# Supplementary Material: A Single-Pass Algorithm for Efficiently Recovering Sparse Cluster Centers of High-dimensional Data

**Jinfeng Yi** JINFENGY@US.IBM.COM

IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA

**Lijun Zhang** ZHANGLJ@LAMDA.NJU.EDU.CN

National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China

**Jun Wang** WANGJUN@US.IBM.COM

IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA

**Rong Jin** RONGJIN@CSE.MSU.EDU

**Anil K. Jain** JAIN@CSE.MSU.EDU

Department of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824 USA

**Theorem 1.** *Let* $1{=}(6m)$ *be a parameter to control the success probability. Assume*

$$
{}_{*} \qquad {}^{1} \qquad {}_{\max}; \tag{1}
$$

$$
\frac{{}^{1}}{2\,p\overline{\overline{2s}}} \qquad {}^{1} \qquad c\frac{{}^{1}}{2\,p\overline{\overline{2s}}}; \tag{2}
$$

$$
T \quad \max\left(\frac{18}{0}\ln\frac{2K}{};\ \frac{3c_2\ \ 0}{1};\ \left(\frac{6c_3}{1}\right)^2(\ln n + \ln d)\right) \tag{3}
$$

*where* $c$, $c_2$ *and* $c_3$ *are some universal constants. Then, with a probability at least* $1 \quad 6m$ *, we have*

$$
{}^{m+1} = \max_{1\le i\le K} k\widehat{c}_i^{m+1} \quad c_i k \quad \max\left( \quad {}_*;\ \frac{c\ \ {}^{1}}{p\overline{\overline{2^m}}}\right):
$$

**Corollary 1.** *The convergence rate for* , *the maximum difference between the optimal cluster centers and the estimated ones, is* $O(\sqrt{(s\log d){=}n})$ *before reaching the optimal difference* ${}_*$.

## 1. Proof of Corollary 1

According to the assumption of ${}^{1}$ in (2), we know that $\frac{1}{\lambda^1} \ / \ \frac{\sqrt{s}}{1}$. Since the value of $T$ is dominated by the last term in the right side of (3), we have $T \ / \ \frac{s\log d}{1.\ \ 1}$, which implies

$$
n \ / \ 2^m T \ / \ 2^m \frac{s\log d}{1 \quad 1}:
$$

Combining with the conclusion $_{m+1} / \frac{1}{\sqrt{2^m}}$, we have

$$_{m+1} / \sqrt{\frac{s\log d}{n}}:$$

**Lemma 1.** *Let* $^t$ *be the maximum difference between the optimal cluster centers and the ones estimated from iteration* $t$*, and* $2$ $(0;1)$ *be the failure probability. Assume*

$$^t \quad \frac{1}{2} \qquad \sqrt{5\ln(3K)}\ , \quad _{\max}; \tag{4}$$

$$jS^tj \quad \frac{18}{_0}\ln\frac{2K}{}; \tag{5}$$

$$^t \quad c_1\exp\left(\frac{(1 \quad 2 \quad ^t \quad)^2}{8(1+ \quad ^t)^2 \quad ^2}\right)( \ _0 + \sqrt{\ln jS^tj}) + \frac{c_2 \ _0}{jS^tj} + c_3 \ \frac{\sqrt{\ln jS^tj} + {}^p\overline{\ln d}}{\sqrt{jS^tj}}; \tag{6}$$

*for some constants* $c_1$*,* $c_2$ *and* $c_3$*. Then with a probability* $1 \quad 6$ *, we have*

$$^{t+1} \quad 2^p\overline{s} \ ^t:$$

## 2. Proof of Lemma 1

For the simplicity of analysis, we will drop the superscript $t$ through this analysis.

### 2.1. Preliminaries

We denote by $C_k$ the support of $\mathbf{c}_k$ and $\overline{C}_k = [d]\, n\, C_k$. For any vector $\mathbf{z}$, $\mathbf{z}(C)$ is defined as $[\mathbf{z}(C)]_i = z_i$ if $i \ 2 \ C$ and zero, otherwise.

