

A GENERALIZATION OF THE SPARSE ITERATIVE COVARIANCE-BASED ESTIMATOR

Johan Swärd, Stefan Ingi Adalbjörnsson, and Andreas Jakobsson

Dept. of Mathematical Statistics, Lund University, Sweden

ABSTRACT

In this work, we extend the popular sparse iterative covariance-based estimator (SPICE) by generalizing the formulation to allow for different norm constraint on the signal and noise parameters in the covariance model. For any choice of norms, the resulting generalized SPICE method enjoys the same benefits as the regular SPICE method, including being hyperparameter free, although the choice of norm is shown to govern the sparsity in the resulting solution. Furthermore, we show that there is a connection between the generalized SPICE and a penalized regression problem, both for the case where one allows the noise parameters to differ for each sample, and when treating each noise parameter as being equal. We examine the performance of the method for different choices of norms, and compare the results to the original SPICE method, showing the benefits of using the generalized version. We also provide a way of solving the generalized SPICE using a gridless method, which solves a semi-definite programming problem.

Index Terms— Covariance fitting, sparse reconstruction, convex optimization

1. 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 the 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 [1, 2] and spectroscopy [3, 4].

Historically, there have been two main principles available for solving these kinds of problems: *parametric* and *non-parametric* methods. The latter do not rely on any *a priori* information about the signal, including assumptions on the model structure or order, and are therefore more robust to uncertainties in such model assumptions than the former.

However, this robustness comes with the downside that the non-parametric methods have, in general, larger variance in the estimates compared to the parametric approaches, which typically in turn are less robust [5]. Recently, notable efforts have been made to combine these two approaches, developing so-called semi-parametric approaches, which typically make some model structure assumptions, although restrain from making strong model order assumptions, other than assuming that the solution is sparse. This implies that although the dictionary, $\mathbf{B} \in \mathbb{C}^{N \times M}$, is formed using $M \gg N$ signal candidates, only a few of these candidates are assumed present in the signal. The problem is thus transformed into finding the subset of these M candidates best approximating the measured signal, \mathbf{y} . Many sparse methods do this by enforcing sparsity on the vector \mathbf{x} , creating a trade-off between the model fit and the level of sparsity. Recently, many other sparse methods have been proposed (see, e.g., [6–11] and the references therein). One potential drawback of these methods is the selection of the user parameter, which is often a non-trivial task. Sometimes there are physical aspects that may aid in the choice of this parameter, whereas, in other, some kind of rule of thumb on how to choose it may be found [12]. Other ideas include solving the problem for all different values of the parameter [11], or to use some iterative process for aiding in the choice [6, 13]. Another common way is to use cross-validation to find a suitable regularization parameter (see, e.g., [11]). In [14], a novel approach to form a sparse method was proposed based on a covariance fitting criteria, and was shown to overcome the drawback of selecting the user parameter (see also [15–19]). The minimization criteria that was proposed was

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

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

$$\mathbf{R}(\mathbf{p}) = \mathbf{A}\mathbf{P}\mathbf{A}^* \quad (3)$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{B} & \mathbf{I} \end{bmatrix} \quad (4)$$

$$\tilde{\mathbf{p}} = \begin{bmatrix} p_1 & \dots & p_M \end{bmatrix}^T \quad (5)$$

$$\boldsymbol{\sigma} = \begin{bmatrix} \sigma_1 & \dots & \sigma_N \end{bmatrix}^T \quad (6)$$

$$\mathbf{p} = \begin{bmatrix} \tilde{\mathbf{p}}^T & \boldsymbol{\sigma}^T \end{bmatrix}^T \quad (7)$$

$$\mathbf{P} = \text{diag}(\mathbf{p}) \quad (8)$$

This work was supported in part by the Swedish Research Council and the Crafoord's and Carl Trygger's foundations.

with \mathbf{I} denoting the $N \times N$ identity matrix, $(\cdot)^T$ the transpose, σ_k^2 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 (2) is equivalent to solving

$$\underset{\mathbf{p} \geq 0}{\text{minimize}} \mathbf{y}^* \mathbf{R}^{-1}(\mathbf{p}) \mathbf{y} + \|\tilde{\mathbf{W}} \mathbf{p}\|_1 \quad (9)$$

where $\|\cdot\|_p$ denotes the p -norm, and

$$\tilde{\mathbf{W}} = \text{diag} \left(\begin{bmatrix} w_1 & \dots & w_{M+N} \end{bmatrix} \right) \quad (10)$$

$$w_k = \|\mathbf{a}_k\|_2 \quad (11)$$

with \mathbf{a}_k denoting the k th column of \mathbf{A} . It is clear from the formulation in (9) that the problem promotes a sparse solution, as a result of the ℓ_1 norm, which penalizes both the parameters corresponding to \mathbf{B} and the parameters corresponding to the noise.

