Stochastic Development Regression Using Method of Moments

Line Kühnel(✉) and Stefan Sommer

Department of Computer Science, University of Copenhagen, Copenhagen, Denmark
{kuhnel,sommer}@di.ku.dk

Abstract. This paper considers the estimation problem arising when inferring parameters in the stochastic development regression model for manifold valued non-linear data. Stochastic development regression captures the relation between manifold-valued response and Euclidean covariate variables using the stochastic development construction. It is thereby able to incorporate several covariate variables and random effects. The model is intrinsically defined using the connection of the manifold, and the use of stochastic development avoids linearizing the geometry. We propose to infer parameters using the Method of Moments procedure that matches known constraints on moments of the observations conditional on the latent variables. The performance of the model is investigated in a simulation example using data on finite dimensional landmark manifolds.

Keywords: Frame bundle · Non-linear statistics · Regression · Statistics on manifolds · Stochastic development

1 Introduction

There is a growing interest for statistical analysis of non-linear data such as shape data arising in medical imaging and computational anatomy. Non-linear data spaces lack vector space structure, and traditional Euclidean statistical theory is therefore not sufficient to analyze non-linear data. This paper considers parameter inference for the stochastic development regression (SDR) model introduced in [10] that generalizes Euclidean regression models to non-linear spaces. The focus of this paper is to introduce an alternative estimation procedure which is simple and computationally tractable.

Stochastic development regression is used to model the relation between a manifold-valued response and Euclidean covariate variables. Similar to Brownian motions on a manifold, $\mathcal{M}$, defined as the transport of a Euclidean Brownian motion from $\mathbb{R}^n$ to $\mathcal{M}$, the SDR model is defined as the transport of a Euclidean regression model. A Euclidean regression model can be regarded as a time dependent model in which, potentially, several observations have been

F. Nielsen and F. Barbaresco (Eds.): GSI 2017, LNCS 10589, pp. 3–11, 2017.
https://doi.org/10.1007/978-3-319-68445-1_1

observed over time. Given a response variable $y_t \in \mathbb{R}^d$ and covariate vector $\boldsymbol{x}_t = (x_t^1, \dots, x_t^m) \in \mathbb{R}^m$, the Euclidean regression model can be written as

$$y_t = \alpha_t + \beta_t \boldsymbol{x}_t + \varepsilon_t, \quad t \in [0,1], \tag{1}$$

where $\alpha_t \in \mathbb{R}^d$ and $\beta_t \in \mathbb{R}^{d \times m}$. A regression model can hence be defined as a stochastic process with drift α_t, covariate dependency through $\beta_t \boldsymbol{x}_t$, and a brownian noise ε_t. The SDR model is then defined as the transport of a regression model of the form (1), from $\mathbb{R}^d$ to the manifold $\mathcal{M}$. The transportation is performed by stochastic development described in Sect. 2. Figure 1 visualizes the idea behind the model.

