Algorithms for sparse analysis
Lecture II: Hardness results for sparse approximation problems

Anna C. Gilbert

Department of Mathematics
University of Michigan

Complexity theory: Reductions

- Problem A (efficiently) reduces to B means a(n efficient) solution to B can be used to solve A (efficiently)
- If we have an algorithm to solve B, then we can use that algorithm to solve A; i.e., A is easier to solve than B
- "reduces" does *not* confer simplification here

- Definition
  $\mathrm{A} \leq_P \mathrm{B}$ *if there's polynomial time computable function f s.t.*

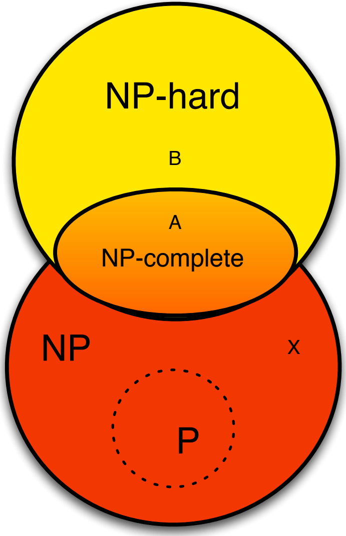
$$w \in A \iff f(w) \in B.$$

- B at least as hard as A

Complexity theory: **NP-hard**

- Definition
  $A \in$ **NP-complete** *if (i)* $A \in$ **NP** *and (ii) for all* $X \in$ **NP**, $X \leq_P A$.

- Definition
  $B \in$ **NP-hard** *if there is* $A \in$ **NP-complete** *s.t.* $A \leq_P B$.



Examples

- RelPrime Are a and b relatively prime?
  - in **P**
  - Euclidean algorithm, simple
- Primes Is x a prime number?
  - in **P**
  - highly non-trivial algorithm, does not determine factors
- Factor Factor x as a product of powers of primes.
  - in **NP**
  - not known to be **NP-hard**
- X3C Given a finite universe $\mathcal{U}$, a collection $\mathcal{X}$ of subsets $X_1, X_2, \ldots, X_N$ s.t. $|X_i| = 3$ for each i, does $\mathcal{X}$ contain a disjoint collection of subsets whose union $= \mathcal{U}$?
  - **NP-complete**

NP-hardness

Theorem

Given an arbitrary redundant dictionary Φ, a signal x, and a sparsity parameter k, it is NP-hard to solve the sparse representation problem $\textsc{D-Exact}$. *[Natarajan'95,Davis'97]*

NP-hardness

Theorem

Given an arbitrary redundant dictionary Φ, a signal x, and a sparsity parameter k, it is NP-hard to solve the sparse representation problem D-EXACT. *[Natarajan'95,Davis'97]*

Corollary

SPARSE, ERROR, EXACT *are all NP-hard.*

NP-hardness

Theorem

Given an arbitrary redundant dictionary Φ, a signal x, and a sparsity parameter k, it is NP-hard to solve the sparse representation problem D-EXACT. *[Natarajan'95,Davis'97]*

Corollary

SPARSE, ERROR, EXACT *are all NP-hard.*

Corollary

Given an arbitrary redundant dictionary Φ and a signal x, it is NP-hard to approximate (in error) the solution of EXACT *to within* any *factor. [Davis'97]*

Exact Cover by 3-sets: $\mathrm{X3C}$

Definition

Given a finite universe $\mathcal{U}$, a collection $\mathcal{X}$ of subsets $X_1, X_2, \ldots, X_N$ s.t. $|X_i| = 3$ for each i, does $\mathcal{X}$ contain a disjoint collection of subsets whose union $= \mathcal{U}$?

Exact Cover by 3-sets: $\mathrm{X3C}$

Definition

Given a finite universe $\mathcal{U}$, a collection $\mathcal{X}$ of subsets $X_1, X_2, \ldots, X_N$ s.t. $|X_i| = 3$ for each i, does $\mathcal{X}$ contain a disjoint collection of subsets whose union $= \mathcal{U}$?

