2014 IEEE International Conference on Internet of Things (iThings 2014), Green Computing and Communications (GreenCom 2014), and Cyber-Physical-Social Computing (CPSCom 2014)

Quantization for Distributed Estimation

Kuan-Yu Chen, and Kwang-Cheng Chen, *Fellow ,IEEE*
Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan
Email: r01942121@ntu.edu.tw and chenkc@cc.ee.ntu.edu.tw

***Abstract*—Distributed Estimation scheme has been widely used in sensor networks. Through various approaches, *sum of bits, Sigma-Delta modulation*, and *source coding*, this paper implements the above techniques for distributed estimation and provides a new one called *Quantization Region Allocation*. In the end, this paper analyzes the performance (minimum mean squared error) and demonstrates performance approximating the soft decision scenario.**

***Index Terms*—distributed estimation, source coding, Sigma-Delta modulation**

I. Introduction

The problem we consider in this paper is associated with *Distributed Source Coding* (DSC) for multiple sources. DSC problems regard the compression of multiple correlated information sources that do not communicate with each other, as Fig. 1a shows. By designing the K encoders for different source and the joint decoder, DSC is able to shift the computational complexity from encoder side to decoder side, therefore provide appropriate frameworks for applications with complexity-constrained sender.

The lossless version of this problem with discrete sources was solved by Slepian and Wolf [1], using a random-binning approach. It is shown that for lossless scenarios, cooperation among encoders does not improve the sum-rate boundary. For $K = 2$, the achievable rate region R_1,R_2 for source Y_1,Y_2 is given by

$$\begin{aligned} R_1 &\geq H(Y_1|Y_2) \\ R_2 &\geq H(Y_2|Y_1) \\ R_1 + R_2 &\geq H(Y_1, Y_2) \end{aligned} \tag{1}$$

Practical distributed lossless coding schemes have been proposed [2], [3] that are close to the Slepian-Wolf bound. However, when lossless coding is not possible, rate distortion theory must be taken into account. This problem has been considered by Wyner and Ziv [4] for the case where the decoder has access to side information about the source.

When $Y_1, Y_2 \cdots, Y_K$ are influenced by a source X , K sensors observe independently corrupted versions of X. The sensors encode their observations without cooperating with one another. This is the so-called *CEO* problem and was first studied by Berger, Zhang, and Viswanathan [5] in the context of discrete memoryless sources and observations. The fusion center and K sensors are referred to as *CEO* and *agents* respectively, which is illustrated in Fig. 1b. Prabhakaran, Tse, Ramchandran [6] have shown that if X is zero mean and if

This work was also supported by National Science Council, National Taiwan University and Intel Corporation under Grants NSC102-2911-I-002-001 and NTU102R7501.

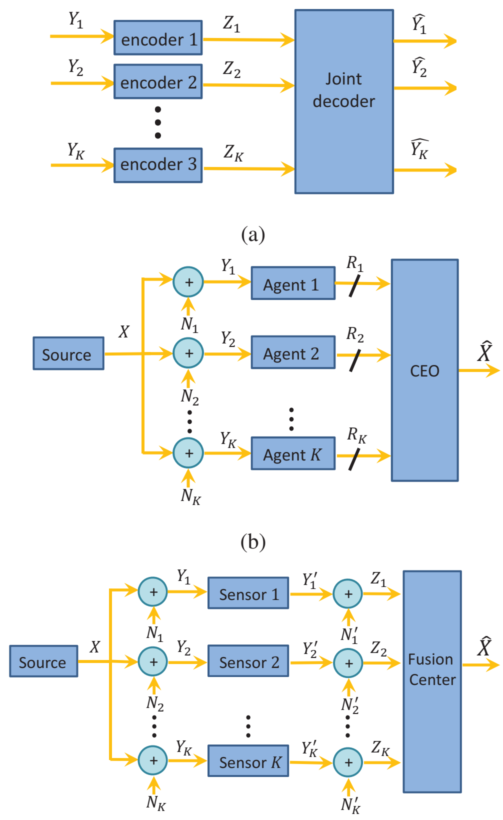


Fig. 1: System structure for (a) distributed source coding (b) *CEO* problem (c) quantized distributed estimation.

the transmission from agents to *CEO* is error-free, the K-tuple rate region $(R_1, R_2 \cdots, R_K)$ is given by

$$\mathcal{R}(D) = \bigcup_{(r_1, r_2 \cdots, r_K) \in \mathcal{F}(D)} \mathcal{R}_D(r_1, r_2 \cdots, r_K) \tag{2}$$

where $\mathcal{R}_D(r_1, r_2, \cdots r_K)$ and $\mathcal{F}(D)$ is given by (3) and (4) respectively.

