

L'interaction entre deux modalités complémentaires

Saloua CHLAILY, Pierre-Olivier AMBLARD, Olivier MICHEL, Christian JUTTEN

GIPSA-Lab, CNRS, Université Grenoble Alpes,
11 rue des Mathématiques, BP46, 38402 Grenoble, France
{saloua.chlaily, pierre-olivier.amblard, olivier.michel,
christian.jutten}@gipsa-lab.grenoble-inp.fr

Résumé – On postule généralement que le traitement conjoint des mesures multimodales conduit à de meilleures performances que celles obtenues en exploitant une seule modalité, ou les deux indépendamment. Mais qu'en est-il vraiment ? A quelles conditions cette affirmation est-elle vérifiée ou au contraire erronée ? Pour y répondre, nous considérons un exemple simple de deux modalités qui produisent des observations bruitées et corrélées via les signaux portant de l'information et les bruits. Dans ce cadre très simple, nous pouvons analyser rigoureusement l'impact de ces corrélations sur les performances en termes d'information et d'erreur d'estimation. Nous montrons également qu'une connaissance erronée de ces corrélations a des effets, parfois tout à fait surprenants, sur les performances.

Abstract – It is generally assumed that the joint processing of multimodal measurements leads to better performance than that obtained by exploiting a single modality, or all modalities independently. Under what conditions is this statement true or false? To answer this, we consider a simple example of two modalities correlated via information-bearing signals and noise. In this very simple framework, we can analyze meticulously the impact of these correlations on performance in terms of information and estimation error. We also show that mismatched correlations has effects, sometimes quite surprising, on performance.

1 Introduction

Il est aujourd'hui courant d'utiliser différents appareils de mesures, c'est-à-dire plusieurs *modalités* pour observer un système complexe. Ce concept est connu sous le nom de *multimodalité*. On affirme généralement que la diversité et la complémentarité de ces données multimodales aboutissent à une meilleure connaissance du système, par exemple à une meilleure performance dans l'estimation des paramètres qui le décrivent.

De façon évidente, la multimodalité soulève plusieurs défis en raison de l'hétérogénéité des données et de la complexité des systèmes de mesure. Suivant l'application, imagerie cérébrale, télédétection, interface homme-machine, etc., nous trouvons dans la littérature de nombreux travaux et algorithmes pour exploiter la multimodalité en affrontant ces défis [1, 2]. Mais, une question fondamentale que peu de travaux ont abordée (par exemple [3, 4, 5, 6]) est la suivante : Quelles conditions garantissent un réel bénéfice à la combinaison de plusieurs modalités ?

L'analyse conjointe de plusieurs mesures garantit en général plus d'information et de meilleures performances que l'analyse séparée, pourvu qu'il existe des liens entre les mesures. Ainsi, la première étape pour répondre à la question ci-dessus consiste à comprendre comment ces modalités interagissent et quel est l'effet de leurs liens mutuels sur leur contenu informationnel. Dans ce travail nous abordons ces questions sur un

exemple simple mais général, qui a l'intérêt de fournir des résultats analytiques faciles à interpréter et à analyser. Ce modèle consiste en deux mesures (c'est-à-dire deux modalités) bruitées d'un même système. Les performances obtenues avec ce modèle, en termes d'information et d'erreur d'estimation, seront étudiées dans deux situations, selon que les liens entre les modalités sont parfaitement ou imparfaitement connus ou modélisés. Dans ce dernier cas, on parlera de modèle *inadéquat*.

Ce papier est organisé comme suit. Le modèle et les notations sont introduits dans la section 2. Dans la section 3, nous rappelons la définition de l'information mutuelle, de l'erreur quadratique moyenne et de leurs généralisations dans le cas d'un modèle inadéquat. L'influence de l'*adéquation* et de l'*inadéquation* du modèle sur les performances attendues est ensuite développée dans les sections 4 et 5, respectivement. Nous concluons ce papier dans la section 6.

