

Principal Flows*

Victor M. Panaretos, Tung Pham & Zhigang Yao

*Section de Mathématiques
Ecole Polytechnique Fédérale de Lausanne*

Technical Report # 04/12, December 2012
Chair of Mathematical Statistics

Abstract

We revisit the problem of extending the notion of principal component analysis (PCA) to multivariate data sets that satisfy non-linear constraints, therefore lying on Riemannian manifolds. Our aim is to determine curves on the manifold that retain their canonical interpretability as principal components, while at the same time being flexible enough to capture non-geodesic forms of variation. We introduce the concept of a principal flow, a curve on the manifold passing through the mean of the data, and with the property that, at any point of the curve, the tangent velocity vector attempts to fit the first eigenvector of a tangent space PCA locally at that same point, subject to a smoothness constraint. That is, a particle flowing along the principal flow attempts to move along a path of maximal variation of the data, up to smoothness constraints. The rigorous definition of a principal flow is given by means of a Lagrangian variational problem, and its solution is reduced to an ODE problem via the Euler-Lagrange method. Conditions for existence and uniqueness are provided, and an algorithm is outlined for the numerical solution of the problem. Higher order principal flows are also defined. It is shown that principal flows can yield the usual principal components on a Euclidean space. By means of examples, it is illustrated that the principal flow is able to capture patterns of variation that can escape other manifold PCA methods.

Keywords: *covariance tensor field; Euler-Lagrange method; intrinsic mean; manifold; principal component analysis; tangent space.*

*Research supported by an ERC Starting Grant Award