An Adaptive Penalty Approach to Multi-Pitch Estimation

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

Published in:
Signal Processing Conference (EUSIPCO), 2015 23rd European

DOI:
10.1109/EUSIPCO.2015.7362339

2015

AN ADAPTIVE PENALTY APPROACH TO MULTI-PITCH ESTIMATION

Ted Kronvall, Filip Elvander, Stefan Ingi Adalbjörnsson, and Andreas Jakobsson

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

ABSTRACT

This work treats multi-pitch estimation, and in particular the common misclassification issue wherein the pitch at half of the true fundamental frequency, here referred to as a sub-octave, is chosen instead of the true pitch. Extending on current methods which use an extension of the Group LASSO for pitch estimation, this work introduces an adaptive total variation penalty, which both enforce group- and block sparsity, and deal with errors due to sub-octaves. The method is shown to outperform current state-of-the-art sparse methods, where the model orders are unknown, while also requiring fewer tuning parameters than these. The method is also shown to outperform several conventional pitch estimation methods, even when these are virtued with oracle model orders.

Index Terms— multi-pitch estimation, block sparsity, adaptive sparse penalty, total variation, ADMM

1. INTRODUCTION

Pitch estimation, i.e., estimating the fundamental frequency of a group of harmonically related sinusoids, is a problem arising in a variety of fields, not least in audio processing. For example, correctly determining the pitches present in a signal is a fundamental building block in many music information retrieval applications, such as automatic music transcription and genre classification [1]. However, pitch estimation for multi-pitch signals is a difficult problem, and although notable efforts have been made to find reliable multi-pitch estimators, (see e.g. [2]), most of the currently available methods which use the harmonic structure depend on *a priori* model order information, i.e., knowing the number of pitches present, as well as the number of harmonic overtones for each pitch. Such information is in general notoriously difficult to obtain. Our approach is instead to solve the problem in a group sparse modeling framework, which allows us to avoid making explicit assumptions on the number of pitches, nor the number of harmonics. Instead, the number of components in the signal is chosen implicitly, by the setting of some tuning parameters. These tuning parameters determine how appropriate a given pitch candidate is to be present in the signal and may be set using some simple heuristics, or by using cross-validation.The sparse modeling approach has earlier been used for audio (see, e.g., [3]), and specifically for sinusoidal components in [4]. We extend on these works by exploiting the harmonic structure of the signals in a block sparse framework, where each block represents a candidate pitch. A similar method was introduced in [5], where block sparsity was enforced using block-norms, penalizing the number of active pitches. As the block-norm penalty, under some circumstances, cannot distinguish a true pitch from its sub-octave, i.e., the pitch with half of the true fundamental frequency, the method is also complemented by a total variation penalty, which is shown to solve such issues. Total variation penalties are often applied in image analysis to obtain block-wise smooth image reconstructions (see, e.g., [6]). For audio data, one can similarly assume that signals often are block-wise smooth, as the harmonics of a pitch are expected to be of comparable magnitude [7]. Enforcing this feature will specifically deal with octave errors (due to present sub-octaves), as, in the noise free case, only every other harmonic of the sub-octave will have non-zero power. In this paper, we show that a total variation penalty, in itself, is enough to enforce a block sparse solution, if utilized efficiently. More specifically, by making the penalty function adaptive, we may improve upon the convex approximation used in [5], allowing us to drop the block-norm penalty altogether, and so reduce the number of tuning parameters. In some estimation scenarios, e.g., when estimating chroma using the approach in [8], this would simplify the tuning procedure significantly. Furthermore, we show that the proposed method performs comparably to that of [5], albeit with the notable improvement of requiring fewer tuning parameters. The method operates by solving a series of convex optimization problems, and so solve these we present an efficient algorithm based on the alternating directions method of multipliers (ADMM) [9].

