Projection-based Multivariate Frequency Estimation

Benedikt Diederichs and Armin Iske
Department of Mathematics, University of Hamburg, Germany
{benedikt.diederichs,armin.iske}@uni-hamburg.de

***Abstract*—The parameter estimation problem for multivariate exponential sums is currently a very active field of research. One idea discussed in the last years is to solve multiple one dimensional parameter estimation problems and combine their results to get a solution of the higher dimensional problem. While these projection-based methods give reasonable results with only very few samples, they cannot efficiently use samples on a grid, which is the most common sampling set. In this paper we demonstrate how to extend the projection-based methods to cover parallel lines and hence samples on a grid.**

I. Introduction

The interest in parameter estimation for multivariate exponential sums has dramatically increased in the last few years, driven by multiple applications like super-resolution [1] or inverse scattering [2]. While multiple algorithms are already available, using tools from convex analysis [1], computational algebra [3], [4] or matrix pencil methods [5], [6], to give a non-exhaustive list, there is still a demand for new methods, giving good result with as few samples as possible. In this paper, we present a new method, meeting this demand.

A multivariate exponential sum is a function of the form

$$f(x) = \sum_{j=1}^{M} c_j e^{iy_j \cdot x} \quad \text{for } x \in \mathbb{R}^d, \tag{1}$$

where $y_j \in (-\pi, \pi]^d$ are the pairwise distinct frequencies and $c_j \in \mathbb{C} \setminus \{0\}$ the corresponding coefficients, and where M is the order of f. The task is now to calculate y_j and c_j from samples of f, usually taken at

$$G_N^d = G_N = \{n \in \mathbb{Z}^d \ : \ \|n\|_\infty \leq N\}.$$

While some other sampling sets were investigated as well, samples of f on G_N are the most relevant in applications. To make only one example, the function values of f on G_N can be interpreted as the Fourier coefficients of a linear combination of Dirac deltas, see for example [1].

The method we present here is based on the simple observation, that restricting a multivariate exponential sum to a line

$$\ell = \{tv + bw \ : \ t \in \mathbb{R}\}, \tag{2}$$

where $v, w \in \mathbb{R}^d$ are orthonormal and $b \in \mathbb{R}$, gives a univariate exponential sum of the form

$$f(tv + bw) = \sum_{j=1}^{M} c_j e^{ibw \cdot y_j} e^{i(v \cdot y_j)t} =: f|_\ell(t) \quad \text{for } t \in \mathbb{R}.$$

Taking samples on the line ℓ, we may apply a univariate method (e.g ESPRIT [7]) and obtain some information about the frequencies of f. This trick has been used multiple times, see for example [2], [8], [9]. But all these methods rely on sampling along several non-parallel lines. This does not fit into the setting of sampling on a grid G_N, where our samples are aligned on parallel lines. We show how nonetheless univariate methods can be used to estimate the frequencies $y_j \in \mathbb{R}^d$.

Our method is quite close to a recently introduced method in [10], but we focus on a pre-determined sampling set. Also, we are able to deal with critical special cases. For instance, the restriction of f to a line ℓ gives an exponential sum where some frequencies $y_j \cdot v$ may not occur, due to cancellations.

The rest of this paper is organized as follow. In Section II we introduce the basic idea of using univariate methods along a few parallel lines $(\ell_j)_{j=1,\dots,K}$ and then once more to the coefficients of $f|_{\ell_j}$ to obtain estimates for the frequencies of f. Also, we give a sampling set $\tilde{G}_N \subset G_N$, where unique reconstruction is possible. In Section III we then present a new ESPRIT type algorithm, which bundles the application of the univariate ESPRIT to a single step, stabilizing our scheme significantly. Numerical examples are finally provided in Section IV.

II. Projected Coefficient Algorithm

Let us briefly recall the univariate case, $d = 1$ in (1). If an upper bound, N, for the order M of f is known, $M \leq N$, then we can reconstruct M, the frequencies y_j and the coefficients c_j from $2N$ samples $f(k),\ k = -N, \dots, N-1$. From the large set of available methods, we prefer to work with ESPRIT [7], which is well-known for its good efficiency and stability.

