# A Note on the SPICE Method

Cristian R. Rojas, *Member, IEEE*, Dimitrios Katselis, *Member, IEEE*, and Håkan Hjalmarsson, *Fellow, IEEE*

***Abstract*—In this article, we analyze the SPICE method developed in [1], and establish its connections with other standard sparse estimation methods such as the Lasso and the LAD-Lasso. This result positions SPICE as a computationally efficient technique for the calculation of Lasso-type estimators. Conversely, this connection is very useful for establishing the asymptotic properties of SPICE under several problem scenarios and for suggesting suitable modifications in cases where the naive version of SPICE would not work.**

***Index Terms*—Irregular sampling, sparse parameter estimation, spectral analysis.**

## I. Introduction

SPECTRAL line estimation, or the problem of estimating the amplitudes and frequencies of a signal composed of a sum of sinusoids contaminated by Gaussian white noise, is a ubiquitous and well studied area in the field of signal processing [2]. Many classes of methods have been devised to solve this problem under several different scenarios like, e.g., uniformly/non-uniformly spaced samples, *a priori* known/unknown number of sinusoids, homoscedastic/heteroscedastic (constant/varying variance) samples, parametric/non-parametric model-based, and so on [2]–[4].

Recently, SPICE (SemiParametric/SParse Iterative Covariance-based Estimator), a new technique for spectral line estimation inspired by ideas from sparse estimation, has been proposed in [1]. This method is capable of handling irregularly sampled data. Similarly, a version of SPICE has also been developed for array signal processing [5], a mathematically almost equivalent problem ([2], Chapter 6).

In this paper, we establish the connection between SPICE and standard sparse estimation methods such as the Lasso [6] and the LAD-Lasso [7]. This connection, based on the so-called Elfving theorem from optimal experiment design [8], puts the SPICE method into perspective, allowing us to examine the asymptotic properties of SPICE under several scenarios by simply applying the existing theory for the Lasso and its variants (see, e.g., the recent book [9]). Conversely, the relationship between SPICE and Lasso-type estimators suggests that SPICE may be used as a (new) numerically efficient technique for computing Lasso estimates.

The manuscript is organized as follows. Section II describes the spectral line estimation problem and the SPICE method. Section III establishes the relation between SPICE and Lasso-type sparse estimation methods. In Section IV a simulation example illustrating the equivalence between SPICE and a version of Lasso is presented. Finally, Section V concludes the paper.

*Notation:* Vectors and matrices are written in bold lowercase and uppercase fonts, respectively. $^T$ and $^H$ denote transposition and complex conjugate transposition, respectively. $\mathrm{Re}z$ and $\mathrm{Im}z$ stand for the real and imaginary parts of the complex number $z$, and $j$ is the square root of $-1$. $\mathbb{R}_0^+$ is the set of non-negative real numbers, and $\mathbb{C}$ is the complex plane. $\|\cdot\|_1$, $\|\cdot\|_2$, $\|\cdot\|_F$ and $|\cdot|$ correspond to the 1-norm, Euclidean norm, Frobenius norm and absolute value, respectively. $\mathrm{diag}\,(a_1,\ldots,a_n)$ is a diagonal matrix whose diagonal is given by $a_1,\ldots,a_n$. $\boldsymbol{I}$ is the identity matrix. $E\{\cdot\}$ denotes mathematical expectation.

## II. Problem Formulation and SPICE Method

Consider the following problem: Let $\boldsymbol{y} \in \mathbb{C}^{N\times 1}$ be given, satisfying the equation

$$\boldsymbol{y} = \sum_{k=1}^{K} \boldsymbol{a}_k s_k + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{\epsilon} \in \mathbb{C}^{N\times 1}$ is a complex Gaussian random vector of zero mean and covariance matrix $\mathrm{diag}\,(\sigma_1,\ldots,\sigma_N)$, and $\{\boldsymbol{a}_k\}_{k=1}^K \in \mathbb{C}^{N\times 1}$ are known complex vectors. $\{s_k\}_{k=1}^K \in \mathbb{C}$ are unknown complex quantities, of the form $s_k = |s_k|e^{j\phi_k}$, where the phases $\{\phi_k\}_{k=1}^K \in [0,2\pi)$ are independent random variables uniformly distributed in $[0,2\pi)$, and the magnitudes $\{|s_k|\}_{k=1}^K \in \mathbb{R}_0^+$ are deterministic parameters to be estimated. The spectral line estimation problem considers a particular case of (1), where the $\boldsymbol{a}_k$'s are vectors of imaginary exponentials of the form $e^{j\omega t}$ [2].

In order to estimate the magnitudes $|s_k|$, let

$$\boldsymbol{R} := E\{\boldsymbol{y}\boldsymbol{y}^H\} = \boldsymbol{A}^H\boldsymbol{P}\boldsymbol{A},$$

where

$$\begin{aligned}\boldsymbol{A}^H &:= [\boldsymbol{a}_1 \cdots \boldsymbol{a}_K \boldsymbol{I}]\\ &=: [\boldsymbol{a}_1 \cdots \boldsymbol{a}_{K+N}]\\ \boldsymbol{P} &:= \mathrm{diag}\,(|s_1|^2,\ldots,|s_K|^2,\sigma_1,\ldots,\sigma_N)\\ &=: \mathrm{diag}\,(p_1,\ldots,p_{K+N}).\end{aligned}$$

The SPICE estimate [1] of the $|s_k|$'s is an iterative procedure of the form:

$$\begin{aligned}\boldsymbol{R}(i) &= \boldsymbol{A}^H \mathrm{diag}\,(p_1(i),\ldots,p_{K+N}(i))\boldsymbol{A}\\ p_k(i+1) &= p_k(i)\frac{\left|\boldsymbol{a}_k^H\boldsymbol{R}^{-1}(i)\boldsymbol{y}\right|}{w_k^{1/2}\rho(i)}, \qquad w_k := \frac{\|\boldsymbol{a}_k\|_2^2}{\|\boldsymbol{y}\|_2^2},\\ \rho(i) &= \sum_{l=1}^{K+N} w_l^{1/2} p_l(i)\left|\boldsymbol{a}_l^H\boldsymbol{R}^{-1}(i)\boldsymbol{y}\right|,\end{aligned} \tag{2}$$

Manuscript received January 27, 2013; revised June 28, 2013; accepted June 28, 2013. Date of publication July 04, 2013; date of current version August 19, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Antonio Napolitano. This work was supported by the European Research Council under the advanced Grant LEARN, Contract 267381, and by the Swedish Research Council under Contracts 621-2009-4017, 621-2010-5819, and 621-2011-5890.

