DOUBLE SHRINKAGE CORRECTION IN SAMPLE LMMSE ESTIMATION

*Jordi Serra** and *Montse Nájar*†

* Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), Castelldefels (Spain). E-mail jserra@cttc.es
† Universitat Politècnica de Catalunya (UPC), Barcelona (Spain). E-mail montse.najar@upc.edu

ABSTRACT

The sample linear minimum mean square error (LMMSE) estimator undergoes high performance degradation in the small sample size regime. Herein a double shrinkage correction is proposed to alleviate this problem. First, an affine transformation of the sample covariance matrix (SCM) is considered within the LMMSE. Second, a linear transformation of that modified filter is proposed. The linear transformation minimizes the asymptotic MSE of the filter given a shrinkage of the SCM. And the shrinkage of the SCM optimizes the asymptotic MSE of the data covariance. Simulations highlight that the proposed estimator outperforms robust methods to the small sample size, namely LMMSE based on diagonal loading (DL) or Ledoit-Wolf (LW) regularizations of the SCM.

Index Terms— LMMSE, shrinkage, small sample size, random matrix theory.

1. INTRODUCTION

The problem of linear estimation of a parameter, observed through a linear model, appears in many signal processing applications, e.g. signal waveform estimation in array processing [1]. Among the linear estimators LMMSE is the optimal method, in an MSE sense, see [2]. In practice, though, it is not realizable as it depends on the unknown covariance of the observations. In order to face this adversity the conventional approach is as follows. The unknown covariance in the LMMSE is estimated using the SCM. This yields the so called sample LMMSE method. The rationale for this approach is that the SCM is the maximum likelihood (ML) estimator for Gaussian data, see [3, Theorem 4.1]. Indeed, it is the minimum variance unbiased estimator (MVUE) in the large sample size regime, i.e. when the number of samples N is large compared to the observation dimension M. Nonetheless, in practical situations N may be comparable to M, which yields an ill conditioned SCM and leads to severe performance degradation of the sample LMMSE. Several methods have been proposed to tackle this problem. Namely, DL techniques regularize the SCM by adding a constant to its diagonal, see [4], though the choice of the constant is rather controversial. In [5], Ledoit and Wolf proposed a general shrinkage of the SCM, i.e. an affine transformation that not only regularizes the SCM but also optimizes the asymptotic MSE of the data covariance. Both DL and LW aim to regularize and improve the SCM estimate but they do not directly deal with the estimation of the parameter of interest, which is the final target herein. This case was tackled in [6], where a shrinkage of the sample LMMSE method that minimizes the asymptotic MSE of the signal of interest, provided that $N > M$, was proposed based on random matrix theory (RMT) tools.

Herein, the benefits of those two approaches are considered. As in [6] a linear transformation of the LMMSE filter is considered, as according to the shrinkage theory by adding a bias to the filter the variance in the estimation of the parameter may diminish and the overall MSE may decrease. However, [6] is improved and extended to deal with $M \geqslant N$. Namely, unlike [6] which uses the SCM to implement the LMMSE filter, herein a shrinkage of the SCM is considered. This is a better estimate than the SCM and is well conditioned in any sample size regime. The scalar controlling the shrinkage of the filter is designed as the one minimizing the asympotic MSE of the filter, given a shrinkage of the SCM. Whereas the scalars governing the shrinkage of the SCM are the ones proposed by LW, i.e. the ones minimizing the asymptotic MSE of the data covariance. The simulations show that the proposed approach not only outperforms dramatically the sample LMMSE, but also the shrinkage LMMSE in [6] and the implementations of the LMMSE when considering DL or LW estimations of the covariance.

The paper is organized as follows. Section 2 deals with the signal model and the relation with prior work. Section 3 derives the proposed double shrinkage LMMSE method. Next, section 4 presents numerical simulations comparing the proposed estimator with the alternatives reviewed above. Finally, section 5 concludes the paper.

This work was partially supported by the Spanish Ministry of Economy and Competitiveness under projects TEC2011-29006-C03-01 (GRE3N-PHY), TEC2011-29006-C03-02 (GRE3N-LINK-MAC), by the Catalan Government under grant 2009-SGR-891, by the European Commission in the framework of the FP7 Network of Excellence in Wireless Communications NEWCOM# (Grant agreement no.318306) and by the European Cooperation in Science and Technology under project COST Action IC0902.

