de l'intervalle de redondance, $I_{red}(s_i|x_1, x_2)$ (3.24) :

$$I_{red}(s_i|x_1, x_2) \Big|_{\alpha_2 < \alpha_1} = \left[0, \rho_n^* = \frac{2\sqrt{\alpha_1\alpha_2}\rho_s(1 + \alpha_j(1 - \rho_s^2))}{\alpha_1\alpha_2(1 - \rho_s^2) + \alpha_i + \alpha_j\rho_s^2} \right] \quad (3.29)$$

Nous remarquons que pour ρ_s donné, l'intervalle de redondance correspondant de la figure 3.5a est plus petit que l'intervalle de redondance correspondant de la figure 3.5b. Ainsi, la corrélation des bruits introduit plus de redondance sur s_2 que sur s_1 . C'est-à-dire que ρ_n est souvent plus bénéfique à l'estimation de s_1 qu'à l'estimation de s_2 .

Maintenant, nous comparons ces intervalles de redondances, $I_{red}(s_i|x_1, x_2) \Big|_{\alpha_2 < \alpha_1}$, avec ceux obtenus pour $\alpha_2 = \alpha_1$, $I_{red}(s_i|x_1, x_2) \Big|_{\alpha_2 = \alpha_1}$, figure 3.3. Nous constatons que :

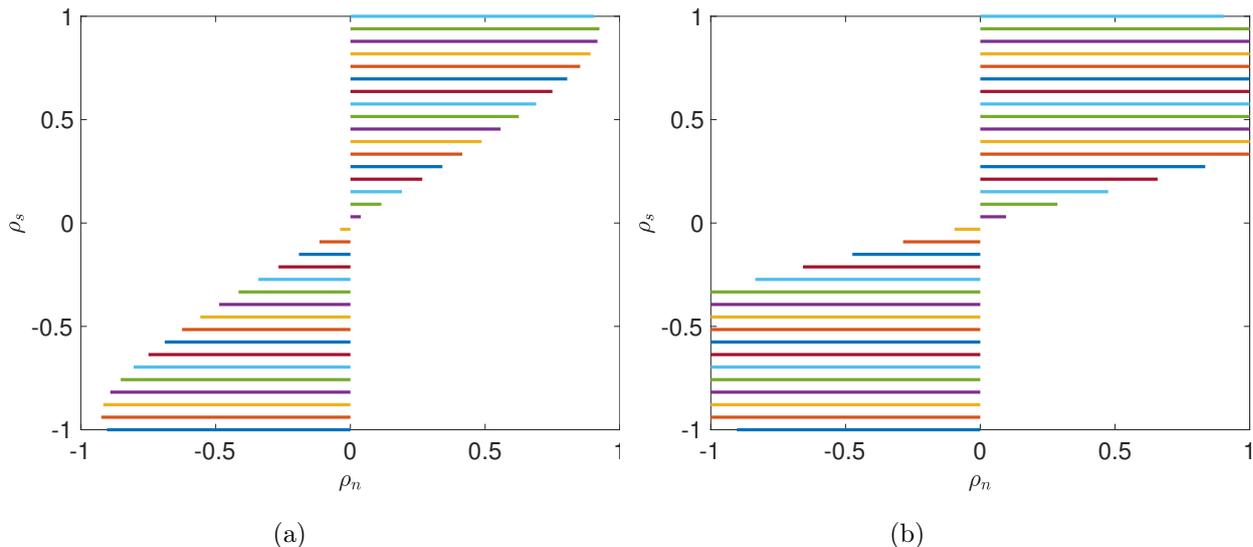


FIGURE 3.5 – Variation de l'intervalle de redondance introduite par la corrélation des bruits entre les modalités sur s_1 (a) et sur s_2 (b), dans le cas où $\alpha_1 = 5$ et $\alpha_2 = 2$. Chaque segment présente, pour ρ_s donné, les valeurs de ρ_n qui introduisent de la redondance entre les modalités pour l'estimation de s_1 (a) et pour l'estimation de s_2 (b).

- Pour l'estimation de s_1 , en comparant la figure 3.5a avec la figure 3.3 pour ρ_s donné, l'intervalle de redondance est plus petit dans le cas de modalités de différentes qualités : $I_{red}(s_i|x_1, x_2)]_{\alpha_2 < \alpha_1} \subseteq I_{red}(s_i|x_1, x_2)]_{\alpha_2 = \alpha_1}$. Parce que, puisque $\alpha_2 < \alpha_1$, la modalité x_2 apporte de l'information sur le bruit qui permet de réduire le bruit dans la modalité x_1 .
- Pour l'estimation de s_2 , en comparant la figure 3.5b avec la figure 3.3 pour ρ_s donné, l'intervalle de redondance est plus large dans le cas de modalités de différentes qualités : $I_{red}(s_i|x_1, x_2)]_{\alpha_2 < \alpha_1} \supseteq I_{red}(s_i|x_1, x_2)]_{\alpha_2 = \alpha_1}$, notamment pour de faibles valeurs de ρ_s . Parce que x_1 apporte principalement de l'information sur le signal d'intérêt qui est exploitée si s_1 et s_2 sont fortement corrélés. En particulier pour $\rho_s = 1$, l'intervalle de redondance est plus petit par rapport au cas $\alpha_2 = \alpha_1$.

Information mutuelle. Les figures 3.6a et 3.6b illustrent les informations multimodales $I_{p,p}(s_1; (x_1, x_2))$ et $I_{p,p}(s_2; (x_1, x_2))$, respectivement, en fonction de ρ_n pour différentes valeurs de ρ_s . Les lignes horizontales correspondent aux informations multimodales du modèle $\mathcal{M}_{s_0}^2$, $\rho_n = 0$. Les autres courbes correspondent aux informations multimodales du modèle \mathcal{M}_s^2 . Les informations multimodales de $\mathcal{M}_{s_0}^2$ et \mathcal{M}_s^2 associées à un même ρ_s donné sont représentées par des lignes de même style.

Nous constatons que $I_{p,p}(s_i; (x_1, x_2))\big|_{\rho_s=1, \rho_n=0} = I_{p,p}(s_i; (x_1, x_2))\big|_{\rho_s=-1, \rho_n=0}$, car pour des bruits indépendants le signe de ρ_s n'a pas d'impact (3.21).

- $\rho_s = 0$:

Nous constatons que pour $\rho_s = 0$, $I_{p,p}(s_i; (x_1, x_2))\big|_{\rho_s=0} \geq I_{p,p}(s_i; (x_1, x_2))\big|_{\rho_s=0, \rho_n=0}$ avec égalité si et seulement si $\rho_n = 0$, car pour des signaux indépendants la corrélation des bruits

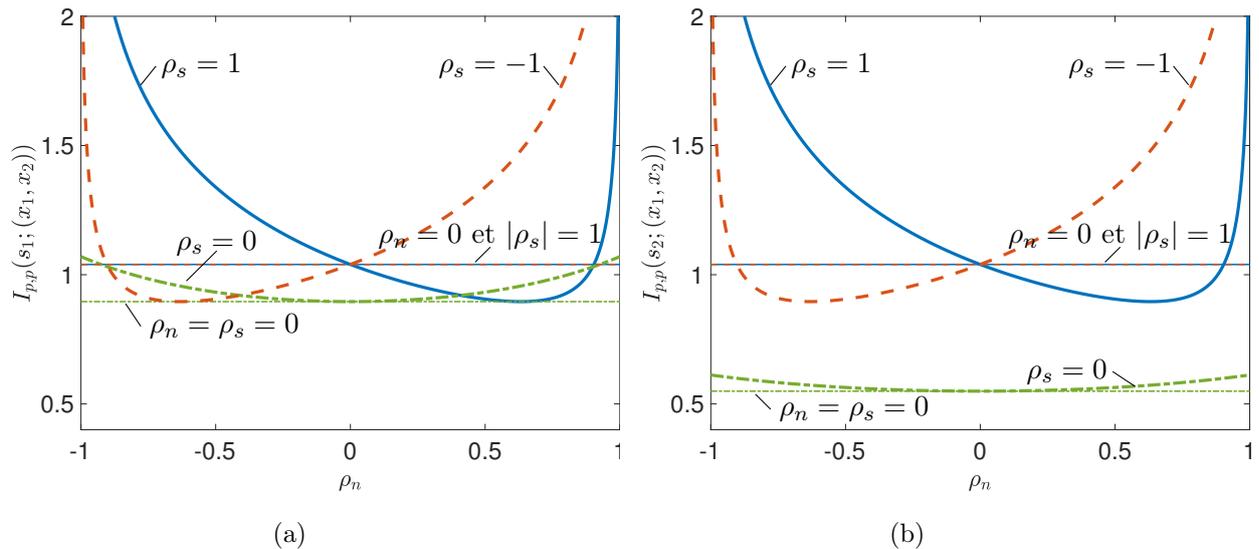


FIGURE 3.6 – Variations, en fonction de ρ_s , des informations mutuelles conjointes entre s_1 et (x_1, x_2) (a) et entre s_2 et (x_1, x_2) (b) dans le cas où $\alpha_1 = 5$ et $\alpha_2 = 2$.

toujours favorise l'interaction.

En particulier, pour l'estimation de s_2 (figure 3.6b), nous remarquons que conjointement x_1 et x_2 communiquent légèrement plus d'informations sur s_2 lorsque $\rho_n \neq 0$. C'est-à-dire que, même si x_1 assure principalement de l'information sur le signal utile ($\alpha_1 > \alpha_2$), qui ne peut pas être exploitée puisque $\rho_s = 0$, x_1 apporte aussi de l'information sur le bruit. Cependant, l'information que x_1 apporte sur le bruit est moins fiable que celle qu'apporte x_2 .

- $|\rho_s| = 1$:

Nous remarquons sur la figure 3.6, que conjointement les deux modalités apportent la même quantité d'information sur s_1 et s_2 , car pour $|\rho_s| = 1$, les signaux d'intérêt sont équivalents en moyenne.

Nous remarquons que $I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=1} \leq I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=1, \rho_n=0}$ sur l'intervalle $I_{red}(s_i|x_1, x_2) = \left[0, \pm \frac{2\sqrt{\alpha_1\alpha_2}}{\alpha_1+\alpha_2}\right]$. Par conséquent, l'intervalle de redondance est plus petit comparé au cas $\alpha_2 = \alpha_1$ où $I_{red}(s_i|x_1, x_2) = [0, \pm 1]$ (figure 3.4), car l'information sur le bruit que x_2 communique permet de compenser le bruit. Notamment pour $\rho_n = \pm 1$, le bruit est complètement compensé et $I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=1} = +\infty$.

La condition pour pleinement bénéficier de la multimodalité est que les corrélations des signaux et des bruits soient de différents signes.

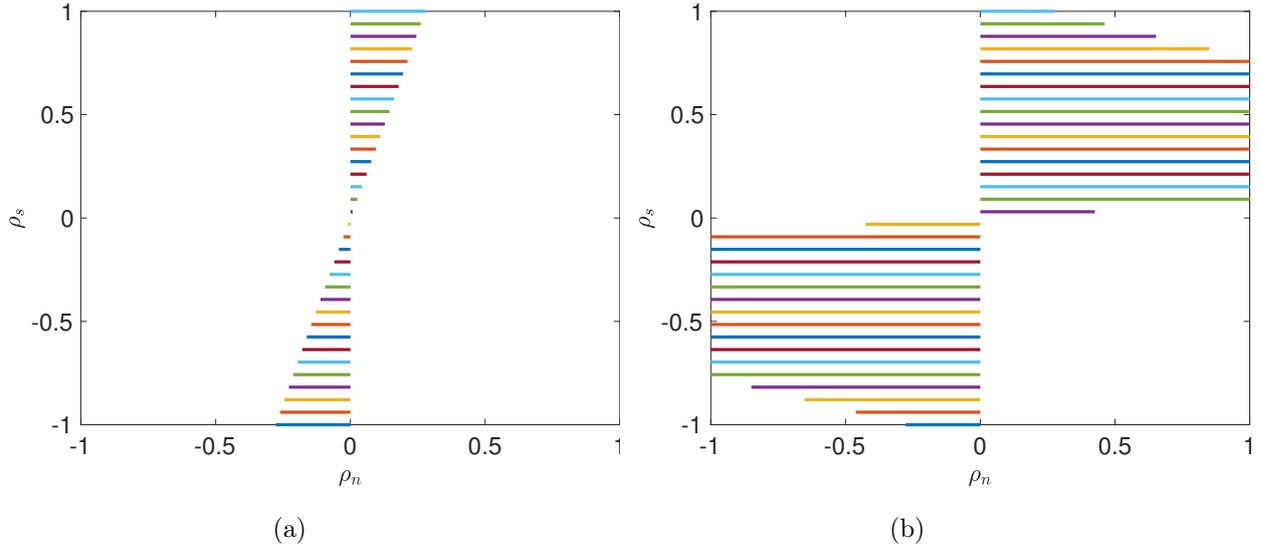


FIGURE 3.7 – Variation de l'intervalle de redondance introduite par la corrélation des bruits entre les modalités sur s_1 (a) et sur s_2 (b), pour $\alpha_1 = 5$ et $\alpha_2 = 0.1$. Chaque segment présente, pour ρ_s donné, les valeurs de ρ_n qui introduisent de la redondance entre les modalités pour l'estimation de s_1 (a) et pour l'estimation de s_2 (b).

3.6.3 Deux modalités de qualités très différentes

Nous terminons cette étude, en considérant un cas particulier où $\alpha_2 \ll \alpha_1$. Si de plus $\alpha_2 \rightarrow 0$ alors la modalité x_2 tend vers une mesure du bruit seul n_2 :

$$\begin{cases} x_1 = \sqrt{\alpha_1} s_1 + n_1 \\ x_2 = \sqrt{\alpha_2} s_2 + n_2 \end{cases} \xrightarrow{\alpha_2 \rightarrow 0} n_2 \quad (3.30)$$

Dans ce cas, x_2 apporte une information négligeable sur s_2 , ainsi s_2 est mal estimé. Cependant, si les bruits sont corrélés l'information que x_2 apporte sur le bruit permet de compenser le bruit en x_1 .

Intervalle de redondance. Les figures 3.7a et 3.7b représentent la variation de l'intervalle de redondance, $I_{red}(s_i|x_1, x_2)$ (3.24) :

$$I_{red}(s_i|x_1, x_2) = \left[0, \rho_n^* = \rho_s \frac{2\sqrt{\alpha_1\alpha_2}(1 + \alpha_j(1 - \rho_s^2))}{\alpha_1\alpha_2(1 - \rho_s^2) + \alpha_i + \alpha_j\rho_s^2} \right] \quad (3.31)$$

Nous remarquons que pour ρ_s donné, l'intervalle de redondance correspondant de la figure 3.7a est très petit par rapport à l'intervalle de redondance correspondant de la figure 3.7b. Ainsi, la corrélation des bruits introduit beaucoup plus de redondance sur s_2 , que sur s_1 . De plus, l'intervalle de redondance sur s_1 s'élargit quand ρ_s augmente alors que l'intervalle de redondance sur s_2 se réduit. C'est-à-dire que les fortes corrélations des signaux introduisent plus de redondance sur s_1 par rapport aux faibles corrélations, contrairement à s_2 .

Si nous comparons ces intervalles avec les intervalles du cas $\alpha_2 < \alpha_1$, figure 3.5, nous constatons que la corrélation des bruits introduit moins de redondance sur s_1 dans le cas $\alpha_2 \ll \alpha_1$.

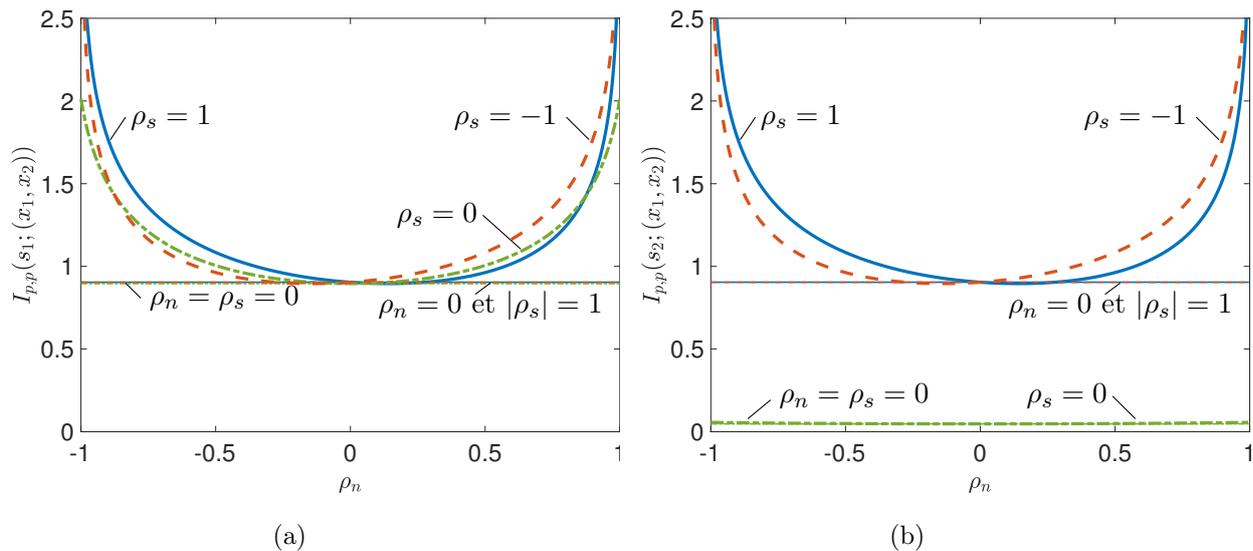


FIGURE 3.8 – Variations, en fonction de ρ_s , des informations mutuelles conjointes entre s_1 et (x_1, x_2) (a) et entre s_2 et (x_1, x_2) (b) pour $\alpha_1 = 5$ et $\alpha_2 = 0.1$.

Réciproquement, elle introduit plus de redondance sur s_2 pour des corrélations faibles des signaux. Cependant, pour de fortes corrélations, l'intervalle de redondance est réduit car l'information que x_1 communique sur le signal est exploitée.

Information mutuelle. Les figures 3.8a et 3.8b illustrent les informations multimodales $I_{p,p}(s_1; (x_1, x_2))$ et $I_{p,p}(s_2; (x_1, x_2))$, respectivement, en fonction de ρ_n pour différentes valeurs de ρ_s . Les lignes horizontales correspondent aux informations multimodales du modèle $\mathcal{M}_{s_0}^2$, $\rho_n = 0$. Les autres courbes correspondent aux informations multimodales du modèle \mathcal{M}_s^2 . Les informations multimodales de $\mathcal{M}_{s_0}^2$ et \mathcal{M}_s^2 associées à un ρ_s donné sont représentées par des lignes de même style.

$I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=1, \rho_n=0}$ et $I_{p,p}(s_i; (x_1, x_2)) \Big|_{\rho_s=-1, \rho_n=0}$ coïncident toujours puisque dans le cas du modèle $\mathcal{M}_{s_0}^2$, le signe de ρ_s n'a pas d'impact (3.21).

- $\rho_s = 0$:

Nous constatons que pour $\rho_s = 0$, $I_{p,p}(s_1; (x_1, x_2)) \Big|_{\rho_s=0} \geq I_{p,p}(s_1; (x_1, x_2)) \Big|_{\rho_s=0, \rho_n=0}$ alors que $I_{p,p}(s_2; (x_1, x_2)) \Big|_{\rho_s=0} \approx I_{p,p}(s_2; (x_1, x_2)) \Big|_{\rho_s=0, \rho_n=0}$, car les signaux d'intérêt sont indépendants et les modalités échangent principalement de l'information sur les bruits. Puisque $\alpha_2 \ll \alpha_1$ alors x_2 communique plus d'information sur le bruit que x_1 . Ainsi, les modalités sont conjointement plus informatives sur s_1 que sur s_2 .

- $|\rho_s| = 1$:

Pour $|\rho_s| = 1$, s_1 et s_2 sont équivalents en moyenne. Ainsi leurs intervalles de redondance sont égaux. Nous remarquons que dans ce cas, les modalités de \mathcal{M}_s^2 sont souvent plus informatives que les modalités du modèle $\mathcal{M}_{s_0}^2$. En effet, pour toutes les valeurs de ρ_s , les modalités de \mathcal{M}_s^2

communiquent souvent plus d'information sur s_1 que les modalités de $\mathcal{M}_{s_0}^2$. Ce qui est pas toujours vrai pour s_2 .

Sur la figure 3.8a, les informations multimodales du modèle $\mathcal{M}_{s_0}^2$ coïncident car l'information que x_2 apporte sur le bruit n'est pas exploitée puisque les bruits sont indépendants, $I_{p,p}(s_1; (x_1, x_2)) \Big|_{\rho_n=0} \approx I_{p,p}(s_1; x_1)$.

Réciproquement, sur la figure 3.8b, $I_{p,p}(s_2; (x_1, x_2)) \Big|_{|\rho_s|=1, \rho_n=0} \geq I_{p,p}(s_2; (x_1, x_2)) \Big|_{\rho_s=\rho_n=0}$, car les signaux sont corrélés ce qui permet de révéler plus d'information sur s_2 en exploitant l'information fournie par x_1 .

- **Synthèse :**

Si nous comparons les trois courbes de la la figure 3.8a correspondantes au modèle \mathcal{M}_s^2 , nous constatons que les informations multimodales sont comparables, sauf si $|\rho_n| \rightarrow 1$ car uniquement pour $|\rho_s| = 1$ les deux modalités communiquent une information infinie sur s_1 .

Cependant, pour s_2 , la courbe correspondante à $\rho_s = 0$ et strictement inférieure aux courbes relatives à $|\rho_s| = 1$. En effet, les deux modalités révèlent plus d'information sur s_2 tant que les signaux sont fortement corrélés, car l'information que x_1 apporte sur s_2 est plus importante pour de fortes corrélations des signaux.

Pour $\alpha_2 \ll \alpha_1$, si les bruits sont corrélés l'information que x_2 apporte sur le bruit assure une meilleure estimation de s_1 mais s_2 est mal estimé. Ce cas est donc intéressant si nous souhaitons avoir de l'information sur la partie s_1 seule.

3.7 Conclusion

Dans ce chapitre, nous avons considéré un modèle avec deux modalités complémentaires. Nous avons supposé qu'il existe des liens faibles entre les signaux d'intérêt et entre les bruits des modalités. Nous avons examiné l'impact de ces deux liens sur l'interaction multimodale.

Cette étude a permis de généraliser les résultats du modèle étudié au chapitre 2. La condition pour tirer le maximum de profit de la multimodalité est que la corrélation des bruits soit de signe opposé à la corrélation des signaux d'intérêt. Ce résultat est cohérent avec celui du chapitre 2 où le maximum d'information est atteint pour des corrélations négatives car $\rho_s = 1$, puisque $s_1 = s_2$.

Nous avons trouvé dans le cas de deux modalités fortement couplées (chapitre 2) qu'une modalité de très faible qualité est parfois plus bénéfique qu'une modalité de bonne qualité. Cependant, dans le cas de deux modalités complémentaires pour avoir de l'information sur les différentes parties observées il faut que les deux modalités soient de bonne qualité. Autrement, la partie observée par la modalité ayant le rapport signal à bruit petit est mal estimée.

Le résultat majeur de ce chapitre est que, même si les signaux d'intérêt ne sont pas corrélés, la corrélation entre les bruits permet aux modalités de s'informer mutuellement ce qui améliore la précision d'estimation.

Dans cette première partie du manuscrit, nous avons pu souligner l'importance des corrélations des bruits et des modalités de faibles qualités. Par contre, tous ces résultats sont valides dans le cadre d'un modèle adéquat. Un modèle adéquat ("*matched model*" en anglais), postule que les densités de probabilité supposées correspondent aux densités de probabilité réelles. Cette hypothèse, largement utilisée en pratique, est peu réaliste en raison de la complexité des systèmes physiques étudiés et des imperfections des mesures. Notamment, en multimodalité, la modélisation des liens entre les modalités est complexe et donc sujette à des erreurs. La partie suivante de ce travail considère l'interaction des modalités dans le cadre d'un modèle inadéquat ("*Mismatched or misspecified model*" en anglais).

Partie II

Modèle inadéquat

Introduction de la partie II

L'adéquation du modèle supposé au modèle des données est une hypothèse fortement répandue. Cependant, en raison de plusieurs facteurs, tels que l'imperfection des mesures, cette hypothèse est loin d'être vraie. Nous consacrons cette partie à l'étude des modèles inadéquats, c'est-à-dire les modèles qui sont supposés exacts alors qu'ils ne correspondent pas aux modèles réels.

Cette partie se divise en trois chapitres. Dans le chapitre 4 sont fournies les mesures d'information et d'estimation relatives aux modèles inadéquats, ainsi que certaines relations établies entre ces mesures dans le cas d'un canal gaussien. Les chapitres 5 et 6 sont consacrés à l'étude de l'interaction des modalités sous inadéquation du modèle. Nous examinons principalement, l'impact des fausses hypothèses sur les liens entre les modalités.

Mesures d'information et d'estimation

Sommaire

4.1	Mesures d'information	78
4.1.1	Entropie inadéquate et entropie différentielle inadéquate	78
4.1.2	Information mutuelle inadéquate	81
4.2	Mesures d'estimation	83
4.2.1	Estimateur optimal au sens de l'erreur quadratique moyenne	83
4.3	Information versus estimation	84
4.3.1	Canal gaussien	85
4.3.2	Modèle gaussien	93
4.3.3	Exemple scalaire	98
4.4	Conclusion	106

Un modèle est dit adéquat si les densités de probabilités supposées correspondent aux densités réelles des données. Autrement, il est dit inadéquat. Dans la littérature, on présume souvent l'adéquation du modèle analytique (supposé) avec le modèle des données. Cependant, en raison des imperfections des mesures, la complexité des systèmes physiques étudiés et des appareils de mesures, etc, ce postulat est loin d'être vrai. C'est pourquoi, il est intéressant d'étudier la sensibilité du modèle des données aux fausses hypothèses et de déterminer, si possible, les modèles inadéquats qui assurent la simplicité des calculs ainsi qu'une performance satisfaisante et tolérable.

L'étude des effets des fausses hypothèses date des années soixante. Depuis, peu de travaux se sont intéressés à cette question. En théorie de l'estimation, la plupart des études des modèles inadéquats se sont concentrées sur l'estimation paramétrique, notamment l'estimateur de maximum de vraisemblance. En 1966, Berk a étudié le comportement limite des distributions a posteriori pour un modèle inadéquat et a établi les conditions sous lesquelles ces distributions convergent vers une distribution dégénérée [31, 32]. Huber [33], Akaike [34] et White [35] ont montré que la distribution de l'estimateur de maximum de vraisemblance dans le cas du modèle inadéquat converge asymptotiquement vers une gaussienne. La moyenne de la gaussienne est donnée par la

valeur qui minimise la divergence de Kullback-Leibler entre la loi exacte et la lois upposée des données, alors que la variance est donnée par la matrice “sandwich” de Huber [33].

Récemment, les travaux de recherche se sont tournés vers la généralisation des bornes d'estimation aux modèles inadéquats. Notamment, la borne de Ziv-Zakai par Xu *et al.* [36] et la borne de Cramér-Rao par Richmond et Horowitz [37]. En théorie de l'information, nous trouvons aussi des études sur les modèles inadéquats, en particulier pour déterminer la capacité d'un canal gaussien dont la covariance du bruit diffère de la covariance exacte [38, 39].

Dans cette deuxième partie du manuscrit, nous examinons la sensibilité de la multimodalité aux fausses hypothèses sur les liens entre les modalités. Cependant, les mesures classiques d'estimation et d'information ne prennent pas en compte l'effet d'une inadéquation de modèle. D'où la nécessité de déterminer de nouvelles mesures. La définition de ces mesures est analogue à la définition des mesures classiques. La différence réside en ce que les mesures analytiques sont fonctions de la densité supposée, donc inadéquate, alors que l'espérance est toujours calculée selon la vraie densité des données, car l'espérance ne peut être approchée qu'à partir des données d'observation distribuées selon la vraie loi.

Nous ajoutons l'indice $_{p,q}$ aux notations classiques des mesures d'estimation et d'information. Le premier indice fait référence à la vraie densité, $p(\cdot)$, alors que le deuxième indice correspond à la densité supposée, $q(\cdot)$. Lorsque $q = p$, le modèle supposé est adéquat. Réciproquement, si $q \neq p$, le modèle est inadéquat.

Ce chapitre est organisé comme suit. Nous présentons tout d'abord les mesures fondamentales d'information sous fausses hypothèses, notamment l'entropie et l'information mutuelle (section 4.1). Ensuite, nous introduisons les mesures d'estimation inadéquates (section 4.2). Finalement, la section 4.3 est consacrée aux relations entre les mesures d'information et d'estimation sous inadéquation du modèle.

4.1 Mesures d'information

Dans le cadre du modèle adéquat, nous avons fait la différence entre l'entropie et l'entropie différentielle, bien que leurs définitions soient identiques, afin de marquer la différence de l'interprétation de ces deux mesures (chapitre 1). Dans ce chapitre, nous ne considérons qu'une seule définition. Ainsi l'espérance fait référence à une somme dans le cas discret et à une intégrale dans le cas continu.

4.1.1 Entropie inadéquate et entropie différentielle inadéquate

Soit \mathbf{x} un vecteur aléatoire de densité $p(\mathbf{x})$. Quand nous supposons que la densité de \mathbf{x} est $q(\mathbf{x})$, au lieu de la vraie densité $p(\mathbf{x})$, l'incertitude moyenne sur \mathbf{x} avant mesure est donnée par l'entropie inadéquate, appelée parfois entropie croisée (“*cross entropy*”). Elle est définie, par analogie à l'entropie classique, comme l'espérance de l'incertitude $\log\left(\frac{1}{q(\mathbf{x})}\right)$ où l'espérance est calculée selon la vraie densité $p(\mathbf{x})$, ainsi :

$$h_{p,q}(\mathbf{x}) \triangleq \mathbb{E}_p \left[\log \left(\frac{1}{q(\mathbf{x})} \right) \right]. \quad (4.1)$$

L'entropie inadéquate est liée à l'entropie classique par la divergence de Kullbak-Leibler entre la densité réelle de \mathbf{x} , $p(\mathbf{x})$, et la densité supposée, $q(\mathbf{x})$:

$$h_{p,q}(\mathbf{x}) = h_{p,p}(\mathbf{x}) + D_{KL}(p(\mathbf{x})||q(\mathbf{x})). \quad (4.2)$$

Démonstration.

$$D_{KL}(p(\mathbf{x})||q(\mathbf{x})) = \mathbb{E}_p \left[\log \left(\frac{p(\mathbf{x})}{q(\mathbf{x})} \right) \right], \quad (4.3)$$

$$= \mathbb{E}_p \left[\log \left(\frac{1}{q(\mathbf{x})} \right) \right] - \mathbb{E}_p \left[\log \left(\frac{1}{p(\mathbf{x})} \right) \right], \quad (4.4)$$

$$= h_{p,q}(\mathbf{x}) - h_{p,p}(\mathbf{x}). \quad (4.5)$$

□

D'après la non-négativité de la divergence de Kullback-Leibler, $h_{p,q}(\mathbf{x}) \geq h_{p,p}(\mathbf{x})$ avec égalité si et seulement si $q(\mathbf{x}) = p(\mathbf{x})$.

L'entropie classique, $h_{p,p}(\mathbf{x})$, est une mesure d'information et d'incertitude. Elle quantifie le taux d'information que \mathbf{x} présente après mesure, et donc réciproquement l'incertitude que nous avons sur \mathbf{x} avant mesure. Cependant, l'entropie croisée $h_{p,q}(\mathbf{x})$, ne quantifie pas l'information que nous allons apprendre sur \mathbf{x} après mesure. Elle représente, uniquement, l'incertitude que nous avons sur \mathbf{x} en supposant une fausse densité. Puisque $h_{p,q}(\mathbf{x}) \geq h_{p,p}(\mathbf{x})$, nous déduisons alors qu'une fausse hypothèse sur la densité d'un vecteur aléatoire augmente son incertitude. En théorie de codage, $h_{p,q}$ quantifie le nombre de bits requis pour coder des événements distribués selon $p(\cdot)$, en utilisant un code optimisé pour une autre densité $q(\cdot)$.

Exemple 4.1. Soit $\mathbf{x} \sim p(\mathbf{x}) = \mathcal{N}(\mu_{\mathbf{x}}, \Sigma_{\mathbf{x}})$ un vecteur gaussien de dimension k . Si nous croyons que la densité de \mathbf{x} est $q(\mathbf{x}) = \mathcal{N}(\hat{\mu}_{\mathbf{x}}, \hat{\Sigma}_{\mathbf{x}})$, alors l'entropie inadéquate de \mathbf{x} est :

$$h_{p,q}(\mathbf{x}) = \frac{1}{2} \log \left((2\pi)^k |\hat{\Sigma}_{\mathbf{x}}| \right) + \frac{1}{2} \text{Tr} \left\{ \hat{\Sigma}_{\mathbf{x}}^{-1} \Sigma_{\mathbf{x}} \right\} + \frac{1}{2} (\mu_{\mathbf{x}} - \hat{\mu}_{\mathbf{x}})^T \hat{\Sigma}_{\mathbf{x}}^{-1} (\mu_{\mathbf{x}} - \hat{\mu}_{\mathbf{x}}). \quad (4.6)$$

Contrairement à l'entropie classique (paragraphe 1.1.2 de la partie I), l'entropie inadéquate d'un vecteur gaussien dépend des moyennes exacte, $\mu_{\mathbf{x}}$, et supposée, $\hat{\mu}_{\mathbf{x}}$, de \mathbf{x} à moins qu'elles ne soient égales. Si $\hat{\mu}_{\mathbf{x}} = \mu_{\mathbf{x}}$, alors :

$$h_{p,q}(\mathbf{x}) = \frac{1}{2} \log \left((2\pi)^k |\hat{\Sigma}_{\mathbf{x}}| \right) + \frac{1}{2} \text{Tr} \left\{ \hat{\Sigma}_{\mathbf{x}}^{-1} \Sigma_{\mathbf{x}} \right\}. \quad (4.7)$$

Si de plus $\hat{\Sigma}_{\mathbf{x}} = \Sigma_{\mathbf{x}}$, le deuxième terme de la partie droite de (4.7), *i.e* $\frac{1}{2} \text{Tr} \left\{ \hat{\Sigma}_{\mathbf{x}}^{-1} \Sigma_{\mathbf{x}} \right\}$, devient $\frac{k}{2}$ et $h_{p,q}(\mathbf{x}) = h_{p,p}(\mathbf{x})$.

Démonstration. Nous avons, $p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^k |\Sigma_{\mathbf{x}}|}} \exp \left(-\frac{1}{2} (\mathbf{x} - \mu_{\mathbf{x}})^T \Sigma_{\mathbf{x}}^{-1} (\mathbf{x} - \mu_{\mathbf{x}}) \right)$ et

$q(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^k |\widehat{\Sigma}_{\mathbf{x}}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \hat{\mu}_{\mathbf{x}})^T \widehat{\Sigma}_{\mathbf{x}}^{-1} (\mathbf{x} - \hat{\mu}_{\mathbf{x}})\right)$, ainsi :

$$h_{p,q}(\mathbf{x}) = -\mathbb{E}_p[\log(q(\mathbf{x}))], \quad (4.8)$$

$$= \frac{1}{2} \mathbb{E}_p \left[\log \left((2\pi)^k |\widehat{\Sigma}_{\mathbf{x}}| \right) \right] + \frac{1}{2} \mathbb{E}_p \left[\text{Tr} \left\{ (\mathbf{x} - \hat{\mu}_{\mathbf{x}})^T \widehat{\Sigma}_{\mathbf{x}}^{-1} (\mathbf{x} - \hat{\mu}_{\mathbf{x}}) \right\} \right], \quad (4.9)$$

$$= \frac{1}{2} \log \left((2\pi)^k |\widehat{\Sigma}_{\mathbf{x}}| \right) + \frac{1}{2} \mathbb{E}_p \left[\text{Tr} \left\{ \widehat{\Sigma}_{\mathbf{x}}^{-1} (\mathbf{x} - \hat{\mu}_{\mathbf{x}}) (\mathbf{x} - \hat{\mu}_{\mathbf{x}})^T \right\} \right], \quad (4.10)$$

$$= \frac{1}{2} \log \left((2\pi)^k |\widehat{\Sigma}_{\mathbf{x}}| \right) + \frac{1}{2} \text{Tr} \left\{ \widehat{\Sigma}_{\mathbf{x}}^{-1} \Sigma_{\mathbf{x}} \right\} + \frac{1}{2} (\mu_{\mathbf{x}} - \hat{\mu}_{\mathbf{x}})^T \widehat{\Sigma}_{\mathbf{x}}^{-1} (\mu_{\mathbf{x}} - \hat{\mu}_{\mathbf{x}}). \quad (4.11)$$

où (4.9) s'ensuit car $(\mathbf{x} - \hat{\mu}_{\mathbf{x}})^T \widehat{\Sigma}_{\mathbf{x}}^{-1} (\mathbf{x} - \hat{\mu}_{\mathbf{x}})$ est une forme quadratique ainsi $(\mathbf{x} - \hat{\mu}_{\mathbf{x}})^T \widehat{\Sigma}_{\mathbf{x}}^{-1} (\mathbf{x} - \hat{\mu}_{\mathbf{x}}) = \text{Tr} \left\{ (\mathbf{x} - \hat{\mu}_{\mathbf{x}})^T \widehat{\Sigma}_{\mathbf{x}}^{-1} (\mathbf{x} - \hat{\mu}_{\mathbf{x}}) \right\}$. La relation (4.10) résulte de l'invariance de la trace à la permutation circulaire. Et la formule (4.11) découle d'une permutation de la trace et l'espérance puisque la trace est linéaire. \square

Nous définissons les entropies inadéquates conditionnelle et conjointe similairement.

Entropie conjointe inadéquate. Soient \mathbf{x} et \mathbf{y} deux variables aléatoires de densité jointe $p(\mathbf{x}, \mathbf{y})$. Si nous supposons que la densité jointe de \mathbf{x} et \mathbf{y} est $q(\mathbf{x}, \mathbf{y}) \neq p(\mathbf{x}, \mathbf{y})$, l'incertitude moyenne de (\mathbf{x}, \mathbf{y}) est l'espérance, selon $p(\mathbf{x}, \mathbf{y})$, de $\log \left(\frac{1}{q(\mathbf{x}, \mathbf{y})} \right)$:

$$h_{p,q}(\mathbf{x}, \mathbf{y}) \triangleq \mathbb{E}_p \left[\log \left(\frac{1}{q(\mathbf{x}, \mathbf{y})} \right) \right]. \quad (4.12)$$

L'entropie conjointe inadéquate est liée à l'entropie (1.2) et (1.14) par la divergence de Kullback-Leibler entre la loi jointe exacte, $p(\mathbf{x}, \mathbf{y})$, et supposée, $q(\mathbf{x}, \mathbf{y})$:

$$h_{p,q}(\mathbf{x}, \mathbf{y}) = h_{p,p}(\mathbf{x}, \mathbf{y}) + D_{KL}(p(\mathbf{x}, \mathbf{y}) || q(\mathbf{x}, \mathbf{y})), \quad (4.13)$$

où $D_{KL}(p(\mathbf{x}, \mathbf{y}) || q(\mathbf{x}, \mathbf{y})) = \int \int p(\mathbf{x}, \mathbf{y}) \log \left(\frac{p(\mathbf{x}, \mathbf{y})}{q(\mathbf{x}, \mathbf{y})} \right) d\mathbf{x} d\mathbf{y}$.

Entropie conditionnelle inadéquate. Soit \mathbf{y} un vecteur aléatoire de densité $p(\mathbf{y})$. Nous supposons que la densité conditionnelle de \mathbf{x} sachant \mathbf{y} est $q(\mathbf{x}|\mathbf{y})$ au lieu de la vraie densité $p(\mathbf{x}|\mathbf{y})$. L'entropie de \mathbf{x} conditionnellement à \mathbf{y} est l'espérance selon la vraie densité jointe $p(\mathbf{x}, \mathbf{y})$ de $\log \left(\frac{1}{q(\mathbf{x}|\mathbf{y})} \right)$:

$$h_{p,q}(\mathbf{x}|\mathbf{y}) \triangleq \mathbb{E}_p \left[\log \left(\frac{1}{q(\mathbf{x}|\mathbf{y})} \right) \right]. \quad (4.14)$$

L'entropie conditionnelle est aussi connectée à l'entropie conditionnelle inadéquate par le biais de la divergence de Kullback-Leibler entre les lois de probabilités supposée et exacte :

$$h_{p,q}(\mathbf{x}|\mathbf{y}) = h_{p,p}(\mathbf{x}|\mathbf{y}) + D_{KL}(p(\mathbf{x}|\mathbf{y}) || q(\mathbf{x}|\mathbf{y})). \quad (4.15)$$

où $D_{KL}(p(\mathbf{x}|\mathbf{y}) || q(\mathbf{x}|\mathbf{y})) = \int \int p(\mathbf{x}, \mathbf{y}) \log \left(\frac{p(\mathbf{x}|\mathbf{y})}{q(\mathbf{x}|\mathbf{y})} \right) d\mathbf{x} d\mathbf{y}$.

Dans le cas du modèle adéquat (paragraphe 1.1.1 de la partie I), l'incertitude de \mathbf{x} est réduite par l'observation d'un vecteur \mathbf{y} qui dépend de \mathbf{x} , *i.e* $h_{p,p}(\mathbf{x}|\mathbf{y}) \leq h_{p,p}(\mathbf{x})$. Ce qui n'est pas toujours vrai pour un modèle inadéquat.

Exemple 4.2. Soit $y = x + n$, avec $x \sim \mathcal{N}(0, 2)$ indépendant de n qui suit la loi $p(n) = \mathcal{N}(0, 1)$ et dont la loi supposée est $q(n) = \mathcal{N}(0, 2)$. Supposons que la densité de x soit bien définie, l'entropie inadéquate de x est égale à l'entropie adéquate (exemple 1.1), ainsi :

$$h_{p,q}(x) = h_{p,p}(x) = \frac{1}{2} \log(4\pi e) \approx 1.26, \quad (4.16)$$

En revanche, l'incertitude de x conditionnellement à y est quantifiée par l'entropie inadéquate (4.7), car en raison de l'erreur sur la loi de n , la loi de $x|y$ est fautive, $q(x|y) = \mathcal{N}(\frac{1}{2}y, 1)$ et $p(x|y) = \mathcal{N}(\frac{2}{3}y, \frac{2}{3})$:

$$h_{p,q}(x|y) = \frac{1}{2} \log(2\pi) + \frac{1}{3} + \frac{1}{18} \approx 1.31 \geq h_{p,q}(x). \quad (4.17)$$

Par conséquent, le conditionnement ne réduit pas forcément l'incertitude sous fausses hypothèses.

Sous inadéquation de modèle, le conditionnement ne réduit pas toujours l'incertitude.

4.1.2 Information mutuelle inadéquate

Dans le chapitre 1, nous avons défini l'information mutuelle classique comme une divergence de Kullback-Leibler (1.18) ou comme une différence d'entropies (1.19). L'information mutuelle est aussi l'espérance du logarithme d'un rapport de vraisemblance.

Soient \mathbf{x} et \mathbf{y} deux vecteurs aléatoires de densité jointe $p(\mathbf{x}, \mathbf{y})$ et de densités marginales $p(\mathbf{x})$ et $p(\mathbf{y})$. L'information mutuelle, dans le cadre d'un modèle adéquat, est l'espérance du logarithme du rapport de vraisemblance, $V_p(\mathbf{x}, \mathbf{y}) = \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})p(\mathbf{y})}$:

$$I_{p,p}(\mathbf{x}, \mathbf{y}) = \mathbb{E}_p \left[\log V_p(\mathbf{x}, \mathbf{y}) \right]. \quad (4.18)$$

Sous inadéquation de modèle, le rapport de vraisemblance $V_p(\mathbf{x}, \mathbf{y})$ devient $V_q(\mathbf{x}, \mathbf{y}) = \frac{q(\mathbf{x}, \mathbf{y})}{q(\mathbf{x})q(\mathbf{y})}$. Par analogie, nous définissons l'information mutuelle inadéquate comme l'espérance (sur la vraie loi) du logarithme du rapport de vraisemblance $V_q(\mathbf{x}, \mathbf{y})$:

$$I_{p,q}(\mathbf{x}; \mathbf{y}) \triangleq \mathbb{E}_p \left[\log V_q(\mathbf{x}, \mathbf{y}) \right] = \int \int p(\mathbf{x}, \mathbf{y}) \log \left(\frac{q(\mathbf{x}, \mathbf{y})}{q(\mathbf{x})q(\mathbf{y})} \right) d\mathbf{x}d\mathbf{y}. \quad (4.19)$$

$I_{p,q}(\mathbf{x}; \mathbf{y})$ est symétrique en \mathbf{x} et \mathbf{y} . Contrairement à l'information mutuelle classique, $I_{p,q}$ n'est pas une divergence de Kullback-Leibler. Par conséquent, sa positivité n'est pas garantie. En fait, elle peut être négative. A partir de (4.19) nous pouvons écrire $I_{p,q}$ comme une différence d'entropies inadéquates :

$$I_{p,q}(\mathbf{x}; \mathbf{y}) = h_{p,q}(\mathbf{x}) - h_{p,q}(\mathbf{x}|\mathbf{y}) = h_{p,q}(\mathbf{y}) - h_{p,q}(\mathbf{y}|\mathbf{x}) \quad (4.20)$$

Etant donné que le conditionnement ne réduit pas forcément l'incertitude (exemple 4.2), si $h_{p,q}(\mathbf{x}) < h_{p,q}(\mathbf{x}|\mathbf{y})$ alors $I_{p,q}(\mathbf{x}; \mathbf{y})$ est négative. C'est le cas si l'observation de \mathbf{y} augmente

l'incertitude moyenne sur \mathbf{x} . $I_{p,q}(\mathbf{x}; \mathbf{y})$ n'est pas une mesure d'information puisqu'elle peut être négative. Néanmoins, du fait qu'elle est une généralisation de l'information mutuelle, nous l'appelons "information mutuelle inadéquate".

L'information mutuelle inadéquate $I_{p,q}$ est liée à l'information mutuelle $I_{p,p}$ par les relations suivantes :

$$I_{p,p}(\mathbf{x}; \mathbf{y}) - I_{p,q}(\mathbf{x}; \mathbf{y}) = D_{KL}(p(\mathbf{x}, \mathbf{y})||q(\mathbf{x}, \mathbf{y})) - D_{KL}(p(\mathbf{x})||q(\mathbf{x})) - D_{KL}(p(\mathbf{y})||q(\mathbf{y})) \quad (4.21)$$

$$= D_{KL}(p(\mathbf{y}|\mathbf{x})||q(\mathbf{y}|\mathbf{x})) - D_{KL}(p(\mathbf{y})||q(\mathbf{y})) \quad (4.22)$$

$$= D_{KL}(p(\mathbf{x}|\mathbf{y})||q(\mathbf{x}|\mathbf{y})) - D_{KL}(p(\mathbf{x})||q(\mathbf{x})) \quad (4.23)$$

Démonstration. En utilisant (4.18) et (4.19), on obtient :

$$\begin{aligned} I_{p,p}(\mathbf{x}; \mathbf{y}) - I_{p,q}(\mathbf{x}; \mathbf{y}) &= \mathbb{E}_p \left[\log \left(\frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} \right) \right] - \mathbb{E}_p \left[\log \left(\frac{q(\mathbf{x}, \mathbf{y})}{q(\mathbf{x})q(\mathbf{y})} \right) \right] \end{aligned} \quad (4.24)$$

$$= \mathbb{E}_p \left[\log \left(\frac{p(\mathbf{x}, \mathbf{y})}{q(\mathbf{x}, \mathbf{y})} \right) \right] - \mathbb{E}_p \left[\log \left(\frac{p(\mathbf{x})}{q(\mathbf{x})} \right) \right] - \mathbb{E}_p \left[\log \left(\frac{p(\mathbf{y})}{q(\mathbf{y})} \right) \right] \quad (4.25)$$

$$= D_{KL}(p(\mathbf{x}, \mathbf{y})||q(\mathbf{x}, \mathbf{y})) - D_{KL}(p(\mathbf{x})||q(\mathbf{x})) - D_{KL}(p(\mathbf{y})||q(\mathbf{y})) \quad (4.26)$$

De plus, (4.22) et (4.23) s'ensuivent en utilisant :

$$D_{KL}(p(\mathbf{x}, \mathbf{y})||q(\mathbf{x}, \mathbf{y})) = D_{KL}(p(\mathbf{x}|\mathbf{y})||q(\mathbf{x}|\mathbf{y})) + D_{KL}(p(\mathbf{y})||q(\mathbf{y})), \quad (4.27)$$

$$= D_{KL}(p(\mathbf{y}|\mathbf{x})||q(\mathbf{y}|\mathbf{x})) + D_{KL}(p(\mathbf{x})||q(\mathbf{x})). \quad (4.28)$$

□

La différence $I_{p,p}(\mathbf{x}; \mathbf{y}) - I_{p,q}(\mathbf{x}; \mathbf{y})$ est égale à une différence de divergences de Kullback-Leibler. Ainsi, elle peut être négative, $I_{p,p}(\mathbf{x}; \mathbf{y}) < I_{p,q}(\mathbf{x}; \mathbf{y})$.

Exemple 4.3. Soient \mathbf{x} et \mathbf{z} deux vecteurs aléatoires indépendants de densités de probabilité $p(\mathbf{x})$ et $p(\mathbf{z})$, respectivement. Soit \mathbf{y} un vecteur aléatoire tel que $\mathbf{y} = \mathbf{x} + \mathbf{z}$.

Nous supposons que la densité de \mathbf{z} est adéquate, $q(\mathbf{z}) = p(\mathbf{z})$, alors que la densité de \mathbf{x} est inadéquate, $q(\mathbf{x}) \neq p(\mathbf{x})$. Dans ce cas, la densité de probabilité de \mathbf{y} est inadéquate $q(\mathbf{y}) \neq p(\mathbf{y})$. Cependant, la densité conditionnelle de \mathbf{y} sachant \mathbf{x} est adéquate, $q(\mathbf{y}|\mathbf{x}) = p(\mathbf{y}|\mathbf{x})$.

A partir de (4.22), puisque $D_{KL}(p(\mathbf{y}|\mathbf{x})||q(\mathbf{y}|\mathbf{x})) = 0$, nous trouvons que :

$$I_{p,p}(\mathbf{x}; \mathbf{y}) - I_{p,q}(\mathbf{x}; \mathbf{y}) = -D_{KL}(p(\mathbf{y})||q(\mathbf{y})) \leq 0. \quad (4.29)$$

En neurosciences, un cas particulier de $I_{p,q}$, souvent désigné par I_{shuffle} , est largement utilisé. I_{shuffle} permet d'évaluer l'importance des corrélations des bruits pour extraire de l'information des réponses neuronales à un stimulus [40, 41]. Elle quantifie l'information fournie sur un stimulus par les réponses de deux neurones si la corrélation des bruits est ignorée. L'indice "shuffle" est

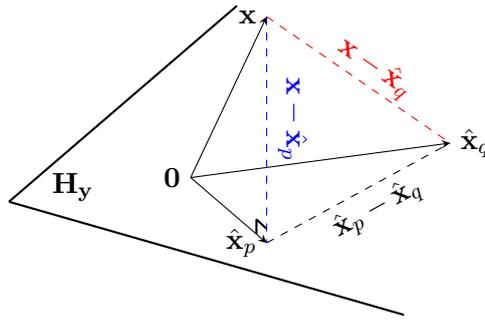


FIGURE 4.1 – Estimateur inadéquat par projection

utilisé car le mélange des données permet de supprimer la corrélation des bruits.

Sous inadéquation de modèle, l'information mutuelle peut être négative.

4.2 Mesures d'estimation

4.2.1 Estimateur optimal au sens de l'erreur quadratique moyenne

Soit $\mathbf{x} \sim p(\mathbf{x})$ un vecteur aléatoire que nous désirons estimer, et $\mathbf{y} \sim p(\mathbf{y})$ une mesure de \mathbf{x} . Nous supposons une densité a posteriori, $q(\mathbf{x}|\mathbf{y})$, différente de la vraie densité, $p(\mathbf{x}|\mathbf{y})$. Dans ce cas, nous calculons analytiquement l'estimateur de \mathbf{x} en fonction de \mathbf{y} , que nous croyons "optimal" au sens de l'erreur quadratique moyenne, comme l'espérance de la densité a posteriori supposée $q(\mathbf{x}|\mathbf{y})$:

$$\hat{\mathbf{x}}_q(\mathbf{y}) \triangleq \mathbb{E}_q[\mathbf{x}|\mathbf{y}] = \int \mathbf{x} q(\mathbf{x}|\mathbf{y}) d\mathbf{x}. \quad (4.30)$$

L'erreur d'estimation inadéquate associée à $\hat{\mathbf{x}}_q(\mathbf{y})$ est la matrice de covariance de l'erreur $\mathbf{x} - \hat{\mathbf{x}}_q(\mathbf{y})$. L'espérance selon la densité de \mathbf{y} peut être approchée par les données. Par conséquent, l'espérance de la matrice de covariance est selon la vraie densité $p(\mathbf{y})$:

$$\mathbf{M}_{p,q}(\mathbf{x}|\mathbf{y}) = \mathbb{E}_p[(\mathbf{x} - \hat{\mathbf{x}}_q(\mathbf{y}))(\mathbf{x} - \hat{\mathbf{x}}_q(\mathbf{y}))^T]. \quad (4.31)$$

Soit $H_{\mathbf{y}}$ le sous-espace des observations muni du produit scalaire $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbb{E}[\mathbf{u}^T \mathbf{v}]$. Dans le cadre d'un modèle adéquat, puisque l'estimateur optimal, $\hat{\mathbf{x}}_p(\mathbf{y}) = \mathbb{E}_p[p(\mathbf{x}|\mathbf{y})]$, minimise l'erreur quadratique moyenne, il est la projection orthogonale de \mathbf{x} sur $H_{\mathbf{y}}$. Cependant, l'estimateur inadéquat, $\hat{\mathbf{x}}_q(\mathbf{y})$, est une projection oblique de \mathbf{x} sur $H_{\mathbf{y}}$ (figure 4.1). La différence entre les erreurs d'estimation associées aux deux estimateurs est donnée par la matrice de covariance :

$$\mathbf{M}_{p,q}(\mathbf{x}|\mathbf{y}) - \mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y}) = \mathbb{E}_p[(\hat{\mathbf{x}}_p(\mathbf{y}) - \hat{\mathbf{x}}_q(\mathbf{y}))(\hat{\mathbf{x}}_p(\mathbf{y}) - \hat{\mathbf{x}}_q(\mathbf{y}))^T]. \quad (4.32)$$

La différence $\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) - \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})$, appelée dans la littérature "excédent de l'erreur quadratique moyenne" ("excess MSE" en anglais), est semi-définie positive et est nulle si et seulement si $q = p$. Par conséquent, les fausses hypothèses sur le modèle conduisent à plus d'erreur d'estimation. L'excédent de l'EQM permet de quantifier les erreurs additionnelles engendrées par le modèle inadéquat.

