ON THE RATE-DISTORTION PERFORMANCE OF COMPRESSED SENSING

Alyson K. Fletcher[1], *Sundeep Rangan*[2], *and Vivek K Goyal*[3]

[1]University of California, Berkeley
[2]QUALCOMM Flarion Technologies
[3]Massachusetts Institute of Technology

ABSTRACT

Encouraging recent results in compressed sensing or compressive sampling suggest that a set of inner products with random measurement vectors forms a good representation of a source vector that is known to be sparse in some fixed basis. With quantization of these inner products, the encoding can be considered universal for sparse signals with known sparsity level. We analyze the operational rate-distortion performance of such source coding both with genie-aided knowledge of the sparsity pattern and maximum likelihood estimation of the sparsity pattern. We show that random measurements induce an additive logarithmic rate penalty, i.e., at high rates the performance with rate $R + O(\log R)$ and random measurements is equal to the performance with rate R and deterministic measurements matched to the source.

Index Terms— compressed sensing, eigenvalue distribution, random matrices, quantization, subspace detection

1. INTRODUCTION AND OVERVIEW

It was recently discovered that when a signal obeys a sparsity or compressibility model, "sensing" can be made blind to the source distribution with remarkably little performance penalty—even while keeping the reconstruction procedure tractable [1, 2]. This is now commonly known as *compressed sensing* or *compressive sampling*.

Consider a signal $x \in \mathbb{R}^L$ with $x = Tu$ and u sparse or compressible. The main ideas in compressed sensing are that

(a) transform coefficients in a *random* basis, $z = \Phi x$, where $\Phi \in \mathbb{R}^{N \times L}$ has an isotropic distribution, are almost as good a representation of x as u; and

(b) the estimate from a convex optimization

$$\hat{x} = T \cdot \operatorname{argmin}_{u\,:\Phi T u = z} \|u\|_1 \tag{1}$$

is almost as good as the optimal estimate of x from z.

This paper focuses on the implications of using random measurements and not on tractable estimation.

More precisely, an orthonormal transform T^T is good for representing a deterministic vector $x \in \mathbb{R}^L$ as $u = T^T x$ when $\|u\|_p = (\sum_i |u_i|^p)^{1/p}$ is small for some fixed $p \in (0, 1)$. The approximation error from keeping only the K largest components of u satisfies

$$\|x - \hat{x}\|_2 \;=\; \|u - \hat{u}\|_2 \;\le\; \zeta_p \cdot \|u\|_p \cdot (K+1)^{1/2-1/p} \tag{2}$$

where ζ_p is a constant depending only on p [2]. So the decay with increasing K gets faster as the smallest p such that $\|u\|_p$ is small gets smaller. When $\|T^T x\|_p$ is small for some $p \in [0, 1]$, x is called "compressible in the basis T."

Compressed sensing exploits compressibility in the basis T even if T is known only in the decoder, not in the encoder. The reconstruction procedure (1) gives an estimate with quality satisfying

$$\|x - \hat{x}\|_2 \;\le\; \xi_p \cdot \|T^T x\|_p \cdot (N/\log(L))^{1/2-1/p} \tag{3}$$

with probability approaching 1 as $L \to \infty$. Comparing (2) and (3), the N "random measurements" Φx give the same reconstruction quality as $N/\log(L)$ best transform coefficients. An intuitive case (not quite covered by the same theory) is for the K best transform coefficients to give an exact representation. Then the decoder (1) recovers the signal exactly from $O(K \log L)$ measurements.

Compressed sensing is robust to noise in the measurements, including quantization noise. This suggests to consider the scalar quantization of $z = \Phi x$ as a universal encoding of sparse or compressible x [1]. This paper investigates universal encoding for $x \in \mathbb{R}^L$ that is known to be K sparse in a fixed basis—without constraints on decoding complexity. We determine the optimal number of measurements N for the case in which the sparsity pattern (positions of nonzero coefficients) is provided to the decoder by a genie; even in this case, the optimal N is strictly larger than K. Then we determine the minimum N that allows a maximum likelihood estimation of the sparsity pattern to succeed with high probability.

