Abstract—This paper puts forth projections designs for compressive classification of Gaussian mixture models. In particular, we capitalize on the asymptotic characterization of the behavior of an (upper bound to the) misclassification probability associated with the optimal Maximum-A-Posteriori (MAP) classifier, which depends on quantities that are dual to the concepts of the diversity gain and coding gain in multi-antenna communications, to construct measurement designs that maximize the diversity-order of the measurement model. Numerical results demonstrate that the new measurement designs substantially outperform random measurements. Overall, the analysis and the designs cast geometrical insight about the mechanics of compressive classification problems.

Index Terms—Compressed Sensing, Compressive Classification

I. INTRODUCTION

Classification of high dimensional signals is fundamental to the broad fields of signal processing and machine learning. The aim is to increase speed and reliability while reducing the complexity of discrimination. An emerging paradigm that has attracted a great deal of recent interest is Compressed Sensing (CS) which seeks to capture important attributes of high-dimensional sparse signals from a small set of linear projections. The observation [1], [2] that captured the imagination of the signal processing community is that it is possible to guarantee fidelity of reconstruction from random linear projections when the source signal exhibits sparsity with respect to some dictionary.

The focus of CS has been predominantly on signal reconstruction rather than classification. However, we are interested in classification rather than estimation, in problems such as hypothesis testing, pattern recognition and anomaly detection that can be viewed as instances of signal classification. It is also natural to employ compressive measurement here since it may be possible to discriminate between signal classes using only partial information about the source signal. The challenge now becomes that of designing measurements that ignore signal features with little discriminative power.

Compressive classification appears in the machine learning literature as feature extraction or supervised dimensionality reduction. Approaches based on geometrical characterizations of the source have been developed, some like linear discriminant analysis (LDA) and principal component analysis (PCA) just depending on second order statistics. Approaches based on higher-order statistics of the source have also been developed [3]–[9].

The common approach in various CS problems [1], [2], is to take the measurement matrix to be such that its elements are drawn independently from a zero-mean Gaussian distribution with a certain fixed variance. This approach was also proposed in [10], in order to study the performance of compressive classification via a "wireless communications-inspired" characterization of the misclassification probability that embodies notions such as the diversity gain and the measurement gain and their relations to the geometry of the classes and the number of measurements.

In this paper, we capitalize on the characterization in [10] to address the problem of measurement matrix design for compressive sensing applications: we offer an insightful analysis that reveals how measurement designs are able to extract salient discriminative features associated with a collection of classes. We also show that such a measurement strategy naturally outperforms the standard random one associated with the conventional CS paradigm.

For reason of space, we relegate the mathematical proofs of our results to an upcoming journal paper [11].

II. THE COMPRESSIVE CLASSIFICATION PROBLEM

We consider a classification problem in the presence of compressive and noisy measurements. In particular, we use the standard measurement model given by:

\[ y = \Phi x + n \]  

where \( y \in \mathbb{R}^M \) represents the measurement vector, \( x \in \mathbb{R}^N \) represents the source vector, \( \Phi \in \mathbb{R}^{M \times N} \) represents the measurement matrix and \( n \sim \mathcal{N}(0, \sigma^2 \cdot I) \in \mathbb{R}^M \) represents standard white Gaussian noise.

We take the source signal to follow a well-known Gaussian mixture model (GMM), which has been shown to lead to state-of-the-art results in various classification applications including hyper-spectral imaging classification and digit recognition [8]. This model assumes that the source signal is drawn from one out of \( L \) classes \( C_i, i = 1, \ldots, L \), with probability \( P_i, i = 1, \ldots, L \), and that the distribution of the source conditioned on \( C_i \) is Gaussian with mean \( \mu_i \in \mathbb{R}^N \) and (possibly rank-deficient) covariance matrix \( \Sigma_i \in \mathbb{R}^{N \times N} \).

The objective is to produce an estimate of the true signal class given the measurement vector. The Maximum-A-Posteriori (MAP) classifier, which minimizes the probability of misclassification [12], produces the estimate given by:
\[
\hat{C} = \arg \max_{C_i} P(C_i \mid y) = \arg \max_{C_i} p(y \mid C_i) P_i.
\] (2)

where \(P(C_i \mid y)\) represents the \textit{a posteriori} probability of class \(C_i\) given the measurement vector \(y\) and \(p(y \mid C_i)\) represents the probability density function of the measurement vector \(y\) given the class \(C_i\).

