Accepted version of the article published in 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), June 2023. DOI:10.1109/ICASSP49357.2023.10095018. Available at: http://ieeexplore.ieee.org

DATA DRIVEN JOINT SENSOR FUSION AND REGRESSION BASED ON GEOMETRIC MEAN SQUARED ERROR

Carlos Alejandro Lopez, Student Member, IEEE, and *Jaume Riba,* Senior Member, IEEE

Signal Theory and Communications Department, Technical University of Catalonia (SPCOM/UPC)
{carlos.alejandro.lopez, jaume.riba}@upc.edu

ABSTRACT

This paper explores the problem of estimating a temporal series measured from multiple independent sensors with unequal and stationary measurement errors with unknown variances. By formulating the data fusion problem as a joint Maximum Likelihood estimation of sensor covariances and a fusion rule, a batch data driven method is derived involving a residual covariance determinant minimization of a diagonal matrix. It is shown that yielding useful learning from data with good generalization properties in the joint regression and fusion approach requires the assumption of some structure on the sensor noises and/or on the temporal series to be estimated. An efficient data driven algorithm is proposed to obtain the best linear sensor combiner, whose performance is numerically analyzed and compared with the Cramer-Rao Lower Bound of the estimated parameters.

***Index Terms*— Blind data fusion, Geometric Mean Squared Error (GMSE), Inverse Wishart distribution (IW), Sensor networks, Non-convex Optimization**

1. INTRODUCTION

Multimodal data fusion is an expanding topic of great interest in the Signal Processing community due to the increasing availability of additional sources of information in several applications that range from multisensor processing in IoT applications [1], audiovisual signal processing [2] or biomedical applications [3], to name a few. Other topics are also tightly related to the multimodal methodology such as in Portfolio Optimization theory [4], where the Portfolio selection can be seen as an estimation based on the combined information of multiple assets.

The general objective of multimodal data processing is to improve the capabilities of classical Signal Processing algorithms by means of exploiting the diversity that appears when several datasets are processed jointly instead of separately. The rationale behind exploiting the diversity found in an ensemble of related datasets comes from the general conception that the joint global dataset carries more information than the sum of independent datasets [5].

Multisensor fusion [6] is an instance of multimodal data fusion and the motivating context of this work. The structure of multisensor fusion problems induces several issues which make them a challenging task. We propose a solution to a known challenge in this framework: identifying differences in quality of sensors while measuring a time variant phenomenon. Other challenges, such as self-calibration, are not considered for clarity of exposition.

This work has been supported by the Spanish Ministry of Science and Innovation through project RODIN (PID2019-105717RB-C22 / AEI / 10.13039/501100011033) and by the Catalan Government (AGAUR) under grant 2017 SGR 578.

The main purpose of this paper is the study of how the spatial diversity is capable of enhancing the performance in joint blind fusion and regression problems. Our approach can be summarized as:

1. Derivation of the Geometric Mean Squared Error (GMSE) criterion for optimal data driven sensor fusion by considering a joint Maximum Likelihood (ML) estimation, which is a known criterion in the literature [7]. However, it is known that GMSE induces sparse solutions, which is an undesired feature in the estimation of multisensor fusion policies as they reject the estimation of informative sensors.
2. A solution to overcome the sparse property of the GMSE criterion in the form of a regularization. We show that the Inverse Wishart (IW) conjugate prior [8] is capable of avoiding an ill conditioned problem.
3. An efficient algorithm to optimize both criteria (GMSE and IW regularized GMSE) by means of iterative local upper bounds.

The particular application that is tightly related to our work is the interference and multipath mitigation in multiple antenna receivers [9]. We show in this paper that the joint estimation of the interferent sources covariance matrix and the system time delays at the receiver in [9] is dual to the joint estimation of the intersensor covariance and the multisensor fusion policy.

