www.ietdl.org

Published in IET Signal Processing
Received on 23rd May 2011
Revised on 28th April 2012
doi: 10.1049/iet-spr.2011.0199

ISSN 1751-9675

Linear coherent distributed estimation with cluster-based sensor networks

C.-A. Lin C.-H. Wu

Department of Electrical and Control Engineering, National Chiao Tung University, Hsinchu, Taiwan
E-mail: chwu.ece92g@nctu.edu.tw

Abstract: The authors consider distributed estimation using sensor network with coherent multiple access channel model and LMMSE fusion rule. The sensors in the network are divided into a number of clusters. Sensors within the same cluster are allowed to collaborate through an amplification matrix to form a message this then transmitted. They formulate the problem of choosing the amplification matrices as an optimal power allocation problem under a total power constraint. The solution gives the optimal amplification matrices as scaled outer products of the observation gain and the channel gain vectors. The authors show that collaboration improves performance and, in simulations, demonstrate that the amount of improvement is closely related to the amount of collaboration.

1 Introduction

Distributed estimation has attracted much attention in signal processing research for sensor networks [1]. In distributed estimation scenario, a certain parameter is measured by spatially distributed sensors and the measurements are sent to a fusion centre (FC) where a final estimate is formed. Owing to energy constraints, power efficiency is an important issue since it is closely related to the network lifetime. To enhance power efficiency, many research works focus on cluster-based sensor networks in which the problem is to efficiently organise sensors into clusters so that network lifetime can be improved [2, 3]. For example, Wimalajeewa and Jayaweera [4] introduced sensor selection schemes to minimise the estimation distortion, whereas Heinzelman *et al.* [5] developed a communication protocol to save power. Recently, analogue transmission schemes aiming at minimising the estimation distortion by optimally allocating power for each sensor have been studied based on the coherent multiple access channel (MAC) model [6–9], the orthogonal MAC model [10–15], as well as the hybrid MAC model [16]. Among them, some works consider distributed estimation of a scalar parameter [13, 14] or a vector parameter [9, 10] with spatially correlated sensor observations. The work in [15] addresses robust estimation that takes account of the uncertainty in the local observing noise variance. Fang and Li [12] considered a cluster-based network architecture in which closely located sensors are able to collaborate to form local messages for transmission through the orthogonal MAC.

In this paper, we consider distributed estimation of a scalar parameter by optimally allocating power based on cluster-based wireless sensor networks with the coherent MAC model. We assume that the sensors in the network are already divided into a number of clusters. The sensors in the same cluster are allowed to collaborate, while collaboration is prohibited for sensors in different clusters. The collaboration is through an amplification matrix, for each cluster, that forms a message from the measurements for transmission to the FC. In other words, collaboration means the measurements within a cluster are linearly combined locally. At the FC, the parameter is estimated based on the linear minimum mean-squared error (LMMSE) rule. The mean-squared error (MSE) depends on the choice of the amplification matrices. We study the problem of choosing the amplification matrices so that the corresponding MSE is minimised. We formulate the problem as one of optimal power allocation under a total power constraint. The solution shows that the optimal amplification matrices are scaled outer products of the observation gain and the channel gain vectors. For comparison, two special cases are also considered: the full collaboration case, in which all sensors are in the same cluster, and the non-collaboration case, in which each cluster has only one sensor. We show that with the optimal amplification matrices, collaboration indeed improves performance in terms of MSE. We demonstrate through simulation results that the amount of improvement is closely related to the amount of collaboration.

The rest of this paper is organised as follows. In Section 2, we describe the model of cluster-based sensor network and the problem we address. In Section 3, we solve the optimisation problem to obtain the optimal amplification matrices and show that with optimal amplification matrices, collaboration indeed improves performance. In Section 4, simulation results are given to verify the analytical result. Section 5 is a brief conclusion.

2 System model and problem formulation

We consider a wireless sensor network consisting of K spatially deployed sensors for estimating a random source signal θ. The sensors in the network are divided into L clusters, as shown in Fig. 1. The lth cluster has K_l sensors and the measurement at the kth sensor is given by

$$x_{l,k} = f_{l,k}\theta + n_{l,k}, \quad 1 \le l \le L, 1 \le k \le K_l \tag{1}$$

where $f_{l,k}$ is the observation gain and $n_{l,k}$ is the measurement noise. In vector form, (1) becomes

$$\boldsymbol{x}_l = \boldsymbol{f}_l\theta + \boldsymbol{n}_l, \quad 1 \le l \le L \tag{2}$$

where $\boldsymbol{x}_l = [x_{l,1} \cdots x_{l,K_l}]^{\mathrm{T}}, \boldsymbol{f}_l = [f_{l,1} \cdots f_{l,K_l}]^{\mathrm{T}}$ and $\boldsymbol{n}_l = [n_{l,1} \cdots n_{l,K_l}]^{\mathrm{T}}$. The collaboration between sensors in the lth cluster is through an amplification matrix $\boldsymbol{A}_l \in \mathbb{R}^{N_l \times K_l}$, which takes $\boldsymbol{x}_l \in \mathbb{R}^{K_l}$ to form the message vector $\boldsymbol{A}_l\boldsymbol{x}_l \in \mathbb{R}^{N_l}$. The messages are then sent to the FC and the signal y received at the FC can be expressed as

$$y = \sum_{l=1}^{L} \boldsymbol{g}_l^{\mathrm{T}}\boldsymbol{A}_l(\boldsymbol{f}_l\theta + \boldsymbol{n}_l) + \nu \tag{3}$$

where $\boldsymbol{g}_l = [g_{l,1} \cdots g_{l,N_l}]^{\mathrm{T}}$ is the channel gain vector and ν is the additive noise at the receiver. In practice, the sensors which are geographically closely located can compose a cluster. The collaboration between sensors in the same cluster can be implemented by choosing one sensor as the cluster head whose task is to collect and process information sent from other sensors to form a message vector and transmit it to the FC.

In this paper, we assume that (i) $E[\theta] = 0$ and $E[\theta^2] = \sigma_\theta^2$, (ii) the measurement noises are zero-mean and mutually uncorrelated, specifically $E[\boldsymbol{n}_l] = 0$, $E[\boldsymbol{n}_l\boldsymbol{n}_l^{\mathrm{T}}] = \sigma_n^2\boldsymbol{I}_{K_l}$ and $E[\boldsymbol{n}_l\boldsymbol{n}_m^{\mathrm{T}}] = 0_{K_l \times K_m}$ for $l \ne m$, (iii) $E[\nu] = 0$ and $E[\nu^2] = \sigma_\nu^2$, (iv) the source signal, the measurement noises, and the receiver noise are uncorrelated, that is, $E[\theta\boldsymbol{n}_l] = 0$, $E[\theta\nu] = 0$ and $E[\nu\boldsymbol{n}_l] = 0$, and (v) the observation gain vectors $\boldsymbol{f}_l$ and the channel gain vectors $\boldsymbol{g}_l$ are known to the FC.