2. BASIC NOTATION AND PRELIMINARY CALCULATIONS

2.1. Sparse Signal Model

The signal to be compressed will be denoted by the vector $x \in \mathbb{R}^L$, and is assumed to be of the form,

$$x = Tu, \tag{4}$$

where $T \in \mathbb{R}^{L \times M}$ and $u \in \mathbb{R}^M$ has at most K non-zero coefficients. The vector x is called K*-sparse with respect to* T.

Since the signal x is a combination of at most K of the M frame vectors in T, x must belong to one of $\binom{M}{K}$ subspaces. We will use the following notation: Let $J = \binom{M}{K}$ and enumerate the possible subspaces for x by S_j where $j = 1, \ldots, J$. Also, for each subspace S_j, let $V_j \in \mathbb{R}^{L \times K}$ be a matrix whose columns form an orthonormal basis for S_j. That is, $V_j' V_j = I_K$ and the range space of V_j is equal S_j. With these definitions, any K-sparse signal x of the form (4) can be written as

$$x = V_\theta w, \tag{5}$$

where $\theta \in \{1, \ldots, J\}$, and w is a vector with $w \in \mathbb{R}^K$.

Support provided by a University of California President's Postdoctoral Fellowship, the Texas Instruments Leadership Universities Program, and the Centre Bernoulli, École Polytechnique Fédérale de Lausanne.

For rate-distortion analysis, it is necessary to have a statistical model for the sparse signal x. Instead of directly specifying the probability distribution on u, it will be convenient to describe x in terms of the subspace model (5) as follows:

Assumption 1 *The signal $x \in \mathbb{R}^L$ is of the form (5), where*

(a) the matrices $V_1, \ldots, V_J$ are fixed (i.e. deterministic) with $V_j \in \mathbb{R}^{L \times K}$ and $V_j' V_J = I_K$ for all j;

(b) the subspace index θ is uniformly distributed on the set $\{1, \ldots, J\}$; and

(c) the vector w is Gaussian with mean zero and unit variance I_K. The vector w is independent of θ.

The model simply states that all the subspaces are equally likely and that, given that x is on a subspace S_j, its conditional distribution on the subspace is Gaussian.

2.2. Adaptive Quantization

The key motivation for compressed sensing is that the set of matrices V_j need not be known at the encoder. The matrices need only be known at the decoder. To evaluate the compressed sensing method, it is useful to first evaluate the performance of a simple baseline non-universal compression scheme, where both the encoder and decoder know the set of matrices V_j. We will call this *adaptive* quantization since the encoder can "adapt" to the signal structure.

To analyze adaptive quantization, suppose that we have a total of RK bits to quantize the vector x. Throughout this work, we will let $\hat{x}$ denote the best estimate for x at the decoder given the quantized values from the encoder. We let D denote the corresponding per-component average distortion, $D = \frac{1}{K}\mathbf{E}\|x - \hat{x}\|^2$, where the expectation is with respect to the signal model in Assumption 1.

First consider the case with $J = 1$. That is, the signal x belongs to a single K-dimensional subspace, known to the both the encoder and decoder. In this case, the encoder is simply quantizing a K-dimensional Gaussian random vector with RK bits. The resulting minimal distortion is given by the well-known rate-distortion formula for a Gaussian source,

$$D_{adapt-genie}(R) = 2^{-2R}. \quad (6)$$

We have used the subscript "adapt-genie", since the assumption that $J = 1$ is equivalent to a "genie" telling the encoder and decoder which subspace the true signal belongs to. That is, in the notation of (5), the encoder and decoder know the subspace index θ.

Of course, for sparse signal encoding, the subspace index θ is not known, in general, to the decoder. Since there are J possible subspaces, the encoder can represent the subspace index by $\log_2 J$ bits. We will let

$$R_V = \frac{1}{K}\log_2 J, \quad (7)$$

which represents the rate to represent the subspace normalized by signal dimension. We will call R_V the *subspace rate*. If $M \to \infty$ and K/M is constant, it can be verified that

$$\lim_{M\to\infty} R_V = \lim_{M\to\infty} \frac{1}{K}\log_2 \binom{M}{K} = \frac{M}{K}h(K/M), \quad (8)$$

where $h(p) = -p\log_2 p - (1-p)\log_2(1-p)$ is the binary entropy.

