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 $d(v, W) := \inf\{\|v - w\|_2 : w \in W\} = \|v - \Pi_W v\|$ where Π_W is the projection operator onto subspace W. 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]

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

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

$$\dim(W) = \dim(V) \implies \sup_{\substack{v \in V \\ \|v\|=1}} d(v, W) = \sup_{\substack{w \in W \\ \|w\|=1}} d(w, V). \tag{2}$$

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* $\theta_F(V, W) \in [0, \frac{\pi}{2}]$ between subspaces V and W of $\mathbb{R}^m$ is the minimal angle formed between unit vectors in $V \cap (V \cap W)^\perp$ and $W \cap (W \cap V)^\perp$, that is

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

In our proof we make use of the following quantity defined for a sequence $V_1, \ldots, V_p$ of closed subspaces of $\mathbb{R}^m$:

$$c(V_1, \ldots, V_p) := 1 - \left[1 - \prod_{i=1}^{r-1} \left(1 - \cos^2\left[\theta_F(V_i, \cap_{j=i+1}^p V_j)\right]\right)\right]^{1/2} \tag{4}$$

Since we will be solely working with subspaces spanned by subsets of columns of a matrix $M \in \mathbb{R}^{n\times m}$, we make the following definition for notational convenience. Given a sequence of supports $S_1, \ldots, S_p \in \binom{[m]}{k}$, let

$$c_M(S_1, \ldots, S_p) := c\left(\text{Span}\{M_{S_1}\}, \ldots, \text{Span}\{M_{S_p}\}\right). \tag{5}$$

Theorem 1: Fix positive integers n and $k < m \leq m'$. Fix $\alpha \in (0, 1]$. There exist $N = mk\binom{m'}{k}$ 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 $\|\mathbf{y}_i - A\mathbf{a}_i\|_2 \leq \varepsilon$ for all $i \in \{1, \ldots, N\}$ 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 $i \in \{1, \ldots, N\}$ 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 $\mathbf{a}_1, \ldots, \mathbf{a}_N \in \mathbb{R}^m$ with supports $S_j = \{j, \ldots, (j + k - 1) \mod m\}$ for $j \in [m]$ (partitioning the a_i evenly among these supports, i.e. for each support S_j there are $k\binom{[m']}{k}$ vectors a_i with that support) by setting the nonzero values of vector $\mathbf{a}_i$ to be those contained in the ith column of V. Note that by 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:

$$\| \sum_{j=1}^{k} c_j \mathbf{a}_{i_j} \|_2 \geq \delta \|c\|_1 \quad \forall c = (c_1, ..., c_k) \in \mathbb{R}^m. \tag{6}$$

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 (6) 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 (6) 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, 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

$$\|\mathbf{z} - \mathbf{y}\|_2 = \| \sum_{i \in \mathcal{I}} c_i A\mathbf{a}_i - \sum_{i \in \mathcal{I}} c_i \mathbf{y}_i \|_2 \leq \sum_{i \in \mathcal{I}} \|c_i (A\mathbf{a}_i - \mathbf{y}_i)\|_2 = \sum_{i \in \mathcal{I}} |c_i| \|A\mathbf{a}_i - \mathbf{y}_i\|_2 \leq \varepsilon \sum_{i \in \mathcal{I}} |c_i|. \tag{7}$$

The alternate factorization for the $\mathbf{y}_i$ implies (by a manipulation identical to that of (7)) 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

$$\|\mathbf{z} - \mathbf{z}'\|_2 \leq \|\mathbf{z} - \mathbf{y}\|_2 + \|\mathbf{y} - \mathbf{z}'\|_2 = 2\varepsilon \sum_{i \in \mathcal{I}} |c_i|. \tag{8}$$

