On Subspace Structure in Source and Channel Coding

Alyson K. Fletcher
University of California, Berkeley
Berkeley, CA 94720, USA
Email: alyson@eecs.berkeley.edu

Sundeep Rangan
Qualcomm Technologies
Bedminster, NJ 07921, USA
Email: s.rangan@qualcomm.com

Vivek K Goyal
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
Email: vgoyal@mit.edu

***Abstract*—The use of subspace structure in source and channel coding is studied. We show that for source coding of an i.i.d. Gaussian source, restriction of the codebook to a union of subspaces need not induce any performance penalty. In fact, in N-dimensional space, a two-stage quantization of first projecting to the nearest of J subspaces of dimension K in a random first-stage codebook of subspaces, followed by quantizing to the nearest of codewords in a second-stage codebook within the K-dimensional subspace induces no performance loss. This structure allows the rate-distortion bound to be approached asymptotically with block length N. The dual results for channel coding are explicitly described: For an additive white Gaussian noise channel, we introduce a particular subspace-based codebook that induces no rate loss, and the Shannon capacity is achieved. While this has complexity exponential in N, it is reduced from an unstructured search.**

I. Introduction

Most of the theory of source and channel coding is examined in the linear space $\mathbb{R}^N$ equipped with the ordinary Euclidean (ℓ^2) norm; norms being the tool needed for development of packings, coverings, etc. In practice though, source coding and approximation theory often use Hilbert space structure where signals are approximated with linear combinations of basis vectors, i.e. subspaces. This encourages investigation of the possible role of subspace structure in coding theory.

Consider the standard block source coding problem of quantizing a random vector. The rate-distortion results for a Gaussian source are well-known. Codebooks are generally generated through a random construction using the source distribution. This results in a codebook with no structure and hence high encoding complexity.

Support provided by a University of California President's Post-doctoral Fellowship, NSF CAREER Grant 643836, and the Centre Bernoulli of École Polytechnique Fédérale de Lausanne.

In this paper, we first examine the quantization of random vectors when the encoding procedure includes a subspace projection step. We show that the restriction of a source coding codebook to a union of subspaces need not induce any performance penalty. In fact, a two-stage quantization for $\mathbb{R}^N$ of

(a) projecting to the nearest of J_N subspaces of dimension K_N in a random first-stage codebook of subspaces; followed by
(b) quantizing to the nearest of T_N codewords in a second-stage codebook within the subspace

induces no performance loss. That is, this structure allows the rate-distortion bound to be approached as $N \to \infty$ when J_N, K_N and T_N sequences are chosen appropriately.

After establishing the source coding results, we study the dual problem of channel coding. Consider the following signaling strategy for an additive white Gaussian noise (AWGN) channel: For N channel uses, pick a *subspace codebook* of J_N subspaces of dimension K_N and an *in-subspace codebook* of cardinality T_N in $\mathbb{R}^{K_N}$. We show that the communication of $\frac{1}{N}(\log_2 J_N + \log_2 T_N)$ bits per channel use in this manner induces no rate loss in that the Shannon capacity can be achieved as $N \to \infty$ while maintaining any desired ratio between $\log_2 J_N$ and $\log_2 T_N$. Furthermore, the decoding need not be exactly maximum likelihood over the codebook of cardinality $J_N T_N$; it suffices to use a two-stage decoder that first finds the nearest of the J_N subspaces and then finds the nearest of the T_N codewords in the selected subspace. While this has complexity exponential in N, it is reduced from an unstructured search.

A precedent to our channel coding results is the determination by Wyner [1] that antipodal codebooks can be used without a performance penalty. Our work extends the optimality from codebooks containing pairs

978-1-4244-2571-6/08/$25.00 ©2008 IEEE

of points on one-dimensional subspaces to sets of T_N points on K_N-dimensional subspaces.

The paper is structured to first define random sets of subspaces in Section II and then develop the source and channel coding results in Sections III and IV respectively. Our results use the geometry of sets of subspaces (the Grassmann manifold) and explicit calculations made with random matrix theory. Some of these extend earlier results from [2], [3].

