Analysis ℓ^1-Minimization in Imaging

Jonathan Fell

Lehrstuhl C für Mathematik (Analysis)
RWTH Aachen University

Joint work with Holger Rauhut
SIAM IS16 - MS12

23.05.2016

European Research Council
Established by the European Commission

RWTH AACHEN UNIVERSITY

Compressed Sensing

Compressed Sensing

Compressed Sensing aims at solving

$$Ax = y \text{ where } x \in \mathbb{C}^N, y \in \mathbb{C}^m, A \in \mathbb{C}^{m \times N}$$

for $m \ll N$ under the assumption that x is s-sparse, i.e. $\|x\|_0 = \sharp\{i : x_i \neq 0\} \leq s$.

Compressed Sensing

Compressed Sensing

Compressed Sensing aims at solving

$$Ax = y \text{ where } x \in \mathbb{C}^N, y \in \mathbb{C}^m, A \in \mathbb{C}^{m\times N}$$

for $m \ll N$ under the assumption that x is s-sparse, i.e. $\|x\|_0 = \sharp\{i : x_i \neq 0\} \leq s$.

Basis Pursuit

Algorthmic approach:

$$\min \|x\|_0 \text{ subject to } Ax = y \tag{BP}$$

Compressed Sensing

Compressed Sensing

Compressed Sensing aims at solving

$$Ax = y \text{ where } x \in \mathbb{C}^N, y \in \mathbb{C}^m, A \in \mathbb{C}^{m \times N}$$

for $m \ll N$ under the assumption that x is s-sparse, i.e. $\|x\|_0 = \sharp\{i : x_i \neq 0\} \leq s$.

Basis Pursuit

Algorthmic approach:

$$\min \|x\|_1 \text{ subject to } Ax = y \tag{BP}$$

Compressed Sensing

Compressed Sensing

Compressed Sensing aims at solving

$$Ax = y \text{ where } x \in \mathbb{C}^N, y \in \mathbb{C}^m, A \in \mathbb{C}^{m \times N}$$

for $m \ll N$ under the assumption that x is s-sparse, i.e. $\|x\|_0 = \sharp\{i : x_i \neq 0\} \leq s$.

Basis Pursuit

Algorthmic approach:

$$\min \|x\|_1 \text{ subject to } Ax = y \tag{BP}$$

Robust approach:

$$\min \|x\|_1 \text{ subject to } \|Ax - y\|_2 \leq \eta \tag{BPDN}$$

where η is an estimate for the noise level.

Sparsity and Modifications

Union of subspaces

Basic idea behind sparsity: the signal x belongs to a union of low-dimensional subspaces:

$$x \in \Sigma_s := \{x \in \mathbb{C}^N ; \|x\|_0 \leq s\} = \bigcup_{\substack{S \subset [N] \\ \sharp S \leq s}} \{x \in \mathbb{C}^N ; \operatorname{supp}(x) \subset S\}$$

Sparsity and Modifications

Union of subspaces

Basic idea behind sparsity: the signal x belongs to a union of low-dimensional subspaces:

$$x \in \Sigma_s := \{x \in \mathbb{C}^N \,;\, \|x\|_0 \leq s\} = \bigcup_{\substack{S \subset [N] \\ \sharp S \leq s}} \{x \in \mathbb{C}^N \,;\, \mathrm{supp}(x) \subset S\}$$

Modifications

- Since x may not be sparse in the standard basis, one may employ an orthonormal operator $\Theta \in \mathrm{O}(n)$, i.e.

$$\min \|z\|_1 \text{ subject to } A\Theta^* z = y$$

which is the same as

$$\min \|\Theta x\|_1 \text{ subject to } Ax = y.$$

Sparsity and Modifications

Union of subspaces

Basic idea behind sparsity: the signal x belongs to a union of low-dimensional subspaces:

$$x \in \Sigma_s := \{x \in \mathbb{C}^N \, ; \, \|x\|_0 \leq s\} = \bigcup_{\substack{S \subset [N] \\ \sharp S \leq s}} \{x \in \mathbb{C}^N \, ; \, \mathrm{supp}(x) \subset S\}$$

Modifications

- Since x may not be sparse in the standard basis, one may employ an orthonormal operator $\Theta \in \mathrm{O}(n)$, i.e.

  $$\min \|z\|_1 \text{ subject to } A\Theta^* z = y$$

  which is the same as

  $$\min \|\Theta x\|_1 \text{ subject to } Ax = y.$$

- If x is not sparse, its *discrete gradient* ∇x often is (e.g. for images), hence we minimize

  $$\min \|\nabla x\|_1 \text{ subject to } Ax = y. \tag{TV}$$

Compressed Sensing

Restricted Isometry Property

A possesses the Restricted Isometry Property (RIP) of order s if

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2 \text{ for all } x \in \Sigma_s \text{ and some } \delta \in (0,1).$$

The smallest such δ is called the *Restricted Isometry Constant* δ_s.

Compressed Sensing

Restricted Isometry Property

A possesses the Restricted Isometry Property (RIP) of order s if

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2 \text{ for all } x \in \Sigma_s \text{ and some } \delta \in (0,1).$$

The smallest such δ is called the *Restricted Isometry Constant* δ_s.

Recovery result

If A has the RIP of order $2s$ with $\delta_{2s} \leq 0.6248$ then the minimizer $x^\sharp$ of BPDN fulfils

$$\|x - x^\sharp\|_2 \leq C\frac{\sigma_s(x)_1}{\sqrt{s}} + D\eta$$

where $\sigma_s(x)_1 = \inf_{z:\|z\|_0 \leq s} \|z-x\|_1$ is the *error of best s-term approximation*.

Compressed Sensing

Standard/Benchmark Theorem

If $m \geq C\delta^{-2} s \ln(eN/s)$, then a Gaussian random matrix A possesses the RIP of order s with constant δ with probability exceeding $1 - 2\exp(-\delta^2 m/2C)$.

