



HAL
open science

Statistiques exactes et approchées sur les normales aléatoires

Sébastien Granger, Xavier Pennec

► **To cite this version:**

Sébastien Granger, Xavier Pennec. Statistiques exactes et approchées sur les normales aléatoires. RR-4533, INRIA. 2002. inria-00072055

HAL Id: inria-00072055

<https://inria.hal.science/inria-00072055>

Submitted on 23 May 2006

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.

*Statistiques exactes et approchées sur les normales
aléatoires*

Sébastien Granger — Xavier Pennec

N° 4533

Aout 2002

THÈME 3



*Rapport
de recherche*

Statistiques exactes et approchées sur les normales aléatoires

Sébastien Granger^{*†}, Xavier Pennec^{*}

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet Epidaure

Rapport de recherche n° 4533 — Aout 2002 — 65 pages

Résumé : Ce rapport de recherche aborde le problème de la manipulation des normales (vecteurs unitaires) et des directions (vecteurs unitaires non-orientés) aléatoires.

Nous présentons dans un premier temps les méthodes de statistiques rigoureuses, en résumant la théorie de Pennec qui montre comment définir, à partir de la distance intrinsèque, les notions statistiques usuelles (moyenne, matrice de covariance, distance de Mahalanobis et loi gaussienne) sur les variétés différentielles.

Ces notions étant souvent peu maniables, nous introduisons dans un deuxième temps un cadre théorique permettant de développer des méthodes statistiques approchées mais plus simples à manipuler. Nous montrons, théoriquement et expérimentalement, que ces méthodes sont robustes et que leur précision est bonne, de l'ordre de l'écart-type de la normale aléatoire traitée. Nous présentons des exemples de ces méthodes approchées, pour les normales et les rotations. Nous introduisons en particulier une méthode basée sur le tenseur d'une normale aléatoire, qui permet de calculer et manipuler de façon linéaire, donc facilement, l'approximation de la moyenne et de la covariance d'une normale.

Nous présentons enfin une première application de ce tenseur, l'extrapolation des normales dans un nuage de points non structurés et non-orientés issu d'une courbe ou d'une surface. Cette méthode, d'essence probabiliste, s'avère extrêmement robuste en présence de beaucoup de points aberrants. Elle est directement inspirée du tensor voting introduit de façon heuristique par Médioni, auquel elle apporte une première explication théorique, moyennant quelques corrections. Cette approche théorique permet aussi de mieux situer les limites de la méthode et savoir quelle technique utiliser lorsqu'elles sont dépassées.

Ce rapport de recherche est en fait une version préliminaire du chapitre Normales Aléatoires de la thèse de S. Granger [7].

Mots-clés : normales aléatoires, interpolation de normales, nuages de points non-structurés

^{*} Epidaure

[†] CIFFRE Ph.D fellowship from AREALL, Neuilly-sur-Seine, France

Statistics on normals

Abstract: We investigate in this research report the manipulation of random normals (unit vectors) and directions (non-oriented unit vectors).

We first give a summary of Penneec's works, who showed how to define, using the intrinsic distance, the statistical notions (mean value, covariance matrix, Mahalanobis distance and Gaussian law) on normals.

Unfortunately, these notions proved to be difficult to compute and manipulate. Then, we show how to approximate them easily, using an M-Estimate of the intrinsic distance. We show experimentally and theoretically that these methods are robust and precise and give examples with normals and rotations. We especially introduce the tensor of a random normal, which gives a linear (and thus easy) way to compute mean and covariance of a random normal.

We finally present a first application of this tensor for the interpolation of normals on a non-structured set of points sampled from a surface. Experimental results show that this probabilistic-based technic is very robust to large number of outliers. This work is in fact inspired from the tensor-voting technics introduced by Medioni et al., and provides a theoretical explanation to their success and their limits.

Key-words: random normals, random directions, interpolation of normals, non-structured set of points

Table des matières

1	Introduction	5
2	Géométrie des ensembles de primitives	6
2.1	Groupe de transformation - ensemble de primitives invariant homogène	7
2.2	Éléments de géométrie différentielle	9
2.2.1	Sous-Variété différentielle d'un espace euclidien et carte locale	10
2.2.2	Trajectoires & Courbes	12
2.2.3	Espace Tangent	12
2.3	Distance induite par une métrique riemannienne sur une sous-variété différentielle connexe géodésiquement complète - Cas de la métrique invariante et la distance intrinsèque	15
2.3.1	Métrique Riemannienne - Métrique invariante	15
2.3.2	Calcul tensoriel dans un sous-espace vectoriel	16
2.3.3	Distances et Géodésiques	17
2.4	Carte exponentielle et lieu de coupure	19
2.4.1	Cas invariant homogène : carte principale	21
2.5	Forme volume et intégration d'une fonction réelle	23
3	Statistiques intrinsèques sur les primitive aléatoire	25
3.1	Primitive aléatoire	25
3.1.1	Primitive aléatoire - application d'une fonction mesurable	25
3.1.2	Densité de probabilité	26
3.1.3	Espérance d'une observable	26
3.2	Espérance et Covariance	29
3.2.1	Espérance au sens de Fréchet	29
3.2.2	Espérance au sens de Karcher	30
3.2.3	Propriété et recherche de l'espérance au sens de Karcher	30
3.2.4	Matrice de covariance	31
3.2.5	Propagation de l'espérance et de la matrice de covariance dans le cas invariant	32
3.3	Approximation par une loi Gaussienne et distance de Mahalanobis	32
3.3.1	Entropie d'une primitive aléatoire	32
3.3.2	Lois gaussiennes	33
3.3.3	Distance de Mahalanobis	34
3.3.4	Test du χ^2	35
4	Statistiques approchées sur les normales aléatoires	35
4.1	Principe	36
4.1.1	Approximation de l'espérance	36
4.1.2	Approximation de la covariance	37
4.1.3	Caractérisation de l'espérance de Karcher	37

4.1.4	Propagation des approximations	38
4.1.5	Approximation de l'espérance d'une gaussienne	38
4.1.6	Cas général	38
4.2	Utilisation de la distance vectorielle	39
4.2.1	Calcul de l'approximation vectorielle de l'espérance	39
4.2.2	Approximation de la matrice de covariance	40
4.3	Tenseur d'une normale aléatoire	40
4.3.1	Tenseur associé à une direction	40
4.3.2	Tenseur associé à une direction aléatoire	41
4.3.3	Propriétés du tenseur associé à une direction aléatoire	42
4.3.4	Représentation sous forme d'ellipsoïdes	43
4.3.5	Exemples de tenseurs associés à une direction aléatoire	44
4.3.6	Tenseurs conditionnels	48
4.4	Utilisation de la distance tensorielle	49
4.4.1	Approximation tensorielle de l'espérance	50
4.4.2	Approximation tensorielle de la covariance	50
4.4.3	Synthèse dans le cas général	51
4.4.4	Approximation tensorielle de la distance de Mahalanobis	51
4.5	Comparaison des statistiques exactes et approchées	52
4.6	Statistiques approchées sur les rotations aléatoires	55
5	Calcul des normales pour un nuage de points non structuré	56
5.1	Vote d'un point en un autre point	56
5.2	Intégration des votes en un point	58
5.3	Extraction de la normale	58
5.4	Pertinence de la normale	58
5.5	Implémentation pratique	59
5.6	Comparaison avec la technique Médioni	60
5.7	Comparaison avec d'autres techniques	61
5.8	Résultats	61
5.8.1	Données synthétiques	61
5.8.2	Données réelles	63
6	Conclusion	63
7	Acknowledgements	64

1 Introduction

L'espace des normales (i.e. l'espace des directions) ne forme pas un espace vectoriel, mais une variété différentielle analogue à une sphère en 3D (ou une hypersphère en dimension quelconque). Le calcul des probabilités et des statistiques usuel ne peut donc s'appliquer simplement dans un tel espace.

Les travaux de Pennec [14] ont permis de dégager un cadre rigoureux permettant de généraliser les notions usuelles de probabilités et statistiques sur une telle variété différentielle. Il montre en particulier qu'il existe une généralisation cohérente dans les variétés différentielles globalement invariantes par un groupe de transformation (rigide, affine, ...). Nous présenterons un résumé plus intuitif que rigoureux de ces travaux, et détaillerons le cas des transformations rigides, des normales orientées et non-orientées.

Nous verrons alors que la moyenne et la covariance ainsi définies pour les normales sont difficiles à calculer et manipuler (le calcul de la moyenne est itératif, la représentation de la covariance varie suivant cette moyenne). Nous reviendrons alors sur deux techniques couramment utilisées pour calculer la normale "moyenne" : la somme normalisée (dans le cas orienté), et la technique du tenseur symétrique (dans le cas non-orienté). Nous justifierons ces techniques comme des approximations au deuxième ordre, et verrons en particulier que la technique du tenseur permet de calculer et manipuler à la fois la moyenne et la covariance de façon extrêmement pratique. Ensuite, nous comparerons expérimentalement ces techniques en faisant le parallèle avec un article présentant des travaux analogues sur les rotations.

Enfin, nous verrons une première application du tenseur symétrique qui permettra d'introduire le vote de tenseur. Il s'agit de l'estimation des normales pour un nuage de points non structurés et non orientés provenant d'une variété quelconque (une courbe en 2D, une courbe ou une surface en 3D). Cette technique est très fortement inspiré des travaux de Médioni sur la reconstruction de courbes et surfaces par vote de tenseur [11]. En fait, c'est en essayant de modéliser son approche efficace mais très heuristique que nous avons développé les notions de ce chapitre. Et la technique d'extrapolation des normales que l'on justifie ici théoriquement est, à quelques corrections près, identique à la sienne.

Notations Nous noterons le produit tensoriel : $a \otimes b = a.b^t$, et le produit tensoriel contracté $a \overline{\otimes} b = a^t.b$, qui est équivalent au produit scalaire $(a|b)$. On utilisera souvent l'égalité suivante : $a \overline{\otimes} b = Tr(a \otimes b)$.

Nous noterons D la dimension de l'espace ambiant E et d la dimension de l'ensemble de primitives \mathcal{M} . Nous distinguerons les points de $E = \mathbb{R}^D$, notés x , les primitives de \mathcal{M} notées \mathbf{x} , leur représentation dans une carte centrée en \mathfrak{o} , notée $\overrightarrow{o\mathbf{x}}$, et les primitives aléatoires seront notées en gras : \mathbf{x} .

Les autres notations seront introduites au fur et à mesure.

2 Géométrie des ensembles de primitives

La théorie développée par [14] s'appuie sur une métrique Riemannienne, qui permet d'identifier localement la variété à un espace euclidien. Par cet intermédiaire, on peut intégrer des fonctions réelles sur des trajectoires (en particulier la vitesse, et donc calculer la longueur de ces trajectoires, ce qui permet de définir une distance entre primitives) ou sur tout ou partie de la variété (ce qui permet de définir une densité de probabilité et calculer l'espérance d'une variable aléatoire réelle).

Un cas particulier est celui des variétés usuelles dans le monde du recalage et de la vision : l'ensemble des normales orientées \mathcal{S}_{D-1} ou non-orientées \mathcal{P}_{D-1} (analogues respectivement à une sphère et un hémisphère), l'ensemble des rotations \mathcal{SO}_D , et quelques autres, et par conséquent l'ensemble des points orientés $\mathbb{R}^D \times \mathcal{S}_{D-1}$ ou semi-orientés $\mathbb{R}^D \times \mathcal{P}_{D-1}$ et l'ensemble des transformations rigides $\mathbb{R}^D \times \mathcal{SO}_D$. Ces ensembles de primitives sont vis-à-vis du groupe des transformations rigides globalement invariants (l'image d'une primitive par une transformation rigide est une primitive) et homogènes (on peut atteindre toute primitive à partir d'une autre primitive par application d'une transformation rigide). La théorie devra, pour être cohérente, être elle aussi invariante vis-à-vis de ce groupe de transformation rigide. Il suffira pour cela de choisir une métrique Riemannienne invariante.

Nous présenterons dans cette section un résumé concis de la mise en place de cette théorie, passant parfois sous silence certaines hypothèses et complications techniques. Il serait en effet bien trop long de présenter la théorie dans tout ces détails et toute sa généralité, et le lecteur pourra se reporter à une littérature extrêmement variée pour une présentation plus complète des variétés différentielles [9; 2].

Nous commencerons par nous restreindre au cas où la variété \mathcal{M} est une sous-variété d'un espace euclidien E^1 , globalement invariante et homogène pour un groupe de transformation \mathcal{G} , les transformations rigides par exemple.

Nous verrons alors comment définir les courbes (i.e. les trajectoires), et serons ainsi amené à considérer l'espace tangent en un point de la variété, espace vectoriel où sont regroupées les dérivés des courbes (i.e. les vitesses sur les trajectoires).

Une métrique Riemannienne sur ces espaces tangent permettra alors de définir la longueur d'une courbe (en intégrant la norme de la vitesse sur l'ensemble du parcours), puis la distance entre points (la longueur de la courbe la plus courte, appelée géodésique).

En utilisant les géodésiques partant d'un point pour "aplatir" la variété sur son espace tangent en ce point, nous définirons ensuite la carte exponentielle, qui généralisera la notion de coordonnées curvilignes (définie pour les courbes) aux variétés de dimension quelconque.

On pourra enfin, à l'aide de la métrique, définir la forme volume qui permettra d'intégrer une fonction directement sur la variété ou en se ramenant, grâce à la carte exponentielle, à un espace euclidien.

1. le théorème de Whitney prouve en fait que l'on peut toujours plonger une variété différentielle de dimension d dans un espace vectoriel de dimension au plus $2d + 1$

2.1 Groupe de transformation - ensemble de primitives invariant homogène

Nous utiliserons très souvent dans la suite la notion d'invariance et d'homogénéité d'un ensemble de primitives vis à vis d'un groupe de transformation donnée. Nous allons ici détailler ce que cette notion implique.

Donnons nous donc un ensemble de transformation de E , i.e. d'applications bijectives de E dans lui-même. Cette ensemble devra être un groupe pour la composition. Nous noterons alors l'ensemble \mathcal{G} , la composition \circ , l'inversion $^{-1}$ et l'élément neutre Id . Sauf mention contraire, \mathcal{G} sera l'ensemble des transformations rigides en 3D, qui seront notées $T = (R, t)$, et appliquées à un point de la manière suivante :

$$(R, t) \star x = R.x + t$$

La composition, l'inversion et l'élément neutre sont définis comme suit :

$$(R, t) \circ (R', t') = (R.R', t + R.t')$$

$$(R, t)^{-1} = (R^t, -R^t.t)$$

$$Id = (Id, 0)$$

Cet ensemble ne sera intéressant que lorsqu'on pourra définir l'application d'une transformation f de \mathcal{G} à une primitive x de \mathcal{M} . Le résultat sera alors noté $f \star x$.

Nous nous intéresserons aux ensembles de primitives globalement invariants vis à vis de \mathcal{G} : l'image d'une primitive de \mathcal{M} par une transformation de \mathcal{G} devra être elle-même une primitive de \mathcal{M} :

$$\forall x \in \mathcal{M}, \forall f \in \mathcal{G} \quad f \star x \in \mathcal{M}$$

Ce sera en particulier le cas de du groupe de transformation \mathcal{G} lui même (en prenant $f \star g = f \circ g$), et, pour les transformations rigides, le cas de l'ensemble des points, des normales orientées ou non orientées, des repères orientés, semi-orientés ou non-orientés. Nous donnons ici les formules pour la composition à gauche avec les normales et les rotations :

$$(R, t) \star n = R.n$$

$$(R, t) \star R' = R.R'$$

Nous exigerons de plus que l'ensemble de primitives soit homogène vis à vis de \mathcal{G} : nous choisirons un point de \mathcal{M} que nous appellerons origine et noterons o , et devrons pouvoir atteindre tous les autres points de \mathcal{M} par application d'une certaine transformation à notre origine o :

$$\forall x \in \mathcal{M} \quad \exists f \in \mathcal{G} / x = f \star o$$

Notons que le choix de l'origine sera parfois canonique (l'identité s'impose dans le cas des transformations, par exemple), parfois arbitraire (dans le cas des normales, aucune n'a de propriété particulière).

Les transformations permettant d'atteindre la primitive \mathbf{x} à partir de l'origine seront appelées fonctions de placement en \mathbf{x} , et rassemblées dans l'ensemble $F_{\mathbf{x}}$ appelé coset de \mathbf{x} :

$$F_{\mathbf{x}} = \{f \in \mathcal{G} / f \star \mathbf{o} = \mathbf{x}\}$$

Nous utiliserons souvent dans les calculs une de ces fonctions de placement en \mathbf{x} , que nous appellerons fonction de placement canonique et noterons $\mathbf{f}_{\mathbf{x}}$. On veillera bien-sûr à choisir la fonction la plus simple possible.

Un cas particulier est le coset de l'origine (i.e. l'ensemble des transformations qui ne déplacent pas l'origine) noté \mathcal{H} :

$$\mathcal{H} = F_{\mathbf{o}} = \{h \in \mathcal{G} / h \star \mathbf{o} = \mathbf{o}\}$$

On montre facilement que ce coset particulier est un sous-groupe de \mathcal{G} appelé sous groupe d'isotropie. On peut retrouver les autres cosets à partir du sous groupe d'isotropie et de la fonction de placement canonique:

$$F_{\mathbf{x}} = \mathbf{f}_{\mathbf{x}} \circ \mathcal{H}$$

On peut donc identifier l'espace quotient \mathcal{G}/\mathcal{H} avec \mathcal{M} , et aussi identifier le coset $F_{\mathbf{x}} = \mathbf{f}_{\mathbf{x}} \circ \mathcal{H}$, la fonction de placement standard $\mathbf{f}_{\mathbf{x}}$ et la primitive \mathbf{x} .

Lorsque \mathcal{M} est l'ensemble des transformations lui même, il n'existe qu'une seule fonction de placement associée à une transformation \mathbf{g} qui est \mathbf{g} elle même. Le coset de \mathbf{g} est donc réduit au singleton $F_{\mathbf{g}} = \{\mathbf{g}\}$ et le sous groupe d'isotropie est réduit à l'identité: $\mathcal{H} = \{Id\}$.

Il est bien évident que tout ce qui suit doit, pour avoir un comportement satisfaisant, être invariant vis à vis du groupe de transformation. Par exemple, nous n'utiliserons que des distances invariante, qui vérifient:

$$\forall \mathbf{x}, \mathbf{y} \in \mathcal{M}, \forall f \in \mathcal{G} \quad d(f \star \mathbf{x}, f \star \mathbf{y}) = d(\mathbf{x}, \mathbf{y})$$

Cette propriété d'invariance devra en particulier être vérifiée pour le sous-groupe d'isotropie. On scindera donc en général cette propriété d'invariance en deux parties, en faisant apparaître le sous-groupe d'isotropie et les fonctions de placement canoniques:

$$\forall \mathbf{x}, \mathbf{y} \in \mathcal{M}, \forall h \in \mathcal{H} \quad d(h \star \mathbf{x}, h \star \mathbf{y}) = d(\mathbf{x}, \mathbf{y})$$

$$\forall \mathbf{x}, \mathbf{y} \in \mathcal{M}, \forall \mathbf{f}_{\mathbf{x}} \quad d(\mathbf{f}_{\mathbf{x}} \star \mathbf{x}, \mathbf{f}_{\mathbf{x}} \star \mathbf{y}) = d(\mathbf{x}, \mathbf{y})$$

On pourra en particulier se ramener à la distance à l'origine (qui peut être considérée comme une norme, notée N , bien qu'elle ne soit pas homogène²) à l'aide de la fonction de placement en \mathbf{x} :

$$d(\mathbf{f}_{\mathbf{x}}^{-1} \star \mathbf{x}, \mathbf{f}_{\mathbf{x}}^{-1} \star \mathbf{y}) = d(\mathbf{o}, \mathbf{f}_{\mathbf{x}}^{-1}(\mathbf{y})) = N(\mathbf{f}_{\mathbf{x}}^{-1}(\mathbf{y}))$$

2. Une norme est dite homogène si elle vérifie $N(\lambda \mathbf{x}) = |\lambda| N(\mathbf{x})$. Hors $\lambda \mathbf{x}$ n'est tout simplement pas défini dans la variété.

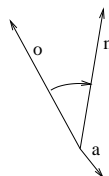
Fonction de placement pour les normales en 3D Nous avons déjà vu, que dans le cas des normales, l'application d'une transformation rigide prenait la forme suivante: $(R,t) \star \mathbf{n} = R.\mathbf{n}$. Dans le cas 3D, les rotations peuvent être définies par leur axe a (lui aussi une direction, une normale) et leur angle θ , ou directement par leur vecteur $v = \theta a$. Nous noterons une telle rotation sous la forme $R_{a,\theta}$ ou R_v .

Choisissons une origine arbitraire à l'ensemble des normales, $\mathbf{o} = (1,0,0)$ par exemple. Le sous-groupe d'isotropie est constitué de l'ensemble des transformations rigides qui laissent cette normale inchangée. La translation n'ayant aucune importance, reste à savoir quelles rotations laissent une normale inchangée. Il s'agit bien-sûr des rotations ayant pour axe cette normale, et un angle quelconque, qui seront donc notées $R_{\mathbf{o},\theta}$. En résumé, le sous-groupe d'isotropie est :

$$\mathcal{H} = \{(R_{\mathbf{o},\theta},t); t,\theta \text{ quelconque}\}$$

Reste à trouver une fonction de placement simple permettant d'atteindre la normale \mathbf{n} à partir de la normale origine \mathbf{o} . Nous prendrons une translation nulle, et la rotation autour d'un axe perpendiculaire à ces deux normales est un choix simple, car cet axe est obtenu par produit vectoriel: $a = \frac{\mathbf{o} \wedge \mathbf{n}}{\|\mathbf{o} \wedge \mathbf{n}\|}$, et l'angle de la rotation est alors l'angle entre les deux normales $(\widehat{\mathbf{o},\mathbf{n}})$. Finalement, la fonction de placement canonique est :

$$\mathbf{f}_{\mathbf{n}} = R_{\left(\frac{\mathbf{o} \wedge \mathbf{n}}{\|\mathbf{o} \wedge \mathbf{n}\|}, (\widehat{\mathbf{o},\mathbf{n}})\right)}$$



Notons que dans le cas où la normale \mathbf{n} est égale à \mathbf{o} ou à son opposée, le vecteur $\mathbf{o} \wedge \mathbf{n}$ est nul, et la formule ci dessus ne peut être appliquée. Lorsque $\mathbf{n} = \mathbf{o}$, l'identité est la fonction de placement la plus simple. Lorsque $\mathbf{n} = -\mathbf{o}$, n'importe quelle rotation d'angle 180 degrés et d'axe perpendiculaire à \mathbf{o} peut convenir :

$$\mathbf{f}_{\mathbf{o}} = Id$$

$$\mathbf{f}_{-\mathbf{o}} = R_{(a,\pi \text{ rad})} \quad \text{avec } a \perp \mathbf{o}$$

2.2 Éléments de géométrie différentielle

Nous allons maintenant voir de plus près comment aborder la géométrie non-linéaire de notre ensemble de primitives \mathcal{M} , qui sera, dans les exemples, l'ensemble des normales ou l'ensemble des transformations rigides. Nous verrons en particulier les cartes locales, qui permettent de se ramener à un ouvert de \mathbb{R}^d , plus pratique pour faire les calculs.

C'est dans cette partie que nous ferons, sans même le préciser, le plus de simplifications. Le lecteur se reportera à tout bon cours ou livre sur les variétés différentielles pour plus de rigueur [9; 2].

2.2.1 Sous-Variété différentielle d'un espace euclidien et carte locale

Nous ne présentons pas la définition générale, mais complexe, d'une variété différentielle, préférant utiliser l'espace euclidien ambiant E , qui est une variété différentielle. Il suffit donc de caractériser les sous-variétés différentielles de E . Intuitivement, il s'agit des courbes (i.e. de fonctions d'un réel t), des surfaces (i.e. des fonctions de couple de réel (u,v)), et plus généralement d'objets que l'on peut mettre sous la forme de fonction d'un n -uplet de réels, avec n fixe (c'est la dimension de la variété). La définition générale suit ces idées, de façon locale et donc sans aucune contrainte topologique (comme la connexité). Elle prend en revanche en compte les aspects analytiques, comme les possibilités de différentiation :

Définition 1 *Sous-variété différentielle - classe - dimension*

Une sous-variété différentielle de E est une partie \mathcal{M} de E localement difféomorphe en tout point \mathbf{x} à un espace euclidien $T_{\mathbf{x}}$ de dimension d .

d doit être constant et est appelé dimension de la sous-variété.

S'il est possible de trouver en tout point un difféomorphisme de classe C^k , la sous-variété est dite de classe C^k .

Une sous-variété différentielle est elle-même une variété différentielle.

Ceci signifie qu'en tout point \mathbf{x} de \mathcal{M} , il existe un difféomorphisme de classe C^k (i.e. une application bijective, C^k , et d'inverse C^k , où $k \geq 1$) défini entre un ouvert D de $T_{\mathbf{x}}$ et un voisinage \mathcal{D} de \mathbf{x} dans \mathcal{M} . Une telle application permet de définir les coordonnées d'un point de \mathcal{D} . Elle est appelée carte locale :

Définition 2 *Carte locale*

Une carte locale $\pi_{\mathbf{x}}$ de \mathcal{M} est un difféomorphisme de classe C^k défini comme suit :

$$D_{\pi_{\mathbf{x}}} \text{ ouvert de } T_{\mathbf{x}} \rightarrow \mathcal{D}_{\pi_{\mathbf{x}}} \subset \mathcal{M}$$

$$\pi_{\mathbf{x}} : \quad \vec{x\bar{y}} = \begin{bmatrix} \vec{x\bar{y}}_1 \\ \vdots \\ \vec{x\bar{y}}_d \end{bmatrix} \mapsto \mathbf{y} = \pi_{\mathbf{x}}(\vec{x\bar{y}})$$

De plus, la carte est dite centrée en \mathbf{x} lorsque $0 \in D_{\pi}$ et $\pi(0) = \mathbf{x}$.

Si $\mathbf{y} \in \mathcal{D}_{\pi_{\mathbf{x}}}$, il existe un unique $\vec{x\bar{y}} = \begin{bmatrix} \vec{x\bar{y}}_1 \\ \vdots \\ \vec{x\bar{y}}_d \end{bmatrix}_i \in D_{\pi_{\mathbf{x}}}$ tel que $\mathbf{y} = \pi_{\mathbf{x}}(\vec{x\bar{y}})$.

