Sparse Solutions of Underdetermined Linear Equations by Linear Programming

David Donoho & Jared Tanner

Stanford University, Department of Statistics
University of Utah, Department of Mathematics

Arizona State University: March 6^{th} 2006

Underdetermined systems, dictionary perspective

- Underdetermined system, infinite number of solutions

$$Ax = b, \qquad A \in \mathbb{R}^{d \times n}, \quad d < n$$

Underdetermined systems, dictionary perspective

- Underdetermined system, infinite number of solutions

$$Ax = b, \qquad A \in \mathbb{R}^{d \times n}, \quad d < n$$

- Least squares solution via "canonical dual" $(A^T A)^{-1} A^T$
  - Linear reconstruction, not signal adaptive
  - Solution vector full, n nonzero elements in x

Underdetermined systems, dictionary perspective

- Underdetermined system, infinite number of solutions

$$Ax = b, \qquad A \in \mathbb{R}^{d \times n}, \quad d < n$$

- Least squares solution via "canonical dual" $(A^T A)^{-1} A^T$
  - Linear reconstruction, not signal adaptive
  - Solution vector full, n nonzero elements in x
- Eschew redundancy, find *simple* model of data from A

Underdetermined systems, dictionary perspective

- Underdetermined system, infinite number of solutions

$$Ax = b, \qquad A \in \mathbb{R}^{d\times n}, \quad d < n$$

- Least squares solution via "canonical dual" $(A^T A)^{-1} A^T$
  - Linear reconstruction, not signal adaptive
  - Solution vector full, n nonzero elements in x
- Eschew redundancy, find *simple* model of data from A
- Seek sparsest solution, $\|x\|_{\ell^0} := \#$ nonzero elements

$$\min \|x\|_{\ell^0} \quad \text{subject to} \quad Ax = b$$

Underdetermined systems, dictionary perspective

- Underdetermined system, infinite number of solutions

$$Ax = b, \qquad A \in \mathbb{R}^{d \times n}, \quad d < n$$

- Least squares solution via "canonical dual" $(A^T A)^{-1} A^T$
  - Linear reconstruction, not signal adaptive
  - Solution vector full, n nonzero elements in x
- Eschew redundancy, find *simple* model of data from A
- Seek sparsest solution, $\|x\|_{\ell^0} := \#$ nonzero elements

$$\min \|x\|_{\ell^0} \quad \text{subject to} \quad Ax = b$$

- Combinatorial cost for naive approach

Underdetermined systems, dictionary perspective

- Underdetermined system, infinite number of solutions

$$Ax = b, \qquad A \in \mathbb{R}^{d \times n}, \quad d < n$$

- Least squares solution via “canonical dual” $(A^T A)^{-1} A^T$
  - Linear reconstruction, not signal adaptive
  - Solution vector full, n nonzero elements in x
- Eschew redundancy, find *simple* model of data from A
- Seek sparsest solution, $\|x\|_{\ell^0} := \#$ nonzero elements

$$\min \|x\|_{\ell^0} \quad \text{subject to} \quad Ax = b$$

- Combinatorial cost for naive approach
- Efficient nonlinear (signal adaptive) methods
  - Greedy (local) and Basis Pursuit (global)

Greedy [Temlyakov, DeVore, Tropp, ...]

- Orthogonal Matching Pursuit: initial $r = b$, $\tilde{A} = []$
  while $r \neq 0$

  $$\max_{\ell\infty} A^T r =: a_j^T r$$

Greedy [Temlyakov, DeVore, Tropp, ...]

- Orthogonal Matching Pursuit: initial $r = b$, $\tilde{A} = []$
  while $r \neq 0$

  $$\max_{\ell\infty} A^T r =: a_j^T r \qquad \tilde{A} = [\tilde{A} \;\; a_j]$$

Greedy [Temlyakov, DeVore, Tropp, ...]

- Orthogonal Matching Pursuit: initial $r = b$, $\tilde{A} = []$
  while $r \neq 0$

$$\max_{\ell^\infty} A^T r =: a_j^T r \qquad \tilde{A} = [\tilde{A} \;\; a_j] \qquad r = b - A(\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T b$$

Greedy [Temlyakov, DeVore, Tropp, ...]

- Orthogonal Matching Pursuit: initial $r = b$, $\tilde{A} = []$
  while $r \neq 0$

$$\max_{\ell^\infty} A^T r =: a_j^T r \qquad \tilde{A} = [\tilde{A} \;\; a_j] \qquad r = b - A(\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T b$$

- Nonlinear selection of basis, $x = (\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T b$; $\|x\|_{\ell^0} \leq d$

Greedy [Temlyakov, DeVore, Tropp, ...]

- Orthogonal Matching Pursuit: initial $r = b$, $\tilde{A} = []$
  while $r \neq 0$

$$\max_{\ell^\infty} A^T r =: a_j^T r \qquad \tilde{A} = [\tilde{A} \;\; a_j] \qquad r = b - A(\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T b$$

- Nonlinear selection of basis, $x = (\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T b$; $\|x\|_{\ell^0} \leq d$
- Highly redundant dictionary often give fast decay of residual

Greedy [Temlyakov, DeVore, Tropp, ...]

- Orthogonal Matching Pursuit: initial $r = b$, $\tilde{A} = []$
  while $r \neq 0$

$$\max_{\ell^\infty} A^T r =: a_j^T r \qquad \tilde{A} = [\tilde{A} \;\; a_j] \qquad r = b - A(\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T b$$

- Nonlinear selection of basis, $x = (\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T b$; $\|x\|_{\ell^0} \leq d$
- Highly redundant dictionary often give fast decay of residual
- Recovery of sparsest solution?
  • examples of arbitrary sub-optimality for a fixed dictionary A
  [Temlyakov, DeVore, S. Chen, Tropp, . . .]

Greedy [Temlyakov, DeVore, Tropp, ...]

