# Joint DOA and Multi-Pitch Estimation Via Block Sparse Dictionary Learning

Kronvall, Ted; Adalbjörnsson, Stefan Ingi; Jakobsson, Andreas

*Published in:*
European Signal Processing Conference

2014

# JOINT DOA AND MULTI-PITCH ESTIMATION VIA BLOCK SPARSE DICTIONARY LEARNING

*Ted Kronvall, Stefan Ingi Adalbjörnsson, and Andreas Jakobsson*

Centre for Mathematical Sciences, Lund University, Sweden.
email: {ted, sia, aj}@maths.lth.se

**ABSTRACT**

In this paper, we introduce a novel sparse method for joint estimation of the direction of arrivals (DOAs) and pitches of a set of multi-pitch signals impinging on a sensor array. Extending on earlier approaches, we formulate a novel dictionary learning framework from which an estimate is formed without making assumptions on the model orders. The proposed method alternatively uses a block sparse approach to estimate the pitches, using an alternating direction method of multipliers framework, and alternatively a nonlinear least squares approach to estimate the DOAs. The preferable performance of the proposed algorithm, as compared to earlier methods, is shown using numerical examples.

***Index Terms—*** multi-pitch estimation, group sparsity, block sparsity, dictionary learning, ADMM, direction-of-arrival.

## 1. INTRODUCTION

The estimation of fundamental frequencies, or pitches, of harmonically related, and often acoustic, signals is a common problem occurring in various forms of applications, and perhaps most notably so in audio processing (see, e.g., [1] and the references therein). Due to the importance of such applications, there have been notable contributions on pitch estimation for signals containing both single and multiple pitches (see e.g., [2–5]). By using an array of several sensors, one may exploit the relative time-delay information at the different sensors to determine the location of the impinging sound sources. Commonly, existing techniques, as the ones in, e.g., [6–8], make strong *a priori* assumptions on the model structure of the impinging signals, such as the number of pitches, as well as the number of harmonics in each pitch. Alternatively, model order information criterias may be used to determine the appropriate model order, such as in [9, 10], or by applying an optimal filtering approach reminiscent to the one proposed in [11]. In this work, we extend on the method presented in [5], and propose a novel joint DOA and pitch estimation technique, formed by using a novel sparse signal reconstruction framework. The technique is reminiscent to the one presented in [12], wherein the solution space is expanded to a large dictionary of candidate fundamental frequencies, from where a small number of pitches which have the best fit to the data are chosen. As the data is measured with several sensors, where each has a phase offset according the specific geometry of the array and the location of the sound source, both the pitches and the sensor phases must be estimated jointly. Such a joint estimation typically requires solving a non-convex optimization problem. Herein, we avoid this difficult by applying a dictionary learning technique, reminiscent to the ones presented in [13, 14]. We thereby split the problem into two subproblems, allowing for an iterative refinement of the pitch estimates, formed using an alternating direction method of multipliers (ADMM) framework, and of the DOA estimates, using a nonlinear least squares (NLS) formulation. The method allows for the estimation of the DOAs and pitches from multi-pitch signals originating from one or more locations, without having to know the number of sources, pitches, or their respective number of harmonics. Our claims are illustrated using numerical simulations of audio signals, comparing the achieved performance to other recent estimators.

This work was supported in part by the Swedish Research Council, Crafoord's and Carl Trygger's foundations, and the Royal Physiographic Society in Lund.

