TWO DIMENSIONAL COMPRESSIVE CLASSIFIER FOR SPARSE IMAGES

Armin Eftekhari, Hamid Abrishami Moghaddam — K.N. Toosi University of Technology, Tehran, Iran

Massoud Babaie-Zadeh — Sharif University of Technology, Tehran, Iran

Mohammad-Shahram Moin — Iran Telecom. Research Center, Tehran, Iran

ABSTRACT

The theory of compressive sampling involves making random linear projections of a signal. Provided signal is sparse in some basis, small number of such measurements preserves the information in the signal, with high probability. Following the success in signal reconstruction, compressive framework has recently proved useful in classification. In this paper, conventional random projection scheme is first extended to the image domain and the key notion of concentration of measure is studied. Findings are then employed to develop a 2D compressive classifier (2D-CC) for sparse images. Finally, theoretical results are validated within a realistic experimental framework.

Index Terms— Compressive sampling, random projections, retinal identification.

1. INTRODUCTION

The recently developed theory of compressive sampling (CS) involves making random linear projections of a signal by multiplying a random matrix. Provided the signal is sparse in some basis, few random projections preserve the information in the signal, with high probability [1]. This remarkable result is rooted in the concentration of measure phenomenon [1], which implies that the Euclidean length of a vector is uniformly "shrunk" under a variety of random projection matrices, with high probability. Due to tangible advantages, CS framework has found many promising applications in signal and image acquisition, compression, and medical image processing [1, 2, 4]. The CS community, however, has mainly focused on the signal reconstruction problem from random projections to date [3] and other applications of CS have yet remained unexplored. In particular, few recent studies showed that classification can be accurately accomplished using random projections, which suggests random projections as an effective and yet universal feature extraction tool. In this context, compressive detection, hypothesis testing, and manifold-aided classification of one-dimensional (1D) signals have been studied [1, 2]. However, as shown later in this paper, direct extension of these results to the image domain (2D) is computationally prohibitive, which strongly hinders the application of conventional compressive classifier (1D-CC) in real-world scenarios. To overcome this drawback, the idea of 2D compressive classification is developed in this paper. First, 2D random projection scheme is introduced in Section 2 and associated concentration properties are studied. It is then observed that using Gaussian random matrices, as the most common choice in 1D compressive framework, in the 2D random projection scheme, does not have feasible theoretical support. Then, by adding an assumption of so-called 2D sparsity (in some basis) to images, desirable concentration properties are proved for the same set of admissible random matrices as in 1D framework. This assumption is not restrictive, as most images become sparse under well-known transformations, like DCT, or have a sparse edge map. In Section 3, these findings are exploited to develop a 2D compressive classifier (2D-CC) for sparse images, along with derivation of error bound for an important special case. Finally, 2D-CC is applied to retinal identification within a realistic setting. It is observed that, at worst, 2D-CC provides significant saving on computational load and memory requirements compared to 1D-CC, at the cost of negligible loss in performance. This performance loss, however, can be avoided by "wise" selection of parameters. We note that, due to space limitations, some intermediate steps of the derivations have been omitted, and the interested reader is referred to [5] for details.

2. CONCENTRATION OF MEASURE FOR 2D RANDOM PROJECTION SCHEME

In order to linearly project a given image $X \in \mathbb{R}^{n_1 \times n_2}$ to a lower dimensional space, columns of X are stacked into a vector $x = \mathrm{vec}(X)$. This process ignores the intrinsic

This research was supported in part by Iran Telecommunication Research Center under Grant T-500-10530.