$\vec{x\bar{y}}_1, \dots, \vec{x\bar{y}}_d$ sont appelées coordonnées de \mathbf{y} dans la carte.

Nous nous restreindrons dans la suite aux sous-variétés suffisamment régulières, i.e. de classe C^k où $k \geq 2$.

Les cartes locales sont en pratiques très importantes. Elles permettent en-effet de se ramener à un ouvert d'un espace vectoriel, et donc d'effectuer certains calculs. Il convient cependant de bien distinguer un point de la sous-variété et ses coordonnées co- ou contra-variantes (voir 2.3.2) dans l'espace ambiant (représenté en général par une minuscule droite : $y, n, y_i, n_i, y^i, n^i, \dots$), qui ne dépendent pas de la carte utilisée, et sa représentation dans une carte locale centrée en x (notée $\vec{x}\vec{y}$, pour bien indiquer que l'on travaille dans la carte centrée en x) et ses coordonnées dans cette carte (notées $\vec{x}\vec{y}_i, \dots$), qui sont relatives à la carte utilisée.

Il faut de plus noter qu'il n'est pas toujours possible de définir une carte couvrant la totalité de la sous-variété. Ainsi, dans la cas de la sphère, on doit exclure au moins un point. C'est l'exemple typique de la cartographie de la terre, où on est obligé d'utiliser un atlas regroupant plusieurs cartes locales.

Cas d'un groupe de Lie Nous avons vu que le groupe des transformations \mathcal{G} était globalement invariant et homogène vis à vis de lui même. Un cas particulier est la cas où ce groupe est un groupe de Lie. Il possède alors une structure de variété différentielle, si bien qu'on peut utiliser dans la suite $\mathcal{M} = \mathcal{G}$. Ce sera le cas des transformations rigides.

Exemples

Espace Affine Un sous-espace affine F de E est bien-entendu une sous-variété différentielle. En-effet, si on dispose d'une base f_1, \dots, f_n de F , on peut définir très facilement une carte locale centrée en un point x de F et couvrant la totalité de l'espace avec la fonction :

$$\vec{x}\vec{y} = \begin{bmatrix} \vec{x}\vec{y}_1 \\ \vdots \\ \vec{x}\vec{y}_d \end{bmatrix} \in \mathbb{R}^d \mapsto y = x + \vec{x}\vec{y}_1 \cdot f_1 + \dots + \vec{x}\vec{y}_d \cdot f_d.$$

Courbe Les sous-variétés différentielles de E de dimension 1 et de classe C^k ne sont autres que les courbes de E de classe C^k . La courbe est souvent connexe et donnée sous forme paramétrée : $\gamma : [a, b] \rightarrow E$. On utilise en général une fonction γ ayant des propriétés de régularités suffisantes pour qu'elle soit une carte locale.

Normales Il va sans dire que l'ensemble des normales orientées et non-orientées de \mathbb{R}^D sont tous deux des sous-variétés C^∞ de dimension $d = D - 1$. Nous donnerons un exemple de carte locale un peu plus loin.

Rotations L'ensemble des rotations est lui une sous-variété C^∞ de $L(\mathbb{R}^D)$ (espace des applications linéaires de \mathbb{R}^D dans lui-même), de dimension $d = \frac{D(D-1)}{2}$ et même un groupe de Lie.

En dimension 3, on peut exhiber une carte locale centrée autour de l'identité. En effet, toute rotation est caractérisée par un axe, représenté par un vecteur unitaire n et un angle

θ , et notée $R(n, \theta)$. On peut donc représenter toute rotation par le vecteur $\theta.n$, situé dans \mathbb{R}^3 . On vérifie que ceci définit bien une carte locale, lorsqu'on impose $\theta < \pi$.

2.2.2 Trajectoires & Courbes

La notion de courbe est simple à établir sur une variété différentielle. Il convient simplement de faire la distinction entre une trajectoire et une courbe. Une trajectoire est intuitivement une fonction régulière du temps, i.e. d'un intervalle réel $[a, b]$ dans \mathcal{M} . Deux trajectoires peuvent suivre la même courbe et différer simplement par la vitesse de parcours. On définit donc les courbes et trajectoires de la manière suivante :

Définition 3 Trajectoires et courbes

Soit \mathcal{M} une variété différentielle.

Une trajectoire de \mathcal{M} est une fonction γ de $[a, b]$ dans \mathcal{M} de classe au moins C^2 , sans point fixe.

On définit une relation d'équivalence entre deux trajectoires de la façon suivante : $\gamma_1 : [a, b] \rightarrow \mathcal{M} \approx \gamma_2 : [c, d] \rightarrow \mathcal{M} \Leftrightarrow \exists f : [a, b] \rightarrow [c, d]$ croissante – surjective – $C^2/\gamma_1(t) = \gamma_2(f(t))$.

Les courbes de \mathcal{M} sont les classes d'équivalences de l'ensemble des trajectoires de \mathcal{M} pour \approx .

Si nous disposons d'une carte locale π , nous pouvons identifier la trajectoire γ dans D_π (donc dans \mathcal{M}) et $\vec{\gamma}$ sa représentation dans D_π (i.e. dans la carte locale), les deux étant reliées par $\gamma(t) = \pi(\vec{\gamma}(t))$.

2.2.3 Espace Tangent

L'espace tangent correspond à la notion de tangente dans le cas d'une courbe. Cette notion est relative à un point et s'appuie sur la dérivé de la courbe en ce point. Pour la généraliser, il faut donc se donner un point de la variété différentielle, et voir comment se répartissent les dérivés des courbes en ce point.

Considérons donc l'ensemble des trajectoires de \mathcal{M} passant par un point x (on supposera $\gamma(0) = x$ pour simplifier). Ces trajectoires peuvent être vues comme des trajectoires dans E . La notion de vecteur tangent est alors usuellement définie et il s'agit d'un vecteur de E . L'ensemble de ces vecteurs forme un sous-espace vectoriel de E , de dimension d et noté $T_x\mathcal{M}$. On peut s'en convaincre et en exhiber une base en utilisant une carte locale π telle que $x \in D_\pi$. x est alors identifié à $\vec{x} \in D_\pi$. Soit alors une trajectoire quelconque définie à l'aide de la carte

locale : $\gamma(t) = \pi \left(\begin{bmatrix} \vec{\gamma}_1(t) \\ \vdots \\ \vec{\gamma}_d(t) \end{bmatrix} \right)$, on montre facilement que $\frac{d\gamma}{dt}(0) = \sum_i \frac{\partial \pi}{\partial \vec{x}_i}(\vec{x}) \frac{d\vec{\gamma}_i}{dt}(0)$. Les

vecteurs $\frac{\partial \pi}{\partial \vec{x}_i}(\vec{x})$ forment donc une famille génératrice pour notre espace tangent. On montre aussi que le déterminant de cette famille, égal au jacobien de la carte locale, ne peut pas être

nul. Il s'agit alors d'une famille libre, donc d'une base, prouvant ainsi que l'espace tangent est bien de dimension d .

Ces quelques remarques sont rassemblées dans la définition suivante :

Définition 4 *Espace tangent*

Soit \mathbf{x} un point de \mathcal{M} .

L'ensemble $\{\dot{\gamma}(t_0); \gamma(t_0) = \mathbf{x}\}$ forme un sous-espace vectoriel de E , de dimension d (dimension de \mathcal{M}), appelé *espace tangent* à \mathcal{M} au point \mathbf{x} et noté $T_{\mathbf{x}}\mathcal{M}$.

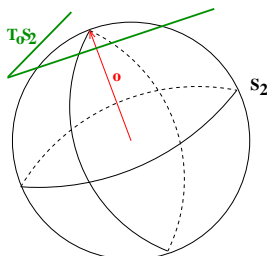
Étant donné une carte locale π et \vec{x} tel que $\pi(\vec{x}) = \mathbf{x}$, les vecteurs $\frac{\partial \pi}{\partial x_i}(\vec{x})$ forment une base de $T_{\mathbf{x}}\mathcal{M}$, appelée *base naturelle* dans la carte locale π .

Il faut remarquer que la base naturelle dans une carte locale est rarement orthonormée.

Exemples

Espace Affine Dans un espace affine F , l'espace tangent en un point \mathbf{x} n'est autre que l'espace vectoriel associé à l'espace affine en \mathbf{x} , i.e. l'ensemble $\{\mathbf{y} - \mathbf{x}; \mathbf{y} \in F\}$.

Normales L'espace des normales orientées est identifié à la sphère unité. L'espace tangent en une normale \mathbf{n} est, comme le suggère l'intuition, le plan tangent à la sphère, d'équation $(\mathbf{x}|\mathbf{n}) = 1$. Nous identifierons, pour simplifier, ce plan au plan d'équation $(\mathbf{x}|\mathbf{n}) = 0$ (i.e. le plan parallèle mais passant par l'origine de l'espace vectoriel ambiant).



Rotations Nous allons ici détailler les calculs permettant de mettre en évidence le plan tangent.

Soit donc une trajectoire $R(t)$ sur SO_D . Elle vérifie, pour tout t : $R(t) \cdot R(t)^t = R(t)^t \cdot R(t) = Id$. En dérivant :

$$\left(\frac{dR(t)}{dt} \cdot R(t)^t\right) + \left(\frac{dR(t)}{dt} \cdot R(t)^t\right)^t = 0 \quad \left(R(t)^t \cdot \frac{dR(t)}{dt}\right) + \left(R(t)^t \cdot \frac{dR(t)}{dt}\right)^t = 0$$

Les matrices $\left(\frac{dR(t)}{dt} \cdot R(t)^t\right)$ et $\left(R(t)^t \cdot \frac{dR(t)}{dt}\right)$ sont donc antisymétriques, et il existe donc deux matrices antisymétriques M_g et M_d telles que :

$$\frac{dR(t)}{dt} = M_d \cdot R = R \cdot M_g$$

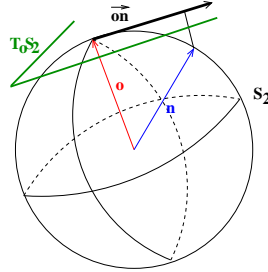
L'espace tangent à l'identité $T_{Id}SO_D$ est donc l'ensemble des matrices antisymétriques (le raisonnement précédent prouve l'inclusion, il suffit alors d'utiliser l'égalité des dimensions des deux espaces pour prouver l'égalité). On retrouve l'espace tangent en R en appliquant R à $T_{Id}SO_D$:

$$T_R SO_D = R \circ T_{Id} SO_D = \{M.R/M + M^t = 0\} = \{R.M/M + M^t = 0\} \quad (1)$$

Espace tangent et carte locale L'espace tangent en un point x est l'espace de départ idéal pour définir une carte locale centrée en x . Il faut cependant alors faire attention, car toutes les cartes ne sont pas définies avec le même espace de départ. Par exemple, $\vec{x}\vec{y}$ - qui représente y dans la carte centrée en x - sera un vecteur de $T_x\mathcal{M}$, alors que $\vec{y}\vec{x}$ - qui représente x dans la carte centrée en y - sera un vecteur de $T_y\mathcal{M}$. Toute égalité du genre $\vec{x}\vec{y} = -\vec{y}\vec{x}$ serait donc absurde...

Espace affine Pour un espace affine F , la carte locale centrée en x la plus simple définie à partir du plan tangent, i.e. de l'espace vectoriel E associé est la translation t_x qui amène l'origine de E sur x .

Normales Essayons de définir une carte locale centrée autour d'une normale o , dans le plan tangent en o . Soit une normale n . Il est naturel de représenter cette normale, dans la carte locale, par un vecteur colinéaire, et de même sens que le projeté orthogonal $\Pi_{o\perp}(n) = n - o(n|o)$ de n sur le plan tangent ($\Pi_{o\perp}$ est la projection orthogonale d'axe o). Ce vecteur sera donc de la forme : $\vec{o}\vec{n} = \lambda.\Pi_{o\perp}(n) = \lambda.(n - o(n|o))$ (avec $\lambda > 0$) :



Plusieurs possibilités sont offertes pour la longueur de ce vecteur :

1. La distance angulaire : $\|\vec{o}\vec{n}\| = (\widehat{o, n})$.
2. La distance vectorielle, définie à partir de la distance quadratique de \mathbb{R}^D : $\|\vec{o}\vec{n}\| = \|o - n\| = 2 \sin(\frac{(\widehat{o, n})}{2})$.
3. La distance tensorielle, définie par $\lambda = 1$, soit $\vec{o}\vec{n} = \Pi_{o\perp}(n) = (n - o(n|o))$ et $\|\vec{o}\vec{n}\| = \sin(\widehat{o, n})$.

Nous utiliserons dans un premier temps la distance angulaire, qui est la distance intrinsèque car elle est la seule à correspondre à une métrique invariante. Nous reviendrons sur les deux autres distances dans la partie traitant des statistiques approchées.

Rotations 3D En 3D, les matrices antisymétriques, qui forment l'espace tangent en l'identité (voir equ. 1), peuvent être écrites sous la forme :

$$\theta.S_n \text{ avec } S_n = \begin{bmatrix} 0 & -n_z & n_y \\ n_z & 0 & -n_x \\ -n_y & n_x & 0 \end{bmatrix} \text{ et } \|n\| = 1 \quad (2)$$

On a ainsi obtenu une carte locale définie à partir de l'espace tangent. On retrouve la carte définie dans la section précédente en identifiant $\theta.S_n$ et $\theta.n$.

La formule de Rodrigues permet de plus de faire le lien avec la matrice de la rotation :

$$R(n, \theta) = Id + \sin \theta.S_n + (1 - \cos \theta).S_n^2 = \cos \theta.Id + \sin \theta.S_n + (1 - \cos \theta).n.n^t \quad (3)$$

et inversement :

$$\theta = \arccos \left(\frac{\text{Tr}(R) - 1}{2} \right) \text{ et } S_n = \frac{R - R^t}{2 \cdot \sin \theta} \quad (4)$$

2.3 Distance induite par une métrique riemannienne sur une sous-variété différentielle connexe géodésiquement complète - Cas de la métrique invariante et la distance intrinsèque

Nous voudrions maintenant pouvoir mesurer des choses dans notre variété (des vitesses, des longueurs, des volumes). Étant donné que les espaces tangents sont vectoriels, on peut commencer par définir une métrique euclidienne sur ces espaces au travers d'un produit scalaire. Nous verrons ensuite comment l'utiliser.

2.3.1 Métrique Riemannienne - Métrique invariante

Définition 5 *Métrique Riemannienne*

Pour définir une métrique Riemannienne sur \mathcal{M} il faut définir une métrique euclidienne (ou de manière équivalente un produit scalaire $(\cdot|\cdot)$ ou un tenseur métrique G) sur chaque espace tangent. On adoptera les notations suivantes :

$$\begin{aligned} \forall v_x, w_x \in T_x \mathcal{M} \quad (v_x | w_x) &= v_x \overline{\otimes} G_{T_x \mathcal{M}} \overline{\otimes} w_x \\ \forall v_x \in T_x \mathcal{M} \quad \|v_x\| &= v_x \overline{\otimes} G_{T_x \mathcal{M}} \overline{\otimes} v_x \end{aligned}$$

Il faudra de plus que la collection de produits scalaires ainsi formée soit continue et différentiable.

Le seul cas qui nous intéressera ici sera le cas où l'on utilise la restriction de la métrique euclidienne usuelle de l'espace ambiant E dans chaque espace tangent (soit pour les normales $(u|v) = \text{Tr}(u.v^t)$ et pour les rotations $(M_1|M_2) = \text{Tr}(M_1 M_2^t)$).

La section suivante permet de se familiariser avec cette métrique, en exposant comment mener les calculs dans la base naturelle d'une carte π avec le formalisme du calcul tensoriel.

2.3.2 Calcul tensoriel dans un sous-espace vectoriel

Nous avons vu que l'espace tangent $T_x\mathcal{M}$ était un sous-espace vectoriel de E . Nous avons de plus obtenu une base de cet espace (la base naturelle dans une carte donnée, en général quelconque). Nous voulons définir un produit scalaire sur $T_x\mathcal{M}$, comme la restriction du produit scalaire de E . Les vecteurs de $T_x\mathcal{M}$ peuvent alors être vus comme des tenseurs d'ordre 1, et nous souhaiterions pouvoir travailler avec les outils de base de la théorie des tenseurs euclidiens (base duale, coordonnées co- et contra-variantes, tenseur métrique, voir [6] ou [4]) dans cette base. Ce travail sera d'autant plus utile que cette base n'est en général pas orthonormée.

Pour simplifier les notations, nous noterons F le sous-espace vectoriel. Il faudra bien évidemment distinguer les coordonnées dans une base de E et les coordonnées dans une base de F . Pour cela, tous les indices seront précédés du nom de l'espace auquel ils correspondent. En particulier, un indice E_i (relatif à E) pourra varier entre 1 et D (dimension de E), tandis qu'un indice F_i ne pourra varier qu'entre 1 et d (dimension de F , $d \leq D$). Lorsque tous les indices correspondent au même espace, on pourra noter qu'une seule fois le nom de l'espace. Il conviendra aussi de distinguer le produit scalaire de E , noté $(|)$ et représenté par le tenseur métrique G , de coordonnées G_{Eij} et celui de F , noté $(|)_F$ et représenté par le tenseur métrique G_F , de coordonnées G_{Fij} .

Soit donc une base $\{e_{Ei}\}$ de E (en général orthonormée, auquel cas $[G_{Eij}] = Id$), et une base $\{f_{Fi}\}$ de F . La seconde est définie en fonction de la première par les relations suivantes : $f_{Fk} = e_{Ei} A_{Fk}^{Ei}$. La matrice $[A_{F\bullet}^{E\bullet}]$ rassemble donc les coordonnées contra-variantes de f_{Fk} dans sa colonne k . On peut aussi écrire ces relations sous la forme $f_{Fi} = A_{Fi}^{Ek} e_{Ek}$, la matrice associée vérifie : $[A_{F\bullet}^{E\bullet}] = [A_{F\bullet}^{E\bullet}]^t$, et rassemble donc les coordonnées contra-variantes de f_{Fi} dans sa ligne i .

On peut aussi utiliser les relations suivantes : $f_{Fi} = A_{FiEk} e_{Ek}^{Ei}$ ou $f_{Fk} = e_{Ei}^{Ei} A_{EiFk}$. La matrice $[A_{F\bullet E\bullet}]$ rassemble donc les coordonnées covariantes de f_{Fi} dans sa ligne i , et la matrice $[A_{E\bullet F\bullet}]$ rassemble les coordonnées covariantes de f_{Fk} dans sa colonne k .

Soient $[G_{E\bullet\bullet}]$ et $[G^{E\bullet\bullet}]$ les matrices du tenseur métrique dans notre base de E . On vérifie facilement que les règles de "l'ascenseur d'indice" sont respectées : $f_{Fi} = A_{Fi}^{Ek} e_{Ek} = A_{Fi}^{Ek} G_{Eki} e^{Ei}$ soit, sous forme matricielle, $[A_{F\bullet}^{E\bullet}] = [A_{F\bullet E\bullet}][G^{E\bullet\bullet}]$, $[A^{F\bullet}_{E\bullet}] = [G^{E\bullet\bullet}][A_{E\bullet F\bullet}]$, et ainsi de suite.

Considérons maintenant le tenseur métrique dans F . On peut calculer ses coordonnées dans la base $\{f_{Fi}\}$ de la façon suivante : $G_{Fij} = (f_{Fi}|f_{Fj})_F = (A_{Fi}^{Ei} e_{Ei} | e^{Ej} A_{EjFj})_E = A_{Fi}^{Ek} A_{EjFj} = A_{Fi}^{Ei} g_{Eij} A_{Fj}^{Ej}$, soit, sous forme matricielle : $[G_{F\bullet\bullet}] = [A_{F\bullet}^{E\bullet}][G_{E\bullet\bullet}][A_{E\bullet F\bullet}]$. Les résultats classiques de la théorie des tenseurs restent valables lorsqu'on travaille dans F , considéré comme un espace vectoriel isolé. On peut donc calculer $[G^{F\bullet\bullet}] = [G_{F\bullet\bullet}]^{-1}$, et appliquer la règle de l'ascenseur d'indice : La base duale est définie par $f^{Fi} = G^{Fik} f_{Fk}$ et donc $[A^{F\bullet}_{E\bullet}] = [G^{F\bullet\bullet}][A_{F\bullet E\bullet}]$ par exemple.

Exemple d'utilisation Nous allons illustrer concrètement l'utilisation de ces résultats pour simplifier les calculs, lorsqu'on travaille dans une carte locale π . Nous prendrons l'exemple de la norme de la vitesse instantanée, exploitée dans la partie suivante. Le sous-

espace vectoriel F sera l'espace tangent en un point $\pi(\vec{x})$, noté $T_{\pi(\vec{x})}\mathcal{M}$. Sa base sera la base naturelle $f_{Fi} = \frac{\partial \pi}{\partial \vec{x}_i}$, notée plus simplement $f_{\pi(\vec{x})i} = \frac{\partial \pi}{\partial \vec{x}_i}(\vec{x})$. Les coordonnées covariantes du tenseur métrique dans cette base seront notées plus simplement : $G_{\pi(\vec{x})ij}$, et calculable par la simple relation $g_{\pi(\vec{x})ij} = (f_{\pi(\vec{x})i} | f_{\pi(\vec{x})j}) = (\frac{\partial \pi}{\partial \vec{x}_i}(\vec{x}) | \frac{\partial \pi}{\partial \vec{x}_j}(\vec{x}))$.

Supposons donc que nous ayons l'équation de cette trajectoire dans une carte locale : $\gamma(t) = \pi(\vec{\gamma}(t))$. Lorsqu'on dérive, on trouve $\dot{\gamma}(t) = \frac{d\vec{\gamma}_i}{dt}(t) f_{\pi(\vec{x})i}$. On a alors : $\|\dot{\gamma}(t)\|^2 = \frac{d\vec{\gamma}_i}{dt}(t) (f_{\pi(\vec{x})i} | f_{\pi(\vec{x})j}) \frac{d\vec{\gamma}_j}{dt}(t)$, soit $\|\dot{\gamma}(t)\|^2 = \frac{d\vec{\gamma}_i}{dt}(t) G_{\pi(\vec{x})ij} \frac{d\vec{\gamma}_j}{dt}(t)$. La méthode de calcul est donc simple : calculer les coordonnées de $G_{\pi(\vec{x})ij}$ dans la base naturelle, et considérer les dérivées $\frac{d\vec{\gamma}_i}{dt}(t)$ comme les coordonnées dans cette base.

2.3.3 Distances et Géodésiques

Maintenant que nous savons mesurer une vitesse en un point d'une courbe, on peut en déduire la longueur de la courbe, tout simplement en intégrant la norme de la vitesse. On peut alors chercher la courbe la plus courte entre deux points et définir ainsi la distance entre ces points (comme dans un espace vectoriel, où la distance est la longueur du segment de droite, étant bien connu que le plus court chemin entre deux points est la ligne droite).

Donnons nous deux points \mathbf{x} et \mathbf{y} de \mathcal{M} . Nous allons donc ici considérer l'ensemble des trajectoires reliant ces deux points, que l'on notera dans cette section $\Theta_{\mathbf{xy}} = \{\gamma : [a,b] \rightarrow \mathcal{M} / \gamma(a) = \mathbf{x}, \gamma(b) = \mathbf{y}\}$. Nous savons, grâce à la métrique Riemannienne, calculer la vitesse instantanée en un point d'une trajectoire. Ceci permet de définir la longueur d'une trajectoire : $L_a^b(\gamma) = \int_a^b \|\dot{\gamma}(t)\| dt$. On vérifie que cette définition ne dépend en fait pas de la trajectoire choisie mais uniquement de la courbe suivie. La distance entre deux points \mathbf{x} et \mathbf{y} de \mathcal{M} est définie comme la borne inférieure des longueurs des courbes joignant ces deux points : $d(\mathbf{x}, \mathbf{y}) = \inf_{\gamma \in \Theta_{\mathbf{xy}}} L_a^b(\gamma)$. Les courbes qui ont effectivement cette longueur seront par la suite importantes, mais il est difficile de les calculer directement.

Or, le calcul des variations permet de calculer explicitement les courbes qui optimisent *localement* un second critère : $E_a^b(\gamma) = \int_a^b \|\dot{\gamma}(t)\|^2 dt$. Deux résultats permettent de faire le lien entre les deux critères. D'une part, on montre que, pour une courbe donnée, les trajectoires qui minimisent *localement* $E_a^b(\gamma)$ sont les trajectoires parcourues à vitesse constante. D'autre part, on montre aussi que si une courbe minimise *localement* $L_a^b(\gamma)$ sur $\Theta_{\mathbf{xy}}$, les trajectoires associées parcourues à vitesse constante minimisent *localement* $E_a^b(\gamma)$ sur $\Theta_{\mathbf{xy}}$.

Les trajectoires qui minimisent *localement* $E_a^b(\gamma)$ seront appelées géodésiques. D'après le calcul des variations, elles sont solutions - dans une carte locale donnée - d'une équation différentielle du deuxième ordre, qui fait intervenir les variations du tenseur métrique dans cette carte dues à la courbure de la surface. Ce résultat est important tant sur le plan pratique (il permet de calculer explicitement les géodésiques et donc les distances) que sur le plan théorique (il permet, comme on le verra dans la section suivante, de définir la carte exponentielle).

Les géodésiques qui minimisent *globalement* $L_a^b(\gamma)$ sur $\Theta_{\mathbf{xy}}$ seront appelées géodésiques minimisantes.

Prenons pour exemple la sphère unité dans \mathbb{R}^n , qui représente les normales. On montre facilement que les géodésiques suivent les grands cercles de la sphère. Elles peuvent effectuer plusieurs fois le tour de ces grands cercles. Les géodésiques minimisantes sont celles qui effectuent moins d'un demi-tour.

La variété est dite géodésiquement complète si le domaine de définition de toute géodésique peut-être étendu à \mathbb{R} . On peut par exemple prolonger une géodésique de la sphère en faisant un nombre infini de fois le tour du grand cercle qu'elle suit. Enlevons maintenant le pôle nord de notre sphère, les géodésiques suivent toujours les grands cercles, mais, dans le cas des méridiens, elles ne peuvent passer par le pôle manquant. Leur domaine de définition ne peut dépasser un tour, et ne peut donc pas être étendu à \mathbb{R} (la vitesse est constante sur une géodésique). On constate en outre qu'il n'existe pas de géodésique minimisante entre deux points d'un même méridien situés de part et d'autre du pôle manquant. Ceci correspond en fait à un résultat général : lorsque la variété est géodésiquement complète, le théorème de Hopf-Rinow-De Rham assure l'existence d'une géodésique entre chaque points, et la variété est complète pour la distance. Nous supposons cette propriété vérifiée dans la suite.

