LINE SPECTRAL ESTIMATION WITH PALINDROMIC KERNELS

Dieter Verbeke and Ivan Markovsky

Vrije Universiteit Brussel

ABSTRACT

Estimation of line spectra is a classical problem in signal processing and arises in many applications. The problem is to estimate the frequencies and corresponding amplitudes of a sum of sinusoidal components from noisy measurements. It can be solved with maximum likelihood methods or with suboptimal subspace methods. The constraint that the model does not have damping is difficult to impose in subspace methods. We develop an equivalent formulation as a structured low-rank approximation problem and present a necessary condition for the model to be undamped. The condition is that a vector in the kernel of a Hankel matrix of observations has palindromic structure and it leads to a linear equality constraint which is easily incorporated into a numerical algorithm. Simulations show that even for relatively high noise-to-signal ratios, the necessary condition is in practice also sufficient, i.e., the identified model does not have damping.

Index Terms— Line spectral estimation, structured low-rank approximation, palindromic kernels, subspace methods

1. INTRODUCTION

Spectral analysis concerns the problem of determining the spectral content of a time series from finite data [1]. In several applications the encountered signals can be well described as a sum of damped or undamped complex-valued sinusoidal signals. The latter occurs in e.g. telecommunications, radar, sonar and seismology, and the signal is said to have a *line spectrum* [1, Chapter 4]. The former occurs e.g. in magnetic resonance spectroscopy [1, Sec. 5.6.3]. The main task is to estimate the model parameters, namely the angular frequencies, and the corresponding amplitudes (and damping factors), from noisy measurements. Signals can be real-valued, in which case both the sine and the corresponding cosine must be present, or complex-valued, for example as a sub-problem in the reconstruction of finite rate of innovation signals [2].

There are two main approaches, maximum likelihood identification [3, 4] and subspace identification [5, 6, 7].

Subspace approaches are based on the low-rank property of a Hankel structured matrix of noise-free observations, and a decomposition of this matrix with two full-rank matrices that have a particular shift structure. Both properties are lost when measurements are corrupted by noise. Classical methods perform *unstructured* low-rank approximation of the measurement matrix using a singular value decomposition, followed by a least squares fit of the signal components to the estimated subspace basis. More recently, *structured* low-rank approximation methods were developed [8, 9].

The rank deficiency of the Hankel matrix is not sufficient for the signal to be a sum of *undamped* exponentials. The constraint that poles of the model are simple and lie on the unit circle is difficult to impose when the model parameters are not the poles themselves. We show how the relaxed constraint that the kernel of the Hankel matrix has a palindromic structure can effectively restrict the location of the identified poles, or equivalently, ensure that the signal is undamped.

The rest of this paper is organized as follows. Section 2 states the problem and describes different model representations. Section 3 introduces the concept of palindromic kernels which are then exploited in the solution methods proposed in Section 4. Section 5 supports the claims with numerical evidence and Section 6 concludes the paper.

2. PROBLEM STATEMENT AND MODEL REPRESENTATIONS

Consider a finite-horizon discrete-time real-valued signal that is a sum of k real-valued sinusoids and therefore $n = 2k$ complex exponentials

$$y(t) = \underbrace{\sum_{\ell=1}^{k} c_\ell \sin(\omega_\ell t + \phi_\ell)}_{\bar{y}(t)} + e(t), \quad \text{for } t = 1, \dots, T, \quad (1)$$

with $c_\ell \in \mathbb{R}, \phi_\ell \in [-\pi, \pi]$, and where $\bar{y}(t)$ is the noise-free sum-of-sines signal and $e(t)$ additive observation noise which is assumed to be zero-mean, white Gaussian noise.

Denote by $\mathcal{M}_n$ the model class of sum-of-exponential models of order (the number of exponentials) at most n. Adhering to the behavioral approach in system theory [10], a model is defined as a set of trajectories $\mathcal{B} \in \mathcal{M}_n$. The fact that y is a trajectory of a model $\mathcal{B}$ is concisely written as $y \in \mathcal{B}$. Let σ denote the backward shift operator $(\sigma y)(t) := y(t+1)$.

