



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.. 2022. hal-03355082v3

HAL Id: hal-03355082

<https://hal.science/hal-03355082v3>

Preprint submitted on 29 Sep 2022

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.
 - Version 3. -

Henri Lantéri

19 septembre 2022

Table des matières

1	Avertissement concernant les Versions 2 et 3	4
2	Introduction et sommaire	4
3	Rappels (très brefs) sur les problèmes inverses.	6
4	Rappels (très brefs) sur l'entropie.	7
5	Entropie de Shannon - Logarithme naturel.	8
5.1	Divergence de Csiszär associée à l'entropie de Shannon - Kullback-Leibler.	9
5.2	Divergence de Csiszär/Shannon duale.	9
5.3	Divergences de Bregman/Shannon.	10
5.4	Divergences de Jensen/Shannon.	10
6	Entropie de Tsallis [19].	11
6.1	Divergence de Csiszär fondée sur l'entropie de Tsallis.	12
6.2	Divergence de Csiszär duale fondée sur l'entropie de Tsallis.	12
6.3	Divergence de Bregman fondée sur l'entropie de Tsallis.	13
6.4	Divergence de Bregman duale fondée sur l'entropie de Tsallis.	13
7	Entropie de Kaniadakis [8], [9], "K" entropie.	13
7.1	Divergence de Csiszär construite sur la "K" entropie.	14
7.2	Divergence de Csiszär duale construite sur la "K" entropie.	14
7.3	Divergence de Bregman construite sur la "K" entropie.	15
7.4	Divergence de Bregman duale construite sur la "K" entropie.	15

8 Entropie de Abe [1].	15
8.1 Divergence de Csiszär fondée sur l'entropie de Abe.	16
8.2 Divergence de Csiszär duale fondée sur l'entropie de Abe.	16
8.3 Divergence de Bregman fondée sur l'entropie de Abe.	17
8.4 Divergence de Bregman duale fondée sur l'entropie de Abe.	17
9 Gamma (γ) entropie.	18
9.1 Divergence de Csiszär construite sur la " γ " entropie.	18
9.2 Divergence de Csiszär duale construite sur la " γ " entropie.	19
9.3 Divergence de Bregman construite sur la " γ " entropie.	19
9.4 Divergence de Bregman duale construite sur la " γ " entropie.	19
10 Entropie à 2 paramètres [10], K.L.S.	20
10.1 Divergence de Csiszär construite sur l'entropie à 2 paramètres.	20
10.2 Divergence de Csiszär duale construite sur l'entropie à 2 paramètres.	21
10.3 Divergence de Bregman construite sur l'entropie à 2 paramètres.	21
10.4 Divergence de Bregman duale construite sur l'entropie à 2 paramètres.	22
11 Autres définitions de l'entropie et du Logarithme déformé.	22
12 Divergences fondées sur l'entropie générale (101).	24
12.1 Divergence de Csiszär construite sur l'entropie générale.	24
12.2 Divergence de Csiszär duale construite sur l'entropie générale.	25
12.3 Divergence de Bregman construite sur l'entropie générale.	26
12.4 Divergence de Bregman duale construite sur l'entropie générale.	27
12.5 Divergence de Jensen construite sur l'entropie générale.	28
12.6 Divergence de Jensen duale construite sur l'entropie générale.	29
13 Quelques autres expressions du Logarithme déformé.	29
13.1 Logarithme déformé de Newton [16].	30
13.1.1 Divergence de Csiszär fondée sur l'entropie de Newton.	30
13.1.2 Divergence de Csiszär duale fondée sur l'entropie de Newton.	30
13.1.3 Divergence de Bregman fondée sur l'entropie de Newton.	31
13.1.4 Divergence de Bregman duale fondée sur l'entropie de Newton.	32
13.1.5 Divergence de Jensen fondée sur l'entropie de Newton.	32
13.1.6 Divergence de Jensen duale fondée sur l'entropie de Newton.	33
13.2 Autre expression du Logarithme déformé [9].	33
13.2.1 Divergence de Csiszär correspondante.	34
13.2.2 Divergence de Csiszär duale correspondante.	34
14 Divergences invariantes.	34
14.1 Quelques rappels sur les facteurs d'invariance et leurs propriétés.	35
14.2 Divergences de Csiszär invariantes	36
14.2.1 Facteur d'invariance.	36
14.2.2 Divergences de Csiszär invariantes - Cas général.	37

14.2.3	Cas particulier : divergence de Csiszär/Tsallis invariante.	37
14.3	Divergences de Csiszär duales invariantes.	38
14.3.1	Facteur d'invariance.	38
14.3.2	Divergences de Csiszär duales invariantes - Cas général.	39
14.3.3	Cas particulier : divergence de Csiszär/Tsallis duale invariante.	41
14.4	Divergences de Bregman invariantes.	41
14.4.1	Facteur d'invariance.	41
14.4.2	Divergences de Bregman invariantes - Cas général.	42
14.4.3	Cas particulier - Divergence de Bregman/Tsallis invariante.	43
14.5	Divergences de Bregman duales - Forme invariante.	43
14.5.1	Facteur d'invariance.	43
14.5.2	Divergences de Bregman duales invariantes - Cas général.	44
14.5.3	Cas particulier de la divergence de Bregman/Tsallis duale invariante.	45
14.6	Divergence de Jensen invariante.	46
14.6.1	Facteur d'invariance.	46
14.6.2	Divergence de Jensen invariante.	46
14.6.3	Cas particulier : Divergence de Jensen/Shannon invariante.	47
14.7	Divergence de Jensen duale invariante.	48
15	Entropie de Newton - Formes invariantes des divergences.	49
15.1	Divergences de Csiszär invariantes fondées sur l'entropie de Newton.	49
15.2	Divergences de Csiszär/Newton duales invariantes	49
15.3	Divergences de Bregman invariantes fondées sur l'entropie de Newton.	50
15.4	Divergences de Bregman/Newton duale invariante.	51
15.5	Divergences de Jensen invariantes fondées sur l'entropie de Newton.	51
16	Algorithmique.	52
16.1	Généralités sur la méthode de construction des algorithmes.	52
16.2	Algorithmes fondés sur l'entropie générale.	55
16.2.1	Divergence de Csiszär construite sur l'entropie générale.	55
16.2.2	Divergence de Bregman construite sur l'entropie générale.	56
16.2.3	Divergence de Jensen construite sur l'entropie générale.	56
16.2.4	Divergence de Csiszär invariante construite sur l'entropie générale.	57
16.2.5	Divergence de Bregman invariante construite sur l'entropie générale.	58
16.2.6	Divergence de Jensen invariante construite sur l'entropie générale.	59
16.3	Divergences invariantes fondées sur l'entropie Newton. Algorithmes.	60
16.3.1	Divergence de Csiszär/Newton invariante.	61
16.3.2	Divergence de Bregman/Newton invariante.	61
16.3.3	Divergence de Jensen/Newton invariante.	61

17 Annexe 1	62
17.1 Divergences de Csiszär.	62
17.2 Divergences de Csiszär duales.	62
17.3 Divergences de Bregman.	63
17.4 Divergences de Bregman duales.	64
18 Annexe 2	64
19 Annexe 3	65
20 Annexe 4	66

1 Avertissement concernant les Versions 2 et 3

Dans la version 2, on a complété la version 1, en introduisant la divergence de Jensen fondée sur la forme la plus générale de l'entropie ainsi que la forme invariante de cette divergence ; les algorithmes associés sont développés.

Par ailleurs, les divergences fondées sur l'entropie de Newton ainsi que leurs formes invariantes qui n'entrent pas dans le cadre de la forme la plus générale de l'entropie sont présentées et analysées.

Ces divergences présentent certaines particularités spécifiques.

Dans la version 3, on a précisé certains points concernant la divergence de Jensen/Shannon qui présente quelques difficultés ; les problèmes algorithmiques sont précisés et ils sont détaillés dans l'Annexe 3.

Par ailleurs, la forme invariante de cette divergence est proposée afin de voir si les problèmes algorithmiques demeurent, ce qui n'est pas le cas.

Dans l'Annexe 4, on a indiqué quelques erreurs à éviter lors de l'application du Logarithme déformé à des divergences qui sont déjà sous forme Logarithmique, comme c'est le cas de la "Gamma Divergence".

2 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.

Après avoir précisé dans la section 1, les particularités de la Version 2 de ce travail, une introduction et un sommaire sont donnés dans la section 2. Nous donnons dans les sections 3 et 4, quelques brefs rappels sur les problèmes inverses et sur l'entropie.

Dans les sections 5 à 10, 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 11, 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 5 à 10, mais permet surtout de mettre en évidence une forme très générale de l'entropie qui synthétise toutes les précédentes.

Dans section 12, nous analysons les divergences construites sur la base de cette forme générale de l'entropie.

La section 13 est consacrée à des formes d'entropie qui sortent du cadre général précédemment décrit dans la section 11 ; les divergences associées à ces entropies particulières sont analysées, et plus spécifiquement les divergences fondées sur l'entropie de Newton.

Dans la section 14, nous étudions les formes invariantes par changement d'échelle des divergences fondées sur l'entropie générale présentées dans la section 12 ; les différentes formes de divergences sont considérées et le cas particulier de l'entropie de Tsallis est analysé en détails.

La section 15 traite des formes invariantes des divergences fondées sur l'entropie de Newton.

La section 16 est consacrée à l'aspect algorithmique.

Après 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 algorithmes correspondant aux formes invariantes des divergences fondées sur l'entropie de Newton sont explicités.

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 5 à 10 comme des applications spécifiques du cas général. Les algorithmes correspondant aux divergences invariantes fondées sur l'entro-

pie de Tsallis sont explicitées à titre de comparaison.

Nous résumons dans l'Annexe 1 l'opposé des gradients correspondant aux différentes formes d'entropie, afin de mettre clairement en évidence les analogies entre celles ci.

Les Annexes 2 et 3 détaillent les difficultés liés à la décomposition de l'opposé des gradients de certaines divergences.

3 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 construction 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 convexité par rapport à "q" ce qui entraîne 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 :

$$f_c(x) = f(x) - f(1) - (x - 1)f'(1) \quad (1)$$

Les divergences considérées dans ce travail sont les divergences de Csiszär et les divergences de Bregman ; les divergences de Jensen analysées dans [11], [12], sont considérées dans la présente version de ce travail.

On rappelle que la divergence de Csiszär construite sur la fonction convexe $f(x)$ s'écrit :

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

De même, la divergence de Bregman s'écrit :

$$B(p||q) = \sum_i f(p_i) - f(q_i) - (p_i - q_i)f'(q_i) \quad (3)$$

Enfin, la divergence de Jensen s'écrit sous sa forme la plus générale, avec $0 < \alpha < 1$:

$$J(p||q) = \sum_i \alpha f(p_i) + (1 - \alpha)f(q_i) - f[\alpha p_i + (1 - \alpha)q_i] \quad (4)$$

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 leur convexité par rapport au deuxième argument dépend de la fonction convexe “ f ” ; ceci justifie une analyse des divergences duales.

Les divergences de Jensen construites sur une fonction convexe simple ou sur une fonction convexe standard conduisent à la même expression ; la convexité des divergences de Jensen rapport à “ q ” dépend des propriétés de la fonction convexe “ f ”.

En ce qui concerne l'aspect algorithmique, nous nous appuyerons 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.

4 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) :

$$S = - \sum_i p_i \Lambda(p_i) = - \sum_i f(p_i) \quad (5)$$

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 $\Lambda(p_i)$ 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 (5), 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 (5) 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 pour 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.

5 Entropie de Shannon - Logarithme naturel.

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

$$S(p) = - \sum_i p_i \log p_i \quad (6)$$

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

$$f(x) = x \log x \quad (7)$$

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

$$f_c(x) = x \log x + 1 - x \quad (8)$$

5.1 Divergence de Csiszär associée à l'entropie de Shannon - Kullback-Leibler.

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

$$C(p\|q) = KL(p\|q) = \sum_i \left[p_i \log \frac{p_i}{q_i} - p_i + q_i \right] \quad (9)$$

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

$$-\frac{\partial KL(p\|q)}{\partial q_j} = \frac{p_j}{q_j} - 1 \quad (10)$$

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.

5.2 Divergence de Csiszär/Shannon duale.

Elle peut être construite à partir de (9) en permutant les rôles de "p" et "q" ou bien en construisant une divergence de Csiszär sur la fonction convexe miroir $\hat{f}_c(x) = x f_c(\frac{1}{x})$.

Elle s'écrit :

$$C(q\|p) = KL(q\|p) = \sum_i \left[q_i \log \frac{q_i}{p_i} - q_i + p_i \right] \quad (11)$$

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

$$-\frac{\partial KL(q\|p)}{\partial q_j} = \log \frac{p_j}{q_j} \quad (12)$$

On remarquera que cette expression ne peut pas se décomposer 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".

C'est à ma connaissance le seul cas où il en est ainsi, même si à la convergence c'est à dire pour $p_j = q_j \quad \forall j$, le gradient devient nul.

Ainsi, comme on le verra dans la section dédiée à l'algorithmique, l'écriture d'un algorithme multiplicatif ne sera pas possible pour cette divergence.

En revanche, il sera tout à fait possible d'écrire un algorithme ne faisant pas appel à une décomposition de l'opposé du gradient.

5.3 Divergences de Bregman/Shannon.

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.

5.4 Divergences de Jensen/Shannon.

La divergence de Jensen construite sur $f_c(x)$ ou sur $f(x)$ conduit à partir de (4) avec $0 < \alpha < 1$, à la divergence :

$$JS(p||q) = \sum_i \alpha p_i \log p_i + (1 - \alpha) q_i \log q_i - \sum_i [\alpha p_i + (1 - \alpha) q_i] \log [\alpha p_i + (1 - \alpha) q_i] \quad (13)$$

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

$$-\frac{\partial JS(p||q)}{\partial q_j} = (1 - \alpha) \log \left[\frac{\alpha p_j + (1 - \alpha) q_j}{q_j} \right] \quad (14)$$

Si on considère le terme $(1 - \alpha)$ comme un bloc positif, on se trouve alors dans une situation proche de celle de la section (6.2), et on peut en tirer les mêmes conclusions.

Cependant, une décomposition de cette expression sous forme d'une différence de 2 termes positifs est possible, mais dépend de la composante considérée.

En effet, on peut écrire :

$$-\frac{\partial JS(p||q)}{\partial q_j} = \alpha \log \left[\frac{q_j}{\alpha p_j + (1 - \alpha) q_j} \right] - \log \left[\frac{q_j}{\alpha p_j + (1 - \alpha) q_j} \right] \quad (15)$$

On voit ainsi apparaître une différence de 2 termes de même signe.

Si on veut que ces 2 termes soient positifs, on doit envisager les situations suivantes :

En notant $LL_j = \log \left[\frac{q_j}{\alpha p_j + (1 - \alpha) q_j} \right]$:

*** Pour les composantes telles que $\frac{q_j}{\alpha p_j + (1 - \alpha) q_j} > 1 \Rightarrow LL_j > 0$, on a :**

$$R_j = \alpha LL_j > 0 ; S_j = LL_j > 0 \quad (16)$$

*** Pour les composantes telles que $\frac{q_j}{\alpha p_j + (1 - \alpha) q_j} < 1 \Rightarrow LL_j < 0$, on a :**

$$R_j = -LL_j > 0 ; S_j = -\alpha LL_j > 0 \quad (17)$$

En vue de leur utilisation ultérieure, ces 2 cas se regroupent en une décomposition unique qui s'écrit :

$$\begin{aligned} R_j &= \frac{\alpha}{2} LL + \frac{\alpha}{2} |LL_j| - \frac{1}{2} LL_j + \frac{1}{2} |LL_j| > 0 \\ S_j &= \frac{1}{2} LL_j + \frac{1}{2} |LL_j| - \frac{\alpha}{2} LL_j + \frac{\alpha}{2} |LL_j| > 0 \end{aligned} \quad (18)$$

Ainsi, pour toutes les composantes, l'opposé du gradient s'écrira sous la forme $R_j - S_j$, avec $R_j > 0$ et $S_j > 0$.

Remarque : Dans l'Annexe 3 on analyse les algorithmes décrits dans la section (17) appliqués à cette divergence particulière.

6 Entropie de Tsallis [19].

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

$$S_T(p) = - \sum_i \frac{p_i^t - p_i}{t-1} \quad ; \quad t > 0 \quad (19)$$

On observera qu'en faisant le passage à la limite $t \rightarrow 1$ on retrouve l'entropie de Shannon.

Avec la définition (5) et le Logarithme déformé de Tsallis s'écrit :

$$\Lambda(x) = \log_T(x) = \frac{x^{t-1} - 1}{t-1} \quad (20)$$

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

$$\log_{TT}(x) = \frac{x^{1-t} - 1}{1-t} \quad (21)$$

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 :

$$S(p) = - \sum_i p_i^t \log_{TT}(p_i) \quad (22)$$

On retrouve l'entropie notée S_T (19).

Dans ce qui suit, on utilisera la forme (20) pour le logarithme déformé de Tsallis et on l'associera à la forme classique de l'entropie (5).

La fonction convexe permettant de construire des divergences s'écrira :

$$f(x) = \frac{x^t - x}{t-1} \quad (23)$$

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 utilisables dans notre problème.

La fonction convexe standard construite sur $f(x)$ s'écrit :

$$f_c(x) = \frac{x^t - x}{t-1} - x + 1 \quad (24)$$

6.1 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 :

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

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

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

Ce qu'on peut aussi écrire :

$$-\frac{\partial C_T(p||q)}{\partial q_j} = \frac{p_j}{q_j} \left(\frac{p_j}{q_j} \right)^{t-1} - 1 \quad (27)$$

Avec cette écriture, en faisant $t = 1$, on retrouve immédiatement l'expression de l'opposé du gradient correspondant à la divergence de Kullback-Leibler (10) fondée sur l'entropie de Shannon.

