Gaussian Approximation of Quantization Error for Estimation from Compressed Data

Alon Kipnis* and Galen Reeves†
* Department of Statistics, Stanford University
† Department of Electrical and Computer Engineering, Department of Statistical Science, Duke University

***Abstract*—We consider the statistical connection between the quantized representation of a high dimensional signal X using a random spherical code and the observation of X under an additive white Gaussian noise (AWGN). We show that given X, the conditional Wasserstein distance between its bitrate-R quantized version and its observation under AWGN of signal-to-noise ratio $2^{2R}-1$ is sub-linear in the problem dimension. We then utilize this fact to connect the mean squared error (MSE) attained by an estimator based on an AWGN-corrupted version of X to the MSE attained by the same estimator when fed with its bitrate-R quantized version.**

I. Introduction

Due to the disproportionate size of modern datasets compared to available computing and communication resources, many inference techniques are applied to a compressed representation of the data rather than the data itself. In the attempt to develop and analyze inference techniques based on a degraded version of the data, it is tempting to model inaccuracies resulting from lossy compression as an additive noise. Indeed, there exists a rich literature devoted to the characterization of this "noise", i.e., the difference between the original data and its compressed representation [1]. Nevertheless, due to the difficulty of analyzing non-linear compression operations, this characterization is generally limited to the high bit compression regime [2]–[4].

In this paper, we show that quantizing the data using a random spherical code (RSC) leads to a strong and relatively simple characterization of the distribution of the quantization noise. Specifically, consider the output codeword of a bitrate-R RSC and the output of a Gaussian noise channel with signal-to-noise ratio (SNR) $2^{2R}-1$ illustrated in Fig. 1; our main result says that the conditional Wasserstein distance [5] between these two outputs is independent of the problem dimension, hence the normalized risk in estimating from the compressed data is asymptotically equivalent to risk in estimating under a Gaussian noise. This connection allows us to adapt inference procedures designed for AWGN-corrupted data to be used with the quantized data, as well as to analyze the performances of such procedures.

Spherical codes have multiple theoretical and practical uses in numerous fields [6]. In the context of information theory,

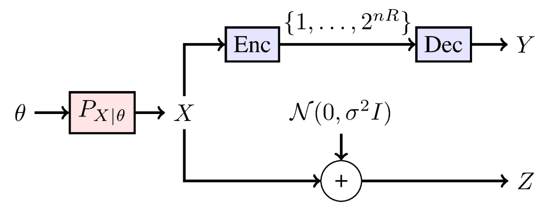

Fig. 1. Inference about the parameter vector θ is based on degraded observations of the data X. The effects of bitrate constraints are compared to the effect of additive gaussian noise by studying the Wassertein distance between $P_{Y|X}$ and $P_{Z|X}$. Under random spherical encoding, we show that this distance is order one, and thus independent of the problem dimensions.

Sakrison [7] and Wyner [8] provided a geometric understanding of random spherical coding in a Gaussian setting, and our main result can be seen as an extension of their insights. Specifically, consider the representation of an n-dimensional standard Gaussian vector X using 2^{2R} codewords uniformly distributed over the sphere of radius $r=\sqrt{n(1-2^{-2R})}$. The left side of Fig. 2, adapted from [7], shows that the angle α between X and its nearest codeword $\hat{X}$ concentrates as n increases so that $\sin(\alpha)=2^{-R}$ with high probability. As a result, the quantized representation of X and the error $X-\hat{X}$ are orthogonal, and thus the MSE between X and its quantized representation averaged over all random codebooks converges to the Gaussian distortion-rate function (DRF) $D_G(R) \triangleq 2^{-2R}$. In fact, as noted in [9], this Gaussian coding scheme[1] achieves the Gaussian DRF when X is generated by any ergodic source of unit variance, implying that the distribution of $\|X-\hat{X}\|^2$ is independent of the distribution of X as the problem dimension n goes to infinity.

