Quasi-linear Compressed Sensing

Massimo Fornasier

Fakultät für Mathematik
Technische Universität München
massimo.fornasier@ma.tum.de
http://www-m15.ma.tum.de/

Johann Radon Institute (RICAM)
Österreichische Akademie der Wissenschaften
massimo.fornasier@oeaw.ac.at
http://hdsparse.ricam.oeaw.ac.at/

Matheon Workshop CSA2013
Berlin, December 12, 2013
A joint work with Martin Ehler and Juliane Sigl

Outline

Outline

Introduction

Compressed sensing by linear measurements

RIP condition and recovery guarantees

Quasi-linear Compressed Sensing

Compressed sensing by nonlinear measurements

First application example: asteroseismology

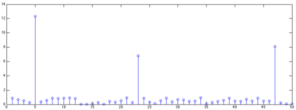
Second application example: phase retrieval

Two generalizations of the RIP and greedy algorithms

ℓ_1-minimization

Outline

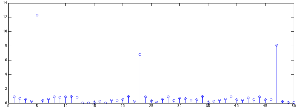
Compressed sensing



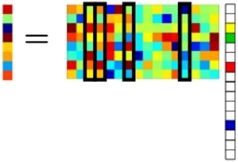
Nearly sparse signal x

Compressive sensing focuses on the robust recovery of (nearly) sparse vectors from the minimal amount of measurements obtained by a linear process.

Compressed sensing



Nearly sparse signal x



Random linear measurements

Compressive sensing focuses on the robust recovery of (nearly) sparse vectors from the minimal amount of measurements obtained by a linear process.

One typically considers model problems of the type

$$Ax = y$$

where $x \in \mathbb{R}^N$ is a (nearly) sparse vector, $A \in \mathbb{R}^{m \times N}$ is the linear measurement matrix, $m \ll N$, $y \in \mathbb{R}^m$ is the result of the measurement.

Outline

Restricted Isometry Property

Definition (Restricted Isometry Property (RIP))

A matrix $A \in \mathbb{R}^{m \times N}$ has the Restricted Isometry Property of order k if there exists $0 < \delta_k < 1$ such that

$$(1 - \delta_k) \|x\|_{\ell_2} \leq \|Ax\|_{\ell_2} \leq (1 + \delta_k) \|x\|_{\ell_2}$$

for all x with $\# \operatorname{supp}(x) \leq k$.

Restricted Isometry Property

Definition (Restricted Isometry Property (RIP))

A matrix $A \in \mathbb{R}^{m \times N}$ has the Restricted Isometry Property of order k if there exists $0 < \delta_k < 1$ such that

$$(1 - \delta_k) \|x\|_{\ell_2} \leq \|Ax\|_{\ell_2} \leq (1 + \delta_k) \|x\|_{\ell_2}$$

for all x with $\# \operatorname{supp}(x) \leq k$.

Under RIP one has guarantees of stable recovery for

- Greedy algorithms (OMP, Orthogonal Least Squares, CoSaMP, ...);
- ℓ_p-minimization (iterative hard-/soft-thresholding for $p = 0, 1$).

Outline

Nonlinear compressed sensing: state of the art

Natural Sciences, Engineering :

Many real-life measurements are nonlinear. Examples later ....

Nonlinear compressed sensing: state of the art

Natural Sciences, Engineering :
Many real-life measurements are nonlinear. Examples later ....

Need of the extension of the applicability of Compressed Sensing towards nonlinear measurements!

Nonlinear compressed sensing: state of the art

Natural Sciences, Engineering :
Many real-life measurements are nonlinear. Examples later ....

Need of the extension of the applicability of Compressed Sensing towards nonlinear measurements!

For general measurements:

- Greedy algorithms - Blumensath and Davies (2008), iterative hard thresholding - Blumensath (2012)

Nonlinear compressed sensing: state of the art

Natural Sciences, Engineering :
Many real-life measurements are nonlinear. Examples later ....

Need of the extension of the applicability of Compressed Sensing towards nonlinear measurements!

For general measurements:

- Greedy algorithms - Blumensath and Davies (2008), iterative hard thresholding - Blumensath (2012)
- Iterative hard thresholding and greedy algorithms - Beck and Eldar (2013)

Nonlinear compressed sensing: state of the art

Natural Sciences, Engineering :
Many real-life measurements are nonlinear. Examples later ....

Need of the extension of the applicability of Compressed Sensing towards nonlinear measurements!

For general measurements:

- Greedy algorithms - Blumensath and Davies (2008), iterative hard thresholding - Blumensath (2012)
- Iterative hard thresholding and greedy algorithms - Beck and Eldar (2013)

Rather straightforward generalizations of known methods for linear measurements: no recovery guarantees though!

Nonlinear compressed sensing: state of the art

Natural Sciences, Engineering :
Many real-life measurements are nonlinear. Examples later ....

Need of the extension of the applicability of Compressed Sensing towards nonlinear measurements!

For general measurements:

- Greedy algorithms - Blumensath and Davies (2008), iterative hard thresholding - Blumensath (2012)
- Iterative hard thresholding and greedy algorithms - Beck and Eldar (2013)

Rather straightforward generalizations of known methods for linear measurements: no recovery guarantees though!

- Sparse solution of polynomial optimization via lifting - Ohlsson et al. (2011-)

Nonlinear compressed sensing: state of the art

Natural Sciences, Engineering :
Many real-life measurements are nonlinear. Examples later ....

Need of the extension of the applicability of Compressed Sensing towards nonlinear measurements!

For general measurements:

- Greedy algorithms - Blumensath and Davies (2008), iterative hard thresholding - Blumensath (2012)
- Iterative hard thresholding and greedy algorithms - Beck and Eldar (2013)

Rather straightforward generalizations of known methods for linear measurements: no recovery guarantees though!

