The Analysis Cosparse Model for Signals and Images

Raja Giryes
Computer Science Department, Technion.

The research leading to these results has received funding from the European Research Council under European Union’s Seventh Framework Program, ERC Grant agreement no. 320649.
Outline

• Classical synthesis model versus the new analysis model.
• A recipe for generating algorithms for the new model.
• New set of tools that provide a theory with uniform recovery guarantees.
• Some open problems.
Agenda

- **Analysis and Synthesis - Introduction**
- From Synthesis to Analysis
- Experimental Results
- Theoretical Guarantees
Problem Setup

• Given the following linear measurements

\[ y = Mx_0 + e, \quad y \in \mathbb{R}^m, \|e\|_2 \leq \varepsilon \]

Recover \( x_0 \in \mathbb{R}^d \) from \( y \).

• \( e \) is the noise term.
• \( M \in \mathbb{R}^{m \times d} \) is the measurement matrix \((m<d)\).
• In the noiseless case \( e = 0 \).
Synthesis Sparsity Prior

- Given that $x_0$ is $k$-sparse we can recover it stably from $y$ under some proper assumption on $M$.
- The same holds if $x_0$ has a $k$-sparse representation under certain types of dictionaries $D \in \mathbb{R}^{d \times n}: x_0 = Dz_0, \|z_0\|_0 \leq k$
Synthesis Sparsity Prior

\[ x_0 = Dz_0, \|z_0\|_0 \leq k \]

\[ x_0 = \begin{array}{c}
\end{array} \]

\[ D \]

Zero location
Non-zero location

\[ z_0, k = 5 \]
Synthesis Minimization Problem

The problem we aim at solving in synthesis is:

$$\hat{z} = \arg \min \|z\|_0 \quad \text{s.t.} \quad \|y - MDz\|_2 \leq \varepsilon$$

$$\hat{x} = D\hat{z}$$

The above can be approximated stably in a polynomial time under some proper assumptions on $M$ and $D$. 
Analysis Model

- Another way to model the sparsity of the signal is using the analysis model [Elad, Milanfar, and Rubinstein, 2007].
- Looks at the coefficients of $\Omega x_0$.
- $\Omega \in \mathbb{R}^{p \times d}$ is a given transform – the analysis dictionary.
Analysis Model - Example

- Assume $\Omega \in \mathbb{R}^{44 \times 22}$, $x_0 \in \mathbb{R}^{22}$

$$\Omega x_0 = x_0$$

Zero location
Non-zero location

$\|\Omega x\|_0 = 23$ non-zeros

$l = p - \|\Omega x\|_0 = 21$ zeros

What is the dimension of the subspace in which $x_0$ “resides”? 
Cosparseity and Corank

- We are interested in the zeros in $\Omega x_0$.
- Cosparseity $l$: $\Omega x_0$ contains $l$ zeros.
- Cosupport $\Lambda$: The zeros’ locations.
- Corank $r$: The rank of $\Omega_\Lambda$
- If $\Omega$ is in general position then $l=r$.
- In general $x_0$ resides in a subspace of dimension $d-r$. 
Synthesis versus Analysis

Synthesis

- Sparsity $k$ – non-zeros

Analysis

- Cosparsivity $l$ – zeros
Synthesis versus Analysis

**Synthesis**
- Sparsity $k$ – non-zeros
- Synthesis dictionary $D \in \mathbb{R}^{d \times n}$

**Analysis**
- Cosparse $l$ – zeros
- Analysis dictionary $\Omega \in \mathbb{R}^{p \times d}$
## Synthesis versus Analysis