On a donc le résultat suivant :

Définition 6 *Distance et géodésiques*

Soient \mathbf{x} et \mathbf{y} deux points de \mathcal{M} , et $\Theta_{\mathbf{x}\mathbf{y}}$ l'ensemble des trajectoires qui relient ces deux points.

La distance entre ces deux points est : $d(\mathbf{x}, \mathbf{y}) = \inf_{\gamma \in \Theta_{\mathbf{x}\mathbf{y}}} L_a^b(\gamma)$ où $L_a^b(\gamma) = \int_a^b \|\dot{\gamma}(t)\| dt$.

Les trajectoires de $\Theta_{\mathbf{x}\mathbf{y}}$ qui minimisent localement $E_a^b(\gamma) = \int_a^b \|\dot{\gamma}(t)\|^2 dt$ sont appelées géodésiques. La vitesse y est constante.

Les trajectoires de $\Theta_{\mathbf{x}\mathbf{y}}$ qui minimisent globalement $L_a^b(\gamma)$ et sont parcourues à vitesse constantes sont des géodésiques particulières, appelées géodésiques minimisantes.

Leur existence est assuré si \mathcal{M} est géodésiquement complète.

Exemples

Espace affine Les géodésiques suivent les lignes droites, et la distance induite est la distance euclidienne.

Courbes Les géodésiques suivent la courbe à vitesse constante. La distance entre deux points correspond à la longueur de la courbe entre ces deux points.

Normales Comme nous l'avons dit, les géodésiques suivent les grands cercles de la sphère. Une géodésique minimisante est donc un arc d'un de ces cercles unitaires. Sa longueur est proportionnelle à l'angulation entre les deux extrémités. La distance intrinsèque entre normale est donc l'angulation :

$$d(\mathbf{n}_1, \mathbf{n}_2) = (\widehat{\mathbf{n}_1, \mathbf{n}_2}) \tag{5}$$

Rotations 3D Lorsque la variété différentielle et sa mesure sont invariantes vis à vis du groupe de transformation \mathcal{G} , et a elle-même une structure de groupe de Lie compact (ce qui est par exemple le cas des rotations), les géodésiques partant de l'origine peuvent être caractérisées autrement : elles correspondent aux sous-groupes à un paramètre, et vérifient :

$$\gamma(s+t) = \gamma(s) \circ \gamma(t) = \gamma(t) \circ \gamma(s)$$

Dans le cas des rotations, les sous-groupes à un paramètre sont les ensembles de rotations de même axe :

$$R(t) = R(n_0, t.\theta_0)$$

Elles sont minimisantes pour $-\pi < t.\theta_0 < \pi$.

La distance entre une rotation $R(n, \theta)$ et l'identité est donc l'angle de cette rotation :

$$d(R(n, \theta), Id) = |\theta| \quad \text{avec } \theta \in [-\pi, \pi] \quad (6)$$

La distance entre deux rotations quelconque peut toujours être calculée en utilisant la propriété d'invariance pour ramener l'une d'elle à l'identité :

$$d(R_1, R_2) = d(\mathbf{f}_{R_1}^{-1}(R_2), Id) = d(R_1^t \cdot R_2, 0) \quad (7)$$

2.4 Carte exponentielle et lieu de coupure

Nous allons ici mettre en évidence, pour toute primitive \mathbf{x} de la variété, une carte locale particulière centrée en \mathbf{x} et définie à partir du plan tangent $T_{\mathbf{x}}\mathcal{M}$, appelée carte exponentielle. Cette carte locale sera obtenue en "aplatissant" la variété sur le plan tangent, le long des géodésiques, et recouvrira la presque totalité de la variété, et pourra être étendue (en perdant éventuellement la propriété d'injectivité ou d'ouverture de l'espace de départ) à la variété entière.

Soit donc une primitive \mathbf{x} de \mathcal{M} . Les géodésiques sont solutions d'une équation différentielle du second ordre dans une carte locale. Il existe donc une unique géodésique vérifiant $\gamma(0) = \mathbf{x}$ et $\dot{\gamma}(0) = \vec{v}_{\mathbf{x}}$ où $\vec{v}_{\mathbf{x}} \in T_{\mathbf{x}}\mathcal{M}$, i.e. partant de \mathbf{x} avec une vitesse $\vec{v}_{\mathbf{x}}$ donnée. Nous noterons cette géodésique $\gamma_{(\mathbf{x}, \vec{v}_{\mathbf{x}})}$. Son domaine de définition peut-être étendu à \mathbb{R} , et le point $\gamma_{(\mathbf{x}, \vec{v}_{\mathbf{x}})}(1)$ est parfaitement défini. Ceci permet de définir l'application suivante, pour \mathbf{x} fixé :

$$\exp_{\mathbf{x}} : \begin{array}{ccc} T_{\mathbf{x}}\mathcal{M} & \rightarrow & \mathcal{M} \\ \vec{v}_{\mathbf{x}} & \mapsto & \gamma_{(\mathbf{x}, \vec{v}_{\mathbf{x}})}(1) \end{array}$$

Comme il existe toujours une géodésique minimisante entre deux points de \mathcal{M} , cette application est surjective. Cependant, elle doit, pour définir une carte locale, être un difféomorphisme, donc être injective et définie sur un ouvert. Nous allons donc la restreindre au plus grand domaine de définition ouvert où elle est injective. Pour cela, nous devons définir le lieu de coupure.

Définition 7 *Lieu de coupure*

Soit \mathbf{x} un point de \mathcal{M} .

Considérons la géodésique $\gamma_{(\mathbf{x}, \vec{v}_\mathbf{x})}$. Elle est minimisante jusqu'à un temps $0 < t_0 \leq +\infty$, et ne l'est plus jamais après. Si $t_0 < +\infty$, le point $\mathbf{y} = \exp_\mathbf{x}(t_0 \cdot \vec{v}_\mathbf{x})$ est appelé point de coupure, et le vecteur tangent $t_0 \cdot \vec{v}_\mathbf{x}$ est appelé point de coupure tangent.

On regroupera les points de coupure tangents dans lieu de coupure tangentiel $C(\mathbf{x})$, et les points de coupure dans le lieu de coupure $\mathcal{C}(\mathbf{x})$.

Ces deux ensembles sont de mesure nulle.

On peut de même regrouper les points du type $t \cdot \vec{v}_\mathbf{x}$ avec $t < t_0$ dans le domaine $D(\mathbf{x})$, qui est donc ouvert. L'existence d'une géodésique minimisante entre tous points permet alors d'affirmer que la restriction de l'exponentielle à $D(\mathbf{x}) \cup C(\mathbf{x})$ couvre la totalité de la variété. L'unicité des géodésiques minimisante permet d'affirmer que sa restriction à $D(\mathbf{x})$ est injective. Nous pouvons donc couvrir presque toute la variété de façon injective à partir de cet ouvert. On montre que la fonction ainsi définie est bel et bien un difféomorphisme et définissons ainsi la carte exponentielle, qui couvre presque toute la variété.

Définition 8 *Carte exponentielle*

Soit \mathbf{x} un point de \mathcal{M} . On note $D(\mathbf{x})$ le domaine de $T_\mathbf{x}\mathcal{M}$ contenant 0 et délimité par le lieu de coupure $C(\mathbf{x})$ (exclu). L'application suivante définit alors une carte locale centrée en \mathbf{x} :

$$\exp_\mathbf{x} : \begin{array}{ccc} D(\mathbf{x}) \in T_\mathbf{x}\mathcal{M} & \rightarrow & \mathcal{M} \setminus \mathcal{C}(\mathbf{x}) \\ \vec{x}\vec{y} & \mapsto & \exp_\mathbf{x}(\vec{x}\vec{y}) = \gamma_{(\mathbf{x}, \vec{x}\vec{y})}(1) \end{array}$$

L'application inverse est notée $\log_\mathbf{x}$ avec $\log_\mathbf{x}(\mathbf{y}) = \vec{x}\vec{y}$.

La géodésique suivie pour atteindre \mathbf{y} depuis \mathbf{x} est par définition parcourue pendant un temps unité à vitesse constante égale à $\|\vec{x}\vec{y}\|$. Sa longueur est donc $\|\vec{x}\vec{y}\|$. Hors cette longueur est par définition la distance $d(\mathbf{x}, \mathbf{y})$ entre les deux extrémités. On a donc la propriété suivante :

Proposition 1 *Norme dans la carte exponentielle*

La norme dans la carte exponentielle correspond à la distance dans la variété :

$$\|\vec{x}\vec{y}\| = \|\log_\mathbf{x}(\mathbf{y})\| = d(\mathbf{x}, \mathbf{y}) \quad (8)$$

Attention Cette égalité n'est vrai que pour la norme dans la carte, i.e. la distance au centre de la carte. Il n'existe en revanche aucune relation simple entre la distance entre deux points quelconques dans la carte et leur distance dans la variété.

Une conséquence est que les coordonnées du tenseur métrique dans la base naturelle à l'origine de la carte sont simples :

Proposition 2 *Coordonnées du tenseur métrique dans la base naturelle de la carte exponentielle à l'origine*

Les coordonnées du tenseur métrique de l'espace tangent en \mathbf{x} $T_\mathbf{x}\mathcal{M}$ dans la base induite par la carte exponentielle centrée en \mathbf{x} sont égales à l'identité :

$$[G_{\exp_\mathbf{x}(\vec{0}) \bullet \bullet}] = Id \quad (9)$$

Exemples

Espace affine La carte exponentielle centrée en \mathbf{x} est la translation qui amène l'origine en \mathbf{x} . Elle couvre la totalité de l'espace (son lieu de coupure est à l'infini dans toutes les directions).

Courbes La carte exponentielle pour une courbe correspond à l'abscisse curviligne à partir d'un point donné (la géodésique, donc la courbe, est parcourue à vitesse constante). Bien entendu, l'espace tangent est une droite identifiée à la droite des réels, et on paramètre directement la carte à partir de \mathbb{R} . L'égalité entre coordonnée curviligne et distance sur une courbe n'est qu'un cas particulier de l'égalité entre norme dans la carte exponentielle et distance dans la variété ($\|\overrightarrow{xy}\| = d(x,y)$).

Normales Nous avons présenté trois cartes locales pour l'ensemble des normales (fin partie 2.2.3). La première (avec la distance intrinsèque $\|\overrightarrow{on}\| = (\widehat{\mathbf{o}, \mathbf{n}})$) est la carte exponentielle. Nous avons aussi expliqué à cette occasion comment calculer \overrightarrow{on} à partir de \mathbf{o} et \mathbf{n} , i.e. comment calculer le logarithme :

$$\overrightarrow{on} = \log_{\mathbf{o}}(\mathbf{n}) = (\widehat{\mathbf{o}, \mathbf{n}}) \frac{\mathbf{n} - \mathbf{o}(\mathbf{n}|\mathbf{o})}{\|\mathbf{n} - \mathbf{o}(\mathbf{n}|\mathbf{o})\|} \quad (10)$$

Pour exprimer l'exponentiel, i.e. calculer \mathbf{n} à partir de \mathbf{o} et \overrightarrow{on} , il faut faire tourner la normale \mathbf{o} d'un angle de $\|\overrightarrow{on}\| = (\widehat{\mathbf{o}, \mathbf{n}})$, dans la direction indiquée par \overrightarrow{on} . Pour cela, il suffit d'appliquer à \mathbf{o} la rotation d'axe $\mathbf{o} \wedge \overrightarrow{on}$ et d'angle $\|\overrightarrow{on}\|$. Étant donné que \mathbf{o} est de norme unitaire et que \mathbf{o} et \overrightarrow{on} sont par construction perpendiculaires, le produit vectoriel $\mathbf{o} \wedge \overrightarrow{on}$ a effectivement pour norme $\|\overrightarrow{on}\|$. Finalement, il suffit d'appliquer à \mathbf{o} la rotation de vecteur $\mathbf{o} \wedge \overrightarrow{on}$:

$$\mathbf{n} = \exp_{\mathbf{o}}(\overrightarrow{on}) = R_{\mathbf{o} \wedge \overrightarrow{on}}(\mathbf{o}) \quad (11)$$

Le lieu de coupure de cette carte exponentielle correspond à un quart (pour les normales non-orienté) ou un demi-tour (dans le cas orienté) sur la sphère, i.e. à $|\overrightarrow{on}| = \pi/2$ (l'ensemble des normales perpendiculaires à \mathbf{o}) pour les normales non-orienté et $|\overrightarrow{on}| = \pi$ (la normale $-\mathbf{o}$) pour le cas orienté.

Rotations 3D Nous n'avons présenté qu'une seule carte pour l'ensemble des rotations en 3D. Il s'agit effectivement de la carte exponentielle centrée autour de l'identité.

2.4.1 Cas invariant homogène : carte principale

Dans le cas invariant homogène, on peut toujours se ramener à l'origine, à l'aide des fonctions de placement. Nous n'utiliserons donc qu'une seule carte, la carte centrée autour de l'origine, appelée carte principale, et écrivons la quasi-totalité des calculs dans cette carte. Il faudra préalablement identifier les espaces tangents.

Lien entre les espaces tangents Nous allons ici montrer comment identifier l'espace tangent à l'origine \mathfrak{o} et l'espace tangent en un point \mathbf{x} par le biais de la fonction de placement $\mathbf{f}_{\mathbf{x}}$.

Soit une courbe γ avec $\gamma(0) = \mathbf{x}$. On peut écrire $\gamma(t) = \mathbf{f}_{\mathbf{x}}(\gamma'(t))$ avec $\gamma'(0) = \mathfrak{o}$. Dérivons par rapport à t en 0 : $\dot{\gamma}(0) = \left. \frac{d\mathbf{f}_{\mathbf{x}}(\mathbf{e})}{d\mathbf{e}} \right|_{\mathbf{e}=\mathfrak{o}} \cdot \dot{\gamma}'(0)$. On peut ainsi faire le lien entre le vecteur tangent $v_{\mathbf{x}}$ dans $T_{\mathbf{x}}\mathcal{M}$ et v dans $T_{\mathfrak{o}}\mathcal{M}$, en les considérant comme des vecteurs de E :

$$\begin{aligned} v_{\mathbf{x}} &= J(\mathbf{f}_{\mathbf{x}}) \cdot v \\ v &= J(\mathbf{f}_{\mathbf{x}})^{-1} \cdot v_{\mathbf{x}} \end{aligned} \quad \text{avec } J(\mathbf{f}_{\mathbf{x}}) = \left. \frac{d\mathbf{f}_{\mathbf{x}}(\mathbf{e})}{d\mathbf{e}} \right|_{\mathbf{e}=\mathfrak{o}}$$

On peut donc écrire quel doit être le lien entre les métriques des différents espaces tangents pour respecter la propriété d'invariance :

$$[G_{T_{\exp_{\mathbf{x}}}\mathcal{M}}] = J(\mathbf{f}_{\mathbf{x}})^{-t} \cdot [G_{T_{\exp_{\mathfrak{o}}}\mathcal{M}}] \cdot J(\mathbf{f}_{\mathbf{x}})^{-1}$$

Cette propriété doit en particulier être vérifiée pour le sous groupe d'isotropie \mathcal{H} (section 2.1) :

$$\forall \mathbf{h} \in \mathcal{H} \quad [G_{T_{\exp_{\mathfrak{o}}}\mathcal{M}}] = J(\mathbf{h})^{-t} \cdot [G_{T_{\exp_{\mathfrak{o}}}\mathcal{M}}] \cdot J(\mathbf{h})^{-1}$$

Carte principale Nous allons ici montrer comment travailler exclusivement dans la carte principale (qui est, rappelons le, la carte exponentielle centrée autour de l'origine de la variété). Les vecteurs dans cette carte seront simplement notés \vec{x} (au lieu de $\vec{\mathbf{o}\mathbf{x}}$), et les fonctions associées seront simplement notées \exp et \log :

$$\begin{aligned} \vec{x} &= \log(\mathbf{x}) \\ \mathbf{x} &= \exp(\vec{x}) \end{aligned}$$

On peut alors définir, dans la carte principale, l'action d'une transformation, et les fonctions de placements :

$$\begin{aligned} \mathbf{f} \star \vec{x} &= \log(\mathbf{f} \star \mathbf{x}) & \mathbf{f} \star \mathbf{x} &= \mathbf{f} \star \exp(\vec{x}) \\ \mathbf{f}_{\vec{x}} \star \vec{y} &= \log(\mathbf{f}_{\mathbf{x}} \star \exp(\vec{y})) \end{aligned}$$

On peut de plus définir directement les vecteurs tangents par leurs coordonnées dans la base induite par la carte principale :

$$v_{\mathbf{x}} = \sum_i \vec{v}_{\vec{x}} \cdot \frac{\partial \exp}{\partial \vec{x}_i}(\vec{x}) \quad v_{\mathbf{x}} \in T_{\mathbf{x}}\mathcal{M} \quad \vec{v}_{\vec{x}} \in T_{\vec{x}}\mathcal{M}$$

Le lien entre le vecteur tangent $\vec{v}_{\vec{x}}$ de $T_{\vec{x}}\mathcal{M}$ et le vecteur \vec{v} de $T_{\mathfrak{o}}\mathcal{M}$ est alors :

$$\begin{aligned} \vec{v}_{\vec{x}} &= J(\mathbf{f}_{\vec{x}}) \cdot \vec{v} \\ \vec{v} &= J(\mathbf{f}_{\vec{x}})^{-1} \cdot \vec{v}_{\vec{x}} \end{aligned} \quad \text{avec } J(\mathbf{f}_{\vec{x}}) = \left. \frac{d\mathbf{f}_{\vec{x}}(\vec{e})}{d\vec{e}} \right|_{\vec{e}=\mathfrak{o}}$$

Et comme $[G_{\exp_{\vec{\mathfrak{o}}}}] = Id$, la propriété d'invariance de la métrique impose :

$$[G_{\exp_{\vec{x}}}] = J(\mathbf{f}_{\vec{x}})^{-t} \cdot [G_{\exp_{\vec{\mathfrak{o}}}}] \cdot J(\mathbf{f}_{\vec{x}})^{-1} = J(\mathbf{f}_{\vec{x}})^{-t} \cdot J(\mathbf{f}_{\vec{x}})^{-1}$$

Lien entre la carte principale et les autres cartes exponentielles On peut finalement faire le lien entre les coordonnées dans la carte principale et dans les autres cartes exponentielles :

$$\vec{x}\vec{y} = \log_{\mathbf{x}}(\mathbf{y}) = J(\mathbf{f}_{\vec{x}}) \cdot (\mathbf{f}_{\vec{x}}^{-1} \star \vec{y})$$

On peut aussi écrire la distance dans une carte exponentielle quelconque, et, bien-sûr, dans la carte principale :

$$\text{dist}(\mathbf{x}, \mathbf{y}) = \vec{x}\vec{y}^t \cdot [G_{\exp \vec{x} \bullet \bullet}] \cdot \vec{x}\vec{y} = (\mathbf{f}_{\vec{x}}^{-1} \star \vec{y})^t \cdot (\mathbf{f}_{\vec{x}}^{-1} \star \vec{y})$$

Cas des transformations rigides elles-mêmes Dans le cas où $\mathcal{M} = \mathcal{G}$, il faut adapter les équations ci-dessus, en remarquant qu'une transformation est sa propre fonction de placement.

2.5 Forme volume et intégration d'une fonction réelle

Nous souhaitons maintenant pouvoir intégrer explicitement une fonction $f : \mathcal{M} \rightarrow \mathbb{R}$ intégrable. Le tenseur métrique de $T_{\mathbf{x}}\mathcal{M}$ permet de définir l'élément de volume infinitésimal (le raisonnement est analogue à celui permettant de calculer la norme de la dérivée) :

Proposition 3 *Forme volume*

L'élément infinitésimal de volume au point \mathbf{x} de \mathcal{M} est donné par : $d\mathcal{M}(\mathbf{x}) = \sqrt{|[G_{T_{\mathbf{x}}\mathcal{M}} \bullet \bullet]|} dx$.

Cette écriture est cependant peu pratique, et nous passerons par une carte locale π , la carte principale de préférence.

Soit donc une carte locale π . Nous supposons pour simplifier qu'elle couvre la presque-totalité de la variété. Si notre fonction est donnée sous la forme $\vec{x} \rightarrow f(\pi(\vec{x}))$, on peut intégrer directement dans la carte, en utilisant la base naturelle. Il faudra cependant faire attention, car la carte ne couvre pas toujours la totalité de l'espace.

Proposition 4 *Intégration d'une fonction réelle à l'aide d'une carte locale*

Soit π une carte locale couvrant presque toute la variété et une fonction $f : \mathcal{M} \rightarrow \mathbb{R}$ intégrable, d'intégrale nulle en dehors du domaine de la carte $\int_{\mathcal{M}/D_{\pi}} f(\mathbf{x}) \cdot d\mathcal{M}(\mathbf{x}) = 0$.

On a :

$$\int_{\mathcal{M}} f(\mathbf{x}) \cdot d\mathcal{M}(\mathbf{x}) = \int_{D_{\pi}} f(\pi(\vec{x})) \cdot \sqrt{|[G_{\pi(\vec{x}) \bullet \bullet}]|} \cdot d\vec{x} \quad (12)$$

Dans le cas homogène invariant, on a $[G_{\exp \vec{x} \bullet \bullet}] = J(\mathbf{f}_{\vec{x}})^{-1} \cdot J(\mathbf{f}_{\vec{x}})^{-t}$ et donc, en utilisant la carte principale :

$$\int_{\mathcal{M}} f(\mathbf{x}) \cdot d\mathcal{M}(\mathbf{x}) = \int_{D_{\pi}} \mathbf{f}(\exp(\vec{x})) \cdot \frac{d\vec{x}}{|J(\mathbf{f}_{\vec{x}})|} \quad (13)$$

Exemple avec les normales 3D Détaillons le cas de \mathcal{P}_2 , ensemble des normales non orientées de \mathbb{R}^3 . On utilisera la base canonique de \mathbb{R}^3 , notée $\{i, j, k\}$.

Cet ensemble peut-être représenté par une demi-sphère de rayon unitaire. Nous fixons l'origine \circ au pôle, de coordonnées $(0,0,1)$, et le plan tangent en \circ est le plan horizontal passant par \circ . Nous le munissons bien évidemment de la base $\{i, j\}$.

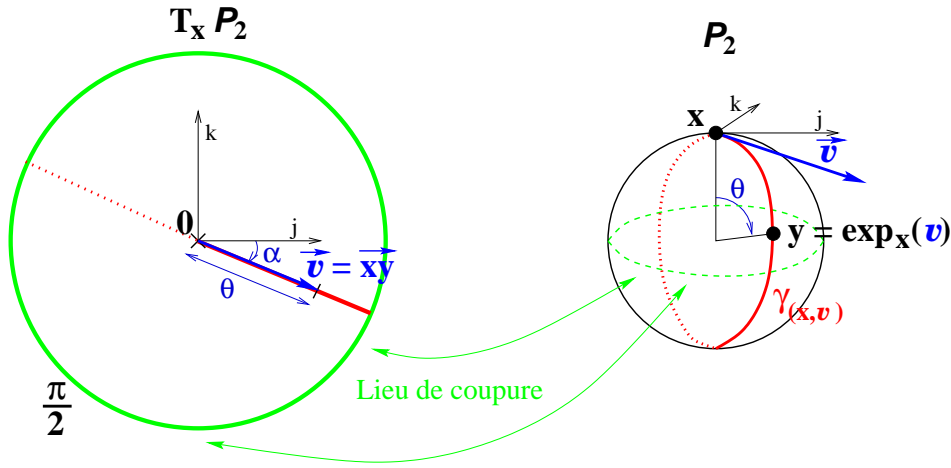
La carte principale est donc définie de la manière suivante :

$$\exp : \begin{cases} \{\vec{n} = (x_1, x_2) / x_1^2 + x_2^2 < \frac{\pi}{2}\} & \rightarrow & \mathcal{P}_2 \\ \vec{n} = (\theta \cos \alpha, \theta \sin \alpha) & \mapsto & \exp(\vec{n}) = (\sin \theta \cos \alpha, \sin \theta \sin \alpha, \cos \theta) \end{cases}$$

Pour alléger les notations, on utilisera plutôt la carte équivalente (en coordonnées polaires) :

$$\mathbf{n} : \begin{cases} \vec{n} = (\theta, \alpha) / \theta \in [0, \pi/2[, \alpha \in]-\pi, \pi] & \rightarrow & \mathcal{P}_2 \\ \vec{n} = (\theta, \alpha) & \mapsto & (\sin \theta \cos \alpha, \sin \theta \sin \alpha, \cos \theta) \end{cases}$$

Tout-ceci est illustré sur la figure suivante :



On peut alors calculer $\sqrt{|G_{\mathbf{n}(\theta, \alpha) \bullet \bullet}|}$:

$$\frac{\partial \mathbf{n}}{\partial \theta}(\theta, \alpha) = (\cos \theta \cos \alpha, \cos \theta \sin \alpha, -\sin \theta) \text{ et } \frac{\partial \mathbf{n}}{\partial \alpha}(\theta, \alpha) = (-\sin \theta \sin \alpha, \sin \theta \cos \alpha, 0)$$

$$[G_{\mathbf{n}(\theta, \alpha) \bullet \bullet}] = \begin{bmatrix} \cos^2 \theta \cos^2 \alpha + \cos^2 \theta \sin^2 \alpha + \sin^2 \theta & \dots \\ -\cos \theta \cos \alpha \sin \theta \sin \alpha + \cos \theta \sin \alpha \sin \theta \cos \alpha & \sin^2 \theta \sin^2 \alpha + \sin^2 \theta \cos^2 \alpha \end{bmatrix}$$

$$[G_{\mathbf{n}(\theta, \alpha) \bullet \bullet}] = \begin{bmatrix} 1 & 0 \\ 0 & \sin^2 \theta \end{bmatrix} \text{ et donc}$$

$$\sqrt{|[G_{\mathbf{n}(\theta, \alpha) \bullet \bullet}]|} = \sin \theta \quad (14)$$

On peut alors intégrer sans problème les fonctions données sous la forme $f(\theta, \alpha)$.

