Accepted for publication in *SIAM Journal on Scientific Computing*

SPARSE MULTIDIMENSIONAL EXPONENTIAL ANALYSIS WITH AN APPLICATION TO RADAR IMAGING*

ANNIE CUYT†, YUAN HOU†, FERRE KNAEPKENS†, AND WEN-SHIN LEE‡

Abstract. We present a d-dimensional exponential analysis algorithm that offers a range of advantages compared to other methods. The technique does not suffer the curse of dimensionality and only needs $O((d+1)n)$ samples for the analysis of an n-sparse expression. It does not require a prior estimate of the sparsity n of the d-variate exponential sum. The method can work with sub-Nyquist sampled data and offers a validation step, which is very useful in low SNR conditions. A favourable computation cost results from the fact that d independent smaller systems are solved instead of one large system incorporating all measurements simultaneously. So the method also lends itself easily to a parallel execution. Our motivation to develop the technique comes from 2D and 3D radar imaging and is therefore illustrated on such examples.

Key words. Exponential analysis, parametric method, multidimensional, sparse model, sparse data, inverse problems

AMS subject classifications. 41A21, 42A15, 42B99, 65D05 65Z05, 97M50

1. Introduction.

In the past few years multidimensional exponential analysis has attracted considerable attention, as researchers were trying to solve the problem from the theoretical minimal number of samples, which equals the product of n, the number of terms in the multivariate exponential sum (2.1), and $d+1$ where d indicates the dimension (see among others [10, 30, 11]). Contrary to other approaches, the method presented in [10] does not need a full d-dimensional grid of data, nor does it require the solution of the full-size structured generalized eigenvalue problem and linear system of interpolation conditions. Instead, in [10] the linear algebra problems are split up in smaller better conditioned problems. In this paper, we develop a reliable numerical implementation, making use of this divide-and-conquer flavour and adding a sub-Nyquist sparse sampling feature to the basic theory, to deal with noisy data, tackle the numerical sensitivity in case of closely spaced exponential terms, offer output validation and provide automatically a reliable estimate of n, being the total number of terms.

Multidimensional exponential analysis is a fundamental inverse problem in signal processing, as it appears in magnetic resonance spectroscopy, MIMO radar, sonar, wireless communication, antenna array processing, sensor networks, RFID, GNSS and automotive radar, to name just a few. It is also at the basis of inverse synthetic aperture radar (ISAR) imaging, where the challenge is to extract high resolution information from noisy data, if possible using a cost effective algorithmic solution rather than an expensive advanced radar system. This application will serve as a guiding example throughout the paper.

ISAR imaging is a system that consists of a real-aperture radar, emitting a sequence of high frequency bursts, and a moving target in the far field of the radar,

*Submitted to the editors DATE.

Funding: Author Yuan Hou was financially supported by the China Scholarship Council under file no. 201706140133 and author Ferre Knaepkens was financially supported by FWO-Flanders through research project G019316N on sparse modeling.

†Department of Mathematics and Computer Science, University of Antwerp, Antwerp, Belgium (annie.cuyt@uantwerpen.be, yuan.hou@uantwerpen.be, ferre.knaepkens@uantwerpen.be).

‡Division of Computing Science and Mathematics, University of Stirling, Stirling, United Kingdom (wen-shin.lee@stir.ac.uk).

causing backscattering. When the target is hit by an electromagnetic wave, a limited number of locations on the object, such as edges and surface discontinuities, scatter the energy back toward the observation point. The locations of these concentrated sources of scattering energy are called scattering centers, each of which can be described by a multivariate complex exponential. ISAR is widely used and plays an important role in target identification, commercial aircraft classification, military surveillance and the like.

So the scattering center model in ISAR consists of a finite linear combination of complex exponentials that describe the different scattering centers of the radar target, where the number of these scattering centers is considerably less than the number of image pixels. Although the model is both simple and sparse, the inverse problem of reliably extracting the location of the scattering centers is rather sensitive to noise [37]. Therefore the problem has attracted a lot of research, which we roughly summarize below.

Fourier-based methods require a large densely sampled 2- or 3-dimensional data set, which may require a relatively long time to collect. Also, these techniques are trapped in the dilemma of time versus frequency resolution and cannot distinguish closely spaced scatterers, as mentioned in [20]. So several researchers have turned their attention to Prony-like spectral estimation or exponential analysis algorithms. In [27] the authors also conclude that the latter are much more accurate than Fourier based methods. But the performance of exponential analysis methods can be seriously affected by a low signal-to-noise ratio (SNR), leading to misclassifying noise as signals.

Here we present another Prony-like technique which allows to overcome this drawback. Also, the number of scatterers must not be estimated a priori, as pointed out in [1] for other parametric methods. In addition, the new technique does not suffer the well-known curse of dimensionality. A d-dimensional exponential analysis of an n-term model can now be carried out from a mere $O((d+1)n)$ regularly collected samples, which is substantially less than in other Prony-based methods [28, 37, 30, 16, 24, 26], where the sample usage explodes exponentially. In [37] the entailed complexity of these numerical algorithms is improved by the use of a slicing technique. The computation cost of the new method here compares much more favourably, as we solve several smaller systems instead of one large system dealing with all measurements at the same time.

The theory of compressive sensing also works with sparsely located data, which are however, collected randomly instead of regularly. Moreover, in radar imaging the results may be severely affected if the scattering centers on the target do not match the pre-discretized scene grid which makes up the dictionary [5]. We emphasize that methods of the Prony family do not work with a discretized grid and hence do not suffer from this drawback.

Other optimisation based ISAR techniques include genetic and evolutionary algorithms [19, 6]. While they are quite robust and can work completely automatic, without estimation of the model order, they require a lot of computation time, a disadvantage shared by most optimisation based methods. Several 2-dimensional compressive sensing or other optimisation approaches [34, 36, 1] may not be feasible in higher dimensions.

The paper is organized as follows. The proposed d-dimensional exponential analysis is presented in Section 2. An additional validation of the computed results, which proves to be very useful when working with low SNR, is introduced in Section 3. The details of the exponential model governing ISAR imaging are given in Section 4, together with a first application and comparison of the new method to [16]. A way to

recondition and subsequently regularize the d-dimensional exponential analysis is explained in Section 5. The full-blown method, including validation and reconditioning, is illustrated in Section 6, where it is further compared to [30]. Among the existing d-dimensional exponential analysis generalizations, we chose to compare our 2-D and 3-D numerical illustrations to [16] and [30] for the following reasons. In Section 4.2 we compare to the 2-D Prony-like algorithm `MEMP` from [16] to illustrate the need for an automatic pairing of the separately computed 1-D results, as is available in the new method. In Section 6 we compare to the multidimensional `ESPRIT` algorithm from [30] to illustrate the importance of obtaining an automatic estimation of the sparsity n, which is considered to be a difficult problem but is solved here.

2. Multidimensional exponential analysis. The problem of d-dimensional exponential analysis consists in retrieving the linear parameters $\alpha_j \in \mathbb{C}$ and the nonlinear parameters $\phi_j \in \mathbb{C}^d$ in the exponential model

$$f(x) = \sum_{j=1}^{n} \alpha_j \exp\left(\langle \phi_j, x \rangle\right), \qquad x = (x_1, \ldots, x_d), \qquad \phi_j = (\phi_{j1}, \ldots, \phi_{jd}) \tag{2.1}$$

from as few function samples as possible. Until recently, algorithms to solve the problem required a number of samples of the order $O(n^d)$ [16, 18, 24, 26] or $O(2^d n)$ [30] or at most $(d+1)n^2 \log^{2d-2} n$ [31], all growing exponentially with the dimension of the problem statement. In this section we present a reliable implementation which is based on [10] and requires only $O((d+1)n)$ regularly gathered samples.

Let $\Delta_1 = (\Delta_{11}, \ldots, \Delta_{1d}) \neq (0, \ldots, 0)$ and [22, 32]

$$|\Im(\langle \phi_j, \Delta_1 \rangle)| < \pi, \qquad j = 1, \ldots, n, \tag{2.2}$$

where the function $\Im(\cdot)$ returns the imaginary part of a complex number. Let us sample $f(x)$ at the points $s\Delta_1$:

$$F_s := f(s\Delta_{11}, \ldots, s\Delta_{1d}), \qquad s = 0, \ldots, 2n-1. \tag{2.3}$$

For simplicity we also assume that the sampling direction Δ_1 is such that the values $\exp(\langle \phi_j, \Delta_1 \rangle)$, $j = 1, \ldots, n$ are mutually distinct. How to deal with collisions in these values is described in [10].

The expressions $\exp(\langle \phi_j, \Delta_1 \rangle), j = 1, \ldots, n$ are retrieved as the generalized eigenvalues λ_j of the problem

$$\begin{pmatrix} F_1 & F_2 & \cdots & F_n \\ F_2 & \cdots & & F_{n+1} \\ \vdots & & & \vdots \\ F_n & F_{n+1} & \cdots & F_{2n-1} \end{pmatrix} v_j = \lambda_j \begin{pmatrix} F_0 & F_1 & \cdots & F_{n-1} \\ F_1 & \cdots & & F_n \\ \vdots & & & \vdots \\ F_{n-1} & F_n & \cdots & F_{2n-2} \end{pmatrix} v_j, \qquad v_j \in \mathbb{C}^n \tag{2.4}$$

where the v_j denote the right eigenvectors. For the sake of completeness and for use further on, we point out that the upper left element in the left and right hand side matrices need not carry the indices 1 and 0 respectively. We can start with a higher index number instead of 0, as long as we have $2n$ consecutive samples lined up in (2.4) [7]. So the sampling of $f(x)$ in the direction of Δ_1 need not start at the origin.

