Sequential Functional Quantization

Vinith Misra
Department of Electrical Engineering
Stanford University
Email: vinith@stanford.edu

Krishnamurthy Viswanathan
Hewlett-Packard Company
Palo Alto, USA
Email: krishnamurthy.viswanathan@hp.com

***Abstract*—We consider the problem of lossy estimation of an arbitrary smooth function of correlated data in a stream. In this problem, a user sequentially observes correlated random variables and wants to construct an estimate of the specified function so that the mean squared estimation error is small. Techniques from high resolution quantization theory are applied and expanded for this problem, and the optimal distortion-rate exponent for companding quantization is determined. In the process, connections are established to sufficient statistics and to sensitivity matrices, as introduced by Linder et al. in the context of companding quantization under non-difference distortion measures. These results are applied to several example statistical functions, including the sample mean, sample variance, and the p-th order statistic.**

I. Introduction

We consider the problem of computing functions of streaming data under a memory constraint. In many modern computing problems, high-volume data streams need to be processed to answer queries in real time. The volume of data is often so large that one cannot store the entire data stream in memory. As such, a synopsis is often computed and stored instead. An information-theoretic analysis of the amount of memory required to losslessly compute a function of the data stream was presented in [1]. There it was shown that the problem was identical to that of source coding on the cascade network (see Fig. 1) where the synopses in the streaming problem correspond to the messages transmitted between nodes. In this paper, we consider the lossy analog of the problem; namely, when the function can be computed with error. Unlike [1] we do not consider the lossy problem in a Shannon-theoretic setting.

Shannon theoretic analysis of source coding systems can be precise, elegant, and meaningful in its results. When applied to multiterminal networks, however, this elegance frequently breaks down and many problems prove too difficult to solve. This is partly a result of information theory's traditional reliance on large-block asymptotics, which can be a liability when latency is important to the network, and partly a result of asking "too much" in the form of precise rate-distortion functions.

High-resolution quantization analysis is in some sense an alternative to rate-distortion theory. Rather than fixing the rate and allowing the blocklength to grow, one fixes the blocklength and allows the rate to grow. Closed-form distortion-rate functions are obtained in this manner for a variety of distortion measures. Despite its benefits, there has been relatively little use of high resolution quantization in the types of multiterminal source coding problems that resist Shannon theoretic analysis.

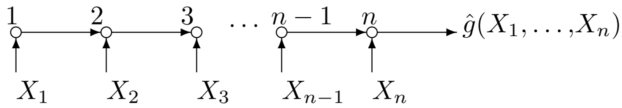

Fig. 1. Cascade network

In this work, to characterize the memory required for the streaming computation problem we examine the cascade functional source coding problem (Fig. 1). It involves a source X^n that is acquired sample-by-sample. After each acquisition, the system must compress all it has observed so far into a fixed amount of memory R. Once the last sample is acquired, an estimate $\widehat{g(X^n)}$ is produced for a function of the data $g(X^n)$. Note the absence of block coding in this problem description.

Naturally, the unique structure of this network complicates the analysis. Specifically, the cascade network introduces "requantization," wherein a source is compressed multiple times. At node 1 in Fig. 1, X_1 is compressed to $Q_1(X_1)$, and at node 2, the joint source $(Q_1(X_1), X_2)$ is compressed to $Q_2((Q_1(X_1), X_2)$. By focusing on the exponent in the distortion-rate function, high-resolution answers are kept clean and meaningful even in this nonstandard scenario.

The cascade network problem was proposed in [2] where the number of nodes $n = 2$. Inner and outer bounds were derived for the Shannon-theoretic rate region for this case. But the techniques are hard to extend to for larger values of n. A version of this problem when $n = 3$ and $g(X_1, X_2, X_3) = X_1$ was considered in [3]. In [4] an identical network is considered where the observed sources are independent noisy observations of an unknown random variable which is to be estimated. Similar network function computation problems were considered in [5], which involves a cascade communication setting with three nodes one of which is a relay, that has access to a degraded version of the side-information available at the receiver, and in [6] which also deals with degraded side-information but on a slightly different network. In [1] and [7] the Shannon-theoretic rate region for lossless computation on this network was derived. More results on the Shannon-theoretic rate region were derived in [8]. The streaming version of the problem has been studied extensively in computer science literature (see [9] and references therein). There the model assumed is of an adversarial data stream and the focus is on reconstructing specific functions (such as frequency moments) of discrete valued random variables upto a multiplicative constant.

