Efficient Linear Fusion of Partial Estimators

David Luengo[a], Luca Martino[b], Víctor Elvira[c], Mónica Bugallo[d]

[a] *Dep. of Signal Theory and Communications, Universidad Politécnica Madrid (Spain)*
[b] *Image Processing Lab., Universitat de València (Spain)*
[c] *IMT Lille Douai & CRIStAL (UMR CNRS 9189), Villeneuve d'Ascq (France)*
[d] *Dep. of Electrical and Computer Eng., Stony Brook University, NY (USA)*

Abstract

Many signal processing applications require performing statistical inference on large datasets, where computational and/or memory restrictions become an issue. In this big data setting, computing an exact global centralized estimator is often either unfeasible or impractical. Hence, several authors have considered distributed inference approaches, where the data are divided among multiple workers (cores, machines or a combination of both). The computations are then performed in parallel and the resulting partial estimators are finally combined to approximate the intractable global estimator. In this paper, we focus on the scenario where no communication exists among the workers, deriving efficient linear fusion rules for the combination of the distributed estimators. Both a constrained optimization perspective and a Bayesian approach (based on the Bernstein-von Mises theorem and the asymptotic normality of the estimators) are provided for the derivation of the proposed linear fusion rules. We concentrate on finding the minimum mean squared error (MMSE) global estimator, but the developed framework is very general and can be used to combine any type of unbiased partial estimators (not necessarily MMSE partial estimators). Numerical results show the good performance of the algorithms developed, both in problems where analytical expressions can be obtained for the partial estima-

*Corresponding author: David Luengo
Email addresses: david.luengo@upm.es (David Luengo), luca.martino@uv.es (Luca Martino), victor.elvira@telecom-lille.fr (Víctor Elvira), monica.bugallo@stonybrook.edu (Mónica Bugallo)

tors, and in a wireless sensor network localization problem where Monte Carlo methods are used to approximate the partial estimators.

Keywords: distributed estimation, linear fusion, constrained minimization, big data, minimum mean squared error (MMSE) estimators, statistical inference, Bayesian estimation, Bernstein-von Mises theorem, Monte Carlo methods

---

1. Introduction

Estimation theory addresses the problem of inferring a set of unknown variables of interest given a collection of available data [1, 2]. This is a central problem in statistical signal processing, where a parametric model for the data is often assumed and its parameters have to be inferred from the observations [3, 4, 5]. Indeed, even non-parametric approaches typically have a reduced set of hyper-parameters that need to be estimated from the data [6, 7, 8]. Unfortunately, determining the *global estimator* of these parameters using all the available information is often unfeasible or impractical for most real-world scenarios. Many current signal processing applications require performing statistical inference on large datasets, where the amount of data at hand imposes computational and/or storage constraints that impede the global estimation process [9]. Furthermore, even when approximate numerical solutions working directly on the whole dataset can be computed, they may not provide a satisfactory performance either. For example, Monte Carlo (MC) methods are often used to attain asymptotically exact estimators when closed-form analytical expressions cannot be otained [10, 11, 12]. However, large datasets pose a challenge for MC-based estimators, since the posterior density tends to be concentrated on a relatively small portion of the state space as the number of data increases [13]. Consequently, MC algorithms may have trouble locating this area (especially if the dimension of the state space is also large), and thus can lead to a poor performance in practice.

An alternative to *global estimation* consists of dividing the available data into groups of manageable information, and distributing them among multiple

workers (cores, machines or a combination of both). The computations are then performed in parallel (with or without communication among the different workers) and *distributed* or *partial* estimators of the unknown parameters are obtained. In this setting, two extreme situations may arise, namely the multi-core and the multi-machine scenarios. On the one hand, in the *multi-core* case, the estimation is performed using several cores of a single machine (e.g., inside a graphics processing unit [GPU]) and communication among the cores can be considered costless [14, 15]. This approach allows for communication among workers, can provide significant speed-ups (if synchronization issues are properly addressed), and solves the computational cost problem, but not the memory/disk storage bottleneck. On the other hand, in the *multi-machine* case, the estimation is distributed among several machines (typically lying inside a large cluster), and the cost of inter-machine communications cannot be ignored. This approach can alleviate all the issues associated to big data signal processing (i.e., both computational and memory/storage issues), but requires each machine to work independently without any communication among workers (which typically communicate only to the central node at the beginning and the end of their tasks) [16]. Finally, note that a combination of both scenarios often occurs in practice (i.e., a large cluster where each machine may have several cores), thus resulting in situations where a moderate amount of communications may be acceptable.

In this paper, we focus on the scenario where no communication exists among the workers, deriving efficient linear fusion rules for the combination of the partial estimators. Our main goal is finding an optimal combination of these partial estimators that allows us to achieve a performance which is as close as possible to the performance of the global estimator that has access to all the information. We concentrate on minimum mean squared error (MMSE) global estimators, but the developed framework is very general and can be used to combine any type of unbiased partial estimators. In the following, we review related works (Section 1.1), detail our main contributions (Section 1.2) and summarize the structure of the whole paper (Section 1.3).

1.1. Related Works

The fusion of different models or estimators has been studied in many different areas including control, signal processing, econometrics and digital communications. The literature on the subject is rather vast, so here we only mention the most important results related to the problem addressed. On the one hand, a related field in the statistical literature is the combination of forecasts [17]. Indeed, the optimal linear combination for the single parameter case was already derived in [18, 19], a Bayesian perspective was provided in [20], and a general procedure to combine estimators in the multiple parameter case has been proposed very recently in [21]. However, there are two important differences with respect to the scenario addressed here: (1) each forecaster is assumed to have access to the whole dataset; (2) the computational complexity issue is not addressed. Therefore, problems related to the scarcity of data per estimator (when the number of data is large but the ratio data/workers is not so large), such as the so-called *small sample bias* [22], the choice of the appropriate number of partial estimators or the feasibility of the optimal combination rules when the number of parameters to be estimated is also large, have never been investigated in this context as far as we know.

On the other hand, in wireless sensor networks the focus has been on distributed learning/estimation under communication constraints [23, 24]. The optimal linear fusion rule for the multi-dimensional case has also been derived in this context [24, 25], but the focus has been on developing optimal compression rules to restrict the amount of information being transmitted, rather than on obtaining efficient fusion schemes. Unfortunately, this compression is of limited use in the multi-machine learning scenario, since passing messages among multiple machines is expensive regardless of their size. Distributed fusion approaches, obtained by adapting methods developed for graphical models, have also been proposed [26], as well as many different consensus, gossip or diffusion algorithms [27, 28, 29]. However, all of these methods still require a non-negligible amount of communication that constitutes a burden for multi-machine signal processing.

Finally, there is currently a great deal of interest in parallel Bayesian computation using MC methods [30], and a few communication-free parallel Markov chain Monte Carlo (MCMC) algorithms working on disjoint partial datasets have been developed following the so-called *embarrassingly parallel* architecture [31]. In [32], four alternatives were proposed to combine the samples drawn from the partial posteriors using either a Gaussian approximation or importance resampling. Then, [33] derived the optimal linear combination of weights required to obtain samples approximately from the full posterior, noting that the approach is optimal when both the full and the partial posteriors are Gaussian. This was followed by [34], where three different approaches to approximate the full posterior from the partial posteriors were proposed: a simple parametric approach, a non-parametric estimator and a semi-parametric method. In [35], an improvement of the quality of the approximation to the full posterior was proposed by using the Weierstrass transform. A variational aggregation approach has been derived in [36], whereas an approach based on space partitioning and density aggregation has been introduced in [37]. However, none of these previous works addresses the potentially large dimension of the optimal combiners in practical problems, which demands the transmission of large matrices across the network. Furthermore, all of the aforementioned works focus on the generation of valid Monte Carlo samples at the fusion node (thus requiring the transmission of all the samples generated in the distributed nodes, which can be an excessive burden in some environments like wireless sensor networks), whereas we concentrate on the problem of obtaining a good global estimator using a single estimate from each distributed node.

1.2. Main Contributions

The main contribution of this work is the derivation of two novel efficient linear schemes for the fusion of the partial estimators. Although we focus on minimum mean squared error (MMSE) partial estimators throughout the paper, the proposed fusion schemes are independent from the specific approach followed to

obtain those partial estimators (they are only assumed to be unbiased). The motivation comes from the optimal linear combination, which involves the calculation of one weighting matrix per partial estimator and thus may be too computationally demanding for large dimensional systems (both in the number of unknowns and observations), as it requires as many weighting matrices (whose size depends quadratically on the number of unknowns) as partial estimators (whose number is typically a fraction of the number of observations). For instance, in a setting where the number of parameters to be estimated is D and the N observations available are equally distributed among L partial estimators, the optimal linear fusion approach requires computing one $D \times D$ matrix per partial estimator (L matrices and LD^2 parameters in total), which must be estimated from the partial dataset composed of N/L samples. In order to reduce the computational complexity, we propose two linear approaches: the single coefficient MMSE (SC-MMSE) fusion rule, that requires only a single weighting coefficient per partial estimator (i.e., L weights in total), and the independent linear MMSE (IL-MMSE) fusion, which requires one weighting coefficient per parameter and partial estimator (i.e., LD weights in total), respectively.

Another important contribution of the paper is providing both a constrained optimization perspective and a Bayesian point of view (based on the Bernstein-von Mises theorem and the asymptotic normality of the estimators) for the derivation of all the linear fusion rules considered. These two complementary visions help to explain the good performance of the derived fusion rules, even when the normality assumption required by the Bayesian approach is not strictly fulfilled. The optimal linear combination, derived first, provides the global MMSE estimator only when the partial MMSE estimators have a Gaussian distribution. Under certain regularity conditions, this is ensured by the Bernstein-von Mises theorem in the large-sample size limit for each partial estimator (i.e., when N/L is large). However, even when this theorem is not fulfilled and the partial estimators do not follow a Gaussian distribution, the optimal linear fusion rule provides the best linear unbiased estimator (in the sense of minimizing the MSE) given the unbiased partial estimates, as shown by the constrained optimization

formulation. This explains the good performance of the optimal fusion rule even when the underlying distributions are not Gaussian. The efficient SC-MMSE and IL-MMSE linear fusion rules, derived next, can then be seen as the optimal restricted linear fusion rules corresponding to a single coefficient and a diagonal matrix, respectively. From a Bayesian point of view, this is equivalent to assuming independence among all the parameters and equal quality in the estimation of all the parameters, respectively. Again, even if these restrictive conditions are not fulfilled, the constrained optimization perspective ensures us that the derived estimators are the best ones that can be obtained given the restrictions. Moreover, it allows us to see that these fusion rules can still provide a good performance when the constraints are approximately met (e.g., if the dependence among the parameters is weak and the quality in the estimation of the different parameters is similar, respectively).

Finally, we analyze the performance of all the fusion rules on several numerical examples. Firstly, we perform a detailed study on two examples where the exact closed-form expressions for the partial and the global estimators can be obtained: a univariate Gamma distribution and a multi-variate Gaussian model. This allows us to rule out any approximation effects (e.g., due to slow convergence and poor mixing in MC methods) and to analyze the effect of the number of samples, the number of estimators, the prior, and the dimensionality of the state space. Then, we apply the proposed algorithms to the problem of target localization in a wireless sensor network using measurements acquired by several sensors with different noise characteristics. In this scenario, MC partial estimators (based on parallel chains) are used to deal with the groups of measurements, showing that the performance of the novel fusion rules is close to that of the optimal fusion rule (or even better in some cases) with only a fraction of its computational cost.

A preliminary version of this work has been published in [38]. In this paper we elaborate on that work, introducing the following main novel contributions:

- A more extensive and updated review of the literature, which includes a

more detailed analysis of the connections with related fields.

- A thorough Bayesian analysis of the different fusion rules and the conditions under which they are optimal: only the asymptotic optimality of the best linear unbiased fusion rule was justified in [38] (using the Bernstein-von Mises theorem), whereas here we prove its optimality in the non-asymptotic regime for the multi-variate Gaussian model and discuss the optimality of the other two efficient fusion rules proposed.

- A set of appendices where we provide the technical derivations for the Bayesian analysis, as well as the solution of the constrained minimization problems considered for the different fusion rules. Note that none of these appendices could be included in [38] due to lack of space.

- Two additional numerical examples, where we show the performance of the different fusion rules as the dimension of the state space increases (note that the dimension of the state space in [38] was just $D = 2$): a multivariate Gaussian case (where the dimension of the state space is changed from $D = 1$ up to $D = 20$), and an extended localization example in wireless sensor networks (where the dimension of the state space is $D = 8$).

1.3. Organization

The reminder of the paper is structured as follows. The notation and the problem statement are provided first in Section 2, which briefly recalls the statistical inference framework to derive parameter estimators (Section 2.1), and compares the global and partial estimators (Section 2.2). This is followed by Section 3, where the optimal linear combination method is obtained by solving a constrained minimization problem (Section 3.1), and two novel efficient linear fusion rules are also derived following this approach (Sections 3.2 and 3.3). An alternative approach is then pursued in Section 4, which provides a Bayesian perspective on linear fusion rules: the best linear unbiased fusion rule (i.e., the one that minimizes the MMSE) is derived for the Gaussian case (Section 4.1);

the asymptotic optimality of this fusion rule in other cases is proved through the asymptotic normality of the partial MMSE estimators as formulated by the Bernstein-von Mises theorem (Section 4.2); and some particular cases that lead to the proposed efficient fusion rules are finally discussed (Section 4.3). Several numerical experiments are analyzed and discussed in Section 5: a one-dimensional problem using Gamma distributions, where analytical expressions for the partial MMSE estimators can be obtained and we analyze the effect of the bias (Section 5.1); a multi-variate Gaussian model, where closed-form estimators can still be found and we study the effect of increasing the number of parameters to be estimated (Section 5.2); and two localization problems in wireless sensor networks, where MCMC methods have to be used to obtain the partial and global estimators (Section 5.3). Finally, some concluding remarks and future lines are provided in Section 6.

2. Problem Statement

2.1. Statistical Inference

Many applications in statistical signal processing require inferring a set of variables of interest or unknowns given a collection of observations or measurements. Let us consider a D-dimensional vector of unknowns, $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^D$, and let $\mathbf{y} \in \mathcal{Y} \subseteq \mathbb{R}^N$ be a collection of N i.i.d. observed data. The two major paradigms in statistical inference, the frequentist approach and the Bayesian perspective [3, 4], are briefly reviewed in the sequel.

2.1.1. Frequentist Parameter Estimation

From a *frequentist* point of view, the optimal estimator of $\mathbf{x}$ is a function of the observations, $\hat{\mathbf{x}}^* = \mathbf{f}(\mathbf{y})$, that minimizes the expected value of some given *loss* or *cost function*, $C(\mathbf{x}, \hat{\mathbf{x}})$, i.e.,

$$\hat{\mathbf{x}}^* = \arg\min_{\hat{\mathbf{x}}} \mathbb{E}\left(C(\mathbf{x}, \hat{\mathbf{x}})\right), \tag{1}$$

where $\mathbb{E}(\cdot)$ denotes the mathematical expectation, in this case w.r.t. the PDF of the data, $p(\mathbf{y})$. Although many different cost functions have been considered, for statistical estimation the most common one is the squared loss, $C(\mathbf{x}, \hat{\mathbf{x}}) = (\hat{\mathbf{x}} - \mathbf{x})^\top (\hat{\mathbf{x}} - \mathbf{x})$, which leads to the well-known least squares (LS) or frequentist minimum mean squared error (MMSE) estimator:

$$\hat{\mathbf{x}}^{(\mathrm{LS})} = \arg\min_{\hat{\mathbf{x}}} \mathrm{MSE}(\hat{\mathbf{x}}), \tag{2}$$

where

$$\mathrm{MSE}(\hat{\mathbf{x}}) = \mathbb{E}\left((\hat{\mathbf{x}} - \mathbf{x})^\top (\hat{\mathbf{x}} - \mathbf{x})\right) = \int_{\mathcal{Y}} (\hat{\mathbf{x}} - \mathbf{x})^\top (\hat{\mathbf{x}} - \mathbf{x}) p(\mathbf{y}|\mathbf{x}) d\mathbf{y}. \tag{3}$$

Note that the MSE in (3) can be expressed as

$$\mathrm{MSE}(\hat{\mathbf{x}}) = \mathrm{Tr}\left(\mathbb{E}\left((\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^\top\right)\right) = \mathrm{Tr}\left(\mathbf{C}_{\hat{\mathbf{x}}}\right) + \mathbf{b}_{\hat{\mathbf{x}}}^\top \mathbf{b}_{\hat{\mathbf{x}}}, \tag{4}$$

with $\mathrm{Tr}(\cdot)$ denoting the trace of a matrix (i.e., the sum of the elements along its main diagonal), $\mathbf{C}_{\hat{\mathbf{x}}} = (\hat{\mathbf{x}} - \mathbb{E}(\hat{\mathbf{x}}))(\hat{\mathbf{x}} - \mathbb{E}(\hat{\mathbf{x}}))^\top$ indicating the covariance matrix of $\hat{\mathbf{x}}$, and $\mathbf{b}_{\mathbf{x}} = \mathbb{E}(\hat{\mathbf{x}}) - \mathbf{x}$ denoting the bias of $\hat{\mathbf{x}}$. Moreover, the estimator $\hat{\mathbf{x}}$ is usually required to be unbiased (i.e., $\mathbb{E}(\hat{\mathbf{x}}) = \mathbf{x}$), implying that $\mathrm{MSE}(\hat{\mathbf{x}}) = \mathrm{Tr}\left(\mathbf{C}_{\hat{\mathbf{x}}}\right)$ and thus the LS estimator is simply the one that minimizes the trace of the covariance matrix of $\hat{\mathbf{x}}$.