<table>
<thead>
<tr>
<th>Synthesis</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sparsity</strong> $k$ – non-zeros</td>
<td><strong>Cosparsity</strong> $l$ – zeros</td>
</tr>
<tr>
<td><strong>Synthesis dictionary</strong> $D \in \mathbb{R}^{d \times n}$</td>
<td><strong>Analysis dictionary</strong> $\Omega \in \mathbb{R}^{p \times d}$</td>
</tr>
<tr>
<td><strong>Sparse representation</strong> $x = Dz$</td>
<td><strong>Cosparse representation</strong> $\Omega x$ with (at least) $l$ zeros</td>
</tr>
<tr>
<td>$z$ with (at most) $k$ non-zeros</td>
<td>$|\Omega x|_0 \leq p - l$</td>
</tr>
</tbody>
</table>
Synthesis versus Analysis

**Synthesis**
- Sparsity $k$ – non-zeros
- Synthesis dictionary $D \in \mathbb{R}^{d \times n}$
- Sparse representation $x = Dz$
- $z$ with (at most) $k$ non-zeros
- $D_i$ - $i$-th column in $D$

**Analysis**
- Cosparsity $l$ – zeros
- Analysis dictionary $\Omega \in \mathbb{R}^{p \times d}$
- Cosparse representation $\Omega x$
- $x$ with (at least) $l$ zeros
- $\Omega_i$ - $i$-th row in $\Omega$
## Synthesis versus Analysis

<table>
<thead>
<tr>
<th>Synthesis</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Support $T \subseteq {1..n}$ - indices of the non-zero elements in $z$</td>
<td>• Cosupport $\Lambda \subseteq {1..p}$ - indices of the zeros in $\Omega_x$</td>
</tr>
</tbody>
</table>
Synthesis versus Analysis

**Synthesis**
- **Support** $T \subseteq \{1..n\}$ - indices of the non-zeros in $z$
- $x \textit{ resides in a } k \textit{ dimensional subspace spanned by} \ D_T = \{D_i, i \in T\}$

**Analysis**
- **Cosupport** $\Lambda \subseteq \{1..p\}$ - indices of the zeros in $\Omega x$
- $x \textit{ resides in a } d-r \textit{ dimensional subspace orthogonal to the subspace spanned by} \ \Omega_\Lambda = \{\Omega_i, i \in \Lambda\}$
Analysis Minimization Problem

• We “work directly” with $x$ and not with its representation.
• Generates a different family of signals.
• The signals are characterized by their behavior and not by their building blocks.
• The problem we aim at solving in analysis is

$$\hat{x} = \arg \min_{x} \|\Omega x\|_0 \quad \text{s.t.} \quad \|y - Mx\|_2 \leq \varepsilon$$
Phantom and Fourier Sampling

Shepp-Logan phantom

12 sampled Fourier radial lines
Naïve recovery

Naïve recovery – $l_2$ minimization result with no prior
Shepp-Logan Derivatives

Result of applying 2D-Finite difference operator
Approximation Methods

- The analysis problem is hard just like the synthesis one and thus approximation techniques are required.
\( l_1 \) relaxation

- For synthesis

\[
\hat{z} = \arg\min_z \|z\|_1 \quad \text{s.t.} \quad \|y - MDz\|_2 \leq \varepsilon
\]

- For analysis

\[
\hat{x} = \arg\min_x \|\Omega x\|_1 \quad \text{s.t.} \quad \|y - Mx\|_2 \leq \varepsilon
\]
More Approximation Techniques

- In synthesis we have many more approximation techniques
  - OMP
  - ROMP [Needell and Vershynin 2009]
  - CoSaMP [Needell and Tropp 2009]
  - SP [Dai and Milenkovic 2009]
  - IHT [Blumensath and Davies, 2009]
  - HTP [Foucart, 2010]

- Can we convert them for the analysis framework?
Agenda

- Analysis and Synthesis - Introduction
- **From Synthesis to Analysis**
- Experimental Results
- Theoretical Guarantees
Today’s algorithms

- We present a general scheme for “translating” synthesis operations into analysis ones.
- This provides us with a recipe for “converting” the algorithms.
- The recipe is general and easy to use.
- We apply this recipe on CoSaMP, SP, IHT and HTP.
- Theoretical study of analysis techniques is more complicated.
### (Co)Sparse vectors addition