II. Independent Subspace Sets

Our analysis is based on certain large, random sets of subspaces. To describe these precisely, we define as a (J, N, K)-*subspace set* any set

$$\mathbf{S} = \{S_1, \ldots, S_J\} \tag{1}$$

where each element S_j is a K-dimensional subspace of $\mathbb{R}^N$. The *sparsity ratio* of the set is $\alpha = K/N$. We call the union of the subspaces

$$\mathbf{X} = \left\{x \in \mathbb{R}^N \mid x \in S_j, \text{ for some } j = 1, \ldots, J\right\} \tag{2}$$

the (J, N, K)-*subspace signal set* generated by $\mathbf{S}$. The random subspace sets are described in the following definition.

Definition 1: A (J, N, K)-subspace set (1) is called *independent and uniformly generated* if

(a) the subspaces S_j are random, independently and identically distributed; and

(b) each S_j is the range of a random $N \times K$ matrix A_j, where the NK components of A_j are i.i.d. Gaussian scalars with zero mean and unit variance.

Part (b) of the definition is equivalent to the subspaces S_js being *rotationally invariant* in that each K-dimensional subspace, S_j, is identically distributed to US_j for any $N \times N$ orthogonal matrix U.

It is important to distinguish our sparse signal model above in Definition 1 from the more common random frame models considered in many recent papers such as [4]–[6]. Specifically, suppose that one chooses $M \geq N$ vectors $\{\varphi_i\}_{i=1}^M \subset \mathbb{R}^N$ such that the vectors span $\mathbb{R}^N$, making the set of vectors a frame [7]. With the additional (weak) condition that every subset of size K is linearly independent, one may generate a (J, N, K)-subspace set with $J = \binom{M}{K}$ subspaces of the form

$$\text{span}(\{\varphi_i\}_{i \in \mathcal{I}}) \qquad \text{where } |\mathcal{I}| = K.$$

Even if the φ_is are generated independently, such a subspace set is *not* independent and uniformly generated because the subspaces share generating vectors.

III. Subspace-Based Source Coding

Our first result shows that codebooks constructed from large collections of random subspaces can achieve the rate-distortion bound for a Gaussian signal. However, to prove this result, we will need to first characterize the approximation error between a Gaussian vector and a random (J, N, K)-signal set.

To this end, let $y \sim \mathcal{N}(0, I_N)$ be an N-dimensional Gaussian vector with zero mean and unit variance. Given any (J, N, K)-subspace signal set, $\mathbf{X}$, we can define the *relative approximation error* of y with respect to $\mathbf{X}$ as

$$\rho_{\min} = \frac{1}{\|y\|^2} \min_{x \in \mathbf{X}} \|y - x\|^2. \tag{3}$$

The approximation error $\rho_{\min}$ is between 0 and 1, with a value closer to 0 implying a lower approximation error.

Given N and J, define the *subspace rate* as

$$R_J = \frac{1}{N} \log_2 J, \tag{4}$$

which represents the number of bits per dimension to index the J subspaces. Also, for scalars p and $q \in (0, 1)$, let $D(p, q)$ denote the binary *Kullback-Leibler distance* [8] given by

$$D(p, q) = p \log_2 \left(\frac{p}{q}\right) + (1 - p) \log_2 \left(\frac{1 - p}{1 - q}\right). \tag{5}$$

With these two definitions, we have the following result that provides a simple analytic expression for the distance between a Gaussian vector and a random independent and uniformly generated subspace set.

Theorem 1 ([3]): Let $y \sim \mathcal{N}(0, I_N)$, and suppose $\mathbf{X}$ is an independent and uniformly generated (J, N, K)-subspace set as in Definition 1. Then, with $\alpha = K/N$ and R_J in (4) held constant as $N \to \infty$,

$$\lim_{N \to \infty} \rho_{\min} = \rho^*_{\min} \tag{6}$$

in distribution, where $\rho^*_{\min}$ is the unique solution to

$$2R_J = D(\alpha, 1 - \rho^*_{\min}), \quad \rho^*_{\min} \in (0, 1 - \alpha). \tag{7}$$

