Chaz's Theorem

Abstract

Extension of theorems in HS2011 to noisy measurements of approximately sparse vectors.

Index Terms

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

I. Introduction

INTRODUCTORY sentence.

II. Definitions

In what follows, we will use the notation $[m]$ for the set $\{1, ..., m\}$, and $\binom{[m]}{k}$ for the subsets of $[m]$ of cardinality k. For a subset $S \subseteq [m]$ and matrix A with columns $\{A_1, ..., A_m\}$ we define

$$\text{Span}\{A_S\} = \text{Span}\{A_s : s \in S\}.$$

Definition 1: Let V, W be subspaces of $\mathbb{R}^m$ and let $d(v, W) := \inf\{\|v - w\|_2 : w \in W\}$. Denote by $\mathcal{S}$ the unit sphere in $\mathbb{R}^m$. The *gap* metric Θ defined on [see Theory of Linear Operators in a Hilbert Space p. 69 who cites first reference] closed linear subspaces of $\mathbb{R}^m$ is defined as

$$\Theta(V, W) := \max\left(\sup_{v \in \mathcal{S} \cap V} d(v, W), \sup_{w \in \mathcal{S} \cap W} d(w, V)\right). \tag{1}$$

We note the following useful fact [ref: Morris, Lemma 3.3]: If we know a priori that $\dim(W) = \dim(V)$ then

$$\Theta(V, W) = \sup_{v \in \mathcal{S} \cap V} d(v, W) = \sup_{w \in \mathcal{S} \cap W} d(w, V). \tag{2}$$

Definition 2: We say that $A \in \mathbb{R}^{n \times m}$ satisfies the (ℓ, α)-*lower-RIP* [ref: Restricted Isometry Property first introduced in "Decoding by linear programming" by Candes and Tao] when for some $\alpha \in (0, 1]$,

$$\|Aa\|_2 \geq \alpha \|a\|_2 \quad \text{for all } \ell\text{-sparse } a \in \mathbb{R}^m.$$

We note here the following useful relationship between the gap metric and the lower-RIP, the proof of which is contained in Lemma 3: If a matrix A satisfies an $(\ell+1, \alpha)$-lower-RIP, then for all $S \neq S' \in \binom{[m]}{\ell}$, we have $\Theta(\text{Span}\{A_S\}, \text{Span}\{A_{S'}\}) \geq \alpha$.

Definition 3: The *Friedrichs angle* $\theta_F \in [0, \frac{\pi}{2}]$ between two subspaces V and W, is the minimal angle between $V \cap (V \cap W)^\perp$ and $W \cap (W \cap V)^\perp$:

$$\cos\theta_F := \max\left\{\frac{\langle v, w\rangle}{\|v\|\|w\|} : v \in V \cap (V \cap W)^\perp, w \in W \cap (V \cap W)^\perp\right\} \tag{3}$$

III. Robust Deterministic Uniqueness Theorem

Definition 4: We say a dataset $Y = \{y_1, ..., y_N\}$ has a *sparse representation* with respect to some matrix A when for some

Definition 5: We say a dataset $Y = \{y_1, ..., y_N\}$ has a *C-stable sparse representation* when it satisfies the following property: there exists some $C > 0$ such that any pair of matrices $A, B \in \mathbb{R}^{n \times m}$, A having unit norm columns, for which $\|y_i - Aa_i\| \leq \varepsilon$ and $\|y_i - Bb_i\| \leq \varepsilon$ for some k-sparse $a_i, b_i \in \mathbb{R}^m$ for all $i \in 1, \ldots, N$ are such that $\|(A - BPD)e_i\| \leq C\varepsilon$ for some permutation matrix $P \in \mathbb{R}^m$ and invertible diagonal matrix $D \in \mathbb{R}^m$, provided ε is small enough.

Theorem 1: Fix $k \leq n < m$ and $\alpha \in (0, 1]$. There exist sets of $N = k\binom{m}{k}^2$ k-sparse $a_i \in \mathbb{R}^m$ and $C > 0$ with the following property: if $Y = \{y_1, ..., y_N\}$ is a dataset for which, for some $A \in \mathbb{R}^{n \times m}$ with unit norm columns satisfying a $(2k, \alpha)$-lower-RIP, $\|y_i - Aa_i\|_2 \leq \varepsilon$ for all $i = 1, \ldots, N$, then Y has a C-stable spare representation.

Corollary 1: Theorem 1 in HS2011. *Proof:* The spark condition implies $(2k, \alpha)$-lower-RIP for some $\alpha > 0$. Set $\varepsilon = 0$.

Proof of Theorem 1: First, we produce a set of $N = k\binom{m}{k}^2$ vectors in general linear position (i.e. any set of k of them are linearly independent). Specifically, let $\sigma_1, ..., \sigma_N$ be any distinct numbers. Then the columns of the $k \times N$ matrix $V = (\sigma_j^i)_{i,j=1}^{k,N}$ are in general linear position (since the σ_j are distinct, any $k \times k$ "Vandermonde" sub-determinant is nonzero). Next, form the k-sparse vectors $a_1, ..., a_N \in \mathbb{R}^m$ by setting the nonzero values of vector a_i to be those contained in the ith column of V while partitioning the a_i evenly among the $\binom{m}{k}$ possible supports.

We will demonstrate that the existence of these a_i proves the theorem. First, we claim that there exists some $\delta > 0$ such that for any set of k vectors $a_{i_1}, ..., a_{i_k}$, the following is true:

$$\|\sum_{j=1}^{k} c_j a_{i_j}\|_2 \geq \delta \|c\|_1 \quad \text{for all} \quad c = (c_1, ..., c_k) \in \mathbb{R}^m. \tag{4}$$

To see why, consider the compact set $\mathcal{C} = \{c : \|c\|_1 = 1\}$ and the continuous map

$$\phi : \mathcal{C} \to \mathbb{R}$$
$$(c_1, ..., c_k) \mapsto \|\sum_{j=1}^{k} c_j a_{i_j}\|_2.$$

By general linear position of the a_i, we know that $0 \notin \phi(\mathcal{C})$. Since $\mathcal{C}$ is compact, we have by continuity of ϕ that $\phi(\mathcal{C})$ is also compact; hence it is closed and bounded. Therefore 0 can't be a limit point of $\phi(\mathcal{C})$ and there must be some $\delta > 0$ such that the neighbourhood $\{x : x < \delta\} \subseteq \mathbb{R} \setminus \phi(\mathcal{C})$. Hence $\phi(c) \geq \delta$ for all $c \in \mathcal{C}$. The property (4) follows by the association $c \mapsto \frac{c}{\|c\|_1}$ and the fact that there are only finitely many subsets of k vectors a_i, hence there is some minimal δ satisfying (4) for all of them.

