Signal Processing ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Two-dimensional random projection ☆

Armin Eftekhari [a,c,*], Massoud Babaie-Zadeh [b], Hamid Abrishami Moghaddam [c]

[a] Division of Engineering, Colorado School of Mines, USA
[b] Department of Electrical Engineering, Sharif University of Technology, Iran
[c] Department of Electrical Engineering, K.N. Toosi University of Technology, Iran

ARTICLE INFO

Article history:
Received 18 January 2010
Received in revised form
22 December 2010
Accepted 3 January 2011

Keywords:
Random projection
Concentration of measure
Sparse signal reconstruction

ABSTRACT

As an alternative to adaptive nonlinear schemes for dimensionality reduction, linear random projection has recently proved to be a reliable means for high-dimensional data processing. Widespread application of conventional random projection in the context of image analysis is, however, mainly impeded by excessive computational and memory requirements. In this paper, a two-dimensional random projection scheme is considered as a remedy to this problem, and the associated key notion of concentration of measure is closely studied. It is then applied in the contexts of image classification and sparse image reconstruction. Finally, theoretical results are validated within a comprehensive set of experiments with synthetic and real images.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

The need for efficient collection, storage and processing of large, high-dimensional data has increased drastically over the past decade. Unfortunately, the high-dimensionality of data, in particular, jeopardizes the performance of inference tasks, due to the so-called "curse of dimensionality" phenomenon [1]. Luckily, dimensionality reduction techniques are often helpful in reducing this burden by extracting key low-dimensional information about the original high-dimensional signals, from which we can later infer key properties of the original data. It is therefore desirable to formulate a method that efficiently reduces the dimensionality efficiently, while preserving as much information from the original data as possible [2]. There are two main scenarios in which dimensionality reduction is successful: (1) Low-complexity inference, where only a small amount of information is required to make an inference about data. Examples include function estimation, signal detection, and classification [3,4]. (2) Low-dimensional signal models, in which signals of interest have few degrees of freedom. In fact, it frequently happens in real-world applications that high-dimensional data actually obey some sort of concise low-dimensional model. Examples include signals with finite rate of innovation, manifolds, etc [5,6]. While most conventional dimensionality reduction techniques are adaptive and involve nonlinear mappings to preserve certain desirable properties of data, a linear non-adaptive technique based on random projections (RP's) of data has recently been introduced [7]. In fact, random projections have been successfully utilized in low-complexity inference tasks, such as classification and estimation [3,4,8,9]. RP has also demonstrated remarkable performance in obtaining a faithful low-dimensional representation of data belonging to low-complexity signal models, as in acquisition and reconstruction of sparse signals and manifolds [10,11,2]. Remarkable properties of RP stem from a simple concentration of measure inequality which states that, with high probability, the norm of a signal is well-preserved under a random dimensionality-reducing projection [12]. This seminal fact allows us to show that in many settings the

☆ This work has been partially funded by the Iran Telecom Research Center (ITRC) and the Iran National Science Foundation (INSF).

* Corresponding author.
E-mail address: aeftekha@mines.edu (A. Eftekhari).

0165-1684/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2011.01.002

distinguishing characteristics of a signal can be encoded by a few random measurements. In particular, using the simple union bound in combination with the above result leads us to Johnson–Lindenstrauss (JL) Lemma, which implies that the geometric structure of a point cloud is preserved under a random dimensionality reduction projection [13].As shown in [14], these results can be further extended to infinite sets with low-complexity geometrical structure, such as sparse signals and manifolds. Despite these impressive results, application of conventional RP to high-dimensional data, such as images and videos faces severe computational and memory difficulties, due to the so-called vector space model [15–17]. Under this model, each datum is modeled as a vector, i.e. columns (or rows) of each two-dimensional signal (2D-signal) are initially stacked into a large vector, as a result of which the row/column-wise structure of the image is ignored and storage and computational requirements are drastically increased. To alleviate the expensive conventional RP (1D-RP) scheme, the so-called two-dimensional random projection (2D-RP) has been recently proposed, which directly leverages the matrix structure of images and represents each datum as a matrix, instead of a vector [15]. In fact, similar ideas have previously appeared, for instance, in the context of 2D principal component analysis (2D-PCA) [18] and 2D linear discriminant analysis (2D-LDA) [19], in which the extensions of conventional PCA and LDA on 1D-signals to the image domain have demonstrated substantial improvements in memory and computational efficiency. In this paper, the idea of 2D-RP is studied and the corresponding concentration properties are closely analyzed. It is observed that desirable properties of 1D-RP extends to 2D analogue, while significantly gaining in computational and storage requirements. This gain, essentially due to the reduction in the number of degrees of freedom of the projection matrices, comes at the cost of extra measurements to obtain the same accuracy. 2D-RP is then applied to two important applications: (1) 2D-compressive classification, which is concerned with classification of images based on random measurements provided by 2D-RP. In particular, we consider multiple hypothesis testing given only random measurements of possibly noisy images, and (2) sparse 2D-signal reconstruction, which addresses the problem of accurate acquisition and reconstruction of sparse images from relatively few random measurements. In accordance with our expectations, comprehensive experiments verify the comparable performance and remarkable computational and storage advantages of 2D-RP compared to the 1D counterpart. Preliminary steps towards this work have been presented in ICIP2009 [20], in which the application of 2D-RP to classification of sparse images was studied briefly, along with a study of 2D-RP with Gaussian random matrices.

The rest of this paper is organized as follows. Section 2 offers a brief review on 1D-RP and corresponding technical results. 2D-RP and its implications for 2D-signals, finite sets, and infinite sets with low-complexity signal models are discussed in Section 3. Section 4 presents two main applications of 2D-RP and offers detailed performance analysis. In Section 5, these findings are validated through comprehensive experiments with synthetic and real images.

2. 1D random projection

Consider making m linear measurements of 1D-signals in $\mathbb{R}^n$, $m<n$. Equivalently, we can represent this measurement process in terms of linear projection onto $\mathbb{R}^m$ by an $m\times n$ matrix A. Successful statistical inference or stable recovery in $\mathbb{R}^m$ then mostly depends on the preservation of the geometric structure of data after projection [21]. This, in turn, requires a stable embedding of data in $\mathbb{R}^m$, which is commonly characterized using the following notion of isometry [10,14].

Definition 1 (*Baraniuk and Wakin [10, Section 3.2.1]*). Given $x\in\mathbb{R}^n$, a matrix $A\in\mathbb{R}^{m\times n}$ is said to have isometry constant ε for x, if the following holds[1]:

$$\sqrt{\frac{m}{n}}(1-\varepsilon)\|x\|_2\le\|Ax\|_2\le\sqrt{\frac{m}{n}}(1+\varepsilon)\|x\|_2, \tag{1}$$

in which $\|\cdot\|_2$ denotes the ℓ_2-norm.

Also, we say that an $m\times n$ random matrix is admissible if its entries are independently drawn from a zero-mean sub-Gaussian[2] probability distribution with variance $1/n$. Examples include random Gaussian and Bernoulli matrices, as well as orthoprojectors.[3] The well-known concentration of measure inequality then implies that, with high probability, (1) holds for all admissible random (AR) matrices [14,22]. This is formally stated as follows.

Theorem 1 (*Baraniuk and Wakin, Baraniuk et al. [10,14]*). *Suppose that $\varepsilon\in(0,1)$ and $x\in\mathbb{R}^n$ are given. Then, there exists a positive constant c depending only on ε, such that an AR matrix $A\in\mathbb{R}^{m\times n}$ has the isometry constant ε for x, with probability exceeding $1-e^{-cm}$.*

In addition, c is shown to be $\varepsilon^2/400\,v^2$, where $v>0$ is the Gaussian standard of the distribution of the entries of A. These arguments about random projection of 1D-signals (1D-RP) easily extend to any finite set of signals. In particular, we say that a matrix $A\in\mathbb{R}^{m\times n}$ has isometry constant ε on a set $\{x_i\}_{i=1}^N\subset\mathbb{R}^n$, if (1) holds for every point in the set [10, Section 3.2.1]. Using a simple union bound in combination with the above results, it is straightforward to show that, AR matrices have desired isometry constant on an arbitrary finite set with high probability, provided that sufficient number of measurements are acquired. This result is formally stated in terms of the JL Lemma, and is concerned with stable embedding of a finite set of points under a random dimensionality-reducing projection. JL Lemma implies that with high probability the geometry of a point cloud is preserved

[1] Similar definitions may differ in scaling.

[2] For a sub-Gaussian random variable y we have $\Pr\{|y|>u\}\le Ke^{-\delta u^2}$ for every u and some $K,\delta>0$. Equivalently, a sub-Gaussian random variable satisfies $\mathbb{E}e^{uy}\le e^{vu^2}$ for every $u\in\mathbb{R}$ and some $v>0$, which we refer to the infimum of such v as the Gaussian standard of u.

[3] By an orthoprojector, we mean an orthogonal projection from $\mathbb{R}^n$ to $\mathbb{R}^m$, $m\le n$, that can be expressed as an $m\times n$ matrix with orthonormal rows.

Please cite this article as: A. Eftekhari, et al., Two-dimensional random projection, Signal Process. (2011), doi:10.1016/j.sigpro.2011.01.002

by random linear projection onto a space with dimension that only logarithmically grows in the number of points. In particular, the pair-wise distances are uniformly shrunk by a factor of $\sqrt{m/n}$ [13,14].

These results can be further extended to infinite sets with low-complexity geometric structures, such as sparse (or nearly sparse) signals and manifolds. Let Σ_k denote the set of signals in $\mathbb{R}^n$ with at most k nonzero entries. With careful application of JL Lemma and simple covering arguments, it has been shown that linear random projection stably embeds Σ_k into the lower-dimensional space $\mathbb{R}^m$ with high probability, provided that the number of measurements m is linear in k and logarithmic in n [14]. This result is formally stated below.

Theorem 2 (*Baraniuk et al. [14]*). *Given $\varepsilon \in (0,1)$, there exist constants $c_1, c_2 > 0$ depending on ε, such that an AR matrix $A \in \mathbb{R}^{m\times n}$ has the isometry constant ε for Σ_k with probability exceeding $1-e^{-c_2 m}$, provided $k \le c_1 m/\log n/k$.*

Theorem 2 implies that if the signal is sparse (or nearly sparse) in some basis, then linear random projection encodes the salient information in the signal with high probability, and enables signal reconstruction within a controllable mean-squared error, even when the observations are corrupted by additive noise [23,24]. Several tractable algorithms, such as basis pursuit [25,26], matching pursuit [27–29], and smoothed ℓ_0-norm algorithm (SL0) [30], have been proposed for efficient sparse signal reconstruction based on such non-adaptive linear measurements.

3. 2D random projection

Traditionally, to collect a set of linear measurements of a 2D-signal (image), columns of the 2D-signal are first stacked into a large column vector. This so-called vector space model for signal processing [16], however, ignores the intrinsic row/column-wise structure of the 2D-signal and, even for moderately sized signals, involves prohibitive computational and memory requirements for collecting linear measurements and for applying statistical inference and reconstruction algorithms after projection. More specifically, to linearly project an $n \times n$ image X onto $\mathbb{R}^{m^2}$ $(m < n)$, 1D-RP produces $y \triangleq Ax$, in which $A \in \mathbb{R}^{m^2\times n^2}$ is an AR matrix and $x = vec(X)$ is the $n^2 \times 1$ vector obtained by stacking the columns of X. This projection requires $\mathcal{O}(m^2n^2)$ operations and m^2n^2 memory units to store A. Therefore, direct application of 1D-RP to high-dimensional data, such as images and videos, quickly reaches practical computational limits.