The system structure for our problem is depicted in Fig. 1c. In the first approach, we simply use the number of 1's $(+A)$ and 0's $(-A)$ sent from K sensors to estimate the value of X, namely,

$$\widehat{X} = \mu + (K_1 - K_0)B \tag{5}$$

where K_1, K_0 denote the number of 1's and 0's respectively and B is a constant needed to be determined. This can give a satisfactory result. To improve the performance, *Sigma-Delta modulation* (Σ-Δ modulation) [7] is introduced. In addition,

$$\mathcal{R}_D(r_1, r_2 \cdots, r_K) = \Bigg\{ (R_1, R_2 \cdots, R_K) : \sum_{k \in A} R_k \geq \sum_{k \in A} r_k + \frac{1}{2} \log \frac{1}{D} - \frac{1}{2} \log \left(\frac{1}{\alpha^2} + \sum_{k \in A^c} \frac{1 - \exp(-2r_k)}{\sigma^2} \right), \forall \text{ non empty } A \subseteq \{1, \cdots, L\} \Bigg\} \quad (3)$$

$$\mathcal{F}(D) = \left\{ (r_1, r_2 \cdots, r_K) \in \mathbb{R}_+^K : \frac{1}{\alpha^2} + \sum_{k=1}^{K} \frac{1 - \exp(-2r_k)}{\sigma^2} = \frac{1}{D} \right\} \quad (4)$$

when it comes to quantization at each sensor, jointly optimize the quantization region would be considered. However, to simplify the computation, we combine Lloyd Algorithm with weighted bits, providing a new approach called *Quantization Region Allocation* in section IV.

II. Problem Statement

As Fig. 1c shows, K sensors is used to detect the source parameter X in the following way. Assume that the source X is Gaussian distributed with mean μ and variance α^2. The k-th sensor has noisy observation $Y_k = X + N_k$, where N_k is i.i.d. Gaussian with mean 0 and variance σ^2. In addition, the data sent from each sensor Y_k' also suffers from noise N_k', which is assumed i.i.d. and has same distribution as N_k. The fusion center finally determines $\widehat{X}$ based on $Z_k = Y_k' + N_k', k = 1, 2, \cdots K$. When each sensor noiselessly sends its observation to Fusion Center, there are two possible ways: to quantize (i.e. to determine between $\{A, -A\}$, also known as hard decision) or not (i.e. just to send observed value, also known as soft decision).

III. Basic Approach

A. *Sum of Bits*

Assume that each sensor sends A to the receiver if $Y_k > \mu$, and sends $(-A)$ otherwise. The fusion center first detect whether A or $(-A)$ is sent from each path. Define the Bernoulli random variables $\{I_i\}$, $i = 0, 1 \cdots, K$, where $I_i = 0$ if A is decided and $I_i = 1$ if $(-A)$ is decided. Our estimator is given by

$$\widehat{X} = \mu + \sum_{i=1}^{K} (-1)^{I_i} B = \mu + NB \quad (6)$$

Furthermore, if we define the random variable $M = \sum_{i=1}^{K} I_i$, then$M|X \sim B(K, p)$, where $p = \Pr(I_k = 1|X)$. Since $N = K - 2M$, the mean squared error can be written as

$$\begin{aligned} &E[(X - \mu - NB)^2] \\ =&E\big[\ E[(X - \mu - (K - 2M)B)^2 | X]\ \big] \end{aligned} \quad (7)$$

Differentiate (7) with respect to B yeild the optimal value

$$B = \frac{-2E[(X - \mu)p]}{(8 - 4K)E[p] - 4E[p^2] + K} \quad (8)$$

The result is shown in Fig. 2. Here we set $\mu = 1$ and $A = B$. For the simulation curves, different values of B has