The authors are with the Automatic Control Laboratory and the ACCESS Linnaeus Center, School of Electrical Engineering, KTH Royal Institute of Technology, S-100 44 Stockholm, Sweden (e-mail: cristian.rojas@ee.kth.se; dimitrios.katselis@ee.kth.se; hakan.hjalmarsson@ee.kth.se)

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2013.2272291

where $i$ is the iteration number, and $p_k(i)$ is the estimate of $p_k$ at iteration $i$. This method is initialized by any initial estimate of the $p_k$'s, and its estimate $\boldsymbol{R}(i)$ converges to the matrix $\boldsymbol{R}$ minimizing

$$f(\boldsymbol{R}) := \|\boldsymbol{R}^{-1/2}(\boldsymbol{y}\boldsymbol{y}^H - \boldsymbol{R})\|_F^2. \tag{3}$$

The $p_k$'s that give $\boldsymbol{R}$ correspond to the limits $\lim_{i\to\infty} p_k(i)$.

*Remark 1:* The presence of the inverse of $\boldsymbol{R}(i)$ in the SPICE method may in principle lead to complications if such a matrix becomes singular. However, if the $p_k(0)$'s are chosen to be strictly positive, then $\boldsymbol{R}(i+1)$ is generically non-singular (since $\boldsymbol{a_k}$ is generically in the column range of $\boldsymbol{R}(i)$, and $\boldsymbol{y}$ is a Gaussian random vector which lies in the null space of $\boldsymbol{R}(i)$ with probability 0). Because of this, here and in the sequel we will implicitly assume for the derivations that $\boldsymbol{R}$ is non-singular.

*Remark 2:* In [5], SPICE was defined based on a slightly different $f(\boldsymbol{R})$. We will not consider that version of SPICE, because such a version can only be defined in a multi-snapshot case. However, similar steps as the ones described in the following sections can be applied to the method in [5] to arrive at an equivalent Lasso-type formulation.

## III. Analysis of SPICE

The first version of SPICE in [1] allows the variances $\sigma_k$ to be different, while a variant of the method imposes the constraint that $\sigma_1 = \cdots \sigma_N =: \sigma$ ([1], Section III.D). We will treat these cases separately, starting with the case where the variances can be different.

### A. Different Variances

As shown in [1], the function $f$ in (3) can be written as

$$\begin{aligned} f(\boldsymbol{R}) &= \mathrm{tr}\,\{[\boldsymbol{R}^{-1/2}(\boldsymbol{y}\boldsymbol{y}^H - \boldsymbol{R})]^H \boldsymbol{R}^{-1/2}(\boldsymbol{y}\boldsymbol{y}^H - \boldsymbol{R})\} \\ &= \|\boldsymbol{y}\|_2^2 \boldsymbol{y}^H \boldsymbol{R}^{-1}\boldsymbol{y} - 2\|\boldsymbol{y}\|_2^2 + \mathrm{tr}\,\boldsymbol{R}, \end{aligned}$$

hence minimizing $f(\boldsymbol{R})$ is equivalent to minimizing

$$\begin{aligned} g(\boldsymbol{R}) &:= \boldsymbol{y}^H \boldsymbol{R}^{-1}\boldsymbol{y} + \frac{1}{\|\boldsymbol{y}\|_2^2}\mathrm{tr}\,\boldsymbol{R} \\ &= \boldsymbol{y}^H \boldsymbol{R}^{-1}\boldsymbol{y} + \sum_{k=1}^{K+N} \frac{\|\boldsymbol{a}_k\|_2^2}{\|\boldsymbol{y}\|_2^2} p_k \\ &= \boldsymbol{y}^H \boldsymbol{R}^{-1}\boldsymbol{y} + \sum_{k=1}^{K+N} w_k p_k, \end{aligned} \tag{4}$$

subject to $p_k \geq 0$, where

$$w_k := \frac{\|\boldsymbol{a}_k\|_2^2}{\|\boldsymbol{y}\|_2^2}.$$

To further simplify the problem, in (Appendix B of [5]) it is argued that the minimization of $g(\boldsymbol{R})$ is equivalent (up to a scaling of the $p_k$'s) to solving

$$\begin{aligned} \min_{p_1,\ldots,p_{K+N}\geq 0} \quad & \boldsymbol{y}^H \boldsymbol{R}^{-1}\boldsymbol{y} \\ \text{s.t.} \quad & \sum_{k=1}^{K+N} w_k p_k = 1 \\ & \sum_{k=1}^{K+N} \boldsymbol{a}_k \boldsymbol{a}_k^H p_k = \boldsymbol{R}. \end{aligned} \tag{5}$$

Equation (5) will be our starting point for the analysis of SPICE. A slight simplification can be achieved by defining $\tilde{p}_k := w_k p_k$ and $\tilde{\boldsymbol{a}}_k := w_k^{-1/2}\boldsymbol{a}_k$ for all $k = 1, \ldots, K+N$. This gives the re-parameterized problem

$$\begin{aligned} \min_{\tilde{p}_1,\ldots,\tilde{p}_{K+N}\geq 0} \quad & \boldsymbol{y}^H \boldsymbol{R}^{-1}\boldsymbol{y} \\ \text{s.t.} \quad & \sum_{k=1}^{K+N} \tilde{p}_k = 1 \\ & \sum_{k=1}^{K+N} \tilde{\boldsymbol{a}}_k \tilde{\boldsymbol{a}}_k^H \tilde{p}_k = \boldsymbol{R}. \end{aligned} \tag{6}$$