2. PROBLEM STATEMENT

Consider N samples of a time series $x(n)$ associated to some phenomenon measured simultaneously by M calibrated sensors:

$$\mathbf{Y} = \mathbf{1}\mathbf{u}^T\mathbf{B}^T + \mathbf{W}, \tag{1}$$

where $\mathbf{Y}, \mathbf{W} \in \mathbb{R}^{M\times N}$, $\mathbf{u} \in \mathbb{R}^D$ are the latent variables (presumably random but considered deterministic and unknown for a particular realization), $\mathbf{B} \in \mathbb{R}^{N\times D}$ with $D \leq N$ is a known matrix of regressors and $\mathbf{x} = \mathbf{B}\mathbf{u}$ is the temporal evolution of the phenomenon. The second order statistics of the noise matrix, $\mathbf{W}$, are given by $\mathbf{Q} = \frac{1}{N}E\left[\mathbf{W}\mathbf{W}^T\right]$, where $\mathbf{Q}$ is referred to as the intersensor noise covariance matrix. We consider the prior knowledge that the noise covariance matrix is diagonal, i.e. $\mathbf{Q} = \text{diag}(\mathbf{q})$ where $\mathbf{q} = [q_1, ..., q_M]^T \succeq 0$ is an unknown vector. The assumption of a diagonal covariance matrix has sense when the sensors are interfered by independent phenomena [10]. The more general case of correlated measurement noise left for future work, as it opens a more challenging scenario.

Regarding the linear dynamical model in (1) for $\mathbf{x}$, even when the actual dynamics are a non-linear function of the latent variables, choosing a linear function is a common practice due to the fact that it is the simplest approximation yielding a reasonable fitting to the data. A classical example of linear approximations in sensor fusion can be found in the known Extended Kalman Filter solutions [11].

Besides, many time series can be well approximated by a class of so-called time series of finite rank [12], motivating the consideration of a generic and low dimensional matrix of regressors in (1). More generally, under the approach proposed in [13], D can be interpreted either as the complexity of a linear model that fits the data, or as a model parameter capable of trading off between accuracy and complexity.

Based on this linear model for the temporal evolution of the measurements, a time redundancy coefficient is defined:

$$\alpha = \frac{D}{N} \leq 1, \tag{2}$$

which is assumed constant for a particular application and sampling rate of sensor signals.

With the previous setting, the multisensor joint blind data fusion and regression problem is cast as estimating $\mathbf{u}$ from the data $\mathbf{Y}$ on the model in (1) where the noise covariance matrix, $\mathbf{Q}$, is unknown and the regression matrix, $\mathbf{B}$, is known. With the proposed formulation, we aim at jointly exploiting the features emerging in the data in time and space domains simultaneously, instead of performing a two-step approach with sensor-by-sensor extraction first and fusion of features later. This particular manner of posing the problem is motivated by the enablement of context awareness, which is in the core of information fusion problems.

2.1. Model driven Data Fusion

In order to establish a benchmark for assessing the performance of the data driven fusion algorithm to be developed in Section 3, let us first assume a full knowledge of $\mathbf{Q}$ with the purpose of unveiling the structure of the benchmark joint fusion rule for the model defined in (1). Assuming statistical independence of the columns of the noise matrix, $\mathbf{W}$, the joint log-likelihood function of the data, up to additive constants, yields the following expression:

$$\ell_{ML}(\mathbf{Y}|\mathbf{Q}) = -\frac{N}{2}\left(\mathrm{tr}(\mathbf{Q}^{-1}\hat{\mathbf{Q}}) + \ln(\det(\mathbf{Q}))\right), \tag{3}$$

where $\hat{\mathbf{Q}}$ is the residuals sample covariance:

$$\hat{\mathbf{Q}} = \frac{1}{N}(\mathbf{Y} - \mathbf{1}\mathbf{u}^T\mathbf{B}^T)(\mathbf{Y} - \mathbf{1}\mathbf{u}^T\mathbf{B}^T)^T. \tag{4}$$

