Sparse Approximation of Signals with Highly Coherent Dictionaries

Bishnu P. Lamichhane and Laura Rebollo-Neira

b.p.lamichhane@aston.ac.uk, rebollol@aston.ac.uk

Support from EPSRC (EP/D062632/1) is acknowledged

http://www.epsrc.co.uk

"New Trends and Directions in Harmonic Analysis, Approximation Theory, and Image Analysis", Inzell Summer School, Germany, Sept. 17–21, 2007
Outline

1. Sparse Approximation of a Signal
   - Introduction
   - Optimized Orthogonal Matching Pursuit
   - Adaptive Biorthogonalization
   - Swapping
   - Controlling the Condition Number

2. Oblique Matching Pursuit
   - Oblique Projections
   - Structured Noise Filtering

3. Conclusion
Sparse Approximation

- The problem of non-linear approximation consists of representing a signal $f$ as a linear superposition of minimal number of basis functions, called \textit{atoms} selected from a redundant set called a \textit{dictionary}, $\mathcal{D} = \{\alpha_i\}_{i \in \mathcal{J}}$.
  We want to minimize $\|x\|_0$, so that $\sum_{i \in \mathcal{J}} \alpha_i x_i = f$ or for given $\epsilon > 0$, we want to obtain $\|\sum_{i \in \mathcal{J}} \alpha_i x_i - f\| \leq \epsilon$. (NP-hard and highly non-linear problem)

- Matching pursuit methodologies yield a trade-off between the optimality and computational complexity.

- Working with a highly coherent \textit{dictionary} the subset selected by matching pursuit methods can have a large condition number.

- We want to control the condition number in a step-wise manner.
Optimized Orthogonal Matching Pursuit

Given a signal $f \in \mathcal{H}$. We select a set of atoms $\{\alpha_{\ell_i}\}_{i=1}^{k-1}$ spanning the space $V_{k-1}$. The new index $l_k$ is the maximizer over all $n \in J \setminus J_{k-1}$ of

$$|\langle \gamma_n, f \rangle| \text{ (OMP)}, \quad \text{or} \quad \frac{|\langle \gamma_n, f \rangle|}{\|\gamma_n\|}, \quad \|\gamma_n\| \neq 0, \text{ (OOMP)}, \quad \gamma_n = \alpha_n - P_{V_n} \alpha_n,$$

where $J_{k-1} = \{\ell_1, \cdots, \ell_{k-1}\}$. The orthogonal projection of the signal $f$ onto $V_{k-1}$

$$P_{V_{k-1}} f = \sum_{i=1}^{k-1} \alpha_{\ell_i} \langle \beta_{i}^{k-1}, f \rangle = \sum_{i=1}^{k-1} \beta_{i}^{k-1} \langle \alpha_{\ell_i}, f \rangle = \sum_{i=1}^{k-1} c_{i}^{k-1} \alpha_{\ell_i},$$

where the set $\{\alpha_{\ell_i}\}_{i=1}^{k-1}$ is biorthogonal to $\{\beta_{i}^{k-1}\}_{i=1}^{k-1}$, and $V_{k-1} = \text{span}\{\beta_{i}^{k-1}\}_{i=1}^{k-1}$.

OOMP criterion is based on minimizing the norm of the residual at each step.

$$\|R_{k-1}\|^2 = \|R_k\|^2 + \frac{|\langle \gamma_n, f \rangle|^2}{\|\gamma_n\|^2}.$$
Recursive Biorthogonalization

We want to update and downdate the orthogonal projection when an atom is added to the set \( \{ \alpha_{\ell_i} \}_{i=1}^{k-1} \), and an atom is deleted from the set \( \{ \alpha_{\ell_i} \}_{i=1}^{k} \).

- **Forward step:** After selecting the \( k \)th atom we have \( \{ \alpha_{\ell_i} \}_{i=1}^{k} \). The biorthogonal functions \( \{ \beta_{i}^{k} \}_{i=1}^{k} \) can be constructed from \( \{ \beta_{i}^{k-1} \}_{i=1}^{k-1} \) using

\[
\beta_{i}^{k} = \frac{\gamma_{\ell_{k}}}{\| \gamma_{\ell_{k-1}} \|^2}; \quad \gamma_{\ell_{k}} = \alpha_{\ell_{k}} - P_{V_{k-1}} \alpha_{\ell_{k}},
\]

\[
\beta_{i}^{k} = \beta_{i}^{k-1} - \beta_{i}^{k} \langle \alpha_{\ell_{k}}, \beta_{i}^{k-1} \rangle, \quad i = 1, \ldots, k.
\]

