VEXPA : Validated EXPonential Analysis through regular sub-sampling

VEXPA: VALIDATED EXPONENTIAL ANALYSIS THROUGH REGULAR SUB-SAMPLING

MATTEO BRIANI*, ANNIE CUYT AND WEN-SHIN LEE

Department Wis-Inf, Universiteit Antwerpen (CMI)
Middelheimlaan 1, 2020 Antwerpen, Belgium

(*) This research is supported by the Instituut voor Weteschap en Technology - IWT

Abstract. In signal processing data are traditionally sampled according to the Shannon-Nyquist theorem in order to prevent aliasing effects. Here we focus on parametric methods and introduce a procedure that allows these methods to work with sub-sampled data. We actually make use of the aliasing effect to regularize the problem statement rather than that we avoid it.

The new approach adds a number of features to a standard exponential analysis, among which output validation, the automatic detection of the exponential model order, robustness against outliers, and the possibilty to parallellize the analysis.

In Section 2 the standard exponential analysis is described, including a sensitivity analysis. In Section 3 the ingredients for the new technique are elaborated, of which good use is made in Section 4 where we essentially bring everything together in what we call VEXPA.

Some numerical examples of the new procedure in Section 5 illustrate that the additional features are indeed realized and that VEXPA is a valuable add-on to any stand-alone exponential analysis. While returning a lot of additional output, it maintains the comparison to the CRLB of the underlying method, for which we here choose ESPRIT.

1. Introduction

Many real-time experiments involve the measurement of signals which fall exponentially with time. The task is then to determine from these measurements the number of terms n and the value of all the parameters in the exponentially damped model

$$\phi(t) = \sum_{i=1}^{n} \alpha_i \exp(\mu_i t), \qquad \alpha_i, \mu_i \in \mathbb{C}. \tag{1}$$

In general, parametric methods as well as the discrete Fourier transform (DFT), sample at a rate dictated by the Shannon-Nyquist theorem [19, 23], which states that the sampling rate needs to be at least twice the maximum bandwitdh of the signal. A coarser time grid than dictated by the theory of Nyquist and Shannon causes aliasing, mapping higher frequencies to lower ones in the analysis. We present a parametric method that samples at a rate below the Shannon-Nyquist one, while maintaining a regular sampling scheme. The new technique is actually exploiting

E-mail address: {matteo.briani,annie.cuyt,wen-shin.lee}@uantwerpen.be.

Key words and phrases. Exponential analysis, sub-Nyquist sampling, uniform sampling, noise handling, Padé-Laplace, Froissart doublets.

aliasing to regularize the problem statement rather than avoiding it. The latter is a useful feature as parametric methods may be more sensitive to noise. On the other hand they have a far superior frequency resolution.

As a consequence of the lower sampling rate it is possible to perform several independent analyses over the original set of samples, each analysis starting from a decimated dataset. If desired, these analyses can be carried out in parallel, thus improving the runing time of the parametric method. The independent solutions are then passed to a cluster detection algorithm in order to add a validation step to the parametric method used, a feature that is lacking in most existing implementations. Thanks to the possibility to work with lower sampling rates, the validation is not at the expense of additional samples.

Making use of the link between Prony-based algorithms and Padé approximation, we are able to separate the added noise from the actual signal and avoid the computation of bogus terms in case of a low signal to noise ratio. In addition, the proposed method detects the number of components n automatically. The latter is a nice side result of working with independent decimations of the given signal data.

Each decimated set of samples is now subject to an independent realization of the noise. While an outlier may skew a single analysis, independent decimations indicate the presence of an outlier. The cluster analysis can eliminate the effect of outliers on the output, which is another desirable feature.

2. The multi-exponential model

Exponential analysis is an inverse problem and may therefore be somewhat sensitive to noise. Besides recalling the basic theory and its connections to some other topics, we also discuss its susceptibility to noise.

2.1. **Exponential analysis.** Let $\phi(t)$ be a sum of complex exponentials with ψ_i, ω_i, β_i and γ_i respectively denoting the damping, frequency, amplitude and phase in each component of the signal $\phi(t)$:

$$\phi(t) = \sum_{i=1}^{n} \alpha_i \exp(\mu_i t),$$

$$\mathrm{i}^2 = -1, \qquad \alpha_i = \beta_i e^{\mathrm{i}\gamma_i}, \qquad \mu_i = \psi_i + \mathrm{i}\omega_i. \tag{2}$$

We sample the function $\phi(t)$ at points $t_j = j\Delta$ for $j = 0, \ldots, 2n-1, \ldots, N-1$ and we set $\Omega = 1/\Delta$. Furthermore, we assume that the frequency content $\omega_i, i = 1, \ldots, n$ in $\phi(t)$ is limited by

$$|\Im(\mu_i)/(2\pi)| = |\omega_i/(2\pi)| < \Omega/2 \quad i = 1, \ldots, n. \tag{3}$$

The aim is to extract the model order n, the parameters $\mu_1, \ldots, \mu_n$ and $\alpha_1, \ldots, \alpha_n$ from the observations $\phi(t_0), \ldots, \phi(t_{2n-1}), \phi(t_{2n}), \ldots$, of $\phi(t)$. When the data are noisefree, the $2n$ parameters α_i and μ_i can be extracted from $2n$ consecutive samples. In order to confirm or reveal the value of n at least one more sample is required. In a noisy context preferably more than the minimal number of samples is provided.

In the sequel we write

$$\phi_j := \phi(t_j), \qquad j = 0, \ldots, N-1,$$
$$\lambda_i := \exp(\mu_i \Delta), \qquad i = 1, \ldots, n,$$

and for integer values s and u, we denote by

$$ {}^{s}_{u}H_n := \begin{pmatrix} \phi_s & \cdots & \phi_{s+(n-1)u} \\ \vdots & \ddots & \vdots \\ \phi_{s+(n-1)u} & \cdots & \phi_{s+(2n-2)u} \end{pmatrix}, \qquad s \geq 0, u \geq 1, \tag{4} $$

the square Hankel matrix of size n constructed from the samples ϕ_j. The left subscript u and left superscript s are respectively called the *undersampling* and the *shift* parameters. Whenever attached to the left of a mathematical notation in the sequel, they need to be interpreted as such.

Note that the Hankel matrices can be decomposed as

$$ {}^{0}_{1}H_n = V_n A_n V_n^T, \qquad {}^{1}_{1}H_n = V_n \Lambda_n A_n V_n^T, $$

$$ V_n = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_n \\ \vdots & \vdots & & \vdots \\ \lambda_1^{n-1} & \lambda_2^{n-1} & \cdots & \lambda_n^{n-1} \end{pmatrix}, \qquad \begin{aligned} A_n &= \operatorname{diag}(\alpha_1, \ldots, \alpha_n), \\ \Lambda_n &= \operatorname{diag}(\lambda_1, \ldots, \lambda_n). \end{aligned} $$

In the standard case $u = 1$ and $s = 0$ or 1. Then the model order n, the coefficients α_i and the parameters μ_i are retrieved from the samples ϕ_j using a variant of Prony's method [22, 15]. Prony's method consists of two stages: first the parameters λ_i are retrieved from which the μ_i can be extracted because of (3), and then the α_i are computed from a linear system of equations. Often the λ_i are obtained from the generalized eigenvalue problem

$$ ({}^{1}_{1}H_n)v = \lambda({}^{0}_{1}H_n)v. \tag{5} $$

Subsequently the α_i are computed from the interpolation conditions

$$ \sum_{i=1}^{n} \exp(\mu_i t_j) = \phi_j, \qquad j = 0, \ldots, 2n-1, \ldots, \tag{6} $$

either by solving the system in the least squares sense, in the presence of noise, or by solving a subset of n interpolation conditions in case of a noisefree $\phi(t)$. Note that $\exp(\mu_i t_j) = \lambda_i^j$ and that the coefficient matrix of (6) is therefore a Vandermonde matrix. In a noisy context the Hankel matrices in (5) can also be extended to rectangular matrices and the generalized eigenvalue problem can be considered in a least squares sense [4].