2 Modèle et notations

Nous considérons deux modalités différentes qui observent le même système physique :

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \sqrt{\alpha_1} & 0 \\ 0 & \sqrt{\alpha_2} \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix}, \quad (1)$$

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. Nous supposons que les bruits sont indépendants de s_1 et s_2 . Le fait que les deux modalités observent le même système physique est modélisé par une

corrélation entre s_1 et s_2 que nous notons ρ_s . Nous supposons aussi une corrélation entre les bruits que nous notons ρ_n , étant donné que les systèmes d'acquisition peuvent être perturbés par des imperfections liées. Par conséquent, les corrélations ρ_s et ρ_n constituent les liens entre les modalités x_1 et x_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.

Dans le cas général, un modèle est inadéquat si les densités de probabilité supposées, $q(\cdot)$, sont différentes des densités réelles, $p(\cdot)$. Toutefois, dans le cadre du modèle (1) gaussien centré, l'inadéquation se manifeste par la différence entre les vraies valeurs des corrélations ρ_s et ρ_n , et leurs valeurs supposées $\hat{\rho}_s$ et $\hat{\rho}_n$. Les deux modalités sont dites analysées conjointement (resp. séparément) si les liens ρ_s et ρ_n sont (resp. ne sont pas) exploités.

3 Quelques mesures d'information et d'estimation

Dans cette section et dans le cas général, nous examinons l'extension de l'information mutuelle et de l'erreur quadratique moyenne dans le cas de modèles inadéquats.

Information. L'information mutuelle mesure la dépendance entre deux vecteurs aléatoires, autrement dit, l'information partagée entre ces deux quantités. L'information mutuelle entre \mathbf{y} et \mathbf{z} est définie par la divergence de Kullback-Leibler entre la loi jointe $p(\mathbf{y}, \mathbf{z})$ et le produit des marginales $p(\mathbf{y})p(\mathbf{z})$ [7].

$$I_{p,p}(\mathbf{y}; \mathbf{z}) = \int \log \left(\frac{p(\mathbf{y}, \mathbf{z})}{p(\mathbf{y})p(\mathbf{z})} \right) p(\mathbf{y}, \mathbf{z}) d\mathbf{y}d\mathbf{z} \quad (2)$$

$$= \mathbb{E}_p \left[\log \left(\frac{p(\mathbf{y}, \mathbf{z})}{p(\mathbf{y})p(\mathbf{z})} \right) \right] \quad (3)$$

$$= \mathbb{E}_p [\log (V_p(\mathbf{y}, \mathbf{z}))], \quad (4)$$

où $\mathbb{E}_p[\cdot]$ est l'espérance par rapport à $p(\cdot)$. Le terme $V_p(\mathbf{y}, \mathbf{z})$ est un rapport de vraisemblance déterminé par le modèle d'observation. Notons que l'espérance ne peut être approchée qu'à partir des données d'observation distribuées selon la vraie loi $p(\cdot)$. Dans le cas d'un modèle inadéquat, $V_p(\mathbf{y}, \mathbf{z})$ devient $V_q(\mathbf{y}, \mathbf{z})$, mais les données restent distribuées selon la vraie loi $p(\cdot)$. La mesure d'information est alors l'espérance par rapport à $p(\cdot)$ du logarithme de $V_q(\mathbf{y}, \mathbf{z})$, notée :

$$I_{p,q}(\mathbf{y}; \mathbf{z}) = \mathbb{E}_p [\log (V_q(\mathbf{y}, \mathbf{z}))], \quad (5)$$

Contrairement à $I_{p,p}$, $I_{p,q}$ n'est pas une divergence de Kullback-Leibler et par conséquent sa non-négativité n'est pas garantie. La perte en information induite par un modèle inadéquat peut être quantifiée par la différence :

$$I_{p,p}(\mathbf{y}; \mathbf{z}) - I_{p,q}(\mathbf{y}; \mathbf{z}) = D_{KL}(p(\mathbf{y}, \mathbf{z})||q(\mathbf{y}, \mathbf{z})) - D_{KL}(p(\mathbf{y})||q(\mathbf{y})) - D_{KL}(p(\mathbf{z})||q(\mathbf{z})), \quad (6)$$