It is well-known in estimation theory that the ML estimator of $\mathbf{u}$ is efficient and the solution for a known $\mathbf{Q}$ is:

$$\hat{\mathbf{u}} = \mathbf{B}^{\dagger}\mathbf{Y}^T\mathbf{f}_b, \tag{5}$$

where $\mathbf{f}_b = \frac{\mathbf{Q}^{-1}\mathbf{1}}{\mathbf{1}^T\mathbf{Q}^{-1}\mathbf{1}}$ is the benchmark fusion rule and $\mathbf{B}^{\dagger} = (\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$ is the Moore-Penrose pseudoinverse of $\mathbf{B}$. As a consequence, the closed-form ML solution for the phenomenon is:

$$\hat{\mathbf{x}} = \mathbf{B}\hat{\mathbf{u}} = \mathbf{Y}_1^T\mathbf{f}_b, \tag{6}$$

where:

$$\mathbf{Y}_1 = \mathbf{Y}\mathbf{P}_1, \tag{7}$$

being $\mathbf{P}_1 = \mathbf{B}\mathbf{B}^{\dagger}$ the projector onto the subspace spanned by $\mathbf{B}$. Clearly, the fusion and regression are decoupled in the case of known covariance as the sensor by sensor regression in (7) does not have any impact on the structure of the fusion in (6). What is more, the previous estimator of $\hat{\mathbf{x}}$ is also known to be the Best Linear Unbiased Estimator (BLUE), which means that assuming normality on the measurement errors is not a huge restriction when linear fusion rules are pursued (as those derived in this paper).

3. PROPOSAL: DATA DRIVEN APPROACH TO DATA FUSION

With the previous considerations, the problem of data driven data fusion can be reformulated as the estimation of the best fusion rule, $\mathbf{f}$, given the input data, where now $\mathbf{Q}$ constitutes an additional unknown. The motive behind reformulating the problem is to derive a more convenient cost function for the joint blind fusion and regression from (3) when $\mathbf{Q}$ is unknown. Indeed, the reformulation of (3) is achieved by computing the intersensor sample covariance estimation from the decomposition of the available data into two orthogonal terms $\mathbf{Y} = \mathbf{Y}_0 + \mathbf{Y}_1$ belonging to the noise and signal subspaces, respectively. Notice that both terms in this decomposition are uncorrelated due to the orthogonality between both subspaces and thus the sample covariance of $\mathbf{Y}$ can be expressed as the sum of covariances of each term. Hence, (4) can be rewritten by plugging an estimation of $\mathbf{u}$ in terms of an arbitrary fusion rule $\mathbf{f}$ in (4), inspired on the structure of (5) and considering the aforementioned decomposition of the data, as:

$$\hat{\mathbf{Q}} = \frac{1}{N}\left(\mathbf{Y}_0 + (\mathbf{Y}_1 - \mathbf{1}\mathbf{f}^T\mathbf{Y}_1)\right)\left(\mathbf{Y}_0 + (\mathbf{Y}_1 - \mathbf{1}\mathbf{f}^T\mathbf{Y}_1)\right)^T = (1-\alpha)\hat{\mathbf{Q}}_0 + \alpha(\mathbf{I} - \mathbf{1}\mathbf{f}^T)\hat{\mathbf{Q}}_1(\mathbf{I} - \mathbf{1}\mathbf{f}^T)^T, \tag{8}$$

where:

$$\mathbf{Q}_i \triangleq \frac{1}{Di + (N-D)(1-i)}\mathbf{Y}\mathbf{P}_i\mathbf{Y}^T \quad i = 0, 1, \tag{9}$$

and $\mathbf{P}_0 = \mathbf{I} - \mathbf{P}_1$ is the projector onto the noise subspace, orthogonal to $\mathbf{P}_1$ defined in (7). It is clear from (8) that $\mathbf{Q}_0$ and $\mathbf{Q}_1$ are the out-space and in-space data correlation matrices, respectively. Note that the overall sample covariance estimator in (8) is still consistent with respect to $\mathbf{Q}$ as long as $\mathbf{f}^T\mathbf{1} = 1$. Although the projectors involved in (9) are non data driven, they retain the information of the subspace spanned by the regressors, which is invariant to the particular values of $\mathbf{B}$. In that sense, the developed method is not sensitive to an exact modeling, since it is exploiting only structural model prior knowledge instead of exact values of their latent parameters.

