16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

SPATIAL POWER SPECTRUM ESTIMATION USING THE PISARENKO FRAMEWORK

Petre Stoica† *Jian Li*‡ *Xing Tan*‡

† Department of Information Technology,
Uppsala University, Uppsala, Sweden
email: ps@it.uu.se

‡ Department of Electrical and Computer Engineering,
University of Florida, Gainesville, FL, USA
email: li@dsp.ufl.edu, tanxing@ufl.edu

ABSTRACT

This paper makes use of the Pisarenko framework, originally devised for temporal power spectrum estimation, to introduce a method for spatial power estimation that outperforms the beamforming method (except in extreme cases with serious calibration errors) as well as the Capon method (except in idealized situations with plentiful data and no miscalibration). An important feature of the proposed method is that it is user parameter-free, unlike most previous proposals with a similar character.

1. INTRODUCTION AND PRELIMINARIES

1.1 Data Model

In this paper, we focus on the spatial power spectrum estimation problem as it appears in array processing applications. The framework we will use to attack this problem is the one introduced in [1]. The said framework was devised to deal with the problem of temporal power spectrum estimation, which is related to but not identical with the problem of interest here. In this sub-section, we discuss briefly the differences between these two problems.

Let $\{x(n)\}$ be a stationary sequence whose spectrum is denoted by $\Phi(\omega)$, and let $\mathbf{R}$ be the following $M \times M$ covariance matrix associated with $\{x(n)\}$:

$$\mathbf{R}=E\left\{\begin{bmatrix} x(n-1) \\ \vdots \\ x(n-M)\end{bmatrix}\begin{bmatrix} x^*(n-1) & \cdots & x^*(n-M)\end{bmatrix}\right\}, \tag{1}$$

where $(\cdot)^*$ denotes the conjugate transpose (or the complex conjugate for scalar variables). It is well-known (see, e.g., [2]) that, as M increases, the eigenvalues of $\mathbf{R}$ approach the set $\left\{\Phi\left(\frac{2\pi}{M}m\right)\right\}_{m=1}^{M}$ and the eigenvectors of $\mathbf{R}$ approach the set of Fourier vectors $\left\{\boldsymbol{\beta}\left(\frac{2\pi}{M}m\right)\right\}_{m=1}^{M}$, where

$$\boldsymbol{\beta}(\omega) = \frac{1}{M^{1/2}}\begin{bmatrix} e^{j\omega} & \cdots & e^{jM\omega}\end{bmatrix}^T. \tag{2}$$

This important result implies that estimating the spectrum $\Phi(\omega)$ can be reduced to the problem of estimating the eigenvalues of $\mathbf{R}$, for a sufficiently large value of M. Note that the eigenvalues must be determined in the correct order, i.e., they must be associated with the corresponding values of the frequency $\left\{\frac{2\pi}{M}m\right\}_{m=1}^{M}$.

The above simple idea lies at the basis of the Pisarenko approach. In a nutshell, this approach can be described as follows. Let $\hat{\mathbf{R}}$ denote a sample estimate of $\mathbf{R}$, and let

$$\hat{\mathbf{R}} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^*; \quad \boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_M\end{bmatrix}; \quad \mathbf{U}^*\mathbf{U} = \mathbf{I} \tag{3}$$

denote the eigenvalue decomposition (EVD) of $\hat{\mathbf{R}}$. Also, let $h(z)$ be a continuous function defined on $(0,\infty)$, and let $g(h)$ denote the inverse function: $g[h(z)] = z$. Assume that $\hat{\mathbf{R}} > 0$, i.e., $\hat{\mathbf{R}}$ is a positive definite matrix, which means that $\{\lambda_m > 0\}_{m=1}^{M}$. The matrix function $h(\hat{\mathbf{R}})$ is defined as:

$$h(\hat{\mathbf{R}}) = \mathbf{U}\begin{bmatrix} h(\lambda_1) & & 0 \\ & \ddots & \\ 0 & & h(\lambda_M)\end{bmatrix}\mathbf{U}^*. \tag{4}$$

Then, with the discussion in the previous paragraph in mind, it is clear that the spectrum can be estimated in the following way:

$$\hat{\Phi}(\omega) = g\left[\boldsymbol{\beta}^*(\omega)h(\hat{\mathbf{R}})\boldsymbol{\beta}(\omega)\right]. \tag{5}$$

This is the class of spectrum estimates introduced in [1]; this class includes the well-known methods of (averaged) periodogram and of Capon as particular cases. We refer to [1], as well as to [3], for detailed discussions on the statistical properties of (5).