As a remedy to these drawbacks, one may use the so-called two-dimensional random projection (2D-RP) to directly leverage the matrix structure of images. 2D-RP of $X \in \mathbb{R}^{n\times n}$ onto $\mathbb{R}^{m\times m}$ produces $Y \triangleq AXB^T$, where A,B are $m \times n$ AR matrices. This can be equivalently shown by $y = (B \otimes A)x$, where $y = vec(Y)$, $x = vec(X)$, and $\otimes$ denotes the Kronecker product [31]. This projection, in contrast to 1D-RP, requires only $\mathcal{O}(mn^2)$ operations and $2mn$ memory units to store the projection matrices. Despite the experimentally verified effectiveness of 2D-RP in the context of sparse images reconstruction [15], theoretical aspects and other applications of this method have mainly remained unexplored. Therefore, as our first result, Theorem 3 focuses on the concentration properties of the Kronecker product of two AR matrices. Note, however, that this is not trivial, since entries of the product are no more independently distributed. Proof of this result is given in Appendix A.

Theorem 3. *Suppose that $\varepsilon \in (0,1)$ and $X \in \mathbb{R}^{n\times n}$ are given. Then, there exists $c = c(\varepsilon) > 0$ depending only on ε, such that with probability exceeding $1-e^{-cm}$, $B \otimes A$ has isometry constant ε for X, where we assume that the entries of the Kronecker product of AR matrices $A,B \in \mathbb{R}^{m\times n}$ are sub-Gaussian random variables.*

In particular, due to the heavy tail of the product of two Gaussian random variables, the Kronecker product of two random Gaussian matrices is not guaranteed to satisfy the concentration inequality (1). Note, however, that the concentration inequality holds for the Kronecker product of matrices with entries of the form $\sqrt{|y|}$, where y is drawn from $\mathcal{N}(0,\pi/2n^2)$, i.e. a Gaussian distribution with zero mean and variance of $\pi/2n^2$. Furthermore, we observe that the concentration inequality is satisfied by the Kronecker product of any two AR matrices with entries drawn from finitely supported probability distributions.

Also, it would be instructional to compare Theorem 3 (which studies the Kronecker product of AR matrices) with that of a single AR matrix. According to Theorem 1, an $m^2 \times n^2$ AR matrix D has isometry constant ε with probability exceeding $1-e^{-cm^2}$, for some $c > 0$. On the other hand, Theorem 3 states that, the Kronecker product of $m \times n$ AR matrices A and B achieves an isometry constant of ε with probability at least $1-e^{-c'm}$, for some $c' > 0$ and provided that a certain condition on probability distribution is met. Therefore, compared to 1D-RP, 2D-RP requires more measurements to ensure the same isometry constant. The factor of $\mathcal{O}(m)$ increase in the required number of observations may be attributed to the reduction in the number of degrees of freedom from $\mathcal{O}(m^2n^2)$ to $\mathcal{O}(mn)$. However, it shall be emphasized that Theorem 3 only states that the sufficient conditions for concentration of measure. In practice, while saving considerably in memory resources and computation time, performance of 2D-RP proves to be comparable to that of 1D-RP most of the times. Before extending the above results to infinite sets with low-complexity structures, it shall be emphasized that Theorem 3 is easily extended to any arbitrary finite set of 2D-signals, $X = \{X_1, \ldots, X_L\} \subset \mathbb{R}^{n\times n}$. Using the union bound, it is straightforward to verify that there exists constants $c_1, c_2 > 0$ depending only on ε, such that $B \otimes A$ has an isometry constant ε for X with probability exceeding $1-e^{-c_2 m}$, provided $m \ge c_1 \ln L$.

Now we extend these results to a well-known example of infinite sets with low-complexity geometric structure, namely sparse 2D-signals, where the signals of interest have few degrees of freedom relative to the dimension of the ambient space. Building upon the ideas presented in [15], we consider the following three definitions for

Please cite this article as: A. Eftekhari, et al., Two-dimensional random projection, Signal Process. (2011), doi:10.1016/j.sigpro.2011.01.002

sparse 2D-signals in $\mathbb{R}^{n\times n}$. Again, since extension of the consequent results to the general case is straightforward, only the symmetric case (i.e. where both A and B are of the same dimension $m\times n$) is assumed for the sake of neatness.

Definition 2. We define the following notations:

Let Σ_k^1 to be the set of $n\times n$ 2D-signals whose nonzero entries are distributed in at most k rows and k columns.

Let Σ_k^2 to be the set of $n\times n$ 2D-signals with no more than k^2 nonzero entries, where the number of nonzero entries in each row and column does not exceed k.

Let Σ_k^3 to be the set of $n\times n$ 2D-signals with no more than arbitrary-distributed k^2 nonzero entries.

Note that $\Sigma_k^1\subset\Sigma_k^2\subset\Sigma_k^3$, and that Σ_k^3 is the direct extension of the concept of sparsity from 1D case and therefore neglects the row/column-wise structure of the 2D-signal. In contrast, Σ_k^1 and Σ_k^2 assume a row/column-wise structure on 2D-signals, which, as shown later, usually allows for better concentration properties. Theorem 4, proved in Appendix B, extends Theorem 2 to the set of sparse 2D-signals.

Theorem 4. *Given $\varepsilon\in(0,1)$, there exist constants $c_{2,1}>c_{2,2}>c_{2,3}>0$ and c_1 depending only on ε, such that with probability exceeding $1-e^{-c_{2,i}m}$, $B\otimes A$ has the isometry constant ε for Σ_k^i, $i\in\{1,2,3\}$, where we assume that the Kronecker product of $m\times n$ AR matrices A and B is sub-Gaussian, and $k\le c_1\sqrt{m}/\log n/k$.*

According to Theorem 4, $B\otimes A$ satisfies stronger concentration inequalities for Σ_k^1 and Σ_k^2, which assume a column/row-wise structure on sparse 2D-signals, compared to Σ_k^3 which is merely the extension of 1D case. Also note that, Theorem 2 states that, with high probability, an $m^2\times n^2$ AR matrix D has the isometry constant ε for Σ_k^3, provided that $m\ge ck\sqrt{\log n^2/k^2}$ for some $c>0$. On the other hand, Theorem 4 states that, with high probability, the Kronecker product of $m\times n$ AR matrices A and B achieves an isometry constant ε for Σ_k^3, provided that $m\ge c'k^2\log^2 n/k$ for some $c'>0$. Again, 2D-RP witnesses an increase in the required number of random measurements compared to 1D-RP.

Finally, Table 1 compares 1D-RP and 2D-RP for 2D-signals. By its definition, 1D-RP is indifferent to the row/column structure of 2D-signals. Thus, in order to have a meaningful comparison, we have only included Σ_k^3 in this table.

Table 1
Comparison of 1D and 2D random projection schemes.

	1D–RP	2D–RP
# of operations to get m^2 measurements	$\mathcal{O}(n^2m^2)$	$\mathcal{O}(nm^2)$
Storage cost for matrices	n^2m^2	2 nm
Failure probability for ε on Σ_3^k if m satisfies	$e^{-c_1m^2}$ $m\ge\mathcal{O}(ck\sqrt{\log\frac{n^2}{k^2}})$	e^{-c_2m} $m\ge\mathcal{O}(k^2\log^2\frac{n}{k})$

4. Applications of 2D random projection

In this section, we use 2D-RP in the context of two representative applications. First, as an example of low-complexity inference tasks, we consider the problem of 2D compressive classification, which is concerned with image classification based on relatively a few 2D random measurements. In particular, we study the problem of multiple hypothesis testing based on (possibly noisy) 2D random measurements. Detailed theoretical analysis along with derivation of an error bound for an important special case is also provided. Next, as an application to low-dimensional signal models, 2D-RP is exploited for compressive 2D-signal reconstruction, in which we rely only on a few non-adaptive linear random measurements for recovery of sparse images [32]. Theoretical requirements for recovery, as well as a fast and effective algorithm for image reconstruction are discussed.

4.1. 2D compressive classification

A few recent studies have shown that classification can be accurately accomplished using random projections [3,4,8,33], which indeed suggests random projections as an effective, reliable, and yet universal feature extraction and dimension reduction tool. Here we apply 2D-RP to the problem of multiple hypothesis testing in an image database. The problem under consideration can be formally described as follows. Let $\mathcal{X}=\{\mathbf{X}_i\}_{i=1}^L$ denote a set of $n\times n$ known images. The "true" image $X_T\in\mathcal{X}$ is contaminated by noise and then projected onto $\mathbb{R}^{m\times m}$ to obtain $Y=A(X_T+N)B^T$, where $N\in\mathbb{R}^{n\times n}$ represents the noise and A and B are $m\times n$ AR matrices. This can be equivalently stated as $y=(B\otimes A)(x_T+n)$, in which $y=vec(Y)$, $x_T=vec(X_T)$, and $n=vec(N)$. Now, given only the low-dimensional random projection Y, we will be concerned with discrimination among the members of $\mathcal{X}$. Given A and B, failure will be quantified in terms of the expected error. For the sake of simplicity, we further assume that noise is Gaussian and white, i.e. $n\sim\mathcal{N}(0,\sigma^2\mathrm{I}_{n^2})$, where I_a denotes the $a\times a$ identity matrix. Moreover, to meet the requirements of Theorem 3 and to preserve the distribution of noise after projection, A and B are chosen to be random orthoprojectors, with entries of $B\otimes A$ being sub-Gaussian. Provided that elements of $\mathcal{X}$ happen equally likely, the Bayes decision rule is [34]:

$$\tilde{x}_l=\underset{x_l\in vec(\mathcal{X})}{\operatorname{argmin}}\|y-(B\otimes A)x_l\|_2=\underset{X_l\in\mathcal{X}}{\operatorname{argmin}}\|Y-AX_lB^T\|_F, \tag{2}$$

in which $\|\cdot\|_F$ denotes the Frobenius norm. The associated expected error would be [34]:

$$\begin{aligned}Err(A,B)&\triangleq 1-\frac{1}{L}\int_y\max_l\{p_l(\tilde{y})\}\,d\tilde{y}\\&=\frac{1}{L}\sum_{l=1}^{L}\int_{R_l^c}p_l(\tilde{y})\,d\tilde{y},\end{aligned} \tag{3}$$

where $p_l(\tilde{y})=\mathcal{N}((B\otimes A)x_l,\sigma^2I_{m^2})$ stands for the conditional density of $\tilde{y}$ given x_l. Also $R_l\subset\mathbb{R}^{m^2}$ is the region in which

$p_l(\cdot)$ achieves the maximum among $\{p_{l'}(\cdot)\}$. Thus, $R_l \triangleq \{\tilde{y} \mid \text{argmax}\, p_{l'}(\tilde{y}) = l\}$. The superscript C denotes the complement of a set. Now let us define $d_{min} \triangleq \min_{l \neq l'} \|x_l - x_{l'}\|_2$. The following result in proved in Appendix C.

