Distributed Quantization of Order Statistics with Applications to CSI Feedback

Matthew Pugh, *Student Member, IEEE*, and Bhaskar D. Rao, *Fellow, IEEE*

Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, CA, 92093-0407, USA
E-mail: {mopugh, brao}@ucsd.edu

Abstract

Feedback of channel state information (CSI) in wireless systems is essential in order to exploit multi-user diversity and achieve the highest possible performace. When each spatially distributed user in the wireless system is assumed to have i.i.d. scalar CSI values, the optimal fixed-rate and entropy-constrained point density functions are established in the high-resolution regime for the quantization of the CSI feedback to a centralized scheduler under the mean square error (MSE) criterion. The spatially distributed nature of the users leads to a distributed functional scalar quantization approach for the optimal high resolution point densities of the CSI feedback. Under a mild absolute moment criterion, it is shown that with a greedy scheduling algorithm at the centralized scheduler, the optimal fixed-rate point density for each user corresponds to a point density associated with the maximal order statistic distribution. This result is generalized to monotonic functions of arbitrary order statistics. Optimal point densities under entropy-constrained quantization for the CSI are established under mild conditions on the distribution function of the CSI metric.

I. Introduction

In many wireless systems, resource allocation and/or scheduling decisions are made based on feedback of channel state information (CSI). In a cellular broadcast channel, cell phones provide CSI feedback to the base station and the base station computes a function of the fed back CSI to determine the appropriate resource allocations or scheduling. For example, in [1], each user in the broadcast channel feeds back their measured signal-to-noise plus interference ratio (SINR) on each random transmit beam, and the base station schedules the user with the highest SINR for each transmit beam. In this case, the SINR represents the CSI metric, and the scheduling function is the $\max(\cdot)$ function, i.e. greedy scheduling. For a comprehensive overview of the role of feedback in wireless system, see [2].

A question that naturally arises in these situations is how to appropriately quantize the CSI for feedback. If the statistics of the CSI are known, then one could naively design a quantizer based on the distribution of the CSI metric. This, however, is sub-optimal if the function that determines the resource allocation or scheduling is known a priori. The problem of interest is finding the optimal high resolution point densities for quantization of the CSI when the users are spatially distributed and a function of the feedback is to be computed at a central controller. This problem is a specific case of the work on distributed functional scalar quantization (DFSQ) established in [3].

This contribution will consider the case where the scalar CSI at each spatially distributed user is i.i.d. and the distribution is known. The main result of this work is to show that the optimal (in the MSE sense) high-resolution point density for quantization of the CSI at each user is based on the distribution of the order statistics for the fixed-rate quantization case and a function of this distribution for the variable-rate quantization case. The MSE distortion criterion is chosen due to its analytical tractability. The organization of this work is as follows. In Section II, the quantization problem is formally stated. The optimal high-resolution fixed-rate point density is considered in Section III and the optimal variable-rate point density is considered in IV. Section V summarizes the results and concludes the contribution.

II. Distributed Functional Scalar Quantization

Distributed functional scalar quantization originated in [3] and the DFSQ problem as it relates to the model considered in this contribution is now stated. Let there be n distributed users in the wireless network where user i observes their CSI, which is a random variable denoted X_i. The observed random variables are assumed to be continuous and i.i.d. from distribution F, i.e. $X_i \sim_{i.i.d.} F$ for $i = 1, \ldots, n$. Let $M_n = \max\{X_1, \ldots, X_n\}$ be the maximum of the observed random variables, which is the output of the greedy scheduling function computed at the central controller. Each user will apply a quantizer Q_i to their observation to produce a quantized version of the observed CSI random variable $\widehat{X_i} = Q_i(X_i)$. The $\widehat{X_i}$ are fed back and the maximum is computed, i.e. $\widehat{M_n} = \max\left\{\widehat{X_1}, \ldots, \widehat{X_n}\right\}$. The goal is to find the optimal high resolution point density for the quantizers Q_i such that the mean squared error (MSE) distortion is minimized, i.e. the point density for the quantizers Q_i such that the quantity $E[(M_n - \widehat{M_n})^2]$ is minimized.