Démonstration. A partir de (1.27) et (4.31) on obtient :

$$\begin{aligned} \mathbf{M}_{p,q}(\mathbf{x}|\mathbf{y}) - \mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y}) &= \mathbb{E}_p[(\mathbf{x} - \hat{\mathbf{x}}_q(\mathbf{y}))(\mathbf{x} - \hat{\mathbf{x}}_q(\mathbf{y}))^T] - \mathbb{E}_p[(\mathbf{x} - \hat{\mathbf{x}}_p(\mathbf{y}))(\mathbf{x} - \hat{\mathbf{x}}_p(\mathbf{y}))^T], \end{aligned} \quad (4.33)$$

$$= \mathbb{E}_p[(\mathbf{x} - \hat{\mathbf{x}}_q(\mathbf{y}))(\mathbf{x} - \hat{\mathbf{x}}_q(\mathbf{y}))^T] - \mathbb{E}_p[(\mathbf{x} - \hat{\mathbf{x}}_p(\mathbf{y}))(\mathbf{x} - \hat{\mathbf{x}}_q(\mathbf{y}))^T], \quad (4.34)$$

$$= \mathbb{E}_p[(\hat{\mathbf{x}}_p(\mathbf{y}) - \hat{\mathbf{x}}_q(\mathbf{y}))(\mathbf{x} - \hat{\mathbf{x}}_q(\mathbf{y}))^T], \quad (4.35)$$

$$= \mathbb{E}_p[(\hat{\mathbf{x}}_p(\mathbf{y}) - \hat{\mathbf{x}}_q(\mathbf{y}))(\mathbf{x} - \hat{\mathbf{x}}_p(\mathbf{y}) + \hat{\mathbf{x}}_p(\mathbf{y}) - \hat{\mathbf{x}}_q(\mathbf{y}))^T], \quad (4.36)$$

$$= \mathbb{E}_p[(\hat{\mathbf{x}}_p(\mathbf{y}) - \hat{\mathbf{x}}_q(\mathbf{y}))(\hat{\mathbf{x}}_p(\mathbf{y}) - \hat{\mathbf{x}}_q(\mathbf{y}))^T]. \quad (4.37)$$

où (4.34) et (4.37) résultent du principe d'orthogonalité. Ainsi, l'erreur $\mathbf{x} - \hat{\mathbf{x}}_p(\mathbf{y})$ est orthogonale à $\hat{\mathbf{x}}_p(\mathbf{y})$ et à $\hat{\mathbf{x}}_q(\mathbf{y})$. C'est-à-dire, $\mathbb{E}_p[(\mathbf{x} - \hat{\mathbf{x}}_p(\mathbf{y}))\hat{\mathbf{x}}_p(\mathbf{y})^T] = \mathbb{E}_p[(\mathbf{x} - \hat{\mathbf{x}}_p(\mathbf{y}))\hat{\mathbf{x}}_q(\mathbf{y})^T] = 0$. \square

L'excédent de l'EQM (excess MSE), $\mathbf{M}_{p,q}(\mathbf{x}|\mathbf{y}) - \mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y})$, est semi-définie positive. Autrement dit, l'estimateur sous une fausse loi conduit à une trace de l'EQM plus grande que sous la vraie loi, $\text{Tr}\{\mathbf{M}_{p,q}(\mathbf{x}|\mathbf{y})\} \geq \text{Tr}\{\mathbf{M}_{p,p}(\mathbf{x}|\mathbf{y})\}$.

4.3 Information versus estimation

Dans le cadre du modèle adéquat, plusieurs relations ont été établies entre les mesures de la théorie de l'information et les mesures de la théorie de l'estimation (section 1.3). Dans le cadre du modèle inadéquat, il existe également des relations généralisant ces liens.

A notre connaissance, c'est à Akaike que nous devons la première relation jamais établie entre la théorie de l'estimation et la théorie de l'information sous inadéquation du modèle. En 1974, Akaike a montré que, sous inadéquation de modèle, l'estimateur de maximum de vraisemblance converge vers la valeur qui minimise la divergence de Kullback-Leibler entre la densité exacte et la densité supposée des données [34].

Dans cette section nous fournissons un aperçu sur une relation récemment proposée par Verdù [42] et qui a été généralisée au cas vectoriel par Chen et Lafferty [43]. Cette relation relie l'excédent de l'EQM (4.32) et la divergence de Kullback-Leibler dans le cas d'inadéquation de la densité du signal d'entrée dans un canal gaussien. Nous proposons, ensuite, une généralisation de cette relation au cas d'inadéquation du bruit et au cas d'une double inadéquation du signal utile et du bruit. Finalement, nous en déduisons une généralisation de la relation entre information mutuelle et l'EQM (I-MMSE), décrite à la section 1.3, au modèle inadéquat.

Cette section se divise en trois parties. La première partie présente deux relations établies entre la théorie de l'information et la théorie de l'estimation dans un canal gaussien et sous trois différentes inadéquations du modèle. La seconde partie exprime ces relations dans le cas particulier du modèle gaussien, *i.e* canal gaussien avec entrée gaussienne. La troisième section analyse ces relations à travers un exemple scalaire.

4.3.1 Canal gaussien

Les relations que nous présentons dans la suite sont développées dans le canal gaussien linéaire bruité schématisé par la figure 4.2 :

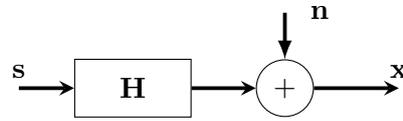


FIGURE 4.2 – Canal gaussien linéaire

Le modèle mathématique correspondant à la figure 4.2 est :

$$\mathbf{x} = \mathbf{H} \mathbf{s} + \mathbf{n}, \quad (4.38)$$

où $\mathbf{s} \in \mathbb{R}^l$ est le signal (vectoriel) d'entrée de densité $p(\mathbf{s})$, $\mathbf{x} \in \mathbb{R}^k$ est le signal (vectoriel) de sortie de densité $p(\mathbf{x})$, $\mathbf{n} \sim \mathcal{N}(0, \mathbf{\Sigma})$ est un bruit gaussien additif de moyenne nulle et de matrice de covariance $\mathbf{\Sigma}$ et $\mathbf{H} \in \mathbb{R}^{k \times l}$ est la matrice du canal.

Dans la suite, nous considérons trois types d'inadéquations du modèle (4.38). Une première inadéquation de la densité du signal utile, $q(\mathbf{s}) \neq p(\mathbf{s})$. Ensuite, une inadéquation sur la matrice de covariance du bruit, $q(\mathbf{n}) = (0, \widehat{\mathbf{\Sigma}}) \neq p(\mathbf{n})$. Finalement, une double inadéquation sur les densités du signal utile et du bruit, $q(\mathbf{s}) \neq p(\mathbf{s})$ et $q(\mathbf{n}) = (0, \widehat{\mathbf{\Sigma}}) \neq p(\mathbf{n})$.

Ces trois inadéquations engendrent différentes inadéquations de la densité de la sortie \mathbf{x} . Cependant, en guise de simplification, nous avons choisi de noter la distribution inadéquate de \mathbf{x} par $q(\mathbf{x})$ indifféremment du type d'inadéquation. Pourtant, le type d'inadéquation devient apparent en exprimant la densité de \mathbf{x} en fonction des densités de \mathbf{s} et de \mathbf{n} , soit :

$$q(\mathbf{x}) = \begin{cases} \int p(\mathbf{x}|\mathbf{s})q(\mathbf{s})d\mathbf{s} & \text{si } q(\mathbf{s}) \neq p(\mathbf{s}) \text{ et } \widehat{\mathbf{\Sigma}} = \mathbf{\Sigma} \\ \int q(\mathbf{x}|\mathbf{s})p(\mathbf{s})d\mathbf{s} & \text{si } q(\mathbf{s}) = p(\mathbf{s}) \text{ et } \widehat{\mathbf{\Sigma}} \neq \mathbf{\Sigma} . \\ \int q(\mathbf{x}|\mathbf{s})q(\mathbf{s})d\mathbf{s} & \text{si } q(\mathbf{s}) \neq p(\mathbf{s}) \text{ et } \widehat{\mathbf{\Sigma}} \neq \mathbf{\Sigma} \end{cases} \quad (4.39)$$

4.3.1.1 Inadéquation du signal d'entrée

Nous supposons la densité du signal d'entrée $q(\mathbf{s})$ différente de la densité exacte, $p(\mathbf{s})$. Nous disons dans ce cas que l'entrée du canal est inadéquate. Suite à cette hypothèse, la densité de la sortie \mathbf{x} est aussi inadéquate, $q(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{s})q(\mathbf{s})d\mathbf{s} \neq p(\mathbf{x})$. Cependant, la densité conditionnelle de \mathbf{x} sachant \mathbf{s} est adéquate car la densité du bruit supposée correspond à la densité réelle, $q(\mathbf{x}|\mathbf{s}) = p(\mathbf{x}|\mathbf{s})$. La figure 4.3 schématise un canal gaussien avec une entrée inadéquate.

Estimation. Sous inadéquation du signal d'entrée, \mathbf{s} , la densité a posteriori supposée est fautive, $q(\mathbf{s}|\mathbf{x}) \neq p(\mathbf{s}|\mathbf{x})$. Par conséquent, l'estimateur que nous croyons optimal est la moyenne de la loi a posteriori inadéquate, $q(\mathbf{s}|\mathbf{x})$ (section 4.2) :

$$\hat{\mathbf{s}}_q(\mathbf{x}) = \int q(\mathbf{s}|\mathbf{x}) \mathbf{s} d\mathbf{s}, \quad (4.40)$$

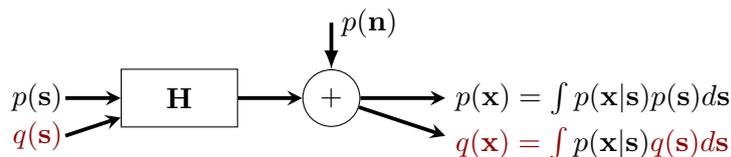


FIGURE 4.3 – Canal gaussien linéaire sous inadéquation de l'entrée

$$= \int \frac{q(\mathbf{s})p(\mathbf{x}|\mathbf{s})}{q(\mathbf{x})} \mathbf{s} d\mathbf{s}. \quad (4.41)$$

L'équation (4.41) est obtenue par la relation de Bayes (1.26) où $p(\mathbf{x}|\mathbf{s}) = \mathcal{N}(\mathbf{H}\mathbf{s}, \mathbf{\Sigma})$ et $q(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{s})q(\mathbf{s})d\mathbf{s}$. L'erreur d'estimation relative à $\hat{\mathbf{s}}_q(\mathbf{x})$ est la matrice de covariance de l'erreur $\mathbf{s} - \hat{\mathbf{s}}_q(\mathbf{x})$ (4.31) :

$$\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) = \mathbb{E}_p[(\mathbf{s} - \hat{\mathbf{s}}_q(\mathbf{x}))(\mathbf{s} - \hat{\mathbf{s}}_q(\mathbf{x}))^T]. \quad (4.42)$$

Information. Quand la densité de l'entrée \mathbf{s} est mal définie, l'information mutuelle entre l'entrée et la sortie du canal est donnée par (4.19) :

$$I_{p,q}(\mathbf{s}; \mathbf{x}) \triangleq \mathbb{E}_p \left[\log \left(\frac{q(\mathbf{s}, \mathbf{x})}{q(\mathbf{s})q(\mathbf{x})} \right) \right] = \mathbb{E}_p \left[\log \left(\frac{p(\mathbf{x}|\mathbf{s})}{q(\mathbf{x})} \right) \right], \quad (4.43)$$

$$= I_{p,p}(\mathbf{s}; \mathbf{x}) + D_{KL}(p(\mathbf{x}|\mathbf{s})||q(\mathbf{x})). \quad (4.44)$$

La relation (4.44) est obtenue à partir de (4.23), avec $D_{KL}(p(\mathbf{x}|\mathbf{s})||q(\mathbf{x}|\mathbf{s})) = 0$.

D'après la non-négativité de la divergence de Kullback-Leibler et à partir de (4.44), $I_{p,q}(\mathbf{s}; \mathbf{x}) \geq I_{p,p}(\mathbf{s}; \mathbf{x})$. Etant donné que $I_{p,q}(\mathbf{s}; \mathbf{x})$ n'est pas une mesure d'information, l'inégalité $I_{p,q}(\mathbf{s}; \mathbf{x}) \geq I_{p,p}(\mathbf{s}; \mathbf{x})$ ne signifie pas que, sous inadéquation du modèle, nous apprenons plus d'information sur s . Essayons de comprendre le sens de cette inégalité, en écrivant $I_{p,q}(\mathbf{s}; \mathbf{x})$ comme la différence d'entropies. Selon (4.20), nous avons :

$$I_{p,q}(\mathbf{s}; \mathbf{x}) = h_{p,q}(\mathbf{s}) - h_{p,q}(\mathbf{s}|\mathbf{x}) \quad (4.45)$$

En considérant une fausse densité du signal d'entrée \mathbf{s} , le résidu d'incertitude moyenne sur \mathbf{s} après l'observation de \mathbf{x} augmente, $h_{p,q}(\mathbf{s}|\mathbf{x}) \geq h_{p,p}(\mathbf{s}|\mathbf{x})$. Ce qui signifie que \mathbf{x} révèle moins d'information sur \mathbf{s} , par rapport au modèle adéquat. Cependant, l'incertitude moyenne de \mathbf{s} augmente également, $h_{p,q}(\mathbf{s}) \geq h_{p,p}(\mathbf{s})$. Par conséquent, la différence $I_{p,q}(\mathbf{s}; \mathbf{x}) = h_{p,q}(\mathbf{s}) - h_{p,q}(\mathbf{s}|\mathbf{x})$ représente la réduction de l'incertitude sur \mathbf{s} suite à l'inadéquation du modèle, mais ne quantifie pas l'information que nous apprenons effectivement sur \mathbf{s} .

Sous inadéquation de l'entrée d'un canal gaussien, l'information mutuelle inadéquate est supérieure à l'information mutuelle classique, $I_{p,q}(\mathbf{s}; \mathbf{x}) \geq I_{p,p}(\mathbf{s}; \mathbf{x})$.

L'entropie relative et l'excédent de l'EQM. Dans le cas d'une entrée inadéquate, Verdú a établi une relation entre l'excédent de l'EQM et la divergence de Kullback-Leibler [42], qui a été généralisée ensuite au cas vectoriel par Chen et Lafferty [43]. Cette relation énonce que le gradient,

par rapport à la matrice du canal \mathbf{H} , de la divergence de Kullback-Leibler entre la loi exacte de la sortie du canal gaussien, $p(\mathbf{x})$, et la loi supposée, $q(\mathbf{x})$, est directement lié à l'excédent de l'EQM :

Théorème 4.1: (D_{kl} -ExcédentEQM sous inadéquation de la densité d'entrée [43])

Soit le canal gaussien :

$$\mathbf{x} = \mathbf{H} \mathbf{s} + \mathbf{n}, \quad (4.46)$$

où $\mathbf{n} \sim \mathcal{N}(0, \mathbf{\Sigma})$. Sous inadéquation de la densité d'entrée, $q(\mathbf{s}) \neq p(\mathbf{s})$, la divergence de Kullback-Leibler entre la densité exacte de la sortie, $p(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{s})p(\mathbf{s})d\mathbf{s}$, et la densité supposée de la sortie, $q(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{s})q(\mathbf{s})d\mathbf{s}$, du canal est liée à l'excédent de l'EQM par la relation :

$$\nabla_{\mathbf{H}} D_{KL}(p(\mathbf{x})||q(\mathbf{x})) = \mathbf{\Sigma}^{-1} \mathbf{H} (\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) - \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})) \quad (4.47)$$

Cette relation relie une distance au sens métrique, qui est l'excédent de l'EQM, à l'écart entre deux densités. Etant donné que certaines mesures d'information sont des cas particuliers de la divergence de Kullback-Leibler, notamment l'entropie mutuelle et l'entropie, la relation (4.47) est considérée comme une généralisation de la plupart des relations existantes entre la théorie de l'information et de l'estimation, notamment l'I-MMSE (paragraphe 1.3.3) :

$$\nabla_{\mathbf{H}} I_{p,p}(\mathbf{s}; \mathbf{x}) = \mathbf{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x}). \quad (4.48)$$

Information mutuelle et erreur quadratique moyenne. En nous appuyant sur la relation (4.47), nous montrons que la relation I-MMSE (paragraphe 1.3.3 de la partie I) reste valable sous inadéquation de l'entrée d'un canal gaussien. En effet, le gradient de $I_{p,q}(\mathbf{s}; \mathbf{x})$ par rapport à la matrice du canal, \mathbf{H} , est directement lié à l'erreur d'estimation $\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x})$ par le théorème suivant :

Théorème 4.2: (I-MMSE sous inadéquation de la densité d'entrée)

Soit le canal gaussien :

$$\mathbf{x} = \mathbf{H} \mathbf{s} + \mathbf{n}, \quad (4.49)$$

où $\mathbf{n} \sim \mathcal{N}(0, \mathbf{\Sigma})$. Sous inadéquation de la densité d'entrée, $q(\mathbf{s}) \neq p(\mathbf{s})$, l'information mutuelle et l'erreur d'estimation sont liées par la relation :

$$\nabla_{\mathbf{H}} I_{p,q}(\mathbf{s}; \mathbf{x}) = \mathbf{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) \quad (4.50)$$

Démonstration. Selon (4.47), nous avons :

$$\mathbf{\Sigma}^{-1} \mathbf{H} (\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) - \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})) = \nabla_{\mathbf{H}} D_{KL}(p(\mathbf{x})||q(\mathbf{x})) \quad (4.51)$$

$$= \nabla_{\mathbf{H}} I_{p,q}(\mathbf{s}; \mathbf{x}) - \nabla_{\mathbf{H}} I_{p,p}(\mathbf{s}; \mathbf{x}) \quad (4.52)$$

où (4.52) résulte de (4.44). Puisque, $\nabla_{\mathbf{H}} I_{p,p}(\mathbf{s}; \mathbf{x}) = \mathbf{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})$ (paragraphe 1.3.3 de la

partie I), on obtient :

$$\mathbf{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) = \nabla_{\mathbf{H}} I_{p,q}(\mathbf{s}; \mathbf{x}) \quad (4.53)$$

□

4.3.1.2 Inadéquation du bruit

Maintenons, nous supposons un deuxième scénario où seul le bruit du canal est inadéquat. Notamment, la matrice de covariance du bruit est mal modélisée, et $\widehat{\mathbf{\Sigma}}$ est considérée au lieu de la matrice de covariance exacte $\mathbf{\Sigma}$. C'est-à-dire que la densité supposée de \mathbf{n} est $q(\mathbf{n}) = \mathcal{N}(0, \widehat{\mathbf{\Sigma}})$ au lieu de $p(\mathbf{n}) = \mathcal{N}(0, \mathbf{\Sigma})$. La figure 4.4 schématise ce cas de canal gaussien :

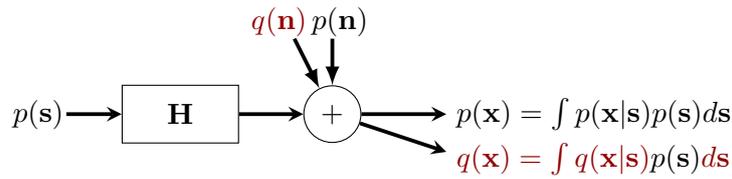


FIGURE 4.4 – Canal gaussien linéaire sous inadéquation du bruit

Estimation. Même si la densité du signal d'entrée est adéquate, $q(\mathbf{s}) = p(\mathbf{s})$, l'inadéquation du bruit fait que la densité a posteriori est inadéquate, $q(\mathbf{s}|\mathbf{x}) \neq p(\mathbf{s}|\mathbf{x})$. Par conséquent, l'estimateur que nous croyons optimal est la moyenne de la loi a posteriori, $q(\mathbf{s}|\mathbf{x})$ (section 4.2) :

$$\hat{\mathbf{s}}_q(\mathbf{x}) = \int q(\mathbf{s}|\mathbf{x}) \mathbf{s} d\mathbf{s}, \quad (4.54)$$

$$= \int \frac{p(\mathbf{s})q(\mathbf{x}|\mathbf{s})}{q(\mathbf{x})} \mathbf{s} d\mathbf{s}. \quad (4.55)$$

L'équation (4.55) est obtenue par la relation de Bayes (1.26) ; où $q(\mathbf{x}|\mathbf{s}) = \mathcal{N}(\mathbf{H}\mathbf{s}, \widehat{\mathbf{\Sigma}})$ et $q(\mathbf{x}) = \int q(\mathbf{x}|\mathbf{s})p(\mathbf{s})d\mathbf{s}$. L'erreur d'estimation relative à $\hat{\mathbf{s}}_q(\mathbf{x})$ est :

$$\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) = \mathbb{E}_p[(\mathbf{s} - \hat{\mathbf{s}}_q(\mathbf{x}))(\mathbf{s} - \hat{\mathbf{s}}_q(\mathbf{x}))^T]. \quad (4.56)$$

Information. Quand nous supposons une fausse densité du bruit, $q(\mathbf{n}) \neq p(\mathbf{n})$, l'information mutuelle entre l'entrée et la sortie du canal est donnée par (4.19) :

$$I_{p,q}(\mathbf{s}; \mathbf{x}) \triangleq \mathbb{E}_p \left[\log \left(\frac{q(\mathbf{s}, \mathbf{x})}{p(\mathbf{s})q(\mathbf{x})} \right) \right] = \mathbb{E}_p \left[\log \left(\frac{q(\mathbf{s}|\mathbf{x})}{p(\mathbf{s})} \right) \right] \quad (4.57)$$

$$= I_{p,p}(\mathbf{s}; \mathbf{x}) - D_{KL}(p(\mathbf{s}|\mathbf{x})||q(\mathbf{s}|\mathbf{x})) \quad (4.58)$$

où (4.58) est obtenue à partir de (4.23), car $D_{KL}(p(\mathbf{s})||q(\mathbf{s})) = 0$. Selon (4.58) et d'après la non-négativité de la divergence de Kullback-Leibler, nous déduisons que $I_{p,q}(\mathbf{s}; \mathbf{x}) \leq I_{p,p}(\mathbf{s}; \mathbf{x})$. Par conséquent, contrairement à l'inadéquation de la densité d'entrée, l'inadéquation du bruit diminue l'information mutuelle. Nous écrivons $I_{p,q}(\mathbf{s}; \mathbf{x})$ comme différence d'entropies :

$$I_{p,q}(\mathbf{s}; \mathbf{x}) = h_{p,p}(\mathbf{s}) - h_{p,q}(\mathbf{s}|\mathbf{x}) \quad (4.59)$$

Sous inadéquation du bruit, l'incertitude moyenne de l'entrée \mathbf{s} demeure invariante, $h_{p,q}(\mathbf{s}) = h_{p,p}(\mathbf{s})$. Cependant, le résidu d'incertitude moyenne sur \mathbf{s} après l'observation de \mathbf{x} augmente, $h_{p,q}(\mathbf{s}|\mathbf{x}) \geq h_{p,p}(\mathbf{s}|\mathbf{x})$. Ainsi, l'incertitude sur \mathbf{s} est moins réduit que dans le cas adéquat. Nous pouvons, seulement dans ce cas, interpréter $I_{p,q}(\mathbf{s}; \mathbf{x})$ comme mesure d'information de \mathbf{s} .

Sous inadéquation du bruit, l'information mutuelle inadéquate est inférieure à l'information mutuelle classique, $I_{p,q}(\mathbf{s}; \mathbf{x}) \leq I_{p,p}(\mathbf{s}; \mathbf{x})$.

Entropie relative et excédent d'erreur quadratique moyenne. La relation établie par Verdú entre l'excédent de l'EQM et la divergence de Kullback-Leibler est limitée à l'inadéquation de l'entrée du canal (théorème 4.1). Nous étudions la relation entre ces deux quantités dans le cas d'inadéquation du bruit et nous montrons que la relation (4.47) n'est plus valable. En effet, le gradient de la divergence de Kullback-Leibler entre les densités exacte et supposée de \mathbf{x} , $D_{KL}(p(\mathbf{x})||q(\mathbf{x}))$, par rapport à la matrice du canal \mathbf{H} , est liée à l'excédent de l'EQM, $\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) - \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})$, par le théorème suivant :

Théorème 4.3: (D_{kl} -ExcédentEQM sous inadéquation du bruit)

Soit le canal gaussien :

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}, \quad (4.60)$$

où $\mathbf{s} \sim p(\mathbf{s})$ et $\mathbf{n} \sim \mathcal{N}(0, \mathbf{\Sigma})$. Sous inadéquation de la matrice de covariance du bruit, $\hat{\mathbf{\Sigma}} \neq \mathbf{\Sigma}$, la divergence de Kullback-Leibler entre la densité exacte de la sortie, $p(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{s})p(\mathbf{s})d\mathbf{s}$, et la densité supposée de la sortie, $q(\mathbf{x}) = \int q(\mathbf{x}|\mathbf{s})p(\mathbf{s})d\mathbf{s}$, est liée à l'excédent de l'EQM par la relation suivante :

$$\nabla_{\mathbf{H}} D_{KL}(p(\mathbf{x})||q(\mathbf{x})) = \hat{\mathbf{\Sigma}}^{-1} \mathbf{H} (\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) - \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})) + \mathbf{R}_D \quad (4.61)$$

$$\text{où } \mathbf{R}_D = -(\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}) \left(\mathbb{E}_p [(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_q^T] - \mathbb{E}_p [(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_p^T] \right).$$

Démonstration. Appendice C

□

Nous remarquons que, comparé à (4.47), l'inadéquation sur le bruit introduit un terme résiduel, \mathbf{R}_D , qui fait que la divergence de Kullback-Leibler n'est plus directement liée à l'excédent de l'EQM. Ce terme résiduel est le produit de deux différences :

- ▶ $\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}$: la différence entre les matrices de covariance des bruits supposée et exacte,
- ▶ $\mathbb{E}_p [(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_q^T] - \mathbb{E}_p [(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_p^T] = \mathbb{E}_p [\mathbf{n}\hat{\mathbf{s}}_q^T] - \mathbb{E}_p [\mathbf{n}\hat{\mathbf{s}}_p^T]$: la différence entre deux matrices de covariance entre le bruit et les estimateurs inadéquats, $\hat{\mathbf{s}}_q$, et adéquat, $\hat{\mathbf{s}}_p$. $\mathbb{E}_p [(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_q^T]$ (resp. $\mathbb{E}_p [(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_p^T]$), reflète la dépendance linéaire entre le bruit et l'estimateur inadéquat (resp. adéquat).

Proposition 4.1. La matrice de covariance $\mathbb{E}_p [(\mathbf{x} - \mathbf{H}\mathbf{s}) \hat{\mathbf{s}}_p^T]$ est directement liée à l'erreur quadratique moyenne de l'estimation adéquate, $\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})$, par la relation :

$$\mathbb{E}_p [(\mathbf{x} - \mathbf{H}\mathbf{s}) \hat{\mathbf{s}}_p^T] = \mathbf{H}\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x}). \quad (4.62)$$

Démonstration. Selon le principe d'orthogonalité :

$$\mathbb{E}_p [\mathbf{x}(\mathbf{s} - \hat{\mathbf{s}}_p)^T] = 0 \Leftrightarrow \mathbb{E}_p [(\mathbf{x} - \mathbf{H}\mathbf{s})(\mathbf{s} - \hat{\mathbf{s}}_p)^T] + \mathbb{E}_p [\mathbf{H}\mathbf{s}(\mathbf{s} - \hat{\mathbf{s}}_p)^T] = 0 \quad (4.63)$$

$$\Leftrightarrow -\mathbb{E}_p [(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_p^T] + \mathbf{H}\mathbb{E}_p [\mathbf{s}(\mathbf{s} - \hat{\mathbf{s}}_p)^T] = 0 \quad (4.64)$$

$$\Leftrightarrow \mathbb{E}_p [(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_p^T] = \mathbf{H}\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x}) \quad (4.65)$$

La deuxième équivalence (4.64) s'ensuit car le bruit $\mathbf{n} = \mathbf{x} - \mathbf{H}\mathbf{s}$ est indépendant de \mathbf{s} . Alors que la dernière équivalence (4.65) est obtenue car $\mathbb{E}_p [\hat{\mathbf{s}}_p(\mathbf{s} - \hat{\mathbf{s}}_p)^T] = 0$. \square

La proposition 4.1 entraîne le corollaire suivant du théorème 4.3 :

Corollaire 4.1. Soit le canal gaussien :

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}, \quad (4.66)$$

où $\mathbf{s} \sim p(\mathbf{s})$ et $\mathbf{n} \sim \mathcal{N}(0, \mathbf{\Sigma})$. Supposons que, la matrice de covariance du bruit soit inadéquate, $\hat{\mathbf{\Sigma}} \neq \mathbf{\Sigma}$, alors :

$$\nabla_{\mathbf{H}} D_{KL}(p(\mathbf{x})||q(\mathbf{x})) = (\hat{\mathbf{\Sigma}}^{-1}\mathbf{H}\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) - \mathbf{\Sigma}^{-1}\mathbf{H}\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})) - (\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1})\mathbb{E}_p [(\mathbf{x} - \mathbf{H}\mathbf{s}) \hat{\mathbf{s}}_q^T]. \quad (4.67)$$

Le corollaire 4.1 relie le gradient de la divergence de Kullback-Leibler, $D_{KL}(p(\mathbf{x})||q(\mathbf{x}))$, à une différence pondérée des erreurs d'estimation : $\hat{\mathbf{\Sigma}}^{-1}\mathbf{H}\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) - \mathbf{\Sigma}^{-1}\mathbf{H}\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})$ plus un terme résiduel.

Information mutuelle et erreur quadratique moyenne. La relation I-MMSE est également non valide sous inadéquation du bruit. En effet, le gradient de $I_{p,q}(\mathbf{s}; \mathbf{x})$ par rapport à la matrice du canal, \mathbf{H} , est liée à l'erreur d'estimation $\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x})$ par le théorème suivant :

Théorème 4.4: (I-MMSE sous inadéquation du bruit)

Soit le canal gaussien :

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}, \quad (4.68)$$

où $\mathbf{s} \sim p(\mathbf{s})$ et $\mathbf{n} \sim \mathcal{N}(0, \mathbf{\Sigma})$. Sous inadéquation du bruit, $q(\mathbf{n}) = \mathcal{N}(0, \hat{\mathbf{\Sigma}}) \neq p(\mathbf{n})$, l'information mutuelle et l'erreur d'estimation sont liées par la relation :

$$\nabla_{\mathbf{H}} I_{p,q}(\mathbf{s}; \mathbf{x}) = \hat{\mathbf{\Sigma}}^{-1}\mathbf{H}\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) + \mathbf{R}_I, \quad (4.69)$$

$$\text{avec, } \mathbf{R}_I = - \left(\widehat{\Sigma}^{-1} - \Sigma^{-1} \right) \mathbb{E}_p[(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_q^T].$$

Démonstration. Selon (4.22), nous avons :

$$\nabla_{\mathbf{H}} I_{p,p}(\mathbf{s}; \mathbf{x}) - \nabla_{\mathbf{H}} I_{p,q}(\mathbf{s}; \mathbf{x}) = \nabla_{\mathbf{H}} D_{KL}(p(\mathbf{x}|\mathbf{s})||q(\mathbf{x}|\mathbf{s})) - \nabla_{\mathbf{H}} D_{KL}(p(\mathbf{x})||q(\mathbf{x})) \quad (4.70)$$

$$= -\widehat{\Sigma}^{-1} \mathbf{H} (\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) - \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})) \quad (4.71)$$

$$+ (\widehat{\Sigma}^{-1} - \Sigma^{-1}) \left(\mathbb{E}_p[(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_q^T] - \mathbb{E}_p[(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_p^T] \right).$$

où (4.71) résulte de (4.61), et du fait que $D_{KL}(p(\mathbf{x}|\mathbf{s})||q(\mathbf{x}|\mathbf{s}))$ ne dépend pas de \mathbf{H} . En utilisant $\nabla_{\mathbf{H}} I_{p,p}(\mathbf{s}; \mathbf{x}) = \Sigma^{-1} \mathbf{H} \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})$ (paragraphe 1.3.3) et (4.62), on obtient :

$$\nabla_{\mathbf{H}} I_{p,q}(\mathbf{s}; \mathbf{x}) = \widehat{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) + \left(\Sigma^{-1} - \widehat{\Sigma}^{-1} \right) \mathbb{E}_p[(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_q^T]. \quad (4.72)$$

□

4.3.1.3 Inadéquation du signal d'entrée et du bruit

Finalement, nous considérons le cas d'une double inadéquation du modèle. A notre connaissance, nous sommes les premiers à étudier les relations estimation-information sous double inadéquation. Nous supposons que la densité du signal d'entrée et la densité du bruit sont mal définies, $q(\mathbf{s}) \neq p(\mathbf{s})$ et $q(\mathbf{n}) \neq p(\mathbf{n})$. L'inadéquation du bruit que nous supposons se limite au cas où seule la matrice de covariance du bruit est fautive, *i.e.* $q(\mathbf{n}) = \mathcal{N}(\mathbf{0}, \widehat{\Sigma})$. Ce cas est schématisé par la figure 4.5 :

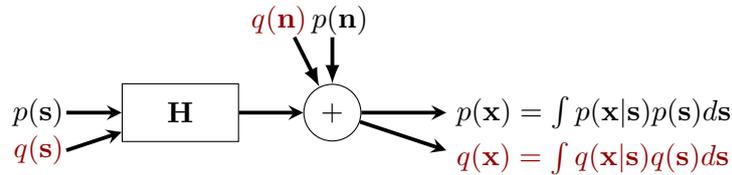


FIGURE 4.5 – Canal gaussien linéaire sous inadéquation du signal d'entrée et du bruit

Estimation. L'estimateur de \mathbf{s} dans le cas d'une double inadéquation du modèle, $q(\mathbf{s}) \neq p(\mathbf{s})$ et $q(\mathbf{n}) \neq p(\mathbf{n})$, est :

$$\hat{\mathbf{s}}_q(\mathbf{x}) = \int \frac{q(\mathbf{s})q(\mathbf{x}|\mathbf{s})}{q(\mathbf{x})} \mathbf{s} d\mathbf{s}, \quad (4.73)$$

où $q(\mathbf{x}|\mathbf{s}) = \mathcal{N}(\mathbf{H}\mathbf{s}, \widehat{\Sigma})$ et $q(\mathbf{x}) = \int q(\mathbf{x}|\mathbf{s})q(\mathbf{s})d\mathbf{s}$. L'erreur quadratique moyenne relative à cet estimateur est :

$$\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) = \mathbb{E}_p[(\mathbf{s} - \hat{\mathbf{s}}_q(\mathbf{x}))(\mathbf{s} - \hat{\mathbf{s}}_q(\mathbf{x}))^T]. \quad (4.74)$$

Information. L'information inadéquate entre l'entrée et la sortie du canal est :

$$I_{p,q}(\mathbf{s}; \mathbf{x}) \triangleq \mathbb{E}_p \left[\log \left(\frac{q(\mathbf{s}, \mathbf{x})}{q(\mathbf{s})q(\mathbf{x})} \right) \right], \quad (4.75)$$

$$= I_{p,p}(\mathbf{s}; \mathbf{x}) + D_{KL}(p(\mathbf{s})||q(\mathbf{s})) - D_{KL}(p(\mathbf{s}|\mathbf{x})||q(\mathbf{s}|\mathbf{x})), \quad (4.76)$$

$$= I_{p,p}(\mathbf{s}; \mathbf{x}) + D_{KL}(p(\mathbf{x})||q(\mathbf{x})) - D_{KL}(p(\mathbf{x}|\mathbf{s})||q(\mathbf{x}|\mathbf{s})). \quad (4.77)$$

La différence entre $I_{p,q}(\mathbf{s}; \mathbf{x})$ et $I_{p,p}(\mathbf{s}; \mathbf{x})$ est égale à une différence de divergences de Kullback-Leibler. Par conséquent, $I_{p,q}(\mathbf{s}; \mathbf{x})$ peut être supérieure ou inférieure à $I_{p,p}(\mathbf{s}; \mathbf{x})$.

Entropie relative et excédent d'erreur quadratique moyenne. Dans le cadre d'une double inadéquation du modèle la divergence de Kullback-Leibler entre les densités d'entrée et de sortie du canal gaussien, $D_{KL}(p(\mathbf{x})||q(\mathbf{x}))$, est liée à l'excédent de l'EQM, $\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) - \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})$, par le théorème suivant :

Théorème 4.5: (D_{kl} -ExcédentEQM sous inadéquation du bruit et du signal d'entrée)

Soit le canal gaussien :

$$\mathbf{x} = \mathbf{H} \mathbf{s} + \mathbf{n}, \quad (4.78)$$

où $\mathbf{s} \sim p(\mathbf{s})$ et $\mathbf{n} \sim \mathcal{N}(0, \mathbf{\Sigma})$. Sous inadéquation du bruit, $q(\mathbf{n}) = \mathcal{N}(0, \widehat{\mathbf{\Sigma}}) \neq p(\mathbf{n}) = \mathcal{N}(0, \mathbf{\Sigma})$, et du signal d'entrée, $q(\mathbf{s}) \neq p(\mathbf{s})$, le gradient de la divergence de Kullback-Leibler entre la densité exacte de \mathbf{x} , $p(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{s})p(\mathbf{s})d\mathbf{s}$, et la densité supposée, $q(\mathbf{x}) = \int q(\mathbf{x}|\mathbf{s})q(\mathbf{s})d\mathbf{s}$, est :

$$\nabla_{\mathbf{H}} D_{KL}(p(\mathbf{x})||q(\mathbf{x})) = \widehat{\mathbf{\Sigma}}^{-1} \mathbf{H} (\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) - \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})) + \mathbf{R}_D \quad (4.79)$$

où $\mathbf{R}_D = -(\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}) \left(\mathbb{E}_p[(\mathbf{x} - \mathbf{H}\mathbf{s})\widehat{\mathbf{s}}_q^T] - \mathbb{E}_p[(\mathbf{x} - \mathbf{H}\mathbf{s})\widehat{\mathbf{s}}_p^T] \right)$.

Démonstration. Voir appendice C

□

Quoique cette relation paraisse similaire à celle obtenue dans le cas d'inadéquation du bruit (4.61), elles sont différentes. En effet, la densité de \mathbf{x} dépend ici de la double inadéquation des distributions de \mathbf{s} et \mathbf{n} , alors qu'en (4.61) elle ne dépendait que de l'inadéquation de la distribution de \mathbf{n} .

Pour $\widehat{\mathbf{\Sigma}} = \mathbf{\Sigma}$, le terme résiduel est nul et nous retrouvons la relation (4.47) proposée par Verdú. Par conséquent, la relation (4.79) que nous proposons est une généralisation de celle de Verdú.

Sous double inadéquation du modèle, la proposition 4.1 et le corollaire 4.1 demeurent valides.

Information mutuelle et erreur quadratique moyenne inadéquates. Nous proposons dans ce paragraphe la généralisation de la relation I-MMSE établie par Palomar et Verdú au cas d'une double inadéquation de modèle.

Théorème 4.6: (I-MMSE sous inadéquation du bruit et du signal d'entrée)

Soit le canal gaussien :

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}, \quad (4.80)$$

où $\mathbf{s} \sim p(\mathbf{s})$ et $\mathbf{n} \sim \mathcal{N}(0, \mathbf{\Sigma})$. Sous inadéquation du bruit, $q(\mathbf{n}) = \mathcal{N}(0, \widehat{\mathbf{\Sigma}}) \neq p(\mathbf{n})$, et du signal d'entrée, $q(\mathbf{s}) \neq p(\mathbf{s})$, l'information mutuelle et l'erreur d'estimation sont liées par la relation :

$$\nabla_{\mathbf{H}} I_{p,q}(\mathbf{s}; \mathbf{x}) = \widehat{\mathbf{\Sigma}}^{-1} \mathbf{H} \mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) + \mathbf{R}_I, \quad (4.81)$$

avec, $\mathbf{R}_I = - \left(\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1} \right) \mathbb{E}_p[(\mathbf{x} - \mathbf{H}\mathbf{s}) \widehat{\mathbf{s}}_q^T]$.

L'inadéquation du signal n'est pas visible dans le relation (4.81), pourtant elle intervient dans $I_{p,q}(\mathbf{s}; \mathbf{x})$ et $\widehat{\mathbf{s}}_q$ à travers les densités supposées de \mathbf{s} , $q(\mathbf{s})$, et \mathbf{x} , $q(\mathbf{x}) = \int q(\mathbf{x}|\mathbf{s})q(\mathbf{s})d\mathbf{s}$.

4.3.2 Modèle gaussien

Maintenant, nous considérons un modèle gaussien qui permettra de :

1. Calculer les termes résiduels \mathbf{R}_D et \mathbf{R}_I dans les théorèmes 4.3, 4.4, 4.5 et 4.6.
2. Discerner la différence entre les relations information-estimation dans le cas d'inadéquation du bruit et celui de double inadéquation.

Soit le canal gaussien linéaire bruitée avec entrée gaussienne :

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}, \quad (4.82)$$

où $\mathbf{s} \in \mathbb{R}^l$ est un vecteur gaussien de dimension k , de moyenne nulle et de matrice de covariance $\mathbf{\Gamma}$, $p(\mathbf{s}) = \mathcal{N}(0, \mathbf{\Gamma})$. $\mathbf{n} \in \mathbb{R}^k$ est un bruit gaussien additif de dimension k , de moyenne nulle et de matrice de covariance $\mathbf{\Sigma}$, $p(\mathbf{n}) = \mathcal{N}(0, \mathbf{\Sigma})$. $\mathbf{H} \in \mathbb{R}^{k \times l}$ est la matrice du canal. Le vecteur de sortie, $\mathbf{x} \in \mathbb{R}^k$, est de densité $p(\mathbf{x}) = \mathcal{N}(0, \mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \mathbf{\Sigma})$.

Nous supposons que les densités inadéquates de \mathbf{s} et de \mathbf{n} sont aussi gaussiennes et que l'inadéquation résulte uniquement des matrices de covariance. C'est-à-dire que les densités inadéquates de \mathbf{s} et \mathbf{n} sont respectivement $q(\mathbf{s}) = \mathcal{N}(0, \widehat{\mathbf{\Gamma}})$ et $q(\mathbf{n}) = \mathcal{N}(0, \widehat{\mathbf{\Sigma}})$.

A la différence de la section précédente, nous considérons uniquement les deux scénarios d'inadéquation où \mathbf{R}_D et \mathbf{R}_I interviennent : l'inadéquation du bruit, *i.e* $\widehat{\mathbf{\Sigma}} \neq \mathbf{\Sigma}$, et la double inadéquation, *i.e* $\widehat{\mathbf{\Gamma}} \neq \mathbf{\Gamma}$ et $\widehat{\mathbf{\Sigma}} \neq \mathbf{\Sigma}$.

4.3.2.1 Inadéquation du bruit

Soit le modèle (4.82). Supposons la densité du bruit $q(\mathbf{n}) = \mathcal{N}(0, \widehat{\mathbf{\Sigma}})$ différente de la densité exacte $p(\mathbf{n}) = \mathcal{N}(0, \mathbf{\Sigma})$.

Estimation. L'estimateur inadéquat du \mathbf{s} est obtenu de (4.73) en appliquant les hypothèses de gaussianité, soit :

$$\widehat{\mathbf{s}}_q = \mathbf{\Gamma}\mathbf{H}^T(\mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \widehat{\mathbf{\Sigma}})^{-1}\mathbf{x}. \quad (4.83)$$

L'EQM d'estimation associée à $\hat{\mathbf{s}}_q$, est :

$$\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) = \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) + \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})\mathbf{H}^T\hat{\Sigma}^{-1}(\Sigma - \hat{\Sigma})\hat{\Sigma}^{-1}\mathbf{H}\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \quad (4.84)$$

où $\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) = \mathbb{E}_q[(\mathbf{s} - \hat{\mathbf{s}}_q)(\mathbf{s} - \hat{\mathbf{s}}_q)^T]$ est la matrice de covariance de l'erreur d'estimation $\mathbf{s} - \hat{\mathbf{s}}_q$ calculée avec les densités supposées $q(\mathbf{s}) = p(\mathbf{s})$ et $q(\mathbf{n})$. C'est-à-dire l'EQM si la densité $q(\mathbf{n})$ était exacte. Concrètement, $\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})$ est l'erreur anticipée en supposant la densité du bruit $q(\mathbf{n})$. Selon les hypothèses de gaussianité $\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})$ s'écrit :

$$\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) = \Gamma - \Gamma\mathbf{H}^T(\mathbf{H}\Gamma\mathbf{H}^T + \hat{\Sigma})^{-1}\mathbf{H}\Gamma. \quad (4.85)$$

Par conséquent, $\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x})$ est égale à l'EQM anticipée plus une erreur due à l'inadéquation du bruit.

| *Démonstration.* Les détails de calcul sont présentés à l'annexe D. □

Information. L'information mutuelle entre \mathbf{s} et \mathbf{x} sous l'hypothèse $\hat{\Gamma} \neq \Gamma$, est obtenue à partir de (4.58) en appliquant les hypothèses de gaussianité :

$$I_{p,q}(\mathbf{s}; \mathbf{x}) = I_{q,q}(\mathbf{s}; \mathbf{x}) - \frac{1}{2}Tr \{ \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})^{-1} \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x}) \} + \frac{l}{2}. \quad (4.86)$$

où $I_{q,q}(\mathbf{s}; \mathbf{x}) = \frac{1}{2} \log (|\Gamma| |\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})^{-1}|)$ est l'information mutuelle entre \mathbf{s} et \mathbf{x} si la densité $q(\mathbf{n})$ était exacte. l est la dimension du vecteur \mathbf{s} . $\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})$ est donnée par l'équation (4.85) et $\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})$ est l'EQM du modèle adéquat :

$$\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x}) = \Gamma - \Gamma\mathbf{H}^T(\mathbf{H}\Gamma\mathbf{H}^T + \Sigma)^{-1}\mathbf{H}\Gamma. \quad (4.87)$$

Entropie relative et excédent d'erreur quadratique moyenne. Le théorème 4.3 exprime la relation établie entre la divergence de Kullback-Leibler et l'erreur quadratique moyenne dans le cas d'inadéquation du bruit. Cette relation fait intervenir un terme résiduel \mathbf{R}_D par rapport à la relation établie par Verdù dans le cas d'inadéquation du signal utile. Dans le cas d'un modèle gaussien, \mathbf{R}_D peut être exprimé en fonction des EQMs.

Dans le cas d'un modèle gaussien (4.82), le théorème 4.3 devient :

Théorème 4.7: (D_{kl} -ExcédentEQM sous inadéquation du bruit - Entrée gaussienne [44])

Soit le canal gaussien :

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}, \quad (4.88)$$

où $p(\mathbf{s}) = \mathcal{N}(0, \Gamma)$ et $p(\mathbf{n}) = \mathcal{N}(0, \Sigma)$. Sous inadéquation de la matrice de covariance du bruit, $\hat{\Sigma} \neq \Sigma$, la divergence de Kullback-Leibler entre les densités exacte et supposée de la sortie du canal est liée à l'excédent de l'EQM par la relation :

$$\begin{aligned} \nabla_{\mathbf{H}} D_{KL}(p(\mathbf{x})||q(\mathbf{x})) &= \hat{\Sigma}^{-1}\mathbf{H}(\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) - \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})) \\ &+ (\hat{\Sigma}^{-1}\Sigma - \mathbf{I})(\Sigma^{-1}\mathbf{H}\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x}) - \hat{\Sigma}^{-1}\mathbf{H}\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})). \end{aligned} \quad (4.89)$$

où

$$\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x}) = \mathbf{\Gamma} - \mathbf{\Gamma}\mathbf{H}^T(\mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \mathbf{\Sigma})^{-1}\mathbf{H}\mathbf{\Gamma}, \quad (4.90)$$

$$\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) = \mathbf{\Gamma} - \mathbf{\Gamma}\mathbf{H}^T(\mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \widehat{\mathbf{\Sigma}})^{-1}\mathbf{H}\mathbf{\Gamma}, \quad (4.91)$$

$$\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) = \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) + \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})\mathbf{H}^T\widehat{\mathbf{\Sigma}}^{-1}(\mathbf{\Sigma} - \widehat{\mathbf{\Sigma}})\widehat{\mathbf{\Sigma}}^{-1}\mathbf{H}\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}). \quad (4.92)$$

En regroupant les termes en $\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})$, la relation (4.89) peut aussi s'écrire autrement comme suit :

$$\nabla_{\mathbf{H}}D_{KL}(p(\mathbf{x})||q(\mathbf{x})) = (\widehat{\mathbf{\Sigma}}^{-1}\mathbf{H}\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) - \mathbf{\Sigma}^{-1}\mathbf{H}\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})) - (\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1})\mathbf{\Sigma}\widehat{\mathbf{\Sigma}}^{-1}\mathbf{H}\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}). \quad (4.93)$$

Démonstration. Selon le théorème 4.3 :

$$\nabla_{\mathbf{H}}D_{KL}(p(\mathbf{x})||q(\mathbf{x})) = \widehat{\mathbf{\Sigma}}^{-1}\mathbf{H}(\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) - \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})) + \mathbf{R}_D. \quad (4.94)$$

où

$$\mathbf{R}_D = -(\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1})\left(\mathbb{E}_p[(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_q^T] - \mathbb{E}_p[(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_p^T]\right). \quad (4.95)$$

Selon la proposition 4.1 :

$$\mathbb{E}_p[(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_p^T] = \mathbf{H}\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x}). \quad (4.96)$$

En nous appuyant sur (4.83), on a :

$$\mathbb{E}_p[(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_q^T] = \mathbb{E}_p[(\mathbf{x} - \mathbf{H}\mathbf{s})\mathbf{x}^T](\mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \widehat{\mathbf{\Sigma}})^{-1}\mathbf{H}\mathbf{\Gamma}, \quad (4.97)$$

$$= \mathbf{\Sigma}\widehat{\mathbf{\Sigma}}^{-1}\mathbf{H}(\mathbf{H}^T\widehat{\mathbf{\Sigma}}^{-1}\mathbf{H} + \mathbf{\Gamma}^{-1})^{-1}, \quad (4.98)$$

$$= \mathbf{\Sigma}\widehat{\mathbf{\Sigma}}^{-1}\mathbf{H}\left(\mathbf{\Gamma} - \mathbf{\Gamma}\mathbf{H}^T(\mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \widehat{\mathbf{\Sigma}})^{-1}\mathbf{H}\mathbf{\Gamma}\right), \quad (4.99)$$

$$= \mathbf{\Sigma}\widehat{\mathbf{\Sigma}}^{-1}\mathbf{H}\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}). \quad (4.100)$$

où les équations (4.98) et (4.99) découlent des identités (E.4) et (E.3), respectivement. L'équation (4.100) est obtenue de l'expression (4.85).

Finalement, en remplaçant $\mathbb{E}_p[(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_q^T]$ et $\mathbb{E}_p[(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_p^T]$ dans (4.95) par leurs expressions en (4.100) et (4.96), on obtient :

$$\mathbf{R}_D = -(\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1})\left(\mathbf{\Sigma}\widehat{\mathbf{\Sigma}}^{-1}\mathbf{H}\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) - \mathbf{H}\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})\right), \quad (4.101)$$

$$= -(\widehat{\mathbf{\Sigma}}^{-1}\mathbf{\Sigma} - \mathbf{I})\left(\widehat{\mathbf{\Sigma}}^{-1}\mathbf{H}\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) - \mathbf{\Sigma}^{-1}\mathbf{H}\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})\right), \quad (4.102)$$

□

Information mutuelle et erreur quadratique moyenne inadéquates. Nous avons montré précédemment que la relation établie par Palomar et Verdù entre l'information mutuelle et l'EQM (1.46) ne s'applique pas sous inadéquation du bruit puisqu'il faut prendre en compte un terme résiduel \mathbf{R}_I (4.69) :

$$\mathbf{R}_I = - \left(\widehat{\Sigma}^{-1} - \Sigma^{-1} \right) \mathbb{E}_p[(\mathbf{x} - \mathbf{H}\mathbf{s})\hat{\mathbf{s}}_q^T]. \quad (4.103)$$

Dans le cas du modèle gaussien (4.82), en nous appuyant sur l'équation (4.100), \mathbf{R}_I s'écrit :

$$\mathbf{R}_I = - \left(\widehat{\Sigma}^{-1} - \Sigma^{-1} \right) \Sigma \widehat{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}), \quad (4.104)$$

$$= - \left(\widehat{\Sigma}^{-1} \Sigma - \mathbf{I} \right) \widehat{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}). \quad (4.105)$$

Ainsi, dans le cas d'une entrée gaussienne, $p(\mathbf{s}) = \mathcal{N}(0, \Gamma)$, le théorème 4.4 s'écrit [44] :

Théorème 4.8: (I-MMSE sous inadéquation du bruit et du signal d'entrée)

Soit le canal gaussien :

$$\mathbf{x} = \mathbf{H} \mathbf{s} + \mathbf{n}, \quad (4.106)$$

où $p(\mathbf{s}) = \mathcal{N}(0, \Gamma)$ et $p(\mathbf{n}) = \mathcal{N}(0, \Sigma)$. Sous inadéquation de la matrice de covariance du bruit, $\widehat{\Sigma} \neq \Sigma$, l'information mutuelle et l'erreur d'estimation sont liées par la relation :

$$\nabla_{\mathbf{H}} I_{p,q}(\mathbf{s}; \mathbf{x}) = \widehat{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) - \left(\widehat{\Sigma}^{-1} \Sigma - \mathbf{I} \right) \widehat{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}), \quad (4.107)$$

avec, $\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) = \Gamma - \Gamma \mathbf{H}^T (\mathbf{H} \Gamma \mathbf{H}^T + \widehat{\Sigma})^{-1} \mathbf{H} \Gamma$.

4.3.2.2 Inadéquation du signal d'entrée et du bruit

Toujours dans le cadre du modèle gaussien, nous supposons maintenant que les matrices de covariance du signal utile et du bruit sont inadéquates, $\widehat{\Gamma} \neq \Gamma$ et $\widehat{\Sigma} \neq \Sigma$.

Estimation. L'estimateur inadéquat de \mathbf{s} est obtenu de (4.73) en appliquant les hypothèses de gaussianité, soit :

$$\hat{\mathbf{s}}_q = \widehat{\Gamma} \mathbf{H}^T (\mathbf{H} \widehat{\Gamma} \mathbf{H}^T + \widehat{\Sigma})^{-1} \mathbf{x}. \quad (4.108)$$

L'EQM d'estimation associée à $\hat{\mathbf{s}}_q$, est :

$$\begin{aligned} \mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) &= \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) + \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \mathbf{H}^T \widehat{\Sigma}^{-1} (\Sigma - \widehat{\Sigma}) \widehat{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \\ &\quad + \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \widehat{\Gamma}^{-1} (\Gamma - \widehat{\Gamma}) \widehat{\Gamma}^{-1} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}). \end{aligned} \quad (4.109)$$

où $\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})$ est l'EQM si les densités $q(\mathbf{s})$ et $q(\mathbf{n})$ étaient exactes.

$$\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) = \widehat{\Gamma} - \widehat{\Gamma} \mathbf{H}^T (\mathbf{H} \widehat{\Gamma} \mathbf{H}^T + \widehat{\Sigma})^{-1} \mathbf{H} \widehat{\Gamma}. \quad (4.110)$$

Nous remarquons que, par rapport au cas précédent de la seule inadéquation du bruit, $\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})$ est fonction de $\widehat{\Gamma}$ au lieu de Γ . De plus, un troisième terme s'ajoute à l'expression de $\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x})$ et correspond à l'inadéquation du signal utile.