Theorem 5. *With a probability of at least* $1-e^{-c_2 m}$ *and provided that* $m \geq c_1 \ln L$, *the average classification error is bounded as*

$$Err(A,B) \leq \sqrt{\frac{2}{\pi}} r^{-1} e^{-r^2/2}, \tag{4}$$

where

$$r \triangleq \sigma^{-1} \sqrt{1-\varepsilon}\, \frac{m}{n} d_{min}.$$

If $m \geq \max(c_1 \ln L, \sqrt{\frac{2}{\pi}} \sigma (\sqrt{1-\varepsilon}\, d_{min})^{-1} n)$, *the above bound can be simplified to*

$$Err(A,B) \leq e^{-(1-\varepsilon)(m^2/n^2)(d_{min}^2/2\sigma^2)}.$$

Here, A and B are random orthoprojectors and we assume that the entries of $B \otimes A$ *are sub-Gaussian. Also,* c_1 *and* c_2 *are constants that depend on* ε *and specified in the proof.*

It is observed that, as the number of observations m^2 increases, the classification error decays exponentially fast. This is also experimentally confirmed in Section 5 with synthetic and real images. Furthermore, the dependence on L is only via d_{min} and the required number of measurements. In the context of 1D-signal classification, estimation, and detection, this exponential rate of decay has previously appeared in [3,4], in which, authors have shown that, despite the loss in information due to non-adaptive projection, statistical inference based on few 1D random measurements achieves a performance comparable to traditional classification using the original images. For the rest of this paper, the above classifiers based on 1D-RP and 2D-RP of signals will be referred to as 2D compressive classifier (2D-CC), and 1D compressive classifier (1D-CC), respectively, where the later simply applies nearest neighbor rule to 1D-RP of signals. Finally, as experimentally verified in Section 5, these remarkable results are not limited to orthoprojectors, but also hold for several other types of random matrices which meet the conditions stated in Theorem 3.

4.2. Sparse 2D-signal reconstruction

In conventional sparse signal reconstruction, random linear projections of sparse (or compressible) 1D-signals have been shown, with high probability, to contain enough information for signal reconstruction within a controllable mean-squared error, even when the observations are corrupted by additive noise [7,23,35,36,24]. The main challenge is then to recover the high-dimensional sparse signal from a few linear random measurements. Although such inverse problems turn out to be ill-posed in general, sparse signal reconstruction algorithms exploit the additional assumption of sparsity to identify the correct signal. In this section, we consider the application of 2D-RP to the problem of sparse image reconstruction. Suppose that a sparse $n \times n$ 2D-signal $X^* \in \Sigma_k^i$ is given, $i \in \{1,2,3\}$. Then, $m \times n$ AR matrices A and B are used to project X^* onto $Y = AX^*B \in \mathbb{R}^{m \times m}$, or equivalently $y = (B \otimes A)x^*$, where $y = vec(Y)$ and $x^* = vec(X^*)$. Now, the goal is to recover X^* given Y. We will observe that, under some conditions on the structure of X^*, solving the following problem uniquely recovers X^* from Y.

$$\mathcal{P}_0 : \underset{x \in \mathbb{R}^{n^2}}{\text{argmin}} \|x\|_0 \quad \text{s.t. } (B \otimes A)x = y, \tag{5}$$

in which $\|x\|_0$ denotes the ℓ_0-norm, i.e. number of non-zero entries, of x. Uniqueness conditions are specified in the following result, which is proved in Appendix D.

Theorem 6. *Suppose that* $Y = AX^*B^T$ *is given, where* $X^* \in \mathbb{R}^{n \times n}$, $Y^* \in \mathbb{R}^{m \times m}$, *and* $A,B \in \mathbb{R}^{m \times n}$. *Then, if any of the following conditions are met, solving* $\mathcal{P}_0$ *uniquely recovers* X^* *from* Y.

$X^* \in \Sigma_k^1$ *and* $B \otimes A$ *has isometry constant* $\varepsilon \in (0,1)$ *for* Σ_{2k}^1.

$X^* \in \Sigma_k^2$ *and* $B \otimes A$ *has isometry constant* $\varepsilon \in (0,1)$ *for* Σ_{2k}^1.

$X^* \in \Sigma_k^3$ *and* $B \otimes A$ *has isometry constant* $\varepsilon \in (0,1)$ *for* $\Sigma_{\sqrt{2}k^3}$.

Since $\Sigma_k^1 \subset \Sigma_k^2 \subset \Sigma_k^3$, provided $B \otimes A$ has isometry constant $\varepsilon \in (0,1)$ for $\Sigma_{\sqrt{2}k^3}$, the accurate recovery of $X^* \in \Sigma_k^i$, $i \in \{1,2\}$, is guaranteed. Similar arguments are also valid. Therefore, in combination with Theorem 4, the above theorem implies that, there exists constants $c_1, c_2 > 0$, such that solving $\mathcal{P}_0$ uniquely recovers $X^* \in \Sigma_k^i$ from Y with probability exceeding $1-e^{-c_2 m}$, provided $k \leq c_1 \sqrt{m}/\log n/k$ and that the entries of $B \otimes A$ remain sub-Gaussian random variables, $i \in \{1,2,3\}$. Directly solving $\mathcal{P}_0$, however, is intractable as it requires a combinatorial search. Moreover, since any small amount of noise completely changes the ℓ_0-norm of a vector, this method is prone to errors in noisy settings [37,38]. In turn, several alternative approaches, such as basis pursuit, matching pursuit, and FOCUSS have been considered to pursue sparse solutions [26,29,27,39–41]. These algorithms essentially attempt to identify a solution which matches the observations, but also has a sparse representation in some basis. Instead of pursuing conventional techniques, we consider the smoothed ℓ_0-norm algorithm for 1D sparse signal reconstruction (1D-SL0) [30]. 1D-SL0 algorithm iteratively minimizes a smoothed version of the ℓ_0-norm and is shown to run much faster than the conventional algorithms, while producing solutions with the same or better accuracy. For a more detailed description of 1D-SL0 algorithm, the interested reader is referred to Appendix E. This appendix also provides the proof of the following theorem, which discusses the application of this algorithm for sparse 2D-signal reconstruction.

Theorem 7. *Suppose* $\varepsilon \in (0,1)$ *is given. There exists constants* $c_1, c_2 > 0$ *depending on* ε, *such that with probability exceeding* $1-e^{-c_2 m}$, *SL0 algorithm uniquely recovers any* $X^* \in \Sigma_k^i$ *from* $Y = AX^*B$, *provided the Kronecker product of the* $m \times n$ *AR matrices A,B remains sub-Gaussian,* $i \in \{1,2,3\}$. *This theorem requires* $k \leq c_1 \sqrt{m}/\log n/k$, *and that the algorithm does not get trapped into local maxima.*

Please cite this article as: A. Eftekhari, et al., Two-dimensional random projection, Signal Process. (2011), doi:10.1016/j.sigpro.2011.01.002

- Initialization:
  - Let $\hat{\mathrm{X}}^0 = \mathrm{A}^{\dagger}\mathrm{Y}\left(\mathrm{B}^{\dagger}\right)^T$, where the superscript † denotes the pseudo-inverse.
- Let $\sigma_1 = 2\max_{i,j}\left|\hat{x}^0_{ij}\right|$, where $\hat{x}^0_{ij}$ is the i,j entry of $\hat{\mathrm{X}}^0$. Choose suitable constants c and J. Then assume a decreasing sequence for σ, as $\left[\sigma_1, c\sigma_1, \ldots, c^J\sigma_1\right]$.
- For $j = 1, \ldots, J$
  - Let $\sigma = \sigma_j$.
  - Maximize (approximately) the function $F_\sigma(\mathrm{X}) = \sum_{i,j}\exp\left(-x^2_{ij}/2\sigma^2\right)$ on the feasible set $\left\{\mathrm{X}|\mathrm{Y} = \mathrm{AXB}^T\right\}$:
    - Initialization: $\mathrm{X} = \hat{\mathrm{X}}^{j-1}$.
    - For $l = 1, \ldots, L$
      - Let $\Delta = [\delta_{ij}]_{ij}$, where $\delta_{ij} \triangleq \exp\left(-x^2_{ij}/2\sigma^2\right)$.
      - Let $\mathrm{X} \leftarrow \mathrm{X} - \mu_0\Delta$, where μ_0 is a small positive constant.
      - Project X onto the feasible set: $\mathrm{X} \leftarrow \mathrm{X} - \mathrm{A}^{\dagger}\left(\mathrm{AXB}^T - \mathrm{Y}\right)\left(\mathrm{B}^{\dagger}\right)^T$.
    - Set $\hat{\mathrm{X}}^j = \mathrm{X}$.
- Final answer if $\hat{\mathrm{X}} = \hat{\mathrm{X}}^J$.

Fig. 1. 2D-SL0 algorithm by Ghaffari et. al. [15].

As further described in experiments, SL0 algorithm produces remarkable experimental results. A detailed convergence analysis guarantees that SL0 finds the unique sparsest solution (and thus avoids the local maxima), if appropriate conditions are met. Interested reader is referred to [42]. Without perturbing the recovery criteria, SL0 algorithm has been adapted to deal with 2D-RP [15]. The resulting 2D-SL0 algorithm accomplishes the signal reconstruction in the matrix domain and hence is much faster and more efficient for images, compared to the 1D-SL0. For convenience, this algorithm is summarized in Fig. 1. Finally, we should emphasize that, as a result of the presence of noise in practical situations, $AX^*B^T = Y$ not exactly but approximately holds and it would be more appropriate to seek for sparse approximate representations, instead. Though not considered here, extension of the above results to the noisy case is easily accomplished using the method presented in [37].

5. Experiments

In this section, the effectiveness of 2D-RP is demonstrated via comprehensive experiments with synthetic and real images. First, we evaluate the performance of 2D-CC (Section 4) for multiple hypothesis testing in databases of synthetically generated random images and real retinal images. Secondly, successful application of 2D-SL0 (Section 4) to synthetic random images illustrates the advantages of 2D-RP in the context of sparse image reconstruction. Our experiments are performed in MATLAB8 environment using an Intel Core 2 Duo, 2.67 GHz processor with 3.24 GB of memory, and under Microsoft Windows XP operating system. Moreover, CPU time is used as a rough indicator of the computational complexity of algorithms.

5.1. 2D compressive classification

5.1.1. Random images

Here, application of 2D-RP to the problem of multiple hypothesis testing is quantitatively assessed on a finite set of synthetically generated random images. Pixels of each image $X_l \in \mathbb{R}^{256\times256}$, $l = 1, \ldots, 400$, are independently obtained from a uniform distribution. To meet the conditions stated in Theorem 3, entries of the $m \times 256$ projection matrices A,B are drawn independently from a Bernoulli distribution $\{1/\sqrt{n}$ with probability 1/2, $-1/\sqrt{n}$ with probability 1/2$\}$, followed by Gram–Schmidt orthonormalization to obtain an orthoprojector. 2D-RP is then used to project the images $\mathcal{X} = \{X_i\}_{i=1}^L$ onto $\mathbb{R}^{m\times m}$ to obtain $\mathcal{Y} = \{AX_iB^T\}_{i=1}^L$. The "true" image is chosen uniformly at random from the set $\mathcal{X}$ and contaminated by additive white Gaussian noise with $\sigma_n^2 = 0.1$. The obtained noisy image is then projected onto $\mathbb{R}^{m\times m}$ with 2D-RP and labeled in consistence with the nearest member of $\mathcal{Y}$.

Alternatively, with entries of the $m^2 \times 256^2$ projection matrix D obtained similarly, $\mathcal{X}$ may be projected onto $\mathbb{R}^{m^2}$ using 1D-RP to obtain $\mathcal{Y} = \{Dx_i\}_{i=1}^L$, where $x_i = vec(X_i)$. 1D compressive classification then assigns a noisy image to the nearest member of $\mathcal{Y}$ in $\mathbb{R}^{m^2}$. In either of the two cases, the average misclassification rates for several values of m are recorded by averaging the empirical errors in 1000 trials. For each trial, independent realizations of noise and projection vectors were generated. Fig. 2 depicts the resulting averaged misclassification rates of 2D-CC and 1D-CC for several values of m. Computational complexity of 2D-CC and 1D-CC are compared in Fig. 3. While explicit calculation of the bound in (4) is intractable [33], we notice that the exponential nature of error is in consistence with our expectations. It is also observed that 2D-CC runs much faster than 1D-CC for all values of m, while producing results with negligible loss in the performance.

