Topics in Sparse Recovery

Piotr Indyk
MIT

Sparse recovery

(approximation theory, learning Fourier coeffs, linear sketching, finite rate of innovation, **compressed sensing...)**

- Setup:
  - Data/signal in n-dimensional space : x
  - Goal: compress x into a "sketch" or "measurement vector" Ax , where A is a m xn matrix, $m << n$
- Goal: want to recover a k-sparse approximation x^* of x
  - Sparsity parameter k
  - Informally: want to recover the largest k coordinates of x
- Guarantees: for some $C>1$
  - L1/L1:
    $$\|x-x^*\|_1 \le C \min_{k\text{-sparse } x''} \|x-x''\|_1$$
  - L2/L2:
    $$\|x-x^*\|_2 \le C \min_{k\text{-sparse } x''} \|x-x''\|_2$$
  - L2/L1:
    $$\|x-x^*\|_2 \le C \min_{k\text{-sparse } x''} \|x-x''\|_1 / k^{1/2}$$

  (other variants later)
- Want:
  - Good compression (small $m=m(k,n)$)
  - Efficient algorithms for encoding and recovery
- Useful for compressed sensing of signals, data stream algorithms, group testing etc….

Scale: Excellent | Very Good | Good | Fair

Results

Paper	R/D	Sketch length	Encode time	Column sparsity	Recovery time	Approx
[CCF'02], [CM'06]	R	k log n	n log n	log n	n log n	l2 / l2
	R	k $\log^c$ n	n $\log^c$ n	$\log^c$ n	k $\log^c$ n	l2 / l2
[CM'04]	R	k log n	n log n	log n	n log n	l1 / l1
	R	k $\log^c$ n	n $\log^c$ n	$\log^c$ n	k $\log^c$ n	l1 / l1
[CRT'04] [RV'05]	D	k log(n/k)	nk log(n/k)	k log(n/k)	n^c	l2 / l1
	D	k $\log^c$ n	n log n	k $\log^c$ n	n^c	l2 / l1
[GSTV'06] [GSTV'07]	D	k $\log^c$ n	n $\log^c$ n	$\log^c$ n	k $\log^c$ n	l1 / l1
	D	k $\log^c$ n	n $\log^c$ n	k $\log^c$ n	$k^2 \log^c$ n	l2 / l1
[BGIKS'08]	D	k log(n/k)	n log(n/k)	log(n/k)	n^c	l1 / l1
[GLR'08]	D	k $\log n^{\log\log\log n}$	kn^{1-a}	n^{1-a}	n^c	l2 / l1
[NV'07], [DM'08], [NT'08], [BD'08], [GK'09], …	D	k log(n/k)	nk log(n/k)	k log(n/k)	nk log(n/k) * log R	l2 / l1
	D	k $\log^c$ n	n log n	k $\log^c$ n	n log n * log R	l2 / l1
[IR'08], [BIR'08],[BI'09]	D	k log(n/k)	n log(n/k)	log(n/k)	n log(n/k)* log R	l1 / l1
[GLSP'09]	R	k log(n/k)	n $\log^c$ n	$\log^c$ n	k $\log^c$ n	l2 / l1

R= $||x||_1$ if x has integer coordinates (dynamic range of signals).

Scale: Excellent | Very Good | Good | Fair

Results

Focus on:
- Deterministic
- k is "large"

Paper	R/D	Sketch length	Recovery time	Approx
[CCF'02], [CM'06]	R	k log n	n log n	l2 / l2
	R	k $\log^c$ n	k $\log^c$ n	l2 / l2
[CM'04]	R	k log n	n log n	l1 / l1
	R	k $\log^c$ n	k $\log^c$ n	l1 / l1
[CRT'04] [RV'05]	D	k log(n/k)	n^c	l2 / l1
	D	k $\log^c$ n	n^c	l2 / l1
[GSTV'06] [GSTV'07]	D	k $\log^c$ n	k $\log^c$ n	l1 / l1
	D	k $\log^c$ n	$k^2 \log^c$ n	l2 / l1
[BGIKS'08]	D	k log(n/k)	n^c	l1 / l1
[GLR'08]	D	k $\log n^{\log\log\log n}$	n^c	l2 / l1
[NV'07], [DM'08], [NT'08], [BD'08], [GK'09], …	D	k log(n/k)	nk log(n/k) * log R	l2 / l1
	D	k $\log^c$ n	n log n * log R	l2 / l1
[IR'08], [BIR'08],[BI'09]	D	k log(n/k)	n log(n/k)+	l1 / l1
[GLSP'09]	R	k log(n/k)	k $\log^c$ n	l2 / l1

R= $||x||_1$ if x has integer coordinates (dynamic range of signals).