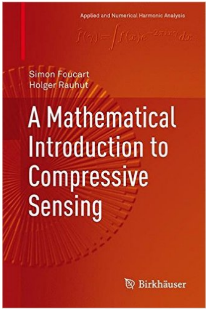
Compressed Sensing

Standard/Benchmark Theorem

If $m \geq C\delta^{-2} s \ln(eN/s)$, then a Gaussian random matrix A possesses the RIP of order s with constant δ with probability exceeding $1 - 2\exp(-\delta^2 m/2C)$.

Further information

For a thorough introduction, see Foucart and Rauhut [2013]:



The big red book

Analysis and Synthesis

Frames

Let $\Omega = \begin{pmatrix} \omega_1^t \\ \vdots \\ \omega_p^t \end{pmatrix} \in \mathbb{C}^{p \times N}$ with $p \geq N$ be a *frame* for $\mathbb{C}^N$, i.e. there exist $A, B > 0$ such that

$$A\|x\|_2^2 \leq \sum_{i=1}^{p} |\langle x, \omega_i \rangle|^2 \leq B\|x\|^2 \text{ for all } x \in \mathbb{C}^N.$$

The sequence $\{\langle x, \omega_i \rangle\}_{i=1}^{p}$ are the *analysis coefficients* of x. A frame is called *tight* if $A = B$ and *Parseval* if $A = B = 1$.

Analysis and Synthesis

Frames

Let $\Omega = \begin{pmatrix} \omega_1^t \\ \vdots \\ \omega_p^t \end{pmatrix} \in \mathbb{C}^{p \times N}$ with $p \geq N$ be a *frame* for $\mathbb{C}^N$, i.e. there exist $A, B > 0$ such that

$$A\|x\|_2^2 \leq \sum_{i=1}^{p} |\langle x, \omega_i \rangle|^2 \leq B\|x\|^2 \text{ for all } x \in \mathbb{C}^N.$$

The sequence $\{\langle x, \omega_i \rangle\}_{i=1}^p$ are the *analysis coefficients* of x. A frame is called *tight* if $A = B$ and *Parseval* if $A = B = 1$. Since $\{\omega_1, \ldots, \omega_N\}$ spans $\mathbb{C}^N$ we can expand

$$x = \sum_{i=1}^{p} c_i \omega_i \text{ for some } c_i \in \mathbb{C}.$$

The c_i are the *synthesis coefficients*. They can be computed via a *dual frame* $\Omega^\dagger \in \mathbb{C}^{N \times p}$, i.e. $c = \Omega^\dagger x$. Those are not unique in general.

Analysis and Synthesis

Frames

Let $\Omega = \begin{pmatrix} \omega_1^t \\ \vdots \\ \omega_p^t \end{pmatrix} \in \mathbb{C}^{p \times N}$ with $p \geq N$ be a *frame* for $\mathbb{C}^N$, i.e. there exist $A, B > 0$ such that

$$A\|x\|_2^2 \leq \sum_{i=1}^{p} |\langle x, \omega_i \rangle|^2 \leq B\|x\|^2 \text{ for all } x \in \mathbb{C}^N.$$

The sequence $\{\langle x, \omega_i \rangle\}_{i=1}^p$ are the *analysis coefficients* of x. A frame is called *tight* if $A = B$ and *Parseval* if $A = B = 1$. Since $\{\omega_1, \ldots, \omega_N\}$ spans $\mathbb{C}^N$ we can expand

$$x = \sum_{i=1}^{p} c_i \omega_i \text{ for some } c_i \in \mathbb{C}.$$

The c_i are the *synthesis coefficients*. They can be computed via a *dual frame* $\Omega^\dagger \in \mathbb{C}^{N \times p}$, i.e. $c = \Omega^\dagger x$. Those are not unique in general.
Remark: The canonical dual frame is $\Omega^\dagger = (\Omega^* \Omega)^{-1} \Omega^*$.

Cosparsity

Frame Minimization

Instead of basis pursuit we consider

$$\min \|\Omega x\|_1 \text{ subject to } Ax = y \tag{Ω-BP}$$

or its robust version

$$\min \|\Omega x\|_1 \text{ subject to } \|Ax - y\|_2 \leq \eta \tag{Ω-BPDN}$$

under the assumption that x is $\Omega - k$-*cosparse*, i.e. $\sharp\{i\,;\, \langle \omega_i, x\rangle \neq 0\} \leq p - k$

Cosparsity

Frame Minimization

Instead of basis pursuit we consider

$$\min \|\Omega x\|_1 \text{ subject to } Ax = y \tag{Ω-BP}$$

or its robust version

$$\min \|\Omega x\|_1 \text{ subject to } \|Ax - y\|_2 \leq \eta \tag{Ω-BPDN}$$

under the assumption that x is $\Omega - k$-*cosparse*, i.e. $\sharp\{i\,;\,\langle \omega_i, x\rangle \neq 0\} \leq p - k$

Union of subspaces

Cosparsity comes from the same idea as sparsity: a $p - k$-cosparse x belongs to

$$\left\{x \in \mathbb{C}^N \,:\, \|\Omega x\|_0 \leq p - k\right\} = \bigcup_{\substack{S \subset [p] \\ \sharp S \leq k}} \{\omega_i \,:\, i \in S\}^{\perp}$$

We often write $s = p - k$.

Previous Results

Ω-Restricted Isometry Property (Ω-RIP)

For the analysis, Candes et al. [2011] introduced the Ω-RIP

$$(1-\delta)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta)\|x\|_2^2 \text{ for all } \Omega - k - \text{ cosparse } x$$

or equivalently

$$(1-\delta)\|\Omega^\dagger c\|_2^2 \leq \|A\Omega^\dagger c\|_2^2 \leq 1+\delta\|\Omega^\dagger c\|_2^2 \text{ for all } c \in \Sigma_s.$$

The smallest $\delta_s \in (0,1)$ fulfilling either of the inequalities is the *restricted isometry constant*.

