Compressed Sensing Recovery: A Survey

Shuangjiang Li
Electrical Engineering and Computer Science
University of Tennessee, Knoxville, TN 37996
shuangjiang@utk.edu

I. RIP

Candes and Tao [1] introduced the following isometry condition on matrices Φ and established its important role in CS. Given a matrix $\Phi \in \mathbb{R}^{m \times n}$ and any set T of column indices, we denote by Φ_T the $m \times \#(T)$ (i.e., $m \times |T|$) matrix composed of these columns. Similarly, for a vector $x \in \mathbb{R}^n$, we denote by x_T the vector obtained by retaining only the entries in x corresponding to the column indices T. We say that a matrix Φ satisfies the *Restricted Isometry Property (RIP)* of order k if there exists a $\delta_k \in (0,1)$ such that

$$(1 - \delta_k)\|x_T\|_2^2 \leq \|\Phi_T x_T\|_2^2 \leq (1 + \delta_k)\|x\|_2^2 \tag{1}$$

holds for all sets T with $\#T \leq k$ (i.e., $|T| \leq k$). The condition (1) is equivalent to requiring that the Grammian matrix $\Phi_T^t \Phi_T$ has all of its eigenvalues in $[1 - \delta_k, 1 + \delta_k]$ (here Φ_T^t denotes the transpose of Φ_T).

Restricted Isometry Condition (RIC) is necessarily a local principle, which concerns not the measurement matrix Φ as a whole, but its submatrices of k columns. All such submatrices $\Phi_I, I \subset \{1, \cdots, n\}, |I| \leq k$, are almost isometries. Therefore, for every k-sparse signal x, the observation vector $u = \Phi^* \Phi v$ approximates v locally, when restricted to a set of cardinality k. The following proposition formalizes these local properties of Φ.

Proposition I.1. (Consequences of Restricted Isometry Condition [2]) *Assume a measurement matrix Φ satisfies the restricted isometry condition with parameters $(2k, \varepsilon)$. Then the following holds:*

1) *(Local approximation) For every k-sparse vector $v \in \mathbb{R}^n$ and and every set $I \subset \{1, \cdots, n\}, |I| \leq k$, the observation vector $u = \Phi^* \Phi v$ satisfies*

$$\|u|_I - v|_I\|_2 \leq 2.03\varepsilon\|v\|_2. \tag{2}$$

2) *(Spectral norm) For any vector $z \in \mathbb{R}^n$ and every set $I \subset \{1, \cdots, n\}, |I| \leq 2k$, we have*

$$\|(\Phi^* z)|_I\|_2 \leq (1 + \varepsilon)\|z\|_2. \tag{3}$$

3) *(Almost orthogonality of columns) Consider two disjoint sets $I, J \subset \{1, \cdots, n\}, |I \cup J| \leq 2k$. Let P_I, P_J denote the orthogonal projections in $\mathbb{R}^N$ onto range(Φ_I) and range(Φ_J), respectively. Then*

$$\|P_I P_J\|_{2 \to 2} \leq 2.2\varepsilon. \tag{4}$$

II. GREEDY

One popular class of sparse recovery algorithms is based on the idea of iterative greedy pursuit. The earliest one include the matching pursuit (MP) by G. Mallat, et al. [3], later advanced by Y. Pati, et al. [4] and G. Davis, et al. [5]. Matching pursuit is related to the field of compressed sensing and has been extended by researchers. Notable extensions are Orthogonal Matching Pursuit (OMP) [6], Stagewise OMP (StOMP) [7], and compressive sampling matching pursuit (CoSAMP).

Matching Pursuit was originally introduced in the signal-processing community as an algorithm *"that decomposes any signal into a linear expansion of waveforms that are selected from a*

Algorithm 1: Matching Pursuit

Input:
  ▶ Measurement matrix $\Phi \in \mathbb{R}^{m \times n}$.
  ▶ Observation vector $y \in \mathbb{R}^m$.
Output:
  ▶ An estimate $\hat{x} \in \mathbb{R}^n$ of the ideal signal x.

1: $\hat{x}_0 = 0$, $r^{(0)} \leftarrow y$, $i = 0$ ▷ Initialization
2: **while** *halting criterion false* **do**
3:      $i \leftarrow i + 1$
4:      $\phi_i \leftarrow \underset{\phi_i \in \Phi}{\mathrm{argmax}} |\langle r^{(i-1)}, \phi_i \rangle|$ ▷ The column of Φ that is most correlated with $r^{(i-1)}$
5:      $\hat{x}_i \leftarrow \langle r^{(i-1)}, \phi_i \rangle$ ▷ From residual new signal estimate
6:      $r^{(i)} = r^{(i-1)} - \phi_i \hat{x}_i$ ▷ Update residual
7: **end while**
8: **return** $\hat{x} \leftarrow \hat{x}_i$

Algorithm 2: Orthogonal Matching Pursuit

Input:
  ▶ Measurement matrix $\Phi \in \mathbb{R}^{m \times n}$.
  ▶ Observation vector $y \in \mathbb{R}^m$.
  ▶ Sparsity level k of the ideal signal $x \in \mathbb{R}^n$.
Output:
  ▶ An estimate $\hat{x} \in \mathbb{R}^n$ of the ideal signal x.
  ▶ A set Λ_k containing the positions of the non-zero elements of $\hat{x}$.
  ▶ An approximation to the measurements y by a_k.
  ▶ The residual $r = y - a_k$.

