EFFICIENT LINEAR COMBINATION OF PARTIAL MONTE CARLO ESTIMATORS

David Luengo[⋆], *Luca Martino*[†], *Víctor Elvira*[‡], *Mónica Bugallo*[♯]

[⋆] Dep. of Signal Theory and Communic., Universidad Politécnica de Madrid, 28031 Madrid (Spain)
[†] Dep. of Mathematics and Statistics, University of Helsinki, 00014 Helsinki (Finland)
[‡] Dep. of Signal Theory and Communic., Universidad Carlos III de Madrid, 28911 Leganés (Spain)
[♯] Dep. of Electrical and Computer Eng., Stony Brook University, Stony Brook, NY 11794 (USA)

ABSTRACT

In many practical scenarios, including those dealing with large data sets, calculating global estimators of unknown variables of interest becomes unfeasible. A common solution is obtaining partial estimators and combining them to approximate the global one. In this paper, we focus on minimum mean squared error (MMSE) estimators, introducing two efficient linear schemes for the fusion of partial estimators. The proposed approaches are valid for any type of partial estimators, although in the simulated scenarios we concentrate on the combination of Monte Carlo estimators due to the nature of the problem addressed. Numerical results show the good performance of the novel fusion methods with only a fraction of the cost of the asymptotically optimal solution.

Index Terms— Global estimator; partial estimator; linear combination; fusion; Monte Carlo estimation.

1. INTRODUCTION

Estimation theory addresses the problem of inferring a set of unknown variables of interest given a collection of observable data [1, 2, 3]. Unfortunately, determining the *global estimator* of these parameters using all the available information is often unfeasible or impractical for many real-world scenarios. For example, in big data applications the amount of data at hand imposes computational and/or storage constraints that impede the global estimation process [4]. Also, large data sets pose a challenge for Monte Carlo estimators, since the posterior density tends to concentrate on a relatively small space as the number of data increases [5].

A possible alternative to *global estimation* reduces to dividing the available data into groups of manageable information and obtaining *partial estimators* of the unknowns. The objective is then to properly combine the partial estimators to achieve the performance of the global one. Fusion of estimates has been widely studied in many different areas. On the one hand, in wireless sensor networks the focus has been on distributed learning/estimation under communication constraints [6, 7] and the adaptation of methods developed for graphical models to distributed fusion [8]. Many different consensus, gossip or diffusion algorithms [9, 10, 11] have been developed, but they require a significant amount of communication that may constitute a burden in big data applications. On the other hand, a related field in the statistical literature is the combination of forecasts [12]. Indeed, the optimal linear combination for the single parameter case was already derived in [13, 14] and a Bayesian perspective was provided in [15]. However, there are two important differences with respect to the scenario addressed here: (1) each forecaster is assumed to have access to the whole data set; (2) computational complexity is not considered an issue in those cases. Finally, there is currently a great interest in parallel Bayesian computation using Monte Carlo methods [16], and a few communication-free parallel Markov chain Monte Carlo (MCMC) algorithms have been developed [17, 18, 19]. However, none of them addresses the potentially large dimension of the optimal combiners.

The main contribution of this work is the derivation of two novel efficient linear schemes for the fusion of *partial MMSE estimators*, which are independent from the methods used to obtain the partial estimates. The motivation comes from the asymptotically optimal linear combination, which involves the calculation of one weighting matrix per partial estimator and thus may be inaccurate and computationally demanding for large dimensional systems (both in number of unknowns and observations).[1] In order to reduce the computational complexity, we propose two linear approaches that require only a single weighting coefficient per partial estimator and one weighting coefficient per parameter and partial estimator respectively. We apply the proposed algorithms to the problem of target localization using measurements acquired by more than one sensor. Monte Carlo partial estimators are used to deal with the groups of measurements.

This work is supported by BBVA Foundation (MG-FIAR project); Spanish government (projects S2010/BMD-2422, TEC2013-41718-R, TEC2012-38800-C03-01, TEC2012-38883-C02-01, TEC2012-38058-C03-01); ERC grant 239784 and AoF grant 251170; and NSF's Award CCF-0953316.

[1] Note that the optimal linear combination requires as many weighting matrices (whose size depends on the number of unknowns) as partial estimators (whose number is related to the number of observations).

