Compressed Sensing and Related Learning Problems

Yingzhen Li

Dept. of Mathematics, Sun Yat-sen University

Advisor: Prof. Haizhang Zhang

Overview

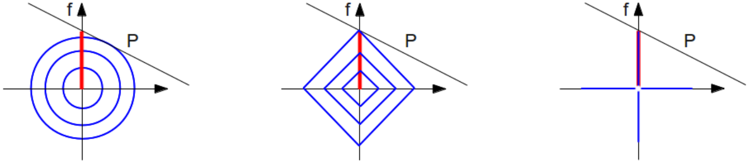
- Background
- Compressed Sensing
    - Norms Indicating Sparsity
    - Restricted Isometry Property (RIP)
    - Recovery Algorithms
    - Applications
- Learning Methods
    - Probabilistic Methods
    - Dictionary Learning
- Conclusion

Please refer to my thesis for details.

Background

- Signal processing everywhere.
  - Entertainment: music, videos, images...
  - Engineering: telecommunication, medical use...
  - Recognition: speech, face, moving object...
- Limitation of the Nyquist-Shannon Sampling Theorem
- Era of the big data.
  - Daily life accompanied with data.
  - Knowledge discovery from data.
- Any fast algorithms?

Why use l_1-Norm?



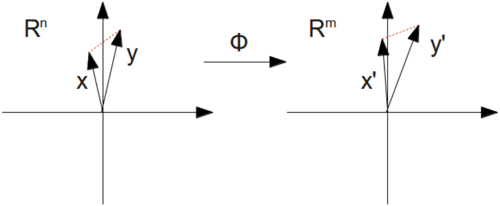
(a) $p = 2$ (b) $p = 1$ (c) $p = 0$

- Recovery of the signal $f \in \mathbb{R}^n$ from $\mathbf{y} \in \mathbb{R}^m$:

$$f^* = \arg\min_f ||f||_p \ s.t. \ \mathbf{y} = \Phi f. \tag{1}$$

- (P_0) $f^* = \arg\min_f ||f||_0 \ s.t. \ \mathbf{y} = \Phi f$ (Matching Pursuit)
- (P_1) $f^* = \arg\min_f ||f||_1 \ s.t. \ \mathbf{y} = \Phi f$ (Basis Pursuit)

Restricted Isometry Property



Definition (Restricted Isometry Property [CT05])

The sensing matrix Φ is said to obey the restricted isometry property of order S if $\exists \delta_S$ s.t. $\forall$ k-sparse f such that $k \leq S$, and f's support $T \subset \{1, 2, ..., n\}(|T| = k)$,

$$(1 - \delta_S)||f||_2^2 \leq ||\Phi_T f||_2^2 \leq (1 + \delta_S)||f||_2^2. \quad (2)$$

Sampling Constraints

Theorem (Equivalence of problem (P_0) and (P_1))

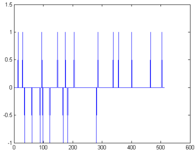
Suppose $S \geq 1$ and $\delta_{2S} < 1$, the solutions of (P_0) and (P_1) coincides if that solution f has its support T satisfying $|T| \leq S$.

Theorem (Noiseless incoherent sampling [CT06])

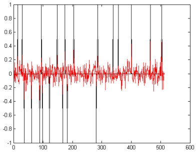
If f is k-sparse, then for any $\beta > 0$, with probability at least $1 - 5/n - e^{-\beta}$ the signal can be perfectly recovered if $m \gtrsim O(k \log n)$

- Random sensing matrix works better.

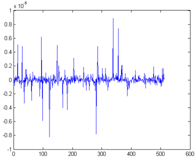
Simulations: solving BP by LP



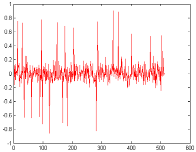
(d) recovery ($p = 1$)



(e) recovery ($p = 2$)



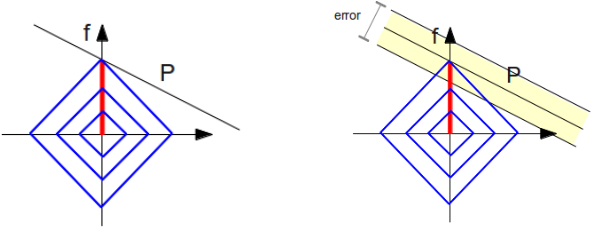
(f) error ($p = 1$)



(g) error ($p = 2$)

Figure: Recovery compared to the raw signal.

Noisy Recovery



(a) noiseless (b) noisy

- In practise $\mathbf{y} = \Phi f + \sigma \mathbf{z}$.
- LASSO: l_2-minimization with l_1-penalty.

$$f^* = \arg\min_f \frac{1}{2} ||\Phi f - \mathbf{y}||_2^2 + \lambda \sigma_m ||f||_1 \quad (3)$$

- Similar (weak) RIP guarantees the recovery.

Single-Pixel Camera

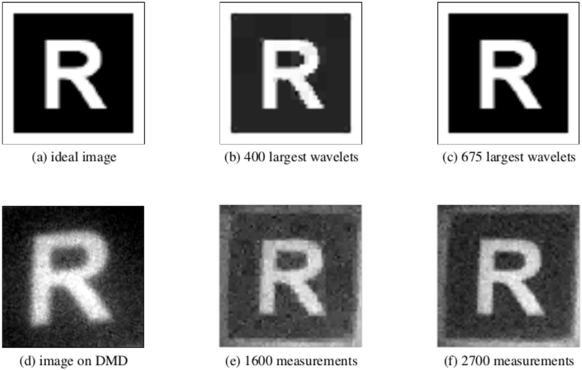


Figure: Compressed sensing v.s. wavelet decomposition [TLW+06].

