

# Réflexions sur le principe d'information de Fisher minimum

Steeve ZOZOR

Univ. Grenoble Alpes, CNRS, Grenoble INP\*, GIPSA-Lab, 38000 Grenoble, France

\*Institute of Engineering Univ. Grenoble Alpes

steeve.zozor@cnrs.fr

**Résumé** – Dans cet article, nous nous intéressons au pendant à la Fisher du principe d'entropie maximum. Ce dernier stipule qu'un système physique fermé soumis à contraintes est nécessairement dans son état de désorganisation maximum étant données les contraintes. En d'autres termes, son entropie est maximale sous les contraintes. L'information de Fisher est elle une mesure d'information issue du monde des statistiques et non de la physique, mais est intimement reliée à la notion d'entropie de Shannon. Il est donc naturel de se poser la question de la revisite du principe d'entropie maximale sous le prisme de l'information de Fisher, objet de cet article. Nous y montrons que le problème est mal posé, et proposons un ensemble de solutions et de réflexions autour des problèmes dérivés bien posés.

**Abstract** – In this paper, we are interested in the Fisher counterpart of the maximum entropy principle. The latter stipulate that a closed physical system subject to constraints is necessarily in its state of maximum disorganization given the constraints. In other words, its entropy is maximal under the constraints. Fisher's information is a measure of information from the world of statistics and not from physics, but is intimately linked to Shannon's notion of entropy. It is therefore natural to ask the question of revisiting the principle of maximum entropy under the lens of Fisher's information, which is the subject of this article. We show that the problem is ill-posed, and propose a set of solutions and reflections around well-posed derived problems.

## 1 Introduction

La mesure de l'information attachée à un système ou vecteur aléatoire ne trouve pas de définition unique et le choix d'une telle mesure est intimement lié en général au contexte dans lequel elle s'applique. Deux mesures toutefois émergent de manière naturelle dans de très nombreux contextes applicatifs à savoir l'entropie de Shannon et l'information de Fisher.

La notion d'entropie, mesure d'incertitude, trouve son origine en mécanique statistique entre autres aux travers des travaux de Boltzman, Gibbs ou Maxwell pour citer les pionniers [1–3]. Cette même mesure, reformalisée, apparaît naturellement dans le domaine des communications suite aux travaux de Hartley, Nyquist, Wiener, Laplume, Clavier ou Shannon d'où elle finit par tirer son nom [2, 3]. L'information de Fisher, elle, est née des statistiques aux travers des travaux d'Edgeworth, Doob, Pearson, ou Fisher, dont l'histoire lui a associé son nom [2–5]. Cette mesure a clairement le sens d'une information, qui se traduit par exemple via l'inégalité de Cramér-Rao où son inverse borne la matrice de covariance [2–5]. Bien qu'issue du monde de l'estimation, cette mesure prend place depuis quelques années dans le monde de la physique, revisitée sous son prisme en complément de la vision entropique [6].

Ces deux quantités issues de mondes différents sont reliées par des identités ou inégalités [2, 3], et diverses inégalités entropiques trouvent leurs pendants en termes d'information de Fisher : l'inégalité de la puissance entropique et le théorème du traitement de l'information en sont deux exemples [2, 3].

Dans ce papier, nous nous focalisons sur le pendant à la

Fisher du principe d'entropie maximum. Ce dernier, défendu entre autres par Jaynes [1, 7], stipule qu'un système physique fermé et soumis à contraintes se désorganise au cours du temps pour atteindre son état d'incertitude maximum étant donné les contraintes. Cette même idée trouve aussi application en inférence Bayésienne, ou le choix d'un a priori étant donné la connaissance de certains moments peut être guidé par la volonté d'être le "moins informatif possible". Par ailleurs, un a priori non informatif, l'a priori de Jeffrey, fait appel à la matrice (paramétrique) de Fisher. Motivés par cette remarque tout comme par les nombreux parallèles entre entropie et information de Fisher, nous nous intéressons dans cet article à la formulation du principe d'incertitude maximum à la Fisher. Une telle formulation a été étudiée anciennement comme problème sans solution sauf cas particulier [8, 9]. Elle a trouvé écho plus récemment dans le monde physique, des statistiques ou du traitement du signal [10–12], mais à notre connaissance uniquement dans le cadre scalaire. Le passage au cadre multivarié n'est pas aussi trivial qu'il peut paraître, et est même mal posé.