2. SIGNAL MODEL

Consider a complex-valued[1] signal consisting of K pitches, where the kth pitch is constituted by a set of L_k harmonically

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

[1]For notational simplicity and computational efficiency, we here use the discrete-time analytical signal formed from the measured (real-valued) signal.

related sinusoids, defined by the component having the lowest frequency ω_k, such that

$$x(t) = \sum_{k=1}^{K} \sum_{\ell=1}^{L_k} a_{k,\ell} e^{i\omega_k \ell t} \tag{1}$$

for $t = 1, \ldots, N$, where $\omega_k \ell$ is the frequency of the ℓth harmonic in the kth pitch, and with $a_{k,l}$ denoting its magnitude and phase. The occurrence of such harmonic signals is often in combination with non-sinusoidal components, such as for instance, colored broadband noise or non-stationary impulses. In the scope of this work, we only treat the narrowband components of the signal, although noting that audio signals often also contain other features of notable perceptual importance such as the signal's timbre. In general, selecting model orders in (1) is a daunting task, with both the number of sources, K, and the number of harmonics in each of these sources, L_k, being unknown, as well as often being structured such that different sources may have spectrally overlapping overtones. In order to remedy this, we propose a relaxation of the model onto a predefined grid of $P \gg K$ candidate fundamentals, each having $L_{\max} \geq \max_k L_k$, harmonics. Here, we chose the candidates so numerous and so finely spaced that the approximation

$$x(t) \approx \sum_{p=1}^{P} \sum_{\ell=1}^{L_{\max}} a_{p,\ell} e^{i\omega_p \ell t} \tag{2}$$

holds sufficiently well. We are only interested in such approximations where few, ideally K, of the fundamentals will have non-zero power, and so steps must be taken to ensure this sparse behavior of the to be estimated amplitudes $a_{p,l}$. This approach may be seen as a sparse linear regression problem reminiscent of [4] and has been thoroughly examined in the context of pitch estimation in, e.g., [5, 10, 11]. For notational convenience, we define the set of all amplitude parameters to be estimated as

$$\mathbf{\Psi} = \left\{ \mathbf{\Psi}_{\omega_1}, \ldots, \mathbf{\Psi}_{\omega_P} \right\} \tag{3}$$

$$\mathbf{\Psi}_{\omega_k} = \left\{ a_{k,1}, \ldots, a_{k,L_{\max}} \right\} \tag{4}$$

where, as described above, most $a_{k,\ell}$ in $\mathbf{\Psi}$ will be zero. It should be noted that the sparse pattern of $\mathbf{\Psi}$ will be group-wise, so that if a pitch with fundamental frequency ω_p is not present, then neither will any of its harmonics, i.e., $\mathbf{\Psi}_{\omega_p} = \mathbf{0}$. Furthermore, when a pitch is present, we may expect that not all $L_{\max}$ harmonics will be non-zero, but only the actual L_k ones. For candidate pitches at fractions of the present pitch, there will be a partial fit of its harmonics, which may render misclassification, which is a cause for errors, which occurs when a present pitch at ω_k may be perfectly modeled by a pitch at $\omega_k/2$ if $L_{\max} \geq 2L_k$, where then every other harmonic, i.e., $\ell = 2, 4, 6, \ldots, 2L_k$, are non-zero and the others equal to zero. To take these attributes into account and to avoid misclassifications, we propose the iterative approach detailed in the next section.

3. MULTI-PITCH ESTIMATION USING AN ADAPTIVE TOTAL VARIATION PENALTY

Considering a measured time-frame of the sought signal, we expect it to be corrupted by noise and perhaps other non-sinusoidal structure, i.e., $y(t) = x(t) + e(t)$, where $e(t)$ is such an additive broadband noise. In order to estimate the parameter set $\mathbf{\Psi}$, one often strives to minimize the squared residual cost function

$$g_1(\mathbf{\Psi}) = \frac{1}{2} \sum_{t=1}^{N} \left| y(t) - \sum_{p=1}^{P} \sum_{\ell=1}^{L_{\max}} a_{p,\ell} e^{i\omega_p \ell t} \right|^2 \tag{5}$$

where $|\cdot|$ denotes the absolute value. However, this function will not enforce said sparsity. As requiring exactly sparse solutions leads to combinatorially infeasible optimization problems, we herein adopt a convex modeling approach using a number of convex cost functions. To discourage spurious harmonics, we introduce a constraint on the ℓ_1-norm of $\mathbf{\Psi}$ by

