# Modifications of Prony's Method for the Recovery and Sparse Approximation with Generalized Exponential Sums

**Ingeborg Keller and Gerlind Plonka**

**Abstract** In this survey we describe some modifications of Prony's method. In particular, we consider the recovery of general expansions into eigenfunctions of linear differential operators of first order. We show, how these expansions can be reconstructed from function samples using generalized shift operators. We derive an ESPRIT-like algorithm for the generalized recovery method and illustrate, how the method can be simplified if some frequency parameters are known beforehand. Furthermore, we present a modification of Prony's method for sparse approximation with exponential sums which leads to a non-linear least-squares problem.

**Keywords** Generalized Prony method · Generalized exponential sums · Shifted Gaussians · Eigenfunctions of linear operators · Sparse signal approximation · Nonstationary signals

## 1 Introduction

The recovery and sparse approximation of structured functions is a fundamental problem in many areas of signal processing and engineering. In particular, exponential sums and their generalizations play an important role in time series analysis and in system theory [13, 15], in the theory of annihilating filters, and for the recovery of signals with finite rate of innovation [3, 10, 26, 35, 37], as well as for linear prediction methods [17, 34]. For system reduction, Prony's method is related to the problem of low-rank approximation of structured matrices (particularly Hankel matrices) and corresponding nonlinear least-squares problems [18, 36]. There is a close relation between Prony's method and Padé approximation [4, 9]. Extended models have also been studied in [16]. Exponential sums started to become more important for sparse approximation of smooth functions, see [5, 6, 12, 23], and

I. Keller · G. Plonka (✉)

University of Göttingen, Institute for Numerical and Applied Mathematics, Göttingen, Germany
e-mail: i.keller@math.uni-goettingen.de; plonka@math.uni-goettingen.de
http://na.math.uni-goettingen.de

this question is closely related to approximation in Hardy spaces and the theory of Adamjan, Arov and Krein, see [1, 2, 22].

## 1.1 The Classical Prony Method

A fundamental problem discussed in many papers is the recovery of exponential sums of the form

$$f(x) := \sum_{j=1}^{M} c_j \, e^{\alpha_j x} = \sum_{j=1}^{M} c_j \, z_j^x, \qquad \text{with} \quad z_j := e^{\alpha_j}, \tag{1}$$

where the coefficients $c_j \in \mathbb{C} \setminus \{0\}$ as well as the pairwise different frequency parameters $\alpha_j \in \mathbb{C}$ or equivalently, $z_j \in \mathbb{C} \setminus \{0\}$ are unknown. For simplicity we assume that the number of terms $M$ is given beforehand. One important question appears: What information about $f$ is needed in order to solve this recovery problem uniquely?

The classical Prony method uses the equidistant samples $f(0), f(1), \ldots, f(2M - 1)$. Indeed, if we suppose that $\operatorname{Im} \alpha_j$, $j = 1, \ldots, M$, lies in a predefined interval of length $2\pi$, as e.g. $[-\pi, \pi)$, these $2M$ samples are sufficient. This can be seen as follows.

We can view $f(x)$ as the solution of a homogeneous linear difference equation of order $M$ with constant coefficients and try to identify these constant coefficients in a first step. We define the characteristic polynomial (Prony polynomial) with the help of its (yet unknown) zeros $z_j = e^{\alpha_j}$, $j = 1, \ldots, M$, and consider its monomial representation,

$$p(z) := \prod_{j=1}^{M} (z - e^{\alpha_j}) = z^M + \sum_{k=0}^{M-1} p_k \, z^k.$$

Then the coefficients $p_k$, $k = 0, \ldots, M - 1$, and $p_M = 1$ satisfy

$$\sum_{k=0}^{M} p_k f(k + m) = \sum_{k=0}^{M} p_k \sum_{j=1}^{M} c_j z_j^{k+m} = \sum_{j=1}^{M} c_j z_j^m \sum_{k=0}^{M} p_k z_j^k = \sum_{j=1}^{M} c_j z_j^m \, p(z_j) = 0$$

for all $m \in \mathbb{Z}$. Thus the coefficients $p_k$ of the linear difference equation can be computed by solving the linear system

$$\sum_{k=0}^{M-1} p_k \, f(k + m) = -f(M + m), \qquad m = 0, \ldots, M - 1.$$

Knowing $p(z)$, we can simply compute its zeros $z_j = e^{\alpha_j}$, and in a further step the coefficients $c_j$, $j = 1, \ldots, M$, by solving the (overdetermined) system

$$f(\ell) = \sum_{j=1}^{M} c_j z_j^{\ell}, \qquad \ell = 0, \ldots, 2M - 1.$$

In practice there are different numerical algorithms available for this method, which take care for the inherit numerical instability of this approach, see e.g. [14, 24, 27, 29, 31]. Note that for a given arbitrary vector $(f_k)_{k=0}^{2M-1}$, the interpolation problem

$$f_k = \sum_{j=1}^{M} c_j z_j^{k}, \qquad k = 0, \ldots, 2M - 1,$$

may not be solvable, see e.g. [8]. The characteristic polynomial $p(z)$ of the homogeneous difference equation $\sum_{k=0}^{M} p_k f_{k+m} = 0$, $m = 0, \ldots, M - 1$, may have zeros with multiplicity greater than 1, whereas the exponential sum in (1) is only defined for pairwise different zeros. In this paper, we will exclude the case of zeros with multiplicity greater than 1. However, the zeros $e^{\alpha_j}$ of the characteristic polynomial $p(z)$ resp. the parameters $\alpha_j$, $j = 1, \ldots, M$, may be arbitrarily close. This may lead to highly ill-conditioned system matrices $(f(k + m))_{k,m=0}^{M-1}$.

## 1.2 Content of This Paper

In this paper, we will particularly consider the following questions.

1. How can we generalize Prony's method in order to recover other expansions than (1)?
2. What kind of information is needed in order to recover the considered expansion?
3. How can we modify Prony's method such that we are able to optimally approximate a given (large) vector of function values in the Euclidean norm by a sparse exponential sum?

To tackle the first question, we introduce the operator based general Prony method in [33] and apply it to study expansions of the form

$$f(x) = \sum_{j=1}^{M} c_j H(x) e^{\alpha_j G(x)}, \qquad x \in [a, b] \subset \mathbb{R}, \tag{2}$$

where $c_j, \alpha_j \in \mathbb{C}$, $c_j \neq 0$, $\alpha_j$ pairwise different, $G, H \in C^{\infty}(\mathbb{R})$ are predefined functions, where $G$ is strictly monotone on $[a, b]$, and $H$ is nonzero on $[a, b]$. This model covers many interesting examples as e.g. shifted Gaussians, generalized

monomial sums and others. For the expansions (2) we will derive different sets of samples which are sufficient for the recovery of all model parameters, thus answering the second question.

In regard to question 3 we will focus on the case of $f$ as in (1) and (2) and show how the methods can be modified for optimal approximation, and how to treat the case of noisy measurements.

The outline of the paper is as follows. First we will introduce the idea of the operator based Prony method by looking at the recovery problem of the classical exponential sum from different angles. In Sect. 3, we study the recovery of the more general expansion $f$ of the form (2). We will show that (2) can be viewed as an expansion into eigenfunctions of a differential operator of first order and thus, according to the generalized Prony method in [21], can be recovered using higher order derivative values of $f$. We will show, how to find a new generalized shift operator possessing the same eigenfunctions. This leads to a recovery method that requires only function values of $f$ instead of derivative values. The idea will be further illustrated with several examples in Sect. 3.3. Section 4 is devoted to the numerical treatment of the generalized recovery method. We will derive an ESPRIT-like algorithm for the computation of all unknown parameters in the expansion (2). This algorithm also applies if the number of terms $M$ in the expansion (2) is not given beforehand. Furthermore, we show in Sect. 4.3, how the recovery problem can be simplified if some frequencies $\alpha_j$ are known beforehand (while the corresponding coefficients $c_j$ are unknown). Finally, in Sect. 5 we study the optimal approximation with exponential sums in the Euclidean norm. This leads to a nonlinear least squares problem which we tackle directly using a Levenberg-Marquardt iteration. Our approach is essentially different from earlier algorithms, as e.g. [7, 19, 20, 38].

## 2   Operator Based View to Prony's Method

In order to tackle the questions 1 and 2 in Sect. 1.2, we start by reconsidering Prony's method. As an introductory example, we study the exponential sum in (1) from a slightly different viewpoint. For $h \in \mathbb{R} \setminus \{0\}$ let $S_h : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ be the shift operator given by $S_h f := f(\cdot + h)$. Then, for any $\alpha \in \mathbb{C}$, the function $e^{\alpha x}$ is an eigenfunction of $S_h$ with corresponding eigenvalue $e^{\alpha h}$, i.e.,

$$(S_h e^{\alpha \cdot})(x) = e^{\alpha(x+h)} = e^{\alpha h} e^{\alpha x}.$$

Therefore, the exponential sum in (1) can be seen as a sparse expansion into eigenfunctions of the shift operator $S_h$. The eigenvalues $e^{\alpha_j h}$ are pairwise different, if we assume that $\operatorname{Im} \alpha_j \in [-\pi/h, \pi/h)$. Now we consider the Prony polynomial

$$p(z) := \prod_{j=1}^{M}(z - e^{\alpha_j h}) = \sum_{k=0}^{M} p_k \, z^k$$

defined by the (unknown) eigenvalues $e^{\alpha_j h}$ corresponding to the active eigenfunctions in the expansion $f$ in (1). Then, for any predefined $x_0 \in \mathbb{R}$ we have

$$\sum_{k=0}^{M} p_k f(x_0 + h(k + m)) = \sum_{k=0}^{M} p_k (S_h^{k+m} f)(x_0) = \sum_{k=0}^{M} p_k \sum_{j=1}^{M} c_j (S_h^{k+m} e^{\alpha_j \cdot})(x_0)$$

$$= \sum_{j=1}^{M} c_j \sum_{k=0}^{M} p_k e^{\alpha_j(hm+hk)} e^{\alpha_j x_0}$$

$$= \sum_{j=1}^{M} c_j e^{\alpha_j hm} p(e^{\alpha_j h}) e^{\alpha_j x_0} = 0, \qquad (3)$$

i.e., we can reconstruct $p(z)$ by solving this homogeneous system for $m = 0, \dots, M-1$. We conclude that the exponential sum in (1) can be recovered from the samples $f(h\ell + x_0)$, $\ell = 0, \dots, 2M - 1$. This is a slight generalization of the original Prony method in Sect. 1.1 as we introduced an arbitrary sampling distance $h \in \mathbb{R} \setminus \{0\}$ and a starting point $x_0 \in \mathbb{R}$.

Moreover, we can also replace the samples $(S_h^{k+m} f)(x_0) = f(h(k + m) + x_0)$ in the above computation (3) by any other representation of the form $F(S_h^{k+m} f)$, where $F : C^\infty(\mathbb{R}) \to \mathbb{C}$ is a linear functional satisfying $F(e^{\alpha \cdot}) \neq 0$ for all $\alpha \in \mathbb{C}$, since

$$\sum_{k=0}^{M} p_k F(S_h^{k+m} f) = \sum_{k=0}^{M} p_k \sum_{j=1}^{M} c_j F(S_h^{k+m} e^{\alpha_j \cdot}) = \sum_{j=1}^{M} c_j e^{\alpha_j hm} p(e^{\alpha_j h}) \, F(e^{\alpha_j \cdot}) = 0.$$

Any set of samples of the form $F(S_h^\ell f)$, $\ell = 0, \dots, 2M - 1$, is sufficient to recover $f$ in (1), and the above set is obtained using the point evaluation functional $F = F_{x_0}$ with $F_{x_0} f := f(x_0)$ with $x_0 \in \mathbb{R}$. For further generalizations of the sampling scheme we refer to [33].

This operator-based view leads us to the generalized Prony method introduced in [21], which can be applied to recover any sparse expansion into eigenfunctions of a linear operator.

