Solving Inverse Problems with Gaussian Mixture Models

with Guoshen Yu and Stephane Mallat

• The Return of PCA
Inverse Problems

\[ y = Uf + w \]
\[ w \sim \mathcal{N}(0, \sigma^2 I_d) \]

Examples

- Inpainting
- Deblurring
- Zooming

\[ \mathbf{U} : \text{masking} \]
\[ \mathbf{U} : \text{subsampling} \]
\[ \mathbf{U} : \text{convolution} \]
Sparse Paradigm

- A signal $f \in \mathbb{R}^N$ and a dictionary $D \in \mathbb{R}^{N \times |\Gamma|}$

$$f = f_\Lambda + \epsilon_\Lambda = Da + \epsilon_\Lambda$$

$|\Lambda| \ll |\Gamma|$, $\Lambda = \text{support}(a)$ and $\|\epsilon_\Lambda\|^2 \ll \|f\|^2$
Gaussian Mixture Models

\[ y_i = U_i f_i + w_i \quad \text{where} \quad w_i \sim \mathcal{N}(0, \sigma^2 I d) \]

- K Gaussian distributions \( \{\mathcal{N}(\mu_k, \Sigma_k)\}_{1 \leq k \leq K} \)
- \( f_i \sim \mathcal{N}(\mu_k, \Sigma_k) \)
Gaussian Mixture Models

\[ y_i = U_i f_i + w_i \quad \text{where} \quad w_i \sim \mathcal{N}(0, \sigma^2 I_d) \]

- Estimate \( \{ (\mu_k, \Sigma_k) \}_{1 \leq k \leq K} \) from \( \{ y_i \}_{1 \leq i \leq I} \)
- Identify the Gaussian \( k_i \) that generates \( f_i \), \( \forall i \)
- Estimate \( \tilde{f}_i \) from \( \mathcal{N}(\mu_{k_i}, \Sigma_{k_i}) \), \( \forall i \)
MAP-EM Algorithm

• Iterate between E- and M-steps.

• E-step
  
  • Assume \( \{(\tilde{\mu}_k, \tilde{\Sigma}_k)\}_{1 \leq k \leq K} \) known.
  
  • Estimate \( \tilde{\kappa}_i \) (clustering) and \( \tilde{f}_i \) with MAP, \( \forall i \).

• M-step
  
  • Assume \( \tilde{\kappa}_i \) (clustering) and \( \tilde{f}_i \) known, \( \forall i \).
  
  • Estimate \( \{(\tilde{\mu}_k, \tilde{\Sigma}_k)\}_{1 \leq k \leq K} \).
E-step: Signal Estimation and Clustering

- MAP (Maximum a Posteriori) estimate

\[
(\tilde{f}_i, \tilde{k}_i) = \arg \max_{f, k} \log p(f | y_i, \tilde{\Sigma}_k) \\
= \arg \min_{f, k} \left( \| U_i f - y_i \|^2 + \sigma^2 f^T \tilde{\Sigma}_k^{-1} f + \sigma^2 \log |\tilde{\Sigma}_k| \right)
\]

\[w \sim \mathcal{N}(0, \sigma^2 I_d) \quad f_i \sim \mathcal{N}(0, \Sigma_k)\]

- Piecewise linear estimation

**Linear** estimation with each Gaussian model

\[\tilde{f}^k_i = \arg \min_f \left( \| U_i f - y_i \|^2 + \sigma^2 f^T \tilde{\Sigma}_k^{-1} f \right) \quad 1 \leq k \leq K\]

\[\Rightarrow \tilde{f}^k_i = W^{k}_i y_i\]

**Nonlinear** Gaussian model selection (clustering)

\[k_i = \arg \min_k \left( \| U_i \tilde{f}^k_i - y_i \|^2 + \sigma^2 (\tilde{f}^k_i)^T \tilde{\Sigma}_k^{-1} \tilde{f}^k_i + \sigma^2 \log |\tilde{\Sigma}_k| \right)\]

\[\tilde{f}_i = \tilde{f}^{k_i}_i\]
M-step: Gaussian Model Estimation

- **ML (Maximum Likelihood) estimate**

\[
(\tilde{\mu}_k, \tilde{\Sigma}_k) = \arg \max_{\mu_k, \Sigma_k} \log p(\{\tilde{f}_i\}_{i \in C_k} | \mu_k, \Sigma_k) \quad 1 \leq k \leq K
\]

