

# Adaptive M-Estimators For Robust Covariance Estimation

Christopher L. Brown, Ramon F. Brcich, and Abdelhak M. Zoubir

Signal Processing Group, Institute of Telecommunications  
Darmstadt University of Technology  
Merckstrasse 25, D-64283 Darmstadt, Germany.  
(e-mail: [chris.brown@ieee.org](mailto:chris.brown@ieee.org), [r.brcich@ieee.org](mailto:r.brcich@ieee.org), [zoubir@ieee.org](mailto:zoubir@ieee.org))

**Abstract.** Robust covariance estimates are required in many applications. Here, a promising adaptive robust scale estimator is extended to this problem and compared to other robust estimators. Often the performance analysis of covariance estimators is performed from the perspective of the final application. However, different applications have different requirements, hence we make a comparison based on some general metrics. Results show that the adaptive scheme shows good performance under the nominal case and graceful degradation in performance with increasing levels of contamination.

**Keywords:** robust estimation, covariance, M-estimators.

## 1 Introduction

Numerous problems in signal processing require estimates of covariance. This occurs, e.g., in array processing where the objective is to either detect the number of sources impinging on an array or their directions of arrival (DOA). Unfortunately, the sample covariance estimator has poor performance when there are model deviations or outliers in the observations [Williams and Johnson, 1993].

Robust estimators protect against this, usually for only a small decrease in performance at the nominal model. Robustness is recognised as a favourable property since, in practice, it is more the norm than the exception that such disturbances exist.

Here we concentrate on robust covariance estimation for multi-dimensional observations. In the context of robust estimation, the covariance matrix is also referred to as the association or scatter matrix to allow for nonexistence of the second order moments. Several approaches have been suggested including:

- i*) FLOM (Fractional Lower Order Moment) estimators based on covariation [Shao and Nikias, 1993, Tsakalides and Nikias, 1996, Liu and Mendel, 2001].
- ii*) Nonparametric estimators using signs or ranks [Visuri et al., 2001, Kendall and Gibbons, 1990].

- iii) Expectation maximisation (EM) applied to Gaussian mixture models [Kozick and Sadler, 2000].
- iv) Ellipsoidal trimming [Cook *et al.*, 1993]
- v) Huber's robust M-estimators [Huber, 1981, Williams and Johnson, 1993].

The first two methods are computationally inexpensive, however they may sacrifice too much performance degradation under nominal conditions in order to be robust. The ellipsoidal trimming procedure and the iterative nature of the EM algorithm make their computational complexity an issue when considering implementation. The last is arguably the method of choice but is difficult to use in practice due to the multi-dimensional optimisation it requires.

To avoid these issues, the simplest class of estimators applies a one-dimensional scale estimator to robustly estimate each element of the covariance matrix. The problem then reduces to one of finding robust estimators of scale. To this end, we will investigate a number of robust scale estimators, including an adaptive M-estimator<sup>1</sup> which was shown to improve upon the existing robust estimators of scale [Breich *et al.*, 2004].

This paper is organised as follows: in Section 2 we define the signal model and describe the scale estimators to be used for element-wise covariance matrix estimation. These methods will be compared with the covariance estimators to be described in Section 3. The final application, e.g. DOA, for these covariance estimates will determine the best metric to be used. However, since we do not wish to restrict our study to one application, we must consider general metrics. Hence, for the simulation results shown in Section 4, comparisons are made using a number of metrics. Finally, conclusions are drawn and directions for future work described.

## 2 Robust covariance estimation using scale estimators

One approach to the estimation of covariance matrices is to estimate individual matrix elements using robust estimators of scale. Consider the following signal model

$$\mathbf{x}(n) = \mathbf{A}\mathbf{u}(n), \quad n = 1, \dots, N \quad (1)$$

where  $\mathbf{x}(n) = [x_1(n), x_2(n), \dots, x_M(n)]^T$  is the observation vector,  $\mathbf{u}(n) = [u_1(n), u_2(n), \dots, u_P(n)]^T$  is a vector of independent and identically distributed (iid) standard (zero mean and unit variance) random variables and  $\mathbf{A}$  is the  $M \times P$  mixing matrix. The true covariance matrix is  $\mathbf{C} = \mathbf{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{A}\mathbf{A}^H$  and each matrix element is

$$C(i, k) = \mathbf{E}[x_i x_k^*]. \quad (2)$$

---

<sup>1</sup> In this paper, the term "adaptive" will be used to refer to techniques that are data-dependent, i.e. parameters used in the procedure are set based on the values of the observations