For a given set of amplification matrices $\boldsymbol{A}_l$, the LMMSE estimate of θ using the received signal y in (3) is [17, p. 382]

$$\begin{aligned}\hat{\theta} &= \frac{E[\theta y]}{E[y^2]}y \\ &= \frac{\sigma_\theta^2\sum_{l=1}^{L}\boldsymbol{g}_l^{\mathrm{T}}\boldsymbol{A}_l\boldsymbol{f}_l}{\sigma_\theta^2\left(\sum_{l=1}^{L}\boldsymbol{g}_l^{\mathrm{T}}\boldsymbol{A}_l\boldsymbol{f}_l\right)^2 + \sigma_n^2\sum_{l=1}^{L}\boldsymbol{g}_l^{\mathrm{T}}\boldsymbol{A}_l\boldsymbol{A}_l^{\mathrm{T}}\boldsymbol{g}_l + \sigma_\nu^2}y\end{aligned} \tag{4}$$

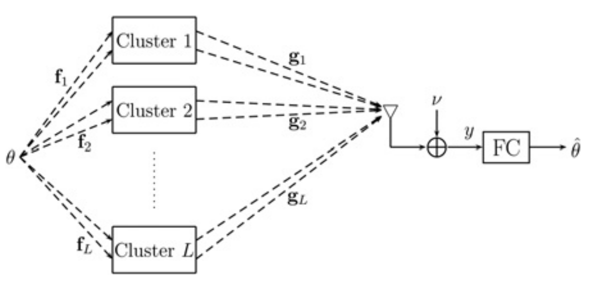

Fig. 1 *Cluster-based sensor network with coherent MAC*

and the corresponding MSE is

$$\begin{aligned}J = E[(\theta - \hat{\theta})^2] &= \sigma_\theta^2 - \frac{(E[\theta y])^2}{E[y^2]} \\ &= \left(\frac{1}{\sigma_\theta^2} + \frac{\left(\sum_{l=1}^{L}\boldsymbol{g}_l^{\mathrm{T}}\boldsymbol{A}_l\boldsymbol{f}_l\right)^2}{\sigma_n^2\sum_{l=1}^{L}\boldsymbol{g}_l^{\mathrm{T}}\boldsymbol{A}_l\boldsymbol{A}_l^{\mathrm{T}}\boldsymbol{g}_l + \sigma_\nu^2}\right)^{-1}\end{aligned} \tag{5}$$

The problem is to minimise the MSE in (5) by choosing optimal amplification matrices $\boldsymbol{A}_l$ under a total power constraint. The total transmitted power of the L clusters is $\sum_{l=1}^{L} E[\boldsymbol{x}_l^{\mathrm{T}}\boldsymbol{A}_l^{\mathrm{T}}\boldsymbol{A}_l\boldsymbol{x}_l]$. Hence if P is the amount of power that the clusters together can used, then we have the following constraint

$$\sum_{l=1}^{L}\mathrm{tr}(E[\boldsymbol{A}_l\boldsymbol{x}_l\boldsymbol{x}_l^{\mathrm{T}}\boldsymbol{A}_l^{\mathrm{T}}]) = \sum_{l=1}^{L}\mathrm{tr}(\sigma_\theta^2\boldsymbol{A}_l\boldsymbol{f}_l\boldsymbol{f}_l^{\mathrm{T}}\boldsymbol{A}_l^{\mathrm{T}} + \sigma_n^2\boldsymbol{A}_l\boldsymbol{A}_l^{\mathrm{T}}) \le P \tag{6}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix and we use $E[\boldsymbol{x}_l\boldsymbol{x}_l^{\mathrm{T}}] = \sigma_\theta^2\boldsymbol{f}\boldsymbol{f}^{\mathrm{T}} + \sigma_n^2\boldsymbol{I}_{K_l}$. From (5) and (6), the optimisation problem under consideration can be written as

$$\begin{aligned}&\min_{\boldsymbol{A}_l, 1\le l \le L} J \\ &\text{subject to} \quad \sum_{l=1}^{L}\mathrm{tr}(\sigma_\theta^2\boldsymbol{A}_l\boldsymbol{f}_l\boldsymbol{f}_l^{\mathrm{T}}\boldsymbol{A}_l^{\mathrm{T}} + \sigma_n^2\boldsymbol{A}_l\boldsymbol{A}_l^{\mathrm{T}}) \le P\end{aligned} \tag{7}$$

where J is given in (5).

Remarks: We had assumed that the measurement noises are mutually uncorrelated across all sensors. If the measurement noises are correlated within the same cluster but uncorrelated across different clusters, the problem can still be formulated in the same form as (7). To see this, suppose $E[\boldsymbol{n}_l\boldsymbol{n}_l^{\mathrm{T}}] = \boldsymbol{R}_{n_l}$, where $\boldsymbol{R}_{n_l} = \boldsymbol{R}_{n_l}^{\mathrm{T}} \in \mathbb{R}^{K_l \times K_l}$ is positive definite and $E[\boldsymbol{n}_l\boldsymbol{n}_m^{\mathrm{T}}] = 0_{K_l \times K_m}$ for $l \ne m$. Let $\boldsymbol{R}_{n_l} = \boldsymbol{U}_{n_l}\Lambda_{n_l}\boldsymbol{U}_{n_l}^{\mathrm{T}}$ be the eigenvalue decomposition with $\Lambda_{n_l} = \mathrm{diag}(\sigma_{n_{l,1}}^2, \ldots, \sigma_{n_{l,K_l}}^2) > 0$, where $\mathrm{diag}(x_1, \ldots, x_M)$ is a diagonal matrix whose mth diagonal element is x_m. By setting $\tilde{\boldsymbol{A}}_l = \boldsymbol{A}_l\boldsymbol{U}_{n_l}\Lambda_{n_l}^{1/2}$ and $\tilde{\boldsymbol{f}}_l = \Lambda_{n_l}^{-1/2}\boldsymbol{U}_{n_l}^{\mathrm{T}}\boldsymbol{f}_l$, the corresponding optimisation problem has the same form as (7) with $\boldsymbol{A}_l$, $\boldsymbol{f}_l$ and σ_n^2 replaced by $\tilde{\boldsymbol{A}}_l$, $\tilde{\boldsymbol{f}}_l$ and 1, respectively.

3 Optimal amplification matrices

In this section, we consider the solution of the optimisation problem (7) with the goal of obtaining a closed form expression for the optimal amplification matrices $\boldsymbol{A}_l$. We first make the following observations:

1. In problem (7), if the '$\le$' sign in the constraint is replaced by the '$=$' sign, the solution does not change. Hence we could consider the optimisation problem with equality constraint. The argument is as follows. Since the constraint function is quadratic in the elements of $\boldsymbol{A}_l$, if a set of $\boldsymbol{A}_l$ is such that strict inequality holds, we can equally scale up each $\boldsymbol{A}_l$ so that equality holds. In addition, if we equally scale up each $\boldsymbol{A}_l$, we obtain a lower function value of J because in (5) the second term inside the parentheses becomes larger. Consequently, with optimal $\boldsymbol{A}_l$, the inequality constraint must be active.

