LUND UNIVERSITY

Online Group-Sparse Regression Using the Covariance Fitting Criterion

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

Published in:
Proceedings of the 25th European Signal Processing Conference (EUSIPCO)

2017

Document Version:
Peer reviewed version (aka post-print)

Link to publication

Citation for published version (APA):
Kronvall, T., Adalbjörnsson, S. I., Nadig, S., & Jakobsson, A. (2017). Online Group-Sparse Regression Using the Covariance Fitting Criterion. In *Proceedings of the 25th European Signal Processing Conference (EUSIPCO)* (Vol. CFP1740S-USB). Article 1570347373 (European Signal Processing Conference (EUSIPCO)). EURASIP. http://www.eurasip.org/Proceedings/Eusipco/Eusipco2017/papers/1570347373.pdf

Total number of authors:
4

Creative Commons License:
Unspecified

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Download date: 18. May. 2024

ONLINE GROUP-SPARSE ESTIMATION USING THE COVARIANCE FITTING CRITERION

Ted Kronvall, Stefan Ingi Adalbjörnsson, Santhosh Nadig, and Andreas Jakobsson

Dept. of Mathematical Statistics, Lund University, Sweden

ABSTRACT

In this paper, we present a time-recursive implementation of a recent hyperparameter-free group-sparse estimation technique. This is achieved by reformulating the original method, termed group-SPICE, as a square-root group-LASSO with a suitable regularization level, for which a time-recursive implementation is derived. Using a proximal gradient step for lowering the computational cost, the proposed method may effectively cope with data sequences consisting of both stationary and non-stationary signals, such as transients, and/or amplitude modulated signals. Numerical examples illustrates the efficacy of the proposed method for both coherent Gaussian dictionaries and for the multi-pitch estimation problem.

***Index Terms*—** Online estimation, covariance fitting, group sparsity, multi-pitch estimation.

1. INTRODUCTION

Estimating a sparse parameter support for a high-dimensional regression problem has been the focus of much scientific attention during the last two decades, as this methodology has shown its usefulness in many applications, ranging from spectral analysis [1–3], array- [4–6] and audio processing [7–9], to biomedical modeling [10], and magnetic resonance imaging [11, 12]. In its vanguard, notable contributions were done by, among others, Donoho et al. [13] and Tibshirani et al. [14]. Their methods are effectively equivalent but are termed differently; the basis pursuit de-noising (BPDN) and the least absolute selection and shrinkage operator (LASSO), respectively, are nowadays a common component in the standard scientific toolboxes. These methods will estimate a parameter vector which reconstructs the signal using only a small number of regressors from the regressor matrix, i.e., a small number of columns from an (often highly underdetermined) linear system. More recently, a methodology termed the group-LASSO [15] was developed for modeling a signal where the sparse parameter support is assumed to be clustered into predefined groups, with the justification that some signal sources are better modeled by a group of regressors rather than just one. The above mentioned methods, as well as the vast majority of sparse estimators, have in common the requirement of selecting one or several hyperparameters, controlling the degree of sparsity in the solution. This may be done using, e.g., application-specific heuristics, cross-validation, or using some information criteria, which may often be computational burdensome and/or inaccurate. The discussed sparse estimation approaches typically assume having access to one or more offline frames of data, each having time-stationary signal support. For many applications, such as, for instance, audio processing, data is often generated online, with large correlation between consecutive frames, and with a varying degree of non-stationarity. To better accommodate these conditions, one may use a sparse recursive least squares (RLS) approach (see, e.g., [16, 17]), such as the one derived in [18] for the multi-pitch estimation problem. In a recent effort, the sparse iterative covariance-based estimator (SPICE) [19] utilizes a criteria for covariance fitting, originally developed within array processing, to form sparse estimates without the need of selecting hyperparameters. In fact, SPICE may shown to be equivalent to the square root (SR) LASSO [20]; in a covariance fitting sense, SPICE may be as a result be viewed as the optimal selection of the SR LASSO hyperparameter [21]. In this paper, we extend the method proposed in [22], which generalizes SPICE for grouped variables, along the lines of [23] to form recursive estimates in an online-fashion, reminiscent to the approach used in [24]. By first reformulating group-SPICE as an SR-LASSO, we then derive an efficient method for sparse recursive estimation formed via proximal gradient iterations, enabling recursive estimation of non-stationary signals. We justify the proposed method accordingly by numerical examples, illustrating its performance as on par with group-SPICE for stationary signals, and outperforming an online SPICE for group-sparse non-stationary signals.