Condition (3) guarantees that the μ_i can be extracted from the λ_i without ambiguity. However, when $|\omega_i/(2\pi)| \geq \Omega/2$, then ω_i is identified with a smaller frequency, an effect known as *aliasing*. In this case the computed λ_i represents an entire set of possible μ_i. How to solve the aliasing problem is addressed in [7, 8] and recalled in Section 3.

What can be said about the number of terms n in (2), which is also called the *sparsity*? From [14, p. 603] and [16] we know that

$$ \begin{aligned} &\det {}^{0}_{1}H_\nu = 0 \text{ accidentally}, \qquad \nu < n, \\ &\det {}^{0}_{1}H_n \neq 0, \\ &\det {}^{0}_{1}H_\nu = 0, \qquad \nu > n. \end{aligned} $$

A standard approach to make use of this statement is to compute a singular value decomposition of the Hankel matrix ${}^{0}_{1}H_\nu$ and this for increasing values of $\nu > n$. In the presence of noise and/or very similar eigenvalues, this technique is not always

reliable and we need to consider rather large values of ν for a correct estimate of n [5]. The method proposed in Section 4 allows to automatically detect n while processing the samples ϕ_j without having to resort to a separate singular value decomposition of ${}_1^0H_\nu$.

2.2. **The Padé and Froissart connections.** There is an interesting but somewhat unknown connection between Padé approximation, Froissart doublets and the Prony problem, which we briefly recall from [1, 5]. Consider the function $f(z)$ defined by

$$f(z) = \sum_{j=0}^{\infty} \phi_j z^j.$$

For ϕ_j given by (2) we can write

$$f(z) = \sum_{i=1}^{n} \frac{\alpha_i}{1 - \lambda_i z}. \tag{7}$$

The partial fraction decomposition (7) is related to both the Laplace transform and the Z-transform of (2) as described in [1]. It is a rational function of degree $n-1$ in the numerator and degree n in the denominator with poles $1/\lambda_i$. Now let us perturb $f(z)$ with white circular Gaussian noise to obtain

$$f(z) + \epsilon(z) = \sum_{j=0}^{\infty} (\phi_j + \epsilon_j) z^j.$$

The theorem of Nuttall-Pommerenke states that if $f(z) + \epsilon(z)$ is analytic throughout the complex plane, except for a countable number of poles [18] and essential singularities [21], then its sequence of Padé approximants $\{r_{\nu-1,\nu}(z)\}_{\nu\in\mathbb{N}}$ of degree $\nu - 1$ over ν converges to $f(z) + \epsilon(z)$ in measure on compact sets. This means that for sufficiently large ν the measure of the set where the convergence is disrupted, so where $|f(z) + \epsilon(z) - r_{\nu-1,\nu}(z)| \geq \tau$ for some given threshold τ, tends to zero as ν tends to infinity. Pointwise convergence is disrupted by $\nu - n$ unwanted pole-zero combinations of the Padé approximants that are added to the n true poles and $n-1$ true zeros of $f(z)$ [10, 12], the pole and zero in the pair almost cancelling each other locally. These pole-zero combinations are also referred to as Froissart doublets. In practice, these Froissart doublets offer a way to separate the noise $\epsilon(z)$ from the underlying $f(z)$. Because of the Padé convergence theorem, the true (physical) poles can be identified as stable poles in successive $r_{\nu-1,\nu}(z)$, while the spurious (noisy) poles are distinguished by their instability. When increasing ν we compute a larger set of poles, of which the noisy ones are moving around in the neighbourhood of the complex unit circle [11, 12] with every different realization of the noise $\epsilon(z)$. The latter is illustrated in Figure 1 where we show the results of the analysis of a test signal perturbed by a large number of independent noise realizations: the true λ_i are forming clusters while the ones related to noise are scattered around [2, 20]. In addition, around each λ_i-cluster one empirically finds 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 $\nu > n$ generalized eigenvalues λ_i, we discard the unstable ones and focus on the stable ones. We now describe in more detail the precise influence of noise in the data ϕ_j on the λ_i.

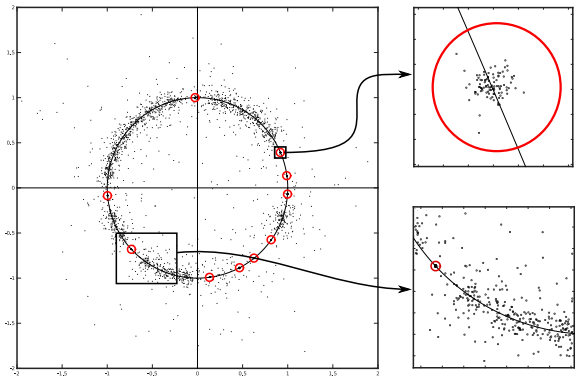

FIGURE 1. *Analysis results* ($n = 10, \nu = 30$) *of test signal* $\phi(t)$ *after several perturbations* $\epsilon(z)$*: the true* λ_i *are drawn as red circles.*

2.3. **Sensitivity to noise.** The exponential analysis of $\phi(t)$, being an inverse problem, is known to be sensitive to noise. Here we briefly recall what is known and in the next section we explain how the new method is able to deal with outliers on the one hand and normally distributed noise on the other.

In [13] the authors explain that the roundoff errors in the computation of the generalized eigenvalues are amplified by mainly three sources:

- the scaling of the problem (the λ_i should lie as closely as possible to the complex unit circle),
- the size of the $|\alpha_i|$ relative to the noise (λ_i with smaller amplitude are more challenging to retrieve),
- the relative position of the λ_i with respect to each other.

The first problem is addressed in [13] by means of a diagonal preconditioning matrix, and in [3] using a suitably chosen invertible upper triangular matrix. The second problem can be tackled with the use of linear time invariant filters which preserve model (2). A solution for the third problem is proposed in [7, 8] and accomplishes a redistribution of the λ_i. Our new method is based on this approach. We now briefly recall the basics of the analysis in [3] to understand the effect of noise and how this is related to the method presented in [7, 8].

