OPTIMIZATION OF THE LOADING FACTOR OF REGULARIZED ESTIMATED SPATIAL-TEMPORAL WIENER FILTERS IN LARGE SYSTEM CASE.

G.T. Pham, P. Loubaton[*]

Université Paris-Est Marne-la-Vallée
LIGM, UMR CNRS 8049
5 Boulevard Descartes
77454 Marne la Vallee Cedex 2, France

ABSTRACT

In this paper, it is established that the signal to interference plus noise ratio (SINR) produced by a trained regularized Wiener spatio-temporal filter can be estimated consistently in the asymptotic regime where the number of receivers and the number of snapshots converge to infinity at the same rate. The optimal regularization parameter is estimated as the argument of the maximum of the estimated SINR. Numerical simulations show that the proposed optimum regularized Wiener filter outperforms the existing regularized spatio-temporal Wiener filters.

1. INTRODUCTION

Finite impulse response spatio-temporal Wiener filter estimation using a training sequence is a very classical problem. When the useful signal is corrupted by an additive temporally and spatially white Gaussian noise, the optimal estimator is known to be the standard least-squares estimate defined as the action of the inverse of the empirical spatio-temporal covariance matrix on the empirical cross correlation between the observation and the training sequence. However, it is known for a long time that regularizing the empirical spatio-temporal covariance matrix by a multiple of the identity matrix may enhance the performance of the estimate because this matrix can be ill-conditioned or even non invertible when the size of the training sequence is smaller than the dimension of the vector associated to the Wiener filter. The choice of the regularization parameter appears to be a crucial issue that was addressed in a heuristic manner in a number of references (see e.g. [10], [13, p. 748], [11] and [12]) because classical figures of merit such as the signal to noise plus interference ratio (SINR) produced by the estimated Wiener filter are difficult to estimate in the general case. In the context of large dimension systems where the number of sensors and the lenght of training sequence are both large, the situation appears more favourable due to some subtle self-averaging effects. The existing related works addressed the purely spatial context. Ledoit and Wolf proposed in [9] to find the loading factor so as to minimize the mean-square error of the estimated empirical covariance matrix, and showed that the optimal value can be estimated consistently. This approach was generalized in [3] to the Tyler estimator in the context of robust estimation. Mestre and Lagunas [7] considered the case where the array response is a priori known (no training sequence) and where the noise plus interference covariance matrix is unknown. It is shown in [7] that the SINR produced by the regularized estimated Wiener filter can be consistently estimated from the available observations, and proposed to estimate the loading factor as the argument of the SINR maximization. The optimization of the SINR was also considered in [14] in the context of robust estimation.

In the present paper, we assume that the observation is a M–dimensional time series defined as a noisy output of an unknown SIMO finite impulse response system driven by the sequence of interest. We assume that a lenght N training sequence is available at the receiver side in order to estimate a regularized degree $L-1$ FIR spatio-temporal Wiener filter from the N M–dimensional observations collected during the transmission of the training sequence. In the large system context in which M and N both converge towards $+\infty$ at the same rate and where L remains fixed, we establish that the SINR produced by the regularized estimated Wiener filter, which, in principle, depends on the additive noise corrupting the N available observations, converges towards a deterministic term depending on the loading factor, the noise variance, assumed to be known, and the unknown filter. We show that, while the channel filter is unknown, the above limit SINR can be estimated consistently from the N available observations for each value of the regularization parameter, and propose to estimate the loading factor as the argument of its minimum.

This paper is organized as follows. In section 2, we present the signal models and the underlying assumptions. In section 3, we present some useful technical results proved in [6] and [8]. In section 4, we establish that the SINR converges towards a deterministic term, and section 5 addresses the consistent estimation of the limit SINR. Finally, section 6 presents numerical experiments sustaining our theoretical results, and comparing our proposal to the Ledoit-Wolf ([9]) estimator of the regularization parameter and to other empirical schemes proposed in the past ([10], [13, p. 748], [11] and [12]).

2. PROBLEM FORMULATION.

We assume that the observation is a M–dimensional time series $(\mathbf{y}_n)_{n\in\mathbb{Z}}$ defined by

$$\mathbf{y}_n = \sum_{p=0}^{P-1} \mathbf{h}_p s_{n-p} + \mathbf{v}_n, n = 1, ..., N \tag{1}$$

where $\mathbf{h}(z) = \sum_{p=0}^{P-1} \mathbf{h}_p z^{-p}$ represents the transfer function of the unknown FIR SIMO system and $(\mathbf{v}_n)_{n\in\mathbb{Z}}$ is an i.i.d sequence of complex Gaussian random vectors with spatial covariance matrix $\sigma^2\mathbf{I}$. Although $\mathbf{h}(z)$ is not known, we assume that P is known, i.e. in practice, that an upper bound of the support of the impulse response associated to $\mathbf{h}(z)$ is available. We assume that a lenght N training sequence $(s_n)_{n=1,...,N}$ is available at the receiver side, and

