Sum-of-exponentials modeling via Hankel low-rank approximation with palindromic kernel structure

Ivan Markovsky* and Dieter Toon Verbeke
Vrije Universiteit Brussel
Pleinlaan 2, 1050 Brussels, Belgium
Email: Ivan.Markovsky@vub.be

Abstract

Estimating a sum-of-damped-exponentials signal from noisy samples of the signal is a classic signal processing problem. It can be solved by maximum likelihood as well as suboptimal subspace methods. In this paper, we consider the related problem of sum-of-exponentials modeling, in which the model is constrained to be undamped. This constraint is difficult to impose in subspace methods. We develop solution methods using an equivalent formulation of the problem as Hankel matrix low-rank approximation. A necessary condition for the model to have no damping is that a vector in the kernel of the Hankel matrix has palindromic structure. This condition leads to a linear equality constraint, which is easy to impose in numerical algorithms. Simulation results show that even for a relatively high noise-to-signal ratios the necessary condition is in fact also sufficient, *i.e.*, the identified model has no dumping. Another contribution of the paper is a method for sum-of-exponentials modeling based on circulant embedding: low-rank approximation of a circulant matrix constructed from the given data. This method imposes the constraint that the model has no damping plus an addition constraint that the model frequencies are on the discrete-Fourier transform's grid.

Keywords: System identification, Sum-of-exponentials modeling, Low-rank approximation, Behavioral approach, Subspace methods, Circulant embedding.

*Corresponding author, member of EURASIP

1 Introduction

Consider a finite-horizon discrete-time real-valued signal that is a sum of $n = 2k$ complex exponentials and a zero mean white Gaussian signal

$$y_{\mathrm{d}}(t) = \underbrace{c_1 \sin(\omega_1 t + \phi_1) + \cdots + c_k \sin(\omega_k t + \phi_k)}_{\bar{y}(t)} + e(t), \quad \text{for } t = 1, \ldots, T. \tag{1}$$

We call $\bar{y}$ the "true signal" and e the "noise". The problem considered is to estimate the true signal's frequencies $\omega_1, \ldots, \omega_k$ from the "noisy signal" y_{d}. Equivalently, we aim to find a sum-of-exponentials model for y_{d}.

Sum-of-exponentials modeling is a basic problem in signal processing (**sp-book-stoica**; **KM81**; *Modern spectral estimation: Theory and applications*). Two main solution approaches are subspace identification (**VODM96**; "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood") and maximum likelihood identification (**sp-bresler**; **sp-abatzoglou**; "Superresolution by structured matrix approximation"). In (**sp-tufts-structured**; **slra**; *Low Rank Approximation: Algorithms, Implementation, Applications*), it is shown that existing identification methods are based on the fact that a Hankel matrix

$$\mathscr{H}_L(\bar{y}) := \begin{bmatrix} \bar{y}(1) & \bar{y}(2) & \cdots & \bar{y}(T-L+1) \\ \bar{y}(2) & \bar{y}(3) & \cdots & \bar{y}(T-L+2) \\ \bar{y}(3) & \bar{y}(4) & \cdots & \bar{y}(T-L+3) \\ \vdots & \vdots & & \vdots \\ \bar{y}(L) & \bar{y}(L+1) & \cdots & \bar{y}(T) \end{bmatrix},$$

constructed from the true signal $\bar{y}$ has rank at most n. Subspace methods do rank-n approximation of the matrix $\mathscr{H}_L(y_{\mathrm{d}})$ constructed from the noisy data y_{d}, using the singular value decomposition, *i.e.*, they perform *unstructured low-rank approximation*. Maximum likelihood methods do *Hankel structured low-rank approximation* of the matrix $\mathscr{H}_L(y_{\mathrm{d}})$. Structured low-rank approximation is a nonconvex optimization problem. In general, it does not admit an analytic solution. Current methods are based on local-optimization, starting from a suboptimal estimate, *e.g.*, one obtained by a subspace method.

The condition $\operatorname{rank}\big(\mathscr{H}_L(\bar{y})\big) = n$, however, is not equivalent to the condition that $\bar{y}$ is a sum-of-exponentials signal. The former imposes that $\bar{y}$ is a sum of polynomials-times-damped-exponentials, while the latter imposes that the poles of the model are simple and lie on the unit circle. Existence of polynomial factors of degree more than zero is related to poles with multiplicity more than one. In modeling from noisy data, this is a nongeneric case, so that it is not of major concern in practical applications. In contrast, the existence of nonzero damping, however, is generic and implies that a model identified with only the constraint on the rank of the Hankel matrix is generically not a sum-of-exponentials model. In addition to the rank constraint, a constraint that the poles are on the unit circle is needed. This latter constraint, however, is difficult to impose in subspace identification methods as well as in optimization-based

methods when the model parameters are not the poles themselves.

In this paper, we show how the relaxed constraint that the kernel of the Hankel matrix has a palindromic structure can effectively constrain the location of the identified poles to the unit circle. The problem of Hankel low-rank approximation with a palindromic kernel is equivalent to Hankel low-rank approximation with an equality constraint on the kernel parameter, as well as Hankel-plus-Toeplitz low-rank approximation, and mosaic-Hankel low-rank approximation. All these formulations, in turn, can be solved by the method of ("Software for weighted structured low-rank approximation"). Imposing the palindromic structure in the subspace approach is also possible and leads to the forward-backward linear prediction method ("Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood") and a modification of Kung's method for for sum-of-exponentials modeling (originally presented in ("A new identification method and model reduction algorithm via singular value decomposition") as a method for system realization).

Apart from imposing polynomial structure on the kernel, another approach for sum-of-exponentials modeling explored in this paper is circulant embedding, *i.e.*, we embed the Hankel matrix in a circulant matrix and do low-rank approximation of the circulant matrix. The motivation for the circulant embedding approach is that circulant structured low-rank approximation admits an analytic solution ("Real-valued, low rank, circulant approximation"). When $\omega_i = \pi n_i/T$, for some $n_i \in \{0, 1, \ldots, T\}$, *i.e.*, the signal frequencies are on the discrete Fourier transform (DFT) grid, the circulant embedding method leads simultaneously to a low-rank *and Hankel-structured* approximation of $\mathscr{H}_L(y_\mathrm{d})$. For noise variance up to a certain (data dependent) threshold, the circulant embedding method recovers the true signal's frequencies exactly.

