

Solving Inverse Problems with Gaussian Mixture Models

with Guoshen Yu and Stephane Mallat

- The Return of PCA



Inverse Problems

$$\mathbf{y} = \mathbf{U}\mathbf{f} + \mathbf{w}$$

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Id})$$



Inpainting



Examples

\mathbf{f}
Zooming



Deblurring



\mathbf{U} : subsampling

2

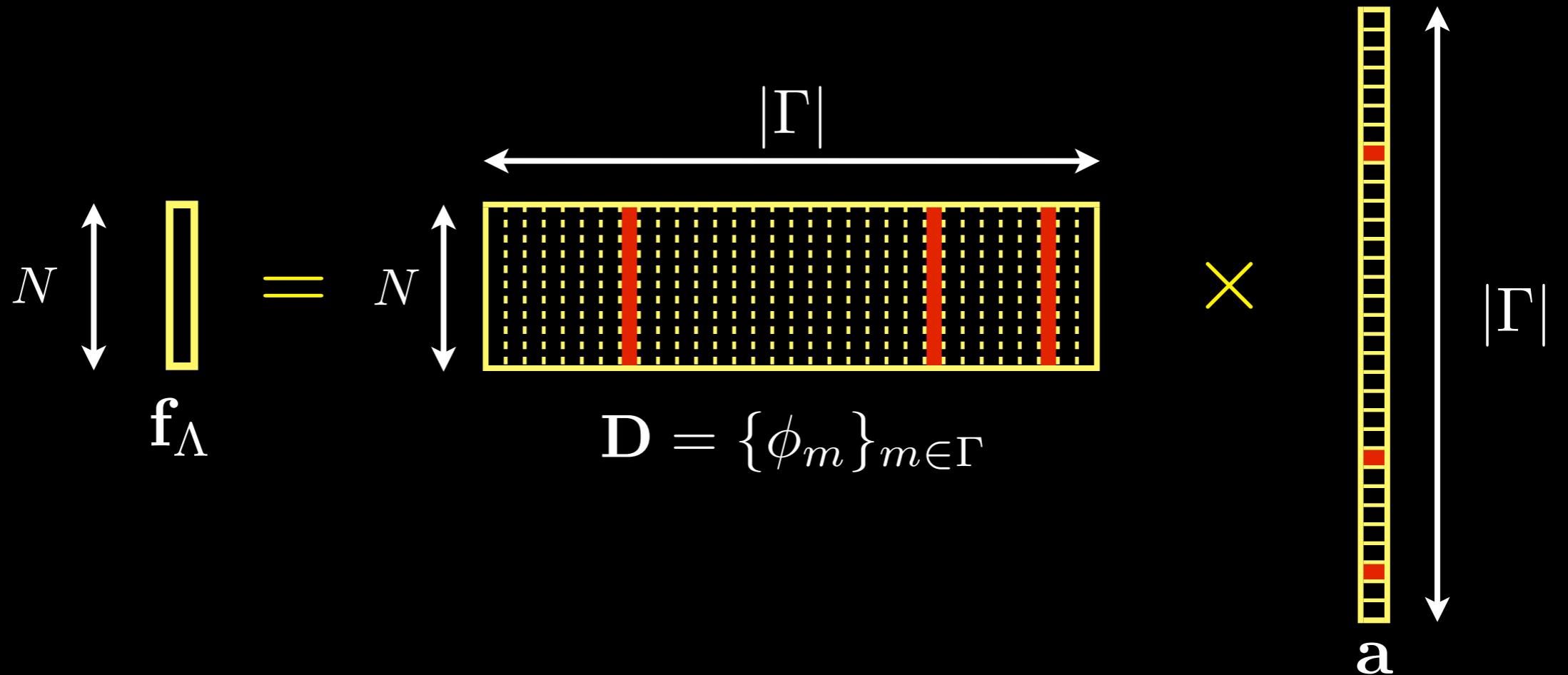
\mathbf{U} : convolution
 $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{Id})$

