Performance of Sparse Representation Algorithms Using Randomly Generated Frames

Mehmet Akçakaya and Vahid Tarokh

***Abstract*—We consider sparse representations of signals with at most L nonzero coefficients using a frame $\mathcal{F}$ of size M in $\mathbb{C}^N$. For any $\mathcal{F}$, we establish a universal numerical lower bound on the average distortion of the representation as a function of the sparsity $\epsilon = L/N$ of the representation and redundancy $(r-1) = M/N - 1$ of $\mathcal{F}$. In low dimensions (e.g., $N = 6, 8, 10$), this bound is much stronger than the analytical and asymptotic bounds given in another of our papers. In contrast, it is much less straightforward to compute. We then compare the performance of *randomly generated frames* to this numerical lower bound and to the analytical and asymptotic bounds given in the aforementioned paper. In low dimensions, it is shown that randomly generated frames perform about 2 dB away from the theoretical lower bound, when the optimal sparse representation algorithm is used. In higher dimensions, we evaluate the performance of randomly generated frames using the greedy orthogonal matching pursuit (OMP) algorithm. The results indicate that for small values of ϵ, OMP performs close to the lower bound and suggest that the loss of the suboptimal search using orthogonal matching pursuit algorithm grows as a function of ϵ. In all cases, the performance of randomly generated frames hardens about their average as N grows, even when using the OMP algorithm.**

***Index Terms*—Distortion, orthogonal matching pursuit, performance bounds, random frames, sparse representations.**

I. Introduction

Consider a set $\mathcal{F}$ of $M \geq N$ nonzero signals in an N-dimensional complex vector space $\mathcal{W}$ such that $\mathcal{F}$ spans $\mathcal{W}$. We refer to $\mathcal{F}$ as a *frame* or a *dictionary* for $\mathcal{W}$. For $\mathbf{r} \in \mathcal{W}$, there are possibly infinitely many ways to represent $\mathbf{r}$ as a linear combination of the elements of $\mathcal{F}$. In various applications [4], we are interested in the sparsest representation of $\mathbf{r}$ with the lowest number of nonzero coefficients (referred to as the $\mathcal{L}_0$ norm of the representation vector).

The *noiseless* sparse representation problem is to find the sparsest representation, whenever $\mathbf{r} \in \mathcal{W}$ is known to have such a sparse representation with the number of nonzero coefficients less than or equal to L. A solution to this problem was given in [1] for a class of frames referred to as the *Vandermonde frames* when $L \leq N/2$.

When $\mathbf{r} \in \mathcal{W}$ is not known to have an exact sparse representation with the number of nonzero coefficients less than or equal to L, then it is expected that any such *noisy* sparse representation suffers from some *distortion*. In this case, two classes of *noisy representation problems* have been considered in the literature, namely, *error constrained sparse approximation* (ECSA) and *sparsity constrained approximation* (SCA) problems. These are, respectively, formulated as

$$\min_{\mathbf{c}\in\mathbb{C}^M} \|\mathbf{c}\|_0 \quad \text{s.t. } \|\mathbf{r}-\mathbf{cF}\|_2^2 \leq \delta\|\mathbf{r}\|^2 \quad \text{(ECSA)} \tag{1}$$

and

$$\min_{\mathbf{c}\in\mathbb{C}^M} \|\mathbf{r}-\mathbf{cF}\|_2^2 \quad \text{s.t. } \|\mathbf{c}\|_0 \leq \epsilon N \quad \text{(SCA)} \tag{2}$$

where $\mathbf{F}$ is the matrix whose rows are the elements of the frame, $\mathcal{F}$.

There is an intimate relationship between the ECSA and SCA problems [1]. Both problems have been extensively studied using either $\mathcal{L}_1$ regularization [2], [4] or by using the greedy orthogonal matching pursuit (OMP) algorithm [6]. The central theme of the vast majority of the existing body of research is to find frame structures and associated sparsity and redundancy criteria for which the $\mathcal{L}_1$ regularization method solves the sparsest representation problem for every $\mathbf{r}$. This *worst-case* approach may be too conservative, since in various practical applications, the performance for a typical signal (i.e., an *average performance measure*) is of interest.

