Metric Analysis Approach for Interpolation and Forecasting of Time Processes

A.V. Kryanev
National Nuclear Research University ”MEPhI”
115409, Moscow, Russia

D.K. Udumyan
National Nuclear Research University ”MEPhI”
115409, Moscow, Russia

G.V. Lukin
National Nuclear Research University ”MEPhI”
115409, Moscow, Russia

V.V. Ivanov
Laboratory of Information Technologies
Joint Institute for Nuclear Research, 141980, Dubna, Russia

Abstract

Interpolation and forecasting schemes are derived in the frame of the metric analysis method for time series described by functions of several variables. The efficiency and reliability of the proposed approach are illustrated on case study problems.

Keywords: metric analysis; time series interpolation; accuracy of forecasting
1 Introduction

The analysis and forecast of time series are well-known problems of interest in applied mathematics. To date, there are known several different methods and schemes resulting in effective solution of forecasting problems in different areas [1-6,8,13].

In the paper we describe a novel approach based on the metric analysis method. This approach results in reliable and efficient solution of problems concerning the interpolation and the extrapolation of a function \( Y \) of several variables. The assumed knowledge of \( Y \) covers some particular set of argument values [7,12].

2 Metric analysis based interpolation and smoothing schemes

Within the metric analysis method, the interpolation scheme starts with a functional dependence:

\[
Y = F(X_1, ..., X_m) = F(\vec{X}),
\]

where the function \( F(\vec{X}) \) is unknown and is subject to recovery at one point or at a set of specified points using the known values of a function \( Y_k, k = 1, ..., n, \vec{Y} = (Y_1, ..., Y_n)^T \) at a set of fixed points, \( \vec{X}_k = (X_{k1}, ..., X_{km})^T \) [9].

Suppose that in the space \( E^m \) there is a metric generated by the norm:

\[
||\vec{X}|| = \sum_{j=1}^{m} w_j X_j^2,
\]

where \( w_j > 0, \sum_{j=1}^{m} w_j = m \) are the metric weights.

Note: an essential feature of the proposed metric analysis is that the weights \( w_j, j = 1, ..., m \) are not predefined, but they are calculated using the specified values \( Y_k, \vec{X}_k, k = 1, ..., n \) (see below). For the time being, we assume known values for the metric weights \( w_j, j = 1, ..., m \).

According to the interpolation schemes based on the metric analysis, the interpolation values are found from a minimum on \( \vec{z} = (z_1, ..., z_n)^T \) measure of the metric uncertainty [9-12]:

\[
\sigma^2_{ND}(Y^*; \vec{z}) = (W(\vec{X}^*; \vec{X}_1; ...; \vec{X}_n)\vec{z}, \vec{z}),
\]

where the interpolated value at \( \vec{X}^* \) is determined from

\[
Y^* = \sum_{i=1}^{n} z_i Y_i,
\]
therefrom
\[ Y^* = \frac{(W^{-1}Y, \bar{1})}{(W^{-1}, \bar{1})}, \]  
(4)

while the matrix of metric uncertainty is determined from
\[ W = \begin{pmatrix}
\rho^2(\bar{X}_1, \bar{X}_1^*)& (\bar{X}_1, \bar{X}_2^*) & \ldots & (\bar{X}_1, \bar{X}_n^*) \\
(\bar{X}_2, \bar{X}_1^*)& \rho^2(\bar{X}_2, \bar{X}_1^*)& \ldots & (\bar{X}_2, \bar{X}_n^*) \\
\vdots & \vdots & \ddots & \vdots \\
(\bar{X}_n, \bar{X}_1^*)& (\bar{X}_n, \bar{X}_2^*)& \ldots & \rho^2(\bar{X}_n, \bar{X}_n^*)
\end{pmatrix}, \]

where
\[ \rho^2_W(\bar{X}_i, \bar{X}_i^*) = \sum_{k=1}^{m} w_k (X_{ik} - X^*_k)^2, \]

\[ (\bar{X}_i, \bar{X}_j)_W = \sum_{k=1}^{m} w_k (X_{ik} - X^*_k) \cdot (X_{jk} - X^*_k), \quad i,j = 1,...,n. \]

If the matrix of the metric uncertainty
\[ W = W(\bar{X}^*/\bar{X}_1,...,\bar{X}_n) \]
is singular, then the equation (4) is replaced by [11]:
\[ Y^* = \frac{(W^{-1}Y, \bar{1})}{(W^{-1}, \bar{1})}. \]
(5)

Here \( W^{-1}_\alpha \) is the inverse matrix of the regularized matrix \( W_\alpha = W + \alpha \cdot B \), where \( \alpha > 0 \) is a regulating parameter and \( B \) is a positive definite regulating matrix [9, 11].

If the values of the function \( Y_k, k = 1,...,n \) are affected by random errors, then the general uncertainty \( \sigma^2(Y^*) \) at the point \( X^* \) is defined as the sum of the metric uncertainty \( \sigma^2_{ND}(Y^*; \bar{z}) \) and the stochastic uncertainty \( (K_{\bar{Y}} \bar{z}, \bar{z}) \), where \( K_{\bar{Y}} \) is the covariance matrix of the random vector \( \bar{Y} = (Y_1,...,Y_n)^T \).

