Fundamental Limits on the Performance of Compressive Classification:

A Characterization Inspired by Dualities Between Classification and Wireless Communications Problems

Miguel Rodrigues

Dept. of Electronic and Electrical Engineering

University College London

Collaborators

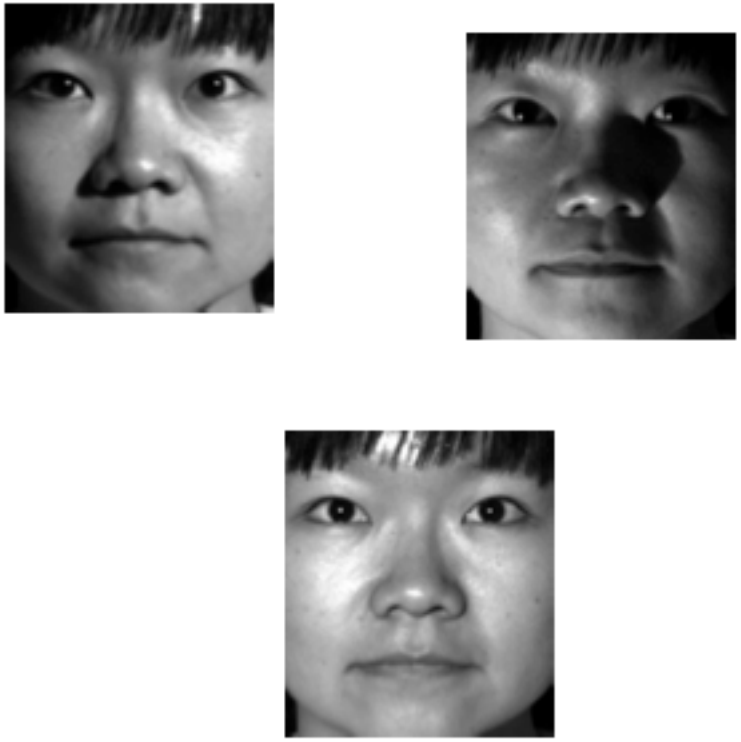
Robert Calderbank
Duke U.

Matthew Nokleby
Duke U.

Hugo Reboredo
U. Porto

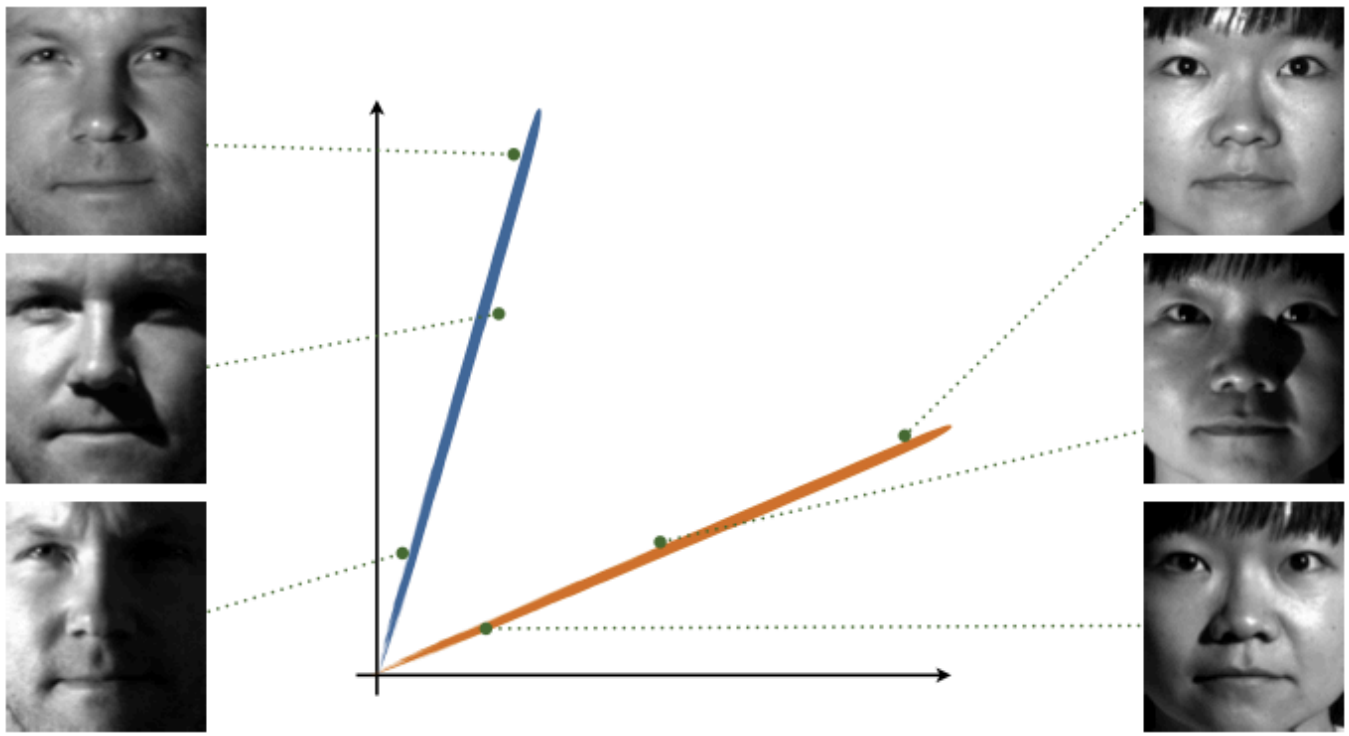
Francesco Renna
U. Porto

Motivation: Face Recognition



source: yale face database b

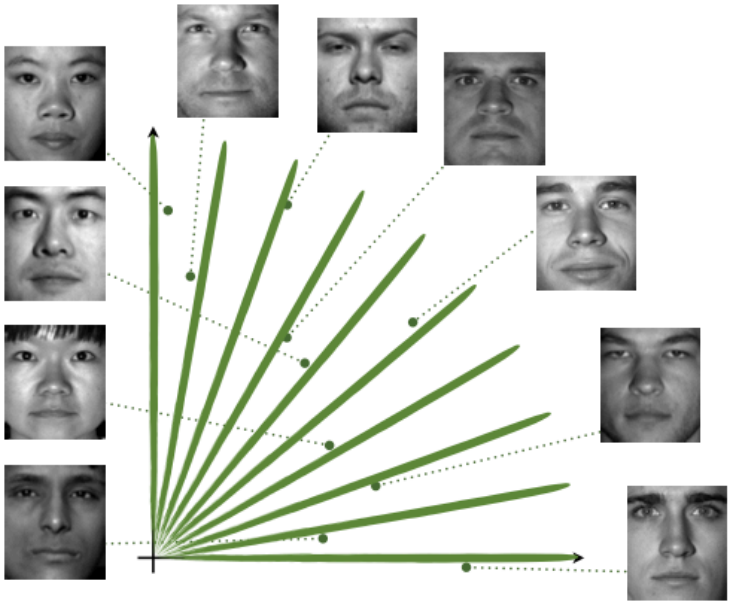
Motivation: Face Recognition



- fixed pose, varying illumination
- images lie in near 9-dimensional **subspaces**

Questions

- Q1: What is the relation between performance and geometry – a "beyond phase transitions" characterization of the misclassification probability
- Q2: What is the number of faces that we can distinguish with arbitrarily low misclassification probability?
- Q3: What is the tradeoff between the number of faces can we distinguish and the misclassification probability?



