Phase Transitions for Greedy Sparse Approximation Algorithms

Andrew Thompson

University of Edinburgh

Joint with Blanchard (Grinnell), Cartis and Tanner (Edinburgh)
Compressed Sensing

Let \( x \in \mathbb{R}^N \) be a given signal.

Suppose we obtain a vector \( b \in \mathbb{R}^n \) of noisy linear measurements

\[
b = Ax + e,
\]

where \( A \in \mathbb{R}^{n \times N} \) is the measurement matrix, and \( e \) is noise.
Compressed Sensing

Let $x \in \mathbb{R}^N$ be a given signal.

Suppose we obtain a vector $b \in \mathbb{R}^n$ of noisy linear measurements

$$b = Ax + e,$$

where $A \in \mathbb{R}^{n \times N}$ is the measurement matrix, and $e$ is noise.

We assume:

- $n < N$ \implies \text{underdetermined system}
- $x$ sparse with $k < n$ non-zeros
Algorithms for Compressed Sensing

- Optimization algorithms for the convex relaxation

\[
\min_{x \in \mathbb{R}^N} \|x\|_1 \quad \text{subject to} \quad \|Ax - b\|_2 \leq \eta
\]
Algorithms for Compressed Sensing

- Optimization algorithms for the **convex relaxation**

\[
\min_{x \in \mathbb{R}^N} \|x\|_1 \quad \text{subject to} \quad \|Ax - b\|_2 \leq \eta
\]

- **Greedy methods**
Algorithms for Compressed Sensing

• Optimization algorithms for the convex relaxation

\[
\min_{x \in \mathbb{R}^N} \|x\|_1 \quad \text{subject to} \quad \|Ax - b\|_2 \leq \eta
\]

• Greedy methods
  • Iterative Hard Thresholding (Blumensath/Davies 2008):

\[
x^{l+1} = H_k \left( x^l + \omega A^* (b - Ax^l) \right)
\]

where \( H_k : \mathbb{R}^N \rightarrow \mathbb{R}^N \) keeps the \( k \) largest entries.

\[\equiv \text{gradient projection for } \min_{x \in \mathbb{R}^N} \|b - Ax\|_2^2 \quad \text{s.t.} \quad \|x\|_0 \leq k \]
Algorithms for Compressed Sensing

- Optimization algorithms for the **convex relaxation**

\[
\min_{x \in \mathbb{R}^N} \|x\|_1 \quad \text{subject to} \quad \|Ax - b\|_2 \leq \eta
\]

- **Greedy methods**
  - Iterative Hard Thresholding (Blumensath/Davies 2008):
    \[
x^{l+1} = H_k (x^l + \omega A^* (b - Ax^l))
    \]

where \(H_k : \mathbb{R}^N \rightarrow \mathbb{R}^N\) keeps the \(k\) largest entries.

\[\equiv\] gradient projection for \(\min_{x \in \mathbb{R}^N} \|b - Ax\|_2^2\) s.t. \(|x|_0 \leq k\)

- We analyse recovery guarantees for three recently proposed greedy algorithms.
Restricted Isometry Property

Restricted Isometry Constants:

\[ L_k := \min_{c \geq 0} c \text{ subject to } (1 - c)\|x\|_2^2 \leq \|Ax\|_2^2 \text{ for all } k\text{-sparse } x \]

\[ U_k := \min_{c \geq 0} c \text{ subject to } (1 + c)\|x\|_2^2 \geq \|Ax\|_2^2 \text{ for all } k\text{-sparse } x \]

\[ R_k := \max\{L_k, U_k\} \]
Restricted Isometry Property

Restricted Isometry Constants:

\[ L_k := \min_{c \geq 0} c \text{ subject to } (1 - c)\|x\|_2^2 \leq \|Ax\|_2^2 \text{ for all } k\text{-sparse } x \]

\[ U_k := \min_{c \geq 0} c \text{ subject to } (1 + c)\|x\|_2^2 \geq \|Ax\|_2^2 \text{ for all } k\text{-sparse } x \]

\[ R_k := \max\{L_k, U_k\} \]

Example: For \( A_\Lambda n \times k \) i.i.d. Gaussian \( \mathcal{N}(0, 1/n) \) entries, as \((k, n) \to \infty\): in expectation,

\[ \lambda_{\min}(A_\Lambda^* A_\Lambda) \to (1 - \sqrt{k/n})^2, \]

\[ \lambda_{\max}(A_\Lambda^* A_\Lambda) \to (1 + \sqrt{k/n})^2. \]
Convergence result for IHT

Let $x$ be $k$-sparse and let $b = Ax + e$ where $A \in \mathbb{R}^{n \times N}$.