2. PRELIMINARIES

2.1. Signal model

Let $x(n) \in \mathbb{C}$ be a parameter to be estimated and let it be observed through the stochastic process $\mathbf{y}(n) \in \mathbb{C}^M$ by means of the next linear model,

$$\mathbf{y}(n) = x(n)\mathbf{s} + \mathbf{n}(n), \quad 1 \leq n \leq N \tag{1}$$

being $\mathbf{s} \in \mathbb{C}^M$ a known deterministic vector, $\mathbf{n}(n) \in \mathbb{C}^M$ a stochastic process and N the number of available measures, e.g. in the context of array processing $\mathbf{y}(n)$ is the output of an antenna array, $\mathbf{s}$ is the steering vector and $\mathbf{n}(n)$ contains the noise plus interference signals [1]. Moreover, the next model assumptions are supposed to hold:

(a) *$x(n)$ and $\mathbf{n}(n)$ are uncorrelated. Moreover, $\mathbb{E}[\mathbf{n}(n)] = \mathbf{0}$ and $\mathbb{E}\left[\mathbf{n}(n)\mathbf{n}(n)^H\right] = \mathbf{R_n}$.*

(b) *As a consequence of (a)* $\mathbf{R} \triangleq \mathbb{E}\left[\mathbf{y}(n)\mathbf{y}(n)^H\right] = \gamma\mathbf{s}\mathbf{s}^H + \mathbf{R_n}$; $\gamma \triangleq \mathbb{E}\left[|x(n)|^2\right]$ is known and $\|\mathbf{s}\|^2 = 1$.

(c) *The set of observations $\{\mathbf{y}(n)\}_{n=1}^N$ are iid.*

2.2. LMMSE in the small sample size regime

Consider an estimate of $x(n)$ based on a linear filtering of $\mathbf{y}(n)$, i.e. $\hat{x}(n) = \mathbf{w}^H\mathbf{y}(n)$. Then, taking into account the data model in (1) the MSE for a given filter $\mathbf{w}$ reads,

$$\mathrm{MSE}(\mathbf{w}) = \mathbf{w}^H\mathbf{R}\mathbf{w} + \gamma(1 - \mathbf{w}^H\mathbf{s} - \mathbf{s}^H\mathbf{w}) \tag{2}$$

The optimization of this expression with respect to $\mathbf{w}$ leads to the well known LMMSE estimator [2],

$$\hat{x}_l(n) = \mathbf{w}_l^H\mathbf{y}(n);\ \mathbf{w}_l = \gamma\mathbf{R}^{-1}\mathbf{s} \tag{3}$$

Nonetheless, this estimator is not realizable as it depends on the unknown noise plus interference covariance $\mathbf{R_n}$, see (1). In order to circumvent this problem the common approach is the so called sample LMMSE, which relies on replacing $\mathbf{R}$ by its sample estimate $\hat{\mathbf{R}} \triangleq \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{y}(n)\mathbf{y}^H(n)$ in (3),

$$\hat{x}_t(n) = \mathbf{w}_t^H\mathbf{y}(n);\ \mathbf{w}_t = \gamma\hat{\mathbf{R}}^{-1}\mathbf{s} \tag{4}$$

The rationale behind this method relies on the optimality properties of $\hat{\mathbf{R}}$. Specifically, it is the ML estimator of $\mathbf{R}$ for Gaussian data [3, Theorem 4.1] and the MVUE for $N \gg M$. Nonetheless, in the small sample size regime the SCM is an ill conditioned estimator [5], i.e. inverting $\hat{\mathbf{R}}$ when $N \approx M$ severely amplifies the estimation error, indeed for $N \leqslant M$ may not be even invertible. Thereby in the small sample size regime the SCM yields large performance degradation of the sample LMMSE. In order to face this problem DL is a popular technique, see [4, Chapter 4] and references therein. It is based on regularizing the SCM by adding a constant to its diagonal, yielding the so called DL-LMMSE herein,

$$\hat{x}_{dl}(n) = \mathbf{w}_{dl}^H\mathbf{y}(n);\ \mathbf{w}_{dl} = \gamma(\hat{\mathbf{R}} + \mu\mathbf{I})^{-1}\mathbf{s} \tag{5}$$

