LUND UNIVERSITY

Off-grid Fundamental Frequency Estimation

J. SWÄRD, H. LI, AND A. JAKOBSSON

Published in: IEEE Transactions on Audio, Speech and Language Processing
doi:10.1109/TASLP.2017.2775800

Lund 2017

Mathematical Statistics
Centre for Mathematical Sciences
Lund University

Off-grid Fundamental Frequency Estimation

Johan Swärd, *Student Member, IEEE*, Hongbin Li, *Senior Member, IEEE*, and Andreas Jakobsson, *Senior Member, IEEE*

***Abstract*—In this paper, we propose a gridless method for estimating an unknown number of fundamental frequencies. Starting with a conventional dictionary matrix, containing sets of candidate fundamental frequencies and their corresponding harmonics, a non-convex log-sum cost function is formed such that it imposes the harmonic structure and treats every fundamental frequency in the dictionary as a parameter. The cost function is iteratively decreased by minimizing a surrogate function, and, in each iteration, the fundamental frequencies are refined, whereas redundant parameters are omitted from the dictionary. The proposed method is tested on both real and simulated data, showing its preferred performance as compared to other state-of-the-art multi-pitch estimators.**

I. Introduction

In areas such as audio, biomedicine, and mechanics, the estimation of fundamental frequencies is often of central importance. In particular, the multi-pitch problem is challenging, as one needs to determine not only the number of fundamental frequencies, but also the number of harmonics related to each fundamental frequency. This problem has historically been addressed by utilizing various forms of model order estimators, or by simply assuming the model order is already known *a-priori* [1]–[4]. Early pitch estimation methods relied on covariance-based methods as the ones presented in [5,6]. Later, filterbank- and subspace-based methods were introduced and MUSIC-like methods were widely used [7]–[12]. Recent contributions include, e.g., [13], where the computational speed is in focus, and [14] where the problem is to estimate the fundamental frequencies in real noise when multiple people speak at the same time. In [15], the Pitch Estimation using ℓ_2 norm and Block Sparsity (PEBS) algorithm was presented, where the fundamental frequency estimation problem was instead solved by using a (block-)sparsity approach, thereby combining the model order estimation with the overall estimation of the fundamental frequencies and their harmonics. Based on the promising performance of the initial PEBS algorithm, several improvements have been suggested, including focusing on the choice of hyperparameters [16], time-updating [17], and computation complexity [18]. The results presented in these works illustrate the benefits of using a sparse framework for solving multi-pitch estimation problems.

Sparse reconstruction methods are used in a vast number of areas and have been intensively studied (see, e.g., [19]–[24]). As in the case of PEBS, the resulting sparse problems have often been expressed using dictionary matrices, containing a large quantity of possible signal candidates, with the assumption that only a small subset of these candidates is needed to approximate the signal well. These candidates are often selected on a pre-defined grid that spans the parameter space of interest. Recently, some concerns have been raised as to how this grid-based selection of candidates affects the performance. In [25], it was shown that since the grid and the true parameters are unlikely to coincide, this may cause the estimation to deteriorate. If one, in an effort to circumvent this, increases the number of grid points to decrease the distance between the grid and the true parameters, the dictionary matrix will become increasingly coherent, i.e., the columns of the dictionary matrix become correlated, which may in turn degrade the performance, and increase the computational complexity of the algorithm. To counter these drawbacks, it has recently been suggested that one may instead solve the sparse problem without applying a grid, using so-called gridless methods. One noticeable example of this is the use of the atomic norm [26]–[30], where the sparse problem is instead formulated as a convex semi-definite program (SDP). The use of the atomic norm can be seen as solving the sparse problem using an infinite grid, but without the problem of a resulting coherent dictionary matrix. Unfortunately, the atomic norm formulation does not easily allow for imposing general data structures to the cost function, and, typically, any additional model constraints will fundamentally change the problem formulation. This is in contrast to the grid-based approaches, where such model structures could easily be accounted for by adding different constraints to the cost function.

In this paper, we aim to combine the benefits of the off-grid methods with the use of a cost function that easily allows for adding structure to the signal of interest. To this effect, we will expand on the PEBS formulation and introduce a method for solving problems involving group sparsity with sparse groups based on the super-resolution iterative reweighted (SURE-IR) method [31]. We then proceed to adress both the computational complexity issue as well as the appropriate choice of hyperparameters for the introduced estimator. Using both simulated and real audio data, we illustrate the preferable performance of the introduced estimator, comparing to several earlier alternative formulations. For the real data case, we test the proposed method using the Bach10 data set, containing 10 musical pieces composed by Johann Sebastian Bach, showing that the proposed method achieves similar performance as state-of-the-art music transcription methods, although without the need of *any* training data, as is typically utilized by such methods.

It should be noted that the proposed method is not limited to audio problems, although this is here the main focus. Indeed, due to the possibility of adding new constraints to the cost

Department of Mathematical Statistics, Lund University, 221 00 Lund, Sweden (emails: js@maths.lth.se, hli@stevens.edu, aj@maths.lth.se). This work was supported in part by the Swedish Research Council and the Crafoord's and Carl Trygger's foundations.

function, the technique may likely be extended to find use in other related fields, such studies of mechanical vibrations (see, e.g., [32]–[35]).

II. Signal model and earlier work

Consider the multi-pitch signal model[1]

$$y(n) = \sum_{k=1}^{K} \sum_{\ell=1}^{L_k} \alpha_{k,\ell} e^{2i\pi f_k \ell t_n} + \epsilon(n) \tag{1}$$

where f_k denotes the kth fundamental frequency (also denoted pitch), $\alpha_{k,\ell}$ the complex amplitude corresponding to the ℓth overtone of the kth fundamental frequency, t_n, for $n = 1, \ldots, N$, the nth time point, and $\epsilon(n)$ any non-tonal audio or noise component, here, for simplicity, being modelled as a complex-valued white Gaussian noise (see also [36]). Often, the problem of interest is that of estimating f_k for $k = 1, \ldots, K$. If this set is known, as well as the number of overtones for each pitch, L_k, the corresponding amplitudes of the overtones may be formed, for instance, using least squares (LS).

Typically, it is non-trivial to determine the required model orders; for simplicity, we will initially consider the problem of only estimating K sinusoids in noise. This corresponds to the case where $L_k = 1$ for all k. To form an efficient estimator, one may then include the model order estimation into the estimation of the frequencies, for instance by forming the sparse optimization problem (see also [37])

$$\underset{\mathbf{z}}{\text{minimize}} \; ||\mathbf{y} - \mathbf{A}\mathbf{z}||_2^2 + \lambda ||\mathbf{z}||_1 \tag{2}$$

where $\mathbf{A}$ is a dictionary matrix, $\mathbf{z}$ a vector containing the complex amplitudes, λ is a hyperparameter that controls the amount of sparsity in the solution, and

$$\mathbf{y} = \begin{bmatrix} y(1) & \ldots & y(n) \end{bmatrix}^T \tag{3}$$

Usually, the dictionary, $\mathbf{A}$, is an $N \times M$ matrix containing $M \gg N$ signal candidates (in this case sinusoids). Thus,

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_M \end{bmatrix} \tag{4}$$

where $\mathbf{a}_k = \begin{bmatrix} e^{2i\pi f_k t_1} & \ldots & e^{2i\pi f_k t_N} \end{bmatrix}^T$.