For any $\mathbf{x}_i \ 2 \ S$, we use $k_i$ to denote the index of the true cluster, and $\widehat{k}_i$ to denote index of the cluster assigned by the nearest neighbor search, i.e.,

$$\begin{aligned} \mathbf{x}_i &= \mathbf{c}_{k_i} + \mathbf{g}_i \text{ and } \mathbf{g}_i \quad N(0; \ ^2 I); \\ \widehat{k}_i &= \arg\max_{j\in[K]} \widehat{\mathbf{c}}_j^\top \mathbf{x}_i: \end{aligned}$$

Then, we can partition data points in $S$ based on either the ground truth or the assigned cluster. Let $S_k$ be the subset of data points in $S$ that belong to the $k$-th cluster, i.e.,

$$S_k = f\mathbf{x}_i \ 2 \ S : \mathbf{x}_i = \mathbf{c}_k + \mathbf{g}_i \text{ and } \mathbf{g}_i \quad N(0; \ ^2 I)g \tag{7}$$

Let $\widehat{S}_k$ be the subset of data points that are assigned to the $k$-th cluster based on the nearest neighbor search, i.e.,

$$\widehat{S}_k = f\mathbf{x}_i \ 2 \ S : k = \arg\max_{j\in[K]} \widehat{\mathbf{c}}_j^\top \mathbf{x}_i g \tag{8}$$

### 2.2. The Main Analysis

Let $L_k(\mathbf{c})$ be the objective function in Step 11 of Algorithm 1. We expand $L_k(\mathbf{c})$ as

$$\begin{aligned} & L_k(\mathbf{c}) \\ = \ & k\mathbf{c}k_1 + k\mathbf{c} \quad \mathbf{c}_k k^2 + \frac{1}{j\widehat{S}_k j}\sum_{\mathbf{x}_i\in\widehat{S}_k} k\mathbf{x}_i \quad \mathbf{c}_k k^2 \quad \frac{2}{j\widehat{S}_k j}\sum_{\mathbf{x}_i\in\widehat{S}_k}(\mathbf{c} \quad \mathbf{c}_k)^\top(\mathbf{x}_i \quad \mathbf{c}_k) \\ = \ & k\mathbf{c}k_1 + k\mathbf{c} \quad \mathbf{c}_k k^2 + \frac{1}{j\widehat{S}_k j}\sum_{\mathbf{x}_i\in\widehat{S}_k} k\mathbf{x}_i \quad \mathbf{c}_k k^2 \\ & \quad 2(\mathbf{c} \quad \mathbf{c}_k)^\top \underbrace{\frac{1}{j\widehat{S}_k j}\sum_{\mathbf{x}_i\in\widehat{S}_k\setminus S_k}(\mathbf{c}_{k_i} \quad \mathbf{c}_k)}_{A_k} \ 2(\mathbf{c} \quad \mathbf{c}_k)^\top \underbrace{\frac{1}{j\widehat{S}_k j}\sum_{\mathbf{x}_i\in\widehat{S}_k}\mathbf{g}_i}_{B_k}: \end{aligned} \tag{9}$$

Let $c_k^*$ be the optimal solution that minimizes $L_k(c)$, and define $f_k = c_k^* \quad c_k$. We have

$$\begin{aligned} & L_k(c_k^*) \quad L_k(c_k) \\ = & \; kf_k + c_kk_1 + kf_kk^2 \quad 2f_k^\top A_k \quad 2f_k^\top B_k \quad kc_kk_1 \\ & kc_kk_1 \quad kf_k(C_k)k_1 + \quad kf_k(\bar{C}_k)k_1 + kf_kk^2 \quad 2f_k^\top A_k \quad 2f_k^\top B_k \quad kc_kk_1 \\ & kf_k(C_k)k_1 + \quad kf_k(\bar{C}_k)k_1 + kf_kk^2 \quad 2kf_kk_1kA_kk_\infty \quad 2kf_kk_1kB_kk_\infty \\ = & \; ( \quad + 2kA_kk_\infty + 2kB_kk_\infty)kf_k(C_k)k_1 + ( \quad 2kA_kk_\infty \quad 2kB_kk_\infty)kf_k(\bar{C}_k)k_1 + kf_kk^2 \\ & \sqrt{jC_kj}( \quad + 2kA_kk_\infty + 2kB_kk_\infty)kf_k(C_k)k + ( \quad 2kA_kk_\infty \quad 2kB_kk_\infty)kf_k(\bar{C}_k)k_1 + kf_kk^2: \end{aligned}$$

