CALCULATING THE VALUE OF A MULTIVARIATE TIME SERIES AT A POINT BY THE PREDICTED MATRIX OF PAIRWISE CORRELATION

Maxim Divilkovskiy
Chair of Data Analysis
MIPT
divilkovskii.mm@phystech.edu

Vadim Strijov
FRC CSC of the RAS
Moscow, Russia
strijov@phystech.edu

Konstantin Yakovlev
Chair of Intellectual Systems
MIPT
iakovlev.kd@phystech.edu

ABSTRACT

We are researching the problem of pointwise forecasting of a set of time series with high covariance and high variance. To solve this problem we propose considering the space of pairwise distances between time series. In this space the pairwise distance matrix is predicted and then the time series values at the next time moment are reconstructed using the known matrix. In this paper we propose several methods for recovering a prediction in time series space from a known pairwise distance matrix. We show the existence of multiple time series values satisfying the same pairwise distance matrix. We propose two algorithms based on the use of matrices constructed over different time intervals using pairwise correlation. Also, the paper derives an explicit view of the recovered values through the pairwise correlation matrix. In addition, we derive an evaluation of the quality of the reconstruction when noise is added to the pairwise distance matrices. The novelty of the method is that the prediction is not done in the original space, but in the space of pairwise distances.

Keywords Metric · Non-Convex Optimization · Correlation · Time Series Forecasting

1   Introduction

Our purpose is to present a new method for pointwise time series forecasting. Considered time series are characterized by high pairwise covariance and high variance. The task of pointwise forecasting is to calculate the values of time series at the next point in time using the available historical data for the previous several points in time. The task is divided into three stages: first, the initial space of time series is transformed into a metric space by constructing a matrix of pairwise distances. Then the prediction of the matrix of pairwise distances at the next moment of time is carried out in this space. In the theoretical part of the paper, we show the necessity of predicting at least two matrices. We also show that the uniqueness of the answer requires the use of several matrices corresponding to different time intervals. In the last step, the predicted matrices are used to return the result into the original space. We focus mostly on the last step. Experiments are presented for the case of accurate matrix prediction with the addition of normal noise to the values.

Existing time series prediction methods such as LSTM [4], SSA [3] and other [8], [1] are based on predicting the value of a single series. These methods can be modified to forecast also a set of time series. For this purpose, it is sufficient to consider a set of series as one multivariate series. However, this approach does not explicitly model the dependencies between different series. In this paper, we propose to analyze the change in *set* of time series. This approach explicitly uses the relationships between them as information. A similar study is carried out in the [5] paper, but it emphasizes on the feature selection task. This task consists in selecting such a subset of the original time series for which it is possible to make a forecast of sufficient quality.

Further, we consider conditions on the distance function between rows under which there is a way to recover the values of time series. The insufficiency of one matrix to recover the answer is proved. Two methods using several matrices are proposed for the case of accurate forecast and for the case of forecast with non-zero noise. Also, we propose a recovery algorithm based on two theorems about the explicit form of the response using pairwise correlation as a function of pairwise distance between rows. Mean Squared Error and Mean Absolute Error are used as quality criteria. It is shown in the article[6] that they are the most suitable for the task of time series forecasting.

2   Formulation of the problem of pointwise forecasting of a set of time series

A set of d time series is given by t vectors:

$$[\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_t], \text{for each } k : \mathbf{x}_k \in \mathbb{R}^d,$$

$\mathbf{x}_{t_i,k} \in \mathbb{R}$ is the value of the series with index k at the time moment t_i.

The task is to predict $\mathbf{x}_{t+1}$ — the set of values of all time series at the time moment $t + 1$.

The matrix $\mathbf{x}$ of size $t \times d$ is treated as a *multidimensional* time series, treating the value at a point as an element of space $\mathbb{R}^d$.

Prediction errors are calculated using the following equations:

$$\text{MAE} = \frac{\sum_{i=1}^{d} |\mathbf{x_{t+1}} - \hat{\mathbf{x}}_{\mathbf{t+1}}|}{d},$$

$$\text{MSE} = \frac{\sum_{i=1}^{d} (\mathbf{x_{t+1}} - \hat{\mathbf{x}}_{\mathbf{t+1}})^2}{d}.$$

3   Algorithm in case of predicting a single distance matrix

1. Distance matrices between a set of time series values are constructed according to the scheme below.

$$[\mathbf{x}_1, \ldots, \mathbf{x}_s] \rightarrow \mathbf{\Sigma}_s,$$
$$[\mathbf{x}_2, \ldots, \mathbf{x}_{s+1}] \rightarrow \mathbf{\Sigma}_{s+1},$$
$$\vdots$$
$$[\mathbf{x}_{t-s}, \ldots, \mathbf{x}_t] \rightarrow \mathbf{\Sigma}_t.$$

Each $\mathbf{\Sigma}_i$ — are matrices of pairwise distances of size $d \times d$. An element of the matrix $\mathbf{\Sigma}_{i,j}$ is the value of the distance between rows i and j. The construction formula as well as the choice of the distance function are described in the following sections.

2. The matrix $\hat{\mathbf{\Sigma}}_{t+1}$ is predicted from these matrices.

3. Find such $\hat{\mathbf{x}}_{t+1}$, that

$$||\hat{\mathbf{\Sigma}}_{t+1} - \bar{\mathbf{\Sigma}}_{t+1}||_2^2$$

is minimal, where $\bar{\mathbf{\Sigma}}_{t+1}$ is the distance matrix, constructed from the set $[\mathbf{x}_{t-s+1}, \ldots, \mathbf{x}_t, \hat{\mathbf{x}}_{t+1}]$. Achieving a minimum by this function would mean that $\hat{\mathbf{\Sigma}}_{t+1}$ are $\bar{\mathbf{\Sigma}}_{t+1}$ equal. In turn, this means that the found continuation of the series at time $t + 1$ has a distance matrix equal to the prediction. In general, this function is non-convex and there can be several minimums. We address this problem in the following sections.

