Chaz's Theorem: The Return of Hillar
Robust Identifiability in Sparse Dictionary Learning

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 fragment.

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 on subspaces of $\mathbb{R}^m$ is [see Theory of Linear Operators in a Hilbert Space p. 69 who cites first reference]

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

We note the following useful fact [ref: Morris, Lemma 3.3]:

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

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

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

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

Theorem 1: Fix $k \leq n < m$ and $\alpha \in (0, 1]$. There exist $N = k\binom{m}{k}^2$ k-sparse vectors $\mathbf{a}_1, \ldots, \mathbf{a}_N \in \mathbb{R}^m$ and $C > 0$ such that if $Y = \{\mathbf{y}_1, ..., \mathbf{y}_N\}$ is a dataset for which, for some $A \in \mathbb{R}^{n \times m}$ with unit norm columns satisfying the $(2k, \alpha)$-lower-RIP, $\|\mathbf{y}_i - A\mathbf{a}_i\|_2 \leq \varepsilon$ for all $i \in \{1, \ldots, N\}$, then the following proposition is true: any matrix $B \in \mathbb{R}^{n \times m}$ with unit norm columns satisfying the (k, α)-lower-RIP and for which $\|\mathbf{y}_i - B\mathbf{b}_i\| \leq \varepsilon$ for some k-sparse $\mathbf{b}_i \in \mathbb{R}^m$ for all $i \in \{1, \ldots, N\}$ is such that $\|(A - BPD)\mathbf{e}_i\| \leq C\varepsilon$ for some permutation matrix $P \in \mathbb{R}^m$ and invertible diagonal matrix $D \in \mathbb{R}^m$, provided ε is small enough.

Remark 1: : The assumption that the matrix B satisfy the (k, α)-lower-RIP allows us to place an upper bound on C in terms of the given variables, but the existence of such a $C > 0$ is not predicated on this assumption. To see why, note that when proving the bijectivity of π in Lemma 1 we do not require this property, and from this bijectivity it follows that every k columns of B are linearly independent. The same arguments made to demonstrate (4) can then be used to show that B necessarily satisfies the (k, β)-lower-RIP for some $\beta > 0$. One may therefore apply the Lemma with the substitution $\alpha \mapsto \min(\alpha, \beta)$.

Proof of Theorem 1: First, we produce a set of $N = k\binom{m}{k}^2$ vectors in $\mathbb{R}^k$ in general linear position (i.e. any set of k of them are linearly independent). Specifically, let $\gamma_1, ..., \sigma_N$ be any distinct numbers. Then the columns of the $k \times N$ matrix $V = (\gamma_j^i)_{i,j=1}^{k,N}$ 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 $\mathbf{a}_1, \ldots, \mathbf{a}_N \in \mathbb{R}^m$ by setting the nonzero values of vector $\mathbf{a}_i$ to be those contained in the ith column of V while partitioning the $\mathbf{a}_i$ evenly among the $\binom{m}{k}$ possible supports.

We will show how the existence of these $\mathbf{a}_i$ proves the theorem. First, we claim that there exists some $\delta > 0$ such that for any set of k vectors $\mathbf{a}_{i_1}, ..., \mathbf{a}_{i_k}$, the following property holds:

$$\| \sum_{j=1}^{k} c_j \mathbf{a}_{i_j} \|_2 \geq \sigma \|c\|_1 \quad \forall c = (c_1, ..., c_k) \in \mathbb{R}^m. \tag{4}$$

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

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

By general linear position of the $\mathbf{a}_i$, 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 $\phi(c) \geq \sigma$ 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 $\mathbf{a}_i$ (actually, for our purposes we need only consider those subsets of k vectors $\mathbf{a}_i$ having the same support), hence there is some minimal σ satisfying (4) for all of them. We refer the reader to the Appendix for a lower bound on σ in terms of k and the sequence $\gamma_1, \ldots, \gamma_N$.

