Linear coherent distributed estimation over unknown channels

Chien-Hsien Wu *, Ching-An Lin

Department of Electrical and Control Engineering, National Chiao Tung University, 1001 University Road, Hsinchu 300, Taiwan

Article history:
Received 29 September 2009
Received in revised form
5 July 2010
Accepted 1 October 2010
Available online 8 October 2010

Keywords:
Distributed estimation
Channel estimation
Coherent multiple access
Power allocation
Sensor networks

Abstract

We study linear distributed estimation with coherent multiple access channel model and MMSE fusion rule. The flat fading channels are assumed unknown at the fusion center and need to be estimated. We adopt a two-phase approach, which first estimates channels and then estimates the source signal, to minimize the MSE of the estimated signal. We study optimal power allocation under a total network power constraint. We consider the optimal power allocation scheme in which training power and data power for each sensor are optimized, and the equal power allocation scheme in which training power is optimized while data power for each sensor is set equal. In both schemes, the problem is formulated as a constrained optimization problem and analytical closed-form solution is obtained. Analytic results reveal that (i) with estimated channels, the MSE approaches to a finite nonzero value as the number of sensors increases; (ii) the optimal training powers are the same in both schemes; (iii) the MSE performance compared with the case when channels are known shows the penalty caused by channel estimation becomes worse as the number of sensors increases. Simulation results verify our findings.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

Wireless sensor network (WSN) is composed of a large number of signal processing sensors, each is capable of simple local computation, short range low data-rate communication, and a fusion center (FC) that has more powerful communication and processing capability. The fusion center receives signals transmitted from the sensors over the wireless channels and combines the signals for a specific processing purpose. One example of such a distributed signal processing scheme is distributed estimation. A certain parameter or variable is measured by the sensors and the measurements are sent to the fusion center, and the goal is to estimate the parameter based on the distributed sensor measurements [1,2].

In the distributed estimation scenario, the sensors could transmit measurements to the fusion center based on quantized or unquantized strategy. In the quantized strategy, the measurements sent from the sensors are quantized, encoded, and transmitted via digital modulation. Due to practical limitations, it is important to make efficient use of energy and bandwidth. Some research works attempt to minimize transmitted power via bit length assignment under a predefined MSE constraint [3,4], while others focus on the search of quantization threshold for a fixed bit length [5,6]. In the unquantized strategy, the sensors send raw measurements directly through channels without quantization and thus analog transmission, such as amplify-and-forward approach, is used. It is asserted in [11] that the amplify-and-forward approach is optimal over additive white Gaussian noise channels. Along this line of approach, many papers study the minimization of mean squared estimation error under a total network power constraint by optimally allocating the transmitted power for each sensor [7–10] and others analyze asymptotic behavior as the network power or the number of sensors increases [12,13].
In the amplify-and-forward approach, two types of channel models are used, the orthogonal multiple access channel (MAC) [7–10] and the coherent MAC [7]. Linear minimum mean squared error (LMMSE) estimator for the coherent MAC model and the orthogonal MAC model with channel knowledge at the FC was discussed and its performance analyzed in [7]. The results indicate that the MSE for the orthogonal MAC model reaches a finite nonzero value as the number of sensors is increased without bound. Şenol and Tepedelenlioğlu [8] consider the orthogonal MAC model with unknown flat Rayleigh fading channels. A two-phase approach, which first estimates channels and then estimates source signal, is proposed. The result shows that with unknown fading channels, increasing the number of sensors may eventually lead to a degradation in performance if the total network power is fixed.

In this paper, we consider coherent MAC model with unknown flat fading channels. We derive training-based LMMSE channel estimator. The channel estimate is then used to obtain LMMSE estimation of the source signal. We consider the allocation of power to each sensor for training and for data transmission, under a total network power constraint, so as to minimize the MSE of the estimated source signal. We consider two schemes: (i) the optimal power allocation scheme and (ii) the equal power allocation scheme. In (i), training power and data power for each sensor are optimized, and the power gain (for data) of each sensor is computed based on the respective channel estimate and sent to the sensor from the FC. In (ii), training power is optimized, but the data power for each sensor is set equal, only the phase of the estimated channel is feedback to each sensor from the FC. In both schemes, the problem is formulated as a constrained optimization problem and analytical closed-form solution is obtained. We compare the performance of the distributed estimation scheme with estimated channels to that with actual known channels. The main results of this paper are as follows: (i) the MSE with estimated channels at the FC approaches to a finite nonzero value as the number of sensors increases; (ii) the optimal training powers are the same in both schemes; (iii) compared with the case when channels are known, the penalty caused by the channel estimation error becomes worse as the number of sensors increases.

The rest of this paper is organized as follows. Section 2 describes the system model. Section 3 derives results of the two-phase approach, namely the LMMSE estimation of channels and source. Section 4 formulates the optimal power allocation problem as an optimization problem. The problem is solved for two cases: when channels are known and when channels are estimated. Comparison of performance of two cases are given. Section 5 describes the equal power allocation scheme in which the training power is optimized. Performance analysis of the scheme is also given. In Section 6, simulation results are given to verify the analytical results obtained in Sections 4 and 5. Section 7 is a brief conclusion.

2. System model

We consider a wireless sensor network with $K$ sensors for estimating a random source signal $\theta$, as depicted in Fig. 1.

### 3. LMMSE estimation

#### 3.1. Channel estimation

During the training phase, the sensors send training symbols in sequence: the training period is divided into $K$ time intervals and only the $k$th sensor sends a training symbol $t_k$ over the $k$th time interval. Thus, the received signal at the $k$th time interval can be expressed as

$$y_k = h_k t_k + v_k, \quad k = 1, 2, \ldots, K,$$

where $v_k \sim \mathcal{CN}(0, \sigma_v^2)$ and $E[v_i^* v_j] = 0$ for $i \neq j$. For a given training sequence $t_k$, the LMMSE estimator of $h_k$ is given by [14, p. 382]