Démonstration. Les détails de calcul sont présentés à l'annexe D. □

Information. L'information mutuelle entre \mathbf{s} et \mathbf{x} sous l'hypothèse $\hat{\mathbf{\Gamma}} \neq \mathbf{\Gamma}$ et $\hat{\mathbf{\Sigma}} \neq \mathbf{\Sigma}$, est obtenue de (4.77) en appliquant les hypothèses de gaussianité :

$$I_{p,q}(\mathbf{s}; \mathbf{x}) = I_{q,q}(\mathbf{s}; \mathbf{x}) - \frac{1}{2} \text{Tr} \{ \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})^{-1} \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x}) \} + \frac{1}{2} \text{Tr} \{ \hat{\mathbf{\Gamma}}^{-1} \mathbf{\Gamma} \}, \quad (4.111)$$

où $I_{q,q}(\mathbf{s}; \mathbf{x}) = \frac{1}{2} \log \left(|\hat{\mathbf{\Gamma}}| |\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})^{-1}| \right)$ est l'information mutuelle entre \mathbf{s} et \mathbf{x} si les densités $q(\mathbf{s})$ et $q(\mathbf{n})$ étaient exactes. $\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})$ et $\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})$ sont données par les équations (4.110) et (4.87), respectivement. En substituant $\hat{\mathbf{\Gamma}}$ par $\mathbf{\Gamma}$, $\text{Tr} \{ \hat{\mathbf{\Gamma}}^{-1} \mathbf{\Gamma} \} = l$ et nous trouvons la même expression en (4.86), qui correspond à l'inadéquation du bruit seul.

Intérêt d'une double inadéquation. Dans le cas particulier où $\mathbf{\Gamma} \mathbf{H}^T \mathbf{\Sigma}^{-1} = \hat{\mathbf{\Gamma}} \mathbf{H}^T \hat{\mathbf{\Sigma}}^{-1}$, les deux inadéquations, du bruit et du signal utile, peuvent totalement se compenser, conduisant à une performance égale à celle du modèle adéquat. En effet, $\hat{\mathbf{s}}_q = \hat{\mathbf{s}}_p$, $\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) = \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})$ et $I_{p,q}(\mathbf{s}; \mathbf{x}) = I_{p,p}(\mathbf{s}; \mathbf{x})$. Ce résultat contre-intuitif suggère qu'une inadéquation du signal utile peut être corrigée en introduisant une inadéquation sur le bruit et vice versa.

le modèle inadéquat est équivalent au modèle adéquat si

$$\mathbf{\Gamma} \mathbf{H}^T \mathbf{\Sigma}^{-1} = \hat{\mathbf{\Gamma}} \mathbf{H}^T \hat{\mathbf{\Sigma}}^{-1}, \quad (4.112)$$

car les deux inadéquations du modèle se compensent totalement.

Dans le cas d'une seule inadéquation, *i.e.* $\hat{\mathbf{\Gamma}} \neq \mathbf{\Gamma}$ ou $\hat{\mathbf{\Sigma}} \neq \mathbf{\Sigma}$, la condition (4.112) n'est remplie que par le modèle adéquat.

Entropie relative et excédent d'erreur quadratique moyenne. Sous double inadéquation d'un modèle gaussien, l'expression de \mathbf{R}_D est formellement similaire au cas d'inadéquation du bruit. Pourtant, l'expression de $\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})$, comme nous avons vu, est différente :

$$\mathbf{R}_D = -(\hat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1}) \left(\mathbf{\Sigma} \hat{\mathbf{\Sigma}}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) - \mathbf{H} \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x}) \right), \quad (4.113)$$

$$= -(\hat{\mathbf{\Sigma}}^{-1} \mathbf{\Sigma} - \mathbf{I}) \left(\hat{\mathbf{\Sigma}}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) - \mathbf{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x}) \right). \quad (4.114)$$

Démonstration. Le même enchaînement de la démonstration du cas précédent s'applique, en substituant $\hat{\mathbf{s}}_q$ et $\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})$ par leurs expressions sous double inadéquation, *i.e.* (4.108) et (4.110) respectivement. \square

Par conséquent, le théorème 4.5 devient :

Théorème 4.9: (D_{kl} -ExcédentEQM sous double inadéquation - Entrée gaussienne [44])

Soit le canal gaussien :

$$\mathbf{x} = \mathbf{H} \mathbf{s} + \mathbf{n}, \quad (4.115)$$

où $p(\mathbf{s}) = \mathcal{N}(0, \mathbf{\Gamma})$ et $p(\mathbf{n}) = \mathcal{N}(0, \mathbf{\Sigma})$. Sous double inadéquation du modèle, $\hat{\mathbf{\Sigma}} \neq \mathbf{\Sigma}$ et $\hat{\mathbf{\Gamma}} \neq \mathbf{\Gamma}$, la divergence de Kullback-Leibler est liée à l'excédent de l'EQM par la relation :

$$\begin{aligned} \nabla_{\mathbf{H}} D_{KL}(p(\mathbf{x})||q(\mathbf{x})) &= \hat{\mathbf{\Sigma}}^{-1} \mathbf{H} (\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) - \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})) \\ &\quad + (\hat{\mathbf{\Sigma}}^{-1} \mathbf{\Sigma} - \mathbf{I}) (\mathbf{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x}) - \hat{\mathbf{\Sigma}}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})). \end{aligned} \quad (4.116)$$

où

$$\begin{aligned} \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x}) &= \mathbf{\Gamma} - \mathbf{\Gamma} \mathbf{H}^T (\mathbf{H} \mathbf{\Gamma} \mathbf{H}^T + \mathbf{\Sigma})^{-1} \mathbf{H} \mathbf{\Gamma}, \\ \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) &= \hat{\mathbf{\Gamma}} - \hat{\mathbf{\Gamma}} \mathbf{H}^T (\mathbf{H} \hat{\mathbf{\Gamma}} \mathbf{H}^T + \hat{\mathbf{\Sigma}})^{-1} \mathbf{H} \hat{\mathbf{\Gamma}}, \\ \mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) &= \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) + \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \mathbf{H}^T \hat{\mathbf{\Sigma}}^{-1} (\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}) \hat{\mathbf{\Sigma}}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \\ &\quad + \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \hat{\mathbf{\Gamma}}^{-1} (\mathbf{\Gamma} - \hat{\mathbf{\Gamma}}) \hat{\mathbf{\Gamma}}^{-1} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}). \end{aligned}$$

La relation entre la divergence de Kullback-Leibler et l'excédent de l'EQM est différente du cas précédent, même si elles sont formellement similaires. Cette différence est apparente à travers les expressions de $\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x})$ et $\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})$.

Information mutuelle et erreur quadratique moyenne inadéquates. De même la relation entre l'information mutuelle et l'EQM sous double inadéquation du modèle semble similaire au cas précédent, puisque l'inadéquation du signal utile intervient indirectement à travers les expressions de $I_{p,q}(\mathbf{s}; \mathbf{x})$, $\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x})$ et $\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})$, que nous avons détaillées précédemment grâce aux hypothèses de gaussianité.

Théorème 4.10: (I-MMSE sous inadéquation du bruit et du signal d'entrée)

Soit le canal gaussien :

$$\mathbf{x} = \mathbf{H} \mathbf{s} + \mathbf{n}, \quad (4.117)$$

où $p(\mathbf{s}) = \mathcal{N}(0, \mathbf{\Gamma})$ et $p(\mathbf{n}) = \mathcal{N}(0, \mathbf{\Sigma})$. Sous inadéquation de la matrice de covariance du bruit, $\hat{\mathbf{\Sigma}} \neq \mathbf{\Sigma}$, et inadéquation de la matrice de covariance du signal utile, $\hat{\mathbf{\Gamma}} \neq \mathbf{\Gamma}$, l'information mutuelle et l'erreur d'estimation sont liées par la relation :

$$\nabla_{\mathbf{H}} I_{p,q}(\mathbf{s}; \mathbf{x}) = \hat{\mathbf{\Sigma}}^{-1} \mathbf{H} \mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) - (\hat{\mathbf{\Sigma}}^{-1} \mathbf{\Sigma} - \mathbf{I}) \hat{\mathbf{\Sigma}}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}), \quad (4.118)$$

avec,

$$\begin{aligned} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) &= \hat{\mathbf{\Gamma}} - \hat{\mathbf{\Gamma}} \mathbf{H}^T (\mathbf{H} \hat{\mathbf{\Gamma}} \mathbf{H}^T + \hat{\mathbf{\Sigma}})^{-1} \mathbf{H} \hat{\mathbf{\Gamma}}, \\ \mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) &= \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) + \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \mathbf{H}^T \hat{\mathbf{\Sigma}}^{-1} (\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}) \hat{\mathbf{\Sigma}}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \\ &\quad + \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \hat{\mathbf{\Gamma}}^{-1} (\mathbf{\Gamma} - \hat{\mathbf{\Gamma}}) \hat{\mathbf{\Gamma}}^{-1} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}). \end{aligned}$$

4.3.3 Exemple scalaire

Dans les sections précédentes nous avons présentés deux relations établies entre les mesures de la théorie de l'information et de la théorie de l'estimation dans trois cas d'inadéquation du

modèle. Nous avons montré que sous inadéquation du bruit, ces relations ne sont plus directes car elles introduisent des termes résiduels.

Nous considérons maintenant un exemple scalaire simple qui permettra d'analyser ces relations et de comprendre la raison d'être des différences qui surviennent suite à l'inadéquation du bruit. Soit le canal gaussien scalaire représenté à la figure suivante :

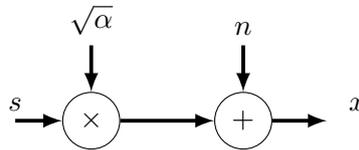


FIGURE 4.6 – Canal gaussien scalaire

Le modèle mathématique correspondant à la figure ci-dessus est :

$$x = \sqrt{\alpha} s + n. \quad (4.119)$$

où $\alpha \in \mathbb{R}^+$ et s et n sont des variables gaussiennes de moyennes nulles et de variances σ_s^2 et σ_n^2 , respectivement, *i.e* $p(s) = \mathcal{N}(0, \sigma_s^2)$ et $p(n) = \mathcal{N}(0, \sigma_n^2)$. Les relations précédemment trouvées peuvent s'appliquer en remplaçant \mathbf{H} par α , $\mathbf{\Gamma}$ par σ_s^2 et $\mathbf{\Sigma}$ par σ_n^2 .

Estimation. L'estimateur optimal au sens de l'erreur quadratique moyenne s'écrit (annexe A) :

$$\hat{\mathbf{s}}_p(x) = \frac{\sqrt{\alpha} \sigma_s^2}{\alpha \sigma_s^2 + \sigma_n^2} x. \quad (4.120)$$

Et l'erreur d'estimation relative à cet estimateur est :

$$\mathbf{M}_{p,p}(s|x) = \frac{\sigma_s^2 \sigma_n^2}{\alpha \sigma_s^2 + \sigma_n^2}. \quad (4.121)$$

Information. Dans le cas du modèle adéquat, l'information mutuelle entre s et x est (annexe B) :

$$I_{p,p}(s; x) = \frac{1}{2} \log \left(\sigma_s^2 \mathbf{M}_{p,p}(s|x)^{-1} \right) = \frac{1}{2} \log \left(1 + \frac{\alpha \sigma_s^2}{\sigma_n^2} \right). \quad (4.122)$$

4.3.3.1 Inadéquation du signal d'entrée

Nous considérons tout d'abord, le cas d'une seule inadéquation, celle de la densité de s due à l'inadéquation de la variance de s . Soit $q(s) = \mathcal{N}(0, \hat{\sigma}_s^2)$ et $q(n) = p(n)$.

Estimation. Quand nous supposons que la variance de s est $\hat{\sigma}_s^2$, $q(s) = \mathcal{N}(0, \hat{\sigma}_s^2)$, nous obtenons selon (4.41) que l'estimateur inadéquat de s est :

$$\hat{\mathbf{s}}_q(x) = \frac{\sqrt{\alpha} \hat{\sigma}_s^2}{\alpha \hat{\sigma}_s^2 + \sigma_n^2} x. \quad (4.123)$$

L'erreur associée à $\hat{\mathbf{s}}_q(x)$ est la variance de $\mathbf{s} - \hat{\mathbf{s}}_q(x)$:

$$\mathbf{M}_{p,q}(s|x) = \underbrace{\frac{\hat{\sigma}_s^2 \sigma_n^2}{\alpha \hat{\sigma}_s^2 + \sigma_n^2}}_{\mathbf{M}_{q,q}(s|x)} + (\sigma_s^2 - \hat{\sigma}_s^2) \frac{\sigma_n^4}{(\alpha \hat{\sigma}_s^2 + \sigma_n^2)^2} \geq \mathbf{M}_{p,p}(s|x). \quad (4.124)$$

Si $\hat{\sigma}_s^2 = \sigma_s^2$, le deuxième terme de la partie droite de (4.124) est nul, et $\mathbf{M}_{p,q}(s|x) = \mathbf{M}_{p,p}(s|x)$.

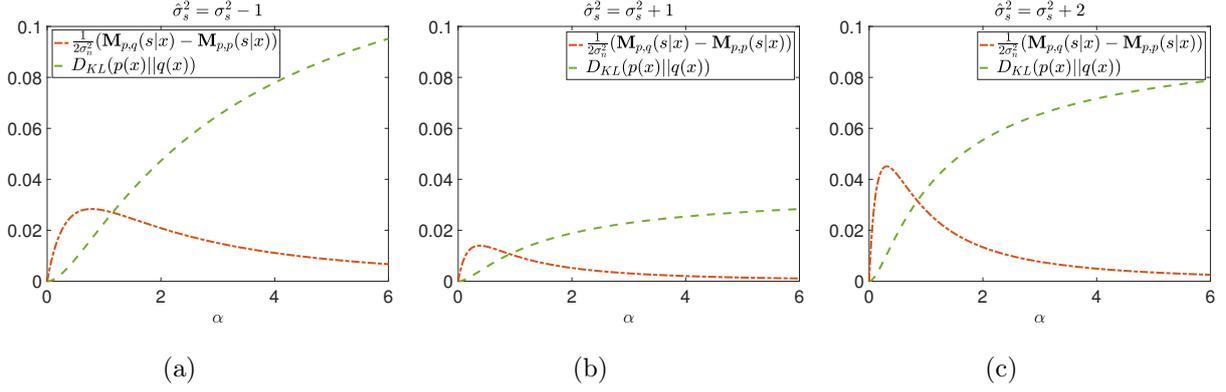


FIGURE 4.7 – $D_{KL}(p(x)||q(x))$ et $\frac{1}{2\sigma_n^2}(\mathbf{M}_{p,q}(s|x) - \mathbf{M}_{p,p}(s|x))$ en fonction de α pour différentes valeurs de $\hat{\sigma}_s^2$: $\hat{\sigma}_s^2 = \sigma_s^2 - 1$ (a), $\hat{\sigma}_s^2 = \sigma_s^2 + 1$ (b), $\hat{\sigma}_s^2 = \sigma_s^2 + 2$ (c). Avec $\sigma_s^2 = 2$ et $\sigma_n^2 = 2$.

Information. L'information mutuelle inadéquate entre s et x est :

$$I_{p,q}(s;x) = \underbrace{\frac{1}{2} \log \left(1 + \frac{\alpha \hat{\sigma}_s^2}{\sigma_n^2} \right)}_{I_{q,q}(s;x)} + \alpha \frac{\hat{\sigma}_s^2 - \sigma_s^2}{(\alpha \hat{\sigma}_s^2 + \sigma_n^2)}. \quad (4.125)$$

Nous remarquons que si $\hat{\sigma}_s^2 = \sigma_s^2$, $I_{p,q}(s;x) = I_{p,p}(s;x)$.

Entropie relative et excédent d'erreur quadratique moyenne. La divergence de Kullback-Leibler entre $p(x) = \mathcal{N}(0, \alpha \sigma_s^2 + \sigma_n^2)$ et $q(x) = \mathcal{N}(0, \alpha \hat{\sigma}_s^2 + \sigma_n^2)$ est :

$$D_{KL}(p(x)||q(x)) = \frac{1}{2} \frac{\alpha(\sigma_s^2 - \hat{\sigma}_s^2)}{\alpha \hat{\sigma}_s^2 + \sigma_n^2} - \frac{1}{2} \log \left(\frac{\alpha \sigma_s^2 + \sigma_n^2}{\alpha \hat{\sigma}_s^2 + \sigma_n^2} \right). \quad (4.126)$$

L'excédent de l'EQM est obtenu à partir de (4.124) et (4.121) :

$$\mathbf{M}_{p,q}(s|x) - \mathbf{M}_{p,p}(s|x) = (\sigma_s^2 - \hat{\sigma}_s^2)^2 \frac{\alpha \sigma_n^4}{(\alpha \hat{\sigma}_s^2 + \sigma_n^2)^2 (\alpha \sigma_s^2 + \sigma_n^2)}. \quad (4.127)$$

Nous pouvons vérifier facilement, analytiquement et graphiquement, l'exactitude de la relation du théorème 4.1 établie par Verdù, soit :

$$\frac{\partial}{\partial \alpha} D_{KL}(p(x)||q(x)) = \frac{1}{2\sigma_n^2} (\mathbf{M}_{p,q}(s|x) - \mathbf{M}_{p,p}(s|x)). \quad (4.128)$$

Sur la figure 4.7 nous représentons, en fonction de α , la divergence de Kullback-Leibler $D_{KL}(p(x)||q(x))$ et $\frac{1}{2\sigma_n^2}(\mathbf{M}_{p,q}(s|x) - \mathbf{M}_{p,p}(s|x))$, pour différentes valeurs de $\hat{\sigma}_s^2$.

Pour $\alpha = 0$, x est une mesure du bruit, $x = n$. Dans ce cas, $D_{KL}(p(x)||q(x)) = D_{KL}(p(n)||q(n)) = 0$ et l'excédent de l'EQM est nul car les deux erreurs d'estimation sont égales : $\mathbf{M}_{p,q}(s|x) = \mathbf{M}_{p,p}(s|x) = \sigma_s^2$.

Quand α croît, nous constatons que l'excédent de l'EQM devient plus important avant de tendre vers zéro pour α assez grand. C'est-à-dire que pour α très grand (rapport signal à bruit important), l'effet de l'inadéquation de la densité d'entrée importe peu à l'estimation contrairement aux faibles valeurs. Alors que, la divergence de Kullback-Leibler est une fonction croissante de α qui tend vers $D_{KL}(p(s)||q(s))$ pour $\alpha \rightarrow +\infty$.

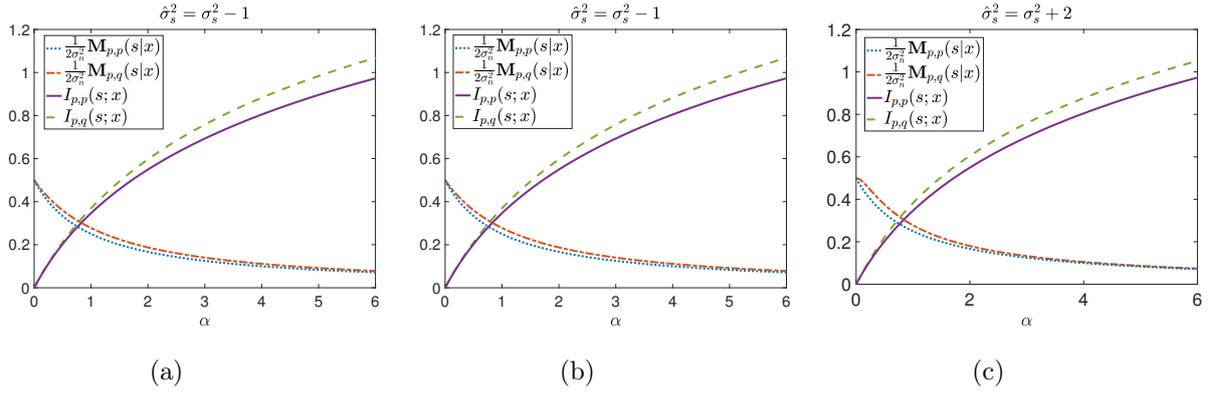


FIGURE 4.8 – $\frac{1}{2\sigma_n^2}\mathbf{M}_{p,p}(s|x)$, $\frac{1}{2\sigma_n^2}\mathbf{M}_{p,q}(s|x)$, $I_{p,p}(s;x)$ et $I_{p,q}(s;x)$ en fonction de α pour différentes valeurs de $\hat{\sigma}_s^2$: $\hat{\sigma}_s^2 = \sigma_s^2 - 1$ (a), $\hat{\sigma}_s^2 = \sigma_s^2 + 1$ (b), $\hat{\sigma}_s^2 = \sigma_s^2 + 2$ (c). Avec $\sigma_s^2 = 2$ et $\sigma_n^2 = 2$.

Si nous comparons les deux courbes, sur les trois figures, nous remarquons que l'excédent de l'EQM est maximal lorsque le taux de croissance de Kullback-Leibler est important. Réciproquement, il est nul si la divergence de Kullback-Leibler est fixe (pour $\alpha \rightarrow +\infty$, $D_{KL}(p(x)||q(x)) \rightarrow D_{KL}(p(s)||q(s))$). De fait, la variation selon α de $D_{KL}(p(x)||q(x))$ est déterminée par l'excédent de l'EQM.

Information mutuelle et erreur quadratique moyenne inadéquates. Le gradient de l'information mutuelle est lié à l'EQM par la relation (4.45), dans le cas du modèle adéquat. Cette relation peut être simplement vérifiée à partir de (4.122) et (4.121) :

$$\frac{\partial}{\partial \alpha} I_{p,p}(s;x) = \frac{1}{2\sigma_n^2} \mathbf{M}_{p,p}(s|x). \quad (4.129)$$

Cette relation est toujours valable sous inadéquation du signal d'entrée (4.2). Facile à vérifier en utilisant (4.124) et (4.125) :

$$\frac{\partial}{\partial \alpha} I_{p,q}(s;x) = \frac{1}{2\sigma_n^2} \mathbf{M}_{p,q}(s|x). \quad (4.130)$$

La figure 4.8 illustre $I_{p,q}(s;x)$ et $\frac{1}{2\sigma_n^2}\mathbf{M}_{p,q}(s|x)$ pour différentes valeurs de $\hat{\sigma}_s^2$ ainsi que $I_{p,p}(s;x)$ et $\frac{1}{2\sigma_n^2}\mathbf{M}_{p,p}(s|x)$. Nous remarquons que, pour une fausse densité de s , $I_{p,q}(s;x) \geq I_{p,p}(s;x)$ et $\mathbf{M}_{p,q}(s|x) \leq \mathbf{M}_{p,p}(s|x)$.

En appliquant l'opérateur d'intégration sur les deux parties des égalités (4.122) et (4.130), on obtient :

$$I_{p,p}(s;x) = \int_0^{+\infty} \frac{1}{2\sigma_n^2} \mathbf{M}_{p,p}(s|x) d\alpha, \quad (4.131)$$

$$I_{p,q}(s;x) = \int_0^{+\infty} \frac{1}{2\sigma_n^2} \mathbf{M}_{p,q}(s|x) d\alpha. \quad (4.132)$$

Ces relations peuvent être interprétées géométriquement ainsi : la surface délimitée par $\frac{1}{2\sigma_n^2}\mathbf{M}_{p,p}(s|x)$ (resp. $\frac{1}{2\sigma_n^2}\mathbf{M}_{p,q}(s|x)$) est égale à l'information mutuelle entre l'entrée et la sortie du canal $I_{p,p}(s;x)$ (resp. $I_{p,q}(s;x)$).

4.3.3.2 Inadéquation du bruit

Nous considérons maintenant, le cas d'inadéquation du bruit. Nous supposons que seule la variance du bruit n est inadéquate. Soit $q(s) = p(s)$ et $q(n) = \mathcal{N}(0, \hat{\sigma}_n^2)$.

Estimation. L'estimateur que nous croyons optimal de s est, selon (4.55) :

$$\hat{s}_q(x) = \frac{\sqrt{\alpha}\sigma_s^2}{\alpha\sigma_s^2 + \hat{\sigma}_n^2} x. \quad (4.133)$$

L'erreur associée à $\hat{s}_q(x)$ est :

$$\mathbf{M}_{p,q}(s|x) = \underbrace{\frac{\sigma_s^2 \hat{\sigma}_n^2}{\alpha\sigma_s^2 + \hat{\sigma}_n^2}}_{\mathbf{M}_{q,q}(s|x)} + \alpha(\sigma_n^2 - \hat{\sigma}_n^2) \frac{\sigma_s^4}{(\alpha\sigma_s^2 + \hat{\sigma}_n^2)^2} \geq \mathbf{M}_{p,p}(s|x), \quad (4.134)$$

avec égalité si et seulement si $\hat{\sigma}_n^2 = \sigma_n^2$.

Information. L'information mutuelle inadéquate entre s et x est obtenue à partir de (4.86) :

$$I_{p,q}(s; x) = \underbrace{\frac{1}{2} \log \left(1 + \frac{\alpha\sigma_s^2}{\hat{\sigma}_n^2} \right)}_{I_{q,q}(s;x)} - \alpha\sigma_s^2 \frac{\sigma_n^2 - \hat{\sigma}_n^2}{2\hat{\sigma}_n^2 (\alpha\sigma_s^2 + \sigma_n^2)} \quad (4.135)$$

Entropie relative et excédent d'erreur quadratique moyenne. La divergence de Kullback-Leibler entre $p(x) = \mathcal{N}(0, \alpha\sigma_s^2 + \sigma_n^2)$ et $q(x) = \mathcal{N}(0, \alpha\sigma_s^2 + \hat{\sigma}_n^2)$ est :

$$D_{KL}(p(x)||q(x)) = \frac{1}{2} \frac{(\sigma_n^2 - \hat{\sigma}_n^2)}{\alpha\sigma_s^2 + \hat{\sigma}_n^2} - \frac{1}{2} \log \left(\frac{\alpha\sigma_s^2 + \sigma_n^2}{\alpha\sigma_s^2 + \hat{\sigma}_n^2} \right). \quad (4.136)$$

Sous inadéquation du bruit, le gradient de la divergence de $D_{KL}(p(x)||q(x))$ n'est plus directement lié à l'excédent de l'EQM :

$$\mathbf{M}_{p,q}(s|x) - \mathbf{M}_{p,p}(s|x) = (\sigma_n^2 - \hat{\sigma}_n^2)^2 \frac{\alpha\sigma_s^4}{(\alpha\sigma_s^2 + \hat{\sigma}_n^2)^2 (\alpha\sigma_s^2 + \sigma_n^2)}, \quad (4.137)$$

puisque il faut prendre en compte le terme résiduel \mathbf{R}_D (théorème 4.7) :

$$\mathbf{R}_D = - (\sigma_n^2 - \hat{\sigma}_n^2)^2 \frac{\sigma_s^2}{2\hat{\sigma}_n^2 (\alpha\sigma_s^2 + \hat{\sigma}_n^2) (\alpha\sigma_s^2 + \sigma_n^2)}, \quad (4.138)$$

En peut vérifier facilement le résultat du théorème 4.7 :

$$\frac{\partial}{\partial \alpha} D_{KL}(p(x)||q(x)) = \frac{1}{2\hat{\sigma}_n^2} (\mathbf{M}_{p,q}(s|x) - \mathbf{M}_{p,p}(s|x)) + \mathbf{R}_D. \quad (4.139)$$

La figure 4.9 illustre, en fonction de α , différents termes de la relation (4.139) : $D_{KL}(p(x)||q(x))$, $\frac{1}{2\hat{\sigma}_n^2} (\mathbf{M}_{p,q}(s|x) - \mathbf{M}_{p,p}(s|x))$ et $\frac{1}{2\hat{\sigma}_n^2} (\mathbf{M}_{p,q}(s|x) - \mathbf{M}_{p,p}(s|x)) + \mathbf{R}_D$ pour différentes valeurs de $\hat{\sigma}_n^2$.

Nous constatons que l'excédent de l'EQM garde le même comportement que dans le cas d'inadéquation du signal (figure 4.7), car on a $\alpha\mathbf{M}_{p,p}(s|x) = \mathbf{M}_{p,p}(n|x)$ [14]. C'est-à-dire que le signal d'intérêt et le bruit ont un effet similaire sur l'erreur d'estimation. Cette relation est valable pour toutes variables aléatoires s et n même si elles sont dépendantes.

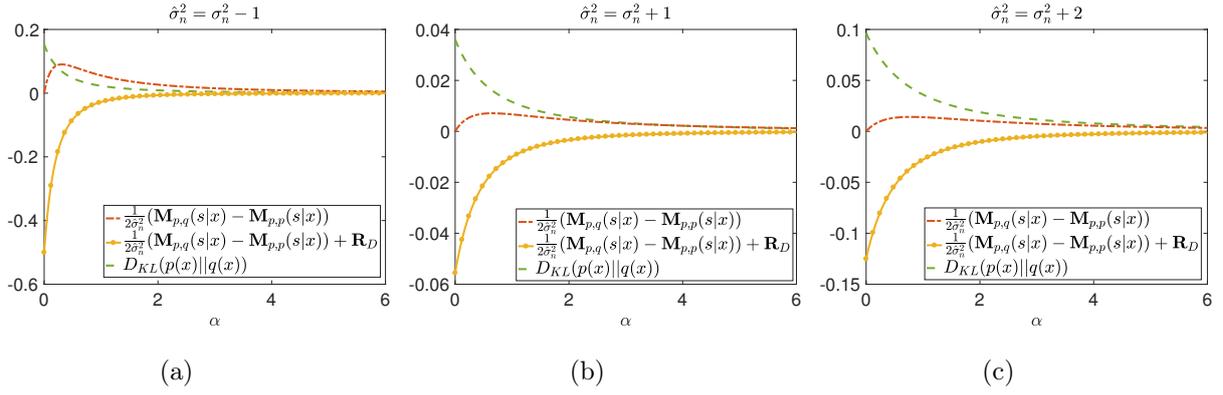


FIGURE 4.9 – $D_{KL}(p(x)||q(x))$, $\frac{1}{2\hat{\sigma}_n^2}(\mathbf{M}_{p,q}(s|x) - \mathbf{M}_{p,p}(s|x))$ et $\frac{1}{2\hat{\sigma}_n^2}(\mathbf{M}_{p,q}(s|x) - \mathbf{M}_{p,p}(s|x)) + \mathbf{R}_D$ en fonction de α pour différentes valeurs de $\hat{\sigma}_n^2$: $\hat{\sigma}_n^2 = \sigma_n^2 - 1$ (a), $\hat{\sigma}_n^2 = \sigma_n^2 + 1$ (b), $\hat{\sigma}_n^2 = \sigma_n^2 + 2$ (c). Avec $\sigma_s^2 = 2$ et $\sigma_n^2 = 2$. L'échelle de l'axe vertical est différente pour chaque figure en vue d'une meilleur visibilité.

Cependant, le comportement de la divergence de Kullback-Leibler change par rapport au cas précédent d'inadéquation du signal d'entrée (figure 4.7). Pour $\alpha = 0$, $D_{KL}(p(\mathbf{x})||q(\mathbf{x})) = D_{KL}(p(n)||q(n)) \neq 0$ car le bruit est inadéquat. Alors que pour $\alpha \rightarrow \infty$, $D_{KL}(p(x)||q(x)) \rightarrow D_{KL}(p(s)||q(s)) = 0$. Par conséquent, la divergence de Kullback-Leibler, dans le cas d'une inadéquation du bruit, est une fonction décroissante et sa dérivée est ainsi négative. Cependant l'excédent de l'EQM est toujours positif. D'où la non-validité de la relation (4.47).

Réciproquement, nous constatons que $\frac{1}{2\hat{\sigma}_n^2}(\mathbf{M}_{p,q}(s|x) - \mathbf{M}_{p,p}(s|x)) + \mathbf{R}_D$ est non-positif et traduit, ainsi, la variation de la divergence de Kullback-Leibler par rapport à α . Nous déduisons alors, que la relation (4.139) s'applique car la prise en compte de \mathbf{R}_D compense la différence de comportement de la divergence de Kullback-Leibler et des erreurs d'estimation envers l'inadéquation du bruit.

Information mutuelle et erreur quadratique moyenne inadéquates. La divergence de l'information mutuelle est également non liée directement à l'EQM sous inadéquation du bruit et un terme résiduel, \mathbf{R}_I , doit être pris en compte :

$$\mathbf{R}_I = -(\sigma_n^2 - \hat{\sigma}_n^2) \frac{\sigma_s^2}{2\hat{\sigma}_n^2(\alpha\sigma_s^2 + \hat{\sigma}_n^2)} \quad (4.140)$$

En effet, selon le théorème (4.8), on a :

$$\frac{\partial}{\partial \alpha} I_{p,q}(s;x) = \frac{1}{2\hat{\sigma}_n^2} \mathbf{M}_{p,q}(s|x) + \mathbf{R}_I. \quad (4.141)$$

La figure 4.10 représente $I_{p,q}(s;x)$, $\frac{1}{2\hat{\sigma}_n^2} \mathbf{M}_{p,q}(s|x)$ et $\frac{1}{2\hat{\sigma}_n^2} \mathbf{M}_{p,q}(s|x) + \mathbf{R}_I$ en fonction de α_2 pour différentes valeurs de $\hat{\sigma}_n^2$. Cette figure représente également $I_{p,p}(s;x)$, $\frac{1}{2\sigma_n^2} \mathbf{M}_{p,p}(s|x)$.

Nous remarquons que, sous inadéquation du bruit, $I_{p,q}(s;x) \leq I_{p,p}(s;x)$ contrairement au cas d'inadéquation du signal. Cependant, $\mathbf{M}_{p,q}(s|x)$ est toujours supérieure à $\mathbf{M}_{p,p}(s|x)$, similairement au cas d'inadéquation du signal. Ainsi, l'information mutuelle et l'EQM agissent différemment aux différents types d'inadéquations du modèle. Cette différence de comportement est compensée par le terme résiduel \mathbf{R}_I .

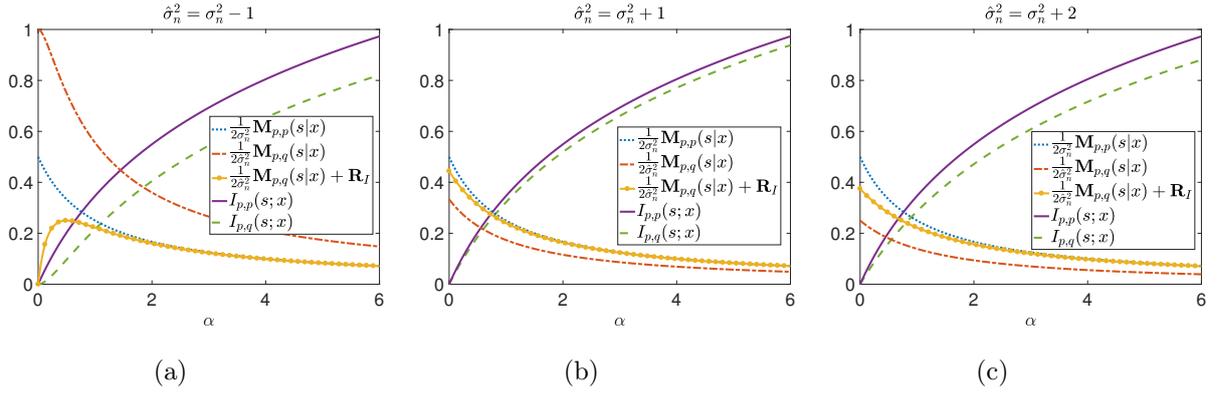


FIGURE 4.10 – $\frac{1}{2\sigma_n^2}\mathbf{M}_{p,p}(s|x)$, $\frac{1}{2\sigma_n^2}\mathbf{M}_{p,q}(s|x)$, $\frac{1}{2\sigma_n^2}\mathbf{M}_{p,q}(s|x) + \mathbf{R}_I$, $I_{p,p}(s;x)$ et $I_{p,q}(s;x)$ en fonction de α pour différentes valeurs de $\hat{\sigma}_n^2$: $\hat{\sigma}_n^2 = \sigma_n^2 - 1$ (a), $\hat{\sigma}_n^2 = \sigma_n^2 + 1$ (b), $\hat{\sigma}_n^2 = \sigma_n^2 + 2$ (c). Avec $\sigma_s^2 = 2$ et $\sigma_n^2 = 2$.

4.3.3.3 Inadéquation du signal d'entrée et du bruit

Nous considérons le cas d'une double inadéquation. Nous supposons que les variances de s et n sont inadéquates, *i.e.* $q(s) = \mathcal{N}(0, \hat{\sigma}_s^2)$ et $q(n) = \mathcal{N}(0, \hat{\sigma}_n^2)$.

Estimation. L'estimateur inadéquat de s dans le cas d'une double inadéquation est obtenu de (4.108) :

$$\hat{\mathbf{s}}_q(x) = \frac{\sqrt{\alpha}\hat{\sigma}_s^2}{\alpha\hat{\sigma}_s^2 + \hat{\sigma}_n^2} x. \quad (4.142)$$

L'erreur associée à $\hat{\mathbf{s}}_q(x)$ est :

$$\mathbf{M}_{p,q}(s|x) = \underbrace{\frac{\hat{\sigma}_s^2\hat{\sigma}_n^2}{\alpha\hat{\sigma}_s^2 + \hat{\sigma}_n^2}}_{\mathbf{M}_{q,q}(s|x)} + (\sigma_s^2 - \hat{\sigma}_s^2) \frac{\hat{\sigma}_n^4}{(\alpha\hat{\sigma}_s^2 + \hat{\sigma}_n^2)^2} + \alpha(\sigma_n^2 - \hat{\sigma}_n^2) \frac{\hat{\sigma}_s^4}{(\alpha\hat{\sigma}_s^2 + \hat{\sigma}_n^2)^2} \geq \mathbf{M}_{p,p}(s|x), \quad (4.143)$$

Information. Selon (4.111), l'information mutuelle inadéquate entre s et x est :

$$I_{p,q}(s;x) = \underbrace{\frac{1}{2} \log \left(1 + \frac{\alpha\hat{\sigma}_s^2}{\hat{\sigma}_n^2} \right)}_{I_{q,q}(s;x)} - \alpha\sigma_s^2 \frac{\sigma_s^2\hat{\sigma}_n^2 - \hat{\sigma}_s^2\sigma_n^2}{\hat{\sigma}_s^2\hat{\sigma}_n^2(\alpha\sigma_s^2 + \sigma_n^2)} \quad (4.144)$$

Dans le cas scalaire, la condition (4.112) pour que le modèle inadéquat soit équivalent au modèle adéquat devient $\frac{\sigma_s^2}{\sigma_n^2} = \frac{\hat{\sigma}_s^2}{\hat{\sigma}_n^2}$. Nous pouvons vérifier facilement que sous cette condition les inadéquations sur la variance du signal utile et la variance du bruit se compensent et que $\hat{\mathbf{s}}_q(x) = \hat{\mathbf{s}}_p(x)$ et $\mathbf{M}_{p,q}(s|x) = \mathbf{M}_{p,p}(s|x)$, $I_{p,q}(s;x) = I_{p,p}(s;x)$.

Entropie relative et excédent d'erreur quadratique moyenne. Sous double inadéquation du modèle, la divergence de Kullback-Leibler est liée à l'EQM par le théorème 4.9 :

$$\frac{\partial}{\partial \alpha} D_{KL}(p(x)||q(x)) = \frac{1}{2\hat{\sigma}_n^2} (\mathbf{M}_{p,q}(s|x) - \mathbf{M}_{p,p}(s|x)) + \mathbf{R}_D. \quad (4.145)$$

où :

$$D_{KL}(p(x)||q(x)) = \frac{1}{2} \frac{(\sigma_n^2 - \hat{\sigma}_n^2)}{\alpha\hat{\sigma}_s^2 + \hat{\sigma}_n^2} + \frac{1}{2} \frac{\alpha(\sigma_s^2 - \hat{\sigma}_s^2)}{\alpha\hat{\sigma}_s^2 + \hat{\sigma}_n^2} - \frac{1}{2} \log \left(\frac{\alpha\sigma_s^2 + \sigma_n^2}{\alpha\hat{\sigma}_s^2 + \hat{\sigma}_n^2} \right), \quad (4.146)$$

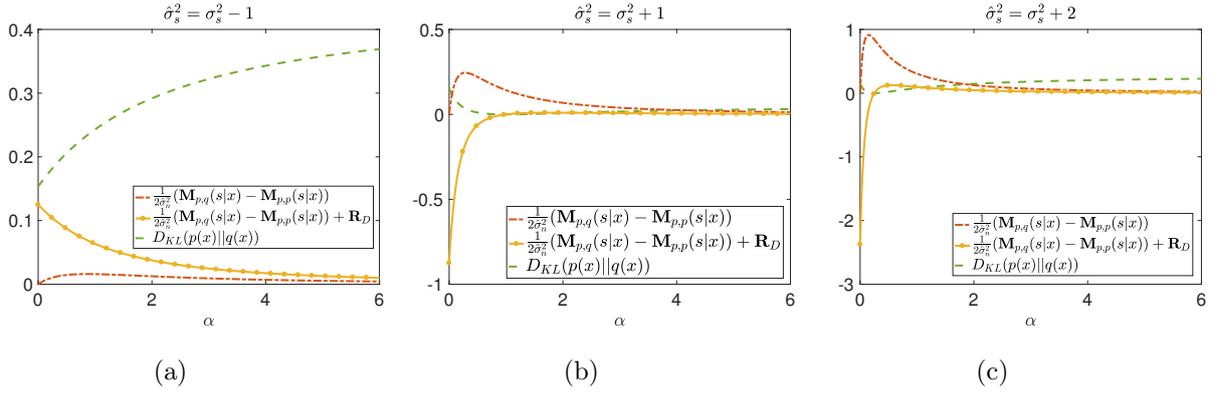


FIGURE 4.11 – $D_{KL}(p(x)||q(x))$, $\frac{1}{2\hat{\sigma}_n^2}(\mathbf{M}_{p,q}(s|x) - \mathbf{M}_{p,p}(s|x))$ et $\frac{1}{2\hat{\sigma}_n^2}(\mathbf{M}_{p,q}(s|x) - \mathbf{M}_{p,p}(s|x)) + \mathbf{R}_D$ en fonction de α pour $\hat{\sigma}_n^2 = \sigma_n^2 - 1$ et différentes valeurs de $\hat{\sigma}_s^2$: $\hat{\sigma}_s^2 = \sigma_s^2 - 1$ (a), $\hat{\sigma}_s^2 = \sigma_s^2 + 1$ (b), $\hat{\sigma}_s^2 = \sigma_s^2 + 2$ (c). Avec $\sigma_s^2 = 2$ et $\sigma_n^2 = 2$. L'échelle de l'axe vertical est différente pour chaque figure en vue d'une meilleur visibilité.

$$\mathbf{R}_D = -(\sigma_n^2 - \hat{\sigma}_n^2)^2 \frac{\sigma_s^2}{2\hat{\sigma}_n^2(\alpha\sigma_s^2 + \hat{\sigma}_n^2)(\alpha\sigma_s^2 + \sigma_n^2)}. \quad (4.147)$$

La figure 4.11 montre, en fonction de α , $D_{KL}(p(\mathbf{x})||q(\mathbf{x}))$, $\frac{1}{2\hat{\sigma}_n^2}(\mathbf{M}_{p,q}(s|x) - \mathbf{M}_{p,p}(s|x))$ et $\frac{1}{2\hat{\sigma}_n^2}(\mathbf{M}_{p,q}(s|x) - \mathbf{M}_{p,p}(s|x)) + \mathbf{R}_D$ pour $\hat{\sigma}_n^2 = \sigma_n^2 - 1$ et différentes valeurs de $\hat{\sigma}_s^2$.

La figure 4.11a correspond à la condition $\sigma_s^2\hat{\sigma}_n^2 = \hat{\sigma}_s^2\sigma_n^2$ pour laquelle le modèle adéquat est équivalent au modèle inadéquat. Nous remarquons que l'excédent de l'EQM est nul et que $D_{KL}(p(\mathbf{x})||q(\mathbf{x}))$ est constante.

Sur les figures 4.11b et 4.11c, nous remarquons que la divergence de Kullback-Leibler est une fonction décroissante pour des valeurs faibles de α où la divergence du bruit est plus influente. Notamment, pour $\alpha \in [0, -\frac{\sigma_n^2 - \hat{\sigma}_n^2}{\sigma_s^2 - \hat{\sigma}_s^2}]$. Au delà de cet intervalle, $D_{KL}(p(x)||q(x))$ devient croissante. En particulier, $D_{KL}(p(x)||q(x))$ est égale à $D_{KL}(p(n)||q(n))$ pour $\alpha = 0$, et elle tend vers $D_{KL}(p(s)||q(s))$ pour $\alpha \rightarrow \infty$. La divergence de Kullback-Leibler est nulle pour $\alpha = -\frac{\sigma_n^2 - \hat{\sigma}_n^2}{\sigma_s^2 - \hat{\sigma}_s^2}$. Attention, une divergence nulle ne signifie pas que le modèle inadéquat est équivalent au modèle adéquat. En effet, pour $\alpha = -\frac{\sigma_n^2 - \hat{\sigma}_n^2}{\sigma_s^2 - \hat{\sigma}_s^2}$, $\mathbf{M}_{p,q}(s|x) \neq \mathbf{M}_{p,q}(s|x)$. Nous remarquons que $\frac{1}{2\hat{\sigma}_n^2}(\mathbf{M}_{p,q}(s|x) - \mathbf{M}_{p,p}(s|x)) + \mathbf{R}_D$ traduit bien la variation de $D_{KL}(p(\mathbf{x})||q(\mathbf{x}))$ par rapport à α . Elle est négative quand $D_{KL}(p(\mathbf{x})||q(\mathbf{x}))$ est décroissante et positive quand $D_{KL}(p(\mathbf{x})||q(\mathbf{x}))$ est croissante.

Information mutuelle et erreur quadratique moyenne inadéquates. En s'appuyant sur le théorème 4.10, sous double inadéquation du modèle, l'information mutuelle et l'EQM sont liée par :

$$\frac{\partial}{\partial \alpha} I_{p,q}(s; x) = \frac{1}{2\hat{\sigma}_n^2} \mathbf{M}_{p,q}(s|x) + \mathbf{R}_I. \quad (4.148)$$

où

$$\mathbf{R}_I = -(\sigma_n^2 - \hat{\sigma}_n^2) \frac{\hat{\sigma}_s^2}{2\hat{\sigma}_n^2(\alpha\hat{\sigma}_s^2 + \hat{\sigma}_n^2)} \quad (4.149)$$

La figure 4.12 représente $I_{p,q}(s; x)$, $\frac{1}{2\hat{\sigma}_n^2} \mathbf{M}_{p,q}(s|x)$ et $\frac{1}{2\hat{\sigma}_n^2} \mathbf{M}_{p,q}(s|x) + \mathbf{R}_I$ en fonction de α_2 pour différentes valeurs de $\hat{\sigma}_n^2$. Cette figure représente également $I_{p,p}(s; x)$ et $\frac{1}{2\sigma_n^2} \mathbf{M}_{p,p}(s|x)$.

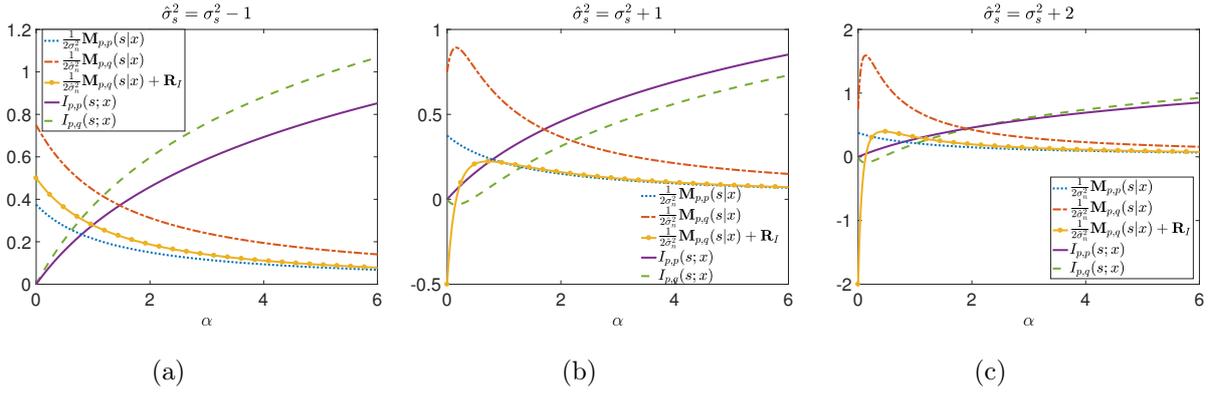


FIGURE 4.12 – $\frac{1}{2\sigma_n^2}\mathbf{M}_{p,p}(s|x)$, $\frac{1}{2\sigma_n^2}\mathbf{M}_{p,q}(s|x)$, $\frac{1}{2\sigma_n^2}\mathbf{M}_{p,q}(s|x) + \mathbf{R}_I$, $I_{p,p}(s;x)$ et $I_{p,q}(s;x)$ en fonction de α pour $\hat{\sigma}_n^2 = \sigma_n^2 - 1$ et différentes valeurs de $\hat{\sigma}_s^2$: $\hat{\sigma}_s^2 = \sigma_s^2 - 1$ (a), $\hat{\sigma}_s^2 = \sigma_s^2 + 1$ (b), $\hat{\sigma}_s^2 = \sigma_s^2 + 2$ (c). Avec $\sigma_s^2 = 1.5$ et $\sigma_n^2 = 2$.

Nous remarquons que l'information inadéquate est négative pour certaines valeur de α , et que $\frac{1}{2\hat{\sigma}_n^2}\mathbf{M}_{p,q}(s|x) + \mathbf{R}_I$ reflète la variation de $I_{p,q}(s;x)$ en fonction de α , contrairement à $\frac{1}{2\hat{\sigma}_n^2}\mathbf{M}_{p,q}(s|x)$.

4.4 Conclusion

Dans ce chapitre, nous avons introduit les mesures d'information et d'estimation adaptées aux modèles inadéquats. Ces mesures sont obtenues à partir des mesures classiques en substituant les densités exactes par les densités supposées à l'exception des espérances selon la densité des observations. Puisque les espérances selon la densité des observations peuvent être approchées par les données.

Nous avons montré que l'information mutuelle inadéquate n'est pas une mesure d'information, car elle peut être négative comme elle peut être supérieure à l'information mutuelle classique. Par conséquent, l'interprétation de cette mesure est complexe. L'entropie inadéquate, quoiqu'elle soit toujours supérieure à l'entropie classique, ne préserve pas certaines propriétés de l'entropie. Notamment, sous inadéquation du modèle, le conditionnement ne réduit pas toujours l'incertitude. Seule l'erreur quadratique moyenne d'estimation inadéquate préserve de bonnes propriétés et elle est facile à interpréter.

Dans la dernière section de ce chapitre, nous avons établi de nouvelles relations information-estimation dans le cas de modèle inadéquat. Nous avons généralisé, aux cas d'inadéquation du bruit et de double inadéquation, la relation I-MMSE par Guo *et al.* et la relation entre la divergence de Kullback-Leibler et l'excédent de l'EQM par Verdù. En effet, l'information mutuelle et la divergence de Kullback-Leibler, réagissent différemment à l'inadéquation du signal d'entrée, à l'inadéquation du bruit et à une double inadéquation. Au contraire, l'erreur d'estimation réagit similairement à tout types d'inadéquation. Par rapport aux relations existantes, les relations que nous avons établies font intervenir des termes résiduels qui compensent la dissimilarité entre les mesures d'information et les mesures d'estimation sous inadéquations de modèle.

Dans le cas particulier de modèles gaussiens, nous avons montré que deux inadéquations de

modèles peuvent totalement s'annuler, conduisant ainsi à une performance similaire, en termes d'information et de précision d'estimation, à celle obtenue en utilisant le modèle adéquat. Ce résultat important et contre-intuitif change notre vision des inadéquations de modèle qui peuvent être bénéfiques.

Modèle inadéquat : Modalités avec un couplage fort

Sommaire

5.1	Modèle	110
5.2	Mesures d'estimation	111
5.3	Liens des bruits ignorés	113
5.3.1	Modèle inadéquat <i>versus</i> modèle adéquat	114
5.3.2	Modèle inadéquat <i>versus</i> une seule modalité	119
5.4	Inadéquation quelconque des liens entre les bruits	121
5.4.1	Modèle inadéquat <i>versus</i> modèle adéquat	121
5.4.2	Modèle inadéquat <i>versus</i> une seule modalité	123
5.5	Conclusion	124

L'idée clé de l'analyse multimodale est de permettre aux modalités de s'informer mutuellement. Nous postulons que cet échange, qui a lieu grâce aux liens existants entre les données, peut révéler plus d'information sur le système physique étudié. De ce fait, la bonne modélisation de ces liens est essentielle en multimodalité.

Cependant, la caractérisation des relations inter-modalités est susceptible d'erreur en raison de la complexité des systèmes physiques et des systèmes de mesures. Dans cette deuxième partie du manuscrit, nous étudions l'interaction des modalités quand leurs liens sont mal définis. En particulier, nous nous intéressons dans ce chapitre aux liens entre les bruits.

Nous considérons un modèle simple de deux modalités qui partagent le même signal d'intérêt. Ce choix permet de mettre le point sur le lien entre les bruits. Ce même modèle, a été étudié précédemment dans le cas du modèle adéquat au chapitre 2, où nous avons mis en lumière le rôle important de la corrélation des bruits pour l'interaction des modalités.

Cependant, puisque en pratique nous supposons souvent l'indépendance des bruits, nous étudions dans ce chapitre l'impact de cette hypothèse si elle est inexacte. Nous considérons aussi le cas général d'un faux lien entre les bruits.

Ce chapitre est organisé comme suit. La section 5.1 présente le modèle considéré. Les mesures utilisées dans l'analyse de ce modèle sont développées dans la section 5.2. Les sections 5.3 et 5.4

examinent l'impact d'un faux lien entre les bruits sur l'interaction des modalités. La section 5.5 conclut ce chapitre.

5.1 Modèle

Nous considérons un modèle de deux mesures bruitées d'une variable $s \in \mathbb{R}$ que nous schématisons par la figure 5.1. Nous notons par $p(\mathbf{n})$ et $q(\mathbf{n})$, les densités jointes exacte et supposée des bruits additifs n_1 et n_2 , $\mathbf{n} = [n_1, n_2]^T$. Lorsque la loi jointe supposée des bruits correspond à la loi jointe exacte, $q(\mathbf{n}) = p(\mathbf{n})$, le modèle est dit adéquat. Autrement, *i.e* $q(\mathbf{n}) \neq p(\mathbf{n})$, la densité jointe des modalités x_1 et x_2 est inadéquate, $q(\mathbf{x}) \neq p(\mathbf{x})$ où $\mathbf{x} = [x_1, x_2]^T$, et le modèle est dit inadéquat.

