Chaz's Theorem: The Return of Hillar

Sufficient Conditions for Robust Dictionary Identification in Sparse Coding

Abstract

Extension of theorems in HS11 to noisy subsamples 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 set of 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 $W \cap (W \cap V)^\perp$:

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

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

Proof of Theorem 1: First, we produce a set of $N = mk\binom{m'}{k}$ vectors in $\mathbb{R}^k$ in general linear position (i.e. any set of k of them are linearly independent). Specifically, let $\gamma_1, ..., \gamma_N$ 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 $\mathbf{a}_1, \dots, \mathbf{a}_N \in \mathbb{R}^m$ with supports in the set $\mathcal{S} = \{\{i, \dots, (i + k - 1) \mod m\} : i \in [m]\} \subseteq \binom{[m]}{k}$ (partitioning the a_i evenly among these supports, i.e. $k\binom{[m']}{k}$ each) by setting the nonzero values of vector $\mathbf{a}_i$ to be those contained in the ith column of V. By this construction, every k vectors a_i are linearly independent.

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:

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

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 $\phi(c) \geq \delta$ 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 $\mathbf{a}_i$ (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 δ satisfying (4) for all of them. (We refer the reader to the Appendix for a lower bound on δ given as a function of k and the sequence $\gamma_1, \ldots, \gamma_N$ used to generate the a_i.)

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 $B \in \mathbb{R}^{n \times m'}$ with unit norm columns satisfying the $(2k, \alpha)$-lower-RIP 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 $S \in \mathcal{S}$, 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 sharing some support $S' \in \binom{[m']}{k}$. Let $\mathcal{Y}$ be a set of k such vectors $\mathbf{y}_i$ which we will index by $\mathcal{I}$, i.e. $\mathcal{Y} = \{\mathbf{y}_i : i \in \mathcal{I}\}$.

Note that any matrix satisfying an (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 these 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 $\text{supp}(\mathbf{a}_i) = S$ for all $i \in \mathcal{I}$ and A satisfies the $(2k, \alpha)$-lower-RIP, we have

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

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 $\|\mathbf{z} - \mathbf{z}'\|_2 \leq \frac{2\varepsilon}{\alpha\delta}\|\mathbf{z}\|_2$. It follows that $d(\mathbf{z}, \text{Span}\{B_{S'}\}) \leq \frac{2\varepsilon}{\alpha\delta}$ for all unit vectors $\mathbf{z} \in \text{Span}\{A_S\}$. Hence,

$$\sup_{\substack{\mathbf{z} \in \text{Span}\{A_S\} \\ \|\mathbf{z}\| = 1}} d(\mathbf{z}, \text{Span}\{B_{S'}\}) \leq \frac{2\varepsilon}{\alpha\delta}. \tag{8}$$

If $\varepsilon < \frac{\alpha\delta}{2}$ then by Lemma 3 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 $\Theta(\text{Span}\{A_S\}, \text{Span}\{\mathcal{B}_{S'}\})$ $\frac{2\varepsilon}{\alpha\delta}$. More specifically, letting $\theta \in [0, \frac{\pi}{2}]$ be the least of all Friedrichs angles formed between pairs of susubspaces for which k columns of A form a basis or pairs of subspaces for which k columns of B form a basis, if

$$\varepsilon < \frac{\alpha^2\delta}{2\sqrt{2}} \left(\frac{1 - \cos^2\theta}{\cos\theta + \sqrt{2 - \cos^2\theta}} \right) \tag{9}$$

then we indeed have $\varepsilon < \frac{\alpha\delta}{2}$ and the association $S \mapsto S'$ defines a map $\pi : \mathcal{S} \to \binom{[m']}{k}$ satisfying

$$\Theta(\text{Span}\{A_S\}, \text{Span}\{\mathcal{B}_{\pi(S)}\}) \leq \frac{2\varepsilon}{\alpha\delta} < \frac{\alpha}{\sqrt{2}} \left(\frac{1 - \cos^2\theta}{\cos\theta + \sqrt{2 - \cos^2\theta}} \right) \quad \text{for all} \quad S \in \mathcal{S}. \tag{10}$$

