

---

# An Analysis of Random Projections in Cancelable Biometrics

---

**Devansh Arpit**

SUNY Buffalo, Buffalo, NY, USA

DEVANSHA@BUFFALO.EDU

**Ifeoma Nwogu**

SUNY Buffalo, Buffalo, NY, USA

INWOGU@BUFFALO.EDU

**Gaurav Srivastava**

Samsung, USA

SRIVASTAVA.G@SAMSUNG.COM

**Venu Govindaraju**

SUNY Buffalo, Buffalo, NY, USA

GOVIND@BUFFALO.EDU

## Abstract

With increasing concerns about security, the need for highly secure physical biometrics-based authentication systems utilizing *cancelable biometric* technologies is on the rise. Because the problem of cancelable template generation deals with the trade-off between template security and matching performance, many state-of-the-art algorithms successful in generating high quality cancelable biometrics all have random projection as one of their early processing steps. This paper therefore presents a formal analysis of why random projections is an essential step in cancelable biometrics. By formally defining the notion of an *Independent Subspace Structure* for datasets, it can be shown that random projection preserves the subspace structure of data vectors generated from a union of independent linear subspaces. The bound on the minimum number of random vectors required for this to hold is also derived and is shown to depend logarithmically on the number of data samples, not only in independent subspaces but in disjoint subspace settings as well. The theoretical analysis presented is supported in detail with empirical results on real-world face recognition datasets.

## 1. Introduction

As physical biometrics-based authentication such as the use of fingerprints, faces, iris scans etc., has gained significant

popularity in last few decades, there is a growing need for *cancelable biometric* technologies. Cancelable biometrics refers to the systematic, intentional, repeatable distortion of biometrics features in order prevent the notion of “stolen biometrics”. A person’s biometrics are stolen for a specific modality, when the feature template used in assigning the biometric to that user is compromised by a masquerading attacker, thus giving the attacker access privileges to the user’s resources. Cancelable biometrics are especially important when there is a need to store biometric templates, because if compromised, it is virtually impossible for a user to regenerate the physical traits that were used in creating the templates during enrollment.

Thus, in an attempt to reduce the vulnerability of such security systems, there has been increased research activity in the areas of cancelable biometrics where the problem deals with the trade-off between template security and matching performance. The state-of-the-art algorithms that have been successful so far in generating high quality cancelable biometrics are all based on random projection (Feng et al., 2010; Teoh et al., 2006; Goh & Ngo, 2003). Of course, the random projection technique alone is not sufficient for generating highly secure and discriminating biometric templates, but is the first fundamental step which occurs before other more complex techniques, such as class-preserving transforms, template hashing, etc., are implemented.

With increasing technological advancements in computational speed and memory, and with increasing volumes of disparate data being collected for security purposes, more high dimensional feature vectors are being used in many

---

*Proceedings of the 31<sup>st</sup> International Conference on Machine Learning*, Beijing, China, 2014. JMLR: W&CP volume 32. Copyright 2014 by the author(s).

---

This work was partially funded by the National Science Foundation under the grant number CNS-1314803

biometrics-driven security systems. However, since computational time increases with dimensionality, real-life biometric systems (employing large volumes of high dimensional feature vectors) are highly susceptible to performance degradation over time. Dimensionality reduction techniques (such as PCA, LDA, LLE (Roweis & Saul, 2000), LPP (He & Niyogi, 2004)) can be employed to overcome this problem, however, for applications that perform tasks such as generating secure and discriminating biometric templates, where the subspace structure of the data should be preserved after dimensionality reduction, many of these techniques will fail. We foresee the use of random projections as a core component of future security systems using biometric modalities such as face recognition for authentication.

For this reason and more, in this paper, we formally define the notion of a *Independent Subspace Structure* for datasets, and based on this definition, we show that random projection preserves the subspace structure of data vectors generated from a union of independent linear subspaces. Thus the technique can be employed as a cancelable transform to project an original biometric template into a subspace and generate a new cancelable template, while maintaining discriminability. While an extensive number of papers in the literature has employed random projection for data dimensionality reduction for tasks such as k-means clustering (Boutsidis et al., 2010), classification (Balcan et al., 2004) (Shi et al., 2012) etc., these papers have shown that for the respective tasks, certain desired properties of the data vectors are preserved under random projection. However, to the best of our knowledge, a more general and formal analysis of linear subspace structure preservation under random projections has not been reported thus far; this is the main thrust for this paper.