row/column-wise structure of the image and, even for moderately sized matrices, involves prohibitive computational load and memory requirements for generation and manipulation of the projection matrix [5, 6]. As a remedy to these drawbacks, one may use the so-called 2D projection scheme, i.e. $\mathrm{Y} = \mathrm{AXB}^{\mathrm{T}}$, in which $\mathrm{Y} \in \mathbb{R}^{m_1 \times m_2}$, $\mathrm{A} \in \mathbb{R}^{m_1 \times n_1}$, $\mathrm{B} \in \mathbb{R}^{m_2 \times n_2}$, with $m_1 < n_1$ and $m_2 < n_2$. This section is devoted to the study of 2D projection scheme in the compressive framework. Clearly, successful inference in this scenario depends on preservation of the structure of samples after projection [1]. For 1D signals, this has received extensive treatment. Particularly, it has been shown that, given a 1D signal, random matrices whose entries are i.i.d. random variables (rv's) with proper tail bounds, e.g. Gaussian rv, uniformly "shrink" the signal length after projection, with high probability. Using the union bound, one may then show that, with high probability, such random matrices preserve the structure of a set of samples after projection, by uniformly shrinking the pair-wise Euclidean distances. Focusing on Gaussian random matrices, as the most common choice in CS framework, a similar strategy is developed here. Consider random Gaussian matrices $\mathrm{A} \in \mathbb{R}^{m_1 \times n_1}$ and $\mathrm{B} \in \mathbb{R}^{m_2 \times n_2}$, whose entries are i.i.d. $\mathcal{N}(0,1/n_1)$ and $\mathcal{N}(0,1/n_2)$ rv's, respectively. Given $\mathrm{X} \in \mathbb{R}^{n_1 \times n_2}$, let us define $\mathrm{x} = \mathrm{vec}(\mathrm{X})$ and $\mathrm{D} = \mathrm{B} \otimes \mathrm{A}$, where $\otimes$ denotes the Kronecker product. Thus, $\mathrm{AXB}^{\mathrm{T}} = \mathrm{Dx}$, where each entry of D has a Bessel-like distribution with zero mean and variance $1/n_1 n_2$ [5]. We are interested in studying the concentration of rv $\|\mathrm{Dx}\|^2$ about its expected value $(m_1 m_2 / n_1 n_2)\|\mathrm{x}\|^2$ by finding an exponentially-fast decreasing bound on $\Pr[|\|\mathrm{Dx}\|^2 - (m_1 m_2/n_1 n_2)\|\mathrm{x}\|^2| \geq (m_1 m_2/n_1 n_2)\epsilon]$, where probability is taken over all matrices A and B. Note that (in contrast to 1D counterpart), entries of D are no more independently distributed. Furthermore, each entry of D is a product of two i.i.d. Gaussian rv's. Properties of the Kronecker product implies that each row of D is dependent with exactly $m_1 + m_2 - 2$ other rows, and that we can partition the rows of D into m nonoverlapping partitions $\{R_i\}_{i=1}^{m}$ with $|R_i| = m_2$, such that rows in each partition are independent. Let us denote by D_{R_i} the $m_1 \times n_1 n_2$ submatrix obtained by retaining the rows of D corresponding to the indices in R_i. Clearly, rows of D_{R_i} are independent, and we have $\|\mathrm{Dx}\|^2 = \sum_{i=1}^{q} \|\mathrm{D}_{R_i}\mathrm{x}\|^2$, where entries of D_{R_i} are zero-mean rv's with variance $1/n_1 n_2$. It is then straightforward to show that strong concentration of D_{R_i} for $i = 1, \ldots, q$, implies that of D [5]. This follows easily from the following fact [5]:

$$\Pr\left[\left|\|\mathrm{Dx}\|^2 - \frac{m_1 m_2}{n_1 n_2}\|\mathrm{x}\|^2\right| \leq \frac{m_1 m_2}{n_1 n_2}\epsilon\right] \leq \sum_{i=1}^{q} \Pr\left[\left|\left\|\mathrm{D}_{R_i}\mathrm{x}\right\|^2 - \frac{|R_i|}{n_1 n_2}\|\mathrm{x}\|^2\right| \leq \frac{|R_i|}{n_1 n_2}\epsilon\right] \tag{1}$$