- Sparse solution of polynomial optimization via lifting - Ohlsson et al. (2011-)

Concept inherited from PhaseLift (see below) with same drawbacks: limited efficiency in high-dimension!

Nonlinear compressed sensing: state of the art

Natural Sciences, Engineering :
Many real-life measurements are nonlinear. Examples later ....

Need of the extension of the applicability of Compressed Sensing towards nonlinear measurements!

For general measurements:

- Greedy algorithms - Blumensath and Davies (2008), iterative hard thresholding - Blumensath (2012)
- Iterative hard thresholding and greedy algorithms - Beck and Eldar (2013)

Rather straightforward generalizations of known methods for linear measurements: no recovery guarantees though!

- Sparse solution of polynomial optimization via lifting - Ohlsson et al. (2011-)

Concept inherited from PhaseLift (see below) with same drawbacks: limited efficiency in high-dimension!

Quasi-linear compressed sensing

In contrast to linear measurements, the nonlinearity actually plays in a disparate manner within different recovery algorithms.

Quasi-linear compressed sensing

In contrast to linear measurements, the nonlinearity actually plays in a disparate manner within different recovery algorithms.

Hence, it does not exists a unique best algorithm for every possible nonlinearity!

Quasi-linear compressed sensing

In contrast to linear measurements, the nonlinearity actually plays in a disparate manner within different recovery algorithms.

Hence, it does not exists a unique best algorithm for every possible nonlinearity!

In this talk we provide a more unified view about algorithms, by restricting the possible nonlinearity to quasi-linear maps fulfilling generalized versions of the classical RIP:

$$A(x)x = y$$

Quasi-linear compressed sensing

In contrast to linear measurements, the nonlinearity actually plays in a disparate manner within different recovery algorithms.

Hence, it does not exists a unique best algorithm for every possible nonlinearity!

In this talk we provide a more unified view about algorithms, by restricting the possible nonlinearity to quasi-linear maps fulfilling generalized versions of the classical RIP:

$$A(x)x = y$$

Any nonlinear map $F : \mathbb{R}^N \to \mathbb{R}^m$ can be written as

$$F(x) = A(x)x,$$

where $x \to A(x) \in \mathbb{R}^{m \times N}$ is a matrix valued function depending on x.

Quasi-linear compressed sensing

In contrast to linear measurements, the nonlinearity actually plays in a disparate manner within different recovery algorithms.

Hence, it does not exists a unique best algorithm for every possible nonlinearity!

In this talk we provide a more unified view about algorithms, by restricting the possible nonlinearity to quasi-linear maps fulfilling generalized versions of the classical RIP:

$$A(x)x = y$$

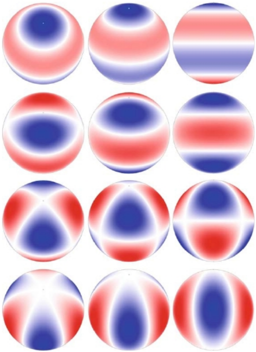
Any nonlinear map $F : \mathbb{R}^N \to \mathbb{R}^m$ can be written as

$$F(x) = A(x)x,$$

where $x \to A(x) \in \mathbb{R}^{m \times N}$ is a matrix valued function depending on x. When the dependency is smooth (e.g., at least Lipschitz continuous) then we say that F is quasi-linear.

Outline

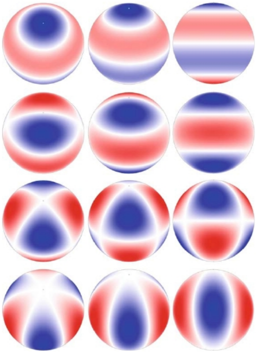
First application example: asteroseismology



Mode of pulsation of a star

- *Asteroseismology studies the oscillation of variable pulsating stars as seismic waves;*

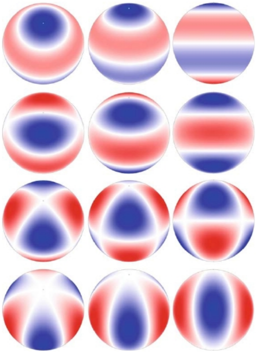
First application example: asteroseismology



Mode of pulsation of a star

- *Asteroseismology studies the oscillation of variable pulsating stars as seismic waves;*
- *Through the analysis of the characteristic pattern of these pulsations it is possible to provide information about the internal layers of the star;*

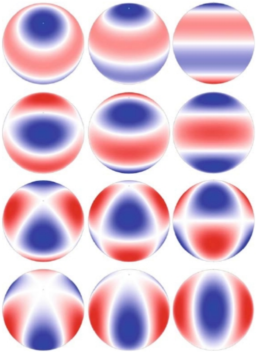
First application example: asteroseismology



Mode of pulsation of a star

- *Asteroseismology studies the oscillation of variable pulsating stars as seismic waves;*
- *Through the analysis of the characteristic pattern of these pulsations it is possible to provide information about the internal layers of the star;*
- *This helps to refine stellar structure and evolution modelling*