2.1.2. Bayesian Inference

From a *Bayesian* perspective, the problem of finding an optimal estimator can be formulated as the minimization of the *Bayesian Expected Loss*,

$$r(\hat{\mathbf{x}}) = \int_{\mathcal{X}} C(\mathbf{x}, \hat{\mathbf{x}}) p(\mathbf{x}|\mathbf{y}) d\mathbf{x}, \tag{5}$$

where $C(\mathbf{x}, \hat{\mathbf{x}})$ is again some suitable *cost function*, and $p(\mathbf{x}|\mathbf{y})$ is the posterior PDF. In the Bayesian framework, this posterior PDF contains all the information available to the user about the unknown variables $\mathbf{x}$, and is given by Bayes

theorem:

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p_0(\mathbf{x})}{p(\mathbf{y})}, \tag{6}$$

where $p(\mathbf{y}|\mathbf{x})$ is the likelihood, $p_0(\mathbf{x})$ is the prior PDF, and $p(\mathbf{y})$ is the PDF of the data. Now, let us consider again the quadratic cost, $C(\mathbf{x}, \hat{\mathbf{x}}) = (\hat{\mathbf{x}} - \mathbf{x})^\top(\hat{\mathbf{x}} - \mathbf{x})$, which is also the most common cost function for regression problems within the Bayesian framework. Then, (5) becomes

$$r(\hat{\mathbf{x}}) = \text{MSE}(\hat{\mathbf{x}}|\mathbf{y}) = \int_{\mathcal{X}} (\hat{\mathbf{x}} - \mathbf{x})^\top(\hat{\mathbf{x}} - \mathbf{x})p(\mathbf{x}|\mathbf{y})d\mathbf{x}, \tag{7}$$

where $\text{MSE}(\hat{\mathbf{x}}|\mathbf{y})$ refers to the *conditional Bayesian MSE* (i.e., the Bayesian MSE for a fixed set of data $\mathbf{y}$), that we will also denote in the sequel as MSE for the sake of simplicity.[1] The optimal estimator (i.e., the one that minimizes Eq. (7)) is the well-known Bayesian minimum mean squared error (MMSE) estimator, which corresponds to the conditional mean, i.e., the expected value of $\mathbf{x}$ w.r.t. the posterior PDF [1, 3, 4, 5]:

$$\hat{\mathbf{x}}^{(\text{MMSE})} = \mathbb{E}(\mathbf{x}|\mathbf{y}) = \int_{\mathcal{X}} \mathbf{x}\ p(\mathbf{x}|\mathbf{y})d\mathbf{x}. \tag{9}$$

2.2. Global vs. Distributed Partial Bayesian Estimators

All the estimators discussed in the previous section are *global estimators*, since they are assumed to have access to all the available data. Hence, their performance is optimal (from the point of view of minimizing their respective cost functions) whenever they can be computed exactly. However, in big data problems we cannot deal with the whole data set globally due to computational and/or memory restrictions. Furthermore, even when we can work with the

[1] The *full Bayesian MSE* is obtained by performing a double integral on both the data and the parameters of interest using the joint PDF $p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{y})$:

$$\text{MSE}(\hat{\mathbf{x}}) = \int_{\mathcal{Y}} \int_{\mathcal{X}} (\hat{\mathbf{x}} - \mathbf{x})^\top(\hat{\mathbf{x}} - \mathbf{x})p(\mathbf{x}|\mathbf{y})p(\mathbf{y})d\mathbf{x}d\mathbf{y}. \tag{8}$$

However, by assuming that the data are fixed (i.e., by conditioning on the data), the outer integral in (8) vanishes and the only integral remaining is the one on $\mathbf{x}$ using $p(\mathbf{x}|\mathbf{y})$.

whole data set globally, splitting it into L data sets may be more efficient and lead to a better performance. For instance, in the Bayesian framework it is well-known that the posterior PDF tends to become more "peaky" as the amount of available data increases. Consequently, the inference process becomes harder (especially for high-dimensional scenarios) when closed form expressions for the Bayesian estimators cannot be found and we have to resort to numerical approximations (like Monte Carlo methods [11, 39]).

A natural solution in these cases is splitting the data into L disjoint groups or clusters, so that the ℓ-th cluster ($1 \leq \ell \leq L$) only has access to N_ℓ samples. Then, we can obtain a *partial estimator* for the ℓ-th cluster (i.e., a partial estimator of $\mathbf{x}$ given only the partial data available within the ℓ-th cluster, $\mathbf{y}_\ell$), $\hat{\mathbf{x}}_\ell$, by constructing a function of the ℓ-th data set that minimizes the corresponding cost function.[2] For instance, the partial MMSE estimator in the Bayesian setting would be given by

$$\hat{\mathbf{x}}_\ell^{(\mathrm{MMSE})} = \mathbb{E}(\mathbf{x}|\mathbf{y}_\ell) = \int_{\mathcal{X}} \mathbf{x}\ p_\ell(\mathbf{x}|\mathbf{y}_\ell) d\mathbf{x}, \tag{10}$$

where $p_\ell(\mathbf{x}|\mathbf{y}_\ell)$ is the partial posterior associated to the ℓ-th dataset (see Table 1 for a summary of the notation used throughout the paper). The challenge now is trying to obtain the exact global MMSE estimator, $\hat{\mathbf{x}}^{(\mathrm{MMSE})}$, from the set of partial MMSE estimators, $\{\hat{\mathbf{x}}_\ell^{(\mathrm{MMSE})}\}_{\ell=1}^{L}$.

In this paper, we concentrate on a communications-free setup for the partial estimators, i.e., we assume that the partial estimators can only transmit their final estimates to the fusion center (FC), altogether with additional side information if needed, and are not allowed to communicate with each other during the estimation process. The FC will then be responsible for combining all the estimates in an efficient way in order to obtain the global MMSE estimator (if it is feasible) or at least the best possible approximation. We consider linear

[2] Note that we use the name partial estimator, instead of distributed or local estimator, to emphasize the fact that $\hat{\mathbf{x}}_\ell$ corresponds to the estimator of the complete set of variables of interest obtained using only the partial information available to the ℓ-th cluster.

Table 1: Summary of the Notation.

$\mathbf{x}$	Unknown parameters to be estimated.
$\hat{\mathbf{x}}$	Global estimator of $\mathbf{x}$.
D	Number of unknowns (i.e., dimension of $\mathbf{x}$).
$\mathbf{y}$	Vector of observations.
N	Number of observations.
L	Number of parallel (partial) estimators.
N_ℓ	Number of data for the ℓ-th estimator.
$\mathbf{y}_\ell$	Data set for the ℓ-th estimator.
$\hat{\mathbf{x}}_\ell$	ℓ-th partial estimator of $\mathbf{x}$ ($\ell = 1, \ldots, L$).
$\mathbf{\Lambda}_\ell$	Weighting matrix used to combine the L partial estimators.
$p(\mathbf{x}\|\mathbf{y})$	Global posterior PDF.
$p(\mathbf{y}\|\mathbf{x})$	Global likelihood.
$p_0(\mathbf{x})$	Global prior PDF.
$p_\ell(\mathbf{x}\|\mathbf{y}_\ell)$	Partial posterior PDF for the ℓ-th estimator.
$p_\ell(\mathbf{y}_\ell\|\mathbf{x})$	Partial likelihood for the ℓ-th estimator.
$p_0^{(\ell)}(\mathbf{x})$	Local prior PDF for the ℓ-th estimator.

fusion rules, implying that the global estimator is given by the following general expression:

$$\hat{\mathbf{x}} = \sum_{\ell=1}^{L} \mathbf{\Lambda}_\ell \hat{\mathbf{x}}_\ell, \tag{11}$$

where the $\mathbf{\Lambda}_\ell$ ($\ell = 1, \ldots, L$) are $D \times D$ weighting matrices. An important case where Eq. (11) leads to the exact global MMSE estimator occurs when both the global and the partial posteriors have Gaussian PDFs, as proved in Section 4.1. Indeed, when the conditions for the Bernstein-von Mises theorem are fulfilled, all the posterior PDFs are Gaussian and (11) becomes asymptotically optimal, as discussed in Section 4.2. Furthermore, even if (11) is not the global MMSE estimator, by choosing the $\mathbf{\Lambda}_\ell$ properly we obtain the best linear unbiased global estimator of $\mathbf{x}$ given the $\hat{\mathbf{x}}_\ell$, as shown in Section 3.1. Imposing additional restrictions on the weighting matrices leads to the novel efficient linear fusion rules described in Sections 3.2 and 3.3, which can still be optimal under certain conditions (as discussed in Section 4.3) and provide good results in many other problems (as shown through numerical simulations in Section 5). In the following sections we discuss all these issues in detail.

3. Linear Fusion of Partial Estimators: A Constrained Minimization Approach

In this section, we first derive the optimal linear fusion rule (Section 3.1) from a constrained minimization perspective, and then provide two alternative efficient fusion rules by restricting the shape of the weighting matrix: the single coefficient fusion rule in Section 3.2 and the independent fusion rule in Section 3.3. These derivations are valid regardless of the approach followed to obtain the partial estimators, since the only assumption performed is their unbiasedness.

3.1. General Case: Optimal Linear Combination

Let us consider the most general linear fusion rule, which is given by

$$\hat{\mathbf{x}} = \sum_{\ell=1}^{L} \mathbf{\Lambda}_\ell \hat{\mathbf{x}}_\ell, \tag{12}$$

where $\hat{\mathbf{x}}_\ell$ can be any partial estimator (not necessarily the MMSE estimator) based on the ℓ-th partial dataset, $\mathbf{y}_\ell$, and $\hat{\mathbf{x}}$ is the corresponding global estimator, obtained by linearly combining all those partial estimators. In this case, assuming that all the partial estimators are unbiased, the mean of the global estimator is given by

$$\mathbb{E}(\hat{\mathbf{x}}) = \sum_{\ell=1}^{L} \mathbf{\Lambda}_\ell \mathbb{E}(\hat{\mathbf{x}}_\ell) = \left(\sum_{\ell=1}^{L} \mathbf{\Lambda}_\ell \right) \mathbf{x}. \tag{13}$$

Thus, in order to obtain an unbiased global estimator (i.e., $\mathbb{E}(\hat{\mathbf{x}}) = \mathbf{x}$) we need to impose the following condition:

$$\sum_{\ell=1}^{L} \mathbf{\Lambda}_\ell = \mathbf{I}. \tag{14}$$

The covariance matrix of the global estimator is now given by

$$\mathbf{C}_{\mathbf{x}} = \mathbb{E}\left((\hat{\mathbf{x}}-\mathbf{x})(\hat{\mathbf{x}}-\mathbf{x})^{\top}\right) = \mathbb{E}\left(\left(\sum_{\ell=1}^{L}\boldsymbol{\Lambda}_{\ell}\hat{\mathbf{x}}_{\ell}-\mathbf{x}\right)\left(\sum_{\ell=1}^{L}\boldsymbol{\Lambda}_{\ell}\hat{\mathbf{x}}_{\ell}-\mathbf{x}\right)^{\top}\right), \quad (15)$$

where we have used (12) and assumed that the global estimator is unbiased. Making use of (14), Eq. (15) can be expressed as

$$\begin{aligned}\mathbf{C}_{\mathbf{x}} &= \mathbb{E}\left(\sum_{\ell=1}^{L}\boldsymbol{\Lambda}_{\ell}(\hat{\mathbf{x}}_{\ell}-\mathbf{x})\sum_{\ell=1}^{L}(\hat{\mathbf{x}}_{\ell}-\mathbf{x})^{\top}\boldsymbol{\Lambda}_{\ell}^{\top}\right)\\ &= \sum_{\ell,k=1}^{L}\boldsymbol{\Lambda}_{\ell}\,\mathbb{E}\left((\hat{\mathbf{x}}_{\ell}-\mathbf{x})(\hat{\mathbf{x}}_{k}-\mathbf{x})^{\top}\right)\boldsymbol{\Lambda}_{k}^{\top}. \quad (16)\end{aligned}$$

Finally, assuming that the partial estimators are independent and taking into account that $\mathbf{C}_{\mathbf{x}}^{(\ell)} = \mathbb{E}((\hat{\mathbf{x}}_{\ell}-\mathbf{x})(\hat{\mathbf{x}}_{\ell}-\mathbf{x})^{\top})$, $\mathbf{C}_{\mathbf{x}}$ can be expressed as a function of the covariance of the partial estimators, $\mathbf{C}_{\mathbf{x}}^{(\ell)}$, and the weighting matrices, $\boldsymbol{\Lambda}_{\ell}$, as

$$\mathbf{C}_{\mathbf{x}} = \sum_{\ell=1}^{L}\boldsymbol{\Lambda}_{\ell}\mathbf{C}_{\mathbf{x}}^{(\ell)}\boldsymbol{\Lambda}_{\ell}^{\top}. \quad (17)$$

The MSE of the global estimator is then given by

$$\mathrm{MSE}(\hat{\mathbf{x}}|\mathbf{y}) = \mathrm{Tr}\left(\mathbf{C}_{\mathbf{x}}\right) = \sum_{\ell=1}^{L}\mathrm{Tr}\left(\boldsymbol{\Lambda}_{\ell}\mathbf{C}_{\mathbf{x}}^{(\ell)}\boldsymbol{\Lambda}_{\ell}^{\top}\right), \quad (18)$$

where $\mathrm{Tr}(\cdot)$ denotes the trace of a matrix.

Hence, in order to obtain the best linear unbiased global estimator (i.e., the linear combination of the partial estimators that minimizes the MSE), we need to solve the following constrained optimization problem:

$$\boldsymbol{\Lambda}^{*} = \arg\min_{\boldsymbol{\Lambda}} \sum_{\ell=1}^{L}\mathrm{Tr}\left(\boldsymbol{\Lambda}_{\ell}\mathbf{C}_{\mathbf{x}}^{(\ell)}\boldsymbol{\Lambda}_{\ell}^{\top}\right), \quad (19a)$$

$$\text{s.t.} \qquad \sum_{\ell=1}^{L}\boldsymbol{\Lambda}_{\ell} = \mathbf{I}, \quad (19b)$$

where $\boldsymbol{\Lambda} = [\boldsymbol{\Lambda}_1, \ \ldots, \ \boldsymbol{\Lambda}_L]^\top$. Since (19a) and (19b) correspond to a convex optimization problem [40], by applying the method of the Lagrange multipliers it can be shown (see Appendix A) that the solution for each of the weighting matrices is given by

$$\boldsymbol{\Lambda}_\ell^* = \left[\sum_{k=1}^{L} \left(\mathbf{C}_{\mathbf{x}}^{(k)}\right)^{-1}\right]^{-1} \left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right)^{-1}. \tag{20}$$

Substituting this expression in (12), the optimal linear MMSE (L-MMSE) fusion rule is finally given by

$$\hat{\mathbf{x}}^{(\text{L-MMSE})} = \left[\sum_{k=1}^{L} \left(\mathbf{C}_{\mathbf{x}}^{(k)}\right)^{-1}\right]^{-1} \sum_{\ell=1}^{L} \left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right)^{-1} \hat{\mathbf{x}}_\ell, \tag{21}$$

regardless of the approach followed to derive the partial estimators.

3.2. Efficient Fusion Rule: Single Coefficient Fusion

Let us consider the particular case in which a single coefficient per estimator is used to construct the global estimator, i.e.,

$$\hat{\mathbf{x}} = \sum_{\ell=1}^{L} \alpha_\ell \hat{\mathbf{x}}_\ell, \tag{22}$$

which is obtained by setting $\boldsymbol{\Lambda}_\ell = \alpha_\ell \mathbf{I}$ in (12). Clearly, this will provide a suboptimal solution in general, but it is a fast and low-cost solution for the combination of estimators, and we can easily obtain a closed-form expression for the optimal weights.

On the one hand, since the partial estimators are unbiased, it is straightforward to see that the mean of the global estimator given by (22) is

$$\mathbb{E}(\hat{\mathbf{x}}) = \sum_{\ell=1}^{L} \alpha_\ell \mathbb{E}(\hat{\mathbf{x}}_\ell) = \left(\sum_{\ell=1}^{L} \alpha_\ell\right) \mathbf{x}. \tag{23}$$

Hence, in order to obtain an unbiased global estimator we need to have

$$\sum_{\ell=1}^{L} \alpha_\ell = 1. \tag{24}$$

On the other hand, the covariance matrix for the global estimator is given by

$$\mathbf{C}_\mathbf{x} = \sum_{\ell=1}^{L} \alpha_\ell^2 \mathbf{C}_\mathbf{x}^{(\ell)}, \tag{25}$$

and the MSE can thus be expressed as

$$\mathrm{MSE}(\hat{\mathbf{x}}|\mathbf{y}) = \mathrm{Tr}(\mathbf{C}_\mathbf{x}) = \sum_{\ell=1}^{L} \alpha_\ell^2 \mathrm{Tr}\left(\mathbf{C}_\mathbf{x}^{(\ell)}\right), \tag{26}$$

where $\mathrm{Tr}(\mathbf{C}_\mathbf{x})$ denotes the trace of the global covariance matrix:

$$\mathrm{Tr}(\mathbf{C}_\mathbf{x}) = \sum_{d=1}^{D} \mathbf{C}_\mathbf{x}[d,d] = \sum_{d=1}^{D} \sigma_{x_d}^2, \tag{27}$$

with $\sigma_{x_d}^2 = \mathbb{E}((\hat{x}_d - x_d)^2)$, and $\mathrm{Tr}(\mathbf{C}_\mathbf{x}^{(\ell)})$ denotes the trace of the ℓ-th partial covariance matrix:

$$\mathrm{Tr}(\mathbf{C}_\mathbf{x}^{(\ell)}) = T_\ell = \sum_{d=1}^{D} \mathbf{C}_\mathbf{x}^{(\ell)}[d,d] = \sum_{d=1}^{D} \sigma_{\ell,d}^2, \tag{28}$$

with $\sigma_{\ell,d}^2 = \mathbb{E}((\hat{x}_d^{(\ell)} - x_d)^2)$.