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

$$\|\mathbf{z}\|_2 = \| \sum_{i \in \mathcal{I}}^{k} c_i A\mathbf{a}_i \|_2 = \|A(\sum_{i \in \mathcal{I}} c_i \mathbf{a}_i)\|_2 \geq \alpha \| \sum_{i \in \mathcal{I}}^{k} c_i \mathbf{a}_i \|_2 \geq \alpha\delta \sum_{i \in \mathcal{I}}^{k} |c_i|, \tag{9}$$

where for the last inequality we have applied the property (6). Combining (8) and (9), 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,

$$\sup_{\substack{\mathbf{z} \in \text{Span}\{A_S\} \\ \|\mathbf{z}\|=1}} d(\mathbf{z}, \text{Span}\{B_{S'}\}) \leq \frac{2\varepsilon}{\alpha\delta}. \tag{10}$$

Suppose $\varepsilon < \frac{\alpha\delta}{2}$. Then $\dim(\text{Span}\{B_{S'}\}) \geq \dim(\text{Span}\{A_S\}) = k$ by Lemma 3 and the fact that every k columns of A are linearly independent . In fact, since $|S'| = k$, we have $\dim(\text{Span}\{B_{S'}\}) = \dim(\text{Span}\{A_S\})$. Recalling (2), we see the association $S \mapsto S'$ thus defines a map $\pi : \{S_1, \ldots, S_m\} \to \binom{[m']}{k}$ satisfying $\Theta(\text{Span}\{A_S\}, \text{Span}\{\mathcal{B}_{\pi(\mathcal{S})}\}) \leq \frac{2\varepsilon}{\alpha\delta}$, Suppose further that $\varepsilon < \frac{\alpha^2\delta\rho}{2k\sqrt{2}}$, where ρ is defined as in equation (13) of Lemma 1. Since $\alpha < 1$ and $\rho < 1$, we then indeed have $\varepsilon < \frac{\alpha\delta}{2}$ so that

$$\Theta(\text{Span}\{A_{S_i}\}, \text{Span}\{\mathcal{B}_{\pi(S_i)}\}) \leq \frac{\rho}{k}\Delta \quad \text{for all} \quad i \in [m], \tag{11}$$

where $\Delta = \frac{2k\varepsilon}{\alpha\delta\rho} < \frac{\alpha}{\sqrt{2}}$. Moreover, it 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 $i \in \{1,\ldots,m\}$, $\|(A - BPD)e_i\|_2 \leq C\varepsilon$ for $C = \frac{2k}{\alpha\delta\rho}$. ■

Lemma 1 (Main Lemma): Fix positive integers n and $k < m \leq m'$. 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. Let $S_i := \{i,\ldots,(i+k-1) \mod m\}$ for $i = 1,\ldots,m$. If there exists a map $\pi : \{S_1,\ldots,S_m\} \to \binom{[m']}{k}$ and some $\Delta < \frac{\alpha}{\sqrt{2}}$ such that

$$\Theta(\text{Span}\{A_{S_i}\}, \text{Span}\{B_{\pi(S_i)}\}) \leq \frac{\rho}{k}\Delta \quad \text{for all} \quad i \in [m] \tag{12}$$

where, for $h(i) = [i-(k-1)] \mod m$,

$$\rho = \min_{i\in[m]} \min\left(c_A(S_{h(i)},\ldots,S_i), c_B(\pi(S_{h(i)}),\ldots,\pi(S_i))\right) \tag{13}$$

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

$$\|(A - BPD)e_i\|_2 \leq \Delta \quad \text{for all} \quad i \in [m]. \tag{14}$$

Proof of Lemma 1: We assume $k \geq 2$, the case $k = 1$ being contained in Lemma 2. Fix $i \in [m]$. Then $\cap_{j=h(i)}^{i} S_j = \{i\}$. (In general, there are combinations of fewer supports S_j with intersection $\{i\}$, e.g. if $m \geq 2k-1$ then $S_{h(i)} \cap S_i = \{i\}$. For brevity, we consider a construction that is valid for any $m > k$.) Condition (12) implies that for all unit vectors $\mathbf{z} \in \text{Span}\{B_{\cap_{j=h(i)}^{i}\pi(S_j)}\} \subseteq \cap_{j=h(i)}^{i}\text{Span}\{B_{\pi(S_j)}\}$ we have $d(\mathbf{z}, \text{Span}\{A_{S_j}\}) \leq \frac{\rho}{k}\Delta$ for all $j = h(i),\ldots,i$. It follows by Lemmas 4 and 5 that