**Theorem:** There exist $\mu_{iht}(k, n, N)$ and $\xi_{iht}(k, n, N)$ which are functions of $L(k, n, N)$ and $U(k, n, N)$, such that, provided $\mu_{iht}(k, n, N) < 1$,

$$
\|x^l - x\|_2 \leq \left[\mu_{iht}(k, n, N)\right]^l \|x\|_2 + \frac{\xi_{iht}(k, n, N)}{1 - \mu_{iht}(k, n, N)} \|e\|_2.
$$
Convergence result for IHT

Let $x$ be $k$-sparse and let $b = Ax + e$ where $A \in \mathbb{R}^{n \times N}$.

**Theorem:** There exist $\mu^{iht}(k, n, N)$ and $\xi^{iht}(k, n, N)$ which are functions of $L(k, n, N)$ and $U(k, n, N)$, such that, provided $\mu^{iht}(k, n, N) < 1$,

$$\|x^l - x\|_2 \leq \left[\mu^{iht}(k, n, N)\right]^l \|x\|_2 + \frac{\xi^{iht}(k, n, N)}{1 - \mu^{iht}(k, n, N)} \|e\|_2.$$

**Corollary:** $e = 0 \Rightarrow x^l \rightarrow x$ at a linear rate.
Convergence result for IHT

Let $x$ be $k$-sparse and let $b = Ax + e$ where $A \in \mathbb{R}^{n \times N}$.

**Theorem:** There exist $\mu^{iht}(k, n, N)$ and $\xi^{iht}(k, n, N)$ which are functions of $L(k, n, N)$ and $U(k, n, N)$, such that, provided $\mu^{iht}(k, n, N) < 1$,

$$
\|x^l - x\|_2 \leq \left[\mu^{iht}(k, n, N)\right]^l \|x\|_2 + \frac{\xi^{iht}(k, n, N)}{1 - \mu^{iht}(k, n, N)} \|e\|_2.
$$

**Corollary:** $e = 0 \Rightarrow x^l \rightarrow x$ at a linear rate.

But when is it true that $\mu^{iht}(k, n, N) < 1$?
Gaussian RIP Upper Bounds

(Blanchard, Cartis and Tanner, 2009)

**Theorem:** Let $A$ be a matrix of size $n \times N$ whose entries are drawn i.i.d. from $\mathcal{N}(0, \frac{1}{n})$.
Let $(k, n, N) \to \infty$ with $\frac{k}{n} \to \rho$ and $\frac{n}{N} \to \delta$.
Then there exist numerically computable functions $L(\delta, \rho)$ and $U(\delta, \rho)$ such that, for any $\epsilon > 0$,

$$\mathbb{P}\{L(k, n, N) < L(\delta, \rho) + \epsilon\} \to 1,$$
$$\mathbb{P}\{U(k, n, N) < U(\delta, \rho) + \epsilon\} \to 1.$$
Gaussian RIP Upper Bounds

$L(\delta, \rho)$

$U(\delta, \rho)$
Gaussian RIP Upper Bounds

Lower bound on $L(k, n, N)$

Lower bound on $U(k, n, N)$

$\Rightarrow$ Gaussian RIP upper bounds are always within a factor of 1.83 of the exact RIP constants.
Gaussian RIP Upper Bounds

⇒ Gaussian RIP upper bounds are always within a factor of 1.83 of the exact RIP constants.

• Bounds further improved upon by Bah & Tanner (2010).
IHT with Gaussian Matrices

Let $x$ be $k$-sparse and let $b = Ax + e$, with entries in $A$ drawn i.i.d. from $\mathcal{N}(0, \frac{1}{n})$. Consider IHT with $\omega = \omega[L, U(\delta, 3\rho)]$.