Now suppose that $Y = \{y_1, ..., y_N\}$ is a dataset for which $\forall i \in [N], \|y_i - Aa_i\| \leq \varepsilon$ for some $A \in \mathbb{R}^{n \times m}$ satisfying the properties stated in the statement of the theorem and that for some alternate $B \in \mathbb{R}^{n \times m}$ there exist k-sparse b_i for which $\forall i \in [N], \|y_i - Bb_i\| \leq \varepsilon$. Since there are $k\binom{m}{k}$ vectors a_i with a given support $S \in \binom{[m]}{k}$, the pigeon-hole principle implies that there are at least k vectors y_i such that $\|y_i - Aa_i\| \leq \varepsilon$ for these a_i and also $\|y_i - Bb_i\| \leq \varepsilon$ for b_i all sharing some support $S' \in \binom{[m]}{k}$. Let $\mathcal{Y} = \{y_i : i \in \mathcal{I}\}$ be a set of k such vectors y_i indexed by $\mathcal{I}$.

Note that any matrix satisfying the $(2k, \alpha)$-lower-RIP is such that any $2k$ columns are linearly independent. It follows from this and the general linear position of the a_i that the set $\{Aa_i : i \in \mathcal{I}\}$ is a basis for $\text{Span}\{A_S\}$. Hence, fixing $z \in \text{Span}\{A_S\}$, there exist a set of $c_i \in \mathbb{R}$ such that $z = \sum_{i \in \mathcal{I}} c_i Aa_i$. Letting $y = \sum_{i \in \mathcal{I}} c_i y_i \in \text{Span}\{\mathcal{Y}\}$, we have by the triangle inequality that

$$\|z - y\|_2 = \|\sum_{i \in \mathcal{I}} c_i Aa_i - \sum_{i \in \mathcal{I}} c_i y_i\|_2 \leq \sum_{i \in \mathcal{I}} \|c_i(Aa_i - y_i)\|_2 = \sum_{i \in \mathcal{I}} |c_i| \|Aa_i - y_i\|_2 \leq \varepsilon \sum_{i \in \mathcal{I}} |c_i|. \tag{5}$$

The alternate factorization for the y_i implies (by a similar manipulation as in (5)) that for $z' = \sum_{i \in \mathcal{I}} c_i Bb_i \in \text{Span}\{B_{S'}\}$ we have $\|y - z'\|_2 \leq \varepsilon \sum_{i \in \mathcal{I}} |c_i|$ as well. It follows again by the triangle inequality that

$$\|z - z'\|_2 \leq \|z - y\|_2 + \|y - z'\|_2 = 2\varepsilon \sum_{i \in \mathcal{I}} |c_i|. \tag{6}$$

Since the a_i with $i \in \mathcal{I}$ all share the same support and A satisfies the $(2k, \alpha)$-lower-RIP, we have

$$\|z\|_2 = \|\sum_{i \in \mathcal{I}}^{k} c_i Aa_i\|_2 = \|A(\sum_{i \in \mathcal{I}} c_i a_i)\|_2 \geq \alpha \|\sum_{i \in \mathcal{I}}^{k} c_i a_i\|_2 \geq \alpha\delta \sum_{i \in \mathcal{I}}^{k} |c_i|. \tag{7}$$

where we have applied the property (4). Combining (6) and (7), we see that for all $z \in \text{Span}\{\mathcal{A}_S\}$ there exists some $z' \in \text{Span}\{B_{S'}\}$ such that

$$\|z - z'\|_2 \leq \tilde{C}\varepsilon \|z\|_2 \quad \text{where} \quad \tilde{C} = \frac{2}{\alpha\delta} \tag{8}$$

It follows that

$$\sup_{\substack{z \in \text{Span}\{A_S\} \\ \|z\|=1}} d(z, \text{Span}\{B_{S'}\}) \leq \tilde{C}\varepsilon. \tag{9}$$

If ε is such that $\tilde{C}\varepsilon < 1$ then by Lemma 6 and the fact that every k columns of A are linearly independent we have $\dim(\text{Span}\{B_{S'}\}) \geq \dim(\text{Span}\{A_S\}) = k$. Since $|S'| = k$, it follows that $\dim(\text{Span}\{B_{S'}\}) = \dim(\text{Span}\{A_S\})$. Recalling (2), it is implied by (9) that $\Theta(\text{Span}\{A_S\}, \text{Span}\{\mathcal{B}_{\mathcal{S}'}\}) \leq \tilde{C}\varepsilon$. Specifically, if

$$\varepsilon < \frac{\alpha^2\delta}{4} \prod_{j=1}^{k} \left[1 + \frac{\cos\theta_j + \sqrt{2 - \cos^2\theta_j}}{1 - \cos^2\theta_j}\right]^{-1} \tag{10}$$

then $\tilde{C}\varepsilon < 1$ and the association $S \mapsto S'$ defines a map $\pi : \binom{[m]}{k} \to \binom{[m]}{k}$ satisfying

$$\Theta(\mathrm{Span}\{A_S\}, \mathrm{Span}\{\mathcal{B}_{\pi(S)}\}) \le \tilde{C}\varepsilon < \frac{\alpha}{2} \prod_{j=1}^{k} \left[1 + \frac{\cos\theta_j + \sqrt{2 - \cos^2\theta_j}}{1 - \cos^2\theta_j} \right]^{-1} \quad \text{for all} \quad S \in \binom{[m]}{k}. \tag{11}$$

from which it follows by Lemma 1 that there exist a permutation matrix $P \in \mathbb{R}^{m\times m}$ and invertible diagonal matrix $D \in \mathbb{R}^{m\times m}$ such that $\|(A - BPD)e_i\| \le \tilde{C}\varepsilon$ for all $i = 1, \dots, m$.