The goal is finding the set of α_ℓ that minimizes (26), subject to Eq. (24) in order to obtain an unbiased estimator. Hence, the optimal selection of the weights can be formulated as the following constrained optimization problem:

$$\boldsymbol{\alpha}^* = \arg\min_{\boldsymbol{\alpha}} \sum_{\ell=1}^{L} \alpha_\ell^2 \mathrm{Tr}\left(\mathbf{C}_\mathbf{x}^{(\ell)}\right), \tag{29a}$$

$$\text{s.t.} \qquad \sum_{\ell=1}^{L} \alpha_\ell = 1, \tag{29b}$$

with $\boldsymbol{\alpha} = [\alpha_1,\ \ldots,\ \alpha_L]^\top$. Eqs. (29a) and (29b) correspond again to a convex optimization problem. Thus, by applying once more the method of the Lagrange multipliers, it can be shown (see Appendix A) that the optimal value of α_ℓ is

$$\alpha_\ell^* = \frac{[\mathrm{MSE}(\hat{\mathbf{x}}_\ell|\mathbf{y}_\ell)]^{-1}}{\sum_{k=1}^{L}[\mathrm{MSE}(\hat{\mathbf{x}}_k|\mathbf{y}_k)]^{-1}} = \frac{T_\ell^{-1}}{\sum_{k=1}^{L} T_k^{-1}}, \tag{30}$$

where we recall that $T_\ell = \mathrm{Tr}(\mathbf{C}_{\mathbf{x}}^{(\ell)})$, and the single coefficient MMSE (SC-MMSE) fusion rule is then given by

$$\hat{\mathbf{x}}^{(\text{SC-MMSE})} = \sum_{\ell=1}^{L} \frac{T_\ell^{-1}}{\sum_{k=1}^{L} T_k^{-1}} \hat{\mathbf{x}}_\ell. \tag{31}$$

3.3. Efficient Fusion Rule: Independent Linear Fusion

The SC-MMSE fusion rule has a substantially reduced computational cost w.r.t. the L-MMSE approach, since it only requires the estimation of L parameters overall instead of the D^2L parameters of the L-MMSE rule. However, noting that the optimal weights in (31) involve the trace of the partial covariance matrices, we introduce an independent linear minimum mean squared error (IL-MMSE) fusion rule, where $\boldsymbol{\Lambda}_\ell = \mathbf{D}_\ell = \mathrm{diag}(\alpha_{\ell,1},\ \ldots,\ \alpha_{\ell,D})$. This approach leads to an independent estimation of each of the D unknowns:

$$\hat{x}_d = \sum_{\ell=1}^{L} \alpha_{\ell,d}\ \hat{x}_{\ell,d}, \tag{32}$$

where $1 \leq d \leq D$, $\hat{x}_d$ denotes the d-th component of the global estimator, and $\hat{x}_{\ell,d}$ indicates the d-th component of the ℓ-th partial estimator. The mean of the d-th component of the global estimator is then given by

$$\mathbb{E}(\hat{x}_d) = \sum_{\ell=1}^{L} \alpha_{\ell,d}\mathbb{E}(\hat{x}_{\ell,d}) = \left(\sum_{\ell=1}^{L} \alpha_{\ell,d}\right) x_d, \tag{33}$$

and the unbiasedness condition becomes

$$\sum_{\ell=1}^{L} \alpha_{\ell,d} = 1, \tag{34}$$

for $d = 1, \ldots, D$. Hence, the covariance matrix for the global estimator is

$$\mathbf{C}_{\mathbf{x}} = \sum_{\ell=1}^{L} \mathbf{D}_\ell \mathbf{C}_{\mathbf{x}}^{(\ell)} \mathbf{D}_\ell^\top, \tag{35}$$

and the MSE can thus be expressed as

$$\mathrm{MSE}(\hat{\mathbf{x}}|\mathbf{y}) = \mathrm{Tr}(\mathbf{C}_{\mathbf{x}}) = \sum_{\ell=1}^{L} \mathrm{Tr}\left(\mathbf{D}_\ell \mathbf{C}_{\mathbf{x}}^{(\ell)} \mathbf{D}_\ell^\top\right) = \sum_{\ell=1}^{L} \sum_{d=1}^{D} \alpha_{\ell,d}^2 \sigma_{\ell,d}^2. \tag{36}$$

From (34) and (36), it becomes apparent that the weights in (32) can be obtained by solving D single parameter constrained optimization problems:

$$\boldsymbol{\alpha}_d^* = \arg\min_{\boldsymbol{\alpha}_d} \sum_{\ell=1}^{L} \alpha_{\ell,d}^2 \sigma_{\ell,d}^2, \tag{37a}$$

$$\text{s.t.} \qquad \sum_{\ell=1}^{L} \alpha_{\ell,d} = 1, \tag{37b}$$

where $\boldsymbol{\alpha}_d = [\alpha_{1,d},\ \ldots,\ \alpha_{L,d}]^\top$, and we recall that $\sigma_{\ell,d}^2 = C_{x_d}^{(\ell)}$ is the d-th element along the main diagonal of $\mathbf{C}_{\mathbf{x}}^{(\ell)}$. The solution (see Appendix A) is now given by

$$\alpha_{\ell,d}^* = \frac{\left[\mathrm{MSE}(\hat{x}_{\ell,d}^{(\mathrm{MMSE})}|\mathbf{y}_\ell)\right]^{-1}}{\sum_{k=1}^{L}\left[\mathrm{MSE}(\hat{x}_{k,d}^{(\mathrm{MMSE})}|\mathbf{y}_k)\right]^{-1}} = \frac{\sigma_{\ell,d}^{-2}}{\sum_{k=1}^{L}\sigma_{k,d}^{-2}}, \tag{38}$$

and thus the IL-MMSE fusion rule is

$$\hat{x}_d^{(\text{IL-MMSE})} = \sum_{\ell=1}^{L} \frac{\sigma_{\ell,d}^{-2}}{\sum_{k=1}^{L}\sigma_{k,d}^{-2}} \hat{x}_{\ell,d}, \tag{39}$$

for $d = 1, \ldots, D$. This approach requires the estimation of DL parameters overall, and thus it can be seen as an intermediate approach between the L-

MMSE and the SC-MMSE fusion rules, both in terms of computational cost and performance (as shown in Section 5).

4. Optimal Linear Fusion: Bayesian Perspective

In this section, we provide an alternative perspective of the optimal linear fusion problem from a Bayesian point of view. First of all, we derive the optimal fusion rule for the multi-variate Gaussian model in Section 4.1, and then we show that this rule is asymptotically optimal under mild conditions in Section 4.2. Finally, in Section 4.3 we address two relevant particular cases that correspond to the SC-MMSE and IL-MMSE fusion rules derived in Sections 3.2 and 3.3, respectively.

4.1. Gaussian Estimators: Optimal Fusion Rule

Let us consider the derivation of the optimal fusion rule at the Fusion Center (FC) from a Bayesian point of view. In this case, our observations are the outputs of each of the L partial estimators, $\hat{\mathbf{x}}_\ell$ for $\ell = 1, \ldots, L$. Let us assume that these estimators are independent, unbiased (i.e., $\mathbb{E}(\hat{\mathbf{x}}_\ell) = \mathbf{x}$), and have Gaussian PDFs with means equal to the true parameter vector $\mathbf{x}$ and covariance matrices $\mathbf{C}_{\mathbf{x}}^{(\ell)}$:

$$\begin{aligned} p(\hat{\mathbf{x}}_\ell|\mathbf{x}) &= \mathcal{N}(\hat{\mathbf{x}}_\ell|\mathbf{x}, \mathbf{C}_{\mathbf{x}}^{(\ell)}) \\ &= \left(2\pi|\mathbf{C}_{\mathbf{x}}^{(\ell)}|\right)^{-D/2} \exp\left(-\frac{1}{2}(\hat{\mathbf{x}}_\ell - \mathbf{x})^\top \left[\mathbf{C}_{\mathbf{x}}^{(\ell)}\right]^{-1} (\hat{\mathbf{x}}_\ell - \mathbf{x})\right). \end{aligned} \tag{40}$$

The full posterior is

$$p(\mathbf{x}|\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_L) \propto p(\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_L|\mathbf{x}) p_0(\mathbf{x}), \tag{41}$$

with the likelihood function given by

$$p(\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_L|\mathbf{x}) = \prod_{\ell=1}^{L} p(\hat{\mathbf{x}}_\ell|\mathbf{x}) = \prod_{\ell=1}^{L} \mathcal{N}(\mathbf{x}|\hat{\mathbf{x}}_\ell, \mathbf{C}_{\mathbf{x}}^{(\ell)}), \tag{42}$$

and a Gaussian prior with mean $\hat{\mathbf{x}}_0$ and covariance matrix $\mathbf{C}_{\mathbf{x}}^{(0)}$,

$$p_0(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\hat{\mathbf{x}}_0, \mathbf{C}_{\mathbf{x}}^{(0)}). \tag{43}$$

Inserting (42) and (43) into (41), the full posterior can be finally expressed as

$$\begin{aligned} p(\mathbf{x}|\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_L) &\propto \prod_{\ell=0}^{L} \mathcal{N}(\mathbf{x}|\hat{\mathbf{x}}_\ell, \mathbf{C}_{\mathbf{x}}^{(\ell)}) \\ &\propto \prod_{\ell=0}^{L} |\mathbf{C}_{\mathbf{x}}^{(\ell)}|^{-1/2} \exp\left(-\frac{1}{2}\sum_{\ell=1}^{L} (\mathbf{x}-\hat{\mathbf{x}}_\ell)^\top \left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right)^{-1} (\mathbf{x}-\hat{\mathbf{x}}_\ell)\right). \end{aligned} \tag{44}$$

After some algebra (see Appendix B), Eq. (44) can be expressed as a single multi-variate Gaussian:

$$p(\mathbf{x}|\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_L) = (2\pi)^{-D/2}|\mathbf{C}_{\mathbf{x}}|^{-1/2} \exp\left(-\frac{1}{2}(\hat{\mathbf{x}}-\boldsymbol{\mu}_{\mathbf{x}})^\top \mathbf{C}_{\mathbf{x}}^{-1}(\hat{\mathbf{x}}-\boldsymbol{\mu}_{\mathbf{x}})\right), \tag{45}$$

where

$$\mathbf{C}_{\mathbf{x}} = \left[\sum_{\ell=0}^{L} \left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right)^{-1}\right]^{-1}, \tag{46a}$$

$$\boldsymbol{\mu}_{\mathbf{x}} = \mathbf{C}_{\mathbf{x}} \sum_{\ell=0}^{L} \left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right)^{-1} \hat{\mathbf{x}}_\ell. \tag{46b}$$

Hence, the global MMSE estimator for the multi-variate Gaussian model, that we denote as the G-MMSE estimator and corresponds to the mean of the full

posterior in Eq. (45), is finally given by

$$\hat{\mathbf{x}}^{(\text{G-MMSE})} = \boldsymbol{\mu}_{\mathbf{x}} = \sum_{\ell=0}^{L} \boldsymbol{\Lambda}_\ell \hat{\mathbf{x}}_\ell, \tag{47}$$

with

$$\boldsymbol{\Lambda}_\ell = \mathbf{C}_{\mathbf{x}} \cdot \left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right)^{-1} = \left[\sum_{k=0}^{L} \left(\mathbf{C}_{\mathbf{x}}^{(k)}\right)^{-1}\right]^{-1} \left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right)^{-1}. \tag{48}$$

Therefore, given any set of partial estimators $\{\hat{\mathbf{x}}_\ell\}_{\ell=1}^{L}$, the linear fusion is always optimal in the multi-variate Gaussian case, with an optimal weighting matrix given by Eq. (48). Furthermore, when the $\hat{\mathbf{x}}_\ell$ are the partial MMSE estimators (i.e., when $\hat{\mathbf{x}}_\ell = \hat{\mathbf{x}}_\ell^{(\text{MMSE})}$ for $\ell = 1, \ldots, L$), then we have $\hat{\mathbf{x}}^{(\text{G-MMSE})} = \hat{\mathbf{x}}^{(\text{MMSE})}$, with $\hat{\mathbf{x}}^{(\text{MMSE})}$ denoting the global MMSE estimator that has access to all the data. Finally, note that the optimal fusion rule derived from the Bayesian perspective is equivalent to the optimal fusion rule obtained following the constrained optimization approach, since (48) is identical to (20).[3]

4.2. Asymptotically Optimal Fusion: Bernstein-von Mises Theorem

The Bernstein-von Mises (a.k.a. Bayesian central limit) theorem states that, under mild regularity conditions, a posterior PDF converges to a Gaussian PDF as the number of samples tends to infinity [41, 42]. Applying this result to the partial posterior PDFs, we have

$$p_\ell(\mathbf{x}|\mathbf{y}_\ell) \to \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{\mathbf{x}}^{(\ell)}, \mathbf{C}_{\mathbf{x}}^{(\ell)}) \qquad \text{as} \quad N_\ell \to \infty, \tag{49}$$

[3] Actually, there is a difference between the optimal fusion rules derived from a Bayesian and a constrained optimization point of view, since the sum in (48) starts at $\ell = 0$, while the sum in (20) starts at $\ell = 1$. This is due to the fact that the sum in (48) includes the prior, whose mean and covariance are $\hat{\mathbf{x}}_0$ and $\mathbf{C}_{\mathbf{x}}^{(0)}$, respectively. However, by using a non-informative prior (i.e., a prior such that both $\hat{\mathbf{x}}_0 = \mathbf{0}$ and $(\mathbf{C}_{\mathbf{x}}^{(0)})^{-1} = \mathbf{0}$), (48) and (20) become identical.

with $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{\mathbf{x}}^{(\ell)}, \mathbf{C}_{\mathbf{x}}^{(\ell)})$ indicating that $\mathbf{x}$ has a Gaussian PDF with a mean vector $\boldsymbol{\mu}_{\mathbf{x}}^{(\ell)} = \hat{\mathbf{x}}_{\ell}^{(\text{MMSE})}$ and a covariance matrix

$$\begin{aligned}\mathbf{C}_{\mathbf{x}}^{(\ell)} &= \mathbb{E}\left((\mathbf{x} - \hat{\mathbf{x}}_{\ell}^{(\text{MMSE})})(\mathbf{x} - \hat{\mathbf{x}}_{\ell}^{(\text{MMSE})})^{\top}\right) \\ &= \int_{\mathcal{X}} (\mathbf{x} - \hat{\mathbf{x}}_{\ell}^{(\text{MMSE})})(\mathbf{x} - \hat{\mathbf{x}}_{\ell}^{(\text{MMSE})})^{\top} p_{\ell}(\mathbf{x}|\mathbf{y}_{\ell}) d\mathbf{x}. \end{aligned} \tag{50}$$

Assuming that we have independent (though not necessarily identically distributed) observations and that each of them can only belong to one cluster (i.e., we have disjoint sets of samples such that $N = \sum_{\ell=1}^{L} N_{\ell}$), the global posterior PDF also converges to a Gaussian PDF as N tends to infinity, i.e.,

$$p(\mathbf{x}|\mathbf{y}) = \prod_{\ell=1}^{L} p_{\ell}(\mathbf{x}|\mathbf{y}_{\ell}) \to \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{C}_{\mathbf{x}}) \qquad \text{as} \quad N \to \infty, \tag{51}$$

with $\mathbf{C}_{\mathbf{x}}$ and $\mu_{\mathbf{x}}$ given by (46a) and (46b), respectively, using $\hat{\mathbf{x}}_{\ell} = \hat{\mathbf{x}}_{\ell}^{(\text{MMSE})}$. Eqs. (49) and (51) state that, even if $p(\mathbf{x}|\mathbf{y})$ and the $p_{\ell}(\mathbf{x}|\mathbf{y}_{\ell})$ are not Gaussian, they will converge to Gaussian PDFs as $N_{\ell} \to \infty$ for $\ell = 1, \ldots, L$ (and thus, $N = \sum_{\ell} N_{\ell} \to \infty$ also). This implies that, when the regularity conditions of the Bernstein-von Mises theorem are fulfilled, the linear fusion rule derived in Section 4.1 is asymptotically optimal, i.e., $\hat{\mathbf{x}}^{(\text{G-MMSE})} \to \hat{\mathbf{x}}^{(\text{MMSE})}$ when $\hat{\mathbf{x}}_{\ell} = \hat{\mathbf{x}}_{\ell}^{(\text{MMSE})}$ and $N_{\ell} \to \infty$ for $\ell = 1, \ldots, L$.

