ARTICLE IN PRESS YJCSS:2999

Journal of Computer and System Sciences ••• (••••) •••–•••

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Sparse approximation is provably hard under coherent dictionaries

A. Çivril *

Antalya International University, Computer Engineering Department, Universite Cad. No:2 Dosemealti, Antalya 07190, Turkey

ARTICLE INFO

Article history:
Received 20 October 2014
Received in revised form 4 May 2016
Accepted 17 July 2016
Available online xxxx

Keywords:
Sparse approximation
Complexity
PCP
Unique games

ABSTRACT

It is well known that sparse approximation problem is NP-hard under general dictionaries. Several algorithms have been devised and analyzed in the past decade under various assumptions on the *coherence* μ of the dictionary represented by an $M \times N$ matrix from which a subset of k column vectors is selected. All these results assume $\mu = O(k^{-1})$. This article is an attempt to bridge the big gap between the negative result of NP-hardness under general dictionaries and the positive results under this restrictive assumption. In particular, it suggests that the aforementioned assumption might be asymptotically the best one can make to arrive at any efficient algorithmic result under well-known conjectures of complexity theory. In establishing the results, we make use of a new simple multilayered PCP which is tailored to give a matrix with small coherence combined with our reduction.

1. Introduction

Given a dictionary Φ with normalized columns, represented by an $M \times N$ matrix ($\Phi \in \mathbb{R}^{M\times N}$) and a target signal $y \in \mathbb{R}^M$ such that the columns of Φ span $\mathbb{R}^M$, *sparse approximation problem* asks to find an approximate representation of y using a linear combination of at most k atoms, i.e. column vectors of Φ. This amounts to finding a coefficient vector $x \in \mathbb{R}^N$ for which one usually solves

$$\min_{\|x\|_0 = k} \|y - \Phi x\|_2 \tag{1}$$

We name the problem with this standard objective function (1) as Sparse. Stated in linear algebraic terms, it is essentially about picking a k-dimensional subspace defined by k column vectors of Φ such that the orthogonal projection of y onto that subspace is as close as possible to y. The reader should note that the problem can be defined with full generality using notions from functional analysis (e.g. Hilbert spaces with elements representing functions), as is usually conceived in signal processing. Indeed, defined in Hilbert and Banach spaces, it has been studied as *highly nonlinear approximation* in functional approximation theory [20,21]. However, we consider linear algebraic language as any kind of generalization is irrelevant to our discussion and the negative results we will present can be readily extended to the general case.

Although mainly studied in signal processing and approximation theory, sparse approximation problem is of combinatorial nature in finite dimensions and the optimal solution can be found by checking all $\binom{N}{k}$ subspaces. It is natural to ask whether one can do better and the answer partly lies in the fact that Sparse is NP-hard even to approximate within any factor [8,18] under general dictionaries. The intrinsic difficulty of the problem under this objective function prevents one

* Fax: +90 242 245 0100.
E-mail address: ali.civril@antalya.edu.tr.

http://dx.doi.org/10.1016/j.jcss.2016.07.001
0022-0000/© 2016 Published by Elsevier Inc.

from designing algorithms which can even *approximate* the optimal solution. Hence, efforts have been towards analyzing algorithms working on a restricted set of dictionaries for which one requires the column vectors to be an almost orthogonal set, namely an incoherent dictionary. Formally, one defines the *coherence* μ of a dictionary Φ as

$$\mu(\Phi) = \max_{i \neq j} |\langle \Phi_i, \Phi_j \rangle|$$

where Φ_i and Φ_j are the ith and jth columns of Φ, respectively and $\langle \cdot, \cdot \rangle$ denotes the usual inner product defined on $\mathbb{R}^M$. Recall that the columns of a dictionary have unit norm. Hence, the coherence μ takes values in the $[0, 1]$ closed interval.

There are roughly three types of algorithmic results regarding sparse approximation problem as listed below.

