$\ell_1$ minimization without amplitude information

Petros Boufounos
petrosb@rice.edu
Classical $\ell_1$ reconstruction problems:

\[
\min_{x} \|x\|_1 \quad \text{s.t.} \quad f(x) = 0
\]

\[
\min_{x} \|x\|_1 \quad \text{s.t.} \quad f(x) \geq 0
\]

\[
\min_{x} \|x\|_1 + \lambda f(x)
\]
The Big $\ell_1$ Picture

Classical $\ell_1$ reconstruction problems:

\[
\begin{align*}
\min_x \|x\|_1 \quad & \text{s.t. } f(x) = 0 \\
\min_x \|x\|_1 \quad & \text{s.t. } f(x) \geq 0 \\
\min_x \|x\|_1 + \lambda f(x) \quad & \text{Sparsity Model}
\end{align*}
\]

Data Fidelity:

\[
\begin{align*}
f(x) &= \|\Phi x - y\|_2 \\
f(x) &= \epsilon - \|\Phi x - y\|_2
\end{align*}
\]
Data without amplitude information

\[
\min_x \|x\|_1 + \lambda f(x)
\]

Q: What if data provide no amplitude information on \(x\)?
Data without amplitude information

$$\min_x \|x\|_1 + \lambda f(x)$$

Q: What if data provide no amplitude information on $x$?

$$f(\alpha x) = \alpha f(x)$$

Minimizing solution is degenerate:

$$x = 0$$

Impose an amplitude constraint.
Case I: Sampling the Zero Crossings
Q: Given only the zero crossings \( \{t_1, t_2, \ldots, t_N\} \) of a signal can we reconstruct it?
Q: Given only the zero crossings \( \{t_1, t_2, \ldots, t_N\} \) of a signal can we reconstruct it?

Easy implementation: only need a comparator and a clock
Q: Given only the zero crossings \( \{t_1, t_2, ..., t_N\} \) of a signal can we reconstruct it?

Logan’s Theorem: YES. Signals bandlimited to \([B,2B)\) are uniquely determined by their zero crossings.

BUT: an arbitrary set of zero crossings might not correspond to a signal bandlimited to \([B,2B)\). Reconstruction is not robust. There is ambiguity.
Q: Given only the zero crossings \( \{t_1, t_2, \ldots, t_N\} \) of a signal can we reconstruct it?

Logan’s Theorem: **YES.** Signals bandlimited to \([B, 2B)\) are uniquely determined by their zero crossings.

**BUT:** an arbitrary set of zero crossings might not correspond to a signal bandlimited to \([B, 2B)\). Reconstruction is not robust. There is ambiguity.

Introduce sparsity to resolve the ambiguity!
Fourier series of $x(t)$:

$$x(t) = \sum_{n \in \mathcal{B}} [a_n \cos(2\pi nt) + b_n \sin(2\pi nt)]$$

Vector of coefficients:

$$\mathbf{x} = \begin{bmatrix} a_{n_1} \\ \vdots \\ a_{n_{N/2}} \\ b_{n_1} \\ \vdots \\ b_{n_{N/2}} \end{bmatrix}$$
Given \( \{t_1,t_2,\ldots,t_N\} \),

\[
\Phi_{\{t_k\}} = \begin{bmatrix}
\cos (2\pi n_1 t_1) & \ldots & \cos (2\pi n_N/2 t_1) & \sin (2\pi n_1 t_1) & \ldots & \sin (2\pi n_N/2 t_1) \\
\cos (2\pi n_1 t_2) & \ldots & \cos (2\pi n_N/2 t_2) & \sin (2\pi n_1 t_2) & \ldots & \sin (2\pi n_N/2 t_2) \\
: & \ddots & : & : & \ddots & : \\
\cos (2\pi n_1 t_N) & \ldots & \cos (2\pi n_N/2 t_N) & \sin (2\pi n_1 t_N) & \ldots & \sin (2\pi n_N/2 t_N)
\end{bmatrix}
\]

Samples the signal at those times:

\[
\Phi_{\{t_k\}} \mathbf{x} = \begin{bmatrix}
x(t_1) \\
: \\
x(t_N)
\end{bmatrix}
\]
Reconstruction Problem

If \( T = \{t_1, t_2, \ldots, t_N\} \) are the zero crossings, then the desired signal is in the nullspace of \( \Phi \): \( \Phi_T x = 0 \).