[*]The work of G.T.Pham and P.Loubaton was supported by project ANR-12-MONU-0003 DIONISOS

estimate from $(\mathbf{y}_n)_{n=1,\dots,N}$ the Wiener spatio-temporal filter $\mathbf{g} = (\mathbf{g}_0^T, \dots, \mathbf{g}_{L-1}^T)^T$, where the M dimensional vectors $(\mathbf{g}_l)_{l=0,\dots,L-1}$ are in principle designed in such a way that $\sum_{l=0}^{L-1} \mathbf{g}_l^* \mathbf{y}_{n+l}$ represents the minimum mean-square estimate of s_n. If we denote by $\mathbf{y}_n^{(L)}$ the ML–dimensional vector defined by

$$\mathbf{y}_n^{(L)} = (\mathbf{y}_{1,n}, \dots, \mathbf{y}_{1,n+L-1}, \dots, \mathbf{y}_{M,n}, \dots, \mathbf{y}_{M,n+L-1})^T$$

we study the performance of the estimated regularized Wiener filter $\hat{\mathbf{g}}_\lambda$ defined by

$$\hat{\mathbf{g}}_\lambda = \left(\frac{1}{N} \sum_{n=1}^{N} \mathbf{y}_n^{(L)} \mathbf{y}_n^{(L)*} + \lambda \mathbf{I}_{ML} \right)^{-1} \left(\frac{1}{N} \sum_{n=1}^{N} \mathbf{y}_n^{(L)} s_n^* \right) \quad (2)$$

$\hat{\mathbf{g}}_\lambda$ is destinated to estimate the unknown transmitted datas $(s_n)_{n>N}$. In the following, for each $m = 1, \dots, M$, we denote by $\mathbf{H}_m$ the $L \times (P+L-1)$ Toeplitz matrix corresponding to the convolution of signal $(s_n)_{n\in\mathbb{Z}}$ with sequence $(\mathbf{h}_{m,p})_{p=0,\dots,P-1}$, and define $ML \times (P+L-1)$ block-Hankel matrix $\mathbf{H}$ by $\mathbf{H} = (\mathbf{H}_1^T, \dots, \mathbf{H}_M^T)^T$. Assuming sequence $(s_n)_{n>N}$ i.i.d., the signal to interference plus noise ratio produced by $\hat{\mathbf{g}}_\lambda^{(L)}$ is easily seen to be equal to

$$\mathsf{SINR}(\hat{\mathbf{g}}_\lambda) = \frac{|\hat{\mathbf{g}}_\lambda^* \mathbf{h}_P|^2}{\hat{\mathbf{g}}_\lambda^* \mathbf{H}_1 \mathbf{H}_1^* \hat{\mathbf{g}}_\lambda + \sigma^2 \|\hat{\mathbf{g}}_\lambda\|^2} \quad (3)$$

where $\mathbf{h}_P$ is column P of $\mathbf{H}$, and matrix $\mathbf{H}_1$ is obtained by deleting column P from matrix $\mathbf{H}$. $\mathsf{SINR}(\hat{\mathbf{g}}_\lambda^{(L)})$ is random in the sense that it depends on the vectors $(\mathbf{y}_n)_{n=1,\dots,N}$, which are random themselves due to the presence of the additive noise. When N goes to ∞ and M, L remain fixed, it is easily to see that if $\lambda = 0$, the filter $\hat{\mathbf{g}}_0$ converges towards Wiener filter $(\mathbf{H}\mathbf{H}^* + \sigma^2 \mathbf{I})^{-1} \mathbf{h}_P$ and that $\mathsf{SINR}(\hat{\mathbf{g}}_0)$ converges towards γ defined by

$$\gamma = \frac{\mathbf{h}_P^* \left(\mathbf{H}\mathbf{H}^* + \sigma^2 \mathbf{I}\right)^{-1} \mathbf{h}_P}{1 - \mathbf{h}_P^* \left(\mathbf{H}\mathbf{H}^* + \sigma^2 \mathbf{I}\right)^{-1} \mathbf{h}_P} \quad (4)$$

Similar results hold when $\lambda > 0$. On the contrary, when M, N are of the same order of magnitude, the analysis of the behaviour of $\mathsf{SINR}(\hat{\mathbf{g}}_\lambda)$ is different and requires much more work. From now on, we assume that

> $M, N \to +\infty$, the ratio $c_N = \frac{M}{N} \to c > 0$, and P and L remain fixed.