## 2. Definitions

A linear subspace in  $\mathbb{R}^n$  of dimensions ( $d$ ) can be represented using a matrix  $B \in \mathbb{R}^{n \times d}$  where the columns of  $B$  form the support of the subspace. Then any vector in this subspace can be represented as  $x = Bw \forall w \in \mathbb{R}^d$ . Let there be  $K$  independent subspaces denoted by  $S_1, S_2, \dots, S_K$ . Any subspace  $S_i$  is said to be independent of all other subspaces if there does not exist any non-zero vector in  $S_i$  which is a linear combination of vectors in the other subspaces. Formally,

$$\sum_{i=1}^K S_i = \oplus_{i=1}^K S_i$$

where,  $\oplus$  denotes direct sum of subspaces.

While the above definition states the condition under which two or more subspaces are independent, it does not specifically tell us quantitatively how well they are separated and

this leads us to the definition of the margin between a pair of subspaces.

### Definition 1 (Subspace Margin)

Subspaces  $S_i$  and  $S_j$  are separated by margin  $\gamma_{ij}$  if

$$\gamma_{ij} = \max_{u \in S_i, v \in S_j} \frac{\langle u, v \rangle}{\|u\|_2 \|v\|_2} \quad (1)$$

Geometrically, the above definition says that margin between any two subspaces is defined as the maximum dot product between two unit vectors, one from either subspace. The vector pair  $u$  and  $v$  that maximize this dot product is known as the *principal vector pair* between the two subspaces while the angle between these vectors is called the *principal angle*. Notice that  $\gamma_{ij} \in [0, 1]$  such that  $\gamma_{ij} = 0$  implies that the subspaces are maximally separated while  $\gamma_{ij} = 1$  implies that the two subspaces are not independent.

Having defined these concepts, our goal is to learn a subspace from any given dataset that is sampled from a union of independent linear subspaces such that this independent subspace structure property is approximately preserved in the dataset. We will make this idea more concrete shortly.

Notice that the above definitions of independent subspaces and separation margin (definition 1) apply explicitly to well defined subspaces. So a natural question is: *How do we define these concepts for datasets?* We define the *Independent Subspace Structure* for a dataset as follows,

### Definition 2 (Independent Subspace Structure)

Let  $X = \{x_j\}_{j=1}^N$  be a  $K$  class dataset of  $N$  data vectors in  $\mathbb{R}^n$  and  $X_i \subset X$  ( $i \in \{1 \dots K\}$ ) such that data vectors in  $X_i$  belong to class  $i$ . Then we say that the dataset  $X$  has *Independent Subspace Structure* if each  $i^{\text{th}}$  class data  $x \in X_i$  is sampled from a linear subspace  $S_i$  ( $i \in \{1 \dots K\}$ ) in  $\mathbb{R}^n$  such that each subspace is independent.

Again, the above definition only specifies that data samples from different classes belong to independent subspaces. To estimate the margin between subspaces these, we define *Subspace Margin for datasets* as follows:

### Definition 3 (Subspace Margin for datasets)

For a dataset  $X$  with *Independent Subspace Structure*, class  $i$  ( $i \in \{1 \dots K\}$ ) data is separated from all the other classes with margin  $\gamma_i$ , if  $\forall x \in X_i$  and  $\forall y \in X \setminus \{X_i\}$ ,  $\frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2} \leq \gamma_i$ , where  $\gamma_i \in [0, 1]$ .

With these definitions, we will now make the idea of independent subspace structure preservation more concrete. Specifically, by subspace structure preservation, we refer to the case where we are originally given a set of data vectors sampled from a union of independent linear subspaces

and subsequently, after projection, the projected data vectors also belong to a union of independent linear subspaces.

Formally, let  $X$  be a  $K$  class dataset in  $\mathbb{R}^n$  with independent subspace structure such that class  $i$  samples ( $x \in X_i$ ) are drawn from subspace  $S_i$ , then the projected data vectors (using projection matrix  $P \in \mathbb{R}^{n \times m}$ ) in the sets  $\bar{X}_i := \{P^T x : x \in X_i\}$  for  $i \in \{1 \dots K\}$  are such that data vectors in each set  $\bar{X}_i$  belong to a linear subspace ( $\bar{S}_i$  in  $\mathbb{R}^m$ ) and the subspaces  $\bar{S}_i$  ( $i \in \{1 \dots K\}$ ) are independent, i.e.,  $\sum_{i=1}^K \bar{S}_i = \bigoplus_{i=1}^K \bar{S}_i$ .

### 3. Random Projections

Random Projection has gained significant popularity in recent years due to its low computational costs and the guarantees it comes with. Specifically, it has been shown in cases of linearly separable data (Shi et al., 2012) (Balcan et al., 2004) and data that lies on a low dimensional compact manifold (Baraniuk & Wakin, 2009) (Hegde et al., 2007), that random projection preserves the linear separability and manifold structure respectively, given that certain conditions are satisfied. Notice that a union of independent linear subspaces is a specific case of manifold structure and hence the results of random projection for manifold structure apply in general to our case. However, as those results are derived for a more general case, their results are weak when applied to our problem setting. Further, to the best of our knowledge, there has not been any prior analysis of random projection on the margin between independent subspaces.