Using this result, we now consider a simple subspace-based quantizer. Suppose $y \sim \mathcal{N}(0, I_N)$ is a Gaussian random vector to be quantized with R bits per dimension. That is, there are a total of RN bits for the vector. Let $\mathbf{X}$ be an independent and uniformly generated (J, N, K)-subspace set. Consider the following simple two-stage quantizer:

- First, project y to the closest of the J subspaces in $\mathbf{S}$. The projection is given by

$$\hat{x} = \underset{x \in \mathbf{X}}{\text{argmin}} \|y - x\|^2.$$

Since there are J possible subspaces for $\hat{x}$, the subspace index can be encoded with $\log_2 J$ bits.

- In each of the K subspaces, design a quantizer with the remaining $RN - \log_2 J$ bits, and transmit the quantized value in the subspace. The final quantized value will be denoted by $\hat{y}$.

Any such two-stage quantizer will be called a *subspace-based quantizer* based on $\mathbf{X}$.

How well does such a quantizer work? Given a sparse signal model $\mathbf{X}$ and rate R, define the minimum average distortion,

$$D_N = \inf \frac{1}{N}\mathbf{E}\|y - \hat{y}\|^2, \tag{8}$$

where the infimum is over all two-stage subspace-based quantizers for the signal model $\mathbf{X}$ as described above. Since we are interested in information theoretic limits, we will allow the quantizer to vector quantize over multiple realizations of y with an arbitrarily long block length. The following result characterizes this minimum average distortion. A proof sketch is given in Section VI.

Theorem 2: For each N, let y be any random vector in $\mathbb{R}^N$ with zero mean and an identity covariance matrix. Fix K and J, and let $\mathbf{X}$ be an independent and uniformly generated (J, N, K)-subspace set generated independently of y. Let R be any quantization rate with $R > R_J$, where R_J is the subspace rate (4). Consider the limit of the distortion D_N in (8) of the subspace-based quantizer as $N \to \infty$ with the sparsity ratio $\alpha = K/N$ and the subspace rate R_J fixed. Then, the minimum of this limit over R_J is bounded by

$$\min_{R_J \leq R} \limsup_{N\to\infty} D_N \leq 2^{-2R}. \tag{9}$$

Suppose that the source vector y in Theorem 2 is Gaussian: $y \sim \mathcal{N}(0, I_N)$. We know from the Gaussian rate distortion function that, for *any* quantized estimate $\hat{y}$ of y with R bits per dimension, the average distortion is bounded below by

$$\frac{1}{N}\mathbf{E}\|y - \hat{y}\|^2 \geq 2^{-2R}.$$

In particular, the quantization error from sparse approximation must also satisfy the bound, $D_N \geq 2^{-2R}$. Theorem 2 shows that sparse approximation can achieve the rate distortion limit as $N \to \infty$ and the subspace rate, R_J, is selected appropriately.

One way to look at this result is a statement of optimality of a certain type of random codebook. The classic random codebook construction for quantizing an N-dimensional Gaussian random vector is to generate a codebook of 2^{NR} random Gaussian vectors. The codebook in Theorem 2 also, in essence, uses random codebook, but with 2^{NR_J} random K-dimensional subspaces, and then a codebook with $2^{N(R-R_J)}$ random vectors in each subspace. The quantization is two steps: first project to the closest subspace, then find the closest codebook vector in the subspace. Theorem 2 shows that for a Gaussian source and proper selection of R_J, this quantization is optimal.

IV. Subspace-Based Channel Coding

We next consider channel coding, the dual problem. However, first we need to establish a certain subspace estimation result. Specifically, suppose $\mathbf{X}$ is a (J, N, K)-subspace signal set. Given an unknown $x \in \mathbf{X}$, let y be a noisy observation of the form

$$y = x + d, \tag{10}$$

where $d \sim \mathcal{N}(0, I_N)$ is additive Gaussian noise. Since $x \in \mathbf{X}$, x belongs to one of J subspaces of dimension K. We consider the problem of detecting the subspace to which x belongs.