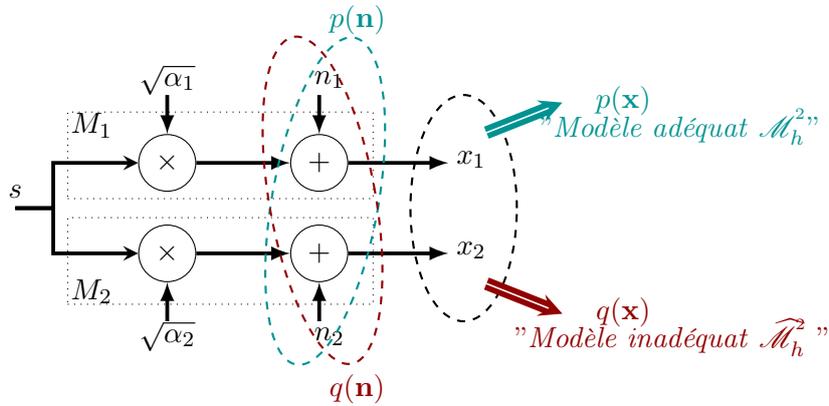


FIGURE 5.1 – Inadéquation sur le lien entre les bruits de deux modalités avec un couplage fort

Le modèle associé au schéma ci-dessus est :

$$\begin{cases} x_1 = \sqrt{\alpha_1} s + n_1 \\ x_2 = \sqrt{\alpha_2} s + n_2 \end{cases}, \quad (5.1)$$

où le signal d'intérêt, s , et les bruits, n_1 et n_2 , sont des variables gaussiennes de moyennes nulles et de variances unitaires, *i.e* $\sim \mathcal{N}(0, 1)$. Etant donné que les variances de s et des bruits sont unitaires, α_1 et α_2 sont les rapports signal à bruit des modalités x_1 et x_2 . Elles reflètent, ainsi, la qualité des modalités. Sans restreindre la généralité, nous supposons que la modalité x_2 a le plus faible rapport signal à bruit, $\alpha_2 \leq \alpha_1$.

Nous supposons que la variable s est indépendante des bruits, alors que les bruits sont corrélés. Nous notons par $\rho_n = \frac{\mathbb{E}_p[n_1 n_2]}{\mathbb{E}_p[n_1^2] \mathbb{E}_p[n_2^2]}$, le coefficient de corrélation des bruits. Ainsi, la densité jointe des bruits n_1 et n_2 est gaussienne de moyenne nulle et de matrice de covariance Σ , *i.e* $p(\mathbf{n}) = \mathcal{N}(0, \Sigma)$, où :

$$\Sigma = \begin{pmatrix} 1 & \rho_n \\ \rho_n & 1 \end{pmatrix}. \quad (5.2)$$

Les liens entre les modalités x_1 et x_2 sont représentés par la variable s et la corrélation des bruits :

Liens exacts entre les modalités :

- **Lien exact entre les signaux d'intérêt :** Un couplage fort est considéré entre les signaux d'intérêt représenté par la variable s , qui est commune aux deux modalités.
- **Lien exact entre les bruits :** Le couplage que nous considérons entre les bruits est plus faible. Il est représenté par le coefficient de corrélation, ρ_n , entre n_1 et n_2 .

Le chapitre 2 était consacré à l'étude du modèle (5.1) dans le cas adéquat. Nous y avons montré l'effet de la corrélation des bruits sur l'interaction des modalités. Dans ce chapitre, nous examinons l'effet d'un faux lien entre les bruits. Nous supposons que la densité de s est bien modélisée, $q(s) = p(s)$, contrairement à la corrélation des bruits. Nous supposons que le coefficient de corrélation des bruits est $\hat{\rho}_n = \frac{\mathbb{E}_q[n_1 n_2]}{\mathbb{E}_p[n_1^2] \mathbb{E}_p[n_2^2]}$ au lieu du coefficient exact ρ_n .

Liens supposés entre les modalités :

- **Lien supposé entre les signaux d'intérêt :** Le lien supposé correspond au lien exact, qui est représenté par la variable commune s . Ainsi la densité du signal utile est adéquate, $q(s) = p(s)$.
- **Lien supposé entre les bruits :** Le coefficient de corrélation $\hat{\rho}_n$ représente le lien supposé entre les bruits n_1 et n_2 .

Si $\hat{\rho}_n \neq \rho_n$ alors la matrice de covariance supposée des bruits est différente de la matrice de covariance exacte, *i.e* $\hat{\Sigma} \neq \Sigma$, où :

$$\Sigma = \begin{pmatrix} 1 & \rho_n \\ \rho_n & 1 \end{pmatrix}, \quad \hat{\Sigma} = \begin{pmatrix} 1 & \hat{\rho}_n \\ \hat{\rho}_n & 1 \end{pmatrix}. \quad (5.3)$$

Ainsi, $q(\mathbf{n}) = \mathcal{N}(0, \hat{\Sigma}) \neq p(\mathbf{n})$. Par conséquent, $q(\mathbf{x}|s) \neq p(\mathbf{x}|s)$ et $q(\mathbf{x}) \neq p(\mathbf{x})$. Ainsi, le modèle est inadéquat et est noté \mathcal{M}_h^2 . Réciproquement, si $\hat{\rho}_n = \rho_n$ le modèle est adéquat et est noté \mathcal{M}_h^2 .

Par le biais de l'erreur quadratique moyenne d'estimation, que nous présentons dans la section suivante, nous étudions, selon les valeurs des rapports signal à bruit, les questions suivantes :

1. Comment les faux liens entre les bruits influencent-ils l'interaction des modalités ?
2. Sous inadéquation du modèle, l'analyse multimodale assure-t-elle toujours une meilleure performance par rapport à une seule modalité ?

5.2 Mesures d'estimation

Dans le chapitre 2, nous avons examiné le modèle \mathcal{M}_h^2 au moyen de l'erreur quadratique moyenne et de l'information mutuelle. Ces deux grandeurs ont permis une interprétation selon deux points de vue des différentes situations de l'analyse multimodale.

Cependant, sous inadéquation du modèle, l'information mutuelle perd certaines de ses propriétés et son interprétation devient plus complexe (paragraphe 4.1.2). Par conséquent, dans ce

chapitre nous choisissons d'utiliser uniquement l'erreur d'estimation pour évaluer la performance de la multimodalité sous inadéquation du bruit.

Nous présentons ici les mesures d'estimation au moyen desquelles nous analysons le modèle (5.1). Dans le cas adéquat, \mathcal{M}_h^2 , nous utilisons les mesures classiques d'estimation (chapitre 1). Alors que dans le cas du modèle inadéquat, $\widehat{\mathcal{M}}_h^2$, nous utilisons les mesures introduites dans le chapitre 4 qui prennent en compte l'effet des fausses hypothèses.

Nous nous intéressons au problème d'estimation de s à partir des modalités x_1 et x_2 :

Modèle adéquat. (chapitre 2) En fonction d'une seule modalité x_i , $i \in \{1, 2\}$, l'estimateur optimal au sens de l'erreur quadratique moyenne de s est la moyenne de la loi a posteriori, $p(s|x_i)$:

$$\hat{s}_p(x_i) \triangleq \int s p(s|x_i) ds = \frac{\sqrt{\alpha_i}}{1 + \alpha_i} x_i. \quad (5.4)$$

L'erreur quadratique moyenne relative à cet estimateur est la variance de l'erreur $s - \hat{s}_p(x_i)$:

$$\mathbf{M}_{p,p}(s|x_i) \triangleq \mathbb{E}_p [(s - \hat{s}_p(x_i))^2] = \frac{1}{1 + \alpha_i}. \quad (5.5)$$

L'estimateur optimal en fonction des deux modalités, x_1 et x_2 , est la moyenne de la loi a posteriori, $p(s|x_1, x_2)$:

$$\hat{s}_p(\mathbf{x}) \triangleq \int s p(s|x_1, x_2) ds = \frac{k_1 x_1 + k_2 x_2}{(1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1 \alpha_2} + \rho_n)^2}. \quad (5.6)$$

avec,

$$k_1 = \sqrt{\alpha_1} - \sqrt{\alpha_2} \rho_n, \quad (5.7)$$

$$k_2 = \sqrt{\alpha_2} - \sqrt{\alpha_1} \rho_n. \quad (5.8)$$

L'erreur quadratique relative à $\hat{s}_p(\mathbf{x})$ est donnée par la variance de l'erreur $s - \hat{s}_p(\mathbf{x})$:

$$\mathbf{M}_{p,p}(s|x_1, x_2) \triangleq \mathbb{E}_p [(s - \hat{s}_p(\mathbf{x}))^2] = \frac{1 - \rho_n^2}{(1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1 \alpha_2} + \rho_n)^2}. \quad (5.9)$$

Modèle inadéquat. Supposons que le coefficient de corrélation entre les bruits est $\hat{\rho}_n$, au lieu de ρ_n . Cette hypothèse influence uniquement les mesures multimodales d'estimation, car les densités marginales de n_1 et de n_2 supposées correspondent aux densités exactes, $q(n_i) = p(n_i)$ ($i \in \{1, 2\}$). Ainsi, $\hat{s}_q(x_i) = \hat{s}_p(x_i)$ et $\mathbf{M}_{p,q}(s|x_i) = \mathbf{M}_{p,p}(s|x_i)$.

Cependant, la densité jointe des bruits est inadéquate, $q(\mathbf{n}) \neq p(\mathbf{n})$. Ainsi, l'estimateur de s en fonction de x_1 et x_2 , que nous croyons optimal, est la moyenne de la loi a posteriori supposée, $q(s|x_1, x_2)$:

$$\hat{s}_q(\mathbf{x}) \triangleq \int s q(s|x_1, x_2) ds, \quad (5.10)$$

$$= \int s \frac{q(\mathbf{x}|s)p(s)}{q(\mathbf{x})} ds, \quad (5.11)$$

$$= \frac{\hat{k}_1 x_1 + \hat{k}_2 x_2}{(1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1 \alpha_2} + \hat{\rho}_n)^2}. \quad (5.12)$$

avec :

$$\hat{k}_1 = \sqrt{\alpha_1} - \sqrt{\alpha_2} \hat{\rho}_n, \quad (5.13)$$

$$\hat{k}_2 = \sqrt{\alpha_2} - \sqrt{\alpha_1} \hat{\rho}_n. \quad (5.14)$$

L'erreur d'estimation associée à $\hat{s}_q(\mathbf{x})$ est donnée par la variance, calculée sur $p(\cdot)$, de l'erreur $s - \hat{s}_q(\mathbf{x})$:

$$\mathbf{M}_{p,q}(s|x_1, x_2) \triangleq \mathbb{E}_p [(s - \hat{s}_q(\mathbf{x}))^2], \quad (5.15)$$

$$= \frac{1 - \hat{\rho}_n^2}{\underbrace{(1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1 \alpha_2} + \hat{\rho}_n)^2}_{\mathbf{M}_{q,q}(s|x_1, x_2)}} \quad (5.16)$$

$$+ 2(\rho_n - \hat{\rho}_n) \frac{(\sqrt{\alpha_2} - \sqrt{\alpha_1} \hat{\rho}_n)(\sqrt{\alpha_1} - \sqrt{\alpha_2} \hat{\rho}_n)}{((1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1 \alpha_2} + \hat{\rho}_n)^2)^2}.$$

$$= a\rho_n + b. \quad (5.17)$$

L'EQM inadéquate $\mathbf{M}_{p,q}(s|x_1, x_2)$ est une fonction affine de ρ_n , dont la pente est a :

$$a = 2 \frac{(\sqrt{\alpha_2} - \sqrt{\alpha_1} \hat{\rho}_n)(\sqrt{\alpha_1} - \sqrt{\alpha_2} \hat{\rho}_n)}{((1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1 \alpha_2} + \hat{\rho}_n)^2)^2} \quad (5.18)$$

Le signe de la pente est le signe du produit $(\sqrt{\alpha_1} - \sqrt{\alpha_2} \hat{\rho}_n)(\sqrt{\alpha_2} - \sqrt{\alpha_1} \hat{\rho}_n)$. Selon l'hypothèse $\alpha_2 < \alpha_1$ et puisque $|\rho_n| < 1$, $(\sqrt{\alpha_1} - \sqrt{\alpha_2} \hat{\rho}_n) \geq 0$ ainsi $\text{sgn}(a) = \text{sgn}(\sqrt{\alpha_2} - \sqrt{\alpha_1} \hat{\rho}_n)$.

Démonstration. Ces expressions sont obtenues des cas généraux du chapitre 4 (paragraphe 4.3.2.1). Les détails des calculs sont fournis à la section D.3 de l'annexe D. \square

5.3 Liens des bruits ignorés

En pratique, nous postulons souvent l'indépendance des bruits des données dans le but de simplifier les calculs. Dans ce paragraphe nous étudions l'effet de cette hypothèse, quand elle est fautive, sur la performance de l'estimation multimodale.

Nous supposons que les bruits sont indépendants alors qu'ils sont dépendants. Nous disons dans ce cas que les liens entre les bruits sont ignorés. Selon l'hypothèse de gaussianité, l'indépendance des bruits correspond à $\hat{\rho}_n = 0$. C'est-à-dire que nous supposons que x_1 et x_2 sont indépendants conditionnellement à s , *i.e* $q(\mathbf{x}|s) = q(x_1|s)q(x_2|s)$. L'estimateur inadéquat relatif à ce cas est obtenu à partir de (5.12) en mettant $\hat{\rho}_n$ à zéro :

$$\hat{s}_q(\mathbf{x}) \Big|_{\hat{\rho}_n=0} = \frac{\sqrt{\alpha_1} x_1 + \sqrt{\alpha_2} x_2}{1 + \alpha_1 + \alpha_2}. \quad (5.19)$$

L'équation (5.19) ne dépend pas de ρ_n . C'est-à-dire que nous utilisons un estimateur qui ne prend pas en compte la corrélation des bruits. En agissant ainsi, nous nous attendons à obtenir l'EQM d'estimation :

$$\mathbf{M}_{q,q}(s|x_1, x_2) \triangleq \mathbb{E}_q [(s - \hat{s}_q(\mathbf{x}))^2] = \frac{1}{1 + \alpha_i + \alpha_j}, \quad (5.20)$$

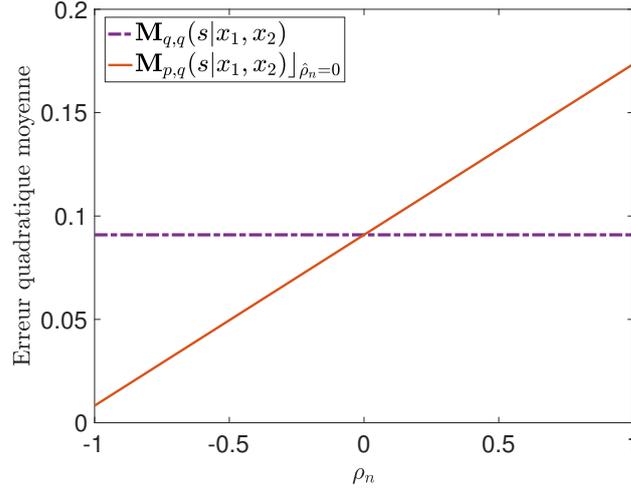


FIGURE 5.2 – Les erreurs d'estimation $\mathbf{M}_{q,q}(s|x_1, x_2)$ et $\mathbf{M}_{p,q}(s|x_1, x_2)$ en fonction de ρ_n , où $\alpha_1 = 5$ et $\alpha_2 = 2$.

où $\mathbf{M}_{q,q}(s|x_1, x_2)$ est l'erreur d'estimation multimodale si le modèle supposé est exact, c'est-à-dire $\hat{\rho}_n = \rho_n = 0$. Cependant, si les bruits n_1 et n_2 sont corrélés, $\rho_n \neq 0$, l'estimateur $\hat{s}_p(\mathbf{x})|_{\hat{\rho}_n=0}$ engendre une EQM d'estimation différente :

$$\mathbf{M}_{p,q}(s|x_1, x_2)|_{\hat{\rho}_n=0} = \underbrace{\frac{1}{1 + \alpha_1 + \alpha_2}}_{\mathbf{M}_{q,q}(s|x_1, x_2)} + \rho_n \frac{2\sqrt{\alpha_1\alpha_2}}{(1 + \alpha_1 + \alpha_2)^2}. \quad (5.21)$$

L'équation (5.21) est obtenue à partir de (5.16) en mettant $\hat{\rho}_n$ à zéro. Le premier terme de la partie droite de (5.21) correspond à $\mathbf{M}_{q,q}(s|x_1, x_2)$. La figure 5.2 représente $\mathbf{M}_{p,q}(s|x_1, x_2)$ (---) et $\mathbf{M}_{q,q}(s|x_1, x_2)$ (—), en fonction du coefficient de corrélation exacte ρ_n .

Pour $\rho_n < 0$, le signe du deuxième terme de la partie droite de (5.21) est négatif. Ainsi, $\mathbf{M}_{p,q}(s|x_1, x_2) < \mathbf{M}_{q,q}(s|x_1, x_2)$. C'est-à-dire que l'erreur obtenue, $\mathbf{M}_{p,q}(s|x_1, x_2)$, est plus faible que l'erreur attendue $\mathbf{M}_{q,q}(s|x_1, x_2)$. Réciproquement, pour $\rho_n > 0$, le signe du deuxième terme de la partie droite de (5.21) est positif et $\mathbf{M}_{p,q}(s|x_1, x_2) > \mathbf{M}_{q,q}(s|x_1, x_2)$.

La perte en performance multimodale due à l'ignorance des corrélations entre les bruits n'est pas quantifiée en comparant $\mathbf{M}_{p,q}(s|x_1, x_2)$ à $\mathbf{M}_{q,q}(s|x_1, x_2)$. Puisque, $\mathbf{M}_{q,q}(s|x_1, x_2)$ quantifie la performance anticipée et non la performance réelle du modèle adéquat.

5.3.1 Modèle inadéquat *versus* modèle adéquat

Dans ce paragraphe nous étudions l'effet d'ignorer la corrélation entre les bruits sur la précision d'estimation multimodale. Nous comparons le modèle adéquat, \mathcal{M}_h^2 , et le modèle inadéquat $\widehat{\mathcal{M}}_h^2$. Les modalités du modèle adéquat \mathcal{M}_h^2 engendrent l'erreur d'estimation $\mathbf{M}_{p,p}(s|x_1, x_2)$ (5.9), alors que les modalités du modèle inadéquat $\widehat{\mathcal{M}}_h^2$ engendrent l'erreur d'estimation $\mathbf{M}_{p,q}(s|x_1, x_2)|_{\hat{\rho}_n=0}$ (5.21).

Nous comparons les erreurs d'estimation multimodales $\mathbf{M}_{p,p}(s|x_1, x_2)$ et $\mathbf{M}_{p,q}(s|x_1, x_2)|_{\hat{\rho}_n=0}$ selon trois scénarios des qualités des modalités x_1 et x_2 : deux modalités de même qualité ($\alpha_2 = \alpha_1$), deux modalités de qualités différentes ($\alpha_2 \leq \alpha_1$) et finalement trois modalités de

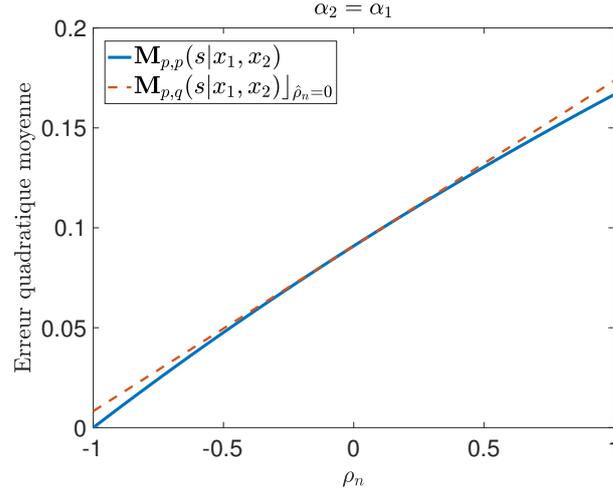


FIGURE 5.3 – Les erreurs d’estimation multimodales $\mathbf{M}_{p,p}(s|x_1, x_2)$ et $\mathbf{M}_{p,q}(s|x_1, x_2)|_{\hat{\rho}_n=0}$ en fonction de ρ_n où $\alpha_2 = \alpha_1 = 5$.

qualités très disparates ($\alpha_2 \ll \alpha_1$).

5.3.1.1 Modalités de même qualité ($\alpha_2 = \alpha_1$)

Nous commençons cette étude par le cas de deux modalités de qualités comparables $\alpha_2 = \alpha_1$. Les deux modalités x_1 et x_2 sont équivalentes puisqu’elles sont aussi précises pour l’estimation de s , $\mathbf{M}_{p,p}(s|x_1) = \mathbf{M}_{p,p}(s|x_2)$. Par conséquent, elles contribuent également à l’estimation multimodale de s . Effectivement, les facteurs de x_1 et x_2 dans les estimateurs adéquat, $\hat{s}_q(\mathbf{x})$, et inadéquat, $\hat{s}_p(\mathbf{x})$, sont égaux, $k_1 = k_2$ et $\hat{k}_1 = \hat{k}_2$:

$$\hat{s}_p(\mathbf{x})|_{\alpha_2=\alpha_1} = \frac{\sqrt{\alpha_1}}{(1 + 2\alpha_1 + \rho_n)} (x_1 + x_2), \quad (5.22)$$

et

$$\hat{s}_q(\mathbf{x})|_{\hat{\rho}_n=0, \alpha_2=\alpha_1} = \frac{\sqrt{\alpha_1}}{1 + 2\alpha_1} (x_1 + x_2). \quad (5.23)$$

La figure 5.3 représente, en fonction de ρ_n , $\mathbf{M}_{p,p}(s|x_1, x_2)$ (—) et $\mathbf{M}_{p,q}(s|x_1, x_2)|_{\hat{\rho}_n=0}$ (---), où :

$$\mathbf{M}_{p,p}(s|x_1, x_2)|_{\alpha_2=\alpha_1} = \frac{1 + \rho_n}{(1 + 2\alpha_1 + \rho_n)}, \quad (5.24)$$

et

$$\mathbf{M}_{p,q}(s|x_1, x_2)|_{\hat{\rho}_n=0, \alpha_2=\alpha_1} = \frac{1}{1 + 2\alpha_1} + \rho_n \frac{2\alpha_1}{(1 + 2\alpha_1)^2}. \quad (5.25)$$

L’EQM adéquate $\mathbf{M}_{p,p}(s|x_1, x_2)$ est une fonction concave croissante de ρ_n et $\mathbf{M}_{p,q}(s|x_1, x_2)|_{\hat{\rho}_n=0}$ est une fonction affine croissante de ρ_n .

Modèle adéquat. L’estimateur adéquat $\hat{s}_p(\mathbf{x})|_{\alpha_2=\alpha_1}$ est une somme pondérée des modalités x_1 et x_2 . Lorsque $\rho_n < 0$, cette somme permet de compenser le bruit. C’est-à-dire que la corrélation

négative favorise l'interaction des modalités. En particulier, pour $\rho_n = -1$, le bruit est totalement supprimé et l'erreur d'estimation est nulle, $\mathbf{M}_{p,p}(s|x_1, x_2) = 0$.

Réciproquement, si $\rho_n > 0$ le bruit est mal compensé car les modalités deviennent partiellement redondantes. Notamment, pour $\rho_n = 1$, les deux modalités sont totalement redondantes et l'erreur $\mathbf{M}_{p,p}(s|x_1, x_2)$ est maximale. Ainsi, les deux modalités sont équivalentes à une seule modalités, $\hat{s}_p(\mathbf{x}) = \hat{s}_p(x_1) = \hat{s}_p(x_2)$.

Modèle inadéquat. ($\hat{\rho}_n = 0$) Similairement à l'estimateur adéquat, l'estimateur inadéquat est une somme pondérée des modalités. Par conséquent, le bruit est compensé pour les corrélations négatives et est mal compensé pour les corrélations positives. C'est-à-dire que lorsque nous ignorons la corrélation des bruits, nous utilisons encore ce lien mais d'une manière sous-optimale car le modèle inadéquat engendre toujours plus d'erreurs que le modèle adéquat.

5.3.1.2 Modalités de différentes qualités ($\alpha_2 < \alpha_1$)

Nous considérons, maintenant, le cas de deux modalités de qualités différentes. Selon l'hypothèse $\alpha_2 < \alpha_1$, la modalité x_1 est plus précise pour estimer s que x_2 , $\mathbf{M}_{p,p}(s|x_1) \leq \mathbf{M}_{p,p}(s|x_2)$. Ainsi, pour l'estimation multimodale de s , la modalité x_1 communique plus d'information sur s que x_2 . Pour cette raison, les facteurs de x_1 dans les estimateurs adéquat et inadéquat sont plus grands que les facteurs de x_2 , $k_1 \geq k_2$ et $\hat{k}_1 \geq \hat{k}_2$:

$$\hat{s}_p(\mathbf{x}) = \frac{(\sqrt{\alpha_1} - \sqrt{\alpha_2}\rho_n)x_1 + (\sqrt{\alpha_2} - \sqrt{\alpha_1}\rho_n)x_2}{(1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1\alpha_2} + \rho_n)^2}, \quad (5.26)$$

et

$$\hat{s}_q(\mathbf{x}) \Big|_{\hat{\rho}_n=0} = \frac{\sqrt{\alpha_1}x_1 + \sqrt{\alpha_2}x_2}{1 + \alpha_1 + \alpha_2}. \quad (5.27)$$

Nous montrons sur la figure 5.4, les erreurs d'estimation multimodales adéquate, $\mathbf{M}_{p,p}(s|x_1, x_2)$ (—), et inadéquate, $\mathbf{M}_{p,q}(s|x_1, x_2) \Big|_{\hat{\rho}_n=0}$ (- - -), en fonction de la corrélation des bruits ρ_n .

$$\mathbf{M}_{p,p}(s|x_1, x_2) \triangleq \mathbb{E}_p [(s - \hat{s}_p(\mathbf{x}))^2] = \frac{1 - \rho_n^2}{(1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1\alpha_2} + \rho_n)^2}, \quad (5.28)$$

$$\mathbf{M}_{p,q}(s|x_1, x_2) = \underbrace{\frac{1}{1 + \alpha_1 + \alpha_2}}_{\mathbf{M}_{q,q}(s|x_1, x_2)} + 2\rho_n \frac{\sqrt{\alpha_1\alpha_2}}{(1 + \alpha_1 + \alpha_2)^2}. \quad (5.29)$$

Nous constatons que $\mathbf{M}_{p,q}(s|x_1, x_2)$ est toujours une fonction croissante de ρ_n , comme c'était le cas dans le paragraphe précédent où $\alpha_2 = \alpha_1$. Cependant, $\mathbf{M}_{p,p}(s|x_1, x_2)$ n'est plus une fonction croissante de ρ_n car elle décroît pour de fortes corrélations du bruit.

Modèle adéquat. Pour $\rho_n < 0$, les facteurs de x_1 et x_2 dans $\hat{s}_p(\mathbf{x})$ sont positifs. Ainsi, le bruit est compensé et est totalement supprimé pour $\rho_n = -1$.

Pour $\rho_n \in \left[0, \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}\right]$, le bruit est mal compensé car les modalités sont partiellement redondantes. Dans ce cas la réduction du bruit peut nuire à la qualité de s . La modalité x_2 devient totalement

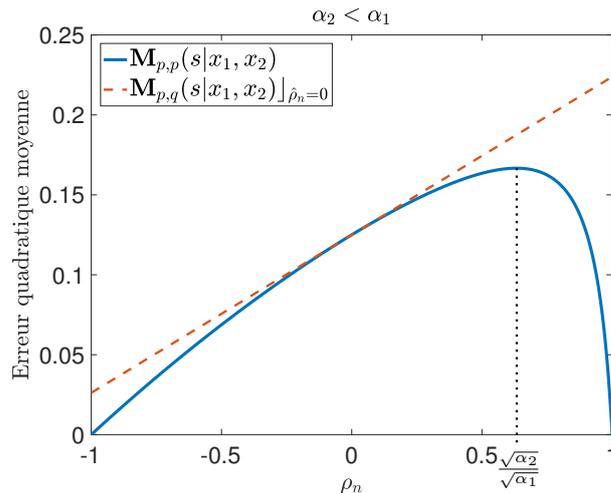


FIGURE 5.4 – Les erreurs d’estimation multimodales $\mathbf{M}_{p,p}(s|x_1, x_2)$ et $\mathbf{M}_{p,q}(s|x_1, x_2)|_{\hat{\rho}_n=0}$ en fonction de ρ_n où $\alpha_2 = 2$ et $\alpha_1 = 5$.

redondante à x_1 si et seulement si $\rho_n = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$. L’erreur d’estimation atteint dans ce cas sa valeur maximale $\mathbf{M}_{p,p}(s|x_1, x_2) = \mathbf{M}_{p,p}(s|x_1)$.

En revanche, pour $\rho_n \in \left[\frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}, 1\right]$, le facteur de x_2 devient négatif, $k_2 < 0$ et l’information sur le bruit que x_2 communique est exploitée pour réduire le bruit sans dégrader le signal s . Ainsi, contrairement au cas $\alpha_2 = \alpha_1$, l’erreur d’estimation décroît sur l’intervalle $\rho_n \in \left[\frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}, 1\right]$, et est nulle pour $\rho_n = 1$ car le bruit est complètement supprimé.

Modèle inadéquat. Quelle que soit la valeur de la corrélation des bruits, l’estimateur inadéquat est toujours une somme pondérée de x_1 et x_2 . Ainsi, $\mathbf{M}_{p,q}(s|x_1, x_2)|_{\hat{\rho}_n=0}$ est toujours une droite croissante, tangente à $\mathbf{M}_{p,p}(s|x_1, x_2)$ en $\rho_n = 0$. C’est-à-dire que le bruit est bien compensé seulement pour les valeurs négatives de ρ_n . Par conséquent, l’information sur le bruit qu’apporte la modalité de faible rapport signal à bruit, x_2 , est non exploitée par le modèle inadéquat.

5.3.1.3 Modalités de qualités très disparates ($\alpha_2 \ll \alpha_1$)

Nous considérons, finalement, le cas de deux modalités ayant des qualités très disparates $\alpha_2 \ll \alpha_1$. Si de plus $\alpha_2 \rightarrow 0$, alors x_2 est une mesure du bruit, $x_2 \rightarrow n_2$ et l’estimateur optimal est :

$$\hat{s}_p(\mathbf{x})|_{\alpha_2 \rightarrow 0} \rightarrow \frac{\sqrt{\alpha_1} x_1 - \sqrt{\alpha_1} \rho_n x_2}{1 + \alpha_1 - \rho_n^2}. \quad (5.30)$$

En revanche, lorsque nous ignorons la corrélation des bruits nous ignorons aussi la mesure du bruit, x_2 . Ainsi, l’estimateur inadéquat est fonction d’une seule modalité x_1 :

$$\hat{s}_q(\mathbf{x})|_{\hat{\rho}_n=0, \alpha_2 \rightarrow 0} \rightarrow \frac{\sqrt{\alpha_1}}{1 + \alpha_1} x_1 = \hat{s}_q(x_1). \quad (5.31)$$

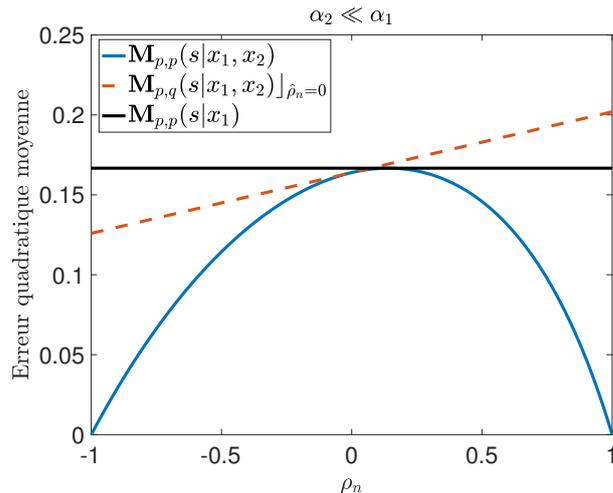


FIGURE 5.5 – Les erreurs d’estimation multimodales $\mathbf{M}_{p,p}(s|x_1, x_2)$ et $\mathbf{M}_{p,q}(s|x_1, x_2)|_{\hat{\rho}_n=0}$ en fonction de ρ_n où $\alpha_2 = 0.1$ et $\alpha_1 = 5$.

La figure 5.5, illustre les erreurs d’estimation multimodales $\mathbf{M}_{p,p}(s|x_1, x_2)$ (—), et $\mathbf{M}_{p,q}(s|x_1, x_2)|_{\hat{\rho}_n=0}$ (---), en fonction de la corrélation des bruits ρ_n .

Modèle adéquat. La modalité x_2 est une mauvaise mesure de s par rapport à x_1 . Réciproquement, elle est une mesure très précise du bruit. Par conséquent, x_2 introduit moins de redondance sur s par rapport à x_1 et permet de compenser le bruit pour presque toutes les valeurs de ρ_n .

Modèle inadéquat. La modalité x_2 apporte principalement de l’information sur le bruit que l’estimateur inadéquat n’exploite pas. Ainsi l’estimateur $\hat{s}_q(\mathbf{x})|_{\hat{\rho}_n=0}$ réduit mal le bruit et tend vers l’estimateur en fonction d’une seule modalité, $\hat{s}_q(x_1)$.

- $\alpha_2 = \alpha_1$ versus $\alpha_2 < \alpha_1$ versus $\alpha_2 \ll \alpha_1$

La différence entre $\mathbf{M}_{p,q}(s|x_1, x_2)|_{\hat{\rho}_n=0}$ et $\mathbf{M}_{p,p}(s|x_1, x_2)$, connue en littérature par excédent de l’EQM, quantifie l’erreur additionnelle due à l’ignorance de la corrélation des bruits. A partir de (5.21) et de (2.5) nous obtenons que :

$$\mathbf{M}_{p,q}(s|x_1, x_2)|_{\hat{\rho}_n=0} - \mathbf{M}_{p,p}(s|x_1, x_2) = \rho_n^2 \frac{(\alpha_1 - \alpha_2)^2 + (\sqrt{\alpha_1} - \sqrt{\alpha_2})^2 + 2\sqrt{\alpha_1\alpha_2}(1 - \rho_n)}{(1 + \alpha_1 + \alpha_2)^2 ((1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1\alpha_2} + \rho_n)^2)}. \quad (5.32)$$

Nous montrons sur la figure 5.6, l’excédent de l’EQM (5.32) en fonction de ρ_n pour différents scénarios selon la valeur des rapports signal à bruit α_1 et α_2 : $\alpha_2 = \alpha_1$ (---), $\alpha_2 < \alpha_1$ (....) et $\alpha_2 \ll \alpha_1$ (-.-.-).

Nous remarquons que l’excédent de l’EQM est minimal pour des modalités ayant le même rapport signal à bruit et devient plus important quand les qualités des modalités deviennent plus disparates. De plus, l’excédent de l’EQM est plus grand pour les corrélations positives des bruits comparé aux corrélations négatives.

En effet, lorsque nous ignorons la corrélation des bruits, nous ne permettons pas aux modalités d’interagir proprement. Ainsi, le modèle inadéquat engendre plus d’erreur que le modèle adéquat.

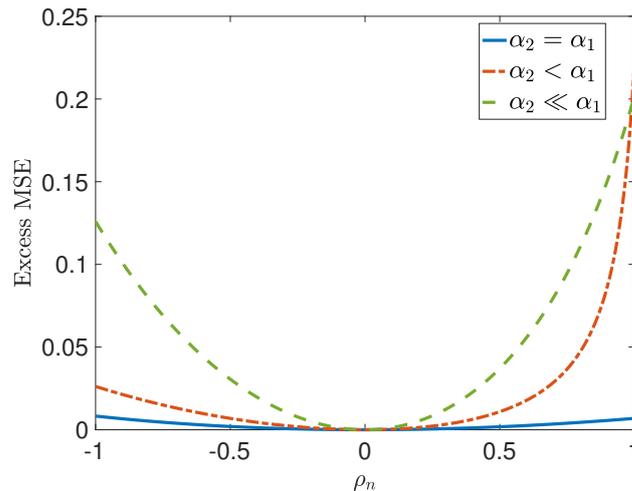


FIGURE 5.6 – $\mathbf{M}_{p,p}(s|x_1)$ et $\mathbf{M}_{p,q}(s|x_1, x_2) \Big|_{\hat{\rho}_n=0}$ pour $\alpha_1 = 5$ et différentes valeurs de α_2 : $\alpha_2 = \alpha_1 = 5$, $\alpha_2 = 2 < \alpha_1$ et $\alpha_2 = 0.1 \ll \alpha_1$, en fonction de ρ_n .

De plus, au moyen de la corrélation du bruit, le modèle adéquat tire profit de la modalité de faible qualité pour avoir de l'information sur le bruit qui permet une meilleure performance surtout pour les valeurs positives des corrélations. Au contraire, dans le cas inadéquat, nous négligeons cette information.

Lorsque nous supposons à tort que les bruits sont indépendants, nous n'exploitons pas au mieux l'information sur le bruit qu'apporte la modalité de faible qualité surtout si les bruits sont, réellement, positivement corrélés.

5.3.2 Modèle inadéquat *versus* une seule modalité

L'intérêt de la multimodalité est de fournir plus d'information sur le signal d'intérêt qu'une seule modalité. Ainsi, quoique le modèle inadéquat $\widehat{\mathcal{M}}_h$ engendre plus d'erreur que le modèle exact \mathcal{M}_h^2 , il demeure néanmoins intéressant tant qu'il assure une bonne performance par rapport à une seule modalité.

Dans ce paragraphe nous comparons la performance du modèle inadéquat $\widehat{\mathcal{M}}_h$ à la performance d'une seule modalité. L'estimateur fondé sur les deux modalités de $\widehat{\mathcal{M}}_h$ donne l'erreur d'estimation $\mathbf{M}_{p,q}(s|x_1, x_2) \Big|_{\hat{\rho}_n=0}$, alors que celui fondé sur la modalité x_i engendre l'erreur d'estimation $\mathbf{M}_{p,p}(s|x_i)$ ($i \in \{1, 2\}$).

5.3.2.1 Modèle inadéquat *versus* x_1

Nous comparons l'erreur d'estimation multimodale inadéquate, $\mathbf{M}_{p,q}(s|x_1, x_2) \Big|_{\hat{\rho}_n=0}$, avec l'erreur d'estimation basée sur la modalité ayant le rapport signal à bruit le plus élevé x_1 , $\mathbf{M}_{p,p}(s|x_1)$.

A partir des équations (5.5) et (5.16) nous obtenons que, le modèle inadéquat est aussi précis qu'une seule modalité x_1 , $\mathbf{M}_{p,q}(s|x_1, x_2) \Big|_{\hat{\rho}_n=0} = \mathbf{M}_{p,p}(s|x_1)$, si et seulement si $\rho_n =$

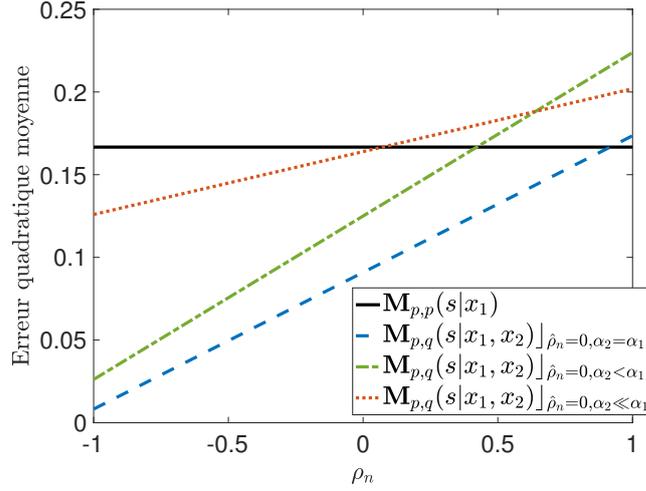


FIGURE 5.7 – Variations, en fonction de ρ_n , de $\mathbf{M}_{p,p}(s|x_1)$ et $\mathbf{M}_{p,q}(s|x_1, x_2)|_{\hat{\rho}_n=0}$ pour différentes valeurs de α_2 : $\alpha_2 = \alpha_1$, $\alpha_2 < \alpha_1$ et $\alpha_2 \ll \alpha_1$.

$\frac{1+\alpha_1+\alpha_2}{2(1+\alpha_1)} \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$. C'est-à-dire qu'en supposant $\hat{\rho}_n = 0$ alors que $\rho_n = \frac{1+\alpha_1+\alpha_2}{2(1+\alpha_1)} \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$, nous négligeons toute l'information qu'apporte la modalité x_2 .

Au delà de cette valeur, *i.e* $\rho_n \geq \frac{1+\alpha_1+\alpha_2}{2(1+\alpha_1)} \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$, le modèle inadéquat engendre plus d'erreur qu'une seule modalité. C'est-à-dire que conjointement les deux modalités sont plus mauvaises que la seule modalité x_1 . Par conséquent, il vaut mieux considérer une seule modalité x_1 que les deux modalités en ignorant les corrélations de l'intervalle :

$$I_{une}(s|x_1) = \left[\frac{1 + \alpha_1 + \alpha_2}{2(1 + \alpha_1)} \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}, 1 \right]. \quad (5.33)$$

La figure 5.7 illustre, en fonction de ρ_n , $\mathbf{M}_{p,p}(s|x_1)$ (—) et $\mathbf{M}_{p,q}(s|x_1, x_2)|_{\hat{\rho}_n=0}$ pour différentes valeurs de α_2 : $\alpha_2 = \alpha_1$ (-.-.), $\alpha_2 < \alpha_1$ (-.-.) et $\alpha_2 \ll \alpha_1$ (.....).

Nous remarquons que l'intervalle (5.33) devient plus large, quand les qualités des modalités x_1 et x_2 deviennent plus disparates. C'est-à-dire que le modèle inadéquat exploite mal l'information sur le bruit qu'apporte la modalité x_2 . Si x_2 est une mesure pure du bruit, $\alpha_2 = 0$, le modèle multimodal inadéquat est équivalent à la seule modalité x_1 .

Les corrélations négatives dans les trois scénarios des rapports signal à bruit garantissent une bonne performance par rapport à une seule modalité. Même sous inadéquation du modèle, la corrélation négative demeure la condition pour une meilleure performance de la multimodalité.

Lorsque nous ignorons une corrélation positive, le modèle inadéquat risque d'être plus mauvais que la seule modalité ayant le rapport signal à bruit le plus élevé. Au contraire, en ignorant une corrélation négative, nous sommes toujours sûrs que le modèle inadéquat est meilleur qu'une seule modalité.

5.3.2.2 Modèle inadéquat *versus* x_2

Le modèle inadéquat engendre parfois plus d'erreur que la modalité de rapport signal à bruit le plus élevé, qu'en est-il de la modalité de faible qualité x_2 ? Nous comparons la performance du faux modèle $\widehat{\mathcal{M}}_h^2$ à la performance de la modalité de faible qualité x_2 . A partir des équations (5.5) et (5.16) nous trouvons que le modèle inadéquat engendre plus d'erreur que la modalité de faible qualité si et seulement si $\rho_n \geq \frac{1+\alpha_1+\alpha_2}{2(1+\alpha_2)} \frac{\sqrt{\alpha_1}}{\sqrt{\alpha_2}}$. Ce qui est impossible parce que $\frac{1+\alpha_1+\alpha_2}{2(1+\alpha_2)} \frac{\sqrt{\alpha_1}}{\sqrt{\alpha_2}} \geq 1$ et $|\rho_n| \leq 1$. Par conséquent, le modèle inadéquat est toujours plus précis que la modalité de faible qualité x_2 .

5.4 Inadéquation quelconque des liens entre les bruits

Nous considérons, dans cette section, le cas général d'inadéquation des liens entre les bruits. Nous supposons que la vraie valeur de corrélation des bruits, ρ_n , est inconnue et que la valeur considérée, $\hat{\rho}_n$, diffère de la valeur exacte. Dans le cas général d'inadéquation des liens entre les bruits l'estimateur inadéquat est donné par l'équation (5.16) :

$$\hat{s}_q(\mathbf{x}) = \frac{(\sqrt{\alpha_1} - \sqrt{\alpha_2}\hat{\rho}_n)x_1 + (\sqrt{\alpha_2} - \sqrt{\alpha_1}\hat{\rho}_n)x_2}{(1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1\alpha_2} + \hat{\rho}_n)^2}. \quad (5.34)$$

Cet estimateur utilise un faux lien pour optimiser l'interaction des modalités. Ainsi, $\hat{s}_q(\mathbf{x})$ engendre l'erreur d'estimation multimodale :

$$\begin{aligned} \mathbf{M}_{p,q}(s|x_1, x_2) &= \frac{1 - \hat{\rho}_n^2}{\underbrace{(1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1\alpha_2} + \hat{\rho}_n)^2}_{\mathbf{M}_{q,q}(s|x_1, x_2)}} \\ &\quad + 2(\rho_n - \hat{\rho}_n) \frac{\overbrace{(\sqrt{\alpha_1} - \sqrt{\alpha_2}\hat{\rho}_n)}^{\hat{k}_1} \overbrace{(\sqrt{\alpha_2} - \sqrt{\alpha_1}\hat{\rho}_n)}^{\hat{k}_2}}{((1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1\alpha_2} + \hat{\rho}_n)^2)^2}, \\ &= a\rho_n + b. \end{aligned}$$

Le premier terme de la partie droite de cette équation correspond à l'erreur d'estimation anticipée si nous supposons que la densité supposée est exacte, $\mathbf{M}_{q,q}(s|x_1, x_2)$. Cependant, puisque $q(\mathbf{n}) \neq p(\mathbf{n})$, l'erreur obtenue peut être supérieure ou inférieure à $\mathbf{M}_{q,q}(s|x_1, x_2)$: en effet, selon l'hypothèse $\alpha_2 \leq \alpha_1$, le signe du deuxième terme correspond au signe du produit $(\rho_n - \hat{\rho}_n)(\sqrt{\alpha_2} - \sqrt{\alpha_1}\hat{\rho}_n)$.

5.4.1 Modèle inadéquat *versus* modèle adéquat

Nous examinons l'effet de l'hypothèse $\hat{\rho}_n$, sur l'interaction des modalités. Nous considérons trois scénarios selon les qualités des modalités et nous comparons, dans chaque scénario, l'erreur d'estimation engendrée par le modèle adéquat \mathcal{M}_h^2 et l'erreur d'estimation obtenue à partir du modèle inadéquat $\widehat{\mathcal{M}}_h^2$.

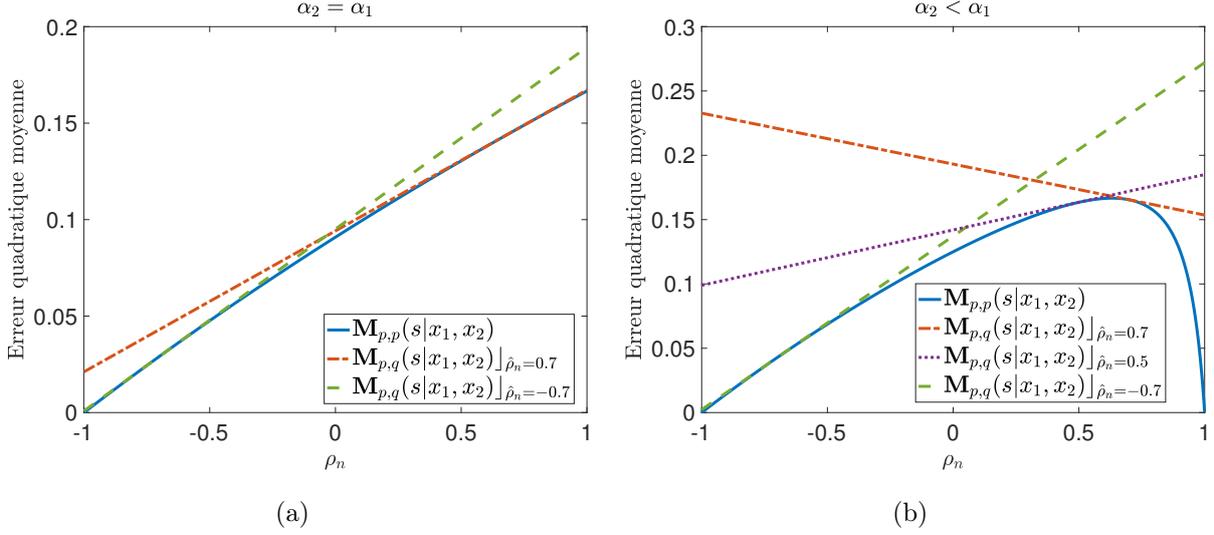


FIGURE 5.8 – Variations, en fonction de ρ_n , de $\mathbf{M}_{p,p}(s|x_1, x_2)$ et $\mathbf{M}_{p,q}(s|x_1, x_2)$ pour différentes valeurs d'inadéquation, avec $\alpha_2 = \alpha_1 = 5$ (a) et $\alpha_2 = 2 < \alpha_1$ (a) .

5.4.1.1 Modalités de même qualité ($\alpha_2 = \alpha_1$)

Nous commençons cette étude par le cas de deux modalités de même rapport signal à bruit. L'EQM inadéquate est :

$$\begin{aligned} \mathbf{M}_{p,q}(s|x_1, x_2) &= \frac{1 + \hat{\rho}_n}{1 + 2\alpha_1 + \hat{\rho}_n} + (\rho_n - \hat{\rho}_n) 2\alpha_1 \frac{1}{(1 + 2\alpha_1 + \hat{\rho}_n)^2}, \\ &= a\rho_n + b. \end{aligned}$$

La pente a est positive $\forall \hat{\rho}_n$, ainsi $\mathbf{M}_{p,q}(s|x_1, x_2)$ est une fonction croissante de ρ_n .

La figure 5.8a illustre, en fonction du paramètre ρ_n , $\mathbf{M}_{p,p}(s|x_1, x_2)$ (—) et $\mathbf{M}_{p,q}(s|x_1, x_2)$ pour deux valeurs d'inadéquation $\hat{\rho}_n = 0.7$ (---) et $\hat{\rho}_n = -0.7$ (-.-).

Nous constatons que $\mathbf{M}_{p,q}(s|x_1, x_2)$ est supérieure à $\mathbf{M}_{p,p}(s|x_1, x_2)$ avec égalité si et seulement si $\hat{\rho}_n = \rho_n$, parce que le modèle inadéquat exploite les liens entre les modalités d'une manière sous optimale.

Les trois erreurs d'estimation sont des fonctions croissantes de ρ_n , puisque le bruit ne peut pas être compensé dans le cas des corrélations positives aussi bien que dans le cas des corrélations négatives sans dégrader la qualité du signal s .

5.4.1.2 Modalités de différentes qualités ($\alpha_2 < \alpha_1$)

La figure 5.8b montre pour $\alpha_2 < \alpha_1$, $\mathbf{M}_{p,p}(s|x_1, x_2)$ (—), $\mathbf{M}_{p,q}(s|x_1, x_2)$ pour trois valeurs d'inadéquation : $\hat{\rho}_n = 0.7$ (---), $\hat{\rho}_n = 0.5$ (....) et $\hat{\rho}_n = -0.7$ (-.-). Nous remarquons que selon la valeur de $\hat{\rho}_n$, $\mathbf{M}_{p,q}(s|x_1, x_2)$ est une fonction croissante ou décroissante de ρ_n .

Si nous examinons les estimateurs adéquat, $\hat{s}_p(\mathbf{x})$ (5.6), et inadéquat, $\hat{s}_q(\mathbf{x})$ (5.12), nous constatons que les facteurs de x_1 dans les deux estimateurs sont toujours positifs, $k_1 \geq 0$ et $\hat{k}_1 \geq 0$, puisque $|\rho_n| \leq 1$, $|\hat{\rho}_n| \leq 1$ et $\alpha_2 < \alpha_1$.

Cependant, les facteurs de x_2 , k_2 et \hat{k}_2 , peuvent être positifs ou négatifs. Si \hat{k}_2 a le même signe que k_2 , l'estimateur inadéquat arrive, d'une manière sous-optimale, à compenser le bruit. Toutefois, si le signe de \hat{k}_2 est différent du signe de k_2 , l'estimateur inadéquat amplifie le bruit et engendre ainsi de larges erreurs d'estimation.

$\hat{\rho}_n < 0$: Nous supposons que le bruit peut être réduit par une somme pondérée des modalités, ainsi $\hat{k}_1 > 0$ et $\hat{k}_2 > 0$. Cependant, si $\rho_n > 0$ cette hypothèse amplifie le bruit. Nous remarquons sur la figure 5.8b que l'erreur d'estimation inadéquate est croissante pour $\hat{\rho}_n = -0.7$, car le bruit est bien compensé pour les corrélations négatives comparé aux corrélations positives.

$0 < \hat{\rho}_n < \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$: Nous supposons que les modalités sont partiellement redondantes et que le bruit ne peut pas être mieux compensé. Ainsi, $\hat{k}_1 > 0$ et $\hat{k}_2 > 0$ pour ne pas dégrader le signal s . Sur la figure 5.8b pour $\hat{\rho}_n = 0.5$ l'erreur d'estimation est une fonction croissante de ρ_n .

$\frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}} < \hat{\rho}_n$: Dans ce cas, l'estimateur inadéquat est une différence pondérée de x_1 et x_2 . Ainsi, le bruit est bien compensé pour $\rho_n > \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$ et l'erreur d'estimation inadéquate est une fonction décroissante de ρ_n . Mais l'excédant de l'EQM si la vraie corrélation est négative peut devenir très important.

5.4.2 Modèle inadéquat *versus* une seule modalité

Nous étudions dans ce paragraphe les inadéquations quelconques du bruit qui engendrent plus d'erreur qu'une seule modalité. A cette fin nous comparons l'erreur d'estimation du modèle inadéquat à l'erreur d'estimation obtenue à partir d'une seule modalité.

5.4.2.1 Modèle inadéquat *versus* x_1

A partir de (5.5) et (5.16) nous trouvons que le modèle inadéquat est équivalent à la seule modalité x_1 si $\hat{\rho}_n = \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$ ou si $\rho_n = \rho_n^*$ †. Dans le premier cas, nous supposons que x_2 est complètement redondante par rapport à x_1 . Ainsi, nous négligeons toute information additionnelle que x_2 peut communiquer. Dans ce cas, le modèle inadéquat est en fonction d'une seule modalité x_1 , $\hat{s}_q(\mathbf{x}) = \hat{s}_q(x_1)$. Dans le deuxième cas, le modèle inadéquat $\widehat{\mathcal{M}}_h^2$ exploite les deux modalités mais engendre la même erreur d'estimation que la seule modalité x_1 .

Le modèle inadéquat engendre plus d'EQM que la modalité x_1 si $\rho_n \in I_{une}(s|x_1)$:

$$I_{une}(s|x_1) = \begin{cases} \left[-1, \rho_n^* \right] & \text{si } \hat{\rho}_n > \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}} \\ \left[\rho_n^*, 1 \right] & \text{si } \hat{\rho}_n < \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}} \end{cases}. \quad (5.35)$$

La figure 5.9 illustre $\mathbf{M}_{p,p}(s|x_1)$ (—), $\mathbf{M}_{p,q}(s|x_1, x_2)$ pour trois valeurs d'inadéquation : $\hat{\rho}_n = 0.7$ (-.-.-), $\hat{\rho}_n = 0.5$ (.....) et $\hat{\rho}_n = -0.7$ (-.-.-), avec $\alpha_2 = \alpha_1$ (a) et $\alpha_2 < \alpha_1$ (a).