## 2. THE PITCH-DOA SIGNAL MODEL

Consider $K$ complex-valued and harmonically related acoustic signals impinging on an array of sensors, corrupted by additive noise and interference, such that the signal measured at the $m$th sensor may be well modelled as [6, 15]

$$y_m(t) = \sum_{k=1}^{K} \sum_{\ell=1}^{L_k} c_m d_{k,\ell} e^{j\omega_k \ell (t+\tau_{k,m})} + e_m(t) \tag{1}$$

where $d_{k,l}$ is the complex-valued amplitude of the $\ell$th harmonic of the $k$th pitch, whereas $L_k$ and $\omega_k$ are the number of harmonics and the pitch of the $k$th signal source, respectively. Furthermore, let $e_m(t)$ denote the additive noise term, $c_m$ the sensor gain, and $\tau_{k,m}$ the time-of-arrival for the $k$th signal source. Define the measurement matrix

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}(1) & \dots & \mathbf{y}(N) \end{bmatrix}^T \tag{2}$$

where, at each time point, $n = 1, \ldots, N$, the data snapshot is found as

$$\mathbf{y}(t) = \begin{bmatrix} y_0(t) & \ldots & y_{M-1}(t) \end{bmatrix}^T$$

with $(\cdot)^T$ denoting the transpose. Then, (2) may be concisely expressed as

$$\mathbf{Y} = \sum_{k=1}^{K} \mathbf{W}_k \text{diag}(\mathbf{d}_k) \mathbf{F}_k(\boldsymbol{\tau}_k) \mathbf{C} + \mathbf{E} \tag{3}$$

where $\mathbf{E}$ denotes the combined noise term constructed in the same manner as $\mathbf{Y}$, and

$$\mathbf{W}_k = \begin{bmatrix} \mathbf{w}_k & \ldots & \mathbf{w}_k^{L_k} \end{bmatrix} \tag{4}$$

$$\mathbf{w}_k = \begin{bmatrix} e^{j\omega_k} & \ldots & e^{j\omega_k N} \end{bmatrix}^T \tag{5}$$

$$\mathbf{d}_k = \begin{bmatrix} d_{k,1} & \ldots & d_{k,L_k} \end{bmatrix}^T \tag{6}$$

$$\mathbf{F}_k(\boldsymbol{\tau}_k) = \begin{bmatrix} e^{j\omega_k \tau_{k,1}} & \ldots & e^{j\omega_k \tau_{k,M}} \\ e^{j\omega_k 2\tau_{k,1}} & \ldots & e^{j\omega_k 2\tau_{k,M}} \\ \vdots & \ddots & \vdots \\ e^{j\omega_k L_k \tau_{k,1}} & \ldots & e^{j\omega_k L_k \tau_{k,M}} \end{bmatrix} \tag{7}$$

$$\boldsymbol{\tau}_k = \begin{bmatrix} \tau_{k,1} & \ldots & \tau_{k,M} \end{bmatrix}^T \tag{8}$$

$$\mathbf{C} = \text{diag}\left( \begin{bmatrix} c_1 & \ldots & c_M \end{bmatrix} \right) \tag{9}$$

such that $\text{diag}(\cdot)$ is a diagonal matrix. One may note that $\mathbf{W}_k$, for $k = 1, \ldots, K$, consists of stacked Fourier vectors, for each harmonic of a pitch in the temporal domain, whereas $\mathbf{F}_k$ consists of stacked Fourier vectors (or array transfer vectors) in the spatial domain with respect to the time-of-arrivals, $\boldsymbol{\tau}_k$, repeated for each pitch $k$ and its $L_k$ harmonics. We proceed to reformulating the problem in (3) using a sparse estimation framework, reminiscent to the one presented in [12], extending the representation of the $K$ pitches onto a large dictionary of $P$ candidate fundamental frequencies, $\omega_1, \ldots, \omega_P$, where $P \gg K$, chosen so large that $K$ of these will reasonably well coincide with the true pitches in the signal. In the same fashion, the number of harmonics of each pitch, $L_p$, is extended to an arbitrary upper level, say $L_{\text{max}}$, for all dictionary elements, $p = 1, \ldots, P$. One can, without loss of generality, assume $\mathbf{C} = \mathbf{I}$, i.e., that the data measurement matrix has been preconditioned to account for different gain at different sensors. The signal model may thus be expressed as

