Quelques propriétés des courbes principales avec contrainte de longueur

Aurélie Fischer & Sylvain Delattre

Laboratoire de Probabilités et Modèles Aléatoires
Université Paris Diderot
75013 Paris
aurelie.fischer@univ-paris-diderot.fr
sylvain.delattre@univ-paris-diderot.fr

Résumé. Les courbes principales sont des courbes paramétrées passant au milieu d'une loi de probabilité dans $\mathbb{R}^d$. Outre la définition originelle basée sur la notion d'auto-consistance, plusieurs points de vue ont été considérés, dont un problème de minimisation de type moindres carrés avec contrainte. Nous étudions les propriétés théoriques de courbes principales de longueur au plus L et montrons notamment qu'elles sont toujours de courbure finie.

A partir de la condition d'ordre 1, exprimant qu'une courbe est un point critique pour le critère, nous obtenons une équation faisant intervenir la courbe, sa courbure, ainsi qu'une variable aléatoire jouant le rôle du paramètre de la courbe paramétrée. Cette équation permet de proposer une nouvelle démonstration de l'injectivité d'une courbe principale contrainte en dimension 2.

Mots-clés. Courbe principale, contrainte de longueur, courbure, critère de type moindres carrés, auto-consistance.

Abstract. Principal curves are defined as parametric curves passing through the middle of a probability distribution in $\mathbb{R}^d$. In addition to the original definition based on self-consistency, several points of view have been considered, among which a least square type constrained minimization problem. We study theoretical properties satisfied by principal curves with length at most L and show in particular that they always have finite curvature.

We derive from the order 1 condition, expressing that a curve is a critical point for the criterion, an equation involving the curve, its curvature, as well as a random variable playing the role of the curve parameter. This equation allows to propose a new proof of the fact that a constrained principal curve in dimension 2 has no multiple points.

Keywords. Principal curve, length constraint, curvature, least-square-type criterion, self-consistency.

1 Introduction

Les courbes principales sont des courbes paramétrées passant au milieu d'une loi de probabilité dans $\mathbb{R}^d$, $d \geq 1$. Elles fournissent en quelque sorte un résumé de dimension 1 de cette loi. La définition originale, basée sur la propriété d'auto-consistance, a été introduite par Hastie et Stuetzle (1989). Une courbe paramétrée f est dite auto-consistante pour un vecteur aléatoire X ayant moment d'ordre 2 si elle vérifie, pour presque tout t,

$$f(t) = \mathbb{E}[X|t_f(X) = t],$$

où l'indice de projection t_f est donné par

$$t_f(x) = \max \arg\min_t \|x - f(t)\|^2.$$

La définition de courbe principale suppose en fait quelques hypothèses de régularité supplémentaires : la courbe est C^∞, elle n'a pas de point double et est de longueur finie dans toute boule de $\mathbb{R}^d$. Observons que la propriété d'auto-consistance prend une forme implicite puisque la courbe f dépend de l'indice de projection t_f, qui lui dépend de f.

D'autres points de vue, plus ou moins proches de la définition originale, ainsi que plusieurs algorithmes, ont été proposés ensuite dans la littérature (Tibshirani (1992),Kégl et al. (2000),Verbeek et al. (2001), Delicado (2001),Sandilya et Kulkarni (2002), Einbeck et al. (2005a),Ozertem et Erdogmus (2011),Gerber et Whitaker (2013)). Remarquons aussi que, dans leur version empirique, lorsque le vecteur aléatoire est remplacé par un nuage de points, les courbes principales ont de nombreuses applications, dans des domaines variés (voir par exemple Hastie et Stuetzle (1989), Friedsam et Oren (1989) pour des applications en physique, Kégl et Krzyżak (2002), Reinhard et Niranjan (1999) en reconnaissance de caractères et reconnaissance de la parole, Brunsdon (2007), Stanford et Raftery (2000), Banfield et Raftery (1992), Einbeck et al. (2005a,b) en cartographie et géologie, De'ath (1999), Corkeron et al. (2004), Einbeck et al. (2005a) en sciences naturelles, Caffo et al. (2008) en pharmacologie, et Wong et Chung (2008), Drier et al. (2013) en médecine, dans l'étude de maladies cardio-vasculaires ou de cancers).