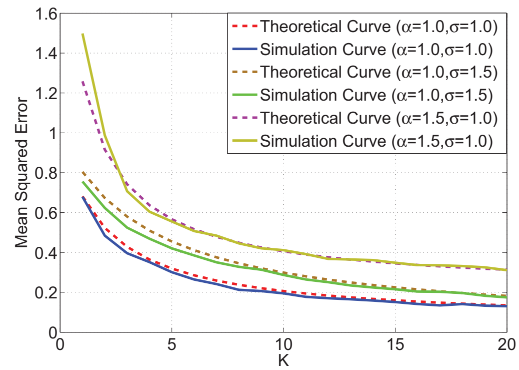


Fig. 2: MMSE by sum-of-bit approach.

been tested, where B varies from 0.05 to 0.5 with step size 0.05, and the optimal value of B is selected.

From Fig. 2 it can be found that the MMSE rises a little when σ increases, while the MMSE rises more when α increases. This is intuitive since the variance of noise can be recovered with large K. But it is harder to estimate X when the source itself has large variance.

B. *Σ-Δ Modulation Approach*

Sigma-Delta Modulation can also be used to deal with the problem. This is done by adding a filter and oversampling the observations in front of the fusion center from each path, as Fig. 3 shows. After collecting the 1-bit information from each path, we make the estimation by

$$\widehat{X} = \mu + (-k_1 \cdot \alpha) + k_2 \cdot \sum_{i=1}^{K} \frac{b_i}{2^M} \quad (9)$$

where k_1, k_2 are parameters needed to be defined, 2^M is the number of decimation, and b_i is the number of ones density of input Z_i.

The value of k_1, k_2, M has influence on the performance. To minimize MSE, we first fix k_1. Then optimize k_2 by simulation. Fig. 4a shows the performance for different k_1. In fact, the optimal value of k_1 is about 0.8. This may be associated with the optimal 1-bit quantization level for Gaussian random variable, i.e. the quantization level for zero-mean Gaussian with variance σ^2 is $\pm\sigma\sqrt{2/\pi}$. On the other hand, the MMSE for different value of M is depicted in Fig.

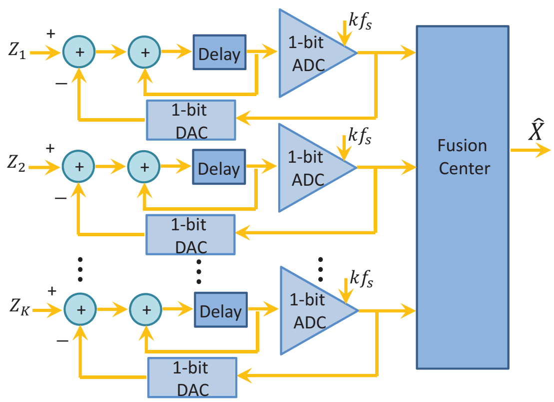
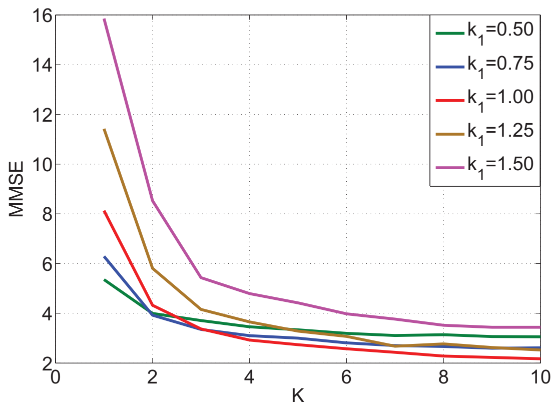
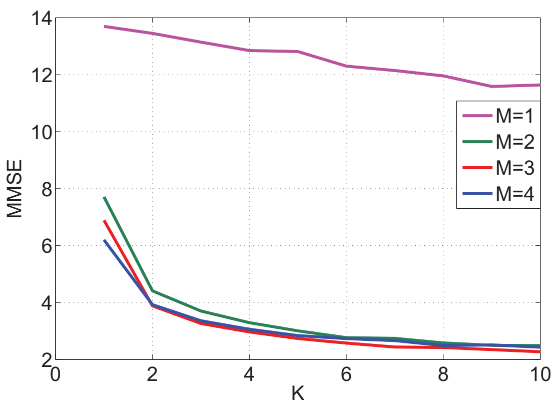


Fig. 3: The modified fusion center.