The paper is organized as follows, in Section 2, we define the sum-of-exponentials model as a set of trajectories, introduce the kernel representation, and define the identification problem. The identification problem is then solved via three different approaches: maximum likelihood, subspace, and circulant embedding. The maximum likelihood and subspace methods, presented in Section 3, use the constraint that the parameter in a kernel representation of the model is palindromic. The circulant embedding method, presented in Section 4, leverages the prior knowledge that the true signal's frequencies are on the DFT grid. Simulation examples, illustrating the performance of the methods, are shown in Section 5.

2 Problem formulation

Denote by $\mathscr{M}_n$ the model class of sum-of-exponential models of order (number of exponentials) at most n. Following the behavioral approach in system theory (*Introduction to Mathematical Systems Theory*; "The behavioral approach to open and interconnected systems: Modeling by tearing, zooming, and linking"; *Exact and Approximate Modeling of Linear Systems: A Behavioral Approach*), we define a model $\mathscr{B} \in \mathscr{M}_n$ as a set of trajectories. The fact that y is a trajectory of a model $\mathscr{B}$ is concisely written as $y \in \mathscr{B}$.

A model $\mathscr{B} \in \mathscr{M}_n$ can be *represented* by equations. This leads to different *parameterizations* of the model. In this paper, we use three representations and their corresponding model parameters.

1. *Sum-of-sines representation*

$$\mathscr{B}(\omega) = \{ y \mid y(t) = c_1 \sin(\omega_1 t + \phi_1) + \cdots + c_k \sin(\omega_k t + \phi_k), \text{ for all } t, \text{ where } c, \phi \in \mathbb{R}^k \}.$$

The vector of the frequencies $\omega = (\omega_1, \ldots, \omega_k) \in \mathbb{R}^k$ parametrizes the model. The amplitudes $c_1, \ldots, c_k$ and phases $\phi_1, \ldots, \phi_k$ parameterize the trajectories y of the model.

2. *Difference equation representation*

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

where σ be the *backwards shift operator*

$$(\sigma y)(t) := y(t+1), \quad \text{for all } t.$$

The difference equation representation is defined by a polynomial

$$R(z) := R_0 z^0 + R_1 z^1 + \cdots + R_n z^n.$$

It is also refered to as a *kernel representation* because the model $\mathscr{B}(R)$ is defined by the kernel $\ker R(\sigma)$ of the polynomial operator $R(\sigma)$. The vector of coefficients

$$R := \begin{bmatrix} R_0 & R_1 & \cdots & R_n \end{bmatrix}$$

is not unique. Indeed, $\mathscr{B}(R) = \mathscr{B}(\alpha R)$, for any $\alpha \neq 0$. However, the roots $z_1, \ldots, z_n$ of $R(z)$ are invariant of the representation and are called the *poles* of $\mathscr{B}$. The trajectories y of a model with a kernel representation can be parameterized by the n "past samples" $\big(y(-n+1), \ldots, y(-1), y(0)\big)$.

3. *State-space representation*

$$\mathscr{B}(A, C) = \{ y \mid \text{ there is } x, \text{ such that } y = Cx, \sigma x = Ax \}.$$

The model parameters in a state space representation are the matrix $A \in \mathbb{R}^{n\times n}$ and the vector $c \in \mathbb{R}^{1\times n}$. They are not unique due to a change of basis transformation:

$$\mathscr{B}(A,C) = \mathscr{B}(VAV^{-1}, cV^{-1}), \qquad \text{for all nonsingular } V \in \mathbb{R}^{n\times n}.$$

The trajectories y of a model with a state space representation $\mathscr{B}(A,C)$ are conveniently parameterized by the initial state vector $x(1) \in \mathbb{R}^n$.

The parameters of a model $\mathscr{B} \in \mathscr{M}_n$ in the frequencies, kernel, and state space representations are related. Let $\mathbf{i}$ be the imaginary unit. Up to a reordering of the poles, we have that $z_i = e^{\mathbf{i}\omega_i}$, for $i = 1,\ldots,k$ and $z_i = e^{-\mathbf{i}\omega_i}$, for $i = k+1,\ldots,2k$. Also, the eigenvalues of A are equal to the poles of the model.

Under the assumption that the given data is generated from a true model $\bar{\mathscr{B}} \in \mathscr{M}_n$ with additive zero-mean, white, Gaussian noise, *i.e.*

$$y_{\mathrm{d}} = \bar{y} + e, \quad \text{where } \bar{y} \in \mathscr{B} \in \mathscr{M}_n \text{ and } e \sim \mathrm{N}(0, s^2 I),$$

the maximum likelihood estimator $\widehat{\mathscr{B}}$ of $\bar{\mathscr{B}}$ is defined by the optimization problem

$$\text{minimize} \quad \text{over } \widehat{y} \text{ and } \widehat{\mathscr{B}} \quad \|y_{\mathrm{d}} - \widehat{y}\| \quad \text{subject to} \quad \widehat{y} \in \widehat{\mathscr{B}} \in \mathscr{M}_n. \tag{2}$$

Here as well as in the rest of the paper, $\|\cdot\|$ denotes the Euclidean norm.

3 Methods based on Hankel low-rank approximation with palindromic kernel

A model $\mathscr{B} = \ker R(\sigma)$ is a sum-of-exponentials model if and only if the roots of $R(z)$ are simple and are on the unit circle. The constraint on $R(z)$ that is equivalent to the constraint that its roots are on the unit circle, however, can not be expressed analytically for polynomials of degree more than four (Abel's impossibility theorem). In Section 3.1, first, we introduce a simple necessary condition and show its system theoretic meaning. The condition is that $R(z)$ is palindromic and its meaning is that the trajectories of the model $\ker R(\sigma)$ are *time-reversible* ("Palindromic polynomials, time-reversible systems, and conserved quantities"). In Section 3.2, we use the necessary condition in order to solve a relaxed version of the maximum likelihood problem (2). Finally, in Section 3.3, we use the condition for deriving subspace identification methods.