Previous Results

Ω-Restricted Isometry Property (Ω-RIP)

For the analysis, Candes et al. [2011] introduced the Ω-RIP

$$(1-\delta)\|x\|_2^2 \le \|Ax\|_2^2 \le (1+\delta)\|x\|_2^2 \text{ for all } \Omega - k - \text{ cosparse } x$$

or equivalently

$$(1-\delta)\|\Omega^\dagger c\|_2^2 \le \|A\Omega^\dagger c\|_2^2 \le 1+\delta\|\Omega^\dagger c\|_2^2 \text{ for all } c \in \Sigma_s.$$

The smallest $\delta_s \in (0,1)$ fulfilling either of the inequalities is the *restricted isometry constant*.

Reconstruction Guarantee

If A has the Ω-RIP with constants $\delta_{2s} \le 0.08$, then the minimizer $x^\sharp$ of BPDN fulfils

$$\|x - x^\sharp\|_2 \le C\frac{\sigma_s(\Omega x)_1}{\sqrt{s}} + D\eta$$

where $\sigma_s(\Omega x)_1 = \inf_{z:\|\Omega z\|_0 \le p-k} \|\Omega z - \Omega x\|_1$.

Previous Results

Ω-RIP for Gaussian Random Matrices

- Candes et al. [2011] showed that if Ω is *Parseval* and $m \gtrsim s \ln(p/s)$, then a Gaussian random matrix possesses the Ω-RIP with high probability.

Previous Results

Ω-RIP for Gaussian Random Matrices

- Candes et al. [2011] showed that if Ω is *Parseval* and $m \gtrsim s \ln(p/s)$, then a Gaussian random matrix possesses the Ω-RIP with high probability.
- The same paper argued, that all matrices classically used in Compressed Sensing (e.g. Rademacher, Bernoulli, Steinhaus, etc.) obey the Ω-RIP if an additional random sign flip is applied.

Previous Results

Ω-RIP for Gaussian Random Matrices

- Candes et al. [2011] showed that if Ω is *Parseval* and $m \gtrsim s \ln(p/s)$, then a Gaussian random matrix possesses the Ω-RIP with high probability.
- The same paper argued, that all matrices classically used in Compressed Sensing (e.g. Rademacher, Bernoulli, Steinhaus, etc.) obey the Ω-RIP if an additional random sign flip is applied.
- Rauhut and Kabanava [2013] showed that if $m \gtrsim \frac{Bs}{A} \ln\left(\frac{ep}{s}\right)$, then a Gaussian random matrix possesses the Ω-RIP with high probability.

Previous Results

Ω-RIP for Gaussian Random Matrices

- Candes et al. [2011] showed that if Ω is *Parseval* and $m \gtrsim s \ln(p/s)$, then a Gaussian random matrix possesses the Ω-RIP with high probability.
- The same paper argued, that all matrices classically used in Compressed Sensing (e.g. Rademacher, Bernoulli, Steinhaus, etc.) obey the Ω-RIP if an additional random sign flip is applied.
- Rauhut and Kabanava [2013] showed that if $m \gtrsim \frac{Bs}{A} \ln\left(\frac{ep}{s}\right)$, then a Gaussian random matrix possesses the Ω-RIP with high probability.
- What about other types of measurement matrices?

Previous Results

Ω-RIP for Gaussian Random Matrices

- Candes et al. [2011] showed that if Ω is *Parseval* and $m \gtrsim s \ln(p/s)$, then a Gaussian random matrix possesses the Ω-RIP with high probability.
- The same paper argued, that all matrices classically used in Compressed Sensing (e.g. Rademacher, Bernoulli, Steinhaus, etc.) obey the Ω-RIP if an additional random sign flip is applied.
- Rauhut and Kabanava [2013] showed that if $m \gtrsim \frac{Bs}{A} \ln\left(\frac{ep}{s}\right)$, then a Gaussian random matrix possesses the Ω-RIP with high probability.
- What about other types of measurement matrices?
  "We will see easily that Gaussian matrices and other random compressed sensing matrices satisfy the Ω-RIP" Candes et al. [2011]

Interlude: Sampling in Bounded Orthonormal Systems

Bounded Orthonormal Systems (BOS)

Let $\mathcal{D}$ a non-empty set endowed with a probability measure ν and $\Psi := \{\psi_1 \ldots, \psi_N\}$ be a system of pairwise orthonormal functions on $\mathcal{D}$ with respect to ν that is

$$\int_{\mathcal{D}} \psi_i(t)\overline{\psi_j}(t) d\nu(t) = \delta_{i,j}$$

Ψ is an *bounded orthonormal system* if there exists a constant $K \geq 0$ such that

$$\max_{i\in[N]} \sup_{t\in\mathcal{D}} |\psi_i(t)| \leq K$$

Interlude: Sampling in Bounded Orthonormal Systems

Bounded Orthonormal Systems (BOS)

Let $\mathcal{D}$ a non-empty set endowed with a probability measure ν and $\Psi := \{\psi_1 \ldots, \psi_N\}$ be a system of pairwise orthonormal functions on $\mathcal{D}$ with respect to ν that is

$$\int_{\mathcal{D}} \psi_i(t)\overline{\psi_j}(t)d\nu(t) = \delta_{i,j}$$

Ψ is an *bounded orthonormal system* if there exists a constant $K \geq 0$ such that

$$\max_{i\in[N]} \sup_{t\in\mathcal{D}} |\psi_i(t)| \leq K$$

Examples

- Trigonometric polynomials $x \mapsto e^{-2\pi i\langle x,\xi\rangle}$ on $\mathcal{D} = [0,1]^d$ are a BOS with $K = 1$ with respect to the Lebesgue-measure.
- Fourier matrices (or any other type of orthonormal matrices) $\mathcal{F}$ with $\mathcal{F}_{j,k} = \frac{1}{\sqrt{N}} e^{-2\pi i(j-1)(k-1)/N}$ renormalized by a factor $\sqrt{N}$ over $\mathbb{C}^N$ (here: $\mathcal{D} = [N]$) with $\nu(B) = \frac{\sharp B}{N}$.