## 2 Définitions et rappel

Tout au long de ce papier nous nous plaçons dans le cadre de vecteurs aléatoires continus, et de vecteurs et matrices à valeur dans l'ensemble des réels  $\mathbb{R}$ , et nous noterons :

- $\mathcal{M}_{d,d'}$ ,  $\mathcal{S}_d$ ,  $\mathcal{P}_d$ ,  $\mathcal{P}_d^+$  respectivement ensemble des matrices  $d \times d'$ , puis  $d \times d$  symétriques, symétriques définies non négatives, symétriques définies positives.

- Les vecteurs sont notés en minuscule gras (ex.  $\mathbf{x} \in \mathfrak{M}_{d,1}$ ), et les matrices en majuscule gras (ex.  $\mathbf{A}$ ).
- $\mathbf{B} \geq \mathbf{A}$  signifie que  $\mathbf{B} - \mathbf{A} \in \mathfrak{P}_d$  (ordre de Loewner),  $\mathbf{B} > \mathbf{A}$  signifie que  $\mathbf{B} - \mathbf{A} \in \mathfrak{P}_d^+$ .
- Pour toute fonction  $g : \mathfrak{M}_{d,d'} \rightarrow \mathbb{R}$ ,  $\frac{\partial g}{\partial \mathbf{M}} \in \mathfrak{M}_{d,d'}$  est la matrice de composantes  $\frac{\partial g}{\partial M_{i,j}}$ .
- $\mathbf{s} : \mathcal{X} \rightarrow \mathbb{R}^k$  statistique  $k$ -dimensionnelle sur  $\mathcal{X} \subset \mathbb{R}^d$ .
- $\mathcal{P}_s$  ensemble des densités de probabilité sur  $\mathcal{X} \subset \mathbb{R}^d$ , supposé convexe, de même statistique  $\mathbf{s}$  et  $\tilde{\mathcal{P}}_s$  sa restriction aux densités différentiables sur l'adhérence de  $\mathcal{X}$ .

Nous utiliserons par ailleurs les définitions suivantes de fonctions matricielles [13, 14] :

- $g : \mathfrak{S}_d \rightarrow \mathfrak{S}_{d'}$  est *matrice-croissante* si  $\forall \mathbf{B} > \mathbf{A} \Rightarrow g(\mathbf{B}) \geq g(\mathbf{A})$ ; *strictement* si l'inégalité est stricte.
- $g : \mathfrak{M}_{d,d'} \rightarrow \mathfrak{S}_{d'}$  est *matrice-convexe* si  $\forall \mathbf{A}, \mathbf{B}, \alpha \in [0; 1]$ ,  $g(\alpha \mathbf{A} + (1 - \alpha)\mathbf{B}) \leq \alpha g(\mathbf{A}) + (1 - \alpha)g(\mathbf{B})$ ; *strictement* si l'inégalité est stricte pour  $0 < \alpha < 1$ .

Sont strictement matrice-croissantes la trace  $\text{Tr}$ , le déterminant  $|\cdot|$ , les valeurs propres  $\lambda_i$  classées par ordre croissant de matrices de  $\mathfrak{P}_d$ ; sont matrice-convexes les fonctions  $\text{Tr}$ ,  $-|\cdot|^{\frac{1}{d}}$ ,  $-\log|\cdot|$ ,  $-\lambda_1, \lambda_d$  [13, 14].

Par la suite, nous nous intéressons au pendant à la Fisher du principe d'entropie maximale sous contrainte de statistique  $\mathbf{s}$ , défendu par Jaynes et qui s'énonce sous la forme [1–5, 7]  $p_{\text{em}} = \text{argmax}_{p \in \mathcal{P}_s} H(p)$  avec  $H(p) = -\int_{\mathcal{X}} p(\mathbf{x}) \log p(\mathbf{x}) d\mathbf{x}$  entropie différentielle de Shannon de  $p$ , et qui a pour solution  $p_{\text{em}}(\mathbf{x}) = e^{\boldsymbol{\eta}^t \mathbf{s}(\mathbf{x}) - \varphi(\boldsymbol{\eta})}$  loi de la famille exponentielle [2, 4, 5], avec  $\boldsymbol{\eta} \in \mathbb{R}^k$  et  $\varphi$  tels que contraintes de normalisation et de statistique soient satisfaites. La preuve fait appel à la technique d'optimisation variationnelle sous contraintes [2, 15]. Une preuve alternative, informationnelle, part de la solution et s'appuie sur la positivité de la divergence de Kullback-Leibler de  $p$  vis-à-vis de  $p_{\text{em}}$  [2].

