Keep the phase!
Signal recovery in phase-only compressive sensing

Laurent Jacques and Thomas Feuillen
ISPGroup, UCLouvain, Belgium
Feb. 4th, 2021, CASI Meetings (Xlim, UMR CNRS Limoges)
Oppenheim and Lim, 1981:

"What’s the most important information between the spectral amplitude and phase of signals?"
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) Fourier transform

Original image $f \in \mathbb{R}^{N_x \times N_y}$

Image reconstructed with spectral amplitude
$$f' = \mathcal{F}^{-1}(|\mathcal{F}f|)$$

Image reconstructed with spectral phase
$$f' = \mathcal{F}^{-1}\left(\frac{\mathcal{F}f}{|\mathcal{F}f|}\right)$$
Fact: ∃ algorithm to recover band-limited images from their spectral phase (up to a global amplitude).

⇒ Use alternate projections onto convex sets, i.e.,

- given $z_0 = \mathcal{F}(f)/|\mathcal{F}(f)|$, the observes spectral phase,
- assuming $f \in \mathcal{B} := $ set of band-limited images (for some cutoff freq.).

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

**Challenges:**

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

**Questions:**

- Which systems are compatible with phase-only signal estimation?
- *(this talk)* Is complex compressive sensing compatible?

Why asking?

- If compatible/robust, quantize the spectral phase for compression!
- Robust to large observation amplitudes; easy quantizers (*e.g.*, over $[0, 2\pi]$).
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 ⇒ different compression impact.

**Questions:** Which systems are compatible with phase-only signal estimation?

▷ (this talk) Is complex compressive sensing compatible?

Why asking?

- If compatible/robust, quantize the spectral phase for compression!
- Robust to large observation amplitudes; easy quantizers (over $[0, 2\pi]$).
Let’s collect $m < n$ measurements about $x$ from this **linear** model:

\[
y = Ax + n \in \mathbb{C}^m,
\]

with:

- a **low-complexity vector** $x \in \mathcal{L} \subset \mathbb{C}^n$ (with $\mathcal{L}$ the set of, e.g., sparse signals, low-rank matrices, ...),
- a **complex** sensing matrix $A \in \mathbb{C}^{m \times n}$,
- a given (additive) noise $n \in \mathbb{C}^m$ and $\|n\|_2 \leq \varepsilon$. 

**Compressive sensing:**

If $m$ larger than $\mathcal{L}$’s “dimension”, and $A$ is “random”, the vector $x$ can be exactly recovered, or estimated (if noise). 

[Candes and Tao, 2005; Foucart and Rauhut, 2013]
Let’s collect $m < n$ measurements about $x$ from this linear model:

$$y = Ax + n \in \mathbb{C}^m,$$

(CS)

with:

- a low-complexity vector $x \in \mathcal{L} \subset \mathbb{C}^n$
  (with $\mathcal{L}$ the set of, e.g., sparse signals, low-rank matrices, ...),
- a complex sensing matrix $A \in \mathbb{C}^{m \times n}$,
- a given (additive) noise $n \in \mathbb{C}^m$ and $\|n\|_2 \leq \varepsilon$.

Compressive sensing:

If $m$ larger than $\mathcal{L}$’s "dimension", and $A$ is "random", the vector $x$ can be exactly recovered, or estimated (if noise).

[Candes 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

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

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

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

$$(1 - \delta^2)\|v\|^2 \leq \|Av\|_2^2 \leq (1 + \delta^2)\|v\|^2, \quad \forall k\text{-sparse } v. \quad \text{(RIP}(k, \delta))$$
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

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

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

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

$$(1 - \delta^2)\|v\|^2 \leq \|Av\|^2 \leq (1 + \delta^2)\|v\|^2, \quad \forall k\text{-sparse } v. \quad \text{(RIP}(k, \delta))$$

So, why does CS work?

- RIP$(2k, \delta) \Rightarrow \|y - Au\|_2^2 = \|A(x - u)\|_2^2 \approx \|x - u\|_2^2$, for all $k$-sparse $x, u$.
- $A$ is *essentially* invertible over the set of sparse vectors.
The RIP supports (one of) the "fundamental theorem(s) of CS"

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

\[
\hat{x} = \arg \min_{u \in \mathbb{C}^n} \| u \|_1 \quad \text{s.t.} \quad \| y - Au \|_2 \leq \varepsilon,
\]

(BPDN)

satisfies the instance optimality

\[
\| x - \hat{x} \|_2 \leq C \left( \frac{\| x - x_k \|_1}{\sqrt{k}} + D \varepsilon \right).
\]