#### Synthesis
- Given two vectors $z_1$ and $z_2$ with supports $T_1$ and $T_2$ of sizes $k_1$ and $k_2$
- Task: find the support of $z_1 + z_2$
- Solution:
  $$\text{supp}(z_1 + z_2) \subseteq T_1 \cup T_2$$
- The maximal size of the support $T_1 \cup T_2$ is $k_1 + k_2$

#### Analysis
- Given two vectors $x_1$ and $x_2$ with cosupports $\Lambda_1$ and $\Lambda_2$ of sizes $l_1$ and $l_2$
- Task: find the cosupport of $x_1 + x_2$
- Solution:
  $$\text{cosupp}(x_1 + x_2) \supseteq \Lambda_1 \cap \Lambda_2$$
- The minimal size of the cosupport $\Lambda_1 \cap \Lambda_2$ is $l_1 + l_2 - p$
Orthogonal Projection

<table>
<thead>
<tr>
<th>Synthesis</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Given a representation $z$ and support $T$</td>
<td>• Given a vector $x$ and cosupport $\Lambda$</td>
</tr>
<tr>
<td>• Task: Orthogonal projection onto the subspace supported on $T$</td>
<td>• Task: Orthogonal projection onto the subspace orthogonal to the one spanned by $\Omega_\Lambda$</td>
</tr>
<tr>
<td>• Solution: keep elements in $z$ supported on $T$ and zero the rest</td>
<td>• Solution: calculate projection to the subspace spanned by $\Omega_\Lambda$ and remove it from the vector</td>
</tr>
<tr>
<td>• $z_T$</td>
<td>• $Q_\Lambda = I - \Omega_\Lambda^\dagger \Omega_\Lambda$</td>
</tr>
</tbody>
</table>
### Objective Aware Projection

**Synthesis**

- Given a vector $z$ and support $T$
- Task: Perform an objective aware projection onto the subspace spanned by $D_T$
- Solution: solve
  \[
  \arg\min_z \|y - MDz\|_2^2 \quad \text{s.t.} \quad z_{T^c} = 0
  \]
- The last has a simple closed form solution $(MD_T)^\dagger z$

**Analysis**

- Given a vector $x$ and cosupport $\Lambda$
- Task: Perform an objective aware projection onto the subspace orthogonal to the one spanned by $\Omega_\Lambda$
- Solution: solve
  \[
  \arg\min_x \|y - Mx\|_2^2 \quad \text{s.t.} \quad \Omega_\Lambda x = 0
  \]
- The last has a closed form solution
(Co)Support Selection

### Synthesis
- Given a non sparse vector $z$
- Task: Find the support $T$ of the closest $k$-sparse vector
- Solution: Select the indices of largest $k$ elements in $z$
  - $T = \text{supp}(z, k)$
- Optimal solution.

### Analysis
- Given a non cosparse vector $x$
- Task: Find the cosupport $\Lambda$ of the closest $l$-cosparse vector
- Solution: Select the indices of smallest $l$ elements in $\Omega x$
  - $\Lambda = \text{cosupp}(\Omega x, l)$
- This solution is suboptimal
Cosupport Selection

- Task: Find the cosupport $\Lambda$ of the closest $l$-cosparse vector to $x$
  \[ \Lambda = \arg \min_{\Lambda} \| x - Q_\Lambda x \|_2^2 \]

- In some cases efficient algorithms exist:
- Simple thresholding for the unitary case.
- Dynamic programming for the one dimensional finite difference operator ($\Omega_{1D-DIF}$) [Han et al., 2004] and the Fused-Lasso operator ($\Omega_{FUS} = [\Omega_{1D-DIF}; I]$) [Giryes et al., 2013].
Near Optimal Cosupport Selection

• In general the cosupport selection problem for a general $\Omega$ is an NP-hard problem [Gribonval et al., 2013].