Nous nous intéressons à la définition introduite par Kégl et al. (2000), qui considèrent des courbes principales contraintes. Plus précisément, les courbes principales sont obtenues dans ce cas comme les solutions d'un problème de minimisation des moindres carrés avec contrainte de longueur. Une motivation pour l'introduction de cette définition, plus facile à manipuler, est que l'existence de courbes principales selon la définition de Hastie et Stuetzle (1989) n'avait pu être démontrée que pour des lois très particulières (voir Duchamp et Stuetzle (1996a) et Duchamp et Stuetzle (1996b) pour quelques résultats, en dimension 2). Plus formellement, Kégl et al. (2000) proposent de minimiser la quantité $\mathbb{E}\left[\min_t \|X - f(t)\|^2\right]$ sur toutes les courbes dont la longueur ne dépasse pas une certaine valeur prédéfinie et montrent qu'il existe un minimiseur de ce critère dès que X possède un moment d'ordre 2. Contrairement à la définition originale, les courbes ne sont pas

supposées différentiables, ce qui permet en particulier de considérer des lignes polygonales. Ces dernières jouent un rôle important dans Kégl et al. (2000), en particulier d'un point de vue algorithmique.

Notons qu'une telle contrainte est tout à fait pertinente dans le cas empirique : en pratique, certains paramètres reflétant la complexité de la courbe doivent être calibrés soigneusement afin de réaliser un compromis entre une courbe passant par tous les points des données et une courbe trop grossière. Cette question de sélection de modèle a été étudiée par exemple dans Biau et Fischer (2012), Fischer (2013) et Gerber et Whitaker (2013).

Notons que notre cadre correspond au problème dit de distance moyenne dans une partie de la communauté mathématique (voir Lu et Slepčev (2016) par exemple).

2 Definitions and notations

Pour tout $d \geq 1$, on munit $\mathbb{R}^d$ de la norme euclidienne standard, notée $\|\cdot\|$. Le produit scalaire entre deux vecteurs u and v est noté $\langle u, v \rangle$.

Pour tout $x \in \mathbb{R}^d$, soit x^j sa j^e composante, $j = 1, \ldots, d$, de sorte que $x = (x^1, \ldots, x^d)$. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et X un vecteur aléatoire sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans $\mathbb{R}^d$, tel que $\mathbb{E}[\|X\|^2] < +\infty$.

Nous considérons des courbes paramétrées, c'est-à-dire des fonctions

$$\begin{aligned} f : [0,1] &\to \mathbb{R}^d \\ t &\mapsto (f^1(t), \ldots, f^d(t)), \end{aligned}$$

chaque coordonnée $t \mapsto f^j(t)$ étant continue. Pour une telle courbe $f : [0,1] \to \mathbb{R}^d$, soit $\mathscr{L}(f)$ sa longueur, définie par

$$\mathscr{L}(f) = \sup \sum_{i=1}^{n} \|f(t_i) - f(t_{i-1})\|,$$

où la borne supérieure est prise sur toutes les subdivisions $0 = t_0 \leq \cdots \leq t_n = 1$, $n \in \mathbb{N}^*$ (voir par exemple Alexandrov et Reshetnyak (1989)).