4.3. Particular Cases

Note that the fusion rule of Eq. (47) requires the computation of a $D \times D$ weighting matrix, given by Eq. (48), for each of the L estimators. This implies calculating $D^2 L$ weights overall, which may be unfeasible (or at least very costly from a computational/storage point of view) when D and/or L is large. Moreover, even if the optimal weighting matrix can be computed, the numerical errors that arise in its computation (since it depends on the $\mathbf{C}_{\mathbf{x}}^{(\ell)}$, which are usually unknown and must be estimated from the data) may hinder the performance of the optimal fusion rule, especially when D is large and D^2 is

comparable to N_ℓ.

In these cases, structured weighting matrices (which require computing a reduced number of coefficients) can be used to obtain an approximation of the optimal case. Indeed, in certain cases the optimum weighting matrix may already contain a reduced number of different elements. As a first example, let us consider the situation where the parameters to be inferred are not interrelated. In this scenario, the covariance matrix for the ℓ-th estimator becomes diagonal,

$$\mathbf{C}_{\mathbf{x}}^{(\ell)} = \mathrm{diag}\left(\sigma_{\ell,1}^2,\ \ldots,\ \sigma_{\ell,D}^2\right), \tag{52}$$

with

$$\sigma_{\ell,d}^2 = \int_{\mathcal{X}_d} (\hat{x}_{\ell,d} - x_d)^2 p(x_d|\mathbf{y}_\ell) dx_d \tag{53}$$

for $d = 1, \ldots, D$. Hence, the optimal weighting matrix becomes

$$\mathbf{\Lambda}_\ell = \mathrm{diag}\left(\alpha_{\ell,1},\ \ldots,\ \alpha_{\ell,D}\right), \tag{54}$$

with

$$\alpha_{\ell,d} = \frac{\sigma_{\ell,d}^{-2}}{\sum_{k=1}^{L} \sigma_{k,d}^{-2}}. \tag{55}$$

Note that only D parameters are required now for each of the L estimators (i.e., DL parameters in total).

As a second example, let us assume that, in addition to having independent parameters to be inferred, the uncertainty in their estimation is the same for a given partial estimator, i.e., $\sigma_{\ell,1}^2 = \sigma_{\ell,2}^2 = \cdots = \sigma_{\ell,D}^2 = \sigma_\ell^2$. Then, the covariance matrix for the ℓ-th partial estimator becomes $\mathbf{C}_{\mathbf{x}}^{(\ell)} = \sigma_\ell^2 \mathbf{I}$, where $\mathbf{I}$ denotes the $D \times D$ identity matrix, and the optimal weighting matrix is now given by

$$\mathbf{\Lambda}_\ell = \alpha_\ell \mathbf{I}, \tag{56}$$

with

$$\alpha_\ell = \frac{\sigma_\ell^{-2}}{\sum_{k=1}^{L} \sigma_k^{-2}}. \tag{57}$$

Note that this approach only requires a single parameter per estimator (i.e., L parameters in total), and thus it is very efficient from a computational point of view. Furthermore, if all the variances are equal (i.e., $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_L^2 = \sigma^2$) we obtain the so-called *equal weights fusion* (EWF) rule, which assigns the same weight to all the partial estimators,

$$\alpha_1 = \alpha_2 = \cdots = \alpha_L = \frac{1}{L}, \tag{58}$$

and thus it does not require the transmission of any coefficient to perform the fusion at the FC.

Obviously, in real applications it is very unlikely that all the parameters to be inferred are completely independent, and even more unlikely that the uncertainty in their estimation is exactly equal. However, it can be shown (see Appendix C) that the coefficients in Eqs. (54) and (56) correspond, respectively, to the best independent and isotropic Gaussian approximations (in terms of minimizing the Kullback-Leibler (KL) divergence) to the asymptotic Gaussian PDFs of the partial estimators, $p_\ell(\mathbf{x}|\hat{\mathbf{x}}_\ell) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_\mathbf{x}, \mathbf{C}_\mathbf{x}^{(\ell)})$. Hence, even if the independence and/or equal quality assumptions are not fulfilled exactly, Eqs. (54) and (56) can still lead to good fusion rules (as shown in Section 5) if the approximations of the partial PDFs are accurate enough.

5. Numerical Experiments

In this section, we provide three different numerical examples where we compare the four linear fusion rules derived in the previous two sections:

1. **Linear MMSE (L-MMSE)** fusion rule, which uses the full weighting matrix and the global estimator is given by Eq. (21).
2. **Independent Linear MMSE (IL-MMSE)** fusion rule, which uses a diagonal weighting matrix and the global estimator is given by Eq. (31).
3. **Single Coefficient MMSE (SC-MMSE)** fusion rule, which uses a scaled identity matrix for the weights and the global estimator is given

by Eq. (39).

4. **Equal Weights Fusion (EWF)**, which simply averages the results of all the partial estimators.

5.1. Univariate Gamma Distributions

Let us consider first a univariate example, where exact calculations may be performed. This allows us to rule out any potential issue with the underlying MC methods typically used to approximate the MMSE estimators (e.g., slow convergence and poor mixing), and concentrate on the performance of the proposed fusion rules. Let us assume that we have N i.i.d. observations distributed according to a Gamma PDF with known shape parameter $\alpha > 0$ and unknown rate parameter $\beta > 0$. Although we do not consider any particular application, the Gamma distribution has been used, for instance, to represent the amount of rainfall in meteorological applications. In this context, the value of the shape parameter is related to the dryness of the area under study [43]: wet areas correspond to large values of α, while dry areas are associated to small values of α. Hence, this example might correspond to the estimation of the dryness of an area from distributed measurements obtained using a sensor network.

In this scenario, the likelihood is given by

$$p(\mathbf{y}|\beta) = \prod_{n=1}^{N} p(y_n|\beta), \tag{59}$$

with

$$p(y_n|\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} y_n^{\alpha-1} \exp(-\beta y_n). \tag{60}$$

The conjugate prior is also a Gamma PDF over β with shape parameter $\alpha_0 > 0$ and rate parameter $\beta_0 > 0$, and thus the global posterior density is another Gamma PDF with parameters $\alpha^* = \alpha_0 + N\alpha$ and $\beta^* = \beta_0 + \sum_{n=1}^{N} y_n$ [44]. The global MMSE estimator is given by the mean of the posterior PDF,

$$\hat{\beta}^{(\mathrm{MMSE})} = \frac{\alpha^*}{\beta^*} = \frac{\alpha_0 + N\alpha}{\beta_0 + \sum_{n=1}^{N} y_n}, \tag{61}$$

and its variance is given by

$$\sigma_\beta^2 = \frac{\alpha^*}{(\beta^*)^2} = \frac{\alpha_0 + N\alpha}{\left(\beta_0 + \sum_{n=1}^{N} y_n\right)^2}. \tag{62}$$

For the partial estimators, the MMSE estimates and their variances are still given by (61) and (62), respectively, but replacing N by N_ℓ and taking the sum only over the N_ℓ samples available to each of the ℓ estimators:

$$\hat{\beta}_\ell^{(\text{MMSE})} = \frac{\alpha_0 + N_\ell\alpha}{\beta_0 + \sum_{n\in\mathcal{Y}_\ell} y_n}, \tag{63a}$$

$$\sigma_\ell^2 = \frac{\alpha_0 + N_\ell\alpha}{\left(\beta_0 + \sum_{n\in\mathcal{Y}_\ell} y_n\right)^2}, \tag{63b}$$

where $\mathcal{Y}_\ell$ denotes the set of data available to the ℓ-th partial estimator.

We are interested now in analyzing the effect of the sample size, the number of partial estimators and the number of samples per estimator. Therefore, we test $N \in \{10^3, 10^4, 10^5, 10^6, 5\cdot 10^6\}$ with an equal number of samples per partial estimator ranging from $N_\ell = 1$ (i.e., as many partial estimators as observations) up to $N_\ell = N$ (i.e., a single estimator that corresponds to the global estimator). For each case, 1000 simulations are performed to average the results.

Figure 1 shows the typical performance of the optimal linear fusion rule (since we only have one parameter, all the fusion rules discussed in the paper are equivalent), the EWF rule (that assigns the same weight to all the partial estimators) and an empirical estimator that combines the optimal and the EWF estimates at the fusion center, $\hat{\beta}^{(\text{Avg})} = \frac{\hat{\beta}^{(\text{L-MMSE})} + \hat{\beta}^{(\text{EWF})}}{2}$. In this example, the true parameters are $\alpha = 2$ and $\beta = 5$, and an improper prior is used by setting $\alpha_0 = \beta_0 = \epsilon$ with $\epsilon \to 0$.[4] First of all, note that the optimal linear fusion rule always performs better than the EWF (as expected), especially when the number of partial estimators (a.k.a. filters) increases. The unexpected result is that combining the optimal fusion strategy and the EWF approach leads to a

[4] In practice, we used $\epsilon = 0.01$ for the simulations.

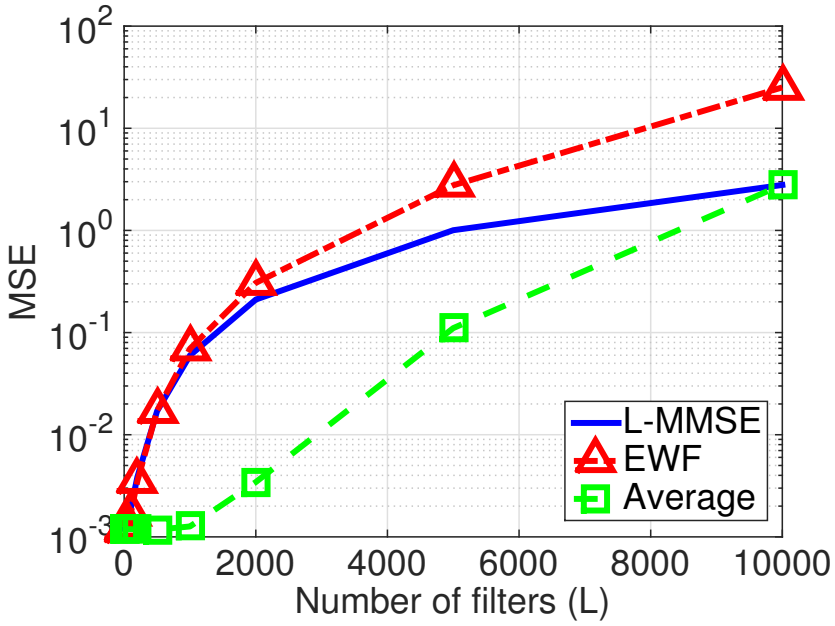


Figure 1: **Univariate Gamma Example:** MSE as a function of L using an improper prior for the three fusion rules considered: L-MMSE, EWF and their average.

better performance than any of the two individual strategies. The reason for this good performance can be seen in Figure 2, which shows the estimated posterior PDFs of the estimators for the three fusion rules considered (L-MMSE, EWF and their average), compared to the posterior of the global estimator.[5] It can be seen that the optimal linear fusion rule introduces a negative bias, whereas the EWF introduces a positive bias with a similar magnitude. Therefore, combining the two estimators leads to an average estimator with a reduced bias and thus a better performance.

The second important issue in Figure 1 is related to the increase of the MSE as the number of partial estimators increases. This is precisely due to the fact that the bias increases as the number of samples per partial estimator decreases (i.e., as the number of partial estimator increases for a fixed number

[5]Note that, given the large number of data available to the global estimator, the global posterior can be considered to be the "ground truth" for the evaluation of the PDFs obtained using the other fusion rules.

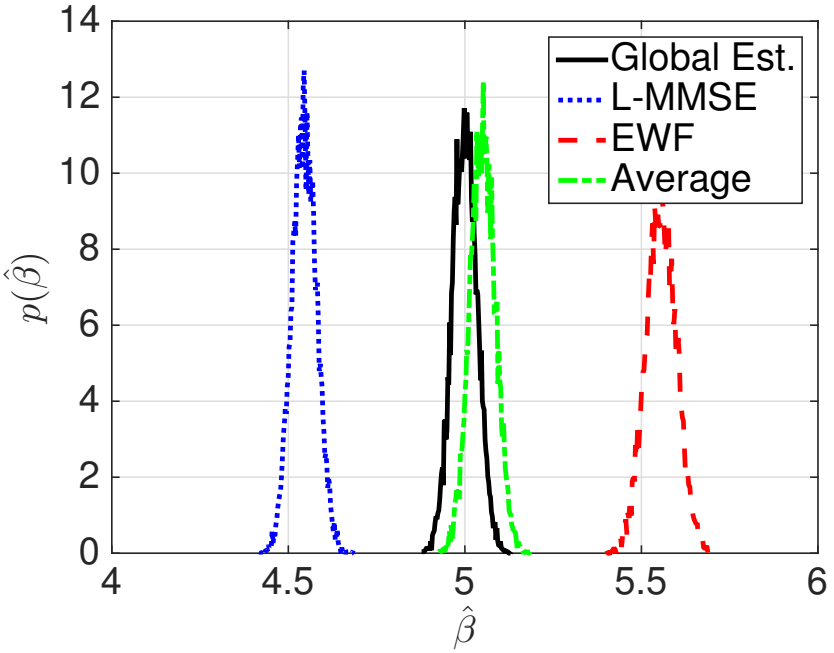


Figure 2: **Univariate Gamma Example:** Posterior density of the estimator for the three fusion rules considered (L-MMSE, EWF and their average), compared to the posterior of the global estimator.

of data).[6] This bias is caused by the mismatch between the "true" prior (in the simulated scenario, an impulse centered around the true value $\beta = 5$) and the prior assumed by the model (a Gamma PDF with parameters α_0 and β_0). In order to see this, Figure 3 shows the evolution of the MSE with the number of filters using a narrow prior (obtained setting $\beta_0 = \frac{\beta}{\varepsilon}$ and $\alpha_0 = \beta \times \beta_0$ for $\varepsilon = 0.01$) centered around the true value of β. In this case, all the estimators are unbiased and the MSE decreases as we increase the number of partial estimators. These results, in an example where the exact MMSE estimator can be obtained, highlight the importance of the prior in the Bayesian distributed inference approach. Although this falls outside of the scope of this paper, let us remark also that some approaches to reduce the bias in this type of distributed

[6]Note that we assumed unbiased partial estimators in the derivation of all the fusion rules. However, when the number of samples per partial estimator is small this is no longer true.

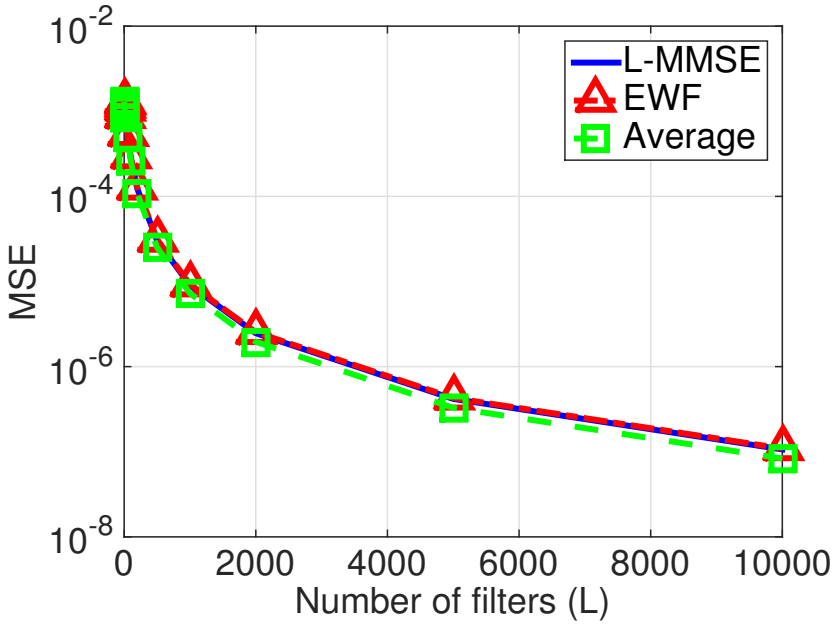


Figure 3: **Univariate Gamma Example:** MSE as a function of L using a narrow prior centered around the true value of β for the three fusion rules considered: L-MMSE, EWF and their average.

inference problems have been recently proposed [45, 46].

Finally, Table 2 provides the complete picture regarding the evolution of the MSE with the number of data and the number of data per partial estimator for all the fusion rules considered. On the one hand, note that a minimum amount of samples per estimator are required in order to attain a performance that decreases linearly as a function of N for the optimal linear fusion and the EWF. Otherwise, the bias dominates and nothing is gained by increasing N (e.g., when $N_\ell = 10$ the MSE of the L-MMSE fusion rule is 0.0637 for $N = 10^3$ and 0.0568 for $N = 10^6$, implying that the MSE is only lowered by $\approx 10\%$ when the total number of data increases by a factor of 1000). On the other hand, note the excellent behaviour of the average fusion rule for all the cases: a linear decrease in the MSE as a function of N is already clearly observed for $N_\ell = 20$.

Table 2: **Univariate Gamma example:** Conditional MSE (averaged over 1000 independent runs) for the three fusion methods considered when $N \in \{10^3, 10^4, 10^5, 10^6\}$ and $N_\ell = N/L \in \{5, 10, 20, 50, 100, 200, N\}$.

