Signal Processing 120 (2016) 249–254

Contents lists available at ScienceDirect

Signal Processing

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

Fast communication

Upper bounds on the error of sparse vector and low-rank matrix recovery ☆

Mohammadreza Malek-Mohammadi [a,*], Cristian R. Rojas [a], Magnus Jansson [a], Massoud Babaie-Zadeh [b]

[a] ACCESS Linnaeus Centre, KTH, Stockholm, 10044, Sweden
[b] Electrical Engineering Department, Sharif University of Technology, Tehran 1458889694, Iran

ARTICLE INFO

Article history:
Received 13 April 2015
Received in revised form
7 September 2015
Accepted 9 September 2015
Available online 21 September 2015

Keywords:
Approximately sparse solutions
Low-rank matrix recovery
Restricted isometry property
Sparse vector recovery

ABSTRACT

Suppose that a solution $\tilde{\mathbf{x}}$ to an underdetermined linear system $\mathbf{b}=\mathbf{A}\mathbf{x}$ is given. $\tilde{\mathbf{x}}$ is approximately sparse meaning that it has a few large components compared to other small entries. However, the total number of nonzero components of $\tilde{\mathbf{x}}$ is large enough to violate any condition for the uniqueness of the sparsest solution. On the other hand, if only the dominant components are considered, then it will satisfy the uniqueness conditions. One intuitively expects that $\tilde{\mathbf{x}}$ should not be far from the true sparse solution $\mathbf{x}_0$. It was already shown that this intuition is the case by providing upper bounds on $\|\tilde{\mathbf{x}}-\mathbf{x}_0\|$ which are functions of the magnitudes of small components of $\tilde{\mathbf{x}}$ but independent from $\mathbf{x}_0$. In this paper, we tighten one of the available bounds on $\|\tilde{\mathbf{x}}-\mathbf{x}_0\|$ and extend this result to the case that $\mathbf{b}$ is perturbed by noise. Additionally, we generalize the upper bounds to the low-rank matrix recovery problem.

1. Introduction

Let $\mathbf{x}_0 \in \mathbb{R}^m$ denote a sparse solution of an underdetermined system of linear equations

$$\mathbf{b}=\mathbf{A}\mathbf{x} \tag{1}$$

in which $\mathbf{b} \in \mathbb{R}^n$ and $\mathbf{A} \in \mathbb{R}^{n\times m}, m>n$. Suppose that $\|\mathbf{x}_0\|_0=k$, where $\|\mathbf{x}_0\|_0$ designates the number of nonzero components or the ℓ_0 norm of $\mathbf{x}_0$. Further, let spark $(\mathbf{A})$ represent the spark of $\mathbf{A}$, defined as the minimum number of columns of $\mathbf{A}$ which are linearly dependent, and let $\delta_{2k}(\mathbf{A})$ denote the restricted isometry constant of order $2k$ for the matrix $\mathbf{A}$ [1]. It is well known that if $k<\text{spark}(\mathbf{A})/2$ or $\delta_{2k}(\mathbf{A})<1$, then $\mathbf{x}_0$ is the unique sparsest solution of the above set of equations [1,2].

When the sparsest solution of (1) is sought, one needs to solve

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{A}\mathbf{x}=\mathbf{b}. \tag{2}$$

However, the above program is generally NP-hard [3] and becomes very intractable when the dimensions of the problem increase. Since finding the sparse solution of (1) has many applications in various fields of science and engineering (cf. [4] for a comprehensive list of applications), many practical alternatives for (2) have been proposed [5–8]. If the solution obtained by these algorithms satisfies one of the above sufficient conditions, then, assuredly, this solution is the sparsest one.

Now, consider the case that the solution given by an algorithm is only approximately sparse meaning that it has some dominant components, while other components are very small but not equal to zero. If the total number of

☆ This work was supported by the ACCESS Linnaeus Centre, KTH (Royal Institute of Technology), the Strategic Research Area ICT-TNG program, and the Swedish Research Council under contract 621-2011-5847.