First application example: asteroseismology

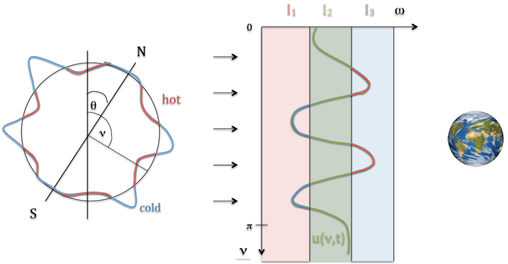


Mode of pulsation of a star

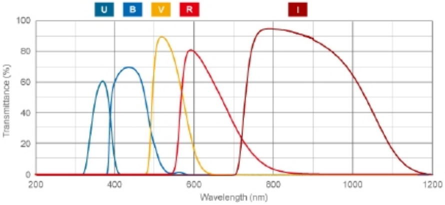
- *Asteroseismology studies the oscillation of variable pulsating stars as seismic waves;*
- *Through the analysis of the characteristic pattern of these pulsations it is possible to provide information about the internal layers of the star;*
- *This helps to refine stellar structure and evolution modelling*

Problem: Instantaneous stellar shape identification from light intensity measurements $\Rightarrow$ Finding the sparse coefficient vector for a Fourier (spherical harmonic) expansion of the star's surface

Asteroseismology measurements



Light intensity measurements at different light frequencies



Typical light filters used by telescopes

Quasi-linear modelling for asteroseismology measurements in 2D

- Description of the shape contour by a function $u(\varphi)$, depending on a parameter $-1 \leq \varphi \leq 1$ and some inclination angle θ.

Quasi-linear modelling for asteroseismology measurements in 2D

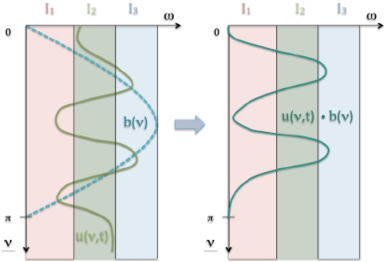
- Description of the shape contour by a function $u(\varphi)$, depending on a parameter $-1 \leq \varphi \leq 1$ and some inclination angle θ.
  Good approximation of its oscillatory behaviour with the sine expansion
  $$u(\varphi) = \sum_{i=1}^{d} x_i \sin((2\pi\varphi + \theta)i),$$
  with sparse coefficient vector $x = (x_1, \ldots, x_d)$.

Quasi-linear modelling for asteroseismology measurements in 2D

- Description of the shape contour by a function $u(\varphi)$, depending on a parameter $-1 \leq \varphi \leq 1$ and some inclination angle θ.
  Good approximation of its oscillatory behaviour with the sine expansion
  $$u(\varphi) = \sum_{i=1}^{d} x_i \sin((2\pi\varphi + \theta)i),$$
  with sparse coefficient vector $x = (x_1, \ldots, x_d)$.
- Modelling of the data acquisition process by a quasilinear relationship with measurement matrix

$$A(x)_{l,i} := \frac{\sqrt{\pi}}{2d+1} \sum_{j=-d}^{d} \omega_l \big(f_j \sum_{k=1}^{d} x_k \sin((2\pi \frac{j}{d} + \theta)k) \big) f_j \sin((2\pi \frac{j}{d} + \theta)i)$$

Breaking the symmetry by limb darkening



Asymmetry introduced by limb darkening

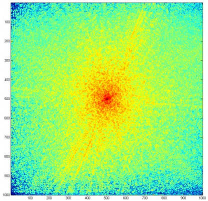
Outline

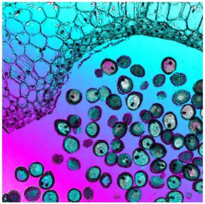
Second application example: phase retrieval

Reconstruct $x \in \mathbb{R}^N$ from measurements $y = \left(|\langle b_i, x\rangle|^2\right)_{i=1}^m$, where $\{b_i : i = 1, \dots m\} \subset \mathbb{R}^N$ is a set of measurement vectors.

Application fields:

- X-ray
- crystallography
- electron microscopy
- coherence theory
- diffraction imaging and optics
- speach enhancement





Some literature and our view

- Fienup (1978, 1982), Hayes (1982), Bates, Mnyama (1986), Liu (1990)

Some literature and our view

- Fienup (1978, 1982), Hayes (1982), Bates, Mnyama (1986), Liu (1990)
- Balan, Casazza, and Edidin (2006)

Some literature and our view

- Fienup (1978, 1982), Hayes (1982), Bates, Mnyama (1986), Liu (1990)
- Balan, Casazza, and Edidin (2006)
- Candes, Strohmer, Voroninski (2013) and Li, Voroninski (2013)
- Eldar, Mendelson (2012) and Schechtman, Beck, Eldar (2013)
- Goldfarb, Ma (2010)
- Ohlsson et al. (2011-)

Some literature and our view

- Fienup (1978, 1982), Hayes (1982), Bates, Mnyama (1986), Liu (1990)
- Balan, Casazza, and Edidin (2006)
- Candes, Strohmer, Voroninski (2013) and Li, Voroninski (2013)
- Eldar, Mendelson (2012) and Schechtman, Beck, Eldar (2013)
- Goldfarb, Ma (2010)
- Ohlsson et al. (2011-)

Recast the problem in a quasi-linear model with measurement matrix

$$A(x) = \begin{pmatrix} x^* B_1 \\ \vdots \\ x^* B_m \end{pmatrix},$$

where $B_1 = b_1 b_1^*, \cdots, B_m = b_m b_m^*$.

