A Study of Hyperplane-Based Vector Quantization for Distributed Estimation

Jun Fang
Department of Electrical and Computer Engineering
Stevens Institute of Technology
Hoboken, NJ 07030
Email: jun.fang@stevens.edu

Hongbin Li
Department of Electrical and Computer Engineering
Stevens Institute of Technology
Hoboken, NJ 07030
Email: hongbin.li@stevens.edu

Abstract—We consider the problem of distributed estimation of a vector parameter in wireless sensor networks (WSNs). Due to stringent power and bandwidth constraints, vector quantization is performed at each sensor to convert its local noisy vector observation into one bit of information. The one bit quantized data is then sent to the fusion center (FC), where a final estimate of the vector parameter is formed. The vector quantization problem is studied in such a distributed estimation context. Specifically, our study focuses on a class of hyperplane-based vector quantizers which linearly convert the observation vector into a scalar by using a compression vector and then carry out a scalar quantization. Under the framework of the Cramér-Rao bound (CRB) analysis, we study the choice of the quantization thresholds and the design of the compression vectors.

Index Terms – Distributed estimation, hyperplane-based vector quantization, wireless sensor network (WSN).

I. INTRODUCTION

Distributed parameter estimation is a fundamental problem arising from sensor network applications. One of the most commonly used network settings for distributed estimation involves a set of spatially distributed sensors linked with a fusion center (FC). Each sensor makes a noisy observation of the phenomenon of interest and transmits its processed information to the FC, where a final estimate is formed. To address the stringent power and bandwidth constraints inherent in WSNs, the noisy observation at each sensor has to be quantized into one or a few bits of information. In this setup, quantization becomes an integral part of the estimation process and is critical to the estimation performance. The quantization design in such a distributed estimation context has been extensively studied in many works, e.g. [1]–[7]. So far most of previous studies focused on the scalar parameter case. For the general vector parameter scenario, the problem becomes much more complicated because, unlike the scalar case which is concerned about only the choice of the quantization threshold, vector quantization involves partitioning of a high dimensional space. Although important, such a problem has not received adequate attention as its scalar counterpart.

In this paper, we consider the problem of distributed quantization and estimation of a deterministic vector parameter, where the noisy vector observation of each sensor is quantized into only one bit of information. We study how to design the vector quantizer for each sensor, aimed at achieving the best estimation performance at the FC. Specifically, as in [7], we consider hyperplane-based vector quantization which utilizes a compression vector to convert the high-dimensional observation vector into a one-dimensional scalar, and then compares the resultant scalar with a quantization threshold to generate one bit quantized data. Our CRB analysis shows that the estimation performance is dependent on the quantization thresholds as well as the compression vectors. The optimal choice of the quantization thresholds, as in the scalar case, is dependent on the unknown parameter. In contrast, the design of the compression vectors is independent of the unknown parameter and hence is the focus of this paper. We develop an efficient iterative algorithm for the compression vector design for a general case and propose optimal/near-optimal solutions for some specific but important noise scenarios.

II. PROBLEM FORMULATION

Consider a WSN consisting of $N$ spatially distributed sensors. Each sensor makes a noisy observation of the unknown vector parameter $\theta \in \mathbb{R}^p$:

$$x_n = \theta + w_n, \quad n = 1, \ldots, N$$  \hspace{1cm} (1)

where $w_n \in \mathbb{R}^p$ denotes the additive multivariate Gaussian noise with zero mean and auto-covariance matrix $R_{w,n}$, and the noise is assumed independent (but not necessarily identically distributed) across sensors. To meet stringent bandwidth/power budgets in WSNs, we consider the case where each sensor quantizes its vector observation into one bit binary data $b_n$ which is sent to the FC to form an estimate of $\theta$. The problem of interest is to determine the vector quantizers for each sensor, and to develop a maximum likelihood (ML) estimator to estimate $\theta$ given $\{b_n\}_{n=1}^N$ for the FC.