The various applications of random projection for dimensionality reduction are rooted in the following version of the Johnson-Lindenstrauss (JL) lemma (Vempala, 2004):

**Lemma 4** For any vector  $x \in \mathbb{R}^n$ , matrix  $R \in \mathbb{R}^{m \times n}$  where each element of  $R$  is drawn i.i.d. from a standard Gaussian distribution,  $R_{ij} \sim \frac{1}{\sqrt{m}} \mathcal{N}(0, 1)$  and any  $\epsilon \in (0, 1/2)$

$$\Pr((1 - \epsilon)\|x\|^2 \leq \|Rx\|^2 \leq (1 + \epsilon)\|x\|^2) \geq 1 - 2 \exp\left(-\frac{m}{4}(\epsilon^2 - \epsilon^3)\right) \quad (2)$$

This lemma states that the  $\ell_2$  norm of any randomly projected vector is approximately equal to the  $\ell_2$  norm of the original vector. While conventionally, elements of the random matrix are generated from a Gaussian distribution, it has been proved (Achlioptas, 2003) (Li et al., 2006) that one can indeed use sparse random matrices (with most of the elements being zero with high probability) to achieve the same goal.

Aside, in relation to adopting random projection in the preliminary steps to providing template cancelability, if given a cancelable biometric template  $\bar{x} = Rx$  constructed from an original template  $x$  with the projection matrix  $R$ , and the

initial cancelable template  $\bar{x}$  is compromised, a new template  $\bar{x}' = R'x$  is issued with a new projection matrix  $R'$  as a replacement. Lemma 4 indicates that discriminability of the original feature vector is preserved for each template, however the conditions required for this still need to be investigated.

Before studying the conditions required for independent subspace structure preservation for a multiclass problem, we first state our cosine preservation lemma which simply states that the cosine of angle between any two fixed vectors is approximately preserved under random projection. A similar angle preservation theorem is stated in (Shi et al., 2012), but we will state the difference between the two after presenting the lemma.

**Lemma 5 (Cosine preservation)**

For all  $x, y \in \mathbb{R}^n$ , any  $\epsilon \in (0, 1/2)$  and matrix  $R \in \mathbb{R}^{m \times n}$  where each element of  $R$  is drawn i.i.d. from a standard Gaussian distribution,  $R_{ij} \sim \frac{1}{\sqrt{m}} \mathcal{N}(0, 1)$ , one of the following inequalities holds true

$$\begin{aligned} \frac{1}{(1 - \epsilon)} \frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2} - \frac{\epsilon}{1 - \epsilon} &\leq \frac{\langle Rx, Ry \rangle}{\|Rx\|_2 \|Ry\|_2} \\ &\leq \frac{1}{(1 + \epsilon)} \frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2} + \frac{\epsilon}{1 + \epsilon} \end{aligned} \quad (3)$$

if  $\frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2} < -\epsilon$ ,

$$\begin{aligned} \frac{1}{(1 - \epsilon)} \frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2} - \frac{\epsilon}{1 - \epsilon} &\leq \frac{\langle Rx, Ry \rangle}{\|Rx\|_2 \|Ry\|_2} \\ &\leq \frac{1}{(1 - \epsilon)} \frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2} + \frac{\epsilon}{1 - \epsilon} \end{aligned} \quad (4)$$

if  $-\epsilon \leq \frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2} < \epsilon$ , and

$$\begin{aligned} \frac{1}{(1 + \epsilon)} \frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2} - \frac{\epsilon}{1 + \epsilon} &\leq \frac{\langle Rx, Ry \rangle}{\|Rx\|_2 \|Ry\|_2} \\ &\leq \frac{1}{(1 - \epsilon)} \frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2} + \frac{\epsilon}{1 - \epsilon} \end{aligned} \quad (5)$$

if  $\frac{\langle x, y \rangle}{\|x\|_2 \|y\|_2} \geq \epsilon$ . Further the inequality holds true with probability at least  $1 - 8 \exp\left(-\frac{m}{4}(\epsilon^2 - \epsilon^3)\right)$ .