Now suppose that $Y = \{\mathbf{y}_1, \ldots, \mathbf{y}_N\}$ is a dataset for which for all $i \in \{1, \ldots, N\}$ we have $\|\mathbf{y}_i - A\mathbf{a}_i\| \leq \varepsilon$ for some $A \in \mathbb{R}^{n \times m}$ with unit norm columns satisfiying the $(2k, \alpha)$-lower-RIP and that for some alternate $B \in \mathbb{R}^{n \times m}$ there exist k-sparse $\mathbf{b}_i \in \mathbb{R}^m$ for which $\|\mathbf{y}_i - B\mathbf{b}_i\| \leq \varepsilon$ for all $i \in \{1, \ldots, N\}$. Since there are $k\binom{m}{k}$ vectors $\mathbf{a}_i$ with a given support $S \in \binom{[m]}{k}$, the pigeon-hole principle implies that there are at least k vectors $\mathbf{y}_i$ such that $\|\mathbf{y}_i - A\mathbf{a}_i\| \leq \varepsilon$ for these $\mathbf{a}_i$ and also $\|\mathbf{y}_i - B\mathbf{b}_i\| \leq \varepsilon$ for $\mathbf{b}_i$ all with supports contained in some $S' \in \binom{[m]}{k}$. Let $\mathcal{Y} = \{\mathbf{y}_i : i \in \mathcal{I}\}$ be a set of k such vectors $\mathbf{y}_i$ indexed by $\mathcal{I}$.

Note that any matrix satisfying the (ℓ, α)-lower-RIP is such that any ℓ of its columns are linearly independent. It follows from this and the general linear position of the $\mathbf{a}_i$ that the set $\{A\mathbf{a}_i : i \in \mathcal{I}\}$ is a basis for $\text{Span}\{A_S\}$. Hence, fixing $\mathbf{z} \in \text{Span}\{A_S\}$, there exists a unique set of $c_i \in \mathbb{R}$ (for notational convenience we index the c_i with $\mathcal{I}$ as well) such that $\mathbf{z} = \sum_{i \in \mathcal{I}} c_i A\mathbf{a}_i$. Letting $\mathbf{y} = \sum_{i \in \mathcal{I}} c_i \mathbf{y}_i \in \text{Span}\{\mathcal{Y}\}$, we have by the triangle inequality that

$$\|\mathbf{z} - \mathbf{y}\|_2 = \| \sum_{i \in \mathcal{I}} c_i A\mathbf{a}_i - \sum_{i \in \mathcal{I}} c_i \mathbf{y}_i \|_2 \leq \sum_{i \in \mathcal{I}} \|c_i(A\mathbf{a}_i - \mathbf{y}_i)\|_2 = \sum_{i \in \mathcal{I}} |c_i| \|A\mathbf{a}_i - \mathbf{y}_i\|_2 \leq \varepsilon \sum_{i \in \mathcal{I}} |c_i|. \tag{5}$$

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

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

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

$$\|\mathbf{z}\|_2 = \| \sum_{i \in \mathcal{I}}^{k} c_i A\mathbf{a}_i \|_2 = \|A(\sum_{i \in \mathcal{I}} c_i \mathbf{a}_i)\|_2 \geq \alpha \| \sum_{i \in \mathcal{I}}^{k} c_i \mathbf{a}_i \|_2 \geq \alpha\sigma \sum_{i \in \mathcal{I}}^{k} |c_i|. \tag{7}$$

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

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

It follows that $d(\mathbf{z}, \text{Span}\{B_{S'}\}) \leq \tilde{C}\varepsilon$ for all unit vectors $\mathbf{z} \in \text{Span}\{A_S\}$. Hence,

$$\sup_{\substack{\mathbf{z} \in \text{Span}\{A_S\} \\ \|\mathbf{z}\|=1}} d(\mathbf{z}, \text{Span}\{B_{S'}\}) \leq \tilde{C}\varepsilon. \tag{8}$$