$$d\left(\mathbf{z}, \text{Span}\{A_i\}\right) \leq \frac{\rho}{k}\Delta\left(\frac{k}{c_A(S_{h(i)},\ldots,S_i)}\right) \leq \Delta \tag{15}$$

and by Lemma 3 we therefore have $\dim(\text{Span}\{B_{\cap_{j=h(i)}^{i}\pi(S_j)}\}) \leq 1$. Similarly, we have from (12) and Lemmas 4 and 5 that

$$d\left(\mathbf{z}, \text{Span}\{B_{\cap_{j=h(i)}^{i}\pi(S_j)}\}\right) \leq \frac{\rho}{k}\Delta\left(\frac{k}{c_B(\pi(S_{h(i)}),\ldots,\pi(S_i))}\right) \leq \Delta. \tag{16}$$

for all unit vectors $\mathbf{z} \in \text{Span}\{A_i\}$. It follows by Lemma 3 that $\dim(\text{Span}\{B_{\cap_{j=h(i)}^{i}\pi(S_j)}) = \dim(\text{Span}\{A_i\})$; hence $|\cap_{j=h(i)}^{i}\pi(S_j)| = 1$. The association $i \mapsto \cap_{j=h(i)}^{i}\pi(S_j)$ thus defines a map $\tau : [m] \to [m']$ such that for all unit vectors $\mathbf{z} \in \text{Span}\{A_i\}$ we have $d\left(\mathbf{z}, \text{Span}\{B_{\tau(i)}\}\right) \leq \Delta$ and, since $\Delta < \frac{\alpha}{\sqrt{2}}$, the result follows by Lemma 2

Lemma 2: Fix positive integers n and $m \leq m'$. 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

$$d\left(Ae_i, \text{Span}\{B\mathbf{e}_{\pi(i)}\}\right) \leq \Delta \quad \text{for all} \quad i \in [m] \tag{17}$$

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 $\|(A - BPD)\mathbf{e}_i\|_2 \leq \Delta$ 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. First, note that (17) implies that there exist $c_1,\ldots,c_m \in \mathbb{R}$ such that

$$\|Ae_i - c_iBe_{\pi(i)}\| \leq \Delta \quad \text{for all} \quad i \in [m]. \tag{18}$$

Now suppose that $\pi(i) = \pi(j) = \ell$ for some $i \neq j$ and $\ell \in [m]$. Then $\|A\mathbf{e}_i - c_iB\mathbf{e}_\ell\|_2 \leq \Delta$ and $\|A\mathbf{e}_j - c_jB\mathbf{e}_\ell\|_2 \leq \Delta$. (Note that $c_i \neq 0$ and $c_j \neq 0$ since A has unit norm columns and $\alpha \leq 1$.) Summing and scaling these two inequalities by c_j and c_i, respectively, we apply the triangle inequality and the $(2,\alpha)$-lower-RIP on A to yield

$$\alpha\|c_je_i + c_ie_j\|_2 \leq \|c_jAe_i + c_iAe_j\|_2 \tag{19}$$

$$\leq |c_j|\|A\mathbf{e}_i - c_iB\mathbf{e}_\ell\|_2 + |c_i|\|c_jB\mathbf{e}_\ell - A\mathbf{e}_j\|_2 \tag{20}$$

$$\leq (|c_i| + |c_j|)\Delta \tag{21}$$

which is in contradiction with the fact that $\|x\|_1 \leq \sqrt{2}\|x\|_2$ for all $x \in \mathbb{R}^2$ and $\Delta < \frac{\alpha}{\sqrt{2}}$. Hence, π is injective and the matrix $P \in \mathbb{R}^{m'\times m'}$ whose i-th column is $e_{\pi(i)}$ for all $1 \leq i \leq m$ and $\mathbf{0}$ for all $m < i \leq m'$ is a partial permutation matrix. Letting $D \in \mathbb{R}^{m'\times m}$ be the diagonal matrix with diagonal elements $c_1,\ldots,c_m$, (18) becomes $||(A - BPD)\mathbf{e}_i|| \leq \Delta$ for all $i \in [m]$. ■

