Information-theoretic Limits on the Classification of Gaussian Mixtures: Classification on the Grassmann Manifold

Matthew Nokleby*, Robert Calderbank*, and Miguel R. D. Rodrigues†
*Duke University, Durham, NC, email: {matthew.nokleby,robert.calderbank}@duke.edu
†University College London, email: m.rodrigues@ucl.ac.uk

Abstract—Motivated by applications in high-dimensional signal processing, we derive fundamental limits on the performance of compressive linear classifiers. By analogy with Shannon theory, we define the classification capacity, which quantifies the maximum number of classes that can be discriminated with low probability of error, and the diversity-discrimination tradeoff, which quantifies the tradeoff between the number of classes and the probability of classification error. For classification of Gaussian mixture models, we identify a duality between classification and communications over non-coherent multiple-antenna channels. This duality allows us to characterize the classification capacity and diversity-discrimination tradeoff using existing results from multiple-antenna communication. We also identify the easiest possible classification problems, which correspond to low-dimensional subspaces drawn from an appropriate Grassmann manifold.

I. INTRODUCTION

Efficient classification of high-dimensional signals is a fundamental component of machine learning. To this end, a variety of classic methods, such as linear discriminant analysis and principal component analysis [1], is used to produce low-dimensional descriptions of signals. Herein we consider an approach to dimensionality reduction inspired by compressed sensing [2]–[4]. High-dimensional signals are transformed into a low-dimensional feature set via a short, fat measurement matrix, after which they are classified. We focus on the classification of Gaussian mixture models (GMMs), in which classes correspond to multivariate Gaussian distributions. GMMs have been shown to be effective in image processing [5], [6], matrix completion [7], and audio signal processing [8].

A key question is the extent to which compression obscures distinctions between classes, thereby making classification difficult. A high-SNR characterization of the classification error probability as a function of the source and measurement geometry was carried out in [9]. In this paper, we investigate the number of classes that can be distinguished reliably and the statistical structure of the easiest classes to distinguish. Somewhat surprisingly, we show that the Shannon theory of wireless communication provides fundamental limits.

We focus on two classification scenarios. The first is batch classification, in which the classifier receives a sequence of high-dimensional signals to be classified jointly. This scenario corresponds, for example, to a classifier making decisions from multiple compressive measurements. For this scenario, we define the classification capacity, which, by analogy with Shannon capacity, quantifies the number of distinct sequences of classes that can be distinguished with vanishing probability of error as the batch size grows to infinity. Finding the classification capacity entails the maximization of mutual information, where the optimum distribution characterizes the classes that are easiest to discriminate. The second scenario is one-shot classification, in which the classifier receives a single signal to be classified. For this scenario, we define the diversity-discrimination tradeoff (DDT), which, by analogy with the diversity-multiplexing tradeoff (DMT) in wireless communication [10], quantifies the tradeoff between scaling up the number of classes to be discerned and driving down the misclassification probability as the SNR becomes large.

Our main contribution is a characterization of the classification capacity and DDT over zero-mean Gaussian mixture models. We identify a duality between classification over GMMs and communications over non-coherent multiple-input multiple-output (MIMO) channels. We observe that the discrimination of covariance matrices from noisy, compressed samples bears a structural resemblance to the recovery of codewords passed through a Rayleigh-fading channel matrix. The non-coherent MIMO channel has been studied extensively [11]–[14], and reasonably tight bounds on the capacity and DMT are known. The duality identified permits us to leverage existing results—with appropriate modifications—to establish bounds on classification performance.

Furthermore, the duality between classification and communication provides insight into the structure of the easiest classes to discriminate. In [13] it is shown that the capacity-achieving codewords are low-dimensional subspaces drawn from the appropriate Grassmann manifold; this type of signaling is also optimal with respect to the DMT. Similarly, the easiest classes to discriminate correspond to low-dimensional subspaces of the ambient signal space, a fact that should inform the selection of dictionaries intended for compressive classification.

The remainder of the paper is organized as follows. In Section II we specify the signal model. In Section III we study batch classification, defining formally the classification capacity and proving upper and lower bounds. In Section IV we study one-shot classification in the high-SNR regime, defining

1Our analysis can be extended to mixtures with non-zero means via appeals to coherent MIMO channels. For brevity’s sake we focus only on the zero-mean case.
the diversity-discrimination tradeoff and again proving upper and lower bounds. Finally, in Section V we offer concluding remarks and discuss future work.