$$\begin{aligned} \mathbf{Y} &= \sum_{p=1}^{P} \mathbf{W}_p \text{diag}(\mathbf{d}_k) \mathbf{F}_p(\tau_p) + \boldsymbol{\mathcal{E}} && (10) \\ &= \boldsymbol{\mathcal{W}} \, \text{diag}(\mathbf{d}) \, \boldsymbol{\mathcal{F}}(\boldsymbol{\tau}) + \boldsymbol{\mathcal{E}} && (11) \end{aligned}$$

where the block dictionary matrices are formed by stacking the matrices such that

$$\begin{aligned} \boldsymbol{\mathcal{W}} &= \begin{bmatrix} \mathbf{W}_1 & \ldots & \mathbf{W}_P \end{bmatrix} && (12) \\ \boldsymbol{\mathcal{F}}(\boldsymbol{\tau}) &= \begin{bmatrix} \mathbf{F}_1(\tau_1)^T & \ldots & \mathbf{F}_P(\tau_P)^T \end{bmatrix}^T && (13) \end{aligned}$$

where $\boldsymbol{\mathcal{W}} \in \mathbb{C}^{N \times PL_{\text{max}}}$, $\boldsymbol{\mathcal{F}}(\boldsymbol{\tau}) \in \mathbb{C}^{PL_{\text{max}} \times M}$, and

$$\begin{aligned} \mathbf{d} &= \begin{bmatrix} \mathbf{d}_1{}^T & \ldots & \mathbf{d}_P{}^T \end{bmatrix}^T && (14) \\ \boldsymbol{\tau} &= \begin{bmatrix} \boldsymbol{\tau}_1 & \ldots & \boldsymbol{\tau}_P \end{bmatrix}^T && (15) \end{aligned}$$

with $\mathbf{d} \in \mathbb{C}^{PL_{\text{max}} \times 1}$ and $\boldsymbol{\tau} \in \mathbb{R}^{P \times M}$. The resulting signal formulation provides a more structured framework than the one presented in [15], separating the complex-valued amplitudes, $\mathbf{d}$, and the sensor offsets in $\boldsymbol{\mathcal{F}}(\boldsymbol{\tau})$. If the sensor array is assumed to be a uniform linear array (ULA), the time-of-arrivals may be related to the corresponding DOA as [9]

$$\tau_{k,m} = (m-1)\delta \sin(\theta_k) \gamma^{-1} \tag{16}$$

with $\delta, \gamma$, and $\theta$ denoting the uniform distance between sensors, the wave propagation velocity, and the DOA respectively. The $P \times M$ time-of-arrivals may thus be expressed as a function of the set of DOAs

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \ldots & \theta_P \end{bmatrix}^T \tag{17}$$

In the interest of notational simplicity, we hereafter use only the dependency of $\boldsymbol{\theta}$ instead of $\boldsymbol{\tau}(\boldsymbol{\theta})$. For other array geometries, one may replace (16) with another function mapping from directionality or location to the time-of-arrival.

## 3. DICTIONARY LEARNING APPROACH

In order to form the estimate of the unknown DOAs and pitches, we formulate the estimates as the solution to a group sparse minimization reminiscent to the scheme presented in [5], such that

$$\begin{aligned} \underset{\boldsymbol{\theta}, \mathbf{d}}{\text{minimize}} \;\; & \frac{1}{2} \left\| \mathbf{Y} - \boldsymbol{\mathcal{W}} \, \text{diag}(\mathbf{d}) \, \boldsymbol{\mathcal{F}}(\boldsymbol{\theta}) \right\|_{\mathcal{F}}^2 \\ & + \lambda\mu \sum_{p=1}^{P} \left\| \mathbf{d}_k \right\|_2 + \lambda(1-\mu) \left\| \mathbf{d} \right\|_1 \end{aligned} \tag{18}$$

where block sparsity is imposed on $\mathbf{d}$, such that the number of pitches, as well as the number of harmonics within each pitch, are sparse. Here, we set $\lambda > 0$ as a parameter weighting the degree of sparsity to the fit of the solution, while $\mu \in [0\,,\,1]$ prioritizes between sparsity and block sparsity. In order to simplify the minimization, one may formulate (18) equivalently as