The first part of (2) is thus a data-fitting term, whereas the second term is a sparsity enhancing term, penalizing the magnitude of $\mathbf{z}$, thus promoting a sparse solution, containing only a few signal candidates. This methodology is widely used in signal processing and has been popular for many years (see, e.g., [19]). However, it has in recent times been argued that using a pre-defined grid may cause the estimation to deteriorate, mainly because of the fact that the true parameter value will typically not exactly coincide with any of the grid points. Trying to increase the grid size, in an effort to minimize the distance from the grid points to the true values, may further harm the estimation as the dictionary matrix then becomes more coherent. To address this issue, a gridless method based on the use of the atomic norm was proposed in [27]. Instead of solving a problem based on a dictionary matrix, the authors proposed the gridless formulation (for the noiseless case)

$$\begin{aligned} &\underset{x,\boldsymbol{u}}{\text{minimize}} \; \frac{1}{2}(x + u_1) \\ &\text{subject to} \; \begin{bmatrix} x & \mathbf{y}^H \\ \mathbf{y} & T(\boldsymbol{u}) \end{bmatrix} \geq 0 \end{aligned} \tag{5}$$

where $T(\boldsymbol{u})$ forms a Hermitian Toeplitz matrix with the vector $\boldsymbol{u}$ on its first row, and where u_1 denotes the first element in $\boldsymbol{u}$. The corresponding frequencies are then obtained using a Vandermonde decomposition of $T(\boldsymbol{u}^*)$, where $\boldsymbol{u}^*$ denotes the value of $\boldsymbol{u}$ at the solution of (5). The atomic norm enjoys many benefits (for a more detailed discussion on the topic, see, e.g., [26]–[30]), but it is generally hard to generalize the method to accommodate for other model restrictions, such as block sparsity or, e.g., spectral smoothness [38]. As an alternative, another gridless approach was suggested in [31], which was based on the formulation of a non-convex optimization problem. The proposed problem utilized a logarithmic penalty to enforce sparsity, such that

$$\underset{\mathbf{z},\boldsymbol{\theta}}{\text{minimize}} \; ||\mathbf{y} - \mathbf{A}(\boldsymbol{\theta})\mathbf{z}||_2^2 + \lambda \sum_{m=1}^{M} \log\left(|z_m|^2 + \eta\right) \tag{6}$$

where $\eta > 0$ is a parameter ensuring that the function is not evaluated at zero, and z_m denotes the mth element of $\mathbf{z}$. It should be noted that the dictionary matrix is now parameterized over the parameter vector $\boldsymbol{\theta}$, containing the sought fundamental frequencies. Thus, instead of using a fixed grid, the grid points are selected as to minimize the cost function in (6). Using a logarithmic penalty will enhance the sparsity, but, at the same time, render the problem non-convex. To solve the problem, a majorization-minimization (MM) approach was proposed in [31] and the optimization problem was reformulated using a surrogate function, thus yielding a simplified version of the original problem. This allows the problem to be solved using an analytic solution of the amplitudes as a function of $\boldsymbol{\theta}$, such that

$$\mathbf{z}^*(\boldsymbol{\theta}) = \left(\mathbf{A}^H(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) + \lambda \mathbf{D}^{(i)}\right)^{-1} \mathbf{A}^H(\boldsymbol{\theta})\mathbf{y} \tag{7}$$

where

$$\mathbf{D}^{(i)} = \text{diag}\left(\frac{1}{|z_1^{(i)}|^2 + \eta}, \ldots, \frac{1}{|z_M^{(i)}|^2 + \eta}\right) \tag{8}$$

with $z_m^{(i)}$ denoting the mth element of $\mathbf{z}$ at iteration i. Using this closed-form solution, the frequencies may then be found using a gradient descent method. The resulting algorithm starts with an initial grid and then iteratively refines the grid points to find the correct solution. This results in a dynamic grid, where the redundant grid points are removed, and the grid points closest to the true solution are refined. The initial grid may here be much coarser than the grid needed to solve (2) with a classic grid-based solution. In the following, we will extend on the SURE-IR algorithm to allow for the incorporation of block penalties, as well as sparsity within each block, showing how the resulting technique may be used to solve the multi-pitch problem.

[1] For notational and computational simplicity, we here consider the discrete-time analytic signal of the (real-valued) measured signal.

III. Proposed method

To take the harmonic structure in (1) into consideration and generalize the above discussed SURE-IR algorithm, we need to reformulate the problem so that it allows for a closed form solution similar to (7). In order to do so, let $\mathbf{A}(\boldsymbol{\theta})$ denote the $N \times M$ dictionary matrix with

$$\mathbf{A}(\boldsymbol{\theta}) = \left[\begin{array}{ccc} \mathbf{A}_1(\boldsymbol{\theta}_1) & \dots & \mathbf{A}_G(\boldsymbol{\theta}_G) \end{array}\right] \tag{9}$$

$$\mathbf{A}_g(\theta_g) = \left[\begin{array}{cccc} \mathbf{a}(\theta_g) & \mathbf{a}(2\theta_g) & \dots & \mathbf{a}(L_g\theta_g) \end{array}\right] \tag{10}$$

$$\mathbf{a}(\ell\theta_g) = \left[\begin{array}{ccc} \mathrm{e}^{i2\pi\ell\theta_g t_1} & \dots & \mathrm{e}^{i2\pi\ell\theta_g t_N} \end{array}\right]^T / \sqrt{(N)} \tag{11}$$

where θ_g denotes the fundamental frequency for the gth pitch-group, for $g = 1, \dots, G$, with G denoting the number of considered groups, and $M = \sum_{g=1}^{G} L_g$, i.e., the total number of frequencies considered in the initial grid. Note that by dividing with $\sqrt{N}$, the columns of the matrix $\mathbf{A}(\boldsymbol{\theta})$ are normalized. Using the logarithmic penalty for a group penalty, and at the same time allowing for sparsity within the groups, one may consider the cost function

$$\underset{\mathbf{z},\boldsymbol{\theta}}{\text{minimize}}\ \lambda \sum_{g=1}^{G}\sum_{\ell=1}^{L_g} \log\left(|z_{g,\ell}|^2 + \eta\right) + \mu \sum_{g=1}^{G} \log\left(||\mathbf{z}_g||_2^2 + \eta\right)/L_g + ||\mathbf{y} - \mathbf{A}(\boldsymbol{\theta})\mathbf{z}||_2^2 \tag{12}$$

