Greedy pursuit methods

$\min \|x\|_0$ st $Ax = b$ $\quad A \in \mathbb{R}^{m \times n}$ $\quad b = Ax_0$ $\quad \|x_0\| = s \ll n$

where desired x_0 is sparse

~~[illegible]~~

$A = \begin{pmatrix} | & | & | & | & | & | \end{pmatrix}$

b is linear combination of a few cols of A

Idea: select "best" column(s) of A
update coeff's for those colums

Best column of A has largest dot product with a residual

Eg. Find max $A^t b$ (col w/ greatest dot prodct w/ b)
Solve a least squars problem to get coef $\to \hat{x}^{[1]}$
Find max $A^t(b - A\hat{x}^{[1]})$
solve least squas systm $\to \hat{x}^{[2]}$
repeat

Matching pursuit

Idea: - At any iteration, select largest dot product of col of A w/ residual (even if repeat)

- perform least squares only over new coef

Algorithm:

(assume columns of A are A_j and are unit length)

$$g^{(i)} = A^t r^{(i-1)}$$
$$j^{(i)} = \operatorname{argmax}_j |g_j^{(i)}|$$
$$\hat{x}_{j^{(i)}}^{(i)} = \hat{x}_{j^{(i)}}^{(i-1)} + g_{j^{(i)}}^{(i)}$$
$$r^{(i)} = r^{(i-1)} - A_{j^{(i)}} g_{j^{(i)}}^{(i)}$$

Note: $\hat{x}$ update step minimizes one variable least squares problem

$$\min_t \|r - A_j t\|_2^2 \rightarrow A_j^t A_j t = A_j^t r$$
$$t = \underbrace{A_j^t r}_{j^{th} \text{ entry of } A^t r} \quad \text{if } \|A_j\|_2 = 1$$

Expensive step: Applying A^t

Good when A - is like FFT
- has sparse columns

Comments: $\|r^{(i)}\| \rightarrow 0$ Can use $\|r^{(i)}\|$ as stopping criterion
will be final r

Can reselect same column

Orthogonal Matching pursuit

Idea:
- At each iteration select largest dot prod of col of A w residual (no repeats)
- perform least squares over all coefs selected

Algorithm (where $\|A_j\|_2 = 1$)

$$g^{(i)} = A^t r^{(i-1)}$$

$$j^{(i)} = \arg\max_j |g_j^{(i)}|$$

$$T^{(i)} = T^{(i-1)} \cup j^{(i)}$$ — add new coef to active set

$$\hat{x}^{(i)} = \arg\min \|A_T x - b\|_2$$ — complete least squares over all coefs

$$r^{(i)} = y - A\hat{x}^{(i)}$$ — projects b onto orthogonal complement of colums of A_T

Note: never reselects same coef twice.

Performance typically better than MP

More expensive than MP

When fast transforms used to compute $A^t r$, bottleneck will be orthogonalizen step in least squares update. (QR, cholesky)

Performance guarantees for greedy pursuit algorithms

(Donoho; Rauhut)

Thm: For certain random $A \in \mathbb{R}^{m \times n}$ w/ $m \sim k \log n$,
whp $\exists\ x_0$ st $\|x_0\|_0 \leq k$ and $\arg\max_j |(A^t b)_j| \notin \operatorname{supp} x_0$

Greedy pursuit methods get off to wrong start. and selects wrong elt.

Could still be fine if recovery considered successful if it is a superset of true set.

Could still be fine if we do not seek successful recovery for all signals simultaneously.

Performance guarantee

Fix $x_0 \in \mathbb{R}^n$ st $\|x_0\| = k$. Let $A \in \mathbb{R}^{m \times n}$ iid $N(0,1)$ entries

Let $b = Ax_0$. If $m \geq C k \log(n/\sqrt{\delta})$ then OMP

succeeds w/ probability at least $1-\delta$. C depends on A

(≈ 2 as $n \to \infty$)

Note: Success only proven for a fixed x_0. Not for all x_0 simultaneously.

Scaling is linear in k.

Iterative Hard Thresholding

$$\min_{x} \; \|Ax - b\|_2^2 \quad \text{st} \quad \|x\|_0 \leq k \qquad (*)$$

Hard thresholdin: $H_k(x)$ – keeps largest k coefs of x.

Algorithm

$$X^{(i+1)} = \underbrace{H_k}_{\text{nonlinear projection}} \Big(\underbrace{X^{(i)} - \mu A^T (Ax^{(i)} - b)}_{\text{gradient descen}} \Big)$$

~~Theorem (Blumensath + Davies)~~

When do we expect convergence?

When μ is small relative to "A".

When μ is small relative to RIP constant of A

Let β_{2k} be smallest # such that

$$\|A(x_1 - x_2)\|_2^2 \leq \beta_{2k} \|x_1 - x_2\|_2^2 \qquad \forall \; k\text{-sparse } x_1 \,\&\, x_2.$$

The $\beta_{2k} \leq (1 + \delta_{2k})$

Thm (Blumensath + Davies)

If $\mu \leq \dfrac{1}{\beta_{2k}}$ and $k \leq m$ and A full rank

then IHT converges to a local minimizer of $(*)$

Performance of IHT:

Let X_k be top k coeffs of X.

Let $\tilde{e} = A(X_0 - X_{0,k}) + e$

Thm: Fix X_0, Let $b = AX_0 + e$ w/ A having nonsymmetric RIP constants $\alpha_{2k} \|X_1 - X_2\|_2^2 \leq \|A(X_1 - X_2)\|_2^2 \leq \beta_{2k} \|X_1 - X_2\|_2^2$ $\forall$ k sparse X_1 & X_2

st $\beta_{2k} \leq \frac{1}{\mu} \leq \frac{3}{2}\alpha_{2k}$

then
$$\|X_k - X^{(i)}\|_2^2 \leq \left[2\left(\frac{1}{\mu\alpha_{2k}} - 1\right)\right]^i \|X_k\|_2^2 + c\|\tilde{e}\|_2^2$$

w/ $c \leq \dfrac{4}{3\alpha_{2k} - 2/\mu}$.

So IHT correct for small enough step size, relative to nonsymm. RIP consts

CoSaMP (compressive sampling matching pursuit)

Needell + Tropp

Gist: Keep track of active set of indices
add to them and prune them

At each iteration
- $\hat{x}^{(i)}$ is k-sparse estimate
- compute largest 2k entries of $A^*(b - A\hat{x}^{(i)})$ — adding indices
- consider indices in supp of $\hat{x}^{(i)}$ and these 2k
- Solve least squares on those coeffs — pruning
- Take top k coeffs.

Theorem (Needell + Tropp)

Let A be $m \times n$ w/ RIP $\delta_{2k} < c$. Let $b = Ax + e$.
Fix precision param η. CoSaMP finds a ~~2k sparse~~ vec a after $O\left(\log \frac{\|x_0\|_2}{\eta}\right)$ iterations
st $\|x_0 - a\|_2 \leq C \max\left(\eta, \frac{1}{\sqrt{k}} \|x_0 - x_{0,k}\|_1 + \|e\|_2\right)$