The correct choice of μ is rather controversial, a possibility is to fix μ equal to the standard deviation of the diagonal entries of $\hat{\mathbf{R}}$, see [7]. This is the value considered below for simulation purposes. More general is the approach of Ledoit and Wolf (LW), see [5]. They propose a shrinkage of the SCM that not only regularizes and improves the SCM, but also minimizes the asymptotic MSE of the data covariance, i.e. $\breve{\mathbf{R}}_{lw} = \beta_1^{lw}\hat{\mathbf{R}} + \beta_2^{lw}\mathbf{I}$. Plugging $\breve{\mathbf{R}}_{lw}$ into (3) yields the so called LW-LMMSE herein,

$$\hat{x}_{lw}(n) = \mathbf{w}_{lw}^H\mathbf{y}(n);\ \mathbf{w}_{lw} = \gamma(\beta_1^{lw}\hat{\mathbf{R}} + \beta_2^{lw}\mathbf{I})^{-1}\mathbf{s} \tag{6}$$

For the complex case, β_1^{lw} and β_2^{lw} may be implemented as follows, see [8],

$$\beta_1^{lw} = 1 - \beta \quad , \quad \beta_2^{lw} = \beta\frac{\mathrm{Tr}(\hat{\mathbf{R}}_s)}{2M} \tag{7}$$

$$\beta = \frac{\sum_{i=1}^{N}\left\|\mathbf{y}_s(i)\mathbf{y}_s^T(i) - \hat{\mathbf{R}}_s\right\|_F^2}{N^2\left[\mathrm{Tr}(\hat{\mathbf{R}}_s^2) - \frac{Tr^2(\hat{\mathbf{R}}_s)}{2M}\right]} \tag{8}$$

where $\hat{\mathbf{R}}_s = \frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_s(i)\mathbf{y}_s^T(i)$, $\|\cdot\|_F$ denotes the Frobenius norm, $\mathbf{y}_s(i) = [\mathrm{Re}(\mathbf{y}(i)), \mathrm{Im}(\mathbf{y}(i))]^T$ and $\mathrm{Tr}[\cdot]$ denotes the trace operator. LMMSE implementation based on either DL or LW aims to improve the estimate of $\mathbf{R}$, but do not directly deal with the estimation of $x(n)$, which is the final target. This was the point of view in [6]. Specifically, the authors proposed the next shrinkage of the sample LMMSE that minimizes the asymptotic MSE of $\hat{x}(n)$, provided that $N > M$, and that gives much better performance than the sample LMMSE,

$$\hat{x}_s(n) = \mathbf{w}_s^H\mathbf{y}(n);\ \mathbf{w}_s = \gamma(1 - c_f)^2\hat{\mathbf{R}}^{-1}\mathbf{s} \tag{9}$$

where $c_f = M/N$. In this paper the benefits of both strategies are adopted by proposing a double shrinkage structure of the filter. Namely, recall that shrinkage theory relies on introducing a bias in the estimation of a parameter such that its estimation variance decreases and thereby the overall MSE diminishes. Thus, as in [6] a linear transformation of the filter is proposed as a correction controlling the minimization of the MSE of the parameter of interest. Moreover, a shrinkage of the SCM is proposed, which permits, unlike in [6], to deal with $M > N$ and to expect better performance, as the shrinkage of the SCM is a better estimate of the data covariance than the SCM. Moreover, we embrace DL-LMMSE or LW-LMMSE methods and improve them due to the additional shrinkage of the filter as shown below. Thus, the proposed estimator has the next structure,

$$\hat{x}_{ds}(n) = \mathbf{w}_{ds}^H \mathbf{y}(n);\ \mathbf{w}_{ds} = \alpha(\beta_1 \hat{\mathbf{R}} + \beta_2 \mathbf{I})^{-1}\mathbf{s} \tag{10}$$