We complete the proof by showing that the dataset Y has a C-stable sparse representation for $C = 2\tilde{C}$. Suppose that some matrix $A' \in \mathbb{R}^{n\times m}$ with unit norm columns is such that $\|y_i - A'a_i'\| \le \varepsilon$ for some k-sparse $a_i' \in \mathbb{R}^m$ for all $i \in 1, \dots, N$. By the preceding arguments, there exists a permutation matrix $P' \in \mathbb{R}^{m\times m}$ and invertible diagonal matrix $D' \in \mathbb{R}^{m\times m}$ such that $\|(A - A'P'D')e_i\| \le \tilde{C}\varepsilon$. By the triangle inequality, we have for all $i \in 1, \dots, N$ that

$$\|(A'P'D' - BPD)e_i\| \le \|(BPD - A)e_i\| + \|(A - A'P'D')e_i\| \le 2\tilde{C}\varepsilon \tag{12}$$

FINISH THIS ■

Lemma 1 (Main Lemma): Fix positive integers $k \le n < m$. Let $A, B \in \mathbb{R}^{n\times m}$ with A having the $(2k, \alpha)$-lower-RIP and unit norm columns and let $\theta_j \in [0, \frac{\pi}{2}]$ be the least of all Friedrichs angles formed between pairs of subspaces for which j columns of A form a basis. If there exists a map $\pi : \binom{[m]}{k} \to \binom{[m]}{k}$ and $\Delta > 0$ such that for all $S \in \binom{[m]}{k}$,

$$\Theta(\mathrm{Span}\{A_S\}, \mathrm{Span}\{B_{\pi(S)}\}) \le \Delta < \frac{\alpha}{2} \prod_{j=1}^{k} \left[1 + \frac{\cos\theta_j + \sqrt{2 - \cos^2\theta_j}}{1 - \cos^2\theta_j} \right]^{-1} \tag{13}$$

then there exist a permutation matrix $P \in \mathbb{R}^{m\times m}$ and an invertible diagonal matrix $D \in \mathbb{R}^{m\times m}$ such that for all $i \in [m]$,

$$\|(A - BPD)e_i\|_2 \le \Delta \prod_{j=1}^{k} \left(\frac{\cos\theta_j + \sqrt{2 - \cos^2\theta_j}}{1 - \cos^2\theta_j} \right) \tag{14}$$

Proof of Lemma 1: We prove the following equivalent statement: If there exists a map $\pi : \binom{[m]}{k} \to \binom{[m]}{k}$ and $\Delta > 0$ such that for all $S \in \binom{[m]}{k}$,

$$\Theta(\mathrm{Span}\{A_S\}, \mathrm{Span}\{B_{\pi(S)}\}) \le f_k(\Delta) < \Delta_k := \frac{\alpha}{2} \prod_{j=1}^{k} \left[1 + \frac{\cos\theta_j + \sqrt{2 - \cos^2\theta_j}}{1 - \cos^2\theta_j} \right]^{-1} \tag{15}$$

where

$$f_k(\Delta) = \Delta \prod_{j=1}^{k} \left(\frac{1 - \cos^2\theta_j}{\cos\theta_j + \sqrt{2 - \cos^2\theta_j}} \right) \tag{16}$$

then there exist a permutation matrix $P \in \mathbb{R}^{m\times m}$ and an invertible diagonal matrix $D \in \mathbb{R}^{m\times m}$ such that for all $i \in [m]$,

$$\|(A - BPD)e_i\|_2 \le \Delta. \tag{17}$$

We shall induct on k, the base case $k = 1$ being contained in Lemma 2. First, we demonstrate that π is injective (and thus bijective) for $k \ge 2$. Suppose $\pi(S_1) = \pi(S_2) = S^*$ for some $S_1, S_2 \in \binom{[m]}{k}$. We have by the triangle inequality and (15) that

$$\Theta(\mathrm{Span}\{A_{S_1}\}, \mathrm{Span}\{A_{S_2}\}) \le \Theta(\mathrm{Span}\{A_{S_1}\}, \mathrm{Span}\{B_{S^*}\}) + \Theta(\mathrm{Span}\{B_{S^*}\}, \mathrm{Span}\{A_{S_2}\}) < 2\Delta_k < \alpha$$

from which it follows by Lemma 3 (setting $\ell = k + 1$) that $S_1 = S_2$. Hence π is bijective. Moreover, from this bijectivity of π and the fact that every k columns of A are linearly independent, it follows by Lemma 6 that every k columns of B are linearly independent. (Fix $S \in \binom{[m]}{k}$. Then $\Theta(\mathrm{Span}\{A_{\pi^{-1}(S)}\}, \mathrm{Span}\{B_S\}) < \Delta_k < 1$ and we have $k \ge \dim(\mathrm{Span}\{B_S\}) \ge \dim(\mathrm{Span}\{A_{\pi^{-1}(S)}\}) = k$.)

The bijectivity of π actually implies another constraint on the columns of B which we demonstrate now to make use of later. When we consider that not only do the columns of A form linearly independent subsets, but satisfy a $(2k, \alpha)$-lower-RIP, we have

$$\Theta(\mathrm{Span}\{B_{S_1}\}, \mathrm{Span}\{B_{S_2}\}) \ge \alpha - 2f_k(\Delta) > \alpha - 2\Delta_k > 0 \quad \text{for all} \quad S_1 \ne S_2 \in \binom{[m]}{k}. \tag{18}$$

This follows from Lemma 3 and the triangle inequality, since

$$\begin{aligned} \alpha \le \Theta(\mathrm{Span}\{A_{\pi^{-1}(S_1)}\}, \mathrm{Span}\{A_{\pi^{-1}(S_2)}\}) &\le \Theta(\mathrm{Span}\{A_{\pi^{-1}(S_1)}\}, \mathrm{Span}\{B_{S_1}\}) + \Theta(\mathrm{Span}\{B_{S_1}\}, \mathrm{Span}\{B_{S_2}\}) \\ &\quad + \Theta(\mathrm{Span}\{B_{S_2}\}, \mathrm{Span}\{A_{\pi^{-1}(S_2)}\}) \\ &\le 2f_k(\Delta) + \Theta(\mathrm{Span}\{B_{S_1}\}, \mathrm{Span}\{B_{S_2}\}). \end{aligned}$$

