Convex optimization for matrix completion with application to forecasting

Jonathan Gillard* and Konstantin Usevich†

*Cardiff School of Mathematics, Cardiff University, Senghennydd Road, Cardiff, UK, CF24 4AG.
†Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France

Abstract. We consider convex relaxations for the low-rank matrix completion problem with specific application to forecasting time series and consider how close the solution of the convex relaxed low-rank matrix completion problem is to the original global optimization problem. This is a fashionable approach in the statistics of big data, difficult non-convex optimization problems are 'convexified' to make them tractable. The question then is: how close is the solution of the convex optimization problem to the non-convex one (which is the one we really want to solve)? We consider a matrix completion problem for Hankel matrices and investigate some cases when the proposed approach can work through theoretical and empirical results.

Keywords: Structured matrix completion, Hankel matrices, nuclear norm
PACS: 02.60.Pn, 02.60.Cb.

PROBLEM STATEMENT

Introduction

In this paper we consider the low-rank matrix completion problem (an intractable, global optimization problem) and consider a convex relaxation to it. The natural question to ask is how close the solution of the convex relaxation is to the original global optimization problem. The main application we consider in this paper is forecasting in time series analysis. Towards the end of the paper we consider forecasting 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. Our forecasting approach produces superior forecasts compared to many classical methods.

Hankel matrices

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

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

We pose the problem of forecasting a given time series as the low-rank matrix completion of a Hankel matrix. Let

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

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

(parameterized by $\mathbf{p}$) we consider is

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

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

In the following subsections, we describe the formal statement of the matrix completion problem, and the convex relaxation approach that we propose.

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 the Hankel matrix structure (2), the Structured Low-Rank Matrix Completion (SLRMC) problem is posed as

$$\tilde{\mathbf{p}} = \arg\min_{\mathbf{p}\in\mathbb{R}^{(n+m)}} \operatorname{rank} \mathscr{S}(\mathbf{p}) \ \text{ subject to } \ \mathbf{p}_{(1:n)} = \mathbf{p}_0\,. \tag{3}$$

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 [5].

The matrix completion problem (3) for general matrix structures is NP-hard (see [3] and [8]). A convex relaxation of this problem based on the nuclear norm became increasingly popular recently. 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 $\mathbf{X}$. A convex relaxation of (3) is obtained by replacing the rank with the nuclear norm:

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}\in\mathbb{R}^{(n+m)}} \|\mathscr{S}(\mathbf{p})\|_* \ \text{ subject to } \ \mathbf{p}_{(1:n)} = \mathbf{p}_0\,. \tag{4}$$

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 (a low-rank solution). The nuclear norm is the convex envelope of the rank of a matrix. The minimum of the convex envelope can serve as a lower bound on the true minimum, and the minimizing argument can serve as an initial point for a more complicated non-convex local search, if needed.

For the Hankel matrix case, the solution of (3) is known. Still, the performance of the nuclear norm relaxation is important for understanding the behavior of forecasting in the approximate case to be introduced in the next subsection.

Approximate low-rank matrix completion

Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ be a vector of length n. We denote

$$\|\mathbf{x}\|_W = \sqrt{\sum_{i=1}^{n} w_i x_i^2} \tag{5}$$

where $W = (w_1, w_2, \ldots, w_n)$, is a vector of weights, $w_i > 0$ for $i = 1, \ldots, n$. For advice on the suitable selection of weights, see [7]. The approximate rank minimization can be posed as follows. Given $\mathbf{p}_0$, W, $m \geq 0$ and $\tau \geq 0$ find

$$\tilde{\mathbf{p}} = \arg\min_{\mathbf{p}\in\mathbb{R}^{(n+m)}} \operatorname{rank} \mathscr{S}(\mathbf{p}) \ \text{ subject to } \ \|\mathbf{p}_{(1:n)} - \mathbf{p}_0\|_W \leq \tau \tag{6}$$

The parameter τ controls the precision of approximation. Unlike (3), the problem (6) does not have a known solution. In fact, it is a dual problem to structured low-rank approximation [10] which is known to be a difficult optimization problem [4, 6].

In order to circumvent the complexity of the problem, we consider the following relaxation of (6), using the nuclear norm:

$$\mathbf{p}_* = \underset{\mathbf{p}\in\mathbb{R}^{(n+m)}}{\arg\min} \|\mathscr{S}(\mathbf{p})\|_* \text{ subject to } \|\mathbf{p}_{(1:n)} - \mathbf{p}_0\|_W \le \tau\,. \tag{7}$$

EXAMPLES OF THEORETICAL RESULTS

The performance of the nuclear norm (i.e. when the solution of (4) coincides with the solution of (3)), was studied recently for a special case of the structure (2):

$$\mathscr{S}(\mathbf{p}) = \begin{pmatrix} p_1 & p_2 & \cdots & p_n \\ p_2 & p_3 & \iddots & p_{n+1} \\ \vdots & \iddots & \iddots & \vdots \\ p_n & p_{n+1} & \cdots & p_{n+m} \end{pmatrix}, \tag{8}$$

i.e. when the matrix is square ($L = K$), and all the values below the main antidiagonal of the Hankel matrix $\mathscr{S}(\mathbf{p})$ are missing (i.e. $L = n = m+1$). The first result, appeared in [2] and refined in [11], treats the rank-one case and is given below.