High resolution quantization was first introduced in its most basic forms by Bennett in 1948, and thereafter elaborated by a variety of sources. Rather than referencing the original works, we direct readers to the detailed review by Gray and Neuhoff [10], and instead draw attention to more recent and relevant developments. Linder et al [11] carefully and rigorously analyze the performance of high-resolution companding quantizer structures under "locally quadratic" distortion measures. The functional distortion measures considered in the present paper are closely related to the locally quadratic class, and indeed the sensitivity matrix introduced by Linder et al plays an important role in Theorem 2. Furthermore, their companding quantizer approach to high-resolution analysis parallels the structures introduced in Sec. II-A.

Application of high resolution techniques to multiterminal settings has been limited. Zamir and Berger [12] solved the two-source lossy Slepian-Wolf problem with high resolution techniques, and in [13], a high-resolution version of Wyner and Ziv's result for lossy source coding with decoder side information is solved. More relevantly, in [14] the authors consider the high resolution distributed quantization problem for the purposes of function computation.

In Sec. II, the problem setting is rigorously defined and several important constructs are introduced. The primary results of this work are presented in Sec. III, followed by examples in Sec. IV. Proof details are provided in Sec. V.

II. Preliminaries

We first define some quantization related terminology.

A. Compander Quantization

It is relatively straightforward to analyze the high-resolution limits of uniform scalar quantization. More complex quantization strategies can be handled by viewing the quantization as a two-stage process. In the first stage, a "compressing" function is applied to data, and then a simple uniform quantization operation is performed on the compressed data. This view permits both flexibility in quantizer design and a mathematically rigorous description of the high-resolution limit.

The canonical uniform scalar quantizer $Q_U^1 : \mathbb{R} \to \mathbb{Z} + \frac{1}{2}$, where $\mathbb{Z} + \frac{1}{2}$ is the set $\{\ldots, -\frac{3}{2}, -\frac{1}{2}, \frac{1}{2}, \frac{3}{2}, \ldots\}$, is given by $Q_U^1(x) \stackrel{\text{def}}{=} \lfloor x \rfloor + 1/2$. The canonical m-dimensional square-lattice uniform vector quantizer may then be expressed as $Q_U^m(x^m) \stackrel{\text{def}}{=} \left(Q_U^1(x_1), \ldots Q_U^1(x_m)\right)$ for all $x \in \mathbb{R}^m$. To capture the high-resolution limit, a scale factor $\alpha \in \mathbb{R}^+$ and an associated scaled uniform vector quantizer $Q_\alpha^m(x^m) \stackrel{\text{def}}{=} \alpha Q_U^m\left(\frac{x^m}{\alpha}\right)$ are both introduced.

Observe that superior quantization lattices are usually possible for Q_U^m — for instance, the hexagonal lattice in two dimensions, or sphere-like cells as the dimension m grows large. However, the choice of lattice has no influence on the distortion exponent, which is the primary concern of this work.

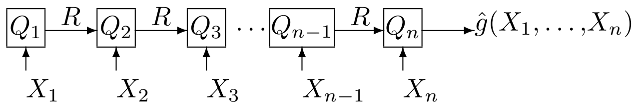

Fig. 2. Cascade quantization network

Definition 1: A compression function $F : \mathbb{R}^m \to \mathbb{R}^k$ is *admissible* if its Jacobian F' is defined and bounded everywhere, and if $k \leq m$.

The *companding quantizer* corresponding to compression F and scale α is given by $Q_{\alpha,F}(x^m) = Q_\alpha^k(F(x^m))$. Note that if $k \leq m$, reconstruction of the source X^m with diminishing distortion is not possible under traditional distortion measures. However, this is not necessarily true under functional distortion measures. For our problem, the more relevant question is whether the computation $g(X^n)$ can be reconstructed from the quantized values.

Definition 2: An estimation function $\widehat{g}(\cdot) : \mathbb{R}^k \to \mathbb{R}$ is *perfect* for a compression $F(x^n)$ if $g(X^n) = \widehat{g}(F(X^n))$ with probability one.