Suppose that rather than being distributed, the $X_1, \ldots, X_n$ were known at one source (i.e. not distributed), and that M_n was to be transmitted to some sink. In this case, M_n is known exactly at the source, and an optimal point density based on the distribution of M_n can be applied which results in minimum MSE distortion. This contribution will show that under the i.i.d. observation assumption and the goal of computing the maximum at the centralized controller, the optimal high resolution fixed-rate point density is equivalent to the fixed-rate point density of the maximum order statistic. In the entropy constrained case, the distributed point density function differs from the non-distributed point density function. With the results established for the maximal order statistic, optimal point densities will be found for arbitrary order statistics, and monotonic functions of the order statistics.

III. Fixed-Rate Quantizers

The question that is addressed in this section is finding the optimal fixed-rate point density function. A high-resolution approach will be taken. Consider the case mentioned at the end of Section II where the X_i are not distributed, so that M_n is known exactly at the source. Since the X_i are drawn i.i.d. from some distribution F, the distribution of the maximum order statistic M_n is given by $F_{\max}(x) = F^n(x)$ ([4]). If the optimal point density function for the fixed-rate quantizer exists, then under the high-resolution assumption it is given by the well known result ([5]):

$$\lambda(x) = \frac{f^{\frac{1}{3}}(x)}{\int f^{\frac{1}{3}}(x)dx}, \tag{1}$$

where $f(x)$ is the density of the random variable to be quantized, and the integral in the denominator of Equation (1) is taken over the support of the random variable. In this case, $f(x) = \frac{d}{dx}F^n(x)$.

Equation (1) states the optimal point density for a high-resolution fixed-rate quantizer provided it exists. Existence is considered in [6] and [7] where it is shown that a simple sufficient condition for the convergence of the Bennett integral is that the random variable to be quantized under an r^{th} power distortion function (i.e. minimizing $E[|X-\hat{X}|^r]$) satisfies $E\left[|X|^{r+\epsilon}\right]<\infty$ for some $\epsilon>0$. Thus, under the MSE criterion and the non-distribued X_i, the point density function $\lambda(x)$ exists provided $E\left[|M_n|^{2+\epsilon}\right]<\infty$ for some $\epsilon>0$. The following lemma establishes when this condition is satisfied.

Lemma 1: Let $X_1,\ldots,X_n$ be drawn i.i.d from some distribution F such that $E\left[|X_i|^{2+\epsilon}\right]<\infty$ for some $\epsilon>0$. Then $M_n=\bigvee_{i=1}^n X_i$, $E\left[|M_n|^{2+\epsilon}\right]<\infty$, and the optimal fixed-rate point density function $\lambda(x)$ exists.

Proof: From [4], it is known that if $E[g(X)]$ exists, then so does $E[g(X_{r:n})]$, where $X_{r:n}$ is the r^{th} order statistics. Thus, letting $g(x)=|x|^{2+\epsilon}$ and $r=n$, if $E[|X|^{2+\epsilon}]<\infty$ for some $\epsilon>0$, then $E[|M_n|^{2+\epsilon}]<\infty$ and the optimal point density function $\lambda(x)$ exists by [6] and [7]. ■

So far only the case where M_n is known at the source has been considered. Now consider the case of interest, where the X_i are distributed and individually quantized and fed back to a centralized controller which is to compute the maximum of the X_i. Let $\lambda_i(x)$ be the point density function for the quantizer Q_i. The optimal point density function $\lambda_i(x)$ for $i=1,\ldots,n$ in this distributed case is given in the following theorem:

Theorem 1: Let the scalar CSI observed at each distributed user be a continuous i.i.d. random variable X_i with density f and distribution F for $i=1,\ldots,n$ and satisfy the same conditions as Lemma 1. The optimal point density function for the fixed-rate quantizer Q_i under greedy scheduling and the MSE distortion is then given by

$$\lambda_i(x)=\frac{\left(nf(x)F^{n-1}(x)\right)^{\frac{1}{3}}}{\int\left(nf(x)F^{n-1}(x)\right)^{\frac{1}{3}}\,dx}\text{ for } i=1,\ldots,n. \tag{2}$$