Experiment		N_ℓ						
N	**Estimator**	5	10	20	50	100	200	N
10^3	EWF	0.3480	0.1011	0.0369	0.0193	0.0159	0.0148	0.0143
	L-MMSE	0.2042	0.0637	0.0243	0.0147	0.0141	0.0142	
	Average	0.0191	0.0152	0.0144	0.0143	0.0143	0.0143	
10^4	EWF	0.3067	0.0695	0.0172	0.0035	0.0017	0.0013	0.0012
	L-MMSE	0.2104	0.0598	0.0170	0.0038	0.0019	0.0014	
	Average	0.0034	0.0013	0.0012	0.0012	0.0012	0.0012	
10^5	EWF	0.3086	0.0695	0.0166	0.0027	0.0008	0.0003	0.0001
	L-MMSE	0.2058	0.0566	0.0149	0.0025	0.0007	0.0003	
	Average	0.0027	0.0003	0.0002	0.0001	0.0001	0.0001	
10^6	EWF	0.3081	0.0691	0.0164	0.0025	0.0006	0.0002	$1.2 \cdot 10^{-5}$
	L-MMSE	0.2069	0.0568	0.0149	0.0025	0.0006	0.0002	
	Average	0.0025	0.0002	$2 \cdot 10^{-5}$	$1.2 \cdot 10^{-5}$	$1.2 \cdot 10^{-5}$	$1.2 \cdot 10^{-5}$	

5.2. Multi-Variate Gaussian Distributions

Let us assume that we want to estimate a D-dimensional parameter vector, $\mathbf{x} \in \mathbb{R}^{D\times 1}$ for $D \geq 1$, from N independent D-dimensional observations that are divided into L groups, $\mathbf{Y} = [\mathbf{Y}_1, \ldots, \mathbf{Y}_L]$, where $\mathbf{Y}_\ell = [\mathbf{y}_1^{(\ell)}, \ldots, \mathbf{y}_{N_\ell}^{(\ell)}] \in \mathbb{R}^{D\times N_\ell}$ (with $\mathbf{y}_n^{(\ell)} \in \mathbb{R}^{D\times 1}$) is the set of observations available to the ℓ-th partial estimator. Let us consider a linear relationship among the $\mathbf{y}_n^{(\ell)}$ and $\mathbf{x}$:

$$\mathbf{y}_n^{(\ell)} = \mathbf{A}\mathbf{x} + \mathbf{w}_n^{(\ell)}, \tag{64}$$

where $\mathbf{w}_n^{(\ell)}$ follows a multi-variate Gaussian distribution, $\mathbf{w}_n^{(\ell)} \sim \mathcal{N}(\mathbf{w}_n^{(\ell)}|\mathbf{0}, \mathbf{\Sigma}_\ell)$. The likelihood is then another multi-variate Gaussian,

$$p(\mathbf{y}_1^{(\ell)}, \ldots, \mathbf{y}_{N_\ell}^{(\ell)}|\mathbf{x}) = \prod_{n=1}^{N_\ell} \mathcal{N}(\mathbf{y}_n^{(\ell)}|\mathbf{A}\mathbf{x}, \mathbf{\Sigma}_\ell) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{\mathbf{y}}^{(\ell)}, \mathbf{C}_{\mathbf{y}}^{(\ell)}), \tag{65}$$

where $\boldsymbol{\mu}_{\mathbf{y}}^{(\ell)} = \mathbf{A}^{-1}\cdot\left(\frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\mathbf{y}_n^{(\ell)}\right)$ and $\mathbf{C}_{\mathbf{y}}^{(\ell)} = \frac{1}{N_\ell}\left(\mathbf{A}^\top\mathbf{\Sigma}_\ell\mathbf{A}\right)^{-1}$. Let us consider also a multi-variate Gaussian prior, $p_0(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{\mathbf{x}}^{(0)}, \mathbf{C}_{\mathbf{x}}^{(0)})$. The ℓ-th partial

posterior is then another multi-variate Gaussian:

$$p(\mathbf{x}|\mathbf{y}_1^{(\ell)},\ldots,\mathbf{y}_{N_\ell}^{(\ell)}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{\mathbf{x}}^{(\ell)}, \mathbf{C}_{\mathbf{x}}^{(\ell)}), \tag{66}$$

with

$$\mathbf{C}_{\mathbf{x}}^{(\ell)} = \left(\left(\mathbf{C}_{\mathbf{x}}^{(0)}\right)^{-1} + \left(\mathbf{C}_{\mathbf{y}}^{(\ell)}\right)^{-1}\right)^{-1}, \tag{67a}$$

$$\boldsymbol{\mu}_{\mathbf{x}|\mathbf{y}}^{(\ell)} = \mathbf{C}_{\mathbf{x}}^{(\ell)}\left(\left(\mathbf{C}_{\mathbf{x}}^{(0)}\right)^{-1}\boldsymbol{\mu}_{\mathbf{x}}^{(0)} + \left(\mathbf{C}_{\mathbf{y}}^{(\ell)}\right)^{-1}\boldsymbol{\mu}_{\mathbf{y}}^{(\ell)}\right). \tag{67b}$$

Once more, we do not concentrate on any particular application, but this multi-variate Gaussian model arises in many statistical signal processing applications.

For the experiments, we set $\mathbf{A} = \mathbf{I}$ (implying that $\boldsymbol{\mu}_{\mathbf{y}}^{(\ell)} = \frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\mathbf{y}_n^{(\ell)}$ and $\mathbf{C}_{\mathbf{y}}^{(\ell)} = \frac{1}{N_\ell}\boldsymbol{\Sigma}_\ell^{-1}$), and randomly generate the parameter vectors $\mathbf{x}$ and the noise vectors $\mathbf{w}_n^{(\ell)}$. On the one hand, the elements of $\mathbf{x}$ are randomly and independently drawn from a zero-mean Gaussian with variance equal to 100, i.e., $\mathbf{x}_d \sim \mathcal{N}(0, 100)$. On the other hand, the covariance matrix for the noise is given by $\boldsymbol{\Sigma}_\ell = \sigma_\ell^2\mathbf{I} + \rho\mathbf{b}_\ell^\top\mathbf{b}_\ell$, with the elements of $\mathbf{b}_\ell$ randomly and independently drawn from a zero-mean and unit-variance Gaussian, $\sigma_\ell = |\varsigma_\ell|$ with $\varsigma_\ell \sim \mathcal{N}(0, 10)$, and $\rho \geq 0$ a user-dependent parameter that allows us to control the correlation among the outputs. In the simulations, we test the performance of the different fusion rules (EWF, SC-MMSE, IL-MMSE and L-MMSE) for different dimensions of the state space ($1 \leq D \leq 20$), correlation factors ρ ($\rho \in \{0, 0.2, 0, 4\}$), and number of partial estimators ($L \in \{1, 2, 5, 10, 20, 50, 100, 200, 500, 1000, 2000, 5000\}$) in two different scenarios:

Sc1: The matrix $\mathbf{C}_{\mathbf{y}}^{(\ell)}$ is assumed to be known.

Sc2: The matrix $\mathbf{C}_{\mathbf{y}}^{(\ell)}$ is unknown and is substituted by the empirical estimate obtained from the data.

In all cases, we set $\boldsymbol{\mu}_0 = \mathbf{0}$ and $\mathbf{C}_0 = \mathbf{I}$ for the prior, which lead to the following

expressions for the mean and covariance of the posterior PDF:

$$\mathbf{C}_{\mathbf{x}}^{(\ell)} = \left(\mathbf{I} + \left(\mathbf{C}_{\mathbf{y}}^{(\ell)}\right)^{-1}\right)^{-1}, \tag{68a}$$

$$\boldsymbol{\mu}_{\mathbf{x}}^{(\ell)} = \mathbf{C}_{\mathbf{x}}^{(\ell)} \left(\mathbf{C}_{\mathbf{y}}^{(\ell)}\right)^{-1} \boldsymbol{\mu}_{\mathbf{y}}^{(\ell)} = \boldsymbol{\mu}_{\mathbf{y}}^{(\ell)} - \mathbf{C}_{\mathbf{y}}^{(\ell)} \left(\mathbf{I} + \mathbf{C}_{\mathbf{y}}^{(\ell)}\right)^{-1} \boldsymbol{\mu}_{\mathbf{y}}^{(\ell)}, \tag{68b}$$

where we have used the matrix inversion lemma [6] in order to obtain the final expression of Eq. (68b).

The results (averaged over 100 realizations for each combination of parameters tested) are shown in Tables 3 and 4 for Sc1 and Sc2, respectively. The total number of data generated is $N = 10^5$, leading to a number of data per partial estimator $N_\ell \in \{10^5, 5 \cdot 10^4, 2 \cdot 10^4, 10^4, 5 \cdot 10^3, 2 \cdot 10^3, 10^3, 500, 200, 100, 50, 20\}$. Note that the first case (i.e., $L = 1$ and thus $N_\ell = N = 10^5$) corresponds to the global estimator that has access to all the data. From these two tables we can extract the following conclusions:

- For $\rho = 0$ the IL-MMSE fusion rule shows the same performance as the L-MMSE approach when the matrix $\mathbf{C}_{\mathbf{y}}^{(\ell)}$ is known (Sc1), and even better when $\mathbf{C}_{\mathbf{y}}^{(\ell)}$ is unknown (Sc2), due to the increased robustness in the estimation of D parameters per partial estimator instead of D^2. This improvement can be appreciated especially when N_ℓ is small (e.g., when $N_\ell = 200$ and $\mathbf{C}_{\mathbf{y}}^{(\ell)}$ has to be estimated from the data, the IL-MMSE scheme provides an improvement of 2.56 dB w.r.t. the L-MMSE rule).

- Both the IL-MMSE and the L-MMSE fusion rules outperform the SC-MMSE approach and the EWF substantially for $\rho = 0$. For example, when $\mathbf{C}_{\mathbf{y}}^{(\ell)}$ is known (Sc1) the L-MMSE provides an improvement w.r.t. the EWF ranging from 1.25 dB when $N_\ell = 5 \cdot 10^4$ up to 6.79 dB for $N_\ell = 500$.

- For $\rho = 0.2$ the performance of all the fusion rules decreases in general, but the L-MMSE approach still outperforms the EWF by a similar amount as before (ranging from 1.41 dB when $N_\ell = 5 \cdot 10^4$ up to 6.31 dB for

$N_\ell = 500$ in Sc1). The IL-MMSE rule decreases its performance w.r.t. the L-MMSE scheme, especially in Sc2, but its performance is still better than the EWF in general (up to 2.66 dB when $N_\ell = 200$ in Sc1, and up to 2.25 dB when $N_\ell = 10^4$ in Sc2).

- As the correlation increases (i.e., for $\rho = 0.4$), the L-MMSE fusion rule still outperforms the EWF by a similar amount as in the two previous cases. The performance of the IL-MMSE fusion rule deteriorates w.r.t. $\rho = 0.2$, but it can still be advantageous w.r.t. the EWF (e.g., 1.76 dB are gained in Sc1 when $N_\ell = 200$ and 1.19 dB are gained in Sc2 when $N_\ell = 10^4$).

- The SC-MMSE fusion rule attains a worse performance than the EWF in all cases. However, in Section 5.3 we will show that it outperforms the EWF in two localization examples.

- Regarding the behavior of the different approaches as the number of data per partial estimator (N_ℓ) decreases (i.e., the number of partial estimators (L) increases for a fixed total amount of data N), it seems to be rather constant for the SC-MMSE and EWF schemes, no pattern can be clearly seen for the IL-MMSE fusion rule, and there seems to be an optimal value around $N_\ell = 100$ (i.e., $L = 1000$) for the L-MMSE approach.

Finally, the performance of the different methods as the dimension of the state space (D) increases is shown in Figures 4 and 5. Note that the performance of all the methods remains quite stable (after an initial change for small values of D when $\rho = 0.2$) as the dimension of the state space increases.

5.3. Localization in a Wireless Sensor Network

5.3.1. Case 1: Estimation of the Target's Position

In this section, we address the problem of positioning a static target in the two-dimensional space of a wireless sensor network using only range measurements.

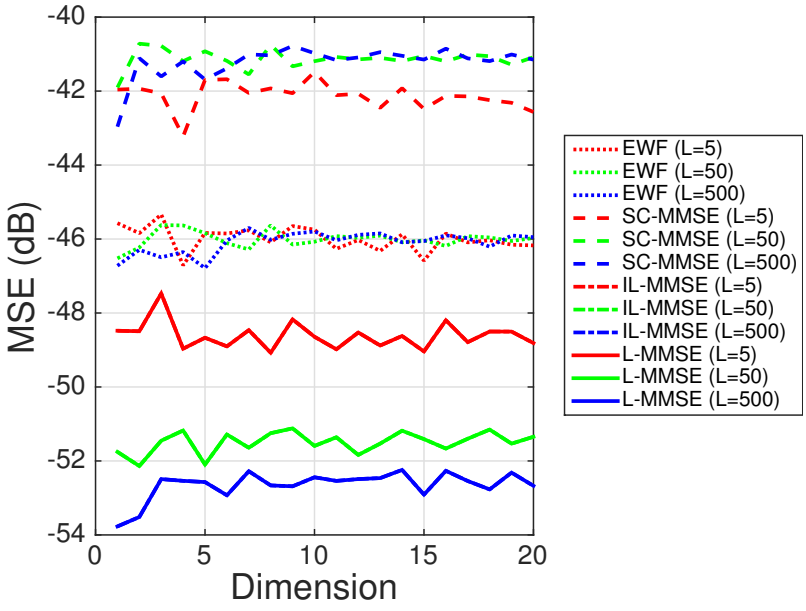
Table 3: **Multi-variate Gaussian Example:** MSE in dB (averaged over 100 independent runs) for the four fusion methods considered when the matrix $\mathbf{\Sigma}_\ell$ is known, $N = 10^5$, $L \in \{2, 10, 50, 200, 1000, 5000\}$ (i.e., $N_\ell = N/L \in \{5 \cdot 10^4, 10^4, 2 \cdot 10^3, 500, 100, 20\}$), $D = 10$, and $\rho \in \{0, 0.2, 0.4\}$.

Experiment		$\boldsymbol{N_\ell}$					
Correlation	**Estimator**	$5 \cdot 10^4$	10^4	2000	500	100	200
$\rho = 0$	EWF	-46.43	-45.98	-46.07	-45.97	-46.07	-45.78
	SCMSE	-43.97	-41.59	-41.19	-41.07	-40.86	-41.01
	ILMSE	-47.68	-49.48	-51.59	-52.76	-52.45	-51.32
	LMSE	-47.68	-49.48	-51.59	-52.76	-52.45	-51.32
$\rho = 0.2$	EWF	-45.09	-45.35	-45.68	-45.61	-45.64	-45.81
	SCMSE	-42.75	-41.27	-40.89	-40.74	-40.80	-41.23
	ILMSE	-46.38	-47.60	-46.32	-44.07	-46.62	-48.47
	LMSE	-46.50	-48.72	-50.83	-51.92	-51.93	-51.15
$\rho = 0.4$	EWF	-45.16	-45.26	-45.45	-45.22	-45.34	-45.31
	SCMSE	-43.38	-41.21	-41.08	-40.49	-40.78	-40.65
	ILMSE	-45.77	-46.45	-45.66	-40.47	-44.16	-47.07
	LMSE	-45.99	-48.92	-50.58	-52.07	-52.12	-50.90

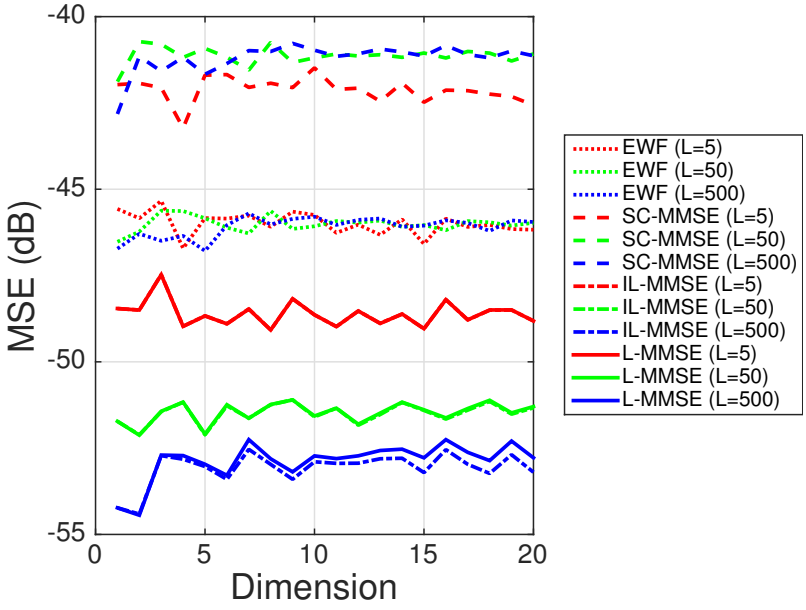
Table 4: **Multi-variate Gaussian Example:** MSE in dB (averaged over 100 independent runs) for the four fusion methods considered when the matrix $\mathbf{\Sigma}_\ell$ is unknown, $N = 10^5$, $L \in \{2, 10, 50, 200, 1000, 5000\}$ (i.e., $N_\ell = N/L \in \{5 \cdot 10^4, 10^4, 2 \cdot 10^3, 500, 100, 20\}$), $D = 10$, and $\rho \in \{0, 0.2, 0.4\}$.