3.1. Geometric Mean Squared Error criterion

Assuming that $\mathbf{Q}$ is unknown and considering the sample covariance estimation in (8), the log-likelihood function in (3) defines a joint optimization problem with respect to $\mathbf{f}$ and $\mathbf{Q}$, being the core of our proposal. Given the fact that the structure of the intersensor covariance matrix is diagonal, the maximization of (3) with respect to $\mathbf{Q}$ is derived explicitly as:

$$\frac{\partial}{\partial q_m}\ell_{ML}(\mathbf{Y}|\mathrm{diag}(\mathbf{q})) = -q_m^{-2}\left[\hat{\mathbf{Q}}\right]_{m,m} + q_m^{-1} = 0, \tag{10}$$

yielding the ML estimation of $\mathbf{Q}$:

$$\hat{\mathbf{Q}}_{ML}(\mathbf{f}) = \hat{\mathbf{Q}} \odot \mathbf{I}., \tag{11}$$

where $\mathbf{I}$ is the identity matrix and $\odot$ is the elementwise product. It can be seen that the estimate of $\mathbf{Q}$ is ensured to be diagonal by ignoring the off-diagonal entries of $\hat{\mathbf{Q}}$. By plugging (11) into (3) and ignoring additive and negative multiplicative constants, the resulting criterion to minimize is:

$$\ell_{GMSE}(\mathbf{f}) = \ln(\det(\hat{\mathbf{Q}}_{ML}(\mathbf{f}))), \tag{12}$$

involving a minimization of the sample covariance with respect to $\mathbf{f}$ and thus coupling the fusion and regression due to (8). As the determinant of a diagonal matrix is a product of sample residuals, (12) becomes a minimization of the GMSE of sensor noise estimates, which has been also formulated in other contexts as a robust criterion [7].

Still, the criterion in (12) is ill-posed in the case of no temporal redundancy, i.e. $\alpha = 1$. In order to illustrate this latter scenario, the cost function in (12) when $\alpha = 1$ is rewritten as follows:

$$\ell_{GMSE}(\mathbf{f}) = \sum_{m=1}^{M} \ln\left(\sum_{n=1}^{N} |y_m(n) - \sum_{j=1}^{M} f_j y_j(n)|^2\right), \quad (13)$$

where f_i is the i-th component of $\mathbf{f}$ and $y_m(n) = [\mathbf{Y}]_{m,n}$ denotes the m, n-th entry of $\mathbf{Y}$. It is now evidenced from (13) that any solution selecting only a single sensor, i.e. $f_i = \delta_{m-i}$ where δ_n denotes the Kronecker delta, achieves effortlessly the minimum of the criterion by nulling the overall determinant. This is also the case when α is close to 1, being easily verified by simulations. In that case, the undesired sparsity in $\mathbf{f}$ emerges as a consequence of the data randomness along with the fact of being too much close to the ill-posed scenario in (13). As a consequence, in order to derive a robust data driven solution for the challenging case of α close to 1, it is advisable to regularize the GMSE criterion in (12).

3.2. Regularized GMSE

With the aim of avoiding sparse solutions of $\mathbf{f}$, we propose the utilization of the Inverse Wishart (IW) prior on $\mathbf{Q} \sim \mathcal{W}^{-1}(\frac{\beta}{2}\mathbf{I}, \upsilon)$ as a regularization to (3). The rationale behind choosing an IW prior is the initial belief that the variances cannot be too high nor too small, as informed with the considered shape parameter. The IW prior has the advantage of being a conjugate prior of the Gaussian distribution and it is classically used in the MAP estimation of scale parameters such as variances [8]. As a result, this regularization avoids sparse solutions on $\mathbf{f}$ and hence can be seen as an anti sparsity regularization for this problem. The logarithm of the IW probability density function, up to additive constants that do not depend on $\mathbf{Q}$, is expressed as [14]:

$$\ln(f_{\mathbf{Q}}(\mathbf{Q})) = -\left(\frac{\upsilon + M + 1}{2}\ln(\det(\mathbf{Q})) + \frac{\beta}{2}\mathrm{tr}(\mathbf{Q}^{-1})\right), \quad (14)$$

being necessary to derive the log-posterior function. By adding (14) to (3) and rearranging terms, the negated log-posterior (up to additive constants) defining a joint optimization problem with respect to $\mathbf{f}$ and $\mathbf{Q}$ is expressed as:

$$\ell_{MAP}(\mathbf{Q}|\mathbf{Y}, \mathbf{f}) = \frac{N}{2}\left(C\ln(\det(\mathbf{Q})) + \mathrm{tr}\left(\mathbf{Q}^{-1}\tilde{\mathbf{Q}}\right)\right), \quad (15)$$

where $\tilde{\mathbf{Q}} \triangleq \frac{\beta}{N}\mathbf{I} + \hat{\mathbf{Q}}$ and $C = \left(1 + \frac{\upsilon+M+1}{N}\right)$. Then, the Maximum a Posteriori (MAP) estimation of the covariance derived from the minimization of (15) with respect to $\mathbf{Q}$, recalling the diagonal structure of the covariance of the intersensor covariance matrix, is:

$$\hat{\mathbf{Q}}_{MAP} = \frac{1}{C}\left(\frac{\beta}{N}\mathbf{I} + \hat{\mathbf{Q}}_{ML}(\mathbf{f})\right). \quad (16)$$

With a similar procedure as in the GMSE derivation, plugging (16) onto (15) and ignoring additive constants, the IW regularized GMSE criterion to minimize yields:

$$\ell_{RGMSE}(\mathbf{f}) = \ln\left(\det\left(\frac{\beta}{N}\mathbf{I} + \hat{\mathbf{Q}}_{ML}(\mathbf{f})\right)\right), \quad (17)$$

with $\hat{\mathbf{Q}}_{ML}(\mathbf{f})$ given in (11). It is noted that (17) is independent of the degrees of freedom of the IW distribution, υ. Note that the RGMSE cost function in (17) is equivalent to the non-regularized GMSE in two cases. The first case is obtained for $\beta = 0$, yielding a non-informative prior. Alternatively, as the number of observations tends to infinity, the regularizing term also tends to zero, so the IW regularization converges asymptotically to the GMSE solution.

3.3. Iterative solution

Given that (12) is a particular case of (17), we focus on the derivation of an efficient iterative solution to the RGMSE criterion. Our proposal consists on the optimization of local upper bounds [15] of the log-determinant in (17), as it is a way to avoid the NP-hardness that is found when one considers the minimization of a concave objective function [16].

Assuming, $\mathbf{f}_k$, a feasible fusion policy as a guess, the first order Taylor expansion is known to fulfill, $\forall \mathbf{f}, \mathbf{f}_k$:

$$\ell_{RGMSE}(\mathbf{f}) \leq \ln(\det(\mathbf{W}_k^{-1})) + \mathrm{tr}\left(\mathbf{W}_k\left(\hat{\mathbf{Q}}_{ML}(\mathbf{f}) - \hat{\mathbf{Q}}_{ML}(\mathbf{f}_k)\right)\right), \quad (18)$$

where $\mathbf{W}_k = \left(\frac{\beta}{N}\mathbf{I} + \hat{\mathbf{Q}}_{ML}(\mathbf{f}_k)\right)^{-1}$. By minimizing the right hand side (RHS) of (18), the resulting iterative criterion is equivalent to the minimization of $\ell_{RGMSE}(\mathbf{f})$. After rearranging the RHS terms of (18) that depend on $\mathbf{f}$ and ignoring additive constants [17], the optimization problem for obtaining the next guess, $\mathbf{f}_{k+1}$, yields:

$$\begin{aligned} \mathbf{f}_{k+1} = \arg\min_{\mathbf{f}} & -2\mathbf{1}^T\mathbf{Q}_1\mathbf{W}_k\mathbf{f} + (\mathbf{f}^T\mathbf{Q}_1\mathbf{f})(\mathbf{1}^T\mathbf{W}_k\mathbf{1}), \\ & \text{s.t. } \mathbf{f}^T\mathbf{1} = 1, \end{aligned} \quad (19)$$

whose solution is easily obtained by invoking the Karush-Kuhn-Tucker conditions [18]:

$$\mathbf{f}_{k+1} = \frac{\mathbf{W}_k\mathbf{1}}{\mathbf{1}^T\mathbf{W}_k\mathbf{1}}. \quad (20)$$

The resulting algorithm consists on iterative estimations of $\mathbf{f}_k$ and $\mathbf{W}_k$. Notice that $\mathbf{W}_k$ is an iterative estimation of the intersensor precision matrix, $\mathbf{Q}^{-1}$, so the update equation (20) is mimicking the benchmark fusion rule in a data driven manner.