$$\begin{aligned} \underset{\boldsymbol{\theta}, \mathbf{d}}{\text{minimize}} \;\; & \frac{1}{2} \sum_{m=1}^{M} \left\| \mathbf{y}_m - \boldsymbol{\mathcal{W}} \, \text{diag}\left( \boldsymbol{f}_m(\boldsymbol{\theta}) \right) \mathbf{d} \right\|_2^2 \\ & + \lambda\mu \sum_{p=1}^{P} \left\| \mathbf{d}_k \right\|_2 + \lambda(1-\mu) \left\| \mathbf{d} \right\|_1 \end{aligned} \tag{19}$$

such that the minimization is formed by summing the squared residual errors sensor by sensor, where $\boldsymbol{f}_m(\cdot)$ is the $m$th column of $\boldsymbol{\mathcal{F}}(\cdot)$, and where we have used that $\text{diag}\left( \boldsymbol{f}_m(\boldsymbol{\theta}) \right) \mathbf{d} =$

**Algorithm 1** The IAPEBS algorithm

1: Initiate $\mathbf{d}^{(0)}$ by taking steps 4-11 for data $\mathbf{y}_1$ only.
2: Set $k = 0$
3: **repeat** {Dictionary learning scheme}
4: Take NLS step $\boldsymbol{\theta}^{(k+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, q(\boldsymbol{\theta}, \mathbf{d}^{(k)})$
5: Initiate $\mathbf{u}^{(0)} = \mathbf{d}^{(k)}, \mathbf{z}^{(0)} = \mathbf{z}^{(\text{save})}, i = 0$
6: **repeat** {ADMM scheme}
7: $\mathbf{z}^{(i+1)} = \underset{\mathbf{z}}{\operatorname{argmin}}\, L_\kappa(\mathbf{z}, \mathbf{u}^{(i)}, \mathbf{d}^{(i)})$
8: $\mathbf{u}^{(i+1)} = \underset{\mathbf{u}}{\operatorname{argmin}}\, L_\kappa(\mathbf{z}^{(i+1)}, \mathbf{u}, \mathbf{d}^{(i)})$
9: $\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} - (\mathbf{z}^{(i+1)} - \mathbf{u}^{(i+1)})$
10: $i \leftarrow i + 1$
11: **until** convergence
12: Set $\mathbf{d}^{(k+1)} = \mathbf{u}^{(\text{end})}$, and $\mathbf{z}^{(\text{save})} = \mathbf{z}^{(\text{end})}$
13: $k \leftarrow k + 1$
14: **until** convergence

$\operatorname{diag}(\mathbf{d})\boldsymbol{f}_m(\boldsymbol{\theta})$. However, solving (19) is a hard problem, as $\boldsymbol{f}(\cdot)$ is a non-convex function of $\boldsymbol{\theta}$, as is its product with $\mathbf{d}$. On the other hand, for a fixed $\boldsymbol{\theta}$, the minimization is the ordinary LASSO with block sparsity for complex sinusoids (see, e.g., [16]), where $\boldsymbol{\mathcal{W}} \operatorname{diag}(\boldsymbol{f}_m(\boldsymbol{\theta}))$ may be seen as a phase-shifted dictionary at sensor $m$ with respect to the corresponding DOA. Adopting a dictionary learning framework reminiscent to the one used in [13, 14], the problem is split in two sub-problems. In the first, we fix the DOAs, and (19) may be solved via one of the freely available interior point solvers, such as SeDuMi [17] and SDPT3 [18]. However, such solvers will typically scale poorly with increasing data length, the use of a finer grid of candidate pitches, and/or the number of sensors. Such methods may thus in many cases be computationally cumbersome, and we here introduce an efficient ADMM-based formulation of (19). To do so, one splits the objective function into two parts, where we let one contain the squared residual error, and the second the sparsity constraints, whereafter an auxiliary variable is introduced, such that

