Using Hankel structured low-rank approximation for sparse signal recovery

Ivan Markovsky[1] and Pier Luigi Dragotti[2]

Department ELEC
Vrije Universiteit Brussel (VUB)
Pleinlaan 2, Building K, B-1050
Brussels, Belgium
Email: imarkovs@vub.ac.be

EEE Department
Imperial College London
Exhibition Road, SW7-2AZ
London, United Kingdom
Email: p.dragotti@imperial.ac.uk

Abstract. Structured low-rank approximation is used in model reduction, system identification, and signal processing to find low-complexity models from data. The rank constraint imposes the condition that the approximation has bounded complexity and the optimization criterion aims to find the best match between the data—a trajectory of the system—and the approximation. In some applications, however, the data is sub-sampled from a trajectory, which poses the problem of sparse approximation using the low-rank prior. This paper considers a modified Hankel structured low-rank approximation problem where the observed data is a linear transformation of a system's trajectory with reduced dimension. We reformulate this problem as a Hankel structured low-rank approximation with missing data and propose a solution methods based on the variable projections principle. We compare the Hankel structured low-rank approximation approach with the classical sparsity inducing method of ℓ_1-norm regularization. The ℓ_1-norm regularization method is effective for sum-of-exponentials modeling with a large number of samples, however, it is not suitable for damped system identification.

Keywords: low-rank approximation, Hankel structure, sparse approximation, missing data estimation, sum-of-exponentials modeling, ℓ_1-norm regularization.

1 Introduction

The problem considered is defined as follows: Given

- full row rank matrix $A \in \mathbb{R}^{n_\mathrm{g} \times n_p}$ with $n_\mathrm{g} < n_p$,
- vector of measurements b,
- structure specification $\mathscr{S} : \mathbb{R}^{n_p} \to \mathbb{R}^{m \times n}$, and
- rank constraint r,

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{p} \quad \|b - A\widehat{p}\|_2 \\ \text{subject to} & \operatorname{rank}\left(\mathscr{S}(\widehat{p})\right) \leq r. \end{array} \tag{1}$$

The measurements b are obtained as

$$b = A\bar{p} + \widetilde{b},$$

where $\bar{p}$ is a vector that we aim to estimate (the "true value") and $\widetilde{b}$ is zero mean Gaussian measurement noise with covariance matrix that is a multiple of the identity. The prior knowledge that makes the estimation of $\bar{p}$ a well posed problem is that it is sparse in the sense that the matrix $\mathscr{S}(\bar{p})$ has low rank:

$$\operatorname{rank}\left(\mathscr{S}(\bar{p})\right) \leq r. \tag{2}$$

Therefore, we impose the low-rank prior knowledge on the estimate $\widehat{p}$ in problem (1).

Problem (1) is a structured low-rank approximation problem. Its novel element with respect to related problems considered in the literature [4, 5, 3, 2, 14, 12] is that a subset of n_{g} samples are observed. In structured low-rank approximation problem formulations considered in the literature, all n_{p} samples are available for the estimation of $\bar{p}$. Our main result, presented in Section 2, is a reformulation of problem (1) as an equivalent structured low-rank approximation problem with missing data [12]. Section 3 presents a solution method based on the variable projection principle [6].

Section 4 considers the special case of (1) when the structure $\mathscr{S}$ is Hankel. Hankel structured low-rank approximation has applications in computer algebra, system theory, and signal processing. In the case of a Hankel matrix structure, the rank constraint (2) is equivalent to the constraint that the to-be-estimated vector $\bar{p}$ satisfies a recursive relation [9, 11]

$$a_0 p_t + a_1 p_{t+1} + \cdots + a_r p_{t+r} = 0, \quad \text{for } t = 1, \ldots, n_p - r.$$

Equivalently, $(\bar{p}_1, \ldots, \bar{p}_{n_p})$ is a sum-of-polynomials-times-damped-exponentials discrete-time signal [13]. In system theoretic terms, $(\bar{p}_1, \ldots, \bar{p}_{n_p})$ is the output of a discrete-time autonomous linear time-invariant system of order at most r. Therefore, (1) can be viewed as the problem of identifying an autonomous linear time-invariant system from partial noisy measurements that are a linear transformation of a system's output.

