Hankel low-rank approximation and completion in time series analysis and forecasting: a brief review

Jonathan Gillard* and Konstantin Usevich

In this paper we offer a review and bibliography of work on Hankel low-rank approximation and completion, with particular emphasis on how this methodology can be used for time series analysis and forecasting. We begin by describing possible formulations of the problem and offer commentary on related topics and challenges in obtaining globally optimal solutions. Key theorems are provided, and the paper closes with some expository examples.

1. INTRODUCTION

Low-rank techniques have become one of the great success stories in computational mathematics, linear algebra, engineering and statistics over the last decade. They now constitute the dominant paradigm for solving large-scale problems from computational fluid dynamics, through computational meteorology, to molecular modelling. They are also used in inverse problems, where restricting the rank acts as a kind of regularisation to make ill-posed problems stable and tractable. Arguably one of the most exciting developments in statistics in recent years has been the exploitation of (approximate) low-rank structure of observed data. This is the idea that underpins the success of compressed sensing and matrix completion.

In statistics, approximating a given matrix by one of small rank, is closely related to principal component analysis and variations of it [103], factor analysis [106] and total least squares [87]. Such low-rank approximations are now essential for the application of kernel methods to large-scale learning problems [71]. Their motivation and need is well described in the paper [109], where it is asserted that matrices of low-rank are pervasive in data science and that any sufficiently large matrix with small spectral norm can be well approximated by a low-rank matrix.

There has been much recent work in the application and development of low-rank approximation methods to solve typical problems of time series analysis, namely those of

*Corresponding author.

modelling and forecasting. Generally speaking the main idea of the methodology is to embed an observed time series into a Hankel matrix, the rank of which is taken to be a measure of the complexity of the time series. There are rich and historical results connecting properties of a Hankel matrix and its embedded original time series which directly mean that finding low-rank approximations of the Hankel matrix has sensible meaning in yielding a time-series which approximates the original one, but is of smaller complexity.

In this paper we offer a review and bibliography of work on Hankel low-rank approximation and completion, with particular emphasis on how this methodology can be used for time series analysis and forecasting. The structure of the paper is as follows. Section 2 introduces the problem of Hankel low-rank approximation, describing possible choices of norm to define the optimization problem, as well its possible formulation via a vector or matrix form. Unstructured low-rank approximation and its solution via the classic Eckart-Young-Mirsky-Schmidt theorem is described as well as related problems. The important relationship between low-rank Hankel matrices and linear recurrence sequences is provided in Section 3. Algorithms to approximate a solution of the Hankel low-rank approximation problem are described in Section 4, and the problem of forecasting is framed as one of low-rank matrix completion in Section 5 where use of the nuclear norm as a convex relaxation of the rank constraint is explained. The paper closes with some numerical examples in Section 6.

2. HANKEL LOW-RANK APPROXIMATION

2.1 Problem statement

Let

$$\mathbf{p} = (p_1, p_2, \ldots, p_N), \tag{1}$$

be a time series of length $N > 1$, taking real or complex values. We mostly focus on real-valued time series in this paper since they more readily appear in statistical application.

Most low-rank based methods of time series analysis embed the vector $\mathbf{p}$ into a Hankel matrix. Let L, K be integers such that $L + K - 1 = N$. For simplicity of notation we will assume $L \leq K$ throughout. The $L \times K$ Hankel matrix $\mathcal{S}(\mathbf{p})$

parameterized by the vector $\mathbf{p}$ is defined as

(2)
$$\mathcal{S}(\mathbf{p}) = \mathcal{H}_L(\mathbf{p}) = \begin{pmatrix} p_1 & p_2 & p_3 & \cdots & \cdots & p_K \\ p_2 & p_3 & \cdot^{\cdot^{\cdot}} & \cdots & \cdots & \vdots \\ p_3 & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \vdots \\ \vdots & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & p_{N-1} \\ p_L & \cdots & \cdots & \cdots & p_{N-1} & p_N \end{pmatrix}.$$

The parameter L is specified by the user, and it referred to as the window length in techniques such as singular spectrum analysis (SSA) [51]. Its choice depends on the low-rank approximation (LRA) problem under consideration. The most recent book on singular spectrum analysis is [51] and advice on the choice of L for SSA is given in [66] for this technique, with some more general discussion available in [46]. Commonly L is chosen so that the resulting Hankel matrix is (close to) square [83]. Another possible strategy is to choose the window length as small as possible in order to make the matrix "fat".

Let the given vector of N observations (time series) be

$$\mathbf{p}_0 = (p_{0,1}, p_{0,2}, \ldots, p_{0,N})$$

and $\mathbf{X}_0 = S(\mathbf{p}_0)$ be the associated Hankel structured matrix. Low-rank based methods of approximating/forecasting a given time series aim to obtain a so-called structured low-rank approximation (SLRA) of the matrix $\mathbf{X}_0$. The problem of SLRA in the literature is typically formulated in two ways: vector and matrix form, which informally speaking refers to the domain in which the SLRA optimization problem is defined. In fact, these formulations can be shown to be equivalent, but both of them are used in the literature. We begin our paper with the vector form, which seems to be more suited to time series analysis.

Vector (parameter) form For a given time series $\mathbf{p}_0$ and fixed rank r, the SLRA problem is to find the closest time series $\widehat{\mathbf{p}}$, such that the corresponding structured matrix $\mathcal{S}(\widehat{\mathbf{p}})$ is low-rank:

(3)
$$\min_{\widehat{\mathbf{p}}} \|\widehat{\mathbf{p}} - \mathbf{p}_0\| \quad \text{subject to} \quad \operatorname{rank} \mathcal{S}(\widehat{\mathbf{p}}) \leq r,$$

where $\| \cdot \|$ is a vector norm (or, in general, an extended semi-norm). The SLRA problem thus can be interpreted as an approximation of a time series by a simpler time series having some structure. The class of time series corresponding to low-rank Hankel matrices is described in Section 3.

The norm $\|\cdot\|$ measures the distance between the time series and its approximation. Typical choices of norm include:

- (Euclidean) 2-norm $\| \cdot \|_2$, which puts equal weights to all observations;
- elementwise weighted extended semi-norm

$$\|\mathbf{p}\|_{\mathbf{w}} = \sqrt{\sum_k w_k p_k^2},$$

defined by a nonnegative vector of weights $w_k \in [0, +\infty]$; more precisely,

- $w_k \in (0, +\infty)$ quantifies importance of observations;
- $w_k = \infty$ corresponds to fixed values (i.e., an additional constraint $\widehat{p}_k = p_{0,k}$ in (3));
- $w_k = 0$ implies that the value $p_{0,k}$ is not important, and therefore can encode missing values;

in particular, if all $w_k \in (0, +\infty)$, then $\|\mathbf{p}\|_{\mathbf{w}}$ is a norm.
- General weighted norm

$$\|p\|_{\mathbf{W}} := \sqrt{p^\top \mathbf{W} p},$$

where $\mathbf{W}$ is a symmetric positive semidefinite matrix. In particular,

- the choice $\mathbf{W} = \operatorname{diag}(w_1, \ldots, w_N)$ corresponds to the elementwise weighted extended norm ($w_k \in (0, +\infty)$);
- a general matrix $\mathbf{W}$ can be taken as an inverse covariance matrix of the noise (if an additive noise model is assumed).

Other types of norms can be considered (see, e.g., examples in the matrix formulation described next), but the SLRA problem becomes much more difficult in that case.

Matrix form It is the matrix form that was used in many of the original papers and classical algorithms on SLRA [24]. For a given matrix $\mathbf{X}_0 = \mathcal{S}(\mathbf{p}_0) \in \mathscr{S}$ belonging to the set of structured matrices $\mathscr{S} = \{\mathcal{S}(\mathbf{p}) \,|\, \mathbf{p} \in \mathbb{R}^N\}$ and a chosen rank r we seek its best rank-r structured approximation, i.e.

(4)
$$\min_{\widehat{\mathbf{X}}} \|\widehat{\mathbf{X}} - \mathbf{X}_0\| \quad \text{subject to} \quad \operatorname{rank} \widehat{\mathbf{X}} \leq r \text{ and } \widehat{\mathbf{X}} \in \mathscr{S},$$

where $\| \cdot \|$ is some matrix norm or, in general, an extended semi-norm. Examples of common choices of norms are:

- the classic Frobenius norm

$$\|\mathbf{X}\|_F = \sqrt{\sum_{i,j} X_{ij}^2};$$

- the elementwise weighted Frobenius norm

$$\|\mathbf{X}\|_{\mathbf{Z}} = \sqrt{\sum_{i,j} Z_{ij} X_{ij}^2},$$

where $\mathbf{Z}$ is a matrix of nonnegative weights encoding the importance of the elements in the approximation. As in the vector formulation described above, the weights can take the values ∞ and 0, as to encode fixed or missing values respectively, which makes $\| \cdot \|_{\mathbf{Z}}$ an (extended) semi-norm.
- so-called (Q,R)-norm [43]:

$$\|\mathbf{X}\|_{\mathbf{Q},\mathbf{R}} = \sqrt{\operatorname{Trace}(\mathbf{Q} \mathbf{X} \mathbf{R} \mathbf{X}^T)},$$

where $\mathbf{Q}$ and $\mathbf{R}$ are positive definite matrices.

- Generalized weighted Frobenius norm, defined by a symmetric positive definite matrix $\mathbf{W}_m \in \mathbb{R}^{LK \times LK}$,

$$(5) \qquad \|\mathbf{X}\|_{\mathbf{W}_m} = \sqrt{(\text{vec}\,\mathbf{X})^T \mathbf{W}_m \,\text{vec}\,\mathbf{X}}.$$

In fact, the general weighted Frobenius norm generalizes the previous two norms, with

- $\mathbf{W}_m = \text{diag}(\text{vec}\,\mathbf{Z})$ for the elementwise weighted norm;
- $\mathbf{W}_m = \mathbf{R} \otimes \mathbf{Q}$ for the (Q,R)-norm, with $\otimes$ being the Kronecker product;
- in particular, a norm that is both an elementwise and a (Q,R)-norm corresponds to a rank-one matrix $\mathbf{Z}$ and diagonal matrices $\mathbf{Q}$ and $\mathbf{R}$:

$$\mathbf{Z} = \mathbf{q}\mathbf{r}^T, \quad \mathbf{Q} = \text{diag}(\mathbf{q}) \quad \text{and} \quad \mathbf{R} = \text{diag}(\mathbf{r}).$$

- Other norms can be considered, e.g., the spectral norm, but this results in a much more difficult problem and very few results and practical algorithms are available [68, 8, 101, 38]. The spectral norm is the dual of the nuclear norm [33], and so is often studied when considering properties of nuclear norm minimization. The spectral norm underpins the work of Adamyan, Arov and Krein which has come to be known as AAK theory [2]: a series of results connecting the theory of complex functions on the unit circle to infinite Hankel matrices. Optimal approximations of Hankel matrices in the spectral norm is gaining traction in developing modern model reduction methods necessary for recurrent neural networks [9].

General remarks

- For the Frobenius norm, unstructured LRA (without the structure constraint) is solved by the singular value decomposition (see the following subsection).
- The SLRA problem is formulated exactly in the same manner as in for general (affine) matrix structures. In this paper, however, we focus on the Hankel SLRA, but discuss related structures in Section 3.6.
- In the general case, the SLRA problem is a difficult NP-hard global optimization problem [44]. For Hankel SLRA (as shown in [93, Corollary 3.9] for the case of Frobenius norm), the number of stationary points grows polynomially in N and exponentially in r. The minimizers possess narrow regions of attraction and the effect of observing noisy observations further compounds the difficulty, dampening and shifting the location of the global minimum [41].
- The two formulations (matrix and vector), in general, are equivalent in the case of the general weighted semi-norm [63]. We review this correspondence in Section 2.4.

- All definitions and problem formulations given thus far are naturally extended to the complex case, whereupon we replace transpose by the Hermitian transpose, and squares (e.g., p_i^2) by the squares of the absolute values (e.g., $|p_i|^2$) in the definitions of the norms.

2.2 Unstructured low-rank approximation and global solutions

Before discussing properties of SLRA, we recall results on unstructured low-rank approximation problems. The first result is the classic Eckart-Young-Mirsky-Schmidt theorem (see e.g., [45]) which treats the case of unitarily invariant norms, and the Frobenius norm in particular.