- Orthogonal Matching Pursuit: initial $r = b$, $\tilde{A} = []$
  while $r \neq 0$

$$\max_{\ell^\infty} A^T r =: a_j^T r \qquad \tilde{A} = [\tilde{A} \;\; a_j] \qquad r = b - A(\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T b$$

- Nonlinear selection of basis, $x = (\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T b$; $\|x\|_{\ell^0} \leq d$
- Highly redundant dictionary often give fast decay of residual
- Recovery of sparsest solution?
  • examples of arbitrary sub-optimality for a fixed dictionary A [Temlyakov, DeVore, S. Chen, Tropp, . . .]
  • residual nonzero for steps $< d$, irregardless of sparsity [Chen]

Greedy [Temlyakov, DeVore, Tropp, ...]

- Orthogonal Matching Pursuit: initial $r = b$, $\tilde{A} = []$
  while $r \neq 0$

$$\max_{\ell^\infty} A^T r =: a_j^T r \qquad \tilde{A} = [\tilde{A} \;\; a_j] \qquad r = b - A(\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T b$$

- Nonlinear selection of basis, $x = (\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T b$; $\|x\|_{\ell^0} \leq d$
- Highly redundant dictionary often give fast decay of residual
- Recovery of sparsest solution?
  • examples of arbitrary sub-optimality for a fixed dictionary A [Temlyakov, DeVore, S. Chen, Tropp, ...]
  • residual nonzero for steps $< d$, irregardless of sparsity [Chen]
- Recover sparsest if sufficiently sparse, $\mathcal{O}(\sqrt{d})$ [Tropp]

Greedy [Temlyakov, DeVore, Tropp, ...]

- Orthogonal Matching Pursuit: initial $r = b$, $\tilde{A} = []$
  while $r \neq 0$

$$\max_{\ell^\infty} A^T r =: a_j^T r \qquad \tilde{A} = [\tilde{A} \;\; a_j] \qquad r = b - A(\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T b$$

- Nonlinear selection of basis, $x = (\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T b$; $\|x\|_{\ell^0} \leq d$
- Highly redundant dictionary often give fast decay of residual
- Recovery of sparsest solution?
  • examples of arbitrary sub-optimality for a fixed dictionary A [Temlyakov, DeVore, S. Chen, Tropp, ...]
  • residual nonzero for steps $< d$, irregardless of sparsity [Chen]
- Recover sparsest if sufficiently sparse, $\mathcal{O}(\sqrt{d})$ [Tropp]
- More sophisticated variants; weak greedy, swapping, etc ...

Greedy [Temlyakov, DeVore, Tropp, ...]

- Orthogonal Matching Pursuit: initial $r = b$, $\tilde{A} = []$
while $r \neq 0$

$$\max_{\ell^\infty} A^T r =: a_j^T r \qquad \tilde{A} = [\tilde{A} \;\; a_j] \qquad r = b - A(\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T b$$

- Nonlinear selection of basis, $x = (\tilde{A}^T \tilde{A})^{-1} \tilde{A}^T b$; $\|x\|_{\ell^0} \leq d$
- Highly redundant dictionary often give fast decay of residual
- Recovery of sparsest solution?
  • examples of arbitrary sub-optimality for a fixed dictionary A [Temlyakov, DeVore, S. Chen, Tropp, ...]
  • residual nonzero for steps $< d$, irregardless of sparsity [Chen]
- Recover sparsest if sufficiently sparse, $\mathcal{O}(\sqrt{d})$ [Tropp]
- More sophisticated variants; weak greedy, swapping, etc ...
- More about OMP for random sampling later

Basis Pursuit

- Rather than solve ℓ^0 (combinatorial), solve ℓ^1, (use LP)

$$\min \|x\|_{\ell^1} \quad \text{subject to} \quad Ax = b$$

  - Global basis selection rather than greedy local selection

Basis Pursuit

- Rather than solve ℓ^0 (combinatorial), solve ℓ^1, (use LP)

$$\min \|x\|_{\ell^1} \quad \text{subject to} \quad Ax = b$$

  - Global basis selection rather than greedy local selection
- Example, $A = [A_1 A_2]$ two ONB with coherence $\max_{ij}(a_i, a_j)$
  - If $\|x\|_{\ell^0} \lessapprox .914(1 + \mu^{-1})$ then $\ell^1 \to \ell^0$ [Elad, Bruckstein]
  - Coherence, $\mu := \max_{ij}(a_i, a_j) \geq 1/\sqrt{d}$, [Candes, Romberg]

Basis Pursuit

- Rather than solve ℓ^0 (combinatorial), solve ℓ^1, (use LP)

$$\min \|x\|_{\ell^1} \quad \text{subject to} \quad Ax = b$$

  - Global basis selection rather than greedy local selection
- Example, $A = [A_1 A_2]$ two ONB with coherence $\max_{ij}(a_i, a_j)$
  - If $\|x\|_{\ell^0} \lessapprox .914(1 + \mu^{-1})$ then $\ell^1 \to \ell^0$ [Elad, Bruckstein]
  - Coherence, $\mu := \max_{ij}(a_i, a_j) \geq 1/\sqrt{d}$, [Candes, Romberg]
- Ensures convergence for only the most sparse, $\mathcal{O}(\sqrt{d})$, signals.

Basis Pursuit

- Rather than solve ℓ^0 (combinatorial), solve ℓ^1, (use LP)

$$\min \|x\|_{\ell^1} \quad \text{subject to} \quad Ax = b$$

  - Global basis selection rather than greedy local selection
- Example, $A = [A_1 A_2]$ two ONB with coherence $\max_{ij}(a_i, a_j)$
  - If $\|x\|_{\ell^0} \lessapprox .914(1+\mu^{-1})$ then $\ell^1 \rightarrow \ell^0$ [Elad, Bruckstein]
  - Coherence, $\mu := \max_{ij}(a_i, a_j) \geq 1/\sqrt{d}$, [Candes, Romberg]
- Ensures convergence for only the most sparse, $\mathcal{O}(\sqrt{d})$, signals.
- Examples of failure: Dirac's Comb [Candes, ...]