We complete the proof of the lemma, inductively, by producing a map $\tau : \binom{[m]}{k-1} \to \binom{[m]}{k-1}$ such that

$$\Theta(\mathrm{Span}\{A_S\}, \mathrm{Span}\{B_{\tau(S)}\}) \leq f_{k-1}(\Delta) < \Delta_{k-1} \tag{19}$$

holds for all $S \in \binom{[m]}{k-1}$. Fix $S \in \binom{[m]}{k-1}$ and set $S_1 = S \cup \{q\}$ and $S_2 = S \cup \{p\}$ for some $q, p \notin S$ with $q \neq p$ (we know such a pair must exist since $k < m$) so that $\pi^{-1}(S_1) \neq \pi^{-1}(S_2)$ by injectivity of π. Condition (15) implies that for all unit vectors $z \in \mathrm{Span}\{B_{S_1}\} \cap \mathrm{Span}\{B_{S_2}\}$, we have $d(z, \mathrm{Span}\{A_{\pi^{-1}(S_1)}\}) \leq f_k(\Delta)$ and $d(z, \mathrm{Span}\{A_{\pi^{-1}(S_2)}\}) \leq f_k(\Delta)$. It follows by Lemmas 4 and 5 that

$$d\left(z, \mathrm{Span}\{A_{\pi^{-1}(S_1)\cap\pi^{-1}(S_2)}\}\right) \leq f_k(\Delta)\left(\frac{\cos\theta_k + \sqrt{2-\cos^2\theta_k}}{1-\cos^2\theta_k}\right) = f_{k-1}(\Delta) \tag{20}$$

Noting that $f_k(\Delta) < \Delta_k$, we have:

$$\begin{aligned} f_{k-1}(\Delta) &< \Delta_k \left(\frac{\cos\theta_k + \sqrt{2-\cos^2\theta_k}}{1-\cos^2\theta_k}\right) \\ &= \Delta_{k-1}\left[1 + \frac{\cos\theta_k + \sqrt{2-\cos^2\theta_k}}{1-\cos^2\theta_k}\right]^{-1}\left(\frac{\cos\theta_k + \sqrt{2-\cos^2\theta_k}}{1-\cos^2\theta_k}\right) \\ &< \Delta_{k-1} \end{aligned} \tag{21}$$

Since (20) holds for all unit vectors $z \in \mathrm{Span}\{B_{S_1}\} \cap \mathrm{Span}\{B_{S_2}\} \supseteq \mathrm{Span}\{B_S\}$, it follows that

$$\sup_{\substack{z \in \mathrm{Span}\{B_S\} \\ \|z\|=1}} d\left(z, \mathrm{Span}\{A_{\pi^{-1}(S_1)\cap\pi^{-1}(S_2)}\}\right) \leq f_{k-1}(\Delta) < \Delta_{k-1}, \tag{22}$$

We will show that, in fact, $\Theta\left(\mathrm{Span}\{B_S\}, \mathrm{Span}\{A_{\pi^{-1}(S_1)\cap\pi^{-1}(S_2)}\}\right) \leq f_{k-1}(\Delta) < \Delta_{k-1}$. Recalling (2), it suffices to show that $\dim(\mathrm{Span}\{B_S\}) = \dim(\mathrm{Span}\{A_{\pi^{-1}(S_1)\cap\pi^{-1}(S_2)}\})$. Since every $k-1$ columns of B are linearly independent, we know $\dim(\mathrm{Span}\{B_S\}) = k-1$. Since $\Delta_{k-1} < 1$, it follows from (22) and Lemma 6 that $\dim(\mathrm{Span}\{A_{\pi^{-1}(S_1)\cap\pi^{-1}(S_2)}\}) \geq k-1$, and the number of elements in $\pi^{-1}(S_1) \cap \pi^{-1}(S_2)$ is then either $k-1$ or k. Knowing $\pi^{-1}(S_1) \neq \pi^{-1}(S_2)$, it must be that $|\pi^{-1}(S_1) \cap \pi^{-1}(S_2)| = k-1$; hence, since every $k-1$ columns of A are linearly independent, $\dim(\mathrm{Span}\{A_{\pi^{-1}(S_1)\cap\pi^{-1}(S_2)}\}) = \dim(\mathrm{Span}\{B_S\}) = k-1$.

The association $\gamma : S \mapsto \pi^{-1}(S_1) \cap \pi^{-1}(S_2)$ thus defines a function $\gamma : \binom{[m]}{k-1} \to \binom{[m]}{k-1}$ with $\Theta(\mathrm{Span}\{B_S\}, \mathrm{Span}\{A_{\gamma(S)}\} \leq f_{k-1}(\Delta) < \Delta_{k-1}$. We now show that γ is injective, which implies that $\tau = \gamma^{-1}$ is the map desired for the induction. Suppose $\gamma(S) = \gamma(S') = S^*$ for some $S, S' \in \binom{[m]}{k-1}$. By the triangle inequality,

$$\Theta(\mathrm{Span}\{B_S\}, \mathrm{Span}\{B_{S'}\}) \leq \Theta(\mathrm{Span}\{B_S\}, \mathrm{Span}\{A_{S^*}\}) + \Theta(\mathrm{Span}\{A_{S^*}\}, \mathrm{Span}\{B_{S'}\}) \leq 2f_{k-1}(\Delta) \tag{23}$$

Recalling (20), we have:

$$f_{k-1}(\Delta) + f_k(\Delta) = f_k(\Delta)\left[1 + \frac{\cos\theta_k + \sqrt{2-\cos^2\theta_k}}{1-\cos^2\theta_k}\right] < \Delta_k\left[1 + \frac{\cos\theta_k + \sqrt{2-\cos^2\theta_k}}{1-\cos^2\theta_k}\right] = \Delta_{k-1} < \frac{\alpha}{2} \tag{24}$$

Hence, we have $2f_{k-1}(\Delta) < \alpha - 2f_k(\Delta)$, implying by (23) that $\Theta(\mathrm{Span}\{B_S\}, \mathrm{Span}\{B_{S'}\}) < \alpha - 2f_k(\Delta)$. Recalling (18), we have by Lemma 9 (setting $\ell = k-1$) **NEEDS PROOF** that $S = S'$. Thus, γ is injective. ■

Lemma 2: Fix positive integers $n < m$ and let $A, B \in \mathbb{R}^{n\times m}$ with A having the $(2,\alpha)$-lower-RIP and unit norm columns. If there exists a map $\pi : [m] \to [m]$ such that for all $i \in \{1, ..., m\}$,

$$\Theta\left(\mathrm{Span}\{Ae_i\}, \mathrm{Span}\{Be_{\pi(i)}\}\right) \leq \Delta \quad \text{for some} \quad \Delta < \frac{\alpha}{\sqrt{2}} \tag{25}$$

then there exist a permutation matrix $P \in \mathbb{R}^{m\times m}$ and an invertible diagonal matrix $D \in \mathbb{R}^{m\times m}$ such that $b_i = PDa_i$ and $\|A_i - BPD_i\|_2 \leq \Delta$ for all $i \in [m]$.