We analyze the following ML estimator. Let θ be the index of the subspace containing x. Under the assumption that d is additive Gaussian noise, the ML estimate for θ is given by

$$\hat{\theta} = \underset{j\in\{1,\ldots,J\}}{\operatorname{argmax}} \|P_j y\|, \tag{11}$$

where P_j is the orthogonal projection operator onto S_j.

Assumption 1: The signal x in (10) is of the form

$$x = V_\theta u, \tag{12}$$

where θ is an unknown subspace index, uniformly distributed on the set $\{1, \ldots, J\}$; each V_j is an orthogonal $N \times K$ matrix, with independent and rotationally invariant distributions; and u is a Gaussian random vector with $u \sim \mathcal{N}(0, I_K\gamma/\alpha)$ for some constant $\gamma > 0$.

Here, the constant $\gamma > 0$ represents the SNR, since it can be easily verified that $\gamma = \mathbf{E}\|x\|^2/\mathbf{E}\|d\|^2$. The following result provides a bound on the average probability of error in terms of the signal dimensions J, N and K and the SNR γ.

Theorem 3: Consider the subspace detection problem above. Let $\gamma_{\rm crit}$ be the solution to

$$R_J = \frac{1}{2}\log_2(1 + \gamma_{\rm crit}) - \frac{\alpha}{2}\log_2\left(1 + \frac{\gamma_{\rm crit}}{\alpha}\right), \tag{13}$$

where R_J is defined in (4). Then, if $\mathbf{X}$ is independent and uniformly generated, the limit of the error probability

$P_{\rm err}$ as $N \to \infty$ with γ, $\alpha = K/N$ and R_J held constant is given by

$$\lim_{N\to\infty} P_{\rm err} = \begin{cases} 0, & \gamma > \gamma_{\rm crit}; \\ 1, & \gamma < \gamma_{\rm crit}. \end{cases} \qquad (14)$$

The result shows that there is a critical SNR, $\gamma_{\rm crit}$, above which the subspace that contains x can be reliably detected. The expression for the critical SNR, γ_{crit} in (13) has a simple communications interpretation. Suppose $\mathbf{S}$ is an independent and uniformly generated (J, N, K)-subspace set. Consider a two-stage communication system using $\mathbf{S}$:

- Select one of the J subspaces, $S_j \in \mathbf{S}$.
- Transmit a signal on the K-dimensional subspace, S_j, with a Gaussian codebook.

The decoder reverses the process. First, it detects the transmitted subspace, S_j. Then, assuming the detected subspace is correct, it decodes the data transmitted on the subspace.

How well does this communication system do? In the first stage, assuming the decoder correctly detects the subspace, the communication system can communicate $\log_2 J$ bits in the subspace dimension. For the second stage, suppose that the noise is Gaussian and the total SNR is γ. Since the system concentrates its power on K of the N dimensions, the maximum possible rate in the second stage is given by the Shannon limit,

$$\frac{K}{2}\log_2\left(1 + \frac{\gamma N}{K}\right) = \frac{K}{2}\log_2\left(1 + \frac{\gamma}{\alpha}\right).$$

So, the total rate normalized by the dimension N is

$$\begin{aligned} R &= \frac{1}{N}\log_2 J + \frac{K}{2N}\log_2\left(1 + \frac{\gamma}{\alpha}\right) \\ &= R_J + \frac{\alpha}{2}\log_2\left(1 + \frac{\gamma}{\alpha}\right). \end{aligned} \qquad (15)$$

Now, if we let the dimension $N \to \infty$, Theorem 3 shows that for any subspace rate R_J with

$$R_J \le \frac{1}{2}\log_2(1+\gamma) - \frac{\alpha}{2}\log_2\left(1 + \frac{\gamma}{\alpha}\right),$$

the decoder with exhaustive ML detection can detect the correct subspace. Maximizing the rate in (15) over such R_Js we can achieve a rate of

$$R = \frac{1}{2}\log_2(1+\gamma).$$

This is precisely the capacity for a Gaussian channel. We conclude that the two-stage communication system, with the appropriate subspace rate, R_J, is optimal.

