A Note on Sparsification by Frames

Christopher A Baker
Dept. of EE & CS, Univ. of Wisconsin-Milwaukee
Milwaukee, WI 53211
email: cabaker2@uwm.edu

May 11, 2014

Abstract

This note proposes a proof of a sharp generalized Dictionary-Restricted Isometry Property (D-RIP) sparsity bound constant for compressed sensing. For fulfilling D-RIP, the constant $\delta_k$ is used in the definition: $(1 - \delta_k)\|Dv\|_2^2 \leq \|\Phi Dv\|_2^2 \leq (1 + \delta_k)\|Dv\|_2^2$. Since a sharp bound for $\delta_k$ has been proved for RIP with $\delta_{2k} < \frac{\sqrt{2}}{2}$ by Cai and Zhang, we prove the same sharp bound for D-RIP. This idea can also be extended for proving other sharp D-RIP bounds (i.e., $\delta_k$).

Let $\Phi \in \mathbb{R}^{n \times p}$ and $\beta \in \mathbb{R}^p$ be a signal such that

$$y = \Phi \beta + z$$

with $\|z\|_2 \leq \varepsilon$. In compressed sensing, one can find a good stable approximation (in terms of $\varepsilon$ and the tail of $\beta$ consisting of $p - k$ smallest entries) of $\beta$ from the measurement matrix $\Phi$ and the measurement $y$ through solving an $\ell_1$-minimization, provided that $\Phi$ belongs to a family of well behaved matrices. A subclass of this family of matrices can be characterized by the well known restrictive isometry property (RIP) of Candès, Romberg, and Tao. [7, 8]. This property requires the following relation for $\Phi$

$$\sqrt{1 - \delta_k}\|c\|_2 \leq \|\Phi c\|_2 \leq \sqrt{1 + \delta_k}\|c\|_2$$

for every $k$-sparse vector $c$ (namely, $c$ has at most $k$ nonzero components), for some small constant $\delta_k$. Some bounds on $\delta$ have been determined, e.g., [2, 8] [1, 5]. Cai and Zhang recently have established several sharp RIP bounds that cover the most interesting cases of $\delta_k$ and $\delta_{2k}$ [3, 4], showing $\delta_k \leq \frac{1}{3}$, $\delta_{2k} < \frac{\sqrt{2}}{2}$.

The requirement of a signal being sparse or approximately sparse is a key in this setting. Many families of integrating signals indeed have sparse representations under suitable bases. Recently an interesting sparsifying scheme was proposed by Candes, Eldar, Needel, and Randall [6]. In their scheme, instead of bases, tight frames are used to sparsify signals.

Let $D \in \mathbb{R}^{p \times d}$ ($d \geq p$) be a tight frame and $k \leq d$. [6] suggests that one use the following optimization to approximate the signal $\beta$:

$$\hat{\beta} = \arg\min_{\gamma \in \mathbb{R}^p} \|D^* \gamma\|_1 \quad \text{subject to} \quad \|y - \Phi \gamma\|_2 \leq \varepsilon.$$  (1)
The traditional RIP is no longer effective in the generalized setting. Candes, Eldar, Needel, and Randall defined the \textit{D-restricted isometry property} which extends RIP [6]. Here we shall use the formulation of D-RIP in [10].

**Definition 1.** The measurement matrix $\Phi$ obeys the D-RIP with constant $\delta_k$ if
\[
(1 - \delta_k) \|Dv\|_2^2 \leq \|\Phi Dv\|_2^2 \leq (1 + \delta_k) \|Dv\|_2^2
\]
holds for all $k$-sparse vector $v \in \mathbb{R}^d$.

[6, 10] have determined some bound for the D-RIP constant $\delta_{2k}$. The purpose of this note is to remark, using the clever ideas of Cai and Zhang [4], that one can get sharp bound for D-RIP constant $\delta_{2k}$ without much difficulty.

**Theorem 2.** Let $D$ be an arbitrary tight frame and let $\Phi$ be a measurement matrix satisfying D-RIP with $\delta_{2k} < \sqrt{2}$, then the solution $\hat{\beta}$ to (1) satisfies
\[
\|\beta - \hat{\beta}\|_2 \leq C_0 \varepsilon + C_1 \frac{\|D^* \beta - (D^* \beta)_{\max(k)}\|_1}{\sqrt{k}}
\]
where the constants $C_0, C_2$ are constants that depend on $\delta_{2k}$, $(D^* \beta)_{\max(k)}$ is the vector $D^* \beta$ with all but the $k$ largest components (in magnitude) set to zero.

Before proving this theorem, let us make some remarks. Firstly, this bound is sharp in general as the counter examples are reported for $D = I$ in [9, 4]. Secondly, following the ideas of [3, 4], more general results (other sharp D-RIP bounds) can be obtained in parallel.