Let $(\epsilon_0, \ldots, \epsilon_{2n-1}, \ldots, \epsilon_{N-1})$ again denote the noise vector added to the samples $(\phi_0, \ldots, \phi_{2n-1}, \ldots, \phi_{N-1})$. We express the noise terms ϵ_j as $\epsilon_j = \epsilon e_j$ where the square Hankel matrices ${}_1^0E_n$ and ${}_1^1E_n$ of size n, filled as in (4) but now with the e_j instead of the ϕ_j, satisfy

$$||{}_1^0E_n||_2 \leq 1, \qquad ||{}_1^0E_n||_2 \leq 1.$$

Let $L_i(\lambda)$ denote the Lagrange basis polynomial of degree n with roots $\lambda_1, \ldots, \lambda_{i-1}, \lambda_{i+1}, \ldots, \lambda_n$ and $L_i(\lambda_i) = 1$. The coefficients of $L_i(\lambda)$ make up the vector ℓ_i of size $n+1$. When the samples ϕ_j are perturbed by the noise terms ϵe_j, then the computed generalized eigenvalues also depend on the noise magnitude ϵ and so we can write $\lambda_i(\epsilon)$ and consider them as functions of ϵ (for fixed ϵ we continue to use the notation λ_i). Then the *disposedness* of the generalized eigenvalue λ_i is measured by

$$\rho_i := \left| \frac{d\lambda_i}{d\epsilon}(0) \right|$$

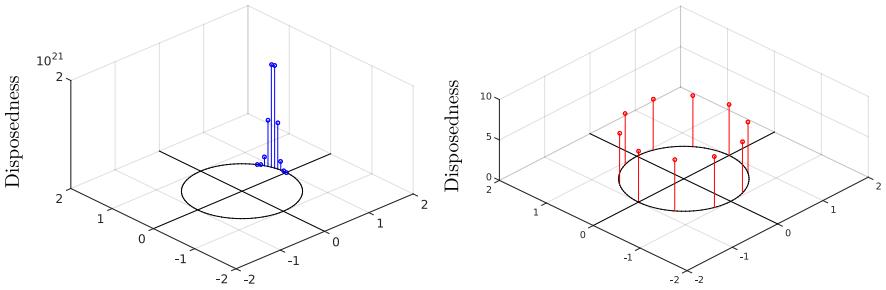

FIGURE 2. *Ill-disposed* ${}_1\lambda_i = \exp(\mathrm{i}2\pi(i-1)/100)$ *at the left and well-disposed* ${}_{10}\lambda_i = \exp(\mathrm{i}2\pi(i-1)/10)$ *at the right,* $i = 1, \ldots, 10$.

and satisfies

$$\rho_i \leq \frac{|\lambda_i|+1}{|\alpha_i|}\ ||\ell_i||_2^2\ (||{}_1^1H_n||_2 + ||{}_1^0H_n||_2). \tag{8}$$

A generalized eigenvalue λ_i is ill-disposed when ρ_i is large. Larger ρ_i imply higher susceptibility to noise. Besides the Froissart phenomenon described earlier, the disposedness ρ_i of the generalized eigenvalues is another tool to use when inspecting the λ_i. In Figure 2 we illustrate the relationship between the ρ_i and the relative position of the λ_i with respect to each other. For our toy problem we choose $\Omega = 100, n = 10, \alpha_i = 1, \mu_i = \mathrm{i}2\pi(i-1)$. At the left the values ρ_i are plotted at the locations of the generalized eigenvalues $\lambda_i = \exp(\mu_i/\Omega), i = 1, \ldots, 10$. When changing the undersampling parameter u in ${}_u^sH_n$ in (4) and (8) from $u = 1$ to $u = 10$ and recomputing the generalized eigenvalues $\exp(10\mu_i/\Omega)$ and the disposedness, which we now denote by ${}_u\rho_i$, the result, which is shown at the right, changes dramatically. Actually, taking $u > 1$ is equivalent to replacing Δ by $u\Delta$ or replacing Ω by Ω/u.

Another important tool for inspecting the λ_i is the Cramèr-Rao lower bound (CRLB) [17]. For any given unbiased estimator of the parameters in (2) and a specific amount and type of noise, the CRLB returns the minimal variance that the estimator suffers. In our case, the estimator is any implementation of Prony's method and the type of noise is white circular Gaussian noise. The CRLB depends on the number of samples N, the variance and type of noise and the set of parameters $\beta_i, \gamma_i, \psi_i$ and ω_i. The bound is often used to compare the variance of a specific estimator to this theoretical lower bound. The closer an estimator is to the CRLB, the more efficient it is said to be. Another way to use the CRLB is for the comparison of different theoretical bounds for the same amount of noise but different parameters $N, \beta_i, \gamma_i, \psi_i$ and $\omega_i, i = 1, \ldots, n$.

We consider the practical computation of the CRLB provided in [24] and illustrate the relationship between the CRLB and the disposedness ρ_i of $\lambda_i, i = 1, \ldots, n$. Take the same toy example and add white circular Gaussian noise of varying signal to noise ratio (SNR). In Figure 3 we graph the root mean square of the vector of CRLB's for the parameters $\omega_i, i = 1, \ldots, 10$, and this for decreasing SNR in three different situations:

- $\Delta = 1/\Omega, N = 200$ samples ϕ_j (blue triangles),
- $\Delta = 10/\Omega, N = 200$ samples ϕ_j (green squares),
- $\Delta = 10/\Omega, N = 20$ samples ϕ_j (red circles).

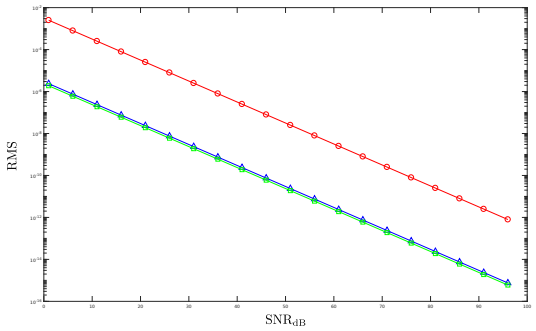

FIGURE 3. *Root mean square of the CRLB vector of the* $\omega_i, i = 1, \ldots, 10$*, respectively for* $\Omega = 100, N = 200$ *(blue),* $\Omega = 10, N = 200$ *(green),* $\Omega = 10, N = 20$ *(red).*

Note that multiplying Δ by $u = 10$ while maintaining $N = 150$ implies that the signal is sampled over a larger time interval, while multiplying Δ by $u = 10$ and dividing N by $u = 10$ does not enlarge the observation window.

3. Recovering from aliasing after decimation

So we know that choosing $u > 1$ may positively impact the disposedness of the λ_i, without negatively impacting the CRLB if the total number of samples can approximately be maintained. Since introducing u impacts Δ or Ω, aliasing may occur when (3) is violated. We now explain how to deal with this effect: the goal is to enjoy the positive influence of a larger u without suffering the aliasing effect introduced by it.