See, e.g., Candès, 2008; Foucart and Rauhut, 2013.
Inspired by Oppenheim and Lim, 1981; Boufounos, 2013, in the context of CS, let’s consider the phase-only (non-linear) sensing model:

\[ \mathbf{z} = \text{sign}_\mathbb{C}(\mathbf{A}\mathbf{x}) + \mathbf{\epsilon} \in \mathbb{C}^m, \quad \text{(PO-CS)} \]

with:

- \( \mathbf{x} \) is real and \( k \)-sparse;
- \( \text{sign}_\mathbb{C}(re^{i\theta}) := e^{i\theta} \) (and 0 if \( r = 0 \)), applied pointwise;
- and \( \mathbf{\epsilon} \in \mathbb{C}^m \) a bounded noise with \( ||\mathbf{\epsilon}||_\infty \leq \tau \) for some \( \tau \geq 0 \).
Inspired by Oppenheim and Lim, 1981; Boufounos, 2013, in the context of CS, let’s consider the phase-only (non-linear) sensing model:

\[ z = \text{sign}_C(Ax) + \epsilon \in \mathbb{C}^m, \]  

(PO-CS)

with:
- \( x \) is real and \( k \)-sparse;
- \( \text{sign}_C(re^{i\theta}) := e^{i\theta} \) (and 0 if \( r = 0 \)), applied pointwise;
- and \( \epsilon \in \mathbb{C}^m \) a bounded noise with \( \|\epsilon\|_\infty \leq \tau \) for some \( \tau \geq 0 \).

Key observations:

1. If \( x \to Cx \) with \( C > 0 \), \( z \) is unchanged (Signal amplitude is lost)

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

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

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

Since signal amplitude is lost, we still have $\text{sign}_C(Ax) = \text{sign}_C(Ax^*)$ with

$$x^* := \frac{\kappa \sqrt{m}}{\|Ax\|_1} x, \quad \text{with } \kappa := \sqrt{\frac{\pi}{2}}.$$ 

⇒ We now focus on the recovery of $x^*$ (→ encodes signal direction).

**Rationale:**

- Useful for our proofs;
- For complex Gaussian $\sqrt{m} A \sim \mathbb{CN}^{m \times n}(0, 2)$ and $g \sim \mathcal{N}(0, 1)$,

$$\mathbb{E}|g| = \kappa \Rightarrow \mathbb{E}\|Ax\|_1 = \kappa \sqrt{m} \|x\|_2 \Rightarrow \|x^*\|_2 \approx 1.$$ 

⇒ $x^*$ is (almost) a unit length vector, a direction
Principle: Turn the non-linear PO model into linear one. Step by step . . .

B. Estimate constraints: From the noiseless model

\[ z = \text{sign}_C(Ax^*), \]

we see that \( u = x^* \in \mathbb{R}^n \) respects both:

\[
\begin{align*}
\langle z, Au \rangle &= \kappa \sqrt{m} \quad \iff \quad \left\langle \frac{1}{\kappa \sqrt{m}} A^* z, u \right\rangle = 1
\end{align*}
\]

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

B. Estimate constraints: From the noiseless model

\[ z = \text{sign}_C(Ax^*) , \]

we see that \( u = x^* \in \mathbb{R}^n \) respects both:

\[
\begin{align*}
\langle z, Au \rangle &= \kappa \sqrt{m} \quad \Leftrightarrow \quad \langle \frac{1}{\kappa \sqrt{m}} A^* z, u \rangle = 1 \\
\text{(normalization)} \\
\text{if } u = x^* \\
\text{and } \\
\text{(phase consistency)} \\
\text{if } u = x^*
\end{align*}
\]

\[
\text{diag}(z)^* A u = \left( \begin{array}{c}
z_1^* \cdot (Au)_1 \\
\vdots \\
z_m^* \cdot (Au)_m
\end{array} \right)^\top \in \mathbb{R}^m_+.
\]
Principle: Turn the non-linear PO model into linear one. Step by step . . .