If we restrict a multivariate exponential function to a line ℓ, as given in (2), then we will obtain a univariate function

$$f|_\ell(t) = \sum_{y \in \{y_j \cdot v\}} c_y^\ell e^{iyt} \quad \text{for } t \in \mathbb{R} \tag{3}$$

with coefficients

$$c_y^\ell = \sum_{j \,:\, y = y_j \cdot v} c_j e^{ibw \cdot y_j}.$$

Now two possible difficulties may occur. Firstly, the number of frequencies of $f|_\ell$ may be smaller than M because several frequency vectors y_j may feature the same projection $y_j \cdot v$. Secondly, it may happen that $c_y^\ell = 0$, which results in hidden frequencies y_j, which are seemingly invisible in $f|_\ell$.

While the first problem is not too severe, the second one is not covered in previous methods relying on univariate ideas in

the multivariate setting [8], [10]. Although in [9] we discuss how to circumvent these problems theoretically, the solution is computational unfeasible.

To this end, we show how parallel lines can be used to solve these issues. To suppress some technical details, we start with the important two dimensional case. Without loss of generality, we assume $v = e_1 = (1,0)^T$ and $b = 0$ in (2). For $L \in \mathbb{N}$ let $(\ell_k)_{k=-L}^{L-1}$ be a collection of parallel lines

$$\ell_k = \{te_1 + ke_2 \ : \ t \in \mathbb{R}\}.$$

This is the case covering the sample set G_N.

The key idea is now as follows. The coefficients $c_y^{\ell_k}$ are again samples of a univariate exponential sum

$$c_y^{\ell_k} = c_y(k) = \sum_{j \,:\, y_{j1} = y} c_j e^{iy_{j2}k},$$

where we let $y_j = (y_{j1}, y_{j2})^T$. This idea is used in [10] as well. Now if $L > M$, then we have enough samples of all c_y to reconstruct their frequencies. This has the nice feature of matching the different projections of the frequency vectors automatically, which is otherwise a difficult problem when trying to use one dimensional methods. Indeed, for any frequency y_{j2} of c_y, we know that $(y, y_{j2})^T \in \mathbb{R}^2$ is a frequency vector of f with coefficient c_j and all frequencies of f are obtained in such a way. In fact, we can even reduce the sampling set a little bit, as the order of all c_y sum up to the order of f and hence we either have a lot of frequencies in $f|_\ell$ or in some of the c_y. This leads us to our first result.

Theorem 1. *Let f be an exponential sum as in* (1) *for dimension $d = 2$. Moreover, assume $N \geq M$ is known. Further consider the set*

$$\tilde{G}_N = \{(n + \delta_1, m + \delta_2)^T \in \mathbb{Z}^2 \ : \ \delta_j \in \{0,1\} \text{ and } |n| + |m| < N\}.$$

Then f is uniquely determined by $f|_{\tilde{G}_N}$ and its parameters can be reconstructed using univariate methods.

Proof: Denote the frequencies of f by Y and let

$$Y_1 = \{y \ : \ y \text{ is a frequency of a } f_{\ell_k},\ k = 0, 1\}.$$

This set can be calculated by applying any univariate method along the lines ℓ_0, ℓ_1. Denote the corresponding coefficients by $c_y(k),\ k = 0, 1$. Assuming that these are exponential sums with only one frequency each (namely the second component of the frequency vector of f with first component y), we use a univariate method and obtain $\tilde{c}_y \in \mathbb{C}$ and $\tilde{y} \in (-\pi, \pi]$ with

$$c_y(k) = \tilde{c}_y e^{i\tilde{y}k}.$$

This gives an exponential sum

$$f_1(x) = \sum_{y \in Y_1} \tilde{c}_y e^{i(y,\tilde{y})^T \cdot x} \quad \text{for } x \in \mathbb{R}^2.$$

If f_1 agrees on all samples with f, we are done. Otherwise at least two frequency vectors of f have to have the same first component. But then only $2N - 2$ samples suffice to reconstruct all frequencies of f_{ℓ_k}. We repeat the procedure with two more lines. Let

$$Y_2 = \{y \ : \ y \text{ is a frequency of a } f_{\ell_k},\ k = -1, 0, 1, 2\}.$$