Theorem 2.1. *Let $\mathbf{A} \in \mathbb{R}^{L \times K}$ be a given matrix ($L \leq K$) with singular value decomposition (SVD) given by $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$, where $\mathbf{U} \in \mathbb{R}^{L \times L}$, $\mathbf{V} \in \mathbb{R}^{K \times K}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{L \times K}$, with elements $\boldsymbol{\Sigma}_{i,i} = \sigma_i$ being the singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_L \geq 0$. Let $\boldsymbol{\Sigma}^{(r)} \in \mathbb{R}^{L \times K}$ with elements*

$$\boldsymbol{\Sigma}_{i,j}^{(r)} = \begin{cases} \sigma_i, & 1 \leq i \leq r \text{ and } j = i \\ 0, & otherwise. \end{cases}$$

Then

$$\mathbf{A}^{(r)} := \mathbf{U}\boldsymbol{\Sigma}^{(r)}\mathbf{V}^T = \underset{\mathbf{B}:\text{rank}(\mathbf{B}) \leq r}{\arg\min} \|\mathbf{A} - \mathbf{B}\|_\diamond^2,$$

where $\|\cdot\|_\diamond$ is an arbitrary unitarily invariant norm.

Remark 2.1. The best low-rank approximation is not uniquely defined in the case $\sigma_r = \sigma_{r+1}$, which is due to nonuniqueness of the singular vectors corresponding to repeating singular values. In such a case, all the possible SVDs give the set of all best rank-r approximations, according to Theorem 2.1.

An extension of the above theorem to the (Q,R)-norms is as follows.

Lemma 2.1. *Let $\mathbf{A} \in \mathbb{R}^{L \times K}$, $\mathbf{B} \in \mathbb{R}^{L \times K}$, $\mathbf{Q}$ and $\mathbf{R}$ be positive definite matrices of dimension $L \times L$ and $K \times K$ respectively. Then*

$$\|\mathbf{A} - \mathbf{B}\|_{\mathbf{Q},\mathbf{R}}^2 = \|\mathbf{Q}^{1/2}(\mathbf{A} - \mathbf{B})\mathbf{R}^{1/2}\|_F^2.$$

A proof of Lemma 2.1 and a discussion of its importance in relating problems of time-series analysis and low-rank approximation is given in [83, Theorem 4.12] (see also [43] for a detailed study). Note that as the matrices $\mathbf{Q}$ and $\mathbf{R}$ are assumed to be positive definite, then $\mathbf{Q}^{1/2}$ and $\mathbf{R}^{1/2}$ exist, as do their inverses.

Theorem 2.2. *Let $\mathbf{A} \in \mathbb{R}^{L \times K}$ be a given matrix ($L \leq K$) with $\mathbf{Q}$ and $\mathbf{R}$ be given positive definite matrices of dimension $L \times L$ and $K \times K$ respectively, and SVD of $\widetilde{\mathbf{A}} := \mathbf{Q}^{1/2}\mathbf{A}\mathbf{R}^{1/2}$ as described in Theorem 2.1. Then*

$$\mathbf{A}_{\mathbf{Q},\mathbf{R}}^{(r)} := (\mathbf{Q}^{1/2})^{-1}\widetilde{\mathbf{U}}\widetilde{\boldsymbol{\Sigma}}^{(r)}\widetilde{\mathbf{V}}(\mathbf{R}^{1/2})^{-1}$$
$$= \underset{\mathbf{B}:\text{rank}(\mathbf{B}) \leq r}{\arg\min} \|\mathbf{A} - \mathbf{B}\|_{\mathbf{Q},\mathbf{R}}^2$$

These results do not consider structured approximations, in that there are no constraints on the elements of the matrix approximation. However, these results are necessary for SLRA algorithms to obtain either: an initial point to begin iterations of some numerical algorithm and/or to provide projections to the 'closest' matrix of rank $\leq r$. Indeed we will use these results to introduce some algorithms which aim to approximate a solution of (3) in Section 4.

2.3 Related problems

In some literature, instead of low-rank matrix approximation, the problem is formulated as one of rank minimization. For example, in the vector formulation, for a given approximation error ε, we wish to minimize the rank

$$(6) \qquad \min_{\widehat{\mathbf{p}}} \operatorname{rank} \mathcal{S}(\widehat{\mathbf{p}}) \text{ subject to } \|\widehat{\mathbf{p}} - \mathbf{p}_0\| \leq \varepsilon.$$

This formulation is dual to that of SLRA (3), as both problems explore the same Pareto optimality front (see e.g., [80] or [83] for a discussion on this topic). In fact, if we are able to solve one of the problems we can find a solution of the other one. One particular drawback of the formulation in (6) is that the solution is in most cases non-unique, and this is one reason why we prefer the SLRA formulation.

Another possible formulation is the "regularized" form,

$$(7) \qquad \min_{\widehat{\mathbf{p}}} \operatorname{rank} \mathcal{S}(\widehat{\mathbf{p}}) + \lambda \|\widehat{\mathbf{p}} - \mathbf{p}_0\|^2,$$

but this formulation is much less intuitive and makes it difficult to find the regularization parameter. Yet another related problem is trace regression [99] (or matrix sensing), but we leave it out of the scope of this paper.

2.4 Parameterizations and norms

The Hankel matrix structure belongs to the class of affine matrix structures [83], i.e. matrix structures which can be parameterized as:

$$(8) \qquad \mathcal{S}(\mathbf{p}) = \mathbf{S}_0 + \sum_{i=1}^{N} p_i \mathbf{S}_i$$

where $\mathbf{S}_i$, $i \in \{0, 1, \ldots, N\} \in \mathbb{R}^{L \times K}$ are given linearly independent basis matrices. In particular, for the Hankel matrix structure (30), the basis matrices in (8) are given as $\mathbf{S}_0 = 0$,

$$(9) \quad \mathbf{S}_1 = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdot^{\cdot^{\cdot}} & 0 & 0 \\ \vdots & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \vdots \\ 0 & 0 & \cdot^{\cdot^{\cdot}} & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \mathbf{S}_2 = \begin{pmatrix} 0 & 1 & \cdots & 0 & 0 \\ 1 & 0 & \cdot^{\cdot^{\cdot}} & 0 & 0 \\ \vdots & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \vdots \\ 0 & 0 & \cdot^{\cdot^{\cdot}} & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \end{pmatrix}, \ldots,$$
$$\mathbf{S}_{N-1} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdot^{\cdot^{\cdot}} & 0 & 0 \\ \vdots & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \vdots \\ 0 & 0 & \cdot^{\cdot^{\cdot}} & 0 & 1 \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \mathbf{S}_N = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdot^{\cdot^{\cdot}} & 0 & 0 \\ \vdots & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \vdots \\ 0 & 0 & \cdot^{\cdot^{\cdot}} & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}.$$

The optimization problems (3) and (4) are equivalent for particular choices of vector and matrix norm. For example

$$\|\mathcal{S}(\mathbf{p}) - \mathcal{S}(\mathbf{p}_0)\|_F^2 = \sum_{j=1}^{N} \kappa_j (p_j - p_{0,j})^2$$

where the weights

$$(10) \qquad \kappa_j = \begin{cases} j, & \text{if } 1 \leq j < L, \\ L, & \text{if } L \leq j \leq K, \\ N - j + 1, & \text{if } K < j \leq N. \end{cases}$$

This remark has a large implication on the use of SLRA for time series applications. Should the Frobenius norm be used in (3), then the weights as given in (10) are (possibly unintentionally) bestowed upon the time series $\mathbf{p}_0$. Work to suggest solutions to rectify this problem is covered in [119], whereby given desired weights w_j methods to obtain matrices $\mathbf{Q}$ and $\mathbf{R}$ such that

$$\|\mathcal{S}(\mathbf{p}) - \mathcal{S}(\mathbf{p}_0)\|_{\mathbf{Q},\mathbf{R}}^2 \approx \sum_{j=1}^{N} w_j (p_j - p_{0,j})^2,$$

are described. For certain cases of elementwise weights w_j it is possible to find exact equivalencies between $\|\mathcal{S}(\mathbf{p}) - \mathcal{S}(\mathbf{p}_0)\|_{\mathbf{Q},\mathbf{R}}^2$ and $\sum_{j=1}^{N} w_j (p_j - p_{0,j})^2$, but in general only approximations are possible. However, if the general weight matrices are allowed, as in (5), then the vector and matrix formulations (3) and (4) can be shown to be equivalent [63].

3. LOW-RANK HANKEL MATRICES AND LINEAR RECURRENCE SEQUENCES

Rank-deficient Hankel matrices are well-studied [59], in particular, they are tightly linked to linear recurrence sequences. Such sequences, can be characterized by a so-called canonical representation of sums of products of polynomial, exponential and cosine functions, which is a very useful class of time series. In this section, we provide a summary of these connections, together with references to the relevant literature.

3.1 Rank properties of Hankel matrices and time series of finite rank

The algebraic theory of Hankel matrices goes back to Sylvester [108], who studied them in the context of decompositions of binary forms, and Prony [95]. The most complete reference is the book of Heing & Rost [59, Ch.5 and 8], however, this reference was not known very well until recently for historical reasons. While not pretending to provide a comprehensive treatment of the subject, we briefly review the key points that are particularly useful in the context of time series analysis, following the details provided in [110] and [40].
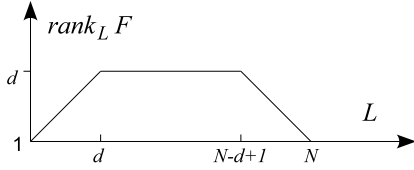
Figure 1. Rank of the Hankel matrix $\mathcal{H}_L$ for varying L.

The key property is that of how the rank varies with respect to the window length. Assume that we fix the length of the time series N and we vary the window length L from 1 to N (allowing in this case that L may be bigger than K).

Lemma 3.1 ([59, Corollary 5.2]). *For any sequence $\mathbf{p} \in \mathbb{C}^N$ there exists a number $d \leq \lfloor \frac{N+1}{2} \rfloor$ such that*

$$\mathrm{rank}\,\mathcal{H}_L(\mathbf{p}) = \min(L, K, d).$$

Figure 1 illustrates the behaviour of ranks as described in Lemma 3.1. As it can be seen from Fig. 1, the rank of the matrix is maximal, excepting a "plateau" at d. The case $d = \frac{N+1}{2}$ (for time series of odd length) is special, as in this case the Hankel matrix is full rank for any value of L (and is square nonsingular for $L = \frac{N+1}{2}$ in particular). For other cases ($d < \frac{N+1}{2}$), the matrices have a very particular structure; we will call such $\mathbf{p}$ *time series of finite rank*, as in the SSA literature [49].

There is a following important consequence of Lemma 3.1: if a Hankel matrix is rank-deficient (i.e., $\mathrm{rank}\,\mathcal{H}_L(\mathbf{p}) \leq r < \min(L, K)$, and this is a time series of finite rank), then $\mathrm{rank}\,\mathcal{H}_L(\mathbf{p}) = \mathrm{rank}\,\mathcal{H}_{r+1}(\mathbf{p})$, where $\mathcal{H}_{r+1}(\mathbf{p})$ is a Hankel matrix of built from the same vector but with $r + 1$ rows. This leads to the following corollary.

Corollary 3.1. *In the vector formulation (3), the SLRA problem with $r < L$ and $\mathcal{S} = \mathcal{H}_L$ is equivalent to the SLRA problem for the rank r and structure $\mathcal{S} = \mathcal{H}_{r+1}$, for the same choice of the (semi-)norm $\|\cdot\|$ (i.e., the search space and the optimal solutions of both problems coincide).*

Thanks to the Corollary 3.1 mentioned above, some authors (e.g., in the work of Markovsky and coauthors, [83]) argue that it makes no sense to consider Hankel SLRA with other window length than $L = r + 1$ (and therefore ignore other window lengths). This, however, leads to a confusion in the literature as it leaves an impression that the approach of [83] cannot handle other cases than rank reduction by 1.

3.2 Linear recurrence relations (LRR)

Low-rank Hankel matrices are tightly linked to linear recurrence relations. We say that a (real- or complex-valued) time series $\mathbf{p} \in \mathbb{C}^N$ satisfies a linear recurrence relation (LRR) of order r if there exists a non-zero vector $\boldsymbol{\theta} = (\theta_0, \ldots, \theta_r) \in \mathbb{C}^{r+1} \setminus \{0\}$ such that

$$(11) \quad \theta_0 p_k + \ldots + \theta_r p_{k+r} = 0, \quad \text{for all } k = 1, \ldots, N - r.$$

For real-valued $\mathbf{p}$, the coefficients of the linear recurrence relation can be also be chosen to be real-valued.

Next, we remark that the linear recurrence property is related to rank-deficiency of a Hankel matrix.

Remark 3.1. In the matrix form, we can rewrite (11) as

$$(12) \qquad \begin{pmatrix} \theta_0 & \cdots & \theta_r \end{pmatrix} \mathcal{H}_{r+1}(\mathbf{p}) = 0.$$

Here, as in Lemma 3.1, the Hankel matrix may have more rows than columns. This implies that if a Hankel matrix is rank-deficient, then the underlying time series satisfies an LRR.

Note that, in general, θ_0 and θ_r may be zero. If the last coefficient is non-zero, then the linear recurrence (11) can be represented in a slightly different form, which is more common in time series analysis (e.g., [49]).