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 $\dim(\text{Span}\{B_{S'}\}) = \dim(\text{Span}\{A_S\})$ and, recalling (2), that $\Theta(\text{Span}\{A_S\}, \text{Span}\{\mathcal{B}_{S'}\}) \leq \tilde{C}\varepsilon$. Specifically, letting $\theta_j \in [0, \frac{\pi}{2}]$ be the least of all Friedrichs angles formed between pairs of subspaces for which j columns of A form a basis, if

$$\varepsilon < \frac{\alpha^2\sigma}{2\sqrt{2}} \prod_{j=1}^{k} \frac{1 - \cos^2\theta_j}{\cos\theta_j + \sqrt{2 - \cos^2\theta_j}} \tag{9}$$

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

$$\Theta(\text{Span}\{A_S\}, \text{Span}\{\mathcal{B}_{\pi(S)}\}) \leq \tilde{C}\varepsilon < \frac{\alpha}{\sqrt{2}} \prod_{j=1}^{k} \frac{1-\cos^2\theta_j}{\cos\theta_j + \sqrt{2-\cos^2\theta_j}} \quad \text{for all} \quad S \in \binom{[m]}{k}. \tag{10}$$

The result then follows by Lemma 1, yielding

$$C = \frac{2}{\alpha\sigma} \prod_{j=1}^{k} \frac{\cos\theta_j + \sqrt{2-\cos^2\theta_j}}{1-\cos^2\theta_j}. \quad \blacksquare \tag{11}$$

Lemma 1 (Main Lemma): Fix positive integers $k \leq n < m$ and let $A \in \mathbb{R}^{n\times m}$ be a matrix having unit norm columns satisfying the $(2k, \alpha)$-lower-RIP. 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 and let

$$f_\ell(A) = \prod_{j=1}^{\ell} \frac{1-\cos^2\theta_j}{\cos\theta_j + \sqrt{2-\cos^2\theta_j}}. \tag{12}$$

If $B \in \mathbb{R}^{n\times m}$ is a matrix with unit norm columns satisfying the (k,α)-lower-RIP and there exists a map $\pi : \binom{[m]}{k} \to \binom{[m]}{k}$ and some $\Delta < \frac{\alpha}{\sqrt{2}} f_k(A)$ such that

$$\Theta(\text{Span}\{A_S\}, \text{Span}\{B_{\pi(S)}\}) \leq \Delta \quad \forall S \in \binom{[m]}{k}, \tag{13}$$

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

$$\|(A - BPD)e_i\|_2 \leq f_k(A)^{-1}\Delta \quad \forall i \in \{1,\ldots,m\}. \tag{14}$$

Proof of Lemma 1: We prove the following equivalent statement: If there exists a map $\pi : \binom{[m]}{k} \to \binom{[m]}{k}$ and $\Delta < \frac{\alpha}{\sqrt{2}}$ such that

$$\Theta(\text{Span}\{A_S\}, \text{Span}\{B_{\pi(S)}\}) \leq f_k(A)\Delta \quad \forall S \in \binom{[m]}{k}, \tag{15}$$

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 $i \in [m]$,

$$\|(A - BPD)\mathbf{e}_i\|_2 \leq \Delta. \tag{16}$$

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). 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

$$\Theta(\text{Span}\{A_{S_1}\}, \text{Span}\{A_{S_2}\}) \leq \Theta(\text{Span}\{A_{S_1}\}, \text{Span}\{B_{S^*}\}) + \Theta(\text{Span}\{B_{S^*}\}, \text{Span}\{A_{S_2}\}) \leq 2f_k(A)\Delta. \tag{17}$$

Since $\theta_j \in [0, \frac{\pi}{2}]$ for all $j \in [k]$ we have $f_k(A) < \left(\frac{1}{\sqrt{2}}\right)^k$. Hence by (17) we have $\Theta(\text{Span}\{A_{S_1}\}, \text{Span}\{A_{S_2}\}) < \alpha$ and it follows by Lemma 3 (setting $\ell = k+1$) that $S_1 = S_2$. Thus π is bijective.

