Uniqueness in Noisy Dictionary Learning

Abstract

Extension of theorems in HS2011 to noisy measurements of compressible vectors.

Index Terms

Dictionary learning, sparse coding, sparse matrix factorization, uniqueness, compressed sensing, combinatorial matrix theory

I. Working questions

When $k > 1$ (i.e. we cannot write $b_i = c_i e_{\pi(i)}$), what can we say about $||D^{-1}P^T b_i - a_i||$ given $||(A - BPD)a_i|| \leq \epsilon$? By the $(2,\delta)$-lower-RIP,

$$||D^{-1}P^T b_i - a_i|| \leq \frac{1}{\sqrt{1+\delta}}||A(D^{-1}P^T b_i - a_i)|| \tag{1}$$

Applying the triangle inequality yields

$$\begin{aligned} ||A(D^{-1}P^T b_i - a_i)|| &\leq (||(A - BPD)D^{-1}P^T b_i|| + ||Aa_i - Bb_i||) \\ &\leq \varepsilon(||D^{-1}P^T b_i|| + 1) \end{aligned}$$

where the second inequality comes from the recovery condition and the fact that $D^{-1}P^T b_i \in S_{1,m}$. Hence,

$$||D^{-1}P^T b_i - a_i|| \leq \frac{\varepsilon}{\sqrt{1-\delta}}(||D^{-1}P^T b_i|| + 1)$$

II. Introduction

INDEPENDENT component analysis [?], [?] and dictionary learning with a sparse coding scheme [?] have become important tools for revealing underlying structure in many different types of data.

III. Statement of Results

Definition 1: We say that $A \in \mathbb{R}^{n\times m}$ has $(2k,\delta)$-lower-RIP when

$$\text{for all } a_1, a_2 \in S_{p,k},\ \|A(a_1 - a_2)\| \geq \sqrt{1-\delta}\|a_1 - a_2\|. \tag{2}$$

Let $S_{k,m}$ denote the set of all k-sparse vectors in $\Re p$, i.e. $S_{k,m} = \{x \in \Re^m : ||x||_0^0 \leq k\}$.
Conjecture 1: Generalize from proof of k = 1.

IV. Proofs

Case k = 1. Set $a_i = e_i$ $(i = 1, ..., m)$ to be the standard basis columns in $\Re^m$. Suppose that $A \in \Re^{n\times m}$ satisfies the $(2,\delta)$-lower-RIP for some $\delta \in (0,1)$ and that for some matrix $B \in \Re^{n\times m}$ and set of 1-sparse b_i we have the recovery condition $||Aa_i - Bb_i|| \leq \varepsilon$ $(i = 1, ..., m)$ for some $\varepsilon > 0$. We will show that if $\varepsilon < \sqrt{\frac{1-\delta}{2}}$, then $b_i = PDa_i$ for some permutation matrix $P \in \Re^{m\times m}$ and invertible diagonal matrix $D \in \Re^{m\times m}$.

To begin, note that since the b_i are 1-sparse, then $b_i = c_i e_{\pi(i)}$ for some $c_i \in \Re$ and $\pi : \{1, ..., m\} \mapsto \{1, ..., m\}$. It cannot be the case that $c_i = 0$ for some i, since otherwise the recovery condition would imply that $||Ae_i|| \leq \varepsilon < \sqrt{1-\delta}$ which is in contradiction with the $(2,\delta)$-lower-RIP. We will now show that π is necessarily injective (and thus a permutation). Suppose $\pi(i) = \pi(j)$. By the recovery condition,

$$||Ae_i - Bc_i e_{\pi(i)}|| = ||Ae_i - Bc_i e_{\pi(j)}|| = ||Ae_i - \frac{c_i}{c_j}Bc_j e_{\pi(j)}|| \leq \varepsilon \tag{3}$$

Also by the recovery condition,

$$||\frac{c_i}{c_j}Ae_j - \frac{c_i}{c_j}Bc_j e_{\pi(j)}|| \leq \frac{|c_i|}{|c_j|}\varepsilon \tag{4}$$

Combining these by the triangle inequality, we get

$$||A(e_i - \frac{c_i}{c_j}e_j)|| \leq \varepsilon(1 + \frac{|c_i|}{|c_j|}) \tag{5}$$

By the $(2, \delta)$-lower-RIP, we have

$$||A(e_i - \frac{c_i}{c_j}e_j)|| \geq \sqrt{1-\delta}||(e_i - \frac{c_i}{c_j}e_j)|| = \sqrt{1-\delta}\sqrt{1 + \frac{c_i^2}{c_j^2}} \tag{6}$$

whenever $i \neq j$. Now, since $\forall x \in \Re$, $1 + |x| \leq \sqrt{2}\sqrt{1+x^2}$, we have (for $x = \frac{c_i}{c_j}$)

$$\varepsilon = \sqrt{1-\delta}\frac{1+x^2}{1+|x|} \geq \sqrt{\frac{1-\delta}{2}} \tag{7}$$

which is in contradiction with our assumption on ε. Hence, π is injective and the matrix $P \in \Re^{m \times m}$ whose i-th column is $e_{\pi(i)}$ is a permutation matrix. Letting $D \in \Re^{m \times m}$ be the (invertible) diagonal matrix with elements $c_1, ..., c_m$, we have that $b_i = c_i e_{\pi(i)} = PDe_i = PDa_i$ for all $i \in \{1, ..., m\}$. Furthermore, the recovery condition becomes $||(A - BPD)e_i|| \leq \varepsilon$ for all $i \in \{1, ..., m\}$ (or, more generally, $||(A - BPD)x|| \leq \varepsilon||x||$ for all $x \in S_{1,m}$).

V. Discussion

Acknowledgment

We thank the following people for helpful discussions: . This work was supported by grant IIS-1219212 from the National Science Foundation. CH was also partially supported by an NSF All-Institutes Postdoctoral Fellowship administered by the Mathematical Sciences Research Institute through its core grant DMS-0441170. FTS was also supported by the Applied Mathematics Program within the Office of Science Advanced Scientific Computing Research of the U.S. Department of Energy under contract No. DE-AC02-05CH11231.