Multiple Descriptions with Codebook Reuse

Young-Han Kim

Abstract— **Multiple description coding, in general, requires separate codebooks for each description. Motivated by the problem of sparse linear representation, we propose a simple coding scheme that recursively describes successive description errors using the same codebook, resulting in a sparse linear combination of codewords that achieves the Gaussian rate distortion region. This result, in particular, provides an elementary proof of successive refinability and additive refinability of white Gaussian sources.**

I. Introduction

Recently, recovery of sparse signals via linear measurements has drawn much attention in the literature. Most notably, Candes and Tao [2] and Donoho [5] showed that sparse signals can be reconstructed *efficiently* from an underdetermined system of linear equations, opening up the exciting field of compressed sensing. There have been several follow-up discussions that connect compressed sensing and information theory; we refer the reader to [13], [1], [8], [16], [19], [10] for various aspects of the connection between two fields.

In this paper, we consider a problem that is, in a sense, dual to the problem of sparse signal recovery (in particular, the problem of sparse signal position recovery):

> How (well) can one represent a signal as a sparse linear combination of codewords in an overcomplete dictionary?

More formally, let $\mathcal{C} = \{\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_M\}$ be a collection of M vectors in $\mathbb{R}^n$. We call $\mathcal{C}$ a dictionary (or a codebook) of size (M, n) and call $\boldsymbol{\phi}_m$, $m = 1, \ldots, M$ codewords.

For each $\mathbf{y} \in \mathbb{R}^n$, we define its best k-linear combination $\hat{\mathbf{y}}_k$ as

$$\hat{\mathbf{y}}_k = x_1\boldsymbol{\phi}_{m_1} + x_2\boldsymbol{\phi}_{m_2} + \ldots + x_k\boldsymbol{\phi}_{m_k},$$

where $x_1, \ldots, x_k \in \mathbb{R}$ and $m_1, \ldots, m_k \in \{1, 2, \ldots, M\}$ are chosen to minimize the mean squared error distortion

$$d_k(\mathbf{y}) = \|\mathbf{y} - x_1\boldsymbol{\phi}_{m_1} + x_2\boldsymbol{\phi}_{m_2} + \ldots + x_k\boldsymbol{\phi}_{m_k}\|^2.$$

We then define the k-sparse distortion $d_k^*(\mathcal{C})$ of the codebook as

$$d_k^*(\mathcal{C}) = \sup_{\mathbf{y}: \|\mathbf{y}\|^2 \le 1} d_k(\mathbf{y}),$$

where the supremum is taken over all n-vectors $\mathbf{y}$ in the (closed) unit sphere.

This work was supported in part by the National Science Foundation CAREER Award CCF-0747111.

The author is with the Department of Electrical and Computer Engineering, University of California, San Diego, La Jolla, CA 92093-0407 (email: yhk@ucsd.edu).

Clearly, if $M < n$, then $d_k^*(\mathcal{C}) = 1$ for any dictionary $\mathcal{C}$ of size (M, n). Hence, we consider the case $M \ge n$, that is, the case in which the dictionary is *overcomplete*.

The main result of this paper is the following:

Theorem 1. *Let M_n be a sequence of integers with*

$$\liminf_{n\to\infty} \frac{\log M_n}{n} > \frac{1}{2k} \log\left(\frac{1}{D}\right)$$

for some $0 < D < 1$. Then there exists a sequence of (M_n, n) dictionaries $\mathcal{C}_n$ such that

$$\limsup_{n\to\infty} d_k^*(\mathcal{C}_n) \le D.$$

Note that $D \ge 1$ is trivial, for a dictionary with a single element $\boldsymbol{\phi}_1 = \mathbf{0}$ would suffice. Also note that the case $k = 1$ is equivalent to the standard Gaussian rate distortion theorem [14], [4, Section 10.3.2]. Indeed, we will show that a simple iterative representation method, which is a special case of successive refinement coding [7], achieves the desired distortion level. The basic idea is to represent the signal, the error, the error of the error, etc. recursively. This representation method, incidentally, gives a very simple proof of successive refinability [7] and additive successive refinability [15] of white Gaussian sources under the mean squared distortion.

II. Proof Sketch of Theorem 1

Define $R(\delta) = (1/2)\log(1/\delta)$ for $0 < \delta < 1$ and let $R > R(\delta)$. We recall that a random codebook (dictionary) $\mathcal{C}_1$ of size $(2^{nR}, n)$ with codewords $\phi_j^{(1)}$ independently generated according to $N(0, 1/n)$ can asymptotically cover the surface of the unit sphere with expected distortion δ; see, for example, [12]. In particular, any $\mathbf{y}$ in the unit sphere can be represented as

$$\mathbf{y} = \hat{\mathbf{y}}_1 + \mathbf{z}_1 = \sqrt{(1-\delta)||\mathbf{y}||^2}\boldsymbol{\phi}_{m_1}^{(1)} + \mathbf{z}_1 \tag{1}$$