• A cosupport selection scheme $\hat{S}_l$ is near optimal with a constant $C_l$ if for any vector $v \in \mathbb{R}^d$

$$\left\| Q_{\hat{S}_l(z)} v - v \right\|_2^2 \leq C_l \inf_{\tilde{x} \text{ is } l\text{-cosaprse}} \left\| \tilde{x} - v \right\|_2^2$$

Reminder: $Q_{\hat{S}_l(z)} v = \left( I - \Omega_{\hat{S}_l(z)}^\dagger \Omega_{\hat{S}_l(z)} \right) v$
Cosupport Selection Open Problems

- Find new types of analysis operators with optimal or near optimal cosupport selection schemes.
- An efficient global near optimal cosupport selection scheme for a given $C_l$. 
Subspace Pursuit (SP)

\( r^0 = y, t = 0 \)
\( T = \emptyset \)

**Find new support elements**
\[ T_\Delta = \text{supp}(D^* M^* r^{t-1}, k) \]

\( t = t + 1 \)

**Update the residual**
\[ r^t = y - M\hat{x}^t \]

**Output**
\( \hat{x} = \hat{x}^t \)
\( \hat{x}^t = D\hat{z}^t \)

**Support update**
\[ \tilde{T} = T \cup T_\Delta \]

**Calculate temporal representations**
\[ z_p = (MD_{\tilde{T}})^\dagger y \]
\[ = \arg \min \| y - MDz \|_2^2 \]
\[ \text{s.t. } z_{T^c} = 0 \]

**Estimate k-sparse support**
\[ T = \text{supp}(z_p, k) \]

[Dai and Milenkovic 2009]
Analysis SP (ASP)

\[ r^0 = y, t = 0 \]
\[ \Lambda = [1..p] \]

Find new cosupport elements:
\[ \Lambda_\Delta = \hat{S}_{al} \left( \Omega M^* r^{t-1} \right) \]

Cosupport update:
\[ \tilde{\Lambda} = \Lambda \cap \Lambda_\Delta \]

\[ t = t + 1 \]

Update the residual:
\[ r^t = y - M\hat{x}^t \]

Compute new solution:
\[ \hat{x}' = \arg\min_x \|y - Mx\|_2 \]
\[ \text{s.t. } \Omega_{\tilde{\Lambda}} x = 0 \]

Output: \( \hat{x} = \hat{x}' \)

Calculate temporal solution:
\[ x_p = \arg\min_x \|y - Mx\|_2 \]
\[ \text{s.t. } \Omega_{\tilde{\Lambda}} x = 0 \]

Estimate \( l \)-cosparse cosupport:
\[ \Lambda = \hat{S}_i \left( \Omega x_p \right) \]

[Giryes and Elad 2012]
Compressive Sampling Matching Pursuit (CoSaMP)

Find new support elements:

\[ T_\triangle = \text{supp}(D^*M^*r^{t-1}, 2k) \]

Support update:

\[ \tilde{T} = T \cup T_\triangle \]

Calculate temporal representations:

\[ z_p = (M D_{\tilde{T}})^\dagger y \]

\[ = \arg \min \| y - M D z \|^2_2 \]

s.t. \( z_{T^c} = 0 \)

Estimate k-sparse support:

\[ T = \text{supp}(z_p, k) \]

Update the residual:

\[ r^t = y - M \hat{x}^t \]

Output:

\[ \hat{x} = \hat{x}^t \]

\[ \hat{x}^t = D \hat{z}^t \]

Output:

\[ x = \hat{x} \]

\[ t = t + 1 \]

\[ r^0 = y, t = 0 \]

\[ T = \emptyset \]
Analysis CoSaMP (ACoSaMP)

\[ r^0 = y, t = 0 \]
\[ \Lambda = [1..p] \]

