



LUND UNIVERSITY

Generalized Sparse Covariance-based Estimation

J. SWÄRD, S. I. ADALBJÖRNSSON, AND A. JAKOBSSON

Published in: Elsevier Signal Processing
doi:10.1016/j.sigpro.2017.09.010

Lund 2018

Mathematical Statistics
Centre for Mathematical Sciences
Lund University

Generalized Sparse Covariance-based Estimation

Johan Swärd, Stefan I. Adalbjörnsson, and Andreas Jakobsson

Abstract—In this work, we generalize the recent sparse iterative covariance-based estimator (SPICE) by extending the problem formulation to allow for different norm constraints on the signal and noise parameters in the covariance model. The resulting extended SPICE algorithm offers the same benefits as the regular SPICE algorithm, including being hyper-parameter free, but the choice of norms allows further control of the sparsity in the resulting solution. We also show that the proposed extension is equivalent to solving a penalized regression problem, providing further insight into the differences between the extended and original SPICE formulations. The performance of the method is evaluated for different choices of norms, indicating the preferable performance of the extended formulation as compared to the original SPICE algorithm. Finally, we introduce two implementations of the proposed algorithm, one gridless formulating for the sinusoidal case, resulting in a semi-definite programming problem, and one grid-based, for which an efficient implementation is given.

I. INTRODUCTION

Many problems in signal processing may be well described using a linear model, such that

$$\mathbf{y} = \mathbf{B}\mathbf{x} + \mathbf{e} \quad (1)$$

where $\mathbf{y} \in \mathbb{C}^N$ is a vector of measurements, \mathbf{B} a matrix of regressors, \mathbf{x} the parameter vector, and \mathbf{e} denotes an additive (complex-valued) noise term, typically assumed to have zero mean and covariance matrix Σ . This model occurs in a wide range of applications, such as in, e.g., audio and speech processing [2,3] and spectroscopy [4]–[8].

Earlier works have primarily focused on *parametric* and *non-parametric* solutions to this problem. The latter kind of estimators typically do not assume any *a-priori* information about the signal, including assumptions on the model order or the signal structure. As a result, such techniques are more robust to uncertainties in the model assumptions that parametric solvers generally impose. However, this robustness also implies that non-parametric methods are, in general, not able to achieve the same level of performance as parametric approaches, given that the made model assumptions hold [9]. Recently, notable efforts have been made to combine these two approaches, developing so-called *semi-parametric* approaches, which typically only make some weak model structure assumptions, such that assuming that the solution is sparse, although restrain from making any stronger model order assumptions. This is done by forming the dictionary, $\mathbf{B} \in \mathbb{C}^{N \times M}$, using $M \gg N$ signal candidates, whereof only a few are assumed present in the signal. This allows the problem

to be reformulated as one of the subset of these M candidates best approximating the measured signal \mathbf{y} . This is typically done by enforcing sparsity on the vector \mathbf{x} , trading off model fit and the resulting level of sparsity. In [10], this was done by introducing the LASSO optimization problem

$$\underset{\mathbf{x}}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{y} - \mathbf{B}\mathbf{x}\|_2^2 + \mu \|\mathbf{x}\|_1 \quad (2)$$

where the first term penalizes the ℓ_2 -norm distance between the model and the signal, whereas the second term enforces sparsity upon the vector \mathbf{x} , with μ being a user parameter governing the trade-off between the two terms. During recent years, many other sparse techniques have been proposed (see, e.g., [11]–[16] and the references therein). Many of these methods suffer from the drawback of requiring the selection of one or many user parameters, often being a non-trivial task. In some cases, the user parameters may be selected using physical aspects, or via some kind of rule of thumb (see, e.g., [17]). Other ideas include solving the problem for all different values of the parameter [16,18], or to use some iterative process for aiding in the choice [11,19,20]. Another common way is to use cross-validation to find a suitable regularization parameter (see, e.g., [16]).

In [21], a novel sparse technique based on a covariance fitting criteria was proposed, avoiding the requirement of selecting any user parameters (see also [22]–[26]). The proposed minimization criteria was there formed as

$$\underset{\tilde{\mathbf{p}} \geq 0}{\text{minimize}} \quad \left\| \mathbf{R}^{1/2}(\tilde{\mathbf{p}}) (\mathbf{R}(\tilde{\mathbf{p}}) - \mathbf{y}\mathbf{y}^*) \right\|_F^2 \quad (3)$$

where $\|\cdot\|_F$ denotes the Frobenius norm, $(\cdot)^*$ the conjugate transpose, and where

$$\mathbf{R}(\tilde{\mathbf{p}}) = \mathbf{A}\mathbf{P}\mathbf{A}^* \quad (4)$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{I} \end{bmatrix} \quad (5)$$

$$\mathbf{P} = \begin{bmatrix} p_1 & \dots & p_M \end{bmatrix}^T \quad (6)$$

$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma_1 & \dots & \sigma_N \end{bmatrix}^T \quad (7)$$

$$\tilde{\mathbf{p}} = \begin{bmatrix} \mathbf{p}^T & \boldsymbol{\sigma}^T \end{bmatrix}^T \quad (8)$$

$$\mathbf{P} = \text{diag}(\tilde{\mathbf{p}}) \quad (9)$$

with \mathbf{I} denoting the $N \times N$ identity matrix, $(\cdot)^T$ the transpose, σ_k the noise variance for sample k , and $\text{diag}(\mathbf{z})$ the diagonal matrix with the vector \mathbf{z} along its diagonal, and zeros elsewhere. It was further shown that solving (3) is equivalent with solving [21]

$$\underset{\tilde{\mathbf{p}} \geq 0}{\text{minimize}} \quad \mathbf{y}^* \mathbf{R}^{-1}(\tilde{\mathbf{p}}) \mathbf{y} + \|\tilde{\mathbf{W}}\tilde{\mathbf{p}}\|_1 \quad (10)$$

where

$$\tilde{\mathbf{W}} = \text{diag} \left(\begin{bmatrix} w_1 & \dots & w_{M+N} \end{bmatrix} \right) \quad (11)$$

$$w_k = \|\mathbf{a}_k\|_2^2 / \|\mathbf{y}\|_2^2, \text{ for } k = 1, \dots, N + M \quad (12)$$

with \mathbf{a}_k denoting the k th column of \mathbf{A} . Clearly, both (2) and (10) minimize a signal fitting criteria. In the former case, this is done by minimizing the distance between the model and the data, whereas the latter measures the distance through the inverse of the (model) covariance matrix. Both models also impose an ℓ_1 norm constraint, with the first one penalizing the parameters corresponding to the different candidates in the dictionary \mathbf{B} , whereas the second, the so-called SPICE formulation, penalizes both the parameters corresponding to \mathbf{B} and the parameters corresponding to the noise.

