A Differential Geometric Approach to Time Series Forecasting

Babak Emami[a,1,*]

[a] 1335 Filbert St., 303, San Francisco, CA, United States 94109

Abstract

A differential geometry based approach to time series forecasting is proposed. Given observations over time of a set of correlated variables, it is assumed that these variables are components of vectors tangent to a real differentiable manifold. Each vector belongs to the tangent space at a point on the manifold, and the collection of all vectors forms a path on the manifold, parametrized by time. We compute a manifold connection such that this path is a geodesic. The future of the path can then be computed by solving the geodesic equations subject to appropriate boundary conditions. This yields a forecast of the time series variables.

Keywords: time series, forecast, manifolds, geodesic

1. Introduction

Forecasting of time series arises in various fields, such as economics [1, 2]. The conventional approach is to regress the response time series, that is the dependent variable, against the time series of a set of variables which are called predictors or independent variables. The choice of regression model depends on the nature of relationship between response and predictors and can be selected from a wide range of models including linear regression, neural networks, and kernel methods such as the Gaussian process regression [3, 4]. Once the parameters of a model are estimated, the future of response can be calculated given the future values of predictors.

There are two issues with the above-mentioned approach. First, the future values of predictors are needed to forecast response. In most cases, future values of predictors are not known, making the forecasts dependent on potentially inaccurate estimations. For example in economics, the macroeconomic indicators are often used as predictors. The estimate forecasts of these economic indicators are often based on qualitative methods, such

* Corresponding author
[1] The author is an independent researcher.

as those in Blue Chip Economic Indicators and Blue Chip Financial Forecasts [5]. Second, the above-mentioned approach is based on the assumption of a causal relationship between response (as effect) and predictors (as cause). In many cases, however, the relationship between a set of correlated time dependent variables is not causal, and considering some variables as predictors is rather arbitrary. The approach proposed in the current work addresses these two problems.

With the increasing availability of complex datasets in various fields over the recent years, topological methods have been used to analyze complex data and to build models. For instance, concepts from algebraic topology such as homology has been used to study dominant topological properties of large complex datasets [6, 7, 8]. This approach provides a powerful tool with several desirable properties. For example, one of the benefits of reducing complex data to topological objects is that these geometric concepts do not depend on the chosen coordinates, but rather they are intrinsic geometric properties [6]. Topological data analysis has been combined with neural networks methods [9].

In the present work, another branch of topology, namely differentiable manifolds, is used to propose a framework to forecast time series. Here, the realized values of all variables at each time t are assumed to be components of a vector tangent to a real differentiable manifold M at a point $p(t) \in M$. With this assumption, the time series correspond to a set of vectors tangent to M at different points. These vectors form a path on M. Assuming that this path is a geodesic on manifold M, one can infer some of the structural properties of M from the data. This is then used to forecast all variables, without depending on future values of any predictors. Rather than taking a causal viewpoint by dividing variables into responses and predictors, the proposed approach searches for the logical relationship between variables by building the above-mentioned manifold.

In what follows, the mathematical approach is discussed in detail. This is followed by applying the proposed method to a simple problem as a canonical case. It is worth noting that the present work is aimed at performing a proof of concept of the ideas discussed rather than a performance comparison to the widely used models.

2. Methodology

Let M be an n dimensional real differentiable manifold equipped with a topology $\mathbb{O}$, and a differentiable atlas $\mathbb{A}$ which contains a chart (U, x) where $U \in \mathbb{O}$ is an open set, and $x : M \to \mathbb{R}^n$ is a homeomorphism.

Moreover, let us assume that we are given N observations $u^m(t_i)$ over a period of time $[0, T]$ for a set of n correlated time dependent variables u^m where $m \in [1, n] \cap \mathbb{N}$, $i \in [1, N] \cap \mathbb{N}$; $t_i \in [0, T]$ denotes the time of observation i.

We assume that observed values of these n variables at time t_i are components of a vector $\mathbf{u}_i$ tangent to manifold M, at a point $p(t_i) \in U \subseteq M$, with respect to the basis $\{\frac{\partial}{\partial x^m}\}$ in tangent space $T_{p(t_i)}M$ induced by chart $(U, x) \in \mathbb{A}$.

Without loss of generality, we can assume that the N point $p(t_i)$ lie on a smooth curve $\gamma : [0, T] \to M$, and that the N vectors $\mathbf{u}_i \in T_{p(t_i)}M$ are tangent to this curve. We denote the coordinates of the curve γ with respect to the chart (U, x) by $x^m(t)$. This is visualized in figure 1 which shows the coordinates x^1, x^2, and x^3 of the curve corresponding to three arbitrary time dependent variables u^1, u^2, and u^3. Here the vector tangent to the curve at $p(t)$ is naively illustrated as an arrow. The components of this vector are $u^1(t)$, $u^2(t)$, and $u^3(t)$.