Sparse Paradigm

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

$$\mathbf{f} = \mathbf{f}_\Lambda + \epsilon_\Lambda = \mathbf{D}\mathbf{a} + \epsilon_\Lambda$$

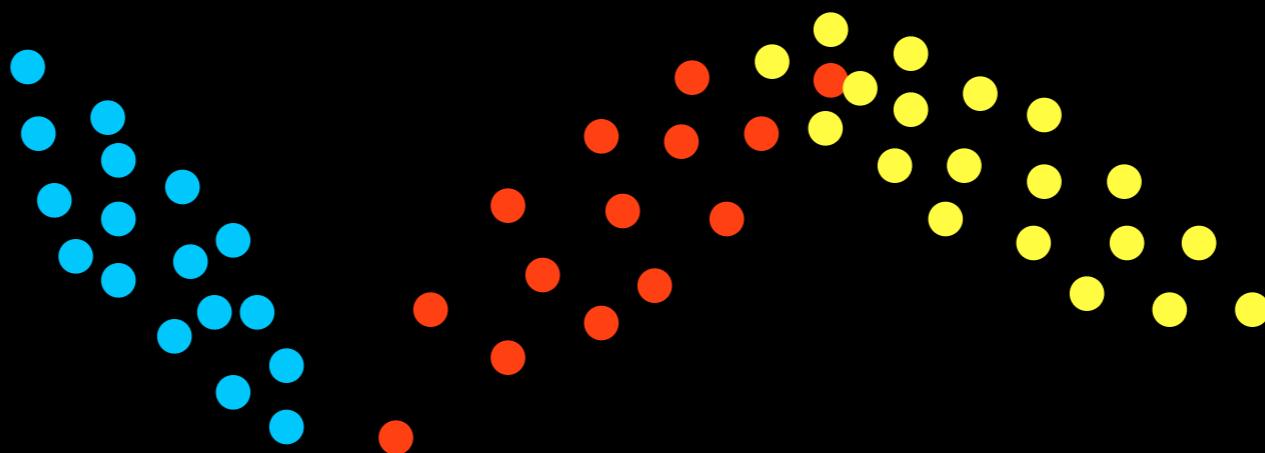
$$|\Lambda| \ll |\Gamma|, \Lambda = \text{support}(\mathbf{a}) \quad \text{and} \quad \|\epsilon_\Lambda\|^2 \ll \|\mathbf{f}\|^2$$



Gaussian Mixture Models

$$\mathbf{y}_i = \mathbf{U}_i \mathbf{f}_i + \mathbf{w}_i \quad \text{where} \quad \mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 Id)$$

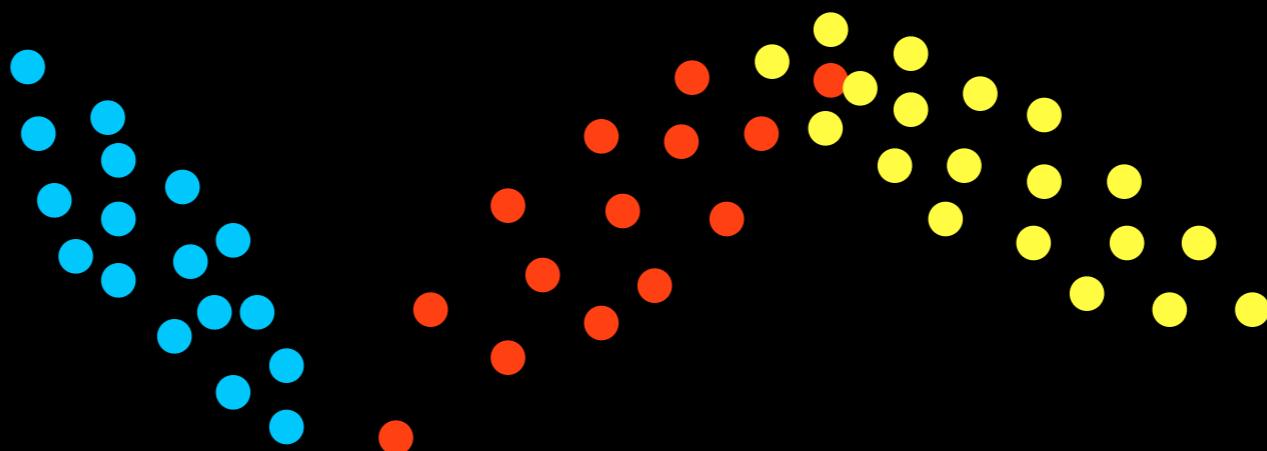
- K Gaussian distributions $\{\mathcal{N}(\mu_k, \Sigma_k)\}_{1 \leq k \leq K}$
- $\mathbf{f}_i \sim \mathcal{N}(\mu_k, \Sigma_k)$