Matrices A

- Restricted Isometry Property (RIP-2) [Candes-Tao'04]

$$\Delta \text{ is } k\text{-sparse} \Rightarrow ||\Delta||_2 \leq ||A\Delta||_2 \leq C\,||\Delta||_2$$

- Holds w.h.p. for random Gaussian/Bernoulli matrices with $m=O(k \log (n/k))$
  - JL lemma applied to the set of k-sparse vectors
    (after discretization it has size about (n choose k))

R/ D	Sketch length	Recovery time	Approx
D	k log(n/k)	nk log(n/k) * log R	l2 / l1

- Holds w.h.p. for matrices consisting of $m=O(k \log^{O(1)} n)$ random rows of Fourier matrix
  - [Candes-Tao'04], [Rudelson-Vershynin'05]

D	k log^c n	n log n * log R	l2 / l1

- RIP-1 property [Berinde-Gilbert-Indyk-Karloff-Strauss'08]:

$$\Delta \text{ is } k\text{-sparse} \Rightarrow d\,(1-\varepsilon)\,||\Delta||_1 \leq ||A\Delta||_1 \leq d||\Delta||_1$$

- Sufficient (and necessary) condition: A is an adjacency matrix of a $(k, d(1-\varepsilon/2))$-expander

D	k log(n/k)	n log(n/k)+	l1 / l1

R/D	Sketch length	Recovery time	Approx
D	k log(n/k)	nk log(n/k) * log R	l2 / l1
D	k $\log^c$ n	n log n * log R	l2 / l1
D	k log(n/k)	n log(n/k)+	l1 / l1

D	k log(n/k)	n $\log^c$ n * log R	l2 / l1

?

- Limitations:
  - The bound $m=O(k \log(n/k))$ is tight
    - Deterministic (essentially) from Kolmogorov width [Foucart-Pajor-Rauhut-Ullrich'10]
    - Randomized: [DoBa-Indyk-Price-Woodruff'10]
  - For l2/l2 and deterministic need $m=\Omega(n)$ [Cohen-Dahmen-Devore]
- Best of all worlds ?
- Could be achieved from fast and dimension-efficient JL transform (if we had it) ?
  - Reduce dimension to $O(\log(1/\delta)/\varepsilon^2)$ with failure probability δ
  - In $n \log^{O(1)} n$ time
- [Ailon-Chazelle'10, Krahmer-Ward'10]:
  - $n \log n$ time
  - But dimension $O(\log(1/\delta)/\varepsilon^2 * \log^{O(1)} n)$

Model-based compressed sensing

- We have seen sketches with
  - Guarantee: $||x-x^*||_1 \leq C \min_{k\text{-sparse } x''} ||x-x''||_1$
  - Sketch length $m=O(k \log (n/k))$

- An alternative

$$||x-x^*||_1 \leq C \min_{\text{supp}(x'') \text{ in } M} ||x-x''||_1$$

  where M is a family of sets of size k.
- Example model: M = l blocks, each of length b, $k=bl$
  - $|M|= (n/b \text{ choose } l)2^k$
  - Can achieve $m=O(k + l \log (n/(bl)))$ (JL applied to a smaller set of points)
- Other models: connected trees

Results

Paper	R/ D	Recovery time	Approx	Factor
[EM'09]+	R	n^c	l2 / l2	O(1)
[BCDH'09]+	R	nk * log R	l2 / l2	O(1)
[BCDH'09]+	D	nk * log R	l1 / l1	$(\log n)^{1/2}$
[P'11]	R	k + n/b * log n	l2 / l2	O(1)