Proof: Because the X_i are i.i.d. and distributed by F, M_n is distributed as $F^n(x)$. If the user experiencing the maximum was known, the optimal quantizer could be applied, and that quantizer would be designed for the distribution $M_n=F^n(x)$ since it is the maximum. Since $F^n(x)$ is the distribution function, the density function is given by

$$f_{M_n}(x)=\frac{d}{dx}F^n(x)=nf(x)F^{n-1}(x). \tag{3}$$

By Lemma 1, it is known that the optimal point density function exists since it is assumed the conditions of the lemma are satisfied, thus substituting Equation (3) into Equation (1) yields Equation (2), the optimal point density function for the user experiencing the maximum. It is not known which user is experiencing the maximum, but one of the n users must be the maximum. Thus applying the quantizer with point density function given by Equation (2) to all the users will optimally quantize the maximum. The centralized controller will select the codeword associated with the largest fed back value, and since all the codebooks are the same, this will yield the global maximum. ■

Theorem 1 shows that the optimal point densities for the quantizers at each user is equivalent to the optimal point density as if the users were not distributed. In [3], the design of optimal high-resolution fixed-rate quantizers is considered for general functions under distributed users. In that work, the maximum function is considered. The assumptions made in [3] are more restrictive, with the biggest constraint being that the random variables are defined on the interval $[0,1]$. The conditions for Theorem 1 are quite general and encompass random variables not confined to the unit interval. For example, in [1], the SINR represents the CSI, and the SINR is an unbounded random variable. Thus it is of interest to optimally quantize random variables of this nature, and Theorem 1 gives the solution. It will now be shown that the results in [3] can be extended to unbounded random variables, and in fact the method in [3] leads to the same conclusion as Theorem 1.

One of the examples considered in [3] is computing the maximum of n distributed i.i.d. uniform random variables. This will provide the basis to extending the computation of the maximum of more general i.i.d. random variables. Before showing the extension, some definitions must be established. Let $g : \mathbb{R}^n \to \mathbb{R}$ be a function that is continuous in each variable, with bounded gradient when defined, and whose first and second derivatives are defined except on a set of zero Jordan measure. Let $Y_1, \ldots, Y_n$ have continuous joint density $f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n)$ which is bounded and supported on $[0, 1]^n$. The following definition is from [3]:

Definition 1: ([3]) The j^{th} functional sensitivity profile of g is defined as

$$\gamma_j(x) = \left(E\left[\left| \frac{\partial}{\partial x_j} g\left(Y_1, \ldots, Y_n\right) \right|^2 |Y_j = y \right] \right)^{\frac{1}{2}}. \tag{4}$$

With the functional sensitivity profile defined, [3] proves the optimal fixed-rate point density function for the j^{th} user in the following theorem:

Theorem 2: ([3]) The optimal point density for the j^{th} distributed user under MSE distortion for the computation of the function g at a central controller is given by

$$\lambda_j(y) = \frac{\left(\gamma_j^2(y) f_{Y_j}(y)\right)^{1/3}}{\int \left(\gamma_j^2(t) f_{Y_j}(t)\right)^{1/3} dt} \text{ for } j = 1, \ldots, n \tag{5}$$

where f_{Y_j} is the marginal density for Y_j.

The goal is to show that the point density function given in Theorem 2 is equivalent to the results of Theorem 1. This equivalence is proven in the following corollary:

Corollary 1: Let $X_1, \ldots, X_n$ be distributed, independent and identically distributed continuous random variables with distribution F, continuous density f and continuous f' such that f is strictly postive on the support. Let F^{-1} have second derivatives except on a set of zero measure. Also let the X_i satisfy the requirements of Lemma 1. Then the point density function given in Equation (2) is equivalent to the point density function in Equation (5) for the computation of the function $g(X_1, \ldots, X_n) = \max(X_1, \ldots, X_n)$.

Proof: Let $U_1, \ldots, U_n$ be i.i.d. uniform random variables on $[0, 1]$. Then $X_j =_d F^{-1}(U_j)$ where $=_d$ means equal in distribution. Thus

$$\begin{aligned} g(X_1, \ldots, X_n) &= \max(X_1, \ldots, X_n) \\ &=_d \max\left(F^{-1}(U_1), \ldots, F^{-1}(U_n)\right) \\ &\triangleq \widetilde{g}(U_1, \ldots, U_n). \end{aligned}$$