3.1 Palindromic kernel representations and time-reversible systems

A kernel representation of a sum-of-exponentials model has structured parameter $R(z)$. If the model has no offset, *i.e.*, $\omega_i \neq 0 \pmod{2\pi}$, for $i = 1,\ldots,k$ or, equivalently $z_i \neq 1$, for $i = 1,\ldots,n$, then $R(z)$ is palindromic.

Proposition 1 (("Palindromic polynomials, time-reversible systems, and conserved quantities")). *Let $\mathscr{B} \in \mathscr{M}_n$ and let* $\ker R(\sigma)$ *be a kernel representation of $\mathscr{B}$. Assuming that* 1 *is not a pole of $\mathscr{B}$, R is a palindromic polynomial,* i.e.*, $R_{n-i} = R_i$, for $i = 0,1,\ldots,k$.*

If the model has an offset, $R(z)$ is anti-palindromic, *i.e.*, $R_{n-i} = -R_i$, for $i = 0,1,\ldots,k$. In the rest of the paper, we assume that the model has no offset, so that as a necessary condition for sum-of-exponentials modeling, we have that the parameter $R(z)$ of a kernel representation is palindromic.

The condition that $R(z)$ is palindromic is not sufficient for the poles of the model to be on the unit circle. It is sufficient for a weaker condition: symmetry with respect to the unit circle, *i.e.*, if z is a pole of the system, then $1/z$ is also a pole of the system. In addition, since the signal is real-valued, the poles must be symmetric with respect to the real axis, *i.e.*, if z is a pole, its complex conjugate $\bar{z}$ is also a pole. Therefore, the poles of a model $\mathscr{B} = \ker R(\sigma)$ with $R(z)$ palindromic, occur in quadruples $(z, \bar{z}, 1/z, 1/\bar{z})$.

Define the time-reversal operator

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

The palindromic structure of $R(z)$ implies time-reversibility of the trajectories y of the model $\mathscr{B} = \ker R(\sigma)$.

Proposition 2 (("Palindromic polynomials, time-reversible systems, and conserved quantities")). *Let $\mathscr{B} = \ker R(\sigma)$ with $R(z)$ palindromic. Then, for all $y \in \mathscr{B}$, rev $y \in \mathscr{B}$.*

3.2 Optimization method

Proposition 1 leads to the following relaxation of the maximum likelihood identification problem (2):

$$\begin{array}{lll} \text{minimize} & \text{over } \widehat{y} \text{ and } R \in \mathbb{R}^{1\times(n+1)} & \|y_{\mathrm{d}} - \widehat{y}\| \\ \text{subject to} & \widehat{y} \in \mathscr{B} = \ker R(\sigma), \quad R \neq 0, \quad \text{and} & R_i = R_{n-i}, \text{ for } i = 1,\ldots,k. \end{array} \tag{3}$$

Since a palindromic vector $R \in \mathbb{R}^{1\times(n+1)}$ can be expressed in terms of half of its coefficients:

$$R = \underbrace{\begin{bmatrix} R_0 & R_1 & \cdots & R_k \end{bmatrix}}_{R'} \underbrace{\begin{bmatrix} 1 & & & 1 \\ & \ddots & \iddots & \\ & & 1 & \end{bmatrix}}_{\Psi}, \tag{4}$$

the optimization in (3) is actually over the vector $R' \in \mathbb{R}^{1\times(k+1)}$.

The link between sum-of-exponentials-modeling (3) and low-rank approximation is given by

$$\widehat{y} \in \mathscr{B} = \ker R(\sigma) \quad \Longleftrightarrow \quad R\mathscr{H}_{n+1}(\widehat{y}) = 0 \quad \text{and} \quad R \neq 0. \tag{5}$$

Using (4) and (5), we obtain the following Hankel structured low-rank approximation problem that is equivalent to (3)

$$\begin{array}{lll} \text{minimize} & \text{over } \widehat{y} \text{ and } R' \in \mathbb{R}^{1\times(k+1)} & \|y_{\mathrm{d}} - \widehat{y}\| \\ \text{subject to} & R'\Psi\mathscr{H}_{n+1}(\widehat{y}) = 0 \quad \text{and} & R' \neq 0. \end{array} \tag{6}$$

This problem can be solved by the variable projection method and software of (**varpro**; "Software for weighted structured low-rank approximation").

```
function [sys, yh, info] = sem_slra1(y, n, sys_ini)
if exist('sys_ini'), opt.Rini = poly(eig(sys_ini)); end
s.m = n + 1; I = eye(n / 2 + 1); opt.psi = [I I(:, end-1:-1:1)];
[yh, info] = slra(y, s, n, opt);
sys = ss(tf(1, info.Rh, -1)); sys = ss(sys.a, [], sys.c, [], -1);

function [sys, yh, info] = sem_slra2(y, n, sys_ini)
if exist('sys_ini'), opt.sys0 = sys_ini; else, opt = []; end
[sys, info, yh] = ident({y(:) flipud(y(:))}, 0, n, opt);
```

3.3 Subspace methods for sum-of-exponentials modeling

Problem (3) is a nonconvex optimization problem that does not admit an analytic solution. The method of ("Software for weighted structured low-rank approximation") is based on local optimization. In this section, we present two other heuristics for solving problem (3). These methods ignore the Hankel structure, *i.e.*, they do unstructured low-rank approximation and subsequently estimate the model parameters from the resulting low-rank matrix factorization.

Ignoring the Hankel structure, we can compute a one-dimensional approximate left kernel of $\Psi\mathscr{H}_{n+1}(y_{\mathrm{d}})$ by the singular value decomposition. A basis vector R' then leads to an approximate model with kernel parameter $R'\Psi$. Equivalently, the kernel parameter can be computed from the singular value decomposition of the mosaic-Hankel

matrix

$$\begin{bmatrix} \mathscr{H}_{n+1}(y_\mathrm{d}) & \mathscr{H}_{n+1}(\operatorname{rev} y_\mathrm{d}) \end{bmatrix}. \tag{7}$$

The resulting methods (see Algorithms 1 and 2) are known as forward-backwards linear prediction due to the time-reversibility of the identified model $\ker R(\sigma)$.

Algorithm 1 Forward-backward linear prediction, version 1

Input: Sequence y_d and an even natural number n.