3.1. **Decimation.** Instead of using the consecutive set of samples $\phi_j, j = 0, \ldots, 2n-1, \ldots$, we consider the decimated set ϕ_{uj} which is obtained by considering one sample every u samples, thus sampling $\phi(t)$ at $t_{ju} = j(u\Delta)$. The square generalized eigenvalue problem

$$({}^{u}_{u}H_n)v = \lambda({}^{0}_{u}H_n)v,$$

leads to a new set of generalized eigenvalues

$${}_u\lambda_i := \exp(\mu_i u\Delta) = \lambda_i^u, \qquad i = 1, \ldots, n.$$

From ${}_u\lambda_i$ we cannot directly retrieve λ_i, due to the disruption of (3). We are left with a set of possible values for λ_i given by

$$U_i := \left\{ \exp\left(\mu_i\Delta + \frac{2\pi\mathrm{i}}{u}\ell\right), \ell = 0, \ldots, u-1 \right\}.$$

Despite this, we can already compute the coefficients α_i by solving the linear system

$$\phi_{uj} = \sum_{i=1}^{n} \alpha_i ({}_u\lambda_i)^j, \qquad j = 0, \ldots, 2n-1, \ldots \tag{9}$$

Now we consider a shifted set of samples ϕ_{s+uj} consisting of at least n samples, for instance at $j = k, \ldots, k+n-1, 0 \le k \le n$, and we choose s coprime with u. Since

$$\phi_{s+uj} = \sum_{i=1}^{n} (\alpha_i \lambda_i^s)({}_u\lambda_i)^j, \qquad j = k, \ldots, k+n-1, \tag{10}$$

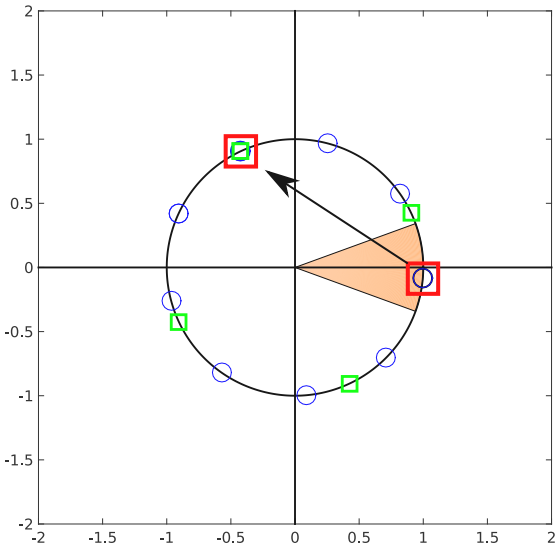

FIGURE 4. *Intersection of U_i (blue circles, $u = 9$) and S_i (green squares, $s = 4$), relocating the aliased ${}_u\lambda_i$ (red square).*

we denote the coefficient of $({}_u\lambda_i)^j$ in the shifted sample ϕ_{s+uj} by

$$ {}^s\alpha_i := \alpha_i \lambda_i^s, \qquad i = 1, \ldots, n. $$

We can solve the interpolation conditions (10) for the second set of coefficients ${}^s\alpha_i$. Note that the linear systems (10) and (9) have the same coefficient matrix. From α_i and ${}^s\alpha_i$ we obtain

$$ {}^s\alpha_i / \alpha_i = \lambda_i^s, $$

which we can denote by ${}^s\lambda_i$. Due to the same disruption of condition (3), ${}^s\lambda_i$ also stands for a set of possible values for λ_i, namely

$$ S_i := \left\{ \exp\left(\mu_i \Delta + \frac{2\pi \mathrm{i}}{s} \ell \right), \ell = 0, \ldots, s-1 \right\}. $$

Both sets U_i and S_i contain the solution λ_i. Since u and s are coprime they share one and only one element which is the non-aliased λ_i [6]. In Figure 4 we graphically sketch what happens. There $u = 9$, the elements in U_i are shown using blue circles, $s = 4$, the elements in S_i are shown using green squares and the arrow points to the unique non-aliased λ_i in their intersection. The orange portion is the region where the aliased ${}_u\lambda_i$ lies (red square), from which we have to recover the correct λ_i, the aliasing being the consequence of the decimation of the collected samples by a factor u.

3.2. **Recovery.** While we know theoretically that U_i and S_i have only one element in their intersection, we still need to find a way to compute this element in practice. In [6] the following two options are presented. Here we develop a more robust third approach.

An obvious approach is to compute all distances between elements of U_i and elements of S_i and select the pair that lies closest. This simple approach does not deliver satisfactory results though, because of noise issues. For increasing noise levels, the sets S_i may be too perturbed, thus leading to a wrong match of the candidate values for λ_i.

A less obvious approach is to use the Euclidean algorithm and compute two integers w and r satisfying $wu + rs = 1$ for the coprime u and s. Then λ_i can be retrieved as

$$({}_u\lambda_i)^w \, ({}^s\lambda_i)^r = \exp((wu + rs)\mu_i\Delta) = \lambda_i.$$

The downside of this method is that if w and r are not small, any noise present in ${}_u\lambda_i$ and ${}^s\lambda_i$ is amplified.

We propose to solve a small number of additional systems of the form (10), in order to stabilize the location of the elements in S_i before building the distance matrix. We continue the use of shifted samples:

$$\phi_{ms+uj} = \sum_{i=1}^{n} ({}^{ms}\alpha_i) \, ({}_u\lambda_i)^j, \qquad m = 0, \ldots, M-1. \tag{11}$$

From each shift we compute the coefficients ${}^{ms}\alpha_i$ and we set up the sequence of values

$$\alpha_i, {}^s\alpha_i, \ldots, {}^{ms}\alpha_i, \ldots, {}^{(M-1)s}\alpha_i,$$

satisfying

$${}^{ms}\alpha_i = \alpha_i({}^s\lambda_i)^m = \alpha_i \exp(\mu_i m(s\Delta)), \qquad m = 0, \ldots, M-1, \tag{12}$$

where ${}^0\alpha_i = \alpha_i$. So for fixed i the values ${}^{ms}\alpha_i$ follow the exponential model (12) consisting of only one term. We can therefore use a Prony-like method to extract ${}^s\lambda_i$ from the expressions ${}^{ms}\alpha_i$, just as described in the previous section on basic exponential analysis. This approach stabilizes the location of ${}^s\lambda_i = \lambda_i^s$ by the use of extra estimates.

At this point we want to point out and stress, that the whole procedure of decimation and recovery can be used on top of any Prony-like method. Retrieving ${}_u\lambda_i$, ${}^s\lambda_i$ or ${}^{ms}\alpha_i$ for chosen u and s does not require a specific parametric method. In fact, the current procedure offers a way to parallelize existing Prony-like methods, as the decimated signals can be treated independently of each other. In the next section we explain how the combination of the decimated results adds a validation step to the method, which is mostly lacking in existing Prony-like algorithms.

3.3. **Frequency collision.** A problem that may occur when decimation causes aliasing is the possible collision of frequencies. For instance, two distinct eigenvalues λ_1 and λ_2 may be aliased to the same eigenvalue ${}_u\lambda_1 = {}_u\lambda_2$. However unlikely, we want to discuss how to deal with this situation. We explain the remedy on an example.