2. Consider the optimal MSE in (7), say, J^* as a function of the power P, then J^* is a strictly decreasing function of P, that is, if $P_2 > P_1$, then $J^*(P_2) < J^*(P_1)$. The argument is similar: if the power level increases, we can equally scale up $\boldsymbol{A}_l$ to obtain a lower value of J and thus a lower value of optimal MSE J^* can be obtained.
3. Since the function $J^*(P)$ is $1-1$ and decreasing, the inverse function $P(J^*)$ is also $1-1$ and decreasing. Hence instead of finding the matrices $\boldsymbol{A}_l$ that minimise J in (5) under an equality constraint on power level, we can find the matrices $\boldsymbol{A}_l$ that minimise the power level subject to an equality constraint on MSE. If the constraint value on MSE is such that the resulting minimum power level matches the given value P in (7), the corresponding matrices $\boldsymbol{A}_l$ are the optimal ones we set out to find. We thus consider the following optimisation problem

$$\begin{cases} \min\limits_{\boldsymbol{A}_l, 1\le l\le L} & \sum\limits_{l=1}^{L} \mathrm{tr}(\sigma_\theta^2 \boldsymbol{A}_l \boldsymbol{f}_l \boldsymbol{f}_l^{\mathrm{T}} \boldsymbol{A}_l^{\mathrm{T}} + \sigma_n^2 \boldsymbol{A}_l \boldsymbol{A}_l^{\mathrm{T}}) \\ \text{subject to} & \left(\dfrac{1}{\sigma_\theta^2} + \dfrac{\left(\sum_{l=1}^{L} \boldsymbol{g}_l^{\mathrm{T}} \boldsymbol{A}_l \boldsymbol{f}_l\right)^2}{\sigma_n^2 \sum_{l=1}^{L} \boldsymbol{g}_l^{\mathrm{T}} \boldsymbol{A}_l \boldsymbol{A}_l^{\mathrm{T}} \boldsymbol{g}_l + \sigma_\nu^2} \right)^{-1} = J^* \end{cases} \tag{8}$$

where $0 < J^* \le \sigma_\theta^2$. We note that both the objective function and the constraint function in (8) are quadratic in the elements of $\boldsymbol{A}_l$. This problem is considerably easier to solve than the original one (7). The main result based on solving (8) is in the following proposition whose proof is given in Appendix 1.

Proposition 1: Consider the sensor network model described by (2) and (3). Suppose the total transmitted power from all sensors is no greater than P, then using the LMMSE estimator, the optimal amplification matrix of the lth cluster is given by

$$\boldsymbol{A}_l^{\mathrm{opt}} = \sqrt{\left(\sum_{i=1}^{L} \frac{\|\boldsymbol{f}_i\|^2 \|\boldsymbol{g}_i\|^2 (\sigma_\theta^2 \|\boldsymbol{f}_i\|^2 + \sigma_n^2)}{\phi_i^2} \right)^{-1} \frac{P}{\phi_l^2}}\, \boldsymbol{g}_l \boldsymbol{f}_l^{\mathrm{T}}, \quad l = 1, \ldots, L \tag{9}$$

where $\phi_i = \sigma_\nu^2(\sigma_\theta^2 \|\boldsymbol{f}_i\|^2 + \sigma_n^2) + \sigma_n^2 \|\boldsymbol{g}_i\|^2 P$ and $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}}$, and the corresponding minimum MSE is

$$J_{\mathrm{M}} = \left(\frac{1}{\sigma_\theta^2} + \sum_{l=1}^{L} \frac{\|\boldsymbol{f}_l\|^2 \|\boldsymbol{g}_l\|^2 P}{\sigma_\nu^2(\sigma_\theta^2 \|\boldsymbol{f}_l\|^2 + \sigma_n^2) + \sigma_n^2 \|\boldsymbol{g}_l\|^2 P} \right)^{-1} \tag{10}$$

The optimal amplification matrix $\boldsymbol{A}_l^{\mathrm{opt}}$ is a rank one matrix, which is a scaled outer product of $\boldsymbol{g}_l$ and $\boldsymbol{f}_l$. As expected the optimal MSE J_{M} decreases as P increases. Moreover, as $P \to \infty$, we have

$$\lim_{P\to\infty} J_{\mathrm{M}} = \frac{\sigma_\theta^2}{1 + (\sigma_\theta^2/\sigma_n^2)\sum_{l=1}^{L} \|\boldsymbol{f}_l\|^2} \tag{11}$$

The limit does not go to zero but approaches a finite value which depends on the signal-to-noise ratio $\sigma_\theta^2 \sum_{l=1}^{L} \|\boldsymbol{f}_l\|^2/\sigma_n^2$, since the measured signal $\boldsymbol{f}_l\theta + \boldsymbol{n}_l$ is amplified by $\boldsymbol{A}_l$, $1 \le l \le L$.

For comparison, we consider two special cases: $L = 1$ and $L = K$. When $L = 1$, there is full collaboration among the K sensors. The observation gain is $\boldsymbol{f} \in \mathbb{R}^K$ and the channel gain is $\boldsymbol{g} \in \mathbb{R}^N$, $N \le K$. With the optimal amplification matrix $\boldsymbol{A}^{\mathrm{opt}} \in \mathbb{R}^{N\times K}$ given by (9), the minimum MSE in (10) becomes

$$J_{\mathrm{C}} = \left(\frac{1}{\sigma_\theta^2} + \frac{\|\boldsymbol{f}\|^2 \|\boldsymbol{g}\|^2 P}{\sigma_\nu^2(\sigma_\theta^2 \|\boldsymbol{f}\|^2 + \sigma_n^2) + \sigma_n^2 \|\boldsymbol{g}\|^2 P} \right)^{-1} \tag{12}$$

When $L = K$, each sensor is a cluster and no collaboration between sensor exists. The scalar observation gains and channel gains are respectively f_k and g_k, $1 \le k \le K$. With the K scalar amplification gains given by (9), the minimum MSE becomes

$$J_{\mathrm{N}} = \left(\frac{1}{\sigma_\theta^2} + \sum_{k=1}^{K} \frac{f_k^2\, g_k^2 P}{\sigma_\nu^2(\sigma_\theta^2 f_k^2 + \sigma_n^2) + \sigma_n^2\, g_k^2 P} \right)^{-1} \tag{13}$$

To compare the performance of the general case and the two special cases, we assume $N_l = K_l$, $1 \le l \le L$, in (10) and $N = K$ in (12), that is, the number of measurements is equal to the number of transmitters in each cluster. Hence the observation gain and the channel gain vectors can be written as $\boldsymbol{f} = [\boldsymbol{f}_1^{\mathrm{T}} \boldsymbol{f}_2^{\mathrm{T}} \cdots \boldsymbol{f}_L^{\mathrm{T}}]^{\mathrm{T}} = [f_1 f_2 \cdots f_K]^{\mathrm{T}}$ and $\boldsymbol{g} = [\boldsymbol{g}_1^{\mathrm{T}} \boldsymbol{g}_2^{\mathrm{T}} \cdots \boldsymbol{g}_L^{\mathrm{T}}]^{\mathrm{T}} = [g_1\, g_2 \cdots g_K]^{\mathrm{T}}$, respectively, where $\boldsymbol{f}_l, \boldsymbol{g}_l \in \mathbb{R}^{K_l}$ and $K_1 + \cdots + K_L = K$. In terms of the MSE, it is not unexpected that collaboration improves performance. Indeed, we have the following proposition.

Proposition 2: The minimum MSEs J_{M} in (10), J_{C} in (12), and J_{N} in (13) satisfy

$$J_{\mathrm{C}} \le J_{\mathrm{M}} \le J_{\mathrm{N}} \tag{14}$$

The proof of Proposition 2 is based on the following lemma.

Lemma 1: For $\boldsymbol{x} = [\boldsymbol{x}_1^{\mathrm{T}} \boldsymbol{x}_2^{\mathrm{T}} \cdots \boldsymbol{x}_L^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^n$ and $\boldsymbol{y} = [\boldsymbol{y}_1^{\mathrm{T}} \boldsymbol{y}_2^{\mathrm{T}} \cdots \boldsymbol{y}_L^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^n$, where $\boldsymbol{x}_i$ and $\boldsymbol{y}_i$ are non-zero vectors of dimension ≥ 1, we have

$$\frac{\|\boldsymbol{x}\|^2 \|\boldsymbol{y}\|^2}{\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2} \ge \sum_{i=1}^{L} \frac{\|\boldsymbol{x}_i\|^2 \|\boldsymbol{y}_i\|^2}{\|\boldsymbol{x}_i\|^2 + \|\boldsymbol{y}_i\|^2} \tag{15}$$

Proof: Please see Appendix 2.

