

Correlation-based Algorithm for multivariate time series prediction

Maxim Divilkovskiy*, Konstantin Yakovlev†, Vadim Strijov‡

October 28, 2024

Abstract

The paper investigates a time series prediction problem. It constructs a pointwise prediction model for a set of time series. These time series have high variance and high covariance. Paper introduces space of pairwise distances between time series and analyses its properties. In this space, the pairwise distance matrix is predicted in its dynamics. The time series values are reconstructed at the next time moment with this matrix. The authors propose several methods for the pointwise prediction of time series space using the reconstruction reconstructed space of the pairwise distance matrices. We prove the existence of multiple time series values satisfying the same pairwise distance matrix. It presents two algorithms based on the use of matrices constructed over different time intervals using pairwise correlation. The paper derives an explicit view of the reconstructed values through the pairwise correlation matrix. Also, it derives 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, Time Series Forecasting, Singular Value Decomposition, Pearson Correlation Coefficient

*Forecsys, mdivilkovskij@gmail.com

†Forecsys, bicst808@gmail.com

‡Forecsys, strijov@gmail.com

1 Introduction

In this research the authors present a new method for pointwise time series prediction. Time series have high pairwise covariance and high variance. Pointwise prediction consists of calculating the values of the time series at the next point in time using the available historical data for the previous several points in time. A set of d time series, each of which consists of t time moments forms a multivariate time series. We call the multivariate time series space isomorphic to $\mathbb{R}^{d \times t}$ the source time series space. The prediction problem divides into three stages. First, transform the source space of the time series into a metric space by constructing a matrix of pairwise distances. Second, do the prediction of the pairwise distances matrix at the next moment of time in the metric space. Third, use the predicted matrix to reconstruct the result into the source space. In the theoretical part of the paper, we prove the necessity of predicting at least two matrices. We show that the uniqueness of the reconstructed result in the source space requires the use of several matrices corresponding to different time intervals. We focus mostly on the last step. We present experiments for the case of accurate matrix prediction and for the matrix prediction with the addition of normal noise to the values.

Existing time series prediction methods such as LSTM [1], SSA [2] and other [3, 4] predict the value of a univariate series. These methods can be modified to predict also a set of time series. For this purpose, it is sufficient to present a set of series as one multivariate series. This approach does not explicitly model the dependencies between different series. In contrast, we propose to analyze the change in *set* of time series. Our 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 problem. It consists in selecting such a subset of the original time series for which it is possible to make a predict of sufficient quality.

Recent studies [6, 7, 8, 9] use popular transformer-based models. Transformer models were originally proposed for natural language processing problems, such as translation and text-completion [10]. However, since many language problems deal with text as a sequence in time, the same approaches may be used for time series predicting. The Crossformer model [11] uses cross-dimensional dependency. However, it does not explicitly model the distance function between time series.

Further, we study conditions on the distance function between rows under which there is a way to reconstruct the values of the time series. We prove the insufficiency of one matrix to reconstruct the answer. We propose two methods for the value prediction using several matrices for the case of accurate prediction and for the case of prediction with non-zero noise. Also, we propose a reconstruction algorithm based on two theorems about the explicit formula of the result in time series space 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 [12] that they are the most suitable for the task of time series prediction.

We show that Multidimensional Scaling Algorithm [13] and Metric Multidimensional Scaling Algorithm [14] cannot uniquely reconstruct the correlation matrix to the source time series space. This algorithm is often used to reconstruct objects from their pairwise distances.

In the experiments section we measure losses of our Algorithm on synthetic dataset and Electricity Transformer Temperature dataset. We show that as the count of matrices used increases, the loss decreases.

2 Point-wise prediction of a set of time series

The time series with t moments in time is a set of t elements:

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

where $\mathbf{x}_{t_i,k} \in \mathbb{R}$ is the value of the series with index k at the time moment t_i . One has to predict \mathbf{x}_{t+1} , which is the value of time series at the time moment $t + 1$. The matrix \mathbf{x} of size $t \times d$ is treated as a *multivariate* time series, treating the value at a point as an element of space \mathbb{R}^d .

The prediction errors are calculated with two functions,

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

3 Reconstruction of time series by a predicted distance matrix

In this section we formulate general algorithm of reconstruction the time series values in the moment $t + 1$ from information about time moments $1, 2, \dots, t$. Novelty of our algorithm consists in usage of *pairwise distance matrices*.

