MTH 995-003: Intro to CS and Big Data	Spring 2014
Lecture 19 — Mar 11th, 2014	
Inst. Mark Iwen	*Scribe: Xianfeng Hu*

1 Bounded Orthonormal Systems(BONS) and the RIP

Let $D \subset \mathbb{R}^d$, ν be a probability measure on D, and $\Phi = \{\phi_1, \phi_2, \cdots, \phi_N\}$ be an orthonormal set of functions, $\phi_j : D \to \mathbb{C}$, $j \in [N]$, with respect to ν. That is, suppose that

$$\int_D \phi_j(\vec{t})\overline{\phi_k(\vec{t})}\, d\nu(\vec{t}) = \begin{cases} 0 & \text{if j}\neq\text{k} \\ 1 & \text{if j=k} \end{cases} .$$

Definition 1. *We will call an ONS Φ a bounded ONS with constant K if*

$$\max_{j\in[N]} \|\phi_j\|_\infty := \max_{j\in[N]} \left(\sup_{\vec{t}\in D} |\phi_j(\vec{t})| \right) \leq K$$

HW:

- Problems one and two can be found in Lecture 15.
- 3.) Prove $K \geq 1$ must hold;
- 4.) Do 12.1 in page 431;
- 5.) Do 12.2 in page 431.

1.1 Examples of Bounded ONS

Example 1. *Trigonometric polynomials are BONS (with $K = 1$).*
Let $D = [0, 1]$ and set $\phi_\omega(t) := e^{2\pi i\omega t}$ for any $\omega \in \mathbb{Z}$. Let ν be the uniform (Lebesgue measure) on $[0, 1]$, and restrict $\omega \in [N]$ (for example).

Then $|\phi_\omega(t)| = 1 \ \ \forall\omega, t \Rightarrow \Phi := \{\phi_1, \ldots, \phi_N\}$ is a BONS with $K = 1$.

Example 2. *Consider DFT matrix $F \in \mathbb{C}^{N\times N}$,*

$$F_{l,k} := \tfrac{1}{\sqrt{N}} e^{-2\pi i(l-1)(k-1)/N},\ \forall\, l, k \in [N]$$

Let ν be discrete uniform measure on $[N]$, s.t $\nu(B) = |B|/N$, $\forall\, B \subset [N]$, and $D = [N]$. Set $\phi_\omega(t) := \sqrt{N} F_{t,\omega}$ (i.e., our functions are the columns of F). Once again, $|\phi_\omega(t)| = 1 \ \ \forall\omega, t \Rightarrow K = 1$ works for our bound. And, the system is still orthonormal since

$$\int_D \phi_\omega(t)\overline{\phi_{\omega'}(t)}\, d\nu(t) = \frac{1}{N}\sum_{t=1}^N e^{-2\pi i(\omega-\omega')t/N} = \delta(\omega,\omega')$$

Finally, the Fast Fourier Transform (FFT) allows any subset of F's rows to be multiplied by a given vector in $O(N \log N)$ time.

Example 3. *Any unitary matrix $U \in \mathbb{C}^{N\times N}$ can be represented as a Bounded ONS with $\phi_k(l) := U_{l,k}$, and $\nu :=$ the discrete uniform measure on [N]. The only difference from above is that we should set $K = \max_{l,k} |\sqrt{N}\cdot U_{l,k}|$.*

Theorem 1 (Thm 12.31 from [1])**.** *Let $A \in \mathbb{C}^{m\times N}$ be a matrix formed by sampling m points, $\vec{t}_1, \cdots, \vec{t}_m \in D$ independently, w.r.t. ν for any given BOS $\Phi = \{\phi_1, \cdots, \phi_N\}$, and then setting $A_{l,k} := \phi_k(\vec{t}_l)$ for $l \in [m]$, $k \in [N]$, $l \in [m]$. If, for $\varepsilon \in (0,1)$ and $k \in [N]$, we have*

$$m \geq (CK/\varepsilon^2)\cdot k \cdot \ln^4 N,$$

then with probability at least $1 - N^{-\ln^3 N}$ the restricted isometry constant $\varepsilon_k(\tilde{A}) \leq \varepsilon$ for $\tilde{A} = \frac{1}{\sqrt{m}}A$. The constant $C > 0$ is universal (i.e. independent of $k, K, N, \varepsilon, \dots$).

