Decay properties of restricted isometry constants

Citation for published version:
Blanchard, JD, Cartis, C & Tanner, J 2009, 'Decay properties of restricted isometry constants' IEEE Signal Processing Letters, vol. 16, no. 7, pp. 572-575. DOI: 10.1109/LSP.2009.2020882

Digital Object Identifier (DOI):
10.1109/LSP.2009.2020882

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Early version, also known as pre-print

Published In:
IEEE Signal Processing Letters

Blanchard, JD., Cartis, C. & Tanner, J. 2009, 'Decay Properties of Restricted Isometry Constants', IEEE Signal Processing Letters, Vol. 16, No. 7, pp. 572-575.

Decay Properties of Restricted Isometry Constants

Jeffrey D. Blanchard*, Coralia Cartis, and Jared Tanner

***Abstract*—Many sparse approximation algorithms accurately recover the sparsest solution to an underdetermined system of equations provided the matrix's restricted isometry constants (RICs) satisfy certain bounds. There are no known large deterministic matrices that satisfy the desired RIC bounds; however, members of many random matrix ensembles typically satisfy RIC bounds. This experience with random matrices has colored the view of the RICs' behavior. By modifying matrices assumed to have bounded RICs, we construct matrices whose RICs behave in a markedly different fashion than the classical random matrices; RICs can satisfy desirable bounds and also take on values in a narrow range.**

***Index Terms*—Compressed sensing, restricted isometry constants, RIP, sparse approximation**

I. Introduction

A central task in sparse approximation and compressed sensing [1], [2], [3], is to approximate or recover a compressible or sparse signal from only a limited number of linear observations. Using an underdetermined measurement matrix and having knowledge of these measurements, the sparsest vector giving rise to these measurements is sought. In this context, Candès and Tao [2] introduced the *restricted isometry constants* of a matrix, otherwise known as *restricted isometry property* (RIP) constants.

Definition 1. Let A be an $n \times N$ matrix with $n < N$. The *k-restricted isometry constant of A*, δ_k^A, is the smallest number such that

$$\left(1-\delta_k^A\right)\|x\|_2^2 \le \|Ax\|_2^2 \le \left(1+\delta_k^A\right)\|x\|_2^2 \qquad (1)$$

for every vector $x \in \chi^N(k) := \left\{x \in \mathbb{R}^N : \|x\|_0 \le k\right\}$, where $\|x\|_0$ counts the number of nonzero entries in x.

Since $\chi^N(k) \subset \chi^N(k+1)$, it is clear that $\delta_k^A \le \delta_{k+1}^A$ for any k. For sparse approximation and compressed sensing, it is desirable to have matrices with bounded k-restricted isometry constants for k proportional to n as n grows. Computing the restricted isometry constants of a matrix is a combinatorial problem and thus intractable for large matrices. Fortunately many random matrix ensembles, for example Gaussian, typically have bounded k-restricted isometry constants for k proportional to n as n grows; moreover, bounds on these constants are known [2], [4].

J.D. Blanchard is with the Department of Mathematics, University of Utah, 155 South 1400 East, Room 233, Salt Lake City, Utah, 84112-0090 USA. T:+1 801 585 1644, F:+1 801 581 4148 (jeff@math.utah.edu)

C. Cartis and J. Tanner are with the School of Mathematics and the Maxwell Institute, University of Edinburgh, Edinburgh, UK. (coralia.cartis, jared.tanner@ed.ac.uk)

EDICS: DSP-TFSR

Unpublished references are available at http://www.dsp.ece.rice.edu/cs/

By determining the magnitude of the restricted isometry constants it is possible to make quantitative statements as to when various sparse approximation algorithms are guaranteed to recover the sparsest solution; here we focus on ℓ_1-regularization. We show (constructively) that there are matrices whose restricted isometry constants have strikingly different decay rates (with respect to k as k decreases) than are observed for the random matrix ensembles typically used in sparse approximation.