We compare the approach of solving the autonomous linear time-invariant system identification problem via (1) with method based on ℓ_1-norm regularization. This latter approach imposes sparsity on the frequency domain representation of the signal. Indeed, an r-sparse frequency domain signal is a sum of r complex exponentials in the time-domain. However, the frequencies are constrained to belong to the grid $\{ k\omega_0 \mid k \in \mathbb{Z} \}$, where $\omega_0 := 2\pi/n_p$. Therefore, the accuracy of the ℓ_1-norm regularization method for autonomous linear time-invariant system identification is limited. Another essential difference between (1) and the ℓ_1-norm approach is that the ℓ_1-norm approach can not deal with damped exponentials and polynomials. Indeed, damping gives rise to "skirts" in the frequency domain, so that the signal is no longer k-sparse in the frequency domain, however, it is sparse in the sense of (2). Section 5 shows numerical examples.

2 Link to missing data estimation

We use the notation $p_{1:n_g}$ for the subvector $\begin{bmatrix} p_1 & \cdots & p_{n_g} \end{bmatrix}^\top$ consisting of the first n_g elements of p.

Theorem 1. *Problem (1) is equivalent to the structured low-rank approximation problem with missing values*

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{p}' \quad \|b - \widehat{p}'_{1:n_g}\|_2 \\ \text{subject to} & \operatorname{rank}\left(\mathscr{S}'(\widehat{p}')\right), \end{array} \tag{3}$$

where

$$\mathscr{S}'(\cdot) := \mathscr{S}(V\cdot) \quad \text{and} \quad \widehat{p}' = V^{-1}\widehat{p},$$

with a nonsingular matrix V, such that $AV = \begin{bmatrix} I_{n_g} & 0 \end{bmatrix}$.

Proof. Using the change of variables $\widehat{p}' = V^{-1}\widehat{p}$, where V is a nonsingular matrix, problem (1) becomes

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{p}' \quad \|b - A'\widehat{p}'\|_2 \\ \text{subject to} & \operatorname{rank}\left(\mathscr{S}'(\widehat{p}')\right) \leq r, \end{array} \tag{4}$$

where $A' = AV$ and $\mathscr{S}'(\cdot) := \mathscr{S}(V\cdot)$. By the full row rank assumption, we can choose V, so that

$$A' = AV = \begin{bmatrix} I_{n_g} & 0 \end{bmatrix}. \tag{5}$$

With this choice of V, problem (4) becomes (3).

Note that if the original structure $\mathscr{S}$ is affine, the new structure $\mathscr{S}'$ is also affine.

Example 1. Let A consists of the first n_g rows of the $n_p \times n_p$ discrete cosine transform matrix C. Since C is orthonormal, we have that $V = C^\top$ satisfies condition (5). The change of variables $\widehat{p}' = V^\top \widehat{p}$ then transforms the problem into the frequency domain.

3 Solution method

Next, we present a local optimization method for solving problem (3). First, we represent the rank constraints in the kernel form

$$\operatorname{rank}\left(\mathscr{S}(\widehat{p})\right) \leq r \quad \Longleftrightarrow \quad \text{there is } R \in \mathbb{R}^{(m-r)\times m}, \text{ such that } R\mathscr{S}(\widehat{p}) = 0 \text{ and } R \text{ is full row rank.} \tag{6}$$

Then, we use the variable projection principle to eliminate $\widehat{p}$, which results in a nonlinear least-squares in R.