(a) $M = 4$, $k_1 = 0.50, 0.75, 1.00, 1.25, 1.50$.



(b) $k_1 = 0.8$, $M = 1, 2, 3, 4$.

Fig. 4: MMSE for (1) different k_1's, (2) different M's by $\Sigma - \Delta$ modulation approach.

4b. It can be seen that the performance for $M \geq 2$ is about the same.

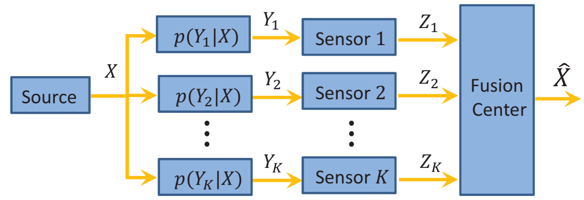


Fig. 5: A distributed estimation system with general observation distribution.

C. Source Coding Approach

A more general system model for distributed source coding has been considered by Gubner [8], as Fig. 5 shows. Each sensor k processes its measurement Y_k to obtain an output Z_k. Each Z_k is then transmitted to the fusion center. Assume that the use of error-correcting codes permits us to view the channel as noiseless. For each Y_k, $A_{k1}, \cdots A_{kN}$ is a partition of the real line, $\mathbb{R}$. In other words, it quantizes Y_k in N levels. The sensor output Z_k is then defined by

$$Z_k \triangleq \sum_{i=1}^{K} (i-1) I_{A_{ki}}(Y_k) \tag{10}$$

where $I_A(y)$ is the indicator function given by

$$I_A(y) = \begin{cases} 1\,, y \in A \\ 0\,, y \notin A \end{cases} \tag{11}$$

At the fusion center, $Z_1, \cdots Z_K$ is collected and for simplicity, the estimator $\widehat{X}$ is defined by

$$\widehat{X} = \sum_{k=1}^{K} \sum_{i=1}^{N} c_{ki} I_{\{i-1\}}(Z_k) \tag{12}$$

The algorithm for optimizing the partitions $\{A_{1i}\}_{i=1}^N, \{A_{2i}\}_{i=1}^N, \cdots \{A_{Ki}\}_{i=1}^N$ and the parameters c_{ki}, $i = 1, 2, \cdots, N$, $k = 1, 2, \cdots, K$ is given in [8], which is straight-forward. However, the computation makes it not realistic for implementation. Thus, we introduce another method to define the partitions $\{A_{ki}\}_{i=1}^N$. Consider a quantization problem: A continuous random variable X are needed to be quantized in K bits. How to minimize the mean squared error by specifying the quantization levels and step sizes? If the step sizes are identical for each quantization region, it is called uniform quantization. Nevertheless, to minimize MSE, nonuniform quantization must be used, in which the step sizes may not be of equal length. This can be done by the well-known *Lloyd algorithm* [9].

The algorithm operates as follows. Given an initial quantization region $\mathcal{R}_1, \mathcal{R}_2, \cdots, \mathcal{R}_M$, where $M = 2^K$, we calculate the conditional mean for each region.

$$a_i = \frac{\displaystyle\int_{\mathcal{R}_i} x f_X(x)\, dx}{P(x \in \mathcal{R}_i)}\,, i = 1, 2, \cdots, 2^K \tag{13}$$

where a_i is the quantization level for region i. Then the region $\mathcal{R}_i = \{x | b_{i-1} \leq x < b_i\}$ is updated by a_i's.