It then follows by Lemma 1 that there exists a partial permutation matrix $P \in \mathbb{R}^{m' \times m'}$ and a diagonal matrix $D \in \mathbb{R}^{m' \times m}$ such that for all $i \in \{1, \ldots, m\}$, $\|(A - BPD)e_i\|_2 \leq C\varepsilon$, where

$$C = \frac{2}{\alpha\delta} \left(\frac{\cos\theta + \sqrt{2 - \cos^2\theta}}{1 - \cos^2\theta} \right). \quad \blacksquare \tag{11}$$

Lemma 1 (Main Lemma): Fix positive integers $n < m \le m'$ and k such that $2k-1 \le m$. Let $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{n \times m'}$ be matrices having unit norm columns satisfying the $(2k, \alpha)$-lower-RIP. Denote by $\theta \in [0, \frac{\pi}{2}]$ the least of all Friedrichs angles formed between pairs of subspaces for which k columns of A form a basis or pairs of subspaces for which k columns of B form a basis and let

$$\rho := \frac{1-\cos^2\theta}{\cos\theta + \sqrt{2-\cos^2\theta}}. \tag{12}$$

Let $\mathcal{S} := \{\{i, \ldots, (i+k-1) \mod m\} : i \in [m]\} \subseteq \binom{[m]}{k}$. If there exists a map $\pi : \mathcal{S} \to \binom{[m']}{k}$ and some $\Delta < \frac{\alpha}{\sqrt{2}}$ such that

$$\Theta(\text{Span}\{A_S\}, \text{Span}\{B_{\pi(S)}\}) \le \rho\Delta \quad \forall S \in \mathcal{S}, \tag{13}$$

then there exist a permutation matrix $P \in \mathbb{R}^{m'\times m'}$ and a diagonal matrix $D \in \mathbb{R}^{m' \times m}$ such that

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

Proof of Lemma 1: We assume $k \ge 2$, the case $k=1$ being contained in Lemma 2. Fix $i \in \{1, \ldots, m\}$. Since $m \ge 2k-1$, there exist $S_1 \neq S_2 \in \binom{[m]}{k}$ such that $S_1 \cap S_2 = \{i\}$ (specifically, $S_1 = \{i, \ldots, (i+k-1) \mod m\}$ and $S_2 = \{i-k+1, \ldots, i \mod m\}$). Condition (13) implies that for all unit vectors $\mathbf{z} \in \text{Span}\{B_{\pi(S_1)\cap\pi(S_2)}\} \subseteq \text{Span}\{B_{\pi(S_1)}\} \cap \text{Span}\{B_{\pi(S_2)}\}$ we have $d(\mathbf{z}, \text{Span}\{A_{S_1}\}) \le \rho\Delta$ and $d(\mathbf{z}, \text{Span}\{A_{S_2}\}) \le \rho\Delta$. It follows by Lemmas 4 and 5 that $d\left(\mathbf{z}, \text{Span}\{A_i\}\right) \le \rho\rho^{-1}\Delta = \Delta$. Similarly, we have from (13) and Lemmas 4 and 5 that $d\left(\mathbf{z}, \text{Span}\{B_{\pi(S_1 \cap \pi(S_2)}\}\right) \le \Delta$ for all unit vectors $\mathbf{z} \in \text{Span}\{A_i\}$. It follows by Lemma 3 that $\dim(\text{Span}\{B_{\pi(S_1)\cap\pi(S_2)}\}) = \dim(\text{Span}\{A_i\})$, hence $|\pi(S_1) \cap \pi(S_2)| = 1$. The association $i \mapsto \pi(S_1) \cap \pi(S_2)$ thus defines a map $\tau : [m] \to [m']$ such that for all unit vectors $\mathbf{z} \in \text{Span}\{A_i\}$ we have $d\left(\mathbf{z}, \text{Span}\{B_{\tau(i)}\}\right) \le \Delta$ and the result follows by Lemma 2

