



# **- Généralisation des divergences par application du Logarithme déformé - Applications aux problèmes inverses linéaires - Algorithmes d'inversion**

Henri Lantéri

## **► To cite this version:**

Henri Lantéri. - Généralisation des divergences par application du Logarithme déformé - Applications aux problèmes inverses linéaires - Algorithmes d'inversion. 2023. <hal-04055699>

**HAL Id: hal-04055699**

**<https://hal.science/hal-04055699v1>**

Preprint submitted on 3 Apr 2023

**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.



Distributed under a Creative Commons CC0 1.0 - Universal - International License

- Généralisation des divergences par application  
du Logarithme déformé.
- Applications aux problèmes inverses linéaires.
  - Algorithmes d'inversion.
  - Version 1. -

Henri Lantéri

3 avril 2023

## Table des matières

<b>1</b>	<b>Rappels sur le contexte.</b>	<b>2</b>
<b>2</b>	<b>But de cette étude.</b>	<b>3</b>
<b>3</b>	<b>Rappels sur le Logarithme déformé.</b>	<b>3</b>
<b>4</b>	<b>Rappels sur les divergences invariantes - Facteur d'invariance.</b>	<b>6</b>
<b>5</b>	<b>Formes générales des divergences et application du Logarithme déformé.</b>	<b>7</b>
5.1	Formes classiques des divergences. . . . .	7
5.2	Formes particulières des divergences. . . . .	8
5.2.1	Divergences faisant apparaitre le Logarithme naturel. . .	8
5.2.2	Formes particulières des divergences invariantes. . . . .	9
5.3	Expressions des Divergences et des Gradients par rapport à "q".	9
5.3.1	Divergence de type LD. . . . .	9
5.3.2	Divergence de type LD1. . . . .	10
5.3.3	Divergence de type LD2. . . . .	10
<b>6</b>	<b>Application aux divergences <math>(\alpha)</math>, <math>(\beta)</math> et <math>(\alpha\beta)</math>.</b>	<b>11</b>
6.1	Alpha $(\alpha)$ divergence. . . . .	11
6.2	Alpha $(\alpha)$ divergence invariante. . . . .	13
6.3	Beta $(\beta)$ divergence. . . . .	15
6.4	Beta $(\beta)$ divergence invariante. . . . .	17
6.5	Alpha-Beta $(\alpha\beta)$ divergence. . . . .	18
6.6	Alpha-Beta $(\alpha\beta)$ divergence invariante. . . . .	21

<b>7</b>	<b>Application aux divergences entre moyennes.</b>	<b>23</b>
7.1	Divergence Géométrique-Harmonique. . . . .	23
7.2	Divergence Géométrique-Harmonique invariante. . . . .	24
7.3	Divergence Arithmique-Géométrique. . . . .	25
7.4	Divergence Arithmique-Géométrique invariante. . . . .	26
7.5	Divergence Arithmétique-Harmonique. . . . .	27
7.6	Divergence Arithmétique-Harmonique invariante. . . . .	27
<b>8</b>	<b>Application aux divergences F et G.</b>	<b>28</b>
8.1	Divergence F. . . . .	29
8.2	Divergence F invariante. . . . .	29
8.3	Divergence G. . . . .	30
8.4	Divergence G invariante. . . . .	31
<b>9</b>	<b>Algorithmique.</b>	<b>32</b>
9.1	Rappels sur la méthode de construction des algorithmes. . . . .	32
<b>10</b>	<b>Annexe 1</b>	<b>35</b>
<b>11</b>	<b>Annexe 2</b>	<b>36</b>
<b>12</b>	<b>Annexe 3</b>	<b>36</b>
12.1	Cas particulier de la divergence de Kullback-Leibler duale. . . . .	36
12.2	Forme invariante. . . . .	38

## 1 Rappels sur le contexte.

Ce travail s'inscrit dans le contexte général des problèmes inverses [5], [12]. La résolution de ce type de problème implique de minimiser (éventuellement sous contraintes), une fonction d'écart (Divergence)  $D(p||q)$  entre les mesures " $p$ " et un modèle physique " $q$ " du phénomène considéré; les inconnues du problème étant les paramètres du modèle.

Dans le cadre des problèmes de déconvolution d'images, par exemple, les contraintes considérées sont typiquement la contrainte de non-négativité des valeurs des intensités des pixels de l'image reconstruite, 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 [18], [19].

Afin de traiter de façon simple la contrainte de somme des paramètres inconnus, nous avons introduit dans ces précédents travaux, les divergences invariantes par changement d'échelle sur les inconnues, et la notion de facteur d'invariance.

## 2 But de cette étude.

Le but de cette note est d'étendre les divergences analysées dans un précédent travail [18], [19], par application du Logarithme déformé sous sa forme la plus générale.

Dans une étude portant sur les divergences entropiques [20] [21], nous avons analysé les différentes formes du Logarithme déformé et leurs applications à ce type de divergence.

Une forme très générale du logarithme déformé a été mise en évidence; elle permet de regrouper les différentes expressions du Logarithme déformé sous une forme unique qui englobe toutes les autres.

Au delà des applications liées aux divergences fondées sur l'entropie, on envisage ici d'étendre les divergences classiques par application du Logarithme déformé généralisé.

## 3 Rappels sur le Logarithme déformé.

Les différentes expressions du Logarithme déformé trouvées dans la littérature ont été présentées dans un précédent travail [20] [21] et les propriétés principales de cette fonction sont développées dans [24].

Nous rappelons ici la démarche permettant d'obtenir la forme la plus générale du Logarithme déformé que nous utiliserons par la suite.

Dans le contexte de la Physique statistique, par analogie avec l'entropie de Shannon [25], une expression générale de l'entropie s'écrit sous la forme :

$$S(p) = - \sum_i p_i \Lambda(p_i) \quad (1)$$

Dans cette expression, la fonction  $\Lambda(p_i)$  est le Logarithme déformé ou bien le logarithme naturel dans le cas de l'entropie de Shannon qui s'écrit :

$$S(p) = - \sum_i p_i \log p_i \quad (2)$$

Cette dernière expression peut être obtenue en utilisant la définition proposée par Abe [1] :

$$S(p) = \left[ -\frac{d}{d\alpha} \sum_i p_i^\alpha \right]_{\alpha=1} \equiv \left[ -\frac{d}{d\alpha} f(\alpha) \right]_{\alpha=1} \quad (3)$$

Dans cette expression la dérivée mise en jeu est la dérivée classique.

A partir de (3), il est proposé de déformer la dérivée classique et de la remplacer par la "q" différentielle de Jackson [13] [14] qui s'écrit :

$$\frac{df(\alpha)}{d(\alpha : t)} = \frac{f(\alpha t) - f(\alpha)}{\alpha t - \alpha} \quad (4)$$

Avec cette définition, on obtient :

$$S(p) = \left[ -\frac{\sum_i p_i^{\alpha t} - \sum_i p_i^{\alpha}}{\alpha t - \alpha} \right]_{\alpha=1} \quad (5)$$

Ce qui correspond à l'entropie de Tsallis [29] :

$$S_T(p) = -\frac{\sum_i p_i^t - \sum_i p_i}{t-1} = -\frac{\sum_i p_i^t - 1}{t-1} = -\sum_i p_i \left( \frac{p_i^{t-1} - 1}{t-1} \right) \quad (6)$$

Le Logarithme déformé correspondant s'écrit :

$$\text{Log}_T(x) = \frac{x^{t-1} - 1}{t-1} \quad (7)$$

Une alternative consiste, toujours à partir de (3), à adopter une autre définition de la différentielle donnée par Mc Anally [22] qui s'écrit :

$$\frac{df(\alpha)}{d(\alpha : t)} = \frac{f(\alpha t) - f(\alpha t^{-1})}{\alpha t - \alpha t^{-1}} \quad (8)$$

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

$$S(p) = \left[ -\frac{\sum_i p_i^{\alpha t} - \sum_i p_i^{\alpha t^{-1}}}{\alpha t - \alpha t^{-1}} \right]_{\alpha=1} = -\frac{\sum_i p_i^t - \sum_i p_i^{t^{-1}}}{t - t^{-1}} \quad (9)$$

On retrouve ainsi l'entropie de Abe [1]; le Logarithme déformé correspondant s'écrit :

$$\text{Log}_A(x) = \frac{x^{t-1} - x^{t^{-1}-1}}{t - t^{-1}} \quad (10)$$

Une autre forme de différentielle généralisée à 2 paramètres, initialement utilisée par Chakrabarti et Jagannathan [7], est proposée par Borges et Roditi [6]; elle consiste à utiliser la définition suivante :

$$\frac{d_{ab}f(\alpha)}{d_{ab}(\alpha)} = \frac{f(a\alpha) - f(b\alpha)}{a\alpha - b\alpha} \quad (11)$$

Cette forme étend les définitions précédemment proposées, elle conduit, avec la définition (3) à l'expression de l'entropie ;

$$\begin{aligned} S_{ab}(p) &= \left[ -\frac{\sum_i p_i^{a\alpha} - \sum_i p_i^{b\alpha}}{a\alpha - b\alpha} \right]_{\alpha=1} = -\frac{\sum_i p_i^a - \sum_i p_i^b}{a - b} \\ &= -\sum_i p_i \left( \frac{p_i^{a-1} - p_i^{b-1}}{a - b} \right) \end{aligned} \quad (12)$$

Cette forme d'entropie très générale (et le logarithme déforme associé) est mentionnée par Wada et Scarfone [30]; elle est fondée sur les travaux plus anciens

de Sharma et Taneja [26] et Mittal [23] et permet bien entendu de retrouver les expressions citées précédemment.

Les domaines de valeurs des paramètres “ $a$ ” et “ $b$ ” indiqués dans [11] et [6] sont les suivants :

$$0 < a \leq 1 \leq b \quad (13)$$

ou bien :

$$0 < b \leq 1 \leq a \quad (14)$$

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

$$Log_{ab}(x) \equiv Log_d(x) = \frac{x^{a-1} - x^{b-1}}{a - b} \quad (15)$$

Le logarithme déformé ainsi défini possède les propriétés de concavité du Logarithme classique dans un domaine de valeurs des paramètres précité.

On notera qu’en tenant compte de (13) et (14), on a :

$$\frac{a-1}{a-b} > 0 \ ; \ \frac{b-1}{a-b} < 0 \quad (16)$$

Les cas particuliers développés dans [20] se retrouvent en effectuant les adaptations suivantes :

- Entropie de Shannon [25] :  $a \rightarrow 1$  et  $b \rightarrow 1$  en respectant (13) ou (14)
- Entropie de Tsallis [29]  $a = t$ ,  $b = 1$
- Entropie de Kaniadakis [15] :  $a = 1 + K$ ,  $b = 1 - K$
- Entropie de Abe [1] :  $a = z$ ,  $b = \frac{1}{z}$
- Entropie “ $\gamma$ ” [16] :  $a = 2\gamma + 1$ ,  $b = 1 - \gamma$
- Entropie à 2 paramètres (KLS) [16] :  $a = 1 + r + K$ ,  $b = 1 + r - K$

Par ailleurs, alors que pour le Logarithme naturel on a :

$$Log(x.y) = Log(x) + Log(y) \quad (17)$$

, pour le Logarithme déformé, cette relation n’est plus vraie, et on a :

$$Log_d(x.y) \neq Log_d(x) + Log_d(y) \quad (18)$$

On notera que le Logarithme naturel est obtenu à partir de la relation (15) en prenant par exemple  $a = 1 + \epsilon$ ,  $b = 1 - \epsilon$  et en effectuant le passage à la limite  $\epsilon \rightarrow 0$ .

Le détail du calcul permettant cette transition est indiqué dans l’Annexe 1.

Les relations (17) et (18) impliquent des traitements différents selon que les

divergences logarithmiques sont obtenues par application du logarithme naturel ou du Logarithme déformé.

Nous mettons en évidence dans les développements qui suivent les différences entre les deux approches.

## 4 Rappels sur les divergences invariantes - Facteur d'invariance.

Cette notion a été longuement présentée dans [18] et [19], nous ne rappelons ici que les points principaux.

Partant d'une divergence  $D(p||q)$ , on se propose de construire une divergence  $D(p||Kq) = DI(p||q)$  (avec " $K(p, q)$ " scalaire positif), telle que  $DI(p||q)$  reste invariante lorsque " $q$ " est multiplié par une constante positive.

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 :

$$\sum_i \frac{\partial D(p_i||Kq_i)}{\partial K} = 0 \quad (19)$$

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 :

$$K(p, q) + \sum_j q_j \frac{\partial K(p, q)}{\partial q_j} = 0 \quad (20)$$

Les facteurs d'invariance nominaux sont bien entendu solutions de l'équation différentielle ci-dessus.

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

$$\sum_j q_j \frac{\partial D(p||K(p, q).q)}{\partial q_j} = 0 \quad (21)$$

## 5 Formes générales des divergences et application du Logarithme déformé.

### 5.1 Formes classiques des divergences.

Une divergence est par définition même, une quantité positive. Ainsi que nous l'avons indiqué dans la section 1, elle exprime un écart entre 2 champs de données : le champ " $p$ " qui pour nous représentera toujours les mesures du phénomène considéré, et le champ " $q$ " qui représentera le modèle physique du même phénomène.

Les paramètres du modèle sont les inconnues du problème.

Ceci étant, de très nombreuses divergences, qu'elles soient non-invariantes ou bien qu'elles soient rendues invariantes par introduction d'un facteur d'invariance quel qu'il soit, peuvent s'écrire sous la forme d'une différence de 2 termes positifs.

**\*Remarque :** Lorsque le facteur d'invariance nominal n'a pas d'expression explicite, on sera amenés à utiliser systématiquement, en guise de facteur d'invariance une forme spécifique ; les divergences invariantes ainsi obtenues seront discutées au cas par cas.

