Sur une généralisation de l'identité de DeBruijn

Steeve ZOZOR et Jean-Marc BROSSIER

GIPSA-Lab, 11 rue des mathématiques, B. P. 46, 38402 Saint Martin d'Hères Cedex, France
steeve.zozor@gipsa-lab.grenoble-inp.fr, jean-marc.brossier@gipsa-lab.grenoble-inp.fr

Résumé – En sortie d'un canal à bruit gaussien additif, l'identité de DeBruijn relie l'entropie de Shannon et l'information de Fisher. En utilisant des $f-$divergences, qui généralisent l'entropie relative au sens de Shannon (divergence de Kullback-Leibler), nous en proposons ici une extension. Celle-ci induit la définition de divergences de Fisher généralisées dont certaines existent déjà dans la littérature. Nous sommes ainsi à même de caractériser un canal bruité à l'aide de mesures générales d'information mutuelle, que ce canal soit gaussien ou non.

Abstract – We propose a generalization of the DeBruijn identity that links the Shannon entropy and the Fisher information for a gaussian channel. The generalization makes use of $f-$divergences, generalizations of the relative Shannon entropy (Kullback-Leibler divergence). The generalized DeBruijn identity induces the definition of generalized Fisher divergences, some of such generalizations existing in the literature. We are then able to characterize a noisy channel using general measures of mutual information, both for gaussian and non gaussian channels.

1 Introduction

Considérons un canal à bruit ξ_t additif, d'entrée X et de sortie $Y_t = X + \xi_t$ (fig. 1). L'objectif de cette contribution est d'étendre l'identité de DeBruijn [1] concernant le cas ξ_t gaussien,

$$\frac{\partial H(Y_t)}{\partial t} = \frac{1}{2} J(Y_t)$$

liant l'entropie différentielle de Shannon $H(Y_t) = -\mathbb{E}\left[\log p_{Y_t}(Y_t)\right]$ de la sortie Y_t (de loi p_{Y_t}) du canal gaussien à son information de Fisher $J(Y_t) = \mathbb{E}\left[\left(\frac{\partial}{\partial t} \log p_{Y_t}(Y_t)\right)^2\right]$.

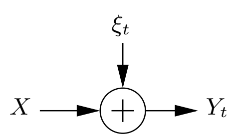

FIGURE 1 – Canal de communication bruité, avec X indépendant du paramètre t et où le bruit ξ_t est indépendant de l'entrée X. Pour un canal gaussien, $\xi_t = \sqrt{t}N$ avec N gaussien centré réduit. Pour un canal de Lévy (voir plus loin), $\xi_t = t^2 L$ avec L Lévy de paramètre d'échelle unité. Pour un canal de Cauchy (voir plus loin), $\xi_t = tC$ avec C Cauchy de paramètre d'échelle unité.

Les mesures $J(Y_t)$ et $H(H_t)$ sont fondamentales dans le domaine des communications ainsi qu'en estimation ; elles quantifient un degré d'incertitude ou d'information contenu dans une variable aléatoire. Comme l'illustre par exemple l'inégalité de Stam [1], ces mesures sont complémentaires. L'information de Fisher joue un rôle important dans l'estimation d'un paramètre θ attaché à une variable aléatoire, que ce soit au travers de l'inégalité de Cramér-Rao, bornant inférieurement la variance d'estimation, ou via la courbure de l'entropie relative (ou divergence de Kullback-Leibler) de la densité de la loi paramétrée par θ, relativement à celle du paramètre "vrai".

L'identité de DeBruijn est importante à plusieurs titres :
(i) en un sens, elle quantifie la perte ou le gain d'entropie de la sortie du canal dû aux variations de la variance du bruit
(ii) elle est utilisée dans la preuve de l'identité de Stam, dans l'inégalité de la puissance entropique ou encore dans celle du théorème de la limite centrale [1, 2, 3].

Nous nous intéressons ici à des mesures généralisées d'entropies relatives à une densité de référence : les $f-$divergences de type Ali-Silvey ou Czisàr [4, 5]. En dépit du caractère fondamental de l'information de Fisher, y compris dans le contexte des $f-$divergences [6], on trouve dans la littérature diverses généralisations de cette mesure d'information, donnant lieu à de possibles extensions des inégalités entropiques classiques liant entropie de Shannon, information de Fisher et loi gaussienne [7].