Representing the constraint of (3) in the kernel form (6), leads to the double minimization problem

$$\text{minimize} \quad \text{over } R \in \mathbb{R}^{(m-r)\times m} \quad f(R) \quad \text{subject to} \quad R \text{ is full row rank}, \tag{7}$$

where

$$f(R) := \min_{\widehat{p}} \|p - \widehat{p}\|_2 \quad \text{subject to} \quad R\mathscr{S}(\widehat{p}) = 0. \tag{8}$$

The computation of $f(R)$, called "inner" minimization, is over the estimate $\widehat{p}$ of p. The minimization over the kernel parameter $R \in \mathbb{R}^{(m-r)\times m}$ is called "outer". The inner minimization problem is a projection of the columns of $\mathscr{S}(p)$ onto the model $\mathscr{B} := \ker(R)$. Note that, the projection depends on the parameter R, which is the variable in the outer minimization problem. Thus, the name "variable projection".

The general linear structure

$$\mathscr{S} : \mathbb{R}^{n_p} \to \mathbb{R}^{m\times n}, \quad \mathscr{S}(\widehat{p}) = \sum_{k=1}^{n_p} S_k \widehat{p}_k \tag{9}$$

is specified by the n_p matrices $S_1, \ldots, S_{n_p} \in \mathbb{R}^{m\times n}$. Let

$$\mathbf{S} := \left[\operatorname{vec}(S_1) \cdots \operatorname{vec}(S_{n_p})\right] \in \mathbb{R}^{mn\times n_p},$$

so that

$$\operatorname{vec}\left(\mathscr{S}(\widehat{p})\right) = \mathbf{S}\widehat{p}, \quad \text{or} \quad \mathscr{S}(\widehat{p}) = \operatorname{vec}^{-1}(\mathbf{S}\widehat{p}). \tag{10}$$

Define the change of variables

$$\widehat{p} \;\mapsto\; \Delta p = p - \widehat{p}.$$

Then, the constraint of the optimization problem becomes

$$\begin{aligned}
R\mathscr{S}(\widehat{p}) = 0 \quad &\iff \quad R\mathscr{S}(p - \Delta p) = 0 \\
&\iff \quad R\mathscr{S}(p) - R\mathscr{S}(\Delta p) = 0 \\
&\iff \quad \operatorname{vec}\left(R\mathscr{S}(\Delta p)\right) = \operatorname{vec}\left(R\mathscr{S}(p)\right) \\
&\iff \quad \underbrace{\left[\operatorname{vec}(RS_1) \cdots \operatorname{vec}(RS_{n_p})\right]}_{G(R)} \Delta p = \underbrace{G(R)p}_{h(R)} \\
&\iff \quad G(R)\Delta p = h(R).
\end{aligned}$$

Assuming that

$$n_p \le (m-r)n \tag{11}$$

the inner minimization problem (8) with respect to the new variable Δp is a generalized linear least norm problem

$$f(R) = \min_{\Delta p} \quad \|\Delta p_{1:n_g}\|_2 \quad \text{subject to} \quad G(R)\Delta p = h(R). \tag{12}$$

(12) is not a standard least norm problem due to the presence of missing data (or equivalently singularity of the cost function), however, it has an analytic solution [12, Theorem 2.1].

For the outer minimization problem in (7), *i.e.*, the minimization of M over R, subject to the constraint that R is full row rank, we use general purpose constrained local optimization methods [15], representing the full row rank constraint as $RR^\top = I_{m-r}$. This is a nonconvex optimization problem, so that there is no guarantee that a globally optimal solution is found.

4 Hankel structured sparse approximation problems and ℓ_1-norm regularization

In this section, we consider the special case of problem (1) when the structure $\mathscr{S}$ is Hankel

$$\mathscr{H}_m(p) := \begin{bmatrix} p_1 & p_2 & p_3 & \cdots & p_{n_p-m+1} \\ p_2 & p_3 & \iddots & & p_{n_p-m+2} \\ p_3 & \iddots & & & p_{n_p-m+3} \\ \vdots & & & & \vdots \\ p_m & p_{m+1} & \cdots & \cdots & p_{n_p} \end{bmatrix}. \tag{13}$$

By the result of Theorem 1, problem (1) is a Hankel structured low-rank approximation with missing data. In turn, Hankel structured low-rank approximation is a linear time-invariant system identification problem with missing data. Therefore, equivalently, we consider a problem of system identification with missing data.

