Notes from Compressed Sensing and Contemporary Sampling

Cambridge Part III Mathematical Tripos 2012-2013

Lecturer: Anders Hansen

Vivak Patel

May 1, 2013

Contents

Part I
Compressed Sensing

1 Sparse Linear Algebra

1.1 Basic Definitions and Notation

1. $\Sigma_k = \{x \in \mathbb{R}^n : \|x\|_0 \leq k\}$ is the set of vectors in $\mathbb{R}^n$ with k or fewer non-zero components

2. k-term Approximation

   (a) Motivation: sparsity is key in determining x uniquely from an underdetermined system. However, x may not be sparse, but we can approximate it as being sparse:

   (b) The k-term approximation for a vector x is:

   $$\sigma_k(x)_p = \inf_{\tilde{x} \in \Sigma_k} \|x - \tilde{x}\|_p$$

3. Null Space of $A \in \mathbb{R}^{m \times n}$ is $\mathcal{N}(A) = \{x \in \mathbb{R}^n : Ax = 0\}$

1.2 Spark

1. Definitions and Properties of Spark

   (a) Definition

   Definition 1.1. *Let $A \in \mathbb{R}^{m \times n}$. The* $\operatorname{spark}(A) = \inf_{z \neq 0}\{\|z\|_0 : Az = 0\}$*. Equivalently, the* $\operatorname{spark}(A)$ *is the minimum number of linearly dependent columns of A.*

   (b) If we assume that $m < n$, which it is for an underdetermined system, $\operatorname{spark}(A) \in [2, m+1]$

2. Recovering at most a unique x given a signal y

   Theorem 1.1. *Let $y \in \mathbb{R}^m$. There exists at most one $x \in \Sigma_k$ such that $y = Ax$ if and only if* $\operatorname{spark}(A) > 2k$

   Proof. The proof relies on the following idea: Suppose $x, x' \in \Sigma_k$ and $Ax' = y = Ax \implies x - x' \in \mathcal{N}(A)$. Moreover, $\|x - x'\|_0 \leq 2k$. However, if we restrict $\operatorname{spark}(A) > 2k$ then no 2k-sparse vector (or more sparse) can be in the null space, hence $x' = x$.

   (a) Let $y \in \mathbb{R}^m$. Suppose $x \in \Sigma_k$ uniquely satisfies $y = Ax$. suppose $\operatorname{spark}(A) \leq 2k$. Then $\Sigma_{2k} \cap \mathcal{N}(A) \neq \emptyset$. Let $h \in \Sigma_{2k} \cap \mathcal{N}(A)$. Then, $\exists \tilde{x} \in \Sigma_k$ such that $h = x - \tilde{x}$ (this is like completing the basis). Then $Ah = 0 \implies Ax = A\tilde{x}$ a contradiction.

   (b) Suppose $\operatorname{spark}(A) > 2k$ Suppose $\exists x, x' \in \Sigma_k$ such that $y = Ax = Ax'$. Then, $x - x' \in \mathcal{N}(A) \cap \Sigma_{2k}$. However, $\forall h \in \mathcal{N}(A)$, $\|h\|_0 \geq \operatorname{spark}(A) > 2k \geq \|x - x'\|_0$. So this is a contradiction.

   □

1.3 Null Space Property

1. Definition of NSP

   Definition 1.2. *A matrix A satisfies the NSP of order k if $\exists c \in \mathbb{R}$ such that on some active set Λ with $|\Lambda| \leq k$, $\forall h \in \mathcal{N}(A)$,*

$$\|h_\Lambda\|_2 \leq \frac{c\,\|h_{\Lambda^c}\|_1}{\sqrt{k}}$$

2. Motivation: when a signal is exactly sparse, the previous theorem is sufficient, however, when it is approximatley sparse, we may not have unique solutions. The null space may contain vectors that are very sparse or very compressible (well approximated by sparse vectors). Therefore, to make sure that $\mathcal{N}(A)$ is well behaved, the null space property guarantees that: $h \in \mathcal{N}(A)$, if $\|h_{\Lambda^c}\|_1 = 0$ then $\|h_\Lambda\|_2 = 0$ so that $h = 0$.

3. Instance-Optimality: Given a sensing matrix A and recovery algorithm Δ, the following guarantees the optimal performance of Δ based on $\sigma_k(x)_1$:

$$\|\Delta(Ax) - x\|_2 \leq \frac{c}{\sqrt{k}}\sigma_k(x)_1$$

4. Instance-Optimality implies NSP

   Theorem 1.2. *Let $A : \mathbb{R}^n \to \mathbb{R}^m$ be a sensing matrix and $\Delta : \mathbb{R}^m \to \mathbb{R}^n$ be a recovery algorithm. If (A, Δ) is instance-optimal, then A satisfies the NSP of order $2k$*

   Proof. Let $h \in \mathcal{N}(A)$ and decompose it into $x - x'$ with $x' \in \Sigma_k$. The result follows:

   (a) Let $h \in \mathcal{N}(A)$ and let Λ be the index set of h's $2k$ largest components. Let $\Lambda = \Lambda_0 \cup \Lambda_1$ where $|\Lambda_0| = |\Lambda_1| = k$. Set $x = h_{\Lambda_1} + h_{\Lambda^c}$ and $x' = -h_{\Lambda_0}$. Then

$$x - x' = h_{\Lambda_1} + h_{\Lambda^c} + h_{\Lambda_0} = h$$

   (b) Since $\|x'\|_0 \leq k$, then $\sigma_k(x')_1 = 0$. By Instance-optimality,

$$\|\Delta(Ax') - x'\|_2 \leq 0 \implies \Delta(Ax') = x'$$

   Moreover,

$$0 = Ah = A(x - x') \implies Ax = Ax'$$

   (c) Therefore:

$$\begin{aligned}\|h_\Lambda\|_2 \leq \|h\|_2 &= \|x - x'\|_2 \\ &= \|x - \Delta(Ax')\|_2 \\ &= \|x - \Delta(Ax)\| \\ &\leq \frac{c}{\sqrt{k}}\sigma_k(x)_1\end{aligned}$$

(d) Noting that $\sigma_k(x)_1 = \inf_{\tilde{x}\in\Sigma_k} \|x - \tilde{x}\|$. By the definition of x, $\tilde{x} = h_{\Lambda_1}$, implying that $\sigma_k(x)_1 = \|h_{\Lambda^c}\|_1$. Therefore:

$$\|h_\Lambda\|_2 \leq \frac{c}{\sqrt{k}}\sigma_k(x)_1 \leq \frac{c'}{\sqrt{2k}} \|h_{\Lambda^c}\|_1$$

□

5. Note that if an algorithm satisfies the optimality condition, it satisfies the NSP. So for an approximately sparse signal (i.e. a compressed signal), we have a unique x given by $Ax = y$

1.4 Restricted Isometry Property

1. Motivation: when a signal has noise or is corrupted, NSP is not strong enough of a condition to guarantee a unique x. The Restrict Isometry Property, on the other hand, is.

2. Definition of RIP

Definition 1.3. *A matrix A satisfies the RIP of order k if $\exists \delta_k \in (0,1)$ such that $\forall x \in \Sigma_k$:*

$$(1-\delta_k)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_k)\|x\|_2^2$$

1.4.1 RIP and Stability

1. Definition of Stability

Definition 1.4. *Let $A \in \mathbb{R}^{m\times n}$ and $\Delta : \mathbb{R}^m \to \mathbb{R}^n$. (A, Δ) is c-stable if for any $x \in \Sigma_k$ and error $e \in \mathbb{R}^m$,*

$$\|\Delta(Ax + e) - x\|_2 \leq c\|e\|_2$$

2. C-Stability implies part of the RIP

Theorem 1.3. *Suppose (A, Δ) is c-stable. Then $\frac{1}{c}\|x\|_2 \leq \|Ax\|_2$ for all $x \in \Sigma_{2k}$*

Proof. We note that we are in Σ_{2k} which can be produced by any two vectors in Σ_k. Let $x, z \in \Sigma_k$. We use these to cleverly define errors:

(a) Let $x, z \in \Sigma_k$. et $e_x = \frac{A(z-x)}{2}$ and $e_z = \frac{A(x-z)}{2}$. Then $Ax + e_x = Az + e_z = \frac{A(x+z)}{2}$. Let the recovered value be $\Delta(Ax + e_x) = \Delta(Az + e_z) = y \in \mathbb{R}^n$.

