Les structures géométriques de l'Information de Jean-Louis Koszul

FRÉDÉRIC BARBARESCO[1]

[1] THALES LAND & AIR SYSTEMS
Voie Pierre-Gilles de Gennes, F91470 Limours , France

[1]frederic.barbaresco@thalesgroup.com

Résumé - Nous allons dans cet article rendre hommage à une partie de l'œuvre du Professeur Jean-Louis Koszul en rappelant les apports fondamentaux des modèles mathématiques de ce grand algébriste et géomètre dans le domaine des sciences géométriques de l'information, qui ont de nombreuses applications dans le domaine de l'Intelligence Artificielle où les algorithmes les plus performants sont bâtis sur le gradient naturel de la géométrie de l'Information déduit de la matrice de Fisher, comme l'a montré récemment Yann Ollivier. Le Professeur Koszul a introduit une 2-forme pour l'étude des domaines bornés homogènes qui est liée à généralisation de la métrique de Fisher pour les cônes convexes homogènes.

Abstract - In this article, we will pay tribute to part of the work of Professor Jean-Louis Koszul by recalling the fundamental contributions of the mathematical models of this great algebraist and geometer in the field of geometric information sciences, which have many applications in the field of Artificial Intelligence where the most powerful algorithms are built on the natural gradient of the information geometry deduced from the Fisher matrix, as Yann Ollivier recently showed. Professor Koszul introduced a 2-form for the study of homogeneous bounded domains that is related to the generalization of the Fisher metric for homogeneous convex cones..

1 Les travaux de Jean-Louis Koszul et la Géométrie de l'Information

Les structures géométriques élémentaires découvertes par Jean-Louis Koszul constituent les fondations de la Géométrie de l'Information. Ces structures communes ont été pour la première fois mises en regard par le Professeur Hirohiko Shima. Ce socle commun a été en particulier cristallisé dans l'ouvrage de Shima de 2007 « The Geometry of Hessian Structures », qui est dédié au Professeur Koszul. L'origine de cet ouvrage fait suite au voyage de Koszul au Japon en 1964, en mission pour le gouvernement français. Koszul donna des cours sur la théorie des variétés plates à l'université d'Osaka. Hirohiko Shima était alors élève et assista à ces cours avec les professeurs Matsushima and Murakami. Ce cours fut à l'origine de la notion de structures hessiennes et le début des travaux de Hirohiko Shima. Le livre de Shima est une introduction systématique à la théorie des structures hessiennes (données par une paire d'une connexion plate *D* et d'une métrique hessienne *g*). Koszul a étudié les variétés plates dotées d'une 1-forme fermée α, telle que $D\alpha$ soit définie positive, selon laquelle $D\alpha$ est une métrique hessienne. Cependant, toutes les métriques hessiennes ne sont pas globalement de la forme $g=D\alpha$. Shima y introduit la notion de structure de Codazzi pour une paire (D,g), avec *D* une connexion sans torsion, vérifiant les équations de Codazzi $(D_X g)(Y,Z)=(D_Y g)(X,Z)$. Une structure hessienne est une structure de Codazzi pour laquelle la connexion *D* est plate. Il s'agit d'une extension de la géométrie Riemannienne. Il est alors possible de définir une connexion *D'* et une structure de Codazzi *(D',g)* duales avec $D'=\nabla-D$ avec ∇ la connexion de Levi-Civita. Pour une structure hessienne (D,g) avec $g=Dd\varphi$, la structure de Codazzi duale (D',g) est aussi une structure hessienne et $g=D'd\varphi'$, ou φ' est la transformée de Legendre de φ : $\varphi'=\sum_i x^i \frac{\partial\varphi}{\partial x^i}-\varphi$. Shima observa que la Géométrie de l'Information a pour cadre l'étude de la théorie de l'information du point de vue des connections duales, comme l'avaient initiée Fréchet, Rao et Chentsov. Une structure hessienne *(D, g)* est de type Koszul, s'il existe une 1-forme fermée ω telle que $g=D\omega$. En utilisant *D* et l'élément de volume de *g*, Koszul a introduit une 2nd forme, qui joue un rôle similaire au tenseur de Ricci pour une métrique Kählérienne. Soit υ l'élément de volume de g, on définit une 1-forme fermée α telle que $D_X\upsilon=\alpha(X)\upsilon$ et une forme bilinéaire symétrique $\gamma=D\alpha$. Les formes α et γ sont appelées 1ère et 2nd formes de Koszul pour une structure hessienne (D,g). On peut considérer les formes associées à la structure hessienne duale (D',g) par $\alpha'=-\alpha$ et $\gamma'=\gamma-2\nabla\alpha$. Dans le cas d'un cône convexe régulier homogène Ω, les formes de Koszul α et γ pour la structure hessienne canonique $(D,g=Dd\psi)$ sont données par $\alpha=d\log\psi$ et $\gamma=g$. L'élément de volume υ déterminé par g est invariant sous l'action du groupe des automorphismes G de Ω..