for some $\boldsymbol{\phi}_{m_1}^{(1)} \in \mathcal{C}_1$ and for some $\mathbf{z}_1$ with $\|\mathbf{z}_1\|^2 \le \delta\|\mathbf{y}\|^2$ with high probability. Now consider another codebook $\mathcal{C}_2 = \{\boldsymbol{\phi}_1^{(2)}, \ldots, \boldsymbol{\phi}_{2^{nR}}^{(2)}\}$ generated independently from $\mathcal{C}_1$, but in the exactly same manner. Then by using (1) once again, $\mathbf{z}_1$ can be represented as

$$\mathbf{z}_1 = \hat{\mathbf{z}}_1 + \mathbf{z}_2 = \sqrt{(1-\delta)||\mathbf{z}_1||^2}\boldsymbol{\phi}_{m_2}^{(2)} + \mathbf{z}_2$$

for some $\boldsymbol{\phi}_{m_2}^{(2)} \in \mathcal{C}_2$ and for some $\mathbf{z}_2$ with $\|\mathbf{z}_2\|^2 \le \delta\|\mathbf{z}_\|^2$ with high probability. In general, we can consider a sequence of randomly generated codebooks $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k$ such that the error $\mathbf{z}_j$ after stage j, $j = 1, \ldots, k-1$, can be represented as

$$\mathbf{z}_j = \hat{\mathbf{z}}_j + \mathbf{z}_{j+1} = \sqrt{(1-\delta)||\mathbf{z}_j||^2}\boldsymbol{\phi}_{m_2}^{(j+1)} + \mathbf{z}_{j+1}$$

for some $\phi^{(j+1)}_{m_{j+1}} \in \mathcal{C}_{j+1}$ and for some $\mathbf{z}_{j+1}$ with $\|\mathbf{z}_{j+1}\|^2 \leq \delta\|\mathbf{z}_j\|^2$ with high probability.

Substituting $\mathbf{z}_j$ recursively, with high probability

$$\begin{aligned}\mathbf{y} = \sqrt{(1-\delta)||\mathbf{y}||^2}\boldsymbol{\phi}^{(1)}_{m_1} + \sqrt{(1-\delta)||\mathbf{z}_1||^2}\boldsymbol{\phi}^{(2)}_{m_2} \\ + \ldots + \sqrt{(1-\delta)||\mathbf{z}_{k-1}||^2}\boldsymbol{\phi}^{(k)}_{m_k} + \mathbf{z}_k\end{aligned}$$

with $\|\mathbf{z}_k\|^2 \leq \delta^k$. Thus, the super-codebook

$$\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \cdots \cup \mathcal{C}_k$$

of size $k2^{nR}$ can represent each $\mathbf{y}$ with distortion $\leq \delta^k$, with high probability.

Now the above coding technique can be augmented in two directions by a result of Wyner [17] on sphere covering, which states that for n sufficiently large and $R > R(\delta)$, there exists a codebook $\mathcal{C}^*$ of size $(2^{nR}, n)$ such that $\mathcal{C}^*$ *uniformly* covers the entire unit sphere with distortion δ. Incorporating Wyner's lemma with the above iterative coding scheme, we see immediately that every $\mathbf{y}$ in the unit sphere has the k-spare linear representation

$$\begin{aligned}\mathbf{y} = \boldsymbol{\phi}_{m_1} + \sqrt{1-\delta}\boldsymbol{\phi}_{m_2} \\ + \ldots + \sqrt{(1-\delta)^{k-1}}\boldsymbol{\phi}_{m_k} + \mathbf{z}_k\end{aligned}$$

with $\|\mathbf{z}_k\|^2 \leq \delta^k$ and $\boldsymbol{\phi}_{m_1}, \boldsymbol{\phi}_{m_2}, \ldots, \boldsymbol{\phi}_{m_k} \in \mathcal{C}^*$. (Recall that after stage j, the resulting error is in the sphere of radius $\delta^{j/2}$, allowing recursive application of the same codebook $\mathcal{C}^*$, scaled by $\delta^{j/2}$.) Taking $D = \delta^k$ and noting that

$$R(\delta) = \frac{1}{2}\log\left(\frac{1}{\delta}\right) = \frac{1}{2k}\log\left(\frac{1}{D}\right),$$

we have the desired proof of the theorem.

III. Discussion

The proof in the previous section leads to a simple proof of successive refinability of white Gaussian sources. Indeed, because the (expected) distortion after stage j is given by δ^j, while the total description rate is $(j/2)\log(1/\delta)$, the rate-distortion tradeoff for each stage traces the Gaussian rate distortion function $R(D) = (1/2)\log(1/D)$. (Recall that we don't need to describe the scaling factors $1, \sqrt{1-\delta}, \ldots$, since these are constants independent of n. Furthermore, the same argument easily extends to the case in which incremental rates $R_1, R_2, \ldots, R_k$ are not necessarily identical; one can even prove the existence of *nested* codebooks (up to scaling) that uniformly cover the unit sphere. Finally, we can use the law of large numbers to see that lossy source coding of a white Gaussian source is asymptotically (in n) equivalent to lossy source coding of a source that is generated uniformly on the sphere. Hence, we have shown the successive refinability of white Gaussian sources.

