On MMSE Estimation from Quantized Observations in the Nonasymptotic Regime

Jaeho Lee, Maxim Raginsky, and Pierre Moulin

Abstract—This paper studies MMSE estimation on the basis of quantized noisy observations. It presents nonasymptotic bounds on MMSE regret due to quantization for two settings: (1) estimation of a scalar random variable given a quantized vector of n conditionally independent observations, and (2) estimation of a p-dimensional random vector given a quantized vector of n observations (not necessarily independent) when the full MMSE estimator has a subgaussian concentration property.

Index Terms—MMSE estimation, vector quantization, indirect rate distortion problems, nonasymptotic bounds.

I. INTRODUCTION

Minimum mean-square error (MMSE) estimation is a fundamental primitive in communications, signal processing, and data analytics. In today's applications, the estimation task is often performed on high-dimensional data collected, and possibly preprocessed, at multiple remote locations. Consequently, attention must be paid to communication constraints and their impact on the MMSE.

One strategy for reducing the communication burden is to compress the observations using vector quantization (VQ). The idea of quantization for estimation and control can be traced back to the work of Curry [1], who considered in detail the jointly Gaussian case, derived a modification of the Kalman filter for use with quantized inputs, and developed approximation-based schemes for nonlinear systems. Because VQ inevitably introduces loss, it is of interest to characterize the resulting *MMSE regret*, i.e., the difference between the optimal performance achievable with quantized observations and the optimal performance that can be attained without quantization. The problem of designing an optimal vector quantizer to minimize the MMSE regret is equivalent to a noisy source coding problem with the quadratic fidelity criterion, where the compressor acts on the vector of observations and the decompressor generates an estimate of the target random vector. This equivalence was systematically studied by Wolf and Ziv [2], who showed that there is no loss of optimality if we first compute the full MMSE estimate and then compress it using an optimal quantizer. Ephraim and Gray [3] later extended this result to a more general class of weighted quadratic distortion functions and gave conditions for convergence of the Lloyd-type iterative algorithm for quantizer design.

The main message of the above works is that the problem of quantizer design for minimum MMSE regret is a *functional compression* problem: the compressed representation of X should retain as much information as possible about the conditional mean $\eta(X) \triangleq \mathbb{E}[Y|X]$ of the target Y given X, also known as the *regression function*. However, there is an interesting tension between the information-theoretic and the *statistical* aspects of the problem: if the MMSE without quantization is sufficiently small, then the dominant contribution to the quantized MMSE should come from the quantization error; in the opposite regime of high-rate quantization, the quantized MMSE should not differ much from its unquantized counterpart.

In this paper, we study both the statistical and the functional-compression aspects of MMSE estimation with quantized observation. In particular, we obtain sharp upper bounds on the MMSE regret in two scenarios:

- MMSE estimation of a scalar random variable Y, where the vector $X = (X_1, \ldots, X_n)$ of conditionally i.i.d. observations is passed through a k-ary vector quantizer. In this setting, under mild regularity conditions on the conditional distribution $P_{X_1|Y}$, we obtain nonasymptotic bounds on the MMSE regret. These bounds exhibit two distinct behaviors depending on whether the number of observations n is larger than the square of the codebook size k^2. In the asymptotic regime of $n, k \to \infty$, we recover an existing result of Samarov and Has'minskii [6], who were the first to address this problem.
- MMSE estimation of a p-dimensional random vector Y when the n-dimensional vector of observations X (not necessarily independent) is passed through a k-ary VQ. In this setting, we derive an upper bound on the MMSE regret under the assumption that the ℓ_2 norm of the regression function exhibits subgaussian concentration around its expected value $\mathbb{E}\|\eta(X)\|$. Unlike some of the existing literature on high-resolution quantization for functional compression (e.g., [4]), we do not require smoothness of $\eta(X)$.

Notation. We will always denote by $\|\cdot\|$ the ℓ_2 (Euclidean) norm. A *k-ary quantizer* on the Euclidean space $\mathbb{R}^d$ is a Borel-measurable mapping $q: \mathbb{R}^d \to [k]$, where $[k]$ is shorthand for the set $\{1, \ldots, k\}$. Any such q is characterized by its cells or bins, i.e., the Borel sets $C_j = q^{-1}(\{j\}) = \{v \in \mathbb{R}^d : q(v) = j\}$, $j \in [k]$. The set of all k-ary quantizers on $\mathbb{R}^d$ will be denoted by $\mathcal{Q}_k^d$. A *k-ary reconstruction function* on $\mathbb{R}^d$ is a mapping $f: [k] \to \mathbb{R}^d$. Any such f is characterized by its