Basis Pursuit

- Rather than solve ℓ^0 (combinatorial), solve ℓ^1, (use LP)

$$\min \|x\|_{\ell^1} \quad \text{subject to} \quad Ax = b$$

  - Global basis selection rather than greedy local selection
- Example, $A = [A_1 A_2]$ two ONB with coherence $\max_{ij}(a_i, a_j)$
  - If $\|x\|_{\ell^0} \lessapprox .914(1+\mu^{-1})$ then $\ell^1 \to \ell^0$ [Elad, Bruckstein]
  - Coherence, $\mu := \max_{ij}(a_i, a_j) \geq 1/\sqrt{d}$, [Candes, Romberg]
- Ensures convergence for only the most sparse, $\mathcal{O}(\sqrt{d})$, signals.
- Examples of failure: Dirac's Comb [Candes, ...]
- Is the story over? Can $\mathcal{O}(\sqrt{d})$ threshold be overcome? yes!

Basis Pursuit

- Rather than solve ℓ^0 (combinatorial), solve ℓ^1, (use LP)

$$\min \|x\|_{\ell^1} \quad \text{subject to} \quad Ax = b$$

  - Global basis selection rather than greedy local selection
- Example, $A = [A_1 A_2]$ two ONB with coherence $\max_{ij}(a_i, a_j)$
  - If $\|x\|_{\ell^0} \lessapprox .914(1 + \mu^{-1})$ then $\ell^1 \to \ell^0$ [Elad, Bruckstein]
  - Coherence, $\mu := \max_{ij}(a_i, a_j) \geq 1/\sqrt{d}$, [Candes, Romberg]
- Ensures convergence for only the most sparse, $\mathcal{O}(\sqrt{d})$, signals.
- Examples of failure: Dirac's Comb [Candes, ...]
- Is the story over? Can $\mathcal{O}(\sqrt{d})$ threshold be overcome? yes!
- Examples of success: partial Fourier and Laplace, $\|x\|_{\ell^0} \lessapprox \lfloor \frac{d}{2} \rfloor$

Basis Pursuit

- Rather than solve ℓ^0 (combinatorial), solve ℓ^1, (use LP)

$$\min \|x\|_{\ell^1} \quad \text{subject to} \quad Ax = b$$

  - Global basis selection rather than greedy local selection
- Example, $A = [A_1 A_2]$ two ONB with coherence $\max_{ij}(a_i, a_j)$
  - If $\|x\|_{\ell^0} \lessapprox .914(1 + \mu^{-1})$ then $\ell^1 \to \ell^0$ [Elad, Bruckstein]
  - Coherence, $\mu := \max_{ij}(a_i, a_j) \geq 1/\sqrt{d}$, [Candes, Romberg]
- Ensures convergence for only the most sparse, $\mathcal{O}(\sqrt{d})$, signals.
- Examples of failure: Dirac's Comb [Candes, ...]
- Is the story over? Can $\mathcal{O}(\sqrt{d})$ threshold be overcome? yes!
- Examples of success: partial Fourier and Laplace, $\|x\|_{\ell^0} \lessapprox \lfloor \frac{d}{2} \rfloor$
- More to come for typical (random) matrices

Sparsity threshold and the sampling matrix, A

Deterministic:

- Uncorrelated frame expansions require $\|x\|_{\ell^0} < \mathcal{O}(\sqrt{d})$ that is, success only for highly sparse signals:
- Some special cases of success, partial Fourier and Laplace, $\|x\|_{\ell^0} \lessapprox \lfloor \frac{d}{2} \rfloor$ [Donoho, T]

Sparsity threshold and the sampling matrix, A

Deterministic:

- Uncorrelated frame expansions require $\|x\|_{\ell^0} < \mathcal{O}(\sqrt{d})$ that is, success only for highly sparse signals:
- Some special cases of success, partial Fourier and Laplace, $\|x\|_{\ell^0} \lessapprox \lfloor \frac{d}{2} \rfloor$ [Donoho, T]

Random:

- Overcoming the $\sqrt{d}$ threshold for most A with *randomness*

Sparsity threshold and the sampling matrix, A

Deterministic:

- Uncorrelated frame expansions require $\|x\|_{\ell^0} < \mathcal{O}(\sqrt{d})$ that is, success only for highly sparse signals:
- Some special cases of success, partial Fourier and Laplace, $\|x\|_{\ell^0} \lessapprox \lfloor \frac{d}{2} \rfloor$ [Donoho, T]

Random:

- Overcoming the $\sqrt{d}$ threshold for most A with *randomness*
- Recent order bounds for random ortho-projectors
  - $\ell^1 \to \ell^0$ if $\|x\|_{\ell^0} \lessapprox \mathcal{O}(d/\log(n/d))$ [Candes, Tao, Romberg; Vershinin, Rudelson]
  - OMP$\to \ell^0$ if $\|x\|_{\ell^0} \lessapprox \mathcal{O}(d/\log(n))$ [Tropp]

Sparsity threshold and the sampling matrix, A

Deterministic:

- Uncorrelated frame expansions require $\|x\|_{\ell^0} < \mathcal{O}(\sqrt{d})$ that is, success only for highly sparse signals:
- Some special cases of success, partial Fourier and Laplace, $\|x\|_{\ell^0} \lessapprox \lfloor \frac{d}{2} \rfloor$ [Donoho, T]

Random:

- Overcoming the $\sqrt{d}$ threshold for most A with *randomness*
- Recent order bounds for random ortho-projectors
  • $\ell^1 \to \ell^0$ if $\|x\|_{\ell^0} \lessapprox \mathcal{O}(d/\log(n/d))$
  [Candes, Tao, Romberg; Vershinin, Rudelson]
  • OMP$\to \ell^0$ if $\|x\|_{\ell^0} \lessapprox \mathcal{O}(d/\log(n))$ [Tropp]
- What is the precise ℓ^1 sparsity threshold for random matrices?

