SPICE : a sparse covariance-based estimation method for array processing

Petre Stoica, Fellow, IEEE, Prabhu Babu∗, Jian Li, Fellow, IEEE

Abstract

This paper presents a novel sparse iterative covariance-based estimation approach, abbreviated as SPICE, to array processing. The proposed approach is obtained by the minimization of a covariance matrix fitting criterion and is particularly useful in many-snapshot cases but can be used even in single-snapshot situations. SPICE has several unique features not shared by other sparse estimation methods: it has a simple and sound statistical foundation, it takes account of the noise in the data in a natural manner, it does not require the user to make any difficult selection of hyperparameters, and yet it has global convergence properties.

Index Terms: array processing, direction of arrival (DOA) estimation, sparse parameter estimation, covariance fitting.

I. INTRODUCTION AND PRELIMINARIES

Consider an array processing scenario in which the main problem is to estimate the location parameters of a number of narrowband sources that are present in the array’s viewing field. Let Ω denote the set of possible locations, and let θ be a generic location parameter. Also, let \{θ_k\}_{k=1}^K denote a grid that covers Ω. We assume that the grid is fine enough such that the true location parameters of the existing sources lie on (or, practically, close to) the grid. Under this reasonable assumption we can use the following non-parametric model for the output of the array (see, e.g. [1]) :

\[ y(t) = \sum_{k=1}^{K} a_k s_k(t) + e(t) \quad t = 1, \cdots , M \quad (N \times 1) \]  

This work was supported in part by the Swedish Research Council (VR), the European Research Council (ERC), the Office of Naval Research (ONR) under Grants No. N00014-09-1-0211 and N00014-10-1-0054, and the National Science Foundation (NSF) under Grant No. ECCS-0729727.

Petre Stoica and Prabhu Babu are with the Dept. of Information Technology, Uppsala University, Uppsala, SE 75105, Sweden. Jian Li is with the Dept. of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA.

∗Please address all the correspondence to Prabhu Babu, Phone: (46) 18-471-3394; Fax: (46) 18-511925; Email: prabhu.babu@it.uu.se

August 9, 2010 DRAFT
where $M$ is the total number of snapshots, $N$ is the number of sensors in the array, $y(t) \in \mathbb{C}^{N \times 1}$ is the \textit{t}-th observed snapshot, $a_k \in \mathbb{C}^{N \times 1}$ denotes the array transfer vector (aka manifold or steering vector) corresponding to $\theta_k$, $s_k(t) \in \mathbb{C}$ is the unknown signal impinging on the array from a possible source at $\theta_k$, and $\epsilon(t) \in \mathbb{C}^{N \times 1}$ is a noise term.

A \textit{sparse} (or semi-parametric) estimation method makes the assumption, reminiscent of the parametric approach, that only a small number of sources exists and therefore that only a few rows of the signal matrix

$$S = \begin{bmatrix} s_1(1) & \cdots & \cdots & s_1(M) \\ \vdots & \vdots & \vdots & \vdots \\ s_K(1) & \cdots & \cdots & s_K(M) \end{bmatrix}$$  \hspace{1cm} (2)$$

are different from zero. The estimation problem is then to decide from the data $\{y(t)\}$ which rows of the above matrix are non-zero. Indeed, once this is done, the solution to the location estimation problem, which is usually the main goal of array processing, is immediate: if the row $\tilde{k}$ (let us say) of (2) is deemed to be different from zero then we can infer that there is a corresponding source at an estimated location equal to $\theta_{\tilde{k}}$.

The previous formulation of the location problem begs for the use of basic ideas from the area of sparse parameter estimation, or rather simple extensions of those ideas to the present multi-snapshot case. To describe these ideas briefly, let:

$$Y^* = [y(1), \cdots, y(M)] \in \mathbb{C}^{N \times M}$$  \hspace{1cm} (3)$$

$$S = \begin{bmatrix} s_1^* \\ \vdots \\ s_K^* \end{bmatrix} \in \mathbb{C}^{K \times M}$$  \hspace{1cm} (4)$$

$$B^* = [a_1 \cdots a_K] \in \mathbb{C}^{N \times K}$$  \hspace{1cm} (5)$$

where the superscript $*$ denotes the conjugate transpose (we denote the matrix in (5) by $B$ to reserve the notation $A$ for an extended form of (5) that will be introduced in the next section). A direct application of the $\ell_1$-norm minimization principle (see, e.g. [2]) to the present scenario described by (1)-(5) consists of estimating the matrix $S$ as the solution to the following constrained minimization problem:

$$\min_S \sum_{k=1}^K \|s_k\| \quad \text{s.t.} \quad \|Y^* - B^* S\| \leq \eta$$  \hspace{1cm} (6)$$

where $\|\cdot\|$ denotes the Euclidean norm for vectors and the Frobenius norm for matrices, and $\eta$ is a threshold that must be chosen by the user. Note that the objective in (6) is equal to the $\ell_1$-norm of the vector $\{\|s_k\|\}_{k=1}^K$, an observation that shows clearly that this approach is a direct extension of the standard single-snapshot approach of, e.g., [2].
A method for array processing based on (6) was pursued in [3]. Note that (6) is easily recognized to be an SOCP (second order cone program), see, e.g., [4], which can be efficiently solved provided that $N$, $M$ and $K$ do not take on too large values. However, in the array processing application this is not necessarily the case: indeed, while $N \approx 10^{-10^2}$ (tens to hundreds) is reasonably small, $M \approx 10^2 - 10^3$ and $K \approx 10^2 - 10^6$ (depending on the desired resolution and the dimension of $\theta$ (1D, 2D, 3D etc)) can be rather large. For such large dimensional problems the currently available SOCP software (see, e.g., [5]) is too slow to use. In an attempt to overcome this computational problem, [3] suggested a way to reduce the number of columns of $Y^*$ and $S$ in (6) to manageable values by means of a singular value decomposition operation.

The approaches that rely on (6), such as the one in [3], suffer from a number of problems. First, the motivation of (6) is more clearly established in the noise-free case than in the more practical noisy data case. Second, and likely related to the first problem, there exist no clear-cut guidelines for the selection of $\eta$ in (6) (see, e.g., [6] for a critical discussion on this aspect). Finally, solving (6) as an SOCP may be too time consuming for some applications in which $N$, $M$ and especially $K$ take on large values, and the available techniques for reducing the dimensions in (6) (such as that suggested in [3]) require the choice of further hyperparameters, besides $\eta$, and even so they are only approximate.