For ease of notation, let $\mathrm{M} = \mathrm{D}_{R_i}$ and $R = |R_i|$. Now, it suffices to find an exponentially-fast decreasing bound for $\Pr\{|\|\mathrm{Mx}\|^2 - R/n_1| \geq R\epsilon/n_1 n_2\}$, where, due to linearity, each column of X is assumed to have unit length. Consider the probability of $\|\mathrm{Mx}\|^2 - R/n_1 \geq R\epsilon/n_1 n_2$. Then, for any $h > 0$, invoking the Chernoff bounding technique and using the inherent symmetries in M [5] gives:

$$\Pr\left[\|\mathrm{Mx}\|^2 - \frac{R}{n_1} \geq \frac{R}{n_1 n_2}\epsilon\right] \leq \mathbb{E}\left[e^{h\left(\sum_1^{n_1 n_2} M_{1,j} x_j\right)^2}\right]^R e^{-\frac{R}{n_1 n_2}h(n_2+\epsilon)} \tag{2}$$

where, without loss of generality, first row of M is used in (2). Further simplifications and exploiting the moments of Gaussian and Chi-square distributions [5] leads to:

$$\mathbb{E}\left[e^{h\left(\sum_1^{n_1 n_2} M_{1,j} x_j\right)^2}\right] \leq \sum_{k=0}^{\infty} \left(\frac{2h}{n_1 n_2}\right)^k \frac{2k!}{(k!)^2} \frac{\Gamma\left(k + \frac{n_1}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)} \tag{3}$$

in which $\Gamma(\cdot)$ denotes the Gamma function. The sum on the right hand side of (3) apparently fails to converge. Consequently, the probability of $\|\mathrm{Mx}\|^2 - R/n_1 \geq R\epsilon/n_1 n_2$ fails to decrease exponentially fast, in general; not letting us to find a solid theoretical base for 2D random projection with random Gaussian matrices. In the following, we prefer to insert an additional constraint on X, which would allow for successful operation of 2D random projection scheme with the same set of admissible random matrices as in 1D case. Let Σ_k denote the set of all signals of length n with at most k nonzero entries. We say that $\mathrm{A} \in \mathbb{R}^{m \times n}$ satisfies restricted isometry property (RIP) of order k, if for every $\mathrm{v} \in \Sigma_k$, the following holds for some $\epsilon_k \in [0,1)$.

$$(1 - \epsilon_k)\frac{m}{n}\|\mathrm{v}\|^2 \leq \|\mathrm{Av}\|^2 \leq \|\mathrm{v}\|^2 \frac{m}{n}(1 + \epsilon_k) \tag{4}$$

It is shown that many random matrices satisfy RIP condition with high probability [1, 4]. For instance, $m \times n$ matrices whose entries are i.i.d. $\mathcal{N}(0,1/n)$ rv's satisfy RIP of order k with probability exceeding $1 - \mathrm{e}^{-c_2(\epsilon_k)m}$, provided $k \leq c_1(\epsilon_k) m \log n/k$, where $c_1(\epsilon_k)$ and $c_2(\epsilon_k)$

depend only on ϵ_k [1, 4]. Also random orthoprojectors, i.e. matrices with random orthonormal rows, satisfy the RIP condition similarly. Extending these ideas to the image domain, let Σ_{k_r,k_c} denote the set of all images $\mathrm{V} \in \mathbb{R}^{n_1 \times n_2}$, whose nonzero entries are distributed in at most k_r rows and k_c columns. A matrix with this property will be called 2D sparse matrix. Now, we have the following observation [5].

Observation 1. Suppose projection matrices $\mathrm{A} \in \mathbb{R}^{m_1 \times n_1}$ and $\mathrm{B} \in \mathbb{R}^{m_2 \times n_2}$, respectively, satisfy the RIP conditions of orders $2k_r$ and $2k_c$, for some $\epsilon_{2k_r}, \epsilon_{2k_c} \in [0,1)$. Then, for any $\mathrm{X} \in \Sigma_{2k_r,2k_c}$, we have:

$$
\begin{aligned}
\left(1-\epsilon_{2k_r}\right)\left(1-\epsilon_{2k_c}\right)\frac{m_1 m_2}{n_1 n_2}\|\mathrm{X}\|^2 &\le \|\mathrm{AXB}^{\mathrm{T}}\|^2 \\
&\le \|\mathrm{X}\|^2 \frac{m_1 m_2}{n_1 n_2}\left(1+\epsilon_{2k_r}\right)\left(1+\epsilon_{2k_c}\right)
\end{aligned} \tag{5}
$$