Gaussian Mixture Models

$$\mathbf{y}_i = \mathbf{U}_i \mathbf{f}_i + \mathbf{w}_i \quad \text{where} \quad \mathbf{w}_i \sim \mathcal{N}(\mathbf{0}, \sigma^2 Id)$$

- Estimate $\{(\mu_k, \Sigma_k)\}_{1 \leq k \leq K}$ from $\{\mathbf{y}_i\}_{1 \leq i \leq I}$
- Identify the Gaussian k_i that generates $\mathbf{f}_i, \forall i$
- Estimate $\tilde{\mathbf{f}}_i$ 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{k}_i (clustering) and $\tilde{\mathbf{f}}_i$ with MAP, $\forall i$.
- M-step
 - Assume \tilde{k}_i (clustering) and $\tilde{\mathbf{f}}_i$ 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

$$\begin{aligned}(\tilde{\mathbf{f}}_i, \tilde{k}_i) &= \arg \max_{\mathbf{f}, k} \log p(\mathbf{f} | \mathbf{y}_i, \tilde{\Sigma}_k) \\&= \arg \min_{\mathbf{f}, k} \left(\frac{\| \mathbf{U}_i \mathbf{f} - \mathbf{y}_i \|^2 + \sigma^2 \mathbf{f}^T \tilde{\Sigma}_k^{-1} \mathbf{f}}{\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \sigma^2 Id)} + \frac{\sigma^2 \log |\tilde{\Sigma}_k|}{\mathbf{f}_i \sim \mathcal{N}(\mathbf{0}, \Sigma_k)} \right)\end{aligned}$$

- Piecewise linear estimation

Linear estimation with each Gaussian model

$$\begin{aligned}\tilde{\mathbf{f}}_i^k &= \arg \min_{\mathbf{f}} \left(\| \mathbf{U}_i \mathbf{f} - \mathbf{y}_i \|^2 + \sigma^2 \mathbf{f}^T \tilde{\Sigma}_k^{-1} \mathbf{f} \right) \quad 1 \leq k \leq K \\&\Rightarrow \tilde{\mathbf{f}}_i^k = \mathbf{W}_i^k \mathbf{y}_i\end{aligned}$$

Nonlinear Gaussian model selection (clustering)

$$\begin{aligned}k_i &= \arg \min_k \left(\| \mathbf{U}_i \tilde{\mathbf{f}}_i^k - \mathbf{y}_i \|^2 + \sigma^2 (\tilde{\mathbf{f}}_i^k)^T \tilde{\Sigma}_k^{-1} \tilde{\mathbf{f}}_i^k + \sigma^2 \log |\tilde{\Sigma}_k| \right) \\&\tilde{\mathbf{f}}_i = \tilde{\mathbf{f}}_i^{k_i}\end{aligned}$$

M-step: Gaussian Model Estimation

- ML (Maximum Likelihood) estimate

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

- Empirical estimate

$$\tilde{\mu}_k = \frac{1}{|\mathcal{C}_k|} \sum_{i \in \mathcal{C}_k} \tilde{\mathbf{f}}_i \quad \text{and} \quad \tilde{\Sigma}_k = \frac{1}{|\mathcal{C}_k|} \sum_{i \in \mathcal{C}_k} (\tilde{\mathbf{f}}_i - \tilde{\mu}_k)(\tilde{\mathbf{f}}_i - \tilde{\mu}_k)^T$$