To illustrate this idea further, let us consider now the differential operator $D : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ given by $(Df)(x) := f'(x)$ with $f'$ denoting the first derivative of $f$. Due to

$$(De^{\alpha \cdot})(x) = \alpha \, e^{\alpha x}$$

we observe that exponentials $e^{\alpha x}$ are eigenfunctions of $D$ corresponding to the eigenvalues $\alpha \in \mathbb{C}$. Thus, the sum of exponentials in (1) can also be seen as a sparse expansion into eigenfunctions of the differential operator $D$. Similarly as before let now

$$\widetilde{p}(z) := \prod_{j=1}^{M}(z - \alpha_j) = \sum_{k=0}^{M} \widetilde{p}_k \, z^k$$

be the characteristic polynomial being defined by the eigenvalues $\alpha_j$ corresponding to the active eigenfunctions of $D$ in (1), where again $\widetilde{p}_M = 1$ holds. Choosing the functional $Ff := f(x_0)$ for some fixed $x_0 \in \mathbb{R}$, we find for any integer $m \geq 0$

$$\sum_{k=0}^{M} \widetilde{p}_k F(D^{k+m} f) = \sum_{k=0}^{M} \widetilde{p}_k f^{(k+m)}(x_0) = \sum_{k=0}^{M} \widetilde{p}_k \sum_{j=1}^{M} c_j \alpha_j^{k+m} \, e^{\alpha_j x_0}$$

$$= \sum_{j=1}^{M} c_j \alpha_j^m \, \widetilde{p}(\alpha_j) \, e^{\alpha_j x_0} = 0.$$

Thus we can determine $\widetilde{p}_k$, $k = 0, \ldots, M - 1$, from $\sum_{k=0}^{M} \widetilde{p}_k f^{(k+m)}(x_0) = 0$ for $m = 0, \ldots, M - 1$ and $\widetilde{p}_M = 1$, and recover the zeros $\alpha_j$ of $\widetilde{p}$ in a first step. The $c_j$ are computed in a second step the same way as in the classical case. We conclude that also the sample set $f^{(\ell)}(x_0)$, $\ell = 0, \ldots, 2M - 1$, for any fixed value $x_0 \in \mathbb{R}$, is sufficient to recover $f$. Note that here we do not have any restrictions regarding $\mathrm{Im}\,\alpha_j$.

This example already shows, that there exist many different sample sets that may be used to recover the exponential sum. In particular, each sample set of the form $F(A^{\ell}h)$, $\ell = 0, \ldots, 2M - 1$, where $A : C^{\infty}(\mathbb{R}) \to C^{\infty}(\mathbb{R})$ is a linear operator with eigenfunctions $e^{\alpha x}$ corresponding to pairwise different eigenvalues $\alpha$ (covering the range of $\alpha_j$ in (1)), and where $F$ is an arbitrary (fixed) linear functional satisfying $F(e^{\alpha \cdot}) \neq 0$ for all $\alpha \in \mathbb{C}$, can be employed for recovery.

However, in practice it is usually much easier to provide function samples of the form $f(x_0 + h\ell)$ than higher order derivative values $f^{\ell}(x_0)$ for $\ell = 0, \ldots, 2M - 1$. Therefore, for more general expansions, for example of the form (2), we will raise the following question which has also been investigated in [33]: Suppose we already found a set of samples which is (theoretically) sufficient to recover the expansion at hand. Is it possible to find other sets of samples which can be more easily acquired and also admit a unique recovery of the sparse expansion? In terms of linear operators, we can reformulate this idea: Suppose that we have already found an operator $A$, such that a considered expansion $f$ is a sparse expansion into $M$ eigenfunctions of $A$ (corresponding to pairwise different eigenvalues). Is it possible to find another operator $B$ that possesses the same eigenfunctions, such that the samples $\widetilde{F}(B^{\ell})f$ (with some suitable linear functional $\widetilde{F}$) can be easier obtained than $F(A^{\ell})f$ for $\ell = 0, \ldots, 2M - 1$?

Back to our introductory example for the exponential sum (1). Let the linear functional $F$ be given as $Ff := f(0)$. Assume that we have found the recovery of (1) from the samples $f^{(\ell)}(0)$, $\ell = 0, \ldots, 2M - 1$ first. This sampling set corresponds to the linear differential operator $A = D$ with $Df = f'$. How can we find the shift operator $B = S_h$, knowing just the fact, that (1) can be viewed as a sparse expansion into eigenfunctions of $D$? Is there a simple link between the linear differential operator $D$ and the shift operator $S_h$?

This is indeed the case. Taking $\varphi \in C^\infty(\mathbb{R})$ with $\varphi(x) = \mathrm{e}^{hx}$, and applying $\varphi$ (formally) to $D$, we observe for each exponential $\mathrm{e}^{\alpha x}$, $\alpha \in \mathbb{C}$,

$$\varphi(D)\mathrm{e}^{\alpha \cdot} = \mathrm{e}^{hD}\mathrm{e}^{\alpha \cdot} = \sum_{\ell=0}^{\infty} \frac{h^\ell}{\ell!} D^\ell \mathrm{e}^{\alpha \cdot} = \left(\sum_{\ell=0}^{\infty} \frac{h^\ell}{\ell!} \alpha^\ell\right) \mathrm{e}^{\alpha \cdot} = \mathrm{e}^{\alpha h} \, \mathrm{e}^{\alpha \cdot} = S_h \mathrm{e}^{\alpha \cdot}.$$

Therefore, we have $\varphi(D)f = S_h f$ for $f$ in (1). We note that $\varphi$ also maps the eigenvalues of the differential operator onto the eigenvalues of the shift operator. This idea to switch from differential operators to other more suitable operators will be also applied to general sparse expansions in the next section.

## 3   Recovery of Generalized Exponential Sums

In this section we focus on the recovery of more general sparse expansions. Let $G : \mathbb{R} \to \mathbb{R}$ be a given function in $C^\infty(\mathbb{R})$, which is strictly monotone in a given interval $[a, b] \subset \mathbb{R}$, and let $H : \mathbb{R} \to \mathbb{R}$ be in $C^\infty(\mathbb{R})$ and nonzero in $[a, b]$. We consider expansions of the form

$$f(x) = \sum_{j=1}^{M} c_j \, H(x) \, \mathrm{e}^{\alpha_j G(x)}, \qquad x \in [a, b] \subset \mathbb{R}, \tag{4}$$

with $c_j \in \mathbb{C} \setminus \{0\}$ and pairwise different $\alpha_j \in \mathbb{C}$. Obviously, (1) is a special case of (4) with $G(x) = x$ and $H(x) \equiv 1$. In order to recover $f$, we need to identify the parameters $c_j$ and $\alpha_j$, $j = 1, \ldots, M$.

## 3.1   *Expansion into Eigenfunctions of a Linear Differential Operator*

According to our previous considerations in Sect. 2, we want to apply the so-called generalized Prony method introduced in [21], where we view (4) as an expansion into eigenfunctions of a linear operator.

**Step 1** First we need to find a linear operator $A$ that possesses the functions $H(x)e^{\alpha_j G(x)}$ as eigenfunctions for any $\alpha_j \in \mathbb{C}$. For this purpose, let us define the functions

$$g(x) := \frac{1}{G'(x)}, \qquad \eta(x) := -g(x)\frac{H'(x)}{H(x)} = -\frac{H'(x)}{G'(x)H(x)}, \qquad (5)$$

which are well defined on $[a, b]$, since $G'$ and $H$ have no zeros in $[a, b]$. Then the differential operator $A : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ with

$$Af(x) := g(x)f'(x) + \eta(x)f(x) \qquad (6)$$

satisfies

$$A\left(H(\cdot)e^{\alpha_j G(\cdot)}\right)(x) = g(x)\left(\alpha_j G'(x)H(x) + H'(x)\right)e^{\alpha_j G(x)} + \eta(x)\,H(x)\,e^{\alpha_j G(x)}$$

$$= \alpha_j\,H(x)\,e^{\alpha_j G(x)}, \qquad \alpha_j \in \mathbb{C},$$

i.e., the differential operator $A$ indeed possesses the eigenfunctions $H(x)\,e^{\alpha_j G(x)}$ with corresponding eigenvalues $\alpha_j \in \mathbb{C}$.

**Step 2** To reconstruct $f$ in (4), we can apply a similar procedure as in Sect. 2. Let

$$\widetilde{p}(z) := \prod_{j=1}^M (z - \alpha_j) = \sum_{k=0}^M \widetilde{p}_k\, z^k, \qquad \widetilde{p}_M = 1, \qquad (7)$$

be the characteristic polynomial defined by the (unknown) eigenvalues $\alpha_j$ that correspond to the active eigenfunctions of the operator $A$ in the expansion (4). Let $F : C^\infty(\mathbb{R}) \to \mathbb{C}$ be the point evaluation functional $Ff := f(x_0)$ with $x_0 \in [a, b]$, such that $H(x_0) \neq 0$ and $G'(x_0) \neq 0$. Then, for $f$ as in (4) we observe that

$$\sum_{k=0}^M \widetilde{p}_k\, F(A^{m+k}f) = \sum_{k=0}^M \widetilde{p}_k \sum_{j=1}^M c_j\, F\left(A^{k+m}\left(H(\cdot)\,e^{\alpha_j G(\cdot)}\right)\right)$$

$$= \sum_{k=0}^M \widetilde{p}_k \sum_{j=1}^M c_j\, \alpha_j^{k+m}\, F\left(H(\cdot)\,e^{\alpha_j G(\cdot)}\right)$$

$$= \sum_{j=1}^M c_j\, \alpha_j^m \left(\sum_{k=0}^M \widetilde{p}_k\, \alpha_j^k\right)\left(H(x_0)\,e^{\alpha_j G(x_0)}\right) = 0 \qquad (8)$$

for all integers $m \geq 0$. Thus we can compute the coefficients $\widetilde{p}_k$, $k = 0, \ldots, M-1$, using the values $F(A^\ell f)$, $\ell = 0, \ldots, 2M - 1$. Having determined the polynomial

$\widetilde{p}(z)$ in (7), we can compute its zeros $\alpha_j$, and afterwards solve a linear equation system to reconstruct the complex coefficients $c_j$ in (4).

However, the question remains, how to obtain the needed data $F(A^\ell f)$, $\ell = 0, \ldots, 2M - 1$. We obtain

$$F(A^0 f) = f(x_0), \tag{9}$$

$$F(A^1 f) = g(x_0) f'(x_0) + \eta(x_0) f(x_0),$$

$$F(A^2 f) = g(x_0)^2 f''(x_0) + [g(x_0)g'(x_0) + 2g(x_0)\eta(x_0)] f'(x_0)$$

$$+ [g(x_0)\eta'(x_0) + \eta(x_0)^2] f(x_0).$$

Since $g$ and $\eta$ (and their derivatives) are known beforehand, it is sufficient to provide the first $2M$ derivative values of $f$ at one point $x_0 \in [a, b]$ in order to reconstruct $f$. Therefore we can conclude.