where μ and λ are hyperparameters that govern the group sparsity and the overall sparsity, respectively, $\eta > 0$ are constants ensuring that the functions are not evaluated over zero, and where $\mathbf{z}_g$ denotes the amplitudes related to group g in $\mathbf{A}$. As expected, the problem in (12) is not convex and difficult to solve. To allow for a closed form solution for $\mathbf{z}$, the second term in (12) is rewritten as

$$\sum_{g=1}^{G} \log(||\mathbf{z}_g||_2^2 + \eta)/L_g = \sum_{g=1}^{G} \log(||\mathbf{F}_g\mathbf{z}||_2^2 + \eta)/L_g \tag{13}$$

where $\mathbf{F}_g$ is a $\sum_{g=1}^{G} L_g \times \sum_{g=1}^{G} L_g$ diagonal matrix with ones on the diagonal corresponding to group g, and zeros elsewhere. Thus, $\mathbf{F}_g\mathbf{z}$ is not equal to $\mathbf{z}_g$. However, their non-zero elements are equal, and $\mathbf{z}_g$ is a subvector in the resulting vector $\mathbf{F}_g\mathbf{z}$. To solve (12), we then follow the same approach as in [31] and use an MM approach. To do so, a surrogate function, $Q(\mathbf{z}|\mathbf{z}^{(i)})$, which is much simpler than the original function, is devised such that it coincides with the original function at the current point $\mathbf{z}^{(i)}$, and is greater than or equal to the original function everywhere else. It can be shown that minimizing (or even just decreasing) $Q(\mathbf{z}, \mathbf{z}^{(i)})$ then yields a non-increasing updating step in the original function, thus yielding a method of minimizing the more complex function, using simpler functions. An appropriate surrogate function to (12) may be selected as

$$\psi_1(\mathbf{z}|\mathbf{z}^{(i)}) = \sum_{g=1}^{G} L_g^{-1} \Bigg(\frac{||\mathbf{F}_g\mathbf{z}||_2^2 + \eta}{||\mathbf{F}_g\mathbf{z}^{(i)}||_2^2 + \eta} + \log(||\mathbf{F}_g\mathbf{z}^{(i)}||_2^2 + \eta) - 1 \Bigg) \tag{14}$$

for the second term in (12) and

$$\psi_2(\mathbf{z}|\mathbf{z}^{(i)}) = \sum_{g=1}^{G}\sum_{\ell=1}^{L_g} \left(\frac{|z_{g,\ell}|^2 + \eta}{|z_{g,\ell}^{(i)}|^2 + \eta} + \log(|z_{g,\ell}^{(i)}|^2 + \eta) - 1 \right) \tag{15}$$

for the first term, thus yielding

$$Q(\mathbf{z}|\mathbf{z}^{(i)}) = \mu\psi_1(\mathbf{z}|\mathbf{z}^{(i)}) + \lambda\psi_2(\mathbf{z}|\mathbf{z}^{(i)})$$

Removing terms that are independent of $\mathbf{z}$ and $\boldsymbol{\theta}$, the surrogate cost function may be re-written as

$$\underset{\mathbf{z},\boldsymbol{\theta}}{\text{minimize}}\ S(\mathbf{z}, \boldsymbol{\theta}|\mathbf{z}^{(i)}) \tag{16}$$

where

$$S(\mathbf{z}, \boldsymbol{\theta}|\mathbf{z}^{(i)}) = \lambda\mathbf{z}^H\mathbf{D}_0^{(i)}\mathbf{z} + \mu\sum_{g=1}^{G} \mathbf{z}^H\mathbf{F}_g^H D_g^{(i)}\mathbf{F}_g\mathbf{z}/L_g + ||\mathbf{A}(\boldsymbol{\theta})\mathbf{z} - \mathbf{y}||_2^2 \tag{17}$$

with

$$\mathbf{D}_0^{(i)} = \text{diag}\left(\frac{1}{|z_1^{(i)}|^2 + \eta}, \dots, \frac{1}{|z_M^{(i)}|^2 + \eta} \right) \tag{18}$$

$$D_g^{(i)} = \frac{1}{||\mathbf{F}_g\mathbf{z}^{(i)}||_2^2 + \eta}, \quad \text{for } g = 1, \dots, G \tag{19}$$

Furthermore, let

$$\mathbf{H}^{(i)} = \sum_{g=1}^{G} \mathbf{F}_g^H D_g^{(i)} \mathbf{F}_g / L_g \tag{20}$$

Differentiating $S(\mathbf{z}, \boldsymbol{\theta}|\mathbf{z}^{(i)})$ with respect to $\mathbf{z}$, setting it equal to zero, yields

$$\frac{\partial S(\mathbf{z}, \boldsymbol{\theta}|\mathbf{z}^{(i)})}{\partial \mathbf{z}} = 0 \Leftrightarrow \tag{21}$$

$$\mathbf{z}(\boldsymbol{\theta})^* = \left(\lambda\mathbf{D}_0^{(i)} + \mu\mathbf{H}^{(i)} + \mathbf{A}(\boldsymbol{\theta})^H\mathbf{A}(\boldsymbol{\theta})\right)^{-1} \mathbf{A}(\boldsymbol{\theta})^H\mathbf{y} \tag{22}$$

Using (22), one may then find the $\boldsymbol{\theta}$ that minimizes (16) by searching for the best $\boldsymbol{\theta}$ using, e.g., a steepest descent method, by substituting (22) in (16), yielding

$$\underset{\boldsymbol{\theta}}{\text{minimize}}\ S(\mathbf{z}^*, \boldsymbol{\theta}|\mathbf{z}^{(i)}) = - \mathbf{y}^H\mathbf{A}(\boldsymbol{\theta}) \left(\lambda\mathbf{D}_0^{(i)} + \mu\mathbf{H}^{(i)} + \mathbf{A}(\boldsymbol{\theta})^H\mathbf{A}(\boldsymbol{\theta})\right)^{-1} \mathbf{A}(\boldsymbol{\theta})^H\mathbf{y} \tag{23}$$

Following the reasoning in [31], one may show that the original cost function, $\Gamma(\boldsymbol{\theta}, \mathbf{z})$, will be non-increasing when one decreases the surrogate function, thus showing that $\Gamma(\boldsymbol{\theta}^{(i+1)}, \mathbf{z}^{(i+1)}) \leq \Gamma(\boldsymbol{\theta}^{(i)}, \mathbf{z}^{(i)})$. This proof has been presented in [31] for the problem in (6); the corresponding proof for the here considered case follows directly, and is, in the interest of brevity, thus omitted.