Remark 3.2. If[1] $\theta_r \neq 0$, then we can define $a_k = -\theta_k/\theta_r$, and we can rewrite (11) as

$$(13) \qquad p_j = a_{r-1} p_{j-1} + a_{r-2} p_{j-2} + \cdots + a_0 p_{j-r}.$$

The LRR (13) (linear recurrent formula in the terminology of [49]) gives a way to recursively compute the next values of the time series based on the previous ones. This serves as a base for time series forecasting, explained in Section 5.3.1.

3.3 Linear recurrence relations and sums of exponentials

In this subsection, we explain the connection between time series of finite rank and sums of (modulated) complex exponentials. This connection is well-known for infinite sequences (or infinite time series), see, for example, [56], but in this and subsequent sections, we show what happens for finite-length time series as well.

The key tool is the *characteristic polynomial* of the LRR, which is

$$\theta(z) = \theta_r z^r + \theta_{r-1} z^{r-1} + \cdots + \theta_1 z + \theta_0 \,.$$

Assuming that the first and last coefficients are nonzero, the characteristic polynomial can be factorized thanks to the fundamental theorem of algebra,

$$(14) \qquad \theta(z) = (z - \lambda_1)^{\nu_1} (z - \lambda_2)^{\nu_2} \cdots (z - \lambda_s)^{\nu_s},$$

with $\nu_1 + \ldots + \nu_s = r$. Moreover, the form of the time series can be determined from its roots, as shown by the following lemma.

---

[1]The case $\theta_r \neq 0$ is called "continuable time series" in the SSA theory [49]. Indeed, from Remark 3.1, the case $\theta_r \neq 0$ corresponds to the case when $\mathbf{e}_L \notin \mathrm{span}(\mathcal{H}_L)$ for $L > r$, which is a necessary and sufficient condition of continuability in [49].

Theorem 3.1 ([56, Theorem 3.1.1]). *A complex-valued time series[2] $\mathbf{p}$ satisfies a linear recurrence with $\theta_0 \neq 0$ and $\theta_s \neq 0$ if and only if it can be expressed as*

$$(15) \qquad p_k = \sum_{j=1}^{s} P_j(k)\lambda_j^k, \ \ k = 1, 2, \ldots$$

where $P_j(k)$ are complex polynomials of degrees at most $\nu_j - 1$ and $\lambda_j \in \mathbb{C} \setminus \{0\}$ are the roots of the polynomial in (14). The coefficients of the polynomials in (15) are determined by the first s values of the time series.

Roughly speaking, Theorem 3.1 implies that the Hankel SLRA problem becomes a problem of approximating a given time series by time series of the form (15). Indeed, thanks to Remark 3.1, the Hankel matrix is rank-deficient if and only if it satisfies a linear recurrence relation, and therefore in most cases its form is given by Theorem 3.1. (The general case will be discussed at the end of this subsection.) In the case when the characteristic polynomial has simple roots ($s = r$ and $\nu_j = 1$ in (14)), we have that

$$(16) \qquad p_k = \sum_{j=1}^{r} c_k \lambda_j^k, \ \ k = 1, 2, \ldots$$

which is known as a sum-of-exponential model in signal processing, see also Section 3.6.

In time series analysis, the real-valued time series are more common. Therefore, we formulate a corollary of Theorem 3.1.

Corollary 3.2. *A real-valued time series satisfies a linear recurrence (11) with real coefficients $\boldsymbol{\theta} \in \mathbb{R}^{r+1}$, $\theta_0, \theta_r \neq 0$ if and only if it can be written as the following sum [76]:*
$$(17)$$
$$p_k = \sum_{j=1}^{q} P_j(k) \exp(d_k k) \sin(2\pi\omega_j k + \phi_j), \ \ k = 1, \ldots, N$$

where $\{P_j(k)\}$ are real polynomials (which can be just constant amplitudes), $\{d_l\}$ are the dampings, $\{\omega_l\}$ are the frequencies and $\{\phi_l\}$ are the phases, to use terminology common in signal processing.

Corollary 3.2 follows from the fact that if the coefficients of the linear recurrence are real, then the non-real roots of the characteristic polynomial must come in complex conjugate pairs. The number of terms q is equal to the number of real roots of $\theta(z)$ plus the number of complex conjugate pairs (i.e., for the case of $q = s/2$ in (16) if there are no real roots). More details on Corollary 3.2 can be found, e.g., in [49, 110, 40, 120].

3.4 Minimal linear recurrences and complete characterisation of finite-length case

While Theorem 3.1 and Corollary 3.2 gives us a form of the time series for several cases, it does not give any information whether this representation is unique. Another drawback is that it does not give a complete characterization, as it does not treat the cases when $\theta_0 = 0$ and $\theta_r = 0$.

Let us first discuss the uniqueness issue. In fact, we can define the linear recurrence relation of minimal length [70], which we can find as a corollary of Lemma 3.1 and Remark 3.1

Lemma 3.2. *For a time series of finite rank d, the minimal length linear recurrence relation is of length d with coefficients $(q_0, \ldots, q_d)$. It exists and is unique up to multiplication by a constant factor, and can be found from the left kernel of the corresponding Hankel matrix*

$$(q_0, \ldots, q_d)\mathcal{H}_{d+1}(\mathbf{p}) = 0.$$

Indeed, a shorter linear recurrence is not possible (which would imply $\mathcal{H}_d$ to be rank-deficient), and the left kernel of $\mathcal{H}_{d+1}(\mathbf{p})$ is generated by a single vector, which is defined uniquely up to normalization and sign. Such a vector $(q_0, \ldots, q_d)$ is related to the first characteristic polynomial in the terminology of [59]. For a range of lengths, all LRRs can be obtained from $\mathbf{q}$, as explained in Section 3.5.

Moreover, the time series $\mathbf{p}$ possesses a unique minimal representation similar to (16). In what follows, we formulate the general case in the following theorem (as to allow the coefficients q_0 and q_d to be zero). Let $\lambda_1, \ldots, \lambda_s \in \mathbb{C} \setminus \{0\}$, $\nu_1, \ldots, \nu_s$ be some positive integers, and $\nu_0, \nu_\infty \in \mathbb{N}$ be nonnegative integers such that

$$(18) \qquad \nu_0 + \nu_1 + \cdots + \nu_s + \nu_\infty = d.$$

Consider a signal $\mathbf{p} \in \mathbb{C}^N$ such that

$$(19) \quad p_k = \underbrace{\sum_{j=1}^{\nu_0} a_j \delta_{j-1}(k) + \sum_{l=1}^{\nu_\infty} b_l \delta_{N-l}(k)}_{\text{transient terms (beginning and end)}} + \sum_{j=1}^{s} P_j(k)\lambda_j^k,$$

where

1. each $P_j(k)$ is a polynomial of degree $\nu_k - 1$;
2. $\delta_x(k)$ is the (Kronecker) delta-function:

$$\delta_x(k) = \begin{cases} 1, & x = k, \\ 0, & x \neq k; \end{cases}$$

3. and $a_{\nu_0} \neq 0$, $b_{\nu_\infty} \neq 0$.

Finally, let $(q_0, \ldots, q_d)$ be coefficients of the characteristic polynomial of the minimal linear recurrence

$$(20) \ \ q(z) = q_0 + q_1 z + \cdots + q_d z^d = z^{\nu_0}(z - \lambda_1)^{\nu_1} \cdots (z - \lambda_s)^{\nu_s};$$

note ν_0 denotes the multiplicity of the root 0 (if present) and ν_∞ encodes the number[3] of zero leading coefficients (q_d, $q_{d-1},...$). The $q(z)$ is precisely the characteristic polynomial of the minimal linear recurrence relation $\mathbf{q}$ in Lemma 3.2.

Theorem 3.2 (Special case of [59, Theorem 8.2]). *The time series* $\mathbf{p}$ *is of finite rank* d *if and only if the signal* $\mathbf{p}$ *has the form* (19). *Moreover, the representation* (19) *with* d *satisfying* (18) *(called the* canonical representation *in [59]), is unique.*

For real-valued time series, a corollary of Theorem 3.2 can be derived giving a complete characterisation of real values time series of finite rank in terms of their canonical representation, in a similar manner to Corollary 3.2.

3.5 Non-minimal linear recurrences and left kernel of Hankel matrices

In this section, we give some details about the way the other linear recurrences (11) (for $r > d$) can be obtained from the minimal one. The following proposition gives a complete characterisation for a wide range of values of r.

Proposition 3.1 (Special case of [59, Theorem 5.1]). *Let* $\mathbf{p}$ *be a complex-valued time series of finite rank* d *and* $\mathbf{q} \in \mathbb{C}^{d+1}$ *be the coefficients of its minimal LRR. Then for any* $r \in \{d, \ldots, N-d\}$ *all linear recurrence relations* $\begin{pmatrix} \theta_0 & \cdots & \theta_r \end{pmatrix}$ *of length* r *satisfied by* $\mathbf{p}$ *must have the form*

$$
(21) \qquad \begin{pmatrix} \theta_0 \\ \vdots \\ \theta_r \end{pmatrix} = \underbrace{\begin{pmatrix} q_0 & 0 & \cdots & 0 \\ q_1 & q_0 & \ddots & \vdots \\ \vdots & q_1 & \ddots & 0 \\ q_d & \vdots & \ddots & q_0 \\ 0 & q_d & & q_1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & q_d \end{pmatrix}}_{\mathbf{M}_{r-d}(\mathbf{q})} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_{r-d} \end{pmatrix},
$$

where b_k *are some complex coefficients (in the case of real-valued* $\mathbf{p}$, *the coefficients* b_k *can be chosen to be real as well).*

In fact, (21) can be conveniently interpreted in terms of the characteristic polynomial of the linear recurrence. Indeed, the equation (21) is equivalent to

$$
\theta(z) = q(z)c(z);
$$

this is why $\mathbf{M}_{r-d}(\mathbf{q})$ is often called a multiplication matrix, because it represents multiplication by the polynomial $q(z)$. Thanks to Remark 3.1, such linear recurrences correspond to all the vectors in the left kernel of the matrix $\mathcal{H}_{r+1}(\mathbf{p})$ (i.e., the set of all vectors such that $\boldsymbol{\theta}^\top \mathcal{H}_{r+1}(\mathbf{p}) = 0$). Thus, for $d \leq r \leq N-d$, the span of columns of the matrix $\mathbf{M}_{r-d}(\mathbf{q})$

---

[3]This corresponds to the multiplicity of ∞ root in the terminology of [59].

in (21) coincides with the left kernel of the matrix $\mathcal{H}_{r+1}(\mathbf{p})$ (which is a linear subspace of $\mathbb{C}^{r+1}$ of dimension $r - d + 1$).

Note that for $N - d < r \leq N - 1$, the situation is more complicated: a second generating linear recurrence comes into play, as explained in [59, Theorem 5.1, Definition 5.4]. This is due to the fact that as $r > N - d$, the rank defect of $\mathcal{H}_{r+1}(\mathbf{p})$ grows twice faster as r increases (see Fig. 1). Note that in the case of time series of infinite length, the second polynomial does not appear, and all possible linear recurrences are given by (21), see e.g. [56].

Finally, we make a remark that there is a representation similar to (21) for the tangent space to the manifold of time series of finite rank d (of dimension $2d$). In fact, the tangent space is spanned by the matrix $\mathbf{M}_{N-1-2d}(\mathbf{q^2})$, where $\mathbf{q^2}$ is the vector of coefficients of squared polynomial $q^2(z)$ [120, Theorem 2.3]. Similar results are known in the field of algebraic geometry in the study of so-called secant varieties (see e.g. [61, Theorem 3.9] or [14, Theorem 2.1]).

3.6 Other uses of Hankel SLRA and other matrix structures

As already remarked, the SLRA problem (3) can be formulated in the same manner for an arbitrary affine structure. We list below some common problems and structures.

Hankel structure in signal processing Fitting and analysing the sum-of-exponentials model is very common in signal processing and so there has been much attention on Hankel SLRA to obtain solutions in problems such as: spectral estimation [96, 107], sparse signal recovery [84], and denoising particular types of signal [15, 27]. Other uses of the sum-of-exponentials model include: finite rate of innovation [30], shape-from-moments [31], and superresolution [18] problems. Note that in spectral estimation it is customary to restrict to unit-norm exponents $|\lambda_k| = 1$, which requires an additional constraint on positive semidefiniteness of the corresponding Toeplitz matrix [18]. Some literature exists on developing approximate Prony methods (see [118] and references therein), which are closely linked to Hankel SLRA. The Hankel structure is a special case of the block-Hankel structure which is discussed later in this subsection.