Basically, vector quantization can be viewed as a space partitioning problem. For each sensor, the binary data $b_n$ is given by

$$b_n = 1\{x_n \in B_n\}$$  \hspace{1cm} (2)

where $b_n$ takes the value 1 when $x_n$ belongs to the region $B_n \subset \mathbb{R}^p$, and 0 otherwise. In this paper, to simplify the problem, the region $B_n$ is confined to be a half-space whose
border is a hyperplane defined by a compression vector $c_n$ and a quantization threshold $\tau_n$, i.e.

$$B_n = \{ x \in \mathbb{R}^p | c_n^T x > \tau_n \}$$  \hspace{1cm} (3)$$

The vector quantization problem therefore reduces to finding a set of compression vectors $\{ c_n \}$ and thresholds $\{ \tau_n \}$. Another important reason for us to consider half-space regions, as we will discuss later, is that the likelihood function of $\theta$ given $\{ b_n \}_{n=1}^N$ is concave when a hyperplane-based quantizer (3) is adopted. Thus the ML estimation of $\theta$ is a well-behaved numerical problem: any gradient-based search starting from a random initial estimate is guaranteed to converge to the global maximum, and many efficient routines exist for this type of work.

In the following, we will firstly develop the ML estimator and carry out a corresponding CRB analysis. The vector quantization design is then studied based on the CRB matrix of the unknown vector parameter $\theta$.

III. MLE AND CRB ANALYSIS

A. MLE

By combining (2) and (3), we have

$$b_n = \text{sgn}(c_n^T x_n - \tau_n), \hspace{1cm} n = 1, 2, \ldots, N$$  \hspace{1cm} (4)$$

where

$$\begin{cases} \text{sgn}(x) = 1, & \text{if } x > 0 \\ \text{sgn}(x) = 0, & \text{otherwise} \end{cases}$$

It can be readily shown that the probability mass function (PMF) of $b_n$ is given by

$$P(b_n; \theta) = [F_{v_n}(\tau_n - c_n^T \theta)]^{b_n} [1 - F_{v_n}(\tau_n - c_n^T \theta)]^{1-b_n}$$  \hspace{1cm} (5)$$

where $p_{v_n}(x)$ and $F_{v_n}(x)$ denote the probability density function (PDF) and the complementary cumulative density function (CCDF) of $v_n \triangleq c_n^T w_n$, respectively, and $v_n$ is a Gaussian random variable with zero mean and variance $\sigma_v^2 \triangleq c_n^T R_n c_n$. Since $\{ b_n \}$ are independent, the log-PMF or log-likelihood function is

$$L(\theta) \triangleq \log P(b_1, \ldots, b_N; \theta)$$

$$= \sum_{n=1}^N \left\{ b_n \log[F_{v_n}(\tau_n - c_n^T \theta)] + (1-b_n) \log[1 - F_{v_n}(\tau_n - c_n^T \theta)] \right\}$$  \hspace{1cm} (6)$$

The ML estimate of $\theta$, therefore, is given as

$$\hat{\theta} = \arg \max_{\theta} L(\theta)$$  \hspace{1cm} (7)$$

Although the ML estimation often suffers from drawbacks of high computational complexity and local maxima, this is not true for our case. In fact, it can be proved that the log-likelihood function $L(\theta)$ is a concave function. Hence computationally efficient search algorithms can be used to find the global maximum. The proof of the concavity is given in [8].

B. CRB

We carry out a CRB analysis of our proposed scheme. By defining

$$g_n(\tau_n, c_n) \triangleq \frac{p_{v_n}^2(\tau_n - c_n^T \theta)}{F_{v_n}(\tau_n - c_n^T \theta)(1 - F_{v_n}(\tau_n - c_n^T \theta))}$$  \hspace{1cm} (8)$$

the Fisher information matrix (FIM) can be re-expressed as (please see [8] for the details of the derivation)

$$J(\theta) = \sum_{n=1}^N g_n(\tau_n, c_n)c_n c_n^T$$  \hspace{1cm} (9)$$

and consequently the CRB matrix is

$$\text{CRB}(\theta) = J^{-1}(\theta) = \left( \sum_{n=1}^N g_n(\tau_n, c_n)c_n c_n^T \right)^{-1}$$  \hspace{1cm} (10)$$