However, rather than simply replacing the expectation operation in (2) with the sample average to estimate matrix elements, a more robust operation is to use [Huber, 1981]

$$\hat{C}_\sigma^*(i, k) = \frac{\hat{\sigma}^2(x_i + x_k) - \hat{\sigma}^2(x_i - x_k)}{\hat{\sigma}^2(x_i + x_k) + \hat{\sigma}^2(x_i - x_k)} \hat{\sigma}(x_i) \hat{\sigma}(x_k) \quad (3)$$

where  $\hat{\sigma}(\cdot)$  is a robust scale estimator. We will now investigate a number of possible robust scale estimators for this procedure.

### 2.1 Sample estimator

The sample estimator of standard deviation is known [Huber, 1981] to have poor resistance to outliers. Despite this, we will include it in this study to provide a frame of reference.

### 2.2 Median absolute deviation

The median absolute deviation (MAD)

$$MAD(\mathbf{x}) = \text{median}(|\mathbf{x} - \text{median}(\mathbf{x})|) \quad (4)$$

has been described as a ‘candidate for being the “most robust estimate of scale” ’ [Huber, 1981]. For symmetric distributions, this is approximately half the interquartile range. Hence, to convert this measure to a true scale estimate, it must be normalised. For nominally Gaussian distributions, a MAD-scale estimator is given by  $\hat{\sigma}_{\text{MAD}}(\mathbf{x}) = \frac{MAD(\mathbf{x})}{\Phi^{-1}(0.75)}$  where  $\Phi^{-1}(\cdot)$  is the inverse Gaussian cdf.

### 2.3 M-estimators of scale

The ML estimate of scale may be found by solving the log-likelihood equation,

$$\sum_{n=1}^N \psi\left(\frac{x^{(n)}}{\sigma}\right) = 0 \quad (5)$$

for  $\sigma$  where

$$\psi(x) = -1 - x \frac{\dot{f}_X(x)}{f_X(x)} \quad (6)$$

is the scale score function associated with the density  $f_X(x)$  and  $\dot{f}_X(x)$  denotes the derivative of  $f_X(x)$ . By contrast, an M-estimator [Huber, 1981] replaces the nominal score function  $\psi(x)$  with a similarly behaved function  $\varphi(x)$  chosen to confer robustness on the estimator under deviations from the

assumed density. With this in mind, Huber proposed that a clipped quadratic score function

$$\varphi_H(x; k) = \min(x^2, k^2) - \delta = \begin{cases} x^2 - \delta, & |x| \leq k \\ k^2 - \delta, & |x| > k, \end{cases} \quad (7)$$

be used in the M-estimator for scale as it minimises the maximum relative asymptotic variance of the scale estimate in the case of a contaminated Gaussian distribution.  $\delta$  is determined such that the estimator is unbiased for the nominal Gaussian distribution. The parameter  $k$  controls the sensitivity of the estimator to the contamination and should decrease as the proportion of outliers increases.

### 2.4 Adaptive M-estimators of scale

One of the drawbacks of the M-estimators described above is that the best value of the cut-off parameter  $k$  is dependent on the degree of contamination [Breich and Zoubir, 2002, Brown *et al.*, 2003]. In [Breich *et al.*, 2004], an adaptive scheme was presented that sought to relieve this restriction. There, the score function is composed of a family of basis functions, the weights of which are chosen adaptively from the data. By using bases that were appropriate for a range of levels of contamination, the adaptive scheme was able to maintain high performance for a wider range of scenarios than the “static” M-estimators. Of course, as well as finite sample performance, the asymptotic performance of the adaptive scheme will also be dependent on selecting appropriate bases that can adequately represent the optimum score function. For a full description of the adaptive algorithm, see [Breich *et al.*, 2004].

## 3 Alternative Robust Covariance Estimators

Together with the element-wise scale estimation based approaches described in the previous section, we will also consider FLOM and sign covariance matrix (SCM) methods.