†. $\rho_n^* = \hat{\rho}_n + \frac{(\sqrt{\alpha_2} - \sqrt{\alpha_1} \hat{\rho}_n)((1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1} \alpha_2 + \hat{\rho}_n)^2)}{2(1 + \alpha_1)(\sqrt{\alpha_1} - \sqrt{\alpha_2} \hat{\rho}_n)}$.

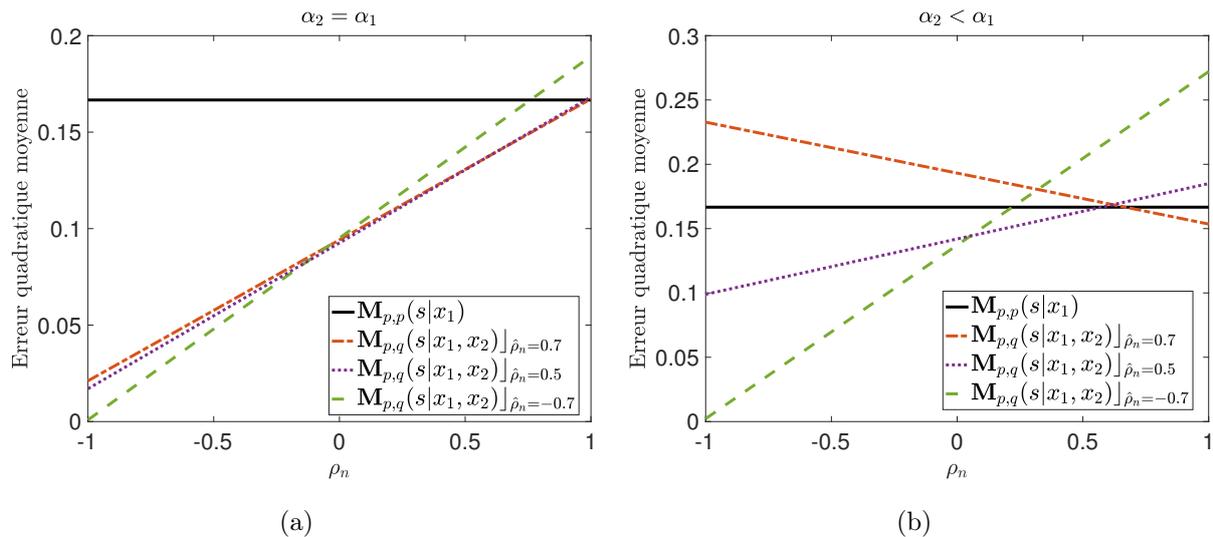


FIGURE 5.9 – $\mathbf{M}_{p,p}(s|x_1, x_2)$ et $\mathbf{M}_{p,q}(s|x_1, x_2)$ pour trois valeurs d'inadéquation : $\hat{\rho}_n = 0.7$, $\hat{\rho}_n = 0.5$ et $\hat{\rho}_n = -0.7$, avec $\alpha_2 = \alpha_1 = 5$ (a) et $\alpha_2 = 2 < \alpha_1$ (b).

Nous remarquons que pour des modalités de même qualité, l'estimateur inadéquat est moins précis que x_1 pour des valeurs positives des corrélations ρ_n proches de 1. Cependant, pour des modalités de différentes qualités, le modèle inadéquat peut introduire plus d'erreur d'estimation que x_1 dans le cas de corrélations exactes négatives, contrairement au cas $\hat{\rho}_n = 0$, notamment si $\hat{\rho}_n \geq \frac{\sqrt{\alpha_2}}{\sqrt{\alpha_1}}$ ($\hat{\rho}_n = 0.7$ sur la figure 5.9b).

Pour $\hat{\rho}_n \neq 0$, la corrélation négative exacte n'assure pas toujours une meilleure performance qu'une seule modalité, sauf si $\alpha_2 = \alpha_1$.

5.4.2.2 Modèle inadéquat *versus* x_2

Nous avons trouvé qu'en ignorant la corrélation du bruit, l'estimateur inadéquat est toujours plus précis que la modalité ayant le rapport signal à bruit le plus petit. Cependant, dans le cas général d'inadéquation, le modèle multimodal inadéquat peut engendrer plus d'erreur que la modalité de faible qualité pour $\rho_n \geq \hat{\rho}_n + \frac{(\sqrt{\alpha_1} - \sqrt{\alpha_2}\hat{\rho}_n)((1+\alpha_1)(1+\alpha_2) - (\sqrt{\alpha_1\alpha_2} + \hat{\rho}_n)^2)}{2(1+\alpha_2)(\sqrt{\alpha_1} - \sqrt{\alpha_2}\hat{\rho}_n)}$, car on a alors $\mathbf{M}_{p,q}(s|x_1, x_2) \geq \mathbf{M}_{p,p}(s|x_2)$.

Pour $\hat{\rho}_n \neq 0$, le modèle multimodal inadéquat peut être moins précis que la modalité de plus faible rapport signal à bruit.

5.5 Conclusion

Dans ce chapitre nous avons examiné l'effet des fausses hypothèses sur les liens inter-modalités. Dans un premier temps, nous avons étudié l'hypothèse la plus répandue en pratique qui est l'indépendance des bruits, avant d'étudier l'effet d'une inadéquation quelconque de la corrélation

des bruits. Nous avons montré que les modalités n'interagissent pas efficacement sous inadéquation du modèle et peuvent engendrer plus d'erreur d'estimation qu'une seule modalité, même que celle de faible rapport signal à bruit. Ce qui va complètement à l'encontre du but de la multimodalité. Le modèle simple que nous avons considéré dans ce chapitre, nous montre l'importance d'une bonne modélisation des liens entre les modalités afin d'en tirer le maximum de profit. Dans le chapitre suivant, nous étudions l'effet d'une double inadéquation du modèle sur la multimodalité.

Modèle inadéquat : Modalités avec des couplages faibles

Sommaire

6.1	Introduction	127
6.2	Modèle	128
6.3	Calculs d'estimation	129
6.4	Inadéquation du bruit	131
6.4.1	Corrélation des bruits ignorée	131
6.4.2	Inadéquation quelconque du bruit	138
6.5	Inadéquation du signal	141
6.6	Double inadéquation	145
6.7	Conclusion	146

6.1 Introduction

Dans le chapitre précédent, nous avons étudié l'impact d'un faux lien entre les bruits sur l'interaction de deux modalités avec des signaux fortement couplés. Ce chapitre porte sur un modèle plus général où le couplage entre les signaux d'intérêt est faible.

Nous examinons trois situations d'inadéquation des liens entre les modalités :

1. L'inadéquation du lien entre les bruits seuls, qui permettra de généraliser les résultats du chapitre précédent.
2. L'inadéquation du lien entre les signaux d'intérêts seuls.
3. L'inadéquation du lien entre les bruits et l'inadéquation du lien entre les signaux d'intérêts (double inadéquation).

Ce chapitre est organisé comme suit. Le modèle et les hypothèses considérés sont présentés à la section 6.2. Les mesures d'estimation associées à ce modèle sont détaillées dans la section 6.3. Nous étudions, dans les sections 6.4, 6.5 et 6.6, les différents scénarios d'inadéquation du modèle. La section 6.7 conclut ce chapitre.

6.2 Modèle

Le modèle que nous étudions dans ce chapitre est schématisé par la figure 6.1 : deux modalités complémentaires, chacune d'elles mesure une partie différente du vecteur $\mathbf{s} = [s_1, s_2]^T$.

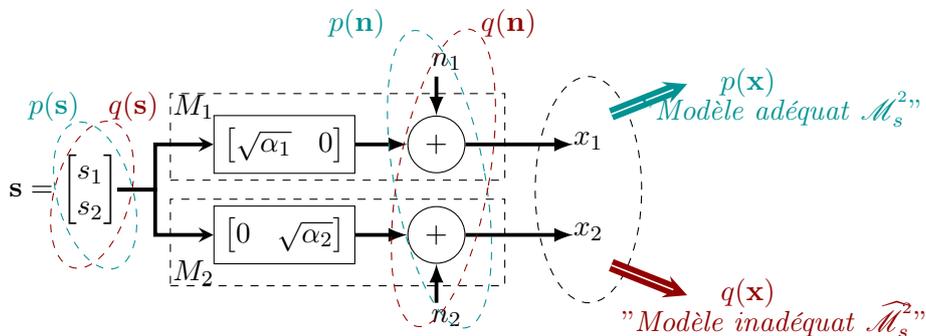


FIGURE 6.1 – Inadéquation sur les liens entre deux modalités avec des couplages faibles

Nous notons les densités exacte et supposée de \mathbf{s} par $p(\mathbf{s})$ et $q(\mathbf{s})$, respectivement, et les densités exacte et supposée de $\mathbf{n} = [n_1, n_2]^T$ par $p(\mathbf{n})$ et $q(\mathbf{n})$, respectivement. Si, $q(\mathbf{s}) = p(\mathbf{s})$ et $q(\mathbf{n}) = p(\mathbf{n})$, $q(\mathbf{x}) = p(\mathbf{x})$ (où $\mathbf{x} = [x_1, x_2]^T$) et le modèle est dit adéquat. Autrement, si $q(\mathbf{s}) \neq p(\mathbf{s})$ et/ou $q(\mathbf{n}) \neq p(\mathbf{n})$, $q(\mathbf{x}) \neq p(\mathbf{x})$ et le modèle est dit inadéquat.

Le modèle mathématique associé au schéma ci-dessus est :

$$\begin{cases} x_1 = \sqrt{\alpha_1} s_1 + n_1 \\ x_2 = \sqrt{\alpha_2} s_2 + n_2 \end{cases}, \quad (6.1)$$

où s_1 , s_2 , n_1 et n_2 sont des variables gaussiennes. Sans perte de généralité, nous supposons que ces variables sont de moyennes nulles et de variances unitaires, *i.e.* $\sim \mathcal{N}(0, 1)$. Les variances des signaux d'intérêt et des bruits étant unitaires, α_1 et α_2 représentent les rapports signal à bruit des modalités. Nous supposons que $\alpha_2 \leq \alpha_1$.

Nous supposons que s_1 et s_2 sont corrélés, mais indépendants des bruits n_1 et n_2 . Nous notons le coefficient de corrélation des signaux utiles par $\rho_s = \frac{\mathbb{E}_p[s_1 s_2]}{\mathbb{E}_p[s_1^2] \mathbb{E}_p[s_2^2]}$. Nous supposons que les bruits sont aussi corrélés et nous notons leur coefficient de corrélation par $\rho_n = \frac{\mathbb{E}_p[n_1 n_2]}{\mathbb{E}_p[n_1^2] \mathbb{E}_p[n_2^2]}$. Ainsi, $p(\mathbf{s}) = \mathcal{N}(0, \mathbf{\Gamma})$ et $p(\mathbf{n}) = \mathcal{N}(0, \mathbf{\Sigma})$, avec :

$$\mathbf{\Gamma} = \begin{pmatrix} 1 & \rho_s \\ \rho_s & 1 \end{pmatrix}, \quad \mathbf{\Sigma} = \begin{pmatrix} 1 & \rho_n \\ \rho_n & 1 \end{pmatrix}. \quad (6.2)$$

Liens exacts entre les modalités :

- **Lien exact entre les signaux d'intérêt :** Un couplage faible est considéré entre les signaux d'intérêt représenté par le coefficient de corrélation, ρ_s , entre s_1 et s_2 .
- **Lien exact entre les bruits :** Le couplage que nous considérons entre les bruits est représenté par le coefficient de corrélation, ρ_n , entre n_1 et n_2 .

L'analyse du modèle (6.1) dans le cas du modèle adéquat a fait l'objet du chapitre 3. Dans ce chapitre, nous étudions le modèle (6.1) dans le cas du modèle inadéquat. Nous supposons que les liens supposés entre les modalités peuvent être différents des liens exacts.

Liens supposés entre les modalités :

- **Lien supposé entre les signaux d'intérêt :** Nous notons par $\hat{\rho}_s = \frac{\mathbb{E}_q[s_1 s_2]}{\mathbb{E}_p[s_1^2] \mathbb{E}_p[s_2^2]}$ le coefficient de corrélation supposé entre s_1 et s_2 .
- **Lien supposé entre les bruits :** $\hat{\rho}_n = \frac{\mathbb{E}_q[n_1 n_2]}{\mathbb{E}_p[n_1^2] \mathbb{E}_p[n_2^2]}$ dénote le coefficient de corrélation supposé entre n_1 et n_2 .

Les fausses hypothèses sur les liens entre les modalités entraînent l'inadéquation des densités jointes des signaux utiles, $p(\mathbf{s})$, et des bruits, $p(\mathbf{n})$. En effet, les densités jointes supposés des signaux d'intérêt et des bruits sont $q(\mathbf{s}) = \mathcal{N}(0, \hat{\mathbf{\Gamma}})$ et $q(\mathbf{n}) = \mathcal{N}(0, \hat{\mathbf{\Sigma}})$, où :

$$\hat{\mathbf{\Gamma}} = \begin{pmatrix} 1 & \hat{\rho}_s \\ \hat{\rho}_s & 1 \end{pmatrix}, \quad \hat{\mathbf{\Sigma}} = \begin{pmatrix} 1 & \hat{\rho}_n \\ \hat{\rho}_n & 1 \end{pmatrix}. \quad (6.3)$$

Dans le cas adéquat, *i.e* $\hat{\rho}_s = \rho_s$ et $\hat{\rho}_n = \rho_n$, nous avons noté le modèle (6.1) par \mathcal{M}_s^2 , où l'exposant "2" et l'indice "s" font référence au nombre de modalités et au couplage faible ("soft") entre les signaux d'intérêt, respectivement. Dans le cas inadéquat, *i.e* $\hat{\rho}_s \neq \rho_s$ et/ou $\hat{\rho}_n \neq \rho_n$, nous notons le modèle (6.1) par $\widehat{\mathcal{M}}_s^2$.

Dans la suite, au moyen des mesures d'estimation (section 6.3), nous proposons de répondre aux questions suivantes :

1. Quel est l'impact d'une seule inadéquation sur l'interaction des modalités ?
2. Comment les modalités interagissent sous double inadéquation ?

6.3 Calculs d'estimation

Nous considérons le problème d'estimation de la variable s_i , $i \in \{1, 2\}$, à partir des modalités x_1 et x_2 .

Modèle adéquat. Nous rappelons ici les résultats du chapitre 3 concernant le modèle adéquat. L'estimateur optimal de s_i en fonction d'une seule modalité x_i , $i \in \{1, 2\}$, est :

$$\hat{s}_{i_p}(x_i) \triangleq \int s_i p(s_i | x_i) ds_i = \frac{\sqrt{\alpha_i}}{1 + \alpha_i} x_i. \quad (6.4)$$

L'erreur quadratique moyenne relative à cet estimateur est la variance de l'erreur $s - \hat{s}_{i_p}(x_i)$:

$$\mathbf{M}_{p,p}(s_i | x_i) \triangleq \mathbb{E}_p [(s_i - \hat{s}_{i_p}(x_i))^2] = \frac{1}{1 + \alpha_i}. \quad (6.5)$$

Si nous utilisons les deux modalités conjointement, *i.e* en exploitant les liens exacts entre les modalités, l'estimateur optimal de s_i est :

$$\hat{s}_{i_p}(x_1, x_2) \triangleq \int s_i p(s_i | x_1, x_2) ds_i, \quad (6.6)$$

$$= \frac{k_{i1} x_i + k_{i2} x_j}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n)^2} \quad (6.7)$$

Avec :

$$k_{i1} = \sqrt{\alpha_i}(1 + \alpha_j) - \sqrt{\alpha_j} \rho_s (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n), \quad (6.8)$$

$$k_{i2} = \sqrt{\alpha_j} \rho_s - \sqrt{\alpha_i} \rho_n. \quad (6.9)$$

L'erreur quadratique relative à $\hat{s}_{i_p}(x_1, x_2)$ est donnée par la variance de l'erreur $s_i - \hat{s}_{i_p}(x_1, x_2)$:

$$\begin{aligned} \mathbf{M}_{p,p}(s_i|x_1, x_2) &\triangleq \mathbb{E} [(s_i - \hat{s}_{i_p}(x_1, x_2))^2], \\ &= \frac{\alpha_j(1 - \rho_s^2) + 1 - \rho_n^2}{(1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n)^2}. \end{aligned} \quad (6.10)$$

Modèle inadéquat. Nous supposons maintenant que les liens entre x_1 et x_2 sont décrits par les corrélations $\hat{\rho}_n$ et $\hat{\rho}_s$ au lieu des liens exacts ρ_n et ρ_s , respectivement.

Les fausses hypothèses sur les liens ne concernent pas les lois marginales de s_i et n_i , *i.e* $q(s_i) = p(s_i)$ et $q(n_i) = p(n_i)$. Il s'ensuit que, $q(x_i) = p(x_i)$ et $q(s_i|x_i) = p(s_i|x_i)$ et :

$$\hat{s}_{i_q}(x_i) = \hat{s}_{i_p}(x_i), \quad \mathbf{M}_{p,q}(s_i|x_i) = \mathbf{M}_{p,p}(s_i|x_i). \quad (6.11)$$

Cependant, les lois jointes des bruits et des signaux d'intérêt sont inadéquates, $q(\mathbf{s}) \neq p(\mathbf{s})$ et $q(\mathbf{n}) \neq p(\mathbf{n})$. Ainsi, l'estimateur multimodal de s_i est, selon (4.30), la moyenne de la densité a posteriori supposée $q(s_i|x_1, x_2)$ pour $i \in \{1, 2\}$:

$$\hat{s}_{i_q}(x_1, x_2) \triangleq \int s_i q(s_i|x_1, x_2) ds_i, \quad (6.12)$$

$$= \frac{\hat{k}_{i1} x_i + \hat{k}_{i2} x_j}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1 \alpha_2} \hat{\rho}_s + \hat{\rho}_n)^2}. \quad (6.13)$$

Avec :

$$\hat{k}_{i1} = \sqrt{\alpha_i}(1 + \alpha_j) - \sqrt{\alpha_j} \hat{\rho}_s (\sqrt{\alpha_1 \alpha_2} \hat{\rho}_s + \hat{\rho}_n), \quad (6.14)$$

$$\hat{k}_{i2} = \sqrt{\alpha_j} \hat{\rho}_s - \sqrt{\alpha_i} \hat{\rho}_n. \quad (6.15)$$

L'EQM de l'estimation multimodale de s_i , sous inadéquation du modèle, est la variance de $s_i - \hat{s}_{i_q}(x_1, x_2)$ calculée par une espérance sur la loi des échantillons, donc la vraie loi $p(\cdot)$:

$$\begin{aligned} \mathbf{M}_{p,q}(s_i|x_1, x_2) &\triangleq \mathbb{E}_p [(s_i - \hat{s}_{i_q}(x_1, x_2))^2], \\ &= \mathbf{M}_{q,q}(s_i|x_1, x_2) \\ &\quad - 2\sqrt{\alpha_j}(\rho_s - \hat{\rho}_s) \hat{k}_{i2} \frac{(1 + \alpha_j) - \hat{\rho}_n(\hat{\rho}_n + \sqrt{\alpha_1 \alpha_2} \hat{\rho}_s)}{((1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1 \alpha_2} \hat{\rho}_s)^2)^2} \\ &\quad + 2(\rho_n - \hat{\rho}_n) \hat{k}_{i2} \frac{(\sqrt{\alpha_i}(1 + \alpha_j) - \sqrt{\alpha_j} \hat{\rho}_s (\sqrt{\alpha_1 \alpha_2} \hat{\rho}_s + \hat{\rho}_n))}{((1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1 \alpha_2} \hat{\rho}_s)^2)^2}. \end{aligned} \quad (6.16)$$

où $\mathbf{M}_{q,q}(s_i|x_1, x_2)$ est l'erreur d'estimation que nous prévoyons obtenir, lorsque nous pensons que les corrélations supposées sont exactes. C'est-à-dire l'EQM si $\rho_n = \hat{\rho}_n$ et $\rho_s = \hat{\rho}_s$:

$$\mathbf{M}_{q,q}(s_i|x_1, x_2) = \frac{\alpha_j(1 - \hat{\rho}_s^2) + 1 - \hat{\rho}_n^2}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1\alpha_2}\hat{\rho}_s + \hat{\rho}_n)^2}. \quad (6.17)$$

Les deuxième et troisième termes de la partie droite de l'équation (6.16), correspondent aux inadéquations du signal utile et du bruit, respectivement.

Démonstration. Ces expressions sont déduites des cas généraux du chapitre 4 (paragraphe 4.3.3.3). Les détails des calculs sont fournis à la section D.2 de l'annexe D. \square

6.4 Inadéquation du bruit

Nous commençons notre étude par l'analyse de l'impact d'une seule inadéquation du modèle. Nous supposons que seul le lien entre les bruits est faussement modélisé, *i.e.* $\hat{\rho}_n \neq \rho_n$ et $\hat{\rho}_s = \rho_s$. Nous traitons dans un premier temps l'hypothèse d'indépendance des bruits, $\hat{\rho}_n = 0$, avant d'aborder le cas d'inadéquation quelconque, $\hat{\rho}_n \in [-1, 1]$.

6.4.1 Corrélation des bruits ignorée

L'indépendance des bruits est une hypothèse largement répandue en pratique. Cette hypothèse a pour finalité essentielle de simplifier les calculs complexes. Or, puisque le lien entre les bruits a un rôle important dans l'interaction des modalités (chapitres 2 et 3), quelles sont les implications de cette hypothèse ?

La section 5.3 du chapitre 5 analyse les conséquences de cette hypothèse sur l'interaction des modalités, mais dans le cas des signaux d'intérêts fortement couplés. Les résultats de cette étude suggèrent qu'ignorer une corrélation positive des bruits engendre de larges erreurs qui risquent de ruiner le but de la multimodalité à assurer une meilleure performance qu'une seule modalité.

Dans cette section nous examinons l'effet de cette hypothèse dans le cas de deux modalités complémentaires. Nous comparons la performance, en terme de précision d'estimation, de deux modalités avec des bruits supposés à tort indépendants avec :

1. la performance de deux modalités avec des liens parfaitement modélisés,
2. la performance d'une seule modalité x_i .

6.4.1.1 Modèle inadéquat versus modèle adéquat

Soit \mathcal{M}_s^2 le modèle de deux modalités dont les liens sont correctement modélisés. En utilisant les modalités de \mathcal{M}_s^2 , l'estimateur optimal de s_i , $i \in \{1, 2\}$, et l'erreur d'estimation associée sont donnés par les équations (6.7) et (6.10), respectivement.

Soit $\widehat{\mathcal{M}}_s^2$ le modèle de deux modalités dont la corrélation entre les bruits est ignorée, *i.e.* $\hat{\rho}_n = 0$. L'estimateur multimodale de s_i , $i \in \{1, 2\}$ en fonction des modalités de $\widehat{\mathcal{M}}_s^2$ est :

$$\hat{s}_{i_q}(x_1, x_2) \Big|_{\hat{\rho}_n=0, \hat{\rho}_s=\rho_s} = \frac{\overbrace{\sqrt{\alpha_i} (1 + \alpha_j - \alpha_j \rho_s^2)}^{\hat{k}_{i1}} x_i + \overbrace{\sqrt{\alpha_j} \rho_s}^{\hat{k}_{i2}} x_j}{(\alpha_1 + 1)(\alpha_2 + 1) - \alpha_1 \alpha_2 \rho_s^2} \quad (6.18)$$

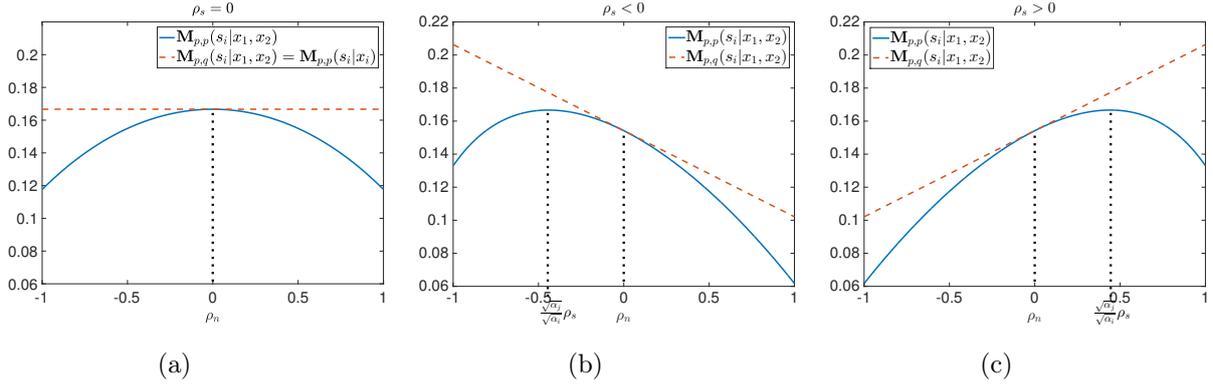


FIGURE 6.2 – $\mathbf{M}_{p,p}(s_i|x_1, x_2)$ et $\mathbf{M}_{p,q}(s_i|x_1, x_2) \Big|_{\hat{\rho}_n=0}$, en fonction de ρ_n , pour $\rho_s = 0$ (a), $\rho_s = -0.7 < 0$ (b) et $\rho_s = 0.7 > 0$ (c), où $\alpha_1 = 5$ et $\alpha_2 = 2$.

Démonstration. $\hat{s}_{i_q}(x_1, x_2) \Big|_{\hat{\rho}_n=0, \hat{\rho}_s=\rho_s}$ est obtenu à partir de (6.13) pour $\hat{\rho}_n = 0$ et $\hat{\rho}_s = \rho_s$. \square

L'erreur d'estimation associée à $\hat{s}_{i_q}(x_1, x_2) \Big|_{\hat{\rho}_n=0, \hat{\rho}_s=\rho_s}$ est :

$$\begin{aligned} \mathbf{M}_{p,q}(s_i|x_1, x_2) \Big|_{\hat{\rho}_n=0} &= \underbrace{\frac{\alpha_j(1 - \rho_s^2) + 1}{(\alpha_1 + 1)(\alpha_2 + 1) - \alpha_1\alpha_2\rho_s^2}}_{\mathbf{M}_{q,q}(s_i|x_1, x_2) \Big|_{\hat{\rho}_n=0}} \\ &\quad + 2\sqrt{\alpha_1\alpha_2}\rho_n\rho_s \frac{\alpha_j(1 - \rho_s^2) + 1}{((1 + \alpha_1)(1 + \alpha_2) - \alpha_1\alpha_2\hat{\rho}_s^2)^2} \end{aligned} \quad (6.19)$$

Démonstration. $\mathbf{M}_{p,q}(s_i|x_1, x_2) \Big|_{\hat{\rho}_n=0}$ est obtenue à partir de (6.16) pour $\hat{\rho}_n = 0$ et $\hat{\rho}_s = \rho_s$. \square

Le premier terme de la partie droite de (6.19) est égal à $\mathbf{M}_{q,q}(s_i|x_1, x_2) \Big|_{\hat{\rho}_n=0}$, qui est l'erreur quadratique d'estimation si $\rho_n = \hat{\rho}_n = 0$.

Le signe du deuxième terme du côté droit de (6.19) est le signe du produit $\rho_n\rho_s$. Ainsi, si la corrélation des bruits est du même signe que la corrélation des signaux, *i.e* $\text{sgn}(\rho_n) = \text{sgn}(\rho_s)$, le modèle inadéquat engendre une erreur supérieure à $\mathbf{M}_{q,q}(s_i|x_1, x_2) \Big|_{\hat{\rho}_n=0}$. Autrement, *i.e* $\text{sgn}(\rho_n) \neq \text{sgn}(\rho_s)$, $\mathbf{M}_{p,q}(s_i|x_1, x_2) \Big|_{\hat{\rho}_n=0} \leq \mathbf{M}_{q,q}(s_i|x_1, x_2) \Big|_{\hat{\rho}_n=0}$.

- $\rho_s = 0$:

Nous commençons cette étude par le cas où s_1 et s_2 sont totalement indépendants, *i.e* $\rho_s = 0$. Sur la figure 6.2a nous présentons $\mathbf{M}_{p,p}(s_i|x_1, x_2)$ (—) et $\mathbf{M}_{p,q}(s_i|x_1, x_2) \Big|_{\hat{\rho}_n=0}$ (---).

Modèle adéquat. Dans le cas adéquat, bien que s_1 et s_2 soient indépendants, les deux modalités sont plus précises pour l'estimation de s_i qu'une seule modalité, *i.e* $\mathbf{M}_{p,p}(s_i|x_1, x_2) \geq \mathbf{M}_{p,p}(s_i|x_i)$, puisqu'elles interagissent toujours à travers le lien entre les bruits.

Modèle inadéquat. Dans le cas du modèle inadéquat, en négligeant la corrélation du bruit nous supprimons le seul lien entre les modalités qui rend possible l'interaction des modalités. C'est-à-dire que, nous supposons une totale indépendance entre les modalités. Nous utilisons, ainsi, la seule modalité x_i pour estimer s_i . En effet, $\hat{k}_{i2} = 0$ et $\mathbf{M}_{p,q}(s_i|x_1, x_2)|_{\hat{\rho}_n=0, \rho_s=0} = \mathbf{M}_{p,p}(s_i|x_i)$.

Modèle inadéquat versus modèle adéquat. Nous constatons, sur la figure 6.2a, que $\mathbf{M}_{p,q}(s_i|x_1, x_2)$ est tangente à $\mathbf{M}_{p,p}(s_i|x_1, x_2)$ au point $\rho_n = 0$, pour lequel les deux modèles \mathcal{M}_s^2 et $\widehat{\mathcal{M}}_s^2$ sont égaux. En s'éloignant de cette valeur, c'est-à-dire quand la corrélation du bruit devient de plus en plus importante, la perte en performance due à l'hypothèse d'indépendance des bruits devient plus grande, car le modèle \mathcal{M}_s^2 tire de plus en plus profit de l'information sur le bruit présente dans la modalité x_j alors que le modèle $\widehat{\mathcal{M}}_s^2$ ignore cette information.

- $\rho_s \neq 0$:

Considérons le cas $\rho_s \neq 0$. Les figures 6.2b et 6.2c illustrent $\mathbf{M}_{p,p}(s_i|x_1, x_2)$ et $\mathbf{M}_{p,q}(s_i|x_1, x_2)|_{\hat{\rho}_n=0}$ pour $\rho_s = -0.7$ et $\rho_s = 0.7$, respectivement.

Modèle adéquat. Etant donné que les bruits sont corrélés, le modèle \mathcal{M}_s^2 exploite la modalité x_j pour extraire de l'information sur le signal d'intérêt et sur le bruit.

Sur les figures 6.2b et 6.2c, nous remarquons que l'interaction entre les deux corrélations entraîne de larges erreurs pour $\text{sgn}(\rho_n) = \text{sgn}(\rho_s)$. En particulier, pour $\rho_n = \frac{\sqrt{\alpha_j}}{\sqrt{\alpha_i}} \rho_s$, $\mathbf{M}_{p,p}(s_i|x_1, x_2)$ atteint sa valeur maximale. Au delà de cette valeur, *i.e* $|\rho_n| \geq \frac{\sqrt{\alpha_j}}{\sqrt{\alpha_i}} |\rho_s|$, on peut compenser le bruit sans nuire au signal d'intérêt, puisque les bruits sont fortement corrélés. Au contraire, pour $\text{sgn}(\rho_s) \neq \text{sgn}(\rho_n)$, le bruit peut être toujours réduit sans dégrader la qualité du signal d'intérêt.

Modèle inadéquat. Si $\rho_s \neq 0$, puisque nous supposons que les bruits sont indépendants, le modèle inadéquat $\widehat{\mathcal{M}}_s^2$ exploite la modalité x_j comme source d'information sur le signal d'intérêt uniquement et non sur le bruit. Par conséquent, le signe de \hat{k}_{i2} est le signe de ρ_s . Cependant, les bruits sont en fait corrélés et l'interaction des deux corrélations engendrent des résultats différents de ce qui est anticipés, *i.e* $\mathbf{M}_{q,q}(s_i|x_1, x_2)|_{\hat{\rho}_n=0}$. En effet, le modèle inadéquat permet intentionnellement de réduire le bruit pour $\text{sgn}(\rho_s) \neq \text{sgn}(\rho_n)$. Ainsi, comme nous le remarquons $\mathbf{M}_{p,q}(s_i|x_1, x_2)$ est une fonction croissante de ρ_n si $\rho_s > 0$ (figure 6.2c) et est décroissante si $\rho_s < 0$ (figure 6.2b).

Modèle inadéquat versus modèle adéquat. Sur les figures 6.2b et 6.2c, $\mathbf{M}_{p,q}(s_i|x_1, x_2)$ est tangente à $\mathbf{M}_{p,p}(s_i|x_1, x_2)$ au point $\rho_n = 0$. Alors que pour $\rho_n \neq 0$, $\widehat{\mathcal{M}}_s^2$ peut engendrer de grandes erreurs d'estimation, *i.e* $\mathbf{M}_{p,q}(s_i|x_1, x_2)|_{\hat{\rho}_n=0} \geq \max\{\mathbf{M}_{p,p}(s_i|x_1, x_2)\}$. Notamment, pour de fortes corrélations des bruits ayant le même signe que ρ_s , car $\widehat{\mathcal{M}}_s^2$ n'exploite alors pas l'information sur le bruit présente dans x_j qui permet de compenser le bruit.

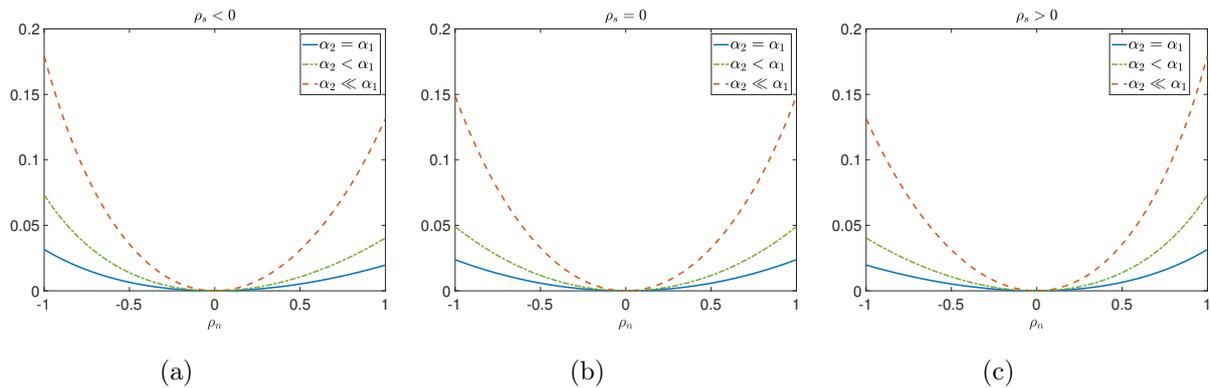


FIGURE 6.3 – $\mathbf{M}_{p,q}(s_1|x_1, x_2) - \mathbf{M}_{p,p}(s_1|x_1, x_2)$, en fonction de ρ_n , pour $\rho_s = -0.7$ (a), $\rho_s = 0$ (b) et $\rho_s = 0.7$ (c), pour $\alpha_1 = 5$ et différentes valeurs de α_2 : $\alpha_2 = \alpha_1 = 5$ (a), $\alpha_2 = 2$ (b) et $\alpha_2 = 0.1$ (c).

• Effet des rapports signal à bruit :

Nous étudions à présent l'excédent de l'EQM, *i.e* $\mathbf{M}_{p,q}(s_1|x_1, x_2)|_{\hat{\rho}_n=0} - \mathbf{M}_{p,p}(s_1|x_1, x_2)$, selon les différents scénarios des rapports signal à bruit des modalités.

La figure 6.3 montre l'excédent de l'EQM en fonction de ρ_n pour $\alpha_1 = 5$ et différentes valeurs du rapport signal à bruit α_2 : $\alpha_2 = \alpha_1$ (—), $\alpha_2 < \alpha_1$ (---) et $\alpha_2 \ll \alpha_1$ (-.-). Les sous-figures 6.3a, 6.3b et 6.3c correspondent à différentes valeurs de ρ_s .

A partir de ces figures, nous remarquons que pour les différentes valeurs de ρ_s , l'excédent de l'EQM est plus important quand les qualités des modalités sont disparates. En effet, les modalités de différentes qualités sont plus sensibles à l'inadéquation de ρ_n , car nous faisons plus confiance à l'information sur le bruit qui est erronée. Ce résultat demeure valide pour $\hat{\rho}_n \neq 0$.

L'excédent de l'EQM, dû à l'inadéquation du bruit, est plus large dans le cas de modalités ayant des rapports signal à bruit disparates.

6.4.1.2 Modèle inadéquat versus une seule modalité

Dans ce paragraphe, nous examinons les cas où le modèle inadéquat $\widehat{\mathcal{M}}_s^2$ est moins précis pour l'estimation de s_i qu'une seule modalité. En comparant les erreurs d'estimation $\mathbf{M}_{p,q}(s_i|x_1, x_2)|_{\hat{\rho}_n=0}$ et $\mathbf{M}_{p,p}(s_i|x_i)$, nous trouvons que le modèle multimodal inadéquat $\widehat{\mathcal{M}}_s^2$ est moins précis que x_i si $\rho_n \in I_{une}(s_i|x_i)|_{\hat{\rho}_n}$, avec :

$$I_{une}(s_i|x_i)|_{\hat{\rho}_n=0} = \begin{cases} [-1, \rho_n^*]_{\hat{\rho}_n=0} & \text{si } \rho_s < 0 \\ [-1, 1] & \text{si } \rho_s = 0 \\ [\rho_n^*]_{\hat{\rho}_n=0}, 1 & \text{si } \rho_s > 0 \end{cases} \quad (6.20)$$

où :

$$\rho_n^*]_{\hat{\rho}_n=0} = \rho_s \frac{\sqrt{\alpha_j} (1 + \alpha_1)(1 + \alpha_2) - \alpha_1 \alpha_2 \rho_s}{\sqrt{\alpha_i} 2(1 + \alpha_i)(1 + \alpha_j(1 - \rho_s^2))}. \quad (6.21)$$

Les figures 6.4 et 6.5 illustrent, pour différentes valeurs de ρ_s , la variation de l'intervalle $I_{une}(s_1|x_1)|_{\hat{\rho}_n=0}$ (resp. $I_{une}(s_2|x_2)|_{\hat{\rho}_n=0}$) (6.20) sur lequel une seule modalité x_1 (resp. x_2) est

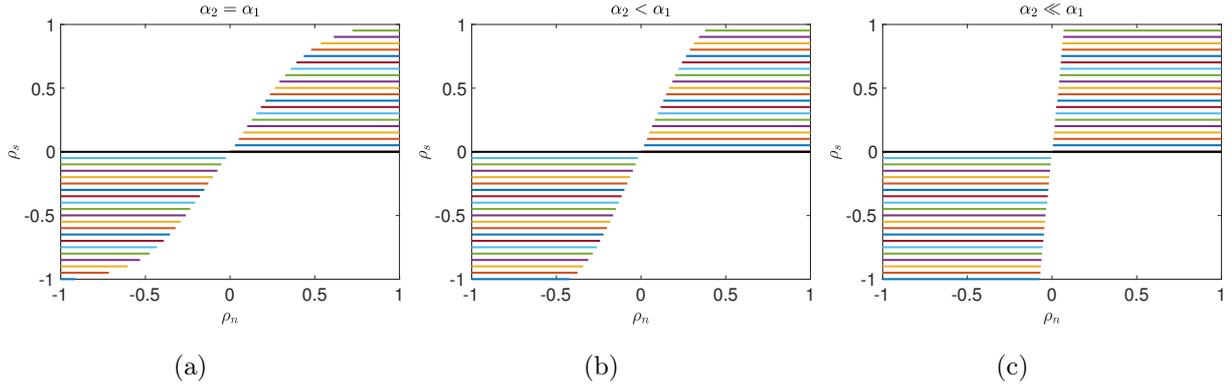


FIGURE 6.4 – Variation de $I_{une}(s_1|x_1)$, en fonction de ρ_s , pour $\alpha_1 = 5$ et différentes valeurs de α_2 : $\alpha_2 = \alpha_1 = 5$ (a), $\alpha_2 = 2$ (b) et $\alpha_2 = 0.1$ (c), avec $\hat{\rho}_n = 0$. Chaque segment présente, pour ρ_s donné, les valeurs exactes de ρ_n qui, lorsqu'elles sont ignorées, engendrent plus d'EQM que la seule modalité x_1 , pour l'estimation de s_1 .

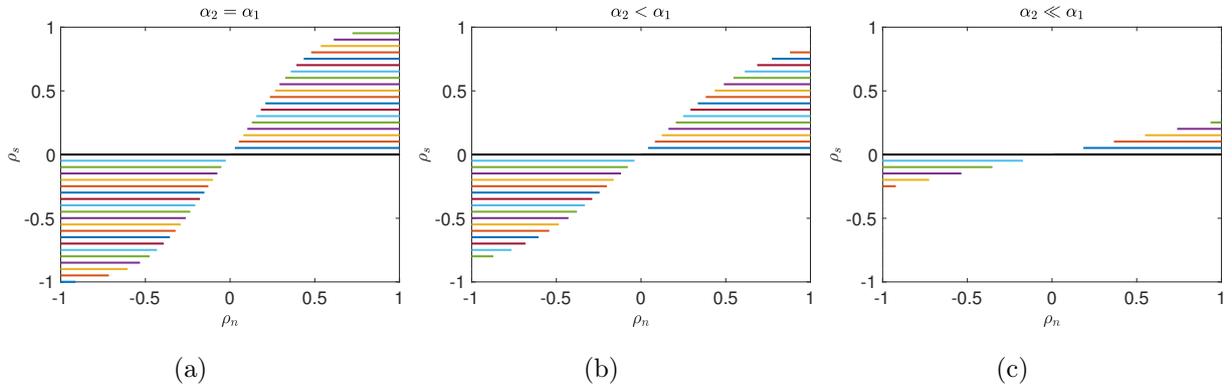


FIGURE 6.5 – Variation de $I_{une}(s_2|x_2)$, en fonction de ρ_s , pour $\alpha_1 = 5$ et différentes valeurs de α_2 : $\alpha_2 = \alpha_1 = 5$ (a), $\alpha_2 = 2$ (b) et $\alpha_2 = 0.1$ (c), avec $\hat{\rho}_n = 0$. Chaque segment présente, pour ρ_s donné, les valeurs exactes de ρ_n pour lesquelles, lorsqu'elles sont ignorées, le modèle inadéquat engendre plus d'EQM que la seule modalité x_2 , pour l'estimation de s_2 .

plus précise pour l'estimation de s_1 (resp. s_2) que le modèle inadéquat $\widehat{\mathcal{M}}_s^2$.

- $\rho_s = 0$: Nous remarquons que pour $\rho_s = 0$, $I_{une}(s_1|x_1)|_{\hat{\rho}_n=0} = [-1, 1]$ et $I_{une}(s_2|x_2)|_{\hat{\rho}_n=0} = [-1, 1]$. En effet, si $\rho_s = 0$, l'hypothèse $\hat{\rho}_n = 0$ suppose une totale indépendance entre les modalités. Par conséquent, puisque nous croyons qu'il n'y a pas d'interaction, les modalités sont utilisées séparément $\hat{s}_{i_q}(x_1, x_2)|_{\hat{\rho}_n=0, \rho_s=0} = \hat{s}_{i_q}(x_i)$, et le modèle inadéquat est donc équivalent à une seule modalité, $\mathbf{M}_{p,q}(s_i|x_1, x_2)|_{\hat{\rho}_n=0} = \mathbf{M}_{p,p}(s_i|x_i)$.

- $\rho_s \neq 0$: Nous constatons que les longueurs des intervalles $I_{une}(s_1|x_1)|_{\hat{\rho}_n=0}$ et $I_{une}(s_2|x_2)|_{\hat{\rho}_n=0}$ sont plus petites par rapport au cas $\rho_s = 0$ et diminuent au fur et à mesure que la corrélation des signaux augmente. En effet, le modèle multimodal inadéquat exploite de l'information sur le signal qui améliore sa précision d'estimation par rapport à une seule modalité.

Nous constatons également que les intervalles $I_{une}(s_1|x_1)|_{\hat{\rho}_n=0}$ et $I_{une}(s_2|x_2)|_{\hat{\rho}_n=0}$ sont situés sur les régions : $\text{sng}(\rho_s) = \text{sgn}(\rho_n)$. Autrement dit, le modèle inadéquat est toujours plus précis

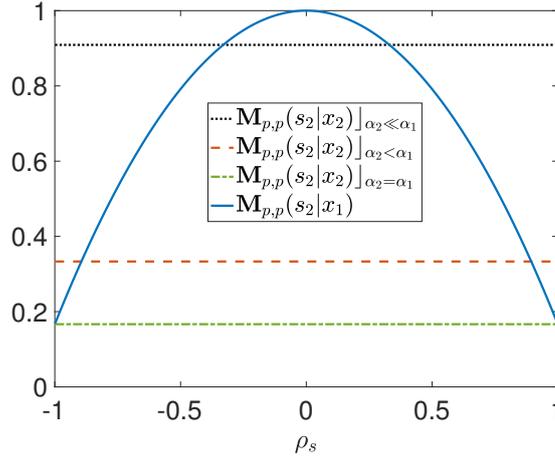


FIGURE 6.6 – $\mathbf{M}_{p,p}(s_2|x_1)$ et $\mathbf{M}_{p,q}(s_2|x_2)$, en fonction de ρ_s , pour $\hat{\rho}_n = 0$, $\alpha_1 = 5$, et différentes valeurs de α_2 : $\alpha_2 = \alpha_1 = 5$, $\alpha_2 = 2$ et $\alpha_2 = 0.1$.

que la seule modalité x_i si $\text{sng}(\rho_s) \neq \text{sng}(\rho_n)$.

Lorsqu'on ignore la corrélation du bruit, le modèle $\widehat{\mathcal{M}}_s^2$ suppose que la modalité x_j est uniquement une source d'information sur s_i . Ainsi, un terme pondéré de x_j est ajouté à x_i (si $\rho_s > 0$) ou soustrait de x_i (si $\rho_s < 0$), afin de tirer profit de l'information sur le signal d'intérêt d'une façon que nous croyons optimale. Cependant, puisque les bruits sont corrélés, ce que nous obtenons diffère de ce que nous anticipons. Si $\text{sng}(\rho_s) \neq \text{sng}(\rho_n)$, l'estimateur inadéquat arrive à réduire le bruit. Alors que si $\text{sng}(\rho_s) = \text{sng}(\rho_n)$, l'information sur le bruit est mal exploitée, ce qui peut engendrer plus d'erreur d'estimation qu'une seule modalité.

Le modèle inadéquat $\widehat{\mathcal{M}}_s^2$, avec $\hat{\rho}_n = 0$ et $\hat{\rho}_s = \rho_s$, est toujours plus précis pour l'estimation de s_i que la seule modalité x_i si $\text{sng}(\rho_s) \neq \text{sng}(\rho_n)$.

• **Effet des rapports signal à bruit :**

Estimation de s_1 . En comparant les figures 6.4a, 6.4b et 6.4c, nous constatons que les longueurs de l'intervalle $I_{une}(s_1|x_1)$ s'élargissent lorsque les qualités des modalités deviennent plus disparates et tendent, pour $\alpha_2 \ll \alpha_1$, vers $[-1, 0]$ si $\rho_s < 0$ et $[0, 1]$ si $\rho_s > 0$. En effet, le modèle inadéquat devient plus sensible à l'inadéquation de la corrélation du bruit quand x_2 (modalité supplémentaire) fournit plus d'information sur le bruit que sur le signal utile.

Estimation de s_2 . Le comportement de $I_{une}(s_2|x_2)]_{\hat{\rho}_n=0}$ (figure 6.5), par rapport aux différents cas des rapports signal à bruit, est différent de $I_{une}(s_1|x_1)]_{\hat{\rho}_n=0}$ (figure 6.4). En effet, lorsque les qualités des modalités deviennent plus disparates, la longueur de $I_{une}(s_1|x_2)]_{\hat{\rho}_n=0}$ se rétrécit. Notamment, pour de fortes corrélations du bruit le modèle inadéquat est toujours plus précis pour l'estimation de s_2 que la seule modalité x_2 , $I_{une}(s_2|x_2)]_{\hat{\rho}_n=0} = \emptyset$. Le modèle inadéquat est moins sensible à l'inadéquation du bruit pour $\alpha_2 \ll \alpha_1$, car il utilise x_1 (moda-

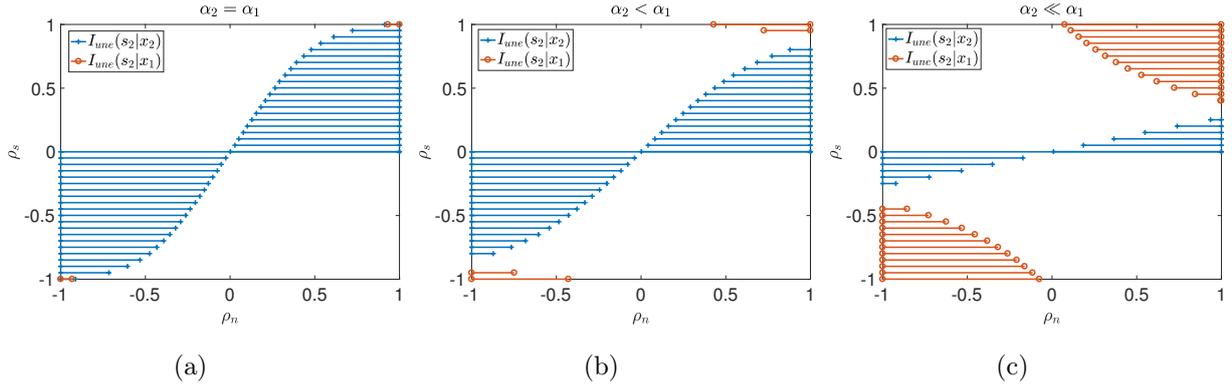


FIGURE 6.7 – Variation de $I_{une}(s_2|x_2)$ et $I_{une}(s_2|x_1)$, en fonction de ρ_s , pour $\alpha_1 = 5$ et différentes valeurs de α_2 : $\alpha_2 = \alpha_1 = 5$ (a), $\alpha_2 = 2$ (b) et $\alpha_2 = 0.1$ (c). Chaque segment présente, pour ρ_s donné, les valeurs exactes de ρ_n qui, lorsqu'elles sont ignorées, engendrent plus d'EQM que la seule modalité x_2 ou x_1 , pour l'estimation de s_2 .

lité supplémentaire) principalement pour avoir de l'information sur le signal utile et non sur le bruit.

Toutefois, puisque la qualité de x_2 est faible, car $\alpha_2 < \alpha_1$, x_1 conduit à une estimation plus précise de s_2 que x_2 pour de fortes corrélations du signal (bien que x_1 n'observe pas s_2). En effet, $\mathbf{M}_{p,p}(s_2|x_1) \leq \mathbf{M}_{p,p}(s_2|x_2)$ si $|\rho_s| \geq \sqrt{\frac{\alpha_2(1+\alpha_1)}{\alpha_1(1+\alpha_2)}}$, où $\mathbf{M}_{p,p}(s_2|x_1)$ est l'EQM de l'estimation de s_2 en fonction de x_1 :

$$\mathbf{M}_{p,p}(s_2|x_1) = \frac{\alpha_1(1 - \rho_s^2) + 1}{(1 + \alpha_1)}. \quad (6.22)$$

La figure 6.6 illustre, en fonction de ρ_s , $\mathbf{M}_{p,p}(s_2|x_1)$ (---) et $\mathbf{M}_{p,p}(s_2|x_2)$ pour différentes valeurs de α_2 : $\alpha_2 = \alpha_1$ (-.-.-), $\alpha_2 < \alpha_1$ (-.-) et $\alpha_2 \ll \alpha_1$ (.....).

$\alpha_2 = \alpha_1$: Nous remarquons que $\mathbf{M}_{p,p}(s_2|x_1) \geq \mathbf{M}_{p,p}(s_2|x_2)$ avec égalité si et seulement si $\rho_s = \pm 1$. C'est-à-dire que x_1 est aussi précise que x_2 pour l'estimation de s_2 si et seulement si $\rho_s = \pm 1$, car dans ce cas s_1 et s_2 sont équivalents.

$\alpha_2 < \alpha_1$: Pour de fortes corrélations du signal, x_1 arrive à communiquer plus d'information sur s_2 et elle est plus précise à son estimation que x_2 .

$\alpha_2 \ll \alpha_1$: L'information que x_2 apporte sur s_2 est faible, par conséquent x_1 est souvent plus précise pour l'estimation de s_2 . Notamment, pour $\alpha_2 \rightarrow 0$, $\mathbf{M}_{p,p}(s_2|x_1) \leq \mathbf{M}_{p,p}(s_2|x_2) \forall \rho_s \neq 0$.

Par conséquent, même si le modèle inadéquat est plus précis pour l'estimation de s_2 que la modalité x_2 pour de fortes valeurs de ρ_s , il risque d'être moins précis que la modalité x_1 . En effet, $\mathbf{M}_{p,p}(s_2|x_1) \leq \mathbf{M}_{p,q}(s_2|x_1, x_2)$ si $\rho_n \in I_{une}(s_2|x_1)$, avec :

$$I_{une}(s_2|x_1)|_{\hat{\rho}_n=0} = \begin{cases} \left[-1, \frac{\sqrt{\alpha_2}(1+\alpha_1(1-\rho_s^2))((1+\alpha_1)(1+\alpha_2)-\alpha_1\alpha_2\rho_s^2)}{2\sqrt{\alpha_1}\rho_s(1+\alpha_1)} \right] & \text{si } \rho_s < 0 \\ \left[\frac{\sqrt{\alpha_2}(1+\alpha_1(1-\rho_s^2))((1+\alpha_1)(1+\alpha_2)-\alpha_1\alpha_2\rho_s^2)}{2\sqrt{\alpha_1}\rho_s(1+\alpha_1)}, 1 \right] & \text{si } \rho_s > 0 \end{cases}. \quad (6.23)$$

Nous montrons sur la figure 6.7, les intervalles $I_{une}(s_2|x_2)]_{\hat{\rho}_n=0}$ et $I_{une}(s_2|x_1)]_{\hat{\rho}_n=0}$ en fonction de ρ_s . Nous constatons que l'intervalle $I_{une}(s_2|x_1)]_{\hat{\rho}_n=0}$ a lieu pour de fortes corrélations du signal. Quand les modalités deviennent plus disparates cet intervalle devient plus large.

La modalité x_1 de rapport signal à bruit le plus élevé, peut être plus précise pour l'estimation de s_2 que x_2 et qu'un modèle multimodal inadéquat.