reconstruction points $c_j = f(j) \in \mathbb{R}^d$, $j \in [k]$. The set of all k-ary reconstruction functions on $\mathbb{R}^d$ will be denoted by $\mathcal{R}_k^d$. Any finite set of points $\mathcal{C} = \{c_1, \ldots, c_k\} \subset \mathbb{R}^d$ defines a quantizer $q_{\mathcal{C}} \in \mathcal{Q}_k^d$ and a reconstruction function $f_{\mathcal{C}} \in \mathcal{R}_k^d$ by

$$q_{\mathcal{C}}(v) = \arg\min_{j \in [k]} \|v - c_j\|, \qquad \forall u \in \mathbb{R}^d$$

and $f_{\mathcal{C}}(j) = c_j$ for all $j \in [k]$. The composite mapping $q_{\mathcal{C}} \triangleq f_{\mathcal{C}} \circ q_{\mathcal{C}} : \mathbb{R}^d \to \mathbb{R}^d$ is called a *k-point nearest-neighbor quantizer* with *codebook* $\mathcal{C}$. A well-known result is that, for any random vector V with $\mathbb{E}\|V\|^2 < \infty$,

$$\inf_{f \in \mathcal{R}_k^d} \inf_{q \in \mathcal{Q}_k^d} \mathbb{E}\|V - f(q(V))\|^2 = \inf_{\mathcal{C} \subset \mathbb{R}^d : |\mathcal{C}| = k} \mathbb{E}\|V - q_{\mathcal{C}}(V)\|^2,$$

and the infimum is actually a minimum. We refer the reader to the survey article by Gray and Neuhoff [5] for more details.

We use the following asymptotic order notation: for two sequences $\{a_m\}$ and $\{b_m\}$, we write $a_m \preceq b_m$ if $a_m = O(b_m)$, and $a_m \asymp b_m$ if $a_m \preceq b_m$ and $b_m \preceq a_m$.

II. PERFORMANCE CRITERIA AND SOME BASIC RESULTS

Suppose two random vectors X in $\mathbb{R}^n$ and Y in $\mathbb{R}^p$ are jointly distributed according to a given probability law P_{XY}. The MMSE in estimating Y as a function of X is

$$\mathsf{mmse}(P_{XY}) \triangleq \inf_f \mathbb{E}\|Y - f(X)\|^2,$$

where the infimum over all Borel-measurable functions $f : \mathbb{R}^n \to \mathbb{R}^p$ is achieved by the regression function $\eta(x) = \mathbb{E}[Y|X = x]$. We consider the problem of MMSE estimation of Y under the circumstances where only $q(X)$, a quantized version of X, is accessible. Thus, for each $k \in \mathbb{Z}^+$, we are interested in the MMSE functional

$$\mathsf{mmse}_k(P_{XY}) \triangleq \inf_{f \in \mathcal{R}_k^p} \inf_{q \in \mathcal{Q}_k^n} \mathbb{E}\|Y - f(q(X))\|^2.$$

This problem can be cast as one-shot fixed-rate lossy coding of Y with squared-error distortion when the encoder only has access to X (see, e.g., [2], [3]). We recall some known results that explicitly involve the regression function $\eta(X)$. The first one, due to Wolf and Ziv [2], is a useful decomposition of $\mathsf{mmse}_k(P_{XY})$:

Proposition 1. *For every $k \in \mathbb{Z}^+$,*

$$\mathsf{mmse}_k(P_{XY}) = \mathsf{mmse}(P_{XY}) + \mathsf{reg}_k(P_{XY}), \qquad (1)$$

where

$$\mathsf{reg}_k(P_{XY}) = \inf_{q \in \mathcal{Q}_k^n} \mathbb{E}\|\mathbb{E}[Y|X] - \mathbb{E}[Y|q(X)]\|^2$$

is the MMSE regret *due to quantization.*

Remark 1. A special case of this result for jointly Gaussian X and Y was obtained by Curry [1, Sec. 2.4].

The second result, which was proved by Ephraim and Gray [3] for a more general class of weighted quadratic distortion functions, shows that there is no loss of optimality if we restrict our attention to schemes of the following type: given X, we first compute the regression function $\eta(X)$, quantize it using a

k-ary quantizer, and then estimate Y by its conditional mean given the cell index of X.