A model $\mathscr{B} \in \mathscr{M}_n$ can be *represented* by by equations, leading to different *parameterizations*:

1. *sum-of-sines*

$$\mathscr{B}(\boldsymbol{\omega}) = \{y \mid y(t) \text{ satisfies (1) for all } c_\ell, \phi_\ell \in \mathbb{R}\}$$

   parameterized by vector of frequencies $\boldsymbol{\omega} \in \mathbb{R}^k$;

2. *state-space representation*

$$\mathscr{B}(A, c) = \{y \mid \text{there is } s, \text{ such that } y = cx, \boldsymbol{\sigma}x = Ax\},$$

   parameterized by matrices $A \in \mathbb{R}^{n \times n}$ and $c \in \mathbb{R}^{1 \times n}$;

3. *difference equation (kernel representation)*

$$\mathscr{B}(R) = \{y \mid R_0 + R_1 \boldsymbol{\sigma} y + \cdots + R_n \boldsymbol{\sigma}^n y = 0\},$$

   parameterized by polynomial

$$R(z) := R_0 + R_1 z + \cdots + R_n z^n.$$

The model $\mathscr{B}(R)$ is determined by the kernel $\ker R(\boldsymbol{\sigma})$ of the polynomial operator $R(\boldsymbol{\sigma})$, hence the term *kernel representation*. The vector of coefficients $R := \begin{bmatrix} R_0 & \cdots & R_n \end{bmatrix}$, is obviously not unique, since $\mathscr{B}(\alpha R) = \mathscr{B}(R)$ for any $\alpha \neq 0$. However, the roots of $z_1, \ldots, z_n$, are invariant w.r.t. the representation and are called the *poles* of the model $\mathscr{B}$. The trajectories y of the model $\mathscr{B}$ with a kernel representation $\ker R(\boldsymbol{\sigma})$ can be parameterized by n "past" samples

$$\big(y(-n+1), \ldots, y(-1), y(0)\big).$$

The parameters of the model $\mathscr{B} \in \mathscr{M}_n$ in the three representations are related. Up to a reordering, $z_i = e^{j\omega_i}$, for $i = 1, \ldots, k$, and $z_i = e^{-i\omega_k}$ for $i = k+1, \ldots, 2k$, while the eigenvalues of the matrix A are equal to the poles of the model.

3. PALINDROMIC KERNELS

Model $\mathscr{B}$ with kernel representation $\ker R(\boldsymbol{\sigma})$ is a sum-of-exponentials model if and only if the roots of the polynomial $R(z)$ lie on the unit circle. Unfortunately, expressing the constraint on the roots of $R(z)$ as a constraint on its coefficient vector R is analytically impossible for polynomials of degree more than four (see Abel's impossibility theorem).

A necessary condition for the poles of $\mathscr{B}$ to lie on the unit circle is that their locations must be *symmetric* w.r.t. the unit circle. This condition can be expressed in terms of the kernel coefficients. More precisely, $\ker R(\boldsymbol{\sigma})$ must be palindromic. A polynomial $R(z)$ is *palindromic* if $R_{n-i} = R_i$, for $i = 0, \ldots, k$, and *anti-palindromic* if $R_{n-i} = -R_i$, for $i = 0, \ldots, k$. The property that the kernel is palindromic has a system-theoretic interpretation, namely that the trajectories of the model it represents are *time-reversible*. Later, in Section 5, it will be demonstrated empirically that imposing the necessary condition improves the performance.