Next we discuss the spatial power spectrum estimation problem, which is the one of main interest in this paper. Under certain assumptions, indicated below, this problem is similar to that of temporal spectrum estimation. Let $\{\mathbf{y}(n)\}$ denote the output sequence of an array of M narrowband sensors, and let $\mathbf{a}(\theta)$ be the known steering vector of the array, where θ denotes the direction-of-arrival (DOA) variable. Also, let $\mathbf{R}$ denote the covariance matrix of $\mathbf{y}(n)$,

$$\mathbf{R} = E[\mathbf{y}(n)\mathbf{y}^*(n)]. \tag{6}$$

Under the assumption that the signals impinging on the array are uncorrelated with one another, $\mathbf{R}$ can be written as:

$$\mathbf{R} = \sigma(\theta_1)\mathbf{a}(\theta_1)\mathbf{a}^*(\theta_1) + \sigma(\theta_2)\mathbf{a}(\theta_2)\mathbf{a}^*(\theta_2) + \cdots, \tag{7}$$

where $\{\theta_k\}$ are the DOA's of the said signals, and $\{\sigma(\theta_k)\}$ are their powers. Under the additional assumption that the DOA's $\{\theta_k\}$ are sufficiently separated from one another so that the vectors $\{\mathbf{a}(\theta_k)\}$ are (nearly) orthogonal to each other, and that $\mathbf{a}(\theta)$ is normalized such that

$$\|\mathbf{a}(\theta)\| = 1, \tag{8}$$

(which is no restriction), where $\|\cdot\|$ denotes the Frobenius norm for matrices and the Euclidean norm for vectors, the right-hand-side of Equation (7) becomes the EVD of $\mathbf{R}$, with $\{\sigma(\theta_k)\}$ being the corresponding eigenvalues and $\{\mathbf{a}(\theta_k)\}$ the associated eigenvectors. Consequently, under the assumptions made, there is a perfect analogy between the spatial and the temporal problems of spectrum estimation, and the power spectrum estimation formula in (5) can be used in a verbatim manner in the spatial case.

In practical array processing problems, the above assumptions are rarely satisfied: the signals are usually at least mildly correlated with one another, and their steering vectors are almost never exactly orthogonal to one another. However, even in such practical cases, the power spectrum estimate under discussion can still be used, albeit in an approximate sense. In the next section, we will provide further support to using (5) in such practical cases, by making use of a *covariance matrix fitting framework*. First, however, we exploit the said framework to derive the well-known methods of beamforming and of Capon for estimating the spatial power spectrum $\sigma(\theta)$, see the next subsections.

To simplify the notation, we omit the argument θ of $\sigma(\theta)$ and of $\mathbf{a}(\theta)$. We also let $\mathbf{R}$ denote either the theoretical covariance matrix or its estimate $\hat{\mathbf{R}}$. Even so, the reader should keep in mind the fact that σ and $\mathbf{a}$ are functions of θ, and that in applications $\mathbf{R}$ is always $\hat{\mathbf{R}}$.

1.2 Beamforming (B)

Invoking the additive decomposition of $\mathbf{R}$ in (7), even if it holds only approximately, we can estimate the spatial power σ by solving the following fitting problem:

$$\min_{\sigma} \|\sigma \mathbf{a}\mathbf{a}^* - \mathbf{R}\|^2. \tag{9}$$

The solution to (9) is readily verified to be the following:

$$\sigma_{\mathrm{B}} = \mathbf{a}^*\mathbf{R}\mathbf{a}, \tag{10}$$

which is, of course, nothing but the beamforming spatial power spectrum estimator. The B estimator is known to suffer from considerable leakage problems, which have a twofold character: i) global leakage that leads to "false alarms" (the power of a strong signal at θ leaks into a DOA region, that can be rather far from θ, which might suggest the presence of "false targets" in that region); and ii) local leakage that causes poor resolution and hence "misses" (the power of two adjacent sources at θ_1 and θ_2 leaks into the DOA region between θ_1 and θ_2 with the result that only one source, in lieu of two, is detected).