Answers provided for a stylized model that captures the fact that data lives in a union of subspaces (or a union of affine spaces)

Related Work

- Compressive information processing

  [Davenport,2007], [Davenport,2010], [Duarte,2006], [Duarte,2007], [Haupt,2006], [Haupt,2007], [Krishnamurthy,2010a], [Krishnamurthy,2010b], [Wimalajeewa, 2012], [Wright,2009]

- Task-specific sensing

  [Neifeld,2007], [Ashok,2008], [Baheti,2008], [Baheti,2009], [Ke,2010]

- Sparse support recovery

  [Aeron,2010; Akcakaya,2009; Fletcher,2009; Rad,2011; Reeves, 2011; Reeves, 2012; Reeves,2013; Tang,2010; Tulino,2013; Wang,2010; Wainwright,2009a; Wainwright,2009b; Zhao,2006]

- Machine learning

  Feature extraction or supervised dimensionality reduction

Model

Measurement model

Standard measurement model:

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{w}, \quad \mathbf{w} \sim \mathcal{N}\left(0, \sigma^2\mathbf{I}\right)$$

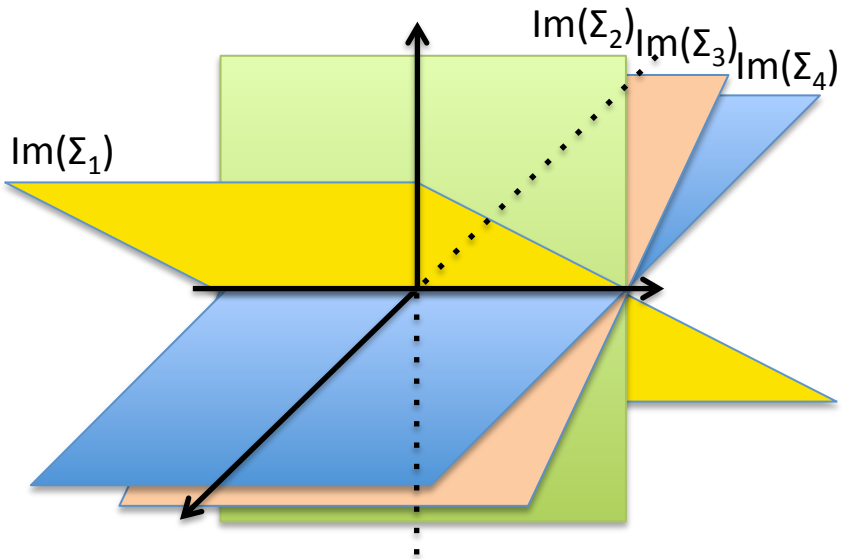
where $\Phi \in \Re^{M\times N}$ is the measurement matrix, $\mathbf{y} \in \Re^{M}$ is the measurement vector, $\mathbf{x} \in \Re^{N}$ is the signal vector, and $\mathbf{w} \in \Re^{N}$ is standard Gaussian noise.

Signal model (Gaussian mixture model)

The signal belongs to one out of C classes, where $\Pr(c) = p_c$ and $p(\mathbf{x} \mid c) = N(\mu_c, \Sigma_c)$.

Objective

The objective is to determine the signal class c from the signal projections $\mathbf{y}$.



Signal lives on
union of subspaces or more generally
union of affine spaces

Unveiling Dualities

Consider that

$$\mathbf{x} \sim \mathcal{N}\left(0, \Sigma_c\right)$$

Then, the compressive classification model

$$\mathbf{y} = \Phi \Sigma_c^{1/2} \mathbf{h} + \mathbf{w}, \quad \mathbf{h} \sim \mathcal{N}\left(0, \mathbf{I}\right), \quad \mathbf{w} \sim \mathcal{N}\left(0, \sigma^2 \mathbf{I}\right)$$

reduces to the non-coherent multiple-antenna communications model

$$\mathbf{y}^t = \mathbf{h}^t \Sigma_c^{1/2} \Phi^t + \mathbf{w}^t, \quad \mathbf{h} \sim \mathcal{N}\left(0, \mathbf{I}\right), \quad \mathbf{w} \sim \mathcal{N}\left(0, \sigma^2 \mathbf{I}\right)$$

(a model where transmitter and receiver do not know the channel realization, only the channel statistics)

Unveiling Dualities

Consider that

$$\mathbf{x} \sim \mathcal{N}\left(0, \Sigma_c\right)$$

Then, the compressive classification model

$$\mathbf{y} = \Phi \Sigma_c^{1/2} \mathbf{h} + \mathbf{w}, \quad \mathbf{h} \sim \mathcal{N}\left(0, \mathbf{I}\right), \quad \mathbf{w} \sim \mathcal{N}\left(0, \sigma^2 \mathbf{I}\right)$$

reduces to the non-coherent multiple-antenna communications model

$$\mathbf{y}^t = \mathbf{h}^t \Sigma_c^{1/2} \Phi^t + \mathbf{w}^t, \quad \mathbf{h} \sim \mathcal{N}\left(0, \mathbf{I}\right), \quad \mathbf{w} \sim \mathcal{N}\left(0, \sigma^2 \mathbf{I}\right)$$

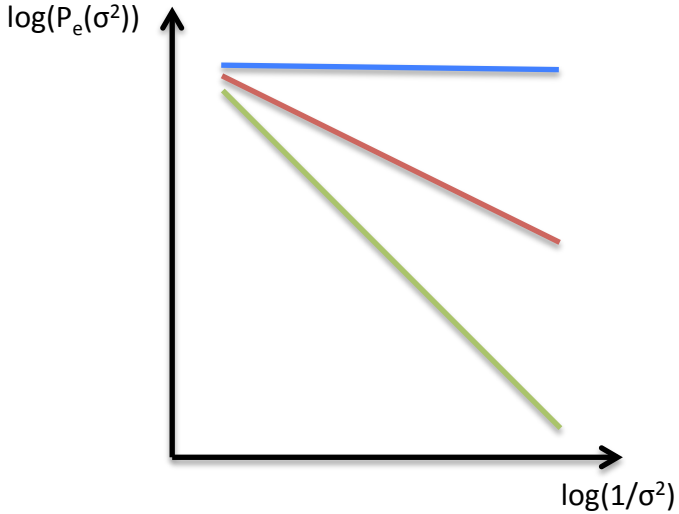
(a model where transmitter and receiver do not know the channel realization, only the channel statistics)

Correspondences Between the Communications and the Classification Problem

Parameter	Classification Model	Communications Model
N	Signal Dimension	Number of Transmit Antennas
M	Signal Measurements	Channel Coherence Time
T=1	Class Coherence Time	Number of Receive Antennas

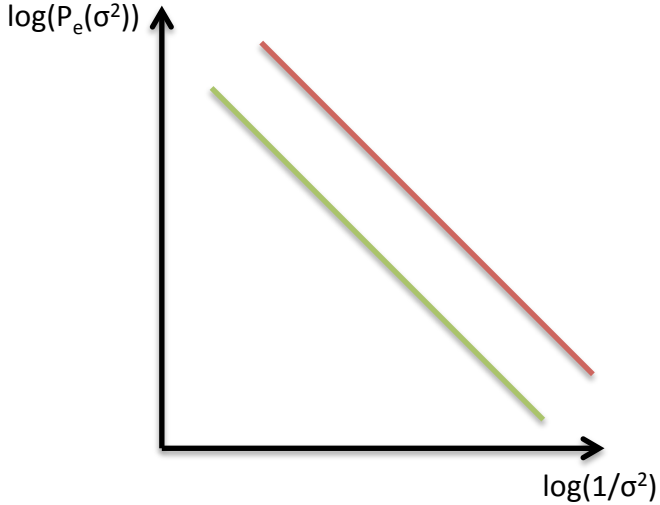
Communications-Theoretic Inspired Performance Characterization

A more refined characterization of the behavior of the misclassification probability



Diversity gain

Determines the slope of the misclassification probability (in the log $(1/\sigma^2)$ scale) at low noise levels



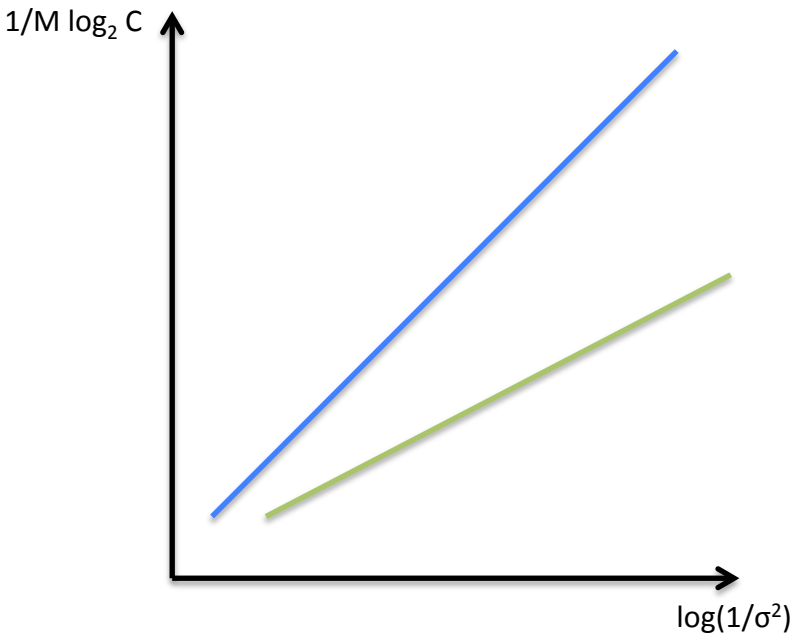
Measurement gain

Determines the offset of the misclassification probability (in the log $(1/\sigma^2)$ scale) at low noise levels