Operationally, the recursive coding scheme for successive refinement (i.e., describing the error, the error of the error, ...) can be viewed as a dual procedure to *succesive cancellation* [3], [18] for the Gaussian multiple access channels. For both cases, one strives to best solve the single-user source [channel] coding problem at each stage and progresses recursively by subtracting off the encoded [decoded] part of the source [channel output] $\mathbf{y}$. This duality can be complemented by an interesting connection between the orthogonal matching pursuit and the sucessive cancellation [11] and the duality between signal recovery and signal representation. Note, however, that the duality here is mostly conceptual and cannot be made more precise. For example, while we can use a single codebook for each of k successive descriptions (again up to scaling) as shown above, one cannot use the same codebook for all k users in the Gaussian multiple access channel. If the channel gains are identical among users, it is impossible to distinguish who sent which message (from the same codebook), even without any additive noise!

While this paper has studied the case in which every successive combination of descriptions is good, one can in general consider the problem of multiple descriptions [6], in which every (not necessarily successive) combination of descriptions should be good (with some tradeoff). As a partial step to solve this problem, one can show that there exists a multiple description codebook that uniformly covers the unit sphere under no excess marginal rate case [20].

Finally, Theorem 1 shows that the k-sparse linear representation from a dictionary of size $(2^{nR}, n)$ can achieve the distortion 2^{-2kR}. Is this the best distortion? The answer is positive and it can be shown by tweaking the converse proof of the rate distortion theorem. The details will be reported elsewhere [9].

References

[1] S. Aeron, M. Zhao, and V. Saligrama, "Fundamental tradeoffs between sparsity, sensing diversity and sensing capacity," in *Proceedings of 40th Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, 2006.

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

[3] T. M. Cover, "Some advances in broadcast channels," in *Advances in Communication Systems*, A. J. Viterbi, Ed. San Francisco: Academic Press, 1975, vol. 4, pp. 229–260.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

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

[6] A. El Gamal and T. M. Cover, "Achievable rates for multiple descriptions," *IEEE Trans. Inform. Theory*, vol. IT-28, no. 6, pp. 851–857, 1982.

[7] W. H. R. Equitz and T. M. Cover, "Successive refinement of information," *IEEE Trans. Inform. Theory*, vol. IT-37, no. 2, pp. 269–275, 1991.

[8] A. K. Fletcher, S. Rangan, and V. K. Goyal, "On the rate-distortion performance of compressed sensing," in *Proc. ICASSP*, Honolulu, HI, 2007.

[9] H. Jeong and Y.-H. Kim, "Successive linear representation," in preparation.

[10] Y. Jin and B. Rao, "Insights into the stable recovery of sparse solutions in overcomplete representations using network information theory," in *Proc. ICASSP*, Las Vegas, NV, 2008.

[11] ——, "Performance limits of matching pursuit algorithms," in *Proc. IEEE International Symposium on Information Theory*, Toronto, ON, July 2008.

[12] A. Lapidoth, "On the role of mismatch in rate distortion theory," *IEEE Trans. Inform. Theory*, vol. IT-43, no. 1, pp. 38–47, 1997.

[13] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Measurements vs. bits: Compressed sensing meets information theory," in *Proceedings of 44th Allerton Conf. Comm., Ctrl., Computing*, Monticello, IL, 2006.

[14] C. E. Shannon, "Coding theorems for a discrete source with a fidelity criterion," in *IRE Int. Conv. Rec., part 4*, 1959, vol. 7, pp. 142–163, reprinted with changes in *Information and decision processes,* R. E. Machol, Ed. New York: McGraw-Hill, 1960, pp. 93-126.

[15] E. Tuncel and K. Rose, "Additive successive refinement," *IEEE Trans. Inform. Theory*, vol. IT-49, no. 8, pp. 1983–1991, 2003.

[16] M. Wainwright, "Information-theoretic bounds on sparsity recovery in the high-dimensional and noisy setting," in *Proc. IEEE International Symposium on Information Theory*, Nice, France, June 2007.

[17] A. D. Wyner, "Random packings and coverings of the unit n-sphere," *Bell System Tech. J.*, vol. 46, pp. 2111–2118, 1967.

[18] ——, "Recent results in the Shannon theory," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 2–10, 1974.

[19] F. Zhang and H. D. Pfister, "Compressed sensing and linear codes over real numbers," in *Proceedings of the UCSD Information Theory and Applications Workshop*, La Jolla, CA, 2008.

[20] Z. Zhang and T. Berger, "Multiple description source coding with no excess marginal rate," *IEEE Trans. Inform. Theory*, vol. IT-41, no. 2, pp. 349–357, 1995.