Theorem 1. *See [11, Thm. 2]. Consider the following conditions: i) every root $\lambda \in \mathbb{C}$ of R(z) is either on the unit circle or R(z) has a root $1/\lambda$ with the same multiplicity as λ; ii) 1 is a root of R(z) with even multiplicity (multiplicity 0 means that 1 is not a root of R(z)); iii) 1 is a root of R(z) with odd multiplicity. The polynomial R(z) is palindromic if and only if i) and ii) hold, and is anti-palindromic if and only if i) and iii) hold.*

Corollary 2. *Let $\mathscr{B} \in \mathscr{M}_n$ be a set of trajectories belonging to the class of sum-of-exponentials models and $\ker R(\boldsymbol{\sigma})$ a kernel representation of $\mathscr{B}$. Assuming that 1 is not a pole of $\mathscr{B}$, R(z) is palindromic.*

Since the signals are real-valued, poles must come in pairs that are symmetric w.r.t. the real axis, *i.e.* if λ is a pole, its complex conjugate $\bar{\lambda}$ is necessarily also a pole. This fact and Theorem 1 imply that the complex roots of palindromic and anti-palindromic polynomials that do not fall on the unit circle can be divided into quadruples $(\lambda, 1/\lambda, \bar{\lambda}, 1/\bar{\lambda})$. Roots that lie on the unit circle or the real line, except possibly roots at ± 1, can be grouped into tuples $(\lambda, 1/\lambda)$.

There is an interesting link between palindromic/anti-palindromic kernels and the concept of time-reversibility [11]. Let rev denote the time-reversal operator

$$\text{rev}\big(y(1), y(2), \ldots, y(t)\big) := \big(y(T)y(T-1), \ldots, y(1)\big).$$

Then, time-reversibility of the trajectories means that for all $y \in \mathscr{B}$ also $\text{rev}(y) \in \mathscr{B}$, and a time-reversible system is defined as a system for which this property holds.

Proposition 1. If $\mathscr{B}(R)$ is defined by a palindromic or anti-palindromic polynomial $R(z)$, then for all $y \in \mathscr{B}$, it is true that $\text{rev}(y) \in \mathscr{B}$, *i.e.*, the trajectories are time-reversible.

If the sum-of-exponentials model does not contain an offset, then $\omega_i \neq 0$ modulo 2π for $i = 1, \ldots, k$, or equivalently, $\lambda_i \neq 1$ for $i = 1, \ldots, n$. If the model contains an offset, the polynomial $R(z)$ will be anti-palindromic. In the remainder of the paper, it is assumed that the model does not have an offset and only palindromic kernels are considered.

4. SOLUTION METHODS

4.1. Structured low-rank approximation

$$\begin{aligned} \text{minimize} \quad & \text{over } \widehat{y} \text{ and } R \neq 0 \quad \|y_d - \widehat{y}\| \\ \text{subject to} \quad & \widehat{y} \in \mathscr{B} = \ker R(\boldsymbol{\sigma}) \text{ and } R_i = R_{n-i}. \end{aligned} \quad (2)$$

Note that the palindromic vector $R \in \mathbb{R}^{1 \times (n+1)}$ can be expressed as

$$R = \underbrace{\begin{bmatrix} R_0 & R_1 & \ldots & R_k \end{bmatrix}}_{R'} \underbrace{\begin{bmatrix} 1 & & & 1 \\ & \ddots & & \ldots \\ & & 1 & \end{bmatrix}}_{\Psi} \quad (3)$$

and the optimization is actually over $R' \in \mathbb{R}^{1 \times (k+1)}$, using only half the number of coefficients. The link between the sum-of-exponentials model (1) and low-rank approximation is given by

$$\widehat{y} \in \mathscr{B} = \ker R(\sigma) \iff R\mathscr{H}_{n+1}(\widehat{y}) = 0 \quad \text{and} \quad R \neq 0. \quad (4)$$

Combining (2)–(4), an equivalent Hankel-structured low-rank approximation problem is obtained

$$
\begin{aligned}
&\text{minimize} \quad \text{over } \widehat{y} \text{ and } R' \neq 0 \quad ||y_{\mathrm{d}} - \widehat{y}|| \\
&\text{subject to} \quad R'\Psi\mathscr{H}_{n+1}(\widehat{y}) = 0,
\end{aligned}
\quad (5)
$$

which can be solved with a variable projection method [12]. The proposed formulations as structured low-rank approximation can be regarded as relaxed versions of a maximum likelihood problem [13].