Lemma 3: Let V, W be closed subspaces of $\mathbb{R}^m$. If $d(v,W) < \|v\|_2$ for all $v \in V$ then $\dim(V) \leq \dim(W)$.

Proof of Lemma 3: The condition implies that for all $v \in V$ there exists some $w \in W$ such that

$$\|v - w\|_2 < \|v\|_2. \tag{22}$$

If $\dim(V) > \dim(W)$ then there exists some $v' \in V \cap W^\perp$. By Pythagoras' Theorem, $\|v' - w\|_2^2 = \|v'\|_2^2 + \|w\|_2^2 \geq \|v\|_2^2$ for all $w \in W$, which is in contradiction with (22). ■.

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

$$y \in \text{Span}\{M_{\cap \mathcal{S}}\} \iff y \in \bigcap_{S \in \mathcal{S}} \text{Span}\{M_S\}. \tag{23}$$

Proof of Lemma 4: The forward direction is trivial; we prove the reverse direction by induction. Enumerate $\mathcal{S} = (S_1, \ldots, S_{|\mathcal{S}|})$ and let $y \in \bigcap_i \text{Span}\{M_{S_i}\}$. Then for all $S_i \in \mathcal{S}$ there exists some x_i with support contained in S_i such that $y = Mx_i$. Suppose there exists some x with support contained in $\cap_{i=1}^{k-1} S_i$, $k \leq |\mathcal{S}|$ such that $y = Mx$. Then $y = Mx = Mx_k$, implying $x = x_k$ (since every $2k$ columns of M are linearly independent). Hence the support of x is also contained in S_k, i.e. $\text{supp}(x) \subseteq \cap_{i=1}^k S_i$. ■

Lemma 5: For $p \geq 2$ let $V_1, \ldots, V_p$ be closed linear subspaces of $\mathbb{R}^m$, let $V = \cap_{i=1}^p V_i$. Then

$$\|x - \Pi_V x\| \leq \frac{1}{c(V_1, \ldots, V_p)} \sum_{i=1}^p \|x - \Pi_{V_i} x\| \quad \text{for all} \quad x \in \mathbb{R}^m. \tag{24}$$

Proof of Lemma 5: We will prove by induction on p that

$$\|x - \Pi_V x\| = \sum_{i=1}^p \|x - \Pi_{V_i} x\| + \|\Pi_{V_1}\Pi_{V_2}\cdots\Pi_{V_p} x - \Pi_V x\| \tag{25}$$

and show that (24) follows from this. Fix $x \in \mathbb{R}^m$ and suppose $p = 2$. Then since $\Pi_V x \in V_1$ for all $x \in \mathbb{R}^m$, we have by the triangle inequality that

$$\begin{aligned}
\|x - \Pi_V x\| &= \|x - \Pi_{V_1} x\| + \|\Pi_{V_1} x - \Pi_{V_1}\Pi_{V_2} x\| + \|\Pi_{V_1}\Pi_{V_2} x - \Pi_V x\| & (26)\\
&\leq \|x - \Pi_{V_1} x\| + \|x - \Pi_{V_2} x\| + \|\Pi_{V_1}\Pi_{V_2} x - \Pi_V x\|, & (27)
\end{aligned}$$

where we have used the fact that $\|\Pi\| \leq 1$ for any projection operator Π. Suppose now that

$$\|x - \Pi_V x\| = \sum_{i=1}^{p-1} \|x - \Pi_{V_i} x\| + \|\Pi_{V_1}\Pi_{V_2}\cdots\Pi_{V_{p-1}} x - \Pi_V x\| \tag{28}$$

