Algorithms for sparse analysis
Lecture III: Dictionary geometry, greedy algorithms, and convex relaxation

Anna C. Gilbert

Department of Mathematics
University of Michigan

Convergence of OMP

Theorem

Suppose Φ is a complete dictionary for $\mathbb{R}^d$. For any vector x, the residual after t steps of OMP satisfies

$$\|r_t\|_2 \leq c \frac{1}{\sqrt{t}}.$$

[Devore-Temlyakov]

Convergence of OMP

Theorem

Suppose Φ is a complete dictionary for $\mathbb{R}^d$. For any vector x, the residual after t steps of OMP satisfies

$$\|r_t\|_2 \leq c\frac{1}{\sqrt{t}}.$$

[Devore-Temlyakov]

- Even if x can be expressed sparsely, OMP may take d steps before the residual is zero.

Convergence of OMP

Theorem

Suppose Φ is a complete dictionary for $\mathbb{R}^d$. For any vector x, the residual after t steps of OMP satisfies

$$\|r_t\|_2 \leq c\frac{1}{\sqrt{t}}.$$

[Devore-Temlyakov]

- Even if x can be expressed sparsely, OMP may take d steps before the residual is zero.
- *But*, sometimes OMP correctly identifies sparse representations.

Sparse representation with OMP

- Suppose x has k-sparse representation

$$x = \sum_{\ell \in \Lambda} c_\ell \varphi_\ell \quad \text{where } |\Lambda| = k$$

i.e., c_{opt} is non-zero on Λ.

Sparse representation with OMP

- Suppose x has k-sparse representation

$$x = \sum_{\ell \in \Lambda} c_\ell \varphi_\ell \quad \text{where } |\Lambda| = k$$

  i.e., c_{opt} is non-zero on Λ.
- Sufficient to find Λ—When can OMP do so?

Sparse representation with OMP

- Suppose x has k-sparse representation

$$x = \sum_{\ell \in \Lambda} c_\ell \varphi_\ell \quad \text{where } |\Lambda| = k$$

  i.e., c_{opt} is non-zero on Λ.
- Sufficient to find Λ—When can OMP do so?
- Define

$$\Phi_\Lambda = \begin{bmatrix} \varphi_{\ell_1} & \varphi_{\ell_2} & \cdots & \varphi_{\ell_k} \end{bmatrix}_{\ell_s \in \Lambda} \quad \text{and}$$
$$\Psi_\Lambda = \begin{bmatrix} \varphi_{\ell_1} & \varphi_{\ell_2} & \cdots & \varphi_{\ell_{N-k}} \end{bmatrix}_{\ell_s \notin \Lambda}$$

Sparse representation with OMP

- Suppose x has k-sparse representation

$$x = \sum_{\ell \in \Lambda} c_\ell \varphi_\ell \quad \text{where } |\Lambda| = k$$

  i.e., c_{opt} is non-zero on Λ.
- Sufficient to find Λ—When can OMP do so?
- Define

$$\Phi_\Lambda = \begin{bmatrix} \varphi_{\ell_1} & \varphi_{\ell_2} & \cdots & \varphi_{\ell_k} \end{bmatrix}_{\ell_s \in \Lambda} \quad \text{and}$$
$$\Psi_\Lambda = \begin{bmatrix} \varphi_{\ell_1} & \varphi_{\ell_2} & \cdots & \varphi_{\ell_{N-k}} \end{bmatrix}_{\ell_s \notin \Lambda}$$

- Define *greedy selection ratio*

$$\rho(r) = \frac{\max_{\ell \notin \Lambda} |\langle r, \ \varphi_\ell \rangle|}{\max_{\ell \in \Lambda} |\langle r, \ \varphi_\ell \rangle|} = \frac{\left\| \Psi_\Lambda^T r \right\|_\infty}{\left\| \Phi_\Lambda^T r \right\|_\infty} = \frac{\text{max i.p. bad atoms}}{\text{max i.p. good atoms}}$$