Cette différence $I_{p,p}(\mathbf{y}; \mathbf{z}) - I_{p,q}(\mathbf{y}; \mathbf{z})$ peut être négative. Par exemple, si $q(\mathbf{z}|\mathbf{y}) = p(\mathbf{z}|\mathbf{y})$ alors que $q(\mathbf{z}) \neq p(\mathbf{z})$, on obtient

$I_{p,p}(\mathbf{y}; \mathbf{z}) - I_{p,q}(\mathbf{y}; \mathbf{z}) = -D_{KL}(p(\mathbf{z})||q(\mathbf{z})) \leq 0$. Dans le modèle (1), si $\mathbf{y} = [s_1, s_2]^T$ et $\mathbf{z} = [x_1, x_2]^T$, ce cas correspond alors à la seule inadéquation du modèle du signal \mathbf{y} .

Estimation. Nous nous intéressons à l'erreur quadratique moyenne qui caractérise la précision d'un estimateur. L'estimateur optimal de \mathbf{y} en fonction de \mathbf{z} au sens de l'erreur quadratique moyenne est la moyenne de la loi a posteriori [8] :

$$\hat{\mathbf{y}}_p(\mathbf{z}) \triangleq \int \mathbf{y}p(\mathbf{y}|\mathbf{z})d\mathbf{y} = \frac{\int p(\mathbf{y})p(\mathbf{z}|\mathbf{y}) \mathbf{y} d\mathbf{y}}{\int p(\mathbf{y})p(\mathbf{z}|\mathbf{y}) d\mathbf{y}}, \quad (7)$$

L'erreur quadratique moyenne relative à cet estimateur est la matrice de covariance de l'erreur $\mathbf{y} - \hat{\mathbf{y}}_p$:

$$\mathbf{M}_{p,p}(\mathbf{y}; \mathbf{z}) = \mathbb{E}_p \left[(\mathbf{y} - \hat{\mathbf{y}}_p(\mathbf{z})) (\mathbf{y} - \hat{\mathbf{y}}_p(\mathbf{z}))^T \right]. \quad (8)$$

Comme pour l'information mutuelle, l'estimateur au sens de l'erreur quadratique moyenne n'est pas adapté au modèle inadéquat. L'estimateur qu'on utilise plutôt est donné par la moyenne de la loi a posteriori supposée $q(\mathbf{y}|\mathbf{z}) \neq p(\mathbf{y}|\mathbf{z})$:

$$\hat{\mathbf{y}}_q(\mathbf{z}) \triangleq \int \mathbf{y}q(\mathbf{y}|\mathbf{z})d\mathbf{y} = \frac{\int q(\mathbf{y})q(\mathbf{z}|\mathbf{y}) \mathbf{y} d\mathbf{y}}{\int q(\mathbf{y})q(\mathbf{z}|\mathbf{y}) d\mathbf{y}}, \quad (9)$$

L'erreur quadratique moyenne relative à cet estimateur est la matrice de covariance de l'erreur $\mathbf{y} - \hat{\mathbf{y}}_q$ par rapport à la vraie loi :

$$\mathbf{M}_{p,q}(\mathbf{y}; \mathbf{z}) = \mathbb{E}_p \left[(\mathbf{y} - \hat{\mathbf{y}}_q(\mathbf{z})) (\mathbf{y} - \hat{\mathbf{y}}_q(\mathbf{z}))^T \right]. \quad (10)$$

Le surplus d'erreur induit par un modèle inadéquat, étant donné par une matrice de covariance, est semi-défini positif :

$$\mathbf{M}_{p,q}(\mathbf{y}; \mathbf{z}) - \mathbf{M}_{p,p}(\mathbf{y}; \mathbf{z}) = \mathbb{E}_p \left[(\hat{\mathbf{y}}_p - \hat{\mathbf{y}}_q) (\hat{\mathbf{y}}_p - \hat{\mathbf{y}}_q)^T \right]. \quad (11)$$

Information versus estimation. Dans le cas gaussien étudié ici, les deux mesures sont équivalentes et nous les utiliserons indifféremment. En revanche, dans un cas général, l'information mutuelle est plus riche que l'erreur quadratique moyenne qui se limite à l'ordre deux. Toutefois, des liens forts existent entre les deux dans les canaux gaussiens [9, 10].