2. PROBLEM STATEMENT: GLOBAL VS. PARTIAL ESTIMATORS

In many applications, we are interested in inferring a variable of interest given a set of observations or measurements. Let us consider the variable of interest, $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^{D\times 1}$, and let $\mathbf{y} \in \mathcal{Y} \subseteq \mathbb{R}^{N\times 1}$ be the observed data. The posterior probability density function (PDF) (i.e., the conditional PDF of the variables of interest given the data) is $p(\mathbf{x}|\mathbf{y}) = \frac{1}{Z(\mathbf{y})}\pi(\mathbf{x},\mathbf{y}) \propto \pi(\mathbf{x},\mathbf{y})$, where $Z(\mathbf{y})$ is the model evidence (a.k.a. partition function) and $\pi(\mathbf{x},\mathbf{y})$ is the joint PDF of $\mathbf{x}$ and $\mathbf{y}$. A common approach for the estimation of $\mathbf{x}$ given $\mathbf{y}$ is trying to find an estimator, $\hat{\mathbf{x}} = \mathbf{f}(\mathbf{y})$, that minimizes the mean squared error (MSE). Mathematically, the minimum mean squared error (MMSE) estimator of $\mathbf{x}$ is obtained as

$$\hat{\mathbf{x}}^{(\text{MMSE})} = \arg\min_{\hat{\mathbf{x}}} \text{MSE}(\hat{\mathbf{x}}|\mathbf{y}), \tag{1}$$

where

$$\begin{aligned}\text{MSE}(\hat{\mathbf{x}}|\mathbf{y}) &= \mathbb{E}\left((\hat{\mathbf{x}}-\mathbf{x})^\top(\hat{\mathbf{x}}-\mathbf{x})\right)\\ &= \int_{\mathcal{X}} (\hat{\mathbf{x}}-\mathbf{x})^\top(\hat{\mathbf{x}}-\mathbf{x})p(\mathbf{x}|\mathbf{y})d\mathbf{x}.\end{aligned} \tag{2}$$

Note that solving (1) is equivalent to minimizing the Bayesian risk under a quadratic loss function. It is well-known that the MMSE estimator is given by the conditional mean [1, 2, 3]:

$$\hat{\mathbf{x}}^{(\text{MMSE})} = \mathbb{E}(\mathbf{x}|\mathbf{y}) = \int_{\mathcal{X}} \mathbf{x}\, p(\mathbf{x}|\mathbf{y})d\mathbf{x}. \tag{3}$$

Unfortunately, obtaining this *global estimator* is often unfeasible or impractical. A possible solution then consists of splitting the data into L groups/clusters, so that the ℓ-th cluster ($1 \leq \ell \leq L$) only has access to N_ℓ samples. In this situation we can obtain the partial MMSE estimator for each cluster (i.e., the MMSE estimator of $\mathbf{x}$ given all the data available to the ℓ-th estimator, $\mathbf{y}_\ell$) as

$$\hat{\mathbf{x}}_\ell^{(\text{MMSE})} = \arg\min_{\hat{\mathbf{x}}_\ell} \text{MSE}(\hat{\mathbf{x}}_\ell|\mathbf{y}_\ell), \tag{4}$$

where

$$\text{MSE}(\hat{\mathbf{x}}_\ell|\mathbf{y}_\ell) = \int_{\mathcal{X}} (\hat{\mathbf{x}}_\ell-\mathbf{x})^\top(\hat{\mathbf{x}}_\ell-\mathbf{x})p_\ell(\mathbf{x}|\mathbf{y}_\ell)d\mathbf{x}, \tag{5}$$

and $p_\ell(\mathbf{x}|\mathbf{y}_\ell)$ denotes the partial posterior induced by the ℓ-th subset of data, $\mathbf{y}_\ell$. Like the global MMSE estimator given by (3), the partial MMSE estimator corresponds to the conditional mean given the ℓ-th subset of data:

$$\hat{\mathbf{x}}_\ell^{(\text{MMSE})} = \mathbb{E}(\mathbf{x}|\mathbf{y}_\ell) = \int_{\mathcal{X}} \mathbf{x}\, p_\ell(\mathbf{x}|\mathbf{y}_\ell)d\mathbf{x}. \tag{6}$$

The objective is obtaining the *global* MMSE estimator from the set of *partial* MMSE estimators.