Again, we obtain Y_2 and the corresponding coefficients $c_y(k),\ -1 \leq k \leq 2$, by applying univariate methods. We build f_2 as we built f_1. If f_2 agrees with f on our sampling set, we are done. Otherwise at least three frequency vectors of f have the same first component. We iterate this argument. Note that in the case where all frequency vectors of f have the same first component, we have two samples on $2N$ parallel lines and the coefficient sum has order N. Now we have exactly the number of samples needed to cover this case. ∎

Now we sketch the higher dimensional case, again choosing lines that are parallel to the standard axes. For each $k \in G_N^{d-1}$ we consider the line

$$\ell_k = \left\{ te_1 + \sum_{j=1}^{d-1} k_j e_{j+1} \ : \ t \in \mathbb{R} \right\},$$

where $e_1 = (1, 0, \ldots, 0) \in \mathbb{R}^d$. Again, f_{ℓ_k} takes the form (3) with frequencies y being equal to the first component of the frequency vectors of f. The coefficients $c_y^{\ell_k}$ are exponential sums of dimension $d - 1$, sampled at $k \in \mathbb{Z}^{d-1}$, as each line gives us one such sample. In the generic case, where all frequency vectors of f have different first components, all $c_y(k)$ have actually only one frequency vector, namely the last $d - 1$ components. Otherwise, when we sample f at the grid $G_N^d = G_N$, we can use univariate methods to obtain samples for each of the exponential sums c_y at G_N^{d-1}. Therefore, we can recursively reduce the multivariate problem to multiple one dimensional problems.

As we use samples on G_N^d, we need $\mathcal{O}(N^d)$ samples. Even when reducing the sampling set with a more sophisticated calculation, as in Theorem 1, we still need $\mathcal{O}(N^d)$ samples. On the other hand, algebraic techniques prove that $\mathcal{O}(N^2 \log^{2d-2}(N))$ are sufficient [11]. This rather poor efficiency is shared by all methods operating on G_N^d. But in the relevant case of two dimensions, this is not critical at all.

Although the above construction provides an algorithm, several problems still remain, in particular with noisy samples.

- We use univariate methods very often. In the bivariate case we start using the univariate method $2N$ times to obtain the frequencies of $f|_{\ell_k}$ and up to N times for the c_y's. As the dimension increases, the number of utilized univariate methods increases rapidly, as we already start with $(2N)^{d-1}$ different lines.
- After using the univariate method along the first collection of lines, we have to decide when two frequencies $y,\ \tilde{y}$ are considered equal. Therefore, some level of tolerance is required, i.e., we need to select one $\delta > 0$ such that for

$$\min_{k \in \mathbb{Z}} |y - \tilde{y} - 2\pi k| = |y - \tilde{y}|_{\mathbb{T}} < \delta$$

we consider y and $\tilde{y}$ as equal. But the selection of the tolerance δ is obviously critical, as small changes in δ may change the output of the procedure significantly.

We address these issues in the next section. We propose an algorithm, which estimates the frequencies along several parallel lines simultaneously. Hence, no matching is required.

III. Simultaneous Frequency Estimation

When we have a multivariate exponential sum f sampled along several parallel lines and perform a univariate method along each of the lines, we expect to find the same frequencies along each line except for a few frequencies missing along some of the lines due to cancellation. And as we are only interested in the set of all frequencies it is better to estimate them in a joint procedure. To this end, we propose an extension of the well known ESPRIT algorithm [7], which covers the following situation. Let

$$f_l(x) = \sum_{y \in Y_l} c_y^l e^{iyx} \quad l = 1, \dots, L$$

be a collection of univariate exponential sums with frequency sets Y_l, $M_l = |Y_l|$, coefficients $c^l = (c_y^l)_{y \in Y_l} \in (\mathbb{C} \setminus \{0\})^{M_l}$.