Dans ce travail, nous nous concentrons tout d'abord sur l'identité de DeBruijn, réécrite en termes de divergence de Kullback-Leibler D_{kl}, rappelant pour cela la définition de la divergence de Fisher. Nous étendons ensuite cette identité dans le cadre de divergences généralisées. Cette extension amène à généraliser la notion de divergence de Fisher. Nous nous focalisons d'abord sur le canal gaussien scalaire pour ensuite exposer quelques résultats dans un cadre plus général.

2 Formulation à l'aide de la divergence de Kullback-Leibler et de l'information de Fisher, en gaussien

Nous considérons ici un canal gaussien où l'entrée X est supposée indépendante du bruit gaussien $\xi_t = \sqrt{t}\,N$ de variance t, et du paramètre t. Le point clé dans la preuve de l'identité de DeBruijn est que la densité de probabilité p de la sortie du canal gaussien vérifie l'équation de la chaleur [1] $\frac{\partial p}{\partial t} = \frac{1}{2} \frac{\partial^2 p}{\partial x^2}$. Cette relation a été reformulée par Barron [2, Lemma 1] ou Johnson [3, Th. C.1] à l'aide de la divergence de Kullback-Leibler $D_{\text{kl}}(p\|g) = \mathbb{E}_g \left[\frac{p(X)}{g(X)} \log\left(\frac{p(X)}{g(X)}\right) \right]$ avec p la loi de la sortie du canal excité par X, g celle de la sortie pour une entrée gaussienne de même variance que X et $\mathbb{E}_g$ l'espérance par rapport à la loi g. De manière plus générale, pour deux entrées $X^{(0)}$ et $X^{(1)}$ et leurs sorties respectives $Y^{(0)}$ (de loi p_0) et $Y^{(1)}$ (de loi p_1), la divergence de Kullback-Leibler entre les lois de sortie vérifie

$$\frac{\partial}{\partial t} D_{\text{kl}}(p_1\|p_0) = -\frac{1}{2} J(p_1\|p_0) \qquad (1)$$

où

$$J(p\|q) = \int \left(\frac{\partial}{\partial x} \log\left(\frac{p(x)}{q(x)}\right) \right)^2 p(x)\, dx \qquad (2)$$

est appelée *divergence non-paramétrique de Fisher* [3, 8]. Plus généralement, pour deux densités p et q paramétrées par un même paramètre θ, on définit la divergence de Fisher paramétrique par

$$J_\theta(p\|q) = \int \left(\frac{\partial}{\partial \theta} \log\left(\frac{p(x)}{q(x)}\right) \right)^2 p(x)\, dx \qquad (3)$$

Quand θ est un paramètre de centrage, dériver en θ ou en x donne le même résultat, aboutissant à $J(p\|q)$.

Dans la suite, deux généralisations de la relation (1) sont envisagées : (i) Pour un canal gaussien, la généralisation à la classe plus large des $f-$divergences, (ii) Généralisation à des canaux non gaussiens.

3 Extension aux $f-$divergences, en gaussien

Rappel sur les $f-$divergences. La $f-$divergence entre p et q relativement à q (Csizár [4], Ali-Silvey [5]) est définie par

$$D_f(p\|q) = \mathbb{E}_q \left[f\left(\frac{p(X)}{q(X)}\right) \right] \qquad (4)$$

où f est une fonction convexe sur $[0; +\infty)$. La classe des $f-$divergences [2] contient comme membres éminents la divergence de Kullback-Leibler, de Jensen-Shannon [4, 9], de Rényi, de Tsallis-Havrda-Charvát ou de Hellinger (à une fonction monotone près) [4, 9, 10, 11], les divergences de Vajda [6, 9], etc. Ces divergences partagent un certain nombre de propriétés qui en font des mesures largement utilisées dans des problèmes de discrimination ou de classification [12] ou en physique statistique [9].

Dans la suite, nous nous limitons au cas des fonctions f convexes C^2, de sorte que la convexité se traduit par $f'' \geq 0$.