In this paper we present a new method of sparse parameter estimation in models of the type of (1). This method, which is called SPICE (sparse iterative covariance-based estimation), is obtained using a novel covariance-based fitting approach that was recently introduced in [7] where the focus was on time-series data (the single-snapshot case) as opposed to the array data (the multi-snapshot scenario) considered here. SPICE has a number of useful features that are hardly shared by other sparse estimation methods. In particular, SPICE does not suffer from the problems described in the previous paragraph. Indeed, as will be shown in the next section, SPICE has a sound (covariance-based) statistical motivation which makes it possible to use the method in noisy data scenarios without the need for choosing any hyperparameters. Additionally, the SPICE algorithm has a simple form, and yet it enjoys global convergence properties. We will show that the covariance fitting problem that SPICE solves can be reformulated as an SOCP of the form of (6) but with $\eta = 0$, with $Y^*$ and $S$ replaced by other matrices with only $N$ columns (typically $M \gg N$), and with row-augmented matrices $A$ and $C$ that, unlike $B$ and $S$ in (6), take account of the noise term in the data equation (1). This SOCP formulation of the SPICE estimation criterion shows that the $\ell_1$-norm minimization problem in (6) can be given a simple statistical motivation based on covariance fitting, provided that the matrices in (6) are suitably defined, see the rest of this paper for details.
II. SPICE ESTIMATION CRITERION

Let us assume that

\[
E[\epsilon(t)\epsilon^*(\bar{t})] = \begin{bmatrix}
\sigma_1 & 0 & \cdots & 0 \\
0 & \sigma_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & \sigma_N \\
\end{bmatrix} \delta_{t,\bar{t}}
\] (7)

where \( E \) stands for the expectation operator, and

\[
\delta_{t,\bar{t}} = \begin{cases} 
1 & \text{if } t = \bar{t} \\
0 & \text{elsewhere.}
\end{cases}
\] (8)

This assumption on the noise term in (1) is reasonable in most applications. Let us also assume that the signals \( \{s_k(t)\} \) and the noise \( \epsilon(\bar{t}) \) are uncorrelated with each other for any \((t, \bar{t})\) and that:

\[
E[s_k(t)s^*_\bar{k}(\bar{t})] = p_k \delta_{k,\bar{k}} \delta_{t,\bar{t}}
\] (9)

Then the data snapshots \( \{y(1), \cdots, y(M)\} \) are uncorrelated with one another and have the following covariance matrix :

\[
R = E[y(t)y^*(t)] = \sum_{k=1}^{K} p_k a_k a^*_k + \begin{bmatrix}
\sigma_1 & 0 & \cdots & 0 \\
0 & \sigma_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & \sigma_N \\
\end{bmatrix} \\
\begin{bmatrix}
p_1 & 0 & \cdots & \cdots & \cdots & 0 \\
0 & p_2 & 0 & \cdots & \cdots & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & \sigma_1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & \sigma_N \\
\end{bmatrix} \\
\begin{bmatrix}
a_1^* \\
\vdots \\
a_K^* \\
I
\end{bmatrix}
\] (10)

\[
\Delta \triangleq A^* PA
\]

where

\[
A^* \triangleq [a_1, \ldots, a_K I] \triangleq [a_1, \ldots, a_K a_{K+1}, \ldots, a_{K+N}]
\] (11)
The assumption in (9) that the source signals are spatially uncorrelated, which led to the above expression for the covariance matrix $R$, does not always hold: in some applications the signals can be correlated or even coherent. However the SPICE method proposed in this paper is robust to this assumption - we refer to [7] (see also [1]) for a theoretical explanation of this robustness property, which will be illustrated numerically in Section V.

We will consider the following covariance fitting criterion for the purpose of parameter estimation (see, e.g., [8] and the references therein):

$$ f = \|R^{-1/2}(\hat{R} - R)\hat{R}^{-1/2}\|^2 $$

where $R^{-1/2}$ denotes the positive definite square-root of $R^{-1}$,

$$ \hat{R} = Y^*Y/M $$

and the inverses of $\hat{R}$ and $R$ are assumed to exist. Note that $R^{-1}$ exists under weak conditions: indeed, while the “true” values of many $\{p_k\}$ in $R$ may be zero, one typically has $\sigma_k > 0$ ($k = 1, \cdots, N$) which renders $R$ nonsingular. Regarding $\hat{R}$, the inverse of this matrix exists with probability one as long as $M > N$. However, for $M < N$ the sample covariance matrix $\hat{R}$ is singular and therefore (13) cannot be used. In the latter case one can estimate the parameters $\{p_k\}$ by minimizing the following criterion, instead of (13),

$$ \|R^{-1/2}(\hat{R} - R)\|^2 $$

The above criterion is the one used in [7] where the focus was on the time series case with $M = 1$. In the present array processing scenario, however, we prefer to consider (13) because usually the condition $M > N$ is satisfied and because (13) has a statistically stronger motivation than (15): indeed, under certain conditions specified in [8], it can be shown that the parameter estimates minimizing (13) are statistically asymptotically (in $M$) efficient, whereas the estimates obtained from (15) are sub-optimal. Nevertheless, we should note that what we do in the following for (13) applies with minor modifications to (15) as well, see Remark 1 below - this observation can be useful in those array processing applications in which $M < N$ and therefore in which (15) should be used in lieu of (13).
A simple calculation shows that:

\[ f = \text{tr} \left[ R^{-1} (\hat{R} - R) \hat{R}^{-1} (\hat{R} - R) \right] = \text{tr} \left[ (R^{-1} \hat{R} - I)(I - R^{-1} \hat{R}) \right] = \text{tr}(R^{-1} \hat{R}) + \text{tr}(\hat{R}^{-1} R) - 2N \]  

(16)

where

\[ \text{tr} \left( \hat{R}^{-1} R \right) = \sum_{k=1}^{K+N} p_k a_k^* \hat{R}^{-1} a_k \]  

(17)

It follows from (16) and (17) that the minimization of \( f \) is equivalent to the minimization of the function:

\[ g = \text{tr} \left( \hat{R}^{1/2} R^{-1/2} \hat{R}^{1/2} \right) + \sum_{k=1}^{K+N} (a_k^* \hat{R}^{-1} a_k) p_k \]  

(18)

Remark 1: A similar calculation shows that in the case of (15) the function to be minimized with respect to the unknown powers \( \{p_k\} \) is:

\[ \text{tr} \left( \hat{R} R^{-1} \hat{R} \right) + \sum_{k=1}^{K+N} \|a_k\|^2 p_k \]  

(19)

Owing to the analogy between (18) and (19), it should come as no surprise that what we do in the following for (18) can also be done for (19) if necessary (e.g., for \( M < N \)).

