Statistics and Its Interface Volume 10 (2017) 19–25

Simple nuclear norm based algorithms for imputing missing data and forecasting in time series

Holly Butcher and Jonathan Gillard*

There has been much recent progress on the use of the nuclear norm for the so-called matrix completion problem (the problem of imputing missing values of a matrix). In this paper we investigate the use of the nuclear norm for modelling time series, with particular attention to imputing missing data and forecasting. We introduce a simple alternating projections type algorithm based on the nuclear norm for these tasks, and consider a number of practical examples.

AMS 2000 subject classifications: Primary 62M10, 62M15; secondary 62P99.
Keywords and phrases: Nuclear norm, Time series analysis, Structured low rank approximation.

1. INTRODUCTION

Many modern approaches of time series analysis involve an initial embedding of a vector of observations into a structured (typically Hankel) matrix [17]. Naturally, properties of this matrix infer properties of the vector of observations. For example, the rank of this structured matrix can be considered as the complexity of a model that would fit the data exactly. The problem of structured low rank approximation involves finding a low rank approximation of a given structured matrix which also preserves the original structure. The rank of the approximation can be selected by the user to balance: (i) the complexity of the approximation and (ii) the accuracy of the approximation. Typically as the rank of the approximation increases, so does its complexity and accuracy. The rank of the approximation needs to be selected to keep the complexity small with accuracy acceptable for the user.

Existing methods for structured low rank approximation are formulated as highly non-convex optimization problems, for which there are no efficient methods to approximate the global optimum [18]. Typically problems are difficult with the objective functions used possessing many local minima and large Lipschitz constants [10, 11].

The nuclear norm as an alternative for the rank has recently been proposed as a possible convex relaxation. Briefly, the nuclear norm of a matrix is the sum of its singular values. The nuclear norm used as a heuristic for solving rank minimization problems was described in [19]. The nuclear norm has the desirable property in that it is the tightest relaxation of the rank.

The objective of this paper is to investigate the potential of the nuclear norm heuristic for imputing missing data and forecasting time series. There are a few recent papers on the theoretical properties of the nuclear norm for the so-called matrix completion problem but to the best of the authors' knowledge there is little to nothing on the potential of the nuclear norm for practical problems of time series analysis.

The structure of the paper is as follows. Section 2 defines the problem considered in this paper, and introduces much of the notation used. We define the optimization problem to be solved under two possible feasible domains and motivate the use of each feasible domain. Section 3 introduces the methodology investigated in this paper, with Section 4 containing examples and discussion. We conclude the paper in Section 5.

2. STATEMENT OF THE PROBLEM

2.1 Notation

Let $Y = (y_1, y_2, \ldots, y_N)^T \in \mathbb{R}^N$ be a vector of observations such that

$$y_n = s_n + \varepsilon_n,\ n = 1, \ldots, N,$$

where $S = (s_1, s_2, \ldots, s_N)^T \in \mathbb{R}^N$ are the unobserved 'true' values, and $\{\varepsilon_n, n = 1, \ldots, N\}$ is a realization of a white noise process. Some observations in Y may be missing. Let Ω denote the set of indices of Y that are observed i.e. $\Omega = \{i : y_i \text{ is observed}, i = 1, \ldots, N\}$. In this paper we do not formally distinguish between missing data and forecasting. Very informally we consider the problem of forecasting to be imputing missing data at the end of the vector.

Let L, K and r be given positive integers such that $1 \leq r \leq L \leq K$. Denote the set of all real-valued $L \times K$ matrices by $\mathbb{R}^{L\times K}$. Let $\mathcal{M}_r = \mathcal{M}_r^{L\times K} \subset \mathbb{R}^{L\times K}$ be the subset of $\mathbb{R}^{L\times K}$ containing all matrices with rank $\leq r$, and $\mathcal{H} = \mathcal{H}^{L\times K} \subset \mathbb{R}^{L\times K}$ be the subset of $\mathbb{R}^{L\times K}$ containing matrices of some known structure. The set of structured $L \times K$ matrices of rank $\leq r$ is $\mathcal{A} = \mathcal{M}_r \cap \mathcal{H}$.

*Corresponding author.

In this paper we only consider the case where $\mathcal{H}$ is the set of Hankel matrices. Recall that a matrix $\mathbf{X} = (x_{lk}) \in \mathbb{R}^{L\times K}$ is called Hankel if $x_{lk} = \text{const}$ for all pairs (l,k) such that $l+k = \text{const}$; that is, all elements on the anti-diagonals of $\mathbf{X}$ are equal. There is a one-to-one correspondence between $L \times K$ Hankel matrices and vectors of size $N = L+K-1$. This correspondence is fundamental to the proposed methodology. For a vector $Y = (y_1, \ldots, y_N)^T$, the matrix $\mathbf{X} = \mathbb{H}(Y) = (x_{lk}) \in \mathbb{R}^{L\times K}$ with elements $x_{lk} = y_{l+k-1}$ is Hankel and vise-versa: for any matrix $\mathbf{X} \in \mathcal{H}$, we may define $Y = \mathbb{H}^{-1}(\mathbf{X})$ so that $\mathbf{X} = \mathbb{H}(Y)$. The non-missing elements of $\mathbf{X} = (x_{lk}) = \mathbb{H}(Y)$ are given by the indices (l,k) such that $l+k-1 \in \Omega, l = 1,\ldots,L,\ k = 1,\ldots,K$.