In this paper, we show that a much stronger statement holds for a scaled version of the quantized representation (Y in Fig. 2): in the limit of high dimension, the distribution of $Y-X$ is independent of the distribution of X and is approximately Gaussian. This property of $Y-X$ suggests that the underlying signal or parameter vector θ can now be estimated as if X is observed under AWGN. This paper formalizes this intuition by showing that estimators of θ from an AWGN-corrupted version of X (Z in Fig. 1) attain similar performances if applied to the scaled representation Y.

The work of G. Reeves was supported in part by funding from the NSF under Grant No. 1750362. The work of A. Kipnis was supported in part by funsing from the NSF under Grant No. DMS-1418362 and DMS-1407813.

[1]We denote this scheme as *Gaussian* since r is chosen according to the distribution attaining the Gaussian DRF.

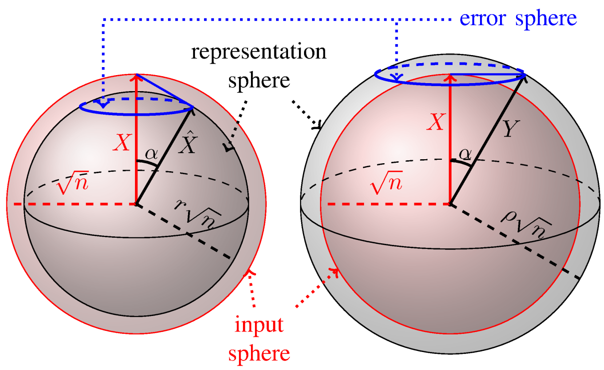

Fig. 2. Left: In the standard source coding problem [7], [8], the representation sphere is chosen such that the error $\hat{X} - X$ is orthogonal to the reconstruction $\hat{X}$. Right: In this paper the representation sphere is chosen such that the error $Y - X$ is orthogonal to the input X.

We emphasize that the scaling of the quantized representation and the estimation of θ are done at the decoder, and thus X is essentially represented by the codeword maximizing the cosine similarity. In particular, no additional information about X or θ is required by the encoder. This situation is in contrast to optimal quantization schemes in indirect source coding [10], [11] and in problems involving estimation from compressed information [12]–[16] in which the encoder depends on $P_{\theta|X}$. As a result, the random spherical coding scheme we rely on is in general sub-optimal, although it can be applied in situations where θ or the model that generated X are unspecified or unknown. Coding schemes with similar properties were studied in the context of indirect source coding under the name *compress-and-estimate* in [17]–[20]. Finally, we note that our approach is similar in spirit to dither based quantization [4] in the sense that it relies on randomness in the coding procedure.

To summarize the contributions of this paper, consider the vectors Y and Z illustrated in Fig. 1, describing the output of a bitrate-R RSC applied to X and the output of an AWGN channel with input X, respectively. Our main result shows that the quadratic Wasserstein distance between Y and Z given the input X is sub-linear in the problem dimension n. We further use this result to establish the equivalence

$$D_{\mathsf{sp}}(R) \approx \mathcal{M}(\mathsf{snr}), \qquad \mathsf{snr} = 2^{2R} - 1, \tag{1}$$

where $D_{\mathsf{sp}}(R)$ is the MSE of estimating θ from Y with input X as a function of the bitrate, and $\mathcal{M}(\mathsf{snr})$ is the MSE of estimating θ under additive Gaussian noise as a function of the SNR in the channel from X to Z. Next, we use (1) to analyze the MSE of Lipschitz continuous estimators of θ from Y by considering their MSE in estimating θ from Z. The benefit is twofold: Allowing inference techniques from AWGN-corrupted measurements to be applied directly on the quantized measurements to recover θ, as well as providing the expected MSE in this recovery. Finally, we apply our results to two special cases where the measurement model $P_{X|\theta}$ is Gaussian. The first case demonstrates how RSC can be used in conjunction with the prior distribution to attain MSE smaller than $D_G(R)$ when this prior is not Gaussian. The second case demonstrates how to derive an achievable distortion for the quantized compressed sensing problem [17], [21] using the approximate message passing (AMP) algorithm and its state evolution property [22], [23].

Due to space limitation, the proofs are omitted and can be found in [24].