We now establish Proposition 2. Let $\boldsymbol{x} = \sqrt{\sigma_\theta^2}\,\boldsymbol{f}$, $\boldsymbol{y} = \sqrt{(\sigma_n^2/\sigma_\nu^2)P}\,\boldsymbol{g}$, $\boldsymbol{x}_i = \sqrt{\sigma_\theta^2}\,\boldsymbol{f}_i$ and $\boldsymbol{y}_i = \sqrt{(\sigma_n^2/\sigma_\nu^2)P}\,\boldsymbol{g}_i$. Then by Lemma 1, we obtain

$$\begin{aligned} &\frac{\|\boldsymbol{f}\|^2 \|\boldsymbol{g}\|^2}{\sigma_\theta^2 \|\boldsymbol{f}\|^2 + (\sigma_n^2/\sigma_\nu^2)P\|\boldsymbol{g}\|^2 + \sigma_n^2} \\ &\ge \sum_{i=1}^{L} \frac{\|\boldsymbol{f}_i\|^2 \|\boldsymbol{g}_i\|^2}{\sigma_\theta^2 \|\boldsymbol{f}_i\|^2 + (\sigma_n^2/\sigma_\nu^2)P\|\boldsymbol{g}_i\|^2 + \sigma_n^2} \end{aligned}$$

and thus $J_{\mathrm{C}}^{-1} \ge J_{\mathrm{M}}^{-1}$, or equivalently, $J_{\mathrm{C}} \le J_{\mathrm{M}}$. The second inequality in (14) follows similarly: apply Lemma 1 to each $\boldsymbol{x}_i$, $\boldsymbol{y}_i$, and their respective scalar components, and their sums give the desired inequality.

4 Numerical results

In this section, we use numerical simulations to verify the analytical result established in Section 3. In all simulations, the random parameters, θ, $n_{l,\ k}$ and ν, are zero-mean

doi: 10.1049/iet-spr.2011.0199

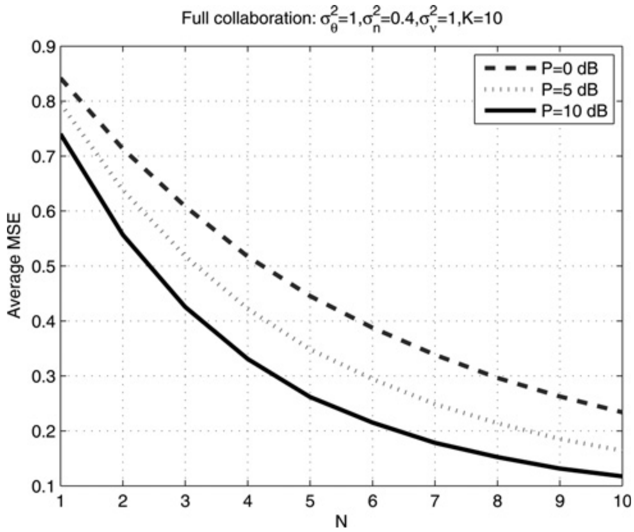

Fig. 2 *MSE of full collaboration case with different numbers of transmitters*

Gaussian, and we assume that $\sigma_\theta^2 = \sigma_\nu^2 = 1$. The observation gains $f_{l,k}$ are assumed to be uniformly distributed in the interval [0.5, 1]. The channel gains are taken as $c_g d^{-3.5}$, where d is uniformly drawn from the interval [1, 10] and $c_g = 22.6$ is a normalisation constant to make $E[g_{l,n}] = 1$ as in [7].

We first consider the effect of different numbers of transmitters N, where $N \le K$, in the full collaboration case. We set $K = 10$ and $\sigma_n^2 = 0.4$. In Fig. 2, we plot the average MSE against N with power levels $P = 0$, 5 and 10 dB. We note that as N increases, the MSEs decrease; also, large power levels result in smaller MSEs.

In all the simulations to follow, the number of sensors and the number of transmitters are set equal. Fig. 3 shows the average MSE against P for the full collaboration and non-collaboration cases with different observation noises, $\sigma_n^2 = 0.4$ and $\sigma_n^2 = 0.8$. We set $K = 20$. For $\sigma_n^2 = 0.4$, the case with full collaboration performs better than the non-collaboration case. Moreover, as the transmitted power increases, the MSEs for both two cases decrease. In fact, from (11), these two cases approach identical MSE as

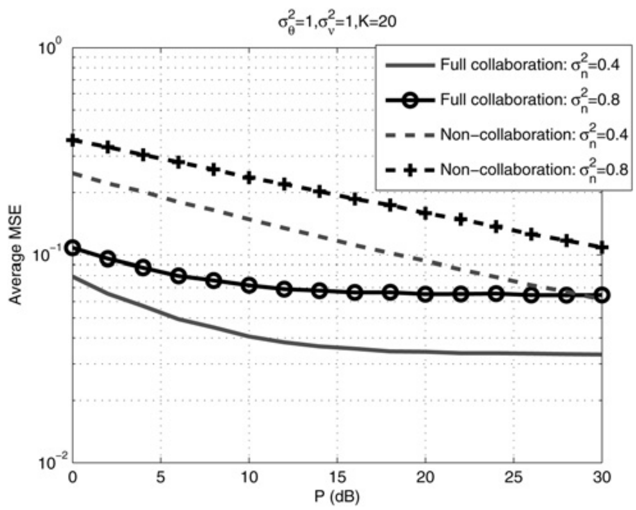

Fig. 3 *MSE of full collaboration and non-collaboration cases with different power levels*

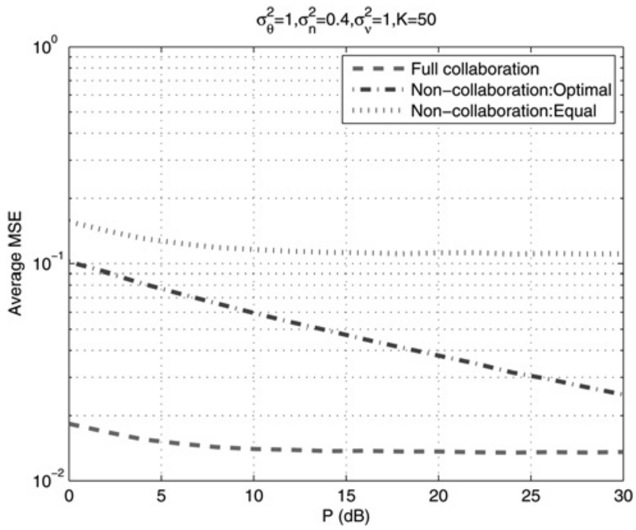

Fig. 4 *MSE of full collaboration and non-collaboration cases with different power levels*