To project a matrix $\mathbf{X} \in \mathbb{R}^{L\times K}$ onto $\mathcal{M}_r$ we may use the singular value decomposition. Let $\sigma_i = \sigma_i(\mathbf{X})$, the singular values of $\mathbf{X}$, be ordered so that $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_L$. Denote $\mathbf{\Sigma}_0 = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_L)$ and $\mathbf{\Sigma} = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r, 0, \ldots, 0)$. Then the SVD of $\mathbf{X}$ can be written as $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}_0\mathbf{V}^T$, where columns U_l of the matrix $\mathbf{U} \in \mathbb{R}^{L\times L}$ are the left singular vectors of $\mathbf{X}$ and columns V_l of the matrix $\mathbf{V} \in \mathbb{R}^{K\times L}$ are the right orthonormal singular vectors $V_l = \mathbf{X}^T U_l/\sigma_l$. The matrix

$$\pi^{(r)}(\mathbf{X}) = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{i=1}^{r} U_i Z_i^T \quad \text{with} \quad Z_i^T = U_i^T\mathbf{X}$$

belongs to $\mathcal{M}_r$ and minimizes the squared distance $||\mathbf{X} - \tilde{\mathbf{X}}||_F^2$ over $\tilde{\mathbf{X}} \in \mathcal{M}_r$, see [9] or [12, 14]. The projection $\pi^{(r)}(\mathbf{X})$ of $\mathbf{X}$ onto $\mathcal{M}_r$ is uniquely defined if and only if $\sigma_r > \sigma_{r+1}$. It follows that

$$||\mathbf{X} - \pi^{(r)}(\mathbf{X})||_F^2 = \sum_{i=r+1}^{L} \sigma_i^2(\mathbf{X})\,.$$

The nuclear norm of $\mathbf{X}$, denoted $||\mathbf{X}||_*$ is simply the sum of the singular values of $\mathbf{X}$. Using the notation already given, this is

$$||\mathbf{X}||_* = \sum_{i=1}^{L} \sigma_i(\mathbf{X})\,.$$

2.2 Use of the nuclear norm

The nuclear norm has gained much recent attention in a variety of applications. One of the earliest papers which used the nuclear norm to much success was [7] where they describe the so-called problem of matrix completion. This is the problem of recovering, often exactly, a matrix with only selected values observed (subject to some conditions). The overwhelming majority of algorithms in matrix completion are based on finding the matrix with minimum nuclear norm that fits the observed data. The nuclear norm is the closest convex relaxation of the non-convex rank of a matrix (as described earlier). The logic of minimizing the nuclear norm is because in many instances the matrix we wish to recover is at least approximately of low rank. We describe two classic examples below, described in more detail in [7]. Other key papers which develop the theoretical underpinnings of the nuclear norm include [5], [6] and [20]. There are many papers which now provide practical algorithms for minimizing the nuclear norm, but there are too many to give a complete list in this paper.

Example 1. *Netflix.*

Netflix users (rows of a matrix) rate a small amount of movies (columns of a matrix). The amount of movies a user can rate is very small compared to the amount available on Netflix, but nevertheless it is desirable to 'complete the matrix' i.e. impute the missing ratings so that Netflix may recommend movies to a particular user. It can be argued that the matrix of all user ratings is at least approximately of low rank as perhaps only a few (perhaps latent) factors contribute to a viewer's preferred type of movie.

Example 2. *Triangulation from incomplete data.*

Suppose we wish to locate sensors placed on a two-dimensional plane. Each sensor is only able to construct distance estimates from other nearest sensors. From these distance estimates we can produce an estimated distance matrix where some distances are missing. Since the sensors are located in the plane then the rank of the true distance matrix is two. Again we wish to 'complete the matrix' to reconstruct the positions of the objects.

2.3 Statement of the optimization problem

In this paper we consider an optimization problem with two possible feasible domains. Assume we are given a matrix $\mathbf{X}_\bullet = (x_{\bullet,l,k}) \in \mathcal{H}^{L\times K}$. The optimization problem can be defined as:

$$\min_{\mathbf{X}\in\mathcal{D}} ||\mathbf{X}||_* \tag{1}$$

where, in the literature, there are usually two choices of the feasible domain $\mathcal{D}$ which we will call feasible domain 1 and feasible domain 2 respectively.