2. NOTATIONAL CONVENTIONS

In this paper, we use the mathematical convention of letting lower-case letters, e.g., y, denote scalars, while lower-case bold-font letters, $\mathbf{y}$, denote column vectors and upper-case bold-font letters, $\mathbf{Y}$, denote matrices. Furthermore, $\mathbb{E}$ denotes the expectation operator, ∇ the first order derivative, and $(\cdot)^\top$ and $(\cdot)^H$ the transpose and hermitian transpose, respectively. Also, $|\cdot|$ denotes the absolute value of a complex number, while $\|\cdot\|_q$ and $\|\cdot\|_F$ denotes the ℓ_q-norm for

This work was supported in part by the Swedish Research Council and the Royal Physiographic Society in Lund.

$q \geq 1$ and the Frobenius norm, respectively. We let $\text{diag}(\mathbf{a})$ denote the diagonal matrix with diagonal vector $\mathbf{a}$, and $\text{tr}(\mathbf{A})$ the matrix trace of $\mathbf{A}$. We describe the structure of a matrix or vector by ordering elements within hard brackets, e.g., $\mathbf{y} = \begin{bmatrix} y(1) & y(2) \end{bmatrix}^\top$, while a set of elements is described using curly brackets, e.g., $\mathbb{N} = \{1, 2, \ldots\}$ denotes the set of natural numbers. We use subscripts to denote a subgroup of a vector or matrix, while time indices are indicated within parentheses, e.g., $\mathbf{x}_k(t)$ denotes the variables in subgroup $\mathbf{x}_k$ at time t. Superscript typically denotes a power operation, except for when the exponent is within parentheses, e.g., $x^{(j)}$, which denotes the j:th iteration of x. Finally, we also make use of notations $(x)_+ = \max(0, x)$, $\text{unit}(\mathbf{x}) = \mathbf{x}/\left\|\mathbf{x}\right\|_2$, and $x \in \text{Bin}(n, p)$, where the latter denotes that x is binomally distributed with n independent trials and probability parameter p.

3. GROUP-SPARSE ESTIMATION VIA THE COVARIANCE FITTING CRITERION

Here, we consider an N sample signal frame which may be reasonably well approximated by a select few variables in the linear signal model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e} \tag{1}$$

where $\mathbf{A} \in \mathbb{C}^{N \times M}$ and $\mathbf{x} \in \mathbb{C}^M$ denote the regressor matrix (or dictionary) and the response variable vector, respectively, and where $\mathbf{e}$ denotes the approximation error and noise. In our signal model, we assume that a possible signal source is represented by a sum of column vectors from the dictionary rather than just one, such that it may be clustered into K predefined groups,

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \ldots & \mathbf{A}_K \end{bmatrix} \tag{2}$$

$$\mathbf{A}_k = \begin{bmatrix} \mathbf{a}_{k,1} & \ldots & \mathbf{a}_{k,L_k} \end{bmatrix} \tag{3}$$

where the k:th group thus have L_k basis vectors, and consequently the dictionary has altogether $M = \sum_{k=1}^K L_k$ columns. By construction, we consider a group-sparse regression problem, where only a small number of the K possible groups are represented in the signal. We assume that $\mathbf{e}$ is reasonably homoscedastic, i.e., $\mathbb{E}(\mathbf{e}\mathbf{e}^H) = \sigma \boldsymbol{I}$, as well as that the variables in $\mathbf{x}$ are independent and identically distributed with a random phase, uniformly distributed over $[0, 2\pi)$. The covariance matrix may thus be expressed as