Please cite this article as: A. Eftekhari, et al., Two-dimensional random projection, Signal Process. (2011), doi:10.1016/j.sigpro.2011.01.002

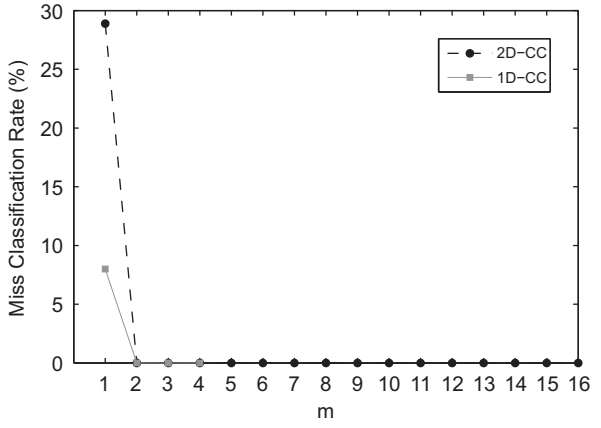

Fig. 2. Misclassification rate (%) of 2D-CC and 1D-CC on random images using m^2 random observations and $\sigma_n^2 = 0.1$. Due to limited memory resources, 1D-CC was not applicable to $m > 4$.

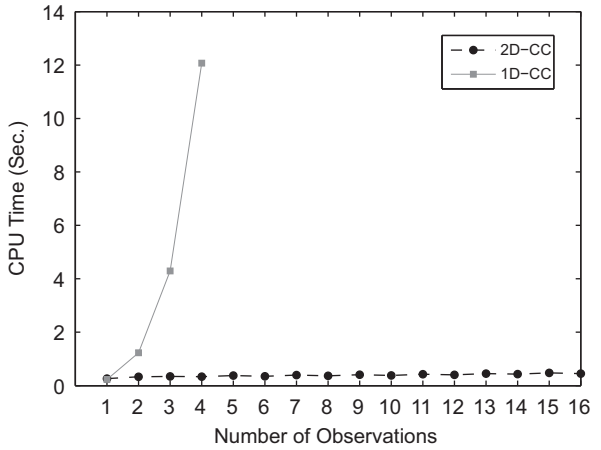

Fig. 3. CPU time (S) for 2D-CC and 1D-CC on random images using m^2 observations and $\sigma_n^2 = 0.1$. Due to limited memory resources, 1D-CC was not applicable to $m > 4$.

In addition, 2D-CC enjoys significantly less memory requirements. Now, to study the effect of noise level, σ_n^2 is varied between 0 and 0.5 and the misclassification rates of 2D-CC and 1D-CC are depicted in Fig. 4, which shows reasonable robustness against noise.

We next conduct an experiment to study the performance of 2D-CC in the general case of non-symmetric projection matrices. For $\sigma_n^2 = 0.1$, we calculate the misclassification rates of 1D-CC and 2D-CC on $\mathcal{X}$ for several values of m_1,m_2. Results, depicted in Fig. 5, verify the similar advantages of 2D-RP with asymmetric left and right projection matrices. Finally, note that the above results apply to other classes of random matrices without any notable difference in performance. This is shown in Fig. 6, using few types of random matrices and setting $\sigma_n^2 = 0.2$.

5.1.2. Retinal images

Retinal biometrics refers to identity verification of individuals based on their retinal images. The retinal vessel distribution pattern, as a biometric trait, has several desirable properties such as uniqueness, time-invariancy, and noninvasiveness, which places it as one of

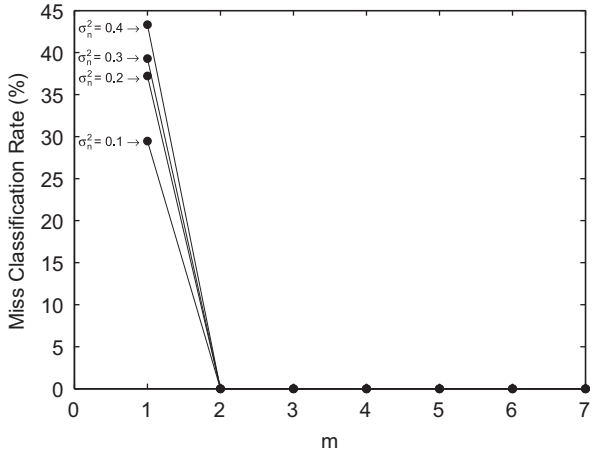

Fig. 4. Performance of 2D-CC on random images using m^2 random observations for different noise levels.

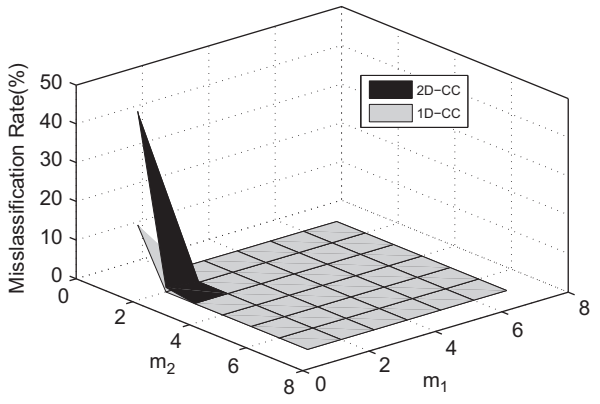

Fig. 5. Misclassification rate (%) of 2D-CC and 1D-CC on random images using $m_1 m_2$ observations and $\sigma_n^2 = 0.2$. Due to limited memory resources, 1D-CC was not applicable to $m > 4$.

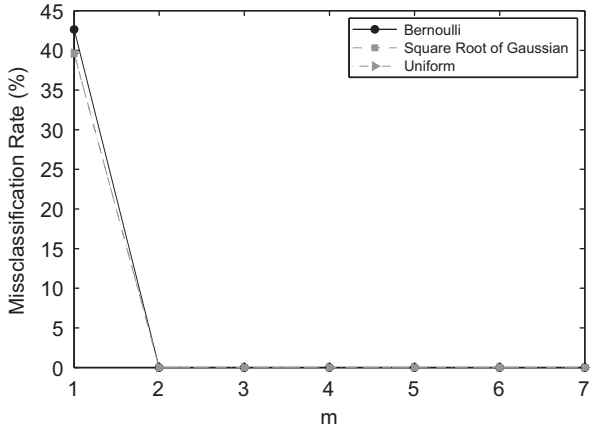

Fig. 6. Performance of 2D-CC using m^2 observations, $\sigma_n^2 = 0.2$, and for several types of random projection matrices.

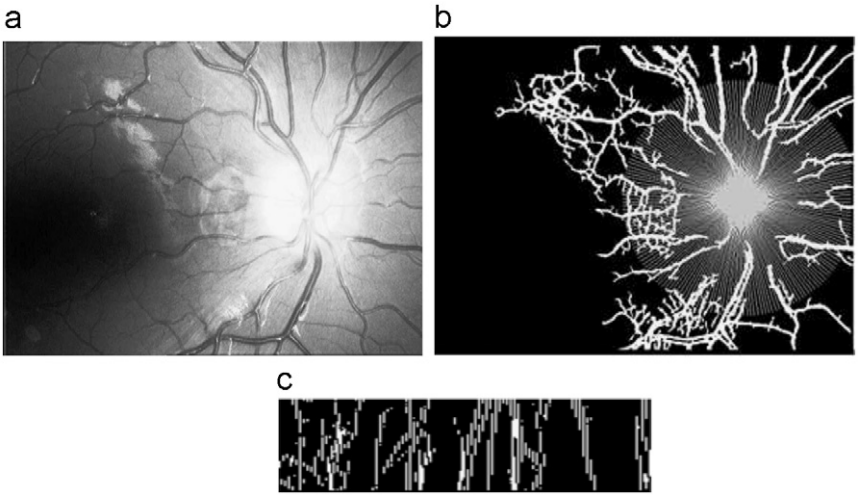

Fig. 7. (a) Retinal vessel map and OD (bright area), (b) Vessel tree and ring-shape mask. (c) Feature matrix for $n_1=100$, $n_2=300$.

the most accurate biometric feature [43]. This pattern is unique for each individual, and does not change through the individual's life, unless a serious pathology appears in the eye. Also, the location of vessels makes it almost impossible to forge. Fig. 7a depicts the retinal vessels and optic disc (OD). We note that the optic disc is usually used as a reference point in retina and vessels converge to it. Retinal-based biometric system relies on feature extraction from retinal vessel map for identification. This system consists of two important steps: (1) Enrollment, in which template characteristics of all individuals are extracted from their retinal images and stored in the database, and (2) Identification or verification, where the identity of the user is, respectively, determined or verified by comparing the feature vector (or matrix) of the user to the database. Our experiment is conducted on VARIA database which contains 153 retinal images of 59 individuals [44]. Few samples of the database are shown in Fig. 8. In this figure, images in the same row belong to the same person. For each retinal image, following the methodology presented in [45], OD is first localized and vessels are segmented. Then, a ring-shaped mask with proper radii centered at OD is used to form the $n_1 \times n_2$ feature matrix by collecting the intensities along $n_2=200$ beams of length $n_1=100$ originating from OD. This process is depicted in Figs. 7b–c. Once all images are processed, the set of feature matrices is obtained. 2D-CC is then used to project the feature matrices onto $\mathbb{R}^{m_1 \times m_2}$. Similar to previous experiments, random admissible matrices $A \in \mathbb{R}^{m_1 \times n_1}$ and $B \in \mathbb{R}^{m_2 \times n_2}$ with independent Bernoulli entries are used for 2D-RP. Let X, $Y=AXB^T$, and M_i denote the new feature matrix, its projection, and the mean of lth class, $l=1,\ldots,59$, respectively. Then, we will classify the new feature matrix X according to the following rule:

$$\hat{l} = \underset{l=1,\ldots,59}{\operatorname{argmin}} \|Y - AM_lB^T\|_F.$$

The above classification rule assigns Y to the class with the closest mean and enhances our performance by reducing the effect of noise. The identification error is measured using the leave-one-out scheme, i.e. for each retinal image, the identity is predicted by the algorithm trained on the rest of the dataset. Similarly, we perform the dimension reduction and classification in $\mathbb{R}^{m_1 m_2}$ using 1D-RP. The average error rates of two algorithms are compared over 100 independent repetitions for a wide range of values for m_1 and m_2 (Fig. 9). Again, we notice the exponential nature of the error.

Moreover, due to highly redundant nature of feature matrices along their columns, for "wise" choices of m_1 and m_2 which consider this redundancy, 2D-CC outperforms 1D-CC and exhibits slightly better performance. In other words, 2D-CC, unlike 1D-CC, can take the available redundancy into account by picking m_1 small (for a fixed $m_1 \cdot m_2$). In contrast, for "careless" choices of m_1 and m_2, 2D-CC performs worse than 1D-CC, as depicted in Fig. 10. 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, may disappear with proper choices for m_1 and m_2 which takes the prior knowledge into account. In addition, 2D-CC enjoys significantly less memory requirements.