Recently, the authors have derived an analytical lower bound on the average distortion of any frame, as a function of the sparsity $\epsilon = L/N$ of the representation and redundancy $(r-1) = M/N - 1$ of $\mathcal{F}$ [1]. However, this analytical bound is not tight in low dimensions. In this letter, we first establish a numerical lower bound of the same spirit, that is much tighter than the bound of [1] in low dimensions. Then we focus on *randomly generated Gaussian frames* and compare their performance to the numerical lower bound, as well as the analytical and asymptotic lower bounds of [1]. We numerically show that in low dimensions, using the optimal sparse representation algorithm, these random frames, in the average sense, perform close to the theoretical lower bound.

II. Numerical Lower Bound on Average Distortion

As in [1], we define the average distortion for any frame $\mathcal{F}$ and sparsity factor $0 \leq \epsilon \leq 1$ by

$$D(\mathcal{F}) = \frac{1}{N}\mathbb{E}_{\mathbf{r}} \min \|\mathbf{r}-\mathbf{cF}\|^2$$

where the minimum is taken over all representations $\mathbf{c}$ of $\mathbf{r}$ with $\|\mathbf{c}\|_0 \leq \epsilon N$, and the expectation is for $\mathbf{r}$ uniformly distributed on the N dimensional complex hypersphere of radius

Manuscript received January 18, 2007; revised April 4, 2007. This work was presented in part at the Conference on Information Sciences and Systems, Johns Hopkins University, Baltimore, MD, Mar.2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Shahram Shahbazpanahi.

The authors are with the School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138 USA (e-mail: akcakaya@deas.harvard.edu; vahid@deas.harvard.edu).

Digital Object Identifier 10.1109/LSP.2007.901683

$\sqrt{N}$ centered at the origin. This definition is motivated by the observation that given a probability distribution $F(\cdot)$ on $\mathbf{r}$, one must design frames that minimize the average distortion for vectors generated according to this distribution. When no *a priori* knowledge of $\mathbf{r}$ is assumed, by using a simple scaling transformation, it can be assumed that $\mathbf{r}$ is uniformly distributed on the complex hypersphere of radius $\sqrt{N}$ centered at the origin [1].

Let $T = \binom{M}{L}$. We consider all the L-dimensional subspaces of $\mathcal{W}$ that are spanned by all subsets of size L of $\{\boldsymbol{\phi}_j\}_{j\in\mathcal{I}_k}$. There are $T_* \leq T$ such distinct L-dimensional subspaces, which we denote by $\{\mathcal{P}_k, k = 1, 2, \cdots, T_*\}$. Given a vector $\mathbf{r}$ on the N-dimensional complex hypersphere, the SCA algorithm finds the closest $\mathcal{P}_k$, $k = 1, 2, \cdots, T_*$ to $\mathbf{r}$. In other words, it minimizes $\|\mathbf{r} - \Pi_{\mathcal{P}_k}\mathbf{r}\|^2$, where $\Pi_{\mathcal{P}_k}$ is the projection operator onto $\mathcal{P}_k$.

Using this geometric interpretation, we define an L-dimensional complex generalized cap (GC) of radius $\sqrt{\rho}$ around an L-dimensional plane $\mathcal{P}_k$ as

$$GC_L(\rho, \mathcal{P}_k) = \left\{\mathbf{x} \in \mathbb{S}^N : \|\mathbf{x} - \Pi_{\mathcal{P}_k}\mathbf{x}\|^2 \leq \rho\right\} \quad (3)$$

where $\mathbb{S}^N$ is the N-dimensional complex unit hypersphere

$$\left\{\mathbf{x} \in \mathbb{C}^N : \|\mathbf{x}\|^2 = 1\right\}.$$

In order to calculate $D(\mathcal{F})$, we are interested in the distribution of $d^2(\mathbf{x}, \mathcal{P}_k) \triangleq \|\mathbf{x} - \Pi_{\mathcal{P}_k}\mathbf{x}\|^2$ and $\mathbf{x}$ is uniformly distributed on $\mathbb{S}^N$. We note that this is a rescaling of the vector $\mathbf{r}$ by a factor of $1/\sqrt{N}$. Clearly, for any given $\mathbf{x} \in \mathbb{S}^N$, we have

$$\begin{aligned}\mathbb{P}\left(\min_k d^2(\mathbf{x}, \mathcal{P}_k) \leq \eta\right) &= \mathbb{P}(\exists \text{ a plane within } \sqrt{\eta} \text{ of } \mathbf{x}) \\ &= \mathbb{P}(\mathbf{x} \text{ is in the area covered by the GCs of radius } \sqrt{\eta}) \\ &= \mathbb{P}\left(\mathbf{x} \in \bigcup_{k=1}^{T_*} GC_L(\mathcal{P}_k, \eta) | \mathbf{x}\right).\end{aligned}$$