$$\hat{h}_k = \frac{\sigma_h^2}{|t_k|^2 \sigma_h^2 + \sigma_v^2} t_k^* y_k$$

Fig. 1. Coherent MAC wireless sensor network.
and the corresponding mean squared error (MSE) is
\[
\hat{\theta}_k = E[|\hat{h}_k - \hat{b}_k|^2] = \frac{\sigma_{\hat{h}_k}^2 \sigma_{\theta_k}^2}{\sigma_{\hat{h}_k}^2 + \sigma_{\theta_k}^2}, \quad k = 1, \ldots, K
\]
(3)
The MSE of \(\hat{b}_k\) decreases as the power of the training symbol \(|t_k|^2\) increases. The LMMSE problem under the training power constraint \(\sum_{k=1}^{K} |t_k|^2 \leq P_t\) can be formulated as
\[
\min \frac{1}{P_t} \sum_{k=1}^{K} \hat{b}_k^2 \quad \text{subject to} \quad \sum_{k=1}^{K} p_k \leq P_t \quad \text{and} \quad p_k \geq 0, \quad k = 1, \ldots, K
\]
where \(p_k = |t_k|^2\) is the training power of the kth sensor. The problem can be solved except as different \(\text{Karush–Kuhn–Tucker} (\text{KKT})\) condition and the solution is \(|t_k|^2 = P_t/K, \forall k, \) as expected since the channels are independent and identically distributed. In particular, we choose the training symbol to be real and positive, that is, \(t_k = \sqrt{P_t/K}\), and the resulting channel estimate is
\[
\hat{h}_k = \frac{\sigma_{\hat{h}_k}^2 \sqrt{K} \sigma_{\theta_k}}{\sigma_{\hat{h}_k}^2 + K \sigma_{\theta_k}^2}, \quad k = 1, 2, \ldots, K
\]
(4)
with the corresponding MSE
\[
\text{\hat{\theta}}_k = \frac{K \sigma_{\hat{h}_k}^2 \sigma_{\theta_k}^2}{\sigma_{\hat{h}_k}^2 + K \sigma_{\theta_k}^2}, \quad k = 1, \ldots, K
\]
(5)
We note that with such choices of training symbols, both the received signal \(y_k\) and the channel estimate \(\hat{h}_k\) are circular Gaussian.

3.2. Source estimation

During the second phase, channel estimates \(\hat{h}_k\) are available at the FC, although the actual channels are unknown. We express the received signal \(y\) in (1) in terms of \(\hat{h}_k\) as
\[
y = \sum_{k=1}^{K} \hat{h}_k z_k \theta + \sum_{k=1}^{K} \hat{h}_k z_k n_k + \epsilon + v
\]
(6)
where \(\epsilon = \sum_{k=1}^{K-1} (\hat{h}_k - \hat{h}_k) \hat{h}_k (\theta + n_k)\) is contributed by channel estimation error. Let \(\hat{h} = [\hat{h}_1, \hat{h}_2, \ldots, \hat{h}_K]^T\) be the vector of channel estimates. The LMMSE estimate of \(\theta\) given \(\hat{h}\) is
\[
\hat{\theta} = \frac{a \text{E}[|\theta|^2 | \hat{h}] \hat{h} \text{E}[|y|^2 | \hat{h}]}{a \text{E}[|y|^2 | \hat{h}]^{-1}}
\]
(7)
From (6) it follows that
\[
E[|\theta|^2 | \hat{h}] = E\left[ \theta \left( \sum_{k=1}^{K} \hat{h}_k^* z_k^2 \theta^* + \sum_{k=1}^{K} \hat{h}_k^* z_k n_k^* + e^* + v^* \right) \right]
\]
(8)
where the last equality is from the assumptions that the source signal is uncorrelated with the measurement noise and the receiver noise, and that \(\hat{h}_k = E[\hat{h}_k | y_k] = E[\hat{h}_k | \hat{h}_k]\) since \(\hat{h}_k\) is a linear function of \(y_k\). It is derived in Appendix A that
\[
\text{E}[|y|^2 | \hat{h}] = \sum_{k=1}^{K} \hat{h}_k z_k^2 \sigma_{\theta_k}^2 + \sum_{k=1}^{K} |\hat{h}_k|^2 |z_k|^2 \sigma_{\theta_k}^2 + (\sigma_{\theta_k}^2 + \sigma_{\theta_k}^2) \sum_{k=1}^{K} |z_k|^2
\]
(9)

The MSE incurred by (7) is
\[
J = E[\theta - \hat{\theta} | \hat{h}] = \sigma_{\theta_k}^2 - \text{E}[|\theta|^2 | \hat{h}] + \sigma_{\theta_k}^2 + |\theta|^2 |\hat{h}^{-1} = \left( \frac{1}{\sigma_{\theta_k}^2} + \sum_{k=1}^{K} |\hat{h}_k|^2 |z_k|^2 \sigma_{\theta_k}^2 + (\sigma_{\theta_k}^2 + \sigma_{\theta_k}^2) \sum_{k=1}^{K} |z_k|^2 \right)^{-1}
\]
(10)
When the channel \(h_k\) is available at the FC, we can set \(\hat{h}_k = h_k\) and \(\text{\hat{\theta}}_k = 0\) in (10), and the corresponding MSE becomes
\[
J_o = \left( \frac{1}{\sigma_{\theta_k}^2} + \sum_{k=1}^{K} |h_k|^2 |z_k|^2 \sigma_{\theta_k}^2 + (\sigma_{\theta_k}^2 + \sigma_{\theta_k}^2) \sum_{k=1}^{K} |z_k|^2 \right)^{-1}
\]
(11)
The MSE \(J_o\) is a lower bound of \(J\) in (10) and can serve as a benchmark against which the performance of the estimator (7) can be compared.

4. Optimal power allocation

During the training phase, each sensor uses the same training symbol and thus consumes the same amount of training power \(P_t/K\), where \(P_t\) is the total allocated training power. From (5), it is clear that as \(P_t\) increases, the MSE in channel estimation decreases. In a sensor network, there is likely a total power constraint, that is, there is an upper bound imposed on the sum of training power and the power used to transmit data. Hence, when more power is allocated for training, less power is available for data transmission and vice versa. Under the total power constraint, the minimum MSE of \(\theta\), that is, \(J\) in (10), depends on the training power \(P_t\) and how the remaining network power is allocated to each sensor for data transmission. In the following, we consider the optimal power allocation problem, that is, to choose \(P_t\) and data power for each sensor to minimize \(J\) under a total power constraint. For comparison, we will also consider the case when channel information is available, no training, no channel error, and all power is used for data transmission. The comparison of the two cases will show the penalty incurred due to the fact that the channel is unknown.

4.1. When channels are known

If the channels are known at the FC, the phase of \(z_k\) is chosen as \(\angle z_k = -\angle h_k\), so that \(h_k z_k = |h_k||z_k|\) and the MSE \(J_o\) in (11) becomes
\[
J_o = \left( \frac{1}{\sigma_{\theta_k}^2} + \frac{\zeta}{\sigma_{\theta_k}^2 \sum_{k=1}^{K} |g_k|^2 (|z_k|^2)} + 1 \right)^{-1}
\]
(12)
where \(\zeta = \sigma_{\theta_k}^2 / \sigma_{\theta_k}^2\) is the channel SNR, and \(g_k = |h_k| / \sigma_{\theta_k}\) is the normalized channel gain for the kth sensor. Such choices of phases make \(J_o\) smallest among \(z_k\)'s of the same magnitude. Note that \(g_k\) has a Rayleigh distribution with density function \(f_{g_k}(x) = 2x \exp(-x^2), x \geq 0\), and \(E[g_k^2] = 1 / \pi^2 / 4\); \(g_k\) has an exponential distribution with density function \(f_{g_k}(x) = \exp(-x), x \geq 0\), and \(E[g_k^2] = [1, p. 51]\). The signal transmitted from the kth sensor is \(z_k(\theta + n_k)\) with power
From (13), we make the following observations:

(i) If the inequality sign in the constraint of problem (13) is replaced by the equality sign, the solution does not change. Hence we could consider the optimization problem with equality constraint. The argument is as follows. Since the constraint function is quadratic in \( |z_k| \), if a set of \( |z_k| \) is such that strict inequality holds, we can equally scale up each \( |z_k| \), so that equality holds. And if we equally scale up each \( |z_k| \), we get a lower function value of \( J_0 \) because in (12) the second term inside the parentheses becomes larger. Consequently, with optimal \( |z_k| \), the inequality constraint must be active.