To simplify the notations, $N \to +\infty$ should be understood as the above asymptotic regime. In the following, it appears convenient to define by $\mathbf{\Sigma}$ and $\mathbf{W}$ the normalized $ML \times N$ block-Hankel matrices defined by

$$\mathbf{\Sigma} = \frac{1}{\sqrt{N}} (\mathbf{y}_1^{(L)}, \dots, \mathbf{y}_N^{(L)}), \mathbf{W} = \frac{1}{\sqrt{N}} (\mathbf{v}_1^{(L)}, \dots, \mathbf{v}_N^{(L)})$$

respectively. Then, the relations between the available observations and sequence $(s_n)_{n=1,\dots,N}$ can be expressed as

$$\mathbf{\Sigma} = \mathbf{H}\mathbf{U} + \mathbf{W} \quad (5)$$

where $\mathbf{U}$ is the $(P + L - 1) \times N$ Hankel matrix defined by $(\mathbf{U}^{(L)})_{i,n} = s_{n+i-P}/\sqrt{N}$ Without loss of generality, we can assume that $\mathbf{U}\mathbf{U}^* = \mathbf{I}$, because it is possible to replace $\mathbf{H}$ by $\mathbf{H}(\mathbf{U}\mathbf{U}^*)^{1/2}$ and $\mathbf{U}$ by $\mathbf{U}(\mathbf{U}\mathbf{U}^*)^{-1/2}$ without modifying the model.

In the following, we define $\mathbf{Q}(z)$ as the resolvent of matrix $\mathbf{\Sigma}\mathbf{\Sigma}^*$ defined by $\mathbf{Q}(z) = (\mathbf{\Sigma}\mathbf{\Sigma}^* - z\mathbf{I}_{ML})^{-1}$, and remark that the estimated Wiener filter $\hat{\mathbf{g}}_\lambda$ can be written as

$$\hat{\mathbf{g}}_\lambda^{(L)} = \mathbf{Q}(-\lambda) \mathbf{\Sigma}^{(L)} \mathbf{u}^*$$

where $\mathbf{u} = \frac{1}{\sqrt{N}}(s_1, \dots, s_N)$ is the P-th row of matrix $\mathbf{U}$. To evaluate the behaviour of the SINR given by formula (3) when $N \to +\infty$, it is necessary to study $|\mathbf{h}_P^* \hat{\mathbf{g}}_\lambda|^2$, $\|\mathbf{H}_1^* \hat{\mathbf{g}}_\lambda\|^2$, and $\|\hat{\mathbf{g}}_\lambda\|^2$. These terms depend on bilinear forms of matrices $\mathbf{Q}(-\lambda)$ and $\mathbf{Q}(-\lambda)^2$ whose asymptotic behaviour have thus to be evaluated. Model (5) can be interpreted as an additive spiked information plus noise model, in the sense that $\mathbf{H}\mathbf{U}$ is a deterministic matrix whose rank $P + L - 1$ does not scale with N and that $\mathbf{W}$ is a random matrix with zero mean elements. If $\mathbf{W}$ was a Gaussian random matrix with i.i.d. elements, the behaviour of the bilinear forms of $\mathbf{Q}(-\lambda)$ and $\mathbf{Q}(-\lambda)^2$ would appear as a consequence of the results of [2] and [4]. In our context, however, the elements of matrix $\mathbf{W}$ are of course not i.i.d. In the present paper, we use recent results of [6] and [8] to establish that the bilinear forms of $\mathbf{Q}(-\lambda)$ and $\mathbf{Q}(-\lambda)^2$ behave as if $\mathbf{W}$ was a random matrix with i.i.d. elements.

3. BACKGROUND ON THE BEHAVIOUR OF MATRIX $\mathbf{W}\mathbf{W}^*$.

This paper is based on a technical result which establishes that, in a certain sense, the eigenvalues of matrix $\mathbf{W}\mathbf{W}^*$ behave as if the entries of $\mathbf{W}$ were i.i.d. In order to state the corresponding result, we recall that the Marcenko-Pastur distribution (see for example [1]) μ_d with parameters (σ^2, d) is the probability distribution defined by

$$d\mu_d(x) = \delta_0 [1 - d^{-1}]_+ + \frac{\sqrt{(x - x^-)(x^+ - x)}}{2\sigma^2 d \pi x} \mathbb{1}_{[x^-, x^+]}(x)\, dx$$