Construct distance matrices between a set of time series values according to the scheme below:

$$\begin{aligned} [\mathbf{x}_1, \dots, \mathbf{x}_s] &\rightarrow \Sigma_s, \\ [\mathbf{x}_2, \dots, \mathbf{x}_{s+1}] &\rightarrow \Sigma_{s+1}, \\ &\vdots \\ [\mathbf{x}_{t-s}, \dots, \mathbf{x}_t] &\rightarrow \Sigma_t. \end{aligned}$$

Each Σ_i is a matrix of pairwise distances of size $d \times d$. An element of the matrix $\Sigma_{i,j}$ is the value of the distance between rows i and j . We describe the construction formula as well as the choice of the distance function in the following sections.

Predict the matrix $\hat{\Sigma}_{t+1}$ from matrices $\Sigma_s, \Sigma_{s+1} \dots, \Sigma_t$.

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

$$||\hat{\Sigma}_{t+1} - \bar{\Sigma}_{t+1}||_2^2 \rightarrow \min_{\hat{\mathbf{x}}_{t+1}}, \quad (1)$$

where $\bar{\Sigma}_{t+1}$ is the distance matrix, constructed from the set $[\mathbf{x}_{t-s+1}, \dots, \mathbf{x}_t, \hat{\mathbf{x}}_{t+1}]$. Achieving a minimum by this function would mean that $\hat{\Sigma}_{t+1}$ are $\bar{\Sigma}_{t+1}$ 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 (1) is the central problem we consider in this paper. In this section we show that using one matrix constructed by

an arbitrary metric it is possible to reconstruct 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.

We now describe the reconstruction of the prediction from the matrix Σ_{t+1} to source time series space, where Σ_{t+1} is the matrix of pairwise distances corresponding to the multivariate series $\mathbf{x} = [\mathbf{x}_1, \dots, \mathbf{x}_{t+1}]$.

Given a pairwise distance matrix Σ 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

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

defined on time series, satisfies the properties of the metric.

$$\Sigma_{i,j} = \rho((\mathbf{x}_{1\dots t,i}, \hat{\mathbf{x}}_{t+1,i}), (\mathbf{x}_{1\dots t,j}, \hat{\mathbf{x}}_{t+1,j})).$$

One of the fundamental metrics for calculating distance of objects in \mathbb{R} is the Euclidean metric. With this example, we show that there can be several reconstructions with the same distance matrices. In the case of the Euclidean metric:

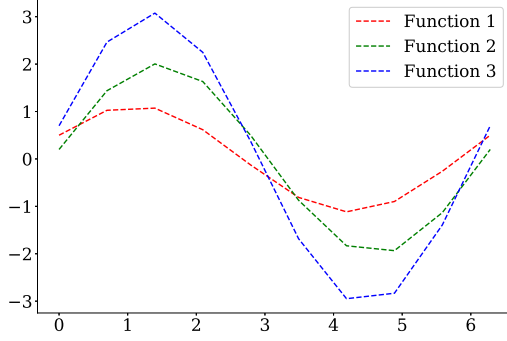
$$\Sigma_{i,j} = \rho(\mathbf{x}_{1\dots t,i} \circ \hat{\mathbf{x}}_{t+1,i}, \mathbf{x}_{1\dots t,j} \circ \hat{\mathbf{x}}_{t+1,j}) = \sqrt{\left(\sum_{k=1}^t (\mathbf{x}_{k,i} - \mathbf{x}_{k,j})^2\right) + (\hat{\mathbf{x}}_{t+1,i} - \hat{\mathbf{x}}_{t+1,j})^2},$$

the number of answers is infinite, which we show below. 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:

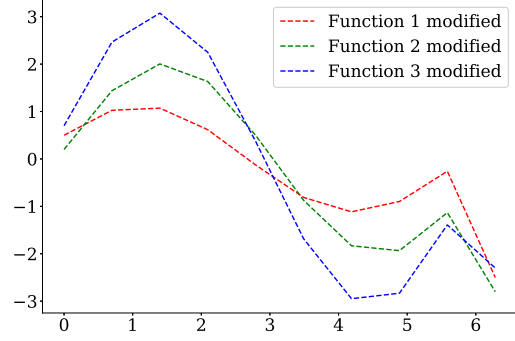
$$\begin{aligned} & \sqrt{\left(\sum_{k=1}^t (\mathbf{x}_{k,i} - \mathbf{x}_{k,j})^2\right) + (\hat{\mathbf{x}}_{t+1,i} - \hat{\mathbf{x}}_{t+1,j})^2} = \\ &= \sqrt{\left(\sum_{k=1}^t (\mathbf{x}_{k,i} - \mathbf{x}_{k,j})^2\right) + [(\hat{\mathbf{x}}_{t+1,i} + C) - (\hat{\mathbf{x}}_{t+1,j} + C)]^2}.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

In the case of the time series prediction problem, this property is critical because even if the matrix Σ is predicted accurately, there are infinitely many time series values at time $t + 1$ corresponding to this matrix. See Figure 1 for the example of this effect.

