HAL open science

-Logarithmes déformés. -Divergences entropiques associées. -Applications aux problèmes inverses linéaires. - Algorithmes d'inversion.

Henri Lantéri

To cite this version:

Henri Lantéri. -Logarithmes déformés. -Divergences entropiques associées. -Applications aux problèmes inverses linéaires. - Algorithmes d'inversion.. 2021. hal-03355082v1

HAL Id: hal-03355082

https://hal.science/hal-03355082v1

Preprint submitted on 27 Sep 2021 (v1), last revised 29 Sep 2022 (v3)

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.

-Logarithmes déformés.
-Divergences entropiques associées.
-Applications aux problèmes inverses linéaires.
- Algorithmes d'inversion.

Henri Lantéri

27 septembre 2021

Table des matières

1 Introduction et sommaire **3**

2 Rappels (très brefs) sur les problèmes inverses. **4**

3 Rappels (très brefs) sur l'entropie. **6**

4 Entropie de Shannon - Logarithme naturel. **7**
4.1 Divergence de Csiszär associée - Kullback-Leibler. 7
4.2 Divergence de Csiszär duale. 8
4.3 Divergences de Bregman associées. 8

5 Entropie de Tsallis [19]. **8**
5.1 Divergence de Csiszär fondée sur l'entropie de Tsallis. 9
5.2 Divergence de Csiszär duale fondée sur l'entropie de Tsallis. . . . 9
5.3 Divergence de Bregman fondée sur l'entropie de Tsallis. 10
5.4 Divergence de Bregman duale fondée sur l'entropie de Tsallis. . . 10

6 Entropie de Kaniadakis [8], [9], "K" entropie. **11**
6.1 Divergence de Csiszär construite sur la "K" entropie. 11
6.2 Divergence de Csiszär duale construite sur la "K" entropie. . . . 12
6.3 Divergence de Bregman construite sur la "K" entropie. 12
6.4 Divergence de Bregman duale construite sur la "K" entropie. . . 12

7 Entropie de Abe [1]. **13**
7.1 Divergence de Csiszär fondée sur l'entropie de Abe. 13
7.2 Divergence de Csiszär duale fondée sur l'entropie de Abe. 14
7.3 Divergence de Bregman fondée sur l'entropie de Abe. 14
7.4 Divergence de Bregman duale fondée sur l'entropie de Abe. . . . 15

8 Gamma (γ) entropie. 15
8.1 Divergence de Csiszär construite sur la "γ" entropie. 16
8.2 Divergence de Csiszär duale construite sur la "γ" entropie. 16
8.3 Divergence de Bregman construite sur la "γ" entropie. 16
8.4 Divergence de Bregman duale construite sur la "γ" entropie. . . . 17

9 Entropie à 2 paramètres [10], K.L.S. 17
9.1 Divergence de Csiszär construite sur l'entropie à 2 paramètres. . 18
9.2 Divergence de Csiszär duale construite sur l'entropie à 2 paramètres. 18
9.3 Divergence de Bregman construite sur l'entropie à 2 paramètres. 19
9.4 Divergence de Bregman duale construite sur l'entropie à 2 paramètres. 19

10 Autres définitions de l'entropie et du Logarithme déformé. 20

11 Divergences fondées sur l'entropie générale (94). 21
11.1 Divergence de Csiszär construite sur l'entropie générale. 22
11.2 Divergence de Csiszär duale construite sur l'entropie générale. . . 23
11.3 Divergence de Bregman construite sur l'entropie générale. 23
11.4 Divergence de Bregman duale construite sur l'entropie générale. . 24

12 Quelques autres expressions du Logarithme déformé. 24
12.1 Logarithme déformé de Newton [16]. 24
12.1.1 Divergence de Csiszär correspondant à l'entropie de Newton. 25
12.1.2 Divergence de Csiszär duale correspondant à l'entropie de Newton. 25
12.1.3 Divergence de Bregman correspondant à l'entropie de Newton. 25
12.1.4 Divergence de Bregman duale correspondant à l'entropie de Newton. 25
12.2 Autre expression du Logarithme déformé [9]. 26
12.2.1 Divergence de Csiszär correspondante. 26
12.2.2 Divergence de Csiszär duale correspondante. 27

13 Divergences invariantes. 27
13.1 Quelques rappels sur les facteurs d'invariance et leurs propriétés. 27
13.2 Divergences de Csiszär invariantes 28
13.2.1 Divergences de Csiszär invariantes - Cas général. 29
13.2.2 Cas particulier : divergence de Csiszär/Tsallis invariante. 30
13.3 Divergences de Csiszär duales invariantes. 30
13.3.1 Divergences de Csiszär duales invariantes - Cas général. . 31
13.3.2 Cas particulier : divergence de Csiszär/Tsallis duale invariante. 32
13.4 Divergences de Bregman invariantes. 32
13.4.1 Divergences de Bregman invariantes - Cas général. 33
13.4.2 Cas particulier - Divergence de Bregman/Tsallis invariante. 34

13.5 Divergences de Bregman duales - Forme invariante. 34
13.5.1 Divergences de Bregman duales invariantes - Cas général. 35
13.5.2 Cas particulier de la divergence de Bregman/Tsallis duale invariante. 36

14 Algorithmique **36**
14.1 Généralités sur la méthode de construction des algorithmes. . . . 36
14.2 Algorithmes fondés sur l'entropie générale. 38
14.2.1 Divergence de Csiszär construite sur l'entropie générale. . 39
14.2.2 Divergence de Bregman construite sur l'entropie générale. 39
14.2.3 Divergence de Csiszär invariante construite sur l'entropie générale. 40
14.2.4 Divergence de Bregman invariante construite sur l'entropie générale. 41

15 Annexe **42**
15.1 Divergences de Csiszär. 42
15.2 Divergences de Csiszär duales. 42
15.3 Divergences de Bregman. 43
15.4 Divergences de Bregman duales. 44

1 Introduction et sommaire

Ce travail s'inscrit dans le cadre des problèmes inverses avec modèle linéaire. La résolution de ce type de problème consiste à minimiser (éventuellement sous contraintes), une fonction d'écart entre les mesures et un modèle physique du phénomène considéré.
Dans le cadre des problèmes de déconvolution d'images, par exemple, les contraintes considérées sont typiquement la contrainte de non-négativité et la contrainte de somme de l'intensité totale de l'image reconstruite.
Cette démarche implique d'une part la définition d'une fonction d'écart et d'autre part la mise en oeuvre d'une méthode algorithmique de minimisation sous contraintes d'une telle fonction.
Ces deux aspects ont été développés dans un précédent travail [11], [12].
Dans le chapitre 3 du travail précité les divergences fondées sur les formes les plus classiques de l'entropie ont été analysées.
Nous étendrons ici cette étude à des fonctions d'écart fondées sur des formes d'entropie développées dans le domaine de la Physique Statistique.
Ces diverses formes d'entropie s'appuient sur "*l'algèbre déformée*" [9] et en particulier sur la notion de "*Logarithme déformé*" [15] [9].
Il est bien entendu que l'analyse des propriétés de ces entropies n'entre pas dans le cadre de notre travail et que nous nous contenterons d'utiliser les résultats trouvés dans la littérature.
En conséquence le présent document est organisé comme suit.
Nous donnons dans les sections 2 et 3, quelques brefs rappels sur les problèmes

inverses et sur l'entropie.
Dans les sections 4 à 9, nous passons en revue les principales formes d'entropie intervenant dans le domaine de la Physique statistique et nous analysons les divergences construites sur de telles fonctions.
Dans la section 10, nous considérons d'autres définitions de l'entropie fondées sur les notions de dérivée généralisées. Ceci permet de retrouver les différentes formes d'entropie proposées dans les sections précedentes, mais permet surtout de mettre en évidence une forme très générale de l'entropie qui synthétise toutes les précédentes.
Dans section 11, nous analysons les divergences construites sur la base de cette forme générale de l'entropie.
La section 12 est consacrée à des formes d'entropie qui sortent du cadre général précédemment décrit dans la section 11 et les divergences associées à ces entropies particulières sont analysées.
Dans la section 13, nous étudions les formes invariantes par changement d'échelle des divergences fondées sur l'entropie générale présentées dans la section 11 ; les différentes formes de divergences sont considérées et le cas particulier de l'entropie de Tsallis est analysé en détails.
La section 14 est consacrée à l'aspect algorithmique.
Aprés avoir reppellé brièvement la méthode SGM plus largement développée dans [11], [12], nous explicitons les algorithmes correspondants permettant la minimisation des divergences générales dans le cas d'une contrainte de non négativité, puis, dans le cas ou on impose en plus, une contrainte sur la somme des inconnues nous appliquons la même méthode aux divergences générales invariantes.
Les formes multiplicatives et non multiplicatives de ces algorithmes sont indiquées dans chaque cas.
Ceci permet d'obtenir les algorithmes correspondant aux divergences développées dans les sections 4 à 9 comme des applications spécifiques du cas général.
Les algorithmes correspondant aux divergences invariantes fondées sur l'entropie de Tsallis sont explicitées à titre de comparaison.
Nous résumons en annexe l'opposé des gradients correspondant aux différentes formes d'entropie, afin de mettre clairement en évidence les analogies entre celles ci.

2 Rappels (très brefs) sur les problèmes inverses.

Dans ces problèmes, on dispose d'une mesure d'une grandeur physique notée "y" qui sera notée "p" et d'un modèle décrivant cette grandeur ; dans notre travail on considèrera que ce modèle est linéaire "Hx", il sera noté "q".
Ici, "H" est une matrice connue et "x" est la véritable inconnue du problème.
Il est bien clair que dans ce contexte, "p" et "q" ne sont pas des densités de probabilités contrairement aux grandeurs analogues qui interviennent dans le domaine de la Physique Statistique.
Ce point particulier constitue l'un des problèmes qui se posent lors de la construc-

tion des divergences.
Les paramètres inconnus "x" sont soumis à des contraintes de 2 types :
1 - les composantes de "x" doivent être positives ou nulles,
2 - dans certains cas une contrainte supplémentaire sera du type $\sum_i x_i = C$, "C" étant une valeur connue.
La résolution du problème implique de minimiser sous contraintes, par rapport à "x", une fonction d'écart ou "*divergence*" entre mesure et modèle ; on notera de façon générale $D(p\|q)$ une telle divergence.
Cette fonction n'est pas nécessairement symétrique $D(p\|q) \neq D(q\|p)$, et en toute généralité, le point important est la stricte convèxité par rapport à "q" ce qui entraine stricte la convexité par rapport à "x".
La divergence $D(q\|p)$ est désignée par le terme divergence duale de $D(p\|q)$.
La construction des différents type de divergence s'appuie sur la connaissance d'une fonction strictement convexe $f_c(x)$ (*fonction convexe standard*) possédant les propriétés ; $f_c(1) = 0$ et $f'_c(1) = 0$.
A partir d'une fonction convexe $f(x)$, on obtient la "*fonction convexe standard*" $f_c(x)$, par la relation :

$$f_c(x) = f(x) - f(1) - (x-1)f'(1) \tag{1}$$

Les divergences considérées dans ce travail sont les divergences de Csiszär et les divergences de Bregman ; les divergences de Jensen proposées dans [11], [12], ne seront pas traitées ici.
On rappelle que la divergence de Csiszär construite sur la fonction convexe $f(x)$ s'écrit :

$$C(p\|q) = \sum_i q_i f\left(\frac{p_i}{q_i}\right) \tag{2}$$

De même, la divergence de Bregman s'écrit :

$$B(p\|q) = \sum_i \left[f(p_i) - f(q_i) - (p_i - q_i)f'(q_i)\right] \tag{3}$$

On remarquera que ces divergences sont séparables (Trace form divergences).
Pour notre problème, les divergences de Csiszär doivent être construites sur une fonction convexe standard, elles sont conjointement convexes par rapport aux deux arguments (p et q).
En revanche, les divergences de Bregman construites sur une fonction convexe simple ou sur une fonction convexe standard conduisent à la même expression, elles sont convexes par rapport au premier argument, mais peuvent ne pas être convexes par rapport au deuxième argument, ce qui justifie une analyse des divergences duales.
En ce qui concerne l'aspect algorithmique, nous nous appuierons sur la méthode SGM exposée dans [11], [12] ; cette méthode permet de prendre en compte facilement la contrainte de non négativité, en revanche, la contrainte de somme pourra être prise en compte très simplement avec la même méthode, à condition d'utiliser des divergences possédant la propriété d'invariance par changement

d'échelle sur la variable "q", c'est à dire sur la variable "x", compte tenu du modèle linéaire.
Dans ce but, nous développerons les formes "*invariantes*" des divergences proposées.

3 Rappels (très brefs) sur l'entropie.

Dans cette note, on s'intéresse aux différentes expressions de l'entropie et du "*Logarithme déformé*" qui apparaissent dans la littérature dédiée à la Physique Statistique [20] ; les propriétés spécifiques des entropies sont pas considérées dans ce travail.
Dans ce contexte, par analogie avec l'entropie de Shannon, on considère que l'expression générale de l'entropie est de la forme séparable (Trace form entropy) :

$$S = -\sum_i p_i \Lambda\left(p_i\right) = -\sum_i f(p_i) \tag{4}$$

Par définition, les "p_i" sont des densités de probabilités, elles sont donc comprises entre 0 et 1 et on a $\sum_i p_i = 1$.
La fonction $\Lambda\left(p_i\right)$ désigne le Logarithme déformé ou le Logarithme traditionnel pour l'entropie de Shannon.
Par ailleurs, l'entropie étant une fonction concave, les fonctions $f(p_i)$ sont convexes par rapport à "p_i".
De plus, les expressions de l'entropie apparaissant dans le domaine de la Physique Statistique dépendent d'un ou plusieurs paramètres dont les domaines de valeurs sont fixés afin que les propriétés classiques de l'entropie soient respectées (spécifiquement dans notre cas, la concavité).
Partant de ces expressions de l'entropie, en utilisant la définition (4), on en déduit les expressions des Logarithmes déformés Log_d correspondants.
Dans le domaine de la Physique Statistique, apparaissent les fonctions d'entropie relative entre 2 densités de probabilités, alors que pour notre usage, nous devons définir des fonctions d'écart analogues à des entropies relatives, mais qui portent sur des champs de données qui ne sont pas des densités de probabilités, ceci implique quelques modifications dans le mode de construction de ces fonctions.
Pour notre usage particulier, nous utilisons les expressions des différentes fonctions $f(x)$ définies dans (4) pour $x \geq 0$ et non pas pour $0 \leq x \leq 1$.
La première question est : les fonctions correspondantes (spécifiques des entropies) restent-elles convexes sur tout le domaine étendu ?
La deuxième question est : permettent elles de construire des fonctions d'écart entre des champs de données qui ne sont pas des densités de probabilités ?
La réponse à ces 2 questions est cruciale, car c'est essentiellement sur ces deux points qu'on s'appuiera pout construire des divergences utilisables ici.
La troisième question est : les expressions de $Log_d(x)$ déduites des différentes entropies sont elles définies pour tout $x > 0$ avec les même paramètres que les

entropies associées, plus précisément, les propriétés de base du logarithme classique sont elles maintenues (concavité, comportement au voisinage de 0, etc) ? Cet aspect est traité en détails dans [15].
Ce dernier point est plus anecdotique, car les propriétés des fonctions Logarithme déformé n'interviennent pas directement dans notre travail, seules les expressions des entropies et des fonctions $f(x)$ correspondantes sont importantes.
Dans les sections suivantes, on passe en revue les expressions des entropies apparaissant dans le domaine de la Physique Statistique en reprenant d'abord le cas de l'entropie de Shannon qui servira de référence.