II. SYSTEM MODEL

A. Signal Model

Here we introduce the batch signal model, in which the receiver obtains a sequence of signals to be classified. At time index \(k\), for \(1 \leq k \leq K\), the received signal is

\[
Y[k] = \sqrt{\text{SNR}} \Phi X[k] + Z[k],
\]

(1)

where \(\Phi \in \mathbb{C}^{M \times N}\) is a fixed sensing matrix, \(X[k] \in \mathbb{C}^{N \times T}\) is the signal of interest, \(Z[k] \in \mathbb{C}^{M \times T}\) is unit-variance i.i.d. complex AWGN, and SNR is the signal-to-noise ratio. We assume that \(M \leq N\) throughout. We call \(N\) the ambient signal dimension, \(M\) the received signal dimension, and \(T\) the coherence time of each index \(k\). That is, for each time index \(k\), the receiver obtains \(T\) vectors in \(\mathbb{C}^M\) corresponding to the class, an assumption motivated by video.

We further assume that the signals \(X[k]\) are drawn from a zero-mean Gaussian mixture model. That is, at each index \(k\), each column of \(X[k]\) is drawn independently from \(\mathcal{N}(0, U[k] U^H[k])\), where \(U[k] \in \mathbb{C}^{N \times N}\) characterizes the covariance. Then, (1) can be rewritten as

\[
Y[k] = \sqrt{\text{SNR}} \Phi U[k] H[k] + Z[k],
\]

(2)

where \(H[k] \in \mathbb{C}^{N \times T}\) has i.i.d. entries drawn from a unit-variance, complex Gaussian distribution. So that the SNR remains meaningful, we impose energy constraints\(^2\) on both \(\Phi\) and \(X[k]\):

\[
\|\Phi\|_F^2 \leq M,
\]

(3)

\[
\frac{1}{K} \sum_{k=1}^{K} \|U[k]\|_F^2 \leq N,
\]

(4)

where \(\|\cdot\|_F\) is the Frobenius norm.

B. Duality with non-coherent MIMO communication

Our analysis is rooted in a duality between classification of GMMs and communication over non-coherent MIMO channels. To illustrate this duality, we briefly review the non-coherent MIMO channel as studied in [13], [14]. It consists of a transmitter having \(N\) antennas, a receiver having \(T\) antennas, and an unknown, i.i.d. complex Gaussian channel matrix \(H \in \mathbb{C}^{T \times N}\) unknown to transmitter and receiver and persisting for \(M\) symbol times. The signal model is

\[
Y[k] = \sqrt{\text{SNR}} H[k] X[k] + Z[k],
\]

(5)

where \(Y[k]\) and \(Z[k]\) are \(T \times M\) matrices, and \(X[k]\) is \(N \times M\). The noise signal \(Z[k]\) is, as before, i.i.d. unit-variance noise. The transmit signal \(X[k]\) is subject to an energy constraint:

\[
\frac{1}{K} \sum_{k=1}^{K} \|X[k]\|_F^2 \leq MN.
\]

(6)

\(^2\) In compressive sensing, it is common to impose \(\ell_2\) norm constraints on the matrices, whereas we constrain the Frobenius norm in order to make explicit the connections to Shannon theory. One could derive similar results with \(\ell_2\) constraints with only small modification to the upper and lower bounds.

Equation (5) is similar to (2). In fact, taking the transpose of (2), we see that the signal model for classification over GMMs is nearly identical to that of communications over the non-coherent MIMO channel:

\[
Y^H[k] = \sqrt{\text{SNR}} H^H[k] U^H[k] \Phi + Z^H[k].
\]

(7)

Now the received signal is essentially identical to the non-coherent MIMO channel. In Table II-B we list the correspondence between dimensionalties in the two problems. The only difference is that the transmitted signal \(X[k]\), which is an arbitrary \(N \times M\) matrix, is replaced by \(U^H[k] \Phi H \in \mathbb{C}^{N \times M}\). The norm of the product is bounded by

\[
\|U^H[k] \Phi H\|_F^2 \leq \|U^H[k]\|_F^2 \|\Phi H\|_F^2 \leq NM.
\]

(8)

Roughly speaking, the performance of any compressive classifier therefore cannot exceed that of the corresponding MIMO system. We make this notion precise in the next two sections.

