A General Framework For Reconstruction and Classification From Compressive Measurements with Side Information

Liming Wang*, Xin Yuan*, Francesco Renna†, Miguel R.D. Rodrigues†, Robert Calderbank* and Lawrence Carin*

*Department of Electrical and Computer Engineering, Duke University, Durham, NC
†Department of Electronic and Electrical Engineering, University College London, London

Introduction

- Dimensionality reduction plays a pivotal role in various machine-learning applications, including compressive feature extraction, supervised/unsupervised dimensionality reduction.
- The signal of interest is often accompanied by additional information. How to leverage side information for the following tasks:
  - Reconstruction
  - Classification
  - Reminiscent of well-known challenges in source-coding with side information (Wyner-Ziv bound) and distributed source coding (Slepian-Wolf bound).

System Model and Problem Statements

- \( x_1 \in \mathbb{R}^n \) denotes the principal signal and \( x_2 \in \mathbb{R}^m \) denotes side information.
- \( C_1 = \{1, \ldots, K_1\} \) and \( C_2 = \{1, \ldots, K_2\} \) represent latent underlying indicator variables associated with \( x_1 \) and \( x_2 \), respectively.
- \( C_1 \) and \( C_2 \) are drawn from an arbitrary discrete joint distribution \( p(C_1, C_2); x_1 \) and \( x_2 \) are drawn from a mixture model \( p(x_1, x_2) = \sum_{k=1}^{K} p(C_1 = i, C_2 = k) p(x_1 | C_1 = i, C_2 = k) p(x_2 | C_1 = i, C_2 = k) \).
- Principal signal: \( y_1 \in \mathbb{R}^m \), where typically \( m \ll n \).
- Side information: \( y_2 \in \mathbb{R}^m \), with \( m \ll n \).

Joint GMM Signal Model

- Conditioned on \( (C_1, C_2) = (i, k) \), the source distribution is \( p(x_1, x_2 | i, C_2) = \mathcal{N}(\mu_{ix}, \Sigma_{ix}) \), where
  \[
  \mu_{ix} = \sum_{k=1}^{K} a_{ik} \mu_{xk}, \quad \Sigma_{ix} = \sum_{k=1}^{K} a_{ik} \Sigma_{xk},
  \]
- Model can be compacted expressed as
  \[
  y = \Phi x + w, \quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, \quad w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix},
  \]
- We present a theoretical result fully characterizing the behavior of MMSE\(E(r^2)\) in the low-noise regime, i.e., when \( \sigma^2 \to 0 \).
- It is interesting to note that the necessary conditions for the phase transition of the MMSE are one feature away from the corresponding sufficient conditions. Following theorem provides a sharp characterization of the region associated to the phase transition of the MMSE of the joint GMM.
- The performance of MMSE estimator can be established via the following classification result.

Joint GMM Signal Model (Cont’d)

Theorem

Assume the joint GMM. If

\[
\begin{align*}
\mu_{ix} > \mu_{ix}^* & \quad \text{or} \\
\mu_{ix} > \mu_{ix}^* - \alpha (\Sigma_{ix}^{1/2}) & \quad \forall i, k \in S,
\end{align*}
\]
then, with probability \( 1 \), we have \( \lim_{\rho \to 0} \text{MMSE}(r^2) = 0 \), where \( S = (i, k) \in \{1, \ldots, K_1\} \times \{1, \ldots, K_2\} \),

\[
\begin{align*}
p(C_1 = i, C_2 = k) > 0. \quad \Sigma_{ix}^{1/2} = \text{rank}(\Sigma_{ix}^{1/2}), \\
\mu_{ix} > \mu_{ix}^* - \alpha (\Sigma_{ix}^{1/2}) \quad \text{and} \quad \mu_{ix} > \mu_{ix}^* - \alpha (\Sigma_{ix}^{1/2}).
\end{align*}
\]
Conversely, if we have \( \lim_{\rho \to 0} \text{MMSE}(r^2) = 0 \), then, with probability \( 1 \), we have

\[
\begin{align*}
\mu_{ix} > \mu_{ix}^* & \quad \text{or} \\
\mu_{ix} > \mu_{ix}^* - \alpha (\Sigma_{ix}^{1/2}) & \quad \forall i, k \in S.
\end{align*}
\]

Exponential Family Side Information Model

Assume \( p(x_1, x_2 | C_1, C_2) = p(x_1 | C_1, C_2) p(x_2 | C_1, C_2) \), where \( p(x_1 | C_1 = i, C_2 = j) \) is a point mass.