$$\underset{\mathbf{z},\mathbf{u}}{\text{minimize}}\; g_1(\mathbf{z}) + g_2(\mathbf{u}) \text{ subj. to } \mathbf{z} - \mathbf{u} = \mathbf{0} \tag{20}$$

since only $\mathbf{z} = \mathbf{u}$ is a feasible point, and where

$$g_1(\mathbf{z}) = \frac{1}{2}\sum_{m=1}^{M} \left\| \mathbf{y}_m - \boldsymbol{\mathcal{W}} \operatorname{diag}\left(\boldsymbol{f}_m(\boldsymbol{\theta})\right) \mathbf{z} \right\|_2^2 \tag{21}$$

$$g_2(\mathbf{u}) = \lambda\mu \sum_{p=1}^{P} \left\|\mathbf{u}_k\right\|_2 + \lambda(1-\mu)\left\|\mathbf{u}\right\|_1 \tag{22}$$

are convex functions. Under the assumption that there is no duality gap, which, for a fixed $\boldsymbol{\theta}$, is true for (18), one may solve the optimization problem via the dual function, defined as the infimum of the augmented Lagrangian with respect to

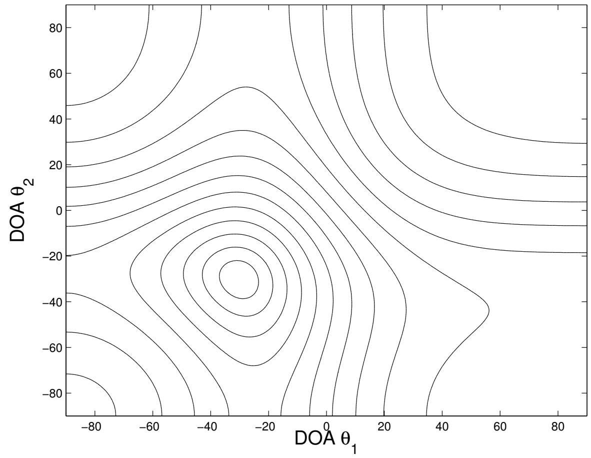


**Fig. 1**. Level curves for the function in (28), for a multi pitch signal containing two pitches, both originating from $-30°$.

$\mathbf{z}$ and $\mathbf{u}$, i.e., [19]

$$L_\kappa(\mathbf{z}, \mathbf{u}, \mathbf{x}) = g_1(\mathbf{z}) + g_2(\mathbf{u}) + \mathbf{x}^T(\mathbf{z} - \mathbf{u}) + \frac{\kappa}{2}||\mathbf{z} - \mathbf{u}||_2^2$$

where $\mathbf{x}$ is the dual variable. The ADMM method solves this iteratively by, at step $i+1$, minimizing the Lagrangian for one primal variable while holding the other fixed at its previous value, and then updating the dual variable by taking a gradient ascent step and maximizing the dual function, i.e.,

$$\mathbf{z}^{(i+1)} = \arg\min_{\mathbf{z}} L_\kappa(\mathbf{z}, \mathbf{u}^{(i)}, \mathbf{d}^{(i)}) \tag{23}$$

$$\mathbf{u}^{(i+1)} = \arg\min_{\mathbf{u}} L_\kappa(\mathbf{z}^{(i+1)}, \mathbf{u}, \mathbf{d}^{(i)}) \tag{24}$$

$$\tilde{\mathbf{x}}^{(i+1)} = \tilde{\mathbf{x}}^{(i)} - \kappa(\mathbf{z}^{(i+1)} - \tilde{\mathbf{x}}^{(i+1)}) \tag{25}$$

where $\kappa$ is the step size for maximizing the dual function, and $\tilde{\mathbf{x}} = \mathbf{x}/\kappa$ is the scaled version of the dual variable, which is more convenient for implementation (see [19] for further details on these aspects). The function in (23), which is quadratic, can be solved in closed form as