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

Elle se déduit immédiatement de l'expression (25) en effectuant la permutation $p \leftrightarrow q$, mais elle peut être construite directement au sens de Csiszär en s'appuyant sur la fonction convexe standard $\hat{f}_c(x)$ (fonction miroir), déduite de $f_c(x)$ par la transformation :

$$\hat{f}_c(x) = x f_c\left(\frac{1}{x}\right) \quad (28)$$

Elle s'écrit :

$$C_T(q||p) = \frac{1}{t-1} \left[\sum_i q_i^t p_i^{1-t} - t \sum_i q_i + (t-1) \sum_i p_i \right] \quad (29)$$

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

$$-\frac{\partial C_T(q||p)}{\partial q_j} = \frac{t}{1-t} \left[\left(\frac{p_j}{q_j} \right)^{1-t} - 1 \right] \quad (30)$$

Ce qui peut aussi s'écrire :

$$-\frac{\partial C_T(q||p)}{\partial q_j} = \frac{1}{1-t} \left[t \left(\frac{p_j}{q_j} \right)^{1-t} - 1 \right] + 1 \quad (31)$$

La décomposition de cette expression en une différence de 2 termes positifs dépend bien entendu du signe de $1-t$; cet aspect sera précisé ultérieurement.

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

6.3 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 :

$$B_T(p||q) = \sum_i \left[q_i^t - \frac{1}{1-t} p_i^t + \frac{t}{1-t} p_i q_i^{t-1} \right] \quad (32)$$

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

$$-\frac{\partial B_T(p||q)}{\partial q_j} = t q_j^{t-1} \left(\frac{p_j}{q_j} - 1 \right) \quad (33)$$

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

Elle se déduit immédiatement de l'expression (32) en effectuant la permutation $p \leftrightarrow q$; elle s'écrit :

$$B_T(q||p) = \sum_i \left[p_i^t - \frac{1}{1-t} q_i^t + \frac{t}{1-t} q_i p_i^{t-1} \right] \quad (34)$$

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

$$-\frac{\partial B_T(q||p)}{\partial q_j} = \frac{t}{1-t} q_j^{t-1} \left[\left(\frac{p_j}{q_j} \right)^{t-1} - 1 \right] \quad (35)$$

Comme dans la section (7.2), la décomposition de cette expression en une différence de 2 termes positifs dépend bien entendu du signe de $1-t$; cet aspect sera précisé ultérieurement.

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

L'expression de l'Entropie s'écrit :

$$S_K(p) = - \sum_i p_i \log_K(p_i) = - \sum_i \frac{p_i^{1+K} - p_i^{1-K}}{2K} \quad ; \quad -1 < K < 1 \quad (36)$$

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 $K \rightarrow 0$.

La fonction convexe déduite de cette expression sera :

$$f(x) = \frac{x^{1+K} - x^{1-K}}{2K} \quad (37)$$

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 :

$$f_c(x) = \frac{x^{1+K} - x^{1-K}}{2K} - x + 1 \quad (38)$$

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

$$\log_K(x) = \frac{x^K - x^{-K}}{2K} \quad (39)$$

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

7.1 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 :

$$C_K(p||q) = \frac{1}{2K} \left[\sum_i p_i^{1+K} q_i^{-K} - \sum_i p_i^{1-K} q_i^K \right] - \sum_i p_i + \sum_i q_i \quad (40)$$

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

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

Ce qui peut encore s'écrire :

$$-\frac{\partial C_K(p||q)}{\partial q_j} = \frac{p_j}{q_j} \left[\frac{1}{2} \left(\frac{p_j}{q_j} \right)^K + \frac{1}{2} \left(\frac{p_j}{q_j} \right)^{-K} \right] - 1 \quad (42)$$

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.

7.2 Divergence de Csiszär duale construite sur la “ K ” entropie.

Elle se déduit de la relation (40) en faisant la permutation $p \leftrightarrow q$; elle s'écrit :

$$C_K(q||p) = \frac{1}{2K} \left[\sum_i q_i^{1+K} p_i^{-K} - \sum_i q_i^{1-K} p_i^K \right] - \sum_i q_i + \sum_i p_i \quad (43)$$

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

$$-\frac{\partial C_K(q||p)}{\partial q_j} = \left[\frac{1-K}{2K} \left(\frac{p_j}{q_j} \right)^K - \frac{1+K}{2K} \left(\frac{p_j}{q_j} \right)^{-K} \right] + 1 \quad (44)$$

7.3 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 :

$$B_K(p||q) = \frac{1}{2K} \sum_i [p_i^{1+K} - p_i^{1-K} - q_i^{1+K} + q_i^{1-K} - (1+K)p_i q_i^K + (1-K)p_i q_i^{-K} + (1+K)q_i^{1+K} - (1-K)q_i^{1-K}] \quad (45)$$

L’opposé du gradient par rapport à “ q ” s’écrit tous calculs faits $\forall j$:

$$-\frac{\partial B_K(p||q)}{\partial q_j} = \left(\frac{p_j}{q_j} - 1 \right) \left[\frac{1+K}{2} q_j^K + \frac{1-K}{2} q_j^{-K} \right] \quad (46)$$

7.4 Divergence de Bregman duale construite sur la “ K ” entropie.

Elle est construite à partir de la relation (45) en faisant la permutation $p \leftrightarrow q$; on obtient :

$$B_K(q||p) = \frac{1}{2K} \sum_i [q_i^{1+K} - q_i^{1-K} - p_i^{1+K} + p_i^{1-K} - (1+K)q_i p_i^K + (1-K)q_i p_i^{-K} + (1+K)p_i^{1+K} - (1-K)p_i^{1-K}] \quad (47)$$

L’opposé du gradient par rapport à “ q ” s’écrit tous calculs faits $\forall j$:

$$-\frac{\partial B_K(q||p)}{\partial q_j} = \left[\frac{1+K}{2K} (p_j^K - q_j^K) - \frac{1-K}{2K} (p_j^{-K} - q_j^{-K}) \right] \quad (48)$$

8 Entropie de Abe [1].

L’entropie correspondante s’écrit :

$$S_A(p) = - \sum_i p_i \log_A(p_i) = - \sum_i \frac{p_i^z - p_i^{\frac{1}{z}}}{z - \frac{1}{z}} \quad (49)$$

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 $z \rightarrow 1$.

La fonction convexe permettant de construire des divergences s’écrit :

$$f(x) = \frac{x^z - x^{\frac{1}{z}}}{z - \frac{1}{z}} \quad (50)$$

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 utilisables dans notre problème.

La fonction convexe standard construite sur $f(x)$ s'écrit :

$$f_c(x) = \frac{x^z - x^{\frac{1}{z}}}{z - \frac{1}{z}} - x + 1 \quad (51)$$

Avec la définition (5) l'expression du Logarithme déformé de Abe s'écrit :

$$\log_A(x) = \frac{x^{z-1} - x^{\frac{1}{z}-1}}{z - \frac{1}{z}} \quad (52)$$

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

8.1 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 :

$$C_A(p||q) = \frac{z}{z^2 - 1} \left[\sum_i p_i^z q_i^{1-z} - \sum_i p_i^{\frac{1}{z}} q_i^{1-\frac{1}{z}} \right] - \sum_i p_i + \sum_i q_i \quad (53)$$

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

$$-\frac{\partial C_A(p||q)}{\partial q_j} = \frac{z}{z+1} \left(\frac{p_j}{q_j} \right)^z + \frac{1}{z+1} \left(\frac{p_j}{q_j} \right)^{\frac{1}{z}} - 1 \quad (54)$$

Ce qui peut aussi s'écrire :

$$-\frac{\partial C_A(p||q)}{\partial q_j} = \frac{p_j}{q_j} \left[\frac{z}{z+1} \left(\frac{p_j}{q_j} \right)^{z-1} + \frac{1}{z+1} \left(\frac{p_j}{q_j} \right)^{\frac{1}{z}-1} \right] - 1 \quad (55)$$

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.

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

Elle se déduit de l'expression (53) en effectuant la permutation $p \leftrightarrow q$; on obtient :

$$C_A(q||p) = \frac{z}{z^2 - 1} \left[\sum_i q_i^z p_i^{1-z} - \sum_i q_i^{\frac{1}{z}} p_i^{1-\frac{1}{z}} \right] - \sum_i q_i + \sum_i p_i \quad (56)$$

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

$$-\frac{\partial C_A(q||p)}{\partial q_j} = \frac{z}{1-z^2} \left[z \left(\frac{p_j}{q_j} \right)^{1-z} - \frac{1}{z} \left(\frac{p_j}{q_j} \right)^{1-\frac{1}{z}} \right] + 1 \quad (57)$$

Ce qui peut aussi s'écrire :

$$-\frac{\partial C_A(q||p)}{\partial q_j} = \frac{p_j}{q_j} \left[\frac{z^2}{1-z^2} \left(\frac{p_j}{q_j} \right)^{-z} - \frac{1}{1-z^2} \left(\frac{p_j}{q_j} \right)^{-\frac{1}{z}} \right] + 1 \quad (58)$$

8.3 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 :

$$\begin{aligned} B_A(p||q) = & \frac{1}{z - \frac{1}{z}} \sum_i \left[p_i^z - p_i^{\frac{1}{z}} - q_i^z + q_i^{\frac{1}{z}} \right. \\ & \left. - z p_i q_i^{(z-1)} + \frac{1}{z} p_i q_i^{(\frac{1}{z}-1)} \right. \\ & \left. + z q_i^z - \frac{1}{z} q_i^{\frac{1}{z}} \right] \end{aligned} \quad (59)$$

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

$$-\frac{\partial B_A(p||q)}{\partial q_j} = \left(\frac{p_j}{q_j} - 1 \right) \left[\frac{z^2}{z+1} q_j^{(z-1)} + \frac{1}{z(z+1)} q_j^{(\frac{1}{z}-1)} \right] \quad (60)$$

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

Elle se déduit de l'expression (59) en effectuant la permutation $p \leftrightarrow q$; elle s'écrit :

$$\begin{aligned} B_A(q||p) = & \frac{1}{z - \frac{1}{z}} \sum_i \left[q_i^z - q_i^{\frac{1}{z}} - p_i^z + p_i^{\frac{1}{z}} \right. \\ & \left. - z q_i p_i^{(z-1)} + \frac{1}{z} q_i p_i^{(\frac{1}{z}-1)} \right. \\ & \left. + z p_i^z - \frac{1}{z} p_i^{\frac{1}{z}} \right] \end{aligned} \quad (61)$$

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

$$-\frac{\partial B_A(q||p)}{\partial q_j} = \left[\frac{z^2}{z^2-1} (p_j^{z-1} - q_j^{z-1}) - \frac{1}{z^2-1} (p_j^{\frac{1}{z}-1} - q_j^{\frac{1}{z}-1}) \right] \quad (62)$$

9 Gamma (γ) entropie.

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

$$S_\gamma(p) = - \sum_i p_i \log_\gamma(p_i) = - \sum_i \frac{p_i^{1+2\gamma} - p_i^{1-\gamma}}{3\gamma} \quad (63)$$

Cette fonction est définie pour $-0.5 \leq \gamma \leq 1$.

Le passage à la limite $\gamma \rightarrow 0$ permet de retrouver l'entropie de Shannon.

La fonction convexe permettant de construire des divergences s'écrit :

$$f(x) = \frac{x^{1+2\gamma} - x^{1-\gamma}}{3\gamma} \quad (64)$$

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 utilisables dans notre problème.

La fonction convexe standard construite sur $f(x)$ s'écrit :

$$f_c(x) = \frac{x^{1+2\gamma} - x^{1-\gamma}}{3\gamma} + 1 - x \quad (65)$$

Avec la définition (5) l'expression du " γ " Logarithme déformé s'écrit :

$$\log_\gamma(x) = \frac{x^{2\gamma} - x^{-\gamma}}{3\gamma} \quad (66)$$

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

9.1 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 :

$$C_\gamma(p||q) = \frac{1}{3\gamma} \left[\sum_i p_i^{1+2\gamma} q_i^{-2\gamma} - \sum_i p_i^{1-\gamma} q_i^\gamma \right] - \sum_i p_i + \sum_i q_i \quad (67)$$

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

$$-\frac{\partial C_\gamma(p||q)}{\partial q_j} = \frac{2}{3} \left(\frac{p_j}{q_j} \right)^{1+2\gamma} + \frac{1}{3} \left(\frac{p_j}{q_j} \right)^{1-\gamma} - 1 \quad (68)$$

Ce qui peut aussi s'écrire :

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

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.

9.2 Divergence de Csiszär duale construite sur la “ γ ” entropie.

Elle est construite à partir de l’expression (67), en effectuant la permutation $p \leftrightarrow q$

$$C_\gamma(q\|p) = \frac{1}{3\gamma} \left[\sum_i q_i^{1+2\gamma} p_i^{-2\gamma} - \sum_i q_i^{1-\gamma} p_i^\gamma \right] - \sum_i q_i + \sum_i p_i \quad (70)$$

L’opposé du gradient par rapport à “ q ” s’écrit $\forall j$:

$$-\frac{\partial C_\gamma(q\|p)}{\partial q_j} = \frac{1}{3\gamma} \left[(1-\gamma) \left(\frac{p_j}{q_j} \right)^\gamma - (1+2\gamma) \left(\frac{p_j}{q_j} \right)^{-2\gamma} \right] + 1 \quad (71)$$

Ou encore :

$$-\frac{\partial C_\gamma(q\|p)}{\partial q_j} = \frac{p_j}{q_j} \left[\frac{1-\gamma}{3\gamma} \left(\frac{p_j}{q_j} \right)^{\gamma-1} - \frac{1+2\gamma}{3\gamma} \left(\frac{p_j}{q_j} \right)^{-2\gamma-1} \right] + 1 \quad (72)$$

9.3 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 :

$$\begin{aligned} B_\gamma(p\|q) = \frac{1}{3\gamma} \sum_i & \left[p_i^{2\gamma+1} - p_i^{1-\gamma} - q_i^{2\gamma+1} + q_i^{1-\gamma} \right. \\ & - (2\gamma+1)p_i q_i^{2\gamma} + (1-\gamma)p_i q_i^{-\gamma} \\ & \left. + (2\gamma+1)q_i^{2\gamma+1} - (1-\gamma)q_i^{1-\gamma} \right] \end{aligned} \quad (73)$$

L’opposé du gradient par rapport à “ q ” s’écrit $\forall j$:

$$-\frac{\partial B_\gamma(p\|q)}{\partial q_j} = \left(\frac{p_j}{q_j} - 1 \right) \left[\frac{2(2\gamma+1)}{3} q_j^{2\gamma} + \frac{1-\gamma}{3} q_j^{-\gamma} \right] \quad (74)$$

9.4 Divergence de Bregman duale construite sur la “ γ ” entropie.

Elle est construite à partir de la relation (73), en effectuant la permutation $p \leftrightarrow q$; elle s’écrit :

$$\begin{aligned} B_\gamma(q\|p) = \frac{1}{3\gamma} \sum_i & \left[q_i^{2\gamma+1} - q_i^{1-\gamma} - p_i^{2\gamma+1} + p_i^{1-\gamma} \right. \\ & - (2\gamma+1)q_i p_i^{2\gamma} + (1-\gamma)q_i p_i^{-\gamma} \\ & \left. + (2\gamma+1)p_i^{2\gamma+1} - (1-\gamma)p_i^{1-\gamma} \right] \end{aligned} \quad (75)$$

L’opposé du gradient par rapport à “ q ” s’écrit $\forall j$:

$$-\frac{\partial B_\gamma(q\|p)}{\partial q_j} = \left[\frac{2\gamma+1}{3\gamma} (p_j^{2\gamma} - q_j^{2\gamma}) - \frac{1-\gamma}{3\gamma} (p_j^{-\gamma} - q_j^{-\gamma}) \right] \quad (76)$$

10 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 :

$$S_{K,r}(p) = - \sum_i p_i \text{Log}_{(K,r)}(p_i) = - \sum_i p_i^{1+r} \frac{p_i^K - p_i^{-K}}{2K} \quad (77)$$

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 $K \rightarrow 0$.

Avec la définition (5), l'expression du Logarithme déformé à 2 paramètres (K, r) proposée par Kaniadakis et al. [10] s'écrit :

$$\text{Log}_{(K,r)}(x) = x^r \frac{x^K - x^{-K}}{2K} \quad (78)$$

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 :

$$f(x) = \frac{x^{1+K+r} - x^{1-K+r}}{2K} \quad (79)$$

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

$$f_c(x) = \frac{x^{1+K+r} - x^{1-K+r}}{2K} + 1 - x \quad (80)$$

10.1 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 :

$$C_{KLS}(p||q) = \frac{1}{2K} \left[\sum_i p_i^{1+r+K} q_i^{-r-K} - \sum_i p_i^{1+r-K} q_i^{-r+K} \right] - \sum_i p_i + \sum_i q_i \quad (81)$$

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

$$-\frac{\partial C_{KLS}(p||q)}{\partial q_j} = \frac{1}{2K} \left[(K+r) \left(\frac{p_j}{q_j} \right)^{1+r+K} + (K-r) \left(\frac{p_j}{q_j} \right)^{1+r-K} \right] - 1 \quad (82)$$

Ce qui peut aussi s'écrire :

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

Si, à partir de cette expression, on fait $r = 0$, on retrouve ce qu'on a obtenu pour le Logarithme déformé de Kaniadakis (41), puis en faisant $K \rightarrow 0$ 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 : $-K \leq r \leq K$.