3. PROPOSED DOUBLE SHRINKAGE LMMSE

Next, the design of the shrinkage coefficients α, β_1 and β_2 of the proposed filter in (10) is dealt with. Namely, first the optimal α is obtained, i.e. β_1 and β_2 are assumed known and the MSE is optimized with respect to α. Thus, defining $\check{\mathbf{R}} = \beta_1 \hat{\mathbf{R}} + \beta_2 \mathbf{I}$, and after straightforward manipulations, one finds that the minimization of the MSE in (2) is obtained for the next α,

$$\alpha_o = \arg\min_{\alpha} \mathrm{MSE}\,(\mathbf{w}) = \gamma \frac{\mathbf{s}^H \check{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^H \check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s}} \tag{11}$$

This highlights that the shrinkage of the filter is not superfluous, i.e. $\alpha_o \neq 1$ and thereby does not lead to the case of just considering $\mathbf{w}_{ds} = (\beta_1 \hat{\mathbf{R}} + \beta_2 \mathbf{I})^{-1}\mathbf{s}$. Nonetheless, this leads to an unrealizable filter. To circumvent this problem, an (M, N)-consistent estimate of (11), denoted by $\hat{\alpha}_o$ will be obtained by means of RMT and G-estimation tools, i.e. $\hat{\alpha}_o$ will converge in probability towards α_o when $M, N \to \infty$ and $M/N \to c \in (0, \infty)$, in compact notation $\hat{\alpha}_o \asymp \alpha_o$. Thereby $\hat{\alpha}_o$ will minimize the asymptotic MSE. Note that this general asymptotic regime naturally deals with small sample size situations. To achieve our aim, the procedure is as follows,

1. Find the asymptotic deterministic expressions of $\mathbf{s}^H \check{\mathbf{R}}^{-1}\mathbf{s}$ and $\mathbf{s}^H \check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s}$, denoted by $f(\mathbf{R}, \beta_1, \beta_2)$ and $g(\mathbf{R}, \beta_1, \beta_2)$, respectively.
2. Obtain (M, N)-consistent estimates of $f(\cdot)$ and $g(\cdot)$, denoted by $\hat{f}(\hat{\mathbf{R}}, \beta_1, \beta_2)$ and $\hat{g}(\hat{\mathbf{R}}, \beta_1, \beta_2)$.
3. Estimate $\mathbf{s}^H \check{\mathbf{R}}^{-1}\mathbf{s}$ and $\mathbf{s}^H \check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s}$ using $\hat{f}\,(\cdot)$ and $\hat{g}(\cdot)$, respectively, i.e. $\hat{\alpha}_o = \gamma \frac{\hat{f}(\hat{\mathbf{R}}, \beta_1, \beta_2)}{\hat{g}(\hat{\mathbf{R}}, \beta_1, \beta_2)} \asymp \alpha_o$.

For the sake of clarity of presentation, the convergence of the numerator and denominator in (11) is given in the appendix. It is based on the results in [9, appendix I] and substituted in (11) and after some manipulations yields that α_o converges in probability to the next expression,

$$\alpha_o \asymp \gamma \beta_1 \frac{1 - c\xi}{1 + cb} \frac{\mathbf{s}^H (\mathbf{R} + \rho \mathbf{I})^{-1}\mathbf{s}}{\mathbf{s}^H (\mathbf{R} + \rho \mathbf{I})^{-1}\mathbf{R}(\mathbf{R} + \rho \mathbf{I})^{-1}\mathbf{s}} \tag{12}$$

where $\xi = \frac{1}{M}\sum_{i=1}^{M} \frac{\lambda_i^2}{(\lambda_i + \rho)^2}$, λ_i are the eigenvalues of $\mathbf{R}$, $\rho = \delta(1 + cb)$, $\delta \triangleq \beta_2/\beta_1$ and b is the positive solution to the next equation $b = \frac{1}{M}\sum_{i=1}^{M} \frac{\lambda_i (1+cb)}{\lambda_i + \delta(1+cb)}$. Therefore, in order to obtain an (M, N)-consistent estimate of α_o, one must obtain consistent estimates of b, ξ, $\eta_d = \mathbf{s}^H (\mathbf{R} + \rho \mathbf{I})^{-1}\mathbf{s}$ and $\eta_n = \mathbf{s}^H (\mathbf{R} + \rho \mathbf{I})^{-1}\mathbf{R}(\mathbf{R} + \rho \mathbf{I})^{-1}\mathbf{s}$. These are provided in [10, Appendix A], and summarized below in the appendix. Now, in order to obtain the (M, N)-consistent estimate of the optimal shrinkage factor α_o, one needs to substitute the estimates of b, ξ, η_n and η_d into (12). After this step, one needs some manipulations based on the matrix inversion lemma and the next identity [4, p. 248],