$$\mathbf{R} = \mathbb{E}(\mathbf{y}\mathbf{y}^H) = \mathbf{A}\mathbf{P}\mathbf{A}^H + \sigma \boldsymbol{I} \tag{4}$$

where $\mathbf{P}$ is a diagonal matrix with the diagonal vector

$$\mathbf{p} = \begin{bmatrix} \mathbf{p}_1 & \ldots & \mathbf{p}_K \end{bmatrix}^\top \tag{5}$$

$$\mathbf{p}_k = \begin{bmatrix} p_{k,1} & \ldots & p_{k,L_k} \end{bmatrix}^\top \tag{6}$$

which corresponds to the squared magnitude of the response variables, i.e.,

$$p_{k,\ell} = |x_{k,\ell}|^2 \tag{7}$$

for the ℓ:th component in the k:th dictionary group. To account for the group-sparse structure, we have relaxed the orginal covariance fitting criterion used in [19] by following the lines of [22] and thus seek to minimize

$$g(\mathbf{p}, \sigma) = \mathbf{y}^H \left(\mathbf{A}\mathbf{P}\mathbf{A}^H + \sigma \mathbf{I}\right)^{-1} \mathbf{y} + \sum_{k=1}^K v_k \left\|\mathbf{p}_k\right\|_\infty \tag{8}$$

with respect to the unknown variables $\mathbf{p}$ and σ, where

$$v_k = \sqrt{\text{tr}(\mathbf{A}_k^H \mathbf{A}_k)} = \left\|\mathbf{A}_k\right\|_F \tag{9}$$

Following the derivations in [22], minimizing (8) with respect to $\mathbf{p}$ and σ is equivalent of minimizing

$$g(\mathbf{x}) = \left\|\mathbf{y} - \mathbf{A}\mathbf{x}\right\|_2 + \sum_{k=1}^K \sqrt{\frac{v_k}{N}} \left\|\mathbf{x}_k\right\|_2 \tag{10}$$

with respect to the original variable $\mathbf{x}$, which is a square root group-LASSO [25] with the regularization parameter individually set for each group as $\sqrt{v_k/N}$.

4. RECURSIVE ESTIMATION VIA PROXIMAL GRADIENT

To allow for a recursive estimation of $\mathbf{x}$ which can be improved or changed adaptively as new samples are added, let $\mathbf{x}(n)$ denote a linear filter for some time point $n \in \{1 \leq n \leq N\}$. Also, we reformulate the first term in (10), here denoted $q(\cdot)$, such that a forgetting factor, $0 < \lambda \leq 1$, is utilized to give older samples less importance than newer samples, i.e.,

$$q(\mathbf{y}(n), \mathbf{x}(n)) = \sqrt{\sum_{t=1}^n \lambda^{n-t} |y(t) - \boldsymbol{\alpha}(t)^\top \mathbf{x}(t)|^2} \tag{11}$$

where $\mathbf{y}(n)$ and $\boldsymbol{\alpha}(t)^\top$ denote the vector of samples up to n and the t:th row of $\mathbf{A}$, respectively. On matrix form, $q(\cdot)$ may be equivalently formulated as

$$q(\mathbf{y}(n), \mathbf{x}(n)) = \left\| \sqrt{\boldsymbol{\Lambda}(n)} \Big(\mathbf{y}(n) - \mathbf{A}(n)\mathbf{x}(n) \Big) \right\|_2 \tag{12}$$

