SCIENCE CHINA
Information Sciences

• LETTER •

September 2021, Vol. 64 199101:1–199101:3
https://doi.org/10.1007/s11432-019-9945-2

Semi-blind compressed sensing via adaptive dictionary learning and one-pass online extension

Di MA & Songcan CHEN*

College of Computer Science and Technology, MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, *China*

Received 25 January 2019/Revised 29 March 2019/Accepted 22 June 2019/Published online 17 July 2020

Citation Ma D, Chen S C. Semi-blind compressed sensing via adaptive dictionary learning and one-pass online extension. Sci China Inf Sci, 2021, 64(9): 199101, https://doi.org/10.1007/s11432-019-9945-2

Dear editor,

Compressed sensing (CS) [1], as an efficient data acquisition paradigm, has attracted much attentions since it came up. The fundamental principle of CS is that a signal, which is sparse under some sparsity basis, can be efficiently acquired and accurately recovered via far fewer measurements than the traditional Shannon-Nyquist sampling. Since the number of measurements is much smaller than the number of signal coefficients, there potentially exist multiple solutions for the unknown signal, leading to so-called ill-posedness. To solve this problem, CS adopts a common mechanism that regularizes the recovery problem by constraining the original signal to be sparse under some sparsity basis [2]. Thus, the prior knowledge about the sparsity basis is critical for CS recovery. The often-used sparsity bases include the discrete cosine transform (DCT) and discrete wavelet transform (DWT). Although CS achieves high efficiency under these sparsity bases, these sparsity bases are data-independent, thus not optimal for CS recovery at hand [3].

For the above reason, Ref. [4] introduces the concept of blind compressed sensing (BCS), which replaces the requirement for a prior sparsity basis in the recovery with an assumption that the data is sparse under some existing yet unknown sparsity basis. BCS aims to simultaneously learn the sparsity basis and the corresponding sparse representation, which is similar to dictionary learning [5]. Once BCS succeeds, it will obtain a more appropriate, task-dependent sparsity basis than the task-independent one, thereby a more accurate recovery of the data. However, BCS suffers from much higher computational cost than CS due to the additionally learned sparsity basis. What's more, the theoretical analysis [6] has shown that BCS needs sufficient number of data samples to ensure a successful recovery, which is not required in CS at all.

In order to take respective advantages of CS and BCS while overcome their disadvantages, we combine CS with BCS and propose the concept of semi-blind compressed sensing (S-BCS). An assumption of S-BCS is that we are given some prior about the sparsity basis (as in CS), and use the prior to learn a more appropriate sparsity basis for the task at hand (as in BCS). To this end, we propose a sparse dictionary learning based method that adaptively transfers the prior sparsity basis to the given task in the process of recovery. Specifically, we utilize the current learned dictionary (starting from the prior sparsity basis) to obtain an intermediate recovery of the data and the corresponding sparse representation, and then in turn update the dictionary using the intermediate recovered data and the sparse representation. In order to incorporate prior sparsity basis into learning process, instead of directly updating the whole dictionary as in BCS, we consider the way of incremental learning [7] that gradually adapts the prior sparsity basis to the given task. Figure 1 illustrates the transition from the prior sparsity basis $\boldsymbol{D}_0$ to the task-dependent sparsity basis $\boldsymbol{D}_T$.

In summary, the contributions of this study are highlighted as follows:

(1) A novel concept, termed as semi-blind compressed sensing, is proposed. Under this concept, CS and BCS are combined to exert respective advantages while overcome their disadvantages.

(2) A sparse dictionary learning based method is proposed to adaptively transfer the prior sparsity basis to the task-dependent one. An incremental learning strategy is employed to keep smooth transition of the dictionary rather than abrupt updating.

(3) Compared with CS, our method inherits the advantage of BCS that learns a task-dependent dictionary to obtain better recovery.

(4) Compared with BCS, our method utilizes the prior knowledge about the sparsity basis that ensures an acceptable recovery performance even with a small number of data samples. Thus our method can be adjusted to address online tasks with mini-batch. Besides, our method enjoys much lower computational complexity than BCS due to the incremental learning strategy.

(5) Experimental results demonstrate that our method performs better than conventional CS algorithms regardless of the number of data samples. As for comparison with

* Corresponding author (email: s.chen@nuaa.edu.cn)

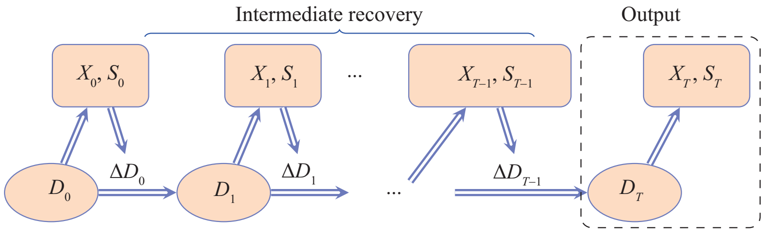

Figure 1 (Color online) Illustration of the transition process from the prior sparsity basis $\boldsymbol{D}_0$ to the task-dependent sparsity basis $\boldsymbol{D}_T$. $\{\Delta\boldsymbol{D}_t\}_{t=0}^{T-1}$ characterize the gradual transition between the dictionaries.