1.3 Capon (C)

The Capon spatial power spectrum estimator [4] can be derived in at least six different ways (see [5]). Making use of the same covariance matrix fitting framework as above, we can obtain the Capon estimate of the spatial power by solving the following problem:

$$\min_{\sigma} \left\|(\sigma \mathbf{a}\mathbf{a}^*)^{\dagger} - \mathbf{R}^{-1}\right\|^2, \tag{11}$$

where $(\cdot)^{\dagger}$ denotes the Moore-Penrose pseudo-inverse.

The solution to (11) is easily seen to be given by the Capon formula of spatial power spectrum estimation,

$$\sigma_{\mathrm{C}} = \frac{1}{\mathbf{a}^*\mathbf{R}^{-1}\mathbf{a}}. \tag{12}$$

Given an exact knowledge of $\mathbf{a}$ and an accurate sample covariance matrix, σ_{C} is a more accurate power estimate than σ_{B}. However, in practical cases in which $\mathbf{a}$ is imprecisely known and the sample covariance matrix is a poor estimate of the theoretical covariance matrix, σ_{C} can be much less accurate than σ_{B}.

1.4 Diagonally Loaded Capon (DL)

There have been many, both old and recent, attempts to modify σ_{C} to make it more robust to errors in $\mathbf{a}$ and $\mathbf{R}$, see, e.g., [6]. Most of these attempts take the form of diagonal loading:

$$\sigma_{\mathrm{DL}} = \frac{\mathbf{a}^*(\mathbf{R}+\alpha\mathbf{I})^{-1}\mathbf{R}(\mathbf{R}+\alpha\mathbf{I})^{-1}\mathbf{a}}{[\mathbf{a}^*(\mathbf{R}+\alpha\mathbf{I})^{-1}\mathbf{a}]^2}, \tag{13}$$

or, equivalently,

$$\sigma_{\mathrm{DL}} = \frac{1}{\mathbf{a}^*(\mathbf{R}+\alpha\mathbf{I})^{-1}\mathbf{a}} - \alpha\frac{\mathbf{a}^*(\mathbf{R}+\alpha\mathbf{I})^{-2}\mathbf{a}}{[\mathbf{a}^*(\mathbf{R}+\alpha\mathbf{I})^{-1}\mathbf{a}]^2}. \tag{14}$$

The diagonal loading factor, $\alpha \geq 0$, is sometimes chosen in an ad-hoc manner as a constant (independent of θ) [7]. In more elaborate methods, on the other hand, α is chosen as a function of both θ and the available data (i.e., $\mathbf{R}$) [6], using information on the errors in $\mathbf{a}$ and in $\mathbf{R}$ either explicitly or implicitly. In any case, by varying α in the DL, one can achieve the whole range of performance from B to C: indeed, for $\alpha = 0$ (14) reduces to σ_{C} whereas for $\alpha \to \infty$ (14) tends to σ_{B}.

The problem of DL is that most schemes for choosing α are ad-hoc, and that schemes which attempt selecting α in a more qualified manner require information that might not be available in applications.

1.5 Contributions

The contributions of this paper can be summarized as follows:

- Making use of the covariance matrix fitting framework, we provide a new perspective on the advantages and downsides of the B and C methods.
- We use the Pisarenko framework to discuss a whole class of estimates of σ that includes B and C. We show that DL does not belong to this class, and so we introduce a modified DL (M-DL) that is included in the said class.
- We advocate the use of a method in the aforementioned class that has a "mid-way position" between B and C (see the next section for details), and which we call the mid-way (MW) method. We argue that MW should be more robust than C and that it should suffer from less leakage than B, and this without the need for choosing any user parameter.
- We compare B, C, DL, M-DL and MW in a number of spatial power and signal estimation examples.