Experiment		$\boldsymbol{N_\ell}$					
Correlation	**Estimator**	$5 \cdot 10^4$	10^4	2000	500	100	200
$\rho = 0$	EWF	-46.43	-45.98	-46.07	-45.97	-46.07	-45.78
	SCMSE	-43.97	-41.59	-41.19	-41.07	-40.83	-40.70
	ILMSE	-47.68	-49.47	-51.59	-52.84	-53.63	-51.62
	LMSE	-47.68	-49.45	-51.57	-52.69	-53.15	-49.06
$\rho = 0.2$	EWF	-45.09	-45.35	-45.68	-45.61	-45.64	-45.81
	SCMSE	-42.75	-41.27	-40.89	-40.74	-40.77	-40.94
	ILMSE	-46.38	-47.60	-46.22	-42.54	-37.37	-45.20
	LMSE	-46.50	-48.72	-50.78	-51.84	-52.56	-48.75
$\rho = 0.4$	EWF	-45.16	-45.26	-45.45	-45.22	-45.34	-45.31
	SCMSE	-43.38	-41.21	-41.08	-40.49	-40.74	-40.21
	ILMSE	-45.77	-46.45	-45.61	-38.65	-33.75	-42.90
	LMSE	-45.99	-48.90	-50.56	-52.08	-52.78	-48.22

More specifically, we consider a random vector $\mathbf{X} = [X_1, X_2]^\top$ to denote the target's position in the $\mathbb{R}^2$ plane. The position of the target is then a specific realization $\mathbf{x}$. The measurements are obtained from 6 range sensors located at

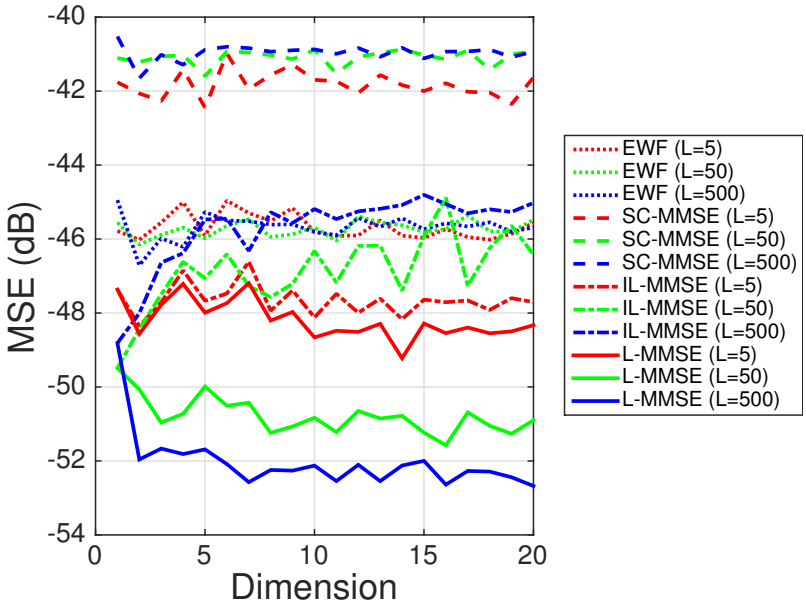


(a)

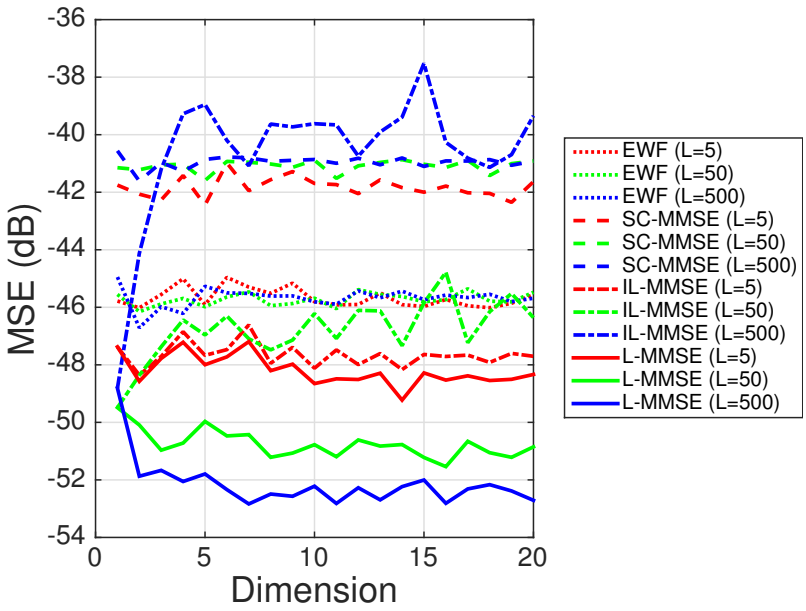


(b)

Figure 4: **Multi-variate Gaussian Example:** MSE as a function of D for $\rho = 0$, $N = 10^5$ and $N_\ell = N/L$. (a) $\mathbf{\Sigma}_\ell$ known. (b) $\mathbf{\Sigma}_\ell$ estimated from the data.



(a)



(b)

Figure 5: **Multi-variate Gaussian Example:** MSE as a function of D for $\rho = 0.2$, $N = 10^5$ and $N_\ell = N/L$. (a) $\mathbf{\Sigma}_\ell$ known. (b) $\mathbf{\Sigma}_\ell$ estimated from the data.

$\mathbf{h}_1 = [1, -8]^\top$, $\mathbf{h}_2 = [8, 10]^\top$, $\mathbf{h}_3 = [-15, -7]^\top$, $\mathbf{h}_4 = [-8, 1]^\top$, $\mathbf{h}_5 = [10, 0]^\top$, and $\mathbf{h}_6 = [0, 10]^\top$. The measurement equations are

$$Y_j = -10 \log\left(\|\mathbf{x} - \mathbf{h}_j\|_2^2\right) + \Theta_j, \quad j = 1, \ldots, 6, \tag{69}$$

where $\Theta_j \sim \mathcal{N}(\theta_j|\mathbf{0}, \omega_j^2\mathbf{I})$, with $\omega_j = 5$ for $j \in \{1, 2, 3\}$ and $\omega_j = 20$ for $j \in \{4, 5, 6\}$. We simulate $N = 6000$ observations from the model ($\frac{N}{6} = 1000$ observations from each sensor) fixing $x_1 = x_2 = 3.5$. We consider a varying number of partial estimators L with $N_\ell = N/L$ for $1 \leq \ell \leq L$, and three scenarios for splitting the data:

Sc1: Exactly $\frac{N}{6L}$ measurements from each sensor are provided to each partial estimator.

Sc2: The first $L/2$ estimators contain an equal number of observations from the first 3 sensors (the best ones), whereas the remaining $L/2$ estimators work with measurements from the last 3 sensors (the noisiest ones).

Sc3: Measurements are randomly assigned to the estimators.

For each scenario, we run $M_C^{(\ell)} = 100$ MCMC independent parallel chains with length $T_C^{(\ell)} = 5000$, compute the MMSE estimates $\hat{x}_1^{(\ell)}$ and $\hat{x}_2^{(\ell)}$, and fuse these estimates into the final result. We compare the Equal Weights Fusion (EWF) method, where each estimator is given the same weight, $1/L$, and the three fusion methods described in the paper. We repeat the experiments 50 times and average the results. The results, shown in Table 5 and Figures 6–8, confirm the good performance of the SC-MMSE and IL-MMSE estimators, which outperform the naive EWF and show an MSE similar to the optimal and more costly L-MMSE. Regarding the three scenarios considered, we note that the best performance is obtained in the second case (with $\text{MSE}(\hat{\mathbf{x}}^{(\text{L-MMSE})}|\mathbf{y}) = 0.0021$), i.e., splitting the data into separate filters according to their quality. This opens up the possibility of performing a "smart" division of the data in order to optimize the performance.

Finally, in order to study the scaling behaviour of the fusion rules as N increases, we have also simulated the three scenarios for $N = 30000$, as well

Table 5: **Localization Example (Case 1):** MSE (averaged over 50 independent runs) for the three scenarios and the four fusion methods considered when $N = 6000$, $L \in \{5, 10, 25, 100, 200, 500, 1000\}$, and $N_\ell = N/L \in \{6, 12, 30, 60, 240, 600, 1200\}$.

Experiment		N_ℓ						
Scenario	**Estimator**	6	12	30	60	240	600	1200
Sc1	EWF	0.0041	0.0049	0.0065	0.0090	0.0167	0.0590	0.1192
	SCMSE	0.0039	0.0046	0.0063	0.0089	0.0166	0.0587	0.1191
	ILMSE	0.0038	0.0046	0.0063	0.0089	0.0166	0.0586	0.1188
	LMSE	0.0037	0.0045	0.0062	0.0088	0.0165	0.0584	0.1183
Sc2	EWF	0.0087	0.0053	0.0064	0.0104	0.0343	0.0648	0.1681
	SCMSE	0.0057	0.0034	0.0047	0.0092	0.0328	0.0628	0.1623
	ILMSE	0.0052	0.0031	0.0043	0.0085	0.0304	0.0588	0.1521
	LMSE	0.0037	0.0021	0.0028	0.0057	0.0210	0.0410	0.1107
Sc3	EWF	0.0078	0.0061	0.0068	0.0092	0.0169	0.0587	0.1181
	SCMSE	0.0060	0.0053	0.0066	0.0091	0.0168	0.0584	0.1180
	ILMSE	0.0055	0.0051	0.0065	0.0090	0.0168	0.0583	0.1177
	LMSE	0.0051	0.0048	0.0064	0.0090	0.0167	0.0582	0.1174

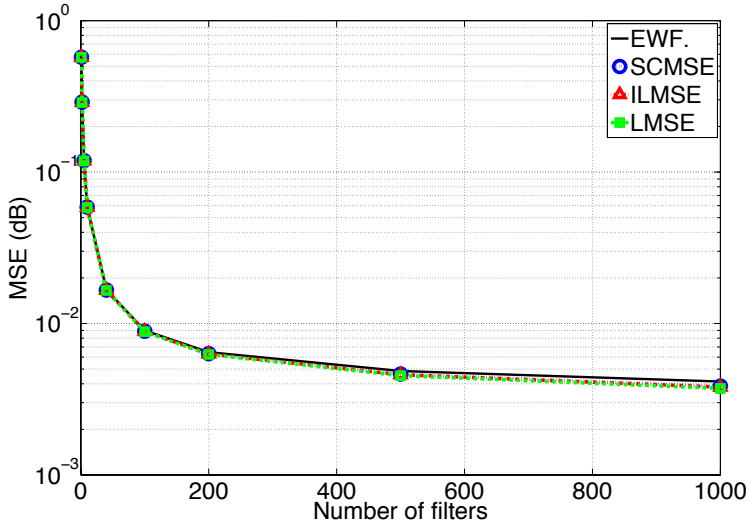


Figure 6: **Localization Example (Case 1):** MSE as a function of L for Scenario 1 (Sc1) when $N = 6000$.

as Scenario 2 for $N = 600000$. The results, displayed in Table 6 and Figure 9, respectively, show that the performance of all the fusion rules scales roughly as a function of the number of samples, N.

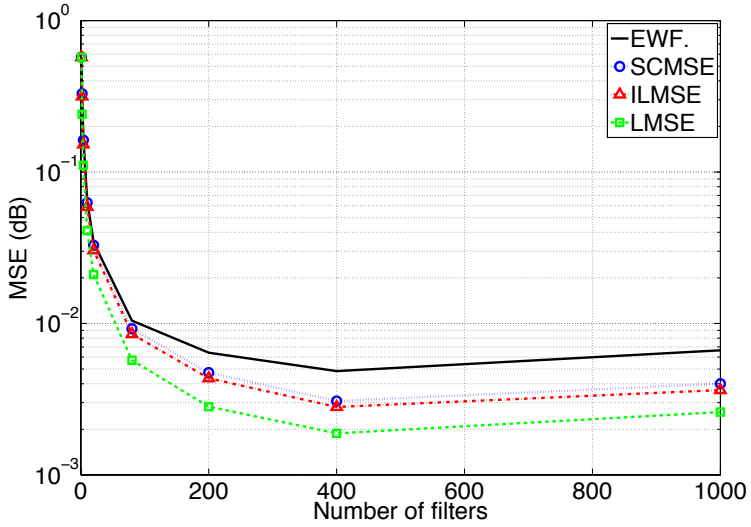


Figure 7: **Localization Example (Case 1):** MSE as a function of L for Scenario 2 (Sc2) when $N = 6000$.

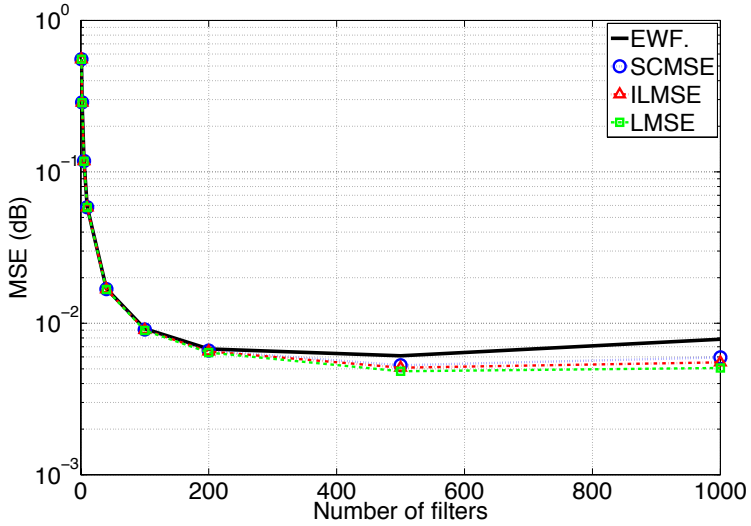


Figure 8: **Localization Example (Case 1):** MSE as a function of L for Scenario 3 (Sc3) when $N = 6000$.

Table 6: **Localization Example (Case 1):** MSE (averaged over 50 independent runs) for the three scenarios and the four fusion methods considered when $N = 30000$, $L \in \{25, 50, 125, 500, 1000, 2500, 5000\}$, and $N_\ell = N/L \in \{6, 12, 30, 60, 240, 600, 1200\}$.

Experiment		N_ℓ						
Scenario	**Estimator**	6	12	30	60	240	600	1200
Sc1	EWF	0.0008	0.001	0.0013	0.0018	0.0033	0.0117	0.0231
	SC-MMSE	0.0008	0.0009	0.0013	0.0018	0.0033	0.0117	0.0230
	IL-MMSE	0.0008	0.0009	0.0013	0.0018	0.0033	0.0117	0.0230
	L-MMSE	0.0007	0.0009	0.0012	0.0017	0.0033	0.0116	0.0229
Sc2	EWF	0.0007	0.0009	0.0012	0.0018	0.0036	0.0131	0.0335
	SC-MMSE	0.0004	0.0006	0.0009	0.0015	0.0033	0.0125	0.0323
	IL-MMSE	0.0004	0.0005	0.0009	0.0014	0.0031	0.0118	0.0304
	L-MMSE	0.0003	0.0003	0.0006	0.0009	0.0021	0.0082	0.0214
Sc3	EWF	0.0018	0.0011	0.0013	0.0018	0.0033	0.0118	0.0229
	SC-MMSE	0.0014	0.001	0.0013	0.0018	0.0033	0.0118	0.0228
	IL-MMSE	0.0013	0.001	0.0013	0.0018	0.0033	0.0118	0.0228
	L-MMSE	0.0012	0.001	0.0013	0.0018	0.0033	0.0117	0.0227

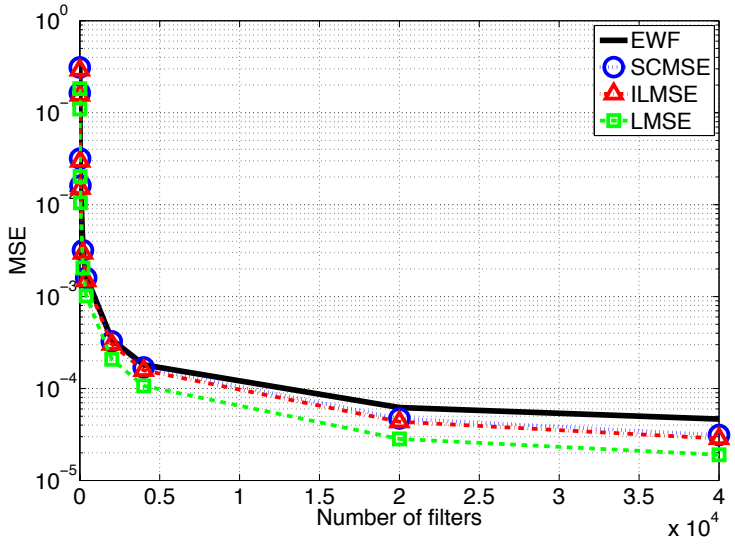


Figure 9: **Localization Example (Case 1):** Conditional MSE as a function of L for Scenario 3 (Sc2) when $N = 600000$.