In this paper, we generalize the SPICE approach to allow for different penalties for $\tilde{\mathbf{p}}$ and $\boldsymbol{\sigma}$ given in (5) and (6), respectively. For the case where all the σ_k are equal, we show that the choice of the norm for the noise parameters corresponds to the different choices of the regularizing parameter, μ , for a penalized regression problem. In the case when all σ_k are allowed to be different, the choice of norm is similarly shown to also affect the sparsity level. This results in the fact that even if the different SPICE formulations are hyper-parameter free, one may interpret the choice of norm as a hyper-parameter for the sparseness of the solution.

2. THE $\{R, Q\}$ -SPICE FORMULATION

It is worth noting that the second term in (9) penalizes the magnitude of each p_j and σ_k , thus promoting a sparse solution with only a few of the terms in \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 one is interested in finding a sparse solution from the columns of the dictionary \mathbf{B} , which would be the case for, e.g., the LASSO, setting some of the noise parameters σ_k to zero makes little sense. Another intuition is given if one interprets (9) as an implicit restriction on \mathbf{R} such that it must be invertible. Thus the total number of σ_k and p_j that can be set to zero is restricted, and in fact, setting any σ_k to zero is problematic as the resulting covariance matrix, \mathbf{R} , loses rank, unless some of the p_j are non-zero. Similar conclusions were stated in [20], where a gridless formulation of SPICE was derived. It was shown that for the gridless version of SPICE, \mathbf{R} had full rank with probability one, which in turn made the method overestimate the model order. Consequently, setting many σ_k to zero, will force the resulting $\tilde{\mathbf{p}}$ to be less sparse, thus increasing the estimated model order. Thus, in the SPICE formulation, σ_k and p_j are competing for the

sparseness allowed in the solution of (9). This suggests that if one could allow for other norms constraining the parameters in (9), the result might be improved. In this work, we investigate such a generalization to the SPICE method, and show how this affects the solution. One possible extension may be formulated as, where $\mathbf{R} = \mathbf{R}(\mathbf{p})$ for notational convenience

$$\underset{\mathbf{p} \geq 0}{\text{minimize}} \mathbf{y}^* \mathbf{R}^{-1} \mathbf{y} + \|\mathbf{W} \mathbf{p}\|_r + \|\mathbf{W}_\sigma \boldsymbol{\sigma}\|_q \quad (12)$$

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

$$\|\mathbf{W} \boldsymbol{\sigma}\|_r = \left[\sum_{k=1}^M w_k^q p_k^r \right]^{1/r} \quad (13)$$

$$\|\mathbf{W}_\sigma \boldsymbol{\sigma}\|_q = \left[\sum_{k=1}^N w_{M+k}^q \sigma_k^q \right]^{1/q} \quad (14)$$

$$\mathbf{W} = \text{diag} \left(\begin{bmatrix} w_1 & \dots & w_M \end{bmatrix} \right) \quad (15)$$

$$\mathbf{W}_\sigma = \text{diag} \left(\begin{bmatrix} w_{M+1} & \dots & w_{M+N} \end{bmatrix} \right) \quad (16)$$

Thus, setting $q = 1$ and $r = 1$ yields the original SPICE formulation. Note that, more general regularization functions could also be used, but in this paper, we confine our attention to the r - and q -norm cases, which we hereafter term the $\{r, q\}$ -SPICE formulation.

3. LINKING $\{R, Q\}$ -SPICE TO THE LASSO

To demonstrate the effects of introducing the r - and q -norms to SPICE, we follow the derivation in [17, 18], and proceed to examine the connection between $\{r, q\}$ -SPICE and a penalized regression problem. In order to do so, we distinguish between two cases, namely the case when each σ_k are allowed to have distinct values, and the case when all σ_k are equal. First, we recall a lemma that will be helpful for the following derivation (see also [18]).