$$g_2(\mathbf{\Psi}) = \sum_{p=1}^{P} \sum_{\ell=1}^{L_{\max}} |a_{p,\ell}| \tag{6}$$

which is a convex approximation of the ℓ_0 penalty. Parameter estimation using a weighted sum of g_1 and g_2 is widely used in the literature, being referred to as the *lasso* [12]. Taking the block-wise sparse behavior described above into account, we further introduce

$$g_3(\mathbf{\Psi}) = \sum_{p=1}^{P} \sqrt{\sum_{\ell=1}^{L_{\max}} a_{p,\ell}^2} \tag{7}$$

which also is a convex function. The inner sum corresponds to the ℓ_2-norm, and does not enforce sparsity within each pitch, whereas instead the outer sum, corresponding to the ℓ_1-norm, enforces sparsity between pitches. Thereby, adding the $g_3(\mathbf{\Psi})$ constraint will penalize the number of non-zero pitches. However, if we for some p have $2L_p \leq L_{\max}$, the above penalties have no way of discriminating between the correct pitch candidate ω_p and the spurious sub-octave candidate $\omega_p/2$. However, as the sub-octave will only contribute to the harmonic signal at every other frequency in its block, one may reduce the risk of such a misclassification by further adding the penalty

$$\breve{g}_4(\mathbf{\Psi}) = \sum_{q=1}^{PL_{\max}-1} \Big| |a_{q+1}| - |a_q| \Big| \tag{8}$$

where the reparametrization is $q = (p-1)L_{\max} + \ell$, which would add a cost to blocks where there are notable magnitude variations between neighboring harmonics. Regrettably, (8)

is not convex, but a simple convex approximation would be $\tilde{g}_4$, detailed as

$$\tilde{g}_4(\mathbf{\Psi}) = \sum_{q=1}^{PL_{\max}-1} \left| a_{q+1} - a_q \right| \tag{9}$$

which would be a good approximation of (8) if all the harmonics had the same phase. Clearly, this may not be the case, resulting in that the penalty in (9) would also penalize the correct candidate. An illustration of this is found by considering the worst-case scenario, when all the adjacent harmonics are completely out of phase and have the same magnitudes, i.e., $a_{p,\ell+1} = a_{p,\ell} e^{i\pi}$ with magnitude $|a_{p,\ell}| = r$, for $\ell = 1, \ldots, L_p - 1$. Then, the penalty in (9) will yield a cost of $\tilde{g}_4(\mathbf{\Psi}_{\omega_p}) = 2rL_p$ rather than the desired $\breve{g}_4(\mathbf{\Psi}_{\omega_p}) = 2r$. The cost may also be compared with that of (6), which is $g_2(\mathbf{\Psi}_{\omega_p}) = rL_p$, suggesting that this would add a relatively large penalty. More interestingly, for the sub-octave candidate, the cost will be just as large, i.e. if $\omega_{p'} = \omega_p/2$, then $\tilde{g}_4(\mathbf{\Psi}_{\omega_{p'}}) = 2rL_p$ provided that $L_{\max} \geq 2L_p$, thereby offering no possibility of discriminating between the true pitch and its sub-octave. Obviously, such a worst case scenario is just as unlikely as having all harmonics same-phased, if assuming that the phases are evenly distributed on $[0, 2\pi)$. Instead, the $\tilde{g}_4$ penalty of the true pitch will be slightly smaller than its sub-octave, on average, and together with (7), the scale tips in favour of the true pitch, as shown in [5]. We may thus conclude that the combination of g_3 and $\tilde{g}_4$ provides a block sparse solution where sub-octaves are usually discouraged. However, it should be noted that such a solution requires the tuning of two functions to control the block sparsity. In this work, we propose to simplify the described algorithm by improving the approximation in (9), by using an adaptive penalty approach. In order to do so, let $\varphi_{k,\ell}$ denote the phase of the component with frequency $\omega_{k,\ell}$ and collect these phases in the parameter set

$$\mathbf{\Phi} = \left\{ \mathbf{\Phi}_{\omega_1}, \ldots, \mathbf{\Phi}_{\omega_P} \right\} \tag{10}$$

$$\mathbf{\Phi}_{\omega_k} = \left\{ \varphi_{k,1}, \ldots, \varphi_{k,L_{\max}} \right\} \tag{11}$$

