International Journal of Forecasting 34 (2018) 582–597

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Structured low-rank matrix completion for forecasting in time series analysis

Jonathan Gillard [a], Konstantin Usevich [b,*]

[a] Cardiff School of Mathematics, Cardiff University, Senghennydd Road, Cardiff, CF24 4AG, UK
[b] Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France

ARTICLE INFO

Keywords:
Hankel matrices
Low-rank matrix completion
Forecasting
Nuclear norm

ABSTRACT

This paper considers the low-rank matrix completion problem, with a specific application to forecasting in time series analysis. Briefly, the low-rank matrix completion problem is the problem of imputing missing values of a matrix under a rank constraint. We consider a matrix completion problem for Hankel matrices and a convex relaxation based on the nuclear norm. Based on new theoretical results and a number of numerical and real examples, we investigate the cases in which the proposed approach can work. Our results highlight the importance of choosing a proper weighting scheme for the known observations.

1. Introduction

Dokumentov and Hyndman (2014) showed low-rank representations and approximations to be very useful tools for time series forecasting. One of the popular approaches is singular spectrum analysis (SSA) forecasting (Golyandina, Nekrutkin, & Zhigljavsky, 2001), which embeds the time series into a Hankel matrix and uses a low-rank approximation and continuation to compute the next values of a time series. SSA uses the fact that many time series can be approximated well by a class of so-called time series of finite rank. However, despite many successful examples (Hassani, Heravi, & Zhigljavsky, 2009; Khan & Poskitt, 2017; Papailias & Thomakos, 2017), SSA forecasting has a number of disadvantages.

This paper develops a method based on Hankel matrix completion. We follow the approach of Butcher and Gillard (2017), who proposed that a time series be embedded into a Hankel matrix, with the missing data (to be forecasted) being stored in the bottom right-hand corner of this matrix. The method of Butcher and Gillard (2017) is based on minimising the nuclear norm, which provides a convex relaxation of a low-rank matrix completion problem that is non-convex and NP-hard in general (see for example Fazel, 2002; Gillis & Glineur, 2011).

The nuclear norm (the sum of singular values) is a popular convex surrogate for the rank (Fazel, 2002), and is similar to using the ℓ_1-norm for sparse approximation (Rish & Grabarnik, 2014). It has been shown to be a successful tool for imputing the missing values of a matrix (see for example Candes & Plan, 2010; Candès & Recht, 2009; Chen & Chi, 2014; Fazel, 2002; Recht, Fazel, & Parrilo, 2010). Nuclear norm relaxation has been a very popular tool for spectral estimation (Chen & Chi, 2014), recommender systems (Fazel, 2002), and system identification (Blomberg, 2016; Liu & Vandenberghe, 2009; Verhaegen & Hansson, 2016). One advantage of the nuclear norm relaxation considered in this paper is the ability to build more complex models to represent potentially more complex behaviors in the observed time series.

An important question is when the convex relaxation solves the original low-rank matrix completion problem. A lot of famous research has been conducted on this topic, but most of the available research (Candes & Plan, 2010; Candès & Recht, 2009; Chen & Chi, 2014; Fazel, 2002; Recht et al., 2010) has assumed that the position of the missing

* Corresponding author.
E-mail addresses: GillardJW@Cardiff.ac.uk (J. Gillard), konstantin.usevich@univ-lorraine.fr (K. Usevich).

https://doi.org/10.1016/j.ijforecast.2018.03.008
0169-2070/

entries in the matrix is random, and often that the known entries are also random; in general, unstructured matrices are considered. Thus, these results do not apply, due to the special arrangement of missing data in our problem and to the Hankel structure. Moreover, the case of structured matrices is much more challenging, as was noted by Markovsky (2012a, b).

There are few available results for the completion of Hankel matrices with a fixed pattern of missing values. Dai and Pelckmans (2015) analysed a special case, square real-valued Hankel matrices with nearly half of their values missing, and showed that the nuclear norm relaxation gives the correct rank-one completion only when the embedded time series can be written as a sum of decreasing exponentials. Usevich and Comon (2016) then extended this analysis to the rank-r case for the same pattern of missing values.

This paper makes several contributions. First, like Butcher and Gillard (2017), we consider the general case of rectangular Hankel matrices with potentially fewer missing values. We show that, when there are only a few missing values, the convex relaxation of the low-rank matrix completion using the nuclear norm will give identical solutions without using the convex relaxation for time series with undamped or exponentially increasing periodic components, and establish bounds on the number of missing values. We also study the question of choosing the optimal shape of the Hankel matrix (parameterized by the so-called window length). Second, we suggest a new (relative to Butcher & Gillard, 2017) formulation of the low-rank matrix completion problem for Hankel matrices, which allows the possibility of allocating different weights to past observations. In particular, exponential weighting is designed to overcome the problems related to the performance of the nuclear norm for time series that can be expressed as a sum of increasing exponentials.