R: for each x a random matrix A works with probability >0.9

D: there exists a matrix A which works for each x

Fast and dimension-optimal JL would imply

=>	D	$n \log^c n$	l1 / l1	$(\log n)^{1/2}$
=>	R	$n \log^c n$	l2 / l2	O(1)

Sparse Recovery under EMD

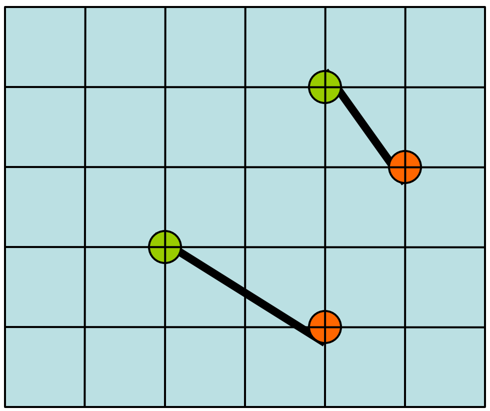
[Indyk, Price'10]

- The guarantee:

$$\|x-x^*\|_1 \le C \min_{k\text{-sparse } x''} \|x-x''\|_1$$

- Lp norms are a natural choice, but there are *many* distance functions out there
  - Edit distance in computational biology
  - Earth-Mover Distance (a.k.a. transportation metric) in computer vision
  - …

Earth-Mover Distance: EMD(x,y)



- Vectors x and y represent two-dimensional sxs arrays ($n=s^2$)
- Consider the case when:
  - $||x||_1=||y||_1$
  - x and y are binary, representing subsets X and Y of $[s]^2$

  Then EMD(x,y) is the min-cost matching between X and Y
- Generalizations to non-binary vectors with unequal L1 mass

EMD: properties

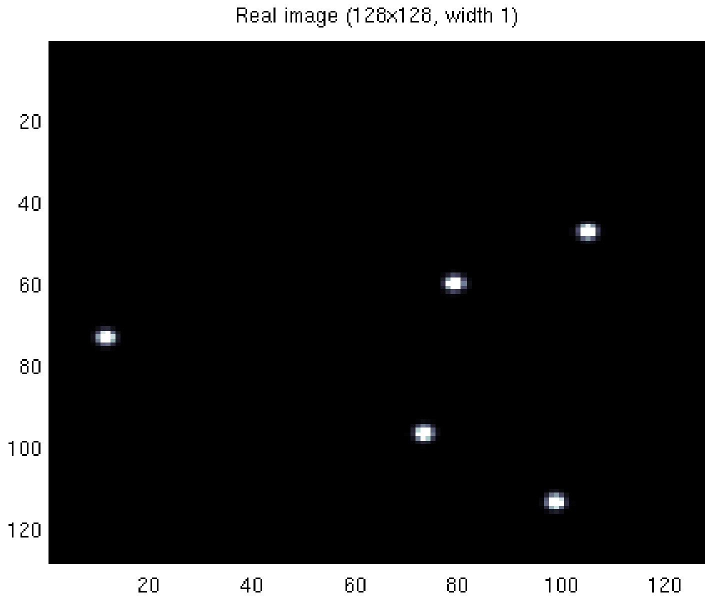
- Differentiates between “small” and “large” perturbations
  - In contrast, L1 distance is the same in both cases
- Implicitly defines correspondences
  - Useful when dealing with image features
- Popular in computer vision [Rubner et al’00, …]
  - In many scenarios well-approximates the perceived difference between images

Sparse Recovery under EMD

- Goal: given Ax, find x* s.t.