Ces situations particulières étant mise à part (pour l'instant), de très nombreuses divergences peuvent s'écrire sous la forme générale suivante :

$$D(p||q) = [A(p, q) - B(p, q)] \quad ; \quad A(p, q) > 0 \quad ; \quad B(p, q) > 0 \quad (22)$$

A partir de cette relation la plus classique, 2 cas apparaissent : soit  $B = X + Y$ , soit  $B = X.Y$  ce qui conduit aux expressions classiques :

$$D1(p||q) = \{A(p, q) - [X(p, q) + Y(p, q)]\} \quad (23)$$

$$D2(p||q) = \{A(p, q) - [X(p, q) . Y(p, q)]\} \quad (24)$$

De plus, si les expressions  $A$ ,  $B$ ,  $X$ ,  $Y$  sont séparables, on aura :

$$\begin{aligned} A(p, q) &= \sum_i A(p_i, q_i) = \sum_i A_i \quad ; \quad A_i > 0 \quad \forall i \\ B(p, q) &= \sum_i B(p_i, q_i) = \sum_i B_i \quad ; \quad B_i > 0 \quad \forall i \\ X(p, q) &= \sum_i X(p_i, q_i) = \sum_i X_i \quad ; \quad X_i > 0 \quad \forall i \\ Y(p, q) &= \sum_i Y(p_i, q_i) = \sum_i Y_i \quad ; \quad Y_i > 0 \quad \forall i \end{aligned} \quad (25)$$

Lorsqu'on applique Logarithme déformé sur de telles expressions, dans la mesure où cette transformation doit conserver la positivité de



la divergence obtenue, les précautions à prendre sont les suivantes :

**1 - Un terme de la forme  $[X(p, q) + Y(p, q)]$  doit être considéré comme un bloc indissociable (ce qui semble évident).**

**2 - Compte tenu de (18), un terme de la forme  $[X(p, q) \cdot Y(p, q)]$  doit être considéré comme un bloc indissociable.**

Pour les divergences de type  $D$ , les formes logarithmiques proposées dans la littérature sont obtenues en appliquant le logarithme, naturel ou déformé sur chacun des termes de la divergence ce qui conduit avec le Logarithme naturel, à :

$$LD(p||q) = [\log A(p, q) - \log B(p, q)] \quad (26)$$

ou bien avec le Logarithme déformé :

$$L_d D(p||q) = [\log_d A(p, q) - \log_d B(p, q)] \quad (27)$$

Pour la forme D1, cette opération n'entraîne pas de problème particulier ; en appliquant le logarithme naturel sur les deux termes de la divergence,  $A$  et  $X + Y$ , il vient :

$$LD1(p||q) = \{\log A(p, q) - \log [X(p, q) + Y(p, q)]\} \quad (28)$$

De même, en appliquant le Logarithme déformé, on aura :

$$L_d D1(p||q) = \{\log_d A(p, q) - \log_d [X(p, q) + Y(p, q)]\} \quad (29)$$

Le cas de la divergence  $D2(p||q)$  conduit par application du Logarithme naturel à :

$$LD2(p||q) = [\log A(p, q) - \log X(p, q) - \log Y(p, q)] \quad (30)$$

En revanche, avec le Logarithme déformé, la décomposition du terme produit est impossible compte tenu de (18), et on aura :

$$L_d D2(p||q) = \{\log_d A(p, q) - \log_d [X(p, q) \cdot Y(p, q)]\} \quad (31)$$

## 5.2 Formes particulières des divergences.

### 5.2.1 Divergences faisant apparaître le Logarithme naturel.

Certaines divergences proposées dans la littérature font apparaître la fonction Logarithme naturel, c'est par exemple le cas bien connu de la divergence de Kullback-Leibler [17] fondée sur l'entropie de Shannon et de toutes les divergences qui en sont déduites.

Dans ce cas, le Logarithme déformé s'introduit simplement en remplacement du Logarithme naturel.

Cette extension a été considérée dans un précédent travail [20] portant spécifiquement sur les divergences entropiques et peut être mise en oeuvre systématiquement.

\* **ATTENTION** : Certaines divergences, c'est le cas notamment de la "Gamma divergence" [8] sont obtenues en appliquant la fonction Logarithme naturel sur des divergences de la forme  $D2(p||q)$  (24) selon l'expression (30). Pour de telles divergences, il est bien entendu incorrect de remplacer brutalement le Logarithme naturel par le Logarithme déformé car cela serait équivalent à considérer que la propriété (17) spécifique du Logarithme naturel est aussi valable pour le Logarithme déformé (ce qui est faux).

### 5.2.2 Formes particulières des divergences invariantes.

Ainsi qu'on l'a signalé précédemment, pour certaines divergences, il est impossible d'obtenir une expression explicite du facteur d'invariance nominal. Dans ces cas particuliers, on utilisera systématiquement le facteur d'invariance  $K^* = \frac{\sum_j p_j}{\sum_j q_j}$  qui est le facteur d'invariance nominal pour la divergence de Kullback-Leibler.

Lorsqu'on utilise un tel facteur d'invariance, à un facteur multiplicatif près, qui ne dépend que de  $\sum_j p_j$ , la divergence invariante obtenue a la même expression que la divergence initiale à condition de remplacer les variables "p" et "q" par des variables normalisées  $\bar{p}_i = \frac{p_i}{\sum_j p_j}$  et  $\bar{q}_i = \frac{q_i}{\sum_j q_j}$ .

Ceci étant, tout ce qui figure dans la Section (5.1) reste valable avec les nouvelles variables.

## 5.3 Expressions des Divergences et des Gradients par rapport à "q".

### 5.3.1 Divergence de type LD.

Par application du Logarithme généralisé donné par (15), la divergence (22) devient :

$$L_d D(p||q) = \left\{ \frac{A^{a-1}(p, q) - A^{b-1}(p, q)}{a - b} - \frac{B^{a-1}(p, q) - B^{b-1}(p, q)}{a - b} \right\} \quad (32)$$

Son gradient par rapport à "q" s'écrit,  $\forall j$  :

$$\begin{aligned} \frac{\partial L_d D(p||q)}{\partial q_j} = & \left\{ \left[ \frac{a-1}{a-b} A^{a-2} - \frac{b-1}{a-b} A^{b-2} \right] \frac{\partial A}{\partial q_j} \right\} \\ & - \left\{ \left[ \frac{a-1}{a-b} B^{a-2} - \frac{b-1}{a-b} B^{b-2} \right] \frac{\partial B}{\partial q_j} \right\} \end{aligned} \quad (33)$$

Cette expression se simplifie lorsque les divergence de base considérées sont séparables, on a en effet :

$$\begin{aligned} \frac{\partial A}{\partial q_j} &= \frac{\partial A(p_j, q_j)}{\partial q_j} \equiv \frac{\partial A_j}{\partial q_j} \\ \frac{\partial B}{\partial q_j} &= \frac{\partial B(p_j, q_j)}{\partial q_j} \equiv \frac{\partial B_j}{\partial q_j} \end{aligned} \quad (34)$$

### 5.3.2 Divergence de type LD1.

Par application du Logarithme généralisé donné par (15), la divergence (23) devient :

$$L_d D1(p||q) = \left[ \frac{A^{a-1}(p, q) - A^{b-1}(p, q)}{a - b} \right] - \left[ \frac{(X + Y)^{a-1}(p, q) - (X + Y)^{b-1}(p, q)}{a - b} \right] \quad (35)$$

Son gradient par rapport à "q" s'écrit,  $\forall j$  :

$$\frac{\partial L_d D1(p||q)}{\partial q_j} = \left\{ \left[ \frac{a-1}{a-b} A^{a-2} - \frac{b-1}{a-b} A^{b-2} \right] \frac{\partial A}{\partial q_j} \right\} - \left\{ \left[ \frac{a-1}{a-b} (X + Y)^{a-2} - \frac{b-1}{a-b} (X + Y)^{b-2} \right] \frac{\partial (X + Y)}{\partial q_j} \right\} \quad (36)$$

Cette expression se simplifie lorsque les divergence de base considérées sont séparables, en effet, on aura :

$$\frac{\partial A}{\partial q_j} = \frac{\partial A(p_j, q_j)}{\partial q_j} \quad ; \quad \frac{\partial (X + Y)}{\partial q_j} = \frac{\partial X(p_j, q_j)}{\partial q_j} + \frac{\partial Y(p_j, q_j)}{\partial q_j} \quad (37)$$

### 5.3.3 Divergence de type LD2.

Ici, le problème se complique un peu compte tenu de la propriété (ou de la non-propriété) (18) du Logarithme déformé, en effet, la divergence considérée a été écrite (31) :

$$L_d D2(p||q) = \{\log_d A(p, q) - \log_d [X(p, q) \cdot Y(p, q)]\} \quad (38)$$

En explicitant la fonction Logarithme déformé, il vient :

$$L_d D2(p||q) = \left\{ \frac{A^{a-1} - A^{b-1}}{a - b} - \left[ \frac{(X \cdot Y)^{a-1} - (X \cdot Y)^{b-1}}{a - b} \right] \right\} \quad (39)$$

Son gradient par rapport à "q" s'écrit,  $\forall j$  :

$$\frac{\partial L_d D2(p||q)}{\partial q_j} = \left\{ \left[ \frac{a-1}{a-b} A^{a-2} - \frac{b-1}{a-b} A^{b-2} \right] \frac{\partial A}{\partial q_j} \right\} - \left\{ \left[ \frac{a-1}{a-b} (X \cdot Y)^{a-2} - \frac{b-1}{a-b} (X \cdot Y)^{b-2} \right] \frac{\partial (X \cdot Y)}{\partial q_j} \right\} \quad (40)$$

Dans certains cas, compte tenu de la séparabilité des termes  $A(p, q)$ ,  $X(p, q)$  et  $Y(p, q)$ , on a :

$$\frac{\partial A(p, q)}{\partial q_j} = \frac{\partial A(p_j, q_j)}{\partial q_j} = \frac{\partial A_j}{\partial q_j} \quad (41)$$

$$\frac{\partial (X \cdot Y)}{\partial q_j} = X \frac{\partial Y(p_j, q_j)}{\partial q_j} + Y \frac{\partial X(p_j, q_j)}{\partial q_j} = X \frac{\partial Y_j}{\partial q_j} + Y \frac{\partial X_j}{\partial q_j} \quad (42)$$

Dans le cas du Logarithme naturel explicité par (30) le gradient s'obtient sans difficulté soit par calcul direct, soit à partir des relations précédentes en suivant la procédure indiquée dans l'Annexe 1 et conduit à :

$$\frac{\partial LD2(p\|q)}{\partial q_j} = \left\{ \frac{1}{A} \frac{\partial A}{\partial q_j} - \left[ \frac{1}{X} \frac{\partial X}{\partial q_j} + \frac{1}{Y} \frac{\partial Y}{\partial q_j} \right] \right\} \quad (43)$$

## 6 Application aux divergences $(\alpha)$ , $(\beta)$ et $(\alpha\beta)$ .

### 6.1 Alpha $(\alpha)$ divergence.

Cette divergence déjà analysée dans [18], [19] et dans les travaux qui y sont cités [2],[8],[9] s'écrit sous la forme :

$$D_\alpha(p\|q) = \frac{1}{\alpha(\alpha-1)} \left\{ \sum_i p_i^\alpha q_i^{1-\alpha} - \left( \alpha \sum_i p_i + (1-\alpha) \sum_i q_i \right) \right\} \quad (44)$$

Elle peut s'écrire sous la forme équivalente :

$$D_\alpha(p\|q) = \frac{1}{\alpha(\alpha-1)} \sum_i p_i^\alpha q_i^{1-\alpha} - \frac{1}{\alpha-1} \sum_i p_i + \frac{1}{\alpha} \sum_i q_i \quad (45)$$

La décomposition de cette divergence sous la forme  $A - B$  avec  $A > 0$  et  $B > 0$  implique de distinguer 3 cas selon les valeurs du paramètre " $\alpha$ ".