(b) Then

$$\begin{aligned}\|x - z\|_2 &\leq \|x - y\|_2^2 + \|y - z\|_2 \\ &\leq c\|e_x\|_2 + c\|e_z\|_2 \\ &= c\|A(x - z)\|_2\end{aligned}$$

□

1.4.2 RIP and Measurement Bounds

1. Motivation: We can recover exactly k sparse vectors if we have that $\operatorname{spark}(A) \geq 2k$, where $m+1 \geq \operatorname{spark}(A)$. Given noise in the signal, we will typically need more measurements.

2. RIP implies number of measurements m needed:

   Theorem 1.4. *Let $A \in \mathbb{R}^{m \times n}$ satisfy the RIP of order $2k$ with $\delta_{2k} \in (0, 1/2]$. Then $m \geq c \log(n/k)$ and $c = \frac{1}{2}\left[\log(\sqrt{24}+1)\right] \approx 0.28$*

3. Important Lemma used to prove Theorem 1.4

   Lemma 1.1. *Suppose k and n satisfy $k < n/2$. Then $\exists X \subset \Sigma_k$ such that $\forall x, z \in X$ with $x \neq z$*

$$
\begin{aligned}
\|x\|_2 &\leq \sqrt{k} \\
\|x - z\|_2 &\geq \sqrt{\frac{k}{2}} \\
\log|X| &\geq \frac{k}{2}\log\left(\frac{n}{k}\right)
\end{aligned}
$$

   Proof. We first create a set satisfying the first requirement. Then we pick a subset satisfying the second requirement. However, this limits the size of the subset, resulting in the third requirement.

   (a) Let $U = \{x \in \{0, -1, 1\}^n \subset \mathbb{R}^n : \|x\|_0 = k\}$. Then $|U| = \binom{n}{k}2^k$ and $\|x\|_2 = \sqrt{k}$. This satisfies the first requirement.

   (b) From this set, fix $x \in U$. Since $\|x-z\|_0 \leq \|x-z\|_2^2$, if we bound $\|x-z\|_2^2 \leq k/2$ then:

$$
\begin{aligned}
|\{x \in U : \|x-z\|_2^2 \leq k/2\}| &\leq |\{z \in U : \|x-z\|_0 \leq k/2\}| \\
&\leq \binom{n}{k/2}2^{k/2} \\
&\leq \binom{n}{k/2}3^{k/2}
\end{aligned}
$$

   (c) If we remove the above set from U we are left with only $\binom{n}{k}2^k - \binom{n}{k/2}2^{k/2}$ points to choose from. Suppose we choose j points for which $\|x_r - x_p\|_2^2 \geq k/2$ for $r \neq q$ and $1 \leq r, q \leq j$. Then we have at least $\binom{n}{k}2^k - j\binom{n}{k/2}3^{k/2}$ points to choose from. If we find the maximum value of j possible we will have that

$$
|X| \leq \frac{\binom{n}{k}}{\binom{n}{k/2}}\left(\frac{4}{3}\right)^{k/2}
$$

□

   Proof of Theorem 1.4. We want to establish a (loose) lower bound on m for sparse signals in Σ_{2k}. Again we can decompose this space as two vectors from Σ_k. We use the special subset $X \subset \Sigma_k$ of the lemma, which has nice properties to get the lower bound.

(a) Let $x, z \in X$. By RIP, assumptions on δ_{2k}, and Lemma 1.1:

$$\|A(x-z)\|_2 \geq \frac{1}{2}\|x-z\|_2 \geq \sqrt{\frac{k}{4}}$$

(b) Let $x \in X$. By the RIP and the lemma:

$$\|Ax\|_2 \leq \sqrt{\frac{3k}{2}}$$

(c) Thus, all Ax for $x \in X$ are a distance $\sqrt{\frac{k}{4}}$, and that all Ax are within a distance $\sqrt{\frac{3k}{2}}$ from the origin. Therefore, for any $z \in X$

$$\begin{aligned}
\text{vol}\left[B_O\left(\sqrt{3k/2}+\sqrt{k/16}\right)\right] &\geq |X| \,\text{vol}\left[B_{Az}\left(\sqrt{k/16}\right)\right] \\
\left(\sqrt{3k/2}+\sqrt{k/16}\right)^m &\geq |X|\left(\sqrt{k/16}\right)^m \\
\left(\sqrt{24}+1\right)^m &\geq |X| \\
m &\geq \log(|X|)\frac{1}{\log\left(\sqrt{24}+1\right)} \\
m &\geq \frac{1}{2\log\left(\sqrt{24}+1\right)} k \log\left(\frac{n}{k}\right)
\end{aligned}$$

□

1.4.3 RIP and NSP

1. Satisfying the RIP implies satisfying the NSP

   Theorem 1.5. *Suppose A satisfies the RIP of order $2k$ with $\delta_{2k} \leq \sqrt{2}-1$. Then A satisfies the NSP of order $2k$ with $c = \frac{2}{1-(1+\sqrt{2}\delta_{2k})}$*

   Proof. We use the Lemmas below to prove this result.

   (a) By RIP and Lemma 1.6. Let $h \in \mathcal{N}(A)$. Let Λ be the $2k$ largest components of h. Let Λ_0 be the k largest components of h and Λ_1 be the next k largest so that $\Lambda = \Lambda_0 \cup \Lambda_1$.

      i. Noting that $Ah = 0$ and since we satisfy the conditions of 1.6:

      $$\|h_\Lambda\|_2 \leq \alpha\frac{\left\|h_{\Lambda_0^c}\right\|_1}{\sqrt{k}} = \frac{\alpha}{\sqrt{k}}\left(\|h_{\Lambda^c}\|_1 + \|h_{\Lambda_1}\|_1\right)$$

      ii. By Lemma 1.2,

      $$\|h_\Lambda\|_2 \leq \frac{\alpha}{\sqrt{k}}\left(\|h_{\Lambda^c}\|_1 + \sqrt{k}\,\|h_{\Lambda_1}\|_2\right) \leq \frac{\alpha}{\sqrt{k}}\left(\|h_{\Lambda^c}\|_1 + \sqrt{k}\,\|h\|_2\right)$$

      iii. Rearranging and noting that $1-\alpha > 0$:

      $$\|h_\Lambda\|_2 \leq \frac{\sqrt{2}\alpha}{1-\alpha}\frac{\|h_{\Lambda^c}\|_1}{\sqrt{2k}}$$

□

2. Important Lemmas

   (a) Relation between essential norms

   Lemma 1.2. *Suppose* $u \in \Sigma_k$ *then* $\|u\|_1 \leq \sqrt{k}\,\|u\|_2 \leq k\,\|u\|_\infty$

   Proof. By Cauchy-Schwarz and noting that $\|u\|_1 = \langle u, \operatorname{sgn}(u)\rangle$ and $u_i^2 \leq \|u\|_\infty^2$:

$$\begin{aligned}\|u\|_1 &= \langle u, \operatorname{sgn}(u)\rangle \\ &\leq \|\operatorname{sgn}(u)\|_2\,\|u\|_2 \\ &\leq \sqrt{k}\,\|u\|_2 \\ \|u\|_2^2 &= \sum_{i=1}^{n} u_i^2 \\ &\leq \sum_{i=1}^{k} \|u\|_\infty^2 \\ &= k\,\|u\|_\infty^2\end{aligned}$$

□

   (b) Perpendicular vectors

   Lemma 1.3. *Suppose* $u \perp v$. *Then* $\|u\|_2 + \|v\|_2 \leq \sqrt{2}\,\|u+v\|_2$

   Proof. Using the inequality from 1.2 on the vector: $A = \begin{pmatrix} \|u\|_2 \\ \|v\|_2 \end{pmatrix}$. We have that $\|A\|_1 \leq \sqrt{2}\,\|A\|_2 = \sqrt{2}\,\|u+v\|_2$ □