Let $\phi(t)$ be specified by $n = 2$, $\alpha_1 = \alpha_2 = 1$, $\mu_1 = 2\pi\mathrm{i}13$, $\mu_2 = 2\pi\mathrm{i}33$. We set $\Omega = 100$ and consider one sample $\phi_j = \phi(j/\Omega)$ every ten samples ($u = 10$) thus changing Ω to be 10. Due to aliasing, λ_1 and λ_2 are mapped to another location in the complex plane. In particular, we have

$${}_u\lambda_1 = {}_u\lambda_2 = \exp\left(\frac{2\pi\mathrm{i}3}{10}\right),$$

because

$$\exp\left(\frac{2\pi\mathrm{i}33}{10}\right) = \exp\left(\frac{2\pi\mathrm{i}13}{10}\right) = \exp\left(\frac{2\pi\mathrm{i}3}{10}\right).$$

So in the decimation step (9) Prony's method retrieves a single frequency with associated coefficient $\alpha_1 + \alpha_2$.

It is however still possible to retrieve the original values λ_1 and λ_2 in the recovery step. As explained, the generalized eigenvalue ${}_u\lambda_1 = {}_u\lambda_2$ stands for a set of values $U_1 = U_2$ that now contains both the correct λ_1 and λ_2. We choose s coprime with

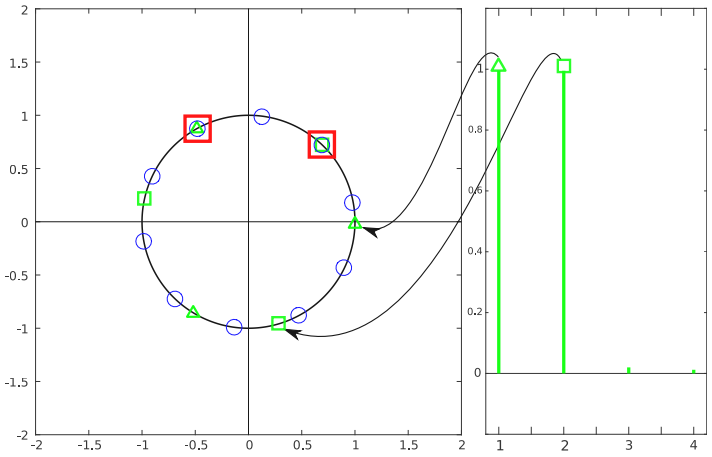

FIGURE 5. *The $|\alpha_1|$ of ${}^s\lambda_1$ (green triangle) and $|\alpha_2|$ of ${}^s\lambda_2$ (green square) at the right ($s = 3, M = 8$), identifying $U_1 \cap S_1 = \{\lambda_1\}$ and $U_2 \cap S_2 = U_1 \cap S_2 = \{\lambda_2\}$ from $u = 10$ at the left (red squares).*

u and compute the values ${}^{ms}\alpha_1$ (remember that the computed ${}^0\alpha_1 = 2$ now equals the sum of the true coefficients). Since s is coprime with u, no frequency collision occurs in ${}^{ms}\alpha_1$ which is following the model

$$ {}^{ms}\alpha_1 = \alpha_1 \exp(\mu_1 ms\Delta) + \alpha_2 \exp(\mu_2 ms\Delta), \qquad m = 0, \ldots, M-1. \tag{13}$$

So in the analysis of (13) Prony's method reveals two contributions ${}^s\lambda_1$ and ${}^s\lambda_2$ which bring forth the sets S_1 and S_2 respectively containing λ_1 and λ_2. The intersections $U_1 \cap S_1$ and $U_2 \cap S_2 = U_1 \cap S_2$ reveal the original λ_1 and λ_2.

Of course the above can also be applied to the more general case of several collisions in a signal $\phi(t)$ containing more terms. The key element is that the value M in (13) is chosen large enough to allow the identification of all the collided eigenvalues. In particular, M should be at least twice the number of collided eigenvalues. Since this number is unknown, the standard procedure is to take M even and fit the ${}^{ms}\alpha_i$ with a model of size $M/2$. If less than $M/2$ frequencies have collided, then some of the terms in the expression for ${}^{ms}\alpha_i$ model the noise and can easily be discarded, as explained in Section 2. We show a typical situation in Figure 5, which applies to the $n = 2$ example above: the set $U_1 = U_2$ is depicted using blue circles ($u = 10$), the sets S_1 and S_2 using green triangles and squares respectively ($s = 3$). We choose $M = 8$. The intersections $U_1 \cap S_1$ and $U_2 \cap S_2$ are indicated using red squares.

4. Validated exponential analysis

A quite robust Prony-like implementation, which approaches the theoretical CRLB (depicted using blue triangles in Figure 3), is for instance found in [22] and is called ESPRIT. Our aim now is to maintain the same accuracy, but add the following features to the implementation by making a detour via decimation:

- validation of the output,
- automatic estimation of the model order n,
- robustness against outliers,
- parallellism in the algorithm.

In other words, while the sub-sampling of a signal usually leads to cruder estimates of the already aliased frequencies (upper CRLB curve in Figure 3), the method explained below still achieves the desired CRLB curve (middle curve in Figure 3), while adding a number of desirable features that become available through the technique described in Section 3.

Given a fixed undersampling parameter u, we can consider u decimated sample sets $\Phi_k, k = 0, \ldots, u-1$, starting respectively at $t_k = t_0, \ldots, t_{u-1}$. The first set contains $\lfloor N/u \rfloor$ samples and all subsequent sets contain either the same number of samples or one less:

$$\Phi_k := \{\phi_{uj+k} : j = 0, \ldots, \min(\lfloor N/u \rfloor, \lfloor (N-k)/u \rfloor) - 1\}, \quad k = 0, \ldots, u-1.$$

From each decimated set Φ_k we extract ${}_u\lambda_i, {}^s\lambda_i, i = 1, \ldots, n$ which should carry a second index k now to indicate from which decimation Φ_k the values was obtained. The same holds for the coefficients α_i. For the sequel we therefore introduce the notations $({}_u\lambda_{i,k}), ({}^s\lambda_{i,k}), {}^{ms}\alpha_{i,k}$ with obvious meanings. We also introduce

$${}_uL := \cup_{i=1,k=0}^{n,u-1} \{{}_u\lambda_{i,k}\},$$

$${}^sL := \cup_{i=1,k=0}^{n,u-1} \{{}^s\lambda_{i,k}\}.$$

We remark that the index i also runs from 1 to n even if the undersampling has caused collisions. Then some ${}_u\lambda_{i,k}$ are merely duplicated.

Each dataset Φ_k is now a decimation of the set of samples $\{\phi_0, \phi_1, \ldots, \phi_{N-1}\}$. From this section on each sample ϕ_j is always perturbed by noise (we choose however to abuse the notation ϕ_j in order to not overload the presentation). Each set Φ_k is subject to an independent realization of the noise because the latter affects each decimated signal in a different and independent way. Thanks to the connection with the theory of Padé approximation and Froissart doublets, we know that the ${}_u\lambda_{i,k}$ and ${}^s\lambda_{i,k}$ form clusters in the sets ${}_uL$ and ${}^sL$ respectively, around the true ${}_u\lambda_i = \lambda_i^u$ and ${}^s\lambda_i = \lambda_i^s$. Any generalized eigenvalues retrieved from overestimating the model order n by $\nu > n$, model the noise and are found scattered around the complex unit disk, as explained in Section 2. To detect the clusters in ${}_uL$ and ${}^sL$ we propose to use the density based cluster algorithm DBSCAN [9].