2 Etude de la Géométrie des cônes convexes

Jean-Louis Koszul et Ernest B. Vinberg ont introduit une métrique hessienne affinement invariante sur un cône convexe saillant Ω par l'intermédiaire d'une fonction appelée fonction caractéristique ψ. Dans la suite, Ω est un cône convexe saillant dans un espace vectoriel E de dimension finie sur R (un cône convexe est saillant s'il ne contient aucune droite affine entière). Dans l'espace dual E^* de E, Ω^* est l'ensemble des formes linéaires strictement positives sur $\overline{\Omega}-\{0\}$. Ω^* est le cône dual de Ω est aussi un cône convexe saillant. Si $\xi \in \Omega^*$, alors l'intersection $\Omega \cap \{x \in E / \langle x,\xi \rangle = 1\}$ est bornée. $G = Aut(\Omega)$ est le groupe des transformations linéaires de E qui préserve Ω (groupe des automorphismes). $G = Aut(\Omega)$ opère sur Ω^* comme suit, $\forall g \in G = Aut(\Omega), \forall \xi \in E^*$ alors $\tilde{g}.\xi = \xi \circ g^{-1}$. Koszul introduit une intégrale de type transformée de Laplace sur le cône convexe saillant dual, comme suit :

Définition de la fonction caractéristique de Koszul-Vinberg:

Soit $d\xi$ la mesure de Lebesgue sur E^*, l'intégrale suivante:

$$\psi_\Omega(x) = \int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi \quad \forall x \in \Omega \qquad (1)$$

*avec Ω^* le cône dual, est une fonction analytique sur Ω, avec $\psi_\Omega(x) \in]0,+\infty[$, appelée **fonction caractéristique de Koszul-Vinberg du cône Ω**.*

Note : Le logarithme de la fonction caractéristique est appelé « fonction barrière » pour les algorithmes d'optimisation convexe. Yurii Nesterov and Arkadii Nemirovskii ont montré dans le cadre de la théorie moderne du point intérieur, en utilisant la fonction $\Theta_\Omega(x) = \log\left(vol_n\{s \in \Omega^* / \langle s, x \rangle \le 1\}\right)$, que tous les cônes convexes dans R^n admettent des barrières auto-concordantes, en lien avec les travaux de Koszul.

La fonction caractéristique de Koszul-Vinberg a les propriétés suivantes:

- Le noyau de Bergmann de $\Omega + iR^{n+1}$ étant noté $K_\Omega(\mathrm{Re}(z))$ à une constante près où K_Ω est défini par l'intégrale: $K_\Omega(x) = \int_{\Omega^*} e^{-\langle \xi, x \rangle} \psi_{\Omega^*}(\xi)^{-1} d\xi$ (2)
- ψ_Ω est une fonction analytique définie dans l'intérieur de Ω et $\psi_\Omega(x) \to +\infty$ lorsque $x \to \partial\Omega$. Si $g \in Aut(\Omega)$ alors $\psi_\Omega(gx) = |\det g|^{-1} \psi_\Omega(x)$ et comme $tI \in G = Aut(\Omega)$ pour tout $t > 0$, nous avons :
  $$\psi_\Omega(tx) = \psi_\Omega(x) / t^n \qquad (3)$$
- ψ_Ω est strictement log convexe, c'est à dire $\phi_\Omega(x) = \log(\psi_\Omega(x))$ est strictement convexe

A partir de cette fonction caractéristique, Koszul introduisit 2 formes qui portent maintenant son nom:

***1ère forme de Koszul α** : la 1-forme différentielle*

$$\alpha = d\phi_\Omega = d \log \psi_\Omega = d\psi_\Omega / \psi_\Omega \qquad (4)$$

est invariante suivant l'ensemble des automorphismes $G = Aut(\Omega)$ de Ω. Si $x \in \Omega$ et $u \in E$ alors

$$\langle \alpha_x, u \rangle = -\int_{\Omega^*} \langle \xi, u \rangle . e^{-\langle \xi, x \rangle} d\xi \ \text{ et } \ \alpha_x \in -\Omega^* \qquad (5)$$

***2nd forme de Koszul γ** : La 2-forme symétrique différentielle*

$$\gamma = D\alpha = Dd \log \psi_\Omega \qquad (6)$$

est une forme bilinéaire symétrique définie positive sur E invariante sous l'action de $G = Aut(\Omega)$ et $D\alpha > 0$

La positivité est donnée par l'inégalité de Schwarz et :

$$Dd \log \psi_\Omega(u,v) = \int_{\Omega^*} \langle \xi, u \rangle \langle \xi, v \rangle e^{-\langle \xi, u \rangle} d\xi \qquad (7)$$

Koszul a montré qu'à partir de cette 2nd forme, il était possible d'introduire une métrique Riemannienne invariante sous l'action des automorphismes du cône:

***Métrique de Koszul**: $D\alpha$ définit une structure Riemannienne invariante par $Aut(\Omega)$, et la métrique Riemannienne est donnée par :*

$$g = Dd \log \psi_\Omega \qquad (8)$$

$$\left(Dd \log \psi(x)\right)(u) = \frac{1}{\psi(u)^2} \left[\int_{\Omega^*} F(\xi)^2 d\xi . \int_{\Omega^*} G(\xi)^2 d\xi - \left(\int_{\Omega^*} F(\xi).G(\xi) d\xi \right)^2 \right] > 0 \qquad (9)$$

avec $F(\xi) = e^{-\frac{1}{2}\langle x, \xi \rangle}$ et $G(\xi) = e^{-\frac{1}{2}\langle x, \xi \rangle} \langle u, \xi \rangle$

On peut en effet montrer la positivité via l'inégalité de Schwarz, $d \log \psi = \frac{d\psi}{\psi}$ et $Dd \log \psi = \frac{Dd\psi}{\psi} - \left(\frac{d\psi}{\psi}\right)^2$ où

$$(d\psi(x))(u) = -\int_{\Omega^*} e^{-\langle x, \xi \rangle} \langle u, \xi \rangle d\xi$$

et $\left(Dd\psi(x)\right)(u) = \int_{\Omega^*} e^{-\langle x, \xi \rangle} \langle u, \xi \rangle^2 d\xi$.

Koszul utilise le difféomorphisme suivant pour définir des coordonnées duales :

$$x^* = -\alpha_x = -d \log \psi_\Omega(x) \qquad (10)$$

avec $\langle df(x), u \rangle = D_u f(x) = \left.\frac{d}{dt}\right|_{t=0} f(x + tu)$. Quand le cône Ω est symétrique, l'application $x \mapsto x^* = -\alpha_x$ est une bijection et une isométrie avec un unique point fixe (la variété est un espace Riemannien symétrique donné par son isométrie):

$(x^*)^* = x$, $\left\langle x, x^* \right\rangle = n$ et $\psi_\Omega(x)\psi_{\Omega^*}(x^*) = cste$ (11)

x^* est caractérisé par la définition suivante $x^* = \arg\min\left\{\psi(y) / y \in \Omega^*, \langle x, y\rangle = n\right\}$ et x^* ***est le centre de gravité de la coupe transverse*** $\left\{y \in \Omega^*, \langle x, y\rangle = n\right\}$ de Ω^* :

$$x^* = \int_{\Omega^*} \xi . e^{-\langle \xi, x\rangle} d\xi \Big/ \int_{\Omega^*} e^{-\langle \xi, x\rangle} d\xi$$

$$\text{et} \quad \left\langle -x^*, h \right\rangle = d_h \log \psi_\Omega(x) \tag{12}$$

$$\left\langle -x^*, h \right\rangle = -\int_{\Omega^*} \langle \xi, h\rangle e^{-\langle \xi, x\rangle} d\xi \Big/ \int_{\Omega^*} e^{-\langle \xi, x\rangle} d\xi$$

