DISTRIBUTED ESTIMATION OVER PARALLEL FADING CHANNELS WITH CHANNEL ESTIMATION ERROR

*Habib Şenol**

Kadir Has University
Department of Computer Engineering
Istanbul, Turkey

Cihan Tepedelenlioğlu †

Arizona State University
Department of Electrical Engineering
Tempe, Arizona

ABSTRACT

We consider distributed estimation of a source observed by sensors in additive Gaussian noise, where the sensors are connected to a fusion center with unknown orthogonal (parallel) flat Rayleigh fading channels. We adopt a two-phase approach of (i) channel estimation with training, and (ii) source estimation given the channel estimates, where the total power is fixed. We prove that allocating half the total power into training is optimal, and show that compared to the perfect channel case, a performance loss of at least 6 dB is incurred. In addition, we show that unlike the perfect channel case, increasing the number of sensors will lead to an eventual degradation in performance. We characterize the optimum number of sensors as a function of the total power and noise statistics. Simulations corroborate our analytical findings.

Index Terms— Sensor Networks, Distributed Estimation, Fading Channels, Channel Estimation

1. INTRODUCTION

A wireless sensor network (WSN) consists of spatially distributed sensors which are capable of monitoring physical phenomena. Sensors typically have limited processing and communication capability because of their limited battery power. In most WSNs a fusion center (FC) which has less limitations in terms of processing and communication, receives transmissions from the sensors over the wireless channels so as to combine the received signals to make inferences on the observed phenomenon.

Especially over the past few years, research on distributed estimation has been evolving very rapidly [1]. Universal decentralized estimators of a source over additive noise have been considered in [2,3]. Much of the literature has focused in finite-rate transmissions of quantized sensor observations [1]. The observations of the sensors can be delivered to the FC by analog or digital transmission methods. Amplify-and-forward is one analog option, whereas in digital transmission, observations are quantized, encoded and transmitted via digital modulation. The optimality of amplify and forward in several settings described in [4], [5]. In [5], amplify-and-forward over orthogonal parallel MAC with perfect channel knowledge at the FC is considered, where increasing the number of sensors is shown to improve performance.

In this work, we consider unknown fading channels where we follow a two-step procedure to first estimate the fading channel coefficients with pilots, and use those estimates in constructing the estimator for the source signal with linear minimum mean square error (LMMSE) estimators. We characterize the effect of channel estimation error on performance for equal power scheduling at the sensors, and imperfect estimated channels at the FC. We show that when the total power for channel estimation and wireless transmission is fixed, increasing the number of sensors will eventually lead to a degradation in performance. Hence, in the absence of channel information, deploying more sensors might not necessarily lead to better performance. We also find approximate expressions for the optimum number of sensors to achieve minimum MSE performance and we characterize the penalty paid for estimating the channel to be factor of at least 4 (6 dB).

2. SYSTEM MODEL AND CHANNEL ESTIMATION

We assume the wireless sensor network (WSN) has K sensors and the k^{th} sensor observes an unknown zero-mean complex random source signal θ with zero mean and variance σ_θ^2, corrupted by a zero-mean additive complex Gaussian noise $n_k \sim \mathcal{CN}(0, \sigma_n^2)$ as shown in Fig.1. Since we assume the amplify-forward analog transmission scheme, the k^{th} sensor amplifies its incoming analog signal $\theta + n_k$ by a factor of α_k and transmits it on the k^{th} flat fading orthogonal channel to the fusion center (FC). In Fig.1, $g_k \sim \mathcal{CN}(0, \sigma_g^2)$ and $v_k \sim \mathcal{CN}(0, \sigma_v^2)$ are the flat fading channel gain and the channel noise of the k^{th} channel path, respectively. The ampli-

*First author's work is supported by The Scientific and Technological Research Council of Turkey (TUBITAK)

†Second author's work is supported in part by NSF Career Grant No. CCR-0133841

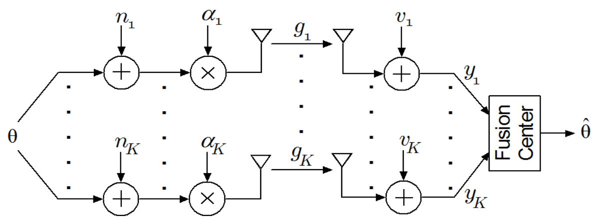

Fig. 1. Wireless Sensor Network with Orthogonal MAC

fication factor α_k is the same for all sensors since there is no channel status information (CSI) is available at the sensor side. The k^{th} received signal at the FC is given as

$$y_k = g_k \, \alpha_k \, (\theta + n_k) + v_k \; , k = 1, \cdots , K \; . \quad (1)$$