4.2. Subspace methods

Optimization problems (2) & (5) do not admit an analytical solution, nor is it convex. The discussed method only guarantees the convergence to a local optimum. Two alternative sub-optimal heuristic approaches will be considered. These methods split the identification problem into two steps: first, an unstructured low-rank approximation, and second, least squares estimation of the model parameters from the obtained matrix factorization.

4.2.1. Kung's method

Kung's method is based on realization theory. For exact data, a rank-revealing factorization

$$\mathscr{H}_L(y) = \mathscr{O}\mathscr{C} \text{ with } \mathscr{O} \in R^{L \times n}, \mathscr{C} \in R^{n \times (T-L)}, \quad (6)$$

exists where L is a user-defined natural number between $n+1$ and $T-n$. A system matrix $\widehat{A}$ is recovered as the solution of a system of linear equations $\overline{\mathscr{O}}\widehat{A} = \underline{\mathscr{O}}$, where $\overline{\mathscr{O}}$ is the matrix $\mathscr{O}$ with the first row removed, and $\underline{\mathscr{O}}$ is the matrix $\mathscr{O}$ with the last row removed. The eigenvalues of $\widehat{A}$ are then the poles of the system.

For inexact data, both steps involve an approximation. The rank-revealing factorization is computed as a rank-n truncation of the singular value decomposition, *i.e.*, an unstructured low-rank approximation. Solving the system of linear equations delivers an approximate solution in the ordinary least-squares sense.

Empirical observations show that subspace methods give more accurate estimates of the model parameters when the number of rows L of the Hankel matrix $\mathscr{H}(y_d)$ is chosen such that $\mathscr{H}(y_d)$ is as square as possible. Therefore L should be the nearest integer to $T/2$.

4.2.2. Modifications to Kung's method for time-reversible systems

Ignoring the Hankel structure, an approximate one-dimensional left kernel of the Hankel-plus-Toeplitz matrix $H := \Psi\mathscr{H}_{n+1}(y_d)$ can be found as the left singular vector of H corresponding to the smallest eigenvalue. A basis vector R' then leads to a model with kernel parameter $R'\Psi$. Equivalently, the kernel parameter can be obtained as the left singular vector of the mosaic-Hankel matrix

$$H := \begin{bmatrix} \mathscr{H}_{n+1}(y_d) & \mathscr{H}_{n+1}(\mathrm{rev}(y_d)) \end{bmatrix}. \quad (7)$$

corresponding to the smallest singular value.

Two versions of the procedure are thus obtained. Due to the time-reversibility of the identified model $\ker R(\sigma)$, the resulting methods can be seen as forward-backward linear prediction. Both procedures have $L = n+1$, such that with $T \gg n$, H in (7) is a highly skewed rectangular matrix. To mitigate this issue another modification is made. Firstly, a rank-revealing factorization of the mosaic Hankel matrix is computed $H := \begin{bmatrix} \mathscr{H}_L(y_d) & \mathscr{H}_L(\mathrm{rev}(y_d)) \end{bmatrix}$, *i.e.*, $H = \mathscr{O}\mathscr{C}$, where $L \in \mathbb{N}$ and $n+1 < L < T-n$. Secondly, a kernel parameter $R \in \mathbb{R}^{1 \times (n+1)}$ must be found for the forward-backward linear prediction method applied to the set of time series $y^{(i)} := (\mathscr{O}_{1i}, \dots, \mathscr{O}_{Li})$, for $i = 1, \dots, n$. That is, compute a one-dimensional approximate left kernel of the mosaic Hankel matrix

$$
\begin{aligned}
H' = \big[\ &\mathscr{H}_{n+1}\big(y^{(1)}\big) \quad \mathscr{H}_{n+1}\big(\mathrm{rev}(y^{(1)})\big) \quad \dots \\
&\dots \quad \mathscr{H}_{n+1}\big(y^{(n)}\big) \quad \mathscr{H}_{n+1}\big(\mathrm{rev}(y^{(n)})\big) \ \big].
\end{aligned}
$$