Logan’s theorem \( \Rightarrow \) \( \Phi_T \) has a one-dimensional nullspace.
Signal Acquisition and Reconstruction

\[ x(t) \]

\[ \{ t_1, t_2, \ldots, t_N \} \]

Clock
Signal Acquisition and Reconstruction

\[ x(t) \]

\[ \{ t_1, t_2, \ldots, t_N \} \]

\[ \Phi_T \]

Build \( \Phi_T \)

Find 1-D nullspace (e.g., SVD)

\[ x \]
Signal Acquisition and Reconstruction

$x(t) \rightarrow $ Clock $\rightarrow \{t_1, t_2, ..., t_N\} \rightarrow$ Build $\Phi_T$ \rightarrow Find 1-D nullspace (e.g. SVD) \rightarrow $ x

$T = \{t_1, t_2, ..., t_N\}$

In practice: noise and quantization. No nullspace! Many small singular values. Ambiguity!

$\text{SVD}(\Phi_T)$
Sparse Reconstruction

\[ \ell_1 \text{ minimization:} \]

\[ \hat{x} = \arg \min_x \|x\|_1 \]

subject to \( \Phi x = 0 \)
Sparse Reconstruction

$\ell_1$ minimization:

$$\hat{x} = \arg \min_x \|x\|_1$$

subject to $\Phi x = 0$

Relaxation:

$$\hat{x} = \arg \min_x \|x\|_1 + \frac{\lambda}{2} \|\Phi x\|_2^2$$
Sparse Reconstruction

\[ \ell_1 \text{ minimization:} \]
\[ \hat{x} = \arg \min_x \|x\|_1 \]
subject to \( \Phi x = 0 \)

\[ \text{Relaxation:} \]
\[ \hat{x} = \arg \min_x \|x\|_1 + \frac{\lambda}{2} \|\Phi x\|_2^2 \]

\[ \text{Unit energy constraint:} \]
\[ \hat{x} = \arg \min_x \|x\|_1 + \frac{\lambda}{2} \|\Phi x\|_2^2 \]
subject to \( \|x\|_2 = 1 \)
\[ \hat{x} = \arg \min_x \|x\|_1 + \frac{\lambda}{2} \|\Phi x\|_2^2 \]
subject to \( \|x\|_2 = 1 \)

**Unconstrained minimization:**
\[
\text{Cost}(x) = g(x) + \frac{\lambda}{2} f(\Phi x)
\]
\[
\text{Cost'}(x) = g'(x) + \frac{\lambda}{2} \Phi^* f'(\Phi x)
\]

where:
\[
(g'(x))_i = \begin{cases} 
-1 & x_i < 0 \\
[-1, 1] & x_i = 0 \\
+1 & x_i > 0 
\end{cases}
\]

No change if gradients are projected on unit sphere
Minimization Algorithm (based on FPC [Hale, Yin, Zhang, ‘07])

Big Picture: Gradient descent until equilibrium.

Initialization parameters: \( \hat{x}, \tau \)

1. Compute quadratic gradient: \( h = \Phi^T \Phi \hat{x} \)
2. Project onto sphere: \( h_p = h - \langle \hat{x}, h \rangle \)
3. Quadratic gradient descent: \( \hat{x} \leftarrow \hat{x} - \tau h_p \)
4. Shrink (\( \ell_1 \) gradient descent):
   \[
   \hat{x}_i \leftarrow \text{sign}(\hat{x}_i) \max \left\{ |\hat{x}_i| - \frac{\tau}{\lambda}, 0 \right\}
   \]
5. Normalize: \( \hat{x} \leftarrow \frac{\hat{x}}{\|\hat{x}\|} \)
6. Iterate until equilibrium.
Optimization on the Sphere

\[ \hat{x} = \arg \min_x \|x\|_1 + \frac{\lambda}{2} \|\Phi x\|_2^2 \]

subject to \( \|x\|_2 = 1 \)

Optimization is not convex.

Convergence to global optimum not guaranteed.
Optimization on the Sphere

\[ \hat{x} = \arg \min_x \|x\|_1 + \frac{\lambda}{2}\|\Phi x\|_2^2 \]
subject to \( \|x\|_2 = 1 \)

Optimization is \textit{not convex}.

Convergence to global optimum \textit{not guaranteed}.

Exploit \textit{randomness}:

• Execute \( L \) times with random initializations.
• Pick best solution.