Algorithm 1: Regularized GMSE algorithm

Data: $\mathbf{Q}_0 = \mathbf{Y}\mathbf{P}_0\mathbf{Y}^T$, $\mathbf{Q}_1 = \mathbf{Y}\mathbf{P}_1\mathbf{Y}^T$
Result: $\mathbf{f}_{opt}$
Initialization: Set $\mathbf{f}_0 = \frac{1}{M}\mathbf{1}$, K_{max}
for $k = 1, \ldots, K_{max}$ **do**
 $\mathbf{A}_k = \left((1-\alpha)\mathbf{Q}_0 + \alpha(\mathbf{I} - \mathbf{1}\mathbf{f}_{k-1}^T)\mathbf{Q}_1(\mathbf{I} - \mathbf{1}\mathbf{f}_{k-1}^T)^T\right) \odot \mathbf{I}$
 $\mathbf{W}_k = \left(\frac{\beta}{N}\mathbf{I} + \mathbf{A}_k\right)^{-1}$
 $\mathbf{f}_k = \frac{\mathbf{W}_k\mathbf{1}}{\mathbf{1}^T\mathbf{W}_k\mathbf{1}}$
 if $\frac{(\mathbf{f}_k - \mathbf{f}_{k-1})^T\mathbf{W}_k^{-1}(\mathbf{f}_k - \mathbf{f}_{k-1})}{\mathbf{f}_{k-1}^T\mathbf{W}_k^{-1}\mathbf{f}_{k-1}} < \epsilon$ **then**
  End
end

Lastly, the remaining consideration to complete the description of the proposed algorithm is the initialization, $\mathbf{f}_0$. It is a known result in the field of GNSS interference and multipath mitigation that in order to ensure the overall convergence of an iterative algorithm from (18), $\mathbf{W}_0$ must be a consistent estimator [19]. In fact, taking into consideration the discussion from (8), choosing any $\mathbf{f}_0$ fulfilling the constraint guarantees the convergence. To sum up, in Algorithm 1 we outline the proposed solution to the joint sensor fusion and regression problems where the use of a stopping criterion is implemented for computational efficiency. The presented stopping criterion is a relative distance between $\mathbf{f}_k$ and $\mathbf{f}_{k-1}$ weighted with $\mathbf{W}_k^{-1}$ in such a way that it takes into account the relative quality between different sensors.

4. NUMERICAL RESULTS

In order to evaluate the performance and limitations of the presented approaches, we simulate sensor networks which have two kinds of sensors: informative and corrupted sensors. In this sense, this distinction can be seen on the vector of variances:

$$\mathbf{q} = [q_1, ..., q_1, q_2, ..., q_2]^T, \tag{21}$$

where $q_1 < q_2$ and there are M_g sensors with variance q_1 and $M - M_g$ sensors with variance q_2. The motivation behind this setting is to emphasize the difference of sensors quality such that additional corrupted sources of information contaminate the overall data fusion. In this way, the difference in performance of suboptimal fusion rules is accentuated.

As a performance metric in the simulations, we introduce the use of a relative measure between the Cramer-Rao Lower Bound (CRLB) and the evaluated fusion rule variance. Having the ideas presented in Section 2 in mind, it can be easily shown that the variance of an arbitrary fusion policy, $\mathbf{f}$, yields:

$$\gamma(\mathbf{f}) \triangleq \frac{1}{N} E\left[||\mathbf{x} - \hat{\mathbf{x}}(\mathbf{f})||^2\right] = \alpha \mathbf{f}^T \mathbf{Q} \mathbf{f}, \tag{22}$$

from which the relative measure is defined as:

$$\mathrm{i}(\mathbf{f}) = \frac{\gamma(\mathbf{f}_b)}{\gamma(\mathbf{f})}, \tag{23}$$

where $\gamma(\mathbf{f}_b)$ is the CRLB. Notice that (23) lies between 0 and 1. The statistical mean of the denominator in (23) is obtained by 1000 MonteCarlo simulations.