We complete the proof of the lemma, inductively, by producing a map $\tau : \binom{[m]}{k-1} \to \binom{[m]}{k-1}$ (assuming $k \geq 2$) such that

$$\Theta(\text{Span}\{A_S\}, \text{Span}\{B_{\tau(S)}\}) \leq f_{k-1}(A)\Delta \quad \forall S \in \binom{[m]}{k-1}. \tag{18}$$

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 $\mathbf{z} \in \text{Span}\{B_{S_1}\} \cap \text{Span}\{B_{S_2}\}$ we have $d(\mathbf{z}, \text{Span}\{A_{\pi^{-1}(S_1)}\}) \leq f_k(A)\Delta$ and $d(\mathbf{z}, \text{Span}\{A_{\pi^{-1}(S_2)}\}) \leq f_k(A)\Delta$. It follows by Lemmas 4 and 5 that

$$d\left(\mathbf{z}, \text{Span}\{A_{\pi^{-1}(S_1)\cap\pi^{-1}(S_2)}\}\right) \leq \Delta f_k(A) \left(\frac{\cos\theta_k + \sqrt{2-\cos^2\theta_k}}{1-\cos^2\theta_k}\right) = f_{k-1}(A)\Delta \tag{19}$$

Since (19) holds for all unit vectors $\mathbf{z} \in \text{Span}\{B_{S_1}\} \cap \text{Span}\{B_{S_2}\} \supseteq \text{Span}\{B_S\}$, it follows that

$$\sup_{\substack{\mathbf{z}\in\text{Span}\{B_S\} \\ \|\mathbf{z}\|=1}} d\left(\mathbf{z}, \text{Span}\{A_{\pi^{-1}(S_1)\cap\pi^{-1}(S_2)}\}\right) \leq f_{k-1}(A)\Delta. \tag{20}$$

We will show that, in fact, $\Theta\left(\text{Span}\{B_S\}, \text{Span}\{A_{\pi^{-1}(S_1)\cap\pi^{-1}(S_2)}\}\right) \leq f_{k-1}(A)\Delta$. Recalling (2), it suffices to show that $\dim(\text{Span}\{B_S\}) = \dim(\text{Span}\{A_{\pi^{-1}(S_1)\cap\pi^{-1}(S_2)}\})$. Since every k columns of B are linearly independent, we know $\dim(\text{Span}\{B_S\}) = k-1$. Since $f_{k-1}(A)\Delta < 1$, it follows from (20) and Lemma 6 that $\dim(\text{Span}\{A_{\pi^{-1}(S_1)\cap\pi^{-1}(S_2)}\}) \geq k-1$, 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 $\dim(\text{Span}\{A_{\pi^{-1}(S_1)\cap\pi^{-1}(S_2)}\}) = \dim(\text{Span}\{B_S\}) = k-1$.

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 $\Theta(\text{Span}\{B_S\}, \text{Span}\{A_{\gamma(S)}\} \leq f_{k-1}(A)\Delta$. 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,

$$\Theta(\text{Span}\{B_S\}, \text{Span}\{B_{S'}\}) \leq \Theta(\text{Span}\{B_S\}, \text{Span}\{A_{S^*}\}) + \Theta(\text{Span}\{A_{S^*}\}, \text{Span}\{B_{S'}\}) \leq 2f_{k-1}(A)\Delta. \tag{21}$$

Since for $k \geq 2$ we have $2f_{k-1}(A)\Delta < \alpha$ and since B satisfies a (k,α)-lower-RIP with unit norm columns, we have by Lemma 3 (setting $\ell = k$) 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 $\pi : \{1,\ldots,m\} \to \{1,\ldots,m\}$ and some $\Delta < \frac{\alpha}{\sqrt{2}}$ such that

$$\Theta\left(\text{Span}\{Ae_i\}, \text{Span}\{B\mathbf{e}_{\pi(i)}\}\right) \leq \Delta \quad \text{for all} \quad i \in \{1,...,m\} \tag{22}$$

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 $\mathbf{b}_i = PD\mathbf{a}_i$ and $\|(A - BPD)\mathbf{e}_i\|_2 \leq \Delta$ for all $i \in \{1,\ldots,m\}$.

