Signal Processing 151 (2018) 45–55

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Optimized projections for compressed sensing via direct mutual coherence minimization

Canyi Lu[a], Huan Li[b], Zhouchen Lin[b,c,*]

[a] Department of Electrical and Computer Engineering, National University of Singapore, Singapore
[b] Key Laboratory of Machine Perception (MOE), School of EECS, Peking University, China
[c] Cooperative Medianet Innovation Center, Shanghai Jiao Tong University, China

ARTICLE INFO

Article history:
Received 26 December 2016
Revised 11 April 2018
Accepted 23 April 2018
Available online 26 April 2018

Keywords:
Mutual coherence minimization
Compressed sensing
Convergence guarantee

ABSTRACT

Compressed Sensing (CS) is a new data acquisition theory based on the existence of a sparse representation of a signal and a projected dictionary **PD**, where $\mathbf{P} \in \mathbb{R}^{m\times d}$ is the projection matrix and $\mathbf{D} \in \mathbb{R}^{d\times n}$ is the dictionary. To recover the signal from a small number m of measurements, it is expected that the projected dictionary **PD** is of low mutual coherence. Several previous methods attempt to find the projection **P** such that the mutual coherence of **PD** is low. However, they do not minimize the mutual coherence directly and thus they may be far from optimal. Their used solvers lack convergence guarantee and thus the quality of their solutions is not guaranteed. This work aims to address these issues. We propose to find an optimal projection matrix by minimizing the mutual coherence of **PD** directly. This leads to a nonconvex nonsmooth minimization problem. We approximate it by smoothing, solve it by alternating minimization and prove the convergence of our algorithm. To the best of our knowledge, this is the first work which directly minimizes the mutual coherence of the projected dictionary and has convergence guarantee. Numerical experiments demonstrate that our method can recover sparse signals better than existing ones.

1. Introduction

Compressed Sensing (CS) [1,2] is a new sampling/data acquisition theory asserting that one can exploit sparsity or compressibility when acquiring signals of interest. It shows that signals which have a sparse representation with respect to appropriate bases can be recovered from a small number of measurements. A fundamental problem in CS is how to construct a measurement matrix such that the number of measurements is near minimal.

Consider a signal $\mathbf{x} \in \mathbb{R}^d$ which is assumed to have a sparse representation with respect to a fixed overcomplete dictionary $\mathbf{D} \in \mathbb{R}^{d\times n}$ $(d < n)$. This can be described as

$$\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}, \tag{1}$$

where $\boldsymbol{\alpha} \in \mathbb{R}^n$ is a sparse representation coefficient, i.e., $\|\boldsymbol{\alpha}\|_0 \ll n$. Here $\|\boldsymbol{\alpha}\|_0$ denotes the ℓ_0-norm which counts the number of nonzero elements in $\boldsymbol{\alpha}$. The solution to problem (1) is not unique since $d < n$. To find an appropriate solution in the solution set of (1), we need to use some additional structures of **D** and $\boldsymbol{\alpha}$. Considering that $\boldsymbol{\alpha}$ is sparse, we are interested in finding the sparsest representation coefficient $\boldsymbol{\alpha}$. This leads to the following sparse representation problem

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0, \ \text{s.t.} \ \mathbf{x} = \mathbf{D}\boldsymbol{\alpha}. \tag{2}$$

However, the above problem is NP-hard [3] and thus is challenging to solve. Some algorithms, such as Basis Pursuit (BP) [4] and Orthogonal Matching Pursuit (OMP) [5], can be used to find suboptimal solutions.

An interesting theoretical problem is that under what conditions the optimal solution to (2) can be computed. If the solution is computable, can it be exactly or approximately computed by BP or OMP? Some previous works answer the above questions based on the mutual coherence of the dictionary **D** [6].

Definition 1. Given $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_n] \in \mathbb{R}^{d\times n}$, its mutual coherence is defined as the largest absolute and normalized inner product between different columns of **D**, i.e.,

$$\mu(\mathbf{D}) = \max_{\substack{1\le i,j\le n \\ i\neq j}} \frac{|\mathbf{d}_i^T \mathbf{d}_j|}{\|\mathbf{d}_i\| \|\mathbf{d}_j\|}.$$

* Corresponding author at: Key Laboratory of Machine Perception (MOE), School of EECS, Peking University, China.
E-mail addresses: canyilu@gmail.com (C. Lu), zlin@pku.edu.cn (Z. Lin).

https://doi.org/10.1016/j.sigpro.2018.04.020
0165-1684/

The mutual coherence measures the highest correlation between any two columns of $\mathbf{D}$. It is expected to be as low as possible in order to find the sparest solution to (2).