The penalty function in (9) may then be modified to

$$g_4(\mathbf{\Psi}, \mathbf{\Phi}) = \sum_{q=1}^{PL_{\max}} \left| a_{q+1} e^{-\varphi_{q+1}} - a_q e^{-\varphi_q} \right| \tag{12}$$

thus penalizing only differences in magnitude. In order to do so, the phases $\varphi_{k,\ell}$ need to be estimated as the arguments of the latest available amplitude estimates $a_{k,\ell}$. As a result, (12) yields an improved approximation of (8), avoiding the issues of (9) described above, and also promotes a block sparse solution. And so, the block-norm penalty function g_3 may be omitted, which simplifies the algorithm noticeably. Thus, we form the parameter estimates by solving

$$\hat{\mathbf{\Psi}} = \arg\min_{\mathbf{\Psi}} \sum_{j=1,2} \lambda_j g_j(\mathbf{\Psi}) + \lambda_4 g_4(\mathbf{\Psi}, \mathbf{\Phi}) \tag{13}$$

where $\lambda_1 = 1$, and where λ_i, for $i = 2, 4$, are user-defined regularization parameters that weigh the importance of each penalty function and the residual cost. To form the convex criteria and to facilitate the implementation, consider the signal expressed in matrix notation as

$$\mathbf{y} = \begin{bmatrix} y(1) & \ldots & y(N) \end{bmatrix}^T \tag{14}$$

$$= \sum_{p=0}^{P} \mathbf{W}_p \, \mathbf{a}_p + \mathbf{e} \triangleq \mathbf{W}\mathbf{a} + \mathbf{e} \tag{15}$$

where

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & \ldots & \mathbf{W}_P \end{bmatrix} \tag{16}$$

$$\mathbf{W}_p = \begin{bmatrix} \mathbf{z}^1 & \ldots & \mathbf{z}^{L_{\max}} \end{bmatrix} \tag{17}$$

$$\mathbf{z}_p = \begin{bmatrix} e^{i\omega_p 1} & \ldots & e^{i\omega_p N} \end{bmatrix}^T \tag{18}$$

$$\mathbf{a} = \begin{bmatrix} \mathbf{a}_1^T & \ldots & \mathbf{a}_P^T \end{bmatrix}^T \tag{19}$$

$$\mathbf{a}_p = \begin{bmatrix} a_{p,1} & \ldots & a_{p,L_{\max}} \end{bmatrix}^T \tag{20}$$

The dictionary matrix $\mathbf{W}$ is constructed of P horizontally stacked blocks, or dictionary atoms $\mathbf{W}_p$, where each is a matrix with $L_{\max}$ columns and N rows. In order to obtain an acceptable approximation of (8), the problem must be solved iteratively, where the last solution is used to improve the next. To pursue an even sparser solution, a re-weighting procedure is simultaneously used for g_2, similar to that in [13]. The solution is thus found at the k-th iteration by solving

$$\hat{\mathbf{a}}^{(k)} = \arg\min_{\mathbf{a}} \sum_{j=1,2,4} g_j(\mathbf{H}_j^{(k)}\mathbf{a}, \lambda_j) \tag{21}$$

where $\mathbf{H}_1^{(k)} = \mathbf{W}$, $\mathbf{H}_2^{(k)} = \mathrm{diag}\big(1/(\big\|\hat{\mathbf{a}}^{(k-1)}\big\|_1 + \epsilon)\big)$, $\mathbf{H}_4^{(k)} = \mathbf{F}\,\mathrm{diag}(\arg\big(\hat{\mathbf{a}}^{(k-1)}\big))^{-1}$, and with

$$g_1(\mathbf{H}_1^{(k)}\mathbf{a}, 1) = \frac{1}{2} ||\mathbf{y} - \mathbf{W}\mathbf{a}||_2^2 \tag{22}$$

$$g_2(\mathbf{H}_2^{(k)}\mathbf{a}, \lambda_2) = \lambda_2 \left|\left| \mathbf{H}_2^{(k)}\mathbf{a} \right|\right|_1 \tag{23}$$