5.2. Sparse image reconstruction

This section presents the experimental results for sparse image reconstruction from 2D random measurements. Our simulations were conducted using synthetically generated random sparse $n \times n$ images. Given $k < n$, sparse image X^* was randomly selected from Σ_k^i, $i \in \{1,2,3\}$, where nonzero entries of X^* were independently drawn from the uniform density over [0,1]. Generation of sample images from Σ_k^1 and Σ_k^2 is then straightforward i.e. each sample of Σ_k^3 was generated by fixing the rows and selecting k random positions in each row for nonzero entries. Also, projection matrices $A,B \in \mathbb{R}^{m \times n}$ were independently drawn from the Bernoulli distribution $\{1/\sqrt{n}$ with probability 1/2, $-1/\sqrt{n}$ with probability 1/2$\}$, which clearly meets the conditions stated in Theorem 6. 2D-RP was then used to obtain the observation Y under the noisy model $Y=AX^*B^T+N$, where the entries of N were independently drawn from $\mathcal{N}(0,\sigma_n^2)$.

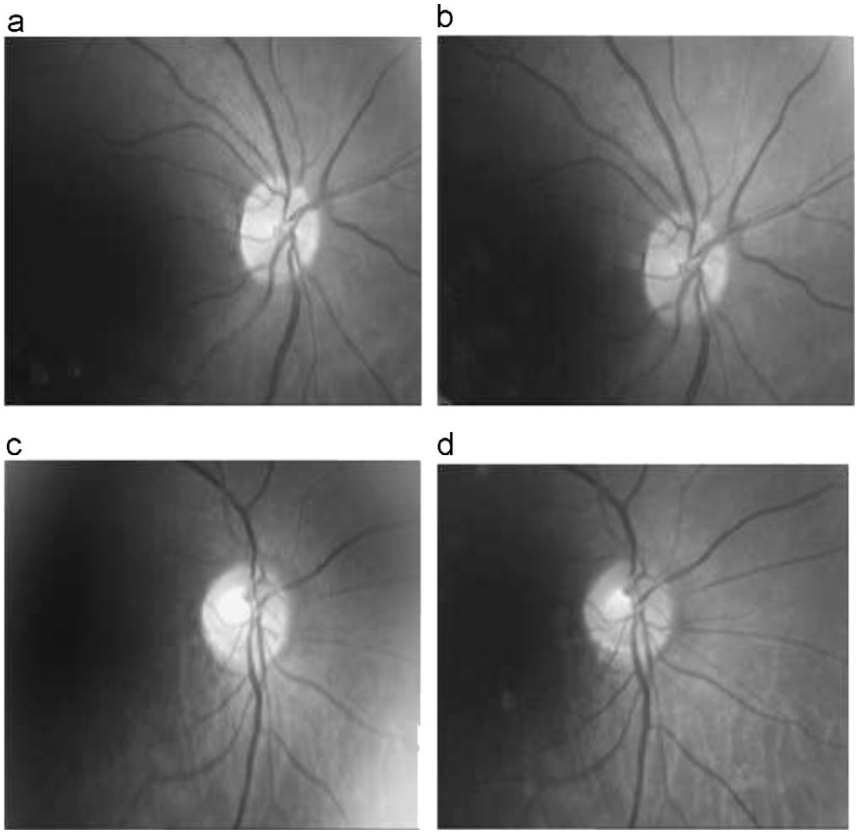

Fig. 8. Sample retinal images from the VARIA database. Images in each row belong to one person.

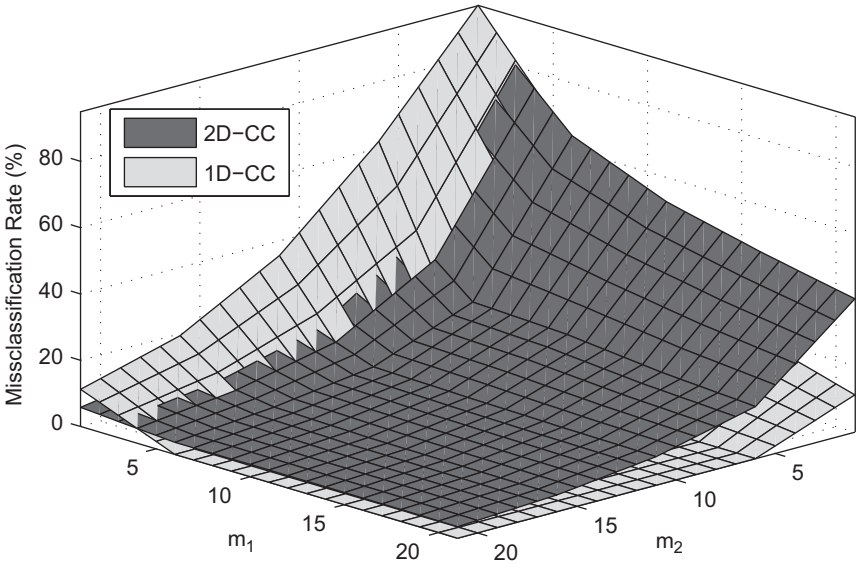

Fig. 9. Failure in retinal identification in VARIA database using 2D-CC and 1D-CC with m_1 and m_2 random measurements.

Finally, 2D-SL0 algorithm was used to recover X^* from Y. The signal to noise ratio (SNR), defined as $20\log(\|X^*\|_F/\|X^*-\hat{X}\|_F)$ with $\hat{X}$ denoting the obtained estimation, was used as the measure of performance. The following set of parameters were used for 2D-SL0: $L=10$, $\mu_0=1$, $c=0.5$, $\sigma_{min}=0.005$. Similarly, we used 1D-RP to project X^* onto $\mathbb{R}^{m^2}$ using an $m^2\times n^2$ projection matrix with independent Bernoulli entries. 1D-SL0 algorithm was then applied for

Please cite this article as: A. Eftekhari, et al., Two-dimensional random projection, Signal Process. (2011), doi:10.1016/j.sigpro.2011.01.002

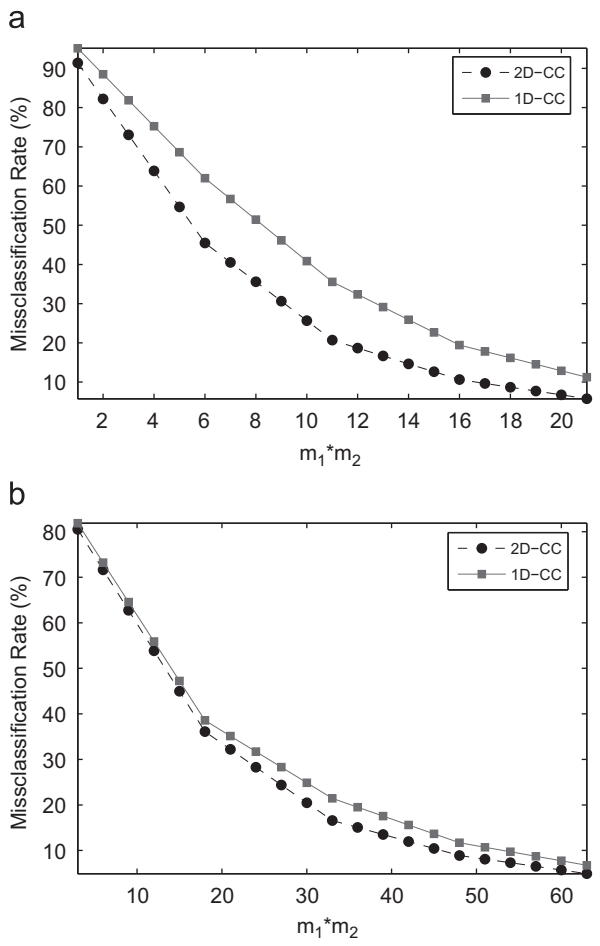

Fig. 10. Two examples of "wise" choices which consider the redundancy along columns: $m_1=1$ (a) and $m_1=3$ (b).

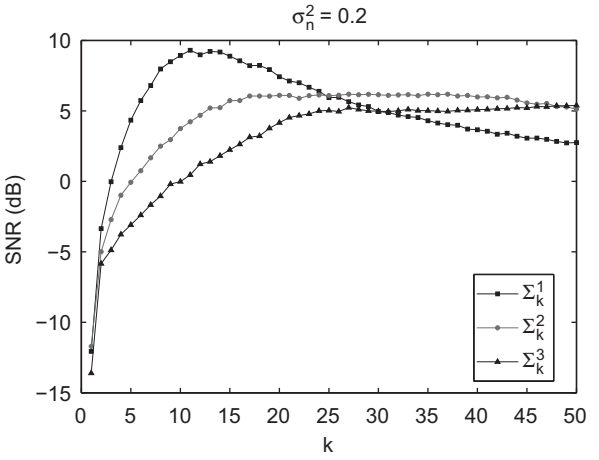

Fig. 11. Average performance in sparse image reconstruction, obtained by 2D-SL0 when sparse random image belongs to Σ_k^i, $i \in 1,2,3$, and $\sigma_n^2 = 0.001$.

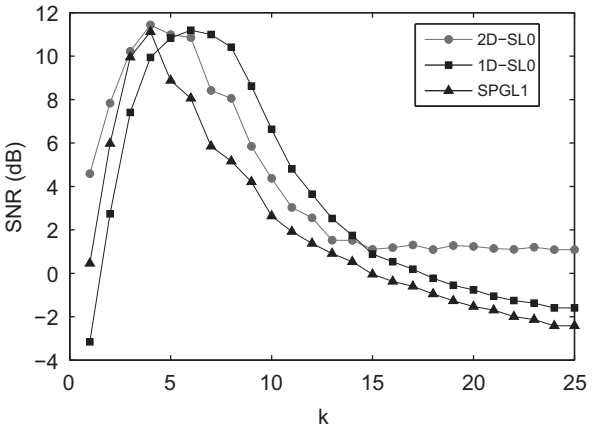

Fig. 12. Average performances in sparse image reconstruction, obtained by 2D-SL0, 1D-SL0, and SPGL1, when sparse random image belongs to Σ_k^3, and $\sigma_n^2 = 0.001$.

recovery with the same set of parameters as of 2D-SL0. In the absence of numerical inaccuracies, both 1D-SL0 and 2D-SL0 have been proved to be mathematically equivalent, and their difference is only in the speed and memory that they need [15]. Results were also quantitatively compared to SPGL1, as a novel, fast and accurate sparse reconstruction algorithm based on the vector space point of view [46]. Parameters of this algorithm were set to their default values. As our first experiment, setting $n=100$, $m=50$, $\sigma_n^2 = 0.2$, and varying k from 1 to 50, we studied the performance of 2D-SL0 for reconstruction of sparse images in Σ_k^i, $i \in \{1,2,3\}$. The average performance over 100 trials is reported in Fig. 11, where for each trial, independent realizations of noise, sparse image, and projection matrices were generated. For all meaningful values of k, results demonstrate the better performance of 2D-SL0 on images in Σ_k^1 and Σ_k^2, which take advantage of the row/column-wise structure of the image.