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

Elle est construite à partir de l'expression (81) en effectuant la permutation $p \leftrightarrow q$; elle s'écrit :

$$C_{KLS}(q||p) = \frac{1}{2K} \left[\sum_i q_i^{1+r+K} p_i^{-r-K} - \sum_i q_i^{1+r-K} p_i^{-r+K} \right] - \sum_i q_i + \sum_i p_i \quad (84)$$

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

$$-\frac{\partial C_{KLS}(q||p)}{\partial q_j} = \frac{1}{2K} \left[(1+r-K) \left(\frac{p_j}{q_j}\right)^{K-r} - (1+r+K) \left(\frac{p_j}{q_j}\right)^{-K-r} \right] + 1 \quad (85)$$

Ou encore :

$$-\frac{\partial C_{KLS}(q||p)}{\partial q_j} = \frac{p_j}{q_j} \left[\frac{1+r-K}{2K} \left(\frac{p_j}{q_j}\right)^{K-r-1} - \frac{1+r+K}{2K} \left(\frac{p_j}{q_j}\right)^{-K-r-1} \right] + 1 \quad (86)$$

10.3 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 :

$$B_{KLS}(p||q) = \frac{1}{2K} \sum_i [p_i^{1+r+K} - p_i^{1+r-K} - q_i^{1+r+K} + q_i^{1+r-K} - (1+r+K)p_i q_i^{r+K} + (1+r+K)q_i^{1+r+K} + (1+r-K)p_i q_i^{r-K} - (1+r-K)q_i^{1+r-K}] \quad (87)$$

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

$$-\frac{\partial B_{KLS}(p||q)}{\partial q_j} = \left(\frac{p_j}{q_j} - 1\right) \left[\frac{(1+r+K)(r+K)}{2K} q_j^{r+K} - \frac{(1+r-K)(r-K)}{2K} q_j^{r-K} \right] \quad (88)$$

En posant : $a = 1 + r + K$ et $b = 1 + r - K$ on obtient une divergence générale (120) qui sera analysée ultérieurement

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

Elle est construite à partir de la relation (87) en effectuant la permutation $p \leftrightarrow q$.

$$B_{KLS}(q||p) = \frac{1}{2K} \sum_i [q_i^{1+r+K} - q_i^{1+r-K} - p_i^{1+r+K} + p_i^{1+r-K} - (1+r+K)q_i p_i^{r+K} + (1+r+K)p_i^{1+r+K} + (1+r-K)q_i p_i^{r-K} - (1+r-K)p_i^{1+r-K}] \quad (89)$$

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

$$-\frac{\partial B_{KLS}(q||p)}{\partial q_j} = \left(\frac{1+r+K}{2K}\right) (p_j^{r+K} - q_j^{r+K}) - \left(\frac{1+r-K}{2K}\right) (p_j^{r-K} - q_j^{r-K}) \quad (90)$$

11 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 5 à 10 et de donner une forme très générale de l'entropie qui regroupe celles-ci.

Les expressions de l'entropie qui seront mentionnées dans le paragraphe 13 n'entrent pas dans ce contexte.

Dans les sections précédentes la forme de l'entropie utilisée était définie par la relation :

$$S(p) = - \sum_i p_i \Lambda(p_i) \quad (91)$$

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 :

$$S(p) = \left[-\frac{d}{d\alpha} \sum_i p_i^\alpha \right]_{\alpha=1} \equiv \left[-\frac{d}{d\alpha} f(\alpha) \right]_{\alpha=1} \quad (92)$$

Dans cette expression la dérivée mise en jeu est la dérivée habituelle. A partir de (92), il est proposé de déformer la dérivée classique et de la remplacer par la “ q ” différentielle de Jackson [6] [7] qui s’écrit, en changeant la notation ($q \leftrightarrow t$)

$$\frac{df(\alpha)}{d(\alpha : t)} = \frac{f(\alpha t) - f(\alpha)}{\alpha t - \alpha} \quad (93)$$

(On pourrait presque dire qu’on a remplacé l’accroissement additif de la variable dans la dérivée classique, par un accroissement multiplicatif de la variable.)

Avec cette définition, on obtient :

$$S(p) = \left[-\frac{\sum_i p_i^{\alpha t} - \sum_i p_i^\alpha}{\alpha t - \alpha} \right]_{\alpha=1} \quad (94)$$

Ce qui correspond à l’entropie de Tsallis [19], déjà mentionnée en (19) :

$$S_T(p) = -\frac{\sum_i p_i^t - \sum_i p_i}{t - 1} = -\frac{\sum_i p_i^t - 1}{t - 1} \quad (95)$$

Le Logarithme déformé correspondant s’écrit (20) :

$$\text{Log}_T(x) = \frac{x^{t-1} - 1}{t - 1} \quad (96)$$

Une alternative consiste, toujours à partir de (92), à adopter une autre définition de la différentielle donnée par Mc Anally [13] qui s’écrit :

$$\frac{df(\alpha)}{d(\alpha : t)} = \frac{f(\alpha t) - f(\alpha t^{-1})}{\alpha t - \alpha t^{-1}} \quad (97)$$

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

$$S(p) = \left[-\frac{\sum_i p_i^{\alpha t} - \sum_i p_i^{\alpha t^{-1}}}{\alpha t - \alpha t^{-1}} \right]_{\alpha=1} = -\frac{\sum_i p_i^t - \sum_i p_i^{t^{-1}}}{t - t^{-1}} \quad (98)$$

On retrouve ainsi l’entropie de Abe (49) ; le Logarithme déformé correspondant s’écrit (52) :

$$\text{Log}_A(x) = \frac{x^{t-1} - x^{t^{-1}-1}}{t - t^{-1}} \quad (99)$$

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 :

$$\frac{d_{ab}f(\alpha)}{d_{ab}(\alpha)} = \frac{f(a\alpha) - f(b\alpha)}{a\alpha - b\alpha} \quad (100)$$

Cette forme étend les définitions précédemment proposées, elle conduit, avec la définition (92) à l’expression de l’entropie ;

$$S_{ab}(p) = \left[-\frac{\sum_i p_i^{a\alpha} - \sum_i p_i^{b\alpha}}{a\alpha - b\alpha} \right]_{\alpha=1} = -\frac{\sum_i p_i^a - \sum_i p_i^b}{a - b} \quad (101)$$

Cette forme d'entropie très générale (et le logarithme déformé 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 5 à 10.

Les domaines de valeurs des paramètres “ a ” et “ b ” indiqués dans [5] et [3] sont les suivants :

$$0 \leq a \leq 1 \leq b \quad (102)$$

ou bien :

$$0 \leq b \leq 1 \leq a \quad (103)$$

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

$$\text{Log}_{ab}(x) = \frac{x^{a-1} - x^{b-1}}{a - b} \quad (104)$$

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

$$0 \leq a \leq 1 \leq b \leq 2 \quad ; \quad 0 \leq b \leq 1 \leq a \leq 2 \quad (105)$$

12 Divergences fondées sur l'entropie générale (101).

On développe dans les paragraphes suivants les différentes formes de divergences fondées sur l'entropie générale (101).

Les cas particuliers développés dans les paragraphes 5 à 10 se retrouvent en effectuant les adaptations suivantes :

- Entropie de Shannon : $a \rightarrow 1$ et $b \rightarrow 1$ en respectant (102) ou (103)
- Entropie de Tsallis : $a = t$, $b = 1$
- Entropie de Kaniadakis : $a = 1 + K$, $b = 1 - K$
- Entropie de Abe : $a = z$, $b = \frac{1}{z}$
- Entropie “ γ ” : $a = 2\gamma + 1$, $b = 1 - \gamma$
- Entropie à 2 paramètres (KLS) : $a = 1 + r + K$, $b = 1 + r - K$

12.1 Divergence de Csiszär construite sur l'entropie générale.

La fonction de base convexe déduite de la forme générale de l'entropie (101) s'écrit :

$$f(x) = \frac{x^a - x^b}{a - b} \quad (106)$$

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

$$f_c(x) = \frac{x^a - x^b}{a - b} - x + 1 \quad (107)$$

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

$$C_{ab}(p||q) = \frac{1}{a - b} \left[\sum_i p_i^a q_i^{1-a} - \sum_i p_i^b q_i^{1-b} \right] - \sum_i p_i + \sum_i q_i \quad (108)$$

A partir de cette expression, en s'appuyant sur les inégalités (102) ou (103), on retrouve la divergence de Kullback-Leibler en faisant le passage à la limite $a \rightarrow 1$ et $b \rightarrow 1$.

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

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

Ce qui peut encore s'écrire :

$$-\frac{\partial C_{ab}(p||q)}{\partial q_j} = \left(\frac{p_j}{q_j} \right) \left[\frac{a - 1}{a - b} \left(\frac{p_j}{q_j} \right)^{a-1} + \frac{1 - b}{a - b} \left(\frac{p_j}{q_j} \right)^{b-1} \right] - 1 \quad (110)$$

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

$$\frac{a - 1}{a - b} > 0 \quad ; \quad \frac{1 - b}{a - b} > 0 \quad (111)$$

Ces inégalités sont satisfaites si "a" et "b" appartiennent aux domaines de valeurs donnés par (102) ou (103).

On observera que pour la divergence de Csiszar/Tsallis qui correspond à ($a = t$) et ($b = 1$), l'expression de l'opposé du gradient se simplifie.

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

Elle est construite à partir de l'expression (108) en effectuant la permutation $p \leftrightarrow q$; elle s'écrit :

$$C_{ab}(q||p) = \frac{1}{a - b} \left[\sum_i q_i^a p_i^{1-a} - \sum_i q_i^b p_i^{1-b} \right] - \sum_i q_i + \sum_i p_i \quad (112)$$

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

$$-\frac{\partial C_{ab}(q||p)}{\partial q_j} = \frac{1}{a - b} \left[b \left(\frac{p_j}{q_j} \right)^{1-b} - a \left(\frac{p_j}{q_j} \right)^{1-a} \right] + 1 \quad (113)$$

Ce qui peut s'écrire sous la forme générale :

$$-\frac{\partial C_{ab}(q||p)}{\partial q_j} = \frac{1}{a-b} [A_j - B_j] + 1 \quad (114)$$

Avec $A_j > 0$ et $B_j > 0$.

Ou encore :

$$-\frac{\partial C_{ab}(q||p)}{\partial q_j} = \frac{p_j}{q_j} \left[\frac{b}{a-b} \left(\frac{p_j}{q_j} \right)^{-b} - \frac{a}{a-b} \left(\frac{p_j}{q_j} \right)^{-a} \right] + 1 \quad (115)$$

On observera que pour la divergence de Csiszar/Tsallis duale qui correspond à ($a = t$) et ($b = 1$), l'expression de l'opposé du gradient se simplifie.

Quoiqu'il en soit, la décomposition de l'opposé du gradient en une différence de 2 termes positifs change suivant le signe de $a - b$.

Une méthode générale permettant de résoudre ce problème est développée dans l'Annexe 2 ; on détaille cet aspect à titre d'exemple pour le cas considéré ici.

On veut écrire l'opposé du gradient sous la forme $R_j - S_j$; $R_j > 0$; $S_j > 0$; alors,

* **Si $a - b > 0$, on a :**

$$R_j = \frac{b}{a-b} \left(\frac{p_j}{q_j} \right)^{1-b} + 1 ; S_j = \frac{a}{a-b} \left(\frac{p_j}{q_j} \right)^{1-a} \quad (116)$$

* **Si $a - b < 0$, on a :**

$$R_j = \frac{a}{b-a} \left(\frac{p_j}{q_j} \right)^{1-a} + 1 ; S_j = \frac{b}{b-a} \left(\frac{p_j}{q_j} \right)^{1-b} \quad (117)$$

Pour globaliser ces deux cas, on peut écrire quel que soit le signe de $a - b$:

$$R_j = 1 + \frac{b}{2} \left(\frac{p_j}{q_j} \right)^{1-b} \left[\frac{1}{a-b} + \frac{1}{|a-b|} \right] + \frac{a}{2} \left(\frac{p_j}{q_j} \right)^{1-a} \left[\frac{1}{b-a} + \frac{1}{|b-a|} \right] \quad (118)$$

$$S_j = \frac{a}{2} \left(\frac{p_j}{q_j} \right)^{1-a} \left[\frac{1}{a-b} + \frac{1}{|a-b|} \right] + \frac{b}{2} \left(\frac{p_j}{q_j} \right)^{1-b} \left[\frac{1}{b-a} + \frac{1}{|b-a|} \right] \quad (119)$$

12.3 Divergence de Bregman construite sur l'entropie générale.

Cette divergence peut être construite soit sur la base de la fonction $f(x)$ (106), soit sur la base de $f_c(x)$ (107) ; l'expression obtenue est identique et

s'écrit tous calculs effectués :

$$B_{ab}(p||q) = \frac{1}{a-b} \sum_i [p_i^a - p_i^b + (a-1)q_i^a - (b-1)q_i^b - ap_iq_i^{a-1} + bp_iq_i^{b-1}] \quad (120)$$

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

$$-\frac{\partial B_{ab}(p||q)}{\partial q_j} = \left(\frac{p_j}{q_j} - 1\right) \left[\frac{(a^2 - a)}{a-b} q_j^{a-1} + \frac{(b-b^2)}{a-b} q_j^{b-1} \right] \quad (121)$$

On observera que pour la divergence de Bregman/Tsallis qui correspond à ($a = t$) et ($b = 1$), l'expression de l'opposé du gradient se simplifie.

Compte tenu des domaines de valeurs de "a" et "b" définies dans (102) ou (103), on a :

$$\frac{a^2 - a}{a-b} > 0 \quad ; \quad \frac{b-b^2}{a-b} \geq 0 \quad (122)$$

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

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

Elle se déduit de l'expression (120) par permutation de "p" et "q" ; elle s'écrit :

$$B_{ab}(q||p) = \frac{1}{a-b} \sum_i [q_i^a - q_i^b + (a-1)p_i^a - (b-1)p_i^b - aq_i p_i^{a-1} + bq_i p_i^{b-1}] \quad (123)$$

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

$$-\frac{\partial B_{ab}(q||p)}{\partial q_j} = \frac{1}{a-b} [a(p_j^{a-1} - q_j^{a-1}) - b(p_j^{b-1} - q_j^{b-1})] \quad (124)$$

Ce qui peut aussi s'écrire :

$$-\frac{\partial B_{ab}(q||p)}{\partial q_j} = \frac{1}{a-b} (A_j - B_j) \quad (125)$$

Avec :

$$A_j = ap_j^{a-1} + bq_j^{b-1} > 0 \quad (126)$$

et,

$$B_j = aq_j^{a-1} + bp_j^{b-1} > 0 \quad (127)$$

On observera que pour la divergence de Bregman/Tsallis duale qui correspond à ($a = t$) et ($b = 1$), l'expression de l'opposé du gradient se simplifie, et, par voie de conséquence les expressions de A_j et B_j .

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

En notant $R_j - S_j$ cette décomposition, on peut préciser ce dernier point comme suit :

*** Si $a - b > 0$, on a :**

$$R_j = \frac{1}{a-b} A_j > 0 \quad ; \quad S_j = \frac{1}{a-b} B_j > 0 \quad (128)$$

*** Si $a - b < 0$, on a :**

$$R_j = \frac{1}{b-a} B_j > 0 \quad ; \quad S_j = \frac{1}{b-a} A_j > 0 \quad (129)$$

On peut globaliser les deux cas en écrivant :

$$R_j = \frac{A_j}{2} \left[\frac{1}{a-b} + \frac{1}{|a-b|} \right] + \frac{B_j}{2} \left[\frac{1}{b-a} + \frac{1}{|b-a|} \right] > 0 \quad (130)$$

$$S_j = \frac{B_j}{2} \left[\frac{1}{a-b} + \frac{1}{|a-b|} \right] + \frac{A_j}{2} \left[\frac{1}{b-a} + \frac{1}{|b-a|} \right] > 0 \quad (131)$$

12.5 Divergence de Jensen construite sur l'entropie générale.

Cette divergence peut être construite soit sur la base de la fonction $f(x)$ (106), soit sur la base de $f_c(x)$ (107) ; l'expression obtenue est identique. On développe ici la forme la plus générale de la divergence de Jensen ("α" Jensen), qui s'écrit, avec $0 \leq \alpha \leq 1$:

$$J_\alpha(p||q) = \sum_i \alpha \frac{p_i^a - p_i^b}{a-b} + (1-\alpha) \frac{q_i^a - q_i^b}{a-b} - \frac{[\alpha p_i + (1-\alpha)q_i]^a}{a-b} + \frac{[\alpha p_i + (1-\alpha)q_i]^b}{a-b} \quad (132)$$

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

$$-\frac{\partial J_\alpha(p||q)}{\partial q_j} = \frac{1-\alpha}{a-b} (A_j - B_j) \quad (133)$$

Avec :

$$A_j = b q_j^{b-1} + a [\alpha p_j + (1-\alpha)q_j]^{a-1} > 0 \quad (134)$$

et,

$$B_j = a q_j^{a-1} + b [\alpha p_j + (1-\alpha)q_j]^{b-1} > 0 \quad (135)$$

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

Pour préciser ce point, en notant l'opposé du gradient $R_j - S_j$, $R_j > 0$, $S_j > 0$, on a :

* Si $a - b > 0$:

$$R_j = \frac{1 - \alpha}{a - b} A_j > 0 \ ; \ S_j = \frac{1 - \alpha}{a - b} B_j > 0 \quad (136)$$

* Si $a - b < 0$:

$$R_j = \frac{1 - \alpha}{b - a} B_j > 0 \ ; \ S_j = \frac{1 - \alpha}{b - a} A_j > 0 \quad (137)$$

Tout cela peut se globaliser avec :

$$R_j = \frac{1 - \alpha}{2} A_j \left[\frac{1}{a - b} + \frac{1}{|a - b|} \right] + \frac{1 - \alpha}{2} B_j \left[\frac{1}{b - a} + \frac{1}{|b - a|} \right] > 0 \quad (138)$$

$$S_j = \frac{1 - \alpha}{2} B_j \left[\frac{1}{a - b} + \frac{1}{|a - b|} \right] + \frac{1 - \alpha}{2} A_j \left[\frac{1}{b - a} + \frac{1}{|b - a|} \right] > 0 \quad (139)$$

Remarques :

* 1 - Compte tenu de (102) ou (103), le passage à la limite $a \rightarrow 1$ et $b \rightarrow 1$ (avec par exemple $a = 1 + \epsilon$ et $b = 1 - \epsilon$, $\epsilon \rightarrow 0$), permet de retrouver le cas de la divergence de Jensen/Shannon (section 5.4).