4   Existence of several values of a series satisfying same distance matrix

The existence of several solutions to the minimization problem described above is the central problem we consider in this paper. In this section we show that by one matrix constructed by an arbitrary metric it is possible to recover several different values of the series at the next moment of time. A solution to this

problem in the case of using pairwise correlation as a function of distance between certain time series is proposed in the next section.

Consider the recovery of the prediction from the matrix $\mathbf{\Sigma}_{t+1}$ to time series space, where $\mathbf{\Sigma}_{t+1}$ is the matrix of pairwise distances corresponding to the multivariate series $\mathbf{x} = [\mathbf{x_1}, \ldots, \mathbf{x_{t+1}}]$.

Given a pairwise distance matrix $\mathbf{\Sigma}$ of size $d \times d$ for multivariate time series $\mathbf{x} \in \mathbb{R}^{d \times (t+1)}$. $\hat{\mathbf{x}}_{t+1} \in \mathbb{R}^d$ is predicted. Also, the metric

$$\rho : \mathbb{R}^{t+1} \times \mathbb{R}^{t+1} \to \mathbb{R},$$

defined on time series, which satisfies the properties of the metric. Which are

$$\mathbf{\Sigma}_{i,j} = \rho(\mathbf{x}_{1\ldots t,i} \circ \hat{\mathbf{x}}_{t+1,i}, \mathbf{x}_{1\ldots t,j} \circ \hat{\mathbf{x}}_{t+1,j}),$$

where $\circ$ denotes the concatenation of vectors.

One of the frequently used metrics for calculating distance of objects in $\mathbb{R}$ is the Euclidean metric. Using its example, we will show that there can be several suitable answers with the same values of distance matrix. Moreover, in the case of the Euclidean metric, the number of answers is infinite,

$$\mathbf{\Sigma}_{i,j} = \rho(\mathbf{x}_{1\ldots t,i} \circ \hat{\mathbf{x}}_{t+1,i}, \mathbf{x}_{1\ldots t,j} \circ \hat{\mathbf{x}}_{t+1,j}) = \sqrt{\left(\sum_{k=1}^{t}(\mathbf{x}_{k,i} - \mathbf{x}_{k,j})^2\right) + (\hat{\mathbf{x}}_{t+1,i} - \hat{\mathbf{x}}_{t+1,j})^2}.$$

Using this metric results in the fact that adding some constant C to all $\hat{\mathbf{x}}_{t+1,i}$ does not change the answer:

$$\sqrt{\left(\sum_{k=1}^{t}(\mathbf{x}_{k,i} - \mathbf{x}_{k,j})^2\right) + (\hat{\mathbf{x}}_{t+1,i} - \hat{\mathbf{x}}_{t+1,j})^2} = \sqrt{\left(\sum_{k=1}^{t}(\mathbf{x}_{k,i} - \mathbf{x}_{k,j})^2\right) + [(\hat{\mathbf{x}}_{t+1,i} + C) - (\hat{\mathbf{x}}_{t+1,j} + C)]^2}.$$

In the case of the time series prediction problem, this property is critical because even if the matrix $\mathbf{\Sigma}$ is predicted accurately, there are infinitely many time series values at time $t + 1$ corresponding to this matrix. See 1 for the example of this effect.
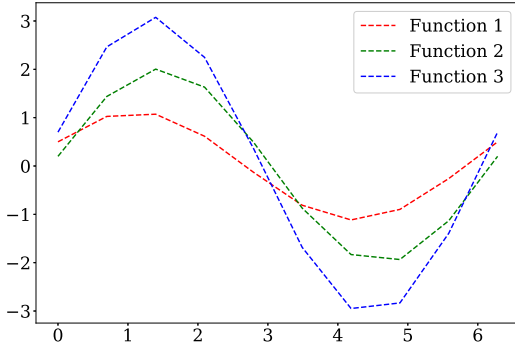
(a) Reference functions
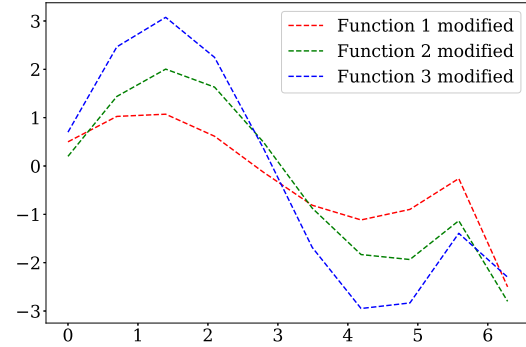
(b) Modified functions where 3 is subtracted from the last values

Figure 1: These sets of time series have the same distance matrix constructed on the Euclidean metric.

However, even using other metrics does not get rid of the problem. This results in the inability to use the classical Multidimensional Scaling [2] algorithm to recover the response to the source time series space. This algorithm is often used to recover objects from their pairwise distance. Metric MDS may be used if distances are not euclidean.

However, even using other metrics does not get rid of the problem. In theorem below we show that usage of only one matrix is insufficient for the problem of recovering values of time series.

Theorem 1. *For any metric ρ defined in the time series space $\mathbb{R}^t$ there is more than one way to recover the original time series from the pairwise distance matrix constructed by the given metric.*

Note 1. This statement does not use information about the first $t-1$ values of the series. In fact, the series in this case can be thought of as a point in $\mathbb{R}^t$ space. The usage of information about the previous moments of time is given after this section.

Proof. We only have to show that the metric is not a bijection. This will mean that there are several different pairs of rows whose distance between them is the same.