CS-MRI

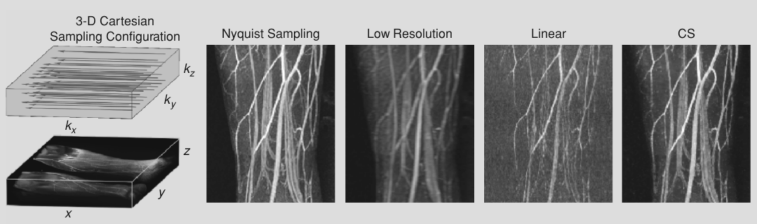


Figure: Applying CS techniques to MRI [LDSP08].

MAP

$$f^* = \arg\max_f (\log P(\mathbf{y}|f;\Phi) + \log P(f)). \tag{4}$$

- $\mathbf{y} - \Phi f = \mathbf{z} \sim \mathcal{N}(0, \sigma^2)$
- Prior $P(f) = \frac{1}{Z} e^{-\lambda_p ||f||_p}$

$$\begin{aligned} &\Rightarrow f = \arg\min_f ||\mathbf{y} - \Phi f||_2^2 + \lambda ||f||_p \\ &\Rightarrow f = \arg\min_f ||f||_p \ s.t. \ ||\mathbf{y} - \Phi f||_2 \leq \epsilon \end{aligned} \tag{5}$$

- $p = 1$: standard LASSO
- $\epsilon \to 0$: perfect recovery.

Dictionary Learning

- Finding the dictionary minimizing the error of representation:

$$\Psi^* = \arg\min_{\Psi} ||\mathbf{Y} - \Phi\Psi\mathbf{X}||_F^2 \tag{6}$$

- $\mathbf{X}$ is supposed to be sparse.
- K-SVD [AEB06].
- Expectation-Maximization.

Maximum Likelihood

- Signal can be represented by some fixed codes from an over-complete dictionary:

$$f = \Psi \mathbf{x} + \mathbf{v},\ \mathbf{v} \sim \mathcal{N}(0, \sigma'^2) \tag{7}$$

- Best dictionary gives sparse representations $||\mathbf{x}||_1 \leq \epsilon$:

$$\Psi^* = \arg\max_{\Psi} \sum_{f} \int_{||\mathbf{x}||_1 \leq \epsilon} e^{-\frac{1}{2\sigma'^2}||f - \Psi \mathbf{x}||_2^2 - \lambda_1 ||\mathbf{x}||_1} \tag{8}$$

- Trick: fast online methods.

Conclusion

- Revolutionary sensing theories without Nyquist rate constraints.
- Benefited from data sparsity.
- The RIP helps yield perfect recovery with high probability.
- Random sensing matrix works better.
- Learning algorithms benefits CS methods.

Future Research

- Can the algorithms keep high performance when n grows?
- Can $P(\Phi)$ be learned?
- Can we loose the restriction of (weak) RIP?
- Any other learning approaches?

Acknowledgements

- Thesis supervisor Prof. Haizhang Zhang.
- Prof. Guocan Feng, Dr. Lei Zhang and Mr. Zhihong Huang.
- Professors who have taught me or given me advices.
- My friends, my family, and myself.

[AEB06] Michal Aharon, Michael Elad, and Alfred Bruckstein.
K -svd : An algorithm for designing overcomplete dictionaries for sparse representation.
IEEE Transactions on Signal Processing, 54(11):4311–4322, 2006.

[Can08] E. J. Candès.
The restricted isometry property and its implications for compressed sensing.
Comptes Rendus Mathematique, pages 589 – 592, 2008.

[CDMS98] Scott Shaobing Chen, David L. Donoho, Michael, and A. Saunders.
Atomic decomposition by basis pursuit.
SIAM Journal on Scientific Computing, 20:33–61, 1998.

[CT05] E.J. Candès and T. Tao.
Decoding by linear programming.

IEEE Transactions on Information Theory, pages 4203–4215, 2005.

[CT06] E.J. Candès and T. Tao.
Near-optimal signal recovery from random projections: Universal encoding strategies?
IEEE Transactions on Information Theory, 52(12):5406–5425, 2006.

[DH99] David L. Donoho and Xiaoming Huo.
Uncertainty principles and ideal atomic decomposition, 1999.

[LDSP08] M. Lustig, D.L. Donoho, J.M. Santos, and J.M. Pauly.
Compressed sensing mri.
Signal Processing Magazine, IEEE, 25(2):72–82, 2008.

[TLW$^+$06] Dharmpal Takhar, Jason N. Laska, Michael B. Wakin, Marco F. Duarte, Dror Baron, Shriram Sarvotham, Kevin F. Kelly, and Richard G. Baraniuk.
A new compressive imaging camera architecture using optical-domain compression.
In *Proceedings of Computational Imaging IV at SPIE Electronic Imaging*, pages 43–52, 2006.

[Wei] Fred Weinhaus.
Fourier transforms.
http://www.imagemagick.org/Usage/fourier/.