**\*Cas 1 :  $0 < \alpha < 1$ , on a alors :**

$$A = \frac{1}{1-\alpha} \sum_i p_i + \frac{1}{\alpha} \sum_i q_i ; \quad B = \frac{1}{\alpha(1-\alpha)} \sum_i p_i^\alpha q_i^{1-\alpha} \quad (46)$$

Par application du Logarithme déformé sur chaque terme de la différence, on obtient (32) dont le gradient par rapport à " $q$ " est donné par (33) ; en tenant compte de la séparabilité des termes  $A(p, q)$  et  $B(p, q)$ , il vient :

$$\frac{\partial A}{\partial q_j} = \frac{\partial A_j}{\partial q_j} = \frac{1}{\alpha} ; \quad \frac{\partial B}{\partial q_j} = \frac{\partial B_j}{\partial q_j} = \frac{1}{\alpha} p_j^\alpha q_j^{-\alpha} \quad (47)$$

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

$$\begin{aligned} -\frac{\partial L_d D_\alpha(p\|q)}{\partial q_j} &= -\frac{1}{\alpha} \left[ \frac{a-1}{a-b} A^{a-2} - \frac{b-1}{a-b} A^{b-2} \right] \\ &\quad + \frac{1}{\alpha} \left[ \frac{a-1}{a-b} B^{a-2} - \frac{b-1}{a-b} B^{b-2} \right] p_j^\alpha q_j^{-\alpha} \end{aligned} \quad (48)$$

Pour un usage algorithmique, l'expression de l'opposé du gradient s'écrit sous la forme  $U - V$  avec  $U > 0$  et  $V > 0$ , en conséquence en tenant compte des

conditions (13) et (14), il vient :

$$\begin{aligned} U_j &= \frac{1}{\alpha} \left[ \frac{a-1}{a-b} B^{a-2} - \frac{b-1}{a-b} B^{b-2} \right] p_j^\alpha q_j^{-\alpha} \\ V_j &= \frac{1}{\alpha} \left[ \frac{a-1}{a-b} A^{a-2} - \frac{b-1}{a-b} A^{b-2} \right] \end{aligned} \quad (49)$$

**\*Cas 2 :**  $\alpha > 1$ , on a alors :

$$A = \frac{1}{\alpha} \sum_i q_i + \frac{1}{\alpha(\alpha-1)} \sum_i p_i^\alpha q_i^{1-\alpha} ; \quad B = \frac{1}{\alpha-1} \sum_i p_i \quad (50)$$

Par application du Logarithme déformé sur chaque terme de la différence, on obtient (32) dont le gradient par rapport à "q" est donné par (33) ; en tenant compte de la séparabilité des termes  $A(p, q)$  et  $B(p, q)$ , il vient :

$$\frac{\partial A}{\partial q_j} = \frac{\partial A_j}{\partial q_j} = -\frac{1}{\alpha} p_j^\alpha q_j^{-\alpha} + \frac{1}{\alpha} ; \quad \frac{\partial B}{\partial q_j} = \frac{\partial B_j}{\partial q_j} = 0 \quad (51)$$

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

$$-\frac{\partial L_d D_\alpha(p||q)}{\partial q_j} = \frac{1}{\alpha} \left[ \frac{a-1}{a-b} A^{a-2} - \frac{b-1}{a-b} A^{b-2} \right] (p_j^\alpha q_j^{-\alpha} - 1) \quad (52)$$

Pour un usage algorithmique, l'expression de l'opposé du gradient s'écrit sous la forme  $U - V$  avec  $U > 0$  et  $V > 0$ , en conséquence en tenant compte des conditions (13) et (14), il vient :

$$\begin{aligned} U_j &= \frac{1}{\alpha} \left[ \frac{a-1}{a-b} A^{a-2} - \frac{b-1}{a-b} A^{b-2} \right] p_j^\alpha q_j^{-\alpha} \\ V_j &= \frac{1}{\alpha} \left[ \frac{a-1}{a-b} A^{a-2} - \frac{b-1}{a-b} A^{b-2} \right] \end{aligned} \quad (53)$$

**\*Cas 3 :**  $\alpha < 0$ , ce cas qui est un peu surprenant doit malgré tout être considéré car il fait partie des développements concernant la "α" divergence. On a alors :

$$A = \frac{1}{1-\alpha} \sum_i p_i + \frac{1}{\alpha(\alpha-1)} \sum_i p_i^\alpha q_i^{1-\alpha} ; \quad B = -\frac{1}{\alpha} \sum_i q_i \quad (54)$$

Par application du Logarithme déformé sur chaque terme de la différence, on obtient (32) dont le gradient par rapport à "q" est donné par (33) ; en tenant compte de la séparabilité des termes  $A(p, q)$  et  $B(p, q)$ , il vient :

$$\frac{\partial A}{\partial q_j} = \frac{\partial A_j}{\partial q_j} = -\frac{1}{\alpha} p_j^\alpha q_j^{-\alpha} ; \quad \frac{\partial B}{\partial q_j} = \frac{\partial B_j}{\partial q_j} = -\frac{1}{\alpha} \quad (55)$$

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

$$-\frac{\partial L_d D_\alpha(p||q)}{\partial q_j} = \frac{1}{\alpha} \left[ \frac{a-1}{a-b} A^{a-2} - \frac{b-1}{a-b} A^{b-2} \right] p_j^\alpha q_j^{-\alpha} - \frac{1}{\alpha} \left[ \frac{a-1}{a-b} B^{a-2} - \frac{b-1}{a-b} B^{b-2} \right] \quad (56)$$

Pour un usage algorithmique, l'expression de l'opposé du gradient s'écrit sous la forme  $U - V$  avec  $U > 0$  et  $V > 0$ , en conséquence en tenant compte des conditions (13) et (14), il vient :

$$U_j = -\frac{1}{\alpha} \left[ \frac{a-1}{a-b} B^{a-2} - \frac{b-1}{a-b} B^{b-2} \right] \quad (57)$$

$$V_j = -\frac{1}{\alpha} \left[ \frac{a-1}{a-b} A^{a-2} - \frac{b-1}{a-b} A^{b-2} \right] p_j^\alpha q_j^{-\alpha}$$

## 6.2 Alpha ( $\alpha$ ) divergence invariante.

Cette divergence a été étudiée dans un précédent travail [18], [19]. En introduisant son facteur d'invariance nominal qui est calculable explicitement, et s'écrit :

$$K_{0,\alpha} = \left[ \frac{\sum_i p_i^\alpha q_i^{1-\alpha}}{\sum_i q_i} \right]^{\frac{1}{\alpha}} \quad (58)$$

On obtient la forme invariante :

$$D_\alpha I(p||q) = \underbrace{\frac{1}{1-\alpha}}_T \left[ \underbrace{\sum_i p_i}_A - \underbrace{\left( \sum_i p_i^\alpha q_i^{1-\alpha} \right)^{\frac{1}{\alpha}}}_X \times \underbrace{\left( \sum_i q_i \right)^{1-\frac{1}{\alpha}}}_Y \right] \quad (59)$$

En faisant abstraction du facteur multiplicatif "T", cette divergence est de la forme D2, par ailleurs il est inutile de distinguer les différentes valeurs possibles de " $\alpha > 0$ " qui conduisent toutes à des résultats identiques.

**Note : Le cas " $\alpha < 0$ " fera l'objet d'une remarque à la fin de ce paragraphe.**

Les termes "X" et "Y" n'étant pas séparables, le calcul des dérivées partielles

conduit à :

$$\begin{aligned}
\frac{\partial A}{\partial q_j} &= 0 \\
\frac{\partial X}{\partial q_j} &= \frac{1-\alpha}{\alpha} \left( \sum_i p_i^\alpha q_i^{1-\alpha} \right)^{\frac{1}{\alpha}-1} p_j^\alpha q_j^{-\alpha} \\
\frac{\partial Y}{\partial q_j} &= \frac{\alpha-1}{\alpha} \left( \sum_i q_i \right)^{-\frac{1}{\alpha}} \cdot 1
\end{aligned} \tag{60}$$

Par application du Logarithme déformé, en suivant le développement de la section (5.3.3), compte tenu de l'expression de "T", on obtient l'expression de l'opposé du gradient par rapport à "q" :

$$\begin{aligned}
-\frac{\partial L_d D_\alpha I(p||q)}{\partial q_j} &= -\frac{1}{\alpha} \left[ \frac{a-1}{a-b} (X.Y)^{a-2} - \frac{b-1}{a-b} (X.Y)^{b-2} \right] \times \\
&\quad \left[ \left( \frac{\sum_i p_i^\alpha q_i^{1-\alpha}}{\sum_i q_i} \right)^{\frac{1}{\alpha}} - \left( \frac{\sum_i p_i^\alpha q_i^{1-\alpha}}{\sum_i q_i} \right)^{\frac{1}{\alpha}-1} p_j^\alpha q_j^{-\alpha} \right]
\end{aligned} \tag{61}$$

La décomposition de l'opposé du gradient sous la forme  $U - V$  avec  $U > 0$  et  $V > 0$  s'écrit :

$$\begin{aligned}
U_j &= \frac{1}{\alpha} \left[ \frac{a-1}{a-b} (X.Y)^{a-2} - \frac{b-1}{a-b} (X.Y)^{b-2} \right] \times \left( \frac{\sum_i p_i^\alpha q_i^{1-\alpha}}{\sum_i q_i} \right)^{\frac{1}{\alpha}-1} p_j^\alpha q_j^{-\alpha} \\
V_j &= \frac{1}{\alpha} \left[ \frac{a-1}{a-b} (X.Y)^{a-2} - \frac{b-1}{a-b} (X.Y)^{b-2} \right] \times \left( \frac{\sum_i p_i^\alpha q_i^{1-\alpha}}{\sum_i q_i} \right)^{\frac{1}{\alpha}}
\end{aligned} \tag{62}$$

**Lorsque  $\alpha < 0$ , on aura :**

$$\begin{aligned}
U_j &= -\frac{1}{\alpha} \left[ \frac{a-1}{a-b} (X.Y)^{a-2} - \frac{b-1}{a-b} (X.Y)^{b-2} \right] \times \left( \frac{\sum_i p_i^\alpha q_i^{1-\alpha}}{\sum_i q_i} \right)^{\frac{1}{\alpha}} \\
V_j &= -\frac{1}{\alpha} \left[ \frac{a-1}{a-b} (X.Y)^{a-2} - \frac{b-1}{a-b} (X.Y)^{b-2} \right] \times \left( \frac{\sum_i p_i^\alpha q_i^{1-\alpha}}{\sum_i q_i} \right)^{\frac{1}{\alpha}-1} p_j^\alpha q_j^{-\alpha}
\end{aligned} \tag{63}$$

On vérifie sans difficulté que, comme pour toutes les divergences invariantes, (logarithmiques ou non), on a ;

$$\sum_j q_j \frac{\partial L_d D_\alpha I(p||q)}{\partial q_j} = 0 \tag{64}$$

Par comparaison, on rappelle ici les résultats correspondant à l'application du Logarithme naturel sur la " $\alpha$  divergence" invariante (59) ; ces résultats ont été

mentionnés dans un travail antérieur précité.  
Compte tenu de la propriété (17), il vient :

$$LD_\alpha I(p||q) = T (\log A - \log X - \log Y) \quad (65)$$

D'ou :

$$-\frac{\partial LD_\alpha I(p||q)}{\partial q_j} = -T \left( \frac{1}{A} \frac{\partial A}{\partial q_j} - \frac{1}{X} \frac{\partial X}{\partial q_j} - \frac{1}{Y} \frac{\partial Y}{\partial q_j} \right) \quad (66)$$

Compte tenu de (60) il vient :

$$-\frac{\partial LD_\alpha I(p||q)}{\partial q_j} = -\frac{1}{\alpha} \left[ -\frac{1}{X} \left( \sum_i p_i^{\frac{1}{\alpha}} q_i^{1-\alpha} \right)^{\frac{1}{\alpha}-1} p_j^\alpha q_j^{-\alpha} + \frac{1}{Y} \left( \sum_i q_i \right)^{-\frac{1}{\alpha}} \right] \quad (67)$$

Et finalement, en explicitant "X" et "Y" :

$$-\frac{\partial LD_\alpha I(p||q)}{\partial q_j} = -\frac{1}{\alpha} \left[ \left( \sum_i q_i \right)^{-1} - \left( \sum_i p_i^\alpha q_i^{1-\alpha} \right)^{-1} p_j^\alpha q_j^{-\alpha} \right] \quad (68)$$

Cette expression peut aussi être obtenue à partir de la relation (61) en suivant la méthode indiquée dans l'Annexe 1.

La décomposition de cette expression sous la forme  $U - V$  change selon le signe de " $\alpha$ ".

Comme pour toutes les divergences invariantes, on aura :

$$\sum_j q_j \frac{\partial LD_\alpha I(p||q)}{\partial q_j} = 0 \quad (69)$$

### 6.3 Beta ( $\beta$ ) divergence.

Cette divergence déjà analysée dans [18], [19] et dans les travaux qui y sont cités [2],[8],[9] s'écrit sous la forme :

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

Elle peut s'écrire sous la forme équivalente :

$$D_\beta(p||q) = \frac{1}{\beta(\beta-1)} \sum_i p_i^\beta - \frac{1}{\beta-1} \sum_i p_i q_i^{\beta-1} + \frac{1}{\beta} q_i^\beta \quad (71)$$

Ainsi qu'il est mentionné dans [10], cette divergence est convexe pour  $1 < \beta \leq 2$ , et on peut ajouter qu'elle est pseudo-convexe dans les autres cas.

La décomposition de cette divergence sous la forme  $A - B$  avec  $A > 0$  et  $B > 0$  implique de distinguer 3 cas, selon les valeurs du paramètre " $\beta$ ".



**\*Cas 1 :  $0 < \beta < 1$  :**

$$A = \frac{1}{1-\beta} \sum_i p_i q_i^{\beta-1} + \frac{1}{\beta} q_i^\beta ; \quad B = \frac{1}{\beta(1-\beta)} \sum_i p_i^\beta \quad (72)$$

En tenant compte de la séparabilité des termes  $A(p, q)$  et  $B(p, q)$ , il vient :

$$\begin{aligned} \frac{\partial A}{\partial q_j} &= \frac{\partial A_j}{\partial q_j} = -p_j q_j^{\beta-2} + q_j^{\beta-1} \\ \frac{\partial B}{\partial q_j} &= \frac{\partial B_j}{\partial q_j} = 0 \end{aligned} \quad (73)$$

Par application du Logarithme déformé sur chaque terme de la différence, on obtient (32) dont le gradient par rapport à "q" est donné par (33).