1. Construct the Hankel-plus-Toeplitz matrix $H := \Psi\mathscr{H}_{n+1}(y_\mathrm{d})$.
2. Compute the left singular vector R' of H, related to the smallest singular value.
3. Define $R := R'\Psi$.

Output: Parameter R of a kernel representation of the model.

```
function sys = sem_ker1(y, n)
I = eye(n / 2 + 1); psi = [I I(:, end-1:-1:1)];
r = psi * lra(psi * blkhank(y, n + 1), n);
sys = ss(tf(1, r, -1)); sys = ss(sys.a, [], sys.c, [], -1);
```

Algorithm 2 Forward-backward linear prediction, version 2

Input: Sequence y_d and an even natural number n.

1. Construct the mosaic-Hankel matrix $H := \begin{bmatrix} \mathscr{H}_{n+1}(y_\mathrm{d}) & \mathscr{H}_{n+1}(\operatorname{rev} y_\mathrm{d}) \end{bmatrix}$.
2. Compute the left singular vector R of H related to the smallest singular value.

Output: Parameter R of a kernel representation of the model.

```
function sys = sem_ker2(y, n)
r = lra(moshank({y(:) flipud(y(:))}, n + 1), n);
sys = ss(tf(1, r, -1)); sys = ss(sys.a, [], sys.c, [], -1);
```

Empirical observations show that subspace methods give more accurate estimates of the model parameters when the number of rows L of the Hankel matrix $\mathscr{H}_L(y_\mathrm{d})$ is chosen, so that $\mathscr{H}_L(y_\mathrm{d})$ is as square as possible. In the forward-backward linear prediction method, $L = n+1$, so that with $T \gg n$, (7) is a highly skewed rectangular matrix. To avoid this issue, next, we present a modification of Kung's method.

Kung's method ("A new identification method and model reduction algorithm via singular value decomposition") is based on realization theory ("Effective construction of linear state-variable models from input/output functions").

It splits the identification problem into two steps. For exact data,

1. computation of a rank revealing factorization

$$\mathscr{H}_L(y_\mathrm{d}) = \mathscr{O}\mathscr{C}, \quad \text{with } \mathscr{O} \in \mathbb{R}^{L\times n} \text{ and } \mathscr{C}^{n\times(T-L)},$$

   where L is a user defined parameter of the method, chosen as a natural number between $n+1$ and $T-n$,

2. solution of a system of linear equations $\overline{\mathscr{O}}\widehat{A} = \underline{\mathscr{O}}$, for $\widehat{A}$, 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 the poles of the identified system. For noisy data, both Steps 1 and 2 involve approximation. Step 1 is an approximate rank revealing factorization, computed by a rank-n truncation of the singular value decomposition, *i.e.*, unstructured low-rank approximation and Step 2 is an approximate solution in the ordinary least-squares sense. Choosing the parameter L as the nearest integer to $T/2$ results in as square as possible Hankel matrix $\mathscr{H}_L(y_\mathrm{d})$.

```
function sys = kung(y, n)
T = length(y); L = round(T / 2);
[r, O] = lra(blkhank(y, L), n);
a = O(1:end-1, :) \ O(2:end, :);
sys = ss(a, [], O(1, :), [], -1);
```

The modification of Kung's method for time-reversible systems is:

1. computation of a rank revealing factorization $H = \mathscr{O}\mathscr{C}$ of the mosaic-Hankel matrix $H := \begin{bmatrix}\mathscr{H}_L(y_\mathrm{d}) & \mathscr{H}_L(\operatorname{rev} y_\mathrm{d})\end{bmatrix}$, where L is a user defined parameter of the method, chosen as a natural number between $n+1$ and $T-n$,

2. computation of a kernel parameter $R \in \mathbb{R}^{1\times(n+1)}$ of the model using the forward-backwards linear prediction method applied to the set of time series $y^i := \big(\mathscr{O}_{1i}, \ldots, \mathscr{O}_{Li}\big)$, for $i = 1, \ldots, n$, *i.e.*, computation of a one-dimensional approximate left kernel of the mosaic-Hankel matrix

$$\begin{bmatrix}\mathscr{H}_{n+1}(\mathrm{y}^1) & \mathscr{H}_{n+1}(\operatorname{rev}\mathrm{y}^1) & \cdots & \mathscr{H}_{n+1}(\mathrm{y}^n) & \mathscr{H}_{n+1}(\operatorname{rev}\mathrm{y}^n)\end{bmatrix}.$$

A basis vector R for the left kernel defines a kernel representation of the identified model. As in the classical Kung's method, with noisy data Steps 1 and 2 are performed approximately, using the singular value decomposition and least-squares approximation, respectively. Choosing the parameter L as the nearest integer to $T/2$ then tends to improves the performance of the method.

```
function sys = sem_kung(y, n)
T = length(y); L = round(T / 2);
[r, O] = lra(moshank({y(:), y(end:-1:1)}, L), n); revO = flipud(O);
H = []; for i = 1:n, H = [H blkhank(O(:, i), n+1) blkhank(revO(:, i), n+1)]; end
r = lra(H, n); sys = ss(tf(1, r, -1)); sys = ss(sys.a, [], sys.c, [], -1);
```

4 Circulant embedding

Let $y = \big(y(1),\ldots,y(T)\big)$ be a T-samples long discrete-time signal. The circulant matrix defined by y is

$$\mathscr{T}(y) := \begin{bmatrix} y(1) & y(T) & y(T-1) & \cdots & y(2) \\ y(2) & y(1) & y(T-2) & \cdots & y(3) \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ y(T) & y(T-1) & y(T-2) & \cdots & y(1) \end{bmatrix}.$$

```
function c = y2c(y), c = toeplitz(y, [y(1); vec(y(end:-1:2))]);
```

Assume that the *periodic extension* of y on the interval $[1,\infty)$

$$\mathrm{PE}(y) := \big(y(1),\ldots,y(T),y(1),\ldots,y(T),\ldots\big)$$

is exact for a model $\mathscr{B} \in \mathscr{M}_n$. It follows then from the linearity and time-invariance properties of $\mathscr{B}$, that the columns of the circulant matrix $\mathscr{T}(y)$ are all T-samples long trajectories of $\mathscr{B}$. Therefore, the rank of $\mathscr{T}(y)$ is at most equal to n.