However, even using other metrics does not get rid of the problem. This results in the inability to use Multidimensional Scaling algorithm to reconstruct the result to the source



(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.

time series space. This algorithm is often used to reconstruct objects from their pairwise distances. Metric MDS Algorithm may be used if distances are not euclidean.

In the theorem below we show that usage of only one matrix is insufficient for the problem of reconstructing 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 reconstruct 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. We return to the information about the previous moments of time 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 series whose distance between them is the same.

Show that the metric is a continuous function. Take the sequence

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

Then,

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

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

$$\rho(\mathbf{x}_n, \mathbf{y}_n) \leq \rho(\mathbf{x}_n, \mathbf{x}) + \rho(\mathbf{x}, \mathbf{y}) + \rho(\mathbf{y}_n, \mathbf{y}) \rightarrow \rho(\mathbf{x}, \mathbf{y}),$$

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

Therefore the metric is a continuous mapping from $\mathbb{R}^t \times \mathbb{R}^t$ to \mathbb{R} . We will show that such mapping cannot be a homeomorphism. Assume that $f : \mathbb{R} \rightarrow \mathbb{R}^t \times \mathbb{R}^t$ 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 Σ_{t+1} using the value of the time series before the time moment t : $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_t]$. The problem is reformulated as follows. There are n objects in \mathbb{R}^{t+1} , their first t coordinates are known. The distance matrix $\Sigma_{t+1} \in \mathbb{R}^{(t+1) \times (t+1)}$. One has to calculate the $(t+1)$ -th coordinate of each of the objects. In time series terms, the $(t+1)$ -th coordinate is the value of each of the series at that moment in time.

5 Pairwise correlation between time series

In this section we investigate the time series values reconstruction using the pairwise correlation matrix. This distance function is used because it is shown in the article [15] that the pairwise correlation estimate of a sample approximates its mathematical expectation.

The pairwise distance matrix is constructed as follows:

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

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

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

$$\begin{aligned}\hat{\mathbf{y}}_i &= \mathbf{y}_i, \\ \hat{\mathbf{y}}_i &= \frac{2}{T-1} \sum_{k=1}^{T-1} \mathbf{x}_{ki} - \mathbf{y}_i,\end{aligned}$$

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

Proof. Let us denote by Σ the true matrix at the moment of time T , and $\hat{\Sigma}$ the predicted one. $\boldsymbol{\mu} = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t$. By construction, $\Sigma = \frac{1}{T} \sum_{k=1}^T (\mathbf{x}_k - \boldsymbol{\mu}_T)(\mathbf{x}_k - \boldsymbol{\mu}_T)^\top$. Then, consider what the elements of the matrices Σ and $\hat{\Sigma}$ are equal to.

$$\Sigma_{ij} = \frac{1}{T} \sum_{k=1}^T (\mathbf{x}_{ki} - \boldsymbol{\mu}_i)(\mathbf{x}_{kj} - \boldsymbol{\mu}_j),$$

separating last term,

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

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: $\Sigma_{ii} = \hat{\Sigma}_{ii}$,

$$\sum_{k=1}^T (\mathbf{x}_{ki} - \boldsymbol{\mu}_i)^2 = \sum_{k=1}^{T-1} (\mathbf{x}_{ki} - \hat{\boldsymbol{\mu}}_i)^2 + (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^2,$$

since $\boldsymbol{\mu}_i$ and $\hat{\boldsymbol{\mu}}_i$ are equal up to $(T-1)$ -th coordinate,

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

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

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

$$\boldsymbol{\mu}_i = \frac{1}{T} \sum_{k=1}^T \mathbf{x}_{ki}.$$

Substitute the expressions for $\hat{\boldsymbol{\mu}}$ into $(\mathbf{x}_{Ti} - \boldsymbol{\mu}_i)^2 = (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^2$:

$$\begin{cases} \frac{T-1}{T} \mathbf{y}_i - \frac{1}{T} \sum_{k=1}^{T-1} \mathbf{x}_{ki} = \mathbf{x}_{Ti} - \boldsymbol{\mu}_i \\ \frac{T-1}{T} \mathbf{y}_i - \frac{1}{T} \sum_{k=1}^{T-1} \mathbf{x}_{ki} = \boldsymbol{\mu}_i - \mathbf{x}_{Ti} \end{cases},$$