B. Problem Statement

The sequential functional quantization problem (Fig. 2) consists of a cascade network of n nodes. The first node quantizes the random variable X_1 and transmits the quantized value using R bits. Each subsequent node, i, $2 \leq i \leq n-1$, receives R bits from node $i-1$, observes X_i, computes a new message, and transmits R bits of data to node $i+1$. At the final node n, a function $g(X^n)$ is estimated.

Source Constraints: The source $X^n = X_1, X_2, \ldots, X_n$ is constrained to have a continuous density function with compact support, which is assumed without loss of generality to be contained in $[0,1]^n$.

Computation Constraints: A computation $g : \mathbb{R}^n \to \mathbb{R}$ is *admissible* if all its first and second derivatives,

$$g_i(x^n) \stackrel{\text{def}}{=} \frac{\partial}{\partial x_i} g(x^n); g_{i,j}(x^n) \stackrel{\text{def}}{=} \frac{\partial^2}{\partial x_i \partial x_j} g(x^n)$$

are defined and bounded for all i, j and all x^n

Compression Constraints: A companding quantization operation is performed at nodes 1 to $n-1$. This requires a choice of $n-1$ compression functions $F_1 : \mathbb{R} \to \mathbb{R}^{k_1}$ and for $1 < m < n-1$, $F_m : \mathbb{R}^{k_{m-1}} \times \mathbb{R} \to \mathbb{R}^{k_m}$, and $n-1$ scale factors α_m, $1 \leq m \leq n-1$. Node 1 computes an estimate $\widehat{F}_1$ of $F_1(x_1)$ to be $\widehat{F}_1(x_1) = Q_{\alpha_1,F_1}(x_1)$ and transmits to node 2. For $2 \leq m \leq n-1$, node m computes and transmits

$$\widehat{F}_m(x^m) = Q_{\alpha_m,F_{m-1}}\Big(\widehat{F}_{m-1}(x^{m-1}), x_m\Big).$$

The vectors of compressors and scale factors are denoted $\mathbf{F} = (F_1, F_2, \ldots, F_{n-1})$ and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_{n-1})$. Finally, node n uses an estimation function $g_{\mathbf{F}} : \mathbb{R}^{k_{n-1}} \times \mathbb{R} \to \mathbb{R}$ to compute the estimate of $g(x^n)$ to be $g_{\mathbf{F}}\Big(\widehat{F}_{n-1}(x^{n-1}), x_n\Big)$.

Distortion and Rate: For a given $\mathbf{F}$ and $\boldsymbol{\alpha}$, the distortion $D(\mathbf{F}, \boldsymbol{\alpha})$ in computation of the function g is measured in terms of mean-squared error i.e.

$$D(\mathbf{F},\boldsymbol{\alpha}) = \mathrm{E}\left[(g(X^n) - g_{\mathbf{F}}(\widehat{F}_{n-1}(X^{n-1}), X_n))^2\right]$$
$$\geq \mathrm{E}\left[\mathrm{var}\left(g(X^n)|\widehat{F}^{(n-1)}, X_n\right)\right],$$

where $\mathrm{var}(A|B)$ is the conditional variance of random variable A conditioned on B, and the lower bound is achieved by the MMSE estimator $\widehat{g}(\widehat{F}_{n-1}(X^{n-1}), X_n) = \mathrm{E}\left[g(X^n)|\widehat{F}_{n-1}(X^{n-1}), X_n\right]$.

We would like to minimize the distortion $D(\mathbf{F},\boldsymbol{\alpha})$ subject to a constraint on the output entropies of the first $n-1$ nodes:

$$H(\widehat{F}_m(X^m)) \leq R \text{ for all } m \in \{1,\ldots,n-1\}. \tag{1}$$

This represents a variable rate encoding of the node outputs $\{\widehat{F}_{(m)}\}_{m=1}^{n-1}$. The less-than-one bit gap between the performance of an optimized variable-rate source code and this entropic quantity is inconsequential to the rate exponent. Similarly, the potential benefits of employing decoder side information at the $(m+1)$-th node (X_{m+1} for transmitting $\widehat{F}^{(m)}$) can also be seen to at most allow for an extra $I(\widehat{F}^{(m)}; X_{m+1}) \leq I(X^m; X_{m+1})$ bits of rate, which also has no effect on the rate exponent. Note that the entropy constraint corresponds to the constraint on the memory in the streaming computation problem.