Find new cosupport elements:
\[ \Lambda_\Delta = \hat{S}_{al} \left( \Omega M^* r^{t-1} \right) \]

Cosupport update:
\[ \tilde{\Lambda} = \Lambda \cap \Lambda_\Delta \]

Update the residual:
\[ r^t = y - M\hat{x}^t \]

Calculate temporal solution:
\[ x_p = \arg\min_{x} \| y - Mx \|_2 \]
\[ \text{s.t. } \Omega_{\tilde{\Lambda}} x = 0 \]

Output: \[ \hat{x} = \hat{x}' \]

Compute new solution:
\[ \hat{x}' = Q_{\Lambda} x_p \]

Estimate \( l \)-cosparse cosupport:
\[ \Lambda = \hat{S}_i \left( \Omega x_p \right) \]

[Giynes and Elad 2012]
AIHT and AHTP

- Applying the same scheme on iterative hard thresholding (IHT) and hard thresholding pursuit (HTP) result with their analysis versions AIHT and AHTP.

[Giryes, Nam, Gribonval and Davies 2011]
Algorithms Variations

- Relaxed ACoSaMP (RACoSaMP) and relaxed ASP (RASP)

\[
\text{arg min}_{x} \| y - Mx \|_{2}^{2} + \lambda \| \Omega_{\Lambda} x \|_{2}^{2}
\]

instead of

\[
\text{arg min}_{x} \| y - Mx \|_{2}^{2} \quad \text{s.t.} \quad \Omega_{\Lambda} x = 0
\]

- Easier to solve in high dimensional problems.

[Giryes and Elad 2012]
Other Analysis Existing Methods

- Greedy analysis pursuit (GAP) [Nam, Davies, Elad and Gribonval, 2013]
  - Selects the cosupport iteratively
  - In each iteration removes elements from the cosupport in a greedy way
- Simple Thresholding
Agenda

• Analysis and Synthesis - Introduction
• From Synthesis to Analysis
• Experimental Results
• Theoretical Guarantees
Noiseless Signal Recovery – Setup

• Synthetic noiseless compressed sensing experiment.
• $\Omega$ is a transpose of a random tight frame.
• $d=200$, $p=240$ (reminder: $\Omega \in \mathbb{R}^{p \times d}$, $M \in \mathbb{R}^{m \times d}$).
• $M$ is drawn from a random Gaussian ensemble.
• We draw a phase diagram: a grid of $\delta = m/d$ versus $\rho = (d - l)/m$.
• We repeat each experiment 50 times.
• We select $a=1$. 
High Dimensional Image Recovery

- $\Omega$ is the 2D-finite differences operator.
- Recovering the *Shepp-Logan Phantom*.
- Few radial lines from two dimensional Fourier transform.
- $15/12$ radial lines suffice for RACoSaMP/RASP - less than $5.77\%/4.63\%$ of the data.
- Performance similar to GAP.
- Better than $l_1$. 


High Dimensional Image Recovery

Shepp-Logan phantom

12 sampled radial lines
High Dimensional Image Stable Recovery

- An additive zero-mean white Gaussian noise.
- SNR of 20.
- Naïve recovery PSNR = 18dB
- Using 22 radial lines for RASP and 25 radial lines for RACoSaMP.
- PSNR of 36dB.
- Same performance as GAPN.
High Dimensional Stable Recovery

Naïve recovery

RASP recovery
Agenda

- Analysis and Synthesis - Introduction
- From Synthesis to Analysis
- Experimental Results
- Theoretical Guarantees
Restricted Isometry Property (RIP)

- We say that a matrix $M$ satisfies the RIP \cite{Candes2006} with constants $\delta_k$ if for every $k$-sparse vector $u$ it holds that

$$
(1 - \delta_k) \|u\|_2^2 \leq \|Mu\|_2^2 \leq (1 + \delta_k) \|u\|_2^2.
$$
Algorithm Guarantees

- Given that $z_0$ is a $k$-sparse vector and that $MD$ satisfies the Restricted Isometry Property (RIP) with a constant $\delta_{bk} \leq \delta_{\text{algorithm}}$ then after a constant number of iterations $i^*$, SP, CoSaMP, IHT and HTP satisfy:

$$\|\hat{z}_{i^*} - z_0\|_2^2 \leq c_1 \|e\|_2^2$$

where $c_1$ is a given constant (the constant of each algorithm is different).