If \( P = P(\text{success for 1 execution}) \), then
\[ P(\text{overall success}) = 1 - (1 - P)^L \]
Results

Probability of Success

$L =$ number of random initializations

$N =$ 256 coefficients

$L =$ number of random initializations
$N =$ 256 coefficients
Results

Probability of Success

L=number of random initializations
N=256 coefficients
Optimization on sphere:

\[ \hat{x} = \arg \min_{x} \|x\|_1 + \frac{\lambda}{2} \|\Phi x\|_2^2 \]

subject to \( \|x\|_2 = 1 \)

Relaxation of sphere constraint:

\[ \hat{x} = \arg \min_{x} \|x\|_1 + \frac{\lambda_1}{2} \|\Phi x\|_2^2 + \lambda_2 \|\|x\|_2^2 - 1\|^2 \]

We can now use standard \( \ell_1 \) algorithms!
\[ \hat{x} = \arg \min_x \|x\|_1 + \frac{\lambda_1}{2} \|\Phi x\|_2^2 + \lambda_2 \|x\|_2^2 - 1 \|^2 \]

Let:

\[
\tilde{\Phi} = \begin{bmatrix} cx \\ \Phi \end{bmatrix}, \quad c = \left( \frac{\lambda_2}{\lambda_1} \right)^2
\]

At equilibrium:

\[ \hat{x} = \arg \min_x \|x\|_1 + \frac{\lambda_1}{2} \|\tilde{\Phi} x - \begin{bmatrix} c \\ 0 \end{bmatrix}\|^2_2 \]
Reweighted FPC algorithm

Initialization parameters: \( \hat{x}, \lambda_1, \lambda_2 \)

1. **Build**
   \[
   \tilde{\Phi} = \begin{bmatrix} c\hat{x} \\ \Phi \end{bmatrix}, \quad c = \left( \frac{\lambda_2}{\lambda_1} \right)^2
   \]

2. **Estimate** using FPC:
   \[
   \hat{x} = \arg\min_x \|x\|_1 + \frac{\lambda_1}{2} \left\| \tilde{\Phi} x - \begin{bmatrix} c \\ 0 \end{bmatrix} \right\|_2^2
   \]

3. **Iterate** until equilibrium.
Results

Probability of Success

$L =$ number of random initializations
$N =$ 256 coefficients

$L = 1$
$L = 5$
$L = 10$
$L = 20$

$\%$ Signal Sparsity ($2K/N$)

Probability of success

---

L = number of random initializations
N = 256 coefficients
Results

Probability of Success

% Signal Sparsity (2K/N) vs. Probability of Success

L=number of random initializations
N=256 coefficients

Reweighted FPC
Case II:
1-bit Compressive Sensing
Q: Can we quantize measurements to 1-bit:

\[ y = \text{sign}(\Phi x) \]

\[ y_i = \text{sign}(\langle \phi_i, x \rangle) \]

and recover the signal (within a positive scaling factor)?
Q: Can we quantize measurements to 1-bit:

\[ y = \text{sign}(\Phi x) \]
\[ y_i = \text{sign}(\langle \phi_i, x \rangle) \]

and recover the signal (within a positive scaling factor)?

1-bit measurements are inexpensive.

Focus on bits rather than measurements.

Exact recovery is not possible.
Sign information from 1-bit measurements:

\[ y_i = \text{sign}(\Phi x)_i \iff y_i \cdot (\Phi x)_i \geq 0 \]

Reconstruction should enforce \textit{model}.
Reconstruction should be \textit{consistent} with measurements.

\[ \hat{x} = \arg\min_x \|x\|_1 \]
\[ \text{subject to } y_i \cdot (\Phi x)_i \geq 0 \]
Reconstruction from 1-bit Measurements

Sign information from 1-bit measurements:

\[ y_i = \text{sign}(\Phi x)_i \iff y_i \cdot (\Phi x)_i \geq 0 \]

Reconstruction should enforce model.
Reconstruction should be consistent with measurements.
Reconstruction should enforce a non-trivial solution.

\[ \hat{x} = \arg \min_x \|x\|_1 \]
subject to \[ y_i \cdot (\Phi x)_i \geq 0 \]
and \[ \|x\|_2 = 1 \]
Information in 1-bit Measurements
Information in 1-bit Measurements
Information in 1-bit Measurements
Information in 1-bit Measurements
Information in 1-bit Measurements
Information in 1-bit Measurements
\[
\hat{x} = \arg \min_x \|x\|_1 \\
\text{subject to } y_i \cdot (\Phi x)_i \geq 0 \\
\text{and } \|x\|_2 = 1
\]
Constraint Relaxation

\[ \hat{x} = \arg \min_x \|x\|_1 \]

subject to \( y_i \cdot (\Phi x)_i \geq 0 \)

and \( \|x\|_2 = 1 \)

