The importance of phase in $\mathbb{C}$omplex compressive sensing


Laurent Jacques and Thomas Feuillen

ISPGroup, INMA-ELEN, UCLouvain, Belgium

"AI for Signal and Image Processing"
(virtual) Paris-Saclay, September 10th, 2021

The importance of ph$\measuredangle$se in $\mathbb{C}$omplex compressive sensing


No added “Deep Learning”
Totally **0%** CNN
Purely Classical
Compressive Sensing

Laurent Jacques and Thomas Feuillen

ISPGroup, INMA-ELEN, UCLouvain, Belgium

"AI for Signal and Image Processing"
(virtual) Paris-Saclay, September 10th, 2021

The importance of ph∡se in $\mathbb{C}$omplex compressive sensing

No added “Deep Learning”
Totally **0%** CNN
Purely Classical
Compressive Sensing

<ins>However</ins>, example of
non-linear sensing model
with perfect recovery

Laurent Jacques and Thomas Feuillen

ISPGroup, INMA-ELEN, UCLouvain, Belgium

"AI for Signal and Image Processing"
(virtual) Paris-Saclay, September 10th, 2021

Amplitude and Phase of Image Frequencies

A bit of history . . . Oppenheim and Lim, 1981:

"*What's the most important information between the spectral amplitude and phase of signals?*"

Amplitude and Phase of Image Frequencies

A bit of history ... Oppenheim and Lim, 1981:

"*What's the most important information between the spectral amplitude and phase of signals?*"

A simple experiment: Let $\mathcal{F}$ the (2-D) discrete Fourier transform (DFT)

Original image $f \in \mathbb{R}^{N_x \times N_y}$;
& we compute $\mathcal{F}f \in \mathbb{C}^{N_x \times N_y}$

Amplitude and Phase of Image Frequencies

A bit of history ... Oppenheim and Lim, 1981:

"*What's the most important information between the spectral amplitude and phase of signals?*"

A simple experiment: Let $\mathcal{F}$ the (2-D) discrete Fourier transform (DFT)

Original image $f \in \mathbb{R}^{N_x \times N_y}$;
& we compute $\mathcal{F}f \in \mathbb{C}^{N_x \times N_y}$

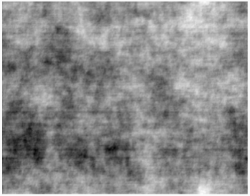


Image reconstructed with spectral amplitude

$$f' = \mathcal{F}^{-\mathbf{1}}(\underbrace{|\mathcal{F}f|}_{*})$$

*: applied component-wise

Amplitude and Phase of Image Frequencies

A bit of history ... Oppenheim and Lim, 1981:

"*What's the most important information between the spectral amplitude and phase of signals?*"

A simple experiment: Let $\mathcal{F}$ the (2-D) discrete Fourier transform (DFT)

Original image $f \in \mathbb{R}^{N_x \times N_y}$; & we compute $\mathcal{F}f \in \mathbb{C}^{N_x \times N_y}$

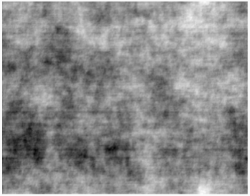


Image reconstructed with spectral amplitude

$$f' = \mathcal{F}^{-\mathbf{1}}(\underbrace{|\mathcal{F}f|}_{*})$$

*: applied component-wise

Image reconstructed with spectral phase

$$f' = \mathcal{F}^{-\mathbf{1}}(\underbrace{\frac{\mathcal{F}f}{|\mathcal{F}f|}}_{*})$$

Reconstructing an image from its spectral phases (1/2)

Fact: $\exists$ algorithm to recover *certain* images from their spectral phase (up to a global amplitude).

Fact: $\exists$ algorithm to recover *certain* images from their spectral phase (up to a global amplitude).

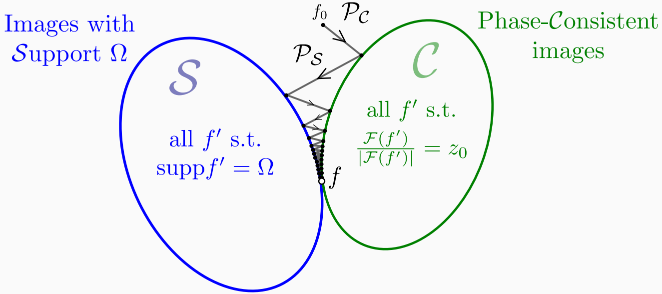
$\Rightarrow$ Use *alternate projections onto convex sets*, *i.e.*,

- given $z_0 = \mathcal{F}(f)/|\mathcal{F}(f)|$, the observed spectral phases,
- assuming $f \in \mathcal{S} :=$ set of images supported on $\Omega \subset \mathbb{R}^2$ (with $|\Omega| \leqslant N_x N_y$).

Fact: $\exists$ algorithm to recover *certain* images from their spectral phase (up to a global amplitude).

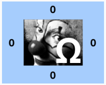
$\Rightarrow$ Use *alternate projections onto convex sets*, *i.e.*,

- given $z_0 = \mathcal{F}(f)/|\mathcal{F}(f)|$, the observed spectral phases,
- assuming $f \in \mathcal{S} :=$ set of images supported on $\Omega \subset \mathbb{R}^2$ (with $|\Omega| \leqslant N_x N_y$).



$$f^{(n+1)} = \mathcal{P}_\mathcal{B}\mathcal{P}_\mathcal{C} f^{(n)},\ f^{(0)} = f_0 \in \mathbb{C}^n, \quad \lim_{n\to+\infty} f^{(n)} = c\,f.$$

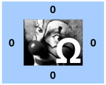
Reconstructing an image from its spectral phases (2/2)

Applying this method to our example … (*i.e.*, with Ω set from 

*)

***: This is equivalent to oversampling the Fourier domain of f.**

Reconstructing an image from its spectral phases (2/2)

Applying this method to our example . . . (*i.e.*, with Ω set from 

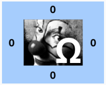
*)

1st iteration (init with ones)
(normalized SNR: 6.6 dB)

10 iterations
(normalized SNR: 11.6 dB)

***: This is equivalent to oversampling the Fourier domain of f.**

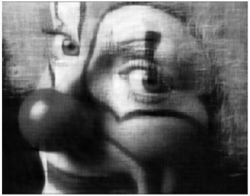
Reconstructing an image from its spectral phases (2/2)

Applying this method to our example . . . (*i.e.*, with Ω set from 

*)

1st iteration (init with ones)
(normalized SNR: 6.6 dB)

10 iterations
(normalized SNR: 11.6 dB)



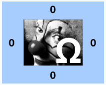
100 iterations
(normalized SNR: 19.6 dB)

1 000 iterations
(normalized SNR: 41 dB)

original

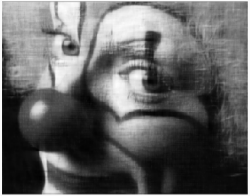
***: This is equivalent to oversampling the Fourier domain of** f.