(ii) Consider the optimal MSE in (13), say, \( J_0 \) as a function of the power \( P \), then \( J_0 \) is a strictly decreasing function of \( P \), that is, if \( P_2 > P_1 \), then \( J_0(P_2) < J_0(P_1) \). The argument is similar: if the power level increases, we can equally scale up \( |z_k| \) to obtain a lower value of \( J_0 \) and thus a lower value of optimal MSE \( J_0 \) can be obtained.

(iii) Since the function of \( J_0(P) \) is one-to-one and decreasing, the inverse function \( P(J_0) \) is also one-to-one and decreasing. Hence instead of finding \( |z_k| \) that minimize \( J_0 \) in (12) under an equality constraint on power level, we can find \( |z_k| \) that minimize the power level subject to an equality constraint on MSE. And if the constraint value on MSE is such that the resulting minimum power level matches the given value \( P \) in (13), the corresponding \( |z_k| \) are the optimal ones we set out to find. We thus consider the following optimization problem:

\[
\begin{align*}
\min_{|z_k|:1 \leq k \leq K} & \sum_{k=1}^{K} |z_k|^2 (\sigma_a^2 + \sigma_z^2) \\
\text{subject to} & \left( \frac{1}{\sigma_a^2} + \frac{\zeta}{\sigma_a^2} \sum_{k=1}^{K} \frac{g_k^2 |z_k|^2}{\sum_{k=1}^{K} g_k^2 |z_k|^2} + 1 \right)^{-1} = J_0 \\
& 0 < J_0 \leq \sigma_z^2.
\end{align*}
\]

where \( \zeta \) satisfies

\[
\sum_{k=1}^{K} \frac{g_k^2 (\sigma_a^2 + \sigma_z^2)}{\sigma_a^2 (\sigma_a^2 + \sigma_z^2) + \sigma_z^2 g_k^2 P} = \frac{1}{J_0} \frac{1}{\sigma_z^2}
\]

The multiplier \( \mu \) is the total network power since from (15) we have \( \sum_{k=1}^{K} |z_k|^2 (\sigma_a^2 + \sigma_z^2) = \mu \). It is also clear from (16) that there is a one-to-one correspondence between the total power \( \mu \) and the constraint \( J_0 \), and that \( \mu \) increases as \( J_0 \) decreases and vice versa. Hence the original problem (13) is solved if we choose \( J_0 \) so that \( \mu \) is equal to \( P \), and the choice is

\[
J_0 = \left( \frac{1}{\sigma_a^2} + \frac{\sum_{k=1}^{K} g_k^2}{\sigma_a^2 \sum_{k=1}^{K} g_k^2 + (\sigma_a^2 + \sigma_z^2) P} \right)^{-1}
\]

which is the minimum MSE of (13). Accordingly, the optimal power allocation is, for \( 1 \leq k \leq K \),

\[
P_k = |z_k|^2 (\sigma_a^2 + \sigma_z^2) = \left( \sum_{k=1}^{K} \frac{g_k^2 (\sigma_a^2 + \sigma_z^2)}{\sigma_a^2 (\sigma_a^2 + \sigma_z^2) + \sigma_z^2 g_k^2 P} \right)^{-1}
\]

and \( |z_k|^2 = P_k/(\sigma_a^2 + \sigma_z^2) \). Since the minimum MSE depends on the total network power \( P \) and the number of sensors \( K \), we hereafter write the MSE \( J_0 \) in (17) as \( J_0(P,K) \).

As the power \( P \) increases, we expect \( J_0 \) to decrease, which is easy to see from (17). For a fixed \( K \), as \( P \to \infty \), we have

\[
\lim_{P \to \infty} J_0(P,K) = \frac{\sigma_a^2}{1 + K \beta}
\]

where \( \beta = \sigma_a^2/\sigma_z^2 \) is the observation SNR. The limit dose not go to zero but is roughly proportional to \( 1/K \) as we would expect. On the other hand, for a fixed \( P > 0 \), as \( K \) increases, we have

\[
\lim_{K \to \infty} J_0(P,K) = \left( \frac{1}{\sigma_a^2} + \lim_{K \to \infty} \left[ \frac{g_k^2}{\sigma_z^2 (\sigma_a^2 + \sigma_z^2) + \sigma_z^2 g_k^2 P} \right] \right)^{-1}
\]

\[
= \lim_{K \to \infty} \frac{1}{K} \left[ \frac{\sum_{k=1}^{K} g_k^2 P}{\sum_{k=1}^{K} g_k^2 (\sigma_a^2 + \sigma_z^2) P} \right] = 0
\]

where in the first equality we used the law of large numbers [17]. From (20), we conclude that in the coherent MAC model, the MSE decreases in the order of \( 1/K \) as \( K \) goes to infinity even though the total network power \( P \) is finite. Similar conclusion for the unit variance case, \( \sigma_a^2 = \sigma_z^2 = 1 \), appeared in [7].

4.2. When channels are estimated

Suppose training for channel estimation consumes power \( P_t \), then the remaining power for data transmission is \( P - P_t \). The power allocation problem now is to optimally choose training power \( P_t \) and data power for each sensor. The phase of \( z_k \) is chosen to match that of \( h_k \), i.e., \( \angle z_k = -\angle h_k \). Write \( h_k = h_k + (h_k - h_k) \), since \( h_k \) and \( h_k - h_k \) are
uncorrelated we have \( \sigma_k^2 = \sigma^2 + \delta_k^2 \), where \( \delta_k^2 = E[|h_k - \hat{h}_k|^2] \).

Use (5) and \( \sigma_k^2 = \sigma^2 - \delta_k^2 \), we can express the MSE in (10) as

\[
J = \left( \frac{1}{\sigma_0^2} + \frac{\zeta^2 (\sum_{k=1}^{K} |\hat{g}_k|^2 |z_k|^2)^2 P_t}{\zeta^2 |\sum_{k=1}^{K} g_k^2 |z_k|^2 |z_k|^2 + \sum_{k=1}^{K} |z_k|^2 + K} \right)^{-1}
\]

(21)

where \( \hat{g}_k = |\hat{h}_k|/\sigma_k \) is the normalized estimated channel gain for the kth sensor. Since \( \hat{h}_k \) is circular Gaussian, \( \hat{g}_k \) and \( g_k \) have identical distribution. From (21), the MMSE optimization problem under a total network power constraint can be formulated as

\[
\begin{align*}
\min_{P_t, |z_k|^2: 1 \leq k \leq K} & \left( \frac{1}{\sigma_0^2} + \frac{\zeta^2 (\sum_{k=1}^{K} |\hat{g}_k|^2 |z_k|^2)^2 P_t}{\zeta^2 |\sum_{k=1}^{K} g_k^2 |z_k|^2 |z_k|^2 + \sum_{k=1}^{K} |z_k|^2 + K} \right)^{-1} \\
\text{subject to} & \quad \sum_{k=1}^{K} |z_k|^2 (\sigma_0^2 + \sigma_k^2) + P_t = P.
\end{align*}
\]

(22)

Again instead of solving problem (22) directly, we consider a problem in which the roles of objective function and constraint are interchanged. The solution to problem (22) is given in the following proposition, the proof of which is given in Appendix C.