Let us show that the metric is a continuous function. Take the sequence

$$\{(\mathbf{x}_n, \mathbf{y}_n)\} \subset \mathbb{R}^t \times \mathbb{R}^t, (\mathbf{x}_n, \mathbf{y}_n) \to (\mathbf{x}, \mathbf{y}).$$

Then,

$$\mathbf{x}_n \to \mathbf{x}, \mathbf{y}_n \to \mathbf{y} \Rightarrow \rho(\mathbf{x}_n, \mathbf{x}) \to 0, \rho(\mathbf{y}_n, \mathbf{y}) \to 0,$$

$n \to \infty$. Using the triangle inequality for the metric, we obtain

$$\rho(\mathbf{x}_n, \mathbf{y}_n) \leqslant \rho(\mathbf{x}_n, \mathbf{x}) + \rho(\mathbf{x}, \mathbf{y}) + \rho(\mathbf{y}_n, \mathbf{y}) \to \rho(\mathbf{x}, \mathbf{y}),$$

therefore, $\rho(\mathbf{x}_n, \mathbf{y}_n) \to \rho(\mathbf{x}, \mathbf{y})$.

Therefore the metric is a continuous mapping from $\mathbb{R}^t \times \mathbb{R}^t$ to $\mathbb{R}$. We will show that such a mapping cannot be a homeomorphism. Assume that $f : \mathbb{R} \to \mathbb{R}^t \times \mathbb{R}^t$ is the desired homeomorphism. Take arbitrary point $a \in \mathbb{R}$ and $f(a)$. Removing point a, $\mathbb{R}$ is no longer connected, but $\mathbb{R}^t \times \mathbb{R}^t$ is still connected. Therefore it is not a homeomorphism. We got a contradiction. ∎

Note 2. Essentially, the proof uses only the continuity of the function. This means that even non-metric continuous functions will give the multiplicity of the answer. For example, pairwise correlation of series is also a continuous function.

Thus, knowing only the distance matrix it is impossible to unambiguously reconstruct the original series.

Now consider the same problem, in addition to the matrix $\mathbf{\Sigma}_{t+1}$ using the value of the time series before the time moment t: $\mathbf{X} = [\mathbf{x_1}, \ldots, \mathbf{x_t}]$. The problem is reformulated as follows:

There are n objects in $\mathbb{R}^{t+1}$, their first t coordinates are known. We also know the distance matrix $\mathbf{\Sigma}_{t+1} \in \mathbb{R}^{(t+1) \times (t+1)}$. It is required to recover the $t+1$ coordinate of each of the objects. In time series terms, the $t+1$-coordinate is the value of each of the series at that moment in time.

5   Pairwise correlation between time series as the distance function

This section investigates the time series values recovery using the pairwise correlation matrix. This distance function is used because it is shown in the article[9] that the pairwise correlation estimate of a sample approximates its mathematical expectation.

The pairwise distance matrix is constructed as follows:

$$\mathbf{\Sigma}_T = \frac{1}{T} \sum_{t=1}^{T} (\mathbf{x}_t - \boldsymbol{\mu}_T)(\mathbf{x}_t - \boldsymbol{\mu}_T)^{\mathsf{T}},$$

$$\boldsymbol{\mu}_T = \frac{1}{T} \sum_{t=1}^{T} \mathbf{x}_t.$$

Theorem 2. *In the case where we have accurately predicted the distance matrix, the function $||\hat{\mathbf{\Sigma}}_{t+1} - \bar{\mathbf{\Sigma}}_{t+1}||_2^2$ will have two minimums, defined explicitly as follows:*

$$\hat{\mathbf{y}}_i = \mathbf{y}_i,$$

$$\hat{\mathbf{y}}_i = \frac{2}{T-1} \sum_{k=1}^{T-1} \mathbf{a}_{ik} - \mathbf{y}_i,$$

where $\hat{\mathbf{y}}_i$ — i-th coordinate of the predicted value of the series at the moment $T + 1$, $\mathbf{A} = (\mathbf{a}_{ik})$ is given multivariate time series, y_i are true values of the series at the moment $T + 1$.

Proof. Let us denote by $\mathbf{\Sigma}$ the true matrix at the moment of time T, and $\hat{\mathbf{\Sigma}}$ the predicted one. By construction, $\mathbf{\Sigma} = \frac{1}{T} \sum_{k=1}^{T} (\mathbf{a}_k^T - \boldsymbol{\mu}_T)(\mathbf{a}_k^T - \boldsymbol{\mu}_T)^{\intercal}$. The matrix $\mathbf{A}$ is a transposed matrix $\mathbf{X}$ of time series, where the first dimension is the number of the series, not the moment of time as in the case of $\mathbf{X}$. Then, consider what the elements of the matrices $\mathbf{\Sigma}$ and $\hat{\mathbf{\Sigma}}$ are equal to.

$$\mathbf{\Sigma}_{ij} = \frac{1}{T} \sum_{k=1}^{T} (\mathbf{a}_{ik} - \boldsymbol{\mu}_i)(\mathbf{a}_{jk} - \boldsymbol{\mu}_j),$$

$$\hat{\mathbf{\Sigma}}_{ij} = \frac{1}{T} \sum_{k=1}^{T-1} (\mathbf{a}_{ik} - \hat{\boldsymbol{\mu}}_i)(\mathbf{a}_{jk} - \hat{\boldsymbol{\mu}}_j) + (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)(\mathbf{y}_j - \hat{\boldsymbol{\mu}}_j).$$

Since we minimise the norm of the difference, both matrices are equal to each other. Consider the equality of diagonal elements.

