Sur l'extremum d'entropie pour les transformations linéaires

Olivier RIOUL

LTCI, Télécom ParisTech, Université Paris Saclay, 75013, Paris, France
olivier.rioul@telecom-paristech.fr

Résumé – Comme pour le fameux principe de maximum d'entropie, le principe de *minimum* d'entropie permet de mesurer le caractère gaussien d'un signal. On présente une nouvelle démonstration très simple de ce principe appliqué aux transformations linéaires, qui s'exprime par une inégalité de puissance entropique initialement due à Shannon et généralisée par Zamir et Feder. Cette démonstration est basée sur un argument de transport optimal qui prend la forme d'un simple changement de variables et permet d'établir facilement la condition d'égalité.

Abstract – Like the well-known maximum entropy principle, the *minimum* entropy principle makes it possible to measure the Gaussian character of a signal. We present a new, simple proof of this minimum entropy principle applied to linear transformations, which is expressed by an entropy power inequality originally due to Shannon and generalized by Zamir and Feder. The proof is based on an optimal transport argument which takes the form of a simple change of variables and easily settles the equality case.

1 Introduction

Considérons un signal aléatoire X de longueur n dont les composantes $X_1, X_2, \ldots, X_n$ sont des variables aléatoires continues statistiquement *indépendantes*. Par commodité on représentera X comme un vecteur aléatoire sous forme de *vecteur colonne* $X = (X_1, X_2, \ldots, X_n)^t$. Ce vecteur X modélise par exemple la donnée de sources mutuellement indépendantes, ou, dans le cas d'un processus i.i.d. où les échantillons X_i suivent la même loi, un bruit blanc ou une source sans mémoire. En général le signal X n'est pas gaussien.

On considère dans cet article l'entropie différentielle [8, § 1.3] $H(\mathbf{A}X)$ du signal après une transformation linéaire *quelconque* donnée par une matrice $\mathbf{A} = ((a_{i,j}))_{i,j}$ de taille $m \times n$. Plus précisément, on s'intéresse à évaluer le caractère *gaussien* du signal à l'aide de cette entropie. Pour ce faire, on va comparer $H(\mathbf{A}X)$ à $H(\mathbf{A}X')$ où X' est un signal gaussien. On peut d'ores et déjà faire la remarque suivante.

Lemme 1. *Si $\mathbf{A}$ est de rang $= m$, alors $H(\mathbf{A}X) - H(\mathbf{A}X')$ est invariant par opérations élémentaires sur les lignes de $\mathbf{A}$. Sinon, on a trivialement $H(\mathbf{A}X) = H(\mathbf{A}X') = -\infty$.*

On supposera désormais que $\mathbf{A}$ est de rang $m \leqslant n$ pour éviter le cas trivial. On pourra donc toujours supposer, par élimination de Gauss, que $\mathbf{A}$ est sous forme canonique (échelonnée réduite).

Démonstration. Par définition de l'entropie différentielle [8], si le rang des lignes de $\mathbf{A}$ n'est pas plein, il existe une relation de dépendance linéaire entre deux composantes de $\mathbf{A}X$ de sorte que $H(\mathbf{A}X) = -\infty$. Sinon, en notant $Y = \mathbf{A}X$ de densité f_Y, l'entropie de Y est l'intégrale $H(Y) = -\int f_Y(y) \log f_Y(y)\, \mathrm{d}y$. Pour tout matrice inversible $\mathbf{P}$ de taille $m \times m$, un changement de variable montre que $H(\mathbf{P}Y) = H(Y) + \log|\det \mathbf{P}|$. Le terme $\log|\det \mathbf{P}|$ disparaît dans la différence $H(\mathbf{PA}X) - H(\mathbf{PA}X') = H(\mathbf{A}X) - H(\mathbf{A}X')$, qui est ainsi de même valeur pour des matrices $\mathbf{A}$ ligne-équivalentes. □

1.1 Énoncés des deux principes entropiques

Il existe deux principes d'extremum d'entropie qui peuvent être vus comme duaux :

Principe du maximum d'entropie (MaxEnt). *Notons $X^* = (X_1^*, X_2^*, \ldots, X_n^*)^t$ un signal gaussien dont les composantes indépendantes ont même variance que X :*

$$\mathrm{Var}(X_i^*) = \mathrm{Var}(X_i) \qquad (i = 1, 2, \ldots, n)$$

Alors

$$H(\mathbf{A}X) \leqslant H(\mathbf{A}X^*) \tag{1}$$

avec égalité si et seulement si X est gaussien.