$$\begin{aligned}\frac{1}{M}\mathrm{Tr}[\mathbf{R}^2(\mathbf{R} + \delta \mathbf{I})^{-2}] \quad &= \quad 1 - \frac{2}{M}\mathrm{Tr}[(\delta^{-1}\mathbf{R} + \mathbf{I})^{-1}] + \\ &\qquad \frac{1}{M}\mathrm{Tr}[(\delta^{-1}\mathbf{R} + \mathbf{I})^{-2}]\end{aligned}$$

Finally, this process leads to obtain the desired (M, N)-consistent estimate of α_o,

$$\begin{aligned}\hat{\alpha}_o = \gamma \beta_1 (1 - \frac{c}{M}\mathrm{Tr}[\hat{\mathbf{R}}(\hat{\mathbf{R}} + \delta \mathbf{I})^{-1}])^2 \\ \cdot \frac{\mathbf{s}^H (\hat{\mathbf{R}} + \delta \mathbf{I})^{-1}\mathbf{s}}{\mathbf{s}^H (\hat{\mathbf{R}} + \delta \mathbf{I})^{-1}\hat{\mathbf{R}}(\hat{\mathbf{R}} + \delta \mathbf{I})^{-1}\mathbf{s}}\end{aligned} \tag{13}$$

At this point, the design of β_1 and β_2 must be tackled. The optimal approach would be to substitute (11) in (10) and minimize the MSE in (2) with respect to β_1 and β_2. Nonetheless, β_1,β_2 could not be isolated due to their presence within the inverse. To circumvent that and to obtain a realizable filter, one could proceed as in [9], find the (M, N)-consistent estimate for the asymptotic MSE($\alpha = \alpha_o$,β_1, β_2) and find the β_1, β_2 minimizing it. Though a realizable filter is obtained, one still must carry out an extensive search to find β_1 and β_2. Instead of this approach, herein we propose to substitute β_1 and β_2 by the estimates proposed by Ledoit and Wolf in [5], i.e. β_1^{lw} and β_2^{lw} in (7). In this way, the numerical search is avoided. Moreover this fits perfectly in our framework as β_1^{lw} and β_2^{lw} are (M, N)-consistent estimates of the β_1 and β_2 that optimize the MSE of the covariance, when using the shrinkage estimator $\check{\mathbf{R}} = \beta_1 \hat{\mathbf{R}} + \beta_2 \mathbf{I}$, $\beta_2 = (1 - \beta_1)\,\mathrm{Tr}(\hat{\mathbf{R}}_s)/2M$, and $\hat{\mathbf{R}}_s$ defined above in (8). Next section compares this approach to the lower MSE bound when considering the optimal α_o in (11) and β_1, β_2 obtained by a 2D search in (2). Recalling that $\check{\mathbf{R}} = \beta_1 \hat{\mathbf{R}} + \beta_2 \mathbf{I}$ and that $\delta = \beta_2/\beta_1$, we can manipulate (13) to obtain the final expression of the proposed algorithm, which is summarized next.

Proposed estimator

The proposed estimator for the parameter $x(n)$, given the data model (1) is given by the next expression,

$$\hat{x}_{ds}(n) = \mathbf{w}_{ds}^H \mathbf{y}(n); \mathbf{w}_{ds} = \hat{\alpha}_o \check{\mathbf{R}}^{-1}\mathbf{s}$$

$$\hat{\alpha}_o = \gamma(1 - \frac{c\beta_1^{lw}}{M}\,\mathrm{Tr}[\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}])^2 \frac{\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^H\check{\mathbf{R}}^{-1}\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}\mathbf{s}} \qquad (14)$$

$$\check{\mathbf{R}} = \beta_1^{lw}\hat{\mathbf{R}} + \beta_2^{lw}\mathbf{I}$$

and the expression of β_1^{lw}, β_2^{lw} is given by (7).

