MTH 995-003: Intro to CS and Big Data	Spring 2014
Lecture 7 – January 28, 2014	
Inst. Mark Iwen	*Scribe: Kishavan Bhola*

1 Overview

In this lecture, we continue the discussion of subspace approximation with $\tau = \infty$. We give a probabalistic method with fast-implementation guarantees that comes close to the optimal solution (see [2] for more details).

2 Problem

Given $P = \{\vec{x_1}, \ldots, \vec{x}_N\} \subset \mathbb{R}^N$, compute $\mathcal{A} \in \Pi_d(\mathbb{R}^N)$ such that

$$R_\infty(\mathcal{A}, P) \approx R_d(P) = \inf_{\widetilde{\mathcal{A}} \in \Pi_d(\mathbb{R}^N)} \left(\max_{j=1,\ldots,N} \|\vec{x_j} - \Pi_{\widetilde{\mathcal{A}}}\ \vec{x_j}\|_2 \right) \tag{1}$$

3 Solution

Lemma 1 (last time)**.** *If $\widetilde{Y}$ is a solution to SPD(2), then $\widetilde{Y}$ has $r \geq D-d$ orthonormal eigenvectors $\vec{v_1}, \ldots, \vec{v}_r$ and*

$$\sum_{l=1}^{r} \lambda_l \langle \vec{x}_1, \vec{v_l} \rangle^2 \leq R_d^2(P)$$

Lemma 2. *The eigenvalues of $\widetilde{Y}$, $\lambda_1, \ldots, \lambda_r \in (0, 1]$, can be partitioned into $D-d$ disjoint sets $I_1, \ldots, I_{D-d} \subset [r] := \{1, \ldots, r\}$ such that*

$$\sum_{l \in I_k} \lambda_l \geq \frac{1}{2} \quad \forall k \in [D-d]$$

Homework 2.4: Prove Lemma 2.

Remark We will use these $D-d$ groups of eigenvectors to construct a basis for $\mathcal{A}^\perp := S_{\mathcal{A}}^\perp$.

Definition Let $\vec{\Psi} \in \mathbb{R}^r$ be a random vector with independent identically distributed Bernoulli entries from $\{1, -1\}$ and, $\forall k \in [D-d]$, define $\vec{z_k} \in \mathbb{R}^D$ by

$$\vec{z_k} := \frac{\sum_{l \in I_k} \Psi_l \sqrt{\lambda_l} \vec{v_l}}{\sqrt{\sum_{l \in I_k} \lambda_l}}$$

Homework 2.5 Show that $\vec{z}_1, \ldots, \vec{z}_{D-d}$ defined above are orthonormal.

Theorem 1. *(Bernstein; see, e.g., Chapter 7 of [1]) Let $\vec{\Psi} \in \mathbb{R}^r$ be a random vector with i.i.d. Bernoulli entries from $\{1, -1\}$ (each selected with probability $1/2$). Then for any $\vec{e} \in \mathbb{R}^r$ and $\beta > 0$,*

$$\mathbb{P}[\langle \vec{\Psi}, \vec{e} \rangle^2 > \beta \|\vec{e}\|_2^2] \leq 2\exp(-\beta/2)$$

Defintion Denote the elements of each $I_k \subset [r]$ by $l_{k,1}, \ldots, l_{k,|I_k|}$. Now define $\vec{\Psi}|_{I_k} \in \mathbb{R}^{|I_k|}$ to be

$$(\vec{\Psi}|_{I_k})_h := \vec{\Psi}_{l_{k,h}}$$

Remark Here, $\vec{\Psi} \in \mathbb{R}^r$.

Definition Let $k \in [D-d]$ and $j \in [N]$. We define the **error vector** $\vec{e}_{k,j} \in \mathbb{R}^{|I_k|}$ to be the vector with, for each $\bar{l} \in [|I_k|]$,

$$(\vec{e}_{k,j})_{\bar{l}} := \sqrt{\lambda_{l_{k,\bar{l}}}} \langle \vec{x}_j,\ \vec{v}_{l_{k,\bar{l}}} \rangle$$

Remark $\|\vec{e}_{k,j}\|_2^2 = \sum_{l \in I_k} \lambda_l \langle \vec{x}_1, \vec{v}_l \rangle^2$ (note the similarity of this sum to the one in Lemma 1).