**Proposition 1.** For \( K > 1 \), the solution to (22) gives the optimal training power

\[
p_{t_{op}} = \frac{K(\zeta P + 1) - \sqrt{K(\zeta P + 1)(\zeta P + K)}}{\zeta (K - 1)}
\]

(23)

where \( \zeta = \sigma_k^2/\sigma_0^2 \) is the channel SNR, and the associated optimal data power for the kth sensor is

\[
p_{k_{op}} = \left( \sum_{k=1}^{K} \hat{g}_k^2 / \phi_k^2 \right)^{-1} \frac{\hat{g}_k^2}{\phi_k^2} \left( P - np_{t_{op}} \right)
\]

(24)

where \( \phi_k = (\sigma_0^2 + \sigma_k^2) + \mu^2 \sigma_k^2 \hat{g}_k^2 P_{t_{op}} + \mu K (\sigma_0^2 + \sigma_k^2) \) \( K \) and \( \mu = P_{t_{op}} / (K + K(\zeta P - P_{t_{op}})) \). The incurred MSE is

\[
j(P,K) = \left( \frac{1}{\sigma_0^2} + \sum_{k=1}^{K} \frac{\hat{g}_k^2}{\sigma_0^2 |\hat{g}_k|^2 + (\sigma_0^2 + \sigma_k^2)} \left( \frac{K - 1}{\sqrt{K(\zeta P + 1) - \sqrt{\zeta P + K}}} \right)^2 \right)^{-1}
\]

(25)

Note that the optimal training power in (23) depends on the number of sensors \( K \), the channel SNR \( \zeta \), and the total network power \( P \). From (25), we see that the MSE decreases as the power \( P \) increases. For a fixed \( K \), as \( P \to \infty \), we obtain

\[
\lim_{P \to \infty} j(P,K) = \frac{\sigma_0^2}{1 + K \beta}
\]

(26)

which is the same as (19). This makes sense since \( P \to \infty \) implies \( p_{t_{op}} \to \infty \) and thus the MSE of channel estimation in (5) approaches to zero, that is, \( \hat{h}_k \to h_k \) as \( P \to \infty \) in the mean square sense. It is shown in Appendix D that, for a fixed \( P \),

\[
\lim_{K \to \infty} j(P,K) = \sigma_0^2 \left( 1 + \frac{\beta}{1 + \beta} (\sqrt{P + 1} - 1)^2 \right)^{-1}
\]

(27)

The MSE does not approach to zero. The reason is that the order of \( 1/K \) decrease in MSE in (20) is offset by the order of \( K \) increase in the power of the error term \( E[|\hat{g}_k|^2] \) in (A.3) in Appendix A. Therefore, in the presence of the channel estimation error, the MSE reaches a finite nonzero value as \( K \) goes to infinity.

### 4.3. Comparison of two cases

If the total network power and number of sensors are fixed, with estimated channel, the estimation performance is worse than when channel information is available due to the presence of channel estimation error. To quantitatively compare the two cases, we set the same MSE objective, use optimal power allocation for both cases, and determine the respective total network power that would be required. Suppose to achieve the selected MSE, total network power \( P^a \) is required when channel information is available and the required total network power is \( P^e \) when channels are estimated. The ratio \( P^e/P^a \) gives an indication of the penalty incurred by the consumption of training power and the presence of channel estimation error. A small ratio would imply a heavy penalty. But the MSE expressions in (17) and (25) are random variables, we instead derive the condition on \( P^a \) and \( P^e \) under which the distributions of MSEs are identical. This is possible due to the fact that the random variables \( g_k \) and \( \hat{g}_k \) have identical Rayleigh distribution. From (17) and (25), the distributions of MSE expressions are identical if the deterministic terms in the denominator are equal, that is,

\[
\left( \frac{K - 1}{\sqrt{K(\zeta P^a + 1) - \sqrt{\zeta P^a + K}}} \right)^2 = \frac{1}{\zeta P^e}
\]

(28)

Rearranging (28), we get

\[
P^e \left( \sqrt{K(1 + 1/(\zeta P^e)) - \sqrt{1 + K/(\zeta P^e)}} \right)^2 \frac{1}{K - 1}
\]

(29)

Note that the ratio in (29) is less than one, and for \( P^e \) large

\[
\frac{P^a}{P^e} \approx \frac{1}{(\sqrt{K + 1})^2}
\]

(30)

The ratio decreases as the number of sensors \( K \) increases. This means that the penalty caused by channel estimation becomes worse as the number of sensor increases.
5. Equal power allocation

The optimal power allocation scheme discussed in the previous section requires that the complex $z_k$ be computed based on channel estimate $h_k$ (or $h_k^*$) and sent to the $k$th sensor through feedback channel from the FC. If the gains are not computed and feedback, in order to reduce computations and save feedback bandwidth, a reasonable strategy is to allocate equal power for each sensor for data transmission. In the following, we study the performance of the equal power allocation scheme. We again consider two cases: (i) channels are known at the FC and (ii) channels are estimated. In the latter case, we consider the optimal choice of training power $p^*$ to achieve the smallest MSE. We compare performance of the two cases in terms of the power ratio $P^*/P^e$ as in the previous section.

5.1. When channels are known

We set the phase of $z_k$ as $\angle z_k = - \angle h_k$. This requires feedback of a real number from the FC. With equal power allocation, we have $|z_k|^2 = P/(K(\sigma_0^2 + \sigma_k^2))$, for $k = 1, \ldots , K$ and the MSE in (12) can be rewritten as

$$J_0(p, K) = \sigma_0^2 \left(1 + \frac{1}{1 + \beta} \left(\frac{1}{K} \sum_{k=1}^{K} g_k^2 - 1 \frac{g_k^2}{K} + 1 \frac{p}{K} \right)^2 \right)^{-1}$$

(31)

It is easy to see from (31) that $J_0$ decreases as $P$ increases. For a fixed $K$, as $P \to \infty$, we have

$$\lim_{P \to \infty} \frac{1}{J_0(p, K)} = \sigma_0^{-2} \left(1 + \frac{1}{1 + \beta} \left(\frac{1}{K} \sum_{k=1}^{K} g_k^2 - 1 \frac{g_k^2}{K} + 1 \frac{p}{K} \right)^2 \right) \leq \sigma_0^{-2} (1 + K\beta)$$

where the last inequality uses the Cauchy–Schwarz inequality and the equal sign holds if and only if $g_1 = \cdots = g_K$. Therefore, as $P \to \infty$, we have a MSE lower bound as follows:

$$\lim_{P \to \infty} J_0(p, K) \geq \frac{\sigma_0^2}{1 + K\beta}$$

(32)