$$\mathbf{z}^{(i+1)} = \left(\sum_{m=1}^{M} \tilde{\mathbf{W}}_m^H \tilde{\mathbf{W}}_m + \kappa \mathbf{I}_{PL_{\max}}\right)^{-1} \times \left(\sum_{m=1}^{M} \tilde{\mathbf{W}}_m^H \mathbf{y}_m + \mathbf{u}^{(i)} + \mathbf{x}^{(i)}\right) \tag{26}$$

where $\tilde{\boldsymbol{\mathcal{W}}}_m = \boldsymbol{\mathcal{W}} \operatorname{diag}\left(\boldsymbol{f}_m(\boldsymbol{\theta})\right)$ denotes the phase-shifted dictionary at sensor $m$. The function in (23), i.e., the primal variable for the sparsity constraints, is obtained by solving sub-differential equations, yielding

$$\mathbf{u}^{(i+1)} = h\left(h'\left(\mathbf{z}^{(i+1)} - \mathbf{x}^{(i)}, \lambda\mu\right), \lambda(1-\mu)\right) \tag{27}$$

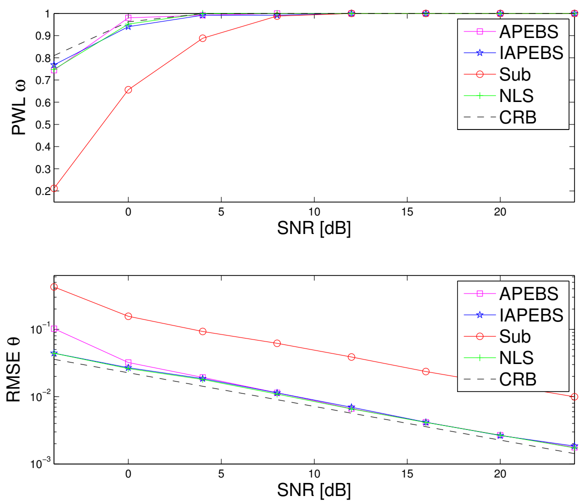


**Fig. 2**. The PWL and RMSE for a single-pitch signal as compared with the optimal performance of an estimator reaching the CRB.

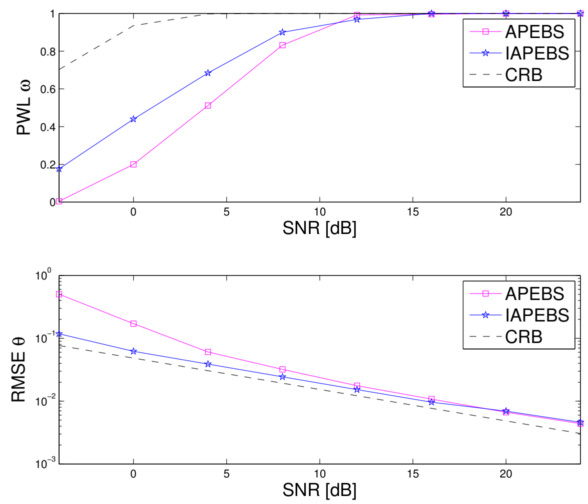


**Fig. 3**. The PWL and RMSE for a multi-pitch signal with two pitches, as compared to the corresponding CRB.

where $h(\mathbf{b}, \xi) = \mathbf{b}\left(1 - \xi / \|\mathbf{b}\|_2\right)^+$, for a vector $\mathbf{b}$ and a positive scalar $\xi$, with $(\cdot)^+$ denoting the identity function for finite values and zero otherwise, and $h'(\cdot)$ defined similarly but operate element-wise on $\mathbf{b}$ (see also [5]). The resulting estimate of $\mathbf{d}^{(k)}$ is then inserted into the second subproblem of the dictionary learning scheme, i.e.,

$$q(\boldsymbol{\theta}, \mathbf{d}^{(k)}) = \frac{1}{2}\left\|\mathbf{Y} - \mathcal{W}\,\mathrm{diag}(\mathbf{d}^{(k)})\,\mathcal{F}(\boldsymbol{\theta})\right\|_{\mathcal{F}}^2 \tag{28}$$