- **Empirical estimate**

\[
\tilde{\mu}_k = \frac{1}{|C_k|} \sum_{i \in C_k} \tilde{f}_i \quad \text{and} \quad \tilde{\Sigma}_k = \frac{1}{|C_k|} \sum_{i \in C_k} (\tilde{f}_i - \tilde{\mu}_k)(\tilde{f}_i - \tilde{\mu}_k)^T
\]

- **Regularization**

\[
\tilde{\Sigma}_k \leftarrow \tilde{\Sigma}_k + \varepsilon I_d
\]
GMM = Structured Sparsity

- **PCA (Principal Component Analysis)**
  
  \[ \Sigma_k = B_k S_k B_k^T \]

  - \( B_k = \{ \phi_m^k \}_{1 \leq m \leq N} \) PCA basis, orthogonal.
  
  - \( S_k = diag(\lambda_1^k, \ldots, \lambda_N^k) \), \( \lambda_1^k \geq \lambda_2^k \geq \cdots \geq \lambda_N^k \) eigenvalues.

- **PCA transform**
  
  \[ \tilde{f}_i^k = B_k \tilde{a}_i^k \]

- **MAP with PCA**
  
  \[ \tilde{f}_i^k = \arg \min_{f_i} \left( \| U_i f_i - y_i \|^2 + \sigma^2 f_i^T \tilde{\Sigma}_k^{-1} f_i \right) \]

  \[ \iff \]

  \[ \tilde{a}_i^k = \arg \min_{a_i} \left( \| U_i B_k a_i - y_i \|^2 + \sigma^2 \sum_{m=1}^N \frac{|a_i[m]|^2}{\lambda_m^k} \right) \]
\[ \tilde{a}_i = \arg \min_{a_i} \|UDa_i - y_i\|^2 + \lambda \sum_{m=1}^{|\Gamma|} |a_i[m]| \]
\[ \tilde{a}_i^k = \arg \min_{a_i} \left( \|U_kB_k a_i - y_i\|^2 + \sigma^2 \sum_{m=1}^N |a_i[m]|^2 \lambda_m^{-k} \right) \]

- Full degree of freedom in atom selection \( |\Gamma| \)

- Linear collaborative filtering in each basis/block/PCA.

- Nonlinear basis selection, degree of freedom \( K \)
Initial Experiments: Evolution

Clustering 1st iteration

Clustering 2nd iteration
Experiments: Inpainting

Original

20% available

MCA 24.18 dB
[Elad, Starck, Querre, Donoho, 05]

ASR 21.84 dB
[Guleryuz, 06]

KR 21.55 dB
[Takeda, Farsiu, Milanfar, 06]

FOE 21.92 dB
[Roth and Black, 09]

BP 25.54 dB
[Zhou, Sapiro, Carin, 10]

PLE 27.65 dB
Experiments: Inpainting

Zoom (original)

20% available 6.69 dB

PLE 30.07 dB

[K-SVD, Mairal, Elad, Sapiro, 08, 29.65 dB]

[BP, Zhou, Sapiro, Carin, 10, 29.12 dB]
Experiments: Zooming

Original

Bicubic 28.47 dB

SAI 30.32 dB

SR 23.85 dB

PLE 30.64 dB

Low-resolution

SR [Yang, Wright, Huang, Ma, 09]

SAI [Zhang and Wu, 08]
Experiments: Zooming Deblurring

$\text{f}$

$\text{Uf}$

$\text{y} = \text{SUf}$

$I_y \ 29.40 \ dB$

$\text{PLE} \ 30.49 \ dB$

$\text{SR} \ 28.93 \ dB$

[Yang, Wright, Huang, Ma, 09]
Experiments: Denoising

Original  Noisy 28.14 dB  PLE 35.37 dB
Beyond images (with G. Yu and F. Leger)