The performance characterization in [10] – in line with the standard practice in multiple-antenna communications systems [13], [14] – is based on an upper bound to the probability of misclassification of the MAP classifier \(P_{err}^{UB}\), rather than the exact probability of misclassification \(P_{err}\). The characterization in [10] is also based on two fundamental metrics that describe the asymptotics of the upper bound to the probability of misclassification in the low-noise regime, which is relevant to various emerging classification tasks [8]. In particular, we define the diversity-order of the measurement model in (1) as:

\[
d = \lim_{\sigma^2 \to 0} \frac{\log P_{err}^{UB}(\sigma^2)}{\log \sigma^2},
\] (3)

that determines how (the upper bound to) the misclassification probability decays (in the \(\log \sigma^2\) scale) at low noise levels [15], [16]. We also define the measurement gain of the measurement model in (1) as:

\[
g_m = \lim_{\sigma^2 \to 0} \sigma^2 \cdot \frac{1}{\sqrt{P_{err}^{UB}(\sigma^2)}},
\] (4)

that determines the power offset of (the upper bound to) the misclassification error probability at low noise levels: note that the measurement gain refines the asymptotic description of the upper bound to the misclassification probability, by distinguishing further characteristics that exhibit identical diversity gain. These quantities admit a counterpart in multiple-antenna communications – for example, the measurement gain corresponds to the standard coding gain.

It will be useful to define various quantities, which relate to the geometry of the measurement model, that determine the behavior of the performance measures in (3) and (4) (see also [10]):

- \(r_i = \text{rank}(\Phi \Sigma_i \Phi^T)\) and \(v_i = \text{pdet}(\Phi \Sigma_i \Phi^T), \ i = 1, \ldots, L\), which measure the dimension and volume of the sub-space spanned by the linear transformation of the signals in class \(C_i\); and \(r_{ij} = \text{rank}(\Phi \Sigma_i + \Sigma_j \Phi^T)\) and \(v_{ij} = \text{pdet}(\Phi \Sigma_i + \Sigma_j \Phi^T), i, j = 1, \ldots, L, i \neq j\), which measure the dimension and volume of the direct sum of sub-spaces spanned by the linear transformation of the signals in classes \(C_i\) or \(C_j\).

- \(r_{\Sigma_i} = \text{rank}(\Sigma_i), \ i = 1, \ldots, L\), which relates to the dimension of the sub-space spanned by source signals in \(C_i\) and \(r_{\Sigma_{ij}} = \text{rank}(\Sigma_i + \Sigma_j), i, j = 1, \ldots, L, i \neq j\), which relates to the dimension of the direct sum of sub-spaces spanned by source signals in \(C_i\) or \(C_j\).

It will also be useful to define the quantity:

\[
NO_{Dim} = r_{\Sigma_{ij}} - \left[ (r_{\Sigma_i} + r_{\Sigma_j}) - r_{\Sigma_{ij}} \right]
\] (5)

that relates to the number of non-overlapping dimensions between the sub-spaces spanned by the eigenvectors of the covariance matrices \(\Sigma_i\) and \(\Sigma_j\). This number represents the sum of the two differences between the dimension of each linear space corresponding to one of the two classes and the dimension of the intersection of such two spaces.

The focus is on the design of a measurement matrix \(\Phi \in \mathbb{R}^{M \times N}\) that maximizes the diversity-order \(d\) for the classification problem, i.e.:

\[
\max_{\Phi} d(\Phi),
\] (6)

where we express the diversity as a function of the measurement matrix, subject to an available measurement budget:

\[
\text{rank}(\Phi) \leq M
\] (7)

In the sequel we consider measurement designs for a two class scenario, proposing solutions for the problem in (6) – (7). We also put forth algorithmic approaches in order to maximize the diversity-order for the multiple class scenario, which do not take into account the measurement budget.

III. TWO-CLASS CASE

The upper bound to the probability of misclassification in a two class classification problem can be defined via the Bhattacharyya bound [10]:

\[
P_{err}^{UB} = \sqrt{P_1 P_2} e^{-K_{ij}},
\] (8)

\[
K_{ij} = \frac{1}{8} \left[ \Phi \left( \mu_i - \mu_j \right) \right]^T \left[ \frac{\Phi (\Sigma_i + \Sigma_j) \Phi^T}{2} + 2\sigma^2 I \right]^{-1} \left[ \Phi \left( \mu_i - \mu_j \right) \right] \det \left( \frac{\Phi (\Sigma_i + \Sigma_j) \Phi^T + 2\sigma^2 I}{2} \right)
+ \frac{1}{2} \log \frac{\det(\Phi \Sigma_i \Phi^T + \sigma^2 I)}{\det(\Phi \Sigma_j \Phi^T + \sigma^2 I)}.
\] (9)

This bound leads immediately to characterizations of the upper bound to the misclassification probability defined via (3) and (4) both for zero-mean classes classification problems and non-zero-mean classes problems [10], that are the basis for the ensuing projections designs.

A. Zero-Mean Classes

The following Theorem considers measurement designs in the case of two zero mean classes.