II. Preliminaries

Our basic setting considers an n-dimensional random vector X with distribution $P_{X|\theta}$, where θ is a d-dimensional random vector with prior P_θ. We refer to X as the measurements and to θ as the underlying signal, and consider the problem of recovering θ from a bitrate-R quantized version of X.

A. Random Spherical Codes

We quantize X using a random spherical code. Such code consists of a list of M codewords (codebook):

$$\mathcal{C}_{\mathsf{sp}}(M,n) \triangleq \{C(1), \ldots, C(M)\},$$

where each $C(i)$, $i = 1, \ldots, M$, is drawn independently from the uniform distribution over the sphere of radius[2] $\sqrt{n}$ in $\mathbb{R}^n$. The quantized representation of X is given by the codeword $c(i^*)$ that maximizes the cosine similarity between X and the codewords, namely:

$$\mathsf{Enc}(X) = c(i^*), \qquad i^* = \arg \max_{1 \le i \le M_n} \frac{\langle c(i), X\rangle}{\|c(i)\|\|X\|}.$$

Let the cosine similarity between the vector X and its quantized representation be denoted by the random variable

$$S \triangleq \frac{\langle X, \mathsf{Enc}(X)\rangle}{\|X\|\|\mathsf{Enc}(X)\|}.$$

Proposition 1 ([7], [8]). *The cosine similarity S is independent of X and has distribution function*

$$\mathbb{P}[S \le s] = (1 - Q_n(s))^M, \tag{2}$$

where

$$Q_n(s) = \kappa_n \int_s^1 (1 - t^2)^{(n-3)/2} \, \mathrm{d}t, \tag{3}$$

with $\kappa_n = \Gamma(\frac{n}{2})/(\sqrt{\pi}\Gamma(\frac{n-1}{2}))$.

Proof. Proposition 1 follows from [7, Eq. (12)]. □

Proposition 2. *Let $M \ge 4n$ and $r = M^{1/(n-1)}$. Then*

$$\mathbb{E}\left[\left|S - \sqrt{1 - r^{-2}}\right|^2\right] = O\left(\frac{\log^2 n}{r^2(r^2-1)n^2} + \frac{1}{n^2}\right) \tag{4}$$

$$\mathbb{E}\left[\left|\sqrt{1 - S^2} - r^{-1}\right|^2\right] = O\left(\frac{\log^2 n}{r^2 n^2} + \frac{1}{n^2}\right). \tag{5}$$

[2]Our results remain essentially unchanged if the codewords are drawn instead from the standard normal distribution in $\mathbb{R}^n$.

B. Wasserstein Distance and Lipschitz Continuity

The p-Wasserstein distance between $\mathbb{R}^n$-valued random vectors X and Y is defined as

$$W_p(X,Y) = W_p(P_X, P_Y) \triangleq \inf\left(\mathbb{E}\left[\|X-Y\|_p^p\right]\right)^{1/p},$$

where the infimum is over all joint distributions (X,Y) satisfying the marginal constraints $X \sim P_X$ and $Y \sim P_Y$. For $p \geq 1$ the p-Wasserstein distance is a metric on the space of distributions with finite p-th moments.

For the purposes of this paper, it is also useful to introduce the conditional Wasserstein distance as follows. Given conditional distributions $P_{X|U}$ and $P_{Y|U}$ and P_U, we define

$$W_p^p(X,Y \mid U) \triangleq \int W_p^p(P_{X|U=u}, P_{Y|U=u}) \, \mathrm{d}P_U. \quad (6)$$

Equivalently, we have $W_p^p(X,Y \mid U) = \inf \mathbb{E}\left[\|X-Y\|_p^p\right]$, where the infimum is over all couplings of (U,X,Y) satisfying the marginal constraints $(U,X) \sim P_{U,X}$ and $(U,Y) \sim P_{U,Y}$. Notice that

$$W_p(X,Y) \leq W_p(P_{U,X}, P_{U,Y}) \leq W_p(X,Y \mid U). \quad (7)$$

Remark 1. The difference between the conditional and unconditional Wasserstein distances in (7) can be arbitrarily large. For example, suppose that U is uniform on $\{\mu, -\mu\}$ and let $P_{X|U=u} = \mathcal{N}(u, I)$ and $P_{Y|U=u} = \mathcal{N}(-u, I)$. Then $W_2(X,Y) = 0$ but $W_2(X,Y \mid U) = W_2(\mathcal{N}(2\mu, I), \mathcal{N}(0,I))$, which increases without bound as $\|\mu\|$ increases .