Based on this receive model, we will estimate the source signal θ. Our two-step strategy is first to estimate parallel channels, and then estimate the source signal given the channel estimates. We will use a LMMSE approach [6] for both steps. In the first phase, the sensors send training symbols of total power P_{trn} to estimate the parallel channels $\{g_k\}_{k=1}^{K}$. In the second phase the sensors transmit their amplified data, which bear information about θ, with a power of $P_{dat} := |\alpha_k|^2 \, (\sigma_\theta^2 + \sigma_n^2) = (P_{tot} - P_{trn})/K$, same for each sensor. Note that the total power in the two phases add to P_{tot}. The fusion center uses the received signal in the second phase and the channel estimates from the first phase to estimate the source signal θ.

To estimate the parallel fading channels $\{g_k\}_{k=1}^{K}$ in the training phase, we consider pilot-based channel estimation, where each sensor sends a pilot symbol to the FC over its own fading channel. The receive model for a pilot s transmitted over the k^{th} channel is

$$x_k = g_k \, s + \nu_k \; , \quad (2)$$

where x_k is the received signal over k^{th} channel and ν_k is zero-mean additive complex Gaussian channel noise, $\nu_k \sim \mathcal{CN}(0, \sigma_v^2)$. Since the total transmitted training power is P_{trn}, we have $P_{trn} = K|s|^2$. According to our observation model in (2), the linear minimum mean square error (LMMSE) estimate $\hat{g}_k$ of the channel g_k is given as follows [6]

$$\hat{g}_k = \frac{E_{\{g_k, x_k\}}[\, g_k \, x_k^* \,]}{E_{\{x_k\}}[\, |x_k|^2 \,]} \, x_k = \frac{\sigma_g^2 \, s^*}{\sigma_v^2 + \sigma_g^2 \, |s|^2} \, x_k \; , \quad (3)$$

where $(\cdot)^*$ denotes the complex conjugate and the channel estimation error variance δ^2 is given as

$$\delta^2 = \left(\frac{1}{\sigma_g^2} + \frac{|s|^2}{\sigma_v^2} \right)^{-1} = \frac{\sigma_v^2 \, \sigma_g^2}{\sigma_v^2 + \sigma_g^2 \, |s|^2} \; . \quad (4)$$

3. MSE OF SOURCE ESTIMATOR

In this section, we describe the estimation of the source signal θ, and the resulting MSE which will be our figure of merit. We use the LMMSE source estimator given the channel estimates $\{\hat{g}_k\}_{k=1}^{K}$ in (3), and the received signal $y_1, \ldots, y_K$ in (1). By doing this, we obtain the source estimator $\hat{\theta}$ in the presence of channel estimation error (CEE). Using the orthogonality principle of the LMMSE estimator, it is possible to show that the minimum MSE in the presence of CEE is given by [7]

$$D = \sigma_\theta^2 \left(1 + \sum_{k=1}^{K} \frac{\gamma \, \hat{\eta}_k \, (\sigma_g^2 - \delta^2) P_{dat}}{\left(\hat{\eta}_k \, (\sigma_g^2 - \delta^2) + \zeta \delta^2 \right) P_{dat} + \sigma_g^2} \right)^{-1} \quad (5)$$

with the following definitions:

Observation SNR	$\gamma := \sigma_\theta^2 / \sigma_n^2$
Variance of $\hat{g}_k$	$\sigma_{\hat{g}}^2 = \sigma_g^2 - \delta^2$
Total training power	$P_{trn} := K \, \lvert s \rvert^2$
Data power, every sensor	$P_{dat} := (P_{tot} - P_{trn})/K$ $= \lvert\alpha_k\rvert^2 \, \sigma_\theta^2 \, (1 + \gamma^{-1})$
Channel SNR	$\zeta := \sigma_g^2 / \sigma_v^2$
k^{th} estimated channel power	$\hat{\eta}_k := \frac{\zeta \, \lvert\hat{g}_k\rvert^2}{\sigma_{\hat{g}}^2 \, (\gamma+1)}$
k^{th} channel power	$\eta_k := \frac{\zeta \, \lvert g_k\rvert^2}{\sigma_g^2 \, (\gamma+1)}$

and we express the channel estimator variance δ^2 using (4) and $P_{trn} = K|s|^2$ as $\delta^2 = (K\sigma_g^2)/(K + \zeta P_{trn})$. Substituting this into (5), it is straightforward to verify that (5) is a convex function of P_{trn} by taking the second derivative. Before we optimize the training power, we will briefly review the perfect CSI case.