For any i, j such that $i = j$ the following is true: $\mathbf{\Sigma}_{ii} = \hat{\mathbf{\Sigma}}_{ii}$,

$$\sum_{k=1}^{T} (\mathbf{a}_{ik} - \boldsymbol{\mu}_i)(\mathbf{a}_{ik} - \boldsymbol{\mu}_i) = \sum_{k=1}^{T-1} (\mathbf{a}_{ik} - \hat{\boldsymbol{\mu}}_i)(\mathbf{a}_{ik} - \hat{\boldsymbol{\mu}}_i) + (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i),$$

$$(\mathbf{a}_{iT} - \boldsymbol{\mu}_i)^2 = (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^2.$$

Consider $\hat{\boldsymbol{\mu}}_i$ and $\hat{\boldsymbol{\mu}}$:

$$\hat{\boldsymbol{\mu}}_i = \frac{1}{T} \sum_{k=1}^{T-1} \mathbf{a}_{ik} + \frac{1}{T} \mathbf{y}_i,$$

$$\boldsymbol{\mu}_i = \frac{1}{T} \sum_{k=1}^{T} \mathbf{a}_{ik},$$

$$\left[\begin{array}{l} \frac{T-1}{T} \mathbf{y}_i - \frac{1}{T} \sum_{k=1}^{T-1} \mathbf{a}_{ik} = \mathbf{a}_{iT} - \boldsymbol{\mu}_i \\ \frac{T-1}{T} \mathbf{y}_i - \frac{1}{T} \sum_{k=1}^{T-1} \mathbf{a}_{ik} = \boldsymbol{\mu}_i - \mathbf{a}_{iT} \end{array} \right. ,$$

$$\left[\begin{array}{l} \frac{T-1}{T} \mathbf{y}_i = \frac{T-1}{T} \mathbf{a}_{iT} \\ \frac{T_1}{T} \mathbf{y}_i = \frac{2}{T} \sum_{k=1}^{T-1} \mathbf{a}_{ik} - \frac{T-1}{T} \mathbf{a}_{iT} \end{array} \right. ,$$

$$\left[\begin{array}{l} \mathbf{y}_i = \mathbf{a}_{iT} \\ \mathbf{y}_i = \frac{2}{T-1} \sum_{k=1}^{T-1} \mathbf{a}_{ik} - \mathbf{a}_{iT} \end{array} \right. .$$

Redefining, we obtain

$$\hat{\mathbf{y}}_i = \mathbf{y}_i,$$

$$\hat{\mathbf{y}}_i = \frac{2}{T-1} \sum_{k=1}^{T-1} \mathbf{a}_{ik} - \mathbf{y}_i.$$

∎

Consequence. (A trivial method for obtaining a pair of possible answers.) This theorem shows that using pairwise correlation as a distance function gives at most *two* different answers when recovering. Moreover, having obtained one, one can explicitly find the second one. Then, to find both possible answers, it is proposed to apply any non-convex optimisation method to find at least one of the minimum of the function. Therefore with the formula above we are able to find another minimum.
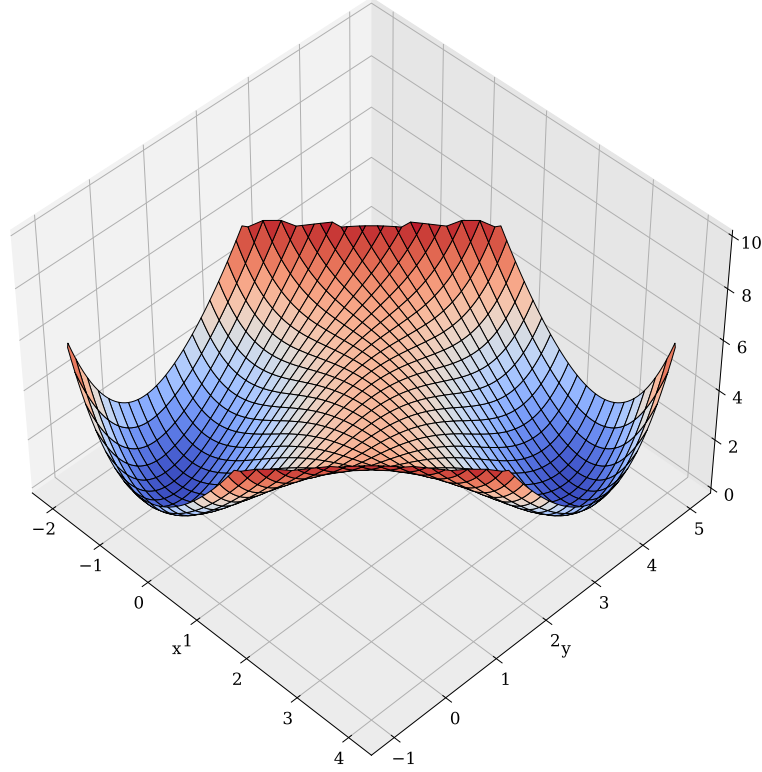
Figure 2: Function $||\hat{\mathbf{\Sigma}}_{t+1} - \bar{\mathbf{\Sigma}}_{t+1}||_2^2$ for following series: $(1, 3)$ and $(2, 4)$. Minumums: $(3; 4)$ is desired and $(-1; 0)$ is alternative.

The problem with this method is the computational cost of non-convex optimisation methods. As an alternative, we propose the following method using only singular matrix decomposition.

Theorem 3. (Efficient method for obtaining a pair of possible answers.) *The minimum of the function* $||\hat{\mathbf{\Sigma}}_{t+1} - \bar{\mathbf{\Sigma}}_{t+1}||_2^2$ *is reached on*

$$\pm\sqrt{\lambda_1}\mathbf{u}_1 + \boldsymbol{\mu}_t,$$

where λ_1 *is the first singular value,* $\mathbf{u}_1$ *is the first left singular vector of matrix* $\mathbf{A} = \left(\hat{\mathbf{\Sigma}}_{t+1} - \frac{t}{t+1} \cdot \mathbf{\Sigma}_t\right) \cdot \frac{(t+1)^2}{t}$

Proof. The notation $\mathbf{x}_i$ is used below to denote the *multidimensional* value of the time series at time i. The proof expresses $\mathbf{\Sigma}_{t+1}$ through $\mathbf{\Sigma}_t$. After that, the operator norm and rank property of the matrix is used. All expressions below are true for arbitrary $\boldsymbol{\mu}_T$ and $\mathbf{\Sigma}_T$ constructed by the definition above.

