Separable Cosparse Analysis Operator Learning

Julian Wörmann

In collaboration with Matthias Seibert, Rémi Gribonval, and Martin Kleinsteuber

Technische Universität München
Dep. Electrical Engineering and Information Technology
Research Group for Geometric Optimization and Machine Learning

September 3rd, 2014
Contents

Sparse signal models

Separable Analysis Operator Learning

Experiments and conclusion
Informative signals are structured
Informative signals are structured
Informative signals are structured
Structure can be captured in adequate signal representation
Synthesis versus Analysis

\[
s \approx Dz
\]

\[
\alpha \approx \Omega s
\]
Unsupervised Analysis Operator Learning

Learn analysis operator from given set of training samples \( \{s_i\}_{i=1}^T \)

\[
\begin{align*}
\text{minimize} \quad & \sum_{i=1}^{T} g(\Omega s_i) \\
\text{with the sparsity promoting function} \\
\end{align*}
\]

\[
g(\alpha) := \sum_{k} \log \left( 1 + \nu \alpha_k^2 \right).
\]

The operator \( \Omega \) is restricted to a set of constraints \( \mathcal{C} \)
Modeling the constraints

- **normalized rows** $||\omega_i|| = 1$

  Feasible set has product of spheres structure (oblique manifold)
  \[
  \Omega \in \text{OB}(m, n) := \{ \Omega \in \mathbb{R}^{m \times n} : (\Omega \Omega^\top)_{ii} = 1, i = 1, \ldots, m \} 
  \]

  → enforced by optimization

- **full rank**

  We can enforce full rank with
  \[
  h(\Omega) = -\frac{1}{n \log(n)} \log \det \left( \frac{1}{m} \Omega^\top \Omega \right) 
  \]

  → controls the condition number of the operator

- **no row repetitions** $\omega_k \neq \pm \omega_l, k \neq l$

  \[
  r(\Omega) = -\sum_{k < l} \log(1 - (\omega_k^\top \omega_l)^2) 
  \]

  → related to the mutual coherence of the operator
Separable structure constraint

\[ \Omega = \left( \Omega^{(1)} \otimes \ldots \otimes \Omega^{(N)} \right) \]

Separable filters dramatically reduce the numerical complexity.
Separable Co-sparse Analysis Operator Learning

\[ A = \Omega^{(1)} S \Omega^{(2)\top} \]
\[ \iff \text{vec}(A) = \left( \Omega^{(2)} \otimes \Omega^{(1)} \right) \cdot \text{vec}(S) \]

- If the rows of \( \Omega^{(1)} \) and \( \Omega^{(2)} \) have unit norm, then the rows of \( \Omega^{(2)} \otimes \Omega^{(1)} \) have unit norm as well.
  \( \rightarrow \) each \( \Omega^{(i)} \) is an element of the oblique manifold

- The rank of \( \Omega^{(2)} \otimes \Omega^{(1)} \) is the product of the rank of \( \Omega^{(1)} \) and \( \Omega^{(2)} \)
  \( \rightarrow \) apply full rank penalty on each \( \Omega^{(i)} \)

- If the operators \( \Omega^{(1)} \) and \( \Omega^{(2)} \) do not exhibit trivially linearly dependent rows, then neither does \( \Omega^{(2)} \otimes \Omega^{(1)} \)
  \( \rightarrow \) coherence of the Kronecker product of each \( \Omega^{(i)} \) is equal to the maximum of the individual mutual coherences
Separable Co-sparse Analysis Operator Learning

- Enforcing all the constraints on the operator components $\Omega^{(i)}$ is equivalent to enforcing them on the *separable* operator $\left(\Omega^{(1)} \otimes \ldots \otimes \Omega^{(N)}\right)$.

- Learning a separable operator via:

$$\Omega^{(i)} \ast \in \arg \min_{\Omega^{(i)}} f(\Omega^{(1)}, \ldots, \Omega^{(i)}, \ldots, \Omega^{(N)}) \quad \text{for } i = 1, \ldots, N$$

$$f(\Omega^{(1)}, \ldots, \Omega^{(i)}, \ldots, \Omega^{(N)}) = \sum_{j=1}^{T} g(S_j \times_1 \Omega^{(1)} \ldots \times_N \Omega^{(N)})$$

$$+ \mu \sum_{j=1}^{N} r(\Omega^{(j)}) + \kappa \sum_{j=1}^{N} h(\Omega^{(j)})$$

subject to: $\Omega^{(i)} \in \text{OB}(m_i, n_i), \quad i = 1, \ldots, N.$

$g$: sparsity objective $\quad r$: incoherence penalty $\quad h$: full-rank penalty
Optimization on manifolds

Optimization task is tackled using a geometric conjugate gradient on manifolds approach\(^1\) \(^2\)

- Euclidean gradient is projected onto the manifold
- Search direction determined in tangent space
- Optimization along geodesics

---


Reconstruction of volumetric MRI signals

Given: \( \{ \Omega^{(1)}, \Omega^{(2)}, \Omega^{(3)} \} \), with \( \Omega^{(i)} \in \mathbb{R}^{6 \times 5} \) learned from 20,000 training examples.
Reconstruction with standard conjugate gradient optimization.

MRI volume data reconstruction from measurements corrupted by additive white Gaussian noise with standard deviation \( \sigma_{\text{noise}} \).

<table>
<thead>
<tr>
<th>Method</th>
<th>( \sigma_{\text{noise}} )</th>
<th>\textit{PSNR (dB)}</th>
<th>\textit{MSSIM}</th>
<th># entries in ( \Omega )</th>
<th>time factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>AKSVD(^1)</td>
<td>5</td>
<td>38.41</td>
<td>0.968</td>
<td>27,000</td>
<td>\approx 50</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>28.83</td>
<td>0.798</td>
<td>216 \times 125</td>
<td></td>
</tr>
<tr>
<td>our method</td>
<td>5</td>
<td>38.55</td>
<td>0.971</td>
<td>90</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>30.64</td>
<td>0.851</td>
<td>( 3 \times (6 \times 5) )</td>
<td></td>
</tr>
</tbody>
</table>

Reconstruction quality from Gaussian noise corrupted measurements.

Conclusion

- Interesting alternative to the synthesis model
- Separable structure of the filters is highly desirable for computational efficiency
- The separability constraint is easily integrable into the manifold optimization framework
- Properties of the operator can be enforced directly during optimization

Preprint and MATLAB Code (will be provided soon) can be found at:

www.gol.ei.tum.de/