The above quantities may be combined to define a distortion-rate function for a given collection of compression functions $\mathbf{F}$. The scale factors $\boldsymbol{\alpha}$ are optimized to minimize the distortion subject to an entropy constraint on each quantizer:

$$D(\mathbf{F}, R) = \min_{\boldsymbol{\alpha}: H(\widehat{F}_m(X^m)) \leq R, m \in \{1,\ldots,n-1\}} D(\mathbf{F},\boldsymbol{\alpha}). \tag{2}$$

We now define our primary quantity of interest. For a given $\mathbf{F}$, the *distortion-rate exponent* is

$$D^*(\mathbf{F}) \stackrel{\text{def}}{=} \lim_{R\to\infty} -\frac{\log D(\mathbf{F}, R)}{R}.$$

C. Sufficient Statistics

The notion of sufficient statistics may be specialized to the case of this problem.

Definition 3: A differentiable mapping $T : \mathbb{R}^m \to \mathbb{R}^k$ defines an *m-sufficient statistic* $T(X^m)$ if $\mathrm{E}\left[\mathrm{var}\left(g(X^n)|T(X^m), X_{m+1}^n\right)\right] = 0$. This implies the existence of a map $g_T(t, x_{m+1}^n) : \mathbb{R}^k \times \mathbb{R}^{n-m} \to \mathbb{R}$ such that $g_T(T(X^m), X_{m+1}^n) = g(X^n)$ with probability one.

Definition 4: If furthermore the singular values of $T'(x^m)$, the Jacobian of T, can be bounded both from above and away from zero, T is *invertibly m-sufficient*.

Let $\mathcal{T}(g,m)$ be the set of invertibly m-sufficient statistics for $g(X^n)$, and let $\dim(T)$ yield the output dimensionality of a mapping T.

Definition 5: The *minimum sufficient dimension* of X^m is

$$\underline{k}_m = \min_{T\in\mathcal{T}(g,m)} \dim(T), \tag{3}$$

and T in $\mathcal{T}(g,m)$ is a *minimum sufficient statistic* of X^m if $\dim(T) = \underline{k}_m$.

D. Sensitivity Matrix

Sensitivity matrices are important in the quantification of asymptotic quantization performance. In the context of this particular problem, they measure the correlation between the computation's dependencies on each source variable.

Definition 6: For a given computation $g(X^n)$ and source distribution $f(X^n)$, the sensitivity matrix M^m of the first m samples X^m is an $m \times m$ matrix whose (i,j)-th entry is

$$M_{i,j}^{(m)}(x^m) = \mathrm{E}\left[g_i(X^n)g_j(X^n)|X^m = x^m\right], \tag{4}$$

where g_i denotes $\partial g(x^n)/\partial x_i$.

Suppose one defines a distortion measure to X^m as $d_g(x^m, y^m) = \mathrm{E}\left[(g(X^n) - g(Y^n))^2|X^m = x^m, Y^m = y^m\right]$. The matrix $M^{(m)}$ is then a special case of the sensitivity matrix M introduced by Linder et al. [11] for the locally quadratic distortion measure d_g.

III. Main Results

Our main result Theorems 1 and 2 relate the optimized distortion-rate exponent to the minimum sufficient statistics of $g(X^n)$ and the sensitivity matrices of $g(X^n)$ respectively. Statements are provided in this section; see Sec. V for proofs of these theorems.

Theorem 1: For all admissible computations g and admissible sources X^n

$$\max_{\mathbf{F}} D^*(\mathbf{F}) = \min_{m\in\{1,\ldots,n-1\}} 2\underline{k}_m^{-1} \tag{5}$$

To interpret this result, observe that the distortion-rate exponent for uniformly quantizing a distribution in $\mathbb{R}^k$ has rate exponent $2/k$. This exponent is therefore tied to the dimensionality of the quantity being quantized. Effectively, Theorem 1 states that the quantization stage(s) with the largest minimum sufficient dimension dominates the error. In other words, re-quantization has no effect on the rate exponent.

