FREQUENCY ESTIMATION BASED ON HANKEL MATRICES AND THE ALTERNATING DIRECTION METHOD OF MULTIPLIERS

Fredrik Andersson[1], Marcus Carlsson[1], Jean-Yves Tourneret[2], Herwig Wendt [2]

[1]Centre for Mathematical Sciences, Lund University, Lund, Sweden
`fa@maths.lth.se, marcus.carlsson@math.lu.se`
[2]University of Toulouse, CNRS UMR 5505, INP-ENSEEIHT, Toulouse, France
`jean-yves.tourneret@enseeiht.fr, herwig.wendt@irit.fr`

ABSTRACT

We develop a parametric high-resolution method for the estimation of the frequency nodes of linear combinations of complex exponentials with exponential damping. We use Kronecker's theorem to formulate the associated nonlinear least squares problem as an optimization problem in the space of vectors generating Hankel matrices of fixed rank. Approximate solutions to this problem are obtained by using the alternating direction method of multipliers. Finally, we extract the frequency estimates from the con-eigenvectors of the solution Hankel matrix. The resulting algorithm is simple, easy to implement and can be applied to data with equally spaced samples with approximation weights, which for instance allows cases of missing data samples. By means of numerical simulations, we analyze and illustrate the excellent performance of the method, attaining the Cramér-Rao bound.

Index Terms— frequency estimation, nonlinear least squares, Hankel matrices, Kronecker's theorem, missing data, alternating direction method of multipliers

1. INTRODUCTION

Spectral estimation constitutes a classical problem that has found applications in a large variety of fields (including astronomy, radar, communications, economics, medical imaging, spectroscopy, . . . , to name but a few). One important category of spectral estimation problems arises for signals that can be well represented by the parametric model

$$f_c(t) = \sum_{p=0}^{P-1} c_p e^{\zeta_p t} + e(t), \quad c_p, \zeta_p \in \mathbb{C}, \qquad (1)$$

where $e(t)$ is an additive noise term. Given a vector of (typically equally spaced) samples,

$$f(j) = f_c(t_0 + jt_s), \quad 0 \le j \le 2N$$

where t_s is the sampling period, the goal is to estimate the complex frequency nodes $\zeta_p = 2\pi(\gamma_p + i\nu_p)$, i.e., the damping and frequency parameters γ_p and ν_p. Note that once the parameters $\boldsymbol{\zeta} = (\zeta_0, \cdots, \zeta_{P-1})^T$ have been computed, determining $\mathbf{c} = (c_0, \cdots, c_{P-1})^T$ reduces to a simple linear regression problem. Thus, the focus lies on the estimation of the nodes $\boldsymbol{\zeta}$. The signal model (1) includes the case of exponentially damped signals defined by $\gamma_p < 0$. This case has recently received significant interest, notably in applications involving nuclear quadrupole resonance signals for which the γ_p (together with ν_p) are used as a signature of the chemical composition of an analysed substance (see for instance [1] and references therein).

The literature on methods for estimating $\boldsymbol{\zeta}$ in this setting is rich, cf. [2] for an overview. Here, we focus on so-called *high-resolution* methods, i.e., methods whose frequency resolution are not tied to a pre-defined grid (and can attain numerical precision in the noise free case).

One important class of techniques is based on the principle of nonlinear least squares (NLS) [3] and aims at solving the nonlinear regression problem associated with (1), i.e., at minimizing $\theta(\boldsymbol{c}, \boldsymbol{\zeta}) = \sum_{j=0}^{2N} \left| f(j) - \sum_{p=0}^{P-1} c_p e^{\zeta_p j} \right|^2$ with respect to the parameter vectors $\boldsymbol{c}$ and $\boldsymbol{\zeta}$. Although NLS enjoy desirable theoretical properties (e.g., those of maximum likelihood estimators when $e(t)$ is a Gaussian white noise; robustness to coloured noise), their practical use remains limited due to the extreme difficulty of globally minimizing $\theta(\boldsymbol{c}, \boldsymbol{\zeta})$ due to its pronounced multimodal shape with one very sharp global minimum (cf., e.g., [2] and references therein).