**Theorem 1** *Let $G, H \in C^\infty([a, b])$, such that $G'$ and $H$ have no zeros on $[a, b]$, and let $x_0 \in [a, b]$ be fixed. Then $f$ in (4) can be viewed as an expansion into eigenfunctions of the differential operator $A$ as in (6), and can be uniquely reconstructed from the derivative samples $f^{(\ell)}(x_0)$, $\ell = 0, \ldots, 2M - 1$.*

**Proof** As seen from the above computations, the operator $A$ of the form (6) indeed possesses the eigenfunctions $H(x) e^{\alpha_j G(x)}$. In order to reconstruct the parameters $\alpha_j$, we first have to compute the required values $F(A^\ell f) = (A^\ell f)(x_0)$, $\ell = 0, \ldots, 2M - 1$. For this purpose, we need to determine the lower triangular matrix $\mathbf{L} = (\lambda_{m,\ell})_{m,\ell=0}^{2M-1} \in \mathbb{R}^{2M \times 2M}$ such that

$$\left( F(A^\ell f) \right)_{\ell=0}^{2M-1} = \left( (A^\ell f)(x_0) \right)_{\ell=0}^{2M-1} = \mathbf{L} \left( f^{(\ell)}(x_0) \right)_{\ell=0}^{2M-1}.$$

As seen in (9), we have already $\lambda_{0,0} := 1$, $\lambda_{1,0} := g(x_0)$, $\lambda_{1,1} := \eta(x_0)$. Generally, to obtain the entries of $\mathbf{L}$, we have to consider the elements $\lambda_{m,\ell}$ as functions in $x$, starting with $\lambda_{0,0}(x) \equiv 1$. By induction, it follows from

$$A^\ell f(x) = \sum_{r=0}^{\ell} \lambda_{\ell,r}(x) f^{(r)}(x)$$

that

$$A^{\ell+1} f(x) = \sum_{r=0}^{\ell} g(x) \left( \lambda'_{\ell,r}(x) f^{(r)}(x) + \lambda_{\ell,r}(x) f^{(r+1)}(x) \right) + \eta(x) \lambda_{\ell,r}(x) f^{(r)}(x)$$

$$= \sum_{r=0}^{\ell} \left( g(x) \lambda'_{\ell,r}(x) + \eta(x) \lambda_{\ell,r}(x) \right) f^{(r)}(x) + g(x) \lambda_{\ell,r}(x) f^{(r+1)}(x).$$

We conclude the recursion

$$
\lambda_{\ell+1,r}(x) := \begin{cases}
g(x)\,\lambda'_{\ell,r}(x) + \eta(x)\,\lambda_{\ell,r}(x) & r = 0, \\
g(x)\,(\lambda'_{\ell,r}(x) + \lambda_{\ell,r-1}(x)) + \eta(x)\,\lambda_{\ell,r}(x) & r = 1, \dots, \ell, \\
g(x)\,\lambda_{\ell,r-1}(x) & r = \ell + 1.
\end{cases}
$$

The matrix entries $\lambda_{\ell,k} := \lambda_{\ell,k}(x_0)$ are well-defined by assumption on $H$ and $G$. In a second step, we solve the homogeneous equation system (8),

$$
\sum_{k=0}^{M} \widetilde{p}_k\, F(A^{k+m} f) = 0, \qquad m = 0, \dots, M-1.
$$

Then we can determine the characteristic polynomial $\widetilde{p}$ in (7) and extract its zeros $\alpha_j$. Finally, the coefficients $c_j$ can be computed from the linear system

$$
F(A^\ell f) = (A^\ell f)(x_0) = \sum_{j=1}^{M} c_j\,(A^\ell(H(\cdot)\,\mathrm{e}^{\alpha_j G(\cdot)}))(x_0) = H(x_0) \sum_{j=1}^{M} c_j \alpha_j^\ell \mathrm{e}^{\alpha_j G(x_0)}
$$

for $\ell = 0, \dots, 2M-1$.                                                                                      □

However, the values $f^{(r)}(x_0)$, $r = 0, \dots, 2M-1$, may not be easily accessible, and we need some extra effort to compute $F(A^\ell f)$ from the derivatives of $f$.

## 3.2 Expansion into Eigenfunctions of a Generalized Shift Operator

Our goal is to find a different set of sample values for the recovery of $f$ in (4), which is easier to obtain but also sufficient for a unique reconstruction. Thus we need to find an operator $B$ which has the same eigenfunctions as $A$ in (6). In addition, we require that $F(B^\ell f)$ (with some point evaluation functions $F$) can be easily obtained from function values of $f$. Similarly as in Sect. 2, we consider the linear operator $B = \varphi(A) = \exp(hA)$ with $A$ in (6) and $h \in \mathbb{R} \setminus \{0\}$. We observe for $f$ in (4),

$$
\exp(hA)f = \sum_{\ell=0}^{\infty} \frac{h^\ell}{\ell!}\,A^\ell f = \sum_{\ell=0}^{\infty} \frac{h^\ell}{\ell!} \sum_{j=1}^{M} c_j\, A^\ell\left(H(\cdot)\mathrm{e}^{\alpha_j G(\cdot)}\right)
$$

$$
= \sum_{\ell=0}^{\infty} \frac{h^\ell}{\ell!} \sum_{j=1}^{M} c_j \alpha_j^\ell \left(H(\cdot)\,\mathrm{e}^{\alpha_j G(\cdot)}\right) = \sum_{j=1}^{M} c_j \left(\sum_{\ell=0}^{\infty} \frac{h^\ell}{\ell!}\alpha_j^\ell\right)\left(H(\cdot)\,\mathrm{e}^{\alpha_j G(\cdot)}\right)
$$

$$= \sum_{j=1}^{M} c_j \, e^{\alpha_j h} \left( H(\cdot) \, e^{\alpha_j G(\cdot)} \right) = H(\cdot) \sum_{j=1}^{M} c_j \, e^{\alpha_j (h + G(\cdot))}$$

$$= H(\cdot) \sum_{j=1}^{M} c_j \, e^{\alpha_j G(G^{-1}(h + G(\cdot)))}$$

$$= \frac{H(\cdot)}{H(G^{-1}(h + G(\cdot)))} \sum_{j=1}^{M} c_j H(G^{-1}(h + G(\cdot))) \, e^{\alpha_j G(G^{-1}(h + G(\cdot)))}$$

$$= \frac{H(\cdot)}{H(G^{-1}(h + G(\cdot)))} \, f \left( G^{-1}(h + G(\cdot)) \right).$$

Therefore, we define the generalized shift operator

$$S_{H,G,h} f(x) := \frac{H(x)}{H(G^{-1}(h + G(x)))} f \left( G^{-1}(h + G(x)) \right), \qquad (10)$$

which depends on the functions $H$, $G$, and the step size $h \in \mathbb{R} \setminus \{0\}$. This shift operator has also been introduced in [25] and satisfies the properties

$$S_{H,G,h_2} \left( S_{H,G,h_1} f \right) = S_{H,G,h_1} \left( S_{H,G,h_2} f \right) = S_{H,G,h_1 + h_2} f$$

for all $h_1, h_2 \in \mathbb{R}$, and

$$S_{H,G,h}^{k} f = S_{H,G,kh} f \qquad (11)$$

for $k \in \mathbb{Z}$, see Theorem 2.1 in [25]. Observe that the generalized shift operator in (10) is already well defined for continuous functions $H$, $G$, and we don't need to assume that $G$ and $H$ are in $C^{\infty}(\mathbb{R})$. We only need to ensure that $G^{-1}$ and $1/H$ are well defined within the considered sampling interval. We summarize this in the following theorem.

**Theorem 2** *Let $G$, $H$ be continuous functions on an interval $[a, b]$, such that $G$ is strictly monotone in $[a, b]$ and $H$ has no zeros in $[a, b]$. Assume that the pairwise different parameters $\alpha_j$ in the expansion*

$$f(x) = \sum_{j=1}^{M} c_j \, H(x) \, e^{\alpha_j G(x)}, \qquad x \in [a, b] \subset \mathbb{R}, \qquad (12)$$

*satisfy* $\operatorname{Im} \alpha_j \in (-T, T]$ *and that* $c_j \in \mathbb{C} \setminus \{0\}$. *Then $f$ can be uniquely reconstructed from the sample values* $f(G^{-1}(h\ell + G(x_0) + h\ell))$, $\ell = 0, \ldots, 2M - 1$, *where $x_0$, $h$ are taken such that* $0 < |h| < \frac{\pi}{T}$ *and* $G(x_0 + h\ell) \in [G(a), G(b)]$ *for* $G(a) < G(b)$ *or* $G(x_0) + h\ell \in [G(b), G(a)]$ *for* $G(a) > G(b)$.

***Proof*** From the arguments above, we can conclude that $H(x)\,\mathrm{e}^{\alpha_j G(x)}$ is an eigenfunction of the generalized shift operator $S_{H,G,h}$ in (10) associated with the eigenvalue $\mathrm{e}^{\alpha_j h}$, since

$$
S_{H,G,h}(H(\cdot)\mathrm{e}^{\alpha_j G(\cdot)}) = \frac{H(\cdot)}{H(G^{-1}(h + G(\cdot)))}\left(H(G^{-1}(h + G(\cdot)))\mathrm{e}^{\alpha_j G(G^{-1}(h+G(\cdot)))}\right)
$$

$$
= H(\cdot)\,\mathrm{e}^{\alpha_j (h+G(\cdot))} = \mathrm{e}^{\alpha_j h}\,H(\cdot)\,\mathrm{e}^{\alpha_j G(\cdot)}.
$$

Further, for $\operatorname{Im}\alpha_j \in (-T, T]$, and $0 < |h| < \frac{\pi}{T}$, the eigenvalues $\mathrm{e}^{\alpha_j h}$ corresponding to active eigenfunctions in (4) are pairwise different, such that we can uniquely derive the active eigenfunctions $H(x)\mathrm{e}^{\alpha_j G(x)}$ in (12) from the corresponding active eigenvalues. We define the Prony polynomial

$$
p(z) := \prod_{j=1}^{M}(z - \mathrm{e}^{\alpha_j h}) = \sum_{k=0}^{M} p_k\, z^k \quad \text{with} \quad p_M = 1, \tag{13}
$$

using the (unknown) eigenvalues $\mathrm{e}^{\alpha_j h}$, where $p_k$, $k = 0, \ldots, M - 1$, are the (unknown) coefficients of the monomial representation of $p(z)$. Then, we conclude

$$
\sum_{k=0}^{M} p_k\, (S_{H,G,h}^{k+m} f)(x_0) = \sum_{k=0}^{M} p_k \sum_{j=1}^{M} c_j\, (S_{H,G,h}^{k+m} H(\cdot)\,\mathrm{e}^{\alpha_j G(\cdot)})(x_0)
$$

$$
= \sum_{k=0}^{M} p_k \sum_{j=1}^{M} c_j\, \mathrm{e}^{\alpha_j h(k+m)} H(x_0)\, \mathrm{e}^{\alpha_j G(x_0)}
$$

$$
= H(x_0) \sum_{j=1}^{M} c_j\, \mathrm{e}^{\alpha_j hm}\, \mathrm{e}^{\alpha_j G(x_0)} \sum_{k=0}^{M} p_k\, (\mathrm{e}^{\alpha_j h})^k
$$

$$
= H(x_0) \sum_{j=1}^{M} c_j\, \mathrm{e}^{\alpha_j hm}\, \mathrm{e}^{\alpha_j G(x_0)}\, p(\mathrm{e}^{\alpha_j h}) = 0 \tag{14}
$$

for all integers $m$, where by definition

$$
(S_{H,G,h}^{k+m} f)(x_0) = \frac{H(x_0)}{H(G^{-1}(h(k + m) + G(x_0)))}\, f(G^{-1}(h(k + m) + G(x_0))).
$$

Thus, we can compute the coefficients $p_k, k = 0, \ldots, M-1$, from the homogeneous linear system

$$\sum_{k=0}^{M} p_k \, (S_{H,G,h}^{k+m} f)(x_0) = H(x_0) \sum_{k=0}^{M} p_k \, \frac{f(G^{-1}(h(k+m) + G(x_0)))}{H(G^{-1}(h(k+m) + G(x_0)))} = 0,$$

$$(15)$$

for $m = 0, \ldots, M - 1$, and $p_M = 1$, or equivalently from

$$\sum_{k=0}^{M-1} p_k \, \frac{f(G^{-1}(h(k+m) + G(x_0)))}{H(G^{-1}(h(k+m) + G(x_0)))} = -\frac{f(G^{-1}(h(M+m) + G(x_0)))}{H(G^{-1}(h(M+m) + G(x_0)))},$$

$$(16)$$

for $m = 0, \ldots, M - 1$. The conditions on $h$ and $x_0$ in the theorem ensure that we only use samples of $f$ in $[a, b]$. The equation system (16) is always uniquely solvable, since the coefficient matrix is invertible. This can be deduced as follows. For $f$ in (12),

$$\left( \frac{f(G^{-1}(h(k+m) + G(x_0)))}{H(G^{-1}(h(k+m) + G(x_0)))} \right)_{m,k=0}^{M-1} = \left( \sum_{j=1}^{M} c_j \, e^{\alpha_j (h(k+m) + G(x_0))} \right)_{m,k=0}^{M-1}$$

$$= \left( e^{\alpha_j h m} \right)_{m=0, j=1}^{M-1, M} \, \text{diag} \left( c_1 e^{\alpha_1 G(x_0)}, \ldots, c_M e^{\alpha_M G(x_0)} \right) \left( e^{\alpha_j h k} \right)_{j=1, m=0}^{M, M-1}. \quad (17)$$

The first and the last matrix factor are invertible Vandermonde matrices with pairwise different nodes $e^{\alpha_j h}$, and the diagonal matrix is invertible, since $c_j \neq 0$.