3. ASYMPTOTICALLY OPTIMAL COMBINATION OF PARTIAL ESTIMATORS

In general, the MMSE estimator is a non-linear function of the whole data set and the exact global MMSE estimator cannot be attained by any combination of partial MMSE estimators.[2] However, the Bernstein-von Mises (a.k.a. Bayesian central limit) theorem states that, under suitable regularity conditions, the partial posterior PDFs, $p_\ell(\mathbf{x}|\mathbf{y}_\ell)$, converge to Gaussian PDFs as N_ℓ tends to infinity [20, 21], i.e.,

$$p_\ell(\mathbf{x}|\mathbf{y}_\ell) \to \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{\mathbf{x}}^{(\ell)}, \mathbf{C}_{\mathbf{x}}^{(\ell)}) \qquad \text{as} \quad N_\ell \to \infty, \tag{7}$$

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

$$\begin{aligned}\mathbf{C}_{\mathbf{x}}^{(\ell)} &= \mathbb{E}\left((\hat{\mathbf{x}}_\ell^{(\text{MMSE})}-\mathbf{x})(\hat{\mathbf{x}}_\ell^{(\text{MMSE})}-\mathbf{x})^\top\right)\\ &= \int_{\mathcal{X}} (\hat{\mathbf{x}}_\ell^{(\text{MMSE})}-\mathbf{x})(\hat{\mathbf{x}}_\ell^{(\text{MMSE})}-\mathbf{x})^\top p_\ell(\mathbf{x}|\mathbf{y}_\ell)d\mathbf{x}.\end{aligned} \tag{8}$$

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.,

$$p(\mathbf{x}|\mathbf{y}) = \prod_{\ell=1}^L p_\ell(\mathbf{x}|\mathbf{y}_\ell) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_{\mathbf{x}}, \mathbf{C}_{\mathbf{x}}) \qquad \text{as} \quad N \to \infty, \tag{9}$$

with

$$\mathbf{C}_{\mathbf{x}} = \left[\sum_{\ell=1}^L \left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right)^{-1}\right]^{-1}, \tag{10a}$$

$$\boldsymbol{\mu}_{\mathbf{x}} = \mathbf{C}_{\mathbf{x}} \sum_{\ell=1}^L \left(\mathbf{C}_{\mathbf{x}}^{(\ell)}\right)^{-1} \hat{\mathbf{x}}_\ell^{(\text{MMSE})}. \tag{10b}$$

This result has been recently exploited in [18] to obtain asymptotically exact samples from the global posterior using a parallel MCMC algorithm.

4. EFFICIENT LINEAR COMBINATION OF PARTIAL ESTIMATORS

4.1. Asymptotically Optimal Combination

Let us consider a linear fusion approach, where the global estimator is obtained as a weighted linear combination of the partial MMSE estimators:

$$\hat{\mathbf{x}}^{(\text{LMSE})} = \sum_{\ell=1}^L \boldsymbol{\Lambda}_\ell \hat{\mathbf{x}}_\ell^{(\text{MMSE})}, \tag{11}$$

[2] An exception occurs when the global MMSE estimator is "separable in the data". For instance, this happens when the global posterior PDF is Gaussian with a mean that is a weighted linear combination of the data. In this case, a properly weighted linear combination of the partial MMSE estimators leads to the exact global MMSE estimator.

where $\mathbf{\Lambda}_\ell$ is a $D \times D$ matrix of weights. Noting that the MSE in (2) can be alternatively expressed as

$$\text{MSE}(\hat{\mathbf{x}}|\mathbf{y}) = \text{Tr}\left(\mathbb{E}\left((\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^\top\right)\right), \quad (12)$$

it is straightforward to show that the MSE of (11) is given by

$$\text{MSE}(\hat{\mathbf{x}}^{(\text{LMSE})}|\mathbf{y}) = \sum_{\ell=1}^{L} \text{Tr}\left(\mathbf{\Lambda}_\ell \mathbf{C}_\mathbf{x}^{(\ell)} \mathbf{\Lambda}_\ell^\top\right), \quad (13)$$

where $\text{Tr}(\mathbf{A})$ denotes the trace of matrix $\mathbf{A}$ and $\mathbf{C}_\mathbf{x}^{(\ell)}$ is given by (8). Given the L partial MMSE estimators, the best linear unbiased global estimator is obtained solving the following constrained optimization problem:

$$\mathbf{\Lambda}^* = \arg\min_{\mathbf{\Lambda}} \sum_{\ell=1}^{L} \text{Tr}\left(\mathbf{\Lambda}_\ell \mathbf{C}_\mathbf{x}^{(\ell)} \mathbf{\Lambda}_\ell^\top\right), \quad (14a)$$

$$\text{s.t.} \quad \sum_{\ell=1}^{L} \mathbf{\Lambda}_\ell = \mathbf{I}, \quad (14b)$$