D'où l'expression de l'opposé du gradient par rapport à "q" :

$$-\frac{\partial L_d D_\beta(p||q)}{\partial q_j} = \left[ \frac{a-1}{a-b} A^{a-2} - \frac{b-1}{a-b} A^{b-2} \right] \left[ p_j q_j^{\beta-2} - q_j^{\beta-1} \right] \quad (74)$$

La décomposition de l'opposé du gradient sous la forme  $U - V$  avec  $U > 0$  et  $V > 0$  s'écrit :

$$\begin{aligned} U_j &= \left[ \frac{a-1}{a-b} A^{a-2} - \frac{b-1}{a-b} A^{b-2} \right] p_j q_j^{\beta-2} \\ V_j &= \left[ \frac{a-1}{a-b} A^{a-2} - \frac{b-1}{a-b} A^{b-2} \right] q_j^{\beta-1} \end{aligned} \quad (75)$$

**\*Cas 2 :  $\beta > 1$  :**

$$A = \frac{1}{\beta(\beta-1)} \sum_i p_i^\beta + \frac{1}{\beta} q_i^\beta ; \quad B = \frac{1}{\beta-1} \sum_i p_i q_i^{\beta-1} \quad (76)$$

En tenant compte de la séparabilité des termes  $A(p, q)$  et  $B(p, q)$ , il vient :

$$\begin{aligned} \frac{\partial A}{\partial q_j} &= \frac{\partial A_j}{\partial q_j} = q_j^{\beta-1} \\ \frac{\partial B}{\partial q_j} &= \frac{\partial B_j}{\partial q_j} = p_j q_j^{\beta-2} \end{aligned} \quad (77)$$

Par application du Logarithme déformé sur chaque terme de la différence, on obtient (32) dont le gradient par rapport à "q" est donné par (33).

D'où l'expression de l'opposé du gradient par rapport à "q" :

$$-\frac{\partial L_d D_\beta(p||q)}{\partial q_j} = - \left[ \frac{a-1}{a-b} A^{a-2} - \frac{b-1}{a-b} A^{b-2} \right] q_j^{\beta-1} + \left[ \frac{a-1}{a-b} B^{a-2} - \frac{b-1}{a-b} B^{b-2} \right] p_j q_j^{\beta-2} \quad (78)$$

La décomposition de l'opposé du gradient sous la forme  $U - V$  avec  $U > 0$  et  $V > 0$  s'écrit :

$$\begin{aligned} U_j &= \left[ \frac{a-1}{a-b} B^{a-2} - \frac{b-1}{a-b} B^{b-2} \right] p_j q_j^{\beta-2} \\ V_j &= \left[ \frac{a-1}{a-b} A^{a-2} - \frac{b-1}{a-b} A^{b-2} \right] q_j^{\beta-1} \end{aligned} \quad (79)$$

**\*Cas 3 :**  $\beta < 0$ . Cette situation n'est pas classique, mais on doit néanmoins la considérer.

On a ici :

$$A = \frac{1}{\beta(\beta-1)} \sum_i p_i^\beta - \frac{1}{\beta-1} \sum_i p_i q_i^{\beta-1} ; \quad B = -\frac{1}{\beta} q_i^\beta \quad (80)$$

En tenant compte de la séparabilité des termes  $A(p, q)$  et  $B(p, q)$ , il vient :

$$\begin{aligned} \frac{\partial A}{\partial q_j} &= \frac{\partial A_j}{\partial q_j} = -p_j q_j^{\beta-2} \\ \frac{\partial B}{\partial q_j} &= \frac{\partial B_j}{\partial q_j} = -q_j^{\beta-1} \end{aligned} \quad (81)$$

Par application du Logarithme déformé sur chaque terme de la différence, on obtient (32) dont le gradient par rapport à "q" est donné par (33).

D'où l'expression de l'opposé du gradient par rapport à "q" :

$$-\frac{\partial L_d D_\beta(p||q)}{\partial q_j} = \left[ \frac{a-1}{a-b} A^{a-2} - \frac{b-1}{a-b} A^{b-2} \right] p_j q_j^{\beta-2} - \left[ \frac{a-1}{a-b} B^{a-2} - \frac{b-1}{a-b} B^{b-2} \right] q_j^{\beta-1} \quad (82)$$

La décomposition de l'opposé du gradient sous la forme  $U - V$  avec  $U > 0$  et  $V > 0$  s'écrit :

$$\begin{aligned} U_j &= \left[ \frac{a-1}{a-b} A^{a-2} - \frac{b-1}{a-b} A^{b-2} \right] p_j q_j^{\beta-2} \\ V_j &= \left[ \frac{a-1}{a-b} B^{a-2} - \frac{b-1}{a-b} B^{b-2} \right] q_j^{\beta-1} \end{aligned} \quad (83)$$

#### 6.4 Beta ( $\beta$ ) divergence invariante.

Cette divergence a été étudiée dans un précédent travail déjà cité [18], [19]. En introduisant son facteur d'invariance nominal qui est calculable explicitement et s'écrit :

$$K_{0,\beta} = \frac{\sum_i p_i q_i^{\beta-1}}{\sum_i q_i^\beta} \quad (84)$$

On obtient la forme invariante :

$$D_\beta I(p||q) = \underbrace{\frac{1}{\beta(\beta-1)}}_T \left[ \underbrace{\sum_i p_i^\beta}_A - \underbrace{\left( \sum_i p_i q_i^{\beta-1} \right)^\beta}_X \times \underbrace{\left( \sum_i q_i^\beta \right)^{1-\beta}}_Y \right] \quad (85)$$

Les termes "X" et "Y" n'étant pas séparables, le calcul des dérivées partielles conduit à :

$$\begin{aligned}\frac{\partial A}{\partial q_j} &= 0 \\ \frac{\partial X}{\partial q_j} &= \beta(\beta-1) \left( \sum_i p_i q_i^{\beta-1} \right)^{\beta-1} p_j q_j^{\beta-2} \\ \frac{\partial Y}{\partial q_j} &= \beta(1-\beta) \left( \sum_i q_i^\beta \right)^{-\beta} \cdot q_j^{\beta-1}\end{aligned}\tag{86}$$

Par application du Logarithme déformé, en suivant le développement de la section (5.3.3), on obtient l'expression de la divergence (39) et celle du gradient par rapport à "q" (40).

L'opposé du gradient s'écrit avec (86) :

$$-\frac{\partial L_d D_\beta I(p||q)}{\partial q_j} = T \left\{ \underbrace{\left[ \frac{a-1}{a-b} (X.Y)^{a-2} - \frac{b-1}{a-b} (X.Y)^{b-2} \right]}_{Z>0} \frac{\partial(X.Y)}{\partial q_j} \right\} \tag{87}$$

Avec :

$$\frac{\partial(X.Y)}{\partial q_j} = \beta(1-\beta) \left[ \left( \frac{\sum_i p_i q_i^{\beta-1}}{\sum_i q_i^\beta} \right)^\beta q_j^{\beta-1} - \left( \frac{\sum_i p_i q_i^{\beta-1}}{\sum_i q_i^\beta} \right)^{\beta-1} p_j q_j^{\beta-2} \right] \tag{88}$$

Et finalement, en tenant compte de l'expression de "T", l'opposé du gradient s'écrit :

$$-\frac{\partial L_d D_\beta I(p||q)}{\partial q_j} = Z \left[ \left( \frac{\sum_i p_i q_i^{\beta-1}}{\sum_i q_i^\beta} \right)^{\beta-1} p_j q_j^{\beta-2} - \left( \frac{\sum_i p_i q_i^{\beta-1}}{\sum_i q_i^\beta} \right)^\beta q_j^{\beta-1} \right] \tag{89}$$

La décomposition de l'opposé du gradient sous la forme  $U - V$  avec  $U > 0$  et  $V > 0$  est immédiate, et on aura comme pour toutes les divergences invariantes :

$$\sum_j q_j \frac{\partial L_d BEI(p||q)}{\partial q_j} = 0 \tag{90}$$

## 6.5 Alpha-Beta ( $\alpha\beta$ ) divergence.

Cette divergence déjà analysée dans [18], [19] et dans les travaux qui y sont cités [2],[8],[9] s'écrit :

$$D_{\alpha\beta}(p||q) = T \left[ \sum_i p_i^{\alpha+\beta-1} + \frac{\beta-1}{\alpha} \sum_i q_i^{\alpha+\beta-1} - \frac{\alpha+\beta-1}{\alpha} \sum_i p_i^\alpha q_i^{\beta-1} \right] \tag{91}$$

Avec :

$$T = \frac{1}{(\beta - 1)(\alpha + \beta - 1)} \quad (92)$$

On peut écrire de façon équivalente :

$$\begin{aligned} D_{\alpha\beta}(p||q) &= \frac{1}{(\beta - 1)(\alpha + \beta - 1)} \sum_i p_i^{\alpha+\beta-1} \\ &+ \frac{1}{\alpha(\alpha + \beta - 1)} \sum_i q_i^{\alpha+\beta-1} \\ &- \frac{1}{\alpha(\beta - 1)} \sum_i p_i^\alpha q_i^{\beta-1} \end{aligned} \quad (93)$$

On notera que la variable " $\beta$ " de la référence [9] est remplacée par  $(\beta - 1)$  dans notre expression.

En écrivant cette divergence sous la forme  $A - B$ ,  $A > 0$ ,  $B > 0$  et en appliquant le Logarithme déformé, il vient :

$$L_d D_{\alpha\beta}(p||q) = \frac{1}{a - b} [A^{a-1} - A^{b-1} - (B^{a-1} - B^{b-1})] \quad (94)$$

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

$$\begin{aligned} -\frac{\partial L_d D_{\alpha\beta}(p||q)}{\partial q_j} &= - \left[ \underbrace{\left( \frac{a-1}{a-b} A^{a-2} - \frac{b-1}{a-b} A^{b-2} \right)}_{Z_A > 0} \frac{\partial A}{\partial q_j} \right] \\ &+ \left[ \underbrace{\left( \frac{a-1}{a-b} B^{a-2} - \frac{b-1}{a-b} B^{b-2} \right)}_{Z_B > 0} \frac{\partial B}{\partial q_j} \right] \end{aligned} \quad (95)$$

La décomposition de la divergence  $D_{\alpha\beta}(p||q)$  sous forme  $(A - B)$  avec  $A$  et  $B$  positifs dépend des signes de  $\alpha$ ,  $\beta - 1$  et de  $\alpha + \beta - 1$ ; cela implique une analyse particulière.

**Si  $\alpha > 0$ , quatre situations peuvent se présenter :**

Pour chacune des situations considérées, la décomposition de l'opposé du gradient sous la forme  $U - V$  avec  $U > 0$  et  $V > 0$  sont indiquées.

**\*1 -  $\alpha > 0$ ,  $\beta - 1 > 0$  donc  $\alpha + \beta - 1 > 0$ , on a alors :**

$$\begin{aligned} A &= \frac{1}{(\beta - 1)(\alpha + \beta - 1)} \sum_i p_i^{\alpha+\beta-1} + \frac{1}{\alpha(\alpha + \beta - 1)} \sum_i q_i^{\alpha+\beta-1} \\ B &= \frac{1}{\alpha(\beta - 1)} \sum_i p_i^\alpha q_i^{\beta-1} \end{aligned} \quad (96)$$

On aura alors :

$$\begin{aligned}\frac{\partial A}{\partial q_j} &= \frac{1}{\alpha} q_j^{\alpha+\beta-2} \\ \frac{\partial B}{\partial q_j} &= \frac{1}{\alpha} p_j^\alpha q_j^{\beta-2}\end{aligned}\tag{97}$$

Ainsi :

$$-\frac{\partial L_d D_{\alpha\beta}(p||q)}{\partial q_j} = \underbrace{\frac{Z_B}{\alpha} p_j^\alpha q_j^{\beta-2}}_{U_j} - \underbrace{\frac{Z_A}{\alpha} q_j^{\alpha+\beta-2}}_{V_j}\tag{98}$$

**\*2** -  $\alpha > 0$ ,  $\beta - 1 < 0$  et  $\alpha + \beta - 1 > 0$ , on a alors :

$$\begin{aligned}A &= \frac{1}{\alpha(\alpha + \beta - 1)} \sum_i q_i^{\alpha+\beta-1} + \frac{1}{\alpha(1 - \beta)} \sum_i p_i^\alpha q_i^{\beta-1} \\ B &= \frac{1}{(1 - \beta)(\alpha + \beta - 1)} \sum_i p_i^{\alpha+\beta-1}\end{aligned}\tag{99}$$

On aura alors :

$$\begin{aligned}\frac{\partial A}{\partial q_j} &= \frac{1}{\alpha} q_j^{\alpha+\beta-2} - \frac{1}{\alpha} p_j^\alpha q_j^{\beta-2} \\ \frac{\partial B}{\partial q_j} &= 0\end{aligned}\tag{100}$$

En conséquence :

$$-\frac{\partial L_d D_{\alpha\beta}(p||q)}{\partial q_j} = \underbrace{\frac{Z_A}{\alpha} p_j^\alpha q_j^{\beta-2}}_{U_j} - \underbrace{\frac{Z_A}{\alpha} q_j^{\alpha+\beta-2}}_{V_j}\tag{101}$$

**\*3** -  $\alpha > 0$ ,  $\beta - 1 > 0$  et  $\alpha + \beta - 1 < 0$ , ce cas n'est pas à considérer, car il est de toute évidence impossible.