In what follows, we adapt the best linear unbiased estimator (BLUE) in [5] to the LMMSE case, since this will serve as a benchmark to the CEE case we derive later. With CSI at the FC, the variance of the channel estimation error is zero $\delta^2 = 0$ and the normalized estimated channel powers are equal to the normalized channel powers $\hat{\eta}_k = \eta_k \; \forall k$. By substituting $\delta^2 = 0$ and $\hat{\eta}_k = \eta_k$ in (5), the MSE expression for the perfect CSI case is obtained as follows

$$D^{(per)}(P_{tot}, K) = \sigma_\theta^2 \left(1 + \sum_{k=1}^{K} \frac{\gamma \, \eta_k}{\eta_k + \frac{K}{P_{tot}}} \right)^{-1} \; . \quad (6)$$

It is straightforward to verify that (6) is a monotonically decreasing function of the number of sensors K. In contrast to this perfect CSI case, we will later see that when the channel is estimated, increase in the number of sensors will not always improve performance.

We now consider the case where the FC has the LMMSE estimates of the channel without feeding back the CSI to the sensors, which transmit with equal power.

3.1. Optimum Training Power

It is clear that if the training power is too small, the resulting unreliable channel estimates will increase the MSE. On the other hand, if the training power P_{trn} is too close to P_{tot}, then each sensor transmits with a small power $P_{dat} = (P_{tot} -$

$P_{trn})/K$ and the FC does not receive much information about θ in the data transmission phase. To find the optimal P_{trn} we note that minimizing (5) and minimizing the sum in (5) are equivalent. Using the definitions in the table and expression for δ^2 below the table, we obtain the following convex optimization problem for the training power

$$\min_{0\leq P_{trn}\leq P_{tot}} \quad -\sum_{k=1}^{K} \frac{\gamma\,\hat{\eta}_k\,\zeta\,(P_{tot}-P_{trn})P_{trn}}{\hat{\eta}_k\,\zeta\,(P_{tot}-P_{trn})P_{trn}+K\zeta P_{tot}+K^2} \tag{7}$$

Using Lagrange multipliers and the Kuhn Tucker conditions for this one dimensional convex optimization problem, the optimum value of the training power $P^{\star}_{trn}$ can be shown to be half of the total power: $P^{\star}_{trn} = P_{tot}/2$ [7]. We stress that the optimum total training power P_{trn} is always half of the total power, regardless of the number of sensors, or the noise level. Substituting this optimum value into (7), we reach the following MSE expression

$$D^{(est)}(P_{tot},K) = \sigma^2_\theta \left(1+\sum_{k=1}^{K} \frac{\gamma\,\hat{\eta}_k}{\hat{\eta}_k + \frac{4K}{P_{tot}}(1+\frac{K}{\zeta P_{tot}})}\right)^{-1} . \tag{8}$$

It is easy to verify that the MSE performance of the source estimator is going to degrade as $K \to \infty$. To see this more clearly, note that (8) increases to its highest value σ^2_θ as the number of sensor goes to infinity: $\lim_{K\to\infty} D^{(est)}(P_{tot},K) = \sigma^2_\theta$. Recalling that σ^2_θ is the worst possible variance for $\hat{\theta}$, it is clear that increasing the number of sensors does not indefinitely improve performance, but rather degrades it after a certain number of sensors. This means that a finite optimum number of sensors minimizing the MSE exists in this imperfect CSI case.