Yet another and classical use of the model (16) in pure mathematics appeared in the field of algebraic geometry and tensor decompositions. In fact, (16) corresponds exactly to the decomposition of a $2 \times \cdots \times 2$ symmetric tensor into a sum of rank-one symmetric tensors [108, 61, 26], see also [93] (equivalently, as a decomposition of a binary form into sums of powers of linear forms, the so-called Waring problem). This is why many aspects of the theory of Hankel matrices trace back to classic works in algebraic geometry and are in parallel studied in the works on tensor decompositions. Low-rank approximation of Hankel matrices is also a useful tool for function approximation [10, 52].

Block-Hankel and mosaic Hankel structure This structure often appears when we wish to model a multivariate time series, i.e., a collection of M time series $\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(M)}$. A typical way is to stack the Hankel matrices corresponding to different time series into one structure

$$(22) \qquad \begin{bmatrix} \mathcal{H}_L(\mathbf{p}^{(1)}) & \cdots & \mathcal{H}_L(\mathbf{p}^{(M)}) \end{bmatrix}.$$

A low-rank approximation of such series is performed when we assume some common structure of time series [94], which is used, for example, in multivariate SSA [47]. Another use of such matrices occurs when the different time series are inputs and outputs of a dynamical systems, which allows for solving system identification problems, thanks to behavioural systems theory [88, 80, 83, 116].

If the time series are of the same length, then the matrix (22) can be reshaped to have a block-Hankel structure. In the general case (of different lengths), (22) is a special case of so-called mosaic Hankel structure [86, 57]. The block-Hankel matrix structure has the form which in general has the following block form:

$$(23) \qquad \begin{pmatrix} \mathbf{H}_1 & \mathbf{H}_2 & \mathbf{H}_3 & \cdots & \mathbf{H}_K \\ \mathbf{H}_2 & \mathbf{H}_3 & \mathbf{H}_4 & \cdots & \mathbf{H}_{K+1} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \mathbf{H}_L & \mathbf{H}_{L+1} & \mathbf{H}_{L+2} & \cdots & \mathbf{H}_N \end{pmatrix},$$

where $\mathbf{H}_k$ are $p \times q$ Hankel matrices. Low-rank approximations of block-Hankel matrices are widely used in linear system theory, where it is related to the problems of realization and model reduction for linear systems [72, 60, 80]. There is also a link between block-Hankel matrices and the short-time Fourier transform [112].

The algebraic theory for finite-size block-Hankel matrices (similar to that given in Sections 3.1-3.5) is also linked to linear recurrences, however, several linear recurrences are needed in this case [58], which makes the explicit parameterization more complicated [37].

Multilevel Hankel matrices Multilevel Hankel matrices (also called quasi-Hankel matrices [90]) appear in problems dealing with higher-dimensional data arrays (e.g., images), where the model of sum-of-exponentials [89, 117] is often applicable. A special case (for two-dimensional data) concerns Hankel-block-Hankel matrices, which have the structure as in (23), but with block also having Hankel structure. Some application areas include: magnetic resonance imaging [104], seismic data analysis [64, 92], two-dimensional extensions of SSA [50, 47, 105], and symmetric tensor decompositions in the general case [13]. A complete algebraic theory for finite-size multilevel Hankel matrices is not available; the characterization is available in the infinite case, where the multivariate linear recurrences are linked to polynomial ideals [89, 73].

Sylvester matrices The Sylvester matrix structure is composed of (two or more) multiplication matrices (i.e., the matrix in (21)) stacked together. The rank of the Sylvester matrix is related to the degree of the greatest common divisor of the polynomials whose co-efficients populate the elements of the Sylvester matrix. Consequently, Sylvester low-rank approximation appears in the problem of finding approximate common divisors of polynomials [23, 114], and in some signal processing applications, e.g., blind deconvolution [35].

4. ALGORITHMS

4.1 Alternating projections and subspace-based approaches

4.1.1 Cadzow's iterations and singular spectrum analysis

One of the simplest (and earliest) algorithms to describe are the so-called Cadzow iterations [17] which can broadly be described as an alternating projections algorithm, which alternates approximations between the closest low-rank matrix and the closest Hankel matrices. Each of these projections are described next.

Projection to the closest low-rank matrix Suppose we are given $\mathbf{A} = \mathcal{S}(\mathbf{p}_0)$. Denote

$$\pi_{(\mathbf{Q},\mathbf{R})}^{(r)}(\mathbf{A}) = \underset{\mathbf{B}:\text{rank}(\mathbf{B})\leq r}{\arg\min} \|\mathbf{A} - \mathbf{B}\|_{\mathbf{Q},\mathbf{R}}^2.$$

The matrix $\pi_{(\mathbf{Q},\mathbf{R})}^{(r)}(\mathbf{A})$ is obtained from the SVD of $\mathbf{A}$, as described in Theorem 2.1.

Projection to the closest Hankel matrix For a given matrix $\mathbf{A} = \mathcal{S}(\mathbf{p}_0)$ we denote

$$(24) \qquad \pi_{(\mathbf{Q},\mathbf{R})}^{\mathcal{H}}(\mathbf{A}) = \underset{\mathbf{B}:\mathbf{B}\in\mathscr{S}}{\arg\min} \|\mathbf{A} - \mathbf{B}\|_{\mathbf{Q},\mathbf{R}}^2,$$

where $\mathscr{S}$ is the set of structured (Hankel) matrices. In the special case when $\mathbf{Q}$ and $\mathbf{R}$ are identity matrices then (24) has an explicit solution. In this case, let $\mathbf{A}^{\mathcal{H}} = \pi_{(\mathbf{I},\mathbf{I})}^{\mathcal{H}}(\mathbf{A})$ and the projection is given by averaging over the antidiagonals

$$\mathbf{A}_{l,k}^{\mathcal{H}} = \kappa_{l+k-1}^{-1} \sum_{l+k=l'+k'} \mathbf{A}_{l',k'}$$

with κ_j as defined in (10). The advantage of this case (Frobenius norm) that the projection can be efficiently implemented with the Fast Fourier Transform, which brings the complexity of the Cadzow iterations down to $O(rN \log N)$ (see e.g. [69, 47]); in addition, the whole Hankel matrix need not be stored.

In the general case, thanks to Lemma 2.1, the projection can be found as $\pi_{(\mathbf{Q},\mathbf{R})}^{\mathcal{H}}(\mathbf{A}) = \mathcal{S}(\widetilde{\mathbf{p}})$ with

$$(25) \qquad \widetilde{\mathbf{p}} = \underset{\mathbf{p}}{\arg\min} \|\widetilde{\mathcal{S}}(\mathbf{p}) - \mathbf{Q}^{1/2}\mathbf{A}\mathbf{R}^{1/2}\|_F,$$

where $\widetilde{\mathcal{S}}(\mathbf{p}) := \mathbf{Q}^{1/2}\mathcal{S}(\mathbf{p})\mathbf{R}^{1/2}$ is a modified matrix structure, which is still linear. The minimizer of (25) has an explicit solution as shown in [63, Lemma 1].

Alternating projections   Using the projections $\pi_{(\mathbf{Q},\mathbf{R})}^{(r)}$ and $\pi_{(\mathbf{Q},\mathbf{R})}^{\mathcal{H}}$ then we can construct the following alternating projections algorithm.

Set $\mathbf{A}^{(0)} = \mathbf{A} = \mathcal{S}(\mathbf{p}_0)$. For $i = 1, 2, \ldots$

$$(26) \qquad \mathbf{A}^{(i)} = \pi_{(\mathbf{Q},\mathbf{R})}^{\mathcal{H}}\left(\pi_{(\mathbf{Q},\mathbf{R})}^{(r)}(\mathbf{A}^{(i-1)})\right)$$

If $\mathbf{Q}$ and $\mathbf{R}$ are identity matrices then iterations of (26) are known as Cadzow iterations, with the special case of one iteration corresponding to the basic version of singular spectrum analysis (SSA). Numerical studies of Cadzow iterations and SSA are provided in [39]. Versions and the relative merits of SSA and Cadzow iterations where observations are allocated different weights are described in [119].

4.1.2 Convergence of Cadzow iterations and improvements

As seen from the form of the Cadzow iterations, the matrices $\mathbf{A}^{(i)}$ get closer to both the set $\mathcal{S}$ of structured matrices and the set $\mathcal{M}_{\leq r}$ of the manifold of fixed rank; but they are not, in general guaranteed to converge to a point in the intersection of $\mathcal{S}$ and $\mathcal{M}_{\leq r}$. However, if a point in the intersection is well-behaved (satisfies some regularity conditions, see [77] and [5]), then the Cadzow iterations converge linearly; as argued in [5, §7], for Hankel matrices almost all matrices in the intersection are well-behaved. Also, for rank-1 matrices, it was shown in [68] that the Cadzow iterations always converge. For rank-r matrices, as shown [41, 119] there always exists a convergent subsequence of Cadzow iterations.

Another issue is that the Cadzow iterations do not minimize the cost function (4) of interest (i.e., it is not guaranteed that the point in the intersection is the closest to $\mathcal{S}(\mathbf{p}_0)$); a classic example of convergence to a suboptimal solution is given in [29]. An easy and cheap scalar correction is described in [42] (see also [119]) is as follows. Suppose $\mathbf{Z}$ is an approximation of $\mathbf{A} = \mathcal{S}(\mathbf{p}_0)$. Then, if we take c as the solution

$$(27) \qquad c = \underset{b \in \mathbb{R}}{\arg\min} \|\mathbf{A} - b\mathbf{Z}\|_{\mathbf{Q},\mathbf{R}}^2$$

then this defines a better approximation in the sense $\|c\mathbf{Z} - \mathcal{S}(\mathbf{p}_0)\|_{\mathbf{Q},\mathbf{R}}^2 \leq \|\mathbf{Z} - \mathcal{S}(\mathbf{p}_0)\|_{\mathbf{Q},\mathbf{R}}^2$. Note that such a correction preserves the rank and the structure of the matrix (e.g., Hankel). In case of the Frobenius norm ($\mathbf{Q}, \mathbf{R} = \mathbf{I}$), the correction can be simply computed as

$$c = \frac{\text{Trace}(\mathbf{Z}^T \mathcal{S}(\mathbf{p}_0))}{\text{Trace}(\mathbf{Z}^T \mathbf{Z})}.$$

The correction can be applied at each step or at the end of the algorithm.

A further improvement was proposed in [102], where instead of $\pi_{(\mathbf{Q},\mathbf{R})}^{(r)}(\mathbf{A}^{(i-1)})$ on the space $\mathcal{S}$ of structured matrices, a Newton-like step was suggested: the next iteration $\mathbf{A}^{(i-1)}$ is computed as a projection of $\mathbf{A}^{(0)}$ on the intersection of $\mathcal{S}$ and the tangent space to the manifold of $\mathcal{M}_r$ of fixed rank matrices at $\pi_{(\mathbf{Q},\mathbf{R})}^{(r)}(\mathbf{A}^{(i-1)})$. It was shown that this algorithm is quadratically convergent, but the cost of each iteration becomes higher. Another recently proposed modification of Cadzow [27] is based on mimicking proximal splitting methods from convex optimization. Yet another method which can be seen as an modification of Cadzow, is the alternating direction method of multipliers proposed in [7]; it requires however storing the dual variables, which increases the storage complexity compared to Cadzow iterations.

4.1.3 Multistart/global optimization

With suitable modifications, (26) can be used as the basis to construct algorithms known to (at least linearly) converge to the global optima. Possible options include running iterations of (26) from several starting points (akin to random multistart), as a multi-stage algorithm and also as an evolutionary method.

The main algorithm studied within the family constructed by the authors in [42] is the so-called multistart APBR (alternating projections with backtracking and randomization) where theoretical conditions are also given for its convergence to the global optima. Within this algorithm, backtracking to the initial data is beneficial to ensure that the approximation remains close to it, whilst adding randomisation assists in moving the iterations away from possible local minima.

The multistart APBR is briefly described as follows. Let $\mathbf{X}$ denote a random Hankel matrix which corresponds to a realization of a white noise Gaussian process $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_N)$, that is, $\mathbf{X} = \mathcal{S}(\boldsymbol{\xi}) = \mathcal{H}_L(\boldsymbol{\xi})$. We assume each ξ are independent Gaussian random variables with zero mean and some variance to be set by the user.