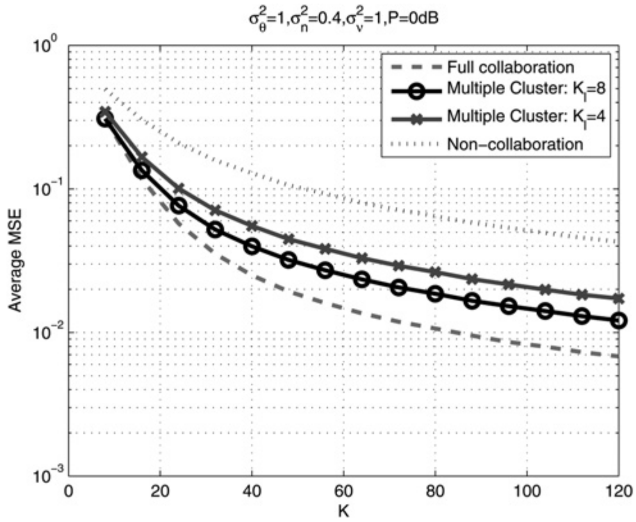

Fig. 5 *MSEs for $K_l = 1$, 4, 8, and K with different number of sensors*

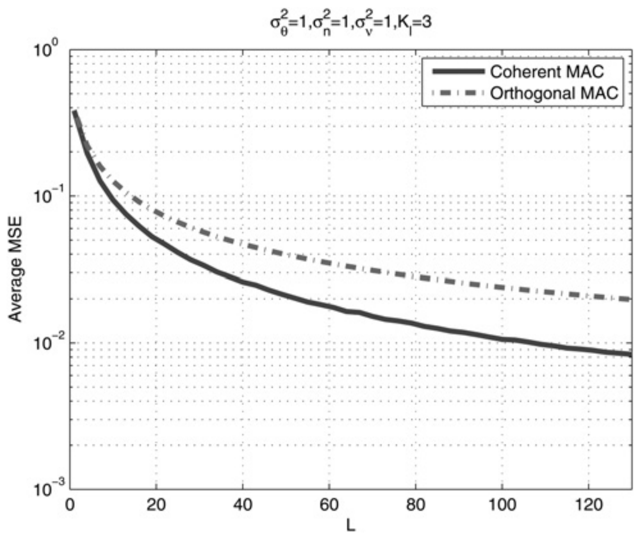

Fig. 6 *Comparison of the coherent MAC model to that of the orthogonal MAC model*

Table 1 Different number of sensors in clusters

					$K = 30$	$P = 0$ dB				
clusters, L	4	5	5	6	6	6	7	7	8	9
number of entries	252	226	218	218	200	184	184	166	162	124
MSE, J_M	0.0566	0.0603	0.0617	0.0635	0.0642	0.0659	0.0687	0.0690	0.0704	0.0790

$P \to \infty$. We also see that the MSE of the case with $\sigma_n^2 = 0.4$ is smaller than that of the case with $\sigma_n^2 = 0.8$, that is, a large signal-to-noise ratio results in a good performance.

Fig. 4 shows the comparison for the full collaboration case and two non-collaboration cases. The first non-collaboration case uses the optimal power allocation scheme and the second case uses equal power allocation scheme, in which the amplification gains are chosen as $a_k = \sqrt{P/K}$, $1 \le k \le K$. We set $K = 50$ and $\sigma_n^2 = 0.4$. Clearly, optimal power allocation improves performance over equal power allocation. The reduction in MSE by full collaboration with optimal power allocation is about 10 dB compared with the equal power allocation scheme.

We now consider two multiple cluster cases: in case 1, each cluster consists of 4 sensors, and in case 2, each cluster consists of eight sensors. Hence for a fixed number of sensors K, case 1 has $K/4$ clusters and case 2 has $K/8$ clusters. We compare their performance with the full collaboration and non-collaboration cases. We set $P = 0$ dB and $\sigma_n^2 = 0.4$. Fig. 5 shows that MSE of case 2 is less than that of case 1 since for a fixed K, case 2 has a smaller number of clusters and thus more collaboration among sensors. That the full collaboration has the lowest MSE and the non-collaboration case has the highest MSE is as predicted by (14).

For comparison, we also simulate the scheme proposed in [12] based on the orthogonal MAC model, where the measurement vector for lth cluster is $\boldsymbol{x}_l = \boldsymbol{f}_l\theta + \boldsymbol{n}_l$ which then transmits to the lth receiver through a diagonal channel gain matrix $\boldsymbol{D}_l$ after multiplying by an amplification matrix $\boldsymbol{A}_l$; at the FC, the received signal vector from the lth cluster is $\boldsymbol{y}_l = \boldsymbol{D}_l\boldsymbol{A}_l\boldsymbol{x}_l + \nu_l$, $l = 1, \ldots, L$, where the additive noise ν_l is assumed to be $E[\nu_l] = 0$, $E[\boldsymbol{\nu}_l\boldsymbol{\nu}_l^{\mathrm{T}}] = \sigma_\nu^2\boldsymbol{I}_{K_l}$, $E[\boldsymbol{\nu}_l\boldsymbol{\nu}_j^{\mathrm{T}}] = 0_{K_l \times K_l}$ for $j \ne l$. After collecting L signal vectors at the FC, the LMMSE fusion rule is used for estimating the source signal. The performance comparison for the orthogonal and coherent MAC models is plotted in Fig. 6, in which we take $P = 10$ dB, $K_l = 3$ for all clusters, and $\sigma_n^2 = 1$. We see that the MSE of the coherent MAC model performs better than that of the orthogonal MAC model. This is because by using the orthogonal MAC model, the number of receiver noises increases as the number of clusters increases. However, by using the coherent MAC model, there is only one receiver noise regardless of the number of clusters.

Finally to see quantitatively the relation between collaboration and MSE, we consider a network with 30 sensors and $P = 0$ dB. We perform ten simulations with the number of cluster ranging from 4 to 9. The number of sensors in each cluster is randomly chosen from 1 to 10. In each case, we count the total number of entries in the amplification matrices $\boldsymbol{A}_l$. For example, in the first case there are four clusters, the numbers of sensors in the clusters are respectively 9, 9, 9 and 3, and the number of entries is $9^2 + 9^2 + 9^2 + 3^2 = 252$. Table 1 shows the number of cluster, the number of entries, and the corresponding MSE for each case. From the table, we see that the MSE decreases as the number of entries increases. For comparison, the MSE for the two special cases are respectively $J_{\mathrm{N}} = 0.1689$ and $J_{\mathrm{C}} = 0.0392$.

5 Conclusion

We study optimal collaboration for distributed estimation in cluster-based wireless sensor network. We show that the optimal amplification matrix of each cluster is a rank one matrix obtained as a scaled outer product of the observation gain and the channel gain vectors. We also show that with optimal amplification matrices, estimation performance is improved compared with the non-collaboration case. We demonstrate, through simulation results, that the amount of improvement is closely related to the amount of collaboration.

6 Acknowledgment

We thank the reviewers for their helpful suggestions that improve the paper. Research sponsored by National Science Council under grant NSC 97-2221-E009-046-MY3.

7 References

1 Xiao, J.-J., Ribeiro, A., Luo, Z.-Q., Giannakis, G.B.: ‘Distributed compression-estimation using wireless sensor networks’, *IEEE Signal Process. Mag.*, 2006, **23**, (4), pp. 27–41

2 Abbasi, A.A., Younis, M.: ‘A survey on clustering algorithms for wireless sensor networks’, *Elsevier, J. Comput. Commun.*, 2007, **30**, pp. 2826–2841

3 Younis, O., Krunz, K., Ramasubramanian, S.: ‘Node clustering in wireless sensor networks: recent developments and deployment challenges’, *IEEE Netw. Mag.*, 2006, **20**, (3), pp. 20–25

4 Wimalajeewa, T., Jayaweera, S.K.: ‘Distributed node selection for sequential estimation over noisy communication channels’, *IEEE Trans. Wirel. Commun.*, 2010, **9**, (7), pp. 2290–2301

5 Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: ‘An application-specific protocol architecture for wireless microsensor networks’, *IEEE Trans. Wirel. Commun.*, 2002, **1**, (4), pp. 660–669

6 Khajehnouri, N., Sayed, A.H.: ‘Distributed MMSE relay strategies for wireless sensor networks’, *IEEE Trans. Signal Process.*, 2007, **55**, (7), pp. 3336–3348