where $\mathbf{\Lambda} = [\mathbf{\Lambda}_1, \ldots, \mathbf{\Lambda}_L]$, (14a) corresponds to a standard MSE minimization problem and (14b) is required to guarantee that the resulting global estimator is unbiased. Applying the method of Lagrange multipliers [22], we obtain the solution for each of the weight matrices as[3]

$$\mathbf{\Lambda}_\ell = \left[\sum_{k=1}^{L} \left(\mathbf{C}_\mathbf{x}^{(k)}\right)^{-1}\right]^{-1} \left(\mathbf{C}_\mathbf{x}^{(\ell)}\right)^{-1} = \mathbf{C}_\mathbf{x} \left(\mathbf{C}_\mathbf{x}^{(\ell)}\right)^{-1}. \quad (15)$$

Substituting (15) into (11), we note that the LMSE estimator is given exactly by (10b), i.e., $\hat{\mathbf{x}}^{(\text{LMSE})} = \boldsymbol{\mu}_\mathbf{x}$. Thus, the LMSE estimator is asymptotically optimal as $N \to \infty$.

4.2. Restricted Linear Combination

Unfortunately, the LMSE estimator described in the previous section requires obtaining a $D \times D$ weighting matrix for each of the L partial estimators. In practice, this implies estimating $D^2 L$ parameters overall. When D is large this can be problematic in terms of statistical accuracy (especially when N is not so large compared to $D^2 L$) and results in high computational and storage costs.

In order to reduce the number of parameters to be estimated, here we consider a restricted LMSE estimator, where a single coefficient per partial estimator is used to construct the global estimator. This single coefficient MSE (SCMSE) estimator is given by[4]

$$\hat{\mathbf{x}}^{(\text{SCMSE})} = \sum_{\ell=1}^{L} \alpha_\ell \hat{\mathbf{x}}_\ell^{(\text{MMSE})}, \quad (16)$$

where the coefficients α_ℓ are obtained solving the following constrained optimization problem:

$$\boldsymbol{\alpha}^* = \arg\min_{\boldsymbol{\alpha}} \sum_{\ell=1}^{L} \alpha_\ell^2 \text{Tr}\left(\mathbf{C}_\mathbf{x}^{(\ell)}\right), \quad (17a)$$

$$\text{s.t.} \quad \sum_{\ell=1}^{L} \alpha_\ell = 1, \quad (17b)$$

with $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_L]$. Using again the method of Lagrange multipliers, the closed-form solution for the ℓ-th weight ($1 \leq \ell \leq L$) is given by

$$\alpha_\ell = \frac{\left[\text{Tr}(\mathbf{C}_\mathbf{x}^{(\ell)})\right]^{-1}}{\sum_{k=1}^{L} \left[\text{Tr}(\mathbf{C}_\mathbf{x}^{(k)})\right]^{-1}} = \frac{\left[\text{MSE}(\hat{\mathbf{x}}_\ell^{(\text{MMSE})}|\mathbf{y}_\ell)\right]^{-1}}{\sum_{k=1}^{L} \left[\text{MSE}(\hat{\mathbf{x}}_k^{(\text{MMSE})}|\mathbf{y}_k)\right]^{-1}}, \quad (18)$$

where the last expression in (18) comes directly from (12).

The SCMSE estimator has a substantially reduced computational cost w.r.t. the LMSE estimator, since it only requires the estimation of L parameters overall, instead of the $D^2 L$ parameters of the LMSE estimator. However, noting that the optimal weights in (18) involve the trace of the partial covariance matrices, we also introduce an independent linear minimum mean squared estimator (ILMSE), where $\mathbf{\Lambda}_\ell = \text{diag}(\alpha_{\ell,1}, \ldots, \alpha_{\ell,D})$. This approach leads to an independent estimation of each of the D variables of interest:

$$\hat{x}_d^{(\text{ILMSE})} = \sum_{\ell=1}^{L} \alpha_{\ell,d}\, \hat{x}_{\ell,d}^{(\text{MMSE})}, \quad (19)$$

where $1 \leq d \leq D$ and $\hat{x}_{\ell,d}^{(\text{MMSE})}$ denotes the d-th component of the ℓ-th partial MMSE estimator. In practice, the weights in (19) can be obtained by solving D single parameter constrained optimization problems:

$$\boldsymbol{\alpha}_d^* = \arg\min_{\boldsymbol{\alpha}_d} \sum_{\ell=1}^{L} \alpha_{\ell,d}^2 C_{x_d}^{(\ell)}, \quad (20a)$$

$$\text{s.t.} \quad \sum_{\ell=1}^{L} \alpha_{\ell,d} = 1, \quad (20b)$$

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

$$\alpha_{\ell,d} = \frac{\left[\text{MSE}(\hat{x}_{\ell,d}^{(\text{MMSE})}|\mathbf{y}_\ell)\right]^{-1}}{\sum_{k=1}^{L} \left[\text{MSE}(\hat{x}_{k,d}^{(\text{MMSE})}|\mathbf{y}_k)\right]^{-1}}. \quad (21)$$

This approach requires the estimation of DL parameters overall, and thus it can be seen as an intermediate approach between the LMSE and the SCMSE.