DBSCAN requires two additional 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, which is useful in case of a higher noise level. Smaller values of δ allow to detect denser clusters, which appear in case of very stable estimates ${}_u\lambda_{i,k}$ or low levels of noise. A value for m_δ smaller than u allows to discard bogus estimates appearing as a consequence of outliers in the data. When m_δ is set equal to u, each ${}_u\lambda_i$ needs to be confirmed by all the decimated analyses. Remember that, through the coefficient matrix shared between (9) and (10), each element from ${}^sL$ is connected to an element in ${}_uL$. So any cluster detected in ${}^sL$ is tied to a set of elements from ${}_uL$ of the same size. We also point out that the introduction of decimation parallellizes the exponential analysis. Each Φ_k is analyzed independently and the computation of the ${}_u\lambda_{i,k}$ and ${}^s\lambda_{i,k}$ does not need data from other decimations. All the results are collected after the individual runs and then passed to the cluster analysis.

Essentially three different DBSCAN scenario's can occur, which are sketched in Figure 6: at the left we find the result of running DBSCAN on the set ${}_uL$ and at the right the result on the set ${}^sL$.

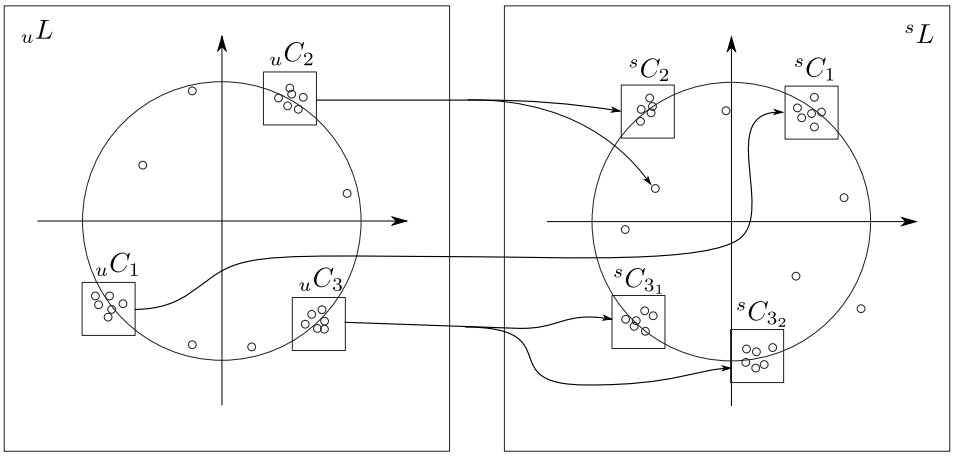

FIGURE 6. *The cluster algorithm as explained in Section 4, with the set ${}_uL$ at the left and the set ${}^sL$ at the right.*

4.1. **Standard scenario.** A cluster ${}_uC_1$ is detected in the set ${}_uL$ and its center of gravity can serve as an estimate of one of the λ_i^u. The elements ${}^s\lambda_{i,k}$ tied to the generalized eigenvalues ${}_u\lambda_{i,k} \in {}_uC_1$ also form a cluster, which we denote by ${}^sC_1$. Its center of gravity then returns an estimate of λ_i^s. From both centers of gravity a reliable estimate of λ_i can be extracted as described in Section 3. With each identified λ_i we can return a list of extra informational items:

- the number of elements validating ${}_u\lambda_i$ in the ${}_uL$ cluster,
- the number of elements validating ${}^s\lambda_i$ in the ${}^sL$ cluster,
- the actual radius of the ${}_uL$ cluster around ${}_u\lambda_i$,
- the actual radius of the ${}^sL$ cluster around ${}^s\lambda_i$,

The cardinality of the ${}_uL$ cluster, which indicates how many decimated analyses succeeded in retrieving λ_i^u, indicates the level of validation of the retrieved λ_i, while that of the ${}^sL$ cluster, in combination with its radius, reflects the correct or poor resolution from the aliasing. The radius of the ${}_uL$ on the one hand and the ${}^sL$ cluster on the other, is a measure of the perturbation suffered by respectively λ_i^u and λ_i^s. Small clusters with large radii indicate that the conclusion may be wrong because of the inherent noise that spread the generalized eigenvalues apart. The total number of clusters detected in ${}^sL$ is automatically a good estimate of the model order n, as can be seen in Figure 6.

4.2. **Outlier scenario.** It may happen that not all elements ${}^s\lambda_{i,k}$ tied to the ${}_u\lambda_{i,k}$ in a detected cluster ${}_uC_2$ belong to a cluster ${}^sC_2$. In that case the outliers in ${}^sL$ are discarded and an estimate for ${}^s\lambda_i = \lambda_i^s$ is still the center of gravity of ${}^sC_2$. Here the number of decimated analyses validating λ_i is different in ${}_uL$ and ${}^sL$.

4.3. **Collision scenario.** In cluster ${}_uC_3$ a collision is involved. As pictured in Figure 5, the ${}^{ms}\alpha_{i,k}$ have identified more than one exponential contribution, as in (13). In ${}^sL$ different clusters are identified instead of one large cluster. The centers of gravity of these individual clusters serve to identify the different generalized eigenvalues that have collided as a consequence of the aliasing.

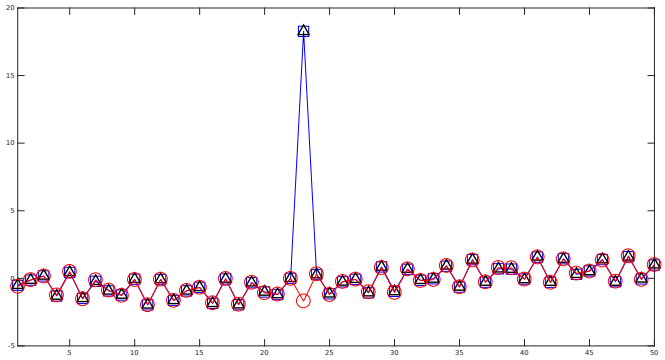

FIGURE 7. *Outlier experiment with original data (black triangles), ESPRIT reconstruction (blue squares) and VEXPA reconstruction (red circles).*

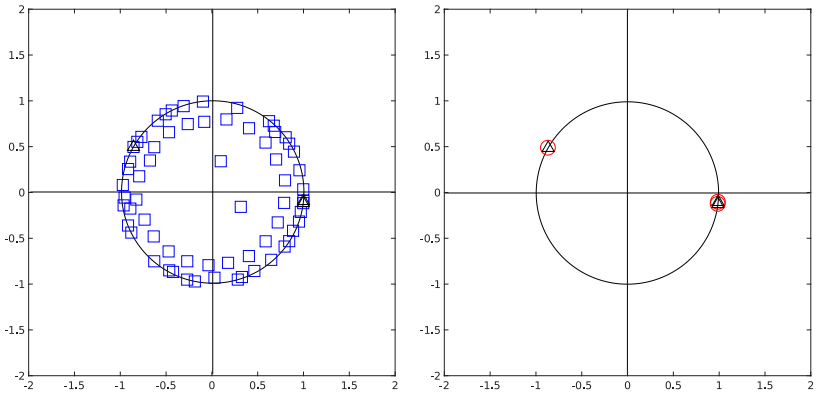

FIGURE 8. *The (ω_i, β_i) output from ESPRIT (left) and VEXPA (right).*

5. Numerical illustration

At this moment we introduce the acronym VEXPA for the new procedure that validates an exponential analysis carried out by a Prony-like method applied to each of the decimated signals. In order to see the proposed method at work, we present the results of two experiments, with the main aim to illustrate the extra features listed in Section 4, which can now be added to whatever Prony-like method used for each separate decimated analysis. For our experiments we use ESPRIT as the method of choice to compute the aliased results ${}_u\lambda_{i,k}$ modelling the data Φ_k and to compute the recovery values ${}^s\lambda_{i,k}$ modelling the ${}^{ms}\alpha_{i,k}$. We then compare the VEXPA results to those of the stand-alone ESPRIT method.