In case of noisy data $y_\mathrm{d} = \bar{y} + e$, the rank of $\mathscr{T}(y_\mathrm{d})$ is generically equal to T. The circulant embedding method does a rank-n approximation of $\mathscr{T}(y_\mathrm{d})$ by truncation of the singular value decomposition. The resulting approximation matrix is also circulant. The signal $\widehat{y}$ that approximates the data y_d is extracted from the approximation matrix by selecting its first column.

```
function y = c2y(c), y = c(:, 1);
```

Let $\mathscr{M}_n' \subset \mathscr{M}_n$ be the class of models $\mathscr{B} \in \mathscr{M}_n$, which have frequencies $\widehat{\omega}_i$, $i = 1,\ldots,n$ of on the DFT grid, *i.e.*,

$$\omega_i \in \left\{ 0\frac{2\pi}{T}, 1\frac{2\pi}{T}, \ldots, (T-1)\frac{2\pi}{T} \right\}. \tag{8}$$

Note that $\mathscr{M}_n'$ is a finite set. A model $\widehat{\mathscr{B}}$ identified by the circulant embedding method with rank-n specification is in $\mathscr{M}_n'$.

Proposition 3. *The approximation $\widehat{y}$ obtained by the circulant embedding method with rank-n specification is such that $\widehat{y} \in \widehat{\mathscr{B}} \in \mathscr{M}_n$.*

The implication of Proposition 3 is that the approximation $\widehat{y}$, computed by the circulant embedding method is an exact trajectory of a model in the model class $\mathscr{M}_n$. Therefore, the parameters of this model can be obtained from $\widehat{y}$ by a standard exact identification method, *e.g.*, the classical Kung's method.

The complete circulant embedding method is summarized in Algorithm 3.

Algorithm 3 Circulant embedding method

Input: Sequence y_d and an even natural number n.

1. Compute the low-rank approximation $\widehat{Y}$ of the circulant matrix $Y = \mathscr{T}(y_\mathrm{d})$.
2. Let $\widehat{y}$ be the first column of $\widehat{Y}$.
3. Compute the model parameters from $\widehat{y}$, using the classical Kung's method.

Output: Parameters (A, C) of a state-space representation of the model.

Note 4. Since by construction the signal $\widehat{y}$ computed by the circulant embedding method is exact for a model $\widehat{\mathscr{B}}$ in the model class $\mathscr{M}_n'$, Step 3 of Algorithm 3 does not involve approximation.

```
function [sys, yh] = sem_circ(y, n)
c = y2c(y); [r, p, ch] = lra(c, n); yh = c2y(ch); sys = kung(yh, n);
```

The circulant matrix is diagonalized by the DFT and its eigenvalues are equal to the discrete Fourier coefficients. This fact, allows us to do fast low-rank approximation of a circulant matrix by using the fast Fourier transform. This shows that the circulant embedding method is just a formalization of the classical frequency analysis approach for detection of a periodic signal in noise by looking for peaks in the spectrum.

```
function [sys, yh] = sem_circ_fast(y, n)
yf  = fft(y);  [~, I] = sort(abs(yf), 1, 'descend');
yfh = zeros(size(yf)); yfh(I(1:n)) = yf(I(1:n)); yh = real(ifft(yfh));
sys = kung(yh, n);
```

Algorithm 4 Fast circulant embedding method

Input: Sequence y_d and an even natural number n.

1. Compute the DFT y_f of y_d, using the fast Fourier transform.
2. Let $\widehat{y}_\mathrm{f}$ be the signal obtained from y_f by setting to zero all but the n largest elements.
3. Compute the inverse DFT $\widehat{y}$ of $\widehat{y}_\mathrm{f}$.
4. Compute the model parameters from $\widehat{y}$, using the classical Kung's method.

Output: Parameters (A, C) of a state-space representation of the model.

5 Simulation examples

In this section, we compare empirically the four methods presented in the paper:

- `sem_slra` — the relaxed maximum likelihood method of Section 3.2, implemented by the variable projections method of ("Software for weighted structured low-rank approximation"),
- `kung` — the classical Kung's method ("A new identification method and model reduction algorithm via singular value decomposition"),
- `sem_kung` — the modified Kung's method, described in Section 3.3,
- `sem_circ` — the circulant embedding method, described in Section 4.

Section 5.1 describes the simulation setup and shows the results of a Monte Carlo simulation in two cases:

1. true model's frequencies are on the DFT grid, and
2. true model's frequencies are unconstrained.

Section 5.2 verifies empirically that the estimates obtained with the relaxed maximum likelihood estimator (3) coincide with the ones obtained with the true maximum likelihood estimator (2). Finally, Section 5.3 verifies the fact that the circulant embedding method tends to the maximum likelihood method when the sample size tends to infinity.

5.1 Simulation setup and comparison of the methods

The simulation setup is as in (1), *i.e.*, output error identification with a true system $\bar{\mathscr{B}} \in \mathscr{M}_n$ and a zero-mean white Gaussian measurement noise. The noise-to-signal ratio is varied from 0 to 0.5 and the estimation results are averaged over $N = 100$ Monte-Carlo repetitions. We consider two experiments: 1) the true signal's frequencies are on the DFT grid (8), 2) the true signal's frequencies are real numbers between 0 and 2π, *i.e.*, they are unrestricted.

