ASYMPTOTIC APPROXIMATION OF OPTIMAL QUANTIZERS FOR ESTIMATION

Rodrigo Cabral Farias, Jean-Marc Brossier

GIPSA-lab, Images and Signal Department
University of Grenoble and Grenoble Institute of Technology
11, rue des Mathématiques, BP 46, 38402 Saint Martin d'Hères, France

ABSTRACT

In this paper, the asymptotic approximation of the Fisher information for the estimation of a scalar parameter based on quantized measurements is studied. As the number of quantization intervals tends to infinity, it is shown that the loss of Fisher information due to quantization decreases exponentially as a function of the number of quantization bits. The optimal quantization interval density and the corresponding maximum Fisher information are obtained. Comparison between optimal nonuniform and uniform quantization for the location estimation problem indicates that nonuniform quantization is slightly better. At the end of the paper, an adaptive algorithm for jointly estimating and setting the thresholds is used to show that the theoretical results can be approximately obtained in practice.

Index Terms— Parameter estimation, quantization, adaptive algorithm.

1. INTRODUCTION

Sensor arrays and sensor networks have attracted attention of the signal processing and communication research communities, as the cost, consumption and size of sensors and communication devices are reducing with recent technology advances. A plethora of sensing system applications ranging from military to commercial areas [1] also motivates the increasing research in multisensor approaches.

Passing from a single sensor context to a multisensor context creates new difficulties for algorithm design. Complexity and constrained bandwidth problems have to be taken into account. A direct approach to tackle these problems is to quantize the sensors measurements. Even though quantization theory is well developed for the reconstruction of measurements, with extensive literature [2], the extension of the theory for the reconstruction of a parameter embedded in quantized noisy measurements is much less developed. This extension, which can be called quantization for estimation, is clearly more related to the main objectives of a sensing system than quantization for measurement reconstruction.

Analysis of the performance of location parameter estimation based on multiple bit quantized measurements is presented in [3], where the effect of the quantizer input offset on the Cramér–Rao bound is studied. In [4], the same problem is analyzed, but in a binary quantization context. In this case, the problem of setting the binary thresholds when the parameter is random is studied in detail. In both works the total number of quantization bits per sensor is considered to be finite.

Asymptotic high-rate approximations for inference based on quantized measurements are presented in [5]. Asymptotic performance for parameter estimation is detailed in the uniform quantization case. A high-rate approximation is also proposed in [6], where the problem of estimating a random parameter based on scalarly quantized measurements is considered. The optimal companding function (quantizer input nonlinear function) and mean squared estimation errors are obtained considering that the quantizers output entropy sum is minimized, thus giving a characterization of the rate-distortion function for Bayesian estimation under the high-rate regime.

In this work, the analysis of asymptotic nonuniform scalar quantization for parameter estimation will be detailed, extending the results presented for uniform quantization in [5]. Differently from [6], the parameter is considered to be deterministic and an asymptotic approximation of the Fisher information (FI) will be written as a function of a density of quantization intervals (the point density). The optimal interval densities and asymptotic FI will then be obtained. The results will be applied to the location estimation problem for Gaussian and Cauchy distributions. A comparison between the theoretical approximation of the asymptotic maximum FI, the FI for uniform quantization and the FI for a practical approximation of the interval density is presented. Finally, as the optimal quantizer thresholds are shown to depend on the parameter to be estimated, an adaptive algorithm for setting them is used to show that the theoretical performance can be approximately achieved.

2. ASYMPTOTIC APPROXIMATION

The problem considered here is the estimation of a scalar deterministic parameter $x \in \mathbb{R}$ of a continuous distribution F based on N independent measurements from this distribution $\mathbf{Y} = [Y_1\ Y_2\ \cdots\ Y_N]^T$. Due to the constraints explained above, estimation of x will not be done based on $\mathbf{Y}$, instead it will be based on a scalarly quantized version of $\mathbf{Y}$, that will be denoted

$$\mathbf{i} = [i_1\ i_2\ \cdots\ i_N]^T = [Q(Y_1)\ Q(Y_2)\ \cdots\ Q(Y_N)]^T,$$

the function Q represents the scalar quantizer and is given by

$$Q(Y) = i, \quad \text{if} \quad Y \in \Delta_i = [\tau_{i-1}, \tau_i), \tag{1}$$

where $i \in \{1, \cdots, N_I\}$, N_I is the number of quantization intervals Δ_i and τ_i are the quantizer thresholds. The first and last thresholds will be set to be $\tau_0 = \tau_{\text{min}}$ and $\tau_{N_I} = \tau_{\text{max}}$.

