ENHANCED COVARIANCE MATRIX ESTIMATORS IN ADAPTIVE BEAMFORMING

Richard Abrahamsson, Yngve Selén, and Petre Stoica

Uppsala University, Dept. of Information Technology, P.O. Box 337, SE-751 05 Uppsala, Sweden.

ABSTRACT

In this paper a number of covariance matrix estimators suggested in the literature are compared in terms of their performance in the context of array signal processing. More specifically they are applied in adaptive beamforming which is known to be sensitive to errors in the covariance matrix estimate and where often only a limited amount of data is available for estimation. As many covariance matrix estimators have the form of diagonal loading or eigenvalue adjustments of the sample covariance matrix and as they sometimes offer robustness to array imperfections and finite sample error, they are compared to a recent robustified adaptive Capon beamforming (RCB) method which also has a diagonal loading interpretation. Some of the covariance estimators show a significant improvement over the sample covariance matrix and in some cases they match the performance of the RCB even when *a priori* knowledge, which is not available in practice, is used for choosing the user parameter of RCB.

***Index Terms*—** Array signal processing, Covariance matrices, Direction of arrival estimation

1. INTRODUCTION

The Maximum Likelihood (ML) estimate of the covariance matrix under the Gaussian i.i.d. assumption, i.e., the sample covariance matrix, $\hat{\boldsymbol{R}}_{\mathrm{S}} = \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{H}}(t)$, performs poorly when the number of observations, N, is approaching the dimension, M, of the observables, or snapshots, $\boldsymbol{x}(t)$. Considering that the number of unknown parameters grows quadratically in M it is understandable that estimating large covariance matrices from limited data is problematic.

Since the pioneering work in [1], where it was shown that the the sample covariance matrix can be improved upon, numerous enhanced covariance matrix estimators have been proposed in the statistical literature. The various covariance matrix estimators are often derived and compared based on different costs or loss functions. An estimator that is optimal or nearly so for a specific loss is not necessarily well behaved with respect to another loss. Furthermore, the behavior in a specific scenario or in a specific application may be hard to predict from the corresponding loss function used to derive it and assess its performance.

Our interest lies in array signal processing and minimum variance beamforming for signal reception and spatial spectrum estimation. These are applications where the performance is severely affected by errors in the covariance matrix estimate and where data sometimes is hard to obtain in sufficient quantity. Hence, our comparison of the covariance matrix estimates is based on how they behave in these applications.

In this study a multitude of covariance matrix estimators were implemented and tested. However, due to space limitations only a few of the more interesting ones are reported upon here. In Chapter 9 of [2] a more extensive report of our results is given along with a description of all estimators tested. Therein, also a space-time adaptive processing (STAP) [3] detection problem is used to evaluate the covariance matrix estimators.

2. THE CAPON BEAMFORMER AND THE ROBUST CAPON BEAMFORMER

Assume that a narrowband signal of interest impinges on an array of M sensors. The vector of array measurements can be modeled as

$$\boldsymbol{x}(t) = \boldsymbol{a}(\boldsymbol{\theta}_s)s(t) + \boldsymbol{n}(t) \in \mathbb{C}^{M\times 1} \quad \text{for} \quad t = 1,\dots N, \qquad (1)$$

where $\boldsymbol{\theta}_s$ is a parameter vector determining the location of the signal source and $\boldsymbol{a}(\boldsymbol{\theta})$ is the array response for a generic source location, $\boldsymbol{\theta}$. The noise/interference term $\boldsymbol{n}(t)$ is zero mean and temporally white with spatial covariance matrix $\boldsymbol{Q}$. We model the unknown $s(t)$ as a zero mean white random process.