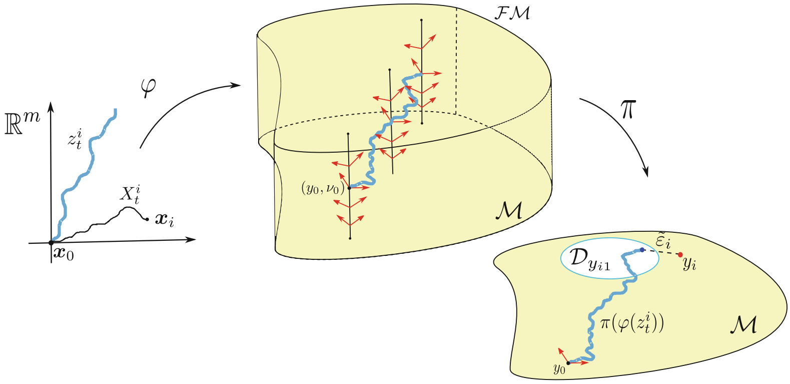

Fig. 1. The idea behind the model. Normal linear regression process z_t^i defined in (1) is transported to the manifold through stochastic development, φ. Here $\mathcal{FM}$ is the frame bundle, π a projection map, and $\mathcal{D}_{y_{i1}}$ the transition distribution of $y_{it} = \pi(\varphi(z_t^i))$. The tangent bundle of $\mathcal{FM}$ can be split in a horizontal and vertical subspace. Changes on $\mathcal{FM}$ in the vertical direction corresponds to fixing a point $y \in \mathcal{M}$ while changing the frame, ν, of the tangent space, $T_y\mathcal{M}$. Changes in the horizontal direction is fixing the frame for the tangent space and changing the point on the manifold. The frame is in this case parallel transported to the new tangent space.

In [10], Laplace approximation was applied for estimation of the parameter vector. However, this method was computational expensive and it was difficult to obtain results for detailed shapes. Alternatively, a Monte Carlo Expectation Maximization (MCEM) method has been considered, but, with this method, high probability samples were hard to obtain, which led to an unstable objective function. As a consequence, this paper examines the Method of Moments (MM) procedure for parameter estimation. The MM procedure is easy to apply and not as computationally expensive as the Laplace approximation. It is a well-known method for estimation in Euclidean statistics (see for example [3,6,14]), where it has been proven in general to provide consistent parameter estimates.

Several versions of the generalized regression model have been proposed in the case of manifold-valued response and Euclidean covariate variables. Local regression is considered in [11,19]. The former defines an intrinsic local regression model, while [11] constructs an extrinsic model. For global regression models, [5,12,16] consider geodesic regression, which is a generalization of the Euclidean linear regression model. There have been several approaches for defining non-geodesic regression models on manifolds. An example is kernel based regression models, in which the model function is estimated by a kernel representation [1,4,13]. In [7,8,17], the non-geodesic relation is modelled by a polynomial or piecewise cubic spline function. Moreover, [2,15] propose estimation of a parametric link function by minimization of the total residual sum of squares and the generalized method of moments procedure respectively.

The paper will be structured as follows. Section 2 gives a brief description of stochastic development and the frame bundle $\mathcal{FM}$. Section 3 introduces the SDR model and Sect. 4 describes the estimation procedure, Method of Moments. At the end, a simulation example is performed in Sect. 5.

2 Stochastic Development

This section gives a brief introduction to frame bundle and stochastic development. For a more detailed description and a reference for the following see [9]. Consider a d-dimensional Riemannian manifold $(\mathcal{M}, g)$ and a probability space $(\Omega, \mathcal{F}, P)$. Stochastic development is a method for transportation of stochastic processes in $\mathbb{R}^d$ to stochastic processes on $\mathcal{M}$. Let $z_t \colon \Omega \to \mathbb{R}^d$ denote a stochastic process for $t \in [0,1]$. In order to define the stochastic development of z_t it is necessary to consider a connection on $\mathcal{M}$. A connection, ∇, defines transportation of vectors along curves on the manifold, such that tangent vectors in different tangent spaces can be compared. A frequently used connection, which will also be used in this paper, is the Levi-Civita connection of a Riemannian metric. Consider a point $q \in \mathcal{M}$ and let ∂_i for $i = 1, \dots, d$ denote a coordinate frame at q, i.e. an ordered basis for $T_q\mathcal{M}$, with dual frame dx^i. A connection ∇ is locally determined by the Christoffel symbols defined by $\nabla_{\partial_i}\partial_j = \Gamma^k_{ij}\partial_k$. The Christoffel symbols for the Levi-Civita connection are given by $\Gamma^k_{ij} = \frac{1}{2}g^{kl}\left(\partial_i g_{jl} + \partial_j g_{il} - \partial_l g_{ij}\right)$, where g_{ij} denotes the coefficients of the metric g in the dual frame dx^i, i.e. $g = g_{ij}dx^i dx^j$, and g^{ij} are the inverse coefficients.