6.4.2 Inadéquation quelconque du bruit

Maintenant, nous examinons l'impact d'une inadéquation quelconque du bruit $\hat{\rho}_n \in [-1, 1]$ sur l'estimation multimodale. L'estimateur multimodal associé aux modalités de $\widehat{\mathcal{M}}_s^2$ sous inadéquation quelconque du bruit est :

$$\hat{s}_{i_q}(x_1, x_2) = \frac{\overbrace{(\sqrt{\alpha_i}(1 + \alpha_j) - \sqrt{\alpha_j}\rho_s(\sqrt{\alpha_1\alpha_2}\rho_s + \hat{\rho}_n))}^{\hat{k}_{i1}} x_i + \overbrace{(\sqrt{\alpha_j}\rho_s - \sqrt{\alpha_i}\hat{\rho}_n)}^{\hat{k}_{i2}} x_j}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1\alpha_2}\rho_s + \hat{\rho}_n)^2}. \quad (6.24)$$

C'est-à-dire que :

$$\hat{s}_{1_q}(x_1, x_2) = \frac{\hat{k}_{11} x_1 + \hat{k}_{12} x_2}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1\alpha_2}\rho_s + \hat{\rho}_n)^2}, \quad (6.25)$$

$$\hat{s}_{2_q}(x_1, x_2) = \frac{\hat{k}_{21} x_2 + \hat{k}_{22} x_1}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1\alpha_2}\rho_s + \hat{\rho}_n)^2}. \quad (6.26)$$

L'EQM associée à $\hat{s}_{i_q}(x_1, x_2)$ est :

$$\mathbf{M}_{p,q}(s_i|x_1, x_2) = \frac{\overbrace{\alpha_j(1 - \rho_s^2) + 1 - \hat{\rho}_n^2}^{\mathbf{M}_{q,q}(s_i|x_1, x_2)}}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1\alpha_2}\rho_s + \hat{\rho}_n)^2} \quad (6.27)$$

$$+ 2(\rho_n - \hat{\rho}_n) \frac{\hat{k}_{i1}\hat{k}_{i2}}{((1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\rho_s)^2)^2},$$

$$= a + b\rho_n. \quad (6.28)$$

$\mathbf{M}_{p,q}(s_i|x_1, x_2)$ est une fonction affine croissante de ρ_n . Le signe de la pente b , est le signe du produit $\hat{k}_{i1}\hat{k}_{i2}$. Par conséquent, si $\text{sgn}(\hat{k}_{i1}) = \text{sgn}(\hat{k}_{i2})$, $\mathbf{M}_{p,q}(s_i|x_1, x_2)$ est une fonction affine croissante de ρ_n . Autrement, elle est décroissante.

Estimation de s_1 . Selon l'hypothèse $\alpha_2 \leq \alpha_1$, \hat{k}_{11} est toujours positif. Ainsi la variation de l'erreur d'estimation inadéquate $\mathbf{M}_{p,q}(s_i|x_1, x_2)$ par rapport à ρ_n dépend uniquement du signe de \hat{k}_{12} . Elle est croissante si $\hat{k}_{12} > 0$ et décroissante si $\hat{k}_{12} < 0$.

Estimation de s_2 . Le facteur de x_2 , \hat{k}_{21} , n'est pas toujours positif. Par conséquent, dans le cas d'estimation de s_2 , la variation de $\mathbf{M}_{p,q}(s_i|x_1, x_2)$ par rapport à ρ_n dépend du signe du produit $\hat{k}_{21}\hat{k}_{22}$.

Pour analyser l'effet d'une inadéquation quelconque du bruit, nous procédons comme précédemment. Nous comparons la performance pour l'estimation de s_i du modèle $\widehat{\mathcal{M}}_s^2$, avec :

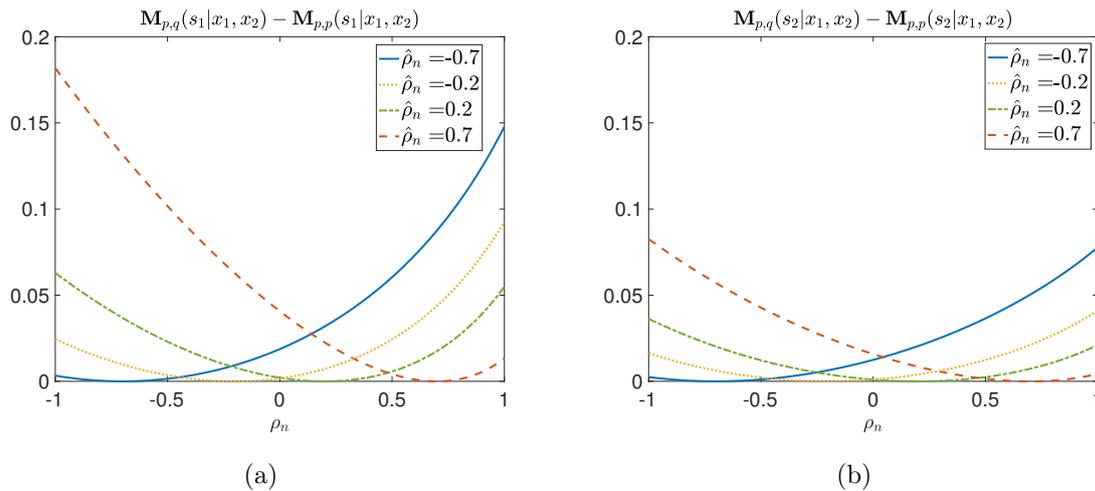


FIGURE 6.8 – $\mathbf{M}_{p,q}(s_1|x_1, x_2) - \mathbf{M}_{p,p}(s_1|x_1, x_2)$ (a) et $\mathbf{M}_{p,q}(s_2|x_1, x_2) - \mathbf{M}_{p,p}(s_2|x_1, x_2)$ (b) pour différentes valeurs de $\hat{\rho}_n$, avec $\rho_s = 0.7$, $\alpha_1 = 5$ et $\alpha_2 = 2$.

1. la performance de deux modalités dont les liens sont parfaitement modélisés,
2. la performance d'une seule modalité x_i .

6.4.2.1 Modèle inadéquat versus modèle adéquat

Nous commençons par quantifier la perte en performance pour l'estimation de s_i du modèle inadéquat $\tilde{\mathcal{M}}_s^2$ par rapport au modèle adéquat \mathcal{M}_s^2 .

Les figures 6.8a et 6.8b montrent les excédents des EQMs relatives pour l'estimation de s_1 et de s_2 , $\mathbf{M}_{p,q}(s_1|x_1, x_2) - \mathbf{M}_{p,p}(s_1|x_1, x_2)$ et $\mathbf{M}_{p,q}(s_2|x_1, x_2) - \mathbf{M}_{p,p}(s_2|x_1, x_2)$, pour différentes valeurs d'inadéquation du bruit.

Nous remarquons que l'excédent de l'EQM est nul pour $\rho_n = \hat{\rho}_n$, car les deux modèles sont égaux, et devient plus large en s'éloignant de cette valeur exacte. Par rapport aux hypothèses de faibles corrélations, les hypothèses de fortes corrélations, *i.e* $|\hat{\rho}_n| = 0.7$ sur la figure, introduisent de larges erreurs car nous accordons à tort confiance à la deuxième modalité. Finalement, l'inadéquation du bruit introduit de plus larges erreurs pour l'estimation de s_1 que pour l'estimation de s_2 car, selon l'hypothèse $\alpha_2 \leq \alpha_1$ x_2 communique plus d'information sur le bruit que x_1 , et par conséquent s_1 bénéficie plus de l'échange de l'information sur le bruit entre les modalités et est plus sensible à l'inadéquation de ρ_n .

L'inadéquation du bruit engendre plus d'erreur pour l'estimation de s_1 dont la mesure, x_1 , a le rapport signal à bruit le plus élevé.

6.4.2.2 Modèle inadéquat versus une modalité

Nous comparons maintenant, la performance du modèle inadéquat $\tilde{\mathcal{M}}_s^2$ à la performance obtenue avec une seule modalité x_i . A partir des EQMs $\mathbf{M}_{p,q}(s_i|x_1, x_2)$ et $\mathbf{M}_{p,p}(s_i|x_i)$, nous trouvons que la modalité x_i seule est autant ou plus précise pour l'estimation de s_i que le modèle

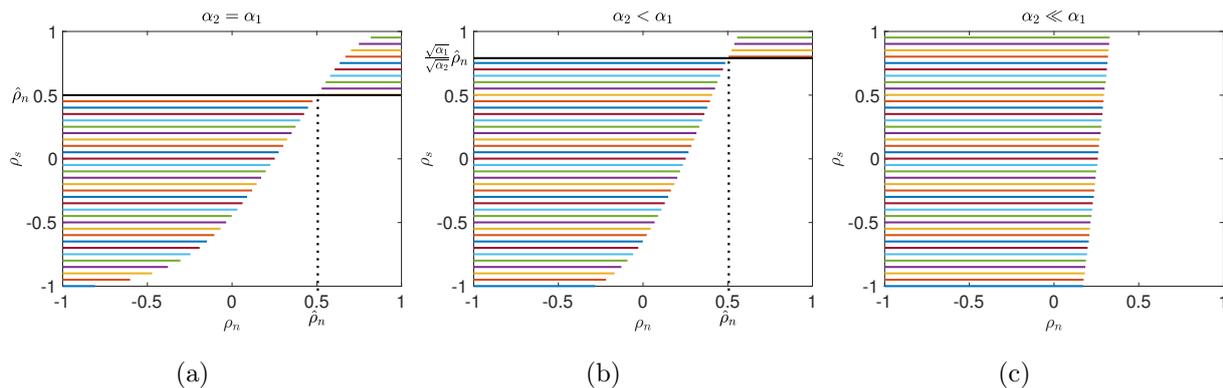


FIGURE 6.9 – Variation de $I_{une}(s_1|x_1)$, en fonction de ρ_s , pour $\alpha_1 = 5$ et différentes valeurs de α_2 : $\alpha_2 = \alpha_1 = 5$ (a), $\alpha_2 = 2$ (b) et $\alpha_2 = 0.1$ (c), avec $\hat{\rho}_n = 0.5$. Chaque segment présente, pour ρ_s donné, les valeurs exactes ρ_n qui, lorsqu'elles sont erronées, engendrent plus d'EQM que la seule modalité x_1 , pour l'estimation de s_1 .

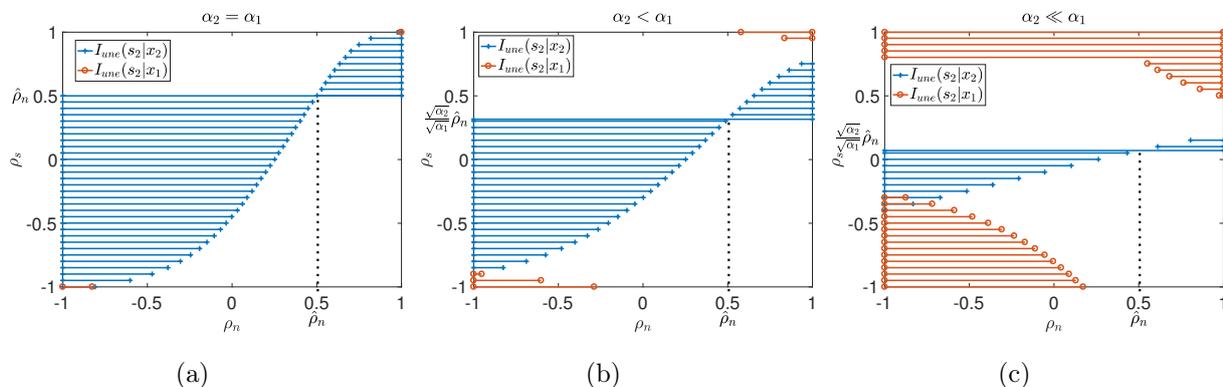


FIGURE 6.10 – Variation de $I_{une}(s_2|x_2)$ et $I_{une}(s_2|x_1)$, en fonction de ρ_s , pour $\alpha_1 = 5$ et différentes valeurs de α_2 , α_2 : $\alpha_2 = \alpha_1 = 5$ (a), $\alpha_2 = 2$ (b) et $\alpha_2 = 0.1$ (c), avec $\hat{\rho}_n = 0.5$. Chaque segment présente, pour ρ_s donné, les valeurs exactes de ρ_n qui, lorsqu'elles sont erronées, engendrent plus d'EQM que la seule modalité x_2 ou x_1 , pour l'estimation de s_2 .

multimodal inadéquat $\widehat{\mathcal{M}}_s^2$ si et seulement si $\rho_n \in I_{une}(s_i|x_i)]_{\hat{\rho}_n}$:

$$I_{une}(s_i|x_i)]_{\hat{\rho}_n} = \begin{cases} [-1, \rho_n^*] & \text{si } \rho_s < \frac{\sqrt{\alpha_i}}{\sqrt{\alpha_j}} \hat{\rho}_n \\ [-1, 1] & \text{si } \rho_s = \frac{\sqrt{\alpha_i}}{\sqrt{\alpha_j}} \hat{\rho}_n \\ [\rho_n^*, 1] & \text{si } \rho_s > \frac{\sqrt{\alpha_i}}{\sqrt{\alpha_j}} \hat{\rho}_n \end{cases} \quad (6.29)$$

où

$$\rho_n^* = \hat{\rho}_n + \frac{\hat{k}_{i2}((1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1 \alpha_2} \hat{\rho}_s)^2)}{2\hat{k}_{i1}(1 + \alpha_i)} \quad (6.30)$$

Les figure 6.9 et 6.10 illustrent la variation des intervalles $I_{une}(s_1|x_1)$ et $I_{une}(s_2|x_2)$, respectivement, selon ρ_s pour différents scénarios des rapports signal à bruit. Nous introduisons l'intervalle $I_{une}(s_2|x_1)$ de la figure 6.10 dans la suite.

- $\rho_s = \frac{\sqrt{\alpha_i}}{\sqrt{\alpha_j}} \hat{\rho}_n$: Nous remarquons que pour $\rho_s = \frac{\sqrt{\alpha_i}}{\sqrt{\alpha_j}} \hat{\rho}_n$ ($\rho_s = \hat{\rho}_n$ si $\alpha_2 = \alpha_1$), $I_{une}(s_1|x_1)]_{\hat{\rho}_n} = [-1, 1]$. En effet, dans ce cas, nous supposons que la modalité x_j est totalement

redondante à x_i sur l'estimation de s_i . C'est-à-dire que nous ne pouvons tirer aucune nouvelle information sur s_i à partir de x_j ayant déjà observé x_i . Ainsi, nous pouvons utiliser la seule modalité x_i pour estimer s_i . Par conséquent, $\hat{s}_{i_q}(x_1, x_2) = \hat{s}_{i_q}(x_i)$ et $\mathbf{M}_{p,q}(s_i|x_1, x_2) = \mathbf{M}_{p,p}(s_i|x_i)$ ($\rho_s = \frac{\sqrt{\alpha_i}}{\sqrt{\alpha_j}}\hat{\rho}_n \Leftrightarrow \hat{k}_{i1} = 0 \Rightarrow b = 0$). Cette condition n'est pas satisfaite si $\alpha_2 \ll \alpha_1$ puisque $|\rho_s| \leq 1 \ll \frac{\sqrt{\alpha_1}}{\sqrt{\alpha_2}}\hat{\rho}_n$.

- $\rho_s \neq \frac{\sqrt{\alpha_i}}{\sqrt{\alpha_j}}\hat{\rho}_n$: Nous constatons que la longueur de $I_{une}(s_1|x_1)]_{\hat{\rho}_n}$ est plus courte, dans ce cas, que pour $\rho_s = \frac{\sqrt{\alpha_i}}{\sqrt{\alpha_j}}\hat{\rho}_n$. En effet, le modèle inadéquat exploite les deux modalités et l'information sur le signal utile fournie par la modalité supplémentaire améliore la performance par rapport à une seule modalité. Effectivement, cet intervalle est plus rétréci pour les fortes valeurs de ρ_s . Cependant, le modèle inadéquat pour $\rho_s \neq \frac{\sqrt{\alpha_i}}{\sqrt{\alpha_j}}\hat{\rho}_n$ n'est plus équivalent à la seule modalité x_i , sur l'intervalle $I_{une}(s_1|x_1)]_{\hat{\rho}_n}$ mais plus mauvais, sauf si $\rho_n = \rho_n^*$.

- **Effet des rapports signal à bruit :**

Estimation de s_1 En comparant les figures 6.9a, 6.9b et 6.9c nous constatons que la longueur de $I_{une}(s_1|x_1)]_{\hat{\rho}_n}$ est plus large pour des modalités de qualités disparates, car nous accordons plus confiance à l'information sur le bruit fournie par la modalité x_2 .

Estimation de s_2 En revanche, à partir des figures 6.10a, 6.10b et 6.10c, nous remarquons que la longueur de $I_{une}(s_2|x_2)]_{\hat{\rho}_n}$ est plus courte pour des modalités de qualités disparates, et devient nulle pour de fortes corrélations du signal. Ce qui veut dire que le modèle inadéquat est tout le temps plus précis que la seule modalité x_2 pour l'estimation de s_2 . En effet, le modèle inadéquat tire plus d'information sur le signal utile à partir de la modalité x_1 et devient moins sensible à l'inadéquation du bruit.

En revanche, puisque la modalité x_1 est parfois plus précise pour l'estimation de s_2 que x_2 , c'est possible qu'elle soit plus précise que le modèle inadéquat. En comparant les erreurs d'estimation $\mathbf{M}_{p,q}(s_2|x_1, x_2)$ et $\mathbf{M}_{p,p}(s_2|x_1)$, nous trouvons que la modalité x_1 est plus précise pour l'estimation de s_2 que le modèle inadéquat \mathcal{M}_s^2 si et seulement si $\rho_n \in I_{une}(s_2|x_1)]_{\hat{\rho}_n}$:

$$I_{une}(s_2|x_1)]_{\hat{\rho}_n} = \begin{cases} \left[-1, \hat{\rho}_n + \frac{[(\alpha_1+1)(\alpha_2+1) - (\sqrt{\alpha_1\alpha_2}\rho_s + \hat{\rho}_n)^2]\hat{k}_{21}}{2\hat{k}_{22}} \right] & \text{si } \rho_s < \frac{\sqrt{\alpha_i}}{\sqrt{\alpha_j}}\hat{\rho}_n \\ \left[\hat{\rho}_n + \frac{[(\alpha_1+1)(\alpha_2+1) - (\sqrt{\alpha_1\alpha_2}\rho_s + \hat{\rho}_n)^2]\hat{k}_{21}}{2\hat{k}_{22}}, 1 \right] & \text{si } \rho_s > \frac{\sqrt{\alpha_i}}{\sqrt{\alpha_j}}\hat{\rho}_n \end{cases}. \quad (6.31)$$

La figure 6.10 montre la variation de l'intervalle $I_{une}(s_2|x_1)]_{\hat{\rho}_n}$ qui a lieu pour de fortes corrélations du signal. En particulier, pour $\rho_s \rightarrow 1$ et $\alpha_2 \ll \alpha_1$, la modalité x_1 est tout le temps plus précise, pour l'estimation de s_2 , que le modèle inadéquat \mathcal{M}_s^2 .

6.5 Inadéquation du signal

Supposons à présent que seul le lien entre les signaux utiles est inadéquat ($\hat{\rho}_n = \rho_n$). Soit $\hat{\rho}_s$ la corrélation supposée entre s_1 et s_2 , qui est différente de la corrélation exacte ρ_s . A partir de

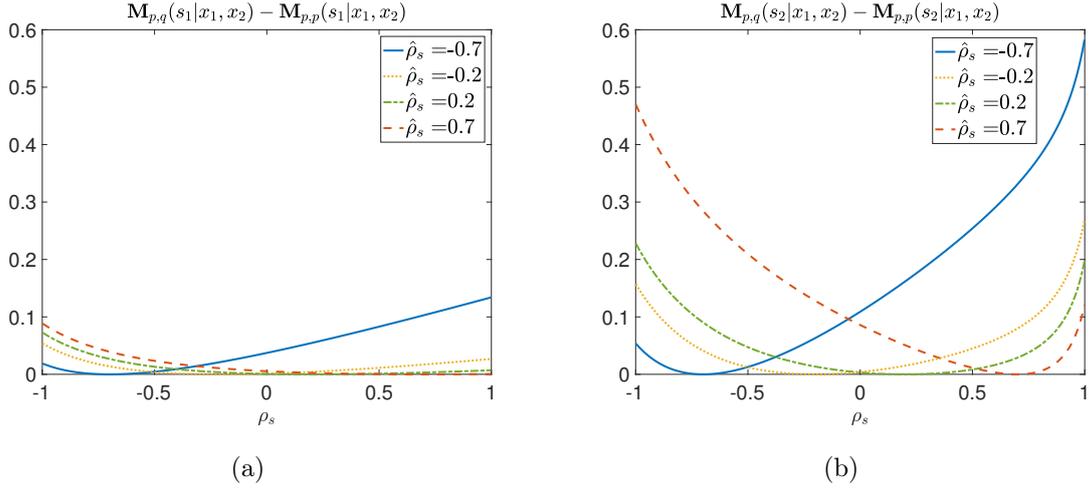


FIGURE 6.11 – $\mathbf{M}_{p,q}(s_1|x_1, x_2) - \mathbf{M}_{p,p}(s_1|x_1, x_2)$ (a) et $\mathbf{M}_{p,q}(s_2|x_1, x_2) - \mathbf{M}_{p,p}(s_2|x_1, x_2)$ (b) pour différentes valeurs de $\hat{\rho}_s$, $\hat{\rho}_s \in \{-0.7, -0.2, 0.2, 0.7\}$, où $\rho_n = 0.7$, $\alpha_1 = 5$ et $\alpha_2 = 2$.

l'estimateur obtenu au paragraphe 6.3, l'estimateur multimodal de s_i sous inadéquation du signal utile est :

$$\hat{s}_{i_q}(x_1, x_2) = \frac{\overbrace{(\sqrt{\alpha_i}(1 + \alpha_j) - \sqrt{\alpha_j}\hat{\rho}_s(\sqrt{\alpha_1\alpha_2}\hat{\rho}_s + \rho_n))}^{\hat{k}_{i1}} x_i + \overbrace{(\sqrt{\alpha_j}\hat{\rho}_s - \sqrt{\alpha_i}\rho_n)}^{\hat{k}_{i2}} x_j}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1\alpha_2}\hat{\rho}_s + \rho_n)^2} \quad (6.32)$$

De même, l'EQM associée à l'estimateur inadéquat $\hat{s}_{i_q}(x_1, x_2)$ est :

$$\begin{aligned} \mathbf{M}_{p,q}(s_i|x_1, x_2) &= \frac{\alpha_j(1 - \hat{\rho}_s^2) + 1 - \rho_n^2}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1\alpha_2}\hat{\rho}_s + \rho_n)^2} \\ &\quad - 2\sqrt{\alpha_j}(\rho_s - \hat{\rho}_s)\hat{k}_{i2} \frac{(1 + \alpha_j) - \rho_n(\rho_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)}{((1 + \alpha_1)(1 + \alpha_2) - (\rho_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)^2)^2}, \\ &= a + c \rho_s. \end{aligned} \quad (6.34)$$

$\mathbf{M}_{p,q}(s_i|x_1, x_2)$ est une fonction affine de ρ_s . Le signe de la pente de $\mathbf{M}_{p,q}(s_i|x_1, x_2)$, c , est le signe du produit $-\hat{k}_{i2}((1 + \alpha_j) - \rho_n(\rho_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s))$.

Dans le cas de l'estimation de s_2 , selon l'hypothèse $\alpha_2 \leq \alpha_1$, $((1 + \alpha_1) - \rho_n(\rho_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s))$ est toujours positif. Ainsi le sens de variation de $\mathbf{M}_{p,q}(s_2|x_1, x_2)$ est déterminé par le signe de $-\hat{k}_{22}$.

6.5.0.1 Modèle inadéquat versus modèle adéquat

Nous comparons la performance du modèle inadéquat, $\widehat{\mathcal{M}}_s^2$, et la performance du modèle adéquat, \mathcal{M}_s^2 , selon les différents types d'inadéquation du signal utile.

Les figures 6.11a et 6.11b montrent les excédents de l'EQM relatives à l'estimation de s_1 et de s_2 respectivement, $\mathbf{M}_{p,q}(s_1|x_1, x_2) - \mathbf{M}_{p,p}(s_1|x_1, x_2)$ et $\mathbf{M}_{p,q}(s_2|x_1, x_2) - \mathbf{M}_{p,p}(s_2|x_1, x_2)$, pour différentes valeurs d'inadéquation du signal utile.

Nous remarquons que, l'excédent de l'EQM est nul pour $\rho_s = \hat{\rho}_s$ et devient plus important en s'éloignant de cette valeur. De plus, les hypothèses de fortes corrélations du signal utile, *i.e* $|\hat{\rho}_s| = 0.7$ sur la figure, engendrent de larges erreurs car nous accordons à tort plus de confiance

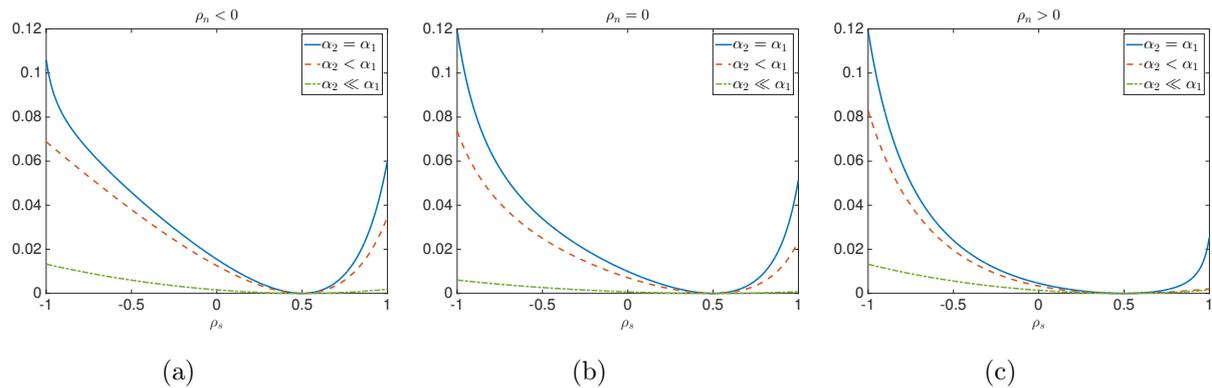


FIGURE 6.12 – $\mathbf{M}_{p,q}(s_1|x_1, x_2) - \mathbf{M}_{p,p}(s_1|x_1, x_2)$, en fonction de ρ_s , pour $\rho_n = -0.7$ (a), $\rho_n = 0$ (b) et $\rho_n = 0.7$ (c), pour $\hat{\rho}_s = 0.5$, $\alpha_1 = 5$ et différentes valeurs de α_2 : $\alpha_2 = \alpha_1 = 5$, $\alpha_2 = 2$ et $\alpha_2 = 0.1$.

à la modalité supplémentaire. Au contraire du cas d'inadéquation du bruit, l'inadéquation de la corrélation du signal engendre plus d'erreur pour l'estimation de s_2 que pour l'estimation de s_1 . En effet, l'estimation de s_2 est plus sensible à l'inadéquation du signal utile car, suivant l'hypothèse $\alpha_2 < \alpha_1$, elle bénéficie le plus de l'échange de l'information sur le signal utile entre les modalités. Réciproquement, puisque s_1 bénéficie le plus de l'échange sur le bruit, elle est plus sensible à l'inadéquation du bruit.

L'inadéquation du signal utile engendre plus d'erreur pour l'estimation de s_2 dont la mesure, x_2 , a le rapport signal à bruit le plus faible.

• **Effet des rapports signal à bruit :**

La figure 6.12 illustre l'excédent de l'EQM en fonction de ρ_s pour $\alpha_1 = 5$ et différentes valeurs du rapport signal à bruit α_2 : $\alpha_2 = \alpha_1$ (—), $\alpha_2 < \alpha_1$ (-.-) et $\alpha_2 \ll \alpha_1$ (-.-.). Les sous-figures 6.12a, 6.12b et 6.12c correspondent à différentes valeurs de ρ_n .

Nous remarquons que pour les différentes valeurs de ρ_n , l'excédent de l'EQM est plus important pour des modalités de même qualité et décroît quand les qualités deviennent disparates, car nous accordons plus de confiance à l'information sur le signal utile lorsque les qualités des modalités sont égales.

Ces résultats sont à l'opposé des résultats obtenus pour l'inadéquation du bruit où l'excédent de l'EQM est plus important pour des modalités de qualités très disparates, puisque dans ce cas nous sommes plus confiant de l'information sur le bruit.

L'excédent de l'EQM, dû à l'inadéquation du signal utile, est plus grand dans le cas de modalités ayant des rapports signal à bruit égaux.

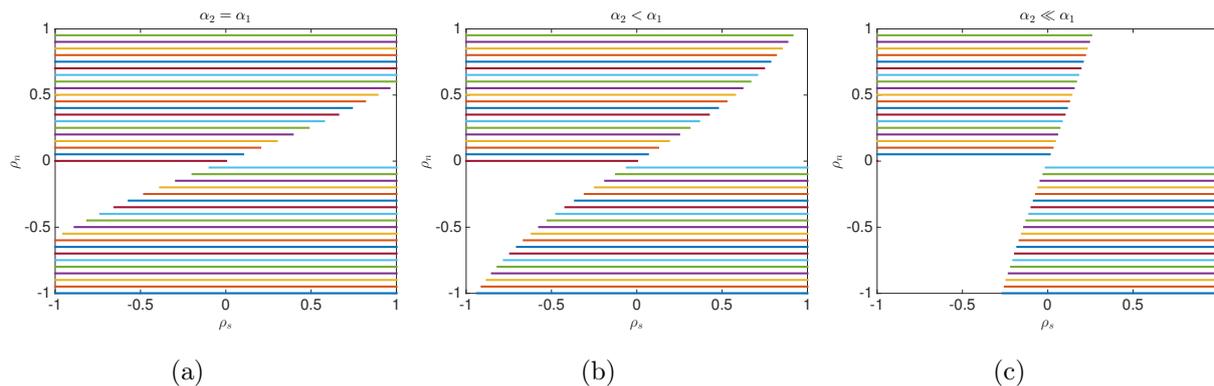


FIGURE 6.13 – Variation de $I_{une}(s_1|x_1)$, en fonction de ρ_s , pour $\alpha_1 = 5$ et différentes valeurs de α_2 : $\alpha_2 = \alpha_1 = 5$ (a), $\alpha_2 = 2$ (b) et $\alpha_2 = 0.1$ (c). Chaque segment présente, pour ρ_n donné, les valeurs exactes de ρ_s qui, lorsqu'elles sont erronées, engendrent plus d'EQM que la seule modalité x_1 , pour l'estimation de s_1 .

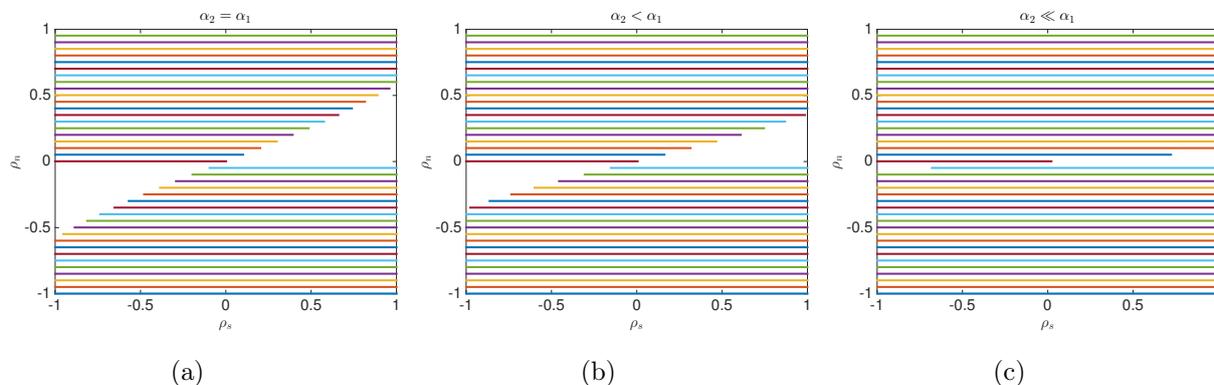


FIGURE 6.14 – Variation de $I_{une}(s_2|x_2)$, en fonction de ρ_s , pour $\alpha_1 = 5$ et différentes valeurs de α_2 : $\alpha_2 = \alpha_1 = 5$ (a), $\alpha_2 = 2$ (b) et $\alpha_2 = 0.1$ (c). Chaque segment présente, pour ρ_n donné, les valeurs exactes de ρ_s qui, lorsqu'elles sont erronées, engendrent plus d'EQM que la seule modalité x_2 , pour l'estimation de s_2 .

6.5.0.2 Modèle inadéquat versus une modalité

Déterminons les conditions sous lesquelles une inadéquation de la corrélation du signal utile engendre plus d'erreur d'estimation qu'une seule modalité. En comparant $\mathbf{M}_{p,q}(s_i|x_1, x_2)$ et $\mathbf{M}_{p,p}(s_i|x_i)$, nous obtenons qu'une seule modalité x_i est plus précise pour l'estimation de s_i qu'un modèle inadéquat si $\rho_s \in I_{une}(s_i|x_i)]_{\hat{\rho}_s}$:

$$I_{une}(s_i|x_i)]_{\hat{\rho}_s} = \begin{cases} [-1, \rho_s^*] & \text{si } \rho_n > \frac{\sqrt{\alpha_i}}{\sqrt{\alpha_j}} \hat{\rho}_s \\ [\rho_s^*, 1] & \text{si } \rho_n > \frac{\sqrt{\alpha_i}}{\sqrt{\alpha_j}} \hat{\rho}_s \end{cases} \quad (6.35)$$

où

$$\rho_s^* = \hat{\rho}_s + \frac{(\sqrt{\alpha_i} \rho_n - \sqrt{\alpha_j} \hat{\rho}_s) ((1 + \alpha_1)(1 + \alpha_2) - (\rho_n + \sqrt{\alpha_1 \alpha_2} \hat{\rho}_s)^2)}{2((1 + \alpha_j) - \rho_n(\rho_n + \sqrt{\alpha_1 \alpha_2} \hat{\rho}_s))(1 + \alpha_i)} \quad (6.36)$$

Les figures 6.13 et 6.14 montrent les variations des intervalles $I_{une}(s_1|x_1)]_{\hat{\rho}_s}$ et $I_{une}(s_2|x_2)]_{\hat{\rho}_s}$ selon les valeurs de ρ_n , pour $\hat{\rho}_s = 0$.

Nous remarquons que, par rapport au cas d'inadéquation du bruit, le modèle inadéquat est plus souvent moins précis qu'une seule modalité. C'est-à-dire que l'inadéquation du signal engendre plus d'erreur que l'inadéquation du bruit, ce qui est logique puisque, pour l'estimation, l'information sur le signal utile est plus importante que celle sur le bruit.

Estimation de s_1 Nous constatons, sur la figure 6.13, que la longueur de $I_{une}(s_1|x_1)|_{\hat{\rho}_s}$ est plus courte pour des modalités de qualité disparates, car x_2 apporte plus d'information sur le bruit ce qui permet souvent d'améliorer $\hat{s}_{1_q}(x_1, x_2)$.

Estimation de s_2 . En revanche, nous remarquons sur la figure 6.14 que la seule modalité x_2 est souvent plus précise que le modèle inadéquat dans le cas de modalités avec des qualités très disparates. Puisque, dans ce cas, *i.e* $\alpha_2 \ll \alpha_1$, l'estimation multimodale accorde plus de confiance à l'information utile qu'apporte la modalité x_1 (ayant le rapport signal à bruit le plus élevé) à travers la corrélation du signal, qui est inadéquate.

Dans le cas de modalités avec des qualités très disparates, *i.e* $\alpha_2 \ll \alpha_1$, le modèle inadéquat engendre souvent plus d'erreur pour l'estimation de s_2 , que la seule modalité x_2 .

6.6 Double inadéquation

Dans cette dernière partie de ce chapitre, nous étudions le cas de double inadéquation du modèle. Nous supposons que la corrélation des bruits et la corrélation des signaux utiles sont inadéquates, $\hat{\rho}_n \neq \rho_n$ et $\hat{\rho}_s \neq \rho_s$.

L'estimateur supposé optimal sous double inadéquation s'écrit :

$$\hat{s}_{i_q}(x_1, x_2) = \frac{(\sqrt{\alpha_i}(1 + \alpha_j) - \sqrt{\alpha_j}\hat{\rho}_s(\sqrt{\alpha_1\alpha_2}\hat{\rho}_s + \hat{\rho}_n)) x_i - \overbrace{(\sqrt{\alpha_i}\hat{\rho}_n - \sqrt{\alpha_j}\hat{\rho}_s)}^{\hat{k}_{i2}} x_j}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1\alpha_2}\hat{\rho}_s + \hat{\rho}_n)^2} \quad (6.37)$$

L'EQM associée à l'estimateur inadéquat $\hat{s}_{i_q}(x_1, x_2)$ est :

$$\begin{aligned} \mathbf{M}_{p,q}(s_i|x_1, x_2) &= \frac{\alpha_j(1 - \hat{\rho}_s^2) + 1 - \hat{\rho}_n^2}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1\alpha_2}\hat{\rho}_s + \hat{\rho}_n)^2} \\ &\quad + 2\sqrt{\alpha_j}(\rho_s - \hat{\rho}_s)\hat{k}_{i2} \frac{(1 + \alpha_j) - \hat{\rho}_n(\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)}{((1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)^2)^2} \\ &\quad - 2(\rho_n - \hat{\rho}_n)\hat{k}_{i2} \frac{(\sqrt{\alpha_i}(1 + \alpha_j) - \sqrt{\alpha_j}\hat{\rho}_s(\sqrt{\alpha_1\alpha_2}\hat{\rho}_s + \hat{\rho}_n))}{((1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\rho_s)^2)^2}, \\ &= a + b\rho_n + c\rho_s. \end{aligned} \quad (6.38)$$

Les deuxième et troisième termes à droite de l'équation (6.38), correspondent aux inadéquations du signal et du bruit respectivement. Ces deux termes sont de signes opposés, si $(1 + \alpha_j) - \hat{\rho}_n(\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s) \geq 0$ et $(\sqrt{\alpha_i}(1 + \alpha_j) - \sqrt{\alpha_j}\hat{\rho}_s(\sqrt{\alpha_1\alpha_2}\hat{\rho}_s + \hat{\rho}_n)) \geq 0$. Ce qui suggère que les deux inadéquations peuvent se compenser.

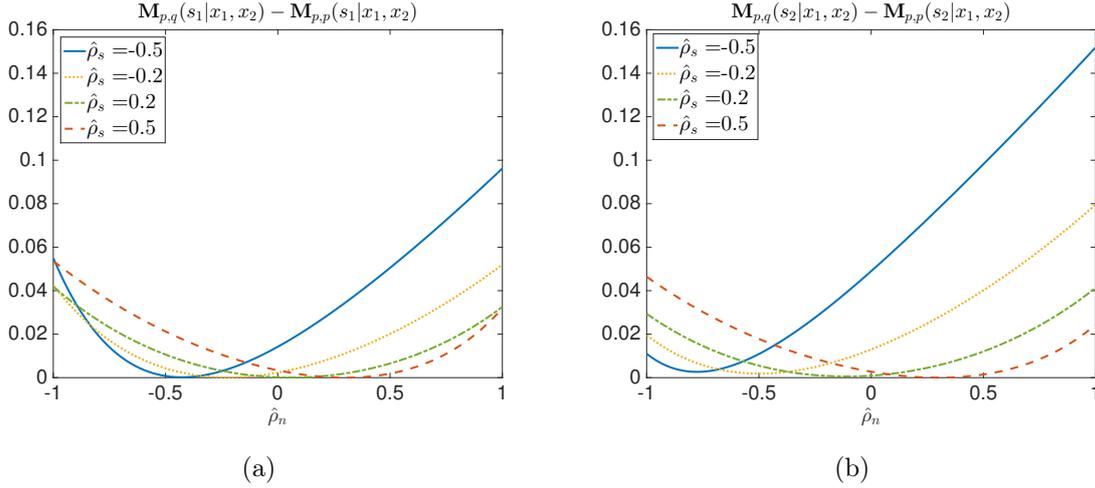


FIGURE 6.15 – $\mathbf{M}_{p,q}(s_1|x_1, x_2) - \mathbf{M}_{p,p}(s_1|x_1, x_2)$ (a) et $\mathbf{M}_{p,q}(s_2|x_1, x_2) - \mathbf{M}_{p,p}(s_2|x_1, x_2)$ (b), en fonction de $\hat{\rho}_n$, pour différentes valeurs de $\hat{\rho}_s$, $\hat{\rho}_s \in \{-0.5, -0.2, 0.2, 0.5\}$, où $\rho_s = 0.5$, $\rho_n = 0.3$, $\alpha_1 = 5$ et $\alpha_2 = 2$.

Les figures 6.15a et 6.15b illustrent les excédents des EQMs $\mathbf{M}_{p,q}(s_1|x_1, x_2) - \mathbf{M}_{p,p}(s_1|x_1, x_2)$ et $\mathbf{M}_{p,q}(s_2|x_1, x_2) - \mathbf{M}_{p,p}(s_2|x_1, x_2)$ respectivement, en fonction de $\hat{\rho}_n$ pour des corrélations exactes fixées, $\rho_s = 0.5$ et $\rho_n = 0.3$, et différentes valeurs de $\hat{\rho}_s$.

Nous remarquons que pour $\hat{\rho}_s = \rho_s$, les excédents des EQMs s'annulent pour $\hat{\rho}_n = \rho_n$. En revanche, pour les autres valeurs de $\hat{\rho}_s \neq \rho_s$, $\hat{\rho}_s \in \{-0.5, -0.2, 0.2\}$, les excédents des EQMs atteignent leurs valeurs minimales, non pour $\hat{\rho}_n = \rho_n$, mais pour d'autres inadéquations du bruit. C'est-à-dire que deux inadéquations du modèle peuvent engendrer moins d'erreurs qu'une seule inadéquation.

Nous concluons alors qu'une inadéquation du bruit permet de compenser une inadéquation du signal, et vice versa. Ceci est dû aux effets opposés des corrélations des bruits et des corrélations du signal utile sur l'interaction des modalités.

Deux inadéquations de modèle peuvent engendrer moins d'erreur qu'une seule inadéquation.

6.7 Conclusion

Dans ce chapitre, nous avons considéré différents types d'inadéquations du modèle : l'inadéquation du bruit et/ou l'inadéquation du signal utile. Les résultats montrent que les inadéquations peuvent détériorer l'estimation multimodale et la rendre plus mauvaise que l'estimation fondée sur une seule modalité.

De façon plus surprenante, ce chapitre met en évidence qu'une double inadéquation, c'est-à-dire une inadéquation du signal utile et des bruits permet d'améliorer la précision de l'estimation. Les deux inadéquations peuvent se compenser partiellement ou complètement, conduisant à des erreurs d'estimation minimales !

Conclusions et perspectives

Conclusions

Dans cette thèse, nous avons remis en question le gain réel de la multimodalité. Nous avons examiné, à travers des exemples simples, les conditions sous lesquelles l'analyse multimodale est effectivement bénéfique. A la différence des travaux existants, nous avons considéré la corrélation des bruits et nous avons étudié l'interaction des modalités non seulement dans le cas de modèles adéquats, mais également dans le cas de modèles inadéquats.

Afin de tirer des conclusions selon deux points de vue, nous avons choisi les mesures de l'information et de l'estimation pour l'analyse des modèles considérés. Il existe dans la littérature des relations simples établies entre ces deux théories, dans un canal gaussien, même sous inadéquation du signal utile. Nous avons proposé, dans le chapitre 4, des généralisations de certaines relations au cas d'inadéquation du bruit et au cas de double inadéquation. Les relations que nous avons établies mettent en évidence que, ces deux théories ne sont plus directement liées sous inadéquation du bruit.

La partie I de ce manuscrit concerne les modèles adéquats. Nous y avons étudié l'interaction des modalités à travers des liens bien modélisés entre les signaux d'intérêts et entre les bruits. Nos résultats montrent que deux modalités couplées à travers un lien unique (entre les signaux utiles ou entre les bruits) sont, conjointement, toujours plus performantes en termes d'information et de précision d'estimation qu'une seule modalité.

En revanche, dans certains cas, deux modalités avec deux liens, fournissent conjointement moins d'information que deux modalités avec un lien unique. D'ailleurs, elles peuvent être équivalentes à une seule modalité. Dans d'autres cas, ces deux modalités assurent de très bonnes performances, notamment lorsque les liens sont forts, et elles peuvent même complètement rétablir le signal d'intérêt.

Par conséquent, nous avons conclu que, dans le cas de modèles adéquats, la multimodalité est souvent plus bénéfique qu'une seule modalité et nous avons établi, pour les modèles considérés, les conditions sur les liens et les rapports signal à bruits pour tirer le maximum d'information possible de l'analyse multimodale.

L'adéquation du modèle supposé au modèle des données est peu probable en pratique. En particulier, la modélisation des liens entre les modalités est complexe et est sujette à erreur. C'est la raison pour laquelle nous avons consacré la partie II de ce manuscrit aux modèles inadéquats, où nous avons étudié l'interaction des modalités en utilisant des liens erronés. Nous avons considéré l'effet des liens mal estimés et l'effet des liens non utilisés (ignorés), notamment l'hypothèse de bruits indépendants.

Nos résultats montrent que, les fausses hypothèses sur les liens entre les modalités engendrent des erreurs qui risquent de fausser le but de la multimodalité. En effet, un modèle multimodal inadéquat risque d'être plus mauvais qu'une seule modalité, même que celle de mauvaise qualité (rapport signal à bruit le plus faible). La multimodalité est, donc bénéfique tant que les liens entre les modalités sont parfaitement modélisés. Autrement, il est recommandé d'utiliser une seule modalité.

Pourtant, ce qui est surprenant est que, l'interaction des modalités à travers des faux liens entre les bruits et entre les signaux utiles (double inadéquation) peut engendrer moins d'erreur que deux modalités avec une seule inadéquation. Ce qui signifie que, les deux inadéquations peuvent se compenser et que le modèle inadéquat peut être équivalent au modèle adéquat.

Dans ce travail, à travers des modèles simples, nous avons tirés d'importants résultats qui infirment certaines idées admises. Contrairement à ce que nous pouvons croire, le lien entre les bruits des modalités permet souvent d'améliorer la performance de la multimodalité. D'ailleurs, en négligeant l'existence de ce lien nous risquons de commettre de graves erreurs. De ce fait, une modalité de très faible qualité est aussi bénéfique, puisqu'elle peut être utilisée comme une référence de bruit et ainsi permettre de supprimer très efficacement le bruit sur l'autre modalité. Finalement, nous avons souligné l'intérêt éventuel d'une double inadéquation à corriger une seule inadéquation de modèle et assurer, donc, une bonne performance.

Cette thèse, bien que limitée à des modèles simples, constitue un guide pour les utilisateurs de la multimodalité. L'analyse détaillée que nous avons conduite et ses différents résultats montrent les conditions sous lesquelles l'analyse multimodale est bénéfique et souligne ainsi les causes conduisant possiblement à une dégradation de performances. Elle nous avise, également, de l'importance de bien modéliser les liens entre les modalités.

Perspectives

Les questions abordées dans cette thèse ont été jusqu'alors peu considérées, en particulier dans la communauté signal et images. Nos résultats sont parmi les premiers et ouvrent de nombreuses perspectives.

Quelques unes des éventuelles perspectives consisteraient à :

- considérer d'autres modèles qui reflètent d'autres caractéristiques des modalités, telles que le retard.
- étudier l'impact de d'autres inadéquations sur la performance multimodale, telles que l'inadéquation de la matrice du canal ou l'inadéquation du retard estimé entre deux

modalités, etc. Dans notre équipe, nous avons déjà obtenu des résultats préliminaires sur l'impact du retard sur l'estimation multimodale.

- proposer d'autres approches qui permettent d'étudier des modèles plus complexes et de généraliser facilement les résultats obtenus à plusieurs modalités.
- étendre les relations que nous avons établies entre la théorie de l'estimation et la théorie de l'information, sous double inadéquation du modèle, à d'autres canaux de communication.
- vérifier l'intérêt de la corrélation des bruits et de la double inadéquation sur des modèles non gaussien.

Annexes

Modèle adéquat : Calcul d'estimation

Dans cette annexe nous détaillons les calculs d'estimation. Nous commençons tout d'abord par un modèle gaussien général. Les résultats obtenus sont ensuite appliqués aux différents modèles que nous étudions dans ce manuscrit.

A.1 Modèle gaussien général

Dans le cas d'un modèle gaussien, l'estimateur optimal au sens de l'erreur quadratique moyenne et l'erreur associée sont donnés par le théorème suivant :

Théorème A.1: (Estimateur optimal - Modèle gaussien [45, ch. 10])

Soit le modèle linéaire suivant :

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n} \quad (\text{A.1})$$

où $\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma})$ est le signal d'intérêt, $\mathbf{x} \in \mathbb{R}^k$ est le vecteur d'observations, $\mathbf{n} \sim \mathcal{N}(0, \mathbf{\Sigma})$ est le bruit gaussien additif et $\mathbf{H} \in \mathbb{R}^{k \times l}$ est la matrice du canal. Nous supposons que \mathbf{s} est indépendante du bruit \mathbf{n} . L'estimateur optimal de \mathbf{s} au sens de l'erreur quadratique moyenne est donné par la moyenne de la loi a posteriori (Théorème 1.2.1), soit :

$$\hat{\mathbf{s}}_p = (\mathbf{\Gamma}^{-1} + \mathbf{H}^T \mathbf{\Sigma}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{\Sigma}^{-1} \mathbf{x} \quad (\text{A.2})$$

$$= \mathbf{\Gamma} \mathbf{H}^T (\mathbf{H} \mathbf{\Gamma} \mathbf{H}^T + \mathbf{\Sigma})^{-1} \mathbf{x} \quad (\text{A.3})$$

Et l'erreur quadratique moyenne associée est :

$$\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x}) = (\mathbf{H}^T \mathbf{\Sigma}^{-1} \mathbf{H} + \mathbf{\Gamma}^{-1})^{-1} \quad (\text{A.4})$$

$$= \mathbf{\Gamma} - \mathbf{\Gamma} \mathbf{H}^T (\mathbf{H} \mathbf{\Gamma} \mathbf{H}^T + \mathbf{\Sigma})^{-1} \mathbf{H} \mathbf{\Gamma} \quad (\text{A.5})$$

Démonstration. D'après les hypothèses précitées, alors $p(\mathbf{x}|\mathbf{s}) = \mathcal{N}(\mathbf{H}\mathbf{s}, \Sigma)$, $p(\mathbf{s}) = \mathcal{N}(\mathbf{0}, \Gamma)$ et $p(\mathbf{x}) = \mathcal{N}(\mathbf{0}, \mathbf{H}\Gamma\mathbf{H}^T + \Sigma)$. En utilisant la relation de Bayes, nous obtenons :

$$\begin{aligned} p(\mathbf{s}|\mathbf{x}) &= \frac{p(\mathbf{x}|\mathbf{s})p(\mathbf{s})}{p(\mathbf{x})} \\ &= \left(\frac{(2\pi)^k |\mathbf{H}\Gamma\mathbf{H}^T + \Sigma|}{(2\pi)^{k+l} |\Gamma\Sigma|} \right)^{\frac{1}{2}} \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{H}\mathbf{s})^T \Sigma^{-1}(\mathbf{x} - \mathbf{H}\mathbf{s})\right) \exp\left(-\frac{1}{2}\mathbf{s}^T \Gamma^{-1}\mathbf{s}\right)}{\exp\left(-\frac{1}{2}\mathbf{x}^T (\mathbf{H}\Gamma\mathbf{H}^T + \Sigma)^{-1}\mathbf{x}\right)} \end{aligned}$$

En utilisant la relation (E.2), on a :

$$\begin{aligned} \frac{(2\pi)^k |\mathbf{H}\Gamma\mathbf{H}^T + \Sigma|}{(2\pi)^{k+l} |\Gamma\Sigma|} &= \frac{1}{(2\pi)^l |\mathbf{H}\Gamma\mathbf{H}^T \Sigma^{-1} + \mathbf{I}|^{-1} |\Gamma|} \\ &= \frac{1}{(2\pi)^l |\mathbf{H}^T \Sigma^{-1} \mathbf{H}\Gamma + \mathbf{I}|^{-1} |\Gamma|} \\ &= \frac{1}{(2\pi)^l |\mathbf{H}^T \Sigma^{-1} \mathbf{H} + \Gamma^{-1}|^{-1}} \end{aligned} \quad (\text{A.6})$$

Ensuite,

$$\begin{aligned} &\frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{H}\mathbf{s})^T \Sigma^{-1}(\mathbf{x} - \mathbf{H}\mathbf{s})\right) \exp\left(-\frac{1}{2}\mathbf{s}^T \Gamma^{-1}\mathbf{s}\right)}{\exp\left(-\frac{1}{2}\mathbf{x}^T (\mathbf{H}\Gamma\mathbf{H}^T + \Sigma)^{-1}\mathbf{x}\right)} \\ &= \frac{\exp\left(-\frac{1}{2}\mathbf{s}^T (\Gamma^{-1} + \mathbf{H}^T \Sigma^{-1} \mathbf{H})\mathbf{s} + \frac{1}{2}\mathbf{x}^T \Sigma^{-1} \mathbf{H}\mathbf{s} + \frac{1}{2}\mathbf{s}^T \mathbf{H}^T \Sigma^{-1} \mathbf{x}\right)}{\exp\left(-\frac{1}{2}\mathbf{x}^T \left((\mathbf{H}\Gamma\mathbf{H}^T + \Sigma)^{-1} - \Sigma^{-1}\right)\mathbf{x}\right)} \\ &= \frac{\exp\left(-\frac{1}{2}\mathbf{s}^T (\Gamma^{-1} + \mathbf{H}^T \Sigma^{-1} \mathbf{H})\mathbf{s} + \frac{1}{2}\mathbf{x}^T \Sigma^{-1} \mathbf{H}\mathbf{s} + \frac{1}{2}\mathbf{s}^T \mathbf{H}^T \Sigma^{-1} \mathbf{x}\right)}{\exp\left(\frac{1}{2}\mathbf{x}^T \left(\Sigma^{-1} \mathbf{H}^T (\Gamma^{-1} + \mathbf{H}^T \Sigma^{-1} \mathbf{H})^{-1} \mathbf{H}\Sigma^{-1}\right)\mathbf{x}\right)} \end{aligned} \quad (\text{A.7})$$

$$\begin{aligned} &= \exp\left(-\frac{1}{2}(\mathbf{s} - (\Gamma^{-1} + \mathbf{H}^T \Sigma^{-1} \mathbf{H})^{-1} \mathbf{H}\Sigma^{-1} \mathbf{x})^T (\Gamma^{-1} + \mathbf{H}^T \Sigma^{-1} \mathbf{H})(\mathbf{s} - (\Gamma^{-1} + \mathbf{H}^T \Sigma^{-1} \mathbf{H})^{-1} \mathbf{H}\Sigma^{-1} \mathbf{x})\right) \\ &= \exp\left(-\frac{1}{2}(\mathbf{s} - \Gamma\mathbf{H}^T (\mathbf{H}\Gamma\mathbf{H}^T + \Sigma)^{-1} \mathbf{x})^T (\Gamma^{-1} + \mathbf{H}^T \Sigma^{-1} \mathbf{H})(\mathbf{s} - \Gamma\mathbf{H}^T (\mathbf{H}\Gamma\mathbf{H}^T + \Sigma)^{-1} \mathbf{x})\right) \end{aligned} \quad (\text{A.8})$$

Les équations (A.7) et (A.8) s'obtiennent en appliquant respectivement (E.3) et (E.4).