In order to recover the signal of interest the Capon beamformer [4] linearly combines the array output of (1) using a vector of weights, $\boldsymbol{w}^{\mathrm{H}}$, according to

$$\arg\min_{\boldsymbol{w}} \boldsymbol{w}^{\mathrm{H}}\boldsymbol{R}\boldsymbol{w} \quad \text{s.t.} \quad \boldsymbol{w}^{\mathrm{H}}\boldsymbol{a}(\boldsymbol{\theta}_s) = 1 \qquad (2)$$

where ideally $\boldsymbol{R} = \boldsymbol{Q}$. In practice when $\boldsymbol{Q}$ is unknown, $\boldsymbol{R}$ is replaced by an estimate $\hat{\boldsymbol{R}}$ of the full data covariance matrix, usually $\hat{\boldsymbol{R}}_{\mathrm{S}}$. The solution to (2) is readily found (see, e.g., [5]) as

$$\boldsymbol{w} = \frac{\boldsymbol{R}^{-1}\boldsymbol{a}(\boldsymbol{\theta}_s)}{\boldsymbol{a}(\boldsymbol{\theta}_s)^{\mathrm{H}}\boldsymbol{R}^{-1}\boldsymbol{a}(\boldsymbol{\theta}_s)}.$$

When the covariance matrix is inaccurately estimated, such as when the number of data samples available is limited, it is well known that the Capon beamformer has a tendency to cancel the desired signal leading to a suboptimal signal to interference plus noise ratio (SINR) and an underestimated signal power. Signal cancellation is also a problem when the interferences are correlated with the signal of interest and when the steering vector is subject to unknown perturbations.

In this paper we will use the Capon beamformer together with different estimates of the covariance matrix from the literature in order to compare the performances of the latter estimates. We will also compare them to a robustified version of the Capon beamformer (RCB) [6, 7] originally developed for mitigating signal cancellation in the case of an uncertain array steering vector, $\boldsymbol{a}(\boldsymbol{\theta}_s)$. RCB has also showed robustness to errors in the covariance matrix due to limited data. The design criterion for RCB can be written

$$\begin{aligned} \min_{\boldsymbol{w},\boldsymbol{a}_0(\boldsymbol{\theta}_s)} \boldsymbol{w}^{\mathrm{H}}\boldsymbol{R}\boldsymbol{w} \quad &\text{s.t.} \quad |\boldsymbol{w}^{\mathrm{H}}\boldsymbol{a}_o(\boldsymbol{\theta}_s)| \geq 1 \\ \text{and} \quad &\text{s.t.} \quad \|\boldsymbol{a}_0(\boldsymbol{\theta}_s) - \boldsymbol{a}(\boldsymbol{\theta}_s)\|^2 \leq \epsilon \end{aligned} \qquad (3)$$

where $\boldsymbol{a}(\boldsymbol{\theta})$ is the presumed steering vector and $\boldsymbol{a}_0(\boldsymbol{\theta})$ is the true and unknown steering vector, which is assumed to belong to an uncertainty set defined by the inequality constraint in (3). A drawback of the RCB method however is that the size of the uncertainty set ϵ has to be provided by the user.

1-4244-0728-1/07/$20.00 ©2007 IEEE

3. COVARIANCE MATRIX ESTIMATORS

In the literature most estimators of $\boldsymbol{R}$ are derived for real valued data only. In our application, however, the data and the noise processes are complex valued. By stacking the real and imaginary parts of the data and letting $\boldsymbol{R}_r$ denote the covariance matrix of the so-obtained vector

$$\boldsymbol{R}_r \triangleq \mathcal{E}\left\{\begin{bmatrix} \boldsymbol{x}_r(t) \\ \boldsymbol{x}_i(t) \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_r^\top(t) & \boldsymbol{x}_i^\top(t) \end{bmatrix}\right\} \triangleq \begin{bmatrix} \boldsymbol{R}_{rr} & \boldsymbol{R}_{ri} \\ \boldsymbol{R}_{ir} & \boldsymbol{R}_{ii} \end{bmatrix}$$

the complex covariance matrix can be formed according to

$$\boldsymbol{R}_c \triangleq \mathcal{E}\left\{\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{H}}(t)\right\} = \boldsymbol{R}_{rr} + \boldsymbol{R}_{ii} + j\left(\boldsymbol{R}_{ir} - \boldsymbol{R}_{ri}\right). \tag{4}$$