Stochastic development uses the frame bundle, $\mathcal{FM}$, defined as the fiber bundle of tuples (y, ν), $y \in \mathcal{M}$ with $\nu \colon \mathbb{R}^d \to T_y\mathcal{M}$ being a frame for the tangent space $T_y\mathcal{M}$. Given a connection on $\mathcal{FM}$, the tangent bundle of the frame bundle, $T\mathcal{FM}$, can be split into a horizontal, $H\mathcal{FM}$, and vertical, $V\mathcal{FM}$, subspace, i.e. $T\mathcal{FM} = H\mathcal{FM} \oplus V\mathcal{FM}$. Figure 1 shows a visualization of the frame bundle and the horizontal and vertical tangent spaces. The horizontal subspace determines changes in $y \in \mathcal{M}$ while fixing the frame ν, while $V\mathcal{FM}$ fixes $y \in \mathcal{M}$ and describes the change in the frame for $T_y\mathcal{M}$. Given the split of the tangent bundle $T\mathcal{FM}$, an isomorphism $\pi_{\star,(y,\nu)} \colon H_{(y,\nu)}\mathcal{FM} \to T_y\mathcal{M}$ can be defined. The inverse map $\pi^{\star}_{(y,\nu)}$ is called the horizontal lift and pulls a tangent vector in $T_y\mathcal{M}$ to $H_{(y,\nu)}\mathcal{FM}$. The horizontal lift of $v \in T_y\mathcal{M}$ is here denoted $v^{\star} \in H_{(y,\nu)}\mathcal{FM}$.

Let $e_1, \dots, e_d$ be the canonical basis of $\mathbb{R}^d$ and consider a point $(y, \nu) \in \mathcal{FM}$. Define the horizontal vector fields, $H_1, \dots, H_d$, by $H_i(\nu) = (\nu e_i)^\star$. The vector fields $H_1, \dots, H_d$ then form a basis for the subspace $H\mathcal{FM}$. Given this basis for $H\mathcal{FM}$, the stochastic development of a Euclidean stochastic process, z_t, to the frame bundle $\mathcal{FM}$ can be found by the solution to the Stratonovich differential equation $dU_t = H_i(U_t) \circ dz_t^i$, where Einsteins summation notation is used and $\circ$ specifies that it is a Stratonovich differential equation. The stochastic development of a process $z_t \in \mathbb{R}^d$ with reference point (y, ν) will be denoted $\varphi_{(y,\nu)}(z_t)$. A stochastic process on $\mathcal{M}$ can then be obtained by the projection of U_t to $\mathcal{M}$ by the projection map $\pi \colon \mathcal{FM} \to \mathcal{M}$.

3 Model

Consider a d-dimensional manifold $\mathcal{M}$ equipped with a connection ∇ and let $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n$ be n realizations of the response $y \in \mathcal{M}$. Notice that the realizations are assumed to be measured with additive noise, which might pull the observations to an ambient space of $\mathcal{M}$. An example of such additive noise for landmark data is given in Sect. 5. Denote for each observation $i = 1, \dots, n$, $\boldsymbol{x}_i = (x_{i1}, \dots, x_{im}) \in \mathbb{R}^m$ the covariate vector of $m \leq d$ covariate variables. The SDR model is defined as a stochastic process on $\mathcal{M}$ based on the definition of Euclidean regression models regarded as stochastic processes (see (1)). Assume therefore that the response $y \in \mathcal{M}$ is the endpoint of a stochastic process y_t in $\mathcal{M}$ and the covariates, $\boldsymbol{x}_i$, the endpoint of a stochastic process $X_t = (X_{1t}, \dots, X_{mt})$ in $\mathbb{R}^m$. The process X_{jt} is for random covariate variables assumed to be a Brownian motion in $\mathbb{R}$, while for fixed covariate effects it is modelled as a fixed drift. The process y_{it} for each observation $i = 1, \dots, n$ is defined as the stochastic development of a Euclidean model on $\mathbb{R}^m$. Consider the stochastic process, z_{it}, in $\mathbb{R}^m$ defined by the stochastic differential equation equivalent to the Euclidean regression model defined in (1),

$$dz_{it} = \alpha dt + W dX_{it} + d\varepsilon_{it}, \quad t \in [0, 1]. \tag{2}$$