Finally, we compared our approach, namely 2D-RP + 2D-SL0, to other algorithms over completely generic 2D-signals. Of interest was to show that using the Kronecker product of random matrices (rather than full matrices) produces very reliable performance with lower computational complexity and memory requirements. To achieve this goal, we used Σ_k^3 (sparse 2D signals with arbitrary distributed nonzero entries) in this experiment. Setting $n=50$, $m=20$, $\sigma_n^2 = 0.01$ and varying k from 1 to 25, the average performances of 1D-SL0 and 2D-SL0 for sparse image reconstruction over 100 trials are reported in Fig. 12. Also, Fig. 13 compares the computational complexities of these algorithms. It is observed that, though roughly equal in terms of reconstruction accuracy, a significant gain in memory and computational cost is obtained by using 2D-SL0. We shall emphasize that, due to extreme memory requirements of 1D-SL0 algorithm, using larger values for n was not feasible in this experiment. Finally, in an attempt to study the robustness of 2D-SL0 against noise, n, m, and k are set to 50, 20 and 5, respectively, and the noise power σ_n^2 is varied from 0 to 0.5. Average performances of different sparse reconstruction algorithms are demonstrated in Fig. 14. It is observed that the performances of all the algorithms degrade when noise increases. We also note that the dependence of 2D-SL0 on

Please cite this article as: A. Eftekhari, et al., Two-dimensional random projection, Signal Process. (2011), doi:10.1016/j.sigpro.2011.01.002

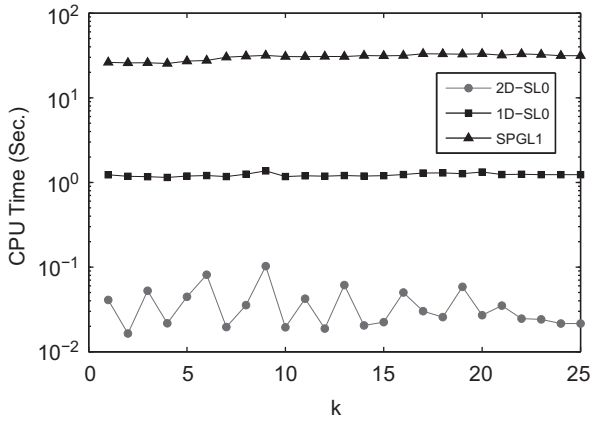

Fig. 13. CPU time (S) for 2D-SL0, 1D-SL0, and SPGL1, when sparse random image belongs to Σ_k^3, and $\sigma_n^2 = 0.01$.

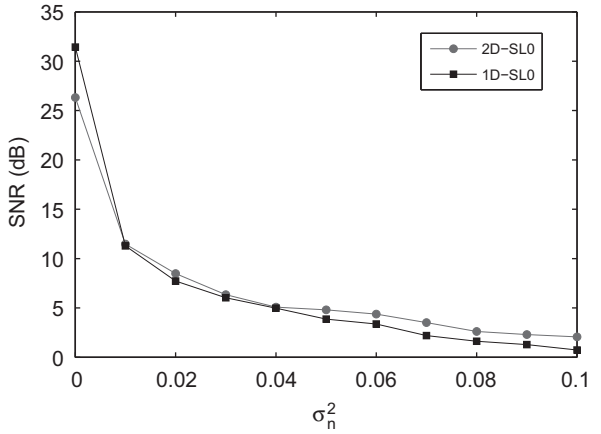

Fig. 14. Performances of 2D-SL0 and 1D-SL0 for different noise levels, when sparse random image belongs to Σ_k^3.

its parameters, though not examined here, is similar to 1D-SL0, which has been thoroughly explored in [30].

6. Conclusions

In this paper, random projection technique was extended to directly leverage the matrix structure of images. We then studied the proposed 2D-RP and its implications for signals, arbitrary finite sets, and infinite sets with low-complexity signal models. These findings were then used to develop 2D-CC for image classification, along with an error bound for an important special case. The proposed classifier proved to be successful in experiments with arbitrary finite sets of synthetic and real images. In addition, 2D-RP was used in the context of sparse image reconstruction. Corresponding theoretical recovery conditions, as well as the recovery conditions of 2D-SL0 algorithm were discussed. Comprehensive validation of these results with synthetic and real images demonstrates significant gain in memory and processing requirements, at the cost of moderate or negligible loss in the performance. Provided results are yet limited to a class of random variables that satisfy certain conditions. Promising experiments with other types of random variables encourages us to seek more general statements, which indeed requires further study.

Acknowledgment

A. Eftekhari gratefully acknowledges Alejandro Weinstein for his help with preparing the revised document.

Appendix A. Proof of theorem 3

Our proof employs some of the ideas and techniques presented in [1,47]. Recall that a sub-Gaussian random variable y satisfies $\mathbb{E}e^{uy} \le e^{vu^2}$, for some $v > 0$, and all $u \in \mathbb{R}$. Also, let g be a zero-mean, unit-variance Gaussian random variable, which is assumed to be independent from all other random variables that appear in the proof. We note that, for all $t \in \mathbb{R}$, $\mathbb{E}e^{tg} = e^{t^2/2}$ and, $\mathbb{E}e^{tg^2} = 1/\sqrt{1-2t}$ for all $t \in (0,1/2)$. We will also use the following inequality for nonnegative random variables $y_1, \ldots, y_K$, which is a simple consequence of the Holder's inequality [48].

$$\mathbb{E}\prod_{k=1}^{K} y_k \le \left(\prod_{k=1}^{K} \mathbb{E}y_k^K\right)^{1/K}. \tag{6}$$

Before proceeding to the main proof, we shall prove the following observation.

Lemma 1. *Suppose that A and B satisfy the conditions stated in Theorem2. Let $a \triangleq [a_1, \ldots, a_n]^T$ and $b \triangleq [b_1, \ldots, b_n]^T$ denote a row of A and B, respectively. Then $d \triangleq n[b_1 a^T, \ldots, b_n a^T] \in \mathbb{R}^{n^2}$ denotes the corresponding row of $nB \otimes A$. Also, given $x \in \mathbb{R}^{n^2}$ with $\|x\|_2 = 1$, we partition it to obtain $x = [x_1^T, \ldots, x_n^T]^T$, where $x_k \in \mathbb{R}^n$, $k = 1, \ldots, n$. Now, let us define $u = n\sum_{l=1}^{n} b_l a^T x_l$. Then, for some $v > 0$, the following holds for all $0 \le \alpha \le 1/8nv$:*

$$\mathbb{E}e^{\alpha u^2} \le e^{\alpha + 100n^2 v^2 \alpha^2}. \tag{7}$$

Proof. First, using the fact that the entries of a and b are i.i.d. sub-Gaussian random variables with zero mean and variance $1/n$, we can write

$$\begin{aligned}\mathbb{E}u^2 &= n^2 \mathbb{E}_b \mathbb{E}_a \sum_{k=1}^{n} b_k^2 (x_k^T a a^T x_k) + n^2 \mathbb{E}_b \mathbb{E}_a \sum_{k \ne l} b_k b_l (x_k^T a a^T x_l) \\ &= n\mathbb{E}_b \sum_{k=1}^{n} b_k^2 \|x_k\|^2 + n\mathbb{E}_b \sum_{k \ne l} b_k b_l (x_k^T x_l) = \sum_{k=1}^{n} \|x_k\|_2^2 = 1.\end{aligned} \tag{8}$$

Now we find a simple upper bound on $\mathbb{E}e^{\alpha u^2}$, which will be refined later

$$\begin{aligned}\mathbb{E}e^{\alpha u^2} &= \mathbb{E}_u \mathbb{E}_g e^{\sqrt{2\alpha} u g} = \mathbb{E}_u \mathbb{E}_g e^{n\sqrt{2\alpha} g \sum_{k=1}^{n} b_k a^T x_k} \\ &\le \mathbb{E}_g \left(\prod_{k=1}^{n} \mathbb{E}e^{n^2 \sqrt{2\alpha} g b_k a^T x_k}\right)^{1/n} \\ &\le \mathbb{E}_g \left(\prod_{k,l=1}^{n} \mathbb{E}e^{n^2 \sqrt{2\alpha} g x_{kl} a_l b_k}\right)^{1/n},\end{aligned} \tag{9}$$

where x_{kl} is the lth entry of x_k. Using the hypothesis that $a_l b_k$ is a sub-Gaussian random variable and denoting the

corresponding Gaussian standard by v/n^2, (9) is simplified for $0 \le \alpha \le 1/8nv$ to get

$$\mathbb{E}e^{\alpha u^2} \le \mathbb{E}_g\left(\prod_{k,l=1}^{n} e^{2vn^2\alpha x_{kl}^2 g^2}\right)^{1/n} = \mathbb{E}_g e^{2vn\alpha g^2} = \frac{1}{\sqrt{1-4vn\alpha}} \le \sqrt{2}, \tag{10}$$

Now, using (8), the bound in (10) is refined for $0 \le \alpha \le 1/8nv$ as follows:

$$\begin{aligned}\mathbb{E}e^{\alpha u^2} &= \sum_{p=0}^{\infty}\frac{\alpha^p \mathbb{E}u^{2\,m}}{p!} = 1+\alpha+\sum_{p=2}^{\infty}\frac{\alpha^p \mathbb{E}u^{2\,m}}{p!}\\ &= 1+\alpha+\sum_{p=2}^{\infty}\frac{(8nv\alpha)^p(8nv)^{-p}\mathbb{E}u^{2\,m}}{p!}\\ &= 1+\alpha+(8nv\alpha)^2\sum_{p=2}^{\infty}\frac{(8nv\alpha)^{p-2}(8nv)^{-p}\mathbb{E}u^{2\,m}}{p!}\\ &\le 1+\alpha+(8nv\alpha)^2\sum_{p=2}^{\infty}\frac{(8nv)^{-p}\mathbb{E}u^{2\,m}}{p!}\\ &\le 1+\alpha+(8nv\alpha)^2\,\mathbb{E}e^{(u^2/8nv)}\\ &\le 1+\alpha+100n^2v^2\alpha^2\\ &\le e^{\alpha+100n^2v^2\alpha^2}.\end{aligned} \tag{11}$$

This completes the proof of this lemma. □

Now we can complete the proof of Theorem 3.

Proof. (Theorem 3) Note that, due to linearity, it suffices to prove the theorem for the case $\|x\|_2^2 = 1$. We first find an exponentially-decreasing upper bound for

$$\Pr\left\{\|(B\otimes A)x\|_2^2 \ge \frac{m^2}{n^2}(1+\varepsilon)\right\}. \tag{12}$$

Let us define $D \triangleq nB\otimes A$. Then, by hypothesis, the entries of D are zero-mean unit-variance sub-Gaussian random variables and (12) can be equivalently written as

$$\Pr\{\|Dx\|_2^2 \ge m^2(1+\varepsilon)\}. \tag{13}$$

Invoking the Chernoff bounding technique [39], for any $t>0$, we have

$$\Pr\{\|Dx\|_2^2 \ge m^2(1+\varepsilon)\} \le \frac{\mathbb{E}e^{t\|Dx\|_2^2}}{e^{tm^2(1+\varepsilon)}}. \tag{14}$$

Therefore, it suffices to bound the expectation on the right hand side of (14). Properties of the Kronecker product imply that each row of D is dependent with exactly $2(m-1)$ other rows, and that we can partition the rows of D into m nonoverlapping partitions $\{R_i\}_{i=1}^m$ with $|R_i| = m$, such that the rows in each partition are independent. Let us denote by D_{R_i} the $m\times n^2$ submatrix obtained by retaining the rows of D corresponding to the indices in R_i. Clearly, the rows of D_{R_i} are independent, and we have

$$\|Dx\|_2^2 = \sum_{i=1}^{m}\|D_{R_i}x\|_2^2. \tag{15}$$

Defining $u_{ij} \triangleq (D_{R_i}x)_j$, $j=1,\ldots,m$, we have

$$\mathbb{E}e^{t\|Dx\|_2^2} = \mathbb{E}e^{t\sum_{i=1}^m \|D_{R_i}x\|_2^2} \le \left(\prod_{i=1}^{m}\mathbb{E}e^{mt\|D_{R_i}x\|_2^2}\right)^{1/m} = \left(\prod_{i=1}^{m}\mathbb{E}e^{mt\sum_{j=1}^m u_{ij}^2}\right)^{1/m} = \left(\prod_{i,j=1}^{m}\mathbb{E}e^{mtu_{ij}^2}\right)^{1/m}, \tag{16}$$

where we have used the independence of the rows of D_{R_i} in the second line. Note u_{ij} is the dot product of x and a row of D, we may set $\alpha = mt$ in Lemma 1, to further bound the last term in (16):

$$\mathbb{E}e^{t\|Dx\|_2^2} \le \left(\prod_{i,j=1}^{m}\mathbb{E}e^{mtu_{ij}^2}\right)^{1/m} \le e^{(m^2t+100n^2v^2m^3t^2)}. \tag{17}$$

Using (17) in combination with (14), we finally obtain

$$\Pr\left\{\|Dx\|_2^2 \ge m^2(1+\varepsilon)\right\} \le e^{-(\varepsilon^2m/400n^2v^2)}. \tag{18}$$

Using similar arguments, we can show that

$$\Pr\left\{\|Dx\|_2^2 \le m^2(1-\varepsilon)\right\} \le e^{-(\varepsilon^2m/400n^2v^2)}. \tag{19}$$

Combining (18) with (19) completes our proof. □

Appendix B. Proof of theorem 4

Our proof follows a similar strategy as in [14]. Let $T \subset \{1,\ldots,n^2\}$ be a subset of indices. We first observe that if $B\otimes A$ has isometry constant ε for all signals which are zero outside T, then $B\otimes A$ also has isometry constant ε for all signals which are zero outside any $T' \subset T$. Consequently, it suffices to prove Theorem 4 for all $T \subset \{1,\ldots,n^2\}$ with $|T| = k^2$.