[3] See our technical report for a detailed derivation of the coefficients in Sections 4.1 and 4.2 [23].

[4] Note that the SCMSE estimator can be obtained by setting $\mathbf{\Lambda}_\ell = \alpha_\ell \mathbf{I}$ in (11), with $\mathbf{I}$ denoting a $D \times D$ identity matrix.

Experiment		N_ℓ								
Scenario	**Estimator**	6	12	30	60	240	600	1200	3000	6000
Sc1	EWF	0.0041	0.0049	0.0065	0.0090	0.0167	0.0590	0.1192	0.2899	0.5540
	SCMSE	0.0039	0.0046	0.0063	0.0089	0.0166	0.0587	0.1191	0.2899	
	ILMSE	0.0038	0.0046	0.0063	0.0089	0.0166	0.0586	0.1188	0.2886	
	LMSE	0.0037	0.0045	0.0062	0.0088	0.0165	0.0584	0.1183	0.2878	
Sc2	EWF	0.0087	0.0053	0.0064	0.0104	0.0343	0.0648	0.1681	0.3392	0.5540
	SCMSE	0.0057	0.0034	0.0047	0.0092	0.0328	0.0628	0.1623	0.3290	
	ILMSE	0.0052	0.0031	0.0043	0.0085	0.0304	0.0588	0.1521	0.3159	
	LMSE	0.0037	0.0021	0.0028	0.0057	0.0210	0.0410	0.1107	0.2406	
Sc3	EWF	0.0078	0.0061	0.0068	0.0092	0.0169	0.0587	0.1181	0.2877	0.5540
	SCMSE	0.0060	0.0053	0.0066	0.0091	0.0168	0.0584	0.1180	0.2877	
	ILMSE	0.0055	0.0051	0.0065	0.0090	0.0168	0.0583	0.1177	0.2867	
	LMSE	0.0051	0.0048	0.0064	0.0090	0.0167	0.0582	0.1174	0.2861	

Table 1. MSE (averaged over 50 independent runs) for the three scenarios and the four fusion methods considered.

5. NUMERICAL EXPERIMENTS

We address the problem of positioning a target in the two-dimensional space of a wireless sensor network with only range measurements [24]. 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 $\mathbf{h}_1 = [1, -8]^\top$, $\mathbf{h}_2 = [8, 10]^\top$, $\mathbf{h}_3 = [-15, -7]^\top$, $\mathbf{h}_4 = [-8, 1]^\top$, $\mathbf{h}_5 = [10, 0]^\top$ and $\mathbf{h}_6 = [0, 10]^\top$. The measurement equations are given by

$$Y_j = -20 \log \left(||\mathbf{x} - \mathbf{h}_j||^2 \right) + \Theta_j, \quad j = 1, \ldots, 6, \qquad (22)$$

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), setting $x_1 = x_2 = 3.5$. We consider a varying number of partial estimators L, with $N_\ell = N/L$ observations per estimator 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 Section 4. The results, shown in Table 1 and Fig. 1, confirm the good performance of the SCME and ILMSE estimators, which outperform the naive EWF and show an MSE similar to the optimal and more costly LMSE. Note that the poor performance of all the estimators for small values of L is due to the slower convergence of the parallel chains when the number of data in the posterior is large (e.g., for $T_C^{(\ell)} = 20000$ the MSE decreases to 0.1624 when $L = 1$). This shows the importance of splitting the data even when a single estimator is able to deal with them.