Ce principe est bien connu depuis le XIXe siècle où il est apparu en mécanique statistique avant d'être préconisé pour une utilisation plus large par Jaynes [6] et appliqué à l'estimation spectrale par Burg [2]. C'est une inégalité fondamentale de la théorie de l'information qui se généralise d'ailleurs au cas où les composantes du signal sont dépendantes ; elle se démontre simplement en remarquant que la différence $H(\mathbf{A}X^*) - H(\mathbf{A}X)$ est une divergence de Kullback-Leibler (entropie relative) qui est toujours positive [8, Chap. 4], [3, Chap. 11]. Le principe suivant est moins connu :

Principe du minimum d'entropie (MinEnt). *Notons $\widetilde{X} = (\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_n)^t$ un signal gaussien dont les composantes indépendantes ont même entropie que X :*

$$H(\widetilde{X}_i) = H(X_i) \qquad (i = 1, 2, \ldots, n)$$

Alors

$$H(\mathbf{A}X) \geqslant H(\mathbf{A}\widetilde{X}) \tag{2}$$

avec égalité si et seulement si ou bien X est gaussien, ou bien la transformation $\mathbf{A}$ est "triviale" au sens où sous forme canonique, $\mathbf{A}$ contient les m colonnes de la matrice identité, les autres colonnes étant nulles.

Dans ce dernier cas, appliquer $\mathbf{A}$ revient simplement à sélectionner m composantes $i_1, i_2, \ldots, i_m$ parmi n, et on a trivialement $H(\mathbf{A}X) = H(X_{i_1}, \ldots, X_{i_m}) = H(X_{i_1}) + \ldots + H(X_{i_m}) = H(\widetilde{X}_{i_1}) + \ldots + H(\widetilde{X}_{i_m}) = H(\widetilde{X}_{i_1}, \widetilde{X}_{i_2}, \ldots, \widetilde{X}_{i_m}) = H(\mathbf{A}\widetilde{X})$.

Noter que le résultat énoncé ci-dessus est évident si $m = n$ par la relation $H(\mathbf{A}X) = H(X) + \log|\det \mathbf{A}| = H(\widetilde{X}) + \log|\det \mathbf{A}| = H(\mathbf{A}\widetilde{X})$ (auquel cas la forme canonique de $\mathbf{A}$ est l'identité $\mathbf{I}$) ; on supposera désormais que $m < n$.

1.2 Historique et applications

L'inégalité (2) est énoncée dans le cas $(m, n) = (1, 2)$ dès 1948 par Shannon dans son article fondateur de la théorie de l'information [12] avant d'être généralisée en 1993 pour (m, n) quelconque par Zamir et Feder [16]. Elle est connue sous le nom d'*inégalité de la puissance entropique* pour la raison que lorsque les signaux sont centrés, alors que (1) permet de maximiser l'entropie à puissance donné, (2) permet de minimiser l'entropie à *puissance entropique* donnée, où la puissance entropique est définie comme la puissance d'un bruit gaussien de même entropie. Dans les deux cas (1), (2), l'extremum est atteint pour un signal gaussien, mais pour des contraintes différentes.