Set $\mathbf{A}^{(0)} = \mathbf{A} = \mathcal{S}(\mathbf{p}_0)$, In multistart APBR, we run M independent trajectories starting at random Hankel matrices $\mathbf{A}^{(0,j)} = (1 - s_0)\mathbf{A}^{(0)} + s_0\mathbf{X}$, with some $0 \leq s_0 \leq 1$, where for trajectory $j = 1, \ldots, M$,
$$(28)$$
$$\mathbf{A}^{(i+1,j)} = c_{i,j}\left[(1 - \delta_i)\pi^{\mathcal{H}}\left(\pi^{(r)}(\mathbf{A}^{(i,j)})\right) + \delta_i\mathbf{A}^{(0)} + \sigma_i\mathbf{X}\right],$$

where $\pi^{\mathcal{H}} = \pi_{(\mathbf{Q},\mathbf{R})}^{\mathcal{H}}$, $\pi^{(r)} = \pi_{(\mathbf{Q},\mathbf{R})}^{(r)}$ and $c_{i,j}$ is the correction defined in (27) for the matrix in brackets. Each trajectory is either run until convergence or for a pre-specified number of iterations. If $s_0 = \delta_i = \sigma_i = 0$ then the iterations (28) co-incide with those of (26) albeit with some local improvement as already described. If $s_0 > 0$ then the j-th trajectory of multistart APBR starts at a random matrix in the neighbourhood of $\mathbf{A}^{(0)}$, with the width of this neighbourhood controlled by the parameter s_0. If $\sigma_i > 0$ then there is a 'random mutation' at iteration i. When $\delta_i > 0$ the i-th

iteration 'backtracks' towards $\mathbf{A}^{(0)}$ At some point it is recommended to set $\delta_i = \sigma_i = 0$ and directly have iterations of the form (26) to accelerate convergence to the set of rank-r Hankel matrices. Recommendations as to the selection of the parameters δ_i and σ_i are also offered in [42].

4.2 Local optimization

There is a significant body of work adressing the problem (3) focused on developing computationally efficient methods for computing a locally optimal solution. Two types of parameterization are typically used: kernel and image representation, where the latter is related to optimizing the manifold of rank-r matrices. We provide below a summary of these approaches, with a focus on Hankel SLRA (see also the introduction of [63] for more details on kernel vs. image representation).

We note that there is no guarantee that a globally optimal solution is found, and the solution will depend on the initial point selected. Many existing algorithms are highly sensitive to the choice of the initial point, and due to the existence of several local minima often do not move significantly from it [42]. However, these methods can be improved by choosing different (random) starting points.

4.2.1 Kernel representation and variable projection

Consider the Hankel low-rank approximation problem (3) in the vector formulation with $\mathcal{S} = \mathcal{H}_{r+1}$ (so that $L = r+1$), which aims at reducing the rank by one[4]. Then thanks to Remark 3.1, we can replace the rank constraint with a constraint on the kernel of the Hankel matrix, which leads to an equivalent formulation

$$(29) \quad \min_{\boldsymbol{\theta} \in \mathbb{R}^{r+1}, \boldsymbol{\theta} \neq 0, \widehat{\mathbf{p}} \in \mathbb{R}^{r+1}} \|\widehat{\mathbf{p}} - \mathbf{p}_0\|^2$$
$$\text{subject to} \quad \boldsymbol{\theta}^\top \mathcal{H}_{r+1}(\widehat{\mathbf{p}}) = 0,$$

where $\boldsymbol{\theta}^\top \mathcal{H}_{r+1}(\mathbf{p}) = 0$ is a shortcut for the constraint (12), and the norm is squared to make the objective smooth. The reformulation (29) is called the *kernel representation*, because the search space is augmented by an additional vector parameterizing the kernel. The kernel representation is also closely related to desingularization in algebraic geometry [67]. Also note the kernel representation was described here only for rank reduction of the given matrix by 1, rank reduction by more than one is also possible (which leads to minimization of a cost function on a Grassmann manifold, see [86]).

In the kernel representation, we are free to choose a scaling of the vector $\boldsymbol{\theta}$, as it does not change the cost function. Earlier approaches addressed Hankel SLRA from a structured total least squares [1] viewpoint (see [81] or [88] and for a survey), which corresponds to a particular choice of

---

[4]Note that by virtue of Corollary 3.1, any Hankel SLRA problem with $L > r$ can be reduced to the case $L = r+1$

$\theta_r = 1$ (fixing one of the coordinates to 1, which parameterizes a part of the search space). This was the approach of De Moor [29], who developed a non-linear extension of the SVD named the Riemannian SVD, which aims to solve a Lagrange multipliers formulation of (3), and also Rosen et al. [100], who proposed a linear approximation to (29) and then use Newton's method. A modification of this approach for rank reduction more than by 1 was described in Van Huffel [115], but this involves Kronecker products which dramatically inflate the dimension of the involved matrices. Structured total least squares can be also viewed as the problem of estimating an errors-in-variables regression model which has constraints on the matrix structure.

One of the most successful approaches is the variable projection approach, initially proposed in [88], which was developed in a series of papers [86, 85]). It uses the fact that for a fixed $\boldsymbol{\theta}$, the minimum of (29) (denoted as $f(\boldsymbol{\theta})$) has a closed form solution and thus the SLRA problem is equivalent to the minimization of $f(\boldsymbol{\theta})$ (and thus the variable $\widehat{\mathbf{p}}$ was eliminated). The cost function $f(\boldsymbol{\theta})$ can be minimized by general purpose constrained optimization methods, such as Gauss-Newton and Levenberg-Marquardt. For elementwise weights, the complexity of evaluating $f(\boldsymbol{\theta})$ and its gradient is $O(rN)$ and the complexity of evaluating an approximation of the Hessian is $O(r^2N)$ [113]. However, this approach is more difficult to implement efficiently in case of missing data [85] (unlike other approaches, presented in the previous and the following subsections).

There are some further improvements of the variable projection approach and other methods that operate in the kernel representation. For example, in [120, 121], a modified Gauss-Newton iteration was proposed, that makes use of the parameterization of the tangent space and a fast projection onto it. A convex relaxation based on the kernel representation was proposed in [25] that globally solves the SLRA if $\mathcal{S}(\mathbf{p}_0)$ is in a small neighborhood of a low-rank and structured matrix. Another recent approach using kernel representation is based on an iterative reweighting scheme [118]. An augmented Lagrangian approach in kernel representation was proposed in [12, Ch. 5].

4.2.2 Image representation and other methods

The image representation approach (according to the categorization in [63, 83]), consists of using the fact that any rank-r matrix $\mathbf{X} \in \mathcal{M}_r$ can be factorized as $\mathbf{X} = \mathbf{PL}$, where $\mathbf{P} \in \mathbb{R}^{L \times r}$ and $\mathbf{L} \in \mathbb{R}^{r \times L}$ and hence the SLRA problem can be posed as optimization over $\mathbf{P}$ and $\mathbf{L}$. The difficulty in using the image representation for the SLRA problem is that an additional constraint $\mathbf{X} \in \mathcal{S}$ needs to be handled. In [63], it was proposed to minimize the following objective:

$$\min_{\mathbf{P},\mathbf{L}} \|\mathbf{PL} - \mathbf{X}_0\|^2_{\mathbf{W}_m} + \lambda \|\mathbf{PL} - \pi^{\mathcal{H}}(\mathbf{PL})\|^2_F,$$

where the second term penalizes the distance to structured matrices and λ is increased adaptively, from small to large.

For fixed λ, the objective function is minimized using the alternating least squares strategy (for fixed $\mathbf{P}$, the problem becomes a least squares problem in $\mathbf{L}$, and vice versa). An extension of this approach based on Lagrangian formulation was proposed in [55]. Another extension based on proximal methods was proposed in [78] in order to handle multiple rank constraints. In general, other methods for optimization on the set (manifold) of rank-r matrices can be applied to such a reformulation of SLRA.

Recently, in a series of papers [54, 35, 36] it was shown that a solution to SLRA can be found as a solution of a gradient system for eigenvalues or singular values. Therefore, it was suggested to solve SLRA by integrating the corresponding ordinary differential equation. While the approach works very well for some matrix structures (e.g., Sylvester [54, 35]), its behaviour in the Hankel case is not completely clear and ad-hoc adjustments to the algorithm are needed [36].

A class of convex relaxation methods have been proposed, whereby a common strategy is to replace the rank constraint with the nuclear norm [32, 34, 11] (these approaches will be mentioned in the next section in the context of time series forecasting). Tighter relaxations were also proposed based on convex envelopes [21, 6, 4], [53, §VI] which take into account the particular cost function under consideration. Finally, several iteratively reweighted least squares methods based on surrogates for rank were proposed, which possess local convergence and theoretical guarantees [75, Ch.3], [74], and allow for a fast implementation [91].

5. CONVEX RELAXATIONS FOR FORECASTING

5.1 Forecasting as matrix completion

For forecasting, we introduce some additional notation for simplicity of the presentation. We denote $N = n + m$, where n is be the length of the given time series that we wish to forecast and m is the number of observations to be forecast. We use the notation $\mathbf{p}_{(1:n)} = (p_1, p_2, \ldots, p_n)$ for the first 'known' n elements of $\mathbf{p}$. Two obvious important cases are when $m = 0$ (where there are no observations to forecast) and $m > 1$. The Hankel matrix with $\mathbf{p}$ embedded takes the form

$$(30) \quad \mathcal{S}(\mathbf{p}) = \begin{pmatrix} p_1 & p_2 & \cdots & \cdots & \cdots & p_K \\ p_2 & p_3 & \cdots & \cdot^{\cdot^{\cdot}} & \cdots & \vdots \\ \vdots & \cdots & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \cdots & p_n \\ \vdots & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & p_{n+1} \\ \vdots & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \vdots \\ p_L & \cdots & p_n & p_{n+1} & \cdots & p_{n+m} \end{pmatrix}.$$

In (30), the grey-shaded values are "known" and others are "missing".

5.2 Nuclear norm relaxation

Often the nuclear norm relaxation of the rank constraint is used, since the nuclear norm is a convex envelope of the rank function. Formally, for a matrix $\mathbf{X} \in \mathbb{C}^{L \times K}$ its nuclear norm is defined as

$$\|\mathbf{X}\|_* = \sum_{k=1}^{\min(L,K)} |\sigma_k(\mathbf{X})|,$$

where $\sigma_k(\mathbf{X})$ are the singular values of $\mathbf{X}$. This idea has a rich recent history and has been discussed by several eminent authors, see [20, 19, 97] for example.

There is an analogy between the use of the nuclear norm in matrix completion problems and the use of the ℓ_1 norm for sparse approximation (see [98] for an overview of this topic) and its use has underpinned several successful works describing how to impute the missing values of a matrix with astonishing accuracy. More generally nuclear norm relaxation has proved a useful tool in: spectral estimation [3], recommender systems [65], system identification [79] and several other application areas.

In what follows we will describe the low-rank matrix completion problem and the remainder of this review describes work which followed the initial proposal of Butcher and Gillard [16] who proposed that a time series be embedded into a Hankel matrix, with the data to be forecasted being missing elements stored in the bottom right-hand corner of this matrix.

As well as low-rank approximation offering the potential for many new algorithms for forecasting, the main theoretical question is when the nuclear norm relaxation solves the original low-rank matrix completion problem [82]. Existing work mainly assumes that the position of the missing elements is random [22], and often that the known entries are non-random. In general, unstructured matrices are studied, yielding many of the theoretical results in famous works such as [19] redundant for our problem of Hankel structure with missing values in a particular section of the matrix.

5.3 Exact low-rank matrix completion

Let $\mathbf{p}_0 = (p_{0,1}, p_{0,2}, \ldots, p_{0,n})$ be a given vector of observations (a time series). For a given matrix structure (8), the exact Structured Low-Rank Matrix Completion (SLRMC) problem is posed as

$$(31) \quad \tilde{\mathbf{p}} = \underset{\mathbf{p} \in \mathbb{R}^{(n+m)}}{\arg\min} \operatorname{rank} \mathcal{S}(\mathbf{p}) \quad \text{subject to} \quad \mathbf{p}_{(1:n)} = \mathbf{p}_0.$$

The implicit low-rank assumption of the Hankel matrix corresponds to the class of time series of so-called finite rank, which are described in [110]. Here the notation $\mathbf{p}_{(1:n)}$ represents the first n (known) elements of $\mathbf{p}$ and so the constraint $\mathbf{p}_{(1:n)} = \mathbf{p}_0$ ensures that the known elements are fixed in the approximation. The optimization problem (31) thus acts on m variables, corresponding to the m missing values.

A convex relaxation of (31) is obtained by replacing the rank with the nuclear norm:

$$(32) \qquad \hat{\mathbf{p}} = \underset{\mathbf{p}\in\mathbb{R}^{(n+m)}}{\arg\min} \|\mathcal{S}(\mathbf{p})\|_* \text{ subject to } \mathbf{p}_{(1:n)} = \mathbf{p}_0.$$

The intuition behind this relaxation is the same as for using the ℓ_1 norm in compressed sensing: the nuclear norm is expected to force all but a few singular values to be zero hopefully rendering solutions of (32) at least close to (31).

5.3.1 Solution of the exact matrix completion

For the Hankel matrix case, the solution of (31) is known. Before describing the approximate case we recall the class of so-called time series of finite rank and hence give the solution of (31). The performance of the nuclear norm relaxation is important in this exact case and is crucial to understand the behaviour of forecasting in the approximate case to be introduced later.