On the other hand, the GMSE and the IW regularized GMSE are compared with additional reference fusion rules. The first considered estimator is the naive fusion rule, defined in Algorithm 1 as the initialization $\mathbf{f}_0$, which is the best estimator in the case of equal measurement errors. The idea behind using $\mathbf{f}_0$ is to assess the difficulty of the simulated scenario.

Finally, the remaining reference estimators are the oracles, which are built on the structure of the benchmark fusion rule presented in (6). The blind oracle fusion policy is obtained by estimating the covariance without the presence of any signal in the input data, $\hat{\mathbf{Q}}_b = \frac{1}{N}\mathbf{W}\mathbf{W}^T$, in contrast to the proposed approaches. Similarly, the super oracle fusion rule with known covariance structure is obtained by adding the diagonal constraint, $\hat{\mathbf{Q}}_s = \frac{1}{N}\mathbf{W}\mathbf{W}^T \odot \mathbf{I}$, which is considered the ideal estimation in these simulations.

In Figure 1 we simulate a sensor network which has high spatial diversity, where it is shown that both the GMSE and the RGMSE solutions outperform the blind oracle and they even yield a similar performance to the super oracle fusion policy with known covariance structure, especially the RGMSE solution.

In contrast, in Figure 2 the simulated sensor network has just $M_g = 4$ informative sensors. There is a degradation in performance with respect to the previous setting, being the non-regularized

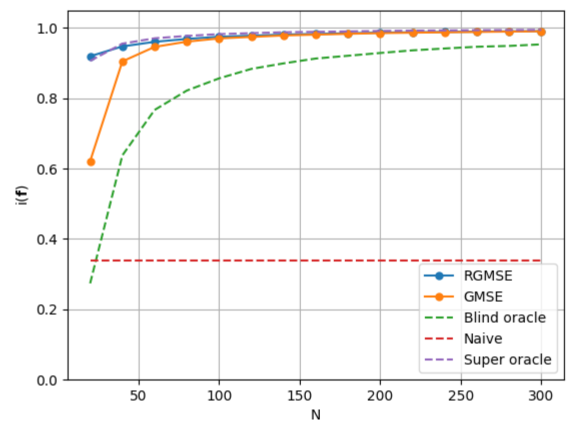

Fig. 1: Fusion performance relative to the benchmark: high spatial diversity.

GMSE criterion the one with a greater loss while the RGMSE solution is still a close competitor to the best possible estimation of $\mathbf{f}$. In spite of that, both criteria have asymptotically better performance than the blind oracle. The comparison between both figures manifests the importance of spatial diversity in a data fusion problem as it enhances the performance of data driven techniques, especially in the small data regime.

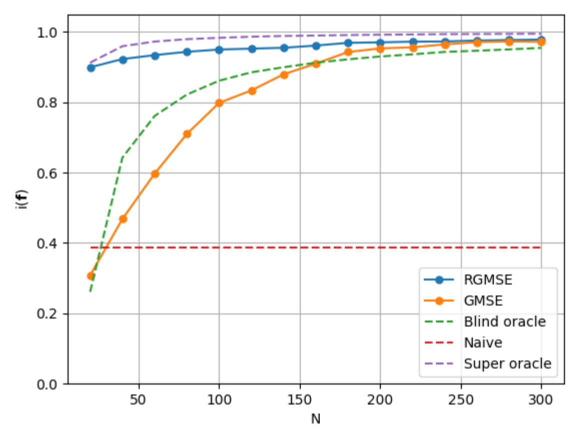

Fig. 2: Fusion performance relative to the benchmark: low spatial diversity.

5. CONCLUSIONS

We show a data driven criterion and its regularized version that yield similar results to the ideal oracle estimator while exhibiting practical properties. We show that it is possible to achieve a computationally efficient data driven algorithm that solves the problem of different quality between sensor measures. We find that providing structure to the estimations, and hence reducing the overall degrees of freedom, improves the performance in the small data regime, whereas classical approaches tend to fail. As a final remark, the fact that the GMSE and RGMSE approaches do not require a training phase is a practical advantage with respect to the oracle estimations.

There is still future work to be done. For instance, the consideration of unknown temporal redundancy (linear or not) with known D is still an open problem. In addition, there is a great interest in the solution of the self-calibrating sensor networks.