Reconstructing an image from its spectral phases (2/2)

Applying this method to our example . . . (*i.e.*, with Ω set from  *)

1st iteration (init with ones)
(normalized SNR: 6.6 dB)

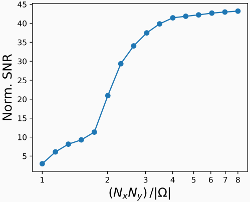
10 iterations
(normalized SNR: 11.6 dB)



100 iterations
(normalized SNR: 19.6 dB)

1 000 iterations
(normalized SNR: 41 dB)

original



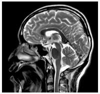
SNR vs oversampling
(1 000 iter. per point)

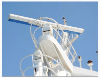
***: This is equivalent to oversampling the Fourier domain of f.**

Why could it be useful?

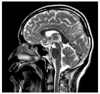
Numerous Fourier/spectral sensing applications:

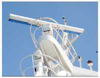
- Magnetic resonance imaging (MRI);
- Radar systems;
- Michelson interferometry / Fourier transform imaging;
- Aperture synthesis by radio interferometry.





Why could it be useful?





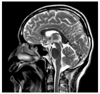
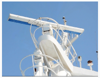
Numerous Fourier/spectral sensing applications:

- Magnetic resonance imaging (MRI);
- Radar systems;
- Michelson interferometry / Fourier transform imaging;
- Aperture synthesis by radio interferometry.

Challenges:

- Massive data stream imposes new data compression strategies.
- Compress but keep useful information (*e.g.*, for subsequent imaging).
- Large magnitude variations $\Rightarrow$ different compression impact.

Why could it be useful?


Numerous Fourier/spectral sensing applications:

- Magnetic resonance imaging (MRI);
- Radar systems;
- Michelson interferometry / Fourier transform imaging;
- Aperture synthesis by radio interferometry.

Challenges:

- Massive data stream imposes new data compression strategies.
- Compress but keep useful information (*e.g.*, for subsequent imaging).
- Large magnitude variations $\Rightarrow$ different compression impact.

Questions: Which systems are compatible with phase-only signal estimation?
$\triangleright$ (this talk) Is $\mathbb{C}$omplex compressive sensing compatible?

Why asking?

- If compatible, insensitive to large amplitudes variations (by definition).
- If robust, easy to compress information: quantize the spectral phase!

Let's collect $m < n$ measurements about $\boldsymbol{x}$ from this linear model:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon} \in \mathbb{C}^m, \tag{CS}$$

with:
- a low-complexity vector $\boldsymbol{x} \in \mathcal{L} \subset \mathbb{C}^n$ (*e.g.*, a vectorized image) with $\mathcal{L}$ the set of, such as sparse signals, low-rank matrices, ...
- a complex sensing matrix $\boldsymbol{A} \in \mathbb{C}^{m \times n}$,
- a given (additive) noise $\boldsymbol{\epsilon} \in \mathbb{C}^m$ and $\|\boldsymbol{\epsilon}\| \leqslant \varepsilon$.

Let's collect $m < n$ measurements about $\boldsymbol{x}$ from this linear model:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\epsilon} \in \mathbb{C}^m, \qquad \text{(CS)}$$

with:
- a low-complexity vector $\boldsymbol{x} \in \mathcal{L} \subset \mathbb{C}^n$ (*e.g.*, a vectorized image) with $\mathcal{L}$ the set of, such as sparse signals, low-rank matrices, ...
- a complex sensing matrix $\boldsymbol{A} \in \mathbb{C}^{m \times n}$,
- a given (additive) noise $\boldsymbol{\epsilon} \in \mathbb{C}^m$ and $\|\boldsymbol{\epsilon}\| \leqslant \varepsilon$.

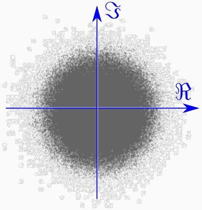
Compressive sensing:

Despite $m < n$, if m larger than $\mathcal{L}$'s "dimension", and $\boldsymbol{A}$ is "random", the vector $\boldsymbol{x}$ can be exactly recovered (or estimated if $\boldsymbol{\epsilon} \neq 0$).

[Candès and Tao, 2005; Foucart and Rauhut, 2013]

Let's be more specific ... let's focus on the Gaussian case.

Restricted isometry property



For some $0 < \delta < 1$ and $k < m < n$, if

$$m \geqslant C\,\delta^{-2} k\, \log(n/k),$$

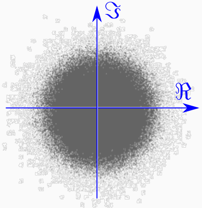
and $\sqrt{m}\,A_{ij} \sim_{\text{i.i.d.}} \mathbb{C}\mathcal{N}(0,2) \sim \mathcal{N}(0,1) + \mathrm{i}\,\mathcal{N}(0,1)$,

then, *with high probability* (w.h.p.),

$$(1-\delta)\|\boldsymbol{v}\|^2 \leqslant \|\boldsymbol{A}\boldsymbol{v}\|^2 \leqslant (1+\delta)\|\boldsymbol{v}\|^2, \quad \forall k\text{-sparse } \boldsymbol{v}. \qquad (\mathrm{RIP}(k,\delta))$$

Let's be more specific . . . let's focus on the Gaussian case.

Restricted isometry property



For some $0 < \delta < 1$ and $k < m < n$, if

$$m \geqslant C\,\delta^{-2} k\,\log(n/k),$$

and $\sqrt{m}\,A_{ij} \sim_{\text{i.i.d.}} \mathbb{C}\mathcal{N}(0,2) \sim \mathcal{N}(0,1) + \mathrm{i}\,\mathcal{N}(0,1)$,

then, *with high probability* (w.h.p.),

$$(1-\delta)\|\boldsymbol{v}\|^2 \leqslant \|\boldsymbol{A}\boldsymbol{v}\|^2 \leqslant (1+\delta)\|\boldsymbol{v}\|^2, \quad \forall k\text{-sparse } \boldsymbol{v}. \qquad (\mathrm{RIP}(k,\delta))$$

So, why does CS work? Given $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ with $\boldsymbol{x}$ a k-sparse signal, we have

- $\mathrm{RIP}(2k,\delta) \Rightarrow \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{u}\|^2 = \|\boldsymbol{A}(\boldsymbol{x}-\boldsymbol{u})\|^2 \approx \|\boldsymbol{x}-\boldsymbol{u}\|^2$, for all k-sparse $\boldsymbol{u}$.

$\Rightarrow$ $\boldsymbol{A}$ is *essentially* invertible over the set of sparse vectors;
just estimate $\boldsymbol{x}$ by finding a sparse $\boldsymbol{u}$ zeroing or minimizing $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{u}\|^2$!