Given a Markov chain $\theta \to X \to (Y,Z)$, the conditional Wasserstein distance associated with $P_{Y,Z|X}$ provides a natural upper bound on the difference between the marginals $P_{\theta,Y}$ and $P_{\theta,Z}$. For the following result, recall that a function $f : \mathbb{R}^n \to \mathbb{R}^d$ is L-Lipschitz if $\|f(u) - f(v)\| \leq L\|u-v\|$ for all $u, v \in \mathbb{R}^n$

Proposition 3. *Let $P_{\theta,X}$ be a distribution on $\mathbb{R}^d \times \mathbb{R}^n$ and let $P_{Y|X}$ and $P_{Z|X}$ be conditional distributions from $\mathbb{R}^n$ to $\mathbb{R}^n$. For any $p \geq 1$ and L-Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}^d$, we have*

$$\left|(\mathbb{E}[\|\theta - f(Y)\|^p])^{\frac{1}{p}} - (\mathbb{E}[\|\theta - f(Z)\|^p])^{\frac{1}{p}}\right| \leq L\, W_p(Y,Z \mid X),$$

provided that the expectations exist.

III. Quantization Error versus Gaussian Noise

This section addresses the extent to which the quantization error induced by random spherical coding can be approximated by isotropic Gaussian noise. Given an integer M and positive number ρ, let $P_{Y|X}$ be the conditional distribution of

$$Y = \rho \, \mathsf{Enc}(X), \quad (8)$$

obtained by marginalizing over the randomness in the codebook. By construction, Y is supported on the sphere of radius $\sqrt{n}\rho$.

For comparison, we define the Gaussian noise observation model $P_{Z|X}$ according to

$$Z = X + \sigma W, \quad (9)$$

where $W \sim \mathcal{N}(0, I)$ is standard Gaussian independent of X and σ^2 is the noise power.

A. Finite-Length Bounds

Theorem 4. *Let $P_{Y|Z}$ and $P_{Z|X}$ be the conditional distribution defined by* (8) *and* (9)*, respectively. For any distribution P_X on $\mathbb{R}^n$, there exists a joint distribution $P_{X,Y,Z}$ with marginals P_X, $P_{Y|Z}$, and $P_{Z|X}$, such that*

$$\|Y - Z\|^2 = \left(\sqrt{n}\rho S - \|X\| - \sigma U\right)^2 + \left(\sqrt{n}\rho\sqrt{1-S^2} - \sigma V\right)^2,$$

where (S, U, V, X) are mutually independent with $U \sim \mathcal{N}(0,1)$, $V \sim \chi_{n-1}$, and S is distributed according to (2)*.*

Theorem 4 is general in the sense that it holds for any choice of the parameters (M, ρ, σ). Optimizing over (ρ, σ) as a function of $\mathbb{E}[\|X\|]$ and M leads to the following upper bound on the conditional Wasserstein distance.

Theorem 5. *Let P_X be a distribution on $\mathbb{R}^n$ with finite second moments. Given $R > 0$, let $M = \lceil 2^{nR} \rceil$ and put*

$$\rho = \frac{\mathbb{E}[\|X\|]}{\sqrt{n(1-2^{-2R})}}, \qquad \sigma = \frac{\mathbb{E}[\|X\|]}{\sqrt{n(2^{2R}-1)}}.$$

Then, the conditional Wasserstein distance between the distributions defined in (8) *and* (9) *satisfies*

$$W_2^2(Y, Z \mid X) \leq \mathsf{Var}(\|X\|) + 2\sigma^2 + C_R \frac{\mathbb{E}[\|X\|]^2 \log^2 n}{n^2},$$

where C_R is a constant that depends only on R.