Since $T_* \leq T$

$$\mathbb{P}\left(\mathbf{x} \in \bigcup_{k=1}^{T_*} GC_L(\mathcal{P}_k, \eta)\right) \leq T\frac{\mathcal{A}(GC_L(\eta))}{\mathcal{A}(\mathbb{S}^N)}. \quad (4)$$

Thus

$$\mathbb{P}\left(\min_k d^2(\mathbf{x}, \mathcal{P}_k) \geq \eta\right) \geq \max\left(1 - T\frac{\mathcal{A}(GC_L(\eta))}{\mathcal{A}(\mathbb{S}^N)}, 0\right). \quad (5)$$

Since the area of the generalized cap $\mathcal{A}(GC_L(\eta))$ is a strictly increasing function of η, the equation $1 - T\mathcal{A}(GC_L(\eta))/\mathcal{A}(\mathbb{S}^N) = 0$ has a *unique root*. We also note that

$$\mathbb{E}\left(\min_k d^2(\mathbf{x}, \mathcal{P}_k)\right) = \int_0^1 \mathbb{P}\left(\min_k d^2(\mathbf{x}, \mathcal{P}_k) \geq \eta\right) d\eta.$$

The following theorem now follows easily from the above inequalities.

Theorem 2.1: Let $0 \leq \rho_c \leq 1$ be the *unique root* of the equation

$$\frac{\mathcal{A}(GC_L(\eta))}{\mathcal{A}(\mathbb{S}^N)} = \frac{1}{T} \quad (6)$$

then for any frame $\mathcal{F}$ of size M in $\mathbb{C}^N$, the average distortion of sparse representations using at most $L = \epsilon N$ nonzero coefficients satisfies

$$D(\mathcal{F}) \geq \int_0^{\rho_c} \left(1 - T\frac{\mathcal{A}(GC_L(\eta))}{\mathcal{A}(\mathbb{S}^N)}\right) d\eta. \quad (7)$$

□

The following lemma has been proven in [1] and can be used to numerically compute the value of ρ_c in the above theorem.

Lemma 2.2: For any $0 \leq \eta \leq 1$, we have

$$\frac{\mathcal{A}(GC_L(\eta))}{\mathcal{A}(\mathbb{S}^N)} = \int_0^{\eta} \frac{\Gamma(N)}{\Gamma(L)\Gamma(N-L)} x^{N-L-1}(1-x)^{L-1} dx.$$

□

Since $\rho_c \in [0, 1]$ is unique and the left side of (6) is an increasing function of η, the value of ρ_c can be found using standard numerical techniques. The lower bound of Theorem 2.1 can also be computed using numerical integration.

III. Performance of Randomly Generated Frames

It has been pointed out in [3] that randomly generated frames are asymptoticaly good in the worst case, when using $\mathcal{L}_1$ regularization. Motivated by these results, we next study the performance of randomly generated frames using the aforementioned average distortion criteria. For $N = 6, M = 12$; $N = 8, M = 16$; and $N = 10, M = 20$, we generate 32 random frames according to an N-dimensional complex Gaussian distribution with mean zero and covariance matrix I_N. Using Monte Carlo simulations and 1000 uniformly generated points on the complex hypersphere of radius $\sqrt{N}$ (the uniform distribution on the hypersphere is obtained via normalizing N-dimensional complex Gaussian vectors with mean zero and covariance matrix I_N [5]), we calculate the average distortion when using the optimal sparse representation algorithm. We tabulate $D(\mathcal{F})$ for $0 \leq L \leq N$ and compare the results to the numerical bound of Theorem 2.1 and the analytical bound of [1]. The results are presented in Figs. 1–3.