**Theorem:** There exist $\mu^{iht}(\delta, \rho)$ and $\xi^{iht}(\delta, \rho)$ which are functions of $L(\delta, 3\rho)$ and $U(\delta, 3\rho)$, such that for any $\epsilon > 0$, as $(k, n, N) \to \infty$ with $n/N \to \delta \in (0, 1)$ and $k/n \to \rho$, there is an exponentially high probability on the draw of $A$ that

$$\|x^l - x\|_2 \leq \left[\mu^{iht}(\delta, \rho)\right]^l \|x\|_2 + \frac{\xi^{iht}(\delta, \rho)}{1 - \mu^{iht}(\delta, \rho)}\|e\|_2,$$

provided that $\rho < (1 - \epsilon)\rho^{iht}(\delta)$ where $\rho^{iht}(\delta)$ is defined to be the solution of $\mu^{iht}(\delta, \rho) = 1$. 
Lower bounds on IHT phase transition

Recovery guaranteed with exponentially high probability for Gaussian matrices with $(\delta, \rho)$ values below the curve.
Inverse of phase transition for IHT

At least $n = 907k$ measurements needed to guarantee recovery $\Rightarrow$ pessimistic result compared with average-case behaviour.
Stability to noise for IHT

(a) $\mu^{iht}(\delta, \rho)$

(b) $\xi^{iht}(\delta, \rho)/(1 - \mu^{iht}(\delta, \rho))$
Comparison of greedy algorithms

We performed similar analysis for two other greedy algorithms:

- **CoSaMP** (Needell/Tropp, 2009):
  A more sophisticated algorithm which employs a projection step to find the ‘best’ approximation to the signal for a given support.

- **Subspace Pursuit** (Dai/Milenkovic, 2008):
  Differs from CoSaMP only in the size of the support sets ($2k \rightarrow k$); and includes an extra projection step.
CoSaMP algorithm

(Needell/Tropp, 2009)

\( T_s : \mathbb{R}^N \rightarrow \mathbb{R}^N \) keeps \( s \) largest entries

Inputs: \( b, A \) and \( k \).

Initialize \( x^0 = 0 \) and \( y^0 = b \), and choose \( \eta > 0 \).

For \( l = 0, 1, 2, \ldots \), until \( \|Ax^l - b\|_2 < \eta \), do:

1. Form \( g = -A^*(Ax^l - b) \)

2. Let \( \Omega = \text{supp}(x^l) \cup \text{supp}(T_{2k}(g)) \) \hspace{1cm} |\Omega| \leq 3k

3. Let \( x_{\Omega}^{l+1} = T_k(P_{A\Omega}(b)) \) and set \( x_{\Omega^C}^{l+1} = 0 \)

End; output \( \hat{x} = x^l \).
Greedy phase transitions

\[ \rho \times 10^{-3} \]

\[ \delta \]

Graph showing different phases and transitions with \( \rho \) and \( \delta \) axes, and various curves labeled with different notations such as \( \rho_S^l(\delta) \), \( \rho_S^{int}(\delta) \), \( \rho_S^{sp}(\delta) \), and \( \rho_S^{csp}(\delta) \).
Inverse of the phase transition

\[ (\rho_S^L(\delta))^{-1} - (\rho_S^{ih}(\delta))^{-1} + (\rho_S^{sp}(\delta))^{-1} - (\rho_S^{csp}(\delta))^{-1} \]
RIP Conditions for $l_1$ Recovery

$$\min_{x \in \mathbb{R}^N} \|x\|_1 \quad \text{subject to} \quad Ax = b$$

Chartrand (2007):

$$bL([b+1]k, n, N) + U(bk, n, N) < b - 1; \quad b > 2$$

Candès (2008):

$$(1 + \sqrt{2})L(2k, n, N) + U(2k, n, N) < \sqrt{2}$$

Foucart, Lai (2009):

$$\frac{1 + U(2k, n, N)}{1 - L(2k, n, N)} < 4\sqrt{2} - 3$$
RIP Conditions for $l_1$ Recovery

\[
\min_{x \in \mathbb{R}^N} \|x\|_1 \text{ subject to } Ax = b
\]