L'image d'une courbe f est notée $f([0,1])$. Soit

$$\Delta(f) = \mathbb{E}\left[\min_{t \in [0,1]} \|X - f(t)\|^2\right],$$

et, pour $L \geq 0$,

$$G(L) = \min\{\Delta(f), f : [0,1] \to \mathbb{R}^d, \mathscr{L}(f) \leq L\}.$$

Remarquons que $G(L)$, pour $L \geq 0$, est le minimum de la quantité

$$\mathbb{E}[\|X - \widehat{X}\|^2]$$

sur tous les vecteurs aléatoires possibles $\widehat{X}$ à valeurs dans l'image $f([0,1])$ d'une courbe $f : [0,1] \to \mathbb{R}^d$ de longueur $\mathscr{L}(f) \leq L$.

Notons qu'une courbe $f : [0,1] \to \mathbb{R}^d$ de longueur $\mathscr{L}(f) \leq L$ peut être paramétrée de sorte que la fonction f soit lipschitzienne avec constante L : c'est cette paramétrisation que l'on considère dans ce travail.

3 Résultat principal

Notre résultat principal s'énonce comme suit.

Théorème 3.1. *Soit $L > 0$ telle que $G(L) > 0$ et soit $f : [0,1] \to \mathbb{R}^d$ telle que $\mathscr{L}(f) \leq L$ et $\Delta(f) = G(L)$. Supposons que f soit paramétrée de façon à être L-Lipschitz. Alors, $\mathscr{L}(f) = L$,*

— *f est dérivable à droite sur $[0,1)$, $\|f'_r(t)\| = L \ \forall t \in [0,1)$,*

— *f est dérivable à gauche sur $(0,1]$, $\|f'_\ell(t)\| = L \ \forall t \in (0,1]$,*

et il existe une unique mesure signée f'' on $[0,1]$ (à valeurs dans $\mathbb{R}^d$) telle que

— *$f'_r(t) = f''([0,t]) \ \forall t \in [0,1)$,*

— *$f'_\ell(t) = f''([0,t)) \ \forall t \in (0,1]$,*

— *$f''([0,1]) = 0$.*

De plus, il existe un unique $\lambda > 0$ et, quitte à considérer une extension de l'espace $(\Omega, \mathcal{F}, \mathbb{P})$, il existe une variable aléatoire $\widehat{t}$ à valeurs dans $[0,1]$ telle que

— *$\|X - f(\widehat{t})\| = \min_{t \in [0,1]} \|X - f(t)\|$ p.s.,*

— *pour toute fonction borélienne bornée $g : [0,1] \to \mathbb{R}^d$,*

$$\mathbb{E}\left[\langle X - f(\widehat{t}), g(\widehat{t})\rangle\right] = -\lambda \int_{[0,1]} \langle g(t), f''(dt)\rangle. \tag{1}$$

Pour démontrer le théorème, on utilise notamment le fait que la propriété $f(t) = \mathbb{E}[X \mid t_f(X) = t]$ n'est pas vérifiée presque sûrement dans notre contexte : dès lors que la contrainte est active, une courbe principale de longueur bornée ne peut pas être auto-consistante !

La formule 1 peut ensuite être employée pour montrer que si la contrainte est active, une courbe principale de longueur bornée en dimension 2 est toujours injective.

Bibliographie

Hastie, T. et Stuetzle, W. (1989). Principal curves. *Journal of the American Statistical Association*, 84, 502–516.

Tibshirani, R. (1992). Principal curves revisited. *Statistics and Computing*, 2, 183–190.

Kégl, B., Krzyżak, A., Linder, T., et Zeger, K. (2000). Learning and design of principal curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 281–297.

Verbeek, J. J., Vlassis, N., et Kröse, B. (2001). A soft k-segments algorithm for principal curves. In *Proceedings of International Conference on Artificial Neural Networks 2001*, pages 450–456.

Delicado, P. (2001). Another look at principal curves and surfaces. *Journal of Multivariate Analysis*, 77, 84–116.

Sandilya, S. et Kulkarni, S. R. (2002). Principal curves with bounded turn. *IEEE Transactions on Information Theory*, 48, 2789–2793.

Einbeck, J., Tutz, G., et Evers, L. (2005a). Local principal curves. *Statistics and Computing*, 15, 301–313.