Here αdt is a fixed drift, W the $m \times m$ coefficient matrix and ε_{it} the random error modelled as a Brownian motion in $\mathbb{R}^m$. The response process y_{it} is then given as the stochastic development of z_{it}, i.e. $y_{it} = \varphi_{(\boldsymbol{y}_0, \nu_0)}(z_{it})$ for a reference point $\boldsymbol{y}_0$ and frame $\nu_0 \in T_{\boldsymbol{y}_0}\mathcal{M}$ (see Fig. 1). The realizations are modelled as noisy observations of the endpoints of y_{it}, $\boldsymbol{y}_i = y_{i1} + \tilde{\boldsymbol{\varepsilon}}_i$ in which $\tilde{\boldsymbol{\varepsilon}}_i \sim \mathcal{N}(0, \tau^2 I)$ denotes iid. additive noise. There is a natural relation between W and the frame ν_0. If ν_0 is assumed to be an orthonormal basis and U the $d \times m$-matrix with columns of basis vectors of ν_0, then the matrix $\tilde{W} = UW$ explains the gathered effect of W and ν_0 through U. However, this decomposition is not unique and hence the $\tilde{W}$ matrix is estimated instead of U and W individually.

4 Method of Moments

In this section the MM procedure is introduced for the estimation of the parameters in the regression model. The MM procedure uses known moment conditions to define a set of equations which can be optimized to find the true parameter vector $\theta = (\tau, \alpha, \tilde{W}, \boldsymbol{y}_0)$, see [3,6,14]. Here τ^2 is the additive noise variance, α the drift, $\tilde{W}$ combined effect of covariates and ν_0, and $\boldsymbol{y}_0$ the initial point on $\mathcal{M}$.

In the SDR model the known moment conditions are based on the moments of the additive noise $\tilde{\boldsymbol{\varepsilon}}_i$ and the fact that $\tilde{\boldsymbol{\varepsilon}}_i$ is independent of the covariate variables x_{ik} for each $k = 1, \ldots, m$. Hence, the moment conditions are,

$$\mathbb{E}\left[\tilde{\varepsilon}_{ij}\right] = 0,\ \mathbb{E}\left[\tilde{\varepsilon}_{ij} x_{ik}\right] = 0,\ \mathbb{E}\left[\tilde{\varepsilon}_{ij}^2\right] = \tau^2\ \forall j = 1, \ldots, d, \text{ and } k = 1, \ldots, m.$$

Known consistent estimators for these moments are the sample means. Consider the residuals, $\hat{\varepsilon}_{ij} = y_{ij} - \hat{y}_{ij}$, in which the dependency of the parameter vector, θ, lies in the predictions, $\hat{y}_{ij}$ for $i = 1, \ldots, n,\ j = 1, \ldots, d$. For a proper choice of parameter vector θ, the sample means will approach the true moments. Therefore, the set of equations used to optimize the parameter vector θ are,

$$\frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_{ij} = 0, \quad \frac{1}{n}\sum_{i=1}^{n} x_{ik}\hat{\varepsilon}_{ij} = 0, \quad \text{and} \quad \frac{1}{n-2}\sum_{i=1}^{n} \hat{\varepsilon}_{ij}^2 = \hat{\tau}^2,$$

for all $j = 1, \ldots, d$ and $k = 1, \ldots, m$ and where $\hat{\tau}^2$ is the estimated variance. In Euclidean statistics, the method of moments is known to provide consistent estimators, but these estimators might be biased.