**\*4** -  $\alpha > 0$ ,  $\beta - 1 < 0$  et  $\alpha + \beta - 1 < 0$ , on a alors :

$$\begin{aligned}A &= \frac{1}{(\beta - 1)(\alpha + \beta - 1)} \sum_i p_i^{\alpha+\beta-1} - \frac{1}{\alpha(\beta - 1)} \sum_i p_i^\alpha q_i^{\beta-1} \\ B &= \frac{1}{\alpha(1 - \alpha - \beta)} \sum_i q_i^{\alpha+\beta-1}\end{aligned}\tag{102}$$

On aura alors :

$$\begin{aligned}\frac{\partial A}{\partial q_j} &= -\frac{1}{\alpha} p_j^\alpha q_j^{\beta-2} \\ \frac{\partial B}{\partial q_j} &= -\frac{1}{\alpha} q_j^{\alpha+\beta-2}\end{aligned}\tag{103}$$

Donc :

$$-\frac{\partial L_d D_{\alpha\beta}(p||q)}{\partial q_j} = \underbrace{\frac{Z_A}{\alpha} p_j^\alpha q_j^{\beta-2}}_{U_j} - \underbrace{\frac{Z_B}{\alpha} q_j^{\alpha+\beta-2}}_{V_j} \quad (104)$$

**Si, maintenant, on considère les cas correspondant à  $\alpha < 0$ , les 4 situations précédentes doivent être à nouveau examinées.**

**\*1bis** -  $\alpha < 0$ ,  $\beta - 1 > 0$  et  $\alpha + \beta - 1 > 0$  ; les expressions de  $A$ ,  $B$ ,  $\frac{\partial A}{\partial q_j}$ ,  $\frac{\partial B}{\partial q_j}$  sont analogue au (**cas \*4**) ci dessus, en revanche, compte tenu du signe de " $\alpha$ ", la décomposition de l'opposé du gradient de la divergence est modifiée et s'écrit sous la forme  $U - V$ ,  $U > 0$ ,  $V > 0$  en prenant :

$$U_j = -\frac{Z_B}{\alpha} q_j^{\alpha+\beta-2} ; \quad V_j = -\frac{Z_A}{\alpha} p_j^\alpha q_j^{\beta-2} \quad (105)$$

**\*2bis** -  $\alpha < 0$ ,  $\beta - 1 < 0$  et  $\alpha + \beta - 1 > 0$  ; ce cas est de toute évidence impossible.

**\*3bis** -  $\alpha < 0$ ,  $\beta - 1 > 0$  et  $\alpha + \beta - 1 < 0$  ; les expressions de  $A$ ,  $B$ ,  $\frac{\partial A}{\partial q_j}$ ,  $\frac{\partial B}{\partial q_j}$  sont analogue au (**cas \*2**) ci dessus, en revanche, compte tenu du signe de " $\alpha$ ", la décomposition de l'opposé du gradient de la divergence est modifiée et s'écrit sous la forme  $U - V$ ,  $U > 0$ ,  $V > 0$  en prenant :

$$U_j = -\frac{Z_A}{\alpha} q_j^{\alpha+\beta-2} ; \quad V_j = -\frac{Z_B}{\alpha} q_j^{\alpha+\beta-2} \quad (106)$$

**\*4bis** -  $\alpha < 0$ ,  $\beta - 1 < 0$  donc  $\alpha + \beta - 1 < 0$  ; les expressions de  $A$ ,  $B$ ,  $\frac{\partial A}{\partial q_j}$ ,  $\frac{\partial B}{\partial q_j}$  sont analogue au (**cas \*1**) ci dessus, en revanche, compte tenu du signe de " $\alpha$ ", la décomposition de l'opposé du gradient de la divergence est modifiée et s'écrit sous la forme  $U - V$ ,  $U > 0$ ,  $V > 0$  en prenant :

$$U_j = -\frac{Z_A}{\alpha} q_j^{\alpha+\beta-2} ; \quad V_j = -\frac{Z_B}{\alpha} p_j^\alpha q_j^{\beta-2} \quad (107)$$

## 6.6 Alpha-Beta ( $\alpha\beta$ ) divergence invariante.

Partant de la ( $\alpha\beta$ ) divergence (93), le facteur d'invariance nominal peut se calculer explicitement ; il s'écrit :

$$K_{0,\alpha\beta} = \left[ \frac{\sum_i p_i^\alpha q_i^{\beta-1}}{\sum_i q_i^{\alpha+\beta-1}} \right]^{\frac{1}{\alpha}} \quad (108)$$

En introduisant ce facteur d'invariance dans (93), on obtient la  $(\alpha\beta)$  divergence invariante déjà considérée dans un précédent travail [18], [19], qui s'écrit :

$$D_{\alpha\beta}I(p||q) = T \left[ \underbrace{\sum_i p_i^{\alpha+\beta-1}}_{A(p,q)} - \underbrace{\left( \sum_i p_i^\alpha q_i^{\beta-1} \right)^{\frac{\alpha+\beta-1}{\alpha}}}_{X(p,q)} \times \underbrace{\left( \sum_i q_i^{\alpha+\beta-1} \right)^{\frac{1-\beta}{\alpha}}}_{Y(p,q)} \right] \quad (109)$$

Avec :

$$T = \frac{1}{(\beta-1)(\alpha+\beta-1)} \quad (110)$$

Par application du Logarithme déformé, il vient :

$$L_d D_{\alpha\beta}I(p||q) = T \left\{ \frac{A^{a-1} - A^{b-1}}{a-b} - \left[ \frac{(X.Y)^{a-1} - (X.Y)^{b-1}}{a-b} \right] \right\} \quad (111)$$

Avec :

$$\begin{aligned} \frac{\partial A}{\partial q_j} &= \frac{\partial A_j}{\partial q_j} = 0 \\ \frac{\partial X}{\partial q_j} &= \frac{(\alpha+\beta-1)(\beta-1)}{\alpha} \left( \sum_i p_i^\alpha q_i^{\beta-1} \right)^{\frac{\beta-1}{\alpha}} p_j^\alpha q_j^{\beta-2} \\ \frac{\partial Y}{\partial q_j} &= \frac{(\alpha+\beta-1)(1-\beta)}{\alpha} \left( \sum_i q_i^{\alpha+\beta-1} \right)^{\frac{1-\alpha-\beta}{\alpha}} q_j^{\alpha+\beta-2} \end{aligned} \quad (112)$$

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

$$-\frac{\partial L_d D_{\alpha\beta}I(p||q)}{\partial q_j} = T \left[ \left( \frac{a-1}{a-b} \right) (X.Y)^{a-2} - \left( \frac{b-1}{a-b} \right) (X.Y)^{b-2} \right] \frac{\partial(X.Y)}{\partial q_j} \quad (113)$$

Avec :

$$\begin{aligned} \frac{\partial(X.Y)}{\partial q_j} &= \frac{(1-\beta)(\alpha+\beta-1)}{\alpha} \underbrace{\left( \frac{\sum_i p_i^\alpha q_i^{\beta-1}}{\sum_i q_i^{\alpha+\beta-1}} \right)^{\frac{\alpha+\beta-1}{\alpha}}}_{M>0} q_j^{\alpha+\beta-2} \\ &\quad - \frac{(1-\beta)(\alpha+\beta-1)}{\alpha} \underbrace{\left( \frac{\sum_i p_i^\alpha q_i^{\beta-1}}{\sum_i q_i^{\alpha+\beta-1}} \right)^{\frac{\beta-1}{\alpha}}}_{N>0} p_j^\alpha q_j^{\beta-2} \end{aligned} \quad (114)$$

D'où, finalement, en tenant compte de l'expression de "T", il vient :

$$-\frac{\partial L_d D_{\alpha\beta}I(p||q)}{\partial q_j} = -\frac{1}{\alpha} \left[ \left( \frac{a-1}{a-b} \right) (X.Y)^{a-2} - \left( \frac{b-1}{a-b} \right) (X.Y)^{b-2} \right] [M-N] \quad (115)$$

On peut constater que cette expression ne dépend pas des valeurs relatives de " $\alpha$ " et " $\beta$ " ce qui rend inutile toute discussion sur ce point.

Une discussion selon le signe de " $\alpha$ " doit toutefois intervenir, en effet, la décomposition de l'opposé du gradient sous la forme  $U - V$  avec  $U > 0$  et  $V > 0$  dépend du signe de " $\alpha$ ", et on aura :

Si  $\alpha > 0$  :

$$\begin{aligned} U_j &= \frac{1}{\alpha} \left[ \left( \frac{a-1}{a-b} \right) (X.Y)^{a-2} - \left( \frac{b-1}{a-b} \right) (X.Y)^{b-2} \right] \times N \\ V_j &= \frac{1}{\alpha} \left[ \left( \frac{a-1}{a-b} \right) (X.Y)^{a-2} - \left( \frac{b-1}{a-b} \right) (X.Y)^{b-2} \right] \times M \end{aligned} \quad (116)$$

Si  $\alpha < 0$  :

$$\begin{aligned} U_j &= -\frac{1}{\alpha} \left[ \left( \frac{a-1}{a-b} \right) (X.Y)^{a-2} - \left( \frac{b-1}{a-b} \right) (X.Y)^{b-2} \right] \times M \\ V_j &= -\frac{1}{\alpha} \left[ \left( \frac{a-1}{a-b} \right) (X.Y)^{a-2} - \left( \frac{b-1}{a-b} \right) (X.Y)^{b-2} \right] \times N \end{aligned} \quad (117)$$

## 7 Application aux divergences entre moyennes.

Ces divergences ont été développées dans le Chapitre 7 des références [18], [19], elles sont fondées sur les travaux de Tanéja [27] et de Ben-Tal [4], elles sont toutes de la forme " $D$ " (22) ; nous n'analysons ici que quelques cas spécifiques à titre d'exemple.

Il est à noter qu'une généralisation de ces divergences a été proposée par Tanéja [28] sous une forme utilisant le Logarithme déformé de Tsallis ; nous en rappellerons les points principaux dans l'Annexe 2.

### 7.1 Divergence Géométrique-Harmonique.

Avec  $0 \leq \alpha \leq 1$  :

$$\begin{aligned} MG &= \sum_i p_i^\alpha q_i^{1-\alpha} = \sum_i MG_i \\ MH &= \sum_i \frac{p_i q_i}{(1-\alpha)p_i + \alpha q_i} = \sum_i MH_i \end{aligned} \quad (118)$$

La divergence s'écrit :

$$DGH(p||q) = MG - MH \quad (119)$$

Par application du Logarithme déformé, on obtient :

$$L_d DGH(p||q) = \frac{(MG)^{a-1} - (MG)^{b-1}}{a-b} - \left[ \frac{(MH)^{a-1} - (MH)^{b-1}}{a-b} \right] \quad (120)$$



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

$$-\frac{\partial L_d DGH(p\|q)}{\partial q_j} = - \left[ \underbrace{\frac{a-1}{a-b} (MG)^{a-2} - \frac{b-1}{a-b} (MG)^{b-2}}_Z \right] \frac{\partial (MG)}{\partial q_j} + \left[ \underbrace{\frac{a-1}{a-b} (MH)^{a-2} - \frac{b-1}{a-b} (MH)^{b-2}}_Y \right] \frac{\partial (MH)}{\partial q_j} \quad (121)$$

Avec :

$$\begin{aligned} \frac{\partial (MG)}{\partial q_j} &= \frac{\partial (MG_j)}{\partial q_j} = (1-\alpha) \frac{p_j^\alpha}{q_j^\alpha} > 0 \\ \frac{\partial (MH)}{\partial q_j} &= \frac{\partial (MH_j)}{\partial q_j} = \frac{(1-\alpha) p_j^2}{[(1-\alpha) p_j + \alpha q_j]^2} > 0 \end{aligned} \quad (122)$$

La décomposition de l'opposé du gradient sous la forme  $U - V$  avec  $U > 0$  et  $V > 0$  est immédiate.

## 7.2 Divergence Géométrique-Harmonique invariante.

Pour cette divergence, le facteur d'invariance nominal n'a pas d'expression explicite, on utilisera donc le facteur d'invariance  $K^* = \frac{\sum_j p_j}{\sum_j q_j}$  ; en utilisant les variables normalisées  $\bar{p}_i = \frac{p_i}{\sum_j p_j}$  et  $\bar{q}_i = \frac{q_i}{\sum_j q_j}$ , la divergence sous forme invariante s'écrit :

$$DGHI(p\|q) = \sum_j p_j [\overline{MG} - \overline{MH}] \quad (123)$$

Avec :

$$\begin{aligned} \overline{MG} &= \sum_i \overline{MG}_i = \sum_i \bar{p}_i^\alpha \bar{q}_i^{1-\alpha} \\ \overline{MH} &= \sum_i \overline{MH}_i = \sum_i \frac{\bar{p}_i \bar{q}_i}{(1-\alpha) \bar{p}_i + \alpha \bar{q}_i} \end{aligned} \quad (124)$$

Par application du Logarithme déformé, on obtient :

$$\begin{aligned} L_d DGHI(p\|q) &= \frac{\sum_j p_j}{a-b} \left[ (\overline{MG})^{a-1} - (\overline{MG})^{b-1} \right] \\ &\quad - \frac{\sum_j p_j}{a-b} \left[ (\overline{MH})^{a-1} - (\overline{MH})^{b-1} \right] \end{aligned} \quad (125)$$

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

$$\begin{aligned}
-\frac{\partial L_d DGH I(p||q)}{\partial q_j} = & -\sum_j p_j \left[ \underbrace{\frac{a-1}{a-b} (\overline{MG})^{a-2} - \frac{b-1}{a-b} (\overline{MG})^{b-2}}_Z \right] \frac{\partial \overline{MG}}{\partial q_j} \\
& + \sum_j p_j \left[ \underbrace{\frac{a-1}{a-b} (\overline{MH})^{a-2} - \frac{b-1}{a-b} (\overline{MH})^{b-2}}_Y \right] \frac{\partial \overline{MH}}{\partial q_j}
\end{aligned} \tag{126}$$

Avec :

$$\frac{\partial \overline{MG}}{\partial q_j} = \frac{1-\alpha}{\sum_j q_j} \left[ \bar{p}_j^\alpha \bar{q}_j^{1-\alpha} - \sum_i \bar{p}_i^\alpha \bar{q}_i^{1-\alpha} \right] = \frac{1-\alpha}{\sum_j q_j} \left[ \frac{\overline{MG}_j}{\bar{q}_j} - \sum_i \overline{MG}_i \right] \tag{127}$$

Et :

$$\begin{aligned}
\frac{\partial \overline{MH}}{\partial q_j} = & \frac{1-\alpha}{\sum_j q_j} \left\{ \frac{\bar{p}_j^2}{[(1-\alpha)\bar{p}_j + \alpha\bar{q}_j]^2} - \sum_i \bar{q}_i \frac{\bar{p}_i^2}{[(1-\alpha)\bar{p}_i + \alpha\bar{q}_i]^2} \right\} \\
= & \frac{1-\alpha}{\sum_j q_j} \left[ \frac{\overline{MH}_j^2}{\bar{q}_j^2} - \sum_i \frac{\overline{MH}_i^2}{\bar{q}_i} \right]
\end{aligned} \tag{128}$$

La décomposition de l'opposé du gradient sous la forme  $U - V$ , avec  $U > 0$  et  $V > 0$  s'écrit :

$$\begin{aligned}
U_j = & (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j} \left[ Y \frac{\overline{MH}_j^2}{\bar{q}_j^2} + Z \sum_i \overline{MG}_i \right] \\
V_j = & (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j} \left[ Z \frac{\overline{MG}_j}{\bar{q}_j} + Y \sum_i \frac{\overline{MH}_i^2}{\bar{q}_i} \right]
\end{aligned} \tag{129}$$

Comme pour toutes les divergences invariantes, on aura :

$$\sum_j q_j \frac{\partial L_d DGH I(p||q)}{\partial q_j} = 0 \tag{130}$$

### 7.3 Divergence Arithmique-Géométrique.

A un facteur multiplicatif près, cette divergence est égale à la Alpha divergence développée dans la section (6.1) ; en conséquence, nous ne reviendrons pas sur ce point.