with $x^- = \sigma^2(1 - \sqrt{d})^2$ and $x^+ = \sigma^2(1 + \sqrt{d})^2$. Here, δ_0 represents the Dirac measure at the origin and $\mathbb{1}_{[x^-, x^+]}(x) = 1$ if $x \in [x^-, x^+]$ and 0 elsewhere. We denote by $m_d(z)$ its Stieltjes transform defined by $m_d(z) = \int_{\mathbb{R}} \frac{\mathrm{d}\mu_d(\lambda)}{\lambda - z}$ and by $\tilde{m}_d(z)$ the function $\tilde{m}_d(z) = d m_d(z) - (1 - d)/z$. These Stieltjes transform satisfy the Marchenko-Pastur canonical equations :

$$\begin{aligned} m_d(z) &= \frac{-1}{z(1 + \sigma^2 \tilde{m}_d(z))} \\ \tilde{m}_d(z) &= \frac{-1}{z(1 + \sigma^2 d \tilde{m}_d(z))} \end{aligned} \quad (6)$$

for each $z \in \mathbb{C} - \mathbb{R}^+$. We denote by $\mathbf{Q}_W(z)$ and $\tilde{\mathbf{Q}}_W(z)$ the resolvent of matrices $\mathbf{W}\mathbf{W}^*$ and $\mathbf{W}^*\mathbf{W}$ defined by $\mathbf{Q}_W(z) = (\mathbf{W}\mathbf{W}^* - z\mathbf{I}_{ML})^{-1}$ and $\tilde{\mathbf{Q}}_W(z) = (\mathbf{W}^*\mathbf{W} - z\mathbf{I}_N)^{-1}$ Then, when $N \to +\infty$, the following result holds.

Proposition 1. *([6],[8]) The eigenvalue distribution of matrix $\mathbf{W}\mathbf{W}^*$ converges almost surely towards the Marcenko-Pastur distribution μ_{cL}. Moreover, if $\mathbf{a}_N, \mathbf{b}_N$ are 2 unit norm (ML)–dimensional deterministic vectors, then it holds that for each $z \in \mathbb{C}^+$*

$$\mathbf{a}_N^* \left(\mathbf{Q}_W(z) - m_{cL}(z)\mathbf{I}\right) \mathbf{b}_N \to 0 \ a.s. \quad (7)$$

Similarly, if $\tilde{\mathbf{a}}_N, \tilde{\mathbf{b}}_N$ are 2 unit norm N–dimensional deterministic vectors, then for each $z \in \mathbb{C}^+$, it holds that

$$\tilde{\mathbf{a}}_N^* \left(\tilde{\mathbf{Q}}_W(z) - \tilde{m}_{cL}(z)\mathbf{I}\right) \tilde{\mathbf{b}}_N \to 0 \ a.s. \quad (8)$$

and that

$$\mathbf{a}_N^* \left(\mathbf{Q}_W(z) \mathbf{W}_N\right) \tilde{\mathbf{b}}_N \to 0 \ a.s. \quad (9)$$

Finally, for each $\epsilon > 0$, convergence properties (7, 8, 9) hold uniformly w.r.t. z on each compact subset of $\mathbb{C} - [0, x^+ + \epsilon]$.

We recall that, roughly speaking, the convergence of the eigenvalue distribution of $\mathbf{W}\mathbf{W}^*$ towards distribution μ_{cL} means that the histogram of the eigenvalues of any realization of $\mathbf{W}\mathbf{W}^*$ tend to accumulate around the graph of the probability density of μ_{cL}. The statements of Proposition 1 are well known when $L = 1$ and that M and N converge towards $+\infty$ at the same rate. The convergence towards μ_{cL} and (7) appear as consequences of the results of [6], (8) and (9) are proved in [8].

4. ASYMPTOTIC BEHAVIOUR OF THE SINR

We recall that $\mathbf{Q}(z) = (\mathbf{\Sigma}\mathbf{\Sigma}^* - z\mathbf{I}_{ML})^{-1}$, and define $\tilde{\mathbf{Q}}(z)$ by $\tilde{\mathbf{Q}}(z) = (\mathbf{\Sigma}^*\mathbf{\Sigma} - z\mathbf{I}_N)^{-1}$. In order to simplify the notations, we omit to mention the dependency w.r.t. complex variable z, and put $d = cL$. We introduce function $\omega(z)$ defined by $\omega(z) = \frac{1}{z m_d(z)\tilde{m}_d(z)}$. Using (6), it is easy to obtain that $m_d(z) = \frac{-1}{\omega(z)+\sigma^2 d}$ and $\tilde{m}_d(z) = \frac{-1}{\omega(z)+\sigma^2}$. If the noise matrix $\mathbf{W}$ was i.i.d., existing results (see [2], [4]) would imply that $\mathbf{Q}(z)$ could be approximated by

$$-\omega(z)m_d(z)\left(\mathbf{H}\mathbf{H}^* - \omega(z)\mathbf{I}\right)^{-1} \quad (10)$$

in the sense that each bilinear form of these 2 matrices have the same asymptotic behaviour. In the same way, $\tilde{\mathbf{Q}}$ could be approximated by

$$-\omega(z)\tilde{m}_d(z)\left(\mathbf{U}^*\mathbf{H}^*\mathbf{H}\mathbf{U} - \omega(z)\mathbf{I}\right)^{-1} \quad (11)$$