   (c) RIP and bounding the Inner Product

   Lemma 1.4. *Suppose A satisfies the RIP of order* $2k$*, then for any pair of vectors* $u, v \in \Sigma_k$ *with disjoint support,* $|\langle Au, Av\rangle| \leq \delta_{2k}\,\|u\|_2\,\|v\|_2$

   Proof. Let $u', v' \in \Sigma_k$ with disjoint support. Let $u = u'/\,\|u'\|_2$ and $v = v'/\,\|v'\|_2$ so that $\|u\|_2 = \|v\|_2 = 1$. The result follows from the parallelogram law and the RIP:

   i. By the Parallelogram Law:

$$\begin{aligned}|\langle Au, Av\rangle| &= \frac{1}{4}\left[\|Au+Av\|_2^2 - \|Au-Av\|_2^2\right] \\ &\leq \frac{1}{4}\left[(1+\delta_{2k})\,\|u+v\|_2^2 - (1-\delta_{2k})\,\|u-v\|_2^2\right]\end{aligned}$$

   ii. Since $u \perp v$, $\|u+v\|_2^2 = \|u-v\|_2^2$. Therefore:

$$|\langle Au, Av\rangle| \leq \frac{\delta_{2k}}{4} = \frac{\delta_{2k}}{2}\left[\|u\|_2^2 + \|v\|_2^2\right] = \delta_{2k}$$

□

(d) k-sparse partitioning of vectors

Lemma 1.5. *Let $\Lambda_0 \subset \{1, \dots, n\}$ such that $|\Lambda_0| \leq k$. For $u \in \mathbb{R}^n$, let Λ_1 be the k largest index set in $h_{\Lambda_0^c}$. Define Λ_2 to be the next largest, and so on. Then:*

$$\sum_{j \geq 2} \left\| u_{\Lambda_j} \right\| \leq \left\| u_{\Lambda_0^c} \right\|_1 / \sqrt{k}$$

Proof. Notice that for $j \geq 2$ that the largest component of h_{Λ_j} is smaller than the smallest component of $h_{\Lambda_{j-1}}$ and the average of the components in modulus of h_{Λ_j}. Therefore:

i. $\left\| h_{\Lambda_j} \right\|_\infty \leq \frac{1}{k} \left\| h_{\Lambda_{j-1}} \right\|_1$ for $j \geq 2$.

ii. So for $j \geq 2$ and by 1.2:

$$\begin{aligned} \sum_{j \geq 2} \left\| h_{\Lambda_j} \right\|_2 &\leq \sum_{j \geq 2} \sqrt{k} \left\| h_{\Lambda_j} \right\|_\infty \\ &\leq \sum_{j \geq 2} \frac{1}{\sqrt{k}} \left\| h_{\Lambda_{j-1}} \right\|_1 \\ &= \frac{1}{\sqrt{k}} \left\| h_{\Lambda_0^c} \right\|_1 \end{aligned}$$

□

(e) RIP implies less strict NSP

Lemma 1.6. *Let $\Lambda_0 \subset \{1, \dots, n\}$ such that $|\Lambda_0| \leq k$. Then for fixed h let Λ_1 be the next k largest index set in $h_{\Lambda_0^c}$. Let $\Lambda = \Lambda_0 \cup \Lambda_1$. Suppose A satisfies the RIP of order $2k$. then:*

$$\left\| h_\Lambda \right\|_2 \leq \alpha \frac{\left\| h_{\Lambda_0^c} \right\|_1}{\sqrt{k}} + \beta \frac{|\langle A h_\Lambda, A h \rangle|}{\left\| h_\Lambda \right\|_2}$$

Proof. Note that $h_\Lambda \in \Sigma_{2k}$, and $h_\Lambda = h - \sum_{j \geq 2} h_{\Lambda_j}$, by the RIP, and

Lemmas 1.4, 1.3, and 1.5:

$$
\begin{aligned}
(1-\delta_{2k})\left\|h_\Lambda\right\|_2^2 &\leq \left\|Ah_\Lambda\right\|_2^2 \\
&= \langle Ah_\Lambda, Ah\rangle - \sum_{j\geq 2}\langle Ah_\Lambda, Ah_{\Lambda_j}\rangle \\
&\leq |\langle Ah_\Lambda, Ah\rangle| + \left|\sum_{j\geq 2}\langle Ah_\Lambda, Ah_{\Lambda_j}\rangle\right| \\
&= |\langle Ah_\Lambda, Ah\rangle| + \left|\sum_{j\geq 2}\langle Ah_{\Lambda_1}, Ah_{\Lambda_j}\rangle + \langle Ah_{\Lambda_0}, Ah_{\Lambda_j}\rangle\right| \\
&\leq |\langle Ah_\Lambda, Ah\rangle| + \left|\sum_{j\geq 2}\left(\left\|h_{\Lambda_1}\right\|_2 + \left\|h_{\Lambda_0}\right\|_2\right)\delta_{2k}\left\|h_{\Lambda_j}\right\|_2\right| \\
&\leq |\langle Ah_\Lambda, Ah\rangle| + \sqrt{2}\left\|h_\Lambda\right\|_2\delta_{2k}\sum_{j\geq 2}\left\|h_{\Lambda_j}\right\|_2 \\
&\leq |\langle Ah_\Lambda, Ah\rangle| + \sqrt{2}\left\|h_\Lambda\right\|_2\delta_{2k}\frac{\left\|h_{\Lambda_0^c}\right\|_1}{\sqrt{k}}
\end{aligned}
$$

Rearranging, we have:

$$
\left\|h_\Lambda\right\|_2 \leq \frac{\sqrt{2}\delta_{2k}}{1-\delta_{2k}}\frac{\left\|h_{\Lambda_0^c}\right\|_1}{\sqrt{k}} + \frac{1}{1-\delta_{2k}}\frac{|\langle Ah_\Lambda, Ah\rangle|}{\left\|h_\Lambda\right\|_2}
$$

□

1.5 Noise-Free Signal Recovery

1. Spark, NSP and RIP are properties of a matrix that determine the number of samples needed to reasonably recover a sparse signal with noise

2. Motivation: to actually compute the signal, we use L^1 minimisation, which can be easily computed. We need to determine an upper bound in the error for the recovered signal $\hat{x}$ and the true signal x. So we explore the value of $\left\|x-\hat{x}\right\|_2$ in the noise free context

3. The L_1 minimisation algorithm for a sensing matrix which satisfies the RIP is instance-optimal

Theorem 1.6. *Suppose A satisfies the RIP of order $2k$ with $\delta_{2k} < \sqrt{2}-1$. Suppose we obtain measurements of the form $y = Ax$. Let $\hat{x} \in \arg\min \left\|z\right\|_1 : Az = y$. Then $\left\|\hat{x}-x\right\|_2 \leq \frac{C_0}{\sqrt{k}}\sigma_k(x)_1$*

Proof. By assumption $\left\|\hat{x}\right\|_1 \leq \left\|x\right\|_1$. Applying Lemma 1.7:

$$
\left\|\hat{x}-x\right\|_2 \leq \frac{C_0}{\sqrt{k}}\sigma_k(x)_1 + \frac{C_1}{\left\|h_\Lambda\right\|_2}\left|\langle A(\hat{x}-x)_\Lambda, A(\hat{x}-x)\rangle\right|
$$

Noting that $Ax = A\hat{x} = y$, then $A(\hat{x}-x) = 0$. The result follows. □

4. The following lemma is similar to Lemma 1.6, and is essential to proving the above result.

Lemma 1.7. *Suppose A satisfies the RIP of order $2k$. Let Λ_0 be the set of the index set for k largest components of x, and Λ_1 be the index set of the next k-largest components of $h_{\Lambda_0^c}$ where $h = \hat{x} - x$ for $x, \hat{x} \in \mathbb{R}^n$. Let $\Lambda = \Lambda_0 \cup \Lambda_1$. If $\|\hat{x}\|_1 \leq \|x\|_1$ then:*

$$\|h\|_2 \leq \frac{C_0}{\sqrt{k}}\sigma_k(x)_1 + \frac{C_1}{\|h_\Lambda\|_2} |\langle Ah_\Lambda, Ah\rangle|$$