which is minimized for $\boldsymbol{\theta}$, and is equivalent to performing a dictionary learning update to the phase-shifted dictionary, $\tilde{\mathcal{W}}_m$, which was used in the ADMM procedure, i.e., (20)-(27). Figure 1 illustrates the cost function in (28) after a few dictionary learning iterations of the proposed algorithm, showing that although the cost function will not be convex, it is unimodal for DOAs in the range $[-90, 90]^\circ$ and may thus be easily solved using a few iterations of, for instance, Newton-Raphson's method. To summarize, an algorithm outline of the proposed metod is stated in Algorithm 1, where it may be noted that the inner ADMM scheme takes fewer and fewer steps at every iteration of the outer dictionary learning scheme, until convergence is reached and only a single ADMM step is taken. The sparsity parameter $\lambda$ is chosen with cross validation in a similar fashion as performed in [20], but the estimates are rather unsensitive with respect to this choice. The proposed method requires estimating a total of $PL_{\max} + M$ parameters, which is considerably fewer than the recent sparse method presented in [15], which required estimating $PL_{\max}M$ parameters.

## 4. NUMERICAL RESULTS

We proceed to illustrate the performance of our proposed method, as compared to other recent methods, using synthetic audio signals. As the fundamental frequencies are estimated on a discrete dictionary grid, comparison is made using a percentage within limits (PWL) metric, defined as the ratio of pitch estimates within a range of $\pm 1/4$ Hz from the true values. For DOA comparison, the total root mean square error (RMSE) is used for all sources, defined as

$$\mathrm{RMSE}_\theta = \sqrt{\frac{1}{nK}\sum_{k=1}^{K}\sum_{i=1}^{n}\left(\hat{\theta}_{k,i} - \theta_k\right)^2} \tag{29}$$

where n is the number of Monte Carlo (MC) simulation estimates, and K is the number of pitches in the signal. Figure 2 shows the PWL of the fundamental frequency, as well as the RMSE for the DOA, for a signal containing a single pitch with $f_1 = 220$ Hz and $L_1 = 7$ harmonics, impinging on a 5-sensor ULA from direction $\theta_1 = -30°$. These results have been computed using 250 MC simulations, assuming a sampling frequency of $f_s = 8820$ Hz, a sound wave propagation velocity of $\gamma = 324.3$ m/s, and a sensor spacing of $\delta = \gamma / f_s = 3.84$ cm. The sensor gains may be obtained from a covariance matrix estimate on the measurement matrix $\mathbf{Y}$, but are, in these simulations and without loss of generality, set to $c_1 = \cdots = c_M = 1$. The figures show the performance for growing signal-to-noise ratios (SNRs), defined as

$$\mathrm{SNR} = 10 \cdot \log\left(\frac{P_{\text{signal}}}{P_{\text{noise}}}\right) \quad (\mathrm{dB}) \tag{30}$$

As is clear from the figures, the proposed method, here termed the iterative array DOA and pitch estimator using block sparsity (IAPEBS), performs similarly to the recently proposed APEBS estimator [15], and the NLS-based estimator proposed in [6], achieving a performance close to the Cramér-Rao lower bound (CRB). The subspace-based method (Sub), also introduced in [6], is found to yield a somewhat lower performance. Figure 3 shows the corresponding performance for a multi-pitch signal consisting of two pitches, with $[\omega_1, \omega_2] = [150, 220]$ Hz, and with $[L_1, L_2] = [7, 6]$ harmonics, impinging from directions $[\theta_1, \theta_2] = [-30, -30]^{\circ}$. As the NLS and Sub estimators only allow for single pitch signals, the figure only shows the performance of IAPEBS, as compared with APEBS and the corresponding CRB. As is clear from the figures, the IAPEBS estimator yields highly accurate parameter estimates, almost reaching the CRBs, notably improving the achievable performance as compared to the APEBS estimator, which decouples the estimation into first estimating the pitches, whereafter the DOAs are determined in a second step. This should be compared with the here proposed iterative estimation scheme, which enables a better joint estimation of pitch and DOA.