4 Modèle adéquat

Dans cette partie et dans le cadre du modèle (1) adéquat, nous envisageons l'effet des liens et des rapports signal à bruit des modalités sur la précision de l'estimation ainsi que sur l'information.

Une ou deux modalités ? L'information apportée conjointement par x_1 et x_2 sur s_i

$$I_{p,p}(s_i; x_1, x_2) = \frac{1}{2} \log \left(\frac{(1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1 \alpha_2} \rho_s + \rho_n)^2}{1 - \rho_n^2 + \alpha_j(1 - \rho_s^2)} \right) \quad (i \neq j) \quad (12)$$

est plus grande que l'information apportée par x_i

$$I_{p,p}(s_i; x_i) = \frac{1}{2} \log(1 + \alpha_i), \quad (13)$$

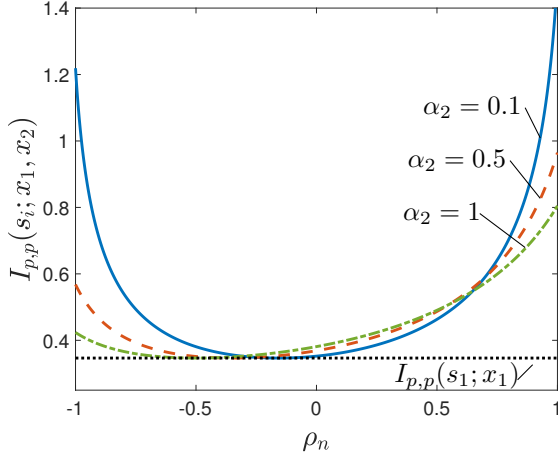


FIGURE 1 – Variation de $I_{p,p}(s_i; x_1, x_2)$ en fonction de ρ_n pour différentes valeurs de α_2 , où $\alpha_1 = 1$ et $\rho_s = -0.5$.

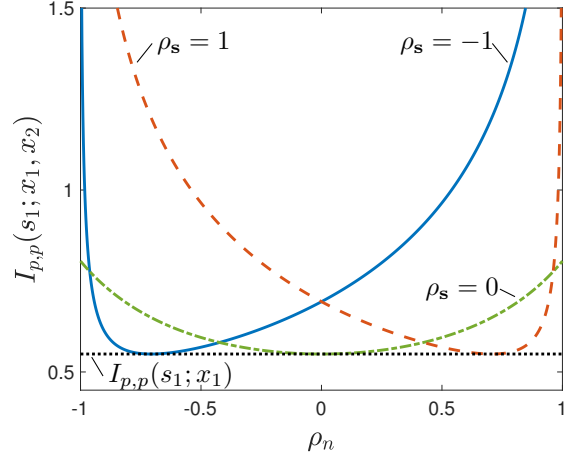


FIGURE 2 – Variations de $I_{p,p}(s_1; x_1, x_2)$ en fonction de ρ_n pour différentes valeurs de ρ_s où $\alpha_1 = 2$ et $\alpha_2 = 1$.

sauf si x_j n'apporte pas de nouveauté sur s_i . Ceci est vrai si $I_{p,p}(s_i; x_1, x_2) = I_{p,p}(s_i; x_i)$ qui correspond à la condition $\sqrt{\alpha_j} \rho_s = \sqrt{\alpha_i} \rho_n$ (l'intersection des courbes avec les lignes horizontales dans les figures 1 et 2). Dans ce cas, x_i est suffisante, seule, pour estimer s_i . Si de plus $\alpha_1 = \alpha_2$, alors x_j est aussi suffisante pour estimer s_j , $I_{p,p}(s_j; x_1, x_2) = I_{p,p}(s_j; x_i)$. Par conséquent, l'analyse séparée est équivalente à l'analyse conjointe si $\alpha_1 = \alpha_2$ et $\rho_s = \rho_n$.