Misha Gromov s'est aussi intéressé à ces structures. Si l'on pose $\Phi(x) = -\log \psi_\Omega(x)$, Gromov a observé que $x^* = d\Phi(x)$ est une injection où la fermeture de l'image est égale à l'enveloppe convexe du support et le volume de cette enveloppe est le volume n-dimensionnel défini par l'intégrale du déterminant du hessien de cette fonction, $\Phi(x)$, ou l'application $\Phi \mapsto M(\Phi) = \int_\Omega \det\left(Hess(\Phi(x))\right).dx$ obéit à une relation de convexité non triviale donnée par l'inégalité de Brunn-Minkowsky :

$$\left[M(\Phi_1 + \Phi_2)\right]^{1/n} \geq \left[M(\Phi_1)\right]^{1/n} + \left[M(\Phi_2)\right]^{1/n} \tag{13}$$

Ces relations ont été également abondement exploitées en physique statistique. Comme le fit le physicien Jean-Marie Souriau, il est en effet possible de définir le concept d'Entropie de Shannon via la transformée de Legendre associée à l'opposé du logarithme de cette fonction caractéristique de Koszul-Vinberg, reprenant les idées séminales de François Massieu en Thermodynamique (condisciple du Corps des Mines, c'est François Massieu qui influença Henri Poincaré qui introduisit la fonction caractéristique en Probabilité, avec une transformée de Laplace, et non pas une transformée de Fourier comme le fit ensuite Paul Levy), qui furent développées plus récemment par Roger Balian en physique quantique, en remplaçant l'Entropie de Shannon par l'Entropie de von Neumann. Nous noterons également les travaux de Jean-Leray sur les extensions de la transformée de Laplace. Partant de la fonction caractéristique de Koszul-Vinberg, il est ainsi possible d'introduire une entropie de Koszul définie comme la transformée de Legendre de la fonction $\Phi(x)$, qui est l'opposée du logarithme de la fonction caractéristique de Koszul-Vinberg (un logarithme lie la fonction caractéristique de Massieu et la fonction caractéristique de Koszul ou de Poincaré).

En partant de la fonction $\Phi(x)$ de Koszul, sa transformée de Legendre donne une fonction potentielle duale $\Phi^*(x^*)$ dans le système de coordonnées duales x^* :

$$\Phi^*(x^*) = \left\langle x, x^* \right\rangle - \Phi(x) \text{ avec } x^* = D_x\Phi \text{ et } x = D_{x^*}\Phi^* \text{ où}$$

$$\Phi(x) = -\log \psi_\Omega(x) \tag{14}$$

Concernant la transformée de Legendre, Darboux donne dans son livre une interprétation de Chasles : « *Ce qui revient suivant une remarque de M. Chasles, à substituer à la surface sa polaire réciproque par rapport à un paraboloïde* ». Remarque que l'on retrouve dans le cours « *Leçons sur le calcul des variations* », de Jacques Hadamard, repris par M.E. Vessiot, qui utilise la « figuratrice », comme polaire réciproque de la « figurative ».

Il est possible de n'exprimer cette transformée de Legendre qu'à partir du système de coordonnées duales x^*, en utilisant le fait que $x = D_{x^*}\Phi^*$. On obtient alors la célèbre équation de Clairaut :

$$\Phi^*(x^*) = \left\langle \left(D_x\Phi\right)^{-1}(x^*), x^* \right\rangle - \Phi\left[\left(D_x\Phi\right)^{-1}(x^*)\right] \quad \forall x^* \in \left\{D_x\Phi(x) / x \in \Omega\right\} \tag{15}$$

Cette équation avait été découverte par Maurice Fréchet dans son article de 1943, dans lequel il introduisait pour la $1^{\text{ère}}$ fois la borne sur la variance de tout estimateur statistique via la matrice de Fisher, borne attribuée à tort à Cramer et Rao .

Pour faire le lien entre la fonction caractéristique de Koszul-Vinberg et l'Entropie de Shannon, nous allons détailler les formules de Koszul dans les développements qui suivent. En utilisant le fait que $-\langle \xi, x\rangle = \log e^{-\langle \xi, x\rangle}$,on peut écrire:

$$-\left\langle x^*, x \right\rangle = \int_{\Omega^*} \log e^{-\langle \xi, x\rangle} . e^{-\langle \xi, x\rangle} d\xi \Big/ \int_{\Omega^*} e^{-\langle \xi, x\rangle} d\xi \tag{16}$$

et ainsi développer la transformée de Legendre, pour faire apparaître la densité à maximum d'Entropie dans l'expression de $\Phi^*(x^*)$, dont émerge naturellement l'expression classique de l'Entropie de Shannon. Dans l'équation suivante, $p_x(\xi) = e^{-\langle \xi, x\rangle} \Big/ \int_{\Omega^*} e^{-\langle \xi, x\rangle} d\xi$ joue le rôle de la densité à maximum d'Entropie comme l'introduisit Jaynes.