For the sample covariance estimate, finding the estimate of $\boldsymbol{R}_c$ by first estimating $\boldsymbol{R}_r$ this way is equivalent to the direct estimation of $\boldsymbol{R}_c$. Note however that for a general covariance matrix estimator the circular symmetry property of the complex data,

$$\mathcal{E}\left\{\boldsymbol{x}(t)\boldsymbol{x}^\top(t)\right\} \Leftrightarrow \begin{cases} \boldsymbol{R}_{rr} = \boldsymbol{R}_{ii} \\ \boldsymbol{R}_{ri} = -\boldsymbol{R}_{ir} \end{cases}, \tag{5}$$

is not exploited by doing so. For the ML estimate this does not matter. For other estimators, however, this may give a suboptimal estimator when applied to stacked real and imaginary parts of complex circularly symmetric Gaussian data. Moreover, several alternative estimators are not applicable if the number of data is less than the number of elements in each snapshot. Hence, the methodology above is sometimes not applicable for small N. On the other hand it is not clear whether the optimality claimed for the different estimators carries over from real to imaginary data if the estimators are applied naively directly to complex data either. Hence we choose to use the methodology described above rather than apply the methods directly in the naive manner.

In [8] James and Stein considered an estimator of the form

$$\hat{\boldsymbol{R}} = \boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^\top \tag{6}$$

where $\boldsymbol{L}$ is the lower triangular Cholesky factor of

$$\boldsymbol{S} = \sum_{t=1}^{N} \boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{H}}(t)$$

with positive diagonal elements and $\boldsymbol{D}$ is a diagonal matrix with diagonal elements $d_1,\ d_2,\ldots,\ d_M$. The estimator $\hat{\boldsymbol{R}}_{\mathrm{JS}}$ which uses the diagonal elements

$$d_m^{(\mathrm{JS})} = \frac{1}{N+M+1-2m} \tag{7}$$

of $\boldsymbol{D}$ in (6) was shown to give the, on average, best estimator of this form (and also better than the sample covariance matrix) with respect to the *Stein's* loss:

$$L_{\mathrm{St}}(\boldsymbol{R}, \hat{\boldsymbol{R}}) = \mathrm{Tr}\left\{\hat{\boldsymbol{R}}\boldsymbol{R}^{-1}\right\} - \ln\left|\hat{\boldsymbol{R}}\boldsymbol{R}^{-1}\right| - M. \tag{8}$$

Similarly it was shown in [9] that the estimator $\hat{\boldsymbol{R}}_{\mathrm{EO}}$ using

$$d_m^{(\mathrm{EO})} = 2\left(\frac{\Gamma\left(\frac{N-m+1}{2}\right)}{\Gamma\left(\frac{N-m}{2}\right)}\right)^2 \frac{(N-1)^2}{(N+M+1-2m)^2(N-m)^2} \tag{9}$$

where $\Gamma(\cdot)$ is the gamma function, is the best estimator of the form given by (6) with respect to the quadratic loss

$$L_{\mathrm{Q}}(\boldsymbol{R}, \hat{\boldsymbol{R}}) = \mathrm{Tr}\left\{(\hat{\boldsymbol{R}}^{-1}\boldsymbol{R} - \boldsymbol{I})^{\mathrm{H}}(\hat{\boldsymbol{R}}^{-1}\boldsymbol{R} - \boldsymbol{I})\right\}.$$

It is well known that the eigenvalue spread of the sample covariance matrix has a tendency to be larger than that of the true covariance matrix of the underlying data [10], in particular if the number of data is small. According to, e.g., [11], it was suggested in a lecture series by C. Stein to consider a sample covariance estimator with adjusted eigenvalues

$$\hat{\boldsymbol{R}} = \boldsymbol{U}\boldsymbol{\Phi}\left(\boldsymbol{\Lambda}\right)\boldsymbol{U}^{\mathrm{H}} \tag{10}$$

where $\boldsymbol{S} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^{\mathrm{H}}$ is the eigenvalue decomposition of $\boldsymbol{S}$ with the eigenvalues, l_m, in decreasing order on the diagonal of $\boldsymbol{\Lambda}$ and

$$\boldsymbol{\Phi}(\boldsymbol{\Lambda}) = \mathrm{diag}([\phi_1(\boldsymbol{\Lambda}) \quad \phi_2(\boldsymbol{\Lambda}) \quad \cdots \quad \phi_M(\boldsymbol{\Lambda})])$$