The problem of minimizing \( g \) in (18) with respect to \( \{p_k\} \) can be easily shown to be an SDP (semi-definite program), see Appendix A, and is therefore convex (see, e.g., [9]). Solving (18) as an SDP, however, is not recommended due to the fact that the available SDP solvers are too computationally intensive for the values of \( N, M \) and \( K \) encountered in many array processing applications. Consequently we adopt a different line of attack that consists of replacing the problem of minimizing (18) by a related problem, as described next.

It follows from (17) that a consistent (in \( M \)) estimate of the right-hand side of this equation is given by \( N \). Therefore we can think of reformulating the problem of minimizing \( g \) as the following constrained minimization:

\[ \min_{\{p_k \geq 0\}} \text{tr} \left( \hat{R}^{1/2} R^{-1/2} \hat{R}^{1/2} \right) \quad \text{s.t.} \sum_{k=1}^{K+N} w_k p_k = 1 \]  

(20)

where

\[ w_k = a_k^* \hat{R}^{-1} a_k / N \]  

(21)

Interestingly, the problems (18) and (20) are not only asymptotically equivalent (as \( M \) increases, and under the condition that \( R \) in (10) can represent the true covariance matrix) but they are exactly equivalent (in general) in the sense that their solutions are scaled versions of each other (note that a scaling of \( \{p_k\} \) has no effect on source location estimation). This equivalence property is proved in Appendix B where we also show that the problem obtained from (18) by constraining the first term to \( N \) (a constraint suggested, once again, by asymptotic considerations) is equivalent to (18) and (20) as well.

The problem (20) is also an SDP, and therefore convex. Furthermore, note that the linear constraint in (20) is of the (weighted) \( \ell_1 \)-norm type, and thus it can be expected to be sparsity inducing for the
solution to (20). Apparently, the criterion in (20) was never considered in the previous literature on sparse parameter estimation. On the other hand, interestingly enough, this type of criterion occurs frequently in the seemingly unrelated literature on optimal experiment design (see, e.g. [10] and the references of that paper). The latter literature (in particular [10]) has served as a source of inspiration for [7] and, to some extent, for this paper too.

III. SPICE UPDATING FORMULAS

Let $C \in \mathbb{C}^{(K+N) \times N}$ and consider the following problem:

$$\min_C \text{ tr } (C^* P^{-1} C) \quad \text{s.t. } A^* C = \hat{R}^{1/2} \tag{22}$$

The solution to (22) is given (for fixed $P$) by

$$C_0 = P A R^{-1} \hat{R}^{1/2} \tag{23}$$

and the corresponding minimum value of the function in (22) is

$$\text{tr} \left( C_0^* P^{-1} C_0 \right) = \text{tr} \left( \hat{R}^{1/2} R^{-1} \hat{R}^{1/2} \right) \tag{24}$$

To prove the above assertion, observe that (23) follows if we can show that (hereafter the notation $X \succeq Y$, with $X$ and $Y$ being Hermitian matrices of appropriate dimensions, means that the difference matrix $X - Y$ is positive semi-definite):

$$C^* P^{-1} C \succeq C_0^* P^{-1} C_0 = \hat{R}^{1/2} R^{-1} \hat{R}^{1/2} \quad \text{s.t. } A^* C = \hat{R}^{1/2} \tag{25}$$

By standard properties of partitioned matrices (see, e.g., [11], Lemma A.3) and the fact that $R > 0$, (25) holds if and only if the following partitioned matrix is positive semi-definite:

$$\begin{bmatrix} C^* P^{-1} C & \hat{R}^{1/2} \\ \hat{R}^{1/2} & R \end{bmatrix} = \begin{bmatrix} C^* P^{-1} C & C^* A \\ A^* C & A^* P A \end{bmatrix} \geq 0 \tag{26}$$

The central matrix in (26) can be re-written as

$$\begin{bmatrix} P^{-1} I \\ I & P \end{bmatrix} = \begin{bmatrix} P^{-1/2} \\ P^{1/2} \end{bmatrix} \begin{bmatrix} P^{-1/2} & P^{1/2} \end{bmatrix} \tag{27}$$

and thus it is obviously positive semi-definite (because it has the form $X^* X$, with $X = [P^{-1/2} P^{1/2}]$). Therefore (25) (and hence (23)) is proved and then (24) follows via substitution.

To summarize, we have proved above (see (24)) that the minimization of the objective in (22) with respect to $C$, for any fixed $P \succeq 0$, leads to the original function of $P$ in (20). Then it follows that the minimization of (22) with respect to $C$ and $\{p_k\}$ yields the same $\{p_k\}$ as the minimization of (20).
usefulness of this observation lies in the fact that the minimization of the augmented function in (22) can be conveniently done by means of a cyclic algorithm that minimizes (22) with respect to $C$, for fixed $P$, then minimizes (22) with respect to $P$, for given $C$, and so forth until convergence. The solution to the first step of this algorithm was already derived, see (23). The solution needed in the second step can also be obtained in closed form. To show how this can be done, let

$$
C = \begin{bmatrix}
c_1^* \\
\vdots \\
c_{K+N}^*
\end{bmatrix} 
$$

and observe that

$$
\text{tr} \left( C^* P^{-1} C \right) = \text{tr} \left( P^{-1} C C^* \right) = \sum_{k=1}^{K+N} \|c_k\|^2 / p_k 
$$

A. The case of different $\{\sigma_k\}$

By the Cauchy-Schwarz inequality

$$
\left( \sum_{k=1}^{K+N} w_k^{1/2} \|c_k\| \right)^2 \leq \left[ \sum_{k=1}^{K+N} \|c_k\|^2 / p_k \right] \left[ \sum_{k=1}^{K+N} w_k p_k \right] = \sum_{k=1}^{K+N} \|c_k\|^2 / p_k
$$