- Regularization

$$\tilde{\Sigma}_k \leftarrow \tilde{\Sigma}_k + \varepsilon Id$$

GMM = Structured Sparsity

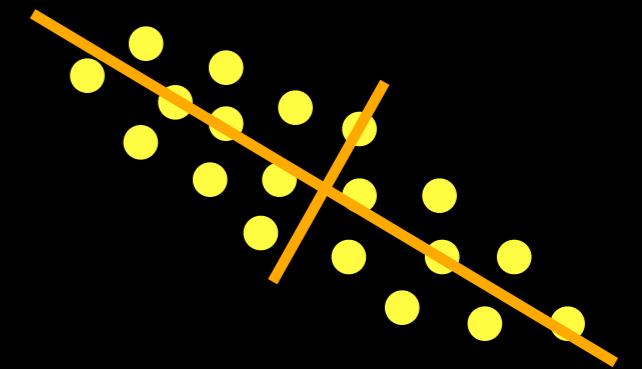
- PCA (Principal Component Analysis)

$$\Sigma_k = \mathbf{B}_k \mathbf{S}_k \mathbf{B}_k^T$$

- $\mathbf{B}_k = \{\phi_m^k\}_{1 \leq m \leq N}$ PCA basis, orthogonal.
- $\mathbf{S}_k = \text{diag}(\lambda_1^k, \dots, \lambda_N^k)$, $\lambda_1^k \geq \lambda_2^k \geq \dots \geq \lambda_N^k$ eigenvalues.

- PCA transform

$$\tilde{\mathbf{f}}_i^k = \mathbf{B}_k \tilde{\mathbf{a}}_i^k$$



- MAP with PCA

$$\tilde{\mathbf{f}}_i^k = \arg \min_{\mathbf{f}_i} \left(\|\mathbf{U}_i \mathbf{f}_i - \mathbf{y}_i\|^2 + \sigma^2 \mathbf{f}_i^T \tilde{\Sigma}_k^{-1} \mathbf{f}_i \right)$$

\Leftrightarrow

$$\tilde{\mathbf{a}}_i^k = \arg \min_{\mathbf{a}_i} \left(\|\mathbf{U}_i \mathbf{B}_k \mathbf{a}_i - \mathbf{y}_i\|^2 + \sigma^2 \sum_{m=1}^N \frac{|\mathbf{a}_i[m]|^2}{\lambda_m^k} \right)$$

Structured Sparsity

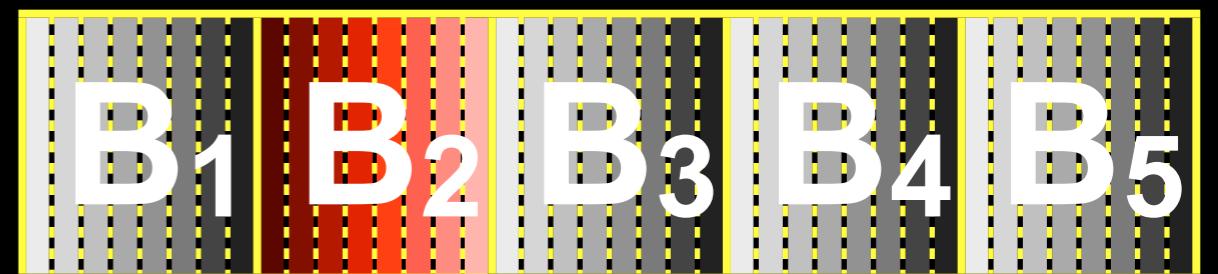
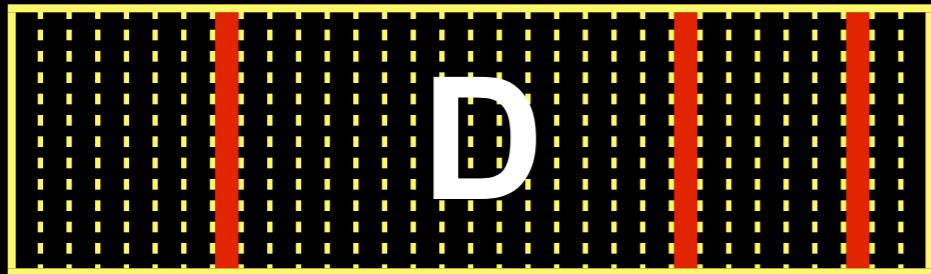
Sparse estimate