Previous Results

Ω-RIP for Bounded Orthonormal Systems

Krahmer et al. [2015] showed the following theorems for Parseval Frames Ω:

- If Ω and Ψ are incoherent, that is $\max_{i,j} |\langle \omega_i, \psi_j \rangle| \leq \frac{K}{\sqrt{N}}$, and if

$$m \geq CsK^2\lambda^2 \ln(\lambda^2 s) \ln(p)$$

where $\lambda = \sup_{\substack{\|z\|_2=1 \\ \|z\|_0 \leq s}} \frac{\|\Omega^\dagger \Omega z\|_1}{\sqrt{s}}$ is the *localization factor*, then the rescaled sampling matrix $\sqrt{\frac{N}{m}}\Phi$, where the rows of Φ are chosen at uniformly at random from Ψ, then with probability exceeding $1 - p^{-\ln(2s)}$, Φ, exhibits uniform recovery via BPDN for $s = p - k$-cosparse vectors.

RWTH AACHEN UNIVERSITY

Previous Results

Ω-RIP for Bounded Orthonormal Systems

Krahmer et al. [2015] showed the following theorems for Parseval Frames Ω:

- If Ω and Ψ are incoherent, that is $\max_{i,j} |\langle \omega_i, \psi_j \rangle| \leq \frac{K}{\sqrt{N}}$, and if

$$m \geq CsK^2\lambda^2 \ln(\lambda^2 s) \ln(p)$$

  where $\lambda = \sup_{\substack{\|z\|_2=1 \\ \|z\|_0 \leq s}} \frac{\|\Omega^\dagger \Omega z\|_1}{\sqrt{s}}$ is the *localization factor*, then the rescaled sampling matrix $\sqrt{\frac{N}{m}}\Phi$, where the rows of Φ are chosen at uniformly at random from Ψ, then with probability exceeding $1 - p^{-\ln(2s)}$, Φ, exhibits uniform recovery via BPDN for $s = p - k$-cosparse vectors.

- If $\max_i |\langle \omega_i, \psi_j \rangle| \leq \kappa_j$ and we construct Φ by choosing rows at random from Ψ according to the probability measure given by $\left(\frac{\kappa_j^2}{\|\kappa\|_2^2}\right)_{j\in[N]}$, then the matrix $\frac{1}{\sqrt{m}}\mathrm{diag}\left(\frac{\|\kappa\|_2}{\kappa_j}\right)\Phi$ exhibits uniform recovery for $s = p - k$-cosparse vectors via BPDN with probability exceeding $1 - p^{-\ln(2s)}$.

Previous Results

Ω-RIP for Bounded Orthonormal Systems

Krahmer et al. [2015] showed the following theorems for Parseval Frames Ω:

- If Ω and Ψ are incoherent, that is $\max_{i,j} |\langle \omega_i, \psi_j \rangle| \leq \frac{K}{\sqrt{N}}$, and if

$$m \geq CsK^2\lambda^2 \ln(\lambda^2 s) \ln(p)$$

  where $\lambda = \sup_{\substack{\|z\|_2=1 \\ \|z\|_0 \leq s}} \frac{\|\Omega^\dagger \Omega z\|_1}{\sqrt{s}}$ is the *localization factor*, then the rescaled sampling matrix $\sqrt{\frac{N}{m}}\Phi$, where the rows of Φ are chosen at uniformly at random from Ψ, then with probability exceeding $1 - p^{-\ln(2s)}$, Φ, exhibits uniform recovery via BPDN for $s = p - k$-cosparse vectors.
- If $\max_i |\langle \omega_i, \psi_j \rangle| \leq \kappa_j$ and we construct Φ by choosing rows at random from Ψ according to the probability measure given by $\left(\frac{\kappa_j^2}{\|\kappa\|_2^2} \right)_{j \in [N]}$, then the matrix $\frac{1}{\sqrt{m}} \mathrm{diag}\left(\frac{\|\kappa\|_2}{\kappa_j} \right) \Phi$ exhibits uniform recovery for $s = p - k$-cosparse vectors via BPDN with probability exceeding $1 - p^{-\ln(2s)}$.
- These theorems employ the Ω-RIP.

The Idea: Null Space Properties

The Null Space Property

- Φ is said to possess the *null space property of order k with respect to* Ω if for all $S \subset I$ with $\sharp S < p - k$

$$\|\Omega_S x\|_1 < \|\Omega_{\overline{S}} x\|_1 \text{ for all } x \in \ker(\Phi) \setminus \{0\} \qquad (\Omega\text{-NSP})$$

The Idea: Null Space Properties

The Null Space Property

- Φ is said to possess the *null space property of order k with respect to* Ω if for all $S \subset I$ with $\sharp S < p - k$

$$\|\Omega_S x\|_1 < \|\Omega_{\overline{S}} x\|_1 \text{ for all } x \in \ker(\Phi) \setminus \{0\} \tag{Ω-NSP}$$

- Φ is said to possess the ℓ^2-*robust null space property of order k with respect to* Ω with constants $\theta \in (0,1)$ and $\tau \geq 0$ if for all $S \subset I$ with $\sharp S < p - k$

$$\|\Omega_S x\|_1 < \frac{\theta}{\sqrt{s}} \|\Omega_{\overline{S}} x\|_1 + \tau \|\Phi x\|_2. \tag{Ω-RNSP}$$