En raison de la complémentarité de x_1 et x_2 , l'exploitation des deux modalités est nécessaire pour avoir de l'information sur s_1 et s_2 . En revanche, si s_1 et s_2 sont parfaitement corrélés ($\rho_s = \pm 1$) et si $\sqrt{\alpha_j} \rho_s = \sqrt{\alpha_i} \rho_n$, alors $I_{p,p}(s_i; x_1, x_2) = I_{p,p}(s_i; x_i)$ et $I_{p,p}(s_j; x_1, x_2) = I_{p,p}(s_j; x_i)$. Une seule modalité x_i suffit donc à tirer le maximum d'information possible sur $\mathbf{s} = [s_1, s_2]^T$.

La corrélation entre les signaux d'intérêt induit-elle de la redondance ou favorise-t-elle l'interaction ? Une corrélation nulle entre les signaux d'intérêt, illustrée par la figure 2 pour $\rho_s = 0$, signifie que les deux modalités observent des parties totalement différentes du système physique.

$$I_{p,p}(s_i; x_1, x_2) \Big|_{\rho_s=0} = \frac{1}{2} \log \left(\frac{(1 + \alpha_1)(1 + \alpha_2) - \rho_n^2}{1 + \alpha_j - \rho_n^2} \right) \geq I_{p,p}(s_i; x_i) \quad (14)$$

Par conséquent, la deuxième modalité x_j n'apporte pas d'information sur le signal s_i , mais elle en apporte sur le bruit (tant que $\rho_n \neq 0$) : l'analyse conjointe des deux modalités apporte donc plus d'information que l'analyse séparée. Si de plus les bruits sont décorrélés, $\rho_n = 0$, les modalités seront indépendantes et aucune interaction ne peut avoir lieu : l'analyse conjointe est, dans ce cas, équivalente à l'analyse séparée.

Réciproquement, une corrélation non nulle entre s_1 et s_2 représente le taux de chevauchement entre les parties observées par x_1 et x_2 . Cette corrélation induit de la redondance entre les modalités et donc moins d'information, $I_{p,p}(s_i; x_1, x_2) \leq$

$I_{p,p}(s_i; x_1, x_2) \Big|_{\rho_s=0}$, si $\rho_s \in [0, \rho_{s_o}]^*$. Par contre, au delà de cet intervalle cette corrélation est exploitée pour mieux réduire le bruit et donc assurer une meilleure estimation surtout si ρ_s et ρ_n ont des signes différents. La figure 2 illustre ces situations extrêmes, pour $\rho_s = -1$ et $\rho_s = 1$. En particulier, nous atteignons une estimation parfaite du signal d'intérêt, $I_{p,p}(s_i; x_1, x_2) = \infty$ et $\mathbf{M}_{p,p}(s_i; x_1, x_2) = 0$, si $|\rho_s| = |\rho_n| = 1$. Une analyse détaillée de ce cas est dans [3].

Doit-on omettre une modalité de très faible rapport signal à bruit ? Si $\alpha_1 = \alpha_2$ ou si $|\rho_s| = |\rho_n|$, les deux modalités s'influencent d'une manière équivalente. Autrement dit, la quantité d'information apportée par x_2 sur s_1 est égale à la quantité apportée par x_1 sur s_2 . Par contre si $\alpha_2 < \alpha_1$, x_1 apporte une information plus fiable sur le signal d'intérêt et x_2 en apporte sur le bruit. Par conséquent, si les signaux d'intérêt sont plus corrélés que les bruits, $|\rho_s| > |\rho_n|$, x_1 apporte plus d'information sur s_2 que celle apportée par x_2 sur s_1 . Inversement si les bruits sont plus corrélés que les signaux d'intérêt, $|\rho_s| < |\rho_n|$, x_2 apporte plus d'information sur s_1 que celle apportée par x_1 sur s_2 . Par ailleurs si $\alpha_2 \ll 1$, l'information que x_2 apporte sur le bruit assure une meilleure estimation de s_1 mais s_2 est mal estimé (cf figure 1).