The construction in the second step leverages Lemma 3 and the fact that for any other couple of trajectories $y', y'' \in \mathscr{B}$ that enter the mosaic Hankel matrix $\begin{bmatrix} \mathscr{H}_{n+1}(y') & \mathscr{H}_{n+1}(y'') \end{bmatrix}$ y' and y'' satisfy the same difference equation.

Lemma 3. *If y_d is a trajectory of T samples, $y_d \in \mathscr{B}$ that is assumed to be persistently exciting of order n, then any L samples long trajectory y of $\mathscr{B}$ can be written as a linear combination of the columns of $\mathscr{H}_L(y_d)$ and any linear combination of these columns is in turn a trajectory of $\mathscr{B}$.*

5. SIMULATION RESULTS

Three methods are compared in a numerical example: the structured-low rank approach of Section 4.1, Kung's method of Section 4.2.1, and the modified version of Kung's method introduced in Section 4.2.2.

The setup is one of output error identification with a true system $\overline{\mathscr{B}} \in \mathscr{M}_n$, and additive zero-mean white Gaussian measurement noise $e(t) \sim \mathscr{N}(0, \sigma^2)$. Let $\bar{\mathscr{B}}$ be the true model and $\bar{z}_1, \bar{z}_2, \dots, \bar{z}_n$ its poles. The accuracy of an estimation method is evaluated by the average relative error

$$\frac{1}{N_{mc}} \sum_k^{N_{mc}} \frac{||\bar{\theta} - \widehat{\theta}^{(k)}||}{||\bar{\theta}||},$$
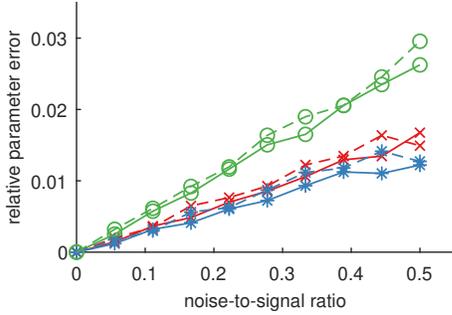
Fig. 1. Kung's method (green ○), modified version of Kung's method (red ×) and structured low-rank approximation method (blue ∗). Solid curves correspond to a sinusoidal component with a frequency on the DFT grid, dashed curves correspond to an unrestricted frequency.

Fig. 2. Up to a noise-to-signal ratio of 1 the results of the (brute-force) maximum likelihood method and structured low-rank approximation method are indistuingishable.

where $\bar{\theta}$ is the true parameter vector $\bar{\theta} \triangleq \begin{bmatrix} 1 & \bar{\theta}_1 & \bar{\theta}_2 & \ldots & \bar{\theta}_n \end{bmatrix}$ with $1 + \bar{\theta}_1 z + \bar{\theta}_2 z^2 + \ldots + \bar{\theta}_n z^n = \prod_{i=1}^{n}(z - \bar{z}_i)$, and $\widehat{\theta}^{(k)}$ is the parameter vector of the model $\mathscr{B}^{(k)}$ obtained in the k-th out of N_{mc} Monte Carlo trials.

The reported results are obtained for a signal with only $k = 1$ random sinusoidal component and $T = 25$ samples of a random trajectory of that model. The noise-to-signal ratio is varied from 0 to 0.5, and estimation results (for a given signal and fixed settings) are averaged over $N_{mc} = 100$ Monte Carlo repetitions. Figure 1 shows the estimation errors of the methods for two different experiments: one where the true model's frequency is restricted to the grid of the discrete Fourier transform, *i.e.*, multiples of $2\pi/N$, and one where the true model's frequency is unconstrained, *i.e.*, real numbers between 0 and 2π. The modified version of Kung's method outperforms the Kung's classical method and the structured low-rank approximation method outperforms the modified version of Kung's method.