Theorem 2: Suppose that for a given computation $g(x^n)$ and any $m \in \{1,\ldots,n-1\}$, the ranks of the sensitivity matrices are constants that do not depend on x^m, *i.e.*, $\forall x^m, \mathrm{rank}(M_m(x^m)) = r_m$,. Then for any m-sufficient statistic T_m,

$$r_m \leq \dim(T_m), \tag{6}$$

and therefore by Theorem 1,

$$\max_{\mathbf{F}} D^*(\mathbf{F}) \leq \min_{m\in\{1,\ldots,n-1\}} 2r_m^{-1}. \tag{7}$$

While this result is presented as lower-bounding the dimensionality of minimum sufficient statistics, it is frequently satisfied with equality. As such, one can often verify whether a sufficient statistic is minimal by comparing its dimension to the rank of M_m.

IV. Examples

For several example functions, sufficient statistics are constructed and sensitivity matrices are computed. By comparing the dimensionality of the former with the rank of the latter, the distortion exponent can sometimes be determined.

A. Sample mean

The function to be computed is the sample mean of the observed random variables, *i.e.*, $g(x^n) = \frac{1}{n}\sum_{i=1}^n x_i$. For all $m \le n$, $T(x^m) = \sum_{i=1}^m x_i$ is m-sufficient for g, and has dimensionality 1. It is easy to see that $M_m(x^m)$ is the all-1 matrix and the rank is therefore 1. It therefore follows that the optimum distortion-rate exponent is $\mathbf{2}$.

B. Sample variance

The function to be computed is the sample variance of the observed random variables, *i.e.*,

$$g(x^n) = \frac{1}{n-1}\sum_{i=1}^n \left(x_i - \frac{1}{n}\sum_{j=1}^n x_j\right)^2$$

It can be verified that for all $k \le n$, $T_m(x^m) = \left(\sum_i x_i, \sum_i x_i^2\right)$ is m-sufficient for g. Therefore, by Theorem 2, the rank of the sensitivity matrix $M_m(x^m)$ is at most 2. Since the rank is clearly greater than 1, it therefore follows that the distortion exponent is $\mathbf{1}$.

C. Order Statistics

In this subsection, we consider computation of order statistics, including the median, the largest element, and the smallest element. Specifically, if the function of interest is $g(x^n) = x_{(p)}$, the p-th smallest of the numbers $x_1, x_2, \ldots, x_n$, then for $m \le n-p$, the smallest p numbers in the set $\{x_1, x_2, \ldots, x_m\}$ are m-sufficient for $g(\cdot)$. For $m > n-p$, the $p-(n-m)$-th smallest to the p-th smallest numbers are sufficient for $g(\cdot)$.

This yields a possible distortion exponent of p, but one may also track the $(n-p+1)$-th largest numbers, which yields a distortion exponent of $n-p+1$. It therefore follows that the distortion-rate exponent is at least $2(\min\{p, n-p+1\})^{-1}$.

V. Proof of Results

A. Proof Outline of Theorem 1

Proof: For convenience, we introduce the notation $F_m^{(x)}(x^m) = F_m(F_{m-1}^{(x)}(x^{m-1}), x_m)$ for the evaluation of an entire compression chain up to node m, in the absence of quantization error.

Achievability. Let $\{T_m\}_{m=1}^{n-1}$ be a sequence of minimum m-sufficient statistics. With the following lemma, we are able to construct from $\{T_m\}_{m=1}^{n-1}$ a vector of compressors $\mathbf{F}$ and a perfect estimation function $g_\mathbf{F}$.

Lemma 3: If $\{T_m\}_{m=1}^{n-1}$ is a sequence of sufficient statistics, there exists a vector $\mathbf{F}$ of compressors with bounded derivatives and dimensionality $\dim(F_m) = \dim(T_m)$, and a perfect estimation function $g_\mathbf{F}$ with bounded derivatives.

Proof: First, recall that by definition for all $m \ge 2$, the following two properties hold:

1) T'_{m-1}, the Jacobian of T_m is well-defined and bounded.
2) The singular values of T'_{m-1} possess a strictly positive lower bound. Let $\underline{\sigma}_{m-1}$ be one such lower bound.