Remark: The proposed estimator does not rely on any assumption regarding the distribution of the observed data $\mathbf{y}(n)$ in (1).

4. NUMERICAL SIMULATIONS

This section conducts an MSE comparison, between the proposed double shrinkage LMMSE in (14), the shrinkage LMMSE in (9), the sample LMMSE in (4), the DL-LMMSE in (5), the LW-LMMSE in (6) and the lower bound i.e. the LMMSE in (3). The lower MSE bound for the type of double shrinkage filters considered herein, i.e. (10), is also plotted, considering the optimal α_o in (11) and β_1, β_2 obtained by a 2D search assessing (2) in the grid $\beta_1, \beta_2 \in (0, 1)$. An array signal processing application is considered. Namely, $\mathbf{y}(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$ in (1), where $\mathbf{R} = \gamma\mathbf{s}\mathbf{s}^H + \mathbf{R_n}$. Without loss of generality $\gamma = 1$. A uniform linear array is assumed to specify the steering vector $\mathbf{s}$, i.e. $[\mathbf{s}]_m = \frac{e^{j\pi \sin\theta_0 m}}{\sqrt{M}}$, where θ_0 is the Direction of Arrival (DOA) of the signal of interest, see [1]. For the simulation purposes $\theta_0 = 0°$. $\mathbf{R_n}$ is modelled as $\mathbf{R_n} = \mathbf{S}\mathbf{P}\mathbf{S}^H + \sigma^2\mathbf{I}$ [1]. Where, $[\mathbf{S}]_{m,k} = \frac{e^{j\pi \sin\theta_k m}}{\sqrt{M}}$, $m = 0, \ldots, M-1$ is the antenna index, $k = 1, \ldots, K$ defines a set of interferers and θ_k is the DOA of the *k-th* interferer. $\mathbf{P}$ is the covariance matrix of the interferers and σ^2 is the power of an AWGN. For the simulations 4 interferers are considered, whose DOAs are $\{\theta_k\}_{k=1}^4 = \{45°, -45°, 85°, -85°\}$. $\mathbf{P}$ is considered to be diagonal and the elements of the diagonal are set to 1. Regarding σ^2, it is set to $\sigma^2 = \gamma 10^{-SNR/10}$, where $SNR = 5$ dB is the signal to noise ratio. Finally, $M = 16$ and N is specified below.

In figure 1, the proposed double shrinkage LMMSE (14) is compared to the theoretical LMMSE (3), its sample implementation (4) and the shrinkage LMMSE (9) proposed in [6]. $N \in [16, 300]$, because the two latter methods can not deal with $M > N$. The proposed method dramatically outperforms the sample LMMSE in the small sample size regime i.e. $c_f \to 1$. Indeed it improves the shrinkage LMMSE (9), since it relies on the shrinkage of the SCM, which is a better estimate of $\mathbf{R}$ than the SCM, the one used by (9). In the large sample size regime all the methods converge, as in this situation the SCM becomes a good estimate of $\mathbf{R}$.

In figure 2, the proposed shrinkage estimator (14) is compared to the DL-LMMSE (5) and the LW-LMMSE (6), i.e. to the other methods that are robust to the small sample size and that support $M > N$, thus in this figure $N \in [4, 300]$. The estimator proposed in this paper clearly outperforms the other alternative methods. This is due to the additional shrinkage of the filter governed by $\hat{\alpha}_o$ (14) and a design based on facing the minimization of the MSE of the estimate of $x(n)$. Moreover, LW-LMMSE outperforms DL-LMMSE as its regularization of $\hat{\mathbf{R}}$ is asymptotically optimal.

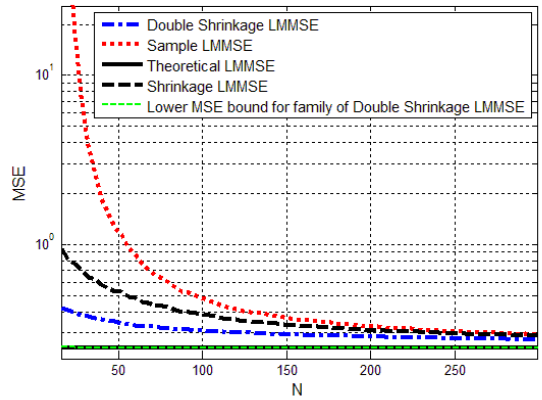

Fig. 1: Performance comparison between proposed double shrinkage LMMSE (14), theoretical LMMSE estimator (3), sample LMMSE (4) and shrinkage LMMSE (9).