Proposition 2.

$$\mathsf{reg}_k(P_{XY}) = \inf_{f \in \mathcal{R}_k^p} \inf_{q \in \mathcal{Q}_k^p} \mathbb{E}\|\eta(X) - f(q(\eta(X)))\|^2 \quad (2a)$$

$$= \inf_{\mathcal{C} \subset \mathbb{R}^p : |\mathcal{C}| = k} \mathbb{E}\|\eta(X) - q_{\mathcal{C}}(\eta(X))\|^2. \quad (2b)$$

III. THE CASE OF n CONDITIONALLY INDEPENDENT OBSERVATIONS

We now consider the situation when the coordinates $X_1, \ldots, X_n$ of X are independent and identically distributed (i.i.d.) conditionally on Y. To keep things simple, we consider the case when Y is a scalar random variable (i.e., $p = 1$), and its marginal distribution P_Y has a probability density function f_Y supported on a compact interval $\mathcal{Y} = [-A, A]$. In the asymptotic regime as $k \to \infty$ and $n \to \infty$, this setting was investigated by Samarov and Has'minskii [6], who showed the following under mild regularity conditions:

- If $n/k^2 \to \infty$, then the dominant contribution to $\mathsf{mmse}_k(P_{XY})$ comes from the minimum expected distortion

$$\inf_{\mathcal{C} \subset \mathcal{Y} ; |\mathcal{C}| = k} \mathbb{E}|Y - q_{\mathcal{C}}(Y)|^2 \asymp \frac{1}{k^2}\left(\int_{\mathcal{Y}} f_Y(y)^{1/3} \mathrm{d}y\right)^3$$

incurred on Y by any k-point quantizer [7], [8].
- If $n/k^2 \to 0$, then the dominant contribution to $\mathsf{mmse}_k(P_{XY})$ comes from

$$\mathsf{mmse}(P_{XY}) \asymp \frac{1}{n}\int_{\mathcal{Y}} \frac{f_Y(y)}{I(y)} \mathrm{d}y,$$

where $I(y)$ is the Fisher information in the conditional distribution $P_{X_1|Y=y}$ [9] (see below for definitions).

The proofs in [6] rely on the deep result, commonly referred to as the *Bernstein–von Mises theorem*, which says that the posterior distribution of the normalized error $\sqrt{nI(Y)}(\eta(X) - Y)$ is asymptotically standard normal (see, e.g., [10, Sec. 10.2]).

In this section, we establish a finite n, finite k version of the results of Samarov–Has'minskii using a recent nonasymptotic generalization of the Bernstein–von Mises theorem due to Spokoiny [11]. Our result requires a number of regularity assumptions. The first two are standard:

(C.1) The conditional distributions $P_{X_1|Y=y}$, $y \in \mathcal{Y}$, are dominated by a common σ-finite measure μ on the real line. The corresponding log-density

$$\ell(u, y) \triangleq \log \frac{\mathrm{d}P_{X_1|Y=y}}{\mathrm{d}\mu}(u)$$

is twice differentiable in y for every u. We will denote the first and second derivatives with respect to y by ∂ and ∂^2.

(C.2) For each $y \in \mathcal{Y}$, the Fisher information [12]

$$I(y) \triangleq -\partial^2 \mathbb{E}\left[\ell(X_1, Y)|Y = y\right]$$

exists and is positive.

The remaining assumptions (see [13, Sec. 5.1]) are slightly stronger than the classical ones. For each $r > 0$, define the local neighborhoods

$$\mathcal{N}_r(y) \triangleq \left\{ y' \in \mathcal{Y} : \sqrt{I(y)}|y' - y| \leq r \right\}.$$

The regularity assumptions can be split into two groups: identifiability conditions and exponential moment conditions. We start with the former:

(I.1) There are positive constants $r_0 > 0$ and $\delta^* > 0$, such that, for every $r \leq r_0$,

$$\sup_{y \in \mathcal{Y}} \sup_{y' \in \mathcal{N}_r(y)} \left| \frac{2D(P_{X_1|Y=y}\|P_{X_1|Y=y'})}{I(y)(y - y')^2} - 1 \right| \leq \delta^* r.$$

(I.2) There is a constant $b > 0$, such that, for every $r > 0$,

$$\inf_{y \in \mathcal{Y}} \inf_{y' \in \partial \mathcal{N}_r(y)} \frac{D(P_{X_1|Y=y}\|P_{X_1|Y=y'x})}{I(y)(y - y')^2} \geq b.$$

The exponential moment assumptions pertain to the random variables

$$\zeta_y(X_1, y') \triangleq \ell(X_1, y') - \mathbb{E}[\ell(X_1, y')|Y = y], \quad y, y' \in \mathcal{Y}$$

[note that $\zeta_y(Z_1, y) \equiv \ell(X_1, y)$]. Let

$$\xi_y(X_1, y') \triangleq \frac{\partial}{\partial w} \zeta_y(X_1, w) \Big|_{w=y'}.$$

(E.1) There exist constants $g_1 > 0$ and $v_0 > 0$, such that

$$\sup_{y \in \mathcal{Y}} \log \mathbb{E}\left[\exp\left(\frac{\lambda \xi_y(X_1, y)}{\sqrt{I(y)}} \right) \Big| Y = y \right] \leq \frac{v_0^2 \lambda^2}{2}$$

for all $|\lambda| \leq g_1$.