Au-delà de l'idée physique qui sous-tend ce principe, il trouve aussi application en traitement du signal, par exemple pour choisir un a priori en inférence Bayésienne [5], ou encore en estimation spectrale ou de processus autorégressifs [2, 16].

### 3 Fisher minimum sous contraintes

Nous nous intéressons ici à la matrice de Fisher non paramétrique, caractérisant la loi vis-à-vis de l'état  $\mathbf{x}$ , et non la version paramétrique qui la caractérise vis-à-vis d'un paramètre où le problème de minimisation se réduit à la recherche du paramètre d'information minimum.

Contrairement à l'entropie, la matrice information de Fisher est fonctionnelle de la loi  $p$  à valeur matricielle et non scalaire :

**Définition 1** (Matrice information de Fisher non paramétrique [2, 4, 6]). Soit  $p$  une densité de probabilité sur  $\mathcal{X} \subset \mathbb{R}^d$ , différentiable sur l'adhérence de  $\mathcal{X}$ . Sa matrice de Fisher non paramétrique, de dimension  $d \times d$ , est définie par

$$\mathbf{J}(p) = \int_{\mathcal{X}} \frac{\partial \log p(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial \log p(\mathbf{x})}{\partial \mathbf{x}^t} p(\mathbf{x}) d\mathbf{x},$$

matrice de covariance de la fonction score non paramétrique  $\boldsymbol{s} = \frac{\partial \log p}{\partial \mathbf{x}}$ . Par construction,  $\mathbf{J}(p) \in \mathfrak{P}_d$ .

$\mathbf{J}$  est une mesure d'information sur l'état  $\mathbf{x}$ , ce qui peut se voir au travers de l'inégalité de Cramér-Rao non paramétrique qui borne inférieurement la matrice de covariance d'un vecteur aléatoire de loi  $p$  par  $\mathbf{J}(p)^{-1}$  [2–4, 6].

#### 3.1 Un problème mal posé

La matrice de Fisher et l'ordre matriciel de Loewner interviennent dans l'inégalité de Cramér-Rao [2, 4], ou encore dans les versions à la Fisher du second principe de la thermodynamique [6, 17]. Par soucis de cohérence, le problème que nous nous posons est celui de la recherche de la loi de matrice de Fisher minimum suivant :

$$p_{\text{fm}} = \underset{p \in \tilde{\mathcal{P}}_s}{\text{argmin}} \mathbf{J}(p) \iff \forall p \in \tilde{\mathcal{P}}_s, \mathbf{J}(p) \geq \mathbf{J}(p_{\text{fm}}) \quad (1)$$

On peut montrer que  $\mathbf{J}(p)$  est matrice-convexe, en suivant pas à pas la preuve donnée dans [18] traitant du cas scalaire  $d = 1$ , et  $\mathcal{P}_s$  est un ensemble convexe. Toutefois, ces propriétés ne garantissent pas l'existence et l'unicité d'une solution dans la mesure où l'ordre de Loewner n'est pas un ordre total [13, 14].

#### 3.2 Ensemble de versions scalaires, bien posées

La relation d'ordre de Loewner se réduisant à l'ordre naturel, total, dans le cas scalaire, nous proposons comme alternative de considérer des fonctions scalaires  $g : \mathfrak{P}_d \rightarrow \mathbb{R} \equiv \mathfrak{S}_1$  matrice-croissantes de la matrice de Fisher, ainsi

$$\mathbf{J}(p_1) \geq \mathbf{J}(p_2) \implies g(\mathbf{J}(p_1)) \geq g(\mathbf{J}(p_2)), \quad (2)$$

et de s'attaquer à un problème de type

$$p_{\text{fm},g} = \underset{p \in \tilde{\mathcal{P}}_s}{\text{argmin}} g(\mathbf{J}(p)) \quad (3)$$

Ce problème est bien posé et si de plus  $g$  est matrice-convexe il a une solution unique. Par ailleurs, si  $g$  est strictement croissante, si le problème (1) a une solution, alors la minimisante du problème (3) est aussi solution du problème (1) (dans le cas contraire, l'implication (2) serait violée). Par ailleurs, si pour deux fonctions  $g$  différentes nous obtenons des minimisantes respectives différentes, le problème initial n'admet pas de solution. En effet, de par l'implication (2), l'ensemble des minimisantes de  $\mathbf{J}$  est ensemble minimisant n'importe quel  $g(\mathbf{J})$ .