Feasible domain 1 is given by:

$$\mathcal{D} = \{\mathbf{X} \in \mathcal{H} : ||\mathbf{X} - \mathbf{X}_\bullet||_F < \tau\} \tag{2}$$

Feasible domain 2 is given by:

$$\begin{aligned}\mathcal{D} = \{&\mathbf{X} \in \mathcal{H} : ||\mathbf{X} - \mathbf{X}_\bullet||_F < \tau,\ x_{l,k} = x_{\bullet,l,k} \\ &\text{for } l+k-1 \in \Omega, l = 1,\ldots,L,\ k = 1,\ldots K\}\end{aligned} \tag{3}$$

Note that if Ω is the empty set, then problem (1) is the same with either of the feasible domains given above. The difference from a modelling point of view between these feasible domains is the following. Suppose we wish to simultaneously impute the missing data and then find a lower complexity approximation of the given data. In this scenario, feasible domain 1 is appropriate. It may be the case that all non-missing observations have been observed exactly, or we may have some other motivation for leaving them unchanged. We do however wish to impute the missing data. In this scenario, feasible domain 2 is appropriate.

3. METHODOLOGY

3.1 Minimizing the nuclear norm

Consider the optimization problem (1) with feasible domain (2) or (3). Let $\widehat{\mathbf{X}}$ be the solution of (1) with feasible domain (2), and let $\widehat{\mathbf{X}}^{\Omega}$ be the solution of (1) with feasible domain (3). For convenience of notation we suppose the optimization problem (1) defines the mappings $\mathbf{X} \to \widehat{\mathbf{X}}$ and $\mathbf{X} \to \widehat{\mathbf{X}}^{\Omega}$ for the feasible domains (2) and (3) respectively. We denote the mapping $\mathbf{X} \to \widehat{\mathbf{X}}$ by $\pi_*(\mathbf{X})$ and the mapping $\mathbf{X} \to \widehat{\mathbf{X}}^{\Omega}$ by $\pi_*^{\Omega}(\mathbf{X})$.

There is some guidance on how to select a value for τ in (1). For example, one potential choice (which will be used in this paper) is to take

$$\tau = c\,||\mathbf{X}_* - \pi^{(r)}(\mathbf{X}_\bullet)||_F \tag{4}$$

for some r and constant $c \geq 1$. The constant c may be used to soften the constraint but we will consider $c = 1$ in all examples that follow. This choice of τ, in some sense, ensures that the nuclear norm approximation is at least as close to the original matrix $\mathbf{X}_\bullet$ as the rank r approximation to $\mathbf{X}_\bullet$. This choice of τ, as well as some alternatives, are described in [18].

Modern approaches in approximating a solution to (1) typically involve the use of semi-definite programming. Specific details of how (1) may be framed as a semi-definite program are included in [21]. There now exists an array of different convex optimization solvers which are capable of solving semi-definite programs. In this paper we use the popular CVX software for MATLAB [8, 15].

3.2 Projection to the space of Hankel matrices

The space $\mathcal{H} = \mathcal{H}^{L\times K}$ of $L\times K$ Hankel matrices is a linear subspace of $\mathbb{R}^{L\times K}$. The closest Hankel matrix (in Frobenius norm) to any given matrix is obtained by using the diagonal averaging procedure.

Every $L \times K$ Hankel matrix $\mathbf{X} \in \mathcal{H}$ is in a one-to-one correspondence with some vector $Y = (y_1, \ldots, y_N)^T$ described by the function $\mathbb{H} : \mathbb{R}^N \to \mathcal{H}^{L\times K}$ which is defined by $\mathbb{H}(Y) = ||y_{l+k-1}||_{l,k=1}^{L,K}$ for $Y = (y_1, \ldots, y_N)^T$. Each element of the vector Y is repeated in $\mathbf{X} = \mathbb{H}(Y)$ several times. Let $\mathbf{E} = (e_{lk}) \in \mathbb{R}^{L\times K}$ be the matrix consisting entirely of ones. We can compute the sum of each anti-diagonal of $\mathbf{E}$, denoted t_n, as

$$t_n = \sum_{l+k=n+1} e_{lk} = \begin{cases} n & \text{for } n = 1, \ldots, L-1, \\ L & \text{for } n = L, \ldots, K-1, \\ N-n+1 & \text{for } n = K, \ldots, N. \end{cases}$$

The value t_n is the number of times the element y_n of the vector Y is repeated in the Hankel matrix $\mathbb{H}(Y)$. Let $\pi_{\mathcal{H}}(\mathbf{X})$ denote the projection of $\mathbf{X} \in \mathbb{R}^{L\times K}$ onto the space $\mathcal{H}$. Then the element $\tilde{x}_{ij}$ of $\pi_{\mathcal{H}}(\mathbf{X})$ is given by

$$\tilde{x}_{ij} = t_{i+j-1}^{-1} \sum_{l+k=i+j} x_{lk}\,.$$

3.3 Proposed algorithm

Assume we are given a matrix $\mathbf{X}_\bullet \in \mathcal{H}$ and a set Ω of observed indices. Set $\mathbf{X}_0 = \mathbf{X}_\bullet$. In this paper we investigate the following algorithm, which we call A1.