L'inégalité de la puissance entropique est un outil devenu classique pour borner des régions de capacité ou de distorsion associées à des problèmes de codage faisant intervenir des sources ou des bruits gaussiens. Pour toutes ces applications, les variables X_i sont toutes gaussiennes sauf une [9]. Dans le cas général de variables X_i quelconques, Donoho [4] a montré que le principe MinEnt permet de caractériser élégamment la *proximité à la normalité* qui croît par transformation linéaire (en particulier par filtrage linéaire). Ce résultat est établi rigoureusement pour la divergence à la normalité comme une conséquence immédiate de (2) dans [16, Théorème 2]. Pour X *non* gaussien, $H(\mathbf{A}X)$ est minimum uniquement dans le cas d'un mélange linéaire $\mathbf{A}$ trivial, ce qui permet de déconvoluer l'effet du système linéaire [1, 4] ou d'effectuer une séparation aveugle de sources indépendantes [1, 14].

1.3 Contribution de cette étude

La démonstration de (2) est beaucoup plus délicate que celle de (1). Elle résulte classiquement d'une intégration sur un chemin de perturbation gaussienne additive, à l'aide de l'identité de Bruijn et d'une inégalité portant sur l'information de Fisher [15]. Une variante utilise une identité reliant information mutuelle et erreur quadratique moyenne minimale [5] ; dans le cas d'un bruit gaussien additif l'information de Fisher et l'erreur quadratique moyenne minimale se trouvent être des quantités complémentaires [7]. L'auteur a déjà proposé une preuve plus simple au moyen d'une inégalité satisfaite par l'information mutuelle [9]. Néanmoins, la complexité des démonstrations disponibles rend jusqu'à présent le principe de minimum d'entropie assez mystérieux. De plus, la condition d'égalité énoncée est clairement suffisante [16], mais son caractère nécessaire n'a jamais été démontré jusqu'ici dans le cas général.

Dans cet article, on étend les résultats récents de [11] pour établir une nouvelle démonstration très simple de (2) ainsi que du cas d'égalité. Alors que [11] se focalise sur une autre inégalité équivalente uniquement dans le cas $(m, n) = (1, 2)$, l'argument développé ici donne une preuve directe de (2) dans le cas général. Il s'agit d'un argument de "transport" qui peut s'interpréter comme un transport optimal pour un problème de Monge-Kantorovitch [13] mais il n'est nul besoin de connaître la théorie du transport optimal pour comprendre la preuve : l'argument de transport prend la forme d'un simple changement de variables réelles $X_i = T_i(\widetilde{X_i})$ pour $i = 1$ à n.

2 Une nouvelle preuve du principe de minimum d'entropie

Elle se fonde sur quelques lemmes connus.

Lemme 2 (Transport)**.** *Pour toutes variables aléatoires continues X et Y, il existe une transformation croissante $T : \mathbb{R} \to \mathbb{R}$ telle que Y a la même loi que $T(X)$.*

Ce résultat est bien connu dans le cas où X suit une distribution uniforme sur $[0, 1]$: c'est la méthode de l'inversion de la fonction de répartition.

Démonstration. On choisit $T = F_Y^{-1} \circ F_X$ où F_X (resp. F_Y) désigne la fonction de répartition de X (resp. Y) et où l'inverse est prise éventuellement au sens généralisé $F^{-1}(\omega) = \inf\{x \in \mathbb{R} \mid F(x) \geqslant \omega\}$. La loi de $T(X)$ est caractérisée par sa fonction de répartition : $\mathbb{P}(T(X) \leqslant y) = \mathbb{P}(F_X(X) \leqslant F_Y(y)) = F_X(F_X^{-1}(F_Y(y)) = F_Y(y)$. □

On peut généraliser le lemme 2 aux vecteurs aléatoires [11, Lemme 1], mais ce n'est pas nécessaire ici.

Lemme 3 (Changement de variable dans l'entropie)**.** *Soit $\mathbf{T} : \mathbb{R}^n \to \mathbb{R}^n$ un difféomorphisme appliqué au vecteur aléatoire X dont la matrice jacobienne est noté $\mathbf{T}'(X)$. On a :*

$$H\big(\mathbf{T}(X)\big) = H(X) + \mathbb{E}\log|\det \mathbf{T}'(X)| \tag{3}$$

Cela résulte d'un simple changement de variable $Y = \mathbf{T}(X)$ de jacobien $\det \mathbf{T}'(X)$ dans l'intégrale définissant l'entropie de $\mathbf{T}(X)$. Le détail de la preuve est confié au lecteur (voir [11, Lemme 2]). Le cas d'une transformation linéaire $\mathbf{T}(X) = \mathbf{P}X$ a déjà été vu dans la démonstration du Lemme 1.