Sparsity threshold and the sampling matrix, A

Deterministic:

- Uncorrelated frame expansions require $\|x\|_{\ell^0} < \mathcal{O}(\sqrt{d})$ that is, success only for highly sparse signals:
- Some special cases of success, partial Fourier and Laplace, $\|x\|_{\ell^0} \lessapprox \lfloor \frac{d}{2} \rfloor$ [Donoho, T]

Random:

- Overcoming the $\sqrt{d}$ threshold for most A with *randomness*
- Recent order bounds for random ortho-projectors
  • $\ell^1 \to \ell^0$ if $\|x\|_{\ell^0} \lessapprox \mathcal{O}(d/\log(n/d))$ [Candes, Tao, Romberg; Vershinin, Rudelson]
  • OMP$\to \ell^0$ if $\|x\|_{\ell^0} \lessapprox \mathcal{O}(d/\log(n))$ [Tropp]
- What is the precise ℓ^1 sparsity threshold for random matrices?
- Computing random inner products, "correlation with noise"

Sparsity threshold and the sampling matrix, A

Deterministic:

- Uncorrelated frame expansions require $\|x\|_{\ell^0} < \mathcal{O}(\sqrt{d})$ that is, success only for highly sparse signals:
- Some special cases of success, partial Fourier and Laplace, $\|x\|_{\ell^0} \lessapprox \lfloor \frac{d}{2} \rfloor$ [Donoho, T]

Random:

- Overcoming the $\sqrt{d}$ threshold for most A with *randomness*
- Recent order bounds for random ortho-projectors
  • $\ell^1 \rightarrow \ell^0$ if $\|x\|_{\ell^0} \lessapprox \mathcal{O}(d/\log(n/d))$
  [Candes, Tao, Romberg; Vershinin, Rudelson]
  • OMP$\rightarrow \ell^0$ if $\|x\|_{\ell^0} \lessapprox \mathcal{O}(d/\log(n))$ [Tropp]
- What is the precise ℓ^1 sparsity threshold for random matrices?
- Computing random inner products, “correlation with noise”
  Why solve this problem? Are there applications?

Motivation for systems with random A

Compressed Sensing [Donoho; Candes, Tao]:

- Transform Φ with *sparse* signal coefficients, $\hat{x} = \Phi x$.
  Can $\hat{x}$ be recovered with few measurements of x?

Motivation for systems with random A

Compressed Sensing [Donoho; Candes, Tao]:

- Transform Φ with *sparse* signal coefficients, $\hat{x} = \Phi x$.
  Can $\hat{x}$ be recovered with few measurements of x?

- Yes, from nonadaptive measurements recover sparse coef.
  Sample the signal with $A\Phi$ where A is random $d \times n$, $d < n$.

$$\min \|\Phi x\|_{\ell^1} \quad \text{subject to measurements} \quad A\Phi x = b$$

Motivation for systems with random A

Compressed Sensing [Donoho; Candes, Tao]:

- Transform Φ with *sparse* signal coefficients, $\hat{x} = \Phi x$.
  Can $\hat{x}$ be recovered with few measurements of x?

- Yes, from nonadaptive measurements recover sparse coef.
  Sample the signal with $A\Phi$ where A is random $d \times n$, $d < n$.

$$\min \|\Phi x\|_{\ell^1} \quad \text{subject to measurements} \quad A\Phi x = b$$

- Coming to a digital camera near you [Baranuik]

Motivation for systems with random A

Compressed Sensing [Donoho; Candes, Tao]:

- Transform Φ with *sparse* signal coefficients, $\hat{x} = \Phi x$.
  Can $\hat{x}$ be recovered with few measurements of x?
- Yes, from nonadaptive measurements recover sparse coef.
  Sample the signal with $A\Phi$ where A is random $d \times n$, $d < n$.

$$\min \|\Phi x\|_{\ell^1} \quad \text{subject to measurements} \quad A\Phi x = b$$

- Coming to a digital camera near you [Baranuik]

Phase transition as function of measurements (aspect ratio):

- Fix aspect ratio, $\delta = d/n \in (0,1)$, where $A \in \mathbb{R}^{d \times n}$
  Sparsity threshold, $\|x\|_{\ell^0} \leq \rho(\delta)d$, $\rho(\delta) \in (0,1)$

Motivation for systems with random A

Compressed Sensing [Donoho; Candes, Tao]:

- Transform Φ with *sparse* signal coefficients, $\hat{x} = \Phi x$.
  Can $\hat{x}$ be recovered with few measurements of x?

- Yes, from nonadaptive measurements recover sparse coef.
  Sample the signal with $A\Phi$ where A is random $d \times n$, $d < n$.

$$\min \|\Phi x\|_{\ell^1} \quad \text{subject to measurements} \quad A\Phi x = b$$

- Coming to a digital camera near you [Baranuik]

Phase transition as function of measurements (aspect ratio):

- Fix aspect ratio, $\delta = d/n \in (0,1)$, where $A \in \mathbb{R}^{d \times n}$
  Sparsity threshold, $\|x\|_{\ell^0} \leq \rho(\delta)d$, $\rho(\delta) \in (0,1)$

- Phase transition as $n \to \infty$, overwhelming probability $\ell^1 \to \ell^0$

Neighborliness and constrained ℓ^1 minimization

Theorem

Let A be a $d \times n$ matrix, $d < n$. The two properties of A are equiv.:

- *The polytope AT has n vertices and is outwardly k-neighborly,*
- *Whenever $y = Ax$ has a nonnegative solution x_0 having at most k nonzeros, x_0 is the unique nonnegative solution to $y = Ax$ and so the unique solution to the constrained ℓ^1 minimization problem.*

Neighborliness and constrained ℓ^1 minimization

Theorem

Let A be a $d \times n$ matrix, $d < n$. The two properties of A are equiv.:

- *The polytope AT has n vertices and is outwardly k-neighborly,*
- *Whenever $y = Ax$ has a nonnegative solution x_0 having at most k nonzeros, x_0 is the unique nonnegative solution to $y = Ax$ and so the unique solution to the constrained ℓ^1 minimization problem.*

Lemma (Neighborliness and face numbers)

Suppose the polytope $P = AT$ has n vertices and is outwardly k-neighborly. Then

$$\forall \ell = 0, \ldots, k-1, \quad \forall\, F \in \mathcal{F}_\ell(T^{n-1}), \quad AF \in \mathcal{F}_\ell(AT).$$

Conversely, suppose that the above equation holds; then $P = AT$ has n vertices and is outwardly k-neighborly.