In applications, the generalized eigenvalue problem (2.4) is often solved as part of a classic one-dimensional exponential analysis algorithm. In our numerical illustrations

we use the combination of the matrix pencil method studied in [17, 35] with the rank reduction step described in [29]. In the literature this combination is often referred to as the ESPRIT method, although the rank reduction is performed on the Hankel matrices directly instead of the covariance matrices. For the practical details concerning this aspect, the reader is referred to the Sections 4 and 6. In the Sections 2, 3 and 5 the mathematical backbone of the new method is developed.

Because of (2.2), we can uniquely recover the inner products

$$\Phi_j := \langle \phi_j, \Delta_1 \rangle, \qquad j = 1, \ldots, n \tag{2.5}$$

from the computed $\exp(\Phi_j)$. Although we have not yet identified the individual $\phi_{ji}, j = 1, \ldots, n, i = 1, \ldots, d$, nothing prevents us from already computing the linear coefficients α_j from the $2n \times n$ Vandermonde problem

$$\begin{pmatrix} 1 & \cdots & 1 \\ \exp(\Phi_1) & \cdots & \exp(\Phi_n) \\ \vdots & & \vdots \\ \exp((2n-1)\Phi_1) & \cdots & \exp((2n-1)\Phi_n) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = \begin{pmatrix} F_0 \\ \vdots \\ F_{2n-1} \end{pmatrix} \tag{2.6}$$

Note that (2.6) reduces to a square Vandermonde system in the noisefree case, because then n of the linear equations are linearly dependent as a consequence of the fact that the values $\exp(\Phi_j)$ already satisfy (2.4).

In order to extract the $\phi_{ji}, j = 1, \ldots, n, i = 1, \ldots, d$ from the $\Phi_j, j = 1, \ldots, n$, still under the assumption that the values $\exp(\Phi_j), j = 1, \ldots, n$ are mutually distinct, some extra samples are required. We choose an additional $d-1$ linearly independent vectors $\Delta_2, \ldots, \Delta_d$ such that the set $\{\Delta_1, \Delta_2, \ldots, \Delta_d\}$ is a basis. The additional samples are then taken along a linear combination of Δ_1 and some $\Delta_i, i = 2, \ldots, d$:

$$F_{si} := f(s\Delta_1 + \Delta_i), \qquad s = 0, \ldots, n-1, \quad i = 2, \ldots, d. \tag{2.7}$$

Note that only n additional samples are taken per Δ_i-shift and that they are placed equidistantly along independent shifts Δ_i with respect to the original vector Δ_1. At the same time we assume the Nyquist constraint [22, 32]

$$|\Im(\langle \phi_j, \Delta_i \rangle)| < \pi, \qquad j = 1, \ldots, n, \quad i = 2, \ldots, d. \tag{2.8}$$

We call these vectors $\Delta_i, i = 2, \ldots, d$ identification shifts because they will allow to identify the individual ϕ_{ji} in the computed Φ_j from samples taken at shifted locations. For this last step we make use of the fact that the ϕ_{ji} appear linearly in the Φ_j.

For i fixed, the additional samples F_{si} can be written as

$$\begin{aligned} F_{si} = f(s\Delta_1 + \Delta_i) &= \sum_{j=1}^{n} \alpha_j \exp\left(\langle \phi_j, \Delta_i \rangle\right) \exp\left(\langle \phi_j, s\Delta_1 \rangle\right), \qquad s = 0, \ldots, n-1 \\ &= \sum_{j=1}^{n} A_{ji} \exp\left(s\Phi_j\right), \qquad A_{ji} = \alpha_j \exp\left(\langle \phi_j, \Delta_i \rangle\right). \end{aligned}$$

So for i fixed, the $A_{ji}, j = 1, \ldots, n$ are obtained from the Vandermonde system

$$\begin{pmatrix} 1 & \cdots & 1 \\ \exp(\Phi_1) & \cdots & \exp(\Phi_n) \\ \vdots & & \vdots \\ \exp((n-1)\Phi_1) & \cdots & \exp((n-1)\Phi_n) \end{pmatrix} \begin{pmatrix} A_{1i} \\ \vdots \\ A_{ni} \end{pmatrix} = \begin{pmatrix} F_{1i} \\ \vdots \\ F_{ni} \end{pmatrix} \tag{2.9}$$

of which the coefficient matrix is part of the Vandermonde structured coefficient matrix in (2.6). From the A_{ji} and the α_j we obtain for i fixed,

$$\frac{A_{ji}}{\alpha_j} = \exp\left(\langle \phi_j, \Delta_i \rangle\right), \qquad j = 1, \ldots, n, \tag{2.10}$$

where in the sequel we denote

$$\Phi_{ji} := \langle \phi_j, \Delta_i \rangle, \qquad j = 1, \ldots, n.$$

Note that we have no problem to pair the Φ_{ji} to the $\Phi_j, j = 1, \ldots, n$ since for each i the A_{ji} are paired to the $\alpha_j, j = 1, \ldots, n$ through the Vandermonde systems (2.6) and (2.9).

These A_{ji} and $\exp(\Phi_{ji})$ can be computed for each $i = 2, \ldots, d$. The fact that the vectors Δ_1 and $\Delta_i, i = 2, \ldots, d$ are linearly independent then leads for each separate $j = 1, \ldots, n$ to the $d \times d$ regular linear system

$$\begin{pmatrix} \Delta_{11} & \cdots & \Delta_{1d} \\ \Delta_{21} & \cdots & \Delta_{2d} \\ \vdots & & \vdots \\ \Delta_{d1} & \cdots & \Delta_{dd} \end{pmatrix} \begin{pmatrix} \phi_{j1} \\ \vdots \\ \phi_{jd} \end{pmatrix} = \begin{pmatrix} \Phi_j \\ \Phi_{j2} \\ \vdots \\ \Phi_{jd} \end{pmatrix} \tag{2.11}$$

from which the individual $\phi_{ji}, j = 1, \ldots, n, i = 1, \ldots, d$ are computed.

So all unknowns in (2.1) can be obtained at the expense of $2n$ evaluations F_s in (2.3) and $(d-1)n$ evaluations F_{si} in (2.7), or a mere total of $(d+1)n$ samples. In practice, when dealing with noisy data, the value of n is overestimated by $\eta > n$, as discussed in the next section. The minimal number of samples in an η-term exponential model of the form (2.1), in the directions Δ_1 and $\Delta_i, i = 2, \ldots, d$, which is respectively 2η and η, is often again overestimated by $N \geq 2\eta$ and $\mathfrak{n} \geq \eta$. The square $n \times n$ generalized eigenvalue problem (2.4), the $2n \times n$ Vandermonde system (2.6) and the $n \times n$ Vandermonde system (2.9) then respectively take the sizes $(N-\eta) \times \eta$, $N \times \eta$ and $\mathfrak{n} \times \eta$ and are all solved in the least squares sense. Sometimes some of the samples are used in a preprocessing step, such as the computation of an intermediate $(N-\eta) \times \nu$ structured lower rank approximation to the Hankel matrices, where $\nu < \eta$.

In the next sections we describe how this technique is combined with convergence theorems from approximation theory on the one hand and sparse interpolation from computer algebra on the other hand, in order to:

- filter unstructured noise in the data out of the structured exponential model (2.1) via a connection to Padé approximation theory,
- automatically deduce and validate the sparsity n of expression (2.1), which is usually regarded to be a hard problem,
- separate exponential components that are contained in a cluster of similar components, using a connection with sparse interpolation,
- and as a result of all of the above, tighten the numerical estimates for the parameters ϕ_j and α_j in case of a low signal-to-noise ratio.

3. Connection with Padé approximation: validation. From the theoretical mathematical presentation in Section 2, we now switch to the practical situation where the samples F_s and F_{si} are contaminated by noise. For the reliable computation of

the parameters ϕ_j and α_j we need to add some additional steps to the algorithm. The first change is that we are going to interpret the samples as if they are coming from an η-term exponential model of the form (2.1), where $\eta > n$ is a safe overestimate of n. A connection with Padé approximation theory will then allow us to separate the noise from the actual signal content.