The significance of Theorem 5 is that the second and third terms are inversely proportional to the dimension of the signal. Dividing both sides by the expected norm yields:

$$\frac{W_2(Y, Z \mid X)}{\mathbb{E}[\|X\|]} \leq \frac{\sqrt{\mathsf{Var}(\|X\|)}}{\mathbb{E}[\|X\|]} + O\left(\frac{1}{\sqrt{n}}\right).$$

It is also interesting to note that the signal-to-noise ratio of the Gaussian observation model (9) depends primarily on the code bitrate R, with

$$\frac{\frac{1}{n}\mathbb{E}[\|X\|]^2}{\sigma^2} = 2^{2R} - 1.$$

Combining Proposition 3 and Theorem 5 provides a meaningful comparison whenever the Lipschitz constant times the variance of $\|X\|$ is sufficiently small. For example, one setting of interest is if $L = o(\sqrt{d})$ and $\mathsf{Var}\,\|\theta\| = O(1)$, since in this case the difference between the normalized MSE in estimating θ using $f(Y)$ and $f(Z)$ converges to zero.

B. Asymptotic Equivalence

In the context of the problem of estimating θ from a quantized version of X, our results can be stated as follows. We consider a sequence of distributions on (θ, X) where both d and n scale to infinity with

$$\frac{1}{n}\mathbb{E}\left[\|X\|^2\right] = \gamma + o(1), \qquad \mathsf{Var}(\|X\|) = O(1). \quad (10)$$

We further suppose that there exists a nonincreasing function $\mathcal{M} : \mathbb{R}_+ \to \mathbb{R}_+$ with $\mathcal{M}(0) < \infty$ and a sequence of estimators $\hat{\theta} = \hat{\theta}_{d,n,\sigma}$ such that for almost all $\sigma \in \mathbb{R}_+$, we have

$$\mathrm{Lip}(\hat{\theta}) = o(\sqrt{d_n}) \tag{11}$$

$$\frac{1}{d_n}\mathbb{E}\left[\left\|\theta - \hat{\theta}(Z)\right\|^2\right] = \mathcal{M}(\mathsf{snr}) + o(1), \tag{12}$$

where $\mathsf{snr} = \gamma/\sigma^2$, and we assume that the dimension d_n of θ is an explicit function of n.

The following result follows from Proposition 3 and Theorem 5.

Theorem 6. *Let $\mathcal{M}^-(\mathsf{snr})$ and $\mathcal{M}^+(\mathsf{snr})$ be the left and right limits, respectively, of $\mathcal{M}(s)$ at the point $s = \mathsf{snr}$. For almost all $R \in \mathbb{R}_+$, the MSE distortion of random spherical encoding with decoding function $\hat{\theta}$ satisfies*

$$\liminf_{n\to\infty} \frac{1}{d_n}\mathbb{E}\left[\left\|\theta - \hat{\theta}(Y)\right\|^2\right] \geq \mathcal{M}^+(2^{2R} - 1) \tag{13}$$

$$\limsup_{n\to\infty} \frac{1}{d_n}\mathbb{E}\left[\left\|\theta - \hat{\theta}(Y)\right\|^2\right] \leq \mathcal{M}^-(2^{2R} - 1). \tag{14}$$

In particular

$$\lim_{n\to\infty} \frac{1}{d_n}\mathbb{E}\left[\left\|\theta - \hat{\theta}(Y)\right\|^2\right] = \mathcal{M}(2^{2R} - 1)$$

if 2^{2R-1} is a continuity point of $\mathcal{M}(s)$.

Remark 2. The assumption that (12) holds almost everywhere allow for the possibility that $\mathcal{M}(\mathsf{snr})$ has a countable number of jump discontinuities. This can arise, for example, in the context of random linear estimation [25].

IV. Applications and Examples

This section provides some specific examples of problem settings satisfying the assumptions of Theorem 6. Given jointly random vectors (θ, U) on $\mathbb{R}^d \times \mathbb{R}^n$, denote by

$$\mathsf{mmse}(\theta \mid U) \triangleq \frac{1}{d}\mathbb{E}\left[\|\theta - \mathbb{E}[\theta \mid U]\|^2\right]$$

the quadratic Bayes optimal risk in estimating θ from U. In the following we give some examples where the Bayes optimal estimator of θ from Z satisfies the Lipschitz condition (11), and thus $\mathcal{M}(\mathsf{snr})$ corresponds to $\mathsf{mmse}(\theta \mid Z)$.