- 2.8 millions ratings
- 1,648 movies
- 74,424 users
- 4.3% data available

- 1 million ratings
- 3,900 movies
- 6,040 users
- 4.6% data available

<table>
<thead>
<tr>
<th>EachMovie</th>
<th>Weak NMAE</th>
<th>Strong NMAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>URP [10]</td>
<td>0.4422</td>
<td>0.4557</td>
</tr>
<tr>
<td>Attitude [11]</td>
<td>0.4520</td>
<td>0.4550</td>
</tr>
<tr>
<td>MMMF [14]</td>
<td>0.4397</td>
<td>0.4341</td>
</tr>
<tr>
<td>IPCF [12]</td>
<td>0.4382</td>
<td>0.4365</td>
</tr>
<tr>
<td>E-MMMF [4]</td>
<td>0.4287</td>
<td>0.4301</td>
</tr>
<tr>
<td>GPLVM [7]</td>
<td>0.4179</td>
<td>0.4134</td>
</tr>
<tr>
<td>M³F [8]</td>
<td>0.4293</td>
<td>n/a</td>
</tr>
<tr>
<td>NBMC [17]</td>
<td><strong>0.4109</strong></td>
<td><strong>0.4091</strong></td>
</tr>
<tr>
<td>GM (proposed)</td>
<td>0.4164</td>
<td>0.4163</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IM MovieLens</th>
<th>Weak NMAE</th>
<th>Strong NMAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>URP [10]</td>
<td>0.4341</td>
<td>0.4444</td>
</tr>
<tr>
<td>Attitude [11]</td>
<td>0.4320</td>
<td>0.4375</td>
</tr>
<tr>
<td>MMMF [14]</td>
<td>0.4156</td>
<td>0.4203</td>
</tr>
<tr>
<td>IPCF [12]</td>
<td>0.4096</td>
<td>0.4113</td>
</tr>
<tr>
<td>E-MMMF [4]</td>
<td>0.4029</td>
<td>0.4071</td>
</tr>
<tr>
<td>GPLVM [7]</td>
<td>0.4026</td>
<td>0.3994</td>
</tr>
<tr>
<td>M³F [8]</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>NBMC [17]</td>
<td><strong>0.3916</strong></td>
<td>0.3992</td>
</tr>
<tr>
<td>GM (proposed)</td>
<td>0.3959</td>
<td><strong>0.3928</strong></td>
</tr>
</tbody>
</table>
Sampling Theory

• Band limited signals
  • Nyquist sampling + Linear reconstruction

• Sparse signals
  • CS + Non-linear reconstruction

• GMM/PCA
  • ??? + Piecewise-linear reconstruction
Theory (G. Yu and GS)

• Statistical compressed sensing
  • Signals sampled from a distribution
  • Accurate reconstruction on average

• Signals from a Gaussian with Gaussian of Bernoulli sensing
  • $O(k)$ measurements (vs. $O(k \log(N/k))$ in standard CS)
  • Optimal MAP linear decoder
  • Reconstruction equivalent to best $k$-term approximation with high probability
  • Probability of failure $<<$ standard CS

• Considering a probabilistic RIP condition
  • Any sensing matrix
  • Error upper bounded by best $k$-term approximation with probability one
  • Bound constant efficiently computable

• GMM
  • Close formulae bounds on error of selection and error of recovery
Summary

- Gaussian mixture models and MAP-EM work well for image inverse problems.

- Piecewise linear estimation, connection to structured sparsity.
  - Collaborative linear filtering.
  - Nonlinear best basis selection, small degree of freedom.

- Faster computation than sparse estimation.
- Results in the same ballpark of the state-of-the-art.

- Beyond images: recommender systems and audio (Sprechmann & Cancela)

- Statistical compressed sensing
• For papers see arxiv.org

• Immediate openings for post-docs and long/short term visitors

Thank you!
Structured Representation and Estimation

Overcomplete dictionary

- Dictionary: union of PCAs
  - Union of \textit{orthogonal} bases \( D = \{ B_k \}_{1 \leq k \leq K} \)
  - In each basis, the atoms are \textit{ordered}: \( \lambda_1^k \geq \lambda_2^k \geq \cdots \geq \lambda_N^k \)

- Piecewise linear estimation (PLE)
  - A \textit{linear} estimator per basis
  - \textit{Non-linear} basis selection: a best linear estimator is selected

- Small degree of freedom, fast computation, state-of-the-art performance