Consider the function

$$\mathfrak{f}(z) = \sum_{s=0}^{\infty} F_s z^s.$$

With F_s given by (2.3) we can write [33, 2]

$$\mathfrak{f}(z) = \sum_{j=1}^{n} \frac{\alpha_j}{1 - \exp(\Phi_j) z}. \tag{3.1}$$

The partial fraction decomposition (3.1) is related to the one-dimensional Laplace transform and the Z-transform of (2.1), where the inner product $\langle \phi_j, x \rangle$ is regarded as the unknown. It is a rational function of degree $n-1$ in the numerator and degree n in the denominator with poles $1/\exp(\Phi_j)$. Now let us perturb $\mathfrak{f}(z)$ with Gaussian noise to obtain

$$\mathfrak{f}(z) + \epsilon(z) = \sum_{s=0}^{\infty} (F_s + \epsilon_s) z^s.$$

The theorem of Nuttall-Pommerenke states that if $\mathfrak{f}(z) + \epsilon(z)$ is analytic throughout the complex plane, except for a countable number of poles [21] and essential singularities [25], then its sequence of Padé approximants $\{r_{\eta-1,\eta}(z)\}_{\eta \in \mathbb{N}}$ of degree $\eta - 1$ over η converges to $\mathfrak{f}(z) + \epsilon(z)$ in measure on compact sets. This means that for sufficiently large η, the measure of the set where the convergence is disrupted, so where $|\mathfrak{f}(z) + \epsilon(z) - r_{\eta-1,\eta}(z)| \geq \tau$ for some given threshold τ, tends to zero as η tends to infinity. Pointwise convergence is disrupted by $\eta - n$ unwanted pole-zero combinations of the Padé approximants that are added to the n true poles and $n-1$ true zeros of $\mathfrak{f}(z)$ [13, 15], the pole and zero in the pair almost cancelling each other locally. These pole-zero combinations are referred to as Froissart doublets. In practice, these Froissart doublets offer a way to separate the noise $\epsilon(z)$ from the underlying $\mathfrak{f}(z)$ [14, 15]. Because of the Padé convergence theorem, the n true (physical) poles are identified as stable poles in successive $r_{\eta-1,\eta}(z)$, while the $\eta - n$ spurious (noisy) poles are distinguished by their instability. For different η [3, 23]:

- the noisy poles lie scattered in the area around the complex unit circle, and this for every different realization of the noise $\epsilon(z)$,
- and the true poles $\exp(-\Phi_j), j = 1, \ldots, n$ are forming clusters with around each $\exp(-\Phi_j)$ cluster an almost Froissart doublet-free zone.

This characteristic of the true poles is precisely the key point on which our method is based: after the computation of $\eta > n$ generalized eigenvalues $\lambda_j = \exp(\Phi_j)$, we discard the unstable ones and focus on the stable ones. Note that:

- In order to safely rely on this convergence result, it is clear that η should be sufficiently large, as the result is more numerically accurate for η large. We usually take η to be a multiple of (the so far unknown) n.

- To decide which generalized eigenvalues are the unstable ones, the computational scheme needs to be repeated a number of times with different sets of $N + (d-1)\mathfrak{n}$ data, which can be achieved as follows.

We discuss the sampling along the Δ_1 direction first. Instead of collecting $F_s, s = 0, \ldots N-1$ in the direction of Δ_1, we collect some additional $F_s, s = 0, \ldots, N + (\kappa - 1)\lfloor \mathfrak{p} N \rfloor - 1$. Here $0 \leq \mathfrak{p} \leq 1$ and $1 \leq \kappa \in \mathbb{N}$. From these samples we construct κ snapshots of N samples each, snapshot number $k = 0, \ldots, \kappa - 1$ starting at $k\lfloor \mathfrak{p} N \rfloor$ with an overlap of roughly $(1-\mathfrak{p})N$ points with the previous and the next snapshot. The case $\mathfrak{p} = 0$ and $\kappa = 1$ delivers the single snapshot situation of the previous section.

When putting all $\kappa\eta$ generalized eigenvalues of the κ different eigenvalue problems (2.4) together, then theoretically κn of them cluster together in n clusters of each κ elements and the other $\kappa(\eta - n)$ generalized eigenvalues lie scattered around as they do not reflect true terms in the exponential model (2.1). Of course, the noise may be such that the method does not work perfectly and that in an apparent cluster somewhat less than κ elements are found. We therefore accept a cluster as soon as a sufficiently large fraction of the κ expected elements is found.

In the numerical examples we found it most useful to use a density-based cluster analysis such as `DBSCAN` [12]. The `DBSCAN` implementation requires two parameters: the density δ of the clusters and the minimum number m_δ of required cluster elements. These parameters are chosen in terms of the noise in the signal:

- Larger values of δ allow the detection of wider clusters, for instance in case of a higher noise level. Smaller values of δ lead to denser clusters with very stable estimates for the generalized eigenvalues, for instance in case of lower levels of noise.
- A value for m_δ smaller than κ allows to discard bogus estimates appearing as a consequence of outliers in the data or too high noise levels. It makes perfect sense, depending on the application, to relax m_δ to for instance $\kappa - 1, \kappa - 2$ or $\lfloor 0.9\kappa \rfloor, \lceil 0.8\kappa \rceil$.

A very desirable side result of the technique described in this section, is the fact that the method automatically reveals the true number n of terms in the expression (2.1) underlying all the samples: n equals the number of clusters detected by the cluster analysis.

Remains to discuss the sampling along the linearly independent shifts of Δ_1. Here also, the data set needs to be enlarged in order to support the processing of κ snapshots. So at most we collect for each $i = 2, \ldots, d$ the samples $F_{si}, s = 0, \ldots, \mathfrak{n} + (\kappa - 1)\lfloor \mathfrak{p} N \rfloor - 1$ (for some choices of the parameters $N, \kappa, \mathfrak{n}, \mathfrak{p}$ not all consecutive samples are used). Remember that each of the computed $\Phi_{ji}, j = 1, \ldots, \eta, i = 2, \ldots, d$ is connected to its $\Phi_j, j = 1, \ldots, \eta$ from the solution of the generalized eigenvalue problem (2.4). For i fixed, we therefore know which Φ_{ji} are linked to a cluster element Φ_j and which belong to a scattered Φ_j. When taking the m_δ values Φ_{ji} together that are linked to a cluster element Φ_j, then we can improve the estimate for $\Phi_{ji}, j = 1, \ldots, n, i = 2, \ldots, d$ by considering the center of gravity of the m_δ values Φ_{ji} that go together. As the Φ_{ji} are obtained from the solution of two Vandermonde structured linear systems through (2.10), their estimates are usually found to be somewhat less accurate than the estimates computed for the clustered Φ_j.

Analysis of the Φ_{ji} values when taking κ snapshots can also serve an additional purpose. Sometimes it is useful to run `DBSCAN` a consecutive number of times with

increasing values for δ. In this way, very condensed clusters are detected right from the start and more relaxed clusters are picked up in some later run. In case δ is relaxed too much, an inspection of the (at least) m_δ values Φ_{ji} associated with the (at least) m_δ estimates for a particular Φ_j in the candidate cluster, helps to accept or refute the relaxed cluster. The latter can be done by looking at the spread (standard deviation) of the associated Φ_{ji}. If this exceeds an acceptable threshold, the candidate cluster is rejected. So while a cluster of m_δ estimates for some Φ_j is "identified", it is "confirmed" by the analysis of the m_δ associated values Φ_{ji} and this for each $i = 2 \ldots, d$.

Let us illustrate the procedure described in the Sections 2 and 3 on some small-scale numerical examples. In Section 5 we further explain how to deal with the situation where some of the clusters around the true $\Phi_j, j = 1, \ldots, n$ partially overlap, for instance because of very similar $\phi_j, j = 1, \ldots, n$ in the exponential model (2.1).

4. Application to ISAR imaging. High frequency scattering toward an observation point is often modeled by means of a finite number of concentrated sources of scattering energy, also called scattering centers. A radar signal backscattered from a far-field target with n scattering centers at locations (x_j, y_j, z_j), $j = 1, \ldots, n$ in a cartesian coordinate system, is then decomposed into n contributions, each with a different phase and magnitude.

Assume the radar system emits a signal at frequency ω_h in the direction or line of sight with azimuth angle θ_g and elevation angle ϕ_m. The backscattered signal $f(h, g, m)$ with $(h, g, m) \in \mathbb{R}^3_+$ is approximated by the following sum of complex exponentials,

$$f(h,g,m) \approx \sum_{j=1}^{n} \beta_j \exp\left(-\frac{4\pi \mathrm{i}}{c}(\omega_h x_j + \omega_c \theta_g y_j + \omega_c \phi_m z_j)\right), \tag{4.1}$$

where β_j is the scattering amplitude of the j-th scattering center, c is the speed of light, ω_c is the central frequency $\omega_c = (\omega_0 + \omega_{(N-1)h})/2$ and the parameters ω_h, θ_g and ϕ_m are defined by

$$\omega_h = \omega_0 + h\delta_\omega, \qquad \theta_g = \theta_0 + g\delta_\theta, \qquad \phi_m = \phi_0 + m\delta_\phi.$$

The remaining values $\omega_0, \theta_0, \phi_0$ and $\delta_\omega, \delta_\theta, \delta_\phi$ are set by the user and are system dependent. We rewrite the exponential model (4.1) as

$$f(h,g,m) \approx \sum_{j=1}^{n} \alpha_j \exp\left(-\frac{4\pi \mathrm{i}}{c}(h\delta_\omega x_j + \omega_c g\delta_\theta y_j + \omega_c m\delta_\phi z_j)\right),$$

$$\alpha_j = \beta_j \exp\left(-\frac{4\pi \mathrm{i}}{c}(\omega_0 x_j + \omega_c \theta_0 y_j + \omega_c \phi_0 z_j)\right).$$

By means of the Prony-like method presented in Section 2, the computation of the unknown scattering locations $(x_j, y_j, z_j), j = 1, \ldots, n$ and the unkown scattering amplitudes $\beta_j, j = 1, \ldots, n$ is then neatly separated, with the scattering locations being delivered first after applying (2.9) and (2.10).