A second prominent and popular class is given by subspace methods (such as MUSIC [4], ESPRIT [5] and min-norm [6]). These methods are based on estimates of the covariance of f and rely on the assumption that the noise e is white.

In the present contribution, we develop a novel high-resolution methodology for the estimation of $\boldsymbol{\zeta}$ in (1). Similar to NLS, we aim at approximating f as well as possible by a linear combination of complex exponentials. However, the proposed methodology fundamentally departs from any of the above methods in the following ways:

First, it is based on *Kronecker's theorem* for complex symmetric matrices, which essentially states that if a function f_c

is uniformly sampled, then the *Hankel matrix* that is generated by the vector of samples f has rank P if and only if f_c coincides at the sample points with a function that is a linear combination of P exponential functions. This fact has been used for the sparse approximation of functions by sums of exponentials in the con-eigenvalue approach [7] and the alternating projections method [8]. Here, it is used to formulate an NLS type minimization problem in which the model (and its parameters c and ζ) does not enter explicitly. Instead, the model is imposed *implicitly* by constraining the rank of the Hankel matrix generated by the vector g approximating f to equal P. Consequently, residual minimization is not performed over the parameter space $\{c, \zeta\}$ directly, but over the space of vectors g which generate Hankel matrices of rank P. The frequency nodes ζ are then obtained by considering the con-eigenvectors of the solution Hankel matrix.

Second, we reformulate the minimization problem such that it can be effectively solved by the *alternating direction method of multipliers* (ADMM) [9]. ADMM is an iterative technique that is recently gaining popularity due to its robustness, versatility and applicability to large-scale problems. Moreover, this technique enjoys performance comparable to problem-specific state-of-the-art methods. While the optimization problem considered here is nonconvex, it is shown numerically that the solutions obtained with the ADMM procedure generically provide excellent approximations and parameter estimates for the model (1).

The resulting Hankel matrix ADMM frequency estimation procedure is practically extremely appealing due to its simplicity and ease of implementation. It enjoys excellent performance, outperforming subspace methods while at the same time alleviating the practical limitations of classical NLS. Unlike subspace methods, it does not rely on a specific noise model. Another interesting property is that it applies to situations with *missing* (or censored) data.

The remainder of this work is organized as follows. In Section 2, we define an optimization problem based on Kronecker's theorem for the approximation by sums of complex exponentials and formulate the ADMM-based procedure for obtaining its solution. Section 3 summarizes the procedure for obtaining frequency node estimates from the solution Hankel matrix. In Section 4, we analyze the performance of the method by means of numerical simulations. Section 5 concludes this contribution and points to future work.

2. APPROXIMATION BY SUMS OF EXPONENTIALS USING ADMM

In this section, we approximate the vector f as well as possible by a vector g that is a linear combination of complex exponential functions. Using Kronecker's theorem, we express the solution to this approximation problem in terms of the vector g whose Hankel matrix has rank P and minimizes the residual. The resulting optimization problem will then be reformulated and solved by ADMM.

2.1. Hankel matrices and Kronecker's theorem

For completeness, we recall that a Hankel matrix A has constant values on the anti-diagonals, i.e., $A(j, k) = A(j', k')$ if $j + k = j' + k'$. It can thus be generated elementwise from a vector $g(j)$, such that $A(j, k) = g(j + k), 0 \le j, k \le N$. We denote this by $A = Hg$.

Kronecker's theorem [10, 11] states that if the Hankel matrix $A = Hg$ generated by the vector g is of rank P then, with the exception of degenerate cases, there exists $\{\zeta_p\}_{p=0}^{P-1}$ and $\{c_p\}_{p=0}^{P-1}$ such that

$$g(j) = \sum_{p=0}^{P-1} c_p e^{\zeta_p j}, \quad c_p, \zeta_p \in \mathbb{C}. \tag{2}$$

The converse holds as well.