Set $\mathbf{X}_1 = \pi_{\mathcal{H}}(\pi_*^{\Omega}(\mathbf{X}_0))$. For $n = 2, 3, \ldots$

$$\mathbf{X}_n = \pi_{\mathcal{H}}(\pi_*(\mathbf{X}_{n-1})).$$

Algorithm A1 is an example of a so-called alternating projection algorithm [1]. Such algorithms are often used for computing or approximating a point in the intersection of two or more convex sets. Both of the projections used in algorithms A1 are to convex spaces and hence algorithm A1 will linearly converge to an approximation of a point in the intersection of the two convex spaces [3]. Note however that we may not converge to an optimal point [16]. This will be investigated in another forthcoming paper. It is possible that algorithm A1 converges to a matrix of low rank, but this is often not the case in practical examples. To a certain extent the proposed methodology is 'model-free' although it is known that one is more likely to obtain an approximation of low rank if the $\mathbf{X}_\bullet$ is (approximately) already of low rank.

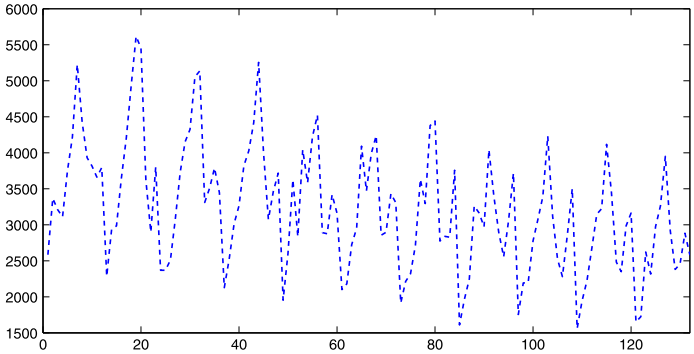
The logic for the nature of our algorithm is as follows. If algorithm A1 is run for just one iteration, then we assume that the non-missing observations are 'exact' and they are not to be changed. Only the missing values are then imputed. If the algorithm is run for more than one iteration, then we also change the non-missing observations. This is to allow our algorithm to deal with the two possible modelling scenarios as motivated in Section 2.3. The user may select which scenario is appropriate for their application by controlling the number of iterations the algorithm is run for.

4. EXAMPLES

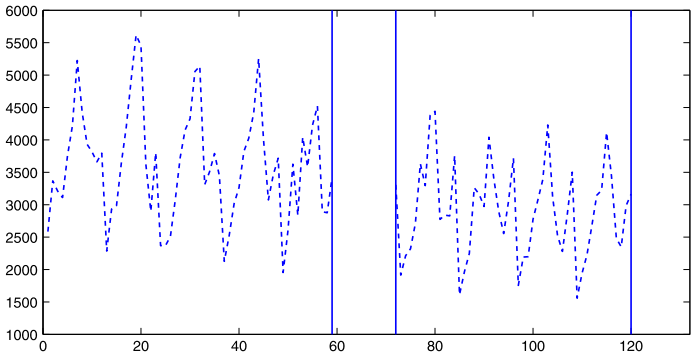
4.1 Fortified wine

In this section we consider the classical time series 'Fortified Wine', where 132 observations depict the monthly volumes of wine sales in the period from January 1984 until January 1994. In order to demonstrate our methodology, we removed 12 known values, starting at the 60th point, that is, we assume that the values for 12 months are unknown. Also, to imitate a forecast we replace the last 12 observed values by 12 missing values. Figure 1 contains a plot of the time series, with (and without) the missing data.

In this example we compute the mean square error (MSE) obtained by imputing the missing values by algorithm A1 with $L = 36$. Denote by $\mathbf{X}_\bullet$ the Hankel matrix of our observations. We consider two values of τ (see (4)), namely $\tau_1 = ||\mathbf{X}_* - \pi^{(10)}(\mathbf{X}_\bullet)||_F$ and $\tau_2 = ||\mathbf{X}_* - \pi^{(15)}(\mathbf{X}_\bullet)||_F$. Ta-



(a) Initial complete time series



(b) Time series with missing data

Figure 1. Monthly volumes of fortified wine sales from January 1984 until January 1994.