In particular, if A and B are admissible random matrices in the 1D compressive framework, (5) holds with probability exceeding $1-e^{-c_2(\epsilon_{2k_r})m_1 - c_2(\epsilon_{2k_c})m_2}$, provided $2k_r \le c_1(\epsilon_{2k_r}) m_1 \log n_1 / 2k_r$ and $2k_c \le c_1(\epsilon_{2k_c})$ $m_2 \log n_2 / 2k_c$. This means that, if $\mathrm{X} \in \Sigma_{2k_r,2k_c}$, then $\|\mathrm{AXB}^{\mathrm{T}}\|^2 / \|\mathrm{X}\|^2$ is strongly concentrated about its expected value for variety of random matrices; hence establishing concentration of measure in 2D random projection scheme.

3. 2D COMPRESSIVE CLASSIFIER FOR SPARSE IMAGES

In many applications, images are sparse in some pixel domain or have a sparse representation in some basis, such that their nonzero entries are concentrated in a small number of rows/columns [5]. For this class of images, the results obtained in the previous section are applicable; enabling us to develop the proposed 2D-CC. Formally, assume that $\mathcal{X} = \{\mathrm{X}_1, \dots, \mathrm{X}_L\}$ denotes a set of $n_1 \times n_2$ known 2D sparse images, i.e. $\mathrm{X}_l \in \Sigma_{k_r,k_c}$, $l = 1, \dots, L$, for some integers $k_r < n_1$, $k_c < n_2$. The (possibly noisy) "true" image $\mathrm{X}_T \in \mathcal{X}$ undergoes 2D random projection to obtain $\mathrm{Y} = \mathrm{A}(\mathrm{X}_T + \mathrm{N})\mathrm{B}^{\mathrm{T}}$, where $\mathrm{N} \in \mathbb{R}^{n_1 \times n_2}$ represents the noise. Now, we will be concerned with discrimination among the members of $\mathcal{X}$, given only low-dimensional random projections. Given A and B, failure will be quantified in terms of expected error. Let $\mathrm{y} = \mathrm{vec}(\mathrm{Y})$, $\mathrm{x} = \mathrm{vec}(\mathrm{X})$, $\mathrm{n} = \mathrm{vec}(\mathrm{N})$ and $\mathrm{D} = \mathrm{B} \otimes \mathrm{A}$. For simplicity of analysis, we further assume $\mathrm{n} \sim \mathcal{N}\left(0, \sigma^2 \mathrm{I}_{n_1 n_2}\right)$, where I_a denotes the $a \times a$ identity matrix. Also, A and B are selected to be random orthoprojectors, which implies that the distribution of noise remains unchanged under projection. Now, provided X_l's happen equally likely, the Bayes decision rule and the associated expected error would be:

$$
\hat{\mathrm{x}}_l = \underset{\mathrm{x}_l \in \mathrm{vec}(\mathcal{X})}{\operatorname{argmin}} \|\mathrm{y} - \mathrm{Dx}_l\| = \underset{\mathrm{X}_l \in \mathcal{X}}{\operatorname{argmin}} \|\mathrm{Y} - \mathrm{AX}_l\mathrm{B}^{\mathrm{T}}\| \tag{6}
$$

$$
\mathrm{Err}(\mathrm{A},\mathrm{B}) = 1 - \frac{1}{L}\int_{\mathrm{y}} \max_{l}\{p_l(\mathrm{y})\}\, dy \tag{7}
$$

where $p_l(\mathrm{y}) = \mathcal{N}\left(\mathrm{Dx}_l, \sigma^2 \mathrm{I}_{n_1 n_2}\right)$. Note that, for any nonnegative sets $\{a_l\}$ and $\{s_l\}$ with $\sum_1^L s_l = 1$, we have $\max_l\{a_l\} \ge \prod_{l=1}^{L} a_l^{s_l}$. Consequently, we may write:

$$
\mathrm{Err}(\mathrm{A},\mathrm{B}) \le 1 - \frac{1}{L}\int_{\mathrm{y}} \prod_{l=1}^{L} \left(p_l(\mathrm{y})\right)^{s_l} d\mathrm{y} \tag{8}
$$