Since it takes $\log_2 J = KR_V$ bits to represent the subspace index, the remaining $KR - KR_V$ bits can be used to quantize the Gaussian signal on the K-dimensional subspace. This results in the distortion

$$D_{adapt}(R) = \begin{cases} 2^{-2(R-R_V)}, & R > R_V; \\ 1, & \text{otherwise.} \end{cases} \quad (9)$$

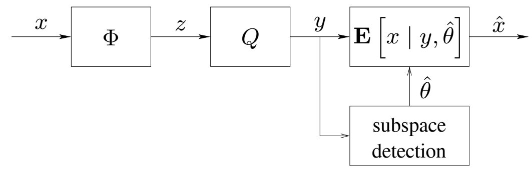

Fig. 1. Compressed sensing encoding with optimal decoding

This is the distortion of the adaptive encoder and will serve as our baseline to compare compressed sensing against.

3. COMPRESSED SENSING WITH QUANTIZATION

3.1. Encoder and Decoder Modeling Assumptions

We consider the compressed sensing encoder and decoder shown in Fig. 1. Our modeling assumptions are summarized as follows:

- *Signal model:* The signal $x \in \mathbb{R}^L$ is modeled as a random vector satisfying Assumption 1.
- *Encoder linear transformation:* The encoder first applies a random linear transformation $\Phi \in \mathbb{R}^{N\times L}$ to yield a vector $z = \Phi x \in \mathbb{R}^N$. We assume that $K \le N \le L$. The ratio $\alpha = K/N$ will be called the *sampling ratio*. We will assume that Φ is *uniformly orthogonal* in that it is a random matrix, uniformly distributed on the set of matrices with $\Phi\Phi' = I_N$.
- *Encoder quantization:* The encoder quantizes the transformed signal z to produce the quantized vector $y = Q(z) \in \mathbb{R}^N$. We will assume that the quantizer has a total of RK bits to quantize the vector $z \in \mathbb{R}^N$. Thus, R is the rate per sparse signal dimension. To simplify the analysis, we will approximate the effect of the quantization with a simple linear additive noise model,

  $$y = \rho z + \sigma_v v, \quad (10)$$

  where $\rho \in [0, 1]$ represents a linear gain, and v is additive noise independent of z. We further assume that v is zero-mean Gaussian with unit variance I_N, and σ_v is a scaling factor representing the quantization error variance. While this model is not exact, it is widely-used in quantization analyses. It can be made exact by assuming vector quantization across instances.
- *Decoder subspace detection:* The decoder first estimates the subspace index θ in Assumption 1 with the estimate

  $$\hat{\theta} = \mathrm{argmax}_{j=1,\ldots,J} \|P_j y\|, \quad (11)$$

  where P_j is the projection onto subspace S_j. That is, the estimator simply selects the subspace with the maximum energy of the received signal. This is the ML estimate for the subspace under the additive Gaussian noise model.
- *Linear signal estimate:* The decoder computes the final estimate with the conditional MMSE estimate $\hat{x} = \mathbf{E}\left[x \mid y, \theta = \hat{\theta}\right]$.

The key difference between this model and standard compressed sensing is in the decoder: In the subspace detection step, the above decoder may exhaustively search over all J possible subspaces. Since $J = \binom{M}{K}$, this search grows exponentially with the problem dimension. Consequently, this decoder is not practical to implement. We consider this exhaustive decoder since our interest is in determining an information theoretic limit of performance, rather than the performance of a particular suboptimal decoder.