Table 1. MSE with $\tau = \tau_1$

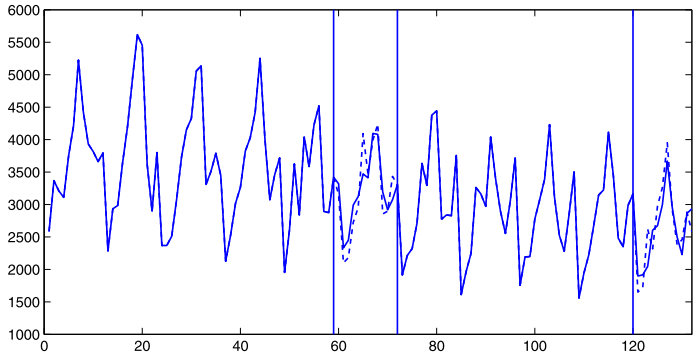
Iteration	Middle	End	Total
1	277.0879	280.8282	118.9508
2	274.1852	281.0847	259.4018
3	283.8793	307.8987	326.3798
4	301.1693	328.9423	359.8569
5	315.6676	343.0752	380.0125

Table 2. MSE with $\tau = \tau_2$

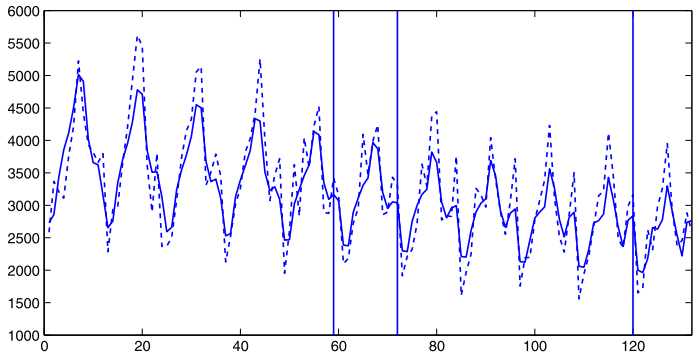
Iteration	Middle	End	Total
1	277.0879	280.8282	118.9508
2	274.5200	284.0636	205.9513
3	268.5309	296.3649	262.2150
4	268.0295	306.9925	291.0720
5	269.8995	315.9035	307.9741

bles 1 and 2 contain the MSE values for the approximated missing middle section of the data, the approximated missing end section of the data, and for the entire approximation with $\tau = \tau_1$ and $\tau = \tau_2$. We present the MSE values for five iterations of algorithm A1. The total MSE is small after one iteration as the non-missing observations are not altered, and they thus add nothing to the MSE.

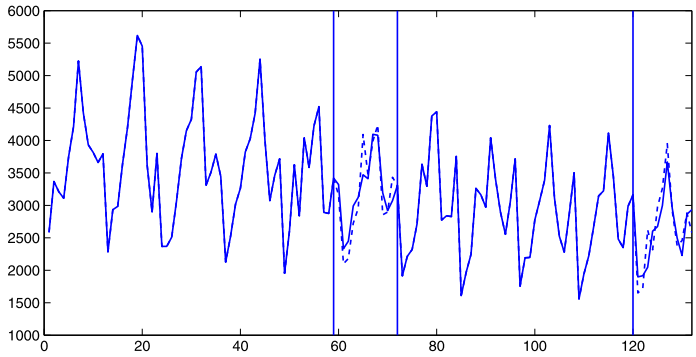
Figure 2 contains plots of the original data and approximations for $\tau = \tau_1$ and $\tau = \tau_2$ after one iteration, and after



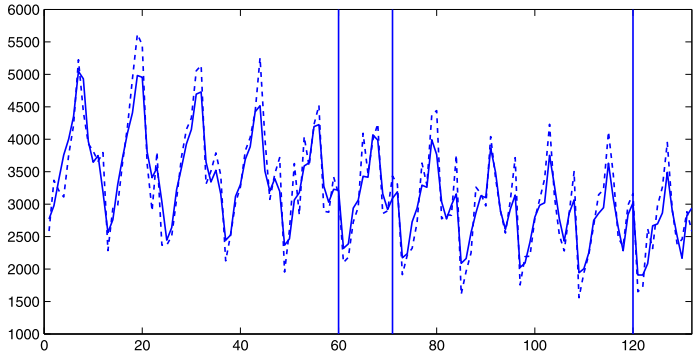
(a) $\tau = \tau_1$, 1 iteration



(b) $\tau = \tau_1$, 5 iterations



(c) $\tau = \tau_2$, 1 iteration



(d) $\tau = \tau_2$, 5 iterations

Figure 2. Monthly volumes of fortified wine sales from January 1984 until January 1994 with different τ and iterations. Original data in dashed line, approximation in bold line.

Table 3. *MSEs from [14] using the methodology in [13]*

	Middle	End	Total
Method 1, $L = 36$	255.9	292.8	275.0
Method 2, $L = 36$	221.2	333.0	282.7
Method 2, $L = 60$	216.2	419.3	333.6

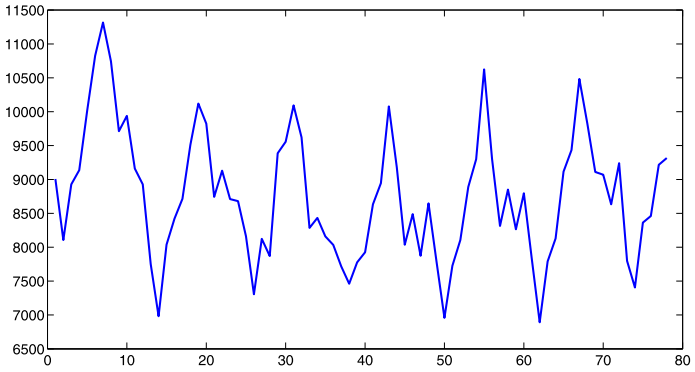


Figure 3. Monthly accidental deaths in the USA between 1973 and 1978.

five iterations of algorithm A1. The missing sections are also highlighted.