3 Statistiques intrinsèques sur les primitive aléatoire

Nous pouvons maintenant construire la base du calcul des probabilités et des statistiques sur nos variétés à partir de deux éléments indépendants :

- La forme volume : la possibilité d'intégrer une fonction (ou une distribution) réelle définie à partir de l'ensemble des primitives permet de définir une densité de probabilité (à valeur réelles positive, d'intégrale unité), qui permettra de modéliser facilement une primitive aléatoire. Remarquons qu'on pourrait s'en passer, par exemple pour une primitive aléatoire n'ayant qu'un nombre fini de réalisations possibles car il suffit alors de connaître la probabilité de chacune (autrement dit, la densité de probabilité est une combinaison de diracs). Mais nous utiliserons quasi-systématiquement la densité par la suite.
- La distance intrinsèque : on pourra calculer l'espérance du carré de la distance entre une primitive aléatoire et une primitive fixe, et généraliser ainsi la notion de variance, puis de moyenne (le minimum de la la variance, comme dans le cas vectoriel). On pourra aussi calculer une covariance (dans l'espace tangent), calculer l'information, et donc généraliser la notion de gaussienne (la loi qui minimise l'information pour une moyenne et une covariance donnée).

3.1 Primitive aléatoire

Le rôle de cette partie est de définir une primitive aléatoire, sa densité de probabilité, et de calculer l'espérance d'une observable, c'est à dire une variable aléatoire réelle définie comme l'image d'une primitive aléatoire par une fonction.

Il est important de noter que les définitions de cette partie reposent toutes directement sur la métrique utilisée. Ces définitions ne seront donc invariantes que lorsqu'on utilisera la métrique invariante.

3.1.1 Primitive aléatoire - application d'une fonction mesurable

Les notions classiques d'espace probabilisé, qui regroupe l'univers, la tribu d'évènements et la mesure de probabilité est très générales, et restent valables dans une variété différentielle. Ainsi, pour une primitive aléatoire, l'univers (i.e. l'ensemble des évènements élémentaires, i.e. des réalisations possibles de la primitive) est l'ensemble des primitives \mathcal{M} . On utilisera la tribu Borélienne \mathbf{F} . On définira enfin une mesure de probabilité, i.e. une fonction σ -additive de \mathbf{F} dans $[0,1]$ exprimant, pour tout évènement complexe $A \in \mathbf{F}$, sa probabilité, i.e. la probabilité de trouver la primitive aléatoire \mathbf{x} dans A :

$$Pr_{\mathbf{x}}(A) = Pr(\mathbf{x} \in A)$$

Rappelons qu'on peut, à l'aide d'une fonction mesurable f définie de l'univers \mathcal{M} dans un autre ensemble B , associer les réalisations possibles de la primitive aléatoire \mathbf{x} définie

dans \mathcal{M} et les réalisations possibles de $f(\mathbf{x})$. En veillant à utiliser les tribus adéquates, on définit ainsi une variable aléatoire $\mathbf{y} = f(\mathbf{x})$ et sa mesure de probabilité est donnée par :

$$Pr_{\mathbf{y}}(B) = Pr_{\mathbf{x}}(f^{-1}(B))$$

3.1.2 Densité de probabilité

La mesure de probabilité pourrait suffire pour définir l'espérance d'une observable. Il est cependant peu commode de travailler avec cette mesure car il s'agit d'une fonction réelle de la tribu (i.e. un ensemble de sous parties de l'univers). Il serait bien plus pratique de travailler, comme dans le cas vectoriel avec une densité de probabilité, i.e. une fonction réelle de l'univers qui synthétise l'information de la mesure de probabilité.

Nous supposons donc, comme dans le cas vectoriel, que nos primitives aléatoires sont représentées par une densité de probabilité $p_{\mathbf{x}}(\mathbf{y})$ qui vérifie :

$$Pr(\mathbf{x} \in A) = \int_A p_{\mathbf{x}}(\mathbf{y}).d\mathcal{M}(\mathbf{y})$$

soit $dPr(\mathbf{y}) = p_{\mathbf{x}}(\mathbf{y}).d\mathcal{M}(\mathbf{y})$. La contrainte de normalisation impose : $\int_{\mathcal{M}} p_{\mathbf{x}}(\mathbf{y}).d\mathcal{M}(\mathbf{y}) = 1$.

Nous avons parlé de la densité de probabilité comme d'une fonction. Mais il pourra s'agir plus généralement d'une distribution, qui permet de représenter dans le même cadre les densités volumiques et les concentrations linéiques, surfaciques ou ponctuelles. Attention cependant, car ces distributions peuvent poser un certain nombre de problèmes, par exemple pour l'intégration dans une carte.

Exemple : densité uniforme La loi uniforme sur un sous-ensemble $X \subset \mathcal{M}$ de volume fini a une densité donnée par : $p_{\mathbf{x}}(\mathbf{y}) = \frac{1_X(\mathbf{y})}{\int_X d\mathcal{M}}$.

Exemple : densité ponctuelle Si une variable aléatoire n'a qu'un nombre fini de réalisations possibles $\{\mathbf{x}_i\}$, on peut la représenter directement en utilisant la probabilité de ces réalisations $Pr(\mathbf{x} = \mathbf{x}_i)$, ou utiliser une densité de probabilité qui est une mixture de diracs : $p_{\mathbf{x}}(\mathbf{y}) = \sum_i Pr(\mathbf{x} = \mathbf{x}_i).\delta_{\mathbf{x}_i}(\mathbf{y})$. Les intégrations deviendront alors des sommes pondérées, comme ici avec la contrainte de normalisation : $\int_{\mathcal{M}} p_{\mathbf{x}}(\mathbf{y}).d\mathcal{M}(\mathbf{y}) = \sum_i Pr(\mathbf{x} = \mathbf{x}_i) = 1$.

3.1.3 Espérance d'une observable

L'espérance d'une variable aléatoire définie comme l'image d'une primitive aléatoire par une fonction φ de \mathcal{M} dans \mathbb{R} (appelée observable en référence à la physique quantique) peut être calculée à l'aide de la densité de probabilité, comme dans le cas vectoriel :

$$E_{\mathbf{x}}[\varphi] = \int_{\mathcal{M}} \varphi(\mathbf{y}).p_{\mathbf{x}}(\mathbf{y}).d\mathcal{M}(\mathbf{y})$$

Nous avons déjà expliqué comment utiliser une carte locale pour calculer de telles intégrales (voir equ. 12), lorsque l'intégrale est nulle en dehors du domaine de la carte $\int_{\mathcal{M}/D_x} f(\mathbf{x}).d\mathcal{M}(\mathbf{x}) =$

0 (en particulier avec des distributions non nulles sur le lieu de coupure lorsqu'on utilise la carte exponentielle). On a dans ce cas :

$$E_{\mathbf{x}}[\varphi] = \int_{D_{\pi}} \varphi(\vec{x}) \cdot p_{\mathbf{x}}(\pi(\vec{x})) \cdot \sqrt{|[G_{\pi(\vec{x})}]|} \cdot d\vec{x}$$

Et en particulier (c'est la contrainte de normalisation) :

$$E_{\mathbf{x}}[1] = \int_{D_{\pi}} p_{\mathbf{x}}(\pi(\vec{x})) \cdot \sqrt{|[G_{\pi(\vec{x})}]|} \cdot d\vec{x} = 1$$

On peut même définir directement la densité de probabilité de \mathbf{x} dans la carte locale :

$$p_{\vec{x}}(\vec{x}) = p_{\mathbf{x}}(\mathbf{x}) \cdot \sqrt{|[G_{\pi(\vec{x})}]|} \quad (15)$$

Si bien que :

$$E_{\mathbf{x}}[\varphi] = \int_{D_{\pi}} \varphi(\vec{x}) \cdot p_{\vec{x}}(\vec{x}) \cdot d\vec{x}$$

$$E_{\mathbf{x}}[1] = \int_{D_{\pi}} p_{\vec{x}}(\vec{x}) \cdot d\vec{x} = 1$$

On peut facilement généraliser ces résultats pour une observable vectorielle, en isolant les coordonnées.

Propriétés On vérifie que l'espérance d'une observable est, comme dans le cas vectoriel, linéaire :

$$E_{\mathbf{x}}[\alpha \cdot \varphi] = \alpha \cdot E_{\mathbf{x}}[\varphi]$$

$$E_{\mathbf{x}}[\varphi_1 + \varphi_2] = E_{\mathbf{x}}[\varphi_1] + E_{\mathbf{x}}[\varphi_2]$$

Carte principale Dans le cas invariant homogène, on peut écrire les formules dans la carte principale (voir equ. 13) :

$$p_{\vec{x}}(\vec{x}) = \frac{p_{\mathbf{x}}(\mathbf{x})}{|J(\mathbf{f}_{\vec{x}})|}$$

$$E(\varphi(\mathbf{x})) = \int_{D_{\pi}} \varphi(\vec{x}) \cdot p_{\vec{x}}(\vec{x}) \cdot d\vec{x}$$

$$Pr(\mathbf{x} \in \mathcal{M}) = \int_{D_{\pi}} p_{\vec{x}}(\vec{x}) \cdot d\vec{x} = 1$$

Exemple: normale aléatoire uniforme en 3D La demi-sphère des normales non-orientées est de volume (en l'occurrence surface) fini. On peut donc définir une normale aléatoire uniforme sur la totalité de l'espace des normales.

Sa densité de probabilité sera de la forme $p_n(n) = k$. Nous travaillerons dans la carte habituelle (la carte exponentielle en coordonnées polaires), où cette densité prend la forme $p_{\theta,\alpha}(\theta,\alpha) = k \cdot \sin \theta$ (voir equ. 15 et 14), on peut alors calculer la constante de normalisation :

$$\int_{\theta=0}^{\frac{\pi}{2}} \int_{\alpha=-\pi}^{\pi} p_{\theta,\alpha}(\theta,\alpha) d\theta d\alpha = 1$$

$$\int_{\theta=0}^{\frac{\pi}{2}} \int_{\alpha=-\pi}^{\pi} k \cdot \sin \theta d\theta d\alpha = k \cdot 2\pi \cdot \int_{\theta=0}^{\frac{\pi}{2}} \sin \theta d\theta = k \cdot 2\pi = 1$$

Finalement : $k = 1/(2\pi)$.

On peut aussi calculer sa variance (que nous définirons dans la section suivante) par rapport à la normale correspondant à $\theta = 0$. Il s'agit de l'espérance de l'observable θ^2 :

$$E(\theta^2) = \frac{1}{2\pi} \int_{\theta=0}^{\frac{\pi}{2}} \int_{\alpha=-\pi}^{\pi} \theta^2 \sin \theta d\theta d\alpha = \int_{\theta=0}^{\frac{\pi}{2}} \theta^2 \sin \theta d\theta$$

$$= [-x^2 \cos x + 2x \cdot \sin x + 2 \cos x]_0^{\pi/2} = \pi - 2 \quad (16)$$

une normale uniforme a donc un écart-type (non-infini !) de $\sqrt{\pi - 2} = 61.2 \text{ deg}$.

Algorithme: tirage d'une normale aléatoire uniforme en nD Pour obtenir une normale aléatoire uniforme, la méthode la plus simple consiste à effectuer un tirage de vecteur uniforme à l'intérieur d'un volume à symétrie sphérique (comme une boule), et normaliser ensuite ce vecteur. Les directions étant réparties uniformément dans le volume, les normales ainsi obtenues le seront aussi.

Reste donc à tirer un vecteur aléatoire dans ce volume. Pour cela, il suffit de tirer un vecteur aléatoire dans n'importe quel ensemble contenant ce volume, et rejeter les vecteurs situés à l'extérieur.

L'ensemble où il sera le plus simple de tirer un vecteur aléatoire est le cube, car on peut tirer les coordonnées facilement (il s'agit d'un tirage uniforme) et indépendamment les unes des autres. Le volume à symétrie sphérique le plus simple sera la boule inscrite dans ce cube auquel on aura retiré une petite boule concentrique (pour éviter les divisions par une norme trop faible lors de la normalisation).

L'algorithme est donc le suivant :

Faire

- . tirer un vecteur aléatoire uniforme : $x = (UniformRnd([-1,1]), UniformRnd([-1,1]), \dots)$
- . calculer sa norme au carré : $r^2 = \sum_i x_i^2$

Jusqu'à ($r^2 > 0.01$) et ($r^2 < 1$)

Normaliser $n = x/r$

3.2 Espérance et Covariance

Nous allons ici voir comment généraliser les notions statistiques de bases aux primitives. Nous nous baserons pour cela sur l'espérance d'une observable.

3.2.1 Espérance au sens de Fréchet

Dans un espace vectoriel, la moyenne possède deux définitions équivalentes :

- On peut définir la moyenne comme la somme des différentes réalisations possibles pondérées par leurs probabilités : cette définition a l'avantage d'être explicite, donc de donner une méthode de calcul directe, mais s'appuie sur la notion de combinaison linéaire (i.e. de somme pondérée), qui est propre aux espaces vectoriels.
- On peut aussi définir la moyenne comme le vecteur minimisant la variance $\sigma_{\mathbf{x}}^2(\mathbf{y}) = E(\|\mathbf{y} - \mathbf{x}\|^2) = E(d(\mathbf{y}, \mathbf{x})^2)$. Cette définition a l'avantage d'être implicite et de ne s'appuyer que sur l'espérance d'une observable.

Fréchet a proposé de généraliser directement cette seconde définition de la moyenne (rebaptisée espérance pour éviter les confusions) à un espace métrique quelconque de la façon suivante :

Définition 9 *Espérance au sens de Fréchet*

Soit \mathbf{x} une primitive aléatoire de \mathcal{M} .

La variance est l'espérance du carré de la distance entre la primitive aléatoire et une primitive fixe \mathbf{y} :

$$\sigma_{\mathbf{x}}^2(\mathbf{y}) = E[d(\mathbf{y}, \mathbf{x})^2] = E[\|\vec{x\mathbf{y}}\|^2] \quad (17)$$

L'espérance au sens de Fréchet est l'ensemble des primitives qui minimisent la variance : $\mathbb{E}[\mathbf{x}] = \operatorname{argmin}_{\mathbf{y}} \sigma_{\mathbf{x}}^2(\mathbf{y})$. Une telle primitive est notée $\bar{\mathbf{x}}$.

Important Il est important de remarquer que, contrairement au cas vectoriel, il n'y a plus unicité de l'espérance. Par exemple, la sphère des normales étant de volume fini, on peut définir une loi uniforme sur sa totalité. La variance est alors la même vis-à-vis de tous les points de la sphère, et l'espérance au sens de Fréchet rassemble toute la sphère.

Généralisation : Primitive centrale à l'ordre α On peut généraliser cette définition aux différentes primitives centrales : les primitives centrales d'ordre α sont les primitives qui minimisent la variance à l'ordre α :

$$\sigma_{\mathbf{x}}^{\alpha}(\mathbf{y}) = E(d(\mathbf{y}, \mathbf{x})^{\alpha})$$

Les primitives centrales d'ordre 0 sont les modes de la variable aléatoire (i.e. les points de densité de probabilité maximale), les primitives d'ordre 1 correspondent à la notion de médiane.

3.2.2 Espérance au sens de Karcher

L'espérance au sens de Fréchet correspond à un minimum global. Il est souvent plus simple de gérer un minimum local, car on peut alors le caractériser par une différentielle nulle. On définit ainsi l'espérance au sens de Karcher :

Définition 10 *Espérance au sens de Karcher*

Soit \mathbf{x} une primitive aléatoire de \mathcal{M} .

L'espérance au sens de Karcher est l'ensemble des primitives qui minimisent localement la variance $\sigma_{\mathbf{x}}^2(\mathbf{y})$.

Elle contient donc l'espérance au sens de Fréchet.

3.2.3 Propriété et recherche de l'espérance au sens de Karcher

L'espérance de Karcher a été définie comme minimisant localement la variance. Elle peut donc être caractérisée par une dérivé nulle par rapport à la primitive \mathbf{y} . Nous le ferons dans la carte principale : on dérivera la variance, sous sa forme $E_{\mathbf{x}}[\|\overrightarrow{x\mathbf{y}}\|^2]$, et on l'annulera.

Commençons par la dérivé, dans la carte locale, de la distance $\|\overrightarrow{x\mathbf{y}}\|^2$:

$$\frac{\partial \|\overrightarrow{x\mathbf{y}}\|^2}{\partial \overrightarrow{\mathbf{y}}} = 2.[G_{\exp \overrightarrow{\mathbf{y}} \bullet \bullet}].\overrightarrow{x\mathbf{y}} = -2.[G_{\exp \overrightarrow{\mathbf{y}} \bullet \bullet}].\overrightarrow{\mathbf{y}\mathbf{x}} = -2.J(\mathbf{f}_{\overrightarrow{\mathbf{y}}})^{-t} . (\mathbf{f}_{\overrightarrow{\mathbf{y}}}^{-1} \star \overrightarrow{\mathbf{x}})$$

On montre (difficilement [13]) que l'on peut alors écrire :

$$\begin{aligned} \frac{\partial \sigma_{\mathbf{x}}^2(\exp(\overrightarrow{\mathbf{y}}))}{\partial \overrightarrow{\mathbf{y}}} &= \frac{\partial E_{\mathbf{x}}[\|\overrightarrow{x\mathbf{y}}\|^2]}{\partial \overrightarrow{\mathbf{y}}} = E_{\mathbf{x}}\left[\frac{\partial \|\overrightarrow{x\mathbf{y}}\|^2}{\partial \overrightarrow{\mathbf{y}}}\right] \\ &= -2.[G_{\exp \overrightarrow{\mathbf{y}} \bullet \bullet}].E_{\mathbf{x}}[\overrightarrow{\mathbf{y}\mathbf{x}}] = -2.J(\mathbf{f}_{\overrightarrow{\mathbf{y}}})^{-t} . E_{\mathbf{x}}[(\mathbf{f}_{\overrightarrow{\mathbf{y}}}^{-1} \star \overrightarrow{\mathbf{x}})] \end{aligned}$$

Cette dérivé doit être nulle en toute primitive $\overrightarrow{\mathbf{x}} = \log \overline{\mathbf{x}}$ appartenant à l'espérance de Karcher. $J(\mathbf{f}_{\overrightarrow{\mathbf{y}}})^{-t}$ étant de rang maximum, il faut et il suffit pour cela que $E_{\mathbf{x}}[(\mathbf{f}_{\overrightarrow{\mathbf{y}}}^{-1} \star \overrightarrow{\mathbf{x}})]$ soit nul :

Proposition 5 *Caractérisation de l'espérance de Fréchet*

Une primitive faisant partie de l'espérance de Karcher (et donc de l'espérance de Fréchet) vérifie :

$$E_{\mathbf{x}}[\overrightarrow{\mathbf{x}\mathbf{x}}] = E_{\mathbf{x}}[\mathbf{f}_{\overrightarrow{\mathbf{x}}}^{-1} \star \overrightarrow{\mathbf{x}}] = 0 \quad (18)$$

Cette condition est nécessaire, mais pas suffisante.

On peut rechercher une de ces primitives en faisant une descente de gradient de type Newton sur le critère. Nous savons pour cela que, à partir d'une estimation $\overrightarrow{\mathbf{y}}$ de l'espérance, nous devons aller dans la direction inverse du gradient de la variance $\Phi = \frac{\partial \sigma_{\mathbf{x}}^2(\exp(\overrightarrow{\mathbf{y}}))}{\partial \overrightarrow{\mathbf{y}}}$ pour annuler celle-ci. Une façon de la pondérer est d'utiliser l'inverse de la matrice hessienne $H = \frac{\partial \Phi(\exp(\overrightarrow{\mathbf{y}}))}{\partial \overrightarrow{\mathbf{y}}}$. En effet l'approximation au premier ordre de Φ est donnée par $\Phi(\overrightarrow{\mathbf{y}} + \overrightarrow{\delta\mathbf{y}}) = \Phi(\overrightarrow{\mathbf{y}}) + H.\overrightarrow{\delta\mathbf{y}} + O(\|\overrightarrow{\mathbf{y}}\|^2)$, et peut être annulée avec $\overrightarrow{\delta\mathbf{y}} = -H^{-1}.\Phi$.

La matrice Hessienne peut ici être approximée par $-2.J(\mathbf{f}_{\vec{y}})^{-t}.J(\mathbf{f}_{\vec{y}})^{-1}$ (la distance est de la forme $x^t.x$, sa dérivé de la forme $2.x^t.\dot{x}$, et sa dérivé seconde doit donc être de la forme $2.\dot{x}^t.\dot{x} + 2.x^t.\ddot{x} \simeq 2.\dot{x}^t.\dot{x}$). Finalement, on peut mettre à jour l'espérance en la déplaçant de $\vec{\delta y} = -H^{-1}.\Phi = J(\mathbf{f}_{\vec{y}}).E_{\mathbf{x}}[(\mathbf{f}_{\vec{y}}^{-1} \star \vec{x})]$, soit $\vec{y}' = \mathbf{f}_{\vec{y}} \star (E_{\mathbf{x}}[(\mathbf{f}_{\vec{y}}^{-1} \star \vec{x})])$. Autrement dit, la correction apportée à \vec{y} correspond à la moyenne dans la carte exponentielle centrée en \vec{y} .

Algorithme : Calcul de l'espérance de Karcher par descente de gradient L'algorithme est donc le suivant :

- Initialisation : \vec{y}^0
- A l'itération I : on calcule la moyenne $E_{\mathbf{x}}[(\mathbf{f}_{\vec{y}^I}^{-1} \star \vec{x})]$ dans la carte exponentielle centrée en \vec{y}^I , et on en déduit la nouvelle estimée : $\vec{y}^{I+1} = \mathbf{f}_{\vec{y}^I} \star (E_{\mathbf{x}}[(\mathbf{f}_{\vec{y}^I}^{-1} \star \vec{x})])$
- Jusqu'à convergence : $d(\vec{y}^{I+1}, \vec{y}^I) < \varepsilon$

On pourra de plus calculer la matrice de covariance sur le résultat obtenu.

3.2.4 Matrice de covariance

La matrice de covariance d'une primitive aléatoire par rapport à une primitive fixe \mathbf{y} est simplement définie en utilisant la covariance de $\vec{y}\vec{x}$ (qui est une observable vectorielle) définie dans la carte exponentielle centrée en \mathbf{y} :

Définition 11 *Matrice de covariance*

Soit \mathbf{x} une primitive aléatoire de \mathcal{M} .

Sa matrice de covariance par rapport à une primitive fixe \mathbf{y} est définie dans $T_{\mathbf{y}}\mathcal{M}$ par :

$$\Sigma_{\mathbf{xx}}(\mathbf{y}) = E[\vec{y}\vec{x} \otimes \vec{y}\vec{x}] = E[\vec{y}\vec{x}.\vec{y}\vec{x}^t] \quad (19)$$

La matrice de covariance par rapport à l'espérance $\bar{\mathbf{x}}$ (lorsque celle-ci est unique) est notée $\Sigma_{\mathbf{xx}} = \Sigma_{\mathbf{xx}}(\bar{\mathbf{x}})$.

Propriété En écrivant la trace de la matrice de covariance $Tr(\Sigma_{\mathbf{xx}}(\mathbf{y})) = Tr(E[\vec{y}\vec{x}.\vec{y}\vec{x}^t]) = E[Tr(\vec{y}\vec{x}.\vec{y}\vec{x}^t)] = E[||\vec{y}\vec{x}||^2]$, et en se rappelant des équations 8 et 17, on constate que le lien entre la matrice de covariance et la variance est analogue au cas vectoriel :

$$\sigma_{\mathbf{x}}^2(\mathbf{y}) = Tr(\Sigma_{\mathbf{xx}}(\mathbf{y}))$$

Cas invariant On peut se ramener à la carte principale :

$$\Sigma_{\mathbf{xx}}(\mathbf{y}) = J(\mathbf{f}_{\vec{y}}).E[(\mathbf{f}_{\vec{y}}^{-1} \star \vec{x}).(\mathbf{f}_{\vec{y}}^{-1} \star \vec{x})^t].J(\mathbf{f}_{\vec{y}})^t$$

3.2.5 Propagation de l'espérance et de la matrice de covariance dans le cas invariant

Dans le cas où on utilise la mesure invariante, on peut calculer la densité de probabilité d'une variable aléatoire définie par l'action d'une transformation rigide fixe \mathbf{f} sur une autre variable aléatoire : si $\mathbf{y} = \mathbf{f} \star \mathbf{x}$, $p_{\mathbf{y}}(\mathbf{y}) = p_{\mathbf{x}}(\mathbf{f}^{-1} \star \mathbf{y})$.

On vérifie donc, et c'est heureux, que, pour une transformation rigide fixe \mathbf{f} , l'espérance de la transformée est bien la transformé de l'espérance :

$$\mathbb{E}[\mathbf{f} \star \mathbf{x}] = \mathbf{f} \star \mathbb{E}[\mathbf{x}]$$

La matrice de covariance se propage en utilisant les formules permettant de passer d'un espace tangent à l'autre :

$$\mathbf{y} = \mathbf{f} \star \mathbf{x} \quad \Rightarrow \quad \Sigma_{\mathbf{y}\mathbf{y}} = J(\mathbf{f}) \cdot \Sigma_{\mathbf{x}\mathbf{x}} \cdot J(\mathbf{f})^t$$

Il n'existe en revanche aucun lien simple dans le cas où la transformation est elle aussi aléatoire.

3.3 Approximation par une loi Gaussienne et distance de Mahalanobis

Dans le cadre d'une implémentation informatique des primitives aléatoires, il est bien évident que l'on ne peut toujours stocker la densité de probabilité. On se limitera souvent à une approximation au deuxième ordre, en stockant l'espérance et la covariance de la primitive aléatoire.

Tout ceci rappelle les approximations gaussiennes dans le cas vectoriel, qui sont utilisées car elles maximisent l'entropie (et minimisent donc l'information) pour une moyenne et une covariance donnée. Nous allons donc généraliser la notion de gaussienne en généralisant la notion d'entropie.

3.3.1 Entropie d'une primitive aléatoire

Nous généralisons directement la définition dans le cas vectoriel : l'entropie d'une primitive aléatoire est définie par :

$$H[\mathbf{x}] = -E_{\mathbf{x}}[\log p_{\mathbf{x}}(\mathbf{x})]$$

On vérifie que la loi qui maximise l'entropie est la loi uniforme.

On vérifie aussi que l'entropie est invariante : pour une transformation déterministe \mathbf{f} , $H[\mathbf{f} \star \mathbf{x}] = H[\mathbf{x}]$.