The results indicate that for $N = 6$, $N = 8$, and $N = 10$ randomly generated frames perform, respectively, about 2.1, 2, and 1.7 dB from the numerical lower bound of Theorem 2.1 for a sparsity of $\epsilon = 1/2$. It is also seen that the performance of randomly generated frames hardens about their average as N grows. For comparison, we tabulate the numerical lower bound of Theorem 2.1 and the analytical bound of [1] in low dimensions. Clearly the numerical bound is much tighter for small N.

In higher dimensions, finding the optimal sparse representation is complex. We thus evaluate the performance of randomly generated frames using the OMP algorithm. For $N = 256$ and

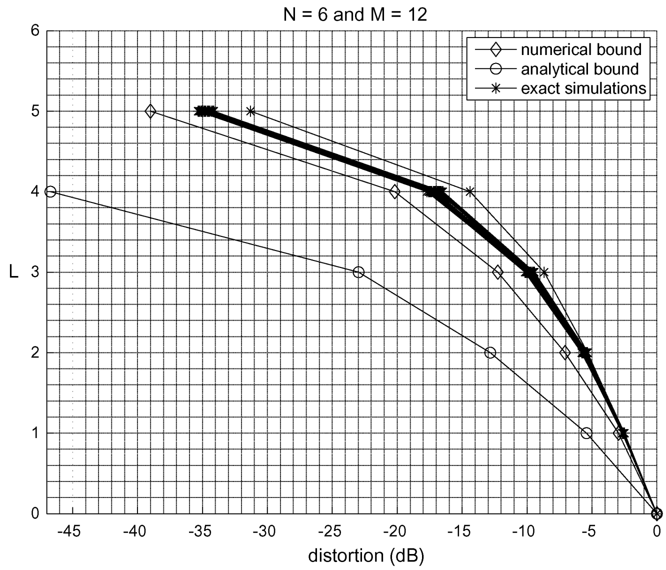

Fig. 1. Comparison of the lower bounds and optimal solutions for $N = 6$, $M = 12$ for 32 randomly generated Gaussian frames.

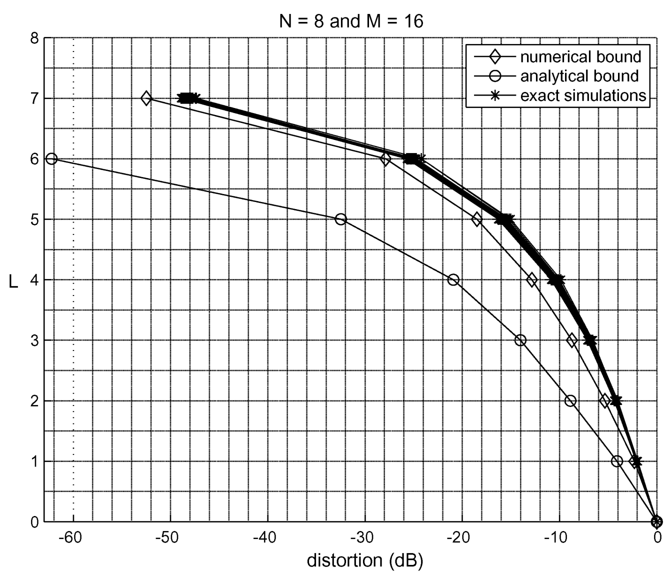

Fig. 2. Comparison of the lower bounds and optimal solutions for $N = 8$, $M = 16$ for 32 randomly generated Gaussian frames.

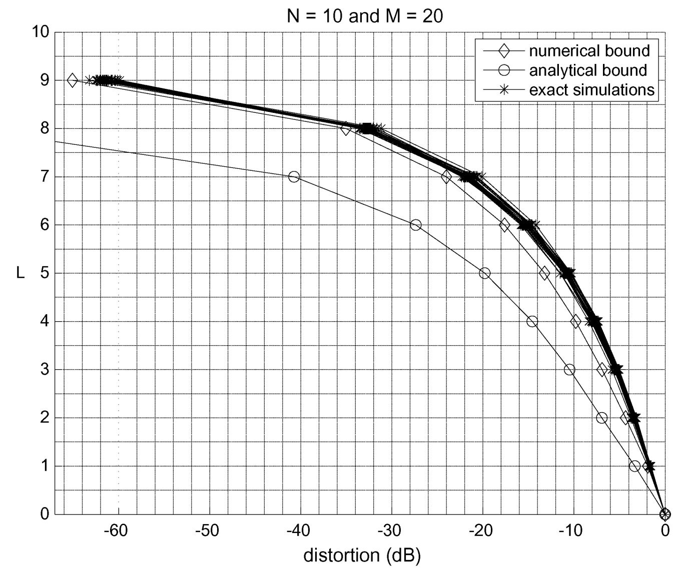

Fig. 3. Comparison of the lower bounds and optimal solutions for $N = 10$, $M = 20$ for 32 randomly generated Gaussian frames.