It can be seen that the results found from using algorithm A1 are at the least competitive with those already reported in the literature. Table 3 contains the MSE's obtained from [14] using two variations of the methodology described in [13] for the same data considered in this section. It can be seen that our algorithm is at least as competitive as other algorithms that have been applied to the same data.

4.2 Accidental deaths in the USA

In this section we consider forecasting the famous 'death' series recording the monthly accidental deaths in the USA between 1973 and 1978. This data has been studied by many authors and can be found in a number of time series data libraries. We wish to replicate the exercise given in [4] which aimed to forecast the final six values of this series. The time series contains a total of 78 observations. We truncate the series to the first 72 observations and will forecast the remaining six observations. Figure 3 contains a plot of the entire time series.

Table 4 contains forecasts of the final six data points of the data series by several methods along with the square root of the mean square error. These results are taken from [4] and full details of the fitted models can be found within. In summary Model I and Model II are examples of SARIMA models as described by [2]. Model I is given by

$$\nabla_{12} y_n = 28.831 + (1 - 0.478B)(1 - 0.588B^{12})Z_n$$

and Model II is given by

Table 4. *Forecasted 6 observations with their MSE's*

	1	2	3	4	5	6	$\sqrt{MSE}$
Orig. data	7798	7406	8363	8460	9217	9316	
Mean	8665	8706	8565	8906	9028	9335	673.2
Model I	8227	7970	8283	8773	9913	10019	514.0
Model II	8245	7802	8206	8503	9833	9910	431.3
HWS	7995	7432	7630	7910	9175	9214	385.3
ARAR	8054	7748	8162	8700	9127	9466	227.7
Alg. A1	7970	7670	7895	8095	9000	9277	288.7

$$(1 - B)(1 - B^{12})y_n = 28.831 + Z_n - 0.596Z_{n-1} - 0.407Z_{n-6} - 0.685Z_{n-12} + 0.460Z_{n-13}$$

where Z_n is a realisation of white noise with zero mean and variance 0.9439, B is the backward shift operator defined as: $B^j Z_n = Z_{n-j}$, and $\nabla_{12} y_n = y_n - y_{n-12}$. HWS represents the model as fitted by the Holt-Winter seasonal algorithm. ARAR represents the model as fitted by transforming the data prior to fitting an autoregressive model. Full details of this model are omitted in this paper, but are included in [4].

For this example we also forecast ahead six points using algorithm A1 with $L = 24$ and $\tau = ||\mathbf{X}_* - \pi^{(15)}(\mathbf{X}_\bullet)||_F$, where $\mathbf{X}_\bullet$ is the Hankel matrix of our observations. Our forecasted values are given in Table 4. Figure 4 contains plots of the end section of the 'death' series with forecasts from a number of methods. In this example algorithm A1 has outperformed many of the classical methods of time series analysis.

4.3 Climate

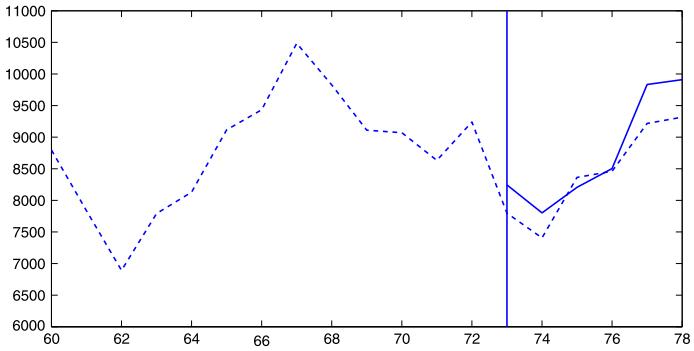
In this example, we simply offer some forecasts of the Earth temperature recordings downloaded from vortex.nsstc.uah.edu/data/msu/t2lt/oldversions/uahncdc_lt_5.5.txt. A plot of the full data is given in Figure 5.

We have taken the last 97 points of the global earth temperature recordings (June 2006–June 2014). This data is provided by the National Space Science & Technology Center and was suggested by the editors of this special edition as a time series to analyse. Using algorithm A1 we predict the next 36 months of data with varying window length and τ chosen so that the nuclear norm approximation is at least as good as a rank 20 approximation. That is $\tau = ||\mathbf{X}_* - \pi^{(20)}(\mathbf{X}_\bullet)||_F$, where $\mathbf{X}_\bullet$ is the Hankel matrix of our observations.