7 Xiao, J.-J., Cui, S., Luo, Z.-Q., Goldsmith, A.J.: ‘Linear coherent decentralized estimation’, *IEEE Trans. Signal Process.*, 2008, **56**, (2), pp. 757–770

8 Guo, W., Xiao, J.-J., Cui, S.: ‘An efficient water-filling solution for linear coherent joint estimation’, *IEEE Trans. Signal Process.*, 2008, **56**, (10), pp. 5301–5305

9 Behbahani, A.S., Eltawil, A.M., Jafarkhani, H.: ‘Linear decentralized estimation of correlated data for wireless sensor networks’. Proc. IEEE Conf. Sensor, Mesh, and Ad Hoc Communications and Networks, Utah, USA, June 2011, pp. 73–79

10 Bahceci, I., Khandani, A.J.: ‘Linear estimation of correlated data in wireless sensor networks with optimum power allocation and analog modulation’, *IEEE Trans. Commun.*, 2008, **56**, (7), pp. 1146–1156

11 Cui, S., Xiao, J.-J., Goldsmith, A.J., Luo, Z.-Q., Poor, H.V.: ‘Estimation diversity and energy efficiency in distributed sensing’, *IEEE Trans. Signal Process.*, 2007, **55**, (9), pp. 4683–4695

12 Fang, J., Li, H.: ‘Power constrained distributed estimation with cluster-based sensor collaboration’, *IEEE Trans. Wirel. Commun.*, 2009, **8**, (7), pp. 3822–3832

13 Fang, J., Li, H.: ‘Power constrained distributed estimation with correlated sensor data’, *IEEE Trans. Signal Process.*, 2009, **57**, (8), pp. 3292–3297

doi: 10.1049/iet-spr.2011.0199

14 Chaudhary, M.H., Vandendorpe, L.: 'Power constrained linear estimation in wireless sensor networks with correlated data and digital modulation', *IEEE Trans. Signal Process.*, 2012, **60**, (2), pp. 570–584
15 Wu, J.-Y., Wang, T.-Y.: 'Power allocation for robust distributed best-linear-unbiased estimation against sensing noise variance uncertainty'. Proc. IEEE Int. Workshop Signal Processing and Advances Wireless Communication, CA, USA, June 2011, pp. 186–190
16 Liu, J.-H., Chung, C.-D.: 'Distributed estimation in a wireless sensor network using hybrid MAC', *IEEE Trans. Veh. Technol*, 2011, **60**, (7), pp. 3424–3435
17 Kay, S.M.: 'Fundamentals of statistical signal processing: estimation theory' (Prentice-Hall PTR, 1993)
18 Bernstein, D.S.: 'Matrix mathematics: theory, facts, and formulas with application to linear systems theory' (Princeton University Press, 2005)

8 Appendix 1: Proof of Proposition 1

Write $\boldsymbol{A}_l^{\mathrm{T}} = [\boldsymbol{a}_{l,1}\boldsymbol{a}_{l,2}\cdots\boldsymbol{a}_{l,N_l}]$, where $\boldsymbol{a}_{l,j}\in\mathbb{R}^{K_l}$. Define slack variables $t_l = \boldsymbol{g}_l^{\mathrm{T}}\boldsymbol{A}_l\boldsymbol{f}_l = \sum_{n=1}^{N_l} g_{l,n}\boldsymbol{a}_{l,n}^{\mathrm{T}}\boldsymbol{f}_l$, $1\le l\le L$. The problem (8) is rewritten as (see (16))

The Lagrangian function for (16) is

$$\begin{aligned} L(\boldsymbol{a}_{l,n}, t_l, \lambda_l, \lambda_0) = {} & \sum_{l=1}^{L}\left(\sigma_\theta^2\sum_{n=1}^{N_l}(\boldsymbol{a}_{l,n}^T\boldsymbol{f}_l)^2 + \sigma_n^2\sum_{n=1}^{N_l}(\boldsymbol{a}_{l,n}^{\mathrm{T}}\boldsymbol{a}_{l,n})\right) \\ & + \sum_{l=1}^{L}\lambda_l\left(t_l - \sum_{n=1}^{N_l} g_{l,n}\boldsymbol{a}_{l,n}^{\mathrm{T}}\boldsymbol{f}_l\right) \\ & + \lambda_0\left[\sigma_n^2\sum_{l=1}^{L}\left(\sum_{n=1}^{N_l} g_{l,n}\boldsymbol{a}_{l,n}^{\mathrm{T}}\right)\left(\sum_{m=1}^{N_l} g_{l,m}\boldsymbol{a}_{l,m}\right)\right. \\ & \left. + \sigma_\nu^2 - \left(\frac{1}{J^*}-\frac{1}{\sigma_\theta^2}\right)^{-1}\left(\sum_{l=1}^{L} t_l\right)^2\right] \end{aligned}$$

where $\lambda_l, \lambda_0 \in \mathbb{R}$, and the associated necessary conditions for optimality are

$$\begin{aligned} \frac{\partial L}{\partial \boldsymbol{a}_{l,n}} = {} & 2\boldsymbol{a}_{l,n}^{\mathrm{T}}(\sigma_\theta^2\boldsymbol{f}_l\boldsymbol{f}_l^T + \sigma_n^2\boldsymbol{I}_{K_l}) - \lambda_l g_{l,n}\boldsymbol{f}_l^{\mathrm{T}} \\ & + 2\lambda_0\sigma_n^2 g_{l,n}\boldsymbol{g}_l^{\mathrm{T}}\boldsymbol{A}_l = 0_{1\times K_l}^{\mathrm{T}}, \quad 1\le n\le N_l,\ 1\le l\le L \\ \Rightarrow {} & 2\boldsymbol{A}_l(\sigma_\theta^2\boldsymbol{f}_l\boldsymbol{f}_l^{\mathrm{T}} + \sigma_n^2\boldsymbol{I}_{K_l}) - \lambda_l\boldsymbol{g}_l\boldsymbol{f}_l^{\mathrm{T}} \\ & + 2\lambda_0\sigma_n^2\boldsymbol{g}_l\boldsymbol{g}_l^{\mathrm{T}}\boldsymbol{A}_l = 0_{N_l\times K_l}, \quad 1\le l\le L \end{aligned} \tag{17}$$

$$\frac{\partial L}{\partial t_l} = \lambda_l - 2\lambda_0\left(\frac{1}{J^*}-\frac{1}{\sigma_\theta^2}\right)^{-1}\left(\sum_{i=1}^{L} t_i\right) = 0, \quad 1\le l\le L \tag{18}$$

$$\frac{\partial L}{\partial \lambda_l} = t_l - \boldsymbol{g}_l^{\mathrm{T}}\boldsymbol{A}_l\boldsymbol{f}_l = 0, \quad 1\le l\le L \tag{19}$$

$$\begin{aligned} \frac{\partial L}{\partial \lambda_0} & = \sigma_n^2\sum_{l=1}^{L}\boldsymbol{g}_l^{\mathrm{T}}\boldsymbol{A}_l\boldsymbol{A}_l^{\mathrm{T}}\boldsymbol{g}_l + \sigma_\nu^2 - \left(\frac{1}{J^*}-\frac{1}{\sigma_\theta^2}\right)^{-1}\left(\sum_{l=1}^{L} t_l\right)^2 = 0 \\ & \Rightarrow \sigma_n^2\sum_{l=1}^{L}\boldsymbol{g}_l^{\mathrm{T}}\boldsymbol{A}_l\boldsymbol{A}_l^{\mathrm{T}}\boldsymbol{g}_l + \sigma_\nu^2 = \left(\frac{1}{J^*}-\frac{1}{\sigma_\theta^2}\right)^{-1}\left(\sum_{l=1}^{L} t_l\right)^2 \end{aligned} \tag{20}$$

