A Match Pursuit Based Method Adapted to Overcomplete Dictionary for Compressive Spectral Imaging

Jianchen Zhu* Shengjie Zhao* and Rongqing Zhang*
* Tongji University, Shanghai, China
E-mail: 1610486@tongji.edu.cn, shengjiezhao@tongji.edu.cn, rongqingz@tongji.edu.cn

***Abstract*—This paper proposes to introduce greedy pursuit, a conceptually simple hard thresholding pursuit algorithm called Signal Space Subspace Pursuit (SSSP) for calculating spare signal representations with overcomplete dictionaries whenever the sensing matrix (sampling operator) satisfies the Restricted Isometry Property adapted to Dictionary (D-RIP). Signal space greedy method has the ability to optimally compute (near) best projections that allow one to identify the most related a small number of dictionary atoms of an arbitrary signal in this setting. More practically, standard recovery algorithms are applicable to such projections while maintaining accurate signal recovery. Simulation results with a typical hyperspectral data set demonstrate the superiority of the proposed approach.**

I. Introduction

Compressive sensing (CS) [1], [2] has emergenced as a novel paradigm to recovery signals, which are assumed to be sparse in a certain domain, from a small set of compressed measurements. Until now, applications of CS are widespread and range from signal processing applications [3], [4], which includes image [5], [6], audio [7], video [8], and so forth.

CS claims that a k sparse signal $x \in R^n$ is sampled by using a small amount of m compressive measurements, or in matrix notation

$$y = Ax + e, \tag{1}$$

where $y \in R^m$ denotes the measurement vector, $A \in R^{m \times n}$ stands for the sensing matrix with $m \ll n$ whose entries are random Gaussian variables, and e is the measurement error vector due to the real signal architecture. Assume that x is sparse in some basis such that $x = \psi a$ with a has at most k nonzero entries. CS freamework indicates that such unknown signal can be recovered correctly provided by only $m = O(klogn)$ using y and A, which results in low computational complexity and cost reduction.

The recovery of seeking any sparse signal from linear measurements can be considered a highly non-linear process obtained by the sampling process. A nature approach to mathematically capture the sparse solution by solving the following l_1-minimization problem

$$\hat{x} = \arg\min \|x\|_1 \ s.t. \ \|y - Ax\|_2 \leq \varepsilon, \tag{2}$$

where $\|x\|_1 = \sum |x_i|$ refers to the l_1-norm, $\|\cdot\|_2$ is the norm of standard Enclidean, and $\varepsilon > 0$ is decided by the upper bound of $\|e\|_2$. The restricted isometry property (RIP) provides a sufficient condition for accurate recovery of a signal x with not entirely sparse but nearly sparse representation from the observation y via problem (2).

***Definition* 1 [9]:** The sensing matrix $A \in R^{m \times n}$ follows the RIP of order k if the smallest restricted isometry constant (RIC) $\delta_k \in (0, 1)$

$$(1 - \delta_k) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_k) \|x\|_2^2, \tag{3}$$

holds for all k sparse signals. It is well known that many random matrices have very small RICs with high probability when m is large reasonably.

Assuming that A describes the RIP, the problem (2) is ensured to stability and robustness recover it from the noisy measurements $y = Ax + e$. Then the solution $\hat{x}$ of (2) follows

$$\begin{aligned} &\|\hat{x} - x\|_2 \leq c_1 \frac{\|x - x_k\|_1}{\sqrt{k}} + c_2 \varepsilon, \\ &\text{with,} \\ &c_1 = \frac{2 + (2\sqrt{2} - 2)\delta_{2k}}{1 - (\sqrt{2} + 1)\delta_{2k}}, \ c_2 = \frac{4\sqrt{1 + \delta_{2k}}}{1 - (\sqrt{2} + 1)\delta_{2k}}, \end{aligned} \tag{4}$$