Empirical comparisons show that, with the proper choice of weights, our novel formulation performs well relative to a number of classical techniques. For the numerical examples in this paper, we use CVX, a MATLAB package for specifying and solving convex programs (Grant & Boyd, 2008,2014). The reproducible examples are hosted at https://github.com/kdu/nucnorm-forecasting/.

This paper has the following structure. Section 2 formally defines the problems to be considered. We start by defining exact matrix completion, then consider an approximate version. This section also describes the settings used throughout the paper. Some known theoretical results that must be stated are reviewed in Section 3. First, the time series of finite rank are recalled and the solution of the exact minimal rank completion is summarized. Next, known results on time series of finite rank are recalled. Section 4 contains the main results of the paper. First, we give theoretical bounds for matrix completion in the case of an arbitrary shape of the matrix and the number of missing values. We check the tightness of our bounds through numerical experiments. Second, we establish the connection between exponential weighting and the preprocessing of time series. Finally, the forecasting examples that involve real and model time series and that demonstrate the advantages of the proposed methodology are provided in Section 5.

2. Problem statement

2.1. Hankel matrices

For a vector $\mathbf{f} = (f_1, \dots, f_n)$ with $n > 1$ and a so-called window length L, the $L \times (n-L+1)$ Hankel matrix is defined as

$$\mathscr{H}_L(\mathbf{f}) = \begin{pmatrix} f_1 & f_2 & \cdots & f_{n-L+1} \\ f_2 & f_3 & \iddots & f_{n-L+2} \\ \vdots & \iddots & \iddots & \vdots \\ f_L & f_{L+1} & \cdots & f_n \end{pmatrix}.$$

In what follows, we are going to pose the problem of forecasting a given time series as the low-rank matrix completion of a Hankel matrix. Formally, let

$$\mathbf{p} = (p_1, p_2, \dots, p_{n+m}) \tag{1}$$

be a vector of length $(n+m)$, with $m \geq 0$. In what follows, m will be the number of observations to be forecast, and n will be the length of the time series that we wish to forecast. We use the notation $\mathbf{p}_{(1:n)} = (p_1, p_2, \dots, p_n)$ for the first n elements of $\mathbf{p}$. Next, let L and K be integers such that $L + K - 1 = m + n$. Then, the matrix structure $\mathcal{S}(\mathbf{p})$ (parameterized by $\mathbf{p}$) that we consider is

$$\mathcal{S}(\mathbf{p}) = \mathscr{H}_L(\mathbf{p}) = \begin{pmatrix} p_1 & p_2 & \cdots & \cdots & \cdots & p_K \\ p_2 & p_3 & \cdots & \iddots & \cdots & \vdots \\ \vdots & \cdots & \iddots & \iddots & \cdots & p_n \\ \vdots & \iddots & \iddots & \iddots & \cdots & p_{n+1} \\ \vdots & \iddots & \iddots & \iddots & \iddots & \vdots \\ p_L & \cdots & p_n & p_{n+1} & \cdots & p_{n+m} \end{pmatrix}. \tag{2}$$

In Eq. (2), the grey-shaded values are "known" and the others are "missing".

The Hankel matrix structure belongs to the class of affine matrix structures (Markovsky, 2012b, Section 3.3) having the form:

$$\mathcal{S}(\mathbf{p}) = \mathbf{S}_0 + \sum_{i=1}^{(n+m)} p_i \mathbf{S}_i, \tag{3}$$

where $\mathbf{S}_i, i \in \{0, 1, \dots, (n+m)\}$, are given linearly independent basis matrices, and in particular, for the Hankel matrix structure in Eq. (2), the basis matrices in Eq. (3) are given as $\mathbf{S}_0 = 0$,

$$\mathbf{S}_1 = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \iddots & 0 & 0 \\ \vdots & \iddots & \iddots & \iddots & \vdots \\ 0 & 0 & \iddots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \mathbf{S}_2 = \begin{pmatrix} 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \iddots & 0 & 0 \\ \vdots & \iddots & \iddots & \iddots & \vdots \\ 0 & 0 & \iddots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \dots,$$

$$\mathbf{S}_{n+m-1} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \iddots & 0 & 0 \\ \vdots & \iddots & \iddots & \iddots & \vdots \\ 0 & 0 & \iddots & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \mathbf{S}_{n+m} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \iddots & 0 & 0 \\ \vdots & \iddots & \iddots & \iddots & \vdots \\ 0 & 0 & \iddots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$