where $\mathbf{A}(n) = \begin{bmatrix} \boldsymbol{\alpha}(1) & \ldots & \boldsymbol{\alpha}(n) \end{bmatrix}^\top$ denotes the first n rows in $\mathbf{A}$, and where $\boldsymbol{\Lambda}(n) = \text{diag}(\begin{bmatrix} \lambda^{n-1} & \ldots & \lambda^0 \end{bmatrix})$. Our aim is to implement a proximal gradient algorithm reminiscent of [26] to estimate $\mathbf{x}(n), \forall n$, to which end one may iteratively upper-bound $q(\cdot)$ by centering it around the previous iteration's estimate, $\mathbf{x}^{(j-1)}(n)$, i.e.,

$$\begin{aligned} q(\mathbf{y}(n), \mathbf{x}^{(j)}(n)) \leq\ & q(\mathbf{y}(n), \mathbf{x}^{(j-1)}(n)) \\ & + (\mathbf{x}^{(j)}(n)) - \mathbf{x}^{(j-1)}(n)))^T \nabla q(\mathbf{y}(n), \mathbf{x}^{(j)}(n))) \\ & + \frac{1}{2h} \left\| \mathbf{x}^{(j)}(n)) - \mathbf{x}^{(j-1)}(n)) \right\|_2^2 \end{aligned} \tag{13}$$

Algorithm 1 The proposed online group-SPICE algorithm

Intitiate $n \leftarrow 0, \mathbf{R} \leftarrow \mathbf{0}, \mathbf{r} \leftarrow \mathbf{0}, \gamma \leftarrow 0$ and set $\mathbf{x}(n) = \mathbf{0}$
while $n < (N - \tau)$ **do**
  Reset $j \leftarrow 0$ and warm start $\mathbf{u}^{(j)} \leftarrow \mathbf{x}(n)$
  Add τ new samples and set $n \leftarrow n + \tau$
  Update $\mathbf{R}, \mathbf{r}$, and γ using (23)
  repeat {proximal gradient iterations}
    Update gradient $\nabla q(\mathbf{y}(n), \mathbf{u}^{(j)}))$ using (18)
    Take a gradient step, from $\mathbf{u}^{(j)}$ to $\mathbf{z}$, using (16)
    Apply group-wise shrinkage $\mathbf{u}_k^{(j+1)}$ using (15)
    $j \leftarrow j + 1$
  until convergence
  Save $\mathbf{x}(n) = \mathbf{u}_k^{(j)}$
end while

for some step size $h > 0$, and instead of minimizing (10) one may equivalently instead iteratively minimize [26]

$$\tilde{g} = \frac{1}{2h} \left\| \mathbf{x}^{(j)}(n)) - h\nabla q(\mathbf{y}(n), \mathbf{x}^{(j-1)}(n)) \right\|_2^2 + \sum_{k=1}^{K} \mu_k \left\| \mathbf{x}_k^{(j)}(n) \right\|_2 \tag{14}$$

for a suitable choice of regularization μ_k. By solving the sub-gradient equations of (14), previously shown in, e.g., [27] and here omitted due to page restrictions, one obtains the closed-form solution for the k:th group as

$$\mathbf{x}_k^{(j)} = (\|\mathbf{z}_k\| - h\mu_k)_+ \, \mathrm{unit}(\mathbf{z}_k) \tag{15}$$

where $\mathbf{z} = \begin{bmatrix} \mathbf{z}_1^\top & \dots & \mathbf{z}_K^\top \end{bmatrix}^\top$ is formed as

$$\mathbf{z} = \mathbf{x}^{(j-1)}(n) - h\nabla q \left(\mathbf{y}(n), \mathbf{x}^{(j-1)}(n) \right) \tag{16}$$

in which the gradient of (12) becomes

$$\nabla q(\mathbf{y}(n), \mathbf{x}(n)) = \frac{-\mathbf{A}(n)^H \mathbf{\Lambda}(n) \left(\mathbf{y}(n) - \mathbf{A}(n)\mathbf{x}(n) \right)}{\left\| \sqrt{\mathbf{\Lambda}(n)} \left(\mathbf{y}(n) - \mathbf{A}(n)\mathbf{x}(n) \right) \right\|_2} \tag{17}$$

wherein the superscript of $\mathbf{x}^{(j-1)}(n)$ was temporarily omitted for notational convenience.