The RIP supports (one of) the "fundamental theorem(s) of CS"

Theorem: If $\boldsymbol{A}$ is RIP$(2k, \delta)$ with $0 < \delta < \delta_0$ (*e.g.*, $\delta_0 = 1/\sqrt{2}$), then the *basis pursuit denoise* estimate:

$$\hat{\boldsymbol{x}} = \operatorname*{arg\,min}_{\boldsymbol{u} \in \mathbb{C}^n} \underbrace{\|\boldsymbol{u}\|_1}_{\text{sparsity promoting}} \quad \text{s.t.} \quad \underbrace{\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{u}\| \leqslant \varepsilon}_{\text{data fidelity}}, \qquad \text{(BPDN)}$$

See, *e.g.*, Candès, 2008; Foucart and Rauhut, 2013; Cai and Zhang, 2013.

The RIP supports (one of) the "fundamental theorem(s) of CS"

Theorem: If $\boldsymbol{A}$ is RIP$(2k, \delta)$ with $0 < \delta < \delta_0$ (*e.g.*, $\delta_0 = 1/\sqrt{2}$), then the *basis pursuit denoise* estimate:

$$\hat{\boldsymbol{x}} = \operatorname*{arg\,min}_{\boldsymbol{u} \in \mathbb{C}^n} \underbrace{\|\boldsymbol{u}\|_1}_{\text{sparsity promoting}} \quad \text{s.t.} \quad \underbrace{\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{u}\| \leqslant \varepsilon}_{\text{data fidelity}}, \qquad \text{(BPDN)}$$

satisfies the *instance optimality*

$$\underbrace{\|\boldsymbol{x} - \hat{\boldsymbol{x}}\|}_{\text{Rec. error; } \approx \text{ MSE}} \leqslant C \underbrace{\frac{\|\boldsymbol{x} - \boldsymbol{x}_k\|_1}{\sqrt{k}}}_{\text{deviation to sparsity}} + \underbrace{D\,\varepsilon}_{\text{noise}}.$$

See, *e.g.*, Candès, 2008; Foucart and Rauhut, 2013; Cai and Zhang, 2013.

Phase-Only Sensing Model for CS

Inspired by Oppenheim and Lim, 1981; Boufounos, 2013,

let's consider the phase-only (non-linear) compressive sensing model:

$$\boldsymbol{z} = \mathrm{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}) + \boldsymbol{\epsilon} \in \mathbb{C}^m, \qquad \text{(PO-CS)}$$

where:

- $\boldsymbol{x}$ is real and k-sparse[†];
- $\mathrm{sign}_{\mathbb{C}}(r e^{\mathrm{i}\theta}) := e^{\mathrm{i}\theta}$ (and 0 if $r = 0$), applied pointwise;
- and $\boldsymbol{\epsilon} \in \mathbb{C}^m$ a bounded noise with $\|\boldsymbol{\epsilon}\|_\infty \leqslant \tau$ for some $\tau \geqslant 0$.

Phase-Only Sensing Model for CS

Inspired by Oppenheim and Lim, 1981; Boufounos, 2013,

let's consider the phase-only (non-linear) compressive sensing model:

$$\boldsymbol{z} = \operatorname{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}) + \boldsymbol{\epsilon} \in \mathbb{C}^m, \qquad \text{(PO-CS)}$$

where:
- $\boldsymbol{x}$ is real and k-sparse[†];
- $\operatorname{sign}_{\mathbb{C}}(r e^{\mathrm{i}\theta}) := e^{\mathrm{i}\theta}$ (and 0 if $r = 0$), applied pointwise;
- and $\boldsymbol{\epsilon} \in \mathbb{C}^m$ a bounded noise with $\|\boldsymbol{\epsilon}\|_\infty \leqslant \tau$ for some $\tau \geqslant 0$.

Key observations:

1. If $\boldsymbol{x} \to C\boldsymbol{x}$ with $C > 0$, $\boldsymbol{z}$ is unchanged (Signal amplitude is lost)

Phase-Only Sensing Model for CS

Inspired by Oppenheim and Lim, 1981; Boufounos, 2013,

let's consider the phase-only (non-linear) compressive sensing model:

$$\boldsymbol{z} = \mathrm{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}) + \boldsymbol{\epsilon} \in \mathbb{C}^m, \tag{PO-CS}$$

where:
- $\boldsymbol{x}$ is real and k-sparse†;
- $\mathrm{sign}_{\mathbb{C}}(r e^{\mathrm{i}\theta}) := e^{\mathrm{i}\theta}$ (and 0 if $r = 0$), applied pointwise;
- and $\boldsymbol{\epsilon} \in \mathbb{C}^m$ a bounded noise with $\|\boldsymbol{\epsilon}\|_\infty \leqslant \tau$ for some $\tau \geqslant 0$.

Key observations:

1. If $\boldsymbol{x} \to C\boldsymbol{x}$ with $C > 0$, $\boldsymbol{z}$ is unchanged (Signal amplitude is lost)

2. If both $\boldsymbol{A}$ and $\boldsymbol{x}$ are real, then $\boldsymbol{z} \in \{\pm 1\}^m$ (Real PO-CS $\to$ 1-bit CS)

Fact: In noiseless 1-bit CS, best estimate s.t. $\|\hat{\boldsymbol{x}} - \boldsymbol{x}\| = \Omega(1/m)$ if $m \uparrow$.

[Boufounos and Baraniuk, 2008; Jacques et al., 2013; Plan and Vershynin, 2012]

(Note†: simplifying hypotheses applied here)

1. We consider the sensing of real vectors $\boldsymbol{x} \in \mathbb{R}^n$.

   Note: If complex signal $\boldsymbol{x}$, we can always rewrite

$$\boldsymbol{A}\,\boldsymbol{x} = (\boldsymbol{A}^{\Re} + \mathrm{i}\boldsymbol{A}^{\Im})(\boldsymbol{x}^{\Re} + \mathrm{i}\boldsymbol{x}^{\Im}) = \big(\boldsymbol{A},\, \mathrm{i}\boldsymbol{A}\big) \begin{pmatrix} \boldsymbol{x}^{\Re} \\ \boldsymbol{x}^{\Im} \end{pmatrix} = \overline{\boldsymbol{A}}\,\bar{\boldsymbol{x}},$$

   with $\bar{\boldsymbol{x}} \in \mathbb{R}^{2n}$ and $\overline{\boldsymbol{A}} \in \mathbb{C}^{m \times 2n}$.