Proof. Using the triangle inequality, $\|h\|_2 \leq \|h_{\Lambda^c}\|_2 + \|h_\Lambda\|_2$. We bound each term individually:

(a) By Lemma 1.5, we have that

$$\|h_{\Lambda^c}\|_2 \leq \sum_{j\geq 2} \left\|h_{\Lambda_j}\right\| \leq \frac{1}{\sqrt{k}} \left\|h_{\Lambda_0^c}\right\|_1$$

Therefore, we will use $\sigma_k(x)_1 = \|x - x_{\Lambda_0}\|_1 = \left\|x_{\Lambda_0^c}\right\|$ to upper bound $\left\|h_{\Lambda_0^c}\right\|_1$.

$$\begin{aligned}\|x\|_1 &\geq \|\hat{x}\|_1 = \|x + h\|_1 \\ &\geq \|x_{\Lambda_0}\|_1 - \|h_{\Lambda_0}\|_1 + \left\|h_{\Lambda_0^c}\right\|_1 - \left\|x_{\Lambda_0^c}\right\|_1\end{aligned}$$

Rearranging, and using Lemma 1.2:

$$\begin{aligned}\left\|h_{\Lambda_0^c}\right\|_1 &\leq \|x\|_1 - \|x_{\Lambda_0}\|_1 + \|h_{\Lambda_0}\|_1 + \left\|x_{\Lambda_0^c}\right\|_1 \\ &\leq \|x - x_{\Lambda_0}\|_1 + \|h_{\Lambda_0}\|_1 + \left\|x_{\Lambda_0^c}\right\|_1 \\ &\leq 2\sigma_k(x)_1 + \|h_{\Lambda_0}\|_1 \\ &\leq 2\sigma_k(x)_1 + \sqrt{k}\,\|h_{\Lambda_0}\|_2\end{aligned}$$

Therefore:

$$\|h_{\Lambda^c}\|_2 \leq \|h_\Lambda\|_2 + \frac{2}{\sqrt{k}}\sigma_k(x)_1$$

(b) By Lemma 1.6 and the previous part of this proof:

$$\begin{aligned}\|h_\Lambda\|_2 &\leq \frac{\alpha}{\sqrt{k}} \left\|h_{\Lambda_0^c}\right\|_1 + \beta\frac{|\langle Ah_\Lambda, Ah\rangle|}{\|h_\Lambda\|_2} \\ &\leq \frac{2\alpha}{\sqrt{k}}\sigma_k(x)_1 + \alpha\,\|h_\Lambda\|_2 + \beta\frac{|\langle Ah_\Lambda, Ah\rangle|}{\|h_\Lambda\|_2} \\ &\leq \frac{2\alpha}{\sqrt{k}(1-\alpha)}\sigma_k(x)_1 + \frac{\beta}{1-\alpha}\frac{|\langle Ah_\Lambda, Ah\rangle|}{\|h_\Lambda\|_2}\end{aligned}$$

(c) Combining the two results:

$$\begin{aligned}\|h\|_2 &\leq 2\|h_\Lambda\|_2 + 2\frac{\sigma_k(x)_1}{\sqrt{k}} \\ &\leq \left(\frac{4\alpha}{1-\alpha}+2\right)\frac{\sigma_k(x)_1}{\sqrt{k}} + \frac{2\beta}{1-\alpha}\frac{|\langle Ah_\Lambda, Ah\rangle|}{\|h_\Lambda\|_2} \\ &= 2\frac{1+\alpha}{1-\alpha}\frac{\sigma_k(x)_1}{\sqrt{k}} + \frac{2\beta}{1-\alpha}\frac{|\langle Ah_\Lambda, Ah\rangle|}{\|h_\Lambda\|_2}\end{aligned}$$

□

5. Remarks
   (a) If x is exactly k-sparse, $\|\hat{x}-x\|_2 = 0$ since $\sigma_k(x)_1 = 0$. Therefore, we would recover x exactly.
   (b) Since we are not dealing with noise, we can simplify Lemma 1.7 and the theorem by assuming an NSP of order $2k$.

1.6 Noisy Signal Recovery

1. Motivation: More often than note, $y = Ax + e$ is our sample where e is some sort of corruption, in which case, we want to estimate z to satisfy $\|Ax - y\|_2 \leq \|e\|_2$. If $\|e\|_2$ is bounded, we can bound the error between the L^1 estimator and the signal.
2. Error bound between the L^1 estimator and the signal

   Theorem 1.7. *Suppose $A \in \mathbb{R}^{m\times n}$ satisfies the RIP of order $2k$ with $\delta_{2k} < \sqrt{2}-1$. Let $y = Ax + e$, where $\|e\|_2 < \epsilon$. Let $\hat{x} \in \arg\min \|z\|_1 : \|Ax - y\|_2 \leq \epsilon$. Then:*

$$\|\hat{x}-x\|_2 \leq \frac{C_0}{\sqrt{k}}\sigma_k(x)_1 + C_2\epsilon$$

   Proof. Let $h = \hat{x} - x$. Let Λ_0 be the index set of the k-largest components of x, and Λ_1 be the index set for the k largest components of $h_{\Lambda_0^c}$. Let $\Lambda = \Lambda_0 \cup \Lambda_1$.

   (a) By assumption, $\|\hat{x}\|_1 \leq \|x\|_1$. Applying Lemma 1.7:

$$\|h\|_2 \leq \frac{C_0}{\sqrt{k}}\sigma_k(x)_1 + C_1\frac{|\langle Ah_\Lambda, Ah\rangle|}{\|h_\Lambda\|_2}$$

   (b) By Cauchy-Schwarz: $|\langle Ah_\Lambda, Ah\rangle| \leq \|Ah_\Lambda\|_2 \|Ah\|_2$. By RIP,

$$\frac{\|Ah_\Lambda\|_2}{\|h_\Lambda\|_2} \leq \sqrt{1+\delta_{2k}}$$

   (c) By triangle-inequality, definition of the estimator, and our assumption on the error:

$$\|Ah\|_2 \leq \|Ax - y\| + \|A\hat{x} - y\| \leq 2\|e\|_2 \leq 2\epsilon$$

(d) Therefore,

$$\|\hat{x} - x\|_2 \leq \frac{C_0}{\sqrt{k}} \sigma_k(x)_1 + C_1 2\sqrt{1+\delta_{2k}}\epsilon$$

□

1.7 Coherence

1. Motivation: Spark, NSP, and RIP all provide guarantees for recovering k-sparse signals, but verifying these properties is difficult. The property of coherence is easily computed and can be related to spark, NSP and RIP under certain conditions.

2. Note: we will be using two definitions of coherence. One is presented below. The other is the maximal modulus element of the matrix.

3. Definition:

   Definition 1.5. *The coherence of A with columns a_i is:*

   $$\mu(A) = \max_{1 \leq i < j \leq n} \frac{|\langle a_i, a_j \rangle|}{\|a_i\|_2 \|a_j\|_2}$$

4. Important Properties (without proof):

   Theorem 1.8. *Let $\mu(A)$ be the coherence defined above of a matrix A.*

   (a) For any matrix A, $\operatorname{spark}(A) \geq 1 + \frac{1}{\mu(A)}$

   (b) If $k < \frac{1}{2}\left(1 + \frac{1}{\mu(A)}\right)$ then for all $y \in \mathbb{R}^m$, there exists at most 1 $x \in \Sigma_k$ such that $y = Ax$.

   (c) If A has unit norm columns and coherence $\mu(A)$, then A satisfies the RIP of order k with $\delta_k = (k-1)\mu(A)$ for all $k < \frac{1}{\mu(A)}$.