Proof of Lemma 2: We will show that π is injective (and thus a permutation) by supposing $\pi(i) = \pi(j)$ for some $i \neq j \in [m]$ and reaching a contradiction. It follows from (25) that for all basis vectors e_i we have $d(Ae_i, \mathrm{Span}\{Be_{\pi(i)}\}) \leq \Delta$. Equivalently, for any $a_1 = c_1 e_i \in \mathbb{R}^m$ there exists some $b_1 = \tilde{c}_1 e_{\pi(i)} \in \mathbb{R}^m$ such that

$$\|Aa_1 - Bb_1\|_2 = \|A(c_1 e_i) - B(\tilde{c}_1 e_{\pi(i)})\|_2 = \|c_1 Ae_i - \tilde{c}_1 Be_{\pi(j)}\|_2 \leq \Delta|c_1|, \tag{26}$$

where we have used $\pi(i) = \pi(j)$. Similarly, for any $a_2 = c_2 e_j \in \mathbb{R}^m$ there exists some $b_1 = \tilde{c}_2 e_{\pi(j)} \in \mathbb{R}^m$ such that

$$\|Aa_2 - Bb_2\|_2 = \|A(c_2 e_j) - B(\tilde{c}_2 e_{\pi(j)})\|_2 = \|c_2 Ae_j - \tilde{c}_2 Be_{\pi(j)}\|_2 \leq \Delta|c_2|. \tag{27}$$

Note that for $c_1 \neq 0$, if $\tilde{c}_1 = 0$ then equation (26) implies that $|c_1| = \|Aa_1\|_2 \leq \Delta|c_1|$, which is impossible since $\Delta < 1$; likewise, if $c_2 \neq 0$ then $\tilde{c}_2 \neq 0$ as well. Scaling equation (26) by $\|b_2\|_1 = |\tilde{c_2}|$, we get

$$|\tilde{c_2}|\|c_1 Ae_i - \tilde{c}_1 Be_{\pi(j)}\|_2 = \|c_1\tilde{c_2}Ae_i - \tilde{c}_1\tilde{c_2}Be_{\pi(j)}\|_2 \leq \Delta|c_1||\tilde{c_2}|, \tag{28}$$

whereas scaling equation (27) by $\|b_2\|_1 = |\tilde{c}_2|$ we have

$$|\tilde{c_1}|\|c_2 Ae_j - \tilde{c}_2 Be_{\pi(j)}\|_2 = \|c_2\tilde{c}_1 Ae_j - \tilde{c}_1\tilde{c}_2 Be_{\pi(j)}\|_2 \leq \Delta|\tilde{c_1}||c_2|. \tag{29}$$

Summing (28) and (29) and applying the triangle inequality, we get

$$\begin{aligned}\Delta(|c_1||\tilde{c_2}| + |\tilde{c_1}||c_2|) &\geq \|c_1\tilde{c}_2 Ae_i - c_2\tilde{c}_1 Ae_j\|_2 \\ &\geq \alpha\|c_1\tilde{c}_2 e_i - c_2\tilde{c}_1 e_j\|_2 \\ &\geq \frac{\alpha}{\sqrt{2}}(|c_1||\tilde{c_2}| + |c_2||\tilde{c}_1|),\end{aligned}$$

where we have also applied the $(2,\alpha)$-lower-RIP on A. This is in contradiction with (25) which states that $\Delta < \frac{\alpha}{\sqrt{2}}$. Hence, π must be injective and the matrix $P \in \mathbb{R}^{m\times m}$ whose i-th column is $e_{\pi(i)}$ for all $i \in [m]$ is a permutation matrix. For any set of $a_i = c_i e_i \neq 0$, if we let $D \in \mathbb{R}^{m\times m}$ be the (invertible) diagonal matrix with corresponding elements $\frac{\tilde{c}_1}{c_1}, ..., \frac{\tilde{c}_m}{c_m}$, we then have that $b_i = \tilde{c}_i e_{\pi(i)} = PD(c_i e_i) = PDa_i$ for all $i \in [m]$. Furthermore, proximity condition (**??**) becomes $||(A - BPD)e_i|| \leq \Delta$ for all $i \in [m]$ or, more generally, $||(A - BPD)x|| \leq \Delta||x||$ for all 1-sparse $x \in \mathbb{R}^m$. ■

Lemma 3: Suppose $M \in \mathbb{R}^{n\times m}$ satisfies the $(\ell+1,\alpha)$-lower-RIP. Then for all $S_1, S_2 \in \binom{[m]}{\ell}$,

$$\Theta(\text{Span}\{M_{S_1}\}, \text{Span}\{M_{S_2}\}) < \alpha \implies S_1 = S_2. \tag{30}$$

Proof of Lemma 3: If $\ell = m$ then the result is vacuously true. Suppose $S_1 \neq S_2 \in \binom{[m]}{\ell}$ for some $\ell < m$ and let $r \in S_1 \setminus S_2$. Since M satisfies the $(\ell+1,\alpha)$-lower-RIP we have we have $\dim(M_{S_1}) = \dim(M_{S_2})$, hence by definition of the gap metric,

$$\Theta(\text{Span}\{M_{S_1}\}, \text{Span}\{M_{S_2}\}) = \sup_{\substack{z\in\text{Span}\{M_{S_1}\}\\ \|z\|_2=1}} d(z, \text{Span}\{M_{S_2}\}).$$

Since $Me_r \in \text{Span}\{M_{S_1}\}$ and M has unit norm columns,

$$\sup_{\substack{z\in\text{Span}\{M_{S_1}\}\\ \|z\|_2=1}} d(z, \text{Span}\{M_{S_2}\}) \geq d(Me_r, \text{Span}\{M_{S_2}\}).$$

Finally, by the $(\ell+1,\alpha)$-lower-RIP on M and noting that $e_r \in \text{Span}\{e_i : i \in S_2\}^\perp$,

$$\begin{aligned}d(Me_r, \text{Span}\{M_{S_2}\}) &= \inf\{\|Me_r - Mx\|_2 : x \in \text{Span}\{e_i : i \in S_2\}\} \\ &\geq \inf\{\alpha\|e_r - x\|_2 : x \in \text{Span}\{e_i : i \in S_2\}\} \\ &= \inf\{\alpha\sqrt{1 + \|x\|_2^2} : x \in \text{Span}\{e_i : i \in S_2\}\} \\ &= \alpha,\end{aligned}$$

which is the contrapositive of the assertion. ■.