5. EFFICIENT RECURSIVE UPDATES FOR NEW SAMPLES

One may facilitate an efficient estimation process when new samples are introduced by reusing old computations. To that end, the derivative (17) may be expressed as

$$\frac{\mathbf{R}(n)\mathbf{x}(n) - \mathbf{r}(n)}{\sqrt{\gamma(n) - 2\,\Re\left(\mathbf{r}(n)^H \mathbf{x}(n)\right) + \mathbf{x}(n)\mathbf{R}(n)\mathbf{x}(n)}} \tag{18}$$

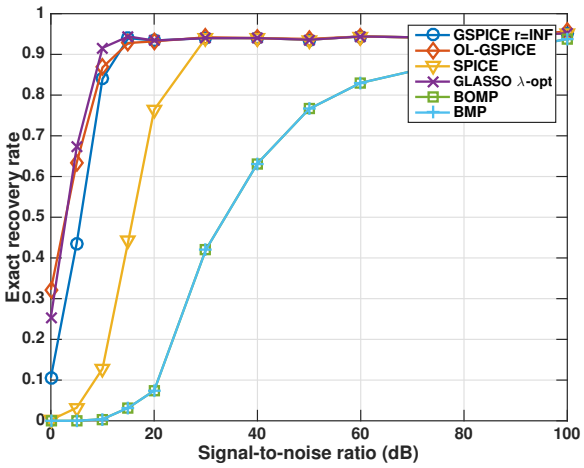

Fig. 1. Exact recovery rates from 500 Monte-Carlo samples estimated with online group-SPICE (OL-GSPICE) in comparison to other methods, where an coherent Gaussian dictionary is used, with $C = 3$ active groups.

where

$$\begin{aligned} \mathbf{r}(n) &= \mathbf{A}(n)^H \mathbf{\Lambda}(n) \mathbf{y}(n) \\ \mathbf{R}(n) &= \mathbf{A}(n)^H \mathbf{\Lambda}(n) \mathbf{A}(n) \\ \boldsymbol{\gamma}(n) &= \mathbf{y}(n)^H \mathbf{\Lambda}(n) \mathbf{y}(n) \end{aligned} \tag{19}$$

Let $\begin{bmatrix} y(n+1) & \dots & y(n+\tau) \end{bmatrix}^\top$, $\tau \in \mathbb{N}$ denote a vector of τ new samples available for estimation, and $(+\tau)$ the time indices from $n+1$ to $n+\tau$. Then,

$$\mathbf{y}(n+\tau) = \begin{bmatrix} \mathbf{y}(n)^\top & \mathbf{y}(+\tau)^\top \end{bmatrix}^\top \tag{20}$$

$$\mathbf{A}(n+\tau) = \begin{bmatrix} \mathbf{A}(n) \\ \mathbf{A}(+\tau) \end{bmatrix} \tag{21}$$

$$\mathbf{\Lambda}(n+\tau) = \begin{bmatrix} \lambda^\tau \mathbf{\Lambda}(n) & \mathbf{0} \\ \mathbf{0}^\top & \mathbf{\Lambda}(\tau) \end{bmatrix} \tag{22}$$

which, if inserted into (19), yields the updating formulas

$$\begin{aligned} \mathbf{r}(n+\tau) &= \lambda^\tau \mathbf{r}(n) + \mathbf{A}(+\tau)^H \mathbf{\Lambda}(\tau) \mathbf{y}(+\tau) \\ \mathbf{R}(n+\tau) &= \lambda^\tau \mathbf{R}(n) + \mathbf{A}(+\tau)^H \mathbf{\Lambda}(\tau) \mathbf{A}(+\tau) \\ \boldsymbol{\gamma}(n+\tau) &= \lambda^\tau \boldsymbol{\gamma}(n) + \mathbf{y}(+\tau)^H \mathbf{\Lambda}(\tau) \mathbf{y}(+\tau) \end{aligned} \tag{23}$$