5 Modèle inadéquat

Dans cette section et dans le cadre du modèle (1) inadéquat, nous examinons les conséquences d'une mauvaise estimation des liens entre les modalités sur la précision d'estimation.

Une ou deux modalités ? Quand nous supposons à tort l'indépendance, $\hat{\rho}_s = 0$ et $\hat{\rho}_n = 0$, ou la redondance entre les modalités, $\sqrt{\alpha_j} \hat{\rho}_s = \sqrt{\alpha_i} \hat{\rho}_n$, nous négligeons l'information que peut porter la deuxième modalité et $\mathbf{M}_{p,q}(s_i; x_1, x_2) = \mathbf{M}_{p,p}(s_i; x_i)$. L'analyse conjointe est donc équivalente à l'ana-

*. $\rho_{s_o} = \frac{2\sqrt{\alpha_1\alpha_2}\rho_n(1+\alpha_j-\rho_n^2)}{\alpha_j(1-\rho_n^2+\alpha_j+\alpha_i\rho_n^2)}$

$$\begin{aligned} \mathbf{M}_{p,q}(s_i; x_1, x_2) &= \frac{1 - \hat{\rho}_n^2 + (1 - \hat{\rho}_s^2)\alpha_j}{(1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1\alpha_2}\hat{\rho}_s + \hat{\rho}_n)^2} \\ &+ (\sqrt{\alpha_i}\hat{\rho}_n - \sqrt{\alpha_j}\hat{\rho}_s) \frac{(1 + \alpha_j)(\sqrt{\alpha_j}(\rho_s - \hat{\rho}_s) - \sqrt{\alpha_i}(\rho_n - \hat{\rho}_n)) + \sqrt{\alpha_j}(\hat{\rho}_s\rho_n - \rho_s\hat{\rho}_n)(\sqrt{\alpha_1\alpha_2}\hat{\rho}_s + \hat{\rho}_n)}{((1 + \alpha_1)(1 + \alpha_2) - (\sqrt{\alpha_1\alpha_2}\hat{\rho}_s + \hat{\rho}_n)^2)} \end{aligned} \quad (15)$$

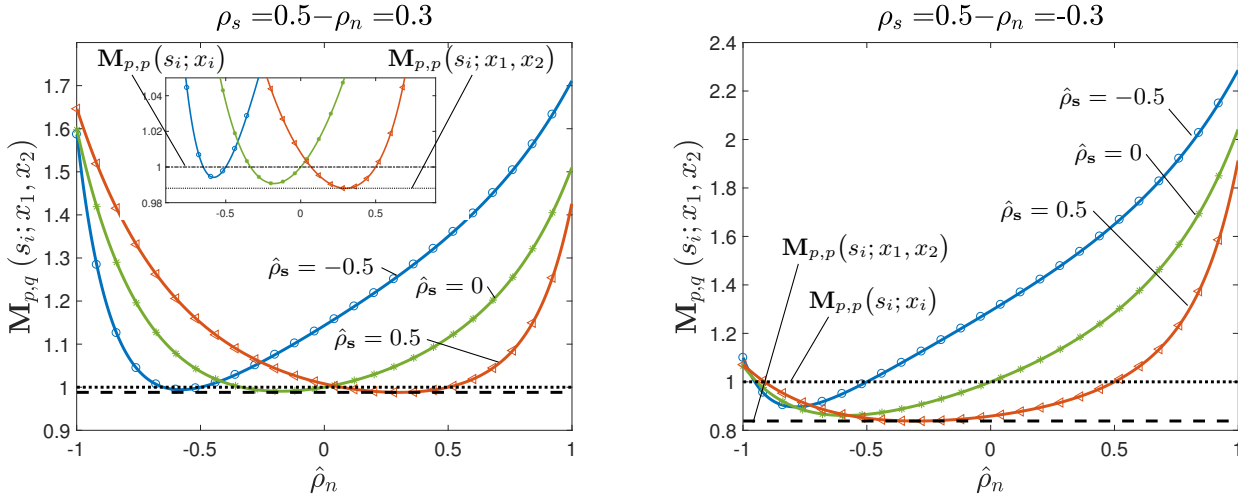


FIGURE 3 – Variations de $\mathbf{M}_{p,q}(s_i; x_1, x_2)$ en fonction de $\hat{\rho}_n$ pour différentes valeurs de $\hat{\rho}_s$ où $\alpha_1 = 1$ et $\alpha_2 = 1$. à gauche $\text{sgn}(\rho_s) = \text{sgn}(\rho_n)$, à droite $\text{sgn}(\rho_s) = -\text{sgn}(\rho_n)$