2 Signal Recovery in Random Sampling

2.1 Bernstein Inequalities

1. Bernstein Inequality

   Theorem 2.1. *Suppose $Y_1, \ldots, Y_n$ are independent zero-mean random variables and $|Y_j| \leq R$ a.s. If $\mathbb{E}\left[|Y_j|^2\right] \leq \sigma_j^2$ then for all $t > 0$:*

   $$\mathbb{P}\left[\left|\sum Y_j\right| \geq t\right] \leq 2\exp\left(\frac{-t^2/2}{\sum \sigma_j^2 + \frac{1}{3}RT}\right)$$

2. Vector Bernstein Inequality

Theorem 2.2. *Let $Y_1, \ldots, Y_n \in \mathbb{C}^d$ be independent zero-mean random vectors such that $\|Y_j\|_2 \leq R$ a.s. Let $\sigma^2 \geq \sum \mathbb{E}\left[\|Y_j\|_2^2\right]$. Then for $0 < t$:*

$$\mathbb{P}\left[\left\|\sum Y_j\right\|_2 \geq \sigma + t\right] \leq \exp\left(\frac{-t^2/2}{\sigma^2 + (6\sigma + t)R/3}\right)$$

3. Square Matrix Bernstein Inequality

   Theorem 2.3. *Let $Y_1, \ldots, Y_n \in \mathbb{C}^{d \times d}$ are self-adjoint (equal to its conjugate transposed), zero-mean random matrices. Suppose the largest eigenvalue of Y_j satisfies $\lambda_{max}(Y_j) \leq R$ a.s.. Let $\sigma^2 = \|\sum \mathbb{V}[Y_j]\|_{2\to 2}$. Then*

$$\mathbb{P}\left[\lambda_{max}(\sum Y_j) \geq t\right] \leq d \exp\left(\frac{-t^2/2}{\sigma^2 + Rt/3}\right)$$

4. Rectangular Matrix Bernstein Inequality

   Theorem 2.4. *Let $Y_1, \ldots, Y_n \in \mathbb{C}^{d_1 \times d_2}$ such that $\|Y_l\|_{2\to 2} \leq R$ for all l. Let $\sigma^2 = \max\{\|\sum \mathbb{E}[Y_l Y_l^*]\|, \|\sum \mathbb{E}[Y_l^* Y_l]\|\}$. Then for $t > 0$:*

$$\mathbb{P}\left[\left\|\sum Y_l\right\|_{2\to 2} \geq t\right] \leq 2(d_1 + d_2) \exp\left(\frac{-t^2/2}{\sigma^2 + Rt/3}\right)$$

2.2 RIP Less Theory of Compressed Sensing

1. Motivation: in some situations, the RIP theory of compressive sensing requires an impossible number of measurements to be taken. The RIP-less theory holds in a probabilistic setting, allowing us to circumvent these issues. Even in the RIPless setting, we can still construct sensing matrices A, and must determine when unique solutions can be recovered.

2. Notation

   (a) For $a \in \mathbb{C}$, let $\operatorname{sgn}(a) = \begin{cases} a/|a| & \text{if } a \neq 0 \\ 0 & \text{if } a = 0 \end{cases}$

   (b) For $x \in \mathbb{C}^n$, let $\operatorname{sgn}(x) = \begin{pmatrix} \operatorname{sgn}(x_1) \\ \vdots \\ \operatorname{sgn}(x_n) \end{pmatrix}$

   (c) For $S \subset \{1, \ldots, n\}$ and $A \in \mathbb{C}^{m \times n}$, $A_S = AP_S$, which is a restriction to columns in S.

3. Inexact Dual Certificate: Properties of a sensing matrix A which ensure a unique L^1 minimiser.

   Theorem 2.5. *Let $A \in \mathbb{C}^{m \times n}$ with columns a_l for $l \in \{1, \ldots, n\}$. let $x \in \mathbb{C}^n$ with $\operatorname{supp}(x) = S$ (i.e. the non-zero components of x are indexed by S). Let $\alpha, \beta, \gamma, \theta > 0$ such that:*

   (a) Firstly:

$$\left\|\left(A_s^* A_s|_{\mathrm{Range}(P_S)}\right)^{-1}\right\|_{2\to 2} \leq \alpha \qquad \max_{l \in S^c} \|A_s^* a_l\|_2 \leq \beta$$

(b) *Secondly, suppose* $\exists u \in \mathbb{C}^n$ *such that* $u = A^* h$ *for* $h \in \mathbb{C}^m$ *where*

$$\|u_s - \operatorname{sgn}(x_s)\|_2 \leq \gamma \qquad \|u_{S^c}\| \leq \theta$$

If $\theta + \alpha\beta\gamma < 1$ *then* x *is the unique minimiser of* $\min \|z\|_1 : Ax = Az$

Proof. Let $\hat{x} = \arg\min \|z\|_1 : Ax = Az$. Let $v = \hat{x} - x$. We want to show that $v = 0$

(a) We do this by showing $\|\hat{x}\|_1 \geq \|x\|_1 + Constant \times \|v_{S^c}\|_1$ which implies that $\|v_s\|_1 = 0$ since, by construction, $\|\hat{x}\|_1 \leq \|x\|_1$

(b) Note that using the definition of $\operatorname{sgn}(a + bi)$ and using generic vectors $a + bi$ and $c + di$ we can show the last line:

$$\begin{aligned}\|\hat{x}\|_1 &= \|x + v\|_1 \\ &= \|x + v_S\|_1 + \|v_{S^c}\|_1 \\ &= \langle \operatorname{sgn}(x + v_S), x + v_S \rangle + \|v_{S^c}\|_1 \\ &\geq \|x\|_1 - \operatorname{Re}(\langle \operatorname{sgn}(x), v_S \rangle) + \|v_{S^c}\|_1\end{aligned}$$

(c) Now we want to upper bound $\operatorname{Re}(\langle \operatorname{sgn}(x), v_S \rangle)$. We use u_s to do so. Note that:

$$\begin{aligned}\langle u_S, v_S \rangle &= \langle u, v \rangle - \langle u_S, v_{S^c} \rangle - \langle u_{S^c}, v_S \rangle - \langle u_{S^c}, v_{S^c} \rangle \\ &= \langle u, v \rangle - \langle u_{S^c}, v_{S^c} \rangle \\ &= \langle A^* h, v \rangle - \langle u_{S^c}, v_{S^c} \rangle \\ &= \langle h, Av \rangle - \langle u_{S^c}, v_{S^c} \rangle \\ &= -\langle u_{S^c}, v_{S^c} \rangle\end{aligned}$$

Therefore, by Cauchy-Schwarz:

$$\begin{aligned}|\langle \operatorname{sgn}(x), v_S \rangle| &= |\langle \operatorname{sgn}(x) - u_S, v_S \rangle + \langle u_S, v_S \rangle| \\ &\leq |\langle \operatorname{sgn}(x) - u_S, v_S \rangle| + |\langle u_{S^c}, v_{S^c} \rangle| \\ &\leq \|\operatorname{sgn}(x) - u_S\|_2 \|v_S\|_2 + \|u_{S^c}\|_\infty \|v_{S^c}\|_1 \\ &\leq \gamma \|v_S\|_2 + \theta \|v_{S^c}\|_1\end{aligned}$$

So we have:

$$\|\hat{x}\|_1 \geq \|x\|_1 - \gamma \|v_S\|_2 + (1 - \theta) \|v_{S^c}\|_1$$

(d) Now we want to get $\|v_S\|_2$ in terms of $\|v_{S^c}\|_1$. Noting that $Av = 0 \implies A_S v_S + A_{S^c} v_{S^c} = 0$,

$$\|v_S\|_2 = \left\| (A_S^* A_S)^{-1} [A_S^* (A_{S^c} v_{S^c})] \right\|_2 \leq \alpha \|A_S^* A_{S^c} v_{S^c}\|_2$$

Moreover,

$$\begin{aligned}\|A_S^* A_{S^c} v_{S^c}\|_2 &= \left\| \sum_{l \in S^c} A_S^* a_l v_l \right\|_2 \\ &\leq \sum_{l \in S^c} |v_l| \|A_S^* a_l\|_2 \\ &\leq \beta \sum_{l \in S^c} |v_l| \\ &= \beta \|v_{S^c}\|_1\end{aligned}$$

(e) Therefore, we have that $\|\hat{x}\| \geq \|x\|_1 + (1 - \theta - \alpha\beta\gamma) \|v_{S^c}\|_1$. By assumption, $1 > \theta + \alpha\beta\gamma$. So $v_{S^c} = 0$ which implies that $v_S = 0$ since $A_S v_S = -A_{S^c} v_{S^c} = 0$. So $v = 0$. And we recover x exactly.