By the properties imposed on F and the properties of the $\max$ function (see Example 4 in [3]), the conditions required of g in Theorem 2 are satisfied. Clearly, the joint distribution of n i.i.d. uniform random variables is continuous, bounded, and supported on $[0, 1]^n$. Therefore the requirements of Theorem 2 are satisfied. Evaluating the functional sensivity profile as in Example 4 in [3], yields

$$\left(\frac{d}{du} F^{-1}(u)\right)^2 u^{n-1} \text{ for } u \in [0, 1]. \tag{6}$$

Therefore the optimal point density function for quantizing U_j is given by plugging in Equation (6) into Equation (5), producing

$$\lambda_j(u) = \frac{\left(\left(\frac{d}{du}F^{-1}(u)\right)^2 u^{n-1}\right)^{1/3}}{\int \left(\left(\frac{d}{du}F^{-1}(u)\right)^2 u^{n-1}\right)^{1/3} du}. \tag{7}$$

From now on define the normalizing constant in the denominator of Equation (7) as c, i.e. $\int \left(\left(\frac{d}{du}F^{-1}(u)\right)^2 u^{n-1}\right)^{1/3} du = c$. The point density given in Equation (7) is for the underlying uniform random variable. Since $\lambda_j(u)$ is a proper density, the point density function for X_j can be found by a transformation of random variables. Let Z be a random variable with density $\lambda(u)$ given by Equation (7). Let W be a random variable defined as $W = F^{-1}(Z)$. Then the density of W, $f_W(w)$, is the optimal point density for X. Proceeding with the transformation of random variables produces the following distribution function for W:

$$\begin{aligned} F_W(x) &= \Pr(W \leq x) = \Pr(F^{-1}(Z) \leq x) = \Pr(Z \leq F(x)) \\ &= \frac{1}{c}\int_0^{F(x)} \left(\frac{d}{du}F^{-1}(u)\right)^{2/3} u^{\frac{n-1}{3}} du. \end{aligned} \tag{8}$$

To find the density function of W, differentiate the distribution function in the above equation. Utilizing the Fundamental Theorem of Calculus yields the desired density:

$$\begin{aligned} f_W(x) &= \frac{d}{dx}\frac{1}{c}\int_0^{F(x)} \left(\frac{d}{du}F^{-1}(u)\right)^{2/3} u^{\frac{n-1}{3}} du \\ &= \frac{1}{c}f(x)\left(\left[\frac{d}{du}F^{-1}(u)\right]_{u=F(x)}\right)^{2/3} F^{\frac{n-1}{3}}(x), \end{aligned} \tag{9}$$

where $f(x) = \frac{d}{dx}F(x)$ is the density function of X. By the Inverse Function Theorem, $\left[\frac{d}{du}F^{-1}(u)\right]_{u=F(x)} = \frac{1}{f(x)}$. Plugging this into Equation (9) produces the optimal point density function for X_j:

$$\lambda_j(x) = \frac{1}{c}f^{\frac{1}{3}}(x)F^{\frac{n-1}{3}}(x) \text{ for } i = 1, \dots, n. \tag{10}$$

This is equivalent to Equation (2) when all of the terms independent of x are absorbed into the normalizing constant, which is finite since the X_i satisfy Lemma 1. ■

The results of Theorem 1 can be extended to find the optimal fixed-rate point density function of any order statistic when the observations are distributed. This extension is stated in the following corollary:

Corollary 2: Let the distributed random variables X_i satisfy the same conditions as Lemma 1 and be continuous with density f. Let $f_{(r)}(x)$ be the density of the r^{th} order statistic. Then the optimal point density function for the fixed-rate quantizer Q_i under the MSE distortion for computing the r^{th} order statistic under distributed observations is given by

$$\lambda_i(x) = \frac{\left(f_{(r)}(x)\right)^{\frac{1}{3}}}{\int \left(f_{(r)}(x)\right)^{\frac{1}{3}} dx} \text{ for } i = 1, \dots, n. \tag{11}$$

Proof: The proof follows the same argument as Theorem 1. One user has to experience the r^{th} largest value, and quantizing that value with the point density function given in Equation (11) is optimal. At the central controller, the codeword associated with the r^{th} largest value is selected. It is possible multiple codewords that are fed back lie in the same quantization region as the r^{th} order statistic, but this results in the same distortion. ■