1. Results relating the quality of the solution $\|y - \Phi x\|_2$ produced by the algorithm to the quality of the optimal solution $\|y - \Phi x^*\|_2$ given that $\mu(\Phi)$ is a slowly growing or decreasing function of k.
2. Results showing the rate of convergence of an algorithm for elements from a specific set related to the dictionary (e.g. [15]).
3. Results stating conditions under which an algorithm optimally recovers a signal either via norms of certain matrices related to the dictionary (e.g. [23]) or the Restricted Isometry Property (RIP) (e.g. [7]).

The results of this paper are immediately related to the first kind, which are expressed via Lebesgue-type inequalities as named by Donoho et al. [9]. Accordingly, we shall define the following:

Definition 1.1. An algorithm is an $(f(k), g(k))$-approximation algorithm for Sparse under coherence $h(k)$ if it selects a vector x with at most $g(k)$ nonzero elements from the dictionary Φ with $\mu(\Phi) \leq h(k)$ such that

$$\|y - \Phi x\|_2 \leq f(k) \cdot \|y - \Phi x^*\|_2$$

where x^* is the optimal solution with at most k nonzero elements.

Orthogonal Matching Pursuit (OMP) is a well studied greedy algorithm yielding such approximation guarantees. There is also a slight variant of this algorithm named Orthogonal Least Squares (OLS). In the last decade, the following results were found in a series of papers by different authors:

Theorem 1.2. *([11]) OMP is an $(8\sqrt{k}, k)$-approximation algorithm for* Sparse *under coherence $\frac{1}{8\sqrt{2}(k+1)}$.*

Theorem 1.3 *([23]). OMP is a $(\sqrt{1+6k}, k)$-approximation algorithm for* Sparse *under coherence $\frac{1}{3k}$.*

Theorem 1.4 *([9]). OMP is a $(24, \lfloor k \log k \rfloor)$-approximation algorithm for* Sparse *under coherence $\frac{1}{90k^{3/2}}$.*

Theorem 1.5 *([22]). OMP is a $(3, 2^{\lfloor \frac{1}{\delta} \rfloor} k)$-approximation algorithm for* Sparse *under coherence $\frac{1}{14\left(2^{\lfloor \frac{1}{\delta} \rfloor} k\right)^{1+\delta}}$, for any fixed $\delta > 0$.*

Theorem 1.6 *([16]). OLS is a $(3, 2k)$-approximation algorithm for* Sparse *under coherence $\frac{1}{20k}$.*

In this paper, we are particularly interested in an approximation of the form $(f(k), k)$ which implies a solution to the standard sparse approximation problem. We will investigate the possibility of such an approximation with respect to $h(k)$. First, let us discuss these algorithmic results qualitatively. First, notice that all the results assume a coherence of $O(k^{-1})$. Indeed, it is a curious question whether such a restrictive assumption is needed for approximating the problem. There is no trivial answer. Another peculiarity is that all the results are essentially due to OMP (or a slight variant), which is a simple and intuitive greedy algorithm reminiscent of the greedy method for the well known Set Cover problem in combinatorial optimization. This method is optimal due to a result of Feige [10] with respect to approximating the best solution. In essence, sparse approximation can also be considered as a covering problem where we want to cover a target vector using vectors from a given set. Of course, unlike Set Cover which does not bear any contextual information on the elements, one also needs to take the linear algebraic content into account. Hence, thinking in purely analogical manner, one would expect that OMP is probably the best algorithm one can hope for under the definition of approximation we have provided and the assumption of $\mu = O(k^{-1})$ is most likely necessary. Although our results are not exactly tight and there is still some room for improvement (algorithmic and/or complexity theoretic), as we will see, this intuition is correct to a certain extent.

We would like to note that intuitive reasonings about why one might need coherence have already been discussed in the literature. The question whether one needs coherence is explicitly articulated in [4] where it is pointed out that “for if two columns are closely correlated, it will be impossible in general to distinguish whether the energy in the signal comes from one or the other”. However strong this intuition is, there is no complexity theoretic barrier for solving the sparse

Please cite this article in press as: A. Çivril, Sparse approximation is provably hard under coherent dictionaries, J. Comput. Syst. Sci. (2016), http://dx.doi.org/10.1016/j.jcss.2016.07.001