The Idea: Null Space Properties

The Null Space Property

- Φ is said to possess the *null space property of order k with respect to* Ω if for all $S \subset I$ with $\sharp S < p - k$

$$\|\Omega_S x\|_1 < \|\Omega_{\overline{S}} x\|_1 \text{ for all } x \in \ker(\Phi) \setminus \{0\} \tag{Ω-NSP}$$

- Φ is said to possess the ℓ^2*-robust null space property of order k with respect to* Ω with constants $\theta \in (0,1)$ and $\tau \geq 0$ if for all $S \subset I$ with $\sharp S < p - k$

$$\|\Omega_S x\|_1 < \frac{\theta}{\sqrt{s}} \|\Omega_{\overline{S}} x\|_1 + \tau \|\Phi x\|_2. \tag{Ω-RNSP}$$

- The robust ℓ^2-robust-NSP implies recovery via BPDN with an error bound for the reconstruction $x^\sharp$

$$\|x - x^\sharp\|_2 \leq \frac{C}{\sqrt{s}} \sigma_s(\Omega x)_1 + D\eta$$

  where the constants C, D only depend on the parameters θ, τ as well as the frame bounds.

Contribution

Theorem [F,Rauhut, '16]

If $\Phi \in \mathbb{C}^{m \times N}$ is a random subsampling of an orthogonal operator $\Psi \in \mathbb{C}^{N \times N}$ and

$$\|\Omega^\dagger \Psi\|_\infty \leq \frac{K}{\sqrt{N}},$$

where $\Omega^\dagger$ denotes *some* dual frame, and

$$\frac{m}{\ln^3(m)} \geq C \frac{Bs}{A\theta^2(1-\delta)^2} \ln(p)$$

then with probability exceeding $1 - C\exp\left(-c\frac{m\delta A}{K^2 sB}\right)$ the matrix Φ, if obtained from Ψ by choosing rows uniformly at random, possesses the ℓ^2-robust NSP of order s for the frame Ω with $\tau = \sqrt{\frac{N}{m\delta}}$.

Contribution

Theorem [F,Rauhut, '16]

If $\Phi \in \mathbb{C}^{m \times N}$ is a random subsampling of an orthogonal operator $\Psi \in \mathbb{C}^{N \times N}$ and

$$\|\Omega^\dagger \Psi\|_\infty \leq \frac{K}{\sqrt{N}},$$

where $\Omega^\dagger$ denotes *some* dual frame, and

$$\frac{m}{\ln^3(m)} \geq C \frac{Bs}{A\theta^2(1-\delta)^2} \ln(p)$$

then with probability exceeding $1 - C \exp\left(-c \frac{m\delta A}{K^2 sB}\right)$ the matrix Φ, if obtained from Ψ by choosing rows uniformly at random, possesses the ℓ^2-robust NSP of order s for the frame Ω with $\tau = \sqrt{\frac{N}{m\delta}}$.

Remark

The quantity $\|\Omega^\dagger \Psi\|_\infty$ can be seen as a generalization of the (local) incoherence.

ℓ^1-minimization and fourier subsampling

Fourier matrices

We consider

$$\min \|x\|_1 \text{ subject to } \Phi x = y$$

where Φ is a subsampling of the Fourier matrix $\mathcal{F} = \left(\frac{1}{\sqrt{N}} e^{-2\pi i (j-1)(k-1)/N}\right)_{1 \le j,k \le N}$ and $\Omega = \mathrm{Id}_N$. Then we have

$$\|\Omega^\dagger \Phi\|_\infty = \frac{1}{\sqrt{N}}$$

hence $K = 1$ in the theorem, which is optimal.

ℓ^1-minimization and fourier subsampling

Fourier matrices

We consider

$$\min \|x\|_1 \text{ subject to } \Phi x = y$$

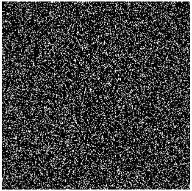
where Φ is a subsampling of the Fourier matrix $\mathcal{F} = \left(\frac{1}{\sqrt{N}} e^{-2\pi i (j-1)(k-1)/N}\right)_{1 \le j,k \le N}$ and $\Omega = \mathrm{Id}_N$. Then we have

$$\|\Omega^\dagger \Phi\|_\infty = \frac{1}{\sqrt{N}}$$

hence $K = 1$ in the theorem, which is optimal.

Features

Sampling can be done uniformly at random.



Uniform sampling pattern in fourier domain

ℓ^1-minimization and fourier subsampling

Fourier matrices

We consider

$$\min \|x\|_1 \text{ subject to } \Phi x = y$$

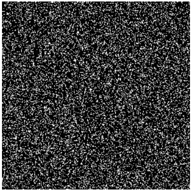
where Φ is a subsampling of the Fourier matrix $\mathcal{F} = \left(\frac{1}{\sqrt{N}} e^{-2\pi i(j-1)(k-1)/N}\right)_{1\leq j,k\leq N}$ and $\Omega = \mathrm{Id}_N$. Then we have

$$\|\Omega^\dagger \Phi\|_\infty = \frac{1}{\sqrt{N}}$$

hence $K = 1$ in the theorem, which is optimal.

Features

- Sampling can be done uniformly at random.
- Example lacks application in imaging problems since the image itself must be sparse.



Uniform sampling pattern in fourier domain

Wavelet minimization

Wavelet Transformation

Let $\Omega = \mathcal{W}$ be the orthonormal Wavelet transform and Φ a subsampled Fourier transform and consider

$$\min \|\mathcal{W}x\|_1 \text{ subject to } \Phi x = y$$

or equivalently $\min \|c\|_1$ subject to $\Phi \mathcal{W}^* c = y$. Then $(\Phi \mathcal{W}^*)_{i,(jk)} = \widehat{\psi}_{j,k}(x_i)$ where j is the scale for the wavelet transform and the $(x_i)_{1 \leq i \leq m}$ are the sampling points.