We say that a matrix $M$ satisfies the $\Omega$ -RIP with a constant $\delta^\Omega_l$ if for every $l$-cosparse vector $v$ it holds that

$$
(1 - \delta^\Omega_l) \|v\|_2^2 \leq \|Mv\|_2^2 \leq (1 + \delta^\Omega_l) \|v\|_2^2.
$$
Near Optimal Cosupport Selection

• Reminder:
A cosupport selection scheme $\hat{S}_l$ is near optimal with a constant $C_l$ if for any vector $v \in \mathbb{R}^d$

$$\left\| Q_{\hat{S}_l(z)} v - v \right\|_2^2 \leq C_l \inf_{\tilde{x} \text{ is } l\text{-cosapprse}} \left\| \tilde{x} - v \right\|_2^2$$

$$Q_{\hat{S}_l(z)} v = \left( I - \Omega_{\hat{S}_l(z)}^{\dagger} \Omega_{\hat{S}_l(z)} \right) v$$
Reconstruction Guarantees

- **Theorem**: Apply either ACoSaMP or ASP with $a=(2l-p)/l$, obtaining $\hat{x}^t$ after $t$ iterations. For an appropriate value of $\tilde{C} = \max(C_1, C_{2l-p})$ there exists a reference constant $\delta_2(\tilde{C}, \sigma_M^2)$ greater than zero such that if $\delta_{4l-3p}^{\Omega} < \delta_2(\tilde{C}, \sigma_M^2)$ then after a finite number of iterations $t^*$

$$\|x_0 - \hat{x}^t^*\|_2 \leq c_2 \|e\|_2,$$

where $c_2<1$ is a function of $\delta_{4l-3p}^{\Omega}, C_1, C_{2l-p}$ and $\sigma_M^2$.

[Giryes, Nam, Elad, Gribonval and Davies, 2013]
Special Cases – Optimal Projection

• Given an optimal projection scheme the condition of the theorem is simply
  \[ \delta_{4l-3p}^\Omega < 0.0156. \]

• When \( \Omega \) is unitary the RIP and \( \Omega\)-RIP coincide and the condition becomes
  \[ \delta_{4k} \left( M\Omega^* \right) < 0.0156. \]
AIHT and AHTP

- AIHT and AHTP have a similar theorem to ASP and ACoSaMP.
- Given an optimal projection scheme AIHT and AHTP have stable recovery if
  \[ \delta_{2l-p}^{\Omega} < 1/3. \]
- In the unitary case
  \[ \delta_{2k} \left( M \Omega^* \right) < 1/3. \]
\( \Omega \) -RIP Matrices

- Theorem: for a fixed \( \Omega \in \mathbb{R}^{p \times d} \), if \( M \in \mathbb{R}^{m \times d} \) satisfies for any \( z \)
  \[
P\left( \left| \| Mz \|_2^2 - \| z \|_2^2 \right| \geq \tilde{\epsilon} \| z \|_2^2 \right) \leq e^{-\frac{C_m \tilde{\epsilon}}{\delta}}
\]
  then, for any constant \( \epsilon_l > 0 \) we have \( \delta_l \leq \epsilon_l \) with probability exceeding \( 1 - e^{-t} \) if
  \[
m \geq \frac{32}{C_M \epsilon_l^2} \left( (p - l) \log \left( \frac{9p}{(p - l) \epsilon_l} \right) + t \right)
\]