Lemma 2: Fix positive integers $n < m \le m'$. Let $A \in \mathbb{R}^{n\times m}$ and $B \in \mathbb{R}^{n\times m'}$ have unit norm columns and suppose that A satsifies the $(2, \alpha)$-lower-RIP. If there exists a map $\pi : \{1, \ldots, m\} \to \{1, \ldots, m'\}$ and some $\Delta < \frac{\alpha}{\sqrt{2}}$ such that

$$d\left(Ae_i, \text{Span}\{B\mathbf{e}_{\pi(i)}\}\right) \le \Delta \quad \text{for all} \quad i \in \{1, ..., m\} \tag{15}$$

then there exist a partial permutation matrix $P \in \mathbb{R}^{m'\times m'}$ and diagonal matrix $D \in \mathbb{R}^{m'\times m}$ such that $\|(A - BPD)\mathbf{e}_i\|_2 \le \Delta$ for all $i \in \{1, \ldots, m\}$.

Proof of Lemma 2: We will show that π is injective (and thus defines a permutation when its codomain is restricted to its image). Suppose that $\pi(i) = \pi(j) = \ell$ for some $i \neq j \in \{1, \ldots, m'\}$. By (15), there exists some $\mathbf{b}_i = c_i\mathbf{e}_{\pi(i)} \in \mathbb{R}^{m'}$ such that $\|A\mathbf{e}_i - c_iB\mathbf{e}_\ell\|_2 < \frac{\alpha}{\sqrt{2}}$. Similarly, by (15) there exists some $b_j = c_j\mathbf{e}_{\pi(j)} \in \mathbb{R}^{m'}$ such that $\|A\mathbf{e}_j - c_jB\mathbf{e}_\ell\|_2 < \frac{\alpha}{\sqrt{2}}$. (Note that $c_i \neq 0$ and $c_j \neq 0$ since A has unit norm columns and $\alpha < 1$.) Summing and scaling these two inequalities, we apply the triangle inequality and the $(2, \alpha)$-lower-RIP on A to yield

$$\alpha\|c_je_i + c_ie_j\|_2 \le \|c_jAe_i + c_iAe_j\|_2 < (|c_i| + |c_j|)\frac{\alpha}{\sqrt{2}}, \tag{16}$$

which is a contradiction due to the fact that $\|x\|_1 \le \sqrt{2}\|x\|_2$ for all 2-sparse $x \in \mathbb{R}^m$. Hence, π is injective and the matrix $P \in \mathbb{R}^{m'\times m'}$ whose i-th column is $e_{\pi(i)}$ for all $1 \le i \le m$ and $\mathbf{0}$ for all $m < i \le m'$ is a partial permutation matrix. Letting $D \in \mathbb{R}^{m'\times m}$ be the diagonal matrix with diagonal elements $c_1, ..., c_m$, we have that $\mathbf{b}_i = c_i\mathbf{e}_{\pi(i)} = PD\mathbf{e}_i$ for all $i \in \{1, \ldots, m\}$, or more generally, $\mathbf{b}_i = PD\mathbf{a}_i$ for all 1-sparse $\mathbf{a}_i$. Furthermore, (15) implies that $||(A - BPD)\mathbf{e}_i|| \le \Delta$ for all $i \in \{1, \ldots, m\}$. ■

Lemma 3: 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 3: Since linear subspaces of $\mathbb{R}^m$ are closed we can assume there exists some $w \in W$ such that

$$\|v - w\|_2 < \|v\|_2. \tag{17}$$

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 (17). ■.