We aim to recover $Y = \cup_{l=1}^L Y_l$ from given samples $f_l(k)$, $k = 0, \dots, 2N$, where we assume $M = |Y| \leq N$. This fits (up to a simple transformation) into our setting of sampling on G_N. To solve this problem, we propose a solution close to the matrix enhancement method in [5], adapted to our particular situation. To this end, just as in the univariate ESPRIT algorithm, we consider the Hankel matrix

$$\mathbf{H}^l_{2N-P+1,P+1} = \begin{pmatrix} f_l(0) & \dots & f_l(P) \\ f_l(1) & \dots & f_l(P+1) \\ \vdots & \ddots & \vdots \\ f_l(2N-P) & \dots & f_l(2N) \end{pmatrix},$$

$\mathbf{H}^l_{2N-P+1,P+1} \in \mathbb{C}^{(2N-P+1)\times(P+1)}$. We assume $P \geq N$. We can rely on the factorization

$$\mathbf{H}^l_{2N-P+1,P+1} = \mathbf{V}_{2N-P+1,M_l}(z^l)\mathrm{diag}(c^l)\mathbf{V}_{P+1,M_l}(z^l)^T.$$

Here, $z^l = (e^{iy})_{y \in Y_l} \in \mathbb{C}^{M_l}$ and

$$\mathbf{V}_{m,n}(z_1, ..., z_n) = (z_p^q)_{\substack{q=0,\dots,m \\ p=0,\dots n}} \in \mathbb{C}^{m \times n}$$

is a Vandermonde matrix. Clearly, we can insert some additional frequencies with coefficients zero (again, we denote the new coefficient vector as c^l for notational convenience) to obtain

$$\mathbf{H}^l_{2N-P+1,P+1} = \mathbf{V}_{2N-P+1,M}(z)\mathrm{diag}(c^l)\mathbf{V}_{P+1,M}(z)^T,$$

where $z = (e^{iy})_{y \in Y} \in \mathbb{C}^M$. We then stack these Hankel matrices in a big matrix

$$\mathbf{H}_{N,P+1} = \left((\mathbf{H}^1_{2N-P+1,P+1})^T \dots (\mathbf{H}^L_{2N-P+1,P+1})^T\right)^T.$$

Due to the common factorization we have

$$\mathbf{H}_{N,P+1} = \tilde{\mathbf{V}}\tilde{\mathbf{D}}\mathbf{V}_{P+1,M}(z)^T \tag{4}$$

where $\tilde{\mathbf{V}}$ is a matrix consisting of L $\mathbf{V}_{2N-P+1,M}(z)$ blocks on the diagonal and $\tilde{\mathbf{D}} = \left(\mathrm{diag}(c^1) \dots \mathrm{diag}(c^L)\right)^T$. By deleting the first or the last column of $\mathbf{H}_{N,P+1}$ we obtain the matrices $\mathbf{H}_{N,P+1}(0)$ and $\mathbf{H}_{N,P+1}(1)$, respectively. They feature a factorization like (4), where $\mathbf{V}_{P+1,M}(z)^T$ is replaced by $\mathbf{V}_{P,M}(z)^T$ and for $\mathbf{H}_{N,P+1}(1)$ each diagonal matrix of $\tilde{\mathbf{D}}$ is multiplied with $\mathrm{diag}(z)$. Hence, the matrix pencil

$$\mathbf{H}_{N,P+1}(1) - \mu\mathbf{H}_{N,P+1}(0) \tag{5}$$

has rank reducing numbers equal to $\{e^{iy},\ y \in Y\}$, just as in the univariate case of the ESPRIT algorithm. To actually solve this matrix pencil, we use its special structure. We start with the SVD of $\mathbf{H}_{N,P+1}$

$$\mathbf{H}_{N,P+1} = \mathbf{U}\Sigma\mathbf{W}. \tag{6}$$

This matrix has rank equal to the order of f, as (4) shows. Hence, calculating the SVD we obtain M. When we have noisy samples, the rank has to be estimated, for example by thresholding the singular values. Furthermore, we can delete all up to the first M rows of $\mathbf{W}$ and all but the first M columns of Σ. Now by deleting the first (resp. last) column of $\mathbf{W}$ we obtain a SVD of $\mathbf{H}_{N,P+1}(0)$ (resp. $\mathbf{H}_{N,P+1}(1)$), given by

$$\mathbf{H}_{N,P+1}(s) = \mathbf{U}\tilde{\Sigma}\tilde{\mathbf{W}}(s) \qquad s \in \{0, 1\}.$$