On supposera par la suite que  $g$  est différentiable et on appellera  $g(\mathbf{J}(p))$  *g-information de Fisher* par raccourci sémantique.

#### 3.3 Résolution via la fonction d'onde

Une réécriture de la densité de probabilité usuelle en physique des ondes ou mécanique quantique est de la forme  $p = \psi^2$ , la fonction  $\psi$  étant la fonction d'onde décrivant un système

(on oubliera ici la phase), paramétrisation utilisée par ailleurs dans la littérature de physique traitant du problème d'information de Fisher minimum dans le cadre scalaire [10]. Un rapide calcul permet de montrer que la matrice de Fisher s'écrit

$$\mathbf{J}(p) \equiv \mathbf{J}(\psi) = 4 \int_{\mathcal{X}} \frac{\partial \psi(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial \psi(\mathbf{x})}{\partial \mathbf{x}^t} d\mathbf{x} \quad (4)$$

$\mathbf{J}$  est également matrice-convexe vis-à-vis de  $\psi$  en raison de la linéarité de la dérivée et de la convexité de  $\mathbf{a} \mapsto \mathbf{a}\mathbf{a}^t$  [13]. Toutefois, l'ensemble  $\tilde{\mathcal{P}}_s$  n'est pas convexe vis-à-vis de  $\psi$ .

En introduisant les facteurs de Lagrange  $\eta_0$  et  $\boldsymbol{\eta} \in \mathbb{R}^k$  pour les contraintes de normalisation et de statistique [15], on cherche à résoudre le problème de minimisation

$$\psi_{\text{fm},g} = \underset{\psi}{\operatorname{argmin}} \left[ g \left( 4 \int_{\mathcal{X}} \frac{\partial \psi(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial \psi(\mathbf{x})}{\partial \mathbf{x}^t} d\mathbf{x} \right) + \int_{\mathcal{X}} (\eta_0 + \boldsymbol{\eta}^t \mathbf{s}(\mathbf{x})) \psi^2(\mathbf{x}) d\mathbf{x} \right] \quad (5)$$

Le problème variationnel consiste à minimiser une expression du type  $C(f) = g \left( \int_{\mathcal{X}} \mathcal{L} \left( \mathbf{x}, f(\mathbf{x}), \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \right) d\mathbf{x} \right) + \int_{\mathcal{X}} \mathcal{K} \left( \mathbf{x}, f(\mathbf{x}), \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} \right) d\mathbf{x}$  avec  $\mathcal{L} : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{M}_{d,d}$  et  $\mathcal{K} : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^d \rightarrow \mathbb{R}$  différentiables. En suivant pas à pas l'approche conduisant à l'équation d'Euler-Lagrange [15] qui consiste à écrire que si  $f$  est minimisante, quelque soit  $\delta f$   $C(f + \varepsilon \delta f)$  est minimum en  $\varepsilon = 0$ , et donc la dérivée de  $C(f + \varepsilon \delta f)$  s'annule en  $\varepsilon = 0$ , on aboutit après des calculs élémentaires mais fastidieux, que  $f$  doit être solution de

$$\operatorname{Tr} \left( \frac{\partial g}{\partial \mathbf{M}} \left( \int_{\mathcal{X}} \mathcal{L} \right) \cdot \left( \frac{\partial \mathcal{L}}{\partial f} - \sum_i \frac{d}{dx_i} \frac{\partial \mathcal{L}}{\partial f'_i} \right) \right) + \frac{\partial \mathcal{K}}{\partial f} - \sum_i \frac{d}{dx_i} \frac{\partial \mathcal{K}}{\partial f'_i} = 0 \quad (6)$$

où, par soucis de simplicité, la variable  $\mathbf{x}$  est omise et  $f'_i = \frac{\partial f}{\partial x_i}$ .