* Corresponding author.
E-mail addresses: mohamma@kth.se (M. Malek-Mohammadi), crro@kth.se (C.R. Rojas), janssonm@kth.se (M. Jansson), mbzadeh@yahoo.com (M. Babaie-Zadeh).

http://dx.doi.org/10.1016/j.sigpro.2015.09.003
0165-1684/

nonzero components is large such that neither of the mentioned conditions hold, it is not clear whether this solution is close to the true sparse solution or not. However, intuitively, one expects that if the number of effective components is small, then the obtained solution should not be far away from the true solution. Immediately, the following questions may be raised. Is this solution still close to the unique sparse solution of $\mathbf{b}=\mathbf{A}\mathbf{x}$? Is it possible in this case to establish a bound on the error of finding $\mathbf{x}_0$ without knowing $\mathbf{x}_0$? Similar questions can be asked when there is error or noise in (1). Taking the noise into account, (1) is updated to

$$\mathbf{b}=\mathbf{A}\mathbf{x}+\mathbf{e}, \tag{3}$$

where $\mathbf{e}$ is the vector of noise or error. In this setting, to estimate $\mathbf{x}_0$ given $\mathbf{b}$ and $\mathbf{A}$, the equality constraint in (2) is relaxed, and the following optimization problem should be solved:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \|\mathbf{A}\mathbf{x}-\mathbf{b}\| \le \epsilon, \tag{4}$$

where $\epsilon \ge \|\mathbf{e}\|$ is some constant and $\|\cdot\|$ designates the ℓ_2 norm.

The answers to the above questions were firstly given in [9]. Let $\tilde{\mathbf{x}}$ denote the output of an algorithm to find or estimate $\mathbf{x}_0$ from (1) or (3). Particularly, [9] provides two upper bounds on the error $\|\mathbf{x}_0-\tilde{\mathbf{x}}\|$. The first one is rather simple to compute but turns out to be loose. On the other hand, while the second bound is tight, generally, it is much more complicated to compute.

Herein, in the spirit of the loose bound in [9], we provide a better bound which is based on the same parameter of the matrix $\mathbf{A}$, but it is *strictly tighter* than the loose bound in [9]. Moreover, our proposed bound is obtained in a much simpler way with a *shorter* algebraic manipulation. The proposed bound is extended to the noisy setting defined in (3). Furthermore, these results are also generalized to the problem of low-rank matrix recovery from compressed linear measurements [10].

The bounds introduced in this paper can be used in analyzing the *theoretical* performance of algorithms in sparse vector and low-rank matrix recovery that provide approximately sparse or low-rank solutions such as [7,11,12].[1] However, the bounds are obtained without any assumption on the recovery algorithm, and it is possible to improve them by exploiting properties of a specific algorithm. A similar upper bound on the error of sparse recovery in the noisy case has been proposed in [13]. This upper bound, however, is only applicable when the given solution has a sparsity level, the number of nonzero components, not greater than that of the true solution, while our bounds are obtained under the opposite assumption on the sparsity level of the given solution.

The rest of this paper is organized as follows. After introducing the notations used throughout the paper, in Section 2, we first present the upper bounds on the error of sparse vector recovery and, next, generalize them to the low-rank matrix recovery problem. Section 3 is devoted to the proofs of the results in Section 2, followed by conclusions in Section 4.