Wavelet minimization

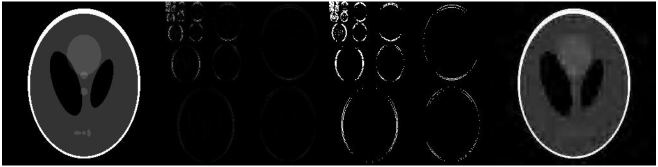
Wavelet Transformation

Let $\Omega = \mathcal{W}$ be the orthonormal Wavelet transform and Φ a subsampled Fourier transform and consider

$$\min \|\mathcal{W}x\|_1 \text{ subject to } \Phi x = y$$

or equivalently $\min \|c\|_1$ subject to $\Phi \mathcal{W}^* c = y$. Then $(\Phi \mathcal{W}^*)_{i,(jk)} = \widehat{\psi}_{j,k}(x_i)$ where j is the scale for the wavelet transform and the $(x_i)_{1 \leq i \leq m}$ are the sampling points.

Pros Natural images are often sparse/compressible in wavelets.



Reconstruction form the largest 6% of wavelet coefficients

Wavelet minimization

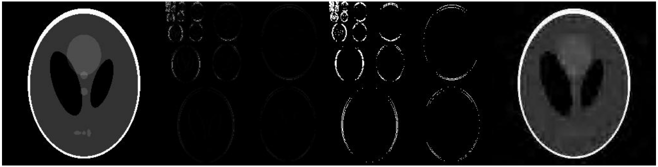
Wavelet Transformation

Let $\Omega = \mathcal{W}$ be the orthonormal Wavelet transform and Φ a subsampled Fourier transform and consider

$$\min \|\mathcal{W}x\|_1 \text{ subject to } \Phi x = y$$

or equivalently $\min \|c\|_1$ subject to $\Phi \mathcal{W}^* c = y$. Then $(\Phi\mathcal{W}^*)_{i,(jk)} = \widehat{\psi}_{j,k}(x_i)$ where j is the scale for the wavelet transform and the $(x_i)_{1\leq i\leq m}$ are the sampling points.

Pros Natural images are often sparse/compressible in wavelets.



Reconstruction form the largest 6% of wavelet coefficients

Cons The sampling points $(x_i)_{1\leq i\leq m}$ need to be drawn uniformly from $\mathbb{R}^d$ but also according to Lebesgue-meassure.

Cons We have $K = 2^J$, where J is the maximal scale employed.

Wavelet-minimization and Preconditioning

Preconditioning

Instead, consider measurements $\left(\phi(x_i)\widehat{\psi}_{j,k}(x_i)\right)=_{1\leq i\leq m}$ where ϕ is chosen such that

- $\phi(x) = C\left(1+|x|\right)^{1/2+\kappa}$ for some $\kappa > 0$ (intuitively: $\kappa \in (0,1)$)
- $\int_{\mathbb{R}^2} \frac{dx}{\phi^2(x)} = 1.$

Then, the preconditioned system $\left\{\phi\widehat{\psi}_{j,k}\ :\ j,k \text{ as chosen before}\right\}$ is a BOS with respect to the *orthogonalization measure* $\frac{dx}{\phi^2(x)}$.

Wavelet-minimization and Preconditioning

Preconditioning

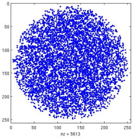
Instead, consider measurements $\left(\phi(x_i)\widehat{\psi}_{j,k}(x_i)\right) =_{1\leq i\leq m}$ where ϕ is chosen such that

- $\phi(x) = C\,(1+|x|)^{1/2+\kappa}$ for some $\kappa > 0$ (intuitively: $\kappa \in (0,1)$)
- $\int_{\mathbb{R}^2} \frac{dx}{\phi^2(x)} = 1.$

Then, the preconditioned system $\left\{\phi\widehat{\psi}_{j,k}\ :\ j,k \text{ as chosen before}\right\}$ is a BOS with respect to the *orthogonalization measure* $\frac{dx}{\phi^2(x)}$.

Effect

- Sampling can to be carried out according to $\frac{dx}{\phi^2(x)}$.



Sampling pattern in $[-1,1]^2$

Wavelet-minimization and Preconditioning

Preconditioning

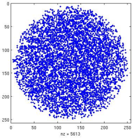
Instead, consider measurements $\left(\phi(x_i)\widehat{\psi}_{j,k}(x_i)\right) =_{1\leq i\leq m}$ where ϕ is chosen such that

- $\phi(x) = C\,(1+|x|)^{1/2+\kappa}$ for some $\kappa > 0$ (intuitively: $\kappa \in (0,1)$)
- $\int_{\mathbb{R}^2} \frac{dx}{\phi^2(x)} = 1.$

Then, the preconditioned system $\left\{\phi\widehat{\psi}_{j,k}\ :\ j,k \text{ as chosen before}\right\}$ is a BOS with respect to the *orthogonalization measure* $\frac{dx}{\phi^2(x)}$.

Effect

- Sampling can to be carried out according to $\frac{dx}{\phi^2(x)}$.
- We can apply the usual reconstruction theorems and now have $K \lesssim \frac{2^{\kappa J}}{\sqrt{\kappa}}$.



Sampling pattern in $[-1,\,1]^2$

Wavelet-minimization and Preconditioning

Preconditioning

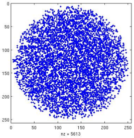
Instead, consider measurements $\left(\phi(x_i)\widehat{\psi}_{j,k}(x_i)\right) =_{1\leq i\leq m}$ where ϕ is chosen such that

- $\phi(x) = C\,(1+|x|)^{1/2+\kappa}$ for some $\kappa > 0$ (intuitively: $\kappa \in (0,1)$)
- $\int_{\mathbb{R}^2} \frac{dx}{\phi^2(x)} = 1$.