In this work, we propose to generalize the SPICE formulation to allow for different penalties on \mathbf{p} and $\boldsymbol{\sigma}$, as given in (6) and (7), respectively, for two different cases. The first case considers the situation when all noise variances, σ_k , are equal, whereas the second considers the case when they are allowed to differ. In the case of equal noise variances, we show that the choice of norm for the noise parameters corresponds to different choices of the regularizing parameter, μ . In the case when the noise variances are allowed to be different, the choices of norms are similarly shown to affect the sparsity level. This results in the fact that even if the different SPICE formulations are hyper-parameter free, one may interpret the choices of norms as the equivalence of selecting hyper-parameters dictating the sparseness of the solution, and that the original SPICE version is equivalent to one particular choice of norms. We also provide an efficient grid-based implementation of the proposed method, which, indirectly, allows for solving (weighted) square-root LASSO problems for a wide choice of regularizing parameters. Additionally, we state a semi-positive programming (SDP) problem that allows for solving the proposed SPICE extension, for the sinusoidal case, without the use of a grid search.

II. THE $\{r, q\}$ -SPICE FORMULATION

It is worth noting that the second term in (10) penalizes the magnitude of each p_j and σ_k , thus promoting a sparse solution with only a few of the terms in $\tilde{\mathbf{p}}$ being non-zero. However, since the penalty does not distinguish between setting the different terms to zero, one may expect that some of the σ_k may be forced to be zero as a part of the minimization. If this happens, the result will be solutions that are less sparse than desired. Intuitively, this may be understood by interpreting (10) to require that \mathbf{R} is invertible. Thus, setting some σ_k to zero will cause the resulting covariance matrix, \mathbf{R} , to lose rank, unless some of the p_j are non-zero. This was also observed in [27], wherein a gridless formulation of SPICE was presented. For this formulation, it was shown that \mathbf{R} had full rank with probability one, resulting in an overestimation of the model order. As a result, forcing any σ_k to zero will yield a less sparse \mathbf{p} , thus increasing the estimated model order. This implies that, in the original SPICE formulation, σ_k and p_j are competing for the sparseness allowed in the solution of (10). In this work, we propose to treat the σ_k terms different from the rest of the p_j terms. A naive way of doing this could be to omit σ_k from the cost function of (10), but this would result in all the p_j terms being set to zeros, as σ_k may then take on any value which will make \mathbf{R} full rank, and will thus make

Algorithm 1 The $\{r, q\}$ -SPICE estimator with $r = 1$

- 1: Initiate $p_k^{(0)} = |\mathbf{b}_k^* \mathbf{y}|^2 / \|\mathbf{b}_k\|^4$, for $k = 1, \dots, M$, $\sigma_k^{(0)} = |y_k|$, for $k = 1, \dots, N$, and set $i = 1$
 - 2: **while** the termination criteria is not fulfilled **do**
 - 3: Let $\mathbf{R}^{(i)} = \mathbf{A} \mathbf{P}^{(i)} \mathbf{A}^*$
 - 4: Form λ from (44)
 - 5: Update $p_k^{(i)}$ from (42), for each $k = 1, \dots, M$
 - 6: Update $\sigma_k^{(i)}$ from (43), for each $k = 1, \dots, N$
 - 7: Set $i = i + 1$
 - 8: **end while**
-

the p_j terms redundant. Clearly, the σ_k terms must instead be penalized to produce a meaningful solution to (1). This may be done in different ways, for instance using

$$\underset{\mathbf{p} \geq 0, \boldsymbol{\sigma} \geq 0}{\text{minimize}} \mathbf{y}^* \mathbf{R}^{-1} \mathbf{y} + \|\mathbf{W} \mathbf{p}\|_r + \|\mathbf{W}_\sigma \boldsymbol{\sigma}\|_q \quad (13)$$

where $r, q \geq 1$, such that

$$\|\mathbf{W} \mathbf{p}\|_r = \left[\sum_{k=1}^M w_k^r p_k^r \right]^{1/r} \quad (14)$$

$$\|\mathbf{W}_\sigma \boldsymbol{\sigma}\|_q = \left[\sum_{k=1}^N w_{M+k}^q \sigma_k^q \right]^{1/q} \quad (15)$$

$$\mathbf{W} = \text{diag} \left([w_1 \quad \dots \quad w_M] \right) \quad (16)$$

$$\mathbf{W}_\sigma = \text{diag} \left([w_{M+1} \quad \dots \quad w_{M+N}] \right) \quad (17)$$

Thus, using $r = 1$ and $q = 1$ yields the original SPICE formulation. More general regularization functions could also be used. Furthermore, one could use an approach reminiscent of the one presented in [28], considering also the case when all $0 < r, q < 1$, resulting in a concave penalty term. However, in this work, we restricted our attention to the $\{r, q\}$ -norm case, using $r \geq 1$ and $q \geq 1$, terming the result the $\{r, q\}$ -SPICE formulation.

III. LINKING $\{R, Q\}$ -SPICE TO PENALIZED REGRESSION

To examine the implications of introducing the r - and q -norms in the SPICE formulation, we examine the connection between $\{r, q\}$ -SPICE and a penalized regression problem, such as the LASSO expression in (2). In doing so, we follow the derivation in [24,25], distinguishing between the case when each σ_k is allowed to have a distinct value, and the case when all σ_k are equal. To do so, we recall the following lemma (see also [25]):

Lemma 1. *Let*

$$\tilde{\mathbf{P}} = \text{diag} \left([p_1 \quad \dots \quad p_M] \right) \quad (18)$$

and

$$\boldsymbol{\Sigma} = \text{diag} \left([\sigma_1 \quad \dots \quad \sigma_N] \right) \quad (19)$$

Then,

$$\mathbf{y}^* \mathbf{R}^{-1} \mathbf{y} = \underset{\mathbf{x}}{\text{minimize}} (\mathbf{y} - \mathbf{B} \mathbf{x})^* \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{B} \mathbf{x}) + \sum_{k=1}^M |x_k|^2 / p_k \quad (20)$$

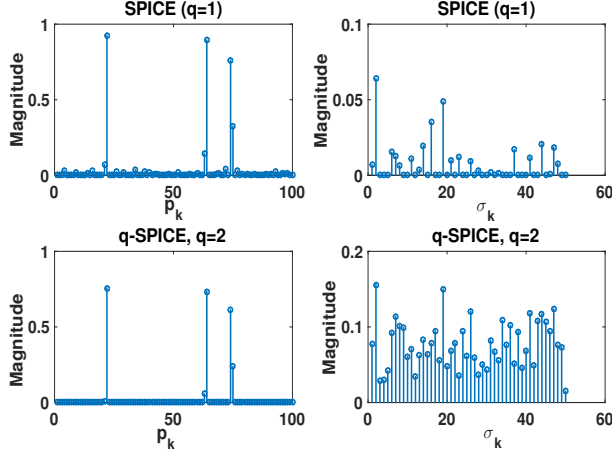


Fig. 1. The resulting estimates of $\tilde{\mathbf{p}}$ and σ from the SPICE and the q -SPICE estimator ($q=2$). Note that q -SPICE is sparser in $\tilde{\mathbf{p}}$, whereas SPICE is sparser in σ . In this example r is set to $r = 1$.

with the minimum occurring at

$$\hat{\mathbf{x}} = \Sigma \mathbf{B}^* \mathbf{R}^{-1} \mathbf{y} \quad (21)$$

A. Varying noise variance

Using Lemma 1, one may express (13) as

$$\begin{aligned} \text{minimize}_{\mathbf{x}, \mathbf{p}, \sigma} & \sum_{k=1}^N |y_k - \mathbf{b}_k^* \mathbf{x}|^2 / \sigma_k + \sum_{k=1}^M |x_k|^2 / p_k \\ & + \left(\sum_{k=1}^M w_k^r p_k^r \right)^{1/r} + \left(\sum_{k=1}^N w_{M+k}^q \sigma_k^q \right)^{1/q} \end{aligned} \quad (22)$$