Outline

A class of fast decaying signals

The nonincreasing rearrangement of $x \in \mathbb{R}^d$ is defined as

$$r(x) = (|x_{j_1}|, \ldots, |x_{j_d}|)^\top, \quad \text{where} \quad |x_{j_i}| \geq |x_{j_{i+1}}|, \text{ for } i = 1, \ldots, N{-}1.$$

A class of fast decaying signals

The nonincreasing rearrangement of $x \in \mathbb{R}^d$ is defined as

$$r(x) = (|x_{j_1}|, \ldots, |x_{j_d}|)^\top, \quad \text{where} \quad |x_{j_i}| \geq |x_{j_{i+1}}|, \text{ for } i = 1, \ldots, N{-}1.$$

For $0 < \kappa < 1$, we define the class of κ-rapidly decaying vectors in $\mathbb{R}^N$ by

$$\mathcal{D}_\kappa = \{x \in \mathbb{R}^d : r_{j+1}(x) \leq \kappa r_j(x), \text{ for } j = 1, \ldots, N-1\}.$$

A class of fast decaying signals

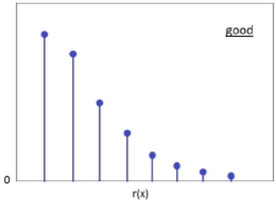
The nonincreasing rearrangement of $x \in \mathbb{R}^d$ is defined as

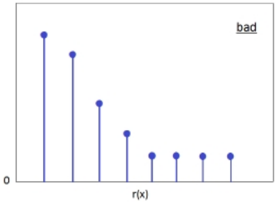
$$r(x) = (|x_{j_1}|, \ldots, |x_{j_d}|)^\top, \quad \text{where} \quad |x_{j_i}| \geq |x_{j_{i+1}}|, \text{ for } i = 1, \ldots, N-1.$$

For $0 < \kappa < 1$, we define the class of κ-rapidly decaying vectors in $\mathbb{R}^N$ by

$$\mathcal{D}_\kappa = \{x \in \mathbb{R}^d : r_{j+1}(x) \leq \kappa r_j(x), \text{ for } j = 1, \ldots, N-1\}.$$

Given $x \in \mathbb{R}^N$, the vector $x_{\{j\}} \in \mathbb{R}^N$ is the best j-sparse approximation of x, i.e., it consists of the j largest entries of x in absolute value and zeros elsewhere.





ℓ_p-greedy solver

Greedy algorithm:
Input: $A : \mathbb{R}^N \to \mathbb{R}^{m \times N}$ nonlinear, $y \in \mathbb{R}^m$
Initialize $x^{(0)} = 0 \in \mathbb{R}^M$, $\Lambda^{(0)} = \emptyset$
for $j = 1, 2, \ldots$ until some stopping criterion is met **do**
  for $l \notin \Lambda^{(j-1)}$ **do**
    $\Lambda^{(j-1,l)} := \Lambda^{(j-1)} \cup \{l\}$

$$x^{(j,l)} := \arg \min_{\{x : \mathrm{supp}(x) \subset \Lambda^{(j-1,l)}\}} \left\| A(x)x - y \right\|_{\ell_p}$$

  end
  Find index that minimizes the error:

$$l_j := \arg\min_l \left\| A(x^{(j,l)})x^{(j,l)} - y \right\|_{\ell_p}$$

  Update: $x^{(j)} := x^{(j,l_j)}$, $\Lambda^{(j)} := \Lambda^{(j-1,l_j)}$
end
Output: $x^{(1)}, x^{(2)}, \ldots$

Recovery result based on a generalized RIP I

Theorem (Ehler, F., Sigl)

Let $b = A(\hat{x})\hat{x} + e$, *where* $\hat{x} \in \mathbb{R}^N$ *is the signal to be recovered and* $e \in \mathbb{R}^m$ *is a noise term. Suppose further that* $1 \leq k \leq N$, $r_k(\hat{x}) \neq 0$, *and* $1 \leq p < \infty$. *If the following conditions hold,*

(i) *there are* $\alpha_k, \beta_k > 0$ *such that, for all* k*-sparse* $z \in \mathbb{R}^N$,

$$\alpha_k \|\hat{x}_{\{k\}} - z\| \leq \|A(\hat{x}_{\{k\}})\hat{x}_{\{k\}} - A(z)z\|_{\ell_p} \leq \beta_k \|\hat{x}_{\{k\}} - z\|,$$

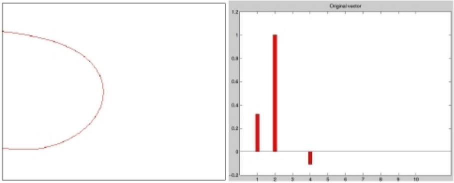
(ii) $\hat{x} \in \mathcal{D}_\kappa$ *such that* $\kappa < \frac{\tilde{\alpha}_k}{\sqrt{\tilde{\alpha}_k^2 + (\beta_k + 2L_k)^2}}$, *where* $0 < \tilde{\alpha}_k \leq \alpha_k - 2\|e\|_{\ell_p}/r_k(\hat{x})$ *and* $L_k \geq 0$ *with* $\|A(\hat{x})\hat{x} - A(\hat{x}_{\{k\}})\hat{x}_{\{k\}}\|_{\ell_p} \leq L_k \|\hat{x} - \hat{x}_{\{k\}}\|$,

then the ℓ_p*-greedy Algorithm yields a sequence* $(x^{(j)})_{j=1}^k$ *satisfying* $\operatorname{supp}(x^{(j)}) = \operatorname{supp}(\hat{x}_{\{j\}})$ *and*

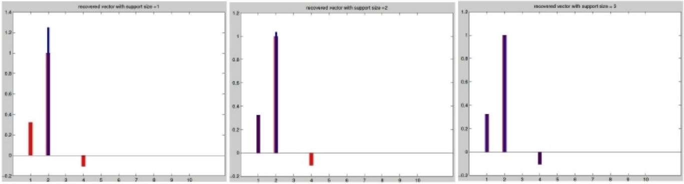
$$\|x^{(j)} - \hat{x}\| \leq \|e\|_{\ell_p}/\alpha_k + \kappa^j r_1(\hat{x})\sqrt{2}\left(1 + \frac{\beta_k + 2L_k}{\alpha_k}\right).$$

If $\hat{x}$ *is* k*-sparse, then* $\|x^{(k)} - \hat{x}\| \leq \|e\|_{\ell_p}/\alpha_k$.