Noter que le lemme 3 s'étend immédiatement au cas d'une entropie *conditionnelle* pour une transformation T_Z qui dépend d'une autre variable Z :

$$H\big(\mathbf{T}_Z(X)|Z\big) = H(X|Z) + \mathbb{E}\log|\det \mathbf{T}'_Z(X)| \tag{4}$$

en appliquant (3) à $Z = z$ fixé puis en moyennant sur Z [10, Lemme 3].

Lemme 4 (Inégalité de concavité du logarithme généralisée)**.** *Si les lignes de* $\mathbf{A}$ *(de taille* $m \times n$*) sont orthonormées (*$\mathbf{A}\mathbf{A}^t = \mathbf{I}$*), pour toute matrice diagonale* $\mathbf{\Lambda}$ *à* n *éléments diagonaux* $\lambda_i > 0$*,*

$$\log\det(\mathbf{A}\mathbf{\Lambda}\mathbf{A}^t) \geqslant \operatorname{tr}(\mathbf{A}\log(\mathbf{\Lambda})\mathbf{A}^t) \tag{5}$$

où $\log(\mathbf{\Lambda})$ *est la matrice diagonale dont les* n *éléments diagonaux sont les* $\log\lambda_i$*.*

Ce lemme apparaît dans [15] et [5]. On en donne ici une preuve simple pour être exhaustif.

Démonstration. Remarquons que pour tout matrice orthogonale $\mathbf{U}$ de taille $m \times m$, les deux membres de (5) ne changent pas si on remplace $\mathbf{A}$ par $\mathbf{U}\mathbf{A}$. En prenant $\mathbf{U}$ la matrice des vecteurs propres de la matrice symétrique $\mathbf{A}\mathbf{\Lambda}\mathbf{A}^t$ – de sorte que $\mathbf{U}\mathbf{A}\mathbf{\Lambda}\mathbf{A}^t\mathbf{U}^t$ est diagonale – on voit qu'on peut se limiter au cas où $\mathbf{A}\mathbf{\Lambda}\mathbf{A}^t$ est diagonale. Ses éléments diagonaux sont $\sum_j a_{i,j}^2\lambda_j$ et on a bien $\log\prod_i\sum_j a_{i,j}^2\lambda_j = \sum_i\log\sum_j a_{i,j}^2\lambda_j \geqslant \sum_i\sum_j a_{i,j}^2\log\lambda_j$ par concavité du logarithme. □

Nous disposons maintenant de tous les ingrédients de la démonstration de l'inégalité (2) :

Démonstration de l'inégalité (2)*.* On peut tout d'abord supposer que toutes les composantes de X ont même entropie. Si ce n'est pas le cas, on peut toujours les diviser par des constantes δ_i de sorte que les X_i/δ_i soient de même entropie (par exemple $\delta_i = \exp H(X_i)$), puis appliquer le résultat à la matrice $\mathbf{A}\Delta$ où Δ est la matrice diagonale dont les éléments diagonaux sont les δ_i. Il en résulte que les composantes de $\widetilde{X}$ sont aussi de même entropie, donc de même variance ; on se ramène ainsi au cas où ces composantes sont i.i.d.

Le procédé de Gram-Schmidt appliqué aux lignes de la matrice $\mathbf{A}$ transforme cette matrice en une matrice ligne-équivalente dont les lignes sont orthonormées. D'après le Lemme 1, on peut donc également se restreindre à ce cas ($\mathbf{A}\mathbf{A}^t = \mathbf{I}$).