Proof of Lemma 2: We first note that since A has unit norm columns and all linear subspaces of $\mathbb{R}^m$ are closed, (22) implies that for all 1-sparse $\mathbf{a} \in \mathbb{R}^m$ with support $i \in \{1,\ldots,m\}$ there exists some 1-sparse $\mathbf{b} \in \mathbb{R}^m$ with support $\pi(i)$ such that

$$\|A\mathbf{a} - B\mathbf{b}\| \leq \Delta\|\mathbf{a}\|. \tag{23}$$

We will show that π is injective (and thus a permutation). Suppose that $\pi(i) = \pi(j) = \pi^*$ for some $i \neq j \in \{1,\ldots,m\}$. By (23), for any $\mathbf{a}_1 = c_1\mathbf{e}_i \in \mathbb{R}^m$ there exists some $\mathbf{b}_1 = \tilde{c}_1\mathbf{e}_{\pi(i)} \in \mathbb{R}^m$ such that

$$\|A\mathbf{a}_1 - B\mathbf{b}_1\|_2 = \|c_1A\mathbf{e}_i - \tilde{c}_1B\mathbf{e}_{\pi^*}\|_2 \leq \Delta|c_1|, \tag{24}$$

Similarly, for any $\mathbf{a}_2 = c_2\mathbf{e}_j \in \mathbb{R}^m$ there exists some $b_1 = \tilde{c}_2\mathbf{e}_{\pi(j)} \in \mathbb{R}^m$ such that

$$\|A\mathbf{a}_2 - B\mathbf{b}_2\|_2 = \|c_2A\mathbf{e}_j - \tilde{c}_2B\mathbf{e}_{\pi^*}\|_2 \leq \Delta|c_2|. \tag{25}$$

Note that for $c_1 \neq 0$, if $\tilde{c}_1 = 0$ then equation (24) implies that $|c_1| = \|A\mathbf{a}_1\|_2 \leq \Delta|c_1|$, which is impossible since $\Delta < 1$; likewise, if $c_2 \neq 0$ then $\tilde{c}_2 \neq 0$ as well. Scaling (24) by $|\tilde{c}_2|$ and (25) by $|\tilde{c}_1|$ we have

$$|\tilde{c}_2|\|A\mathbf{a}_1 - B\mathbf{b}_1\|_2 = \|c_1\tilde{c}_2A\mathbf{e}_i - \tilde{c}_1\tilde{c}_2B\mathbf{e}_{\pi^*}\|_2 \leq \Delta|c_1||\tilde{c}_2| \tag{26}$$

and

$$|\tilde{c}_1|\|A\mathbf{a}_2 - B\mathbf{b}_2\|_2 = \|c_2\tilde{c}_1A\mathbf{e}_j - \tilde{c}_1\tilde{c}_2B\mathbf{e}_{\pi^*}\|_2 \leq \Delta|\tilde{c}_1||c_2|. \tag{27}$$

Summing (26) and (27) and applying the triangle inequality, we get

$$\begin{aligned}
\Delta(|c_1||\tilde{c_2}| + |\tilde{c_1}||c_2|) &\geq \|c_1\tilde{c_2}Ae_i - c_2\tilde{c}_1Ae_j\|_2 \\
&\geq \alpha\|c_1\tilde{c_2}e_i - c_2\tilde{c}_1e_j\|_2 \\
&\geq \frac{\alpha}{\sqrt{2}}(|c_1||\tilde{c_2}| + |c_2||\tilde{c}_1|),
\end{aligned}$$