## 5. ACKNOWLEDGEMENTS

We would like to express our most sincere gratitude towards the authors of [6] for sharing their Matlab implementations with us.

## 6. REFERENCES


[1] M. Christensen and A. Jakobsson, *Multi-Pitch Estimation*, Morgan & Claypool, 2009.

[2] L. Rabiner, "On the use of autocorrelation analysis for pitch detection," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 25, no. 1, pp. 24–33, 1977.

[3] E. Benetos and S. Dixon, "Joint Multi-Pitch Detection Using Harmonic Envelope Estimation for Polyphonic Music Transcription," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 6, pp. 1111–1123, Oct. 2011.

[4] C. Lee, Y. Yang, and H. H. Chen, "Multipitch Estimation of Piano Music by Exemplar-Based Sparse Representation," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 608–618, 2012.

[5] S. I. Adalbjörnsson, A. Jakobsson, and M. G. Christensen, "Estimating Multiple Pitches Using Block Sparsity," in *38th IEEE Intern. Conf. on Acoustics, Speech, and Signal Processing*, 2013.

[6] J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Nonlinear Least Squares Methods for Joint DOA and Pitch Estimation," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 21, no. 5, pp. 923–933, 2013.

[7] S. Gerlach, S. Goetze, J. Bitzer, and S. Doclo, "Evaluation of joint position-pitch estimation algorithm for localising multiple speakers in adverse acoustical environments," in *Proc. German Annual Conference on Acoustics (DAGA)*, Düsseldorf, Germany, 2011, vol. Mar. 2011, pp. 633–634.

[8] J. X. Zhang, M. G. Christensen, S. H. Jensen, and M. Moonen, "Joint DOA and Multi-pitch Estimation Based on Subspace Techniques," *EURASIP J. on Advances in Signal Processing*, vol. 2012, no. 1, pp. 1–11, 2012.

[9] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, Upper Saddle River, N.J., 2005.

[10] M. G. Christensen, A. Jakobsson, and S. H. Jensen, "Sinusoidal Order Estimation using Angles between Subspaces," *EURASIP Journal on Advances in Signal Processing*, 2009, Article ID 948756, 11 pages.

[11] M. G. Christensen, J. .L Højvang, A. Jakobsson, and S. H. Jensen, "Joint fundamental frequency and order estimation using optimal filtering," *EURASIP Journal on Advances in Signal Processing*, vol. 13, pp. 1–18, 2011.

[12] J. J. Fuchs, "On the Use of Sparse Representations in the Identification of Line Spectra," in *17th World Congress IFAC*, Seoul,, jul 2008, pp. 10225–10229.

[13] C. D. Austin, J. N. Ash, and R. L. Moses, "Dynamic Dictionary Algorithms for Model Order and Parameter Estimation," *IEEE Transactions on Signal Processing*, vol. 61, no. 20, pp. 5117–5130, October 2013.

[14] I. Tosic and P. Frossard, "Dictionary Learning," *Signal Processing Magazine, IEEE*, vol. 28, no. 2, pp. 27–38, March 2011.

[15] T. Kronvall, S. Adalbjörnsson, and A. Jakobsson, "Joint DOA and Multi-Pitch Estimation Using Block Sparsity," in *Proceedings of the 39th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence, 2014.

[16] J. Friedman, T. Hastie, and R. Tibshirani, "A Note on the Group Lasso and a Sparse Group Lasso," Unpublished manuscript, 2010.

[17] J. F. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11-12, pp. 625–653, August 1999.

[18] R. H. Tutuncu, K. C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," *Mathematical Programming Ser. B*, vol. 95, pp. 189–217, 2003.

[19] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[20] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society B*, vol. 58, no. 1, pp. 267–288, 1996.