Strong threshold, random A and all x_0

Expected number of faces, random ortho-projector:

$$\mathcal{E} f_k(AT) = f_k(T) - 2 \sum_{s \geq 0} \sum_{F \in \mathcal{F}_k(T)} \sum_{G \in \mathcal{F}_{d+1+2s}(T)} \beta(F,G)\gamma(G,T)$$

where β and γ are internal and external angles respectively [Affentranger, Schneider]

Strong threshold, random A and all x_0

Expected number of faces, random ortho-projector:

$$\mathcal{E} f_k(AT) = f_k(T) - 2 \sum_{s \geq 0} \sum_{F \in \mathcal{F}_k(T)} \sum_{G \in \mathcal{F}_{d+1+2s(T)}} \beta(F,G)\gamma(G,T)$$

where β and γ are internal and external angles respectively [Affentranger, Schneider]

Theorem (Strong threshold)

Let $\rho < \rho_N(\delta)$ and let $A = A_{d,n}$ be a uniformly-distributed random projection from $\mathbb{R}^n$ to $\mathbb{R}^d$, with $d \geq \delta n$. Then

$$Prob\{f_\ell(AT^{n-1}) = f_\ell(T^{n-1}), \quad \ell = 0, \ldots, \lfloor \rho d \rfloor\} \to 1, \quad \text{as } n \to \infty.$$

$\implies$ P is k neighborly for $k = \lfloor (\rho_N(\delta) - \epsilon) d \rfloor$

Strong threshold, random A and all x_0

Expected number of faces, random ortho-projector:

$$\mathcal{E} f_k(AT) = f_k(T) - 2 \sum_{s \geq 0} \sum_{F \in \mathcal{F}_k(T)} \sum_{G \in \mathcal{F}_{d+1+2s}(T)} \beta(F,G)\gamma(G,T)$$

where β and γ are internal and external angles respectively [Affentranger, Schneider]

Theorem (Strong threshold)

Let $\rho < \rho_N(\delta)$ and let $A = A_{d,n}$ be a uniformly-distributed random projection from $\mathbb{R}^n$ to $\mathbb{R}^d$, with $d \geq \delta n$. Then

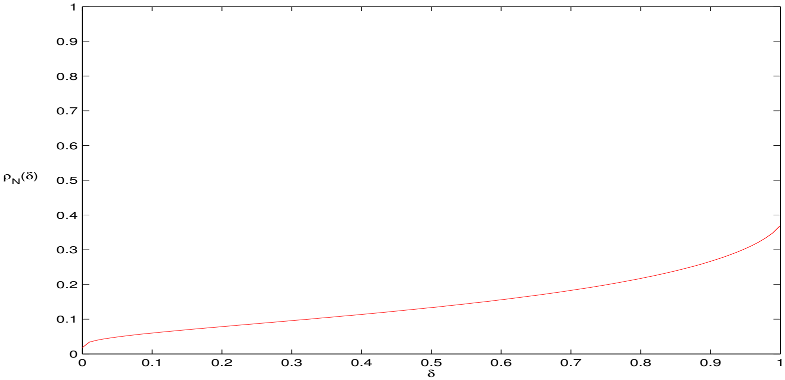
$$Prob\{f_\ell(AT^{n-1}) = f_\ell(T^{n-1}), \quad \ell = 0, \ldots, \lfloor \rho d \rfloor\} \to 1, \quad as\ n \to \infty.$$

$\Longrightarrow$ P is k neighborly for $k = \lfloor (\rho_N(\delta) - \epsilon) d \rfloor$

$\Rightarrow$ With overwhelming probability $(f_\ell(T^{n-1}) - \mathcal{E} f_\ell(AT^{n-1}) \leq \pi_n e^{-\epsilon n})$ on A, for every x_0 with $\|x_0\|_{\ell^0} \leq \lfloor (\rho_N(\delta) - \epsilon) d \rfloor$, $y = Ax_0$ generates an instance of the constrained ℓ^1 minimization problem with x_0 as its unique solution.

Phase Transition, Strong (non-negative)

$$\ell^1 \rightarrow \ell^0 \quad \text{if} \quad \|x_0\|_{\ell^0} \leq \lfloor (\rho_N(\delta) - \epsilon) d \rfloor \quad \text{and} \quad x \geq 0$$



- As $\delta \uparrow 1$ $\rho_N(\delta) \approx .371$
- As $\delta \rightarrow 0$ $\rho_N(\delta) \sim [2e \log(1/\delta)]^{-1}$

Weak threshold, random A and most x_0

Theorem (Vershik-Sporyshev)

Let $d = d(n) \sim \delta n$ and let $A = A_{d,n}$ be a uniform random projection from $\mathbb{R}^n$ to $\mathbb{R}^d$. Then for a sequence $k = k(n)$ with $k/d \sim \rho$, $\rho < \rho_{VS}(\delta)$, we have

$$f_k(AT^{n-1}) = f_k(T^{n-1})(1 + o_P(1)).$$

Weak threshold, random A and most x_0

Theorem (Vershik-Sporyshev)