3.3.2 Loïs gaussiennes

Les lois gaussiennes sont, dans le cas vectoriel, les lois qui maximisent l'entropie pour une moyenne et une covariance imposée. Nous conservons la même définition pour les primitives. On montre alors qu'elles sont de la même forme que dans le cas vectoriel :

Proposition 6 *Loïs gaussiennes*

La loi gaussienne d'espérance \bar{x} et de covariance Σ est définie par la densité de probabilité de la forme :

$$N_{(\bar{x}, \Gamma)}(y) = \frac{e^{-\frac{\bar{y}^t \Gamma \bar{y}}{2}}}{k} \quad (20)$$

où k (constante de normalisation) est solution de

$$k = \int_{\mathcal{M}} e^{-\frac{\bar{y}^t \Gamma \bar{y}}{2}} . d\mathcal{M}(y)$$

et Γ (matrice de concentration) est solution de

$$\Sigma = \frac{\int_{\mathcal{M}} \bar{x} \bar{y} . \bar{x} \bar{y}^t . e^{-\frac{\bar{y}^t \Gamma \bar{y}}{2}} . d\mathcal{M}(y)}{k}$$

Dans le cas où notre variété différentielle est de volume fini $V = \int_{\mathcal{M}} d\mathcal{M}$, on peut utiliser une matrice de concentration non-inversible (i.e. de rang non maximal). Ainsi, la loi gaussienne de matrice de concentration nulle (centrée autour d'une espérance arbitraire) n'est autre que la loi uniforme. La loi uniforme est donc un cas particulier de loi gaussienne.

Il est important de noter que la constante de normalisation et la matrice de covariance ne sont pas - comme dans le cas d'un espace vectoriel - simplement reliées à la matrice de concentration. On montre cependant que dans le cas d'une *faible covariance*, les formules du cas vectoriel constituent de bonnes approximations :

$$\Sigma \text{ faible} \quad \Rightarrow \quad \begin{aligned} \Gamma &\simeq \Sigma^{-1} \\ k &\simeq (2\pi)^{\binom{d}{2}} . \sqrt{|G_{T_{\bar{x}}\mathcal{M}}| . \Sigma} \end{aligned} \quad (21)$$

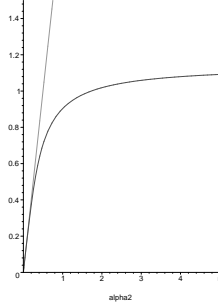
Exemple : Lien entre Variance et concentration pour les normales 3D gaussiennes isotrope Nous allons ici illustrer le lien entre covariance et concentration pour des normales en 3D qui suivent une loi gaussienne isotrope. Il s'agit des lois pour lesquelles la matrice de concentration est proportionnelle à l'identité : $\Gamma = \frac{1}{\alpha^2} . Id$. Leur loi est donc de la forme :

$$p(\mathbf{m}) = \frac{1}{k} \exp \left(-\frac{(\widehat{\mathbf{n}}, \mathbf{m})^2}{2 . \alpha^2} \right)$$

Nous pouvons calculé la variance par la formule suivante (au dénominateur : formule analogue à l'équation 16, au numérateur : la constante de normalisation) :

$$Tr(\Sigma) = E(\theta^2) = \frac{\int_{\theta=0}^{\frac{\pi}{2}} \int_{\alpha=-\pi}^{\pi} \theta^2 \sin \theta \exp \left(-\frac{\theta^2}{2 . \alpha^2} \right) d\theta d\alpha}{\int_{\theta=0}^{\frac{\pi}{2}} \int_{\alpha=-\pi}^{\pi} \sin \theta \exp \left(-\frac{\theta^2}{2 . \alpha^2} \right) d\theta d\alpha}$$

Nous ne donnons pas ici les formules explicites (obtenues avec un logiciel de calcul symbolique), trop complexes, mais avons tracé $Tr(\Sigma)$ en fonction de α^2 :



On constate alors que $Tr(\Sigma)$ et $Tr(\Gamma^{-1}) = 2\alpha^2$ sont bien équivalents au voisinage de 0, comme le prévoyait l'equ. 21. En revanche, la variance tend vers une valeur finie lorsque α tend vers l'infini, i.e. lorsque la loi de probabilité tend vers une distribution uniforme. On vérifie que la limite est la variance d'une distribution uniforme $\sigma^2 = \pi - 2$, que nous avons déjà calculé (equ. 16).

Exemple : Tirage d'une normales 3D gaussiennes isotrope D'après ce qui précède, il faut, pour tirer une normale 3D gaussienne centrée autour de $\bar{\sigma}$ et de concentration Γ , tirer le vecteur $\vec{\sigma n}$ de l'espace tangent suivant une loi gaussienne de covariance Γ^{-1} (voir equ. 20) et calculer $\mathbf{n} = \exp_{\bar{\sigma}}(\vec{\sigma n}) = R_{\bar{\sigma} \wedge \vec{\sigma n}}(\bar{\sigma})$. Dans le cas d'une loi isotrope avec $\Gamma = \frac{1}{\alpha_0^2} Id$, on peut encore simplifier les choses. En effet, dans ce cas $v = \bar{\sigma} \wedge \vec{\sigma n}$ suit aussi une loi gaussienne isotrope d'écart-type α_0^2 dans l'espace tangent. On peut donc tirer directement v , sans passer par $\vec{\sigma n}$. Pour cela, on tire un vecteur aléatoire dans l'espace 3D, et on le projette dans le plan tangent pour éliminer la composante perpendiculaire à celui-ci. L'algorithme est donc le suivant :

Algorithme Tirage d'une normale aléatoire gaussienne isotrope

- Tirer un vecteur aléatoire a suivant une gaussienne centrée isotrope d'écart-type α_0^2 .
- Le projeter dans l'espace tangent : $v = a - \bar{\sigma} \cdot (a | \bar{\sigma})$.
- Calculer la normale : $\mathbf{n} = R_v(\bar{\sigma})$

3.3.3 Distance de Mahalanobis

Rappelons ici que la distance de Mahalanobis simple est définie, dans le cas vectoriel par $\mu^2(\mathbf{x}, \mathbf{y}) = \vec{\mathbf{x}} \mathbf{y}^t \cdot \Sigma_{\mathbf{x}}^{-1} \cdot \vec{\mathbf{x}} \mathbf{y}$. On peut alors se poser la question de savoir si on utilise la matrice de concentration Γ où l'inverse de la matrice de covariance Σ^{-1} pour la généraliser.

Deux arguments permettrons de nous décider pour ce dernier cas :

- Nous pouvons toujours calculer la covariance d'une primitive aléatoire, même si elle ne suit pas une loi gaussienne. La définition à partir de la matrice de covariance ne

repose donc pas sur ces lois. En revanche, une primitive aléatoire non-gaussienne n'a pas de matrice de concentration associée. On pourrait envisager d'utiliser la matrice de concentration de la gaussienne qui possède la même moyenne et la même covariance, faisant ainsi jouer un rôle particulier à ces lois. Les calculs seraient de plus compliqués.

- Nous verrons que, en utilisant la matrice de covariance, l'espérance de la distance de Mahalanobis entre une primitive aléatoire quelconque et sa moyenne est égale à la dimension de l'espace, comme dans le cas vectoriel.

Nous définirons donc la distance de Mahalanobis simple ainsi :

$$\mu^2(\mathbf{x}, \mathbf{y}) = \overrightarrow{\mathbf{x}}\mathbf{y}^t \cdot \Sigma_{\mathbf{xx}}^{-1} \cdot \overrightarrow{\mathbf{x}}\mathbf{y} \quad (22)$$

Propriétés La distance de Mahalanobis est invariante : pour une transformation rigide f fixée :

$$\mu^2(f \star \mathbf{x}, f \star \mathbf{y}) = \mu^2(\mathbf{x}, \mathbf{y})$$

On peut donc se ramener à la carte principale :

$$\mu^2(\mathbf{x}, \mathbf{y}) = \mu^2(\mathbf{f}_y^{-1} \star \mathbf{x}, 0)$$

Calculons maintenant l'espérance de la distance de Mahalanobis entre une primitive aléatoire *quelconque* et son espérance (lorsque celle-ci est unique) :

$$\begin{aligned} E[\mu^2(\mathbf{x}, \overline{\mathbf{x}})] &= \int_{\mathcal{M}} \overrightarrow{\mathbf{x}}\mathbf{y}^t \cdot \Sigma_{\mathbf{xx}}^{-1} \cdot \overrightarrow{\mathbf{x}}\mathbf{y} \cdot p_{\mathbf{x}}(\mathbf{y}) \cdot d\mathcal{M}(\mathbf{y}) \\ &= \int_{\mathcal{M}} \text{Tr}(\Sigma_{\mathbf{xx}}^{-1} \cdot \overrightarrow{\mathbf{x}}\mathbf{y} \cdot \overrightarrow{\mathbf{x}}\mathbf{y}^t) \cdot p_{\mathbf{x}}(\mathbf{y}) \cdot d\mathcal{M}(\mathbf{y}) \\ &= \text{Tr} \left(\Sigma_{\mathbf{xx}}^{-1} \cdot \int_{\mathcal{M}} \overrightarrow{\mathbf{x}}\mathbf{y} \cdot \overrightarrow{\mathbf{x}}\mathbf{y}^t \cdot p_{\mathbf{x}}(\mathbf{y}) \cdot d\mathcal{M}(\mathbf{y}) \right) \\ &= \text{Tr}(\Sigma_{\mathbf{xx}}^{-1} \cdot \Sigma_{\mathbf{xx}}) = \text{Tr}(Id) \\ &= d \end{aligned}$$

Le résultat est bien identique à celui du cas vectoriel.

3.3.4 Test du χ^2

La distance de Mahalanobis pour une primitive gaussienne suit approximativement une loi du χ^2 à d degrés de libertés (en particulier lorsque la covariance est faible). On peut donc utiliser le test du χ^2 comme dans le cas vectoriel.

4 Statistiques approchées sur les normales aléatoires

Les notions statistiques rigoureuses que nous venons de définir se révèlent assez peu maniables dans le cas des normales : l'espérance ne peut-être calculée que de façon itérative, des doutes subsistent lorsqu'elle n'est pas unique, et la matrice de covariance est définie dans l'espace tangent à l'espérance, qui est variable.

Il a donc semblé utile de justifier ou créer des méthodes de calcul approchée mais plus maniable de l'espérance et la covariance d'une primitive. Ces méthodes sont basées sur l'utilisation d'une approximation de la distance intrinsèque dans les formules développées dans la partie 3. Précisons que nous ne modifierons pas la formule donnant l'espérance d'une observable (nous utiliserons toujours la métrique invariante, même si nous n'utilisons plus la distance associée).

Nous présenterons d'abord un cadre général à ce genre d'approximation, permettant d'apprécier leurs qualités. Nous verrons ensuite rapidement le cas de la distance vectorielle, qui permet de justifier l'approximation la plus simple de l'espérance au sens de Fréchet d'une normale aléatoire. Nous définirons ensuite le tenseur d'une normale aléatoire et verrons comment l'utiliser pour retrouver l'approximation de l'espérance au sens de Fréchet et de la covariance relative à la distance tensorielle. Nous comparerons enfin ces différentes méthodes.

4.1 Principe

Nous définirons par la suite une distance entre normales notée d_a et reliée à la distance intrinsèque d par une formule du type :

$$d_a = \Psi_a(d) \quad (23)$$

Pour être, comme d , une distance, d_a doit être :

- symétrique : ce sera toujours le cas ici car d est symétrique.
- définie positive : d étant elle-même définie positive, il faut et il suffit que Ψ_a soit définie positive ($(d > 0) \Rightarrow (\Psi_a(d) > 0)$ et $\Psi_a(0) = 0$).
- conforme à l'inégalité triangulaire : il suffit que Ψ_a soit croissante et de dérivé décroissante (et vérifie donc $\Psi_a(x + y) \leq \Psi_a(x) + \Psi_a(y)$). On a alors $d_a(\mathbf{n}_1, \mathbf{n}_2) = \Psi_a(d(\mathbf{n}_1, \mathbf{n}_2)) \leq \Psi_a(d(\mathbf{n}_1, \mathbf{n}_3) + d(\mathbf{n}_3, \mathbf{n}_2)) \leq \Psi_a(d(\mathbf{n}_1, \mathbf{n}_3)) + \Psi_a(d(\mathbf{n}_3, \mathbf{n}_2)) = d_a(\mathbf{n}_1, \mathbf{n}_3) + d_a(\mathbf{n}_3, \mathbf{n}_2)$. On remarquera que dans ce cas, notre distance approchée est une M-estimée [17] de la distance intrinsèque, et est donc censée être plus robuste.

On supposera en général la dérivé de Ψ_a égale à 1 en 0 : $\frac{\partial \Psi_a}{\partial d}(0) = 1$, pour que les deux distances d et d_a soient directement comparables au voisinage de 0.

Nous utiliserons par la suite aussi la fonction ψ_a définie par $\psi_a(d) = \Psi_a(d)/d$. On a donc :

$$d_a = \psi_a(d).d$$

La distance d étant invariante, la distance d_a sera elle aussi invariante. En revanche, elle n'est reliée à aucune métrique Riemannienne invariante.

4.1.1 Approximation de l'espérance

L'espérance de Fréchet (resp. Karcher) d'une normale aléatoire \mathbf{n} est définie comme l'ensemble des normales \mathbf{m} minimisant globalement (resp. localement) la variance $\sigma_{\mathbf{n}}^2(\mathbf{m}) = E_{\mathbf{n}}[d^2(\mathbf{n}, \mathbf{m})]$ (voir 3.2.1). Ces normales sont regroupées dans l'ensemble $\mathbb{E}[\mathbf{n}]$.

Nous approximerons donc l'espérance au sens de Fréchet (resp. Karcher) en prenant l'ensemble des normales \mathbf{m} minimisant globalement (resp. localement) la variance approchée

$$\sigma_{a \mathbf{n}}^2(\mathbf{m}) = E_{\mathbf{n}}[d_a^2(\mathbf{n}, \mathbf{m})] \quad (24)$$

et regrouperons ces normales dans l'ensemble $\mathbb{E}_a[\mathbf{n}]$.

4.1.2 Approximation de la covariance

La covariance d'une primitive aléatoire \mathbf{x} par rapport à une primitive fixe \mathbf{y} est définie par : $\Sigma_{\mathbf{nn}}(\mathbf{m}) = E_{\mathbf{n}}[\overrightarrow{\mathbf{nm}} \cdot \overrightarrow{\mathbf{nm}}^t]$. Ainsi, elle vérifie $\sigma_{\mathbf{n}}^2(\mathbf{m}) = Tr(\Sigma_{\mathbf{nn}}(\mathbf{m}))$.

Or la variance approchée peut se ré-écrire $\sigma_{a \mathbf{n}}^2(\mathbf{m}) = E_{\mathbf{n}}[(\Psi_a(d(\mathbf{n}, \mathbf{m})))^2] = E_{\mathbf{n}}[(\Psi_a(\|\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}\|))^2] = E_{\mathbf{n}}[\psi_a^2(\|\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}\|) \cdot \|\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}\|^2]$. Il semble alors assez naturel de généraliser la fonction Ψ_a au plan tangent en \mathbf{m} en écrivant :

$$\Psi_a(\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}) = \overrightarrow{\mathbf{m}\hat{\mathbf{n}}} \cdot \frac{\Psi_a(\|\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}\|)}{\|\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}\|} = \overrightarrow{\mathbf{m}\hat{\mathbf{n}}} \cdot \psi_a(\|\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}\|) \quad (25)$$

On a alors $\sigma_{\mathbf{n}}^2(\mathbf{m}) = E_{\mathbf{n}}[Tr(\Psi_a(\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}) \cdot \Psi_a(\overrightarrow{\mathbf{m}\hat{\mathbf{n}}})^t)] = Tr(E_{\mathbf{n}}[\Psi_a(\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}) \cdot \Psi_a(\overrightarrow{\mathbf{m}\hat{\mathbf{n}}})^t])$. On définit donc la matrice de covariance approchée de la façon suivante :

$$\begin{aligned} \Sigma_{a \mathbf{nn}}(\mathbf{m}) &= E_{\mathbf{n}}[\Psi_a(\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}) \cdot \Psi_a(\overrightarrow{\mathbf{m}\hat{\mathbf{n}}})^t] \\ \Sigma_{a \mathbf{nn}} &= E_{\mathbf{n}}[\Psi_a(\overrightarrow{\mathbf{n}\hat{\mathbf{n}}}) \cdot \Psi_a(\overrightarrow{\mathbf{n}\hat{\mathbf{n}}})^t] \end{aligned} \quad (26)$$

4.1.3 Caractérisation de l'espérance de Karcher

Nous avons montré (equ. 18) que l'espérance $\overline{\mathbf{n}}$ au sens de Karcher de la normale vérifiait :

$$E[\overrightarrow{\mathbf{n}\hat{\mathbf{n}}}] = 0$$

Nous pensons que ceci s'adapte pour notre approximation (la démonstration doit être analogue³, bien que certaines difficultés pourraient apparaître au niveau du lieu de coupure). L'approximation de l'espérance $\overline{\mathbf{n}}$ au sens de Karcher de la normale vérifierait alors :

$$E[\dot{\Psi}_a(\|\overrightarrow{\mathbf{n}\hat{\mathbf{n}}}\|) \cdot \Psi_a(\overrightarrow{\mathbf{n}\hat{\mathbf{n}}})] = 0 \quad (27)$$

3. Il nous faut la dérivée, dans la carte locale, de la distance $d_a^2(\mathbf{n}, \mathbf{m}) = \Psi_a(\|\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}\|)^2$:

$$\begin{aligned} \frac{\partial \Psi_a(\|\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}\|)^2}{\partial \overrightarrow{\mathbf{n}}} &= 2 \cdot \Psi_a(\|\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}\|) \cdot \frac{\partial \Psi_a(\|\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}\|)}{\partial \overrightarrow{\mathbf{n}}} = 2 \cdot \Psi_a(\|\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}\|) \cdot \dot{\Psi}_a(\|\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}\|) \frac{\partial \|\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}\|}{\partial \overrightarrow{\mathbf{n}}} \\ &= 2 \cdot \Psi_a(\|\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}\|) \cdot \dot{\Psi}_a(\|\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}\|) \cdot [G_{\text{exp}} \overrightarrow{\mathbf{y}} \bullet \bullet] \cdot \frac{\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}}{\|\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}\|} = 2 \cdot \dot{\Psi}_a(\|\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}\|) \cdot [G_{\text{exp}} \overrightarrow{\mathbf{y}} \bullet \bullet] \cdot \Psi_a(\overrightarrow{\mathbf{m}\hat{\mathbf{n}}}) \end{aligned}$$

La suite du raisonnement est alors identique.

4.1.4 Propagation des approximations

Avec des distances de ce type, les propriétés de propagation par une transformation fixe de l'espérance et de la covariance sont conservées :

$$\begin{aligned}\mathbb{E}_a[\mathbf{f} \star \mathbf{n}] &= \mathbf{f} \star \mathbb{E}_a[\mathbf{n}] \\ \Sigma_{\mathbf{f} \star \mathbf{n} \mathbf{f} \star \mathbf{n}} &= J(\mathbf{f}) \cdot \Sigma_{\mathbf{nn}} \cdot J(\mathbf{f})^t\end{aligned}$$

4.1.5 Approximation de l'espérance d'une gaussienne

Nous allons essayer de montrer que l'approximation de l'espérance se révèle être exacte pour une loi gaussienne.

Loi symétrique par rapport à une normale Soit une normale aléatoire \mathbf{n} définie par une densité de probabilité $p_{\mathbf{n}}(\mathbf{n})$. Nous dirons qu'elle suit une loi symétrique décroissante s'il existe une normale $\tilde{\mathbf{n}}$ telle que pour tout \mathbf{m} on ait : $p_{\mathbf{n}}(\exp_{\tilde{\mathbf{n}}} \tilde{\mathbf{n}}\mathbf{m}) = p_{\mathbf{n}}(-\exp_{\tilde{\mathbf{n}}} \tilde{\mathbf{n}}\mathbf{m})$, i.e. si la loi est invariante par une symétrie d'axe $\tilde{\mathbf{n}}$.

Or la symétrie de la loi de \mathbf{n} implique que $E \left[\dot{\Psi}_a(\|\tilde{\mathbf{n}}\|) \cdot \Psi_a(\tilde{\mathbf{n}}) \right] = E \left[-\dot{\Psi}_a(\|\tilde{\mathbf{n}}\|) \cdot \Psi_a(\tilde{\mathbf{n}}) \right] = -E \left[\dot{\Psi}_a(\|\tilde{\mathbf{n}}\|) \cdot \Psi_a(\tilde{\mathbf{n}}) \right]$ et donc :

$$E \left[\dot{\Psi}_a(\|\tilde{\mathbf{n}}\|) \cdot \Psi_a(\tilde{\mathbf{n}}) \right] = 0$$

La normale $\tilde{\mathbf{n}}$ a donc des chances de faire partie de l'espérance de Karcher voire de l'espérance de Fréchet. Il faut cependant vérifier qu'il s'agit bien d'un minimum local ou global (et non d'un maximum par exemple) de la variance.

Loi symétrique mono-modale Nous pensons que, lorsque la loi est mono-modale et symétrique, le mode, qui vérifie $E \left[\dot{\Psi}_a(\|\tilde{\mathbf{n}}\|) \cdot \Psi_a(\tilde{\mathbf{n}}) \right] = 0$, est nécessairement le minimum de la variance et de la variance approchée.

Conclusion Lorsque la loi est symétrique mono-modale (ce qui est en particulier le cas des gaussiennes), on a donc vraisemblablement l'équivalence entre les différentes approximations de l'espérance de Fréchet et sa véritable valeur.

4.1.6 Cas général

Dans le cas général (loi de probabilité quelconque), on peut donner un majorant de l'erreur commise en utilisant nos statistiques approchées. Il faut pour cela faire appel à des travaux mathématiques plus généraux.

Nous nous appuyerons ici sur l'article [15]. Cet article est basé sur le concept de connecteur, qui généralise les cartes locales définies dans le plan tangent : en résumé, un connecteur

est une fonction γ qui à un couple (\mathbf{x}, \mathbf{y}) de primitives associe un vecteur de l'espace tangent \overrightarrow{xy} . Il peut s'agir de la fonction log (on parle alors de connecteur géodésique), d'une fonction convexe sur les géodésiques (on parle alors de connecteur convexe), ou de fonctions plus générales. Dans notre cas, $\gamma(\mathbf{m}, \mathbf{n}) = \Psi_a(\overrightarrow{m\hat{n}})$ définit effectivement un connecteur convexe (car la fonction Ψ_a est elle-même convexe). L'article définit la notion de γ -barycentre d'une primitive aléatoire (partie 1), qui est en fait équivalente à notre espérance de Fréchet au sens de la distance d_a . Il montre ensuite (partie 2) que l'erreur commise sur le γ -barycentre est de la forme $O(\sigma)$, i.e. inférieure à $k\sigma$, où σ est l'écart-type de la primitive aléatoire.

En résumé, l'erreur commise sur l'estimation de l'espérance de Fréchet d'une normale aléatoire par les méthodes de statistiques approchées est de l'ordre de σ où σ est l'écart-type de la normale aléatoire.

4.2 Utilisation de la distance vectorielle

Nous allons voir ici quels sont les effets de l'utilisation de la distance vectorielle entre deux normales définie comme la distance au sens vectoriel entre les deux représentations unitaires des normales. Il s'agit donc, pour deux normales \mathbf{m} et \mathbf{n} de :

$$d_v(\mathbf{m}, \mathbf{n}) = \|\mathbf{m} - \mathbf{n}\| = 2 \sin\left(\frac{\widehat{(\mathbf{m}, \mathbf{n})}}{2}\right) = d(\mathbf{m}, \mathbf{n}) + O(d(\mathbf{m}, \mathbf{n})^3)$$

où $\|\cdot\|$ est la norme quadratique usuelle de \mathbb{R}^d . Cette distance vérifie aussi :

$$d_v(\mathbf{m}, \mathbf{n})^2 = \|\mathbf{m} - \mathbf{n}\|^2 = 2(1 - (\mathbf{m}|\mathbf{n}))$$

où $(\cdot|\cdot)$ est le produit scalaire usuel de \mathbb{R}^d .

On a donc :

$$\Psi_v(d) = 2 \sin\left(\frac{d}{2}\right) \quad \psi_v(d) = \frac{\sin\left(\frac{d}{2}\right)}{\frac{d}{2}}$$

Remarquons au passage que cette distance est bien adaptée au cas orienté.

4.2.1 Calcul de l'approximation vectorielle de l'espérance

Nous allons effectuer ce calcul dans le cas discret, et retrouver ainsi une technique souvent utilisée pour calculer la moyenne de plusieurs normales. Soit donc un ensemble de normales \mathbf{n}_i représentées par les vecteurs unitaires de coordonnées \mathbf{n}_i^k et munies de leur probabilité $P(\mathbf{n}_i)$. L'approximation vectorielle de l'espérance est l'ensemble des normales \mathbf{m} représentées par les vecteurs unitaires de coordonnées \mathbf{m}^k minimisant :

$$\sum_i P(\mathbf{n}_i) \cdot \|\mathbf{m} - \mathbf{n}_i\|_{\mathbb{R}^d}^2 = \sum_i P(\mathbf{n}_i) \cdot (\mathbf{m} - \mathbf{n}_i | \mathbf{m} - \mathbf{n}_i)$$

En utilisant un multiplicateur de Lagrange α pour prendre en compte la contrainte $\|\mathbf{m}\| = 1$, on a donc :

$$\mathbf{m} = \arg \min_{\mathbf{m}, \alpha} \frac{\left[\sum_i P(\mathbf{n}_i) \cdot (\mathbf{m} - \mathbf{n}_i | \mathbf{m} - \mathbf{n}_i) + \alpha \cdot (\|\mathbf{m}\|^2 - 1) \right]}{2}$$

En dérivant par rapport à \mathbf{m} , on obtient : $\sum_i P(\mathbf{n}_i) \cdot (\mathbf{m} - \mathbf{n}_i) + \alpha \cdot \mathbf{m} = 0$ et donc : $(1 + \alpha)\mathbf{m} = \sum_i P(\mathbf{n}_i) \cdot \mathbf{n}_i$. On calcule alors la valeur de α grâce à la contrainte $\|\mathbf{m}\| = 1$. Finalement :

$$\mathbf{m} = \frac{\sum_i P(\mathbf{n}_i) \cdot \mathbf{n}_i}{\|\sum_i P(\mathbf{n}_i) \cdot \mathbf{n}_i\|}$$

On a donc la proposition suivante :

Proposition 7 *Espérance de Fréchet au sens de la distance vectorielle*

On obtient l'approximation vectorielle de l'espérance au sens de Fréchet d'une normale aléatoire en normalisant la moyenne des normales considérées comme de simples vecteurs de \mathbb{R}^d .