Then, the preconditioned system $\left\{\phi\widehat{\psi}_{j,k} \,:\, j,k \text{ as chosen before}\right\}$ is a BOS with respect to the *orthogonalization measure* $\frac{dx}{\phi^2(x)}$.

Effect

- Sampling can to be carried out according to $\frac{dx}{\phi^2(x)}$.
- We can apply the usual reconstruction theorems and now have $K \lesssim \frac{2^{\kappa J}}{\sqrt{\kappa}}$.
- We only need the father/mother wavelet $\psi \in C^2(\mathbb{R}^2)$.



Sampling pattern in $[-1,1]^2$

Examples of wavelet minimization

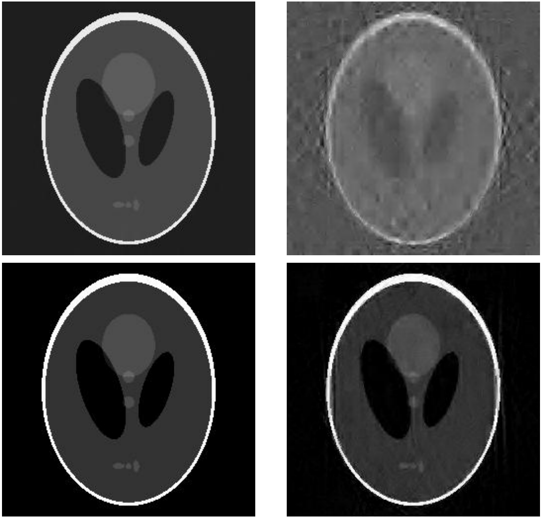
7.6%

Sampling rate $\frac{m}{n^2}$ TV Wavelet Minimization

Examples of wavelet minimization

7.6%

28.5%



Sampling rate $\frac{m}{n^2}$ TV Wavelet Minimization

Examples of wavelet minimization

7.6%  

Sampling rate $\frac{m}{n^2}$ TV Wavelet Minimization

Examples of wavelet minimization

7.6%

28.5%

Sampling rate $\frac{m}{n^2}$ TV Wavelet Minimization

RWTH AACHEN UNIVERSITY

Shearlet Frames and Anisotropy

Shearlets

- Main feature: Detects anisotropic features.

Shearlet Frames and Anisotropy

Shearlets

- Main feature: Detects anisotropic features.
- $\psi_{j,k,m}(x) = 2^{3j/4}\psi(S_{-k}A_{2^j}x - m)$ where

$$S_k = \begin{pmatrix} 1 & -k \\ 0 & 1 \end{pmatrix} \text{ and } A_a = \begin{pmatrix} a & 0 \\ 0 & \operatorname{sgn}(a)\sqrt{|a|} \end{pmatrix}$$

for some suitable $\psi \in L^2(\mathbb{R}^2)$ and parameters j, k, m.

Shearlet Frames and Anisotropy

Shearlets

- Main feature: Detects anisotropic features.
- $\psi_{j,k,m}(x) = 2^{3j/4}\psi(S_{-k}A_{2^j}x - m)$ where

$$S_k = \begin{pmatrix} 1 & -k \\ 0 & 1 \end{pmatrix} \text{ and } A_a = \begin{pmatrix} a & 0 \\ 0 & \operatorname{sgn}(a)\sqrt{|a|} \end{pmatrix}$$

  for some suitable $\psi \in L^2(\mathbb{R}^2)$ and parameters j, k, m.
- Obtain a frame $\mathcal{SH}$ via suitable truncation and exchange of variables. Those are not tight in general!

Shearlet Frames and Anisotropy

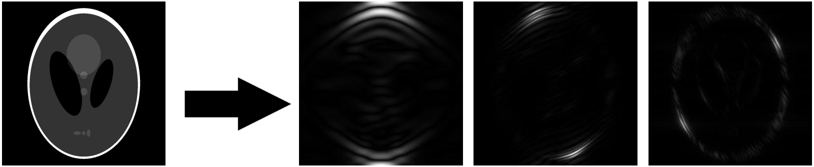
Shearlets

- Main feature: Detects anisotropic features.
- $\psi_{j,k,m}(x) = 2^{3j/4}\psi(S_{-k}A_{2^j}x - m)$ where

$$S_k = \begin{pmatrix} 1 & -k \\ 0 & 1 \end{pmatrix} \text{ and } A_a = \begin{pmatrix} a & 0 \\ 0 & \operatorname{sgn}(a)\sqrt{|a|} \end{pmatrix}$$

  for some suitable $\psi \in L^2(\mathbb{R}^2)$ and parameters j, k, m.

- Obtain a frame $\mathcal{SH}$ via suitable truncation and exchange of variables. Those are not tight in general!



Shearlet Frames and Anisotropy

Problems & Solutions

We consider min $\|\mathcal{SH}x\|_1$ subject to $\|\Phi x - y\|_2 \leq \eta$.

- Computing $\mathcal{SH}^\dagger$, in order to estimate $\|\mathcal{SH}^\dagger \Psi\|_\infty$, is extremely computational expensive in general.

Shearlet Frames and Anisotropy

Problems & Solutions

We consider $\min \|\mathcal{SH}x\|_1$ subject to $\|\Phi x - y\|_2 \leq \eta$.

- Computing $\mathcal{SH}^\dagger$, in order to estimate $\|\mathcal{SH}^\dagger \Psi\|_\infty$, is extremely computational expensive in general.
- Kutyniok and Lim [2014] constructed a shearlet frame which
  - involves functions which are compactly supported and form a frame for $L^2(\mathbb{R}^2)$,
  - which's associated dual frame can be stated in closed form and efficiently computed and
  - is moreover composed of orthonormal bases.
  - There is an implementation of this frame.
  - This frame is <u>not</u> tight in general,

Shearlet Frames and Anisotropy

Problems & Solutions

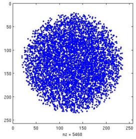
We consider min $\|\mathcal{SH}x\|_1$ subject to $\|\Phi x - y\|_2 \leq \eta$.