Ozertem, U. et Erdogmus, D. (2011). Locally defined principal curves and surfaces. *Journal of Machine Learning Research*, 12, 1249–1286.

Gerber, S. et Whitaker, R. (2013). Regularization-free principal curve estimation. *Journal of Machine Learning Research*, 14, 1285–1302.

Friedsam, H. et Oren, W. A. (1989). The application of the principal curve analysis technique to smooth beamlines. In *Proceedings of the 1st International Workshop on Accelerator Alignment*.

Kégl, B. et Krzyżak, A. (2002). Piecewise linear skeletonization using principal curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24, 59–74.

Reinhard, K. et Niranjan, M. (1999). Parametric subspace modeling of speech transitions. *Speech Communication*, 27, 19–42.

Brunsdon, C. (2007). Path estimation from GPS tracks. In *Proceedings of the 9th International Conference on GeoComputation, National Centre for Geocomputation, National University of Ireland, Maynooth, Eire*.

Stanford, D. C. et Raftery, A. E. (2000). Finding curvilinear features in spatial point patterns : principal curve clustering with noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22, 2237–2253.

Banfield, J. D. et Raftery, A. E. (1992). Ice floe identification in satellite images using mathematical morphology and clustering about principal curves. *Journal of the American Statistical Association*, 87, 7–16.

Einbeck, J., Tutz, G., et Evers, L. (2005b). Exploring multivariate data structures with local principal curves. In C. Weihs et W. Gaul, éditeurs, *Classification – The Ubiquitous Challenge, Proceedings of the 28th Annual Conference of the Gesellschaft für Klassifikation, University of Dortmund*, Studies in Classification, Data Analysis, and Knowledge Organization, pages 256–263. Springer, Berlin, Heidelberg.

De'ath, G. (1999). Principal curves : a new technique for indirect and direct gradient analysis. *Ecology*, 80, 2237–2253.

Corkeron, P. J., Anthony, P., et Martin, R. (2004). Ranging and diving behaviour of two 'offshore' bottlenose dolphins, Tursiops sp., off eastern Australia. *Journal of the Marine Biological Association of the United Kingdom*, 84, 465–468.

Caffo, B. S., Crainiceanu, C. M., Deng, L., et Hendrix, C. W. (2008). A case study in pharmacologic colon imaging using principal curves in single photon emission computed tomography. *Journal of the American Statistical Association*, 103, 1470–1480.

Wong, W. C. K. et Chung, A. C. S. (2008). Principal curves to extract vessels in 3D angiograms. In *Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'08)*, pages 1–8.

Drier, Y., Sheffer, M., et Domany, E. (2013). Pathway-based personalized analysis of cancer. *PNAS*, 110, 6388–6393.

Duchamp, T. et Stuetzle, W. (1996a). Extremal properties of principal curves in the plane. *The Annals of Statistics*, 24, 1511–1520.

Duchamp, T. et Stuetzle, W. (1996b). Geometric properties of principal curves in the plane. In H. Rieder, éditeur, *Robust Statistics, Data Analysis, and Computer Intensive Methods : in Honor of Peter Huber's 60th Birthday*, volume 109 de *Lecture Notes in Statistics*, pages 135–152. Springer-Verlag, New York.

Biau, G. et Fischer, A. (2012). Parameter selection for principal curves. *IEEE Transactions on Information Theory*, 58, 1924–1939.

Fischer, A. (2013). Selecting the length of a principal curve within a Gaussian model. *Electronic Journal of Statistics*, 7, 342–363.

Lu, X. Y. et Slepčev, D. (2016). Average-distance problem for parameterized curves. *ESAIM : Control, Optimisation and Calculus of Variations*, 22, 404–416.

Alexandrov, A. D. et Reshetnyak, Y. G. (1989). *General Theory of Irregular Curves*. Mathematics and its Applications. Kluwer Academic Publishers, Dordrecht.