Sparse representation with OMP

- Suppose x has k-sparse representation

$$x = \sum_{\ell \in \Lambda} c_\ell \varphi_\ell \quad \text{where } |\Lambda| = k$$

  i.e., c_{opt} is non-zero on Λ.
- Sufficient to find Λ—When can OMP do so?
- Define

$$\Phi_\Lambda = \begin{bmatrix} \varphi_{\ell_1} & \varphi_{\ell_2} & \cdots & \varphi_{\ell_k} \end{bmatrix}_{\ell_s \in \Lambda} \quad \text{and}$$
$$\Psi_\Lambda = \begin{bmatrix} \varphi_{\ell_1} & \varphi_{\ell_2} & \cdots & \varphi_{\ell_{N-k}} \end{bmatrix}_{\ell_s \notin \Lambda}$$

- Define *greedy selection ratio*

$$\rho(r) = \frac{\max_{\ell \notin \Lambda} |\langle r, \ \varphi_\ell \rangle|}{\max_{\ell \in \Lambda} |\langle r, \ \varphi_\ell \rangle|} = \frac{\left\| \Psi_\Lambda^T r \right\|_\infty}{\left\| \Phi_\Lambda^T r \right\|_\infty} = \frac{\text{max i.p. bad atoms}}{\text{max i.p. good atoms}}$$

- OMP chooses good atom iff $\rho(r) < 1$

Exact Recovery Condition

Theorem (ERC)

A sufficient condition for OMP *to identify* Λ *after* k *steps is that*

$$\max_{\ell \notin \Lambda} \left\| \Phi_\Lambda^+ \varphi_\ell \right\|_1 < 1$$

where $A^+ = (A^T A)^{-1} A^T$. *[Tropp'04]*

Exact Recovery Condition

Theorem (ERC)

A sufficient condition for OMP *to identify* Λ *after* k *steps is that*

$$\max_{\ell \notin \Lambda} \left\| \Phi_\Lambda^+ \varphi_\ell \right\|_1 < 1$$

where $A^+ = (A^T A)^{-1} A^T$. *[Tropp'04]*

- A^+x is a coefficient vector that synthesizes best approximation of x using atoms in A.
- $P = AA^+$ orthogonal projector produces this best approximation

Proof.

$r_0 = x \in \mathrm{range}(\Phi_\Lambda)$

Proof.

$r_0 = x \in \mathrm{range}(\Phi_\Lambda)$

At iteration $t + 1$ assume $\ell_1, \ell_2, \ldots, \ell_t \in \Lambda$, thus $a_t \in \mathrm{range}(\Phi_\Lambda)$ and $r_t = x - a_t \in \mathrm{range}(\Phi_\Lambda)$

Proof.

$r_0 = x \in \text{range}(\Phi_\Lambda)$

At iteration $t+1$ assume $\ell_1, \ell_2, \ldots, \ell_t \in \Lambda$, thus $a_t \in \text{range}(\Phi_\Lambda)$ and $r_t = x - a_t \in \text{range}(\Phi_\Lambda)$

Express orthogonal projector onto range(Φ_Λ) as $(\Phi_\Lambda^+)^T \Phi_\Lambda^T$, therefore

$$(\Phi_\Lambda^+)^T \Phi_\Lambda^T r_t = r_t.$$

Proof.

$r_0 = x \in \text{range}(\Phi_\Lambda)$

At iteration $t+1$ assume $\ell_1, \ell_2, \ldots, \ell_t \in \Lambda$, thus $a_t \in \text{range}(\Phi_\Lambda)$ and $r_t = x - a_t \in \text{range}(\Phi_\Lambda)$