\textit{Theorem 1:} Consider the measurement model in (1) where \(x \sim N(0, \Sigma_1)\) with probability \(P_1\) and \(x \sim N(0, \Sigma_2)\) with probability \(P_2 = 1 - P_1\). Assume that the measurement budget is such that \(M \geq NO_{Dim}\). Then, the maximum achievable diversity-order is given by:

\[
d_{max} = \frac{1}{4} NO_{Dim}
\] (10)

which is achieved by a measurement matrix design that obeys the following necessary and sufficient condition:

\[
2r_{12} - r_1 - r_2 = NO_{Dim}
\] (11)

One possible construction of the measurement matrix \(\Phi\) that achieves the maximum diversity-order is such that:

\[
\Phi = \left[ v_1, \ldots, v_{n_{\Sigma_1}}, w_1, \ldots, w_{n_{\Sigma_2}} \right]^T
\] (12)

where the set of vectors \([u_1, \ldots, u_{n_{\Sigma_1}}]_i \in \mathbb{R}^N, \ [v_1, \ldots, v_{n_{\Sigma_2}}]_i \in \mathbb{R}^N\) constitute an orthonormal basis of the linear spaces \(\text{Null} (\Sigma_1) \cap \text{Null} (\Sigma_2), \ \text{Null} (\Sigma_1)\) and \(\text{Null} (\Sigma_2)\), respectively.
Assume now that the measurement budget is such that \( M < NO_{\text{Dim}} \). Then, the maximum achievable diversity-order is given by:

\[
d = d_{\max} = \left( \frac{NO_{\text{Dim}} - M}{4} \right) = \frac{1}{4} M
\]  

which is achieved by a measurement matrix design such that \( r_{12} = M \) and \( r_1 + r_2 = r_{12} \). The measurement matrix \( \Phi \) that achieves such a diversity order can be obtained from the measurement matrix \( \Phi \) in (12), by taking only \( M \) vectors.

Note that, as expected, an insufficient measurement budget penalizes the maximum achievable diversity-order in a two class compressive classification problem via optimized projections: of particular relevance, the very maximum diversity order in (10) requires a measurement construction that preserves the number of non-overlapping dimensions between the two classes in (5). Note also that the measurement designs disclose the salient discriminative features associated with each individual class, which are associated with linear subspaces that are contained in the spaces spanned by the individual classes but are not contained in their intersection.

### B. Nonzero-mean classes

The next Theorem addresses the projection matrix design in the case of two non-zero mean classes.

**Theorem 2**: Consider the measurement model in (1) where \( x \sim N(\mu_1, \Sigma_1) \) with probability \( P_1 \) and \( x \sim N(\mu_2, \Sigma_2) \) with probability \( P_2 = 1 - P_1 \) and \( \mu_1 \neq \mu_2 \). Assume that:

\[
(\mu_1 - \mu_2) \notin \text{im}(\Sigma_1 + \Sigma_2),
\]  

Then, the maximum diversity-order is \( d = \infty \) and a matrix design that achieves such a diversity-order is:

\[
\Phi = [\phi_1 \ldots \phi_M]^T
\]  

where \( \phi \in \text{Null}(\Sigma_1 + \Sigma_2) \).

Assume now that:

\[
(\mu_1 - \mu_2) \in \text{im}(\Sigma_1 + \Sigma_2),
\]  

Then, the maximum diversity-order is given by (10) or (13), depending on the number of available measurements and a matrix design that achieves such a diversity-order is given by Theorem 1.

The most important feature about this Theorem is that under certain conditions associated with the geometry of the classification problem it is possible to attain a diversity-order \( d = \infty \), which entails exponential decay of the error probability, by taking a single designed measurement. This feature is unique to non-zero mean classes.

### IV. MULTIPLE-CLASSES CASE

The upper bound to the probability of misclassification in a multiple class classification problem, where \( L \geq 3 \), can be defined via the union bound together with the Bhattacharyya bound [10]:

\[
P_{\text{err, Mult}}^{UB} = \sum_{i=1}^{L} \sum_{j=1}^{L} P_i e^{-K_{ij}}
\]

where \( K_{ij} \) is given by (9).

This bound also leads immediately to characterizations defined via (3) and (4) both for zero-mean classes classification problems and non-zero-mean classes classification problems.

Therefore, we argue that the diversity-order of the multiple-class misclassification probability corresponds to the lowest of the diversity-orders of the pairwise classification problems: this offers immediately a route to carry out projections designs via the results encapsulated in the previous Theorems. We, thus, propose algorithmic approaches, inspired in the two-class case designs to achieve the maximum possible diversity-order: note that these algorithms do not place a direct restriction on the number of measurements.