Since equality holds in (19), we see that the performance of the equal power scheme is usually worse than that of the optimal power scheme as $P \to \infty$. On the other hand, for a fixed $P$, as $K \to \infty$, we have $(1/K) \sum_{k=1}^{K} g_k^2 = \mathbb{E}[g_k^2] = \sqrt{\pi}/4$ and $(1/K) \sum_{k=1}^{K} g_k^2 = \mathbb{E}[|g_k|^2] = 1$, thus (31) becomes

$$\lim_{K \to \infty} J_0(p, K) = \lim_{K \to \infty} \frac{\sigma_0^2}{K} \left(1 + \frac{1 + \beta}{1 + \beta} \left(\frac{1}{K} \sum_{k=1}^{K} g_k^2 - 1 \frac{g_k^2}{K} + 1 \frac{p}{K} \right)^2 \right)^{-1} = 0$$

(33)

Hence, the MSE decreases in the order of $1/K$ and approaches to zero as $K \to \infty$ even though the total power $P$ is finite. Similar conclusion appeared in [7] for the unit variance case.

5.2. When channels are estimated

If the power $P^e$ is used for channel estimation, the transmitted data power for the $k$th sensor is $P_k = (P - P^e)/K$, or equivalently, $|z_k|^2 = (P - P^e)/(K(\sigma_0^2 + \sigma_k^2))$. Again the phase of $z_k$ is chosen as $\angle z_k = - \angle h_k$ and the MSE derived from (21) is

$$J(P, K) = \sigma_0^2 \left(1 + \frac{1 + \beta}{1 + \beta} \left(\frac{1}{K} \sum_{k=1}^{K} g_k^2 - 1 \frac{g_k^2}{K} + 1 \frac{P^e}{K} \right)^2 \right)^{-1}$$

(34)

From (34), the optimization problem becomes to choose $P^e$ so that the MSE $J$ is minimum under the total network power constraint. From (34) the MMSE optimization problem can be formulated equivalently as

$$\begin{align*}
\min_{P^e} \quad & - \sigma_0^2 \left(1 + \frac{1 + \beta}{1 + \beta} \left(\frac{1}{K} \sum_{k=1}^{K} g_k^2 - 1 \frac{g_k^2}{K} + 1 \frac{P^e}{K} \right)^2 \right) \\
\text{subject to} \quad & 0 \leq P^e \leq P.
\end{align*}$$

(35)

It can be shown that the second derivative of the objective function in (35) with respect to $P^e$ is positive. Hence, the optimization problem (35) is convex since the objective function is convex and the constraint is linear. The following proposition gives the optimal training power and the corresponding MSE.

**Proposition 2.** For $K > 1$, the solution to (35) gives the optimal training power

$$p^\text{opt} = \frac{K(P+1) - \sqrt{K(P+1)(K+P)}}{K(K-1)}$$

(36)

where $\zeta = \frac{\sigma_0^2}{\sigma_k^2}$, and the incurred MSE

$$J(P, K) = \sigma_0^2 \left(1 + \frac{1 + \beta}{1 + \beta} \left(\frac{1}{K} \sum_{k=1}^{K} g_k^2 - 1 \frac{g_k^2}{K} + 1 \frac{K\zeta}{K-1} \right)^2 \right)^{-1}$$

(37)

**Proof.** Please see Appendix E.

Note that the optimal training powers for both the equal and optimal power allocation schemes are the same. From (37) with fixed $K$, as $P \to \infty$, we obtain

$$\lim_{P \to \infty} (1/J(P, K)) \leq \sigma_0^{-2} (1 + K\beta)$$

and thus

$$\lim_{P \to \infty} J(P, K) \geq \frac{\sigma_0^2}{1 + K\beta}$$

(38)

which is the same as (32). On the other hand, for a fixed $P$, when $K \to \infty$, we obtain

$$\lim_{K \to \infty} J(P, K) = \sigma_0^2 \left(1 + \frac{\pi}{4 \beta} \frac{K}{1 + \beta} \left(\frac{P}{K} \right)^1 \right)^{-1}$$

(39)

which is worse than (27). Note that the MSE in (39) also approaches a finite nonzero value as the number of sensors goes to infinity due to the same reason as stated in Section 4.2.
5.3. Comparison of two cases

To compare performance of the two cases, we set $P_a$ and $P_e$ respectively so that the MSE expressions in (31) and (37) have the same distribution as in Section 4.3. From (31) and (37), the distributions of the MSE are identical if the deterministic terms in the denominator are equal, that is,

$$\frac{K-1}{\sqrt{K(P_e+1) - \sqrt{P_e+K}}} = \frac{1}{\sqrt{P_e}}$$

This equation is the same as (28) and thus we have the same ratio of penalty incurred by the training power consumption and the channel estimation error as shown in (29) and (30).

6. Numerical results

In this section, we use a number of numerical simulations to verify the analytical results obtained in previous sections. All random parameters, $\theta$, $n_k$, $h_k$, and $v$, are set as zero-mean circular Gaussian. The parameter $\theta$ and the channel $h_k$ are assumed to have unit variance, that is, we set $\sigma_\theta^2 = \sigma_h^2 = 1$ (0 dB). The observation noise variance $\sigma_n^2 = 10$ dB and the receiver noise variance $\sigma_n^2 = 1$ dB, so that the observation SNR $\beta = \sigma_n^2/\sigma_n^2$ and the channel SNR $\zeta = \sigma_h^2/\sigma_n^2$ are 10 and 1 dB, respectively.

We first compute the average MSE of the optimal power allocation scheme. The average MSE is the average of $10^5$ independent runs. The theoretical MSE is given in (25), where only the normalized channel gains $\hat{g}_k$ are random. To obtain the simulation MSE, we use the LMMSE estimators in (2) and (7) with all random variables independently generated, and take the average MSE of $\hat{\theta}$. It is clear from Fig. 2 that the theoretical and simulation values of MSE are very close. For a fixed $P$, we see that the MSE decreases as the number of sensors $K$ increases and approaches to the lower bound (27). The results for $P=14$ and 17 dB show that the 3 dB difference in total power leads to about 3 dB difference in MSE for $K \geq 20$.

The comparison between theoretical and simulation average MSES for the equal power scheme is shown in Fig. 3, where the theoretical result averages the MSE in (37). Again the figure shows that the theoretical and simulation values are very close. As $K$ increases, the average MSE decreases and approaches to the lower bound in (39). The results for $P=14$ and 17 dB also show roughly 3 dB difference in MSE for $K \geq 20$.

Fig. 4 shows the comparison of MSES between the equal and optimal power schemes for a fixed $P=16$ dB. It shows that the optimal power scheme performs better than the equal power scheme. For $K \geq 20$, the difference in MSE between the two schemes approaches to a constant value 0.007 (approximately 20% difference), which is about the difference between the respective low bounds.