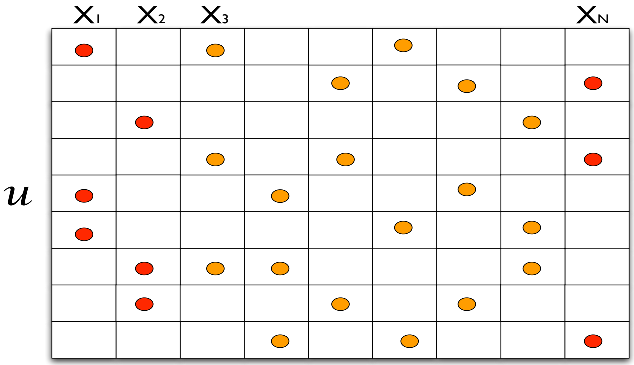
NP-complete problem.

Exact Cover by 3-sets: $\mathrm{X3C}$

Definition

Given a finite universe $\mathcal{U}$, a collection $\mathcal{X}$ of subsets $X_1, X_2, \ldots, X_N$ s.t. $|X_i| = 3$ for each i, does $\mathcal{X}$ contain a disjoint collection of subsets whose union $= \mathcal{U}$?

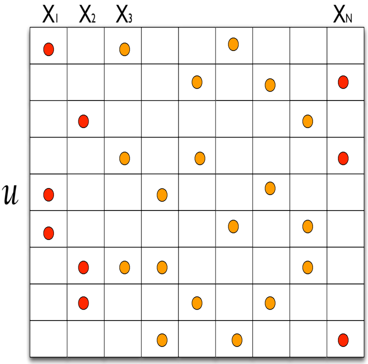
NP-complete problem.



Proposition

Any instance of X3C *is reducible in polynomial time to* D-EXACT.

X3C $\leq_P$ D-EXACT



Proof.

- Let $\Omega = \{1, 2, \ldots, N\}$ index Φ. Set $\varphi_i = \mathbf{1}_{X_i}$.

  Select $x = (1, 1, \ldots, 1)$, $k = \frac{1}{3}|\mathcal{U}|$.

- Suppose have solution to X3C. Sufficient to check if SPARSE solution has zero error.

  Assume solutions of X3C indexed by Λ. Set $c_{\mathrm{opt}} = \mathbf{1}_\Lambda$.

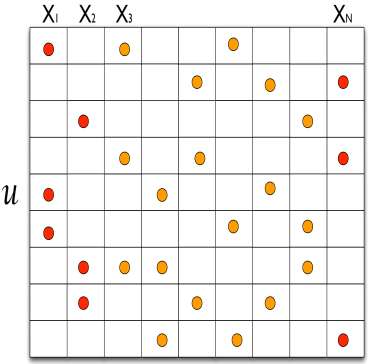
  $$\Phi c_{\mathrm{opt}} = x.$$

  $\Longrightarrow$ SPARSE solution has zero error and D-Exact returns YES.

Proposition

Any instance of $\mathrm{X3C}$ *is reducible in polynomial time to* D-EXACT.

$\mathrm{X3C} \leq_P$ D-EXACT



Proof.

- Let $\Omega = \{1, 2, \ldots, N\}$ index Φ. Set $\varphi_i = \mathbf{1}_{X_i}$.

  Select $x = (1, 1, \ldots, 1)$, $k = \frac{1}{3}|\mathcal{U}|$.

- Suppose have solution to X3C. Sufficient to check if SPARSE solution has zero error.

  Assume solutions of X3C indexed by Λ. Set $c_{\mathrm{opt}} = \mathbf{1}_\Lambda$.

  $$\Phi c_{\mathrm{opt}} = x.$$

  $\implies$ SPARSE solution has zero error and D-Exact returns YES.

- Suppose c_{opt} is optimal solution of SPARSE

  $$\Phi c_{\mathrm{opt}} = x$$

  then c_{opt} contains $k \leq \frac{1}{3}|\mathcal{U}|$ nonzero entries and D-Exact returns YES.

  Each column of Φ has 3 nonzero entries

  $\implies$ $\{X_i \mid i \in \mathrm{supp}(c_{\mathrm{opt}})\}$ is disjoint collection covering $\mathcal{U}$.

□

What does this mean?