For time series of finite rank, the solution to (31) with Hankel structure (30) can be found by the so-called minimal rank extension of Hankel matrices, solved by [62] for square matrices and [59] for rectangular matrices. We summarise their results in the following theorem.

Theorem 5.1. *Consider the Hankel matrix structure* (30) *and let* $r \leq \min(L-1, K-1, \frac{n}{2})$. *For any complex-valued time series* $(p_{0,1}, p_{0,2}, \ldots, p_{0,n}, \ldots)$ *of finite rank* r *and vector* $\mathbf{p}_0 = (p_{0,1}, p_{0,2}, \ldots, p_{0,n})$ *then the unique solution of* (31) *is given by*

$$\tilde{\mathbf{p}} = (p_{0,1}, p_{0,2}, \ldots, p_{0,n}, p_{0,n+1}, \ldots, p_{0,n+m})$$

where $p_{0,n+1}, \ldots, p_{0,n+m}$ *are found from the linear recurrence formula* (13).

A proof is available in, for example, [111].

5.3.2 Known results on the performance of the nuclear norm

Recent results are described in [111, 40] with the first described by [28]. We summarise some of the known results.

Theorem 5.2. *Let* $\mathbf{p}_0 = (p_{0,1}, p_{0,2}, \ldots, p_{0,n})$ *be a complex-valued vector given as*

$$p_{0,k} = c\lambda^k, \quad k = 1, \ldots, n,$$

where $\lambda \in \mathbb{C}$.

- *If* $|\lambda| \leq 1$ *the solution of* (31), *i.e.*

$$p_{0,k} = c\lambda^k, \quad k = n+1, \ldots, n+m$$

is also a solution of (32), *in particular, if* $|\lambda| < 1$, *then the solution of* (32) *is unique.*

- *In the case* $L = K = n = m+1$, *if* $|\lambda| > 1$, *then the unique solution of* (32) *is given by*

$$p_{0,n+k} = c\frac{\lambda^n}{\overline{\lambda}^k}, \quad k = 1, \ldots, m.$$

A version of Theorem 3 was originally given in [28], and has been later refined in [111, Theorem 6] and [40, Propositions 4.2 and 4.5]. In this case the exponential needs to be damped for the solution of (31) to equal that of (32). Despite concerning what might be a simple case of a "rank 1" Hankel matrix it yields many clues as to when the nuclear norm relaxation will coincide with the original problem when there is no convex relaxation, and this is described in the next theorem, originally stated in [111] and extended for rectangular Hankel matrices in [40].

Theorem 5.3 ([40, Proposition 4.5])**.** *Fix* L, K, n, m, r *as in Theorem* 5.1, *and let* $r \leq \frac{\min(L,K)+1}{2}$. *Then there exists a number* $0 < \rho_{max,r,m} < 1$ *such that for any complex-valued time series of finite rank* r *with*

$$|\lambda_j| < \rho_{max,r,m},$$

the solution of (32) *is unique and coincides with the solution as described in Theorem* 5.1.

5.4 Approximate matrix completion

The approximate rank minimization can be posed as follows. Given $\mathbf{p}_0$, vector of weights $\mathbf{w} \in \mathbb{R}^n$, $m \geq 0$ and $\tau \geq 0$ find

$$(33)$$
$$\tilde{\mathbf{p}} = \underset{\mathbf{p}\in\mathbb{R}^{(n+m)}}{\arg\min} \text{rank } \mathcal{S}(\mathbf{p}) \text{ subject to } \|\mathbf{p}_{(1:n)} - \mathbf{p}_0\|_W \leq \tau,$$

which is a special case of the rank minimization problem (6). The parameter τ controls the precision of approximation and two extreme cases can be distinguished:

- if $\tau = 0$, then (33) is equivalent to the exact matrix completion problem (31), i.e. there is no approximation;
- if $m = 0$ in (33), then $\tilde{\mathbf{p}}$ is an approximation to the given vector $\mathbf{p}_0$ with no forecast.

Unlike (31), the problem (33) does not have a known solution.

In [40], the following relaxation of (33) using the nuclear norm was introduced and studied:

$$(34)$$
$$\mathbf{p}_* = \underset{\mathbf{p}\in\mathbb{R}^{(n+m)}}{\arg\min} \|\mathcal{S}(\mathbf{p})\|_* \text{ subject to } \|\mathbf{p}_{(1:n)} - \mathbf{p}_0\|_W \leq \tau.$$

There are alternative ways to extend the problem (32) to approximate versions. For example, consider the following equivalent formulations:

$$(35) \qquad \underset{\mathbf{p}\in\mathbb{R}^{(n+m)}}{\min} \|\mathbf{p}_{(1:n)} - \mathbf{p}_0\|_W \text{ subject to } \|\mathcal{S}(\mathbf{p})\|_* \leq \delta,$$

$$(36) \qquad \underset{\mathbf{p}\in\mathbb{R}^{(n+m)}}{\min} \|\mathbf{p}_{(1:n)} - \mathbf{p}_0\|_W + \gamma\|\mathcal{S}(\mathbf{p})\|_*,$$

where δ and γ are regularisation parameters for each of the formulations.

Problems (34), (35) and (36) are equivalent in the following sense: for any value of τ, there exist δ and γ such that
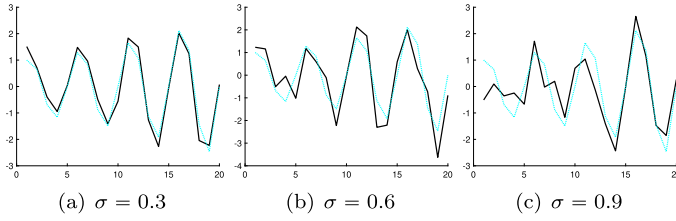
Figure 2. Plots of **p** for different σ, plot of time-series without noise in light colour, plot of time-series with noise in black.

Figure 3. Boxplots of $\log \|\mathbf{p} - \mathbf{p}_{approx}\|_2$ taken over 1000 simulations for the case of white noise.

the solutions to (34), (35) and (36) coincide. However, the relation between "equivalent" τ, δ and γ is not known (a priori).

6. NUMERICAL EXAMPLES

6.1 Example 1: structured low-rank approximations

Here we provide some commentary on the methods described earlier, via a simple example, to find a structured low-rank approximation of a Hankel matrix in which a time series has been embedded.

Assume we have observed a time-series ($n = 20$, $m = 0$) of the form $\mathbf{p} = (p_1, p_2, \ldots, p_{20})$ where $p_j = s_j + \varepsilon_j$. Here $s_j = \exp(\lambda j)\sin(2\pi\omega j)$ and ε_j is a noise term. We take $L = 10$ and compare $r = 3$ structured low-rank approximations to $\mathcal{S}(\mathbf{p})$ using Cadzow iterations (26), multistart APBR (as described in [42] and in Section 4.1.3), and the variable projections algorithm of Markovsky using the software provided in https://slra.github.io/software-slra.html and described in [86], which we will abbreviate as VP. We take $\lambda = 0.05$ and $\omega = 0.2$. We consider different forms of the noise term ε_j.

Case 1: white noise, $\varepsilon_j \sim N[0, \sigma^2]$ We first consider the case of so-called white noise, that is $\varepsilon_j \sim N[0, \sigma^2]$ for $j = 1, 2, \ldots, 20$, with some σ^2, and $cov[\varepsilon_i, \varepsilon_j] = 0$ for $i \neq j$. Plots of typical time-series with different values of σ^2 are given in Figure 2.

We vary σ and below report the norm $\|\mathbf{p} - \mathbf{p}_{approx}\|_2$ taken over 1000 simulations, where $\mathbf{p}_{approx}$ is the solution to (3) as found by the three methods described above. Figure 3 contains boxplots for each of these methods for different σ.

Case 2: deterministic 'noise' We now consider the case $\varepsilon_j = c(-1)^j$ for some $c > 0$. Table 1 reports the norm $\|\mathbf{p} - \mathbf{p}_{approx}\|_2$ obtained from the three methods considered.

Case 3: red noise, $\varepsilon_j = \alpha\varepsilon_{j-1} + \eta_j$ We now consider the case $\varepsilon_j = \sigma(\alpha\varepsilon_{j-1} + \eta_j)$, where $\eta_j \sim N[0, 1 - \alpha^2]$ for $j = 1, 2, \ldots, 20$, $\alpha = 0.5$, $cov[\eta_i, \eta_j] = 0$ for $i \neq j$, and $\sigma > 0$. This is red noise, more precisely, an autoregressive process of order 1. We vary σ and below report the norm $\|\mathbf{p} -$
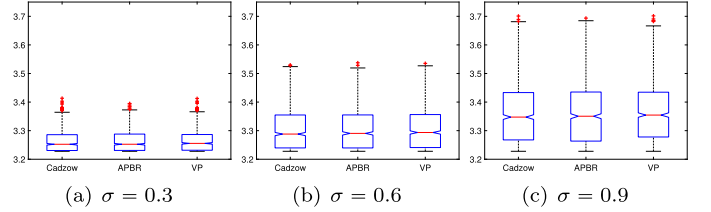
Table 1. The norm $\|\mathbf{p} - \mathbf{p}_{approx}\|_2$ obtained from three methods, for different c.

	0.2	0.4	0.6	0.8
Cadzow	25.6373	26.6507	28.2001	30.2832
APBR	25.4312	26.4329	28.1950	30.1874
VP	25.6373	26.6507	28.2001	30.2832

$\mathbf{p}_{approx}\|_2$ taken over 1000 simulations, where $\mathbf{p}_{approx}$ is the solution to (3) as found by the three methods described above. Figure 3 contains boxplots for each of these methods for different σ.

We make the following remarks. We note that for the examples considered, there are only marginal differences between each of the methods described. Improvements are found when using the multistart APBR method, which as noted, aims to approximate the global minimum of the problem. One may postulate that in bigger, more complex examples, perhaps this difference would be more pronounced. However it is a pleasing validation of each method that in this example each of the approximations they yield are close. More generally, each method seems relatively robust to the type and amount of additive noise it is exposed to. Finally, the performance of the VP approach can be improved by choosing different (random) starting points.

6.2 Example 2: structured low-rank matrix completion

To illustrate structured low-rank matrix completion and its application to forecasting we consider a small classical example "Cowtemp" which records the daily morning tem-
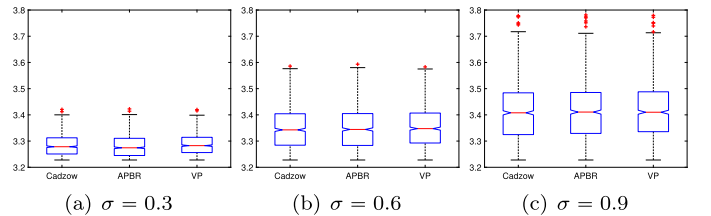

Figure 4. Boxplots of $\log \|\mathbf{p} - \mathbf{p}_{approx}\|_2$ taken over 1000 simulations for the case of red noise.
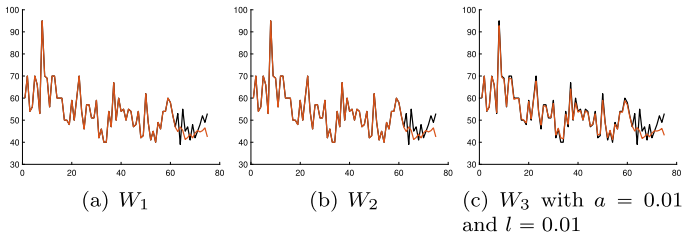
(a) W_1 (b) W_2 (c) W_3 with $a = 0.01$ and $l = 0.01$

Figure 5. Forecasts of the Cowtemp time series, with original data (black), for different weighting schemes.

(a) Surface plot   (b) Contour plot

Figure 6. Root mean square error of the forecast against a and l.

Figure 7. Forecast when $a = 0.001$ and $l = 0.017$ for weighting scheme W_3.

an active area of pursuit, with both rich theoretical work, several algorithms and much possibility for practical application. Open questions remain around when the solution of the nuclear norm relaxation coincides with the 'equivalent' rank minimization problem, and this is the motivator behind much present research. It remains to be the case to see such technology embraced by forecasting practitioners, but one would assume, that in time, the work described in this paper will be. Future big data applications on modelling several time series simultaneously are surely on the horizon.

REFERENCES

[1] ABATZOGLOU, T. J., MENDEL, J. M. and HARADA, G. A. (1991). The constrained total least squares technique and its applications to harmonic superresolution. *IEEE Transactions on Signal Processing* **39** 1070-1087.

[2] ADAMJAN, V. M., AROV, D. Z. and KREĬN, M. G. (1971). Analytic properties of Schmidt pairs for a Hankel operator and the generalized Schur-Takagi problem. *Mathematics of the USSR-Sbornik* **15** 31. MR0298453

[3] AKÇAY, H. (2014). Subspace-based spectrum estimation in frequency-domain by regularized nuclear norm minimization. *Signal Processing* **99** 69–85.