4 Entropie de Shannon - Logarithme naturel.

L'expression classique de l'entropie de Shannon reliée à la statistique de Boltzman-Gibbs s'écrit :

$$S(p) = -\sum_i p_i \log p_i \tag{5}$$

Ici, le logarithme qui intervient est le logarithme naturel.
La fonction convexe qui intervient s'écrit :

$$f(x) = x \log x \tag{6}$$

Ce n'est pas une fonction convexe standard ; la fonction convexe standard déduite de $f(x)$ s'écrit :

$$f_c(x) = x \log x + 1 - x \tag{7}$$

4.1 Divergence de Csiszär associée - Kullback-Leibler.

En s'appuyant sur la fonction $f_c(x)$ définie dans la section précédente, on obtient par application de (2) la divergence :

$$C(p\|q) = KL(p\|q) = \sum_i \left[p_i \log \frac{p_i}{q_i} - p_i + q_i \right] \tag{8}$$

C'est la divergence de Kullback-Leibler.
L'opposé de son gradient par rapport à "q" s'écrit $\forall j$:

$$-\frac{\partial KL(p\|q)}{\partial q_j} = \frac{p_j}{q_j} - 1 \tag{9}$$

Cette expression particulière se retrouvera pour les différentes divergences examinées ici et permettra de mettre en évidence les modifications liées à l'utilisation du "*Logarithme déformé*" à la place du Logarithme naturel.

4.2 Divergence de Csiszär duale.

Elle peut être construite à partir de (8) en permuttant les rôles de "p" et "q" ou bien en construisant une divergence de Csiszär sur la fonction convexe miroir $\hat{f}_c(x) = xf_c\left(\frac{1}{x}\right)$.
Elle s'écrit :

$$C(q\|p) = KL(q\|p) = \sum_i \left[q_i \log \frac{q_i}{p_i} - q_i + p_i\right] \tag{10}$$

L'opposé de son gradient par rapport à "q" s'écrit :

$$-\frac{\partial KL(q\|p)}{\partial q_j} = \log \frac{p_j}{q_j} \tag{11}$$

On remarquera que cette expression ne peut pas se décomposer de façon unique en une différence de 2 termes de même signe car cette décomposition change selon que "p" et "q" sont compris entre "0" et "1", ou supérieurs à "1".

4.3 Divergences de Bregman associées.

La divergence de Bregman construite sur "$f_c(x)$" (ou sur "$f(x)$") conduit à partir de (3) à la divergence de Kullback-Leibler donnée dans la section précédente ; il en est de même pour les divergences duales. C'est le (seul) point commun entre les divergences de Csiszär et les divergences de Bregman.

5 Entropie de Tsallis [19].

Avec le changement de notation ($q \leftrightarrow t$), l'entropie correspondante s'écrira :

$$S_T(p) = -\sum_i \frac{p_i^t - p_i}{t-1} \quad ; \quad t > 0 \tag{12}$$

On observera qu'en faisant le passage à la limite $t \to 1$ on retrouve l'entropie de Shannon.
Avec la définition (4) et le Logarithme déformé de Tsallis s'écrit :

$$\Lambda(x) = \log_T(x) = \frac{x^{t-1} - 1}{t-1} \tag{13}$$

Cette fonction est concave pour $t < 2$; elle est bien sur nulle pour $x = 1$.
Cependant, on trouve dans la litterature [2] une autre expression du Logarithme déformé de Tsallis qui s'écrit :

$$\log_{TT}(x) = \frac{x^{1-t} - 1}{1-t} \tag{14}$$

Cette fonction est équivalente à un logarithme pour $t > 0$.
Si on associe à cette forme une autre expression de l'entropie qui s'écrit :

$$S(p) = -\sum_i p_i^t \log_{TT}(p_i) \tag{15}$$

On retrouve l'entropie notée S_T (12).
Dans ce qui suit, on utilisera la forme (13) pour le logarithme déformé de Tsallis et on l'associera à la forme classique de l'entropie (4).
La fonction convexe permettant de construire des divergences s'écrira :

$$f(x) = \frac{x^t - x}{t-1} \tag{16}$$

Pour cette fonction, on a $f(1) = 0$ mais en revanche $f'(1) = 1$, ce n'est pas une fonction convexe standard, elle est par conséquent inadéquate pour construire des divergences de Csiszär utiisables dans notre problème.
La fonction convexe standard construite sur $f(x)$ s'écrit :

$$f_c(x) = \frac{x^t - x}{t-1} - x + 1 \tag{17}$$

5.1 Divergence de Csiszär fondée sur l'entropie de Tsallis.

La divergence de Csiszär construite sur la fonction convexe $f_c(x)$ définie dans la section précédente s'écrit, toutes simplifications faites :

$$C_T(p\|q) = \frac{1}{t-1}\left[\sum_i p_i^t q_i^{1-t} - t\sum_i p_i + (t-1)\sum_i q_i\right] \tag{18}$$

L'opposé du gradient s'écrit, tous calculs faits $\forall j$:

$$-\frac{\partial C_T(p\|q)}{\partial q_j} = \left(\frac{p_j}{q_j}\right)^t - 1 \tag{19}$$

Ce qu'on peut aussi écrire :

$$-\frac{\partial C_T(p\|q)}{\partial q_j} = \frac{p_j}{q_j}\left(\frac{p_j}{q_j}\right)^{t-1} - 1 \tag{20}$$

Avec cette écriture, en faisant $t = 1$, on retrouve immédiatement l'expression de l'opposé du gradient correspondant à la divergence de Kullback-Leibler (9) fondée sur l'entropie de Shannon.

5.2 Divergence de Csiszär duale fondée sur l'entropie de Tsallis.

Elle se déduit immédiatement de l'expression (18) en effectuant la permutation $p \leftrightarrow q$, mais elle peut être construite directement au sens de Csiszâr en s'appuyant sur la fonction convexe standard $\hat{f}_c(x)$ (fonction miroir), déduite de $f_c(x)$ par la transformation :

$$\hat{f}_c(x) = x f_c\left(\frac{1}{x}\right) \tag{21}$$

Elle s'écrit :

$$C_T(q\|p) = \frac{1}{t-1}\left[\sum_i q_i^t p_i^{1-t} - t\sum_i q_i + (t-1)\sum_i p_i\right] \tag{22}$$

L'opposé de son gradient par rapport à q est donné $\forall j$ par :

$$-\frac{\partial C_T(q\|p)}{\partial q_j} = \frac{t}{1-t}\left[\left(\frac{p_j}{q_j}\right)^{1-t} - 1\right] \tag{23}$$

Ce qui peut aussi s'écrire :

$$-\frac{\partial C_T(q\|p)}{\partial q_j} = \frac{1}{1-t}\left[t\left(\frac{p_j}{q_j}\right)^{1-t} - 1\right] + 1 \tag{24}$$

Le passage à la limite $t \to 1$ permet de retrouver le résultat obtenu pour la divergence de Csiszär duale correspondant à l'entropie de Shannon.

5.3 Divergence de Bregman fondée sur l'entropie de Tsallis.

La divergence de Bregman correspondante peut être construite indifféremment sur la fonction convexe $f(x)$ ou sur la fonction $f_c(x)$; elle s'écrit :

$$B_T(p\|q) = \sum_i \left[q_i^t - \frac{1}{1-t}p_i^t + \frac{t}{1-t}p_i q_i^{t-1}\right] \tag{25}$$

L'opposé du gradient par rapport à "q" s'écrit tous calculs faits $\forall j$:

$$-\frac{\partial B_T(p\|q)}{\partial q_j} = t\, q_j^{t-1}\left(\frac{p_j}{q_j} - 1\right) \tag{26}$$

5.4 Divergence de Bregman duale fondée sur l'entropie de Tsallis.

Elle se déduit immédiatement de l'expression (25) en effectuant la permutation $p \leftrightarrow q$; elle s'écrit :

$$B_T(q\|p) = \sum_i \left[p_i^t - \frac{1}{1-t}q_i^t + \frac{t}{1-t}q_i p_i^{t-1}\right] \tag{27}$$

L'opposé du gradient par rapport à "q" s'écrit tous calculs faits $\forall j$:

$$-\frac{\partial B_T(q\|p)}{\partial q_j} = \frac{t}{1-t}\, q_j^{t-1}\left[\left(\frac{p_j}{q_j}\right)^{t-1} - 1\right] \tag{28}$$

6 Entropie de Kaniadakis [8], [9], "K" entropie.

L'expression de l'Entropie s'écrit :

$$S_K(p) = -\sum_i p_i \log_K(p_i) = -\sum_i \frac{p_i^{1+K} - p_i^{1-K}}{2K} \quad ; \quad -1 < K < 1 \tag{29}$$

Elle est invariante lors de la transformation $K \leftrightarrow -K$, ce qui permet de limiter le domaine de variations de K à l'intervalle $0 < K < 1$.
L'entropie de Shannon est obtenue en faisant le passage à la limite $K \to 0$.
La fonction convexe déduite de cette expression sera :

$$f(x) = \frac{x^{1+K} - x^{1-K}}{2K} \tag{30}$$

Pour cette fonction, on a : $f(1) = 0$, mais $f'(1) = 1$; elle ne permet donc pas de construire au sens de Csiszär, des divergences utilisables dans nos problèmes. La fonction convexe standard déduite de celle-ci, permettant de construire des divergences utilisable dans nos problèmes s'écrira :

$$f_c(x) = \frac{x^{1+K} - x^{1-K}}{2K} - x + 1 \tag{31}$$

L'expression du logarithme déformé de Kaniadakis s'écrit :

$$\log_K(x) = \frac{x^K - x^{-K}}{2K} \tag{32}$$

C'est une fonction nulle si $x = 1$; elle est croissante et concave si $-1 < K < 1$.

6.1 Divergence de Csiszär construite sur la "K" entropie.

La divergence de Csiszär construite sur la fonction convexe standard $f_c(x)$ s'écrit, toutes simplifications faites :

$$C_K(p\|q) = \frac{1}{2K}\left[\sum_i p_i^{1+K} q_i^{-K} - \sum_i p_i^{1-K} q_i^K\right] - \sum_i p_i + \sum_i q_i \tag{33}$$

L'opposé du gradient par rapport à "q" s'écrit $\forall j$:

$$-\frac{\partial C_K(p\|q)}{\partial q_j} = \frac{1}{2}\left[\left(\frac{p_j}{q_j}\right)^{1+K} + \left(\frac{p_j}{q_j}\right)^{1-K}\right] - 1 \tag{34}$$

Ce qui peut encore s'écrire :

$$-\frac{\partial C_K(p\|q)}{\partial q_j} = \frac{p_j}{q_j}\left[\frac{1}{2}\left(\frac{p_j}{q_j}\right)^{K} + \frac{1}{2}\left(\frac{p_j}{q_j}\right)^{-K}\right] - 1 \tag{35}$$

Avec cette écriture, en faisant $K = 0$, on retrouve immédiatement l'expression de l'opposé du gradient correspondant à la divergence de Kullback-Leibler fondée sur l'entropie de Shannon dans laquelle le Logarithme déforme devient le Logarithme classique.

6.2 Divergence de Csiszär duale construite sur la "K" entropie.

Elle se déduit de la relation (33) en faisant la permutation $p \leftrightarrow q$; elle s'écrit :

$$C_K(q\|p) = \frac{1}{2K}\left[\sum_i q_i^{1+K} p_i^{-K} - \sum_i q_i^{1-K} p_i^K\right] - \sum_i q_i + \sum_i p_i \tag{36}$$

L'opposé du gradient par rapport à "q" s'écrit $\forall j$:

$$-\frac{\partial C_K(q\|p)}{\partial q_j} = \left[\frac{1-K}{2K}\left(\frac{p_j}{q_j}\right)^K - \frac{1+K}{2K}\left(\frac{p_j}{q_j}\right)^{-K}\right] + 1 \tag{37}$$

6.3 Divergence de Bregman construite sur la "K" entropie.

La divergence de Bregman correspondante peut être construite sur la fonction convexe $f(x)$ ou sur la fonction $f_c(x)$; elle s'écrit :

$$\begin{aligned} B_K(p\|q) =& \frac{1}{2K}\sum_i \left[p_i^{1+K} - p_i^{1-K} - q_i^{1+K} + q_i^{1-K}\right. \\ & -(1+K)p_i q_i^K + (1-K)p_i q_i^{-K} \\ & \left.+(1+K)q_i^{1+K} - (1-K)q_i^{1-K}\right] \end{aligned} \tag{38}$$

L'opposé du gradient par rapport à "q" s'écrit tous calculs faits $\forall j$:

$$-\frac{\partial B_K(p\|q)}{\partial q_j} = \left(\frac{p_j}{q_j} - 1\right)\left[\frac{1+K}{2} q_J^K + \frac{1-K}{2} q_J^{-K}\right] \tag{39}$$

6.4 Divergence de Bregman duale construite sur la "K" entropie.

Elle est construite à partir de la relation (38) en faisant la permutation $p \leftrightarrow q$; on obtient :

$$\begin{aligned} B_K(q\|p) =& \frac{1}{2K}\sum_i \left[q_i^{1+K} - q_i^{1-K} - p_i^{1+K} + p_i^{1-K}\right. \\ & -(1+K)q_i p_i^K + (1-K)q_i p_i^{-K} \\ & \left.+(1+K)p_i^{1+K} - (1-K)p_i^{1-K}\right] \end{aligned} \tag{40}$$

L'opposé du gradient par rapport à "q" s'écrit tous calculs faits $\forall j$:

$$-\frac{\partial B_K(q\|p)}{\partial q_j} = \left[\frac{1+K}{2K}\left(p_j^K - q_j^K\right) - \frac{1-K}{2K}\left(p_j^{-K} - q_j^{-K}\right)\right] \tag{41}$$

7 Entropie de Abe [1].

L'entropie correspondante s'écrit :

$$S_A(p) = -\sum_i p_i \log_A(p_i) = -\sum_i \frac{p_i^z - p_i^{\frac{1}{z}}}{z - \frac{1}{z}} \tag{42}$$

Elle est invariante dans la transformation $z \leftrightarrow z^{-1}$ et de ce fait, le domaine de valeurs du paramètre "z" peut être restreint à l'intervalle $0 < z \leq 1$.
L'entropie de Shannon se retrouve en faisant le passage à la limite $z \to 1$.
La fonction convexe permettant de construire des divergences s'écrira :

$$f(x) = \frac{x^z - x^{\frac{1}{z}}}{z - \frac{1}{z}} \tag{43}$$

Pour cette fonction, on a $f(1) = 0$ mais en revanche $f'(1) = 1$, ce n'est pas une fonction convexe standard, elle est par conséquent inadéquate pour construire des divergences de Csiszär utiisables dans notre problème.
La fonction convexe standard construite sur $f(x)$ s'écrit :

$$f_c(x) = \frac{x^z - x^{\frac{1}{z}}}{z - \frac{1}{z}} - x + 1 \tag{44}$$

Avec la définition (4) l'expression du Logarithme déformé de Abe s'écrit :

$$\log_A(x) = \frac{x^{z-1} - x^{\frac{1}{z}-1}}{z - \frac{1}{z}} \tag{45}$$

Cette fonction est croissante et concave pour $0.5 < z < 2$; elle est invariante dans la transformation $z \leftrightarrow z^{-1}$.