Having solved (16), we can reconstruct $p(z)$ and extract all its zeros $z_j = e^{\alpha_j h}$. In a second step we can compute the coefficients $c_j$ from the overdetermined system

$$f(G^{-1}(h\ell + G(x_0))) = \sum_{j=1}^{M} c_j \, H(G^{-1}(h\ell + G(x_0))) \, e^{\alpha_j (h\ell + G(x_0))}, \quad (18)$$

for $\ell = 0, \ldots, 2M - 1$.                                                          $\square$

## 3.3   Application to Special Expansions

The model (4) covers many special expansions, and we want to illustrate some of them.

### 3.3.1   Classical Exponential Sums

Obviously, the model (1) is a special case of (4) with $G(x) := x$ and $H(x) := 1$. In this case, we have

$$g(x) \equiv 1, \quad \eta(x) \equiv 0$$

in (5) such that $A$ in (6) reduces to $Af = f'$. The generalized shift operator in (10) with $G^{-1}(x) = x$ is of the form $S_{1,x,h} f(x) = f(h + x)$ and is therefore just the usual shift operator $S_h$ in Sect. 2. By Theorem 1, the sample values $f^{(\ell)}(x_0)$, $\ell = 0, \ldots, 2M - 1$ are sufficient for recovery of $f$, where in this case the interval $[a, b]$ can be chosen arbitrarily in $\mathbb{R}$ and thus also $x_0$. Theorem 2 provides the set of sample values $f(x_0 + h\ell)$ similarly as we had seen already in Sect. 2.

### 3.3.2 Expansions into Shifted Gaussians

We want to reconstruct expansions of the form

$$f(x) = \sum_{j=1}^{M} c_j \, e^{-\beta(x-\alpha_j)^2}, \tag{19}$$

where $\beta \in \mathbb{R} \setminus \{0\}$ is known beforehand, and we need to find $c_j \in \mathbb{C} \setminus \{0\}$ and pairwise different $\alpha_j \in \mathbb{C}$, see also [25, 37].

First, we observe that the functions

$$e^{-\beta(x-\alpha_j)^2} = e^{-\beta\alpha_j^2} \, e^{-\beta x^2} \, e^{2\beta\alpha_j x},$$

are of the form $H(x) \, e^{\alpha_j G(x)}$, with

$$H(x) := e^{-\beta\alpha_j^2} e^{-\beta x^2}, \qquad G(x) := 2\beta x.$$

Using the results in Sects. 3.1 and 3.2, (5) yields

$$g(x) = \frac{1}{G'(x)} = \frac{1}{2\beta}, \qquad \eta(x) = -g(x)\frac{H'(x)}{H(x)} = -\frac{1}{2\beta}(-2\beta x) = x.$$

Therefore, the operator $A$ defined in (6) simplifies to $Af(x) := \frac{1}{2\beta} f'(x) + x \, f(x)$ and

$$A\left(e^{-\beta(\cdot-\alpha_j)^2}\right)(x) = \left(\frac{1}{2\beta}(-2\beta(x - \alpha_j)) + x\right) e^{-\beta(x-\alpha_j)^2} = \alpha_j \, e^{-\beta(x-\alpha_j)^2}.$$

Thus, we can reconstruct $f$ in (19) according to Theorem 1 from the derivative samples $f^{(\ell)}(x_0)$, $\ell = 0, \ldots, 2M - 1$. Here, $x_0$ can be chosen arbitrarily in $\mathbb{R}$, since $G'(x) = 2\beta \neq 0$ and $H(x) \neq 0$ for all $x \in \mathbb{R}$, which means that the interval $[a, b]$ can be chosen arbitrarily in Theorem 1.

Another sampling set is obtained by Theorem 2. The generalized shift operator $S_{H,G,h}$ in (10) reduces to

$$S_{H,G,h} f(x) = \frac{e^{-\beta x^2}}{e^{-\beta((h+2\beta x)/2\beta)^2}} \, f\left(\frac{h+2\beta x}{2\beta}\right) = e^{h(x+h/4\beta)} \, f\left(x + \frac{h}{2\beta}\right). \tag{20}$$

Then

$$S_{H,G,h}(e^{-\beta(\cdot - \alpha_j)^2})(x) = e^{h(x+h/4\beta)} \, e^{-\beta(x + \frac{h}{2\beta} - \alpha_j)^2}$$
$$= e^{h\alpha_j} \, e^{-\beta(x-\alpha_j)^2}.$$

Therefore, the expansion in (19) is an expansion into eigenfunctions of the generalized shift operator in (20) and can be reconstructed from the equidistant samples

$$f\left(x_0 + \frac{h\ell}{2\beta}\right), \qquad \ell = 0, \dots, 2M - 1,$$

where $x_0 \in \mathbb{R}$ can be chosen arbitrarily and $0 < |h| < \frac{\pi}{T}$, where $T$ is the a priori known bound satisfying $|\alpha_j| < T$ for all $j = 1, \dots, M$. Since the interval $[a, b]$ occurring in Theorem 2 can be taken arbitrarily large, we can always take it such that

$$\frac{|G(b) - G(a)|}{2M} = \frac{2|\beta|(b-a)}{2M} > \frac{\pi}{T},$$

and therefore, there is no further condition on the choice of $h$. We note that the procedure also applies for $\beta \in \mathbb{C} \setminus \{0\}$. In this case we can use the substitution $\widetilde{\alpha}_j = \alpha_j 2\beta$ and take $G(x) = x$.

*Remark 1* In particular, the model (19) includes expansions into modulated shifted Gaussians

$$f(x) = \sum_{j=1}^{M} c_j \, e^{2\pi i x \kappa_j} \, e^{-\beta(x-s_j)^2}$$

with $\kappa_j \in [0, 1)$ and $s_j \in \mathbb{R}$ which have been considered in [25]. Since

$$e^{2\pi i x \kappa_j} \, e^{-\beta(x-s_j)^2} = e^{-\beta s_j^2} \, e^{-\beta x^2} \, e^{-x(2\beta s_j + 2\pi i \kappa_j)},$$

we choose $\alpha_j := 2\beta s_j + 2\pi i \kappa_j$, $j = 1, \dots, M$. Then the reconstruction of the $\alpha_j$ is sufficient to find the parameters $s_j$ and $\kappa_j$ from the real and the imaginary part of $\alpha_j$, respectively.

*Example 1* We illustrate the recovery of expansions into shifted Gaussians and consider $f$ of the form (19) with $M = 10$ and $\beta = i$. The original parameters in Table 1 have been obtained by applying a uniform random choice from the intervals

**Table 1** Parameters $c_j$ and $\alpha_j$ for $f(x)$ in (19) with $M = 10$, see Fig. 1

|            | $j = 1$ | $j = 2$ | $j = 3$ | $j = 4$ | $j = 5$ | $j = 6$ | $j = 7$ | $j = 8$ | $j = 9$ | $j = 10$ |
|------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|----------|
| Re $c_j$   | −1.754  | −1.193  | 0.174   | −1.617  | 2.066   | −1.831  | −1.644  | −1.976  | −1.634  | −0.386   |
| Im $c_j$   | −0.756  | 1.694   | −0.279  | −1.261  | 1.620   | 1.919   | −0.245  | −1.556  | −0.968  | −0.365   |
| $\alpha_j$ | 0.380   | −0.951  | 0.411   | 0.845   | −1.113  | −1.530  | −0.813  | −0.725  | −0.303  | −0.031   |
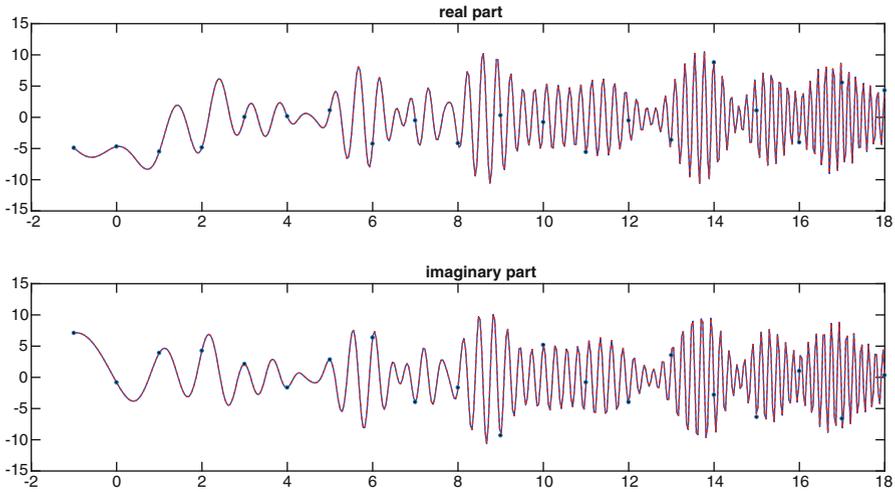


**Fig. 1** Real and imaginary part of the signal $f(x)$ consisting of shifted Gaussians as given in Example 1. The black dots indicate the used signal values. Here the reconstructed signal is shown in red and cannot be distinguished from the original signal $f(x)$

$(-3, 3) + i(-2, 2)$ for $c_j$ and from $(-2, 2)$ for $\alpha_j$. Since $\beta$ is complex, we use $G(x) = x$ and the substitution $\widetilde{\alpha}_j = 2i\alpha_j$. Further, we choose $x_0 = -1$ and $h = 1$.

Figure 1 represents the outcome of such reconstruction. The numerical treatment of the generalized Prony method is studied in more detail in Sect. 4. For the computation of this example we have used Algorithm 1 (see Sect. 4.1) with the minimal number of 20 samples $f(k)$, $k = -1, \ldots, 18$. The samples are represented as black dots in Fig. 1. The obtained maximal reconstruction error for the parameters $\alpha_j$ parameters $c_j$ are

$$err_\alpha = 1.518622755454592 \cdot 10^{-11}, \qquad err_c = 5.286537816367291 \cdot 10^{-10}.$$

### 3.3.3 Expansions into Functions of the Form $\exp(\alpha_j \sin x)$

We want to reconstruct expansions of the form

$$f(x) = \sum_{j=1}^{M} c_j \, e^{\alpha_j \sin x}, \tag{21}$$

where we need to find $c_j \in \mathbb{C} \setminus \{0\}$ and pairwise different $\alpha_j \in \mathbb{C}$. Here, $e^{\alpha_j \sin x}$ is of the form $H(x) \, e^{\alpha_j G(x)}$ with $H(x) := 1$ and $G(x) := \sin(x)$. To ensure that $G(x)$ is strictly monotone, we choose the interval $[-\frac{\pi}{2} + \delta, \ \frac{\pi}{2} - \delta]$ with some small $\delta > 0$. With $g(x) = (G'(x))^{-1} = (\cos(x))^{-1}$ and $\eta(x) = 0$ the operator $A$ defined in (6) simplifies to $Af(x) = (\cos(x))^{-1} f'(x)$ and

$$A(e^{\alpha_j \sin(\cdot)})(x) = \frac{1}{\cos(x)} (\alpha_j \cos(x) \, e^{\alpha_j \sin(x)}) = \alpha_j \, e^{\alpha_j \sin(x)}.$$

According to Theorem 1 we can therefore reconstruct $f$ in (21) from the derivative samples $f^{(\ell)}(x_0)$ for some $x_0 \in [-\frac{\pi}{2} + \delta, \frac{\pi}{2} - \delta]$.