Information-Theoretic Inspired Performance Characterization

Characterization of the "data volume" that can be handled by the classifier with arbitrarily low misclassification probability

Classification Capacity

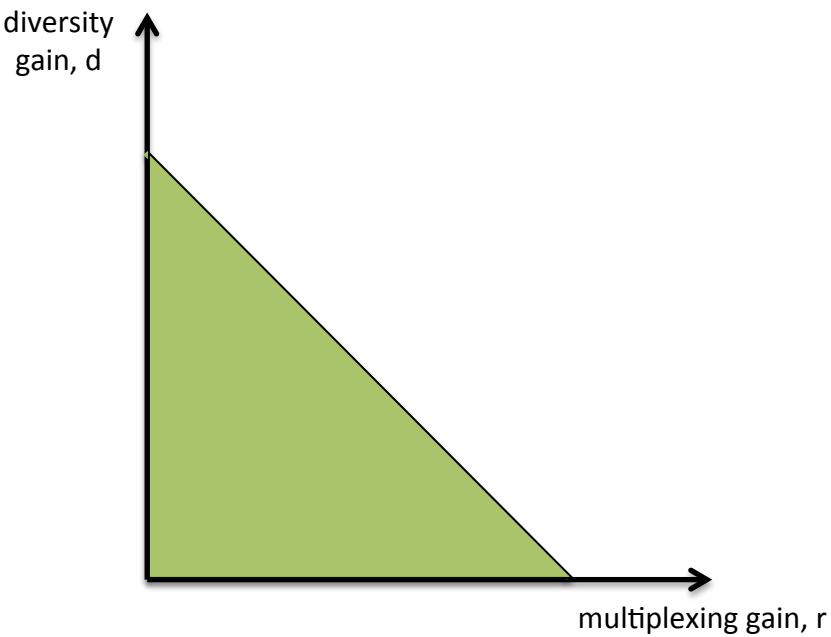


- Defines the number of unique classes that can be discerned by the classifier with arbitrarily low misclassification probability in the limit of high dimension
- Defines phase transitions in the misclassification probability in terms of the (logarithm of the) number of classes normalized by signal dimension in the limit of high dimension
- Akin to **channel capacity** in communications

Information-Theoretic Inspired Performance Characterization

Characterization of tradeoffs between the "data volume" that can be handled by the classifier and its performance

Diversity-Discrimination Tradeoff (DDT)



- Characterizes the relationship between the number of classes that can be discerned by the classifier and the associated misclassification probability in the limit of low noise
- Characterizes a region of achievable exponent pairs for the number of classes and the misclassification probability in the limit of low noise
- Akin to the **diversity-multiplexing tradeoff** in wireless communications

Characterizations in the Compressive Classification Domain and the Wireless Communications Domain

Compressive Classification Domain	Wireless Communications Domain	References
Diversity Gain and Measurement Gain	Diversity Gain and Coding Gain	[Tarokh, Seshadri and Calderbank' 98]
Classification Capacity	Channel Capacity	[Zheng and Tse'2002]
Diversity-Discrimination Tradeoff	Diversity-Multiplexing Tradeoff	[Zheng and Tse'2003]

[Tarokh, Seshadri and Calderbank'98] V. Tarokh, N. Seshadri and A. R. Calderbank. Space-time codes for high data rate wireless communication: performance criterion and code construction. IEEE Transactions on Information Theory, 1998.

[Zheng and Tse'03] L. Zheng and D. Tse. Diversity and mutiplexing: A fundamental tradeoff in multiple-antenna channels. IEEE Transactions on Information Theory, 2003.

[Zheng and Tse'02] L. Zheng and D. Tse. Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel. IEEE Transactions on Information Theory, 2002.

Communications-Theoretic Inspired Characterization of Fundamental Limits

Enter "Phase Transitions", Diversity Gain and Measurement Gain

Model

Measurement model

Recall the measurement model:

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{w}, \quad \mathbf{w} \sim \mathcal{N}\left(0, \sigma^2\mathbf{I}\right)$$

where $\Phi \in \Re^{M\times N}$ is the measurement matrix, $\mathbf{y} \in \Re^{M}$ is the measurement vector, $\mathbf{x} \in \Re^{N}$ is the signal vector, and $\mathbf{w} \in \Re^{N}$ is standard Gaussian noise.

Signal model

The signal belongs to one out of two classes where

$\Pr\left(c = 1\right) = P_1$

$\Pr\left(c = 2\right) = P_2$

$p\left(\mathbf{x} \mid c = 1\right) = \mathcal{N}\left(0, \Sigma_1\right)$

$p\left(\mathbf{x} \mid c = 2\right) = \mathcal{N}\left(0, \Sigma_2\right)$

Model

Measurement model

Recall the measurement model:

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{w}, \quad \mathbf{w} \sim \mathcal{N}\left(0, \sigma^2\mathbf{I}\right)$$

where $\Phi \in \Re^{M\times N}$ is the measurement matrix, $\mathbf{y} \in \Re^{M}$ is the measurement vector, $\mathbf{x} \in \Re^{N}$ is the signal vector, and $\mathbf{w} \in \Re^{N}$ is standard Gaussian noise.

Signal model

The signal belongs to one out of two classes where

$$\Pr\left(c=1\right) = P_1$$

$$\Pr\left(c=2\right) = P_2$$

$$p\left(\mathbf{x} \mid c=1\right) = \mathcal{N}\left(0, \Sigma_1\right)$$

$$p\left(\mathbf{x} \mid c=2\right) = \mathcal{N}\left(0, \Sigma_2\right)$$

Quantity	Interpretation
$r_{\Sigma_i} = \mathrm{rank}\left(\Sigma_i\right)$	Dimension of subspace i
$r_{\Sigma_{ij}} = \mathrm{rank}\left(\Sigma_i + \Sigma_j\right)$	Dimension of direct sum of subspaces i and j
$v_i = \mathrm{pdet}\left(\Phi\Sigma_i\Phi^t\right)$	"Volume" of project subspace i
$v_{ij} = \mathrm{pdet}\left(\Phi\left(\Sigma_i + \Sigma_j\right)\Phi^t\right)$	"Volume" of direct sum of projected subspaces i and j
$NO_{DIM} = \left(r_{\Sigma_{ij}} - r_{\Sigma_i}\right) + \left(r_{\Sigma_{ij}} - r_{\Sigma_j}\right)$	Number of non-overlapping dimensions

Performance Metrics

Misclassification probability: Two-Class Case

$$P_e = \int \min\left(P_1 \cdot p\left(\mathbf{y} \mid c=1\right), P_2 \cdot p\left(\mathbf{y} \mid c=2\right)\right) d\mathbf{y}$$

Upper Bound to Misclassification Probability: Two-Classes

$$P_e^{UB} = \int \sqrt{P_1 \cdot P_2} \sqrt{p\left(\mathbf{y} \mid c=1\right) \cdot p\left(\mathbf{y} \mid c=2\right)} d\mathbf{y}$$

(the two-class bound generalizes to multiple classes by using a combination of the union bound together with the Bhattacharya bound)

Diversity gain:

$$d = \lim_{\sigma^2 \to \infty} \frac{\log P_e^{UB}\left(\sigma^2\right)}{\log \sigma^2}$$

Measurement gain

$$g_m = \lim_{\sigma^2 \to \infty} \sigma^2 \frac{1}{\sqrt[d_{UB}]{P_e^{UB}\left(\sigma^2\right)}}$$