□

2.3 Random Sampling in Bounded Orthonormal Systems

1. Motivation: A random sensing matrix confers unique advantages to reconstructing signals. However, structured random matrices, those generated by a random choice of parameters, have many computational advantages over completely unstructured random matrices

2. Definition of Bounded Orthonormal System (BOS)

   Definition 2.1. *Suppose $D \subseteq \mathbb{R}^d$ with a probability measure ν. Let $\Phi = \{\phi_1, \ldots, \phi_n\}$ be an orthonormal system of complex valued functions on D with respect to ν (i.e. $\int_D \phi_i(t)\overline{\phi_i(t)} d\nu(t) = \delta_{ij}$). Φ is a Bounded Orthonormal System if $\exists K > 0$ such that $\|\phi_i\|_\infty \leq K$ for $i = 1, \ldots, n$*

3. Goal: consider a function $f(t) = \sum_{k=1}^n x_k \phi_k(t)$. Suppose $t_1, \ldots, t_m \in D$ are randomly sampled points with measurements $y_i = f(t_i)$. Then we have measurements $y = \begin{pmatrix} y_1 & \cdots & y_m \end{pmatrix}^T$, sensing matrix $A_{ij} = \phi_j(t_i)$, and unknown parameters $x = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}^T$. By determining x we can construct the function f.

2.4 Sampling for Recovery

1. Motivation: Ultimately, we want a lower bound guarantee on the number of sampled points m that allows us to recover x with some probability

2. Applying the Bernstein Inequality

   Lemma 2.1. *Let $A \in \mathbb{C}^{m \times n}$ be a random sampling matrix with respect to a BOS. Let $\tilde{A} = \frac{1}{\sqrt{m}} A$. Let $v \in \mathbb{C}^n$ with $\operatorname{supp}(v) = S$, and $|S| = s$. Then for $t > 0$*

$$\mathbb{P}\left[\left\|\tilde{A}_{S^c}^* \tilde{A} v\right\|_\infty > t \|v\|_2\right] \leq 4n \exp\left(\frac{-m}{4K^2} \frac{t^2}{1 + \sqrt{\frac{s}{18}} t}\right)$$

   Proof. We will be applying the Bernstein Inequality to prove this solution. So we need to find a suitable RV and prove it has 0 mean, are bounded and have bounded variance.

   (a) Note that $\left\|\tilde{A}_{S^c}^* \tilde{A} v\right\|_\infty = \max_{k \in S^c} \left|\langle e_k, \tilde{A}^* A v \rangle\right|$. Let $X_l = \{\overline{\phi_j(t_l)}\}_{j=1}^n$. Then X_l are the columns of A^*. Without loss of generality, let $\|v\|_2 = 1$ and let $Y_l = \langle e_k, X_l X_l * v \rangle$

   (b) We compute the expectation of Y, noting that

$$\mathbb{E}\left[(X_l X_l^*)_{ij}\right] = \mathbb{E}\left[\overline{\phi_i(t_l)} \phi_j(t_l)\right] = \delta_{ij}$$

   Implies, since $k \in S^c$ and $\operatorname{supp}(v) = S$:

$$\mathbb{E}[Y_l] = \langle e_k, \mathbb{E}[X_l X_l^*] v \rangle = \langle e_k, Iv \rangle = 0$$

(c) We now bound Y_l:

$$\begin{aligned}
|Y_l| &= |\langle e_k, X_l X_l^* v\rangle| \\
&= |\langle e_k, X_l\rangle|\,|\langle X_l|_{k\in S}, v\rangle| \\
&= |\phi_k(t_l)|\,|\langle X_l|_{k\in S}, v\rangle| \\
&\le K \,\|X_l|_{k\in S}\|_2 \,\|v\|_2 \\
&= K \,\|X_l|_{k\in S}\|_2 \\
&= K \sqrt{\sum_{s\in K} |\phi_k(t_l)|^2} \\
&\le K^2 \sqrt{\sum_{s\in K} 1} \\
&= K^2 \sqrt{s}
\end{aligned}$$

(d) We now bound the variance:

$$\begin{aligned}
\mathbb{E}\left[Y_l Y_l^*\right] &= \mathbb{E}\left[\langle e_k, X_l\rangle\langle X_l, v\rangle\langle v, X_l\rangle\langle X_l, e_k\rangle\right] \\
&= \mathbb{E}\left[|\phi_k(t_l)|^2 \, v^* X_l X_l^* v\right] \\
&\le K^2 v^* \mathbb{E}\left[X_l X_l^*\right] v \\
&= K^2 v^* v \\
&= K^2
\end{aligned}$$

(e) Before we apply the Bernstein Inequality, we note that for any $z \in \mathbb{C}$

$$|z|^2 = \mathrm{Re}(z)^2 + \mathrm{Im}(z)^2 \le 2\mathrm{Re}(z)^2 \wedge 2\mathrm{Im}(z)^2$$

(f) In light of the previous fact, and applying Bernstein's Inequality:

$$\begin{aligned}
\mathbb{P}\left[\left|\langle e_k, \tilde{A}^* \tilde{A} v\rangle\right| > t\right] &\le \mathbb{P}\left[\frac{1}{m}\sum_{l=1}^{m} \mathrm{Re}(Y_l) > \frac{t}{\sqrt{2}}\right] \\
&\quad + \mathbb{P}\left[\frac{1}{m}\sum_{l=1}^{m} \mathrm{Im}(Y_l) > \frac{t}{\sqrt{2}}\right] \\
&\le 4\exp\left(\frac{-m^2 t^2/4}{mK^2 + \sqrt{s}K^2 mt/\sqrt{18}}\right) \\
&= 4\exp\left(\frac{-m}{4K^2}\frac{t^2}{1 + \sqrt{\frac{s}{18}}t}\right)
\end{aligned}$$

(g) Noting that $\mathbb{P}\left[\max_{k\in S^c} Z_l > t\right] \le \sum_{k=1}^{n} \mathbb{P}\left[Z_l > t\right]$ for some random variable Z_l, the result follows.

□

3. Applying the Matrix Bernstein Inequality

Lemma 2.2. *Let $A \in \mathbb{C}^{m\times n}$ random matrix corresponding to a BOS with $K \geq 1$. Let $S \subseteq \{1,\ldots,n\}$ and $|S| = s$. Then for $\delta \in (0,1)$, $\tilde{A} = \frac{1}{\sqrt{m}}A$, and $I = P_S$:*

$$\mathbb{P}\left[\left\|\tilde{A}_S^*\tilde{A}_S - I\right\|_{2\to 2} > \delta\right] \leq 2s\exp\left(\frac{-3m\delta^2}{8K^2s}\right)$$

Proof. We want to use matrix Bernstein inequality for which we need Y_l independent, self-adjoin, zero-mean, bounded, and bounded summed covariance in spectral norm.