The CRB places a lower bound on the estimation error of any unbiased estimator and is asymptotically attained by the ML estimator [9]. Specifically, the covariance matrix of any unbiased estimate $\hat{\theta}$ satisfies $\text{cov}(\hat{\theta}) - \text{CRB}(\theta) \succeq 0$. Also, the variance of each component is bounded by the corresponding diagonal element of $\text{CRB}(\theta)$, i.e. $\text{var}(\hat{\theta}) \succeq [\text{CRB}(\theta)]_{ii}$. It is observed from (10) that the CRB matrix of $\theta$ depends on the compression vectors $\{ c_n \}$ and the quantization thresholds $\{ \tau_n \}$. Naturally, we may wish to optimize $\{ c_n \}$ and $\{ \tau_n \}$ by minimizing the trace of the CRB matrix, i.e. the overall estimation error asymptotically achieved by the ML estimator. The optimization therefore is formulated as follows

$$\min_{\{ c_n \}, \{ \tau_n \}} \text{tr} \{ \text{CRB}(\theta) \} = \text{tr} \left\{ \left( \sum_{n=1}^N g_n(\tau_n, c_n)c_n c_n^T \right)^{-1} \right\}$$  \hspace{1cm} (11)$$

Such an optimization is examined in the following section, where it is shown that the optimization of the compression vectors $\{ c_n \}$ can be decoupled from the choice of the thresholds $\{ \tau_n \}$ and is the key to the vector quantization design.

IV. VECTOR QUANTIZATION DESIGN: THRESHOLD

DETERMINATION

Note that for the Gaussian random variable $v_n$, $g_n(\tau_n, c_n)$ defined in (8) is a unimodal, positive and symmetric function attaining its maximum when $\tau_n = c_n^T \theta$. Hence, given a fixed set of compression vectors $\{ c_n \}$, the optimal quantization thresholds conditional on $\{ c_n \}$ can be readily solved from (11) and are given by

$$\tau_n^* = c_n^T \theta, \hspace{1cm} \forall n \in \{ 1, \ldots, N \}$$  \hspace{1cm} (12)$$

The result (12) comes directly by noting that

$$\sum_{n=1}^N g_n(\tau_n^*, c_n)c_n c_n^T - \sum_{n=1}^N g_n(\tau_n, c_n)c_n c_n^T \succeq 0$$  \hspace{1cm} (13)$$

and resorting to the convexity of $\text{tr}(A^{-1})$ over the set of positive definite matrix, i.e. for any $A \succ 0$, $B \succ 0$, and $A - B \succeq 0$, the following inequality $\text{tr}(A^{-1}) \leq \text{tr}(B^{-1})$ holds (see [10]).
We see that, as in the scalar parameter case, the optimal choice of the quantization threshold \( \tau_n \) is dependent on the parameter \( \theta \). If \( \theta \) is perfectly known, then the optimization (11) simply reduces to finding a set of compression vectors \( \{ c_n \} \), with \( \tau_n = c_n^T \theta \), i.e.

\[
\min_{\{ c_n \}} \text{tr} \left\{ \sum_{n=1}^{N} g_n(\tau_n, c_n)c_n c_n^T \right\}^{-1} \\
\text{s.t. } \tau_n = c_n^T \theta, \quad \forall n
\] (14)

By noting that

\[
g_n(\tau_n^*, c_n) = \frac{\sigma_{\nu_n}^2(0)}{F_{\nu_n}(0)[1 - F_{\nu_n}(0)]} = \frac{2}{\pi c_n^T R_{w,k} c_n} = \frac{2}{\pi c_n^T R_{w,k} c_n}
\] (15)

(11) becomes an optimization independent of \( \theta \)

\[
\min_{\{ c_n \}} \text{tr} \left\{ \sum_{n=1}^{N} \frac{c_n c_n^T}{c_n^T R_{w,k} c_n} \right\}^{-1} \\
\] (16)