Throughout, let A be an $n \times N$ matrix with $n < N$. Let $x \in \chi^N(k)$ for $k < n$ and $y = Ax$. We seek to recover the sparsest vector x from (y, A), namely,

$$\min \|x\|_0 \quad \text{subject to } y = Ax. \qquad (2)$$

Rather than solve (2) directly through a combinatorial search, the problem is relaxed to solving [5]

$$\min \|x\|_1 \quad \text{subject to } y = Ax. \qquad (3)$$

If (2) and (3) both have a unique solution which is x, we call *x a point of ℓ_1/ℓ_0-equivalence.* A major endeavor in compressed sensing is determining when every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence.

Donoho [6] has provided a necessary and sufficient (geometric) condition on the measurement matrix A so that every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence. Consider C^N, the ℓ_1-ball in $\mathbb{R}^N$, whose $2N$ vertices are the canonical basis vectors $\{\pm e_j : j = 1, \dots, N\}$. Associated to the matrix A, there is a convex polytope P_A obtained by applying A to C^N; $P_A = AC^N$. A polytope P is *k-central-neighborly* when every set of $k+1$ vertices (which do not include an antipodal pair) span a k-dimensional face of P.

Theorem 1. (Donoho [6]) *Every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence if and only if $P_A = AC^N$ has $2N$ vertices and is k-central-neighborly.*

Random matrices with Gaussian entries typically provide k-central-neighborly measurement matrices for k proportional to n as n grows [6], [7]. This geometric perspective inspires the proofs of theorems of Section 2 concerning the RIP and ℓ_1/ℓ_0-equivalence.

The restricted isometry approach of Candès and Tao [2] provides sufficient conditions for when every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence. The following is a small sample of the various conditions placed on the restricted isometry constants of A.

Theorem 2. (Candès, Tao [2]) *If $\delta_k^A + \delta_{2k}^A + \delta_{3k}^A < 1$, then every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence.*

Theorem 3. (Candès, Romberg, Tao [8]) *If $\delta_{3k}^A + 3\delta_{4k}^A < 2$,* *then every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence.*

Theorem 4. (Chartrand, Staneva [9]) *For $b > 1$ with bk an integer, if $\delta_{bk}^A + b\delta_{(b+1)k}^A < b - 1$, then every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence.*

Theorem 5. (Candès [10]) *If $\delta_{2k}^A < \sqrt{2} - 1$, then every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence.*

This paper explores the possible behavior of the restricted isometry constants for an arbitrary matrix. Understanding how δ_k^A may vary with k for a general matrix A is essential in the search for suitable compressed sensing matrices and for the comparison of RIP statements involving multiple sparsity levels. In Section 2 we state the main results and discuss their implications for compressed sensing. We present the proofs of the main results and elaborate on their implications in Section 3.

II. Main Results

In order to ensure that any k-sparse vector can be recovered from (y, A), even via an exhaustive search, no two k-sparse vectors may be mapped by A to the same observation y. When $\delta_{2k}^A < 1$, we are assured that A will return a unique observation y for every $x \in \chi^N(k)$ and therefore guarantee[1] a unique solution to (2). Restricting δ_{2k}^A to be bounded by a constant smaller than one, e.g. Theorem 5, is sufficient to ensure ℓ_1/ℓ_0-equivalence; however, the largest bound on δ_{2k}^A which guarantees ℓ_1/ℓ_0-equivalence is not known. Our first theorem states that $\delta_{2k}^A < 1$ is not sufficient to guarantee ℓ_1/ℓ_0-equivalence. In fact, no restricted isometry constant being less than one will ensure that 1-sparse vectors can be recovered; results of a similar nature were also derived by Davies and Gribonval [11].