1. Express $\boldsymbol{\mu}$ through the values of the time series:

$$\boldsymbol{\mu}_t = \frac{1}{t}\sum_{i=1}^{t}\mathbf{x}_i \Rightarrow \sum_{i=1}^{t}\mathbf{x}_i = t\boldsymbol{\mu}_t.$$

6

2. Similarly, express $\boldsymbol{\Sigma_t}$ through the values of the series:

$$\boldsymbol{\Sigma}_t = \frac{1}{t}\sum_{i=1}^{t}(\mathbf{x}_i - \boldsymbol{\mu}_t)(\mathbf{x}_i - \boldsymbol{\mu}_t)^{\mathsf{T}};$$

$$\boldsymbol{\Sigma}_t = \frac{1}{t}\sum_{i=1}^{t}\mathbf{x}_i\mathbf{x}_i^{\mathsf{T}} - \frac{1}{t}\left(\sum_{i=1}^{t}\mathbf{x}_i\right)\boldsymbol{\mu}_t^{\mathsf{T}} - \boldsymbol{\mu}_t\frac{1}{t}\left(\sum_{i=1}^{t}\mathbf{x}_i^{\mathsf{T}}\right) + \boldsymbol{\mu}_t\boldsymbol{\mu}_t^{\mathsf{T}} = \frac{1}{t}\sum_{i=1}^{t}\mathbf{x}_i\mathbf{x}_i^{\mathsf{T}} - \boldsymbol{\mu}_t\boldsymbol{\mu}_t^{\mathsf{T}} \Rightarrow$$

$$\sum_{i=1}^{t}\mathbf{x}_i\mathbf{x}_i^{\mathsf{T}} = t\boldsymbol{\Sigma}_t + t\boldsymbol{\mu}_t\boldsymbol{\mu}_t^{\mathsf{T}}.$$

3. Express $\boldsymbol{\Sigma_{t+1}}$ through $\boldsymbol{\Sigma_t}$ using the previous expressions:

$$\boldsymbol{\Sigma}_{t+1} = \frac{1}{t+1}\left(\sum_{i=1}^{t}\mathbf{x}_i\mathbf{x}_i^{\mathsf{T}} + \mathbf{x}_{t+1}\mathbf{x}_{t+1}^{\mathsf{T}}\right) - \boldsymbol{\mu}_{t+1}\boldsymbol{\mu}_{t+1}^{\mathsf{T}} =$$

$$= \frac{t}{t+1}\boldsymbol{\Sigma}_t + \frac{t}{t+1}\boldsymbol{\mu}_t\boldsymbol{\mu}_t^{\mathsf{T}} + \frac{1}{t+1}\mathbf{x}_{t+1}\mathbf{x}_{t+1}^{\mathsf{T}} - \frac{1}{(t+1)^2}(t\boldsymbol{\mu}_t + \mathbf{x}_{t+1})(t\boldsymbol{\mu}_t + \mathbf{x}_{t+1})^{\mathsf{T}} =$$

$$= \frac{t}{t+1}\boldsymbol{\Sigma}_t + \frac{t}{t+1}\boldsymbol{\mu}_t\boldsymbol{\mu}_t^{\mathsf{T}} + \frac{t}{(t+1)^2}\mathbf{x}_{t+1}\mathbf{x}_{t+1}^{\mathsf{T}} - \frac{t^2}{(t+1)^2}\boldsymbol{\mu}_t\boldsymbol{\mu}_t^{\mathsf{T}} - \frac{t}{(t+1)^2}\boldsymbol{\mu}_t\mathbf{x}_{t+1}^{\mathsf{T}} - \frac{t}{(t+1)^2}\mathbf{x}_{t+1}\boldsymbol{\mu}_t^{\mathsf{T}} =$$

$$= \frac{t}{t+1}\boldsymbol{\Sigma}_t + \frac{t}{t+1}\boldsymbol{\mu}_t\boldsymbol{\mu}_t^{\mathsf{T}} - \frac{t}{(t+1)^2}\left(-\mathbf{x}_{t+1}\mathbf{x}_{t+1}^{\mathsf{T}} + t\boldsymbol{\mu}_t\boldsymbol{\mu}_t^{\mathsf{T}} + \boldsymbol{\mu}_t\mathbf{x}_{t+1}^{\mathsf{T}} + \mathbf{x}_{t+1}\boldsymbol{\mu}_t^{\mathsf{T}}\right) =$$

$$= \frac{t}{t+1}\boldsymbol{\Sigma}_t + \frac{t}{t+1}\boldsymbol{\mu}_t\boldsymbol{\mu}_t^{\mathsf{T}} - \frac{t}{(t+1)^2}\left(-(\mathbf{x}_{t+1} - \boldsymbol{\mu}_t)(\mathbf{x}_{t+1} - \boldsymbol{\mu}_t)^{\mathsf{T}} + (t+1)\boldsymbol{\mu}_t\boldsymbol{\mu}_t^{\mathsf{T}}\right) =$$

$$= \frac{t}{t+1}\boldsymbol{\Sigma}_t + \frac{t}{t+1}\boldsymbol{\mu}_t\boldsymbol{\mu}_t^{\mathsf{T}} - \frac{t(t+1)}{(t+1)^2}\boldsymbol{\mu}_t\boldsymbol{\mu}_t^{\mathsf{T}} + \frac{t}{(t+1)^2}(\mathbf{x}_{t+1} - \boldsymbol{\mu}_t)(\mathbf{x}_{t+1} - \boldsymbol{\mu}_t)^{\mathsf{T}} =$$

$$= \frac{t}{t+1}\boldsymbol{\Sigma}_t + \frac{t}{(t+1)^2}(\mathbf{x}_{t+1} - \boldsymbol{\mu}_t)(\mathbf{x}_{t+1} - \boldsymbol{\mu}_t)^{\mathsf{T}}.$$

This equality expresses $\boldsymbol{\Sigma}_{t+1}$ through $\boldsymbol{\Sigma}_t$. For further proof it is useful to derive the following equality for our problem:

$$(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^{\mathsf{T}} = \left(\bar{\boldsymbol{\Sigma}}_{t+1} - \frac{t}{t+1}\cdot\boldsymbol{\Sigma}_t\right)\cdot\frac{(t+1)^2}{t}.$$