Lemma 1. *Let*

$$\tilde{\mathbf{P}} = \text{diag} \left(\begin{bmatrix} p_1 & \dots & p_M \end{bmatrix} \right) \quad (17)$$

and

$$\boldsymbol{\Sigma} = \text{diag} \left(\begin{bmatrix} \sigma_1 & \dots & \sigma_N \end{bmatrix} \right) \quad (18)$$

Then,

$$\begin{aligned} \mathbf{y}^* \mathbf{R}^{-1} \mathbf{y} = \underset{\mathbf{x}}{\text{minimize}} & (\mathbf{y} - \mathbf{B} \mathbf{x})^* \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{B} \mathbf{x}) \\ & + \sum_{k=1}^M |x_k|^2 / p_k \end{aligned} \quad (19)$$

with the minimum occurring at

$$\hat{\mathbf{x}} = \boldsymbol{\Sigma} \mathbf{B}^* \mathbf{R}^{-1} \mathbf{y} \quad (20)$$

□

3.1. Varying noise variance

Using Lemma 1, one may rewrite (12) as

$$\begin{aligned} \underset{\mathbf{x}, \mathbf{p}, \sigma}{\text{minimize}} \quad & \sum_{k=1}^N |y_k - \mathbf{b}_k^* \mathbf{x}|^2 / \sigma_k + \sum_{k=1}^M |x_k|^2 / p_k \\ & + \left(\sum_{k=1}^M w_k^r p_k^r \right)^{1/r} + \left(\sum_{k=1}^N w_{M+k}^q \sigma_k^q \right)^{1/q} \end{aligned} \quad (21)$$

Differentiating (21) with respect to (w.r.t.) p_j and setting it to zero yields

$$p_j = w_k^{-\frac{r}{r+1}} |x_k|^{\frac{2}{r+1}} \|\mathbf{W}^{1/2} \mathbf{x}\|^{\frac{r-1}{r+1}} \quad (22)$$

Doing the same for σ_k yields

$$\sigma_k = w_{M+k}^{-\frac{q}{q+1}} |r_k|^{\frac{2}{q+1}} \|\mathbf{W}_\sigma^{1/2} \mathbf{r}\|^{\frac{q-1}{q+1}} \quad (23)$$

Finally, inserting (22) and (23) into (21) yields

$$\underset{\mathbf{x}}{\text{minimize}} \quad \left\| \mathbf{W}_\sigma^{1/2} (\mathbf{y} - \mathbf{B}\mathbf{x}) \right\|_{\frac{2q}{q+1}} + \left\| \mathbf{W}^{1/2} \mathbf{x} \right\|_{\frac{2r}{r+1}} \quad (24)$$

From the resulting expression, it may be noted that by properly choosing the r - and q -norms, we may reach a large range of penalized regression problems. This implies that if we are able to solve the $\{r, q\}$ -SPICE minimization problem, we also solve the corresponding penalized regression problem. It might thus be in many cases preferable to solve (12) rather than trying to compute (24). The implications of this will be discussed further below.

Clearly, regardless of the choice of r and q , the corresponding problem in (12) will still be scale invariant. To see this, we follow [18] and scale each p_k and σ_k with a constant c , doing the same for the cost function in (12), thus defining

$$\begin{aligned} g(\mathbf{p}, \sigma) & \triangleq c \mathbf{y}^* (\mathbf{A}c\mathbf{P}\mathbf{A}^*)^{-1} \mathbf{y} \\ & + c \left[\sum_{k=1}^M w_k^q c^r p_k^r \right]^{1/r} + c \left[\sum_{k=M+1}^{N+M} w_k^q c^q p_k^q \right]^{1/q} \\ & = \mathbf{y}^* (\mathbf{A}\mathbf{P}\mathbf{A}^*)^{-1} \mathbf{y} + c^2 \left[\sum_{k=1}^M w_k^r p_k^r \right]^{1/r} + \\ & c^2 \left[\sum_{k=M+1}^{N+M} w_k^q p_k^q \right]^{1/q} \end{aligned} \quad (25)$$

Defining the cost function in (12) as $f(\mathbf{p}, \sigma)$, we may use Lemma 2 in [18] to conclude that if

$$\{\hat{\mathbf{p}}, \hat{\sigma}\} = \arg \min_{\mathbf{p}, \sigma} g(\mathbf{p}, \sigma) \quad (26)$$

and

$$\{\hat{\mathbf{p}}, \hat{\sigma}\} = \arg \min_{\bar{\mathbf{p}}, \bar{\sigma}} f(\bar{\mathbf{p}}, \bar{\sigma}) \quad (27)$$

then

$$\hat{\mathbf{p}} = c\bar{\mathbf{p}} \quad (28)$$

where $c > 0$, which is true in this case as well. Due to this scale invariance, we conclude that the $\{r, q\}$ -SPICE method is also hyper-parameter free, in the same sense as SPICE. Furthermore, it may be noted that when converting the p_k to x_k , any scaling will disappear.