A. Noisy Source Coding

Consider the additive Gaussian noise model

$$X = \theta + \epsilon W$$

where $\theta \in \mathbb{R}^n$ and $W \sim \mathcal{N}(0, I)$. The problem of quantizing X at bitrate R and estimating θ from this quantized version is an instance of the *noisy* (a.k.a. *indirect* or *remote*) source coding problem [10], [11]. An achievable distortion for this problem using RSC is given by the following theorem.

Theorem 7. *Assume that the entries of θ are i.i.d. from a distribution with second moment γ and finite fourth moment. There exists a bitrate-R spherical code achieving MSE distortion*

$$D_{\mathsf{sp}}(R, \epsilon) \triangleq \mathsf{mmse}(\theta_1 \mid \theta_1 + \eta W),$$

where $W \sim \mathcal{N}(0, 1)$ and

$$\eta = \eta(R, \epsilon^2) = \sqrt{\frac{\epsilon^2 + \gamma 2^{-2R}}{1 - 2^{-2R}}}.$$

B. Standard Source Coding

An interesting special case of the setting of Theorem 7 arises in the limit $\gamma/\epsilon^2 \to \infty$, in which the noisy source coding problem reduces to a standard one. The minimal achievable distortion for this problem is known to be Shannon's DRF of θ, denoted here by $D_{\mathsf{Sh}}(R)$. Theorem 7 implies that there exists a spherical code attaining

$$D_{\mathsf{sp}}(R) \triangleq D_{\mathsf{sp}}(R, 0) = \mathsf{mmse}\left(\theta_1 \mid \theta_1 + \frac{\sqrt{\gamma}}{\sqrt{2^{2R} - 1}} W_1\right),$$

hence $D_{\mathsf{sp}}(R)$ is an achievable distortion for this source coding problem. Furthermore, $D_{\mathsf{sp}}(R)$ is achievable using a source code agnostic to the distribution of θ. We also note that if θ is zero mean, then

$$D_{\mathsf{Sh}}(R) \leq D_{\mathsf{sp}}(R) \leq D_G(R) \triangleq 2^{-2R}\gamma, \tag{15}$$

with equality if and only if $P_{\theta_1} = \mathcal{N}(0, \gamma)$. The fact that the Gaussian DRF $D_G(R)$ is achievable for this problem already follows from [7], [9]. Our results shows that one can usually do much better than the Gaussian DRF by applying the Bayes optimal estimator of θ from $\theta + \sqrt{\gamma/(2^{2R} - 1)}W$ to an appropriately scaled version of the encoded representation.

Figure 3 illustrates the distortion attained by RSC compared to Shannon's and the Gaussian DRFs in encoding an i.i.d. signal uniform over $\{-1, 1\}$. This figure reveals that a RSC in this setting does not attain the Shannon limit. Indeed, this is the case whenever θ has any distribution satisfying the moment condition that is not Gaussian. In particular, for P_θ with a finite entropy, $D_{\mathsf{sp}}(R) > 0$ for any R whereas $D_{\mathsf{Sh}}(R)$ equals zero for any R exceeding this entropy. In this sense, RSC can be seen as a compromise between the optimal indirect source coding scheme that requires the full specification of $P_{\theta|X}$ at the encoder [11] and the Gaussian coding scheme that assumes the least favorable prior on X.