Further simplification and setting $s_l = 1/L$, gives the following bound:

$$
\mathrm{Err}(\mathrm{A},\mathrm{B}) \le 1 - \exp\left(-\frac{\sum_{l \ne l'} \|\mathrm{D}(\mathrm{x}_l - \mathrm{x}_{l'})\|_2^2}{4\sigma^2 L^2}\right) \tag{9}
$$

Assuming that A and B satisfy the RIP conditions similar to Observation 1, and defining $d_{min} \triangleq \min_{l \ne l'} \|\mathrm{x}_l - \mathrm{x}_{l'}\|^2$, Observation 1 implies the following bound for classification error [5]:

$$
1 - \exp\left(-\left(1+\epsilon_{2k_r}\right)\left(1+\epsilon_{2k_c}\right)\frac{m_1 m_2}{n_1 n_2}\frac{(L-1)}{8\sigma^2 L} d_{min}\right) \tag{10}
$$

Particularly, if A and B are random orthoprojectors, then above bound holds with conditions noted right after Observation 1. It is observed that the classification error decays exponentially fast as the number of observations $m_1 m_2$ increases.

4. EXPERIMENTS

In this section, the efficacy of the proposed 2D-CC is examined in retinal identification problem. Retinal biometrics refers to identity verification of individuals based on their retinal vessel tree pattern (Fig. 1.a). Our experiments are conducted on VARIA database containing 153 (multiple) retinal images of 59 individuals. To compensate for the variations in the location of optic disc (OD) in retinal images, a ring-shaped region of interest (ROI) in the vicinity of OD is used to construct the feature matrix (Fig. 1.b). To extract the ROI, using the technique presented in [6], OD and vessel tree are extracted. Then, a ring-shaped mask with proper radii centered at OD is used to form the feature matrix $\mathrm{X} \in \mathbb{R}^{n_1 \times n_2}$ by collecting the pixels along $n_2 = 200$ beams of length $n_1 = 100$ originating from OD (Fig. 1.c). Note that feature matrices readily satisfy the requirements of Observation 1 in pixel domain [5]. Due to small number of images per subject (~2) and approximate invariance of feature matrix to the location of OD, feature matrices of the same subject are modeled as noisy deviations from corresponding mean feature matrix. Hence, once all images are processed, $\mathcal{X} = \{\mathrm{X}_1, \ldots, \mathrm{X}_L\}$ is formed, where each X_l is the mean feature matrix of lth subject. Dimension reduction and classification of a test feature matrix is then performed in $\mathbb{R}^{m_1 \times m_2}$ by the proposed 2D-CC, and in $\mathbb{R}^{m_1 m_2}$ by 1D-CC. Error is measured using the leave-one-out scheme and the average results over 100 independent repetitions are depicted in Fig. 2 for a wide range of m_1 and m_2. Although explicit calculation of the bound in (10) is intractable [1], we note that the exponential nature of error is in accordance with our findings. Also, due to highly redundant nature of feature matrices along their columns, "wise" choices for m_1 and m_2, which consider this redundancy, exhibit good performance especially for small values of m_1 and m_2. In contrast, "careless" choices for m_1 and m_2, degrade the performance (Fig. 3). Using an Intel Core 2 Duo, 2.67 GHz processor with 3.24 GB of memory, we found that each repetition of 2D-CC approximately took $0.0098\, m_1 m_2$ seconds, whereas this number was roughly $0.655 m_1 m_2$ for 1D-CC, in MATLAB7 environment. Note that this difference is significant even with our small-sized feature matrices. In sum, for typical choices of m_1 and m_2, 2D-CC runs much faster than 1D-CC, yet producing results with negligible loss in performance. This loss, however, disappears with proper choice of m_1 and m_2 which takes the prior knowledge into account. In addition, 2D-CC enjoys significantly less memory requirements.