Let $d = d(n) \sim \delta n$ and let $A = A_{d,n}$ be a uniform random projection from $\mathbb{R}^n$ to $\mathbb{R}^d$. Then for a sequence $k = k(n)$ with $k/d \sim \rho$, $\rho < \rho_{VS}(\delta)$, we have

$$f_k(AT^{n-1}) = f_k(T^{n-1})(1 + o_P(1)).$$

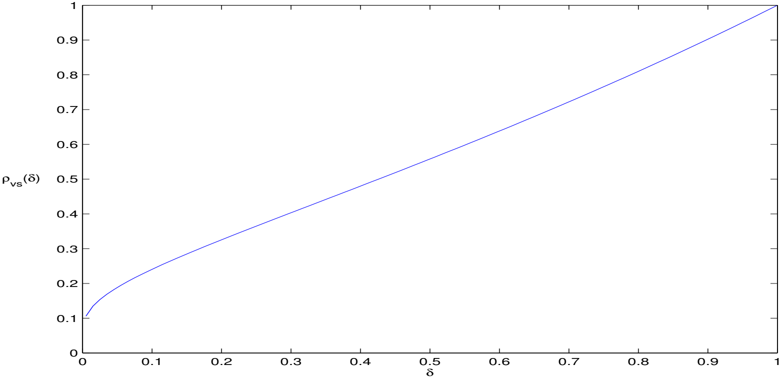
Theorem

Let A be a $d \times n$ matrix, $d < n$ in general position. For $1 \leq k \leq d-1$, these two properties of A are equivalent

- *The polytope $P = AT$ has at least $(1-\epsilon)$ times as many zero-free $(k-1)$-faces as T,*
- *Among all problem instances (y, A) generated by some nonnegative vector x_0 with at most k nonzeros, the constrained ℓ^1 minimization recovers the sparsest solution, except in a fraction $\leq \epsilon$ of instances.*

Phase Transition, Weak (non-negative)

$$\ell^1 \rightarrow \ell^0 \quad \text{if} \quad \|x\|_{\ell^0} \leq \lfloor (\rho_{VS}(\delta) - \epsilon) d \rfloor \quad \text{and} \quad x \geq 0$$

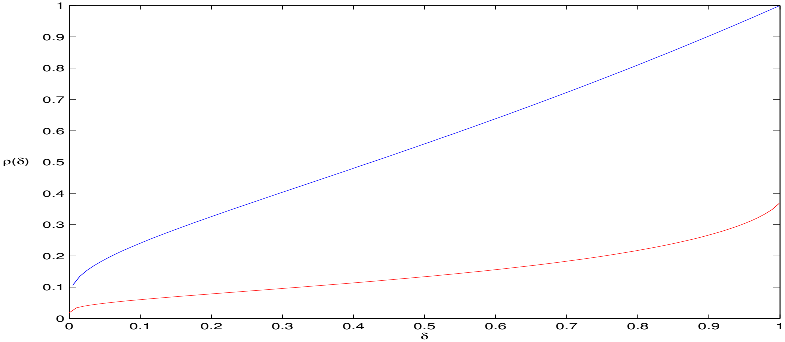


- Asymptotic limit of empirical tests (example shown later)
- As $\delta \rightarrow 0$ $\rho_{VS}(\delta) \sim [2\log(1/\delta)]^{-1}$
- Typically e times less strict sparsity requirement as $\delta \rightarrow 0$

Phase Transitions, $\ell^1 \to \ell^0$ *if* $\|x\|_{\ell^0} < \rho(\delta)d$

Two modalities from the random sampling perspective:

- Weak threshold, random signal and measurement independent
- Strong threshold, worst signal for a given measurement



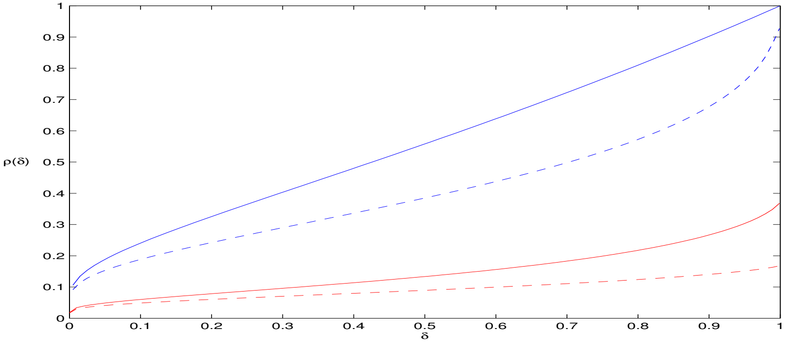
Weak (typical) and Strong (malicious) transition

Non-negative signal, x, [Donoho,T]

Phase Transitions, $\ell^1 \rightarrow \ell^0$ *if* $\|x\|_{\ell^0} < \rho(\delta)d$

Two modalities from the random sampling perspective:

- Weak threshold, random signal and measurement independent
- Strong threshold, worst signal for a given measurement



Weak (typical) and Strong (malicious) transition

Solid for non-negative, dashed for signed signal [Donoho]

Some precise numbers and implications

	$\delta = .1$	$\delta = .25$	$\delta = .5$	$\delta = .75$	$\delta = .9$
ρ_N^+	.060131	.087206	.133457	.198965	.266558
ρ_W^+	.240841	.364970	.558121	.765796	.902596
$\rho_N^\pm$	.048802	.065440	.089416	.117096	.140416
$\rho_W^\pm$	.188327	.266437	.384803	.532781	.677258

- For most A measure $1/10$ of a non-negative signal, recovery every signal if 6% sparse, most if 24% sparse.

Some precise numbers and implications

	$\delta = .1$	$\delta = .25$	$\delta = .5$	$\delta = .75$	$\delta = .9$
ρ_N^+	.060131	.087206	.133457	.198965	.266558
ρ_W^+	.240841	.364970	.558121	.765796	.902596
$\rho_N^\pm$	.048802	.065440	.089416	.117096	.140416
$\rho_W^\pm$	.188327	.266437	.384803	.532781	.677258

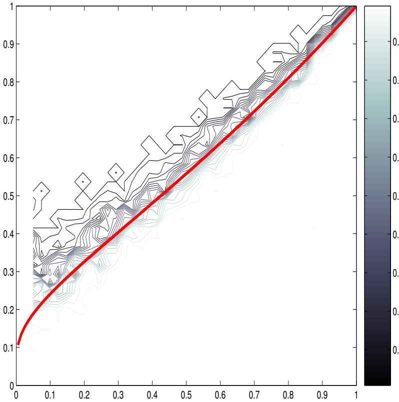
- For most A measure $1/10$ of a non-negative signal, recovery every signal if 6% sparse, most if 24% sparse.
- Half 'under-sampling', i.e., $\delta = 1/2$: apply ℓ^1, if non-negative solution is $<$ 55% sparse then typically is *sparsest* solution.