From this it follows that the minimization of the objective in (22) with respect to $\{p_k\}$ (s.t. $p_k \geq 0$ and $\sum_{k=1}^{K+N} w_k p_k = 1$) gives (for fixed $C$):

$$
p_k = \frac{\|c_k\|}{w_k} \frac{1}{\rho}, \quad \rho = \sum_{m=1}^{K+N} w_m^{1/2} \|c_m\| 
$$

and the corresponding minimum value of the objective is

$$
\left( \sum_{k=1}^{K+N} w_k^{1/2} \|c_k\| \right)^2 
$$

Equation (31) above provides the solution to the second step of the cyclic algorithm, whereas the solution to the first step is given by (23). Combining (23) and (31) leads to the updating formulas of the SPICE algorithm in which only the powers $\{p_k\}$ (that are the quantities of main interest) appear explicitly:

$$
p_k^{i+1} = p_k^i \frac{\|a_k R^{-1}(i) \hat{R}^{1/2} \|}{w_k^{1/2} \rho(i)} \quad k = 1, \cdots, K+N 
$$

$$
\rho(i) = \sum_{m=1}^{K+N} w_m^{1/2} p_m^i \|a_m R^{-1}(i) \hat{R}^{1/2} \| 
$$

where the index $i$ denotes the iteration number, and $R(i)$ is the matrix $R$ made from $\{p_k^i\}$. The algorithm can be initialized with the power estimates obtained by means of the periodogram method (see, e.g., [1]):

$$
p_k^0 = a_k^* \hat{R} a_k / \|a_k\|^4 \quad k = 1, \cdots, K+N 
$$

Remark 2: The SPICE algorithm for the alternative covariance fitting criterion in (15) (or (19)) can be readily derived by paralleling the above calculations. The result is an updating formula similar to (33) above with the only difference that $\hat{R}^{1/2}$ in (33) and (34) should be replaced by $\hat{R}$ and $\{w_k\}$ in (21) by $w_k = \|a_k\|^2 / \text{tr}(\hat{R})$. 

August 9, 2010 DRAFT
The case of identical \( \{\sigma_k\} \)

In some applications it is known that the noise components in the different elements of the array output vector have the same variance:

\[
\sigma_1 = \cdots = \sigma_N \Delta = \sigma
\]  

(36)

Using this information is important as based on it we can reduce the number of powers that need to be estimated. To derive the necessary modification of SPICE, that takes (36) into account, first observe that the minimization of (22) with respect to \( C \) is not affected by (36). However the minimization with respect to \( \{p_k\}_{k=1}^K \) and \( \sigma \), for fixed \( C \), is slightly different. Under the above constraint on \( \{\sigma_k\} \), equation (29) becomes:

\[
\text{tr}(C^*P^{-1}C) = \sum_{k=1}^K \|c_k\|^2/p_k + \sum_{k=K+1}^{K+N} \|c_k\|^2/\sigma
\]  

(37)

This function is to be minimized with respect to \( \{p_k \geq 0\} \) and \( \sigma \geq 0 \), subject to

\[
\sum_{k=1}^K w_k p_k + \gamma \sigma = 1
\]  

(38)

where

\[
\gamma = \frac{\sum_{k=K+1}^{K+N} w_k}{\sum_{k=K+1}^{K+N} \|c_k\|^2}
\]  

(39)

By the same argument as above, see (30)-(32), the solution to this optimization problem is:

\[
p_k = \frac{\|c_k\|^2}{w_k^{1/2} \rho}, \quad k = 1, \cdots, K
\]  

(40)

\[
\sigma = \left[ \frac{\sum_{k=K+1}^{K+N} \|c_k\|^2}{\gamma^{1/2} \rho} \right]^{1/2}
\]  

(41)

where

\[
\rho = \frac{K}{\sum_{k=1}^K w_k^{1/2} \|c_k\|^2} + \gamma^{1/2} \left[ \frac{\sum_{k=K+1}^{K+N} \|c_k\|^2}{\gamma^{1/2} \rho} \right]^{1/2}
\]  

(42)

and the corresponding minimum function (for fixed \( C \)) is

\[
\left( \frac{K}{\sum_{k=1}^K w_k^{1/2} \|c_k\|^2} + \gamma^{1/2} \left[ \frac{\sum_{k=K+1}^{K+N} \|c_k\|^2}{\gamma^{1/2} \rho} \right]^{1/2} \right)^2
\]  

(43)

Inserting the expression (23) for \( C \) in (40) and (41) we obtain the following modified SPICE algorithm for the case of (36):

\[
p_k^{i+1} = \frac{\|a_k^* R^{-1}(i) \hat{R}^{1/2}\|}{w_k^{1/2} \rho(i)} p_k^i, \quad k = 1, \cdots, K
\]  

(44)

\[
\sigma^{i+1} = \frac{\|R^{-1}(i) \hat{R}^{1/2}\|}{\gamma^{1/2} \rho(i)}
\]  

(45)

\[
\rho(i) = \frac{K}{\sum_{k=1}^K w_k^{1/2} p_k^i \|a_k^* R^{-1}(i) \hat{R}^{1/2}\|} + \gamma^{1/2} \sigma^i \|R^{-1}(i) \hat{R}^{1/2}\|
\]  

(46)

Initial estimates of the powers, for \( k = 1, \cdots, K \), can still be obtained with the periodogram (see (35)), and \( \sigma \) can be initialized for instance as the average of the \( N \) smallest values of \( \{p_k^0\}_{k=1}^K \) each
multiplied by $\|a_k\|^2$. To motivate this choice for $\sigma^0$ note that, at least for $M$ sufficiently large, we have $p^0_k = a_k^* \hat{R} a_k / \|a_k\|^2 \geq \sigma / \|a_k\|^2$ with the equality being likely to hold for the smallest values of $p^0_k$, e.g. for the $N$ smallest such values. This observation implies that (below $\{\tilde{p}_k^0\}_{k=1}^N$ are the $N$ smallest values of the set $\{p_k^0\}$ and $\{\tilde{a}_k\}$ the corresponding manifold vectors)

$$\sigma^0 = \frac{N}{\|\tilde{a}_k\|_2^2} \sum_{k=1}^N \tilde{p}_k^0 / N$$

(47)

can be expected to be a reasonable estimate of $\sigma$.

Because SPICE monotonically decreases the objective function (due to its cyclic operation) and as the minimization problem it solves is convex, we can expect that the algorithm has global convergence properties. That this is indeed the case follows from the general analysis in [10] where it was proved that, under weak conditions (essentially requiring that $p^0_k > 0$ and that the matrix $R(i)$ stays positive definite as the iteration proceeds), the limit points of the SPICE iterative process are global solutions to (20).