Theorem 6. *For any $l \in \{1, \ldots, n-1\}$, $\delta_l^A < 1$ does not imply that every $x \in \chi^N(1)$ is a point of ℓ_1/ℓ_0-equivalence.*

There is no known deterministic class of matrices for which there is a fixed $\rho \in (0, 1)$ such that $\delta_{\lceil \rho n \rceil}^A < 1$ as $n \to \infty$ and $n/N \to \tau \in (0, 1)$. However there are random matrix ensembles whose members are shown to typically have bounded restricted isometry constants. In particular, for Gaussian random matrices there exists a constant $\rho^\star \in (0, 1)$ such that $\delta_{\lceil \rho^\star n \rceil}^A < 1$ as $n \to \infty$ with $n/N \to \tau \in (0, 1)$ [2]. For these same random matrices, it is known that $\delta_2^A \sim n^{-1/2}$ [12]. Moreover, the restricted isometry constants δ_l^A decrease rapidly from $\delta_{\lceil \rho^\star n \rceil}^A$ (near 1) to near 0 as l decreases from $\lceil \rho^\star n \rceil$ to 2; we refer to this as the decay rate of the restricted isometry constants of A. In the search for broader classes of matrices with bounded k-restricted isometry constants for k proportional to n as n grows, it may prove beneficial to know that we need not mimic the restricted isometry constant behavior of these random matrix ensembles. Moreover, when making quantitative comparisons of Theorems 2-5, how the restricted isometry constants vary with k plays an important role. The second result states that $\delta_k^A < 1$ does not imply that $\delta_1^A << 1$; indeed δ_1^A may be arbitrarily close to δ_k^A. That is, the restricted isometry constants may not exhibit appreciable decay.

Theorem 7. *Given any $\epsilon \in (0, 1)$ and $k \in \{1, \ldots, n-1\}$, there exists a matrix A such that $\delta_1^A, \ldots, \delta_k^A \in [1 - \epsilon, 1)$.*

At first glance this may seem not to be such a significant obstacle since having $\delta_l^A < 1$ for any l was already not sufficient to recover a 1-sparse vector. However, it is also possible to construct a matrix whose RIP constants are all confined to an interval whose length is equal to the difference between two consecutive restricted isometry constants of another matrix.

Theorem 8. *Suppose there exists a matrix B of size $(n-1) \times (N-1)$ such that $\delta_k^B < 1$. Then there exists a matrix A of size $n \times N$ such that $\delta_1^A, \ldots, \delta_k^A \in \left[\delta_{k-1}^B, \delta_k^B\right]$.*

Although we do not know a way to construct or randomly draw a matrix with $\delta_k^B - \delta_{k-1}^B$ being arbitrarily small for a specific choice of k, this clustering of restricted isometry constants is typical of matrices with δ_k^B bounded for k proportional to n as $n \to \infty$. For any $\epsilon > 0$ and $k = \lceil \rho n \rceil$ for some $\rho \in (0, 1)$, if δ_k is bounded as $n \to \infty$ (as is the case for Gaussian random matrices) then for all but at most a finite number of $j \le k$, $\delta_j^B - \delta_{j-1}^B < \epsilon$ as $n \to \infty$.

From Theorem 8, having a restricted isometry constant, δ_k^A, which is strictly bounded away from 1, even for k arbitrarily large, does not give any indication of the size of the smallest restricted isometry constants. This lack of decay helps interpret some previous RIP results.

In general, there exist matrices such that $\delta_k^A + \epsilon$ and $\delta_{2k}^A + \epsilon$ are greater than δ_{3k}^A. Unless further information is known about the restricted isometry constants of A it is appropriate to collapse Theorem 2 to $\delta_{3k}^A < 1/3$. In fact, collapsing RIP statements to a single sparsity level allows for a more intuitive comparison of the results. In this light, Theorem 5 is a verifiable improvement of Theorem 2 as it simultaneously decreases the sparsity level restriction from $3k$ to $2k$ and increases the bound from $1/3$ to $\sqrt{2} - 1$. With Theorem 4, Chartrand and Staneva point out that the integers 2, 3, and 4 in Theorem 3 can be replaced by $b-1$, b, and $b+1$, respectively. Theorem 8 implies that in the general setting, one may collapse Theorem 4 to the single sparsity level, ck.