Chartrand (2007):

\[
bL(\delta, [b + 1]\rho) + U(\delta, b\rho) < b - 1; \quad b > 2
\]

Candès (2008):

\[
(1 + \sqrt{2})L(\delta, 2\rho) + U(\delta, 2\rho) < \sqrt{2}
\]

Foucart, Lai (2009):

\[
\frac{1 + U(\delta, 2\rho)}{1 - L(\delta, 2\rho)} < 4\sqrt{2} - 3
\]
Comparison of $l_1$ Phase Transitions

The highest phase transitions are obtained by taking $b \approx 11$ in the result by Chartrand: $11L(12k, n, N) + U(11k, n, N) < 10$. 
Conclusions

• It is important to understand what RIP conditions mean quantitatively: the phase transition framework combined with RIP bounds for Gaussian matrices is a useful tool to investigate this.
Conclusions

• It is important to understand what RIP conditions mean quantitatively: the phase transition framework combined with RIP bounds for Gaussian matrices is a useful tool to investigate this.

• Recovery guarantees for the simpler IHT algorithm are in fact superior to those for the more complex CoSaMP and SP, with SP outperforming CoSaMP.
Conclusions

• It is important to understand what RIP conditions mean quantitatively: the phase transition framework combined with RIP bounds for Gaussian matrices is a useful tool to investigate this.

• Recovery guarantees for the simpler IHT algorithm are in fact superior to those for the more complex CoSaMP and SP, with SP outperforming CoSaMP.

• Recovery guarantees for all three greedy algorithms are still inferior to those for convex relaxation.
Conclusions

• It is important to understand what RIP conditions mean quantitatively: the phase transition framework combined with RIP bounds for Gaussian matrices is a useful tool to investigate this.

• Recovery guarantees for the simpler IHT algorithm are in fact superior to those for the more complex CoSaMP and SP, with SP outperforming CoSaMP.

• Recovery guarantees for all three greedy algorithms are still inferior to those for convex relaxation.

• Clear need for algorithm-specific methods of analysis.
Conclusions

• It is important to understand what RIP conditions mean quantitatively: the phase transition framework combined with RIP bounds for Gaussian matrices is a useful tool to investigate this.

• Recovery guarantees for the simpler IHT algorithm are in fact superior to those for the more complex CoSaMP and SP, with SP outperforming CoSaMP.

• Recovery guarantees for all three greedy algorithms are still inferior to those for convex relaxation.

• Clear need for algorithm-specific methods of analysis.

• It is not always quantitatively beneficial to have RIP conditions with the smallest possible support sizes.
Bibliography

- *Phase transitions for greedy sparse approximation algorithms*; J.Blanchard, C.Cartis, J.Tanner, AT (submitted 2009)
- *On support sizes of restricted isometry constants*; J.Blanchard, AT (2010; to appear, ACHA)
- *Improved bounds on restricted isometry constants for Gaussian matrices*; B.Bah, J.Tanner (submitted 2010)

All papers available on the Edinburgh Compressed Sensing website:

http://ecos.maths.ed.ac.uk
Actual form of $\mu$ and $\xi$ for IHT

For a given step-size $\omega$, the functions $\mu^{iht}(k, n, N)$ and $\xi^{iht}(k, n, N)$ take the form:

$$\mu^{iht}(k, n, N) = 2\sqrt{2} \max \{\omega [1 + U(3k, n, N)] - 1, 1 - \omega [1 - L(3k, n, N)]\};$$

$$\xi^{iht}(k, n, N) = 2\omega \sqrt{1 + U(2k, n, N)}.$$

For step-size $\omega = 2/[2 + U(\delta, 3\rho) - L(\delta, 3\rho)]$, the functions $\mu^{iht}(\delta, \rho)$ and $\xi^{iht}(\delta, \rho)$ take the form:

$$\mu^{iht}(\delta, \rho) = \frac{2\sqrt{2}[L(\delta, 3\rho) + U(\delta, 3\rho)]}{2 + U(\delta, 3\rho) - L(\delta, 3\rho)};$$

$$\xi^{iht}(\delta, \rho) = \frac{4[1 + U(\delta, 2\rho)]}{2 + U(\delta, 3\rho) - L(\delta, 3\rho)}.$$