(a) Let X_l be the columns of A_S^* which are independent since t_l are independent. Let $Y_l = X_lX_l^* - I = Y_l^*$. So Y_l are independent, self-adjoint, and:

$$\tilde{A}_S^*\tilde{A}_S^* - I = \frac{1}{m}\sum_{l=1}^m Y_l$$

(b) $\mathbb{E}[Y_l] = \mathbb{E}[X_lX_l^* - I] = 0$

(c) Now we bound Y_l:

$$\begin{aligned}
\|Y_l\|_{2\to 2} &= \|X_lX_l^* - I\|_{2\to 2}\\
&= \max_{\|v\|_2=1}\left|v^*(X_lX_l^*)v - \|v\|_2^2\right|\\
&= \max_{\|v\|_2=1}\left|(\langle X_l, v\rangle)^2 - 1\right|\\
&\leq \max_{\|v\|_2=1}\left|\|X_l\|_2^2\,\|v\|_2^2 - 1\right|\\
&= \left|\|X_l\|_2^2 - 1\right|\\
&\leq \sum_{k\in S}\left|\overline{\phi_k(t_l)}\right|^2\\
&\leq K^2s
\end{aligned}$$

(d) Noting that the covariance of Y_l has the following property (in the usual sense):

$$\begin{aligned}
\mathbb{V}[Y_l] =&\mathbb{E}[Y_lY_l^*]\\
=&\mathbb{E}\left[(X_lX_l^* - I)^2\right]\\
=&\mathbb{E}[X_l(X_l^*X_l)X_l^* - 2X_lX_l^* + I]\\
=&\mathbb{E}\left[\|X_l\|_2^2\,X_lX_L^*\right] - I\\
\leq&K^2sI - I\\
\leq&K^2sI
\end{aligned}$$

Then:

$$\left\|\sum_{l=1}^m \mathbb{V}[X_l]\right\|_{2\to 2} \leq \left\|mK^2sI\right\|_{2\to 2} = mK^2s$$

(e) Applying the Matrix Bernstein Inequality:

$$\begin{aligned}\mathbb{P}\left[\left\|\tilde{A}_S^*\tilde{A}_S - I\right\|_{2\to2} > \delta\right] =&\mathbb{P}\left[\left\|\sum_{l=1}^m Y_l\right\|_{2\to2} > m\delta\right]\\ \leq&2s\exp\left(\frac{-\delta^2 m/2}{K^2 s(1+\delta/3)}\right)\\ \leq&2s\exp\left(\frac{-\delta^2 m}{K^2 s}\frac{3}{6+2\delta}\right)\\ \leq&2s\exp\left(\frac{-3\delta^2 m}{8K^2 s}\right)\end{aligned}$$

□

4. Application of Vector Bernstein Inequality

Lemma 2.3. *Let $S \subseteq \{1,\dots,n\}$ with $|S| = s$. Let $v \in \mathbb{C}^s$ and $\|v\|_2 = 1$. Then for $t > 0$:*

$$\mathbb{P}\left[\left\|\left(\tilde{A}_S^*\tilde{A}_S - I\right)v\right\|_2 \geq \sqrt{\frac{K^2 s}{m}} + t\right] \leq \exp\left(\frac{-mt^2}{2K^2 s}\frac{1}{1+2\sqrt{K^2 s/m}+t/3}\right)$$

Proof. Here we want to apply the vector Bernstein inequality, which requires a random vector with mean zero, bounded in norm, and with $\sigma^2 = \sup_{\|z\|_2=1}\mathbb{E}\left[|\langle Y, z\rangle|^2\right]$ bounded.

(a) Let X_l^* be a column vector of A_S^*. Let $Y_l = (X_l X_l^* - I)v$. Then:

$$\left(\tilde{A}_S^*\tilde{A}_S - I\right)v = \frac{1}{m}\sum_{l=1}^m Y_l$$

(b) Zero Mean:

$$\mathbb{E}\left[Y_l\right] = \mathbb{E}\left[X_l X_l^* - I\right]v = (I - I)v = 0$$

(c) Bounded in Norm:

$$\|Y_l\|_2 = \|(X_l X_l^* - I)v\| \leq K^2 s$$

(d) Bounded σ^2:

$$\begin{aligned}\sigma^2 \leq&m \max_{\|z\|_2=1}\mathbb{E}\left[|\langle Y_l, z\rangle|^2\right]\\ \leq&m \max_{\|z\|_2=1}\mathbb{E}\left[\|Y_l\|_2^2\right]\|z\|_2^2\\ =&mv^*\mathbb{E}\left[\|X_l\|_2^2 X_l X_l^* - 2X_l X_l^* + I\right]v\\ \leq&mv^*(K^2 s - 1)v\\ \leq&mK^2 s\end{aligned}$$

(e) Therefore, by the Vector Bernstein Inequality:

$$\begin{aligned}
&\mathbb{P}\left[\left\|\left\|\left(\tilde{A}_S^*\tilde{A}_S - I\right)v\right\|_2 \geq \sqrt{\frac{K^2 s}{m}} + t\right\|\right] \\
&= \mathbb{P}\left[\left\|\sum Y_l\right\|_2 \geq \sqrt{mK^2 s} + mt\right] \\
&\leq \exp\left(\frac{-m^2t^2/2}{msK^2 + (6\sqrt{msK^2} + mt)sK^2/3}\right) \\
&= \exp\left(\frac{-mt^2}{2K^2 s}\frac{1}{1 + 2\sqrt{K^2 s/m} + t/3}\right)
\end{aligned}$$

□

5. Application of Rectangular Matrix Bernstein Inequality

Lemma 2.4. *For $0 < t \leq 2\sqrt{s}$, and $\tilde{a}_j$ the j^{th} column of $\tilde{A} = \frac{1}{\sqrt{m}}A$:*

$$\mathbb{P}\left[\max_{j \in S^c}\left\|\tilde{A}_S^*\tilde{a}_j\right\|_2 \geq t\right] \leq 2(s+1)n\exp\left(\frac{-3}{10}\frac{mt^2}{K^2 s}\right)$$

Proof. We want to apply the rectangular matrix Bernstein for $s \times 1$ matrices. We need to find Y_l with mean zero, bounded and bounded σ^2

(a) Let X_l be the column vector of A_S^*. Let $Y_l = X_l\phi_j(t_l)$ for $j \in S^c$ fixed.

(b) Zero-mean, since $k \in S$ and $j \in S^c$:

$$\mathbb{E}\left[Y_l\right]_k = \mathbb{E}\left[\overline{\phi_k(t_l)}\phi_j(t_l)\right] = 0$$

(c) Bounded in Norm

$$\|Y_l\|_2 = \|X_l\phi_j(t_l)\|_2 \leq |\phi_j(t_l)|\,\|X_l\|_2 \leq K \times K\sqrt{s}$$

(d) Bounded $\sigma^2 = \max\left[\|\sum_{l=1}^m \mathbb{E}\left[Y_lY_l^*\right]\|_{2\to 2}, \|\sum_{l=1}^m \mathbb{E}\left[Y_l^*Y_l\right]\|\right]$. For the first term:

$$\mathbb{E}\left[Y_lY_l^*\right] = \mathbb{E}\left[|\phi_j(t_l)|^2 X_lX_l^*\right] \leq K^2 I$$

For the second term, since the system is orthonormal:

$$\begin{aligned}
\mathbb{E}\left[Y_l^*Y_l\right] &= \mathbb{E}\left[|\phi_j(t_l)|^2 X_l^*X_l\right] \\
&\leq K^2\sum_{k \in S}\mathbb{E}\left[\phi_k(t_l)\overline{\phi_k(t_l)}\right] \\
&= K^2\sum_{k \in S} 1 \\
&= K^2 s
\end{aligned}$$

Therefore, $\sigma^2 \leq mK^2 s$.

(e) We apply the Rectangular Matrix Bernstein for $t \leq 2\sqrt{s}$

$$\mathbb{P}\left[\left\|\tilde{A}_S^* \tilde{a}_j\right\|_2 \geq t\right] = \mathbb{P}\left[\left\|\sum_{l \in S} \frac{1}{m} Y_l\right\|_2 \geq t\right]$$
$$\leq 2(s+1)\exp\left(\frac{-m^2t^2/2}{\sigma^2 + K^2\sqrt{s}mt/3}\right)$$
$$\leq 2(s+1)\exp\left(\frac{-3mt^2}{6K^2 s + 4K^2 s}\right)$$

Therefore:, the result follows from:

$$\mathbb{P}\left[\max_{j \in S^c}\left\|\tilde{A}_S^* \tilde{a}_j\right\|_2 \geq t\right] \leq \sum_{j=1}^{n} \mathbb{P}\left[\left\|\tilde{A}_S^* \tilde{a}_j\right\|_2 \geq t\right]$$

□

6. Probably of Satisfying the Inexact Dual Certificate for a BOS:

Theorem 2.6. *Let $x \in \mathbb{C}^n$ be an s-sparse vector. Choose $A \in \mathbb{C}^{m \times n}$ random matrix corresponding to a BOS with $K \geq 1$. If*

$$m \geq CK^2 s\left[2\log(4N)\log(12\epsilon^{-1}) + \log(s)\log(12\epsilon^{-1}\log(s))\right]$$

for some constant $C > 0$ then x is the unique minimiser for $\|z\|_1 : Ax = Az$ with probability of at least $1 - \epsilon$.