Bad news

- Given any polynomial time algorithm for SPARSE, there is a dictionary Φ and a signal x such that algorithm returns incorrect answer
- Pessimistic: worst case
- Cannot hope to approximate solution, either

What does this mean?

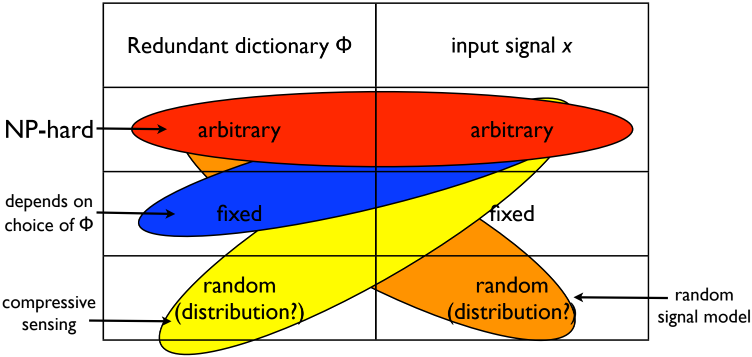
Bad news

- Given any polynomial time algorithm for SPARSE, there is a dictionary Φ and a signal x such that algorithm returns incorrect answer
- Pessimistic: worst case
- Cannot hope to approximate solution, either

Good news

- Natural dictionaries are far from arbitrary
- Perhaps natural dictionaries admit polynomial time algorithms
- Optimistic: rarely see worst case
- Hardness depends on instance type

Hardness depends on instance



Leverage intuition from orthonormal basis

- Suppose Φ is orthogonal, $\Phi^{-1} = \Phi^T$
- Solution to EXACT problem is unique

$$c = \Phi^{-1}x = \Phi^T x \quad \text{i.e.,} \quad c_\ell = \langle x,\ \varphi_\ell \rangle$$

hence, $x = \sum_\ell \langle x,\ \varphi_\ell \rangle \varphi_\ell$.

Leverage intuition from orthonormal basis

Solution to SPARSE problem similar

- Let $\ell_1 \longleftarrow \arg\max_\ell |\langle x, \varphi_\ell \rangle|$

  Set $c_{\ell_1} \longleftarrow \langle x, \varphi_{\ell_1} \rangle$

  Residual $r \longleftarrow x - c_{\ell_1}\varphi_{\ell_1}$

Leverage intuition from orthonormal basis

Solution to SPARSE problem similar

- Let $\ell_1 \longleftarrow \arg\max_\ell |\langle x, \varphi_\ell \rangle|$
  Set $c_{\ell_1} \longleftarrow \langle x, \varphi_{\ell_1} \rangle$
  Residual $r \longleftarrow x - c_{\ell_1}\varphi_{\ell_1}$
- Let $\ell_2 \longleftarrow \arg\max_\ell |\langle r, \varphi_\ell \rangle| =$
  $\arg\max_\ell |\langle x - c_{\ell_1}\varphi_{\ell_1}, \varphi_\ell \rangle| = \arg\max_{\ell \neq \ell_1} |\langle x, \varphi_\ell \rangle|$
  Set $c_{\ell_2} \longleftarrow \langle r, \varphi_{\ell_2} \rangle$.
  Update residual $r \longleftarrow x - (c_{\ell_1}\varphi_{\ell_1} + c_{\ell_2}\varphi_{\ell_2})$

Leverage intuition from orthonormal basis

Solution to SPARSE problem similar

- Let $\ell_1 \longleftarrow \arg\max_\ell |\langle x,\ \varphi_\ell\rangle|$
  Set $c_{\ell_1} \longleftarrow \langle x,\ \varphi_{\ell_1}\rangle$
  Residual $r \longleftarrow x - c_{\ell_1}\varphi_{\ell_1}$
- Let $\ell_2 \longleftarrow \arg\max_\ell |\langle r,\ \varphi_\ell\rangle| =$
  $\arg\max_\ell |\langle x - c_{\ell_1}\varphi_{\ell_1},\ \varphi_\ell\rangle| = \arg\max_{\ell\neq\ell_1} |\langle x,\ \varphi_\ell\rangle|$
  Set $c_{\ell_2} \longleftarrow \langle r,\ \varphi_{\ell_2}\rangle$.
  Update residual $r \longleftarrow x - (c_{\ell_1}\varphi_{\ell_1} + c_{\ell_2}\varphi_{\ell_2})$
- Repeat $k-2$ times.