1: $r^{(0)} \leftarrow y$ ▷ Initialize the residual
2: $\Lambda^{(0)} \leftarrow \varnothing$ ▷ Initialize the indices
3: **for** $i = 1, \cdots, k$ **do**
4:      $\lambda^{(i)} \leftarrow \arg\max_{j=1,\cdots,n} |\langle r^{(i-1)}, \phi_j \rangle|$ ▷ The column of Φ that is most correlated with $r^{(i-1)}$
5:      $\Lambda^{(i)} \leftarrow \Lambda^{(i-1)} \cup \lambda^{(i)}$
6:      $\Phi^{(i)} \leftarrow [\Phi^{(i-1)} \phi_{\lambda^{(i)}}]$
7:      $x^{(i)} \leftarrow \arg\min_{\hat{x}} \|y - \Phi^{(i)} \hat{x}\|_2$ ▷ Solve the Least Squares for new signal estimate
8:      $a^{(i)} \leftarrow \Phi^{(i)} x^{(i)}$ ▷ New data approximation
9:      $r^{(i)} \leftarrow y - a^{(i)}$ ▷ New residual
10: **end for**
11: $\hat{x} \leftarrow x^{(k)}$
12: **return** $\hat{x}, \Lambda^{(k)}, a^{(k)}, r^{(k)}$

---

redundant dictionary of functions" [3]. It is a general, greedy, sparse function approximation scheme with squared error loss, which iteratively adds new functions (i.e., basis functions) to the linear expansion.

The essence of matching pursuit, Algorithm 1 is that, for a given vector x to be approximated, first choose the vector from the dictionary on which x has the longest projection. Then, remove any component of the form of the selected vector from x, i.e., orthogonalize x with respect to the selected dictionary vector, and obtain the residual of x. The selected dictionary vector is in fact the one that results in the residual of x with the smallest energy. Repeat this process for the residual of x with the rest of dictionary vectors until the norm of the residual becomes smaller than the threshold ε.

▷ (From the machine learning point of view [8])
For any function $f \in \mathcal{H}$ we will use $\mathbf{f}$ to represent the l-dimensional vector that corresponds to the evaluation of f on the l training points:

$$\mathbf{f} = (f(x_1), \cdots, f(x_l)). \tag{5}$$

- $\mathbf{y} = (y_1, \cdots, y_l)$ is the *target vector*.
- $\mathbf{r}_n = \mathbf{y} - \mathbf{f}_n$ is the *residue*.

It starts at stage 0 with $\mathbf{f}_0 = 0$, and recursively appends functions to an initially empty basis, at each stage n, trying to reduce the norm of the residue $\mathbf{r}_n = \mathbf{y} - \mathbf{f}_n$.

Given $\mathbf{f}_n$ we build,

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \alpha_{n+1}\mathbf{g}_{n+1} \tag{6}$$

by searching for $\mathbf{g}_{n+1} \in \mathcal{D}$ and for $\alpha_{n+1} \in \mathbb{R}$ that minimize the residual error, i.e., the squared norm of the next residue:

$$\begin{aligned}
\|\mathbf{r}_{n+1}\|^2 &= \|\mathbf{y} - \mathbf{f}_{n+1}\|^2 \\
&= \|\mathbf{y} - (\mathbf{f} + \alpha_{n+1}\mathbf{g}_{n+1})\|^2 \\
&= \|\mathbf{r}_n - \alpha_{n+1}\mathbf{g}_{n+1}\|^2.
\end{aligned}$$

Formally,

$$(g_{n+1}, \alpha_{n+1}) = \operatorname*{arg\,min}_{(g \in \mathcal{D}, \alpha \in \mathbb{R})} \|\mathbf{r}_n - \alpha\mathbf{g}\|^2$$

For any $g \in \mathcal{D}$, the α that minimizes $\|\mathbf{r}_n - \alpha\mathbf{g}\|^2$ is given by

$$\frac{\partial\|\mathbf{r}_n - \alpha\mathbf{g}\|^2}{\partial\alpha} = 0$$

$$-2\langle\mathbf{g}, \mathbf{r}_n\rangle + 2\alpha\|\mathbf{g}\|^2 = 0$$

$$\alpha = \frac{\langle\mathbf{g}, \mathbf{r}_n\rangle}{\|\mathbf{g}\|^2} \tag{7}$$