Let $\bar{\mathscr{B}}$ be the true model and $\bar{z}_1, \ldots, \bar{z}_n$ its poles. We evaluate the accuracy of the estimated models by the average relative parameter's error

$$\frac{1}{N} \sum_{k=1}^{N} \frac{\|\bar{\theta} - \widehat{\theta}^k\|}{\|\bar{\theta}\|},$$

where $\bar{\theta}$ is the true parameter vector $\bar{\theta} := \begin{bmatrix} 1 & \bar{\theta}_1 & \cdots & \bar{\theta}_n \end{bmatrix}$ with

$$1 + \bar{\theta}_1 z + \cdots + \bar{\theta}_n z^n = \Pi_{i=1}^{n} (z - \bar{z}_i),$$

and $\widehat{\theta}^k$ is the parameter vector of the model $\widehat{\mathscr{B}}^k$ obtained in the k-th Monte Carlo repetition.

```
clear all, addpath ../detss
k = 1;  n = 2 * k;      % number of sines and order of the LTI model
T = 25; N = 100;        % number of samples and number of MC repetitions
np = 10; S = linspace(0, 0.5, np); % noise-to-signal ratio

%% generate a random sum-of-sines signal
if 0 % frequencies on the DFT grid
  w = 2 * pi * randi(T, 1, k) / T; t = vec(0:T-1);
  y0 = sin(w(ones(T, 1), :) .* t(:, ones(1, k))) * rand(k, 1);
  sys0 = kung(y0, n);
else
  rng(0); sys0 = dr_ms_ss(k); y0 = initial(sys0, rand(1, n), T - 1);
end
r0 = poly(eig(sys0)); R0 = r0(ones(N, 1),:);

%% MC simulation
for j  = 1:np
  s = S(j); j = j,
  for i = 1:N
```

```
    yt = randn(T, 1); y = y0 + s * yt / norm(yt) * norm(y0);
    sysh1 = sem_circ(y, n);   r1(i, :) = poly(eig(sysh1));
    sysh2 = sem_kung(y, n);   r2(i, :) = poly(eig(sysh2));
    sysh3 = sem_slra1(y, n);  r3(i, :) = poly(eig(sysh3));
    sysh4 = kung(y, n);       r4(i, :) = poly(eig(sysh4));
  end
  res(:, j) = [norm(R0 - r1, 'fro')
               norm(R0 - r2, 'fro')
               norm(R0 - r3, 'fro')
               norm(R0 - r4, 'fro')] / norm(R0, 'fro');
end
plot(S, res, '-o')
legend('sem\_circ', 'sem\_kung', 'sem\_slra', 'kung', 'Location', 'northwest')
ylabel('relative parameter error'), xlabel('noise-to-signal ratio')
ax = axis; axis([S(1) S(end) ax(3:4)])
```

The reported results are for an example with $k = 1$ random sines and $T = 25$ samples of a random trajectory of the model. Figure 1 shows the estimation errors of the methods when the true signal's frequency is on the DFT grid and Figure 2 shows the results estimation errors of the methods when the true signal's frequency are unrestricted. The results confirm the fact that the more prior knowledge a method uses, the more accurate the estimated parameters are. In addition, nonlinear optimization of the maximum likelihood cost function (which can also be viewed as using prior knowledge) improves the suboptimal results of Kung's and modified Kung's methods.

```
clear all, addpath ../detss
k = 1;  n = 2 * k;      % number of sines and order of the LTI model
T = 25; N = 100;        % number of samples and number of MC repetitions
np = 10; S = linspace(0, 1.5, np); % noise-to-signal ratio

%% generate a random sum-of-sines signal
w = 2 * pi * randi(T, 1, k) / T; t = vec(0:T-1);
y0 = sin(w(ones(T, 1), :) .* t(:, ones(1, k))) * rand(k, 1);
sys0 = kung(y0, n); r0 = poly(eig(sys0)); R0 = r0(ones(N, 1),:);

%% MC simulation
for j  = 1:np
  s = S(j); j = j,
  for i = 1:N
    yt = randn(T, 1); y = y0 + s * yt / norm(yt) * norm(y0);
```

```
    sysh1 = sem_circ(y, n);    r1(i, :) = poly(eig(sysh1));
  end
  res(:, j) = [norm(R0 - r1, 'fro')] / norm(R0, 'fro');
end
plot(S, res, '-o')
ylabel('relative parameter error'), xlabel('noise-to-signal ratio')
ax = axis; axis([S(1) S(end) ax(3:4)])
```

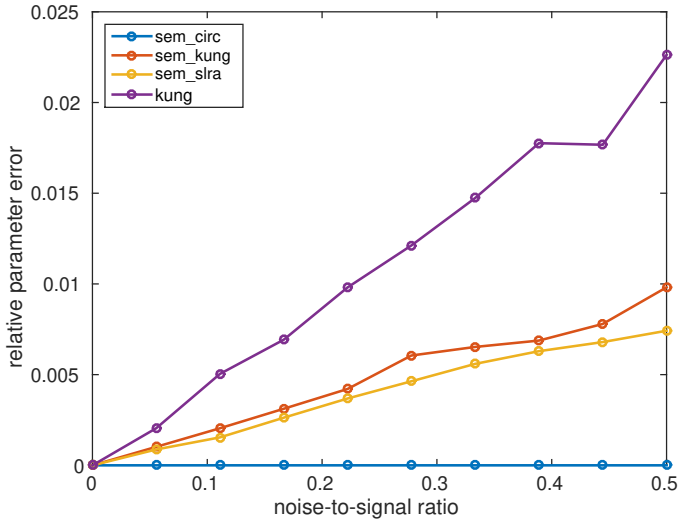

Figure 1: When the true model's frequencies are on the DFT grid, the circulant embedding method gives exact results (for as high noise-to-signal ratio as 1). From the other methods, the modified Kung's method outperforms the classical Kung's method and the structured low-rank approximation method outperforms the modified Kung's method. The results are consistent with the level of prior knowledge used by the methods: only the circulant embedding method uses the prior knowledge that the frequencies are on the DFT grid.

5.2 The relaxed maximum likelihood coincides with the maximum likelihood

In the comparison of methods, presented in Section 5.1, we did not include the maximum likelihood method 2. The reason for this is that up to a certain noise-to-signal level the relaxed maximum likelihood method (3) gives identical results as the maximum likelihood method. This is verified empirically in this section.