#### 7.4 Divergence Arithmique-Géométrique invariante.

Dans ce cas, bien que le facteur d'invariance nominal soit calculable explicitement et conduise à la Alpha divergence invariante déjà analysée dans la section (6.2), nous analysons dans un soucis d'homogénéité, une autre forme de divergence Arithmique-Géométrique invariante obtenue en utilisant le facteur d'invariance  $K^* = \frac{\sum_j p_j}{\sum_j q_j}$ .

Avec :

$$\overline{MG} = \sum_i \bar{p}_i^\alpha \bar{q}_i^{1-\alpha} ; \overline{MA} = \sum_i \alpha \bar{p}_i + (1-\alpha) \bar{q}_i = 1 \quad (131)$$

La divergence invariante s'écrit :

$$DAGI(p||q) = \sum_j p_j [\overline{MA} - \overline{MG}] = \sum_j p_j [1 - \overline{MG}] \quad (132)$$

Par application du Logarithme déformé, on obtient :

$$L_d DAGI(p||q) = - \sum_j p_j \left[ \frac{(\overline{MG})^{a-1} - (\overline{MG})^{b-1}}{a-b} \right] \quad (133)$$

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

$$-\frac{\partial L_d DAGI(p||q)}{\partial q_j} = \sum_j p_j \left[ \underbrace{\frac{a-1}{a-b} (\overline{MG})^{a-2} - \frac{b-1}{a-b} (\overline{MG})^{b-2}}_Z \right] \frac{\partial (\overline{MG})}{\partial q_j} \quad (134)$$

Avec :

$$\frac{\partial \overline{MG}}{\partial q_j} = \frac{(1-\alpha)}{\sum_j q_j} \left[ \bar{p}_j^\alpha \bar{q}_j^{-\alpha} - \sum_i \bar{p}_i^\alpha \bar{q}_i^{1-\alpha} \right] = \frac{(1-\alpha)}{\sum_j q_j} \left[ \frac{\overline{MG}_j}{\bar{q}_j} - \overline{MG} \right] \quad (135)$$

La décomposition de l'opposé du gradient sous la forme  $U - V$ , avec  $U > 0$  et  $V > 0$  s'écrit :

$$\begin{aligned} U_j &= (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j} Z \frac{\overline{MG}_j}{\bar{q}_j} \\ V_j &= (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j} Z \overline{MG} \end{aligned} \quad (136)$$

Comme pour toutes les divergences invariantes, on aura :

$$\sum_j q_j \frac{\partial L_d DAGI(p||q)}{\partial q_j} = 0 \quad (137)$$

## 7.5 Divergence Arithmétique-Harmonique.

Avec :

$$\begin{aligned} MA &= \sum_i \alpha p_i + (1 - \alpha) q_i = \sum_i MA_i \\ MH &= \sum_i \frac{p_i q_i}{(1 - \alpha) p_i + \alpha q_i} = \sum_i MH_i \end{aligned} \quad (138)$$

La divergence s'écrit :

$$DAH(p||q) = MA - MH \quad (139)$$

Par application du Logarithme déformé, on obtient :

$$L_d DAH(p||q) = \frac{(MA)^{a-1} - (MA)^{b-1}}{a-b} - \left[ \frac{(MH)^{a-1} - (MH)^{b-1}}{a-b} \right] \quad (140)$$

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

$$\begin{aligned} -\frac{\partial L_d DAH(p||q)}{\partial q_j} &= - \left[ \underbrace{\frac{a-1}{a-b} (MA)^{a-2} - \frac{b-1}{a-b} (MA)^{b-2}}_{Z>0} \right] \frac{\partial (MA)}{\partial q_j} \\ &\quad + \left[ \underbrace{\frac{a-1}{a-b} (MH)^{a-2} - \frac{b-1}{a-b} (MH)^{b-2}}_{Y>0} \right] \frac{\partial (MH)}{\partial q_j} \end{aligned} \quad (141)$$

Avec :

$$\begin{aligned} \frac{\partial (MA)}{\partial q_j} &= \frac{\partial (MA_j)}{\partial q_j} = (1 - \alpha) > 0 \\ \frac{\partial (MH)}{\partial q_j} &= \frac{\partial (MH_j)}{\partial q_j} = \frac{(1 - \alpha) p_j^2}{[(1 - \alpha) p_j + \alpha q_j]^2} > 0 \end{aligned} \quad (142)$$

La décomposition de l'opposé du gradient sous la forme  $U - V$ , avec  $U > 0$  et  $V > 0$  s'écrit immédiatement :

$$\begin{aligned} U_j &= \frac{(1 - \alpha) p_j^2}{[(1 - \alpha) p_j + \alpha q_j]^2} Y \\ V_j &= (1 - \alpha) Z \end{aligned} \quad (143)$$

## 7.6 Divergence Arithmétique-Harmonique invariante.

Pour cette divergence, le facteur d'invariance nominal n'a pas d'expression explicite, on utilisera donc le facteur d'invariance  $K^* = \frac{\sum_j p_j}{\sum_j q_j}$  ; en utilisant les

variables normalisées  $\bar{p}_i = \frac{p_i}{\sum_j p_j}$  et  $\bar{q}_i = \frac{q_i}{\sum_j q_j}$ , on a :

$$\overline{MH} = \sum_i \frac{\bar{p}_i \bar{q}_i}{(1-\alpha)\bar{p}_i + \alpha\bar{q}_i} ; \overline{MA} = \sum_i \alpha\bar{p}_i + (1-\alpha)\bar{q}_i = 1 \quad (144)$$

La divergence s'écrit :

$$DAHI(p||q) = \sum_j p_j (\overline{MA} - \overline{MH}) = \sum_j p_j (1 - \overline{MH}) \quad (145)$$

Par application du Logarithme déformé, on obtient :

$$L_d DAHI(p||q) = - \sum_j p_j \left[ \frac{(\overline{MH})^{a-1} - (\overline{MH})^{b-1}}{a-b} \right] \quad (146)$$

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

$$-\frac{\partial L_d DAHI(p||q)}{\partial q_j} = \sum_j p_j \left[ \underbrace{\frac{a-1}{a-b} (\overline{MH})^{a-2} - \frac{b-1}{a-b} (\overline{MH})^{b-2}}_{Z>0} \right] \frac{\partial \overline{MH}}{\partial q_j} \quad (147)$$

Avec :

$$\begin{aligned} \frac{\partial \overline{MH}}{\partial q_j} &= \frac{1-\alpha}{\sum_j q_j} \left\{ \frac{\bar{p}_j^2}{[(1-\alpha)\bar{p}_j + \alpha\bar{q}_j]^2} - \sum_i \bar{q}_i \left[ \frac{\bar{p}_i}{(1-\alpha)\bar{p}_i + \alpha\bar{q}_i} \right]^2 \right\} \\ &= \frac{1-\alpha}{\sum_j q_j} \left[ \frac{\overline{MH}_j^2}{\bar{q}_j^2} - \sum_i \frac{\overline{MH}_i^2}{\bar{q}_i} \right] \end{aligned} \quad (148)$$

Comme pour toutes les divergences invariantes, on aura :

$$\sum_j q_j \frac{\partial L_d DAHI(p||q)}{\partial q_j} = 0 \quad (149)$$

La décomposition de l'opposé du gradient sous la forme  $U - V$ , avec  $U > 0$  et  $V > 0$  s'écrit immédiatement :

$$\begin{aligned} U_j &= (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j} Z \frac{\overline{MH}_j^2}{\bar{q}_j^2} \\ V_j &= (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j} Z \sum_i \frac{\overline{MH}_i^2}{\bar{q}_i} \end{aligned} \quad (150)$$

## 8 Application aux divergences F et G.

Ces divergences présentent la particularité d'être fondées sur l'utilisation de la Divergence de Kullback-Leibler ; leurs expressions contiennent donc explicitement la fonction Logarithme naturel.

L'introduction du Logarithme déformé consiste donc simplement de remplacer dans ces expressions, le Logarithme naturel par le Logarithme déformé.

### 8.1 Divergence F.

Elle s'écrit [27], [18], [19] :

$$F(p||q) = \sum_i p_i \log \frac{p_i}{\alpha p_i + (1-\alpha) q_i} + (1-\alpha)(q_i - p_i) \quad (151)$$

On notera :

$$Z_i = \frac{p_i}{\alpha p_i + (1-\alpha) q_i} \quad (152)$$

En introduisant le Logarithme déformé, il vient :

$$L_d F(p||q) = \sum_i \frac{p_i}{a-b} \left[ (Z_i)^{a-1} - (Z_i)^{b-1} \right] + (1-\alpha)(q_i - p_i) \quad (153)$$

$$\frac{\partial L_d F(p||q)}{\partial q_j} = \left[ \frac{a-1}{a-b} p_j (Z_j)^{a-2} - \frac{b-1}{a-b} p_j (Z_j)^{b-2} \right] \frac{\partial Z_j}{\partial q_j} + (1-\alpha) \quad (154)$$

Avec :

$$\frac{\partial Z_j}{\partial q_j} = -(1-\alpha) \frac{p_j}{[\alpha p_j + (1-\alpha) q_j]^2} \quad (155)$$

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

$$-\frac{\partial L_d F(p||q)}{\partial q_j} = (1-\alpha) \left\{ \left[ \frac{a-1}{a-b} (Z_j)^a - \frac{b-1}{a-b} (Z_j)^b \right] - 1 \right\} \quad (156)$$

La décomposition de l'opposé du gradient sous la forme  $U - V$ , avec  $U > 0$  et  $V > 0$  s'écrit :

$$\begin{aligned} U_j &= (1-\alpha) \left[ \frac{a-1}{a-b} (Z_j)^a - \frac{b-1}{a-b} (Z_j)^b \right] \\ V_j &= (1-\alpha) \end{aligned} \quad (157)$$

### 8.2 Divergence F invariante.

Pour cette divergence, le facteur d'invariance nominal n'a pas d'expression explicite, on est donc conduits à utiliser (par exemple) le facteur d'invariance particulier  $K^* = \frac{\sum_j p_j}{\sum_j q_j}$ .

On obtient ainsi la divergence invariante :

$$FI(p||q) = \sum_j p_j \sum_i \bar{p}_i \log \frac{\bar{p}_i}{\alpha \bar{p}_i + (1-\alpha) \bar{q}_i} \quad (158)$$

Avec les variables normalisées  $\bar{p}_i = \frac{p_i}{\sum_j p_j}$  et  $\bar{q}_i = \frac{q_i}{\sum_j q_j}$ .

Au facteur multiplicatif  $(\sum_j p_j)$  près, on retrouve l'expression de la "F" divergence initiale qui se simplifie compte tenu de la normalisation des variables.