The result is the exact dual of the rate-distortion result in Section III. Similar to that section, we can interpret the coding scheme as a certain random codebook. The codebook uses $2^{R_J N}$ random subspaces and then $2^{(R-R_J)N}$ Gaussian codewords in each subspaces.

V. Interpreting the Results

The results presented here are not intended as practical code constructions. However, one may interpret the subspace structure as yielding a complexity reduction.

Consider the source coding problem in dimension N with rate R bits per component. Naive search for the nearest codeword requires an $O(N)$-complexity distance computation for each of 2^{RN} codewords. The overall complexity is thus $O(N\,2^{RN})$.

The two-stage strategy described in Section III instead requires a search over $J = 2^{R_J N}$ subspaces (for ML subspace detection) followed by a search amongst $2^{(R-R_J)N}$ codewords in a K-dimensional space. Since the computation of distance to a K-dimensional subspace in $\mathbb{R}^N$ has $O(KN)$ complexity, the first stage has overall complexity $O(KN\,2^{R_J N})$. The complexity of the second stage is $O(K\,2^{(R-R_J)N})$. Though this accounting for complexity is obviously coarse, it can be seen that subspace structure and two-stage encoding have lowered the encoding complexity. Either method could be improved using techniques described in [9].

VI. Proof Sketch of Theorem 2

Lemma 1: Suppose R, $R_J > 0$ and α and $\rho \in (0,1)$ satisfy

$$2R_J = D(\alpha, 1-\rho) \qquad (16)$$

and

$$\rho = (1-\alpha)2^{-2R}. \qquad (17)$$

Then, $R > R_J$ and

$$2^{-2R} = \rho + (1-\rho)2^{-2(R-R_J)/\alpha}. \qquad (18)$$

Proof: From the definition of the Kullback-Leibler distance $D(\cdot,\cdot)$ in (5),

$$2^{D(\alpha,1-\rho)} = \left(\frac{\alpha}{1-\rho}\right)^{\alpha}\left(\frac{1-\alpha}{\rho}\right)^{1-\alpha}.$$

With some algebraic manipulations of (16) and (17), (18) can now be shown. Also, rearranging (18) and using the fact that $2^{-2R} < 1$, one can show $R > R_J$. ☐

Proof of Theorem 2: As usual, let S_j, $j = 1, \ldots, J$, denote the K-dimensional subspaces generating the (J, N, K)-subspace signal set $\mathbf{X}$. Given a vector y, let θ denote the index of the subspace closest to y. Therefore, the subspace containing $\hat{x}$ is S_θ.

Now, the quantization error $y - \hat{y}$ can be divided into two terms,

$$y - \hat{y} = (y - \hat{x}) + (\hat{x} - \hat{y}).$$

The first term is the projection error of y onto S_θ, and the second term is the quantization error in the subspace S_θ. Since the projection of y onto S_θ is orthogonal,

$$\|y - \hat{y}\|^2 = \|y - \hat{x}\|^2 + \|\hat{x} - \hat{y}\|^2. \quad (19)$$

We bound the two terms in (19) separately.

The first term on the right hand side of (19) represents the approximation error of representing y with the sparse signal model. Using the definition of $\rho_{\min}$ in (3) implies that this term is given by $\|y - \hat{x}\|^2 = \rho_{\min}\|y\|^2$. A more detailed analysis shows that $\rho_{\min}$ is independent of y and therefore

$$\mathbf{E}\|y - \hat{x}\|^2 = \mathbf{E}\rho_{\min}\mathbf{E}\|y\|^2 = N\mathbf{E}\rho_{\min}, \quad (20)$$

where in the last step we that y has unit variance.

The second term, $\|\hat{x} - \hat{y}\|^2$, on the right hand side of (19) represents the quantization error of $\hat{x}$ in the subspace S_θ. Since $\hat{x}$ is the projection of y onto the closest of J subspaces, its distribution is not, in general, Gaussian. Consequently, the minimum achievable distortion in quantizing $\hat{x}$ is difficult to evaluate exactly. However, it is well-known that, for any random variable, the minimum distortion in quantizing the variable is always less than or equal to the achievable distortion in quantizing a Gaussian of the same variance with the same number of bits. Therefore, using the distortion-rate function of a Gaussian random variable, the distortion in quantizing $\hat{x}$ is bounded above by

$$\mathbf{E}\|\hat{x} - \hat{y}\|^2 \leq \mathbf{E}\|\hat{x}\|^2 2^{2-R_x}, \quad (21)$$

where R_x is the number of bits per dimension in quantizing $\hat{x}$. So, we need to evaluate the energy, $\mathbf{E}\|\hat{x}\|^2$, and the rate, R_x.