Further assume

\[
p(x_1 | C_1 = i, C_2 = j) = \sum_{s=1}^{N} \pi_{ij}^{(s)} N(\mu_{ij}^{(s)}, \Sigma_{ij}^{(s)}),
\]
where \( \pi_{ij}^{(s)} \) are the hyperparameters, \( \Sigma_{ij}^{(s)} \) are the covariance matrices, \( \mu_{ij}^{(s)} \) are the mean vectors, \( \eta_{ij} \) is the normalizing factor, and \( \theta \) is the natural parameter.

Assume above EF side information model. \( \hat{P}_{err} \) can be expressed as

\[
\begin{align*}
\hat{P}_{err} = \sum_{k_1=1}^{K_1} \sum_{k_2=1}^{K_2} \sum_{i=1}^{N} \left( \pi_{ij}^{(s)} \right) \sqrt{p(C_1 = c_1 | k_1) p(C_2 = c_2 | k_2) \phi_B(B_1, B_2)},
\end{align*}
\]
where \( B_1 \) is a function of all the model parameters associated with \( y_1 \), i.e., \( \eta_{y_1}^{(s)}, \mu_{y_1}^{(s)}, \Sigma_{y_1}^{(s)}, N_1, \Phi_1, \) and \( \sigma^2 \). \( B_2 \) can be expressed as

\[
B_2 = \sum_{i=1}^{N} \eta_{j}^{(s)}(\Phi_1 \Sigma_{y_1}^{(s)} \Phi_1) \eta_{j}^{(s)}(\Phi_2 \Sigma_{y_2}^{(s)} \Phi_2),
\]
where \( \Phi_1 \) and \( \Phi_2 \) are the log-normalizer and the natural parameter associated with the underlying exponential family.

In particular, let \( \hat{P}_{err} \) denote the Bhattacharyya misclassification upper bound when the side information is ignored, i.e., the classification is performed via the MAP classifier \( \hat{C}_1 = \arg\max_{c_1} p(C_1 | y_1) \). Then we have \( \hat{P}_{err} \leq \hat{P}_{err} \).

Experiment: GMM Signal Model

- Principal signal: RGB “castle” image with size 480 x 320 x 3.
- Side information: gray-scale “castle” image with size 120 x 80, a low-resolution version of the same subject. \((x_1, x_2)\) are assumed to follow the joint GMM, with \( K_1 = K_2 = 10 \).
- The parameters of the joint GMM are learned from other images in Caltech 101, via the expectation-maximization (EM) algorithm.

- When side information is not available, the phase transition occurs at \( m_1 > \max_{k_1} a_{1k_1} \) \( \approx 20 \), which matches the right plot in the Figure.
- The presence of side information allows for reliable reconstruction in the low-noise regime with a reduced number of features extracted from each patch. Namely, in the case under consideration, the conditions are equivalent to \( m_1 > 16 \), which is shown to match well the PSNR behavior in the left plot of the Figure.

Figure: Results of numerical experiments to demonstrate phase transition, depicting PSNR vs. 1/\( \sigma^2 \) for reconstruction of the RGB image “castle” without and with side information (\( y_2 \in \{0, \sigma^2\} \)).

- The compression sensing ratio (CSR) is defined as \( CSR = m_1 / n_1 \). It can be seen that the classification accuracies improve significantly (\( \approx 20\% \) for training and \( > 15\% \) for testing) due to the presence of side information at various compression ratios, which is consistent with conclusion of Theorem.

Figure: GMM Signal Model

- Consider the image classification problem with the text annotations of each image as side information.
- \( x_1 \) follows a GMM (for a patch model of an image).
- \( x_2 \) follows a Poisson distribution (corresponding to a count vector, associated with a text document).
- The LabelMe data: Five outdoor image classes: “coast”, “highway”, “insidecity”, “mountain”, “street” are used in our testing.
- Bag-of-words model for the side information, and the collected word counts follow \( y_2 \sim \text{Poisson}(\Phi_2 x_2) \).
- For each image class \( C_1 = i, i = 1, \ldots, K_1 \), we associate it with one Poisson rate vector \( \{ x_2^{(c)} \} \), and \( K_1 = 5 \) for this example.

Figure: Classification accuracy against compression ratio, average of 10 trials.