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

$$\text{Span}\{M_{S\cap S'}\} = \text{Span}\{M_S\} \cap \text{Span}\{M_{S'}\} \tag{18}$$

Lemma 5: Let $x \in \mathbb{R}^m$ and suppose V, W are linear subspaces of $\mathbb{R}^m$. Suppose $d(x, V) \le d(x, W) \le \Delta$. Then

$$d(x, V \cap W) \le \Delta\left(\frac{\cos\theta_F + \sqrt{2-\cos^2\theta_F}}{1-\cos^2\theta_F}\right) \tag{19}$$

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

Proof of Lemma 5: 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 (19). Since $\Pi_{V\cap W} x \in W$ for all $x \in \mathbb{R}^m$, we have by Pythagoras' theorem that

$$\|x - \Pi_{V\cap W} x\|^2 = \|x - \Pi_W x\|^2 + \|\Pi_W x - \Pi_{V\cap W} x\|^2. \tag{20}$$

The first term on the RHS of (20) is $d(x, W)$. Applying the triangle inequality to the second term, we have

$$\|\Pi_W x - \Pi_{V\cap W} x\| \leq \|\Pi_W x - \Pi_W \Pi_V x\| + \|\Pi_W \Pi_V x - \Pi_{V\cap W} x\|. \tag{21}$$

The first term on the RHS of (21) 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)"]:

$$\|(\Pi_W \Pi_V) x - \Pi_{V\cap W} x\| \leq \cos\theta_F \|x\| \quad \text{for all} \quad x \in \mathbb{R}^m. \tag{22}$$

First, note that

$$\begin{aligned} \|(\Pi_W \Pi_V)(x - \Pi_{V\cap W} x) - \Pi_{V\cap W}(x - \Pi_{V\cap W} x)\| &= \|\Pi_W \Pi_V x - \Pi_W \Pi_V \Pi_{V\cap W} x - \Pi_{V\cap W} x + \Pi_{V\cap W}^2 x\| \\ &= \|(\Pi_W \Pi_V) x - \Pi_{V\cap W} x\|, \end{aligned} \tag{23}$$

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 (22) and (23) 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, (20), (21) 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 $\rho := \|x - \Pi_{V\cap W} x\|$:

$$\left(1 - cos^2\theta_F\right)\rho^2 - 2\Delta\cos\theta_F \rho - 2\Delta^2 \leq 0 \tag{24}$$

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 (24) 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 6: Let $\gamma_1 < ... < \gamma_N$ be any distinct numbers such that $\gamma_{i+1} = \gamma_i + \delta$ and form the $k \times N$ Vandermonde matrix $V = (\gamma_j^i)_{i,j=1}^{k,N}$. Then for all $S \in \binom{[N]}{k}$,

$$\|V_S x\|_2 > \rho\|x\|_1 \quad \text{where} \quad \rho = \frac{\delta^k}{\sqrt{k}}\left(\frac{k-1}{k}\right)^{\frac{k-1}{2}} \prod_{i=1}^{k}(\gamma_1 + (i-1)\delta) \tag{25}$$

Proof of Lemma 6: The determinant of the Vandermonde matrix is

$$\det(V) = \prod_{1\leq j\leq k} \gamma_j \prod_{1\leq i\leq j\leq k} (\gamma_j - \gamma_i) \geq \delta^k \prod_{i=1}^{k}(\gamma_1 + (i-1)\delta). \tag{26}$$

Since the γ_i are distinct, the determinant of any $k \times k$ submatrix of V is nonzero; hence V_S is nonsingular for all $S \in \binom{[N]}{k}$. Suppose $x \in \mathbb{R}^k$. Then $\|x\|_2 = \|A_S^{-1} A_S x\|_2 \leq \|A_S^{-1}\|\|A_S x\|_2$, implying $\|A_S x\|_2 \geq \|A_S^{-1}\|^{-1}\|x\|_2 \geq \frac{1}{\sqrt{k}}\|A_S\|_2^{-1}\|x\|_1$. For the Euclidean norm we have $\|A_S^{-1}\|_2 = \frac{1}{\sigma_{\min}(A_S)}$, where $\sigma_{\min}$ is the smallest singular value of A_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]:

$$\sigma_{\min}(M) > \left(\frac{k-1}{k}\right)^{\frac{k-1}{2}} |\det M| \tag{27}$$

and the result follows. ■