We need the following $\ell_1$-norm invariant convex $k$-sparse decomposition of Xu and Xu [11], and Cai and Zhang [4] in our proof of theorem 2. We shall take the description from [11].

**Lemma 3.** For positive integers $k \leq n$, and positive constant $C$, let $v \in \mathbb{R}^n$ be a vector with $\|v\|_1 \leq C$ and $\|v\|_\infty \leq \frac{C}{k}$. Then there are $k$-sparse vectors $w_1, \ldots, w_M$ with
\[
\|w_t\|_1 = \|v\|_1 \quad \text{and} \quad \|w_t\|_\infty \leq \frac{C}{k} \quad \text{for} \ t = 1, \ldots, M,
\]
such that
\[
v = \sum_{t=1}^M x_t w_t
\]
for some nonnegative real numbers $x_1, \ldots, x_M$ with $\sum_{t=1}^M x_t = 1$.

Now let us proceed to the proof of theorem 2.

**Proof.** In this proof we mainly follow the clever ideas in the proofs of Theorems 1.1 and 2.1 of [4], incorporating some more simplified steps. We also use some strategies from [11]. We only deal with the $\delta_{2k}$ case so that the key ideas can be conveyed clearly.

Let $h = \hat{\beta} - \beta$.

For a subset $S \subset \{1, 2, \ldots, d\}$, we will denote by $D_S$ the matrix $D$ restricted to the columns indexed by $S$. Let $\Omega$ denote the index set of the largest $k$ components of $D^* \beta$ (in magnitude), i.e., $(D^* \beta)_{\max(k)} = D^*_\Omega \beta$. With this notation we have $D^*_{\Omega^c} \beta = D^* \beta - (D^* \beta)_{\max(k)}$. As in [6], one can easily verify
1. \( \|D_{\Omega}^* h\|_1 \leq 2\|D_{\Omega}^* \beta\|_1 + \|D_{\Omega}^* h\|_1 \);

2. \( \|\Phi h\|_2 < 2\varepsilon \).

Denote \( v_i = \langle D_i, h \rangle \) for \( i = 1, \ldots, d \), where \( D_i \) is the \( i \)-th column of \( D \), then

\[
D^* h = (v_1, \ldots, v_d).
\]

By rearranging the columns of \( D \) if necessary, we may assume \( |v_1| \geq |v_2| \geq \cdots \geq |v_d| \).

Let \( T = \{1, 2, \ldots, k\} \). Since \( \|D_{\Omega}^* h\|_1 \leq \|D_{T}^* h\|_1 \) and \( \|D_{\Omega}^* h\|_1 + \|D_{T}^* h\|_1 = \|D_{\Omega}^* h\|_1 \), the relation \( \|D_{\Omega}^* h\|_1 \leq 2\|D_{\Omega}^* \beta\|_1 + \|D_{0}^* h\|_1 \) yields

\[
\|D_{T}^* h\|_1 \leq 2\|D_{\Omega}^* \beta\|_1 + \|D_{0}^* h\|_1
\]

Note that

\[
\|D_{T}^* h\|_\infty \leq \frac{\|D_{T}^* h\|_1}{k} \leq \frac{2\|D_{\Omega}^* \beta\|_1 + \|D_{0}^* h\|_1}{k}.
\]

Thus by lemma 3, the following \( \ell_1 \)-invariant convex \( k \)-sparse decomposition of \( D_{T}^* h \) is available:

\[
D_{T}^* h = \sum_{t=1}^{M} x_t w_t,
\]

with each \( w_t \in \mathbb{R}^d \) being \( k \)-sparse, \( \|w_t\|_1 = \|D_{T}^* h\|_1 \) and \( \|w_t\|_\infty \leq \frac{2\|D_{\Omega}^* \beta\|_1 + \|D_{0}^* h\|_1}{k} \).

From this and the Cauchy-Schwartz inequality, we have immediately

\[
\|D_{T}^* h\|_2 \leq \sum_{t=1}^{M} x_t \|w_t\|_2 \leq \frac{2\|D_{\Omega}^* \beta\|_1 + \|D_{T}^* h\|_1}{\sqrt{k}} \leq \frac{2\|D_{\Omega}^* \beta\|_1}{\sqrt{k}} + \|D_{0}^* h\|_2.
\]

Note that \( \|\beta - \hat{\beta}\|_2^2 = \|h\|_2^2 = \|D^* h\|_2^2 = \|D_{T}^* h\|_2^2 + \|D_{T}^* h\|_2^2 \) and \( D^* \beta - (D^* \beta)_{\text{max}(k)} = D_{\Omega}^* \beta \). In order to prove the theorem, it suffices to show that there are constants \( C_0', C_1' \) such that

\[
\|D_{T}^* h\|_2 \leq C_0' \varepsilon + C_1' \frac{\|D_{\Omega}^* \beta\|_1}{\sqrt{k}}.
\]