The Cramér–Rao bound (CRB) for estimating x based on $\mathbf{i}$ can be used to give a lower bound on the variance of any unbiased estimator $\hat{X}$ of x. Under the independence assumption and supposing that the support of F does not depend on x, the CRB for $\hat{X}$ is

$$\mathbb{V}\text{ar}\left[\hat{X}\right] \geq CRB_q = \frac{1}{NI_q}, \tag{2}$$

The authors would like to thank the EBW II program for funding this work.

where I_q is the FI for a quantized measurement, it can be written as

$$I_q = \mathbb{E}\left[S_q^2\right] = \mathbb{E}\left\{\left[\frac{\partial \log \mathbb{P}(i;x)}{\partial x}\right]^2\right\} = \sum_{i=1}^{N_I}\left[\frac{\partial \log \mathbb{P}(i;x)}{\partial x}\right]^2 \mathbb{P}(i;x), \quad (3)$$

S_q is the score function for quantized measurements and $\mathbb{P}(i;x)$ is the probability of having the quantizer output i:

$$\mathbb{P}(i;x) = F(\tau_i;x) - F(\tau_{i-1};x). \quad (4)$$

For $N \to \infty$, the CRB is tight and it can be attained asymptotically using maximum likelihood estimation. Thus, the CRB and consequently the FI can be used to approximately assess estimation performance.

The FI in (3) is a function not only of the noise distribution and the parameter, but also in this case, a function of the quantizer thresholds (or equivalently of the quantizer intervals). Clearly, two questions arise: what are the optimal thresholds that maximize the FI? For the optimal thresholds, what is the value of the FI as a function of N_I?

Only in a few cases these questions can be answered easily. In the binary case, for the estimation of a location parameter, under certain conditions on F, it can be shown that the threshold must be placed exactly at x. For uniform quantization, maximizing the FI with respect to (w.r.t.) the quantization step is a one dimensional maximization problem, which can be solved easily using exhaustive search. For nonuniform non binary quantization, maximization of I_q w.r.t. τ_i or Δ_i is a difficult problem.

Similarly to standard quantization for measurement reconstruction, where optimal nonuniform quantization intervals can be approximated for large N_I, an approximation for I_q will now be developed.

It will be assumed that $F(y;x)$ accepts a probability density function (PDF) $f(y;x)$ positive, smooth in both y and x and defined on a bounded support. Following a similar development as in [6], the quantity $\mathbb{E}\left[(S_c - S_q)^2\right]$ will be analyzed. S_c is the score function for estimation based on continuous measurements $S_c = \frac{\partial \log f(y;x)}{\partial x}$. The quantity $\mathbb{E}\left[(S_c - S_q)^2\right]$ can be rewritten as

$$\mathbb{E}\left[(S_c - S_q)^2\right] = I_c + I_q - 2\mathbb{E}\left[S_c S_q\right], \quad (5)$$

where I_c is the FI for continuous measurements. Using the definitions of S_c and S_q, it can be shown that $\mathbb{E}\left[S_c S_q\right] = \mathbb{E}\left[S_q^2\right]$. Therefore,

$$I_q = I_c - \mathbb{E}\left[(S_c - S_q)^2\right]. \quad (6)$$

The right term shows that quantization can only decrease FI and it can be interpreted as the loss of performance due to quantization. The loss, denoted L from now on, must be minimized with respect to the quantization intervals. L, which is an expectation under the measure F, can be rewritten as a sum of integrals, each term of the integral corresponds to the loss produced by a quantization interval:

$$L = \sum_{i=1}^{N_I} \int_{\Delta_i} \left[\frac{\partial \log f(y;x)}{\partial x} - \frac{\partial \log \mathbb{P}(i;x)}{\partial x}\right]^2 f(y;x)\,\mathrm{d}y. \quad (7)$$