[4] AL-SHABILI, A. and SELESNICK, I. (2019). Sharpening Sparse Regularizers. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 4908–4912.

[5] ANDERSSON, F. and CARLSSON, M. (2013). Alternating projections on nontangential manifolds. *Constructive Approximation* **38** 489–525. MR3122280

[6] ANDERSSON, F., CARLSSON, M. and OLSSON, C. (2017). Convex envelopes for fixed rank approximation. *Optimization Letters* **11** 1783–1795. MR3722251

[7] ANDERSSON, F., CARLSSON, M., TOURNERET, J.-Y. and WENDT, H. (2014). A New Frequency Estimation Method for Equally and Unequally Spaced Data. *IEEE Transactions on Signal Processing* **62** 5761–5774. MR3273530

[8] ANTOULAS, A. C. (1997). *On the Approximation of Hankel Matrices* In *Operators, Systems and Linear Algebra: Three Decades of Algebraic Systems Theory* 17–22. Vieweg+Teubner Verlag, Wiesbaden.

perature of a cow. This time-series is available in several online repositories. The time-series consists of 75 observations, and the exercise is to forecast the last $m = 14$ observations, based on the first $n = 61$ observations. This exercise was also performed in [48], comparing forecasts of SSA with various parameter choices. As there, we select $L = 28$.

We solve (36) with $\gamma = 100$ and explore different weight vectors W. Figure 5 contains plots of the forecasts of the Cowtemp time series for three weight vectors W. First, we give each observation unit weight, $W_1 = (1, 1, \ldots, 1)$. Second we introduce the weights such that the weighted vector norm in (36) becomes the Frobenius norm $\|\mathcal{S}(\mathbf{p}_{(1:n)}) - \mathcal{S}(\mathbf{p}_0)\|_F$. Finally we have the exponential weighing, where the jth element of W_3 is given by $a \exp(lj)$ for scalars a and l.

Figure 6 contains plots of the root mean square error (RMSE) of the forecast obtained using exponential weighting, that is (36) with $W = W_3$, for different a and l. Note that the approximately optimal RMSE found from (36) is 4.9928 obtained when $a = 0.001$ and $l = 0.017$. A plot of the data with this forecast is given in Figure 7, and it has approximating rank of 28, in this case. The RMSE obtained by SSA as described in [48] was found to be 5.253602.

7. CONCLUSION

In this paper we have attempted to give a review, commentary and bibliography of papers associated with Hankel low-rank approximation/completion, through the lens of time series modelling and forecasting. It is clear that this is
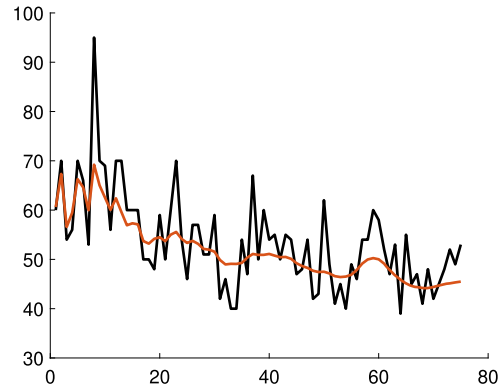
[9] BALLE, B., LACROCE, C., PANANGADEN, P., PRECUP, D. and RABUSSEAU, G. (2021). Optimal Spectral-Norm Approximate Minimization of Weighted Finite Automata. *arXiv preprint arXiv:2102.06860.* MR4288948

[10] BEYLKIN, G. and MONZÓN, L. (2005). On approximation of functions by exponential sums. *Applied and Computational Harmonic Analysis* **19** 17–48. MR2147060

[11] BLOMBERG, N., ROJAS, C. R. and WAHLBERG, B. (2015). Regularization Paths for Re-Weighted Nuclear Norm Minimization. *IEEE Signal Processing Letters* **22** 1980–1984.

[12] BORSDORF, R. (2012). Structured Matrix Nearness Problems: Theory and Algorithms PhD thesis, University of Manchester.

[13] BRACHAT, J., COMON, P., MOURRAIN, B. and TSIGARIDAS, E. (2010). Symmetric tensor decomposition. *Linear Algebra and its Applications* **433** 1851–1872. MR2736103

[14] BRAMBILLA, M. C. and OTTAVIANI, G. (2008). On the Alexander–Hirschowitz theorem. *Journal of Pure and Applied Algebra* **212** 1229–1251. MR2387598

[15] BRESLER, Y. and MACOVSKI, A. (1986). Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **34** 1081–1089. MR0642901

[16] BUTCHER, H. and GILLARD, J. (2017). Simple Nuclear Norm Based Algorithms for Imputing Missing Data and Forecasting in Time Series. *Statistics and Its Interface* **10** 19–25. MR3562108

[17] CADZOW, J. A. (1988). Signal enhancement-a composite property mapping algorithm. *Acoustics, Speech and Signal Processing, IEEE Transactions on* **36** 49–62.

[18] CANDÈS, E. J. and FERNANDEZ-GRANDA, C. (2013). Super-resolution from noisy data. *Journal of Fourier Analysis and Applications* **19** 1229–1254. MR3132912

[19] CANDES, E. J. and PLAN, Y. (2010). Matrix completion with noise. *Proceedings of the IEEE* **98** 925–936.

[20] CANDÈS, E. J. and RECHT, B. (2009). Exact Matrix Completion via Convex Optimization. *Foundations of Computational Mathematics* **9** 717–772. MR2565240

[21] CAPRANI, J. and CARLSSON, M. (2019). Quadratic Envelope Regularization for Structured Low Rank Approximation. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 8217–8221.

[22] CHEN, Y. and CHI, Y. (2014). Robust Spectral Compressed Sensing via Structured Matrix Completion. *IEEE Transactions on Information Theory* **60** 6576-6601. MR3265040

[23] CHIN, P., CORLESS, R. M. and CORLISS, G. F. (1998). Optimization strategies for the approximate GCD problem. In *Proceedings of ISSAC '98* 228–235. ACM. MR1805185

[24] CHU, M. T., FUNDERLIC, R. E. and PLEMMONS, L. J. (2003). Structured low rank approximation. *Linear Algebra Its Appl.* **366** 157–172. MR1987719

[25] CIFUENTES, D. (2021). A Convex Relaxation to Compute the Nearest Structured Rank Deficient Matrix. *SIAM Journal on Matrix Analysis and Applications* **42** 708-729. MR4252075

[26] COMON, P. and MOURRAIN, B. (1996). Decomposition of quantics in sums of powers of linear forms. *Signal Processing* **53** 93–107. Higher Order Statistics.

[27] CONDAT, L. and HIRABAYASHI, A. (2015). Cadzow denoising upgraded: A new projection method for the recovery of Dirac pulses from noisy linear measurements. *Sampling Theory in Signal and Image Processing* **14** 17–47. MR3317200

[28] DAI, L. and PELCKMANS, K. (2015). On the nuclear norm heuristic for a Hankel matrix completion problem. *Automatica* **51** 268–272. MR3284777

[29] DE MOOR, B. (1993). Structured total least squares and L2 approximation problems. *Linear Algebra and its Applications* **188** 163–205. MR1223460

[30] DRAGOTTI, P. L., VETTERLI, M. and BLU, T. (2007). Sampling Moments and Reconstructing Signals of Finite Rate of Innovation: Shannon Meets Strang–Fix. *IEEE Transactions on Signal Processing* **55** 1741-1757. MR2472334

[31] ELAD, M., MILANFAR, P. and GOLUB, G. H. (2004). Shape from moments - an estimation theory perspective. *IEEE Transactions on Signal Processing* **52** 1814-1829. MR2087688

[32] FAZEL, M. (2002). Matrix rank minimization with applications PhD thesis, Stanford University.

[33] FAZEL, M., HINDI, H. and BOYD, S. P. (2001). A rank minimization heuristic with application to minimum order system approximation. In *Proceedings of the 2001 American Control Conference.(Cat. No. 01CH37148)* **6** 4734–4739. IEEE.

[34] FAZEL, M., PONG, T. K., SUN, D. and TSENG, P. (2013). Hankel Matrix Rank Minimization with Applications to System Identification and Realization. *SIAM Journal on Matrix Analysis and Applications* **34** 946-977. MR3073649

[35] FAZZI, A., GUGLIELMI, N. and MARKOVSKY, I. (2018). An ODE based method for computing the Approximate Greatest Common Divisor of polynomials. *Numerical Algorithms* **81** 719–740. MR3953167

[36] FAZZI, A., GUGLIELMI, N. and MARKOVSKY, I. (2021). A gradient system approach for Hankel structured low-rank approximation. *Linear Algebra Appl.* MR4255495

[37] FELDMANN, S. and HEINIG, G. (1999). Parametrization of minimal rank block Hankel matrix extensions and minimal partial realizations. *Integral Equations and Operator Theory* **33** 151–171. MR1670847

[38] FLOATER, M. S., MANNI, C., SANDE, E. and SPELEERS, H. (2021). Best low-rank approximations and Kolmogorov n-widths. *SIAM Journal on Matrix Analysis and Applications* **42** 330–350. MR4226996

[39] GILLARD, J. (2010). Cadzow's basic algorithm, alternating projections and singular spectrum analysis. *Statistics and its Interface* **3** 335–343. MR2720137

[40] GILLARD, J. and USEVICH, K. (2018). Structured low-rank matrix completion for forecasting in time series analysis. *International Journal of Forecasting* **34** 582–597.

[41] GILLARD, J. and ZHIGLJAVSKY, A. (2013). Optimization challenges in the structured low rank approximation problem. *Journal of Global Optimization* **57** 733–751. MR3119378

[42] GILLARD, J. and ZHIGLJAVSKY, A. (2015). Stochastic algorithms for solving structured low-rank matrix approximation problems. *Communications in Nonlinear Science and Numerical Simulation* **21** 70–88. MR3278324

[43] GILLARD, J. and ZHIGLJAVSKY, A. A. (2016). Weighted norms in subspace-based methods for time series analysis. *Numerical Linear Algebra with Applications* **23** 947–967. MR3547538

[44] GILLIS, N. and GLINEUR, F. (2011). Low-rank matrix approximation with weights or missing data is NP-hard. *SIAM Journal on Matrix Analysis and Applications* **32** 1149–1165. MR2854607

[45] GOLUB, G. H. and VAN LOAN, C. F. (2012). *Matrix computations*, 2nd ed. JHU Press. MR3024913

[46] GOLYANDINA, N. (2010). On the choice of parameters in Singular Spectrum Analysis and related subspace-based methods. *Statistics and its Interface* **3** 259–279. MR2720132

[47] GOLYANDINA, N., KOROBEYNIKOV, A., SHLEMOV, A. and USEVICH, K. (2015). Multivariate and 2D Extensions of Singular Spectrum Analysis with the Rssa Package. *Journal of Statistical Software, Articles* **67** 1–78. MR3793637

[48] GOLYANDINA, N., KOROBEYNIKOV, A. and ZHIGLJAVSKY, A. (2018). *Singular spectrum analysis with R.* Springer. MR3793637

[49] GOLYANDINA, N., NEKRUTKIN, V. and ZHIGLJAVSKY, A. A. (2001). *Analysis of time series structure: SSA and related techniques.* CRC press. MR1823012

[50] GOLYANDINA, N. and USEVICH, K. (2010). 2D-extension of Singular Spectrum Analysis: algorithm and elements of theory. In *Matrix Methods: Theory, Algorithms and Applications* (V. Olshevsky and E. Tyrtyshnikov, eds.) 449–474. World Scientific. MR2648635

[51] GOLYANDINA, N. and ZHIGLJAVSKY, A. (2020). *Singular Spectrum Analysis for time series*, 2nd ed. Springer. MR4178480

[52] GONNET, P., GÜTTEL, S. and TREFETHEN, L. N. (2013). Robust Padé Approximation via SVD. *Siam Review* **55** 101–117. MR3089442

[53] GRUSSLER, C., RANTZER, A. and GISELSSON, P. (2018). Low-Rank Optimization With Convex Constraints. *IEEE Transactions on Automatic Control* **63** 4000-4007. MR3875402

[54] GUGLIELMI, N. and MARKOVSKY, I. (2017). An ODE based method for computing the distance of co-prime polynomials to common divisibility. *SIAM Journal on Numerical Analysis* **55** 1456–1482. MR3665204

[55] HAGE, C. and KLEINSTEUBER, M. (2015). Robust Structured Low-Rank Approximation on the Grassmannian. In *International Conference on Latent Variable Analysis and Signal Separation* 295–303. Springer.