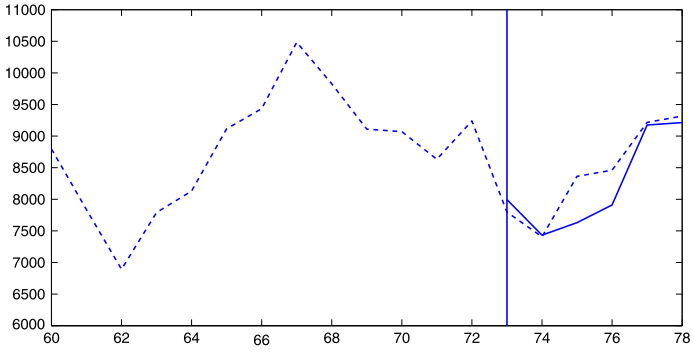
We provide these forecasts without comment, but include them to offer comparisons with other papers in this special edition investigating the same data.

5. CONCLUSION

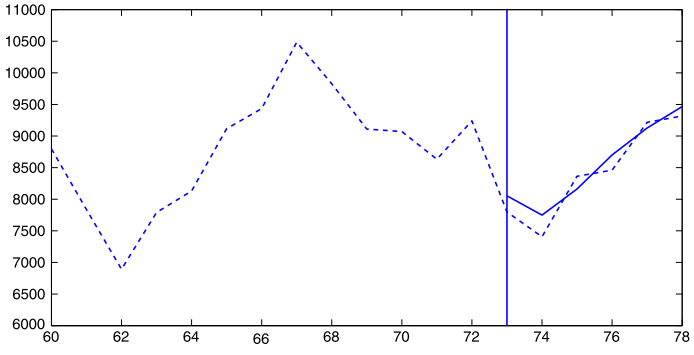
In this paper we have demonstrated the potential of the nuclear norm for two common tasks of time series analysis, namely imputing data and forecasting. We introduced a simple algorithm based on the algorithm of alternating projections but remark that future research will concentrate on



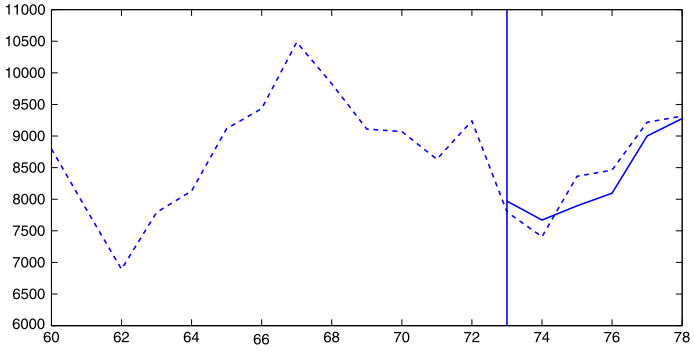
(a) Model II



(b) HWS



(c) ARAR



(d) Algorithm A1

Figure 4. Forecasting monthly accidental deaths in the USA with different methods. Original data in dashed line, forecast in bold line.

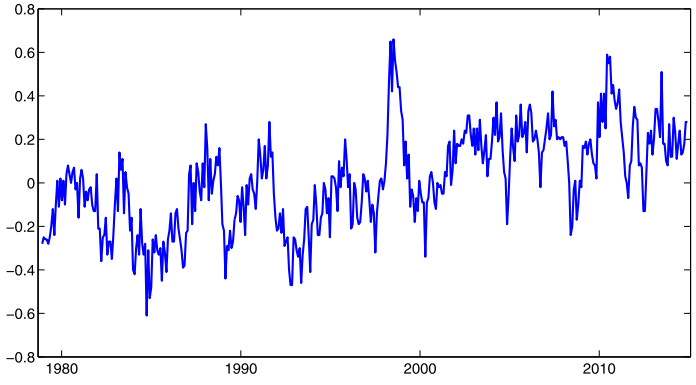
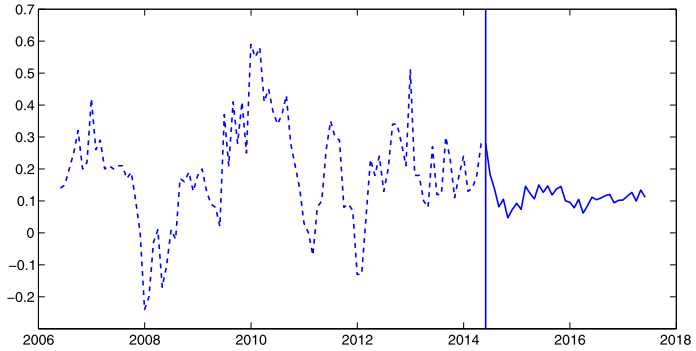
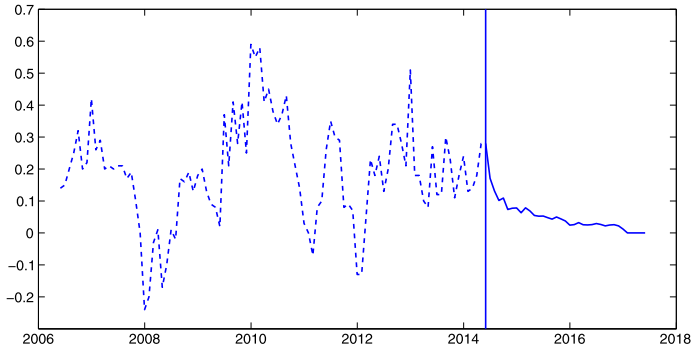


Figure 5. Earth temperature recorded monthly, 1978–2014.