For this optimal value of α, we have

$$\|\mathbf{r}_n - \alpha\mathbf{g}\|^2 = \|\mathbf{r}_n - \frac{\langle\mathbf{g},\mathbf{r}_n\rangle}{\|\mathbf{g}\|^2}\mathbf{g}\|^2$$

$$= \|\mathbf{r}_n\|^2 - 2\frac{\langle\mathbf{g},\mathbf{r}_n\rangle}{\|\mathbf{g}\|^2}\langle\mathbf{g},\mathbf{r}_n\rangle + (\frac{\langle\mathbf{g},\mathbf{r}_n\rangle}{\|\mathbf{g}\|^2})^2\|\mathbf{g}\|^2$$

$$= \|\mathbf{r}_n\|^2 - (\frac{\langle\mathbf{g},\mathbf{r}_n\rangle}{\|\mathbf{g}\|})^2 \qquad (8)$$

So the $g \in \mathcal{D}$ that minimize expression (7) is the one that minimize (8), which corresponds to maximizing $\|\frac{\langle\mathbf{g},\mathbf{r}_n\rangle}{\|g\|}\|^2$. In other words, it is the function in the dictionary whose corresponding vector is "most collinear" with the current residue.

▷ End

In matching pursuit, after a vector in the dictionary is selected, one may remove any component of its form not only from x, but also from all other dictionary vectors before repeating the process. This version of the method is called orthogonal matching pursuit and is computationally more expensive than the nonorthogonal version, but typically gives significantly better results in the context of coding. The basic orthogonal matching pursuit algorithm, is shown in Algorithm 2.

Algorithm 3: Regularized Orthogonal Matching Pursuit

Input:
- ▶ Measurement matrix $\Phi \in \mathbb{R}^{m\times n}$.
- ▶ Observation vector $y \in \mathbb{R}^m$.
- ▶ Sparsity level k of the ideal signal $x \in \mathbb{R}^n$.

Output:
- ▶ Index set $I \subset \{1,\cdots,n\}$.
- ▶ Reconstructed vector $\hat{v} = x \in \mathbb{R}^n$.

1: $r^{(0)} \leftarrow y$ ▷ Initialize the residual
2: $I \leftarrow \varnothing$ ▷ Initialize the indices
3: Repeat until $|I| \geq 2k$
4:     $u \leftarrow \Phi^* r$
5:     $sort(u), J \leftarrow \min\{u(1:k),|u|\}$ ▷ Identify
6:     $|u(i)| \leq 2|u(j)|$ s.t. $\forall i,j \in J_0$ and $J_0 \subset J$, $J_0 \leftarrow \arg\min\|u|_{J_0}\|_2$ ▷ Regularize
7:     $I = I \cap J_0$
8:     $\hat{x} \leftarrow \arg\min_{x^*\in\mathbb{R}^n}\|y - \Phi x^*\|_2$
9:     $r^{(i)} \leftarrow y - \Phi\hat{x}$ ▷ Update residual
10: end
11: $\hat{x} \leftarrow x^{(k)}$
12: return $\hat{x}, \Lambda^{(k)}, a^{(k)}, r^{(k)}$

The reconstruction complexity of these algorithms (OMP, StOMP, ROMP) is around $\mathbb{O}(KMN)$, which is significantly lower than the BP methods. However, they require more measurements for perfect reconstruction and they lack provable reconstruction quality. More recently, greedy algorithms such as the subpace pursuit(SP) [9] and the compressive sampling matching pursuit (CoSaMP) [10] have been proposed by incorporating the idea of backtracking. They offer comparable theoretical reconstruction quality as that of the LP methods and low reconstruction

complexity. However, both the SP and the CoSAMP assume that the sparsity K is known, whereas K may not be available in many practical applications.

Algorithm *sparsity adaptive matching pursuit (SAMP)* [11], could recover signal without prior information of the sparsity. Which make it promising for many practical applications when the number of non-zero (significant) coefficients of a signal is not available.

REFERENCES

[1] E. Candes and T. Tao, "Decoding by linear programming," *Information Theory, IEEE Transactions on*, vol. 51, no. 12, pp. 4203–4215, 2005.

[2] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Foundations of computational mathematics*, vol. 9, no. 3, pp. 317–334, 2009.

[3] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *Signal Processing, IEEE Transactions on*, vol. 41, no. 12, pp. 3397–3415, 1993.

[4] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Signals, Systems and Computers, 1993. 1993 Conference Record of The Twenty-Seventh Asilomar Conference on*.   IEEE, 1993, pp. 40–44.

[5] G. Davis, S. Mallat, and Z. Zhang, "Adaptive time-frequency decompositions with matching pursuits," *Optical Engineering*, vol. 33, 1994.

[6] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655–4666, 2007.

[7] D. Donoho, Y. Tsaig, I. Drori, and J. Starck, "Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 58, no. 2, pp. 1094–1121, 2012.

[8] V. Pascal and B. Yoshua, "Kernel matching pursuit," *Machine Learning*, vol. 48, no. 3, pp. 165–187, 2002.

[9] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *Information Theory, IEEE Transactions on*, vol. 55, no. 5, pp. 2230–2249, 2009.

[10] D. Needell and J. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[11] T. Do, L. Gan, N. Nguyen, and T. Tran, "Sparsity adaptive matching pursuit algorithm for practical compressed sensing," in *Signals, Systems and Computers, 2008 42nd Asilomar Conference on*.   IEEE, 2008, pp. 581–587.