### 3.1 FLOM based estimator

The use of FLOMs has been shown to have strong motivation and impressive performance when impulsive noise exists [Shao and Nikias, 1993]. They estimate the covariation of  $\alpha$ -stable random processes – analogous to the covariance of Gaussian random variables. Recognising this, FLOM based measures of association were proposed in [Tsakalides and Nikias, 1996] for the purpose of determining DOA. The “covariation” matrices are found by

$$\hat{C}_{\text{FLOM}}(i, k; p) = \frac{\sum_{n=1}^N x_i(n) |x_k(n)|^{p-2} x_k^*(n)}{N} . \quad (8)$$

The parameter  $p$  is the order of the moments. Setting  $p = 2$  reduces (8) to a sample covariance – appropriate under the condition of Gaussianity. However, as contamination occurs, to prevent estimator breakdown,  $p$  should be set to a lower value. The lower the value, the greater the degree of robustness, at the cost of less accuracy under the nominal case.

The form of (8) is very similar to that used in ROC-MUSIC [Liu and Mendel, 2001] differing only by the normalisation factor of each column. Further, for identically distributed observations, this normalisation factor will be approximately equal for all columns. Here, only the FLOM-based method will be considered. To ensure Hermitian matrices, the estimated matrix is averaged with its Hermitian, as in [Tsakalides and Nikias, 1996].

### 3.2 Sign covariance matrix

The SCM was suggested as a robust estimate of covariance in [Visuri *et al.*, 2001]. The concept is to take the sample covariance of some function,  $\tilde{\mathbf{x}} = \mathbf{S}(\mathbf{x})$ , of the multi-variate observations. In [Visuri *et al.*, 2001]  $\mathbf{S}(\cdot)$  was the spatial sign function which normalises each observation to a unit vector. Hence, the spatial sign function can be viewed as the multi-dimensional version of the sign function and from this interpretation the robust behaviour of the SCM is clear. Due to the normalisation of the observations, scale information is lost. However, it was also shown that the subspace estimates from the sample SCM converge to the true subspace.

When more than just a good subspace estimate is required, results in [Visuri *et al.*, 2001] showed that for small samples it is better to whiten the observations using the eigenvectors of the sample SCM and then estimate the eigenvalues as the marginal variances of the transformed observations. To estimate the marginal variances the MAD was used.

## 4 Results

Herein, and without loss of generality, we only consider real random variables. In the results shown here, iid samples of  $\mathbf{u}(t)$  for  $P = 4$  and  $N = 100$  were drawn from the selected distribution. The first six distributions were Gaussian mixtures where the nominal distribution was  $\mathcal{N}(0, 1)$  distribution and the contaminating distribution was  $\mathcal{N}(0, 100)$ . The probability of contamination took values  $\varepsilon = 0, 0.01, 0.02, 0.05, 0.1, 0.2$ . The last two distributions were the  $t_3$  and  $t_4$  distributions respectively.

A number of mixing matrices were considered, however, due to space limitations, results are only shown for two representative cases

$$\tilde{A}_1(i, k) = 0.4^{|i-k|} \text{ and } \tilde{A}_2(i, k) = \begin{cases} 1 & i = k \\ \frac{1}{4} & i \neq k \end{cases} .$$

The mixing matrices are then standardised so that the true covariance matrices have unit diagonals,

$$A(i, k) = \frac{\tilde{A}(i, k)}{\sqrt{\sum_{j=1}^P \tilde{A}^2(i, j)}} \quad . \quad (9)$$

Our objective here is to compare the static and adaptive M-estimators of scale with existing methods for the purposes of covariance matrix estimation. The comparison is not straightforward as it can either depend on the final application, i.e., mean squared error (MSE) of DOA estimates, or on more general metrics, such as the Frobenius norm. The former approach is popular in signal processing, however, good performance in one application does not necessarily imply similar performance in others. Hence, we now study the performance of the estimators using three different metrics: the Frobenius norm, relative MSE of the eigenvalues and the sphericity statistic. Average values for the metrics over 500 Monte Carlo runs were calculated.