Applying the triangle inequality, we have

$$\begin{aligned}
\|x - \Pi_V x\| &= \sum_{i=1}^{p-1} \|x - \Pi_{V_i} x\| + \|\Pi_{V_1}\Pi_{V_2}\cdots\Pi_{V_{p-1}} x - \Pi_{V_1}\Pi_{V_2}\cdots\Pi_{V_p} x\| + \|\Pi_{V_1}\Pi_{V_2}\cdots\Pi_{V_p} x - \Pi_V x\| & (29)\\
&\leq \sum_{i=1}^{p} \|x - \Pi_{V_i} x\| + \|\Pi_{V_1}\Pi_{V_2}\cdots\Pi_{V_p} x - \Pi_V x\| & (30)
\end{aligned}$$

which proves (25). We now make use of the following result by [Deutsch, "Best Approximation in Inner Product Spaces, Theorem 9.33 "]:

$$\|\left(\Pi_{V_1}\Pi_{V_2}\cdots\Pi_{V_\ell}\right) x - \Pi_V x\| \leq (1 - c(V_1, \ldots, V_p))\, \|x\| \quad \text{for all} \quad x \in \mathbb{R}^m \tag{31}$$

$$\begin{aligned}
\|(\Pi_{V_1}\Pi_{V_2}\cdots\Pi_{V_k})(x - \Pi_V x) - \Pi_V(x - \Pi_V x)\| &= \|(\Pi_{V_1}\Pi_{V_2}\cdots\Pi_{V_k})x - (\Pi_{V_1}\Pi_{V_2}\cdots\Pi_{V_k})\Pi_V x - \Pi_V x + \Pi_V^2 x\| \\
&= \|(\Pi_{V_1}\Pi_{V_2}\cdots\Pi_{V_\ell})x - \Pi_V x\|, & (32)
\end{aligned}$$

We then have by (31) and (32) that

$$\begin{aligned}
\|(\Pi_{V_1}\Pi_{V_2}\cdots\Pi_{V_\ell})x - \Pi_V x\| &= \|(\Pi_{V_1}\Pi_{V_2}\cdots\Pi_{V_\ell})(x - \Pi_V x) - \Pi_V(x - \Pi_V x)\| \\
&\leq (1 - c(V_1, \ldots, V_p))\|x - \Pi_V x\|
\end{aligned}$$

It follows from this and (25) that

$$\|x - \Pi_V x\| \leq \sum_{i=1}^p \|x - \Pi_{V_i} x\| + (1 - c(V_1, \ldots, V_p))\|x - \Pi_V x\|$$

from which the result follows by solving for $\|x - \Pi_V x\|$. ■

III. APPENDIX

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 $S \in \binom{[N]}{k}$,

$$\|V_S x\|_2 > \rho \|x\|_1 \quad \text{where} \quad \rho = \frac{\delta^k}{\sqrt{k}} \left(\frac{k-1}{k}\right)^{\frac{k-1}{2}} \prod_{i=1}^{k} (\gamma_1 + (i-1)\delta) \quad \text{for all} \quad x \in \mathbb{R}^k \tag{33}$$

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

$$\det(V) = \prod_{1 \le j \le k} \gamma_j \prod_{1 \le i \le j \le k} (\gamma_j - \gamma_i) \ge \delta^k \prod_{i=1}^{k} (\gamma_1 + (i-1)\delta). \tag{34}$$

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 $\|x\|_2 = \|V_S^{-1} V_S x\|_2 \le \|V_S^{-1}\| \|V_S x\|_2$, 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 $\|V_S^{-1}\|_2^{-1} = \sigma_{\min}(V_S)$, 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]:

$$\sigma_{\min}(M) > \left(\frac{k-1}{k}\right)^{\frac{k-1}{2}} |\det M| \tag{35}$$

and the result follows. ■