University of Edinburgh
INFR11156: Algorithmic Foundations of Data Science (2019)
Solutions of the Coursework

Student ID Number: s1764997

Solutions for Part A

Prove that

$$\frac{x_k^T B x_k}{x_k^T x_k} \geq (1-\epsilon)\frac{\lambda_1}{1+4n(1-\epsilon)^{2k}} \tag{1}$$

Lemma 0:
Let Y be a random variable generated according to a χ^2 distribution with n degrees of freedom. Then it holds that $P[Y > n] < \frac{1}{2}$

Lemma 1:
Let $v \in R^n$ such that $||v|| = 1$. Let x be a Gaussian vector. That is, it is a random vector $x \in R^n$ where every x_i is generated independently according to $N(0,1)$ where $N(0,1)$ is a normal distribution with 0 mean and variance equal to 1. Then it holds that:

$$P(| < x, v > | \geq \frac{1}{2}) \geq \frac{3}{16} \tag{2}$$

Lemma 2:
Let $x \in R^n$ be a vector such that $| < x, v_1 > | \geq \frac{1}{2}$ with a constant probability. Then, for every positive integer k and $\epsilon > 0$, it holds that:

$$\frac{x_k^T B x_k}{x_k^T x_k} \geq (1-\epsilon)\frac{\lambda_1}{1+4||x||^2(1-\epsilon)^{2k}} \tag{3}$$

where x_k is the output vector of Algorithm 1 as specified in the coursework handout.

By Lemma 1, with constant probability, a Gaussian vector x where each x_i is sampled randomly and independently according to $N(0,1)$, we have $| < x, v > | \geq \frac{1}{2}$ for any $||v|| = 1$.
Note that $||x||^2$ in Lemma 2 can be written as

$$||x||^2 = \sum_{i=1}^{n} x_i^2 \tag{4}$$

and thus $||x||^2$ is a random variable distributed according to χ^2 distribution with n degrees of freedom where n defines the number of elements in x. This is due to the already mentioned fact that each x_i is distributed according to $N(0,1)$ randomly and independently.
Using Lemma 0 we have that

$$P(||x^2|| > n) < \frac{1}{2} \tag{5}$$

and thus

$$P(||x^2|| \leq n) \geq \frac{1}{2} \tag{6}$$

Conditioning on this and using Lemma 2, we have that:

$$\frac{x_k^T B x_k}{x_k^T x_k} \geq (1-\epsilon)\frac{\lambda_1}{1+4||x^2||(1-\epsilon)^{2k}} \tag{7}$$

where $||x^2|| \leq n$ with a constant probability which is greater than or equal to one-half. Substituting this into equation above, we obtain that for any $\epsilon > 0$ and any positive integer k:

$$\frac{x_k^T B x_k}{x_k^T x_k} \geq (1-\epsilon)\frac{\lambda_1}{1+4n(1-\epsilon)^{2k}} \tag{8}$$

with a constant probability. More specifically, using Lemma 1 and equation (6) above, this probability $Prob$ is $Prob \geq \frac{3}{32}$

Proof of Lemma 1:
Similar to the proof from a tutorial, we define a random variable S which is

$$S =< x, v >= \sum_{i=1}^{n} x_i v_i \tag{9}$$

We can observe that $E(x_i) = 0$ and $Var(x_i) = 1$ for every $1 \leq i \leq n$. Because S is a linear combination of x_i, and thus a linear combination of normal distributions, it is also normally distributed (from courses PwA and MLPR). Hence, we have that $E(S) = 0$ and $Var(S) = 1$. This can be used to show that:

$$E(S^2) = E^2(S) + Var(S) = 0^2 + 1 = 1 \tag{10}$$

Now, we need to show what is $E(S^4)$.

$$\begin{aligned}
E(S^4) &= E[(\sum_{i=1}^{n} x_i v_i)^4] \\
&= E[\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} x_i x_j x_k x_l v_i v_j v_k v_l] \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} E[x_i x_j x_k x_l] v_i v_j v_k v_l \\
&= \sum_{i=1}^{n} E(x_i^4) v_i^4 + \frac{1}{2}\binom{4}{2} \sum_{i \neq j}^{n} E(x_i^2 x_j^2) v_i^2 v_j^2
\end{aligned} \tag{11}$$