*Proof: See appendix.*

We would like to point out that cosine of both acute and obtuse angles are preserved under random projection as is evident from the above lemma. However, if the cosine value is close to zero, the additive error in the inequalities 3, 4 and 5 distorts the cosine significantly after projection. On the other hand, (Shi et al., 2012) in their paper state that obtuse angles are not preserved. As evidence, the authors empirically show cosines with negative value close to zero. However, as already stated, cosine values close to zero are

not well preserved. Hence this does not serve as an evidence that obtuse angles are not preserved under random projection which we show empirically otherwise to be true. Notice that this is not the case for the JL lemma (4) where the error is multiplicative and hence length of vectors are preserved to a good degree invariantly for all vectors.

In general, the inner product between vectors is not well preserved under random projection irrespective of the angle between the two vectors. This can be analyzed using Equation 12. Rewriting this equation in the following form, we have that,

$$\langle x, y \rangle - \epsilon \|x\|_2 \|y\|_2 \leq \langle Rx, Ry \rangle \leq \langle x, y \rangle + \epsilon \|x\|_2 \|y\|_2 \quad (6)$$

holds with high probability. Clearly, because the error term itself depends on the length of the vectors, inner product between arbitrary vectors after random projection is not well preserved. However, as a special case, inner product of vectors with length less than 1 is preserved (corollary 2 in (Arriaga & Vempala, 2006)) because the error term gets diminished in this case.

For ease of representation, in all further analysis, we will use Equation 5 while making use of the cosine preservation lemma. We will now go on to examine the conditions under which independent subspace structure can be preserved for any linearly separable dataset.

### 3.1. Subspace Margin Preservation

In order for independent subspace structure to be preserved for any dataset, we need two conditions to hold simultaneously. First, data sampled from each subspace should continue to belong to a linear subspace after projection. Second, the subspace margin for the dataset should be preserved.

**Remark 6** (*Individual Subspace preservation*)

Let  $X_i$  denote the set of data vectors ( $x$ ) drawn from the subspace  $S_i$ , and let  $R \in \mathbb{R}^{m \times n}$  denote the random projection matrix as defined before. Then after projection, all the vectors in  $X_i$  continue to lie along the linear subspace in the span of  $RB_i$ , where the columns of  $B_i$  denote the span of  $S_i$ .

The above straight forward remark states that the first requirement always holds true. Now we need to derive the condition needed for the second requirement to hold true.

**Theorem 7** (*Multiclass Subspace Preservation*) Let  $X = \{x_j\}_{j=1}^N$  be a  $K$  class dataset with Independent Subspace structure and the  $i^{th}$  class have margin  $\gamma_i$ . Then for any  $\epsilon \in (0, 1/2)$ , the subspace structure of the entire dataset is preserved after random projection using matrix  $R \in \mathbb{R}^{m \times n}$

( $R_{ij} \sim_{i.i.d.} \frac{1}{\sqrt{m}} \mathcal{N}(0, 1)$ ) with margin  $\bar{\gamma}_i$  for class  $i$  as follows

$$Pr \left( \bar{\gamma}_i \leq \frac{1}{(1-\epsilon)} \gamma_i + \frac{\epsilon}{1-\epsilon}, \forall i \in \{1 \dots K\} \right) \geq 1 - 6N^2 \exp \left( -\frac{m}{4} (\epsilon^2 - \epsilon^3) \right) \quad (7)$$

*Proof: See appendix.*

Recall from our discussions on the cosine preservation lemma (5), that cosine values close to zero are not well preserved under random projection. However, from our above error bound on the margin (eq 7), it turns out that this is not a problem - two subspaces separated with a margin close to zero implies that the principal angle between them is almost orthogonal, i.e., they are maximally separated. Therefore, under these circumstances, the projected subspaces are also well separated.

Formally, let  $\gamma = 0$ , so that after projection,  $\bar{\gamma} \leq \frac{\epsilon}{1-\epsilon}$  is further upper bounded by 1 as  $\epsilon$  tends to 0.5. In practice we set  $\epsilon$  to be a much smaller quantity, hence  $\bar{\gamma}$  is well below 1.

While the analysis so far only relates to structure preservation for datasets with independent subspace structure, it is not hard to see that the same bounds also apply to datasets with disjoint subspace structure, i.e., each subspace (class) is pairwise disjoint with each other but not independent overall.