IV. SOCP FORMULATION OF SPICE

SPICE takes account of the noise without introducing any hyperparameters to be selected by the user. Indeed, SPICE is a fully data adaptive approach that estimates both $\{p_k\}$ and $\{\sigma_k\}$ from the available data snapshots. As already mentioned, to our knowledge, the SPICE criterion was never used in the previous literature on sparse parameter estimation. It is therefore somewhat unexpected that SPICE can be related to (an extended version of) the classical $\ell_1$-norm minimization approach in (6), as we explain in the rest of this section.

As indicated in Section II the SPICE estimation problem, (20), can be cast and solved as an SDP. However it can also be reformulated as an SOCP which is a special case of SDP. It is this reformulation of (20) as an SOCP that reveals the connection between SPICE and a certain extended $\ell_1$-norm minimization problem. To reformulate SPICE as an SOCP we proceed in the following way. Instead of minimizing (22) cyclically (or alternately) with respect to $C$ and $P$, we minimize this function with respect to $\{p_k\}$ for arbitrary $C$. The result of this minimization operation is already available from the calculations in the previous section where it was shown that the minimizing powers are given by (31) and the corresponding minimum function by (32). It remains to minimize (32) with respect to $C$ (under the constraint in (22)):

$$\min_C \sum_{k=1}^{K+N} w_k^{1/2} \|c_k\| \quad \text{s.t.} \quad A^* C = \hat{R}^{1/2}$$

(48)

Once $C$ is obtained from (48), $\{p_k\}$ can be calculated using (31).

The above problem, which is an SOCP (see Appendix A), is similar to the $\ell_1$-norm minimization problem (6). However there are also significant differences between these two problems, as well as between the ways in which they have been obtained:

i) The $\ell_1$-norm in (48) has a weighted form, with the weights determined by the data. It is well known
that \( w_k = a_k^* \hat{R}^{-1} a_k / N \) is an estimate of the inverse power corresponding to the \( k \)-th point of the location grid (see, e.g., [1]). Therefore, the smaller the said power the larger the \( w_k^{1/2} \) and consequently more weight is put on minimizing the associated \( \| c_k \| \) in (48), which intuitively is the way it should be. Furthermore, the \( \ell_1 \)-norm objective in (48) follows naturally from the statistically sound covariance fitting criterion in (13), in contrast with the similar objective in (6) whose motivation is essentially heuristical.

ii) The matrices \( A \) and \( C \) in (48) comprise extra rows that account for the noise in the data, unlike the similar quantities \( B \) and \( S \) in (6).

iii) The inequality constraint in (6) is replaced in (48) by a hyperparameter-free equality constraint where \( \hat{R}^{1/2} \) is used in lieu of \( Y^* \) in (6). Note that the matrix \( C \) in (48) has \( N \) columns instead of \( M \) columns as \( S \) in (6), which evidently makes (48) more convenient computationally (typically \( M \gg N \)). This dimensional reduction also follows naturally from the employed covariance fitting criterion, whereas the idea suggested in [3] for reducing the number of columns of \( Y^* \) and \( S \) in (6) was motivated on more or less heuristical grounds and its use depends on the selection of an additional hyperparameter, besides \( \eta \).

The reader might wonder what is the relationship between \( C \) in (48) and \( S \) in (6). To clarify this aspect, observe that we can write the data equation (1) in the following matrix form:

\[
Y^* = B^* S + \Delta 
\]  

(49)

where

\[
\Delta = [\epsilon(1), \ldots, \epsilon(M)]
\]  

(50)

It follows from (49) that we can assume, as in (6), that \( \| Y^* - B^* S \| \leq \eta \) provided that \( \eta \) is "sufficiently large"; furthermore, a proper selection of \( \eta \) would require information about the noise variance(s) and the noise distribution, which is usually unavailable, and even then the inequality constraint in (6) could not in general be made to hold surely for any finite practically convenient value of \( \eta \). On the other hand, it also follows from (49) that

\[
Y^* = A^* \begin{bmatrix} S \\ \Delta \end{bmatrix}
\]  

(51)

and therefore that

\[
A^* \begin{bmatrix} S \\ \Delta \end{bmatrix} Y \hat{R}^{-1/2} / M = \hat{R}^{1/2} 
\]  

(52)

\[
\hat{R}^{1/2} \Delta \nabla \triangleq C
\]  

The above equation clarifies the relationship between \( C \) and \( S \). Unlike (6), the equality constraint (52) holds deterministically and it does not involve any hyperparameter. Note also that if \( S \) is row-sparse then so is the matrix \( S Y \hat{R}^{-1/2} / M \) in (52). Regarding the other matrix appearing in (52), viz. \( \Delta Y \hat{R}^{-1/2} / M \), it tends to zero as the signal-to-noise ratio (SNR) increases; however, for small or medium SNR values...
this matrix is not row-sparse, which may be perceived as a downside of (52), although a relatively minor one as the matrix $C$ can still be considered to be row-sparse because $K \gg N$.

Whenever the constraint $\sigma_1 = \cdots = \sigma_N \triangleq \sigma$ is enforced, the above reformulation of the SPICE estimation problem should be changed as follows. First, $C$ is obtained as the solution of the following SOCP (see (43)):

$$
\min_C \sum_{k=1}^K w_k^{1/2} \|c_k\| + \gamma^{1/2} \left[ \sum_{k=K+1}^{K+N} \|c_k\|^2 \right]^{1/2} \quad \text{s.t. } A^*C = \hat{\mathbf{R}}^{1/2}
$$

(53)

(see Appendix A for the formulation of the above problem as an SOCP). Then $\{p_k\}$ and $\sigma$ are determined using (40)-(42).

According to our albeit limited experience, the SPICE multiplicative algorithm in (33)-(34) (or (44)-(46)) is usually visibly faster than the SOCP-based algorithm that was outlined in this section. However, the SOCP formulation of the SPICE estimation problem remains important for understanding the connection between the proposed approach and the standard one based on $\ell_1$-norm minimization.