Lemma 3. *Choose $k \in [D-d]$ and $\vec{x}_j \in P$. Then*

$$\mathbb{P}\left[\langle \vec{x}_j, \vec{z}_k \rangle^2 > 12\ln(N)\|\vec{e}_{k,j}\|_2^2\right] \leq \frac{2}{N^3}$$

Proof: Lemma 2 tells us that

$$\sum_{l \in I_k} \lambda_l \geq \frac{1}{2} \quad \forall k \in [D-d]$$

Thus,

$$\begin{aligned}
\langle \vec{x}_j, \vec{z}_k \rangle^2 &= \left\langle \vec{x}_j, \frac{\sum_{l \in I_k} \Psi_l \sqrt{\lambda_l} \vec{v}_l}{\sqrt{\sum_{l \in I_k} \lambda_l}} \right\rangle^2 && \text{(by definition)} \\
&\leq 2 \left\langle \vec{x}_j, \sum_{l \in I_k} \Psi_l \sqrt{\lambda_l} \vec{v}_l \right\rangle^2 && \text{(using Lemma 2)} \\
&= 2 \left(\sum_{l \in I_k} \Psi_l \sqrt{\lambda_l} \langle \vec{x}_j, \vec{v}_l \rangle \right)^2 && \text{(linearity of the inner product)} \\
&= 2 \langle \vec{\Psi}|_{I_k}, \vec{e}_{k,j} \rangle^2 && \text{(by definitions)}
\end{aligned}$$

Thus,

$$
\begin{aligned}
\mathbb{P}\left[\langle \vec{x}_j, \vec{z}_k\rangle^2 > 12\ln(N)\|\vec{e}_{k,j}\|_2^2\right] & \\
&\le \mathbb{P}\left[\langle \vec{\Psi}|_{I_k}, \vec{e}_{k,j}\rangle^2 > 6\ln(N)\|\vec{e}_{k,j}\|_2^2\right] \\
&\le 2\exp(-3\ln(N)) && \text{(using Theorem 1)} \\
&= \frac{2}{N^3}
\end{aligned}
$$

$\square$

Remark The lemma above plus Lemma 1 shows this random combination gives us something close to what the best error should be with high probability (w.h.p.).

Theorem 2. *Let $P \subset \mathbb{R}^D$ have $|P| = N$. We can compute in D, N-polynomial time an affine subspace $\mathcal{A}$ of dimension d such that*

$$R_\infty(\mathcal{A}, P) \le 2\sqrt{12 \cdot \ln(2N)}R_d(P)$$

with probability at least $1 - \frac{8}{N}$.

Proof: First, let $\bar{P}$ be the symmetrization of P. Note that $|\bar{P}| \le 2N$. Moreover, if $D > 2N$, we can reduce D by working in $\text{Span}(\bar{P})$.

Now define the statement E as follows:

$$E : \exists \vec{x}_j \in \bar{P}, \exists k_j \in [D-d] \text{ such that } \langle \vec{x}_j, \vec{z}_k\rangle^2 > 12\ln(2N)\|\vec{e}_{k,j}\|^2$$

Let us bound the probability that E is true:

$$
\begin{aligned}
\mathbb{P}[E] &\le \sum_{j\in[|\bar{P}|], k\in[D-d]} \mathbb{P}\left[\langle \vec{x}_j, \vec{z}_k\rangle^2 > 12\ln(2N)\|\vec{e}_{k,j}\|^2\right] && \text{(by the union bound)} \\
&\le 2N(D-d)\frac{2}{N^3} && \text{(using Lemma 4)} \\
&\le \frac{4(D-d)}{N^2} \\
&\le \frac{8}{N} && \text{(since } D \le 2N \text{)}
\end{aligned}
$$

Now, if E is false, then we have for each $\vec{x}_j \in \bar{P}$,

$$
\begin{aligned}
\|\Pi_{\text{Span}(\vec{z}_1,\ldots,\vec{z}_k)}\vec{x}_j\|^2 \equiv \sum_{k=1}^{D-d} \langle \vec{x}_j, \vec{z}_k\rangle^2 & \\
&\le 12\ln(2N)\sum_{k=1}^{D-d}\|\vec{e}_{k,j}\|^2 && \text{(since } E \text{ is false)} \\
&\le 12\ln(2N)R_d^2(\bar{P}) && \text{(by Lemma 1)}
\end{aligned}
$$

Now, taking square roots and, via homework problem 1 from lecture 6, we get the desired error bound. We should choose $\mathcal{A}$ to be,

$$\mathcal{A} := (\text{Span}(\vec{z}_1, \ldots, \vec{z}_{D-d}))^{\perp} + \vec{x}_1$$

And this is computable in polynomial time since we get it from $\widetilde{Y}$ (the solution to an SDP). $\square$

References

[1] Simon Foucart, Holger Rauhut. *A Mathematical Introduction to Compressed Sensing.* Springer, 2013.

[2] K. Varadarajan, S. Venkatesh, Y. Ye, and J. Zhang. *Approximating the Radii of Point Sets.* SIAM J. Comput., Vol. 36, No. 6, pp. 1764-1776.