2. THE PROPOSED POWER ESTIMATION METHOD (MW)

2.1 Some Insights into the Pisarenko Class of Methods

By direct analogy with (5), we consider the following class of spatial power spectrum estimates:

$$\sigma_h = g\left[\mathbf{a}^* h(\mathbf{R})\mathbf{a}\right], \tag{15}$$

where the subscript h indicates the dependence of (15) on the function h. A desirable property of any power estimator is that it should be well-behaved to the scaling of data: mathematically, this means that if $\mathbf{R}$ in (15) is replaced by $\gamma\mathbf{R}$ (for any $\gamma > 0$) then the result should be $\gamma\sigma_h$. As explained in [1], the only class of functions for which this property holds is given by $h(z) = z^r$ for $r \in \mathbb{R}$. The corresponding power estimate, (15), is:

$$\sigma_r = \begin{cases} [\mathbf{a}^*(\mathbf{R})^r\mathbf{a}]^{1/r}, & \text{for } r \neq 0, \\ \exp\left[\mathbf{a}^* \ln(\mathbf{R})\mathbf{a}\right], & \text{for } r = 0. \end{cases} \tag{16}$$

Let (using $\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^*$, see (3))

$$\mathbf{v} = \mathbf{U}^*\mathbf{a}, \tag{17}$$

and let

$$b_m = |v_m|^2, \quad m = 1, \cdots M \tag{18}$$

denote the squared magnitudes of the elements of $\mathbf{v}$. Note that we have

$$\sum_{m=1}^{M} b_m = \|\mathbf{v}\|^2 = 1, \tag{19}$$

(see (8)). Using $\{b_m\}$ we can rewrite (16) as a *weighted power mean*:

$$\sigma_r = \begin{cases} \left(\sum_{m=1}^{M} b_m \lambda_m^r\right)^{1/r}, & \text{for } r \neq 0, \\ \exp\left[\sum_{m=1}^{M} b_m \ln(\lambda_m)\right] = \prod_{m=1}^{M} \lambda_m^{b_m}, & \text{for } r = 0. \end{cases} \tag{20}$$

An interesting property that follows from this weighted power mean interpretation of the estimator in (16) is that σ_r is a monotonically increasing function of r:

$$\sigma_r \leq \sigma_s, \quad \text{for } r < s, \tag{21}$$

where the equality can hold only under rather restrictive conditions (specifically, all λ_m that correspond to $b_m \neq 0$ must be identical), which rarely hold in practice – consequently, *σ_r is usually a strictly monotonically increasing function of r* [1].

2.2 Further Insights and the Selected Method

While (21) is a neat mathematical result, it does not provide a significant insight into the behavior of σ_r as a function of r. To obtain more insights, we remark on the fact that σ_r can be obtained by solving the following covariance matrix fitting problem (analogous to (9) and (11)):

$$\begin{aligned} &\min_{\sigma} \|\sigma^r \mathbf{a}\mathbf{a}^* - (\mathbf{R})^r\|^2 \\ &= \min_{\sigma} \left[(\sigma^r)^2 - 2\sigma^r \mathbf{a}^*(\mathbf{R})^r \mathbf{a} + \|(\mathbf{R})^r\|^2\right], \quad \text{for } r \neq 0, \end{aligned} \tag{22}$$

and

$$\min_{\sigma} \|\ln(\sigma)\mathbf{a}\mathbf{a}^* - \ln(\mathbf{R})\|^2, \quad \text{for } r = 0, \tag{23}$$

(note that (22) with $r < 0$ should be interpreted similarly to (11)).

If $r > 0$ then, with the interpretation (22) in mind, it is clear that σ_r will estimate well the peak values of the spatial power spectrum, whereas the small values will be estimated poorly due to interference from the large values (that are amplified in $(\mathbf{R})^r$ via raising them to the power r). In other words, the estimate σ_r with $r > 0$ suffers from leakage problems (both global and local), and therefore it has low resolution, which is especially true as r increases.