Using Theorem 2, we define with $H(x) := 1$ and $G(x) := \sin(x)$ the generalized shift operator

$$S_{H,G,h} f(x) = f(G^{-1}(h + G(x)) = f(\arcsin(h + \sin(x))).$$

We have to choose $x_0$ and $h$ such that all samples $f(\arcsin(h\ell + \sin(x_0)))$ that we require for the reconstruction are well-defined, i.e., $\sin(x_0) + h\ell \in [-\frac{\pi}{2} + \delta, \frac{\pi}{2} - \delta]$ for $\ell = 0, \ldots, 2M - 1$. This is for example ensured for $x_0 = -\frac{\pi}{2} + \frac{h}{2}$ and $0 < h \le \frac{\pi}{2M+1}$.

*Example 2* We illustrate the reconstruction of a function $f(x)$ of the form (21) with $M = 10$ and with real parameters $c_j$ and $\alpha_j$ in Table 2 that have been obtained by applying a uniform random choice from the intervals $(-3, 3)$ for $c_j$ and from $(-\pi, \pi)$ for $\alpha_j$. We choose a sampling distance $h = \frac{1}{17}$ and a starting point $x_0 = -\frac{\pi}{2} + \frac{h}{2} = -\frac{\pi}{2} + \frac{1}{34}$. The reconstruction is performed using Algorithm 1 in Sect. 4.1.

The reconstruction problem is very ill-posed in this setting, since the measurements all have to be taken from a small interval, see Fig. 2. The possible sampling distance strongly depends on the length of the interval, where $G(x)$ is strictly monotone, as well as on the slope of $G^{-1}(x)$. Therefore, we cannot reconstruct the exact parameters with high precision, however, the reconstructed function is still a very good approximation of $f$, see Fig. 2.

**Table 2** Parameters $c_j$ and $\alpha_j$ for $f(x)$ in (21) with $M = 10$, see Fig. 2

|  | $j = 1$ | $j = 2$ | $j = 3$ | $j = 4$ | $j = 5$ | $j = 6$ | $j = 7$ | $j = 8$ | $j = 9$ | $j = 10$ |
|---|---|---|---|---|---|---|---|---|---|---|
| $c_j$ | 2.104 | 0.363 | 2.578 | 1.180 | 0.497 | 1.892 | 2.274 | 2.933 | −2.997 | 2.192 |
| $\alpha_j$ | 1.499 | 0.540 | −1.591 | 1.046 | −2.619 | 0.791 | 1.011 | 1.444 | 2.455 | 3.030 |
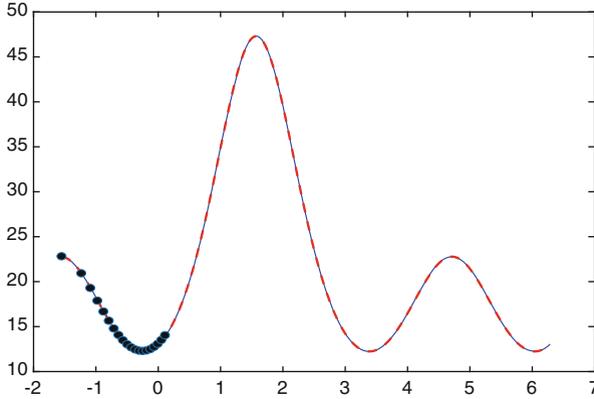
**Fig. 2** Signal $f(x)$ in (21) consisting of $M = 10$ terms according to Table 2. The black dots indicate the used signal values and the reconstructed signal is shown in red

# 4 Numerical Treatment of the Generalized Prony Method

In this section, we consider some numerical procedures to recover the parameters $\alpha_j$, $c_j$, $j = 1, \ldots, M$, in (4) resp. (12).

## 4.1 The Simple Prony Algorithm

First we summarize the direct algorithm for the recovery of $f$ in (12) from the function values $f(G^{-1}(h\ell + G(x_0)))$, $\ell = 0, \ldots, 2M - 1$, according to the proof of Theorem 2.

**Algorithm 1**
**Input:** $M \in \mathbb{N}$, $h > 0$, sampled values $f(G^{-1}(h\ell + G(x_0)))$, $\ell = 0, \ldots, 2M - 1$.

1. Solve the linear system (16) to find the vector $\mathbf{p} = (p_0, \ldots, p_{M-1})^T$.
2. Compute all zeros $z_j \in \mathbb{C}$, $j = 1, \ldots, M$, of $p(z) = \sum\limits_{k=0}^{M-1} p_k z^k + z^M$.
3. Extract the coefficients $\alpha_j := \frac{1}{h} \log z_j$ from $z_j = e^{\alpha_j h}$, $j = 1, \ldots, M$.
4. Solve the system (18) to compute $c_1, \ldots, c_M \in \mathbb{C}$.

**Output:** $\alpha_j \in \mathbb{R} + i[\frac{-\pi}{h}, \frac{\pi}{h})$, $c_j \in \mathbb{C}$, $j = 1, \ldots, M$.

The assumptions of Theorem 2 imply that the coefficient matrix of the linear system (16) is the invertible Hankel matrix,

$$\mathbf{H}_M := \left( \frac{f(G^{-1}(h(k+m)+G(x_0)))}{H(G^{-1}(h(k+m)+G(x_0)))} \right)_{k,m=0}^{M-1}.$$

However, the factorization (17) indicates that $\mathbf{H}_M$ may have very high condition number that particularly depends on the condition number of the Vandermonde matrix $\left( e^{\alpha_j hm} \right)_{m=0,\,j=1}^{M-1,M}$.

## 4.2 ESPRIT for the Generalized Prony Method

We are interested in a more stable implementation of the recovery method and present a modification of the ESPRIT method, see [24, 28, 29, 31] for the classical exponential sum. We assume that the number of terms $M$ in (4) is not given beforehand, but $L$ is a known upper bound of $M$. In the following, we use the notation $\mathbf{A}_{K,N}$ for a rectangular matrix in $\mathbb{C}^{K \times N}$ and $\mathbf{A}_K$ for a square matrix in $\mathbb{C}^{K \times K}$, i.e., the subscripts indicate the matrix dimension.

Let

$$f_\ell := \frac{f(G^{-1}(h\ell + G(x_0)))}{H(G^{-1}(h\ell + G(x_0)))}, \qquad \ell = 0, \dots, 2N-1, \tag{22}$$

be given and well defined, where $N \geq L \geq M$. We consider first the rectangular Hankel matrix

$$\mathbf{H}_{2N-L,L+1} := (f_{\ell+m})_{\ell,m=0}^{2N-L-1,L} \in \mathbb{C}^{(2N-L)\times(L+1)}.$$

For exact data, (14) implies that rank $\mathbf{H}_{2N-L,L+1} = M$. We therefore compute the singular value decomposition of $\mathbf{H}_{2N-L,L+1}$,

$$\mathbf{H}_{2N-L,L+1} = \mathbf{U}_{2N-L}\,\mathbf{D}_{2N-L,L+1}\,\mathbf{W}_{L+1}, \tag{23}$$

with unitary square matrices $\mathbf{U}_{2N-L}$, $\mathbf{W}_{L+1}$ and a rectangular diagonal matrix $\mathbf{D}_{2N-L,L+1}$ containing the singular values of $\mathbf{H}_{2N-L,L+1}$. We determine the numerical rank $M$ of $\mathbf{H}_{2N-L,L+1}$ by inspecting its singular values $\widetilde{\sigma}_1 \geq \widetilde{\sigma}_2 \geq \dots \geq \widetilde{\sigma}_{L+1} \geq 0$. We find $M$ as the number of singular values being larger than a predefined bound $\epsilon$. Usually, we can find a clear gap between $\widetilde{\sigma}_M$ and the further singular values $\widetilde{\sigma}_{M+1}, \dots, \widetilde{\sigma}_{L+1}$, which are close to zero. We redefine the Hankel matrix and consider $\mathbf{H}_{2N-M,M+1} := (f_{\ell+m})_{\ell,m=0}^{2N-M-1,M} \in \mathbb{C}^{(2N-M)\times(M+1)}$ with the corresponding SVD

$$\mathbf{H}_{2N-M,M+1} = \mathbf{U}_{2N-M}\,\mathbf{D}_{2N-M,M+1}\,\mathbf{W}_{M+1}, \tag{24}$$

with unitary matrices $\mathbf{U}_{2N-M}$ and $\mathbf{W}_{M+1}$. For exact data, $\mathbf{H}_{2N-M,M+1}$ has rank $M$, and $\mathbf{D}_{2N-M,M+1}^T = (\mathrm{diag}(\sigma_1, \ldots, \sigma_M, 0), \mathbf{0}) \in \mathbb{R}^{(M+1) \times (2N-M)}$ with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_M > 0$.

We introduce the sub-matrices $\mathbf{H}_{2N-M,M}(0)$ and $\mathbf{H}_{2N-M,M}(1)$ given by

$$\mathbf{H}_{2N-M,M+1} = \left( \mathbf{H}_{2N-M,M}(0), (f_{\ell+M})_{\ell=0}^{2N-M-1} \right) = \left( (f_\ell)_{\ell=0}^{2N-M-1}, \mathbf{H}_{2N-M,M}(1) \right),$$

i.e., we obtain $\mathbf{H}_{2N-M,M}(0)$ be removing the last column of $\mathbf{H}_{2N-M,M+1}$ and $\mathbf{H}_{2N-M,M}(1)$ by removing the first column of $\mathbf{H}_{2N-M,M+1}$. Recalling (16) we have for exact data

$$\mathbf{H}_{2N-M,M}(0) \, \mathbf{p} = - (f_{\ell+M})_{\ell=0}^{2N-M-1}, \tag{25}$$

where $\mathbf{p} = (p_0, \ldots, p_{M-1})^T$ contains the coefficients of the Prony polynomial in (13). Let now

$$\mathbf{C}_M(\mathbf{p}) := \begin{pmatrix} 0 & 0 & \ldots & 0 & -p_0 \\ 1 & 0 & \ldots & 0 & -p_1 \\ 0 & 1 & \ldots & 0 & -p_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & -p_{M-1} \end{pmatrix} \in \mathbb{C}^{M \times M}$$

be the (unknown) companion matrix of $\mathbf{p}$ having the $M$ zeros of $p(z)$ in (13) as eigenvalues. By (25) it follows that

$$\mathbf{H}_{2N-M,M}(0) \, \mathbf{C}_M(\mathbf{p}) = \mathbf{H}_{2N-M,M}(1). \tag{26}$$

This observation leads to the following algorithm. According to (24) we find the factorizations

$$\mathbf{H}_{2N-M,M}(0) = \mathbf{U}_{2N-M} \, \mathbf{D}_{2N-M,M+1} \, \mathbf{W}_{M+1,M}(0),$$

$$\mathbf{H}_{2N-M,M}(1) = \mathbf{U}_{2N-M} \, \mathbf{D}_{2N-M,M+1} \, \mathbf{W}_{M+1,M}(1),$$

where $\mathbf{W}_{M+1,M}(0)$ is obtained by removing the last column of $\mathbf{W}_{M+1}$ and $\mathbf{W}_{M+1,M}(1)$ by removing its first column. Now, (26) implies

$$\mathbf{D}_{2N-M,M+1} \mathbf{W}_{M+1,M}(0) \, \mathbf{C}_M(\mathbf{p}) = \mathbf{D}_{2N-M,M+1} \mathbf{W}_{M+1,M}(1).$$

Multiplication with the generalized inverse

$$\mathbf{D}_{2N-M,M+1}^\dagger = \left( \mathrm{diag}\left( \frac{1}{\sigma_1}, \ldots \frac{1}{\sigma_M}, 0 \right), \mathbf{0} \right) \in \mathbb{R}^{(M+1) \times (2N-M)},$$

finally yields

$$\mathbf{W}_M(0)\,\mathbf{C}_M(\mathbf{p}) = \mathbf{W}_M(1),$$

where the square matrices $\mathbf{W}_M(0)$ and $\mathbf{W}_M(1)$ are obtained from $\mathbf{W}_{M+1,M}(0)$ and $\mathbf{W}_{M+1,M}(1)$, respectively, by removing the last row. Thus, the eigenvalues of $\mathbf{C}_M(\mathbf{p})$ are equal to the eigenvalues of

$$\mathbf{W}_M(0)^{-1}\,\mathbf{W}_M(1),$$

where $\mathbf{W}_M(0)$ is invertible since $\mathbf{C}_M(\mathbf{p})$ is invertible. (We can assume here that $z_j \neq 0$ since $z_j = e^{\alpha_j}$.) We therefore obtain the following new algorithm.