7.1 Divergence de Csiszär fondée sur l'entropie de Abe.

La fonction convexe standard $f_c(x)$ définie dans la section précédente permet de construire une divergence de Csiszär adaptée à notre problème ; elle s'écrit :

$$C_A(p\|q) = \frac{z}{z^2 - 1}\left[\sum_i p_i^z q_i^{1-z} - \sum_i p_i^{\frac{1}{z}} q_i^{1-\frac{1}{z}}\right] - \sum_i p_i + \sum_i q_i \tag{46}$$

L'opposé du gradient par rapport à "q" s'écrit $\forall j$:

$$-\frac{\partial C_A(p\|q)}{\partial q_j} = \frac{z}{z+1}\left(\frac{p_j}{q_j}\right)^z + \frac{1}{z+1}\left(\frac{p_j}{q_j}\right)^{\frac{1}{z}} - 1 \tag{47}$$

Ce qui peut aussi s'écrire :

$$-\frac{\partial C_A(p\|q}{\partial q_j} = \frac{p_j}{q_j}\left[\frac{z}{z+1}\left(\frac{p_j}{q_j}\right)^{z-1} + \frac{1}{z+1}\left(\frac{p_j}{q_j}\right)^{\frac{1}{z}-1}\right] - 1 \tag{48}$$

Avec cette écriture, en faisant $z = 1$, on retrouve immédiatement l'expression de l'opposé du gradient correspondant à la divergence de Kullback-Leibler fondée sur l'entropie de Shannon.

7.2 Divergence de Csiszär duale fondée sur l'entropie de Abe.

Elle se déduit de l'expression (46) en effectuant la permutation $p \leftrightarrow q$; on obtient :

$$C_A\left(q\|p\right) = \frac{z}{z^2-1}\left[\sum_i q_i^z p_i^{1-z} - \sum_i q_i^{\frac{1}{z}} p_i^{1-\frac{1}{z}}\right] - \sum_i q_i + \sum_i p_i \tag{49}$$

L'opposé du gradient par rapport à "q" s'écrit $\forall j$:

$$-\frac{\partial C_A(q\|p)}{\partial q_j} = \frac{z}{1-z^2}\left[z\left(\frac{p_j}{q_j}\right)^{1-z} - \frac{1}{z}\left(\frac{p_j}{q_j}\right)^{1-\frac{1}{z}}\right] + 1 \tag{50}$$

Ce qui peut aussi s'écrire :

$$-\frac{\partial C_A(q\|p)}{\partial q_j} = \frac{p_j}{q_j}\left[\frac{z^2}{1-z^2}\left(\frac{p_j}{q_j}\right)^{-z} - \frac{1}{1-z^2}\left(\frac{p_j}{q_j}\right)^{-\frac{1}{z}}\right] + 1 \tag{51}$$

7.3 Divergence de Bregman fondée sur l'entropie de Abe.

La divergence de Bregman correspondante peut être construite sur la fonction convexe $f(x)$ ou sur la fonction $f_c(x)$; elle s'écrit :

$$\begin{aligned} B_A(p\|q) = & \frac{1}{z-\frac{1}{z}}\sum_i \Big[p_i^z - p_i^{\frac{1}{z}} - q_i^z + q_i^{\frac{1}{z}} \\ & -zp_iq_i^{(z-1)} + \frac{1}{z}p_iq_i^{(\frac{1}{z}-1)} \\ & +zq_i^z - \frac{1}{z}q_i^{\frac{1}{z}}\Big] \end{aligned} \tag{52}$$

L'opposé du gradient par rapport à "q" s'écrit $\forall j$:

$$-\frac{\partial B_A(p\|q)}{\partial q_j} = \left(\frac{p_j}{q_j} - 1\right)\left[\frac{z^2}{z+1}q_j^{(z-1)} + \frac{1}{z(z+1)}q_j^{(\frac{1}{z}-1)}\right] \tag{53}$$

7.4 Divergence de Bregman duale fondée sur l'entropie de Abe.

Elle se déduit de l'expression (52) en effectuant la permutation $p \leftrightarrow q$; elle s'écrit :

$$\begin{aligned} B_A(q\|p) =& \frac{1}{z-\frac{1}{z}} \sum_i \Big[q_i^z - q_i^{\frac{1}{z}} - p_i^z + p_i^{\frac{1}{z}} \\ & -zq_i p_i^{(z-1)} + \frac{1}{z} q_i p_i^{(\frac{1}{z}-1)} \\ & +zp_i^z - \frac{1}{z} p_i^{\frac{1}{z}} \Big] \end{aligned} \tag{54}$$

L'opposé du gradient par rapport à "q" s'écrit $\forall j$:

$$-\frac{\partial B_A(q\|p)}{\partial q_j} = \left[\frac{z^2}{z^2-1}(p_j^{z-1} - q_j^{z-1}) - \frac{1}{z^2-1}(p_j^{\frac{1}{z}-1} - q_j^{\frac{1}{z}-1}) \right] \tag{55}$$

8 Gamma (γ) entropie.

L'entropie "γ" mentionnée dans [10], s'écrit :

$$S_\gamma(p) = -\sum_i p_i \log_\gamma(p_i) = -\sum_i \frac{p_i^{1+2\gamma} - p_i^{1-\gamma}}{3\gamma} \tag{56}$$

Cette fonction est définie pour $-0.5 \leq \gamma \leq 1$.
Le passage à la limite $\gamma \to 0$ permet de retrouver l'entropie de Shannon.
La fonction convexe permettant de construire des divergences s'écrit :

$$f(x) = \frac{x^{1+2\gamma} - x^{1-\gamma}}{3\gamma} \tag{57}$$

Pour cette fonction, on a $f(1) = 0$ mais en revanche $f'(1) = 1$, ce n'est pas une fonction convexe standard, elle est par conséquent inadéquate pour construire des divergences de Csiszär utiisables dans notre problème.
La fonction convexe standard construite sur $f(x)$ s'écrit :

$$f_c(x) = \frac{x^{1+2\gamma} - x^{1-\gamma}}{3\gamma} + 1 - x \tag{58}$$

Avec la définition (4) l'expression du "γ" Logarithme déformé s'écrit :

$$\log_\gamma(x) = \frac{x^{2\gamma} - x^{-\gamma}}{3\gamma} \tag{59}$$

Comme indiqué dans Kaniadakis et al. [10], cette fonction est nulle pour $x = 1$, elle est croissante et concave pour $-0.5 < \gamma < 0.5$.

8.1 Divergence de Csiszär construite sur la "γ" entropie.

La fonction $f_c(x)$ définie dans la section précédente permet de construire la divergence de Csiszär suivante :

$$C_\gamma(p\|q) = \frac{1}{3\gamma}\left[\sum_i p_i^{1+2\gamma} q_i^{-2\gamma} - \sum_i p_i^{1-\gamma} q_i^{\gamma}\right] - \sum_i p_i + \sum_i q_i \tag{60}$$

L'opposé du gradient par rapport à "q" s'écrit $\forall j$:

$$-\frac{\partial C_\gamma(p\|q)}{\partial q_j} = \frac{2}{3}\left(\frac{p_j}{q_j}\right)^{1+2\gamma} + \frac{1}{3}\left(\frac{p_j}{q_j}\right)^{1-\gamma} - 1 \tag{61}$$

Ce qui peut aussi s'écrire :

$$-\frac{\partial C_\gamma(p\|q)}{\partial q_j} = \frac{p_j}{q_j}\left[\frac{2}{3}\left(\frac{p_j}{q_j}\right)^{2\gamma} + \frac{1}{3}\left(\frac{p_j}{q_j}\right)^{-\gamma}\right] - 1 \tag{62}$$

Avec cette écriture, en faisant $\gamma = 0$, on retrouve immédiatement l'expression de l'opposé du gradient correspondant à la divergence de Kullback-Leibler fondée sur l'entropie de Shannon.

8.2 Divergence de Csiszär duale construite sur la "γ" entropie.

Elle est construite à partir de l'expression (60), en effectuant la permutation $p \leftrightarrow q$

$$C_\gamma(q\|p) = \frac{1}{3\gamma}\left[\sum_i q_i^{1+2\gamma} p_i^{-2\gamma} - \sum_i q_i^{1-\gamma} p_i^{\gamma}\right] - \sum_i q_i + \sum_i p_i \tag{63}$$

L'opposé du gradient par rapport à "q" s'écrit $\forall j$:

$$-\frac{\partial C_\gamma(q\|p)}{\partial q_j} = \frac{1}{3\gamma}\left[(1-\gamma)\left(\frac{p_j}{q_j}\right)^{\gamma} - (1+2\gamma)\left(\frac{p_j}{q_j}\right)^{-2\gamma}\right] + 1 \tag{64}$$

Ou encore :

$$-\frac{\partial C_\gamma(q\|p)}{\partial q_j} = \frac{p_j}{q_j}\left[\frac{1-\gamma}{3\gamma}\left(\frac{p_j}{q_j}\right)^{\gamma-1} - \frac{1+2\gamma}{3\gamma}\left(\frac{p_j}{q_j}\right)^{-2\gamma-1}\right] + 1 \tag{65}$$

8.3 Divergence de Bregman construite sur la "γ" entropie.

La divergence de Bregman correspondante peut être construite sur la fonction convexe $f(x)$ ou sur la fonction $f_c(x)$; elle s'écrit :

$$\begin{aligned} B_\gamma(p\|q) = & \frac{1}{3\gamma}\sum_i \Big[p_i^{2\gamma+1} - p_i^{1-\gamma} - q_i^{2\gamma+1} + q_i^{1-\gamma} \\ & -(2\gamma+1)p_i q_i^{2\gamma} + (1-\gamma)p_i q_i^{-\gamma} \\ & +(2\gamma+1)q_i^{2\gamma+1} - (1-\gamma)q_i^{1-\gamma}\Big] \end{aligned} \tag{66}$$

L'opposé du gradient par rapport à "q" s'écrit $\forall j$:

$$-\frac{\partial B_\gamma(p\|q)}{\partial q_j} = \left(\frac{p_j}{q_j} - 1\right)\left[\frac{2(2\gamma+1)}{3}q_j^{2\gamma} + \frac{1-\gamma}{3}q_j^{-\gamma}\right] \tag{67}$$

8.4 Divergence de Bregman duale construite sur la "γ" entropie.

Elle est construite à partir de la relation (66), en effectuant la permutation $p \leftrightarrow q$; elle s'écrit :

$$\begin{aligned} B_\gamma(q\|p) =& \frac{1}{3\gamma}\sum_i \Big[q_i^{2\gamma+1} - q_i^{1-\gamma} - p_i^{2\gamma+1} + p_i^{1-\gamma} \\ & -(2\gamma+1)q_i p_i^{2\gamma} + (1-\gamma)q_i p_i^{-\gamma} \\ & +(2\gamma+1)p_i^{2\gamma+1} - (1-\gamma)p_i^{1-\gamma}\Big] \end{aligned} \tag{68}$$

L'opposé du gradient par rapport à "q" s'écrit $\forall j$:

$$-\frac{\partial B_\gamma(q\|p)}{\partial q_j} = \left[\frac{2\gamma+1}{3\gamma}\left(p_j^{2\gamma} - q_j^{2\gamma}\right) - \frac{1-\gamma}{3\gamma}\left(p_j^{-\gamma} - q_j^{-\gamma}\right)\right] \tag{69}$$

9 Entropie à 2 paramètres [10], K.L.S.

L'expression de l'entropie à 2 paramètres (K, r) proposée par Kaniadakis et al. [10] s'écrit :

$$S_{K,r}(p) = -\sum_i p_i Log_{(K,r)}(p_i) = -\sum_i p_i^{1+r}\frac{p_i^K - p_i^{-K}}{2K} \tag{70}$$

Pour retrouver l'entropie de Shannon, on procède en 2 temps : on fait d'abord $r = 0$, puis on effectue le passage à la limite $K \to 0$.
Avec la définition (4), l'expression du Logarithme déformé à 2 paramètres (K, r) proposée par Kaniadakis et al. [10] s'écrit :

$$Log_{(K,r)}(x) = x^r\frac{x^K - x^{-K}}{2K} \tag{71}$$

Cette fonction est croissante si $-|K| \leq r \leq |K|$.
Elle est concave :
- pour $-|K| \leq r \leq |K|$ si $|K| \leq \frac{1}{2}$
ou bien,
- pour $-|K| \leq r \leq 1 - |K|$ si $|K| \geq \frac{1}{2}$.

La fonction convexe de base déduite de l'expression de l'entropie s'écrit :

$$f(x) = \frac{x^{1+K+r} - x^{1-K+r}}{2K} \tag{72}$$

Ce n'est pas une fonction convexe standard car $f'(1) = 1$; la fonction convexe standard déduite de $f(x)$ s'écrira :

$$f_c(x) = \frac{x^{1+K+r} - x^{1-K+r}}{2K} + 1 - x \tag{73}$$

9.1 Divergence de Csiszär construite sur l'entropie à 2 paramètres.

La divergence de Csiszär construite sur la fonction $f_c(x)$ définie dans la section précédente s'écrit :

$$C_{KLS}(p\|q) = \frac{1}{2K}\left[\sum_i p_i^{1+r+K} q_i^{-r-K} - \sum_i p_i^{1+r-K} q_i^{-r+K}\right] - \sum_i p_i + \sum_i q_i \tag{74}$$

L'opposé du gradient par rapport à "q" s'écrit tous calculs faits $\forall j$:

$$-\frac{\partial C_{KLS}(p\|q)}{\partial q_j} = \frac{1}{2K}\left[(K+r)\left(\frac{p_j}{q_j}\right)^{1+r+K} + (K-r)\left(\frac{p_j}{q_j}\right)^{1+r-K}\right] - 1 \tag{75}$$

Ce qui peut aussi s'écrire :

$$-\frac{\partial C_{KLS}(p\|q)}{\partial q_j} = \left(\frac{p_j}{q_j}\right)\left[\frac{K+r}{2K}\left(\frac{p_j}{q_j}\right)^{r+K} + \frac{K-r}{2K}\left(\frac{p_j}{q_j}\right)^{r-K}\right] - 1 \tag{76}$$

Si, à partir de cette expression, on fait $r = 0$, on retrouve ce qu'on a obtenu pour le Logarithme déformé de Kaniadakis (34), puis en faisant $K \to 0$ on retrouve le cas correspondant au Logarithme classique et à la divergence de KullbacK Leibler.
La forme intéressante de l'algorithme sera obtenue à la condition que : $-K \leq r \leq K$.