The strategy now is to consider a derivation similar to Elfving's theorem, from optimal experiment design [8], to obtain an optimization problem equivalent to (6). First notice that

$$\begin{aligned} \left(\boldsymbol{y}^H \boldsymbol{R}^{-1}\boldsymbol{y}\right)\big|_{\boldsymbol{R}=\sum_{k=1}^{\prime K+N} \tilde{\boldsymbol{a}}_k \tilde{\boldsymbol{a}}_k^H \tilde{p}_k} \\ = \min_{c_1,\ldots,c_{K+N}} \sum_{k=1}^{K+N,\prime} \frac{|c_k|^2}{\tilde{p}_k} \quad \text{s.t.} \quad \tilde{\boldsymbol{A}}^H \boldsymbol{c} = \boldsymbol{y}, \end{aligned} \tag{7}$$

where $\tilde{\boldsymbol{A}}^H := [\tilde{\boldsymbol{a}}_1 \cdots \tilde{\boldsymbol{a}}_{K+N}]$ and $\boldsymbol{c} := [c_1 \cdots, c_{K+N}]^T$. Here the $'$ symbol in the summation sign indicates that the values of $k$ for which $\tilde{p}_k = 0$ should be omitted from the sum. The proof of (7) is given in the Appendix.

The combination of (6) and (7) gives a minimization problem in $\{\tilde{p}_k\}$ and $\{c_k\}$, i.e.,

$$\begin{aligned} \min_{\substack{\tilde{p}_1,\ldots,\tilde{p}_{K+N}\geq 0, \\ c_1,\ldots,c_{K+N}}} \quad & \sum_{k=1}^{K+N,\prime} \frac{|c_k|^2}{\tilde{p}_k} \\ \text{s.t.} \quad & \sum_{k=1}^{K+N} \tilde{p}_k = 1 \\ & \tilde{\boldsymbol{A}}^H \boldsymbol{c} = \boldsymbol{y}, \end{aligned} \tag{8}$$

where the order of the minimizing variables can be exchanged. Now, when the $c_k$'s are kept fixed, the minimization of the cost

in (8) with respect to $\{\tilde{p}_k\}$ can be done explicitly. To see this, notice that by the Cauchy-Schwarz inequality we have

$$\begin{aligned}\sum_{k=1}^{N+k}\frac{|c_k|^2}{\tilde{p}_k} &= \left(\sum_{k=1}^{N+k}\frac{|c_k|^2}{\tilde{p}_k}\right)\left(\sum_{k=1}^{K+N}\tilde{p}_k\right)\\ &\geq \left(\sum_{k=1}^{N+k}\frac{|c_k|}{\sqrt{\tilde{p}_k}}\sqrt{\tilde{p}_k}\right)^2\\ &= \left(\sum_{k=1}^{N+k}|c_k|\right)^2,\end{aligned}$$

where the lower bound is attained if and only if there is an $\alpha \in \mathbb{C}$ such that

$$\frac{|c_k|^2}{\tilde{p}_k} = \alpha\tilde{p}_k, \quad k = 1,\ldots,K+N,$$

or

$$\tilde{p}_k = \frac{|c_k|}{\sqrt{\alpha}}, \quad k = 1,\ldots,K+N.$$

The proportionality constant $\alpha$ can be determined from the condition $\sum_{k=1}^{K+N}\tilde{p}_k = 1$, giving

$$\tilde{p}_k = \frac{|c_k|}{\sum_{i=1}^{K+N}|c_i|}, \quad k = 1,\ldots,K+N. \tag{9}$$

Putting this expression in (8) gives the reduced problem

$$\begin{aligned}\min_{c_1,\ldots,c_{K+N}} &\left(\sum_{k=1}^{K+N}|c_k|\right)^2\\ \text{s.t.}\quad &\tilde{\boldsymbol{A}}^H\boldsymbol{c} = \boldsymbol{y},\end{aligned}$$

or, equivalently,

$$\begin{aligned}\min_{c_1,\ldots,c_{K+N}} &\sum_{k=1}^{K+N}|c_k|\\ \text{s.t.}\quad &\tilde{\boldsymbol{A}}^H\boldsymbol{c} = \boldsymbol{y}.\end{aligned} \tag{10}$$

This is a complex-valued $l_1$-optimization problem, hence it can be expected to give a sparse solution in $\{c_k\}$. This, in turn, gives a sparse solution in $\{\tilde{p}_k\}$ through (9), and thus in

$$p_k = \frac{\tilde{p}_k}{w_k} = \frac{|c_k|\|\boldsymbol{y}\|_2^2}{\|\boldsymbol{a}_k\|_2^2\sum_{i=1}^{K+N}|c_i|}, \quad k = 1,\ldots,K+N.$$

To explore the behavior of SPICE in more detail, we can notice, by denoting first $K$ components of the $k$-th row of $\tilde{\boldsymbol{A}}^H$ as $\boldsymbol{\varphi}_k^H$, i.e., $\boldsymbol{\varphi}_k^H := [(\tilde{\boldsymbol{a}}_1)_k\cdots,(\tilde{\boldsymbol{a}}_K)_k]$, and observing that the constraints in (10) read $c_{k+j} = y_j - \boldsymbol{\varphi}_j^H\tilde{\boldsymbol{c}}$ for $j = 1,\ldots,N$, that (10) is equivalent to

$$\min_{c_1,\ldots,c_K} \quad \sum_{k=1}^{N}\left|y_k - \boldsymbol{\varphi}_k^H\tilde{\boldsymbol{c}}\right| + \sum_{k=1}^{K}|c_k|,$$

where $\tilde{\boldsymbol{c}} := [c_1\cdots c_K]^T$, or more compactly

$$\min_{\tilde{\boldsymbol{c}}} \quad \|\boldsymbol{y} - \boldsymbol{\Phi}\tilde{\boldsymbol{c}}\|_1 + \|\tilde{\boldsymbol{c}}\|_1, \tag{11}$$