Lemma 4: Let $x \in \mathbb{R}^m$ and suppose V, W are linear subspaces of $\mathbb{R}^m$. Suppose $d(x,V) \leq d(x,W) \leq \Delta$. Then

$$d(x, V\cap W) \leq \Delta\left(\frac{\cos\theta_F + \sqrt{2 - \cos^2\theta_F}}{1 - \cos^2\theta_F}\right) \tag{31}$$

where $\theta_F]in[0, \frac{\pi}{2}]$ is the Friedrichs angle between V and W.

Proof of Lemma 4: Recall that for all $x \in \mathbb{R}^m$, $d(x,U) = \|x - \Pi_U x\|$ for all subspaces $U \subseteq \mathbb{R}^m$. Since $\Pi_{V\cap W}x \in W$ for all $x \in \mathbb{R}^m$, we have by Pythagoras' theorem that

$$d(x, V\cap W)^2 = \|x - \Pi_{V\cap W}x\|^2 = \|x - \Pi_W x\|^2 + \|\Pi_W x - \Pi_{V\cap W}x\|^2. \tag{32}$$

The first term on the RHS of (32) is $d(x,W)$. Applying the triangle inequality to the second term,

$$\|\Pi_W x - \Pi_{V\cap W}x\| \leq \|\Pi_W x - \Pi_W\Pi_V x\| + \|\Pi_W\Pi_V x - \Pi_{V\cap W}x\|. \tag{33}$$

The first term on the RHS of (33) can be bounded as follows: $\|\Pi_W x - \Pi_W\Pi_V x\| = \|\Pi_W(I - \Pi_V)x\| \leq \|x - \Pi_V x\| = d(x,V)$. This is because for any projection matrix Π and for all $x \in \mathbb{R}^m$ we have $\langle\Pi x, \Pi x - x\rangle = 0$, hence $\|\Pi x\|^2 = |\langle\Pi x, \Pi x\rangle| = |\langle\Pi x, x\rangle + \langle\Pi x, \Pi x - x\rangle| \leq \|\Pi x\|\|x\|$ by the Cauchy-Schwartz inequality. To bound the second term, we make use of the following result by [Deutsch, "Best Approximation in Inner Product Spaces, Lemma 9.5(7)"]:

$$\|(\Pi_W\Pi_V)x - \Pi_{V\cap W}x\| \leq \cos\theta_F\|x\| \quad \text{for all} \quad x \in \mathbb{R}^m. \tag{34}$$

First, note that

$$\begin{aligned}\|(\Pi_W\Pi_V)(x - \Pi_{V\cap W}x) - \Pi_{V\cap W}(x - \Pi_{V\cap W}x)\| &= \|\Pi_W\Pi_V x - \Pi_W\Pi_V\Pi_{V\cap W}x - \Pi_{V\cap W}x + \Pi_{V\cap W}^2 x\| \\ &= \|(\Pi_W\Pi_V)x - \Pi_{V\cap W}x\|\end{aligned} \tag{35}$$

since $\Pi_V \Pi_{V\cap W} = \Pi_W \Pi_{V\cap W} = \Pi_{V\cap W}$ and $\Pi_{V\cap W}^2 = \Pi_{V\cap W}$ (all projection matrices are idempotent). We then have by (35) and (34) that

$$\begin{aligned}\|(\Pi_W\Pi_V)x - \Pi_{V\cap W}x\| &= \|(\Pi_W\Pi_V)(x - \Pi_{V\cap W}x) - \Pi_{V\cap W}(x - \Pi_{V\cap W}x)\| \\ &\le \cos\theta_F \|x - \Pi_{V\cap W}x\| \\ &= d(x, V\cap W)\cos\theta_F\end{aligned}$$

It follows from this, (32), (33) and the assumption $d(x,V) \le d(x,W) \le \Delta$ that

$$\begin{aligned}d(x, V\cap W)^2 &\le d(x,W)^2 + \left[d(x,V) + d(x,V\cap W)\cos\theta_F\right]^2 \\ &\le \Delta^2 + \left[\Delta + d(x,V\cap W)\cos\theta_F\right]^2\end{aligned}$$

which can be rearranged into the following quadratic inequality in $d(x, V\cap W)$:

$$d(x,V\cap W)^2\left(1 - cos^2\theta_F\right) - d(x,V\cap W)2\Delta\cos\theta_F - 2\Delta^2 \le 0 \tag{36}$$

The zeros of the LHS are

$$\begin{aligned}d(x,V\cap W)_\pm &= \frac{2\Delta\cos\theta_F \pm \sqrt{4\Delta^2\cos^2\theta_F - 4\left(1-\cos^2\theta_F\right)\left(-2\Delta^2\right)}}{2\left(1-\cos^2\theta_F\right)} \\ &= \Delta\left(\frac{\cos\theta_F \pm \sqrt{2-\cos^2\theta_F}}{1-\cos^2\theta_F}\right),\end{aligned}$$

of which, for all $\theta_F \in [0, \frac{\pi}{2}]$, only $d(x,V\cap W)_+$ is positive. Hence (36) implies that

$$0 \le d(x,V\cap W) \le \Delta\left(\frac{\cos\theta_F + \sqrt{2-\cos^2\theta_F}}{1-\cos^2\theta_F}\right). \quad \blacksquare$$

Lemma 5: Let $M \in \mathbb{R}^{n\times m}$. If every $2k$ columns of M are linearly independent, then for $S, S' \in \binom{[m]}{k}$,

$$\text{Span}\{M_{S\cap S'}\} = \text{Span}\{M_S\} \cap \text{Span}\{M_{S'}\} \tag{37}$$

Lemma 6: Let V, W be subspaces of $\mathbb{R}^m$ and suppose that for all $v \in V$ there exists some $w \in W$ such that

$$\|v - w\|_2 < \|v\|_2. \tag{38}$$

Then $\dim(W) \ge \dim(V)$.