5.3.2. Case 2: Estimation of the Target's Position and the Sensors' Noise Variance

In this section, we address a more complicated version of the previous case. We still consider the localization of a single target (placed at $x_1 = x_2 = 3.5$,

as before), using range-only measurements from 6 sensors placed at the same locations as in case 1: $\mathbf{h}_1 = [1, -8]^\top$, $\mathbf{h}_2 = [8, 10]^\top$, $\mathbf{h}_3 = [-15, -7]^\top$, $\mathbf{h}_4 = [-8, 1]^\top$, $\mathbf{h}_5 = [10, 0]^\top$ and $\mathbf{h}_6 = [0, 10]^\top$. However, in addition to the estimation of the target's location, now we also estimate the noise variance of the different sensors, implying that the dimension of the state space in this case is $D = 8$. The ground-truth values of the noise variances used in the simulations are $\omega_1 = 1$, $\omega_2 = 2$, $\omega_3 = 1$, $\omega_4 = 0.5$, $\omega_5 = 3$, and $\omega_6 = 0.2$. Furthermore, we investigate the performance of the proposed approaches when a different number of data is available to each the partial estimators. Namely, we generate 87 data from each sensor (i.e., $N = 87 \times 6 = 522$ data overall) according to the model in Eq. (69) and split them randomly among $L = 5$ partial estimators in such a way that $N_1 = 12$, $N_2 = 30$, $N_3 = 60$, $N_4 = 120$, and $N_5 = 300$.

We test the performance of the optimal L-MMSE estimator, comparing it to the EWF approach and the two novel efficient fusion methods proposed (IL-MMSE and SC-MMSE). Moreover, we also test two heuristic fusion rules that take into account the unequal number of data available to each of the partial estimators: a linearly proportional weights fusion (L-PWF) scheme that assigns weights to the partial estimates as $w_\ell = N_\ell / N$ (i.e., in our case $w_1 = 2/87$, $w_2 = 5/87$, $w_3 = 10/87$, $w_4 = 20/87$, and $w_5 = 50/87$); and a square root PWF (SR-PWF) scheme that assigns weights as $w_\ell = \sqrt{N_\ell}/S$, with $S = \sum_{\ell=1}^{L} \sqrt{N_\ell}$ (i.e., in our case, $S = \sqrt{2} + \sqrt{5} + \sqrt{10} + \sqrt{20} + \sqrt{50} \approx 18.3558$, $w_1 = \sqrt{2}/S \approx 0.0770$, $w_2 = \sqrt{5}/S \approx 0.1218$, $w_3 = \sqrt{10}/S \approx 0.1723$, $w_4 = \sqrt{20}/S \approx 0.2436$, and $w_5 = \sqrt{50}/S \approx 0.3852$). The rationale behind these two heuristic fusion rules is that the estimates provided by estimators that have more data available should be trusted more, since the precision in the estimation is directly proportional to the number of data (L-PWF implicitly assumes that the precision is proportional to N_ℓ, whereas SR-PWF assumes that it is proportional to $\sqrt{N_\ell}$).

Figure 10 shows the MSE for all the different fusion rules considered as a function of the number of samples of the Markov chain. First of all, it is remarkable the increase in performance as the number of iterations of the chain increases: around $5 \cdot 10^3$ iterations are required in order to attain a good per-

formance in all cases. Then, note the similar performance of the three heuristic approaches (EWF, L-PWF and SR-PWF), with EWF performing slightly better than the other two (showing that having weights directly proportional to N_ℓ or $\sqrt{N_\ell}$ is not a good idea in this case). Finally, note also that the IL-MMSE fusion rule always outperforms the optimal L-MMSE fusion, whereas the SC-MMSE approach performs better than the L-MMSE fusion for less than 10^4 iterations. This shows the increased robustness of the two proposed efficient fusion rules, which require the estimation of a reduced number of parameters from the noisy data (in this case, $D = 8$ parameters per partial estimator for the IL-MMSE approach and just one parameter for the SC-MMSE fusion rule, instead of the $D^2 = 64$ parameters per partial estimator of the L-MMSE rule). Furthermore, both of the efficient fusion rules proposed are able to provide an improved performance w.r.t. the naive EWF. For instance, using $5 \cdot 10^4$ iterations of the chain the SC-MMSE fusion rule achieves ≈ 1 dB reduction in MSE w.r.t. the EWF with the transmission of a single weight per partial estimator, whereas the IL-MMSE approach attains a reduction of 1.69 dB by transmitting $D = 8$ coefficients per partial estimator.

6. Conclusions

In this paper, we have addressed the linear fusion of unbiased and independent partial estimators, focusing on the scenario where no communication is allowed among them. The best linear fusion rule has been derived and two novel efficient linear combination schemes, that achieve an excellent trade-off between the amount of information that has to be transmitted to the fusion center and the performance in terms of mean squared error (MSE) of the final estimator, have been proposed. Both a constrained optimization point of view and a Bayesian perspective have been provided for these three fusion rules, allowing us to explore their connections and to explain their good performance even in situations where the assumptions that lead to them are not strictly fulfilled. All the methods were tested through computer simulations by applying them

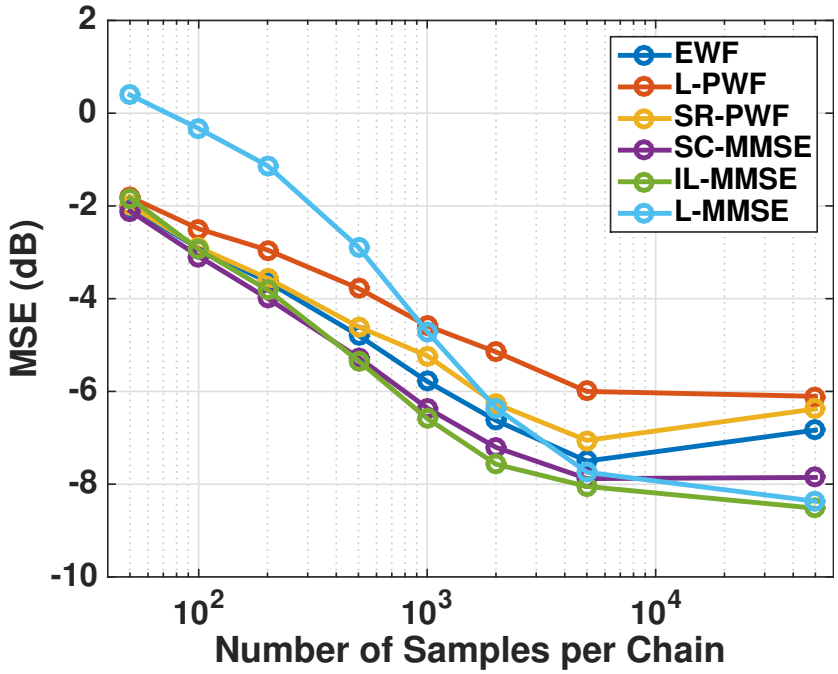


Figure 10: **Localization Example (Case 2):** MSE as a function of the number of samples per chain for all the different fusion rules considered.

to three problems. First of all, we considered a univariate problem, where all the posterior densities followed a Gamma PDF and all the estimators could be computed analytically. We use this example to analyze the performance of the fusion rules as the number of samples per estimator decreases and the partial estimators cannot be considered unbiased any more. Then, we address a parameter estimation problem in a multi-variate Gaussian model, where we increase the dimension of the state space up to $D = 20$, showing the good performance and robustness of the two efficient fusion rules proposed. Finally, we tackle a localization problem with one target and six sensors whose measurements were processed using several parallel filters. Here we consider two cases: localization of the target when the noise characteristics of the sensors are known (dimension of the state space $D = 2$), and joint localization of the target and estimation of the noise characteristics of the sensors (dimension of the state space $D = 8$). The new fusion methods show a performance equivalent to the optimal linear

combination (sometimes even better) with a reduced computational cost. Furthermore, it has been shown that splitting the data can be advantageous in terms of attaining a good MSE with a reduced computational cost, but only when the bias in the partial estimators can be controlled. In future works we plan to address bias correction approaches, as well as optimal linear fusion schemes for biased and/or correlated partial estimators. Some other interesting areas of research are non-linear fusion techniques and the development of fusion schemes where the partial Monte Carlo estimators are allowed to exchange a reduced amount of information.

Acknowledgements

This work has been supported by Ministerio de Economía y Competitividad of Spain through the MIMOD-PLC project (TEC2015-64835-C3-3-R); Ministerio de Educación, Cultura y Deporte of Spain under CAS15/00350 grant; Universidad Politécnica de Madrid through a mobility grant for a short visit of D. Luengo to Stony Brook University; the BBVA Foundation through project MG-FIAR ("I Convocatoria de Ayudas Fundación BBVA a Investigadores, Innovadores y Creadores Culturales"); the National Science Foundation under Award CCF-1617986; and the European Research Council (ERC) through the ERC Consolidator Grant SEDAL ERC-2014-CoG 647423.

References

[1] H. L. Van Trees, Detection, Estimation and Modulation Theory, John Wiley and Sons, Hoboken, NJ (USA), 1968.

[2] G. Casella, R. L. Berger, Statistical inference, Duxbury, 2002.

[3] L. L. Scharf, Statistical Signal Processing, Addison-Wesley, Reading, MA (USA), 1991.

[4] S. M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Prentice Hall, Upper Saddle River, NJ (USA), 1993.

[5] J. M. Mendel, Lessons in estimation theory for signal processing, communications, and control, Pearson Education, New York, NY (USA), 1995.

[6] C. Rasmussen, C. Williams, Gaussian Processes for Machine Learning, MIT Press, 2006.

[7] N. L. Hjort, C. Holmes, P. Müller, S. G. Walker, Bayesian nonparametrics, Vol. 28, Cambridge University Press, Cambridge (UK), 2010.

[8] J. D. Gibbons, S. Chakraborti, Nonparametric statistical inference, Springer, 2011.

[9] G. B. Giannakis, F. Bach, R. Cendrillon, M. Mahoney, J. Neville, Signal processing for big data (special issue), IEEE Signal Processing Magazine 31 (5) (2014) 15–111.

[10] P. M. Djuric, S. J. Godsill, Special issue on Monte Carlo methods for statistical signal processing, IEEE Transactions on Signal Processing 50 (2) (2002) 173–173.

[11] C. P. Robert, G. Casella, Monte Carlo Statistical Methods, Springer, 2004.

[12] A. Doucet, X. Wang, Monte Carlo methods for signal processing: a review in the statistical signal processing context, Signal Processing Magazine, IEEE 22 (6) (2005) 152–170.

[13] N. Chopin, A sequential particle filter method for static models, Biometrika 89 (3) (2002) 539–552.

[14] M. A. Suchard, Q. Wang, C. Chan, J. Frelinger, A. Cron, M. West, Understanding GPU programming for statistical computation: Studies in massively parallel massive mixtures, Journal of Computational and Graphical Statistics 19 (2) (2010) 419–438.

[15] A. Lee, C. Yau, M. B. Giles, A. Doucet, C. C. Holmes, On the utility of graphics cards to perform massively parallel simulation of advanced Monte Carlo methods, Journal of computational and graphical statistics 19 (4) (2010) 769–789.

[16] J. Dean, S. Ghemawat, Mapreduce: simplified data processing on large clusters, Communications of the ACM 51 (1) (2008) 107–113.

[17] K. F. Wallis, Combining forecasts – forty years later, Applied Financial Economics 21 (1–2) (2011) 33–41.

[18] J. M. Bates, C. W. Granger, The combination of forecasts, Operational Research Quarterly 20 (4) (1969) 451–468.

[19] J. P. Dickinson, Some statistical results in the combination of forecasts, Operational Research Quarterly 24 (1975) 253–260.

[20] R. F. Bordley, The combination of forecasts: A Bayesian approach, Journal of the Operational Research Society 33 (2) (1982) 171–174.

[21] F. Lavancier, P. Rochet, A general procedure to combine estimators, arXiv preprint arXiv:1401.6371 (2014) 1–38.

[22] G. M. Allenby, Cross-validation, the Bayes theorem, and small-sample bias, Journal of Business & Economic Statistics 8 (2) (1990) 171–178.

[23] J. B. Predd, S. R. Kulkarni, H. V. Poor, Distributed learning in wireless sensor networks, IEEE Signal Processing Magazine 23 (4) (2006) 56–69.

[24] J.-J. Xiao, A. Ribeiro, Z.-Q. Luo, G. B. Giannakis, Distributed compression-estimation using wireless sensor networks, IEEE Signal Processing Magazine 23 (4) (2006) 27–41.

[25] A. Swami, Q. Zhao, Y.-W. Hong, L. Tong (Eds.), Wireless Sensor Networks: Signal Processing and Communications Perspectives, John Wiley and Sons, 2007.

[26] M. Cetin, L. Chen, J. W. F. III, A. T. Ihler, R. L. Moses, M. J. Wainwright, A. S. Willsky, Distributed fusion in sensor networks, IEEE Signal Processing Magazine 23 (4) (2006) 42–55.

[27] R. Olfati-Saber, J. A. Fax, R. M. Murray, Consensus and cooperation in networked multi-agent systems, Proceedings of the IEEE 95 (1) (2007) 215–233.

[28] A. G. Dimakis, S. Kar, J. F. Moura, M. G. Rabbat, A. Scaglione, Gossip algorithms for distributed signal processing, Proceedings of the IEEE 98 (11) (2010) 1847–1864.

[29] F. S. Cattivelli, A. H. Sayed, Diffusion LMS strategies for distributed estimation, IEEE Transactions on Signal Processing 58 (3) (2010) 1035–1048.

[30] D. J. Wilkinson, Parallel Bayesian computation, Statistics Textbooks and Monographs 184.

[31] B. Wilkinson, M. Allen, Parallel programming: techniques and applications using networked workstations and parallel computers, Prentice-Hall, Inc., 1999.

[32] Z. Huang, A. Gelman, Sampling for Bayesian computation with large datasets, Available at SSRN 1010107.

[33] S. L. Scott, A. W. Blocker, F. V. Bonassi, H. A. Chipman, E. I. George, R. E. McCulloch, Bayes and big data: The consensus Monte Carlo algorithm, International Journal of Management Science and Engineering Management 11 (2) (2016) 78–88.

[34] W. Neiswanger, C. Wang, E. Xing, Asymptotically exact, embarrassingly parallel MCMC, arXiv:1311.4780v2 (21 Mar. 2014) 1–16.

[35] X. Wang, D. B. Dunson, Parallelizing MCMC via Weierstrass sampler, arXiv:1312.4605v2 (25 May 2014) 1–35.

[36] M. Rabinovich, E. Angelino, M. I. Jordan, Variational consensus monte carlo, in: Advances in Neural Information Processing Systems (NIPS), 2015, pp. 1207–1215.

[37] X. Wang, F. Guo, K. A. Heller, D. B. Dunson, Parallelizing mcmc with random partition trees, in: Advances in Neural Information Processing Systems (NIPS), 2015, pp. 451–459.

[38] D. Luengo, L. Martino, V. Elvira, M. Bugallo, Efficient linear combination of partial Monte Carlo estimators, in: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2015, pp. 4100–4104.

[39] A. Doucet, X. Wang, Monte Carlo methods for signal processing, IEEE Signal Processing Magazine 22 (6) (2005) 152–170.

[40] S. Boyd, L. Vandenberghe, Convex optimization, Cambridge university press, 2004.

[41] L. Le Cam, Asymptotic Methods in Statistical Decision Theory, Springer-Verlag, New York, NY (USA), 1986.

[42] A. W. van der Vaart, Asymptotic Statistics, Cambridge University Press, 1998.

[43] G. J. Husak, J. Michaelsen, C. Funk, Use of the gamma distribution to represent monthly rainfall in africa for drought monitoring applications, International Journal of Climatology 27 (7) (2007) 935–944.

[44] D. Fink, A compendium of conjugate priors, Technical Report (May 1997) 1–47.

[45] H. Strathmann, D. Sejdinovic, M. Girolami, Unbiased Bayes for big data: Paths of partial posteriors, arXiv preprint arXiv:1501.03326.

[46] D. Luengo, L. Martino, V. Elvira, M. Bugallo, Bias correction for distributed Bayesian estimators, in: IEEE 6th International Workshop on

Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), IEEE, 2015, pp. 253–256.

[47] T. M. Cover, J. A. Thomas, Elements of information theory, John Wiley & Sons, 2012.