- Computing $\mathcal{SH}^\dagger$, in order to estimate $\|\mathcal{SH}^\dagger \Psi\|_\infty$, is extremely computational expensive in general.
- Kutyniok and Lim [2014] constructed a shearlet frame which
  - involves functions which are compactly supported and form a frame for $L^2(\mathbb{R}^2)$,
  - which's associated dual frame can be stated in closed form and efficiently computed and
  - is moreover composed of orthonormal bases.
  - There is an implementation of this frame.
  - This frame is <u>not</u> tight in general,

Preconditioning

Again, we need preconditioning in order to avoid sampling over $\mathbb{R}^2$ uniformly. Then,

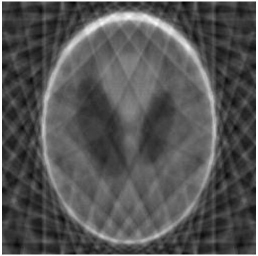
$$K = \max_{x \in \mathbb{R}^2} |\widehat{\psi}^\dagger_{j,k,m}(x)\phi(x)| \lesssim \max_j \frac{2^{j(1/4+2\kappa)}}{\sqrt{\kappa}}.$$



Sampling pattern in $[-1, 1]^2$

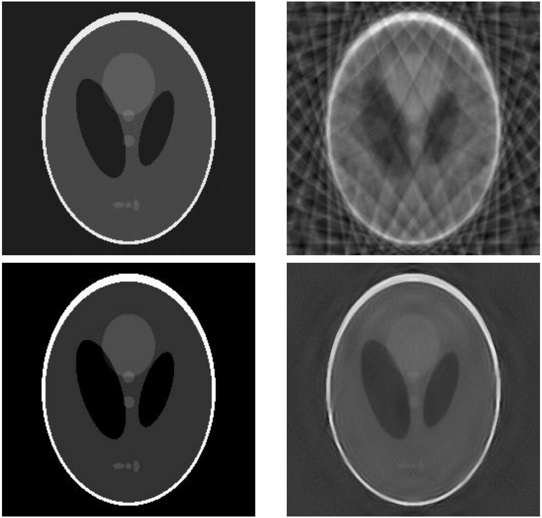
Shearlet Frames

7.6%



Sampling rate $\frac{m}{n^2}$ TV Shearlet Minimization

Shearlet Frames



7.6%

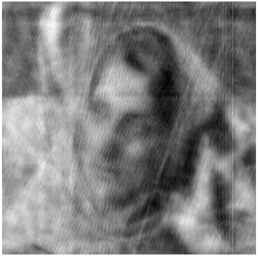
28.5%

Sampling rate $\frac{m}{n^2}$ TV Shearlet Minimization

RWTH AACHEN UNIVERSITY

Shearlet Frames

7.6%



Sampling rate $\frac{m}{n^2}$ TV Shearlet Minimization

Shearlet Frames

7.6%

28.5%

Sampling rate $\frac{m}{n^2}$ TV Shearlet Minimization

Summary & Conclusion

Advantages

- The theorem allows the usage of *arbitrary* frames.

Summary & Conclusion

Advantages

- The theorem allows the usage of *arbitrary* frames.
- Only *analysis* sparsity is needed for the theory to work.

Summary & Conclusion

Advantages

- The theorem allows the usage of *arbitrary* frames.
- Only *analysis* sparsity is needed for the theory to work.
- Subsamplings of orthogonal operators actually appear in applications, e.g. in MRI, CT etc.

Summary & Conclusion

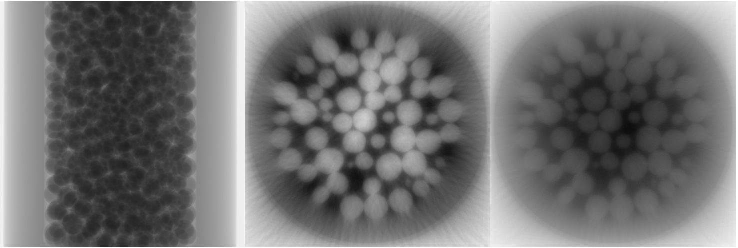
Advantages

- The theorem allows the usage of *arbitrary* frames.
- Only *analysis* sparsity is needed for the theory to work.
- Subsamplings of orthogonal operators actually appear in applications, e.g. in MRI, CT etc.
- The theory also extends to infinite-dimensional spaces, e.g. $L^2(\mathbb{R}^d)$, $H^2(\Omega)$ etc.

Summary & Conclusion

Advantages

- The theorem allows the usage of *arbitrary* frames.
- Only *analysis* sparsity is needed for the theory to work.
- Subsamplings of orthogonal operators actually appear in applications, e.g. in MRI, CT etc.
- The theory also extends to infinite-dimensional spaces, e.g. $L^2(\mathbb{R}^d)$, $H^2(\Omega)$ etc.
- Side project: Application to real-world CT-data.



Papers & Sources

E. J. Candes, Y. C. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis*, 31(1):59–73, 2011.

J. Fell and H. Rauhut. Analysis l1-minimization in imaging. *In preparation*, 2016.

S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013.

F. Krahmer, D. Needell, and R. Ward. Compressive sensing with redundant dictionaries and structured measurements. *SIAM Journal on Mathematical Analysis*, 47(6):4606–4629, 2015.

G. Kutyniok and W.-Q. Lim. Dualizable shearlet frames and sparse approximation. *arXiv preprint arXiv:1411.2303*, 2014.

H. Rauhut and M. Kabanava. Analysis l1-recovery with frames and gaussian measurements. 2013.