(Note†: simplifying hypotheses applied here)

1. We consider the sensing of real vectors $\boldsymbol{x} \in \mathbb{R}^n$.

   Note: If complex signal $\boldsymbol{x}$, we can always rewrite

   $$\boldsymbol{A}\,\boldsymbol{x} = (\boldsymbol{A}^{\Re} + \mathrm{i}\boldsymbol{A}^{\Im})(\boldsymbol{x}^{\Re} + \mathrm{i}\boldsymbol{x}^{\Im}) = \big(\boldsymbol{A},\, \mathrm{i}\boldsymbol{A}\big)\begin{pmatrix}\boldsymbol{x}^{\Re}\\ \boldsymbol{x}^{\Im}\end{pmatrix} = \overline{\boldsymbol{A}}\,\bar{\boldsymbol{x}},$$

   with $\bar{\boldsymbol{x}} \in \mathbb{R}^{2n}$ and $\overline{\boldsymbol{A}} \in \mathbb{C}^{m\times 2n}$.

   Caveat: This can impact the signal model
   e.g., sparse in $\mathbb{C}^n \equiv$ group sparse in $\mathbb{R}^{2n}$.

(Note†: simplifying hypotheses applied here)

1. We consider the sensing of real vectors $\boldsymbol{x} \in \mathbb{R}^n$.

Note: If complex signal $\boldsymbol{x}$, we can always rewrite

$$\boldsymbol{A}\,\boldsymbol{x} = (\boldsymbol{A}^{\Re} + \mathrm{i}\boldsymbol{A}^{\Im})(\boldsymbol{x}^{\Re} + \mathrm{i}\boldsymbol{x}^{\Im}) = \big(\boldsymbol{A},\, \mathrm{i}\boldsymbol{A}\big) \begin{pmatrix} \boldsymbol{x}^{\Re} \\ \boldsymbol{x}^{\Im} \end{pmatrix} = \overline{\boldsymbol{A}}\,\bar{\boldsymbol{x}},$$

with $\bar{\boldsymbol{x}} \in \mathbb{R}^{2n}$ and $\overline{\boldsymbol{A}} \in \mathbb{C}^{m\times 2n}$.

Caveat: This can impact the signal model
e.g., sparse in $\mathbb{C}^n \equiv$ group sparse in $\mathbb{R}^{2n}$.

2. We focus here on the case of sparse vectors in $\mathbb{R}^n$.

However, extension to any low-complexity signals is possible
(with small "dimension", that is *Gaussian mean width*)

Principle: Turn the non-linear PO model into linear one.

Principle: Turn the non-linear PO model into linear one. Step by step ...

A. Let's normalize $\boldsymbol{x}$: Since signal amplitude is lost, the renormalized signal

$$\boldsymbol{x}^\star := \frac{\kappa\sqrt{m}}{\|\boldsymbol{A}\boldsymbol{x}\|_1}\,\boldsymbol{x}, \quad \text{with } \kappa := \sqrt{\tfrac{\pi}{2}}.$$

preserves PO measurements, *i.e.*, $\mathrm{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}) = \mathrm{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}^\star)$.

Therefore, we focus on the recovery of $\boldsymbol{x}^\star$ ($\rightarrow$ encodes signal direction) with

$$\|\boldsymbol{A}\boldsymbol{x}^\star\|_1 = \kappa\sqrt{m}.$$

Principle: Turn the non-linear PO model into linear one. Step by step ...

A. Let's normalize x: Since signal amplitude is lost, the renormalized signal

$$\boldsymbol{x}^\star := \frac{\kappa\sqrt{m}}{\|\boldsymbol{A}\boldsymbol{x}\|_1}\,\boldsymbol{x}, \quad \text{with } \kappa := \sqrt{\tfrac{\pi}{2}}.$$

preserves PO measurements, *i.e.*, $\mathrm{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}) = \mathrm{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}^\star)$.

Therefore, we focus on the recovery of $\boldsymbol{x}^\star$ ($\to$ encodes signal direction) with

$$\|\boldsymbol{A}\boldsymbol{x}^\star\|_1 = \kappa\sqrt{m}.$$

Rationale:

- Well, it's useful for our proofs 🙂

Principle: Turn the non-linear PO model into linear one. Step by step ...

A. Let's normalize x: Since signal amplitude is lost, the renormalized signal

$$\boldsymbol{x}^\star := \frac{\kappa\sqrt{m}}{\|\boldsymbol{A}\boldsymbol{x}\|_1}\,\boldsymbol{x}, \quad \text{with } \kappa := \sqrt{\tfrac{\pi}{2}}.$$

preserves PO measurements, *i.e.*, $\mathrm{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}) = \mathrm{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}^\star)$.

Therefore, we focus on the recovery of $\boldsymbol{x}^\star$ ($\rightarrow$ encodes signal direction) with

$$\|\boldsymbol{A}\boldsymbol{x}^\star\|_1 = \kappa\sqrt{m}.$$

Rationale:

- Well, it's useful for our proofs 🙂
- For complex Gaussian $\sqrt{m}\,\boldsymbol{A} \sim \mathbb{C}\mathcal{N}^{m\times n}(0,2)$ and $g \sim \mathcal{N}(0,1)$,

$$\mathbb{E}|g| = \kappa \quad \Rightarrow \quad \mathbb{E}\|\boldsymbol{A}\boldsymbol{x}\|_1 = \kappa\sqrt{m}\,\|\boldsymbol{x}\| \quad \Rightarrow \quad \|\boldsymbol{x}^\star\| \approx 1.$$

$\Rightarrow$ $\boldsymbol{x}^\star$ is (almost) a unit length vector, a direction

Principle: Turn the non-linear PO model into linear one. Step by step ...

B. Let's find linear constraints: From the noiseless model

$$\boldsymbol{z} = \mathrm{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}^\star),$$

we see that the vector $\boldsymbol{u} = \boldsymbol{x}^\star \in \mathbb{R}^n$ respects both:

$$\underbrace{\langle \boldsymbol{z}, \boldsymbol{A}\boldsymbol{u} \rangle}_{\substack{=\langle \mathrm{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}^\star), \boldsymbol{A}\boldsymbol{x}^\star \rangle \\ =\|\boldsymbol{A}\boldsymbol{x}^\star\|_1 \text{ if } \boldsymbol{u}=\boldsymbol{x}^\star}} = \kappa\sqrt{m} \quad \Leftrightarrow \big\langle \underbrace{\tfrac{1}{\kappa\sqrt{m}} \boldsymbol{A}^* \boldsymbol{z}}_{:=\boldsymbol{\alpha}_{\boldsymbol{z}}}, \boldsymbol{u} \big\rangle = 1 \qquad \textbf{(normalization)}$$

Principle: Turn the non-linear PO model into linear one. Step by step ...