Lorsque l'on applique cette équation de type Euler-Lagrange au problème que nous cherchons à résoudre, il se trouve que  $\frac{1}{4} \sum_i \frac{d}{dx_i} \frac{\partial \mathcal{L}}{\partial \psi'_i} = \frac{\partial^2 \psi}{\partial \mathbf{x} \partial \mathbf{x}^t}$  matrice hessienne de  $\psi$ ,

$\frac{\partial \mathcal{K}}{\partial \psi} = (\eta_0 + \boldsymbol{\eta}^t \mathbf{s}) \psi$ , les autres termes étant nuls. Après reparamétrisation des facteurs de Lagrange, on aboutit au résultat :

**Proposition 1** (*g-information de Fisher minimale sous contraintes*). Soit  $g(\mathbf{M})$  matrice-croissante et matrice-convexe. S'il existe une fonction  $\psi$  vérifiant l'équation intégrodifférentielle non linéaire

$$\operatorname{Tr} \left( \frac{\partial g}{\partial \mathbf{M}} \left( \int_{\mathcal{X}} \frac{\partial \psi(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial \psi(\mathbf{x})}{\partial \mathbf{x}^t} d\mathbf{x} \right) \cdot \frac{\partial^2 \psi}{\partial \mathbf{x} \partial \mathbf{x}^t} \right) + \boldsymbol{\eta}^t \mathbf{s} \psi = \eta_0 \psi,$$

alors, pourvu que l'on puisse déterminer  $\eta_0$  et  $\boldsymbol{\eta}$  pour satisfaire aux contraintes, la ou les solutions  $\psi_{\text{fm},g}$  sont telles que

$$p_{\text{fm},g} = \psi_{\text{fm},g}^2 = \underset{p \in \tilde{\mathcal{P}}_s}{\operatorname{argmin}} g(\mathbf{J}(p))$$

Particularisons à présent ce résultat très formel.

### 3.4 Deux exemples

**Trace de la matrice de Fisher.** La trace de la matrice de Fisher offre un résumé moyen de la matrice de Fisher au travers de la somme de ses valeurs propres [2, 4].  $g = \operatorname{Tr}$  est strictement croissante et convexe [13, 14] et on montre que  $\frac{\partial g}{\partial \mathbf{M}} = \mathbf{I}$  matrice identité. Par conséquent, la fonction d'onde aboutissant à la trace de la matrice de Fisher minimale satisfait à

$$\Delta \psi + \boldsymbol{\eta}^t \mathbf{s} \psi = \eta_0 \psi \quad (7)$$

avec  $\Delta = \operatorname{Tr} \frac{\partial^2}{\partial \mathbf{x} \partial \mathbf{x}^t}$  opérateur Laplacien. Cette équation n'est autre que l'équation de Schrödinger indépendante du temps où  $\eta_0$  représente le niveau d'énergie et  $\boldsymbol{\eta}^t \mathbf{s}$  le potentiel auquel est soumis le système. On retrouve ici un résultat déjà connu dans la littérature, dans le cas scalaire [10]. Par ailleurs, la littérature de physique foisonne de cas où cette équation trouve solution (particule libre, potentiel coulombien/hydrogène, oscillateur harmonique, ...).

**Valeur propre maximale de la matrice de Fisher** La minimisation de la valeur propre maximale de la matrice de Fisher signifie minimiser l'information maximale contenue dans la matrice de Fisher. Il est prouvé que  $g = \lambda_d$ , valeur propre maximale, est également strictement croissante et convexe [13]. Toutefois, cette fonction n'a pas de forme explicite, tout du moins pour  $d \geq 5$ , et même pour  $d = 2, 3, 4$  sa détermination n'est pas une sinécure. Cependant, une valeur propre est fonction implicite issue de l'équation  $|\lambda \mathbf{I} - \mathbf{M}| = 0$ . Suivant le théorème des fonctions implicites [19], de  $\frac{\partial |\mathbf{M}|}{\partial \mathbf{M}} = -|\mathbf{M}| \mathbf{M}^{-t}$  [14], et de la décomposition spectrale de  $\mathbf{M} \in \mathfrak{F}_d$  avec  $\mathbf{v}_i$  vecteur propre associé à la valeur propre  $\lambda_i$  on obtient  $\frac{\partial \lambda_d}{\partial \mathbf{M}} = \lim_{\lambda \rightarrow \lambda_d} \frac{(\lambda \mathbf{I} - \mathbf{M})^{-t}}{\operatorname{Tr}(\lambda \mathbf{I} - \mathbf{M})^{-t}} = \mathbf{v}_d \mathbf{v}_d^t$ . Ainsi, la fonction d'onde aboutissant à la matrice de Fisher de valeur propre maximale minimale satisfait à

$$\mathbf{v}_d(\psi)^t \frac{\partial^2 \psi}{\partial \mathbf{x} \partial \mathbf{x}^t} \mathbf{v}_d(\psi) + \boldsymbol{\eta}^t \mathbf{s} \psi = \eta_0 \psi \quad (8)$$

avec  $\mathbf{v}_d(\psi)$  vecteur propre de  $\int_{\mathcal{X}} \frac{\partial \psi(\mathbf{x})}{\partial \mathbf{x}} \frac{\partial \psi(\mathbf{x})}{\partial \mathbf{x}^t} d\mathbf{x}$  associé à sa valeur propre maximale  $\lambda_d$ .