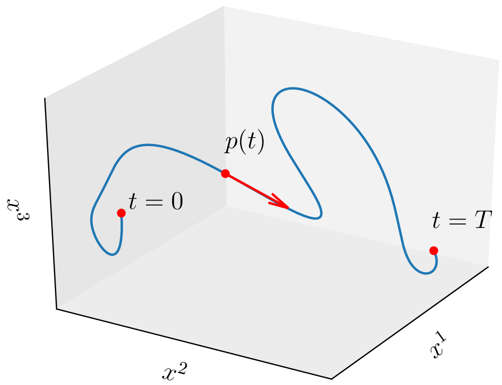

Figure 1: Visualization of a smooth curve on a three dimensional manifold M in coordinates x^1, x^2, and x^3 corresponding to three time dependent variables u^1, u^2, and u^3. The values $u^1(t)$, $u^2(t)$, and $u^3(t)$ at time $t \in [0, T]$ are components of the vector tangent to this curve.

Let us impose a restriction by assuming that this smooth curve is a geodesic on M. The coordinates x^m of the curve should then satisfy the geodesic equations [10, 11],

$$\ddot{x}^m + \Gamma^m_{ab}\dot{x}^a\dot{x}^b = 0 \tag{1}$$

where $x^m(t)$ are coordinates of $p(t) \in M$, $\ddot{x}^m = \frac{dx^m}{dt}$, and $\Gamma^m_{ab} = \Gamma^m_{ab}(p(t))$ are Christoffel symbols. Note that the summation convention is in effect. As u^ms are components of a tangent vector with respect to basis $\{\frac{\partial}{\partial x^m}\}$ in $T_{p(t)}M$, we have $u^m(t) = \frac{dx^m \circ p(t)}{dt}$, and so the geodesic equations can be formulated in terms of u^ms,

$$\dot{u}^m + \Gamma^m_{ab}u^a u^b = 0 \tag{2}$$

We can equip manifold M with a connection, and thus Christoffel symbols in chart (U, x), such that equation 2 holds. Once a connection is fixed, equation 2 subject to an appropriate boundary condition can be solved to forecast $u^m(t)$ for $t \in (T, \infty)$. The choice of connection is not unique. In fact, given that information provided by $u^m(t)$ for $t \in [0, T]$ corresponds to only one geodesic on M, the set of all connections that behave similarly near the geodesic path are equally valid choices.

2.1. Constraints on Christoffel symbols

It may be useful to constrain the manifold connection by assuming conditions on Christoffel symbols. The simplest constraint is to assume that Christoffel symbols are constant. This is a sufficient, but not necessary, condition for the manifold to have a constant curvature. This is a rather strong assumption. Alternatively, one can assume that given a chart, Christoffel symbols at each point on a manifold are linearly dependent on the coordinates of that point. The validity of any assumption depends on the nature of time series variables that define the manifold.

Here on, we assume that Christoffel symbols are constant to simplify the analysis in this preliminary study. What follows can be extended to more complicated forms of Christoffel symbols in a straightforward manner.

One can further constrain the Christoffel symbols by making assumptions about the underlying metric tensor. Assuming that Christoffel symbols are those of a Levi-Civita [10] connection, we have

$$\Gamma^m_{ab} = \frac{g^{ml}}{2}\left(\frac{\partial g_{al}}{\partial x^b} + \frac{\partial g_{lb}}{\partial x^a} - \frac{\partial g_{ab}}{\partial x^l}\right) \tag{3}$$

where g_{ab} are components of the metric tensor, and g^{ab} are the inverse components such that $g_{ab}g^{bc} = \delta^a_c$. Because the metric tensor is symmetric we have,

$$\Gamma^m_{ab} = \Gamma^m_{ba} \tag{4}$$

One can further assume that g_{ab} is diagonal. Looking at equation 3, this assumption indicates,

$$\Gamma^m_{ab}|_{m \neq a \neq b} = 0 \tag{5}$$

In the remainder of this work, it is assumed that Christoffel symbols are constant and correspond to a Levi-Civita connection with an underlying metric that has diagonal components in chart (U, x).

2.2. Choice of a boundary condition

We need to fix a boundary condition to solve the system of ODEs in equation 2. In general, we can impose a condition at any $t \in [0, T]$. Naturally, one may choose to impose a condition either at $t = 0$ or $t = T$. Here on, we assume that a condition is imposed at $t = T$, that is,

$$u^m(t = T) = u_0^m \tag{6}$$

What follows can be readily extended to other choices of a boundary condition.