It follows from (18) that $\lambda_1 = \ldots = \lambda_L$. Let $\lambda_l = \lambda$, $\forall l$. It follows from (17) that

$$\begin{aligned} & \boldsymbol{A}_l + \lambda_0\sigma_n^2\boldsymbol{g}_l\boldsymbol{g}_l^{\mathrm{T}}\boldsymbol{A}_l(\sigma_\theta^2\boldsymbol{f}_l\boldsymbol{f}_l^{\mathrm{T}} + \sigma_n^2\boldsymbol{I}_{K_l})^{-1} \\ & = \frac{\lambda}{2}\boldsymbol{g}_l\boldsymbol{f}_l^{\mathrm{T}}(\sigma_\theta^2\boldsymbol{f}_l\boldsymbol{f}_l^{\mathrm{T}} + \sigma_n^2\boldsymbol{I}_{K_l})^{-1} \\ & \Rightarrow \boldsymbol{A}_l + \lambda_0\boldsymbol{g}_l\boldsymbol{g}_l^{\mathrm{T}}\boldsymbol{A}_l - \frac{\sigma_\theta^2\lambda_0 t_l}{\sigma_n^2+\sigma_\theta^2\|\boldsymbol{f}_l\|^2}\boldsymbol{g}_l\boldsymbol{f}_l^{\mathrm{T}} \\ & = \frac{\lambda}{2(\sigma_n^2+\sigma_\theta^2\|\boldsymbol{f}_l\|^2)}\boldsymbol{g}_l\boldsymbol{f}_l^{\mathrm{T}} \\ & \Rightarrow \boldsymbol{A}_l = \frac{\sigma_\theta^2\lambda_0 t_l}{\sigma_n^2+\sigma_\theta^2\|\boldsymbol{f}_l\|^2}(\boldsymbol{I}_{N_l}+\lambda_0\boldsymbol{g}_l\boldsymbol{g}_l^{\mathrm{T}})^{-1}\boldsymbol{g}_l\boldsymbol{f}_l^{\mathrm{T}} \\ & \quad + \frac{\lambda}{2(\sigma_n^2+\sigma_\theta^2\|\boldsymbol{f}_l\|^2)}(\boldsymbol{I}_{N_l}+\lambda_0\boldsymbol{g}_l\boldsymbol{g}_l^{\mathrm{T}})^{-1}\boldsymbol{g}_l\boldsymbol{f}_l^{\mathrm{T}} \\ & \Rightarrow \boldsymbol{A}_l = \frac{\sigma_\theta^2\lambda_0 t_l}{(\sigma_n^2+\sigma_\theta^2\|\boldsymbol{f}_l\|^2)(1+\lambda_0\|\boldsymbol{g}_l\|^2)}\boldsymbol{g}_l\boldsymbol{f}_l^{\mathrm{T}} \\ & \quad + \frac{\lambda}{2(\sigma_n^2+\sigma_\theta^2\|\boldsymbol{f}_l\|^2)(1+\lambda_0\|\boldsymbol{g}_l\|^2)}\boldsymbol{g}_l\boldsymbol{f}_l^{\mathrm{T}} \end{aligned} \tag{21}$$

where in the second equation we use the matrix inversion lemma [18, p. 45] and $t_l = \boldsymbol{g}_l^{\mathrm{T}}\boldsymbol{A}_l\boldsymbol{f}_l$. Substituting (21) into (19), we obtain $t_l = \lambda\|\boldsymbol{g}_l\|^2\|\boldsymbol{f}_l\|^2/(2\varphi_l)$, where $\varphi_l = \sigma_\theta^2\|\boldsymbol{f}_l\|^2 + \sigma_n^2 + \lambda_0\sigma_n^2\|\boldsymbol{g}_l\|^2$, and thus from (21), we obtain $\boldsymbol{A}_l = \lambda/(2\varphi_l)\boldsymbol{g}_l\boldsymbol{f}_l^{\mathrm{T}}$. From (18) and $t_l = \lambda\|\boldsymbol{g}_l\|^2\|\boldsymbol{f}_l\|^2/2\varphi_l$, we have

$$\frac{1}{J^*}-\frac{1}{\sigma_\theta^2} = \sum_{i=1}^{L}\frac{\lambda_0\|\boldsymbol{g}_i\|^2\|\boldsymbol{f}_i\|^2}{\varphi_i} \tag{22}$$

Substituting $t_l = \lambda\|\boldsymbol{g}_l\|^2\|\boldsymbol{f}_l\|^2/(2\varphi_l)$, $\boldsymbol{A}_l = \lambda/(2\varphi_l)\boldsymbol{g}_l\boldsymbol{f}_l^{\mathrm{T}}$, and (22) into (20), we have

$$\frac{\lambda}{2} = \sqrt{\left(\sum_{i=1}^{L}\frac{\|\boldsymbol{g}_i\|^2\|\boldsymbol{f}_i\|^2(\sigma_\theta^2\|\boldsymbol{f}_i\|^2+\sigma_n^2)}{\varphi_i}\right)^{-1}}\,\sigma_\nu\lambda_0 \tag{23}$$

$$\begin{cases} \min\limits_{\boldsymbol{a}_{l,n},t_l} & \sum\limits_{l=1}^{L}\left(\sigma_\theta^2\sum\limits_{n=1}^{N_l}(\boldsymbol{a}_{l,n}^{\mathrm{T}}\boldsymbol{f}_l)^2 + \sigma_n^2\sum\limits_{n=1}^{N_l}(\boldsymbol{a}_{l,n}^{\mathrm{T}}\boldsymbol{a}_{l,n})\right) \\ \text{subject to} & \sum\limits_{n=1}^{N_l} g_{l,n}\boldsymbol{a}_{l,n}^{\mathrm{T}}\boldsymbol{f}_l = t_l, \quad 1\le l\le L \\ & \sigma_n^2\sum\limits_{l=1}^{L}\left(\sum\limits_{n=1}^{N_l} g_{l,n}\boldsymbol{a}_{l,n}^{\mathrm{T}}\right)\left(\sum\limits_{m=1}^{N_l} g_{l,m}\boldsymbol{a}_{l,m}\right) + \sigma_\nu^2 = \left(\frac{1}{J^*}-\frac{1}{\sigma_\theta^2}\right)^{-1}\left(\sum\limits_{l=1}^{L} t_l\right)^2 \end{cases} \tag{16}$$

From (23) and $\boldsymbol{A}_l = \lambda/(2\varphi_l)\boldsymbol{g}_l\boldsymbol{f}_l^{\mathrm{T}}$, the minimum total power can be written as follows:

$$P_{\min} = \sum_{l=1}^{L} \mathrm{tr}(\sigma_\theta^2 \boldsymbol{A}_l\boldsymbol{f}_l\boldsymbol{f}_l^{\mathrm{T}}\boldsymbol{A}_l^{\mathrm{T}} + \sigma_n^2\boldsymbol{A}_l\boldsymbol{A}_l^{\mathrm{T}}) = \lambda_0\sigma_\nu^2$$

It follows that

$$\lambda_0 = P_{\min}/\sigma_\nu^2 \tag{24}$$

and thus from (22), we obtain

$$\frac{1}{J^*} - \frac{1}{\sigma_\theta^2} = \sum_{i=1}^{L} \frac{P_{\min}\|\boldsymbol{g}_i\|^2\|\boldsymbol{f}_i\|^2}{\sigma_\nu^2(\sigma_\theta^2\|\boldsymbol{f}_l\|^2 + \sigma_n^2) + P_{\min}\sigma_n^2\|\boldsymbol{g}_l\|^2} \tag{25}$$