The hyperparameters, μ_k, from (15), are in (10) defined as $\mu_k = \sqrt{v_k / N}$. In a time-recursive scheme, however, when new samples are added and older samples are given smaller importance, one must choose μ_k accordingly. As the sample size and dictionary matrix increase, governed the forgetting factor, one may select $\mu_k(n)$ as

$$\mu_k(n) = \sqrt{\frac{\sqrt{\mathrm{tr}\left(\mathbf{R}_{k,k}(n)\right)}}{(\lambda^n - 1)/(\lambda - 1)}} \tag{24}$$

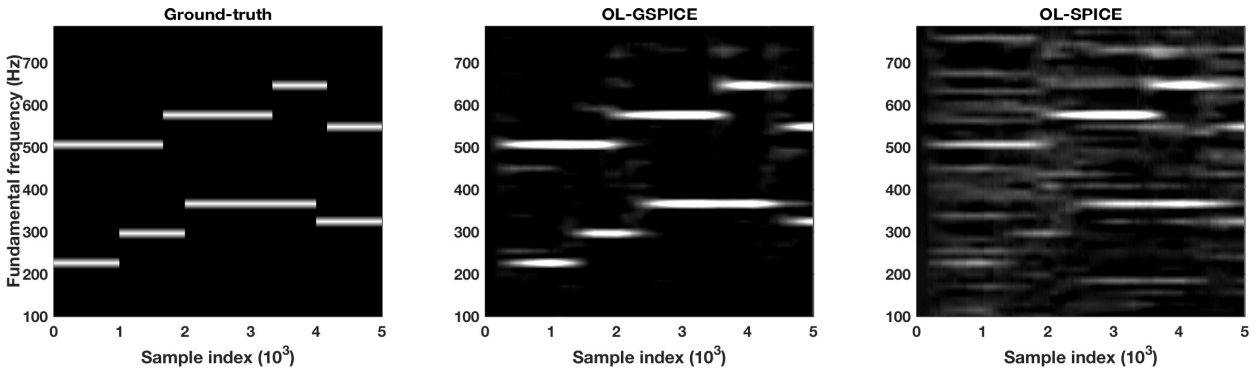

Fig. 2. True parameters for a simulated non-stationary multi-pitch signal (left), with corresponding estimates of the proposed method (middle), in comparison with the online-SPICE estimator (right).

where the denominator results from the geometric sum $\sum_{t=0}^{n-1} \lambda^t$ and $\mathbf{R}_{k,k}(t) = \mathbf{A}_k(t)^H \mathbf{\Lambda}(n) \mathbf{A}_k(n)$, which is obtained by choosing a submatrix of the recursively updated $\mathbf{R}(t)$ with rows and columns corresponding to group k. The step size, h, may, e.g., be chosen along the lines of [27]. Algorithm 1 summarizes the proposed method, which has computational complexity $\mathcal{O}(M^2)$. The main cost occurs at line 5 and is independent of the sample size, n.

6. NUMERICAL RESULTS

In this section, we compare the proposed estimator to relevant estimators for some different scenarios. We begin by examining the case of a cohrerent Gaussian dictionary, which is constructed by letting

$$\mathbf{a}_{k,\ell} = \sum_{\mathcal{I}^{\rho}_{k,\ell}} \mathbf{b}_{k',\ell'}, \quad \mathbf{b}_{k',\ell'} \in \mathcal{N}(\mathbf{0}, \boldsymbol{I}) \tag{25}$$

where $\mathbf{b}_{k,\ell}, \forall (k,\ell)$ are independent and identically distributed Gaussian vectors with zero mean and unit variance. The set $\mathcal{I}^{\rho}_{k,\ell}$ selects a mix of $n \in \text{Bin}\big(M - L_k, \rho\big)$ of these vectors, the indices of which are uniformly drawn from

$$(k', \ell') \in \{k : 1 \leq k \leq K, k' \neq k\} \times \{1 \leq \ell \leq L_k\} \tag{26}$$