C. Inference in High-Dimensional Linear Models

Consider the case where $X = \mathbf{A}\theta + \epsilon W$ with $\mathbf{A} \in \mathbb{R}^{n \times d}$ and $W \sim \mathcal{N}(0, I)$. The corresponding relation between θ and Z is given by

$$Z = \mathbf{A}\theta + \xi W, \tag{16}$$

where $\xi \triangleq \xi(\mathsf{snr}) \triangleq \sqrt{\epsilon^2 + \mathbb{E}[\|X\|^2]/n\,\mathsf{snr}}$. We use the approximate message passing (AMP) algorithm [22] to estimate θ from Z, and evaluate the Bayes risk of this estimator using the state evolution property of AMP. Specifically, let $\theta^T_{\mathsf{AMP}}(z)$ denote the result of T iterations of the AMP algorithm applied to $z \in \mathbb{R}^n$. Under the assumption that θ is iid with a bounded support and the operator norm of $\mathbf{A}$ is uniformly bounded in n, we show in the Appendix that θ^T_{AMP} has a Lipschitz constant

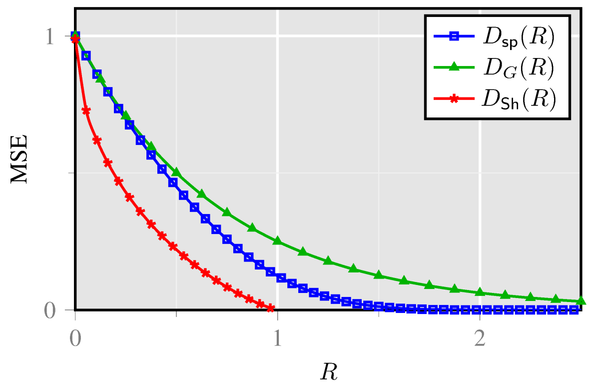

Fig. 3. MSE in encoding an equiprobable binary signal versus coding bitrate R: $D_{\mathsf{sp}}(R)$ is achievable using a RSC, $D_G(R)$ is the Gaussian distortion-rate function achievable using a Gaussian code, and $D_{\mathsf{Sh}}(R)$ is Shannon's distortion-rate function describing the minimal achievable distortion under any bitrate-R code.

independent of n and d. In addition, let $\mathcal{M}^T_{\mathsf{AMP}}(\mathsf{snr}) = \tau_T^2$ be defined by the recursion

$$\tau_t^2 = \begin{cases} \xi^2 + \frac{d}{n}\mathsf{Var}(\theta_1) & t = 0, \\ \xi^2 + \frac{d}{n}\mathbb{E}\left[\left(\mathbb{E}[\theta_1|\theta_1 + \tau_{t-1}W_1] - \theta_1\right)^2\right] & 1 \le t \le T. \end{cases} \tag{17}$$

If we assume that $\mathbf{A}$ has i.i.d. entries $\mathcal{N}(0, 1/n)$ and $n/d \to \delta \in (0, \infty)$, then the main result of [23] says that

$$\frac{1}{d}\mathbb{E}\left[\left\|\theta - \theta^T_{\mathsf{AMP}}(Z)\right\|^2\right] = M^T_{\mathsf{AMP}}(\mathsf{snr}) + o(1).$$

Thm. 6 now implies that for almost all R, δ and ϵ, the limit

$$\lim_{n\to\infty} \frac{1}{d}\mathbb{E}\left[\left\|\theta - \theta^T_{\mathsf{AMP}}(Y)\right\|^2\right]$$

exists and is given by $\mathcal{M}^T_{\mathsf{AMP}}(2^{2R} - 1)$.

V. Conclusions

We considered the problem of estimating an underlying signal from the lossy compressed version of another high dimensional signal. We showed that when the compression is obtained using a RSC of bitrate R, the conditional distribution of the output codeword is close in Wasserstein distance to the conditional distribution of the output of an AWGN channel with SNR $2^{2R} - 1$. This equivalence between the noise associated with lossy compression and an AWGN channel allows us to adapt existing techniques for inference from AWGN-corrupted measurements to estimate the underlying signal from the compressed measurements, as well as to characterize their asymptotic performance.

Acknowledgments

The authors wish to thank Cynthia Rush, David Donoho, and Robert Gray for helpful discussions and comments.