v.s.

Piecewise linear estimate

$$\tilde{\mathbf{a}}_i = \arg \min_{\mathbf{a}_i} \|\mathbf{U}\mathbf{D}\mathbf{a}_i - \mathbf{y}_i\|^2 + \lambda \sum_{m=1}^{|\Gamma|} |\mathbf{a}_i[m]|$$

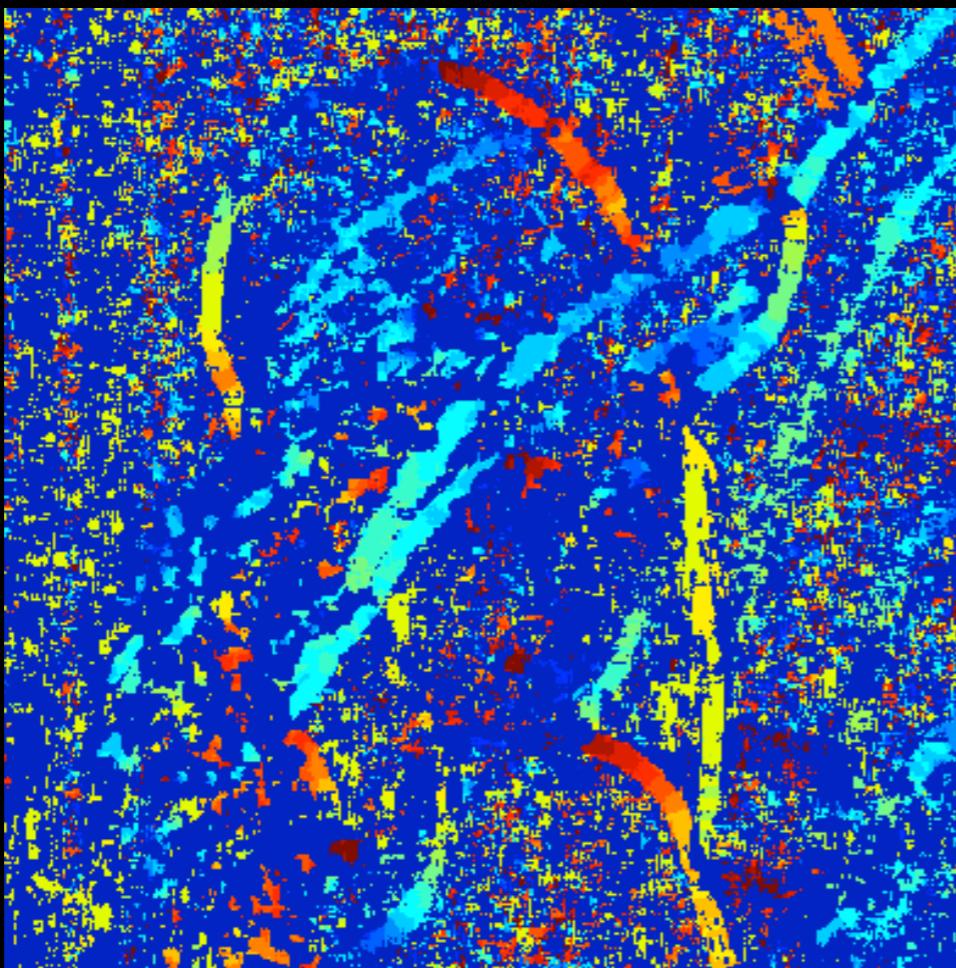
$$\tilde{\mathbf{a}}_i^k = \arg \min_{\mathbf{a}_i} \left(\|\mathbf{U}_i \mathbf{B}_k \mathbf{a}_i - \mathbf{y}_i\|^2 + \sigma^2 \sum_{m=1}^N \frac{|\mathbf{a}_i[m]|^2}{\lambda_m^k} \right)$$