Leverage intuition from orthonormal basis

Solution to SPARSE problem similar

- Let $\ell_1 \longleftarrow \arg\max_\ell |\langle x,\ \varphi_\ell\rangle|$
  Set $c_{\ell_1} \longleftarrow \langle x,\ \varphi_{\ell_1}\rangle$
  Residual $r \longleftarrow x - c_{\ell_1}\varphi_{\ell_1}$
- Let $\ell_2 \longleftarrow \arg\max_\ell |\langle r,\ \varphi_\ell\rangle| =$
  $\arg\max_\ell |\langle x - c_{\ell_1}\varphi_{\ell_1},\ \varphi_\ell\rangle| = \arg\max_{\ell\neq\ell_1} |\langle x,\ \varphi_\ell\rangle|$
  Set $c_{\ell_2} \longleftarrow \langle r,\ \varphi_{\ell_2}\rangle$.
  Update residual $r \longleftarrow x - (c_{\ell_1}\varphi_{\ell_1} + c_{\ell_2}\varphi_{\ell_2})$
- Repeat $k - 2$ times.
- Set $c_\ell \longleftarrow 0$ for $\ell \neq \ell_1, \ell_2, \ldots, \ell_k$.

Leverage intuition from orthonormal basis

Solution to SPARSE problem similar

- Let $\ell_1 \longleftarrow \arg\max_\ell |\langle x,\ \varphi_\ell\rangle|$
  Set $c_{\ell_1} \longleftarrow \langle x,\ \varphi_{\ell_1}\rangle$
  Residual $r \longleftarrow x - c_{\ell_1}\varphi_{\ell_1}$
- Let $\ell_2 \longleftarrow \arg\max_\ell |\langle r,\ \varphi_\ell\rangle| =$
  $\arg\max_\ell |\langle x - c_{\ell_1}\varphi_{\ell_1},\ \varphi_\ell\rangle| = \arg\max_{\ell\neq\ell_1} |\langle x,\ \varphi_\ell\rangle|$
  Set $c_{\ell_2} \longleftarrow \langle r,\ \varphi_{\ell_2}\rangle$.
  Update residual $r \longleftarrow x - (c_{\ell_1}\varphi_{\ell_1} + c_{\ell_2}\varphi_{\ell_2})$
- Repeat $k - 2$ times.
- Set $c_\ell \longleftarrow 0$ for $\ell \neq \ell_1, \ell_2, \ldots, \ell_k$.
- Approximate $x \approx \Phi c = \sum_{t=1}^{k} \langle x,\ \varphi_{\ell_t}\rangle\, \varphi_{\ell_t}$.

Check: algorithm generates list of coeffs of x over basis in *descending* order (by absolute value).

Geometry

- Why is orthogonal case easy?

  inner products between atoms are small

  it's easy to tell which one is the best choice

$$\langle r,\ \varphi_j\rangle = \langle x - c_i\varphi_i,\ \varphi_j\rangle = \langle x,\ \varphi_j\rangle - c_i\langle\varphi_i,\ \varphi_j\rangle$$

- When atoms are (nearly) parallel, can't tell which one is best

Coherence

Definition

The coherence of a dictionary

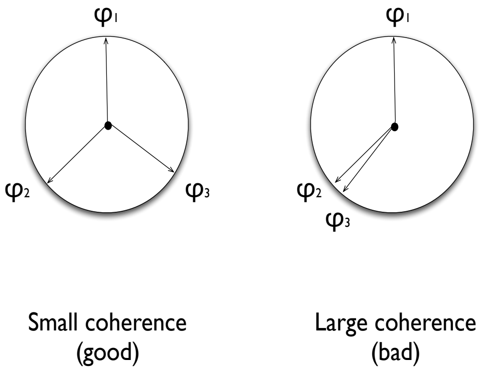
$$\mu = \max_{j \neq \ell} |\langle \varphi_j,\ \varphi_\ell \rangle|$$