For the interval with index i, the PDF can be approximated with a Taylor series around the central point $y_i = \frac{\tau_i + \tau_{i-1}}{2}$:

$$f(y;x) = f_i + f_i^{(y)}(y - y_i) + \frac{f_i^{(yy)}}{2}(y - y_i)^2 + o(y - y_i)^2, \quad (8)$$

where the superscripts indicate the variables for which the function is differentiated. The subscript represents that the function (after differentiation) is evaluated at y_i. It will be assumed that the sequences of intervals for increasing N_I are chosen such that for any $\epsilon > 0$ it is possible to find a N_I^* for which

$$\frac{o(y - y_i)^2}{(y - y_i)^2} < \epsilon, \quad \text{for } N_I > N_I^*,\ y \in \Delta_i. \quad (9)$$

Under the assumption that $f > 0$, the logarithm of f at interval Δ_i can be approximated also using a Taylor series:

$$\log f(y;x) = \log f_i + (\log f)_i^{(y)}(y - y_i) + (\log f)_i^{(yy)}\frac{(y - y_i)^2}{2} + o(y - y_i)^2 \quad (10)$$

and the derivative w.r.t. x is

$$\frac{\partial \log f(y;x)}{\partial x} = (\log f)_i^{(x)} + (\log f)_i^{(yx)}(y - y_i) + (\log f)_i^{(yyx)}\frac{(y - y_i)^2}{2} + o(y - y_i)^2, \quad (11)$$

which is an expression for the continuous score function on Δ_i to be used in (7). Now, the other term in the squared factor must be calculated. Integrating the PDF in (8) on the interval Δ_i (which is denoted in the same form as its length), one gets

$$\mathbb{P}(i,x) = f_i \Delta_i + f_i^{(yy)}\frac{\Delta_i^3}{24} + o\left(\Delta_i^3\right). \quad (12)$$

Note that the term on Δ_i^2 is zero because y_i is the interval central point. The logarithm of $\mathbb{P}(i,x)$ can be obtained by dividing the second and third terms of the right hand side of (12) by the first term and then using the Taylor series for $\log(1+x)$. Differentiating the resulting expression w.r.t. x gives

$$\frac{\partial \log \mathbb{P}(i,x)}{\partial x} = (\log f)_i^{(x)} + \left(\frac{f^{(yy)}}{f}\right)_i^{(x)}\frac{\Delta_i^2}{24} + o\left(\Delta_i^2\right). \quad (13)$$

Subtracting (13) from (11), squaring, then multiplying by the Taylor series of f and integrating gives:

$$L = \sum_{i=1}^{N_I}\left\{\left[(\log f)_i^{(yx)}\right]^2 f_i \frac{\Delta_i^3}{12} + o\left(\Delta_i^3\right)\right\} = \sum_{i=1}^{N_I}\left\{\left(S_{c,i}^{(y)}\right)^2 f_i \frac{\Delta_i^3}{12} + o\left(\Delta_i^3\right)\right\}. \quad (14)$$

To obtain a characterization of the quantization intervals, an interval density function $\lambda(y)$ will be defined:

$$\lambda(y) = \lambda_i = \frac{1}{N_I \Delta_i}, \quad \text{for } y \in \Delta_i. \quad (15)$$

Rewriting (14) with this density gives

$$L = \sum_{i=1}^{N_I}\left\{\left(S_{c,i}^{(y)}\right)^2 f_i \frac{\Delta_i}{12 N_I^2 \lambda_i^2} + o\left(\frac{1}{N_I^2}\right)\Delta_i\right\}. \quad (16)$$

As $N_I \to \infty$, it will be supposed that all Δ_i converge uniformly to zero. Therefore,

$$\lim_{N_I \to \infty} N_I^2 L = \frac{1}{12}\int \frac{\left(\frac{\partial S_c(y;x)}{\partial y}\right)^2 f(y;x)}{\lambda^2(y)}\mathrm{d}y. \quad (17)$$

N_B	Gaussian ($I_c = 2$) Optimal	Uniform	Companding approx.	Cauchy ($I_c = 0.5$) Optimal	Uniform	Companding approx.
1	$1.27323954^{\dagger}$	–	1.27323954	$0.40528473^{\dagger}$	–	0.40528473
2	$1.76503630^{\dagger}$	1.76503630	1.75128300	$0.43433896^{\dagger}$	0.43433896	0.40528473
3	$1.93090199^{\dagger}$	1.92837814	1.92740111	$0.48474865^{\dagger}$	0.45600797	0.47893785
4	$1.97874454^{\star}$	1.97841622	1.98038526	$0.49533850^{\star}$	0.48136612	0.49504170
5	$1.99468613^{\star}$	1.99353005	1.99489906	$0.49883463^{\star}$	0.49204506	0.49879785
6	$1.99867153^{\star}$	1.99807736	1.99869886	$0.49970866^{\star}$	0.49656712	0.49970408
7	$1.99966788^{\star}$	1.99943563	1.99967136	$0.49992716^{\star}$	0.49851056	0.49992659
8	$1.99991697^{\star}$	1.99983649	1.99991741	$0.49998179^{\star}$	0.49935225	0.49998172