Ce procédé, souvent utilisé, est donc une approximation de la véritable espérance d'une normale aléatoire. Il échoue cependant lorsque $\sum_i P(\mathbf{n}_i) \cdot \mathbf{n}_i = 0$.

4.2.2 Approximation de la matrice de covariance

L'approximation de la covariance ne présente ici aucun intérêt pratique. Nous ne détaillons pas les calculs.

4.3 Tenseur d'une normale aléatoire

La distance tensorielle et l'approximation de l'espérance et la covariance qu'elle permet de définir est étroitement liée au tenseur d'une normale aléatoire (d'où son nom). Nous allons donc dans un premier temps nous familiariser avec ce tenseur, avant de revenir, à la section suivante à nos approximations.

Ce tenseur peut être défini comme la covariance de la normale aléatoire, calculée comme pour un vecteur aléatoire de l'espace nD classique, i.e. en oubliant qu'il s'agit d'un vecteur unitaire. Cet oubli est, inutile de le préciser, lourd de conséquence, et nous ne ferons plus allusion à cette matrice de covariance dans la suite, pour éviter les confusions avec la vraie matrice de covariance des normales, définie dans l'espace tangent à la moyenne.

4.3.1 Tenseur associé à une direction

Plaçons nous dans \mathbb{R}^D muni de sa norme canonique. On peut associer à tout vecteur le tenseur suivant :

Définition 12 *Tenseur associé à un vecteur de \mathbb{R}^D*

Soit $x \in \mathbb{R}^D$.

Le tenseur associé à x et noté $T(x)$ est le tenseur d'ordre 2 donné par :

$$T(x) = x \otimes x$$

soit, sous forme matricielle :

$$T(x) = x \cdot x^t$$

$$(T(x))_{ij} = x_i \cdot x_j \quad (T(x))^i_j = x^i \cdot x_j \quad \dots$$

On remarque que $(y \otimes y = x \otimes x) \Leftrightarrow (y = \pm x)$. Autrement dit, l'utilisation des tenseurs nous fait perdre l'information sur l'orientation. Il est alors naturel de définir la relation d'équivalence suivante: $x \sim y \Leftrightarrow (x = \pm y)$, et la définition du tenseur reste valable dans l'espace non orienté \mathbb{R}^D / \sim . Elle y est même plus intéressante, car l'application $x \mapsto T(x)$ est alors injective.

On peut de plus restreindre cette définition à l'ensemble des normales orientées de \mathbb{R}^D qui forment le sous-ensemble \mathcal{S}_{D-1} identifié ensuite à la sphère unité $\{n \in \mathbb{R}^D / \|n\| = 1\}$, ou l'ensemble des directions (i.e. des normales non-orientées) de \mathbb{R}^D , qui forment l'ensemble $\mathcal{P}_{D-1} = \mathcal{S}_{D-1} / \sim$. Nous présentons le cas non-orienté dans la suite, mais ces définitions restent bien-sûr valable (mais ne sont plus biunivoque) dans le cas orienté.

Définition 13 *Tenseur associé à une normale non-orientée*

Soit une direction de \mathcal{P}_{D-1} , représentée par le vecteur unitaire non orienté $\pm n$.
Le tenseur associé à n et noté $T(n)$ est le tenseur d'ordre 2 donné par :

$$T(n) = n \otimes n$$

soit, sous forme matricielle :

$$T(n) = n \cdot n^t$$

$$(T(n))_{ij} = n_i \cdot n_j \quad (T(n))^i_j = n^i \cdot n_j \quad \dots$$

4.3.2 Tenseur associé à une direction aléatoire

Soit maintenant une direction aléatoire \mathbf{n} . On lui associe le tenseur aléatoire $T(\mathbf{n})$. Or l'ensemble des tenseurs d'ordre 2 forment un espace vectoriel réel de dimension finie. Le tenseur aléatoire $T(\mathbf{n})$ est donc un vecteur aléatoire, dont nous pouvons calculer l'espérance (en passant par exemple par les coordonnées de ce tenseur dans une base qui sont des variables aléatoires réelles). Nous utiliserons cette espérance pour représenter \mathbf{n} :

Définition 14 *Tenseur associé à une élément aléatoire*

Soit \mathbf{n} une direction aléatoire de \mathcal{P}_{D-1} .

Le tenseur associé à \mathbf{n} et noté $\overline{T(\mathbf{n})}$ est le tenseur d'ordre 2 donné par :

$$\overline{T(\mathbf{n})} = E[T(\mathbf{n})] = E[\mathbf{n} \otimes \mathbf{n}]$$

soit sous forme matricielle :

$$\overline{T(\mathbf{n})} = E[\mathbf{n} \mathbf{n}^t]$$

$$(\overline{T(\mathbf{n})})_{ij} = E[\mathbf{n}_i \mathbf{n}_j] \quad (\overline{T(\mathbf{n})})^i_j = E[\mathbf{n}^i \mathbf{n}_j]$$

4.3.3 Propriétés du tenseur associé à une direction aléatoire

La première propriété importante d'un tenseur de la forme $\overline{T(\mathbf{n})}$ est la forme de son application à deux vecteurs. On a en effet, par linéarité : $a \overline{T(\mathbf{n})} \overline{\otimes} b = E[a \overline{\otimes} (\mathbf{n} \otimes \mathbf{n}) \overline{\otimes} b]$. Or $a \overline{\otimes} (\mathbf{n} \otimes \mathbf{n}) \overline{\otimes} b = (a \overline{\otimes} \mathbf{n}) \otimes (\mathbf{n} \overline{\otimes} b) = (a|\mathbf{n})(\mathbf{n}|b)$. Finalement :

Proposition 8 *Application de $\overline{T(\mathbf{n})}$ à deux vecteurs*

Soient $\overline{T(\mathbf{n})}$ tenseur associé à une direction aléatoire de \mathcal{P}_{D-1} et a et b deux vecteurs de \mathbb{R}^D . On a :

$$\begin{aligned} a \overline{\otimes} \overline{T(\mathbf{n})} \overline{\otimes} b &= E[(a|\mathbf{n}).(\mathbf{n}|b)] \\ a \overline{\otimes} \overline{T(\mathbf{n})} \overline{\otimes} a &= E[(a|\mathbf{n})^2] \geq 0 \end{aligned}$$

On voit alors immédiatement que $\overline{T(\mathbf{n})}$ est un tenseur symétrique. Or il existe un important résultat sur les tenseurs symétriques réels d'ordre 2 : un tenseur T symétrique réel d'ordre 2 sur un espace euclidien E de dimension D possède exactement D valeurs propres réelles $\lambda_1, \lambda_2, \dots, \lambda_D$ (distinctes ou non). Les espaces propres étant orthogonaux entre-eux, il peut donc être diagonalisé dans une base orthonormale $\{v_i\}$ constituée de vecteurs propres.

De plus, on constate que $\overline{T(\mathbf{n})}$ est positif. Ses valeurs propres sont donc positives, et nous les mettrons sous forme carré : $\lambda_i = \alpha_i^2$.

En résumé :

Proposition 9 *Diagonalisation de $\overline{T(\mathbf{n})}$*

Soient $\overline{T(\mathbf{n})}$ tenseur associé à une direction aléatoire de \mathcal{P}_{D-1} .

$\overline{T(\mathbf{n})}$ est diagonalisable de valeurs propres $\alpha_1^2, \alpha_2^2, \dots, \alpha_D^2$, avec $\alpha_i \in \mathbb{R}^+$.

On peut de plus construire une base orthonormale $\{v_i\}$ de \mathbb{R}^D , où chaque v_i est un vecteur propre associé à la valeur propre α_i^2 . Une telle base est appelée base propre de $\overline{T(\mathbf{n})}$.

L'expression de $\overline{T(\mathbf{n})}$ dans cette base est : $\overline{T(\mathbf{n})} = \sum_i \alpha_i^2 (v_i \otimes v_i) = \sum_i (\alpha_i v_i) \otimes (\alpha_i v_i)$.

Une dernière propriété importante de $\overline{T(\mathbf{n})}$ porte sur sa trace (égale à la somme des valeurs propres) : par linéarité, $Tr(\overline{T(\mathbf{n})}) = E[Tr(\mathbf{n} \otimes \mathbf{n})]$. Or $Tr(\mathbf{n} \otimes \mathbf{n}) = \mathbf{n} \overline{\otimes} \mathbf{n} = (\mathbf{n}|\mathbf{n}) = 1$. On a donc :

Proposition 10 *Trace de $\overline{T(\mathbf{n})}$*

Soient $\overline{T(\mathbf{n})}$ tenseur associé à une direction aléatoire de \mathcal{P}_{D-1} . La trace de $\overline{T(\mathbf{n})}$, qui est aussi la somme de ses valeurs propres, est unitaire :

$$Tr(\overline{T(\mathbf{n})}) = \sum_i \alpha_i^2 = 1$$

4.3.4 Représentation sous forme d'ellipsoïdes

Il est commode de représenter les tenseurs symétriques positifs par des ellipsoïdes. En effet, une fois notre tenseur diagonalisé $\overline{T(\mathbf{n})} = \sum_i \alpha_i^2 (v_i \otimes v_i)$, on peut lui associer de façon biunivoque l'ellipsoïde dont les axes principaux sont colinéaires à v_i et de demi-longueur α_i .

Dans le cas où les valeurs propres sont toutes non nulles, il s'agit donc de l'ellipsoïde d'équation $\sum \frac{(x|v_i)^2}{\alpha_i^2} = 1$. Le tenseur est alors inversible, d'inverse $\overline{T(\mathbf{n})}^{-1} = \sum_i \frac{1}{\alpha_i^2} (v_i \otimes v_i)$, si bien que l'équation de notre ellipsoïde n'est autre que :

$$x \overline{T(\mathbf{n})}^{-1} \overline{x} = Cte$$

Dans le cas où au moins l'une de nos valeurs propres est nulle, il s'agit d'un ellipsoïde dégénéré. Nous donnerons des exemples de tels tenseurs en présentant les tenseurs Stick et Plate de Lee et Médioni (voir pages suivantes).

Prenons un exemple en 2D. Soit le tenseur de matrice $\begin{bmatrix} 1/2 & 1/4 \\ 1/4 & 1/2 \end{bmatrix}$ dans la base canonique. Ses valeurs propres sont $1/4$ et $3/4$, et les vecteurs propres unitaires associés sont par exemple respectivement $\begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$ et $\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$. Ils forment bien une base orthonormale, et le tenseur est alors représenté par l'ellipse de la figure 1 (les vecteurs propres sont représentés par les flèches).

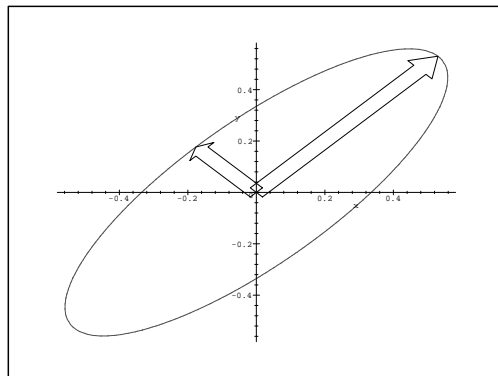


FIG. 1 – Exemple de représentation sous forme d'ellipsoïde du tenseur d'une normale aléatoire

4.3.5 Exemples de tenseurs associés à une direction aléatoire

Le principal intérêt des tenseurs utilisés ici est l'effet tout-à-fait particulier que le passage à l'espérance a sur eux. Avant de le traiter de manière rigoureuse, nous allons essayer d'illustrer cet effet en prenant quelques exemples de normales aléatoires simples, où l'espérance d'une fonction réelle est simple à définir. Commençons par examiner le cas d'une loi binomiale en 2D. Soient donc deux normales parfaitement déterminées. En notant $\{e_1, e_2\}$ la base canonique de \mathbb{R}^2 , on prend par exemple les normales représentées par $n_1 = \sin(-\theta)e_1 + \cos(-\theta)e_2$ et $n_2 = \sin(\theta)e_1 + \cos(\theta)e_2$. Les tenseurs associés ont pour matrice dans la base canonique :

$$[T(n_1)] = \begin{bmatrix} \sin^2 \theta & -\sin \theta \cos \theta \\ -\sin \theta \cos \theta & \cos^2 \theta \end{bmatrix} \quad \text{et} \quad [T(n_2)] = \begin{bmatrix} \sin^2 \theta & \sin \theta \cos \theta \\ \sin \theta \cos \theta & \cos^2 \theta \end{bmatrix}$$

Le tenseur $T(n_1)$ a pour valeurs propres 1 et 0, et les vecteurs propres associés sont n_1 et $n_{1\perp}$, vecteur perpendiculaire à n_1 . Il en va de même pour $T(n_2)$. $T(n_1)$ et $T(n_2)$ sont donc associés aux ellipses dé-générées de la figure 2 gauche (On a ici pris $\theta = \frac{\pi}{12}$, et on a exagéré la largeur du petit axe, qui est en principe nulle). Le tenseur associé à la normale aléatoire \mathbf{n} qui suit une loi binomiale de la forme $Pr(\mathbf{n} = n_1) = Pr(\mathbf{n} = n_2) = 1/2$, est tout simplement la moyenne des deux tenseurs qui précèdent. Sa matrice est donc :

$$[\overline{T(\mathbf{n})}] = \left[\frac{T(n_1)}{2} + \frac{T(n_2)}{2} \right] = \begin{bmatrix} \sin^2 \theta & 0 \\ 0 & \cos^2 \theta \end{bmatrix}$$

Les vecteurs propres et valeurs propres apparaissent immédiatement, l'ellipse associée est représenté sur la figure 2 droite.

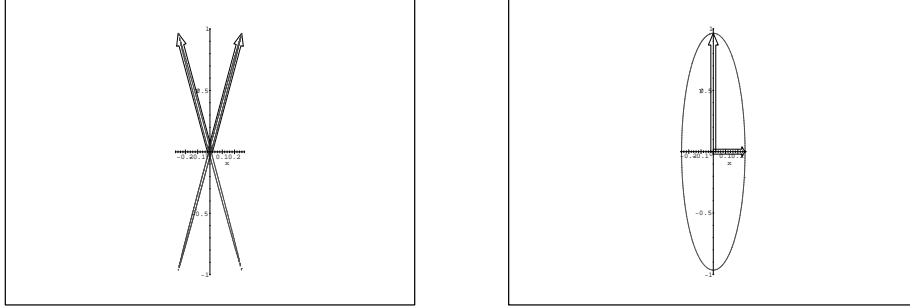


FIG. 2 – Tenseur d'une normale aléatoire suivant une loi binomiale

On constate donc que l'aspect bimodal de \mathbf{n} est voilé. En revanche, le grand-axe de l'ellipse (et donc la vecteur propre associée à la plus grande valeur propre) correspond à l'espérance de la normale telle qu'on se la figure intuitivement. De plus, le fait que le petit

axe ne soit plus de longueur nulle semble exprimer le fait que les réalisations de la variable aléatoire sont dispersées autour d'une valeur. Sa longueur, qui augmente avec la dispersion (lorsque θ augmente dans notre exemple), suggère un lien avec la covariance.

Soit maintenant 4 normales en 3D. On utilise la base canonique $\{e_1, e_2, e_3\}$, et les 4 normales sont de la forme $n_1 = \sin(-\theta_1)e_1 + \cos(-\theta_1)e_3$, $n_2 = \sin(\theta_1)e_1 + \cos(\theta_1)e_3$, $n_3 = \sin(-\theta_2)e_2 + \cos(-\theta_2)e_3$ et $n_4 = \sin(\theta_2)e_2 + \cos(\theta_2)e_3$. Les tenseurs associés sont donnés par leurs matrices :

$$[T(n_1)] = \begin{bmatrix} \sin^2 \theta_1 & 0 & -\sin \theta_1 \cos \theta_1 \\ 0 & 0 & 0 \\ -\sin \theta_1 \cos \theta_1 & 0 & \cos^2 \theta_1 \end{bmatrix}$$

$$[T(n_2)] = \begin{bmatrix} \sin^2 \theta_1 & 0 & \sin \theta_1 \cos \theta_1 \\ 0 & 0 & 0 \\ \sin \theta_1 \cos \theta_1 & 0 & \cos^2 \theta_1 \end{bmatrix}$$

$$[T(n_3)] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sin^2 \theta_2 & -\sin \theta_2 \cos \theta_2 \\ 0 & -\sin \theta_2 \cos \theta_2 & \cos^2 \theta_2 \end{bmatrix}$$

$$[T(n_4)] = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sin^2 \theta_2 & \sin \theta_2 \cos \theta_2 \\ 0 & \sin \theta_2 \cos \theta_2 & \cos^2 \theta_2 \end{bmatrix}$$

Les ellipsoïdes dé-générés sont représentés dans la figure 3 gauche.

Soit maintenant une normale aléatoire \mathbf{n} définie par $Pr(\mathbf{n} = n_1) = Pr(\mathbf{n} = n_2) = Pr(\mathbf{n} = n_3) = Pr(\mathbf{n} = n_4) = 1/4$. Le tenseur associé est donc donné par la matrice :

$$[\overline{T(\mathbf{n})}] = \left[\frac{T(n_1)}{4} + \frac{T(n_2)}{4} + \frac{T(n_3)}{4} + \frac{T(n_4)}{4} \right] = \begin{bmatrix} \frac{\sin^2 \theta_1}{2} & 0 & 0 \\ 0 & \frac{\sin^2 \theta_2}{2} & 0 \\ 0 & 0 & \frac{\cos^2 \theta_1 + \cos^2 \theta_2}{2} \end{bmatrix}$$

Il est représenté par l'ellipsoïde de la figure 3 droite.

On constate qu'une fois de plus, les vecteurs propres associés à la plus grande valeur propre semble correspondre à la notion intuitive d'espérance. Les 2 autres vecteurs propres sont dans la plan perpendiculaire au premier, et les valeurs propres associées semblent aussi représenter une divergence (plus importante suivant l'axe y, dans notre cas). Ceci suggère un lien entre la sous-matrice de $\overline{T(\mathbf{n})}$ dans la base formée par les vecteurs propres obtenue après élimination de la première coordonnée (première ligne et première colonne donc) et une certaine matrice de covariance.

Tenseurs stick, plate et ball Certains tenseurs particuliers peuvent retenir l'attention : il s'agit des tenseurs dont certaines valeurs propres sont nulles et dont les autres valeurs propres sont égales. Nous donnerons ici l'exemple en 3D, avec ce que Médioni a appelé

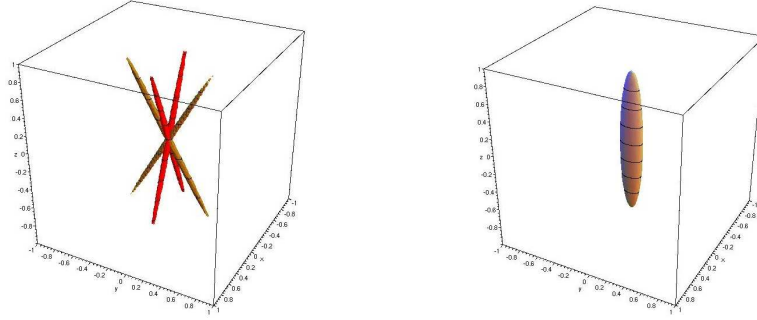


FIG. 3 – Tenseur d'une normale aléatoire suivant une loi quadrimodale

les tenseurs Stick, Plate et Ball [11], mais on pourrait faire de même en n'importe quelle dimension.

Nous travaillerons dans la carte définie dans la partie 2.5. Nous considérerons en général la densité de probabilité sous la forme $p_{\mathbf{n}}(n(\theta, \alpha))$, et non dans la carte locale. On aura alors :

$$\overline{T(\mathbf{n})} = \int_{\theta=0}^{\frac{\pi}{2}} \int_{\alpha=-\pi}^{\pi} \begin{bmatrix} \cos \theta \\ \sin \theta \cos \alpha \\ \sin \theta \sin \alpha \end{bmatrix} \otimes \begin{bmatrix} \cos \theta \\ \sin \theta \cos \alpha \\ \sin \theta \sin \alpha \end{bmatrix} p_{\mathbf{n}}(n(\theta, \alpha)) \sin \theta d\theta d\alpha$$

La matrice de $\overline{T(\mathbf{n})}$ est alors :

$$[\overline{T(\mathbf{n})}] = \int_0^{\frac{\pi}{2}} \int_{-\pi}^{\pi} \begin{bmatrix} \cos^2 \theta & \cos \theta \sin \theta \cos \alpha & \cos \theta \sin \theta \sin \alpha \\ \dots & \sin^2 \theta \cos^2 \alpha & \sin^2 \theta \cos \alpha \sin \alpha \\ \dots & \dots & \sin^2 \theta \sin^2 \alpha \end{bmatrix} p_{\mathbf{n}}(n(\theta, \alpha)) \sin \theta d\theta d\alpha$$

Tenseur associé à une normale déterministe - Tenseur Stick Lee et Médioni associent à une donnée du type normale à une surface le tenseur $n \otimes n$ (qui possède une valeur propre unitaire et les autres nulles). Ceci correspond donc à l'hypothèse que la normale est déterministe, ce qui est le cas. Nous avons représenté dans la figure 4 la surface et sa normale, puis le tenseur correspondant.

Tenseur associé à un élément de courbe - Tenseur Plate Lee et Médioni associent à un élément de courbe, défini par sa tangente t le tenseur $e_1 \otimes e_1 + e_2 \otimes e_2$ (qui possède deux valeurs propres égales et une dernière nulle), où (t, e_1, e_2) forme un repère orthonormal.

Or un élément de courbe déterministe nous apporte l'information suivante : la normale à la courbe est perpendiculaire à la tangente. Autrement dit, dans le repère (t, e_1, e_2) , on a $\theta = \frac{\pi}{2}$, mais on ne dispose d'aucune information sur γ . L'hypothèse la plus faible est alors de supposer que γ suit une loi de probabilité uniforme sur $\{\frac{\pi}{2}\} \times [-\pi, \pi]$.

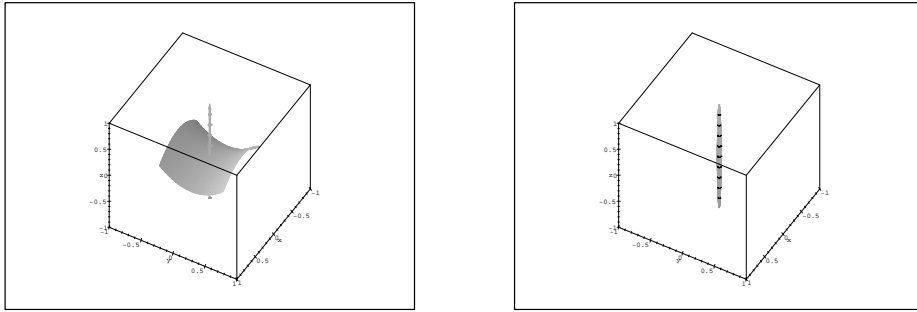


FIG. 4 – Un élément de surface et sa normale (gauche) et le tenseur stick associé à cette normale (droite)

On a donc $p_{\mathbf{n}}(\mathbf{n}(\theta, \gamma)) = \delta_{\frac{\pi}{2}}(\theta) \frac{1}{2\pi}$, et en reportant, on trouve bien $\overline{T(\mathbf{n})} = \frac{e_1 \otimes e_1 + e_2 \otimes e_2}{2}$, soit le tenseur utilisé par Lee et Médioni après normalisation.

La figure 5 représente une courbe et sa tangente, puis le tenseur correspondant.

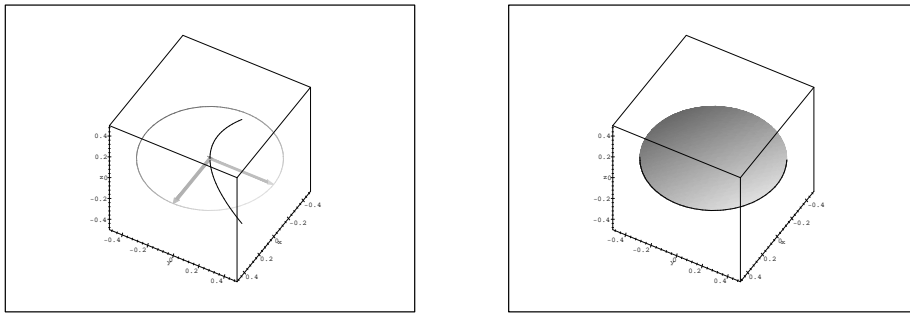


FIG. 5 – Un élément de courbe et sa tangente (gauche) et le tenseur plate associé à cette tangente (droite)

Tenseur associé à un point - Tenseur Ball Dans le cas d'un point, on ne dispose d'aucune information sur la normale. L'hypothèse la plus faible est alors de supposer qu'elle suit une loi de probabilité uniforme: $p_{\mathbf{n}}(m) = \frac{1}{2\pi}$. Le résultat, que l'on peut obtenir par calcul direct ou par simple considération de symétrie, est alors $\overline{T(\mathbf{n})} = \frac{e_1 \otimes e_1 + e_2 \otimes e_2 + e_3 \otimes e_3}{3}$

(qui possède trois valeurs propres égales) pour n'importe quelle base (e_1, e_2, e_3) orthonormée, qui est, au coefficient de normalisation prêt le tenseur utilisé par Lee et Médioni.

Le tenseur ball est représenté sur la figure 6.