Coherence

Definition

The coherence of a dictionary

$$\mu = \max_{j \neq \ell} |\langle \varphi_j, \ \varphi_\ell \rangle|$$



Coherence: lower bound

Theorem

For a $d \times N$ dictionary,

$$\mu \geq \sqrt{\frac{N-d}{d(N-1)}} \approx \frac{1}{\sqrt{d}}.$$

[Welch'73]

Theorem

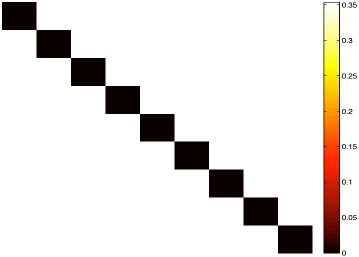
For most pairs of orthonormal bases in $\mathbb{R}^d$, the coherence between the two is

$$\mu = O\Big(\sqrt{\frac{\log d}{d}}\Big).$$

[Donoho, Huo '99]

Large, incoherent dictionaries

- Fourier–Dirac, $N = 2d$, $\mu = \frac{1}{\sqrt{d}}$
- wavelet packets, $N = d \log d$, $\mu = \frac{1}{\sqrt{2}}$
- There are large dictionaries with coherence close to the lower (Welch) bound; e.g., Kerdock codes, $N = d^2$, $\mu = 1/\sqrt{d}$



Approximation algorithms (error)

- SPARSE. Given $k \geq 1$, solve

$$\arg\min_c \|x - \Phi c\|_2 \quad \text{s.t.} \quad \|c\|_0 \leq k$$

i.e., find the best approximation of x using k atoms.

- c_{opt} = optimal solution
- $E_{\text{opt}} = \|\Phi c_{\text{opt}} - x\|_2$ = optimal error

Approximation algorithms (error)

- SPARSE. Given $k \geq 1$, solve

$$\arg\min_{c} \|x - \Phi c\|_2 \quad \text{s.t.} \quad \|c\|_0 \leq k$$

i.e., find the best approximation of x using k atoms.

- c_{opt} = optimal solution
- $E_{\mathrm{opt}} = \|\Phi c_{\mathrm{opt}} - x\|_2$ = optimal error

- Algorithm returns $\widehat{c}$ with
  (1) $\|\widehat{c}\|_0 = k$
  (2) $E = \|\Phi \widehat{c} - x\|_2 \leq C_1 E_{\mathrm{opt}}$

Approximation algorithms (error)

- Sparse. Given $k \geq 1$, solve

$$\arg\min_{c} \|x - \Phi c\|_2 \quad \text{s.t.} \quad \|c\|_0 \leq k$$

  i.e., find the best approximation of x using k atoms.

- c_{opt} = optimal solution
- $E_{\mathrm{opt}} = \|\Phi c_{\mathrm{opt}} - x\|_2$ = optimal error

- Algorithm returns $\widehat{c}$ with
  (1) $\|\widehat{c}\|_0 = k$
  (2) $E = \|\Phi \widehat{c} - x\|_2 \leq C_1 E_{\mathrm{opt}}$

- (Error) approximation ratio: $\frac{E}{E_{\mathrm{opt}}} = \frac{C_1 E_{\mathrm{opt}}}{E_{\mathrm{opt}}} = C_1$

Approximation algorithms (terms)

- Algorithm returns $\widehat{c}$ with
  (1) $\|\widehat{c}\|_0 = C_2 k$
  (2) $E = \|\Phi\widehat{c} - x\|_2 = E_{\mathrm{opt}}$

- (Terms) approximation ratio: $\frac{\|\widehat{c}\|_0}{\|c_{\mathrm{opt}}\|_0} = \frac{C_2 k}{k} = C_2$

Bi-criteria approximation algorithms

- Algorithm returns $\widehat{c}$ with
  (1) $\|\widehat{c}\|_0 = C_2 k$
  (2) $E = \|\Phi\widehat{c} - x\|_2 = C_1 E_{\mathrm{opt}}$

- (Terms, Error) approximation ratio: (C_2, C_1)

Greedy algorithms

Build approximation one step at a time...

Greedy algorithms

Build approximation one step at a time...