* 2 - Par ailleurs, les divergences de Jensen formées sur les entropies des sections 6 à 10 s'obtiennent sans difficulté en appliquant sur les expressions de la sous-section (12.5), les adaptations précisées en début de section (12).

* 3 - Dans le cas particulier de la divergence de Jensen/Tsallis des simplifications apparaissent l'expression de l'opposé du gradient et par voie de conséquence, dans les expressions de A_j et B_j .

12.6 Divergence de Jensen duale construite sur l'entropie générale.

L'analyse de cette divergence est inutile, car elle se déduit de la divergence de Jensen décrite dans la section précédente en modifiant la valeur du paramètre " α "; on pourrait dire que la divergence de Jensen est "*self duale*".

13 Quelques autres expressions du Logarithme déformé.

On trouve dans la littérature 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].

13.1 Logarithme déformé de Newton [16].

Il est donné par l'expression :

$$L_N(x) = \frac{1}{2} (x - 1 + \text{Log} x) \quad (140)$$

En s'appuyant sur la relation (5) on obtient pour l'entropie, l'expression spécifique :

$$S_N(p) = - \sum_i p_i L_N(p_i) = - \sum_i \frac{1}{2} [p_i^2 - p_i + p_i \text{Log} p_i] \quad (141)$$

La fonction convexe de base s'écrit en conséquence :

$$f(x) = \frac{1}{2} (x^2 - x + x \text{Log} x) \quad (142)$$

Ce n'est pas une fonction complète, en effet $f'(1) = 1$.

La fonction convexe standard qu'on en déduit, s'écrit :

$$f_c(x) = \frac{1}{2} (x^2 - 3x + 2 + x \text{Log} x) \quad (143)$$

13.1.1 Divergence de Csiszär fondée sur l'entropie de Newton.

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

$$C_N(p\|q) = \frac{1}{2} \sum_i \left[p_i^2 q_i^{-1} - 3p_i + 2q_i + p_i \text{Log} \left(\frac{p_i}{q_i} \right) \right] \quad (144)$$

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

$$-\frac{\partial C_N(p\|q)}{\partial q_j} = \frac{p_j}{q_j} \left[\frac{1}{2} \frac{p_j}{q_j} + \frac{1}{2} \right] - 1 \equiv R_j - S_j \quad (145)$$

Avec $R_j > 0$ et $S_j > 0$.

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

13.1.2 Divergence de Csiszär duale fondée sur l'entropie de Newton.

Elle se déduit de l'expression (144) en effectuant la permutation $p \leftrightarrow q$; elle s'écrit :

$$C_N(q\|p) = \frac{1}{2} \sum_i \left[q_i^2 p_i^{-1} - 3q_i + 2p_i + q_i \text{Log} \left(\frac{q_i}{p_i} \right) \right] \quad (146)$$

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

$$-\frac{\partial C_N(q\|p)}{\partial q_j} = \frac{1}{2} \log \frac{p_j}{q_j} + 1 - \frac{q_j}{p_j} \quad (147)$$

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

*** Remarque :**

Comme on le verra dans la section dédiée à l'algorithmique, l'écriture d'un algorithme multiplicatif implique la décomposition (valable pour toutes les composantes) de l'opposé du gradient sous forme d'une différence de 2 termes positifs. Cela ne posait aucun problème pour toutes les divergences proposées dans les sections précédentes.

En revanche, dans le cas présent, la difficulté tient au fait qu'une telle décomposition, même si elle est toujours possible, varie d'une composante à l'autre et ne peut pas s'écrire de immédiatement façon simple.

En effet, en écrivant :

$$-\frac{\partial C_N(q||p)}{\partial q_j} = R_j - S_j \quad (148)$$

* Si $\log \frac{p_i}{q_i} > 0 \Leftrightarrow p_j > q_j$, la décomposition s'écrit :

$$R_j = \frac{1}{2} \log \frac{p_j}{q_j} + 1 > 0 \quad ; \quad S_j = \frac{q_j}{p_j} > 0 \quad (149)$$

* Si $\log \frac{p_i}{q_i} < 0 \Leftrightarrow p_j < q_j$, la décomposition s'écrit :

$$R_j = 1 > 0 \quad ; \quad S_j = -\frac{1}{2} \log \frac{p_j}{q_j} + \frac{q_j}{p_j} > 0 \quad (150)$$

Afin d'avoir une décomposition unique, valable pour toutes les composantes de l'opposé du gradient, on écrira :

$$R_j = \frac{\frac{1}{2} \log \frac{p_j}{q_j} + \left| \frac{1}{2} \log \frac{p_j}{q_j} \right|}{2} + 1 > 0 \quad ; \quad S_j = \frac{-\frac{1}{2} \log \frac{p_j}{q_j} + \left| \frac{1}{2} \log \frac{p_j}{q_j} \right|}{2} + \frac{q_j}{p_j} > 0 \quad (151)$$

13.1.3 Divergence de Bregman fondée sur 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 :

$$B_N(p||q) = \frac{1}{2} \sum_i [p_i^2 - p_i + p_i \log p_i + q_i^2 + q_i - 2p_i q_i - p_i \log q_i] \quad (152)$$

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

$$-\frac{\partial B_N(p||q)}{\partial q_j} = \left(\frac{p_j}{q_j} - 1 \right) \left(\frac{1}{2} + q_j \right) \quad (153)$$

Ce qui peut s'écrire :

$$-\frac{\partial B_N(p||q)}{\partial q_j} = \left[\frac{1}{2} \left(\frac{p_j}{q_j} \right) + p_j \right] - \left[\frac{1}{2} + q_j \right] \equiv R_j - S_j \quad (154)$$

Avec $R_j > 0$ et $S_j > 0$.

13.1.4 Divergence de Bregman duale fondée sur l'entropie de Newton.

Elle est déduite de la relation (152) en effectuant la permutation $p \leftrightarrow q$; elle s'écrit ;

$$B_N(q\|p) = \frac{1}{2} \sum_i [q_i^2 - q_i + q_i \log q_i + p_i^2 + p_i - 2q_i p_i - q_i \log p_i] \quad (155)$$

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

$$-\frac{\partial B_N(q\|p)}{\partial q_j} = q_j \left(\frac{p_j}{q_j} - 1 \right) + \frac{1}{2} \log \frac{p_j}{q_j} \quad (156)$$

Ce qui peut aussi s'écrire :

$$-\frac{\partial B_N(q\|p)}{\partial q_j} = \left(p_j + \frac{1}{2} \log p_j \right) - \left(q_j + \frac{1}{2} \log q_j \right) \quad (157)$$

Le problème qui s'est posé dans la section (13.1.2) se pose à nouveau, en effet, l'expression (156) ne peut pas être écrite de façon simple comme une différence de 2 termes positifs valable pour toutes les composantes.

En effet, en écrivant :

$$-\frac{\partial C_N(q\|p)}{\partial q_j} = R_j - S_j \quad (158)$$

* Si $\log \frac{p_j}{q_j} > 0 \Leftrightarrow p_j > q_j$, la décomposition s'écrit :

$$R_j = \frac{1}{2} \log \frac{p_j}{q_j} + p_j > 0 \quad ; \quad S_j = q_j > 0 \quad (159)$$

* Si $\log \frac{p_j}{q_j} < 0 \Leftrightarrow p_j < q_j$, la décomposition s'écrit :

$$R_j = p_j > 0 \quad ; \quad S_j = -\frac{1}{2} \log \frac{p_j}{q_j} + q_j > 0 \quad (160)$$

Afin d'avoir une décomposition unique, valable pour toutes les composantes de l'opposé du gradient, on écrira :

$$R_j = \frac{\frac{1}{2} \log \frac{p_j}{q_j} + \left| \frac{1}{2} \log \frac{p_j}{q_j} \right|}{2} + p_j > 0 \quad ; \quad S_j = \frac{-\frac{1}{2} \log \frac{p_j}{q_j} + \left| \frac{1}{2} \log \frac{p_j}{q_j} \right|}{2} + q_j > 0 \quad (161)$$

13.1.5 Divergence de Jensen fondée sur l'entropie de Newton.

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

$$\begin{aligned} J_N(p\|q) = & \sum_i \alpha(1-\alpha)(p_i - q_i)^2 + \sum_i \alpha p_i \log \frac{p_i}{\alpha p_i + (1-\alpha)q_i} \\ & + \sum_i (1-\alpha)q_i \log \frac{q_i}{\alpha p_i + (1-\alpha)q_i} \end{aligned} \quad (162)$$

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

$$-\frac{\partial J_N(p||q)}{\partial q_j} = (1 - \alpha) \left[2\alpha (p_j - q_j) - \log \frac{q_j}{\alpha p_j + (1 - \alpha) q_j} \right] \quad (163)$$

Comme dans les sections (13.1.2) et (13.1.4), cette expression ne peut pas être écrite de façon simple comme une différence de 2 termes positifs valable pour toutes les composantes.

En effet, en écrivant :

$$-\frac{\partial C_N(q||p)}{\partial q_j} = R_j - S_j \quad (164)$$

* Si $\log \frac{q_j}{\alpha p_j + (1 - \alpha) q_j} > 0$, la décomposition s'écrit :

$$R_j = 2\alpha (1 - \alpha) p_j > 0 ; \quad S_j = (1 - \alpha) \left(2\alpha q_j + \log \frac{q_j}{\alpha p_j + (1 - \alpha) q_j} \right) > 0 \quad (165)$$

* Si $\log \frac{q_j}{\alpha p_j + (1 - \alpha) q_j} < 0$, la décomposition s'écrit :

$$R_j = (1 - \alpha) \left(2\alpha p_j - \log \frac{q_j}{\alpha p_j + (1 - \alpha) q_j} \right) > 0 ; \quad S_j = 2\alpha (1 - \alpha) q_j > 0 \quad (166)$$

Afin d'avoir une décomposition unique, valable pour toutes les composantes de l'opposé du gradient, on écrira :

$$\begin{aligned} R_j &= (1 - \alpha) \left(2\alpha p_j - \frac{1}{2} \log \frac{q_j}{\alpha p_j + (1 - \alpha) q_j} + \frac{1}{2} \left| \log \frac{q_j}{\alpha p_j + (1 - \alpha) q_j} \right| \right) > 0 \\ S_j &= (1 - \alpha) \left(2\alpha q_j + \frac{1}{2} \log \frac{q_j}{\alpha p_j + (1 - \alpha) q_j} + \frac{1}{2} \left| \log \frac{q_j}{\alpha p_j + (1 - \alpha) q_j} \right| \right) > 0 \end{aligned} \quad (167)$$

13.1.6 Divergence de Jensen duale fondée sur l'entropie de Newton.

L'analyse de cette divergence est superflue dans la mesure où elle peut être obtenue par variation du paramètre “ α ” dans la relation (162).

13.2 Autre expression du Logarithme déformé [9].

Il s'écrit :

$$\text{Log}_\alpha(x) = \frac{1}{\alpha} \frac{x^\alpha - x^{-\alpha}}{x^\alpha + x^{-\alpha}} \quad (168)$$

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 $\alpha \rightarrow 0$.

En s'appuyant sur la relation (5) on obtient pour l'entropie, l'expression spécifique :

$$S_\alpha(p) = - \sum_i p_i \text{Log}_\alpha(p_i) = - \frac{1}{\alpha} \sum_i \frac{p_i^{\alpha+1} - p_i^{-\alpha+1}}{p_i^\alpha + p_i^{-\alpha}} \quad (169)$$

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

$$f(x) = \frac{1}{\alpha} \frac{x^{\alpha+1} - x^{-\alpha+1}}{x^\alpha + x^{-\alpha}} \quad (170)$$

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 :

$$f_c(x) = \frac{1}{\alpha} \frac{x^{\alpha+1} - x^{-\alpha+1}}{x^\alpha + x^{-\alpha}} - x + 1 \quad (171)$$

13.2.1 Divergence de Csiszär correspondante.

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

$$C_\alpha(p||q) = \frac{1}{\alpha} \sum_i \frac{p_i^{1+\alpha} q_i^{-\alpha} - p_i^{1-\alpha} q_i^\alpha}{p_i^\alpha q_i^{-\alpha} + p_i^{-\alpha} q_i^\alpha} - \sum_i p_i + \sum_i q_i \quad (172)$$

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

$$-\frac{\partial C_\alpha(p||q)}{\partial q_j} = \frac{p_j}{q_j} \left[\frac{4}{(p_j^\alpha q_j^{-\alpha} + p_j^{-\alpha} q_j^\alpha)^2} \right] - 1 \quad (173)$$

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

13.2.2 Divergence de Csiszär duale correspondante.

Elle se déduit de l'expression (172) en effectuant la permutation $p \leftrightarrow q$; elle s'écrit :

$$C_\alpha(q||p) = \frac{1}{\alpha} \sum_i \frac{q_i^{1+\alpha} p_i^{-\alpha} - q_i^{1-\alpha} p_i^\alpha}{q_i^\alpha p_i^{-\alpha} + q_i^{-\alpha} p_i^\alpha} - \sum_i q_i + \sum_i p_i \quad (174)$$

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

$$-\frac{\partial C_\alpha(q||p)}{\partial q_j} = \frac{1}{\alpha} \left(\frac{p_j^{2\alpha} - q_j^{2\alpha}}{p_j^{2\alpha} + q_j^{2\alpha}} \right) + \left(\frac{p_j^{2\alpha} - q_j^{2\alpha}}{p_j^{2\alpha} + q_j^{2\alpha}} \right)^2 \quad (175)$$

14 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 (108) pour les divergences de Csiszär, (120) pour les divergences de Bregman et (132) pour les

divergences de Jensen.

Les divergences spécifiques des diverses entropies analysées dans les paragraphes 5 à 10 seront vues comme des cas particuliers.

Cependant, les divergences de Csiszär et de Bregman 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 ; en revanche, tel n'est pas le cas pour les divergences de Jensen.

On rappelle d'abord quelques propriétés utiles des facteurs d'invariance qui ont été exposées longuement dans les références [11], [12].

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

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

$$\sum_i \frac{\partial D(p_i||Kq_i)}{\partial K} = 0 \quad (176)$$

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 :

$$K(p, q) + \sum_j q_j \frac{\partial K(p, q)}{\partial q_j} = 0 \quad (177)$$

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 :

$$K(p, q) = \left(\frac{\sum_i p_i^\alpha q_i^\beta}{\sum_i p_i^\delta q_i^\gamma} \right)^\mu \quad (178)$$

Avec :

$$\alpha + \beta = \gamma + \delta \quad (179)$$

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

$$\mu(\gamma - \beta) = 1 \quad (180)$$

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 :

$$\sum_j q_j \frac{\partial D(p\|K(p, q).q)}{\partial q_j} \quad (181)$$

14.2 Divergences de Csiszär invariantes

14.2.1 Facteur d'invariance.

On rappelle que la divergence de base a été écrite (108) :

$$C_{ab}(p\|q) = \frac{1}{a-b} \left[\sum_i p_i^a q_i^{1-a} - \sum_i p_i^b q_i^{1-b} \right] - \sum_i p_i + \sum_i q_i \quad (182)$$

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

$$C_{ab}(p\|Kq) = \frac{1}{a-b} \left[\sum_i K^{1-a} p_i^a q_i^{1-a} - \sum_i K^{1-b} p_i^b q_i^{1-b} \right] - \sum_i p_i + \sum_i K q_i \quad (183)$$

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

$$\frac{\partial C_{ab}(p\|Kq)}{\partial K} = 0 \quad (184)$$

Cette équation s'écrit dans le cas présent :

$$\frac{1-a}{a-b} K^{-a} \sum_i p_i^a q_i^{1-a} - \frac{1-b}{a-b} K^{-b} \sum_i p_i^b q_i^{1-b} + \sum_i q_i = 0 \quad (185)$$

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 :

$$K_0 = \left(\frac{\sum_i p_i^t q_i^{1-t}}{\sum_i q_i} \right)^{\frac{1}{t}} \quad (186)$$

* 2 - Cas des divergences fondées sur l'entropie de Shannon (divergence de Kullback-Leibler) qui s'obtient par exemple à partir de (186) en faisant simplement $t = 1$, ce qui conduit à :

$$K = \frac{\sum_i p_i}{\sum_i q_i} \quad (187)$$

En conséquence, pour toutes les divergences considérées, on utilisera en guise de facteur d'invariance l'expression (187) (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 (186).