On the other hand, if $r < 0$ then σ_r will give accurate estimates of the small values of the spatial spectrum (as the inverses of those small values, σ^r, that appear in (22) will take on large values), but comparatively poor estimates of the peak values of the spectrum (whose inverses, appearing in the fitting criterion, take on small values). Consequently, in this case, σ_r will not have the problem of leakage from peaks to valleys, as in the case of $r > 0$ (in particular, σ_r with $r < 0$ will have better resolution than that for $r > 0$), but the peak values of the spatial spectrum will be poorly estimated – more specifically, underestimated, see (21) – especially so as r decreases.

In the discussion in the two paragraphs above, we recognize the drawbacks and advantages of the B and C methods, respectively, as explained in Section 1. In fact, it is obvious that

$$\sigma_1 = \sigma_{\mathrm{B}}, \quad \text{and} \quad \sigma_{-1} = \sigma_{\mathrm{C}}. \tag{24}$$

Making use of the interpretation (20) of σ_r we can say, based on (24), that σ_{B} is equal to the *weighted arithmetic mean* of the eigenvalues of $\mathbf{R}$, whereas σ_{C} is equal to the *weighted harmonic mean* of the said eigenvalues.

The previous discussion also suggests a possible way to cope with the leakage problem of σ_{B} and the peak underestimation problem of σ_{C}. This way consists of using σ_0, which we will denote by σ_{MW} because it corresponds to $r = 0$ that lies mid-way between $r = -1$ leading to σ_{C} and $r = 1$ yielding σ_{B},

$$\sigma_{\mathrm{MW}} = \exp\left[\mathbf{a}^* \ln(\mathbf{R})\mathbf{a}\right]. \tag{25}$$

Note that σ_{MW} is the solution to the fitting problem in (23). In view of (20), σ_{MW} is equal to the *weighted geometric mean* of the eigenvalues of $\mathbf{R}$.

Owing to its mid-way position between σ_{B} and σ_{C}, one can argue that σ_{MW} should suffer less from the leakage problem of σ_{B} and should be affected less by the peak power underestimation problem of σ_{C}. Indeed, the logarithmic transformation in (25) can be viewed as a "smoothing operation" whose main effect is to *reduce the dynamic range of the spectrum* such that the estimation of small (large) spectral values is not significantly affected by the existence of large (small) spectral values, as is in B (C) (see also [10]).

Remark: Reduction of the dynamic range of the (eigen)spectrum is also the main objective of the so-called shrinkage methods, of which DL is just a special case. These methods, either indirectly (as in DL) or directly (operating right on the eigenvalues of $\mathbf{R}$) attempt to reduce the spreading of the eigenvalues of this matrix, which is particularly useful in the small-sample support case when $\mathbf{R}$ can be rather ill-conditioned [8, 9]. We can see from (25) that MW operates on the eigenvalues of $\mathbf{R}$ and that it uses $\ln(\mathbf{R})$ in lieu of $\mathbf{R}$, which could be interpreted as a direct shrinkage operation. However, unlike a typical shrinkage method, which would use the shrunken $\mathbf{R}$ (e.g., $\ln(\mathbf{R})$) directly in a power estimate (e.g., in B, which would result in $\mathbf{a}^* \ln(\mathbf{R})\mathbf{a}$), MW tries to compensate for the shrinkage operation on $\mathbf{R}$ by using the inverse function associated with that used to "shrink" $\mathbf{R}$ (see, e.g., (25)). This is an important distinction between the method proposed here and the shrinkage methods, which will surface again when we will shortly discuss DL and its modified version advocated in this paper. □

The reader might think that σ_r with a properly selected $r \neq 0$ could be a better power estimate than σ_0 (i.e., σ_{MW}). To defend MW, first note that σ_0 does not require the choice of any user parameter, which is a desirable feature. Secondly, if we attempted choosing r in σ_r to outperform σ_0, then we would run immediately into problems: indeed, it is apparent from the previous discussion that for peak values of the spectrum r should be chosen equal to 1 or larger, whereas for valley values r should be selected equal to -1 or smaller; and perhaps between -1 and 1 for intermediate values of the spectrum. However, selecting r in such a way would require information on the true spectrum that is usually unavailable. This discussion also suggests that in general a single value of r can hardly be optimal for the estimation of an entire spectrum with a relatively large dynamic range. However, unless information on the true spectrum was available, we have little choice but choosing a fixed value of r. In such cases, $r = 0$ appears to be a reasonable choice if we expect "medium" errors in $\mathbf{a}$ and $\mathbf{R}$; if the said errors are expected to be "rather large", then we should use $r = 1$ (i.e., B), whereas we may use $r = -1$ (i.e., C) if we expect the errors in $\mathbf{a}$ and $\mathbf{R}$ to be "small"; finally, when we are uncertain about the size of these errors, then once again we may use $r = 0$ (i.e., MW).