$$g_4(\mathbf{H}_4^{(k)}\mathbf{a}, \lambda_4) = \lambda_4 \left|\left| \mathbf{H}_4^{(k)}\mathbf{a} \right|\right|_1 \tag{24}$$

where $\mathrm{diag}(\cdot)$ denotes a diagonal matrix, $\arg(\cdot)$ is the element-wise complex argument, and $\epsilon \ll 1$. Also, $\mathbf{I}$ denotes the identity matrix, and $\mathbf{F}$ is a first order difference matrix, having elements $\mathbf{F}\{n, n\} = 1$, $\mathbf{F}\{n, n+1\} = -1$, for $n = 1, \ldots, PL_{\max} - 1$, and zeros everywhere else. As intended, the minimization in (21) is convex, and may be solved using one of many convex solvers publicly available, such as, for instance, the interior point methods SeDuMi [14] or SDPT3 [9]. These are, however, quite computationally burdensome and will scale poorly with increased data length and larger grid. Instead, we here propose an efficient implementation using ADMM. In brief, ADMM is a method where the original problem is split into two or more subproblems, using a number of auxiliary variables, which are solved independently in

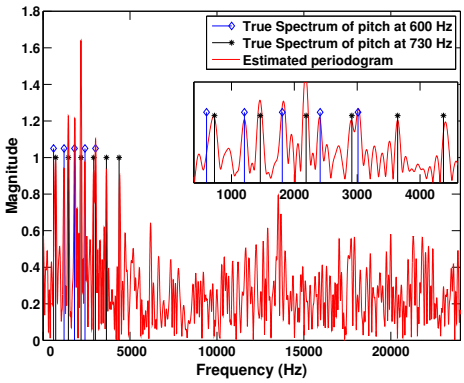

Fig. 1. The periodogram estimate and the true signal studied in Figure 2.

an iterative fashion. The problem in (21) may be implemented in a similar manner as was done [6], thus requiring only two tuning parameters, λ_2 and λ_4. The proposed method compares to PEBS and PEBS-TV introduced in [5] as improving upon the former, and requiring less tuning than the latter. We therefore term the proposed method PEBSI-Lite. An outline of its implementation is given in Algorithm 1 where $\mathbf{z}, \mathbf{u}, \mathbf{d}$ are the introduced auxiliary variables, μ is an inner convergence variable, and

$$\mathbf{G}^{(k)} = \left[\begin{array}{ccc} \mathbf{H}_1^T, & \mathbf{H}_2^{(k)T}, & \mathbf{H}_4^{(k)T} \end{array} \right]^T \tag{25}$$

$$\mathbf{u} = \left[\begin{array}{ccc} \mathbf{u}^{(1)T}, & \mathbf{u}^{(2)T}, & \mathbf{u}^{(3)T} \end{array} \right]^T \tag{26}$$

$$\mathbf{d} = \left[\begin{array}{ccc} \mathbf{d}^{(1)T}, & \mathbf{d}^{(2)T}, & \mathbf{d}^{(3)T} \end{array} \right]^T \tag{27}$$

$$\mathbf{T}\left(\mathbf{x}, \xi\right) = \frac{\max(|\mathbf{x}| - \xi, 0)}{\max(|\mathbf{x}| - \xi, 0) + \xi} \odot \mathbf{x} \tag{28}$$

such that the solution is given as $\hat{\mathbf{a}} = \mathbf{z}(\ell_{\text{end}})$ at iteration k_{end}.