Table 1. Fisher information (FI) for the estimation of Gaussian and Cauchy location parameters based on quantized measurements. N_B is the number of quantization bits. In *Optimal*† the maximum FI obtained by exhaustive search of the thresholds is shown. *Optimal** is the theoretical asymptotic approximation of the FI. *Uniform* shows the value of the FI for optimal uniform quantization and *Companding approx.* gives the FI for the practical approximation of the asymptotically optimal thresholds.

Based on (17), the following approximation can be used for the FI

$$I_q \approx I_c - \frac{1}{12N_I^2}\int \frac{\left(\frac{\partial S_c(y;x)}{\partial y}\right)^2 f\left(y;x\right)}{\lambda^2\left(y\right)}\mathrm{d}y, \tag{18}$$

which is valid for large N_I.

It is possible to find the optimal interval density $\lambda^{\star}$ maximizing (18) and the corresponding maximum FI $I_q^{\star}$ by applying the Hölder's inequality to the integral. This gives

$$\lambda^{\star}\left(y\right) = \frac{\left(\frac{\partial S_c(y;x)}{\partial y}\right)^{\frac{2}{3}} f^{\frac{1}{3}}\left(y;x\right)}{\int\left(\frac{\partial S_c(y;x)}{\partial y}\right)^{\frac{2}{3}} f^{\frac{1}{3}}\left(y;x\right)\mathrm{d}y}, \tag{19}$$

$$I_q^{\star} \approx I_c - \frac{1}{12N_I^2}\left[\int\left(\frac{\partial S_c\left(y;x\right)}{\partial y}\right)^{\frac{2}{3}} f^{\frac{1}{3}}\left(y;x\right)\mathrm{d}y\right]^3. \tag{20}$$

The main difference from standard quantization is the presence of the derivative of the score function in the interval density.

3. LOCATION PARAMETER ESTIMATION

As an application of the results above, estimation of the location parameter x of two different distributions will be considered. The distributions chosen to be analyzed are the Gaussian and the Cauchy distributions, the former is commonly used for modeling thermal noise and the latter for modeling noise with outliers. Even if their support is unbounded, as in standard quantization theory, it is expected that the error caused by incorrect approximation of the extremal regions (overload region) will be small. Their PDFs are given respectively by

$$f_G\left(y;x\right) = \frac{1}{\delta\sqrt{\pi}}\mathrm{e}^{-\left(\frac{y-x}{\delta}\right)^2}, \tag{21}$$

$$f_C\left(y;x\right) = \frac{1}{\delta\pi}\frac{1}{\left[1+\left(\frac{y-x}{\delta}\right)^2\right]}, \tag{22}$$

where δ is scale parameter. Denoting $N_B = \log_2\left(N_I\right)$ the number of quantization bits, one gets for the Gaussian PDF

$$\lambda_G^{\star}\left(y\right) = \frac{1}{\delta\sqrt{3\pi}}\mathrm{e}^{-\left(\frac{y-x}{\sqrt{3}\delta}\right)^2}, \tag{23}$$

$$I_{q,G}^{\star} \approx \frac{2}{\delta^2}\left[1-\pi\sqrt{3}\,2^{-(2N_B-1)}\right], \tag{24}$$