On notera :

$$\bar{Z}_i = \frac{\bar{p}_i}{\alpha \bar{p}_i + (1 - \alpha) \bar{q}_i} \quad (159)$$

En introduisant le Logarithme déformé, il vient :

$$L_d FI(p||q) = \sum_j p_j \sum_i \frac{\bar{p}_i}{a-b} \left[ (\bar{Z}_i)^{a-1} - (\bar{Z}_i)^{b-1} \right] \quad (160)$$

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

$$-\frac{\partial L_d FI(p||q)}{\partial q_j} = - \sum_j p_j \left[ \frac{a-1}{a-b} \sum_i \bar{p}_i (\bar{Z}_i)^{a-2} - \frac{b-1}{a-b} \sum_i \bar{p}_i (\bar{Z}_i)^{b-2} \right] \frac{\partial \bar{Z}_i}{\partial q_j} \quad (161)$$

Avec :

$$\frac{\partial \bar{Z}_i}{\partial q_j} = \frac{\partial \bar{Z}_i}{\partial \bar{q}_i} \frac{\partial \bar{q}_i}{\partial q_j} = - \left[ \frac{(1-\alpha) \bar{p}_i}{(\alpha \bar{p}_i + (1-\alpha) \bar{q}_i)^2} \right] \left[ \frac{\delta_{ij} - \bar{q}_i}{\sum_j q_j} \right] \quad (162)$$

L'opposé du gradient s'écrit finalement :

$$\begin{aligned} -\frac{\partial L_d FI(p||q)}{\partial q_j} = & - (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j} \left( \frac{a-1}{a-b} \right) \left[ \sum_i (\bar{Z}_i)^a \bar{q}_i - \bar{Z}_j^a \right] \\ & + (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j} \left( \frac{b-1}{a-b} \right) \left[ \sum_i (\bar{Z}_i)^b \bar{q}_i - \bar{Z}_j^b \right] \end{aligned} \quad (163)$$

La décomposition de l'opposé du gradient sous la forme  $U - V$ , avec  $U > 0$  et  $V > 0$  s'écrit :

$$\begin{aligned} U_j &= (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j} \left[ \left( \frac{a-1}{a-b} \right) \bar{Z}_j^a - \left( \frac{b-1}{a-b} \right) \bar{Z}_j^b \right] > 0 \\ V_j &= (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j} \left[ \left( \frac{a-1}{a-b} \right) \sum_i \bar{Z}_i^a \bar{q}_i - \left( \frac{b-1}{a-b} \right) \sum_i \bar{Z}_i^b \bar{q}_i \right] > 0 \end{aligned} \quad (164)$$

### 8.3 Divergence G.

Elle s'écrit [27], [18], [19] :

$$G(p||q) = \sum_i [\alpha p_i + (1-\alpha) q_i] \log \frac{\alpha p_i + (1-\alpha) q_i}{p_i} + (1-\alpha) (p_i - q_i) \quad (165)$$

On notera :

$$T_i = \alpha p_i + (1-\alpha) q_i \quad (166)$$

En conséquence :

$$\frac{\partial T_j}{\partial q_j} = 1 - \alpha \quad (167)$$

En introduisant le Logarithme déformé, il vient :

$$L_d G(p||q) = \sum_i \frac{T_i}{a-b} \left[ \left( \frac{T_i}{p_i} \right)^{a-1} - \left( \frac{T_i}{p_i} \right)^{b-1} \right] + (1-\alpha)(p_i - q_i) \quad (168)$$

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

$$-\frac{\partial L_d G(p||q)}{\partial q_j} = (1-\alpha) \left\{ 1 - \left[ \left( \frac{a}{a-b} \right) \left( \frac{T_j}{p_j} \right)^{a-1} - \left( \frac{b}{a-b} \right) \left( \frac{T_j}{p_j} \right)^{b-1} \right] \right\} \quad (169)$$

La décomposition de l'opposé du gradient sous la forme  $U - V$ , avec  $U > 0$  et  $V > 0$  dépendra du signe de  $(a - b)$ .

En effet, on aura :

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

$$U_j = (1-\alpha) \left[ 1 + \left( \frac{b}{a-b} \right) \left( \frac{T_j}{p_j} \right)^{b-1} \right] ; \quad V_j = \left( \frac{a}{a-b} \right) \left( \frac{T_j}{p_j} \right)^{a-1} \quad (170)$$

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

$$U_j = (1-\alpha) \left[ 1 + \left( \frac{a}{b-a} \right) \left( \frac{T_j}{p_j} \right)^{a-1} \right] ; \quad V_j = \left( \frac{b}{b-a} \right) \left( \frac{T_j}{p_j} \right)^{b-1} \quad (171)$$

#### 8.4 Divergence G invariante.

Pour cette divergence, le facteur d'invariance nominal n'a pas d'expression explicite, on est donc conduits à utiliser (par exemple) le facteur d'invariance particulier  $K^* = \frac{\sum_j p_j}{\sum_j q_j}$ .

On obtient ainsi la divergence invariante :

$$GI(p||q) = \sum_j p_j \sum_i [\alpha \bar{p}_i + (1-\alpha) \bar{q}_i] \log \frac{\alpha \bar{p}_i + (1-\alpha) \bar{q}_i}{\bar{p}_i} \quad (172)$$

Avec les variables normalisées  $\bar{p}_i = \frac{p_i}{\sum_j p_j}$  et  $\bar{q}_i = \frac{q_i}{\sum_j q_j}$ .

Au facteur multiplicatif  $(\sum_j p_j)$  près, on retrouve l'expression de la "G" divergence initiale.

On notera :

$$\bar{T}_i = \alpha \bar{p}_i + (1-\alpha) \bar{q}_i \quad (173)$$

En introduisant le Logarithme déformé, il vient :

$$L_d GI(p||q) = \frac{\sum_j p_j}{a-b} \sum_i \left[ \frac{\bar{T}_i^a}{\bar{p}_i^{a-1}} - \frac{\bar{T}_i^b}{\bar{p}_i^{b-1}} \right] \quad (174)$$



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

$$-\frac{\partial L_d GI(p||q)}{\partial q_j} = -\sum_j p_j \left[ \frac{a}{a-b} \sum_i \left( \frac{\bar{T}_i}{\bar{p}_i} \right)^{a-1} - \frac{b}{a-b} \sum_i \left( \frac{\bar{T}_i}{\bar{p}_i} \right)^{b-1} \right] \frac{\partial \bar{T}_i}{\partial q_j} \quad (175)$$

Avec :

$$\frac{\partial \bar{T}_i}{\partial q_j} = \frac{\partial \bar{T}_i}{\partial \bar{q}_i} \frac{\partial \bar{q}_i}{\partial q_j} = (1-\alpha) \frac{\delta_{ij} - \bar{q}_i}{\sum_j q_j} \quad (176)$$

D'où, finalement :

$$\begin{aligned} -\frac{\partial L_d GI(p||q)}{\partial q_j} = & -(1-\alpha) \frac{\sum_j p_j}{\sum_j q_j} \left[ \frac{a}{a-b} \left( \frac{\bar{T}_j}{\bar{p}_j} \right)^{a-1} - \frac{a}{a-b} \sum_i \left( \frac{\bar{T}_i}{\bar{p}_i} \right)^{a-1} \bar{q}_i \right] \\ & + (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j} \left[ \frac{b}{a-b} \left( \frac{\bar{T}_j}{\bar{p}_j} \right)^{b-1} - \frac{b}{a-b} \sum_i \left( \frac{\bar{T}_i}{\bar{p}_i} \right)^{b-1} \bar{q}_i \right] \end{aligned} \quad (177)$$

La décomposition de l'opposé du gradient sous la forme  $U - V$ , avec  $U > 0$  et  $V > 0$  dépendra du signe de  $(a - b)$ .

En effet, on aura :

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

$$\begin{aligned} U_j &= (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j} \left[ \frac{a}{a-b} \sum_i \left( \frac{\bar{T}_i}{\bar{p}_i} \right)^{a-1} \bar{q}_i + \frac{b}{a-b} \left( \frac{\bar{T}_j}{\bar{p}_j} \right)^{b-1} \right] \\ V_j &= (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j} \left[ \frac{b}{a-b} \sum_i \left( \frac{\bar{T}_i}{\bar{p}_i} \right)^{b-1} \bar{q}_i + \frac{a}{a-b} \left( \frac{\bar{T}_j}{\bar{p}_j} \right)^{a-1} \right] \end{aligned} \quad (178)$$

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

$$\begin{aligned} U_j &= (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j} \left[ \frac{b}{b-a} \sum_i \left( \frac{\bar{T}_i}{\bar{p}_i} \right)^{b-1} \bar{q}_i + \frac{a}{b-a} \left( \frac{\bar{T}_j}{\bar{p}_j} \right)^{a-1} \right] \\ V_j &= (1-\alpha) \frac{\sum_j p_j}{\sum_j q_j} \left[ \frac{a}{b-a} \sum_i \left( \frac{\bar{T}_i}{\bar{p}_i} \right)^{a-1} \bar{q}_i + \frac{b}{b-a} \left( \frac{\bar{T}_j}{\bar{p}_j} \right)^{b-1} \right] \end{aligned} \quad (179)$$

## 9 Algorithmique.

### 9.1 Rappels sur la méthode de construction des algorithmes.

On rappelle que dans le cas qui nous intéresse (modèle linéaire), 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 :

$$\frac{\partial D}{\partial x} = H^T \frac{\partial D}{\partial q} \quad (180)$$

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 [18] et [19].

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 :

$$x_j^{k+1} = x_j^k + \alpha_j^k x_j^k \left( -\frac{\partial D}{\partial x} \right)_j^k \quad (181)$$

Cette écriture est toujours possible.

La divergence  $D$  étant convexe par rapport à “ $x$ ”,  $\left( -\frac{\partial D}{\partial x} \right)$  est une direction de descente et le vecteur de composantes  $\left[ x_j \cdot \left( -\frac{\partial D}{\partial x} \right)_j \right]$  est une direction de descente si  $x_j > 0$ .

En conséquence, avec un estimé initial positif, pour autant que le mode opératoire assure la non-négativité des itérés successifs, un tel algorithme converge sous réserve que le pas de descente soit convenablement calculé ; ces deux points seront précisés dans ce qui suit.

A notre connaissance, à la seule exception notable de la divergence de Kullback-Leibler duale qui sera développé dans l’annexe 3, l’opposé du gradient peut toujours s’écrire sous la forme :

$$\left( -\frac{\partial D}{\partial x} \right)_j^k = U_j^k - V_j^k \quad ; \quad U_j^k > 0 \quad ; \quad V_j^k > 0 \quad (182)$$

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

$$x_j^{k+1} = x_j^k + \alpha_j^k x_j^k (U_j^k - V_j^k) \quad (183)$$

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

$$x_j^{k+1} = x_j^k + \alpha_j^k x_j^k \left( \frac{U_j^k}{V_j^k} - 1 \right) \quad (184)$$

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  $\alpha^k$  (valable pour toutes les composantes), assurant la convergence de l'algorithme est alors calculé par une méthode de recherche unidimensionnelle de type Armijo [3] (par exemple), sur l'intervalle  $[0, (\alpha^k)_{Max}]$ . Cette procédure est analysée en détails dans [18] et [19].

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 :

$$x^{k+1} = x^k + \alpha^k x^k (U^k - V^k) \quad (185)$$

**Remarque 1** : 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 :

$$x^{k+1} = x^k + \alpha^k x^k \left( \frac{U^k}{V^k} - 1 \right) \quad (186)$$

Pour cette expression, la remarque précédente s'applique.

**Remarque 2** : 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 :

$$x^{k+1} = x^k \left( \frac{U^k}{V^k} \right) \quad (187)$$

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

**Remarque 3** : Pour les divergences non invariantes, les algorithmes (185) (186) et (187) assurent seulement la non négativité de la solution.

**Remarque 4** : Les divergences invariantes prennent tout leur sens si on exige en plus que la contrainte de somme soit assurée, en effet, un algorithme de type (185) appliqué sur de telles divergences permet d'assurer la propriété :

$$\sum_l x_l^{k+1} = \sum_l x_l^k \quad (188)$$

Partant d'un estimé initial  $x^0$  tel que  $\sum_l x_l^0 = C$ , tout les estimés successifs seront de même somme.

Cependant, lorsque de telles divergences invariantes sont considérées, l'utilisation d'un algorithme pseudo multiplicatif (préconditionné) (186) ou d'un algorithme purement multiplicatif de type (187) (pour autant que sa convergence soit assurée), ne permet pas d'assurer spontanément la contrainte de somme ;

une opération supplémentaire s'impose.  
A chaque itération, on l'algorithme se déroule en 2 étapes :

\* - On calcule d'abord un estimé provisoire :

$$\tilde{x}^{k+1} = x^k + \alpha^k x^k \left( \frac{U^k}{V^k} - 1 \right) \quad (189)$$

ou bien :

$$\tilde{x}^{k+1} = x^k \left( \frac{U^k}{V^k} \right) \quad (190)$$

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

$$x^{k+1} = \frac{\tilde{x}^{k+1}}{\sum_i \tilde{x}_i^{k+1}} C \quad (191)$$

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

## 10 Annexe 1

Le logarithme déformé appliqué sur le produit  $(x.y)$  s'écrit :

$$L_d(x.y) = \frac{(x.y)^{a-1} - (x.y)^{b-1}}{a - b} \quad (192)$$

Avec par exemple  $a = 1 + \epsilon$  et  $b = 1 - \epsilon$ , on obtient la forme simplifiée :

$$LS_d(x.y) = \frac{(x.y)^\epsilon - (x.y)^{-\epsilon}}{2\epsilon} \quad (193)$$

Ce qui peut s'écrire :

$$LS_d(x.y) = \frac{\exp \epsilon \log(x.y) - \exp(-\epsilon) \log(x.y)}{2\epsilon} \quad (194)$$

En prenant le développement limité à l'ordre 1, il vient :

$$LS_d(x.y) = \frac{1 + \epsilon \log(x.y) - (1 - \epsilon \log(x.y))}{2\epsilon} \quad (195)$$

D'où :

$$LS_d(x.y) = \log(x.y) = \log x + \log y \quad (196)$$

C.Q.F.D.

## 11 Annexe 2

La généralisation des divergences entre moyennes proposée par Taneja [28] peut se résumer comme suit :  
 Considérons par exemple l'inégalité entre moyenne arithmétique pondérée  $(MA)_i = \alpha p_i + (1 - \alpha) q_i$  et la moyenne géométrique pondérée  $(MG)_i = p_i^\alpha q_i^{1-\alpha}$  ;

$$(MG)_i - (MA)_i \leq 0 \quad \Rightarrow \quad (MG)_i^{1-r} - (MA)_i^{1-r} \leq 0 \quad ; \quad r \leq 1 \quad (197)$$

Et, par suite :

$$(MG)_i^{1-r} (MA)_i^r - (MA)_i \leq 0 \quad (198)$$

L'extension au champs entier conduit à :

$$\sum_i (MG)_i^{1-r} (MA)_i^r - \sum_i (MA)_i \leq 0 \quad (199)$$

Et finalement :

$$\frac{1}{r-1} \left[ \sum_i (MG)_i^{1-r} (MA)_i^r - \sum_i (MA)_i \right] \geq 0 \quad (200)$$

A partir de là, on peut appliquer sur chacun des termes de la différence, une fonction croissante sans changer le signe de l'inégalité.