An alternative approach for missing data estimation with sparsity prior is ℓ_1-norm regularization. Sparsity of a signal in the frequency domain means that the signal is a sum of a few exponentials. In the paper, we consider real-valued time-domain signals, so that the frequency domain signal has an additional symmetry property.

A signal that is a sum of `n`-complex exponentials can be represented as an output of an autonomous linear time-invariant system of order `n`. Alternatively, such a signal can be represented as the impulse response of a `n`-th order linear time-invariant system. Representing exactly or approximately a given signal as an output of an autonomous linear time-invariant system or as the impulse response of an input/output linear time-invariant system are fundamental problems in system theory and system identification.

Next, we explain the similarities and differences between sparse approximation by ℓ_1-norm minimization in the frequency domain and sparse approximation by Hankel structured low-rank approximation. The underlying assumption for the ℓ_1-norm minimization problem is that the data b is generated as

$$b = Dx + \widetilde{b}, \tag{14}$$

where D is a $n_{\mathrm{g}} \times n_p$ matrix with $n_{\mathrm{g}} < n_p$, x is k-sparse with $k \ll n_p$, and $\widetilde{b}$ is a zero mean Gaussian random vector with covariance matrix $\sigma^2 I$. Moreover, it is assumed that D consists of the first n_{g} rows of the inverse discrete cosine transform matrix $C^\top$. Due to the properties of D (submatrix of the inverse discrete cosine transform) and x (k-sparse vector), $\bar{b} := Dx$ is a sum of k cosines with frequencies on the grid

$$0\frac{2\pi}{n_p}, 1\frac{2\pi}{n_p}, 2\frac{2\pi}{n_p}, \ldots (n_p-1)\frac{2\pi}{n_p}. \tag{15}$$

Assuming that enough observations are available, namely

$$n_{\mathrm{g}} \geq 2r+1, \qquad \text{where } r := 2k,$$

the Hankel matrix $\mathscr{H}_{r+1}(\bar{b})$ with $r+1$ rows and $n_{\mathrm{g}} - r$ columns, constructed from $\bar{b}$ has rank r. Vice versa,

$$\operatorname{rank}\left(\mathscr{H}_{r+1}(\bar{b})\right) \leq r \tag{16}$$

implies that $\bar{b}$ is a sum of at most $2n$ polynomials-times-damped-exponentials.

Note that (16) does not impose a constraint that the frequencies are on (15); they can be any real numbers in the interval $[0, 2\pi)$. Also (16) allows damped cosines while the model $\bar{b} = Dx$ does not allow damping. Therefore, (16) is not equivalent to $\bar{b} = Dx$ with x k-sparse. For large values of n_p, (15) approximates "well" the interval $[0, 2\pi)$.

5 Numerical examples

In this section we consider the Hankel structured low-rank approximation problem

$$\begin{array}{ll} \text{minimize} & \text{over } \widehat{b} \quad \|b - A\widehat{p}\|_2 \\ \text{subject to} & \operatorname{rank}\left(\mathscr{H}_{r+1}(\widehat{p})\right) \leq r \end{array} \tag{17}$$

First, we specialize the variable projections method described in Section 3 to the Hankel structured case and demonstrate on a simulation example that the resulting algorithm allows us to separate signal from noise. Then, we compare numerically the variable projections method with the ℓ_1-norm regularization method in setup of (14).