3.2. Uniform noise variance

If, similar to [17, 18], one instead assumes that all the noise terms have equal variance, thus treating $\sigma_k = \sigma$, $\forall k$, one arrives at an interesting conclusion: with this assumption, it has been shown that the SPICE problem is connected to the square-root LASSO problem [17, 18], i.e.,

$$\underset{\mathbf{x}}{\text{minimize}} \quad \|\mathbf{y} - \mathbf{B}\mathbf{x}\|_2 + \mu \|\mathbf{W}^{1/2} \mathbf{x}\|_1 \quad (29)$$

where $\mu = N^{-1/2}$ for SPICE. Following the derivation in Section 3.1, together with the assumption that all the noise terms have equal variance, yields $\mu = N^{-1/2q}$ and

$$\underset{\mathbf{x}}{\text{minimize}} \quad \|\mathbf{y} - \mathbf{B}\mathbf{x}\|_2 + \mu \|\mathbf{W}^{1/2} \mathbf{x}\|_{\frac{2r}{r+1}} \quad (30)$$

Thus, the choice of q corresponds to the weight that governs the trade-off between the model fitting term and the regularization of the parameters, and the choice of r decides which norm will be used in the regularization of the parameters. For instance, using $r = 1$ means that increasing q corresponds to increasing the sparsity in the square-root LASSO. Again, it might be preferable to solve (30) using the $\{r, q\}$ -SPICE formulation, rather than solving (30) directly.

4. OFF-GRID SOLUTION

Many forms of estimation problems are commonly 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., [21, 22]). However, such a solution might cause the columns of the extended dictionary to be highly correlated, thereby decreasing the performance of the method. In [20] and [23], an off-grid solution to the original SPICE version was presented. In this section, we similarly provide one possible version of off-grid estimation for the proposed $\{r, q\}$ -SPICE method. In order to do so,

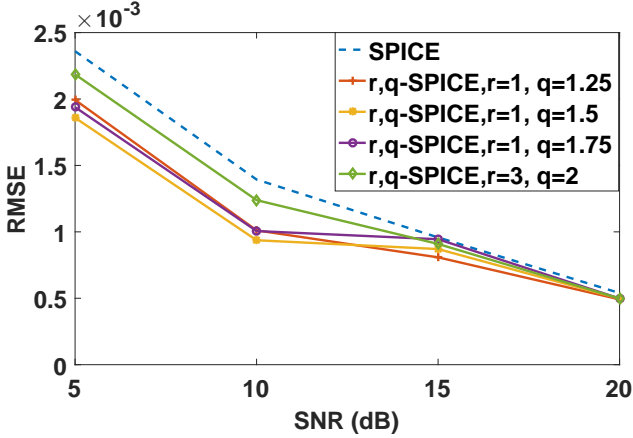


Fig. 1. The RMSE of the frequency estimates, as defined in (33), as a function of SNR for $\{r, q\}$ -SPICE and SPICE.

it may initially be noted that one may separate \mathbf{R} into two different matrices

$$\mathbf{R} = \mathbf{B}^* \text{diag}(\hat{\mathbf{p}}) \mathbf{B} + \text{diag}(\boldsymbol{\sigma}) \triangleq \mathbf{T}(\mathbf{u}) + \text{diag}(\boldsymbol{\sigma}) \quad (31)$$

where $\mathbf{T}(\mathbf{u})$ is a Toeplitz matrix with \mathbf{u} forming the first column of $\mathbf{T}(\mathbf{u})$. Thus, (12) may be expressed as

$$\begin{aligned} & \underset{\mathbf{u}, \boldsymbol{\sigma}, x}{\text{minimize}} \quad \|\mathbf{y}\|_2^2 x + \|\text{diag}(\mathbf{T}(\mathbf{u}))\|_r + \|\mathbf{W}_\sigma \boldsymbol{\sigma}\|_q \\ & \text{subject to} \quad \begin{bmatrix} x & \mathbf{y}^* \\ \mathbf{y} & \mathbf{T}(\mathbf{u}) + \text{diag}(\boldsymbol{\sigma}) \end{bmatrix} \geq 0 \\ & \quad \mathbf{T}(\mathbf{u}) \geq 0 \\ & \quad \mathbf{T}(\mathbf{u}) - \mathbf{T}(\mathbf{u})^* = 0 \\ & \quad \boldsymbol{\sigma} \geq 0 \end{aligned} \quad (32)$$