Theorem 0.1 ([11, Theorem 6]) *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 $|\lambda| \le 1$, the solution of (3), i.e.*

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

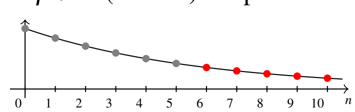
  is also a solution of

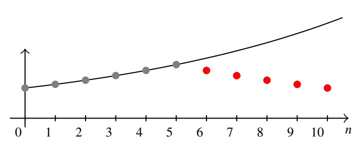
  $$\widehat{\mathbf{p}} = \underset{\mathbf{p}\in\mathbb{C}^{(n+m)}}{\arg\min} \|\mathscr{S}(\mathbf{p})\|_* \text{ subject to } \mathbf{p}_{(1:n)} = \mathbf{p}_0\,; \tag{9}$$

  in particular, if $|\lambda| < 1$, then the solution of (9) is unique.
- *If $|\lambda| > 1$, then the unique solution of (9) is given by*

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

An illustration of the above theorem is given below where forecasts (in red) given by the nuclear norm for the time series $p_k = c\rho^k$, $n = 6$, with $\rho < 1$ (top) and $\rho > 1$ (bottom) are provided.





Theorem 0.2 ([11, Theorem 7]) *Consider a complex-valued infinite time series $(p_1, p_2, \ldots, p_n, \ldots)$ with*

$$p_k = \sum_{j=1}^{s} P_j(k)\lambda_j^k, \quad k = 1, 2, \ldots, \tag{10}$$

where $P_j(k)$ are complex polynomials of degrees $\nu_j - 1$, (ν_j are positive integers), and $\lambda_j \in \mathbb{C}\setminus\{0\}$. Fix the structure (8) (i.e. $L = K = n = m+1$ in (2)) and $r \le \frac{n}{2}$. Then there exists a number $0 < \rho_{max,r,m} < 1$ such that for any complex-valued time series of finite rank r (10) with $|\lambda_j| < \rho_{max,r,m}$, the solution of (9) is unique and coincides with the solution of the non-convex rank minimization problem (6).

Theorem 0.2 implies that the same result holds for the real-valued problems (3) and (4). New results in a similar vein to the two described above are given in [5].

EXAMPLE: FORECASTING DEATHS

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 (such as [9]) and can be found in a number of time series data libraries. We wish to replicate the exercise given in [1] 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. Denote these series of 72 observations by $\mathbf{p}_0$. We consider (7) with $n = 72$ and $m = 6$. The table below 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 [1] and full details of the fitted models can be found within. 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. We compare the results of these forecasts with those obtained from (7) with $L = 24$ and the weights $w_i = \exp(\alpha i)$ with $\alpha = 0.05$. We take τ to be the solution of

$$X_\tau = \operatorname*{arg\,min}_{X \in \mathbb{R}^{L\times K},\, \mathrm{rank}(X)=12} \|\mathscr{H}_L(\mathbf{p}_0) - X\|_F, \;\; \mathbf{p}_\tau = \operatorname*{arg\,min}_{\mathbf{p}\in\mathbb{R}^n} \|\mathscr{H}_L(\mathbf{p}) - X_\tau\|_F, \;\; \tau = \|\mathbf{p}_\tau - \mathbf{p}_0\|_W.$$

so that the forecast is 'as least as good' as that obtained from an unstructured low-rank approximation of rank 12. The forecast obtained from this procedure is labelled 'NM'. For more details and a more through discussion of parameter selection, see [5].

	1	2	3	4	5	6	$\sqrt{MSE}$
Original data	7798	7406	8363	8460	9217	9316	
HWS	8039	7077	7750	7941	8824	9329	401.263
ARAR	8168	7196	7982	8284	9144	9465	253.202
NM	7953	7490	7906	8118	9122	9356	247.380

CONCLUSION

In this paper we have considered matrix completion as a tool for forecasting in time series analysis. We have formulated a nuclear norm relaxation of structured low-rank matrix completion suitable for this purpose, and have demonstrated its practical potential towards the end of the paper. We are able to give specific results for when the solution of the non-convex, global optimization problem (4) coincides with the solution of the convex problem (3).

REFERENCES

1. Peter Brockwell and Richard Davis. *Time series: theory and methods*. Springer Science & Business Media, 2009.
2. Liang Dai and Kristiaan Pelckmans. On the nuclear norm heuristic for a Hankel matrix completion problem. *Automatica*, 51:268–272, 2015.
3. Maryam Fazel. *Matrix rank minimization with applications*. PhD thesis, PhD thesis, Stanford University, 2002.
4. Jonathan Gillard and Dmitri Kvasov. Lipschitz optimization methods for fitting a sum of damped sinusoids to a series of observations. *Statistics and its Interface*, 10(1):59–70, 2017.
5. Jonathan Gillard and Konstantin Usevich. Structured low-rank matrix completion for forecasting in time series analysis. *arXiv preprint arXiv:1802.08242*, 2018.
6. Jonathan Gillard and Anatoly Zhigljavsky. Optimization challenges in the structured low rank approximation problem. *Journal of Global Optimization*, 57(3):733–751, 2013.
7. Jonathan Gillard and Anatoly Zhigljavsky. Weighted norms in subspace-based methods for time series analysis. *Numerical Linear Algebra with Applications*, 23(5):947–967, 2016.
8. Nicolas Gillis and François Glineur. Low-rank matrix approximation with weights or missing data is NP-hard. *SIAM Journal on Matrix Analysis and Applications*, 32(4):1149–1165, 2011.
9. Nina Golyandina, Vladimir Nekrutkin, and Anatoly Zhigljavsky. *Analysis of time series structure: SSA and related techniques*. CRC press, 2001.
10. Ivan Markovsky. *Low Rank Approximation: Algorithms, Implementation, Applications*. Communications and Control Engineering. Springer, 2012.
11. Konstantin Usevich and Pierre Comon. Hankel low-rank matrix completion: Performance of the nuclear norm relaxation. *IEEE Journal of Selected Topics in Signal Processing*, 10(4):637–646, 2016.