2.3 A Modified DL (M-DL)

Apparently the DL method, see (14), does not belong to the class of power estimates defined in (15). However, we can readily obtain a modified DL (M-DL) power estimator that is a member of the said class. To do so, let

$$h(z) = (z+\alpha)^{-1}, \tag{26}$$

for which the inverse function is given by

$$g(h) = \frac{1}{h} - \alpha. \tag{27}$$

Using (26) and (27) in (15) leads to the M-DL power estimator:

$$\sigma_{\text{M-DL}} = \frac{1}{\mathbf{a}^*(\mathbf{R}+\alpha\mathbf{I})^{-1}\mathbf{a}} - \alpha. \tag{28}$$

Note that α in the above equation may depend on the DOA; therefore the second term in (28) may mean more than just a simple vertical translation of the spectral estimate. Next, we remark on the fact that

$$\sigma_{\text{M-DL}} \geq \sigma_{\text{DL}}. \tag{29}$$

Indeed, by the Cauchy-Schwartz inequality,

$$\left[\mathbf{a}^*(\mathbf{R}+\alpha\mathbf{I})^{-1}\mathbf{a}\right]^2 \leq \mathbf{a}^*(\mathbf{R}+\alpha\mathbf{I})^{-2}\mathbf{a}, \tag{30}$$

which shows that the second term in (14) (the minus sign omitted) is larger than the second term in (28), and this observation concludes the proof of (29). In particular, (29) implies that $\sigma_{\text{M-DL}} \geq 0$, as it should.

Finally, we remark on the fact that the transformation of $\mathbf{R}$ used in both M-DL and DL (i.e., $\mathbf{R} \to \mathbf{R} + \alpha\mathbf{I}$) can be interpreted as a shrinkage operation that, similarly to the transformation used in MW, attempts to reduce the dynamic range of the eigenvalues of $\mathbf{R}$. However, an important distinction between M-DL and DL is that M-DL compensates for this transformation when estimating the spectrum, whereas DL does not (like most other shrinkage methods). At any rate, it is the said transformation that confers to both methods an "intermediate" position between B and C (again, like for MW). Concerning this aspect, as we have already mentioned, $\sigma_{\text{DL}} = \sigma_{\text{C}}$ for $\alpha = 0$ and $\sigma_{\text{DL}} \to \sigma_{\text{B}}$ for $\alpha \to \infty$. As somewhat expected, the same is true for M-DL. To show this, rewrite (28) as:

$$\sigma_{\text{M-DL}} = \frac{\mathbf{a}^*\left[\mathbf{I} - (\mathbf{R}+\alpha\mathbf{I})^{-1}\alpha\right]\mathbf{a}}{\mathbf{a}^*(\mathbf{R}+\alpha\mathbf{I})^{-1}\mathbf{a}} = \frac{\mathbf{a}^*(\mathbf{R}+\alpha\mathbf{I})^{-1}\mathbf{R}\mathbf{a}}{\mathbf{a}^*(\mathbf{R}+\alpha\mathbf{I})^{-1}\mathbf{a}}, \tag{31}$$

which clearly has the stated behavior.

3. NUMERICAL EXAMPLES

In this section, we provide several numerical examples that compare the performances of MW, DL, M-DL, B, and C. The diagonal loading level used in DL and M-DL is equal to ten times the noise power (the noise power is assumed to be known to DL and M-DL).

Consider a uniform linear array with $M = 10$ sensors and half-wavelength inter-element spacing. The additive noise is assumed to be a spatially and temporally white complex Gaussian random process with zero-mean and unit variance (0 dB). In addition to the signal of interest (SOI) impinging on the array from 0° relative to the array normal, we assume that there are two strong interferences at 12° and 60°. The power of the SOI is 10 dB and each of the two interferences has a power of 20 dB. The SOI and the interferences are temporally white sequences that are independent of one another.