2.3. Computation of Christoffel symbols

Let us denote the actual observed values of time series by $\hat{u}^m(t)$ to distinguish them from predictions $u^m(t)$ that satisfy equation 2. We need to compute a set of Christoffel symbols Γ^m_{ab} such that solution u^m of equation 2 is an estimation of $\hat{u}^m$. In other words, we aim to determine Γ^m_{ab} by fitting equation 2 over the known values of $\hat{u}^m(t_1)$, ..., $\hat{u}^m(t_N)$. This can be formulated as a constraint optimization problem,

$$\min_{\Gamma^m_{ab}} J(\Gamma^m_{ab}, u^1, ..., u^n, \hat{u}^1, ..., \hat{u}^n) \tag{7}$$

where the objective function J is defined as,

$$J := \frac{1}{2} \int_0^T dt \sum_{m=1}^{n} (u^m(t) - \hat{u}^m(t))^2 \tag{8}$$

Note that the choice an objective function is not unique. The above problem is constrained to the geodesic equation 2 and boundary condition $u^m(t = T) = \hat{u}^m(T)$.

2.4. Solution of the optimization problem

We need an optimization algorithm to solve the above problem numerically. Most optimization algorithms need the gradient of the objective function with respect to the optimization variables, that is $\frac{dJ}{d\Gamma^m_{ab}}$. As the number of optimization variables Γ^m_{ab} is potentially large, the numerical computation of all gradients is expensive. Here, we use an approach which is generally referred to as the continuous adjoint optimization method [12]. Using the adjoint method allows one to compute the gradient of the objective function, constrained by an ODE, with respect to the optimization variables, by solving the ODE and an extra "adjoint" ODE only once per optimization iteration.

Let us define $\tilde{\Gamma}^\alpha$ with $\alpha \in [1, n(2n-1)] \cap \mathbb{N}$ as the unique components of Γ^m_{ab} after applying the constraints in section 2.1. Here on, we consider components of $\tilde{\Gamma}^\alpha$ as the optimization variables. Let us add the constraint ODE to the objective function as a penalty term [12]. We have,

$$J(\tilde{\Gamma}^\alpha) = \frac{1}{2} \int_0^T dt \sum_{m=1}^{n} (u^m(t) - \hat{u}^m(t))^2 + \int_0^T dt v_m \left(\dot{u}^m + \Gamma^m_{ab} u^a u^b \right) \tag{9}$$

where v_m denotes the components of the adjoint variable. These are in fact the continuous Lagrangian multipliers used to apply the ODE constraint. Note that as u^ms transform as components of a tangent vector, v_ms should transform as components of a cotangent vector to keep equation 9 unchanged under a coordinate transformation. As such, one can consider $v_m(t)$ to be components of a cotangent vector $\boldsymbol{v}(t) \in T^*_{p(t)}$ with respect to the dual basis induced by chart (U, x).

Taking the gradient of the objective function J with respect to $\tilde{\Gamma}^\alpha$, integrating by parts, using equation 6, and imposing $v_r(t = 0) = 0$, we get,

$$\frac{dJ}{d\tilde{\Gamma}^\alpha} = \int_0^T \left[\frac{df}{dy^s} + 2\Gamma^m_{\ as} y^a v_m - \dot{v}_s \right] \frac{dy^s}{d\tilde{\Gamma}^\alpha} dt + \int_0^T v_m \frac{d\Gamma^m_{\ ab}}{d\tilde{\Gamma}^\alpha} y^a y^b dt \tag{10}$$

where $f = \frac{1}{2} \sum_{m=1}^n (u^m(t) - \hat{u}^m(t))^2$.

We can choose to satisfy,

$$\dot{v}_s - 2\Gamma^m_{\ as} y^a v_m = \frac{df}{dy^s} \tag{11}$$

with boundary condition,

$$v_r(t = 0) = 0 \tag{12}$$

Equation 10 then becomes,

$$\frac{dJ}{d\tilde{\Gamma}^\alpha} = \int_0^T v_m \frac{d\Gamma^m_{\ ab}}{d\tilde{\Gamma}^\alpha} y^a y^b dt \tag{13}$$

We can calculate the gradient of the objective function using equation 13, where the adjoint variables v_s are calculated by solving the adjoint equation 11 subject to 12.

As we have a method to compute the gradients of the objective function we can use any gradient based optimization method to solve equation 7 subject to 2 as a constraint.

3. Results

As a proof of concept, the proposed model was applied to a simple canonical problem consisting of three time dependent variables, $u^1(t)$, $u^2(t)$, and $u^3(t)$. The time series were generated by adding a random noise to three correlated smooth curves for $t_i \in [0, 1.2]$ where $i \in [1, 120000] \cap \mathbb{N}$; the values corresponding to $t \in [0, 1]$ were used to build a model and the remaining $t \in [1, 1.2]$ were used for out of sample testing of the forecast model.