It follows that the best approximation (in the l_2 sense) of f by a linear combination of P complex exponentials is given by the vector g which satisfies $\text{rank}(Hg) = P$ and minimizes the l_2 norm of the residual $r = f - g$. In other words, g can be obtained as the solution to the optimisation problem

$$\begin{aligned} \underset{g}{\text{minimize}} \quad & \frac{1}{2}\|r\|_2^2 = \frac{1}{2}\|f - g\|_2^2 \\ \text{subject to} \quad & \text{rank}(Hg) = P. \end{aligned} \tag{3}$$

The parameter P can be selected by considering the singular values of the Hankel matrix Hf generated by the samples f, cf., [7]. At this stage, we assume P to be given.

2.2. A solution based on ADMM

Let $\mathcal{R}_P$ be the indicator function for square matrices S given by

$$\mathcal{R}_P(S) = \begin{cases} 0 & \text{if rank}(S) = P, \\ \infty & \text{otherwise.} \end{cases}$$

Then, problem (3) can be reformulated as follows

$$\begin{aligned} \underset{A,r}{\text{minimize}} \quad & \mathcal{R}_P(A) + \frac{1}{2}\|r\|_2^2 \\ \text{subject to} \quad & A(j, k) + r(j + k) = f(j + k), \quad 0 \le j, k \le N. \end{aligned} \tag{4}$$

The optimization problem (4) has a cost function that consists of two terms, each depending only on one of the variables A and r, along with a linear constraint. The problem formulation is therefore well suited to be addressed using ADMM. ADMM is an iterative technique in which a solution to a large global problem is found by coordinating solutions to smaller subproblems. It can be seen as an attempt to merge the benefits of dual decomposition and augmented Lagrangian methods for constrained optimization. For an overview of the ADMM method see [9]. For convex cost function ADMM

is guaranteed to converge to the unique solution. For non-convex problems it can get stuck at local minima, although it can work quite well in practice also in these situations, cf. [9, Chapter 9]. Since the rank constraint $\mathcal{R}_P$ is nonconvex, convergence of an algorithm solving (4) is not guaranteed. Our numerical simulations indicate that the method works substantially better than established high resolution techniques like ESPRIT, and moreover that it can be applied to cases where ESPRIT is non-applicable (e.g., missing data)[1].

An ADMM iterative step (from iteration q to iteration $q+1$) for (4) reads

$$A^{q+1} = \underset{A}{\arg\min}\, L(A, r^q, \Lambda^q), \qquad (5)$$

$$r^{q+1} = \underset{r}{\arg\min}\, L(A^{q+1}, r, \Lambda^q), \qquad (6)$$

$$\Lambda^{q+1}(j,k) = \Lambda^q(j,k) + \rho\big(A^{q+1}(j,k) + r^{q+1}(j+k) - f(j+k)\big) \qquad (7)$$

where $\Lambda = \Lambda_{\mathbb{R}} + i\Lambda_{\mathbb{I}}$ are the Lagrange multipliers, and $L(A, r, \Lambda)$ is the augmented Lagrangian associated with (4),

$$L(A, r, \Lambda) = \mathcal{R}_P(A) + \frac{1}{2}\|r\|_2^2 +$$
$$\sum_{j,k=0}^{N} \Big[\Lambda_{\mathbb{R}}(j,k)\mathsf{Re}\big(A(j,k) + r(j+k) - f(j+k)\big)$$
$$+ \Lambda_{\mathbb{I}}(j,k)\mathsf{Im}\big(A(j,k) + r(j+k) - f(j+k)\big)$$
$$+ \frac{\rho}{2}\big|A(j,k) + r(j+k) - f(j+k)\big|^2 \Big].$$