3.2. Optimum Number of Sensors

In what follows, we obtain an approximate value of the optimum number of sensors. The optimum number of sensors $K^\star$ must be obtained by minimizing the expected value of the MSE $E_{\{\hat{\eta}_k\}}[D]$ since $K^\star$ can not depend on instantaneous channel estimates. Since this expectation is not tractable, we find an approximate value of $K^\star$ by minimizing a tight lower bound on $E_{\{\hat{\eta}_k\}}[D]$. We note that the MSE in (8) is convex with respect to the sum and use the Jensen's inequality

$$E[D^{(est)}(P_{tot},K)] \quad \geq \quad \frac{\sigma^2_\theta}{1+E\left[\frac{K\,\gamma\,\hat{\eta}_k}{\hat{\eta}_k+\frac{4K}{P_{tot}}(1+\frac{K}{\zeta P_{tot}})}\right]} \tag{9}$$

where the expectations are with respect to $\hat{\eta}_k$. To minimize (9) with respect to K, we treat K as a continuous parameter, and differentiate (9) with respect to K to get the following condition:

$$E\left[\frac{\hat{\eta}_k^2 - \frac{4K^2}{\zeta P_{tot}^2}\hat{\eta}_k}{(\hat{\eta}_k + \frac{4K}{P_{tot}}(1+\frac{K}{\zeta P_{tot}}))^2}\right]\Bigg|_{K=K^\star} = 0 \; . \tag{10}$$

Since the expectation above is still intractable, we note that the variance $\hat{\eta}_k$ is very small $\mathrm{var}[\hat{\eta}_k] = (\frac{\zeta}{\gamma+1})^2 \ll \frac{4K}{P_{tot}}(1+\frac{K}{\zeta P_{tot}})$. Treating the denominator as deterministic, and carrying out the required expectations, the optimum number of sensors is approximated as:

$$K^\star \approx \mathrm{round}\left(\frac{\zeta P_{tot}}{\sqrt{2(\gamma+1)}}\right) \; , \tag{11}$$

where the $\mathrm{round}(\cdot)$ is the nearest integer. We note that even though the optimum value in (11) is an approximation, it is quite accurate as shown in the simulations. Moreover, when the total power P_{tot} or the channel SNR ζ are large, the optimum number of sensors increase. This is because when P_{tot} is large, $P_{trn} = P_{tot}/2$ will also be large, leading to almost perfect channel estimates. This is in agreement with the fact that in the perfect channel case in (6), the optimum number of sensors is infinite since the performance always improves with the number of sensors. From (11) we also see that if the sensor observation SNR γ is increased, then it is best to use a smaller number of sensors. To explain this, first recall that in the perfect channel case, the reason the MSE improves monotonically with K is because more sensors average out the observation noise. In the imperfect channel case, however, the favorable averaging effect of having more sensors is offset by having to learn all the channel coefficients $\{g_k\}_{k=1}^K$ with a fixed total training power $P_{trn} = P_{tot}/2$, which results in increased channel estimation error variance δ, which ultimately degrades the MSE of θ. Therefore, the optimum value of K that strikes a balance in this tradeoff, increases when there is more noise to be averaged (smaller γ).

3.3. Comparison of Perfect and Imperfect CSI

In order to compare the MSE performances of the perfect and the estimated CSI cases for a fixed number of sensors K, we first note that the MSE expressions in (6) and (8) are random variables. Hence it is appropriate to derive the conditions under which the distributions of MSEs in (6) and (8) are identical. We will do this by exploiting the fact that the random variables η_k and $\hat{\eta}_k$ have identical distributions (both are exponential with mean $b = \zeta/(\gamma+1)$), and allow the perfect CSI case and the imperfect CSI case to have different total transmit powers ${P_{tot}}^{(per)}$ and ${P_{tot}}^{(est)}$ to see how much more power would be needed in the imperfect CSI case to get the same MSE distribution. The MSE expressions in (6) and (8) have identical distributions if and only if the deterministic terms in the denominator of the sums are equal: $K/{P_{tot}}^{(per)} = 4K/{P_{tot}}^{(est)}(1+K/(\zeta {P_{tot}}^{(est)}))$. Solving for ${P_{tot}}^{(est)}$ we obtain

$${P_{tot}}^{(est)} = 2{P_{tot}}^{(per)} + 2{P_{tot}}^{(per)}\sqrt{1+\frac{K}{\zeta {P_{tot}}^{(per)}}} \; , \tag{12}$$

which ensures that the expected MSE (averaged over the channel distribution) will be the same. From (12) we see that ${P_{tot}}^{(per)}/{P_{tot}}^{(est)} \leq 1/4$, which is a penalty of at least 6 dB for having to estimate the channel. The inequality becomes equal to 6 dB for large total powers ${P_{tot}}^{(est)}$: ${P_{tot}}^{(per)}/{P_{tot}}^{(est)} \rightarrow 1/4$, which is easily seen from (12). Recalling that half of the total power has to be spared for training, we can conclude that another 3 dB is lost due to the effect of estimation error at the FC.