For comparison, we also simulate the two-phase approach proposed in [8] based on the orthogonal model, where the $k$th sensor transmits the measured signal $x_k(\theta+n_k)$ to the $k$th receiver through an unknown fading...
channel $h_k, k=1,...,K$. The $k$th receiver is corrupted by an additive noise $v_k \sim \mathcal{CN}(0,\sigma^2_v)$, where $\sigma^2_v = -1$ dB, and $E[v_i v_j] = 0$ for $i \neq j$; then $K$ received data is collected by the FC with the LMMSE fusion rule for estimating the signal. The coherent model is shown in (1), where the received signal is a linear combination of the $K$ transmitted data corrupted by a noise. Fig. 5 shows that with a fixed $P=17$ dB, the MSE of the orthogonal model exhibits a conspicuous degradation as $K > 40$, while the MSE of the coherent model approaches to a constant value. Also compared with the orthogonal model, the coherent model has a lower average MSE regardless of the number of sensors used. This is a consequence of using orthogonal model, which results in $K$ different receiver noise $v_k$ at the FC so that the increase of $K$ does not reduce the effect of receiver noise; while in the coherent model, only one receiver noise is generated at the FC, which leads to increased signal to noise ratio as $K$ increases. In the figure, we see that as $K$ increases, the MSE of the coherent model is less sensitive to the channel estimation error than that of the orthogonal model.

Fig. 6 shows the ratio of average MSE $E[J_o]/E[J]$ versus the ratio $P_d/P$ for the optimal power allocation scheme with a fixed $K=16$ sensors. In the figure, the curves corresponding to total network power $P=20$, 23 dB, 27 dB, and 30 dB, respectively. The curves all cross the horizontal lines $E[J_o]/E[J]=1$ at about 0.04 very close to the predicted $1/(\sqrt{K}+1)^2$ in (30). Fig. 7 shows the ratio of average MSE $E[J_o]/E[J]$ versus the ratio $P_d/P$ for the equal power scheme. The total network power is fixed at $P=30$ dB and the number of sensors $K=9$, 25, and 36. We see that the curve for $K=9$ crosses $E[J_o]/E[J]=1$ at about 0.06, the curve for $K=25$ at about 0.03, and the curve for $K=36$ at about 0.02. The curves show that the penalty caused by channel estimation becomes worse as the number of sensors $K$ increases.

7. Conclusion

We study distributed estimation with coherent multiple access channel model and MMSE fusion rule. We use a two-phase approach for channel and source signal estimations; in both phases, the MMSE criterion is used. We study optimal power allocation problem under a total network power constraint. We obtain expressions of optimal training power and optimal data power for each sensor and the resulting MSE as a function of total network power $P$ and the number of sensor $K$ when channel estimates are used to compute power gains $a_k$ and feedback to the sensors. For the equal power scheme, we obtain an expression for the optimal training power and the resulting MSE. In both schemes, the optimal training powers are equal. Our results show that with estimated channels, the MSEs approach to finite nonzero values as the number of sensors increases. We note that this is in contrast with the result obtained for orthogonal MAC model [8] which shows the MSE performance eventually deteriorates as the number of sensor...
increases. The MSE performance compared with the case when channels are known shows the penalty caused by channel estimation becomes worse as the number of sensors increases.

**Appendix A. Derivation of (9)**

We first show that given \( \hat{h}, \varepsilon \) is uncorrelated with \( \nu, \sum_{k=1}^{K} \hat{h}_k \nu_k \), and \( \sum_{k=1}^{K} \hat{h}_k \nu_k \). Since \( E[\nu^2] = 0 \), \( E[\nu \varepsilon] = 0 \), and \( E[\nu \varepsilon | h] = 0 \). We show that \( \varepsilon \) and \( \sum_{k=1}^{K} \hat{h}_k \nu_k \) are uncorrelated as follows:

\[
E \left[ \left( \sum_{k=1}^{K} \hat{h}_k \nu_k \right)^* \varepsilon \hat{h} \right] = E \left[ \left( \sum_{k=1}^{K} \hat{h}_k \nu_k \right)^* \left( E[\hat{h}^2(h_k - \hat{h}_k) | \nu_k^2] \right) \hat{h} \right] = \sum_{k=1}^{K} \hat{h}_k E(h_k - \hat{h}_k) | \nu_k^2 \sigma_n^2 = 0 \quad (A.1)
\]

where the first equality uses that \( E[\nu_k^2] = 0 \) and \( E[\nu_k^2 | h_k] = 0 \) for \( k \neq l \). The last equality follows because \( \hat{h}_k = E[h_k | y_k] = E[h_k | \hat{h}_k] \). Similarly,

\[
E \left[ \left( \sum_{k=1}^{K} \hat{h}_k \varepsilon_k \theta \right)^* \varepsilon \hat{h} \right] = E \left[ \left( \sum_{k=1}^{K} \hat{h}_k \varepsilon_k \theta \right)^* \left( E[\hat{h}^2(h_k - \hat{h}_k) | \varepsilon_k \theta_k^2] \right) \hat{h} \right] = \sum_{k=1}^{K} \hat{h}_k E(h_k - \hat{h}_k) | \varepsilon_k \theta_k^2 \sigma_n^2 = 0 \quad (A.2)
\]

Finally the conditional variance

\[
E[\varepsilon^2 | \hat{h}] = E \left[ \sum_{k=1}^{K} (h_k - \hat{h}_k)^*(h_k - \hat{h}_k) | \nu_k^2 \sigma_n^2 \right] = (\sigma_n^2 + \sigma_\nu^2) \sum_{k=1}^{K} E[(h_k - \hat{h}_k)^*(h_k - \hat{h}_k) | \hat{h}] | \nu_k^2 \sigma_n^2 = (\sigma_n^2 + \sigma_\nu^2) \sum_{k=1}^{K} |\nu_k^2 \sigma_n^2 | (A.3)
\]

where the last equality uses \( \delta_1 = \cdots = \delta_k \) in (5). By (A.1)to (A.3) and (6), equality (9) follows.

**Appendix B. Derivation of (15)**

We introduce a slack variable \( t = \sum_{k=1}^{K} g_k | \nu_k | \) and rewrite the problem (14) as

\[
\begin{align*}
\min_{\nu_k, t} & \sum_{k=1}^{K} | \nu_k |^2 (\sigma_n^2 + \sigma_\nu^2) \\
\text{subject to} & \sum_{k=1}^{K} g_k | \nu_k | - t = 0 \\
& 1 + \zeta \sigma_n^2 \left( \sum_{k=1}^{K} g_k^2 | \nu_k |^2 \right) = \left( \frac{1}{f_0} - \frac{1}{\sigma_\nu^2} \right)^{-1} \zeta t^2
\end{align*}
\]