Interestingly, the minimization problem in (23) is very similar to the one in [31]; the difference lies in the introduction of $\mu\mathbf{H}^{(i)}$, which weights the different $z_{g,\ell}$ accordingly to the power of the group they belong to. This indicates how easy it is to extend the SURE-IR algorithm and allow for the modeling of other structures in the signal. For instance, one may consider

Algorithm 1 The BSURE-IR estimator

Input: A grid, $\boldsymbol{\theta}$, of size M over the considered fundamental frequencies, $\lambda = \lambda_0$, $\mu = \mu_0$, $\xi = \xi_0$, $\eta = 1$, $i = 1$, $k = 0$, $\mathbf{z}^{(0)} = \mathbf{0}_{MG}$, $\mathbf{z}^{(1)} = \mathbf{1}_{MG}$, and data vector $\mathbf{y}$.
Output: The estimates of $\mathbf{z}^{(i)}$ and $\boldsymbol{\theta}^{(i)}$.
while $||\mathbf{z}^{(i)} - \mathbf{z}^{(i-1)}||_2 > \xi$ **do**
  Form $\mathbf{H}^{(i)}$ from (18), (19), and (20).
  Update $\mathbf{z}(\boldsymbol{\theta})^{(i)}$ from (22).
  Update $\boldsymbol{\theta}^{(i)}$ by taking a single step in (28).
  Decrease λ, η, and μ, prune the dictionary and remove all columns of $\mathbf{A}(\boldsymbol{\theta})$ corresponding to elements in $\mathbf{z}$ with $|z_{g,\ell}| < 0.05$ and $||\mathbf{z}_g||_2 < 0.05$.
  Set $i = i + 1$
  If $||\mathbf{z}||_0 = 0$, then set $k = k + 1$, $\mu = \mu_0/2^k$, $\mathbf{z}^{(0)} = \mathbf{0}_{MG}$, $\mathbf{z}^{(1)} = \mathbf{1}_{MG}$, and restart the iterations with $i = 1$.
end while

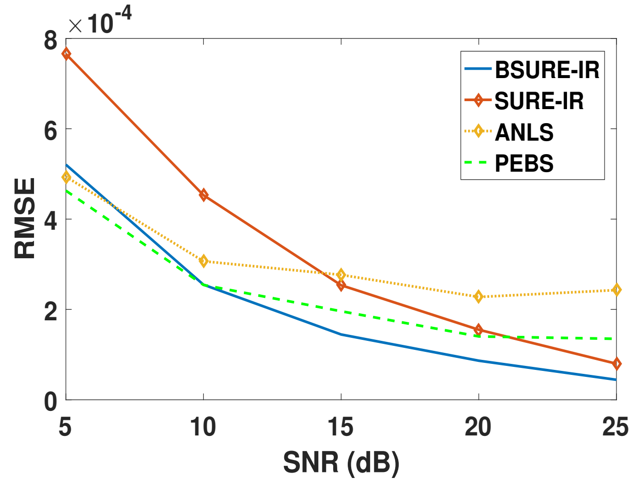

Fig. 1. The RMSE of the frequency estimates, as defined in (30), as a function of SNR, for uniformly sampled data.

adding a logarithmic version of the total variation penalty to (12), which would then simply add another term in (23). This suggests that the SURE-IR approach, in contrast to, e.g., atomic norm, can allow for adding and subtracting different penalties and may thus easily be extended to cover also other model structures.

For the gradient based search, one needs to compute the gradient of $S(\mathbf{z}^*, \boldsymbol{\theta}|\mathbf{z}^{(i)})$ with respect to $\boldsymbol{\theta}$. The gradient for the single sinusoid case was presented in [31] and the reader is referred to that paper for the details. However, we note that, in contrast to the single sinusoidal case, the derivative of one fundamental frequency, $\partial\mathbf{A}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, is in the examined case operating on all the elements of that pitch group; the derivative will thus be a matrix instead of a vector for the here considered case. Thus the direction, d_g, for which the frequency for the pitch group g is moving is

$$d_g = -\mathbf{y}^H \left(\boldsymbol{T}_1 + \mathbf{A}\mathbf{G}\mathbf{A}^H + \boldsymbol{T}_1^H \right) \mathbf{y} \tag{24}$$

where

$$\boldsymbol{T}_1 = \frac{\partial\mathbf{A}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \boldsymbol{T}_2 \mathbf{A}(\boldsymbol{\theta})^H \tag{25}$$

with

$$\boldsymbol{T}_2 = \left(\lambda\mathbf{D}_0 + \mu\mathbf{H} + \mathbf{A}(\boldsymbol{\theta})^H \mathbf{A}(\boldsymbol{\theta}) \right)^{-1} \tag{26}$$

and

$$\mathbf{G} = -\boldsymbol{T}_2 \left(\frac{\partial\mathbf{A}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}^H \mathbf{A}(\boldsymbol{\theta}) + \mathbf{A}(\boldsymbol{\theta})^H \frac{\partial\mathbf{A}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \right) \boldsymbol{T}_2 \tag{27}$$

When forming the gradient step, each harmonic is then multiplied with its corresponding harmonic order, i.e., ℓ. Thus, the updating becomes

$$\theta_g^{(i+1)} = \theta_g^{(i)} - \alpha d_g \tag{28}$$

where α denotes the step length. The harmonics are then updated accordingly by scaling the fundamental frequency with the harmonic order ℓ.

The algorithm starts by first selecting a grid of fundamental frequencies, and then adding the harmonics, thus forming a grid containing G fundamental frequencies and M total grid points (thereby including both the fundamental frequencies and their respective harmonics). In pitch estimation, one has to pay particular attention to the so-called halfling problem [15,16]. This problem stems from the fact that the frequencies corresponding to $\{f_0, 2f_0, \ldots, L_0 f_0\}$ are also present in the group corresponding to $f_0/2$. This ambiguity results in that the algorithms often prefer to choose the lower fundamental frequency. A common solution to this problem is to include a total variation penalty, which can easily be included in the proposed method. However, we opt to overcome this problem by, similarly to [17], instead penalize the amplitudes in each group with the power of the group's fundamental frequency such that the second penalty term in (19) becomes

$$D_g^{(i)} = \frac{1}{|z_{g,1}^{(i-1)}| \left(||\mathbf{F}_g \mathbf{z}^{(i)}||_2^2 + \eta \right)}, \quad \text{for } g = 1, \ldots, G \tag{29}$$

where $z_{g,1}$ denotes the amplitude corresponding to the fundamental frequency of group g. Thereby, if the amplitude of the candidate fundamental frequency is zero, the other amplitudes in that group will be heavily penalized; thus, if there is any competition between f_0 and $f_0/2$ candidates, the method is more likely to choose the higher fundamental frequency. This penalty is not necessary after the algorithm has found some initial estimates of the groups, and may be removed after a couple of iterations. Appropriately setting hyperparameters such as μ and λ is often a difficult problem. In this work, we take a practical stance to this problem. First, we observe that if the true $\boldsymbol{\theta}$ were known, one would solve (23) with $\mu = \lambda = 0$. Thus, we should expect the method to improve if we gradually decreased λ and μ. To this end, we begin setting λ as in [31]. Then, after the first pruning step, we decrease λ by half each iteration, thereby gradually improving the estimates. Similar to the method introduced in [31], the extended algorithm will also decrease η in each iteration. The choice of μ is more critical. A too small value of μ will result in too many groups being