Extension aux $f-$divergences, en gaussien. Dans ce cadre, nous nous intéressons à la généralisation de (1) en remplaçant la divergence de Kullback-Leibler par une $f-$divergence. La généralisation à laquelle nous aboutissons fait appel à une généralisation de l'information de Fisher : la *divergence de $f-$Fisher* entre deux densités p et q paramétrées par θ,

$$J_\theta^{(f)}(p\|q) = \int \left(\frac{\partial \log\left(\frac{p(x)}{q(x)}\right)}{\partial \theta} \right)^2 \left(\frac{p(x)}{q(x)}\right)^2 f''\left(\frac{p(x)}{q(x)}\right) q(x) dx \quad (5)$$

où f est une fonction convexe de classe C^2. Lorsque θ est un paramètre de centrage, dériver en θ revient à dériver en x ; dans ce cas, la *divergence de $f-$Fisher non paramétrique* est simplement notée $J^{(f)}$.

Dans le contexte de la divergence de Kullback-Leibler, la divergence de $f-$Fisher qui correspond à la même fonction f est la divergence de Fisher usuelle. Pour Rényi d'indice α, on obtient le gain $\alpha-$Fisher introduit par Hammad dans [8], tandis que pour Jensen-Shannon nous retrouvons la divergence de Jensen-Fisher $J^{JS}(p\|q)$ définie très récemment par Sánchez-Moreno et al. [13] mais seulement par pure analogie avec la divergence de Jensen-Shannon. On notera que certaines de ces divergences de $f-$Fisher peuvent être reliées à des extensions de l'information de Fisher proposées récemment par Johnson et Vignat [14], Lutwak et al. [15] ou Bercher [7].

Revenons alors au canal gaussien (fig. 1) soumis à une entrée $X^{(0)}$ ou $X^{(1)}$, de sorties respectives $Y^{(0)}$ ou $Y^{(1)}$, de densités p_0 ou p_1. La même approche que celle utilisée pour obtenir la relation de DeBruijn aboutit à une relation en tout point similaire à (1),

$$\frac{\partial}{\partial t} D_f(p_1\|p_0) = -\frac{1}{2} J^{(f)}(p_1\|p_0) \qquad (6)$$

avec la même fonction f pour chacun des termes. La relation de DeBruijn permet de quantifier la variation d'entropie de la sortie du canal gaussien induite par une variation de l'amplitude du bruit : l'information de Fisher de la sortie quantifie la robustesse du canal face au bruit. La formulation (6) à l'aide de divergences nous indique que pour deux entrées différentes d'un même canal gaussien, les sorties tendent à se "rapprocher" –au sens de la divergence considérée– quand le niveau de bruit augmente ($J^{(f)} \geq 0$) et la divergence de $f-$Fisher indique à quelle vitesse elles se rapprochent.

Une question naturelle est de savoir si l'identité de DeBruijn s'étend hors du cadre gaussien scalaire. Dans la section suivante, nous présentons une généralisation possible.

4 Quelques résultats en non-gaussien

Considérons deux densités p_0 et p_1, paramétrées par un paramètre t, vérifiant toutes les deux la même équation différentielle

$$\sum_{i=0}^{2} \alpha_i(t) \frac{\partial^i p(x,t)}{\partial t^i} = \beta_2(t) \frac{\partial^2 p(x,t)}{\partial x^2} + \frac{\partial \left(\beta_1(x,t) p(x,t)\right)}{\partial x} \qquad (7)$$

1. On écrit l'entropie de la sortie, que l'on dérive et, moyennant quelques hypothèses de régularité sur les lois mises en jeu, on aboutit à l'identité de DeBruijn [2, 1].

2. Rigoureusement, $D_f = h(\mathbb{E}_q[f(p/q)]$ avec h croissante, de sorte que certaines $f-$divergences usuelles sont fonctions monotones des divergences définies ici.

On montre (cf. annexe), à l'aide des mêmes étapes que celles qui aboutissent à (6), l'identité

$$\sum_{i=0}^{2}\alpha_i(t)\frac{\partial^i D_f(p_1\|p_0)}{\partial t^i} = \alpha_2(t)J_t^{(f)}(p_1\|p_0) - \beta_2(t)J^{(f)}(p_1\|p_0). \tag{8}$$

Cette identité est difficile à interpréter de manière générale. Toutefois, elle contient à elle seule différentes variantes, rencontrées dans la littérature, de l'identité de DeBruijn :