Given a set of appropriate indices T with $|T| = k^2$, let $\mathcal{X}_T \subset \Sigma_k^i$ denote the set of $n\times n$ 2D-signals that are zero outside of T. Also, let us define $vec(\mathcal{X}_T) \triangleq \{vec(X)|X \in \mathcal{X}_T\}$. We then cover the k^2-dimensional subspace of $vec(\mathcal{X}_T)$ with a finite set of points $Q_T \in vec(\mathcal{X}_T)$, such that $\|q\|_2 \le 1$ for all $q \in Q_T$, and $\min_{q\in Q_T}\|x-q\|_2 \le \varepsilon/4$ for all $x \in vec(\mathcal{X}_T)$ with $\|x\|_2 \le 1$. Simple covering arguments show that we can choose a set with $|Q_T| \le (12/\varepsilon)^{k^2}$. Applying the union bound, we find that $B\otimes A$ has isometry constant $\varepsilon/2$ for Q_T with probability exceeding $1-2|Q_T|e^{-\varepsilon^2m/1600n^6v^2}$, where v is the Gaussian standard of the entries of $B\otimes A$. Noting that $1+\varepsilon/2+(1+\varepsilon)\varepsilon/4 \le 1+\varepsilon$, the following inequality is indeed valid:

$$\|(B\otimes A)x\|_2^2 \le \|(B\otimes A)q\|_2^2 + \|(B\otimes A)(x-q)\|_2^2 \le 1+\frac{\varepsilon}{2}+(1+\varepsilon)\frac{\varepsilon}{4} \le 1+\varepsilon. \tag{20}$$

Similarly, since

$$\|(B\otimes A)x\|_2^2 \ge \|(B\otimes A)q\|_2^2 - \|(B\otimes A)(x-q)\|_2^2 \ge 1-\frac{\varepsilon}{2}-(1+\varepsilon)\frac{\varepsilon}{4} \ge 1-\varepsilon. \tag{21}$$

we conclude that $B\otimes A$ has isometry constant ε for $\mathcal{X}_T$, with probability exceeding $1-2(12/\varepsilon)^{k^2}e^{-\varepsilon^2m/1600n^2v^2}$. There exist $\binom{n}{k}^2$, $\gamma\binom{n^2}{k^2}$, and $\binom{n^2}{k^2}$ such choices for $\mathcal{X}_T$ in Σ_k^1, Σ_k^2, and Σ_k^3, respectively, where[4] $\gamma<1$ is an absolute constant. Application of union bound then implies that $B\otimes A$ has isometry

[4] Finding an explicit formula for the upper bound seems to be difficult. Therefore, we just use the coefficient $\gamma<1$ to remind the fact that the number of choices in Σ_k^2 is less than Σ_k^3.

Please cite this article as: A. Eftekhari, et al., Two-dimensional random projection, Signal Process. (2011), doi:10.1016/j.sigpro.2011.01.002

constant ε for Σ_k^i, with probability exceeding P_i, where

$$P_1 = 1-2\binom{n}{k}^2\left(\frac{12}{\varepsilon}\right)^{k^2} e^{-(\varepsilon^2 m/1600n^6\nu^2)}$$

$$P_2 = 1-2\gamma\binom{n^2}{k^2}\left(\frac{12}{\varepsilon}\right)^{k^2} e^{-(\varepsilon^2 m/1600n^6\nu^2)}$$

$$P_3 = 1-2\binom{n^2}{k^2}\left(\frac{12}{\varepsilon}\right)^{k^2} e^{-(\varepsilon^2 m/1600n^6\nu^2)}. \quad (22)$$

Thus, for a fixed c_1, whenever $k \le c_1\sqrt{m}/\log n/k$, we have that $P_i \ge 1-2e^{-c_{2,i}m}$, provided that $c_{2,i}$ satisfies

$$c_{2,1} \le \frac{\varepsilon^2}{1600n^6\nu^2} - 2c_1\frac{1+\ln\frac{n}{k}}{\ln\frac{n}{k}} - c_1^2\frac{\ln\frac{12}{\varepsilon}}{\ln\frac{n}{k}}$$

$$c_{2,2} \le \frac{\varepsilon^2}{1600n^6\nu^2} - c_1^2\frac{1+2\ln\frac{n}{k}+\ln\frac{12}{\varepsilon}}{\ln^2\frac{n}{k}} + \ln\gamma$$

$$c_{2,3} \le \frac{\varepsilon^2}{1600n^6\nu^2} - c_1^2\frac{1+2\ln\frac{n}{k}+\ln\frac{12}{\varepsilon}}{\ln^2\frac{n}{k}}, \quad (23)$$

where we have used the fact that $\binom{n}{k} \le (en/k)^k$. Hence we can always choose c_1 sufficiently small to ensure that $c_{2,i} > 0$, $i \in \{1,2,3\}$. This completes our proof.

Appendix C. Proof of theorem 5

Since, by hypothesis, A and B are $m \times n$ AR matrices, remarks following Theorem 3 imply that $B \otimes A$ has the isometry property for any prescribed $\varepsilon \in (0,1)$ with high probability. In particular, let $c = c(\varepsilon)$ be as specified in Theorem 3, and define $c_1 \triangleq 2c^{-1}\ln L$ and $c_2 \triangleq c/2$. As a consequence, it is easy to check that provided $m \ge c_1 \ln L$, the following holds except with a probability of at least $1-e^{-c_2 m}$:

$$\frac{1}{2}\min_{l\neq l'}\|(B\otimes A)(x_l - x_{l'})\|_2 \ge \sqrt{1-\varepsilon}\frac{m}{n}d_{min} \triangleq r.$$

Define $R'_l \triangleq \{\tilde{y} \mid \|\tilde{y}-(B\otimes A)x_l\|_2 \ge r\}$, and note that $R_l^C \subset R'_l$, because $\|\tilde{y}-(B\otimes A)x_l\|_2 < r$ implies that $\tilde{y}$ is closest to the lth Gaussian. Now, under the event above, it follows from (3) that

$$\begin{aligned}
Err(A,B) &= \frac{1}{L}\sum_{l=1}^{L}\int_{R_l^C} p_l(\tilde{y})\,d\tilde{y} \\
&\le \frac{1}{L}\sum_{l=1}^{L}\int_{R'_l} p_l(\tilde{y})\,d\tilde{y} \\
&= \frac{1}{L}\sum_{l=1}^{L}\int_{\|\tilde{y}\|_2 \ge r}\mathcal{N}((B\otimes A)x_l,\sigma^2 I_{m^2})\,d\tilde{y} \\
&= (2\pi\sigma^2)^{-m^2/2}\int_{\|\tilde{y}\|_2 \ge r} e^{-(\|\tilde{y}\|_2^2/2\sigma^2)}\,d\tilde{y} \\
&= (2\pi\sigma^2)^{-1/2}\int_{\|\tilde{y}\|_2 \ge r} e^{-(\|\tilde{y}\|_2^2/2\sigma^2)}\,d\|\tilde{y}\|_2 \\
&= \pi^{-1/2}\int_{|u| \ge \frac{r}{\sqrt{2}\sigma}} e^{-u^2}\,du \\
&= 1-\mathrm{erf}\left(\frac{r}{\sqrt{2}}\sigma\right) \le \frac{2\sigma}{\sqrt{2\pi}}re^{-(r^2/2\sigma^2)}.
\end{aligned}$$

The third line above follows because the distributions share the same covariance matrix. Also, $\mathrm{erf}(\cdot)$ is the standard error function and the last line is a well-known bound on the error function ($1-\mathrm{erf}(\alpha) \le e^{-\alpha^2}/\sqrt{\pi}\alpha$ for $\alpha > 0$). In particular, when $r \ge 2\sigma/\sqrt{2\pi}$, we obtain the following more compact result:

$$Err(A,B) \le e^{-(1-\varepsilon)((m^2/n^2)/(d_{min}^2/2\sigma^2))},$$

as claimed.

Appendix D. Proof of theorem 6

We only prove the first part. The proof of the other parts of the theorem is very similar. Assume, in contrast, that $AX_1B^T = Y$ and $AX_2B^T = Y$, for $X_1, X_2 \in \Sigma_k^1$, and $X_1 \neq X_2$. This requires that $A(X_1-X_2)B^T = 0$. On the other hand, X_1-X_2 is clearly a member of Σ_{2k}^1. Therefore, our hypothesis on A and B implies that $0 < m/n(1-\varepsilon)\|X_1-X_2\|_2 \le \|A(X_1-X_2)B^T\|_2$, which contradicts our assumption. This completes the proof.

Appendix E. SL0 algorithm and proof of the theorem 7

SL0 algorithm for reconstruction of sparse 1D-signals is formally stated as follows. Consider the problem $\mathcal{P}_0 : \min_x \|x\|_0$ s.t. $Hx = y$, where $x \in \mathbb{R}^n$, $y \in \mathbb{R}^m$ and $H \in \mathbb{R}^{m\times n}$. SL0 approximates the $\|x\|_0$ with a continuous function $n-F_\sigma(x)$, where we usually set $F_\sigma(x) = \sum_{i=1}^n \exp(-x_i^2/2\sigma^2)$ [30]. Therefore, SL0 attempts to solve the following problem:

$$\mathcal{Q}\ \max_x \lim_{\sigma\to 0} F_\sigma(x) \text{ s.t. } Hx = y. \quad (24)$$

However, to avoid getting trapped into several local maxima of $F_\sigma(\cdot)$ for small σ's, SL0 solves a sequence of problems of the form $\mathcal{Q}_\sigma :\ \max_x F_\sigma(x)$ s.t. $Hx = y$, decreasing σ at each step, and initializing the next step at the maximizer of the previous larger value of σ (external loop). Each $\mathcal{Q}_\sigma$ is approximately solved using a few iterations of gradient ascent (internal loop).

Further analysis of the theoretical aspects of SL0 algorithm requires the concept of spark of a matrix [49]. Given $H \in \mathbb{R}^{n\times m}$, *spark*($H$) is defined as the minimum number of columns of H that are linearly dependent. Application of SL0 algorithm to sparse 1D-signals is discussed by Theorem 8, which is merely a restatement of Theorem 1 of [37] for noiseless situation.