Theorem 1. *[6, 7, 8] For problem (2), if $\boldsymbol{\alpha}$ satisfies*

$$\|\boldsymbol{\alpha}\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu(\mathbf{D})}\right), \tag{3}$$

then the following results hold:

- *$\boldsymbol{\alpha}$ is the solution to (2).*
- *$\boldsymbol{\alpha}$ is also the solution to the following convex ℓ_1-minimization problem*

  $$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1, \ \text{s.t.} \ \mathbf{x} = \mathbf{D}\boldsymbol{\alpha},$$

  where $\|\boldsymbol{\alpha}\|_1 = \sum_i |\alpha_i|$ is the ℓ_1-norm of $\boldsymbol{\alpha}$.
- *$\boldsymbol{\alpha}$ can be obtained by OMP.*

The above theorem shows that if the mutual coherence of $\mathbf{D}$ is low enough, then the sparest solution to (2) is computable. Thus, how to construct a dictionary $\mathbf{D}$ with low mutual coherence is crucial in sparse coding. In CS, to reduce the number of measurements, we face a similar problem on the sensing matrix construction.

The theory of CS guarantees that a signal having a sparse representation can be recovered exactly from a small set of linear and nonadaptive measurements. This result suggests that it may be possible to sense sparse signals by taking far fewer measurements than what the conventional Nyquist–Shannon sampling theorem requires. But note that CS differs from classical sampling in several aspects. First, the sampling theory typically considers infinite-length and continuous-time signals. In contrast, CS is a mathematical theory that focuses on measuring finite-dimensional vectors in $\mathbb{R}^n$. Second, rather than sampling the signal at specific points in time, CS systems typically acquire measurements in the form of inner products between the signal and general test functions. At last, the ways to dealing with the signal recovery are different. Given the signal $\mathbf{x} \in \mathbb{R}^d$ in (1), CS suggests replacing these n direct samples with m indirect ones by measuring linear projections of $\mathbf{x}$ defined by a proper projection or sensing matrix $\mathbf{P} \in \mathbb{R}^{m\times d}$, i.e.,

$$\mathbf{y} = \mathbf{P}\mathbf{x}, \tag{4}$$

such that $m \ll d$. It means that instead of sensing all n elements of the original signal $\mathbf{x}$, we can sense $\mathbf{x}$ indirectly by its compressed form $\mathbf{y}$ in a much smaller size m. Surprisingly, the original signal $\mathbf{x}$ can be recovered from the observed $\mathbf{y}$ by using the sparse representation in (1), i.e, $\mathbf{y} = \mathbf{PD}\boldsymbol{\alpha}$ with the sparest $\boldsymbol{\alpha}$. Thus the reconstruction requires solving the following problem

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0, \ \text{s.t.} \ \mathbf{y} = \mathbf{M}\boldsymbol{\alpha}, \tag{5}$$

where $\mathbf{M} = \mathbf{PD} \in \mathbb{R}^{m\times n}$ is called the effective dictionary. Problem (5) is also NP-hard. As suggested by Theorem 1, if the mutual coherence of $\mathbf{PD}$ is low enough, then the solution $\boldsymbol{\alpha}$ to (5) is computable by OMP or by solving the following convex problem

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_1, \ \text{s.t.} \ \mathbf{y} = \mathbf{M}\boldsymbol{\alpha}. \tag{6}$$

Finally, the original signal $\mathbf{x}$ can be reconstructed by $\mathbf{x} = \mathbf{D}\boldsymbol{\alpha}$. So it is expected to find a proper projection matrix $\mathbf{P}$ such that $\mu(\mathbf{PD})$ is low. Furthermore, many previous works [9,10] show that the required number of measurements for recovering the signal $\mathbf{x}$ by CS can be reduced if $\mu(\mathbf{PD})$ is low.

In summary, the above discussions imply that by choosing an appropriate projection matrix $\mathbf{P}$ such that $\mu(\mathbf{PD})$ is low enough, the true signal $\mathbf{x}$ can be recovered with high probability by efficient algorithms. At the beginning, random projection matrices were shown to be good choices since their columns are incoherent with any fixed basis $\mathbf{D}$ with high probability [11]. However, many previous works [9,10,12] show that well designed deterministic projection matrices can often lead to better performance of signal reconstruction than random projections do. In this work, we focus on the construction of deterministic projection matrices. We first give a brief review on some previous deterministic methods.