In fact, assuming (5) we get

\[
\|h\|_2 = \sqrt{\|D_{T}^* h\|_2^2 + \|D_{T}^* h\|_2^2} \leq \sqrt{(C_0' \varepsilon + C_1' \frac{\|D_{\Omega}^* \beta\|_1}{\sqrt{k}})^2 + \left( \frac{2\|D_{\Omega}^* \beta\|_1}{\sqrt{k}} + \|D_{T}^* h\|_2 \right)^2} \leq C_0' \varepsilon + C_1' \frac{\|D_{\Omega}^* \beta\|_1}{\sqrt{k}} + \frac{2\|D_{\Omega}^* \beta\|_1}{\sqrt{k}} + \|D_{T}^* h\|_2 \leq 2C_0' \varepsilon + 2(C_1' + 1) \frac{\|D_{\Omega}^* \beta\|_1}{\sqrt{k}}.
\]

Now let us prove (5).

Let \( \alpha \) be a positive real number such that

\[
\alpha - \frac{1 + 2\alpha^2}{2} \delta_{2k} > 0.
\]
This parameter will be determined later for the purpose of explaining how the condition \( \delta_{2k} < \frac{\sqrt{2}}{2} \) is derived. Let us denote

\[
\Pi := |\langle \Phi DD_T^* h, \alpha \Phi h \rangle| = |\langle \Phi DD_T^* h, \alpha \Phi D D^* h \rangle|.
\]

First, as \( D_T^* h \) is \( k \) sparse, hence \( 2k \) sparse. we have and \( \delta_k \leq \delta_{2k} \), we have

\[
\Pi \leq \| \Phi DD_T^* h \|_2 \| \alpha \Phi h \|_2 \leq \sqrt{1 + \delta_{2k} \| D_T^* h \|_2^2} 2 \alpha \varepsilon. \tag{7}
\]

On the other hand, as each \( D_T^* h + w_t \) is \( 2k \) sparse, we have

\[
\Pi = |\langle \Phi DD_T^* h, \alpha \Phi DD_T^* h + \alpha \Phi DD_T^* c h \rangle| \\
= |\sum_{t=1}^M x_t \langle \Phi DD_T^* h, \alpha \Phi DD_T^* h + \alpha \Phi Dw_t \rangle| \\
= |\sum_{t=1}^M x_t \left( (1 + \alpha - \frac{\alpha}{2} \Phi D D_T^* h + \alpha \Phi Dw_t) + (1 + \alpha - \frac{\alpha}{2} \Phi D D_T^* h - \alpha \Phi Dw_t) \right) \\
= \sum_{t=1}^M x_t \left( (1 - \delta_{2k}) \| \frac{1 + \alpha}{2} D_T^* h + \frac{\alpha}{2} w_t \|_2^2 - (1 + \delta_{2k}) \| \frac{1 - \alpha}{2} D_T^* h - \frac{\alpha}{2} w_t \|_2^2 \right) \\
= (\alpha - 1 + \frac{\alpha^2}{2} \delta_{2k}) \| D_T^* h \|_2^2 - \frac{\alpha^2}{2} \delta_{2k} \sum_{t=1}^M x_t \| w_t \|_2^2 \\
\geq (\alpha - 1 + \frac{\alpha^2}{2} \delta_{2k}) \| D_T^* h \|_2^2 - \frac{\alpha^2}{2} \delta_{2k} \left( \frac{2 \| D_T^* c \beta \|_1}{\sqrt{k}} + \| D_T^* h \|_1 \right)^2 \\
\geq (\alpha - 1 + \frac{\alpha^2}{2} \delta_{2k}) \| D_T^* h \|_2^2 - \frac{\alpha^2}{2} \delta_{2k} \left( \frac{4 \| D_T^* c \beta \|_1^2}{k} + \frac{4 \| D_T^* c \beta \|_1 \| D_T^* h \|_2}{\sqrt{k}} \right). \\
\]

Now we observe that the expression of the left hand side of (6) naturally becomes the coefficient of \( \| D_T^* h \|_2^2 \). To determine the value of \( \alpha \), we will use the criterion that the \( \alpha \) should be chosen so that the allowed range for \( \delta_{2k} \) is as large as possible. The condition (6) gives that

\[
\delta_{2k} < \frac{2\alpha}{1 + 2\alpha^2}
\]

The maximum value of the right hand side is \( \frac{\sqrt{2}}{2} \) and is achieved at \( \alpha = \frac{\sqrt{2}}{2} \).