– Canal Lévy, $\xi_t = t^2 L$ (cf. fig. 1). La densité de sortie vérifie (7) avec $\alpha_2 = 1$, $\beta_1 = 2$: pour deux entrées $X^{(0,1)}$, la divergence entre les lois de sortie vérifie $\frac{\partial^2}{\partial t^2}D_f(p_1\|p_0) = J_t^{(f)}(p_1\|p_0)$ (voir [3, Th. 5.5] dans le cas $X^{(0)}$ de Lévy et avec la divergence de Kullback-Leibler). L'information de $f-$Fisher paramétrique caractérise la courbure de la divergence ; son intégrale caractérise la vitesse de rapprochement entre lois pour deux entrées différentes (ou de convergence vers la loi de Lévy si $X^{(0)}$ est de Lévy).

– Canal de Cauchy, $\xi_t = tC$ (cf. fig. 1). L'équation vérifiée par la densité du bruit et de la sortie est encore (7), avec cette fois $\alpha_2 = 1$, $\beta_2 = -1$, aboutissant à $\frac{\partial^2}{\partial t^2}D_f(p_1\|p_0) = J^{(f)}(p_1\|p_0) + J_t^{(f)}(p_1\|p_0)$ (voir [3, Th. 5.6] dans le cas Kullback-Leibler et $X^{(0)}$ de Cauchy). Cette fois, c'est la somme des divergences de $f-$Fisher qui caractérise le canal.

– Ceci se généralise à un bruit quelconque vérifiant (7), à condition que β_1 ne dépende pas de l'état x. C'est le cas par exemple des lois dites $q-$gaussiennes [7].

Un cas particulier intéressant réside dans la relation de Guo-Shamai-Verdu qui caractérise la robustesse du canal gaussien, en termes d'information mutuelle, à des variations du rapport signal-sur-bruit (RSB) [16]. L'entrée du canal X est amplifiée par un facteur $\sqrt{s}$, le bruit étant de variance unité. Le canal est caractérisé par la quantité d'information y transitant à savoir l'information mutuelle entre l'entrée X et la sortie Y, en fonction du RSB s. On montre par exemple facilement que la loi de sortie $p_Y(y)$ et la loi conditionnelle $p_{Y|X=x}$ suivent une même équation du type (7), quelques étapes de calculs aboutissant à

$$\begin{aligned}\frac{\partial}{\partial s}D_f(p_{X,Y}\|p_Xp_Y) &= \frac{1}{2}\mathrm{eqmm}^{(f)}(X|Y)\\ &= \frac{1}{2}\int(x-\mathbb{E}[X|Y=y])^2\,\frac{p_{X,Y}^2}{p_Xp_Y}f''\left(\frac{p_{X,Y}}{p_Xp_Y}\right)\end{aligned}$$

Dans le cas de la divergence de Kullback-Leibler le terme de divergence est l'information mutuelle de Shannon, dont les variations sont caractérisées par l'erreur quadratique minimale de l'estimation de X via la sortie Y comme prouvé dans [16]. Nous trouvons ici une généralisation de cette identité, permettant de caractériser le canal gaussien vis-à-vis du RSB, en termes d'information entrée-sortie généralisée et d'erreur quadratique généralisée.

5 Discussion

Après avoir reformulé l'identité de DeBruijn classique en termes de divergence, nous avons vu comment celle-ci pouvait être généralisée dans plusieurs directions : en restant en gaussien mais en utilisant une classe plus large de divergences ($f-$divergences) et dans une famille de cas non gaussiens qui inclue les canaux de Cauchy et de Levy.

La relation (8) peut s'étendre encore un peu dans le cadre multivarié et à paramètre multivarié. Nous n'entrerons pas dans les détails ici. Le point clé est encore basé sur l'équation (7) que doivent vérifier les deux densités, où les dérivées sont remplacées par des gradients, les dérivées secondes par des hessiennes, et les fonctions α_i et β_i par des opérateurs linéaires, ce qui permet de trouver une relation du type (8) liant gradients et hessiens de la divergence et des matrices $f-$Fisher. De telles relations permettent, par exemple, de généraliser le cas du canal gaussien, ou non, au cas multivarié de matrice de covariance quelconque ou, à l'inverse, de mettre de la corrélation sur les composantes d'entrée (ex. pré-filtrage avant transmission). Nous retrouvons les résultats de [16] comme cas particuliers de l'identité générale.