Proof. We want to satisfy the conditions of the inexact dual certificate, which allow us to recover a sparse solution. These conditions are that for some $\alpha, \beta, \gamma, \theta$:

1 For a_l the l^{th} column of A:

$$\left\|(A_S^* A_S)^{-1}\right\|_{2\to 2} \leq \alpha \qquad \max_{l \in S^c}\|A_S^* a_l\|_2 \leq \beta$$

2 There exists a $u \in \mathbb{C}^n$ for which $u = A^*h$ for some $h \in \mathbb{C}^m$ such that:

$$\|u_S - \operatorname{sgn}(x_S)\|_2 \leq \gamma \qquad \|u_{S^c}\| \leq \theta$$

The proof depends on the Golfing Framework, which we set up first. Then we show that property **2** holds with some probability, from which we can show property **1** will hold as well.

(a) Golfing Scheme Construction. Let $m = \sum_{k=1}^{L} m_k$. Let $A^{(1)} \in \mathbb{C}^{m_1 \times N}, \ldots, A^{(L)} \in \mathbb{C}^{m_L \times N}$. Therefore, letting $\tilde{A} = \frac{1}{\sqrt{m}}A$, where A is:

$$A = \begin{bmatrix} A^{(1)} \\ \vdots \\ A^{(L)} \end{bmatrix} \in \mathbb{C}^{m \times N}$$

We select u in a recursive way hoping it will satisfy **2**. Letting $u^{(0)} = 0 \in \mathbb{C}^N$, define the sequence:

$$u^{(n)} = \frac{1}{m_n}\left(A^{(n)}\right)^* A_S^{(n)}(\operatorname{sgn}(x_S) - u_S^{(n-1)}) + u^{(n-1)} \in \mathbb{C}^N$$

i. By construction, $u^{(L)} = A^* h$ for $h \in \mathbb{C}^m$

ii. Defining $w^{(n)} = \operatorname{sgn}(x_S) - u_S^{(n)}$:

$$w^{(n)} = \left(I - \frac{1}{m_n}\left(A_S^{(n)}\right)^* A_S^{(n)}\right) w^{(n-1)}$$

$$w^{(n)} = \left(\prod_{j=1}^{n} I - \frac{1}{m_j}\left(A_S^{(j)}\right)^* A_S^{(j)}\right) \operatorname{sgn}(x_S)$$

$$u^{(n)} = \sum_{i=1}^{n} \frac{1}{m_i}\left(A^{(i)}\right) A_S^{(i)} w^{(i-1)}$$

(b) Demonstrating property **2**.

i. Notice, we want to apply Lemma 2.3 for some choice of $r_n > 0$:

$$\begin{aligned}
\left\|u_S^{(n)} - \operatorname{sgn}((x_S))\right\|_2 &= \left\|w^{(n)}\right\|_2 \\
&\leq \left\|w^{(n)}\right\|_2 \left\|\frac{1}{m_n}\left(A_S^{(n)}\right)^* A_S^{(n)} - I\right\|_{2\to 2} \\
&\leq \left\|w^{(n)}\right\|_2 \left[\sqrt{K^2 s/m_n} + r_n\right] \\
\left\|u_S^{(L)} - \operatorname{sgn}((x_S))\right\|_2 &\leq \|\operatorname{sgn}(x_S)\|_2 \prod \left[\sqrt{K^2 s/m_n} + r_n\right] \\
&\leq \sqrt{s} \prod \left[\sqrt{K^2 s/m_n} + r_n\right]
\end{aligned}$$

Therefore by Lemma 2.3, the first inequality has probability of failure:

$$p_1(n) \leq \exp\left(\frac{-m_n r_n^2}{2K^2 s} \frac{1}{1 + 2\sqrt{K^2 s/m_n} + r_n/3}\right)$$

So the overall probability of failure is going to be

$$\sum_{n=1}^{L} p_1(n)$$

ii. Now to get the second part of property **2**, we want to use Lemma 2.1. Therefore, for some choice of t_n:

$$\begin{aligned}
\left\|u_{S^c}^{(L)}\right\|_\infty &\leq \sum_{i=1}^{L} \left\|\frac{1}{m_n}\left(A_{S^c}^{(n)}\right)^* A_S^{(n)} w^{(n-1)}\right\|_\infty \\
&\leq \sum_{n=1}^{L} t_n \left\|w^{(n-1)}\right\|_2 \\
&\leq \sum_{n=1}^{L} t_n \sqrt{s} \prod_{n'=1}^{n-1} \left[\sqrt{K^2 s/m_{n'}} + r_{n'}\right]
\end{aligned}$$

By Lemma 2.1, we fail at bounding each term with a probability of:

$$p_2(n) \leq 4N \exp\left(\frac{-m_n t_n^2}{4K^2} \frac{1}{1 + \sqrt{s/18}t_n}\right)$$

We fail at abounding the entire sum by:

$$\sum_{L=1}^{n} p_2(n)$$

iii. We now choose m_n, r_n, t_n, L, C:
   A. $m_1 = m_2 \geq CK^2 s \log(4N) \log(2\epsilon^{-1})$ and $m_n \geq CK^2 s \log(2L\epsilon^{-1})$
   B. $r_1 = r_2 = \frac{1}{2e\sqrt{\log 4N}}$ and $r_n = \frac{1}{2e}$
   C. $t_1 = t_2 = \frac{1}{e\sqrt{s}}$ and $t_n = \frac{\log 4N}{e\sqrt{s}}$
   D. $L = \lceil \log(s)/2 \rceil + 2$
   E. $C = 8e^2 \left[1 + 2e^{-1}\left(\frac{1}{\sqrt{8}} + \frac{1}{6}\right)\right]$

iv. These choices imply the following results:
   A. The choice of m_n and r_n implies:

$$\sqrt{K^2 s/m_1} + r_1 \leq \frac{1}{e\sqrt{\log 4N}} \qquad \sqrt{K^2 s/m_n} + r_n \leq \frac{1}{e}$$

   B. $\|u_S - \operatorname{sgn}(x_S)\|_2 \leq e^{-2}$ for which failure is $\sum p_1(n) \leq 2\epsilon$
   C. $\|u_{S^c}\|_\infty \leq \frac{1}{\epsilon - 1}$ for which failure is $\sum p_2(n) \leq 2\epsilon$
   D. So **2** holds with probability $1 - 4\epsilon$

(c) Demonstrating property **1**:

i. By Lemma 2.2, $\left\|\tilde{A}_S^* \tilde{A}_S - I\right\|_{2\to 2} \leq \frac{1}{2}$ with probability of failure $2s \exp\left(\frac{-3m}{32K^2 s}\right)$. Therefore, $\left\|(A_S^* A_S)^{-1}\right\|_{2\to 2} < \frac{2}{m}$ holds with the same probability, which we want to be bounded by ϵ. This requires that:

$$m > 32/3K^2 s \log(2s\epsilon^{-1})$$

which can be satisfied by the value of m selected to demonstrate **2**.

ii. Note that $\alpha = 2/m$, $\gamma = e^{-2}$, $\theta = (e-1)^{-1}$. To satisfy $\theta + \alpha\beta\gamma < 1$, we require $\beta/m < 1.554$ so we choose $\beta/m < 3/2$. By Lemma 2.4

$$\max_{l \in S^c} \left\|\tilde{A}_S^* \tilde{a}_l\right\|_2 < \beta/m$$

with a probability of failure of at most

$$2N^2 \exp\left(\frac{-27m}{40K^2 s}\right)$$

We can make this probability less than ϵ with the selection of m above.

(d) Overall, $\alpha, \beta, \gamma, \theta$ behave with a probability of $1 - 4\epsilon + 1 - \epsilon + 1 - \epsilon$ for the appropriate selection of m. Replacing 6ϵ by ϵ gives us the desired result.

□

Part II
Generalised Sampling