4. NUMERICAL RESULTS

In order to examine the performance of the proposed estimator, we evaluate it using a simulated dual-pitch signal, measured in white Gaussian noise at different Signal-to-Noise Ratios (SNR), ranging from -5 dB to 20 dB in steps of 5 dB. At each level of SNR, 200 Monte Carlo simulations are performed, each simulation generating a signal with fundamental frequencies $[600, 730]$ Hz. To reflect the performance in presence of off-grid effects, the fundamental frequencies are randomly chosen at each simulation uniformly on $\pm d/2$, where d is the grid point spacing. The phases of the harmonics in each pitch are chosen uniformly on $[0, 2\pi)$, whereas all have unit magnitude.The signal is sampled at $f_s = 48$ kHz on a time frame of 10 ms, yielding $N = 480$ samples per frame. As a result, the pitches are spaced by just over f_s/N, which is the resolution limit of the periodogram. This is also seen in Figure 1, illustrating the resolution of the periodogram as well as the frequencies of the harmonics, at SNR $= -5$ dB. From the figure, it may be concluded that the signal contains more than one harmonic source, as the observed peaks are not harmonically related. Furthermore, it is clear that the fundamental frequencies are not separated by the periodogram, indicating that any pitch estimation algorithm based on the periodogram would suffer notable difficulties. In order to form our estimates, we begin by using a coarse dictionary with candidate pitches uniformly distributed on the interval $[280, 1500]$ Hz, thus also including $\omega_p/2$ and $2\omega_p$ for both pitches. The coarse resolution is $d = 10$ Hz, i.e., still a super-resolution of $1/10N$. After estimation on this grid, a zooming step is taken where a new grid with spacing $d/10$ is laid $\pm 2d$ around each pitch having non-zero power. This zooming approach is taken for the proposed method, as well as for PEBS and PEBS-TV. Comparisons are also made with the ANLS, ORTH, and the harmonic Capon estimators, which have been given the oracle model orders (see [15] for more details on these methods). The simulation and estimation procedure is performed for two cases; one where the number of harmonics L_k are set to $[5, 6]$ and one where L_k are set to $[10, 11]$. In the former case, we set $L_{\max} = 10$ and in the latter we set $L_{\max} = 20$, i.e. well above the true number of harmonics. Figures 2 and 3 show the percentage of pitch estimates where both lie within

Algorithm 1 The proposed PEBSI-Lite algorithm

initialize $k := 0$, $\mathbf{H}_4^{(0)} = \mathbf{F}$, and
  $\mathbf{a}^{(0)} = \mathbf{z}_{\text{save}} = \mathbf{d}_{\text{save}} = \mathbf{0}^{PL_{\max} \times 1}$
repeat {adaptive penalty scheme}
  initialize $\ell := 0$, $\mathbf{u}^{(2)}(0) = \mathbf{a}^{(k)}$,
    $\mathbf{z}(0) = \mathbf{z}_{\text{save}}$, and $\mathbf{d}(0) = \mathbf{d}_{\text{save}}$
  repeat {ADMM scheme}
    $\mathbf{z}(\ell) = \left(\mathbf{G}^{(k)H}\mathbf{G}^{(k)}\right)^{-1}\mathbf{G}^{(k)H}\left(\mathbf{u}(\ell) + \mathbf{d}(\ell)\right)$
    $\mathbf{u}^{(1)}(\ell+1)$
      $= \frac{\mathbf{y} - \mu\left(\mathbf{H}_1\mathbf{z}(\ell+1) - \mathbf{d}^{(1)}(\ell)\right)}{1+\mu}$
    $\mathbf{u}^{(2)}(\ell+1)$
      $= \mathbf{T}\left(\mathbf{H}_2\mathbf{z}(\ell+1) - \mathbf{d}^{(2)}(\ell), \frac{\lambda_2}{\mu}\right)$
    $\mathbf{u}^{(3)}(\ell+1)$
      $= \mathbf{T}\left(\mathbf{H}_4^{(k)}\mathbf{z}(\ell+1) - \mathbf{d}^{(3)}(\ell), \frac{\lambda_4}{\mu}\right)$
    $\mathbf{d}(\ell+1)$
      $= \mathbf{d}(\ell) - \left(\mathbf{G}^{(k)}\mathbf{z}(\ell+1) - \mathbf{u}(\ell+1)\right)$
    $\ell \leftarrow \ell + 1$
  until convergence
  store $\mathbf{a}^{(k)} = \mathbf{u}^{(2)}(\text{end})$, $\mathbf{z}_{\text{save}} = \mathbf{z}(\text{end})$, and
    $\mathbf{d}_{\text{save}} = \mathbf{d}(\text{end})$
  update $\mathbf{H}_4^{(k+1)} = \mathbf{F}\text{diag}\left(\arg\left(\mathbf{a}^{(k)}\right)\right)^{-1}$
  $k \leftarrow k + 1$
until convergence