$$EMD(x,x^*) \le C \min_{k\text{-sparse } x''} EMD(x,x'')$$

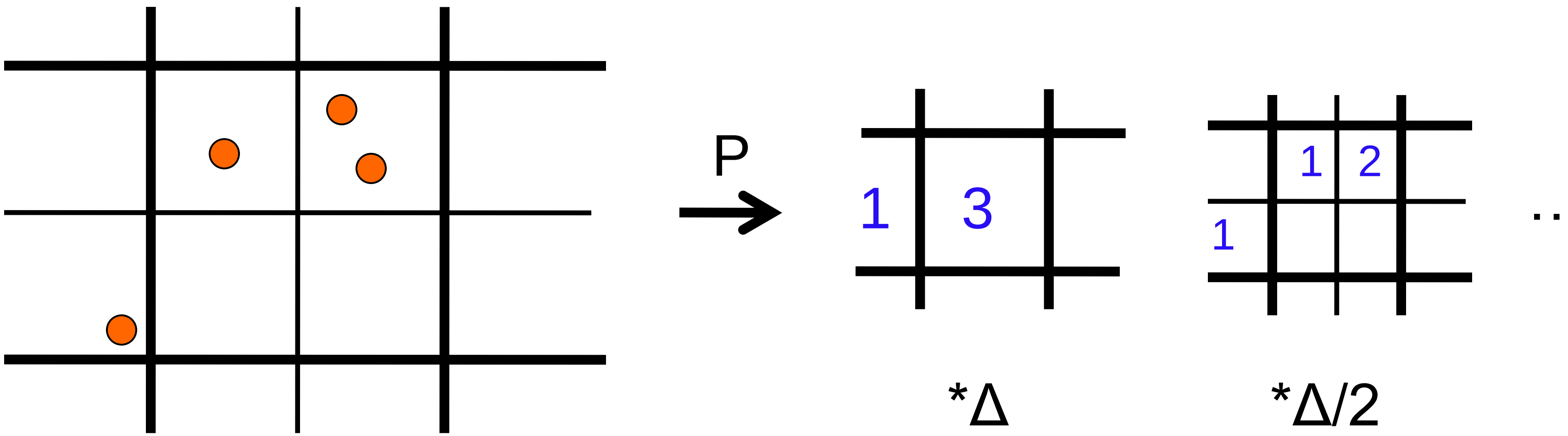
- What does this mean ?
  - The minimizer x" contains k non-zero coordinates/points, centered at the "clusters" in x
  - The values of those coordinates correspond to the cluster weights
- Want short linear sketches for recovering a near-optimal clustering [Indyk'04; Frahling-Sohler'05]



Recovery Schemes

- Two questions:
  - What is the measurement matrix ?
  - What is the algorithm that recovers x^* ?
- High level idea:
  - Reduce the problem under EMD to a problem under L1 norm (distance preserving embedding)
  - Solve the problem in the L1 norm

The pyramid mapping



- Grids $G_0, G_1, \ldots$ with cell lengths $\Delta = 2^0, 2^1, 2^2 \ldots$
- Lemma [Indyk-Thaper'03, following Charikar'02, Bartal]: If the grids are shifted by random vector $v \in \{1 \ldots s\}^2$ then
  - $E[\ \|Px - Py\|_1\ ] \leq O(\log s)\ EMD(x,y)$
  - $\|Px - Py\|_1 \geq EMD(x,y)$

Sparse recovery in EMD via pyramid mapping P

- We have that $||Px - Py||_1 \approx EMD(x,y)$
- If x is k-sparse then Px is $O(k \log n)$ sparse
- Approach:
  - Let A be a measurement matrix enabling $O(k \log n)$-sparse recovery under L1 norm
  - Use $A' = A*P$ for k-sparse recovery under EMD
  - Recover $O(k \log n)$-sparse z that approximately minimizes $||z-Px||_1$
- Additional constraints:
  - x is non-negative
  - If x is k-sparse, the non-zero coefficients of Px form a tree
- Enforcing these reduces the number of measurements (as in *model-based compressive sensing*)

Results

R/D	Sketch length	Recovery time	Factor	Comments
D	k log (n/k) logn	n log n	O(1)	Pyramid mapping +sparse recovery in L1
D	k log (n/k)	nk * log R	$(\log n)^{1/2}$	Pyramid mapping+model-based recovery
R	k log (n/k)	k * log n	O(1)	Pyramid mapping+tailored recovery algorithm

Conclusions

- Connections between:
  - (Model)-based sparse recovery
  - EMD
  - Sketching algorithms for k-median
  - (Fast) JL
- Many open questions