where c_1, c_2 are positive constants, x_k is an approximation to x with k largest nonzero entries with δ_{2k} is the order $2k$ RIC of A. Eq. (4) shows that the resulting recovery error is proportional to the noise level and the signal tail, which means that the coefficients of the compressible signals follows a power law decay such that $\hat{x}$ is said properly to be approximately x. Based on this concept, several methods can address the NP-hard problem and computationally infeasible with signal dimension increasing, which depended on A and y, such as linear programming [1], convex [10], gradient [11], and greedy [12], [13]. Among them, two families of approaches are basis pursuit (BP) based methods and matching pursuit (MP) based method.

However, a real signal is not exactly sparse in an orthonormal [14]–[16]. Specifically, these results do not hold when ψ is an orthonormal basis, but hold for an overcomplete dictionary, which means that the signal $f \in R^n$ can now be expressed as $f = Da$, where $D \in R^{n \times d}$ $(d > n)$ is a given overcomplete dictionary called as sparsifying basis with $a \in R^d$ being the sparse coefficient vector. Therefore, this paper aims to propose a variant of MP, which is known as Signal Space Subspace Pursuit (SSSP). SSSP is a hard thresholding pursuit algorithm nearly likes MP, which minimizes the l_1-norm of the sparse coefficient vector via constructing the representation by selecting atoms from an arbitrary overcomplete dictionary instead of an orthonormal basis per iteration. Ultimately, one can allows us to thoroughly understand that these sample vectors derived from the sparsity-dictionary signals can be recovered from a small set of linear measurements, which is depended on the information level and noise level. Assume that the sampling operator/sensing matrix satisfies the D-RIP that can be seen as the extension of the standard RIP in compressive sensing, two relative simple families of methods, i.e., BP based methods and MP based [17]–[19], can calculate the required near optimal projections to approximate the optimal projections subject to a given iteration number constraint.

II. Problem Formulation

We consider a sparse-dictionary signal recovery problem from a set of noisy measurements like the framework in [20]. Let us denote $x \in R^n$ is the unknown signal vector. Assume that the unknown signal vector is k sparse, where k is sparsity level and known a priori, x can be sparsely expressed related to the linear combination of an overcomplete dictionary

$$x = Da, \tag{5}$$

where $D \in R^{n\times d}$ can be regarded as the sparsifying basis, e.g., an overcomplete dictionary, which is assumed to be a tight frame such that $DD^{\mathrm{T}} = I_n$ with T being the transpose operator. Analogous to the traditional CS, the sampling of each component of the input signal x is realized by using a linear operator A. One can obtain the following random linear projection in matrix notation as

$$y = Ax + e. \tag{6}$$

To the best of our knowledge, it has been shown that the broadly used D-RIP as a useful tool provides one of the most important conditions to alternatively estimate the quality of a sensing matrix for the exactly recovery of a not entirely sparse but nearly sparse signal x corrupted by additional noise from its observations y, which can be regarded as the natural extension of the standard RIP.

***Definition* 2:** Let $D \in R^{n\times d}$ be an overcomplete dictionary and a tight frame. We say that a sensing matrix A follows the restricted isometry property adapted to D (abbreviated D-RIP) with the smallest constant δ_k such that

$$(1-\delta_k)\left\|Da\right\|_2^2 \le \left\|ADa\right\|_2^2 \le (1+\delta_k)\left\|Da\right\|_2^2 \tag{7}$$

holds for all k sparse signals with $\|a\|_0 \le k$. Eq. (7) claims that nearly all the random matrices whose entries are drawn from the Gaussian, subgaussian or Bernoulli distribution with measurement number on the order of $klog(d/k)$ satisfy the D-RIP with very high probability, which exploits the nearly mutually orthogonal of the columns of a sensing matrix.