B. Estimate constraints: From the noiseless model

$$z = \text{sign}_\mathbb{C}(Ax^*)$$,

we see that $$u = x^* \in \mathbb{R}^n$$ respects both:

\[
\begin{align*}
\langle z, Au \rangle &= \kappa \sqrt{m} \iff \langle \frac{1}{\kappa \sqrt{m}} A^* z, u \rangle = 1 \\
&= \|Ax^*\|_1 \text{ if } u = x^* \\
\end{align*}
\]

(normalization)

\[
\begin{align*}
\text{diag}(z)^* Au &= (z_1^* \cdot (Au)_1, \cdots, z_m^* \cdot (Au)_m)^\top \in \mathbb{R}^m \\
&= |(Ax^*)_1| \text{ if } u = x^* \\
&= |(Ax^*)_m| \text{ if } u = x^* \\
\end{align*}
\]

(phase consistency)

Let's relax the phase consistency, i.e., impose \(\text{diag}(z)^* Au \in \mathbb{R}^m\), that is

$$0 = \mathfrak{S}(\text{diag}(z)^* Au) = \left(\text{diag}(z)^\mathbb{R} A^\mathbb{S} - \text{diag}(z)^\mathbb{S} A^\mathbb{R}\right)u =: H_z u.$$

noting also that

$$\langle \alpha_z, u \rangle = 1 \iff \langle \alpha_z^\mathbb{R}, u \rangle = 1, \langle \alpha_z^\mathbb{S}, u \rangle = 0.$$
In summary, \( u = x^* \) respects the relaxed, real \( m + 2 \) constraints . . .

\[
A_z u = e_1 := (1, 0, \cdots, 0)^\top \quad \Rightarrow \quad \text{This is a linear sensing model!}
\]

with

\[
A_z := (\alpha_z^\Re, \alpha_z^\Im, H_z^\top)^\top \in \mathbb{R}^{(m+2) \times n}.
\]

In other words,

- A good estimate \( \hat{x} \) of \( x^* \) should respect the linear model \( A_z \hat{x} = e_1 \) since \( x^* \in \{ u \in \mathbb{R}^n : A_z \hat{u} = e_1 \} \).
- We know this estimate should be sparse (as \( x^* \) is)
In summary, \( u = x^* \) respects the relaxed, real \( m + 2 \) constraints . . .

\[
A_z u = e_1 := (1, 0, \cdots, 0)^\top \quad \Rightarrow \quad \text{This is a linear sensing model!}
\]

with

\[
A_z := (\alpha^R_z, \alpha^\Im_z, H_z^\top)^\top \in \mathbb{R}^{(m+2) \times n}.
\]

In other words,

- A good estimate \( \hat{x} \) of \( x^* \) should respect the linear model \( A_z \hat{x} = e_1 \)
  since \( x^* \in \{ u \in \mathbb{R}^n : A_z \hat{u} = e_1 \} \).
- We know this estimate should be sparse (as \( x^* \) is)

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

\[
\hat{x} = \arg \min_{u \in \mathbb{C}^n} \|u\|_1 \quad \text{s.t.} \quad A_z u = e_1, \quad \text{(BP}(A_z, e_1))
\]

**Question:** How far is \( \hat{x} \) from \( x^* \)? Well, let’s see if \( A_z \) respects the RIP!
RIP for $A_z$?

How could $A_z := (\alpha_z^R, \alpha_z^S, H_z^\top)^\top$ respect the RIP?

For a sparse $v$, $\|A_zv\|_2^2 := |\langle \alpha_z, v \rangle|^2 + \|H_zv\|_2^2$

you can show that, for complex Gaussian $A$:

- $\langle \alpha_z, v \rangle \approx \langle x/\|x\|_2, v \rangle \approx$ projection of $v$ onto $\mathcal{X} := \mathbb{R} \bar{x}$.
- $H_zx = 0$ & $H_z$ RIP on $\mathcal{X}^\perp \cap 2k$-sparse signals.
RIP for $A_z$?

How could $A_z := (\alpha_z^R, \alpha_z^\Im, H_z^\top)^\top$ respect the RIP?

For a sparse $v$, $\|A_z v\|_2^2 := |\langle \alpha_z, v \rangle|^2 + \|H_z v\|_2^2 \approx \langle \frac{x}{\|x\|_2}, v \rangle^2 + \|v^\perp\|_2^2 = \|v\|_2^2$

you can show that, for complex Gaussian $A$:

- $\langle \alpha_z, v \rangle \approx \langle \frac{x}{\|x\|_2}, v \rangle \approx$ projection of $v$ onto $\mathcal{X} := \mathbb{R} \bar{x}$.
- $H_z x = 0$ & $H_z$ RIP on $\mathcal{X}^\perp \cap 2k$-sparse signals.

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

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

then, w.h.p., $A_z$ satisfies the RIP $(k, \delta)$.

**Consequences:**

- For $\hat{x} = \text{BP}(A_z, e_1)$, if $A_z$ is RIP($\delta < \delta_0, 2k$), we get exact reconstruction of signal direction, i.e., $\hat{x} = x^*$!