4.1. 3-dimensional illustration of the new algorithm. To illustrate the method on a synthetic small-scale 3-D example, we consider the 29-term exponential expression (4.1) with (x_j, y_j, z_j) and β_j given in Table 1. We further set the following

radar parameters:

$$\omega_0 = 7.9\text{GHz}, \qquad \delta_\omega = 0.0015\text{GHz},$$
$$\theta_0 = \phi_0 = -0.024, \qquad \delta_\theta = \delta_\phi = 3.75 \times 10^{-4}.$$

We choose (for no specific reason, except that (2.2) and (2.8) need to be satisfied): the linearly independent vectors $\Delta_1 = (1.17, 0.7, 1.87), \Delta_2 = (-1.00, -1.00, -1.00), \Delta_3 = (-2.10, 1.20, 3.29)$ as basis vectors for all (h, g, m), and furthermore $N = 450, \nu = 150, \eta = 100, \mathfrak{n} = 450, \kappa = 10, \mathfrak{p} = 0.1$ and start the collection of the required samples. To each evaluation of (4.1) we add a Gaussian noise term of a fixed prechosen signal-to-noise ratio SNR (in dB). So

$$\begin{aligned} F_s &= f(s\Delta_1) + \epsilon_s, & s &= 0, \ldots, M-1 := N + (\kappa-1)\lfloor \mathfrak{p}N \rfloor - 1 \\ F_{s2} &= f(s\Delta_1 + \Delta_2) + \epsilon_{M+s}, & s &= 0, \ldots, \mathfrak{m}-1 := \mathfrak{n} + (\kappa-1)\lfloor \mathfrak{p}N \rfloor - 1 \\ F_{s3} &= f(s\Delta_1 + \Delta_3) + \epsilon_{M+\mathfrak{m}+s}, & s &= 0, \ldots, \mathfrak{m}-1. \end{aligned}$$

With our choices for the parameters, we sample at 855 points in the direction of Δ_1 and another 855 at each of the shifted locations $s\Delta_1 + \Delta_i, i = 2, 3$, or a total of 2565 points. This number is in sharp contrast with even the simplest $(\Delta_1, \Delta_2, \Delta_3)$-grid structured data set of $35 \times 35 \times 35 = 42875$ points, where we choose 35 as a very mild overestimate of $n = 29$. In addition, a d-dimensional algorithm departing from a grid structured data set [16] does not offer any of the advantages we have discussed so far, among which:

- the natural pairing of $\Phi_{ji}, i = 2, \ldots, d$ to $\Phi_j, j = 1, \ldots, n$,
- the automatic detection of the sparsity n, and
- the validation of the computed locations (x_j, y_j, z_j).

In Figure 1 we show the `DBSCAN` result for SNR = 10 dB with $m_\delta = \kappa - 2$ and δ varying over $5^\ell \times 10^{-4}, \ell = 0, \ldots, 4$: among the 1000 computed generalized eigenvalues ($\eta = 100, \kappa = 10$) 29 clusters are indicated in colour. They identify the stable generalized eigenvalues, which were detected and confirmed by the algorithm outlined in Section 2. None of the groups of m_δ associated values $\exp(\Phi_{ji}), j = 1, \ldots, 29, i = 2, 3$ exhibits a standard deviation larger than 0.25.

We also run the above example for varying noise levels, from 40 dB SNR to 5 dB SNR, now with $\kappa = 20$ and each experiment repeated 100 times as the noise is randomly generated. In Figure 2 we show the average true cluster radius over the 29 scattering locations, for the generalized eigenvalues $\exp(\Phi_j)$ with $m_\delta = \kappa - 2$. This true radius is computed a posteriori with the exact $\exp(\Phi_j)$ in the center. In Figure 3 we respectively show at the left and the right for $i = 2, 3$ the average cluster radius over the 29 scattering locations, for the $\kappa - 5$ estimates closest to the true associated $\exp(\Phi_{ji})$. With $\kappa = 20$, a ratio of $\kappa - 5$ over the maximum number m_δ of associated elements still represents 83.3% of the associated values. In each of the Figures 2 and 3 we also show the smallest and largest cluster radius (dashed lines): they differ by a factor of about 2. It is quite clear that the computation of the $\exp(\Phi_j)$ is more accurate than that of the $\exp(\Phi_{ji})$. The estimates of the latter can be tightened but this is not really important at this point.

4.2. 2-dimensional illustration of the validation aspect. In another experiment we consider the 2-D example with 12 scattering centers (x_j, y_j) from Table 2. The dimension is reduced from three to two for the sole reason that in our figures

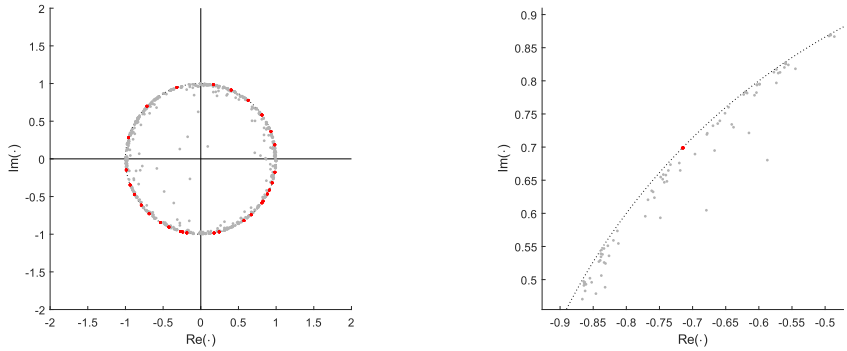

Fig. 1: The 29 clusters (red) identified by DBSCAN (left) and a zoom (right) on the stable clustered (red) versus the unstable scattered (grey) generalized eigenvalues

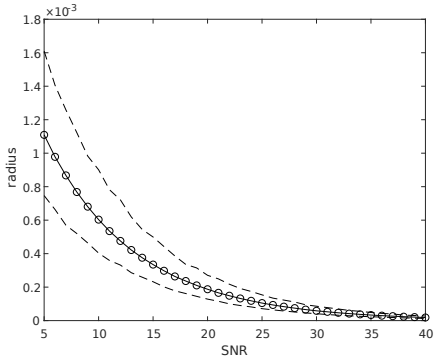

Fig. 2: Average radius of the $\exp(\Phi_j), j = 1, \ldots, 29$ clusters

we want to use the third dimension to graph the impact of the SNR. The radar parameters ω_0, θ_0 and $\delta_\omega, \delta_\theta$ are as in Section 4.1. We further take $N = 150, \nu = \eta = 50, \mathfrak{n} = 100, \kappa = 11, \mathfrak{p} = 0.054$ and $\Delta_1 = (1.38, 4.14), \Delta_2 = (-7.56, 5.67)$.

In order to reduce the cluster radius in the shift direction Δ_2 we perform the shift a number of times, over $\Delta_2, 2\Delta_2, \ldots, 8\Delta_2$ and combine the results, since for s fixed,

$$f(s\Delta_1 + m\Delta_2) = \sum_{j=1}^{n} \alpha_j \exp(\langle \phi_j, m\Delta_2 \rangle) \exp(\langle \phi_j, s\Delta_1 \rangle).$$

So the coefficients extracted from the different shifts are

$$\alpha_j \exp(\langle \phi_j, \Delta_2 \rangle), \alpha_j \exp(\langle \phi_j, 2\Delta_2 \rangle), \ldots, \alpha_j \exp(\langle \phi_j, 8\Delta_2 \rangle).$$

The total number of collected samples then adds up to $(\kappa - 1)\lfloor \mathfrak{p} N \rfloor + N$ in the Δ_1 direction and $8 \times ((\kappa - 1)\lfloor \mathfrak{p} N \rfloor + \mathfrak{n})$ in the Δ_2 shifts, or 1670 samples altogether.

In Figure 4 (right) we show the result of the computations, after applying `DBSCAN` with $m_\delta = \kappa - 1$ and $\delta = 0.00001, 0.002505, 0.005$ to each SNR result for the Φ_j and discarding cluster results when the standard deviation of the Φ_{j2} exceeds 0.2. We let

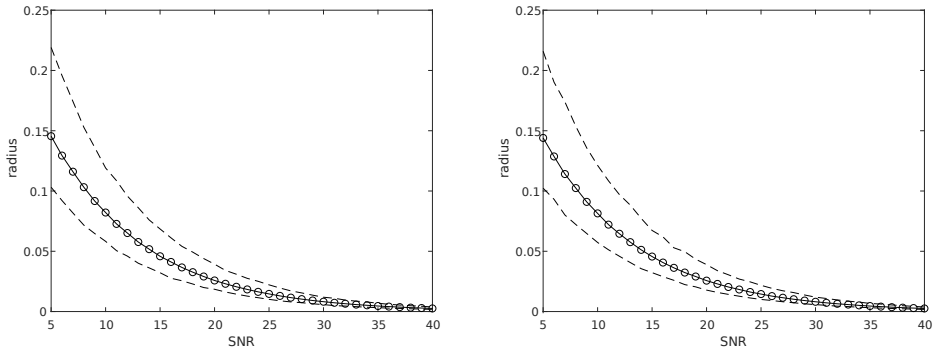

Fig. 3: Average radius of the $\exp(\Phi_{j2}), j = 1, \ldots, 29$ clusters (left) and the $\exp(\Phi_{j3})$, $j = 1, \ldots, 29$ clusters (right)

the SNR vary from 40 dB to 5 dB, top to bottom. The SNR = 10 dB slice is presented in Figure 5 (right) and a separate coordinate view is found in Figure 6 (right) where the SNR varies from right to left. For each SNR the experiment is repeated 250 times.