**Frobenius norm:** The element-wise sum of squared differences between  $\hat{C}$  and  $C$

$$L_F(\hat{C}, C) = \text{trace}\{(\hat{C} - C)(\hat{C} - C)^H\}. \quad (10)$$

This measures the MSE. Results using the Frobenius norm are shown in Table 1 and Table 2 for  $A_1$  and  $A_2$  mixing matrices respectively for the following methods: sample scale estimator, adaptive scale M-estimator with basis functions  $\varphi_H(x; k)$ ,  $k = 1.5, 2, 2.5$ , static scale M-estimator (Huber) with basis functions  $\varphi_H(x; k)$ ,  $k = 1, 1.5, 2, 2.5$ , MAD, FLOM based estimator with  $p = 1, 1.5, 1.8$  and the SCM in its original (SCM1) and whitened (SCM2) forms.

Estimator	Noise distribution							
	1	2	3	4	5	6	7	8
Sample	0.45	3.4	5.8	15	29	55	6.2	3
Adaptive	0.47	0.6	0.82	1.6	3.8	15	2.5	1.9
Huber 1.0	0.64	0.71	0.8	1.3	2.7	8.8	1.9	1.5
Huber 1.5	0.53	0.64	0.74	1.5	3.9	15	2.3	1.5
Huber 2.0	0.48	0.59	0.79	2.1	6.2	24	2.8	1.8
Huber 2.5	0.46	0.66	1.1	3.4	10	35	3.2	2
MAD	0.72	0.78	0.85	1.2	2.7	7.9	2	1.4
FLOM 1.0	0.61	0.47	0.57	1.5	3.2	6.2	0.82	0.46
FLOM 1.5	0.5	1	1.8	4.5	9	18	2.2	1.2
FLOM 1.8	0.43	1.9	3.8	9.3	18	35	3.9	2
SCM1	2.2	2.2	2.2	2.2	2.2	2.1	2.2	2.2
SCM2	0.73	0.74	0.87	1.5	3.6	14	2.6	1.9

**Table 1.** Estimator performance using the Frobenius norm and the  $A_1$  mixing matrix.

Estimator	Noise distribution							
	1	2	3	4	5	6	7	8
Sample	0.46	3.2	6.2	14	28	53	5.5	2.9
Adaptive	0.47	0.59	0.84	1.8	4.7	16	2.5	1.7
Huber 1.0	0.64	0.69	0.83	1.5	3.3	11	2	1.4
Huber 1.5	0.52	0.61	0.79	1.7	4.5	16	2.2	1.5
Huber 2.0	0.48	0.6	0.91	2.3	7.1	23	2.7	1.8
Huber 2.5	0.44	0.68	1.2	3.6	10	31	3.1	2
MAD	0.72	0.79	0.88	1.4	3.1	9.9	1.9	1.3
FLOM 1.0	0.61	0.48	0.61	1.6	3.4	6.6	0.85	0.49
FLOM 1.5	0.51	1	1.8	4.7	9.2	18	2.2	1.2
FLOM 1.8	0.43	2	3.7	9.4	18	34	3.9	2
SCM1	2.2	2.2	2.1	2.1	2.1	2.1	2.1	2.1
SCM2	0.67	0.75	0.85	1.5	3.6	14	2.6	1.8

**Table 2.** Estimator performance using the Frobenius norm and the  $A_2$  mixing matrix.

Though not shown here, when considering the robust scale estimator based techniques as described in Section 2, investigations showed that the use of (3) did indeed improve robustness considerably. No change was observed in the case of the sample estimator. Therefore, all results shown here for these scale estimator based techniques utilised this procedure.

Inspecting the two tables, similar observations can be made,

- The performance when using the sample scale estimator quickly breaks down with contamination.
- Comparing the adaptive and static M-estimators shows that the adaptive scheme tends to follow the best performance of the static schemes – i.e. for low contamination levels, the adaptive scheme shows similar performance to the static case with high  $k$  and for high contamination levels it is similar to the static estimator with low  $k$ .
- MAD is indeed showing very robust performance, with little deterioration in performance, however, the SCM techniques show themselves to be insensitive to contamination, especially SCM1. In both cases, however, poor performance relative to some of the other techniques is observed in the nominal case (Gaussianity) – as expected of nonparametric techniques.
- Both static M-estimator and FLOM techniques can be “tuned” through parameters  $k$  and  $p$  respectively. For low contamination, high parameter values are best, while for high contamination, low parameter settings are best.