### Table I

<table>
<thead>
<tr>
<th>Classification</th>
<th>Dim.</th>
<th>Communication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ambient signal</td>
<td>(N)</td>
<td># of transmit antennas</td>
</tr>
<tr>
<td># of measurements</td>
<td>(M)</td>
<td>Channel coherence time</td>
</tr>
<tr>
<td>Class coherence time</td>
<td>(T)</td>
<td># of receive antennas</td>
</tr>
</tbody>
</table>

III. BATCH CLASSIFICATION

A. Classification Capacity

When \(K \to \infty\), we can employ Shannon theory to quantify fundamental performance limits. Here we define the classification capacity, which, by analogue with the usual Shannon capacity, characterizes the number of unique batches that can be discerned with vanishing probability of error. We start by defining a classification problem.

**Definition 1:** A classification problem \(\mathcal{P}^K\) is a collection of sequences of distributions where the sequences are indexed by \(i\) and the elements of each sequence are indexed by \(k\), i.e.

\[
\mathcal{P}^K = \{\mathcal{U}_i[k]\}, 1 \leq k \leq K, 1 \leq i \leq I.
\]

(9)

Each classification problem is analogous to a codebook in Shannon theory. Instead of a transmitter choosing a codeword, “nature” chooses a particular \(K\)-length batch, indexed by \(i\). Then, the receiver obtains the sequence of signals \(\{Y[k]\}\) according to (2), with \(U_i[k]\) describing the class at index \(k\). The problem of classifying the batch of signals therefore reduces to identifying which \(i\) was chosen. Similarly, in Shannon theory, identifying the transmitted signal reduces to identifying which codeword index was chosen at the transmitter. Each classification problem has a rate:

\[
R^K = \frac{\log_2(I)}{K},
\]

(10)

which is measured in bits per time index. The receiver employs a decoder \(D^K : \mathbb{C}^{M \times T \times K} \to \{1, \ldots, I\}\) mapping received signals \(Y[k]\), \(1 \leq k \leq K\) to an estimate \(i\). Define the probability of error \(P_e = P(i \neq i)\) as the probability that the classifier makes a mistake.
Definition 2: A classification rate $R$ is said to be achievable if there exists a sequence of problems $\mathcal{P}^K$ having rates $R^K > R$ and decoders $D^K$ such that $P_e \to 0$ as $K \to \infty$. Then, by analogy with Shannon theory, we can define the classification capacity.

Definition 3: The classification capacity $C_{N,M,T}(\text{SNR})$ is

$$C_{N,M,T}(\text{SNR}) = \sup \{ R : R \text{ is achievable} \}.$$  \tag{11}

From basic information theory it follows that

$$C_{N,M,T}(\text{SNR}) = \sup_{\Phi, p(U)} I(U; Y),$$  \tag{12}

where $p(U)$ is the distribution over the matrices $U[k]$, and where the supremum is taken over matrices and classes satisfying the constraints. Just as with Shannon capacity, a random coding argument suffices to establish the classification capacity: A classification problem with matrices $U[k]$ drawn randomly from the optimum distribution in (12) will have vanishing probability of error if the rate is below capacity.

At this point it is worthwhile to examine the operational significance of the classification capacity. After all, in classification we do not always have control over the classes to be distinguished, and very rarely will a batch of classification problems exhibit the codebook-like structure necessary to achieve capacity. One might therefore wonder if the classification capacity characterizes a rather unrealistic scenario.

Nevertheless, we contend that the classification capacity provides meaningful insight into the fundamental limits of classification performance. First, the classification capacity gives a hard upper limit on the number of classes that can be discriminated reliably per time index. If there are more than $2^C$ classes per index, then Fano’s inequality guarantees that the probability of error is bounded away from zero, regardless of the class structure or the batch length. Second, analysis of the capacity-achieving distributions reveals which classes are “easiest” to discriminate, and a particular classification problem can be compared to the optimal ones. Finally, as we will see in Section IV, Shannon-type analysis gives insight to the high-SNR performance of one-shot classifiers.