and under the additional constraint that $\mathbf{T}(\mathbf{u})$ is a Toeplitz matrix. The optimization problem in (32) is convex, and may be solved by using, e.g., a publicly available SDP (semidefinite programming) solver, such as the one presented in [24]. The final off-grid estimates may then be found using the celebrated Vandermonde decomposition in combination with, for instance, Prony's method (see [5, 25] for further details on such an approach)¹.

5. NUMERICAL EXAMPLES

Arguably, the most interesting case for $\{r, q\}$ -SPICE is when we restrict $q > 1$, thus enforcing a sparse solution. We will thus, in the numerical examples, focus on this situation, but will include some examples showing the performance when $r > 1$. We investigate two properties of the estimators, namely the resulting root-mean-squared error (RMSE) of the

¹An implementation of the algorithm will be provided online upon publication.

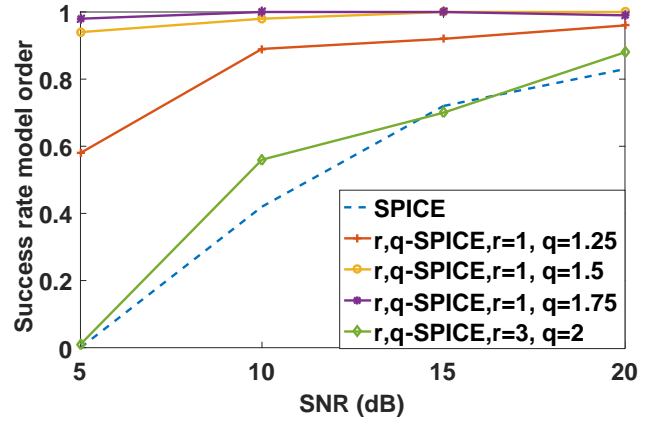


Fig. 2. The probability of finding the correct model order of the signal as a function of SNR for $\{r, q\}$ -SPICE and SPICE.

frequency estimates, defined as

$$\text{RMSE} \triangleq \sqrt{\frac{1}{P} \sum_{k=1}^P |\hat{\theta}_k - \theta_k|^2} \quad (33)$$

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 with circular white Gaussian noise, and where the noise terms were allowed to differ. The solution was obtained by solving (32) for all settings except for the original SPICE, where the estimates were obtained from solving the problem formulated in [23]. In Figure 1, the resulting RMSEs are shown for different values of q and r , as a function of the signal-to-noise-ratio (SNR). To make the figures readable, one respectively two outliers were removed for SPICE and for the $r = 3, q = 2$ case for 5 dB SNR-level. Furthermore, to remove the noise peaks that appear when using small values of q , all peaks smaller than 20 % of the largest found peak was 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 2. 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 are starting 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.

6. REFERENCES

- [1] M. Christensen and A. Jakobsson, *Multi-Pitch Estimation*, Morgan & Claypool, San Rafael, Calif., 2009.
- [2] 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.
- [3] 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.
- [4] 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.
- [5] P. Stoica and R. Moses, *Spectral Analysis of Signals*, Prentice Hall, Upper Saddle River, N.J., 2005.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] I. F. Gorodnitsky and B. D. Rao, “Sparse Signal Reconstruction from Limited Data Using FOCUSS: A Reweighted Minimum Norm Algorithm,” *IEEE Trans. Signal Process.*, vol. 45, no. 3, pp. 600–616, March 1997.
- [10] 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.
- [11] T. Hastie, R. Tibshirani, and M. Wainwright, *Statistical Learning with Sparsity: The Lasso and Generalizations*, Chapman and Hall/CRC, 2015.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] 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.
- [19] D. Zachariah and P. Stoica, “Online Hyperparameter-Free Sparse Estimation Method,” *IEEE Trans. Signal Process.*, vol. 63, no. 13, pp. 3348–3359, July 2015.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] 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.
- [24] Inc. CVX Research, “CVX: Matlab Software for Disciplined Convex Programming, version 2.0 beta,” <http://cvxr.com/cvx>, Sept. 2012.
- [25] 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.