Corollary 9. *For $c > 2$ with ck an integer, if $\delta_{ck} < \frac{c-2}{c}$, then every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence.*

Observe that collapsing Theorems 2 and 3 results in special cases of Corollary 9 with $c = 3$ and $c = 4$, respectively. On one hand, Corollary 9 is less desirable that Theorem 5 as it requires A to act as a restricted isometry on larger support sizes. However, this trade-off allows the bound on the restricted isometry constants to approach 1. For example, if $\delta_{100k}^A < .98$, then every $x \in \chi^N(k)$ is a point of ℓ_1/ℓ_0-equivalence for the matrix A. Even more quantitative notions of the restricted isometry constants [4], [13] suggest that the

[1] Requiring $\delta_{2k}^A < 1$ is not necessary to ensure that there are no two k-sparse vectors which are mapped by A to the same measurement y. For many matrices there exists an $x \in \chi^N(2k)$ such that $\|Ax\|_2 \ge 2\|x\|_2$, i.e. $\delta_{2k}^A \ge 1$, while there are no $2k$-sparse vectors mapped to zero; examples include Gaussian $\mathcal{N}(0, 1/\sqrt{n})$ matrices commonly used in compressed sensing [4].

significantly larger support size is not unrealistic; for example, the current RIP statements involving Gaussian matrices require $n > 317k$ [4].

Needell and Tropp [14] have shown that the restricted isometry constants cannot exceed a linear growth rate as k increases: for positive integers c and k, the restricted isometry constants of A satisfy $\delta_{ck}^A \leq k \cdot \delta_{2c}^A$. In particular, for $c = 1$, $\delta_k^A \leq k \cdot \delta_2^A$. Since restricted isometry constants are nondecreasing, knowledge of either δ_2^A or δ_k^A implies a bound on the other; namely that the restricted isometry constants are contained in the intervals,

$$\delta_2^A, \ldots, \delta_k^A \in \left[\delta_2^A, k\delta_2^A\right] \tag{4}$$

$$\text{and} \quad \delta_2^A, \ldots, \delta_k^A \in \left[\frac{1}{k}\delta_k^A, \delta_k^A\right], \tag{5}$$

respectively. Whereas (4) and (5) indicate large intervals that contain the restricted isometry constants, Theorem 8 states that these constants may in fact be contained in an arbitrarily narrow interval.

Another generic condition on A used in sparse approximation is its *coherence* defined by

$$\mu^A := \sup_{\substack{i,j \in \{1,\ldots,N\} \\ i \neq j}} |\langle Ae_i, Ae_j \rangle| \,. \tag{6}$$

Smaller values of μ^A provide larger values of k for which it is guaranteed that most $x \in \chi^N(k)$ are points of ℓ_1/ℓ_0-equivalence [12]. It is known that $\mu^A \leq \delta_2^A$ [2]. Analogous to Theorem 7, knowledge of $\delta_k^A < 1$ for k large does not imply that μ^A is small.

Theorem 10. *For any $\epsilon > 0$ and any $k \in \{1, \ldots, n-1\}$, there exists a matrix A with $\delta_k^A < 1$ and $\mu^A > 1 - \epsilon$.*

Theorem 7 states that, although $\delta_k^A < 1$ for k large, δ_1^A may be arbitrarily close to one; Theorem 10 tells us that μ^A may also be arbitrarily close to one. Therefore assumptions regarding the coherence of A are additional assumptions to those regarding the restricted isometry constants.

III. Proofs of Main Results

Throughout this section let $z \in \mathbb{R}^{N-1} \setminus \{0\}$ and $\alpha \in \mathbb{R}$. Recall that e_j is the jth standard basis vector of $\mathbb{R}^N$ and C^N denotes the ℓ_1 ball in $\mathbb{R}^N$. We refer to a vector with no more than l nonzero entries as being *at most l-sparse*.

Theorem 6 states that knowledge of $\delta_k^A < 1$ for k large does not even guarantee recovery of 1-sparse vectors by solving (3). This is proved by showing that there are matrices that satisfy $\delta_l < 1$ for any $l \in \{1, \ldots, n-1\}$ which are not 0-central-neighborly.