Express orthogonal projector onto $\text{range}(\Phi_\Lambda)$ as $(\Phi_\Lambda^+)^T \Phi_\Lambda^T$, therefore

$$(\Phi_\Lambda^+)^T \Phi_\Lambda^T r_t = r_t.$$

Bound

$$\begin{aligned}
\rho(r_t) &= \frac{\left\| \Psi_\Lambda^T r_t \right\|_\infty}{\left\| \Phi_\Lambda^T r_t \right\|_\infty} = \frac{\left\| \Psi_\Lambda^T (\Phi_\Lambda^+)^T \Phi_\Lambda^T r_t \right\|_\infty}{\left\| \Phi_\Lambda^T r_t \right\|_\infty} \\
&\leq \left\| \Psi_\Lambda^T (\Phi_\Lambda^+)^T \right\|_\infty \\
&= \left\| \Phi_\Lambda^+ \Psi_\Lambda \right\|_1 \\
&= \max_{\ell \notin \Lambda} \left\| \Phi_\Lambda^T \varphi_\ell \right\|_1 < 1
\end{aligned}$$

Proof.

$r_0 = x \in \text{range}(\Phi_\Lambda)$

At iteration $t+1$ assume $\ell_1, \ell_2, \ldots, \ell_t \in \Lambda$, thus $a_t \in \text{range}(\Phi_\Lambda)$ and $r_t = x - a_t \in \text{range}(\Phi_\Lambda)$

Express orthogonal projector onto $\text{range}(\Phi_\Lambda)$ as $(\Phi_\Lambda^+)^T \Phi_\Lambda^T$, therefore

$$(\Phi_\Lambda^+)^T \Phi_\Lambda^T r_t = r_t.$$

Bound

$$\begin{aligned}
\rho(r_t) &= \frac{\left\|\Psi_\Lambda^T r_t\right\|_\infty}{\left\|\Phi_\Lambda^T r_t\right\|_\infty} = \frac{\left\|\Psi_\Lambda^T (\Phi_\Lambda^+)^T \Phi_\Lambda^T r_t\right\|_\infty}{\left\|\Phi_\Lambda^T r_t\right\|_\infty} \\
&\leq \left\|\Psi_\Lambda^T (\Phi_\Lambda^+)^T\right\|_\infty \\
&= \left\|\Phi_\Lambda^+ \Psi_\Lambda\right\|_1 \\
&= \max_{\ell \notin \Lambda} \left\|\Phi_\Lambda^T \varphi_\ell\right\|_1 < 1
\end{aligned}$$

Then OMP selects an atom from Λ at iteration t and since it chooses a new atom at each iteration,

Proof.

$r_0 = x \in \text{range}(\Phi_\Lambda)$

At iteration $t+1$ assume $\ell_1, \ell_2, \ldots, \ell_t \in \Lambda$, thus $a_t \in \text{range}(\Phi_\Lambda)$ and $r_t = x - a_t \in \text{range}(\Phi_\Lambda)$

Express orthogonal projector onto range(Φ_Λ) as $(\Phi_\Lambda^+)^T \Phi_\Lambda^T$, therefore

$$(\Phi_\Lambda^+)^T \Phi_\Lambda^T r_t = r_t .$$

Bound

$$\begin{aligned}
\rho(r_t) &= \frac{\left\| \Psi_\Lambda^T r_t \right\|_\infty}{\left\| \Phi_\Lambda^T r_t \right\|_\infty} = \frac{\left\| \Psi_\Lambda^T (\Phi_\Lambda^+)^T \Phi_\Lambda^T r_t \right\|_\infty}{\left\| \Phi_\Lambda^T r_t \right\|_\infty} \\
&\leq \left\| \Psi_\Lambda^T (\Phi_\Lambda^+)^T \right\|_\infty \\
&= \left\| \Phi_\Lambda^+ \Psi_\Lambda \right\|_1 \\
&= \max_{\ell \notin \Lambda} \left\| \Phi_\Lambda^T \varphi_\ell \right\|_1 < 1
\end{aligned}$$

Then OMP selects an atom from Λ at iteration t and since it chooses a new atom at each iteration, After k iterations, chosen all atoms from Λ.

□

Coherence Bounds

Theorem

The ERC holds whenever $k < \frac{1}{2}(\mu^{-1} + 1)$. Therefore, OMP can recover any sufficiently sparse signals. [Tropp'04]

For most redundant dictionaries, $k < \frac{1}{2}(\sqrt{d} + 1)$.