Random Measurements:

Two Zero-Mean Gaussian Classes

Theorem:

Consider the measurement model:

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{w}, \qquad \mathbf{w} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$$

where $\mathbf{x} \sim \mathcal{N}\left(\mathbf{0}, \Sigma_1\right)$ with probability P_1 and $\mathbf{x} \sim \mathcal{N}\left(\mathbf{0}, \Sigma_2\right)$ with probability P_2

–When $\dfrac{r_{\Sigma_1} + r_{\Sigma_2}}{2} = r_{\Sigma_{12}}$

$$P_e^{UB}\left(\sigma^2\right) = O(1), \qquad \sigma^2 \to 0$$

Subspaces overlap

–When $\dfrac{r_{\Sigma_1} + r_{\Sigma_2}}{2} < r_{\Sigma_{12}}$

– if $\mathrm{M} \le \min\left(\mathrm{r}_{\Sigma_1}, \mathrm{r}_{\Sigma_2}\right)$ then

$$P_e^{UB}\left(\sigma^2\right) = O(1), \qquad \sigma^2 \to 0$$

Subspaces do not overlap

– if $\mathrm{M} > \min\left(\mathrm{r}_{\Sigma_1}, \mathrm{r}_{\Sigma_2}\right)$ then

$$P_e^{UB}\left(\sigma^2\right) = \left(\frac{g_m}{\sigma^2}\right)^{-d} + o\left(\left(\frac{1}{\sigma^2}\right)^{-d}\right), \qquad \sigma^2 \to 0$$

Measurement gain

$$g_m = \left(2^{\frac{\min\left(M, r_{\Sigma_{12}}\right)}{2}} \sqrt{P_1 P_2} \left(\frac{v_{12}}{\sqrt{v_1 v_2}} \right)^{-\frac{1}{2}} \right)^{-\frac{1}{d}}$$

Diversity gain

Number of Measurements	Diversity Gain
$\min\left(r_{\Sigma_1}, r_{\Sigma_2}\right) < M \leq \max\left(r_{\Sigma_1}, r_{\Sigma_2}\right)$	$d = \frac{1}{2} \cdot \left(\frac{\min\left(r_{\Sigma_1}, r_{\Sigma_2}\right) - M}{2} \right)$
$\max\left(r_{\Sigma_1}, r_{\Sigma_2}\right) < M \leq r_{\Sigma_{12}}$	$d = \frac{1}{2} \cdot \left(\frac{r_{\Sigma_1} + r_{\Sigma_2} - 2M}{2} \right)$
$M > r_{\Sigma_{12}}$	$d = \frac{1}{2} \cdot \left(\frac{r_{\Sigma_1} + r_{\Sigma_2} - 2r_{\Sigma_{12}}}{2} \right)$

Measurements Design:
Framework

Optimization Framework

$$\max_{\Phi} d(\Phi) \quad \text{subject to} \quad \text{rank}(\Phi) \leq M$$

Other Optimization Frameworks

1. Determine the class of measurement matrices that maximize diversity
2. Determine the measurement matrix that maximize measurement gain

(this alternative framework would target both the decay and power offset of the misclassification probability)

Measurements Design:
Two Zero-Mean Gaussian Classes

Theorem:

Consider the measurement model:

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{w}, \qquad \mathbf{w} \sim \mathcal{N}\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$$

where $\mathbf{x} \sim \mathcal{N}\left(\mathbf{0}, \Sigma_1\right)$ with probability P_1 and $\mathbf{x} \sim \mathcal{N}\left(\mathbf{0}, \Sigma_2\right)$ with probability P_2

– When $M \geq NO_{DIM}$

$$P_e^{UB}\left(\sigma^2\right) = \left(\frac{g_m}{\sigma^2}\right)^{-d} + o\left(\left(\frac{1}{\sigma^2}\right)^{-d}\right), \qquad \sigma^2 \to 0$$

where $d = 1/4 \cdot NO_{DIM}$

measurements sufficient to capture all discriminative features

– When $M < NO_{DIM}$

$$P_e^{UB}\left(\sigma^2\right) = \left(\frac{g_m}{\sigma^2}\right)^{-d} + o\left(\left(\frac{1}{\sigma^2}\right)^{-d}\right), \qquad \sigma^2 \to 0$$

where $d = 1/4 \cdot M$

measurements not sufficient to capture all discriminative features

Measurements Design:

Two Zero-Mean Gaussian Classes

Basis for Linear Spaces:

Linear Space	Basis
$\mathrm{Null}(\Sigma_1) \cap \mathrm{Null}(\Sigma_2)$	$w_1, w_2, \ldots, w_{n_{12}}$
$\mathrm{Null}(\Sigma_1)$	$w_1, w_2, \ldots, w_{n_{12}}, u_1^1, u_2^1, \ldots, u_{n_1}^1$
$\mathrm{Null}(\Sigma_2)$	$w_1, w_2, \ldots, w_{n_{12}}, u_1^2, u_2^2, \ldots, u_{n_1}^2$

Optimum Measurement Design

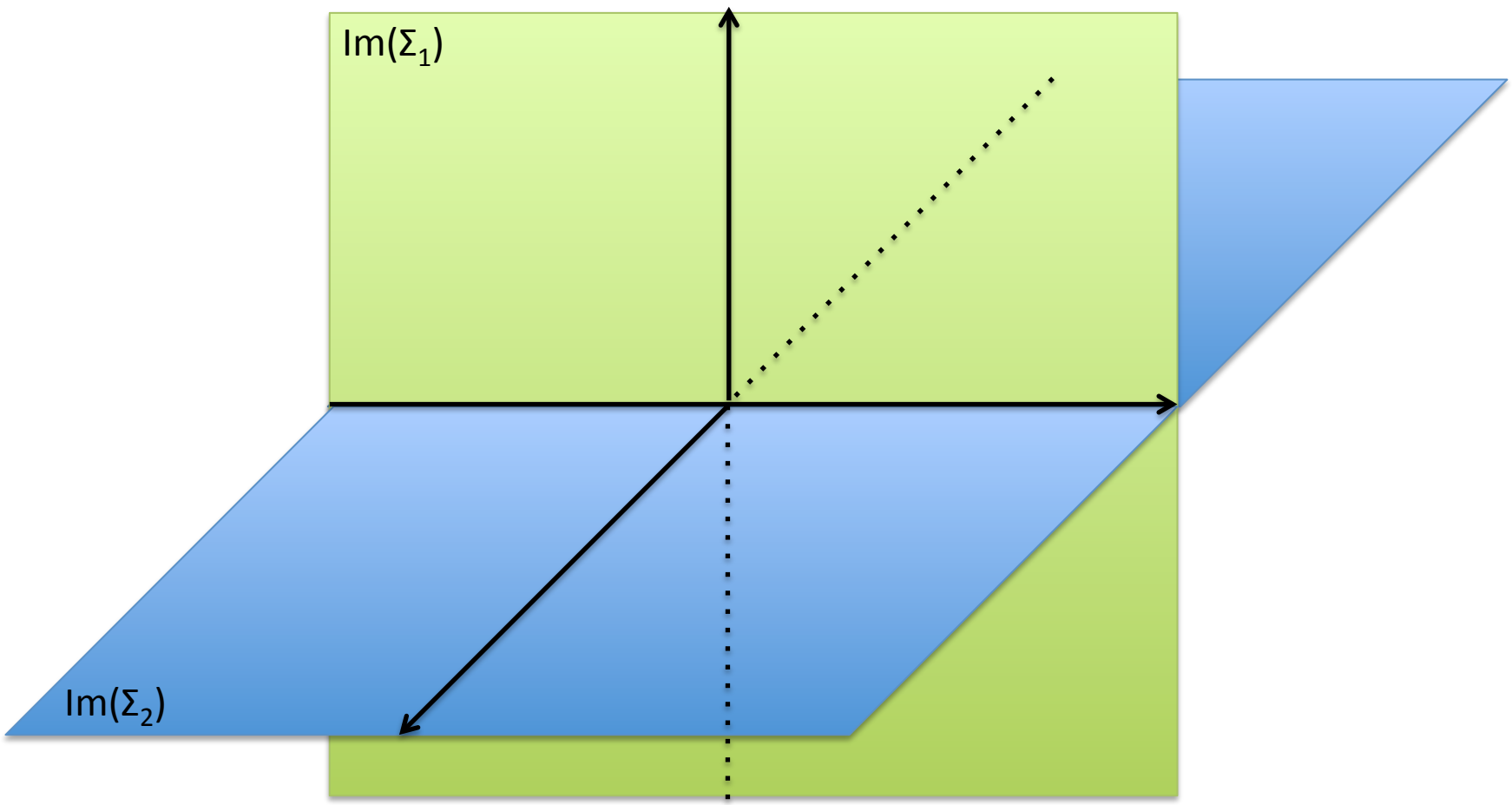
– When $M \geq NO_{DIM}$

$$\Phi = \left[u_1^{1t}, \ldots, u_{n_1}^{1t}, u_1^{2t}, \ldots, u_{n_2}^{2t} \right]^t$$