### Algorithm 2 (ESPRIT for the generalized Prony method)
**Input:** $L, N \in \mathbb{N}$, $L \leq N$, $L$ upper bound for the number $M$ of terms in (12), sample values $f_\ell$, $\ell = 0, \ldots, 2N - 1$ as given in (22), $G(x_0)$.

1. Compute the SVD of the rectangular Hankel matrix $\mathbf{H}_{2N-L,L+1}$ as in (23). Determine the numerical rank $M$ of $\mathbf{H}_{2N-L,L+1}$, and compute the SVD of $\mathbf{H}_{2N-M,M+1} = \mathbf{U}_{2N-M}\,\mathbf{D}_{2N-M,M+1}\,\mathbf{W}_{M+1}$.
2. Build the restricted matrix $\mathbf{W}_M(0)$ by removing the last column and the last row of $\mathbf{W}_{M+1}$ and $\mathbf{W}_M(1)$ by removing the first column and the last row of $\mathbf{W}_{M+1}$. Compute the eigenvalues $z_j$, $j = 0, \ldots, M$, of $\mathbf{W}_M(0)^{-1}\mathbf{W}_M(1)$.
3. Extract the coefficients $\alpha_j := \frac{1}{h}\log z_j$ from $z_j = e^{\alpha_j h}$, $j = 1, \ldots, M$.
4. Solve the overdetermined system

$$f_\ell = \sum_{j=1}^{M} c_j\, z_j^{G(x_0)/h}\, z_j^\ell, \qquad \ell = 0, \ldots, 2N - 1,$$

to compute $c_1, \ldots, c_M \in \mathbb{C}$.

**Output:** $M$, $\alpha_j \in \mathbb{R} + \mathrm{i}[\frac{-\pi}{h}, \frac{\pi}{h})$, $c_j \in \mathbb{C}$, $j = 1, \ldots, M$.

*Example 3* We compare the performance of the classical Prony method in Algorithm 1 with the ESPRIT method in Algorithm 2 and focus on the reconstruction of the frequency parameters for $f$ of the form (21). In our numerical example we choose $M = 5$, $x_0 = -\frac{\pi}{2} + \frac{1}{34}$, $h = \frac{1}{17}$ and the parameter vectors $\alpha = (\alpha_j)_{j=1}^{M}$, $\mathbf{c} = (c_j)_{j=1}^{M}$ as

$$\alpha = (\frac{\pi}{2}, \frac{\mathrm{i}\pi}{4}, 0.4 + \mathrm{i}, -0.5, -1)^T \text{ and } \mathbf{c} = (0.5, 2, -3, 0.4\mathrm{i}, -0.2)^T.$$

For Algorithm 1 we have only used the first $N = 10$ samples. For the ESPRIT Algorithm 2 we have used $N = 15$, i.e., 30 sample values, and have fixed an upper bound $L = 10$. For the rank approximation we have applied a bound $\epsilon = 10^{-8}$. For comparison we also tested Algorithm 2 with an upper bound of $L = 13$. In Table 3,

**Table 3** Reconstructed parameters $\alpha_j$ in Example 3 provided by Algorithms 1 and 2

| $j$ | Exact $\alpha_j$ | $\alpha_j$   (Algorithm 1) | $\alpha_j$   (Algorithm 2, $L = 10$) | $\alpha_j$   (Algorithm 2, $L = 13$) |
|---|---|---|---|---|
| $j = 1$ | $\frac{\pi}{2}$ | $1.57121 + 6.0886 \cdot 10^{-5}$i | $1.57079 - 2.3198 \cdot 10^{-8}$i | $1.57079 - 2.5066$i $\cdot 10^{-8}$ |
| $j = 2$ | $\frac{\mathrm{i}\pi}{4}$ | $0.00231 + 0.7928$i | $2.00492 \cdot 10^{-6} + 0.7854$i | $2.00522 \cdot 10^{-6} + 0.7854$i |
| $j = 3$ | $0.4 + $i | $0.40168 + 0.9982$i | $0.4000 + 1$i | $0.4000 + 1$i |
| $j = 4$ | $-0.5$ | $-0.49944 - 0.0013$i | $-0.5 - 4.3008 \cdot 10^{-7}$i | $-0.5 - 4.5298 \cdot 10^{-7}$i |
| $j = 5$ | $-1$ | $-1.00019 - 0.0042$i | $-1.0 - 1.1763^{-6}$i | $-1.0 - 1.16642 \cdot 10^{-6}$i |

we present the results of parameter reconstruction using Algorithms 1 and 2. The reconstruction of the frequency values using Algorithm 2 is in the case for $L = 10$ as well as in the case $L = 13$ much more accurate than the reconstruction using Algorithm 1. For both upper bounds $L$ the reconstruction error is of the same order. Lemma 3.1 in [30] suggests that a sufficiently large choice of $L \approx N$ is a good choice.

*Remark 2* The Hankel matrices occurring in the considered reconstruction problems can have a very high condition number. However, there are stable algorithms available to compute the SVD for such Hankel matrices, particularly for the square case, see e.g. [11].

## 4.3   Simplification in Case of Partially Known Frequency Parameters

In some applications, one or more of the parameters $\alpha_j$, or equivalently $z_j = e^{\alpha_j h}$ in the expansion (12), may be already known beforehand. However, if the corresponding coefficients $c_j$ are unknown, we cannot just eliminate the term $c_j H(x) e^{\alpha_j G(x)}$ from the sum in (12) to get new measurements of the simplified sum from the original measurements. However, we can use the following approach. Recall that the vector $\mathbf{p} = (p_0, \ldots, p_M)^T$ of coefficients of the Prony polynomial

$$p(z) = \sum_{k=0}^{M} p_k z^k = \prod_{j=1}^{M} (z - z_j)$$

satisfies by (15) and (16)

$$\mathbf{H}_{2N-M,M+1}\,\mathbf{p} = \mathbf{0},$$

where the Hankel matrix $\mathbf{H}_{2N-M,M+1}$ is constructed from $f_\ell$ in (22) as in the previous section. Assume that $z_1$ is already known beforehand, and let

$$q(z) := \prod_{j=2}^{M}(z - z_j) = \sum_{k=0}^{M-1} q_k z^k,$$

with the coefficient vector $\mathbf{q} := (q_0, \ldots, q_{M-1})^T$. Then $p(z) = (z-z_1)q(z)$ implies for the coefficient vectors

$$\mathbf{p} = \begin{pmatrix} 0 \\ q_0 \\ \vdots \\ q_{M-1} \end{pmatrix} - z_1 \begin{pmatrix} q_0 \\ \vdots \\ q_{M-1} \\ 0 \end{pmatrix}$$

and thus

$$\mathbf{H}_{2N-M,M+1}\mathbf{p} = \big(\mathbf{H}_{2N-M,M}(1) - z_1\mathbf{H}_{2N-M,M}(0)\big)\,\mathbf{q} = \mathbf{0},$$

with $\mathbf{H}_{2N-M,M}(0)$ and $\mathbf{H}_{2N-M,M}(1)$ denoting the submatrices of $\mathbf{H}_{2N-M,M+1}$, where either the last column or the first column is removed. Therefore, we easily find the new Hankel matrix

$$\widetilde{\mathbf{H}}_{2N-M,M} = \mathbf{H}_{2N-M,M}(1) - z_1\mathbf{H}_{2N-M,M}(0)$$

for the reduced problem. Observe from (22), that the new components of the matrix $\mathbf{H}_{2N-M,M}(1) - z_1\mathbf{H}_{2N-M,M}(0)$ are of the form

$$\widetilde{f}_\ell = f_{\ell+1} - z_1 f_\ell = \sum_{j=1}^{M} c_j e^{\alpha_j(h(\ell+1)+G(x_0))} - e^{\alpha_1 h}\sum_{j=1}^{M} c_j e^{\alpha_j(h\ell+G(x_0))}$$

$$= \sum_{j=2}^{M} c_j(e^{\alpha_j h} - e^{\alpha_1 h})e^{\alpha_j(h\ell+G(x_0))},$$

i.e., the coefficients $c_j$, $j = 2, \ldots, M$, are changed to $\widetilde{c}_j = c_j(e^{\alpha_j h} - e^{\alpha_1 h})$. Thus, we can use the samples $\widetilde{f}_\ell$ to recover the shorter sum $\sum_{j=2}^{M} \widetilde{c}_j e^{\alpha_j G(x)}$. Once we have computed the remaining $\alpha_j$, $j = 2, \ldots, M$ we obtain the coefficients $c_j$, $j = 1, \ldots, M$, by solving the linear system (18).

# 5 Modified Prony Method for Sparse Approximation

In this section, we want to consider the question, how to approximate a given data vector $\mathbf{y} = (y_k)_{k=0}^N \in \mathbb{C}^{N+1}$ with $N \geq 2M-1$ by a new vector $\mathbf{f} = (f_k)_{k=0}^N \in \mathbb{C}^{N+1}$ whose elements are structured as

$$f_k = \sum_{j=1}^M c_j \, z_j^k,$$

i.e., $\mathbf{f}$ only depends on the parameter vectors $\mathbf{c} = (c_j)_{j=1}^M$ and $\mathbf{z} = (z_j)_{j=1}^M$. In this setting, the length $N$ of the data vector $\mathbf{y}$ is usually much larger than $M$, i.e., $N \gg M$, while $M$ is assumed to be small, say $M < 30$. We assume that for the given data $\mathbf{y}$ the corresponding Hankel matrix $\mathbf{H} := (y_{k+m})_{k=0,m=0}^{N-M-1,M-1}$ has full rank, i.e., that the given data cannot be exactly represented by an exponential sum with less than $M$ terms, as it can be also seen from the factorization (17). Further, we assume that $c_j \in \mathbb{C} \setminus \{0\}$ and that $z_j \in \mathbb{C} \setminus \{0\}$ are pairwise distinct.

