Identification of Matrices Having a Sparse Representation

Holger Rauhut\footnote{Funded by an Individual Marie Curie Fellowship from the EU}
University of Vienna

Trends in Harmonic Analysis, Strobl
June 20, 2007

Joint work with Götz Pfander and Jared Tanner
Let $\Gamma \in \mathbb{C}^{n \times m}$.

**Problem:** Identify $\Gamma$ from its application on a small number of vectors in $\mathbb{C}^m$.

Motivation: Channel estimation problem in wireless communication and sonar.
Let $\Gamma \in \mathbb{C}^{n \times m}$.

**Problem:** Identify $\Gamma$ from its application on a small number of vectors in $\mathbb{C}^m$.

**Motivation:** Channel estimation problem in wireless communication and sonar.

If $h_\ell \in \mathbb{C}^m, \ell = 1, \ldots, m$ forms a basis of $\mathbb{C}^m$ then $\Gamma$ is completely determined by $y_\ell = \Gamma h_\ell, \ell = 1, \ldots, m$. 

Let $\Gamma \in \mathbb{C}^{n \times m}$.

**Problem:** Identify $\Gamma$ from its application on a small number of vectors in $\mathbb{C}^m$.

Motivation: Channel estimation problem in wireless communication and sonar.

If $h_\ell \in \mathbb{C}^m$, $\ell = 1, \ldots, m$ forms a basis of $\mathbb{C}^m$ then $\Gamma$ is completely determined by $y_\ell = \Gamma h_\ell$, $\ell = 1, \ldots, m$.

Without additional knowledge at least $m$ vectors $h_\ell$ are necessary to identify $\Gamma$ from $\Gamma h_\ell$, $\ell = 1, \ldots, m$. 
Let $\Gamma \in \mathbb{C}^{n \times m}$.

**Problem:** Identify $\Gamma$ from its application on a small number of vectors in $\mathbb{C}^m$.

**Motivation:** Channel estimation problem in wireless communication and sonar.

If $h_\ell \in \mathbb{C}^m$, $\ell = 1, \ldots, m$ forms a basis of $\mathbb{C}^m$ then $\Gamma$ is completely determined by $y_\ell = \Gamma h_\ell$, $\ell = 1, \ldots, m$.

Without additional knowledge at least $m$ vectors $h_\ell$ are necessary to identify $\Gamma$ from $\Gamma h_\ell$, $\ell = 1, \ldots, m$. 
A matrix dictionary is a set $\Psi = \{\psi_j\}_{j=1}^N$ of matrices $\psi_j \in \mathbb{C}^{n \times m}$ (or $\mathbb{R}^{n \times m}$).

A matrix $\Gamma$ is said to have a $k$-sparse representation if

$$\Gamma = \sum_{j \in \Lambda} x_j \psi_j, \quad \text{with } |\Lambda| = k.$$ 

Problem: Identify a $k$-sparse $\Gamma \in \mathbb{C}^{n \times m}$ from its action $\Gamma h \in \mathbb{C}^n$ on only one vector $h \in \mathbb{C}^m$. 
A matrix dictionary is a set \( \Psi = \{\psi_j\}_{j=1}^N \) of matrices \( \psi_j \in \mathbb{C}^{n \times m} \) (or \( \mathbb{R}^{n \times m} \)).

A matrix \( \Gamma \) is said to have a \( k \)-sparse representation if

\[
\Gamma = \sum_{j \in \Lambda} x_j \psi_j, \quad \text{with } |\Lambda| = k.
\]

**Problem:** Identify a \( k \)-sparse \( \Gamma \in \mathbb{C}^{n \times m} \) from its action \( \Gamma h \in \mathbb{C}^n \) on only one vector \( h \in \mathbb{C}^m \).

Identification of \( \Gamma \) amounts to determining \( x \).
A matrix dictionary is a set $\Psi = \{\psi_j\}_{j=1}^N$ of matrices $\psi_j \in \mathbb{C}^{n \times m}$ (or $\mathbb{R}^{n \times m}$).

A matrix $\Gamma$ is said to have a $k$-sparse representation if

$$\Gamma = \sum_{j \in \Lambda} x_j \psi_j, \quad \text{with } |\Lambda| = k.$$ 

**Problem:** Identify a $k$-sparse $\Gamma \in \mathbb{C}^{n \times m}$ from its action $\Gamma h \in \mathbb{C}^n$ on only one vector $h \in \mathbb{C}^m$.

Identification of $\Gamma$ amounts to determining $x$.

No knowledge on the support of $x$ is assumed except its maximal size $k$. 
A matrix dictionary is a set $\Psi = \{\psi_j\}_{j=1}^N$ of matrices $\psi_j \in \mathbb{C}^{n \times m}$ (or $\mathbb{R}^{n \times m}$).