The first minimization step (5) is nonconvex due to the projection onto a non convex set but can be solved analytically. Deriving closed form expressions for both minimization steps (5) and (6) is straightforward. The solution A^{q+1} to the first minimization step is equal to the best rank P approximation of the matrix B defined elementwise by

$$B(j,k) = f(j+k) - r^q(j+k) - \rho^{-1}\Lambda^q(j,k).$$

Denoting by $B = U\Sigma V^*$ the singular value decomposition of B, then by the Eckart-Young theorem,

$$A^{q+1} = U\Sigma_P V^*, \qquad (8)$$

where Σ_P is obtained from the diagonal matrix Σ by replacing the diagonal elements $\Sigma(l,l)$ for $l > P$ with zeros. The solution to the second minimization step is given by

$$r^{q+1}(l) = \frac{\rho Q(l) f(l) - \sum_{j+k=l} \big[\Lambda^q(j,k) + \rho A^{q+1}(j,k)\big]}{1 + \rho Q(l)}, \qquad (9)$$

where $\sum_{j+k=l}$ stands for $\sum_{\{0 \le j,k \le N\} \cap \{j+k=l\}}$ and

$$Q(l) = \begin{cases} l+1 & \text{if } l \le N, \\ 2N+1-l & \text{otherwise.} \end{cases}$$

---

[1] A convex problem could be obtained by replacing the rank constraint $\mathcal{R}_P$ by the nuclear norm, at the cost of biased solutions.

With these explicit expressions for the ADMM minimization steps, the procedure for the approximation by sums of exponentials is given by the following simple MATLAB function:

```
1  function f_appro=expo_sum_admm(f,P,rho,it_max)
2    %Best approximation to f using P exponentials.
3    N=(length(f)-1)/2;r=zeros(2*N+1,1);
4    Lambda=zeros(N+1);Q=[1:N+1,N:-1:1]';
5    for iter=1:it_max,
6      B=form_hank(f-r)-Lambda/rho;[u,s,v]=svd(B);
7      s=diag(s);s(P+1:end)=0;A=u*diag(s)*v';
8      r=(rho*Q.*f-sum_hank(Lambda+rho*A))./(1+rho*Q);
9      Lambda=Lambda+rho*(A-form_hank(f-r));
10   end;
11   f_appro=sum_hank(A)./Q;
12
13   function H=form_hank(f)      %form Hankel matrix
14     H=hankel(f(1:(end+1)/2),f((end+1)/2:end));
15   function f=sum_hank(H)       %sum anti-diagonals
16     N=size(H,1)-1; f=zeros(2*N+1,1); H=flipud(H);
17     for j=-N:N, f(j+N+1,1)=sum(diag(H,j)); end;
```

Extension to the missing data case. It is straightforward to replace $\|r\|_2^2$ in problem (4) by a weighted norm

$$\|r\|_{2,w}^2 = \sum_{j=0}^{2N} w(j)|r(j)|^2$$

and to derive the corresponding ADMM procedure. The only change in the above final expressions is that the denominator in the second minimization step (9) (and in the corresponding line 8 of the MATLAB code) is replaced by $w(l) + \rho Q(l)$.

Let $\mathcal{J}$ denote the set of indices j of the missing data samples and set $f(\mathcal{J}) = 0$. The use of weights $w_{\mathcal{J}}$ defined by

$$w(j) = w_{\mathcal{J}}(j) = \begin{cases} 0 & \text{if } j \in \mathcal{J} \\ 1 & \text{otherwise} \end{cases}$$

yields an ADMM procedure for data with missing samples.

Convergence. In order to highlight the simplicity of the proposed algorithm, the ADMM procedure was described above to terminate after a preset number q^* of iterations. Commonly used stopping criteria are based on the magnitude of the primal and dual residuals and are straight-forward to incorporate in the ADMM procedure (cf., [9] for details).