- Instance optimality for the noisy setting (with BPDN) (not covered here)
Let’s plot a *phase-transition curve*: we generate $\sqrt{m}A \sim \mathbb{C}N^{m\times256}(0, 2)$ &

- 20-sparse vectors in $\mathbb{R}^{256}$;
- $m \in [1, 256]$ and average over 100 trials;
- Reconstruction successful if SNR $\geq 60$ dB.
Let’s plot a phase-transition curve: we generate \( \sqrt{m}A \sim \mathbb{C}N^{m \times 256}(0, 2) \) &

- 20-sparse vectors in \( \mathbb{R}^{256} \);
- \( m \in [1, 256] \) and average over 100 trials;
- Reconstruction successful if SNR \( \geq 60 \) dB.
Let’s be a little more daring . . . and forget Gauss
Let’s be a little more daring … and forget Gauss

**Bernoulli random matrix**

\[ A_{ij} \sim_{iid} \{ \pm 1 \} \]

**Random partial Fourier**

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

Interestingly:
- These results are not covered by theory.
- Bernoulli random matrices do not work for 1-bit CS.
- Fourier sensing has PO-CS counter-examples (that cannot be recovered)!

\[ e.g., \quad x' = h^* x \quad \text{with} \quad \hat{h}_k > 0, \quad \forall k, \quad \text{sign} C(Ax') = \text{sign} C(Ax) \]
Let’s be a little more daring . . . and forget Gauss

Bernoulli random matrix

\[ A_{ij} \sim \text{iid} \{ \pm 1 \} \]

Random partial Fourier

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

Interestingly:

- These results are not covered by theory.
- Bernoulli random matrices do not work for 1-bit CS.
- Fourier sensing has PO-CS counter-examples (that cannot be recovered)!

\[ e.g., \text{ for } x' := h \ast x \text{ with } \hat{h}_k > 0, \forall k, \quad \text{sign}_\mathbb{C}(Ax') = \text{sign}_\mathbb{C}(Ax). \]
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 $x^*$).

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 (2020, submitted).

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.

Candes, 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
Extra simulations: noisy case

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

- 20-sparse vectors in $\mathbb{R}^{256}$;
- $m \in [1, 256]$ and average over 100 trials;
- $z = \text{sign}_\mathbb{C}(Ax) + \xi$, with $\xi \in \mathbb{C}^m$ and $\|\xi\|_\infty \leq \tau$. 

![Graph showing Mean SNR vs. m/s with a color scale ranging from -10 dB to 50 dB, and a color bar from 1.0 to 3.0 representing -log10(τ/m).]
Let's first simplify the context...

1. We consider the sensing of real vectors $x \in \mathbb{R}^n$. Note: If complex signal $x$, we can always rewrite $A x = (A \Re + i A \Im)(x \Re + i x \Im) = A \tilde{x}$, with $\tilde{x} \in \mathbb{R}^{2n}$ and $A \in \mathbb{C}^{m \times 2n}$.

Caveat: This can impact the signal model e.g., sparse in $\mathbb{C}^n$ - 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 $17$)...
Phase-only observation in Compressive Sensing?

Let’s first simplify the context . . .

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

   Note: If complex signal $x$, we can always rewrite

   $$A x = (A^\Re + iA^\Im)(x^\Re + ix^\Im) = (A, iA) \begin{pmatrix} x^\Re \\ x^\Im \end{pmatrix} = \overline{A} \overline{x},$$

   with $\overline{x} \in \mathbb{R}^{2n}$ and $\overline{A} \in \mathbb{C}^{m \times 2n}$.
Phase-only observation in Compressive Sensing?

Let’s first simplify the context . . .

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

   Note: If complex signal $x$, we can always rewrite

   $$
   A x = (A^\Re + iA^\Im)(x^\Re + ix^\Im) = (A, iA)\begin{pmatrix} x^\Re \\ x^\Im \end{pmatrix} = \overline{A} \bar{x},
   $$

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

   **Caveat:** This can impact the signal model

   e.g., sparse in $\mathbb{C}^n \equiv$ group sparse in $\mathbb{R}^{2n}$. 


Simplifying hypothesis

Phase-only observation in Compressive Sensing?

Let’s first simplify the context . . .

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

   Note: If complex signal $x$, we can always rewrite

   $$Ax = (A^\Re + iA^\Im)(x^\Re + ix^\Im) = (A, iA) \begin{pmatrix} x^\Re \\ x^\Im \end{pmatrix} = \overline{A} \bar{x},$$

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

   **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*)