lyse séparée. Par contre, pour certaines valeurs de $\hat{\rho}_s$ et $\hat{\rho}_n$ le modèle inadéquat introduit plus d'erreurs que l'analyse séparée $\mathbf{M}_{p,q}(s_i; x_1, x_2) > \mathbf{M}_{p,p}(s_i; x_i)$, comme l'illustre la figure 3. Dans ce cas, contrairement au modèle adéquat, l'analyse conjointe apporte moins d'information que l'analyse séparée. Ces erreurs peuvent être très importantes surtout si nous supposons de fortes corrélations, puisque nous accordons à tort confiance à la deuxième modalité. Il faut noter que si les corrélations exactes, ρ_s et ρ_n , ont des signes différents nous risquons moins de surpasser l'erreur de l'analyse séparée (cf la figure 3).

L'intérêt d'une double inadéquation. Une seule inadéquation du modèle, $\hat{\rho}_s \neq \rho_s$ ou $\hat{\rho}_n \neq \rho_n$, n'assure pas moins d'erreurs qu'une double inadéquation, $\hat{\rho}_s \neq \rho_s$ et $\hat{\rho}_n \neq \rho_n$. Au contraire, la figure 3 montre que pour chaque inadéquation du signal $\hat{\rho}_s \neq \rho_s$ correspond une inadéquation de bruit $\hat{\rho}_n \neq \rho_n$ qui la compense et donc minimise $\mathbf{M}_{p,q}(s_i; x_1, x_2)$.

6 Conclusion

Dans cet article, nous avons étudié l'impact des corrélations entre les modalités sur leurs contenus en information. Nos résultats suggèrent que l'analyse conjointe est plus souvent bénéfique que l'analyse séparée, notamment si les corrélations des signaux d'intérêt et des bruits ont des signes différents. En outre, une corrélation entre les bruits améliore la précision d'estimation même si les signaux d'intérêt ne sont pas corrélés ou si une modalité est de mauvaise qualité. De plus, ce travail met en garde contre le modèle inadéquat, puisqu'il risque de détériorer la performance et fausser l'objectif de la multi-

modalité. Comme il met en évidence l'avantage d'une double inadéquation de modèle à réduire l'erreur d'estimation.

Références

- [1] D. Lahat, T. Adali, and C. Jutten. Multimodal data fusion : An overview of methods, challenges, and prospects. *Proceedings of the IEEE*, 103(9) :1449–1477, 2015.
- [2] 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.
- [3] S. Chlailly, 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.
- [4] A. Bourrier, P. O. Amblard, O. Michel, and C. Jutten. Multimodal kalman filtering. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 4413–4417, March 2016.
- [5] A. Aksenov, P. O. Amblard, O. Michel, and C. Jutten. *Optimal Measurement Times for Observing a Brownian Motion over a Finite Period Using a Kalman Filter*, pages 509–518. Springer International Publishing, Cham, 2017.
- [6] I. Fijalkow, E. Heiman, and H. Messer. Parameter estimation from heterogeneous/multimodal data sets. *IEEE Signal Processing Letters*, 23(3) :390–393, March 2016.
- [7] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley-Interscience, New York, NY, USA, 1991.
- [8] S. M. Kay. *Fundamentals of Statistical Signal Processing : Estimation Theory*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.
- [9] D.P. Palomar and S. Verdu. Gradient of mutual information in linear vector gaussian channels. *Information Theory, IEEE Transactions on*, 52(1) :141–154, Jan 2006.
- [10] S. Chlailly, 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.