The ADMM procedure would converge to the global solution of the problem (4) if the problem was convex. Since it is non convex, it is only guaranteed to converge locally and can converge to different points depending on the choice of ρ and initial points Λ^0 and r^0. Numerical experiments indicate, however, that the ADMM solution $g = f - r^{q^*}$ of the problem (4) provides in general an excellent approximation for f. Here, we have initialized Λ^0 and r^0 with zeros.

3. FREQUENCY ESTIMATION

Once the solution Hankel matrix $A = Hg$ to (4) has been found, the parameter vector ζ can be obtained by considering

ν_p	1.86	6.59	7.49	19.84			
γ_p	0.20	−0.28	0.04	−0.23			
$|c_p|$	1.00	0.40	1.50	0.70			
ν_p	1.86	3.84	5.95	7.49	9.60	19.84	30.08
γ_p	0.05	0.14	0.06	0.01	0.16	0.08	0.07
$|c_p|$	1.00	0.40	1.50	0.70	0.60	1.20	1.00

Table 1. Parameters for the $P = 4$ (top) and $P = 7$ (bottom) complex exponentials used in the numerical simulations.

the so-called con-eigenvectors of A. Hankel matrices belong to the class of complex symmetric matrices satisfying $A = A^T$, which generically can be decomposed as [12]

$$A = \sum_{p=0}^{N} s_p u_p u_p^T, \quad s_p \in \mathbb{R}^+, \quad u_p \in \mathbb{C}^{N+1}, \qquad (10)$$

where $s_0 \geq s_1 \geq \cdots \geq s_N > 0$ are the decreasingly ordered con-eigenvalues of A and where the *con-eigenvectors* $\{u_p\}$ are orthogonal and satisfy the relation $A\overline{u_p} = s_p u_p$. Similarly to the Eckart Young theorem, it can be shown that the best rank $P \leq N + 1$ approximation of A is given by [12]

$$\sum_{p=0}^{P-1} s_p u_p u_p^T. \qquad (11)$$

Since g is a pure sum of P exponentials, Hg can be expressed in the form (11), and therefore each con-eigenvector u_p will also be a sum of the same exponentials. Let $U = (u_0, \ldots, u_{P-1})$. It is then possible to write $U = VG$, where V is the $N + 1 \times P$ Vandermonde matrix generated by e^{ζ_p} (i.e., $V(j,p) = e^{\zeta_p j}$) and G is some (invertible) $P \times P$ matrix. Denote as $\underline{U}$ (resp. $\overline{U}$) the matrix U whose first row (resp. last row) has been dropped. Clearly, we have $\underline{U} = \underline{V}G$, $\overline{U} = \overline{V}G$ and the Vandermonde structure of V leads to $\mathrm{diag}(e^{\zeta_0}, \ldots, e^{\zeta_{P-1}})\overline{V} = \underline{V}$. It follows that

$$(\overline{U}^*\overline{U})^{-1}(\overline{U}^*\underline{U}) = \overline{U}^\dagger \underline{U} = G^{-1}\mathrm{diag}(e^{\zeta_0}, \ldots, e^{\zeta_{P-1}})G,$$

where * stands for conjugate transpose. Therefore, we can compute the nodes $\boldsymbol{\zeta}$ by computing the eigenvalues of $\overline{U}^\dagger \underline{U}$.