Solving (22) for p_j yields

$$p_j = w_k^{-\frac{r}{r+1}} |x_k|^{\frac{2}{r+1}} \|\mathbf{W}^{1/2} \mathbf{x}\|_{\frac{2r}{r+1}}^{\frac{r-1}{r+1}} \quad (23)$$

Differentiating the function to be minimized in (22) with respect to σ_k and setting it to zero yields

$$-\frac{|y_k - \mathbf{b}_k^* \mathbf{x}|^2}{\sigma_k^2} + \frac{w_{M+k}^q \sigma_k^{q-1}}{\|\mathbf{W}_\sigma \boldsymbol{\sigma}\|_{\frac{q}{q-1}}} = 0 \quad (24)$$

Summing over k on both sides and simplifying, one arrives at

$$\|\mathbf{W}_\sigma \boldsymbol{\sigma}\|_q = \|\mathbf{W}_\sigma^{1/2} \mathbf{r}\|_{\frac{2q}{q+1}} \quad (25)$$

Inserting (25) into (24) yields

$$\sigma_k = w_{M+k}^{-\frac{q}{q+1}} |r_k|^{\frac{2}{q+1}} \left\| \mathbf{W}_\sigma^{1/2} \mathbf{r} \right\|_{\frac{2q}{q+1}}^{\frac{q-1}{q+1}} \quad (26)$$

Finally, inserting (23) and (26) into (22) yields

$$\text{minimize}_x \left\| \mathbf{W}_\sigma^{1/2} (\mathbf{y} - \mathbf{B}\mathbf{x}) \right\|_{\frac{2q}{q+1}} + \left\| \mathbf{W}^{1/2} \mathbf{x} \right\|_{\frac{2r}{r+1}} \quad (27)$$

As may be noted from the resulting expression, using $q = 1$ yields the least absolute deviations (LAD) estimate, whereas using $q = \infty$ yields the (unscaled) square-root LASSO. The implications of this is discussed further below. Regardless of the choice of q , the corresponding problem in (13) will still be scale invariant. This may be seen by following the example

Algorithm 2 The $\{r, q\}$ -SPICE estimator for equal σ_k with $r = 1$.

- 1: Initiate $p_k^{(0)} = |\mathbf{b}_k^* \mathbf{y}|^2 / \|\mathbf{b}_k\|^4$, for $k = 1, \dots, M$, $\sigma^{(0)} = \sqrt{\frac{1}{N-1} \sum_{k=1}^N (y_k - \bar{y})^2}$, for $k = 1, \dots, N$, and set $i = 1$
- 2: **while** the termination criteria is not fulfilled **do**
- 3: Let $\mathbf{R}^{(i)} = \mathbf{A} \mathbf{P}^{(i)} \mathbf{A}^*$
- 4: Form λ from (49)
- 5: Update $p_k^{(i)}$ from (47), for each $k = 1, \dots, M$
- 6: Update $\sigma_k^{(i)}$ from (48), for each $k = 1, \dots, N$
- 7: Set $i = i + 1$
- 8: **end while**

in [25], scaling each p_k and σ_k with a constant c and do the same for the cost function in (13), defining

$$\begin{aligned} g(\mathbf{p}, \sigma) & \triangleq c \mathbf{y}^* (\mathbf{A} c \mathbf{P} \mathbf{A}^*)^{-1} \mathbf{y} \\ & + c \left[\sum_{k=1}^M w_k^r c^r p_k^r \right]^{1/r} + c \left[\sum_{k=M+1}^{N+M} w_k^q c^q p_k^q \right]^{1/q} \\ & = \mathbf{y}^* (\mathbf{A} \mathbf{P} \mathbf{A}^*)^{-1} \mathbf{y} + c^2 \left[\sum_{k=1}^M w_k^r p_k^r \right]^{1/r} + \\ & c^2 \left[\sum_{k=M+1}^{N+M} w_k^q p_k^q \right]^{1/q} \end{aligned} \quad (28)$$

Let $f(\mathbf{p}, \boldsymbol{\sigma})$ denote the cost function in (13). Then, one may use Lemma 2 in [25] to conclude that if

$$\{\hat{\mathbf{p}}, \hat{\boldsymbol{\sigma}}\} = \arg \min_{\mathbf{p}, \boldsymbol{\sigma}} g(\mathbf{p}, \boldsymbol{\sigma}) \quad (29)$$

and

$$\{\hat{\tilde{\mathbf{p}}}, \hat{\tilde{\boldsymbol{\sigma}}}\} = \arg \min_{\tilde{\mathbf{p}}, \tilde{\boldsymbol{\sigma}}} f(\tilde{\mathbf{p}}, \tilde{\boldsymbol{\sigma}}) \quad (30)$$

then

$$\hat{\tilde{\mathbf{p}}} = c \hat{\mathbf{p}} \quad (31)$$

where $c > 0$, which is true in the here examined case as well. The observed scale invariance implies that one may view the $\{r, q\}$ -SPICE method as being hyper-parameter free in the same sense as the original SPICE algorithm is. Furthermore, it may be noted that when converting the p_k to x_k , using (21), any scaling will disappear.

B. Uniform noise variance

If, similar to [24,25], one instead assumes that all the noise terms have equal variance, treating the case when $\sigma_k = \sigma$, $\forall k$, one may observe interesting connection to the LASSO. Under these assumptions, it has been shown that the SPICE problem is connected to the (weighted) square-root LASSO problem [24,25], i.e.,

$$\text{minimize}_x \|\mathbf{y} - \mathbf{B}\mathbf{x}\|_2 + \mu \|\mathbf{W}^{1/2} \mathbf{x}\|_1 \quad (32)$$

where $\mu = N^{-1/2}$ yields the SPICE estimator. Following the derivation in Section III-A, together with the assumption that

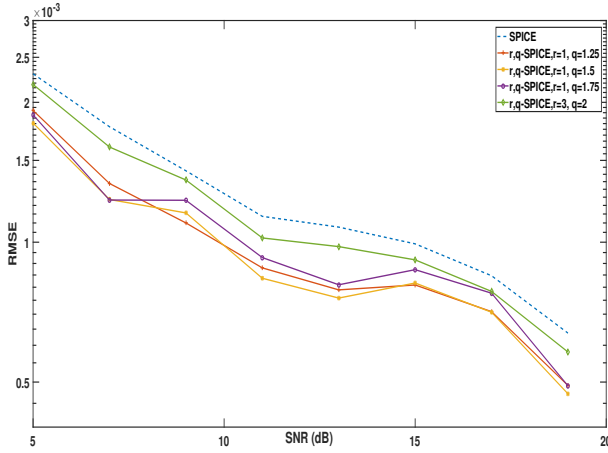


Fig. 2. The RMSE of the frequency estimates, as a function of SNR for $\{r, q\}$ -SPICE and SPICE.

all the noise terms have equal variance, yields $\mu = N^{-1/2q}$ for the $\{r, q\}$ -SPICE formulation, implying the equivalent formulation

$$\underset{\mathbf{x}}{\text{minimize}} \|\mathbf{y} - \mathbf{B}\mathbf{x}\|_2 + \mu \|\mathbf{W}^{1/2}\mathbf{x}\|_{\frac{2r}{r+1}} \quad (33)$$

As a result, the choice of q corresponds to selecting the weight that governs the trade-off between the model fitting term and the regularization of the parameters, whereas the choice of r decides which norm will be used in the regularization of the parameters. Thus, using $r = 1$ means that increasing q corresponds to increasing the sparsity in the (weighted) square-root LASSO; this implies that if the signal at hand is assumed to be sparse, solving $\{r, q\}$ -SPICE with $q > 1$ will yield preferable estimates. Furthermore, setting $r \rightarrow \infty$ yields a ridge regression problem, with q governing the amount of regularization. We note that it might be preferable to solve (33) using the $\{r, q\}$ -SPICE formulation, rather than solving (33) directly.