The cost function considered for optimization with respect to θ is,

$$f(\theta) = \frac{1}{d}\sum_{j}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_{ij}\right)^2 + \frac{1}{dm}\sum_{j,k}\left(\frac{1}{n}\sum_{i=1}^{n} x_{ik}\hat{\varepsilon}_{ij}\right)^2 + \frac{1}{d}\sum_{j}\left(\frac{1}{n-2}\sum_{i=1}^{n}\hat{\varepsilon}_{ij}^2 - \hat{\tau}^2\right)^2. \tag{3}$$

This cost function depends on predictions from the model based on the given parameter vector in each iteration. In order for the objective function to be stable it has to be evaluated for several predictions. Therefore, the function has been averaged for several predictions to obtain a more stable gradient descent optimization procedure.

The initial value of θ can in practice be chosen as parameters estimated from a Euclidean multivariate linear regression model. Here, the estimated covariance matrix would resemble the $\tilde{W}$ effect and the intercept the initial point $\boldsymbol{y}_0$.

5 Simulation Example

The performance of the estimation procedure will be evaluated using simulated data. We will generate landmark data on Riemannian landmark manifolds as defined in the Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework [18], and use the Levi-Civita connection. Shapes in the landmark manifold $\mathcal{M}$ are defined by a finite landmark representation, i.e. $q \in \mathcal{M}$, $q =$

$(x_1^1, x_1^2, \ldots, x_{n_l}^1, x_{n_l}^2)$, where n_l denotes the number of landmarks. The dimension of $\mathcal{M}$ is hence $d = 2n_l$. Using a kernel K, the Riemannian metric on $\mathcal{M}$ is defined as $g(v, w) = \sum_{i,j}^{n_l} vK^{-1}(\boldsymbol{x}_i, \boldsymbol{x}_j)w$ with K^{-1} denoting the inverse of the kernel matrix. In the following, we use a Gaussian kernel for K with standard deviation $\sigma = 0.1$. We will consider a single covariate variable $x \in \mathbb{R}$ drawn from $\mathcal{N}(0, 36)$ and model the relation to two response variables either with 1 or 3 landmarks. The response variables are simulated from a model with parameters given in Table 1 and Fig. 2 for $n_l = 3$. Examples of simulated data for $n_l = 1$ and 3 are shown in Fig. 2. The additive noise is in this case normally distributed iid. random noise added to each coordinate of landmarks. In this example we consider a simplification of the model, as the random error in z_{it}, given in (2), will be disregarded. Estimation of parameters is examined for three different models: one without additive noise and drift, one without drift, and at last the full model. For $n_l = 3$ only estimation of the two first models is studied, and estimation in the model with no drift has been considered for $n = 70$ and $n = 150$.

Table 1. Parameter estimates found with the MM procedure for 1 landmark. First column shows the true values and each column, estimated parameters in each model.

	True	Excl. τ, α $n = 70$	Excl. α $n = 70$	Excl. α $n = 150$	Full model $n = 150$
τ	0.1	$-(\tau = 0)$	0.256	0.226	0.207
α	40	$-(\alpha = 0)$	$-(\alpha = 0)$	$-(\alpha = 0)$	37.19
$\tilde{W}$	$(0, 2)$	$(0, 2.013)$	$(0.004, 1.996)$	$(0, 2.003)$	$(0, 2.004)$
$\boldsymbol{y}_0$	$(1, 0)$	$(1.064, 0.0438)$	$(1.158, 0.162)$	$(1.026, 0.0227)$	$(1.076, 2.708)$

By the results shown in Table 1 and Fig. 2, the procedure makes a good estimate of the frame matrix $\tilde{W}$ in every situation. For the model with no additive noise and no drift, the procedure finds a reasonable estimate of $\boldsymbol{y}_0$. When noise is added, it is seen that a larger sample size is needed in order to get a good estimate of $\boldsymbol{y}_0$. On the contrary, the variance estimate seems biased in each case. For $n_l = 3$ the variance parameters estimated were $\hat{\tau} = 0.306$ for $n = 70$ and $\hat{\tau} = 0.231$ for $n = 150$. However, when drift is added to the model, the estimation procedure has a hard time recapture the true estimates of $\boldsymbol{y}_0$ and α. This difficulty can be explained by the relation between the variables. In normal linear regression, only one intercept variable is present in the model, but in the SDR this intercept variable is split between α and $\boldsymbol{y}_0$.