V. NUMERICAL ILLUSTRATIONS AND CONCLUDING REMARKS

In this section we illustrate numerically the performance of the proposed methods and compare it with the performance of some standard methods from the literature. We will consider both a uniform linear array (ULA) and a non-uniform linear array (NULA). In the ULA case the sensors are uniformly placed with a spacing of $\frac{\lambda}{2}$, where $\lambda$ denotes the wavelength of the sources. The inter-element spacings in the case of NULA are as shown in Fig 1. The number of sensors in the array is $N = 10$ for ULA and $N = 100$ for NULA. In both cases, the number of snapshots is $M = 200$. The steering vector $a_k$ for the ULA, corresponding to a direction of arrival (DOA) equal to $\theta_k$, is given by:

$$
a_k = \begin{bmatrix} e^{i\pi \sin(\theta_k)} \\
\vdots \\
e^{i\pi N \sin(\theta_k)} \end{bmatrix}
$$

(54)

(observe that $\|a_k\|^2 = N$ is a constant). The steering vector for the NULA can be similarly defined. The interval for the DOA is $\Omega = (-90^\circ, 90^\circ]$. We use a uniform grid $\{\theta_k\}_{k=1}^K$ to cover $\Omega$, with a step of $0.1^\circ$, which means that $K = 1800$.

The source signals $\{s_k(t)\}$, see (1), have constant modulus, which is usually the situation in communications applications. We will consider cases with both uncorrelated and coherent sources. The noise term in (1) is chosen to be white, both temporally and spatially, and Gaussian distributed with zero mean and variance $\sigma$. 

August 9, 2010 DRAFT
A. Fixed sources: DOA estimation

In this subsection we consider DOA estimation of fixed sources for both the ULA and the NULA. For the ULA, the data samples were simulated using equation (1) with three sources at \( \theta_1 = 10^\circ, \theta_2 = 40^\circ, \theta_3 = 55^\circ \) and the following signals 
\[
    s_1(t) = 3e^{i\phi_1(t)}, \quad s_2(t) = 10e^{i\phi_2(t)}, \quad s_3(t) = 10e^{i\phi_3(t)},
\]
where the phases \( \{\phi_k(t)\}_{k=1}^3 \) were independently and uniformly distributed in \([0, 2\pi] \). For the NULA, the data samples were simulated using three sources at \( \theta_1 = 10^\circ, \theta_2 = 36^\circ, \) and \( \theta_3 = 37^\circ \) with the same signals as for the ULA. To simulate the case of coherent sources, the phases of the respective sources were set to identical values for all \( t \)'s. In the case of NULA, the sources at \( \theta_2 \) and \( \theta_3 \) were chosen to be closer than for ULA, as the NULA has a much larger aperture and thus a much better resolution than the ULA.

In the simulation with coherent sources, for both ULA and NULA, the sources at \( \theta_1 \) and \( \theta_3 \) were the coherent ones, whereas the source at \( \theta_2 \) was uncorrelated to them. The noise variance was varied to control the signal-to-noise ratio defined as:

\[
    \text{SNR} = 10 \log \left( \frac{100}{\sigma} \right) = 20 - 10 \log \sigma \quad [\text{dB}]
\]

The following methods were used for DOA estimation:

- \( M_1 \): The periodogram (PER), see (35).
- \( M_2 \): The iterative adaptive approach (IAA), see [12].
- \( M_3 \): The multiple signal classification method (MUSIC), see [1].
- \( M_4 \): SPICE, see (15), (19) as well as Remark 2 and [7].
- \( M_5 \): The enhanced SPICE or, for short, SPICE-plus (SPICE+), see (44)-(46).

Note that IAA was shown in [12] to outperform the method of [3] and its variations in the literature, which is why we selected IAA for the present comparison. As a cross-check and for comparison’s sake we have also computed the SPICE+ estimates by means of an SDP solver (see [5] for the latter). The so-obtained estimates were identical (within numerical accuracy) to those provided by the SPICE+ algorithm; however, as expected, the SDP solver was either slower (it required 92 sec for one realization of the ULA example, whereas SPICE+ needed 51 sec) or, worse, it could not be executed for instance due to memory problems (such as in the NULA example in which \( N \) is ten times larger than in the ULA case).

In Figs 2-6 we show the histograms, based on 100 Monte Carlo runs, of the DOA estimates obtained with \( M_1 - M_5 \) in the ULA case for two different SNR values. Similarly, Figs 7-11 show the histograms of DOA estimates obtained in the NULA case. In each of these figures, the two top plots show the histograms for uncorrelated sources and the two bottom plots show the histograms for coherent sources. As can be seen from these figures PER gives biased estimates of the DOAs for the ULA, and fails to resolve the sources at \( \theta_2 \) and \( \theta_3 \) in the NULA case, due to leakage problems. IAA works fine and resolves the DOAs for both uncorrelated and coherent sources, except in the NULA case at the low
SNR. On the other hand, MUSIC estimates accurately all the DOAs in the uncorrelated signal case (even at the low SNR) but fails to work well in the case of coherent sources (as expected); note also that MUSIC requires knowledge of the number of sources present in the data. The proposed methods SPICE and SPICE+ yield the best performance in most of the cases considered. Performance-wise, there is no significant difference in the presented results between SPICE and SPICE+. Computation-wise, however, SPICE+ converged faster than SPICE presumably due to the data-dependent weights used by SPICE+.

B. Mobile sources: DOA tracking

In this subsection, we consider DOA estimation of mobile sources using the NULA. The data samples were generated assuming two mobile sources, one moving linearly from 30° to 60° and the other from 60° to 30°, in steps of 0.03°, over a course of 1000 data snapshots. The uncorrelated signals of these two sources were given by $s_1(t) = 10e^{i\varphi_1(t)}$ and $s_2(t) = 10e^{i\varphi_2(t)}$, where the phases $\{\varphi_k(t)\}$ were independently and uniformly distributed in $[0, 2\pi]$.

In this example, only the methods $M_1$, $M_4$ and $M_5$ are considered. At any given time instant $t$, the sample covariance matrix $\hat{R}$ is formed from the most recent 100 data snapshots and the DOA estimates are re-computed by each of the three methods. As before, the SPICE methods, $M_4$ and $M_5$, are initialized with the PER estimate. Alternatively, one may think of initializing the SPICE methods with their respective estimates obtained from the previous data window; however, as discussed in Section III-B, for global convergence SPICE should be initialized with a dense estimate rather than a sparse one. The SPICE methods were iterated only 5 times except at $t = 100$ where they were applied for the first time and were iterated till convergence. We have also tried a larger number of iterations, such as 20, but noticed no significant improvement in the DOA estimates (compare Figs 12(c) and (d) below). The SNR in this simulation was 20 dB (however very similar results were obtained for SNR = 10 dB and 0 dB).