Note that the interval density in this case is exactly the same as for standard quantization (proportional to $f^{\frac{1}{3}}$). For the Cauchy distribution

$$\lambda_C^{\star}\left(y\right) = \frac{1}{\delta B\left(\frac{1}{2};\frac{5}{6}\right)}\frac{\left[1-\left(\frac{y-x}{\delta}\right)^2\right]^{\frac{2}{3}}}{\left[1+\left(\frac{y-x}{\delta}\right)^2\right]^{\frac{5}{3}}}, \tag{25}$$

$$I_{q,C}^{\star} \approx \frac{1}{2\delta^2}\left[1-\frac{B\left(\frac{1}{2};\frac{5}{6}\right)^3}{3\pi}2^{-2N_B+1}\right], \tag{26}$$

where B is the *beta* function. From the definition of the interval density, the percentage of intervals until interval i, $\frac{i}{N_I}$ must be equal to the integral of the interval density from $\tau_{\min}$ to τ_i. Thus, a practical way of approximating the optimal thresholds is to set

$$\tau_i^{\star} = F_{\lambda}^{-1}\left(\frac{i}{N_I}\right), \tag{27}$$

where F_{λ}^{-1} is the inverse of the cumulative distribution function (CDF) related to λ.

To evaluate the validity of the approximations, the FI (3) under both distributions with $\delta = 1$ was evaluated for

- the optimal set of thresholds for $N_B = \{1,2,3\}$. The optimal thresholds were obtained through exhaustive search. For $N_B = \{4,5,6,7,8\}$ the theoretical results (24) and (26) were used as an approximation.
- uniform quantization considering $N_B = \{1,\cdots,8\}$. After setting the central threshold to x, the optimal quantization interval Δ was found by maximizing the FI also using exhaustive search.
- the approximate optimal set of thresholds given by (27), for $N_B = \{1,\cdots,8\}$.

The results are given in Tab. 1.

Note that in all cases fast convergence to the continuous Fisher information (I_c) with increasing N_B is verified. It seems that for estimation purposes 4 quantization bits are enough. The difference of performance between uniform and nonuniform quantization seems to be higher for the Cauchy distribution. In the Gaussian case this difference is negligible, indicating that in practice uniform quantization should be used (as it is easier to implement). It can also be observed that the theoretical asymptotic approximation of I_q and its true value for the approximation of the optimal thresholds are very close, even for small values of N_B as 4 and 5. Thus, it seems that the asymptotic approximation of I_q given by (18) and that the approximation of the optimal thresholds given by (27) are possible answers for the questions raised at the beginning of Sec. 2, at least for $N_B \geq 4$.

An important issue for evaluating τ_i in (27) is that they depend explicitly on x. A possible solution for this problem is to initially set τ_i with an arbitrary guess of x, then estimate x using an initial set of measurements and finally update the thresholds with the estimate. This procedure can be done in a recursive way to get closer and closer to the optimal thresholds.

4. ADAPTIVE ALGORITHM

A recursive procedure for updating the estimate of x that will produce asymptotically optimal results associated with the threshold placement proposed here (27) can be found in [7]. The procedure is the following: for measurement Y_k, the quantizer thresholds will be set using (27) with x being replaced by its last estimate $\hat{X}_{k-1}$, then Y_k is quantized using the new thresholds, this produces the quantized measurement i_k and a new estimate given by

$$\hat{X}_k = \hat{X}_{k-1} + \frac{1}{kI_q^x}\eta\left(i_k\right), \tag{28}$$

where I_q^x is the Fisher information when the central threshold $\tau_{\frac{N_I}{2}}$ is placed exactly at x (N_I is assumed to be an even number). The function η is given by

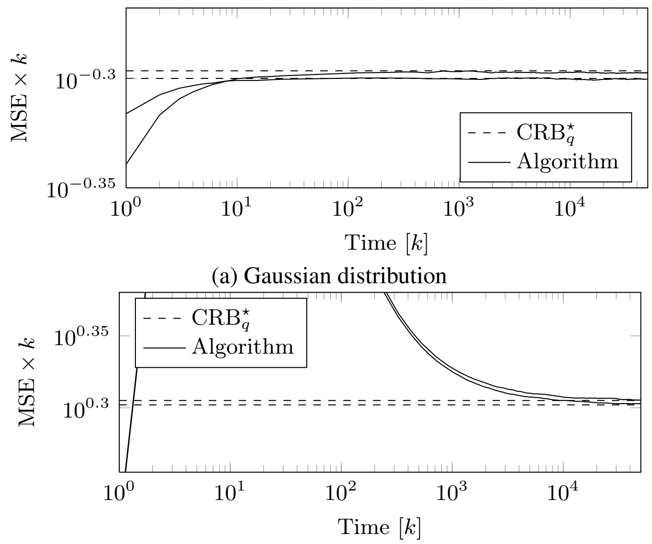
$$\eta\left(i\right) = \frac{f\left(\tau_{i-1}^\star; x\right) - f\left(\tau_i^\star; x\right)}{F\left(\tau_i^\star; x\right) - F\left(\tau_{i-1}^\star; x\right)}, \tag{29}$$

here F is the CDF of Y_k and $\tau_i^\star$ are the thresholds calculated using the true x. Note that for location estimation, η is independent of x.