La question de l'interprétation de l'identité (8) au-delà de la caractérisation d'un canal reste une question ouverte, tout comme les implications de cette identité. Induit-elle, par exemple, des généralisations du théorème de la limite centrale pour des variables sans variances ?

Annexe : preuve de l'identité (8).

Par définition $D_f(p_1\|p_0) = \int f\left(\frac{p_1}{p_0}\right)p_0$. En supposant que l'on puisse permuter intégration et dérivation suivant le paramètre t, nous obtenons alors

$$\frac{\partial D_f}{\partial t} = \int f\left(\frac{p_1}{p_0}\right)\frac{\partial p_0}{\partial t} + \int p_1 f'\left(\frac{p_1}{p_0}\right)\frac{\partial \log\left(\frac{p_1}{p_0}\right)}{\partial t} \tag{9}$$

et par suite

$$\begin{aligned}\frac{\partial^2 D_f(p_1\|p_0)}{\partial t^2} &= \int\left(\frac{\partial\log\left(\frac{p_1}{p_0}\right)}{\partial t}\right)^2\left(\frac{p_1}{p_0}\right)^2 f''\left(\frac{p_1}{p_0}\right)p_0\\ &+\int\left(p_1 f'\left(\frac{p_1}{p_0}\right)\left(\frac{\frac{\partial^2 p_0}{\partial t^2}}{p_0}-\frac{\frac{\partial^2 p_1}{\partial t^2}}{p_1}\right)+f\left(\frac{p_1}{p_0}\right)\frac{\partial^2 p_0}{\partial t^2}\right)\end{aligned}$$

En reconnaissant la divergence de $f-$Fisher dans le premier terme de droite, on obtient donc

$$\begin{aligned}\sum_{i=0}^{2}\alpha_i(t)\frac{\partial^i D_f(p_1\|p_0)}{\partial t^i} &= \alpha_2(t)J_t^{(f)}(p_1\|p_0)\\ &+\int p_1 f'\left(\frac{p_1}{p_0}\right)\sum_{i=0}^{2}\alpha_i(t)\left(\frac{\frac{\partial^i p_1}{\partial t^i}}{p_1}-\frac{\frac{\partial^i p_0}{\partial t^i}}{p_0}\right)\\ &+\int f\left(\frac{p_1}{p_0}\right)\sum_{i=0}^{2}\alpha_i(t)\frac{\partial^i p_0}{\partial t^i}\end{aligned}$$

en ajoutant et enlevant $\alpha_0(t)$ dans la seconde ligne. En injectant alors l'équation différentielle (7) nous obtenons

$$\sum_{i=0}^{2}\alpha_i(t)\frac{\partial^i D_f(p_1\|p_0)}{\partial t^i} = \alpha_2(t)J_t^{(f)}(p_1\|p_0)$$
$$+\int p_1 f'\left(\frac{p_1}{p_0}\right)\sum_{i=1}^{2}\left(\frac{\frac{\partial^i(\beta_i p_1)}{\partial x^i}}{p_1} - \frac{\frac{\partial^i(\beta_i p_0)}{\partial x^i}}{p_0}\right)$$
$$+\int f\left(\frac{p_1}{p_0}\right)\sum_{i=1}^{2}\frac{\partial^i(\beta_i p_0)}{\partial x^i}$$

On remarque alors que

$$\frac{\partial}{\partial x}\left(\beta_1 p_0 f\left(\frac{p_1}{p_0}\right)\right) = \frac{\partial(\beta_1 p_0)}{\partial x} f\left(\frac{p_1}{p_0}\right)$$
$$+p_1\left(\frac{\frac{\partial(\beta_1 p_1)}{\partial x}}{p_1} - \frac{\frac{\partial(\beta_1 p_0)}{\partial x}}{p_1}\right) f'\left(\frac{p_1}{p_0}\right)$$

et donc, en supposant que $\beta_1 p_0 f\left(\frac{p_1}{p_0}\right) \to 0$ sur les bords du domaine d'intégration, nous obtenons

$$\sum_{i=0}^{2}\alpha_i(t)\frac{\partial^i D_f(p_1\|p_0)}{\partial t^i} = \alpha_2(t)J_t^{(f)}(p_1\|p_0)$$
$$+\beta_2\int\left(p_1 f'\left(\frac{p_1}{p_0}\right)\left(\frac{\frac{\partial^2 p_1}{\partial t^2}}{p_1} - \frac{\frac{\partial^2 p_0}{\partial t^2}}{p_0}\right) + f\left(\frac{p_1}{p_0}\right)\frac{\partial^2 p_0}{\partial x^2}\right)$$