where we have also applied the $(2,\alpha)$-lower-RIP of A and the fact that $\|x\|_1 \leq \sqrt{p}\|x\|_2$ for all $x \in \mathbb{R}^p$ to reach a contradiction with our initial assumption that $\Delta < \frac{\alpha}{\sqrt{2}}$. Hence, π is injective and the matrix $P \in \mathbb{R}^{m\times m}$ whose i-th column is $e_{\pi(i)}$ for all $i \in \{1,\ldots,m\}$ is a permutation matrix. For any set of $a_i = c_ie_i \neq 0$, letting $D \in \mathbb{R}^{m\times m}$ be the (invertible) diagonal matrix with corresponding nonzero elements $\frac{\tilde{c}_1}{c_1},\ldots,\frac{\tilde{c}_m}{c_m}$, we have that $\mathbf{b}_i = \tilde{c}_i\mathbf{e}_{\pi(i)} = PD(c_i\mathbf{e}_i) = PD\mathbf{a}_i$ for all $i \in \{1,\ldots,m\}$. Furthermore, (23) implies that $\|(A - BPD)\mathbf{e}_i\| \leq \Delta$ for all $i \in \{1,\ldots,m\}$. ■

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

$$\Theta(\text{Span}\{M_{S_1}\}, \text{Span}\{M_{S_2}\}) < \alpha \implies S_1 = S_2. \tag{28}$$

Proof of Lemma 3: If $\ell = m$ then, trivially, $S_1 = S_2$. 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 then every $\ell+1$ columns of M are linearly independent and $\dim(M_{S_1}) = \dim(M_{S_2})$. Hence, by (2) we have

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

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

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

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

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

Hence, $\Theta(\text{Span}\{M_{S_1}\}, \text{Span}\{M_{S_2}\}) \geq \alpha$, 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

$$d(x, V\cap W) \leq \Delta\left(\frac{\cos\theta_F + \sqrt{2-\cos^2\theta_F}}{1-\cos^2\theta_F}\right) \tag{29}$$

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

Proof of Lemma 4: It can be shown [ref?] that for a given subspace $U \subseteq \mathbb{R}^m$, the projection operator $\Pi_U : \mathbb{R}^m \to U$ is the unique operator for which $d(x,U) = \|x - \Pi_U x\|$ for all $x \in \mathbb{R}^m$. Hence, it suffices to show that $\|x - \Pi_{V\cap W}x\|$ is bounded from above by the RHS of (29). Since $\Pi_{V\cap W}x \in W$ for all $x \in \mathbb{R}^m$, we have by Pythagoras' theorem that

$$\|x - \Pi_{V\cap W}x\|^2 = \|x - \Pi_W x\|^2 + \|\Pi_W x - \Pi_{V\cap W}x\|^2. \tag{30}$$

The first term on the RHS of (30) is $d(x,W)$. Applying the triangle inequality to the second term, we have

$$\|\Pi_W x - \Pi_{V\cap W}x\| \leq \|\Pi_W x - \Pi_W\Pi_V x\| + \|\Pi_W\Pi_V x - \Pi_{V\cap W}x\|. \tag{31}$$

The first term on the RHS of (31) 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)"]:

$$\|(\Pi_W\Pi_V)x - \Pi_{V\cap W}x\| \leq \cos\theta_F\|x\| \quad \text{for all} \quad x \in \mathbb{R}^m. \tag{32}$$

First, note that

$$\begin{aligned} \|(\Pi_W\Pi_V)(x - \Pi_{V\cap W}x) - \Pi_{V\cap W}(x - \Pi_{V\cap W}x)\| &= \|\Pi_W\Pi_V x - \Pi_W\Pi_V\Pi_{V\cap W}x - \Pi_{V\cap W}x + \Pi_{V\cap W}^2 x\| \\ &= \|(\Pi_W\Pi_V)x - \Pi_{V\cap W}x\|, \end{aligned} \tag{33}$$

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 (32) and (33) that

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

It follows from this, (30), (31) and the assumption $d(x,V) \leq d(x,W) \leq \Delta$ that

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

which can be rearranged into the following quadratic inequality in $\rho := \|x - \Pi_{V\cap W}x\|$:

$$\left(1 - cos^2\theta_F\right)\rho^2 - 2\Delta\cos\theta_F\rho - 2\Delta^2 \leq 0 \tag{34}$$