4. We solve the problem of finding the minimum of the function $||\hat{\boldsymbol{\Sigma}}_{t+1} - \bar{\boldsymbol{\Sigma}}_{t+1}||_2^2$. In our case, this is analogous to the equality of this function to zero. Let us write the expression under the norm:

$$\hat{\boldsymbol{\Sigma}}_{t+1} - \bar{\boldsymbol{\Sigma}}_{t+1} = \frac{t}{t+1}\boldsymbol{\Sigma}_t + \frac{t}{(t+1)^2}(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^{\mathsf{T}} - \frac{t}{t+1}\boldsymbol{\Sigma}_t + \frac{t}{(t+1)^2}(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^{\mathsf{T}} =$$

$$= \frac{t}{(t+1)^2}(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^{\mathsf{T}} - \frac{t}{(t+1)^2}(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^{\mathsf{T}}.$$

Then, the problem above is equivalent to finding the minimum of the function

$$||(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^{\mathsf{T}} - (\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^{\mathsf{T}}||_2^2.$$

Denote

$$\mathbf{A} = (\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^{\mathsf{T}} = \left(\bar{\boldsymbol{\Sigma}}_{t+1} - \frac{t}{t+1}\cdot\boldsymbol{\Sigma}_t\right)\cdot\frac{(t+1)^2}{t}.$$

The rank of the matrix $(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^{\mathsf{T}}$ is 1, and since the desired minimum is 0, it turns out that the rank of the matrix $\mathbf{A}$ is also 1.

5. From the previous paragraph, the matrix $\mathbf{A}$ has rank 1. Let us write the singular value decomposition.

$$\mathbf{A} = \sum_{i=1}^{1}\lambda_i\mathbf{u}_i\mathbf{v}_i^{\mathsf{T}} = \lambda_1\mathbf{u}_1\mathbf{v}_1^{\mathsf{T}}.$$

On the other hand, $\mathbf{A} = (\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^{\intercal}$. Then,

$$\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t = \pm\sqrt{\lambda_1}\mathbf{u}_1 \Leftrightarrow \bar{\mathbf{x}}_{t+1} = \pm\sqrt{\lambda_1}\mathbf{u}_1 + \boldsymbol{\mu}_t.$$

The sign $\pm$ comes from the fact that in the case of a symmetric matrix, there are two singular value decompositions: $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\intercal} = (-\mathbf{U})\boldsymbol{\Sigma}(-\mathbf{V})^{\intercal}$.

■

This theorem allows us to find both minimums of a function much faster than with standard non-convex optimisation methods [7].

6   Algorithm for recovering time series values in case of accurate prediction of the matrix

Theorems 2 and 3 show that using a single pairwise correlation matrix and information about the first t moments of time allows us to obtain a *pair* of possible values after recovery. In this section, we propose a method to select the true value from the obtained *pair* $\boldsymbol{\Sigma}_{t+1}$ *predicted accurately*.

The algorithm described below is based on the use of *two* predicted matrices corresponding to different intervals of time. Two different values T, T' are chosen. Two matrices are predicted:

First $\boldsymbol{\Sigma}_{t+1}^1$ pairwise correlation matrix for the multivariate time series $\mathbf{x}$ at time moments from $t - T + 2$ to $t + 1$ (in total T values).

Second $\boldsymbol{\Sigma}_{t+1}^2$ pairwise correlation matrix for the multivariate time series $\mathbf{x}$ at time moments from $t - T' + 2$ to $t + 1$ (in total T' values).

Hence, when we reconstruct answers from these matrices, we obtain two pairs of answers, each of which is a candidate for the true answer. At the same time, a true answer exists in each of the pairs. It is suggested to take the answer from the intersection. We does not consider the case when the intersection size is 2, since the probability of this situation is 0 when using continuous values.

Algorithm scheme:

1. Take T and $T' : T \neq T'$.

2. For T and T' perform the above algorithm and obtain the answer sets:

$$[\hat{\mathbf{x}}_{t+1}^1, \hat{\mathbf{x}}'^1_{t+1}], [\hat{\mathbf{x}}_{t+1}^2, \hat{\mathbf{x}}'^2_{t+1}].$$

3. Find the answer that lies in the intersection. In real data, the probability of matching sets of answers is 0, just as it is in synthetic data with random noise added.
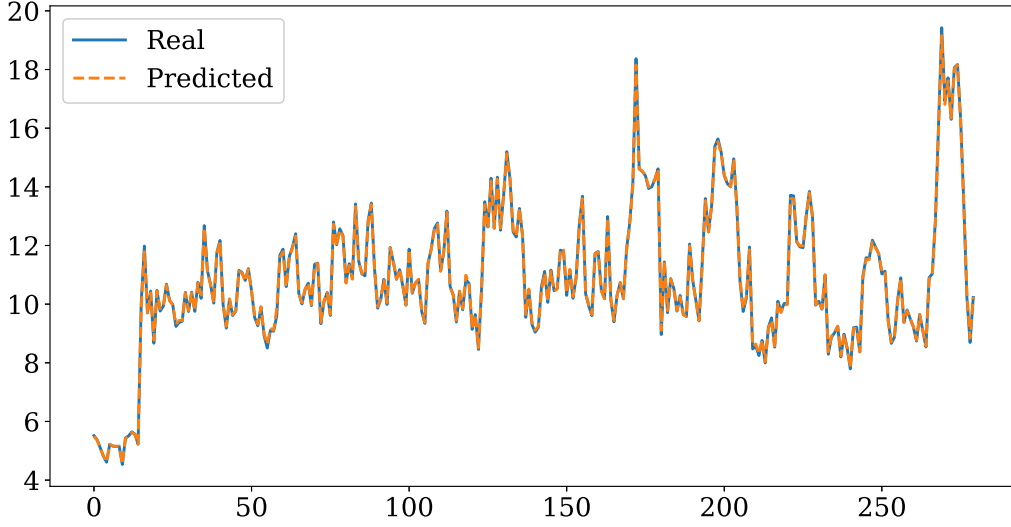
Figure 3: Prediction recovery in case of accurate prediction of correlation matrix $\boldsymbol{\Sigma}$. $T = 20$, $T' = 10$