B. Classification Capacity Bounds

In order to characterize the classification capacity, it is necessary to find the distribution $p(U)$ and the sensing matrix $\Phi$ that maximize the mutual information $I(U; Y)$. In principle this is a daunting task; however, given the duality between classification and non-coherent MIMO described in Section II, it is possible to bound the classification capacity using well-known results. We begin with an upper bound.

Lemma 1: Let $C_{N,M,T}(\text{SNR})$ denote the capacity of the non-coherent MIMO with $N$ transmit antennas, $T$ receive antennas, and a coherence time of $M$. Then,

$$C_{N,M,T}(\text{SNR}) \leq M C_{N,M,T}^{\text{MIMO}}(\text{SNR}).$$ \tag{13}

Proof: From (7) we observe that the transpose of the received signal is identical to that of a non-coherent MIMO channel, except that the transmit signal $X[k]$ is replaced by the product $U^H[k] \Phi^H \in \mathbb{C}^{N \times M}$. From (8) the squared norm of the product cannot exceed $NM$, so at best the product is an arbitrary $N \times M$ matrix satisfying the transmit power constraint. The mutual information $I(U; Y)$ therefore is bounded by the mutual information between transmit and receive signals in the non-coherent MIMO channel. Since the capacity of the non-coherent channel is the mutual information normalized by the coherence time, the claim follows.

In addition to permitting an upper bound, the duality between classification and communication permits a lower bound on the classification capacity, as well as a description of the classes that achieve the lower bound. To show this, we need to review a basic result about the capacity-achieving codewords for non-coherent MIMO.

Definition 4: An $N \times N$ matrix $A$ is said to be isotropically distributed if $A$ is unitary, and

$$p(\mathbf{AV}) = p(\mathbf{A})$$ \tag{14}

for any fixed, unitary matrix $V \in \mathbb{C}^{N \times N}$.

In other words, the distribution of an isotropically distributed matrix is invariant under rotations. Such matrices play a key role in the capacity of non-coherent MIMO channels.

Lemma 2 (\cite{11}, Theorem 2): The capacity-achieving signals for a non-coherent MIMO channel as described in (5) can be written as

$$X[k] = B[k] A[k],$$ \tag{15}

where $A[k] \in \mathbb{C}^{N \times N}$ is an isotropically-distributed matrix, and $B[k] \in \mathbb{R}^{T \times N}$ is a diagonal matrix whose distribution is invariant to permutations of its diagonal elements.

We now show that classes having a similar structure achieve the classification capacity to within a constant gap.

Theorem 1: The classification capacity satisfies

$$M(c_1 - \log_2(M)) \leq C_{N,M,T}(\text{SNR}) - M^*(M - M^*) \log_2(\text{SNR}) \leq M c_2,$$ \tag{16}

where $M^* = \min\{T, \lfloor M/2 \rfloor \}$, and where $c_1$ and $c_2$ are constants not depending on SNR. Furthermore, the lower bound is achieved by choosing $\Phi = [W0]$, for $W$ any $M \times M$ unitary matrix, and

$$U[k] = \begin{bmatrix} \Theta[k] & 0 \\ 0 & D[k] \end{bmatrix},$$ \tag{17}

where $\Theta[k]$ is an $M \times M$ isotropically distributed unitary matrix, and $D[k]$ is an $M \times M$ diagonal matrix, and 0 represents a zero matrix of appropriate size.

Proof: Taking the transpose of the received signal, and substituting in the choices for $\Phi$ and $U[k]$, yields

$$Y^H[k] = \sqrt{\text{SNR}} H^H_{M \times T}[k] D[k] \Theta^H[k] W^H + Z^H[k],$$ \tag{18}

where $H_{M \times T}[k]$ denotes the first $M$ rows of $H[k]$. This signal model is equivalent to a non-coherent MIMO channel with $M$ transmit antennas, $T$ receive antennas, and a coherence time of $M$, except for a few small discrepancies. First, $\Theta^H[k]$ is premultiplied by $W^H$, but since $W^H$ is unitary, the product $\Theta^H[k] W^H$ is unitary, isotropically distributed, and independent of $W$. Second, in the non-coherent MIMO channel, the
matrix $U[k]$ is constrained to have average squared Frobenius norm not exceeding $MN$, whereas in classification the average squared norm must not exceed $N$. The classification problem is therefore equivalent to a non-coherent MIMO channel with signal to noise ratio $SNR/M$.