14.2.2 Divergences de Csiszär invariantes - Cas général.

En utilisant le facteur d'invariance (187), la divergence (183) s'écrit tous calculs faits, avec les notations : $\bar{p}_i = \frac{p_i}{\sum_j p_j}$ et $\bar{q}_i = \frac{q_i}{\sum_j q_j}$:

$$CI_{ab}(p||q) = \frac{\sum_j p_j}{a-b} \left[\sum_i \bar{p}_i^a \bar{q}_i^{1-a} - \sum_i \bar{p}_i^b \bar{q}_i^{1-b} \right] \quad (188)$$

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

$$-\frac{\partial CI_{ab}(p||q)}{\partial q_j} = \frac{\sum_j p_j}{\sum_j q_j} \left[\frac{a-1}{a-b} \left(\frac{\bar{p}_j}{\bar{q}_j} \right)^a + \frac{1-b}{a-b} \left(\frac{\bar{p}_j}{\bar{q}_j} \right)^b - \frac{a-1}{a-b} \sum_i \bar{q}_i \left(\frac{\bar{p}_i}{\bar{q}_i} \right)^a - \frac{1-b}{a-b} \sum_i \bar{q}_i \left(\frac{\bar{p}_i}{\bar{q}_i} \right)^b \right] \quad (189)$$

Cette expression peut s'écrire :

$$\frac{\partial CI_{ab}(p||q)}{\partial q_j} = \frac{\sum_j p_j}{\sum_j q_j} (R_j - S_j) \quad (190)$$

Avec :

$$R_j = \frac{a-1}{a-b} \left(\frac{\bar{p}_j}{\bar{q}_j} \right)^a + \frac{1-b}{a-b} \left(\frac{\bar{p}_j}{\bar{q}_j} \right)^b \quad (191)$$

et :

$$S_j = \frac{a-1}{a-b} \sum_i \bar{q}_i \left(\frac{\bar{p}_i}{\bar{q}_i} \right)^a + \frac{1-b}{a-b} \sum_i \bar{q}_i \left(\frac{\bar{p}_i}{\bar{q}_i} \right)^b \quad (192)$$

Dans cette décomposition, "R_j" et "S_j" sont positifs quelles que soient les valeurs relatives de "a" et "b" comme indiqué en (102) et (103).

De cette forme générale, on déduit facilement les expressions correspondant aux divergences entropiques données dans les sections 5 à 10, en faisant les adaptations déjà précisées au début de la section (12).

Dans le cas de la divergence de Csiszär/Tsallis invariante déduite du cas général, ce qui correspond à (a = t) et (b = 1), l'expression de l'opposé du gradient se simplifie, et, par voie de conséquence les expressions de R_j et S_j.

14.2.3 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 (186).

La divergence invariante obtenue en utilisant ce facteur d'invariance spécifique s'écrit :

$$CI_T^{K_0}(p||q) = \frac{t}{1-t} \left[\sum_i p_i - \left(\frac{\sum_i p_i^t q_i^{1-t}}{\sum_i q_i} \right)^{\frac{1}{t}} \sum_i q_i \right] \quad (193)$$

Ou, plus simplement :

$$CI_T^{K_0}(p||q) = \frac{t}{1-t} \left[\sum_i p_i - K_0 \sum_i q_i \right] \quad (194)$$

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

$$-\frac{\partial CI_T^{K_0}(p||q)}{\partial q_j} = \left(\frac{\sum_i p_i^t q_i^{1-t}}{\sum_i q_i} \right)^{\frac{1}{t}-1} p_j^t q_j^{-t} - \left(\frac{\sum_i p_i^t q_i^{1-t}}{\sum_i q_i} \right)^{\frac{1}{t}} \quad (195)$$

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

$$-\frac{\partial CI_T^{K_0}(p||q)}{\partial q_j} = \frac{\sum_j p_j}{\sum_j q_j} \left(\sum_i \bar{p}_i^t \bar{q}_i^{1-t} \right)^{\frac{1}{t}-1} \bar{p}_j^t \bar{q}_j^{-t} - \frac{\sum_j p_j}{\sum_j q_j} \left(\sum_i \bar{p}_i^t \bar{q}_i^{1-t} \right)^{\frac{1}{t}} \quad (196)$$

Cette expression s'écrit de façon évidente sous la forme $R_j - S_j$ avec $R_j > 0$ et $S_j > 0$.

14.3 Divergences de Csiszär duales invariantes.

14.3.1 Facteur d'invariance.

A partir de la divergence (112) qui a été écrite :

$$C_{ab}(q||p) = \frac{1}{a-b} \left[\sum_i q_i^a p_i^{1-a} - \sum_i q_i^b p_i^{1-b} \right] - \sum_i q_i + \sum_i p_i \quad (197)$$

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

$$C_{ab}(Kq||p) = \frac{1}{a-b} \left[K^a \sum_i q_i^a p_i^{1-a} - K^b \sum_i q_i^b p_i^{1-b} \right] - K \sum_i q_i + \sum_i p_i \quad (198)$$

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

$$\frac{\partial C_{ab}(Kq||p)}{\partial K} = 0 \quad (199)$$

Cette expression de cette dérivée partielle s'écrit :

$$\frac{\partial C_{ab}(Kq||p)}{\partial K} = \frac{1}{a-b} \left[a K^{a-1} \sum_i q_i^a p_i^{1-a} - b K^{b-1} \sum_i q_i^b p_i^{1-b} \right] - \sum_i q_i \quad (200)$$

L'équation (199) 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 :

$$K_0 = \left(\frac{\sum_j q_j}{\sum_j q_j^t p_j^{1-t}} \right)^{\frac{1}{t-1}} \quad (201)$$

Le facteur d'invariance nominal correspondant à la divergence de Kullback-Leibler duale (11) se déduit de cete expression en faisant le passage à la limite $t \rightarrow 1$.

Ceci conduit à l'expression difficilement exploitable :

$$\log K_0 = \frac{\sum_i q_i \log \frac{p_i}{q_i}}{\sum_i q_i} \quad (202)$$

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

$$K = \frac{\sum_j p_j}{\sum_j q_j} \quad (203)$$

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 (201).

14.3.2 Divergences de Csiszär duales invariantes - Cas général.

A partir de (198), et en introduisant le facteur d'invariance (203), cette divergence s'écrit, avec les variables réduites $\bar{q}_i = \frac{q_i}{\sum_j q_j}$ et $\bar{p}_i = \frac{p_i}{\sum_j p_j}$:

$$CI_{ab}(q||p) = \frac{\sum_j p_j}{a-b} \left(\sum_i \bar{q}_i^a \bar{p}_i^{1-a} - \sum_i \bar{q}_i^b \bar{p}_i^{1-b} \right) \quad (204)$$

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

$$-\frac{\partial CI_{ab}(q||p)}{\partial q_j} = \frac{1}{a-b} \frac{\sum_j p_j}{\sum_j q_j} \left[a \sum_i \bar{q}_i \left(\frac{\bar{p}_i}{\bar{q}_i} \right)^{1-a} - a \left(\frac{\bar{p}_j}{\bar{q}_j} \right)^{1-a} - b \sum_i \bar{q}_i \left(\frac{\bar{p}_i}{\bar{q}_i} \right)^{1-b} + b \left(\frac{\bar{p}_j}{\bar{q}_j} \right)^{1-b} \right] \quad (205)$$

Cette expression peut s'écrire :

$$-\frac{\partial CI_{ab}(q||p)}{\partial q_j} = \frac{1}{a-b} \frac{\sum_j p_j}{\sum_j q_j} (A_j - B_j) \quad (206)$$

Avec :

$$A_j = a \sum_i \bar{q}_i \left(\frac{\bar{p}_i}{\bar{q}_i} \right)^{1-a} + b \left(\frac{\bar{p}_j}{\bar{q}_j} \right)^{1-b} > 0 \quad (207)$$

et :

$$B_j = a \left(\frac{\bar{p}_j}{\bar{q}_j} \right)^{1-a} + b \sum_i \bar{q}_i \left(\frac{\bar{p}_i}{\bar{q}_i} \right)^{1-b} > 0 \quad (208)$$

On observera que pour la divergence de Csiszär/Tsallis duale invariante qui correspond à $(a = t)$ et $(b = 1)$, l'expression de l'opposé du gradient se simplifie, et, par voie de conséquence les expressions de A_j et B_j .

Bien entendu, la décomposition de l'opposé du gradient en une différence de 2 termes positifs dépendra du signe de $(a - b)$.

Ce point peut être précisé comme suit :

* **Si** $a - b > 0$, on a :

$$R_j = \frac{\sum_j p_j}{\sum_j q_j} \frac{A_j}{a - b} ; \quad S_j = \frac{\sum_j p_j}{\sum_j q_j} \frac{B_j}{a - b} \quad (209)$$

* **Si** $a - b < 0$, on a :

$$R_j = \frac{\sum_j p_j}{\sum_j q_j} \frac{B_j}{b - a} ; \quad S_j = \frac{\sum_j p_j}{\sum_j q_j} \frac{A_j}{b - a} \quad (210)$$

On peut globaliser cela en écrivant :

$$R_j = \frac{\sum_j p_j}{2 \sum_j q_j} \left\{ \left[\frac{A_j}{a - b} + \frac{A_j}{|a - b|} \right] + \left[\frac{B_j}{b - a} + \frac{B_j}{|b - a|} \right] \right\} > 0 \quad (211)$$

$$S_j = \frac{\sum_j p_j}{2 \sum_j q_j} \left\{ \left[\frac{B_j}{a - b} + \frac{B_j}{|a - b|} \right] + \left[\frac{A_j}{b - a} + \frac{A_j}{|b - a|} \right] \right\} > 0 \quad (212)$$

Remarque : Le cas spécifique de la divergence de Csiszär/Shannon duale invariante (KL duale invariante) est obtenu en faisant le passage à la limite $a \rightarrow 1$; $b \rightarrow 1$ en respectant (102) ou (103) c'est à dire avec par exemple $a = 1 - \epsilon$; $b = 1 + \epsilon$.

Ceci conduit à une expression de l'opposé du gradient qui s'écrit :

$$-\frac{\partial KLI(q||p)}{\partial q_j} = \frac{\sum_j p_j}{\sum_j q_j} (R_j - S_j) \quad (213)$$

Avec :

$$R_j = -\log \bar{q}_j - \sum_i \bar{q}_i \log \bar{p}_i > 0 \quad (214)$$

et :

$$S_j = -\log \bar{p}_j - \sum_i \bar{q}_i \log \bar{q}_i > 0 \quad (215)$$

On pourra donc envisager d'écrire un algorithme multiplicatif, contrairement à ce qu'on a observé pour la divergence de KL duale non-invariante (section 5.2).

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

On utilise dans ce cas, le facteur d'invariance nominal donné par (201) ; à partir de (198), on obtient avec $a = t$ et $b = 1$:

$$CI_T^{K_0}(q||p) = \sum_i p_i - \left(\frac{\sum_j q_j}{\sum_j q_j^t p_j^{1-t}} \right)^{\frac{1}{t-1}} \sum_i q_i \quad (216)$$

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

$$-\frac{\partial CI_T^{K_0}(q||p)}{\partial q_j} = \frac{t}{1-t} \left[\left(\frac{\sum_j q_j}{\sum_j q_j^t p_j^{1-t}} \right)^{\frac{t}{t-1}} \left(\frac{p_j}{q_j} \right)^{1-t} - \left(\frac{\sum_j q_j}{\sum_j q_j^t p_j^{1-t}} \right)^{\frac{1}{t-1}} \right] \quad (217)$$

En distinguant les cas $t < 1$ et $t > 1$ selon la procédure proposée dans l'Annexe 2, cette expression s'écrit de façon évidente sous la forme $R_j - S_j$ avec $R_j > 0$ et $S_j > 0$.

14.4 Divergences de Bregman invariantes.

14.4.1 Facteur d'invariance.

On rappelle que la divergence de Bregman fondée sur la forme générale de l'entropie a été écrite (120) :

$$B_{ab}(p||q) = \frac{1}{a-b} \sum_i [p_i^a - p_i^b + (a-1)q_i^a - (b-1)q_i^b - ap_i q_i^{a-1} + bp_i q_i^{b-1}] \quad (218)$$

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

$$B_{ab}(p||Kq) = \frac{1}{a-b} \sum_i [p_i^a - p_i^b + (a-1)K^a q_i^a - (b-1)K^b q_i^b - aK^{a-1} p_i q_i^{a-1} + bK^{b-1} p_i q_i^{b-1}] \quad (219)$$

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

$$\frac{\partial B_{ab}(p||Kq)}{\partial K} = 0 \quad (220)$$

Avec :

$$\frac{\partial B_{ab}(p||Kq)}{\partial K} = \frac{1}{a-b} \left[a(a-1)K^{a-1} \sum_i q_i^a - b(b-1)K^{b-1} \sum_i q_i^b - a(a-1)K^{a-2} \sum_i p_i q_i^{a-1} + b(b-1)K^{b-2} \sum_i p_i q_i^{b-1} \right] \quad (221)$$

Avec cette expression, l'équation (220) 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 :

$$K_0 = \frac{\sum_i p_i q_i^{t-1}}{\sum_i q_i^t} \quad (222)$$

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

$$K = \frac{\sum_i p_i}{\sum_i q_i} \quad (223)$$

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 (222).

14.4.2 Divergences de Bregman invariantes - Cas général.

En introduisant le facteur d'invariance (223) dans l'expression (219), on obtient, tous calculs faits :

$$BI_{ab}(p||q) = \frac{(\sum_j p_j)^a}{a-b} \left[\sum_i \bar{p}_i^a + (a-1) \sum_i \bar{q}_i^a - a \sum_i \bar{p}_i \bar{q}_i^{a-1} \right] - \frac{(\sum_j p_j)^b}{a-b} \left[\sum_i \bar{p}_i^b + (b-1) \sum_i \bar{q}_i^b - b \sum_i \bar{p}_i \bar{q}_i^{b-1} \right] \quad (224)$$

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

$$-\frac{\partial BI_{ab}(p||q)}{\partial q_j} = \frac{a(a-1)}{(a-b)} \frac{(\sum_j p_j)^a}{\sum_j q_j} \left[\bar{p}_j \bar{q}_j^{a-2} - \sum_i \bar{p}_i \bar{q}_i^{a-1} - \bar{q}_j^{a-1} + \sum_i \bar{q}_i^a \right] - \frac{b(b-1)}{(a-b)} \frac{(\sum_j p_j)^b}{\sum_j q_j} \left[\bar{p}_j \bar{q}_j^{b-2} - \sum_i \bar{p}_i \bar{q}_i^{b-1} - \bar{q}_j^{b-1} + \sum_i \bar{q}_i^b \right] \quad (225)$$

La décomposition de l'opposé du gradient en une différence de 2 termes positifs ne pose pas de problème compte tenu des conditions sur les valeurs relatives de "a" et "b" précisées en (102) et (103) ; ce point sera précisé dans la section dédiée à l'aspect algorithmique.

Comme c'était le cas pour les divergences de Csiszär, les expressions correspondant aux différentes divergences entropiques données dans les sections 5 à 10 se déduisent facilement de cette forme générale en faisant les adaptations déjà précisées en début de section (12).

On observera que pour la divergence de Bregman/Tsallis invariante déduite de la forme générale précédente, ce qui correspond à $(a = t)$ et $(b = 1)$, l'expression de l'opposé du gradient se simplifie, et, par voie de conséquence, la décomposition se simplifie.

14.4.3 Cas particulier - Divergence de Bregman/Tsallis invariante.

Dans ce cas, le facteur d'invariance nominal peut être calculé explicitement ; son expression est donnée par (222) ; la divergence invariante correspondante s'exprime sous la forme :

$$BI_T^{K_0}(p||q) = \frac{1}{1-t} \left[\left(\sum_i p_i q_i^{t-1} \right)^t \left(\sum_i q_i^t \right)^{1-t} - \sum_i p_i^t \right] \quad (226)$$

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

$$-\frac{\partial BI_T^{K_0}(p||q)}{\partial q_j} = t \left[\left(\frac{\sum_i p_i q_i^{t-1}}{\sum_i q_i^t} \right)^{t-1} p_j q_j^{t-2} - \left(\frac{\sum_i p_i q_i^{t-1}}{\sum_i q_i^t} \right)^t q_j^{t-1} \right] \quad (227)$$

Cette expression s'écrit de façon évidente sous la forme $R_j - S_j$ avec $R_j > 0$ et $S_j > 0$. Ou bien, sous forme simplifiée :

$$-\frac{\partial BI_T^{K_0}(p||q)}{\partial q_j} = t [K_0^{t-1} p_j q_j^{t-2} - K_0^t q_j^{t-1}] \quad (228)$$

Ou encore :

$$-\frac{\partial BI_T^{K_0}(p||q)}{\partial q_j} = t K_0^{t-1} q_j^{t-1} \left[\frac{p_j}{q_j} - K_0 \right] \quad (229)$$

14.5 Divergences de Bregman duales - Forme invariante.

14.5.1 Facteur d'invariance.

On reprend la forme de cette divergence écrite en (123) :

$$B_{ab}(q||p) = \frac{1}{a-b} \sum_i [q_i^a - q_i^b + (a-1)p_i^a - (b-1)p_i^b - a q_i p_i^{a-1} + b q_i p_i^{b-1}] \quad (230)$$

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

$$B_{ab}(Kq||p) = \frac{1}{a-b} \sum_i [K^a q_i^a - K^b q_i^b + (a-1)p_i^a - (b-1)p_i^b - a K q_i p_i^{a-1} + b K q_i p_i^{b-1}] \quad (231)$$

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

$$\frac{\partial B_{ab}(Kq||p)}{\partial K} = 0 \quad (232)$$

Avec :

$$\frac{\partial B_{ab}(Kq||p)}{\partial K} = \frac{1}{a-b} \left[aK^{a-1} \sum_i q_i^a - bK^{b-1} \sum_i q_i^b - a \sum_i q_i p_i^{a-1} + b \sum_i q_i p_i^{b-1} \right] \quad (233)$$

Avec cette expression, l'équation (232) 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 :

$$K_0 = \left(\frac{\sum_i q_i^t}{\sum_i q_i p_i^{t-1}} \right)^{\frac{1}{1-t}} \quad (234)$$

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

$$K = \frac{\sum_i p_i}{\sum_i q_i} \quad (235)$$

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 (234).