Some precise numbers and implications

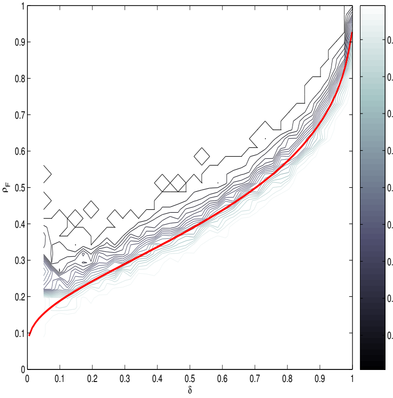
	$\delta = .1$	$\delta = .25$	$\delta = .5$	$\delta = .75$	$\delta = .9$
ρ_N^+	.060131	.087206	.133457	.198965	.266558
ρ_W^+	.240841	.364970	.558121	.765796	.902596
$\rho_N^\pm$	.048802	.065440	.089416	.117096	.140416
$\rho_W^\pm$	.188327	.266437	.384803	.532781	.677258

- For most A measure $1/10$ of a non-negative signal, recovery every signal if 6% sparse, most if 24% sparse.
- Half 'under-sampling', i.e., $\delta = 1/2$: apply ℓ^1, if non-negative solution is $< 55\%$ sparse then typically is *sparsest* solution.
- Encode $(1-\delta)n$ bits of info in signal of length n. Can recover with less than $\delta\rho_W^\pm(\delta)n$ accidental, $\delta\rho_N^\pm(\delta)n$ malicious errors.
  - twice redun., tolerate 19% random error, 4.4% mallitous.

Empirical verification of weak transitions



Non-negative



Signed signal

$n = 200$, 40×40 mesh with 60 random tests per node.

Ingredients of the proofs, non-negative

Proof (main ideas):

- Given $x_0 \geq 0$ with $\|x_0\|_{\ell^1} = 1$ and $\|x_0\|_{\ell^0} = k$

Ingredients of the proofs, non-negative

Proof (main ideas):

- Given $x_0 \geq 0$ with $\|x_0\|_{\ell^1} = 1$ and $\|x_0\|_{\ell^0} = k$
- x_0 is on a $k-1$ face (say F) of unit simplex, T^{n-1}

Ingredients of the proofs, non-negative

Proof (main ideas):

- Given $x_0 \geq 0$ with $\|x_0\|_{\ell^1} = 1$ and $\|x_0\|_{\ell^0} = k$
- x_0 is on a $k-1$ face (say F) of unit simplex, T^{n-1}
- $b = Ax_0$ either on boundary ($AF \in \mathcal{F}_k(AT)$), or inside AT

Ingredients of the proofs, non-negative

Proof (main ideas):

- Given $x_0 \geq 0$ with $\|x_0\|_{\ell^1} = 1$ and $\|x_0\|_{\ell^0} = k$
- x_0 is on a $k-1$ face (say F) of unit simplex, T^{n-1}
- $b = Ax_0$ either on boundary ($AF \in \mathcal{F}_k(AT)$), or inside AT
- if on boundary then unique for $\|x\|_{\ell^1} \leq 1$ and $Ax = b$, $\ell^1 \to \ell^0$

Ingredients of the proofs, non-negative

Proof (main ideas):

- Given $x_0 \geq 0$ with $\|x_0\|_{\ell^1} = 1$ and $\|x_0\|_{\ell^0} = k$
- x_0 is on a $k-1$ face (say F) of unit simplex, T^{n-1}
- $b = Ax_0$ either on boundary ($AF \in \mathcal{F}_k(AT)$), or inside AT
- if on boundary then unique for $\|x\|_{\ell^1} \leq 1$ and $Ax = b$, $\ell^1 \to \ell^0$
- if b in the interior of $P = AT$ then $\exists$ $\|x\|_{\ell^1} < 1$ with $Ax = b$.

Ingredients of the proofs, non-negative

Proof (main ideas):

- Given $x_0 \geq 0$ with $\|x_0\|_{\ell^1} = 1$ and $\|x_0\|_{\ell^0} = k$
- x_0 is on a $k-1$ face (say F) of unit simplex, T^{n-1}
- $b = Ax_0$ either on boundary ($AF \in \mathcal{F}_k(AT)$), or inside AT
- if on boundary then unique for $\|x\|_{\ell^1} \leq 1$ and $Ax = b$, $\ell^1 \to \ell^0$
- if b in the interior of $P = AT$ then $\exists$ $\|x\|_{\ell^1} < 1$ with $Ax = b$.
- If $j \leq k$ faces of T^{n-1} remain in P then $\ell^1 \to \ell^0$ for $\|x\|_{\ell^0} \leq k$

Ingredients of the proofs, non-negative

Proof (main ideas):

- Given $x_0 \geq 0$ with $\|x_0\|_{\ell^1} = 1$ and $\|x_0\|_{\ell^0} = k$
- x_0 is on a $k-1$ face (say F) of unit simplex, T^{n-1}
- $b = Ax_0$ either on boundary ($AF \in \mathcal{F}_k(AT)$), or inside AT
- if on boundary then unique for $\|x\|_{\ell^1} \leq 1$ and $Ax = b$, $\ell^1 \to \ell^0$
- if b in the interior of $P = AT$ then $\exists$ $\|x\|_{\ell^1} < 1$ with $Ax = b$.
- If $j \leq k$ faces of T^{n-1} remain in P then $\ell^1 \to \ell^0$ for $\|x\|_{\ell^0} \leq k$

Ingredients of the proofs, non-negative

Proof (main ideas):