Now, in [13, Section IV.D], it is shown that
\[
c_1 \leq C_{N,T,M}^{MIMO}(SNR) - M^*(1 - M^*/M) \log_2(SNR) \leq c_2,
\]
where $c_1, c_2$ do not depend on $SNR$. It is also shown that the proposed choice for $U[k]$ is sufficient to achieve the lower bound. Therefore,
\[
C_{N,M,T}(SNR) \geq MC_{N,T,M}^{MIMO}(SNR/M).
\]
Combining (19) and (20) with Lemma 1 yields the claim.

Remark 1: The capacity bounds do not depend on the ambient dimension $N$; only the number of measurements $M$ and the coherence time $T$ determine the capacity, except possibly for the coefficients $c_1$ and $c_2$. This is because, for $M \leq N$, the corresponding MIMO channel has short coherence, and there is no significant capacity advantage to signaling over all $N$ antennas. Furthermore, for $T = 1$, $M^* = 1$, and the classification capacity grows approximately as $M \log_2(SNR)$. Therefore, having a longer coherence time considerably improves classification performance up to $T = \lfloor M/2 \rfloor$, at which point the number of measurements becomes the bottleneck. In order for the classification error to vanish, therefore, the number of components per time index in any Gaussian mixture must scale no faster than $SNR^{M^*(M - M^*/M)}$. Furthermore, it is shown in [13] that the bounds are achieved when $M^*$ diagonal entries of $D[k]$ are bounded away from zero. In other words, the GMMs easiest to discriminate are composed of subspaces of $\mathbb{C}^N$, drawn from the Grassmann manifold having dimension no higher than $M/2$. In [10] it is also shown that a training scheme, in which pilot symbols are sent over the first $M^*$ symbol times and Gaussian codewords are sent over the remaining symbol times, achieves near-capacity performance. While such a scheme corresponds to a rather odd Gaussian mixture, the intuition is the same as in the approach shown here: For best performance, classes should correspond to low-dimensional subspaces.

IV. ONE-SHOT CLASSIFICATION

Whereas in the previous section we found limits on the performance of large batches of classification problems, in this section we examine one-shot classification at high SNR, which leads to the diversity-discrimination tradeoff. We first define formally the DDT, after which we prove upper and lower bounds. As with the classification capacity, these bounds are found by leveraging the duality between classification and communication over non-coherent MIMO.

A. Diversity-discrimination Tradeoff

In wireless, diversity-multiplexing tradeoff (DMT) was introduced in [10] to characterize the high-SNR performance of fading coherent MIMO channels. It was shown in [10] that the spatial flexibility provided by multiple antennas can simultaneously increase the achievable rate and decrease the probability of error, but only according to a tradeoff that is tightly characterized at high SNR. For classification, we desire a similar characterization, which unveils the relationship between the decay of error probability and the increase in the number of discernible classes as the SNR becomes large.

Since here we examine one-shot classification, $K = 1$ and we drop the index $k$. Therefore each classification problem $P$ is merely a set $\{U_i\}$ of $I$ matrices. Now, let $P(SNR)$ be a sequence of classification problems defined for each SNR, where $I$ may vary with the SNR. As before, let $P_e = P(i \neq i)$ be the probability of error for each classification problem. Then the diversity-discrimination tradeoff is defined as follows.

Definition 5: A sequence $P(SNR)$ has discrimination gain $r$ and diversity gain $d$ if
\[
\lim_{SNR \to \infty} \frac{\log_2(I)}{\log_2(SNR)} = r
\]
and
\[
\lim_{SNR \to \infty} \frac{\log_2(P_e)}{\log_2(SNR)} = -d.
\]
We say that the sequence $P(SNR)$ has diversity-discrimination function $d^*(r)$. In other words, the sequence of problems $P(SNR)$ has approximately $SNR^d$ classes and a probability of error approximately $SNR^{-d}$. This leads us to the following definition.

Definition 6: The diversity-protection tradeoff is defined as the supremum over all DDT functions, or
\[
d^*(r) = \sup_{P(SNR)} d(r).
\]
where the supremum is over all sequences satisfying the constraints (3) and (4). Therefore, at high SNR, a classification problem with $SNR^{-d}$ classes cannot have probability of error decaying faster than $SNR^{-d^*(r)}$.