We compute the maximum likelihood estimator (2) by brute-force optimization over the frequencies $\omega_1, \ldots, \omega_k$ (using the function `fminsearch` from the Optimization Toolbox of MATLAB). The cost function $\min_{\hat{y} \in \mathscr{B}} \|y_d - \hat{y}\|$ evaluation is a linear least-squares problem (computed using the `misfit` function from (**ident**)). The resulting method is implemented in the function `sem_ml`.

```
function [sys, yh, M] = sem_ml(y, sys_ini, opt)
if ~exist('opt'), opt = []; end
```

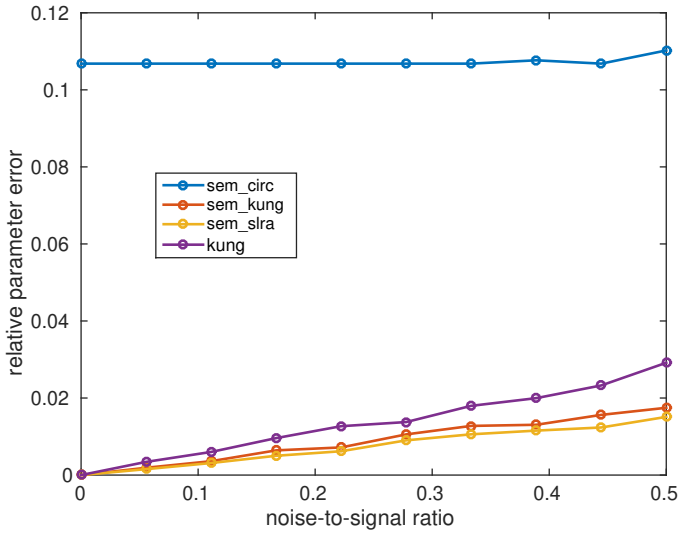

Figure 2: When the true model's frequencies are unconstrained, the circulant embedding method generically results in a biased estimate. Since all other methods are unbiased, for small noise-to-signal ratios the circulant embedding method gives the worst estimates. The performance of the other methods with respect to each other is the same as in the case when the true model's frequencies are on the DFT grid.

```
a_ini = unique(abs(angle(eig(sys_ini))));
a = fminsearch(@(a) cost_fun(a, y), a_ini, opt);
[M, sys, yh] = cost_fun(a, y);

function [M, sys, yh] = cost_fun(a, y)
r = poly(exp(j * [a' -a']));  sys = ss(tf(1, r, -1));
sys = ss(sys.a, [], sys.c, [], -1);  [M, yh] = misfit(y, sys);
```

For the numerical comparison, we use the same simulation setup as in Section 5.1 with unconstrained true model's frequencies. Figure 3 verifies that the solutions obtained by `sem_ml` and `sem_slra` coincide (up to the convergence tolerance of the optimization methods). As initial approximation for the optimization in both cases—maximum likelihood and relaxed maximum likelihood—we use the suboptimal solution obtained by `sem_kung`.

```
clear all
k = 1;  n = 2 * k;      % number of sines and order of the LTI model
T = 25; N = 100;        % number of samples and number of MC repetitions
np = 10; S = linspace(0, 1.5, np); % noise-to-signal ratio

%% generate a random sum-of-sines signal
rng(0); sys0 = dr_ms_ss(k); y0 = initial(sys0, rand(1, n), T - 1);
r0 = poly(eig(sys0)); R0 = r0(ones(N, 1),:);
```

```
%% MC simulation
for j  = 1:np
  s = S(j); j = j,
  for i = 1:N
    yt = randn(T, 1); y = y0 + s * yt / norm(yt) * norm(y0);
    sys_ini = sem_kung(y, n);
    sysh2 = sem_slra1(y, n, sys_ini); r2(i, :) = poly(eig(sysh2));
    sysh1 = sem_ml(y, sysh2);         r1(i, :) = poly(eig(sysh1));
  end
  res(:, j) = [norm(R0 - r1, 'fro')
               norm(R0 - r2, 'fro')] / norm(R0, 'fro');
end
plot(S, res, '-o')
legend('sem\_ml', 'sem\_slra', 'Location', 'northwest')
ylabel('relative parameter error'), xlabel('noise-to-signal ratio')
ax = axis; axis([S(1) S(end) ax(3:4)])
```

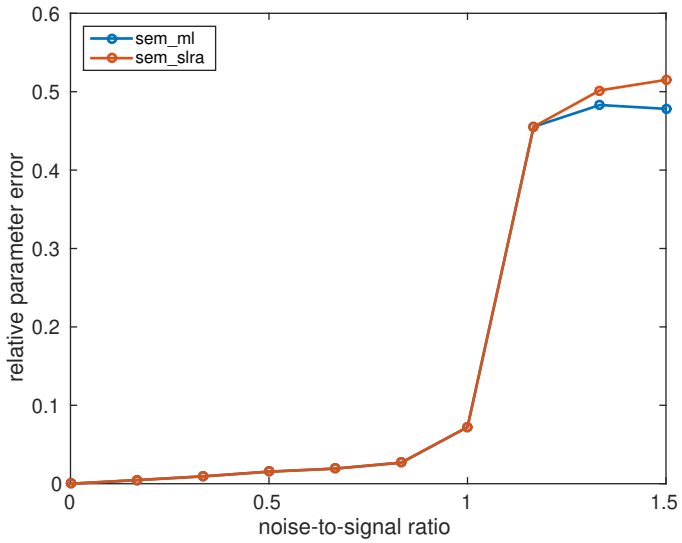

Figure 3: Up to noise-to-signal ratio 1.2, the maximum likelihood and relaxed maximum likelihood methods give identical results.

5.3 Accuracy of the circulant embedding method when the sample size tends to infinity

As illustrated by the numerical example in Section 5.1, when the true model's frequencies are unconstrained, the circulant embedding method (generically) gives a biased estimate. The reason for the bias is that the estimated frequencies are constraint to be on the DFT grid. The bias is therefore determined by the spacing of the DFT grid. The spacing of the DFT grid, in turn, is determined by the number of samples T. We verify empirically that as T increases