Replacing \( \alpha \) by \( \frac{\sqrt{2}}{2} \) and using (7), we get

\[
\left( \frac{\sqrt{2}}{2} - \delta_{2k} \right) \| D_T^* h \|_2^2 - \delta_{2k} \| D_T^* c \beta \|_1^2 \frac{1}{\sqrt{k}} - \delta_{2k} \| D_T^* c \beta \|_1 \| D_T^* h \|_2 \leq \sqrt{1 + \delta_{2k}} \sqrt{2} \| D_T^* h \|_2 \varepsilon.
\]

By making perfect square, we have

\[
\left( \| D_T^* h \|_2^2 - \frac{\sqrt{2} \sqrt{1 + \delta_{2k}} \varepsilon + \delta_{2k} \| D_T^* c \beta \|_1}{\sqrt{2} - 2\delta_{2k}} \right)^2 \leq \left( \frac{\sqrt{2} \sqrt{1 + \delta_{2k}} \varepsilon + \delta_{2k} \| D_T^* c \beta \|_1}{\sqrt{2} - 2\delta_{2k}} \right)^2 + \left( \frac{2\delta_{2k} \| D_T^* c \beta \|_1}{\sqrt{k}} \right)^2.
\]

4
This implies that
\[
\|D^*_T h\|_2 \leq \sqrt{2\sqrt{1 + \delta_{2k}^2 \varepsilon} + \delta_{2k}^2 \frac{\|D^*_C \beta\|_1}{\sqrt{k}}} \cdot \sqrt{2\sqrt{1 + \delta_{2k}^2 \varepsilon} + \delta_{2k}^2 \frac{\|D^*_C \beta\|_1}{\sqrt{k}}} + \sqrt{\frac{2\delta_{2k}}{\sqrt{2} - 2\delta_{2k}}} \frac{\|D^*_C \beta\|_1}{\sqrt{k}},
\]
and finally we get (5):
\[
\|D^*_T h\|_2 \leq \sqrt{2\sqrt{1 + \delta_{2k}^2 \varepsilon} + \delta_{2k}^2 \frac{\|D^*_C \beta\|_1}{\sqrt{k}}} \cdot \sqrt{2\sqrt{1 + \delta_{2k}^2 \varepsilon} + \delta_{2k}^2 \frac{\|D^*_C \beta\|_1}{\sqrt{k}}} + \sqrt{\frac{2\delta_{2k}}{\sqrt{2} - 2\delta_{2k}}} \frac{\|D^*_C \beta\|_1}{\sqrt{k}}.
\]

Acknowledgement:

I would like to thank my academic supervisor, Professor Guangwu Xu, for his assistance in this proof.

References

[1] T. Cai, L. Wang, and G. Xu, New Bounds for Restricted Isometry Constants, IEEE Transactions on Information Theory, 56(2010), 4388-4394.

[2] T. Cai, G. Xu, and J. Zhang, On Recovery of Sparse Signals via \(\ell_1\) Minimization, IEEE Trans. Inf. Theory, 55(2009), 3388-3397.

[3] T. Cai and A. Zhang, Sharp RIP bound for sparse signal and low-rank matrix recovery, Applied and Computational Harmonic Analysis, 35(2013), 74-93.

[4] T. Cai and A. Zhang, Sparse representation of a polytope and recovery of sparse signals and low-rank matrices, 2013, http://www-stat.wharton.upenn.edu/tcai/Papers.html

[5] E. J. Candès, The restricted isometry property and its implications for compressed sensing, Compte Rendus de l’ Academie des Sciences, Paris, Serie I, 346 589-592.

[6] E.J. Candès, Y. Eldar, D. Needel, P. Randall, Compressed sensing with coherent and redundant dictionaries, Appl. Comput. Harmonic Anal. 31(1) (2010) 59-73.

[7] E. J. Candès J. Romberg, and T. Tao, Stable signal recovery from incomplete and inaccurate measurements, Communications on Pure and Applied Mathematics, 59(2006), 1207-1223.

[8] E. J. Candès and T. Tao, Decoding by linear programming, IEEE Transactions on Information Theory, 51(2005), 4203-4215.

[9] M. E. Davies and R. Gribonval, Restricted Isometry Constants where \(\ell_p\) sparse recovery can fail for \(0 < p \leq 1\), IEEE Trans. Inf. Theory, 2009.
[10] J. Lin, S. Li, and Y. Shen, New bounds for restricted isometry constants with coherent tight frames, *IEEE Transactions on Signal Processing*, 61(2013)

[11] G. Xu and Z. Xu, On the $\ell_1$-Norm Invariant Convex $k$-Sparse Decomposition of Signals, 2013, [http://arxiv.org/abs/1305.6021](http://arxiv.org/abs/1305.6021)