B. Let's find linear constraints: From the noiseless model

$$z = \mathrm{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}^\star),$$

we see that the vector $\boldsymbol{u} = \boldsymbol{x}^\star \in \mathbb{R}^n$ respects both:

$$\begin{cases} \underbrace{\langle \boldsymbol{z}, \boldsymbol{A}\boldsymbol{u}\rangle}_{\substack{=\langle \mathrm{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}^\star), \boldsymbol{A}\boldsymbol{x}^\star\rangle \\ =\|\boldsymbol{A}\boldsymbol{x}^\star\|_1 \ \text{if}\ \boldsymbol{u}=\boldsymbol{x}^\star}} = \kappa\sqrt{m} \quad \Leftrightarrow \big\langle \underbrace{\tfrac{1}{\kappa\sqrt{m}}\boldsymbol{A}^*\boldsymbol{z}}_{:=\boldsymbol{\alpha}_{\boldsymbol{z}}}, \boldsymbol{u}\big\rangle = 1 & \text{(normalization)} \\ \mathrm{diag}(\boldsymbol{z})^*\,\boldsymbol{A}\boldsymbol{u} = \big(\underbrace{z_1^* \cdot (\boldsymbol{A}\boldsymbol{u})_1}_{=|(\boldsymbol{A}\boldsymbol{x}^\star)_1|\ \text{if}\ \boldsymbol{u}=\boldsymbol{x}^\star}, \cdots, \underbrace{z_m^* \cdot (\boldsymbol{A}\boldsymbol{u})_m}_{=|(\boldsymbol{A}\boldsymbol{x}^\star)_m|\ \text{if}\ \boldsymbol{u}=\boldsymbol{x}^\star}\big)^\top \in \mathbb{R}_+^m & \text{(phase consistency)} \end{cases}$$

Principle: Turn the non-linear PO model into linear one. Step by step . . .

B. Let's find linear constraints: From the noiseless model

$$\boldsymbol{z} = \operatorname{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}^\star),$$

we see that the vector $\boldsymbol{u} = \boldsymbol{x}^\star \in \mathbb{R}^n$ respects both:

$$\begin{cases} \underbrace{\langle \boldsymbol{z}, \boldsymbol{A}\boldsymbol{u}\rangle}_{\substack{=\langle \operatorname{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}^\star), \boldsymbol{A}\boldsymbol{x}^\star\rangle \\ =\|\boldsymbol{A}\boldsymbol{x}^\star\|_1 \text{ if } \boldsymbol{u}=\boldsymbol{x}^\star}} = \kappa\sqrt{m} \quad \Leftrightarrow \big\langle \underbrace{\tfrac{1}{\kappa\sqrt{m}}\boldsymbol{A}^*\boldsymbol{z}}_{:=\boldsymbol{\alpha}_{\boldsymbol{z}}}, \boldsymbol{u}\big\rangle = 1 & \text{(normalization)} \\ \operatorname{diag}(\boldsymbol{z})^*\boldsymbol{A}\boldsymbol{u} = \big(\underbrace{z_1^* \cdot (\boldsymbol{A}\boldsymbol{u})_1}_{=|(\boldsymbol{A}\boldsymbol{x}^\star)_1| \text{ if } \boldsymbol{u}=\boldsymbol{x}^\star}, \cdots, \underbrace{z_m^* \cdot (\boldsymbol{A}\boldsymbol{u})_m}_{=|(\boldsymbol{A}\boldsymbol{x}^\star)_m| \text{ if } \boldsymbol{u}=\boldsymbol{x}^\star} \big)^\top \in \mathbb{R}^m_{\cancel{+}} & \text{(phase consistency)} \end{cases}$$

Let's relax the *phase consistency*: just impose $\operatorname{diag}(\boldsymbol{z})^*\boldsymbol{A}\boldsymbol{u} \in \mathbb{R}^m$, that is

$$0 = \Im(\operatorname{diag}(\boldsymbol{z})^*\boldsymbol{A}\boldsymbol{u})$$

Principle: Turn the non-linear PO model into linear one. Step by step ...

B. Let's find linear constraints: From the noiseless model

$$\boldsymbol{z} = \mathrm{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}^\star),$$

we see that the vector $\boldsymbol{u} = \boldsymbol{x}^\star \in \mathbb{R}^n$ respects both:

$$\begin{cases} \underbrace{\langle \boldsymbol{z}, \boldsymbol{A}\boldsymbol{u}\rangle}_{\substack{=\langle \mathrm{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}^\star), \boldsymbol{A}\boldsymbol{x}^\star\rangle \\ = \|\boldsymbol{A}\boldsymbol{x}^\star\|_1 \text{ if } \boldsymbol{u}=\boldsymbol{x}^\star}} = \kappa\sqrt{m} \;\Leftrightarrow\; \langle \underbrace{\tfrac{1}{\kappa\sqrt{m}}\boldsymbol{A}^*\boldsymbol{z}}_{:=\boldsymbol{\alpha}_{\boldsymbol{z}}}, \boldsymbol{u}\rangle = 1 & \text{(normalization)} \\ \mathrm{diag}(\boldsymbol{z})^*\,\boldsymbol{A}\boldsymbol{u} = \big(\underbrace{z_1^* \cdot (\boldsymbol{A}\boldsymbol{u})_1}_{=|(\boldsymbol{A}\boldsymbol{x}^\star)_1| \text{ if } \boldsymbol{u}=\boldsymbol{x}^\star}, \cdots, \underbrace{z_m^* \cdot (\boldsymbol{A}\boldsymbol{u})_m}_{=|(\boldsymbol{A}\boldsymbol{x}^\star)_m| \text{ if } \boldsymbol{u}=\boldsymbol{x}^\star}\big)^\top \in \mathbb{R}^m_{\cancel{+}} & \text{(phase consistency)} \end{cases}$$

Let's relax the *phase consistency*: just impose $\mathrm{diag}(\boldsymbol{z})^*\,\boldsymbol{A}\boldsymbol{u} \in \mathbb{R}^m$, that is

$$0 = \Im(\mathrm{diag}(\boldsymbol{z})^*\,\boldsymbol{A}\boldsymbol{u}) = (\mathrm{diag}(z)^{\Re}\boldsymbol{A}^{\Im} - \mathrm{diag}(z)^{\Im}\boldsymbol{A}^{\Re})\boldsymbol{u} =: \boldsymbol{H}_z\boldsymbol{u}.$$

Moreover, our *normalization* means

$$\langle \boldsymbol{\alpha}_{\boldsymbol{z}}, \boldsymbol{u}\rangle = 1 \quad\Leftrightarrow\quad \langle \boldsymbol{\alpha}_{\boldsymbol{z}}^{\Re}, \boldsymbol{u}\rangle = 1,\ \langle \boldsymbol{\alpha}_{\boldsymbol{z}}^{\Im}, \boldsymbol{u}\rangle = 0.$$

In summary, $\boldsymbol{u} = \boldsymbol{x}^\star$ respects the relaxed, real $m + 2$ constraints ...

$$\boldsymbol{A}_{\boldsymbol{z}}\boldsymbol{u} = \boldsymbol{e}_1 := (1, 0, \cdots, 0)^\top \quad \Rightarrow$$