IV. EFFICIENT IMPLEMENTATION

As will be argued later, for sparse problems, the most interesting setting for $\{r, q\}$ -SPICE is when $r = 1$, since, according to (33), this will yield an ℓ_1 regularization. To this end, we will in this section derive an efficient implementation for this case. In [21], an efficient implementation of SPICE was introduced. To derive the steps of this algorithm, it was noted that the original SPICE minimization in (10) could also be expressed as

$$\underset{\{p_k \geq 0\}_{k=1}^M, \{\sigma_k \geq 0\}_{k=1}^N}{\text{minimize}} \mathbf{y}^* \mathbf{R}^{-1} \mathbf{y} \quad \text{subject to} \quad (34)$$

Furthermore, it was noted that one could further rewrite the objective in (34) by considering the optimization problem

$$\underset{\mathbf{Q}}{\text{minimize}} \mathbf{y}^* \mathbf{Q}^* \mathbf{P}^{-1} \mathbf{Q} \mathbf{y} \quad \text{subject to} \quad \mathbf{Q}^* \mathbf{A} = \mathbf{I} \quad (35)$$

which has the solution $\mathbf{Q}_0 = \mathbf{P} \mathbf{A}^* \mathbf{R}^{-1}$. By defining

$$\boldsymbol{\beta} = \mathbf{Q} \mathbf{y} \quad (36)$$

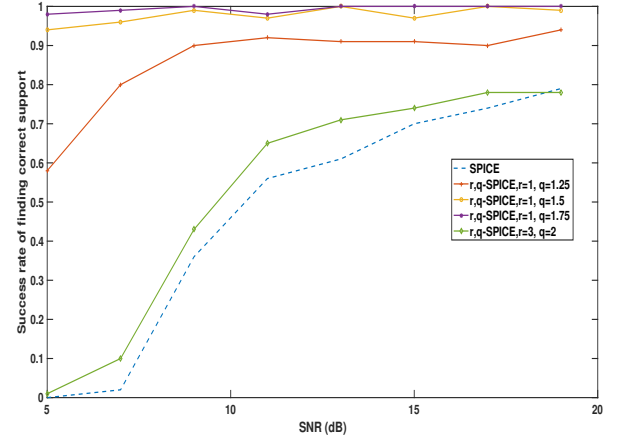


Fig. 3. The probability of finding the correct model order of the signal as a function of SNR for $\{r, q\}$ -SPICE and SPICE.

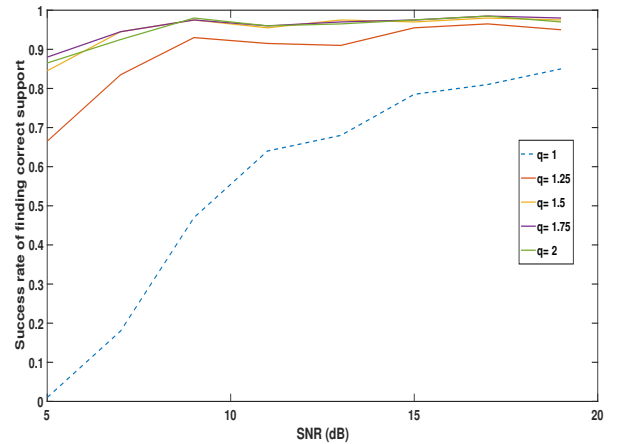


Fig. 4. The probability of finding the correct support of the signal as a function of q and SNR. Here, all the σ_k are assumed to be equal. In this example, $r = 1$.

one may rewrite (34) as

$$\underset{\{p_k \geq 0\}_{k=1}^M, \{\sigma_k \geq 0\}_{k=1}^N}{\text{minimize}} \sum_{k=1}^{M+N} \frac{|\beta_k|^2}{p_k} \quad (37)$$

$$\text{subject to} \quad \sum_{k=1}^M w_k p_k + \sum_{k=1}^N w_k \sigma_k = 1 \quad (38)$$

The estimates may then be found by iteratively updating \mathbf{R} and solving for p_k in (37). For $\{r, q\}$ -SPICE, with $r = 1$, when assuming different values for the σ_k , the same update for \mathbf{R} may be used, but instead of (37), one needs to solve

$$\underset{\{p_k \geq 0\}_{k=1}^M, \{\sigma_k \geq 0\}_{k=1}^N}{\text{minimize}} \sum_{k=1}^M \frac{|\beta_k|^2}{p_k} + \sum_{k=1}^N \frac{|\beta_{M+k}|^2}{\sigma_k} \quad (39)$$

$$\text{subject to} \quad \sum_{k=1}^{M+N} w_k p_k + \left(\sum_{k=1}^N w_{M+k}^q \sigma_k^q \right)^{1/q} = 1$$

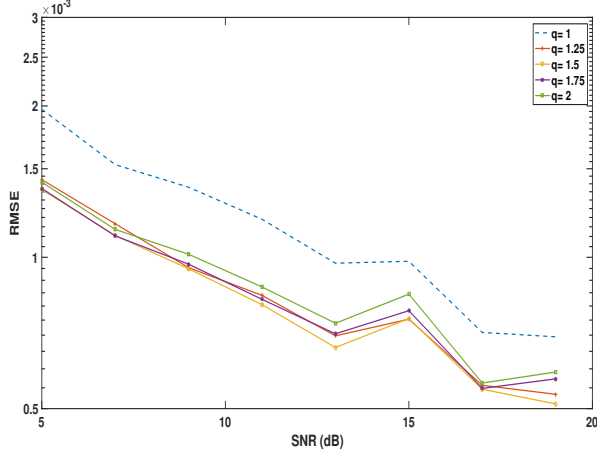


Fig. 5. The RMSE of the frequency estimates, as a function of q and SNR. Here, all the σ_k are assumed to be equal. In this example, $r = 1$.