The parameters ρ and σ_v in (10) can be based on the quantizer accuracy. One reasonable choice of the parameters is $\rho = 1 - \beta$, $\sigma_v^2 = \beta(1-\beta)$, where β is the relative quantization error

$$\beta = \mathbf{E}\|z - Q(z)\|^2 / \mathbf{E}\|z\|^2. \tag{12}$$

The relative quantizer accuracy β in (12) depends on the number of bits of the quantizer and the specific quantizer design. Since there are RK bits to quantize the N-dimensional vector z, the quantizer has RK/N bits per component. If we assume optimal vector quantization, the relative accuracy is given by the well-known formula

$$\beta = 2^{-2KR/N} = 2^{-2\alpha R}. \tag{13}$$

3.2. Asymptotic Average Distortion

Given the assumptions in the previous section, we define the average per component distortion as $D_K = \frac{1}{K}\mathbf{E}\|x - \hat{x}\|^2$, where the expectation is taken over the signal x, quantization noise v and random encoding matrix Φ. To reflect the dependence on rate, we may sometimes write $D_K(R)$ for D_K. Our analyses will apply mostly to large frames, and we will be mostly interested in the *asymptotic average distortion* defined as $D(R) = \lim_{K\to\infty} D_K(R)$. In this limit we assume that, as $K \to \infty$, the ratios K/N and L/N along with R and R_V remain constant.

4. ANALYSIS FOR $J = 1$

When $J = 1$ the decoder effectively knows the K-dimensional subspace that the true signal x belongs to. Consequently, the situation is equivalent to a "genie" telling the decoder the true subspace index θ. The performance of the genie-aided decoder is given by the following result.

Theorem 1 *Consider a random signal $x \in \mathbb{R}^L$ encoded and decoded as in Section 3.1. Assume that $J = 1$. Then the asymptotic per component distortion is given by,*

$$D_{CS-genie} = \tfrac{1}{2}\left(-A + \sqrt{A^2 + 4B}\right), \tag{14}$$

where $A = (1-\alpha)/\alpha$, $B = \beta/(1-\beta)$, and $\alpha = K/N$.

Theorem 1 provides a simple, explicit formula for the asymptotic distortion in terms of the sampling rate α and quantization accuracy β. Due to space considerations, we will not include the proof of the result. The basic idea is to express the MSE in terms of the eigenvalues of a certain random matrix and apply the Marčenko-Pastur asymptotic eigenvalue distribution.

Now, recall that the sampling ratio $\alpha = K/N$ is a design parameter for the system. It is interesting to find the optimal value of α in this "genie-aided" case. Assume that we employ optimal VQ, so that the quantizer accuracy is given by $\beta = 2^{-2\alpha R}$. Then, $D_{CS-genie}$ in (14) is strictly a function of α. Define the minimum value by

$$D_{CS-genie}^{min} = \min_{\alpha\in[0,1]} D_{CS-genie}, \tag{15}$$

and let α^{min} be the corresponding minimizing value of α. For large R, it can be verified that $\alpha^{min} \approx 1 - 1/(2\log(2)R)$, and

$$D_{CS-genie}^{min} \approx (2\log(2)R - 1)e2^{-2R} \approx 3.76R2^{-2R}, \tag{16}$$

where the second approximation occurs when $2\log(2)R \gg 1$. We make the following observations:

- Comparing the minimum distortion in (16) of the genie-aided decoder to the corresponding performance (6) of the adaptive decoder, we see that $D_{CS-genie}^{min} \approx 3.76RD_{adapt-genie}(R)$. The formula shows that, even if the compressed sensing decoder knows the true subspace, there is a multiplicative increase in the distortion by a factor of approximately $3.76R$ at high rates. The loss is a result of the random angle between the K-dimensional subspace on which the signal lies and the N-dimensional subspace the signal is projected to. We will call this a *rotational loss*.
- We can rewrite (16) as

  $$D_{CS-genie}^{min} \approx 2^{-2(R-\log_2(3.76R))} = D_{adapt-genie}(R_{eff})$$

  where $R_{eff} = R - \log_2(3.76R)$. Thus, compressed sensing results in an additive loss in effective rate, where the additive term $\log_2(3.76R)$ grows logarithmically with the rate.