Proof of Theorem 6. Let $1 \leq l < n$, B be a full rank matrix of size $n \times (N-1)$ with $\delta_l^B < 1$, and set $0 < \epsilon < \sqrt{1 - \delta_l^B}$. Select any point u with $\|u\|_2 \leq \epsilon$ that is in the interior of $P_B = BC^{N-1}$ such that, for any $1 \leq m < n$, u has no m-sparse representation in terms of the columns of B. (Such a vector u exists as $0 \in \text{int}P_B$ and the columns of B span $\mathbb{R}^n$. Therefore, when $m < n$, the set of vectors with an m-sparse representation in terms of the columns of B has measure zero in $\mathbb{R}^n$ while P_B has positive measure in $\mathbb{R}^n$.) Define A by appending u to B: $A = [B \ u]$. To show $\delta_l^A < 1$, we first prove $\|Ay\|_2^2/\|y\|_2^2 > 0$, and then $\|Ay\|_2^2/\|y\|_2^2 < 2$. Let $y^T = \begin{bmatrix} z^T & \alpha \end{bmatrix}$ be an at most l-sparse vector. Then z is at most l-sparse and hence $(1 - \delta_l^B)\|z\|_2^2 \leq \|Bz\|_2^2 \leq (1 + \delta_l^B)\|z\|_2^2$. Since $\delta_l^B < 1$, for $\alpha = 0$, we have $\|Ay\|_2 = \|Bz\|_2 > 0$. When $\alpha \neq 0$, again we must have $\|Ay\|_2 = \|Bz + \alpha u\|_2 > 0$ due to our choice of u; otherwise, $Bz + \alpha u = 0$ would imply that u admits an at most $(n-1)$-sparse representation in terms of the columns of B. Thus in all cases, $\|Ay\|_2^2/\|y\|_2^2 > 0$. To prove $\|Ay\|_2^2/\|y\|_2^2 < 2$, note that

$$\|Ay\|_2 \leq \|Bz\|_2 + |\alpha| \cdot \|u\|_2 \leq \sqrt{1 + \delta_l^B}\|z\|_2 + \epsilon|\alpha|,$$

which, by applying Cauchy-Schwarz inequality to the above right-hand side, gives $\|Ay\|_2^2 \leq (1 + \delta_l^B + \epsilon^2)\|y\|_2^2 < 2\|y\|_2^2$ due to our choice of ϵ.

Despite $\delta_l^A < 1$, $Ae_N = u \in \text{int}P_B = \text{int}P_A$, thus P_A is not 0-central-neighborly. By Theorem 1, there exists $x \in \chi^N(1)$, for example e_N, that is not a point of ℓ_1/ℓ_0-equivalence. $\square$

To prove Theorems 7 and 8, we use the following lemma.

Lemma 11. *Let B be an $(n-1) \times (N-1)$ matrix with $\delta_k^B < 1$, and $A = \begin{bmatrix} B & 0 \\ 0 & \sqrt{\beta} \end{bmatrix}$ where $\beta \in (0,1)$. Then*

$$\delta_1^A \geq 1 - \beta \quad \textit{and} \quad \delta_k^A \leq \max\{\delta_k^B, 1 - \beta\}. \tag{7}$$

Proof. Since $Ae_N = \sqrt{\beta}\ e_N$, we have $\beta = \|Ae_N\|_2^2 \geq (1 - \delta_1^A)\|e_N\|_2^2 = 1 - \delta_1^A$. Thus, the first part of (7) follows. To show the second inequality, let $y^T = \begin{bmatrix} z^T & \alpha \end{bmatrix}$ be an at most k-sparse vector; thus z is at most k-sparse and so $(1 - \delta_k^B)\|z\|_2^2 \leq \|Bz\|_2^2 \leq (1 + \delta_k^B)\|z\|_2^2$. Also, $\|Ay\|_2^2 = \|Bz\|_2^2 + \beta\alpha^2$ and hence

$$(1 - \delta_k^B)\|z\|_2^2 + \beta\alpha^2 \leq \|Ay\|_2^2 \leq (1 + \delta_k^B)\|z\|_2^2 + \beta\alpha^2.$$