The problem lies in that the vector parameter \( \theta \), to be estimated, is unknown and unusable in practice. Hence the choice of the thresholds \( \{ \tau_n \} \) is tricky. One strategy is to use a FC feedback-based iterative algorithm [2], [6] in which the thresholds are iteratively refined by the FC based on the previous estimate. Specifically, at iteration \( i \), the FC assigns the quantization thresholds \( \{ \tau_n^{(i)} \}_{n=1}^{N} \) to the sensors. With this assigned quantization threshold, each sensor generates its quantized data \( b_n^{(i)} \) and reports back to the FC. Upon receiving the quantized data \( \{ b_n^{(i)} \}_{n=1}^{N} \), the FC can compute an estimate \( \hat{\theta}^{(i)} \) from the ML estimator (7). This ML estimate is then plugged in (12) to obtain updated quantization thresholds, i.e. \( \tau_n^{(i+1)} = c_n^T \hat{\theta}^{(i)} \), which are assigned to the sensors for subsequent iteration. Note that when computing an ML estimate \( \hat{\theta}^{(i)} \), not only the quantized data from the current but also from all previous iterations can be used (the ML estimator (7) can be easily adapted to accommodate these quantized data since the data are independent across different iterations). Due to the consistency of the ML estimator for large data records, this iterative process will asymptotically lead to conditional optimal quantization thresholds, i.e. \( \tau_n \xrightarrow{\text{t} \to \infty} c_n^T \theta \). A rigorous proof of this asymptotic optimality was provided in [6], where a similar feedback-based iterative algorithm was proposed to adjust the quantization thresholds. Given this asymptotic optimality, the problem of interest, therefore, is to determine the set of compression vectors by assuming the quantization threshold attaining their conditional optimal values \( c_n^T \theta \). Hence we are faced with the optimization (16).

V. VECTOR QUANTIZATION DESIGN: COMPRESSION VECTOR DESIGN

The optimization (16) involves determining \( N \) compression vectors. Clearly, joint searching over the \( N \) compression vectors is practically infeasible since it has a complexity that grows exponential with \( N \). To simplify the problem, we employ a Gauss-Seidel iterative technique to reduce the number of optimization variables. Specifically, we study how to determine the \( k \)th compression vector \( c_k \) when the remaining \( (N - 1) \) compression vectors are fixed, through which we can develop an efficient iterative algorithm to search for an effective, albeit suboptimal, solution. Let

\[
Q_k \triangleq \sum_{n \neq k} \frac{c_n c_n^T}{c_n^T R_{w,k} c_n}
\] (17)

The optimization of \( c_k \) given fixed \( \{ c_n \}_{n \neq k} \) is formulated as

\[
\min_{c_k} \text{tr} \left\{ \left( Q_k + \frac{c_k c_k^T}{c_k^T R_{w,k} c_k} \right)^{-1} \right\}
\] (18)

Recalling the Woodbury identity, the objective function of (18) can be rewritten as

\[
\text{tr} \left\{ \left( Q_k - \frac{c_k c_k^T}{c_k^T R_{w,k} c_k} \right)^{-1} \right\} = \text{tr} \left\{ Q_k^{-1} - \frac{c_k c_k^T Q_k^{-1} c_k^T}{c_k^T R_{w,k} c_k + c_k^T Q_k^{-1} c_k} \right\}
\]

\[
= \text{tr} \left\{ Q_k^{-1} - \frac{c_k^T Q_k^{-1} c_k}{c_k^T (R_{w,k} + Q_k^{-1}) c_k} \right\}
\] (19)

where \( (a) \) comes from the trace identity \( \text{tr}(AB) = \text{tr}(BA) \). Since \( Q_k \) is fixed, (18) becomes

\[
\max_{c_k} \frac{c_k^T Q_k^{-1} c_k}{c_k^T (R_{w,k} + Q_k^{-1}) c_k}
\] (20)

The above optimization is a generalized Rayleigh quotient problem and has a closed form solution which can be obtained as the eigenvector of \( \Gamma \) associated with the largest eigenvalue, where

\[
\Gamma \triangleq (R_{w,k} + Q_k^{-1})^{-\frac{1}{2}} Q_k^{-1} (R_{w,k} + Q_k^{-1})^{-\frac{1}{2}}
\] (21)