$$b_i = \frac{a_i + a_{i+1}}{2}\,, i = 1, 2, \cdots, 2^K \tag{14}$$

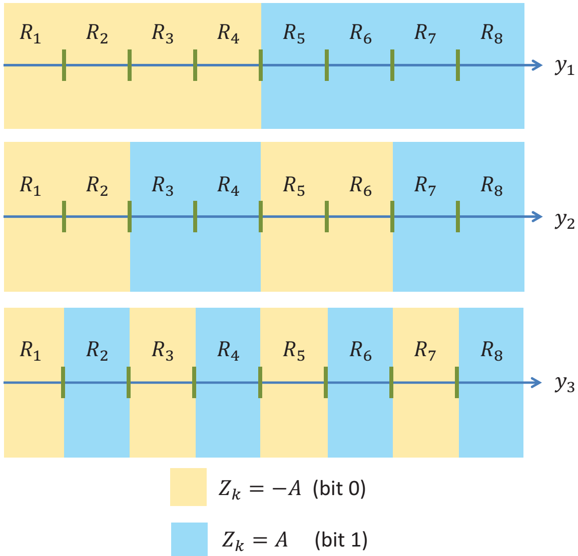


Fig. 6: Detection rule for sensor k, $k = 1, 2, 3$.

with $b_0 \triangleq -\infty$, $b_{2^K+1} = \infty$. After the regions $\mathcal{R}_i, i = 1, 2, \cdots, 2^K$ are decided, repeat the calculation of a_i's. This iteration stops until the difference of MSE between two iterations is small enough.

Next, we could connect this problem to our original problem. This is illustrated by Fig. 6. For instance, if $K = 3$, we first use Lloyd Algorithm to generate the quantization levels and regions for the random variable X. The detection rule at each sensor is described as follows.

$$Z_1 = \begin{cases} -A \text{ (bit 0), if } Y_1 \in \mathcal{R}_1 \bigcup \mathcal{R}_2 \bigcup \mathcal{R}_3 \bigcup \mathcal{R}_4 \\ A \text{ (bit 1), otherwise} \end{cases}$$

$$Z_2 = \begin{cases} -A \text{ (bit 0), if } Y_2 \in \mathcal{R}_1 \bigcup \mathcal{R}_2 \bigcup \mathcal{R}_5 \bigcup \mathcal{R}_6 \\ A \text{ (bit 1), otherwise} \end{cases}$$

$$Z_3 = \begin{cases} -A \text{ (bit 0), if } Y_3 \in \mathcal{R}_1 \bigcup \mathcal{R}_3 \bigcup \mathcal{R}_5 \bigcup \mathcal{R}_7 \\ A \text{ (bit 1), otherwise} \end{cases}$$

In other worlds, we use the quantization rule obtained by Lloyd Algorithm for X to quantize Y_k in sensor k. If the k-th bit is 0, sensor k sends $(-A)$, otherwise sends A. Finally, the fusion center combines these bits and maps back the quantization levels a_i, which is used to estimate X. The simulation result is illustrated in Fig. 7, where $\mu = 5$, $\alpha = 10$, $\sigma = 1$.

IV. Quantization Region Allocation

As the Fig. 7 shows, the MMSE doesn't decrease monotonically while the number of sensors increase. This may be due to the fact that bits sent from various sensors are not equally weighted. Namely, the most significant bit, which decides whether the source X locates on the left of mean μ or on the right, depends on one sensor only. Though every sensor represents different bit, these bits sent by sensors are not trustworthy since the noise exists in two stages of transmission.

Therefore, we must adjust our strategy. Only few bits are used in estimation. Besides, the number of sensors which

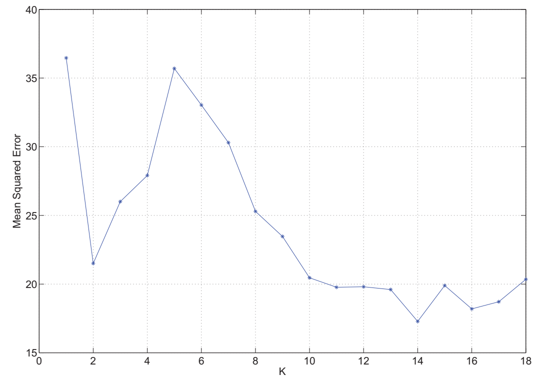
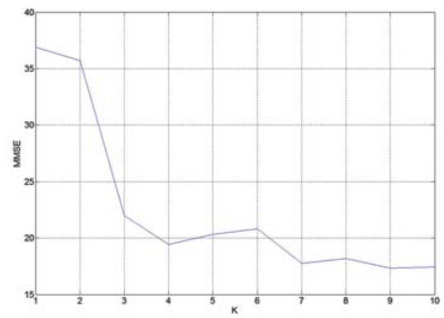
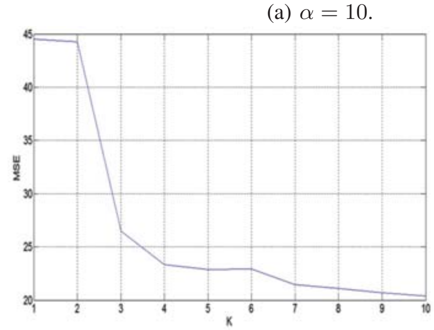