Since $\|y\|_2^2 = \|z\|_2^2 + \alpha^2$, we have

$$\min\{1 - \delta_k^B, \beta\} \leq \frac{\|Ay\|_2^2}{\|y\|_2^2} \leq \max\{1 + \delta_k^B, \beta\},$$

which together with (1), gives bounds on the lower and upper restricted isometry constants, $\min\{1 - \delta_k^B, \beta\} \leq 1 - \delta_k^A$ and $\max\{1 + \delta_k^B, \beta\} \geq 1 + \delta_k^A$, hence

$$\delta_k^A \leq \max\{1 - \min\{1 - \delta_k^B, \beta\}, \max\{1 + \delta_k^B, \beta\} - 1\}$$

which for $\beta \in (0,1)$ and $\delta_k^B \geq 0$ reduces to the second inequality in (7). $\square$

Theorem 7 shows that no assumption can be made in general about how the restricted isometry constants vary with k. In fact these constants may be made arbitrarily close together. To demonstrate this, we perturb a matrix known to have a certain restricted isometry constant less than one. Since these constants are nondecreasing, we construct a matrix that retains this restricted isometry constant less than 1 but has δ_1 arbitrarily close to 1.

Proof of Theorem 7. Let B be an $(n-1) \times (N-1)$ matrix with $\delta_k^B < 1$. Construct A from B as in Lemma 11 with $\beta := \epsilon \in (0,1)$. Then the first inequality in (7) provides $\delta_1^A \geq 1-\epsilon$. Since $\delta_k^B < 1$ by design, the second inequality in (7) yields $\delta_k^A < 1$. □

Proof of Theorem 8. Construct A from the given B as in Lemma 11 with $\beta := 1-\delta_{k-1}^B$. The first inequality in (7) implies $\delta_1^A \geq \delta_{k-1}^B$, while the second gives $\delta_k^A \leq \delta_k^B$. □

The coherence, μ^A, of the measurement matrix A is often used in addition to or independent of the restricted isometry constants to derive results in sparse approximation. While $\mu \leq \delta_2$, the restricted isometry constants can be arbitrarily close together and even arbitrarily close to one, it is natural to ask if the coherence can also be arbitrarily close to one while preserving that the restricted isometry constants are all less than one. Theorem 10 shows this is indeed possible.

Proof of Theorem 10. Let B be a full rank matrix of size $n \times (N-1)$ with $\delta_k^B < 1$, unit norm columns $b_1, \ldots, b_{N-1}$, and let $P_B = BC^{N-1}$, with vertices $\{\pm b_i\}_{i=1}^{N-1}$. Consider $0 < \epsilon \ll 1$. Pick any vertex b_j. Let $\tilde{u} = (1-\frac{\epsilon}{2})b_j$. Then $\tilde{u} \in \mathrm{int}P_B$ and so there exists $\beta \in (0, \epsilon/2)$ so that the ball $\mathtt{B}_\beta(\tilde{u})$ of radius β centered at $\tilde{u}$ satisfies $\mathtt{B}_\beta(\tilde{u}) \subset \mathrm{int}P_B$. Choose $u \in \mathtt{B}_\beta(\tilde{u})$ so that u has no $(n-1)$-sparse or sparser representation in terms of the columns of B (see the proof of Theorem 6 as to why this choice is possible). Define A by appending u to B and scaling: $A = [B \;\; u]/\sqrt{2}$. To show $\delta_k^A < 1$, let $y^T = \begin{bmatrix} z^T & \alpha \end{bmatrix}$ be at most k-sparse. The argument for $\|Ay\|_2^2/\|y\|_2^2 > 0$ follows similarly to the corresponding part of the proof of Theorem 6 (with $l := k$). It remains to show that $\|Ay\|_2^2 < 2\|y\|_2^2$, or equivalently, that $\|Ay\|_2^2 < 2$ for y with $\|y\|_2 \leq 1$. To prove the latter, note that $u = \tilde{u} + (u-\tilde{u})$ and so