5.1. **Outlier experiment.** We consider $\phi(t)$ defined by the parameters β_i, γ_i, ω_i, ψ_i, $i = 1, 2, 3$ listed in Table 1. In addition we add white circular Gaussian noise with SNR= 30 dB and we create an outlier by perturbing sample ϕ_{23} by adding 20 to it. The total number of samples is $N = 300$. The full signal is analyzed using the Prony-like algorithm ESPRIT [22], of which the result can be found in the Figures 7 (time domain) indicated with blue squares, and 8 (frequency domain) in blue at the left. In Figure 7 the signal, perturbed by noise and an outlier, is depicted using black triangles.

TABLE 1. *Outlier experiment with* $n = 3$ *and* $N = 300$.

β_i	γ_i	ω_i	ψ_i
1	0.3342	$2\pi 417.764$	-0.1
1	0.8084	$-2\pi 15.8$	0
0.5	0.5880	$-2\pi 19.5$	0

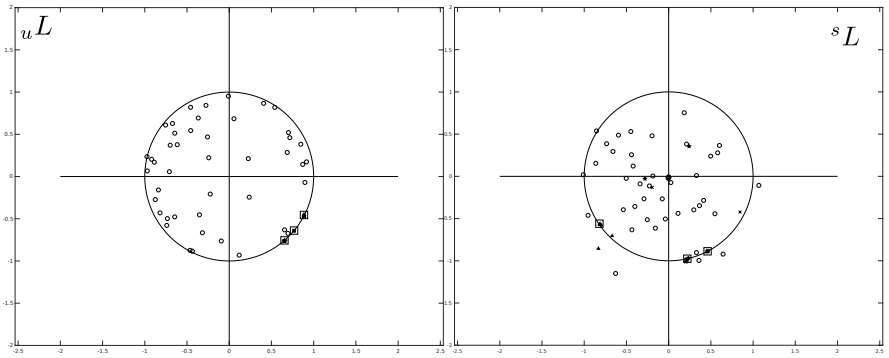

FIGURE 9. *Cluster detection in* $_uL$ *(left) and* ^sL *(right) for the outlier experiment.*

For VEXPA we take $u = 7$ and $s = 11$. So each Φ_k contains 42 or 41 samples. The decimation Φ_2 is the one containing the outlier. So we can expect to find clusters of 6 elements in $_uL$ instead of 7. The values $^{ms}\alpha_{i,k}$ perturbed by the outlier are the ones where $k = 1$ and $k = 5$. So we expect the clusters in ^sL to contain 4 elements (only 6 elements in ^sL are connected to the possible 6 cluster elements in $_uL$ to start with). In Figure 9 we show the results of the DBSCAN cluster analysis on $_uL$ and on ^sL.

The signal reconstructed from the VEXPA output is depicted in Figure 7 using red circles. The spectral output is shown in Figure 8 in red at the right. Both the Figures 7 and 8 illustrate that the stand-alone ESPRIT method suffers from the outlier. The new VEXPA add-on is able to filter out the outlier and reconstruct the original signal because it retrieves the parameters correctly. The ESPRIT implementation introduces, besides the correct frequencies and amplitudes, a lot of additional terms that are hard to discard.

How has the computation of the α_i profited from the cluster analysis as well? Since the clusters in $_uL$ consist of 6 elements, we know which decimation did not contribute to the validation and so we can omit all data points from that decimation in the linear system from which the parameters α_i are computed! So the computation of the α_i starts from already outlier filtered data.

5.2. **High noise experiment.** For our second experiment we consider a signal $\phi(t)$ defined by the parameters $\beta_i, \gamma_i, \omega_i, \psi_i, i = 1, \ldots, 12$ in Table 2. We perturb the samples with white circular Gaussian noise of increasing SNR. The total number of samples is again $N = 300$. The perturbed signal is then analysed using ESPRIT on the one hand and VEXPA (on top of ESPRIT) on the other. For the latter we choose $u = 7$ and $s = 6$. We pass the correct model order n only to ESPRIT. The new VEXPA will detect it automatically.

Up to SNR= 20 dB both ESPRIT and VEXPA closely approach the theoretical CRLB, as can be seen from Figure 10. In Figure 11 we show the retrieved ω_i-values

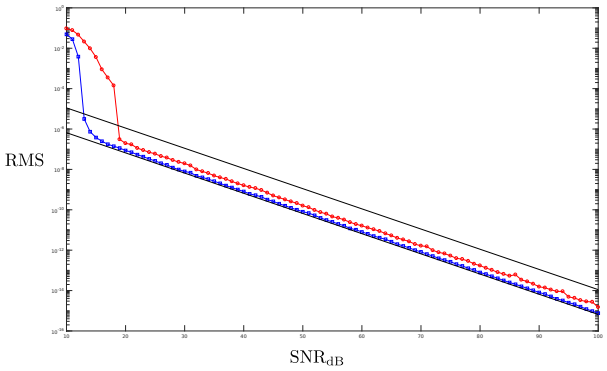

FIGURE 10. *Variance of ESPRIT (blue) and VEXPA (red), compared to the Cramer-Rao lower bound.*

TABLE 2. *Varying SNR experiment with* $n = 12$ *and* $N = 300$.

β_i	γ_i	ω_i	ψ_i
1	0	$-2\pi 5.93$	0
2	π	$-2\pi 4.05$	0
2	$\pi/4$	$-2\pi 3.10$	0
2	$\pi/8$	$-2\pi 1.82$	0
2	$3\pi/4$	$-2\pi 1.31$	0
1	$\pi/10$	$2\pi 1.90$	0
3	$-\pi$	$2\pi 2.97$	0
1.5	$-7\pi/8$	$2\pi 6.05$	0
2	0	$2\pi 6.67$	0
3	$-78\pi/100$	$2\pi 38$	0
1	0	$2\pi 43$	0
1	$\pi/5$	$-2\pi 24$	0

for both ESPRIT (top) and VEXPA (bottom). For higher noise levels (smaller SNR) the stand-alone ESPRIT method returns unreliable results, while the VEXPA method implemented on top of ESPRIT detects that the signal is heavily perturbed as none of the computed results is validated in the cluster analysis. Therefore VEXPA, in its standard implementation, does not return λ_i output.