14.5.2 Divergences de Bregman duales invariantes - Cas général.

En introduisant le facteur d'invariance (235) dans la divergence (231), on obtient tous calculs faits :

$$BI_{ab}(q||p) = \frac{(\sum_j p_j)^a}{a-b} \left[\sum_i \bar{q}_i^a + (a-1) \sum_i \bar{p}_i^a - a \sum_i \bar{q}_i \bar{p}_i^{(a-1)} \right] - \frac{(\sum_j p_j)^b}{a-b} \left[\sum_i \bar{q}_i^b + (b-1) \sum_i \bar{p}_i^b - b \sum_i \bar{q}_i \bar{p}_i^{(b-1)} \right] \quad (236)$$

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

$$-\frac{\partial BI_{ab}(q||p)}{\partial q_j} = \frac{b (\sum_j p_j)^b}{(a-b) \sum_j q_j} \left(\bar{q}_j^{(b-1)} - \sum_i \bar{q}_i^b - \bar{p}_j^{(b-1)} + \sum_i \bar{q}_i \bar{p}_i^{(b-1)} \right) - \frac{a (\sum_j p_j)^a}{(a-b) \sum_j q_j} \left(\bar{q}_j^{(a-1)} - \sum_i \bar{q}_i^a - \bar{p}_j^{(a-1)} + \sum_i \bar{q}_i \bar{p}_i^{(a-1)} \right) \quad (237)$$

Cette expression peut s'écrire :

$$-\frac{\partial BI_{ab}(q||p)}{\partial q_j} = \frac{1}{(a-b)\sum_j q_j} (A_j - B_j) \quad (238)$$

Avec $A_j > 0$; $B_j > 0$.

En distinguant les cas $a - b > 0$ et $a - b < 0$ selon la procédure indiquée dans l'Annexe 2 on obtient la décomposition de l'opposé du gradient sous la forme $R_j - S_j$ ($R_j > 0$; $S_j > 0$) utilisable dans la section dédiée à l'algorithmique.

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.

On observera que pour la divergence de Bregman/Tsallis duale invariante déduite de la forme générale précédente, qui correspond à ($a = t$) et ($b = 1$), l'expression de l'opposé du gradient se simplifie, et, par voie de conséquence les expressions de A_j et B_j .

14.5.3 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 (234).

La divergence invariante correspondante est donnée par :

$$BI_T^{K_0}(q||p) = \sum_i p_i^t - \left(\frac{\sum_i q_i^t}{\sum_i q_i p_i^{t-1}} \right)^{\frac{t}{1-t}} \sum_i q_i^t \quad (239)$$

Ce qui peut s'écrire plus simplement :

$$BI_T^{K_0}(q||p) = \sum_i p_i^t - K_0^t \sum_i q_i^t \quad (240)$$

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

$$-\frac{\partial BI_T^{K_0}(q||p)}{\partial q_j} = \frac{t}{t-1} \left[\left(\frac{\sum_i q_i^t}{\sum_i q_i p_i^{t-1}} \right)^{\frac{1}{1-t}} p_j^{t-1} - \left(\frac{\sum_i q_i^t}{\sum_i q_i p_i^{t-1}} \right)^{\frac{t}{1-t}} q_j^{t-1} \right] \quad (241)$$

Ou bien, sous une autre forme :

$$-\frac{\partial BI_T^{K_0}(q||p)}{\partial q_j} = \frac{t}{t-1} K_0 [p_j^{t-1} - K_0^{t-1} q_j^{t-1}] \quad (242)$$

En distinguant les cas $t < 1$ et $t > 1$ selon la procédure proposée dans l'Annexe 2, cette expression s'écrit de façon évidente sous la forme $R_j - S_j$ avec $R_j > 0$ et $S_j > 0$.

14.6 Divergence de Jensen invariante.

14.6.1 Facteur d'invariance.

En introduisant le facteur d'invariance "K" dans l'expression de la divergence de Jensen (132), on obtient :

$$J_\alpha(p\|Kq) = \sum_i \alpha \frac{p_i^a - p_i^b}{a-b} + (1-\alpha) \frac{K^a q_i^a - K^b q_i^b}{a-b} - \frac{[\alpha p_i + (1-\alpha)Kq_i]^a}{a-b} + \frac{[\alpha p_i + (1-\alpha)Kq_i]^b}{a-b} \quad (243)$$

On obtient alors, tous calculs faits :

$$\frac{\partial J_\alpha(p\|Kq)}{\partial K} = \frac{1-\alpha}{a-b} \sum_i aK^{a-1} q_i^a - bK^{b-1} q_i^b - aq_i [\alpha p_i + (1-\alpha)Kq_i]^{a-1} + bq_i [\alpha p_i + (1-\alpha)Kq_i]^{b-1} \quad (244)$$

Le facteur d'invariance nominal peut être obtenu par résolution en "K" de l'équation :

$$\frac{\partial J_\alpha(p\|Kq)}{\partial K} = 0 \quad (245)$$

Dans le cas présent, cette équation n'a pas de solution explicite, on utilisera donc le facteur d'invariance :

$$K = \frac{\sum_i p_i}{\sum_i q_i} \quad (246)$$

C'est le facteur d'invariance nominal pour la divergence de Kullback-Leibler fondée sur l'entropie de Shannon qui nous sert de référence.

14.6.2 Divergence de Jensen invariante.

La divergence de Jensen invariante est obtenue, en introduisant le facteur d'invariance (246) dans l'expression (243) ; en utilisant les notations $\bar{p}_i = \frac{p_i}{\sum_j p_j}$ et $\bar{q}_i = \frac{q_i}{\sum_j q_j}$, elle s'écrit, tous calculs faits :

$$JI_\alpha(p\|q) = \frac{\left(\sum_j p_j\right)^a}{a-b} \sum_i \alpha \bar{p}_i^a + (1-\alpha) \bar{q}_i^a - [\alpha \bar{p}_i + (1-\alpha) \bar{q}_i]^a - \frac{\left(\sum_j p_j\right)^b}{a-b} \sum_i \alpha \bar{p}_i^b + (1-\alpha) \bar{q}_i^b - [\alpha \bar{p}_i + (1-\alpha) \bar{q}_i]^b \quad (247)$$

Après quelques calculs, l'opposé du gradient par rapport à "q" s'écrit, $\forall j$:

$$-\frac{\partial JI_\alpha(p\|q)}{\partial q_j} = \frac{1-\alpha}{(a-b)\sum_j q_j} (A_j - B_j) \quad (248)$$

Avec $A_j > 0$ et $B_j > 0 \forall j$, donnés par :

$$\begin{aligned}
A_j = & a \left(\sum_j p_j \right)^a \left\{ \sum_i \bar{q}_i^a + [\alpha \bar{p}_j + (1 - \alpha) \bar{q}_j]^{a-1} \right\} \\
& + b \left(\sum_j p_j \right)^b \left\{ \bar{q}_j^{b-1} + \sum_i \bar{q}_i [\alpha \bar{p}_i + (1 - \alpha) \bar{p}_i]^{b-1} \right\}
\end{aligned} \tag{249}$$

et

$$\begin{aligned}
B_j = & a \left(\sum_j p_j \right)^a \left\{ \bar{q}_j^{a-1} + \sum_i \bar{q}_i [\alpha \bar{p}_i + (1 - \alpha) \bar{q}_i]^{a-1} \right\} \\
& + b \left(\sum_j p_j \right)^b \left\{ \sum_i \bar{q}_i^b + [\alpha \bar{p}_j + (1 - \alpha) \bar{q}_j]^{b-1} \right\}
\end{aligned} \tag{250}$$

A partir de cette écriture et en tenant compte de $1 - \alpha > 0$, la décomùposition de l'opposé du gradient en une différence de 2 termes positifs dépend, bien entendu, du signe de $(a - b)$.

Cette décomposition de la forme $R_j - S_j$ ($R_j > 0; S_j > 0$) tenant compte du signe de $a - b$ sera obtenue en suivant la procédure indiquée dans l'Annexe 2.

Remarque 1 : Dans le cas particulier de la divergence de Jensen fondée sur l'entropie de Tsallis (correspondant à $a = t$ et $b = 1$), contrairement à ce qu'on a observé pour les divergences de Csiszär et de Bregman, fondées sur cette même entropie, on ne peut pas calculer d'expression spécifique du facteur d'invariance.

Remarque 2 : Les divergences de Jensen invariantes fondées sur les entropies des sections 5 à 10 s'obtiennent sans difficulté en procédant aux adaptations déjà précisées au début de la section 12.

Remarque 3 : Pour la divergence de Jensen/Tsallis invariante déduite de la forme générale précédente, qui correspond à $(a = t)$ et $(b = 1)$, l'expression de l'opposé du gradient se simplifie, et, par voie de conséquence les expressions de A_j et B_j .

14.6.3 Cas particulier : Divergence de Jensen/Shannon invariante.

Compte tenu des difficultés qui sont apparues dans le cas de la divergence de Jensen/Shannon non-invariantes (Section (6.4)), on développe ici le cas particulier de la divergence de Jensen/Shannon invariante.

Deux voies sont possibles :

* soit on part de la divergence de Jensen invariante développée pour l'entropie

générale de la section (14.6.2) et on effectue le passage à la limite ($\epsilon \rightarrow 0$), avec (par exemple) $a = 1 + \epsilon$ et $b = 1 - \epsilon$; cette méthode donne lieu à des calculs qui, défaut d'être compliqués, sont extrêmement lourds

* soit on effectue le calcul directement à partir de la divergence de Jensen/Shannon de la section (6.4).

Les deux méthodes conduisent bien sur à des résultats identiques.

Dans la mesure ou pour cette divergence, le facteur d'invariance nominal n'a pas de forme explicite, on utilise en guise de facteur d'invariance l'expression

$K = \frac{\sum_j p_j}{\sum_j q_j}$ déjà utilisé par ailleurs.

La divergence de Jensen/Shannon invariante s'écrit :

$$JSI(p||q) = \sum_j p_j \left\{ \sum_i \alpha \bar{p}_i \log \bar{p}_i + \sum_i (1 - \alpha) \bar{q}_i \log \bar{q}_i - \sum_i [\alpha \bar{p}_i + (1 - \alpha) \bar{q}_i] \log [\alpha \bar{p}_i + (1 - \alpha) \bar{q}_i] \right\} \quad (251)$$

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

$$-\frac{\partial JSI(p||q)}{\partial q_j} = (1 - \alpha) \frac{\sum_j p_j}{\sum_j q_j} \left\{ \log \left[\frac{\alpha \bar{p}_j + (1 - \alpha) \bar{q}_j}{\bar{q}_j} \right] - \sum_i \bar{q}_i \log \left[\frac{\alpha \bar{p}_i + (1 - \alpha) \bar{q}_i}{\bar{q}_i} \right] \right\} \quad (252)$$

Contrairement à la divergence de Jensen/Shannon non-invariante, la décomposition de l'opposé du gradient sous la forme $R_j - S_j$ avec $R_j > 0$ et $S_j > 0$ ne pose aucun problème, et on aura :

$$R_j = (1 - \alpha) \frac{\sum_j p_j}{\sum_j q_j} \left\{ -\log \bar{q}_j - \sum_i \bar{q}_i \log [\alpha \bar{p}_i + (1 - \alpha) \bar{q}_i] \right\} \quad (253)$$

et

$$S_j = (1 - \alpha) \frac{\sum_j p_j}{\sum_j q_j} \left\{ -\log [\alpha \bar{p}_j + (1 - \alpha) \bar{q}_j] - \sum_i \bar{q}_i \log \bar{q}_i \right\} \quad (254)$$

Cette décomposition pourra être utilisée dans la section dédiée à l'algorithmique.

14.7 Divergence de Jensen duale invariante.

Comme dans le cas de la divergence de Jensen invariante, il n'est pas possible d'obtenir une expression explicite du facteur d'invariance nominal; on utilisera donc le facteur d'invariance $K = \frac{\sum_j p_j}{\sum_j q_j}$.

Par ailleurs, l'analyse de la divergence de Jensen duale invariante est sans intérêt dans la mesure ou on peut l'obtenir à partir de la divergence de la section précédente en modifiant la valeur du paramètre " α "; on peut dire que la divergence de Jensen invariante est "*self duale*".

15 Entropie de Newton - Formes invariantes des divergences.

Les divergences fondées sur l'entropie de Newton constituent des cas particuliers qui n'entrent pas dans le contexte des divergences déduites de l'entropie générale analysées dans la section précédente.

Les divergences non-invariantes fondées sur l'entropie de Newton ont été données dans la section (13.1) portant sur le Logarithme déformé de Newton et l'entropie associée.

En conséquence, nous analysons ici les formes invariantes de ces divergences.

15.1 Divergences de Csiszär invariantes fondées sur l'entropie de Newton.

Le calcul du facteur d'invariance nominal conduit à une expression difficile à exploiter, en conséquence, on utilise pour le facteur d'invariance, l'expression $K = \frac{\sum_j p_j}{\sum_j q_j}$; à partir de (144), en utilisant les notations $\bar{p}_i = \frac{p_i}{\sum_j p_j}$ et $\bar{q}_i = \frac{q_i}{\sum_j q_j}$, on obtient l'expression de la divergence invariante :

$$CI_N(p||q) = \frac{1}{2} \sum_j p_j \left[\sum_i \bar{p}_i \frac{\bar{p}_i}{\bar{q}_i} - 1 + \sum_i \bar{p}_i \log \frac{\bar{p}_i}{\bar{q}_i} \right] \quad (255)$$

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

$$-\frac{\partial CI_N(p||q)}{\partial q_j} = \frac{1}{2} \frac{\sum_l p_l}{\sum_l q_l} \left(\frac{\bar{p}_j}{\bar{q}_j} + \left(\frac{\bar{p}_j}{\bar{q}_j} \right)^2 \right) - \frac{1}{2} \frac{\sum_l p_l}{\sum_l q_l} \left(\sum_i \bar{p}_i \frac{\bar{p}_i}{\bar{q}_i} + 1 \right) \quad (256)$$

Cette expression s'écrit de façon évidente sous la forme $R_j - S_j$ avec $R_j > 0$ et $S_j > 0$.

15.2 Divergences de Csiszär/Newton duales invariantes .

Le calcul du facteur d'invariance nominal conduit à une expression difficile à exploiter, en conséquence, on utilise pour le facteur d'invariance, l'expression $K = \frac{\sum_j p_j}{\sum_j q_j}$; à partir de (146), en utilisant les notations $\bar{p}_i = \frac{p_i}{\sum_j p_j}$ et $\bar{q}_i = \frac{q_i}{\sum_j q_j}$, on obtient l'expression de la divergence invariante :

$$CI_N(q||p) = \frac{1}{2} \sum_j p_j \left[\sum_i \bar{q}_i \frac{\bar{q}_i}{\bar{p}_i} - 1 + \sum_i \bar{q}_i \log \frac{\bar{q}_i}{\bar{p}_i} \right] \quad (257)$$

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

$$-\frac{\partial CI_N(q||p)}{\partial q_j} = \frac{1}{2} \frac{\sum_l p_l}{\sum_l q_l} \left[\sum_i \frac{\bar{q}_i^2}{\bar{p}_i} - \frac{\bar{q}_j}{\bar{p}_j} + \sum_i \bar{q}_i \log \frac{\bar{q}_i}{\bar{p}_i} - \log \frac{\bar{q}_j}{\bar{p}_j} \right] \quad (258)$$

Cette expression peut aussi s'écrire sous une forme plus adaptée à l'usage qu'on en fera dans la section dédiée à l'algorithmique :

$$-\frac{\partial CI_N(q||p)}{\partial q_j} = \frac{1}{2} \frac{\sum_l p_l}{\sum_l q_l} [R_j - S_j] \quad (259)$$

Avec :

$$R_j = \sum_i \bar{q}_i \frac{\bar{q}_i}{\bar{p}_i} - \sum_i \bar{q}_i \log \bar{p}_i - \log \bar{q}_j > 0 \quad (260)$$

et :

$$S_j = \frac{\bar{q}_j}{\bar{p}_j} - \sum_i \bar{q}_i \log \bar{q}_i - \log \bar{p}_j > 0 \quad (261)$$

15.3 Divergences de Bregman invariantes fondées sur l'entropie de Newton.

Le calcul du facteur d'invariance nominal conduit à une expression difficile à exploiter, en conséquence, on utilise pour le facteur d'invariance, l'expression $K = \frac{\sum_j p_j}{\sum_j q_j}$; à partir de (152), en utilisant les notations $\bar{p}_i = \frac{p_i}{\sum_j p_j}$ et $\bar{q}_i = \frac{q_i}{\sum_j q_j}$, on obtient l'expression de la divergence invariante :

$$BI_N(p||q) = \frac{1}{2} \left(\sum_j p_j \right)^2 \sum_i (\bar{p}_i - \bar{q}_i)^2 + \frac{1}{2} \sum_j p_j \sum_i \bar{p}_i \log \frac{\bar{p}_i}{\bar{q}_i} \quad (262)$$

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

$$\begin{aligned} -\frac{\partial BI_N(p||q)}{\partial q_j} &= \frac{\left(\sum_j p_j \right)^2}{\sum_j q_j} \left[(\bar{p}_j - \bar{q}_j) - \sum_i \bar{q}_i (\bar{p}_i - \bar{q}_i) \right] \\ &\quad + \frac{1}{2} \frac{\sum_j p_j}{\sum_j q_j} \left[1 - \frac{\bar{p}_j}{\bar{q}_j} \right] \end{aligned} \quad (263)$$

Ce qui peut s'écrire :

$$-\frac{\partial BI_N(p||q)}{\partial q_j} = R_j - S_j \quad (264)$$

Avec :

$$R_j = \frac{(\sum_l p_l)^2}{\sum_l q_l} \left(\bar{p}_j + \sum_i \bar{q}_i^2 \right) + \frac{1}{2} \frac{\sum_l p_l}{\sum_l q_l} > 0 \quad (265)$$

et :

$$S_j = \frac{(\sum_l p_l)^2}{\sum_l q_l} \left(\bar{q}_j + \sum_i \bar{p}_i \bar{q}_i \right) + \frac{1}{2} \frac{\sum_l p_l}{\sum_l q_l} \left(\frac{\bar{p}_j}{\bar{q}_j} \right) > 0 \quad (266)$$

15.4 Divergences de Bregman/Newton duale invariante.

Le calcul du facteur d'invariance nominal conduit à une expression difficile à exploiter, en conséquence, on utilise pour le facteur d'invariance, l'expression $K = \frac{\sum_j p_j}{\sum_j q_j}$; à partir de (155), en utilisant les notations $\bar{p}_i = \frac{p_i}{\sum_j p_j}$ et $\bar{q}_i = \frac{q_i}{\sum_j q_j}$, on obtient l'expression de la divergence invariante :

$$BI_N(q||p) = \frac{1}{2} \left(\sum_j p_j \right)^2 \sum_i (\bar{q}_i - \bar{p}_i)^2 + \frac{1}{2} \sum_j p_j \sum_i \bar{q}_i \log \frac{\bar{q}_i}{\bar{p}_i} \quad (267)$$

Bien entendu, cette divergence est la duale de (262).