5. ANALYSIS FOR GENERAL J WITH RANDOM SUBSPACES

5.1. Independent Subspace Signal Model

We now turn to the case of a general value of J. In this case, the performance of the decoder depends on the specific matrices V_j in Assumption 1. Since it is difficult to derive concrete results for a general set of subspaces, we consider a simple model where the subspaces are independent and *uniformly orthogonal*:

Assumption 2 *The signal $x \in \mathbb{R}^L$ satisfies Assumption 1, where the subspace matrices $V_j \in \mathbb{R}^{L\times K}$ are themselves random. Specifically, the matrices V_j are i.i.d. and uniformly orthogonal in that they are uniformly distributed over the set of matrices with $V_j'V_j = I_K$.*

The assumption requires that the distribution of each of the J subspaces S_j in Assumption 1 to be spherically symmetric. This property will occur, for example, if the columns of the matrix T in (4) are themselves i.i.d. and spherically symmetric. However, even in this case, while one can construct uniformly orthogonal basis matrices V_j for the J subspaces, the matrices will not, in general, be independent. Indeed, in the model (4) many of the subspaces will share common random frame vectors and will thus not be independent. Consequently, the assumption in Assumption 2, represent an approximation to the standard frame model.

5.2. Probability of Subspace Misdetection

Having described the signal model, we first evaluate the subspace detection performance. Consider the signal model and encoder and decoder assumptions described in Section 3.1, and assume that the signal x is generated from a set of uniformly orthogonal set of subspaces as described in Assumption 2. Define the *asymptotic probability of subspace misdetection* as $P_{err} = \lim_{N\to\infty} \Pr\left(\hat{\theta} \neq \theta\right)$, where the probability is with respect to the random signal x, the random subspaces V_j, the random encoding matrix Φ and quantization noise v. As in Section 3.2, the limit as $N \to \infty$ is taken with K/N, L/N, R and R_V all constant. Under these assumptions, the asymptotic probability of error is given as follows:

Theorem 2 *Let α^{crit} be the maximum value of $\alpha \in [0,1]$ satisfying*

$$-\tfrac{1}{2}\log_2(\beta) \geq \alpha R_V + \tfrac{1}{2}\alpha\log_2\left(1 + \tfrac{1-\beta}{\alpha\beta}\right). \tag{17}$$

Then, the asymptotic probability of subspace misdetection satisfies

$$P_{err} = \begin{cases} 1, & \alpha > \alpha^{crit}; \\ 0, & \alpha < \alpha^{crit}, \end{cases} \tag{18}$$

where $\alpha = K/N$.

The theorem shows that there is a critical sampling ratio, α^{crit}, below which the ML subspace detection is guaranteed to work. That is, as long as $N > \alpha^{crit}K$, the estimate for the subspace index $\hat{\theta}$ will equal the true subspace index θ with probability one. On the other hand, if $N < \alpha^{crit}K$, then the subspace detection will almost surely fail. We again omit the proof, although the result is proven similar to the style of [3].

When using optimal VQ, $\beta = 2^{-2\alpha R}$, and high rates R, it can be easily verified that $\alpha^{crit} \approx 1 - R_V/R$. In particular, as $R \to \infty$, $\alpha^{crit} \to 1$ and $K/N \to 1$. This can be contrasted with the compressed sensing with ℓ_1 decoding. The best known bounds for compressed sensing with ℓ_1 decoding require $N = O(K \log M)$. Thus, as the problem dimension grows, $K/N \to 0$.

5.3. Asymptotic Distortion

Suppose α is selected so that $\alpha < \alpha^{crit}$ where α^{crit} is defined in Theorem 2. In this case, the theorem states that with probability approaching one the subspace detector will detect the correct subspace. Consequently, the performance of the decoder with the subspace detection will be identical to the "genie" decoder in Section 4 which knew the subspace *a priori*. Thus, if we let $D_{CS}(\alpha)$ be the asymptotic distortion of the compressed sensing decoder, we obtain

$$D_{CS}(\alpha) = D_{CS-genie}(\alpha) \quad \text{when } \alpha < \alpha^{crit}.$$

Minimizing this distortion over all $\alpha \in \alpha^{crit}$, we obtain the minimum distortion

$$D_{CS}^{min} = \inf_{\alpha \in [0, \alpha^{crit}]} D_{CS-genie}(\alpha). \tag{19}$$

We will let α^{min} be the corresponding minimizing value of α.