Since the projection of y to $\hat{x}$ is orthogonal,

$$\|\hat{x}\|^2 = \|y\|^2 - \|y - \hat{x}\|^2 = (1 - \rho_{\min})\|y\|^2.$$

and therefore,

$$\mathbf{E}\|\hat{x}\|^2 = N(1 - \mathbf{E}\rho_{\min}). \quad (22)$$

Also, recall that there are $RN - \log_2 J$ bits to quantize the vector $\hat{x}$. Since the quantization is in one of the K-dimensional subspaces S_j, the number of bits per dimension in the subspace is,

$$R_x = \frac{1}{K}(NR - \log_2 J) = \frac{N(R - R_J)}{K} = (R - R_J)/\alpha. \quad (23)$$

Substituting (22) and (23) into (21), we obtain

$$\mathbf{E}\|\hat{x} - \hat{y}\|^2 \leq N(1 - \mathbf{E}\rho_{\min})2^{-2(R-R_J)/\alpha}. \quad (24)$$

Finally, combining (19) and (20) and (24), we can bound the distortion by

$$D_N \leq \frac{1}{N}\mathbf{E}\|y - \hat{y}\|^2 \leq \mathbf{E}\rho_{\min} + (1 - \mathbf{E}\rho_{\min})2^{-2(R-R_J)/\alpha}. \quad (25)$$

Now, Theorem 1 shows that, for a independent and uniformly generated subspace set, $\rho_{\min} \to \rho^*_{\min}$, where the convergence is in distribution and $\rho^*_{\min}$ is the unique solution to (7), Since $\rho_{\min} \to \rho^*_{\min}$ in distribution, $\mathbf{E}\rho_{\min} \to \rho^*_{\min}$. Substituting this limit into the bound (25) gives

$$\limsup_{N\to\infty} D_N \leq \rho^*_{\min} + (1 - \rho^*_{\min})2^{-2(R-R_J)/\alpha}. \quad (26)$$

To complete the proof of the theorem, we need to minimize this bound over $R_J \leq R$. To this end, let

$$\rho = (1 - \alpha)2^{-2R} \text{ and } 2R_J = D(\rho, 1 - \alpha).$$

By construction, ρ satisfies (7). Since $\rho^*_{\min}$ is the unique solution to this equation, $\rho^*_{\min} = \rho$. Also, Lemma 1 implies that $R_J < R$ and

$$\limsup_{N\to\infty} D_N \leq \rho^*_{\min} + (1 - \rho^*_{\min})2^{-2(R-R_J)/\alpha} = 2^{-2R}.$$

This completes the proof. □

REFERENCES

[1] A. D. Wyner, "Capabilities of bounded discrepancy decoding," *Bell Syst. Tech. J.*, vol. 44, pp. 1061–1122, Jul.–Aug. 1965.

[2] 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. Process.*, vol. 2006, pp. 1–19, Mar. 2006.

[3] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Rate-distortion bounds for sparse approximation," in *IEEE Statist. Sig. Process. Workshop*, Madison, WI, Aug. 2007, pp. 254–258.

[4] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

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

[6] A. K. Fletcher, S. Rangan, and V. K. Goyal, "Necessary and sufficient conditions on sparsity pattern recovery," arXiv:0804.1839v1 [cs.IT]., Apr. 2008.

[7] O. Christensen, *An Introduction to Frames and Riesz Bases*. Boston, MA: Birkhäuser, 2003.

[8] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: John Wiley & Sons, 1991.

[9] N. Moayeri and D. L. Neuhoff, "Time-memory tradeoffs in vector quantizer codebook searching based on decision trees," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 4, pp. 490–506, Oct. 1994.