On remarque ensuite que

$$\frac{\partial^2}{\partial x^2}\left(p_0 f\left(\frac{p_1}{p_0}\right)\right) = f\left(\frac{p_1}{p_0}\right)\frac{\partial^2 p_0}{\partial x^2}$$
$$+p_1 f'\left(\frac{p_1}{p_0}\right)\left(\frac{\frac{\partial^2 p_1}{\partial x^2}}{p_1} - \frac{\frac{\partial^2 p_0}{\partial x^2}}{p_0}\right)$$
$$+\frac{p_1^2}{p_0}\left(\frac{\partial \log\left(\frac{p_1}{p_0}\right)}{\partial x}\right)^2 f''\left(\frac{p_1}{p_0}\right)$$

et donc, en supposant que $\frac{\partial}{\partial x^2}\left(p_0 f\left(\frac{p_1}{p_0}\right)\right) \to 0$ sur les bords du domaine d'intégration, nous aboutissons à

$$\sum_{i=0}^{2}\alpha_i(t)\frac{\partial^i D_f(p_1\|p_0)}{\partial t^i} = \alpha_2(t)J_t^{(f)}(p_1\|p_0)$$
$$-\beta_2\int\frac{p_1^2}{p_0}\left(\frac{\partial \log\left(\frac{p_1}{p_0}\right)}{\partial x}\right)^2 f''\left(\frac{p_1}{p_0}\right)$$

qui n'est rien d'autre que la relation (8), le dernier terme étant la divergence de $f-$Fisher non paramétrique.

Références

[1] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Info. and Control*, 2(2) :101–112, June 1959.

[2] A. R. Barron. Entropy and the central limit theorem. *The Annals of Probability*, 14(1) :336–342, January 1986.

[3] O. Johnson. *Information Theory and The Central Limit Theorem*. Imperial college Press, London, 2004.

[4] I. Csizàr. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2 :299–318, 1967.

[5] S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *J. of the Royal Stat. Society B*, 28(1) :131–142, 1966.

[6] I. Vajda. χ^α-divergence and generalized Fisher's information. In *Transactions of the 6th Prague Conference on Information Theory, Statistitics, Decision Functions and Random Processes*, pages 873–886, 1973.

[7] J.-F. Bercher. On a (β, q)-generalized Fisher information and inequalities invoving q-Gaussian distributions. *J. of Math. Phys.*, 53(6) :063303, june 2012.

[8] P. Hammad. Mesure d'ordre α de l'information au sens de Fisher. *Revue de Statistique Appliquée*, 26(1) :73–84, 1978.

[9] F. Liese and I. Vajda. On divergence and informations in statistics and information theory. *IEEE Trans. on Information Theory*, 52(10) :4394–4412, October 2006.

[10] A. Rényi. On measures of entropy and information. *in Proc. of the 4th Berkeley Symposium on Math. Stat. and Proba.*, 1 :547–561, 1961.

[11] J. Havrda and F. Charvát. Quantification method of classification processes : Concept of structural α-entropy. *Kybernetica*, 3 :30–35, 1967.

[12] M. Basseville. Distance measures for signal processing and pattern recognition. *Signal Processing*, 18(4) :349–369, December 1989.

[13] P. Sánchez-Moreno, A. Zarzo, and J. S. Dehesa. Jensen divergence based on Fisher's information. *J. of Phys. A*, 45(12) :125305, March 2012.

[14] O. Johnson and C. Vignat. Some results concerning maximum Rényi entropy distributions. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 43(3) :339–351, May-June 2007.

[15] E. Lutwak, S. Lv, D. Yang, and G. Zhang. Extension of Fisher information and Stam's inequality. *IEEE Trans. on Information Theory*, 58(3) :1319–1327, March 2012.

[16] D. P. Palomar and S. Verdú. Gradient of mutual information in linear vector Gaussian channels. *IEEE Trans. on Information Theory*, 52(1) :141–154, January 2006.