Fig. 7: MMSE for $\mu = 5$, $\alpha = 10$, $\sigma = 1$ by source coding approach.



bit 1	bit 2	bit 3
1	0	0
2	0	0
2	1	0
3	1	0
3	2	0
4	2	0
4	3	0
5	3	0
5	4	0
7	3	0

(a) $\alpha = 10$.



bit 1	bit 2	bit 3
1	0	0
2	0	0
2	1	0
3	1	0
4	1	0
4	2	0
4	3	2
4	2	2
5	2	2
4	3	3

(b) $\alpha = 11$.

Fig. 8: MMSE for $\mu = 5$, $\sigma = 1$ after bit allocation.

are used to decide specific bit is optimized. The simulation goes as follows. We use three bits to quantize the source. Let a_1, a_2, a_3 be the number of bit 1,2,3 repectively, where a_1 stands for the most significant bit (MSB) and a_3 represents the last significant bit (LSB). Since MSB is more important, assume that $a_1 \geq a_2 \geq a_3$ and running simulation for all possible allocation of bits. The result is shown in Fig. 8.

In Fig. 8, we also attach the table for the bit allocation. It can be seen that two bits are enough for estimation. Namely, Instead of increasing the resolution, it is more important to increase the accuracy of every bits, especially the MSB. In addition, notice that the third bit is used in Fig. 8b. This is due to the larger variance used in Fig. 8b than Fig. 8a. While dealing the source with larger variance, more bits are worthy to use.

V. Comparison to Soft Decision

Since the data sent from each sensor Z_k suffers from two noises, N_k and N'_k, Z_k given X is Gaussian distributed with variance $2\sigma^2$. Therefore, the *a posteriori* probability of X given $\underline{Z} = [Z_1\, Z_2 \cdots Z_K]$ is

$$\begin{aligned} & f_{X|\underline{Z}}(x|\underline{z}) \\ =& \frac{1}{f_{\underline{Z}}(\underline{z})} \prod_{k=1}^{K} \frac{1}{\sqrt{4\pi}\sigma} e^{-\frac{(z_k - x)^2}{4\sigma^2}} \cdot \frac{1}{\sqrt{2\pi}\alpha} e^{-\frac{(x-\mu)^2}{2\alpha^2}} \\ =& g(\underline{z}) \exp\left[-\frac{1}{2\sigma_p^2} \left(x - \sigma_p^2 (\frac{\mu}{\alpha^2} + \frac{1}{2\sigma^2} \sum_{k=1}^{K} z_k) \right)^2 \right] \end{aligned} \tag{15}$$

where $\sigma_p^2 = \left(\frac{K}{2\sigma^2} + \frac{1}{\alpha^2}\right)^{-1}$. Since the exponent in (15) is quadratic function of x, $f_{X|\underline{Z}}(x|\underline{z})$ is Gaussian. Note that the estimator under mean squared error criterion is the conditional mean of the *a posteriori* random variable. Thus the MMSE estimator is given by

$$\widehat{X}(\underline{Z}) = \sigma_p^2 \Big(\frac{\mu}{\alpha^2} + \frac{1}{2\sigma^2} \sum_{k=1}^{K} Z_k \Big) \tag{16}$$

And the corresponding MMSE is

$$E[(X - \widehat{X}(\underline{Z}))^2] = \left(\frac{K}{2\sigma^2} + \frac{1}{\alpha^2} \right)^{-1} \tag{17}$$