From the Karush-Kuhn-Tucker (KKT) conditions [29], it follows that

$$-\frac{|\beta_k|^2}{p_k^2} + \lambda w_k = 0, \text{ for } k = 1, \dots, M \quad (40)$$

$$-\frac{|\beta_{M+k}|^2}{\sigma_k^2} + \lambda \sigma_k^q w_{M+k} \left(\sum_{k=1}^N w_{M+k}^q \sigma_k^{q-1} \right)^{1/q} = 0 \quad (41)$$

where λ denotes the dual variable, for $k = 1, \dots, M$, together with the constraint in (37). Solving these equation for each p_k and σ_k yields

$$p_k = \frac{|\beta_k|}{\sqrt{w_k \lambda^{1/2}}} \quad (42)$$

$$\sigma_\ell = \frac{|\beta_{M+\ell}|^{\frac{2}{q+1}} \|\mathbf{W}_\sigma^{1/2} \beta_\sigma\|^{\frac{q-1}{q+1}}}{w_{M+\ell}^{\frac{q}{q+1}} \lambda^{1/2}} \quad (43)$$

$$\lambda = \left(\|\mathbf{W}^{1/2} \beta\|_1 + \|\mathbf{W}_\sigma^{1/2} \beta_\sigma\|_{\frac{2q}{q+1}} \right)^2 \quad (44)$$

for $k = 1, \dots, M$ and $\ell = 1, \dots, N$, where

$$\beta = [\beta_1 \quad \dots \quad \beta_M]^T \quad (45)$$

$$\beta_\sigma = [\beta_{M+1} \quad \dots \quad \beta_{M+N}]^T \quad (46)$$

This allows for the formulation of an efficient implementation by iteratively forming \mathbf{R} from (4), β_k from (36), and p_k and σ_k from (42) and (43), respectively. Since $\{1, q\}$ -SPICE allows for a more sparse solution than the original SPICE, one may speed up the computations further by removing the zero valued p_k when forming \mathbf{R} and β_k .

When instead assuming that $\sigma_k = \sigma$, $\forall k$, one obtains the steps

$$p_k = \frac{|\beta_k|}{\sqrt{w_k \lambda^{1/2}}} \quad (47)$$

$$\sigma = \frac{\|\beta_M\|_2}{N^{1/2q} \lambda^{1/2}} \quad (48)$$

$$\lambda = \left(\|\mathbf{W}^{1/2} \beta\|_1 + \|\mathbf{W}^{1/2} \beta_\sigma\|_2 \right)^2 \quad (49)$$

for $k = 1, \dots, M$. Algorithms 1 and 2 summarize the $\{1, q\}$ -SPICE implementations for the two settings, with \bar{y} denoting the mean value of the vector \mathbf{y} . Similar to the previous case, since using $q > 1$ will enforce more sparsity than $q = 1$, one may utilize this added sparsity in the implementation of the algorithm. Since most of the elements in \mathbf{p} will be zero, one may form \mathbf{R}^{-1} by only considering the columns and rows of \mathbf{A} and \mathbf{A}^* corresponding to the non-zero entries in \mathbf{p} . Let $\hat{K}^{(i)}$ be the number of non-zero entries in $\mathbf{p}^{(i)}$ at iteration i . Then, if $\hat{K} < N$, one may use the Woodbury matrix identity to efficiently calculate the inverse of \mathbf{R} (see, e.g., [30]).

The termination criterias in Algorithms 1 and 2 can take on many forms. In this work, we have chosen to terminate the algorithms when the percentage of change in \mathbf{p} and σ between two consecutive iterations falls below a certain level, say in the range $[10^{-9}, 10^{-3}]$.

Note that the algorithm described in Algorithm 2 solves a (weighted) square-root LASSO problem, where the different choices of q corresponds to different levels of sparsity, i.e., different values of μ in (32). If one is interested in solving a (weighted) square-root LASSO with $\mu = \mu_0$, then one may instead solve the $\{r, q\}$ -SPICE with $q = -\frac{1}{2 \ln \mu_0}$, as long as $q > 1$, and with $r = 1$. Thus, the algorithm in Algorithm 2 presents an attractive and efficient way of solving the (weighted) square-root LASSO problem, for a large range of different μ .

To give an idea of the running time of the proposed algorithm as compared with a standard SDP solver (see, e.g., [31,32]), the algorithms were tested on a problem with $M = 10000$, $N = 1000$, and with $q = 5$, and $r = 1$, where the data vector, \mathbf{y} , contained 3 sinusoids, using a standard PC (2.6 Ghz Intel Core i7, 16 GB RAM). The corresponding run times were roughly 4 seconds for the Matlab implementation in Algorithm 2 and 4132 seconds for the SDP Matlab solver¹.

V. OFF-GRID SOLUTION

Many forms of estimation problems are solved by evaluating over a grid of the parameters of interest. However, such a solution may cause concerns when the sought solution falls outside the grid or may be found in between grid points. A common solution to this problem is to increase the grid size to thereby minimize the distance from the closest grid point to the true parameter value (see, e.g., [33,34]). However, such a solution might cause the columns of the extended dictionary to be highly correlated, thereby decreasing the performance of the method (we instead refer the interested reader to other works treating this issue, e.g., [34]–[37] and the references therein). In [27] and [38], an off-grid solution to the original SPICE version was presented for the sinusoidal case. In this section, we similarly provide one possible version of off-grid estimation for the proposed $\{r, q\}$ -SPICE method for a signal containing superimposed sinusoids. In order to do so, it may initially be noted that one may separate \mathbf{R} into two different matrices, such that

$$\mathbf{R} = \mathbf{B}^* \text{diag}(\mathbf{p}) \mathbf{B} + \text{diag}(\boldsymbol{\sigma}) \triangleq \mathbf{T}(\mathbf{u}) + \text{diag}(\boldsymbol{\sigma}) \quad (50)$$

¹Our implementation of $\{r, q\}$ -SPICE will be made available on the authors' web-pages upon publication.

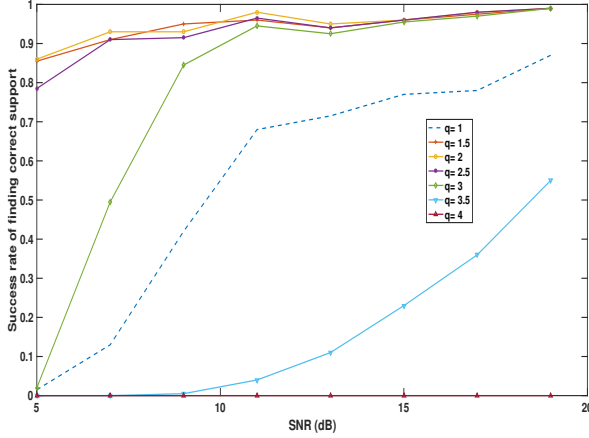


Fig. 6. The probability of finding the correct support of the signal as a function of q and SNR. Here, all the σ_k are assumed to be equal. Here, $r = 1$.

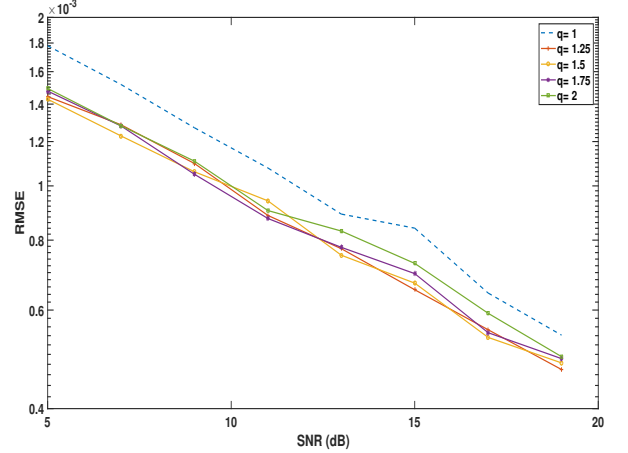


Fig. 8. The RMSE of the frequency estimates, as a function of q and SNR. Here, $r = 1$.