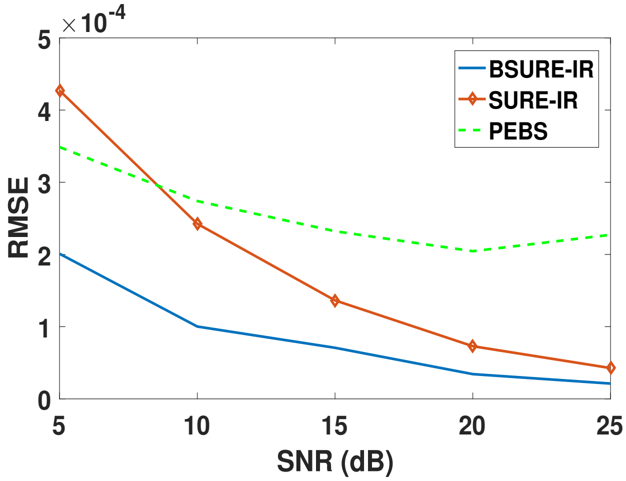

Fig. 2. The RMSE of the frequency estimates as a function of SNR for non-uniformly sampled data.

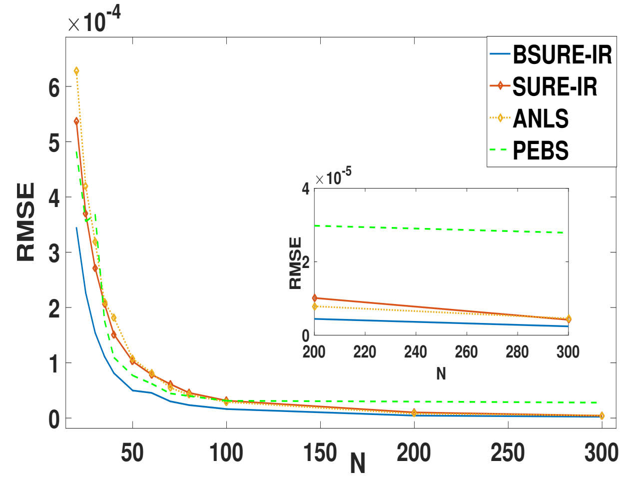

Fig. 3. The RMSE of the frequency estimates as a function of the data length, N.

involved in the solution, and a too large value will suppress true groups and often result in the method breaking down. If one is not able to find a suitable value of μ, one may first run the algorithm by setting a large μ; if the method breaks down, i.e., yields an empty set, the problem is simply resolved using a smaller value of μ, preferably by decreasing the value by a factor 2. As noted above, it may be beneficial to continue to decrease the value of μ through out the iterations. This approach to selecting a good value of μ is possible since with the pruning step, the computational complexity is low, and it can be further decreased by warm-starting the algorithm for each decrease of μ. As shown in the numerical section, the proposed method is notably faster than the SURE-IR algorithm when using a dictionary with the same number of frequencies. This is primarily due to the fact that even though the number of grid points are the same, the proposed method only has the fundamental frequencies as variables; thus, when calculating the gradient, and pruning the dictionary, these steps become more efficient. The value of η is decreased with a factor 10 every time $||\mathbf{z}^{(i)} - \mathbf{z}^{(i-1)}||_2^2 < \eta$. This rule is based on the fact that when the methods starts to converge, η should play a smaller part in (18) and (19). Furthermore, since the dictionary is pruned, it means that as the method converges, a larger ratio of the elements in $\mathbf{z}$ becomes non-zero. Thus, it is reasonable to decrease η to achieve a smaller bias in the $\mathbf{z}$ estimates.

IV. Implementational aspects

The implementation of the proposed algorithm relies on three steps in each iteration: solving (23) using a gradient-based minimization, evaluating $\mathbf{z}$ for the new value of $\boldsymbol{\theta}$ using (22), and removing redundant grid points. The last step is implemented to reduce the computational complexity by decreasing the size of the matrices $\mathbf{A}(\boldsymbol{\theta})$, $\mathbf{D}_0$, and $\mathbf{H}$. To speed up the calculations, one may start pruning the dictionary after merely a few iterations. This is done by removing all the groups and all the individual frequencies in case their magnitudes are below a certain predefined limit, say τ, which we in this paper has selected to be $\tau = 0.05$.

We coin the presented method the block super-resolution iteratively reweighted (BSURE-IR). Algorithm 1 summaries the proposed method, wherein $\mathbf{0}_{MG}$ and $\mathbf{1}_{MG}$ denote an $MG \times 1$ long vector of zeros and ones, respectively, ξ a predefined stopping criteria, and k the number of restarts.

V. Numerical Examples

In this section, we investigate the performance of the proposed method and compare the results to other competing methods. Throughout this section, we will evaluate the methods' ability to correctly estimate the frequencies by measuring the root-mean-squared-error (RMSE), defined as

$$\mathrm{RMSE}(\hat{\boldsymbol{\theta}}) = \sqrt{\frac{1}{\sum_{k=1}^{K} L_k} \sum_{k=1}^{K} \sum_{\ell=1}^{L_k} (\theta_{k,\ell} - \hat{\theta}_{k,\ell})^2} \tag{30}$$

where $\theta_{k,\ell}$ denotes the true parameter value, $\hat{\theta}_{k,\ell}$ the estimated value, and $\hat{\boldsymbol{\theta}}$ the vector of parameters that are estimated. In the following, we compare the methods' RMSE as a function either of the length of the signal, N, or the signal-to-noise-ratio (SNR), defined as

$$\mathrm{SNR} = 10 \log \left(\frac{P}{\sigma^2} \right) \tag{31}$$

where P is the power of the noiseless signal and σ^2 the variance of the noise. For each SNR level or signal length, the presented results are found using 100 Monte-Carlo simulations. In the first example, an $N = 30$ long uniformly sampled signal with a single pitch was considered. The fundamental frequency was uniformly drawn between $[1/7, 1/3)$ for each Monte-Carlo simulation and the number of harmonics were selected as $\left\lfloor \frac{1}{f_0} \right\rfloor$ for each fundamental frequency, f_0, with $\lfloor \cdot \rfloor$ denoting the floor operator. Four algorithms were considered; BSURE-IR, SURE-IR [31], ANLS [9], and the PEBS algorithm [15]. The BSURE-IR method was allowed an initial grid of 15 elements over the fundamental frequencies, ranging from $[0.1, 0.3]$, and the number of harmonics selected as $\left\lfloor \frac{1}{f_0} \right\rfloor$, for each considered fundamental frequency, f_0, thus yielding

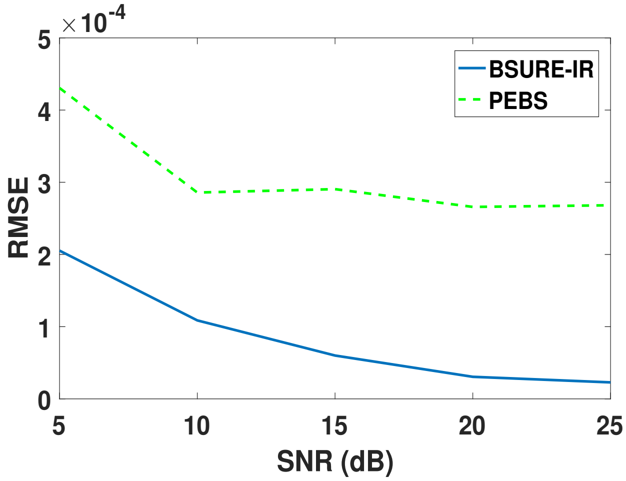

Fig. 4. The RMSE of the frequency estimates of a multi-pitch signal containing two pitches for non-uniformly sampled data.