BCS, our method undoubtedly performs better than BCS for small number of data samples whereas gains comparable performance for sufficient data samples.

(6) Further visualization results on image data show the interpretability of S-BCS that it prefers to transfer low frequency counterparts of the prior sparsity basis to specific tasks while preserves high frequency counterparts almost unchanged. This phenomenon reveals that the dictionary learned by S-BCS comprises both the frequency characteristic of images and task-dependent information for specific tasks, whereas those used/learned by CS and BCS only comprise one of them. Consequently, the dictionary learned by S-BCS indeed builds a bridge at middle of the 'spectrum', where the two ends of the 'spectrum' are the dictionary used/learned by CS and BCS, respectively.

The proposed method. Our method consists of two main parts: (1) learning intermediate recovery of the data and corresponding sparse representations; (2) learning task-dependent dictionary.

Given the data samples $\boldsymbol{x}_i \in \mathbb{R}^d$ and the measurement matrices $\boldsymbol{\Phi}_i \in \mathbb{R}^{p\times d}$, $i = 1,\dots,n$ (generally $p < d$), the observed compressive measurements are expressed as

$$\boldsymbol{y}_i = \boldsymbol{\Phi}_i\boldsymbol{x}_i + \boldsymbol{\epsilon}_i, \quad i = 1,\dots,n, \tag{1}$$

where $\boldsymbol{\epsilon}_i \in \mathbb{R}^p$ represents the acquisition noise independent of $\boldsymbol{x}_i$. According to Eq. (1), we have the general error term as given in CS

$$\sum_{i=1}^{n} \|\boldsymbol{y}_i - \boldsymbol{\Phi}_i\boldsymbol{x}_i\|_2^2. \tag{2}$$

Now when knowing the current learned dictionary $\boldsymbol{D}_t \in \mathbb{R}^{d\times r}$, we can also obtain an expression of $\boldsymbol{X} = [\boldsymbol{x}_1,\dots,\boldsymbol{x}_n] \in \mathbb{R}^{d\times n}$ with respect to sparse representation $\boldsymbol{S} = [\boldsymbol{s}_1,\dots,\boldsymbol{s}_n] \in \mathbb{R}^{r\times n}$ as

$$\boldsymbol{X} = \boldsymbol{D}_t\boldsymbol{S} + \boldsymbol{\epsilon}_{\boldsymbol{X}}, \tag{3}$$

where $\boldsymbol{\epsilon}_{\boldsymbol{X}}$ is the residual term. According to Eq. (3), we have the penalty on the representation error in terms of $\boldsymbol{D}_t$ as

$$\|\boldsymbol{X} - \boldsymbol{D}_t\boldsymbol{S}\|_F^2. \tag{4}$$

By combining Eqs. (2), (4) and adding ℓ_1-norm regularization term on $\boldsymbol{S}$ to constrain its sparsity, we obtain the following optimization problem:

$$\begin{aligned}\{\boldsymbol{X}_t, \boldsymbol{S}_t\} = \arg\min_{\boldsymbol{X},\boldsymbol{S}} \frac{1}{2}\sum_{i=1}^{n}\|\boldsymbol{y}_i - \boldsymbol{\Phi}_i\boldsymbol{x}_i\|_2^2 \\ + \frac{\lambda}{2}\|\boldsymbol{X} - \boldsymbol{D}_t\boldsymbol{S}\|_F^2 + \alpha\|\boldsymbol{S}\|_1,\end{aligned} \tag{5}$$

where λ is a trade-off parameter to balance the two error terms, α is the parameter to control the sparsity constraint on the sparse representation. This optimization problem is not jointly convex in $\{\boldsymbol{X}, \boldsymbol{S}\}$ but biconvex and thus can be solved by alternatively updating $\boldsymbol{X}$ and $\boldsymbol{S}$. When $\boldsymbol{S}$ is fixed, the subproblem about $\boldsymbol{X}$ has closed form solution for each $\boldsymbol{x}_i$ as

$$\boldsymbol{x}_i = (\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{\Phi}_i + \lambda\boldsymbol{I})^{-1}(\boldsymbol{\Phi}_i^{\mathrm{T}}\boldsymbol{y}_i + \lambda\boldsymbol{D}_t\boldsymbol{s}_i). \tag{6}$$

When $\boldsymbol{X}$ is fixed, the subproblem about each $\boldsymbol{s}_i$ can be simply rewritten as

$$\min_{\boldsymbol{s}} \frac{\lambda}{2}\|\boldsymbol{x}_i - \boldsymbol{D}_t\boldsymbol{s}\|_2^2 + \alpha\|\boldsymbol{s}\|_1, \tag{7}$$

which is the classical ℓ_1-norm minimization problem and can be solved via various ℓ_1-norm minimization algorithms. In this study, we choose FIS-TA [8] as the ℓ_1 solver and the solving process is summarized as Algorithm A1 in Appendix A.1. The convergence of the iterative procedure between $\boldsymbol{X}$ and $\boldsymbol{S}$ can be mathematically guaranteed [9].