This results in a dictionary of mixed Gaussian regressors, with no linearly dependent components within the groups, but for every component in a group, there will on average be $(M - L_k)\rho$ components in other groups to which it is linearly dependent. The parameter ρ thus controls the degree of regressor coherence, $0 \leq \rho \leq 1$. Figure 1 verifies the stationary performance of the proposed method in comparison with the non-recursive group-SPICE, the standard SPICE, the group-LASSO with an oracle choice of hyperparameter, as well as the (greedy) block matching pursuit [28] and block orthogonal matching pursuit [29]. The results are based on 500 Monte-Carlo (MC) simulations of $N = 100$ samples with $C = 3$ groups, having $L_1 = L_2 = L_3 = 10$ components, randomly drawn from a dictionary with $K = 200$ blocks of size $L = 10$, with dictionary coherence $\rho = 0.1$. To measure performance, we use the exact recovery rate (ERR) metric, defined as the rate of correct support recovery, i.e.,

$$\text{ERR}^{(i)} = 1\left\{\hat{\mathcal{I}}_C^{(i)} = \mathcal{I}_C^{(i)}\right\} \tag{27}$$

for the i:th MC simulation, averaged over all simulations, where $\mathcal{I}_C^{(i)}$ and $\hat{\mathcal{I}}_C^{(i)}$ denote the true and the estimated support, respectively. To be able to make comparisons with the abovementioned stationary estimators, we use $\lambda = 1$ for the proposed method. As can be seen from the figure, the online group-SPICE performs on par with a group-LASSO, which has been given the oracle hyperparameter, whereas SPICE (without grouping) yields significantly poorer results. Next, we examine estimation results for a multi-pitch dictionary, where the k:th candidate dictionary group in the dictionary is

$$\boldsymbol{\alpha}_k(t) = \begin{bmatrix} e^{i2\pi f_k/f_s 1t)} & \dots & e^{i2\pi f_k/f_s L_k t} \end{bmatrix}$$

at sample point t, i.e., where the regressors are Fourier vectors with frequencies at an integer multiple of the fundamental frequency candidate f_k. Here, we simulate a non-stationary signal by allowing $C = 2$ sources to have a dynamic support changing at random locations over a frame $N = 5 \cdot 10^3$ samples. We let the dictionary contain $K = 50$ candidate fundamental frequencies, f_k, uniformly spaced on $[100,\ 800)$ Hz, with $f_s = 44$ kHz. Figure 2 illustrates the true signal (left), the estimates of the proposed estimator (middle), and the estimates of the online SPICE (right). The figure clearly shows favorable performance of the online group-SPICE, whereas online SPICE is prone to misclassification. This is likely due to the harmonic structure of the multi-pitch dictionary, making it highly coherent, with the consequence that many erroneous candidate groups partly fit the signal.

7. REFERENCES

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

[2] H. C. S. Bourguignon and J. Idier, "A sparsity-based method for the estimation of spectral lines from irregularly sampled data," *IEEE J. Sel. Topics Signal Process.*, December 2007.

[3] 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, pp. 4649–4662, September 2016.

[4] I. F. Gorodnitsky and B. D. Rao, "Sparse Signal Reconstruction from Limited Data Using FOCUSS: A Re-weighted Minimum Norm Algorithm," *IEEE Trans. Signal Process.*, vol. 45, pp. 600–616, March 1997.

[5] D. Malioutov, M. Cetin, and A. S. Willsky, "A Sparse Signal Reconstruction Perspective for Source Localization With Sensor Arrays," *IEEE Trans. Signal Process.*, vol. 53, pp. 3010–3022, August 2005.

[6] S. I. Adalbjörnsson, T. Kronvall, S. Burgess, K. Åström, and A. Jakobsson, "Sparse Localization of Harmonic Audio Sources," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 24, pp. 117–129, Jan. 2016.

[7] 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.

[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] 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.