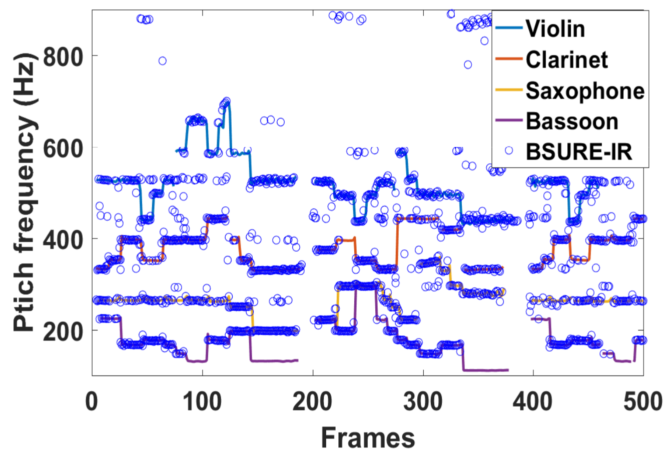

Fig. 5. The resulting estimation of the fundamental frequencies (pitches) of the Bach10 data set.

a dictionary containing a total of 77 spectral lines. The initial value of μ was set to 100. The SURE-IR algorithm was also allowed a dictionary containing 77 elements, although these being unstructured. The ANLS was allowed 2^8 grid points and was given the same range over the fundamental frequency as BSURE-IR, as well as perfect model order knowledge. The PEBS algorithm was given prior information about where the fundamental frequency was positioned, given as a range of ± 0.02 around the true value. In this range, PEBS was given 1000 grid points and the initial user parameters were set to 5 and 30 for the parameter governing the ℓ_1 and the ℓ_2 norms, respectively. Furthermore, for the PEBS algorithm, only the largest peak was selected from the estimates, thus not requiring the algorithm to make a correct model order, thereby avoiding the problem of wrongly setting the hyperparameters. This was not true for the other methods, where each wrong model order estimate was recorded. The resulting RMSE may be seen in Figure 1, where it can be seen that the proposed method outperforms the other methods for SNR-levels of 10 dB and above. Interestingly, it can be seen that the grid-based methods have similar performance to the BSURE-IR for low SNR levels, whereas the two off-grid methods excel for higher SNR levels; even SURE-IR, which does not take the harmonic structure in consideration, actually outperforms the two grid-based methods that actively exploits the harmonic structure. In this setting, the BSURE-IR method failed to correctly estimate the model order 6 times for the lowest SNR level, but managed to correctly do so for the other SNRs. The average run-times for the methods were 3.0 seconds for BSURE-IR, 10.5 seconds for SURE-IR, 0.1 seconds for ANLS, and 4.7 for PEBS.

Proceeding, we investigate how the performance is affected by non-uniformly sampled data. This scenario is not as common for audio samples, but is so in many other areas. As ANLS does not allow for this case, the algorithm is omitted from comparison. Using the same settings as before, but now with non-uniform sampled data with length $N = 30$ sampled from 60 measurements, the RMSE was measured for the methods. Figure 2 shows the result. As expected, BSURE-IR again outperforms the competing methods. Again, comparing SURE-IR with PEBS, the latter seems to benefit from exploiting the harmonic structure for lower SNR levels. However, when the SNR level reaches 10 dB, the unstructured SURE-IR again outperforms the PEBS algorithm. Here, BSURE-IR failed to determine the correct model order 6 times for SNR 5 dB, but estimated it correctly in the other cases. The run times in this setting were 2.5 seconds for BSURE-IR, 11.5 seconds for SURE-IR, and 18.5 seconds for PEBS.

In the third example, we investigate the performance as function of the length of the signal. Figure 3 shows the results when using the same settings as before, but with N ranging from 20 to 300 and with SNR fixed at 15 dB. Once again it may be seen that the purposed method outperforms the competing methods. In this scenario, we had to remove 86 outliers for PEBS to make the figure readable; 55 outliers for $N = 20$, 29 for $N = 25$, and 2 for $N = 30$. BSURE-IR estimated the wrong model order five times, once for $N = 20$, $N = 100$, and $N = 300$, and twice for $N = 200$. The run times for the considered algorithms were 2.3 seconds for BSURE-IR, 8.6 seconds for SURE-IR, and 14.2 seconds for PEBS.

In the fourth example, we look at the case were the signal contains multiple pitches. Here, we consider a signal with length $N = 30$, non-uniformly sampled and with two fundamental frequencies set at $0.15\pi/3$ and $0.26\pi/3$. Figure 4 shows the resulting RMSE for all frequencies in both pitches. For the case when the SNR level is 5 dB, BSURE-IR seems to have problem to get the model order correct, and 41 times the estimated order model was incorrect. This only happened 8 times for the other SNR levels. For PEBS, 42 outliers were removed to make the figure more readable. If disregarding the 5 dB case, one can see that the BSURE-IR method outperforms the PEBS algorithm for the multi-pitch case. Note that, again, PEBS is given K *a priori* and is also zoomed in around the correct fundamental frequencies. Also, PEBS are now allowed 1000 grid points for each fundamental frequency. The run times for this examples are 5.2 seconds for BSURE-

Method	Accuracy	Precision	Recall	Pre-trained
BSURE-IR	0.47	0.71	0.58	No
PEARLS	0.44	0.68	0.54	No
PEBS	0.39	0.56	0.51	No
PEBSI-Lite	0.45	0.63	0.61	No
BW15	0.52	0.68	0.68	Yes
ESACF	0.27	0.47	0.39	No

TABLE I
PERFORMANCE MEASURES FOR THE BSURE-IR, PEARLS, PEBS, PEBSI-LITE, BW15, AND ESACF ALGORITHMS, WHEN EVALUATED ON THE BACH10 DATASET.

IR and 84.3 seconds for PEBS. The increase in run time for PEBS is mainly due to the increase in grid size.