for some set of scalar valued functions $\phi_m(\boldsymbol{\Lambda})$. In particular the set of adjusted eigenvalues (or shrunken eigenvalues)

$$\phi_m^{(\mathrm{St})}(\boldsymbol{\Lambda}) = \frac{l_m}{N - M + 1 + 2l_m \sum\limits_{\substack{i=1 \\ i \neq m}}^{M} \frac{1}{l_m - l_j}} \tag{11}$$

was suggested as an approximate minimization of an unbiased estimate of the average of the Stein loss in (8). In order to preserve the order and positiveness of the eigenvalues an *isotonic* regression scheme [12] has to be used. When referring to $\hat{\boldsymbol{R}}_{\mathrm{St}}$ we use the *isotonized* version of (11).

In [10] a shrinkage method was considered of the form

$$\boldsymbol{R} = \alpha\boldsymbol{S} + \beta\boldsymbol{I} \tag{12}$$

which minimizes the mean square error asymptotically (asymptotically in both M and N with $M/N < \infty$). This covariance matrix estimate is found as

$$\hat{\boldsymbol{R}}_{\mathrm{LW}} = (1-\rho)\hat{\boldsymbol{R}}_{\mathrm{S}} + \rho\mu\boldsymbol{I} \tag{13}$$

where

$$\rho = \min\left(\frac{\sum\limits_{t=1}^{N}\left\|\boldsymbol{x}(t)\boldsymbol{x}(t)^\top - \hat{\boldsymbol{R}}_{\mathrm{S}}\right\|_F^2}{N^2\left\|\hat{\boldsymbol{R}}_{\mathrm{S}} - \mu\boldsymbol{I}\right\|_F^2}, 1\right) \tag{14}$$

and μ is the average eigenvalue of $\hat{\boldsymbol{R}}_{\mathrm{S}}$: $\mu \triangleq \mathrm{Tr}\{\hat{\boldsymbol{R}}_{\mathrm{S}}\}/M$.

Using an empirical Bayes (EB) approach an estimator was given in [10] according to

$$\hat{\boldsymbol{R}}_{\mathrm{EB}} = \frac{N}{N+1}\hat{\boldsymbol{R}}_{\mathrm{S}} + \frac{MN-2N-2}{MN^2}\left|\hat{\boldsymbol{R}}_{\mathrm{S}}\right|^{\frac{1}{M}}\boldsymbol{I}. \tag{15}$$

To handle cases when $M > N$ it was suggested that the geometric mean of the eigenvalues of $\hat{\boldsymbol{R}}_{\mathrm{S}}$, $|\hat{\boldsymbol{R}}_{\mathrm{S}}|^{\frac{1}{M}}$, should be replaced by their algebraic mean, $\mathrm{Tr}\{\hat{\boldsymbol{R}}_{\mathrm{S}}\}/M$, when $N < M$.

3.1. Expected Likelihood Estimates

So far, with the exception of the sample covariance matrix, the covariance matrix estimators above were derived for real valued data. In [13] an interesting class of covariance matrix estimators where derived in the complex Gaussian data framework. The basis of this class is the remarkable fact that the probability distribution of the likelihood ratio

$$\mathrm{LR}(\boldsymbol{X}, \boldsymbol{R}) \triangleq \frac{l(\boldsymbol{X}; \boldsymbol{R})}{l(\boldsymbol{X}; \hat{\boldsymbol{R}}_{\mathrm{ML}})} \leq 1 \tag{16}$$

does not depend on $\boldsymbol{R}$ when $\boldsymbol{R}$ attains its true value (in (16) $l(\boldsymbol{X};\boldsymbol{R})$ is the likelihood of observing $\boldsymbol{X}$ if $\boldsymbol{R}$ is the underlying covariance matrix and $\hat{\boldsymbol{R}}_{\mathrm{ML}}$ is the maximum likelihood estimate). It turns out that the distribution of (16) has most of its mass at a significant distance from 1, and especially so when the number of snapshots approaches the dimension of the data vector. Hence by parameterizing the covariance matrix estimate with a parameter β, e.g., using a diagonal loading parameter, and choosing the parameter so that (16) attains its mean value,