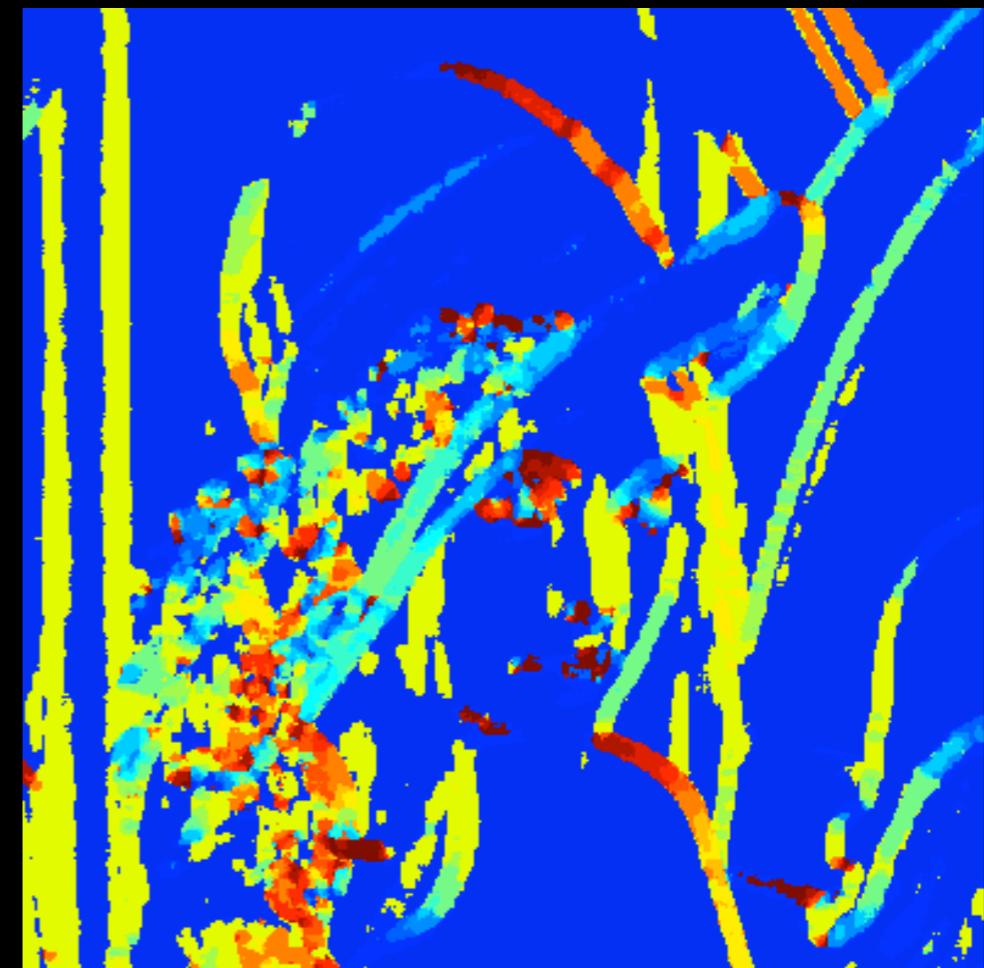
Full degree of freedom
in atom selection $\binom{|\Gamma|}{|\Lambda|}$

- Linear collaborative filtering in each basis/block/PCA.
- Nonlinear basis selection, degree of freedom K .