Proposition 1 implies that (10) and (11) remain valid although $\mathbf{W}$ is not an i.i.d. matrix. In order to establish this, we express express $\mathbf{Q}$ in terms of $\mathbf{Q}_W$ using the matrix inversion lemma, and use Proposition 1 as in [2]. Due to the lack of space, we refer the reader to a forthcoming extended version of this paper. It is possible to establish the following Proposition.

Proposition 2. *When $N \to +\infty$, the three terms $\|\mathbf{h}_P^*\hat{\mathbf{g}}_\lambda\|^2$, $\|\mathbf{H}_1^*\hat{\mathbf{g}}_\lambda\|^2$ et $\|\hat{\mathbf{g}}_\lambda\|^2$ can be approximated (i.e. have the same almost sure behaviour) by the following deterministic quantities:*

- $$\|\mathbf{h}_P^*\hat{\mathbf{g}}_\lambda\|^2 \simeq \left(\mathbf{h}_P^*(\mathbf{H}\mathbf{H}^* - \omega(-\lambda)\mathbf{I})^{-1}\mathbf{h}_P\right)^2 \quad (12)$$

- $$\|\mathbf{H}_1^*\hat{\mathbf{g}}_\lambda\|^2 \simeq \omega(-\lambda)\mathbf{h}_P^*(\mathbf{H}\mathbf{H}^* - \omega(-\lambda)\mathbf{I})^{-2}\mathbf{h}_P + \mathbf{h}_P^*(\mathbf{H}\mathbf{H}^* - \omega(-\lambda)\mathbf{I})^{-1}\mathbf{h}_P\left(1 - \mathbf{h}_P^*(\mathbf{H}\mathbf{H}^* - \omega(-\lambda)\mathbf{I})^{-1}\mathbf{h}_P\right) \quad (13)$$

- $$\|\hat{\mathbf{g}}_\lambda\|^2 \simeq \frac{\sigma^2 d}{\omega(-\lambda) - \sigma^4 d}\left(1 - \mathbf{h}_P^*(\mathbf{H}\mathbf{H}^* - \omega(-\lambda)\mathbf{I})^{-1}\mathbf{h}_P\right) + \frac{\omega(-\lambda)(\sigma^2 d + \omega(-\lambda))}{\omega(-\lambda) - \sigma^4 d}\mathbf{h}_P^*(\mathbf{H}\mathbf{H}^* - \omega(-\lambda)\mathbf{I})^{-2}\mathbf{h}_P \quad (14)$$

Moreover, if we introduce $\alpha(\lambda) = \mathbf{h}_P^*(\mathbf{H}\mathbf{H}^* - \omega(-\lambda)\mathbf{I})^{-1}\mathbf{h}_P$ and $\beta(\lambda) = \mathbf{h}_P^*(\mathbf{H}\mathbf{H}^* - \omega(-\lambda)\mathbf{I})^{-2}\mathbf{h}_P$, it holds that

$$\mathit{SINR}(\hat{\mathbf{g}}_\lambda^{(L)}) - \phi(\lambda) \to 0 \quad (15)$$

almost surely, where $\phi(\lambda)$ is defined by

$$\phi(\lambda) = \frac{\alpha(\lambda)^2}{[1-\alpha(\lambda)][(\alpha(\lambda) + \frac{\sigma^4 d}{\omega^2(-\lambda)-\sigma^4 d}] + \frac{\omega^2(-\lambda)(\omega(-\lambda)+\sigma^2)}{\omega^2(-\lambda)-\sigma^4 d}\beta(\lambda)} \quad (16)$$

When $cL < 1$, it is possible to consider the case where $\lambda = 0$. Using the observation that $\omega(0) = -\sigma^2$, we obtain immediately that

$$\phi(0) = \gamma\,\frac{(1-cL)\gamma}{\gamma + cL} \quad (17)$$

where γ is the SINR corresponding to the true Wiener filter (see formula 4). Consequently, the estimation of the Wiener filter by $\hat{\mathbf{g}}_{(0)}^L$ produces a SINR loss equal to $(1-cL)\frac{\gamma}{\gamma+cL}$, which, of course, is considerable when cL is close from 1. As shown below, the use of a convenient regularization coefficient allows to improve considerably the SINR.