The minimization of the uncertainty given by the general expression
\[ \sigma^2(Y^*) = \sigma^2_{ND}(Y^*; \bar{z}) + \alpha \cdot (K_{\bar{Y}} \bar{z}, \bar{z}) \]
under the condition \( (\bar{z}, \bar{1}) = 1 \), yields a solution of the form (5), where \( W_\alpha = W + \alpha \cdot K_{\bar{Y}} \), and an optimum numerical value of the parameter \( \alpha > 0 \) follows from a residual principle [11].

From (5) it follows that, under \( \alpha \to \infty \), the smoothed restored values of the function \( Y \) at an arbitrary point tends to the value
\[ Y^{**} = \frac{(B^{-1}Y, \bar{1})}{(B^{-1}, \bar{1})}. \]
This means that the maximum degree of spatial smoothness is reached.

In particular, if the distances between the points $\vec{X}_1, ..., \vec{X}_n$ are very small as compared to the possible casual fluctuations of the random values of the function, then, instead of (7), we have the following estimation,

$$Y^{**} = \frac{(K^{-1} \vec{Y}, \vec{1})}{(K^{-1} \vec{1}, \vec{1})}.$$  \hspace{1cm} (8)

Let us consider now the problem of selecting the metric weights which determine the norm in the space $E^m$. The choice of the weights is determined by the degree of the function under consideration when its arguments change. The larger the function change when editing an argument in relation to the changes of other arguments, the larger should be the metric weight corresponding to that argument. Therefore, the matrix of metric uncertainty enables the definition of the degree of dependence of the change of a function in terms of the changes in each of its arguments through the metric weights. Assume now that the degree of the function variation at change of its arguments is known a priori (for example, its partial derivatives are known) in a given subspace of $E^m$. Then the normalized values of the weights (assumed to be defined, such that their sum gets equal to $m$) are chosen proportional to the absolute values of the derivatives. If there is no such an information, then the definition of the metric weights can be done following the scheme provided in [10,11].

Figure 1 provides an instance of chaotic time series smoothing with help of the metric analysis. For the deterministic source process given by the solid line, the use of the smoothing scheme (5) - (6) to the original restoration from the chaotic noise process represented by zigzag line results in the restored process shown by the dashed line.

\section{Forecasting schemes based on the metric analysis}

Let the time process $y = f(t)$ shows the known values $y_1 = f(t_1), ..., y_n = f(t_n)$ at the argument sequence $t_1 < t_2 < ... < t_n$. A predicted value $y_{n+1}$ at a point $t_{n+1}$ is requested.

The problem of finding the predicted value $y_{n+1}$ is reduced to the problem of interpolating a multidimensional function with the help of the nonlinear
Metric analysis approach

1057

Figure 1: Results of restoration a chaotic time series

auto-regression:
\[ y(t_{m+1}) = Y_{m+1} = F(y_1, ..., y_m), \]
\[ y(t_{m+2}) = Y_{m+2} = F(y_2, ..., y_{m+1}), \]
\[ ... \]
\[ y(t_n) = Y_n = F(y_{n-m}, ..., y_{n-1}). \]

The predicted value \( y_{n+1} \) is defined as an interpolation of the \( m \)-dimensional function \( F \) at a point \( \vec{X}^* \):
\[ y_{n+1} = F(\vec{X}^*), \quad (9) \]

where
\[ \vec{X}^* = (Y_{n-m+1}, ..., Y_n)^T. \]

The natural number \( m \) defines the dimension of the space spanned by the vectors \( \vec{X} \) and its value is found as a solution of the optimization problem
\[ m = \arg \min_m \| \vec{Y} - \vec{Y}_{for} \|, \quad (10) \]
where \( \vec{Y}_{for} \) is a vector of predicted values obtained from those of the function \( y = f(t) \) by using the method described in [11].

Figure 2 shows the results of chaotic time series forecasting with help of the metric analysis method. The ratio of correctly predicted falls and growths compared to the analyzed time series is about 80 percent.

Figures 3 and 4 provide two examples of deterministic time series forecasting the with help of the metric analysis approach. Figure 3 shows the results of forecasting 50 steps for the time process \( y = exp(-0.2 \sqrt{|t|}) sin(0.1t^2) \). Figure 4 gives the results of forecasting 150 steps for the time process \( y = exp(t) sin(2.9t) \). Original data are drawn by solid lines while the predicted curves are shown by dashed lines.
4 Conclusions

The presented method of forecasting using the metric analysis approach allows the identification of those arguments which result in the most significant influence on the future values of the investigated temporal process and the inference of this degree of impact on the forecasted values. Numerical results obtained with the help of this method point to a reliable highly accurate retrieval of the values the studied time processes.

Acknowledgements

The authors thank Prof. Gh. Adam for useful comments and assistance with translation of the article.

References


Figure 3: Results of forecasting 50 steps for the first deterministic time process

Figure 4: Results of forecasting 150 steps for the second deterministic time process


[7] A.V. Kryanev, V.V. Ivanov, G.V. Lukin, D.K. Uдумyan, S.G. Klimanov, Mathematical methods and algorithms for predicting time series based


Received: December 1, 2013