## 4. Sparse Representation based Recognition

Sparse representation (SR) has been widely used for classification purposes in various machine learning applications, including face recognition tasks in biometric security applications. The idea of SR is based on the theory of compressed sensing. This theory claims that if a system of linear equations with an overcomplete dictionary has a sparse solution then it can be achieved by solving the basis pursuit algorithm:

$$w^* = \arg \min_w \|w\|_1 \quad s.t. \quad y = Dw \quad (8)$$

where  $y \in \mathbb{R}^m$  is the measurement vector,  $D \in \mathbb{R}^{m \times n}$  is overcomplete dictionary and  $w \in \mathbb{R}^n$  is the variable for which we want a sparse solution. This property is very useful for classification because one can use all the training samples as the columns of the overcomplete dictionary  $D$ , test sample as  $y$  and solve the above optimization to obtain the sparse reconstruction coefficient  $w^*$  over the training samples. The advantage of representing a test sample as a sparse linear combination of the training samples is that fewer non-zero coefficients over the training samples will

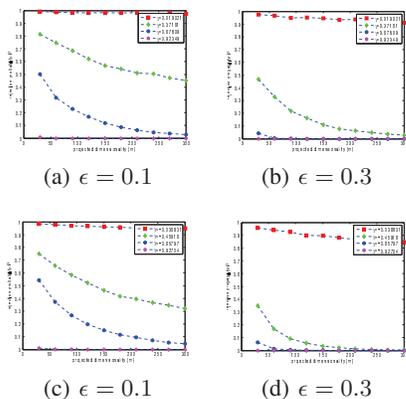


Figure 1. Empirical rejection probability for cosine preservation of acute (a,b) and obtuse angles (c,d). See section 5.1.1 for details.

be more discriminative in terms of the class of the test sample.

More recently, Sparse Subspace clustering (SSC) has been used for subspace clustering applications. The subspace clustering domain assumes that each individual class lies along a linear independent subspace and under this assumption we want to cluster a given set of data samples such that each cluster corresponds to samples from one such subspace. The authors of SSC approach (Elhamifar & Vidal, 2009) show that, the basis pursuit optimization guarantees the correct reconstruction of a test sample ( $y$ ) using an over-complete dictionary of training samples ( $D$ ). Formally this is stated in the following theorem:

**Theorem 8** (Theorem 1 in (Elhamifar & Vidal, 2009))

Let  $D \in \mathbb{R}^{m \times n}$  be a matrix whose columns are drawn from a union of  $K$  independent linear subspaces. Assume that the points within each subspace are in general position. Let  $y$  be a new point in subspace  $i$ . The solution to the  $\ell^1$  problem in 8,  $w^* \in \mathbb{R}^n$  is sparse such that  $w_j \neq 0$  iff  $D_j$  belongs to the  $i^{th}$  subspace and  $w_j=0$  otherwise.

where  $D_j$  denotes the  $j^{th}$  column of matrix  $D$ . This theorem gives us the sufficient condition under which one is guaranteed to recover the correct coefficients for a given test sample using SR. This property is used in the SSC algorithm for clustering. However, this also clearly shows why it makes sense to use sparse representation for the task of classification under the assumption that our classes lie along independent linear subspaces. This assumption is widely used for applications like face recognition and motion segmentation.

Since the above algorithms make use of the underlying subspace assumption for datasets, it is natural to investigate if there exists a dimensionality reduction method that is guar-

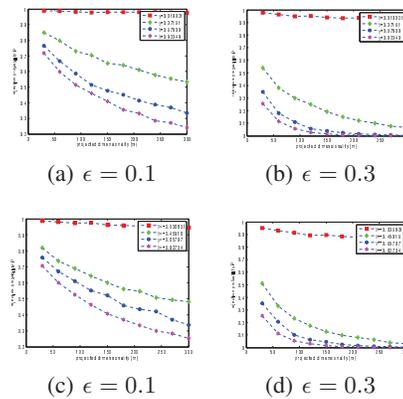


Figure 2. Empirical rejection probability for inner product preservation of acute (a,b) and obtuse angles (c,d). See section 5.1.2 for details.

anteed to preserve this structure in the dataset. If so, we can apply the aforementioned algorithms in a much smaller feature space without losing accuracy while simultaneously being much faster.

In the preceding section, we showed that random projection preserves the underlying structure in datasets and thus can be effectively used for dimensionality reduction. Notice that the advantage of random projections is three fold: it allows for the classification/recognition algorithm to run faster; (ii) it is extremely inexpensive to compute; and (iii) it yields classification results with accuracies at par with that in the original dimensions of the data. While most dimensionality reduction algorithms are expensive in terms of computing the projection vectors (e.g. PCA takes cubic time in the size of feature space), random projection needs each element of its projection vectors to be sampled randomly independent of the data at hand. This non-adaptive nature of random projection makes it a very powerful dimensionality reduction tool. These qualities indicate why random projections is becoming such an essential technique for developing very efficient and highly secure biometric applications.

## 5. Empirical Analysis

In this section, we present empirical evidence to support our theoretical analysis (from Section 3) of why random projections work for cancelable biometrics. We perform experiments to show both cosine preservation and subspace structure preservation under random projections, using different face recognition datasets.