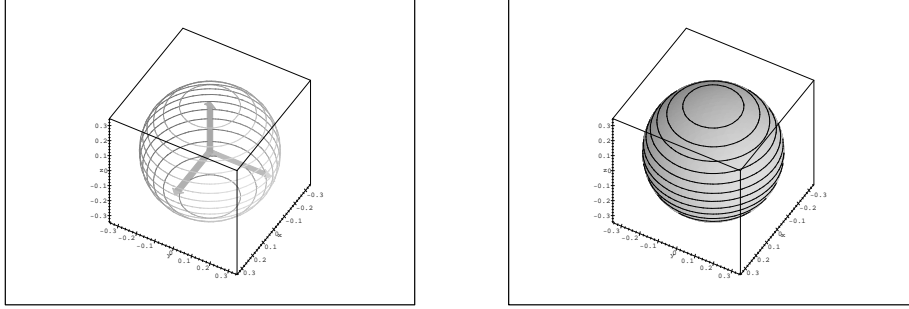


FIG. 6 – *Le tenseur ball*

Décomposition d'un tenseur associée à une normale aléatoire Nous avons déjà vu qu'un tenseur pouvait se décomposer suivant ses 3 valeurs propres : $\overline{T(\mathbf{n})} = \alpha_1^2(v_1 \otimes v_1) + \alpha_2^2(v_2 \otimes v_2) + \alpha_3^2(v_3 \otimes v_3)$. Mais on peut aussi regrouper ces termes pour obtenir des tenseurs Stick, Plate et Ball. Si les valeurs propres sont rangées dans l'ordre décroissant : $\alpha_1 > \alpha_2 > \alpha_3$, on peut écrire :

$$\overline{T(\mathbf{n})} = (\alpha_1^2 - \alpha_2^2 - \alpha_3^2)(v_1 \otimes v_1) + (\alpha_2^2 - \alpha_3^2)(v_1 \otimes v_1 + v_2 \otimes v_2) + \alpha_3^2(v_1 \otimes v_1 + v_2 \otimes v_2 + v_3 \otimes v_3)$$

On obtient ainsi une combinaison d'un tenseur Stick (qui représente la normale la plus probable et la confiance par rapport aux autres possibilités), d'un Plate (qui représente le plan dans lequel la normale est la plus probable, et la confiance par rapport aux autres), et d'un Ball (qui ne représente pas grand chose, si ce n'est l'absence d'information sur l'orientation).

4.3.6 Tenseurs conditionnels

Nous rencontrerons par la suite des normales aléatoires dont la loi de probabilités est conditionnée par d'autres événements aléatoires. De telles normales seront donc données par des lois de probabilités conditionnelles de la forme $P(n/\dots)$ et nous noterons $T(n/\dots)$ le tenseur conditionnel associé. Nous allons ici voir comment sont reliés ces tenseurs.

Le premier cas est celui d'une normale aléatoire pouvant provenir de plusieurs processus aléatoires a_i différents, représentés par les lois $P(n/a = a_i) = P(n_i)$ de tenseurs $\overline{T(\mathbf{n}_i)}$.

La provenance est exprimée au travers d'une loi a-priori sur le processus aléatoire d'origine $P(a = a_i) = \pi_i$. On a alors par le théorème des probabilités totales :

$$P(n) = \sum_i P(n/a = a_i)P(a = a_i) = \sum P(n_i)\pi_i$$

le passage au tenseur étant linéaire par rapport à la loi de probabilité, on a immédiatement :

$$\overline{T(\mathbf{n})} = \sum \overline{T(\mathbf{n}/a = a_i)}P(a_i) = \sum \pi_i \overline{T(\mathbf{n}_i)} \quad (28)$$

on peut alors faire complètement abstraction des loi de probabilités et travailler directement sur les tenseurs.

Un deuxième cas est celui où la loi sur la normale dépend d'une deuxième normale, dont la loi $P(n_0)$ est partiellement connue au travers du tenseur $\overline{T(n_0)}$. La dépendance est exprimée par la loi de probabilité conditionnelle $P(n/n_0)$ et le tenseur conditionnel $\overline{T(n/n_0)}$. on a alors :

$$P(n) = \int P(n/n_0)P(n_0)dn_0$$

on peut alors écrire, par linéarité :

$$\overline{T(n)} = \int \overline{T(n/n_0)}P(n_0)dn_0 \quad (29)$$

on pourrait alors croire que cette égalité implique, pour $\overline{T(n/n_0)}$ donné, une relation linéaire entre les tenseurs $\overline{T(n)}$ et $\overline{T(n_0)}$ ⁴. Il n'en est rien, comme le montre l'exemple 2D de la figure 7, où deux normales aléatoires ayant le même tenseur $\overline{T(n_0)}$ définissent, via l'équation 29, des tenseurs $\overline{T(n)}$ très différents.

4.4 Utilisation de la distance tensorielle

Nous allons maintenant voir comment retrouver, à l'aide de notre tenseur, l'espérance et la covariance approchée au sens de la distance tensorielle. Cette distance est définie, pour deux normales non orientées \mathbf{n} et \mathbf{m} , par :

$$d_t(\mathbf{m}, \mathbf{n})^2 = \|\mathbf{n} - \mathbf{m}(\mathbf{n}|\mathbf{m})\|^2 = \|\Pi_{\mathbf{m}^\perp}(\mathbf{n})\|^2 = (1 - (\mathbf{n}|\mathbf{m})^2) = 1 - \mathbf{n} \otimes \overline{T(\mathbf{m})} \otimes \mathbf{n}$$

où $\|\cdot\|$ est la norme quadratique usuelle de \mathbb{R}^d et $(\cdot|\cdot)$ est son produit scalaire usuel, et où $\Pi_{\mathbf{m}^\perp}$ est la projection orthogonale le long de \mathbf{m} . Elle vérifie aussi :

$$d_t(\mathbf{m}, \mathbf{n}) = \sin((\widehat{\mathbf{m}}, \mathbf{n})) = d(\mathbf{m}, \mathbf{n}) + O(d(\mathbf{m}, \mathbf{n})^3)$$

On a donc :

$$\Psi_t(d) = \sin(d) \quad \psi_t(d) = \frac{\sin(d)}{d} \quad \psi_t(\overline{\mathbf{m}\mathbf{n}}) = \Pi_{\mathbf{m}^\perp}(\mathbf{n})$$

4. C'est ce que fait Médioni lorsque il parle de linéarité de la fonction de vote [11, page 46] : il exprime le tenseur $\overline{T(n_0)}$ sous sa décomposition Stick-Ball-Plate $\overline{T(n_0)} = (\lambda_1 - \lambda_2)S + (\lambda_2 - \lambda_3)P + \lambda_3B$ et en déduit le tenseur $\overline{T(n)}$ en combinant linéairement les tenseurs de vote Stick-Plate-Ball : $\overline{T(n)} = (\lambda_1 - \lambda_2)V_S + (\lambda_2 - \lambda_3)V_P + \lambda_3V_B$

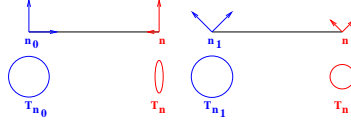


FIG. 7 – *Contre-exemple à la linéarité des votes de Médioni : prenons deux normales aléatoires \mathbf{n}_0 et \mathbf{n}_1 n'ayant que deux possibilités équiprobables, verticale et horizontale pour \mathbf{n}_0 , à 45 degrés pour \mathbf{n}_1 . Elles ont alors le même tenseur isotrope $\overline{T(\mathbf{n}_0)} = \overline{T(\mathbf{n}_1)} = Id/2$. Prenons maintenant un vote calqué sur le vote Stick de Médioni : la normale \mathbf{n}_0 verticale donne une normale \mathbf{n} verticale très probable (avec une faible covariance), la normale horizontale donne une normale horizontale très improbable (avec une forte covariance), les deux normales à 45 degrés donnent des normales symétriques équiprobables. On obtient alors pour \mathbf{n} un tenseur indiquant une normale très probablement verticale en considérant \mathbf{n}_0 , et isotrope en considérant \mathbf{n}_1 .*

4.4.1 Approximation tensorielle de l'espérance

En utilisant l'égalité $d_t(\mathbf{m}, \mathbf{n})^2 = 1 - \mathbf{n} \otimes \overline{T(\mathbf{m})} \otimes \mathbf{n} = 1 - \mathbf{m} \otimes \overline{T(\mathbf{n})} \otimes \mathbf{m}$, on peut écrire : $\sigma_{t, \mathbf{n}}^2(\mathbf{m}) = 1 - E_{\mathbf{n}}[\mathbf{m} \otimes \overline{T(\mathbf{n})} \otimes \mathbf{m}] = 1 - \mathbf{m} \otimes \overline{E_{\mathbf{n}}[T(\mathbf{n})]} \otimes \mathbf{m}$. On retrouve donc le tenseur d'une variable aléatoire tel que défini dans la section 4.3, et nous recherchons les normales \mathbf{m} qui maximisent : $\mathbf{m} \otimes \overline{T(\mathbf{n})} \otimes \mathbf{m}$. Pour les trouver, il faut diagonaliser $\overline{T(\mathbf{n})}$: $\overline{T(\mathbf{n})} = \sum_i (\alpha_i \mathbf{n}_i) \otimes (\alpha_i \mathbf{n}_i)$. On a alors $\mathbf{m} \otimes \overline{T(\mathbf{n})} \otimes \mathbf{m} = \sum_i \alpha_i (\mathbf{n}_i | \mathbf{m})^2 \leq \max_i (\alpha_i) \cdot \sum_i (\mathbf{n}_i | \mathbf{m})^2 = \max_i (\alpha_i)$, avec égalité si et seulement si \mathbf{m} appartient au sous-espace propre associé à la plus grande valeur propre. On a donc la proposition suivante :

Proposition 11 *Espérance de Fréchet au sens de la distance tensorielle*

L'approximation tensorielle de l'espérance au sens de Fréchet d'une normale non-orientée aléatoire est constituée de l'ensemble des vecteurs unitaires du sous-espace propre de $\overline{T(\mathbf{n})}$ associé à sa plus grande valeur propre.

Elle est de plus capable de traiter correctement le cas des distributions de normales du type plate et ball (section 4.3.5).

Enfin, on doit se rappeler que cette espérance n'est pas orientée. On peut l'orienter en utilisant le même sens que la moyenne vectorielle. Pour cela, il suffit de calculer l'espérance du produit scalaire entre la normale trouvée $\overline{\mathbf{n}}$ et la normale aléatoire. Si elle est négative, on inverse la normale $\overline{\mathbf{n}}$.

4.4.2 Approximation tensorielle de la covariance

La covariance par rapport à une primitive fixe \mathbf{m} peut s'écrire $\Sigma_{t, \mathbf{nn}}(\mathbf{m}) = E_{\mathbf{n}}[\psi_t(\overline{\mathbf{m}\mathbf{n}}) \cdot \psi_t(\overline{\mathbf{m}\mathbf{n}})^t] = E_{\mathbf{n}}[\Pi_{\mathbf{m}\perp}(\mathbf{n}) \cdot \Pi_{\mathbf{m}\perp}(\mathbf{n})^t]$, soit, en utilisant la matrice de la projection

$$\Sigma_{t, \mathbf{nn}}(\mathbf{m}) = \Pi_{\mathbf{m}\perp} \cdot E_{\mathbf{n}}[\mathbf{n} \cdot \mathbf{n}^t] \cdot \Pi_{\mathbf{m}\perp}^t = \Pi_{\mathbf{m}\perp} \cdot \overline{T(\mathbf{n})} \cdot \Pi_{\mathbf{m}\perp}^t$$

Le tenseur d'une normale aléatoire contient donc aussi une approximation de la covariance par rapport à n'importe quelle primitive fixe m .

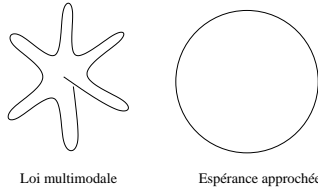
Proposition 12 *Covariance au sens de la distance tensorielle*

L'approximation tensorielle de la covariance d'une normale non-orientée aléatoire est obtenue en restreignant $\overline{T(\mathbf{n})}$ à l'hyperplan orthogonal à la moyenne.

4.4.3 Synthèse dans le cas général

Le tenseur utilisé par Lee et Médioni est permet donc une approximation extrêmement efficace de l'espérance et de la covariance d'une primitive aléatoire. Il suffit en effet de les diagonaliser, et la plus grande valeur propre et ses vecteurs propres unitaires associé indique l'espérance, et les autres valeurs propres et vecteurs propres indiquent la covariance.

Elle trouve cependant ses limites dans certains cas extrêmes de lois multimodales, comme dans l'exemple suivant :



4.4.4 Approximation tensorielle de la distance de Mahalanobis

Le tenseur d'une normale aléatoire contient une approximation de la covariance qui permet de calculer la distance de Mahalanobis. Nous allons donc ici voir comment exprimer directement et simplement une approximation de la distance de mahalanobis en fonction du tenseur.

La vraie distance de Mahalanobis entre une primitive n et une primitive aléatoire centrée en \mathbf{o} et de covariance $\Sigma_{\mathbf{o}\mathbf{o}}$ est donnée par :

$$\mu^2(\mathbf{o}, n) = \overline{\vec{n}} \otimes \Sigma_{\mathbf{o}\mathbf{o}}^{-1} \otimes \overline{\vec{n}}$$

où $\overline{\vec{n}}$ et $\Sigma_{\mathbf{o}\mathbf{o}}$ sont calculés dans la carte exponentielle. Il semble alors naturel de remplacer ces deux éléments par leur approximation tensorielle. On a alors :

$$\mu_t^2(\mathbf{o}, n) = \Psi_t(\overline{\vec{n}}) \otimes \Sigma_{t\mathbf{o}\mathbf{o}}^{-1} \otimes \Psi_t(\overline{\vec{n}})$$

Or $\Psi_t(\overline{\vec{n}})$ et $\Sigma_{t\mathbf{o}\mathbf{o}}^{-1}$ sont respectivement la projection de n et de $\overline{T(\mathbf{o})}^{-1}$ dans le plan perpendiculaire à \mathbf{o} . On peut alors se contenter de ne projeter qu'un seul de ces deux éléments pour obtenir le même résultat :

$$\mu_t^2(\mathbf{o}, n) = n \otimes \Sigma_{t\mathbf{o}\mathbf{o}}^{-1} \otimes \overline{n} = \Psi_t(\overline{\vec{n}}) \otimes \overline{T(\mathbf{o})}^{-1} \otimes \Psi_t(\overline{\vec{n}})$$

On utilisera en général la première formule, car on doit de toute façon diagonaliser le tenseur pour connaître la normale moyenne et savoir ainsi dans quelle direction effectuer la projection. Il est alors aisé de calculer la covariance en annulant la première valeur propre.

Le principal intérêt de cette approximation est qu'elle conserve sa principale propriété : calculons l'espérance de la distance de Mahalanobis approchée entre une primitive aléatoire et son espérance (lorsque celle ci est unique) :

$$\begin{aligned}
 E[\mu_t^2(\mathbf{o}, n)] &= \int_{\mathcal{M}} \Psi_t(\vec{o}\vec{n})^t \cdot \Sigma_{t\mathbf{o}\mathbf{o}}^{-1} \cdot \Psi_t(\vec{o}\vec{n}) \cdot p_{\mathbf{o}}(n) \cdot d\mathcal{M}(n) \\
 &= \int_{\mathcal{M}} \text{Tr}(\Sigma_{t\mathbf{o}\mathbf{o}}^{-1} \cdot \Psi_t(\vec{o}\vec{n}) \cdot \Psi_t(\vec{o}\vec{n})^t) \cdot p_{\mathbf{o}}(n) \cdot d\mathcal{M}(n) \\
 &= \text{Tr} \left(\Sigma_{t\mathbf{o}\mathbf{o}}^{-1} \cdot \int_{\mathcal{M}} \Psi_t(\vec{o}\vec{n}) \cdot \Psi_t(\vec{o}\vec{n})^t \cdot p_{\mathbf{o}}(n) \cdot d\mathcal{M}(n) \right) \\
 &= \text{Tr}(\Sigma_{t\mathbf{o}\mathbf{o}}^{-1} \cdot \Sigma_{t\mathbf{o}\mathbf{o}}) = \text{Tr}(Id) \\
 &= d
 \end{aligned}$$

l'espérance de la distance de Mahalanobis approchée entre une primitive aléatoire et son espérance est donc toujours égale à la dimension de l'espace.

Enfin, on peut utiliser directement cette distance de Mahalanobis pour calculer la densité de probabilité approchée d'une loi gaussienne :

$$p_{t\mathbf{o}}(n) = \exp\left(\frac{\mu_t^2(\mathbf{o}, n)}{2}\right)/k$$

4.5 Comparaison des statistiques exactes et approchées

Nous allons ici comparer les précisions des différentes méthodes d'estimation de la moyenne d'un ensemble de normales. Nous travaillerons avec les normales non-orientées.

Nous allons effectuer une série d'expériences sur le modèle suivant : on tire une première normale aléatoire uniforme, puis n normales suivant une loi gaussienne centrée autour de la première et d'écart-type σ . On estime l'espérance suivant les trois méthodes (l'espérance rigoureuse : min de la distance angulaire par descente de gradient, les deux espérances approchées : min de la dist. vectorielle par moyenne classique+normalisation, min de la dist. tensorielle par calcul du tenseur+extraction du premier vecteur propre). On examine alors l'écart entre le véritable centre de la distribution et les 3 espérances ainsi calculés. On mesure aussi l'écart entre les deux espérances approchées et l'espérance rigoureuse.

Pour chaque paramètres, on réalise ainsi plusieurs expériences (10000), et on calcule la RMS (figure 8) et le maximum de chacun des écarts (figure 9).

On constate alors les faits suivants :

- Jusqu'à 20 degrés, les 3 méthodes d'estimation ont des précisions comparables : les écarts absolus (entre l'espérance estimée et la vraie valeur centrale de notre loi) sont analogues. Ceci explique par le fait que les écarts relatifs (entre les différentes méthodes d'estimations) restent négligeables par rapport aux écarts absolus.
- Au delà, l'espérance vectorielle garde un écart absolu comparable à celui de l'espérance rigoureuse, même si l'écart relatif entre les deux estimés augmente. On constate même

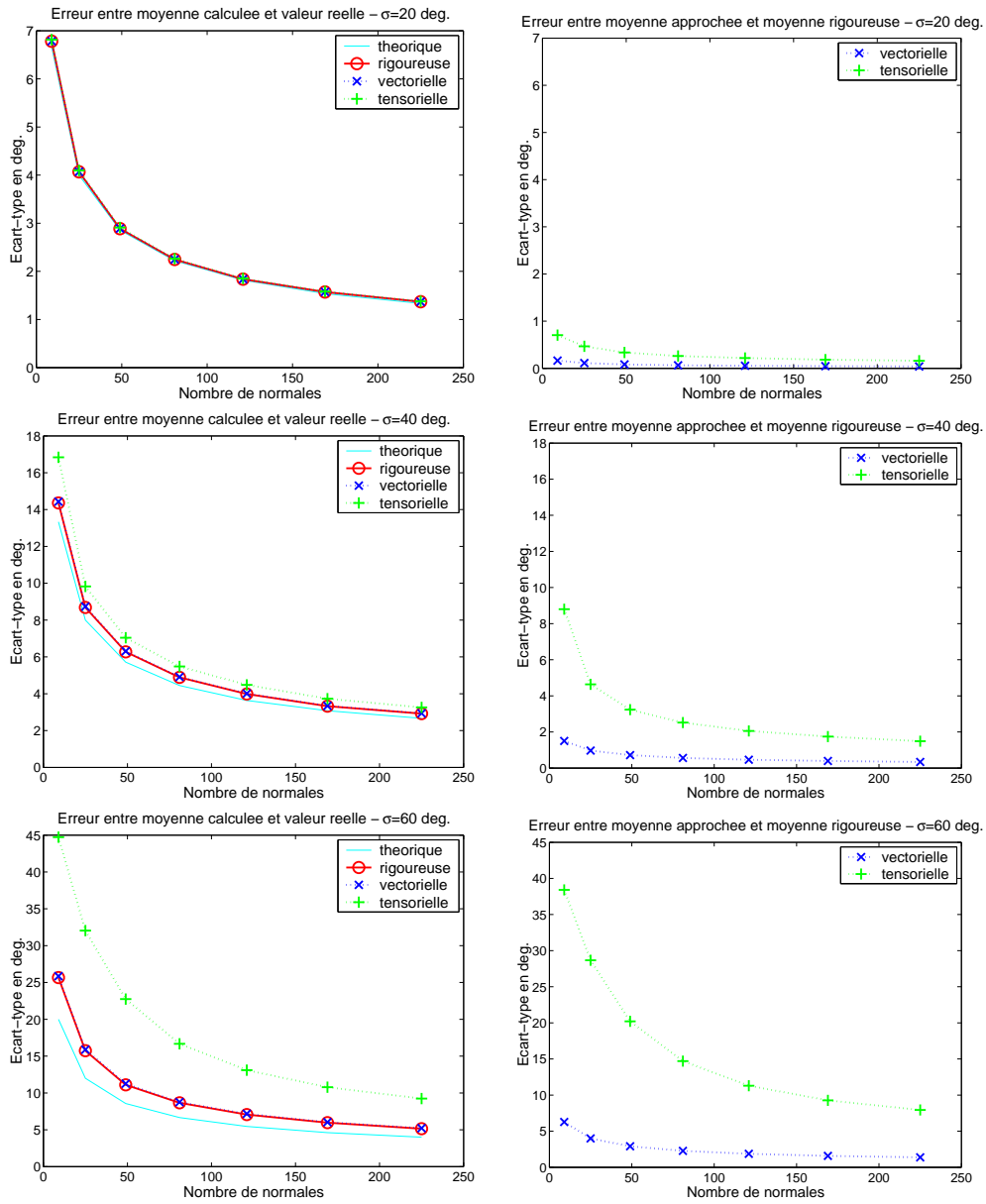


FIG. 8 – Écart-type de l'erreur sur l'estimation de la moyenne d'une normale aléatoire 3D par la méthode rigoureuse et les méthodes approchées. A gauche : erreur des méthodes par rapport à la véritable valeur moyenne de la normale aléatoire. A droite : erreur des méthodes approchées par rapport au résultat de la méthode rigoureuse.

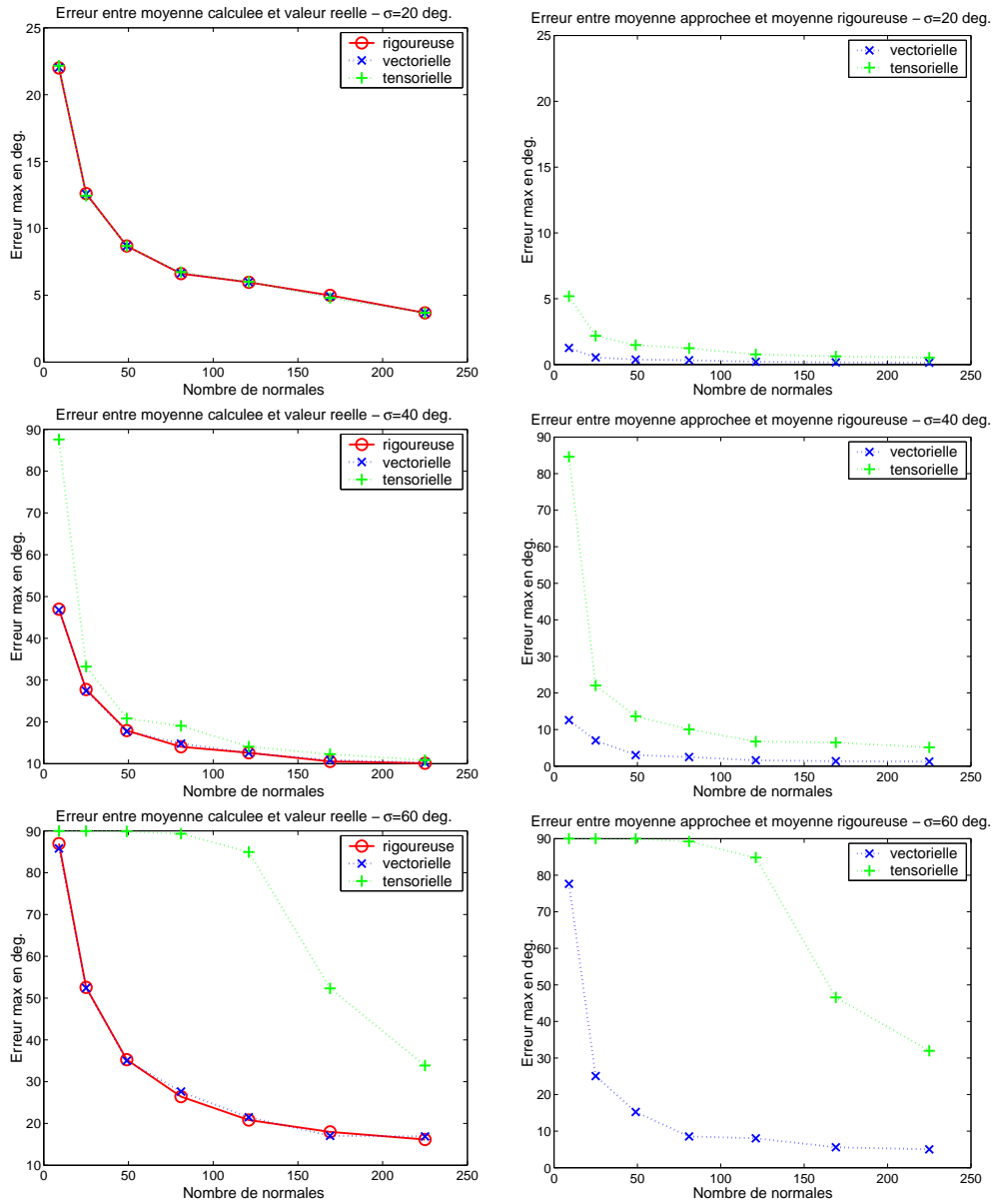


FIG. 9 – Erreur maximale sur l'estimation de la moyenne d'une normale 3D aléatoire par la méthode rigoureuse et les méthodes approchées.

dans certains cas avec forte variance que l'espérance vectorielle est légèrement plus précise. Cela provient probablement du fait que la distance vectorielle est une M-Estimée de la distance angulaire, et est par conséquent plus robuste [17].

- En revanche, l'espérance tensorielle devient inutilisable. A 40 degrés, et pour un faible nombre de normales, les écarts maximaux augmentent, même si les variances des écarts restent comparables. La méthode est donc encore bonne en moyenne, mais peu robuste. Au delà de 60 degré, les écarts deviennent en revanche trop importants.

On ne pourra donc utiliser les tenseurs que pour des lois assez bien définies (écart-type inférieur à 40 degrés). Au delà, il faudra revenir à des techniques plus précises : l'espérance vectorielle, et même si besoin les techniques rigoureuses.

4.6 Statistiques approchées sur les rotations aléatoires

Des méthodes de statistiques approchées sur les rotations ont déjà été développées et largement étudiées. L'objet de cette section n'est donc pas de les représenter en détail, mais simplement de les replacer dans notre cadre mathématique et de faire le parallèle avec les méthodes utilisées pour les normales.