Application to 2D star oscillation retrieval

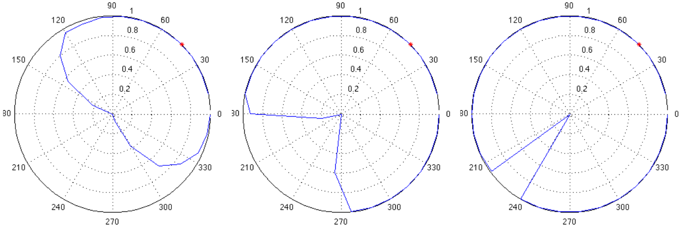


Original shape by means of u and the corresponding 3-sparse Fourier coefficients

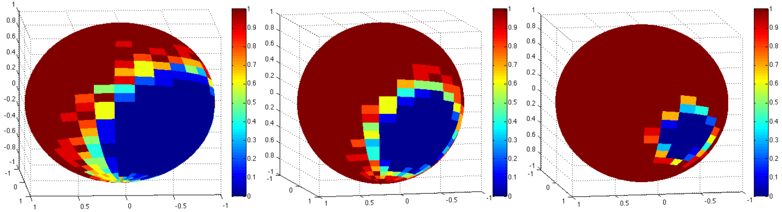


Iterative recovery by the greedy algorithm

RIP fails for phase retrieval measurements!



(a) $k = 2$



(b) $k = 3$

RIP fails for phase retrieval measurements!

The right term in the RIP is not anymore $\|x - z\|_{\ell_2}$ but

$$\|xx^* - zz^*\|_{HS} \leq \|x - z\|\|x + z\| \leq \sqrt{2}\|xx^* - zz^*\|_{HS}.$$

Recovery result based on a generalized RIP II

Theorem (Ehler, F., Sigl)

Let $y = A(\hat{x})\hat{x} + e$, *where* $\hat{x} \in \mathbb{R}^N$ *is the signal to be recovered and* $e \in \mathbb{R}^m$ *is a noise term. Suppose further that* $1 \leq k \leq N$, $r_k(\hat{x}) \neq 0$, *and* $1 \leq p < \infty$. *If the following conditions are satisfied,*

(i) *there are constants* $\alpha_k, \beta_k > 0$, *such that, for all* k*-sparse* $z \in \mathbb{R}^N$,

$$\alpha_k \|\hat{x}_{\{k\}}\hat{x}_{\{k\}}^* - zz^*\|_{HS} \leq \|A(\hat{x}_{\{k\}})\hat{x}_{\{k\}} - A(z)z\|_{\ell_p} \leq \beta_k \|\hat{x}_{\{k\}}\hat{x}_{\{k\}}^* - zz^*\|_{HS},$$

(ii) $\hat{x} \in \mathcal{D}_\kappa$ *with* $\kappa < \frac{\tilde{\alpha}_k}{\sqrt{\tilde{\alpha}_k^2 + 2(\beta_k + 2L_k)^2}}$, *where* $0 < \tilde{\alpha}_k \leq \alpha_k - 2\|e\|_{\ell_p}/r_k(\hat{x})$ *and* $L_k \geq 0$ *with* $\|A(\hat{x})\hat{x} - A(\hat{x}_{\{k\}})\hat{x}_{\{k\}}\|_{\ell_p} \leq L_k \|\hat{x}\hat{x}^* - \hat{x}_{\{k\}}\hat{x}_{\{k\}}^*\|_{HS}$,

then the ℓ_p*-greedy Algorithm yields a sequence* $(x^{(j)})_{j=1}^k$ *satisfying* $\mathrm{supp}(x^{(j)}) = \mathrm{supp}(\hat{x}_{\{j\}})$ *and*

$$\|x^{(j)}x^{(j)*} - \hat{x}\hat{x}^*\|_{HS} \leq \|e\|_{\ell_p}/\alpha_k + \kappa^j r_1(\hat{x})\sqrt{3}\left(1 + \frac{\beta_k + 2L_k}{\alpha_k}\right).$$

If $\hat{x}$ *is* k*-sparse, then* $\|x^{(k)}x^{(k)*} - \hat{x}\hat{x}^*\|_{HS} \leq \|e\|_{\ell_p}/\alpha_k$.

RIP verified probabilistically: complex random measurement vectors

Theorem (Ehler, F., Sigl)

If $\{b_i : i = 1, \ldots, m\}$ *are independent uniformly distributed vectors on the unit sphere, then there is a constant* $\alpha > 0$ *such that, for all* k*-sparse* $x, z \in \mathbb{C}^N$ *and* $m \geq c_1 k \log(eN/k)$*,*

$$\sum_{i=1}^{m} \left| |\langle b_i, x\rangle|^2 - |\langle b_i, z\rangle|^2 \right| \geq \alpha m \|xx^* - zz^*\|_{HS}$$

with probability of failure at most e^{-mc_2}.

RIP verified probabilistically: complex random measurement vectors

Theorem (Ehler, F., Sigl)

If $\{b_i : i = 1, \ldots, m\}$ are independent uniformly distributed vectors on the unit sphere, then there is a constant $\alpha > 0$ such that, for all k-sparse $x, z \in \mathbb{C}^N$ and $m \geq c_1 k \log(eN/k)$,

$$\sum_{i=1}^{m} \left| |\langle b_i, x\rangle|^2 - |\langle b_i, z\rangle|^2 \right| \geq \alpha m \|xx^* - zz^*\|_{HS}$$

with probability of failure at most e^{-mc_2}.