If the distritubtion and density functions of the underlying X_i are known and given by F and f respectively, then from [4], it is well known that the density for the r^{th} order statistic is given by

$$f_{(r)}(x) = \frac{1}{B(r, n-r+1)} F^{r-1}(x) \left[1 - F(x)\right]^{n-r} f(x), \tag{12}$$

where $B(a, b)$ is the beta function. The results of Theorem 1 and Corollary 2 can be extended even further. At the central controller, perhaps a function of the order statistic of interest is to be computed. For example, in the scheduling algorithm of [1], the ultimate quantity of interest is $\log(1 + \max_{1 \leq i \leq n} X_i)$, which is a function of the maximum order statistic and represents the maximum average rate that can be supported. More generally, let the function of the r^{th} order statistic to be computed be denoted $g(X_{(r)})$. The following corollary shows that as long as g is monotonic, optimal point densities can be found in the distributed setting.

Corollary 3: Let g be a continuous montonic function and $X_1, \ldots, X_n$ be distributed i.i.d continuous random variables with distribution F and density f such that $E\left[|g(X_i)|^{2+\epsilon}\right] < \infty$ for some $\epsilon > 0$. Let $g(X_{(r)})$ be computed at the central controller. If g is increasing, then the optimal point density for the high-resolution quantizers Q_i is given by

$$\lambda_i(x) = \frac{\left(f_{g_{(r)}}(x)\right)^{\frac{1}{3}}}{\int \left(f_{g_{(r)}}(x)\right)^{\frac{1}{3}} dx} \text{ for } i = 1, \ldots, n, \tag{13}$$

where $f_{g_{(r)}}(x)$ is the distribution of the r^{th} order statistic of $g(X_1), \ldots, g(X_n)$. If g is decreasing, then the optimal point density for the high-resolution quantizers Q_i is given by

$$\lambda_i(x) = \frac{\left(f_{g_{(n-r+1)}}(x)\right)^{\frac{1}{3}}}{\int \left(f_{g_{(n-r+1)}}(x)\right)^{\frac{1}{3}} dx} \text{ for } i = 1, \ldots, n. \tag{14}$$

Proof: First, suppose that g is monotonically increasing. Let $Y_i = g(X_i)$ be computed at each user. Because g is monotonically increasing, the r^{th} order statistic of the Y_i, $Y_{(r)}$ is given by $g(X_{(r)})$. Thus the result follows by application of Corollary 2 applied to the Y_i, which have densities f_g. These densities exists since g is monotonic and continuous. This yields Equation (13). The point density exists since $E\left[|g(X_i)|^{2+\epsilon}\right] < \infty$ for some $\epsilon > 0$ and by the arguments used in the proof of Lemma 1.

Now suppose that g is monotonically decreasing. Once again let $Y_i = g(X_i)$ be computed at each user. Since g is monotonically decreasing, the r^{th} order statistic of the Y_i, $Y_{(r)}$ is given by $g(X_{(n-r+1)})$ rather than $g(X_{(r)})$ as in the increasing case. Thus the result follows exactly as before but replacing the density $f_{g_{(r)}}$ in the point density expression with $f_{g_{(n-r+1)}}$, which produces Equation (14). ■

The main result of this section is that in the distributed setting, optimal fixed-rate point densities for quantizing the distributed values correspond to the point densities of the order statistic distributions in the MSE sense when the central controller is trying to compute a monotonic function of an order statistic. The previous results are for the high-resolution case, but this relationship allows for the construction of high performance finite-rate codebooks. The Lloyd-Max algorithm ([8],[9]) can be used to design an optimal (or locally optimal) low-rate quantizer by training the algorithm on the distribution of the order statistic of interest. The resulting quantizer is then applied to all the distributed users. Thus, in the fixed-rate quantizer regime, low and high-resolution quantizers can be designed for the computation of order statistics, or montonic functions of order statistics, at a central controller when the observations at the distributed users are i.i.d.