This is a linear sensing model! Like "$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$"

with

$$\boldsymbol{A}_{\boldsymbol{z}} := (\boldsymbol{\alpha}_{\boldsymbol{z}}^{\Re}, \boldsymbol{\alpha}_{\boldsymbol{z}}^{\Im}, \boldsymbol{H}_{\boldsymbol{z}}^\top)^\top \in \mathbb{R}^{(m+2)\times n}.$$

In other words,

- A good estimate $\hat{\boldsymbol{x}}$ of $\boldsymbol{x}^\star$ should respect the linear model $\boldsymbol{A}_{\boldsymbol{z}}\hat{\boldsymbol{x}} = \boldsymbol{e}_1$ since $\boldsymbol{x}^\star \in \{\boldsymbol{u} \in \mathbb{R}^n : \boldsymbol{A}_{\boldsymbol{z}}\hat{\boldsymbol{u}} = \boldsymbol{e}_1\}$.
- We know this estimate should be sparse (as $\boldsymbol{x}^\star$ is)

In summary, $\boldsymbol{u} = \boldsymbol{x}^\star$ respects the relaxed, real $m+2$ constraints ...

$$\boldsymbol{A}_z \boldsymbol{u} = \boldsymbol{e}_1 := (1, 0, \cdots, 0)^\top \quad \Rightarrow$$

This is a linear sensing model! Like "$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{y}$"

with

$$\boldsymbol{A}_z := (\boldsymbol{\alpha}_z^{\Re}, \boldsymbol{\alpha}_z^{\Im}, \boldsymbol{H}_z^\top)^\top \in \mathbb{R}^{(m+2)\times n}.$$

In other words,

- A good estimate $\hat{x}$ of $\boldsymbol{x}^\star$ should respect the linear model $\boldsymbol{A}_z \hat{\boldsymbol{x}} = \boldsymbol{e}_1$ since $\boldsymbol{x}^\star \in \{\boldsymbol{u} \in \mathbb{R}^n : \boldsymbol{A}_z \hat{\boldsymbol{u}} = \boldsymbol{e}_1\}$.
- We know this estimate should be sparse (as $\boldsymbol{x}^\star$ is)

$\Rightarrow$ As in linear CS, we can compute $\hat{x}$ from a *basis pursuit* program (BP)

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{u} \in \mathbb{C}^n} \|\boldsymbol{u}\|_1 \ \text{ s.t. } \quad \boldsymbol{A}_z \boldsymbol{u} = \boldsymbol{e}_1, \qquad (\mathrm{BP}(\boldsymbol{A}_z, \boldsymbol{e}_1))$$

Question: How far is $\hat{x}$ from $\boldsymbol{x}^\star$? Well, let's see if $\boldsymbol{A}_z$ respects the RIP!

Restricted isometry property for $\boldsymbol{A}_z$?

How could $\boldsymbol{A}_{\boldsymbol{z}} := (\boldsymbol{\alpha}_{\boldsymbol{z}}^{\Re}, \boldsymbol{\alpha}_{\boldsymbol{z}}^{\Im}, \boldsymbol{H}_{\boldsymbol{z}}^\top)^\top$ respect the RIP?

For a sparse $\boldsymbol{v}$, $\|\boldsymbol{A}_{\boldsymbol{z}}\boldsymbol{v}\|^2 := |\langle \boldsymbol{\alpha}_{\boldsymbol{z}}, \boldsymbol{v}\rangle|^2 + \|\boldsymbol{H}_{\boldsymbol{z}}\boldsymbol{v}\|^2$

you can show that, for complex Gaussian $\boldsymbol{A}$:

(i) $\langle \boldsymbol{\alpha}_{\boldsymbol{z}}, \boldsymbol{v}\rangle \approx \langle \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}, \boldsymbol{v}\rangle \approx$ projection of $\boldsymbol{v}$ onto $\mathcal{X} := \mathbb{R}\,\boldsymbol{x}$.

(ii) $\boldsymbol{H}_{\boldsymbol{z}}\boldsymbol{x} = 0$, and $\boldsymbol{H}_{\boldsymbol{z}}$ RIP on $\mathcal{X}^\perp \cap$ $2k$-sparse signals.

Restricted isometry property for A_z?

How could $\boldsymbol{A_z} := (\boldsymbol{\alpha}_{\boldsymbol{z}}^{\Re}, \boldsymbol{\alpha}_{\boldsymbol{z}}^{\Im}, \boldsymbol{H}_{\boldsymbol{z}}^{\top})^{\top}$ respect the RIP?

For a sparse $\boldsymbol{v}$, $\|\boldsymbol{A_z v}\|^2 := |\langle \boldsymbol{\alpha_z}, \boldsymbol{v}\rangle|^2 + \|\boldsymbol{H}_z \boldsymbol{v}\|^2 \approx \langle \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}, \boldsymbol{v}\rangle^2 + \|\boldsymbol{v}^{\perp}\|^2 = \|\boldsymbol{v}\|^2$

you can show that, for complex Gaussian $\boldsymbol{A}$:

(i) $\langle \boldsymbol{\alpha_z}, \boldsymbol{v}\rangle \approx \langle \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}, \boldsymbol{v}\rangle \approx$ projection of $\boldsymbol{v}$ onto $\mathcal{X} := \mathbb{R}\,\boldsymbol{x}$.

(ii) $\boldsymbol{H_z x} = 0$, and $\boldsymbol{H_z}$ RIP on $\mathcal{X}^{\perp} \cap$ $2k$-sparse signals.

Restricted isometry property for A_z?

How could $\boldsymbol{A_z} := (\boldsymbol{\alpha_z^{\Re}}, \boldsymbol{\alpha_z^{\Im}}, \boldsymbol{H_z^\top})^\top$ respect the RIP?

For a sparse $\boldsymbol{v}$, $\|\boldsymbol{A_z v}\|^2 := |\langle \boldsymbol{\alpha_z}, \boldsymbol{v}\rangle|^2 + \|\boldsymbol{H_z v}\|^2 \approx \langle \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}, \boldsymbol{v}\rangle^2 + \|\boldsymbol{v}^\perp\|^2 = \|\boldsymbol{v}\|^2$

you can show that, for complex Gaussian $\boldsymbol{A}$:

(i) $\langle \boldsymbol{\alpha_z}, \boldsymbol{v}\rangle \approx \langle \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}, \boldsymbol{v}\rangle \approx$ projection of $\boldsymbol{v}$ onto $\mathcal{X} := \mathbb{R}\,\boldsymbol{x}$.

(ii) $\boldsymbol{H_z x} = 0$, and $\boldsymbol{H_z}$ RIP on $\mathcal{X}^\perp \cap$ $2k$-sparse signals.