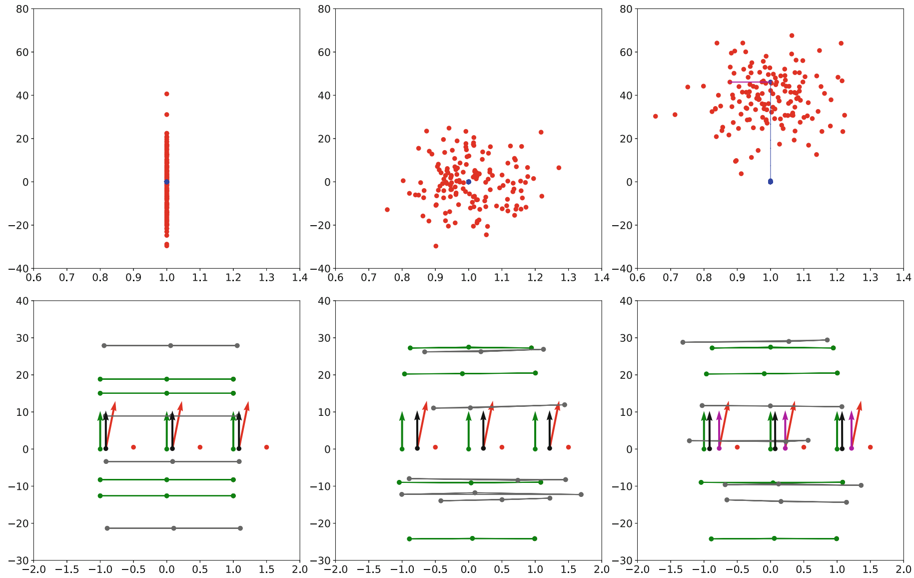

Fig. 2. (upper left) Sample drawn from model without additive noise and drift. (upper center) Sample drawn with additive noise, but no drift. (upper right) Sample drawn from the full model. The vertical lines are the stochastic development of z_{it} and the horizontal corresponds to the additive noise, the blue point is the reference point. (lower left) Model without drift and variance for $n_l = 3$, $n = 70$. (lower center) Model without drift and $n = 70$. (lower right) Model without drift and $n = 150$. These plots show the estimated results. (red) initial, (green) true, and (black) estimated reference point and frame. The gray samples are predicted from the estimated model while the green are a subset of the simulated data. Lower right plot does also show the difference in the estimated parameters for $n = 70$, $n = 150$ for the model with no drift. The magenta parameters in that plot is the estimated parameters for model without drift and $n = 70$, the corresponding black parameters in lower center plot. (Color figure online)

6 Conclusion

Method of Moments procedure has been examined for parameter estimation in the stochastic development regression (SDR) model. The SDR model is a generalization of regression models on Euclidean space to manifold-valued data. This model analyzes the relation between manifold-valued response and Euclidean covariate variables. The performance of the estimation procedure was studied based on a simulation example. The Method of Moments procedure was easier to apply and less computationally expensive than the Laplace approximation considered in [10]. The estimates found for the frame parameters were reasonable, but the procedure had a hard time retrieving the reference point and drift parameter. This is due to a mis-specification of the model as the reference point and drift parameter jointly correspond to the intercept in normal Euclidean regression models and hence there is no unique split of these parameters.

For further investigation, it could be interesting to test the relation between the reference point and drift parameter to be able to retrieve good estimates of these parameters. In the Euclidean case, the Method of Moments procedure has been shown to provide consistent, but sometimes biased estimates. An interesting question for future work could also be, whether the parameter estimates in this model is consistent and biased.

Acknowledgements. This work was supported by the CSGB Centre for Stochastic Geometry and Advanced Bioimaging funded by a grant from the Villum foundation.