Ici, Tanéja utilise la fonction Logarithme déformé de Tsallis (7) avec la puissance  $(s-1/r-1)$  au lieu de  $t-1$ , pour obtenir la divergence :

$$D_r^s(p||q) = \frac{1}{s-1} \left\{ \left[ \sum_i (MG)_i^{1-r} (MA)_i^r \right]^{\frac{s-1}{r-1}} - \left[ \sum_i (MA)_i \right]^{\frac{s-1}{r-1}} \right\} \quad (201)$$

En fait, Taneja [28] propose quelque chose de moins général que ça, car il se limite au cas de variables de somme "1", dans ce cas, le 2° terme de la différence sera égal à "1".

## 12 Annexe 3

### 12.1 Cas particulier de la divergence de Kullback-Leibler duale.

Elle s'écrit :

$$KL(q||p) = \sum_i q_i \log \frac{q_i}{p_i} + p_i - q_i \quad (202)$$

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

$$-\frac{\partial KL(q||p)}{\partial q_j} = \log p_j - \log q_j \quad (203)$$

C'est là que le problème se pose, en effet, une décomposition de cette expression sous la forme  $U - V$  avec les inégalités strictes  $U > 0$  et  $V > 0$  est impossible.

Cette décomposition serait :

\* Si  $q_j > 1$  et  $p_j < 1$ , l'opposé du gradient est négatif,  
 $U_j = 0$  ;  $V_j = -\log p_j + \log q_j$ .

\* Si  $q_j < 1$  et  $p_j > 1$ , l'opposé du gradient est positif,  
 $U_j = \log p_j - \log q_j$  ;  $V_j = 0$ .

\* Si  $q_j > 1$  et  $p_j > 1$ ,  
 $U_j = \log p_j$  ;  $V_j = \log q_j$ .

\* Si  $q_j < 1$  et  $p_j < 1$ ,  
 $U_j = -\log q_j$  ;  $V_j = -\log p_j$ .

Ainsi, si un algorithme de la forme générale (181) peut toujours être utilisé, il n'en est pas de même pour les algorithmes impliquant une décomposition de l'opposé du gradient, tels les algorithmes pseudo multiplicatifs (186) ou bien purement multiplicatifs (187).

Un tel problème ne se pose pas (en apparence) lorsqu'on considère la forme de cette divergence obtenue en introduisant le Logarithme déformé.

En effet en remplaçant le logarithme naturel par la Logarithme déformé dans l'expression (202), on a :

$$L_d KL(q||p) = \frac{1}{a-b} \sum_i \left[ \frac{q_i^a}{p_i^{a-1}} - \frac{q_i^b}{p_i^{b-1}} \right] + \sum_i p_i - \sum_i q_i \quad (204)$$

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

$$-\frac{\partial L_d KL(q||p)}{\partial q_j} = 1 - \left[ \frac{a}{(a-b)} \frac{q_j^{a-1}}{p_j^{a-1}} - \frac{b}{(a-b)} \frac{q_j^{b-1}}{p_j^{b-1}} \right] \quad (205)$$

La décomposition de l'opposé du gradient sous la forme  $U - V$  avec  $U > 0$  et  $V > 0$  dépend du signe de  $(a - b)$ , elle s'écrit :

\* **Si**  $a - b > 0$

$$U_j = 1 + \frac{b}{a-b} \left( \frac{q_j}{p_j} \right)^{b-1} ; \quad V_j = \frac{a}{a-b} \left( \frac{q_j}{p_j} \right)^{a-1} \quad (206)$$

\* **Si**  $a - b < 0$

$$U_j = 1 + \frac{a}{b-a} \left( \frac{q_j}{p_j} \right)^{a-1} ; \quad V_j = \frac{b}{b-a} \left( \frac{q_j}{p_j} \right)^{b-1} \quad (207)$$

## 12.2 Forme invariante.

Le problème mentionné ci-dessus apparaît également pour cette divergence. Le facteur d'invariance nominal est explicitement calculable [18], [19]. Son expression est :

$$K_0 \propto \exp \left( \frac{\sum_i q_i \log \frac{p_i}{q_i}}{\sum_j q_j} \right) \quad (208)$$

La divergence invariante correspondante s'écrit :

$$KLI_n(q||p) = \sum_i p_i - K_0 \sum_i q_i \quad (209)$$

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

$$-\frac{\partial KLI_n(q||p)}{\partial q_j} = K_0 \left[ \log \frac{p_j}{q_j} - \frac{\sum_i q_i \log \frac{p_i}{q_i}}{\sum_j q_j} \right] \quad (210)$$

Ainsi, on peut observer qu'avec ce facteur d'invariance, une décomposition de cette expression sous la forme  $U - V$  avec  $U > 0$  et  $V > 0$  se heurte aux mêmes problèmes que ceux déjà signalés dans la sous section précédente pour la divergence non invariante, avec les mêmes problèmes algorithmiques.

A titre de comparaison, nous développons le cas où on utilise le facteur d'invariance  $K^* = \frac{\sum_j p_j}{\sum_j q_j}$ . Avec  $\bar{p}_i = \frac{p_i}{\sum_j p_j}$  et  $\bar{q}_i = \frac{q_i}{\sum_j q_j}$ , la divergence invariante ainsi obtenue s'écrit :

$$KLI(q||p) = \sum_j p_j \sum_i \bar{q}_i \log \frac{\bar{q}_i}{\bar{p}_i} \quad (211)$$

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

$$-\frac{\partial KLI(q||p)}{\partial q_j} = \frac{\sum_j p_j}{\sum_j q_j} \left( \sum_i \bar{q}_i \log \frac{\bar{q}_i}{\bar{p}_i} - \log \frac{\bar{q}_j}{\bar{p}_j} \right) \quad (212)$$

Dans ce cas, la décomposition de l'opposé du gradient mentionnée précédemment ne pose pas de problème compte tenu de la normalisation des variables.

On aura, en effet  $\forall j$  :

$$U_j = - \sum_i \bar{q}_i \log \bar{p}_i - \log \bar{q}_j \quad ; \quad V_j = - \sum_i \bar{q}_i \log \bar{q}_i - \log \bar{p}_j \quad (213)$$

On montre, à titre d'exemple, que ces problèmes ne se posent pas, non plus lorsqu'on introduit le Logarithme déformé dans l'expression (211).

On a alors :

$$L_d KLI(q||p) = \sum_j p_j \sum_i \frac{\bar{q}_i}{a-b} \left[ \left( \frac{\bar{q}_i}{\bar{p}_i} \right)^{a-1} - \left( \frac{\bar{q}_i}{\bar{p}_i} \right)^{b-1} \right] \quad (214)$$

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

$$-\frac{\partial L_d KLI(q||p)}{\partial q_j} = -\sum_j p_j \sum_i \left[ \frac{a}{a-b} \left( \frac{\bar{q}_i}{\bar{p}_i} \right)^{a-1} - \frac{b}{a-b} \left( \frac{\bar{q}_i}{\bar{p}_i} \right)^{b-1} \right] \frac{\partial \bar{q}_i}{\partial q_j} \quad (215)$$

Ce qui conduit, tous calculs effectués, à :

$$\begin{aligned} -\frac{\partial L_d KLI(q||p)}{\partial q_j} &= \frac{\sum_j p_j}{\sum_j q_j} \left[ \sum_i \frac{a}{a-b} \left( \frac{\bar{q}_i}{\bar{p}_i} \right)^a - \sum_i \frac{b}{a-b} \left( \frac{\bar{q}_i}{\bar{p}_i} \right)^b \right] \\ &\quad - \frac{\sum_j p_j}{\sum_j q_j} \left[ \frac{a}{a-b} \left( \frac{\bar{q}_j}{\bar{p}_j} \right)^{a-1} - \frac{b}{a-b} \left( \frac{\bar{q}_j}{\bar{p}_j} \right)^{b-1} \right] \end{aligned} \quad (216)$$

Avec cette expression la décomposition sous la forme  $U - V$  avec  $U > 0$  et  $V > 0$  ne pose aucun problème mais dépend, bien entendu, du signe de  $(a - b)$ , et on aura  $\forall j$  :

\* **Si**  $a - b > 0$

$$\begin{aligned} U_j &= \frac{\sum_j p_j}{\sum_j q_j} \left[ \sum_i \frac{a}{a-b} \frac{\bar{q}_i^a}{\bar{p}_i^{a-1}} + \frac{b}{a-b} \left( \frac{\bar{q}_j}{\bar{p}_j} \right)^{b-1} \right] \\ V_j &= \frac{\sum_j p_j}{\sum_j q_j} \left[ \sum_i \frac{b}{a-b} \frac{\bar{q}_i^b}{\bar{p}_i^{b-1}} + \frac{a}{a-b} \left( \frac{\bar{q}_j}{\bar{p}_j} \right)^{a-1} \right] \end{aligned} \quad (217)$$

\* **Si**  $a - b < 0$  :

$$\begin{aligned} U_j &= \frac{\sum_j p_j}{\sum_j q_j} \left[ \sum_i \frac{b}{b-a} \frac{\bar{q}_i^b}{\bar{p}_i^{b-1}} + \frac{a}{b-a} \left( \frac{\bar{q}_j}{\bar{p}_j} \right)^{a-1} \right] \\ V_j &= \frac{\sum_j p_j}{\sum_j q_j} \left[ \sum_i \frac{a}{b-a} \frac{\bar{q}_i^a}{\bar{p}_i^{a-1}} + \frac{b}{b-a} \left( \frac{\bar{q}_j}{\bar{p}_j} \right)^{b-1} \right] \end{aligned} \quad (218)$$

## 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.I. Amari. Alpha divergence is unique, belonging to both  $f$ -divergence and Bregman divergence classes. *Information Theory, IEEE Transactions on*, 55(11) :4925–4931, 2009.
- [3] L. Armijo. Minimization of functions having Lipschitz continuous first partial derivatives. *Pacific Journal of mathematics*, 16(1) :1–3, 1966.
- [4] A. Ben-Tal, A. Charnes, and M. Teboulle. Entropic means. *Journal of Mathematical Analysis and Applications*, 139(2) :537–551, 1989.



- [5] M. Bertero and P. Boccacci. *Introduction to inverse problems in imaging*. CRC press, 1998.
- [6] E.P. Borges and I. Roditi. A family of nonextensive entropies. *Physics Letters A*, 246(5) :399–402, 1998.
- [7] 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.
- [8] A. Cichocki and S.I. Amari. Families of  $\alpha$ - $\beta$ -and  $\gamma$ -divergences : Flexible and robust measures of similarities. *Entropy*, 12(6) :1532–1568, 2010.
- [9] A. Cichocki, S. Cruces, and S.I. Amari. Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization. *Entropy*, 13(1) :134–170, 2011.
- [10] C. Févotte and J. Idier. Algorithms for nonnegative matrix factorization with the  $\beta$ -divergence. *Neural computation*, 23(9) :2421–2456, 2011.
- [11] S. Furuichi. An axiomatic characterization of a two-parameter extended relative entropy. *Journal of mathematical physics*, 51(12) :123302, 2010.
- [12] J. Idier, editor. *Approche bayésienne pour les problèmes inverses*. Traité IC2, Série traitement du signal et de l’image, Hermès, Paris, nov. 2001.
- [13] F.H. Jackson. q-form of taylor’s theorem. *Mess. Math*, 3 :57, 1909.
- [14] F.H. Jackson. On q—definite integrals, j. *Pure Appl. Math*, 41 :193–403, 1910.
- [15] G. Kaniadakis. Statistical mechanics in the context of special relativity. *Physical review E*, 66(5) :056125, 2002.
- [16] 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.
- [17] S. Kullback and R.A. Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1) :79–86, 1951.
- [18] H. Lantéri. Divergences. divergences invariantes. applications aux problèmes inverses linéaires, nmf et déconvolution aveugle. *HAL.Archives ouvertes hal-01745256v3*, 2019.
- [19] H. Lantéri. Divergences. scale invariant divergences. applications to linear inverse problems. nmf blind deconvolution. *arXiv preprint arXiv :2003.01411*, 2020.
- [20] H. Lantéri. Deformed logarithms. associated entropic divergences. applications to linear inverse problems. inversion algorithms. *arXiv preprint arXiv :2109.12895*, 2021.
- [21] H. Lantéri. Logarithmes déformés. -divergences entropiques associées. - applications aux problèmes inverses linéaires. - algorithmes d’inversion. *HAL.Archives ouvertes hal-03355082v3*, 2022.
- [22] D.S. McAnally. q-exponential and q-gamma functions. i. q-exponential functionsa. *Journal of Mathematical Physics*, 36(1) :546–573, 1995.

- [23] D.P. Mittal. On some functional equations concerning entropy, directed divergence and inaccuracy. *Metrika*, 22(1) :35–45, 1975.
- [24] J. Naudts. Deformed exponentials and logarithms in generalized thermostatics. *Physica A : Statistical Mechanics and its Applications*, 316(1-4) :323–334, 2002.
- [25] C. E. Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3) :379–423, 1948.
- [26] B.D. Sharma and I.J. Taneja. Entropy of type  $(\alpha, \beta)$  and other generalized measures in information theory. *Metrika*, 22(1) :205–215, 1975.
- [27] I.J. Taneja. Generalized information measures and their applications. on-line book, 2001. URL [www.mtm.ufsc.br/taneja/book/book.html](http://www.mtm.ufsc.br/taneja/book/book.html).
- [28] I.J. Taneja. On generalized information measures and their applications. *Advances in Electronics and Electron Physics*, 76 :327–413, 1989.
- [29] C. Tsallis. Possible generalization of boltzmann-gibbs statistics. *Journal of statistical physics*, 52(1) :479–487, 1988.
- [30] 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.