Appendices

A. Solution of the Constrained Optimization Problems

A.1. Single Coefficient MMSE (SC-MMSE) Fusion

Let us consider first the SC-MMSE fusion rule. From Eqs. (29a) and (29b), applying the method of the Lagrange multipliers, the cost function that has to be minimized is

$$J_{\text{SC-MMSE}} = \sum_{k=1}^{L} \alpha_k^2 T_k + \lambda \left(\sum_{k=1}^{L} \alpha_k - 1 \right), \tag{70}$$

where $T_k = \text{Tr}(\mathbf{C}_{\mathbf{x}}^{(k)})$. Differentiating w.r.t. α_ℓ ($\ell = 1, \ldots, L$) and equating the result to zero we obtain a set of L equations,

$$\frac{\partial J_{\text{SC-MMSE}}}{\partial \alpha_\ell} = 2\alpha_\ell T_\ell + \lambda = 0, \tag{71}$$

whereas differentiating w.r.t. λ and equating the result to zero we obtain one additional equation:

$$\frac{\partial J_{\text{SC-MMSE}}}{\partial \lambda} = \sum_{k=1}^{L} \alpha_k - 1 = 0. \tag{72}$$

In matrix form, these $L+1$ equations can be expressed compactly as

$$\begin{bmatrix} \mathbf{D} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \lambda \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}, \tag{73}$$

with $\mathbf{D} = \mathrm{diag}(T_1, \ldots, T_L)$, $\mathbf{1} = [1, \ldots, 1]^\top$, $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_L]^\top$, and $\mathbf{0} = [0, \ldots, 0]^\top$. The optimal values of $\boldsymbol{\alpha}$ and λ can thus be obtained as

$$\begin{bmatrix} \boldsymbol{\alpha}^* \\ \lambda^* \end{bmatrix} = \begin{bmatrix} \mathbf{D} & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{P} & \mathbf{q} \\ \mathbf{q}^\top & r \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}, \tag{74}$$

where $\mathbf{P}$, $\mathbf{q}$ and r can be obtained from the block matrix inversion lemma [6]:

$$\mathbf{P} = \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{1}\ r\ \mathbf{1}^\top\mathbf{D}^{-1}, \tag{75a}$$

$$\mathbf{q} = -\mathbf{D}^{-1}\mathbf{1}\ r = r \cdot \left[\frac{1}{2T_1},\ \frac{1}{2T_2},\ \ldots,\ \frac{1}{2T_L}\right], \tag{75b}$$

$$r = -\left(\mathbf{1}^\top\mathbf{D}^{-1}\mathbf{1}\right)^{-1} = \frac{2}{\sum_{k=1}^{L} T_k^{-1}}. \tag{75c}$$

Finally, from Eq. (74) it is straightforward to see that the optimal solution is given by $\lambda^* = r = \frac{2}{\sum_{k=1}^{L} T_k^{-1}}$ and $\boldsymbol{\alpha}^* = \mathbf{q}$, implying that

$$\alpha_\ell^* = \frac{T_\ell^{-1}}{\sum_{k=1}^{L} T_k^{-1}}, \tag{76}$$

which are the optimal coefficients given in Eq. (30).

A.2. Independent Linear MMSE (IL-MMSE) Fusion

Let us consider now the IL-MMSE fusion rule. Since the weighting matrix is now given by $\boldsymbol{\Lambda}_\ell = \mathbf{D}_\ell = \mathrm{diag}(\alpha_{\ell,1}, \ldots, \alpha_{\ell,D})$, the global estimator becomes

$$\hat{\mathbf{x}} = \sum_{k=1}^{L} \mathbf{D}_k \hat{\mathbf{x}}_k = \sum_{k=1}^{L} [\alpha_{k,1}\hat{x}_{k,1},\ \ldots,\ \alpha_{k,D}\hat{x}_{k,D}]^\top. \tag{77}$$

From Eq. (77), it is straightforward to see that the global estimator for the d-th parameter ($d = 1, \ldots, D$),

$$\hat{x}_d = \sum_{k=1}^{L} \alpha_{k,d} \hat{x}_{k,d}, \tag{78}$$

depends only on the L partial estimates of the d-th parameter and not on any of the partial estimates of any other parameter. Therefore, from Eqs. (37a) and (37b) and applying the method of the Lagrange multipliers, the cost function that has to be minimized can be expressed as

$$J_{\text{IL-MMSE}} = \sum_{d=1}^{D} J_{\text{IL-MMSE}}^{(d)} \tag{79a}$$

$$J_{\text{IL-MMSE}}^{(d)} = \sum_{k=1}^{L} \alpha_{k,d}^2 \sigma_{k,d}^2 + \lambda_d \left(\sum_{k=1}^{L} \alpha_{k,d} - 1 \right). \tag{79b}$$

Noting that Eqs. (79a) and (79b) correspond to D independent optimization problems and that (79b) is identical to (70) (with $\alpha_{k,d}$, $\sigma_{k,d}^2$ and λ_d in place of α_k, T_k and λ, respectively), it is obvious that the optimal coefficients for the IL-MMSE fusion rule are given by (38):

$$\alpha_{\ell,d}^* = \frac{\sigma_{\ell,d}^{-2}}{\sum_{k=1}^{L} \sigma_{k,d}^{-2}}. \tag{80}$$

A.3. Linear MMSE (L-MMSE) Fusion

Finally, let us consider the optimal L-MMSE fusion rule. From Eqs. (19a) and (19b), and applying once more the method of the Lagrange multipliers, the cost function that has to be minimized is now

$$J_{\text{L-MMSE}} = \sum_{k=1}^{L} \text{Tr}\left(\mathbf{\Lambda}_k \mathbf{C}_{\mathbf{x}}^{(k)} \mathbf{\Lambda}_k^\top \right) + \lambda \left(\sum_{k=1}^{L} \mathbf{\Lambda}_k - \mathbf{I} \right), \tag{81}$$

where $\mathbf{I}$ denotes the $D \times D$ identity matrix. Differentiating w.r.t. $\mathbf{\Lambda}_\ell$ ($\ell = 1, \ldots, L$) and equating the result to zero we obtain a set of LD^2 equations,

$$\frac{\partial J_{\text{L-MMSE}}}{\partial \mathbf{\Lambda}_\ell} = 2\mathbf{C}_{\mathbf{x}}^{(\ell)} \mathbf{\Lambda}_\ell^\top + \lambda \mathbf{I} = \mathbf{0}, \tag{82}$$

whereas differentiating w.r.t. λ and equating the result to zero we obtain one additional equation:

$$\frac{\partial J_{\text{L-MMSE}}}{\partial \lambda} = \sum_{k=1}^{L} \mathbf{\Lambda}_k - \mathbf{I} = \mathbf{0}. \tag{83}$$

And now, it can be easily checked that the weighting matrix given by (48),

$$\mathbf{\Lambda}_\ell^* = \left[\sum_{k=1}^{L} \left(\mathbf{C}_{\mathbf{x}}^{(k)} \right)^{-1} \right]^{-1} \left(\mathbf{C}_{\mathbf{x}}^{(\ell)} \right)^{-1},$$

and the regularization parameter,

$$\lambda^* = 2 \left[\sum_{k=1}^{L} \left(\mathbf{C}_{\mathbf{x}}^{(k)} \right)^{-1} \right]^{-1},$$

fulfill both (82) and (83), and are thus the unique solution of the convex optimization problem posed by Eqs. (19a) and (19b).[7]

[7] Note that $\mathbf{C}_{\mathbf{x}}^{(k)} = (\mathbf{C}_{\mathbf{x}}^{(k)})^\top$ (for $k = 1, \ldots, L$), since $\mathbf{C}_{\mathbf{x}}^{(k)}$ is a covariance matrix, and thus $[(\mathbf{C}_{\mathbf{x}}^{(k)})^{-1}]^\top = (\mathbf{C}_{\mathbf{x}}^{(k)})^{-1}$.

B. Multivariate Gaussians: Derivation of the Global Covariance Matrix

Let us concentrate on the sum in the exponential of Eq. (44):

$$\begin{aligned} S_1 &= \sum_{\ell=1}^{L} (\mathbf{x} - \hat{\mathbf{x}}_\ell)^\top \left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right)^{-1} (\mathbf{x} - \hat{\mathbf{x}}_\ell) \\ &= \mathbf{x}^\top \mathbf{C}_{\mathbf{x}}^{-1} \mathbf{x} - 2\mathbf{x}^\top \sum_{\ell=1}^{L} \left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right)^{-1} \hat{\mathbf{x}}_\ell + \sum_{\ell=1}^{L} \hat{\mathbf{x}}_\ell^\top \left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right)^{-1} \hat{\mathbf{x}}_\ell, \end{aligned} \tag{84}$$

where we have defined $\mathbf{C}_{\mathbf{x}}$ as in Eq. (46a):

$$\mathbf{C}_{\mathbf{x}} = \left[\sum_{\ell=1}^{L} \left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right)^{-1}\right]^{-1}.$$

Now, let us consider the following sum:

$$S_2 = (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^\top \mathbf{C}_{\mathbf{x}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}) = \mathbf{x}^\top \mathbf{C}_{\mathbf{x}}^{-1} \mathbf{x} - 2\mathbf{x}^\top \mathbf{C}_{\mathbf{x}}^{-1} \boldsymbol{\mu}_{\mathbf{x}} + \boldsymbol{\mu}_{\mathbf{x}}^\top \mathbf{C}_{\mathbf{x}}^{-1} \boldsymbol{\mu}_{\mathbf{x}}. \tag{85}$$

By defining the mean as in Eq. (46b),

$$\boldsymbol{\mu}_{\mathbf{x}} = \mathbf{C}_{\mathbf{x}} \sum_{\ell=1}^{L} \left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right)^{-1} \hat{\mathbf{x}}_\ell,$$

it is straightforward to see that

$$S_1 = S_2 + \sum_{\ell=1}^{L} \hat{\mathbf{x}}_\ell^\top \left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right)^{-1} \hat{\mathbf{x}}_\ell - \boldsymbol{\mu}_{\mathbf{x}}^\top \mathbf{C}_{\mathbf{x}}^{-1} \boldsymbol{\mu}_{\mathbf{x}}. \tag{86}$$

And finally, since the last two terms in (86) do not depend on $\mathbf{x}$, we have

$$p(\mathbf{x}|\hat{\mathbf{x}}_1, \ldots, \hat{\mathbf{x}}_L) \propto \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})^\top \mathbf{C}_{\mathbf{x}}^{-1} (\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})\right) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{C}_{\mathbf{x}}),$$

showing that the global posterior PDF in Eq. (44) can be expressed alternatively as the single Gaussian given in Eq. (45).

C. Optimal Independent and Isotropic Multivariate Gaussian Approximations

Let us define $p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\hat{\mathbf{x}}_\ell, \mathbf{\Sigma}_p)$ and $q(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\hat{\mathbf{x}}_\ell, \mathbf{\Sigma}_q)$. The Kullback-Leibler (KL) divergence, $D_{KL} \geq 0$, is a standard measure of the discrepancy between two distributions [47]: the larger the value of D_{KL} the more different the two distributions, with $D_{KL} = 0$ indicating that they are equal almost everywhere. The KL divergence between $p(\mathbf{x})$ and $q(\mathbf{x})$ is defined as

$$D_{KL}(p\|q) = \int_{\mathcal{X}} p(\mathbf{x}) \ln \frac{p(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x}. \tag{87}$$

Since both $p(\mathbf{x})$ and $q(\mathbf{x})$ are Gaussian PDFs, we have

$$\ln \frac{p(\mathbf{x})}{q(\mathbf{x})} = \frac{D}{2} \ln |\mathbf{\Sigma}_q| - \frac{D}{2} \ln |\mathbf{\Sigma}_p| - \frac{1}{2} (\mathbf{x} - \hat{\mathbf{x}}_\ell)^\top (\mathbf{\Sigma}_p^{-1} - \mathbf{\Sigma}_q^{-1})(\mathbf{x} - \hat{\mathbf{x}}_\ell), \tag{88}$$

and the KL divergence becomes

$$D_{KL}(p\|q) = K + \frac{D}{2} \ln |\mathbf{\Sigma}_q| + \frac{1}{2} \int_{\mathcal{X}} (\mathbf{x} - \hat{\mathbf{x}}_\ell)^\top \mathbf{\Sigma}_q^{-1} (\mathbf{x} - \hat{\mathbf{x}}_\ell) \mathcal{N}(\mathbf{x}|\hat{\mathbf{x}}_\ell, \mathbf{\Sigma}_p) d\mathbf{x}, \tag{89}$$

where K is a constant that includes all the terms that do not depend on $\mathbf{\Sigma}_q$. Now, let us consider the two particular cases that we are interested in: the approximation of an arbitrary multi-variate Gaussian PDF using an independent and an isotropic Gaussian PDF, respectively.

C.1. Isotropic Approximation

Let us consider $\mathbf{\Sigma}_p = \mathbf{C}_{\mathbf{x}}^{(\ell)}$ and $\mathbf{\Sigma}_q = \sigma_\ell^2 \mathbf{I}$, so that $|\mathbf{\Sigma}_q| = \sigma_\ell^{2D}$ and $\mathbf{\Sigma}_q^{-1} = \frac{1}{\sigma_\ell^2} \mathbf{I}$. Then, the KL divergence in (89) becomes

$$D_{KL}(p\|q) = K + \frac{D^2}{2} \ln(\sigma_\ell^2) + \frac{1}{2\sigma_\ell^2} \mathrm{Tr}\left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right). \tag{90}$$

In order to minimize the KL divergence in (90), we take the derivative w.r.t. σ_ℓ^2 and equate it to zero:

$$\frac{\partial D_{KL}(p\|q)}{\partial \sigma_\ell^2} = \frac{D^2}{2\sigma_\ell^2} - \frac{1}{2\sigma_\ell^4}\mathrm{Tr}\left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right) = 0. \tag{91}$$

The solution of this equation is

$$\sigma_\ell^2 = \frac{1}{D^2}\mathrm{Tr}\left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right), \tag{92}$$

and it can be easily checked (by taking the second derivative of $D_{KL}(p\|q)$) that it corresponds to a minimum. Hence, the best isotropic approximation of $p(\mathbf{x}|\hat{\mathbf{x}}_\ell) = \mathcal{N}(\mathbf{x}|\hat{\mathbf{x}}_\ell, \mathbf{C}_{\mathbf{x}}^{(\ell)})$ is $q(\mathbf{x}|\hat{\mathbf{x}}_\ell) = \mathcal{N}(\mathbf{x}|\hat{\mathbf{x}}_\ell, \sigma_\ell^2\mathbf{I})$, with σ_ℓ^2 given by Eq. (92). Using this approximation in the expression of the posterior of Eq. (44), it is straightforward to see that the weighting matrix for the linear fusion in this case is $\mathbf{\Lambda}_\ell = \alpha_\ell \mathbf{I}$, with

$$\alpha_\ell = \frac{\mathrm{Tr}\left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right)^{-1}}{\sum_{k=1}^{L}\mathrm{Tr}\left(\mathbf{C}_{\mathbf{x}}^{(k)}\right)^{-1}} == \frac{\sigma_\ell^{-2}}{\sum_{k=1}^{L}\sigma_k^{-2}}. \tag{93}$$

Note that Eq. (93) corresponds to the SC-MMSE fusion rule, obtained as a solution of a constrained optimization problem in Section 3.2 and given by (30), and also derived in Section 4.3 from a Bayesian perspective and given by (57).

C.2. Independent Approximation

Let us consider again $\mathbf{\Sigma}_p = \mathbf{C}_{\mathbf{x}}^{(\ell)}$, but now let us use $\mathbf{\Sigma}_q = \mathrm{diag}(\hat{\sigma}_{\ell,1}^2, \ldots, \hat{\sigma}_{\ell,D}^2)$. Then, $|\mathbf{\Sigma}_q| = \prod_{d=1}^{D}\hat{\sigma}_{\ell,d}^2$, $\ln|\mathbf{\Sigma}_q| = \sum_{d=1}^{D}\ln(\hat{\sigma}_{\ell,d}^2)$, and $\mathbf{\Sigma}_q^{-1} = \mathrm{diag}(\hat{\sigma}_{\ell,1}^{-2}, \ldots, \hat{\sigma}_{\ell,D}^{-2})$. In this scenario, the KL divergence in (89) becomes

$$D_{KL}(p\|q) = K + \frac{D}{2}\sum_{d=1}^{D}\ln(\hat{\sigma}_{\ell,d}^2) + \frac{1}{2}\sum_{d=1}^{D}\hat{\sigma}_{\ell,d}^{-2}\sigma_{\ell,d}^2, \tag{94}$$

where $\sigma^2_{\ell,d} = \mathbf{C}_{\mathbf{x}}^{(\ell)}[d,d]$. In order to minimize the KL divergence in (94), we take again the derivative w.r.t. $\hat{\sigma}^2_{\ell,d}$ and equate it to zero:

$$\frac{\partial D_{KL}(p\|q)}{\partial \hat{\sigma}^2_{\ell,d}} = \frac{D}{2\hat{\sigma}^2_{\ell,d}} - \frac{\sigma^2_{\ell,d}}{2\hat{\sigma}^4_{\ell,d}} = 0. \tag{95}$$

The solution of this equation is

$$\hat{\sigma}^2_{\ell,d} = \frac{1}{D}\sigma^2_{\ell,d}, \tag{96}$$

and it can be easily checked (by taking the second derivative of $D_{KL}(p\|q)$) that it corresponds to a minimum. Hence, the best independent approximation of $p(\mathbf{x}|\hat{\mathbf{x}}_\ell) = \mathcal{N}(\mathbf{x}|\hat{\mathbf{x}}_\ell, \mathbf{C}_{\mathbf{x}}^{(\ell)})$ is $q(\mathbf{x}|\hat{\mathbf{x}}_\ell) = \mathcal{N}(\mathbf{x}|\hat{\mathbf{x}}_\ell, \mathrm{diag}(\hat{\sigma}^2_{\ell,1}, \ldots, \hat{\sigma}^2_{\ell,D}))$, with $\hat{\sigma}^2_{\ell,d}$ given by Eq. (96). Using this approximation in the expression of the posterior of Eq. (44), it is straightforward to see that the weighting matrix for the linear fusion in this case is $\mathbf{\Lambda}_\ell = \mathbf{D}_\ell = \mathrm{diag}(\alpha_{\ell,1}, \ldots, \alpha_{\ell,D})$, with

$$\alpha_{\ell,d} = \frac{\sigma^{-2}_{\ell,d}}{\sum_{k=1}^{L} \sigma^{-2}_{k,d}}. \tag{97}$$

Note that Eq. (97) corresponds now to the IL-MMSE fusion rule, obtained as a solution of a constrained optimization problem in Section 3.3 and given by (38), and also derived in Section 4.3 from a Bayesian perspective and given by (55).