4. ESTIMATION PERFORMANCE

We analyze the estimation performance of the proposed frequency estimation algorithm by considering different numerical simulations conducted with $P = 4$ and $P = 7$ complex exponentials embedded in circular white Gaussian noise (500 independent realisations) for different signal-to-noise ratio (SNR, defined as $10\log_{10}\left(\frac{energy_{signal}}{energy_{noise}}\right)$). The model parameters $\boldsymbol{c}$ and $\boldsymbol{\zeta}$ are summarized in Tab. 1 (units correspond to uniform sampling of the interval $t \in [-\frac{1}{2}, \frac{1}{2}]$). The performance is compared with the theoretical Cramér-Rao bounds (CRBs) for the estimation problem (see, e.g., [1, 2]).
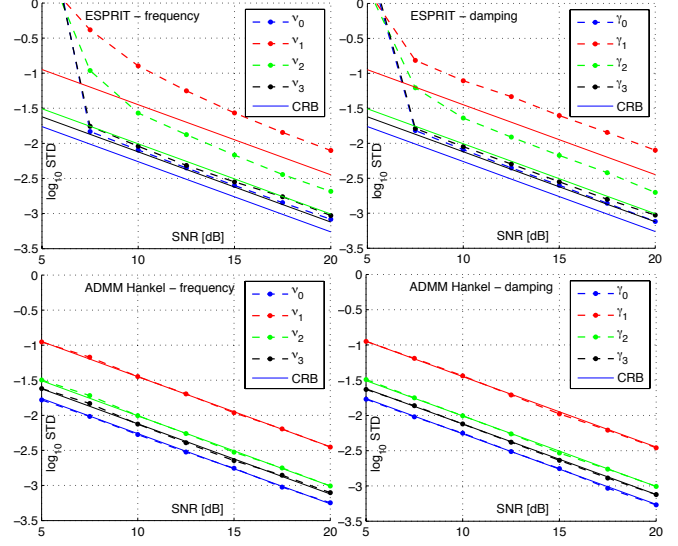

Fig. 1. Performance for the estimation of ν_p (left) and γ_p (right), $p = 0, \cdots, 3$, for ESPRIT (top) and the ADMM procedure (bottom). The CRBs are plotted in solid lines.

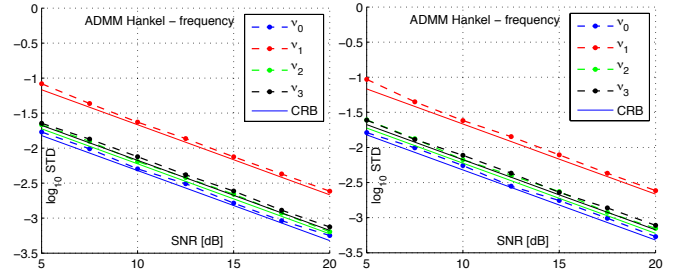**Fig. 2**. Performance for the estimation of ν_p, $p = 0, \cdots, 3$, for the ADMM procedure when 64 out of 321 samples are missing at random positions (left) and as one consecutive block (right). The solid lines indicate the CRBs.

Estimation performance for uniformly sampled data. In Figure 1, the standard deviations (STDs) of the estimates of γ_p and ν_p, $p = 0, \cdots, 3$, obtained with the ADMM procedure and with the ESPRIT method, are plotted as a function of SNR together with the (square root of the) CRBs for $2N + 1 = 257$ samples (bias is found to be much smaller than STD and not reported here). Note that the components $p = 1$ and $p = 2$ are close in frequency but significantly differ in amplitude (cf., Tab. 1). We observe that the proposed ADMM procedure provides estimates attaining the corresponding CRBs, indicating that the procedure realizes the minimum variance estimator. It consistently outperforms the ESPRIT method in terms of STD, and especially so for small SNR and for the two components that are close in frequency.
Estimation performance for missing data. In Fig. 2, we summarize the performance of the ADMM procedure for estimating ν_p, $p = 0, \cdots, 3$, for $2N + 1 = 257 + 64$ samples of which 64 samples are missing at random positions (left) and
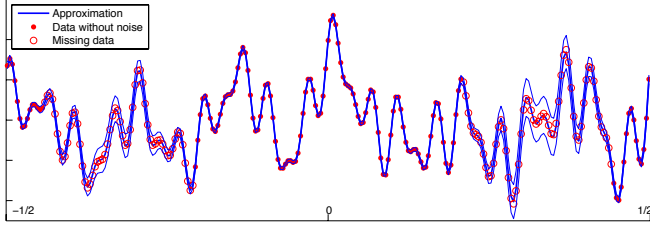
Fig. 3. Approximation of $P = 7$ exponentials for $2N + 1 = 257$ samples of which two blocks of 60 samples are missing ($SNR = 10dB$). Real parts of noise-free available samples (red dots) and missing samples (red circles); mean and ± 1.96 standard deviation error band (blue solid lines).

as one consecutive block centred at random position (right). The STDs are again found to reach the (square root of the) CRBs, demonstrating that the ADMM procedure is not affected by missing data. Similar results are obtained for γ_p and are not reported here for space limitation reasons.