After getting the intermediate recovery $\boldsymbol{X}_t$ and the sparse representation $\boldsymbol{S}_t$, we learn the next dictionary $\boldsymbol{D}_{t+1}$. Specifically, we learn the gradual transition $\Delta\boldsymbol{D}_t$ between $\boldsymbol{D}_t$ and $\boldsymbol{D}_{t+1}$ by minimizing the representation error,

$$\begin{aligned}\Delta\boldsymbol{D}_t &= \arg\min_{\Delta\boldsymbol{D}} \frac{1}{n}\|\boldsymbol{X}_t - \boldsymbol{D}_t\boldsymbol{S}_t - \Delta\boldsymbol{D}\boldsymbol{S}_t\|_F^2 + \beta\|\Delta\boldsymbol{D}\|_F^2 \\ &= \frac{1}{n}(\boldsymbol{X}_t - \boldsymbol{D}_t\boldsymbol{S}_t)\boldsymbol{S}_t^{\mathrm{T}}\left(\frac{1}{n}\boldsymbol{S}_t\boldsymbol{S}_t^{\mathrm{T}} + \beta\boldsymbol{I}\right)^{-1},\end{aligned} \tag{8}$$

where the first term ensures that the gradual transition can further decrease the representation error of the intermediate recovery $\boldsymbol{X}_t$, the extra regularization term $\beta\|\Delta\boldsymbol{D}\|_F^2$ is used to prevent abrupt change between two adjacent dictionaries. Once we have $\Delta\boldsymbol{D}_t$, we update the next dictionary as

$$\boldsymbol{D}_{t+1} = \boldsymbol{D}_t + \Delta\boldsymbol{D}_t. \tag{9}$$

The whole process of S-BCS is summarized as Algorithm A2 in Appendix A.2.

We give the one-pass online extension of our method in Appendix B. The experimental results are shown in Appendix C.

Conclusion. We propose a concept of semi-blind compressed sensing and develop a sparse dictionary learning based algorithm that adaptively transfers a prior, task-independent sparsity basis to a task-dependent one. We have shown that by incorporating prior sparsity basis into BCS framework, the recovery quality can be further improved while the requirement for the number of samples can be reduced. Benefiting from the satisfactory performance of the proposed method on small number of samples, we next extend our method to one-pass online version that only needs to record a small batch of sequential samples, which further reduces the storage cost after data compression. Besides, we show that by adaptively transferring the prior sparsity basis to specific tasks, the learned dictionary enjoys good interpretability and adaptability. This framework can be generalized to other algorithms that require sparsity basis. For example, Bayesian model based compressed sensing approaches.

Acknowledgements This work was supported by Key Program of National Natural Science Foundation of China (Grant No. 61732006).

Supporting information Appendixes A–C. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Donoho D L. Compressed sensing. IEEE Trans Inform Theory, 2006, 52: 1289–1306
2 Zhao J, Bai X, Tao R. Improved RIP-based performance guarantees for multipath matching pursuit. Sci China Inf Sci, 2018, 61: 102303
3 Testa M, Magli E. Compressive Bayesian K-SVD. Signal Process-Image Commun, 2018, 60: 1–5
4 Gleichman S, Eldar Y C. Blind compressed sensing. IEEE Trans Inform Theory, 2011, 57: 6958–6975
5 Serra J G, Testa M, Molina R, et al. Bayesian K-SVD using fast variational inference. IEEE Trans Image Process, 2017, 26: 3344–3359
6 Aghagolzadeh M, Radha H. New guarantees for blind compressed sensing. In: Proceedings of the 53rd Annual Allerton Conference on Communication, Control, and Computing, Monticello, 2015. 1227–1234
7 Ristin M, Guillaumin M, Gall J, et al. Incremental learning of random forests for large-scale image classification. IEEE Trans Pattern Anal Mach Intell, 2016, 38: 490–503
8 Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J Imag Sci, 2009, 2: 183–202
9 Gorski J, Pfeuffer F, Klamroth K. Biconvex sets and optimization with biconvex functions: a survey and extensions. Math Meth Oper Res, 2007, 66: 373–407