6. Conclusion

Exponential analysis methods of the Prony type are more sensitive to noise. We offer an add-on technique that regularizes the problem statement and stabilizes and validates the computed results. As we illustrate in the numerical examples the algorithm works very well. In addition, the method is robust with respect to outliers and estimates the model order while performing the validation analysis. The approach is highly suited for parallellization and hence improves the running time of the underling Prony-like exponential analysis.

References

[1] Zeljko Bajzer et al. “Padé-Laplace method for analysis of fluorescence intensity decay”. In: *Biophys J.* 56.1 (1989), pp. 79–93.

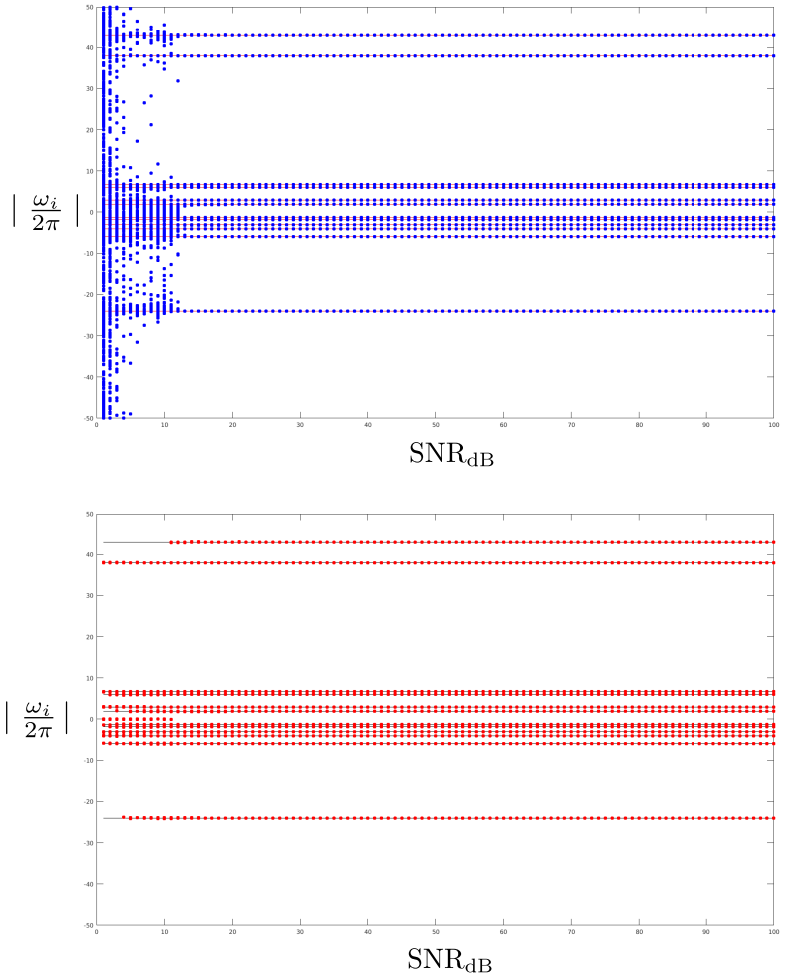

FIGURE 11. *Retrieved ω_i by ESPRIT (top, blue) and VEXPA (bottom, red).*

[2] P. Barone. "On the distribution of poles of Padé approximants to the Z-transform of complex Gaussian white noise". In: *Journal of Approximation Theory* 132.2 (2005), pp. 224 –240.

[3] B. Beckermann, G.H. Golub, and G. Labahn. "On the numerical condition of a generalized Hankel eigenvalue problem". In: *Numer. Math.* 106.1 (2007), pp. 41–68.

[4] Delin Chu and Gene H. Golub. "On a generalized eigenvalue problem for nonsquare pencils". In: *SIAM J. Matrix Anal. Appl.* 28.3 (2006), 770–787 (electronic).

[5] A. Cuyt et al. *Faint and clustered components in exponential analysis.* Tech. rep. Universiteit Antwerpen, 2017.

[6] Annie Cuyt and Wen-shin Lee. *How to get high resolution results from sparse and coarsely sampled data.* Tech. rep. Universiteit Antwerpen, 2017.

[7] Annie Cuyt and Wen-shin Lee. "Smart data sampling and data reconstruction". Pat. PCT/EP2012/066204. Patent PCT/EP2012/066204.

[8] Annie Cuyt and Wen-shin Lee. “Smart data sampling and data reconstruction”. Pat. US 61/611,899. Patent US 61/611,899.

[9] Martin Ester et al. “A density-based algorithm for discovering clusters in large spatial databases with noise”. In: AAAI Press, 1996, pp. 226–231.

[10] J.L. Gammel. “Effect of random errors (noise) in the terms of a power series on the convergence of the Padé approximants”. In: *Padé approximants*. Ed. by P.R. Graves-Morris. 1972, pp. 132–133.

[11] J. Gilewicz and M. Pindor. “Padé approximants and noise: a case of geometric series”. In: *J. Comput. Appl. Math.* 87 (1997), pp. 199–214.

[12] J. Gilewicz and M. Pindor. “Padé approximants and noise: rational functions”. In: *J. Comput. Appl. Math.* 105 (1999), pp. 285–297.

[13] G.H. Golub, P. Milanfar, and J. Varah. “A stable numerical method for inverting shape from moments”. In: *SIAM J. Sci. Comput.* 21 (1999), pp. 1222–1243.

[14] P. Henrici. *Applied and computational complex analysis I.* New York: John Wiley & Sons, 1974.

[15] Yingbo Hua and Tapan K. Sarkar. “Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise”. In: *IEEE Trans. Acoust. Speech Signal Process.* 38 (1990), pp. 814–824.

[16] Erich Kaltofen and Wen-shin Lee. “Early termination in sparse interpolation algorithms”. In: *J. Symbolic Comput.* 36.3-4 (2003). International Symposium on Symbolic and Algebraic Computation (ISSAC’2002) (Lille), pp. 365–400.

[17] Steven M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[18] J. Nuttall. “The convergence of Padé approximants of meromorphic functions”. In: *J. Math. Anal. Appl.* 31 (1970), pp. 147–153.

[19] H. Nyquist. “Certain Topics in Telegraph Transmission Theory”. In: *Trans. Am. Inst. Electr. Eng.* 47.2 (Apr. 1928), pp. 617–644.

[20] Luca Perotti et al. “Identification of gravitational-wave bursts in high noise using Padé filtering”. In: *Physical Review D* 90.12 (2014), p. 124047.

[21] Ch. Pommerenke. “Padé approximants and convergence in capacity”. In: *J. Math. Anal. Appl.* 41 (1973), pp. 775–780.

[22] R. Roy and T. Kailath. “ESPRIT-estimation of signal parameters via rotational invariance techniques”. In: *IEEE Trans. Acoust. Speech Signal Process.* 37.7 (July 1989), pp. 984–995.

[23] Claude E. Shannon. “Communication in the Presence of Noise”. In: *Proc. IRE* 37 (1949), pp. 10–21.

[24] Ying-Xian Yao and Sudhakar M. Pandit. “Cramer-Rao lower bounds for a damped sinusoidal process”. In: *IEEE Trans. Signal Processing* 43 (1995), pp. 878–885.