[56] HALL, M. (1967). *Combinatorial Theory.* Blaisdell Publishing Company. MR0224481

[57] HEINIG, G. (1995). Generalized inverses of Hankel and Toeplitz mosaic matrices. *Linear Algebra Appl.* **216** 43–59. MR1319974

[58] HEINIG, G. and JANKOWSKI, P. (1992). Kernel structure of block Hankel and Toeplitz matrices and partial realization. *Linear Algebra and its Applications* **175** 1–30. MR1179338

[59] HEINIG, G. and ROST, K. (1984). *Algebraic methods for Toeplitz-like matrices and operators.* Birkhäuser, Boston. MR0782105

[60] HO, B. L. and KALMAN, R. E. (1966). Editorial: Effective construction of linear state-variable models from input/output functions. *Regelungstechnik* **14** 545–548.

[61] IAROBBINO, A. and KANEV, V. (1999). *Power sums, Gorenstein Algebras and Determinantal Loci. Lecture Notes in Mathematics* **1721**. Springer. MR1735271

[62] IOHVIDOV, I. (1982). *Toeplitz and Hankel matrices and forms. Algebraic theory.* Birkhäuser, Basel. MR0677503

[63] ISHTEVA, M., USEVICH, K. and MARKOVSKY, I. (2014). Factorization approach to structured low-rank approximation with applications. *SIAM Journal on Matrix Analysis and Applications* **35** 1180–1204. MR3257636

[64] JIA, Y., YU, S., LIU, L. and MA, J. (2016). A fast rank-reduction algorithm for three-dimensional seismic data interpolation. *Journal of Applied Geophysics* **132** 137–145.

[65] KANG, Z., PENG, C. and CHENG, Q. (2016). Top-n recommender system via matrix completion. In *Proceedings of the AAAI Conference on Artificial Intelligence* **30**.

[66] KHAN, M. A. R. and POSKITT, D. S. (2013). A note on window length selection in singular spectrum analysis. *Australian & New Zealand Journal of Statistics* **55** 87–108. MR3079022

[67] KHRULKOV, V. and OSELEDETS, I. (2018). Desingularization of Bounded-Rank Matrix Sets. *SIAM Journal on Matrix Analysis and Applications* **39** 451–471. MR3773762

[68] KNIRSCH, H., PETZ, M. and PLONKA, G. (2021). Optimal Rank-1 Hankel Approximation of Matrices: Frobenius Norm, Spectral Norm, and Cadzow's Algorithm. MR4293739

[69] KOROBEYNIKOV, A. (2010). Computation-and space-efficient implementation of SSA. *Statistics and Its Interface* **3** 357–368. MR2720139

[70] KUIJPER, M. and WILLEMS, J. C. (1997). On constructing a shortest linear recurrence relation. *IEEE Transactions on Automatic Control* **42** 1554-1558. MR1479091

[71] KULIS, B., SUSTIK, M. and DHILLON, I. (2006). Learning low-rank kernel matrices. In *Proceedings of the 23rd International Conference on Machine Learning* 505–512.

[72] KUNG, S. Y. (1978). A new identification and model reduction algorithm via singular value decomposition. In *Proceedings of 12th Asilomar Conference on Circuits, Systems and Computers* 705–714.

[73] KURAKIN, V. L., KUZMIN, A. S., MIKHALEV, A. V. and NECHAEV, A. A. (1995). Linear recurring sequences over rings and modules. *Journal of Mathematical Sciences* **76** 2793–2915. MR1365809

[74] KÜMMERLE, C. (2019). Understanding and Enhancing Data Recovery Algorithms Dissertation, Technische Universität München, München.

[75] KÜMMERLE, C. and VERDUN, C. M. (2019). Completion of Structured Low-Rank Matrices via Iteratively Reweighted Least Squares. In *2019 13th International conference on Sampling Theory and Applications (SampTA)* 1–5.

[76] LEMMERLING, P. and VAN HUFFEL, S. (2001). Analysis of the structured total least squares problem for Hankel/Toeplitz matrices. *Numerical algorithms* **27** 89–114. MR1847986

[77] LEWIS, A. S. and MALICK, J. (2008). Alternating Projections on Manifolds. *Mathematics of Operations Research* **33** 216–234. MR2393548

[78] LIU, T., MARKOVSKY, I., PONG, T. K. and TAKEDA, A. (2020). A Hybrid Penalty Method for a Class of Optimization Problems with Multiple Rank Constraints. *SIAM Journal on Matrix Analysis and Applications* **41** 1260–1283. MR4142769

[79] LIU, Z. and VANDENBERGHE, L. (2010). Interior-point method for nuclear norm approximation with application to system identification. *SIAM Journal on Matrix Analysis and Applications* **31** 1235–1256. MR2558821

[80] MARKOVSKY, I. (2008). Structured low-rank approximation and its applications. *Automatica* **44** 891–909. MR2530933

[81] MARKOVSKY, I. (2010). Bibliography on total least squares and related methods. *Statistics and its Interface* **3** 329–334. MR2720136

[82] MARKOVSKY, I. (2012). How effective is the nuclear norm heuristic in solving data approximation problems? In *Proc. of the 16th IFAC Symposium on System Identification* 316–321.

[83] MARKOVSKY, I. (2019). *Low-Rank Approximation: Algorithms, Implementation, Applications*, 2nd ed. Springer. MR3837540

[84] MARKOVSKY, I. and DRAGOTTI, P. (2018). Using Hankel Structured Low-Rank Approximation for Sparse Signal Recovery. 479–487. Springer.

[85] MARKOVSKY, I. and USEVICH, K. (2013). Structured low-rank approximation with missing data. *SIAM Journal on Matrix Analysis and Applications* **34** 814–830. MR3071404

[86] MARKOVSKY, I. and USEVICH, K. (2014). Software for weighted structured low-rank approximation. *J. Comput. Appl. Math.* **256** 278–292. MR3095704

[87] MARKOVSKY, I. and VAN HUFFEL, S. (2007). Overview of total least-squares methods. *Signal Processing* **87** 2283–2302.

[88] MARKOVSKY, I., WILLEMS, J. C., VAN HUFFEL, S. and DE MOOR, B. (2006). *Exact and Approximate Modeling of Linear Systems: A Behavioral Approach.* SIAM. MR2207544

[89] MOURRAIN, B. (2018). Polynomial-exponential decomposition from moments. *Foundations of Computational Mathematics* **18** 1435–1492. MR3875844

[90] MOURRAIN, B. and PAN, V. Y. (2000). Multivariate polynomials, duality, and structured matrices. *Journal of Complexity* **16** 110–180. MR1762401

[91] ONGIE, G. and JACOB, M. (2017). A Fast Algorithm for Convolutional Structured Low-Rank Matrix Recovery. *IEEE Transactions on Computational Imaging* **3** 535–550. MR3730215

[92] OROPEZA, V. and SACCHI, M. (2011). Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis. *Geophysics* **76** V25-V32.

[93] OTTAVIANI, G., SPAENLEHAUER, P.-J. and STURMFELS, B. (2014). Exact Solutions in Structured Low-Rank Approximation. *SIAM Journal on Matrix Analysis and Applications* **35** 1521–1542. MR3286679

[94] PAPY, J.-M., DE LATHAUWER, L. and VAN HUFFEL, S. (2006). Common pole estimation in multi-channel exponential data modeling. *Signal Processing* **86** 846–858.

[95] PRONY, G. R. B. (an III, 1795). Essai éxperimental et analytique: sur les lois de la dilatabilité de fluides élastique et sur celles de la force expansive de la vapeur de l'alkool, à différentes températures. *Journal de l'École Polytechnique Floréal et Prairial* **1** 24–76. cahier 22.

[96] RAZAVILAR, J., LI, Y. and LIU, K. R. (1996). Spectral estimation based on structured low rank matrix pencil. In *1996 IEEE*

International Conference on Acoustics, Speech, and Signal Processing Conference Proceedings **5** 2503–2506. IEEE.

[97] RECHT, B., FAZEL, M. and PARRILO, P. A. (2010). Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review* **52** 471–501. MR2680543

[98] RISH, I. and GRABARNIK, G. (2014). *Sparse modeling: theory, algorithms, and applications.* CRC press. MR3328718

[99] ROHDE, A. and TSYBAKOV, A. (2011). Estimation of high-dimensional low-rank matrices. *Annals of Statistics* **39** 887–930. MR2816342

[100] ROSEN, J. B., PARK, H. and GLICK, J. (1996). Total least norm formulation and solution for structured problems. *SIAM Journal on Matrix Analysis and Applications* **17** 110–126. MR1372925

[101] RUMP, S. (2003). Structured Perturbations Part I: Normwise Distances. *SIAM J. Matrix Anal. Appl.* **25** 1–30. MR2002898

[102] SCHOST, E. and SPAENLEHAUER, P.-J. (2016). A Quadratically Convergent Algorithm for Structured Low-Rank Approximation. *Found. Comput. Math.* **16** 457–492. MR3464213

[103] SHEN, H. and HUANG, J. Z. (2008). Sparse principal component analysis via regularized low rank matrix approximation. *Journal of Multivariate Analysis* **99** 1015–1034. MR2419336

[104] SHIN, P. J., LARSON, P. E. Z., OHLIGER, M. A., ELAD, M., PAULY, J. M., VIGNERON, D. B. and LUSTIG, M. (2014). Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion. *Magnetic Resonance in Medicine* **72** 959–970.

[105] SHLEMOV, A. and GOLYANDINA, N. (2014). Shaped extension of singular spectrum analysis. In *Proceedings of the 21st International Symposium on Mathematical Theory of Networks and Systems (MTNS 2014).*

[106] SREBRO, N. and JAAKKOLA, T. (2003). Weighted low-rank approximations. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)* 720–727.

[107] STOICA, P., MOSES, R. L. et al. (2005). Spectral analysis of signals.

[108] SYLVESTER, J. J. (1886). Sur une extension d'un théorème de Clebsch relatif aux courbes du quatrième degré. *Comptes Rendus, Math. Acad. Sci. Paris* **102** 1532–1534.

[109] UDELL, M. and TOWNSEND, A. (2019). Why are big data matrices approximately low rank? *SIAM Journal on Mathematics of Data Science* **1** 144–160. MR3949704

[110] USEVICH, K. (2010). On signal and extraneous roots in Singular Spectrum Analysis. *Statistics and Its Interface* **3** 281–295. MR2720133

[111] USEVICH, K. and COMON, P. (2016). Hankel Low-Rank Matrix Completion: Performance of the Nuclear Norm Relaxation. *IEEE Journal of Selected Topics in Signal Processing* **10** 637–646.

[112] USEVICH, K., EMIYA, V., BRIE, D. and CHAUX, C. (2018). Char-

acterization of Finite Signals with Low-Rank STFT. In *2018 IEEE Statistical Signal Processing Workshop (SSP)* 393–397.

[113] USEVICH, K. and MARKOVSKY, I. (2014). Variable projection for affinely structured low-rank approximation in weighted 2-norms. *J. Comput. Appl. Math.* **272** 430–448. MR3227396

[114] USEVICH, K. and MARKOVSKY, I. (2017). Variable projection methods for approximate (greatest) common divisor computations. *Theoretical Computer Science* **681** 176–198. Symbolic Numeric Computation. MR3659423

[115] VAN HUFFEL, S., PARK, H. and ROSEN, J. B. (1996). Formulation and solution of structured total least norm problems for parameter estimation. *IEEE Transactions on Signal Processing* **44** 2464–2474.

[116] WILLEMS, J. C. (1986). From time series to linear system— Part I. Finite dimensional linear time invariant systems. *Automatica* **22** 561-580. MR0863344

[117] YANG, Z., XIE, L. and STOICA, P. (2016). Vandermonde decomposition of multilevel Toeplitz matrices with application to multidimensional super-resolution. *IEEE Transactions on Information Theory* **62** 3685–3701. MR3506757

[118] ZHANG, R. and PLONKA, G. (2019). Optimal approximation with exponential sums by a maximum likelihood modification of Prony's method. *Adv. Comput. Math.* **45** 1657–1687. MR3955732

[119] ZVONAREV, N. and GOLYANDINA, N. (2017). Iterative algorithms for weighted and unweighted finite-rank time-series approximations. *Statistics and its Interface* **10** 5–18. MR3562107

[120] ZVONAREV, N. and GOLYANDINA, N. (2021). Low-rank signal subspace: parameterization, projection and signal estimation. *arXiv preprint arXiv:2101.09779.*

[121] ZVONAREV, N. and GOLYANDINA, N. (2021). Fast and stable modification of the Gauss-Newton method for low-rank signal estimation. *Numer. Linear Algebra Appl.* **e2428**.

Jonathan Gillard
School of Mathematics
Cardiff University
UK
E-mail address: gillardjw@cardiff.ac.uk

Konstantin Usevich
Université de Lorraine, CNRS
CRAN, Nancy
France
E-mail address: konstantin.usevich@univ-lorraine.fr