Figure 2 depicts the estimation errors of the the structured low-rank approximation method with those of a maximum likelihood estimator that searches for the frequencies ω_ℓ by brute force optimization. It is verified that the solutions coincide (up to the convergence tolerance of the optimization methods) up to noise-to-signal ratio of 1. The results confirm that including prior information, here to be understood as imposing the necessary condition on the location of poles, improves the accuracy of the estimated parameters.

6. CONCLUSION

We proposed a relaxed structured low-rank approximation method and a modification of Kung's method. They exploit a necessary condition for the model to be a sum-of-exponentials model: its kernel representation should have palindromic structure. The palindromic kernel structure
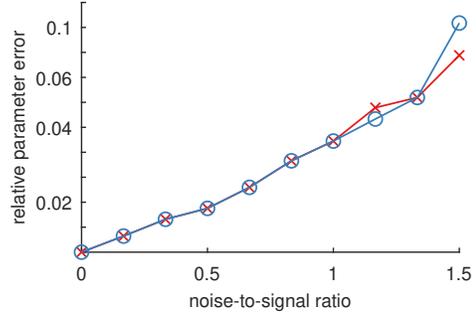
corresponds to time-reversibility of the model. Simulation results show that even for relatively high noise-to-signal ratios, the necessary condition is in practice also sufficient, *i.e.*,, the identified models are in fact sum-of-exponentials models. Imposing constraints on the kernel parameter allows to enforce other properties on the model, e.g. offset, trend, and seasonal components. Combining additional constraints with the time-reversibility property, obtained by a palindromic kernel structure, is an interesting topic for future research, as is implementation for complex-valued sum-of-exponentials.

7. REFERENCES

[1] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, 2005.

[2] Martin Vetterli, Pina Marziliano, and Thierry Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Proc.*, vol. 50, no. 6, pp. 1417, 2002.

[3] D. Rife and R. Boorstyn, "Multiple tone parameter estimation from discrete-time observations," *Bell System Technical Journal*, vol. 55, no. 9, pp. 1389–1410, 1976.

[4] D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids," *Proc. of the IEEE*, vol. 70, no. 9, pp. 975–989, 1982.

[5] B. Ho and Kalman R., "Effective construction of linear state-variable models from input/output functions," *Regelungstechnik*, vol. 14, no. 1-12, pp. 545–548, 1966.

[6] S.-Y. Kung, "A new identification and model reduction algorithm via singular value decomposition," in *Proc. 12th Asilomar Conf. on Circuits, Systems and Computers*, 1978.

[7] R. Roy, A. Paulraj, and T. Kailath, "Esprit–a subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoustics, Speech, and Signal Proc.*, vol. 34, no. 5, pp. 1340–1342, 1986.

[8] J. Klintberg and T. McKelvey, "An improved method for parametric spectral estimation," in *ICASSP 2019*, May 2019, pp. 5551–5555.

[9] I. Markovsky, "Structured low-rank approximation and its applications," *Automatica*, vol. 44, no. 4, pp. 891–909, 2008.

[10] I. Markovsky, J. C. Willems, S. Van Huffel, and B. De Moor, *Exact and Approximate Modeling of Linear Systems: A Behavioral Approach*, SIAM, 2006.

[11] I. Markovsky and S. Rao, "Palindromic polynomials, time-reversible systems, and conserved quantities," in *16th Mediterranean Conf. on Control and Automation*, 2008, pp. 125–130.

[12] I. Markovsky and K. Usevich, "Software for weighted structured low-rank approximation," *J. Comp. and Appl. Math.*, vol. 256, pp. 278–292, 2014.

[13] I. Markovsky and D. Toon Verbeke, "Sum-of-exponentials modeling via hankel low-rank approximation with palindromic kernel structure," Tech. Rep., Dept. ELEC, Vrije Universiteit Brussel, 2018.