Assume now that A and D exhibit the D-RIP. Another novel way based on the property of highly correlated of the overcomplete dictionary D for recovering a sparse-dictionary signal f from its noisy measurements $y = ADa + e$ under a stronger assumption that f is k-analysis-sparse, i.e., $D^{\mathrm{T}}x \in R^d$ is k-sparse, is to acquire the solution via the relaxed l_1-minimization problem, which states

$$\hat{x} = \underset{\tilde{x}\in R^n}{argmin}\left\|D^{\mathrm{T}}\tilde{x}\right\|_1 \;\; s.t.\|y - A\tilde{x}\|_2 \le \varepsilon, \tag{8}$$

where again $\varepsilon > 0$ is likely an upper bound that is proportional to the noise level $\|e\|_2$.

Then the solution of (9) follows

$$\|\hat{x} - x\|_2 \le c_1\frac{\left\|D^{\mathrm{T}}x - (D^{\mathrm{T}}x)_k\right\|_1}{\sqrt{k}} + c_2\varepsilon, \tag{9}$$

where some numerical constants $c_1 > 0$ and $c_2 > 0$ may only rely on δ_{2k}, and $(D^{\mathrm{T}}x)_k$ denotes the best approximation of $D^{\mathrm{T}}x$ within all the k-largest nonzero entries in l_1-norm, which is analogous to the soulution of (4). Thus, the recovery error $\|\hat{x}-x\|_2$ is depends on the noise level $\|e\|_2$ and the "tail" of the analysis vector $\frac{\left\|D^{\mathrm{T}}x-(D^{\mathrm{T}}x)_k\right\|_1}{\sqrt{k}}$, which results in rapidly decays of the error caused by a fixed iteration number. As these results hold even when the columns of a particular dictionary are highly correlated, the general setting related to this paper allows us to mathematically deduce the results similar to MP based methods with the introduction of previous restrictions. Without loss of generality, the convergence for the algorithm is similar to that for the l_1-analysis methods.

III. Signal Space Subspace Pursuit Algorithm

Our proposed algorithm as a adaptation of the MP based methods recover the signal x using an overcomplete dictionary D from its compressive measurements in the noisy case $y = Ax + e$, which is listed below.

TABLE I: SIGNAL SPACE SUBSPACE PURSUIT ALGORITHM

Algorithm: Signal Space Subspace Pursuit
Input: A, D, y, k, stopping criterion ε
Initialization: $l = 0$, $I = \emptyset$; $r^0 = y$, $x^0 = 0$
while halting criterion is not satisfied ***do*** **1**: identify the index set $\Omega = supp_{(D,2k)}(A^{\mathrm{T}}r)$ **2**: find the support estimation $T = \Omega \cup I$ **3**: compute the approximation: $\tilde{x} = \arg\min_z \|y - Az\|_2$ s.t. $z \in \mathbf{R}(D_T)$ **4**: shrink the index $I = supp(\tilde{x}, k)$ ={index corresponding to the largest magnitude entries of estimation $\tilde{x}$} **5**: compute the new signal estimation: $x^{l+1} = P_I\tilde{x}$ **6**: compute the latest residual: $r^{l+1} = y - Ax^{l+1}$ ***end while***: $l = MaxIter$ or $\left\|x^{l+1} - x^l\right\|_2/\left\|x^l\right\|_2 \le \varepsilon$ is satisfied
Output: $\hat{x} = x^{l+1} = SSSP(A, D, y, k)$

The overall procedure for the modified algorithm in term of the MP based methods is consisted of six steps. The first step is called identifying, the second step is called the merging, and the third step is called updating integrated from step 3 to step 6. In the first step, the algorithm is used to identify the support estimation Ω by iteratively introducing multiple indices with $2k$ largest related values in magnitude provided by some inputs A, D and y, and the initialization $x^0 = 0$. In the third step, the cardinality of the updated support set I corresponding to the sparsity level k is updated using shrinking obtained by merging in the second step, which leads to a relatively sparse solution based on the trimmed set and a least square solution. Finally, the stopping criteria selection plays an important role for the algorithm.