9.2 Divergence de Csiszär duale construite sur l'entropie à 2 paramètres.

Elle est construite à partir de l'expression (74) en effectuant la permutation $p \leftrightarrow q$; elle s'écrit :

$$C_{KLS}(q\|p) = \frac{1}{2K}\left[\sum_i q_i^{1+r+K} p_i^{-r-K} - \sum_i q_i^{1+r-K} p_i^{-r+K}\right] - \sum_i q_i + \sum_i p_i \tag{77}$$

L'opposé du gradient par rapport à "q" s'écrit tous calculs faits $\forall j$:

$$-\frac{\partial C_{KLS}(q\|p)}{\partial q_j} = \frac{1}{2K}\left[(1+r-K)\left(\frac{p_j}{q_j}\right)^{K-r} - (1+r+K)\left(\frac{p_j}{q_j}\right)^{-K-r}\right] + 1 \tag{78}$$

Ou encore :

$$-\frac{\partial C_{KLS}(q\|p)}{\partial q_j} = \frac{p_j}{q_j}\left[\frac{1+r-K}{2K}\left(\frac{p_j}{q_j}\right)^{K-r-1} - \frac{1+r+K}{2K}\left(\frac{p_j}{q_j}\right)^{-K-r-1}\right] + 1 \tag{79}$$

9.3 Divergence de Bregman construite sur l'entropie à 2 paramètres.

La divergence de Bregman construite sur les fonctions $f(x)$ ou $f_c(x)$ définies dans la section précédente s'écrit :

$$\begin{aligned} B_{KLS}(p\|q) = \frac{1}{2K}\sum_i \big[&p_i^{1+r+K} - p_i^{1+r-K} - q_i^{1+r+K} + q_i^{1+r-K} \\ &-(1+r+K)p_i q_i^{r+K} + (1+r+K)q_i^{1+r+K} \\ &+(1+r-K)p_i q_i^{r-K} - (1+r-K)q_i^{1+r-K}\big] \end{aligned} \tag{80}$$

L'opposé du gradient par rapport à "q" s'écrit après quelques calculs $\forall j$:

$$\begin{aligned} -\frac{\partial B_{KLS}(p\|q)}{\partial q_j} = \left(\frac{p_j}{q_j} - 1\right)\Bigg[&\frac{(1+r+K)(r+K)}{2K}q_j^{r+K} \\ &-\frac{(1+r-K)(r-K)}{2K}q_j^{r-K}\Bigg] \end{aligned} \tag{81}$$

En posant : $a = 1+r+K$ et $b = 1+r-K$ on obtient une divergence générale (108) qui sera analysée ultèrieurement

9.4 Divergence de Bregman duale construite sur l'entropie à 2 paramètres.

Elle est construite à partir de la relation (80) en effectuant la permutation $p \leftrightarrow q$.

$$\begin{aligned} B_{KLS}(q\|p) = \frac{1}{2K}\sum_i \big[&q_i^{1+r+K} - q_i^{1+r-K} - p_i^{1+r+K} + p_i^{1+r-K} \\ &-(1+r+K)q_i p_i^{r+K} + (1+r+K)p_i^{1+r+K} \\ &+(1+r-K)q_i p_i^{r-K} - (1+r-K)p_i^{1+r-K}\big] \end{aligned} \tag{82}$$

L'opposé du gradient par rapport à "q" s'écrit $\forall j$:

$$-\frac{\partial B_{KLS}(q\|p)}{\partial q_j} = \left(\frac{1+r+K}{2K}\right)\left(p_j^{r+K} - q_j^{r+K}\right) - \left(\frac{1+r-K}{2K}\right)\left(p_j^{r-K} - q_j^{r-K}\right) \tag{83}$$

10 Autres définitions de l'entropie et du Logarithme déformé.

Dans ce paragraphe, on reprend d'autres façons de définir l'entropie qui permettent de retrouver les expressions des paragraphes 4 à 9 et de donner une forme trés générale de l'entropie qui regroupe celles-ci.
Les expressions de l'entropie qui seront mentionnées dans le paragraphe 12 n'entrent pas dans ce contexte.
Dans les sections précédentes la forme de l'entropie utilisée était définie par la relation :

$$S(p) = -\sum_i p_i \Lambda(p_i) \tag{84}$$

ou $\Lambda(p_i)$ désigne une forme du Logarithme généralisé ; le Logarithme classique correspondant à l'entropie de Shannon.
Une autre définition de l'entropie de Shannon proposée par Abe [1] s'écrit :

$$S(p) = \left[-\frac{d}{d\alpha} \sum_i p_i^\alpha \right]_{\alpha=1} \equiv \left[-\frac{d}{d\alpha} f(\alpha) \right]_{\alpha=1} \tag{85}$$

Dans cette expression la dérivée mise en jeu est la dérivée habituelle.
A partir de (85), il est proposé de déformer la dérivée classique et de la remplaçer par la "q" différentielle de Jackson [6] [7] qui s'écrit, en changeant la notation ($q \leftrightarrow t$)

$$\frac{df(\alpha)}{d(\alpha : t)} = \frac{f(\alpha t) - f(\alpha)}{\alpha t - \alpha} \tag{86}$$

(On pourrait presque dire qu'on a remplaçé l'accroissement additif de la variable dans la dérivée classique, par un accroissement multiplicatif de la variable.)
Avec cette définition, on obtient :

$$S(p) = \left[-\frac{\sum_i p_i^{\alpha t} - \sum_i p_i^\alpha}{\alpha t - \alpha} \right]_{\alpha=1} \tag{87}$$

Ce qui correspond à l'entropie de Tsallis [19], déja mentionnée en (12) :

$$S_T(p) = -\frac{\sum_i p_i^t - \sum_i p_i}{t-1} = -\frac{\sum_i p_i^t - 1}{t-1} \tag{88}$$

Le Logarithme déformé correspondant s'écrit (13) :

$$Log_T(x) = \frac{x^{t-1} - 1}{t-1} \tag{89}$$

Une alternative consiste, toujours à partir de (85), à adopter une autre définition de la différentielle donnée par Mc Anally [13] qui s'écrit :

$$\frac{df(\alpha)}{d(\alpha : t)} = \frac{f(\alpha t) - f(\alpha t^{-1})}{\alpha t - \alpha t^{-1}} \tag{90}$$

On a ainsi une invariance $t \leftrightarrow t^{-1}$ et on obtient :

$$S(p) = \left[-\frac{\sum_i p_i^{\alpha t} - \sum_i p_i^{\alpha t^{-1}}}{\alpha t - \alpha t^{-1}} \right]_{\alpha=1} = -\frac{\sum_i p_i^t - \sum_i p_i^{t^{-1}}}{t - t^{-1}} \tag{91}$$

On retrouve ainsi l'entropie de Abe (42) ; le Logarithme déformé correspondant s'écrit (45) :

$$Log_A(x) = \frac{x^{t-1} - x^{t^{-1}-1}}{t - t^{-1}} \tag{92}$$

Une autre forme de différentielle généralisée à 2 paramètres, initialement utilisée par Chakrabarti et Jagannathan [4], est proposée par Borges et Roditi [3] ; elle consiste à utiliser la définition suivante :

$$\frac{d_{ab} f(\alpha)}{d_{ab}(\alpha)} = \frac{f(a\alpha) - f(b\alpha)}{a\alpha - b\alpha} \tag{93}$$

Cette forme étend les définitions précédemment proposées, elle conduit, avec la définition (85) à l'expression de l'entropie ;

$$S_{ab}(p) = \left[-\frac{\sum_i p_i^{a\alpha} - \sum_i p_i^{b\alpha}}{a\alpha - b\alpha} \right]_{\alpha=1} = -\frac{\sum_i p_i^a - \sum_i p_i^b}{a - b} \tag{94}$$

Cette forme d'entropie très générale (et le logarithme déforme associé) est mentionnée par Wada et Scarfone [21] ; elle est fondée sur les travaux plus anciens de Sharma et Taneja [17] et Mittal [14] et permet bien entendu de retrouver les expressions précédentes des paragraphes 4 à 9.
Les domaines de valeurs des paramètres "a" et "b" indiqués dans [5] et [3] sont les suivants :

$$0 \leq a \leq 1 \leq b \tag{95}$$

ou bien :

$$0 \leq b \leq 1 \leq a \tag{96}$$

L'expression du Logarithme déformé déduite de cette forme de l'entropie s'écrit :

$$Log_{ab}(x) = \frac{x^{a-1} - x^{b-1}}{a - b} \tag{97}$$

Le logarithme déformé ainsi défini possède les propriétés de concavité du Logarithme classique dans un domaine de valeurs des paramètres donné par :

$$0 \leq a \leq 1 \leq b \leq 2 \quad ; \quad 0 \leq b \leq 1 \leq a \leq 2 \tag{98}$$

11 Divergences fondées sur l'entropie générale (94).

On développe dans les paragraphes suivants les différentes formes de divergences fondées sur l'entropie générale (94).
Les cas particuliers développés dans les paragraphes 4 à 9 se retrouvent en effectuant les adaptations suivantes :

- Entropie de Shannon : $a \to 1$ et $b \to 1$ en respectant (95) ou (96)
- Entropie de Tsallis : $a = t, \quad b = 1$
- Entropie de Kaniadakis : $a = 1 + K, \quad b = 1 - K$
- Entropie de Abe : $a = z, \quad b = \frac{1}{z}$
- Entropie "γ" : $a = 2\gamma + 1, \quad b = 1 - \gamma$
- Entropie à 2 paramètres (KLS) : $a = 1 + r + K, \quad b = 1 + r - K$

11.1 Divergence de Csiszär construite sur l'entropie générale.

La fonction de base convexe déduite de la forme générale de l'entropie (94) s'écrit :

$$f(x) = \frac{x^a - x^b}{a - b} \tag{99}$$

Ce n'est pas une fonction complète car $f'(1) = 1$; la fonction convexe standard déduite de $f(x)$ s'écrit :

$$f_c(x) = \frac{x^a - x^b}{a - b} - x + 1 \tag{100}$$

La divergence de Csiszär construite sur la base de cette fonction s'écrit :

$$C_{ab}(p\|q) = \frac{1}{a - b}\left[\sum_i p_i^a q_i^{1-a} - \sum_i p_i^b q_i^{1-b}\right] - \sum_i p_i + \sum_i q_i \tag{101}$$

A partir de cette expression, en s'appuyant sur les inégalités (95) ou (96), on retrouve la divergence de Kullback-Leibler en faisant le passage à la limite $a \to 1$ et $b \to 1$.
L'opposé de son gradient par rapport à "q" s'écrit tous calculs faits $\forall j$:

$$-\frac{\partial C_{ab}(p\|q)}{\partial q_j} = \left[\frac{a-1}{a-b}\left(\frac{p_j}{q_j}\right)^a + \frac{1-b}{a-b}\left(\frac{p_j}{q_j}\right)^b\right] - 1 \tag{102}$$

Ce qui peut encore s'écrire :

$$-\frac{\partial C_{ab}(p\|q)}{\partial q_j} = \left(\frac{p_j}{q_j}\right)\left[\frac{a-1}{a-b}\left(\frac{p_j}{q_j}\right)^{a-1} + \frac{1-b}{a-b}\left(\frac{p_j}{q_j}\right)^{b-1}\right] - 1 \tag{103}$$

La décomposition de l'opposé du gradient permettant d'obtenir une forme multiplicative de l'algorithme d'inversion implique :

$$\frac{a-1}{a-b} > 0 \quad ; \quad \frac{1-b}{a-b} > 0 \tag{104}$$

Ces inégalités sont satisfaites si "a" et "b" appartiennent aux domaines de valeurs donnés par (95) ou (96).

11.2 Divergence de Csiszär duale construite sur l'entropie générale.

Elle est construite à partir de l'expression (101) en effectuant la permutation $p \leftrightarrow q$; elle s'écrit :

$$C_{ab}(q\|p) = \frac{1}{a-b}\left[\sum_i q_i^a p_i^{1-a} - \sum_i q_i^b p_i^{1-b}\right] - \sum_i q_i + \sum_i p_i \tag{105}$$

L'opposé de son gradient par rapport à "q" s'écrit tous calculs faits $\forall j$:

$$-\frac{\partial C_{ab}(q\|p)}{\partial q_j} = \frac{1}{a-b}\left[b\left(\frac{p_j}{q_j}\right)^{1-b} - a\left(\frac{p_j}{q_j}\right)^{1-a}\right] + 1 \tag{106}$$

Ou encore :

$$-\frac{\partial C_{ab}(q\|p)}{\partial q_j} = \frac{p_j}{q_j}\left[\frac{b}{a-b}\left(\frac{p_j}{q_j}\right)^{-b} - \frac{a}{a-b}\left(\frac{p_j}{q_j}\right)^{-a}\right] + 1 \tag{107}$$

Attention : la décomposition de l'opposé du gradient en une différence de 2 termes positifs change suivant le signe de $a-b$.

11.3 Divergence de Bregman construite sur l'entropie générale.

Cette divergence peut être construite soit sur la base de la fonction $f(x)$, soit sur la base de $f_c(x)$; l'expression obtenue est identique et s'écrit tous calculs effectués :

$$\begin{aligned} B_{ab}(p\|q) = \frac{1}{a-b}\sum_i \big[&p_i^a - p_i^b + (a-1)q_i^a - (b-1)q_i^b \\ &-ap_i q_i^{a-1} + bp_i q_i^{b-1}\big] \end{aligned} \tag{108}$$

L'opposé du gradient par rapport à "q" s'écrit, $\forall j$, après factorisation :

$$-\frac{\partial B_{ab}(p\|q)}{\partial q_j} = \left(\frac{p_j}{q_j} - 1\right)\left[\frac{(a^2-a)}{a-b}q_j^{a-1} + \frac{(b-b^2)}{a-b}q_j^{b-1}\right] \tag{109}$$

Compte tenu des domaimes de valeurs de "a" et "b" définies dans (95) ou (96), on a :

$$\frac{a^2-a}{a-b} > 0 \quad ; \quad \frac{b-b^2}{a-b} > 0 \tag{110}$$

Ceci permettra d'écrire sans ambiguité une forme purement multiplicative de l'algorithme d'inversion.

11.4 Divergence de Bregman duale construite sur l'entropie générale.

Elle se déduit de l'expression (108) par permutation de "p" et "q" ; elle s'écrit :

$$B_{ab}(q\|p) = \frac{1}{a-b}\sum_i \left[q_i^a - q_i^b + (a-1)p_i^a - (b-1)p_i^b \right.$$
$$\left. -aq_ip_i^{a-1} + bq_ip_i^{b-1}\right] \tag{111}$$

L'opposé du gradient par rapport à "q" s'écrit, $\forall j$:

$$-\frac{\partial B_{ab}(q\|p)}{\partial q_j} = \frac{1}{a-b}\left[a\left(p_j^{a-1} - q_j^{a-1}\right) - b\left(p_j^{b-1} - q_j^{b-1}\right)\right] \tag{112}$$

Ce qui peut aussi s'écrire :

$$-\frac{\partial B_{ab}(q\|p)}{\partial q_j} = \frac{1}{a-b}\left[(ap_j^{a-1} + bq_j^{b-1}) - (aq_j^{a-1} + bp_j^{b-1})\right] \tag{113}$$

Attention : la décomposition de l'opposé du gradient en une différence de 2 termes positifs change suivant le signe de $a-b$.