Theorem: Given $\boldsymbol{x}$ and $0 < \delta < 1$, $\sqrt{m}\boldsymbol{A} \sim \mathbb{C}\mathcal{N}^{m\times n}(0,2)$, if

$$m \geqslant C\delta^{-2} k \log(n/k),$$

then, w.h.p., $\boldsymbol{A_z}$ satisfies the RIP (k, δ).

Consequences:

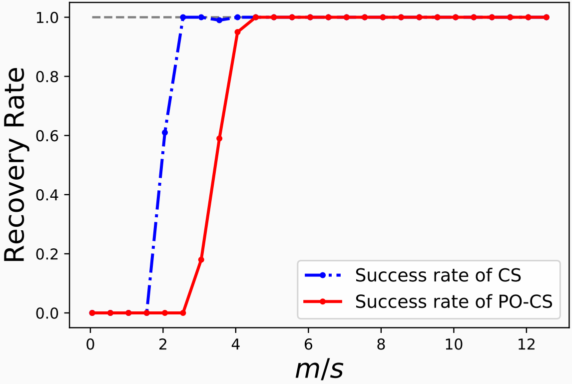
- For $\hat{\boldsymbol{x}} = \mathrm{BP}(\boldsymbol{A_z}, \boldsymbol{e}_1)$, if $\boldsymbol{A_z}$ is RIP$(\delta < \delta_0, 2k)$,
  we get exact reconstruction of signal direction, *i.e.*, $\hat{\boldsymbol{x}} = \boldsymbol{x}^\star$!
- $+$ Stability & robustness (aka *instance optimality*) with BPDN (see paper)

Let's plot a *phase-transition curve*: we generate $\sqrt{m}\boldsymbol{A} \sim \mathbb{C}\mathcal{N}^{m \times 256}(0, 2)$ &

- 20-sparse vectors in $\mathbb{R}^{256}$;
- $m \in [1, 256]$ and average over 100 trials;
- Reconstruction successful if SNR $\geqslant$ 60 dB.

Let's plot a *phase-transition curve*: we generate $\sqrt{m}\boldsymbol{A} \sim \mathbb{C}\mathcal{N}^{m\times 256}(0,2)$ &

- 20-sparse vectors in $\mathbb{R}^{256}$;
- $m \in [1, 256]$ and average over 100 trials;
- Reconstruction successful if SNR $\geqslant$ 60 dB.

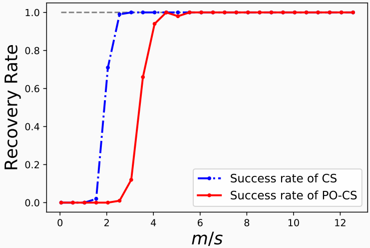


Let's be a little more daring . . . and forget Gauss

Let's be a little more daring ... and forget Gauss

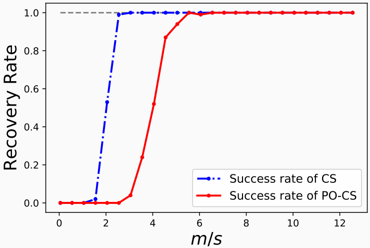
Bernoulli random matrix

$A_{ij} \sim_{\text{iid}} \{\pm 1\}$



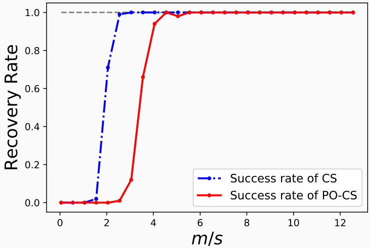
Random partial Fourier

(A = sub-sampled $\mathcal{F}(\boldsymbol{x})$)

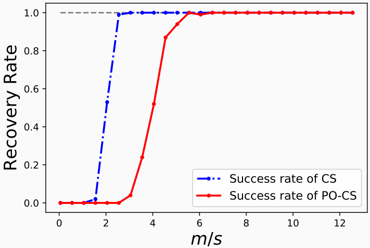


Let's be a little more daring ... and forget Gauss

Bernoulli random matrix
$A_{ij} \sim_{\text{iid}} \{\pm 1\}$



Random partial Fourier
(A = sub-sampled $\mathcal{F}(\boldsymbol{x})$)



Interestingly:

- These results are not covered by theory.
- Bernoulli random matrices do not work for 1-bit CS.
- PO-CS with Fourier $\supset$ unrecoverable counter-examples!

$$\textit{e.g.}, \text{ for } \boldsymbol{x}' := \boldsymbol{h} * \boldsymbol{x} \text{ with } \hat{h}_k > 0, \forall k, \quad \operatorname{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}') = \operatorname{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}).$$

Extra simulations: noisy case

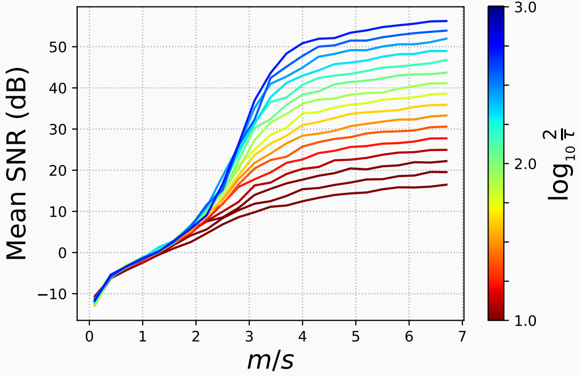
We generate $\sqrt{m}\boldsymbol{A} \sim \mathbb{C}\mathcal{N}^{m\times 256}(0,2)$ and ...

- 20-sparse vectors in $\mathbb{R}^{256}$;
- $m \in [1,256]$ and average over 100 trials;
- $\boldsymbol{z} = \mathrm{sign}_{\mathbb{C}}(\boldsymbol{A}\boldsymbol{x}) + \boldsymbol{\xi}$, with $\boldsymbol{\xi} \in \mathbb{C}^m$ and $\|\boldsymbol{\xi}\|_\infty \leqslant \tau < 2$ (Question: why 2?).

Extra simulations: noisy case

We generate $\sqrt{m}\boldsymbol{A} \sim \mathbb{C}\mathcal{N}^{m\times 256}(0,2)$ and ...

- 20-sparse vectors in $\mathbb{R}^{256}$;
- $m \in [1, 256]$ and average over 100 trials;
- $\boldsymbol{z} = \mathrm{sign}_{\mathbb{C}}(\boldsymbol{Ax}) + \boldsymbol{\xi}$, with $\boldsymbol{\xi} \in \mathbb{C}^m$ and $\|\boldsymbol{\xi}\|_\infty \leqslant \tau < 2$ (Question: why 2?).



Take-Away Messages

1. In Gauss' world, despite:
   - the non-linearity of its sensing model,
   - and the bad example of 1-bit CS (the "real" PO-CS),

   phase-only compressive sensing works "*as well as*" (linear) CS.