1.2 Applications of Theorem 1

Application 1 Suppose that $f(\vec{t}) = \sum_{j=1}^{N} x_j \cdot \phi_j(\vec{t})$ for a BONS, $\Phi = \{\phi_1, \cdots, \phi_N\}$. We assume (or hope) that $\vec{x} =$ the coefficient vector is sparse, or compressible. That is, we hope that $\inf_{\|z\|_0 \leq k} \|\vec{x} - \vec{z}\|_1$ is small.

We can try to learn f by learning $\vec{x}$ as follows: We sample $\vec{t}_1, \cdots, \vec{t}_m$ from D according to ν, and then use $f(\vec{t}_1), \cdots, f(\vec{t}_m)$ to recover $\vec{x}$ (and therefore f).

We have

$$\begin{pmatrix} f(\vec{t}_1) \\ f(\vec{t}_2) \\ \vdots \\ f(\vec{t}_m) \end{pmatrix} = \begin{pmatrix} \phi_1(\vec{t}_1) & \phi_2(\vec{t}_1) & \cdots & \phi_N(\vec{t}_1) \\ \phi_1(\vec{t}_2) & \phi_2(\vec{t}_2) & \cdots & \phi_N(\vec{t}_2) \\ \vdots & \vdots & \vdots & \vdots \\ \phi_1(\vec{t}_m) & \phi_2(\vec{t}_m) & \cdots & \phi_N(\vec{t}_m) \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}.$$

That is, we have

$$\begin{pmatrix} f(\vec{t}_1) \\ f(\vec{t}_2) \\ \vdots \\ f(\vec{t}_m) \end{pmatrix} = A\vec{x}$$

where $A_{j,i} := \phi_i(\vec{t}_j)$. Theorem 1 says that this A has the RIP, so we can interpolate $f(\vec{t}_1), \cdots, f(\vec{t}_m)$ to learn f by

1 Taking $\vec{t}_1, \cdots, \vec{t}_m$ from D for $m \geq \frac{CK^2\cdot k\cdot \log^4 N}{\varepsilon^2}$;

2 Using BP to find the $\vec{z}$ with minimal l_1 norm subject to $A\vec{x} = A\vec{z}$ (Lecture 16 $\rightarrow$ this gives us a good result).

Example 4 (Chebyshev Polynomials of the first kind)**.** *They are defined by* $T_0(x) = 1$*;* $T_1(x) = x$*;* $T_2(x) = 2x^2 - 1$*;* $\cdots$*;* $T_{n+1} = 2xT_n(x) - T_n - 1(x)$*. It is also true that* $T_j(x) = \cos(j \cdot \arccos(x))$ *holds for all* j*.*

Here we have $D = [-1, 1]$*, and* $\nu(A) = \frac{1}{\pi} \int_A \frac{1}{\sqrt{1-x^2}}\, dx$*, for all* $A \subset [-1, 1]$*.*

Thus, Chebyshev polynomials provide a BONS with $\Phi := \{\sqrt{2}T_1(x), \sqrt{2}T_2(x), \cdots, \sqrt{2}T_n(x)\}$*.*

That is, we have $\phi_j(x) = \sqrt{2}\cos(j \cdot \arccos(x))$ *for all* $j \in [N]$*. It is now easy to see that* $K = \sqrt{2}$*.*

Since Chebyschev polynomials form a BONS, we can interpolate Chebyschev-sparse functions using a small number of function samples!

Application 2 Recall from lecture 16, Theorem 3, that RIP matrices $\Rightarrow$ J-L embedding matrices: If A has RIP, take $D = diag(\star, \cdots, \star)$ with random ± 1 on the diagonal, and then AD will serve as a J-L embedding. Note that AD will now be fast to multiply if A is formed using the columns of a DFT matrix. This leads to "fast JL-embedding" matrices. More on this next time...

References

[1] Simon Foucart, Holger Rauhut. A Mathematical Introduction to Compressive Sensing. *Birkhauser Basel*, 2013