where $\boldsymbol{\Phi}^H := [\boldsymbol{\varphi}_1\cdots,\boldsymbol{\varphi}_N]$, i.e., $\boldsymbol{\Phi}$ corresponds to the first $K$ columns of $\tilde{\boldsymbol{A}}^H$. Equation (11) is essentially a simplified (complex-valued) version of the LAD-Lasso [7] or the RLAD [10], where $\tilde{\boldsymbol{c}}$ takes the role of a parameter vector, and the regressors have been scaled by $w_k^{-1/2} = \|\boldsymbol{y}\|_2/\|\boldsymbol{a}_k\|_2$, so that their Euclidean norms are equal to $\|\boldsymbol{y}\|_2$. The fact that the cost function in (11) considers the $\ell_1$ norm of the residuals $(\boldsymbol{y}-\boldsymbol{\Phi}\tilde{\boldsymbol{c}})$ instead of their $\ell_2$ norm suggests that SPICE might be a robust estimator against outliers or errors with heavy-tailed distributions (since, heuristically speaking, it does not penalize large deviations of the residuals from zero, due mainly to outliers, as much as the $\ell_2$ norm); in fact, this is the reason why some authors have proposed the use of the LAD-Lasso instead of the standard Lasso in the presence of outliers [7].

We can summarize these results in the following theorem:

*Theorem 1:* The limit value of the SPICE iterations (allowing for different $\sigma_k$), which corresponds to the minimizer of (3), is also given by the minimizer of (11), by performing the following change of variables:

$$p_k = \frac{\|\boldsymbol{y}\|_2^2|c_k|}{\|\boldsymbol{a}_k\|_2^2\left\{\sum_{i=1}^{K}|c_i| + \sum_{k=1}^{N}\left|y_k - \boldsymbol{\varphi}_k^H\tilde{\boldsymbol{c}}\right|\right\}}, \quad k = 1,\ldots,K+N,$$

where $c_{k+j} = y_j - \boldsymbol{\varphi}_j^H\tilde{\boldsymbol{c}}$ for $j = 1,\ldots,N$.

*Remark 3:* In [11], a slightly different version of SPICE has derived, based on (4) rather than on (5). By performing essentially the same steps as in the derivation of Theorem 1, we obtain the following corollary, which shows that the SPICE method of [11] is equivalent to the same LAD-Lasso problem, but where the relation between the $p_k$'s and $c_k$'s is simpler.

*Corollary 1:* The limit value of the SPICE iterations described in [11] (allowing for different $\sigma_k$) is also given by the minimizer of (11), by performing the following change of variables (where $c_{k+j} = y_j - \boldsymbol{\varphi}_j^H\tilde{\boldsymbol{c}}$ for $j = 1,\ldots,N$):

$$p_k = \frac{\|\boldsymbol{y}\|_2^2}{\|\boldsymbol{a}_k\|_2^2}|c_k|, \qquad k = 1,\ldots,K+N.$$

### B. Equal Variances

Now we will analyze the variant of SPICE where the variances are constrained to be equal. The development in this case is exactly as in Section III-A until (8). At this point, the constraint $\sigma_1 = \cdots = \sigma_N =: \sigma$ implies that $\tilde{p}_{K+1} = \cdots = \tilde{p}_{K+N}$, which allows us to simplify (8) as

$$\begin{aligned}\min_{\substack{p'_1,\ldots,p'_{K+1}\geq 0,\\ c_1,\ldots,c_{K+N}}} &\sum_{k=1}^{K}\frac{|c_k|^2}{p'_k} + \frac{N}{p'_{K+1}}\sum_{k=K+1}^{K+N}|c_k|^2\\ \text{s.t.}\quad &\sum_{k=1}^{K+1}p'_k = 1\\ &\tilde{\boldsymbol{A}}^H\boldsymbol{c} = \boldsymbol{y},\end{aligned}$$

where $p'_k = \tilde{p}_k$ for $k = 1, \ldots, K$, $p'_{K+1} = N\tilde{p}_{K+1}$, and $\boldsymbol{c} := [c_1 \cdots c_{K+N}]^T$. Now, the Cauchy-Schwarz argument used in Section III.A reveals that

$$p'_k = \begin{cases} \dfrac{|c_k|}{\sqrt{\alpha}}, & k = 1, \ldots, K, \\ \sqrt{\dfrac{N}{\alpha} \displaystyle\sum_{k=K+1}^{K+N} |c_k|^2}, & k = K+1, \end{cases}$$

and from the condition $\sum_{k=1}^{K+N} \tilde{p}_k = 1$ we obtain

$$\alpha = \left( \sum_{k=1}^{K} |c_k| + \sqrt{N \sum_{k=K+1}^{K+N} |c_k|^2} \right)^2 .$$

The constants $c_k$, on the other hand, must be the solution of

$$\min_{c_1, \ldots, c_{K+N}} \sum_{k=1}^{K} |c_k| + \sqrt{N \sum_{k=K+1}^{K+N} |c_k|^2} \quad \text{s.t.} \quad \tilde{\boldsymbol{A}}^H \boldsymbol{c} = \boldsymbol{y}. \tag{12}$$

Just as in Section III.A, (12) can be rewritten as

$$\min_{\tilde{\boldsymbol{c}}} \sqrt{N \sum_{k=1}^{N} \left| y_k - \boldsymbol{\varphi}_k^H \tilde{\boldsymbol{c}} \right|^2} + \|\tilde{\boldsymbol{c}}\|_1,$$

where $\tilde{\boldsymbol{c}} := [c_1 \cdots c_K]^T$, or

$$\min_{\tilde{\boldsymbol{c}}} \quad \sqrt{N} \|\boldsymbol{y} - \boldsymbol{\Phi}\tilde{\boldsymbol{c}}\|_2 + \|\tilde{\boldsymbol{c}}\|_1. \tag{13}$$

Equation (13) is essentially a simplified (complex-valued) version of the standard Lasso [6], where $\tilde{\boldsymbol{c}}$ takes the role of a parameter vector, and the Euclidean norms of the regressors have been equalized. We summarize these results as a theorem:

*Theorem 2:* The limit value of the SPICE iterations (imposing the constraint that $\sigma_1 = \cdots = \sigma_N$), which corresponds to the minimizer of (3), is also given by the minimizer of (13), by performing the following change of variables:

$$p_k = \frac{\|\boldsymbol{y}\|_2^2 |c_k|}{\|\boldsymbol{a}_k\|_2^2 (\|\tilde{\boldsymbol{c}}\|_1 + \sqrt{N}\|\boldsymbol{y} - \boldsymbol{\Phi}\tilde{\boldsymbol{c}}\|_2)}, \quad k = 1, \ldots, K$$

$$p_{K+1} = \frac{\|\boldsymbol{y}\|_2^2 \|\boldsymbol{y} - \boldsymbol{\Phi}\tilde{\boldsymbol{c}}\|_2}{\sqrt{N}(\|\tilde{\boldsymbol{c}}\|_1 + \sqrt{N}\|\boldsymbol{y} - \boldsymbol{\Phi}\tilde{\boldsymbol{c}}\|_2)}.$$

The following remarks are appropriate:

*Remark 4:* As for the case of different variances, it is possible to follow the same steps leading to Theorem 2 to derive a Lasso equivalent for the version of SPICE derived in [11], based on (4) rather than on (5). The result is stated in the following corollary, which shows again that the SPICE method of [11] is equivalent to the same Lasso problem as in Theorem 2, but where the relation between the $p_k$'s and $c_k$'s is simpler.

*Corollary 2:* The limit value of the SPICE iterations described in [11] (imposing the constraint that $\sigma_1 = \cdots = \sigma_N$) is also given by the minimizer of (13), by performing the following change of variables:

$$p_k = \frac{\|\boldsymbol{y}\|_2^2}{\|\boldsymbol{a}_k\|_2^2} |c_k|, \quad k = 1, \ldots, K$$

$$p_{K+1} = \frac{\|\boldsymbol{y}\|_2^2}{\sqrt{N}} \|\boldsymbol{y} - \boldsymbol{\Phi}\tilde{\boldsymbol{c}}\|_2.$$

*Remark 5:* The results stated in Theorems 1 and 2 are quite surprising, because they reveal that different assumptions on the noise variance produce versions of SPICE which are equivalent to two quite different but standard sparse estimators, namely the LAD-Lasso and the Lasso.

*Remark 6:* Even though the equivalent Lasso formulations are not given in the same variables as the SPICE method, the required variables transformations (between the $c_k$'s and the $p_k$'s) are simple scalings. This means that the sparsity properties of SPICE are essentially the same as the ones for the equivalent Lasso estimators.

*Remark 7:* The optimization problem given by (13) is not written as a standard Lasso problem, since the first term is a not a *squared* $\ell_2$-norm, but rather as a *square-root Lasso* [12]. These two formulations, however, are equivalent. To see this, notice that $\min_\theta \|Y - \Phi\theta\|_2^2 + \lambda\|\theta\|_1$ is the Lagrangian form of an optimization problem of the form $\min_\theta \|Y - \Phi\theta\|_2^2$ s.t. $\|\theta\|_1 \leq \varepsilon$; this problem is equivalent to $\min_\theta \|Y - \Phi\theta\|_2$ s.t. $\|\theta\|_1 \leq \varepsilon$, whose Lagrangian formulation is $\min_\theta \|Y - \Phi\theta\|_2 + \lambda'\|\theta\|_1$. This means that there is a (possibly data-dependent) bijection $\lambda \mapsto \lambda'$ for which Lasso and the square-root Lasso give the same estimate.

*Remark 8:* The relations between the $c_k$'s and the $p_k$'s given by Theorems 1 and 2 have a nontrivial structure, which comes from the fact that SPICE considers the (unknown) noise variances as parameters to be estimated, and puts them in the same footing as the amplitudes of the spectral lines. This relation is simpler when the version of SPICE from [11] is considered instead, as shown in Corollaries 1 and 2.

*Remark 9:* The cost function $g(\boldsymbol{R})$ minimized by SPICE in (4) can be interpreted as follows: The first term of $g(\boldsymbol{R})$, $\boldsymbol{y}^H \boldsymbol{R}^{-1} \boldsymbol{y}$, is a model fit measure, while the second term, $\|\boldsymbol{y}\|_2^{-2} \mathrm{tr}\, \boldsymbol{R}$, can be interpreted as a trace heuristic or nuclear norm regularization (since $\boldsymbol{R} = \boldsymbol{R}^H \geq 0$, so the trace and nuclear norm coincide) [13]. This regularization term is known to encourage low rank matrices $\boldsymbol{R}$, which, due to its structure, $\boldsymbol{R} = \boldsymbol{A}^H \boldsymbol{P} \boldsymbol{A}$, enforces the vector $[p_1, \ldots, p_{K+N}]^T$ to be sparse. This interpretation thus provides an alternative heuristic justification for the sparsity-inducing behavior of SPICE.

*Remark 10:* Theorems 1 and 2 have been presented for the complex-valued versions of SPICE. However, the derivations in this section apply almost unaltered to real valued problems. This means that Theorems 1 and 2 establish Lasso-type equivalences for the real-valued versions of SPICE as well. Notice, however, that the complex Lasso versions can be seen as real-valued Group Lasso estimators, as explained next.

*Remark 11:* The complex-valued nature of SPICE is inherited by its Lasso equivalents. Thus, for example problem (13) does not behave as the standard (real-valued) Lasso, but as the (real-valued) Group Lasso [14]. To see this, let us define

$$\boldsymbol{y}_R := \begin{bmatrix} \mathrm{Re}\, \boldsymbol{y} \\ \mathrm{Im}\, \boldsymbol{y} \end{bmatrix}, \qquad \tilde{\boldsymbol{c}}_R := \begin{bmatrix} \mathrm{Re}\, \tilde{\boldsymbol{c}} \\ \mathrm{Im}\, \tilde{\boldsymbol{c}} \end{bmatrix}$$

$$\boldsymbol{\Phi}_R := \begin{bmatrix} \mathrm{Re}\, \boldsymbol{\Phi} & -\mathrm{Im}\, \boldsymbol{\Phi} \\ \mathrm{Im}\, \boldsymbol{\Phi} & \mathrm{Re}\, \boldsymbol{\Phi} \end{bmatrix}$$

Based on this notation, (13) can be written as

$$\min_{\tilde{\boldsymbol{c}}_R} \quad \sqrt{N} \|\boldsymbol{y}_R - \boldsymbol{\Phi}_R \tilde{\boldsymbol{c}}_R\|_2 + \sum_{k=1}^{K} \left\| \begin{bmatrix} (\tilde{\boldsymbol{c}}_R)_k \\ (\tilde{\boldsymbol{c}}_R)_{k+K} \end{bmatrix} \right\|_2 . \tag{14}$$

The second term in (14) is a sum of Euclidean norms, which promotes group sparsity, i.e., it tries to enforce that both the real and imaginary parts of individual entries of $\tilde{c}$ become zero simultaneously. Similarly, (11) corresponds to a grouped version of the LAD-Lasso.