Étendons alors $\mathbf{A}$ en rajoutant $n - m$ lignes orthonormées (dans une matrice complémentaire $\mathbf{A}^c$) de façon à obtenir une matrice orthogonale $\mathbf{A}'$ de taille $n \times n$:

$$\mathbf{A}' = \begin{pmatrix} \mathbf{A} \\ \hline \mathbf{A}^c \end{pmatrix} \tag{6}$$

Posons alors $\widetilde{Y} = \mathbf{A}\widetilde{X}$ et $\widetilde{Y}^c = \mathbf{A}^c\widetilde{X}$ de sorte que

$$\widetilde{Y}' = \begin{pmatrix} \widetilde{Y} \\ \widetilde{Y}^c \end{pmatrix} = \mathbf{A}'\widetilde{X} \tag{7}$$

Puisque les espaces engendrés par les lignes de $\mathbf{A}$ et de $\mathbf{A}^c$ sont orthogonaux et que les composantes de $\widetilde{X}$ ont même variance, les vecteurs gaussiens $\widetilde{Y}$ et $\widetilde{Y}^c$ sont décorrélés et indépendants. En inversant $\mathbf{A}'$ on obtient

$$\widetilde{X} = \mathbf{A}'^t\,\widetilde{Y}' \tag{8}$$

Le lemme 2 appliqué à chaque composante de X permet de supposer que $X_i = T_i(\widetilde{X}_i)$ où T_i croissante pour chaque $i = 1$ à n. En notant $\mathbf{T}(\widetilde{X})$ le vecteur de composantes $T_i(\widetilde{X}_i)$, il vient

$$\begin{aligned} H(\mathbf{A}X) &= H(\mathbf{A}\,\mathbf{T}(\widetilde{X})) && (9)\\ &= H(\mathbf{A}\,\mathbf{T}(\mathbf{A}'^t\,\widetilde{Y}')) && (10)\\ &\geqslant H(\mathbf{A}\,\mathbf{T}(\mathbf{A}'^t\,\widetilde{Y}')\,|\,\widetilde{Y}^c) && (11) \end{aligned}$$

la dernière inégalité résultant du fait que le conditionnement réduit l'entropie [3, 8]. Pour $\widetilde{Y}^c$ fixé,

$$\mathbf{T}_{\widetilde{Y}^c}(\widetilde{Y}) = \mathbf{A}\,\mathbf{T}(\mathbf{A}'^t\,\widetilde{Y}') = \mathbf{A}\,\mathbf{T}(\mathbf{A}^t\,\widetilde{Y} + (\mathbf{A}^c)^t\,\widetilde{Y}^c) \tag{12}$$

est une fonction de $\widetilde{Y}$ dont la matrice jacobienne vaut

$$\mathbf{T}'_{\widetilde{Y}^c}(\widetilde{Y}) = \mathbf{A}\mathbf{T}'(\mathbf{A}'^t\,\widetilde{Y}')\mathbf{A}^t = \mathbf{A}\mathbf{T}'(\widetilde{X})\mathbf{A}^t \tag{13}$$

où $\mathbf{T}'(\widetilde{X})$ est diagonale d'éléments diagonaux $T_i'(\widetilde{X}_i)$.

Le lemme 3 (équation (4)) appliqué à (11) donne

$$\begin{aligned} H(\mathbf{A}X) &\geqslant H(\mathbf{A}\,\mathbf{T}(\mathbf{A}'^t\,\widetilde{Y}')\,|\,\widetilde{Y}^c) \\ &= H(\widetilde{Y}|\widetilde{Y}^c) + \mathbb{E}\log\det(\mathbf{A}\mathbf{T}'(\widetilde{X})\mathbf{A}^t) && (14) \end{aligned}$$

où l'on peut simplifier $H(\widetilde{Y}|\widetilde{Y}^c) = H(\widetilde{Y}) = H(\mathbf{A}\widetilde{X})$ par indépendance de $\widetilde{Y}$ et $\widetilde{Y}^c$.