We compare these results to the output in the Figures 4 (left), 5 (left) and 6 (left) of the 2-D Prony-like algorithm `MEMP` [16] using the same number of samples but now laid out in a (Δ_1, Δ_2)-grid of size 40×42. We remark the main differences with the new algorithm:

- the matching in the `MEMP` algorithm between results computed in separate dimensions is definitely not flawless, and as the noise increases erroneous combinations give rise to inexistent locations;
- the matching through the indexing of the variables in (2.6) and (2.9) leaves no room for error and so does not introduce matching errors;
- for increasing noise, meaning decreasing SNR, the unvalidated `MEMP` algorithm may return a few erroneous (x_j, y_j), despite the fact that the sparsity $n = 12$ was passed to the algorithm as well;
- the correct sparsity $n = 12$ need not be passed to the new algorithm, which detects it automatically as the number of identified and confirmed clusters;
- in the new algorithm the results for very small SNR are either somewhat less accurate or absent because of the high validation requirement, which can of course be relaxed by the user.

4.3. 3-dimensional fighter jet example. In a larger scale example of 1000 scatterers depicting the surface of a fighter jet [37], we take the radar parameters as in Section 4.1, add noise with SNR = 20 dB and further choose $N = 6000, \nu = 2000, \eta = 1500, \mathfrak{n} = 6000, \kappa = 11, \mathfrak{p} = 0.4$ with $\Delta_1 = (-2.2371, 0.2796, 0.8389), \Delta_2 = (1.6528, -1.6528, 4.9584), \Delta_3 = (0.4744, 2.1350, 0.5535)$. The density δ in `DBSCAN` was varied over $2^\ell \times 10^{-5}, \ell = 0, \ldots, 10$ while m_δ was kept at $m_\delta = \lceil 0.8\kappa \rceil = 9$.

When dealing with the $\exp(\Phi_{ji}), i = 2, 3$ we discard cluster results with a standard deviation above 0.5. We remark that as the density δ increases, the probability increases that a candidate cluster, detected among the Δ_1-projections, is not confirmed in each and every one of the Δ_i-projections, $i = 2, \ldots, d$. Rejection dominates acceptance from $\ell = 7$ on.

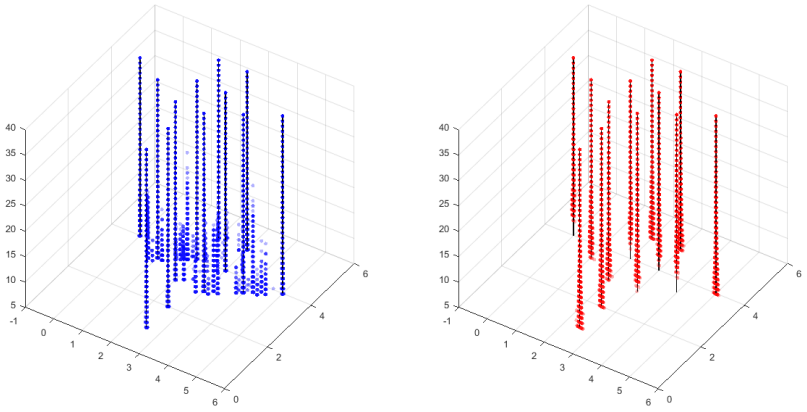

Fig. 4: Unvalidated (x_j, y_j) locations from `MEMP` (left) and validated (x_j, y_j) from the new algorithm (right).

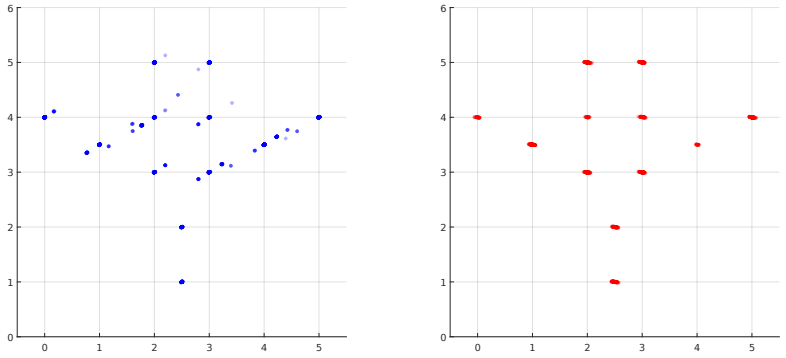

Fig. 5: Slice of Figure 4 (SNR = 10) with the colour intensity indicating the frequency of detection in the 250 runs.

In the end, the above algorithm detects and validates 516 scatterers (out of 1000), but misses out on the scatterers that are located too closely together or for which the inner products in (4.1) are too much alike. Although the overall shape of the fighter is correctly recognized (nose, wing tips, tail, ...), which may be more than satisfactory for many applications, the accuracy of the algorithm can be improved in the region where several scattering centers (x_j, y_j, z_j) are located near one another, such as the windshield. To this end the algorithm needs to be combined with a sub-Nyquist technique, particularly suitable for the exponential analysis of such signals [9]. This final addition to the algorithm is explained in the next section. We also point out that, thanks to the validation step, there are no false results.

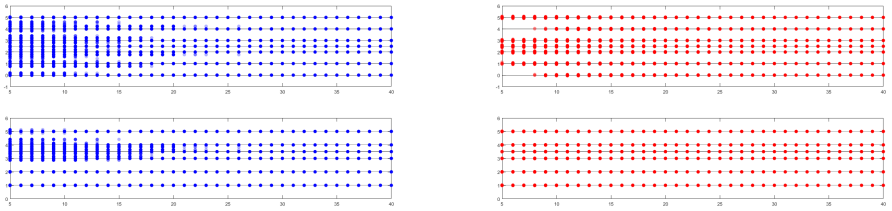

Fig. 6: Unvalidated x_j and y_j coordinates from `MEMP` (left) versus validated x_j and y_j coordinates from the new algorithm (right)

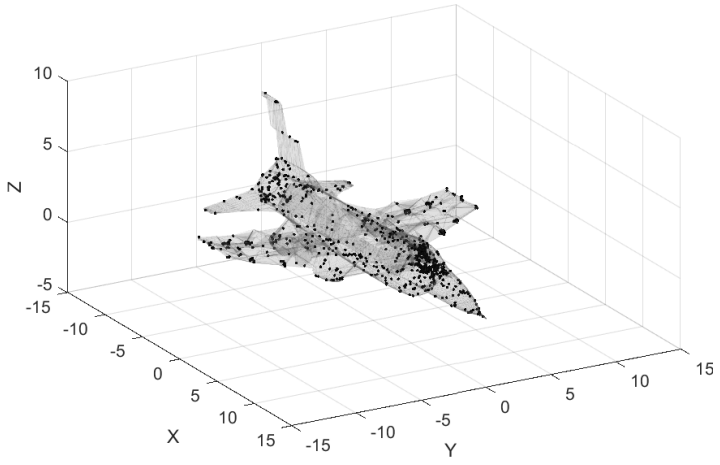

Fig. 7: Fighter jet original 1000 scattering center data

5. Connection with sparse interpolation: superresolution. We return to the notation of Section 2 to continue our presentation. When replacing the primary sampling direction Δ_1 by a multiple

$$\Delta_1(m) := m\Delta_1$$

and sampling at $s\Delta_1(m), s = 0, \ldots, 2n-1$ instead of at $s\Delta_1, s = 0, \ldots, 2n-1$, we are possibly violating the Shannon-Nyquist constraint (2.2) for $\Delta_1(m)$, when $|\Im(\langle \phi_j, \Delta_1 \rangle)| \geq \pi/m, j = 1, \ldots, n$. With

$$F_s := f\left(s\Delta_{11}(m), \ldots, s\Delta_{1d}(m)\right), \qquad s = 0, \ldots, 2n-1,$$

the eigenvalues retrieved from (2.4) are not λ_j but [4]

$$\lambda_j(m) = \exp(m\Phi_j) = \lambda_j^m, \qquad j = 1, \ldots, n.$$

From λ_j^m the imaginary part of $\Phi_j = \langle \phi_j, \Delta_1(m) \rangle$ may not be retrieved uniquely anymore because we can only guarantee that