We relax the inequality constraints:

\[ \hat{x} = \arg \min_x \|x\|_1 + \frac{\lambda}{2} \sum_i f(y_i \cdot (\Phi x)) \]

subject to \( \|x\|_2 = 1 \)

where \( f(x) \) is a one sided quadratic:

\[ f(x) = \begin{cases} 
  x^2 & x \leq 0 \\
  0 & x > 0 
\end{cases} \]
Fixed point equilibrium

\[ \hat{x} = \arg \min_x \|x\|_1 + \frac{\lambda}{2} \sum_i f(y_i \cdot (\Phi x)) \]

subject to \( \|x\|_2 = 1 \)

Unconstrained minimization:

\[ Y \equiv \text{diag}(y) \]

\[ \text{Cost}(x) = g(x) + \frac{\lambda}{2} f(Y \Phi x) \]

\[ \text{Cost}'(x) = g'(x) + \frac{\lambda}{2} (Y \Phi)^T f(Y \Phi x) \]

\( (g'(x))_i = \begin{cases} 
-1 & x_i < 0 \\
[-1, 1] & x_i = 0 \\
+1 & x_i > 0 
\end{cases} \)

and \( \left( \frac{f'(x)}{2} \right)_i = \begin{cases} 
-x_i & x_i \leq 0 \\
0 & x_i > 0 
\end{cases} \)

No change if gradients are projected on unit sphere.
Minimization Algorithm

Big Picture: Gradient descent until equilibrium.

**Initialization parameters:** $\hat{x}$, $\tau$

1. **Compute quadratic gradient:** $h = (Y\Phi)^T f'(Y\Phi x)$

2. **Project onto sphere:** $h_p = h - \langle \hat{x}, h \rangle$

3. **Quadratic gradient descent:** $\hat{x} \leftarrow \hat{x} - \tau h_p$

4. **Shrink ($\ell_1$ gradient descent):**
   $$\hat{x}_i \leftarrow \text{sign}(\hat{x}_i) \max \left\{ |\hat{x}_i| - \frac{\tau}{\lambda}, 0 \right\}$$

5. **Normalize:** $\hat{x} \leftarrow \frac{\hat{x}}{\|\hat{x}\|}$

6. **Iterate until equilibrium.**
Reconstruction Error ($N=512$)

(a) $K=16$

(b) $K=32$

(c) $K=64$

(d) $K=128$
If the signal is an image, we have more information! (i.e., a better signal model)

Images are sparse in wavelets and positive:

\[ x = W \alpha \]
\[ x_i \geq 0 \]
and \( \alpha \) is sparse

Incorporate better model in the reconstruction:

\[ \hat{\alpha} = \arg \min_\alpha \| \alpha \|_1 \]
subject to \( y_i \times (\Phi W)_i \geq 0 \)
and \( (W \alpha)_i \geq 0 \)
and \( \| \alpha \|_2 = 1 \)
Original Image
4096 pixels
256 levels
Results

Original Image
4096 pixels
256 levels

Classical Compressive Sensing, 1 bit per pixel

<table>
<thead>
<tr>
<th>Measurements</th>
<th>Bits per Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>4096</td>
<td>1</td>
</tr>
<tr>
<td>2048</td>
<td>2</td>
</tr>
<tr>
<td>1024</td>
<td>4</td>
</tr>
<tr>
<td>512</td>
<td>8</td>
</tr>
</tbody>
</table>
Results

Reconstruction on unit sphere
1 bit per pixel

Classical Compressive Sensing, 1 bit per pixel

Original Image
4096 pixels
256 levels

4096 measurements
1 bit per measurement

4096 measurements
1 bit per measurement
Results

Original Image
- 4096 pixels
- 256 levels

4096 measurements
1 bit per measurement
4096 bits (1 bit per pixel)

Reconstruction on unit sphere

Classical Compressive Sensing

512 measurements
8 bits per measurement
Results

Classical Compressive Sensing

Original Image
4096 pixels
256 levels

4096 measurements
1 bit per measurement
4096 bits (1 bit per pixel)

Reconstruction on unit sphere

512 measurements
8 bits per measurement

512 measurements
1 bit per measurement
512 bits (0.125 bits per pixel)
Concluding Remarks

- Practical systems may eliminate amplitude information
- Reconstruction on the unit sphere is necessary
- The sphere is a well-behaved manifold
- Unit sphere constraint reduces the search space
- Several still open questions