7 Algorithm for recovering time series values in case of inaccurate matrix prediction

The problem with the above algorithm is that if the prediction is inaccurate, there may be no intersection. This happens because the error in each of the predicted matrices is different. For this, the following algorithm is proposed to amortise the error:

Instead of two values of T and T', K values are taken. Next, each matrix is reduced to the nearest positive semi-definite matrix. Then we get K sets of answers:

$$[\hat{\mathbf{x}}_{t+1}^{1}, \hat{\mathbf{x}}_{t+1}'^{1}],$$
$$[\hat{\mathbf{x}}_{t+1}^{2}, \hat{\mathbf{x}}_{t+1}'^{2}],$$
$$\vdots$$
$$[\hat{\mathbf{x}}_{t+1}^{K}, \hat{\mathbf{x}}_{t+1}'^{K}].$$

Then we propose to search through 2^K sets of answers and choose the set in which the diameter is minimal. The diameter of set is calculated as maximum Euclidean distance between two different points in set. It is a *necessary* condition for the actual answer. In other words, the one that is the intersection of all pairs of answers. In the case of an accurate prediction, the diameter of such a set will always be zero.

The asymptotic complexity of this recovery will be $O(2^K \times K \times N)$ + the complexity of the minimum search algorithm used.
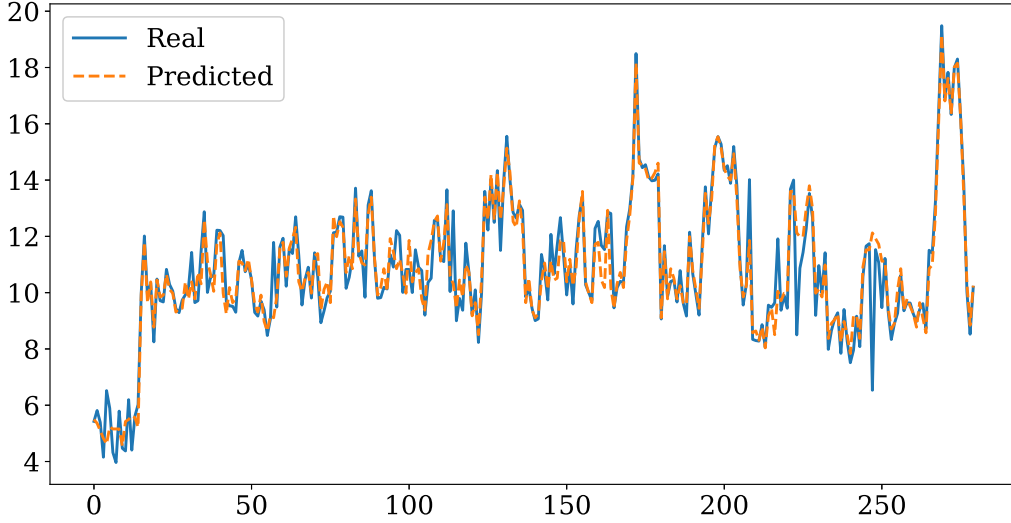
Figure 4: Prediction recovery in case of inaccurate correlation matrix $\mathbf{\Sigma}$ prediction. In addition to the prediction error caused by noise in the correlation matrix a new type of error is added. When selecting a set, it is possible that the diameter is minimised not at the right set, as this is only a necessary condition, but not a sufficient one.

8   Experiment

In this section we test the algorithm when the pairwise correlation matrix prediction from the previous section is inaccurate. Experiments are performed on synthetic data as well as on Electricity Transformer Temperature [10] data. Different values of K are tested, as well as different added noise to the true matrix values.

Synthetic data.

The table below shows the error values after recovery under different conditions. Generated data consisting of a combination of noisy sines and cosines are used.

Table 1: Loss on synthetic data

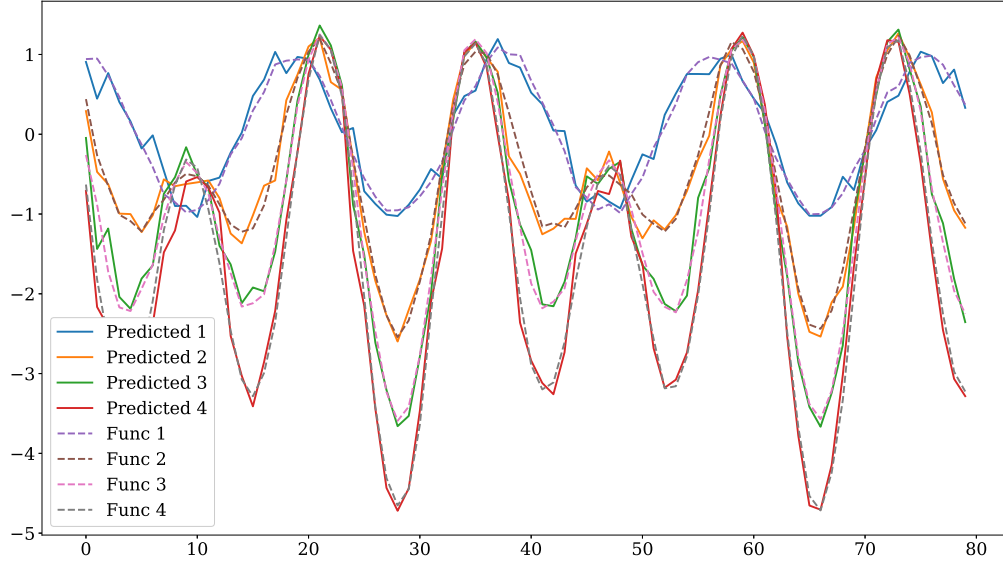| Parameters Noise	Metric	$K = 2$	$K = 4$	$K = 10$
$\mathcal{N}(0, 0.01)$	MAE:	0.070	0.052	0.040
	MSE:	0.010	0.005	0.002
$\mathcal{N}(0, 0.05)$	MAE:	0.316	0.176	0.116
	MSE:	0.295	0.060	0.025
$\mathcal{N}(0, 0.1)$	MAE:	0.530	0.398	0.230
	MSE:	0.635	0.348	0.111