Fig 12 shows the plots of DOA estimates vs. $t$ for the three considered methods. In each plot, the estimates shown at any time $t$ were obtained as the locations of the two largest peaks in the spatial spectrum provided by the method in question. As can be seen from the figure, PER performs poorly as it cannot track either the source moving from 30° to 45° or the one moving from 45° to 30°. Note that at each time instant, the data window (made of the most recent 100 data snapshots) comprises two sets of DOAs each with a width of 3°. For instance, at $t = 400$, the data window will contain signals with DOAs lying in the interval $[39°, 42°]$ and, respectively, in $[48°, 51°]$. PER often picks wrongly the two peaks only from the set with larger DOAs. On the other hand, the SPICE methods correctly pick the two peaks from both sets which leads to the band-like appearance of the corresponding plots in Figs 12(b)-(d); observe that, as expected, the width of the bands in these figures (measured along the DOA axis) is equal to 3°. Note also that the true DOA trajectories are well approximated by the upper edge of the bands in Figs 12(b)-(d) corresponding to the increasing DOA and by the lower edge...
of the bands associated with the decreasing DOA - this behavior, which was expected in view of the above discussion, is in fact quite stable with respect to the SNR (in simulations not shown here, we have observed that decreasing the SNR from 20 dB to 0 dB caused only 1 DOA estimate, out of 2000, to lie outside bands similar to those in Figs 12(b)-(d)). To provide further insight into this type of behavior, a single source moving along a sinusoidal trajectory was considered. Fig 13 shows the DOA estimates for this source obtained with SPICE+. It is clear from this plot that the width of the band decreases till $t = 500$, where it is nearly zero, and then starts increasing again; hence, as expected, the width of the band is proportional to the slope of the DOA variation. Furthermore in this case, too, the true DOA trajectory is well approximated by the upper edge of the band (for increasing DOA) and the lower edge (for decreasing DOA).

![Graph showing distance between elements](image)

Fig. 1. Inter-element spacings (in units of $\lambda/2$) for the NULA with $N = 100$.

**APPENDIX A: SOME SDP AND SOCP FORMULATIONS**

First we show that the problem of minimizing (18), s.t. $p_k \geq 0$ ($k = 1, \cdots, K + N$), can be formulated as an SDP. The proof that the same is true for (20) is similar and therefore its details are omitted. Let

$$
\hat{R}^{1/2} = \begin{bmatrix}
r_1^1 \\
\vdots \\
r_N^1 
\end{bmatrix}
$$

(55)

and let

$$
v_k = (a_k^* \hat{R}^{-1} a_k)
$$

(56)
Fig. 2. The ULA example. Histograms of DOA estimates obtained with PER at two SNR values in 100 Monte Carlo runs: (a)-(b) uncorrelated sources and (c)-(d) coherent sources. The dashed lines show the true DOAs.

Fig. 3. The ULA example. Histograms of DOA estimates obtained with IAA at two SNR values in 100 Monte Carlo runs: (a)-(b) uncorrelated sources and (c)-(d) coherent sources. The dashed lines show the true DOAs.
Fig. 4. The ULA example. Histograms of DOA estimates obtained with MUSIC at two SNR values in 100 Monte Carlo runs: (a)-(b) uncorrelated sources and (c)-(d) coherent sources. The dashed lines show the true DOAs.

Fig. 5. The ULA example. Histograms of DOA estimates obtained with SPICE at two SNR values in 100 Monte Carlo runs: (a)-(b) uncorrelated sources and (c)-(d) coherent sources. The dashed lines show the true DOAs.
Fig. 6. The ULA example. Histograms of DOA estimates obtained with SPICE+ at two SNR values in 100 Monte Carlo runs: (a)-(b) uncorrelated sources and (c)-(d) coherent sources. The dashed lines show the true DOAs.

Fig. 7. The NULA example. Histograms of DOA estimates obtained with PER at two SNR values in 100 Monte Carlo runs: (a)-(b) uncorrelated sources and (c)-(d) coherent sources. The dashed lines show the true DOAs.
Fig. 8. The NULA example. Histograms of DOA estimates obtained with IAA at two SNR values in 100 Monte Carlo runs: (a)-(b) uncorrelated sources and (c)-(d) coherent sources. The dashed lines show the true DOAs.

Fig. 9. The NULA example. Histograms of DOA estimates obtained with MUSIC at two SNR values in 100 Monte Carlo runs: (a)-(b) uncorrelated sources and (c)-(d) coherent sources. The dashed lines show the true DOAs.
Fig. 10. The NULA example. Histograms of DOA estimates obtained with SPICE at two SNR values in 100 Monte Carlo runs: (a)-(b) uncorrelated sources and (c)-(d) coherent sources. The dashed lines show the true DOAs.

Fig. 11. The NULA example. Histograms of DOA estimates obtained with SPICE+ at two SNR values in 100 Monte Carlo runs: (a)-(b) uncorrelated sources and (c)-(d) coherent sources. The dashed lines show the true DOAs.
Fig. 12. Tracking performance of PER, SPICE and SPICE+. The solid lines denote the trajectories of the true DOAs.

Fig. 13. Tracking performance of SPICE+ for a sinusoidal trajectory. The solid line denotes the true DOA trajectory.
Using this notation we can re-write the function $g$ in (18) as:
\[
g = \sum_{k=1}^{N} r_k^* R^{-1} r_k + \sum_{k=1}^{K+N} v_k p_k
\]
(57)

Let $\{\alpha_k\}_{k=1}^{N}$ be auxiliary variables satisfying $\alpha_k \geq r_k^* R^{-1} r_k$, or equivalently
\[
\begin{bmatrix}
\alpha_k \\
r_k
\end{bmatrix} \geq 0
\]
(58)

Then the minimization problem under discussion can be stated as:
\[
\min_{\{\alpha_k\},\{p_k\}} \sum_{k=1}^{N} \alpha_k + \sum_{k=1}^{K+N} v_k p_k \quad \text{s.t.} \quad p_k \geq 0 \quad k = 1, \cdots, K + N
\]
\[
\begin{bmatrix}
\alpha_k \\
r_k
\end{bmatrix} \geq 0 \quad k = 1, \cdots, N
\]
(59)

which is an SDP (as $R$ is a linear function of $\{p_k\}$) [4] [9].

Next, we show that (48) can be cast as an SOCP. Let $\{\beta_k\}$ denote $(K + N)$ auxiliary variables that satisfy:
\[
\beta_k \geq w_k^{1/2} \|c_k\| \quad k = 1, \cdots, K + N
\]
(60)

Then simply observe that (48) can be re-written as:
\[
\min_{\{\beta_k\},\{C\}} \sum_{k=1}^{K+N} \beta_k \quad \text{s.t.} \quad \|c_k\| \leq \beta_k / w_k^{1/2} \quad k = 1, \cdots, K + N
\]
\[
A^* C = \hat{R}^{1/2}
\]
(61)

which is an SOCP [4].