$$\Phi^*(x^*) = \langle x, x^* \rangle - \Phi(x)$$
$$\Phi^*(x^*) = -\int_{\Omega^*} \log e^{-\langle \xi, x \rangle} . e^{-\langle \xi, x \rangle} d\xi \Big/ \int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi + \log \int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi$$
$$\Phi^*(x^*) = \left[\left(\int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi \right) . \log \int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi - \int_{\Omega^*} \log e^{-\langle \xi, x \rangle} . e^{-\langle \xi, x \rangle} d\xi \right] \Big/ \int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi$$
$$\Phi^*(x^*) = \left[\log \int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi - \int_{\Omega^*} \log e^{-\langle \xi, x \rangle} . \frac{e^{-\langle \xi, x \rangle}}{\int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi} d\xi \right]$$
$$\Phi^*(x^*) = \left[\log \int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi . \left(\int_{\Omega^*} \frac{e^{-\langle \xi, x \rangle}}{\int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi} d\xi \right) - \int_{\Omega^*} \log e^{-\langle \xi, x \rangle} . \frac{e^{-\langle \xi, x \rangle}}{\int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi} d\xi \right] \quad (17)$$
$$\text{with } \int_{\Omega^*} \frac{e^{-\langle \xi, x \rangle}}{\int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi} d\xi = 1$$
$$\Phi^*(x^*) = \left[-\int_{\Omega^*} \frac{e^{-\langle \xi, x \rangle}}{\int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi} . \log \left(\frac{e^{-\langle \xi, x \rangle}}{\int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi} \right) d\xi \right]$$

Nous appellerons l'Entropie associée, ***l'Entropie de Koszul***:

$$\Phi^* = -\int_{\Omega^*} p_x(\xi) \log p_x(\xi) d\xi \quad (18)$$

avec $p_x(\xi) = e^{-\langle \xi, x \rangle} \Big/ \int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi = e^{-\langle x, \xi \rangle - \log \int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi} = e^{-\langle x, \xi \rangle + \Phi(x)}$

et $x^* = \int_{\Omega^*} \xi . p_x(\xi) d\xi$ (19)

Cette densité de Koszul $p_x(\xi) = \dfrac{e^{-\langle \xi, x \rangle}}{\int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi}$ permet d'écrire le logarithme de la densité à maximum d'Entropie :

$$\log p_x(\xi) = -\langle x, \xi \rangle - \log \int_{\Omega^*} e^{-\langle \xi, x \rangle} d\xi = -\langle x, \xi \rangle + \Phi(x) \quad (20)$$

dont on déduit directement en prenant l'espérance:

$$E_\xi\left[-\log p_x(\xi)\right] = \langle x, x^* \rangle - \Phi(x) \quad (21)$$

Nous notons $x^* = \hat{\xi}$, pour donner à cette variable duale son sens de moment statistique $\hat{\xi} = \int_{\Omega^*} \xi . p_{\hat{\xi}}(\xi) d\xi$. Pour se ramener au paramétrage lié au 1[er] moment, on inverse la relation $\hat{\xi} = \Theta(x) = \dfrac{d\Phi(x)}{dx}$, en écrivant $x = \Theta^{-1}(\hat{\xi})$ la fonction inverse (qui est donnée par la transformée de Legendre):

$$p_{\hat{\xi}}(\xi) = \frac{e^{-\left\langle \xi \Theta^{-1}(\hat{\xi}) \right\rangle}}{\int_{\Omega^*} e^{-\left\langle \xi \Theta^{-1}(\hat{\xi}) \right\rangle} d\xi} \quad \text{avec} \quad \hat{\xi} = \int_{\Omega^*} \xi . p_{\hat{\xi}}(\xi) d\xi \quad \text{et}$$
$$\Phi(x) = -\log \int_{\Omega^*} e^{-\langle x, \xi \rangle} d\xi \quad (22)$$