IV. Entropy-Constrained Quantizers

In entropy-constrained quantizers, the number of codepoints is not a priori fixed as in the fixed-rate quantizer case. Instead, the entropy induced by the quantizer partitions is constrained to be below some rate R, and the goal is to minimize the distortion given this constraint. In this case, the arguments used in finding the optimal fixed-rate point density functions in the distributed setting fail. To see this, imagine trying to design an entropy-constrained quantizer for the maximum of n i.i.d. uniform random variables. In Section III, the fixed-rate point density function corresponds to the distribution of the maximum since one user is guaranteed to experience the maximum, and the point density function is optimal for that user. If one designs the optimal entropy-constrained quantizer based on the distribution of the maximum order statistic, then on average, the quantizer will be optimal $\frac{1}{n}$ percent of the time, but suboptimal $1-\frac{1}{n}$ percent of the time due to the permutation invariance of order statistics under i.i.d. assumptions ([4]). In the $1-\frac{1}{n}$ percent of the time, the entropy constraint is violated since the codeword lengths were designed for the maximum order statistic, and now longer codewords become more probable and thus the entropy increases. A new approach must be taken to design the optimal entropy-constrained quantizer under distributed observations.

The approach is a modification to a result found in [3] identical to that used in proving Corollary 1. In [3], the optimal variable-rate point density function is shown for a specific class of random variables and functions of those random variables. The class of random variables of interest are the random variables $X_1, \dots, X_n$ with joint density $f_{X_1^n}(x_1, \dots, x_n)$ defined on $[0,1]^n$ such that the density is smooth enough so that the density can be approximated as constant over the joint quantization cells. The class of functions on these random variables that are analyzed are functions $g : \mathbb{R}^n \to \mathbb{R}$ that are continuous in each variable, with bounded gradient when defined, and whose first and second derivatives are defined except on a set of zero Jordan measure. Under these conditions, and using Definition 1, the optimal entropy-constrained point density function is given by the following theorem in [3]:

Theorem 3: (Theorem 14 in [3]) Under the assumptions previously mentioned, the optimal point density for the j^{th} entropy-constrained quantizer is given by

$$\lambda_j = \frac{\gamma_j(x)}{\int \gamma_j(t)dt} \tag{15}$$

To go from random variables defined on $[0,1]^n$ to random variables with more general supports requires an unnormalized point density function for the j^{th} user which is denoted $\Lambda_j(x)$. From Theorem 3, the unnormalized point density is given as $\Lambda_j(x) = \gamma_j(x)$. The need for unnormalized point density for entropy-constrained quantizers is well known ([5],[10]) for random variables with unbounded support. Having established the unnormalized point density, the following corollary extends Theorem 3 for order statistics of a class of i.i.d. random variables:

Corollary 4: Let $X_1, \dots, X_n$ satisfy the same conditions as Corollary 1. The optimal unnormalized point density function for a variable-rate quantizer to minimize the MSE distortion of the computation of the r^{th} order statistic is

$$\Lambda_j(x) = \left(\frac{1}{B(r, n-r+1)}\right)^{1/2} \cdot (F(x))^{\frac{r-1}{2}} (1 - F(x))^{\frac{n-r}{2}} \text{ for } j = 1, \dots, n. \tag{16}$$

Proof: Let $U_1, \dots, U_n$ be i.i.d. uniform random variables. Since the X_i satisfy the same conditions as Corollary 1, as in that corollary a valid selection of the function g is $g(U_1, \dots, U_n) = F^{-1}(U_{(r)})$, where $U_{(r)}$ is the r^{th} order statistic of the i.i.d. uniform random variables. $F^{-1}(U_j)$ has the same distribution as X_j for all j, and since F^{-1} is monotonic, $F^{-1}(U_{(r)}) =_d X_{(r)}$. To find the unnormalized point density function, the functional sensitivity profile must first be computed. Due to the i.i.d. nature of the random variables, for notational simplicity consider the first user without loss of generality. The partial derivative of $g(U_1, \dots U_n)$ is $\frac{d}{dx} F^{-1}(x)|_{x=U_1}$ when U_1 is the r^{th} order statistic, and zero otherwise. Therefore

$$\begin{aligned}
\gamma_1(u) &= \left(E\left[\left|\frac{\partial}{\partial U_1} g(U_1, \dots, U_n)\right|^2 |U_1 = u\right]\right)^{1/2} \\
&= \left(\left(\frac{d}{dx} F^{-1}(x)|_{x=u}\right)^2 \left(\Pr\left[g(U_1, \dots, U_n) = F^{-1}(U_1) \middle| U_1 = u\right]\right)\right)^{1/2} \\
&= \left(\frac{d}{dx} F^{-1}(x)|_{x=u}\right) \left(\frac{1}{B(r, n-r+1)} u^{r-1} (1-u)^{n-r}\right)^{1/2} \\
&= \left(\frac{d}{dx} F^{-1}(x)|_{x=u}\right) \left(\frac{1}{B(r, n-r+1)}\right)^{1/2} u^{\frac{r-1}{2}} (1-u)^{\frac{n-r}{2}}
\end{aligned} \tag{17}$$