the bias of the circulant embedding method decreases. The plot in Figure 4 shows the relative estimation error as a function of the number of samples for the same simulation setup as in Section 5.1.

```
clear all
k = 1;  n = 2 * k;      % number of sines and order of the LTI model
s = 0.25;                % fixed noise-to-signal ratio
np = 10; TT = linspace(25, 250, np); % different sample sizes
T = TT(end); N = 100;
rng(0); sys0 = dr_ms_ss(k); y0 = initial(sys0, rand(1, n), T - 1);
r0 = poly(eig(sys0)); R0 = r0(ones(N, 1),:);

%% MC simulation
for j  = 1:np
  T = TT(j); j = j,
  for i = 1:N
    yt = randn(T, 1); y = y0(1:T) + s * yt / norm(yt) * norm(y0(1:T));
    sysh1 = sem_circ(y, n);    r1(i, :) = poly(eig(sysh1));
    sysh2 = sem_kung(y, n);    r2(i, :) = poly(eig(sysh2));
    sysh3 = sem_slra1(y, n, sysh2);  r3(i, :) = poly(eig(sysh3));
    sysh4 = kung(y, n);        r4(i, :) = poly(eig(sysh4));
  end
  res(:, j) = [norm(R0 - r1, 'fro')
               norm(R0 - r2, 'fro')
               norm(R0 - r3, 'fro')
               norm(R0 - r4, 'fro')] / norm(R0, 'fro');
end
plot(TT, res, '-o')
legend('sem\_circ', 'sem\_kung', 'sem\_slra', 'kung', 'Location', 'northeast')
ylabel('relative parameter error'), xlabel('number of samples')
ax = axis; axis([TT(1) TT(end) ax(3:4)])
```

6 Conclusions

We developed a relaxed maximum likelihood method, a modification of Kung's method, and a circulant embedding method for sum-of-exponentials modeling. The relaxed maximum likelihood and modified Kung's methods 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 identified model.

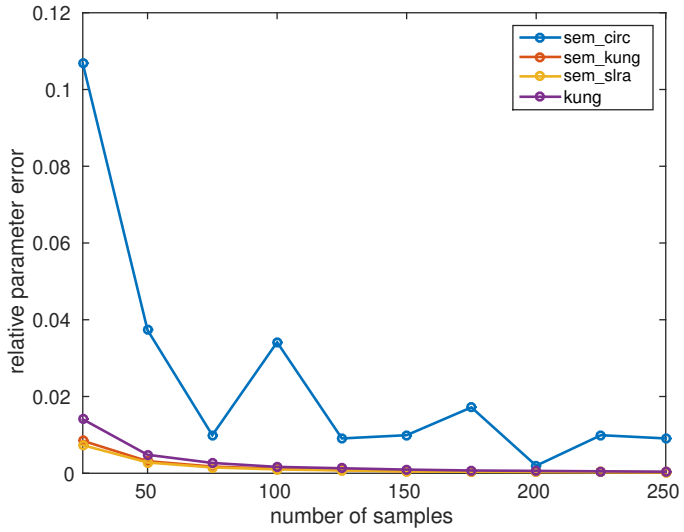

Figure 4: The relative estimation error of all methods decreases as the number of samples increases. For `sem_circ`, this is due to the decreased spacing of the DFT grid.

Simulation results show that for as high noise-to-signal ratios as 1, the necessary condition is also sufficient, *i.e.*, the identified models by the methods are in fact sum-of-exponentials models.

The circulant embedding method does a low-rank approximation of a circulant matrix, which results in a circulant low-rank approximation matrix. This method is closely related to the classical Fourier spectral analysis and can be implemented by the fast Fourier transform. The circulant embedding method, however, imposes the constraint that the frequencies of the identified model are on the DFT grid. When the true model's frequencies are not on the DFT grid, the circulant embedding method results in a biased estimate.

Imposing constraints on the kernel parameter of the model allows us to enforce other properties on the model, *e.g.*, offset, trend, and seasonal components. Combining these additional constraint with the time-reversibility property, obtained by palindromic kernel structure, is an interesting topic for future research.

Acknowledgments

The research leading to these results has received funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007–2013) / ERC Grant agreement number 258581 "Structured low-rank approximation: Theory, algorithms, and applications", Research Foundation Flanders (FWO) projects G028015N "Decoupling multivariate polynomials in nonlinear system identification", G090117N "Block-oriented nonlinear identification using Volterra series", and Dieter Verbeke's PhD grant; and Fonds de la Recherche Scientifique (FNRS) – FWO Vlaanderen under Excellence of Science (EOS) Project number 30468160 "Structured low-rank matrix / tensor approximation: numerical optimization-based algorithms and applications".

References

Chu, M. and R. Plemmons (2003). "Real-valued, low rank, circulant approximation". In: *SIAM J. Matrix Anal. Appl.* 24.3, pp. 645–659.

Ho, B. L. and R. E. Kalman (1966). "Effective construction of linear state-variable models from input/output functions". In: *Regelungstechnik* 14.12, pp. 545–592.

Kay, S. (1988). *Modern spectral estimation: Theory and applications*. Prentice-Hall.

Kumaresan, R. and A. Shaw (1988). "Superresolution by structured matrix approximation". In: *IEEE Trans. Antennas Propagation* 36, pp. 34–44.

Kung, S. (1978). "A new identification method and model reduction algorithm via singular value decomposition". In: *Proc. 12th Asilomar Conf. Circuits, Systems, Comp.* Pp. 705–714.

Markovsky, I. (2012). *Low Rank Approximation: Algorithms, Implementation, Applications*. First. Springer. (Citations: 189 in GS).

Markovsky, I. and P.-L. Dragotti (2018). "Using structured low-rank approximation for sparse signal recovery". In: *Latent Variable Analysis and Signal Separation*. Lecture Notes in Computer Science. Springer, pp. 479–487. DOI: 10.1007/978-3-319-93764-9_44.

Markovsky, I. and S. Rao (25). "Palindromic polynomials, time-reversible systems, and conserved quantities". In: *16th Mediterranean Conf. on Control and Automation*. Ajaccio, France, pp. 125–130.

Markovsky, I. and K. Usevich (2014). "Software for weighted structured low-rank approximation". In: *J. Comput. Appl. Math.* 256, pp. 278–292. (Citations: 11 in WoS, 39 GS).

Markovsky, I. et al. (2006). *Exact and Approximate Modeling of Linear Systems: A Behavioral Approach*. SIAM. (Citations: 106 in GS).

Polderman, J. and J. C. Willems (1998). *Introduction to Mathematical Systems Theory*. New York: Springer-Verlag.

Tufts, D. and R. Kumaresan (1982). "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood". In: *Proc. IEEE* 70.9, pp. 975–989.

Willems, J. C. (2007). "The behavioral approach to open and interconnected systems: Modeling by tearing, zooming, and linking". In: *Control Systems Magazine* 27, pp. 46–99.