$$|\Im(\langle \phi_j, \Delta_1(m) \rangle)| < m\pi. \tag{5.1}$$

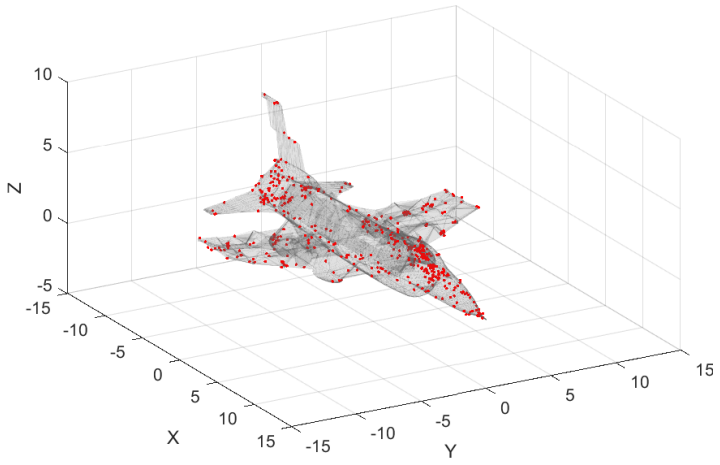

Fig. 8: Fighter jet reconstruction of 516 out of 1000 scatterers

So aliasing may have kicked in: because of the periodicity of $\exp(\Im(\langle\phi_j, m\Delta_1\rangle))$ a total of m values in the $2m\pi$ wide interval (5.1) can be identified as plausible values for $\langle\phi_j, \Delta_1\rangle$. Note that when the original λ_j are clustered, the powered λ_j^m may be distributed quite differently and unclustered. Such a relocation of the generalized eigenvalues, here referred to as superresolution, can seriously improve the conditioning of the Hankel matrices involved. In Figure 9 we show the effect of this powering on a particular example where 20 generalized eigenvalues are clustered in 5 clusters of different size.

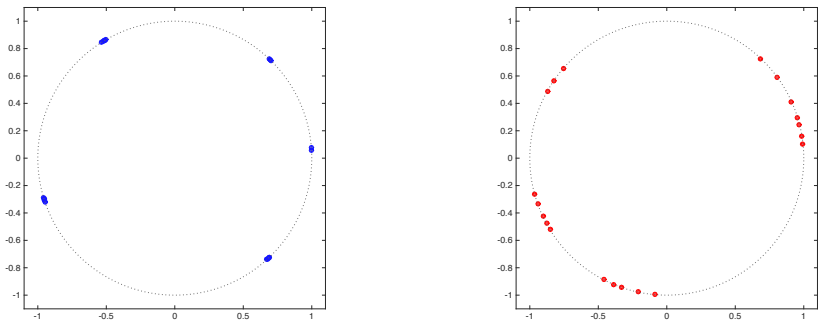

Fig. 9: Example analysis of (2.1) with $n = 20$ generalized eigenvalues: $m = 1$ (left) versus $m = 11$ (right)

What we need to resolve now is the aliasing problem that is possibly introduced by powering the generalized eigenvalues. This aliasing can be fixed at the expense of a small number of additional samples. Remember that in what follows, n can everywhere be replaced by $\eta \geq n$ when using $\eta - n$ additional terms to model the

noise.

To fix the aliasing, we add n samples to the collection $F_0, F_m, \ldots, F_{(2n-1)m}$, namely at the shifted points

$$s\Delta_1(m) + \mu\Delta_1 = (sm + \mu)\Delta_1, \qquad s = r, \ldots, r + n - 1, \quad 0 \le r \le n.$$

An easy choice for μ is a (small) number mutually prime with m (for the most general choice allowed, we refer to [8]). With the additional samples we proceed as follows:

- From the samples $F_0, F_m, \ldots, F_{(2n-1)m}$ we first compute the generalized eigenvalues λ_j^m and the coefficients α_j going with λ_j^m in the model

$$F_{sm} = f(sm\Delta_1) = \sum_{j=1}^{n} \alpha_j \exp(\langle \phi_j, sm\Delta_1 \rangle) \tag{5.2}$$

$$= \sum_{j=1}^{n} \alpha_j \lambda_j^{sm}, \qquad s = 0, \ldots, 2n-1. \tag{5.3}$$

  We know which coefficient α_j goes with which generalized eigenvalue λ_j^m, but we just cannot identify the correct $\Im(\langle \phi_j, \Delta_1 \rangle)$ from λ_j^m.
- Next we deal with the samples at the additional locations $sm\Delta_1 + \mu\Delta_1$, which satisfy

$$F_{sm+\mu} = f(sm\Delta_1 + \mu\Delta_1) = \sum_{j=1}^{n} \alpha_j \exp\left(\langle \phi_j, (sm + \mu)\Delta_1 \rangle\right) \tag{5.4}$$

$$= \sum_{j=1}^{n} (\alpha_j \lambda_j^{\mu}) \lambda_j^{sm}, \qquad s = r, \ldots, r + n - 1. \tag{5.5}$$

  This system is a linear system with a similar Vandermonde structured coefficient matrix as in (5.2), but now with a new left hand side $F_{sm+\mu}$ and unknown coefficients $\alpha_1\lambda_1^{\mu}, \ldots, \alpha_n\lambda_n^{\mu}$ instead of $\alpha_1, \ldots, \alpha_n$. And again we can associate each computed $\alpha_j\lambda_j^{\mu}$ with the proper generalized eigenvalue λ_j^m, because of the indexing of the variables and coefficients.
- Then, by dividing the $\alpha_j\lambda_j^{\mu}$ computed from (5.4) by the α_j computed from (5.2), for $j = 1, \ldots, n$, we obtain from each quotient λ_j^{μ} a second set of μ plausible values for $\langle \phi_j, \Delta_1 \rangle$ in the $2\mu\pi$ wide interval $|\Im(\langle \phi_j, \mu\Delta_1 \rangle)| < \mu\pi$.
- Because of the fact that we choose μ and m relatively prime, the two sets of plausible values for $\langle \phi_j, \Delta_1 \rangle$ have only one value in their intersection [9]. Thus the aliasing problem is solved: each $\langle \phi_j, \Delta_1 \rangle$ is retrieved uniquely from the computation of both λ_j^m and λ_j^{μ} for $j = 1, \ldots, n$.

This multidimensional sub-Nyquist sampling strategy may help us determine the clustered scattering centers occurring in Section 4.3. As suggested in Figure 9, the technique spreads out the generalized eigenvalues, which may recondition the inverse problem. In addition, a variation of scale factors m may be used and the idea can be translated to the sampling at the shifted locations involving the identification shifts $\Delta_2, \ldots, \Delta_d$ which satisfy (2.8).

To illustrate how the combined algorithm, laid out in the Sections 2, 3 and 5 works, we take up the challenging example of Section 4.3 again to return highly accurate

results on about 95% of the scattering locations. The result is also compared to another d-dimensional generalization, called `ND-ESPRIT`, which arranges the samples in multilevel Hankel matrices [30].

6. Full scale ISAR illustration. When returning to the example in Section 4.3, we take the radar parameters, the signal-to-noise ratio and the vectors $\Delta_i, i = 1, 2, 3$ as specified there. We collect 30000 samples $F_s = f(s\Delta_1)$ and 30000 samples $F_{si} = f(s\Delta_1 + \Delta_i), i = 2, 3$ along each of the shifts, so a total of 90000 samples in total. These samples are now reorganized as follows for use with the technique described in Section 5.

With the total of 90000 samples we perform the following analyses. For each of the analyses we take $N = 6000, \nu = 2000, \eta = 1500, \mathfrak{n} = 6000$. The remaining parameters for the sub-Nyquist sampling in the direction Δ_1 are:

- $m = 2, \mu = 1, \kappa = 6, \mathfrak{p} = 0.3$;
- $m = 3, \mu = 1, \kappa = 4, \mathfrak{p} = 2/9$;
- $m = 4, \mu = 1, \kappa = 3, \mathfrak{p} = 0.125$.

In each of the above analyses, the sampling in the direction Δ_1 starts with F_0 and continues with $F_m, F_{2m}, \ldots$ The shifted samples, that serve the purpose of repairing the possible sub-Nyquist aliasing effect, start with F_1 and continue with $F_{m+1}, F_{2m+1}, \ldots$ In order to make good use of the samples inbetween, the procedure can be repeated $m - 1$ times with the same m and μ but now starting the sampling, instead of at F_0, at F_1 and then at F_2 and so on till F_{m-1}. In this way a choice of m produces $m\kappa$ estimates for the $\exp(\Phi_i), i = 1, \ldots, \eta$ instead of κ, and thus provides a sound basis for validation since $m\kappa$ is usually sufficiently large.

For the choices above, we have $m\kappa = 12$ for $m = 2, 3, 4$ and so we can take, for instance, $m_\delta = (5/6)m\kappa = 10$. In Figure 10 (left and right), we show how accurate the scattering centers are reconstructed, under SNR = 20 dB noise: with every scattering center in the original data we associate the $\log_{10}$ of the Euclidean distance to the nearest reconstructed scattering center (in meter on the x-axis) and then accumulate these (tally is on the y-axis). The distinction between the two figures is that Figure 10 (left) is the result for $m = 1$ (516 scatterers reliably identified), without the use of the enhancement given in Section 5, and Figure 10 (right) is the result for $m = 4$ (696 scatterers detected and validated).

The improvement from $m = 1$ to $m = 4$ may not seem very impressive at first sight. But note that the accurately reconstructed scattering centers (say $\log_{10}(\cdot) \leq -1$) from $m = 1$, need not be the same as the accurately reconstructed ones from the use of $m = 4$. Therefore the combination of both results, merely joining the 516 reconstructions from $m = 1$ with the 696 reconstructions from $m = 4$, immediately leads to the improved distance graph shown in Figure 11.