– When $M < NO_{DIM}$

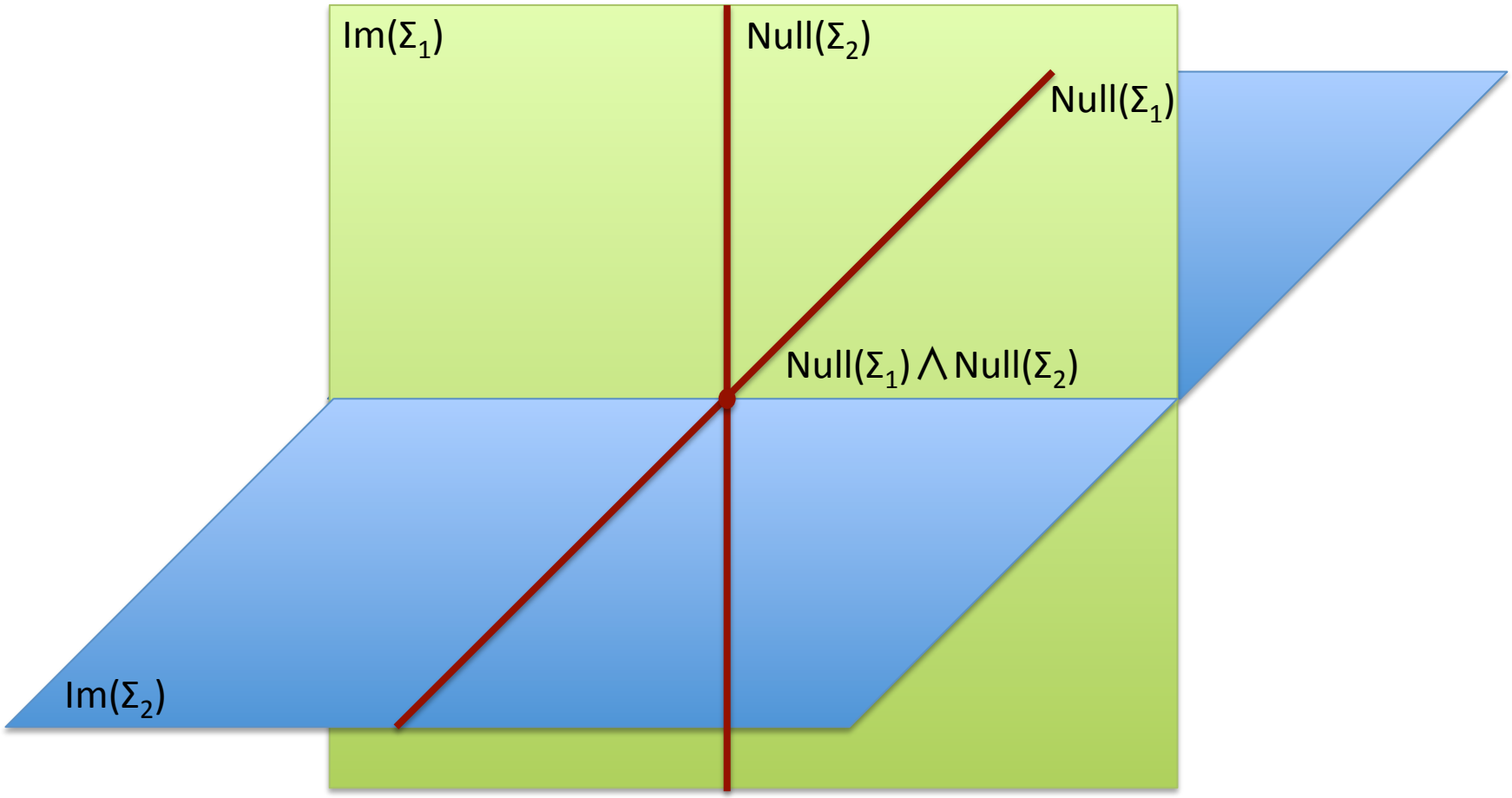
Φ can be constituted by using any M vectors from $u_1^{1t}, \ldots, u_{n_1}^{1t}, u_1^{2t}, \ldots, u_{n_2}^{2t}$

Interpretation



$$NO_{DIM} = 2 \cdot \dim\left(\mathrm{Im}\left(\Sigma_1 + \Sigma_2\right)\right) - \dim\left(\mathrm{Im}\left(\Sigma_1\right)\right) - \dim\left(\mathrm{Im}\left(\Sigma_1\right)\right) = 2 \times 3 - 2 - 2 = 2$$