Similar to the classical MP based methods, Signal space subspace pursuit (SSSP) is a hard thresholding pursuit algorithm that greedily finds the support estimation, i.e., index set, performs such projection, and constructs the sparse representation, which is provided by a finite iteration number. The main idea behind the proposed algorithm depends on the correct selection of the support estimation subject to an overcomplete dictionary constraint during each iteration. As depicted in Algorithm 1, the proposed sparse recovery algorithm is somewhat different from the traditional MP based methods via one of the most crucial steps involved by replacing simple hard thresholding to calculate the projection the, i.e., project x in signal space onto the set of k sparse signals over the dictionary D, which is given by

$$\Lambda_{opt}(x,k) = \underset{\Lambda:\|\Lambda\|_0 \le k}{\arg\min} \|x - P_\Lambda x\|_2, \tag{10}$$

where P_Λ is simply a hard thresholding operator applied to the optimal projection of a general vector $x \in R^n$ onto the columns of D subject to the index set Λ constraint. In particular, the calculation of such projection is obtained by utilizing the proxy $A^T r$ if it has at most k nonzero entries, which provides a probably guarantee to approximate the desired sparse solution.

Recall that in (10) a vector x has a sparse representation in terms of D, which is caused by minimizing $\|x - P_\Lambda x\|_2$. Assume that calculation of the optimal projection $P_\Lambda x$ is generally unworkable and nontrivial provided by any A, D and r, one alternative way allows us to apply near-optimal projection to capture a near approximation of $P_\Lambda x$ such that

$$\begin{aligned} \left\| P_{supp_{(D,k)}(A^{\mathrm{T}}r)} x - x \right\|_2 &\le c_1 \|P_\Lambda x - x\|_2 \\ \left\| P_{supp_{(D,k)}(A^{\mathrm{T}}r)} x - x \right\|_2 &\le c_2 \|P_\Lambda x\|_2 \end{aligned} \tag{11}$$

for some constants $c_1 \ge$ and $c_2 \ge 0$. Notably, near-optimal projection is equal to optimal projection even when the columns of D are highly correlative under the assumption that $c_1 = 0$ or $c_2 = 0$, i.e., $P_{supp_{(D,k)}(A^{\mathrm{T}}r)} x = P_\Lambda x$, which leads to accurate signal recovery. We then consider the least-squares problem based on the calculation of $P_{supp_{(D,k)}(A^{\mathrm{T}}r)} x$ instead of that of $P_\Lambda x$. Specifically, the algorithm establishes the signal estimation $\tilde{x}$ via the dominant least squares problem of the merging step provided by T such that

$$\hat{x} = \underset{z}{argmin} \|y - Az\|_2 \;\; s.t. \; z \in \mathbf{R}(D_T) \tag{12}$$

with $2k$ largest nonzero entries of $D^{\mathrm{T}}(A^{\mathrm{T}}r)$. Similarly, one can obtain the solution of (16) via calculating

$$\hat{x} = D_T \tilde{A}_T^\dagger y = D_T((\tilde{A}_T^{\mathrm{T}} \tilde{A}_T)^{-1} \tilde{A}_T^{\mathrm{T}} y), \tag{13}$$

where A_T is the submatrix of A indexed by T. More practically, numerous empirical researches verify that the above-mentioned standard CS recovery algorithms can be used to calculate such near-optimal projection.