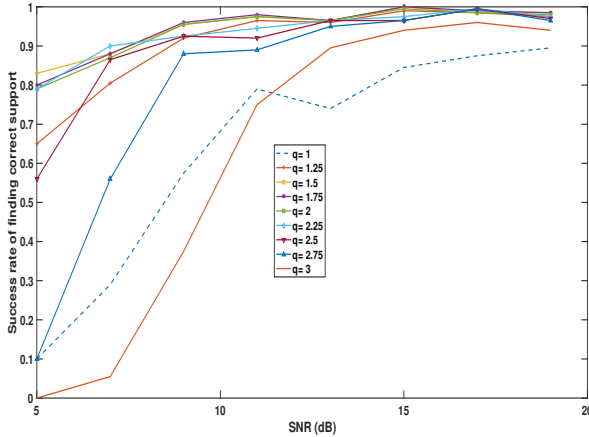


Fig. 7. The probability of finding the correct support of the signal as a function of q and SNR. Here, $r = 1$.

where $\mathbf{T}(\mathbf{u})$ is a Toeplitz matrix with \mathbf{u} forming the first column of $\mathbf{T}(\mathbf{u})$. Thus, (13) may be expressed as (see also [27,38])

$$\begin{aligned}
 & \underset{\mathbf{u}, \boldsymbol{\sigma}, x}{\text{minimize}} \quad \|\mathbf{y}\|_2^2 x + \|\text{diag}(\mathbf{T}(\mathbf{u}))\|_r + \|\mathbf{W}\boldsymbol{\sigma}\|_q \\
 & \text{subject to} \quad \begin{bmatrix} x & \mathbf{y}^* \\ \mathbf{y} & \mathbf{T}(\mathbf{u}) + \text{diag}(\boldsymbol{\sigma}) \end{bmatrix} \geq 0 \\
 & \quad \mathbf{T}(\mathbf{u}) \geq 0 \\
 & \quad \mathbf{T}(\mathbf{u}) - \mathbf{T}(\mathbf{u})^* = 0 \\
 & \quad \boldsymbol{\sigma} \geq 0
 \end{aligned} \tag{51}$$

and under the additional constraint that $\mathbf{T}(\mathbf{u})$ is a Toeplitz matrix. The optimization problem in (51) is convex, and may be solved using, e.g., a publicly available SDP solver, such as the one presented in [31,32]. The final off-grid estimates may then be found using the celebrated Vandermonde decomposition in combination with, for instance, Prony's method (see [9,39] for further details on such an approach).

VI. NUMERICAL EXAMPLES

Using the interpretation provided by the reformulation in Section III, it is clear that the choice of r will decide what kind of regularization that will be used. Thus, choosing $r = 1$ will yield an ℓ_1 norm and letting $r \rightarrow \infty$ will result in the ℓ_2 norm. In this paper, we consider sparse problems, and will therefore mainly confine our attention to the case where $r = 1$, since this will yield the most sparse convex regularizer, namely ℓ_1 .

From the discussion in Section II, one may expect that SPICE will set some of the elements in $\boldsymbol{\sigma}$ to zero, since the sparsity enforcing term in (10) also applies to these parameters. Figure 1 shows the estimated \mathbf{p} and $\boldsymbol{\sigma}$ for the SPICE and the $\{r, q\}$ -SPICE estimators, when applied to a linear signal formed using (1) with three non-zero components. As expected, using $r = 1$, $\{r, q\}$ -SPICE offers a sparser \mathbf{p} vector as compared to SPICE, whereas the solution is more sparse in $\boldsymbol{\sigma}$ for SPICE. As a result, the sparsity constraints on the σ_k terms in $\{r, q\}$ -SPICE are thus relaxed and are instead subjected to a bounding of their power in the q -norm, thus allowing for more sparsity in \mathbf{p} .

We will proceed by showing the difference in performance for different values of r and q , to provide an example on how the different choices of these norms affect the estimates. We investigate two properties of the estimators, namely the resulting root-mean-squared error (RMSE) of the frequency estimates, defined as

$$\text{RMSE} \triangleq \sqrt{\frac{1}{P} \sum_{k=1}^P |\hat{\theta}_k - \theta_k|^2} \tag{52}$$

where θ_k is the true frequency of the k th component, whereas $\hat{\theta}_k$ is the formed estimate, and the ability to correctly estimate the model order. The signal was $N = 50$ samples long and contained 4 sinusoids with unit magnitude and random phase. The simulation was done using 100 Monte-Carlo simulations for each SNR-level, where the signal-to-noise ratio (SNR) is

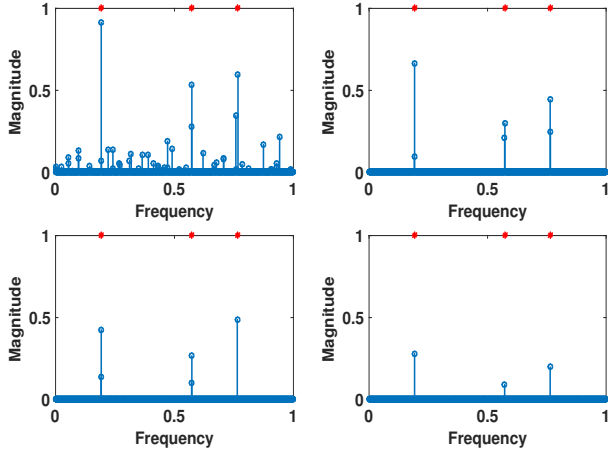


Fig. 9. A typical result from q -SPICE for different values of q . Top left: $q = 1$, top right $q = 1.5$, bottom left $q = 2$, and bottom right $q = 2.5$. The red stars indicate the position and size of the true sinusoids. In this example, $r = 1$.

defined as

$$\text{SNR} = 10 \log_{10} \left(\frac{P_y}{P_\sigma} \right) \quad (53)$$

with P_y denoting the power of the true signal and P_σ the power of the noise. The noise used was circular white Gaussian noise, and the noise terms were allowed to differ.

The solution was obtained by solving (51) for all settings except for the original SPICE, where the estimates were obtained from solving the problem formulated in [38]. In Figure 2, the resulting RMSEs are shown for different values of r and q , as a function of the SNR. To make the figures readable, 11 outliers were removed for SPICE and for the $r = 3, q = 2$ case for $\{r, q\}$ -SPICE each, whereas only 2, 5, and 5 outliers were removed for the case where $q = 1.25$, $q = 1.5$, and $q = 1.75$, respectively. Furthermore, to remove the noise peaks that appear when using small values of q , all peaks smaller than 20 % of the largest found peak were removed. Note, however, that this is not necessary for the case where q is larger. As is clear from the figure, the RMSE is decreased as the sparsity level is increased, with the $\{r, q\}$ -SPICE versions outperforming the original SPICE. This is also true for the resulting model order estimation, which is shown in Figure 3. As may be expected, when increasing q the sparsity is increased and the spurious peaks are removed, but as q is further increased, the true peaks start to disappear. In this setting, it seems to be beneficial to set the norms around $q = 1.5$ and $r = 1$. From these results, we conclude that the generalized version of SPICE allows for better estimation of parameter values, as well as model order. As was expected, using $r > 1$ was not beneficial when confronted with a sparse signal, and we will therefore, in the succeeding example, restrict our attention to the case where $r = 1$, referring to the method as q -SPICE. However, it should be stressed that for certain situations, it might be preferable to use $r > 1$, e.g., in situations when otherwise considering to use ridge regression; we will further examine this aspect in future works.