2. What is recovered/estimated is the signal direction (via $\boldsymbol{x}^\star$).

3. Applications: phase-quantization procedures with bounded distortion
   e.g., in radar, MRI, ...

4. Open questions:
   - (minor) Extension to complex signals.
   - (major) Theoretical extension to other random sensing matrices.

. — . — .

Thank you!

LJ, T. Feuillen, "The importance of phase in complex compressive sensing",
arXiv:2001.02529 (IEEE Tr. Inf. Th., **67**(6):4150 - 4161, June 2021)

Codes and demos on `https://tinyurl.com/phaseonly`

Boufounos, Petros T (2013). "Sparse signal reconstruction from phase-only measurements". In: *Proc. Int. Conf. Sampling Theory and Applications (SampTA)],(July 1-5 2013)*. Citeseer.

Boufounos, Petros T and Richard G Baraniuk (2008). "1-bit compressive sensing". In: *2008 42nd Annual Conference on Information Sciences and Systems*. IEEE, pp. 16–21.

Cai, T Tony and Anru Zhang (2013). "Sparse representation of a polytope and recovery of sparse signals and low-rank matrices". In: *IEEE Transactions on Information Theory* 60.1, pp. 122–132.

Candès, EJ and T Tao (2005). "Decoding by linear programming". In: *IEEE Transactions on Information Theory* 51.12, pp. 4203–4215.

Candès, Emmanuel J. (May 2008). "The restricted isometry property and its implications for compressed sensing". In: *Comptes Rendus Mathematique* 346.9-10, pp. 589–592.

Foucart, Simon and Holger Rauhut (2013). *A Mathematical Introduction to Compressive Sensing*. Springer New York.

Jacques, Laurent et al. (2013). "Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors". In: *IEEE Transactions on Information Theory* 59.4, pp. 2082–2102.

Oppenheim, A.V. and J.S. Lim (1981). "The importance of phase in signals". In: *Proceedings of the IEEE* 69.5, pp. 529–541.

Plan, Yaniv and Roman Vershynin (2012). "Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach". In: *IEEE Transactions on Information Theory* 59.1, pp. 482–494.

Part I

Extra slides

How could $\boldsymbol{A}_{\boldsymbol{z}} := (\boldsymbol{\alpha}_{\boldsymbol{z}}^{\Re}, \boldsymbol{\alpha}_{\boldsymbol{z}}^{\Im}, \boldsymbol{H}_{\boldsymbol{z}}^{\top})^{\top}$ respect the RIP?

1. Limited Projection Distortion (LPD)

Given $\bar{\boldsymbol{x}} := \boldsymbol{x}/\|\boldsymbol{x}\|$, for complex Gaussian matrix $\sqrt{m}\boldsymbol{A} \sim \mathbb{C}\mathcal{N}^{m\times n}(0,2)$, w.h.p.,

$$\langle \boldsymbol{\alpha}_{\boldsymbol{z}}, \boldsymbol{v} \rangle = \tfrac{1}{\kappa\sqrt{m}} \big\langle \mathrm{sign}_{\mathbb{C}}(\boldsymbol{A}\bar{\boldsymbol{x}}), \boldsymbol{A}\boldsymbol{v} \big\rangle \approx \langle \bar{\boldsymbol{x}}, \boldsymbol{v} \rangle, \quad \forall k\text{-sparse } \boldsymbol{v}, \tag{LPD}$$

provided $m = O(k \log n/k)$ and up to an additive error.

$\Rightarrow \langle \boldsymbol{\alpha}_{\boldsymbol{z}}, \boldsymbol{v} \rangle \approx$ projection of $\boldsymbol{v}$ onto the line $\mathcal{X} := \mathbb{R}\,\boldsymbol{x} \subset \mathbb{R}^n$.

How could $\boldsymbol{A_z} := (\boldsymbol{\alpha}_{\boldsymbol{z}}^{\Re}, \boldsymbol{\alpha}_{\boldsymbol{z}}^{\Im}, \boldsymbol{H}_{z}^{\top})^{\top}$ respect the RIP?

2. Since $\boldsymbol{H}_z := (\mathrm{diag}(z)^{\Re}\boldsymbol{A}^{\Im} - \mathrm{diag}(z)^{\Im}\boldsymbol{A}^{\Re})$, we have both:

- $\boldsymbol{H}_{\boldsymbol{z}}\boldsymbol{x} = \boldsymbol{H}_{\boldsymbol{z}}\bar{\boldsymbol{x}} = 0$
- given $\boldsymbol{x}$, $\sqrt{m}\boldsymbol{H}_{\boldsymbol{z}}$ Gaussian over $\mathcal{X}^{\perp} = \{\boldsymbol{v} : \langle\bar{\boldsymbol{x}}, \boldsymbol{v}\rangle = 0\}$ with unit variance.

$\Rightarrow$ $\boldsymbol{H}_{\boldsymbol{z}}$ can be RIP on $\mathcal{X}^{\perp} \cap$ 2k-sparse signals.

How could $\boldsymbol{A_z} := (\boldsymbol{\alpha}_{\boldsymbol{z}}^{\Re}, \boldsymbol{\alpha}_{\boldsymbol{z}}^{\Im}, \boldsymbol{H}_z^\top)^\top$ respect the RIP?

2. Since $\boldsymbol{H}_z := (\mathrm{diag}(z)^{\Re}\boldsymbol{A}^{\Im} - \mathrm{diag}(z)^{\Im}\boldsymbol{A}^{\Re})$, we have both:

- $\boldsymbol{H}_z\boldsymbol{x} = \boldsymbol{H}_z\bar{\boldsymbol{x}} = 0$
- given $\boldsymbol{x}$, $\sqrt{m}\boldsymbol{H}_z$ Gaussian over $\mathcal{X}^\perp = \{\boldsymbol{v} : \langle\bar{\boldsymbol{x}}, \boldsymbol{v}\rangle = 0\}$ with unit variance.

$\Rightarrow$ $\boldsymbol{H}_z$ can be RIP on $\mathcal{X}^\perp \cap$ $2k$-sparse signals.

Therefore, provided m is big enough, w.h.p.

$$\begin{aligned}
\|\boldsymbol{A_z}\boldsymbol{v}\|^2 := |\langle\boldsymbol{\alpha_z}, \boldsymbol{v}\rangle|^2 + \|\boldsymbol{H}_z\boldsymbol{v}\|^2 &\approx |\langle\bar{\boldsymbol{x}}, \boldsymbol{v}\rangle|^2 + \|\boldsymbol{H}_z\boldsymbol{v}^\perp\|^2 \\
&\approx |\langle\bar{\boldsymbol{x}}, \boldsymbol{v}\rangle|^2 + \|\boldsymbol{v}^\perp\|^2 \quad \textit{RIP of } \boldsymbol{H_z} \\
&= \|\boldsymbol{v}\|^2.
\end{aligned}$$