IV. Simulation Results

In this experiment, The cameraman hyperspectral images have a spatial resolution of 256×256 pixels with $L = 24$ bands ranging from 450 nm to 700 nm at steps of 10 nm. Fig. 1 depicts an RGB profile of the test datacube decomposed by the spectral bands. Each spatial information of the images is broken down into 1024 nonoverlapping of size 8×8 pixel patches, which forms the columns of the signal to be reconstructed. We use the proposed algorithm to recover the underlying datacube from the CASSI measurements corrupted by the sensing noise with zero-mean and variance $\delta^2 = 10^{-4}$. The standard baseline methods act as the benchmark to recover the images due to their reasonable computation complexity. The peak signal-to-noise ratio (PSNR) is used to evaluate the recovered image quality, which is given by $PSNR = 10log_{10}\frac{\|\hat{x}\|_2}{\|x-\hat{x}\|_2}$, where x is the original image and $\hat{x}$ is the recovered image in this case.

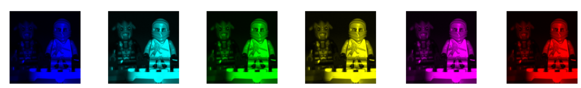

Fig. 1: A RGB decomposed representation of the original bands for the test data set used in simulations.

Fig. 2 and Table II show higher quality of recovered bands obtained from the proposed algorithm provided by the structure of H using a random coded aperture in term of PSNR compared to the baselined methods. Notice that all the entries lie in a random code are following subgaussian distribution. The results visually show the recovered bands with the proposed algorithm are superior. Otherwise, these baseline methods get poor results because they fail to consider the near-optimal projection scheme. Fig.3 further provides a visual comparison of recovered quality of the whole datacube to further proof the validity of the proposed algorithm.

TABLE II: AVERAGE PSNR FOR THE RECOVERED BANDS OF MULTISPECTRAL SCENE IN dB

	PSNR(dB)					
Metohds	**Band 1**	**Band 2**	**Band 3**	**Band 4**	**Band 5**	**Band 6**
OMP	14.25	14.32	14.24	14.93	14.21	14.57
SP	16.32	16.58	16.43	16.59	16.42	16.58
CoSaMP	17.33	17.46	17.44	17.52	17.39	17.55
LP	27.33	27.31	27.75	27.81	27.77	27.41
SSSP	27.32	27.55	27.83	27.41	27.92	27.55

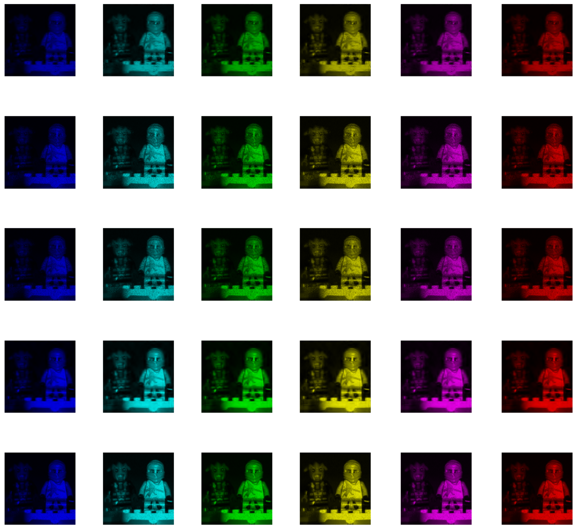

Fig. 2: Simulation results show the reconstruction of a subset of bands for the hyperspectral images with different algorithms: OMP (first row), (2) SP (second row), (3) CoSaMP (third row), (4) LP (fourth row), and (5) SSSP (fifth row).

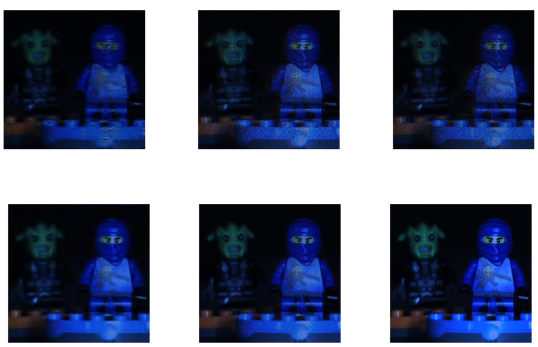

Fig. 3: Simulation results demonstrate the reconstruction of a RGB representation of the hyperspectral dataset using different algorithms: OMP (first one), (2) SP (second one), (3) CoSaMP (third one), (4) LP (fourth one), and (5) SSSP (fifth one) compared to the original spectral datacube (last one).