Approximation for missing data.   Finally, Figure 3 illustrates the approximation performance of the ADMM procedure for $2N + 1 = 257$ samples of $P = 7$ exponentials of which two blocks of 60 consecutive samples are missing ($SNR = 10dB$, mean over 100 realizations). Despite the fact that 47% of the samples are missing, the obtained approximations are notably consistent with the underlying noise-free function. The maximum observed differences $|\hat{\nu}_p - \nu_p|$ are found to be below 10^{-3} times the sampling frequency.

5. CONCLUSIONS

A high-resolution parametric frequency estimation procedure that is based on approximation with sums of P complex exponentials was proposed. In the proposed algorithm, the Kronecker theorem was used to cast the approximation problem in terms of generating functions for Hankel matrices of rank P. The resulting optimization problem was addressed by an ADMM procedure. The Hankel matrices obtained from the ADMM procedure were then used to compute the parameters of the complex exponential model. This is in contrast to other methods, such as classical NLS or subspace methods. Although the optimization problem considered in this paper is non-convex (and the problems of local minima are inherited with non-convexity, as is the case for classical NLS) numerical simulations indicated that the ADMM procedure yields excellent practical performance.

The ADMM procedure can be applied to equally spaced samples, including situations with missing data, and does not rely on explicit noise model assumptions. Despite its versatility, the resulting algorithm is simple and easy to implement. The method has been presented here for complex-valued data. However, its real-valued counterpart can be obtained in a similar fashion as outlined in this work.

The procedure will be applied to frequency node estimation in nuclear quadrupole resonance applications. Future work includes the extension to multiple time series and to the estimation of directions of arrival.

6. REFERENCES

[1] E. Gudmundson, P. Wirfalt, A. Jakobsson, and M. Jansson, "An esprit-based parameter estimator for spectroscopic data," in *IEEE Statistical Signal Processing Workshop (SSP'12)*, 2012, pp. 77 – 80.

[2] P. Stoica and R. Moses, *Spectral analysis of Signals*, Prentice–Hall, 2005.

[3] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoustic, Speech and Signal Process.*, vol. 34, no. 5, pp. 1081–1089, 1986.

[4] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. on Antennas and Propagation*, vol. 34, no. 3, pp. 276 – 280, 1986.

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

[6] R. Kumaresan and D. W. Tufts, "Estimating the angles of arrival of multiple plane waves," *IEEE Trans. Aerospace and Electronic Systems*, vol. 19, pp. 134–139, 1983.

[7] G. Beylkin and L. Monzon, "On approximation of functions by exponential sums," *Applied and Computational Harmonic Analysis*, vol. 19, no. 1, pp. 17–48, July 2005.

[8] F. Andersson and M. Carlsson, "Alternating projections on non-tangential manifolds," *arXiv:1107.4055*, 2011.

[9] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[10] V. M. Adamjan, D. Z. Arov, and M. G. Kreĭn, "Infinite Hankel matrices and generalized problems of Carathéodory-Fejér and F. Riesz," *Funkcional. Anal. i Priložen.*, vol. 2, no. 1, pp. 1–19, 1968.

[11] F. Andersson, M. Carlsson, and M.V. de Hoop, "Sparse approximation of functions using sums of exponentials and aak theory," *Journal of Approximation Theory*, vol. 163, no. 2, pp. 213–248, February 2011.

[12] R. A. Horn and C. R. Johnson, *Topics in matrix analysis*, Cambridge University Press, Cambridge, 1994.