We consider the effect of both the number of snapshots N and the steering vector errors on the SOI power estimate, as well as on the spatial power spectrum estimate. 300 Monte-Carlo trials are carried out to calculate the average SOI power and the average spatial power spectrum. The array steering vector errors for the SOI and for the two interferences are modelled as independent additive white complex Gaussian random vectors with zero-mean and covariance matrix equal to $\eta\mathbf{I}$. The amount of array steering vector error is expressed as $10\log_{10}\eta$ dB. The error-affected steering vectors are normalized such that $\|\tilde{\mathbf{a}}\|^2 = M$.

Figure 1 shows the average of the SOI power estimates as a function of N in the absence of array steering vector errors. Note from Figure 1(a) that MW is clearly a midway approach between B and C. From Figures 1(b), it can be observed that MW performs similarly to DL and M-DL.

Figure 2 shows the average of the SOI power estimates as a function of the steering vector error for $N = 100$. From Figures 2(a), we can see that B and MW are much more robust than C to steering vector errors. Compared with DL and M-DL (we remind the reader that we use idealized versions of these methods based on the true noise variance) MW is better for all values of N and η tested.

Figure 3 shows the average of the spatial power spectrum estimates for $N = 100$ in the absence of array steering vector errors. From Figures 3(a), we note that if the peaks of the power spectrum of B are used to determine the source locations, there will be a significant bias for the SOI location estimate. C has sharp peaks around the true source locations. MW is again clearly a midway approach between B and C.

Figure 4 shows the average of the spatial power spectrum estimates for $N = 100$ when the array steering vector error, η, is -10 dB. MW clearly gives the best spatial power spectrum estimate.

4. CONCLUSIONS

The main findings of this paper can be summarized as follows:

i) The Pisarenko framework for temporal power spectrum estimation can be extended to the spatial power spectrum estimation problem as it appears in array processing applications. Compared with the temporal case, this extension involves a certain approximation as well as a scale correction. For easy reference, let $\mathcal{C}$ denote the extended class of spatial power estimation methods.

ii) The methods in $\mathcal{C}$ can be obtained as solutions of a properly defined covariance matrix fitting problem. This interpretation provides a clear intuitive explanation of the typical performance that can be expected from these methods. In particular, the advantages and drawbacks of the beamforming method (B) and of the Capon method (C), both of which belong to $\mathcal{C}$, become quite transparent in the covariance matrix fitting framework.

iii) B suffers from serious leakage problems that cause, among others, low resolution and spurious peaks. In idealized scenarios with plentiful data and precisely known steering vectors, C does not suffer of these problems and outperforms B significantly. In more practical cases, however, the opposite is true: C grossly underestimates the power if the number of data samples is small or the steering vector is imprecisely known. A plethora of methods have been proposed to correct this misbehavior of C. Most of these so called robust methods take the form of a diagonally loaded Capon approach (DL). We showed that DL does not belong to $\mathcal{C}$ and introduced a modified DL method that is a member of $\mathcal{C}$.

iv) Let σ_B and σ_C denote the power estimates provided by B and C, respectively. By varying the diagonal loading factor in DL we can achieve all power estimates from σ_C to σ_B ($\sigma_C < \sigma_B$). The problem of DL, however, is that the said loading factor must be chosen by the user and this choice is not an easy one, despite some recent advances made in this direction. With this fact in mind we have introduced a method whose power estimate also lies between σ_C and σ_B, which we called the mid-way (MW) method for reasons explained in the paper: the main feature of MW is that, unlike DL, it is user parameter-free.

v) The numerical examples of the paper have shown that the MW power estimates are much more accurate than the C estimates, and also slightly more accurate than the B estimates for small-to-moderate steering vector errors. The MW power estimates were also shown to be more accurate than the estimates obtained with an idealized version of DL that was fed with the true noise variance. The resolution of the MW spatial spectrum estimate was much better than that of B, but slightly worse than the one of C and DL. Compared with C and DL, however, MW provided much more accurate estimates of the spectral-peak heights.