As a result of these properties, there exists a mapping $P_{m-1}: \mathbb{R}^k \to \mathbb{R}^{m-1}$ such that for all t_{m-1}, $P_{m-1}(t_{m-1}) \in T_{m-1}^{-1}(t_{m-1})$ and P'_{m-1} is bounded. Now, for $1 \le m \le n$, define the compression function

$$F_m(f_{m-1}, x_m) = T_m(P_{m-1}(f_{m-1}), x_m). \quad (8)$$

Notice first that by construction $\dim(F_m) = \dim(T_m)$. Furthermore, since T'_m and P'_{m-1} are both bounded and well-defined, F'_m is as well. Define $g_\mathbf{F}(f_{n-1}, x_n) = F_n(f_{n-1}, x_n)$. The derivatives of $g_\mathbf{F}$ are also bounded and well-defined. Further, $g_\mathbf{F}(F_{n-1}^{(x)}(x^{n-1}), x_n) = F_n(P_{n-1}(F_{n-1}^{(x)}(x^{n-1})), x_n) = g(x^n)$, and therefore $g_\mathbf{F}$ is a perfect estimation function. ■

To bound the distortion, we bound the worst-case error $e(R, \mathbf{F})$ in estimating $g(x^n)$. Let C_g denote the bound on the derivative of $g_\mathbf{F}$ from the preceding lemma. Then

$$\begin{aligned} e(R,\mathbf{F}) &= \max_{x^n} \left| g(x^n) - g_F(\widehat{F}_{n-1}(x^{n-1}), x_n)\right| \\ &= \max_{x^n} \left| g_\mathbf{F}(F_{n-1}^{(x)}(x^{n-1}), x_n) - g_\mathbf{F}(\widehat{F}_{n-1}(x^{n-1}), x_n)\right| \\ &\le C_g \max_{x^{n-1}} \left\| F_{n-1}^{(x)}(x^{n-1}) - \widehat{F}_{n-1}(x^{n-1})\right\|_1 \quad (9) \end{aligned}$$

where the equality follows from the perfectness of $g_\mathbf{F}$. To refine (9), let C_F be the upper bound on the derivatives of the compressor functions. Then for $m \in \{1, \ldots, n-1\}$, the worst-case error at the mth stage can be geometrically bounded by the sum of the (scaled) worst case error at the $(m-1)$th stage and the ℓ_1 error from a uniform quantization:

$$\begin{aligned} E_m &\stackrel{\text{def}}{=} \max_{x^{m-1}} \left\| F_{m-1}^{(x)}(x^{m-1}) - \widehat{F}_{m-1}(x^{m-1})\right\|_1 . \\ &\le C_F E_{m-1} + \sqrt{m} 2^{-R/\underline{k}_m} \le \sum_{i=1}^{m-1} \sqrt{i} C_F^{m-i} 2^{-R/\underline{k}_i}, \quad (10) \end{aligned}$$

where recall that $\underline{k}_i = \dim(T_i)$. Combining (9) with (10) one can bound the distortion to obtain that

$$D^*(\mathbf{F}) \ge \min_{m \in \{1,\ldots,n-1\}} 2\underline{k}_m^{-1}.$$

Converse. Let $\mathbf{F}$ be an arbitrary admissible chain of compression functions, and let $g_\mathbf{F}$ be an associated estimation function.

By the law of total variance,

$$\begin{aligned} D(\mathbf{F}, R) &\ge \mathrm{E}\left[\mathrm{var}\left(g(X^n) | \widehat{F}_{n-1}(X^{n-1}), X_n\right)\right] \\ &\ge \mathrm{E}\left[\mathrm{var}\left(g(X^n) | \widehat{F}_{m-1}(X^{m-1}), X_m^n\right)\right] \\ &\ge \mathrm{E}\left[\mathrm{var}\left(g(X^n) | F_{m-1}^{(x)}(X^{m-1}), X_m^n\right)\right], \end{aligned}$$

which is a strictly positive lower bound unless $F_m^{(x)}$ is m-sufficient. Thus $F_m^{(x)}(x^m)$ must be m-sufficient for the distortion $D(\mathbf{F}, R)$ to vanish with R. One may similarly show that g_F must be a perfect estimation function. We may therefore assume that $\{\mathbf{F}\}$ and $g_\mathbf{F}$ satisfy these constraints. As a result, $\dim(F_m) \ge \underline{k}_m$ for every m in $\{1, \ldots, n-1\}$.