L'opposé du gradient par rapport à "q" s'écrit $\forall j$ sous la forme synthétique suivante :

$$-\frac{\partial BI_N(q||p)}{\partial q_j} = \frac{1}{2} \frac{\sum_l p_l}{\sum_l q_l} [R_j - S_j] \quad (268)$$

Avec :

$$R_j = \sum_l p_l \left(2\bar{p}_j + 2 \sum_i \bar{q}_i^2 \right) - \log \bar{q}_j - \sum_i \bar{q}_i \log \bar{p}_i > 0 \quad (269)$$

et :

$$S_j = \sum_l p_l \left(2\bar{q}_j + 2 \sum_i \bar{p}_i \bar{q}_i \right) - \log \bar{p}_j - \sum_i \bar{q}_i \log \bar{q}_i > 0 \quad (270)$$

15.5 Divergences de Jensen invariantes fondées sur l'entropie de Newton.

Le calcul du facteur d'invariance nominal ne permet pas d'obtenir une expression explicite de ce facteur, en conséquence, on utilise pour le facteur d'invariance, l'expression $K = \frac{\sum_j p_j}{\sum_j q_j}$; à partir de (162), en utilisant les notations $\bar{p}_i = \frac{p_i}{\sum_j p_j}$ et $\bar{q}_i = \frac{q_i}{\sum_j q_j}$, on obtient l'expression de la divergence invariante :

$$\begin{aligned} JI_N(p||q) = & \alpha(1-\alpha) \left(\sum_j p_j \right)^2 \sum_i (\bar{p}_i - \bar{q}_i)^2 \\ & + \alpha \left(\sum_j p_j \right) \sum_i \bar{p}_i \log \frac{\bar{p}_i}{\alpha \bar{p}_i + (1-\alpha) \bar{q}_i} \\ & + (1-\alpha) \left(\sum_j p_j \right) \sum_i \bar{q}_i \log \frac{\bar{q}_i}{\alpha \bar{p}_i + (1-\alpha) \bar{q}_i} \end{aligned} \quad (271)$$

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

$$\begin{aligned}
-\frac{\partial JI_N(p||q)}{\partial q_j} = & 2\alpha(1-\alpha) \frac{\left(\sum_j p_j\right)^2}{\sum_j q_j} \left[(\bar{p}_j - \bar{q}_j) - \sum_i \bar{q}_i (\bar{p}_i - \bar{q}_i) \right] \\
& + (1-\alpha) \frac{\left(\sum_j p_j\right)}{\sum_j q_j} \sum_i \bar{q}_i \log \frac{\bar{q}_i}{\alpha \bar{p}_i + (1-\alpha) \bar{q}_i} \\
& - (1-\alpha) \frac{\left(\sum_j p_j\right)}{\sum_j q_j} \log \frac{\bar{q}_j}{\alpha \bar{p}_j + (1-\alpha) \bar{q}_j}
\end{aligned} \tag{272}$$

En vue de son utilisation dans la section Algorithmique, cette expression peut s'écrire sous la forme :

$$-\frac{\partial JI_N(p||q)}{\partial q_j} = R_j - S_j \tag{273}$$

Avec :

$$\begin{aligned}
R_j = & 2\alpha(1-\alpha) \frac{\left(\sum_j p_j\right)^2}{\sum_j q_j} \left[\sum_i \bar{q}_i^2 + \bar{p}_j \right] \\
& - (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j} \left\{ \sum_i \bar{q}_i \log [\alpha \bar{p}_i + (1-\alpha) \bar{q}_i] + \log \bar{q}_j \right\} > 0
\end{aligned} \tag{274}$$

et :

$$\begin{aligned}
S_j = & 2\alpha(1-\alpha) \frac{\left(\sum_j p_j\right)^2}{\sum_j q_j} \left[\sum_i \bar{p}_i \bar{q}_i + \bar{q}_j \right] \\
& - (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j} \left\{ \sum_i \bar{q}_i \log \bar{q}_i + \log [\alpha \bar{p}_j + (1-\alpha) \bar{q}_j] \right\} > 0
\end{aligned} \tag{275}$$

16 Algorithmique.

16.1 Généralités sur la méthode de construction des algorithmes.

On rappelle que dans le cas qui nous intéresse, on a $p = y$, $q = Hx$ donc $q_j^k = (Hx^k)_j$.

Alors, pour une divergence $D(p||q)$ ou $D(q||p)$, on a :

$$\frac{\partial D}{\partial x} = H^T \frac{\partial D}{\partial q} \tag{276}$$

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 :

$$x_j^{k+1} = x_j^k + \alpha_j^k x_j^k \left(-\frac{\partial D}{\partial x} \right)_j^k \quad (277)$$

Cette écriture est toujours possible.

La divergence D étant convexe par rapport à " x ", $(-\frac{\partial D}{\partial x})$ est une direction de descente. Un tel algorithme converge pour autant que le pas de descente est convenablement calculé.

A notre connaissance, à la seule exception notable de la divergence de Kullback-Leibler duale, l'opposé du gradient peut toujours s'écrire sous la forme :

$$\left(-\frac{\partial D}{\partial x} \right)_j^k = U_j^k - V_j^k ; \quad U_j^k > 0 ; \quad V_j^k > 0 \quad (278)$$

Dans certains cas (très rares), cette décomposition demande quelque réflexion. D'où l'écriture presque toujours possible :

$$x_j^{k+1} = x_j^k + \alpha_j^k x_j^k (U_j^k - V_j^k) \quad (279)$$

Puis, en préconditionnant lorsque c'est possible, par $\frac{1}{V_j^k} > 0$, on obtient la forme pseudo-multiplicative :

$$x_j^{k+1} = x_j^k + \alpha_j^k x_j^k \left(\frac{U_j^k}{V_j^k} - 1 \right) \quad (280)$$

Compte tenu de $V_j^k > 0 \quad \forall j$, l'opposé du gradient modifié reste une direction de descente.

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 $[0, (\alpha^k)_{Max}]$. Cette procédure est analysée en détails dans [11] et [12].

On pourra écrire de façon générale, quelle que soit l'expression de l'opposé du gradient, avec un pas de descente indépendant de la composante :

$$x^{k+1} = x^k + \alpha^k x^k (U^k - V^k) \quad (281)$$

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

Avec un gradient modifié (préconditionné), on aura on obtient la forme pseudo-multiplicative :

$$x^{k+1} = x^k + \alpha^k x^k \left(\frac{U^k}{V^k} - 1 \right) \quad (282)$$

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.

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 :

$$x^{k+1} = x^k \left(\frac{U^k}{V^k} \right) \quad (283)$$

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 (281) (282) et (283) 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 (281) appliqué sur de telles divergences permet d'assurer la propriété :

$$\sum_l x_l^{k+1} = \sum_l x_l^k \quad (284)$$

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 pseudo multiplicatif (préconditionné) (282) ou d'un algorithme purement multiplicatif de type (283) (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 :

$$\tilde{x}^{k+1} = x^k \left(\frac{U^k}{V^k} \right) \quad (285)$$

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

$$x^{k+1} = \frac{\tilde{x}^{k+1}}{\sum_l \tilde{x}_l^{k+1}} C \quad (286)$$

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

16.2 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 (102) (103) :

$$0 \leq a \leq 1 \leq b \quad (287)$$

ou bien :

$$0 \leq b \leq 1 \leq a \quad (288)$$

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.

16.2.1 Divergence de Csiszär 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 (281) s'écrit ici :

$$x^{k+1} = x^k + \alpha^k x^k H^T \left\{ \underbrace{\left[\frac{a-1}{a-b} \left(\frac{p}{q^k} \right)^a + \frac{1-b}{a-b} \left(\frac{p}{q^k} \right)^b \right]}_{R^k} - \underbrace{1}_{S^k} \right\} \quad (289)$$

La forme multiplicative correspondante s'en déduit immédiatement (283) avec :

$$U^k = H^T R^k \quad ; \quad V^k = H^T S^k \quad (290)$$

16.2.2 Divergence de Bregman construite sur l'entropie générale.

L'expression de cette divergence est donnée par (120) ; compte tenu de l'expression de l'opposé du gradient donnée en (121), l'algorithme (281) s'écrit ici :

$$x^{k+1} = x^k + \alpha^k x^k H^T \left\{ \left(\frac{p}{q^k} - 1 \right) \left[\frac{(a^2 - a)}{a - b} (q^k)^{a-1} + \frac{(b - b^2)}{a - b} (q^k)^{b-1} \right] \right\} \quad (291)$$

La forme multiplicative correspondante s'en déduit immédiatement par (283), avec :

$$U^k = H^T \left\{ \left(\frac{p}{q^k} \right) \left[\frac{(a^2 - a)}{a - b} (q^k)^{a-1} + \frac{(b - b^2)}{a - b} (q^k)^{b-1} \right] \right\} \quad (292)$$

et

$$V^k = H^T \left[\frac{(a^2 - a)}{a - b} (q^k)^{a-1} + \frac{(b - b^2)}{a - b} (q^k)^{b-1} \right] \quad (293)$$

16.2.3 Divergence de Jensen construite sur l'entropie générale.

L'expression de cette divergence est donnée par (132) ; on rappelle l'expression de l'opposé du gradient donnée en (133).

$$-\frac{\partial J_\alpha(p||q)}{\partial q_j} = \frac{1 - \alpha}{a - b} (A_j - B_j) \quad (294)$$

A l'itération "k", cette expression fait apparaître les quantités :

$$A^k = b(q^k)^{b-1} + a [\alpha p + (1 - \alpha)q^k]^{a-1} > 0 \quad (295)$$

et,

$$B^k = a(q^k)^{a-1} + b [\alpha p + (1 - \alpha)q^k]^{b-1} > 0 \quad (296)$$

Les développements proposés dans l'Annexe 2 permettent de prendre en compte les situations correspondant aux cas $a - b > 0$ et $a - b < 0$.

L'algorithme (281) s'écrit alors, avec un pas de descente "δ" :

$$x^{k+1} = x^k + \delta^k x^k [U^k - V^k] \quad (297)$$

Avec :

$$U^k = (1 - \alpha) H^T \left[\frac{A^k}{2(a - b)} + \frac{A^k}{2|a - b|} + \frac{B^k}{2(b - a)} + \frac{B^k}{2|b - a|} \right] \quad (298)$$

$$V^k = (1 - \alpha) H^T \left[\frac{B^k}{2(a - b)} + \frac{B^k}{2|a - b|} + \frac{A^k}{2(b - a)} + \frac{A^k}{2|b - a|} \right] \quad (299)$$

Dans cette décomposition déjà précisée dans la section (13.5), $U^k > 0$ et $V^k > 0$; elle permet d'écrire les algorithmes multiplicatifs par la relation (283).

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

L'expression de cette divergence est donnée par (188), et l'opposé de son gradient est donnée par (189); l'algorithme (281) s'écrit ici :

$$x^{k+1} = x^k + \alpha^k x^k H^T \frac{\sum_j p_j}{\sum_j q_j^k} \left[\frac{a-1}{a-b} \left(\frac{\bar{p}}{\bar{q}^k} \right)^a + \frac{1-b}{a-b} \left(\frac{\bar{p}}{\bar{q}^k} \right)^b - \frac{a-1}{a-b} \sum_i \bar{q}_i^k \left(\frac{\bar{p}_i}{\bar{q}_i^k} \right)^a - \frac{1-b}{a-b} \sum_i \bar{q}_i^k \left(\frac{\bar{p}_i}{\bar{q}_i^k} \right)^b \right] \quad (300)$$

La décomposition permettant d'écrire un algorithme multiplicatif s'écrit en effectuant successivement les étapes (285) puis (286) :

$$U^k = H^T \frac{\sum_j p_j}{\sum_j q_j^k} \left[\frac{a-1}{a-b} \left(\frac{\bar{p}}{\bar{q}^k} \right)^a + \frac{1-b}{a-b} \left(\frac{\bar{p}}{\bar{q}^k} \right)^b \right] > 0 \quad (301)$$

et

$$V^k = H^T \frac{\sum_j p_j}{\sum_j q_j^k} \left[\frac{a-1}{a-b} \sum_i \bar{q}_i^k \left(\frac{\bar{p}_i}{\bar{q}_i^k} \right)^a + \frac{1-b}{a-b} \sum_i \bar{q}_i^k \left(\frac{\bar{p}_i}{\bar{q}_i^k} \right)^b \right] > 0 \quad (302)$$

* 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 (186).

Ceci conduit à la divergence invariante (193) dont l'opposé du gradient est donné par (196).

L'algorithme (281) s'écrit en conséquence :

$$x^{k+1} = x^k + \alpha^k x^k H^T \frac{\sum_j p_j}{\sum_j q_j^k} \left(\sum_i \bar{p}_i^t (\bar{q}_i^k)^{1-t} \right)^{\frac{1}{t}-1} \left[\left(\frac{\bar{p}}{\bar{q}^k} \right)^t - \sum_i \bar{q}_i^k \left(\frac{\bar{p}_i}{\bar{q}_i^k} \right)^t \right] \quad (303)$$

La décomposition permettant d'écrire un algorithme multiplicatif est immédiate. On explicite ci dessous, l'algorithme analogue déduit de (300) en faisant $a = t$ et $b = 1$.

Cet algorithme correspond à un facteur d'invariance non-nominal pour l'entropie de Tsallis, il s'écrit :

$$x^{k+1} = x^k + \alpha^k x^k H^T \frac{\sum_j p_j}{\sum_j q_j^k} \left[\left(\frac{\bar{p}}{\bar{q}^k} \right)^t - \sum_i \bar{q}_i^k \left(\frac{\bar{p}_i}{\bar{q}_i^k} \right)^t \right] \quad (304)$$

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

L'expression de cette divergence est donnée par (224), et l'opposé de son gradient est donnée par (225) ; l'algorithme (281) s'écrit ici :

$$x^{k+1} = x^k + \alpha^k x^k H^T \left[\frac{a(a-1)}{(a-b)} \frac{(\sum_j p_j)^a}{\sum_j q_j^k} \left(\bar{p}(\bar{q}^k)^{a-2} - \sum_i \bar{p}_i (\bar{q}_i^k)^{a-1} - (\bar{q}^k)^{a-1} + \sum_i (\bar{q}_i^k)^a \right) - \frac{b(b-1)}{(a-b)} \frac{(\sum_j p_j)^b}{\sum_j q_j^k} \left(\bar{p}(\bar{q}^k)^{b-2} - \sum_i \bar{p}_i (\bar{q}_i^k)^{b-1} - (\bar{q}^k)^{b-1} + \sum_i (\bar{q}_i^k)^b \right) \right] \quad (305)$$

La décomposition permettant d'écrire un algorithme multiplicatif s'écrit en effectuant successivement les étapes (285) puis (286) :

$$U^k = H^T \left[\frac{a(a-1)}{(a-b)} \frac{(\sum_j p_j)^a}{\sum_j q_j^k} \left(\bar{p}(\bar{q}^k)^{a-2} + \sum_i (\bar{q}_i^k)^a \right) + \frac{b(1-b)}{(a-b)} \frac{(\sum_j p_j)^b}{\sum_j q_j^k} \left(\bar{p}(\bar{q}^k)^{b-2} + \sum_i (\bar{q}_i^k)^b \right) \right] > 0 \quad (306)$$

$$V^k = H^T \left[\frac{a(a-1)}{(a-b)} \frac{(\sum_j p_j)^a}{\sum_j q_j^k} \left(\sum_i \bar{p}_i (\bar{q}_i^k)^{a-1} + (\bar{q}^k)^{a-1} \right) + \frac{b(1-b)}{(a-b)} \frac{(\sum_j p_j)^b}{\sum_j q_j^k} \left(\sum_i \bar{p}_i (\bar{q}_i^k)^{b-1} + (\bar{q}^k)^{b-1} \right) \right] > 0 \quad (307)$$

* 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 (222).

Ceci conduit à la divergence invariante (226) dont l'opposé du gradient est donné par (227).