References

[1] R. Gray and D. Neuhoff, "Quantization," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2325–2383, Oct 1998.
[2] W. R. Bennett, "Spectra of quantized signals," *Bell System Technical Journal*, vol. 27, no. 3, pp. 446–472, 1948.
[3] D. H. Lee and D. L. Neuhoff, "Asymptotic distribution of the errors in scalar and vector quantizers," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 446–460, 1996.
[4] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Transactions on Information Theory*, vol. 42, no. 4, pp. 1152–1159, Jul 1996.
[5] S. T. Rachev and L. Rüschendorf, *Mass Transportation Problems: Volume I: Theory*. Springer Science & Business Media, 1998, vol. 1.
[6] P. Delsarte, J.-M. Goethals, and J. J. Seidel, "Spherical codes and designs," in *Geometry and Combinatorics*. Elsevier, 1991, pp. 68–93.
[7] D. Sakrison, "A geometric treatment of the source encoding of a gaussian random variable," *IEEE Transactions on Information Theory*, vol. 14, no. 3, pp. 481–486, 1968.
[8] A. D. Wyner, "Communication of analog data from a Gaussian source over a noisy channel," *The Bell System Technical Journal*, vol. 47, no. 5, pp. 801–812, 1968.
[9] A. Lapidoth, "On the role of mismatch in rate distortion theory," *IEEE Transactions on Information Theory*, vol. 43, no. 1, pp. 38–47, 1997.
[10] R. Dobrushin and B. Tsybakov, "Information transmission with additional noise," *IRE Transactions on Information Theory*, vol. 8, no. 5, pp. 293–304, 1962.
[11] T. Berger, *Rate-distortion theory: A mathematical basis for data compression*. Englewood Cliffs, NJ: Prentice-Hall, 1971.
[12] T. S. Han, "Hypothesis testing with multiterminal data compression," *IEEE Transactions on Information Theory*, vol. 33, no. 6, pp. 759–772, 1987.
[13] Z. Zhang and T. Berger, "Estimation via compressed information," *IEEE Transactions on Information Theory*, vol. 34, no. 2, pp. 198–211, 1988.
[14] J. C. Duchi, M. I. Jordan, M. J. Wainwright, and Y. Zhang, "Optimality guarantees for distributed statistical estimation," *arXiv preprint arXiv:1405.0782*, 2014.
[15] Y. Han, P. Mukherjee, A. Ozgur, and T. Weissman, "Distributed statistical estimation of high-dimensional and nonparametric distributions," in *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2018, pp. 506–510.
[16] A. Kipnis and J. C. Duchi, "Mean estimation from adaptive one-bit measurements," in *Communication, Control, and Computing (Allerton), 2017 55nd Annual Allerton Conference on*. IEEE, 2017, to be presented.
[17] A. Kipnis, G. Reeves, and Y. C. Eldar, "Single letter formulas for quantized compressed sensing with gaussian codebooks," 2018, available online: http://www.stanford.edu/~kipnisal/Papers/QuantizedCS_ISIT2018.pdf.
[18] A. Kipnis, S. Rini, and A. J. Goldsmith, "Compress and estimate in multiterminal source coding," 2017, unpublished. [Online]. Available: https://arxiv.org/abs/1602.02201
[19] A. Kipnis, A. J. Goldsmith, and Y. C. Eldar, "The distortion-rate function of sampled wiener processes," *IEEE Transactions on Information Theory*, vol. 65, no. 1, pp. 482–499, Jan 2019.
[20] G. Murray, A. Kipnis, and A. J. Goldsmith, "Lossy compression of decimated gaussian random walks," in *2018 52nd Annual Conference on Information Sciences and Systems (CISS)*, March 2018, pp. 1–6.
[21] A. Kipnis, G. Reeves, Y. C. Eldar, and A. J. Goldsmith, "Fundamental limits of compressed sensing under optimal quantization," in *Information Theory (ISIT), 2017 IEEE International Symposium on*, 2017.
[22] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914–18 919, 2009.
[23] M. Bayati and A. Montanari, "The dynamics of message passing on dense graphs, with applications to compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 764–785, Feb 2011.
[24] A. Kipnis and G. Reeves, "Gaussian approximation of quantization error for estimation from compressed data," 2019, available online: http://www.stanford.edu/~kipnisal/.
[25] G. Reeves and H. D. Pfister, "The replica-symmetric prediction for compressed sensing with gaussian matrices is exact," in *2016 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2016, pp. 665–669.