[10] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and Smoothness via the Fused Lasso," *Journal of the Royal Statistical Society B*, vol. 67, pp. 91–108, January 2005.

[11] A. S. Stern, D. L. Donoho, and J. C. Hoch, "NMR data processing using iterative thresholding and minimum l1-norm reconstruction," *J. Magn. Reson.*, vol. 188, no. 2, pp. 295–300, 2007.

[12] J. Swärd, S. I. Adalbjörnsson, and A. Jakobsson, "High Resolution Sparse Estimation of Exponentially Decaying N-dimensional Signals," *Elsevier Signal Processing*, vol. 128, pp. 309–317, Nov 2016.

[13] D. Donoho, M. Elad, and V. Temlyakov, "Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise," *IEEE Transactions on Information Theory*, vol. 52, pp. 6–18, Jan 2006.

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

[15] M. Yuan and Y. Lin, "Model Selection and Estimation in Regression with Grouped Variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006.

[16] D. Angelosante, J. A. Bazerque, and G. B. Giannakis, "Online Adaptive Estimation of Sparse Signals: Where RLS meets the ℓ_1-Norm," *IEEE Trans. Signal Process.*, vol. 58, 2010.

[17] B. Babadi, N. Kalouptsidis, and V. Tarokh, "SPARLS: The Sparse RLS Algorithm," *IEEE Trans. Signal Process.*, vol. 58, pp. 4013–4025, August 2010.

[18] F. Elvander, J. Swärd, and A. Jakobsson, "Online Estimation of Multiple Harmonic Signals," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, pp. 273–284, February 2017.

[19] P. Stoica, P. Babu, and J. Li, "New method of sparse parameter estimation in separable models and its use for spectral analysis of irregularly sampled data," *IEEE Trans. Signal Process.*, vol. 59, pp. 35–47, Jan 2011.

[20] A. Belloni, V. Chernozhukov, and L. Wang, "Square-Root LASSO: Pivotal Recovery of Sparse Signals via Conic Programming," *Biometrika*, vol. 98, no. 4, pp. 791–806, 2011.

[21] C. R. Rojas, D. Katselis, and H. Hjalmarsson, "A Note on the SPICE Method," *IEEE Trans. Signal Process.*, vol. 61, pp. 4545–4551, Sept. 2013.

[22] T. Kronvall, S. I. Adalbjrnsson, S. Nadig, and A. Jakobsson, "Group-Sparse Regression Using the Covariance Fitting Criterion," *Signal Processing*, vol. 139, pp. 116 – 130, 2017.

[23] D. Zachariah and P. Stoica, "Online Hyperparameter-Free Sparse Estimation Method," *IEEE Trans. Signal Process.*, vol. 63, pp. 3348–3359, July 2015.

[24] E. Eksioglu, "Group sparse RLS algorithms," *International Journal of Adaptive Control and Signal Processing*, vol. 28, pp. 1398–1412, 2014.

[25] F. Bunea, J. Lederer, and Y. She, "The Group Square-Root Lasso: Theoretical Properties and Fast Algorithms," *IEEE Trans. Inf. Theor.*, vol. 60, pp. 1313–1325, Feb. 2014.

[26] N. Parikh and S. Boyd, "Proximal Algorithms," *Found. Trends Optim.*, vol. 1, pp. 127–239, 2014.

[27] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "A Sparse-Group Lasso," *Journal of Computational and Graphical Statistics*, vol. 22, no. 2, pp. 231–245, 2013.

[28] L. Peotta and P. Vandergheynst, "Matching Pursuit with Block Incoherent Dictionaries," *IEEE Transactions on Signal Processing*, vol. 55, pp. 4549–4557, Sept 2007.

[29] Y. V. Eldar, P. Kuppinger, and H. Bolcskei, "Block-Sparse Signals: Uncertainty Relations and Efficient Recovery," *IEEE Transactions on Signal Processing*, vol. 58, no. 6, pp. 3042–3054, 2010.