Autonomous system identification from data with missing values

In the case of Hankel structure (13), the matrices S_k in (9) are

$$S_1 = \begin{bmatrix} 1\,0\,0\,\cdots\,0 \\ 0\,0\,0\,\cdots\,0 \\ 0\,0\,0\,\cdots\,0 \\ \vdots\,\vdots\,\vdots\quad\;\vdots \\ 0\,0\,0\,\cdots\,0 \end{bmatrix}, \begin{bmatrix} 0\,1\,0\,\cdots\,0 \\ 1\,0\,0\,\cdots\,0 \\ 0\,0\,0\,\cdots\,0 \\ \vdots\,\vdots\,\vdots\quad\;\vdots \\ 0\,0\,0\,\cdots\,0 \end{bmatrix}, \begin{bmatrix} 0\,0\,1\,\cdots\,0 \\ 0\,1\,0\,\cdots\,0 \\ 1\,0\,0\,\cdots\,0 \\ \vdots\,\vdots\,\vdots\quad\;\vdots \\ 0\,0\,0\,\cdots\,0 \end{bmatrix}, \ldots, \begin{bmatrix} 0\,0\,0\,\cdots\,0 \\ 0\,0\,0\,\cdots\,0 \\ 0\,0\,0\,\cdots\,0 \\ \vdots\,\vdots\,\vdots\quad\;\vdots \\ 0\,0\,0\,\cdots\,1 \end{bmatrix}$$

and the $G(R)$ matrix in (12) is

$$G(R) = \begin{bmatrix} R_0 & R_1 & \cdots & R_r & & \\ & R_0 & R_1 & \cdots & R_r & \\ & & \ddots & \ddots & & \ddots \\ & & & R_0 & R_1 & \cdots & R_r \end{bmatrix},$$

where all missing elements are zeros. Fast ($O(n_p)$) implementation of the variable projection algorithm, taking into account the structure of $G(R)$ for the cost function and Jacobian evaluation is presented in [16].

Example 2. A random second order ($r = 2$) autonomous linear time-invariant system is generated in Matlab by the function `drss` and a random trajectory $\bar{p}$ of the system with $n_p = 50$ samples is then generated. The sampling matrix A is $\begin{bmatrix} I_{n_\mathrm{g}} & 0 \end{bmatrix}$, where $n_\mathrm{g} = 20$, *i.e.*, only the first 40% of the samples of $\bar{p}$ are observed. Finally, zero mean, white, Gaussian noise with standard deviation s is added to the true samples.

Figure 1 shows the relative estimation error

$$e := \|\bar{b} - \widehat{p}_{1:n_g}\|_2 / \|\bar{b}\|_2$$

from a Monte Carlo experiment with standard deviations varying in the interval $[0, 0.1]$ (signal-to-noise ratio varying from 46dB to infinity). The result shows that the low-rank prior allows us to filters noise from the data. Indeed, the error e in using the noisy data (solid black line) is higher and increases faster than the error e in using the estimate $\widehat{p}$ (the solution of problem (3) obtained with the variable projections algorithm).

Moreover, it can be shown that in the simulation setup of the example, the solution of problem (1) gives the maximum likelihood estimator, so that it is statistically optimal.

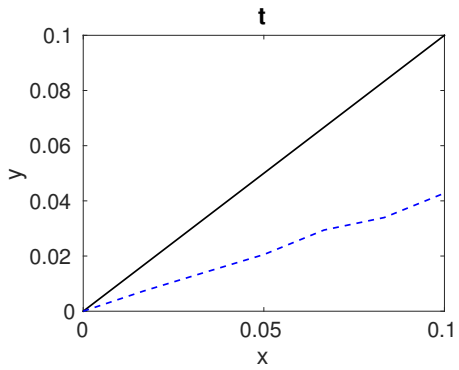

Fig. 1. The maximum likelihood (ML) estimator obtained by solving problem (1) with the variable projections algorithm (blue dashed line) improves the relative estimation error in comparison with the use of the raw noisy data (black solid line).

Comparison with the ℓ_1-norm regularization method

In this section, we consider data generated from the compressive sensing model (14) with $n_p = 100$, $n_g = 20$, $k = 2$, and noise standard deviation $s = 0.1$. The true data $\bar{p}$ is a sum of two sines with frequencies on the grid (15). With this simulation setup, the ℓ_1-norm regularization method recovers the correct frequencies with 100% success rate.