(E.2) There exists a constant $\omega^* > 0$, such that, for every $r \leq r_0$,

$$\sup_{y \in \mathcal{Y}} \sup_{y' \in \mathcal{N}_r(y)} \log \mathbb{E}\left[\exp\left(\frac{\lambda[\xi_y(X_1, y') - \xi_y(X_1, y)]}{\omega^* r \sqrt{I(y)}} \right) \Big| Y = y \right]$$
$$\leq \frac{v_0^2 \lambda^2}{2}$$

for all $|\lambda| \leq g_1$.

(E.3) For every $r > 0$ there exists $g_1(r) > 0$, such that

$$\sup_{y \in \mathcal{Y}} \sup_{y' \in \mathcal{N}_r(y)} \log \mathbb{E}\left[\exp\left(\frac{\lambda \xi_y(X_1, y)}{\sqrt{I(y)}} \right) \Big| Y = y \right] \leq \frac{v_0^2 \lambda^2}{2}.$$

for all $|\lambda| \leq g_1(r)$.

For example, in the case of additive Gaussian noise, i.e., when $P_{X_1|Y=y} = N(y, \sigma^2)$ for all $y \in \mathcal{Y}$, it is easy to verify that all of these assumptions are met.

We are now ready to state and prove the main result of this section. Since both Y and $\eta(X)$ are supported on the bounded interval $[-A, A]$, there is no loss of generality in restricting our attention only to nearest-neighbor quantizers $q_\mathcal{C}$ with codebooks $\mathcal{C} = \{y_1, \ldots, y_k\} \subset [-A, A]$. Moreover, we can assume that $\mathcal{C}$ is ordered in such a way that $-A \equiv$

$y_0 \leq y_1 < y_2 < \ldots < y_k \leq y_{k+1} \equiv A$. Given such an ordered $\mathcal{C}$, we define

$$\Delta_\mathcal{C} \triangleq \max_{0 \leq j \leq k} (y_{j+1} - y_j).$$

Theorem 1. *Suppose that Assumptions (C.1)–(C.2), (I.1)–(I.2), and (E.1)–(E.3) hold. Suppose also that $\log f_Y$ is Lipschitz on $[-A, A]$. Then there exists a constant $L > 0$ that depends only on the constants in the above assumptions, such that, for any k-point nearest-neighbor quantizer $q_\mathcal{C}$ with $\mathcal{C} \subset \mathcal{Y}$, we have*

$$\left| \mathbb{E} |\eta(X) - q_\mathcal{C}(\eta(X))|^2 - \mathbb{E} |Y - q_\mathcal{C}(Y)|^2 \right|$$
$$\leq L\Delta_\mathcal{C}^2 \min \left\{ 1, \frac{1}{\Delta_\mathcal{C} \sqrt{n}} \left(\mathbb{E}\left[\frac{1}{\sqrt{I(Y)}} \right] + \sqrt{\mathsf{mmse}(P_{XY})} \right) \right\}.$$
(3)

In the additive Gaussian noise case, the bound (3) becomes

$$\left| \mathbb{E} |\eta(X) - q_\mathcal{C}(\eta(X))|^2 - \mathbb{E} |Y - q_\mathcal{C}(Y)|^2 \right|$$
$$\leq L\Delta_\mathcal{C}^2 \min \left\{ 1, \frac{\sigma}{\Delta_\mathcal{C} \sqrt{n}} \right\},$$

where σ^2 is the noise variance.