Enfin le lemme 4 s'applique au deuxième terme de (14) :

$$\begin{aligned} H(\mathbf{A}X) &\geqslant H(\mathbf{A}\widetilde{X}) + \mathbb{E}\log\det(\mathbf{A}\mathbf{T}'(\widetilde{X})\mathbf{A}^t) \\ &\geqslant H(\mathbf{A}\widetilde{X}) + \mathbb{E}\operatorname{tr}(\mathbf{A}\cdot\log\mathbf{T}'(\widetilde{X})\cdot\mathbf{A}^t) && (15)\\ &= H(\mathbf{A}\widetilde{X}) + \operatorname{tr}(\mathbf{A}\cdot\mathbb{E}\log\mathbf{T}'(\widetilde{X})\cdot\mathbf{A}^t) && (16) \end{aligned}$$

par linéarité de la trace. Mais le lemme 3 appliqué à T_i donne $\mathbb{E}\log T_i'(\widetilde{X}_i) = H(T_i(\widetilde{X}_i)) - H(\widetilde{X}_i) = H(X_i) - H(\widetilde{X}_i)$ qui s'annule pour tout $i = 1$ à n par hypothèse, par conséquent $\mathbb{E}\log\mathbf{T}'(\widetilde{X})$ égale la matrice nulle dans (16), ce qui démontre (2). □

3 Le cas d'égalité

La méthode exposée ici permet d'obtenir facilement la condition nécessaire et suffisante d'égalité dans (2). D'après la preuve ci-dessus, l'égalité a lieu si et seulement s'il y a égalité à la fois dans (11) et dans (15).

L'égalité dans (15) revient à une égalité presque sûre dans le lemme 4 lorsque les λ_i valent $T_i'(\widetilde{X}_i)$. Ici $\mathbf{A}$ est mise sous une forme telle que ses lignes sont orthonormées ; elle est éventuellement multipliée à gauche par une matrice orthogonale $\mathbf{U}$ ce qui ne change pas cette propriété. D'après la preuve du lemme 4 qui utilise la convexité stricte du logarithme, on aura égalité seulement dans un des deux cas suivants :

1. chaque ligne $(a_{i,j})_j$ correspond à une combinaison linéaire triviale, ce qui revient à dire que la transformation $\mathbf{A}$ est “triviale” au sens indiqué ci-dessus après l'équation (2) ;
2. les λ_j sont tous égaux, ce qui revient à dire que les $T_i'(X_i)$ sont égaux p.s. pour chaque valeur de $i = 1$ à n.

Dans ce dernier cas, les X_i étant indépendants, cela implique que les $T_i'(X_i)$ sont égaux *et constants* p.s., de sorte que chaque transformation T_i est linéaire. Il en résulte bien que les $X_i = T_i(\widetilde{X}_i)$ sont tous *gaussiens*. De plus, comme $\mathbf{T}$ est linéaire (proportionnelle à l'identité), le vecteur $\widetilde{Y}^c$ disparaît dans l'expression

$$\mathbf{A}\,\mathbf{T}(\mathbf{A}'^t\,\widetilde{Y}') \propto \mathbf{A}(\mathbf{A}^t\,\widetilde{Y} + (\mathbf{A}^c)^t\,\widetilde{Y}^c) = \widetilde{Y} \qquad (17)$$

ce qui implique bien que (11) est une égalité.

En résumé on a bien démontré qu'il y a égalité dans (2) seulement si X est gaussien ou si la transformation $\mathbf{A}$ est triviale. Il s'agit bien de la condition énoncée énoncée ci-dessus pour le principe MinEnt de minimum d'entropie.

4 Ouvertures

La nouvelle démonstration du principe de minimum d'entropie exposée ici a le mérite, outre sa simplicité, de fournir un terme additionnel $\mathrm{tr}(\mathbf{A}\cdot\mathbb{E}\log\mathbf{T}'(\widetilde{X})\cdot\mathbf{A}^t)$ où $\mathbb{E}\log T_i'(\widetilde{X}_i) = H(X_i) - H(\widetilde{X}_i)$, dont il faut tenir compte comme un terme de pénalité dans le cas où les composantes de $\widetilde{X}$ ne sont plus de même entropie que celles de X. On obtient alors une autre inégalité (dans le cas de lignes de $\mathbf{A}$ orthonormées) qui se trouve, si les $\widetilde{X}_i$ sont de même variance, coïncider avec [5, Théorème 3], dont les auteurs ont montré l'équivalence avec (2). La même preuve exposée ici s'étend donc pour les deux variantes de l'inégalité de puissance entropique généralisée par Zamir et Feder.