12 Quelques autres expressions du Logarithme déformé.

On trouve dans la litterature des expressions du Logarithme déformé qui ne sont pas en accord avec la forme générale décrite dans les sections précédentes, on les examine ici. Elles sont citées par Trivellato [18].

12.1 Logarithme déformé de Newton [16].

Il est donné par l'expression :

$$L_N(x) = \frac{1}{2}\left(x - 1 + Log x\right) \tag{114}$$

En s'appuyant sur la relation (4) on obtient pour l'entropie, l'expression spécifique :

$$S_N(p) = -\sum_i p_i L_N(p_i) = -\sum_i \frac{1}{2}\left[p_i^2 - p_i + p_i\, Log\, p_i\right] \tag{115}$$

La fonction convexe de base s'écrit en conséquence :

$$f(x) = \frac{1}{2}\left(x^2 - x + x\, Log\, x\right) \tag{116}$$

Ce n'est pas une fonction complète, en effet $f^{'}(1) = 1$.
La fonction convexe standard qu'on en déduit, s'écrit :

$$f_c(x) = \frac{1}{2}\left(x^2 - 3x + 2 + x\, Log\, x\right) \tag{117}$$

12.1.1 Divergence de Csiszär correspondant à l'entropie de Newton.

La divergence de Csiszär fondée sur la fonction $f_c(x)$ précédente s'écrit :

$$C_N(p\|q) = \frac{1}{2}\sum_i \left[p_i^2 q_i^{-1} - 3p_i + 2q_i + p_i Log\left(\frac{p_i}{q_i}\right)\right] \tag{118}$$

L'opposé du gradient par rapport à "q" s"'crit $\forall j$:

$$-\frac{\partial C_N(p\|q)}{\partial q_j} = \frac{p_j}{q_j}\left[\frac{1}{2}\frac{p_j}{q_j} + \frac{1}{2}\right] - 1 \tag{119}$$

Il est bien sur égal à "0" si $p_j = q_j \quad \forall j$.

12.1.2 Divergence de Csiszär duale correspondant à l'entropie de Newton.

Elle se déduit de l'expression (118) en effectuant la permutation $p \leftrightarrow q$; elle s'écrit :

$$C_N(q\|p) = \frac{1}{2}\sum_i \left[q_i^2 p_i^{-1} - 3q_i + 2p_i + q_i Log\left(\frac{q_i}{p_i}\right)\right] \tag{120}$$

L'opposé du gradient par rapport à "q" s'écrit $\forall j$:

$$-\frac{\partial C_N(q\|p)}{\partial q_j} = \frac{1}{2}\log\frac{p_j}{q_j} + 1 - \frac{q_j}{p_j} \tag{121}$$

Il est bien sur égal à "0" si $p_j = q_j \quad \forall j$.

12.1.3 Divergence de Bregman correspondant à l'entropie de Newton.

La divergence de Bregman peut être construite sur la fonction convexe $f(x)$ ou bien sur la fonction convexe standard $f_c(x)$ et on obtient :

$$B_N(p\|q) = \frac{1}{2}\sum_i \left[p_i^2 - p_i + p_i \log p_i + q_i^2 + q_i - 2p_i q_i - p_i \log q_i\right] \tag{122}$$

L'opposé du gradient s'écrit $\forall j$:

$$-\frac{\partial B_N(p\|q)}{\partial q_j} = \left(\frac{p_j}{q_j} - 1\right)\left(\frac{1}{2} + q_j\right) \tag{123}$$

12.1.4 Divergence de Bregman duale correspondant à l'entropie de Newton.

Elle est déduite de la relation (122) en effectuant la permutation $p \leftrightarrow q$; elle s'écrit ;

$$B_N(q\|p) = \frac{1}{2}\sum_i \left[q_i^2 - q_i + q_i \log q_i + p_i^2 + p_i - 2q_i p_i - q_i \log p_i\right] \tag{124}$$

L'opposé du gradient s'écrit $\forall j$:

$$-\frac{\partial B_N(q\|p)}{\partial q_j} = q_j\left(\frac{p_j}{q_j}-1\right)+\frac{1}{2}\log\frac{p_j}{q_j} \tag{125}$$

Ce qui peut aussi s'écrire :

$$-\frac{\partial B_N(q\|p)}{\partial q_j} = \left(p_j+\frac{1}{2}\log p_j\right)-\left(q_j+\frac{1}{2}\log q_j\right) \tag{126}$$

12.2 Autre expression du Logarithme déformé [9].

Il s'écrit :

$$Log_\alpha(x) = \frac{1}{\alpha}\frac{x^\alpha - x^{-\alpha}}{x^\alpha + x^{-\alpha}} \tag{127}$$

Il est défini et concave pour $-0.5 \leq \alpha \leq 0.5$; il est invariant dans la transformation $\alpha \leftrightarrow -\alpha$, ce qui permet de réduire le domaine de valeurs du paramètre à $0 \leq \alpha \leq 0.5$; le logarithme classique est obtenu pour $\alpha \to 0$.
En s'appuyant sur la relation (4) on obtient pour l'entropie, l'expression spécifique :

$$S_\alpha(p) = -\sum_i p_i Log_\alpha(p_i) = -\frac{1}{\alpha}\sum_i \frac{p_i^{\alpha+1}-p_i^{-\alpha+1}}{p_i^\alpha + p_i^{-\alpha}} \tag{128}$$

La fonction de base correspondante, convexe pour $-0.5 \leq \alpha \leq 0.5$ s'écrit :

$$f(x) = \frac{1}{\alpha}\frac{x^{\alpha+1}-x^{-\alpha+1}}{x^\alpha + x^{-\alpha}} \tag{129}$$

Ce n'est pas une fonction convexe standard, car, si $f(1)=0$, on a $f'(1)=1$.
La fonction convexe standard correspondante s'écrira donc :

$$f_c(x) = \frac{1}{\alpha}\frac{x^{\alpha+1}-x^{-\alpha+1}}{x^\alpha + x^{-\alpha}} - x + 1 \tag{130}$$

12.2.1 Divergence de Csiszär correspondante.

La divergence de Csiszär construite sur la fonction $f_c(x)$ précédente s'écrit :

$$C_\alpha(p\|q) = \frac{1}{\alpha}\sum_i \frac{p_i^{1+\alpha}q_i^{-\alpha} - p_i^{1-\alpha}q_i^\alpha}{p_i^\alpha q_i^{-\alpha} + p_i^{-\alpha}q_i^\alpha} - \sum_i p_i + \sum_i q_i \tag{131}$$

Le calcul de l'opposé du gradient par rapport à "q" conduit après quelques calculs simples mais fastidieux, à l'expression suivante $\forall j$:

$$-\frac{\partial C_\alpha(p\|q)}{\partial q_j} = \frac{p_j}{q_j}\left[\frac{4}{\left(p_j^\alpha q_j^{-\alpha} + p_j^{-\alpha}q_j^\alpha\right)^2}\right] - 1 \tag{132}$$

Comme on peut s'y attendre, ce gradient est nul si $p_j = q_j \quad \forall j$, de plus le cas correspondanr à la divergence de Kullback-Leibler se retrouve pour $\alpha = 0$.

12.2.2 Divergence de Csiszär duale correspondante.

Elle se déduit de l'expression (131) en effectuant la permutation $p \leftrightarrow q$; elle s'écrit :

$$C_\alpha(q\|p) = \frac{1}{\alpha}\sum_i \frac{q_i^{1+\alpha}p_i^{-\alpha} - q_i^{1-\alpha}p_i^{\alpha}}{q_i^{\alpha}p_i^{-\alpha} + q_i^{-\alpha}p_i^{\alpha}} - \sum_i q_i + \sum_i p_i \tag{133}$$

L'opposé du gradient s'écrit $\forall j$, après quelques calculs :

$$-\frac{\partial C_\alpha(q\|p)}{\partial q_j} = \frac{1}{\alpha}\left(\frac{p_j^{2a} - q_j^{2a}}{p_j^{2a} + q_j^{2a}}\right) + \left(\frac{p_j^{2a} - q_j^{2a}}{p_j^{2a} + q_j^{2a}}\right)^2 \tag{134}$$

13 Divergences invariantes.

Dans cette section, on développe les divergences invariantes par changement d'échelle sur "q".
En effet, les divergences possédant cette propriété *d'invariance* permettent, lors du processus de minimisation, de prendre en compte de façon simple une contrainte de somme connue sur les variables correspondantes.
On développera les calculs sur les divergences générales (101) pour les divergences de Csiszär et (108) pour les divergences de Bregman.
Les divergences spécifiques des diverses entropies analysées dans les paragraphes 4 à 9 seront vues comme des cas particuliers.
Cependant, les divergences fondées sur l'entropie de Tsallis constituent un cas spécifique qui sera détaillé à chaque étape, en effet, pour ces divergences il est possible de calculer un facteur d'invariance nominal dont l'influence sera examinée.
On rappelle d'abord quelques propriétés utiles des facteurs d'invariance qui ont été exposées longuement dans les références [11], [12].

13.1 Quelques rappels sur les facteurs d'invariance et leurs propriétés.

1- Pour une divergence $D(p\|q)$, le facteur d'invariance nominal $K_0(p,q)$ spécifque de la divergence considérée est obtenu comme solution de l'équation :

$$\sum_i \frac{\partial D(p_i\|Kq_i)}{\partial K} = 0 \tag{135}$$

si une solution explicite existe.

2- Les propriétés générales des facteurs d'invariance sont les suivantes :

* le facteur d'invariance $K(p,q)$ est un scalaire positif.
* Les composantes du vecteur $K(p,q).q$ sont invariantes si "q" est multiplié par un facteur positif constant.

* Tous les facteurs $K(p,q)$ possédant les propriétés ci dessus rendent invariante n'importe quelle divergence ; ils sont solution de l'équation différentielle :

$$K(p,q) + \sum_j q_j \frac{\partial K(p,q)}{\partial q_j} = 0 \tag{136}$$

Les facteurs d'invariance nominaux sont bien sur solutions de l'équation différentielle ci-dessus.
* L'expression générale des facteurs d'invariance s'écrit sous la forme :

$$K(p,q) = \left(\frac{\sum_i p_i^\alpha q_i^\beta}{\sum_i p_i^\delta q_i^\gamma} \right)^\mu \tag{137}$$

Avec :

$$\alpha + \beta = \gamma + \delta \tag{138}$$

Ce qui traduit le fait que $K(p,q)$ est un scalaire positif, et :

$$\mu(\gamma - \beta) = 1 \tag{139}$$

Ce qui traduit le fait que le vecteur $K(p,q).q$ est invariant par rapport à q.

3- Que le facteur d'invariance soit le facteur d'invariance nominal ou non, on a la propriété fondamentale suivante :

$$\sum_j q_j \frac{\partial D(p \| K(p,q).q)}{\partial q_j} \tag{140}$$

13.2 Divergences de Csiszär invariantes

On rappelle que la divergence de base a été écrite (101) :

$$C_{ab}(p\|q) = \frac{1}{a-b} \left[\sum_i p_i^a q_i^{1-a} - \sum_i p_i^b q_i^{1-b} \right] - \sum_i p_i + \sum_i q_i \tag{141}$$

En introduisant le facteur d'invariance dans cette expression, il vient :

$$C_{ab}(p\|Kq) = \frac{1}{a-b} \left[\sum_i K^{1-a} p_i^a q_i^{1-a} - \sum_i K^{1-b} p_i^b q_i^{1-b} \right] - \sum_i p_i + \sum_i K q_i \tag{142}$$

Le facteur d'invariance nominal "K_0" est obtenu comme solution (si elle existe) de l'équation :

$$\frac{\partial C_{ab}(p\|Kq)}{\partial K} = 0 \tag{143}$$

Cette équation s'écrit dans le cas présent :

$$\frac{1-a}{a-b} K^{-a} \sum_i p_i^a q_i^{1-a} - \frac{1-b}{a-b} K^{-b} \sum_i p_i^b q_i^{1-b} + \sum_i q_i = 0 \tag{144}$$

En toute généralité, il n'y a pas de solution explicite en "K" sauf dans deux cas particuliers :
* 1 - Cas de la divergence fondée sur l'entropie de Tsallis.
On a alors $a = t$ et $b = 1$, ce qui conduit à l'expression explicite du facteur d'invariance nominal :

$$K_0 = \left(\frac{\sum_i p_i^t q_i^{1-t}}{\sum_i q_i} \right)^{\frac{1}{t}} \tag{145}$$

* 2 - Cas des divergences fondées sur l'entropie de Shannon (divergence de Kullback-Leibler) qui s'obtient par exemple à partir de (145) en faisant simplement $t = 1$, ce qui conduit à :

$$K = \frac{\sum_i p_i}{\sum_i q_i} \tag{146}$$

En conséquence, pour toutes les divergences considérées, on utilisera en guise de facteur d'invariance l'expression (146) (facteur d'invariance nominal pour la divergence de Kullback-Leibler), il nous servira de référence.
Dans le cas de la divergence fondée sur l'entropie de Tsallis, nous développerons aussi la divergence invariante utilisant le facteur nominal (145).