- **Backward step:** The biorthogonal set \( \{ \beta_{i}^{k} \}_{i=1}^{k} \) spanning \( V_{k} \) is given and one atom, say \( \alpha_{\ell_{j}} \), is to be removed from \( V_{k} \). Denote by \( V_{k \setminus j} \) the reduced subspace \( V_{k \setminus j} = \text{span}\{ \alpha_{\ell_{1}}, \ldots, \alpha_{\ell_{j-1}}, \alpha_{\ell_{j+1}}, \ldots, \alpha_{\ell_{k}} \} \). The biorthogonal set is to be modified according to the equation

\[
\beta_{i}^{k \setminus j} = \beta_{i}^{k} - \frac{\beta_{j}^{k} \langle \beta_{j}^{k}, \beta_{i}^{k} \rangle}{\| \beta_{j}^{k} \|^2}, \quad i = 1, \ldots, j - 1, j + 1, \ldots, k.
\]
The orthogonal projector onto $V_k = V_{k-1} \cup \{\alpha_{\ell_k}\}$, $P_{V_k}$, is then given by

$$P_{V_k} = \sum_{i=1}^{k} \alpha_{\ell_i} \langle \beta_i^k, \cdot \rangle = \sum_{i=1}^{k} \beta_i^k \langle \alpha_{\ell_i}, \cdot \rangle.$$ 

and the orthogonal projector onto $V_k \setminus j$ by

$$P_{V_k \setminus j} = \sum_{\substack{i=1 \\ i \neq j}}^{k} \alpha_{\ell_i} \langle \beta_i^k \setminus j, \cdot \rangle = \sum_{\substack{i=1 \\ i \neq j}}^{k} \beta_i^k \setminus j \langle \alpha_{\ell_i}, \cdot \rangle.$$ 

The coefficients are to be modified as when including the atom $\alpha_{\ell_k}$

$$c_{i}^{k} = \langle \beta_{i}^{k}, f \rangle = c_{i}^{k-1} - c_{k}^{k} \langle \beta_{i}^{k-1}, \alpha_{k} \rangle, \quad i = 1, \ldots, k - 1.$$ 

and when deleting an atom $\alpha_{\ell_j}$

$$c_{i}^{k \setminus j} = \langle \beta_{i}^{k \setminus j}, f \rangle = c_{i}^{k} - \frac{\langle \beta_{i}^{k}, \beta_{j}^{k} \rangle c_{j}^{k}}{||\beta_{j}^{k}||^2}, \quad i = 1, \ldots, j - 1, j + 1, \ldots, k.$$
Swapping

Swapping is based on comparing whether removing one already chosen atom and including a new one improve the approximation. Backward and forward steps are made easy with adaptive biorthogonalization.

**Backward step:** Downdate $P_{V_k} f$ to $P_{V_{k\setminus j}} f$
The index to be removed is the minimizer over all $n = 1, \ldots, k$

$$\frac{|c_n^k|}{||\beta_n^k||}.$$

**Forward step:** Update $P_{V_{k-1}} f$ to $P_{V_k} f$
The new index $l_k$ is the maximizer over all $n \in J_{k-1}$ of

$$|\langle \gamma_n, f \rangle|\text{ (OMP)}, \quad \frac{|\langle \gamma_n, f \rangle|}{||\gamma_n||}, \quad ||\gamma_n|| \neq 0, \quad \text{(OOMP)},$$

where $J_{k-1} = \{\ell_1, \ldots, \ell_{k-1}\}$. 
Theoretical Bound

OMP selects only the optimal atoms if $\| \Phi_{opt}^\dagger \Phi_{npt} \|_1 < 1$ (Tropp 2004). The bound is obtained from the inequality $\frac{\| \Phi_{npt} R_n f \|_\infty}{\| \Phi_{opt} R_n f \|_\infty} < 1$, where $\Phi_{opt}$ and $\Phi_{npt}$ form sets of optimal and non-optimal atoms, and $R_n f = f - P_{V_n} f$. Using $\langle \phi, R_n f \rangle = \langle R_n \phi, f \rangle$, we get $\frac{\| R_n \Phi_{npt} f \|_\infty}{\| R_n \Phi_{opt} f \|_\infty} < 1$, which gives better bound.

Figure 1: Signal independent and signal dependent check for OMP and OOMP
More Coherent B-spline Dictionary

The theoretical bound does not apply at all.

- Signals are generated by selecting atoms randomly and combining them linearly.
- The signal is exactly recovered with swapping.
- Without swapping only the signals up to sparsity 80 are exactly recovered.