Cette équation ressemble de manière troublante à l'équation de Schrödinger, mais est bien plus complexe du fait que nous n'avons pas de forme analytique de  $\mathbf{v}_d$ .

Il se trouve que pour tout  $1 \leq k \leq d$ ,  $\sum_{i=k}^d \lambda_i$  est matrice croissance et matrice convexe [13], de sorte que ce résultat se généralise en remplaçant  $\mathbf{v}_d$  par  $\mathbf{V}_{k:d}$  matrice de vecteurs colonnes  $\mathbf{v}_i$  vecteurs propres associés aux  $\lambda_i$  et en prenant la trace. Le cas  $g = \operatorname{Tr}$  se retrouve pour  $k = 1$ .

### 3.5 Retour au problème matriciel

La proposition 1 concernant toute fonction  $g$  croissante convexe différentiable, on peut penser une résolution possible du problème initial mal posé en supprimant  $g$  et l'opérateur trace dans ce résultat, tout en rendant les facteurs de Lagrange

matriciels. On peut également partir de la technique d'Euler-Lagrange en travaillant comme si la matrice de Fisher était scalaire pour "voir" à quoi on aboutit. Les deux approches sont absolument non rigoureuses, mais aboutissent à un même résultat qu'il est possible de vérifier a posteriori :

**Proposition 2** (Matrice de Fisher minimale sous contraintes). S'il existe une fonction  $\psi$  vérifiant l'équation intégrodifférentielle non linéaire matricielle

$$\frac{\partial^2 \psi}{\partial \mathbf{x} \partial \mathbf{x}^t} + \left( \sum_{i=1}^k \mathbf{L}_i s_i \right) \psi = \mathbf{L}_0 \psi$$

avec  $s_i$  composantes de  $\mathbf{s}$ ,  $\mathbf{L}_i \in \mathfrak{M}_{d,d}$ ,  $i = 0, \dots, k$ , alors, si on peut déterminer des  $\mathbf{L}_i$  pour satisfaire aux contraintes, la ou les solutions  $\psi_{\text{fm}}$  sont telles que

$$\forall p \in \tilde{\mathcal{P}}_{\mathbf{s}}, \mathbf{J}(p) \geq \mathbf{J}(\psi_{\text{fm}})$$

*Démonstration.* Faute de place la démonstration complète ne sera pas donnée dans le détail. À l'instar du problème d'entropie maximale, une preuve de type informationnelle se base sur la fonctionnelle  $\mathbf{C} = 4(\delta - \delta_{\text{fm}})(\delta - \delta_{\text{fm}})^t \psi^2$  avec  $\psi$  telle que  $\psi^2 \in \tilde{\mathcal{P}}_{\mathbf{s}}$  et les fonctions score  $\delta = 2 \frac{\partial \psi}{\partial \mathbf{x}} / \psi$  et  $\delta_{\text{fm}} = 2 \frac{\partial \psi_{\text{fm}}}{\partial \mathbf{x}} / \psi_{\text{fm}}$ . Le travail consiste à écrire  $\mathbf{C}$  en extrayant  $4 \left( \frac{\partial \psi}{\partial \mathbf{x}} \frac{\partial \psi}{\partial \mathbf{x}^t} - \frac{\partial \psi_{\text{fm}}}{\partial \mathbf{x}} \frac{\partial \psi_{\text{fm}}}{\partial \mathbf{x}^t} \right)$ , puis à faire apparaître dans le terme résiduel la hessienne de  $\psi_{\text{fm}}$ , que l'on remplacera par son expression issue de l'équation différentielle. Des calculs fastidieux et hypothèses techniques (permettant d'évoquer entre autre le théorème de la divergence) aboutissent au fait que l'intégrale des termes résiduels s'annule, de sorte que l'intégrale de  $\mathbf{C}$ , matrice divergence de Fisher avec  $\psi_{\text{fm}}^2$  la distribution de référence [20] s'écrit

$$\int_{\mathcal{X}} \mathbf{C}(\mathbf{x}) d\mathbf{x} = \mathbf{J}(\psi) - \mathbf{J}(\psi_{\text{fm}})$$