Note that $\tilde{\mathbf{W}}(s)$ has full rank and (5) is equivalent to the pencil

$$\tilde{\mathbf{W}}(1) - \mu\tilde{\mathbf{W}}(0)$$

where $\tilde{\mathbf{W}}(s) \in \mathbb{C}^{M\times P}$. Hence, every matrix $\mathbf{F} \in \mathbb{C}^{M\times M}$ satisfying

$$\mathbf{F}^T\tilde{\mathbf{W}}(1) = \tilde{\mathbf{W}}(0)$$

has eigenvalues equal to the rank reducing numbers and therefore to $\{e^{iy},\ y \in Y\}$. We use the least square solution

$$\mathbf{F} = (\mathbf{W}(0)^T)^\dagger\mathbf{W}(1)^T, \tag{7}$$

where $\mathbf{A}^\dagger$ denotes the Moore-Penrose pseudoinverse of matrix A. Of course the total least square solution is possible as well.

In summary, we obtain the following algorithm.

Algorithm. *Given: $f(n,k)$, $0 \leq n,k \leq 2N$, upper bound $P \leq N$ on the order of f, threshold $tol > 0$, second threshold $\varepsilon > 0$*

1) *Build the Hankel matrix $\mathbf{H}_{N,P+1}$ and compute its singular value decomposition* (6)*.*
2) *Find M, such that $\sigma_M > tol \cdot \sigma_1 \geq \sigma_{M+1}$, where σ_j is the j-th singular value of $\mathbf{H}_{N,P+1}$. Thus, M estimates the order of f.*
3) *Form $\tilde{\mathbf{W}}(s)$ and calculate $\mathbf{F}$ as given in* (7) *and compute its eigenvalues $\lambda_1, \dots, \lambda_M$.*
4) *Solve for the coefficents c_y^l by solving*

$$\mathbf{V}_{2N,M}(\lambda_1, \dots \lambda_M)c^l = (f(l,k))_{k=0,\dots,2N}$$

5) *For every $j = 1, \dots, M$ apply the univariate ESPRIT algorithm to $(c_j^l)_{l=-N,\dots N}$. Call the observed frequencies $\{y_{jk},\ k = 1, \dots, M_j\}$ with coefficients c_{jk}.*

6) *Delete all frequencies with coefficients of modulus smaller than ε. Recalculate coefficients.*
7) *Let $y_{j1} = \log(\lambda_j)/i$, where* log *is the principal branch of the logarithm. Result: $\{(y_{j1}, y_{jk}),\ j = 1, \dots, M\ k = 1, \dots M_j\}$ with coefficients c_{jk}.*

Remark: One advantage of this algorithm is, that it can be used with significantly less samples, for example with the sampling set $G_{N,2} = \{(k, j),\ k = 0, \dots, 2N\ j = 0, 1\}$. While a recovery cannot be guaranteed, assuming the frequencies are randomly distributed, samples on $\tilde{G}_{N,2}$ are almost surely sufficient, as no cancellation will occur. We demonstrate this in the next section. Also, it should be noted that we only need the right singular vectors and that the linear systems in 4) all have the same system matrix.

Finally, we remark that the tolerance parameter $tol > 0$ essentially needs to reflect the noise level. In fact, tol is the value for the threshold used in the univariate ESPRIT algorithm to determine the number of frequencies. On the other hand, the threshold $\varepsilon > 0$ cuts off small (noisy) coefficients c_j to zero. Hence, possible a priori knowledge on the minimal modulus of the coefficients c_j may be used.

IV. Numerical Examples

Now to illustrate the proposed algorithm, we give numerical examples. We focus on the undersampled case, where many known algorithms are not even directly applicable.

First Example: We choose frequencies and coefficients randomly. M frequencies are uniformly distributed in $(-\pi, \pi]^2$, the coefficients in $\{c \in \mathbb{C} \ :\ 0.25 \leq |c| \leq 2.25\}$. We bound the coefficients from below to avoid numerically undetectable frequencies. The sampling set is given by

$$G_{N,K} = \{(n, k) \ :\ 0 \leq n \leq N-1,\ 0 \leq k \leq K\}.$$

We report the ℓ_2 error in the frequencies

$$e_2(Y, Y_{est}) = \min_{\sigma \text{ permutation}} \sum_{j=1}^{|Y|} \|y_j - y_{\sigma(j),est}\|_2$$

whenever it is smaller than 0.5, otherwise we count the estimate as *failed*. Also, we count all instances as failed, where the number of frequencies does not match. All errors are averages over 1000 runs.