Initial Experiments: Evolution



**Clustering
1st iteration**

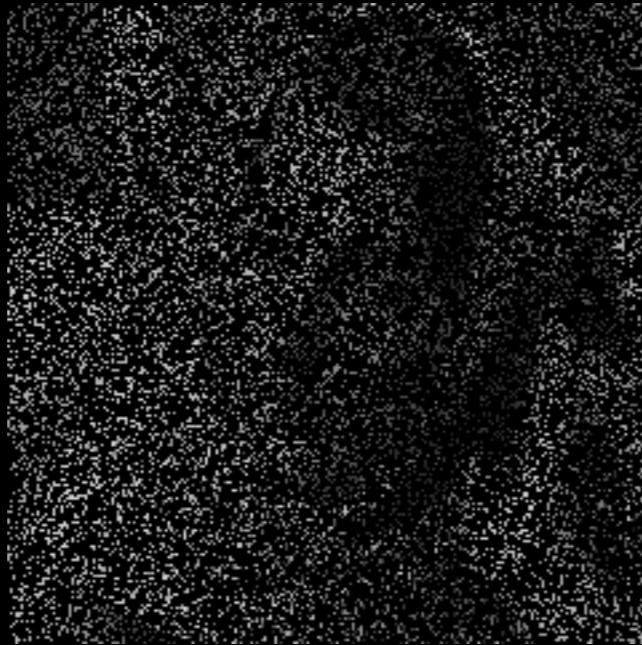


**Clustering
2nd iteration**

Experiments: Inpainting



Original



20% available



MCA 24.18 dB

[Elad, Starck, Querre, Donoho, 05]



ASR 21.84 dB

[Guleryuz, 06]



FOE 21.92 dB

[Takeda, Farsiu, Milanfar, 06]



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



Low-resolution



Original

Bicubic 28.47 dB

SAI 30.32 dB

SR 23.85 dB

PLE 30.64 dB

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

SAI [Zhang and Wu, 08]

Experiments: Zooming Deblurring



f



Uf



$y = SUf$



Iy 29.40 dB



PLE 30.49 dB



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

<i>EachMovie</i>	Weak NMAE	Strong NMAE
URP [10]	0.4422	0.4557
Attitude [11]	0.4520	0.4550
MMMF [14]	0.4397	0.4341
IPCF [12]	0.4382	0.4365
E-MMMF [4]	0.4287	0.4301
GPLVM [7]	0.4179	0.4134
M ³ F [8]	0.4293	n/a
NBMC [17]	0.4109	0.4091
GM (proposed)	0.4164	0.4163

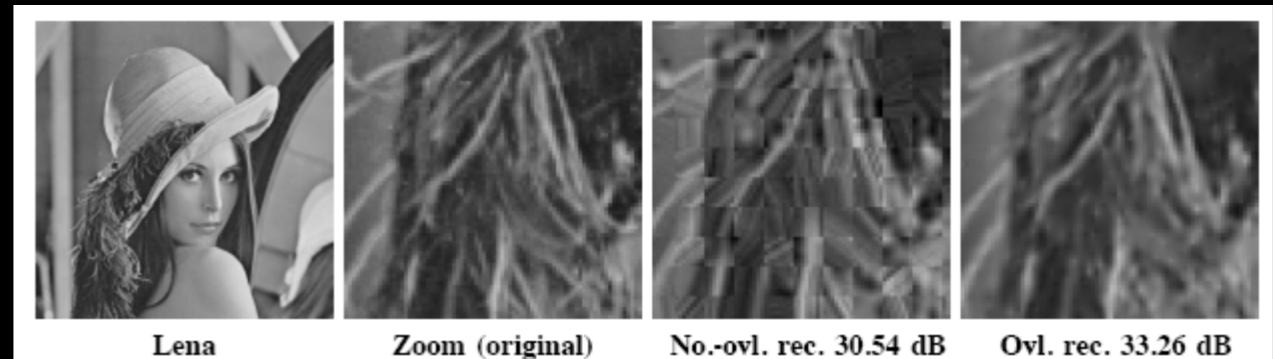
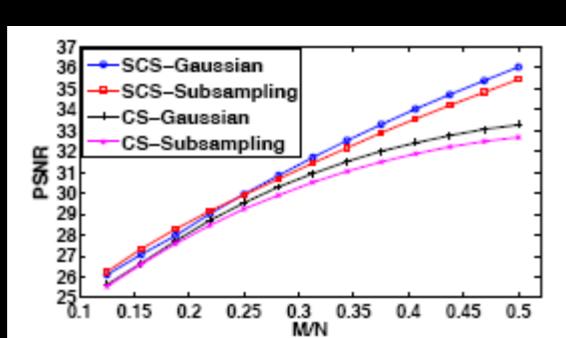
<i>IM MovieLens</i>	Weak NMAE	Strong NMAE
URP [10]	0.4341	0.4444
Attitude [11]	0.4320	0.4375
MMMF [14]	0.4156	0.4203
IPCF [12]	0.4096	0.4113
E-MMMF [4]	0.4029	0.4071
GPLVM [7]	0.4026	0.3994
M ³ F [8]	n/a	n/a
NBMC [17]	0.3916	0.3992
GM (proposed)	0.3959	0.3928