### 5.1. Cosine and Inner product Preservation

#### 5.1.1. COSINE PRESERVATION

In lemma 5, we concluded that the cosine of the angle between any two vectors remains preserved under random

Table 1. Time taken (in msec) for dimensionality reduction on Extended Yale dataset B. See section 5.2 for details.

method/dimensions	100	150	300	500
RP	1.59	2.76	11.11	25.78
PCA	322.04	328.58	332.64	341.10

Table 2. Classification accuracy on Extended Yale dataset B. See section 5.2 for details.

method/dimensions	100	150	300	500
RP	<b>94.61</b>	<b>95.85</b>	<b>96.53</b>	<b>97.93</b>
PCA	94.27	94.95	95.94	96.86

projection irrespective of the angle being acute or obtuse. However, we also stated that cosine values close to zero are not well preserved. Here, we perform empirical analysis on vectors with varying angles (both acute and obtuse) and arbitrary length to verify the same. In order to achieve this, we use settings similar to (Shi et al., 2012). We generate 2000 random projection matrices  $R_i \in \mathbb{R}^{m \times n}$  ( $i = 1$  to 2000) where we vary  $m = \{30, 60, \dots, 300\}$  and  $n = 300$  is the dimension of the original space. We define empirical rejection probability for cosine preservation similar to (Shi et al., 2012) as,

$$\hat{P} = 1 - \frac{1}{2000} \sum_{i=1}^{2000} \mathbf{1}((1 - \epsilon) \leq \frac{\langle R_i x, R_i y \rangle \|x\|_2 \|y\|_2}{\|R_i x\|_2 \|R_i y\|_2 \langle x, y \rangle}) \leq (1 + \epsilon)$$

where we vary  $\epsilon \in \{0.1, 0.3\}$  and  $\mathbf{1}(\cdot)$  is the indicator operator.

For acute angle, we randomly generate vectors  $x$  and  $y$  of arbitrary length but with fixed cosine values  $\gamma = \{0.019021, 0.37161, 0.67809, 0.92349\}$ . For obtuse angle, we similarly generate vectors  $x$  and  $y$  with fixed cosine values  $\gamma = \{-0.036831, -0.45916, -0.65797, -0.92704\}$ . We then compute the empirical rejection probability as mentioned above for different values of  $\epsilon$ . Figure 1 shows the results on these vectors. In the figure, notice that the rejection probability decreases as the absolute value of cosine of the angle ( $\gamma$ ) increases (from 0 to 1), as well as for higher value of  $\epsilon$ . Notice, for cosine values close to zero, the rejection probability is close to 1 even at high dimensions. These results corroborate with our theoretical analysis in lemma 5.

### 5.1.2. INNER PRODUCT UNDER RANDOM PROJECTION

We use the same experimental setting as in section 5.1.1. We define the empirical rejection probability of inner product similar to (Shi et al., 2012) as

$$\hat{P} = 1 - \frac{1}{2000} \sum_{i=1}^{2000} \mathbf{1}((1 - \epsilon) \leq \frac{\langle R_i x, R_i y \rangle}{\langle x, y \rangle}) \leq (1 + \epsilon)$$

We use the same vectors as in 5.1.1 for experiments in this

Table 3. Time taken (in msec) for dimensionality reduction on PIE dataset. See section 5.2 for details.

method/dimensions	30	50	70	100
RP	0.35	0.60	0.73	1.3
PCA	317.3	315.9	318.8	319.6

Table 4. Classification accuracy on PIE dataset. See section 5.2 for details.

method/dimensions	30	50	70	100
RP	96.67	<b>97.45</b>	<b>98.04</b>	<b>98.04</b>
PCA	<b>97.06</b>	97.45	97.45	97.25

section. We then compute the empirical rejection probability as mentioned above for different values of  $\epsilon$ . Figure 2 shows the results on these vectors. As is evident from the figure, inner product between vectors is not well preserved (even when cosine values are close to 1). This result is in line with our theoretical bound in equation 6 as the vector lengths in our experiment are arbitrarily greater than 1.

### 5.1.3. REQUIRED NUMBER OF RANDOM VECTORS

We study the number of random vectors required for subspace preservation by varying different parameters. The lower bound on the number of random vectors required for theorem 7 to hold is given by,

$$m \geq \frac{8}{(\epsilon^2 - \epsilon^3)} \ln \frac{\sqrt{6}N}{1 - \delta} \quad (9)$$

It can be seen that for  $N = 1000$  and  $\epsilon = 0.15$ , random projection to lower dimensions is effective only if  $m > 6000$  while for  $\epsilon = 0.4$ ,  $m > 900$  suffices. The choice of  $\epsilon$  depends on the robustness of the algorithm (for the respective task) towards noise and is a trade-off between noise (allowed) and the number of random vectors ( $m$ ) required.