The Lagrangian is

\[
L(\nu_k, t, \lambda, \mu) = \sum_{k=1}^{K} | \nu_k |^2 (\sigma_n^2 + \sigma_\nu^2) + \lambda \left( \sum_{k=1}^{K} g_k | \nu_k | - t \right) + \mu \left[ 1 + \zeta \sigma_n^2 \left( \sum_{k=1}^{K} g_k^2 | \nu_k |^2 \right) - \left( \frac{1}{f_0} - \frac{1}{\sigma_\nu^2} \right)^{-1} \zeta t^2 \right]
\]

where \( \lambda, \mu \in \mathbb{R} \), and the associated necessary conditions [16] for optimality are

\[
\begin{align*}
\frac{\partial L}{\partial |\nu_k|^2} &= 2(\sigma_n^2 + \sigma_\nu^2) | \nu_k | + \lambda g_k + 2 \mu \zeta \sigma_n^2 g_k^2 | \nu_k | = 0 \\
& \quad (B.1) \\
\frac{\partial L}{\partial t} &= -\lambda - 2 \mu \left( \frac{1}{f_0} - \frac{1}{\sigma_\nu^2} \right)^{-1} \zeta t = 0 \\
& \quad (B.2) \\
\frac{\partial L}{\partial \nu_k} &= \sum_{k=1}^{K} g_k | \nu_k | - t = 0 \\
& \quad (B.3) \\
\frac{\partial L}{\partial \mu} &= 1 + \zeta \sigma_n^2 \sum_{k=1}^{K} g_k^2 | \nu_k |^2 - \left( \frac{1}{f_0} - \frac{1}{\sigma_\nu^2} \right)^{-1} \zeta t^2 = 0 \\
& \quad (B.4)
\end{align*}
\]

**Acknowledgment**

We thank the reviewers, their comments improve the paper.
From (B.1), \(|x_k| = -\frac{\dot{\phi}_k}{2\phi_k}\), where \(\phi_k = (\sigma_0^2 + \sigma_n^2) + \mu \sigma_n^2 g_k^2\), and thus from (B.3), we have \(t = \sum_{k=1}^{K} -\frac{g_k^2}{2\phi_k}\) and then from (B.2), we have

\[
\frac{1}{J_0} - \frac{1}{\sigma_0^2} = \sum_{k=1}^{K} \frac{\mu \zeta g_k^2}{\phi_k} = \sum_{k=1}^{K} \frac{\zeta g_k^2}{\sigma_0^2 + \sigma_n^2 + \mu \sigma_n^2 g_k^2} \tag{B.5}
\]

Hence, we obtain (16). Finally use \(t = -\frac{\dot{\lambda}}{1/J_0 - 1/\sigma_0^2}/(2\mu \zeta)\) from (B.2) and \(|x_k| = -\frac{\dot{\phi}_k}{2\phi_k}\) in (B.4) to get

\[
\frac{\dot{\lambda}^2}{4} = \left( \frac{1}{J_0} - \frac{1}{\sigma_0^2} - \frac{\zeta \sigma_n^2 g_k^2}{\phi_k} \right)^{-1} = \left( \sum_{k=1}^{K} \frac{\sigma_0^2 + \sigma_n^2}{\phi_k} \right)^{-1} \mu \tag{B.6}
\]

where the last equality follows from (B.5). Therefore, we have \(|x_k|^2 = (\dot{\lambda}^2/4)(g_k^2/\phi_k^2)\) and (15) is established.

**Appendix C. Proof of Proposition 1**

Instead of solving (22) directly, we consider the following problem:

\[
\begin{align*}
\min_{\gamma, \eta} & \quad (\sigma_0^2 + \sigma_n^2) \sum_{k=1}^{K} |x_k|^2 + P_t \\
\text{subject to} & \quad \zeta \left( \sum_{k=1}^{K} g_k |x_k|^2 \right)^2 P_t + \zeta \left( \sum_{k=1}^{K} |x_k|^2 \right) + K = \left( \frac{1}{J_0} - \frac{1}{\sigma_0^2} \right)^{-1} \zeta^2 t^2 P_t
\end{align*}
\]

The Lagrangian is

\[
L(|x_k|, P_t, \lambda, \mu)T = (\sigma_0^2 + \sigma_n^2) \sum_{k=1}^{K} |x_k|^2 + P_t + \lambda \left( \sum_{k=1}^{K} g_k |x_k|^2 - t \right) + \mu \left[ \zeta \left( \sum_{k=1}^{K} g_k |x_k|^2 \right)^2 P_t + \zeta \left( \sum_{k=1}^{K} |x_k|^2 \right) + K \right] - \left( \frac{1}{J_0} - \frac{1}{\sigma_0^2} \right)^{-1} \zeta^2 t^2 P_t
\]

where \(\lambda, \mu \in \mathbb{R}\), and the associated necessary conditions for optimality are

\[
\frac{\partial L}{\partial |x_k|} = 2(\sigma_0^2 + \sigma_n^2)|x_k| + \lambda g_k + \mu (2\zeta^2 \sigma_0^2 g_k^2 P_t |x_k| + 2K \zeta (\sigma_0^2 + \sigma_n^2) |x_k|) = 0 \tag{C.1}
\]

\[
\frac{\partial L}{\partial P_t} = 1 + \mu \left[ \zeta^2 \sigma_0^2 \left( \sum_{k=1}^{K} g_k^2 |x_k|^2 \right)^2 + \zeta \left( \frac{1}{J_0} - \frac{1}{\sigma_0^2} \right)^{-1} \zeta^2 t^2 \right] = 0 \tag{C.2}
\]

\[
\frac{\partial L}{\partial t} = -\dot{\lambda} - 2\mu \left( \frac{1}{J_0} - \frac{1}{\sigma_0^2} \right)^{-1} \zeta^2 t^2 P_t = 0 \tag{C.3}
\]

\[
\frac{\partial L}{\partial \zeta} = \sum_{k=1}^{K} \hat{g}_k |x_k| - t = 0 \tag{C.4}
\]

\[
\frac{\partial L}{\partial \mu} = \zeta^2 \sigma_0^2 \left( \sum_{k=1}^{K} g_k^2 |x_k|^2 \right) P_t + \zeta P_t + K \zeta (\sigma_0^2 + \sigma_n^2) \left( \sum_{k=1}^{K} |x_k|^2 \right) + K - \left( \frac{1}{J_0} - \frac{1}{\sigma_0^2} \right)^{-1} \zeta^2 t^2 P_t = 0 \tag{C.5}
\]

From (C.1), \(|x_k| = -\frac{\dot{\phi}_k}{2\phi_k}\), where \(\phi_k = (\sigma_0^2 + \sigma_n^2) + \mu \sigma_n^2 g_k^2\), thus it follows from (C.4) that \(t = -\frac{\dot{\lambda}}{2} \sum_{k=1}^{K} (\hat{g}_k^2/\phi_k)\) and then from (C.3), we have

\[
\frac{1}{J_0} - \frac{1}{\sigma_0^2} = \mu \zeta^2 P_t \sum_{k=1}^{K} \frac{\hat{g}_k^2}{\phi_k} = \zeta^2 P_t \sum_{k=1}^{K} \frac{\hat{g}_k^2}{(\sigma_0^2 + \sigma_n^2)/\mu + \zeta^2 \sigma_0^2 g_k^2 P_t + K \zeta (\sigma_0^2 + \sigma_n^2)} \tag{C.6}
\]
Use $t = -\lambda(1/J - 1/\sigma_n^2)/(2\mu^2P_t)$ from (C.3) and $|z_k| = -\lambda \hat{g}_k/(2\hat{\phi}_k)$ in (C.2) to get