13.2.1 Divergences de Csiszär invariantes - Cas général.

En utilisant le facteur d'invariance (146), la divergence (142) s'écrit tous calculs faits, avec les notations : $\bar{p}_i = \frac{p_i}{\sum_j p_j}$ et $\bar{q}_i = \frac{q_i}{\sum_j q_j}$:

$$CI_{ab}(p\|q) = \frac{\sum_j p_j}{a-b} \left[\sum_i \overline{p}_i^a \overline{q}_i^{1-a} - \sum_i \overline{p}_i^b \overline{q}_i^{1-b} \right] \tag{147}$$

L'opposé de son gradient par rapport à "q" s'écrit $\forall j$, après quelques calculs :

$$\begin{aligned} -\frac{\partial CI_{ab}(p\|q)}{\partial q_j} = \frac{\sum_j p_j}{\sum_j q_j} \Bigg[& \frac{a-1}{a-b} \left(\frac{\overline{p}_j}{\overline{q}_j} \right)^a + \frac{1-b}{a-b} \left(\frac{\overline{p}_j}{\overline{q}_j} \right)^b \\ & - \frac{a-1}{a-b} \sum_i \overline{q}_i \left(\frac{\overline{p}_i}{\overline{q}_i} \right)^a - \frac{1-b}{a-b} \sum_i \overline{q}_i \left(\frac{\overline{p}_i}{\overline{q}_i} \right)^b \Bigg] \end{aligned} \tag{148}$$

De cette forme générale, on déduit facilement les expressions correspondant aux divergences entropiques données dans les sections 4 à 9, en faisant les adaptations suivantes :

- Entropie de Shannon : $a \to 1$ et $b \to 1$ en respectant (95) ou (96)
- Entropie de Tsallis : $a = t, \quad b = 1$
- Entropie de Kaniadakis : $a = 1 + K, \quad b = 1 - K$

– Entropie de Abe : $a = z, \quad b = \frac{1}{z}$

– Entropie "γ" : $a = 2\gamma + 1, \quad b = 1 - \gamma$

– Entropie à 2 paramètres (KLS) : $a = 1 + r + K, \quad b = 1 + r - K$

13.2.2 Cas particulier : divergence de Csiszär/Tsallis invariante.

Pour cette divergence, le facteur d'invariance nominal est calculable de façon explicite ; son expression est donnée par (145).
La divergence invariante obtenue en utilisant ce facteur d'invariance spécifique s'écrit :

$$CI_T^{K_0}(p\|q) = \frac{t}{1-t}\left[\sum_i p_i - \left(\frac{\sum_i p_i^t q_i^{1-t}}{\sum_i q_i}\right)^{\frac{1}{t}} \sum_i q_i\right] \tag{149}$$

Ou, plus simplement :

$$CI_T^{K_0}(p\|q) = \frac{t}{1-t}\left[\sum_i p_i - K_0 \sum_i q_i\right] \tag{150}$$

L'opposé du gradient par rapport à "q" s'écrit $\forall j$:

$$-\frac{\partial CI_T^{K_0}(p\|q)}{\partial q_j} = \left(\frac{\sum_i p_i^t q_i^{1-t}}{\sum_i q_i}\right)^{\frac{1}{t}-1} p_j^t q_j^{-t} - \left(\frac{\sum_i p_i^t q_i^{1-t}}{\sum_i q_i}\right)^{\frac{1}{t}} \tag{151}$$

En introduisant les variables normalisées $\overline{p}_j = \frac{p_j}{\sum_i p_i}$ et $\overline{q}_j = \frac{q_j}{\sum_i q_i}$, ceci peut aussi s'écrire :

$$-\frac{\partial CI_T^{K_0}(p\|q)}{\partial q_j} = \frac{\sum_j p_j}{\sum_j q_j}\left[\left(\sum_i \overline{p}_i^t \overline{q}_i^{1-t}\right)^{\frac{1}{t}-1} \overline{p}_j^t \overline{q}_j^{-t} - \left(\sum_i \overline{p}_i^t \overline{q}_i^{1-t}\right)^{\frac{1}{t}}\right] \tag{152}$$

13.3 Divergences de Csiszär duales invariantes.

A partir de la divergence (105) qui a été écrite : :

$$C_{ab}(q\|p) = \frac{1}{a-b}\left[\sum_i q_i^a p_i^{1-a} - \sum_i q_i^b p_i^{1-b}\right] - \sum_i q_i + \sum_i p_i \tag{153}$$

En introduisant le facteur d'invariance "K", il vient :

$$C_{ab}(Kq\|p) = \frac{1}{a-b}\left[K^a \sum_i q_i^a p_i^{1-a} - K^b \sum_i q_i^b p_i^{1-b}\right] - K\sum_i q_i + \sum_i p_i \tag{154}$$

Le facteur d'invariance nominal est solution (si elle existe) de l'équation :

$$\frac{\partial C_{ab}(Kq\|p)}{\partial K} = 0 \tag{155}$$

Cette expression de cette dérivée partielle s'écrit :

$$\frac{\partial C_{ab}(Kq\|p)}{\partial K} = \frac{1}{a-b}\left[a\ K^{a-1}\sum_i q_i^a p_i^{1-a} - b\ K^{b-1}\sum_i q_i^b p_i^{1-b}\right] - \sum_i q_i \quad (156)$$

L'équation (155) n'a pas de solution explicite en "K", en revanche dans le cas particulier de l'entropie de Tsallis qui correspond à $a = t$ et $b = 1$, une solution explicite peut être obtenue ; on a alors le facteur d'invariance nominal :

$$K_0 = \left(\frac{\sum_j q_j}{\sum_j q_j^t p_j^{1-t}}\right)^{\frac{1}{t-1}} \quad (157)$$

Le facteur d'invariance nominal correspondant à la divergence de Kullback-Leibler duale (10) se déduit de cete expression en faisant le passage à la limite $t \to 1$.
Ceci conduit à l'expression difficilement exploitable :

$$\log K_0 = \frac{\sum_i q_i \log \frac{p_i}{q_i}}{\sum_i q_i} \quad (158)$$

En conséquence, on développera le cas général en utilisant le facteur d'invariance :

$$K = \frac{\sum_j p_j}{\sum_j q_j} \quad (159)$$

qui est le facteur d'invariance nominal pour la divergence de Kullback-Leibler et on traitera le cas particulier de l'entropie de Tsallis en utilisant son facteur d'invariance nominal donné par (157).

13.3.1 Divergences de Csiszär duales invariantes - Cas général.

A partir de (154), et en introduisant le facteur d'invariance (159), cette divergence s'écrit, avec les variables réduites $\overline{q}_i = \frac{q_i}{\sum_j q_j}$ et $\overline{p}_i = \frac{p_i}{\sum_j p_j}$:

$$CI_{ab}(q\|p) = \frac{\sum_j p_j}{a-b}\left(\sum_i \overline{q}_i^a \overline{p}_i^{1-a} - \sum_i \overline{q}_i^b \overline{p}_i^{1-b}\right) \quad (160)$$

L'opposé du gradient par rapport à "q" s'écrit $\forall j$, tous calculs faits :

$$-\frac{\partial CI_{ab}(q\|p)}{\partial q_j} = \frac{1}{a-b}\frac{\sum_j p_j}{\sum_j q_j}\left[a\ \sum_i \overline{q}_i\left(\frac{\overline{p}_i}{\overline{q}_i}\right)^{1-a} - a\ \left(\frac{\overline{p}_j}{\overline{q}_j}\right)^{1-a}\right.$$
$$\left. +b\ \sum_i \overline{q}_i\left(\frac{\overline{p}_i}{\overline{q}_i}\right)^{1-b} - b\ \left(\frac{\overline{p}_j}{\overline{q}_j}\right)^{1-b}\right] \quad (161)$$

13.3.2 Cas particulier : divergence de Csiszär/Tsallis duale invariante.

On utilise dans ce cas, le facteur d'invariance nominal donné par (157) ; à partir de (154), on obtient avec $a = t$ et $b = 1$:

$$CI_T^{K_0}(q\|p) = \sum_i p_i - \left(\frac{\sum_j q_j}{\sum_j q_j^t p_j^{1-t}}\right)^{\frac{1}{t-1}} \sum_i q_i \tag{162}$$

L'opposé de son gradient par rapport à "q" s'écrira $\forall j$:

$$-\frac{\partial CI_T^{K_0}(q\|p)}{\partial q_j} = \frac{t}{1-t}\left[\left(\frac{\sum_j q_j}{\sum_j q_j^t p_j^{1-t}}\right)^{\frac{t}{t-1}} \left(\frac{p_j}{q_j}\right)^{1-t} - \left(\frac{\sum_j q_j}{\sum_j q_j^t p_j^{1-t}}\right)^{\frac{1}{t-1}}\right] \tag{163}$$

13.4 Divergences de Bregman invariantes.

On rappelle que la divergence de Bregman fondée sur la forme générale de l'entropie a été écrite (108) :

$$\begin{aligned} B_{ab}(p\|q) = \frac{1}{a-b}\sum_i \big[&p_i^a - p_i^b + (a-1)q_i^a - (b-1)q_i^b \\ &-ap_iq_i^{a-1} + bp_iq_i^{b-1}\big] \end{aligned} \tag{164}$$

En introduisant le facteur d'invariance "K", il vient :

$$\begin{aligned} B_{ab}(p\|Kq) = \frac{1}{a-b}\sum_i \big[&p_i^a - p_i^b + (a-1)K^aq_i^a - (b-1)K^bq_i^b \\ &-aK^{a-1}p_iq_i^{a-1} + bK^{b-1}p_iq_i^{b-1}\big] \end{aligned} \tag{165}$$

Le facteur d'invariance nomimal s'obtient comme solution de l'équation :

$$\frac{\partial B_{ab}(p\|Kq)}{\partial K} = 0 \tag{166}$$

Avec :

$$\begin{aligned} \frac{\partial B_{ab}(p\|Kq)}{\partial K} = \frac{1}{a-b}\Bigg[&a(a-1)K^{a-1}\sum_i q_i^a - b(b-1)K^{b-1}\sum_i q_i^b \\ &-a(a-1)K^{a-2}\sum_i p_iq_i^{a-1} + b(b-1)K^{b-2}\sum_i p_iq_i^{b-1}\Bigg] \end{aligned} \tag{167}$$

Avec cette expression, l'équation (166) n'a pas de solution explicite en toute généralité.

Toutefois, dans le cas spécifique lié à l'entropie de Tsallis qui correspond à $a = t$ et $b = 1$ une solution explicite existe et le facteur d'invariance nominal s'écrit :

$$K_0 = \frac{\sum_i p_i q_i^{t-1}}{\sum_i q_i^t} \tag{168}$$

En conséquence, pour toutes les divergences considérées, on utilisera en guise de facteur d'invariance l'expression :

$$K = \frac{\sum_i p_i}{\sum_i q_i} \tag{169}$$

C'est le facteur d'invariance nominal pour la divergence de Kullback-Leibler fondée sur l'entropie de Shannon ; il nous servira de référence.
Dans le cas de la divergence fondée sur l'entropie de Tsallis, nous développerons aussi la divergence invariante utilisant le facteur nominal (168).

13.4.1 Divergences de Bregman invariantes - Cas général.

En introduisant le facteur d'invariance (169) dans l'expression (165), on obtient, tous calculs faits :

$$\begin{aligned} BI_{ab}(p\|q) =& \frac{(\sum_j p_j)^a}{a-b}\left[\sum_i \overline{p}_i^a + (a-1)\sum_i \overline{q}_i^a - a\sum_i \overline{p}_i\overline{q}_i^{a-1}\right] \\ &- \frac{(\sum_j p_j)^b}{a-b}\left[\sum_i \overline{p}_i^b + (b-1)\sum_i \overline{q}_i^b - b\sum_i \overline{p}_i\overline{q}_i^{b-1}\right] \end{aligned} \tag{170}$$

L'opposé de son gradient par rapport à "q" s'écrit $\forall j$:

$$\begin{aligned} -\frac{\partial BI_{ab}(p\|q)}{\partial q_j} =& \frac{a(a-1)}{(a-b)}\frac{(\sum_j p_j)^a}{\sum_j q_j}\left[\overline{p}_j\overline{q}_j^{a-2} - \sum_i \overline{p}_i\overline{q}_i^{a-1} - \overline{q}_j^{a-1} + \sum_i \overline{q}_i^a\right] \\ &- \frac{b(b-1)}{(a-b)}\frac{(\sum_j p_j)^b}{\sum_j q_j}\left[\overline{p}_j\overline{q}_j^{b-2} - \sum_i \overline{p}_i\overline{q}_i^{b-1} - \overline{q}_j^{b-1} + \sum_i \overline{q}_i^b\right] \end{aligned} \tag{171}$$

Comme c'était le cas pour les divergences de Csiszär, les expressions correspondant aus différentes divergences entropiques données dans les sections 4 à 9 se déduisent facilement de cette forme générale en faisant les adaptations suivantes :

- Entropie de Shannon : $a \to 1$ et $b \to 1$ en respectant (95) ou (96)
- Entropie de Tsallis : $a = t, \quad b = 1$
- Entropie de Kaniadakis : $a = 1 + K, \quad b = 1 - K$

– Entropie de Abe : $a = z, \quad b = \frac{1}{z}$

– Entropie "γ" : $a = 2\gamma + 1, \quad b = 1 - \gamma$

– Entropie à 2 paramètres (KLS) : $a = 1 + r + K, \quad b = 1 + r - K$

13.4.2 Cas particulier - Divergence de Bregman/Tsallis invariante.

Dans ce cas, le facteur d'invariance nominal peut être calculé explicitement ; son expression est donnée par (168) ; la divergence invariante correspondante s'exprime sous la forme :

$$BI_T^{K_0}(p\|q) = \frac{1}{1-t}\left[\left(\sum_i p_i q_i^{t-1}\right)^t \left(\sum_i q_i^t\right)^{1-t} - \sum_i p_i^t\right] \tag{172}$$

L'opposé de son gradient par rapport à "q" est donné $\forall j$ par :

$$-\frac{\partial BI_T^{K_0}(p\|q)}{\partial q_j} = t\left[\left(\frac{\sum_i p_i q_i^{t-1}}{\sum_i q_i^t}\right)^{t-1} p_j q_j^{t-2} - \left(\frac{\sum_i p_i q_i^{t-1}}{\sum_i q_i^t}\right)^t q_j^{t-1}\right] \tag{173}$$

Ou bien, sous forme simplifiée :

$$-\frac{\partial BI_T^{K_0}(p\|q)}{\partial q_j} = t\left[K_0^{t-1} p_j q_j^{t-2} - K_0^t q_j^{t-1}\right] \tag{174}$$

Ou encore :

$$-\frac{\partial BI_T^{K_0}(p\|q)}{\partial q_j} = t\ K_0^{t-1} q_j^{t-1}\left[\frac{p_j}{q_j} - K_0\right] \tag{175}$$

13.5 Divergences de Bregman duales - Forme invariante.

On reprend la forme de cette divergence écrite en (111) :

$$\begin{aligned} B_{ab}(q\|p) = \frac{1}{a-b}\sum_i \big[&q_i^a - q_i^b + (a-1)p_i^a - (b-1)p_i^b \\ &-aq_i p_i^{a-1} + bq_i p_i^{b-1}\big] \end{aligned} \tag{176}$$

En introduisant le facteur d'invariance "K", il vient :

$$\begin{aligned} B_{ab}(Kq\|p) = \frac{1}{a-b}\sum_i \big[&K^a q_i^a - K^b q_i^b + (a-1)p_i^a - (b-1)p_i^b \\ &-aKq_i p_i^{a-1} + bKq_i p_i^{b-1}\big] \end{aligned} \tag{177}$$

Le facteur d'invariance nomimal s'obtient comme solution de l'équation :

$$\frac{\partial B_{ab}(Kq\|p)}{\partial K} = 0 \tag{178}$$

Avec :

$$\frac{\partial B_{ab}(Kq\|p)}{\partial K} = \frac{1}{a-b}\left[aK^{a-1}\sum_i q_i^a - bK^{b-1}\sum_i q_i^b \right.$$
$$\left. -a\sum_i q_i p_i^{a-1} + b\sum_i q_i p_i^{b-1}\right] \tag{179}$$

Avec cette expression, l'équation (178) n'a pas de solution explicite en toute généralité.
Toutefois, dans le cas spécifique lié à l'entropie de Tsallis qui correspond à $a = t$ et $b = 1$ une solution explicite existe et le facteur d'invariance nominal s'écrit :

$$K_0 = \left(\frac{\sum_i q_i^t}{\sum_i q_i p_i^{t-1}}\right)^{\frac{1}{1-t}} \tag{180}$$

En conséquence, pour toutes les divergences considérées, on utilisera en guise de facteur d'invariance l'expression :

$$K = \frac{\sum_i p_i}{\sum_i q_i} \tag{181}$$

C'est le facteur d'invariance nominal pour la divergence de Kullback-Leibler fondée sur l'entropie de Shannon ; il nous servira de référence.
Dans le cas de la divergence fondée sur l'entropie de Tsallis, nous développerons aussi la divergence invariante utilisant le facteur nominal (180).