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 or 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

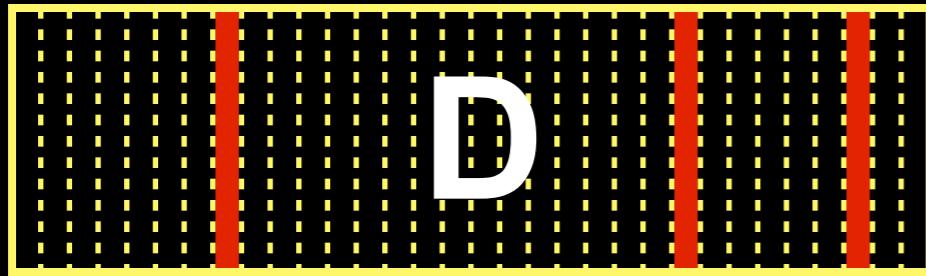
Summary

- 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

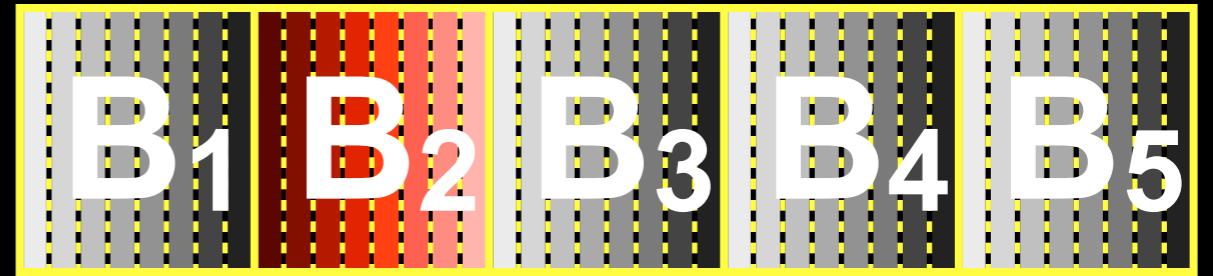
- For papers see arxiv.org
- Immediate openings for post-docs and long/short term visitors

Thank you!

Structured Representation and Estimation



Overcomplete dictionary



Structured overcomplete dictionary

- Dictionary: union of PCAs
 - Union of **orthogonal** bases $D = \{B_k\}_{1 \leq k \leq K}$
 - In each basis, the atoms are **ordered**: $\lambda_1^k \geq \lambda_2^k \geq \dots \geq \lambda_N^k$
- Piecewise linear estimation (PLE)
 - A **linear** estimator per basis
 - **Non-linear** basis selection: a best linear estimator is selected
- Small degree of freedom, fast computation, state-of-the-art performance