Rewriting $\boldsymbol{\mu}$,

$$\begin{cases} \frac{T-1}{T} \mathbf{y}_i = \frac{T-1}{T} \mathbf{x}_{Ti} \\ \frac{T-1}{T} \mathbf{y}_i = \frac{2}{T} \sum_{k=1}^{T-1} \mathbf{x}_{ki} - \frac{T-1}{T} \mathbf{x}_{Ti} \end{cases},$$

$$\begin{cases} \mathbf{y}_i = \mathbf{x}_{Ti} \\ \mathbf{y}_i = \frac{2}{T-1} \sum_{k=1}^{T-1} \mathbf{x}_{ki} - \mathbf{x}_{Ti} \end{cases}.$$

Redefining, we obtain

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

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

■

Figure 2 shows an example of a function $\|\hat{\Sigma}_{t+1} - \bar{\Sigma}_{t+1}\|_2^2$ with two minimum for certain time series.

Corollary. (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 reconstructing. Moreover, having obtained one, we 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.

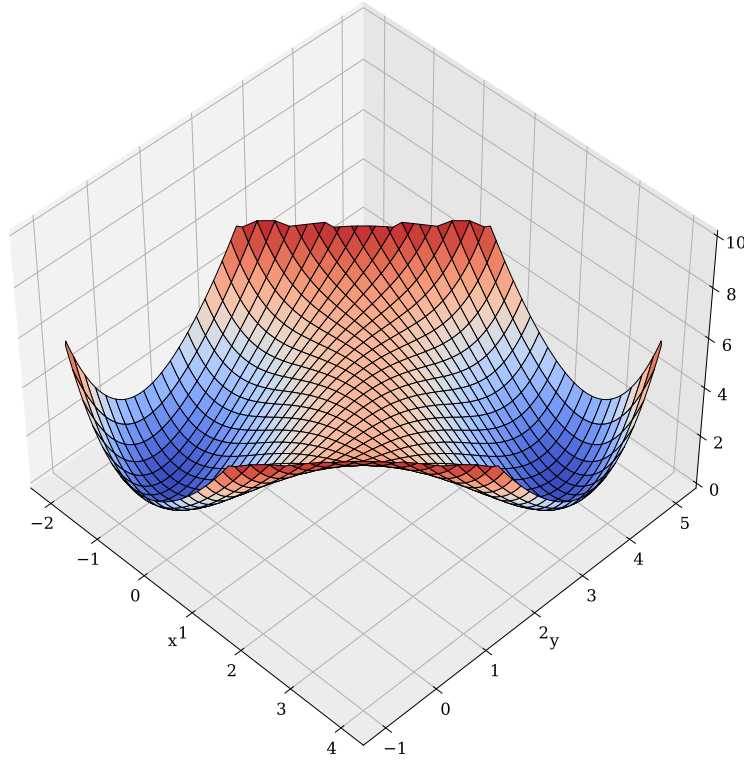


Figure 2: Function $\|\hat{\Sigma}_{t+1} - \bar{\Sigma}_{t+1}\|_2^2$ for following series: (1, 3) and (2, 4). Minumums: (3; 4) is desired and (-1; 0) is alternative.

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 $\|\hat{\Sigma}_{t+1} - \bar{\Sigma}_{t+1}\|_2^2$ 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 $\mathbf{A} = \left(\hat{\Sigma}_{t+1} - \frac{t}{t+1} \cdot \Sigma_t \right) \cdot \frac{(t+1)^2}{t}$

Proof. The notation \mathbf{x}_i is used below to denote the *multivariate* value of the time series at time i . The proof expresses Σ_{t+1} through Σ_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 Σ_T constructed by the definition above.