6. REFERENCES

[1] Pau Ferrer-Cid, Jose M. Barcelo-Ordinas, Jorge Garcia-Vidal, Anna Ripoll, and Mar Viana, "Multisensor Data Fusion Calibration in IoT Air Pollution Platforms," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 3124–3132, 2020.

[2] Shankar T. Shivappa, Mohan Manubhai Trivedi, and Bhaskar D. Rao, "Audiovisual Information Fusion in Human–Computer Interfaces and Intelligent Environments: A Survey," *Proceedings of the IEEE*, vol. 98, no. 10, pp. 1692–1715, 2010.

[3] Randa M. Abdelmoneem, Eman Shaaban, and Abderrahim Benslimane, "A Survey on Multi-Sensor Fusion Techniques in IoT for Healthcare," in *2018 13th International Conference on Computer Engineering and Systems (ICCES)*, 2018, pp. 157–162.

[4] H. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*, Yale University Press, 2008.

[5] Dana Lahat, Tülay Adali, and Christian Jutten, "Multimodal Data Fusion: An Overview of Methods, Challenges and Prospects," *Proceedings of the IEEE*, vol. 103, 09 2015.

[6] Bahador Khaleghi, Alaa Khamis, Fakhreddine O. Karray, and Saiedeh N. Razavi, "Multisensor data fusion: A review of the state-of-the-art," *Information Fusion*, vol. 14, no. 1, pp. 28–44, 2013.

[7] Prasanna Mayilvahanan, "Estimation of regression coefficients using geometric mean of squared error for single index regression model," *International Journal of Artificial Intelligence and Applications (IJAIA)*, vol. 7, no. 6, pp. 75–84, 2016.

[8] Tom Leonard and John S. J. Hsu, "Bayesian Inference for a Covariance Matrix," *The Annals of Statistics*, vol. 20, no. 4, pp. 1669 – 1696, 1992.

[9] Gonzalo Seco-Granados, Juan A. Fernández-Rubio, and Carles Fernández-Prades, "ML Estimator and Hybrid Beamformer for Multipath and Interference Mitigation in GNSS Receivers," *IEEE Trans. Signal Process*, vol. 53, no. 3, pp. 1194–1208, Mar. 2005.

[10] Muhammad Abu Bakr and Sukhan Lee, "Distributed multisensor data fusion under unknown correlation and data inconsistency," *Sensors*, vol. 17, pp. 2472, 10 2017.

[11] Horaţiu George Todoran and Markus Bader, "Extended Kalman filter (EKF)-based local SLAM in dynamic environments: A framework," in *Advances in Robot Design and Intelligent Control*, pp. 459–469. Springer, 2016.

[12] Jonathan Gillard and Konstantin Usevich, "Structured low-rank matrix completion for forecasting in time series analysis," *International Journal of Forecasting*, vol. 34, no. 4, pp. 582–597, 2018.

[13] I. Markovsky, J. C. Willems, S. Van Huffel, and B. De Moor, *Exact and Approximate Modeling of Linear Systems: A Behavioral Approach*, SIAM, 2006.

[14] Zhiyuan Weng and Petar M. Djurić, "A Bayesian approach to covariance estimation and data fusion," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, 2012, pp. 2352–2356.

[15] M. Fazel, H. Hindi, and S.P. Boyd, "Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices," in *Proceedings of the 2003 American Control Conference, 2003.*, 2003, vol. 3, pp. 2156–2162 vol.3.

[16] R. Horst, "On the global minimization of concave functions," *Operations-Research-Spektrum*, vol. 6, no. 4, pp. 195–205, Dec 1984.

[17] A. Graham, *Kronecker Products and Matrix Calculus with Applications*, Dover Books on Mathematics. Dover Publications, 2018.

[18] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*, Cambridge university press, 2004.

[19] Gonzalo Seco Granados, *Antenna arrays for multipath and interference mitigation in GNSS receivers*, Ph.D. thesis, Universitat Politècnica de Catalunya. Departament de Teoria del Senyal i Comunicacions, 2000, `https://upcommons.upc.edu/handle/2117/94187`.