Proof of Lemma 6: If $\dim(W) < \dim(V)$ then $V \cap W^\perp \neq \emptyset$, but for all $v \in V\cap W^\perp$ we would have that $\|v-w\|_2^2 = \|v\|_2^2 + \|w\|_2^2 \ge \|v\|_2^2$ for all $w \in W$, which is in contradiction with (38). $\blacksquare$.

***Note:* I found an equivalent statement in the literature (Corollary 2.6 in Kato, knowing also that the gap function is a metric since the ambient space is a Hilbert space (see footnote 1 p. 196))**.

Lemma 7: Fix matrices $A, \tilde{A} \in \mathbb{R}^{n\times m}$ where $\tilde{A} = AE$ for some invertible diagonal matrix $E = \text{diag}(\lambda_i) \in \mathbb{R}^{m\times m}$, $\lambda_i \in \mathbb{R}$ for all $i \in [m]$. If there exists a matrix $B \in \mathbb{R}^{n\times m}$ such that $\|(A-B)e_i\| \le \varepsilon$ for all $i \in [m]$, then the matrix $\tilde{B} = BE$ satisfies $\|(\tilde{A} - \tilde{B})e_i\| \le \lambda\varepsilon$ for all $i \in [m]$, where $\lambda = \max_i |\lambda_i|$.

This lemma allows us to extend uniqueness guarantees (up to permutation, scaling, and error) for matrices with unit norm columns to those without and vice versa.

Proof of Lemma 7: For all $i \in [m]$, we have:

$$\|(\tilde{A} - \tilde{B})e_i\| = \|(A-B)Ee_i\| = |\lambda_i|\|(A-B)e_i\| \le |\lambda_i|\varepsilon \le \lambda\varepsilon \quad \blacksquare$$

IV. Can we have $N < k\binom{m}{k}$?

The proof relies on the existence of a set of k-sparse $a_i \in \mathbb{R}^m$ in general linear position (every k of them are linearly independent) such that for any given support $S \in \binom{[m]}{k}$ there are $k\binom{m}{k}$ vectors a_i for which $\text{supp}(a_i) \subseteq S$. The very general construction supplied in the proof invokes $N = k\binom{m}{k}^2$ vectors a_i to satisfy this property, but we can reduce N by explicitly including vectors with supports of cardinality less than k. For instance, the proof holds for any set of $N = m + \left[k\binom{m}{k} - k\right]\binom{m}{k}$ vectors a_i in general linear position if it contains the elementary basis vectors e_j for $j \in [m]$. To see why, consider that for each support $S \in \binom{[m]}{k}$ there are k basis vectors e_j for which $\text{supp}(e_j) \subseteq S$. Hence, by constructing for every $S \in \binom{[m]}{k}$ an additional $k\binom{m}{k} - k$ vectors a_i in general linear position for which $\text{supp}(a_i) = S$, we satisfy the required property. (Need to show the union of the a_i and e_j will still be in general linear position...substituting e_j for any a_j in the Vandermonde matrix shouldn't nullify it.)

Alternatively, we can argue for $N < k\binom{m}{k}^2$ as follows. (For simplicity consider the noiseless case.) Construct $k\binom{m}{k}$ k-sparse vectors a_i sharing support $S_1 \in \binom{[m]}{k}$. Then there exist at least k k-sparse vectors b_i in the alternate factorization sharing some support $S_1' \in \binom{[m]}{k}$ and it follows that $\text{Span}\{A_{S_1}\} = \text{Span}\{B_{S_1'}\}$. Suppose now we were to construct $k\binom{m}{k}$ k-sparse vectors a_i

sharing support $S_2 \in \binom{[m]}{k}$ for some $S_2 \neq S_1$. Then we know that there can't be k b_i in the alternate factorization sharing support S_1', since this implies $\text{Span}\{A_{S_1}\} = \text{Span}\{B_{S_1'}\} = \text{Span}\{A_{S_2}\}$ which contradicts the spark condition. So given that there are $k\binom{m}{k}$ vectors a_i sharing support S_1 we need not invoke just as many a_i sharing support S_2 to argue by the pigeon-hole principle that there are at least k alternate vectors b_i sharing some support $S_2' \in \binom{[m]}{k}$. $\text{Span}\{A_{S_1 \cap S_2}\} = \text{Span}\{A_{S_1}\} \cap \text{Span}\{A_{S_2}\}$ plays into this. This may not affect the deterministic theorem, but could be incorporated into the probabilistic extensions.

V. Scrap Paper

Lemma 8: Let U, V, W be subspaces of $\mathbb{R}^m$ such that $V \cap U = W \cap U = 0$. Then $\Theta(V \cup U, W \cup U) \leq \Theta(V, W)$.

Proof of Lemma 8: Let $v \in V \cup U$. Since $V \cap U = 0, v = v_V + v_U$ is a unique decomposition. By definition of Θ, we have

$$\begin{aligned}\Theta(V \cup U, W \cup U) &= \sup_{\substack{v \in V \cup U \\ \|v\|=1}} \inf \{\|v - w\| : w \in W \cup U\} &(39)\\ &= \sup_{\substack{v \in V \cup U \\ \|v\|=1}} \inf \{\|v_V + v_U - w_W - w_U\| : w \in W \cup U\} &(40)\end{aligned}$$

$$\Theta(V \cup U, W \cup U) = \|\Pi_{V \cup U} - \Pi_{W \cup U}\| \tag{41}$$

Lemma 9: Let $M \in \mathbb{R}^{n \times m}$. If for some $\ell < m$, every set of ℓ columns of M are linearly independent and $\Theta(\text{Span}\{M_{S_1}\}, \text{Span}\{M_{S_2}\}) \geq \Delta$ for all $S_1 \neq S_2 \in \binom{[m]}{\ell}$, then for any $S, S' \in \binom{[m]}{\ell-1}$,

$$\Theta(\text{Span}\{B_S\}, \text{Span}\{B_{S'}\}) < \Delta \implies S = S'. \tag{42}$$

Proof of Lemma 9: Since $\ell < m - 1$ and any two distinct supports of cardinality ℓ can share at most $\ell - 1$ indices, we have two possible scenarios given $S \neq S'$: either there exists an index $r' \notin S \cup S'$ such that $S \cup r' \neq S' \cup r' \in \binom{[m]}{\ell+1}$, or there exist a pair of indices $r \in S \setminus S'$ and $r' \in S' \setminus S$ such that $S \cup r' \neq S' \cup r \in \binom{[m]}{\ell+1}$. (*Proof:* If $S \neq S' \in \binom{[m]}{\ell}$ then there must exist some $p \in S \setminus S'$ and $q \in S' \setminus S$. Suppose that $S \cup q = S' \cup p$. Then $S = S' \cup \{p\} \setminus \{q\}$, and $|S' \cap S| = |S' \cap (S' \cup \{p\} \setminus \{q\})| = |S' \setminus \{q\}| = \ell - 1$. Hence $|S \cup S'| = \ell + 1 < m$ and there must exist some $r \in [m] \setminus (S \cup S')$; it then follows from $S \neq S'$ that $S \cup r \neq S' \cup r$.)