- Given $x_0 \geq 0$ with $\|x_0\|_{\ell^1} = 1$ and $\|x_0\|_{\ell^0} = k$
- x_0 is on a $k-1$ face (say F) of unit simplex, T^{n-1}
- $b = Ax_0$ either on boundary ($AF \in \mathcal{F}_k(AT)$), or inside AT
- if on boundary then unique for $\|x\|_{\ell^1} \leq 1$ and $Ax = b$, $\ell^1 \to \ell^0$
- if b in the interior of $P = AT$ then $\exists \ \|x\|_{\ell^1} < 1$ with $Ax = b$.
- If $j \leq k$ faces of T^{n-1} remain in P then $\ell^1 \to \ell^0$ for $\|x\|_{\ell^0} \leq k$
  $\rho_N := \text{Prob}(f_\ell(AT^{n-1}) = f_\ell(T^{n-1}) : \ell = 0, \ldots, \lfloor \rho_N d \rfloor) \to 1$
  That is, $f_\ell(T^{n-1}) - \mathcal{E} f_\ell(AT^{n-1}) \leq \pi_n e^{-\epsilon n}$

Ingredients of the proofs, non-negative

Proof (main ideas):

- Given $x_0 \geq 0$ with $\|x_0\|_{\ell^1} = 1$ and $\|x_0\|_{\ell^0} = k$
- x_0 is on a $k-1$ face (say F) of unit simplex, T^{n-1}
- $b = Ax_0$ either on boundary ($AF \in \mathcal{F}_k(AT)$), or inside AT
- if on boundary then unique for $\|x\|_{\ell^1} \leq 1$ and $Ax = b$, $\ell^1 \to \ell^0$
- if b in the interior of $P = AT$ then $\exists$ $\|x\|_{\ell^1} < 1$ with $Ax = b$.
- If $j \leq k$ faces of T^{n-1} remain in P then $\ell^1 \to \ell^0$ for $\|x\|_{\ell^0} \leq k$
  $\rho_N := \text{Prob}(f_\ell(AT^{n-1}) = f_\ell(T^{n-1}) : \ell = 0, \ldots, \lfloor \rho_N d \rfloor) \to 1$
  That is, $f_\ell(T^{n-1}) - \mathcal{E} f_\ell(AT^{n-1}) \leq \pi_n e^{-\epsilon n}$
  $\rho_W := \mathcal{E} f_\ell(AT^{n-1}) \geq (1-\epsilon) f_\ell(T^{n-1})$, $\ell = 0, \ldots, \lfloor \rho_W d \rfloor$

Ingredients of the proofs, non-negative

Proof (main ideas):

- Given $x_0 \geq 0$ with $\|x_0\|_{\ell^1} = 1$ and $\|x_0\|_{\ell^0} = k$
- x_0 is on a $k-1$ face (say F) of unit simplex, T^{n-1}
- $b = Ax_0$ either on boundary ($AF \in \mathcal{F}_k(AT)$), or inside AT
- if on boundary then unique for $\|x\|_{\ell^1} \leq 1$ and $Ax = b$, $\ell^1 \to \ell^0$
- if b in the interior of $P = AT$ then $\exists\ \|x\|_{\ell^1} < 1$ with $Ax = b$.
- If $j \leq k$ faces of T^{n-1} remain in P then $\ell^1 \to \ell^0$ for $\|x\|_{\ell^0} \leq k$
  $\rho_N := \text{Prob}(f_\ell(AT^{n-1}) = f_\ell(T^{n-1}) : \ell = 0, \ldots, \lfloor \rho_N d \rfloor) \to 1$
  That is, $f_\ell(T^{n-1}) - \mathcal{E} f_\ell(AT^{n-1}) \leq \pi_n e^{-\epsilon n}$
  $\rho_W := \mathcal{E} f_\ell(AT^{n-1}) \geq (1-\epsilon) f_\ell(T^{n-1})$, $\ell = 0, \ldots, \lfloor \rho_W d \rfloor$

Robustness:

Nearby sparse solution, $\|Ax_0 - b\|_2 \leq \epsilon$, then solve

$$\min \|x_{1,\epsilon}\|_{\ell^1} \quad \text{such that} \quad \|Ax_{1,\epsilon} - b\|_2 \leq \epsilon$$

Then $\|x_0 - x_{1,\epsilon}\|_2 \leq C(k, A)\epsilon$ where $k = \|x_0\|_{\ell^0}$.

Summary

- Underdetermined system, $Ax = b$ with $A \in \mathbb{R}^{d \times n}$ where $d < n$
- To obtain sparsest solution, $\|x\|_{\ell^0}$, solve constrained min $\|x\|_{\ell^1}$
- Precise sparsity phase transitions, $\rho(d/n)$ available for $\ell^1 \to \ell^0$
- That is, if $\|x\|_{\ell^0} < \rho(d/n) \cdot d$ then min $\|x\|_{\ell^1} \to$ min $\|x\|_{\ell^0}$
- Surprisingly large transition, effectiveness of Basis Pursuit (ℓ^1)

Summary

- Underdetermined system, $Ax = b$ with $A \in \mathbb{R}^{d \times n}$ where $d < n$
- To obtain sparsest solution, $\|x\|_{\ell^0}$, solve constrained min $\|x\|_{\ell^1}$
- Precise sparsity phase transitions, $\rho(d/n)$ available for $\ell^1 \to \ell^0$
- That is, if $\|x\|_{\ell^0} < \rho(d/n) \cdot d$ then min $\|x\|_{\ell^1} \to$ min $\|x\|_{\ell^0}$
- Surprisingly large transition, effectiveness of Basis Pursuit (ℓ^1)

Associated Papers for non-negative case [Donoho, T]:

- Sparse Nonnegative Solution of Underdetermined Linear Equations by Linear Programming, Proc. Nat. Acc. Sci.
- Neighborliness of Randomly-Projected Simplices in High Dimensions, Proc. Nat. Acc. Sci.

• see also work by Donoho; Candes, Romberg, Tao; Tropp

Thank you for your time