Acknowledgment

This work was sponsored in part by the Swedish Research Council (VR), the Office of Naval Research under Grant No. N00014-07-1-0193, and the National Science Foundation under Grant No. CCF-0634786 and ECCS-0729727.

REFERENCES

[1] V. F. Pisarenko, "On the estimation of spectra by means of non-linear functions of the covariance matrix," *Geophysical Journal of the Royal Astronomical Society*, vol. 28, pp. 511–531, June 1972.

[2] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*. Berkeley, CA: University of California Press, 1958.

[3] C. Vaidyanathan and K. M. Buckley, "Performance analysis of the enhanced minimum variance spatial spectrum estimator," *IEEE Transactions on Signal Processing*, vol. 46, pp. 2202–2206, August 1998.

[4] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, pp. 1408–1418, August 1969.

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

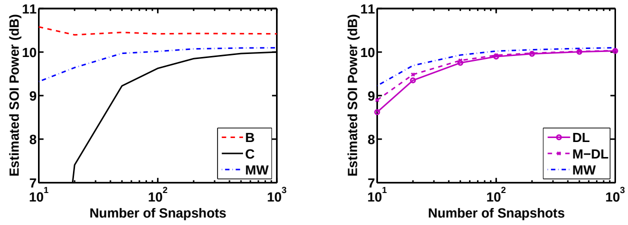
[6] J. Li and P. Stoica, eds., *Robust Adaptive Beamforming*. New York, NY: John Wiley & Sons, 2005.

[7] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, pp. 1365–1376, October 1987.

[8] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, pp. 365–411, 2004.

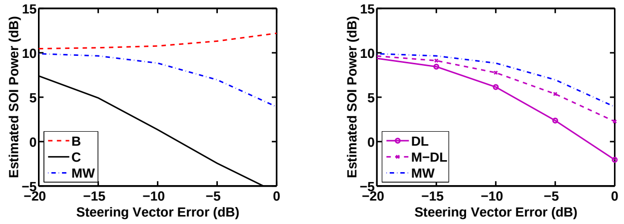
[9] P. Stoica, J. Li, X. Zhu, and J. R. Guerci, "On using a priori knowledge in space-time adaptive processing," to appear in *IEEE Transactions on Signal Processing*, 2008.

[10] M. Lagunas, P. Stoica, and M. Rojas, "ARMA parameter estimation: Revisiting a cepstrum-based method," *The 2008 International Conference on Acoustics, Speech and Signal Processing,* Las Vegas, Nevada, USA, 2008.



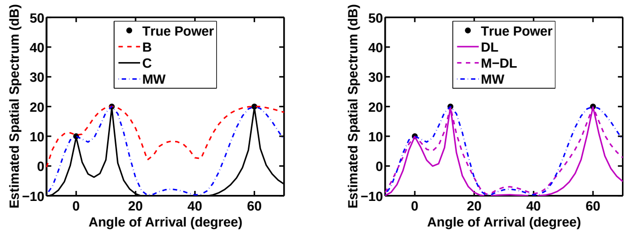
(a) (b)

Figure 1: Average of the SOI power estimates vs. the number of snapshots in the absence of steering vector errors.



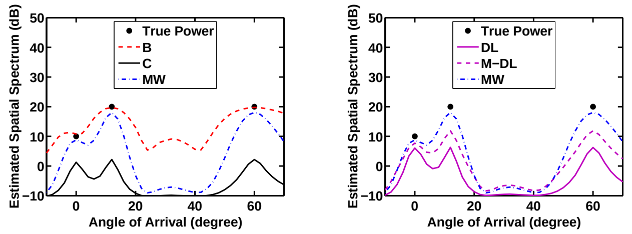
(a) (b)

Figure 2: Average of the SOI power estimates vs. the steering vector error, for $N = 100$.



(a) (b)

Figure 3: Average of the spatial power spectrum estimates in the absence of steering vector errors, for $N = 100$.



(a) (b)

Figure 4: Average of the spatial power spectrum estimates in the presence of steering vector errors, for $N = 100$.