Finalement, à partir de (A.6) et (A.8) nous obtenons :

$$\begin{aligned} p(\mathbf{s}|\mathbf{x}) &= \left(\frac{1}{(2\pi)^l |\mathbf{H}^T \Sigma^{-1} \mathbf{H} + \Gamma^{-1}|^{-1}} \right)^{\frac{1}{2}} \times \\ &\quad \exp\left(-\frac{1}{2}(\mathbf{s} - \Gamma\mathbf{H}^T (\mathbf{H}\Gamma\mathbf{H}^T + \Sigma)^{-1} \mathbf{x})^T (\Gamma^{-1} + \mathbf{H}^T \Sigma^{-1} \mathbf{H})(\mathbf{s} - \Gamma\mathbf{H}^T (\mathbf{H}\Gamma\mathbf{H}^T + \Sigma)^{-1} \mathbf{x})\right) \end{aligned} \quad (\text{A.9})$$

Puisque $p(\mathbf{s}|\mathbf{x})$ est gaussienne, sa moyenne est $\mathbb{E}_p[\mathbf{s}|\mathbf{x}] = \Gamma\mathbf{H}^T (\mathbf{H}\Gamma\mathbf{H}^T + \Sigma)^{-1} \mathbf{x}$ et sa variance est $(\mathbf{H}^T \Sigma^{-1} \mathbf{H} + \Gamma^{-1})^{-1}$. Ainsi d'après le théorème 1.2.1, l'estimateur optimale au sens de l'erreur quadratique et l'erreur associée s'écrivent :

$$\hat{\mathbf{s}}_p(\mathbf{x}) = (\Gamma^{-1} + \mathbf{H}^T \Sigma^{-1} \mathbf{H})^{-1} \mathbf{H}\Sigma^{-1} \mathbf{x} \quad (\text{A.10})$$

$$= \mathbf{\Gamma} \mathbf{H}^T (\mathbf{H} \mathbf{\Gamma} \mathbf{H}^T + \mathbf{\Sigma})^{-1} \mathbf{x} \quad (\text{A.11})$$

$$\mathbf{M}_{p,p}(s|\mathbf{x}) = (\mathbf{H}^T \mathbf{\Sigma}^{-1} \mathbf{H} + \mathbf{\Gamma}^{-1})^{-1} \quad (\text{A.12})$$

$$= \mathbf{\Gamma} - \mathbf{\Gamma} \mathbf{H}^T (\mathbf{H} \mathbf{\Gamma} \mathbf{H}^T + \mathbf{\Sigma})^{-1} \mathbf{H} \mathbf{\Gamma} \quad (\text{A.13})$$

□

A.2 Une modalité scalaire

Soit la modalité scalaire :

$$x = \sqrt{\alpha} s + n \quad (\text{A.14})$$

où $\alpha \in \mathbb{R}^+$, s et n sont des variables gaussiennes centrées réduites indépendantes. L'estimateur optimal au sens de l'erreur quadratique moyenne de s en fonction de x est :

$$\hat{s}_p(x) = \frac{\sqrt{\alpha}}{1 + \alpha} x \quad (\text{A.15})$$

L'erreur d'estimation associée à $\hat{s}_p(x)$ est :

$$\mathbf{M}_{p,p}(s|x) = \frac{1}{1 + \alpha} \quad (\text{A.16})$$

Démonstration. Les formules (A.15) et (A.16) sont obtenues en appliquant le théorème A.1 avec $\mathbf{H} = \alpha$ et $\mathbf{\Gamma} = \mathbf{\Sigma} = 1$. □

A.3 Deux modalités scalaires faiblement couplées

Soit le modèle de deux modalités :

$$\begin{cases} x_1 = \sqrt{\alpha_1} s_1 + n_1 \\ x_2 = \sqrt{\alpha_2} s_2 + n_2 \end{cases}, \quad (\text{A.17})$$

où α_1, α_2 sont des réels non négatifs. s_1, s_2, n_1 et n_2 sont des variables gaussiennes centrées réduites. s_1 et s_2 sont indépendants de n_1 et n_2 . Alors que, s_1 et s_2 (resp. n_1 et n_2) sont corrélés, nous notons leurs coefficients de corrélation par ρ_s (resp. ρ_n).

L'estimateur multimodal de s_1 s'écrit :

$$\hat{s}_{i_p}(\mathbf{x}) = \frac{(\alpha_j \sqrt{\alpha_i} (1 - \rho_s^2) + (\sqrt{\alpha_i} - \sqrt{\alpha_j} \rho_n \rho_s)) x_1 - (\sqrt{\alpha_i} \rho_n - \sqrt{\alpha_j} \rho_s) x_2}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n)^2} \quad (\text{A.18})$$

L'erreur d'estimation de s_1 en fonction de x_1 et x_2 est :

$$\mathbf{M}_{p,p}(s_1|\mathbf{x}) = \frac{\alpha_2(1 - \rho_s^2) + 1 - \rho_n^2}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n)^2} \quad (\text{A.19})$$

L'estimateur de s_2 et l'erreur d'estimation associée peuvent être obtenus à partir de (A.18) et (A.19) en permutant α_1 et α_2 , et x_1 et x_2 , respectivement.

Démonstration. Nous posons $\mathbf{x} = (x_1, x_2)^T$, $\mathbf{s} = (s_1, s_2)^T$ et $\mathbf{n} = (n_1, n_2)^T$. Soient, $\mathbf{\Gamma} = \begin{bmatrix} 1 & \rho_s \\ \rho_s & 1 \end{bmatrix}$, $\mathbf{\Sigma} = \begin{bmatrix} 1 & \rho_n \\ \rho_n & 1 \end{bmatrix}$ et $\mathbf{H} = \begin{bmatrix} \sqrt{\alpha_1} & 0 \\ 0 & \sqrt{\alpha_2} \end{bmatrix}$. Par conséquence, en appliquant le théorème A.1, l'estimateur optimale de \mathbf{s} en fonction de \mathbf{x} est :

$$\begin{aligned}
\hat{\mathbf{s}}_p(\mathbf{x}) &= \mathbf{\Gamma H}^T (\mathbf{H \Gamma H}^T + \mathbf{\Sigma})^{-1} \mathbf{x} \\
&= \begin{bmatrix} 1 & \rho_s \\ \rho_s & 1 \end{bmatrix} \begin{bmatrix} \sqrt{\alpha_1} & 0 \\ 0 & \sqrt{\alpha_2} \end{bmatrix} \begin{bmatrix} \alpha_1 + 1 & \sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n \\ \sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n & \alpha_2 + 1 \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&= \begin{bmatrix} \sqrt{\alpha_1} & \sqrt{\alpha_2} \rho_s \\ \sqrt{\alpha_1} \rho_s & \sqrt{\alpha_2} \end{bmatrix} \frac{\begin{bmatrix} \alpha_2 + 1 & -\sqrt{\alpha_1 \alpha_2} \rho_s - \rho_n \\ -\sqrt{\alpha_1 \alpha_2} \rho_s - \rho_n & \alpha_1 + 1 \end{bmatrix}}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n)^2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&= \frac{\begin{bmatrix} \sqrt{\alpha_1}(\alpha_2 + 1) - \sqrt{\alpha_2} \rho_s (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n) & \sqrt{\alpha_2} \rho_s (\alpha_1 + 1) - \sqrt{\alpha_1} (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n) \\ \sqrt{\alpha_1} \rho_s (\alpha_2 + 1) - \sqrt{\alpha_2} (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n) & \sqrt{\alpha_2} (\alpha_1 + 1) - \sqrt{\alpha_1} \rho_s (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n) \end{bmatrix}}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n)^2} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&= \frac{\begin{bmatrix} (\alpha_2 \sqrt{\alpha_1} (1 - \rho_s^2) + (\sqrt{\alpha_1} - \sqrt{\alpha_2} \rho_n \rho_s)) x_1 - (\sqrt{\alpha_1} \rho_n - \sqrt{\alpha_2} \rho_s) x_2 \\ (\alpha_1 \sqrt{\alpha_2} (1 - \rho_s^2) + (\sqrt{\alpha_2} - \sqrt{\alpha_1} \rho_n \rho_s)) x_2 - (\sqrt{\alpha_2} \rho_n - \sqrt{\alpha_1} \rho_s) x_1 \end{bmatrix}}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n)^2} \tag{A.20}
\end{aligned}$$

L'erreur d'estimation relative à $\hat{\mathbf{s}}_p(\mathbf{x})$ est :

$$\begin{aligned}
\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x}) &= \mathbf{\Gamma} - \mathbf{\Gamma H}^T (\mathbf{H \Gamma H}^T + \mathbf{\Sigma})^{-1} \mathbf{H \Gamma} = (\mathbf{I} - \mathbf{\Gamma H}^T (\mathbf{H \Gamma H}^T + \mathbf{\Sigma})^{-1} \mathbf{H}) \mathbf{\Gamma} \\
&= \left(\mathbf{I} - \frac{\begin{bmatrix} \alpha_2 \sqrt{\alpha_1} (1 - \rho_s^2) + (\sqrt{\alpha_1} - \sqrt{\alpha_2} \rho_n \rho_s) & \sqrt{\alpha_2} \rho_s - \sqrt{\alpha_1} \rho_n \\ \sqrt{\alpha_1} \rho_s - \sqrt{\alpha_2} \rho_n & \alpha_1 \sqrt{\alpha_2} (1 - \rho_s^2) + (\sqrt{\alpha_2} - \sqrt{\alpha_1} \rho_n \rho_s) \end{bmatrix}}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n)^2} \begin{bmatrix} \sqrt{\alpha_1} & 0 \\ 0 & \sqrt{\alpha_2} \end{bmatrix} \right) \mathbf{\Gamma} \\
&= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{\begin{bmatrix} \alpha_2 \alpha_1 (1 - \rho_n^2) + \sqrt{\alpha_1} (\sqrt{\alpha_1} - \sqrt{\alpha_2} \rho_n \rho_s) & \alpha_2 \rho_s - \sqrt{\alpha_1 \alpha_2} \rho_n \\ \alpha_1 \rho_s - \sqrt{\alpha_1 \alpha_2} \rho_n & \alpha_1 \alpha_2 (1 - \rho_n^2) + \sqrt{\alpha_2} (\sqrt{\alpha_2} - \sqrt{\alpha_1} \rho_n \rho_s) \end{bmatrix}}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n)^2} \right) \mathbf{\Gamma} \\
&= \frac{\begin{bmatrix} (1 - \rho_n^2) + \sqrt{\alpha_2} (\sqrt{\alpha_2} - \sqrt{\alpha_1} \rho_n \rho_s) & \sqrt{\alpha_1 \alpha_2} \rho_n - \alpha_2 \rho_s \\ \sqrt{\alpha_1 \alpha_2} \rho_n - \alpha_1 \rho_s & (1 - \rho_n^2) + \sqrt{\alpha_1} (\sqrt{\alpha_1} - \sqrt{\alpha_2} \rho_n \rho_s) \end{bmatrix}}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n)^2} \begin{bmatrix} 1 & \rho_s \\ \rho_s & 1 \end{bmatrix} \\
&= \frac{\begin{bmatrix} \alpha_2 (1 - \rho_s^2) + 1 - \rho_n^2 & \sqrt{\alpha_1 \alpha_2} \rho_n (1 - \rho_s^2) + \rho_s (1 - \rho_n^2) \\ \sqrt{\alpha_1 \alpha_2} \rho_n (1 - \rho_s^2) + \rho_s (1 - \rho_n^2) & \alpha_1 (1 - \rho_s^2) + 1 - \rho_n^2 \end{bmatrix}}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n)^2} \tag{A.21}
\end{aligned}$$

□

A.4 Deux modalités scalaires fortement couplées

Soit le modèle de deux modalités :

$$\begin{cases} x_1 = \sqrt{\alpha_1}s + n_1 \\ x_2 = \sqrt{\alpha_2}s + n_2 \end{cases}, \quad (\text{A.22})$$

où α_1 et α_2 sont des réels positifs. s , n_1 et n_2 sont des variables gaussiennes centrées réduites. s est indépendant de n_1 et n_2 . Les bruits n_1 et n_2 sont corrélés et nous notons leurs coefficients de corrélation par ρ_n .

L'estimateur multimodal de s s'écrit :

$$\hat{s}_p(\mathbf{x}) = \frac{(\sqrt{\alpha_1} - \sqrt{\alpha_2}\rho_n)x_1 + (\sqrt{\alpha_2} - \sqrt{\alpha_1}\rho_n)x_2}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1\alpha_2} + \rho_n)^2}. \quad (\text{A.23})$$

L'erreur d'estimation de s_1 en fonction de x_1 et x_2 est :

$$\mathbf{M}_{p,p}(s|\mathbf{x}) \frac{1 - \rho_n^2}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1\alpha_2} + \rho_n)^2}. \quad (\text{A.24})$$

Démonstration. Le modèle (A.22) est équivalent au modèle (A.17) lorsque s_1 et s_2 sont égaux, *i.e* $\rho_s = 1$. Par conséquent, les formules (A.23) et (A.24) sont obtenues à partir de (A.18) et (A.19), respectivement, pour $\rho_s = 1$. \square

A.5 Trois modalités scalaires

Soit le modèle :

$$\begin{cases} x_1 = \sqrt{\alpha_1}s + n_1 \\ x_2 = \sqrt{\alpha_2}s + n_2 \\ x_3 = \sqrt{\alpha_3}s + n_3 \end{cases}, \quad (\text{A.25})$$

où $\alpha_i \in \mathbb{R}^+$, $i \in \{1, 2, 3\}$. s , n_1 , n_2 et n_3 sont des variables gaussiennes centrées réduites, *i.e* $\sim \mathcal{N}(0, 1)$. Nous supposons que s est indépendante des bruits alors que les bruits sont corrélés. Nous notons par $\rho_{12} = \frac{\mathbb{E}[n_1 n_2]}{\mathbb{E}[n_1^2]\mathbb{E}[n_2^2]}$, $\rho_{13} = \frac{\mathbb{E}[n_1 n_3]}{\mathbb{E}[n_1^2]\mathbb{E}[n_3^2]}$ et $\rho_{23} = \frac{\mathbb{E}[n_2 n_3]}{\mathbb{E}[n_2^2]\mathbb{E}[n_3^2]}$, les coefficients de corrélation de n_2 et n_1 , n_3 et n_1 et de n_3 et n_2 , respectivement.

L'estimateur optimal de s en fonction de (x_1, x_2, x_3) est :

$$\hat{s}_p(x_1, x_2, x_3) = \frac{k_1 x_1 + k_2 x_2 + k_3 x_3}{|\Omega|}, \quad (\text{A.26})$$

avec :

$$k_1 = (1 + \alpha_3)(\sqrt{\alpha_1} - \sqrt{\alpha_2}\rho_{12}) + (\sqrt{\alpha_1\alpha_3} + \rho_{13})(\sqrt{\alpha_2}\rho_{23} - \sqrt{\alpha_3}), + (\sqrt{\alpha_2\alpha_3} + \rho_{23})(\sqrt{\alpha_3}\rho_{12} - \sqrt{\alpha_1}\rho_{23}),$$

$$k_2 = (1 + \alpha_3)(\sqrt{\alpha_2} - \sqrt{\alpha_1}\rho_{12}) + (\sqrt{\alpha_2\alpha_3} + \rho_{23})(\sqrt{\alpha_1}\rho_{13} - \sqrt{\alpha_3}) - (\sqrt{\alpha_1\alpha_3} + \rho_{13})(\sqrt{\alpha_2}\rho_{13} - \sqrt{\alpha_3}\rho_{12}),$$

$$k_3 = (1 + \alpha_2)(\sqrt{\alpha_3} - \sqrt{\alpha_1}\rho_{13}) + (\sqrt{\alpha_2\alpha_3} + \rho_{23})(\sqrt{\alpha_1}\rho_{12} - \sqrt{\alpha_2}) + (\sqrt{\alpha_1\alpha_2} + \rho_{12})(\sqrt{\alpha_2}\rho_{13} - \sqrt{\alpha_3}\rho_{12}).$$

$$|\mathbf{\Omega}| = (1 + \alpha_1)(1 + \alpha_2)(1 + \alpha_3) + 2(\sqrt{\alpha_1\alpha_2} + \rho_{12})(\sqrt{\alpha_1\alpha_3} + \rho_{13})(\sqrt{\alpha_2\alpha_3} + \rho_{23}) - (1 + \alpha_2)(\sqrt{\alpha_1\alpha_3} + \rho_{13})^2 - (1 + \alpha_3)(\sqrt{\alpha_1\alpha_2} + \rho_{12})^2 - (1 + \alpha_1)(\sqrt{\alpha_2\alpha_3} + \rho_{23})^2. \quad (\text{A.27})$$

L'erreur associée à l'estimateur multimodale $\hat{s}_p(x_1, x_2, x_3)$ est :

$$\mathbf{M}_{p,p}(s|x_1, x_2, x_3) = \frac{|\mathbf{\Sigma}|}{|\mathbf{\Omega}|}, \quad (\text{A.28})$$

où $\mathbf{\Sigma}$ est la matrice de covariance des bruits, dont le déterminant est :

$$|\mathbf{\Sigma}| = 1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}. \quad (\text{A.29})$$

Démonstration. Nous posons $\mathbf{x} = (x_1, x_2, x_3)^T$, $\mathbf{n} = (n_1, n_2, n_3)^T$. En appliquant le théorème A.1, avec $\mathbf{\Gamma} = \mathbf{1}$, $\mathbf{\Sigma} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{bmatrix}$ et $\mathbf{H} = \begin{bmatrix} \sqrt{\alpha_1} \\ \sqrt{\alpha_2} \\ \sqrt{\alpha_3} \end{bmatrix}$, nous obtenons l'estimateur multimodale de s :

$$\begin{aligned} \hat{s}(\mathbf{x}) &= \mathbf{\Gamma}\mathbf{H}^T(\mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \mathbf{\Sigma})^{-1}\mathbf{x}, \\ &= \begin{bmatrix} \sqrt{\alpha_1} \\ \sqrt{\alpha_2} \\ \sqrt{\alpha_3} \end{bmatrix}^T \begin{bmatrix} \alpha_1 + 1 & \sqrt{\alpha_1\alpha_2} + \rho_{12} & \sqrt{\alpha_1\alpha_3} + \rho_{13} \\ \sqrt{\alpha_1\alpha_2} + \rho_{12} & \alpha_2 + 1 & \sqrt{\alpha_2\alpha_3} + \rho_{23} \\ \sqrt{\alpha_1\alpha_3} + \rho_{13} & \sqrt{\alpha_2\alpha_3} + \rho_{23} & \alpha_3 + 1 \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \\ &= \begin{bmatrix} \sqrt{\alpha_1} & \sqrt{\alpha_2} & \sqrt{\alpha_3} \end{bmatrix} \begin{bmatrix} a & d & e \\ d & b & f \\ e & f & c \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \\ &= \begin{bmatrix} \sqrt{\alpha_1} & \sqrt{\alpha_2} & \sqrt{\alpha_3} \end{bmatrix} \frac{\begin{bmatrix} bc - f^2 & fe - dc & df - be \\ fe - dc & ac - e^2 & de - af \\ df - be & de - af & ab - d^2 \end{bmatrix}}{|\mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \mathbf{\Sigma}|} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \\ &= \frac{\begin{bmatrix} \sqrt{\alpha_1}(bc - f^2) + \sqrt{\alpha_2}(fe - dc) + \sqrt{\alpha_3}(df - be) \\ \sqrt{\alpha_1}(fe - dc) + \sqrt{\alpha_2}(ac - e^2) + \sqrt{\alpha_3}(de - af) \\ \sqrt{\alpha_1}(df - be) + \sqrt{\alpha_2}(de - af) + \sqrt{\alpha_3}(ab - d^2) \end{bmatrix}^T}{|\mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \mathbf{\Sigma}|} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \\ &= \frac{\left(\begin{array}{l} [\sqrt{\alpha_1}(bc - f^2) + \sqrt{\alpha_2}(fe - dc) + \sqrt{\alpha_3}(df - be)] x_1 \\ + [\sqrt{\alpha_1}(fe - dc) + \sqrt{\alpha_2}(ac - e^2) + \sqrt{\alpha_3}(de - af)] x_2 \\ + [\sqrt{\alpha_1}(df - be) + \sqrt{\alpha_2}(de - af) + \sqrt{\alpha_3}(ab - d^2)] x_3 \end{array} \right)}{|\mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \mathbf{\Sigma}|}. \end{aligned}$$

Avec :

$$\begin{aligned} &\sqrt{\alpha_1}(bc - f^2) + \sqrt{\alpha_2}(fe - dc) + \sqrt{\alpha_3}(df - be) \\ &= (\alpha_3 + 1)(\sqrt{\alpha_1} - \sqrt{\alpha_2}\rho_{12}) + (\sqrt{\alpha_1\alpha_3} + \rho_{13})(\sqrt{\alpha_2}\rho_{23} - \sqrt{\alpha_3}) + (\sqrt{\alpha_2\alpha_3} + \rho_{23})(\sqrt{\alpha_3}\rho_{12} - \sqrt{\alpha_1}\rho_{23}), \end{aligned}$$

$$\begin{aligned} &\sqrt{\alpha_1}(fe - dc) + \sqrt{\alpha_2}(ac - e^2) + \sqrt{\alpha_3}(de - af) \\ &= (\sqrt{\alpha_2\alpha_3} + \rho_{23})(\sqrt{\alpha_1}\rho_{13} - \sqrt{\alpha_3}) + (\alpha_3 + 1)(\sqrt{\alpha_2} - \sqrt{\alpha_1}\rho_{12}) - (\sqrt{\alpha_1\alpha_3} + \rho_{13})(\sqrt{\alpha_2}\rho_{13} - \sqrt{\alpha_3}\rho_{12}), \end{aligned}$$

$$\begin{aligned} & \sqrt{\alpha_1}(df - be) + \sqrt{\alpha_2}(de - af) + \sqrt{\alpha_3}(ab - d^2) \\ &= (\sqrt{\alpha_2\alpha_3} + \rho_{23})(\sqrt{\alpha_1}\rho_{12} - \sqrt{\alpha_2}) + (\sqrt{\alpha_1\alpha_2} + \rho_{12})(\sqrt{\alpha_2}\rho_{13} - \sqrt{\alpha_3}\rho_{12}) + (\alpha_2 + 1)(\sqrt{\alpha_3} - \sqrt{\alpha_1}\rho_{13}). \end{aligned}$$

$$\begin{aligned} |\mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \mathbf{\Sigma}| &= abc - be^2 - cd^2 - af^2 + 2def = d(fe - cd) + e(df - be) + a(bc - f^2), \\ &= (\alpha_1 + 1)(\alpha_2 + 1)(\alpha_3 + 1) + 2(\sqrt{\alpha_1\alpha_2} + \rho_{12})(\sqrt{\alpha_1\alpha_3} + \rho_{13})(\sqrt{\alpha_2\alpha_3} + \rho_{23}) \\ &\quad - (\alpha_2 + 1)(\sqrt{\alpha_1\alpha_3} + \rho_{13})^2 - (\alpha_3 + 1)(\sqrt{\alpha_1\alpha_2} + \rho_{12})^2 - (\alpha_1 + 1)(\sqrt{\alpha_2\alpha_3} + \rho_{23})^2. \end{aligned}$$

Toujours à partir du théorème A.1, l'erreur d'estimation de s s'écrit :

$$\begin{aligned} \mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x}) &= \mathbf{\Gamma} - \mathbf{\Gamma}\mathbf{H}^T(\mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \mathbf{\Sigma})^{-1}\mathbf{H}\mathbf{\Gamma}, \\ &= \mathbf{\Gamma} - \frac{\begin{bmatrix} \sqrt{\alpha_1}(bc - f^2) + \sqrt{\alpha_2}(fe - dc) + \sqrt{\alpha_3}(df - be) \\ \sqrt{\alpha_1}(fe - dc) + \sqrt{\alpha_2}(ac - e^2) + \sqrt{\alpha_3}(de - af) \\ \sqrt{\alpha_1}(df - be) + \sqrt{\alpha_2}(de - af) + \sqrt{\alpha_3}(ab - d^2) \end{bmatrix}}{|\mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \mathbf{\Sigma}|} \mathbf{H}\mathbf{\Gamma}, \\ &= \frac{|\mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \mathbf{\Sigma}| - \begin{bmatrix} \sqrt{\alpha_1}(bc - f^2) + \sqrt{\alpha_2}(fe - dc) + \sqrt{\alpha_3}(df - be) \\ \sqrt{\alpha_1}(fe - dc) + \sqrt{\alpha_2}(ac - e^2) + \sqrt{\alpha_3}(de - af) \\ \sqrt{\alpha_1}(df - be) + \sqrt{\alpha_2}(de - af) + \sqrt{\alpha_3}(ab - d^2) \end{bmatrix}^T \begin{bmatrix} \sqrt{\alpha_1} \\ \sqrt{\alpha_2} \\ \sqrt{\alpha_3} \end{bmatrix}}{|\mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \mathbf{\Sigma}|}, \\ &= \frac{|\mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \mathbf{\Sigma}| - \left(\alpha_1(bc - f^2) + 2\sqrt{\alpha_1\alpha_2}(fe - dc) + 2\sqrt{\alpha_1\alpha_3}(df - be) + \alpha_2^2(ac - e^2) + 2\sqrt{\alpha_2\alpha_3}(de - af) + \alpha_3^2(ab - d^2) \right)}{|\mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \mathbf{\Sigma}|}, \\ &= \frac{1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}}{|\mathbf{H}\mathbf{\Gamma}\mathbf{H}^T + \mathbf{\Sigma}|}. \end{aligned}$$

□

Modèle adéquat : Calcul d'Information

Cet annexe détaille le calcul d'information. Nous commençons par un modèle gaussien général, ensuite nous appliquons les résultats aux différents modèles étudiés dans ce manuscrit.

B.1 Modèle gaussien général

Soit le modèle linéaire suivant :

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n} \quad (\text{B.1})$$

où $\mathbf{s} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma})$ est le signal d'intérêt, $\mathbf{x} \in \mathbb{R}^k$ est le vecteur d'observations, $\mathbf{n} \sim \mathcal{N}(0, \mathbf{\Sigma})$ est le bruit gaussien additif et $\mathbf{H} \in \mathbb{R}^{k \times l}$ est la matrice du canal. Nous supposons que \mathbf{s} est indépendante du bruit \mathbf{n} . L'information mutuelle entre \mathbf{s} et \mathbf{x} est :

$$I_{p,p}(\mathbf{s}; \mathbf{x}) = \frac{1}{2} \log \left(|\mathbf{\Gamma}| |\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x}_1)|^{-1} \right), \quad (\text{B.2})$$

$$= \frac{1}{2} \log \left(|\mathbf{\Gamma}| |\mathbf{H}^T \mathbf{\Sigma}^{-1} \mathbf{H} + \mathbf{\Gamma}^{-1}| \right). \quad (\text{B.3})$$

Démonstration. A partir de la définition de l'information mutuelle (1.19), nous obtenons :

$$I_{p,p}(\mathbf{s}; \mathbf{x}) = h(\mathbf{s}) - h(\mathbf{s}|\mathbf{x}), \quad (\text{B.4})$$

$$= \frac{1}{2} \log \left((2\pi e)^l |\mathbf{\Gamma}| \right) - \frac{1}{2} \log \left((2\pi e)^l |\mathbf{\Sigma}_{\mathbf{s}|\mathbf{x}}| \right), \quad (\text{B.5})$$

$$= \frac{1}{2} \log \left((2\pi e)^l |\mathbf{\Gamma}| \right) - \frac{1}{2} \log \left((2\pi e)^l |\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})| \right), \quad (\text{B.6})$$

$$= \frac{1}{2} \log \left(|\mathbf{\Gamma}| |\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x})|^{-1} \right), \quad (\text{B.7})$$

$$= \frac{1}{2} \log \left(|\mathbf{\Gamma}| |\mathbf{H}^T \mathbf{\Sigma}^{-1} \mathbf{H} + \mathbf{\Gamma}^{-1}| \right). \quad (\text{B.8})$$

La formule (B.5) est obtenue à partir de la définition de l'entropie d'un vecteur gaussien (paragraphe 1.1.2). L'équation (B.6) s'ensuit puisque $\mathbf{M}_{p,p}(\mathbf{s}|\mathbf{x}) = \mathbb{E}_p[\mathbf{\Sigma}_{\mathbf{s}|\mathbf{x}}]$ (théorème 1.2.1)

et puisque $\Sigma_{s|x}$ est constante selon l'hypothèse de gaussianité. Finalement, le résultat (B.8) est obtenu en utilisant le théorème A.1. \square

B.2 Une modalité scalaire

Soit la modalité scalaire :

$$x = \sqrt{\alpha} s + n, \quad (\text{B.9})$$

où $\alpha \in \mathbb{R}^+$. s et n sont des variables gaussiennes centrées réduites indépendantes, $i.e \sim \mathcal{N}(0, 1)$.

L'information mutuelle entre s et x est :

$$I_{p,p}(s; x) = \frac{1}{2} \log(1 + \alpha) \quad (\text{B.10})$$

Démonstration. L'information mutuelle (B.10) est obtenue en appliquant (B.2) pour $\Gamma = \Sigma = 1$ et $\mathbf{H} = \alpha$ et en utilisant (A.16). \square

B.3 Deux modalités scalaires faiblement couplées

Soit le modèle :

$$\begin{cases} x_1 = \sqrt{\alpha_1} s_1 + n_1 \\ x_2 = \sqrt{\alpha_2} s_2 + n_2 \end{cases}, \quad (\text{B.11})$$

où s_1, s_2, n_1 et n_2 sont des variables gaussiennes centrées réduites. s_1 et s_2 sont indépendants de n_1 et n_2 . Alors que, s_1 et s_2 (resp. n_1 et n_2) sont corrélés, nous notons leurs coefficients de corrélations par ζ_{12} (resp. ρ_{12}).

L'information mutuelle entre $s_i, i \in \{1, 2\}$, et (x_1, x_2) est :

$$I_{p,p}(s_1; (x_1, x_2)) = \frac{1}{2} \log \left(\frac{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1 \alpha_2} \zeta_{12} + \rho_{12})^2}{\alpha_2(1 - \zeta_{12}^2) + 1 - \rho_{12}^2} \right). \quad (\text{B.12})$$

Démonstration. Nous posons $\mathbf{x} = (x_1, x_2)^T$, $\mathbf{s} = (s_1, s_2)^T$ et $\mathbf{n} = (n_1, n_2)^T$. Soient, $\Gamma = \begin{pmatrix} 1 & \zeta_{12} \\ \zeta_{12} & 1 \end{pmatrix}$, $\Sigma = \begin{pmatrix} 1 & \rho_{12} \\ \rho_{12} & 1 \end{pmatrix}$ et $\mathbf{H} = \begin{pmatrix} \sqrt{\alpha_1} & 0 \\ 0 & \sqrt{\alpha_2} \end{pmatrix}$.
La formule (B.12) est obtenue en appliquant (B.2) et en utilisant (A.21). \square

B.4 Deux modalités scalaires fortement couplées

Soit le modèle :

$$\begin{cases} x_1 = \sqrt{\alpha_1} s + n_1 \\ x_2 = \sqrt{\alpha_2} s + n_2 \end{cases}, \quad (\text{B.13})$$

où s, n_1 et n_2 sont des variables gaussiennes centrées réduites. s est indépendante de n_1 et n_2 . Les bruits n_1 et n_2 sont corrélés, nous notons leurs coefficients de corrélations par ρ_{12} .

L'information mutuelle entre s et (x_1, x_2) est :

$$I_{p,p}(s_1; (x_1, x_2)) = \frac{1}{2} \log \left(\frac{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1 \alpha_2} \zeta_{12} + \rho_{12})^2}{\alpha_2(1 - \zeta_{12}^2) + 1 - \rho_{12}^2} \right). \quad (\text{B.14})$$

Démonstration. La formule (B.14) est obtenue à partir de (B.12) pour $\zeta_{12} = 1$, car le modèle (B.13) est un cas particulier de (B.11) où $s_1 = s_2$. \square

B.5 Trois modalités scalaires

Soit le modèle :

$$\begin{cases} x_1 = \sqrt{\alpha_1} s + n_1 \\ x_2 = \sqrt{\alpha_2} s + n_2 \\ x_3 = \sqrt{\alpha_3} s + n_3 \end{cases}, \quad (\text{B.15})$$

où $\alpha_i \in \mathbb{R}^+$, $i \in \{1, 2, 3\}$. s , n_1 , n_2 et n_3 sont des variables gaussiennes centrées réduites. s est indépendante de n_1 , n_2 et n_3 . n_1 , n_2 et n_3 sont corrélés, nous notons leurs coefficients de corrélation par ρ_{12} , ρ_{13} et ρ_{23} , respectivement.

L'information mutuelle entre s et (x_1, x_2, x_3) est :

$$I_{p,p}(s; x_1, x_2, x_3) = \frac{1}{2} \log \left(\frac{((\alpha_1 + 1)(\alpha_2 + 1)(\alpha_3 + 1) + 2(\sqrt{\alpha_1 \alpha_2} + \rho_{12})(\sqrt{\alpha_1 \alpha_3} + \rho_{13})(\sqrt{\alpha_2 \alpha_3} + \rho_{23}) - (\alpha_2 + 1)(\sqrt{\alpha_1 \alpha_3} + \rho_{13})^2 - (\alpha_3 + 1)(\sqrt{\alpha_1 \alpha_2} + \rho_{12})^2 - (\alpha_1 + 1)(\sqrt{\alpha_2 \alpha_3} + \rho_{23})^2)}{1 - \rho_{12}^2 - \rho_{13}^2 - \rho_{23}^2 + 2\rho_{12}\rho_{13}\rho_{23}} \right). \quad (\text{B.16})$$

Démonstration. Nous posons $\mathbf{x} = (x_1, x_2, x_3)^T$, $\mathbf{s} = (s_1, s_2, s_3)^T$ et $\mathbf{n} = (n_1, n_2, n_3)^T$.

Soient, $\mathbf{\Gamma} = \mathbf{1}$, $\mathbf{\Sigma} = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}$ et $\mathbf{H} = \begin{pmatrix} \sqrt{\alpha_1} & 0 & 0 \\ 0 & \sqrt{\alpha_2} & 0 \\ 0 & 0 & \sqrt{\alpha_3} \end{pmatrix}$.

L'expression (B.16) est obtenue en appliquant (B.2) et en utilisant (A.28). \square

Modèle inadéquat : Démonstration du théorème 4.5

Nous considérons le modèle général d'un canal gaussien :

$$\mathbf{x} = \mathbf{H} \mathbf{s} + \mathbf{n}, \quad (\text{C.1})$$

où $\mathbf{s} \in \mathbb{R}^l$ est le signal (vectoriel) d'entrée de densité $p(\mathbf{s})$, $\mathbf{x} \in \mathbb{R}^k$ est le signal (vectoriel) de sortie de densité $p(\mathbf{x})$, $\mathbf{n} \sim \mathcal{N}(0, \mathbf{\Sigma})$ est un bruit gaussien additif de moyenne nulle et de matrice de covariance $\mathbf{\Sigma}$ et $\mathbf{H} \in \mathbb{R}^{k \times l}$ est la matrice du canal.

Nous supposons que la densité du signal d'entrée et la densité du bruit sont mal définies, $q(\mathbf{s}) \neq p(\mathbf{s})$ et $q(\mathbf{n}) \neq p(\mathbf{n})$. L'inadéquation du bruit que nous supposons se limite au cas où seule la matrice de covariance du bruit est fautive, *i.e.* $q(\mathbf{n}) = \mathcal{N}(\mathbf{0}, \hat{\mathbf{\Sigma}})$.

La technique de démonstration que nous utilisons est similaire à celle utilisée par Chen et Lafferty dans [43]. Nous introduisons, préalablement, quelques lemmes nécessaires à la démonstration. Pour la clarté du développement, nous justifions les interversions d'intégrales et de dérivées à la fin.

C.1 Lemmes

Lemma C.1.1. Les gradients de $p(\mathbf{x}|\mathbf{s})$ par rapport à \mathbf{H} et par rapport à \mathbf{x} sont liés comme suit

$$\nabla_{\mathbf{H}} p(\mathbf{x}|\mathbf{s}) = -\nabla_{\mathbf{x}} p(\mathbf{x}|\mathbf{s}) \mathbf{s}^T \quad (\text{C.2})$$

Démonstration. Etant donné que $p(\mathbf{x}|\mathbf{s}) = \mathcal{N}(\mathbf{H}\mathbf{s}, \mathbf{\Sigma})$, alors

$$\begin{aligned} \nabla_{\mathbf{H}} p(\mathbf{x}|\mathbf{s}) &= (\nabla_{\mathbf{H}} \log p(\mathbf{x}|\mathbf{s})) p(\mathbf{x}|\mathbf{s}) \\ &= -\frac{1}{2} (\nabla_{\mathbf{H}} (\mathbf{x} - \mathbf{H}\mathbf{s})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \mathbf{H}\mathbf{s})) p(\mathbf{x}|\mathbf{s}) \\ &= \mathbf{\Sigma}^{-1} (\mathbf{x} - \mathbf{H}\mathbf{s}) \mathbf{s}^T p(\mathbf{x}|\mathbf{s}) \\ &= \frac{1}{2} (\nabla_{\mathbf{x}} (\mathbf{x} - \mathbf{H}\mathbf{s})^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \mathbf{H}\mathbf{s})) \mathbf{s}^T p(\mathbf{x}|\mathbf{s}) \end{aligned}$$

$$= -\nabla_{\mathbf{x}} p(\mathbf{x}|\mathbf{s}) \mathbf{s}^T$$

□

Lemma C.1.2. Cette lemme est une variation d'Esposito [46] qui relie le gradient du logarithme du vraisemblance à l'erreur quadratique moyenne

$$\nabla_{\mathbf{x}} \log p(\mathbf{x}) = \Sigma^{-1}(\mathbf{H}\mathbb{E}_p[\mathbf{s}|\mathbf{x}] - \mathbf{x}) \quad (\text{C.3})$$

Démonstration.

$$\begin{aligned} \nabla_{\mathbf{x}} \log p(\mathbf{x}) &= \frac{1}{p(\mathbf{x})} \nabla_{\mathbf{x}} p(\mathbf{x}) \\ &= \left(\frac{1}{p(\mathbf{x})} \nabla_{\mathbf{x}} \int p(\mathbf{x}|\mathbf{s}) \right) p(\mathbf{s}) \, ds \\ &= \int \frac{1}{p(\mathbf{x})} \nabla_{\mathbf{x}} p(\mathbf{x}|\mathbf{s}) p(\mathbf{s}) \, ds \quad (\text{C.4}) \\ &= \int \frac{1}{p(\mathbf{x})} (\nabla_{\mathbf{x}} \log p(\mathbf{x}|\mathbf{s})) p(\mathbf{x}|\mathbf{s}) p(\mathbf{s}) \, ds \\ &= \int -\Sigma^{-1}(\mathbf{x} - \mathbf{H}\mathbf{s}) p(\mathbf{s}|\mathbf{x}) \, ds \\ &= \Sigma^{-1}(\mathbf{H}\mathbb{E}_p[\mathbf{s}|\mathbf{x}] - \mathbf{x}) \end{aligned}$$

□

Lemma C.1.3. Les gradients de $p(\mathbf{x})$ par rapport à \mathbf{H} et par rapport à \mathbf{x} sont liés comme suit

$$\nabla_{\mathbf{H}} p(\mathbf{x}) = -\nabla_{\mathbf{x}} (\mathbb{E}_p[\mathbf{s}|\mathbf{x}]^T p(\mathbf{x})) \quad (\text{C.5})$$

Démonstration.

$$\begin{aligned} \nabla_{\mathbf{H}} p(\mathbf{x}) &= \nabla_{\mathbf{H}} \int p(\mathbf{x}|\mathbf{s}) p(\mathbf{s}) \, ds \\ &= \int (\nabla_{\mathbf{H}} p(\mathbf{x}|\mathbf{s})) p(\mathbf{s}) \, ds \quad (\text{C.6}) \end{aligned}$$

$$= -\int (\nabla_{\mathbf{x}} p(\mathbf{x}|\mathbf{s})) \mathbf{s}^T p(\mathbf{s}) \, ds \quad (\text{Lemme C.1.1}) \quad (\text{C.7})$$

$$= -\nabla_{\mathbf{x}} \int \mathbf{s}^T p(\mathbf{x}|\mathbf{s}) p(\mathbf{s}) \, ds \quad (\text{C.8})$$

$$= -\nabla_{\mathbf{x}} \int \mathbf{s}^T p(\mathbf{s}|\mathbf{x}) p(\mathbf{x}) \, ds$$

$$= -\nabla_{\mathbf{x}} (\mathbb{E}_p[\mathbf{s}|\mathbf{x}]^T p(\mathbf{x}))$$

□

C.2 Démonstration du théorème 4.5

Maintenant nous procédons au calcul du gradient de la divergence de Kullback-Leibler. La divergence de Kullback-Leibler peut être exprimée comme différence entre l'entropie inadéquate et l'entropie :

$$D_{KL}(p(\mathbf{x})||q(\mathbf{x})) = h_{p,q}(\mathbf{x}) - h(\mathbf{x}) \quad (\text{C.9})$$

Par conséquent, le gradient de la divergence de Kullback-Leibler est donné par la différence des gradients de l'entropie inadéquate et de l'entropie :

$$\nabla_{\mathbf{H}} D_{KL}(p(\mathbf{x})||q(\mathbf{x})) = \nabla_{\mathbf{H}} h_{p,q}(\mathbf{x}) - \nabla_{\mathbf{H}} h(\mathbf{x}) \quad (\text{C.10})$$

C.2.0.1 Le gradient de l'entropie inadéquate

$$\begin{aligned} \nabla_{\mathbf{H}} h_{p,q}(\mathbf{x}) &= \nabla_{\mathbf{H}} \left(- \int p(\mathbf{x}) \log q(\mathbf{x}) d\mathbf{x} \right) \\ &= - \int \nabla_{\mathbf{H}} p(\mathbf{x}) \log q(\mathbf{x}) d\mathbf{x} - \int p(\mathbf{x}) \nabla_{\mathbf{H}} \log q(\mathbf{x}) d\mathbf{x} \end{aligned} \quad (\text{C.11})$$

Le premier terme de la partie droite de (C.11), s'écrit

$$\begin{aligned} - \int \nabla_{\mathbf{H}} p(\mathbf{x}) \log q(\mathbf{x}) d\mathbf{x} &= - \int \nabla_{\mathbf{H}} \int p(\mathbf{x}|\mathbf{s}) p(\mathbf{s}) d\mathbf{s} \log q(\mathbf{x}) d\mathbf{x} \\ &= - \int \int (\nabla_{\mathbf{H}} p(\mathbf{x}|\mathbf{s})) p(\mathbf{s}) d\mathbf{s} \log q(\mathbf{x}) d\mathbf{x} \end{aligned} \quad (\text{C.12})$$

$$\begin{aligned} &= \int \int (\nabla_{\mathbf{x}} p(\mathbf{x}|\mathbf{s})) \log q(\mathbf{x}) d\mathbf{x} \mathbf{s}^T p(\mathbf{s}) d\mathbf{s} \quad (\text{Lemme C.1.1}) \\ &= \int \left([p(\mathbf{x}|\mathbf{s}) \log q(\mathbf{x})]_{-\infty}^{+\infty} - \int (\nabla_{\mathbf{x}} \log q(\mathbf{x})) p(\mathbf{x}|\mathbf{s}) d\mathbf{x} \right) \mathbf{s}^T p(\mathbf{s}) d\mathbf{s} \quad (\text{C.13}) \end{aligned}$$

$$= - \int \int \hat{\Sigma}^{-1} (\mathbf{H} \mathbb{E}_q[\mathbf{s}|\mathbf{x}] - \mathbf{x}) p(\mathbf{x}|\mathbf{s}) d\mathbf{x} \mathbf{s}^T p(\mathbf{s}) d\mathbf{s}$$

$$= - \int \hat{\Sigma}^{-1} (\mathbf{H} \mathbb{E}_q[\mathbf{s}|\mathbf{x}] - \mathbf{x}) \left(\int p(\mathbf{s}|\mathbf{x}) \mathbf{s}^T d\mathbf{s} \right) p(\mathbf{x}) d\mathbf{x}$$

$$= - \int \hat{\Sigma}^{-1} (\mathbf{H} \mathbb{E}_q[\mathbf{s}|\mathbf{x}] - \mathbf{x}) \mathbb{E}_p[\mathbf{s}|\mathbf{x}]^T p(\mathbf{x}) d\mathbf{x}$$

$$= - \int \hat{\Sigma}^{-1} (\mathbf{H} \hat{\mathbf{s}}_q - \mathbf{H} \hat{\mathbf{s}}_p - \mathbf{n}) \hat{\mathbf{s}}_p^T p(\mathbf{x}) d\mathbf{x} \quad (\text{C.14})$$

$$= \hat{\Sigma}^{-1} \mathbf{H} (\mathbf{M}_{p,q} - \mathbf{M}_{p,p}) + \hat{\Sigma}^{-1} \mathbb{E}_p[\mathbf{n} \hat{\mathbf{s}}_p^T] \quad (\text{C.15})$$

où (C.13) résulte d'une intégration par partie dont le premier terme s'annule à l'infini et (C.15) résulte du principe d'orthogonalité.

Le second terme de la partie droite de (C.11), s'écrit

$$\begin{aligned} - \int p(\mathbf{x}) \nabla_{\mathbf{H}} \log q(\mathbf{x}) d\mathbf{x} &= - \int \frac{p(\mathbf{x})}{q(\mathbf{x})} \nabla_{\mathbf{H}} q(\mathbf{x}) d\mathbf{x} \\ &= \int \frac{p(\mathbf{x})}{q(\mathbf{x})} \nabla_{\mathbf{x}} \left(\mathbb{E}_q[\mathbf{s}|\mathbf{x}]^T q(\mathbf{x}) \right) d\mathbf{x} \quad (\text{Lemme C.1.3}) \\ &= \left[p(\mathbf{x}) \mathbb{E}_q[\mathbf{s}|\mathbf{x}]^T \right]_{-\infty}^{+\infty} - \int \left(\nabla_{\mathbf{x}} \frac{p(\mathbf{x})}{q(\mathbf{x})} \right) \mathbb{E}_q[\mathbf{s}|\mathbf{x}]^T q(\mathbf{x}) d\mathbf{x} \quad (\text{C.16}) \end{aligned}$$

$$\begin{aligned}
&= - \int \left(\nabla_{\mathbf{x}} \log \frac{p(\mathbf{x})}{q(\mathbf{x})} \right) \mathbb{E}_q [\mathbf{s}|\mathbf{x}]^T p(\mathbf{x}) d\mathbf{x} \\
&= - \int \left(\Sigma^{-1}(\mathbf{H}\hat{\Sigma}_p - \mathbf{x}) - \hat{\Sigma}^{-1}(\mathbf{H}\hat{\Sigma}_q - \mathbf{x}) \right) \hat{\Sigma}_q^T p(\mathbf{x}) d\mathbf{x} \quad (\text{C.17})
\end{aligned}$$

$$= \left(\Sigma^{-1} - \hat{\Sigma}^{-1} \right) \mathbb{E}_p [\mathbf{n}\hat{\Sigma}_q^T] \quad (\text{C.18})$$

où (C.18) résulte du principe d'orthogonalité et (C.16) résulte d'une intégration par partie dont le premier terme s'annule à l'infini. En fait,

$$p(\mathbf{x})\mathbb{E}_q [\mathbf{s}|\mathbf{x}]^T = p(\mathbf{x}) \int \mathbf{s}^T q(\mathbf{s}|\mathbf{x}) d\mathbf{s} = \frac{p(\mathbf{x})}{q(\mathbf{x})} \int \mathbf{s}^T q(\mathbf{s}|\mathbf{x}) q(\mathbf{x}) d\mathbf{s} = \frac{p(\mathbf{x})}{q(\mathbf{x})} \int \mathbf{s}^T q(\mathbf{x}|\mathbf{s}) q(\mathbf{s}) d\mathbf{s} \quad (\text{C.19})$$

$$p(\mathbf{x})\mathbb{E}_q [\mathbf{s}|\mathbf{x}]^T = p(\mathbf{x}) \int \mathbf{s}^T p(\mathbf{s}|\mathbf{x}) d\mathbf{s} = \int \mathbf{s}^T p(\mathbf{s}|\mathbf{x}) p(\mathbf{x}) d\mathbf{s} = \int \mathbf{s}^T p(\mathbf{x}|\mathbf{s}) p(\mathbf{s}) d\mathbf{s} \quad (\text{C.20})$$

En appliquant le théorème de convergence dominée [47] alors

$$\lim_{\mathbf{x} \rightarrow \infty} p(\mathbf{x})\mathbb{E}_q [\mathbf{s}|\mathbf{x}]^T = \int \mathbf{s}^T \lim_{\mathbf{x} \rightarrow \infty} p(\mathbf{x}|\mathbf{s}) p(\mathbf{s}) d\mathbf{s} = 0$$

Finalement, en remplaçant les gradients des deux termes à droite de (C.11) nous obtenons :

$$\nabla_{\mathbf{H}} h_{p,q}(\mathbf{x}) = \hat{\Sigma}^{-1} \mathbf{H} (\mathbf{M}_{p,q} - \mathbf{M}_{p,p}) + \hat{\Sigma}^{-1} \mathbb{E}_p [\mathbf{n}\hat{\Sigma}_p^T] + \left(\Sigma^{-1} - \hat{\Sigma}^{-1} \right) \mathbb{E}_p [\mathbf{n}\hat{\Sigma}_q^T] \quad (\text{C.21})$$

C.2.0.2 Le gradient d'entropie

Le gradient de l'entropie est obtenu à partir du gradient de l'entropie inadéquate (C.21), en remplaçant q par p and $\hat{\Sigma}$ par Σ , ainsi

$$\nabla_{\mathbf{H}} h(\mathbf{x}) = \Sigma^{-1} \mathbb{E}_p [\mathbf{n}\mathbb{E}_p[\mathbf{s}|\mathbf{x}]^T] \quad (\text{C.22})$$

Finalement, en substituant les gradients des entropies par leurs expressions

$$\nabla_{\mathbf{H}} D_{KL}(p(\mathbf{x})||q(\mathbf{x})) = \hat{\Sigma}^{-1} \mathbf{H} (\mathbf{M}_{p,q} - \mathbf{M}_{p,p}) - (\hat{\Sigma}^{-1} - \Sigma^{-1}) (\mathbb{E}_p [\mathbf{n}\mathbb{E}_q[\mathbf{s}|\mathbf{x}]] - \mathbb{E}_p [\mathbf{n}\mathbb{E}_p[\mathbf{s}|\mathbf{x}]])$$

C.3 Permutation d'intégrale et de dérivée

Dans la démonstration on a effectué plusieurs permutations d'intégral et de dérivation. Afin de les justifier, nous rappelons un lemme utilisé par Palomar et al à la même fin.

Lemma C.3.1. Soient $f(\mathbf{y}, z)$ une fonction et $g(\mathbf{y})$ une fonction non négative. S'il existe un voisinage \mathcal{N}_{z_0} pour tout z_0 et une fonction $M(\mathbf{y}; z_0)$ tels que :

$$\sup_{z \in \mathcal{N}_{z_0}} \left| \frac{\partial}{\partial z} f(\mathbf{y}, z) \right| \leq M(\mathbf{y}; z_0) \quad \text{presque partout} \quad (\text{C.23})$$

et

$$\int g(\mathbf{y}) M(\mathbf{y}; z_0) d\mathbf{y} < \infty \quad (\text{C.24})$$

alors

$$\frac{\partial}{\partial z} \int f(\mathbf{y}, z) g(\mathbf{y}) d\mathbf{y} = \int g(\mathbf{y}) \frac{\partial}{\partial z} f(\mathbf{y}, z) d\mathbf{y} \quad (\text{C.25})$$

Démonstration. Voir [21, Appendix B] □

La première et troisième permutations sont justifiées dans [21]. Par souci d'exhaustivité, nous les rappelons ici

C.3.1 La permutation en (C.4)

En (C.4), nous avons effectué la permutation suivante : $\nabla_{\mathbf{x}} \int p(\mathbf{x}|\mathbf{s})p(\mathbf{s})d\mathbf{s} = \int \nabla_{\mathbf{x}}p(\mathbf{x}|\mathbf{s})p(\mathbf{s})d\mathbf{s}$
Par analogie au lemme C.3.1, $z = \mathbf{x}_i$, $\mathbf{y} = \mathbf{s}$, $f(\mathbf{y}, z) = p(\mathbf{x}_i|\mathbf{s})$ et $g(\mathbf{y}) = p(\mathbf{s})$.

On a,

$$\left| \frac{\partial}{\partial \mathbf{x}_i} p(\mathbf{x}|\mathbf{s}) \right| = \left| \left[-\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mathbf{H}\mathbf{s})p(\mathbf{x}|\mathbf{s}) \right]_i \right| \quad (\text{C.26})$$

$$\leq \left| \left[-\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mathbf{H}\mathbf{s}) \right]_i \right| \quad (\text{C.27})$$

$$\leq \sum_k \left| \boldsymbol{\Sigma}_{ik} \right|^{-1} |\mathbf{x}_i| + \sum_k \left| \left[\boldsymbol{\Sigma}^{-1} \mathbf{H} \right]_{lk} \right| |\mathbf{s}_k| \quad (\text{C.28})$$

Ainsi la condition (C.23) est vérifiée, où $M(\mathbf{y}; z_0) = \sum_k \left| \boldsymbol{\Sigma}_{ik} \right|^{-1} |\mathbf{x}_i| + \sum_k \left| \left[\boldsymbol{\Sigma}^{-1} \mathbf{H} \right]_{lk} \right| |\mathbf{s}_k|$. De plus, on a

$$\int \sum_k \left| \boldsymbol{\Sigma}_{ik} \right|^{-1} |\mathbf{x}_k| p(\mathbf{s}) d\mathbf{s} = \left| \boldsymbol{\Sigma}_{ik} \right|^{-1} |\mathbf{x}_k| < \left| \boldsymbol{\Sigma}_{ik} \right|^{-1} |\mathbf{x}_k| + \epsilon \quad (\text{C.29})$$

et

$$\int \sum_k \left| \left[\boldsymbol{\Sigma}^{-1} \mathbf{H} \right]_{lk} \right| |\mathbf{s}_k| p(\mathbf{s}) d\mathbf{s} = \sum_k \left| \left[\boldsymbol{\Sigma}^{-1} \mathbf{H} \right]_{lk} \right| \int |\mathbf{s}_k| p(\mathbf{s}) d\mathbf{s} \quad (\text{C.30})$$

où $\int |\mathbf{s}_k| p(\mathbf{s}) d\mathbf{s}$ est fini, puisque les moments d'ordre 2 de \mathbf{s} sont finis [47]. Par suite la condition (C.24) est confirmé.