The SLSQP optimization algorithm [13] was used to solve 7. The objective function gradients were computed by solving the geodesic and adjoint ODEs as was described in section 2.4. The LSODA algorithm [14] and an explicit Runge-Kutta method of order 5 [15] were used to solve equations 2 and 11, respectively. Note that the geodesic ODEs were solved subject to a boundary condition at $t = 1$. The optimization algorithm converged to a relative error of 1.0e-3 within a few iterations. The computed Christoffel symbols Γ^m_{ab} from optimization were then used to forecast the out of sample values for $t \in [1, 1.2]$ by solving the geodesic ODEs, again subject to a boundary condition at $t = 1$.

Figure 2 shows a comparison of actual values of the three time series to those predicted by the proposed model. The in sample predictions, for $t \in [0, 1]$, are plotted in red, and the out of sample predictions, for $t \in [1, 1.2]$, are plotted in yellow. The model captures the trend of the time series reasonably well.

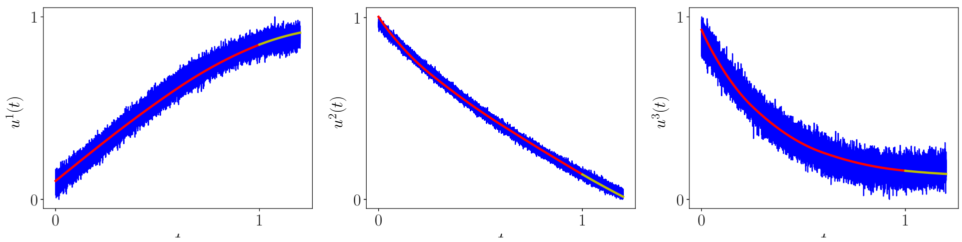

Figure 2: Comparison of actual and predicted time series $u^1(t)$, $u^2(t)$, and $u^3(t)$.

Figure 3 shows the corresponding actual and predicted geodesic curves in coordinates x. This is the chart map that induces the tangent space bases which correspond to $u^1(t)$, $u^2(t)$, and $u^3(t)$. Again, the in sample and out of sample predictions are plotted in red and yellow, respectively, and the agreement between actual and predicted curves is reasonably well.

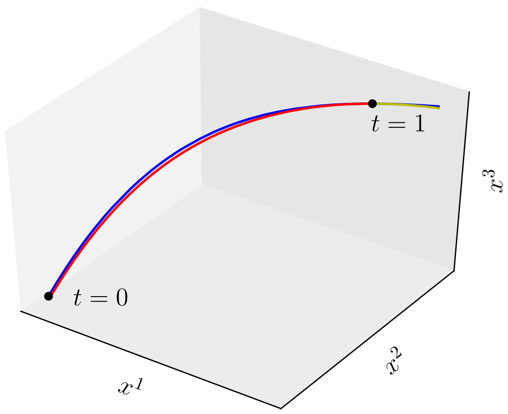

Figure 3: Comparison of actual and predicted geodesics, plotted in chart (U, x).

4. Conclusions

The present study aims at performing a proof of concept for a proposed time series forecast model. The algorithm, under the simplifying assumptions discussed in section 2.1, works reasonably well on a simple canonical problem. The future work may include applying this model to more complex problems and to relax some of the constraints in section 2.1.

References

[1] P. Kennedy, A Guide to Econometrics, The MIT Press, 1998.

[2] F. J. Fabozzi, F. Modigliani, F. J. Jones, Foundations of Financial Markets and Institutions, Pearson Education Limited, 2014.

[3] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer Series in Statistics, Springer New York, 2009.

[4] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning, The MIT Press, 2005.

[5] Blue chip publications (2019).

[6] G. Carlsson, Topology and data, Bulletin of The American Mathematical Society 46 (2009) 255–308.

[7] P. Bubenik, Statistical topological data analysis using persistence landscapes, Journal of Machine Learning Research 16 (3) (2015) 77–102.

[8] L. Wasserman, Topological data analysis, Annual Review of Statistics and Its Application 5 (2018) 501–532.

[9] G. R. Carlsson, G., Topological approaches to deep learning, Topological Data Analysis. Abel Symposia 15 (2020) 119–146.

[10] F. de Felice, C. J. S. Clarke, Relativity on curved manifolds, Cambridge University Press, 1992.

[11] E. Schrödinger, E. Dinger, Space-Time Structure, Cambridge Science Classics, Cambridge University Press, 1985.

[12] M. B. Giles, N. A. Pierce, An introduction to the adjoint approach to design, Flow, Turbulence and Combustion 65 (2000) 393–415.

[13] J. Nocedal, S. Wright, Numerical Optimization, Springer Series in Operations Research and Financial Engineering, Springer New York, 2006.

[14] L. Petzold, Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations, SIAM Journal on Scientific and Statistical Computing 4 (1983) 136–148.

[15] J. R. Dormand, P. J. Prince, A family of embedded runge-kutta formulae, Journal of Computational and Applied Mathematics 6 (1980) 19–26.