A matrix $\Gamma$ is said to have a $k$-sparse representation if

$$\Gamma = \sum_{j \in \Lambda} x_j \psi_j, \quad \text{with } |\Lambda| = k.$$

Problem: Identify a $k$-sparse $\Gamma \in \mathbb{C}^{n \times m}$ from its action $\Gamma h \in \mathbb{C}^n$ on only one vector $h \in \mathbb{C}^m$.

Identification of $\Gamma$ amounts to determining $x$.

No knowledge on the support of $x$ is assumed except its maximal size $k$. 
Let $\Gamma = \sum_j x_j \psi_j$. Then

$$\Gamma h = \sum_j x_j \psi_j h = (\psi_1 h \mid \psi_2 h \mid \ldots \mid \psi_N h) x =: (\Psi h)x.$$  

Hence, sparse matrix identification reduces to the problem of finding a sparse representation of $\Gamma h$ in terms of the vector dictionary $(\Psi h) = (\psi_1 h \mid \ldots \mid \psi_N h) \in \mathbb{C}^{n \times N}$.

Techniques from sparse approximation and compressed sensing apply.
Let $\Gamma = \sum_j x_j \Psi_j$. Then

$$\Gamma h = \sum_j x_j \Psi_j h = (\Psi_1 h | \Psi_2 h | \ldots | \Psi_N h) x =: (\Psi h) x.$$ 

Hence, sparse matrix identification reduces to the problem of finding a sparse representation of $\Gamma h$ in terms of the vector dictionary $(\Psi h) = (\Psi_1 h | \ldots | \Psi_N h) \in \mathbb{C}^{n \times N}$.

Techniques from sparse approximation and compressed sensing apply.

In contrast to sparse approximation we have the additional freedom of choosing $h$. 
Let $\Gamma = \sum_j x_j \psi_j$. Then

$$\Gamma h = \sum_j x_j \psi_j h = (\psi_1 h \mid \psi_2 h \mid \ldots \mid \psi_N h) x =: (\Psi h) x.$$

Hence, sparse matrix identification reduces to the problem of finding a sparse representation of $\Gamma h$ in terms of the vector dictionary $(\Psi h) = (\psi_1 h \mid \ldots \mid \psi_N h) \in \mathbb{C}^{n \times N}$.

Techniques from sparse approximation and compressed sensing apply.

In contrast to sparse approximation we have the additional freedom of choosing $h$. 

Questions

- Good matrix dictionaries $\Psi = \{\psi_j\}_{j=1}^N$?
- Choice of $h$?
- Which recovery algorithm?
- Maximal sparsity $k$ that ensures identification?
Recovery algorithms

\( \ell_0 \)-minimization:

\[
\min_{x'} \|x'\|_0 \quad \text{subject to } (\Psi h)x' = \Gamma h,
\]

where \( \|x'\|_0 = \#\{j : x_j \neq 0\} \).

**Problem:** \( \ell_0 \)-minimization is NP-hard.
Recovery algorithms

$\ell_0$-minimization:

$$\min_{x'} \|x'\|_0 \quad \text{subject to } (\Psi h) x' = \Gamma h,$$

where $\|x'\|_0 = \# \{j : x_j \neq 0\}$.

**Problem:** $\ell_0$-minimization is NP-hard.

Tractable alternatives:

- **Basis Pursuit** ($\ell_1$-minimization)
- Greedy algorithms: **Orthogonal Matching Pursuit**, **Thresholding**.
- ...

Holger Rauhut

Sparse Matrix Identification 6
Recovery algorithms

\( \ell_0 \)-minimization:

\[
\min_{x'} \| x' \|_0 \quad \text{subject to } (\Psi h)x' = \Gamma h,
\]

where \( \| x' \|_0 = \# \{ j : x_j \neq 0 \} \).

Problem: \( \ell_0 \)-minimization is NP-hard.

Tractable alternatives:

- **Basis Pursuit** (\( \ell_1 \)-minimization)
- **Greedy algorithms**: Orthogonal Matching Pursuit, Thresholding.
- ...

Holger Rauhut

Sparse Matrix Identification 6
Basis Pursuit

\( \ell_1 \)-minimization

\[
\min_{x'} \| x' \|_1 \quad \text{subject to } (\Psi h)x' = \Gamma h,
\]

where \( \| x \|_1 = \sum_j |x_j| \).

Convex relaxation of \( \ell_0 \)-minimization problem.

Can be solved efficiently with convex optimization techniques.
Theorem

Let \( h \) be a non-zero vector in \( \mathbb{R}^m \).