5. CONSISTENT ESTIMATORS OF THE SINR

It is clear that function $\lambda \to \phi(\lambda)$ depends on matrix $\mathbf{H}$ which is unknown. We establish in this section that it is possible to estimate $\phi(\lambda)$ consistently for each $\lambda > 0$. For this, it is sufficient to estimate $\alpha(\lambda)$ and $\beta(\lambda)$ (see (16)). It is easy to see that

$$\alpha(\lambda) = 1 + \omega(-\lambda)((\mathbf{H}^*\mathbf{H} - \omega(-\lambda)\mathbf{I})^{-1})_{P,P}$$

By (10) and (11), $((\mathbf{H}^*\mathbf{H} - \omega(-\lambda)\mathbf{I})^{-1})_{P,P}$ can be estimated by $\frac{\mathbf{u}\tilde{\mathbf{Q}}(-\lambda)\mathbf{u}^*}{\omega(-\lambda)\tilde{m}_d(-\lambda)}$. Thus,

$$\hat{\alpha}(\lambda) = 1 - \frac{\mathbf{u}\tilde{\mathbf{Q}}(-\lambda)\mathbf{u}^*}{\tilde{m}_d(-\lambda)} \quad (18)$$

is a consistent estimate of $\alpha(\lambda)$. In order to obtain an estimator $\hat{\beta}(\lambda)$ of $\beta(\lambda)$, we observe that

$$\mathbf{h}_P^*(\mathbf{H}\mathbf{H}^* - \omega(-\lambda)\mathbf{I})^{-2}\mathbf{h}_P = ((\mathbf{H}^*\mathbf{H} - \omega(-\lambda)\mathbf{I})^{-1})_{P,P} + \omega(-\lambda)((\mathbf{H}^*\mathbf{H} - \omega(-\lambda)\mathbf{I})^{-2})_{P,P}$$

By (11), $\mathbf{U}\tilde{\mathbf{Q}}\mathbf{U}^* \simeq \tilde{m}_d(z)\omega(z)\left(\mathbf{H}^*\mathbf{H} - \omega(z)\mathbf{I}\right)^{-1}$. Multiplying by z and taking the derivative w.r.t. z leads to the conclusion that it is possible to estimate consistently $(\mathbf{H}\mathbf{H}^* - \omega(-\lambda)\mathbf{I})^{-2})_{P,P}$ by:

$$\frac{\omega^2(-\lambda) - \sigma^4 d}{\omega^2(-\lambda)(\omega(-\lambda) + \sigma^2 d)}\left[\mathbf{u}\left(\frac{d}{dz}\Big|_{z=-\lambda} z\tilde{\mathbf{Q}}(z)\right)\mathbf{u}^* - \frac{\omega^2(-\lambda)}{\omega^2(-\lambda) - \sigma^4 d}(\mathbf{H}^*\mathbf{H} - \omega(-\lambda)\mathbf{I})^{-1})_{p,p}\right]$$

From this, we obtain that $\beta(\lambda)$ can be estimated consistently by the term $\hat{\beta}(\lambda)$ defined by

$$\hat{\beta}(\lambda) = -\frac{\sigma^2 d}{(\omega(-\lambda) + \sigma^2 d)^2}\lambda\mathbf{u}\tilde{\mathbf{Q}}(-\lambda)\mathbf{u}^* + \frac{\omega^2(-\lambda) - \sigma^4 d}{\omega(-\lambda)(\omega(-\lambda) + \sigma^2 d)}\left[\mathbf{u}(\tilde{\mathbf{Q}}(-\lambda) - \lambda\tilde{\mathbf{Q}}^2(-\lambda)\mathbf{u}^*)\right] \quad (19)$$

Replacing $\alpha(\lambda)$ and $\beta(\lambda)$ by $\hat{\alpha}(\lambda)$ and $\hat{\beta}(\lambda)$ in formula (16), we obtain immediately a consistent estimator $\hat{\phi}(\lambda)$ of the asymptotic SINR $\phi(\lambda)$. Moreover, it is possible to establish that function $\phi(\lambda) - \hat{\phi}(\lambda)$ converges uniformly towards 0 on each compact subset of $\mathbb{R}_*^+$. Therefore, if we denote by λ_{opt} and $\hat{\lambda}_{opt}$ the argument of the maximum of ϕ and $\hat{\phi}$ on a fixed compact of $\mathbb{R}_*^+$, it holds that $\lambda_{opt} - \hat{\lambda}_{opt} \to 0$. Therefore, maximizing function $\lambda \to \hat{\phi}(\lambda)$ allows to estimate a regularization parameter for which the true asymptotic SINR $\phi(\lambda)$ is maximum. We also notice that this approach allows to choose the smoothing factor L: it is sufficient to evaluate $\hat{\phi}(\hat{\lambda}_{opt})$ for each choice of L, and to select the smoothing factor for which the latter term is maximum. This is of course not a computationally efficient procedure because it needs to evaluate matrix $\mathbf{Q}(-\lambda)$ et $\tilde{\mathbf{Q}}(-\lambda)$ for each λ and each integer L.