Note that incorporation of additional bases with smaller  $k$  can increase the robustness of the adaptive scheme. This comes at the expense of a slightly

higher computational burden and reduction in performance for the nominal and lightly contaminated cases.

**Relative MSE of eigenvalues:** Let  $\lambda_{i,\hat{C}}, \lambda_{i,C}, i = 1, \dots, M$  be the ordered eigenvalues of  $\hat{C}$  and  $C$ . This metric measures the relative squared difference between the eigenvalues  $\lambda_{i,\hat{C}}$  and  $\lambda_{i,C}$

$$L_E(\hat{C}, C) = \sum_{i=1}^M \left( \frac{\lambda_{i,\hat{C}} - \lambda_{i,C}}{\lambda_{i,C}} \right)^2 . \tag{11}$$

Results are shown in Table 3 and similar qualitative conclusions could be drawn as those from Tables 1 and 2. This confirms that an estimator with good performance in a Frobenius norm sense will also produce good eigenvalue estimates. In particular, relevant to this investigation, it confirms that the adaptive M-estimator scheme exhibits good performance compared to static schemes. However, for high contamination levels, again, the MAD and SCM1 methods gain strong justification for their use.

Estimator	Noise distribution							
	1	2	3	4	5	6	7	8
Sample	0.078	5.5	16	1e+02	4.1e+02	1.5e+03	33	5.2
Adaptive	0.089	0.15	0.33	1.3	7.4	1e+02	3.2	1.7
Huber 1.0	0.2	0.24	0.28	0.81	3.6	36	1.9	1.1
Huber 1.5	0.12	0.17	0.25	1.2	7.5	1e+02	2.6	1.2
Huber 2.0	0.096	0.17	0.34	2.4	20	2.8e+02	3.9	1.6
Huber 2.5	0.086	0.22	0.63	6.2	51	5.9e+02	5.4	2.2
MAD	0.31	0.34	0.37	0.83	3.3	27	1.8	0.99
FLOM 1.0	0.27	0.27	0.32	0.73	2.2	7.2	0.35	0.22
FLOM 1.5	0.16	0.31	0.85	5.1	21	92	1.5	0.47
FLOM 1.8	0.093	1.4	5.1	30	1.2e+02	4.9e+02	6.4	1.7
SCM1	1.8	1.8	1.8	1.8	1.9	1.9	1.8	1.8
SCM2	0.22	0.25	0.38	1.3	6.3	89	3.5	1.7

**Table 3.** Estimator performance using the relative MSE of the eigenvalues and the  $A_1$  mixing matrix.

**Sphericity statistic:** The ratio of the geometric mean to the arithmetic mean of the eigenvalues,

$$L_{SS}(\hat{C}) = \frac{\left( \prod_i \lambda_{i,\hat{C}} \right)^{1/M}}{\frac{1}{M} \sum_i \lambda_{i,\hat{C}}} . \tag{12}$$

A normalised sphericity metric is then obtained as  $L_S(\hat{C}, C) = L_{SS}(\hat{C})/L_{SS}(C)$ . The sphericity statistic indicates the shape of the distribution. If  $C$  has equal eigenvalues the scale is equal in all directions. It

also appears in the likelihood function of model selection criteria, such as the MDL, for source detection with Gaussian observations. Hence an  $L_S$  nears 1 would indicate good performance of model selection criteria when using robust eigenvalue estimates. Results are shown in Table 4 for  $A_2$ .

- The M-estimator based techniques, both static and adaptive, show steady degeneration with increasing contamination.
- Encouraging results are found for SCM2 and for the higher order FLOMs.
- Results for the sample estimator are seen to be excellent.

Estimator	Noise distribution							
	1	2	3	4	5	6	7	8
Sample	0.98	0.89	0.87	0.89	0.93	0.96	0.95	0.97
Adaptive	0.99	0.96	0.95	0.9	0.79	0.67	0.94	0.96
Huber 1.0	0.97	0.95	0.93	0.89	0.82	0.68	0.92	0.93
Huber 1.5	0.97	0.96	0.94	0.89	0.79	0.68	0.94	0.95
Huber 2.0	0.98	0.96	0.94	0.87	0.76	0.81	0.95	0.96
Huber 2.5	0.99	0.96	0.92	0.83	0.79	0.9	0.95	0.97
MAD	0.94	0.93	0.91	0.88	0.81	0.67	0.9	0.92
FLOM 1.0	0.98	0.89	0.83	0.69	0.56	0.49	0.84	0.88
FLOM 1.5	0.99	0.89	0.84	0.78	0.76	0.79	0.9	0.93
FLOM 1.8	0.98	0.89	0.85	0.82	0.85	0.89	0.92	0.95
SCM1	1.2	1.2	1.2	1.2	1.2	1.1	1.2	1.2
SCM2	0.97	0.96	0.97	0.95	0.91	0.82	0.92	0.94