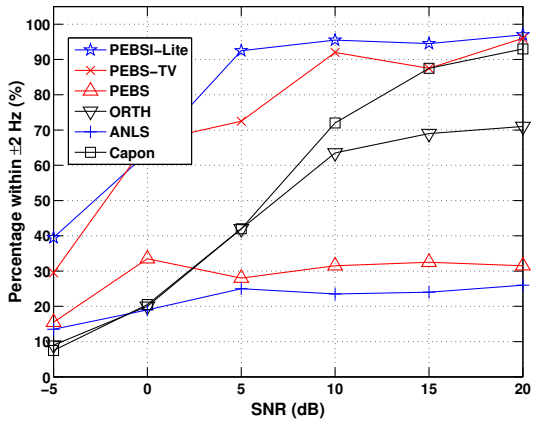

Fig. 2. Percentage of estimated pitches where both fundamental frequencies lie at most 2 Hz, or $d/5 = 1/50N$, from the ground truth, plotted as a function of SNR. Here, the pitches have $[5, 6]$ harmonics, respectively, and $L_{\max} = 10$.

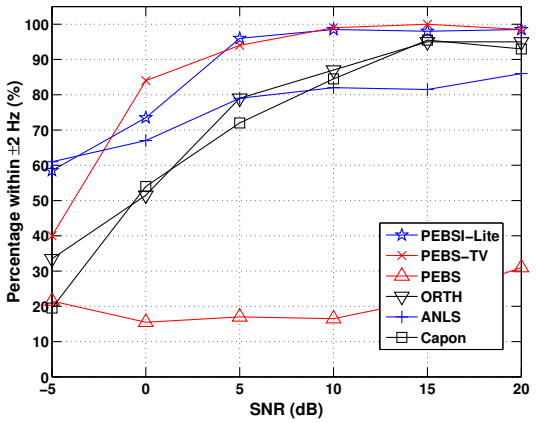

Fig. 3. Percentage of estimated pitches where both fundamental frequencies lie at most 2 Hz, or $d/5 = 1/50N$, from the ground truth, plotted as a function of SNR. Here, the pitches have $[10, 11]$ harmonics, respectively, and $L_{\max} = 20$.

± 2 Hz from the true values for the six compared methods, for the case of $[5, 6]$ and $[10, 11]$ harmonics, respectively. As is clear from the figures, the proposed method performs as well, or better, than the PEBS-TV algorithm, although requiring fewer tuning parameters. In this setting, PEBS performs poorly, as the generous choices of $L_{\max}$ allows it to ambiguously pick the sub-octave, as predicted.

5. REFERENCES

[1] M. Müller, D. P. W. Ellis, A. Klapuri, and G. Richard, "Signal Processing for Music Analysis," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 6, pp. 1088–1110, 2011.

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

[3] R. Gribonval and E. Bacry, "Harmonic decomposition of audio signals with matching pursuit," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 101–111, jan. 2003.

[4] 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.

[5] 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.

[6] M. A. T. Figueiredo and J. M. Bioucas-Dias, "Algorithms for imaging inverse problems under sparsity regularization," in *Proc. 3rd Int. Workshop on Cognitive Information Processing*, May 2012, pp. 1–6.

[7] A. Klapuri, "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 11, no. 6, pp. 804–816, 2003.

[8] T. Kronvall, M. Juhlin, S. I. Adalbjörnsson, and A. Jakobsson, "Sparse Chroma Estimation for Harmonic Audio," in *40th IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, Brisbane, Apr. 19-24 2015.

[9] 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.

[10] T. Kronvall, S. I. Adalbjörnsson, and A. Jakobsson, "Joint DOA and Multi-Pitch Estimation Using Block Sparsity," in *39th IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, Florence, May 4-9 2014.

[11] T. Kronvall, S. I. Adalbjörnsson, and A. Jakobsson, "Joint DOA and Multi-pitch estimation via Block Sparse Dictionary Learning," in *22nd European Signal Processing Conference*, Lisbon, Sept. 1-5 2014.

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

[13] E. J. Candes, 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.

[14] 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.

[15] 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.