La preuve se clôt en remarquant que  $\mathbf{C} \geq 0$ . □

## 4 Discussions

Cet article visait à revisiter le principe d'incertitude maximum, ou information minimum, sous le prisme de la matrice information de Fisher. Cette mesure étant matricielle et non scalaire semble plus riche que l'entropie de Shannon. Par ailleurs, si elle trouve son origine dans le monde de l'estimation ou des statistiques, la physique s'en est emparée ces dernières décennies entre autres sous l'impulsion de Frieden [6], renforçant l'intérêt de son étude dans ce cadre.

D'autre part, de nombreuses relations lient entropie et Fisher, vues comme mesures complémentaires. Dériver un principe de Fisher minimum permet ainsi de compléter le panorama des équivalents à la Fisher des relations entropiques [2, 3, 6]. Les lois d'information de Fisher minimum sont également susceptibles d'offrir des alternatives au principe d'entropie maximum, par exemple en physique ou en inférence Bayésienne dans le choix d'a priori, en remplaçant un choix de loi d'entropie maximum par celle d'information de Fisher minimum.

On notera enfin que si cet article expose quelques réflexions sur la recherche de loi d'information de Fisher minimum sous contraintes, à notre connaissance le problème reste encore ouvert, en particulier traitant de la matrice de Fisher.

## Références

- [1] N. Merhav. *Statistical Physics for Electrical Engineering*. Springer, 2018.
- [2] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2nd edition, 2006.
- [3] O. Rioul. *Théorie de l'information et du codage*. Lavoisier, 2007.
- [4] S. M. Kay. *Fundamentals for Statistical Signal Processing : Estimation Theory*. vol. 1. Prentice Hall, 1993.
- [5] C. P. Robert. *The Bayesian Choice. From Decision-Theoretic Foundations to Computational Implementation*. Springer, 2nd edition, 2007.
- [6] B. R. Frieden. *Science from Fisher Information : A Unification*. Cambridge University Press, 2004.
- [7] E. T. Jaynes. Gibbs vs Boltzmann entropies. *American Journal of Physics*, 33(5) :391–398, May 1965.
- [8] P. J. Huber. Robust estimation of a location parameter. *Annals of Math Stat.*, 35(1) :73–101, March 1964.
- [9] A. M. Kagan, Y. V. Linnik, and C. R. Rao. *Characterization problems in mathematical statistics*. John Wiley & Sons, 1973.
- [10] B.R. Frieden, A. Plastino, A.R. Plastino, and B.H. Soffer. Non-equilibrium thermodynamics and Fisher information : An illustrative example. *Phys. Let. A*, 304(3-4) :73–78, Nov. 2002.
- [11] J. M. Borwein, A. S. Lewis, and D. Noll. Maximum entropy reconstruction using derivative information, part 1 : Fisher information and convex duality. *Math. Op. Res.*, 21(2) :442–468, May 1996.
- [12] V. Živojnović. Minimum Fisher information of moment-constrained distributions with application to robust blind identification. *Sig. Proc.*, 65(2) :297–13, Mar. 1998.
- [13] A. W. Marshall, I. Olkin, and B. C. Arnold. *Inequalities : Theory of Majorization and Its Applications*. Springer Verlag, 2nd edition, 2011.
- [14] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, 3rd edition, 1999.
- [15] B. van Brunt. *The Calculus of Variations*. Springer, 2004.
- [16] J. P. Burg. *Maximum entropy spectral analysis*. PhD thesis, Department of Geophysics, Stanford University, Stanford University, Stanford, CA, May 1975.
- [17] S. Zozor. Une version quantitative du second principe de la thermodynamique au travers de la matrice information de Fisher. In *Colloque GRETSI*, Nancy, France, 5-9 September 2022.
- [18] M. Cohen. The Fisher information and convexity. *IEEE Trans. on Information Theory*, 14(4) :591–592, July 1968.
- [19] J. J. Duistermaat and J. A. C. Kolk. *Multidimensional Real Analysis I : Differentiation*. Cambridge University Press, 2004.
- [20] O. Johnson. *Information Theory and The Central Limit Theorem*. Imperial college Press, 2004.