13.5.1 Divergences de Bregman duales invariantes - Cas général.

En introduisant le facteur d'invariance (181) dans la divergence (177), on obtient tous calculs faits :

$$\begin{aligned} BI_{ab}(q\|p) =& \frac{(\sum_j p_j)^a}{a-b}\left[\sum_i \overline{q}_i^a + (a-1)\sum_i \overline{p}_i^a - a\sum_i \overline{q}_i\overline{p}_i^{(a-1)}\right] \\ &- \frac{(\sum_j p_j)^b}{a-b}\left[\sum_i \overline{q}_i^b + (b-1)\sum_i \overline{p}_i^b - b\sum_i \overline{q}_i\overline{p}_i^{(b-1)}\right] \end{aligned} \tag{182}$$

L'opposé de son gradient par rapport à "q" s'écrit $\forall j$, après quelques calculs :

$$\begin{aligned} -\frac{\partial BI_{ab}(q\|p)}{\partial q_j} =& \frac{b\ (\sum_j p_j)^b}{(a-b)\sum_j q_j}\left(\overline{q}_j^{(b-1)} - \sum_i \overline{q}_i^b - \overline{p}_j^{(b-1)} + \sum_i \overline{q}_i\overline{p}_i^{(b-1)}\right) \\ &- \frac{a\ (\sum_j p_j)^a}{(a-b)\sum_j q_j}\left(\overline{q}_j^{(a-1)} - \sum_i \overline{q}_i^a - \overline{p}_j^{(a-1)} + \sum_i \overline{q}_i\overline{p}_i^{(a-1)}\right) \end{aligned} \tag{183}$$

Les expressions correspondant aux différentes entropies sont obtenues, comme il a été indiqué dans les sections précédentes, en affectant à "a" et "b", les valeurs des paramètres propres à chaque cas.

13.5.2 Cas particulier de la divergence de Bregman/Tsallis duale invariante.

Dans ce cas, le facteur d'invariance nominal se calcule explicitement ; il est donné par la relation (180).
La divergence invariante correspondante est donnée par :

$$BI_T^{K_0}(q\|p) = \sum_i p_i^t - \left(\frac{\sum_i q_i^t}{\sum_i q_i p_i^{t-1}}\right)^{\frac{t}{1-t}} \sum_i q_i^t \tag{184}$$

Ce qui peut s'écrire plus simplement :

$$BI_T^{K_0}(q\|p) = \sum_i p_i^t - K_0^t \sum_i q_i^t \tag{185}$$

L'opposé du gradient par rapport à "q" s'écrit $\forall j$:

$$-\frac{\partial BI_T^{K_0}(q\|p)}{\partial q_j} = \frac{t}{t-1}\left[\left(\frac{\sum_i q_i^t}{\sum_i q_i p_i^{1-t}}\right)^{\frac{1}{1-t}} p_j^{t-1} - \left(\frac{\sum_i q_i^t}{\sum_i q_i p_i^{1-t}}\right)^{\frac{t}{1-t}} q_j^{t-1}\right] \tag{186}$$

Ou bien, sous une autre forme :

$$-\frac{\partial BI_T^{K_0}(q\|p)}{\partial q_j} = \frac{t}{t-1} K_0 \left[p_j^{t-1} - K_0^{t-1} q_j^{t-1}\right] \tag{187}$$

14 Algorithmique

14.1 Généralités sur la méthode de construction des algorithmes.

On rappelle que dans le cas qui nous intéresse, on a $p = y$ et $q = Hx$.
Alors, pour une divergence $D(p\|q)$ ou $D(q\|p)$, on a :

$$\frac{\partial D}{\partial x} = H^T \frac{\partial D}{\partial q} \tag{188}$$

Ceci justifie le fait que tout le long de ce travail, on ait développé les expressions des gradients par rapport à "q".
Les algorithmes proposés ici s'appuient sur la méthode SGM ou ses variantes exposées en détails dans [11] et [12].
Si la divergence considérée est de forme classique non invariante, les algorithmes proposés prennent en compte la contrainte de non-négativité de la solution.
En revanche, si on veut en plus prendre en compte une contrainte de somme sur les composantes de la solution, les divergences invariantes doivent être utilisées.
Quoiqu'il en soit, la forme générale des algorithmes reste la même.
L'algorithme itératif de base s'écrit sous forme générale :

$$x_l^{k+1} = x_l^k + \alpha_l^k x_l^k \left(-\frac{\partial D}{\partial x}\right)_l^k \tag{189}$$

La divergence D étant convèxe par rapport à "x", $\left(-\frac{\partial D}{\partial x}\right)$ est une direction de descente.
L'opposé du gradient peut toujours s'écrire sous la forme :

$$\left(-\frac{\partial D}{\partial x}\right)_l^k = U_l^k - V_l^k \quad ; \quad U_l^k > 0 \quad ; \quad V_l^k > 0 \tag{190}$$

D'ou :

$$x_l^{k+1} = x_l^k + \alpha_l^k x_l^k \left(U_l^k - V_l^k\right) \tag{191}$$

Puis, en préconditionnant par $\frac{1}{V_l^k} > 0$:

$$x_l^{k+1} = x_l^k + \alpha_l^k x_l^k \left(\frac{U_l^k}{V_l^k} - 1\right) \tag{192}$$

Compte tenu de $V_l^k > 0$, le gradient modifié $\frac{U^k}{V^k} - 1$ reste une direction de descente.

Observation concernant les notations : dans cette écriture, $\frac{U^k}{V^k}$ est un vecteur obtenu en faisant le rapport composante à composante des vecteurs U^k et V^k, de même que "1" est le vecteur unité.

En ce qui concerne le pas de descente, pour chacun de ces deux algorithmes, on adopte la procédure suivante :
* 1- A une itération donnée "k", on calcule le pas maximum $(\alpha^k)_{Max}$ assurant la non négativité de l'ensemble des composantes de x^{k+1}.
* 2 - Le pas de descente α^k (valable pour toutes les composantes), assurant la convergence de l'algorithme est alors calculé par une méthode de recherche unidimensionnelle de type Armijo (par exemple), sur l'intervalle $\left[0, (\alpha^k)_{Max}\right]$.
Cette procédure est analysée en détails dans [11] et [12].
On pourra écrire de façon générale, avec un pas de descente indépendant de la composante :

$$x^{k+1} = x^k + \alpha^k x^k \left(U^k - V^k\right) \tag{193}$$

Observation concernant les notations : dans cette expression, l'opération $x^k \left(U^k - V^k\right)$ représente le produit composante à composante des vecteurs x^k et $\left(U^k - V^k\right)$ (produit de Hadamard).

Avec un gradient modifié (préconditionné), on aura :

$$x^{k+1} = x^k + \alpha^k x^k \left(\frac{U^k}{V^k} - 1\right) \tag{194}$$

Dans ce dernier cas, si on utilise un pas de descente $\alpha^k = 1$, $\forall k$, la non-négativité est assurée, et on obtient un algorithme purement multiplicatif qui s'écrit :

$$x^{k+1} = x^k \left(\frac{U^k}{V^k}\right) \tag{195}$$

Bien entendu, en toute généralité, rien ne prouve la convergence des algorithmes purement multiplicatifs, chaque divergence implique une étude particulière.

Commentaire : Les observations concernant les notations sont valables pour tous les algorithmes proposés dans les sections suivantes, plus précisément , les opérations entre vecteurs (produits ou rapports), sont des opérations composante à composante ; le résultat est toujours un vecteur.

Remarques :
* 1 - Pour les divergences non invariantes, les algorithmes (193) (194) et (195) assurent seulement la non négativité de la solution.
* 2 - Si on exige en plus que la contrainte de somme soit assurée, les divergences invariantes prennent tout leur sens, en effet, un algorithme de type (193) appliqué sur de telles divergences permet d'assurer la propriété :

$$\sum_l x_l^{k+1} = \sum_l x_l^k \tag{196}$$

Partant d'un estimé initial x^0 tel que $\sum_l x_l^0 = C$, tout les estimés successifs seront de même somme.
En revanche, l'utilisation d'un algorithme préconditionné (194) ou d'un algorithme purement multiplicatif de type (195) (pour autant que sa convergence soit assurée), ne permet pas d'assurer spontanément la contrainte de somme ; une étape supplémentaire s'impose.
A chaque itération, on effectue l'opération en 2 étapes :

* - On calcule d'abord un estimé provisoire :

$$\widetilde{x}^{k+1} = x^k \left(\frac{U^k}{V^k} \right) \tag{197}$$

* - Puis dans une étape de normalisation, on calcule :

$$x^{k+1} = \frac{\widetilde{x}^{k+1}}{\sum_l \widetilde{x}_l^{k+1}} C \tag{198}$$

Compte tenu des propriétés des divergences invariantes, cette dernière opération ne modifie pas la valeur de la divergence considérée.

14.2 Algorithmes fondés sur l'entropie générale.

On détaille dans cette section les algorithmes fondés sur les formes de divergences liées à la forme générale de l'entropie.
On rappelle que les domaines de valeurs des paramètres "a" et "b" intervenant dans ces divergences sont donnés par (95) (96) :

$$0 \leq a \leq 1 \leq b \tag{199}$$

ou bien :

$$0 \leq b \leq 1 \leq a \tag{200}$$

On n'explicitera pas ici les algorithmes concernant les divergences duales fondées sur l'entropie générale, qu'elles soient invariantes ou non invariantes, mais ceci ne présente aucune difficulté particulière hormis les problèmes spécifiques déjà précisés liés aux expressions des gradients.

14.2.1 Divergence de Csiszär construite sur l'entropie générale.

L'expression de cette divergence est donnée par (101) ; compte tenu de l'expression de l'opposé du gradient donnée en (102), l'algorithme (193) s'écrit ici :

$$x^{k+1} = x^k + \alpha^k x^k H^T \left\{ \underbrace{\left[\frac{a-1}{a-b} \left(\frac{p}{q^k} \right)^a + \frac{1-b}{a-b} \left(\frac{p}{q^k} \right)^b \right]}_{A^k} - \underbrace{1}_{B^k} \right\} \tag{201}$$

La forme multiplicative correspondante s'en déduit immédiatement par (195), avec :

$$U^k = H^T A^k \quad ; \quad V^k = H^T B^k \tag{202}$$

14.2.2 Divergence de Bregman construite sur l'entropie générale.

L'expression de cette divergence est donnée par (108) ; compte tenu de l'expression de l'opposé du gradient donnée en (109), l'algorithme (193) s'écrit ici :

$$x^{k+1} = x^k + \alpha^k x^k H^T \left\{ \left(\frac{p}{q^k} - 1 \right) \left[\frac{(a^2-a)}{a-b} (q^k)^{a-1} + \frac{(b-b^2)}{a-b} (q^k)^{b-1} \right] \right\} \tag{203}$$

La forme multiplicative correspondante s'en déduit immédiatement par (195), avec :

$$U^k = H^T \left\{ \left(\frac{p}{q^k} \right) \left[\frac{(a^2-a)}{a-b} (q^k)^{a-1} + \frac{(b-b^2)}{a-b} (q^k)^{b-1} \right] \right\} \tag{204}$$

et

$$V^k = H^T \left[\frac{(a^2-a)}{a-b} (q^k)^{a-1} + \frac{(b-b^2)}{a-b} (q^k)^{b-1} \right] \tag{205}$$

14.2.3 Divergence de Csiszär invariante construite sur l'entropie générale.

L'expression de cette divergence est donnée par (147), et l'opposé de son gradient est donnée par (148) ; l'algorithme (193) s'écrit ici :

$$x^{k+1} = x^k + \alpha^k x^k H^T \frac{\sum_j p_j}{\sum_j q_j^k} \left[\frac{a-1}{a-b} \left(\frac{\overline{p}}{\overline{q}^k} \right)^a + \frac{1-b}{a-b} \left(\frac{\overline{p}}{\overline{q}^k} \right)^b \right.$$
$$\left. - \frac{a-1}{a-b} \sum_i \overline{q}_i^k \left(\frac{\overline{p}_i}{\overline{q}_i^k} \right)^a - \frac{1-b}{a-b} \sum_i \overline{q}_i^k \left(\frac{\overline{p}_i}{\overline{q}_i^k} \right)^b \right] \tag{206}$$

La décomposition permettant d'écrire un algorithme multiplicatif s'écrit :

$$U^k = H^T \frac{\sum_j p_j}{\sum_j q_j^k} \left[\frac{a-1}{a-b} \left(\frac{\overline{p}}{\overline{q}^k} \right)^a + \frac{1-b}{a-b} \left(\frac{\overline{p}}{\overline{q}^k} \right)^b \right] \tag{207}$$

et

$$V^k = H^T \frac{\sum_j p_j}{\sum_j q_j^k} \left[\frac{a-1}{a-b} \sum_i \overline{q}_i^k \left(\frac{\overline{p}_i}{\overline{q}_i^k} \right)^a + \frac{1-b}{a-b} \sum_i \overline{q}_i^k \left(\frac{\overline{p}_i}{\overline{q}_i^k} \right)^b \right] \tag{208}$$

*** Cas particulier de la divergence de Csiszär/Tsallis invariante.**
Compte tenu du fait que pour ce type d'entropie, le facteur d'invariance nominal peut se calculer explicitement et conduit à un algorithme simplifié, on détaille ici, ce cas spécifique.
On rappelle que l'expression du facteur d'invariance nominal est donné par la relation (145).
Ceci conduit à la divergence invariante (149) dont l'opposé du gradient est donné par (152).
L'algorithme (193) s'écrit en conséquence :

$$x^{k+1} = x^k + \alpha^k x^k H^T \frac{\sum_j p_j}{\sum_j q_j^k} \left(\sum_i \overline{p}_i^t (\overline{q}_i^k)^{1-t} \right)^{\frac{1}{t}-1} \left[\left(\frac{\overline{p}}{\overline{q}^k} \right)^t - \sum_i \overline{q}_i^k \left(\frac{\overline{p}_i}{\overline{q}_i^k} \right)^t \right] \tag{209}$$

La décomposition permettant d'écrire un algorithme multiplicatif est immédiate.
On explicite ci dessous, l'algorithme analogue déduit de (206) en faisant $a = t$ et $b = 1$.
Cet algorithme correspond à un facteur d'invariance non-nomimal pour l'entropie de Tsallis, il s'écrit :

$$x^{k+1} = x^k + \alpha^k x^k H^T \frac{\sum_j p_j}{\sum_j q_j^k} \left[\left(\frac{\overline{p}}{\overline{q}^k} \right)^t - \sum_i \overline{q}_i^k \left(\frac{\overline{p}_i}{\overline{q}_i^k} \right)^t \right] \tag{210}$$