#### A. Zero Mean Classes

Table I puts forth an algorithmic approach to design projections for multiple class classification problems with zero mean classes.

<table>
<thead>
<tr>
<th>Step 1</th>
<th>Identify the maximum diversity-order associated with the “worst” pair of classes, i.e.:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[ d_{\max} = \min_{i \neq j} \left[ \frac{1}{4} NO_{\text{Dim}}(C_i, C_j) \right] ]</td>
</tr>
<tr>
<td></td>
<td>where ( NO_{\text{Dim}}(C_i, C_j) ) represents the number of non-overlapping dimensions associated with classes ( C_i ) and ( C_j ).</td>
</tr>
</tbody>
</table>

| Step 2 | Construct \( \Phi \) such as in Theorem 1, for the “worst” pair of classes \( (C_i, C_j) \). |

| Step 3 | For every other pair of classes, pick \( n_{\Sigma_1} + n_{\Sigma_2} \) vectors accordingly with Theorem 1; check which of these vectors are linear combinations of the ones already present in matrix \( \Phi \) and discard them. Add the remaining vectors to matrix \( \Phi \), in order to achieve \( d_{\max} \). |

#### B. Non-Zero Mean Classes

Table II now puts forth an algorithmic approach to design projections for multiple class classification problems with non-zero mean classes.

| Step 1 | For each pair of classes \( (C_i, C_j) \), \( i = 1, \ldots, L \); \( j = 1, \ldots, L \); \( i \neq j \), pick a vector \( \phi \in \text{Null}(\Sigma_i + \Sigma_j) \). |

| Step 2 | If this vector is a linear combination of the vectors already present in the matrix \( \Phi \), discard it. Otherwise add the vector to the measurement matrix \( \Phi \). |

Otherwise, if \( (\mu_i - \mu_j) \in \text{im}(\Sigma_i + \Sigma_j) \), for some pair of classes \( (i, j) \), the maximum diversity-order is finite and a matrix design that achieves the maximum diversity-order follows from the algorithmic approach in section IV-A.

### V. RESULTS AND CONCLUSIONS

We now compare the behavior of both the true misclassification probability and its upper bound for random measurements and designed measurements.
We focus on a two-class compressive classification problem to illustrate as simply as possible the merit of the designs. In particular, we consider a pair of classes where the distribution of the source conditioned on $C_1$ is Gaussian with mean $\mu_1 = 0$ and covariance matrix $\Sigma_1 = \begin{bmatrix} 1 & 0.25 \\ 0.25 & 0.5 \end{bmatrix}$, and the distribution of the source conditioned on $C_2$ is Gaussian with mean $\mu_2 = 0$ and covariance matrix $\Sigma_2 = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 0.25 \end{bmatrix}$. Note that $r_{\Sigma_1} = 2$, $r_{\Sigma_2} = 2$, $r_{\Sigma_1\Sigma_2} = 3$, and $N = 3$. Note also that there are 2 dimensions where the signals do not overlap and a single common dimension to both signals.

The measurement matrix $\Phi$ is designed by taking the first $M$ rows of the matrix $\begin{bmatrix} 1 \ 0 \ 0 \\ 0 \ 1 \ 0 \\ 0 \ 0 \ 1 \end{bmatrix}$, according to the desired number of measurements. Note that this measurement matrix is consistent with Theorem 1.

We can observe the behavior of the upper bound to the probability of misclassification in Figure 1. With a random measurement matrix we need $M = 3$ measurements in order to drive the probability of misclassification to zero as the noise approaches zero – as described in [10], we require $M > r_{\Sigma_1}$. In contrast, with the designed measurement matrix we need a single measurement to drive the probability of misclassification to zero as the noise approaches zero. The diversity-order depends on the number of designed measurements in accordance with Theorem 1. A single designed measurement offers only a diversity-order equal to $\frac{1}{4}$, but two designed measurements offer the maximum diversity-order equal to $\frac{1}{2} NO_{Dim} = \frac{2}{4}$. By analyzing Figure 2 we can observe that the diversity-order associated with the true probability of misclassification is lower bounded by the diversity-order associated with the upper bound to the probability of misclassification but there is very close agreement between the behavior of the upper bound and the true probability of misclassification.

We conclude that it is the ability to measure dimensions where the signals do not overlap, that allows us to drive the misclassification probability to zero, as the noise level approaches zero, in order to extract the maximum possible diversity. The value of measurement designs in relation to random ones then relates to the ability to single out these unique non-shared dimensions that act as the discriminative features associated with the problem - the higher the number of non-shared dimensions that one can single out via designed measurements the higher the diversity-order. Beyond $M = NO_{Dim}$, a designed matrix offers only additional measurement gain and no additional diversity gain.

**References**