Consider now the first scenario. Suppose $v \in \text{Span}\{M_S, M_{r'}\}$. Since every $\ell + 1$ columns of M are linearly independent, we can write $v = v_S + cMe_{r'}$ for some unique $v_S \in \text{Span}\{M_S\}$ and $c \in \mathbb{R}$. Likewise for any $v' \in \text{Span}\{M_{S'}, M_{r'}\}$ we can write $v' = v'_{S'} + c'Me_{r'}$ for some unique $v'_{S'} \in \text{Span}\{M_{S'}\}$ and $c' \in \mathbb{R}$. We then have

$$\begin{aligned}d(v, \text{Span}\{M_{S'}, M_{r'}\}) &= \inf \{\|v - v'\|_2 : v' \in \text{Span}\{M_{S'}, M_{r'}\}\} \\ &\leq \inf \{\|v_S - v'_{S'}\|_2 + \|(c - c')Me_{r'}\|_2 : v'_{S'} \in \text{Span}\{M_{S'}\}, c' \in \mathbb{R}\} \\ &= \inf \{\|v_S - v'_{S'}\|_2 : v'_{S'} \in \text{Span}\{M_{S'}\}\} + \inf \{|c - c'|\|Me_{r'}\|_2 : c' \in \mathbb{R}\}\end{aligned}$$

By definition of the gap metric,

$$\Theta(\text{Span}\{M_S, M_{r'}\}, \text{Span}\{M_{S'}, M_r\}) = \sup \inf \{\|v_S - v'_{S'}\|_2 : v'_{S'} \in \text{Span}\{M_{S'}\}\} + \inf \{|c - c'|\|Me_{r'}\|_2 : c' \in \mathbb{R}\} \tag{43}$$

$$\tag{44}$$

We now consider the second scenario. Suppose again that $v \in \text{Span}\{M_S, M_{r'}\}$ and write $v = v_S + cMe_{r'}$ for unique $v_S \in \text{Span}\{M_S\}$ and $c \in \mathbb{R}$. Keeping in mind that $r' \in S'$, we have

$$\begin{aligned}d(v, \text{Span}\{M_{S'}, M_r\}) &= \inf \{\|v - v'\|_2 : v' \in \text{Span}\{M_{S'}, M_r\}\} \\ &\leq \inf \{\|v - v'\|_2 : v' \in \text{Span}\{M_{S'}\}\} \\ &= \inf \{\|v_S + cMe_{r'} - v' - c'Me_{r'}\|_2 : v' \in \text{Span}\{M_{S'}\}, c' \in \mathbb{R}\} \\ &\leq \inf \{\|v_S - v'\|_2 + \|(c - c')Me_{r'}\|_2 : v' \in \text{Span}\{M_{S'}\}, c' \in \mathbb{R}\} \\ &= \inf \{\|v_S - v'\|_2 : v' \in \text{Span}\{M_{S'}\}\} + \inf \{|c - c'|\|Me_{r'}\|_2 : c' \in \mathbb{R}\} \\ &= d(v_S, \text{Span}\{M_{S'}\})\end{aligned}$$

We can now proceed to argue for both scenarios simultaneously (in what follows, for the first scenario set $r = r'$), given that in either case we have $r' \notin S$. By definition of the gap metric,

$$\begin{aligned}\Theta(\text{Span}\{M_S, M_{r'}\}, \text{Span}\{M_{S'}, M_r\}) &= \sup\{d(v, \text{Span}\{M_{S'}, M_r\}) : v \in \text{Span}\{M_S, M_{r'}\}, \|v\| = 1\} \\ &= \sup_{\substack{v_S \in \text{Span}\{M_S\} \\ c \in \mathbb{R} \\ \|v_S + cMe_{r'}\| = 1}} \inf\{\|v_S - v'_{S'}\|_2 : v'_{S'} \in \text{Span}\{M_{S'}\}\} \\ &+ \sup_{\substack{v_S \in \text{Span}\{M_S\} \\ c \in \mathbb{R} \\ \|v_S + cMe_{r'}\| = 1}} \inf\{|c - c'|\|Me_{r'}\|_2 : c' \in \mathbb{R}\} \\ &\leq \sup\{d(v_S, \text{Span}\{M_{S'}\}) : v_S \in \text{Span}\{M_S\}, c \in \mathbb{R}, \|v_S + cMe_{r'}\| = 1\} \\ &= \sup\{d(v_S, \text{Span}\{M_{S'}\}) : v_S \in \text{Span}\{M_S\}, \|v_S\| \leq 1\} \\ &= \sup\{d(v_S, \text{Span}\{M_{S'}\}) : v_S \in \text{Span}\{M_S\}, \|v_S\| = 1\} \\ &= \Theta(\text{Span}\{M_S\}, \text{Span}\{M_{S'}\}).\end{aligned}$$

where the second to last equality is due to the fact that every $\ell + 1$ columns of M are linearly independent, since then $\|v_S + cMe_{r'}\| = 1 \iff \|v_S\| \leq 1$ **SHIT...not true**. The last equality is due to the fact that for any subspace $W \subseteq \mathbb{R}^m$ we have $\|x\| \leq 1 \implies d(x, W) \leq d(\frac{x}{\|x\|}, W)$ for all $x \in \mathbb{R}^m$. Hence,

$$S \neq S' \implies \Theta(\text{Span}\{M_S\}, \text{Span}\{M_{S'}\}) \geq \Theta(\text{Span}\{M_S, M_r\}, \text{Span}\{M_{S'}, M_r\}) \geq \Delta, \tag{45}$$

which is the contrapositive of the result. ■