Grâce à l'expression de la fonction caractéristique de Koszul-Vinberg et la forme de Cartan-Killing pour définir le produit de dualité, on peut exprimer la densité à maximum d'Entropie dans des cas très généraux. Par exemple, si on applique ces formules pour le cône Ω (auto-dual : $\Omega^* = \Omega$) des matrices symétriques définies positives $Sym^+(n)$, la forme de Cartan-Killing nous donne le produit de dualité :

$$\langle \eta, \xi \rangle = Tr\left(\eta^T \xi\right), \ \forall \eta, \xi \in Sym^+(n) = \left\{ \xi / \xi^T = \xi, \xi > 0 \right\} \quad (23)$$

La densité à maximum d'entropie est alors donnée directement par les expressions précédentes :

$$\psi_\Omega(\beta) = \int_{\Omega^*} e^{-\langle \beta, \xi \rangle} d\xi = \det(\beta)^{-\frac{n+1}{2}} \psi_\Omega(I_d)$$
$$\text{et} \quad \hat{\xi} = \frac{\partial \Phi(\beta)}{\partial \beta} = \frac{\partial(-\log \psi_\Omega(\beta))}{\partial \beta} = \frac{n+1}{2} \beta^{-1} \quad (24)$$

dont on déduit l'expression finale:

$$p_{\hat{\xi}}(\xi) = e^{-\left\langle \Theta^{-1}(\hat{\xi}), \xi \right\rangle + \Phi\left(\Theta^{-1}(\hat{\xi})\right)} \quad avec \quad \alpha = \frac{n+1}{2}$$
$$p_{\hat{\xi}}(\xi) = \psi_\Omega(I_d) . \left[\det\left(\alpha \hat{\xi}^{-1}\right) \right] . e^{-Tr\left(\alpha \hat{\xi}^{-1} \xi\right)} \quad (25)$$

Dans le cas des densités multivariés gaussiennes, comme l'a remarqué Souriau, on peut réécrire l'expression classique sous forme de Gibbs en modifiant le système de coordonnées et le produit de dualité. La densité Gaussienne multivariée est classiquement écrite suivant le système de coordonnées (m, R), avec m le vecteur moyenne, et R la matrice de covariance du vecteur z :

$$p_{\hat{\xi}}(\xi) = \frac{1}{(2\pi)^{n/2} \det(R)^{1/2}} e^{-\frac{1}{2}(z-m)^T R^{-1} (z-m)}$$
$$m = E(z), R = E\left[(z-m)(z-m)^T \right] \quad (26)$$

En introduisant un crochet de dualité par rapport à $\left(z, zz^T\right)$ et $\left(-R^{-1}m, \dfrac{1}{2}R^{-1}\right)$:

$$p_{\hat{\xi}}(\xi) = \frac{1}{(2\pi)^{n/2} \det(R)^{1/2} e^{\frac{1}{2} m^T R^{-1} m}} e^{-\left[-m^T R^{-1} z + \frac{1}{2} z^T R^{-1} z \right]} \quad (27)$$

On peut alors réécrire la densité, sous la forme de Koszul:

$$p_{\hat{\xi}}(\xi) = \frac{1}{\int_{\Omega^*} e^{-\langle \xi, \beta \rangle} . d\xi} e^{-\langle \xi, \beta \rangle} = \frac{1}{Z} e^{-\langle \xi, \beta \rangle}$$
$$\text{avec} \quad \log(Z) = n \log(2\pi) + \frac{1}{2} \log \det(R) + \frac{1}{2} m^T R^{-1} m \quad (28)$$
$$\xi = \begin{bmatrix} z \\ zz^T \end{bmatrix}, \ \hat{\xi} = E[\xi] = \begin{bmatrix} E[z] \\ E\left[zz^T\right] \end{bmatrix} = \begin{bmatrix} m \\ R + mm^T \end{bmatrix}$$
$$\beta = \begin{bmatrix} a \\ H \end{bmatrix} = \begin{bmatrix} -R^{-1} m \\ \frac{1}{2} R^{-1} \end{bmatrix} \quad \text{avec} \quad \langle \xi, \beta \rangle = Tr\left[z a^T + H^T z z^T \right]$$

3 Références

1. Selected Papers of J L Koszul, Series in Pure Mathematics, Vol. 17, World Scientific, 1994
2. Barbaresco, F. Jean-Louis Koszul and the elementary structures of Information Geometry. In Geometric Structures of Information; Nielsen, F.; Ed.; Springer: Berlin, Germany, 2018