Transition from the second into the third line comes from the linearity of expectation.
The way we got to line 4 is the following: as already stated, $E(S) = 0$ and thus any expression containing this term will be zero as well. Thus, all terms with S raised to an odd power will disappear.
There are $\binom{4}{2}$ ways to select tuples of 2 fixed indices from 4 loops. Because (x_i, x_j) is the same as (x_j, x_i), and these would be accounted for twice in the expectation term, we need to divide their count by 2.

According to Wikipedia and mathematical forums I have come across, the expectation of X to the power of p, where X has a normal distribution, can be computed as follows:

$$\mathrm{E}\left[X^p\right] = \begin{cases} 0 & \text{if } p \text{ is odd,} \\ \sigma^p (p-1)!! & \text{if } p \text{ is even.} \end{cases} \tag{12}$$

where σ is the standard deviation of the distribution. In our case, x_i is distributed according to the standard normal distribution, so $\sigma = \sqrt{Var(x_i)} = \sqrt{1} = 1$. Therefore, for all x_i:

$$E(x_i^4) = 3 \tag{13}$$

Substituting this back into the original equation, we obtain:

$$\begin{aligned}
E(S^4) &= 3\sum_{i=1}^{n} v_i^4 + 3\sum_{i \neq j}^{n} E(x_i^2)E(x_j^2) v_i^2 v_j^2 \\
&= 3\sum_{i=1}^{n} v_i^4 + 3(\sum_{i=1}^{n} E(x_i^2) v_i^2)(\sum_{j=1}^{n} E(x_j^2) v_j^2) - 2\sum_{i=1}^{n} E(x_i^4) v_i^4 \\
&= 3\sum_{i=1}^{n} v_i^4 + 3(\sum_{i=1}^{n} v_i^2)(\sum_{j=1}^{n} v_j^2) - 6\sum_{i=1}^{n} v_i^4 \\
&= 3||v||^4 - 3\sum_{i=1}^{n} v_i^4 \\
&\leq 3
\end{aligned} \tag{14}$$

The first line is using the fact that $E(x_i^2 x_j^2) = E(x_i^2)E(x_j^2) = 1$ as x_i and x_j are independent. In the second line, we are expanding the term with double sum with $i \neq j$. We need to subtract $2\sum_{i=1}^n E(x_i^4)v_i^4$ as this was now created where $i = j$. The third line uses the already mentioned fact that $E(x_i^2) = 1$ and $E(x_i^4) = 3$. In the fourth line we use the definition of the squared norm of v as $\sum_{i=1}^n v_i^2 = ||v||^2$ and replace the double sum with $||v||^4$. Finally, we use the fact that v is a unit vector, so $||v|| = 1$ which implies that $||v||^4 = 1$. v_i^4 cannot be negative, so we end up with stated inequality.

The rest of the proof is exactly the same as in the notes. We are using the Paley-Zygmund inequality:

$$P(Z \geq \delta E(Z)) \geq (1-\delta)^2 \frac{E^2(Z)}{E(Z^2)} \tag{15}$$

where Z is a non-negative random variable, in our case $Z = S^2$. Using $\delta = 1/4$, we obtain:

$$P(S^2 \geq \delta E(S^2)) = P(S^2 \geq \frac{1}{4}) \geq \frac{3^2}{4^2}\frac{1}{3} = \frac{3}{16} \tag{16}$$

We have proven Lemma 1.

Our Lemma 2 is exactly the same as Lemma 3 in Lecture Notes 6. The only difference is that we used slightly different notation. As Lemma 3 in the notes states $y = B^k x$, which is in our case notated as x_k. After setting $x_k = y$, the proof is exactly the same as in the notes and therefore not repeated here. Please note that this is due to the fact that this lemma does not specify vector x, only uses it in the way that was proven by our 1st lemma.

THIS IS THE END OF MY SOLUTIONS.