Eventually, all runs executed with $m = 1, 2, 3, 4$ can be combined, merely joining all the computed scatterer reconstructions: 516 from $m = 1$, 667 from $m = 2$, 673 from $m = 3$ and 696 from $m = 4$, adding up to 2552 in total, with many of them (almost) duplicates. This then leads to highly acurate results for most of the scatterers. In Figure 12 we see that in this combined output 81% of the scatterers is reconstructed within an error of at most 10 cm and 95% is found within a distance of 30 cm! Only 3 scatterers are not reconstructed within a distance of 1 m. The most inaccurately reconstructed scatterer in Figure 13 is near the engine outlet, where one can note that one reconstruction is slightly off. In Figure 13 the 2552 reconstructions are displayed

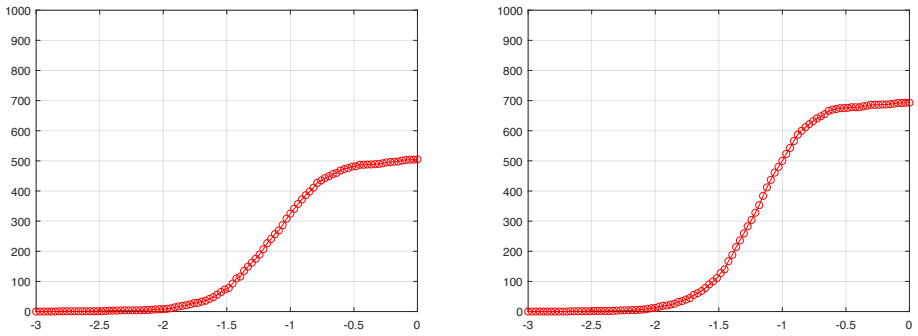

Fig. 10: Accuracy of the reconstructed scatterers for $m = 1$ (left) and $m = 4$ (right) versus the data ($\log_{10}$ of the Euclidean distance)

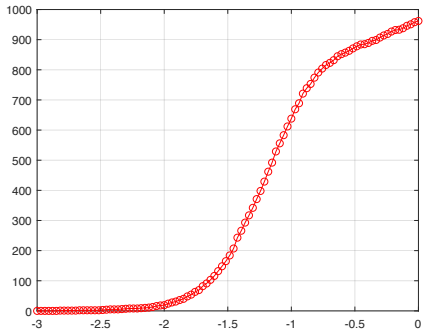

Fig. 11: Accuracy of the reconstructed scatterers ($m = 1$ and $m = 4$ combined) versus the data ($\log_{10}$ of the Euclidean distance)

altogether. Note that, thanks to the validation technique, there are no false results, as also pointed out for Figure 8 where the sub-Nyquist subdivision of the data samples is not yet put to work.

Remains to compare the result to that of a d-dimensional Prony-type algorithm, such as [30] from data laid out in a grid. For instance, a $45 \times 45 \times 45$ grid consists of 91125 samples, which compares nicely to the 90000 samples used in our method. The d-dimensional version considered in [30] starts with the construction of a multilevel Hankel matrix, for which we take 26×20 Hankel blocks on all $d = 3$ levels, thus totalling up to a $26^3 \times 20^3$ or 17576×8000 matrix. A log-plot of its singular values is shown in Figure 14 (left), from which one can deduce that $n \approx 467$ (point of maximal curvature of the plot). With 20 dB noise added to the data, the Euclidean distance log-plot for the 467 reconstructed scatterers is as in Figure 14 (right). This graph somewhat compares to the graphs in Figure 10 but is far from the result displayed in Figure 12 which can be attained with the same sample usage. In Figure 15 we show the actual 467 reconstructed scatterers superimposed on the fighter jet.

One may wonder what the role is, played by the total number of 90000 collected

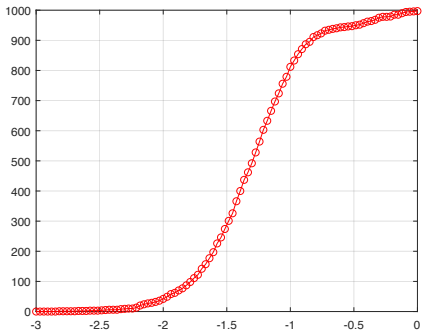

Fig. 12: Accuracy of the reconstructed scatterers ($m = 1, 2, 3, 4$ combined) versus the data ($\log_{10}$ of the Euclidean distance)

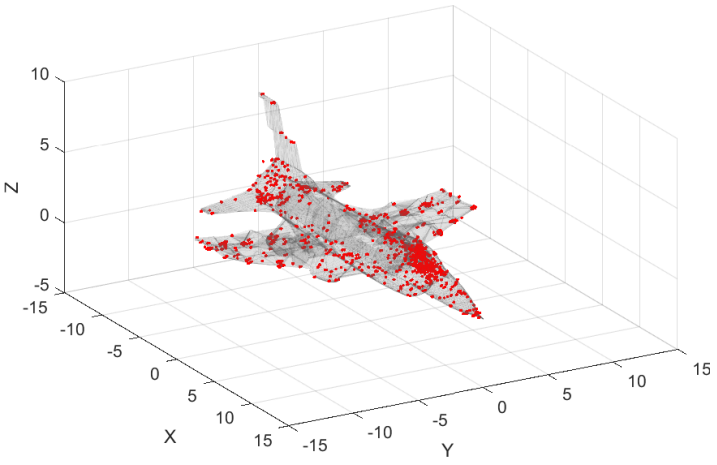

Fig. 13: Fighter jet reconstruction of 934 out of 1000 scatterers

samples for our method, in the Figures 10 to 13. When reducing the sampling from 30000 along each of the 3 directions to 24000, then 71% of the scatterers is found within a distance of 10 cm and 93% within 30 cm. When increasing the sampling from three times 30000 to three times 60000, then as expected, the reconstruction improves, namely 94% is found within 10 cm and 98% within 30 cm.

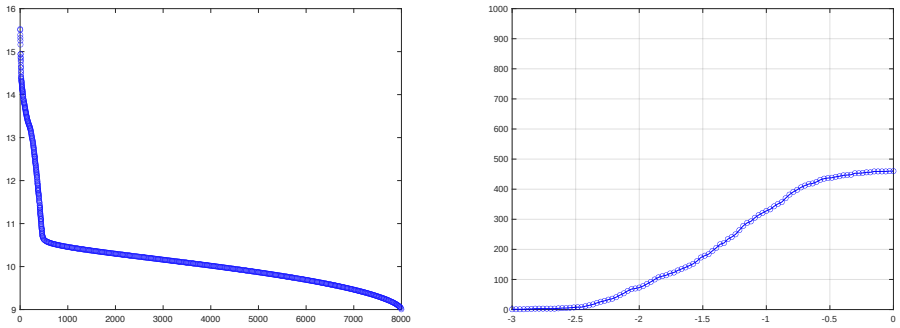

Fig. 14: Singular values of the multilevel Hankel matrix (left log-plot) and distance error of the reconstructed scatterers (right log-plot) using [30]

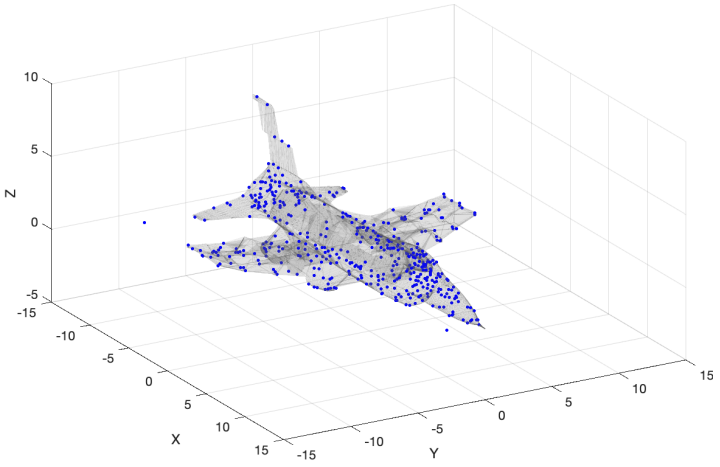

Fig. 15: Fighter jet reconstruction of 467 scatterers using the method in [30]

j	x_j	y_j	z_j	β_j	j	x_j	y_j	z_j	β_j
1	-9.25	0.77	-9.10	53.40	16	-0.59	-8.07	2.44	38.80
2	-8.51	0.77	-7.34	52.50	17	1.93	9.00	4.18	37.6
3	-7.75	0.77	-5.65	51.30	18	1.93	7.22	4.18	36.90
4	-5.15	2.99	-3.19	50.60	19	1.93	-8.07	4.18	35.70
5	-5.15	-2.55	-3.19	49.30	20	1.93	-6.22	4.18	34.90
6	-4.31	4.33	-1.85	48.20	21	2.99	0.77	4.18	33.80
7	-4.31	-4.11	-1.85	47.50	22	4.46	0.77	5.50	32.60
8	-2.61	7.22	0.59	46.30	23	5.87	2.99	6.55	31.50
9	-2.61	-6.22	0.59	45.70	24	5.87	-2.55	6.55	30.80
10	-0.59	9.00	2.44	44.40	25	7.02	2.99	7.52	29.90
11	-0.59	4.33	2.44	43.20	26	7.02	-2.55	7.52	28.70
12	-0.59	2.99	2.44	42.00	27	8.67	2.99	8.42	27.60
13	-0.59	-0.80	2.44	41.20	28	8.67	0.77	8.42	26.50
14	-0.59	-2.55	2.44	40.50	29	8.67	-2.55	8.42	25.10
15	-0.59	-4.11	2.44	39.60					

Table 1: 29-term 3-dimensional ISAR problem.

j	x_j	y_j	β_j
1	0.00	4.00	50.00
2	1.00	3.50	50.00
3	2.00	5.00	50.00
4	2.00	4.00	50.00
5	2.00	3.00	50.00
6	2.50	2.00	50.00
7	2.50	1.00	50.00
8	3.00	5.00	50.00
9	3.00	4.00	50.00
10	3.00	3.00	50.00
11	4.00	3.50	50.00
12	5.00	4.00	50.00

Table 2: 12-term 2-dimensional ISAR problem.