6. NUMERICAL EXPERIMENTS.

In this section, we provide numerical simulations illustrating the results given in the previous sections. We first illustrate the accuracy of the approximation $\text{SINR}(\hat{g}_\lambda) \simeq \phi(\lambda)$ where we recall that $\text{SINR}(\hat{g}_\lambda)$ is the true SINR defined by (3). Matrix $(\mathbf{h}_0, ..., \mathbf{h}_{P-1})$ is a realization of a normalized version (so as to obtain a Frobenius norm equal to 1) of random matrix $(\mathbf{a}(\theta_0), ..., \mathbf{a}(\theta_{P-1}))$, with $\mathbf{a}(\theta) = \frac{1}{\sqrt{M}}(1, ..., e^{i(M-1)\theta})^T$, and where the angles are drawn uniformly on $[0, 2\pi]$. The sequence $(s_n)_{n=1,...,N}$ is a realization of an i.i.d ± 1 sequence with probability $1/2$. The signal to noise ratio SNR is thus equal to $1/\sigma^2$. In the following experiments, $N = 200, M = 40$ and $P = 5$. In figure 1, SNR is equal to 8dB, $L = 5$, and we evaluate by Monte-Carlo simulations (10.000 realizations are generated) function $\lambda \to \text{SINR}(\hat{g}_\lambda^{(L)})$. We represent the graph of the function $\phi(\lambda)$ along with 2 plots representing the lower and upper bounds of the 95% confidence interval of $\lambda \to \text{SINR}(\hat{g}_\lambda^{(L)})$. We can notice that the 3 graphs are close one from each other.

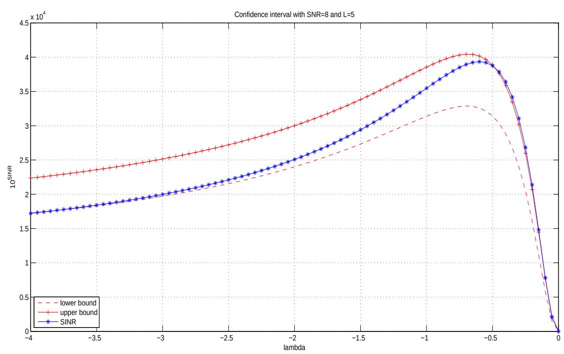

Fig. 1. Confidence region and asymptotic curve of SINR versus λ

We now evaluate the performance of the estimator $\hat{\lambda}_{opt}$ of λ_{opt} and evaluate by Monte-Carlo simulation the root relative least mean squares error of $\phi(\hat{\lambda}_{opt}) - \phi(\lambda_{opt})$. We also evaluate the same quantity, but when λ_{opt} is estimated by other existing schemes: the Ledoit-Wolf estimator ([9]), 3 empirical methods mentioned in [7] to be referred to as M1 ([10], [13, p. 748]), M2 [11], M3 [12] in the figure (2), and the naive estimate obtained by maximizing w.r.t. λ the expression (3) in which matrix $\mathbf{H}$ is replaced by $\mathbf{\Sigma U}^*$, which, of course, is not supposed to be a good estimator when M and N are of the same order of magnitude. The various root relative mean squares errors are given in figure (2) for various values of the smoothing parameter L.

We finally justify that our approach may be used in order to estimate a relevant value of the smoothing parameter L. As mentioned above, we evaluate $\hat{\phi}(\hat{\lambda}_{opt})$ for each possible value of L, and propose to select the value of L for which the latter term is maximum. We keep the same parameters as above. We first represent in Figure 3 function $\lambda \to \phi(\lambda)$ for $L = 1, 2, 3, 4, 5, 6, 7, 8$, and conclude that $L = P = 5$ maximizes $\phi(\lambda_{opt})$, but that $L = 6, 7, 8$ also provide reasonable optimum asymptotic SINR. We can also check that choosing in a convenient way L and λ may improve considerably the SINR.

