ASYMPTOTICALLY OPTIMAL POWER-CONSTRAINED DISTRIBUTED ESTIMATION

M. Guerriero, P. Willett
University of Connecticut
ECE Department
Storrs, CT 06269 USA

S. Marano, V. Matta
Università degli Studi di Salerno
DIIIE
Via Ponte don Melillo I-84084, Fisciano (SA), Italy

ABSTRACT
A random parameter is estimated by a distributed network of sensors that communicate over a common MAC. The channel implies an enforced additive fusion rule, and the goal here is to design a power-constrained forwarding strategy and the post-processing by the fusion center. To get an explicit solution we appeal to asymptotics, meaning that we design the locally optimal scheme for the limiting case that the received power goes to zero.

Index Terms— Distributed Estimation, Nonlinearity, Sensor Networks, MAC, Power Constraint.

1. INTRODUCTION & BACKGROUND
Reliable data delivery from several remote sensors sharing a common transmission medium can be realized by employing source and channel coding strategies borrowed from (or extending) classical point-to-point results. This approach decouples the source and channel coding stages; however, it is known that this separation is not necessarily optimal [1] when the sources are correlated and/or one’s goal is not direct recovery of the observations.

A basic lesson from [2] is that there exist cases in which a simple amplify-and-forward strategy outperforms the best separate scheme by orders of magnitude. This happens, for instance, for Gaussian estimation problems over a power-constrained Gaussian multiple access channel [3], and it is due to the perfect match between the (additive) nature of the optimal MMSE estimator and the (additive) channel structure. In this work, we depart from the Gaussian model, and allow the local encoders to apply a nonlinear transformation to arbitrarily distributed observations (see Fig. 1).

There is joint consideration of the estimation problem and the noisy MAC in [4, 5]: in the asymptote of an increasingly large number of sensors, an optimal communication/estimation scheme (called Type Based Multiple Access, or TBMA) is proposed. A Likelihood Based Multiple Access (LBMA) scheme, suitable for continuous observations, is in [6], and yields asymptotically efficient estimation over a waveform channel.

Especially relevant to our work is the approach pursued in [7], where the authors consider the problem of finding the best relaying strategy using the received (generalized) signal-to-noise ratio as performance metric. We elaborate on [7], and focus on an asymptotic formulation of the communication/estimation problem where, different from [4–6], the limit is in terms of vanishingly small received power.

2. PROBLEM FORMULATION
With reference to Fig. 1, we consider a Wireless Sensor Network (WSN) engaged in the task of estimating a random parameter $\Theta$ from the output $Y$ of an additive Multiple Access Channel (MAC), not necessarily a Gaussian one. We assume that the $i^{th}$ observation $X_i$ is sent over the MAC after being transformed by a zero-memory nonlinearity $g(\cdot)$, and the focus is the design of the best nonlinearity yielding the minimum mean-square error (MMSE) in the low-power regime.

We denote by capital letters the random variables and by bold symbols the vectors. The problem can be formalized as that of estimating a scalar random parameter $\Theta$ from an $N$-vector $X = [X_1, X_2, \ldots, X_N]$, with the interpretation that the entries of the vector are the observations collected at the $N$ sensors of the network. The dependence of $X$ upon the unknown $\Theta$ is encoded in the conditional probability density function (pdf) $f_{X|\Theta}(x|\theta)$, while the a priori knowledge about

...
the parameter is summarized in the pdf \( f_\Theta(\theta) \); these pdfs specify the joint density of the remote nodes’ data \( X \), denoted by \( f_X(x) \). We assume that the observations are exchangeable (a deeper statistical concept than we need, but here meaning that their pdf is invariant to a permutation of them) and share the same first-order pdf \( f_X(x) \).