In the case when $\beta = 2^{-2\alpha R}$ and R is large, we can explicitly evaluate the minimum distortion. It turns out that

$$D_{CS}^{min} \approx \left(\tfrac{R}{a} - 1\right) 2^{-2(R-a)}, \tag{20}$$

where

$$a = \max\{R_V, R_V^{min}\}, \quad R_V^{min} = 1/(2\log 2) \approx 0.72. \tag{21}$$

The minimizing sampling ratio α^{min} is given by $\alpha^{min} = 1 - a/R$.

We make the following comments:

- The minimum distortion shown in (20) and (21) has two regions: (a) when the subspace rate, R_V, is greater than a universal constant R_V^{min}, and (b) when $R_V < R_V^{min}$.
- In the case when $R_V > R_V^{min}$, $a = R_V$ and (20) reduces to

  $$D_{CS}^{min} \approx \left(\frac{R}{R_V} - 1\right) 2^{-2(R-R_V)} \approx \left(\frac{R}{R_V} - 1\right) D_{adapt}(R),$$

  where $D_{adapt}(R)$ in the last step is given in (9). Thus, at high rates R, compressed sensing with ML decoding increases the distortion by a multiplicative factor $R/R_V - 1$ over adaptive encoding. As in Section 4, this is equivalent to an additive penalty in rate of $O(\log R)$. This penalty is significantly less than the multiplicative rate penalty in compressed sensing with ℓ_1 decoding.

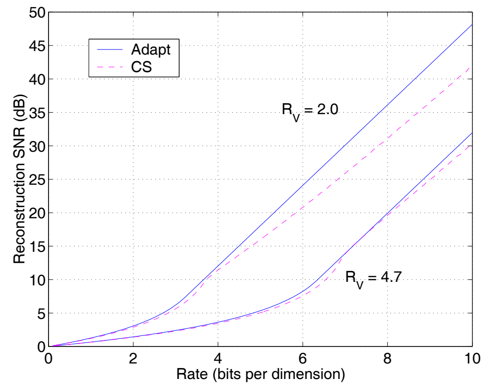

Fig. 2. Theoretical asymptotic performance of compressed sensing ('CS') and adaptive quantization ('Adapt') of a sparse source with two different subspace rates: R_V = 2 and 4.6.

- When $R_V < R_V^{min}$, $a = R_V^{min}$ and it can be verified that

  $$D_{CS}^{min} \approx \left(\frac{R}{2\log(2)} - 1\right) e2^{-2R} \approx D_{CS-genie}^{min}.$$

  Thus, when the number of subspaces is sufficiently small, the compressed sensing decoder with ML detection performs identically to the "genie-aided" decoder in Section 4. The loss in performance in comparison to adaptive coding is due to the *rotational loss* and not subspace misdetection.

To illustrate the above expressions, Fig. 2 plots the asymptotic distortion for the adaptive and compressed sensing schemes as a function of the quantization rate R for two values of the subspace rate R_V = 2 and 4.6. As in (8), the two values correspond to K/M = 0.1 and 0.5, respectively. For the adaptive quantization, the distortion is given by $D(R) = D_{adapt}$ in (9); for compressed sensing $D(R) = D_{CS}^{min}$ in (19). In both cases, the distortion is plotted as the reconstruction SNR given by $10\log_{10}(1/\overline{D}(R))$, where $\overline{D}(R)$ is the convex lower bound to the distortion.

There is an increase in the reconstruction SNR with a lower subspace rate R_V. This occurs since the adaptive encoder requires fewer bits to encode the subspace index, and the CS encoder requires a smaller number of samples N to detect the subspace at the decoder.

The loss in performance is compressed sensing is relatively small. For example, even at a high rate of 10 bits per dimension, with $R_V = 2$, compressed sensing achieves a reconstruction SNR of approximately 42 dB, while adaptive quantization achieves approximately 48 dB. This is equivalent to a gap of a little more than 1 bit. For the higher number of subspaces, $R_V = 4.6$, the gap is even smaller: approximately 0.3 bits.

6. REFERENCES

[1] E. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inform. Theory*, Oct. 2004, Submitted.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] A. K. Fletcher, S. Rangan, V. K Goyal, and K. Ramchandran, "Denoising by sparse approximation: Error bounds based on rate–distortion theory," *EURASIP J. Appl. Sig. Proc.*, vol. 2006, pp. 1–19, Mar. 2006.