By using the above results, we can establish an iterative algorithm to solve (16) by successively optimizing and replacing each compression vector \( c_k \). The algorithm is summarized as follows

1) Randomly generate a set of compression vectors \( \{ c_n^{(0)} \} \) as an initialization.
2) At iteration \( i + 1 \) (\( i = 0, 1, \ldots \)): determine \( c_1^{(i+1)} \) given: \( \{ c_2^{(i)}, \ldots, c_N^{(i)} \} \); determine \( c_k^{(i+1)} \) given: \( \{ c_1^{(i)}, \ldots, c_{k-1}^{(i)}, c_k^{(i)}, c_{k+1}^{(i)}, \ldots, c_N^{(i)} \} \) for \( k = 2, \ldots, N \).
3) Go to Step 2 if \( |f(\{ c_n^{(i+1)} \}) - f(\{ c_n^{(i)} \})| < \epsilon \), where \( f(\cdot) \) denotes the objective function defined in (16), \( \epsilon \) is a prescribed tolerance value; otherwise stop.

Clearly, in this algorithm, every iteration results in a non-increasing objective function value. In this manner, the iterative algorithm converges to a stationary point and finds an effective set of compression vectors. Nevertheless, this algorithm is not guaranteed to converge to the global minimum, and it is unclear how close the achieved stationary point is to the global minimum. In [8], we show that near-optimal or
even optimal solutions can be obtained for some specific but important noise models that characterize most practical scenarios. These optimal/near-optimal solutions render an insight into the compression vector design and the theoretical analysis provides a fundamental understanding of the performance of our proposed one-bit quantization scheme. Due to the space limit, these results are not included in this paper.

VI. SIMULATION RESULTS

We now carry out experiments to illustrate the performance of our one-bit quantization scheme. We compare our scheme with the approach [7] and the clairvoyant estimator using unquantized data. In [7], each sensor employs \( p \) hyperplanes to quantize the vector observation \( x_n \) into \( p \) bits (instead of one bit in our scheme) of information, i.e. one bit per sensor per dimension. The \( p \) hyperplanes are chosen to be perpendicular to the eigenvectors of the noise covariance matrix in order to ensure that the resultant \( p \) binary data are independent.

In our simulations, the noise covariance matrix \( R_{w_n} \in \mathbb{R}^{p \times p} \) for each sensor is diagonal with its diagonal elements \( R_{w_n} = \alpha \nu_i \), where \( \alpha = 1 \) and \( \nu_i \sim \chi^2_1 \), and \( p \) is set to 5. The results are averaged over 500 Monte Carlo runs. Fig. 1 shows the overall estimation distortion (trace of the CRB matrix) of the three schemes as a function of the number of sensors, \( N \). From Fig. 1, we see that our proposed iterative algorithm is very effective. It achieves almost the same estimation performance as that of [7], while transmitting only \( 1/p \) times the total number of bits required by [7]. Also, both our proposed scheme and the scheme [7] incur a mild performance loss compared with the clairvoyant estimator using raw data. Note that in previous example, we assume the thresholds for our scheme and [7] are optimally chosen. Since \( \theta \) to be estimated is unknown in practice, it is interesting to examine the performance of our proposed threshold determination strategy: a FC feedback-based iterative algorithm (see Section IV) which iteratively adjusts the threshold based on the ML estimate. In our simulations, we assume a homogeneous scenario with \( R_{w_n} = 0.1I \). The dimension of the unknown parameter, \( p \), is set to 3 and the number of sensors varies from 100 to 200. Fig. 2 shows the estimation MSE of the ML estimator versus the number of iterations. From Fig. 2, we see that the FC feedback-based iterative algorithm is very effective and can achieve an accurate estimate of the unknown parameter within several iterations.

![Fig. 1. Overall estimation error vs. the number of sensors for three schemes.](image1)

![Fig. 2. MSE of FC-based feedback iterative algorithm for threshold determination.](image2)

REFERENCES