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

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

2. Similarly, express Σ_t through the values of the series:

$$\begin{aligned} \Sigma_t &= \frac{1}{t} \sum_{i=1}^t (\mathbf{x}_i - \boldsymbol{\mu}_t)(\mathbf{x}_i - \boldsymbol{\mu}_t)^\top; \\ \Sigma_t &= \frac{1}{t} \sum_{i=1}^t \mathbf{x}_i \mathbf{x}_i^\top - \frac{1}{t} \left(\sum_{i=1}^t \mathbf{x}_i \right) \boldsymbol{\mu}_t^\top - \boldsymbol{\mu}_t \frac{1}{t} \left(\sum_{i=1}^t \mathbf{x}_i^\top \right) + \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top = \\ &= \frac{1}{t} \sum_{i=1}^t \mathbf{x}_i \mathbf{x}_i^\top - \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top \Rightarrow \\ &\sum_{i=1}^t \mathbf{x}_i \mathbf{x}_i^\top = t\Sigma_t + t\boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top. \end{aligned}$$

3. Express Σ_{t+1} with Σ_t using the previous expressions:

$$\begin{aligned}
\Sigma_{t+1} &= \frac{1}{t+1} \left(\sum_{i=1}^t \mathbf{x}_i \mathbf{x}_i^\top + \mathbf{x}_{t+1} \mathbf{x}_{t+1}^\top \right) - \boldsymbol{\mu}_{t+1} \boldsymbol{\mu}_{t+1}^\top = \\
&= \frac{t}{t+1} \Sigma_t + \frac{t}{t+1} \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top + \frac{1}{t+1} \mathbf{x}_{t+1} \mathbf{x}_{t+1}^\top - \frac{1}{(t+1)^2} (t \boldsymbol{\mu}_t + \mathbf{x}_{t+1})(t \boldsymbol{\mu}_t + \mathbf{x}_{t+1})^\top = \\
&= \frac{t}{t+1} \Sigma_t + \frac{t}{t+1} \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top + \frac{t}{(t+1)^2} \mathbf{x}_{t+1} \mathbf{x}_{t+1}^\top - \\
&\quad - \frac{t^2}{(t+1)^2} \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top - \frac{t}{(t+1)^2} \boldsymbol{\mu}_t \mathbf{x}_{t+1}^\top - \frac{t}{(t+1)^2} \mathbf{x}_{t+1} \boldsymbol{\mu}_t^\top = \\
&= \frac{t}{t+1} \Sigma_t + \frac{t}{t+1} \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top - \frac{t}{(t+1)^2} (-\mathbf{x}_{t+1} \mathbf{x}_{t+1}^\top + t \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top + \boldsymbol{\mu}_t \mathbf{x}_{t+1}^\top + \mathbf{x}_{t+1} \boldsymbol{\mu}_t^\top) = \\
&= \frac{t}{t+1} \Sigma_t + \frac{t}{t+1} \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top - \frac{t}{(t+1)^2} (-(\mathbf{x}_{t+1} - \boldsymbol{\mu}_t)(\mathbf{x}_{t+1} - \boldsymbol{\mu}_t)^\top + (t+1) \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top) = \\
&= \frac{t}{t+1} \Sigma_t + \frac{t}{t+1} \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top - \frac{t(t+1)}{(t+1)^2} \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top + \frac{t}{(t+1)^2} (\mathbf{x}_{t+1} - \boldsymbol{\mu}_t)(\mathbf{x}_{t+1} - \boldsymbol{\mu}_t)^\top = \\
&= \frac{t}{t+1} \Sigma_t + \frac{t}{(t+1)^2} (\mathbf{x}_{t+1} - \boldsymbol{\mu}_t)(\mathbf{x}_{t+1} - \boldsymbol{\mu}_t)^\top.
\end{aligned}$$

This equality expresses Σ_{t+1} through Σ_t . For further proof it is useful to derive the following equality for our problem:

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

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

$$\begin{aligned}
\hat{\Sigma}_{t+1} - \bar{\Sigma}_{t+1} &= \frac{t}{t+1} \Sigma_t + \frac{t}{(t+1)^2} (\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^\top - \\
&\quad - \frac{t}{t+1} \Sigma_t + \frac{t}{(t+1)^2} (\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^\top = \\
&= \frac{t}{(t+1)^2} (\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^\top - \frac{t}{(t+1)^2} (\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^\top.
\end{aligned}$$

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

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

Denote

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

The rank of the matrix $(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^\top$ 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. Consider the singular value decomposition.

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

On the other hand, $\mathbf{A} = (\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^\top$. 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: $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top = (-\mathbf{U})\boldsymbol{\Sigma}(-\mathbf{V})^\top$.

■

This theorem allows us to find both minimums of a function much faster than with standard non-convex optimisation methods [16].