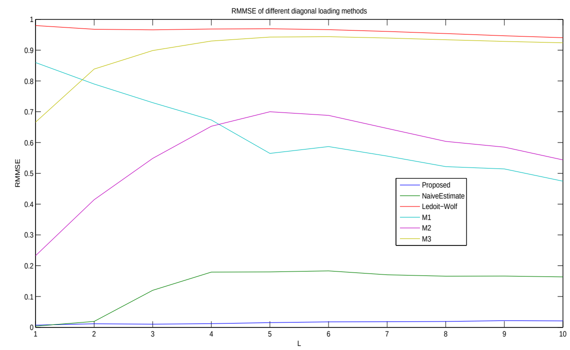

Fig. 2. RMMSE of different diagonal loading methods versus L

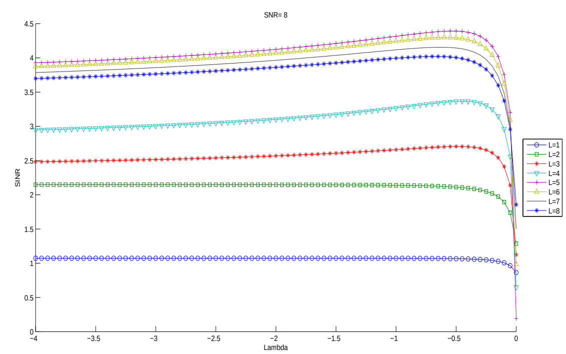

Fig. 3. Asymptotic SINR versus L and λ

7. CONCLUSION

In this paper, we have proposed a new approach to evaluate the loading factor of a regularized estimated spatio-temporal Wiener filter in the context of large systems. Assuming that the number of sensors M and the lenght of the training sequence N are large and of the same order of magnitude, and that the degrees of the Wiener filter and of filter $\mathbf{h}(z)$ do not scale with M, N, we have established that the SINR provided by the estimated regularized Wiener filter converges towards an expression depending on the noise variance, the channel coefficients, and the loading factor. This limit SINR can be consistently estimated, and we have proposed to estimate the regularization factor as the argument of the maximum of the estimated limit SINR. Simulation results have shown that the proposed approach allows to considerably improve the results provided by existing approaches.

8. REFERENCES

[1] L.A Pastur, M.Shcherbina, ”Eigenvalue Distribution of Large Random Matrices”, American Mathematical Society, 2011

[2] F. Benaych-Georges, R.R. Nadakuditi,”The singular values and vectors of low rank perturbations of large rectangular random matrices”, J. Multivariate Anal., Vol. 111 (2012), 120–135.

[3] R. Couillet, M. McKay, ”Large Dimensional Analysis and Optimization of Robust Shrinkage Covariance Matrix Estimators”, Elsevier Journal of Multivariate Analysis, vol. 131, pp. 99-120, 2014

[4] W. Hachem, P. Loubaton, J. Najim, X. Mestre, P. Vallet ”A subspace estimator for fixed rank perturbations of large random matrices”, Journal of Multivariate Analysis, 114 (2013), pp. 427-447, also available on Arxiv (arXiv:1106.1497).

[5] R. Horn, C.R. Johnson, ”Matrix Analysis”, 2nd. Ed., Cambridge Univ. Press, 2013.

[6] P. Loubaton, ”On the almost sure location of the singular values of certain Gaussian block-Hankel large random matrices”, to appear in J. of Theoretical Probability, published on line at http://link.springer.com/article/10.1007/s10959-015-0614-z, also available on Arxiv (http://arxiv.org/abs/1405.2006)

[7] X. Mestre, M. A. Lagunas. ”Finite sample size effect on Minimum Variance beamformers: optimum diagonal loading factor for large arrays”. IEEE Transactions on Signal Processing, vol. 54, no. 1, pp. 69-82, January 2006.

[8] G.T.Pham, P.Loubaton, P.Vallet, ”Performance analysis of spatial smoothing schemes in the context of large arrays”, to appear in IEEE Trans. on Signal Processing, also available on Arxiv (arXiv:1503.08196).

[9] O.Ledoit, M.Wolf, ”A Well-Conditioned Estimator For Large-Dimensional Covariance Matrices”, Journal of Multivariate Analysis 88 (2004) 365411.

[10] E. Hung and R. Turner, A fast beamforming algorithm for large arrays, IEEE Transactions on Aerospace and Electronic Systems, vol. 19, pp. 598607, Jul. 1983.

[11] N.Ma and J. Goh, Efficient method to determine diagonal loading value, in Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (Vol. V), pp. 341344, 2003.

[12] Y. Kim, S. Pillai, and J. Guerci, Optimal loading factor for minimal sample support space-time adaptive radar, in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, (Seattle), pp. 25052508, 1998.

[13] H. Van Trees, Optimum Array Processing. New York: John Wiley and Sons, 2002. Part IV of ”Detection, Estimation and Modulation Theory”.

[14] L. Yang, R. Couillet, M. McKay, ”A Robust Statistics Approach to Minimum Variance Portfolio Optimization”, IEEE Transactions on Signal Processing, vol. 63, no. 24, pp. 6684–6697, 2015