Consider the quantity

$$D_m(F_m, R) = \mathrm{E}\left[\mathrm{var}\left(g(X^n) | Q_{\alpha, F_m}(X^m), X_{m+1}^n\right)\right],$$

which is the distortion when infinite quantization rate is permitted at every node other than node m. It is easily shown with the law of total variance that $D_m(F_m, R) \le D(\mathbf{F}, R)$. In the rest of this outline, we sketch the analysis that produces an asymptotically valid lower bound for $D_m(F_m, R)$.

Since the second derivatives $g_{i,j}$ and the Jacobian F'_m are bounded, one may lower bound the variation of g within most quantization cells. For sufficiently small α (and correspondingly large R),

$$D_m(F_m, R) \ge \beta \mathrm{E}\left[\mathrm{var}\left(F_m(X^m) | Q_{\alpha,F_m}(F_m(X^m))\right)\right],$$

with $\beta > 0$. Again invoking the boundedness of F'_m, one may lower bound this latter quantity for sufficiently small α (sufficiently large R) as

$$D_m(F_m, R) \ge \gamma \alpha^{-2}. \tag{11}$$

To relate the parameter α with the rate $R = H(Q_{F_m,\alpha})$, we use the high-resolution rate approximation given in Proposition 2 from [11]. This relation only applies for one-to-one compression functions, so rather than applying it to source X^m with compressor F_m, we apply it to source $F_m(X^m)$ with the identity compressor. Note that this is permissible because, once again, the Jacobian of the transformation F'_m is bounded. Denoting by C a real-valued constant, the result states that

$$R + \dim(F_m) \log \alpha \to_{\alpha \to 0} C. \tag{12}$$

The converse is proved by combining (11) with (12), and lower bounding $D(R, \mathbf{F})$ with $D_m(R, F_m)$, *i.e.*, for all $m \in \{1, \dots, n-1\}$, $D^*(\mathbf{F}) \ge 2(\dim(F_m))^{-1} \ge 2(\underline{k}_m)^{-1}$. ■

B. Proof of Theorem 2

Proof: Let $p = \dim(T_m)$. If $p = m$, the theorem is trivially true. Let $p < m$. For $1 \le i \le p$, let $T_i(x^m)$ denote the the i-th component of $T(x^m)$. Also let T^p abbreviate the vector $(T_1, T_2, \dots, T_p)$. By the chain rule

$$\begin{aligned} g_i(x_1^n) &= \frac{\partial}{\partial x_i} g_T\big(T_1(x_1^m), T_2(x_1^m), \dots, T_p(x_1^m), x_{m+1}^n\big) \\ &= \sum_{a=1}^{p} \frac{\partial}{\partial T_a} g_T\big(T^p, x_{m+1}^n\big) \frac{\partial}{\partial x_i} T_a(x^m) \end{aligned}$$

where for $1 \le i \le p$, $T_i = T_i(x^m)$. Then observe that

$$\begin{aligned} &M_m(x^m)[i,j] \\ &= \sum_{a=1}^{p} \mathrm{E}\left[\frac{\partial}{\partial T_a} g_T\big(T^p, X_{m+1}^n\big) g_i\big(x_1^m, X_{m+1}^n\big) \middle| X^m = x^m\right] \\ &\qquad \times \frac{\partial}{\partial x_i} T_a(x^m). \end{aligned} \tag{13}$$

For $\ell \le p$, let the ℓ-th row of $M_m(x^m)$ be $M_m(x^m)[\ell, :]$. We will show that there exists $\gamma_1, \dots, \gamma_{p+1}$, not all 0, such that

$$\sum_{\ell=1}^{p+1} \gamma_\ell M_m(x^m)[\ell, :] = \bar{0} \tag{14}$$

where $\bar{0}$ is a $1 \times k$ vector all of whose components are 0. It then follows that $\mathrm{rank}(M_m(x^m)) \le p$. Let N be a $p+1 \times p$ matrix whose (i,j)-th entry, $1 \le i \le p+1$, $1 \le j \le p$ is

$$N(i,j) = \frac{\partial}{\partial x_i} T_j(x^m). \tag{15}$$

Since $\mathrm{rank}(N) \le p$, there exists $\gamma_1, , \dots, \gamma_{p+1}$, not all 0, such that $\sum_{\ell=1}^{p+1} \gamma_\ell N(\ell, :) = 0$ where $N(\ell, :)$ is the ℓ-th row of N. Substituting for N from (15), it follows that for all $1 \le j \le p$

$$\sum_{\ell=1}^{p+1} \gamma_\ell \frac{\partial}{\partial x_\ell} T_j(x^m) = 0. \tag{16}$$