Thus, if

$$2kA_kk_\infty + 2kB_kk_\infty;$$

we have

$$kf_k(C_k)k^2 \quad kf_kk^2 \quad ( \quad + 2kA_kk_\infty + 2kB_kk_\infty)\sqrt{jC_kj}kf_k(C_k)k \quad 2 \quad \sqrt{jC_kj}kf_k(C_k)k \;\; ) \;\; kf_k(C_k)k \quad 2 \quad \sqrt{jC_kj};$$

and thus

$$kf_kk^2 \quad 2 \quad \sqrt{jC_kj}kf_k(C_k)k \quad 4 \; {}^2jC_kj \;\; ) \;\; kf_kk \quad 2 \quad \sqrt{jC_kj}:$$

In summary, if

$$2kA_kk_\infty + 2kB_kk_\infty; 8k \; 2 \; [K]$$

we have

$$\max_{1\le k\le K} kc_k^* \quad c_kk \quad 2^{p}\overline{\ \ }\!s \;:$$

In the following, we discuss how to bound $kA_kk_\infty$ and $kB_kk_\infty$.

### 2.3. Bound for $kA_kk_\infty$

From the definition of $A_k$ in (9), we have

$$kA_kk_\infty \quad 2 \;\; {}_0\frac{j\hat{S}_k \; n \; S_kj}{j\hat{S}_kj}:$$

#### 2.3.1. LOWER BOUND OF $j\hat{S}_kj$

First, we show that the size of $S_k$ is lower-bounded, which means a significant amount of data points in $S$ belong to the $k$-th cluster. Recall that $_1; : : : ; \; _K$ are the weight of the Gaussian mixtures, and $_0 = \min_{1\le i\le K} \; _i$. According to the Chernoff bound (Angluin & Valiant, 1979) provided in Appendix A, we have, with a probability at least 1

$$jS_kj \quad {}_kjSj\left(1 \quad \sqrt{\frac{2}{{}_kjSj}\ln\frac{K}{\ }}\right)^{(5)} \frac{2}{3} \; {}_kjSj; \; 8k \; 2 \; [K]: \tag{10}$$

Next, we prove that a larger amount of data points in $S_k$ belong to $\hat{S}_k$. We begin by analyzing the probability that the assigned cluster $\hat{k}_i$ of $x_i$ is the true cluster $k_i$. The similarity between $x_i$ and the estimated cluster centers can be bounded by

$$\begin{aligned} \hat{c}_{k_i}^\top x_i = & \hat{c}_{k_i}^\top(c_{k_i} + g_i) = kc_{k_i}k^2 + [\hat{c}_{k_i} \quad c_{k_i}]^\top c_{k_i} + \hat{c}_{k_i}^\top g_i \\ & 1 \quad k\hat{c}_{k_i} \quad c_{k_i}k \quad j\hat{c}_{k_i}^\top g_ij \quad 1 \qquad (1 + \quad)\left|g_i^\top\frac{\hat{c}_{k_i}}{k\hat{c}_{k_i}k}\right|; \\ \hat{c}_j^\top x_i = & \hat{c}_j^\top(c_{k_i} + g_i) = c_j^\top c_{k_i} + [\hat{c}_j \quad c_j]^\top c_{k_i} + \hat{c}_j^\top g_i \\ & + k\hat{c}_j \quad c_jk + j\hat{c}_j^\top g_ij \quad + \quad + (1 + \quad)\left|g_i^\top\frac{\hat{c}_j}{k\hat{c}_jk}\right|; \; j \neq k_i: \end{aligned}$$