The low-rank constraint (2) with Hankel structured matrix and rank $n = 4$ imposes the weaker prior that the signal is a sum-of-damped exponentials, *i.e.*, damping is allowed and the frequencies are not assumed to be on the grid (15). Nevertheless, in the above simulation example the estimator defined by problem (17) also recovers the correct frequencies with 100% success rate.

Both the ℓ_1-norm regularization method and (17) fail when the number of given samples n_g is decreased and/or the noise standard deviation s is increased. The ℓ_1-norm regularization method fails for a smaller number of samples and at a higher noise standard deviation s. The reader can reproduce the reported results by downloading the SLRA package (`http://slra.github.io/`) and

`http://homepages.vub.ac.be/~imarkovs/software/ica18.tar`

Acknowledgements

The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC Grant 258581 "Structured low-rank approximation: Theory, algorithms, and applications"; Fund for Scientific Research (FWO-Vlaanderen), FWO projects G028015N "Decoupling multivariate polynomials in nonlinear system identification"; G090117N "Block-oriented nonlinear identification using Volterra series"; and FWO/FNRS Excellence of Science project 30468160 "Structured low-rank matrix / tensor approximation: numerical optimization-based algorithms and applications".

References

1. Beck, A., Eldar, Y.: Structured total maximum likelihood: An alternative to structured total least squares. SIAM Journal on Matrix Analysis and Applications 31(5), 2623–2649 (2010)
2. Bresler, Y., Macovski, A.: Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. IEEE Trans. Acust., Speech, Signal Process. 34, 1081–1089 (1986)
3. Cadzow, J.: Signal enhancement—A composite property mapping algorithm. IEEE Trans. Signal Proc. 36, 49–62 (1988)
4. Chu, M., Funderlic, R., Plemmons, R.: Structured low rank approximation. Linear Algebra Appl. 366, 157–172 (2003)
5. De Moor, B.: Structured total least squares and L_2 approximation problems. Linear Algebra Appl. 188–189, 163–207 (1993)
6. Golub, G., Pereyra, V.: Separable nonlinear least squares: the variable projection method and its applications. Institute of Physics, Inverse Problems 19, 1–26 (2003)
7. Huang, J.J., Dragotti, P.L.: Sparse signal recovery using structured total maximum likelihood. In: Int. Conf. on Sampling Theory and Applications. pp. 639–643 (2017)
8. Ishteva, M., Usevich, K., Markovsky, I.: Factorization approach to structured low-rank approximation with applications. SIAM J. Matrix Anal. Appl. 35(3), 1180–1204 (2014)
9. Markovsky, I.: Structured low-rank approximation and its applications. Automatica 44(4), 891–909 (2008)
10. Markovsky, I.: Recent progress on variable projection methods for structured low-rank approximation. Signal Processing 96PB, 406–419 (2014)
11. Markovsky, I.: Low-Rank Approximation: Algorithms, Implementation, Applications. Springer, second edn. (2018)
12. Markovsky, I., Usevich, K.: Structured low-rank approximation with missing data. SIAM J. Matrix Anal. Appl. 34(2), 814–830 (2013)
13. Polderman, J., Willems, J.C.: Introduction to Mathematical Systems Theory. Springer-Verlag, New York (1998)
14. Rosen, J., Park, H., Glick, J.: Total least norm formulation and solution of structured problems. SIAM J. Matrix Anal. Appl. 17, 110–126 (1996)
15. Usevich, K., Markovsky, I.: Optimization on a Grassmann manifold with application to system identification. Automatica 50, 1656–1662 (2014)
16. Usevich, K., Markovsky, I.: Variable projection for affinely structured low-rank approximation in weighted 2-norms. J. Comput. Appl. Math. 272, 430–448 (2014)
17. Yeredor, A., De Moor, B.: On homogeneous least-squares problems and the inconsistency introduced by mis-constraining. Comput. Stat. Data Anal. 47(3), 455–465 (2004)