The zeros of the LHS are

$$\begin{aligned} \rho_\pm &= \frac{2\Delta\cos\theta_F \pm \sqrt{4\Delta^2\cos^2\theta_F - 4\left(1-\cos^2\theta_F\right)\left(-2\Delta^2\right)}}{2\left(1-\cos^2\theta_F\right)} \\ &= \Delta\left(\frac{\cos\theta_F \pm \sqrt{2-\cos^2\theta_F}}{1-\cos^2\theta_F}\right), \end{aligned}$$

of which, for all $\theta_F \in [0, \frac{\pi}{2}]$, only ρ_+ is positive. Hence (34) implies that

$$0 \leq \rho \leq \Delta\left(\frac{\cos\theta_F + \sqrt{2-\cos^2\theta_F}}{1-\cos^2\theta_F}\right). \quad \blacksquare$$

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

$$\mathrm{Span}\{M_{S\cap S'}\} = \mathrm{Span}\{M_S\} \cap \mathrm{Span}\{M_{S'}\} \tag{35}$$

Lemma 6: Let V, W be subspaces of $\mathbb{R}^m$ and suppose that for all $v \in V$ we have $d(v, W) < \|v\|_2$. Then $\dim(V) \le \dim(W)$.

Proof of Lemma 6: Since linear subspaces of $\mathbb{R}^m$ are closed we can assume there exists some $w \in W$ such that

$$\|v - w\|_2 < \|v\|_2. \tag{36}$$

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 (36). ■.

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)).**

IV. Appendix

Lemma 7: Let $\gamma_1 < \ldots < \gamma_N$ be an arithmetic sequence with common difference δ. Then for all $S \in \binom{[N]}{k}$ the $k \times N$ Vandermonde matrix $V = (\gamma_j^i)_{i,j=1}^{k,N}$ satisfies

$$\|V_S x\|_2 > \rho \|x\|_1 \quad \text{where} \quad \rho = \left(\frac{k-1}{k}\right)^{\frac{k-1}{2}} \delta \prod_{1\le j\le k} \gamma_j \prod_{1\le i<j\le k} (j-i).. \tag{37}$$

Proof of Lemma 7: The determinant of the Vandermonde matrix is

$$\det(V) = \prod_{1\le j\le k} \gamma_j \prod_{1\le i<j\le k} (\gamma_j - \gamma_i). \tag{38}$$

Since the γ_i are distinct, the determinant of any $k\times k$ submatrix of V is nonzero; hence, given $S \in \binom{[N]}{k}$, V_S is nonsingular. Its determinant is

$$\det(V_S) = \prod_{j\in S} \gamma_j \prod_{\substack{i\in S\\ i\le j}} (\gamma_j - \gamma_i) \ge \delta \prod_{1\le j\le k} (\gamma_1 + (j-1)\delta) \prod_{1\le i<j\le k} (j-i). \tag{39}$$

Now suppose $x \in \mathbb{R}^k$. Then $\|x\|_2 = \|V_S^{-1} V_S x\|_2 \le \|V_S^{-1}\| \|V_S x\|_2$, implying $\|V_S x\|_2 \ge \|V_S^{-1}\|^{-1} \|x\|_2 \ge \frac{1}{\sqrt{k}} \|V_S\|_2^{-1} \|x\|_1$. For the Euclidean norm we have $\|V_S^{-1}\|_2 = \frac{1}{\sigma_{\min}(V_S)}$, where $\sigma_{\min}$ is the smallest singular value of V_S. A lower bound for the smallest singular value of a nonsingular matrix $M \in \mathbb{R}^{k\times k}$ is given in [Hong and Pan]:

$$\sigma_{\min}(M) > \left(\frac{k-1}{k}\right)^{\frac{k-1}{2}} |\det M| \tag{40}$$

and the result follows. ■

Lemma 8: Fix matrices $A, \tilde{A} \in \mathbb{R}^{n\times m}$ where $\tilde{A} = AE$ for some invertible diagonal matrix $E = \mathrm{diag}(\lambda_i) \in \mathbb{R}^{m\times m}$, $\lambda_i \in \mathbb{R}$ 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 8: For all $i \in [m]$, we have:

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