B. DDT Bounds

As before, the classification problem corresponds to the non-coherent MIMO channel with short coherence time. In the short coherence time regime, only non-matching upper and lower bounds on the DMT function are known. However, we can still leverage these bounds to characterize the high-SNR performance limits of classification. To do so, we first define the function
\[
G_{X,Y,Z}(r) = \min_{\alpha \in G} \sum_{i=1}^{\min\{X,Y\}} (2i - 1 + |X - Y|)\alpha_i + \left(Z \sum_{i=1}^{\min\{X,Y\}} (1 - \alpha_i) - r \right),
\]
where
\[
G = \{\alpha \in [0, 1]^{\min\{X,Y\}} : \alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_{\min\{X,Y\}} \sum_i (1 - \alpha_i) > r \}.
\]
This function describes the DMT achieved by random Gaussian codes over coherent MIMO channels [15]. Now, we bound the DDT in the following theorem.

**Theorem 2:** The diversity-discrimination tradeoff function $d^*(r)$ is bounded above by the piecewise linear function connecting the points

$$\frac{M - \min\{N,T\}}{M} (T - r/M)(N - r/M),$$

for $r \in \{0, M, 2M, \ldots, M \min\{N,T\}\}$. Furthermore, $d^*(r)$ is bounded below by

$$d^*(r) \geq G_{M,T,M-M^*}(r)$$

$$d^*(r) \geq G_{M,M-M^*,T}(r),$$

where again $M^* = \min\{T, \lfloor M/2 \rfloor\}$.

**Proof:** The proof follows a similar outline to Theorem 1. First, looking at the transpose in (7), we again see that the performance of the classifier is upper-bounded by that of a non-coherent MIMO channel with $T$ receive antennas, $N$ transmit antennas, and a coherence time of $M$, with the proviso that the rates in the MIMO channel are normalized by $M$, whereas those of the classifier are not. In [15, Theorem 31] the upper bound on the DMT is proven, which, after normalizing by $M$, yields the upper bound claimed.

Next, by choosing $U[k]$ to be a zero-padded $M \times M$ matrix and $\Phi$ to be a unitary, zero-padded $M \times M$ matrix, the performance is at least as good as a non-coherent MIMO system having $M$ transmit antennas, $T$ receive antennas, and a coherence time of $M$, with an equivalent signal-to-noise ratio of $\text{SNR}/M$. This SNR gap vanishes in the limit, and we invoke lower bounds proven in [15, Theorem 31].

The lower bounds also can be written as piecewise linear functions, although their forms are more complicated. The lower bounds are achieved via “sending” pilot symbols over $M^*$ signal dimensions, and using random and expurgated Gaussian codes for the remaining. However, it is also shown in [14] that signals of the form specified in Lemma 2 are DMT-optimal. Therefore, the DDT-optimum classes also correspond to low-dimensional subspaces drawn from an appropriate Grassmann manifold.

Unlike with the classification capacity, the upper bound depends on $N$, while the lower bounds do not. The extra dimensions, which remain unused in our achievability scheme, can provide an additional diversity advantage. Future work includes investigation of schemes that exploit the extra dimensions.

V. Conclusion

We have derived fundamental limits on the performance of linear compressive classifiers over Gaussian mixture models, identifying a duality between classification of GMMs and communication over non-coherent MIMO channels. In particular, we have shown that, in the limit of large batch size or high SNR, classifier performance can be characterized via Shannon theory. For batch classification problems, these characterizations provide hard limits on the number of classes that can be discerned with vanishing probability of error. For one-shot classification problems, they provide limits on the tradeoff between the number of classes and the probability of error. In both cases, the easiest classes to discriminate correspond to low-dimensional subspaces drawn from an appropriate Grassmann manifold.

We also point out that one can extend our results to classifiers over arbitrary statistical models. We have omitted the details for brevity, but it turns out that capacity and DMT results from the coherent MIMO channel can be leveraged to bound the performance of any classifier.

Finally, there is significant future work to be done interfacing these results with practical classification techniques. For example, we envision dictionary learning algorithms [16] that take advantage of the characterization of optimal classes. Such methods would be tuned to balance the desire for easily-distinguished classes against the desire for other useful properties, such as sparsity and fidelity of representation.

References