Figure 5: Synthetic data recovery plot at $K = 10$, Additional noise $\mathcal{N}(0, 0.05)$. **MAE: 0.116, MSE: 0.025**

Electricity Transformer Temperature

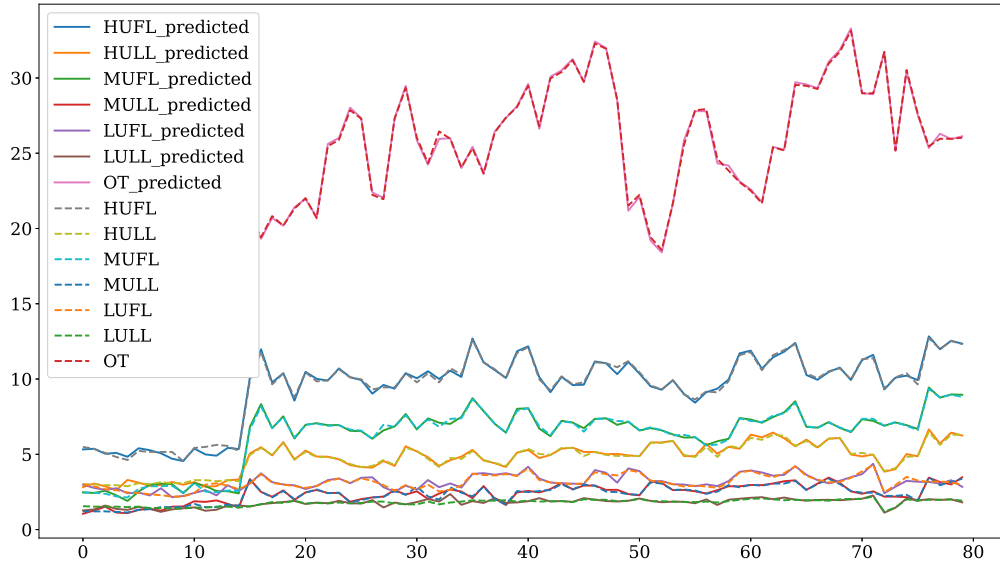The following is a similar table for the Electricity Transformer Temperature dataset.

Figure 6: ETTh1 data recovery plot at $K = 10$, Additional Noise $\mathcal{N}(0, 0.05)$. **MAE: 0.096, MSE: 0.019**

Table 2: Loss on ETTh1 data

Parameters Noise	Metric	$K = 2$	$K = 4$	$K = 10$
$\mathcal{N}(0, 0.01)$	MAE: MSE:	0.071 0.030	0.047 0.004	0.038 0.003
$\mathcal{N}(0, 0.05)$	MAE: MSE:	0.240 0.198	0.153 0.053	0.096 0.019
$\mathcal{N}(0, 0.1)$	MAE: MSE:	0.466 0.719	0.306 0.281	0.217 0.148

In all cases, using more different values of T predictably gave the highest accuracy. However, for $K > 15$ the algorithm becomes too computationally complex due to the exponential dependence on K.

9   Conclusion

The paper investigates an approach to time series forecasting using a pairwise correlation matrix between series. It is shown that the use of only one matrix leads to the existence of a pair of possible answers — the values of the series at the next moment of time. An explicit formula for computing one answer through the other is derived, which allows the problem to be solved using nonconvex optimisation. Moreover, an explicit form of the pair of answers via the singular value decomposition of the pairwise correlation matrix is derived. Two algorithms are proposed to identify the desired answer from a pair of possible answers. The first one relies on the exact prediction of the pairwise correlation matrix. The second one admits the presence of an error in the prediction, but it is more computationally demanding.

The future development of the study is to find a way to predict the pairwise correlation matrix with high accuracy. Using basic regression models give insufficiently accurate results. In such a prediction, errors are often made by incorrectly selecting the set of answers from Algorithm 2.

Also, the side of development can be the estimation of the error radius when recovering the value of time series from the matrix of pair correlations. In addition, it makes sense to consider other functions of pairwise distance as a metric.

References

[1] Stephen Boyd, Enzo Busseti, Steven Diamond, Ronald N. Kahn, Kwangmoo Koh, Peter Nystrup, and Jan Speth. Multi-period trading via convex optimization, 2017.

[2] Mark Davison and Stephen Sireci. Multidimensional scaling. 10 2012.

[3] James B. Elsner and Anastasios A. Tsonis. Singular spectrum analysis: A new tool in time series analysis. 1996.

[4] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9:1735–80, 12 1997.

[5] Roman Isachenko and Vadim Strijov. Quadratic programming feature selection for multicorrelated signal decoding with partial least squares. *Expert Systems with Applications*, 207:117967, 11 2022.

[6] Aryan Jadon, Avinash Patil, and Shruti Jadon. A comprehensive survey of regression based loss functions for time series forecasting, 2022.

[7] V. S. Mikhalevich, A. M. Gupal, and V. I. Norkin. Methods of nonconvex optimization, 2024.

[8] Anastasia Motrenko and Vadim Strijov. Extracting fundamental periods to segment biomedical signals. *IEEE journal of biomedical and health informatics*, 20, 08 2015.

[9] Nikita Puchkin, Fedor Noskov, and Vladimir Spokoiny. Sharper dimension-free bounds on the frobenius distance between sample covariance and its expectation, 2023.

[10] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting, 2021.