Sparse

Theorem

Assume $k \leq \frac{1}{3}\mu^{-1}$. For any vector x, the approximation $\Phi\widehat{c}$ after k steps of OMP satisfies $\|\widehat{c}\|_0 \leq k$ and

$$\|x - \Phi\widehat{c}\|_2 \leq \sqrt{1+6k}\,\|x - \Phi c_{\mathrm{opt}}\|_2$$

where c_{opt} is the best k-term approximation to x over Φ. *[Tropp'04]*

Theorem

Assume $4 \leq k \leq \frac{1}{\sqrt{\mu}}$. Two-phase greedy pursuit produces $\widehat{x} = \Phi\widehat{c}$ s.t.

$$\|x - \widehat{x}\|_2 \leq 3\,\|x - \Phi c_{\mathrm{opt}}\|_2 .$$

Assume $k \leq \frac{1}{\mu}$. Two-phase greedy pursuit produces $\widehat{x} = \Phi\widehat{c}$ s.t.

$$\|x - \widehat{x}\|_2 \leq \left(1 + \frac{2\mu k^2}{(1-2\mu k)^2}\right)\|x - \Phi c_{\mathrm{opt}}\|_2 .$$

[Gilbert, Strauss, Muthukrishnan, Tropp '03]

Convex relaxation: BP

- EXACT: non-convex optimization

$$\arg\min \|c\|_0 \quad \text{s.t.} \quad x = \Phi c$$

Convex relaxation: BP

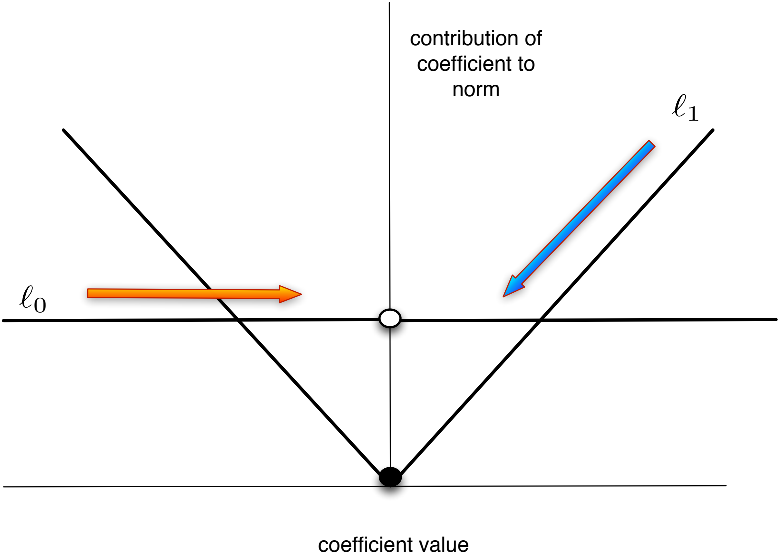
- EXACT: non-convex optimization

$$\arg\min \|c\|_0 \quad \text{s.t.} \quad x = \Phi c$$

- Convex relaxation of non-convex problem

$$\arg\min \|c\|_1 \quad \text{s.t.} \quad x = \Phi c$$

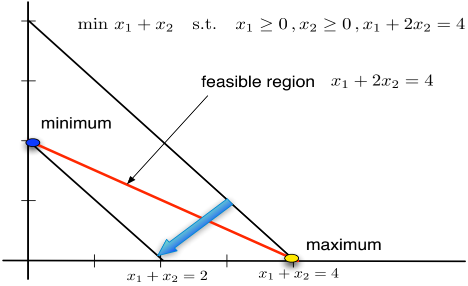
Convex relaxation



Convex relaxation: algorithmic formulation

- Well-studied algorithmic formulation [Donoho, Donoho-Elad-Temlyakov, Tropp, and many others]
- Optimization problem = linear program: linear objective function (with variables c^+, c^-) and linear constraints
- Still need *algorithm* for solving optimization problem
- Hard part of analysis: showing solution to convex problem = solution to original problem