$$
\frac{\lambda^2}{4} = \frac{1}{\mu^2\sigma_n^2 P_t^2} \sum_{k=1}^K \hat{g}_k^4/\hat{\phi}_k^2 = \frac{(1 + \mu \zeta)/(1 + K \mu \zeta)}{\sum_{k=1}^K \hat{g}_k^2(\sigma_n^2 + \sigma_n^2)/\hat{\phi}_k^2}
$$

where the last equality follows from (C.6). Since the data power for the $k$th sensor is $P_k = \sigma_n^2(\sigma_n^2 + \sigma_n^2)/\hat{\phi}_k^2 = (\lambda^2/4)\sum_{k=1}^K \hat{g}_k^2(\sigma_n^2 + \sigma_n^2)/\hat{\phi}_k^2 = (1 + \mu \zeta)P_t/(1 + K \mu \zeta)$. With the total network power constraint $P_t$, it follows from (C.2) and (C.5) that

$$
\mu = \frac{P_t}{K + \mu \zeta(P - P_t)}
$$

where we use $\sum_{k=1}^K P_k = P - P_t$. Moreover, since $\sum_{k=1}^K P_k + P_t = P$, we have

$$
\frac{2 + (K + 1) \mu}{K + \mu \zeta} P_t = P
$$

Substituting (C.8) into (C.9), we get the optimal training power in (23). With $P_t^{opt}$ and $\mu$, we get $P_k^{opt}$ in (24) and the MSE in (25) follows from (C.6).

**Appendix D. Derivation of (27)**

Rewrite (25) as

$$
J(P, K) = \left( \frac{1}{\sigma_n^2} + \frac{1}{\sigma_n^2 + \sigma_n^2} b(K) \sum_{k=1}^K \frac{\hat{g}_k^2}{b(K)\hat{g}_k^2 + K} \right)^{-1}
$$

where $b(K) = [\sqrt{\gamma}P + 1 - \sqrt{1 + \gamma}P/K]/(K - 1)^2$ and $\gamma = \sigma_n^2/(\sigma_n^2 + \sigma_n^2)$. Note that $\lim_{K \to \infty} b(K) = (\sqrt{\gamma}P + 1)^2$. We will show that the sum inside the parentheses converges to $E[\hat{g}_k^2] = 1$ as $K \to \infty$. Since

$$
\frac{\hat{g}_k^2}{K} - \frac{\hat{g}_k^2}{\gamma b(K)\hat{g}_k^2 + K} = \frac{\gamma b(K)\hat{g}_k^2}{K^2} \leq \frac{\gamma b(K)\hat{g}_k^2}{K^2}
$$

we have $\hat{g}_k^2/K - \gamma b(K)\hat{g}_k^2/K \leq \hat{g}_k^2/K - \gamma b(K)\hat{g}_k^2/K \leq \hat{g}_k^2/K$ and thus

$$
\sum_{k=1}^K \frac{\hat{g}_k^2}{K} - \sum_{k=1}^K \frac{\gamma b(K)\hat{g}_k^2}{K^2} \leq \sum_{k=1}^K \frac{\gamma b(K)\hat{g}_k^2}{K^2} \leq \sum_{k=1}^K \frac{\hat{g}_k^2}{K}
$$

It follows from the law of large numbers that as $K \to \infty$, we have $\sum_{k=1}^K \hat{g}_k^2/K = E[\hat{g}_k^2] = 1$.

$$
\sum_{k=1}^K \gamma b(K)\hat{g}_k^2/K = \gamma(\sqrt{\gamma}P + 1)^2 E[\hat{g}_k^2] \quad \text{and} \quad \sum_{k=1}^K \frac{\gamma b(K)\hat{g}_k^2}{K^2} = 0
$$

because $E[\hat{g}_k^2]$ is finite. Therefore,

$$
\sum_{k=1}^K \gamma b(K)\hat{g}_k^2/K = 1, \quad \text{as} \quad K \to \infty
$$

and (27) follows.

**Appendix E. Proof of Proposition 2**

Let

$$
c_1 = K \frac{\beta}{1 + \beta} \left( \frac{1}{K} \sum_{k=1}^K \hat{g}_k \right)^2
$$

and

$$
c_2 = \frac{1}{1 + \beta} \left( \frac{1}{K} \sum_{k=1}^K \hat{g}_k \right)^2
$$
then the Lagrangian of the problem (35) is

\[ L(P_1, \mu_1, \mu_2) = -\frac{\zeta^2 c_1 (P - P_1) P_1}{\zeta^2 c_2 (P - P_1) P_1 + \zeta P_1 + \zeta (K - P_1) + K} + \mu_1 (P_1 - P) - \mu_2 P_1 \]

and the associated KKT conditions are

\[ \frac{\zeta^2 c_1 (K - P_1) P_1^2 - 2K (\zeta + 1) P_1 + \zeta K P_1}{(\zeta^2 c_2 (P - P_1) P_1 + \zeta P_1 + \zeta (K - P_1) + K)^2} + \mu_1 - \mu_2 = 0 \]  
(E.1)

\[ \mu_1 (P_1 - P) = 0, \quad \mu_1 \geq 0 \]  
(E.2)

\[ \mu_2 P_1 = 0, \quad \mu_2 \geq 0 \]  
(E.3)

Since the training power have to be greater than 0, we have \( \mu_2 = 0 \). If \( \mu_1 > 0 \), then \( P_1 = P \), but then (E.1) leads to \( \mu_1 < 0 \) a contradiction. Therefore, we have \( \mu_1 = \mu_2 = 0 \) and \( P > P_1 > 0 \). From (E.1), we have

\[ \zeta (K - 1) P_1^2 - 2K (\zeta + 1) P_1 + \zeta K P_1 = 0 \Rightarrow P_1 = \frac{K (\zeta + 1) \pm \sqrt{K (\zeta + 1) (\zeta K + 1)}}{\zeta (K - 1)} \]

where we take negative term since positive term cannot satisfy the constraint \( P \geq P^* \). Let \( a = K (\zeta + 1) \) and \( b = \zeta K + 1 \), then we have \( P_1^{\text{opt}} = (a - \sqrt{a b}) / (\zeta (K - 1)) \) and \( P - P_1^{\text{opt}} = (b - \sqrt{a b}) / (\zeta (K - 1)) \), and from (34), (37) follows:

\[ J(P, K) = \sigma_0^2 \left( 1 + \frac{c_1 (c_2 + \sqrt{a b - 2 a b})}{c_2 (c_2 + (\sqrt{a b - 2 a b}) + (K - 1)^2 \sqrt{a b})} \right)^{-1} = \sigma_0^2 \left( 1 + \frac{c_1}{c_2 + (K - 1)^2 \sqrt{a b}} \right)^{-1} \]

where the first equality uses that

\[ (K - 1)(a - \sqrt{a b}) + K (K - 1)(\sqrt{a b - b}) + K (K - 1)^2 = (K - 1)(a - \sqrt{a b}) + K (K - 1)^2 - K \]

\[ = 0 \]

References