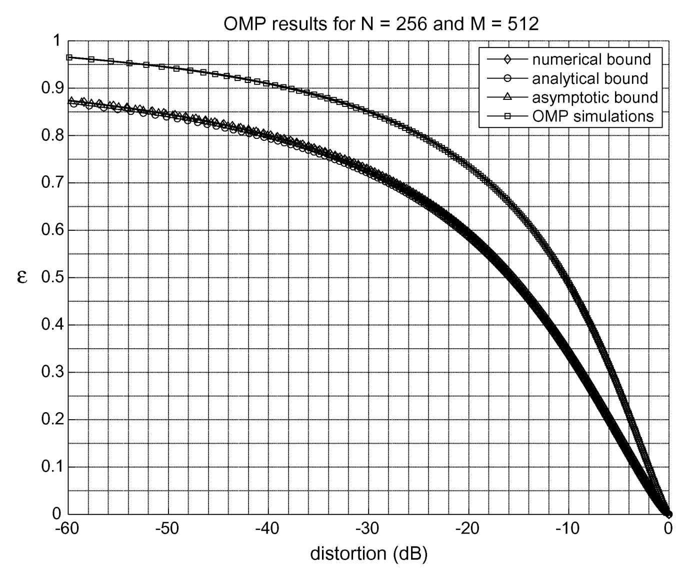

Fig. 4. Comparison of the lower bounds and OMP solutions for $N = 256$, $M = 512$ for 32 randomly generated Gaussian frames.

$M = 512$, we generate 32 randomly generated frames as above. Using Monte Carlo simulations, we compute the average distortion of these frames using the OMP algorithm. The results are presented in Fig. 4 and compared to the numerical bound of Theorem 2.1 and to the analytical and asymptotic bounds of [1]. The results indicate that for small values of ϵ, the performance is close to the theoretical lower bound. In comparison, for $\epsilon = 1/2$, the performance is about 5.1 dB from the numerical lower bound of Theorem 2.1, and distance to the lower bound grows as a function of ϵ.

It can be seen that even in moderate dimensions (such as $N = 256$), the numerical lower bound is very close to the analytical and asymptotic bounds of [1]. Additionally, the performance of the randomly generated frames hardens about their average, even when using the OMP algorithm.

IV. Conclusion

In this letter, we considered sparse representations of signals with at most L nonzero coefficients using a frame $\mathcal{F}$ of size M in $\mathbb{C}^N$. We established a universal lower bound (that can be numerically computed) on the average distortion of the representation as a function of the sparsity $\epsilon = L/N$ of the representation and redundancy $(r-1) = M/N - 1$ of $\mathcal{F}$.

We then compared the performance of *randomly generated frames* to this numerical lower bound and to the analytical and asymptotic bounds of [1]. It was shown that randomly generated frames in low dimensions ($N = 6$, 8, 10) perform close to the numerical lower bound, when the optimal sparse representation algorithm is used, and their performance hardens about their average as N grows. We then evaluated the performance of randomly generated frames using the greedy orthogonal matching pursuit algorithm for $N = 256$. The results indicate that for small values of ϵ, OMP performs close to the lower bound. However, the distance to the lower bound increases as a function of ϵ. Moreover, the performance of randomly generated frames hardens about their average, even when using the OMP algorithm. Finally, it was observed that the numerical bound presented in this letter is significantly stronger than the analytical lower bound of [1] but only in low dimensions. In contrast, it is less straightforward to compute.

REFERENCES

[1] M. Akçakaya and V. Tarokh, "Performance bounds on sparse representations using redundant frames," *IEEE Trans. Signal Process.* [Online]. Available: http://www.arxiv.org/abs/cs/0703045, submitted for publication

[2] E. J. Candès and P. A. Randall, Highly Robust Error Correction by Convex Programming (preprint).

[3] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[4] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[5] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York: Wiley, 1982.

[6] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.