6 Algorithm for reconstructing 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 subsegments of time. Two different values T, T' are chosen. Two matrices are predicted:

First matrix $\boldsymbol{\Sigma}_{t+1}^1$ 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 matrix $\boldsymbol{\Sigma}_{t+1}^2$ 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. We suggest 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:

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

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.

3 shows the example of reconstructed time series with accurate predicted matrix. As you can see in the graph, the reconstruction is lossless.

7 Algorithm for reconstructing 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' , we propose take K values. We get K matrices with some noise that came from inaccuracy in prediction. Thus, each matrix is reduced to the nearest positive semi-definite matrix. Algorithm is explained in [17]. Then, for each value algorithm for obtaining the pair of possible answers is applied.

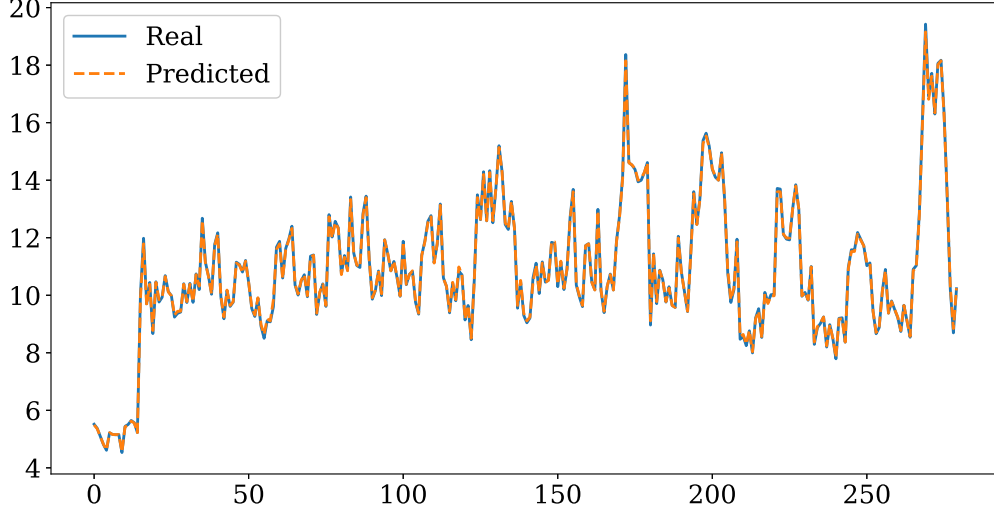


Figure 3: Prediction recovery in case of accurate prediction of correlation matrix Σ . $T = 20$, $T' = 10$

We get K sets of answers:

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

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 the case of an accurate prediction, the diameter of such a set will always be zero.

The asymptotic complexity of this reconstruction will be $O(2^K \times K \times N)$ + the complexity of the minimum search algorithm used.

On 4 you can see an example of reconstructing time series values when the predicted matrix contains some random noise. The reconstruction is no more lossless.

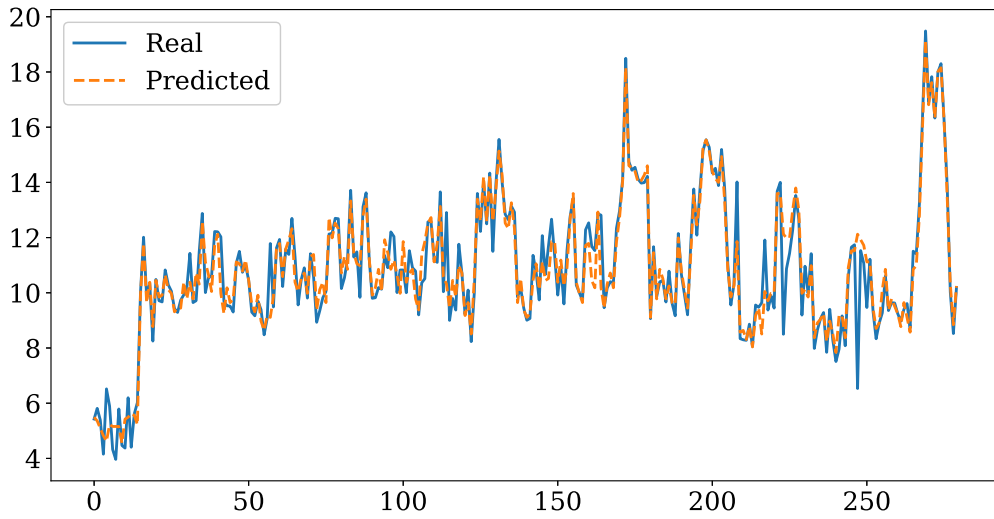


Figure 4: Prediction reconstruction in case of inaccurate correlation matrix Σ prediction. In addition to the prediction error caused by noise in the correlation matrix a new type of error is added. When selecting a set, it is possible that the diameter is minimised not at the right set, as this is only 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 [6] 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 reconstruction the time series values under different conditions. Generated data consisting of a combination of noisy sines and cosines is used.