(a) Forecast using $L = 48$



(b) Forecast using $L = 32$

Figure 6. Forecasting global temperature recordings with varying window length. Original data in dashed line, forecast in bold line.

investigating its theoretical properties. It is possible to develop alternative algorithms on the basis of the projections introduced in this paper. In the comparative examples considered, the introduced algorithm was at least competitive with classical methodologies. There remain open questions which will be investigated in future works. These include: the choice of τ and its influence on the accuracy of the approximation and convergence, and how it appears that the fit of the reconstruction worsens upon additional iterations.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of the London Mathematical Society in providing funding for H. Butcher to undertake this research (Undergraduate Research Bursaries in Mathematics, 2014, URB 14-29). Both authors would like to thank Prof. Anatoly Zhigljavsky for his useful comments and suggestions throughout the writing of this paper.

Received 5 January 2015

REFERENCES

[1] Bauschke, H., and Borwein, J. M. On the convergence of von neumann's alternating projection algorithm for two sets. *Set-Valued Analysis 1*, 2 (1993), 185–212. MR1239403

[2] Box, G. E., Jenkins, G. M., and Reinsel, G. C. *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 2013. MR2419724

[3] Boyd, S., and Vandenberghe, L. *Convex Optimization*. Cambridge University Press, 2009. MR2061575

[4] Brockwell, P. J., and Davis, R. A. *Introduction to Time Series and Forecasting*, vol. 1. Taylor & Francis, 2002. MR1894099

[5] Candes, E., and Plan, Y. Matrix completion with noise. *Proceedings of the IEEE* (2010).

[6] Candès, E., and Tao, T. The power of convex relaxation: Near-optimal matrix completion. *Information Theory, IEEE Transactions on* (2010), 1–51. MR2723472

[7] Candès, E. J. E., and Recht, B. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics 9*, 6 (Apr. 2009), 717–772. MR2565240

[8] CVX Research, I. CVX: Matlab software for disciplined convex programming, version 2.0. http://cvxr.com/cvx, Aug. 2012.

[9] Eckart, C., and Young, G. The approximation of one matrix by another of lower rank. *Psychometrika 1* (1936), 211–218.

[10] Gillard, J., and Zhigljavsky, A. Analysis of structured low rank approximation as an optimization problem. *Informatica 22*, 4 (2011), 489–505. MR2885683

[11] Gillard, J., and Zhigljavsky, A. Optimization challenges in the structured low rank approximation problem. *Journal of Global Optimization 57*, 3 (2013), 733–751. MR3119378

[12] Golyandina, N., Nekrutkin, V., and Zhigljavsky, A. A. *Analysis of Time Series Structure: SSA and Related Techniques*. CRC Press, 2010. MR1823012

[13] Golyandina, N., and Osipov, E. The "caterpillar"-ssa method for analysis of time series with missing values. *Journal of Statistical Planning and Inference 137*, 8 (2007), 2642–2653. MR2326115

[14] Golyandina, N., and Zhigljavsky, A. *Singular Spectrum Analysis for Time Series*. Springer, 2013. MR3024734

[15] Grant, M., and Boyd, S. Graph implementations for nonsmooth convex programs. In *Recent Advances in Learning and Control*, V. Blondel, S. Boyd, and H. Kimura, Eds., Lecture Notes in Control and Information Sciences. Springer-Verlag Limited, 2008, pp. 95–110. http://stanford.edu/~boyd/graph_dcp.html. MR2409077

[16] Lewis, A., and Malick, J. Alternating projections on manifolds. *Mathematics of Operations Research* (2008), 1–25. MR2393548

[17] Markovsky, I. Structured low-rank approximation and its applications. *Automatica 44*, 4 (Apr. 2008), 891–909. MR2530933

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

[19] Recht, B., Fazel, M., and Parrilo, P. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review* (2010), 1–32. MR2680543

[20] Recht, B., Xu, W., and Hassibi, B. Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization. *47th IEEE Conference on Decision and Control*, (2008).

[21] Vandenberghe, L., and Boyd, S. Semidefinite programming. *SIAM Review 38*, 1 (1996), 49–95. MR1379041

Holly Butcher
School of Mathematics
Cardiff University
UK
E-mail address: ButcherHL@Cardiff.ac.uk

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