Sensors do not communicate each other. The \( i^{th} \) sensor simply forwards \( g(X_i) \) to the MAC, where \( g(\cdot) \) is the subject of this paper. We assume perfect synchronization; the MAC output is accordingly

\[
Y = \alpha \sum_{i=1}^{N} g(X_i) + W = \alpha G(X) + W, \quad (1)
\]

where \( W \) is a noise term with arbitrary, but known, pdf \( f_W(w) \). The factor \( \alpha \) scales the power, and we assume

\[
E[g^2(X)] = \int g^2(x) f_X(x) dx = 1.
\]

The overall aim is to estimate \( \Theta \) based on \( Y \). The optimization criterion we adopt is to minimize the estimation mean square error (MSE). Let \( \hat{\theta}_y(y) \) be the MMSE estimator based upon the channel output. From the Markov chain relationship \( \Theta \rightarrow X \rightarrow Y \) the optimal estimator is [8]

\[
\hat{\theta}_y(y) = E[\hat{\theta}_x(X)|y], \quad (2)
\]

the conditional expectation, given the observed data \( y \), of the optimal estimator \( \hat{\theta}_x(x) \) given the (unobserved) vector \( x \).

Now, the MSE pertaining to the optimal estimator \( \hat{\theta}_y(y) \) can be written as [8]

\[
\text{MSE} = E[(\hat{\theta}_y(Y) - \Theta)^2] = \text{VAR}[\Theta] - \text{VAR}[\hat{\theta}_y(Y)], \quad (3)
\]

so that the optimization problem of finding the nonlinearity \( g(\cdot) \) yielding the lowest MSE, can be recast as

\[
\min_{g: E[g^2(X)] = 1} \text{MSE} \leftrightarrow \max_{g: E[g^2(X)] = 1} \text{VAR}[\hat{\theta}_y(Y)]. \quad (4)
\]

It is relatively easy to solve the above (4) in the low-power regime \( \alpha \approx 0 \), as we now show.

### 3. SYSTEM OPTIMIZATION

#### 3.1. Finding the Nonlinearity

Considering that we are interested in the low-power case let us start by a first-order Taylor approximation of \( \hat{\theta}_y(y) \) about \( \alpha = 0 \). We have

\[
\hat{\theta}_y(y) = \int \hat{\theta}_x(x) f_X(x|y) dx
\]

\[
= \frac{\int \hat{\theta}_x(x) f_W(y - \alpha G(x)) f_X(x) dx}{\int f_W(y - \alpha G(x)) f_X(x) dx},
\]

From a first-order Taylor expansion we get

\[
\hat{\theta}_y(y) \approx E[\Theta] - \alpha \frac{\partial \ln f_W(y)}{\partial y} \times \int \left( \hat{\theta}_x(x) - E[\Theta] \right) g(x) f_X(x) dx. \quad (5)
\]

We now focus on the integral that depends upon the nonlinearity \( g(\cdot) \). \( G(x) \) is additive, so we have

\[
\int \left( \hat{\theta}_x(x) - E[\Theta] \right) G(x) f_X(x) dx
\]

\[
= \sum_{i=1}^{N} \left( \hat{\theta}_x(x) - E[\Theta] \right) g(x_i) f_X(x_i) dx. \quad (6)
\]

The same definition of the MMSE estimator \( \hat{\theta}_x(x) \) yields

\[
\int \hat{\theta}_x(x) g(x_i) f_X(x_i) dx = \int \hat{\theta}_x(x) g(x_i) f_X(x_i) dx_i, \quad (7)
\]

where \( \hat{\theta}_x(x) \) is the optimal MMSE estimator computed from one single observation. We thus have

\[
\int \left( \hat{\theta}_x(x) - E[\Theta] \right) g(x) f_X(x) dx
\]

\[
= N \int \left( \hat{\theta}_x(x) - E[\Theta] \right) g(x) f_X(x) dx. \quad (8)
\]

Substituting the above into (5), one finally gets

\[
\hat{\theta}_y(y) \approx E[\Theta] - \alpha N \frac{\partial \ln f_W(y)}{\partial y} \times \int \left( \hat{\theta}_x(x) - E[\Theta] \right) g(x) f_X(x) dx. \quad (9)
\]

We have made no assumption of conditional independence of \( X \); only exchangeability.