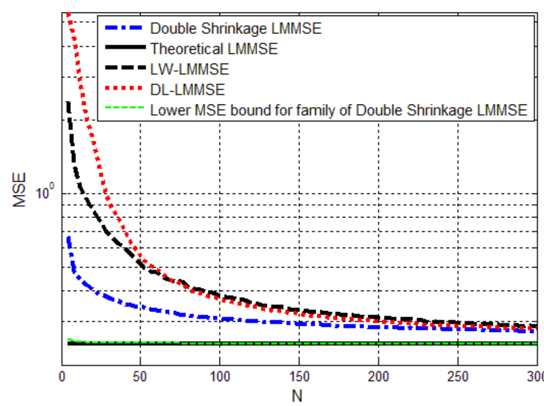

Fig. 2: Performance comparison between proposed double shrinkage LMMSE (14), theoretical LMMSE estimator (3), LW-LMMSE (6) and DL-LMMSE (5).

5. CONCLUSIONS

This paper has dealt with the degradation of the sample LMMSE when $M > N$ or $N \approx M$. The LMMSE is implemented by using a shrinkage of the SCM and a linear scaling of the LMMSE is considered. Given the shrinkage of the SCM, the linear scaling minimizes the asymptotic MSE of the parameter of interest. And then the shrinkage of the SCM is selected to optimize the asymptotic MSE of the data covariance. Simulations highlight that the proposed estimator clearly outperforms alternative methods based on regularizing the SCM or shrinking the LMMSE filter.

A. APPENDIX

A.1 Convergence of the numerator and denominator in (11)

Let define $\delta = \beta_2/\beta_1$, then one can express the numerator and the denominator in (11) as $\beta_1^{-1}\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}$ and $\beta_1^{-2}\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{R}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}$, respectively. The convergence of these expressions was obtained in [9, appendix I] in terms of the eigenvectors, $\mathbf{e}_i$, and eigenvalues, λ_i, of $\mathbf{R}$.

$$\beta_1^{-1}\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s} \asymp \beta_1^{-1}\sum_{i=1}^{M}\frac{(1+cb)\left|\mathbf{s}^H\mathbf{e}_i\right|^2}{\lambda_i+\rho}$$
$$\beta_1^{-2}\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{R}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s} \asymp \beta_1^{-2}((1+cb)^2+cb')\cdot\sum_{i=1}^{M}\frac{\left|\mathbf{s}^H\mathbf{e}_i\right|^2\lambda_i}{(\lambda_i+\rho)^2}$$

where $\rho \triangleq \delta(1+cb)$, $b \triangleq b(z)\mid_{z=0}$ is the positive solution to the next transcendental equation, $b' = \frac{db(z)}{dz}\mid_{z=0}$ is defined next and $b(z)$ is defined in [9, eq. 25],

$$b = \frac{1}{M}\sum_{i=1}^{M}\frac{\lambda_i(1+cb)}{\lambda_i+\rho}$$
$$b' = (1-\frac{1}{M}\sum_{i=1}^{M}\frac{c\lambda_i^2}{(\lambda_i+\rho)^2})^{-1}\frac{1}{M}\sum_{i=1}^{M}\frac{\lambda_i^2(1+cb)^2}{(\lambda_i+\rho)^2}$$

Finally, bearing in ming the next equalities,

$$\sum_{i=1}^{M}\frac{\left|\mathbf{s}^H\mathbf{e}_i\right|^2}{\lambda_i+\rho} = \mathbf{s}^H(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{s}$$
$$\sum_{i=1}^{M}\frac{\left|\mathbf{s}^H\mathbf{e}_i\right|^2\lambda_i}{(\lambda_i+\rho)^2} = \mathbf{s}^H(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{R}(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{s}$$

one obtains that the numerator and denominator in (11) convergence in probability to the next expressions,

$$\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s} \asymp (1+cb)\mathbf{s}^H(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{s}$$
$$\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{R}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s} \asymp ((1+cb)^2+cb')\cdot\mathbf{s}^H(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{R}(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{s}$$