In the final example, we evaluate the performance of the methods on the Bach10 dataset [39]. The data set contains ten excerpts from chorals that were composed by Johann Sebastian Bach. The instruments playing in the pieces are a violin, a clarinet, a saxophone, and a bassoon, and the set contains many sequences where the overtones overlaps. The resulting estimates are compared to ground truth fundamental frequencies, obtained by applying the single pitch estimator YIN [40] to each separate channel. Obvious errors in the ground truth were corrected for manually. Each excerpt is about 25-42 seconds long. Table I presents the performance measures accuracy, precision, and recall, as defined in [41]. In Table I, the performance of the BSURE-IR estimator is compared to four other multi-pitch estimators, namely PEARLS [17], PEBS [15], PEBSI-Lite [16], and ESACF [6], as well as a state-of-the-art music transcription method [42], here denoted BW15 (after the surnames of the authors and the year of publication). For BSURE-IR, the starting value of μ was set to 1 and the number of initial fundamental frequency grid-points 30, and the maximum allowed L was set to 4. PEARLS is a time-recursive multi-pitch estimator, with a dictionary learning scheme that resembles a gridless method, but uses a different cost function, and ESACF is a auto-correlation based multi-pitch estimator. The BW15 method is a music transcription algorithm that uses a probabilistic latent component analysis to produce pitch estimates that are trained on databases of music instruments. We choose to include this method into the comparison to show the performance of a state-of-the-art method that is pre-trained and specifically tailored for music transcription, which is not the case for the other discussed methods. The settings and results from ESACF and PEBSI-Lite were taken from [16] and for PEARLS and BW15, the setting and results were from [17]. The PEBS settings and results were obtained from [18]. Figure 5 shows the resulting BSURE-IR estimates of the fundamental frequencies from an excerpt of J. S. Bach's *Ach, Gott und Herr* performed by a violin, a bassoon, a clarinet, and a saxophone. As can be seen from the figure, BSURE-IR manages to capture most of the fundamental frequencies without too many false positives. Furthermore, from Table I, one may see that the BSURE-IR method scores higher on both accuracy and precision as compared to the other multi-pitch estimators, and has somewhat even score for recall. Not surprisingly, BW15 attains a higher score than BSURE-IR, except for precision, where BSURE-IR attains a slightly higher score. However, it should be stressed that BW15 has been trained on the instruments included in the Bach10 data set, whereas BSURE-IR has not. We note that, as for a future research topic, it would be interesting to try to combine the probabilistic approach of BW15 and the more robust BSURE-IR signal model approach.

VI. CONCLUSIONS

In this paper, we present a novel off-grid multi-pitch estimator. By parameterizing the dictionary containing the candidate pitches and solving a non-convex optimization problem using a majorization-minimization approach, an iterative method is derived. In each iteration, the dictionary is pruned which allows for a decreased computational complexity. The method is evaluated on both simulated and real data. In the real data case, the proposed method is shown to yield similar performance as a specialized music transcription algorithm that is pre-trained on the instruments present in the signal. Furthermore, the proposed method is benchmarked against other popular multi-pitch estimates, showing the preferred performance of the proposed method.

REFERENCES

[1] H. Akaike, "A New Look at Statistical Model Identification," *IEEE Trans. Autom. Control*, vol. 19, pp. 716–723, 1974.

[2] J. Rissanen, "Modeling by Shortest Data Description," *Automatica*, vol. 14, pp. 465–471, 1978.

[3] G. Schwarz, "Estimating the Dimension of a Model," *Ann. Stat.*, vol. 6, pp. 461–464, 1978.

[4] P. Stoica and Y. Selén, "Model-order Selection — A Review of Information Criterion Rules," *IEEE Signal Process. Mag.*, vol. 21, no. 4, pp. 36–47, July 2004.

[5] 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.

[6] T. Tolonen and M. Karjalainen, "A computationally efficient multipitch analysis model," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 6, pp. 708–716, 2000.

[7] 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.

[8] M. G. Christensen, J. H. Jensen, A. Jakobsson, and S. H. Jensen, "On optimal filter designs for fundamental frequency estimation," *IEEE Signal Process. Lett.*, vol. 15, pp. 745–748, 2008.

[9] M. G. Christensen, P. Stoica, A. Jakobsson, and S. H. Jensen, "Multi-pitch estimation," *Signal Processing*, vol. 88, no. 4, pp. 972–983, April 2008.

[10] M. Christensen and A. Jakobsson, *Multi-Pitch Estimation*, Morgan & Claypool, San Rafael, Calif., 2009.

[11] K. W. Chan and H. C. So, "Accurate frequency estimation for real harmonic sinusoids," *IEEE Signal Process. Lett.*, vol. 11, no. 7, pp. 609–612, 2004.

[12] M. G. Christensen, S. H. Jensen, S. V. Andersen, and A. Jakobsson, "Subspace-based Fundamental Frequency Estimation," in *European Signal Processing Conference*, Vienna, September 7-10 2004.

[13] J. K. Nielsen, T. L. Jensen, J. R. Jensen, M. G. Christensen, and S. H. Jensen, "Fast fundamental frequency estimation: Making a ststatistical efficient estimator computationally efficient," *Elsevier Signal Processing*, vol. 135, pp. 188–197, Jan 2017.

[14] J. Zeremdini, M. A. B. Messaoud, and A. Bouzid, "Multiple comb filters and autocorrelation of the multi-scale product for multi-pitch estimation," *Elsevier Signal Processing*, vol. 120, pp. 45–53, May 2017.

[15] S. I. Adalbjörnsson, A. Jakobsson, and M. G. Christensen, "Multi-Pitch Estimation Exploiting Block Sparsity," *Elsevier Signal Processing*, vol. 109, pp. 236–247, April 2015.

[16] F. Elvander, T. Kronvall, S. I. Adalbjörnsson, and A. Jakobsson, "An Adaptive Penalty Multi-Pitch Estimator with Self-Regularization," *Elsevier Signal Processing*, vol. 127, pp. 56–70, October 2016.

[17] F. Elvander, J. Swärd, and A. Jakobsson, "Online Estimation of Multiple Harmonic Signals," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 2, pp. 273–284, February 2017.

[18] S. Lei, F. Elvander, J. Swärd, S. I. Adalbjörnsson, and A. Jakobsson, "Computationally Efficient Multi-Pitch Estimation Using Sparsity," in *11th IMA internation Conference on Mathematics in Signal Processing*, Birmingham, England, Dec 2016.

[19] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations*, Chapman and Hall/CRC, 2015.

[20] E. J. Candès, M. B. Wakin, and S. Boyd, "Enhancing Sparsity by Reweighted l_1 Minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, Dec. 2008.

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

[22] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 707–710, Oct. 2007.

[23] Jean-Jacques Fuchs, "On sparse representation in arbitrary redundant bases," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1341–1344, June 2004.

[24] J. A. Tropp, *Topics in Sparse Approximation*, Computational and applied mathematics, 2004.

[25] Y. Chi, L. L. Scharf, A. Pezeshki, and A. R. Calderbank, "Sensitivity to Basis Mismatch in Compressed Sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182 –2195, May 2011.

[26] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The Convex Geometry of Linear Inverse Problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, Dec 2012.