*Remark 12:* It is well known that the real-valued LAD-Lasso can be written as a linear program (LP), which can be solved in a very efficient manner using existing techniques. For the standard Lasso, there are also very efficient computational methods. However, as Theorems 1 and 2 show, the original (complex-valued) SPICE method is actually equivalent to a group LAD-Lasso estimator (or to a standard group Lasso estimator), which cannot be formulated as an LP, but as a second-order cone program (SOCP). Many of the algorithms developed for the standard Lasso or the LAD-Lasso, such as the homotopy method [15] or LARS [16], cannot be extended to the group (LAD) Lasso, since its solution path is not piecewise affine, even though accelerated proximal methods [17] have been successfully applied to this class of estimators. This means that SPICE may be a potentially attractive technique in this case, to be compared with proximal methods (some of which depend on tuning parameters, to be specified by the user, while SPICE does not require user intervention).

*Remark 13:* It is well known that the (LAD-)Lasso is a biased estimator [9], because while the criterion $\|\boldsymbol{y} - \boldsymbol{\Phi}\tilde{\boldsymbol{c}}\|_{1,2}$ gives unbiased estimates, the addition of an $\ell_1$-norm regularization term pushes the estimates towards zero, in a way that no simple re-scaling of the parameters can correct. This means that SPICE will give in general biased estimates. However, this is not a relevant issue with either (LAD-)Lasso or SPICE, since, in order to correct the presence of bias, it is standard practice with the (LAD-)Lasso to re-estimate the non-zero components using least squares, and this idea can be applied to the SPICE estimates as well. The question of whether SPICE can detect the correct frequencies of the measured multisine is equivalent to the study of the support recovery properties of (LAD-)Lasso [9].

*Remark 14:* The equivalence between SPICE and Lasso allows the use of the well-developed theory for the Lasso to study the asymptotic behavior of SPICE [9], [18]. Take, for example, the spectral line estimation problem described in the next Section IV, where $K = O(N^\alpha)$ for some $\alpha \geq 1$, the samples are taken at uniformly distributed time instants in a fixed interval, and the components of $\boldsymbol{\epsilon}$ have unit variance. Then, by suitably modifying an argument in ([19] Section 2), it can be shown that as the number of samples $N$ tends to $\infty$, if the true amplitudes and number of frequencies are kept constant (but not necessarily the location of the frequencies), then SPICE (assuming equal variances) may not enjoy *persistency*, or $\ell_2$ prediction consistency, i.e., $N^{-1}\|\sum_{k=1}^{K} \boldsymbol{a}_k(s_k - s_k^o)\|_2^2 \to 0$. To achieve persistency, the second term in (13) should be amplified by a positive number $\lambda_N$ such that $\lambda_N/N \to 0$ and $\lambda_N/\ln(N) \to \infty$ (e.g., $\lambda_N = \sqrt{N}$); see [19] for further details. This problem can be solved by suitably modifying SPICE, but this aspect will be properly addressed in a future publication.

*Remark 15:* Recently, a re-weighted version of SPICE, called LIKES, has been proposed in [11]. We will not address here the relation between LIKES and standard sparse estimators (such as Sparse Bayesian Learning (SBL) and Automatic Relevance Determination (ARD) [20]), because this has partly been discussed in [11], and the equivalence to Lasso-type estimators can be formally studied along the lines of [20].

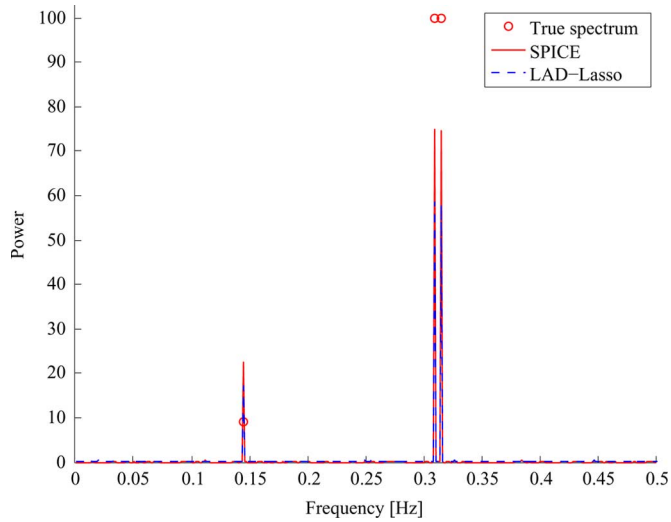


Fig. 1. Spectrum obtained by SPICE and LAD-Lasso.

## IV. Simulation Example

In this section, a numerical example, based on ([1], Section IV), is used to illustrate the equivalence between SPICE and the LAD-Lasso, formally established in Theorem 1.

Let $\boldsymbol{y}_k = y(t_k)$, $k = 1, \ldots, N$, be the $k$-th sample, where the $t_k$'s are irregular time samples, drawn independently from a uniform distribution on $[0, 200]$. The basis functions considered here are of the form

$$\boldsymbol{a}_k = [e^{j\omega_k t_1} \cdots e^{j\omega_k t_N}]^T,$$

where $\omega_k := 2\pi k/1000$. Following [1], we take $N = 100$, and $\boldsymbol{y}$ to be given by (1) with $K = 3$, $s_{145} = 3e^{j\phi_1}$, $s_{310} = 10e^{j\phi_2}$ and $s_{315} = 10e^{j\phi_3}$, and $s_k = 0$ otherwise. The phases $\phi_1$, $\phi_2$ and $\phi_3$ are independent random variables, uniformly distributed in $[0, 2\pi]$. The noise $\epsilon$ is assumed to have a covariance matrix $0.25\boldsymbol{I}$.