(a) Let all entries of the \( N \) matrices \( \Psi_j \in \mathbb{R}^{n \times m}, j = 1, \ldots, N \) be chosen independently according to a standard normal distribution (Gaussian ensemble); or

(b) let all entries of the \( N \) matrices \( \Psi_j \in \mathbb{R}^{n \times m}, j = 1, \ldots, N \) be independent Bernoulli \( \pm 1 \) variables (Bernoulli ensemble).

Then there exists a constant \( c > 0 \) such that

\[
k \leq c \frac{n}{\log \left( \frac{N}{n \varepsilon} \right)}
\]

implies that with probability of at least \( 1 - \varepsilon \) all matrices \( \Gamma \in \mathbb{R}^{n \times m} \) having a \( k \)-sparse representation with respect to \( \Psi = \{\Psi_j\} \) can be recovered from \( \Gamma h \) by Basis Pursuit.
Dictionary of time-frequency-shifts

Translation and Modulation on $\mathbb{C}^n$

$$(T_p h)_q = h(p+q) \mod n \quad \text{and} \quad (M_\ell h)_q = e^{2\pi i \ell q/n} h_q.$$\

$\mathcal{G} = \{ M_\ell T_p : \ell, p = 0, \ldots, n-1 \}$ forms a basis of $\mathbb{C}^{n \times n}$.

$$(\mathcal{G} h) = (M_\ell T_p h)_{\ell, p=0, \ldots, n-1} \in \mathbb{C}^{n \times n^2}$$ is a Gabor system.

Motivation: Wireless communications and sonar. Multipath-propagation of the signal due to reflections at possibly moving scatterers causes time-delays (translation) and Doppler-shifts (modulations).
Dictionary of time-frequency-shifts

Translation and Modulation on \( \mathbb{C}^n \)

\[(T_p h)_q = h_{(p+q) \mod n} \quad \text{and} \quad (M_\ell h)_q = e^{2\pi i \ell q/n} h_q.\]

\( \mathcal{G} = \{M_\ell T_p : \ell, p = 0, \ldots, n-1\} \) forms a basis of \( \mathbb{C}^{n \times n} \).

\((\mathcal{G} h) = (M_\ell T_p h)_{\ell,p=0,\ldots,n-1} \in \mathbb{C}^{n \times n^2} \) is a Gabor system.

Motivation: Wireless communications and sonar.
Multipath-propagation of the signal due to reflections at possibly moving scatterers causes \textit{time-delays} (translation) and \textit{Doppler-shifts} (modulations).
Therefore, the channel in wireless communications and sonar can be modeled as

\[ \Gamma = \sum_{\ell,p} x_{\ell p} M_\ell T_p. \]

Usually only a small number of scatterers, hence \( x \) can be assumed sparse.
Dictionary of time-frequency-shifts

Translation and Modulation on $\mathbb{C}^n$

$$(T_p h)_q = h_{(p+q) \mod n} \quad \text{and} \quad (M_\ell h)_q = e^{2\pi i \ell q/n} h_q.$$  

$G = \{ M_\ell T_p : \ell, p = 0, \ldots, n-1 \}$ forms a basis of $\mathbb{C}^{n \times n}$.  

$(G h) = (M_\ell T_p h)_{\ell, p=0,\ldots,n-1} \in \mathbb{C}^{n \times n^2}$ is a Gabor system.

Motivation: Wireless communications and sonar.  
Multipath-propagation of the signal due to reflections at possibly moving scatterers causes time-delays (translation) and Doppler-shifts (modulations).  
Therefore, the channel in wireless communications and sonar can be modeled as

$$\Gamma = \sum_{\ell, p} x_{\ell p} M_\ell T_p.$$  

Usually only a small number of scatterers, hence $x$ can be assumed sparse.
Two Choices for $h$

- Alltop window (Strohmer and Heath) in prime dimension $n$

\[ h^A_q = \frac{1}{\sqrt{n}} e^{2\pi i q^3/n}, \quad q = 0, \ldots, n-1. \]

- Randomly generated window (for arbitrary $n$)

\[ h^R_q = \frac{1}{\sqrt{n}} \epsilon_q, \quad q = 0, \ldots, n-1, \]

where the $\epsilon_q$ are independent and uniformly distributed on the torus $\{ z \in \mathbb{C}, |z| = 1 \}$. 
Two Choices for $h$

- Alltop window (Strohmer and Heath) in prime dimension $n$

$$h^A_q = \frac{1}{\sqrt{n}} e^{2\pi i q^3 / n}, \quad q = 0, \ldots, n-1.$$  

- Randomly generated window (for arbitrary $n$)

$$h^R_q = \frac{1}{\sqrt{n}} \epsilon_q, \quad q = 0, \ldots, n-1,$$

where the $\epsilon_q$ are independent and uniformly distributed on the torus $\{ z \in \mathbb{C}, |z| = 1 \}$.  