Proof: The idea of the proof of our nonasymptotic result is actually rather simple, unlike that of Samarov and Has'minskii [6], which requires a number of delicate asymptotic approximations and several fairly tedious integrations. For any collection $\mathcal{C} = \{y_1, \ldots, y_k\}$ of k reconstruction points, define the function $e_\mathcal{C} : \mathcal{Y} \to \mathbb{R}^+$ by

$$e_\mathcal{C}(y) \triangleq \min_{j \in [k]} (y - y_j)^2.$$

Then a simple calculation shows that

$$|e_\mathcal{C}(y) - e_\mathcal{C}(y')| \leq \min \left\{ 2\Delta_\mathcal{C}^2, 2\Delta_\mathcal{C}|y - y'| \right\}, \qquad (4)$$

for all $y, y' \in \mathcal{Y}$. The expected reconstruction error of the nearest-neighbor quantizer $q_\mathcal{C}$ can be written as

$$\mathbb{E} |\eta(X) - q_\mathcal{C}(\eta(X))|^2 = \mathbb{E}[e_\mathcal{C}(\eta(X))].$$

Using the law of iterated expectation and the smoothness estimate (4), we obtain

$$|\mathbb{E}[e_\mathcal{C}(\eta(X))] - \mathbb{E}[e_\mathcal{C}(Y)]|$$
$$\leq \mathbb{E} |\mathbb{E}[e_\mathcal{C}(\eta(X)) - e_\mathcal{C}(Y)|Y]|$$
$$\leq 2\Delta_\mathcal{C}^2 \mathbb{E} \min \left\{ 1, \frac{1}{\Delta_\mathcal{C}} \mathbb{E}[|\eta(X) - Y||Y] \right\}. \qquad (5)$$

We now invoke Spokoiny's nonasymptotic Bernstein–von Mises theorem [11]. Let $Z_n \triangleq \eta(X) - (Y + G_n(X, Y))$, where

$$G_n(X, Y) \triangleq \frac{1}{nI(Y)} \sum_{i=1}^{n} \partial \ell(X_i, Y),$$

and for any $L > 0$ consider the event

$$\mathcal{A}_n^L(Y) \triangleq \left\{ \sqrt{nI(Y)}|Z_n| \leq L \left(\frac{\log n}{n} \right)^{1/4} \right\},$$

Then there exists a choice $L = L_0$ that depends only on the constants in the regularity conditions, such that

$$\mathbb{P}\big[\mathcal{A}_n(Y)\big|Y\big] \geq 1 - \frac{C}{n}, \qquad (6)$$

where $\mathcal{A}_n(Y) \equiv \mathcal{A}_n^{L_0}(Y)$, and $C > 0$ is an absolute constant [11, Sec. 2.4.3]. Therefore,

$$\mathbb{E}\big[|\eta(X) - Y|\big|Y\big]$$
$$\overset{(a)}{\leq} \mathbb{E}\big[|Z_n|\big|Y\big] + \mathbb{E}\big[|G_n(X,Y)|\big|Y\big]$$
$$= \mathbb{E}\big[\mathbb{1}\{\mathcal{A}_n(Y)\}|Z_n|\big|Y\big] + \mathbb{E}\big[\mathbb{1}\{\mathcal{A}_n^c(Y)\}|Z_n|\big|Y\big]$$
$$\quad + \mathbb{E}\big[|G_n(X,Y)|\big|Y\big]$$
$$\overset{(b)}{\leq} \frac{L_0}{\sqrt{nI(Y)}} + \sqrt{\frac{C}{n}\mathbb{E}[|Z_n|^2|Y]} + \mathbb{E}\big[|G_n(X,Y)|\big|Y\big]$$
$$\overset{(c)}{\leq} \frac{L_0}{\sqrt{nI(Y)}} + \sqrt{\frac{2C}{n}\mathbb{E}[|Y - \eta(X)|^2|Y]}$$
$$\quad + \sqrt{\frac{2C}{n}\mathbb{E}\big[|G_n(X,Y)|^2\big|Y\big]} + \mathbb{E}\big[|G_n(X,Y)|\big|Y\big], \qquad (7)$$

where (a) follows from the triangle inequality, (b) from (6) and Cauchy–Schwarz, and (c) again from the triangle inequality. Now, since $\mathbb{E}[\partial\ell(X_1,Y)|Y] = 0$ and $\mathrm{Var}[\partial\ell(X_1,Y)|Y] = I(Y)$ [12], we have

$$\mathbb{E}\big[|G_n(X,Y)|^2\big|Y\big] = \mathrm{Var}[G_n(X,Y)|Y] = \frac{1}{nI(Y)}$$

and

$$\mathbb{E}\big[|G_n(X,Y)|\big|Y\big] \leq \frac{1}{\sqrt{nI(Y)}}.$$

Substituting these estimates into (7) gives

$$\mathbb{E}\big[|\eta(X) - Y|\big|Y\big] \leq \frac{L_0 + 1}{\sqrt{nI(Y)}} + \frac{\sqrt{2C}}{n\sqrt{I(Y)}}$$
$$\quad + \sqrt{\frac{2C\,\mathbb{E}[|Y - \eta(X)|^2|Y]}{n}}.$$