V. Conclusions

This paper presents a near-optimal projection scheme realized by some existing MP and BP based methods to solve the sparse signal recovery problem constraint to an overcomplete dictionary. Inspired by this scheme, we thus develop a fast hard threshold pursuit algorithm adapted to the dictionary algorithm to iteratively identify the atoms from the dictionary provided by the correct support set, which leads to a similar recovery guarantees. Under specific assumption on the signal structure, we verify that the signal space method is alternatively used to approximate the optimal projection such that providing theoretical backing to clearly explain the observed phenomena.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 61936014, and in part by the National Key Research and Development Project under Grants 2019YFB2102300 and 2019YFB2102301, and was supported by the Fundamental Research Funds for the Central Universities.

References

[1] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE transactions on information theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[2] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] J. Zhu, S. Zhao, X. Ma, and G. R. Arce, "On theoretical optimization of the sensing matrix for sparse-dictionary signal recovery," in *2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2019, pp. 1–5.

[4] R. M. Castro and E. Tanczos, "Adaptive compressed sensing for estimation of structured sparse sets," *arXiv: Statistics Theory*, 2014.

[5] H. Zhang, X. Ma, D. L. Lau, J. Zhu, and G. R. Arce, "Compressive spectral imaging based on hexagonal blue noise coded apertures," *IEEE Transactions on Computational Imaging*, pp. 1–1, 2020.

[6] J. Zhu, T. Zhang, and S. Zhao, "Low-rank subspace representation from optimal coded-aperture for unsupervised classification of hypersepctral imagery," in *2020 IEEE International Conference on Image Processing (ICIP)*, 2020, pp. 2860–2864.

[7] H. Xu, N. Fu, L. Qiao, W. Yu, and X. Peng, "Compressive blind mixing matrix estimation of audio signals," *IEEE Transactions on Instrumentation and Measurement*, vol. 63, no. 5, pp. 1253–1261, 2014.

[8] Kareth, M., Len-Lpez, Laura, V., Galvis, Carreo, Henry, Arguello, and Fuentes, "Temporal colored coded aperture design in compressive spectral video sensing," *IEEE Transactions on Image Processing*, 2018.

[9] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[10] P. A. Randall, *Sparse recovery via convex optimization*. California Institute of Technology, 2009.

[11] T. Blumensath and M. E. Davies, "Gradient pursuits," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2370–2382, 2008.

[12] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. part i: Greedy pursuit," *Signal Processing*, vol. 86, no. 3, pp. 572–588, 2006.

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

[14] D. L. Donoho, "Sparse components of images and optimal atomic decompositions," *Constructive Approximation*, vol. 17, no. 3, pp. 353–382, 2001.

[15] A. C. Gilbert and J. A. Tropp, "Applications of sparse approximation in communications," in *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on*, 2005, pp. 1000–1004.

[16] B. D. Rao, "Signal processing with the sparseness constraint," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 1998, pp. 1861–1864 vol.3.

[17] R. Giryes and D. Needell, "Greedy signal space methods for incoherence and beyond," *Applied and Computational Harmonic Analysis*, vol. 39, no. 1, pp. 1–20, 2015.

[18] X. GU and S. TU, "On practical approximate projection schemes in signal space methods," 2016.

[19] M. A. Davenport, D. Needell, and M. B. Wakin, "Signal space cosamp for sparse recovery with redundant dictionaries," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6820–6829, 2013.

[20] R. Zhang and S. Li, "Optimal d-rip bounds in compressed sensing," *Acta Mathematica Sinica, English Series*, vol. 31, no. 5, pp. 755–766, 2015.