Equation (25) gives the relation between the achieved minimum power and the constraint J^* on MSE. The optimal amplification matrices are $\boldsymbol{A}_l = \lambda/(2\varphi_l)\boldsymbol{g}_l\boldsymbol{f}_l^{\mathrm{T}}$, where λ and φ_l depends on $P_{\min}$ through (24). In view of observation (iii) in Section 3, if we set $P_{\min} = P$, the corresponding MSE is given in (10) and the corresponding amplification matrices is in (9). □

9 Appendix 2: Proof of Lemma 1

We first show that for $\boldsymbol{x} = [x_1 \cdots x_n]^{\mathrm{T}} \in \mathbb{R}^n$ and $\boldsymbol{y} = [y_1 \cdots y_n]^{\mathrm{T}} \in \mathbb{R}^n$, the following inequality holds

$$\frac{\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2}{\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2} \geq \sum_{i=1}^{n} \frac{x_i^2 y_i^2}{x_i^2 + y_i^2} \tag{26}$$

or equivalently,

$$\frac{\sum_{i=1}^n x_i^2 \sum_{j=1}^n y_j^2}{\sum_{j=1}^n (x_j^2 + y_j^2)} \geq \frac{\sum_{i=1}^n \left(x_i^2 y_i^2 \prod_{k\neq i}^n (x_k^2 + y_k^2)\right)}{\prod_{k=1}^n (x_k^2 + y_k^2)}$$
$$\Leftrightarrow \left(\sum_{i=1}^n x_i^2\right)\left(\sum_{j=1}^n y_j^2\right)\prod_{k=1}^n (x_k^2 + y_k^2)$$
$$- \left[\sum_{j=1}^n (x_j^2 + y_j^2)\right]\left[\sum_{i=1}^n\left(x_i^2 y_i^2 \prod_{k\neq i}^n (x_k^2 + y_k^2)\right)\right] \geq 0$$

The left-hand side of the above inequality can be written as follows: (see equation at the bottom of the page)

Thus, we obtain (26). Now, let $\boldsymbol{x} = [\boldsymbol{x}_1^{\mathrm{T}} \cdots \boldsymbol{x}_L^{\mathrm{T}}]^{\mathrm{T}} = [x_1 \cdots x_n]^{\mathrm{T}}$ and $\|\boldsymbol{x}_l\|^2 = \tilde{x}_l^2$, then we have

$$\|\boldsymbol{x}\|^2 = \sum_{i=1}^n x_i^2 = \sum_{l=1}^L \|\boldsymbol{x}_l\|^2 = \sum_{l=1}^L \tilde{x}_l^2 = \|\tilde{\boldsymbol{x}}\|^2 \tag{27}$$

where $\tilde{\boldsymbol{x}} = [\tilde{x}_1 \cdots \tilde{x}_L]^{\mathrm{T}}$. By the same way, we have

$$\|\boldsymbol{y}\|^2 = \sum_{i=1}^n y_i^2 = \sum_{l=1}^L \|\boldsymbol{y}_l\|^2 = \sum_{l=1}^L \tilde{y}_l^2 = \|\tilde{\boldsymbol{y}}\|^2 \tag{28}$$

From (26)–(28), we have

$$\frac{\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2}{\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2} = \frac{\|\tilde{\boldsymbol{x}}\|^2\|\tilde{\boldsymbol{y}}\|^2}{\|\tilde{\boldsymbol{x}}\|^2 + \|\tilde{\boldsymbol{y}}\|^2} \geq \sum_{l=1}^L \frac{\tilde{x}_l^2\tilde{y}_l^2}{\tilde{x}_l^2 + \tilde{y}_l^2}$$
$$= \sum_{l=1}^L \frac{\|\boldsymbol{x}_l\|^2\|\boldsymbol{y}_l\|^2}{\|\boldsymbol{x}_l\|^2 + \|\boldsymbol{y}_l\|^2}$$

and the result follows. □

$$\left[\sum_{i=1}^n\sum_{j=1}^n x_i^2 y_j^2\right]\prod_{k=1}^n (x_k^2 + y_k^2) - \left[\sum_{i=1}^n\sum_{j=1}^n x_i^2 y_i^2 (x_j^2 + y_j^2)\right]\prod_{k\neq i}^n (x_k^2 + y_k^2)$$
$$= \left[\sum_{i=1}^n\sum_{j=i}^n x_i^2 y_j^2\right]\prod_{k=1}^n (x_k^2 + y_k^2) - \left[\sum_{i=1}^n x_i^2 y_i^2\right]\prod_{k=1}^n (x_k^2 + y_k^2) - \left[\sum_{i=1}^n\sum_{j\neq i}^n x_i^2 y_i^2 (x_j^2 + y_j^2)\right]\prod_{k\neq i}^n (x_k^2 + y_k^2)$$
$$= \left[\sum_{i=1}^n\sum_{j\neq i}^n x_i^2 y_j^2\right]\prod_{k=1}^n (x_k^2 + y_k^2) - \left[\sum_{i=1}^n\sum_{j\neq i}^n x_i^2 y_i^2 (x_j^2 + y_j^2)\right]\prod_{k\neq i}^n (x_k^2 + y_k^2)$$
$$= \left[\sum_{i=1}^n\sum_{j\neq i}^n x_i^2 y_j^2 (x_i^2 + y_i^2) - \sum_{i=1}^n\sum_{j\neq i}^n x_i^2 y_i^2 (x_j^2 + y_j^2)\right]\prod_{k\neq i}^n (x_k^2 + y_k^2) = \left[\sum_{i=1}^n\sum_{j\neq i}^n x_i^2 (x_i^2 y_j^2 - x_j^2 y_i^2)\right]\prod_{k\neq i}^n (x_k^2 + y_k^2)$$
$$= \left[\sum_{i=1}^n\sum_{j>i}^n x_i^2 (x_i^2 y_j^2 - x_j^2 y_i^2)\right]\prod_{k\neq i}^n (x_k^2 + y_k^2) + \left[\sum_{i=1}^n\sum_{j<i}^n x_i^2 (x_i^2 y_j^2 - x_j^2 y_i^2)\right]\prod_{k\neq i}^n (x_k^2 + y_k^2)$$
$$= \left[\sum_{i=1}^n\sum_{j>i}^n x_i^2 (x_i^2 y_j^2 - x_j^2 y_i^2)\right]\prod_{k\neq i}^n (x_k^2 + y_k^2) + \left[\sum_{i=1}^n\sum_{j>i}^n x_j^2 (x_j^2 y_i^2 - x_i^2 y_j^2)\right]\prod_{k\neq j}^n (x_k^2 + y_k^2)$$
$$= \left[\sum_{i=1}^n\sum_{j>i}^n x_i^2 (x_j^2 + y_j^2)(x_i^2 y_j^2 - x_j^2 y_i^2) - x_j^2 (x_i^2 + y_i^2)(x_i^2 y_j^2 - x_j^2 y_i^2)\right]\prod_{k\neq i,j}^n (x_k^2 + y_k^2)$$
$$= \sum_{i=1}^n\sum_{j>i}^n (x_i^2 y_j^2 - x_j^2 y_i^2)^2 \prod_{k\neq i,j}^n (x_k^2 + y_k^2) \geq 0$$

doi: 10.1049/iet-spr.2011.0199