Acknowledgments. The authors are indebted to the colleagues from [37] for emailing them the scattering center data of the fighter jet used in the examples.

REFERENCES

[1] A. Bacci, E. Giusti, D. Cataldo, S. Tomei, and M. Martorella, *ISAR resolution enhancement via compressive sensing: A comparison with state of the art SR techniques*, in 2016 4th International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa), IEEE, 2016, pp. 227–231.

[2] Z. Bajzer, A. C. Myers, S. S. Sedarous, and F. G. Prendergast, *Padé-Laplace method for analysis of fluorescence intensity decay*, Biophys. J., 56 (1989), pp. 79–93.

[3] P. Barone, *On the distribution of poles of Padé approximants to the Z-transform of complex Gaussian white noise*, Journal of Approximation Theory, 132 (2005), pp. 224–240, https://doi.org/10.1016/j.jat.2004.10.014.

[4] M. Briani, A. Cuyt, and W. Lee, *Validated exponential analysis for harmonic sounds*, in DAFX17, 20th international conference on digital audio effects, vol. 20, Edinburgh, United Kingdom, 2017, pp. 222–227.

[5] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, *Sensitivity to basis mismatch in compressed sensing*, IEEE Transactions on Signal Processing, 59 (2011), pp. 2182–2195.

[6] L.-S. Choi and H.-T. Kim, *One-dimensional evolutionary programming-based CLEAN*, Electronics Letters, 37 (2001), pp. 400–401.

[7] A. Cuyt and W.-s. Lee, *Sparse interpolation and rational approximation*, vol. 661 of Contemporary Mathematics, Providence, RI, 2016, American Mathematical Society, pp. 229–242, https://doi.org/10.1090/conm/661/13284.

[8] A. Cuyt and W.-s. Lee, *An analog Chinese Remainder Theorem*, tech. report, Universiteit Antwerpen, 2017.

[9] A. Cuyt and W.-s. Lee, *How to get high resolution results from sparse and coarsely sampled data*, Appl. Comput. Harmon. Anal., (2018), https://doi.org/10.1016/j.acha.2018.10.001. To appear.

[10] A. Cuyt and W.-s. Lee, *Multivariate exponential analysis from the minimal number of samples*, Adv. Comput. Math., 44 (2018), pp. 987–1002, https://doi.org/10.1007/s10444-017-9570-8.

[11] B. Diederichs and A. Iske, *Parameter estimation for bivariate exponential sums*, in IEEE International Conference Sampling Theory and Applications (SampTA2015), 2015, pp. 493–497, https://doi.org/10.1109/SAMPTA.2015.7148940.

[12] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, *A density-based algorithm for discovering clusters in large spatial databases with noise*, in KDD'96 Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD-96, AAAI Press, 1996, pp. 226–231.

[13] J. Gammel, *Effect of random errors (noise) in the terms of a power series on the convergence*

of the Padé approximants, in Padé approximants, P. Graves-Morris, ed., 1972, pp. 132–133.
[14] J. Gilewicz and M. Pindor, *Padé approximants and noise: a case of geometric series*, J. Comput. Appl. Math., 87 (1997), pp. 199–214, https://doi.org/10.1016/S0377-0427(97)00185-4.
[15] J. Gilewicz and M. Pindor, *Padé approximants and noise: rational functions*, J. Comput. Appl. Math., 105 (1999), pp. 285–297, https://doi.org/10.1016/S0377-0427(99)00041-2.
[16] Y. Hua, *Estimating Two-dimensional Frequencies by Matrix Enhancement and Matrix Pencil*, IEEE Transactions on Signal Processing, 40 (1992), pp. 2267–2280.
[17] Y. Hua and T. K. Sarkar, *Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise*, IEEE Trans. Acoust. Speech Signal Process., 38 (1990), pp. 814–824, https://doi.org/10.1109/29.56027.
[18] S. Kunis, T. Peter, T. Römer, and U. von der Ohe, *A multivariate generalization of Prony's method*, Linear Algebra and its Applications, 490 (2016), pp. 31–47.
[19] Q. Li, E. J. Rothwell, K.-M. Chen, and D. P. Nyquist, *Scattering center analysis of radar targets using fitting scheme and genetic algorithm*, IEEE Transactions on Antennas and Propagation, 44 (1996), pp. 198–207.
[20] Q. Liu, A. Liu, Y. Wang, and H. Li, *A super-resolution sparse aperture ISAR sensors imaging algorithm via the MUSIC technique*, IEEE Transactions on Geoscience and Remote Sensing, 57 (2019), pp. 7119–7134.
[21] J. Nuttall, *The convergence of Padé approximants of meromorphic functions*, J. Math. Anal. Appl., 31 (1970), pp. 147–153, https://doi.org/10.1016/0022-247X(70)90126-5.
[22] H. Nyquist, *Certain topics in telegraph transmission theory*, Trans. Am. Inst. Electr. Eng., 47 (1928), pp. 617–644, https://doi.org/10.1109/T-AIEE.1928.5055024.
[23] L. Perotti, T. Regimbau, D. Vrinceanu, and D. Bessis, *Identification of gravitational-wave bursts in high noise using Padé filtering*, Physical Review D, 90 (2014), p. 124047.
[24] T. Peter, G. Plonka, and R. Schaback, *Prony's method for multivariate signals*, in PAMM Proc. Appl. Math. Mech., vol. 15, 2015, pp. 664–666.
[25] C. Pommerenke, *Padé approximants and convergence in capacity*, J. Math. Anal. Appl., 41 (1973), pp. 775–780, https://doi.org/10.1016/0022-247X(73)90248-5.
[26] D. Potts and M. Tasche, *Parameter estimation for multivariate exponential sums*, Electronic Transactions on Numerical Analysis, 40 (2013), pp. 204–224.
[27] A. Quinquis, E. Radoi, and F.-C. Totir, *Some radar imagery results using superresolution techniques*, IEEE Transactions on Antennas and Propagation, 52 (2004), pp. 1230–1244.
[28] S. Rouquette and M. Najim, *Estimation of frequencies and damping factors by two-dimensional ESPRIT type methods*, IEEE Transactions on Signal Processing, 49 (2001), pp. 237–245.
[29] R. Roy and T. Kailath, *ESPRIT-estimation of signal parameters via rotational invariance techniques*, IEEE Trans. Acoust., Speech, Signal Process., 37 (1989), pp. 984–995, https://doi.org/10.1109/29.32276.
[30] S. Sahnoun, K. Usevich, and P. Comon, *Multidimensional ESPRIT for Damped and Undamped Signals: Algorithm, Computations, and Perturbation Analysis*, IEEE Transactions on Signal Processing, 65 (2017), pp. 5897–5910, https://doi.org/10.1109/TSP.2017.2736512.
[31] T. Sauer, *Prony's method in several variables: symbolic solutions by universal interpolation*, Journal of Symbolic Computation, 84 (2018), pp. 95–112.
[32] C. E. Shannon, *Communication in the presence of noise*, Proc. IRE, 37 (1949), pp. 10–21.
[33] L. Weiss and R. McDonough, *Prony's method, Z-transforms, and Padé approximation*, SIAM Rev., 5 (1963), pp. 145–149.
[34] M. Xing, R. Wu, Y. Li, and Z. Bao, *New ISAR imaging algorithm based on modified Wigner–Ville distribution*, IET radar, sonar & navigation, 3 (2009), pp. 70–80.
[35] N. Yilmazer, S. Ari, and T. K. Sarkar, *Multiple snapshot direct data domain approach and ESPRIT method for direction of arrival estimation*, Digital Signal Processing, 18 (2008), pp. 561–567, https://doi.org/10.1016/j.dsp.2007.07.004.
[36] L. Zhang, Z. Qiao, M. Xing, J. Sheng, R. Guo, and Z. Bao, *High-resolution ISAR imaging by exploiting sparse apertures*, IEEE Transactions on Antennas and Propagation, 60 (2012), pp. 997–1008.
[37] J. Zhao, M. Zhang, X. Wang, Z. Cai, and D. Nie, *Three-dimensional super resolution ISAR imaging based on 2D unitary ESPRIT scattering centre extraction technique*, IET Radar, Sonar & Navigation, 11 (2017), pp. 98–106.