Figure 2: Exact signal recovery using and not using swapping operation.
Controlling the Condition Number

- Restrict the search space with \( \| \gamma_n \| \geq \epsilon_1 \) and stop when \( \frac{|\langle \gamma_n, f \rangle|}{\| \gamma_n \|} \leq \epsilon_2 \), \( \gamma_n = \alpha_n - P_{V_k} \alpha_n \). Both depend on the dictionary and the signal.

- If \( \tilde{w}_k = [\beta^k_1, \cdots, \beta^k_k] \), \( \frac{1}{\| \tilde{w}_k \|} \) gives the smallest singular value of the selected basis. The search can be broken if it goes beyond some limit.

- Optimal trade-off between the condition number and the approximation? Need a cheap condition estimator.

- Implementation of the matching pursuit based on the incremental QR-decomposition and the biorthogonal functions \( \Rightarrow \) the use of incremental norm estimator proposed by Duff and Voemel.

- Assume \( w_k = [\alpha_{l_1}, \cdots, \alpha_{l_k}] \), then \( \tilde{w}_k^T w_k = \mathbf{I}_k \).
  Suppose that \( \tilde{w}_k = Q_k R_k \) (\( QR \)-decomposition of \( w_k \)).
  The condition number of the basis \( w_k = \| \tilde{w}_k \| \| w_k \| = \| R_k^{-1} \| \| R_k \| \).
### Controlling the Condition Number

- $W_k = [\alpha_{l_1}, \ldots, \alpha_{l_k}] = Q_k R_k$. Update the QR-decomposition of $W_k$ to get the $QR$-decomposition of $W^*_k = [\alpha_{l_1}, \ldots, \alpha_{l_k}, \alpha_n]$. Hence $W^*_k = Q^*_k R^*_k$ with

$$R^*_k = \begin{pmatrix} R_k & v^*_k \\ 0 & r^*_k \end{pmatrix}.$$  

- Assume that $\sigma_k = \| R_k z \|$ be the norm of the matrix with $\| z \| = 1$. The idea of incremental norm estimator is to maximize the quantity $\| R^*_k \hat{z} \|$ with $\hat{z} = (sz, c)^T$, and $s^2 + c^2 = 1$. An analytical expression for the maximum of $\| R^*_k \hat{z} \|$ is easily obtained [Duff and Voemel].

- $\| (R^*_k)^{-1} \| = 1/\sigma^*_k$ with $\sigma^*_k$ being the smallest singular value of the matrix $R^*_k$. The same incremental norm estimator can be used to compute the smallest singular value of $R^*_k$. 


Controlling the Condition Number

- The smallest and largest singular values of $R_k$ and corresponding right singular vectors can be cheaply computed by applying a few power iterations to $R_k$ and $R_k^{-1}$ initialized by some intelligent guess.

- Let $\mathcal{J}_k^\epsilon = \{n : |\langle \gamma_n, f \rangle| \geq \epsilon_k \parallel \gamma_n \parallel \}$. Assume that $\kappa_k^n$ the estimated condition number of the basis $[\tilde{W}_k, \alpha_n], n \in \mathcal{J}_k^\epsilon$. Select the atom with index $\ell_{k+1} = \arg \min_{n \in \mathcal{J}_k^\epsilon} \kappa_k^n$.

- Alternative way: Assume that $\tilde{\mathcal{D}}_k := \{\tilde{\alpha}_n, n \in \mathcal{J} \setminus \mathcal{J}_k \}$ is sorted out in such a way that $\{\frac{|\langle \gamma_n, f \rangle|}{\parallel \gamma_n \parallel}, n \in \mathcal{J} \setminus \mathcal{J}_k \}$ is in descending order, and $\kappa_k$ is the condition number of the basis at the $k$th step, then starting from the first atom in $\tilde{\mathcal{D}}_k$, estimate the condition number $\kappa_k^n$ of $[\tilde{W}_k, \alpha_n], n \in \mathcal{J} \setminus \mathcal{J}_k$ until $\frac{\kappa_k^n}{\kappa_k}$ is less than some fixed number.

- Swapping can be done to improve the approximation or the condition number of the basis.
Numerical Results

Given a set find a subset with a given condition number $ax^n$, $a = \{1, 2\}, n = \{1, 2, 3, 4, 5\}, x = 10$. (Old Alg. is based on SVD: Golub and Van Loan)