$$\mathrm{LR}(\boldsymbol{X};\hat{\boldsymbol{R}}(\beta)) = \mathcal{E}\{\mathrm{LR}(X;\boldsymbol{R})\} \triangleq \mathrm{LR}_0, \tag{17}$$

which only depends on M and N and can be precalculated, we are likely to be closer to the true $\boldsymbol{R}$ than the maximum likelihood estimate. In this paper the two parameterizations suggested in [13] are considered:

$$\hat{\boldsymbol{R}}_{\mathrm{AS}_1} = \hat{\boldsymbol{R}}_{\mathrm{S}} + \beta \boldsymbol{I} \tag{18}$$

and

$$\hat{\boldsymbol{R}}_{\mathrm{AS}_2} = \boldsymbol{U}_1\boldsymbol{\Lambda}_1\boldsymbol{U}_1^{\mathrm{H}} + \bar{\lambda}_2\boldsymbol{U}_2\boldsymbol{U}_2^{\mathrm{H}} \tag{19}$$

where $\boldsymbol{U} = [\boldsymbol{U}_1 \quad \boldsymbol{U}_2]$ is the matrix of eigenvectors of $\hat{\boldsymbol{R}}_{\mathrm{S}}$ partitioned according to the $M-m$ largest eigenvalues and the m smallest eigenvalues, $\boldsymbol{\Lambda}_1$ is the diagonal matrix of the corresponding $M-m$ largest eigenvalues and $\bar{\lambda}_2$ is the average of the m smallest eigenvalues. In the latter case m is the parameter which is chosen so that $\mathrm{LR}(\boldsymbol{X};\hat{\boldsymbol{R}}_{\mathrm{AS}_2}(m))$ is as close as possible to LR_0.

4. NUMERICAL STUDY

In our first example an unperturbed uniform linear array (ULA) [5] of $M = 10$ sensor elements with half wavelength spacing is used. Three signals simulated as temporally white complex Gaussian noise impinge on the array. The signal of interest has a DOA $\theta_s = 20^\circ$ and a power, σ_s^2, 10 dB above the white complex Gaussian sensor noise. The other two signals are mutually independent interferences, uncorrelated with the signal of interest and each with a power 15 dB above the sensor noise. One interference is located at $\theta_{i_1} = -30^\circ$. The other interference has an angular location corresponding to a spatial frequency

$$\omega_{i_2} = \pi\sin(\theta_s) + 2\pi\frac{\gamma}{M}, \tag{20}$$

where $\gamma = 0.9$. Each plot is based on $K = 5000$ Monte-Carlo simulations. For the RCB we use a radius of the uncertainty set $\epsilon = 0.875$ throughout the simulations (see below).

In the first example we study the effects of having few data snapshots. In Figure 1, the mean signal to interference plus noise ratio (SINR) is shown versus N,

$$\text{mean SINR} = \frac{1}{K}\sum_{k=1}^{K}\frac{\sigma_s^2\left|\hat{\boldsymbol{w}}_k^{\mathrm{H}}\boldsymbol{a}(\theta_s)\right|^2}{\hat{\boldsymbol{w}}_k^{\mathrm{H}}\boldsymbol{Q}\hat{\boldsymbol{w}}_k}$$

where $\hat{\boldsymbol{w}}_k$ is the vector of beamformer weights obtained at the k^{th} realization, $\boldsymbol{Q}$ is the noise plus interference covariance matrix, σ_s^2 is the power of the signal of interest and $\boldsymbol{a}(\theta_s)$ is the array steering vector. Capon using the sample covariance matrix suffers significant signal cancellation at low N. The improvement by using the enhanced estimators is evident in this case.