[27] B. N. Bhaskar, G. Tang, and B. Recht, "Atomic Norm Denoising with Applications to Line Spectral Estimation," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5987 – 5999, July 2013.

[28] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed Sensing Off the Grid," *IEEE Trans. Inform. Theory*, vol. 59, no. 11, pp. 7465–4790, Nov 2013.

[29] Z. Yang and L. Xie, "On Gridless Sparse Methods for Line Spectral Estimation From Complete and Incomplete Data," *IEEE Trans. Signal Process.*, vol. 63, no. 12, pp. 3139–3153, June 2015.

[30] Z. Yang and L. Xie, "Enhancing Sparsity and Resolution via Reweighted Atomic Norm Minimization," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 995–1006, Feb 2016.

[31] J. Fang, F. Wang, Y. Shen, H. Li, and R. S. Blum, "Super-Resolution Compressed Sensing for Line Spectral Estimation: An Iterative Reweighted Approach," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4649–4662, September 2016.

[32] C. S. Sunnersjö, "Rolling Bearing Vibrations - The Effects of Geometrical Imperfections and Wear," *Journal of Sound and Vibration*, vol. 98, no. 4, pp. 455–474, Feb 1985.

[33] A. M. Ahmadi, D. Petersen, and C. Howard, "A nonlinear dynamic vibration model of defective bearings - the importance of modelling the finite size of rolling elements," *Mechanical Systems and Signal Processing*, vol. 52-53, pp. 309–326, Feb 2015.

[34] W. Liu, Y. Zhang, Z. J. Feng, J. S. Zhao, and D. Wang, "A study on waviness induced vibration of ball bearings based on signal coherence theory," *Journal of Sound and Vibration*, vol. 333, no. 23, pp. 6107–6120, Nov 2014.

[35] A. Cubillo, S. Perinpanayagam, and M. Esperon-Miguez, "A review of physics-based models in prognostics: Application to gears and bearings of rotating machinery," *Advances in Mechanical Engineering*, vol. 8, no. 8, pp. 1–21, Aug 2016.

[36] P. C. Hansen and S. H. Jensen, "Subspace-Based Noise Reduction for Speech Signals via Diagonal and Triangular Matrix Decompositions: Survey and Analysis," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 1–24, 2007.

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

[38] S. I. Adalbjörnsson, J. Swärd, and A. Jakobsson, "Enhancing smoothness in amplitude modulated sparse signals," in *11th IMA International Conference on Mathematics in Signal Processing*, Birmingham, England, Dec 2016.

[39] Z. Duan and B. Pardo, "Bach10 dataset," http://music.cs.northwestern.edu/data/Bach10.html, Accessed December 2015.

[40] A. de Cheveigné and H. Kawahara, "YIN, a fundamental frequency estimator for speech and music," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1917–1930, 2002.

[41] M. Bay, A. F. Ehmann, and J. S. Downie, "Evaluation of Multiple-F0 Estimation and Tracking Systems," in *International Society for Music Information Retrieval Conference*, Kobe, Japan, October 2009.

[42] E. Benetos and T. Weyde, "An Efficient Temporally-Constrained Probabilistic Model for Multiple-Instrument Music Transcription," in *Proceedings of the 16th International Society for Music Information Retrieval Conference*, Malaga, Spain, October 2015.

Johan Swärd (S'12) received his M.Sc. in Industrial Engineering and Management, and his Ph.D. in Mathematical Statistics from Lund University, Sweden, in 2012 and 2017, respectively. He has been a visiting researcher at the Department of Systems Innovations at Osaka University, Japan, and Stevens Institute of Technology, New Jersey, USA. His research interests include machine learning and applications of sparse and convex modeling in statistical signal processing and spectral analysis.

Hongbin Li (M'99-SM'08) received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, in 1991 and 1994, respectively, and the Ph.D. degree from the University of Florida, Gainesville, FL, in 1999, all in electrical engineering.

From July 1996 to May 1999, he was a Research Assistant in the Department of Electrical and Computer Engineering at the University of Florida. Since July 1999, he has been with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, where he became a Professor in 2010. He was a Summer Visiting Faculty Member at the Air Force Research Laboratory in the summers of 2003, 2004 and 2009. His general research interests include statistical signal processing, wireless communications, and radars.

Dr. Li received the IEEE Jack Neubauer Memorial Award in 2013 from the IEEE Vehicular Technology Soceity, the Outstanding Paper Award from the IEEE AFICON Conference in 2011, the Harvey N. Davis Teaching Award in 2003 and the Jess H. Davis Memorial Award for excellence in research in 2001 from Stevens Institute of Technology, and the Sigma Xi Graduate Research Award from the University of Florida in 1999. He has been a member of the IEEE SPS Signal Processing Theory and Methods Technical Committee (TC) and the IEEE SPS Sensor Array and Multichannel TC, an Associate Editor for *Signal Processing* (Elsevier), *IEEE Transactions on Signal Processing*, *IEEE Signal Processing Letters*, and *IEEE Transactions on Wireless Communications*, as well as a Guest Editor for *IEEE Journal of Selected Topics in Signal Processing* and *EURASIP Journal on Applied Signal Processing*. He has been involved in various conference organization activities, including serving as a General Co-Chair for the 7th IEEE Sensor Array and Multichannel Signal Processing (SAM) Workshop, Hoboken, NJ, June 17-20, 2012. Dr. Li is a member of Tau Beta Pi and Phi Kappa Phi.

Andreas Jakobsson (S'95-M'00-SM'06) received his M.Sc. from Lund Institute of Technology and his Ph.D. in Signal Processing from Uppsala University in 1993 and 2000, respectively. Since, he has held positions with Global IP Sound AB, the Swedish Royal Institute of Technology, King's College London, and Karlstad University, as well as held an Honorary Research Fellowship at Cardiff University. He has been a visiting researcher at King's College London, Brigham Young University, Stanford University, Katholieke Universiteit Leuven, and University of California, San Diego, as well as acted as an expert for the IAEA. He is currently Professor and Head of Mathematical Statistics at Lund University, Sweden. He has published his research findings in about 200 refereed journal and conference papers, and has filed five patents. He has also authored a book on time series analysis (Studentlitteratur, 2013 and 2015), and co-authored (together with M. G. Christensen) a book on multi-pitch estimation (Morgan & Claypool, 2009). He is a member of The Royal Swedish Physiographic Society, a member of the EURASIP Special Area Team on Signal Processing for Multisensor Systems (2015-), a Senior Member of IEEE, and an Associate Editor for Elsevier Signal Processing. He has previously also been a member of the IEEE Sensor Array and Multichannel (SAM) Signal Processing Technical Committee (2008-2013), an Associate Editor for the IEEE Transactions on Signal Processing (2006-2010), the IEEE Signal Processing Letters (2007-2011), the Research Letters in Signal Processing (2007-2009), and the Journal of Electrical and Computer Engineering (2009-2014). His research interests include statistical and array signal processing, detection and estimation theory, and related application in remote sensing, telecommunication and biomedicine.