References

1. Banerjee, M., Chakraborty, R., Ofori, E., Okun, M.S., Vaillancourt, D.E., Vemuri, B.C.: A nonlinear regression technique for manifold valued data with applications to medical image analysis. In: 2016 IEEE Conference on CVPR, pp. 4424–4432, June 2016
2. Cornea, E., Zhu, H., Kim, P., Ibrahim, J.G.: The Alzheimer's disease neuroimaging initiative. Regression models on Riemannian symmetric spaces. J. Roy. Stat. Soc. B **79**, 463–482 (2017)
3. Cragg, J.G.: Using higher moments to estimate the simple errors-in-variables model. Rand J. Econ. **28**, S71–S91 (1997)
4. Davis, B.C., Fletcher, P.T., Bullitt, E., Joshi, S.: Population shape regression from random design data. In: 2007 IEEE 11th International Conference on Computer Vision, pp. 1–7, October 2007
5. Fletcher, P.T.: Geodesic regression and the theory of least squares on Riemannian manifolds. Int. J. Comput. Vision **105**, 171–185 (2012)
6. Hazelton, M.L.: Methods of moments estimation. In: Lovric, M. (ed.) International Encyclopedia of Statistical Science, pp. 816–817. Springer, Heidelberg (2011). doi:10.1007/978-3-642-04898-2_364
7. Hinkle, J., Muralidharan, P., Fletcher, P.T., Joshi, S.: Polynomial Regression on Riemannian Manifolds. arXiv:1201.2395 (2012)
8. Hong, Y., Kwitt, R., Singh, N., Vasconcelos, N., Niethammer, M.: Parametric regression on the Grassmannian. IEEE Trans. Pattern Anal. Mach. Intell. **38**(11), 2284–2297 (2016)
9. Hsu, E.P.: Stochastic Analysis on Manifolds. American Mathematical Society, Providence (2002)
10. Kühnel, L., Sommer, S.: Stochastic development regression on non-linear manifolds. In: Niethammer, M., Styner, M., Aylward, S., Zhu, H., Oguz, I., Yap, P.-T., Shen, D. (eds.) IPMI 2017. LNCS, vol. 10265, pp. 53–64. Springer, Cham (2017). doi:10.1007/978-3-319-59050-9_5
11. Lin, L., St Thomas, B., Zhu, H., Dunson, D.B.: Extrinsic local regression on manifold-valued data. arXiv:1508.02201, August 2015
12. Niethammer, M., Huang, Y., Vialard, F.-X.: Geodesic regression for image time-series. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011. LNCS, vol. 6892, pp. 655–662. Springer, Heidelberg (2011). doi:10.1007/978-3-642-23629-7_80
13. Nilsson, J., Sha, F., Jordan, M.I.: Regression on manifolds using kernel dimension reduction. In: Proceedings of the 24th ICML, pp. 697–704. ACM (2007)

14. Pal, M.: Consistent moment estimators of regression coefficients in the presence of errors in variables. J. Econom. **14**(3), 349–364 (1980)
15. Niethammer, M., Huang, Y., Vialard, F.-X.: Geodesic regression for image time-series. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011. LNCS, vol. 6892, pp. 655–662. Springer, Heidelberg (2011). doi:10.1007/978-3-642-23629-7_80
16. Singh, N., Hinkle, J., Joshi, S., Fletcher, P.T.: Hierarchical geodesic models in diffeomorphisms. Int. J. Comput. Vision **117**, 70–92 (2016)
17. Singh, N., Vialard, F.-X., Niethammer, M.: Splines for diffeomorphisms. Med. Image Anal. **25**(1), 56–71 (2015)
18. Younes, L.: Shapes and Diffeomorphisms. Springer, Heidelberg (2010)
19. Yuan, Y., Zhu, H., Lin, W., Marron, J.S.: Local polynomial regression for symmetric positive definite matrices. J. Roy. Stat. Soc. Ser. B Stat. Methodol. **74**(4), 697–719 (2012)