Plugging this bound into (5), using Jensen's inequality, and simplifying, we get (3). ∎

Corollary 1. *Under the same assumptions as in Theorem 1,*

$$\left|\mathrm{reg}_k(P_{XY}) - \inf_{\mathcal{C}\subset\mathcal{Y};\,|\mathcal{C}|\leq k}\mathbb{E}|Y - \mathsf{q}_\mathcal{C}(Y)|^2\right|$$
$$\preceq \min\left\{\frac{1}{k^2}, \frac{1}{k\sqrt{n}}\left(\mathbb{E}\left[\frac{1}{\sqrt{I(Y)}}\right] + \sqrt{\mathsf{mmse}(P_{XY})}\right)\right\}. \qquad (8)$$

Remark 2. Using the information inequality [9]

$$\mathsf{mmse}(P_{XY}) \geq \frac{1}{n}\mathbb{E}\left[\frac{1}{I(Y)}\right]$$

and Jensen's inequality, the bound (8) can be weakened to

$$\left|\mathrm{reg}_k(P_{XY}) - \inf_{\mathcal{C}\subset\mathcal{Y};\,|\mathcal{C}|\leq k}\mathbb{E}|Y - \mathsf{q}_\mathcal{C}(Y)|^2\right|$$
$$\preceq \min\left\{\frac{1}{k^2}, \frac{\sqrt{\mathsf{mmse}(P_{XY})}}{k}\right\}.$$

Proof: Let $\mathcal{C}^* = \{y_1^*, \ldots, y_k^*\}$ be the reconstruction points of an optimal k-point quantizer for Y arranged in increasing order. From the work of Panter and Dite [7], we know that these points should be chosen in such a way that

$$\int_{y_j^*}^{y_{j+1}^*} f_Y(y)^{1/3}\mathrm{d}y = \frac{c_j}{k}\int_\mathcal{Y} f_Y(y)^{1/3}\mathrm{d}y,$$

for all $j = 0, 1, \ldots, k-1$, where $c_j = 1/2$ for $j = 0$ and 1 otherwise. Therefore, $\Delta_{\mathcal{C}^*} \preceq 1/k$. Using this and the definition of the MMSE regret in (3), we get (8). ∎

Observe that the value of the right-hand side of (8) is determined by whether the number of observations n is larger or smaller than k^2. This agrees with the asymptotic results of Samarov and Has'minskii [6].

IV. A HIGH-RESOLUTION BOUND FOR FUNCTIONAL COMPRESSION

We now consider a general setting of p-dimensional Y and n-dimensional X without any conditional independence assumptions. Instead, we focus on the scaling of the MMSE regret $\mathrm{reg}_k(P_{XY})$ with k, while the values of n and p stay fixed. Proposition 1 shows that the MMSE regret is precisely the minimum expected distortion attainable by k-ary quantization of X in the problem of *functional compression* of the regression function $\eta(X)$. The results we present in this section require only some regularity assumptions on the ℓ_2 norm $\|\eta(X)\|$:

Assumption 1. *The random variable $\|\eta(X)\|$ has a finite fourth moment:* $\mathbb{E}\|\eta(X)\|^4 < \infty$.

Assumption 2. *The random variable $\|\eta(X)\|$ is subgaussian: there exists a positive constant $v > 0$, such that*

$$\log\mathbb{E}\left[e^{\lambda(\|\eta(X)\| - \mathbb{E}\|\eta(X)\|)}\right] \leq \frac{v\lambda^2}{2}, \qquad \forall\lambda\in\mathbb{R}.$$

One sufficient condition for Assumption 1 to hold is for $\|Y\|$ to have a finite fourth moment. Indeed, in that case, using Jensen's inequality, we have $\mathbb{E}\|\eta(X)\|^4 = \mathbb{E}\|\mathbb{E}[Y|X]\|^4 \leq \mathbb{E}\|Y\|^4 < \infty$. As for Assumption 2, it will be met, for example, if the regression function η is Lipschitz, i.e., if there exists some finite constant $L > 0$, such that

$$\|\eta(x) - \eta(x')\| \leq L\|x - x'\|, \qquad \forall x, x'\in\mathbb{R}^n$$

and if X is a Gaussian random vector with a nonsingular covariance matrix [14].