On distingue 3 méthodes couramment utilisées pour calculer une rotation moyenne :

- Le calcul de la vraie moyenne, au sens de la distance intrinsèque, introduite par [14]. Cette méthode est itérative comme dans le cas des normales (voir section 3.2.3).
- Le calcul direct par la méthode de la SVD [1; 18].
- Le calcul direct par la méthode des quaternions unitaires [8].

Ces deux dernières méthodes ont été d'abord introduite dans le cadre du recalage rigide puis adaptées pour le calcul de la rotation moyenne. Les trois méthodes ont été comparées dans le cadre du recalage [3] et dans le cadre du calcul de la rotation moyenne [5]. Cette dernière comparaison permet de comprendre que les deux dernières méthodes suivent exactement la même démarche que la méthode vectorielle pour les normales (voir 4.2) :

- La méthode des quaternions unitaires calcule dans un premier temps la moyenne vectorielle des quaternions unitaires qui représentent les rotations. On obtient alors un quaternion \bar{q} qui n'est plus forcément unitaire. On retrouve alors la rotation moyenne en prenant le quaternion unitaire \bar{Q} le plus proche de \bar{q} au sens de la distance euclidienne entre quaternions (il suffit pour cela de normaliser \bar{q}).
- La méthode de la SVD calcule dans un premier temps la moyenne des matrices qui représente les rotations. On obtient alors une matrice \bar{M} qui n'est plus forcément orthogonale. On retrouve alors la rotation moyenne en prenant la matrice orthogonale \bar{R} la plus proche de \bar{M} au sens de la distance euclidienne entre matrices.

L'article montre aussi que les deux méthodes directes minimisent la variance au sens des distances suivantes :

$$d_{quat}^2 = 8(1 - \cos(\theta/2)) = (4 \sin(\theta/4))^2$$

$$d_{SVD}^2 = 2(1 - \cos(\theta)) = (2 \sin(\theta/2))^2$$

où θ est l'angle entre les deux rotations, qui est la distance intrinsèque.

Finalement, ces deux méthodes sont donc des méthodes de statistiques approchées pour le calcul des rotations, avec :

$$d_{quat} = \Psi_{quat}(d) = 4 \sin(d/4)$$

$$d_{SVD} = \Psi_{SVD}(d) = 2 \sin(d/2)$$

ces fonctions sont sensiblement les mêmes que celles que nous avons utilisé pour le calcul approché des normales.

Pour finir la parallèle, on peut comparer les résultats expérimentaux des méthodes sur les normales et des méthodes sur les rotations. L'article indique que les résultats commencent à significativement se détériorer pour un écart-type de 40 degrés, et ne va pas au delà. C'est aussi le cas pour nos méthodes.

5 Calcul des normales pour un nuage de points non structuré

Nous allons examiner une première application du tenseur d'une normale aléatoire. Il s'agit d'estimer les normales lorsqu'on dispose d'un nuage de points non structurés et non orientés, mais qui provient d'une surface en 3D (ou d'un contour en 2D ou d'une hypersurface en nD).

Pour cela, on utilise la remarque suivante : si deux points sont sensés appartenir à la même surface, la normale à cette surface a plus de chances d'être perpendiculaire au segment qui relie ces deux points. Aussi, pour chaque point de notre nuage, il suffit de synthétiser l'information donnée par les autres points du nuage (les votes) pour en déduire la normale la plus probable. Cette synthèse des votes est effectuée en utilisant les tenseurs précédemment définis.

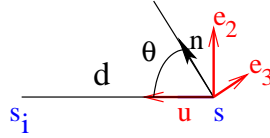
Les calculs seront indépendants pour chaque point du nuage. Nous allons donc détailler les différentes étapes pour un point de la scène donné, les interpréter, et évaluer la qualité du résultat.

5.1 Vote d'un point en un autre point

Fixons nous un point du nuage pour lequel nous allons rechercher la normale la plus probable. Nous le noterons s . Prenons maintenant un des autres points du nuage, noté s_i .

Nous allons essayer de modéliser la loi de probabilité de la normale n_s au point s , sachant que le point s_i appartient aussi à la surface. Cette loi est donc notée $p(n_s/s, s_i)$. Nous allons commencer par utiliser les symétries du problème : il est invariant par rotation (d'angle quelconque) autour du segment ss_i . Notre loi de probabilité ne dépend donc que de la

distance d entre les deux points et de l'angle θ entre la normale n_s et le segment ss_i - elle est de plus symétrique par rapport à θ : $p(n_s/s, s_i) = p(|\theta|, d)$.



On peut alors synthétiser cette loi dans un tenseur (rappelons que le point, et donc d est fixe) : on intègre sur l'ensemble des normales (i.e. sur θ). On travaille pour cela dans le repère orthogonal formé par le vecteur ss_i normalisé noté u et une base de l'espace perpendiculaire notée (e_2, e_3, \dots, e_D) . On obtient alors, par symétrie (les directions e_2, \dots, e_D sont équivalentes), un tenseur nécessairement de la forme suivante :

$$\overline{T(\mathbf{n}_s/s, s_i)} = \alpha^2(u \otimes u) + \sum_j \beta^2(e_j \otimes e_j) = (\alpha^2 - \beta^2)(u \otimes u) + \beta^2(u \otimes u + \sum_j e_j \otimes e_j)$$

avec

$$\alpha^2 = E(\cos^2 \theta) \quad \beta^2 = E(\sin^2 \theta)/(D - 1)$$

On a donc une combinaison d'un tenseur Ball et d'un tenseur Stick. Reste à déterminer α et β en fonction de d , ou plus simplement $(\alpha^2 - \beta^2)$ dans la mesure où le tenseur Ball n'apporte aucune information sur l'orientation.

Intuitivement, on voudrait favoriser les surfaces de forme plus lisses, donc les normales perpendiculaires au segment ss_i . On doit donc avoir une probabilité croissante et maximale pour $\theta = \pi/2$, et symétrique autour de ce mode, qui est donc aussi sa moyenne. En conséquence, $\alpha^2 < \beta^2$.

Ensuite, on peut penser que l'incertitude (i.e. l'écart-type) sur cette moyenne dépend de la distance entre les deux points : s'ils sont proches, la surface n'a pas le temps de se tordre ou se courber, l'écart-type est donc faible ($\alpha^2 \rightarrow 0$ et $\beta^2 \rightarrow 1/(D - 1)$). S'ils sont éloignés, la surface a le temps de se courber, et l'écart-type sur l'orientation de la normale est plus élevé ($\alpha^2 \rightarrow 1/D$ et $\beta^2 \rightarrow 1/D$).

Modélisation gaussienne Nous allons utiliser une modélisation gaussienne sans la justifier : disons qu'on a une loi gaussienne sur la normale n_s , centrée autour de la perpendiculaire au segment, et qui possède une variance proportionnelle à la distance entre les points⁵ (a^2 représente la variance angulaire sur la normale pour une distance d'un mm) :

$$p(\theta, d) = \exp\left(-\frac{(\pi/2 - |\theta|)^2}{2da^2}\right)/k$$

5. Ceci correspond à un modèle stochastique : on suppose que l'on peut suivre, sur la surface, des courbes à vitesse constante, et que la direction sur ces courbes suit un mouvement brownien. La variance sur la direction est alors proportionnelle à la distance au premier point de la courbe, ici s .

En 2D, on trouve alors (I représente le nombre imaginaire : $\sqrt{-1}$) :

$$\begin{bmatrix} \alpha^2 \\ \beta^2 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} - \\ + \end{bmatrix} \frac{1}{2} \exp(-2da^2) \frac{\operatorname{erf}\left(\frac{\sqrt{2}}{4a\sqrt{d}}(\pi + 4Ida^2)\right) + \operatorname{erf}\left(\frac{\sqrt{2}}{4a\sqrt{d}}(\pi - 4Ida^2)\right)}{2 \cdot \operatorname{erf}\left(\frac{\sqrt{2}}{4a\sqrt{d}}\pi\right)}$$

En 3D, on trouve alors :

$$\begin{bmatrix} \alpha^2 \\ \beta^2 \end{bmatrix} = \begin{bmatrix} \frac{1}{8} - \frac{1}{4} * \\ \frac{1}{8} + \frac{1}{8} * \end{bmatrix} \exp(-4da^2) \frac{\operatorname{erf}\left(\frac{\sqrt{2}}{4a\sqrt{d}}(\pi + 6Ida^2)\right) + \operatorname{erf}\left(\frac{\sqrt{2}}{4a\sqrt{d}}(\pi - 6Ida^2)\right)}{\operatorname{erf}\left(\frac{\sqrt{2}}{4a\sqrt{d}}(\pi + 2Ida^2)\right) + \operatorname{erf}\left(\frac{\sqrt{2}}{4a\sqrt{d}}(\pi - 2Ida^2)\right)}$$

5.2 Intégration des votes en un point

On veut maintenant intégrer les votes des différents points de la scène. On va leur donner une probabilité à priori $p(s_i) = \pi_i$, qui sera en général uniforme, mais pourra aussi permettre de prendre en compte une confiance a-priori, si bien que l'on peut écrire la probabilité empirique :

$$p(n_s/S) = \sum_i p(n_s/s, s_i) p(s_i) = \sum_i \pi_i p(n_s/s, s_i)$$

Le passage au tenseur étant linéaire, on trouve alors :

$$\overline{T(\mathbf{n}_s/S)} = \sum_i \pi_i \overline{T(\mathbf{n}_s/s, s_i)} = \sum_i \pi_i (\alpha_i^2 - \beta_i^2) (u_i \otimes u_i) + Cte * Id$$

5.3 Extraction de la normale

Le tenseur que l'on a calculé ci-dessus représente véritablement la loi de probabilité sur la normale en s . On extrait donc la normale la plus probable en calculant le vecteur propre associé à la plus grande valeur propre de ce tenseur. On comprend pourquoi on peut se contenter de calculer le tenseur à une composante Ball près (représentée ici par $Cte * Id$), et une constante multiplicative positive près : en effet, multiplier le tenseur par un scalaire ou lui ajouter un tenseur Ball ne fait que modifier les valeurs propres, sans modifier ni leur ordre ni leurs vecteurs propres associés.

5.4 Pertinence de la normale

Lorsqu'on le souhaite, on peut aussi obtenir une information sur la pertinence de la normale ainsi extraite. En effet le tenseur, lorsqu'il est entièrement calculé (i.e. lorsqu'on n'élimine pas la partie Ball et lorsqu'on veille à bien le normaliser) puis diagonalisé (on connaît alors ses valeurs propres dans l'ordre décroissant : $\lambda_1 > \dots > \lambda_D$), contient une

approximation de la covariance sur la normale, donnée par les valeurs propres secondaires $(\lambda_2, \dots, \lambda_D)$. On peut alors :

- étudier directement cette covariance, ou simplement la variance (la trace de la covariance ou plus simplement la somme des valeurs propres secondaires du tenseur : $\sigma^2 = \lambda_2 + \dots + \lambda_D$).
- étudier le complémentaire de la variance $1 - \sigma^2$, qui sera d’autant plus élevé que la variance et donc le doute sur la direction de la normale sera faible. Cet indice est en fait très simple à calculer : $1 - \sigma^2 = 1 - \lambda_2 - \dots - \lambda_D = \lambda_1$. Il s’agit donc de la plus grande valeur propre, qui indique la différence de pertinence entre la normale la plus probable et les autres normales possibles.
- étudier une variante de ce complémentaire : on peut par exemple penser à la différence entre les deux premières valeurs propres $\lambda_1 - \lambda_2$, qui correspond à l’amplitude de la composante stick du tenseur et indique la différence de pertinence entre la normale la plus probable et la “deuxième” plus probable.
- étudier ce que Médioni appelle la saillance qui est l’indice précédent avant normalisation du tenseur. Cela n’a pas vraiment de sens mathématique, mais permet de prendre en compte le nombre de votes reçus en un point, et de pouvoir accentuer par la suite les zones où beaucoup d’informations étaient disponible.

5.5 Implémentation pratique

Pour l’implémentation pratique, nous utiliserons, comme fonction de vote, une version simplifiée de la modélisation gaussienne.

Vote simplifié en 2D L’équation ci-dessus nous donne en 2D :

$$\alpha^2 - \beta^2 = -\exp(-2da^2) \frac{\operatorname{erf}\left(\frac{\sqrt{2}}{4a\sqrt{d}}(\pi + 4I da^2)\right) + \operatorname{erf}\left(\frac{\sqrt{2}}{4a\sqrt{d}}(\pi - 4I da^2)\right)}{2 \cdot \operatorname{erf}\left(\frac{\sqrt{2}}{4a\sqrt{d}}\pi\right)}$$

Nous obtenons, à un terme correctif près, une belle gaussienne. Or ce terme correctif est proche de 1 pour de faibles valeurs de da^2 , ce qui nous intéresse.

Nous utiliserons donc l’approximation suivante :

$$\alpha^2 - \beta^2 = -\exp(-2da^2)$$

Vote simplifié en 3D L’équation ci-dessus nous donne, en 3D :

$$\alpha^2 - \beta^2 = -\left(\frac{1}{8} + \frac{3}{8} \exp(-4da^2) \frac{\operatorname{erf}\left(\frac{\sqrt{2}}{4a\sqrt{d}}(\pi + 6I da^2)\right) + \operatorname{erf}\left(\frac{\sqrt{2}}{4a\sqrt{d}}(\pi - 6I da^2)\right)}{\operatorname{erf}\left(\frac{\sqrt{2}}{4a\sqrt{d}}(\pi + 2I da^2)\right) + \operatorname{erf}\left(\frac{\sqrt{2}}{4a\sqrt{d}}(\pi - 2I da^2)\right)}\right)$$

Comme dans le cas 2D, le terme correctif tend vers 1 en 0. Une approximation grossière est donc :

$$\alpha^2 - \beta^2 = -\frac{1}{2} \exp(-4da^2)$$

Algorithme & variante Pour des raisons d'efficacité algorithmique, nous n'utiliserons pas, pour un point de la scène donnée, tous les autres points, mais uniquement les points suffisamment proche, que nous rechercherons avec un kD-Tree[16]. Les points plus éloignés auront de toute façon des contributions très faibles, que l'on pourra négliger.

L'algorithme est donc basé sur deux paramètres, la variance de l'angle par unité de distance a^2 et la distance limite de recherche d_{max} . Son pseudo code est le suivant :

Boucle : pour chaque point s_i :

- Recherche des points proches à l'aide d'un kD-Tree : recherche des s_j tel que $j \neq i$ et $|s_i s_j| < d_{max}$.
- Calcul du tenseur : $T = \sum_j -\exp(-4a^2 |s_i s_j|) \cdot \frac{s_i s_j}{|s_i s_j|} \otimes \frac{s_i s_j}{|s_i s_j|}$.
- Recherche des valeurs et vecteurs propres de T
- Détermination de la normale : vecteur propre associé à la plus grande valeur propre

Notons que l'algorithme ici présenté concerne le calcul des normales aux hypersurfaces (courbes en 2D, surface en 3D). Il peut être aisément adapté pour le calcul des tangentes aux courbes (en dimension quelconque), en prenant le vecteur associé à la plus faible valeur propre.

5.6 Comparaison avec la technique Médioni

Lorsqu'il dispose de points non-orientés en entrée, Médioni utilise une première phase de vote pour estimer leurs orientations. Pour cela, il calcule pour chaque point de son nuage un tenseur défini comme la somme des votes des autres points du nuage. L'intégration des votes est donc identique à la nôtre.

Il ne garde ensuite que la composante stick du tenseur obtenu, et s'en sert comme entrée dans les phases suivantes de son moteur de reconstruction. Cette composante stick correspond au plus grand vecteur propre, si bien qu'elle représente implicitement la normale que nous avons extrait explicitement. Elle contient cependant une information supplémentaire, sa norme, qui correspond à la différence entre la première et la deuxième valeur propre. Nous avons déjà présenté cette valeur (section 5.4), appelée saillance, et expliqué quel pouvait être son intérêt pour des traitements ultérieurs.

La seule différence majeure entre notre technique et celle de Médioni est la fonction de vote élémentaire utilisée. Chez Médioni, celle-ci n'est pas donnée explicitement, mais obtenue par intégration numérique de la fonction de vote élémentaire d'un tenseur stick. Nous l'avons essayé dans nos expériences, et avons obtenu des résultats légèrement moins bon qu'avec notre fonction de vote. Ceci semble confirmer l'on peut faire varier la fonction de vote sans différence notables dans les résultats [10].

5.7 Comparaison avec d'autres techniques

Il existe de nombreuses autres méthodes permettant d'interpoler les normales pour un nuage de points non-structurés. Mais qu'elles fassent appel à la triangulation de Delaunay [12] ou à des approximations locales par des modèles de surface, ces méthodes sont d'essence géométrique et non probabiliste, et se révèlent très sensible au bruit, en particulier à la présence d'outliers.

5.8 Résultats

Nous présentons ici les résultats de l'algorithme sur des données synthétiques et sur des surfaces segmentées dans les scanners.

Dans toutes ces expériences, nous avons fait varier une distance caractéristique $CDist$. Le paramètre a^2 a été fixé à une variance de $1rad^2$ par distance caractéristique: $a^2 = 1/CDist$ et nous avons déterminé la distance de recherche maximale à partir d'une distance de Mahalanobis maximale μ_{max}^2 :

$$4a^2d < 2\cdot\mu_{max}^2 \implies d_{max} = \mu_{max}^2/2a^2 = \frac{\mu_{max}^2}{2}CDist$$

Nous avons alors étudié la distance angulaire non-orientée entre les normales trouvées par notre méthode et les vraies normales.

5.8.1 Données synthétiques

Nous avons ici utilisé des nuages de points échantillonnées sur une forme géométrique simple et connaissons donc parfaitement les normales.

La forme utilisée est un ellipsoïde dont les demi-axes font 100, 50 et 30 mm. L'échantillonnage est composé de 1000 points répartis au hasard (fig. 10 gauche), que nous avons détérioré en le bruitant de deux façons :

- En rajoutant un bruit gaussien isotrope sur la position des points (fig. 10 centre).
- En rajoutant des points aberrants, répartis uniformément dans la boîte englobante de l'ellipsoïde (fig. 10 droite).

Nous avons alors tracé la RMS de la distance entre les normales calculées et les normales de référence en fonction de la distance caractéristique (fig. 11). On voit sur les résultats que la principale force de la méthode est sa résistance aux points aberrants (les résultats sont encore significativement bons avec 8 fois plus de points aberrants que de points de l'ellipse) et au bruit de position (jusqu'à 2mm, mais un écart-type de 8mm sur une ellipse de 100mm, c'est énorme).

On peut maintenant se poser la question de savoir comment régler le paramètre de l'algorithme (a^2 ou $CDist$ suivant l'approche). Nous le ferons pour l'instant de manière purement expérimentale, en prenant le paramètre qui minimise l'erreur commise (25 dans les expériences précédentes).

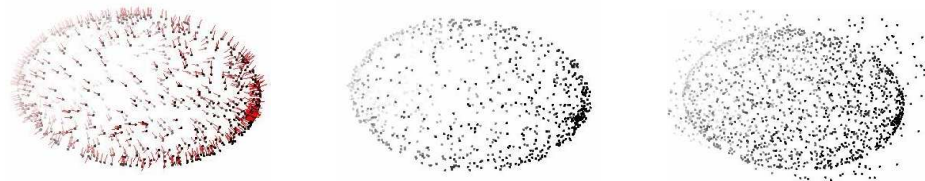


FIG. 10 – Échantillonnage irrégulier d'une ellipse de demi-axes 100,50 et 30 mm. Gauche : avec les normales de référence. Centre : avec un bruit de position (écart-type 1 mm). Droite : avec des points aberrants (1 point aberrant pour 1 point du nuage original)

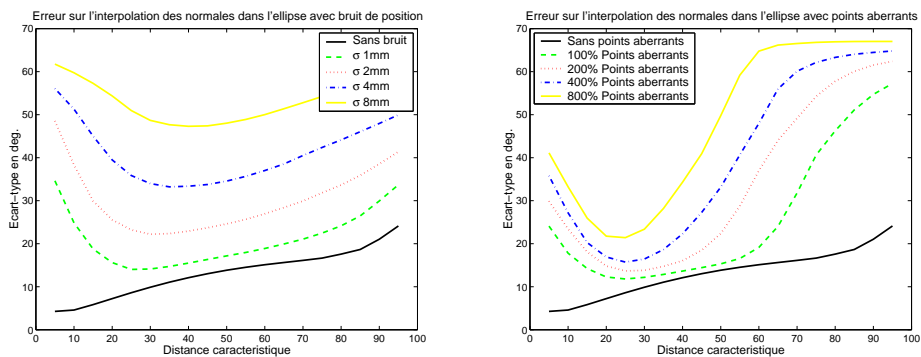


FIG. 11 – RMS de l'erreur commise sur l'estimation des normales pour l'ellipse, en fonction de la distance caractéristique. Gauche : avec un bruit de position (de 1 à 8 mm). Droite : avec des points aberrants (de 100% à 800%).

5.8.2 Données réelles

Nous avons ici testé le calcul des normales sur les surfaces des dents extraites à partir des scanners. On dispose en-effet dans ces surfaces des normales calculées à l'aide du gradient de l'intensité dans l'image scanner. Elles ne sont pas exactes, mais on a pu estimer leur précision : l'écart-type est de l'ordre de 10 degrés.

Les résultats sont présentés sur deux cas figure 12. Ils indiquent que les résultats sont assez précis.

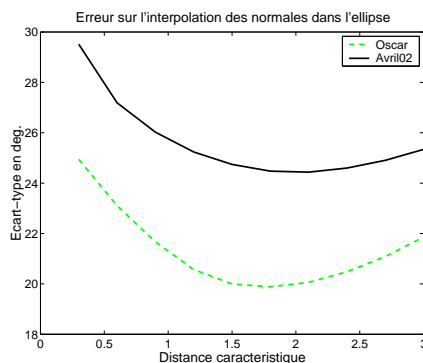


FIG. 12 – RMS de l'erreur commise sur l'estimation des normales pour la surface des dents extraites d'un scanner.

6 Conclusion

Nous avons ici montré commencé par présenter une théorie rigoureuse sur les primitives aléatoires et présenté en particulier différents algorithmes pour les normales aléatoires.

Nous avons ensuite vu comment définir, à partir de M-Estimateur, des méthodes approchées d'estimation de l'espérance et la covariance d'une primitive aléatoire. Nous avons présenté plusieurs exemples de ces méthodes pour les normales et les rotations aléatoires et avons montré leurs qualités (simplicité de calcul, robustesse et précision) et leurs limites (mauvais comportement pour de grandes covariances ou des lois multimodales, incapacité à gérer des lois à espérance multiple).

Nous nous sommes ensuite focalisé sur un cas particulier de ces techniques : le tenseur d'une normale aléatoire, et avons montré comment il permet de synthétiser de façon linéaire l'information sur l'espérance et la covariance d'une normale aléatoire non-orientée.

Ce tenseur a permis, moyennant quelques modifications mineures, de donner une explication théorique cohérente à la technique de tensor-voting introduite de façon heuristique par Médioni et al. pour interpoler les tangentes ou les normales dans un nuage de points non-orientés issu d'une courbe ou d'une surface.

Il reste maintenant à expliquer la suite des techniques de tensor-voting développées par Médioni pour l'estimation des courbures des surfaces et la reconstruction de surfaces. Une telle explication pourrait en-effet permettre de corriger ses algorithmes (en indiquant par exemple quand et comment utiliser des statistiques rigoureuses lorsque le tenseur s'avère inapproprié), ouvrir la voie à une théorie des surfaces probabilistes et permettre le développement de nouvelles applications basées sur le tensor-voting (pour la fusion et le recalage de surfaces notamment).

7 Acknowledgements

Ce travail a été financé par AREALL au travers d'une bourse CIFFRE.

Références

- [1] K. Arun, T. Huang, and S. Blostein. Least-squares fitting of two 3D point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5):698–700, 1987.
- [2] M.d. Carmo. *Riemannian Geometry*. Birkhäuser (Boston, Basel, Berlin), 1992.
- [3] D.W. Eggert, A. Lorusso, and R.B. Fisher. Estimating 3-D rigid body transformations : A comparison of four major algorithms. *Machine Vision and Applications*, 9:272–290, 1997.
- [4] J. Garrigues. Eléments d'algèbre et d'analyse tensorielle à l'usage des mécaniciens, 1999. <http://esm2.imt-mrs.fr/gar/tenseurs.html>.
- [5] Claus Gramkow. On averaging rotations. *International Journal of Computer Vision*, 42(1):7–16, 2001.
- [6] S. Granger. Notes sur les tenseurs euclidiens, 1999. <http://www-sop.inria.fr/epidaure/personnel/Sebastien.Granger/topos/tens/tens.html>.
- [7] S. Granger. *Recalage et reconstruction de surfaces : Une approche statistique multi-échelle. Application à l'implantologie dentaire assistée par ordinateur*. PhD thesis, Ecole des Mines de Paris - France, 2003.
- [8] B. Horn. Closed-form solution of absolute orientation using unit quaternions. *Optical Society of America*, 4(4):629–642, 1987.
- [9] W. Klingenberg. *Riemannian Geometry*. Walter de Gruyter (berlin, New-York), 1982.
- [10] M.S. Lee. Communication personnelle.
- [11] G. Médioni, M.S. Lee, and C.K. Tang. *A Computational Framework for Segmentation and Grouping*. 2000.
- [12] Amenta N. and Bern M. Surface reconstruction by voronoi filtering. *Discrete and Computational Geometry*, 22(04):481–504, 1999.
- [13] X. Pennec. Probabilities and statistics on riemanian manifolds: A geometric approach. unpublished.

-
- [14] X. Pennec. *L'incertitude dans les Problèmes de Reconnaissance et de Recalage - Applications en Imagerie Médicale et Biologie Moléculaire*. PhD thesis, Ecole Polytechnique - France, 1996.
 - [15] Jean Picard. Barycentres et martingales sur une variété. *Annales de l'Institut Henri Poincaré (Probabilités et Statistiques)*, 30(4):647–702, 1994.
 - [16] F. Preparata and M Shamos. *Computational geometry, an introduction*. Springer, 1986.
 - [17] P.J. Rousseeuw and A.M. Leroy. *Robust regression and outlier detection*. Ed. John Wiley and Sons, 1987.
 - [18] S. Umeyama. Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(4):376–380, 1991.



Unité de recherche INRIA Sophia Antipolis

2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)

Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)

Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)

Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Éditeur

INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)

<http://www.inria.fr>

ISSN 0249-6399