$$\begin{aligned}
\|Ay\|_2 &= \|Bz + \alpha(1-\epsilon/2)b_j + \alpha(u-\tilde{u})\|_2/\sqrt{2} \\
&= \|B(z+\alpha e_j) - \alpha\frac{\epsilon}{2}b_j + \alpha(u-\tilde{u})\|_2/\sqrt{2} \\
&\leq \left(\sqrt{1+\delta_k^B}\|z+\alpha e_j\|_2 + \left(\frac{\epsilon}{2}+\beta\right)|\alpha|\right)/\sqrt{2} \\
&\leq \sqrt{1+\delta_k^B} + \epsilon < \sqrt{2}, \text{ for } \epsilon \text{ sufficiently small,}
\end{aligned}$$

where in the second inequality, we used $|\alpha|, \|z+\alpha e_j\|_2 \leq \sqrt{2}$ and $\beta < \epsilon/2$; in the first inequality above, besides using $\|b_j\|_2 = 1$ and $u \in \mathtt{B}_\beta(\tilde{u})$, we argued that in the nontrivial case when $\alpha \neq 0$, z is at most $k-1$ sparse and so $z + \alpha e_j$ is at most k-sparse and hence (1) provides $\|B(z+\alpha e_j)\|_2 \leq \sqrt{1+\delta_k^B}\|z+\alpha e_j\|_2$.

Since b_j and u are both columns of A, then (6) implies

$$\begin{aligned}
\mu^A &\geq |\langle b_j, u\rangle| = |\langle b_j, \tilde{u} + (u-\tilde{u})\rangle| \\
&\geq (1-\frac{\epsilon}{2})\,\|b_j\|_2^2 - |\langle b_j, u-\tilde{u}\rangle| \geq 1-\epsilon,
\end{aligned}$$

with the last inequality due to $\|b_j\|_2^2 = 1$ and $\beta < \frac{\epsilon}{2}$. □

IV. Conclusions

Sparse approximation results derived from bounds on restricted isometry constants, such as Theorems 2-4, are most applicable to matrices (or random matrix ensembles) with significant decay rates of the restricted isometry constants. For a general matrix the restricted isometry constants may exhibit no decay; hence, statements such as Theorem 5 or Corollary 9 are more appropriate where there is no further knowledge of the decay properties of the restricted isometry constants. Finally, an assumption on the coherence of a matrix is additional to assumptions on the restricted isometry constants.

Acknowledgments

The authors thank E. Candès for providing an early draft of [10] and the referees for their useful comments. JDB acknowledges support from NSF DMS (VIGRE) grant 0602219. JT acknowledges support from the Alfred P. Sloan Foundation and thanks John E. and Marva M. Warnock for their generous support in the form of an endowed chair.

References

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[3] E. J. Candès, "Compressive sampling," in *International Congress of Mathematicians. Vol. III*. Eur. Math. Soc., Zürich, 2006, pp. 1433–1452.

[4] J. D. Blanchard, C. Cartis, and J. Tanner, "The restricted isometry property and ℓ^q-regularization: phase transitions for sparse approximation," 2008, submitted.

[5] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp. 129–159 (electronic), 2001, reprinted from SIAM J. Sci. Comput. **20** (1998), no. 1, 33–61.

[6] D. L. Donoho, "Neighborly polytopes and sparse solution of underdetermined linear equations," 2005, technical Report, Department of Statistics, Stanford University.

[7] D. L. Donoho and J. Tanner, "Counting faces of randomly-projected polytopes when the projection radically lowers dimension," *J. of the AMS*, vol. 22, no. 1, pp. 1–53, 2009.

[8] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[9] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 035020, pp. 1–14, 2008.

[10] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *C. R. Math. Acad. Sci. Paris*, vol. 346, no. 9-10, pp. 589–592, 2008.

[11] M. E. Davies and R. Gribonval, "Restricted isometry constants where ℓ^p sparse recovery can fail for $0 < p \leq 1$," *IEEE Trans. Inform. Theory*, 2009, in press.

[12] J. Tropp, "On the conditioning of random subdictionaries," *Appl. Comp. Harm. Anal.*, vol. 25, no. 1, pp. 1–24, 2008.

[13] S. Foucart and M.-J. Lai, "Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leq 1$," *Appl. Comput. Harmon. Anal.*, 2008, in press.

[14] D. Needell and J. Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comp. Harm. Anal.*, in press.