- [Blumensath, Davies, 2009], [Giryes, Nam, Elad, Gribonval and Davies, 2013]
Linear Dependencies in $\Omega$

- For $\Omega$-RIP the number of measurements we need is $m = O((p-l) \log(p))$
- If $\Omega$ is in a general position then $l < d$
- $\Rightarrow$ if $p = 2d$ we have $m > d$
- If we allow linear dependencies then $l$ can be greater than $d$.
- $\Rightarrow$ if $p = 2d$ we can have $m < d$.
- Conclusion: Linear dependencies are permitted in $\Omega$ and even encouraged.
ω-RIP Open Questions

• Johnson-Lindenstrauss matrices are also RIP and ω-RIP matrices.
  • Is there an advantage for the ω-RIP over the RIP?
  • Can we find a family of matrices that has the ω-RIP but not the RIP?
• Can we get a guarantee which is in terms of $d-r$ ($r$ is the corank) instead of $p-l$?
• What is coherence in the analysis model?
Related Work

• $D$-RIP based recovery guarantees for $l_1$-minimization [Candès, Eldar, Needell, Randall, 2011], [Liu, Mi, Li, 2012].

• ERC like based recovery conditions for GAP [Nam, Davies, Elad and Gribonval, 2013] and $l_1$-minimization [Vaiter, Peyre, Dossal, Fadili, 2013].

• $D$-RIP based TV recovery guarantees [Needell, Ward, 2013]

• Thresholding denoising ($M=I$) performance for the case of Gaussian noise [Peleg, Elad, 2013].
Implications for the Synthesis Model

- Classically, in the synthesis model linear dependencies between small groups of columns are not allowed.
- However, in the analysis model, as our target is the signal, linear dependencies are even encouraged.
- Does the same hold for synthesis?
Implication for the Synthesis Model

- Signal space CoSaMP – signal recovery under the synthesis model also when the dictionary is highly coherent [Davenport, Needell, Wakin, 2013].
- An easy extension of our analysis theoretical guarantees also to synthesis [Giryes, Elad, 2013].
- The uniqueness and stability conditions for the signal recovery and the representation recovery are not the same [Giryes, Elad, 2013].
Conclusion

• New sparsity model – the analysis model.
• The zeros contain the information.
• A general recipe for converting standard synthesis techniques into analysis ones.
• New theory and tools.
• Linear dependencies are a good thing in the analysis dictionary.
• Implications on the synthesis model.
Questions?
Additional information on the AIHT and AHTP algorithms and their guarantees
Reconstruction Guarantees

- **Theorem**: Apply either ACoSaMP or ASP with \( a=(2l-p)/l \), obtaining \( \hat{x}^t \) after \( t \) iterations. If
  \[
  \left( \tilde{C}^2 - 1 \right) \sigma^2_M / \tilde{C}^2 < 1 (*),
  \]
  and \( \delta^\Omega_{4l-3p} < \delta_2 (\tilde{C}, \sigma^2_M) \), where \( \tilde{C} = \max (C_l, C_{2l-p}) \) and \( \delta_2 (\tilde{C}, \sigma^2_M) \) is a constant greater than zero whenever (*) is satisfied, then after a finite number of iterations \( t^* \)
  \[
  \left\| x_0 - \hat{x}^{t^*} \right\|_2 \leq c_2 \left\| e \right\|_2,
  \]
  where \( c_2<1 \) is a function of \( \delta^\Omega_{4l-3p}, C_l, C_{2l-p} \) and \( \sigma^2_M \).