5. CONCLUSIONS

In this paper, the idea of random projections is extended to image domain and associated concentration properties are studied. Our findings are then used to develop 2D-CC, along with error bound for an important special case. Finally, results are validated in a realistic application.

6. REFERENCES

[1] M.A. Davenport, M.F. Duarte, M.B. Wakin, J.N. Laska, "The smashed filter for compressive classification and target recognition," Proc. SPIE Computational Imaging V, 2007.

[2] J. Haupt, R. Castro, R. Nowak, G. Fudge, A. Yeh, "Compressive sampling for signal classification," Fortieth Asilomar Conf. on Signals, Systems and Computers, 2006.

[3] D.L. Donoho, M. Elad, V.N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," IEEE Trans. on Info. Theory, vol. 52, pp. 6-18, 2006.

[4] J. Haupt, W.U. Bajwa, G. Raz, "Toeplitz compressed sensing matrices with applications to channel sensing," Submitted, 2008.

[5] A. Eftekhari, H. Abrishami Moghaddam, and M. Babaie-Zadeh, "2D compressive classifier for sparse images," Technical report,http://www.ee.kntu.ac.ir/pow/detlbiog.asp?biouser=moghadam&bioglastn=Abrishami%20Moghaddam

[6] A. Ghaffari, M. Babaie-Zadeh, C. Jutten, "Sparse decomposition of two dimensional signals," in Proceedings of ICASSP2009, pp. 3157-3160, 2009.

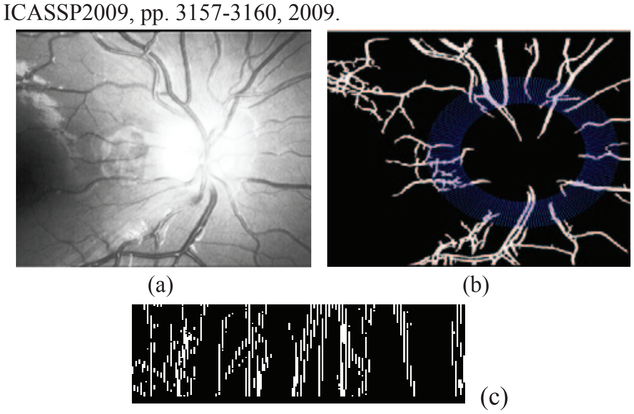

Fig. 1. (a) Retinal image; bright area is OD. (b) Vessel tree (in white) and mask (in blue). (c) Feature matrix for $n_1 = 100$, $n_2 = 300$. Due to limited space, images (a) and (b) are cropped.

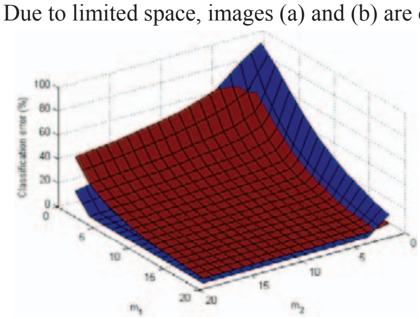

Fig. 2. Average classification error of 2D-CC (red surface) and 1D-CC (blue surface) for a wide range of m_1 and m_2.

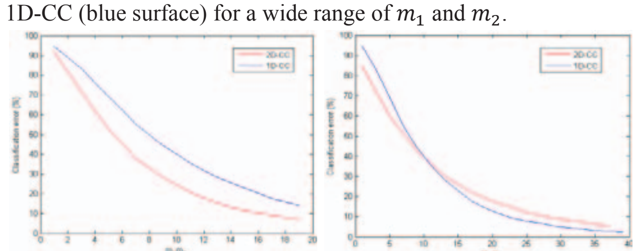

Fig. 3. Two examples of "wise" choices which consider the redundancy along columns: $m_1 = 1$ (left) and $m_1 = 3$ (right)