Table 1: Loss on synthetic data. As expected, the error is less on bigger K value. See 5 for the example of reconstruction with $K = 10$ and noise $\mathcal{N}(0, 0.05)$.

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

Electricity Transformer Temperature

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

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 .

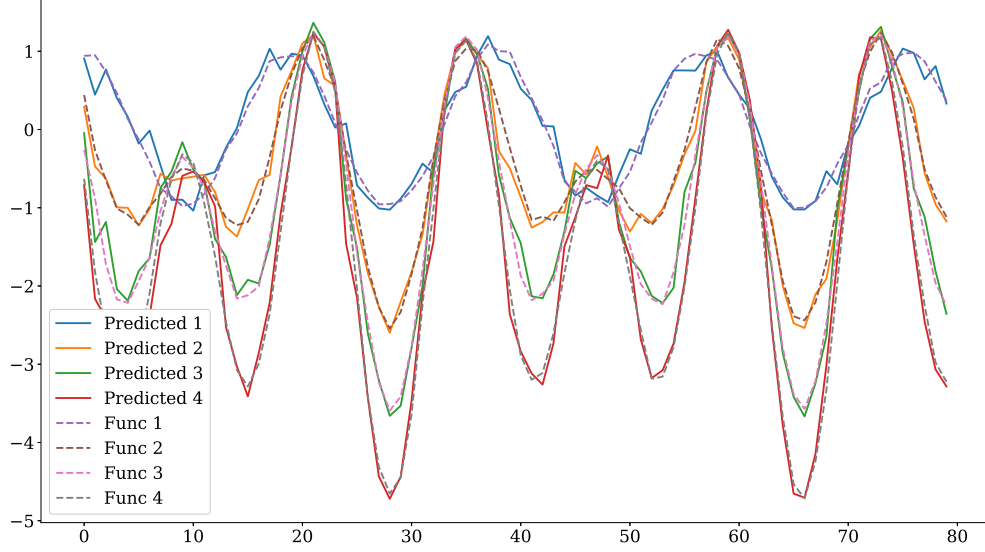


Figure 5: Synthetic data reconstruction plot at $K = 10$, Additional noise $\mathcal{N}(0, 0.05)$. **MAE: 0.116, MSE: 0.025**

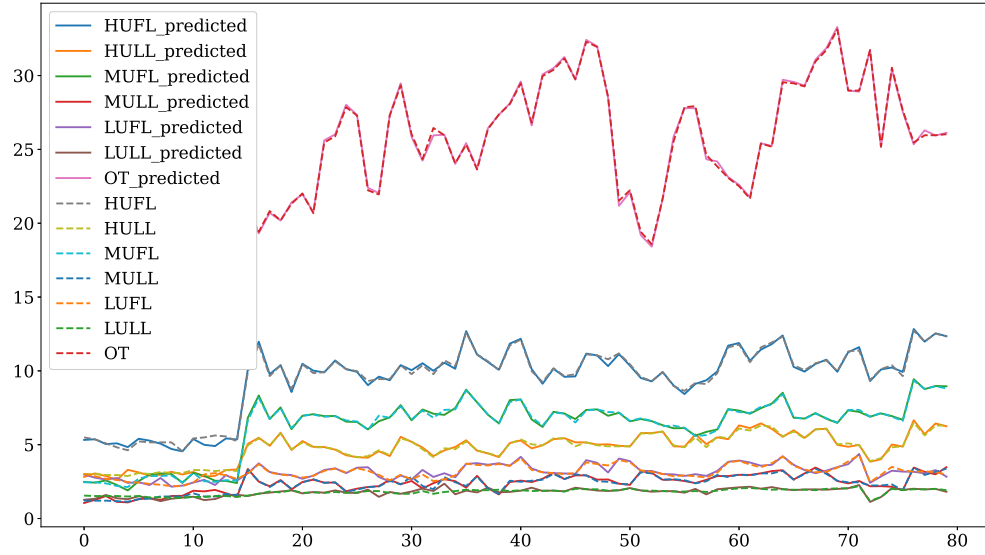


Figure 6: ETTh1 data reconstruction plot at $K = 10$, Additional Noise $\mathcal{N}(0, 0.05)$. **MAE: 0.096, MSE: 0.019**

Table 2: Loss on ETTh1 data. The same dependence of the error on the value of K as on the synthetic data can be seen. See 6 for the example of reconstruction with $K = 10$ and noise $\mathcal{N}(0, 0.05)$.