**Table 4.** Estimator performance using the sphericity ratio and the  $A_2$  mixing matrix

This can be explained as follows. Both the nominal and contaminating components have the same correlation matrix, differing only in their relative powers. Hence large amounts of contamination do not significantly affect this statistic. However, with only small amounts of contamination the sample subspace can be perturbed. If the nominal and contaminating components possessed different correlation structures, we would expect a steady deterioration in performance with respect to the amount of contamination.

## 5 Conclusions

The proposed adaptive scheme shows significant advantages over the static M-estimator. In particular, when considering the possibility of an unknown degree of contamination, performance follows the properties of the appropriate static estimator. When compared to other estimators, the adaptive scheme shows good performance in the nominal case, while also showing

graceful degradation as contamination increases. Other schemes were shown to have either poor nominal performance (e.g. M-estimator with small  $k$ , MAD, SCM, FLOM with small  $p$ ) or more rapid breakdown (e.g. M-estimator with large  $k$  and FLOM with large  $p$ ).

It is observed that SCM2, i.e. the whitened SCM, shows considerable improvement across *all* metrics for light contamination when compared to the unmodified SCM1. This motivates future investigation into the application of a similar transformation for other estimators.

## References

- [Breich and Zoubir, 2002]R. F. Breich and A.M. Zoubir. Robust estimation with parametric score function estimation. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2002*, Orlando, FL, USA, May 2002.
- [Breich *et al.*, 2004]R. F. Breich, C. L. Brown, and A. M. Zoubir. An adaptive robust estimator for scale in contaminated distributions. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2004*, Montreal, Canada, May 2004.
- [Brown *et al.*, 2003]C. L. Brown, R. F. Breich, and A. Taleb. Suboptimal robust estimation using rank score functions. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2003*, Hong Kong, April 2003.
- [Cook *et al.*, 1993]R. D. Cook, D. M. Hawkins, and S. Weisberg. Exact iterative computation of the robust multivariate minimum volume ellipsoid estimator. *Statistics and Probability Letters*, 16:213–218, 1993.
- [Huber, 1981]P. Huber. *Robust Statistics*. John Wiley, 1981.
- [Kendall and Gibbons, 1990]M. Kendall and J. D. Gibbons. *Rank Correlation Methods*. Oxford Univ. Press, 5th edition, 1990.
- [Kozick and Sadler, 2000]R. Kozick and B. Sadler. Maximum likelihood array processing in non-Gaussian noise with Gaussian mixtures. *IEEE Trans. on Signal Processing*, 48(12):3520–35, December 2000.
- [Liu and Mendel, 2001]T. Liu and J. Mendel. A subspace-based direction finding algorithm using fractional lower order statistics. *IEEE Trans. on Signal Processing*, 49(8):1605–13, August 2001.
- [Shao and Nikias, 1993]M. Shao and C. L. Nikias. Signal processing with fractional lower order moments: Stable processes and their applications. *Proc. IEEE*, 81(7):986–1010, July 1993.
- [Tsakalides and Nikias, 1996]P. Tsakalides and C. Nikias. The robust covariation-based MUSIC (ROC-MUSIC) algorithm for bearing estimation in impulsive environments. *IEEE Trans. on Signal Processing*, 44(7):1623–33, July 1996.
- [Visuri *et al.*, 2001]S. Visuri, H. Oja, and V. Koivunen. Subspace-based direction-of-arrival estimation using nonparametric statistics. *IEEE Trans. on Signal Processing*, 49(9):2060–73, September 2001.
- [Williams and Johnson, 1993]D. Williams and D. Johnson. Robust estimation of structured covariance matrices. *IEEE Trans. on Signal Processing*, 41(9):2891–906, September 1993.