To test the stability with respect to noise, we add noise which is uniformly distributed in $\alpha([-1/2, 1/2] + i[-1/2, 1/2])$. The threshold for the rank estimation we denote by tol and the upper bound for the order of f by P. Throughout all instances, we set the threshold for the coefficients to $\varepsilon = 0.1$. The results are reported in Table I. Overall, it is clear that the algorithm runs reasonable with a very small number of samples. If noise is large, more samples are needed to stabilize the procedure.

Second Example: We now consider a specific example, where critical cancellations occur. To this end, we consider the frequencies

$$y_1 = \pi(0.2, 0.2)^T, \qquad y_2 = \pi(0.6, 0.2)^T,$$
$$y_3 = \pi(0.6, 0.2)^T, \qquad y_4 = \pi(0.6, 0.6)^T$$

TABLE I
Results of the First Example

M	P	N	K	tol	α	$e_2(Y, Y_{est})$	fails/1000
5	5	10	2	1e-7	0	1e-09	0
5	5	10	2	1e-5	1e-6	3e-03	24
5	10	20	2	1e-5	1e-6	6e-04	2
5	10	20	4	1e-5	1e-6	3e-05	1
5	10	20	4	1e-2	1e-3	4e-03	34
10	10	20	2	1e-5	1e-6	2e-02	114
10	10	40	4	1e-5	1e-6	5e-03	28
10	20	50	8	1e-5	1e-6	5e-06	2
10	20	50	8	1e-2	1e-2	4e-03	100
10	20	60	15	1e-2	1e-2	1e-03	81
10	20	60	40	1e-2	1e-2	8e-04	50

with coefficients $c_1, c_4 = 1$, $c_2, c_3 = -1$. Due to cancellations, $f(x, 0) = f(0, y) = 0$ for all $x, y \in \mathbb{R}$, which makes the frequency estimation problem rather difficult for many algorithms.

Running the proposed algorithm with $P = 4$, samples on $G_{8,4}$ and a noise level of $\alpha = 1e\text{-}2$ and $tol = 1e\text{-}2$, we observed an average error of $e_2 \approx 1e\text{-}2$ for over 1000 runs with no fails. To give one example where the exact number of frequencies is not given, samples on $G_{20,8}$ with the same noise level and $P = 10$, the algorithm gives an average error of $5e\text{-}3$, where no fail occurred.

References

[1] E. J. Candès and C. Fernandez-Granda, "Towards a mathematical theory of super-resolution," *Communications on Pure and Applied Mathematics*, vol. 67, no. 6, pp. 906–956, 2014.

[2] R. Griesmaier and C. Schmiedecke, "A multifrequency music algorithm for locating small inhomogeneities in inverse scattering," *arXiv preprint arXiv:1607.04017*, 2016.

[3] S. Kunis, T. Peter, T. Roemer, and U. von der Ohe, "A multivariate generalization of Prony's method," *Linear Algebra and its Applications*, vol. 490, pp. 31–47 2016.

[4] T. Sauer, "Pronys method in several variables," *Numerische Mathematik*, pp. 1–28, 2016.

[5] Y. Hua, "Estimating two-dimensional frequencies by matrix enhancement and matrix pencil," *Signal Processing, IEEE Transactions on*, vol. 40, no. 9, pp. 2267–2280, 1992.

[6] S. Rouquette and M. Najim, "Estimation of frequencies and damping factors by two-dimensional ESPRIT type methods," *Signal Processing, IEEE Transactions on*, vol. 49, no. 1, pp. 237–245, 2001.

[7] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, no. 7, pp. 984–995, 1989.

[8] D. Potts and M. Tasche, "Parameter estimation for multivariate exponential sums," *Electronic Transactions on Numerical Analysis*, vol. 40, pp. 204–224, 2013.

[9] B. Diederichs and A. Iske, "Parameter estimation for bivariate exponential sums," *Proc. Sampling Theory and Applications*, pp. 493–497, 2015.

[10] A. Cuyt and A. Lee "Multivariate Exponential Analysis from the Minimal Number of Samples," *arXiv preprint arXiv:1610.06329*, 2017.

[11] T. Sauer, "Prony's method in several variables: symbolic solutions by universal interpolation," *arXiv preprint arXiv:1603.03944*, 2016.