In Figure 2 the root mean square (RMS) error of the spatial frequency estimate, $\hat{\omega}_s = \pi\sin(\hat{\theta}_s)$, versus separation γ (see (20)) is plotted. The number of data snapshots is $N = 22$. The thick near-horizontal dotted line shows when the RMS error corresponds to half

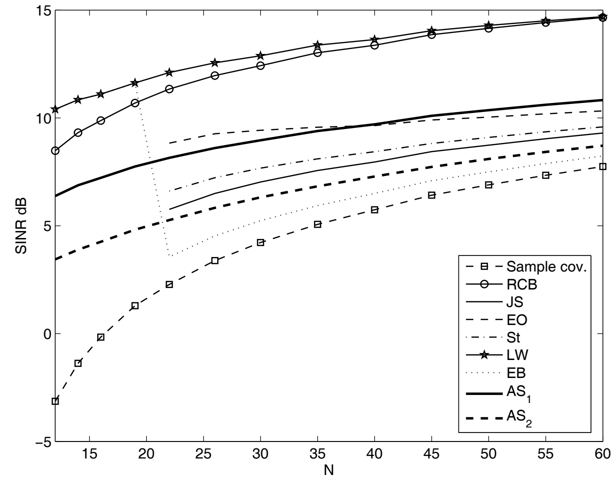

Fig. 1. SINR versus number of snapshots, N.

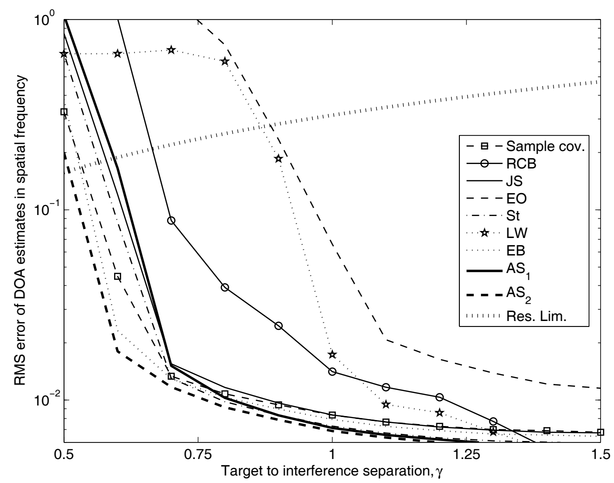

Fig. 2. RMS error of estimated spatial frequency versus separation, γ, between target and the closest interference measured in standard beamformer beamwidth.

the separation between the signal of interest and the closest interference. We see that RCB and the $\hat{\boldsymbol{R}}_{\mathrm{LW}}$ which gave the best SINR performance previously now give the worst resolution of the Capon methods. It is well known that Capon, using the sample covariance estimate, gives accurate location of the peaks despite signal cancellation. Some of the enhanced covariance estimators still tend to give a slight improvement over using the sample covariance matrix.

In the next examples we instead study the effect of calibration errors, using a larger $N = 60$. Calibration errors in the array are simulated as a perturbation $\boldsymbol{\delta}(\theta) \sim \mathcal{CN}\big(\boldsymbol{0},\sigma_{\boldsymbol{\delta}}^2\boldsymbol{I}\big)$ to each antenna element with the true array steering vector, $\boldsymbol{a}_0(\theta) = \boldsymbol{a}(\theta)+\boldsymbol{\delta}(\theta)$, where $\sigma_{\boldsymbol{\delta}}^2 = 0.005$. This corresponds to a standard deviation of the perturbation which is about 7% of the magnitude of each element and 4° of phase. For the RCB this means that $P\big(\|\boldsymbol{a}_0 - \boldsymbol{a}\|^2 < \epsilon = 0.875\big) = 0.98$.

As the target signal power increases, the Capon beamformer spends more of its degrees of freedom to cancel the desired signal which, due to the array imperfections, is not preserved by the unit

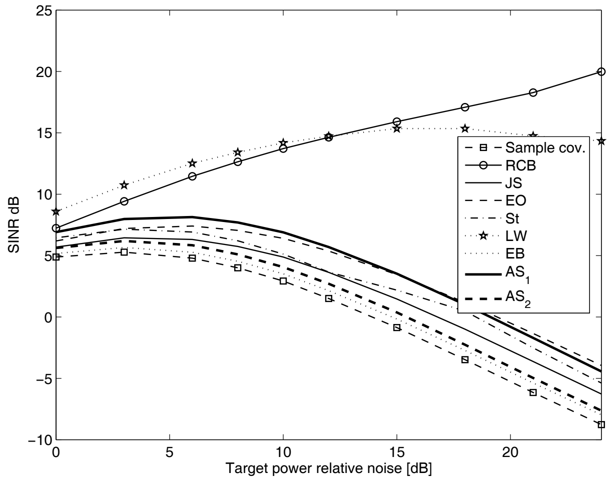

Fig. 3. SINR versus target power relative to noise.