The above approximation can be inserted in the objective function \( \text{VAR}[\hat{\theta}_y(Y)] \) to be maximized, see (4). Then we can compute the expectation using \( f_Y(y) \approx f_Y(y)|_{\alpha=0} = f_W(y) \), so that

\[
\text{VAR}[\hat{\theta}_y(Y)] \approx \alpha^2 N^2 J(f_W)
\]

\[
\times \left( \int \left( \hat{\theta}_x(x) - E[\Theta] \right) g(x) f_X(x) dx \right)^2 \quad (10)
\]

where \( J(f) \) denotes the location Fisher information of the pdf \( f(x) \), see e.g., [1]. The sought optimal nonlinearity is thus

\[
\arg \max_{g: E[g^2(X)] = 1} \left( \int \left( \hat{\theta}_x(x) - E[\Theta] \right) g(x) f_X(x) dx \right)^2.
\]

(11)

Schwartz’s inequality implies that

\[
\left( \int \left( \hat{\theta}_x(x) - E[\Theta] \right) g(x) f_X(x) dx \right)^2
\]

\[
\leq E[g^2(X)] \text{VAR}[\hat{\theta}_x(X)] \quad (12)
\]

2983
implying that the solution to (11) can only be
\[ g(x) = \frac{\widehat{\theta}_x(x) - E[\Theta]}{\sqrt{\text{VAR}[\widehat{\theta}_x(X)]}} \]  \( (13) \)

### 3.2. Performance and Discussion

Let us comment on the resulting MSE. The term \( \text{VAR}[\Theta] \) in (3) can be regarded as the quadratic loss incurred by the a priori MMSE-estimator and therefore uses \( E[\Theta] \) as estimate of \( \Theta \). Accordingly, the difference \( \text{VAR}[\Theta] - \text{MSE} = \text{VAR}[\widehat{\theta}_y(Y)] \) represents the best gain (with respect to the blind estimator) that can be achieved starting from the \( y \)-data. We define the relative gain

\[ \eta_y = \frac{\text{VAR}[\widehat{\theta}_y(Y)]}{\text{VAR}[\Theta]} \in [0, 1], \]  \( (14) \)

such that \( \text{MSE} = \text{VAR}[\Theta](1 - \eta_y) \). The value of the optimal relative gain \( \eta_y \) is now obtained by using the \( g(x) \) provided in (13) into (10), yielding

\[ \eta_y = \alpha^2 N^2 J(f_W) \frac{\text{VAR}[\widehat{\theta}_x(X)]}{\text{VAR}[\Theta]} \]  \( (15) \)

It is useful to consider the estimation problem in which \( \Theta \) has to be estimated from \( x \), that is, in the case where all the noise-uncorrupted remote observations are fully available and no communication issues arise. In the following, we refer to the latter as the original (or communication-free) estimation problem, whose solution is, of course, straightforward. We can now recast (15) as

\[ \eta_y = \alpha^2 N^2 J(f_W) \eta_x, \]  \( (16) \)

where \( \eta_x \) is the relative gain corresponding to the original estimation problem with \( N = 1 \), which can be defined by replacing \( \widehat{\theta}_y(Y) \) with \( \widehat{\theta}_x(X) \) in (14).

Some comments are now in order.

- Equation (9) with the optimal \( g(x) \) becomes
  \[ \widehat{\theta}_y(y) \approx E[\Theta] - \alpha N \sqrt{\text{VAR}[\widehat{\theta}_x(X)]} \frac{\partial \ln f_W(y)}{\partial y} \]  \( (17) \)
  revealing that, in the low-power regime, the optimal estimator amounts to correction of the a priori estimator \( E[\Theta] \) by a term proportional to the score function of the channel-noise pdf, that is, \( \partial \ln f_W(y)/\partial y \).