## 5.2. Subspace Structure preservation

In this section, our goal is to show that random projections achieve accuracy better or at least at par with the most widely used dimensionality reduction technique (PCA). We report comparative analysis on the accuracy results and performance times between random projections and PCA. We selected PCA alone for detailed analysis mainly because we found the performance of the other nonlinear dimensionality reduction techniques to be significantly less than the two techniques. Testing on the Extended Yale dataset B (described below), we initially used LPP (Locality Preserving Projections), NPE (Neighborhood Preserving Embedding) (He et al., 2005), and Laplacian Eigenmaps (Belkin & Niyogi, 2003) to reduce the data to 150 dimensions. The best performing of these reduction techniques yielded a result of only 73% compared to the close to 96% accuracies resulting from random projections and PCA. In fairness, these other techniques make no claim to preserving the original subspace structure of the data, rather they pre-

serve some general manifold structure, and do not necessarily guarantee subspace separability.

With this intent of showing that random projections achieve accuracy better or at least at par with PCA, we use sparse representation based classification (SRC, (Wright et al., 2009)) technique that exploits the subspace structure in the data. One can always use a better classification algorithm that exploits this structure to achieve higher accuracy. However, our aim here is not to compare different classification algorithms but to show that random projection is a computationally inexpensive dimensionality reduction tool with performance guarantees supported by our theoretical analysis.

Cancelable biometrics on face templates is our testbed-of-choice because it is generally assumed that face images with illumination variation lie along linear independent subspaces (Shakhnarovich & Moghaddam, 2004). We use the following datasets for evaluation:

1. *Extended Yale dataset B* (Georghiades et al., 2001): It consists of  $\sim 2414$  frontal face images of 38 individuals ( $K = 38$ ) with 64 images per person. These images were taken under constrained but varying illumination conditions. We crop all the images to  $32 \times 32$  and concatenate all the pixel intensity to form our feature vectors. We use a 50% – 50% train-test split for evaluation.

2. *PIE dataset* (Sim et al., 2002): The CMU pose, illumination, and expression (PIE) database consists of 41368 images of 68 people ( $K = 68$ ) under 13 different poses, 43 illumination conditions and 4 different expressions. However, we utilize only the first 10 classes from this dataset with 70% – 30% train-test split for evaluation. We cropped to size  $32 \times 32$  pixels. The pixel intensities are concatenated to form the feature vectors.

We perform two types of experiments. First, we compare the time taken for dimensionality reduction by PCA and random projections for both datasets. This time is the sum of the time taken by either algorithm to compute its projection vectors and then to project the entire dataset down to these projection vectors. The results are shown in Table 1 and 3 for the Extended Yale B dataset and PIE dataset respectively. The results show that random projections is faster than PCA by at least an order of 10 times.

Secondly, we show classification accuracies on both the datasets after dimensionality reduction. These results are shown in Table 2 and 4 for Extended Yale B dataset and PIE dataset respectively. Clearly, random projections performs better than PCA while being significantly faster.

These results substantiate our claim that random projections preserve the subspace structure of any given dataset. Also notice that even a very low number of random vectors

used for projection yields good accuracy. This observation can be explained using Lemma 10 of (Sarlós, 2006) where the authors show that if a given data lies along a  $d$  dimensional subspace then one only needs  $O(d \log d)$  random vectors. In most real applications the value of  $d$  is usually low, i.e., classes usually lie along a low dimensional subspace. Thus it is not surprising that even small number of random vectors yield high accuracy.

## 6. Discussion

A major advantage of random projections occurs for streaming data where  $N$  is constantly changing. Also, as long as the data lies in a  $d$ -dimensional subspace, as stated in Lemma 10 of (Sarlós, 2006),  $O(d \log d)$  random projection vectors preserve the length of all the vectors in that subspace, hence our structure preservation results still hold true. Thus our results not only hold true for a fixed size dataset, but also for an infinite stream of data vectors, as long as a sufficient (but finite and small,  $O(d \log d)$ ) number of random vectors are used and the underlying data structure remains the same.

As originally stated in Section 1, the random projections technique by itself is not a complete solution to generating highly secure and discriminating biometric templates. Although the random projection step is secure against the brute-force attack because original templates are often real-valued and high-dimensional, if the projection matrix is not well protected, an attacker could construct its pseudo-inverse to recover an approximation to the original data. Nevertheless, with the advantages of random projections namely: allowing for the classification/recognition algorithm to run faster; (ii) being extremely inexpensive to compute; and (iii) yielding classification results with accuracies at par with that in the original dimensions of the data, random projections is quickly becoming an essential early-step technique in the development of very efficient and highly secure biometric applications.