1.1. Related work

In this work, we only consider the case that $\mathbf{D}$ is fixed while $\mathbf{P}$ can be changed. Our target is to find $\mathbf{P}$ by minimizing $\mu(\mathbf{M})$, where $\mathbf{M} = \mathbf{PD}$. If each column of $\mathbf{M}$ is normalized to have unit Euclidean length, then $\mu(\mathbf{M}) = \|\mathbf{G}\|_{\infty,\text{off}}$, where $\mathbf{G} = (g_{ij}) = \mathbf{M}^T\mathbf{M}$ is named as the Gram matrix and $\|\mathbf{G}\|_{\infty,\text{off}} = \max_{i\neq j}|g_{ij}|$ is the largest off-diagonal element of $|\mathbf{G}|$. Several previous works used the Gram matrix to find the projection matrix $\mathbf{P}$ [9,10,12]. We give a review on these methods in the following.

1.1.1. The algorithm of Elad

The algorithm of Elad [9] considers minimizing the t-averaged mutual coherence defined as the average of the absolute and normalized inner products between different columns of $\mathbf{M}$ which are above t, i.e.,

$$\mu_t(\mathbf{M}) = \frac{\sum_{1\le i,j\le k,\ i\neq j} \chi_t(|g_{ij}|)|g_{ij}|}{\sum_{1\le i,j\le k,\ i\neq j} \chi_t(|g_{ij}|)},$$

where $\chi_t(x)$ is the characteristic function defined as

$$\chi_t(x) = \begin{cases} 1, & \text{if } x \ge t, \\ 0, & \text{otherwise}, \end{cases}$$

and t is a fixed threshold which controls the top fraction of the matrix elements of $|\mathbf{G}|$ that are to be considered.

To find $\mathbf{P}$ by minimizing $\mu_t(\mathbf{M})$, some properties of the Gram matrix $\mathbf{G} = \mathbf{M}^T\mathbf{M}$ are used. Assume that each column of $\mathbf{M}$ is normalized to have unit Euclidean length. Then

$$\text{diag}\,(\mathbf{G}) = \mathbf{1}, \tag{7}$$

$$\text{rank}\,(\mathbf{G}) = m. \tag{8}$$

The work [9] proposed to minimize $\mu_t(\mathbf{M})$ by iteratively updating $\mathbf{P}$ as follows. First, initialize $\mathbf{P}$ as a random matrix and normalize each column of $\mathbf{PD}$ to have unit Euclidean length. Second, shrink the elements of $\mathbf{G} = \mathbf{M}^T\mathbf{M}$ (where $\mathbf{M} = \mathbf{PD}$) by

$$g_{ij} = \begin{cases} \gamma g_{ij}, & \text{if } |g_{ij}| \ge t, \\ \gamma t \text{sign}(g_{ij}), & \text{if } t > |g_{ij}| \ge \gamma t, \\ g_{ij}, & \text{if } \gamma t > |g_{ij}|, \end{cases}$$

where $0 < \gamma < 1$ is a down-scaling factor. Third, apply SVD and reduce the rank of $\mathbf{G}$ to be equal to m. At last, build the square root $\mathbf{S}$ of $\mathbf{G}$: $\mathbf{S}^T\mathbf{S} = \mathbf{G}$, where $\mathbf{S} \in \mathbb{R}^{m\times n}$, and find $\mathbf{P} = \mathbf{SD}^\dagger$, where $\dagger$ denotes the Moore–Penrose pseudoinverse.

There are several limitations of the algorithm of Elad. First, it is suboptimal since the t-averaged mutual coherence $\mu_t(\mathbf{M})$ is different from the mutual coherence $\mu(\mathbf{M})$ which is our real target. Second, the proposed algorithm to minimize $\mu_t(\mathbf{M})$ has no convergence guarantee. So the quality of the obtained solution is not guaranteed. Third, the choices of two parameters, t and γ, are crucial for the signal recovery performance in CS. However, there is no guideline for their settings and thus in practice it is usually difficult to find their best choices.

1.1.2. The algorithm of Duarte-Carajalino and Sapiro

The algorithm of Duarte-Carajalino and Sapiro [12] is not a method that is based on mutual coherence. It instead aims to find the sensing matrix $\mathbf{P}$ such that the corresponding Gram matrix is as close to the identity matrix as possible, i.e.,

$$\mathbf{G} = \mathbf{M}^T\mathbf{M} = \mathbf{D}^T\mathbf{P}^T\mathbf{PD} \approx \mathbf{I}, \tag{9}$$