On peut également étendre le raisonnement au cas où $\widetilde{X}$ est non-gaussien, ce qui ne permet plus de simplifier $H(\widetilde{Y}|\widetilde{Y}^c) = H(\widetilde{Y})$ mais permet d'obtenir, à l'instar de [10], une inégalité réciproque faisant intervenir une information mutuelle entre deux vecteurs aléatoires orthogonaux mais dépendants. La question d'une borne inférieure optimale sur cette information mutuelle est l'objet d'une question ouverte dans [16].

De plus, comme indiqué dans [9, § IV] on peut généraliser le principe de minimum d'entropie au cas où les composantes X_i sont elle-mêmes des vecteurs aléatoires, ce qui inclut alors le cas classique de l'inégalité vectorielle de la puissance entropique. Cette généralisation s'obtient avec une preuve identique qui nécessite de généraliser le lemme 2 aux vecteurs aléatoires, comme cela est fait dans [11].

Enfin, dans le cas $m = 1$, cette approche peut être étendue [10] aux entropies de Rényi en relation avec l'inégalité de Young optimale. La généralisation aux transformations linéaires pour $m > 1$ est en cours d'étude.

Références

[1] J. F. Bercher and C. Vignat, "Estimating the entropy of a signal with applications," *IEEE Transactions on Signal Processing*, vol. 48, no. 6, pp. 1687–1694, Jun. 2000.

[2] J. P. Burg, "Maximum entropy spectral analysis," Ph.D. dissertation, Standford University, Dept. of Geophysics, Standford, CA, USA, 1975.

[3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, 2006.

[4] D. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis II*. New York : Academic Press, 1981, pp. 565–608.

[5] D. Guo, S. Shamai (Shitz), and S. Verdú, "Proof of entropy power inequalities via MMSE," in *Proc. IEEE Int. Symp. Information Theory*, Seattle, USA, Jul. 2006, pp. 1011–1015.

[6] E. T. Jaynes, "Information theory and statistical mechanics," *Physical Review*, vol. 106, no. 4, pp. 620–630, 1957.

[7] O. Rioul, "A simple proof of the entropy-power inequality via properties of mutual information," in *Proc. IEEE Int. Symp. Information Theory*, Nice, France, Jun. 2007, pp. 46–50.

[8] ——, *Théorie de l'information et du codage*. London, UK : Hermes Science - Lavoisier, 2007.

[9] ——, "Information theoretic proofs of entropy power inequalities," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 33–55, Jan. 2011.

[10] ——, "Optimal transportation to the entropy-power inequality," in *IEEE Information Theory and Applications Workshop (ITA 2017)*, San Diego, USA, Feb. 2017.

[11] ——, "Yet another proof of the entropy power inequality," *IEEE Trans. Inf. Theory*, to appear., draft available at https ://arxiv.org/abs/1606.05969.

[12] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 623–656, Oct. 1948.

[13] C. Villani, *Topics in Optimal Transportation*. Providence, RI : Amer. Math. Soc., 2003, no. 58.

[14] F. Vrins and M. Verleysen, "On the entropy minimization of a linear mixture of variables for source separation," *Signal Processing*, vol. 85, no. 5, pp. 1029–1044, May 2005.

[15] R. Zamir and M. Feder, "A generalization of information theoretic inequalities to linear transformations of independent vector," in *Proc. Sixth Joint Swedish-Russian International Workshop on Information Theory*, Mölle, Sweden, Aug. 1993, pp. 254–258.

[16] ——, "A generalization of information theoretic inequalities to linear transformations of independent vector," in *Proceedings of the 6-th Joint Swedish-Russian International Workshop on Information Theory*, Möle, Sweden, Aug. 1993, pp. 254–258.