Therefore for all $1 \le i \le m$ and all $1 \le j \le p$

$$\begin{aligned} &\sum_{\ell=1}^{p+1} \gamma_\ell E\left[\frac{\partial}{\partial T_j} g_T\big(T^p, X_{m+1}^n\big) g_i\big(x_1^m, X_{m+1}^n\big)\right] \frac{\partial}{\partial x_\ell} T_j(x^m) \\ &= E\left[\frac{\partial}{\partial T_j} g_T\big(T^p, X_{m+1}^n\big) g_i\big(x_1^m, X_{m+1}^n\big)\right] \times \sum_{\ell=1}^{p+1} \gamma_\ell \frac{\partial}{\partial x_\ell} T_j(x^m) \\ &= 0 \end{aligned}$$

where the second equality follows from (16). Summing both sides over j, we obtain that for all $1 \le i \le m$

$$\begin{aligned} &\sum_{j=1}^{p} \sum_{\ell=1}^{p+1} \gamma_\ell E\left[\frac{\partial}{\partial T_j} g_T\big(T^p, X_{m+1}^n\big) g_i\big(x_1^m, X_{m+1}^n\big)\right] \frac{\partial}{\partial x_\ell} T_j(x^m) \\ &= \sum_{\ell=1}^{p+1} \gamma_\ell \sum_{j=1}^{p} E\left[\frac{\partial}{\partial T_j} g_T\big(T^p, X_{m+1}^n\big) g_i\big(x_1^m, X_{m+1}^n\big)\right] \frac{\partial}{\partial x_\ell} T_j(x^m) \\ &= \sum_{\ell=1}^{p+1} \gamma_\ell M_m(x^m)[\ell, i] = 0 \end{aligned}$$

where the second equality follows from (13). This proves (14). ■

References

[1] K. Viswanathan, "On the memory required to compute functions of streaming data," in *Proc. of ISIT*, 2010.

[2] P. Cuff, H. Su and A. El Gamal, "Cascade multiterminal source coding," in *Proc. of IEEE Symp. on Info. Theory*, pages 1199–1203, 2009.

[3] P. Ishwar and S. S. Pradhan, "A Relay-assisted distributed source coding problem," in *Proc. of ITA Workshop*, 2008.

[4] S. C. Draper and G. W. Wornell, "Side information aware coding strategies for sensor networks," in *IEEE JSAC*, 2004.

[5] D. Vasudevan, C. Tian and S.N. Diggavi, "Lossy source coding for a cascade communication system with side-informations," in *Proceedings of Allerton Conference*, Illinois, September 2006.

[6] W.H. Gu and M. Effros, "On multi-resolution coding and a two-hop network," in *Data Compression Conference*, 2006.

[7] K. Viswanathan, "Information-theoretic analysis of function computation on streams" in Proc. Allerton Conf., 2010.

[8] M. Sedgaran and A. Tchamkerten, "On function computation over a cascade network," in Proc. of Info Th. Workshop, 2012.

[9] S. Muthukrishnan, "Data Streams: Algorithms and Applications," *Foundations and Trends in Theoretical Computer Science*, 1:2, 2005.

[10] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. of Info. Th.*, vol. 44, no. 6, pp. 2325–2383, Oct. 1998.

[11] T. Linder, R. Zamir, and K. Zeger, "High-resolution source coding for non-difference distortion measures: Multidimensional companding," *IEEE Trans. of Info. Th.*, vol. 45, no. 2, pp. 548–561, Mar. 1999.

[12] R. Zamir and T. Berger, "Multiterminal source coding with high resolution," *IEEE Trans. of Info. Th.*, vol. 45, no. 1, pp. 106-117, Jan. 1999.

[13] D. Rebollo-Monedero, S. Rane, A. Aaron, and B. Girod, "High-rate quantization and transform coding with side information at the decoder,"*Signal Processing*, vol. 86, no. 11, pp. 3160–3179, Nov. 2006.

[14] V. Misra, V. Goyal, and L. Varshney, "Distributed scalar quantization for computing: High-resolution analysis and extensions," *IEEE Trans. of Info. Th.*, vol. 57, no. 8, pp. 5298 –5325, aug. 2011.