4. NUMERICAL RESULTS

In Fig. 2 the simulation results indicate the accuracy of the optimum number $K^\star$ of sensors calculated from (11). We found the analytical formula to be very accurate in a wide range of settings. Even when the predicted number of sensors do not match the simulations perfectly (e.g., when $\gamma = 5$ in the figure), the resulting minimum average MSE obtained from (11) is very close to the minimum achievable average MSE.

In Fig. 3 the perfect and imperfect CSI cases are compared. It is seen that, for the two cases to have the same performance (ratio of average MSEs in the y-axis equaling unity), the total power for the estimated case is about 4 times as much as the perfect channel case. This agrees with analytical results mentioned earlier.

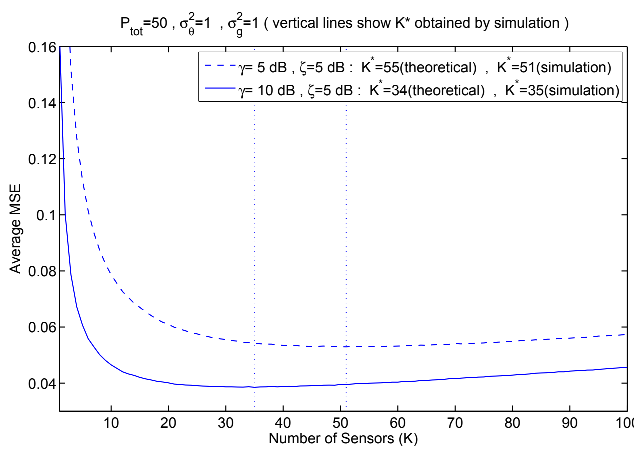

Fig. 2. Optimum number of sensors

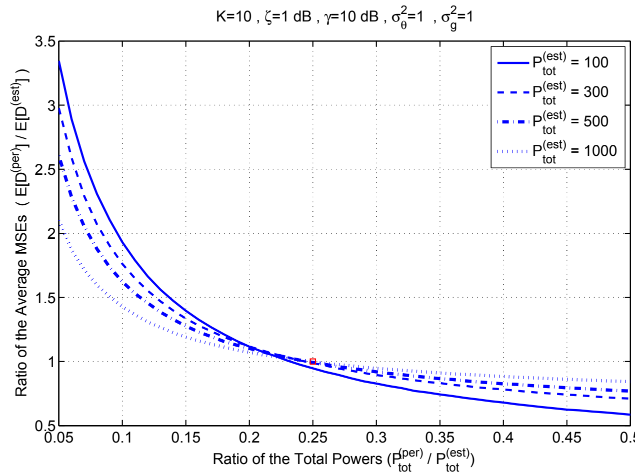

Fig. 3. Power loss due to estimation

5. CONCLUSIONS

To facilitate the estimation of the source θ, we estimated the fading channel coefficients. We found that half the total power is the optimum amount of training to estimate the fading channels, regardless of the SNR or the number of sensors. For the same MSE performance of the source estimator, it was found that at least a factor of 4 more total power is needed when the fading channels are unknown, compared to the case they are known perfectly. Unlike the perfect channel case, there is an optimum number of sensors, and we found an approximate formula to calculate this number.

6. REFERENCES

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

[2] Z.-Q. Luo, "Universal decentralized estimation in a bandwidth constrained sensor network," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2210–2219, June 2005.

[3] Jin-Jun Xiao, Shuguang Cui, Zhi-Quan Luo, and A.J. Goldsmith, "Power scheduling of universal decentralized estimation in sensor networks," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 413–422, February 2006.

[4] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code or not to code: Lossy source-channel communication revisited," *IEEE Transaction on Information Theory*, vol. 49, pp. 1147–1158, May 2003.

[5] S. Cui, J.-J. Xiao, A. J. Goldsmith, Z.-Q. Luo, and H. V. Poor, "Estimation diversity and energy efficiency in distributed sensing," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4683–4695, September 2007.

[6] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Prentice Hall, New Jersey, 1993.

[7] H. Senol and C. Tepedelenlioglu, "Distributed estimation over unknown parallel fading channels," *IEEE Transactions on Signal Processing*, (submitted).