## 7. Conclusion

In this paper, we presented a formal analysis of why random projections are an essential initial step for generating cancelable biometrics, especially in a real-life scenario where security, discriminability and cancelability are required. Using random projections for dimensionality reduction ensures that the independent subspace structure of datasets are preserved. We derived the bound on the minimum number of random vectors required for this to hold (Section 5.1.3) and concluded that this number depends logarithmically on the number of data samples. All the above arguments hold under *disjoint subspace* settings as well. As a side analysis, we also showed that while cosine values (lemma 5) are preserved under random projection for

both acute and obtuse angles, inner product (equation 6) between vectors are not well preserved in general.

Although we describe our work in the context of cancelable biometrics, the discussion and evaluations presented is a detailed analysis of linear subspace structure preservation under random projections, irrespective of the task-at-hand.

## References

- Achlioptas, Dimitris. Database-friendly random projections: Johnson-lindenstrauss with binary coins. *J. Comput. Syst. Sci.*, 66(4):671–687, June 2003.
- Arriaga, Rosa I. and Vempala, Santosh. An algorithmic theory of learning: Robust concepts and random projection. *Machine Learning*, 63:161–182, 2006.
- Balcan, Maria-Florina, Blum, Avrim, and Vempala, Santosh. Kernels as features: On kernels, margins, and low-dimensional mappings. In *In 15th International Conference on Algorithmic Learning Theory (ALT 04)*, pp. 79–94, 2004.
- Baraniuk, Richard G. and Wakin, Michael B. Random projections of smooth manifolds. *Foundations of Computational Mathematics*, 9:51–77, 2009.
- Belkin, Mikhail and Niyogi, Partha. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.*, 15(6):1373–1396, June 2003. ISSN 0899-7667.
- Boutsidis, Christos, Zouzias, Anastasios, and Drineas, Petros. Random projections for  $k$ -means clustering. *CoRR*, abs/1011.4632, 2010.
- Elhamifar, Ehsan and Vidal, René. Sparse subspace clustering. In *CVPR*, pp. 2790–2797, 2009.
- Feng, Yi Cheng, Yuen, Pong Chi, and Jain, Anil K. A hybrid approach for generating secure and discriminating face template. *IEEE Transactions on Information Forensics and Security*, 5(1):103–117, 2010.
- Georghiades, A.S., Belhumeur, P.N., and Kriegman, D.J. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23(6):643–660, 2001.
- Goh, Alwyn and Ngo, David C.L. Computation of cryptographic keys from face biometrics. In *Communications and Multimedia Security. Advanced Techniques for Network and Data Protection*, volume 2828 of *Lecture Notes in Computer Science*, pp. 1–13. Springer Berlin Heidelberg, 2003.
- He, X. and Niyogi, P. Locality preserving projections. *Proc. of the NIPS, Advances in Neural Information Processing Systems. Vancouver: MIT Press*, 103, 2004.
- He, Xiaofei, Cai, Deng, Yan, Shuicheng, and Zhang, Hong-Jiang. Neighborhood preserving embedding. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pp. 1208–1213 Vol. 2, Oct 2005.
- Hegde, Chinmay, Wakin, Michael B., and Baraniuk, Richard G. Random projections for manifold learning. In Platt, John C., Koller, Daphne, Singer, Yoram, and Roweis, Sam T. (eds.), *NIPS*. Curran Associates, Inc., 2007.
- Li, Ping, Hastie, Trevor J., and Church, Kenneth W. Very sparse random projections. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, KDD '06*, pp. 287–296, New York, NY, USA, 2006. ACM.
- Roweis, Sam T. and Saul, Lawrence K. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, December 2000.
- Sarlós, Tamás. Improved approximation algorithms for large matrices via random projections. In *FOCS*, pp. 143–152. IEEE Computer Society, 2006.
- Shakhnarovich, Gregory and Moghaddam, Baback. Face recognition in subspaces. In *In: S.Z. LI, A.K. Jain, Handbook of Face Recognition*, pp. 141–168. Springer, 2004.
- Shi, Qinfeng, Shen, Chunhua, Hill, Rhys, and van den Hengel, Anton. Is margin preserved after random projection? *CoRR*, abs/1206.4651, 2012.
- Sim, Terence, Baker, Simon, and Bsat, Maan. The cmu pose, illumination, and expression (pie) database. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pp. 46–51. IEEE, 2002.
- Teoh, A.B.J., Goh, A., and Ngo, D.C.L. Random multispace quantization as an analytic mechanism for biohashing of biometric and random identity inputs. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 28(12):1892–1901, Dec 2006.
- Vempala, S. *The Random Projection Method*. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 2004.
- Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., and Ma, Yi. Robust face recognition via sparse representation. *IEEE TPAMI*, 31(2):210–227, Feb. 2009.