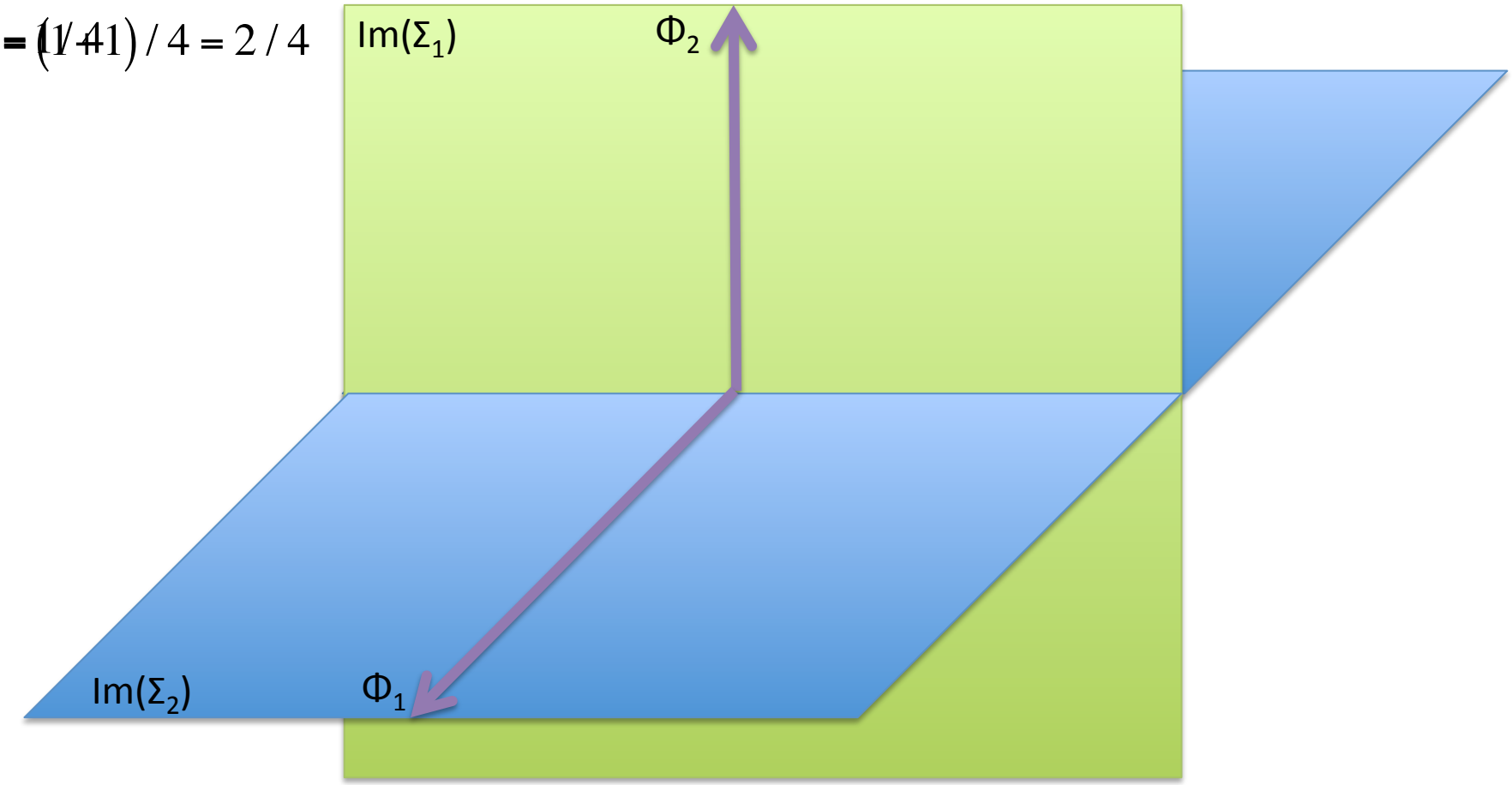
Interpretation



$$NO_{DIM} = 2 \cdot \dim\left(\mathrm{Im}\left(\Sigma_1 + \Sigma_2\right)\right) - \dim\left(\mathrm{Im}\left(\Sigma_1\right)\right) - \dim\left(\mathrm{Im}\left(\Sigma_1\right)\right) = 2 \times 3 - 2 - 2 = 2$$

Interpretation

$d = (1 + 1) / 4 = 2 / 4$

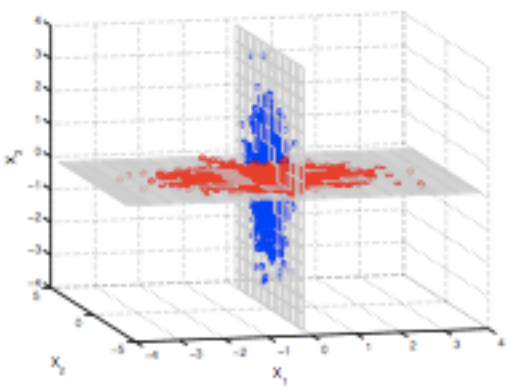


$$NO_{DIM} = 2 \cdot \dim\left(\mathrm{Im}\left(\Sigma_1 + \Sigma_2\right)\right) - \dim\left(\mathrm{Im}\left(\Sigma_1\right)\right) - \dim\left(\mathrm{Im}\left(\Sigma_1\right)\right) = 2 \times 3 - 2 - 2 = 2$$

Representative Results:

Two Zero-Mean Gaussian Classes

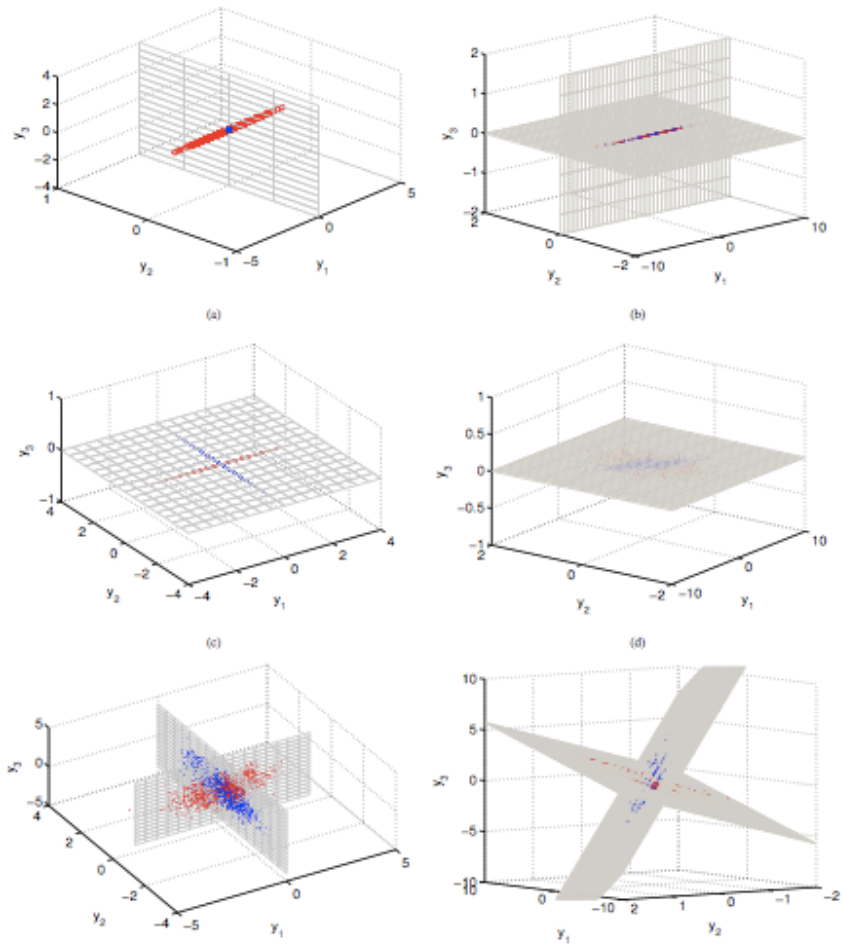
Realizations from the original classes



Realizations from the projected classes

Designed Projections

Random Projections



M=1

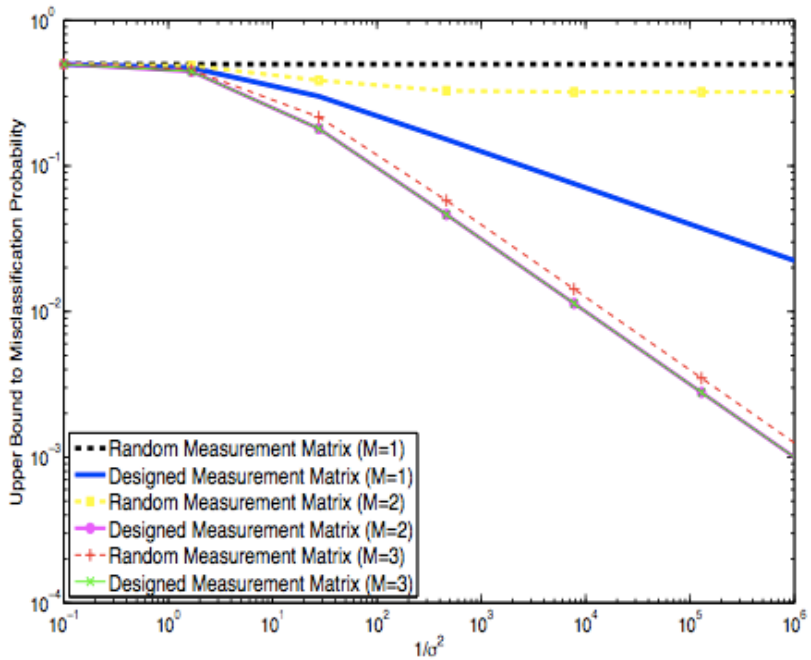
M=2

M=3

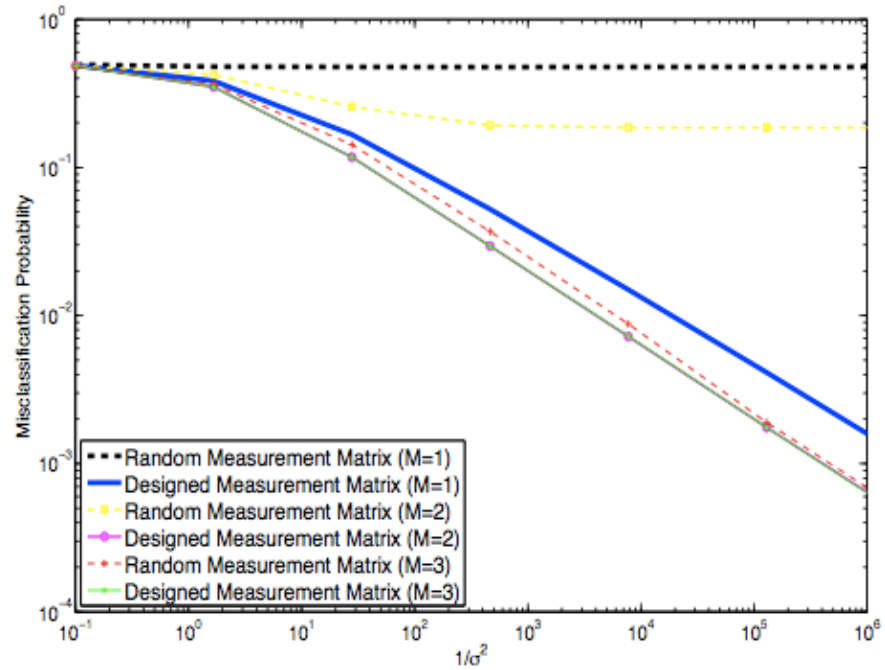
Representative Results:

Two zero-Mean Gaussian Classes

Misclassification Probability Upper Bound



Exact Misclassification Probability



Information-Theoretic Inspired Characterization of Fundamental Limits

Classification Capacity

and

Diversity-Discrimination Tradeoff

Model

Measurement model

Recall the measurement model:

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{w}, \quad \mathbf{w} \sim \mathcal{N}\left(0, \sigma^2\mathbf{I}\right)$$

where $\Phi \in \Re^{M\times N}$ is the measurement matrix, $\mathbf{y} \in \Re^{M}$ is the measurement vector, $\mathbf{x} \in \Re^{N}$ is the signal vector, and $\mathbf{w} \in \Re^{N}$ is standard Gaussian noise.

Signal model

The signal belongs to one out of C classes, where $\Pr(c) = 1/C$ and $p(\mathbf{x} \mid c) = \mathcal{N}(0, \Sigma_c)$ and $\text{rank}(\Sigma_c) = k,\ \text{c} = 1,\ldots,C.$

Measurement and signal constraints

1. $\|\Phi\|_2^2 \le 1$
2. $\mathrm{E}\left\{\|\mathbf{x}\|^2\right\} \le M$

Signal-to-noise ratio is meaningful: *snr*=1/σ²

Definitions: Classification Capacity

Definitions:

$$\lim_{M\to\infty}\frac{N^{(M)}}{M}=\upsilon \qquad \lim_{M\to\infty}\frac{k^{(M)}}{M}=\kappa \qquad \lim_{M\to\infty}\frac{\log_2 C^{(M)}}{M}=\rho$$

Definition (Achievable Classification Rate)

A classification rate ρ is said to be achievable if there exists a sequence of classification problems such that $\lim_{M\to\infty}\rho^{(M)}=\rho$ and $\lim_{M\to\infty}P_e^{(M)}=0$.

Definition (Classification Capacity)

The classification capacity $C\left(\upsilon,\kappa,1/\sigma^2\right)$ is defined to be the supremum over achievable rates ρ.

Remarks

1. Number of classes cannot be greater than $C\approx 2^{M\cdot\rho}$ for successful classification
2. Notion asymptotic in problem dimensionalities

Definitions: Diversity-Discrimination Tradeoff

Definition (DDT Function)

A sequence of classification problems is said to have diversity-discrimination function $d(r)$ if

$$\lim_{1/\sigma^2 \to \infty} \frac{\log_2 C^{(1/\sigma^2)}}{1/2 \cdot \log_2 \left(1/\sigma^2\right)} = r \qquad \lim_{1/\sigma^2 \to \infty} \frac{\log_2 P_e^{(1/\sigma^2)}}{1/2 \cdot \log_2 \left(1/\sigma^2\right)} = -d$$

Definition (DDT Tradeoff)

The diversity-discrimination tradeoff is defined as the supremum over all diversity-discrimination functions:

$$d^*(r) = \sup d(r)$$

Remarks:

1. A classification problem with $snr^{r/2}$ classes cannot have misclassification probability decaying faster than $snr^{-d^*(r)}$
2. Notion asymptotic in the signal-to-noise ratio rather than signal dimensionalities

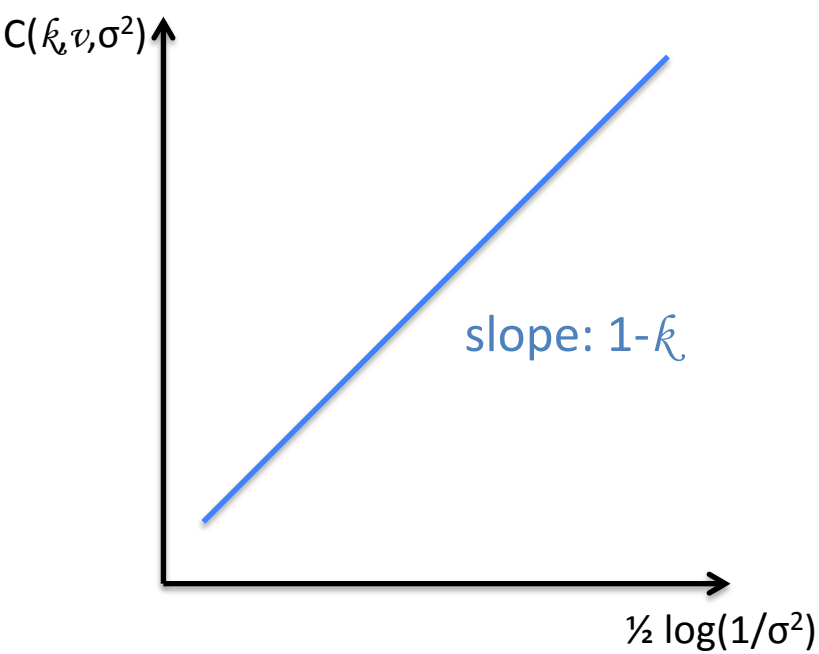
Classification Capacity

Theorem:

Consider the classification of zero-mean Gaussian signals. Then, the classification capacity is bounded as follows:

$$\frac{1-\kappa}{2}\cdot\log_2\left(\frac{1}{\sigma^2}\right)-1+\kappa \le C\left(\kappa,\upsilon,\sigma^2\right) \le \frac{1-\kappa}{2}\cdot\log_2\left(\frac{1}{\sigma^2}\right)+\frac{1}{2}\cdot\log_2\left(\frac{2}{\pi^2 e}\right)+\frac{1}{2\log(2)}$$

Interpretation



Spirit of the Proof:

1. Inspired by the dualities between the classification problem and the non-coherent multiple-antenna communications problem, together with converse and achievability arguments.
2. Subtleties: signal dimensions rather than some sort of block length approaches infinity prevents the use of random code construction via i.i.d. sequences and the use of standard typical sequence decoding arguments.

Proof Outline:

Converse

The **upper bound** follows from the fact that the classification capacity of any classification problem is upper bounded by:

$$C\left(\kappa,\upsilon,\sigma^2\right) \leq \lim_{M\to\infty}\sup \frac{1}{M}\cdot I\left(c;\mathbf{y}\right)$$

and from upper bounds to the mutual information in **non-coherent multiple-antenna fading channels**.

Achievability

The **lower bound** follows a random coding argument, where one constructs random classes inspired by training schemes over **non-coherent multiple-antenna fading channels**:

$$\Phi = \begin{bmatrix} \mathbf{I}_{k\times k} & 0 \\ 0 & \Phi' \end{bmatrix} \qquad \Sigma_c = \begin{bmatrix} \mathbf{I}_{k\times k} & 0 \\ 0 & \mathbf{U}_c\mathbf{U}_c^t \end{bmatrix}$$

where $\mathbf{U}\in\Re^{N-k\times k}$ contains Gaussian i.i.d. entries and $\Phi'\in\Re^{M-k\times N-k}$ contains orthonormal rows

Direct analysis of the error probability using a “channel estimation” and “outage“ type of argument leads to the lower bound.