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

9 Conclusion

The paper investigates an approach to time series prediction 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 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 non-convex 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 reconstructing 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] Sepp Hochreiter and Jürgen Schmidhuber. “Long Short-term Memory”. In: *Neural computation* 9 (Dec. 1997), pp. 1735–80. DOI: 10.1162/neco.1997.9.8.1735.
- [2] Nina Golyandina and Anatoly Zhigljavsky. *Singular Spectrum Analysis for Time Series*. Jan. 2013. ISBN: 978-3-642-34913-3. DOI: 10.1007/978-3-642-34913-3.
- [3] Anastasia Motrenko and Vadim Strijov. “Extracting Fundamental Periods to Segment Biomedical Signals”. In: *IEEE journal of biomedical and health informatics* 20 (Aug. 2015). DOI: 10.1109/JBHI.2015.2466440.
- [4] Stephen Boyd et al. *Multi-Period Trading via Convex Optimization*. 2017. DOI: 10.1561/24000000023.
- [5] Roman Isachenko and Vadim Strijov. “Quadratic programming feature selection for multicorrelated signal decoding with partial least squares”. In: *Expert Systems with Applications* 207 (Nov. 2022), p. 117967. DOI: 10.1016/j.eswa.2022.117967.
- [6] Haoyi Zhou et al. “Informer: ond Efficient Transformer for Long Sequence Time-Series Forecasting”. In: *The Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Virtual Conference*. Vol. 35. 12. AAAI Press, 2021, pp. 11106–11115. DOI: 10.1609/aaai.v35i12.17325.
- [7] Haoyi Zhou et al. “Expanding the prediction capacity in long sequence time-series forecasting”. In: *Artificial Intelligence* 318 (2023), p. 103886. ISSN: 0004-3702.
- [8] Haixu Wu et al. *Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting*. 2022. DOI: 10.5555/3540261.3541978.
- [9] Shizhan Liu et al. “Pyraformer: Low-Complexity Pyramidal Attention for Long-Range Time Series Modeling and Forecasting”. In: *International Conference on Learning Representations*. 2022. DOI: 10.34726/2945.
- [10] Ashish Vaswani et al. “Attention Is All You Need”. In: (June 2017). DOI: 10.48550/arXiv.1706.03762.

- [11] Yunhao Zhang and Junchi Yan. “Crossformer: Transformer Utilizing Cross-Dimension Dependency for Multivariate Time Series Forecasting”. In: *The Eleventh International Conference on Learning Representations*. 2023. URL: <https://openreview.net/forum?id=vSVLM2j9eie>.
- [12] Aryan Jadon, Avinash Patil, and Shruti Jadon. *A Comprehensive Survey of Regression Based Loss Functions for Time Series Forecasting*. 2022. DOI: 10.1007/978-981-97-3245-6_9.
- [13] Mark Davison and Stephen Sireci. “Multidimensional Scaling”. In: (Oct. 2012). DOI: 10.1016/B978-012691360-6/50013-6.
- [14] Nickolay Trendafilov and Michele Gallo. “Metric multidimensional scaling (MDS) and related methods”. In: Sept. 2021, pp. 325–371. ISBN: 978-3-030-76973-4. DOI: 10.1007/978-3-030-76974-1_10.
- [15] Nikita Puchkin, Fedor Noskov, and Vladimir Spokoiny. *Sharper dimension-free bounds on the Frobenius distance between sample covariance and its expectation*. 2023. DOI: 10.48550/arXiv.2308.14739.
- [16] V. S. Mikhalevich, A. M. Gupal, and V. I. Norkin. *Methods of Nonconvex Optimization*. 2024. DOI: 10.48550/arXiv.2406.10406.
- [17] Nicholas J. Higham. “Computing a nearest symmetric positive semidefinite matrix”. In: *Linear Algebra and its Applications* 103 (1988), pp. 103–118. ISSN: 0024-3795. DOI: [https://doi.org/10.1016/0024-3795\(88\)90223-6](https://doi.org/10.1016/0024-3795(88)90223-6).