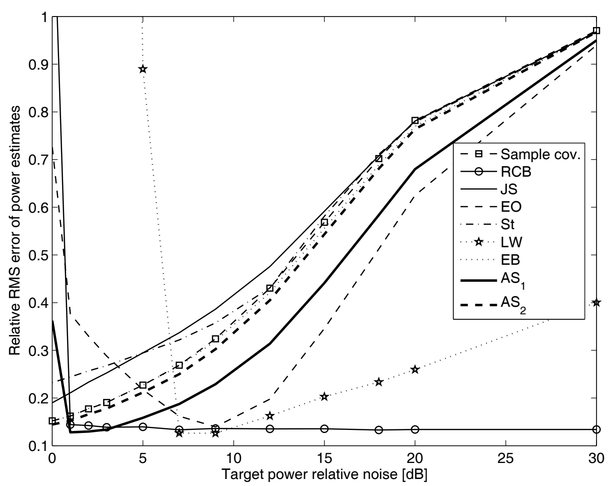

Fig. 4. Relative RMS error of estimated signal power versus target power relative to noise.

gain constraint. This is reflected in the results shown in Figure 3 where the SINR versus power of the impinging signal of interest is plotted and in the last figure (Figure 4) where the RRMSE of the estimated power at the estimated spatial frequency is presented as the signal power is varied.

5. CONCLUDING REMARKS

For the limited data case several of the enhanced covariance matrix estimators improve upon the Capon beamformer. Using $\hat{\boldsymbol{R}}_{\mathrm{LW}}$, which shows the best improvement in SINR for a moderate signal and interference separation, gives a lower resolution in the spatial spectrum, similar to the RCB. The expected likelihood based estimators both offer an improved SINR and a slightly better resolution than the Capon beamformer using the sample covariance matrix.

For the perturbed array case, the enhanced estimators show only a small improvement in SINR and DOA/power estimates with the exception of $\hat{\boldsymbol{R}}_{\mathrm{LW}}$ which appears to give a sufficiently large diagonal load to mitigate signal cancellation at the cost of a lower resolution. The RCB estimates the power of weaker signals better. In this case the size of the uncertainty set, ϵ, has been chosen based on *a priori* information that in practice is not available.

6. REFERENCES

[1] C. Stein, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," in *Proceedings of the 3^{rd} Berkeley Symposium on Mathematical Statistics and probability*, Berkeley, CA, 1956, vol. 1, pp. 197–206, University of California Press.

[2] R. Abrahamsson, *Estimation Problems in Array Signal Processing, System Identification and Radar Imagery*, Ph.D. thesis, Uppsala University, Uppsala, Sweden, October 2006.

[3] J. Ward, "Space-time adaptive processing for airborne radar," Technical Report 1015, Massachusetts Institute of Technology, Lincoln Laboratory, Lexington, MA, December 1994.

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

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

[6] S. A. Vorobyov, A. B. Gershman, and Z. Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: A solution to the signal mismatch problem," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 313–324, February 2003.

[7] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1702–1715, July 2003.

[8] W. James and C. Stein, "Estimation with quadratic loss," in *Proceedings of the 4^{th} Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, CA, 1961, vol. 1, pp. 361–379, University of California Press.

[9] M. L. Eaton and I. Olkin, "Best equivariant estimators of a Cholesky decomposition," *The Annals of Statistics*, vol. 15, no. 4, pp. 1639–1650, 1987.

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

[11] D. K. Dey and C. Srinivasan, "Estimation of a covariance matrix under Stein's loss," *The Annals of Statistics*, vol. 13, no. 4, pp. 1581–1591, 1985.

[12] S. P. Lin and M. D. Perlman, "A Monte-Carlo comparison of four estimators of a covariance matrix," *Multivariate Analysis*, vol. VI, pp. 411–429, 1985.

[13] Y. I. Abramovich and N. K. Spencer, "Expected-likelihood covariance matrix estimation for adaptive detection," in *The 2005 IEEE International Radar Conference*, Arlington, VA, May 2005, pp. 623–628.