- The optimal \( g(x) \), see (13), is simply (a scaled version of) the MMSE estimator using only the \( x \)-data.

- In [9] the optimal nonlinearity with a rate constraint instead of a power one has been derived, in the absence of any channel. It is found that the best shape is essentially governed by a (conditionally averaged version of)

\[ f_{X|\Theta}(x|\theta) = (1 - \epsilon)N(x; \theta, \sigma_1) + \epsilon N(x; \theta, \sigma_2), \]  \( (18) \)

\[ \eta_y = \frac{\text{VAR}[\widehat{\theta}_y(Y)]}{\text{VAR}[\Theta]} \in [0, 1], \]  \( (14) \)

![Fig. 2. Optimal nonlinearity for the model described in Sect. 4. The curves are drawn as prescribed by (13), for several values of \( \epsilon \) and parameters \( \sigma_1 = 1, \sigma_2 = 3 \).](image-url)

The performance increases quadratically with the number of sensors \( N \); the loss with respect to the communication-free scenario is governed by the channel Fisher information \( J(f_W) \).

- In the exact expression for the mean square error one has \( \eta_y \ll 1 \). This suggests that the low-power regime corresponds to

\[ \alpha^2 N^2 J(f_W) \eta_x \ll 1 \iff \alpha^2 \ll \frac{1}{N^2 J(f_W) \eta_x}. \]

This might be refined by retaining more Taylor terms.

### 4. Example

In this section we apply the previous results to the particular case that the observations are drawn according to the Gaussian mixture

\[ f_{X|\Theta}(x|\theta) = (1 - \epsilon)N(x; \theta, \sigma_1) + \epsilon N(x; \theta, \sigma_2), \]  \( (18) \)
where \( N(x; \mu, \sigma) \) stands for a Gaussian pdf with average value \( \mu \) and standard deviation \( \sigma \). The estimated parameter and noise

\[
 f_\theta(\theta) = N(\theta; 0, 1) \quad f_W(w) = N(w; 0, \sigma_w)
\]

are also assumed to be Gaussian.

Using (16) we get

\[
 \text{MSE} = 1 - \left( \frac{\alpha N}{\sigma_w} \right)^2 \eta_\epsilon(\epsilon).
\]

The system performances and the optimal shape for \( g(x) \) can be now drawn from the following closed-form expression for the one-sample estimator

\[
 \tilde{\theta}_x(x) = x \left( \frac{N(x; 0, \sqrt{1+\alpha^2})}{(1-\epsilon)N(x; 0, \sqrt{1+\sigma_w^2}) + \epsilon N(x; 0, \sqrt{1+\sigma_0^2})} \right)
\]

In Fig. 2, the optimal nonlinearities are displayed, for different estimation problems, i.e., for different values of \( \epsilon \). It is worth noting that the purely Gaussian cases, corresponding to \( \epsilon = 0, 1 \), yield a linear shape for \( g(x) \). This result has already been obtained in different contexts [9], [10].

Fig. 3 shows the mean square error as a function of the term \( \alpha N/\sigma_w \), parameterized by \( \epsilon \). The curve is expected to give a reasonable prediction of the best achievable MSE for low values of the ratio \( \alpha N/\sigma_w \): in the low-power region.

\section{5. CONCLUSIONS}

The results in [7] on decentralized estimation in which the inherent additive fusion rule of a MAC is exploited are particularly interesting, in that they show that the SNR-maximizing data-forwarding rule is proportional to the local MMSE, as opposed to the amplify-and-forward procedure that one might expect. Here that result is shown to apply directly to minimum fused-MSE estimation. Assuming vanishing received signal power we are able to show that the optimal sensor-level forwarding rule is proportional to the local MMSEE even for dependent sensors. We also provide the optimal fusion post-processing rule, which, unlike cases in which the asymptotics relate to arbitrarily large numbers of sensors, depends strongly on the channel noise statistics.

\section{6. REFERENCES}