Extension of results of Eldar and Mendelson (2012) for the real case ...
First one uses results from Candes, Strohmer, Voroninski (2013) to show that for fixed $x, z \in \mathbb{R}^d$, there are $c_1, c > 0$ such that, for all $t > 0$,

$$\sum_{i=1}^{m} \left| |\langle b_i, x\rangle|^2 - |\langle b_i, z\rangle|^2 \right| \geq 1/\sqrt(2)(c_1 - t) m \|xx^* - zz^*\|_{HS}$$

with probability of failure at most $2e^{-mct^2} \Rightarrow$ union bound.

Discrepancy in ℓ_1-norm :-(

As a consequence we have that

$$\alpha_k \|\hat{x}_{\{k\}}\hat{x}^*_{\{k\}} - zz^*\|_{HS} \leq \|A(\hat{x}_{\{k\}})\hat{x}_{\{k\}} - A(z)z\|_{\ell_p} \leq \beta_k \|\hat{x}_{\{k\}}\hat{x}^*_{\{k\}} - zz^*\|_{HS},$$

holds for $p = 1$ with high probability!

Discrepancy in ℓ_1-norm :-(

As a consequence we have that

$$\alpha_k \|\hat{x}_{\{k\}}\hat{x}^*_{\{k\}} - zz^*\|_{HS} \leq \|A(\hat{x}_{\{k\}})\hat{x}_{\{k\}} - A(z)z\|_{\ell_p} \leq \beta_k \|\hat{x}_{\{k\}}\hat{x}^*_{\{k\}} - zz^*\|_{HS},$$

holds for $p = 1$ with high probability!

A proof for $p = 2$ is still open!

ℓ_p-greedy solver

Greedy algorithm:

Input: $A : \mathbb{R}^N \to \mathbb{R}^{m\times N}$ nonlinear, $y \in \mathbb{R}^m$

Initialize $x^{(0)} = 0 \in \mathbb{R}^M$, $\Lambda^{(0)} = \emptyset$

for $j = 1, 2, \ldots$ until some stopping criterion is met **do**

for $l \notin \Lambda^{(j-1)}$ **do**

$\Lambda^{(j-1,l)} := \Lambda^{(j-1)} \cup \{l\}$

$$x^{(j,l)} := \arg \min_{\{x : \mathrm{supp}(x) \subset \Lambda^{(j-1,l)}\}} \left\| A(x)x - y \right\|_{\ell_p}$$

end

Find index that minimizes the error:

$$l_j := \arg\min_l \left\| A(x^{(j,l)})x^{(j,l)} - y \right\|_{\ell_p}$$

Update: $x^{(j)} := x^{(j,l_j)}$, $\Lambda^{(j)} := \Lambda^{(j-1,l_j)}$

end

Output: $x^{(1)}, x^{(2)}, \ldots$

Iterative global optimizations

At each iteration j one needs to perform a global nonconvex optimization in dimension j

Iterative global optimizations

At each iteration j one needs to perform a global nonconvex optimization in dimension $j \Rightarrow$ complexity explosion?

Iterative global optimizations

At each iteration j one needs to perform a global nonconvex optimization in dimension $j \Rightarrow$ complexity explosion?

- for $p > 1$ and for $A(\cdot)$ sufficiently smooth, at the j^{th} step, one can use Newton methods locally around previously found approximations $x^{(j)}$;

Iterative global optimizations

At each iteration j one needs to perform a global nonconvex optimization in dimension $j \Rightarrow$ complexity explosion?

- for $p > 1$ and for $A(\cdot)$ sufficiently smooth, at the j^{th} step, one can use Newton methods locally around previously found approximations $x^{(j)}$;
- for $p = 1$ one needs to apply smoothing: for instance using an iteratively reweighted least squares, whose iterations are solved by Newton methods.

Phase retrieval, $p = 2$, and fast algorithms

Assume we were allowed to consider $p = 2$ in the phase retrieval problem (so far theoretically only $p = 1$, despite numerical evidences!) and that $x^{(j)} = \sum_{n=1}^{j} \alpha_{I_n} e_{I_n}$.

Phase retrieval, $p = 2$, and fast algorithms

Assume we were allowed to consider $p = 2$ in the phase retrieval problem (so far theoretically only $p = 1$, despite numerical evidences!) and that $x^{(j)} = \sum_{n=1}^{j} \alpha_{l_n} e_{l_n}$. Although not fully equivalent to

$$x^{(j+1,l)} := \arg \min_{\{x : \operatorname{supp}(x) \subset \Lambda^{(j,l)}\}} \left\| A(x)x - y \right\|_{\ell_2}$$

we consider multiple optimizations for $l \notin \Lambda^{(l)}$

$$\arg\min_{\alpha \in \mathbb{R}} \left(\sum_{i=1}^{m} \left[\left(\sum_{n=1}^{j} \alpha_{l_n} (b_i)_{l_n} + \alpha \cdot (b_i)_l \right)^2 - y_i \right]^2 \right).$$

Phase retrieval, $p = 2$, and fast algorithms

Assume we were allowed to consider $p = 2$ in the phase retrieval problem (so far theoretically only $p = 1$, despite numerical evidences!) and that $x^{(j)} = \sum_{n=1}^{j} \alpha_{l_n} e_{l_n}$. Although not fully equivalent to

$$x^{(j+1,l)} := \arg \min_{\{x : \mathrm{supp}(x) \subset \Lambda^{(j,l)}\}} \left\| A(x)x - y \right\|_{\ell_2}$$

we consider multiple optimizations for $l \notin \Lambda^{(l)}$

$$\arg\min_{\alpha \in \mathbb{R}} \left(\sum_{i=1}^{m} \left[\left(\sum_{n=1}^{j} \alpha_{l_n} (b_i)_{l_n} + \alpha \cdot (b_i)_l \right)^2 - y_i \right]^2 \right).$$

Easily solvable as fourth degree polynomial optimizations in the sole variable α to find the optimal index l_{j+1}.