## *5.1 The Nonlinear Least-Squares Problem*

We want to solve the minimization problem

$$\operatorname*{argmin}_{\mathbf{c},\mathbf{z}\in\mathbb{C}^M} \left\| \mathbf{y} - \left( \sum_{j=1}^M c_j \, z_j^k \right)_{k=0}^N \right\|_2. \tag{27}$$

This problem occurs in two different scenarios. The first one is the problem of parameter estimation in case of noisy data. Assume that we have noisy samples $y_k = f(k) + \epsilon_k$, $k = 0, \ldots, N$, of $f(x) = \sum_{j=1}^M c_j z_j^x$, where $\epsilon_k$ are i.i.d. random variables with $\epsilon_k \in N(0, \sigma^2)$. In the second scenario we consider the sparse nonlinear approximation problem to find a function $f(x) = \sum_{j=1}^M c_j z_j^x$, which minimizes $\sum_{\ell=0}^N |y_\ell - f(\ell)|^2$. With the Vandermonde matrix

$$\mathbf{V_z} := \begin{pmatrix} 1 & 1 & \ldots & 1 \\ z_1 & z_2 & \ldots & z_M \\ z_1^2 & z_2^2 & \ldots & z_M^2 \\ \vdots & \vdots & & \vdots \\ z_1^N & z_2^N & \ldots & z_M^N \end{pmatrix} \in \mathbb{C}^{(N+1)\times M}$$

we have $\mathbf{f} = \mathbf{V_z}\,\mathbf{c}$, and the problem (27) can be reformulated as

$$\underset{\mathbf{c},\mathbf{z}\in\mathbb{C}^M}{\mathrm{argmin}} \|\mathbf{y} - \mathbf{V_z}\mathbf{c}\|_2.$$

For given $\mathbf{z}$, the linear least squares problem $\underset{\mathbf{c}\in\mathbb{C}^M}{\mathrm{argmin}} \|\mathbf{y} - \mathbf{V_z}\mathbf{c}\|_2$ can be directly solved, and we obtain $\mathbf{c} = \mathbf{V_z^+}\mathbf{y} = [\mathbf{V_z^*}\mathbf{V_z}]^{-1}\mathbf{V_z^*}\mathbf{y}$, since $\mathbf{V_z}$ has full rank $M$. Thus (27) can be simplified to

$$\underset{\mathbf{z}\in\mathbb{C}^M}{\mathrm{argmin}} \|\mathbf{y} - \mathbf{V_z}\mathbf{V_z^+}\mathbf{y}\|_2^2 = \underset{\mathbf{z}\in\mathbb{C}^M}{\mathrm{argmin}} \|(\mathbf{I} - \mathbf{P_z})\mathbf{y}\|_2^2$$

$$= \underset{\mathbf{z}\in\mathbb{C}^M}{\mathrm{argmin}}(\mathbf{y^*}\mathbf{y} - \mathbf{y^*}\mathbf{P_z}\mathbf{y}) = \underset{\mathbf{z}\in\mathbb{C}^M}{\mathrm{argmax}}\, \mathbf{y^*}\mathbf{P_z}\mathbf{y},$$

where $\mathbf{P_z} := \mathbf{V_z}\mathbf{V_z^+}$ is the projection matrix satisfying $\mathbf{P_z} = \mathbf{P_z^*} = \mathbf{P_z^2}$, $\mathbf{P_z}\mathbf{V_z} = \mathbf{V_z}$ as well as $\mathbf{V_z^+}\mathbf{P_z} = \mathbf{V_z^+}$. Hence, similarly as for Prony's method, we can concentrate on finding the parameters $z_j$ in $\mathbf{z}$ first.

Let now $\mathbf{r(z)} := \mathbf{P_z}\mathbf{y} \in \mathbb{C}^{N+1}$. Then the optimization problem is equivalent to

$$\underset{\mathbf{z}\in\mathbb{C}^M}{\mathrm{argmax}} \|\mathbf{r(z)}\|_2^2 = \underset{\mathbf{z}\in\mathbb{C}^M}{\mathrm{argmax}} \|\mathbf{P_z}\mathbf{y}\|_2^2. \qquad (28)$$

To derive an iterative algorithm for solving (28), we first determine the Jacobian $\mathbf{J_z}$ of $\mathbf{r(z)} = (r_\ell(\mathbf{z}))_{\ell=0}^N$.

**Theorem 3** *The Jacobian matrix* $\mathbf{J_z} \in \mathbb{C}^{(N+1)\times M}$ *of* $\mathbf{r(z)}$ *in (28) is given by*

$$\mathbf{J_z} := \left(\frac{\partial r_\ell(\mathbf{z})}{\partial z_j}\right)_{\ell=0,j=1}^{N,M}$$

$$= (\mathbf{I}_{N+1} - \mathbf{P_z})\mathbf{V_z'}\,\mathrm{diag}(\mathbf{V_z^+}\mathbf{y}) + (\mathbf{V_z^+})^* \mathrm{diag}\left((\mathbf{V_z'})^*(\mathbf{I}_{N+1} - \mathbf{P}_z)\mathbf{y}\right), \quad (29)$$

*where* $\mathbf{I}_{N+1}$ *denotes the identity matrix of size* $N + 1$,

$$\mathbf{V_z'} := \begin{pmatrix} 0 & 0 & \dots & 0 \\ 1 & 1 & \dots & 1 \\ 2z_1 & 2z_2 & \dots & 2z_M \\ \vdots & \vdots & & \vdots \\ Nz_1^{N-1} & Nz_2^{N-1} & \dots & Nz_M^{N-1} \end{pmatrix} \in \mathbb{C}^{(N+1)\times M},$$

*and* $\mathrm{diag}(\mathbf{q})$ *denotes the diagonal matrix of size* $M \times M$ *for a vector* $\mathbf{q} \in \mathbb{C}^M$. *In particular,*

$$\nabla\|\mathbf{r(z)}\|_2^2 = 2\mathbf{J_z^*}\,\mathbf{r(z)} = \mathrm{diag}((\mathbf{V_z'})^T(\mathbf{I}_{N+1} - \mathbf{P}_z)\overline{\mathbf{y}})\,\mathbf{V_z^+}\mathbf{y}. \qquad (30)$$

***Proof*** First, observe that $\frac{\partial}{\partial z_j}\mathbf{V_z}$ is a rank-1 matrix of the form

$$\frac{\partial}{\partial z_j}\mathbf{V_z} = \mathbf{z}'_j\, \mathbf{e}^*_j \in \mathbb{C}^{(N+1)\times M}, \qquad j = 1,\dots,M,$$

where $\mathbf{z}'_j = (0, 1, 2z_j, 3z_j^2, \dots, Nz_j^{N-1})^T$ and $\mathbf{e}_j$ is the $j$th unit vector of length $M$. Then we obtain

$$\frac{\partial}{\partial z_j}\mathbf{r}(\mathbf{z}) = \frac{\partial}{\partial z_j}(\mathbf{P_z y}) = \frac{\partial}{\partial z_j}\left(\mathbf{V_z}[\mathbf{V_z^* V_z}]^{-1}\mathbf{V_z^* y}\right)$$

$$= (\mathbf{z}'_j\, \mathbf{e}^*_j)\mathbf{V_z^+ y} - (\mathbf{V_z^+})^*\left[(\mathbf{z}'_j\, \mathbf{e}^*_j)^*\mathbf{V_z} + \mathbf{V_z^*}(\mathbf{z}'_j\, \mathbf{e}^*_j)\right]\mathbf{V_z^+ y} + (\mathbf{V_z^+})^*(\mathbf{z}'_j\, \mathbf{e}^*_j)^*\mathbf{y}$$

$$= (\mathbf{V_z^+ y})_j\, \mathbf{z}'_j - ((\mathbf{z}'_j)^*\mathbf{P_z y})(\mathbf{V_z^+})^*\mathbf{e}_j - (\mathbf{V_z^+ y})_j\mathbf{P_z z}'_j + ((\mathbf{z}'_j)^*\mathbf{y})(\mathbf{V_z^+})^*\mathbf{e}_j$$

$$= (\mathbf{V_z^+ y})_j\,(\mathbf{I}_{N+1} - \mathbf{P_z})\mathbf{z}'_j + ((\mathbf{z}'_j)^*(\mathbf{I}_{N+1} - \mathbf{P_z})\mathbf{y})(\mathbf{V_z^+})^*\mathbf{e}_j$$

$$= (\mathbf{V_z^+ y})_j\,(\mathbf{I}_{N+1} - \mathbf{P_z})\mathbf{V}'_z\mathbf{e}_j + ((\mathbf{z}'_j)^*(\mathbf{I}_{N+1} - \mathbf{P_z})\mathbf{y})(\mathbf{V_z^+})^*\mathbf{e}_j,$$

where $(\mathbf{V_z^+ y})_j$ denotes the $j$th component of $\mathbf{V_z^+ y}$. From this observation, we immediately find $\mathbf{J_z}$ in (29). Further, this formula implies

$$\mathbf{J_z^* r}(\mathbf{z}) = \left(\mathrm{diag}\,\overline{\mathbf{V_z^+ y}}\right)(\mathbf{V}'_z)^*(\mathbf{I}_{N+1} - \mathbf{P_z})\mathbf{P_z y} + \left(\mathrm{diag}((\mathbf{V}'_z)^*(\mathbf{I}_{N+1} - \mathbf{P_z})\mathbf{y})\right)^*\mathbf{V_z^+ P_z y}$$

$$= \mathrm{diag}\left((\mathbf{V}'_z)^T(\mathbf{I}_{N+1} - \mathbf{P_z})\overline{\mathbf{y}}\right)\mathbf{V_z^+ y}.$$

$\square$

**Corollary 1** *Let* $\mathbf{y} \in \mathbb{C}^{N+1}$ *be given and assume that* $(y_{k+m})_{k=0,m=0}^{N-M+1,M-1}$ *has full rank* $M$. *Then, a vector* $\mathbf{z} \in \mathbb{C}^M$ *solving (28) necessarily satisfies*

$$(\mathbf{V}'_z)^*(\mathbf{I}_{N+1} - \mathbf{P_z})\mathbf{y} = \mathbf{0}.$$

*Proof* If $\mathbf{z}$ solves (28), then $\nabla\|\mathbf{r}(\mathbf{z})\|_2^2 = 0$. Now, the assertion follows from (30) using the information that $\mathbf{c} = \mathbf{V_z^+ y}$ has no vanishing components. $\square$

*Remark 3*

1. The necessary condition in Corollary 1 can be used to build an iterative algorithm for updating the vector $\mathbf{z}$ where we start with $\mathbf{z}^{(0)}$ obtained from the ESPRIT Algorithm 2. We then search for $\mathbf{z}^{(j+1)}$ by solving

$$(\mathbf{V}'_{\mathbf{z}^{(j+1)}})^*(\mathbf{I}_{N+1} - \mathbf{P}_{\mathbf{z}^{(j)}})\mathbf{y} = \mathbf{0},$$

   i.e., by computing the zeros of the polynomial with coefficient vector

$$\mathrm{diag}(0, 1, 2, \dots, N)\,(\mathbf{I}_{N+1} - \mathbf{P}_{\mathbf{z}^{(j)}})\mathbf{y}$$

and taking the subset of $M$ zeros which is closest to the previous set $\mathbf{z}^{(j)}$. We will further elaborate on this approach in the future.