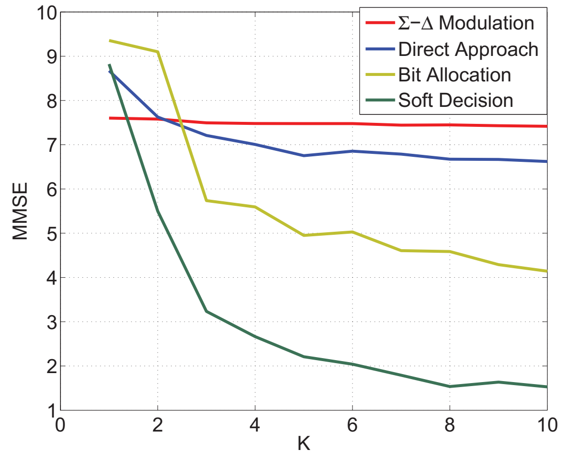
This should be the optimal performance for estimating X since in soft decision, we have assumed that all of the sensors can directly transmit the unquantized data they receive, which is not true in realistic.

Next, we compare the performance among optimal soft decision and the hard decision methods mensioned above. For soft decision, 8-FSK is used to transmit quantized data at each sensor. Thus, in these hard decision methods, the received data is oversampled at each sensor and the 1-bit data is transmitted using 8 BPSK channels simultaneously for fair comparison.

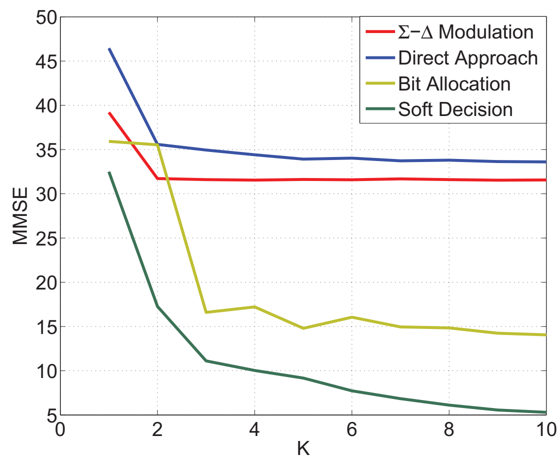
VI. Conclusion

From Fig. 9, we can see that the performance of Σ-Δ method is almost unchanged since it has oversampled the data at each sensor and it oversamples again at the fusion center. Thus, adding more sensor has little effect on the performance. Besides, the MMSE for Σ-Δ method is lower than direct method when the source variance is 100. This is intuitively due to the fact that oversampling for Σ-Δ method has more effect on the MMSE while the variance is large.

For bit allocation method, the MMSE decreases until more sensors are used. Since we let different sensors be responsible for different bits, which increases the resolution, and optimize the quantization region. The performance of bit allocation method is always better than other two hard decision methods when the number of sensors increases. Finally, proper windowing techniques may further improve the performance.



(a) $\alpha = 5$.



(b) $\alpha = 10$.

Fig. 9: MMSE comparison for $\mu = 5$, $\sigma = 1$.

References

[1] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 471–480, Jul. 1973.

[2] J. Garcia-Frias, "Compression of correlated binary sources using turbo codes," *Communications Letters, IEEE*, vol. 5, no. 10, pp. 417–419, Oct 2001.

[3] A. Aaron and B. Girod, "Compression with side information using turbo codes," in *Data Compression Conference, 2002. Proceedings. DCC 2002*, 2002, pp. 252–261.

[4] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, pp. 1–10, Jan 1976.

[5] T. Berger, Z. Zhang, and H. Viswanathan, "The ceo problem," *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 887–902, May 1996.

[6] V. Prabhakaran, D. Tse, and K. Ramachandran, "Rate region of the quadratic gaussian ceo problem," in *Information Theory, 2004. ISIT 2004. Proceedings. International Symposium on*, June 2004, pp. 119–.

[7] W. Kester, *The Data Conversion Handbook*. Newnes, 2005.

[8] J. Gubner, "Distributed estimation and quantization," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1456–1459, Jul 1993.

[9] S. Lloyd, "Least squares quantization in pcm," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar 1982.