This algorithm can be seen as gradient ascent algorithm applied to the log-likelihood. Under some constraints on f, it can be shown [7] that the estimator $\hat{X}_k$ is asymptotically unbiased and that its asymptotic variance attains the Cramér–Rao bound for a fixed thresholding scheme with the central threshold placed at x. Thus, if $N_B > 4$ and the approximation of the optimal thresholds is used, the asymptotic variance of this algorithm will be close to optimal and it will be given approximately by

$$\mathbb{V}\text{ar}\left[\hat{X}_k\right] \approx CRB_q^\star = \frac{1}{kI_q^\star}. \tag{30}$$

This algorithm was tested under both distributions for $N_B = 4$ and 5. The mean squared error (MSE) of estimation was evaluated using Monte Carlo simulation, 4×10^6 realizations of blocks with 5×10^4 samples were used. The initial error $x - \hat{X}_0$ and δ were both set to be 1 in all simulations. The MSE for the algorithm and the approximation given by (30) are both given in Fig. 1, where they are multiplied by k for better visualization. It can be observed that the asymptotic algorithm performance is very close to the approximation. For small k the CRB is not tight and that seems to be the reason for the algorithm to perform better than the bound. In other simulations, it was also observed that using uniform thresholds leads to faster convergence to the asymptotic performance. This indicates that in practice an algorithm with changing thresholds can be used for obtaining better results. In the convergence phase, a uniform set of thresholds is used, then after a given number of samples, the thresholds change to the approximately optimal set.



(b) Cauchy distribution

Fig. 1. Simulated mean squared error (MSE) for the adaptive algorithm considering Gaussian and Cauchy distributions. The numbers of quantization bits are $N_B = 4$ and 5. The initial estimation error and δ were set to 1 in all the cases. The curves that have asymptotically higher values correspond to $N_B = 4$.

5. CONCLUSIONS

In this paper, an asymptotic approximation of the Fisher information for the estimation of a scalar parameter based on quantized measurements is presented. The approximation is given for a large number of quantization intervals. It is shown that the Fisher information loss due to quantization decreases quadratically as a function of the number of quantization intervals, or equivalently, it decreases exponentially as a function of the number of quantization bits. This means that for estimation purposes, the best strategy will probably be based on a low resolution multiple sensor approach. Also, the optimal quantization interval density is obtained and showed to depend not only on $f^{\frac{1}{3}}$ but also on a power of the score function derivative.

Application of the results in location parameter estimation is presented, Gaussian and Cauchy distributions are considered. It is shown that the asymptotic results are valid for 4 quantization bits or more. A comparison with the uniform case shows that specially in the Gaussian case, nonuniform quantization is only slightly better. In the Cauchy case, the performance gain seems to be moderate. As the threshold density is shown to depend on the parameter, a recursive algorithm that jointly estimates the location parameter and sets the quantizer thresholds is presented. Performance results for this algorithm show that the asymptotic results can be approximately achieved in practice.

As a direct extension of this work, vector quantization can be considered. Also, by using the Fisher information asymptotic results, approximate optimal bit allocation for estimation with multiple sensors under constrained bandwidth can be studied.

6. REFERENCES

[1] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *Communications magazine,*

IEEE, vol. 40, no. 8, pp. 102–114, 2002.

[2] R.M. Gray and D.L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325–2383, 1998.

[3] H.C. Papadopoulos, G.W. Wornell, and A.V. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 978–1002, 2001.

[4] A. Ribeiro and G.B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks-part I: Gaussian case," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 1131–1143, 2006.

[5] HV Poor, "Fine quantization in signal detection and estimation," *IEEE Trans. Inf. Theory*, vol. 34, no. 5, pp. 960–972, 1988.

[6] S. Marano, V. Matta, and P. Willett, "Asymptotic design of quantizers for decentralized mmse estimation," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5485–5496, 2007.

[7] R. C. Farias and J.-M. Brossier, "Adaptive quantizers for estimation," arXiv:1210.3583, October 2012.