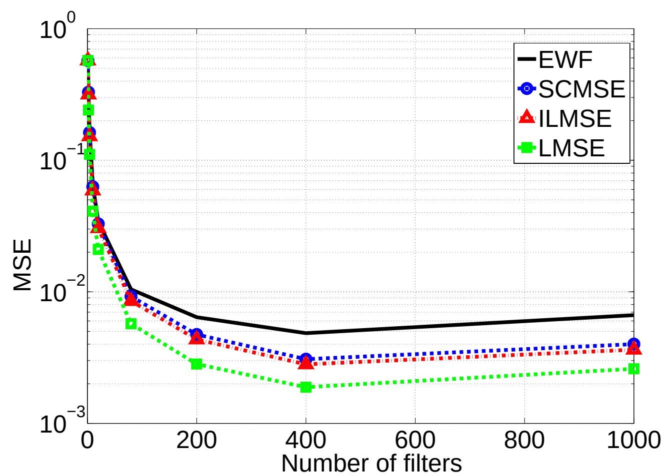

Fig. 1. MSE as a function of L for scenario 2 (Sc2).

6. CONCLUSIONS AND FUTURE LINES

In this paper we have addressed the fusion of partial minimum mean squared error (MMSE) estimators using two novel efficient linear combination schemes. The methods were tested through computer simulations by applying them to a localization problem with one target and six sensors whose measurements were processed using several parallel filters. The new fusion methods show a similar performance to the optimal linear combination with a reduced computational cost.

7. REFERENCES

[1] Harry L. Van Trees, *Detection, Estimation and Modulation Theory*, John Wiley and Sons, 1968.

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

[3] Steven M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, 1993.

[4] Georgios B. Giannakis, Francis Bach, Raphael Cendrillon, Michael Mahoney, and Jennifer Neville, "Signal processing for big data (special issue)," *IEEE Signal Processing Magazine*, vol. 31, no. 5, pp. 15–111, 2014.

[5] Nicolas Chopin, "A sequential particle filter method for static models," *Biometrika*, vol. 89, no. 3, pp. 539–552, 2002.

[6] Jin-Jun Xiao, Alejandro Ribeiro, Zhi-Quan Luo, and Georgios B. Giannakis, "Distributed compression-estimation using wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 27–41, 2006.

[7] Joel B. Predd, Sanjeev R. Kulkarni, and H. Vincent Poor, "Distributed learning in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 56–69, 2006.

[8] Mujdat Cetin, Lei Chen, John W. Fisher III, Alexander T. Ihler, Randolph L. Moses, Martin J. Wainwright, and Alan S. Willsky, "Distributed fusion in sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 42–55, 2006.

[9] Reza Olfati-Saber, J. Alex Fax, and Richard M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 215–233, 2007.

[10] Alexandros G. Dimakis, Soummya Kar, José F. Moura, Michael G. Rabbat, and Anna Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of the IEEE*, vol. 98, no. 11, pp. 1847–1864, 2010.

[11] Federico S. Cattivelli and Ali H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1035–1048, 2010.

[12] Kenneth F. Wallis, "Combining forecasts – forty years later," *Applied Financial Economics*, vol. 21, no. 1–2, pp. 33–41, 2011.

[13] J. M. Bates and Clive W. Granger, "The combination of forecasts," *Operational Research Quarterly*, vol. 20, no. 4, pp. 451–468, 1969.

[14] J. P. Dickinson, "Some statistical results in the combination of forecasts," *Operational Research Quarterly*, vol. 24, pp. 253–260, 1975.

[15] Robert F. Bordley, "The combination of forecasts: A Bayesian approach," *Journal of the Operational Research Society*, vol. 33, no. 2, pp. 171–174, 1982.

[16] Darren J. Wilkinson, "Parallel Bayesian computation," *Statistics Textbooks and Monographs*, vol. 184, 2006.

[17] Steven L. Scott, Alexander W. Blocker, Fernando V. Bonassi, Hugh A. Chipman, Edward I. George, and Robert E. McCulloch, "Bayes and big data: The consensus Monte Carlo algorithm," in *EFaBBayes 250th conference*, 2013, vol. 16.

[18] Willie Neiswanger, Chong Wang, and Eric Xing, "Asymptotically exact, embarrassingly parallel MCMC," *arXiv:1311.4780v2*, pp. 1–16, 21 Mar. 2014.

[19] Xiangyu Wang and David B. Dunson, "Parallelizing MCMC via Weierstrass sampler," *arXiv:1312.4605v2*, 25 May 2014.

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

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

[22] Dimitri P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, 1982.

[23] David Luengo, Luca Martino, Víctor Elvira, and Mónica Bugallo, "Efficient combination of partial Monte Carlo estimators," *viXra*, Oct. 2014.

[24] Luca Martino and Joquín Míguez, "Generalized rejection sampling schemes and applications in signal processing," *Signal Processing*, vol. 90, no. 11, pp. 2981–2995, November 2010.