where $\frac{1}{B(r,n-r+1)} u^{r-1} (1-u)^{n-r}$ is the probability that $r-1$ users have a smaller value than u and $n-r$ users have a value larger than u, thus making $U_1 = U_{(r)}$. The unnormalized point density for the entropy-constrained quantizer on the underlying uniform random variables is $\Lambda_{U_i}(x) = \gamma_1(x)$ from Equation (17). To get the unnormalized point density function for the random variables of interest, $X_1, \dots, X_n$, the same techniques used in Corollary 1 are employed, but utilizing unnormalized densities, yielding the following:

$$\begin{aligned}
\Lambda_1(x) &= \frac{d}{dx} \int_0^{F(x)} \left(\frac{d}{dt} F^{-1}(t)\right) \left(\frac{1}{B(r, n-r+1)}\right)^{1/2} t^{\frac{r-1}{2}} (1-t)^{\frac{n-r}{2}} dt \\
&= \left(\frac{1}{B(r, n-r+1)}\right)^{1/2} \cdot f(x) \cdot \frac{1}{f(x)} \cdot (F(x))^{\frac{r-1}{2}} \cdot (1 - F(x))^{\frac{n-r}{2}} \\
&= \left(\frac{1}{B(r, n-r+1)}\right)^{1/2} \cdot (F(x))^{\frac{r-1}{2}} \cdot (1 - F(x))^{\frac{n-r}{2}}
\end{aligned} \tag{18}$$

where, as before, going from the first to second equality uses the Inverse Function Theorem. Equation (18) is the conclusion of the corollary. ■

To see that this result differs from traditional, i.e. non-distributed, entropy-constrained scalar quantization, consider quantization of the maximum of n exponential random

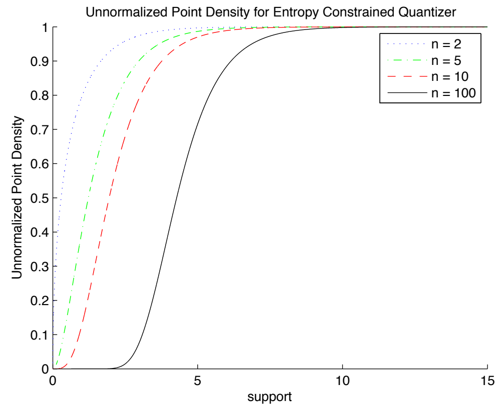

Fig. 1. Unnormalized point density function for the maximum of n exponential random variables

variables under the distributed and non-distributed cases. In the non-distributed case, the quantizer observes the n exponential random variables, selects the largest, and then quantizes it. From [10] and [11] it is known that regardless of the distribution, the optimal unnormalized point density function is constant. The quantity that is dependent on the distribution is the number of codepoints and partitions that can be selected to meet the given entropy constraint. In the distributed case, this quantizer cannot be used, because as mentioned at the beginning of the section, using the entropy-constrained quantizer designed based on the maximum order statistic distribution will violate the entropy constraint. From Corollary 4, the unnormalized point density function for the maximum of n exponential random variables is given by $\Lambda(x) = (1-e^{-x})^{\frac{n-1}{2}}$, which is not constant. Figure 1 shows the unnormalized point density for varying n. The interesting thing to notice is that, the larger n is, the further to the right the "mass" is. Since the point density function cannot be uniform due to the entropy constraint, the optimal entropy-constrained point density function puts more mass in the high values that the maximum is more likely to obtain while putting less mass in the more improbable low values to help save entropy and still meet the constraint. For large values of the argument, the point density function does eventually become uniform, and this is because for the maximum $\Lambda(x) = (F(x))^{\frac{n-1}{2}}$ and the distribution function eventually becomes arbitrarily close to unity.