Arguably, the most important property of a sparse estimator is the ability to return the true support of the signal, as well as yielding reasonable amplitude estimates for this support. However, it seems inevitable that when including a sparsity enforcing penalty, one also introduced a (downwards) bias on the magnitude of the amplitudes. Fortunately, this problem is often easy to overcome by simply re-estimating the amplitudes using, e.g., least squares, once the true support is known. Accordingly, we will in this section focus on the methods ability of finding the true support of the signal. To this end, 200 Monte-Carlo simulation for each SNR level are formed. In each simulation, $N = 50$ samples of a signal containing three sinusoids, each with unit magnitude, and phase uniformly drawn from $(0, 2\pi]$, was created. The normalized frequencies were uniformly selected, but were at least $1/2N$ apart.

The dictionary contained $M = 1000$ candidate sinusoids, selected on a uniform frequency grid from $(0, 1]$. The estimated support was selected to be the elements of the vector \mathbf{x} that had a corresponding absolute value of at least 20% of the largest estimated value in \mathbf{x} . This was done to allow for comparison with the less sparse q -SPICE versions, for cases with small q value (most notably $q = 1$). It may be noted that for values of q that are large, this is not necessary. The support was deemed correctly estimated if the estimated frequencies were at most two grid points away from the true frequencies.

Figure 4 shows the results of applying q -SPICE, for different values of q , assuming that all the σ_k are the same, with $q = 1$ yielding the SPICE estimate. As is clear from the figure, the results improve with increasing q values. From the discussion in Section III-B, we note that this corresponds to increasing the value of μ in (32), thus increasing the sparsity in the estimates. Thus, one could assume that when further increasing q , the estimate of the support should decline. In Figure 6, this behavior can be seen, where now q -SPICE is evaluated over a range of larger q values. It is also apparent from the figure that the best value for q is for this signal somewhere around $q = 2$, which corresponds to using $\mu \approx 0.38$ in (33). Next, we investigate the precision for different values of q , by using the RMSE of the frequency estimates. Figure 5 shows the resulting RMSE of the frequency estimates, for the three largest values of \mathbf{x} . As can be seen in the figure, the RMSE is clearly improving as q is increased, corresponding to sparser solutions. For smaller values of q , the results are not very sparse, and large spurious noise peaks can be found. To improve readability, seven, two, and three outliers were removed from the cases $q = 1$, $q = 1.25$, and $q = 1.5$, respectively. If q is increased too much this will, of course, make the solution too sparse, thus risking setting non-noise peaks to zero. This can also be seen in Figure 6, where for about $q = 3$, the probability of retrieving the true support of the signal starts to decline, and at $q > 3.5$, the solution is too sparse.

We proceed by considering the case when the σ_k parameters are allowed to take on different values, using the same set-up as above. Figures 7 and 8 show the probability of estimating the correct support of the signal and the RMSE of the three largest frequency estimates, respectively. Again, in the interest of readability, three outliers were removed from $q = 1$, six

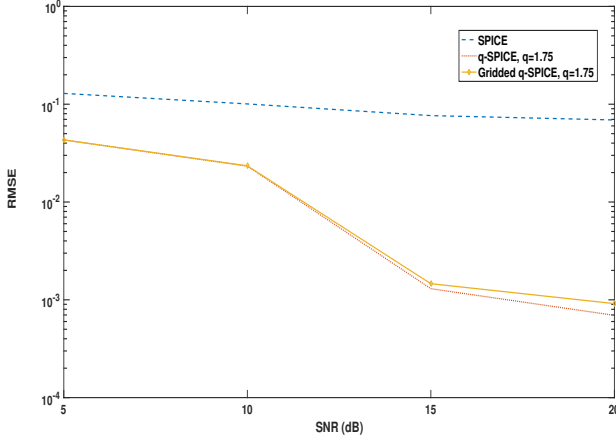


Fig. 10. The RMSE of the frequency estimates, as defined in (52), as a function of SNR for the gridless versions of q -SPICE and SPICE, together with the gridded version of q -SPICE.

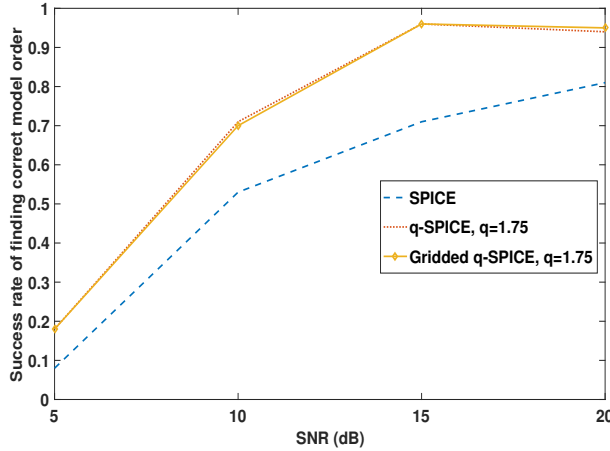


Fig. 11. The probability of finding the correct model order of the signal as a function of SNR for the gridless versions of q -SPICE and SPICE, together with the gridded version of q -SPICE.

outliers from $q = 1.25$, and three outliers for $q = 1.5$. As previously noted, it is clear from the figures that q governs the sparsity enforced on the solution. From the figures, one may also see that for this setup, it is advantageous to choose q in the interval $q = [1.25, 2.25]$. To demonstrate the differences in the solutions obtained from using different values of q , we show a typical simulation result for four different values of q , namely $q = 1, 1.5, 2$, and 2.5 , for the settings above, with $\text{SNR} = 5$ dB. Figure 9 shows the results, where it may again be noted that the sparsity level increases with q .

Finally, we provide a numerical example showing the results from solving the $\{r, q\}$ -SPICE using (51), with $r = 1$ and $q = 1.75$, and for the case where each noise variance are allowed to differ across the samples. In this scenario, we evaluated the gridless version of $\{r, q\}$ -SPICE, given in (51), and the gridless version of SPICE, given in [38], together with the grid-based $\{r, q\}$ -SPICE, given a frequency grid of $M = 500$ grid points. In each of the 100 Monte-Carlo

simulations, the $N = 50$ samples long signal contained four sinusoids, each with random phase, with two peaks having magnitude 4, one peak magnitude 2, and the last one unit magnitude. The frequencies were selected not to be closer than $1/2N$ from each other and were randomly selected in each simulation from the interval $(0, 1]$. Both gridless versions were computed using the SDP-solver in CVX [31,32]. Figure 10 and 11 show the resulting RMSE and probability of finding the correct support as functions of the SNR level. As seen in the figures, the two versions of the q -SPICE outperforms the gridless version of SPICE. It is worth noting that in this scenario, only SPICE had the benefit of removing the smallest peaks. Furthermore, the model order was deemed correct if the method found the true number of peaks, thus there were no limitation on how close an estimated frequency had to be the true value. If the model order was too high, the four largest peaks were selected to compute the RMSE, whereas if the model order was too low, these estimates were omitted from the RMSE evaluations.