<table>
<thead>
<tr>
<th>Thrs</th>
<th>x</th>
<th>2x</th>
<th>$x^2$</th>
<th>$2x^2$</th>
<th>$x^3$</th>
<th>$2x^3$</th>
<th>$x^4$</th>
<th>$2x^4$</th>
<th>$x^5$</th>
<th>$2x^5$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bspline</td>
<td>Old</td>
<td>37</td>
<td>41</td>
<td>49</td>
<td>52</td>
<td>82</td>
<td>93</td>
<td>109</td>
<td>114</td>
<td>122</td>
</tr>
<tr>
<td></td>
<td>New</td>
<td>34</td>
<td>39</td>
<td>48</td>
<td>50</td>
<td>75</td>
<td>87</td>
<td>105</td>
<td>110</td>
<td>119</td>
</tr>
<tr>
<td>(143)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavelet</td>
<td>Old</td>
<td>152</td>
<td>163</td>
<td>182</td>
<td>189</td>
<td>202</td>
<td>207</td>
<td>216</td>
<td>219</td>
<td>226</td>
</tr>
<tr>
<td></td>
<td>New</td>
<td>158</td>
<td>169</td>
<td>185</td>
<td>194</td>
<td>205</td>
<td>209</td>
<td>217</td>
<td>221</td>
<td>228</td>
</tr>
<tr>
<td>(635)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3: The ratios of the thresholds (thrs/cnd) and the computed condition numbers, B-spline dictionary (left), Wavelet dictionary (right)
Numerical Results for OOMP

We apply OOMP with and without condition control to approximate the modulated chirp function $f = \exp(x) \cos(k\pi x^2)$ for $k = 5, 10, 15, 20, 25, 30, 35, 40$ and $45$. The search is stopped when the condition number is greater than $10^6$.

Figure 4: Chirp function (left) error (middle) and number of selected atoms (right) versus $k$ for the approximation of the modulated chirp function
Numerical Results for Swapping

Signals are chosen by selecting the atoms randomly and using linear combination of them.

Figure 5: Error and condition growth with swapping operation
Oblique Projection

- We are given three subspaces $V, W$ and $W^\perp$ of $H$ with $H = W \oplus W^\perp = V + W^\perp$.

- The oblique projector $E_{VW^\perp} : H \to V$ is uniquely defined with the properties $E_{VW^\perp}v = v$, $v \in V$, and $E_{VW^\perp}w = 0$, $w \in W^\perp$ if $V \cap W^\perp = \{0\}$.

- For $f \in H$, Orthogonal projection: $P_{V_k}f = \arg\min_{g \in V_k} \|f - g\|$. Oblique projection: $E_{V_kW^\perp}f = \arg\min_{g \in V_k} \|f - g\|_{P_W}$ with $\langle f, g \rangle_{P_W} = \langle f, P_Wg \rangle$.

- Following error bound holds: $\|f - P_{V_k}f\| \leq \|f - E_{V_kW^\perp}f\| \leq \frac{1}{\cos(\theta)}\|f - P_{V_k}f\|$, where $\theta$ is the angle between the spaces $V_k$ and $W$. 
Oblique Matching Pursuit

- If a signal \( f \in \mathcal{H} \) is corrupted with noise, and the noise is known to lie in \( \mathcal{W}^\perp \) with \( f = g + h \) with \( g \in \mathcal{V} \) and \( h \in \mathcal{W}^\perp \), then \( g = E_{\mathcal{V}\mathcal{W}^\perp} f \).

- In general, only a redundant set \( \mathcal{D} \) of atoms, called a dictionary, is known, which span \( \mathcal{V} \). The signal can be even corrupted with other noise.

- How to select atoms from the set \( \mathcal{D} \) to span \( g \)? Non-uniqueness of \( \mathcal{V} \) in \( \mathcal{H} = \mathcal{W} \oplus \mathcal{W}^\perp = \mathcal{V} + \mathcal{W}^\perp \).

- Use OOMP on the modified dictionary \( P_W \mathcal{D} \) to select atoms and construct oblique projector \( E_{\mathcal{V}_k \mathcal{W}^\perp} \). We call it 'Oblique matching pursuit'.

- The efficient implementation can be done by using biorthogonal functions as before which span the space \( P_W \mathcal{V}_k \).
Figure 6: Oblique matching pursuit applied to the signal corrupted with only structured noise—called background (left) and corrupted with structured and white noise $\sigma = 0.001$ (right).

Note: The direct oblique projection (computed by T-SVD) explodes in case of the signal with white noise.
Conclusion and Remark

• The recursive biorthogonal approach yields an efficient implementation for the Matching Pursuit method.

• Forward and backward basis selection can be combined to introduce swapping operation based on the minimal residual condition.

• The incremental condition estimator can be easily adapted within the recursive biorthogonal approach to deal with the ill-conditioning.

• Oblique Matching Pursuit (Weighted Matching Pursuit) can be used to filter the structured noise.

• Theoretical investigation of the swapping operation.

• Generalizing the result for the multiple selection.

• Develop a domain decomposition algorithm.

• Combining different algorithms for efficiency and flexibility.

Thank You