2. This approach is different from most ideas to solve (27) in the literature, see e.g. [7, 19, 20] and the recent survey [38]. In that papers, one first transfers the problem of finding $\mathbf{z} \in \mathbb{C}^M$ into the problem of finding the vector $\mathbf{p} = (p_k)_{k=0}^M \in \mathbb{C}^{M+1}$ with $\|\mathbf{p}\|_2 = 1$, such that $p(z_j) = \sum_{k=0}^M p_k z_j^k = 0$ for all $j = 1, \ldots, M$, thereby imitating the idea of Prony's method. Introducing the matrix

$$\mathbf{X}_{\mathbf{p}}^T = \begin{pmatrix} p_0 & p_1 & \cdots & & p_M & & \\ & p_0 & p_1 & \cdots & & p_M & \\ & & \ddots & & & & \ddots \\ & & & p_0 & p_1 & \cdots & & p_M \end{pmatrix} \in \mathbb{C}^{(N-M+1)\times(N+1)}$$

that satisfies $\mathbf{X}_{\mathbf{p}}^T \mathbf{V}_{\mathbf{z}} = \mathbf{0}$, we obtain a projection matrix

$$\overline{\mathbf{P}}_{\mathbf{p}} := \overline{\mathbf{X}}_{\mathbf{p}} \overline{\mathbf{X}}_{\mathbf{p}}^+ = \overline{\mathbf{X}}_{\mathbf{p}} [\mathbf{X}_{\mathbf{p}}^T \overline{\mathbf{X}}_{\mathbf{p}}]^{-1} \mathbf{X}_{\mathbf{p}}^T = (\mathbf{I}_{N+1} - \mathbf{P}_{\mathbf{z}}),$$

and (28) can be rephrased as

$$\underset{\substack{\mathbf{p} \in \mathbb{C}^{M+1} \\ \|\mathbf{p}\|_2 = 1}}{\operatorname{argmin}} \|\overline{\mathbf{P}}_{\mathbf{p}} \mathbf{y}\|_2^2 = \underset{\substack{\mathbf{p} \in \mathbb{C}^{M+1} \\ \|\mathbf{p}\|_2 = 1}}{\operatorname{argmin}} \mathbf{y}^* \overline{\mathbf{X}}_{\mathbf{p}} [\mathbf{X}_{\mathbf{p}}^T \overline{\mathbf{X}}_{\mathbf{p}}]^{-1} \mathbf{X}_{\mathbf{p}}^T \mathbf{y}.$$

## *5.2 Gauß-Newton and Levenberg-Marquardt Iteration*

Another approach than given in Remark 3 to solve the non-linear least squares problem (28) is the following. We approximate $\mathbf{r}(\mathbf{z} + \boldsymbol{\delta})$ using its first order Taylor expansion $\mathbf{r}(\mathbf{z}) + \mathbf{J}_{\mathbf{z}} \boldsymbol{\delta}$. Now, instead of maximizing $\|\mathbf{r}(\mathbf{z} + \boldsymbol{\delta})\|_2^2$ we consider

$$\underset{\boldsymbol{\delta} \in \mathbb{C}^M}{\operatorname{argmax}} \|\mathbf{r}(\mathbf{z}) + \mathbf{J}_{\mathbf{z}} \boldsymbol{\delta}\|_2^2 = \underset{\boldsymbol{\delta} \in \mathbb{C}^M}{\operatorname{argmax}} (\|\mathbf{r}(\mathbf{z})\|_2^2 + (\mathbf{r}(\mathbf{z})^* \mathbf{J}_{\mathbf{z}} \boldsymbol{\delta} + \boldsymbol{\delta}^* \mathbf{J}_{\mathbf{z}}^* \mathbf{r}(\mathbf{z}) + \boldsymbol{\delta}^* \mathbf{J}_{\mathbf{z}}^* \mathbf{J}_{\mathbf{z}} \boldsymbol{\delta})$$

which yields

$$2 \operatorname{Re}(\mathbf{J}_{\mathbf{z}}^* \mathbf{r}(\mathbf{z})) + 2 \mathbf{J}_{\mathbf{z}}^* \mathbf{J}_{\mathbf{z}} \boldsymbol{\delta} = \mathbf{0}.$$

Thus, starting with the vector $\mathbf{z}^{(0)}$ obtained from Algorithm 2, the $j$th step of the Gauß-Newton iteration is of the form

$$(\mathbf{J}_{\mathbf{z}^{(j)}}^* \mathbf{J}_{\mathbf{z}^{(j)}}) \boldsymbol{\delta}^{(j)} = -\operatorname{Re}(\mathbf{J}_{\mathbf{z}^{(j)}}^* \mathbf{r}(\mathbf{z}^{(j)}))$$

to get the improved vector $\mathbf{z}^{(j+1)} = \mathbf{z}^{(j)} + \boldsymbol{\delta}^{(j)}$. Since $(\mathbf{I}_{N+1} - \mathbf{P}_{\mathbf{z}^{(j)}})\mathbf{y}$ may already be close to the zero vector, the matrix $(\mathbf{J}^*_{\mathbf{z}^{(j)}} \mathbf{J}_{\mathbf{z}^{(j)}})$ is usually ill-conditioned. Therefore, we regularize by changing the matrix in each step to $(\mathbf{J}^*_{\mathbf{z}^{(j)}} \mathbf{J}_{\mathbf{z}^{(j)}}) + \lambda_j \mathbf{I}_M$ and obtain the Levenberg-Marquardt iteration

$$((\mathbf{J}^*_{\mathbf{z}^{(j)}} \mathbf{J}_{\mathbf{z}^{(j)}}) + \lambda_j \mathbf{I}_M)\, \boldsymbol{\delta}^{(j)} = -\mathrm{Re}\,(\mathbf{J}^*_{\mathbf{z}^{(j)}} \mathbf{r}(\mathbf{z}^{(j)})).$$

In this algorithm, we need to fix the parameters $\lambda_j$, which are usually taken very small. If we arrive at a (local) maximum, then the right-hand side in the Levenberg-Marquardt iteration vanishes, and we obtain $\boldsymbol{\delta}^{(j)} = \mathbf{0}$.

*Remark 4*

1. The considered non-linear least squares problem is also closely related to structured low-rank approximation, see [18, 36]. Further, instead of the Euclidean norm, one can consider the maximum norm, see [6, 12] or the 1-norm, see [32].
2. Some questions remain. How good is the approximation with exponential sums, if $(y_\ell)_{\ell=0}^N$ is known to be a sampling sequence of a function in a given smoothness space, and what is the convergence rate with respect to the number of terms $M$? The authors are not aware of a complete answer to this question. However, in [6] it has been shown that the function $1/x$ can be approximated by an $M$-term exponential sum with an error $\mathcal{O}(\exp(c\sqrt{M})$. Also the results in [5] and [23] indicate that we can hope for an exponential decay of the approximation error for a larger class of functions.

# References

1. Adamjan, V., Arov, D., Krein, M.: Analytic properties of the Schmidt pairs of a Hankel operator and the generalized Schur-Takagi problem. Math. USSR Sb. **86**, 34–75 (1971)
2. Andersson, F., Carlsson, M., de Hoop, M.: Sparse approximation of functions using sums of exponentials and AAK theory. J. Approx. Theory **163**, 213–248 (2011)
3. Baechler, G., Scholefield, A., Baboulaz, L., Vetterli, M.: Sampling and exact reconstruction of pulses with variable width. IEEE Trans. Signal Process. **65**(10), 2629–2644 (2017)
4. Barone, P.: On the distribution of poles of Padé approximants to the Z-transform of complex Gaussian white noise. J. Approx. Theory **132**(2), 224–240 (2005)
5. Beylkin, G., Monzón, L.: On approximation of functions by exponential sums. Appl. Comput. Harmon. Anal. **19**, 17–48 (2005)
6. Braess, D., Hackbusch, W.: Approximation of $1/x$ by exponential sums in $[1, \infty)$. IMA J. Numer. Anal. **25**, 685–697 (2005)
7. Bresler, Y., Macovski, A.: Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. IEEE Trans. Acoust. Speech Signal Process. **34**(5), 1081–1089 (1986)

8. Chunaev, P., Danchenko, V.: Approximation by amplitude and frequency operators. J. Approx. Theory **207**, 1–31 (2016)
9. Cuyt, A., Tsai, M.N., Verhoye, M., Lee, W.S.: Faint and clustered components in exponential analysis. Appl. Math. Comput. **327**, 93–103 (2018)
10. Dragotti, P., Vetterli, M., Blu, T.: Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets Strang–Fix. IEEE Trans. Signal Process. **55**(5), 1741–1757 (2007)
11. Drmač, Z.: SVD of Hankel matrices in Vandermonde-Cauchy product form. Electron. Trans. Numer. Anal. **44**, 593–623 (2015)
12. Hackbusch, W.: Computation of best $l^\infty$ exponential sums for $1/x$ by Remez' algorithm. Comput. Vis. Sci. **20**(1–2), 1–11 (2019)
13. Hauer, J., Demeure, C., Scharf, L.: Initial results in Prony analysis of power system response signals. IEEE Trans. Power Syst. **5**(1), 80–89 (1990)
14. Hua, Y., Sarkar, T.: On the total least squares linear prediction method for frequency estimation. IEEE Trans. Acoust. Speech Signal Process. **38**(12), 2186–2189 (1990)
15. Lang, M.C.: Least-squares design of IIR filters with prescribed magnitude and phase responses and a pole radius constraint. IEEE Trans. Signal Process. **48**(11), 3109–3121 (2000)
16. Levin, D.: Behavior preserving extension of univariate and bivariate functions. In: Hoggan, P. (ed.) Electronic Structure Methods with Applications to Experimental Chemistry, vol. 68, pp. 19–42. Proceedings of MEST 2012. Academic Press, Chennai (2014)
17. Manolakis, D., Ingle, V., Kogon, S.: Statistical and Adaptive Signal Processing. McGraw-Hill, Boston (2005)
18. Markovsky, I.: Low-Rank Approximation: Algorithms, Implementation, Applications, 2nd edn. Springer, Berlin (2018)
19. Osborne, M., Smyth, G.: A modified Prony algorithm for fitting functions defined by difference equations. SIAM J. Sci. Stat. Comput. **12**, 362–382 (1991)
20. Osborne, M., Smyth, G.: A modified Prony algorithm for exponential function fitting. SIAM J. Sci. Comput. **16**(1), 119–138 (1995)
21. Peter, T., Plonka, G.: A generalized Prony method for reconstruction of sparse sums of eigenfunctions of linear operators. Inverse Prob. **29**(2) (2013)
22. Plonka, G., Pototskaia, V.: Application of the AAK theory for sparse approximation of exponential sums (2016). Preprint. http://arxiv.org/pdf/1609.09603
23. Plonka, G., Pototskaia, V.: Computation of adaptive Fourier series by sparse approximation of exponential sums. J. Fourier Anal. Appl. **25**(4), 1580–1608 (2019)
24. Plonka, G., Tasche, M.: Prony methods for recovery of structured functions. GAMM-Mitteilungen **37**(2), 239–258 (2014)
25. Plonka, G., Stampfer, K., Keller, I.: Reconstruction of stationary and non-stationary signals by the generalized Prony method. Anal. Appl. **17**(2), 179–210 (2019)
26. Poh, K., Marziliano, P.: Compressive sampling of EEG signals with finite rate of innovation. EURASIP J. Adv. Signal Process. **2010**, 183105 (2010)
27. Potts, D., Tasche, M.: Parameter estimation for exponential sums by approximate Prony method. Signal Process. **90**(5), 1631–1642 (2010)
28. Potts, D., Tasche, M.: Parameter estimation for multivariate exponential sums. Electron. Trans. Numer. Anal. (40), 204–224 (2013)
29. Potts, D., Tasche, M.: Parameter estimation for nonincreasing exponential sums by Prony-like methods. Linear Algebra Appl. **439**(4), 1024–1039 (2013)
30. Potts, D., Tasche, M.: Error estimates for the ESPRIT algorithm. In: Large Truncated Toeplitz Matrices, Toeplitz Operators, and Related Topics, pp. 621–648. Birkhäuser, Basel (2017)
31. Roy, R., Kailath, T.: Esprit estimation of signal parameters via rotational invariance techniques. IEEE Trans. Acoust. Speech Signal Process. **37**, 984–995 (1989)
32. Skrzipek, M.R.: Signal recovery by discrete approximation and a Prony-like method. J. Comput. Appl. Math. **326**, 193–203 (2017)
33. Stampfer, K., Plonka, G.: The generalized operator-based Prony method. Constr. Approx. (2020). https://doi.org/10.1007/s00365-020-09501-6

34. Stoica, P., Moses, R.L.: Spectral analysis of signals. Pearson Prentice Hall, Upper Saddle River (2005)
35. Urigen, J., Blu, T., Dragotti, P.: FRI sampling with arbitrary kernels. IEEE Trans. Signal Process. **61**(21), 5310–5323 (2013)
36. Usevich, K., Markovsky, I.: Variable projection for affinely structured low-rank approximation in weighted 2-norms. J. Comput. Appl. Math. **272**, 430–448 (2014)
37. Vetterli, M., Marziliano, P., Blu, T.: Sampling signals with finite rate of innovation. IEEE Trans. Signal Process. **50**(6), 1417–1428 (2002)
38. Zhang, R., Plonka, G.: Optimal approximation with exponential sums by a maximum likelihood modification of Pronys method. Adv. Comput. Math. **45**(3), 1657–1687 (2019)