Notations: For a vector $\mathbf{x}$, $\|\mathbf{x}\|$, $\|\mathbf{x}\|_1$, and $\|\mathbf{x}\|_0$ denote the ℓ_2, ℓ_1, and the so-called ℓ_0 norms, respectively. Moreover, $\mathbf{x}^{\downarrow}$ denotes a vector obtained by sorting the elements of $\mathbf{x}$ in terms of magnitude in descending order, and x_i designates the ith component of $\mathbf{x}$. $\mathbf{x}_I$ represents the subvector obtained from $\mathbf{x}$ by keeping components indexed by the set I. A vector is called k-sparse if it has exactly k nonzero components. For a matrix $\mathbf{A}$, $\mathbf{a}_i$ denotes the ith column. Additionally, spark $(\mathbf{A})$ and null $(\mathbf{A})$ designate the minimum number of columns of $\mathbf{A}$ that are linearly dependent and the null space of $\mathbf{A}$, respectively. Similar to the vectors, $\mathbf{A}_I$ represents the submatrix of $\mathbf{A}$ obtained by keeping those columns indexed by I. It is always assumed that the singular values of matrices are sorted in descending order, and $\sigma_i(\mathbf{X})$ denotes the ith largest singular value of $\mathbf{X}$. Let $\mathbf{X}=\sum_{i=1}^{q}\sigma_i\mathbf{u}_i\mathbf{v}_i^T$, where $q=\text{rank}\,(\mathbf{X})$, denote the singular value decomposition (SVD) of $\mathbf{X}$. $\mathbf{X}_{(r)}=\sum_{i=1}^{r}\sigma_i\mathbf{u}_i\mathbf{v}_i^T$ represents a matrix obtained by keeping the r first terms in the SVD of $\mathbf{X}$, and $\mathbf{X}_{(-r)}=\mathbf{X}-\mathbf{X}_{(r)}$. $\|\mathbf{X}\|_F$ denotes the Frobenius norm, and $\|\mathbf{X}\|_* \triangleq \sum_{i=1}^{q}\sigma_i(\mathbf{X})$, in which $q=\text{rank}\,(\mathbf{X})$, stands for the nuclear norm.

2. Upper bounds

In this section, the upper bounds on the error of sparse vector and low-rank matrix recovery are presented.

2.1. Sparse vector recovery

Following the common practice in the literature of compressive sensing (CS), we refer to $\mathbf{b}, \mathbf{A}$, and $\mathbf{e}$ in (3) as the measurement vector, sensing matrix, and noise vector, respectively. Before stating the results, we recall two definitions.

Definition 1 (*Candès [1]*). For a matrix $\mathbf{A}\in\mathbb{R}^{n\times m}$ and all integers $k\le m$, the restricted isometry constant (RIC) of order k is the smallest constant $\delta_k(\mathbf{A})$ such that

$$(1-\delta_k(\mathbf{A}))\|\mathbf{x}\|^2 \le \|\mathbf{A}\mathbf{x}\|^2 \le (1+\delta_k(\mathbf{A}))\|\mathbf{x}\|^2 \tag{5}$$

holds for all vectors $\mathbf{x}$ with sparsity at most k.

Definition 2 (*Babaie-Zadeh et al. [9]*). For a matrix $\mathbf{A}\in\mathbb{R}^{n\times m}$, let $\sigma_{\min,p}(\mathbf{A})>0$ for $p\le spark(\mathbf{A})-1$ be the smallest singular value of all $\binom{m}{p}$ possible $n\times p$ submatrices of $\mathbf{A}$.

The following theorem presents the upper bounds for both noisy and noiseless cases. We deliberately separate the noisy and noiseless cases in order to be able to provide a tighter bound in the noiseless setting.

Theorem 1. *Let $\mathbf{A}\in\mathbb{R}^{n\times m}$, $m>n$, denote a sensing matrix. We have the following upper bounds.*

- *Noiseless case: Suppose that $\mathbf{x}_0$ is a k-sparse solution of $\mathbf{A}\mathbf{x}=\mathbf{b}$, where $k<spark(\mathbf{A})/2$. For all $\tilde{\mathbf{x}}$ solutions of $\mathbf{A}\mathbf{x}=\mathbf{b}$*

[1] It is worth emphasizing that the results presented in this paper are theoretical in nature and can be used only to theoretically justify the effectiveness of an algorithm. More precisely, as they are based on some parameters of the sensing matrix, which cannot be computed in general, these results cannot be used to experimentally evaluate the performance of an algorithm.