Furthermore, one may see that the gridless version of q -SPICE is slightly better than the gridded version. However, this slight improvement from using the gridless q -SPICE version may not be worth the extra computation time; the gridless version took on average 9.4 seconds to execute, whereas the gridded version only took 0.5 seconds. However, it is worth recalling that other works on gridless solutions implicate that faster implementations are available (see, e.g., [40]), and these improvements in implementation can likely also be applied to the gridless q -SPICE.

VII. CONCLUSION

In this paper, we introduced a generalization of the SPICE method, in which we allow for a trade-off between the penalties for the model, using a q -norm, and the noise parameters, using an r -norm. We show that for larger values of q , one achieves a higher level of sparsity and better performance for recovering the support of the signal. Furthermore, we show that the proposed method is equivalent to a penalized regression formulation, with the $\frac{2q}{q+1}$ norm on the model fit, for the case when we let the noise variance vary across all samples. In the case where the noise variance is assumed to be equal for all samples, it is shown that the proposed method is equal to the (weighted) square-root LASSO, where the regularization parameter has a one-to-one correspondence to the choice of q for a given problem. Furthermore, we provide a fast and efficient implementation for both the case when $r = 1$ and the noise variances are equal for all samples, and where they are allowed to differ. As a result of the shown equivalence, the presented implementation offers an attractive alternative for solving $\frac{2q}{q+1}$ -norm problems, and, perhaps more interesting, (weighted) square-root LASSO problems for different regularization parameters. We also present a gridless version of $\{r, q\}$ -SPICE for the sinusoidal signals, which is on the form of an SDP problem. Numerical result show the preferred performance of the $\{r, q\}$ -SPICE as compared to the original SPICE method, both for gridded and for gridless versions for the estimator.

REFERENCES

- [1] J. Swärd, S. I. Adalbjörnsson, and A. Jakobsson, "A Generalization of the Sparse Iterative Covariance-based Estimator," in *42nd IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, New Orleans, USA, March, 5-9 2017.
- [2] M. Christensen and A. Jakobsson, *Multi-Pitch Estimation*, Morgan & Claypool, San Rafael, Calif., 2009.
- [3] 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.
- [4] Y. Li, J. Razavilar, and K. J. R. Liu, "A High-Resolution Technique for Multidimensional NMR Spectroscopy," *IEEE Trans. Biomed. Eng.*, vol. 45, no. 1, pp. 78–86, 1998.
- [5] W. Sun and H. C. So, "Accurate and Computationally Efficient Tensor-Based Subspace Approach for Multidimensional Harmonic Retrieval," *IEEE Trans. Signal Process.*, vol. 60, no. 10, pp. 5077–5088, Oct. 2012.
- [6] S. D. Somasundaram, A. Jakobsson, J. A. S. Smith, and K. Althofer, "Exploiting Spin Echo Decay in the Detection of Nuclear Quadrupole Resonance Signals," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 925–933, April 2007.
- [7] Y. Tan, S. L. Tantum, and L. M. Collins, "Cramér-Rao Lower Bound for Estimating Quadrupole Resonance Signals in Non-Gaussian Noise," *IEEE Signal Process. Lett.*, vol. 11, no. 5, pp. 490–493, May 2004.
- [8] 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.
- [9] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, Upper Saddle River, N.J., 2005.
- [10] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society B*, vol. 58, no. 1, pp. 267–288, 1996.
- [11] 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.
- [12] 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, no. 1, pp. 91–108, January 2005.
- [13] J. Fang, J. Li, Y. Shen, H. Li, and S. Li, "Super-Resolution Compressed Sensing: An Iterative Reweighted Algorithm for Joint Parameter Learning and Sparse Signal Recovery," *IEEE Signal Process. Lett.*, vol. 21, no. 6, pp. 761–765, 2014.
- [14] 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, no. 3, pp. 600–616, March 1997.
- [15] X. Tan, W. Roberts, J. Li, and P. Stoica, "Sparse Learning via Iterative Minimization With Application to MIMO Radar Imaging," *IEEE Trans. Signal Process.*, vol. 59, no. 3, pp. 1088–1101, March 2011.
- [16] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations*, Chapman and Hall/CRC, 2015.
- [17] J. Swärd, J. Brynolfsson, A. Jakobsson, and M. Hansson-Sandsten, "Sparse Semi-Parametric Estimation of Harmonic Chirp Signals," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1798–1807, April 2016.
- [18] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of Statistics*, vol. 32, no. 2, pp. 407–499, April 2004.
- [19] D. Wipf and S. Nagarajan, "Iterative Reweighted ℓ_1 and ℓ_2 Methods for Finding Sparse Solutions," *IEEE J. Sel. Topics in Signal Processing*, vol. 4, pp. 317–329, 2010.
- [20] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, "Iteratively reweighted least squares minimization for sparse recovery," *Comm. Pure Appl. Math.*, vol. 63, 2010.
- [21] 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, no. 1, pp. 35–47, Jan 2011.
- [22] P. Stoica, P. Babu, and J. Li, "SPICE : a novel covariance-based sparse estimation method for array processing," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 629 –638, Feb. 2011.
- [23] P. Stoica and P. Babu, "SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation," *Signal Processing*, vol. 92, no. 7, pp. 1580–1590, July 2012.
- [24] C. R. Rojas, D. Katselis, and H. Hjalmarsson, "A Note on the SPICE Method," *IEEE Trans. Signal Process.*, vol. 61, no. 18, pp. 4545–4551, Sept. 2013.
- [25] P. Stoica, D. Zachariah, and L. Li, "Weighted SPICE: A Unified Approach for Hyperparameter-Free Sparse Estimation," *Digit. Signal Process.*, vol. 33, pp. 1–12, October 2014.
- [26] D. Zachariah and P. Stoica, "Online Hyperparameter-Free Sparse Estimation Method," *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3348–3359, July 2015.
- [27] 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.
- [28] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 707–710, Oct. 2007.
- [29] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.
- [30] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The John Hopkins University Press, 4th edition, 2013.
- [31] Inc. CVX Research, "CVX: Matlab Software for Disciplined Convex Programming, version 2.0 beta," <http://cvxr.com/cvx>, Sept. 2012.
- [32] M. Grant and S. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pp. 95–110. Springer-Verlag Limited, 2008, http://stanford.edu/~boyd/graph_dcp.html.
- [33] P. Stoica and P. Babu, "Sparse Estimation of Spectral Lines: Grid Selection Problems and Their Solutions," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 962–967, Feb. 2012.
- [34] 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.
- [35] 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.
- [36] 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.
- [37] Z. Yang, L. Xie, and C. Zhang, "Off-Grid Direction of Arrival Estimation Using Sparse Bayesian Inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38 – 43, October 2012.
- [38] P. Stoica, G. Tang, Z. Yang, and D. Zachariah, "Gridless Compressed-Sensing Methods for Frequency Estimation: Points of Tangency and Links to Basics," in *22nd European Signal Processing Conference*, Lisbon, Portugal, 2014.
- [39] T. Blu, P. L. Dragotti, M. Vetterli, P. Marziliano, and L. Coulot, "Sparse Sampling of Signal Innovations," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 31 – 40, March 2008.
- [40] 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.