Theorem 2. *Suppose Assumption 1 holds. Then*

$$\mathrm{reg}_k(P_{XY}) \preceq \big(\mathbb{E}\|\eta(X)\|^2\mathbb{E}\|\eta(X)\|^4\big)^{2/3} k^{-2/3p}. \qquad (9)$$

If Assumption 2 also holds, then

$$\mathrm{reg}_k(P_{XY})$$
$$\preceq \inf_{r > \mathbb{E}\|\eta(X)\|}\left\{r^2 k^{-2/p} + \sqrt{\mathbb{E}\|\eta(X)\|^4}\,e^{-(r - \mathbb{E}\|\eta(X)\|)^2/4v}\right\}. \qquad (10)$$

Remark 3. Here, p can be replaced by a suitable intrinsic dimension of the support of $\eta(X)$ (e.g., the rate-distortion dimension [15]). For example, if $\eta(X)$ is linear, i.e., $\mathbb{E}[Y|X] = AX$ for some deterministic matrix $A \in \mathbb{R}^{p \times n}$, then we can replace p by $\mathrm{rank}(A)$. Moreover, using a suboptimal value of r, we can weaken the bound in (10) to

$$\mathrm{reg}_k(P_{XY}) \preceq \frac{\log k}{k^{2/p}},$$

where the hidden constant depends on p, on the first and fourth moments of $\|\eta(X)\|$, and on the subgaussian constant v. Apart from the logarithmic factor, this scaling of the MMSE regret agrees with the high-resolution approximation for VQ [5] and with the Shannon lower bound [15], [16].

Proof: From Assumption 1 and from Jensen's inequality, it follows that the first and second moments of $\|\eta(X)\|$ are also finite.

Fix a positive real constant $r > 0$, which will be optimized later. A simple volumetric estimate shows that the ℓ^2 ball of radius r in $\mathbb{R}^p$ can be covered by at most $\left(1 + \frac{2r}{\epsilon}\right)^p$ balls of radius ϵ. So, for a given $k \in \mathbb{Z}^+$, we can cover the radius-r ball by k balls of radius $\epsilon \asymp rk^{-1/p}$. Let $\{y_i\}_{i=1}^k$ be the centers of these k balls. We now construct a quantizer $q^{(r)} \in \mathcal{Q}_{k+1}^p$ as follows:

$$q^{(r)}(x) = \begin{cases} \arg\min_{j \in [k]} \|\eta(x) - y_j\|^2, & \text{if } \|\eta(x)\| \leq r \\ k+1, & \text{otherwise.} \end{cases}$$

For this quantizer, we have

$$\mathbb{E}\|\eta(X) - \mathbb{E}[\eta(X)|q^{(r)}(X)]\|^2 = T_1 + T_2,$$

where

$$T_1 \triangleq \mathbb{E}\left[\mathbb{1}\{\|\eta(X)\| \leq r\}\|\eta(X) - \mathbb{E}[\eta(X)|q^{(r)}(X)]\|^2\right]$$
$$T_2 \triangleq \mathbb{E}\left[\mathbb{1}\{\|\eta(X)\| > r\}\|\eta(X) - \mathbb{E}[\eta(X)|q^{(r)}(X)]\|^2\right].$$

For the first term, we have

$$T_1 \leq 4\epsilon^2 \mathbb{P}\left[\|\eta(X)\| \leq r\right] \leq 4\epsilon^2 \asymp r^2 k^{-2/p},$$

where the first inequality is true since $\eta(X)$ and $\mathbb{E}[\eta(X)|q^{(r)}(X)]$ are inside the same ϵ-ball in the covering for all X such that $q^{(r)}(X) \in [k]$. For the second term,

$$T_2 \leq \sqrt{\mathbb{P}\left[\|\eta(X)\| > r\right] \mathbb{E}\|\eta(X) - \mathbb{E}[\eta(X)|q^{(r)}(X)]\|^4}$$
$$\leq \sqrt{\mathbb{P}[\|\eta(X)\| > r]}\sqrt{\mathbb{E}(\|\eta(X)\| + \|\mathbb{E}[\eta(X)|q^{(r)}(X)]\|)^4}$$
$$\leq \sqrt{\mathbb{P}[\|\eta(X)\| > r]}\sqrt{8\,\mathbb{E}\|\eta(X)\|^4 + 8\,\mathbb{E}\|\mathbb{E}[\eta(X)|q^{(r)}]\|^4}$$
$$\leq \sqrt{\mathbb{P}[\|\eta(X)\| > r]}\sqrt{16\,\mathbb{E}\|\eta(X)\|^4},$$

where the first line is by Cauchy–Schwarz inequality, and the remaining steps follow from monotonicity and convexity.