LP



- Feasible region is convex polytope
- Linear objective function: convex and concave $\Longrightarrow$ local minimum/maximum are global
- If feasible solution exists and if objective function bounded, then optimum achieved on boundary (possibly many points)

Exact Recovery Condition

Theorem (ERC)

A sufficient condition for BP *to recover the sparsest representation of* x *is that*

$$\max_{\ell \notin \Lambda} \left\| \Phi_\Lambda^+ \varphi_\ell \right\|_1 < 1$$

where $A^+ = (A^T A)^{-1} A^T$. *[Tropp'04]*

Exact Recovery Condition

Theorem (ERC)

A sufficient condition for BP *to recover the sparsest representation of* x *is that*

$$\max_{\ell \notin \Lambda} \left\| \Phi_\Lambda^+ \varphi_\ell \right\|_1 < 1$$

where $A^+ = (A^T A)^{-1} A^T$. [Tropp'04]

Theorem

The ERC holds whenever $k < \frac{1}{2}(\mu^{-1} + 1)$. *Therefore,* BP *can recover any sufficiently sparse signals.* [Tropp'04]

Convex relaxation: BP-denoising

- ERROR: non-convex optimization

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

Convex relaxation: BP-denoising

- ERROR: non-convex optimization

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

- Convex relaxation of non-convex problem

$$\arg\min \|c\|_1 \quad \text{s.t.} \quad \|x - \Phi c\|_2 \leq \delta.$$

- Convex objective function over convex set.

Optimization formulations

- Constrained minimization

$$\arg\min \|c\|_1 \quad \text{s.t.} \quad \|x - \Phi c\|_2 \leq \delta.$$

- Unconstrained minimization (ℓ_1-regularization): minimize

$$L(c; \gamma, x) = \frac{1}{2}\|x - \Phi c\|_2^2 + \gamma \|c\|_1 .$$

- **Constrained minimization**

Theorem

Suppose that $k \leq \frac{1}{3}\mu^{-1}$. Suppose c_{opt} is k-sparse and solves original optimization problem. Then solution $\widehat{c}$ to constrained minimization problem has same sparsity and satisfies

$$\|x - \Phi\widehat{c}\|_2 \leq \left(\sqrt{1+6k}\right)\epsilon.$$

[Tropp '04]

- **Unconstrained minimization:** many algorithms for ℓ_1-regularization (e.g., Bregman iteration, interior point methods, LASSO and LARS)

Optimization vs. Greedy

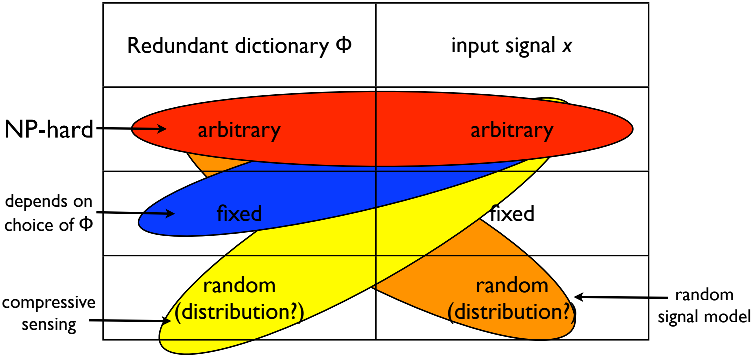
- EXACT and ERROR amenable to convex relaxation and convex optimization

- SPARSE not amenable to convex relaxation

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

but appropriate for greedy algorithms

Hardness depends on instance



Random signal model

Theorem

If Φ has consistent coherence $\mu = 1/\sqrt{d}$, choose $k \sim d/\log d$ atoms for x at random from Φ, then sparse representation is unique and, given x and Φ, convex relaxation finds it. [Tropp'07]

Summary

- Geometry of dictionary is important *but*
- Obtain *sufficient* conditions on the geometry of the dictionary to solve SPARSE problems efficiently.
- Algorithms are *approximation* algorithms (wrt error).
- Greedy pursuit and convex relaxation.
- **Next lecture:** Sublinear algorithms for sparse approximation and compressive sensing