L'algorithme (281) s'écrit en conséquence :

$$x^{k+1} = x^k + \alpha^k x^k H^T t \left[\left(\frac{\sum_i p_i (q_i^k)^{t-1}}{\sum_i (q_i^k)^t} \right)^{t-1} p (q^k)^{t-2} - \left(\frac{\sum_i p_i (q_i^k)^{t-1}}{\sum_i (q_i^k)^t} \right)^t (q^k)^{t-1} \right] \quad (308)$$

La décomposition permettant d'écrire un algorithme multiplicatif est immédiate. 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 (305) en faisant $a = t$ et $b = 1$ et qui correspond à un facteur d'invariance non-nominal pour l'entropie de Tsallis.

$$x^{k+1} = x^k + \alpha^k x^k H^T t \frac{\sum_j (p_j)^t}{\sum_j q_j^k} \left[\bar{p}(\bar{q}^k)^{t-2} - \sum_i \bar{p}_i (\bar{q}_i^k)^{t-1} - (\bar{q}^k)^{t-1} + \sum_i (\bar{q}_i^k)^t \right] \quad (309)$$

ou encore :

$$x^{k+1} = x^k + \alpha^k x^k H^T t \frac{\sum_j (p_j)^t}{\sum_j q_j^k} \left\{ \bar{p}(\bar{q}^k)^{t-2} + \sum_i (\bar{q}_i^k)^t - \left[\sum_i \bar{p}_i (\bar{q}_i^k)^{t-1} + (\bar{q}^k)^{t-1} \right] \right\} \quad (310)$$

La décomposition permettant d'écrire un algorithme multiplicatif est immédiate.

16.2.6 Divergence de Jensen invariante construite sur l'entropie générale.

L'expression de cette divergence est donnée par (247), et l'opposé de son gradient est donnée par (248).

Cette dernière expression fait apparaître les quantités :

$$A^k = a \left(\sum_j p_j \right)^a \left\{ \sum_i (\bar{q}_i^k)^a + [\alpha \bar{p} + (1 - \alpha) \bar{q}^k]^{a-1} \right\} + b \left(\sum_j p_j \right)^b \left\{ (\bar{q}^k)^{b-1} + \sum_i (\bar{q}_i^k) [\alpha \bar{p}_i + (1 - \alpha) (\bar{q}_i^k)]^{b-1} \right\} > 0 \quad (311)$$

et

$$B^k = a \left(\sum_j p_j \right)^a \left\{ (\bar{q}^k)^{a-1} + \sum_i (\bar{q}_i^k) [\alpha \bar{p}_i + (1 - \alpha) (\bar{q}_i^k)]^{a-1} \right\} + b \left(\sum_j p_j \right)^b \left\{ \sum_i (\bar{q}_i^k)^b + [\alpha \bar{p} + (1 - \alpha) \bar{q}^k]^{b-1} \right\} > 0 \quad (312)$$

L'algorithme (281) s'écrit alors, avec un pas de descente "δ"

$$x^{k+1} = x^k + \delta^k x^k [U^k - V^k] \quad (313)$$

Ici, on distingue 2 cas en fonction du signe de $(a-b)$, selon la procédure décrite dans l'Annexe 2 :

* **Si $(a-b) > 0$**

$$x^{k+1} = x^k + \delta^k x^k \frac{1-\alpha}{\sum_j q_j^k} H^T \left[\frac{A^k}{a-b} - \frac{B^k}{a-b} \right] \quad (314)$$

La décomposition permettant d'écrire des algorithmes multiplicatifs en effectuant successivement les étapes (285) puis (286) s'écrit :

$$U^k = \frac{1-\alpha}{\sum_j q_j^k} H^T \left[\frac{A^k}{a-b} \right], \quad V^k = \frac{1-\alpha}{\sum_j q_j^k} H^T \left[\frac{B^k}{a-b} \right] \quad (315)$$

* **Si $(a-b) < 0$**

$$x^{k+1} = x^k + \delta^k x^k \frac{1-\alpha}{\sum_j q_j^k} H^T \left[\frac{B^k}{b-a} - \frac{A^k}{b-a} \right] \quad (316)$$

La décomposition permettant d'écrire des algorithmes multiplicatifs en effectuant successivement les étapes (285) puis (286) s'écrit :

$$U^k = \frac{1-\alpha}{\sum_j q_j^k} H^T \left[\frac{B^k}{b-a} \right], \quad V^k = \frac{1-\alpha}{\sum_j q_j^k} H^T \left[\frac{A^k}{b-a} \right] \quad (317)$$

Afin de regrouper les deux cas, on a l'écriture synthétique suivante :

$$U^k = \frac{(1-\alpha)}{\sum_j q_j^k} H^T \left[\frac{A^k}{2(a-b)} + \frac{A^k}{2|a-b|} + \frac{B^k}{2(b-a)} + \frac{B^k}{2|b-a|} \right] \quad (318)$$

$$V^k = \frac{(1-\alpha)}{\sum_j q_j^k} H^T \left[\frac{B^k}{2(a-b)} + \frac{B^k}{2|a-b|} + \frac{A^k}{2(b-a)} + \frac{A^k}{2|b-a|} \right] \quad (319)$$

A partir de cette écriture, U^k et V^k sont positifs et l'algorithme (313) s'en déduit immédiatement.

16.3 Divergences invariantes fondées sur l'entropie Newton. Algorithmes.

Pour toutes les divergences fondées sur l'entropie de Newton, les algorithmes peuvent être utilisés sous la forme de base (281), ou bien sous la forme pseudo multiplicative (282), soit sous la forme purement multiplicative (283).

Les algorithmes fondés sur les divergences non-invariantes se déduisent sans difficulté des expressions de l'opposé des gradients explicitées dans la section (13.1).

Dans la présente section, on n'explicitera que les algorithmes fondés sur les formes invariantes des divergences.

Dans ce cas, afin de respecter la contrainte de somme, les formes pseudo-multiplicatives ou purement multiplicatives impliquent d'effectuer successivement les étapes (285) puis (286).

16.3.1 Divergence de Csiszär/Newton invariante.

L'expression de cette divergence est donnée par la relation (255), et l'opposé du gradient par rapport à "q" est donnée par l'expression (256) ; en conséquence les algorithmes (281), (282) et (283) se déduisent des expressions :

$$U^k = \frac{1}{2} \frac{\sum_l p_l}{\sum_l q_l^k} H^T \left[\frac{\bar{p}}{\bar{q}^k} + \left(\frac{\bar{p}}{\bar{q}^k} \right)^2 \right] > 0 \quad (320)$$

et

$$V^k = \frac{1}{2} \frac{\sum_l p_l}{\sum_l q_l^k} H^T \left[\sum_i \bar{p}_i \frac{\bar{p}_i}{q_i^k} + 1 \right] > 0 \quad (321)$$

16.3.2 Divergence de Bregman/Newton invariante.

L'expression de cette divergence est donnée par la relation (262), et l'opposé du gradient par rapport à "q" est donnée par l'expression (263) ; en conséquence les algorithmes (281), (282) et (283) se déduisent des expressions :

$$U^k = H^T \left[\frac{1}{2} \frac{\sum_l p_l}{\sum_l q_l^k} + \frac{(\sum_l p_l)^2}{\sum_l q_l^k} \left(\bar{p} + \sum_i (\bar{q}_i^k)^2 \right) \right] > 0 \quad (322)$$

et

$$V^k = H^T \left[\frac{1}{2} \frac{\sum_l p_l}{\sum_l q_l^k} \left(\frac{\bar{p}}{\bar{q}^k} \right) + \frac{(\sum_l p_l)^2}{\sum_l q_l^k} \left(\bar{q}^k + \sum_i \bar{p}_i \bar{q}_i^k \right) \right] > 0 \quad (323)$$

16.3.3 Divergence de Jensen/Newton invariante.

L'expression de cette divergence est donnée par la relation (271), et l'opposé du gradient par rapport à "q" est donnée par l'expression (272) ; en conséquence l'algorithme (281) s'écrit avec un pas de descente "δ" :

$$x^{k+1} = x^k + \delta^k x^k (U^k - V^k) \quad (324)$$

Avec :

$$U^k = H^T \left[2\alpha(1-\alpha) \frac{(\sum_j p_j)^2}{\sum_j q_j^k} \left\{ \sum_i (\bar{q}_i^k)^2 + \bar{p} \right\} - (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j^k} \left\{ \sum_i \bar{q}_i^k \log [\alpha \bar{p}_i + (1-\alpha) \bar{q}_i^k] + \log \bar{q}^k \right\} \right] > 0 \quad (325)$$

et

$$V^k = H^T \left[2\alpha(1-\alpha) \frac{(\sum_j p_j)^2}{\sum_j q_j^k} \left\{ \sum_i \bar{p}_i \bar{q}_i^k + \bar{q}^k \right\} - (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j^k} \left\{ \sum_i \bar{q}_i^k \log \bar{q}_i^k + \log [\alpha \bar{p} + (1-\alpha) \bar{q}^k] \right\} \right] > 0 \quad (326)$$

17 Annexe 1

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.

17.1 Divergences de Csiszär.

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

$$\frac{p_j}{q_j} X_j - 1 \quad (327)$$

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 : $X_j = \frac{1}{2} \left(\frac{p_j}{q_j}\right) + \frac{1}{2}$

17.2 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 l'entropie 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 :

$$\frac{p_j}{q_j} T_j + 1 \tag{328}$$

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 : $T_j = \frac{1}{2} \left(\frac{p_j}{q_j} \right)^{-1} \log \left(\frac{p_j}{q_j} \right) - 1$

17.3 Divergences de Bregman.

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

$$\left[\frac{p_j}{q_j} - 1 \right] Z_j \tag{329}$$

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 : $Z_j = \frac{1}{2} + q_j$

17.4 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 : $\frac{t}{1-t} [p_j^{t-1} - q_j^{t-1}]$
- 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" : $\frac{2\gamma+1}{3\gamma} (p_j^{2\gamma} - q_j^{2\gamma}) - \frac{1-\gamma}{3\gamma} (p_j^{-\gamma} - q_j^{-\gamma})$
- Entropie à 2 paramètres : $\frac{1+r+K}{2K} (p_j^{r+K} - q_j^{r+K}) - \frac{1+r-K}{2K} (p_j^{r-K} - q_j^{r-K})$
- Entropie générale : $\frac{a}{a-b} (p_j^{a-1} - q_j^{a-1}) - \frac{b}{a-b} (p_j^{b-1} - q_j^{b-1})$
- Entropie de Newton : $(p_j - q_j) + \frac{1}{2} \log \frac{p_j}{q_j} = (p_j + \frac{1}{2} \log p_j) - (q_j + \frac{1}{2} \log q_j)$

18 Annexe 2

On précise dans cette section la procédure permettant de tenir compte du signe de $a - b$ lorsque l'opposé du gradient d'une divergence "D" par rapport à "q" s'écrit sous la forme :

$$\frac{\partial D}{\partial q_j} = Z \left[\frac{A_j}{a-b} - \frac{B_j}{a-b} \right] + C \quad (330)$$

Avec $Z > 0, A_j > 0, B_j > 0, C \geq 0$.

Le problème est d'écrire :

$$\frac{\partial D}{\partial q_j} = R_j - S_j \quad (331)$$

On distingue les cas :

* Si $a - b > 0$, on aura :

$$R_j = C + Z \frac{A_j}{a - b} ; S_j = Z \frac{B_j}{a - b} \quad (332)$$

* Si $a - b < 0$, on aura :

$$R_j = C + Z \frac{B_j}{b - a} ; S_j = Z \frac{A_j}{b - a} \quad (333)$$

On peut s'en tenir là, mais si on veut avoir des expressions de R_j et S_j utilisables quelque soit le signe de $a - b$, on peut écrire :

$$R_j = C + \frac{Z}{2} \left[\frac{A_j}{a - b} + \frac{A_j}{|a - b|} \right] + \frac{Z}{2} \left[\frac{B_j}{b - a} + \frac{B_j}{|b - a|} \right] \quad (334)$$

$$S_j = \frac{Z}{2} \left[\frac{B_j}{a - b} + \frac{B_j}{|a - b|} \right] + \frac{Z}{2} \left[\frac{A_j}{b - a} + \frac{A_j}{|b - a|} \right] \quad (335)$$

19 Annexe 3

On développe ici les algorithmes correspondant spécifiquement à la divergence de Jensen/Shannon qui présentent de réelles particularités.

On rappelle l'expression de l'opposé du gradient par rapport à "q" (14) :

$$-\frac{\partial JS(p||q)}{\partial q_j} = (\alpha - 1) \log \left[\frac{q_j}{\alpha p_j + (1 - \alpha) q_j} \right] () \quad (336)$$

Ce qu'on notera :

$$-\frac{\partial JS(p||q)}{\partial q_j} = \alpha LL_j - LL_j \quad (337)$$

Un algorithme non-multiplicatif (277) s'écrira à l'itération "k" :

$$x^{k+1} = x^k + \delta^k x^k H^T \left\{ (\alpha - 1) \log \left[\frac{q^k}{\alpha p_j + (1 - \alpha) q^k} \right] \right\} \quad (338)$$

Afin de proposer d'autres formes algorithmiques, on s'appuiera sur l'analyse de la section (6.4), en se plaçant à l'itération "k"

* Pour les composantes telles que $\frac{q_j^k}{\alpha p_j + (1 - \alpha) q_j^k} > 1 \Rightarrow LL_j^k > 0$, on a :

$$R_j = \alpha LL_j^k > 0 ; S_j = LL_j^k > 0 \quad (339)$$

* Pour les composantes telles que $\frac{q_j^k}{\alpha p_j + (1-\alpha)q_j^k} < 1 \Rightarrow LL_j^k < 0$, on a :

$$R_j = -LL_j^k > 0 ; S_j = -\alpha LL_j^k > 0 \quad (340)$$

Clairement la décomposition de l'opposé du gradient change d'une composante à l'autre ; cela peut se synthétiser en utilisant la décomposition (18) qu'on rappelle ici :

$$\begin{aligned} R_j &= \frac{\alpha}{2}LL + \frac{\alpha}{2}|LL_j| - \frac{1}{2}LL_j + \frac{1}{2}|LL_j| > 0 \\ S_j &= \frac{1}{2}LL_j + \frac{1}{2}|LL_j| - \frac{\alpha}{2}LL_j + \frac{\alpha}{2}|LL_j| > 0() \end{aligned} \quad (341)$$

En conséquence, on peut proposer un algorithme pseudo multiplicatifs qui s'écrira ici :

$$x^{k+1} = x^k + \delta^k x^k \left\{ \frac{H^T [(\alpha - 1)LL^k + (\alpha + 1)|LL^k|]}{H^T [(1 - \alpha)LL^k + (\alpha + 1)|LL^k|]} - 1 \right\} \quad (342)$$

En revanche, un algorithme purement multiplicatif correspondant à " $\delta^k = 1$ " s'écrira :

$$x^{k+1} = x^k \left\{ \frac{H^T [(\alpha - 1)LL^k + (\alpha + 1)|LL^k|]}{H^T [(1 - \alpha)LL^k + (\alpha + 1)|LL^k|]} \right\} \quad (343)$$

Bilan du comportement de ces algorithmes :

* 1 - Il apparait que les algorithmes (338), utilisant l'expression de l'opposé du gradient (336) (non décomposé), ne posent pas de problème particulier, c'est une situation analogue à celle de la section (6.2).

* 2 - Il en est de même pour les algorithmes pseudo-multiplicatifs (342) s'appuyant sur la relation (341) du fait de l'utilisation d'un pas de descente " δ^k " qui permet de contrôler la convergence.

* 3 - En revanche, on a pu observer que les algorithmes purement multiplicatifs (343) pour lesquels le pas de descente est fixé à "1" posent des problèmes et sont inutilisables pour les valeurs de " α " qui ne seront pas proches de "1".

20 Annexe 4

On montre ici une utilisation erronée des propriétés du Logarithme déformé vis à vis du Logarithme naturel.

Ceci tient essentiellement au fait que la propriété du Logarithme naturel :

$$\log(a.b) = \log a + \log b \quad (344)$$

n'est plus vraie pour le Logarithme déformé :

$$\log_d(a.b) \neq \log_d a + \log_d b \quad (345)$$

Un exemple d'erreur que ça pourrait induire est le cas de la Gamma divergence : En faisant référence au document *hal - 01745256* version 3 la Beta divergence invariante a été écrite sous la forme :

$$BI(p||q) = \frac{1}{\lambda(\lambda-1)} \left[\underbrace{\sum_i p_i^\lambda}_A - \underbrace{\left(\frac{\sum_i p_i q_i^{\lambda-1}}{\sum_i q_i^\lambda} \right)^\lambda}_{B} \underbrace{\sum_i q_i^\lambda}_C \right] \quad (346)$$

Si, sur chacun des deux termes de la différence on applique la fonction Logarithme naturel, on obtient :

$$LBI(p||q) = \frac{1}{\lambda(\lambda-1)} [\log A - \log(B.C)] \quad (347)$$

Puis, en utilisant la propriété (344), on obtient l'expression de la Gamma divergence de Fujisawa et Eguchi :

$$LBI(p||q) = \frac{1}{\lambda(\lambda-1)} [\log A - \log B - \log C] \quad (348)$$

En revanche, si à partir de (346), on applique sur les 2 termes de la différence, la fonction Logarithme déformé, on obtient l'expression :

$$L_d BI(p||q) = \frac{1}{\lambda(\lambda-1)} [\log_d A - \log_d(B.C)] \quad (349)$$

Si on s'en tient là, l'algorithme développé sur la base de cette divergence permettra de respecter spontanément la contrainte de somme des inconnues ; c'est la propriété classique des divergences invariantes.

Au contraire, la relation (345) interdit de franchir la dernière étape, c'est à dire plus précisément, que le fait de remplacer brutalement les Logarithmes naturels par des Logarithmes déformés dans la divergence (348) conduit à quelque chose d'incorrect "*a priori*", dans la mesure ou un algorithme construit sur la base d'une telle divergence ne permettra pas de respecter la contrainte de somme.

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 thermo-statistics. *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 A.M. Scarfone. Finite difference and averaging operators in generalized entropies. In *Journal of Physics : Conference Series*, volume 201, page 012005. IOP Publishing, 2010.