Phase retrieval, $p = 2$, and fast algorithms

Assume we were allowed to consider $p = 2$ in the phase retrieval problem (so far theoretically only $p = 1$, despite numerical evidences!) and that $x^{(j)} = \sum_{n=1}^{j} \alpha_{l_n} e_{l_n}$. Although not fully equivalent to

$$x^{(j+1,l)} := \arg \min_{\{x:\mathrm{supp}(x) \subset \Lambda^{(j,l)}\}} \left\| A(x)x - y \right\|_{\ell_2}$$

we consider multiple optimizations for $l \notin \Lambda^{(l)}$

$$\arg\min_{\alpha \in \mathbb{R}} \left(\sum_{i=1}^{m} \left[\left(\sum_{n=1}^{j} \alpha_{l_n} (b_i)_{l_n} + \alpha \cdot (b_i)_l \right)^2 - y_i \right]^2 \right).$$

Easily solvable as fourth degree polynomial optimizations in the sole variable α to find the optimal index l_{j+1}. Then one applies a Newton method starting from $\hat{x}^{(j+1)} = x^{(j)} + \widehat{\alpha}_{l_{j+1}} e_{l_{j+1}}$ to steer locally the guess to the optimal $x^{(j+1)}$.

Conclusion about this greedy strategy

- It can be very competitive with respect to semi-definite programs on matrices;

Conclusion about this greedy strategy

- It can be very competitive with respect to semi-definite programs on matrices;
- It needs careful implementation of the global optimizations at each step: Newton recommended;

Conclusion about this greedy strategy

- It can be very competitive with respect to semi-definite programs on matrices;
- It needs careful implementation of the global optimizations at each step: Newton recommended;
- It does not work for phase retrieval for the Fourier basis; one needs a "group greedy strategy", see Schechtman, Beck, Eldar (2013) for a very efficient algorithm (no guarantees though!)

Outline

What about the popular ℓ_1-minimization?

We consider the problem

$$\arg\min \|x\|_{\ell_1} \quad \text{subject to} \quad A(x)x = y.$$

In the noise case it is also standard to work with an additional relaxation of it and instead solve for $\hat{x}_\alpha$ given by

$$\hat{x}_\alpha := \arg\min_{x\in\mathbb{R}^N} \mathcal{J}_\alpha(x), \qquad \text{where} \qquad \mathcal{J}_\alpha(x) := \|A(x)x - y\|_{\ell_2}^2 + \alpha\|x\|_{\ell_1},$$

where $\alpha > 0$ is sometimes called the relaxation parameter.

Existence result

We define the map

$$\mathscr{S}_\alpha : \mathbb{R}^d \to \mathbb{R}^d, \qquad x \mapsto \mathscr{S}_\alpha(x) := \arg\min_{z \in \mathbb{R}^N} \|A(x)z - y\|^2 + \alpha \|z\|_{\ell_1}.$$

Existence result

We define the map

$$\mathscr{S}_\alpha : \mathbb{R}^d \to \mathbb{R}^d, \qquad x \mapsto \mathscr{S}_\alpha(x) := \arg\min_{z \in \mathbb{R}^N} \|A(x)z - y\|^2 + \alpha \|z\|_{\ell_1}.$$

Theorem (Ehler, F., Sigl)

Given $y \in \mathbb{R}^m$, fix $\alpha > 0$ and $c_1, c_2, c_3, \gamma > 0$ are such that, for all $x, z \in \mathbb{R}^N$,

(i) $\|A(x)\|_2 \leq c_1$,

(ii) *there is $z_x \in \mathbb{R}^d$ such that $\|z_x\|_{\ell_1} \leq c_2 \|y\|$ and $A(x) z_x = y$,*

(iii) $\|A(x) - A(z)\|_2 \leq c_3 \|x - z\|$,

(iv) *if z is $\frac{4}{\alpha^2}(c_1 + c_2 + c_1^2 c_2)^2 \|y\|^2$-sparse, then*

$$(1 - \gamma)\|z\|^2 \leq \|A(x)z\|^2 \leq (1 + \gamma)\|z\|^2,$$

(v) *the constants satisfy $\gamma < 1 - (1 + 2c_1 c_2) c_3 \|y\|$,*

then $\mathscr{S}_\alpha$ is a bounded contraction, so that $x_\alpha^{(j+1)} := \mathscr{S}_\alpha(x_\alpha^{(j)})$ converges towards a point x_α satisfying

$$x_\alpha = \arg\min_{z \in \mathbb{R}^N} \|A(x_\alpha)z - y\|^2 + \alpha \|z\|_{\ell_1}.$$

Iterative soft-thresholding algorithm

We introduce the soft-thresholding operator $\mathbb{S}_\alpha : \mathbb{R}^d \to \mathbb{R}^d$, $x \mapsto \mathbb{S}_\alpha(x)$ given by

$$(\mathbb{S}_\alpha(x))_i = \begin{cases} x_i - \alpha/2, & \alpha/2 \leq x_i \\ 0, & -\alpha/2 < x_i < \alpha/2 \\ x_i + \alpha/2, & x_i \leq -\alpha/2 \end{cases},$$

and the algorithm:

Quasi-linear iterative soft-thresholding:
Input: $A : \mathbb{R}^N \to \mathbb{R}^{m \times N}$, $y \in \mathbb{R}^m$
Initialize $x^{(0)}$ as an arbitrary vector
for $j = 1, 2, \ldots$ until some stopping criterion is met **do**

$$x_\alpha^{(j+1)} = \mathbb{S}_\alpha\big((I - A(x_\alpha^{(j)})^* A(x_\alpha^{(j)}))x_\alpha^{(j)} + A(x_\alpha^{(j)})^* y\big)$$

end
Output: $x_\alpha^{(1)}, x_\alpha^{(2)}, \ldots$

Convergence

Theorem (Ehler, F., Sigl)

Suppose that the assumptions of previuos Theorem are satisfied and let x_α be the k-sparse fixed point. We define $\hat{z}_\alpha := (I - A(x_\alpha)^* A(x_\alpha))x_\alpha) + A(x_\alpha)^* y$ and $K = \frac{4\|x_\alpha\|^2}{\alpha^2} + \frac{4c}{\alpha} C$, where $C = \sup_{1 \leq l < d}(\sqrt{l+1} \|\hat{z}_\alpha - (\hat{z}_\alpha)_{\{l\}}\|_{\ell_2})$ and $c > 0$ sufficiently large. Additionally assume that

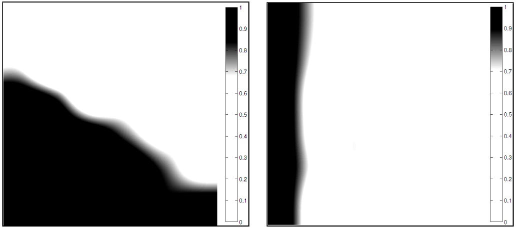
(a) *there is $0 < \tilde{\gamma} < \gamma$ such that, for all $K + k$-sparse vectors $z \in \mathbb{R}^N$,*

$$(1 - \tilde{\gamma})\|z\|^2 \leq \|A(x_\alpha)z\|^2 \leq (1 + \tilde{\gamma})\|z\|^2,$$

(b) *the constants satisfy $\tilde{\gamma} + (1 + 4c_1 c_2)c_3\|b\| < \gamma$.*

Then by using $x_\alpha^{(0)} = 0$ as initial vector, the iterative Algorithm converges towards x_α with

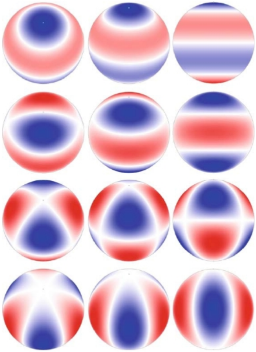
$$\|x_\alpha^{(j)} - x_\alpha\| \leq \gamma^j \|x_\alpha\|, \quad j = 0, 1, 2, \ldots.$$



(c) IST (d) IHT

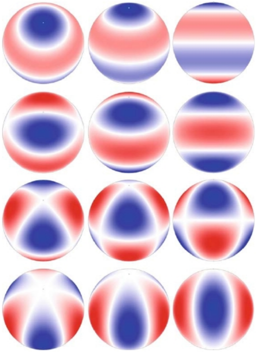
Figure: Recovery rates for iterative hard- and soft-thresholding used with the measurements $A(x) = A_1 + f(\|x - x_0\|) \times I$ with $N = 80$, $m = 20$, A_1 having i.i.d. Gaussian entries. The sparsity parameter k runs on the horizontal axis from 1 to 10, the norm of $\hat{x}$ runs on the vertical axis from 0.01 to 1. As expected, the recovery rates decrease with growing k. Consistent with the theory, we also observe decreased recovery rates for larger signal norms with soft-thresholding. Hard-thresholding appears only successful for these parameters when $k = 1$, but throughout the entire range of considered signal norms.

Conclusion



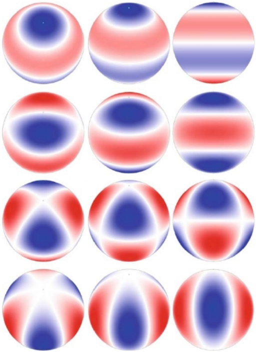
- We illustrated the classical theory of compressed sensing and some of its related algorithms

Conclusion



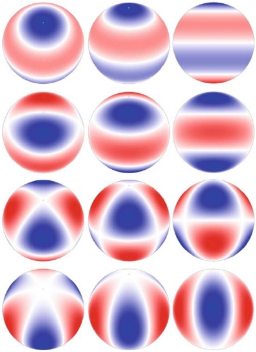
- We illustrated the classical theory of compressed sensing and some of its related algorithms
- We motivated the extension to quasi-linear measurements by two relevant real-life applications

Conclusion



- We illustrated the classical theory of compressed sensing and some of its related algorithms
- We motivated the extension to quasi-linear measurements by two relevant real-life applications
- We introduced and analyze a greedy algorithm to solve quasi-linear compressed sensing problems

Conclusion



- We illustrated the classical theory of compressed sensing and some of its related algorithms
- We motivated the extension to quasi-linear measurements by two relevant real-life applications
- We introduced and analyze a greedy algorithm to solve quasi-linear compressed sensing problems
- We introduced and analyze a iterative soft-thresholding algorithm to solve quasi-linear compressed sensing problems

A few info

- **WWW:** http://www-m15.ma.tum.de/
- **References**:
  - Martin Ehler, Massimo Fornasier, Juliane Sigl, *Quasi-linear compressed sensing*, submitted to Multiscale Modeling and Simulation, July 2013, pp. 23
  - Juliane Sigl, *Quasi-linear compressed sensing*, Master Thesis, Technical University of Munich, March 2013