A.2 (M,N)-consistent estimates of b, ξ, η_n, η_d in (12)

The consistent estimates of these parameters is given in [10, App.A] and it is based on expressing b, ξ, η_n and η_d in terms of the next real Stieljes transforms $t(x)$ and $s(x)$,

$$\xi = 1 - s(x)\mid_{x=\rho^{-1}} + \rho^{-1}\frac{ds(x)}{dx}\mid_{x=\rho^{-1}}$$
$$b/(1+cb) = 1 - s(\delta^{-1}(1+cb)^{-1})$$
$$\eta_d = \rho^{-1}t(\rho^{-1}),\ \ \eta_n = -[x^2 dt(x)/dx]\mid_{x=\rho^{-1}}$$

$$t(x) = \sum_{k=1}^{M}\frac{\left|\mathbf{s}^H\mathbf{e}_k\right|^2}{1+x\lambda_k},\ s(x) = \frac{1}{M}\sum_{k=1}^{M}\frac{1}{1+x\lambda_k},\ x>0$$

The (M,N)-consistent estimators of $t(x)$ and $s(x)$ were obtained in [11], see [10, App.A]. Thereby, this paves the way to obtain the next (M,N)-consistent estimates of b, ξ, η_n and η_d, see [10, App.A] for further details,

$$\hat{b} = \frac{1-\frac{\delta}{M}\operatorname{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}]}{1-c(1-\frac{\delta}{M}\operatorname{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}])}$$

$$\hat{\eta}_d = (1-c+c\frac{\delta}{M}\operatorname{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}])\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}$$

$$\hat{\xi} = \frac{\frac{1}{M}\operatorname{Tr}[\hat{\mathbf{R}}^2(\hat{\mathbf{R}}+\delta\mathbf{I})^{-2}]-\frac{c}{M^2}\operatorname{Tr}^2[\hat{\mathbf{R}}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}]}{1-c+c\delta^2\frac{1}{M}\operatorname{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-2}]}$$

$$\hat{\eta}_n = \frac{(1-c(1-\frac{\delta}{M}\operatorname{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}]))^2}{1-c+\frac{c}{M}\operatorname{Tr}[(\delta^{-1}\hat{\mathbf{R}}+\mathbf{I})^{-2}]}\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\hat{\mathbf{R}}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}$$

B. REFERENCES

[1] H.L. Van Trees, *Optimum Array Processing*, John Wiley and Sons, New York, NY, USA, 2002.

[2] S.M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice-Hall, 1993.

[3] N.R. Goodman, "Statistical analysis based on a certain multivariate complex gaussian distribution," *Ann. Math. Stat.*, vol. 34, pp. 152–177, March 1963.

[4] J. Li and P. Stoica, Eds., *Robust Adaptive Beamforming*, Wiley-Interscience, John Wiley and Sons, 2005.

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

[6] J. Serra and F. Rubio, "Bias corrections in linear mmse estimation with large filters," in *Proc. European Sig. Proc. Conf. (EUSIPCO 2010)*, Aalborg (Denmark), 2010, pp. 23–27.

[7] N. Ma and J. Goh, "Efficient method to determine diagonal loading value," in *Proc. International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2003, vol. V, pp. 341–344.

[8] R. Abrahamsson, Y. Selén, and P. Stoica, "Enhanced covariance matrix estimators in adaptive beamforming," in *Proc. International Conference on Acoustics Speech and Signal Processing (ICASSP) 2007*, 2007, pp. 969–972.

[9] X. Mestre and M.A. Lagunas, "Finite sample size effect on MV beamformers: optimum diagonal loading factor for large arrays," *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 69–82, January 2006.

[10] X. Mestre and M.A. Lagunas, "Estimating the optimum loading factor against the finite sample size effect in minimum variance beamformers," in *Proc. Sensor Array and Multichannel Signal Processing Workshop (SAM) 2004*, 2004, pp. 357–361.

[11] V.L. Girko, *An Introduction to Statistical Analysis of Random Arrays*, VSP, The Netherlands, 1998.