The results of applying 100 iterations of SPICE, (2), and its LAD-Lasso equivalent (11), solved using the CVX package [21], are presented in Fig. 1. As the figure shows, both estimators practically coincide, their differences being mainly due to numerical implementations. Notice also that these estimators correctly detect the location of the peaks of the true spectrum, even though the estimated amplitudes do not approach their true values; this observation is consistent with theoretical results regarding the bias of the Lasso and its variants [9]. On a PC with an 2.53 GHz Intel Core Duo CPU and 4 Gb RAM, 100 iterations of SPICE take 23.0 s, while the implementation of LAD-Lasso using CVX only takes 14.6 s. However, if $N$ is further increased to 1000, CVX is incapable of solving the LAD-Lasso problem, while SPICE can still provide a good (and numerically reliable) estimate.

## V. Conclusion

In this manuscript, the recently proposed SPICE method for sparse estimation has been studied, and its relation to Lasso-type estimators has been established. This connection may enable the use of existing theoretical results for the Lasso to predict the behavior of SPICE in diverse problem settings, and, at the same time, the application of the computationally efficient algorithm developed for SPICE to sparse estimation problems where the Lasso algorithms are currently impractical.

As a interesting future line of research, the relation between SPICE and the Group Lasso suggests that the former method could be modified to deal with general group sparsity problems (instead of only groups with two real variables). In addition, from this relation it is easy to modify SPICE in order to compensate for deficiencies already detected in standard Lasso estimators, such as lack of consistency in sparse support recovery, which can be fixed by adding re-weighting steps (see, e.g., [22]).

## Appendix
## Proof of (7)

In this Appendix we prove (7). Without loss of generality we can assume that the values of $k$ for which $\tilde{p}_k = 0$ have been removed from the sum. We start by rewriting (7) as

$$\boldsymbol{y}^H(\tilde{\boldsymbol{A}}^H\tilde{\boldsymbol{P}}\tilde{\boldsymbol{A}})^{-1}\boldsymbol{y} = \min_{\boldsymbol{c}} \boldsymbol{c}^H\tilde{\boldsymbol{P}}^{-1}\boldsymbol{c} \quad \text{s.t.} \quad \tilde{\boldsymbol{A}}^H\boldsymbol{c} = \boldsymbol{y}, \qquad (15)$$

where $\tilde{\boldsymbol{P}} := \text{diag}\,(\tilde{p}_1, \ldots, \tilde{p}_{K+N})$. We will proceed by establishing the minimum value of the right hand side of (15) and showing that it coincides with its left hand side. To this end, notice that since that optimization problem is convex, $\boldsymbol{c}$ is an optimal solution of the right hand side of (15) if and only if there is a Lagrange multiplier $\boldsymbol{\lambda} \in \mathbb{C}^N$ such that

$$\frac{\partial}{\partial \boldsymbol{c}}[\boldsymbol{c}^H\tilde{\boldsymbol{P}}^{-1}\boldsymbol{c} + \boldsymbol{\lambda}^H(\tilde{\boldsymbol{A}}^H\boldsymbol{c} - \boldsymbol{y})] = 0, \qquad \tilde{\boldsymbol{A}}^H\boldsymbol{c} = \boldsymbol{y},$$

or, equivalently,

$$2\tilde{\boldsymbol{P}}^{-1}\boldsymbol{c} + \tilde{\boldsymbol{A}}\boldsymbol{\lambda} = 0, \qquad \tilde{\boldsymbol{A}}^H\boldsymbol{c} = \boldsymbol{y}.$$

From this set of equations we obtain

$$\boldsymbol{\lambda} = -2(\tilde{\boldsymbol{A}}^H\tilde{\boldsymbol{P}}\tilde{\boldsymbol{A}})^{-1}\boldsymbol{y}$$
$$\boldsymbol{c} = \tilde{\boldsymbol{P}}\tilde{\boldsymbol{A}}(\tilde{\boldsymbol{A}}^H\tilde{\boldsymbol{P}}\tilde{\boldsymbol{A}})^{-1}\boldsymbol{y},$$

and the optimal cost of right hand side of (15) gives $\boldsymbol{c}^H\tilde{\boldsymbol{P}}^{-1}\boldsymbol{c} = \boldsymbol{y}^H(\tilde{\boldsymbol{A}}^H\tilde{\boldsymbol{P}}\tilde{\boldsymbol{A}})^{-1}\tilde{\boldsymbol{A}}^H\tilde{\boldsymbol{P}}\tilde{\boldsymbol{P}}^{-1}\tilde{\boldsymbol{P}}\tilde{\boldsymbol{A}}(\tilde{\boldsymbol{A}}^H\tilde{\boldsymbol{P}}\tilde{\boldsymbol{A}})^{-1}\boldsymbol{y} = \boldsymbol{y}^H(\tilde{\boldsymbol{A}}^H\tilde{\boldsymbol{P}}\tilde{\boldsymbol{A}})^{-1}\boldsymbol{y}$, which corresponds to the left hand side of (15). This concludes the proof of (7).

*Remark 16:* Equation (7) is closely related to the so-called Gauss-Markov theorem, which states that, in a linear regression framework, the least squares estimator is the minimum variance unbiased estimator [23]. In fact, let $\boldsymbol{z} = \tilde{\boldsymbol{A}}\boldsymbol{\theta} + \boldsymbol{e}$, where $\boldsymbol{\theta} \in \mathbb{C}^{K+N}$, $\boldsymbol{e} \sim \mathcal{CN}(\boldsymbol{0}, \tilde{\boldsymbol{P}}^{-1})$. Furthermore, suppose we are interested in estimating $x = \boldsymbol{y}^H\boldsymbol{\theta}$. Then, the cost function in the right hand side of (7) can be interpreted as the variance of an estimate $\hat{x} = \boldsymbol{c}^H\boldsymbol{z}$ of $x$, and the corresponding constraint $\tilde{\boldsymbol{A}}^H\boldsymbol{c} = \boldsymbol{y}$ restricts $\hat{x}$ to be unbiased, while the left hand side of (7) corresponds to the minimum achievable variance, according to the Gauss-Markov theorem.