Holger Rauhut  
Sparse Matrix Identification 10
Theorem

(a) Let $n$ be prime and $h^A$ be the Alltop window. If $k < \frac{\sqrt{n+1}}{2}$ then BP recovers from $\Gamma h^A$ all matrices $\Gamma$ having a $k$-sparse representation with respect to the time–frequency shift dictionary.

(b) Let $n$ be even and choose $h^R$ to be the random unimodular window. Let $t > 0$ and suppose

$$k \leq \frac{1}{4} \sqrt{\frac{n}{C \log(n) + t}} + \frac{1}{2} \tag{1}$$

with $C = 2 \log(4) \approx 2.77$. Then with probability at least $1 - e^{-t}$ BP recovers from $\Gamma h^R$ all matrices $\Gamma \in \mathbb{C}^{n \times n}$ having a $k$-sparse representation.

Relies on coherence estimates for the Gabor systems $G h^A$ (Strohmer and Heath) and $G h^R$ (Pfander, R, Tanner).
Assuming randomly chosen support set $\Lambda$ of $x$ and random phases $\text{sign}(x_{\ell p})$, $(\ell, p) \in \Lambda$ one can show recovery for both $h^A$ and $h^R$ under the condition

$$k \leq c \frac{n}{(\log n)^{1+u}}$$

with $c > 0$ and $u > 0$ governing the probability of success.

Based on recent results by Tropp ("Random subdictionaries of general dictionaries").
Theorem (Deterministic Support and Random Phases)

Choose $h^R$ at random. Let $\Lambda \subset \{0, \ldots, n-1\}^2$ with cardinality $|\Lambda| = k$. Let $x$ with $\text{supp}(x) = \Lambda$ and random phases $(\text{sgn}(x_\lambda))_{\lambda \in \Lambda}$ that are independent and uniformly distributed on the torus $\{z \in \mathbb{C}, |z| = 1\}$. Set $\Gamma = \sum_{(\ell, p)} x_{\ell p} M_\ell T_p \in \mathbb{C}^{n \times n}$. Let $\sigma > 8$. Then with probability at least

$$2(n - k) \exp \left( -\frac{n}{\sigma k} \right) + Ck \exp \left( -\frac{n}{16ek} \right) + 4n^{-(\sigma/4-2)}$$

Basis Pursuit recovers $\Gamma$ from $\Gamma h^R$. The constant $C \approx 1.075$.

The probability estimate becomes effective once

$$k \leq c \frac{n}{\log(n)},$$

with $c$ “slightly smaller” than $1/(16e) \approx 1/(43.49) \approx 0.023$. 

Holger Rauhut 
Sparse Matrix Identification 13
Theorem (Deterministic $\Gamma$)

Let $\Gamma \in \mathbb{C}^{n \times n}$ be $k$-sparse with respect to the time-frequency shift dictionary. Choose $h^R$ at random. Assume that

$$k \leq C \frac{n}{\log(n/\epsilon)}.$$ 

Then with probability at least $1 - \epsilon$ Basis Pursuit recovers $\Gamma$ from $\Gamma h^R$.

The proof reveals the more precise condition

$$n \geq \max\{C_1 k \log(n^2/\epsilon), k(C_2 \log(k^4/\epsilon) + C_3)\}$$

with $C_1 = 284.64$, $C_2 = 235.12$ and $C_3 = 8.35$. 

Holger Rauhut

Sparse Matrix Identification 14
Theorem (Deterministic $\Gamma$)

Let $\Gamma \in \mathbb{C}^{n \times n}$ be $k$-sparse with respect to the time-frequency shift dictionary. Choose $h^R$ at random. Assume that

$$k \leq C \frac{n}{\log (n/\epsilon)}.$$

Then with probability at least $1 - \epsilon$ Basis Pursuit recovers $\Gamma$ from $\Gamma h^R$.

The proof reveals the more precise condition

$$n \geq \max\{C_1 k \log (n^2/\epsilon), k(C_2 \log (k^4/\epsilon) + C_3)\}$$

with $C_1 = 284.64$, $C_2 = 235.12$ and $C_3 = 8.35$. 
Numerical experiments for $h^A$, $n$ prime

Horizontal axis $1/n$, vertical axis $k/n$.

Contours of success probability, 93% success rate, $1/(2 \log(n))$.

Numerical experiments suggest $k \leq \frac{n}{2 \log(n)}$ ensures recovery of most $k$-sparse $\Gamma$. 

Numerical experiments for $h^R$

Horizontal axis $1/n$, vertical axis $k/n$.
Contours of success probability, 93% success rate, $1/(2 \log(n))$.
Numerical experiments suggest $k \leq \frac{n}{2 \log(n)}$ ensures recovery of most $k$-sparse $\Gamma$.