C.3.2 La permutation en (C.8)

En (C.8), nous avons effectué la permutation suivante : $\nabla_{\mathbf{x}} \int p(\mathbf{x}|\mathbf{s})p(\mathbf{s})d\mathbf{s} = \int \nabla_{\mathbf{x}}p(\mathbf{x}|\mathbf{s})p(\mathbf{s})d\mathbf{s}$
Par analogie au lemme C.3.1, $z = \mathbf{x}_i$, $\mathbf{y} = \mathbf{s}$, $f(\mathbf{y}, z) = \mathbf{s}^T p(\mathbf{x}_i|\mathbf{s})$ et $g(\mathbf{y}) = p(\mathbf{s})$. Notez la similarité avec la permutation précédente sauf pour la fonction $f(\mathbf{y}, z)$.

On a,

$$\left| \frac{\partial}{\partial \mathbf{x}_i} \mathbf{s}^T p(\mathbf{x}|\mathbf{s}) \right| = \left| \left[-\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mathbf{H}\mathbf{s}) \mathbf{s}^T p(\mathbf{x}|\mathbf{s}) \right]_i \right| \quad (\text{C.31})$$

$$\leq \left| \left[-\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mathbf{H}\mathbf{s}) \mathbf{s}^T \right]_i \right| \quad (\text{C.32})$$

$$\leq \sum_k \sum_l \left| \boldsymbol{\Sigma}_{ik} \right|^{-1} |\mathbf{x}_k| |\mathbf{s}_l| + \sum_k \sum_l \left| \left[\boldsymbol{\Sigma}^{-1} \mathbf{H} \right]_{lk} \right| |\mathbf{s}_k| |\mathbf{s}_l| \quad (\text{C.33})$$

Ainsi la condition (C.23) est vérifiée, où $M(\mathbf{y}; z_0) = \sum_k \sum_l \left| \boldsymbol{\Sigma}_{ik} \right|^{-1} |\mathbf{x}_k| |\mathbf{s}_l| + \sum_k \sum_l \left| \left[\boldsymbol{\Sigma}^{-1} \mathbf{H} \right]_{lk} \right| |\mathbf{s}_k| |\mathbf{s}_l|$. De plus, on a

$$\int \sum_k \sum_l \left| \boldsymbol{\Sigma}_{ik} \right|^{-1} |\mathbf{x}_k| |\mathbf{s}_l| p(\mathbf{s}) d\mathbf{s} = \sum_k \sum_l \left| \boldsymbol{\Sigma}_{ik} \right|^{-1} |\mathbf{x}_k| \int |\mathbf{s}_l| p(\mathbf{s}) d\mathbf{s} \quad (\text{C.34})$$

et

$$\int \sum_k \sum_l \left| \left[\boldsymbol{\Sigma}^{-1} \mathbf{H} \right]_{ik} \right| |\mathbf{s}_k| |\mathbf{s}_l| p(\mathbf{s}) d\mathbf{s} = \sum_k \sum_l \left| \left[\boldsymbol{\Sigma}^{-1} \mathbf{H} \right]_{ik} \right| \int |\mathbf{s}_k| |\mathbf{s}_l| p(\mathbf{s}) d\mathbf{s} \quad (\text{C.35})$$

où $\int |\mathbf{s}_k| p(\mathbf{s}) d\mathbf{s}$ et $\int |\mathbf{s}_k| |\mathbf{s}_l| p(\mathbf{s}) d\mathbf{s}$ sont finis, puisque les moments d'ordre 2 de \mathbf{s} sont finis [47]. Par suite la condition (C.24) est confirmé.

C.3.3 La permutation en (C.6) et (C.12)

En (C.6) et (C.12), nous avons effectué la permutation suivante :

$$\nabla_{\mathbf{H}} \int p(\mathbf{x}|\mathbf{s}) p(\mathbf{s}) d\mathbf{s} = \int \nabla_{\mathbf{H}} p(\mathbf{x}|\mathbf{s}) p(\mathbf{s}) d\mathbf{s}$$

Par analogie au lemme C.3.1, $z = [\mathbf{H}]_{ij}$, $\mathbf{y} = \mathbf{s}$, $f(\mathbf{y}, z) = p(\mathbf{x}_i|\mathbf{s})$ et $g(\mathbf{y}) = p(\mathbf{s})$.

On a,

$$\left| \frac{\partial}{\partial [\mathbf{H}]_{ij}} p(\mathbf{x}|\mathbf{s}) \right| = \left| \left[\boldsymbol{\Sigma}^{-1} (\mathbf{x} - \mathbf{H}\mathbf{s}) \mathbf{s}^T p(\mathbf{x}|\mathbf{s}) \right]_{ij} \right| \quad (\text{C.36})$$

$$\leq \left| \left[\boldsymbol{\Sigma}^{-1} (\mathbf{x} - \mathbf{H}\mathbf{s}) \mathbf{s}^T \right]_i \right| \quad (\text{C.37})$$

$$\leq \sum_k \left| \boldsymbol{\Sigma}_{ik} \right|^{-1} |\mathbf{x}_i| |\mathbf{s}_j| + \sum_k \left| \left[\boldsymbol{\Sigma}^{-1} \mathbf{H} \right]_{ik} \right| |\mathbf{s}_k| |\mathbf{s}_j| \quad (\text{C.38})$$

Ainsi la condition (C.23) est vérifiée, où $M(\mathbf{y}; z_0) = \sum_k \left| \boldsymbol{\Sigma}_{ik} \right|^{-1} |\mathbf{x}_i| |\mathbf{s}_j| + \sum_k \left| \left[\boldsymbol{\Sigma}^{-1} \mathbf{H} \right]_{ik} \right| |\mathbf{s}_k| |\mathbf{s}_j|$.

De plus, on a

$$\int \sum_k \left| \boldsymbol{\Sigma}_{ik} \right|^{-1} |\mathbf{x}_i| |\mathbf{s}_j| p(\mathbf{s}) d\mathbf{s} = \left| \boldsymbol{\Sigma}_{ik} \right|^{-1} |\mathbf{x}_i| \int |\mathbf{s}_j| p(\mathbf{s}) d\mathbf{s} \quad (\text{C.39})$$

et

$$\int \sum_k \left| \left[\boldsymbol{\Sigma}^{-1} \mathbf{H} \right]_{ik} \right| |\mathbf{s}_k| |\mathbf{s}_j| p(\mathbf{s}) d\mathbf{s} = \sum_k \left| \left[\boldsymbol{\Sigma}^{-1} \mathbf{H} \right]_{ik} \right| \int |\mathbf{s}_k| |\mathbf{s}_j| p(\mathbf{s}) d\mathbf{s} \quad (\text{C.40})$$

où $\int |\mathbf{s}_k| p(\mathbf{s}) d\mathbf{s}$ et $\int |\mathbf{s}_k| |\mathbf{s}_l| p(\mathbf{s}) d\mathbf{s}$ sont finis, puisque les moments d'ordre 2 de \mathbf{s} sont finis [47]. Par suite la condition (C.24) est confirmée.

C.3.4 La permutation en (C.11)

En (C.11), nous avons effectué la permutation suivante :

$$\nabla_{\mathbf{H}} \left(\int p(\mathbf{x}) \log q(\mathbf{x}) d\mathbf{x} \right) = \int \nabla_{\mathbf{H}} (p(\mathbf{x}) \log q(\mathbf{x})) d\mathbf{x}$$

Par analogie au lemme C.3.1, $z = [\mathbf{H}]_{ij}$, $\mathbf{y} = \mathbf{x}$, $f(\mathbf{y}, z) = p(\mathbf{x}) \log q(\mathbf{x})$ et $g(\mathbf{y}) = 1$.

L'expression de $\nabla_{\mathbf{H}} (p(\mathbf{x}) \log q(\mathbf{x}))$ découle de (C.14) et (C.17),

$$\nabla_{\mathbf{H}} (p(\mathbf{x}) \log q(\mathbf{x})) = \left(\boldsymbol{\Sigma}^{-1} (\mathbf{H}\hat{\mathbf{s}}_p - \mathbf{x}) \hat{\mathbf{s}}_q^T + \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{H}\hat{\mathbf{s}}_q - \mathbf{x}) (\hat{\mathbf{s}}_q^T - \hat{\mathbf{s}}_q^T) \right) p(\mathbf{x}) \quad (\text{C.41})$$

On a

$$\left| \frac{\partial}{\partial [\mathbf{H}]_{ij}} (p(\mathbf{x}) \log q(\mathbf{x})) \right| = \left| \left[\left(\boldsymbol{\Sigma}^{-1} (\mathbf{H}\hat{\mathbf{s}}_p - \mathbf{x}) \hat{\mathbf{s}}_q^T + \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{H}\hat{\mathbf{s}}_q - \mathbf{x}) (\hat{\mathbf{s}}_q^T - \hat{\mathbf{s}}_q^T) \right) p(\mathbf{x}) \right]_{i,j} \right|$$

$$\begin{aligned}
&\leq \left| \left[\boldsymbol{\Sigma}^{-1}(\mathbf{H}\hat{\mathbf{s}}_p - \mathbf{x})\hat{\mathbf{s}}_q^T + \hat{\boldsymbol{\Sigma}}^{-1}(\mathbf{H}\hat{\mathbf{s}}_q - \mathbf{x})(\hat{\mathbf{s}}_p^T - \hat{\mathbf{s}}_q^T) \right]_{i,j} \right| \\
&\leq \sum_k \left| [\boldsymbol{\Sigma}^{-1}\mathbf{H} + \hat{\boldsymbol{\Sigma}}^{-1}\mathbf{H}]_{ik} \right| \mathbb{E}_p[\mathbf{s}|\mathbf{x}]_i \mathbb{E}_q[\mathbf{s}|\mathbf{x}]_j + \sum_k \left| \boldsymbol{\Sigma}_{ik}^{-1} + \hat{\boldsymbol{\Sigma}}_{ik}^{-1} \right| \|\mathbf{x}_i\| \mathbb{E}_q[\mathbf{s}|\mathbf{x}]_j \\
&\quad + \sum_k \left| [\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{H}]_{ik} \right| \mathbb{E}_p[\mathbf{s}|\mathbf{x}]_i \mathbb{E}_p[\mathbf{s}|\mathbf{x}]_j + \sum_k \left| \hat{\boldsymbol{\Sigma}}_{ik}^{-1} \right| \|\mathbf{x}_i\| \mathbb{E}_p[\mathbf{s}|\mathbf{x}]_j
\end{aligned}$$

Ainsi la condition (C.23) est vérifiée. Nous pouvons facilement vérifier que la condition (C.24) est remplie.

Modèle inadéquat : Calcul d'estimation

D.1 Modèle gaussien général

Dans le cadre d'un modèle gaussien inadéquat, l'estimateur considéré et l'erreur d'estimation associée sont donnés par le théorème suivant :

Théorème D.1: (Estimateur inadéquat - Modèle gaussien)

Soit le modèle linéaire suivant,

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{n}, \quad (\text{D.1})$$

où $\mathbf{s} \sim p(\mathbf{s}) = \mathcal{N}(\mathbf{0}, \mathbf{\Gamma})$ est le signal d'intérêt, $\mathbf{x} \in \mathbb{R}^k$ est le vecteur d'observations, $\mathbf{n} \sim p(\mathbf{n}) = \mathcal{N}(0, \mathbf{\Sigma})$ est le bruit gaussien additif et $\mathbf{H} \in \mathbb{R}^{k \times l}$ est la matrice du canal. Nous supposons que \mathbf{s} est indépendante du bruit \mathbf{n} .

Nous considérons le cas d'une double inadéquation du modèle. Nous supposons que la densité a priori du signal d'entrée est $q(\mathbf{s}) = \mathcal{N}(0, \hat{\mathbf{\Gamma}})$ et la densité du bruit est $q(\mathbf{n}) = \mathcal{N}(0, \hat{\mathbf{\Sigma}})$.

L'estimateur de \mathbf{s} en fonction de \mathbf{x} est :

$$\hat{\mathbf{s}}_q(\mathbf{x}) = (\hat{\mathbf{\Gamma}}^{-1} + \mathbf{H}^T \hat{\mathbf{\Sigma}}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \hat{\mathbf{\Sigma}}^{-1} \mathbf{x}, \quad (\text{D.2})$$

$$= \hat{\mathbf{\Gamma}} \mathbf{H}^T (\mathbf{H} \hat{\mathbf{\Gamma}} \mathbf{H}^T + \hat{\mathbf{\Sigma}})^{-1} \mathbf{x}. \quad (\text{D.3})$$

Et l'erreur quadratique moyenne associée à $\hat{\mathbf{s}}_q(\mathbf{x})$ est :

$$\begin{aligned} \mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) &= \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) + \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \hat{\mathbf{\Gamma}}^{-1} (\mathbf{\Gamma} - \hat{\mathbf{\Gamma}}) \hat{\mathbf{\Gamma}}^{-1} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \\ &\quad + \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \mathbf{H}^T \hat{\mathbf{\Sigma}}^{-1} (\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}) \hat{\mathbf{\Sigma}}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}). \end{aligned} \quad (\text{D.4})$$

où

$$\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) = (\hat{\mathbf{\Gamma}}^{-1} + \mathbf{H}^T \hat{\mathbf{\Sigma}}^{-1} \mathbf{H})^{-1} = \hat{\mathbf{\Gamma}} - \hat{\mathbf{\Gamma}} \mathbf{H}^T (\mathbf{H} \hat{\mathbf{\Gamma}} \mathbf{H}^T + \hat{\mathbf{\Sigma}})^{-1} \mathbf{H} \hat{\mathbf{\Gamma}}. \quad (\text{D.5})$$

Démonstration. Sous inadéquation du modèle, nous calculons l'estimateur de \mathbf{s} en fonction de \mathbf{x} , que nous croyons optimal, comme la moyenne de la loi a posteriori supposé $q(\mathbf{s}|\mathbf{x})$ (paragraphe 4.2) :

$$\hat{\mathbf{s}}_q(\mathbf{x}) = \mathbb{E}_q[\mathbf{s}|\mathbf{x}]. \quad (\text{D.6})$$

Ainsi, en appliquant le théorème A.1, nous obtenons :

$$\hat{\mathbf{s}}_q(\mathbf{x}) = (\hat{\Gamma}^{-1} + \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{x}, \quad (\text{D.7})$$

$$= \hat{\Gamma} \mathbf{H}^T (\mathbf{H} \hat{\Gamma} \mathbf{H}^T + \hat{\Sigma})^{-1} \mathbf{x}. \quad (\text{D.8})$$

L'erreur quadratique moyenne associée à l'estimateur $\hat{\mathbf{s}}_q(\mathbf{x})$ est donnée par la matrice de covariance de $\mathbf{s} - \hat{\mathbf{s}}_q(\mathbf{x})$:

$$\begin{aligned} \mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) &= \mathbb{E}_p[(\mathbf{s} - \hat{\mathbf{s}}_q(\mathbf{x}))(\mathbf{s} - \hat{\mathbf{s}}_q(\mathbf{x}))^T], \\ &= \mathbb{E}_p[\mathbf{s}\mathbf{s}^T] - \mathbb{E}_p[\hat{\mathbf{s}}_q(\mathbf{x})\mathbf{s}^T] - \mathbb{E}_p[\mathbf{s}\hat{\mathbf{s}}_q(\mathbf{x})^T] + \mathbb{E}_p[\hat{\mathbf{s}}_q(\mathbf{x})\hat{\mathbf{s}}_q(\mathbf{x})^T], \\ &= \Gamma - (\hat{\Gamma}^{-1} + \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H} \Gamma - \Gamma \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H} (\hat{\Gamma}^{-1} + \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H})^{-1} \\ &\quad + (\hat{\Gamma}^{-1} + \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \hat{\Sigma}^{-1} (\mathbf{H} \Gamma \mathbf{H}^T + \Sigma) \hat{\Sigma} \mathbf{H}^T (\hat{\Gamma}^{-1} + \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H})^{-1}, \\ &= \Gamma - \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H} \Gamma - \Gamma \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \\ &\quad + \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \mathbf{H}^T \hat{\Sigma}^{-1} (\mathbf{H} \Gamma \mathbf{H}^T + \Sigma) \hat{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}), \\ &= \Gamma (\mathbf{I} - \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})) - \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H} \Gamma (\mathbf{I} - \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})) \\ &\quad + \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \mathbf{H}^T \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}), \\ &= (\mathbf{I} - \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H}) \Gamma (\mathbf{I} - \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})) \\ &\quad + \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \mathbf{H}^T \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}), \\ &= \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) (\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})^{-1} - \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H}) \Gamma (\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x})^{-1} - \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H}) \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \\ &\quad + \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \mathbf{H}^T \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}), \\ &= \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) (\mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H} + \hat{\Gamma}^{-1} - \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H}) \Gamma (\mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H} + \hat{\Gamma}^{-1} - \mathbf{H}^T \hat{\Sigma}^{-1} \mathbf{H}) \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \\ &\quad + \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \mathbf{H}^T \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}), \\ &= \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \hat{\Gamma}^{-1} \Gamma \hat{\Gamma}^{-1} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) + \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \mathbf{H}^T \hat{\Sigma}^{-1} \Sigma \hat{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}), \\ &= \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) + \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \hat{\Gamma}^{-1} (\Gamma - \hat{\Gamma}) \hat{\Gamma}^{-1} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \\ &\quad + \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) \mathbf{H}^T \hat{\Sigma}^{-1} (\Sigma - \hat{\Sigma}) \hat{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}). \end{aligned}$$

□

D.2 Deux modalités scalaires faiblement couplées

Soit le modèle de deux modalités complémentaires suivant :

$$\begin{cases} x_1 = \sqrt{\alpha_1} s_1 + n_1 \\ x_2 = \sqrt{\alpha_2} s_2 + n_2 \end{cases}, \quad (\text{D.9})$$

où s_1, s_2, n_1 et n_2 sont des variables gaussiennes centrées réduites. s_1 et s_2 sont indépendants de n_1 et n_2 . Alors que, les signaux s_1 et s_2 (resp. les bruits n_1 et n_2) sont corrélés, nous notons leurs coefficients de corrélations par ρ_s (resp. ρ_n).

Nous supposons que les coefficients de corrélation des signaux et des bruits sont $\hat{\rho}_s$ et $\hat{\rho}_n$ au lieu de ρ_s et ρ_n , respectivement. Ainsi, l'estimateur inadéquat de $s_i, i \in \{1, 2\}$, en fonction de x_1 et x_2 est :

$$\hat{s}_{ip}(x_1, x_2) = \frac{(\alpha_j \sqrt{\alpha_i} (1 - \hat{\rho}_s^2) + (\sqrt{\alpha_i} - \sqrt{\alpha_j} \hat{\rho}_n \hat{\rho}_s)) x_i - (\sqrt{\alpha_i} \hat{\rho}_n - \sqrt{\alpha_j} \hat{\rho}_s) x_j}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1 \alpha_2} \hat{\rho}_s + \hat{\rho}_n)^2}. \quad (\text{D.10})$$

L'erreur d'estimation associée à l'estimateur $\hat{s}_{ip}(x_1, x_2)$ est :

$$\begin{aligned} \mathbf{M}_{p,q}(s_i | x_1, x_2) &= \frac{\alpha_j (1 - \hat{\rho}_s^2) + 1 - \hat{\rho}_n^2}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1 \alpha_2} \hat{\rho}_s + \hat{\rho}_n)^2} \\ &\quad + 2\sqrt{\alpha_j} (\rho_s - \hat{\rho}_s) (\sqrt{\alpha_i} \hat{\rho}_n - \sqrt{\alpha_j} \hat{\rho}_s) \frac{(1 + \alpha_j) - \hat{\rho}_n (\hat{\rho}_n + \sqrt{\alpha_1 \alpha_2} \hat{\rho}_s)}{((1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1 \alpha_2} \hat{\rho}_s)^2)^2} \\ &\quad - 2(\rho_n - \hat{\rho}_n) (\sqrt{\alpha_i} \hat{\rho}_n - \sqrt{\alpha_j} \hat{\rho}_s) \frac{\sqrt{\alpha_i} (1 + \alpha_j) - \sqrt{\alpha_j} \hat{\rho}_s (\hat{\rho}_n + \sqrt{\alpha_1 \alpha_2} \hat{\rho}_s)}{((1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1 \alpha_2} \hat{\rho}_s)^2)^2}. \end{aligned} \quad (\text{D.11})$$

Démonstration. Nous posons $\mathbf{x} = (x_1, x_2)^T$, $\mathbf{s} = (s_1, s_2)^T$ et $\mathbf{n} = (n_1, n_2)^T$. Soient, $\mathbf{\Gamma} =$

$$\begin{bmatrix} 1 & \rho_s \\ \rho_s & 1 \end{bmatrix}, \quad \mathbf{\Sigma} = \begin{bmatrix} 1 & \rho_n \\ \rho_n & 1 \end{bmatrix} \quad \text{et} \quad \mathbf{H} = \begin{bmatrix} \sqrt{\alpha_1} & 0 \\ 0 & \sqrt{\alpha_2} \end{bmatrix}.$$

L'estimateur inadéquat de \mathbf{s} en fonction de \mathbf{x} est obtenu à partir du théorème D.1 :

$$\begin{aligned} \hat{\mathbf{s}}_q(\mathbf{x}) &= \begin{bmatrix} \hat{s}_{1q} \\ \hat{s}_{2q} \end{bmatrix} = \hat{\mathbf{\Gamma}} \mathbf{H}^T (\mathbf{H} \hat{\mathbf{\Gamma}} \mathbf{H}^T + \hat{\mathbf{\Sigma}})^{-1} \mathbf{x}, \\ &= \frac{\begin{bmatrix} (\alpha_2 \sqrt{\alpha_1} (1 - \hat{\rho}_s^2) + (\sqrt{\alpha_1} - \sqrt{\alpha_2} \hat{\rho}_n \hat{\rho}_s)) x_1 - (\sqrt{\alpha_1} \hat{\rho}_n - \sqrt{\alpha_2} \hat{\rho}_s) x_2 \\ (\alpha_1 \sqrt{\alpha_2} (1 - \hat{\rho}_s^2) + (\sqrt{\alpha_2} - \sqrt{\alpha_1} \hat{\rho}_n \hat{\rho}_s)) x_2 - (\sqrt{\alpha_2} \hat{\rho}_n - \sqrt{\alpha_1} \hat{\rho}_s) x_1 \end{bmatrix}}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1 \alpha_2} \hat{\rho}_s + \hat{\rho}_n)^2}. \end{aligned} \quad (\text{D.12})$$

L'équation (D.12) est obtenue à partir de (A.20) en substituant ρ_s par $\hat{\rho}_s$ et ρ_n par $\hat{\rho}_n$.

Selon le théorème D.1, l'erreur d'estimation associée à l'estimateur inadéquat $\hat{\mathbf{s}}_q(\mathbf{x})$ est :

$$\begin{aligned} \mathbf{M}_{p,q}(\mathbf{s} | \mathbf{x}) &= \mathbf{M}_{q,q}(\mathbf{s} | \mathbf{x}) + \mathbf{M}_{q,q}(\mathbf{s} | \mathbf{x}) \hat{\mathbf{\Gamma}}^{-1} (\mathbf{\Gamma} - \hat{\mathbf{\Gamma}}) \hat{\mathbf{\Gamma}}^{-1} \mathbf{M}_{q,q}(\mathbf{s} | \mathbf{x}) \\ &\quad + \mathbf{M}_{q,q}(\mathbf{s} | \mathbf{x}) \mathbf{H}^T \hat{\mathbf{\Sigma}}^{-1} (\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}) \hat{\mathbf{\Sigma}}^{-1} \mathbf{H} \mathbf{M}_{q,q}(\mathbf{s} | \mathbf{x}). \end{aligned} \quad (\text{D.13})$$

où

$$\mathbf{M}_{q,q}(\mathbf{s} | \mathbf{x}) = (\hat{\mathbf{\Gamma}}^{-1} + \mathbf{H}^T \hat{\mathbf{\Sigma}}^{-1} \mathbf{H})^{-1} = \hat{\mathbf{\Gamma}} - \hat{\mathbf{\Gamma}} \mathbf{H}^T (\mathbf{H} \hat{\mathbf{\Gamma}} \mathbf{H}^T + \hat{\mathbf{\Sigma}})^{-1} \mathbf{H} \hat{\mathbf{\Gamma}}. \quad (\text{D.14})$$

A partir de (A.21), en substituant ρ_s par $\hat{\rho}_s$ et ρ_n par $\hat{\rho}_n$, nous trouvons :

$$\mathbf{M}_{q,q}(\mathbf{s}|\mathbf{x}) = \frac{\begin{bmatrix} \alpha_2(1 - \hat{\rho}_s^2) + 1 - \hat{\rho}_n^2 & \sqrt{\alpha_1\alpha_2}\hat{\rho}_n(1 - \hat{\rho}_s^2) + \hat{\rho}_s(1 - \hat{\rho}_n^2) \\ \sqrt{\alpha_1\alpha_2}\hat{\rho}_n(1 - \hat{\rho}_s^2) + \hat{\rho}_s(1 - \hat{\rho}_n^2) & \alpha_1(1 - \hat{\rho}_s^2) + 1 - \hat{\rho}_n^2 \end{bmatrix}}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1\alpha_2}\hat{\rho}_s + \hat{\rho}_n)^2} \quad (\text{D.15})$$

Nous calculons, ensuite, le deuxième terme de la partie droite de l'équation (D.13). On a :

$$\begin{aligned} \hat{\Gamma}^{-1}(\mathbf{\Gamma} - \hat{\mathbf{\Gamma}})\hat{\Gamma}^{-1} &= \frac{1}{(1 - \hat{\rho}_s^2)^2} \begin{bmatrix} 1 & -\hat{\rho}_s \\ -\hat{\rho}_s & 1 \end{bmatrix} \begin{bmatrix} 0 & \rho_s - \hat{\rho}_s \\ \rho_s - \hat{\rho}_s & 0 \end{bmatrix} \begin{bmatrix} 1 & -\hat{\rho}_s \\ -\hat{\rho}_s & 1 \end{bmatrix}, \\ &= \frac{1}{(1 - \hat{\rho}_s^2)^2} \begin{bmatrix} -\hat{\rho}_s(\rho_s - \hat{\rho}_s) & \rho_s - \hat{\rho}_s \\ \rho_s - \hat{\rho}_s & -\hat{\rho}_s(\rho_s - \hat{\rho}_s) \end{bmatrix} \begin{bmatrix} 1 & -\hat{\rho}_s \\ -\hat{\rho}_s & 1 \end{bmatrix}, \\ &= \frac{\rho_s - \hat{\rho}_s}{(1 - \hat{\rho}_s^2)^2} \begin{bmatrix} -2\hat{\rho}_s & (1 + \hat{\rho}_s^2) \\ (1 + \hat{\rho}_s^2) & -2\hat{\rho}_s \end{bmatrix}. \end{aligned}$$

Ensuite :

$$\begin{aligned} &\mathbf{M}_{q,q}\hat{\Gamma}^{-1}(\mathbf{\Gamma} - \hat{\mathbf{\Gamma}})\hat{\Gamma}^{-1}\mathbf{M}_{q,q} \\ &= \frac{\begin{bmatrix} (1 - \hat{\rho}_n^2) + (1 - \hat{\rho}_s^2)\alpha_2 & \hat{\rho}_s(1 - \hat{\rho}_n^2) + \sqrt{\alpha_1\alpha_2}(1 - \hat{\rho}_s^2)\hat{\rho}_n \\ \hat{\rho}_s(1 - \hat{\rho}_n^2) + \sqrt{\alpha_1\alpha_2}(1 - \hat{\rho}_s^2)\hat{\rho}_n & (1 - \hat{\rho}_n^2) + (1 - \hat{\rho}_s^2)\alpha_1 \end{bmatrix}}{(1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)^2} \\ &\quad \times \frac{\rho_s - \hat{\rho}_s}{(1 - \hat{\rho}_s^2)^2} \begin{bmatrix} -2\hat{\rho}_s & (1 + \hat{\rho}_s^2) \\ (1 + \hat{\rho}_s^2) & -2\hat{\rho}_s \end{bmatrix} \mathbf{M}_{q,q}, \\ &= \frac{\begin{bmatrix} -(1 - \hat{\rho}_n^2)\hat{\rho}_s + (1 + \hat{\rho}_s^2)\sqrt{\alpha_1\alpha_2}\hat{\rho}_n - 2\alpha_2\hat{\rho}_s & 1 - \hat{\rho}_n^2 + \alpha_2(1 + \hat{\rho}_s^2) - 2\sqrt{\alpha_1\alpha_2}\hat{\rho}_s\hat{\rho}_n \\ 1 - \hat{\rho}_n^2 + \alpha_1(1 + \hat{\rho}_s^2) - 2\sqrt{\alpha_1\alpha_2}\hat{\rho}_s\hat{\rho}_n & -(1 - \hat{\rho}_n^2)\hat{\rho}_s + (1 + \hat{\rho}_s^2)\sqrt{\alpha_1\alpha_2}\hat{\rho}_n - 2\alpha_2\hat{\rho}_s \end{bmatrix}}{(1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)^2} \\ &\quad \times \frac{\rho_s - \hat{\rho}_s}{(1 - \hat{\rho}_s^2)} \frac{\begin{bmatrix} (1 - \hat{\rho}_n^2) + (1 - \hat{\rho}_s^2)\alpha_2 & \hat{\rho}_s(1 - \hat{\rho}_n^2) + \sqrt{\alpha_1\alpha_2}(1 - \hat{\rho}_s^2)\hat{\rho}_n \\ \hat{\rho}_s(1 - \hat{\rho}_n^2) + \sqrt{\alpha_1\alpha_2}(1 - \hat{\rho}_s^2)\hat{\rho}_n & (1 - \hat{\rho}_n^2) + (1 - \hat{\rho}_s^2)\alpha_1 \end{bmatrix}}{(1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)^2}, \\ &= (\rho_s - \hat{\rho}_s) \frac{\begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}}{[(1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)^2]^2}. \quad (\text{D.16}) \end{aligned}$$

où :

$$\begin{aligned} a_{11} &= 2\sqrt{\alpha_2}(\sqrt{\alpha_1}\hat{\rho}_n - \sqrt{\alpha_2}\hat{\rho}_s) ((1 + \alpha_2) - \hat{\rho}_n(\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)), \\ a_{22} &= 2\sqrt{\alpha_1}(\sqrt{\alpha_2}\hat{\rho}_n - \sqrt{\alpha_1}\hat{\rho}_s) ((1 + \alpha_1) - \hat{\rho}_n(\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)). \end{aligned}$$

Nous détaillons maintenant le calcul du troisième terme de la partie droite de l'équation (D.13). On a :

$$\begin{aligned} \mathbf{H}^T \widehat{\Sigma}^{-1} (\Sigma - \widehat{\Sigma}) \widehat{\Sigma}^{-1} \mathbf{H} &= \frac{\rho_n - \hat{\rho}_n}{(1 - \hat{\rho}_n^2)^2} \begin{bmatrix} \sqrt{\alpha_1} & 0 \\ 0 & \sqrt{\alpha_2} \end{bmatrix} \begin{bmatrix} -2\hat{\rho}_n & (1 + \hat{\rho}_n^2) \\ (1 + \hat{\rho}_n^2) & -2\hat{\rho}_n \end{bmatrix} \begin{bmatrix} \sqrt{\alpha_1} & 0 \\ 0 & \sqrt{\alpha_2} \end{bmatrix}, \\ &= \frac{\rho_n - \hat{\rho}_n}{(1 - \hat{\rho}_n^2)^2} \begin{bmatrix} -2\sqrt{\alpha_1}\hat{\rho}_n & \sqrt{\alpha_1}(1 + \hat{\rho}_n^2) \\ \sqrt{\alpha_2}(1 + \hat{\rho}_n^2) & -2\sqrt{\alpha_2}\hat{\rho}_n \end{bmatrix} \begin{bmatrix} \sqrt{\alpha_1} & 0 \\ 0 & \sqrt{\alpha_2} \end{bmatrix}, \\ &= \frac{\rho_n - \hat{\rho}_n}{(1 - \hat{\rho}_n^2)^2} \begin{bmatrix} -2\alpha_1\hat{\rho}_n & \sqrt{\alpha_1\alpha_2}(1 + \hat{\rho}_n^2) \\ \sqrt{\alpha_1\alpha_2}(1 + \hat{\rho}_n^2) & -2\alpha_2\hat{\rho}_n \end{bmatrix}. \end{aligned}$$

Ensuite :

$$\begin{aligned} \mathbf{M}_{q,q} \mathbf{H}^T \widehat{\Sigma}^{-1} (\Sigma - \widehat{\Sigma}) \widehat{\Sigma}^{-1} \mathbf{H} \mathbf{M}_{q,q} &= \frac{\begin{bmatrix} (1 - \hat{\rho}_n^2) + (1 - \hat{\rho}_s^2)\alpha_2 & \hat{\rho}_s(1 - \hat{\rho}_n^2) + \sqrt{\alpha_1\alpha_2}(1 - \hat{\rho}_s^2)\hat{\rho}_n \\ \hat{\rho}_s(1 - \hat{\rho}_n^2) + \sqrt{\alpha_1\alpha_2}(1 - \hat{\rho}_s^2)\hat{\rho}_n & (1 - \hat{\rho}_n^2) + (1 - \hat{\rho}_s^2)\alpha_1 \end{bmatrix}}{(1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)^2} \\ &\quad \times \frac{\rho_n - \hat{\rho}_n}{(1 - \hat{\rho}_n^2)} \begin{bmatrix} -2\alpha_1\hat{\rho}_n & \sqrt{\alpha_1\alpha_2}(1 + \hat{\rho}_n^2) \\ \sqrt{\alpha_1\alpha_2}(1 + \hat{\rho}_n^2) & -2\alpha_2\hat{\rho}_n \end{bmatrix} \mathbf{M}_{q,q}, \\ &= \frac{\begin{bmatrix} \sqrt{\alpha_1\alpha_2}(1 + \hat{\rho}_n^2)\hat{\rho}_s - (1 - \hat{\rho}_s^2)\alpha_1\alpha_2\hat{\rho}_n - 2\alpha_1\hat{\rho}_n & \sqrt{\alpha_1\alpha_2}(1 + \hat{\rho}_n^2) + \sqrt{\alpha_1\alpha_2}^3(1 - \hat{\rho}_s^2) - 2\alpha_2\hat{\rho}_s\hat{\rho}_n \\ \sqrt{\alpha_1\alpha_2}(1 + \hat{\rho}_n^2) + \sqrt{\alpha_1}^3\sqrt{\alpha_2}(1 - \hat{\rho}_s^2) - 2\alpha_1\hat{\rho}_s\hat{\rho}_n & \sqrt{\alpha_1\alpha_2}(1 + \hat{\rho}_n^2)\hat{\rho}_s - (1 - \hat{\rho}_s^2)\alpha_1\alpha_2\hat{\rho}_n - 2\alpha_2\hat{\rho}_n \end{bmatrix}}{(1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)^2} \\ &\quad \times \frac{\rho_n - \hat{\rho}_n}{(1 - \hat{\rho}_n^2)} \frac{\begin{bmatrix} (1 - \hat{\rho}_n^2) + (1 - \hat{\rho}_s^2)\alpha_2 & \hat{\rho}_s(1 - \hat{\rho}_n^2) + \alpha_1(1 - \hat{\rho}_s^2)\hat{\rho}_n \\ \hat{\rho}_s(1 - \hat{\rho}_n^2) + \sqrt{\alpha_1\alpha_2}(1 - \hat{\rho}_s^2)\hat{\rho}_n & (1 - \hat{\rho}_n^2) + (1 - \hat{\rho}_s^2)\alpha_1 \end{bmatrix}}{(1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)^2}, \\ &= (\rho_n - \hat{\rho}_n) \frac{\begin{bmatrix} b_{11} & b_{12} \\ b_{12} & b_{22} \end{bmatrix}}{[(1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)^2]^2}. \end{aligned} \tag{D.17}$$

où :

$$\begin{aligned} b_{11} &= 2(\sqrt{\alpha_1}\hat{\rho}_n - \sqrt{\alpha_2}\hat{\rho}_s)(-\sqrt{\alpha_1}(1 + \alpha_2) + \sqrt{\alpha_2}\hat{\rho}_s(\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)), \\ b_{22} &= 2(\sqrt{\alpha_2}\hat{\rho}_n - \sqrt{\alpha_1}\hat{\rho}_s)(-\sqrt{\alpha_2}(1 + \alpha_1) + \sqrt{\alpha_1}\hat{\rho}_s(\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)). \end{aligned}$$

Finalement, étant donné que :

$$\mathbf{M}_{p,q}(\mathbf{s}|\mathbf{x}) = \begin{bmatrix} \mathbf{M}_{p,q}(s_1|\mathbf{x}) & \mathbb{E}_p[(s_1 - \hat{s}_{1_q})(s_2 - \hat{s}_{2_q})] \\ \mathbb{E}_p[(s_1 - \hat{s}_{1_q})(s_2 - \hat{s}_{2_q})] & \mathbf{M}_{p,q}(s_2|\mathbf{x}) \end{bmatrix}, \tag{D.18}$$

nous trouvons en utilisant les formules (D.15), (D.16) et (D.17) :

$$\mathbf{M}_{p,q}(s_i|x_1, x_2) = \frac{\alpha_j(1 - \hat{\rho}_s^2) + 1 - \hat{\rho}_n^2}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1\alpha_2}\hat{\rho}_s + \hat{\rho}_n)^2} \tag{D.19}$$

$$\begin{aligned}
& -2\sqrt{\alpha_j}(\rho_s - \hat{\rho}_s)(\sqrt{\alpha_i}\hat{\rho}_n - \sqrt{\alpha_j}\hat{\rho}_s) \frac{(1 + \alpha_j) - \hat{\rho}_n(\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)}{((1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)^2)^2} \\
& + 2(\rho_n - \hat{\rho}_n)(\sqrt{\alpha_i}\hat{\rho}_n - \sqrt{\alpha_j}\hat{\rho}_s) \frac{\alpha_j\sqrt{\alpha_i}(1 - \hat{\rho}_s^2) + (\sqrt{\alpha_i} - \sqrt{\alpha_j}\hat{\rho}_n\hat{\rho}_s)}{((1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1\alpha_2}\hat{\rho}_s)^2)^2}.
\end{aligned}$$

□

D.3 Deux modalités scalaires fortement couplées

Soit le modèle de deux modalités complémentaires suivant :

$$\begin{cases} x_1 = \sqrt{\alpha_1}s + n_1 \\ x_2 = \sqrt{\alpha_2}s + n_2 \end{cases}, \quad (\text{D.20})$$

où s , n_1 et n_2 sont des variables gaussiennes centrées réduites. s est indépendant de n_1 et n_2 . Alors que, n_1 et n_2 sont corrélés, nous notons leurs coefficients de corrélations par ρ_n .

Nous croyons que le coefficient de corrélation des bruits est $\hat{\rho}_n$ au lieu de ρ_n . Dans ce cas, l'estimateur inadéquat de s en fonction de x_1 et x_2 est :

$$\hat{s}_q(x_1, x_2) = \frac{(\sqrt{\alpha_1} - \sqrt{\alpha_2}\hat{\rho}_n)x_1 - (\sqrt{\alpha_1}\hat{\rho}_n - \sqrt{\alpha_2})x_2}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1\alpha_2} + \hat{\rho}_n)^2}. \quad (\text{D.21})$$

L'erreur d'estimation associée à l'estimateur $\hat{s}_q(x_1, x_2)$ est :

$$\begin{aligned}
\mathbf{M}_{p,q}(s|x_1, x_2) &= \frac{1 - \hat{\rho}_n^2}{(\alpha_1 + 1)(\alpha_2 + 1) - (\sqrt{\alpha_1\alpha_2} + \hat{\rho}_n)^2} \\
&+ 2(\rho_n - \hat{\rho}_n) \frac{(\sqrt{\alpha_1}\hat{\rho}_n - \sqrt{\alpha_2})(\sqrt{\alpha_2}\hat{\rho}_n - \sqrt{\alpha_1})}{((1 + \alpha_1)(1 + \alpha_2) - (\hat{\rho}_n + \sqrt{\alpha_1\alpha_2})^2)^2}.
\end{aligned} \quad (\text{D.22})$$

Démonstration. Le modèle des modalités fortement couplées (D.20) est équivalent au modèle des modalités faiblement couplées (D.9), pour $\rho_s = 1$. Par conséquent, l'estimateur $\hat{s}_q(x_1, x_2)$ et l'erreur d'estimation associée $\mathbf{M}_{p,q}(s|x_1, x_2)$ sont obtenus à partir des équations (D.10) et (D.11) pour $\rho_s = 1$. □

Calcul matriciel

Soient \mathbf{A} et \mathbf{B} deux matrices positives semi définies et \mathbf{X} et \mathbf{Y} des matrices quelconques.

E.1 Déterminant

$$\begin{vmatrix} \mathbf{A} & \mathbf{X} \\ \mathbf{Y} & \mathbf{B} \end{vmatrix} = |\mathbf{A}| |\mathbf{B} - \mathbf{Y}\mathbf{A}^{-1}\mathbf{X}| = |\mathbf{B}| |\mathbf{A} - \mathbf{X}\mathbf{B}^{-1}\mathbf{Y}| \quad (\text{E.1})$$

D'après le théorème des déterminants de Sylvester [48]

$$|\mathbf{I} + \mathbf{A}\mathbf{X}^T\mathbf{B}\mathbf{X}| = |\mathbf{I} + \mathbf{B}\mathbf{X}\mathbf{A}\mathbf{X}^T| \quad (\text{E.2})$$

E.2 Inverse

Identités de Woodbury [49] :

$$(\mathbf{A} + \mathbf{X}^T\mathbf{B}\mathbf{X})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{X}^T(\mathbf{X}\mathbf{A}^{-1}\mathbf{X}^T + \mathbf{B}^{-1})^{-1}\mathbf{X}\mathbf{A}^{-1} \quad (\text{E.3})$$

$$\mathbf{B}\mathbf{X}(\mathbf{A} + \mathbf{X}^T\mathbf{B}\mathbf{X})^{-1} = (\mathbf{X}\mathbf{A}^{-1}\mathbf{X}^T + \mathbf{B}^{-1})^{-1}\mathbf{X}\mathbf{A}^{-1} \quad (\text{E.4})$$

E.3 Gradient

Soit f une fonction dérivable de la matrice $\mathbf{X} = (x_{ij}) \in \mathbb{R}^{m \times n}$ alors le gradient de f par rapport à \mathbf{X} est définie par

$$\nabla_{\mathbf{X}} f(\mathbf{X}) = \begin{bmatrix} \frac{\partial}{\partial x_{11}} f(\mathbf{X}) & \cdots & \frac{\partial}{\partial x_{1n}} f(\mathbf{X}) \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_{m1}} f(\mathbf{X}) & \cdots & \frac{\partial}{\partial x_{mn}} f(\mathbf{X}) \end{bmatrix} = \sum_{i=1}^m \sum_{j=1}^n \mathbf{e}_i^T \frac{\partial f(\mathbf{X})}{\partial x_{ij}} \mathbf{e}_j \quad (\text{E.5})$$

Bibliographie

- [1] A. K. Katsaggelos, S. Bahaadini, and R. Molina. Audiovisual fusion : Challenges and new approaches. *Proceedings of the IEEE*, 103(9) :1635–1653, Sept 2015.
- [2] J. Sui, Tülay Adali, Q. Yu, J. Chen, and V. D. Calhoun. A review of multivariate methods for multimodal fusion of brain imaging data. *Journal of Neuroscience Methods*, 204(1) :68 – 81, 2012.
- [3] M. Dalla Mura, S. Prasad, F. Pacifici, P. Gamba, J. Chanussot, and J. A. Benediktsson. Challenges and opportunities of multimodality and data fusion in remote sensing. *Proceedings of the IEEE*, 103(9) :1585–1601, Sept 2015.
- [4] H. McGurk, , and J. MacDonald. Hearing lips and seeing voices. *Nature*, 264 :746–748, 12 1976.
- [5] N. Longbotham, F. Pacifici, T. Glenn, A. Zare, M. Volpi, D. Tuia, E. Christophe, J. Michel, J. Inglada, J. Chanussot, and Q. Du. Multi-modal change detection and application to the detection of flooded areas : Outcome of the 2009-2010 data fusion contest. *IEEE J. Sel. Topics Appl. Earth Observ.*, 5(1) :331–342, Feb. 2012.
- [6] D. Lahat, T. Adali, and C. Jutten. Multimodal data fusion : An overview of methods and challenges and and prospects. *Proceedings of the IEEE*, 103(9) :1449–1477, Sept 2015.
- [7] A. Bourrier, P. O. Amblard, O. Michel, and C. Jutten. Multimodal kalman filtering. In *2016 IEEE International Conference on Acoustics and Speech and Signal Processing (ICASSP)*, pages 4413–4417, March 2016.
- [8] I. Fijalkow, E. Heiman, and H. Messer. Parameter estimation from heterogeneous/multimodal data sets. *IEEE Signal Processing Letters*, 23(3) :390–393, March 2016.
- [9] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3) :379–423, 1948.
- [10] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1) :79–86, 03 1951.
- [11] S. Kullback. *Information Theory And Statistics*. Dover Pubns, 1997.

- [12] H. L. Van Trees. *Detection and Estimation and Modulation Theory : Radar-Sonar Signal Processing and Gaussian Signals in Noise*. Krieger Publishing Co. and Inc., Melbourne and FL and USA, 1992.
- [13] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York and NY and USA, 1991.
- [14] D. Guo, S. Shamai, and S. Verdú. The interplay between information and estimation measures. *Foundations and Trends in Signal Processing*, 6(4) :243–429, 2013.
- [15] A.J. Stam. Some inequalities satisfied by the quantities of information of fisher and shannon. *Information and Control*, 2(2) :101 – 112, 1959.
- [16] D. Guo, S. Shamai, and S. Verdu. Mutual information and minimum mean-square error in gaussian channels. *IEEE Transactions on Information Theory*, 51(4) :1261–1282, April 2005.
- [17] M. Costa and T. Cover. On the similarity of the entropy power inequality and the brunn-minkowski inequality (corresp.). *IEEE Transactions on Information Theory*, 30(6) :837–839, Nov 1984.
- [18] O. Johnson and Y. Suhov. Entropy and random vectors. *Journal of Statistical Physics*, 104(1) :145–165, 2001.
- [19] M. F. Huber, T. Bailey, H. Durrant-Whyte, and U. D. Hanebeck. On entropy approximation for gaussian mixture random vectors. In *2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 181–188, Aug 2008.
- [20] O. Rioul. Information theoretic proofs of entropy power inequalities. *IEEE Transactions on Information Theory*, 57(1) :33–55, Jan 2011.
- [21] D.P. Palomar and S. Verdu. Gradient of mutual information in linear vector gaussian channels. *Information Theory and IEEE Transactions on*, 52(1) :141–154, Jan 2006.
- [22] D. Guo, S. Shamai, and S. Verdu. Mutual information and conditional mean estimation in poisson channels. *IEEE Transactions on Information Theory*, 54(5) :1837–1849, May 2008.
- [23] R. Atar and T. Weissman. Mutual information and relative entropy and estimation in the poisson channel. *IEEE Transactions on Information Theory*, 58(3) :1302–1318, March 2012.
- [24] L. Wang, M. Rodrigues, and L. Carin. Generalized bregman divergence and gradient of mutual information for vector poisson channels. In *Information Theory Proceedings (ISIT) and 2013 IEEE International Symposium on*, pages 454–458, July 2013.
- [25] J. Jiao, K. Venkat, and T. Weissman. Relations between information and estimation in scalar lévy channels. In *2014 IEEE International Symposium on Information Theory*, pages 2212–2216, June 2014.
- [26] C. G. Taborda and F. Pérez-Cruz. Mutual information and relative entropy over the binomial and negative binomial channels. In *Information Theory Proceedings (ISIT) and 2012 IEEE International Symposium on*, pages 696–700, July 2012.

- [27] D. Guo. On information-estimation relationships over binomial and negative binomial models. In *Information Theory Proceedings (ISIT) and 2013 IEEE International Symposium on*, pages 459–463, July 2013.
- [28] C. G. Taborda, D. Guo, and F. Perez-Cruz. Information-estimation relationships over binomial and negative binomial models. *IEEE Transactions on Information Theory*, 60(5) :2630–2646, May 2014.
- [29] B. Widrow and M. E. Hoff. Adaptive switching circuits. In *1960 IRE WESCON Convention Record and Part 4*, pages 96–104, New York, 1960. IRE.
- [30] G. T. Gilbert. Positive definite matrices and sylvester’s criterion. *The American Mathematical Monthly*, 98(1) :44–46, 1991.
- [31] R. H. Berk. Limiting behavior of posterior distributions when the model is incorrect. *Ann. Math. Statist.*, 37(1) :51–58, 02 1966.
- [32] R. H. Berk. Consistency a posteriori. *Ann. Math. Statist.*, 41(3) :894–906, 06 1970.
- [33] P. J. Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability and Volume 1 : Statistics*, pages 221–233, Berkeley and Calif., 1967. University of California Press.
- [34] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6) :716–723, Dec 1974.
- [35] H. White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1) :1–25, 1982.
- [36] W. Xu, A. B. Baggeroer, and K. L. Bell. A bound on mean-square estimation error with background parameter mismatch. *IEEE Transactions on Information Theory*, 50(4) :621–632, April 2004.
- [37] C. D. Richmond and L. L. Horowitz. Parameter bounds on estimation accuracy under model misspecification. *IEEE Transactions on Signal Processing*, 63(9) :2263–2278, May 2015.
- [38] C. Baker. Capacity of the mismatched gaussian channel. *IEEE Transactions on Information Theory*, 33(6) :802–812, Nov 1987.
- [39] K. Yanagi. On the capacity of a class of mismatched gaussian channels (corresp.). *IEEE Transactions on Information Theory*, 30(6) :840–842, Nov 1984.
- [40] S. Nirenberg, S. M. Carcieri, A. L. Jacobs, and P. E. Latham. Retinal ganglion cells act largely as independent encoders. *Journal Nature*, 411 :698–701, 06 2001.
- [41] E. Schneidman, W. Bialek, and M. J. Berry. Synergy and redundancy and independence in population codes. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 23(37) :11539–11553, December 2003.
- [42] S. Verdú. Mismatched estimation and relative entropy. *IEEE Transactions on Information Theory*, 56(8) :3712–3720, Aug 2010.

- [43] M. Chen and J. Lafferty. Mismatched estimation and relative entropy in vector gaussian channels. In *Information Theory Proceedings (ISIT) and 2013 IEEE International Symposium on*, pages 2845–2849, July 2013.
- [44] S. Chlaily, C. Ren, P. O. Amblard, O. Michel, P. Comon, and C. Jutten. Information-estimation relationship in mismatched gaussian channels. *IEEE Signal Processing Letters*, 24(5) :688–692, May 2017.
- [45] S. M. Kay. *Fundamentals of Statistical Signal Processing : Estimation Theory*. Prentice-Hall and Inc., Upper Saddle River and NJ and USA, 1993.
- [46] R. Esposito. On a relation between detection and estimation in decision theory. *Information and Control*, 12(2) :116 – 120, 1968.
- [47] P. Billingsley. *Probability and Measure*. John Wiley and Sons, third edition, 1995.
- [48] J.J. Sylvester M.A. F.R.S. On the relation between the minor determinants of linearly equivalent quadratic functions. *Philosophical Magazine Series 4*, 1(4) :295–305, 1851.
- [49] K. B. Petersen and M. S. Pedersen. *The matrix cookbook*, 2007.
- [50] D. Baudois, C. Serviere, and A. Silvent. Soustraction de bruit : analyse et synthèse bibliographiques. *Traitement du signal*, 6(5) :391 – 497, 1989.
- [51] AP. Dawid. Conditional independence in statistical theory. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–31, 1979.
- [52] S. Verdu and Dongning Guo. A simple proof of the entropy-power inequality. *IEEE Transactions on Information Theory*, 52(5) :2165–2166, May 2006.
- [53] H. Akaike. *Information Theory and an Extension of the Maximum Likelihood Principle*, pages 199–213. Springer New York, New York and NY, 1998.
- [54] S. Chlaily, P. O. Amblard, O. Michel, and C. Jutten. Impact of noise correlation on multimodality. In *2016 24th European Signal Processing Conference (EUSIPCO)*, pages 195–199, Aug 2016.
- [55] S. Chlaily, P-O Amblard, O. Michel, and C.Jutten. L’interaction entre deux modalités complémentaires. In *GRETSI 2017 XXVI éme colloque*, Juan-Les-Pins and France, Sep 2017.

Résumé — Bien que le traitement conjoint des mesures multimodales soit supposé conduire à de meilleures performances que celles obtenues en exploitant une seule modalité ou plusieurs modalités indépendamment, il existe des exemples en littérature qui prouvent que c’est pas toujours vrai. Dans cette thèse, nous analysons rigoureusement, en termes d’information mutuelle et d’erreur d’estimation, les différentes situations de l’analyse multimodale afin de déterminer les conditions conduisant à des performances optimales.

Dans la première partie, nous considérons le cas simple de deux ou trois modalités, chacune étant associée à la mesure bruitée d’un signal, avec des liens entre modalités matérialisés par les corrélations entre les parties utiles du signal et par les corrélations des bruits. Nous montrons comment les performances obtenues sont améliorées avec l’exploitation des liens entre les modalités. Dans la seconde partie, nous étudions l’impact sur les performances d’erreurs sur les liens entre modalités. Nous montrons que ces fausses hypothèses dégradent les performances, qui peuvent alors devenir inférieure à celles atteintes avec une seule modalité.

Dans le cas général, nous modélisons les multiples modalités comme un canal gaussien bruité. Nous étendons alors des résultats de la littérature en considérant l’impact d’erreurs sur les densités de probabilité du signal et du bruit sur l’information transmise par le canal. Nous analysons ensuite cette relation dans la cas d’un modèle simple de deux modalités. Nos résultats montrent en particulier le fait inattendu qu’une double inadéquation du bruit et du signal peuvent parfois se compenser et ainsi conduire à de très bonnes performances.

Mots clés : Multimodalité, modèle inadéquat, double inadéquation, relations information-estimation, bruits corrélés, canal gaussien.

Abstract — The joint processing of multimodal measurements is supposed to lead to better performances than those obtained using a single modality or several modalities independently. However, in literature, there are examples that show that is not always true. In this thesis, we analyze, in terms of mutual information and estimation error, the different situations of multimodal analysis in order to determine the conditions to achieve the optimal performances.

In the first part, we consider the simple case of two or three modalities, each associated with noisy measurement of a signal. These modalities are linked through the correlations between the useful parts of the signal and the correlations between the noises. We show that the performances are improved if the links between the modalities are exploited. In the second part, we study the impact on performance of wrong links between modalities. We show that these false assumptions decline the performance, which can become lower than the performance achieved using a single modality.

In the general case, we model the multiple modalities as a noisy Gaussian channel. We then extend literature results by considering the impact of the errors on signal and noise probability densities on the information transmitted by the channel. We then analyze this relationship in the case of a simple model of two modalities. Our results show in particular the unexpected fact that a double mismatch of the noise and the signal can sometimes compensate for each other, and thus lead to very good performances.

Keywords : Multimodality, mismatched model, double mismatch, information-estimation relationships, correlated noises, gaussian channel.