## References


[1] 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.

[2] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Upper Saddle River, NJ, USA: Prentice Hall, 2005.

[3] B. Porat, *Digital Processing of Random Signals*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1994.

[4] Y. Wang, J. Li, and P. Stoica, *Spectral Analysis of Signals: The Missing Data Case*. San Mateo, CA, USA: Morgan & Claypool Publishers, 2005.

[5] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 629–638, Feb. 2011.

[6] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *J. Roy. Statist. Soc. Series B*, vol. 58, no. 1, pp. 267–288, 1996.

[7] H. Wang, G. Li, and G. Jiang, "Robust regression shrinkage and consistent variable selection through the LAD lasso," *J. Bus. Econ. Statist.*, vol. 25, no. 3, pp. 347–355, 2007.

[8] G. Elfving, "Optimum allocation in linear regression theory," *Annal. Math. Statist.*, vol. 23, no. 2, pp. 255–262, 1952.

[9] P. Bühlmann and S. van de Geer, *Statistics for High- Dimensional Data: Methods, Theory and Applications.*. Berlin, Germany: Springer-Verlag, 2011.

[10] L. Wang, M. D. Gordon, and J. Zhu, "Regularized least absolute deviations regression and an efficient algorithm for parameter tuning," in *Proc. Sixth Int. Conf. Data Mining (ICDM06)*, Hong Kong, 2006, pp. 690–700.

[11] P. Stoica and P. Babu, "SPICE and LIKES: Two hyperparameter-free methods for sparse-parameter estimation," *Signal Process.*, vol. 92, no. 7, pp. 1580–1590, 2012.

[12] 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.

[13] M. Fazel, H. Hindi, and S. P. Boyd, "A rank minimization heuristic with application to minimum order system approximation," in *Proc. Amer. Contr. Conf. (ACC'01)*, 2001, vol. 6, pp. 4734–4739.

[14] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Statist. Soc.: Series B (Statist. Methodol.)*, vol. 68, no. 1, pp. 49–67, 2006.

[15] M. R. Osborne, B. Presnell, and B. A. Turlach, "A new approach to variable selection in least squares problems," *IMA J. Numer. Anal.*, vol. 20, no. 3, pp. 389–403, 2000.

[16] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *Annal. Statist.*, vol. 32, no. 2, pp. 407–451, 2004.

[17] J. Liu, S. Ji, and J. Ye, "Multi-task feature learning via efficient l2, 1-norm minimization," presented at the 25th Conf. Uncertainty Artif. Intell. (UAI2009), 2009.

[18] X. Gao and J. Huang, "Asymptotic analysis of high dimensional LAD regression with Lasso," *Statistica Sinica*, vol. 20, pp. 1485–1506, 2010.

[19] E. Greenshtein and Y. Ritov, "Persistence in high-dimensional linear predictor selection and the virtue of overparameterization," *Bernoulli*, vol. 10, no. 6, pp. 971–988, 2004.

[20] D. P. Wipf and S. Nagarajan, "A new view of automatic relevance determination," in *Advances in Neural Information Processing Systems*, J. Platt, D. Koller, Y. Singer, and S. Roweis, Eds. Cambridge, MA, USA: MIT Press, 2008, vol. 20.

[21] M. C. Grant and S. P. Boyd, "Graph implementations for nonsmooth convex programs," in *Recent Adv. Learn. Contr. (Tribute to M. Vidyasagar)*, V. D. Blondel, S. P. Boyd, and H. Kimura, Eds. Berlin, Germany: Springer-Verlag, 2008, pp. 95–110.

[22] H. Zou, "The adaptive Lasso and its oracle properties," *J. Amer. Statist. Assoc.*, vol. 101, no. 476, pp. 1418–1429, 2006.

[23] S. M. Kay, *Fundamentals of Statistical Signal Processing, Estimation Theory.*. Englewood Cliffs, NJ, USA: Prentice Hall, 1993.

**Cristian R. Rojas** (M'13) was born in 1980. He received the M.S. degree in electronics engineering from the Universidad Técnica Federico Santa María, Valparaíso, Chile, in 2004, and the Ph.D. degree in electrical engineering at The University of Newcastle, NSW, Australia, in 2008. Since October 2008, he has been with the Royal Institute of Technology, Stockholm, Sweden, where he is currently an Assistant Professor of the Automatic Control Lab, School of Electrical Engineering. His research interests lie in system identification and signal processing.

**Dimitrios Katselis** (M'12) received the Ph.D. degree from the Department of Informatics and Telecommunications, University of Athens in 2009. After completing his military service, he joined in 2010 the Electrical Engineering Department at KTH Royal Institute of Technology as a postdoctoral researcher, where he currently holds the title of Research Associate. His research interests span the areas of system identification, MIMO and multicarrier communications, signal processing, compressive sensing, information theory and networking.

**Håkan Hjalmarsson** (M'98–SM'11–F'13) was born in 1962. He received the M.S. degree in electrical engineering in 1988, and the Licentiate degree and the Ph.D. degree in automatic control in 1990 and 1993, respectively, all from Linköping University, Sweden. He has held visiting research positions at California Institute of Technology, Louvain University and at the University of Newcastle, Australia. He has served as an Associate Editor for Automatica (1996–2001), and IEEE Transactions on Automatic Control (2005–2007) and been Guest Editor for European Journal of Control and Control Engineering Practice. He is Professor at the School of Electrical Engineering, KTH, Stockholm, Sweden. He is an IEEE Fellow and Chair of the IFAC Coordinating Committee CC1 Systems and Signals. In 2001 he received the KTH award for outstanding contribution to undergraduate education. He is co-recipient of the European Research Council advanced grant. His research interests include system identification, signal processing, control and estimation in communication networks and automated tuning of controllers.