...choose the best atom at each step

Orthogonal Matching Pursuit OMP [Mallat '92], [Davis'97]

Input. Dictionary Φ, signal x, steps k

Output. Coefficient vector c with k nonzeros, $\Phi c \approx x$

Orthogonal Matching Pursuit OMP [Mallat '92], [Davis'97]

Input. Dictionary Φ, signal x, steps k

Output. Coefficient vector c with k nonzeros, $\Phi c \approx x$

Initialize. counter $t = 1$, residual $r_0 = x$, $c = 0$

Orthogonal Matching Pursuit OMP [Mallat '92], [Davis'97]

Input. Dictionary Φ, signal x, steps k

Output. Coefficient vector c with k nonzeros, $\Phi c \approx x$

Initialize. counter $t = 1$, residual $r_0 = x$, $c = 0$

1. **Greedy selection.** Find atom φ_{j_t} s.t.

$$j_t = \mathrm{argmax}_\ell |\langle r_{t-1},\ \varphi_\ell \rangle|$$

Orthogonal Matching Pursuit OMP [Mallat '92], [Davis'97]

Input. Dictionary Φ, signal x, steps k

Output. Coefficient vector c with k nonzeros, $\Phi c \approx x$

Initialize. counter $t = 1$, residual $r_0 = x$, $c = 0$

1. **Greedy selection.** Find atom φ_{j_t} s.t.

$$j_t = \mathrm{argmax}_\ell |\langle r_{t-1},\ \varphi_\ell\rangle|$$

2. **Update.** Find $c_{\ell_1}, \ldots, c_{\ell_t}$ to solve

$$\min \left\| x - \sum_s c_{\ell_s} \varphi_{\ell_s} \right\|_2$$

Orthogonal Matching Pursuit OMP [Mallat '92], [Davis'97]

Input. Dictionary Φ, signal x, steps k

Output. Coefficient vector c with k nonzeros, $\Phi c \approx x$

Initialize. counter $t = 1$, residual $r_0 = x$, $c = 0$

1. **Greedy selection.** Find atom φ_{j_t} s.t.

$$j_t = \mathrm{argmax}_\ell \left| \langle r_{t-1},\ \varphi_\ell \rangle \right|$$

2. **Update.** Find $c_{\ell_1}, \ldots, c_{\ell_t}$ to solve

$$\min \left\| x - \sum_s c_{\ell_s} \varphi_{\ell_s} \right\|_2$$

new residual $r_t \longleftarrow x - \Phi c$

Orthogonal Matching Pursuit OMP [Mallat '92], [Davis'97]

Input. Dictionary Φ, signal x, steps k

Output. Coefficient vector c with k nonzeros, $\Phi c \approx x$

Initialize. counter $t = 1$, residual $r_0 = x$, $c = 0$

1. **Greedy selection.** Find atom φ_{j_t} s.t.

$$j_t = \mathrm{argmax}_\ell | \langle r_{t-1},\ \varphi_\ell \rangle |$$

2. **Update.** Find $c_{\ell_1}, \ldots, c_{\ell_t}$ to solve

$$\min \left\| x - \sum_s c_{\ell_s} \varphi_{\ell_s} \right\|_2$$

new residual $r_t \longleftarrow x - \Phi c$

3. **Iterate.** $t \longleftarrow t + 1$, stop when $t > k$.

Many greedy algorithms with similar outline

- Matching Pursuit: replace step 2. by $c_{\ell_t} \longleftarrow c_{\ell_t} + \langle r_{t-1},\ \varphi_{k_t} \rangle$

- Thresholding

  Choose m atoms where $|\langle x,\ \varphi_\ell \rangle|$ are among m largest

- Alternate stopping rules:

  $\|r_t\|_2 \leq \epsilon$
  $\max_\ell |\langle r_t,\ \varphi_\ell \rangle| \leq \epsilon$

- *Many* other variations

Summary

- Sparse approximation problems are **NP-hard**
- At least as hard as other well-studied problems
- Hardness result of arbitrary input: *dictionary and signal*
- Intuition from orthonormal basis suggests some feasible solutions under certain conditions on redundant dictionary
- Geometric properties and greedy algorithms
- **Next lecture:** rigorous proofs for algorithms