Diversity-Discrimination Tradeoff

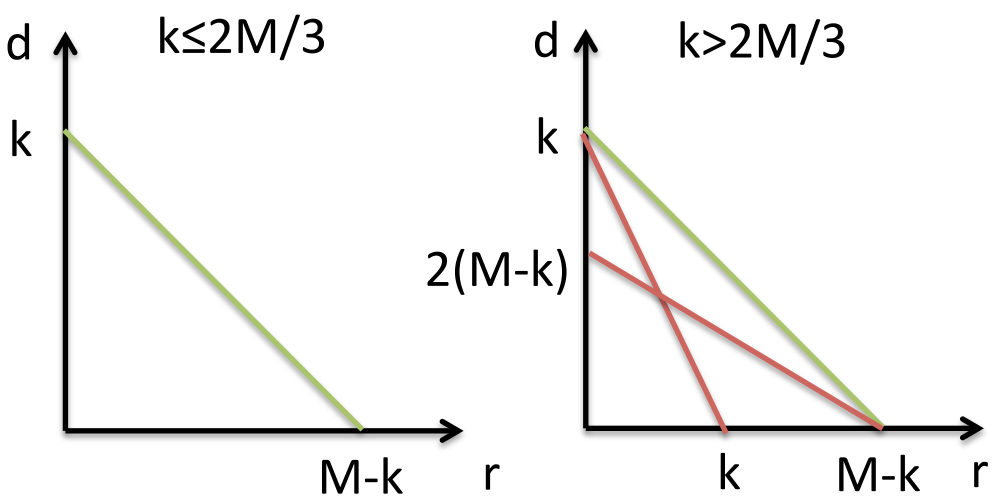
Theorem:

Consider the classification of zero-mean Gaussian signals. Then, the diversity-discrimination tradeoff is bounded as follows:

k≤2M/3: $$k\cdot\left[1-\frac{r}{M-k}\right]^{+}\le d(r)\le k\cdot\left[1-\frac{r}{M-k}\right]^{+}$$

k>2M/3: $$\left[\max\left\{k-r;2\cdot(M-k-r)\right\}\right]^{+}\le d(r)\le k\cdot\left[1-\frac{r}{M-k}\right]^{+}$$

Interpretation

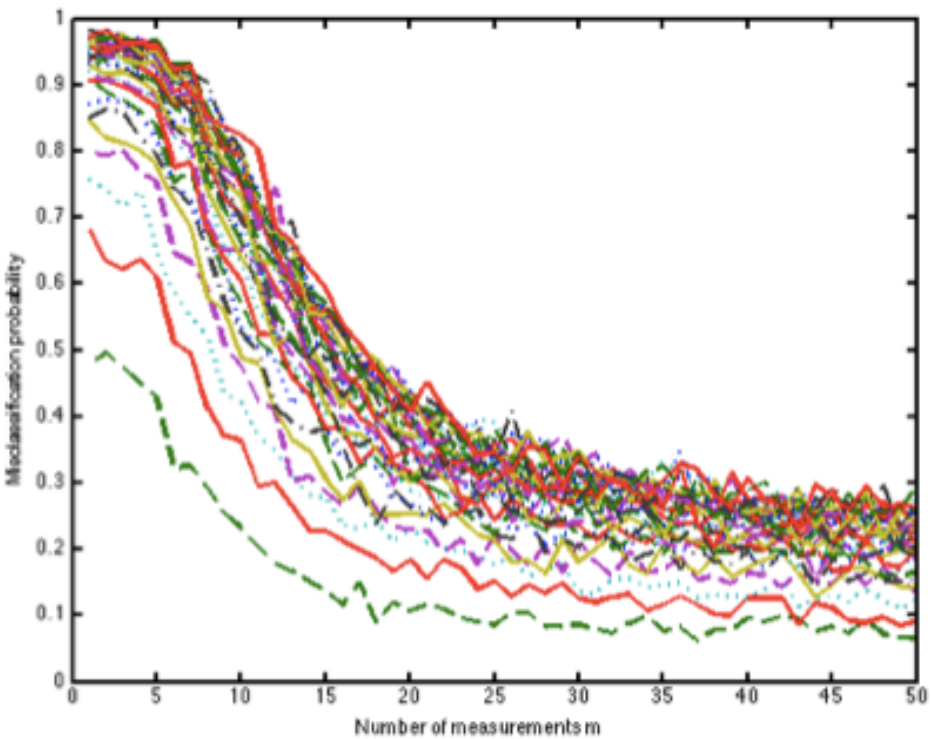


Spirit of the Proof:

1. Inspired by the dualities between the classification problem and the non-coherent multiple-antenna communications problem, together with converse and achievability arguments.
2. Outer bound based on the analysis of the outage probability.
3. Inner bound based on the analysis of the error probability of a "training" scheme.

Theory meets Practice: Face Recognition

- 38 cropped faces from the extended Yale Face Database B
- Database contains a few dozen greyscale pthotographs for each face (under a variety of illuminations)
- Measurement matrix design chosen to maximize signal power
- Face classification via a compressive variation of nine points of light (9PL) algorithm – involves both training and classification

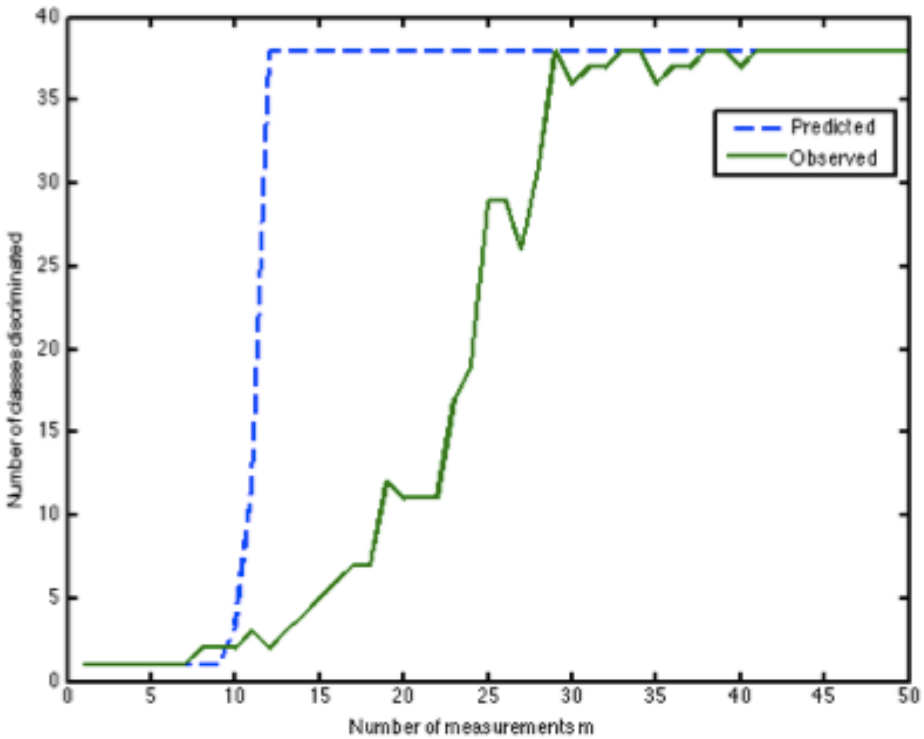


$$\hat{c} = \underset{c}{\operatorname{argmin}} \left\| \mathbf{y} - \Pi_{\Phi \mathbf{U}_c} \mathbf{y} \right\| = \underset{c}{\operatorname{argmin}} \left\| \Phi \mathbf{x} - \Pi_{\Phi \mathbf{U}_c} \Phi \mathbf{x} \right\|$$

Theory meets Practice: Face Recognition

Number of discernible classes
(Theoretical Prediction)

$$\max\left\{1, \min\left\{snr^{\frac{m-9}{2}}, 38\right\}\right\}$$



Concluding Remarks

- Dualities between the classification and communications problem lead to new characterizations of the performance of compressive classifiers:
  - Phase transitions, diversity gain and measurement gain
  - Classification capacity and diversity-discrimination tradeoff

- These characterizations determine the interplay between performance and geometry, as well as tradeoffs between "volume of data" and "performance".

- Theory with reasonable agreement with practice in a face recognition application – in spite of the use of stylized model.

Concluding Remarks

- A final remark:
  - In communications nature designs the channel and the engineer designs the input (codewords)
  - In classification nature determines the input (subspaces) and the engineer determines the channel (measurement system)

- The characterizations, which unveil optimal geometries, can provide guidelines to dictionary design.