Theorem 8 (*Eftekhari et al. [37]*). *Assume that the columns of the projection matrix H are normalized to unit ℓ_2-norm. Suppose also that x^* is given such that $\|x^*\|_0 < 1/2spark(H)$. Then, SL0 algorithm correctly recovers x^*, provided that it is not trapped into local maxima in the internal loop of SL0.*

It should be emphasized that the gradual decrease in σ is aimed avoid the local maxima when maximizing $F_\sigma(\cdot)$ for a fixed σ. Though experimentally studied in [30,37], the question of "how much gradually" is still open for investigation, although we have a convergence proof for SL0 [42]. Using the Gershgorin disc theorem, we may obtain an analogous performance guarantee in terms of isometry constants. This is stated in the following two lemmas.

Please cite this article as: A. Eftekhari, et al., Two-dimensional random projection, Signal Process. (2011), doi:10.1016/j.sigpro.2011.01.002

Lemma 2. *Let $q=q(H)=spark(H)-1$ denote the Kruskal rank of the matrix H. If $\sigma_{min}^{(q)}$ denotes the smallest singular value of all submatrices of H formed by taking q columns of H, then $\sigma_{min}^{(q)}>0$.*

Proof. This follows directly from the definition of spark(H). □

Lemma 3. *Let Σ_k denote the set of 1D-signals $x\in\mathbb{R}^n$ with at most k nonzero entries. Suppose also that the projection matrix $H\in\mathbb{R}^{m\times n}$ has isometry constant $\varepsilon\in(0,1)$ on Σ_{2k}. Then, any given $x^*\in\Sigma_k$ could be uniquely recovered from $y=Hx^*\in\mathbb{R}^m$ using SL0 algorithm, provided that it does not get trapped into the local maxima in the internal loop of S* L0.

Proof. Let H_I denote the column submatrix of H corresponding to the set of indices I with $|I|\le 2k$. Also, let x_I denote the vector obtained by retaining only the entries in x corresponding to I. Then, the hypothesis on H implies that, for $\varepsilon<1$ and for any I with $|I|\le 2k$, we have

$$\sqrt{\frac{m}{n}}(1-\varepsilon)\|x_I\|_2\le\|H_Ix_I\|_2\le\sqrt{\frac{m}{n}}(1+\varepsilon)\|x_I\|_2. \tag{25}$$

Defining $G=\sqrt{n/m}\,H_I$, it is observed that the eigenvalues of G^{TG} belong to the interval (0,2). Assume, without any loss of generality, that the columns of G are normalized to unit ℓ_2-norm. Then, the Gershgorin disc theorem[5] and Lemma 2 together require that the eigenvalues of G^{TG} do not exceed $1+\sigma_{min}^{(q)}(2k-1)$. Therefore $1+\sigma_{min}^{(q)}(2k-1)>2$, or equivalently $1/2k-1<\sigma_{min}^{(q)}$. Consequently, according to Theorem 8, SL0 algorithm correctly recovers x^* when $\|x\|_0\le 1/2(1+(2k-1))=k$, which completes the proof. □

Theorem 4 on Σ_k^3, in combination with Lemma 3 implies Theorem 6.

References

[1] P. Indyk, R. Motwani, Approximate nearest neighbors: towards removing the curse of dimensionality, in: Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing, 1998, pp. 604–613.
[2] C. Hegde, M. Wakin, R. Baraniuk, Random projections for manifold learning, in: Neural Information Processing Systems (NIPS), 2007.
[3] J. Haupt, R. Castro, R. Nowak, G. Fudge, A. Yeh, Compressive sampling for signal classification, in: Proceedings of 40th Asilomar Conference Signals, Systems and Computers, Pacific Grove, CA, 2006, pp. 1430–1434.
[4] M. Davenport, M. Duarte, M. Wakin, J. Laska, D. Takhar, K. Kelly, R. Baraniuk, The smashed filter for compressive classification and target recognition, Proceedings of SPIE, vol. 6498, 2007.
[5] P. Agarwal, S. Har-Peled, H. Yu, Embeddings of surfaces, curves, and moving points in euclidean space, in: Proceedings of the Twenty-Third Annual Symposium on Computational Geometry, 2007.
[6] M. Vetterli, P. Marziliano, T. Blu, Sampling signals with finite rate of innovation, IEEE Transactions on Signal Processing 50 (6) (2002) 1417–1428.
[7] E. Candes, M. Wakin, People hearing without listening: an introduction to compressive sampling, IEEE Signal Processing Magazine 25 (2) (2008) 21–30.
[8] J. Haupt, R. Nowak, Compressive sampling for signal detection, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2007.
[9] J. Lin, D. Gunopulos, Dimensionality reduction by random projection and latent semantic indexing, in: Proceedings of the Text Mining Workshop, at the 3rd SIAM International Conference on Data Mining, 2003.
[10] R. Baraniuk, M. Wakin, Random projections of smooth manifolds, Foundations of Computational Mathematics 9 (1) (2009) 51–77.
[11] E. Bingham, H. Mannila, Random projection in dimensionality reduction: applications to image and text data, in: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2001, pp. 245–250.
[12] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces, Publications Mathematiques de l'IHES 81 (1) (1995) 73–205.
[13] S. Dasgupta, A. Gupta, An elementary proof of the Johnson–Lindenstrauss lemma, Random Structures and Algorithms 22 (1) (2002) 60–65.
[14] R. Baraniuk, M. Davenport, R. DeVore, M. Wakin, A simple proof of the restricted isometry property for random matrices, Constructive Approximation 28 (3) (2008) 253–263.
[15] A. Ghaffari, M. Babaie-Zadeh, C. Jutten, Sparse decomposition of two dimensional signals, Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing2009, pp. 3157–3160.
[16] J. Ye, Generalized low rank approximations of matrices, Machine Learning 61 (1) (2005) 167–191.
[17] A.C. Gurbuz, J.H. McClellan, W.R. Scott Jr., Compressive sensing for subsurface imaging using ground penetrating radar, Signal Processing 89 (10) (2009) 1959–1972.
[18] J. Yang, D. Zhang, A. Frangi, J. Yang, Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 26 (1) (2004) 131–137.
[19] J. Ye, Q. Li, A two-stage linear discriminant analysis via QR-decomposition, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (6) (2005) 929–941.
[20] A. Eftekhari, H. Abrishami Moghaddam, M. Babaie-Zadeh, M. Moin, Two dimensional compressive classifier for sparse images, in: Proceedings of IEEE International Conference on Image Processing, 2009.
[21] A. Magen, Dimensionality reductions that preserve volumes and distance to affine spaces, and their algorithmic applications, in: Lecture Notes in Computer Science, 2002, pp. 239–253.
[22] D. Achlioptas, Database-friendly random projections: Johnson–Lindenstrauss with binary coins, Journal of Computer and System Sciences 66 (4) (2003) 671–687.
[23] D. Donoho, M. Elad, V. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, IEEE Transactions on Information Theory 52 (1) (2006) 6–18.
[24] E. Candes, J. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Communications on Pure and Applied Mathematics 59 (8) (2006) 1207–1223.
[25] A.M.E.D.L. Donoho, On the stability of the basis pursuit in the presence of noise, Signal Processing 86 (3) (2006) 511–532.
[26] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM Review 43 (1) (2001) 129–159.
[27] J. Tropp, A. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, IEEE Transactions on Information Theory 53 (12) (2007) 4655–4666.
[28] P. Xu, D. Yao, Two dictionaries matching pursuit for sparse decomposition of signals, Signal Processing 86 (11) (2006) 3472–3480.
[29] F. Bergeaud, S. Mallat, Matching pursuit of images, Signal and Image Representation in Combined Spaces (1998) 285–288.
[30] G. Mohimani, M. Babaie-Zadeh, C. Jutten, A fast approach for overcomplete sparse decomposition based on smoothed l0 norm, IEEE Transactions on Signal Processing 57 (2009) 289–301.
[31] K. B. Petersen, M. Pedersen, The Matrix Cookbook, 2006 ⟨http://matrixcookbook.com/⟩.
[32] D. Donoho, Compressed sensing, IEEE Transactions on Information Theory 52 (4) (2006) 1289–1306.
[33] M. Davenport, M. Wakin, R. Baraniuk, Detection and estimation with compressive measurements, Technical Report, Rice University ECE Department, 2006.
[34] S. Theodoridis, K. Koutroumbas, Pattern Recognition, Academic Press, New York, 2003.
[35] E. Candes, The restricted isometry property and its implications for compressed sensing, Comptes rendus Mathe-matique 346 (9–10) (2008) 589–592.
[36] P. Bofill, M. Zibulevsky, Underdetermined blind source separation using sparse representations, Signal Processing 81 (2001) 2353–2362.

[5] The Gershgorin disc theorem states that the eigenvalues of a $l\times l$ matrix B all lie in the union of l discs $d_i(c_i,r_i)$, centered at $c_i=b_{ii}$ and with radii $r_i=\sum_{i\neq j}|b_{ij}|$, $i,j=1,\ldots,l$ [48,50].

Please cite this article as: A. Eftekhari, et al., Two-dimensional random projection, Signal Process. (2011), doi:10.1016/j.sigpro.2011.01.002

[37] A. Eftekhari, M. Babaie-Zadeh, C. Jutten, H. Abrishami Moghaddam, Robust-sl0 for stable sparse representation in noisy settings, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP 2009, 2009, pp. 3433–3436.
[38] Y. Tsaig, D.L. Donoho, Extensions of compressed sensing, Signal Processing 86 (3) (2006) 549–571.
[39] I. Gorodnitsky, B. Rao, Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm, IEEE Transactions on Signal Processing 45 (3) (1997) 600–616.
[40] R. Gribonval, R.F.I. Ventura, P. Vandergheynst, A simple test to check the optimality of a sparse signal approximation, Signal Processing 86 (3) (2006) 496–510.
[41] J.A. Tropp, A.C. Gilbert, M.J. Strauss, Algorithms for simultaneous sparse approximation. Part i: greedy pursuit, Signal Processing 86 (3) (2006) 572–588.
[42] H. Mohimani, M. Babaie-Zadeh, I. Gorodnitsky, C. Jutten, Sparse recovery using smoothed l0 (sl0): Convergence analysis, 2010, arxiv: cs.IT/1001.5073 ⟨http://arxiv.org/pdf/1001.5073⟩.
[43] A. Jain, R. Bolle, S. Pankanti (Eds.), Biometrics: Personal Identification in Networked Society, Kluwer Academic Publishers, 1999.
[44] Varia database ⟨http://www.varpa.es/varia.html⟩.
[45] H. Farzin, H. Abrishami-Moghaddam, M.S. Moin, A novel retinal identification system, EURASIP Journal on Advances in Signal Processing. 2008 (11) (2008) ⟨http://www.hindawi.com/journals/asp/2008/280635.abs.html⟩.
[46] E. vanden Berg, M. Friedlander, Probing the Pareto frontier for basis pursuit solutions, SIAM Journal on Scientific Computing 31 (2) (2008) 890–912.
[47] W. Bajwa, J. Haupt, G. Raz, S. Wright, R. Nowak, Toeplitz-structured compressed sensing matrices, in: IEEE Workshop on Statistical Signal Processing (SSP), Madison, Wisconsin, 2007.
[48] A. Householder, The Theory of Matrices in Numerical Analysis, Blaisdell, New York, 1965.
[49] D. Donoho, M. Elad, Optimally sparse representation in general nonorthogonal dictionaries via l1 minimization, Proceedings of National Academy of Science 100 (2003) 2197–2202.
[50] R. Varga, Gersgorin and his Circles, Springer Verlag, 2004.

Please cite this article as: A. Eftekhari, et al., Two-dimensional random projection, Signal Process. (2011), doi:10.1016/j.sigpro.2011.01.002