[Giryes, Nam, Elad, Gribonval and Davies, 2013]
Iterative hard thresholding (IHT)

- Iterative hard thresholding (IHT) [Blumensath and Davies, 2009] has two steps:
  - Gradient step with step size $\mu$:
    \[
    \hat{z}_{i+1} = \hat{z}_i + \mu (MD)^T (y - MD\hat{z}_i).
    \]
  - Projection to the closest $k$-sparse vector:
    \[
    \hat{z}_{i+1} = [\hat{z}_{i+1}]_k.
    \]

$[\cdot]_k$ is a hard thresholding operator that keeps the $k$ largest elements and zeros the rest.
Hard thresholding pursuit (HTP)

• Hard Thresholding Pursuit (HTP) [Foucart, 2010] is similar to IHT but differ in the projection step

• Instead of projecting to the closest $k$-sparse vector, takes the support $T_{i+1}$ of $[\tilde{z}_{i+1}]_k$ and minimizes the fidelity term:

$$\hat{z}_{i+1} = \arg\min_z \| y - MD_{T_{i+1}} z \|^2_2$$

$$= \left( D_{T_{i+1}}^* M^* MD_{T_{i+1}} \right)^{-1} D_{T_{i+1}}^* M^* y$$
Changing step size

• Instead of using a constant step size, one can choose in each iteration the step size that minimizes the fidelity term [V. Cevher, 2011]:

\[ \| y - MD\hat{z}_{i+1} \|_2^2 \]
A-IHT and A-HTP

- **gradient step** \( \tilde{x}_{i+1} = \hat{x}_i + \mu M^T (y - M\hat{x}_i) \)
- **cosupport selection**: \( \hat{\Lambda}_{i+1} = \text{cosupp}(\Omega \tilde{x}_{i+1}, l) \)
- **Projection**:
  - For A-IHT: \( \hat{x}_{i+1} = Q_{\Lambda_{i+1}} \tilde{x}_{i+1} \)
  - **Reminder**: \( Q_{\Lambda_{i+1}} = I - \Omega_{\Lambda_{i+1}}^\dagger \Omega_{\Lambda_{i+1}} \)
  - For A-HTP:

\[
\hat{x}_{i+1} = \arg \min_{x \text{ s.t. } \Omega_{\hat{\Lambda}_{i+1}} x = 0} \| y - Mx \|_2^2 = \begin{bmatrix} Q_{\hat{\Lambda}_{i+1}} & M^T M \\ \Omega_{\hat{\Lambda}_{i+1}} & 0 \end{bmatrix}^\dagger \begin{bmatrix} Q_{\hat{\Lambda}_{i+1}} & M^T y \\ 0 & 0 \end{bmatrix}
\]

[Giureys, Nam, Gribonval and Davies 2011]
Algorithms variations

• Optimal changing stepsize instead of a fixed one:
\[
\mu_{i+1} = \arg\min_{\mu} \| y - M\hat{x}_{i+1} \|^2_2
\]

• Approximated changing stepsize
\[
\mu_{i+1} = \arg\min_{\mu} \| y - M\tilde{x}_{i+1} \|^2_2
\]

• Underestimate the cosparssity and use \( \tilde{l} \leq l \).
Reconstruction Guarantees

• **Theorem**: Apply either AIHT or AHTP with a certain constant step size or optimal step size obtaining $x^t$ after $t$ iterations. *If*

$$\left( C_l - 1 \right) \frac{\sigma_M^2}{C_l} < 1 (*)$$

and $\delta_{2l-p} < \delta_1 \left( C_l, \sigma_M^2 \right)$, where $\delta_1 \left( \tilde{C}, \sigma_M^2 \right)$ is a constant greater than zero whenever (*) is satisfied, then after a finite number of iterations $t^*$

$$\left\| x_0 - \hat{x}^t \right\|_2 \leq c_1 \| e \|_2,$$

where $c_1 < 1$ is a function of $\delta_{2l-p}, C_l$, and $\sigma_M^2$.

[Giryes, Nam, Elad, Gribonval and Davies, 2013]
Special Case – Unitary Transform

- When $\Omega$ is unitary the condition of the first theorems is simply

$$\delta_{2k} \left( M \Omega^* \right) < 1/3$$