Having established the unnormalized point density function for the entropy-constrained quantizer for order statistics, Corollary 4 can be extended to point densities of monotonic functions of order statistics akin to Corollary 3 in the fixed-rate quantization setting. The following corollary presents this extension:

Corollary 5: Let g be a continuous monotonic function and $X_1 \ldots, X_n$ be distributed i.i.d. random variables. Define $Y_i = g(X_i)$. Assume that Y_i has distribution F_Y that satisfies the same conditions as Corollary 1. If g is monotonically increasing, then the optimal unnormalized point density function for the j^{th} variable-rate quantizer to minimize the MSE distortion of the r^{th} order statistic is

$$\Lambda_j(x) = \left(\frac{1}{B(r, n-r+1)}\right)^{1/2} \cdot (F_Y(x))^{\frac{r-1}{2}} (1 - F_Y(x))^{\frac{n-r}{2}} \text{ for } j = 1, \ldots, n. \quad (19)$$

If g is monotonically decreasing, then the optimal unnormalized point density function for quantizing the r^{th} order statistic is:

$$\Lambda_j(x) = \left(\frac{1}{B(n-r+1,r)}\right)^{1/2} \cdot (F_Y(x))^{\frac{n-r}{2}} (1-F_Y(x))^{\frac{r-1}{2}} \text{ for } j=1,\ldots,n. \quad (20)$$

Proof: The results follow by making the same observation used in the proof of Corollary 3, that the r^{th} order statistic of the Y_i, $Y_{(r)}$, is equivalent to $g(X_{(r)})$ when g is monotonically increasing, and $Y_{(r)}$ is equivalent to $g(X_{(n-r+1)})$ when g is monotonically decreasing. Under the assumption that Y_i has distribution F_Y that satisfy the conditions of Corollary 1, allowing the use of Theorem 3, Equations (19) and (20) follow by plugging in the appropriate order statistic of Y into Corollary 4. ∎

V. CONCLUSION

The need to feedback CSI to increase the performance of wireless sytems motivates the question of how to optimally quantize this feedback. In a wireless system where each user feeds back its observed CSI to a centralized controller, it was shown that under the greedy scheduling algorithm, the optimal fixed-rate point density function of each user corresponds to the point density function designed based on the maximum order statistic. The optimal fixed-rate point density function was then found for arbitrary order statistics, and monotonic functions of the order statistics. These results imply a finite-rate codebook can be designed via the Lloyd-Max algorithm trained on the appropriate distribution. The optimal variable-rate point density functions are derived and are functions of the order statistic distributions. The variable-rate point density functions for distributed users differ from the standard high-resolution result of uniform point densities for the non-distributed case.

REFERENCES

[1] M. Sharif and B. Hassibi, "On the capacity of MIMO broadcast channel with partial side information," *IEEE Trans. Inf. Theory*, vol. 51, pp. 506–522, Feb. 2005.

[2] V. L. D. G. B. R. D.J. Love, R.W. Heath Jr. and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE Jour. on Sel. Areas in Comm.*, vol. 26, Oct. 2008.

[3] V. Misra, V. K. Goyal, and L. R. Varshney, "Distributed scalar quantization for computing: High-resolution analysis and extensions," 2010, arXiv:0811.3617v2.

[4] H. A. David, *Order Statistics*. New York: Wiley, 1970.

[5] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Kluwer Academic Publishers, 1992.

[6] J. A. Bucklew and G. L. Wise, "Multidimensional asymptotic quantization theory with rth power distortion measures," *Information Theory, IEEE Transactions on*, vol. 28, pp. 239 – 247, mar. 1982.

[7] S. Cambanis and N. Gerr, "A simple class of asymptotically optimal quantizers," *Information Theory, IEEE Transactions on*, vol. 29, pp. 664 – 676, sep. 1983.

[8] S. Lloyd, "Least squares quantization in pcm," *Information Theory, IEEE Transactions on*, vol. 28, pp. 129 – 137, mar. 1982.

[9] J. Max, "Quantizing for minimum distortion," *Information Theory, IRE Transactions on*, vol. 6, pp. 7 –12, mar. 1960.

[10] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE trans. inform. theoryq*, vol. 44, no. 6, pp. 2325–29, 1998.

[11] H. Gish and J. Pierce, "Asymptotically efficient quantizing," *Information Theory, IEEE Transactions on*, vol. 14, pp. 676 – 683, sep. 1968.