14.2.4 Divergence de Bregman invariante construite sur l'entropie générale.

L'expression de cette divergence est donnée par (170), et l'opposé de son gradient est donnée par (171) ; l'algorithme (193) s'écrit ici :

$$x^{k+1} = x^k + \alpha^k x^k H^T \left[\frac{a(a-1)}{(a-b)} \frac{(\sum_j p_j)^a}{\sum_j q_j^k} \left(\overline{p}(\overline{q}^k)^{a-2} - \sum_i \overline{p}_i (\overline{q}_i^k)^{a-1} - (\overline{q}^k)^{a-1} + \sum_i (\overline{q}_i^k)^a \right) \right.$$
$$\left. - \frac{b(b-1)}{(a-b)} \frac{(\sum_j p_j)^b}{\sum_j q_j^k} \left(\overline{p}(\overline{q}^k)^{b-2} - \sum_i \overline{p}_i (\overline{q}_i^k)^{b-1} - (\overline{q}^k)^{b-1} + \sum_i (\overline{q}_i^k)^b \right) \right] \tag{211}$$

La décomposition permettant d'écrire un algorithme multiplicatif s'écrit :

$$U^k = H^T \left[\frac{a(a-1)}{(a-b)} \frac{(\sum_j p_j)^a}{\sum_j q_j^k} \left(\overline{p}(\overline{q}^k)^{a-2} + \sum_i (\overline{q}_i^k)^a \right) \right.$$
$$\left. + \frac{b(1-b)}{(a-b)} \frac{(\sum_j p_j)^b}{\sum_j q_j^k} \left(\overline{p}(\overline{q}^k)^{b-2} + \sum_i (\overline{q}_i^k)^b \right) \right] \tag{212}$$

$$V^k = H^T \left[\frac{a(a-1)}{(a-b)} \frac{(\sum_j p_j)^a}{\sum_j q_j^k} \left(\sum_i \overline{p}_i (\overline{q}_i^k)^{a-1} + (\overline{q}^k)^{a-1} \right) \right.$$
$$\left. + \frac{b(1-b)}{(a-b)} \frac{(\sum_j p_j)^b}{\sum_j q_j^k} \left(\sum_i \overline{p}_i (\overline{q}_i^k)^{b-1} + (\overline{q}^k)^{b-1} \right) \right] \tag{213}$$

*** Cas particulier de la divergence de Bregman/Tsallis invariante.** Compte tenu du fait que pour ce type d'entropie, le facteur d'invariance nominal peut se calculer explicitement et conduit à un algorithme simplifié, on détaille ici, ce cas spécifique.
On rappelle que l'expression du facteur d'invariance nominal est donné par la relation (168).
Ceci conduit à la divergence invariante (172) dont l'opposé du gradient est donné par (173).
L'algorithme (193) s'écrit en conséquence :

$$x^{k+1} = x^k + \alpha^k x^k H^T \; t \; \left[\left(\frac{\sum_i p_i (q_i^k)^{t-1}}{\sum_i (q_i^k)^t} \right)^{t-1} p \; (q^k)^{t-2} \right.$$
$$\left. - \left(\frac{\sum_i p_i (q_i^k)^{t-1}}{\sum_i (q_i^k)^t} \right)^{t} (q^k)^{t-1} \right] \tag{214}$$

On notera que le terme $(p \; q^{t-2})$ est un vecteur obtenu en faisant le produit composante à composante des vecteurs p et q^{t-2}.

A titre de comparaison, on explicite ci après, l'algorithme analogue déduit de l'expression (211) en faisant $a = t$ et $b = 1$ et qui correspond à un facteur d'invariance non-nominal pour l'entropie de Tsallis.

$$x^{k+1} = x^k + \alpha^k x^k H^T \; t \; \frac{\sum_j (p_j)^t}{\sum_j q_j^k} \left[\overline{p}(\overline{q}^k)^{t-2} - \sum_i \overline{p}_i (\overline{q}_i^k)^{t-1} \right.$$
$$\left. -(\overline{q}^k)^{t-1} + \sum_i (\overline{q}_i^k)^t \right] \tag{215}$$

15 Annexe

Nous résumons ici les expressions de l'opposé des gradients $\frac{\partial D}{\partial q_j}$ des différentes divergences examinées dans les sections précédentes, dans le but de mettre en évidence des formes synthétiques spécifiques de ces expressions.

15.1 Divergences de Csiszär.

Ces expressions peuvent s'écrire $\forall j$ sous la forme synthétique :

$$\frac{p_j}{q_j} X_j - 1 \tag{216}$$

Avec pour X_j, les expressions suivantes :

- Entropie de Shannon : $X_j = 1$
- Entropie de Tsallis : $X_j = \left(\frac{p_j}{q_j}\right)^{t-1}$
- Entropie de Kaniadakis : $X_j = \frac{1}{2}\left(\frac{p_j}{q_j}\right)^{K} + \frac{1}{2}\left(\frac{p_j}{q_j}\right)^{-K}$
- Entropie de Abe : $X_j = \frac{z}{z+1}\left(\frac{p_j}{q_j}\right)^{z-1} + \frac{1}{z+1}\left(\frac{p_j}{q_j}\right)^{\frac{1}{z}-1}$
- Entropie "γ" : $X_j = \frac{2}{3}\left(\frac{p_j}{q_j}\right)^{2\gamma} + \frac{1}{3}\left(\frac{p_j}{q_j}\right)^{-\gamma}$
- Entropie à 2 paramètres (KLS) : $X_j = \frac{K+r}{2K}\left(\frac{p_j}{q_j}\right)^{r+K} + \frac{K-r}{2K}\left(\frac{p_j}{q_j}\right)^{r-K}$
- Entropie générale : $X_j = \frac{a-1}{a-b}\left(\frac{p_j}{q_j}\right)^{a-1} + \frac{1-b}{a-b}\left(\frac{p_j}{q_j}\right)^{b-1}$
- Entropie de Newton : $X_j = \frac{1}{2}\left(\frac{p_j}{q_j}\right) + \frac{1}{2}$

15.2 Divergences de Csiszär duales.

Dans le but d'avoir une écriture synthétique pour les expressions de l'opposé des gradients, on effectue d'abord les changements d'écriture suivants (qui

peuvent sembler un peu tortueux) :
Pour l'entropie de Shannon :

$$\log \frac{p_j}{q_j} = \left[\log \frac{p_j}{q_j} - 1\right] + 1 = \frac{p_j}{q_j}\left[\left(\frac{p_j}{q_j}\right)^{-1} \log \frac{p_j}{q_j} - \left(\frac{p_j}{q_j}\right)^{-1}\right] + 1$$

Pour lntropie de Tsallis :

$$\left[\frac{t}{1-t}\left(\frac{p_j}{q_j}\right)^{1-t} - \frac{1}{1-t}\right] + 1 = \frac{p_j}{q_j}\left[\frac{t}{1-t}\left(\frac{p_j}{q_j}\right)^{-t} - \frac{1}{1-t}\left(\frac{p_j}{q_j}\right)^{-1}\right] + 1$$

Alors les expressions de l'opposé des gradients peuvent s'écrire $\forall j$ sous la forme synthétique :

$$\frac{p_j}{q_j}\, T_j + 1 \tag{217}$$

Avec pour T_j, les expressions suivantes :

- Entropie de Shannon : $T_j = \left(\frac{p_j}{q_j}\right)^{-1} \log \frac{p_j}{q_j} - \left(\frac{p_j}{q_j}\right)^{-1}$
- Entropie de Tsallis : $T_j = \frac{t}{1-t}\left(\frac{p_j}{q_j}\right)^{-t} - \frac{1}{1-t}\left(\frac{p_j}{q_j}\right)^{-1}$
- Entropie de Kaniadakis : $T_j = \frac{1-K}{2K}\left(\frac{p_j}{q_j}\right)^{K-1} - \frac{1+K}{2K}\left(\frac{p_j}{q_j}\right)^{-K-1}$
- Entropie de Abe : $T_j = \frac{z^2}{1-z^2}\left(\frac{p_j}{q_j}\right)^{-z} - \frac{1}{1-z^2}\left(\frac{p_j}{q_j}\right)^{-\frac{1}{z}}$
- Entropie "γ" : $T_j = \frac{1-\gamma}{3\gamma}\left(\frac{p_j}{q_j}\right)^{\gamma-1} - \frac{1+2\gamma}{3\gamma}\left(\frac{p_j}{q_j}\right)^{-2\gamma-1}$
- Entropie à 2 paramètres : $T_j = \frac{1+r-K}{2K}\left(\frac{p_j}{q_j}\right)^{K-r-1} - \frac{1+r+K}{2K}\left(\frac{p_j}{q_j}\right)^{-K-r-1}$
- Entropie générale : $T_j = \frac{b}{a-b}\left(\frac{p_j}{q_j}\right)^{-b} - \frac{a}{a-b}\left(\frac{p_j}{q_j}\right)^{-a}$
- Entropie de Newton : $T_j = \frac{1}{2}\left(\frac{p_j}{q_j}\right)^{-1} \log\left(\frac{p_j}{q_j}\right) - 1$

15.3 Divergences de Bregman.

Ces expressions peuvent s'écrire $\forall j$ sous la forme synthétique :

$$\left[\frac{p_j}{q_j} - 1\right]\, Z_j \tag{218}$$

Avec pour Z_j, les expressions suivantes :

- Entropie de Shannon : $Z_j = 1$

- Entropie de Tsallis : $Z_j = t\ q_j^{t-1}$
- Entropie de Kaniadakis : $Z_j = \frac{1+K}{2} q_j^K + \frac{1-K}{2} q_j^{-K}$
- Entropie de Abe : $Z_j = \frac{z^2}{z+1} q_j^{z-1} + \frac{1}{z(z+1)} q_j^{\frac{1}{z}-1}$
- Entropie "γ" : $Z_j = \frac{2(2\gamma+1)}{3} q_j^{2\gamma} + \frac{1-\gamma}{3} q_j^{-\gamma}$
- Entropie à 2 paramètres : $Z_j = \frac{(r+K)(1+r+K)}{2K} q_j^{r+K} + \frac{(r-K)(1+r-K)}{2K} q_j^{r-K}$
- Entropie générale : $Z_j = \frac{a^2-a}{a-b} q_j^{a-1} + \frac{b-b^2}{a-b} q_j^{b-1}$
- Entropie de Newton : $Z_j = \frac{1}{2} + q_j$

15.4 Divergences de Bregman duales.

Une forme synthétique générale n'est pas simple à écrire, même si on observe une analogie certaine entre ces expressions qui peuvent s'écrire $\forall j$:

- Entropie de Shannon : $\log \frac{p_j}{q_j} = \log p_j - \log q_j$
- Entropie de Tsallis : $\frac{t}{1-t} \left[p_j^{t-1} - q_j^{t-1}\right]$
- Entropie de Kaniadakis : $\frac{1+K}{2K}(p_j^K - q_j^K) - \frac{1-K}{2K}(p_j^{-K} - q_j^{-K})$
- Entropie de Abe : $\frac{z^2}{z^2-1}(p_j^{z-1} - q_j^{z-1}) - \frac{1}{z^2-1}(p_j^{\frac{1}{z}-1} - q_j^{\frac{1}{z}-1})$
- Entropie "γ" : $\frac{2\gamma+1}{3\gamma}\left(p_j^{2\gamma} - q_j^{2\gamma}\right) - \frac{1-\gamma}{3\gamma}\left(p_j^{-\gamma} - q_j^{-\gamma}\right)$
- Entropie à 2 paramètres : $\frac{1+r+K}{2K}\left(p_j^{r+K} - q_j^{r+K}\right) - \frac{1+r-K}{2K}\left(p_j^{r-K} - q_j^{r-K}\right)$
- Entropie générale : $\frac{a}{a-b}\left(p_j^{a-1} - q_j^{a-1}\right) - \frac{b}{a-b}\left(p_j^{b-1} - q_j^{b-1}\right)$
- Entropie de Newton : $(p_j - q_j) + \frac{1}{2}\log\frac{p_j}{q_j} = \left(p_j + \frac{1}{2}\log p_j\right) - \left(q_j + \frac{1}{2}\log q_j\right)$

Références

[1] S. Abe. A note on the q-deformation-theoretic aspect of the generalized entropies in nonextensive physics. *Physics Letters A*, 224(6) :326–330, 1997.

[2] S. Asgarani and B. Mirza. Two-parameter entropies, s(k, r), and their dualities. *Physica A : Statistical Mechanics and its Applications*, 417 :185–192, 2015.

[3] E.P. Borges and I. Roditi. A family of nonextensive entropies. *Physics Letters A*, 246(5) :399–402, 1998.

[4] R. Chakrabarti and R. Jagannathan. A (p, q)-oscillator realization of two-parameter quantum algebras. *Journal of Physics A : Mathematical and General*, 24(13) :L711, 1991.

[5] S. Furuichi. An axiomatic characterization of a two-parameter extended relative entropy. *Journal of mathematical physics*, 51(12) :123302, 2010.

[6] F.H. Jackson. q-form of taylor's theorem. *Mess. Math*, 3 :57, 1909.

[7] F.H. Jackson. On q—definite integrals, j. *Pure Appl. Math*, 41 :193–403, 1910.

[8] G. Kaniadakis. Non-linear kinetics underlying generalized statistics. *Physica A : Statistical mechanics and its applications*, 296(3-4) :405–425, 2001.

[9] G. Kaniadakis. Statistical mechanics in the context of special relativity. *Physical review E*, 66(5) :056125, 2002.

[10] G. Kaniadakis, M. Lissia, and A.M. Scarfone. Two-parameter deformations of logarithm, exponential, and entropy : A consistent framework for generalized statistical mechanics. *Physical Review E*, 71(4) :046128, 2005.

[11] H. Lantéri. Divergences. divergences invariantes. applications aux problèmes inverses linéaires, nmf et déconvolution aveugle. *HAL.Archives ouvertes*, 2019.

[12] H. Lantéri. Divergences. scale invariant divergences. applications to linear inverse problems. nmf blind deconvolution. *arXiv preprint arXiv :2003.01411*, 2020.

[13] D.S. McAnally. q-exponential and q-gamma functions. i. q-exponential functionsa. *Journal of Mathematical Physics*, 36(1) :546–573, 1995.

[14] D.P. Mittal. On some functional equations concerning entropy, directed divergence and inaccuracy. *Metrika*, 22(1) :35–45, 1975.

[15] J. Naudts. Deformed exponentials and logarithms in generalized thermostatistics. *Physica A : Statistical Mechanics and its Applications*, 316(1-4) :323–334, 2002.

[16] N.J. Newton. An infinite-dimensional statistical manifold modelled on hilbert space. *Journal of Functional Analysis*, 263(6) :1661–1681, 2012.

[17] B.D. Sharma and I.J. Taneja. Entropy of type (α, β) and other generalized measures in information theory. *Metrika*, 22(1) :205–215, 1975.

[18] B. Trivellato. Deformed exponentials and applications to finance. *Entropy*, 15(9) :3471–3489, 2013.

[19] C. Tsallis. Possible generalization of boltzmann-gibbs statistics. *Journal of statistical physics*, 52(1) :479–487, 1988.

[20] C. Tsallis. Beyond boltzmann–gibbs–shannon in physics and elsewhere. *Entropy*, 21(7) :696, 2019.

[21] T Wada and AM Scarfone. Finite difference and averaging operators in generalized entropies. In *Journal of Physics : Conference Series*, volume 201, page 012005. IOP Publishing, 2010.