By Markov's inequality,

$$\mathbb{P}[\|\eta(X)\| > r] = \mathbb{P}[\|\eta(X)\|^2 > r^2] \leq \frac{\mathbb{E}\|\eta(X)\|^2}{r^2}.$$

Therefore,

$$\mathrm{reg}_{k+1}(P_{XY}) \leq \mathbb{E}\|\eta(X) - \mathbb{E}[\eta(X)|q^{(r)}]\|^2$$
$$\asymp r^2 k^{-2/p} + \frac{\sqrt{\mathbb{E}\|\eta(X)\|^2 \mathbb{E}\|\eta(X)\|^4}}{r}.$$

Optimizing over all $r > 0$, we get (9). Now suppose Assumption 2 also holds. Let $r = \mathbb{E}\|\eta(X)\| + t$ for some $t > 0$. Since $\|\eta(X)\|$ is subgaussian, the Chernoff bounding technique gives

$$\mathbb{P}\left[\|\eta(X)\| > r\right] = \mathbb{P}\left[\|\eta(X)\| - \mathbb{E}\|\eta(X)\| > t\right]$$
$$\leq e^{-t^2/2v}$$
$$= e^{-(r - \mathbb{E}\|\eta(X)\|)^2/2v}$$

(see, e.g., [14, Chap. 3]). Thus, we have

$$\mathrm{reg}_{k+1}(P_{XY}) \leq \mathbb{E}\|\eta(X) - \mathbb{E}[\eta(X)|q^{(r)}]\|^2$$
$$\asymp r^2 k^{-2/p} + \sqrt{\mathbb{E}\|\eta(X)\|^4}\, e^{-(r - \mathbb{E}\|\eta(X)\|)^2/4v}.$$

for all $r > \mathbb{E}\|\eta(X)\|$. Optimizing over r, we get (10). ∎

REFERENCES

[1] R. E. Curry, *Estimation and Control with Quantized Measurements*, Cambridge, MA: M.I.T. Press, 1970.

[2] J. K. Wolf and J. Ziv, "Transmission of noisy information to a noisy receiver with minimum distortion," *IEEE Trans. Inform. Theory*, vol. 16, no. 4, pp. 406-411, July 1970.

[3] Y. Ephraim and R. M. Gray, "A unified approach for encoding clean and noisy sources by means of waveform and autoregressive model vector quantization," *IEEE Trans. Inform. Theory*, vol. 34, no. 4, pp. 826-834, July 1988.

[4] V. Misra, V. K. Goyal, and L. R. Varshney, "Distributed scalar quantization for computing: High-resolution analysis and extensions," *IEEE Trans. Inform. Theory*, vol. 57, no. 8, pp. 5298-5325, August 2011.

[5] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, October 1998.

[6] A. M. Samarov and R. Z. Has'minskii, "Estimation by means of statistics that take a finite number of values," *Problems of Inform. Transmission*, vol. 13, no. 4, pp. 260-265, 1977.

[7] P. F. Panter and W. Dite, "Quantisation distortion in pulse-count modulation with nonuniform spacing of levels," *Proc. IRE*, vol. 39, pp. 44-48, January 1951.

[8] P. Elias, "Bounds on performance of optimum quantizers," *IEEE Trans. Inform. Theory*, vol. 16, no. 2, pp. 172-184, March 1970.

[9] I. A. Ibragimov and R. Z. Hasminskii, "On information inequalities and superefficient estimates," *Problems of Inform. Transmission*, vol. 9, no. 3, pp. 216-227, 1973.

[10] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, 1998.

[11] V. Spokoiny, "Bernstein–von Mises theorem for growing parameter dimension," arXiv preprint 1302.3430, 2013.

[12] I. A. Ibragimov and R. Z. Has'minskii, *Statistical Estimation: Asymptotic Theory*, Springer, New York, 1981.

[13] V. Spokoiny, "Parametric estimation. Finite sample theory," *Ann. Statist.*, vol. 40, no. 6, pp. 2877-2909, 2012.

[14] M. Raginsky and I. Sason, *Concentration of Measure Inequalities in Information Theory, Communications, and Coding*, 2nd ed., Now Publishers, 2014.

[15] T. Kawabata, and A. Dembo, "The rate-distortion dimension of sets and measures," *IEEE Trans. Inform. Theory*, vol. 40, no. 5, pp. 1564-1572, September 1994.

[16] T. Linder and R. Zamir, "On the asymptotic tightness of the Shannon lower bound," *IEEE Trans. Inform. Theory*, vol. 40, no. 6, pp. 2026-2031, November 1994.