Finally, we show how to re-formulate (53) as an SOCP. Let $\{\beta_k\}$ still denote some auxiliary variables (but now only $K + 1$ of them) and constrain them to satisfy:
\[
\beta_k \geq w_k^{1/2} \|c_k\| \quad k = 1, \cdots, K
\]
\[
\beta_{K+1} \geq \gamma^{1/2} \left[ \sum_{k=1}^{K+N} \|c_k\|^2 \right]^{1/2}
\]
(62)

Using $\{\beta_k\}$ we can re-state (53) in the following form:
\[
\min_{\{\beta_k\},\{C\}} \sum_{k=1}^{K+1} \beta_k \quad \text{s.t.} \quad \|c_k\| \leq \beta_k / w_k^{1/2} \quad k = 1, \cdots, K
\]
\[
\begin{bmatrix}
c_{K+1} \\
\vdots \\
c_{K+N}
\end{bmatrix} \leq \beta_{K+1} / \gamma^{1/2}
\]
\[
A^* C = \hat{R}^{1/2}
\]
(63)

which, once again, is an SOCP.
APPENDIX B : EQUIVALENCE OF (18) AND (20)

Let $P_0$ and $P_1$ denote the minimization problems corresponding to (18) and (20). Therefore:

$$P_0 : \min_{\{p_k \geq 0\}} \text{tr}(R^{-1} \hat{R}) + \text{tr}(\hat{R}^{-1} R)$$

and

$$P_1 : \min_{\{p_k \geq 0\}} \text{tr}(R^{-1} \hat{R}) \quad \text{s.t.} \quad \text{tr}(\hat{R}^{-1} R) = 1$$

Note that we have replaced $N$ by 1 in the right-hand side of the constraint in (65) for notational convenience (this replacement has only a scaling effect on the solution of $P_1$). For completeness, we will also consider the following problem obtained by constraining the first term in (64) to one:

$$P_2 : \min_{\{p_k \geq 0\}} \text{tr}(\hat{R}^{-1} R) \quad \text{s.t.} \quad \text{tr}(\hat{R}^{-1} \hat{R}) = 1$$

First we prove the equivalence of $P_1$ and $P_2$. By making use of an auxiliary variable $\alpha$ we can re-write (65) and (66) as follows:

$$P_1 : \min_{\alpha, \{p_k \geq 0\}} \alpha \quad \text{s.t.} \quad \text{tr}(R^{-1} \hat{R}) = \alpha \quad \text{and} \quad \text{tr}(\hat{R}^{-1} R) = 1$$

$$P_2 : \min_{\alpha, \{p_k \geq 0\}} \alpha \quad \text{s.t.} \quad \text{tr}(\hat{R}^{-1} \hat{R}) = \alpha \quad \text{and} \quad \text{tr}(R^{-1} \hat{R}) = 1$$

Let $\{\tilde{p}_k = \alpha p_k\}$ (note that $\alpha > 0$) and reformulate $P_1$ as:

$$P_1 : \min_{\alpha, \{\tilde{p}_k \geq 0\}} \alpha \quad \text{s.t.} \quad \text{tr}(\hat{R}^{-1} \hat{R}) = \alpha \quad \text{and} \quad \text{tr}(\hat{R}^{-1} \hat{R}) = 1$$

where $\hat{R}$ is made from $\{\tilde{p}_k\}$. It follows from (68) and (69) that if $\{p_k\}$ is the solution to $P_2$ then $\{p_k/\alpha\}$ is the solution to $P_1$, and thus the proof of the equivalence between $P_1$ and $P_2$ is concluded.

Next, we consider $P_0$ and $P_2$. By writing down the Karush-Kuhn-Tucker conditions for $P_0$ it can be shown that the two terms of the objective in (64) must be equal to one another at the minimizing $\{p_k\}$. A more direct proof of this fact runs as follows. Assume that $\{p_k\}$ is the optimal solution of $P_0$ and that at $\{p_k\}$ we have:

$$\text{tr}(R^{-1} \hat{R}) = \rho^2 \text{tr}(\hat{R}^{-1} R)$$

for some $\rho^2 \neq 1$; note that the scaling factor in (70) must necessarily be positive, which is why we wrote it as $\rho^2$; below we will let $\rho > 0$ denote the square root of $\rho^2$. We will show in the following that the previous assumption leads to a contradiction: $\{p_k\}$ cannot be the solution to $P_0$ if $\rho \neq 1$. To do so let $\{\tilde{p}_k = \rho p_k\}$ and observe that, at $\{\tilde{p}_k\}$, the two terms in the objective of $P_0$ are identical:

$$\text{tr}(\hat{R}^{-1} \hat{R}) = \text{tr}(\hat{R}^{-1} \hat{R})$$

Making use of (70) and (71) along with the assumption that $\{p_k\}$ is the solution to $P_0$, we obtain the following inequality:

$$\text{tr}(\hat{R}^{-1} R)(1 + \rho^2) < 2\text{tr}(\hat{R}^{-1} \hat{R}) = 2\rho \text{tr}(\hat{R}^{-1} R)$$

August 9, 2010
or, equivalently,
\[(1 - \rho)^2 < 0\] (73)
which cannot hold, and hence the solution to $P_0$ must satisfy (70) with $\rho = 1$. Using this fact we can re-formulate $P_0$ as follows:

\[
P_0 : \min_{\alpha, \{\rho_k \geq 0\}} \alpha \quad \text{s.t.} \quad \text{tr}(R^{-1} \hat{R}) = \alpha \quad \text{and} \quad \text{tr}(\hat{R}^{-1} R) = \alpha
\] (74)

or, equivalently (as $\alpha > 0$),

\[
P_0 : \min_{\beta, \{\rho_k \geq 0\}} \beta \quad \text{s.t.} \quad \text{tr}(\hat{R}^{-1} \tilde{R}) = \beta \quad \text{and} \quad \text{tr}(\tilde{R}^{-1} \hat{R}) = 1
\] (75)

where $\{\tilde{p}_k = \alpha \rho_k\}$, as before, and $\beta = \alpha^2$ is a new auxiliary variable. Comparing (75) and (68) concludes the proof of the equivalence of $P_0$ and $P_2$.

REFERENCES