Hence, $x_i$ will be assigned to cluster $k_i$ if

$$1 \qquad (1+\quad)\left|g_i^\top \frac{\widehat{c}_{k_i}}{k\widehat{c}_{k_i}k}\right| \qquad + \quad + (1+\quad)\left|g_i^\top \frac{\widehat{c}_j}{k\widehat{c}_j k}\right|; 8j \neq k_i;$$

which leads to the following sufficient condition

$$\max_{1\le j\le K}\left|g_i^\top \frac{\widehat{c}_j}{k\widehat{c}_j k}\right| \quad \frac{1 \quad 2}{2(1+\quad)} \;,\; g_0 \overset{(4)}{\ } \frac{2\ \sqrt{5\ln(3K)}}{3} \qquad \sqrt{2\ln(3K)}: \tag{11}$$

It is easy to verify that for any fixed direction $\widehat{c}$ with $k\widehat{c}k = 1$, $g_i^\top c$ is a Gaussian random variable with mean 0 and variance $^2$. Based on the tail bound for the Gaussian distribution (Chang et al., 2011) provided in Appendix B, we have

$$\Pr\left[\max_{1\le j\le K}\left|g_i^\top \frac{\widehat{c}_j}{k\widehat{c}_j k}\right| \quad g_0\right] \quad 1 \quad K\exp\left(\quad \frac{g_0^2}{2\ ^2}\right):$$

Define

$$= K\exp\left(\quad \frac{g_0^2}{2\ ^2}\right) \overset{(11)}{\ } \frac{1}{3}: \tag{12}$$

In summary, we have proved the following lemma.

**Lemma 2.** *Under the condition in (4), with a probability at least* 1 *,* $x_i = c_{k_i} + g_i \ 2\ S_{k_i} \quad S$ *satisfies*

$$\max_{1\le j\le K}\left|g_i^\top \frac{\widehat{c}_j}{k\widehat{c}_j k}\right| \quad g_0;$$

*and is assigned to the correct cluster* $k_i$ *based on the nearest neighbor search (i.e.,* $\widehat{k}_i = k_i$*).*

Define

$$S_k^1 = \left\{x_i\ 2\ S_k : \max_{1\le j\le K}\left|g_i^\top \frac{\widehat{c}_j}{k\widehat{c}_j k}\right| \quad g_0\right\} \quad \widehat{S}_k \backslash S_k: \tag{13}$$

Since each data point in $S_k$ has a probability at least 1 to be assigned to set $S_k^1$, using the Chernoff bound again, we have, with a probability at least 1 ,

$$\begin{aligned} j\widehat{S}_kj \quad j\widehat{S}_k \backslash S_kj \quad jS_k^1j \quad & \mathrm{E}\left[jS_k^1j\right]\left(1 \quad \sqrt{\frac{2}{\mathrm{E}\left[jS_k^1j\right]}\ln\frac{K}{\ }}\right) \\ & (1 \quad )jS_kj\left(1 \quad \sqrt{\frac{2}{(1 \quad )jS_kj}\ln\frac{K}{\ }}\right) \\ \overset{(12)}{\ } \quad & \frac{2}{3}jS_kj\left(1 \quad \sqrt{\frac{3}{jS_kj}\ln\frac{K}{\ }}\right) \overset{(5),\,(10)}{\ } \frac{1}{3}jS_kj; 8k\ 2\ [K]: \end{aligned} \tag{14}$$

### 2.3.2. Upper bound of $j\widehat{S}_k\ n\ S_kj$

Define

$$O = [_{k=1}^K S_k^1 \quad S \text{ and } \overline{O} = [_{k=1}^K\left(\widehat{S}_k\ n\ S_k^1\right) = S\ n\ O \quad S:$$

From Lemma 2, we know that with a probability at least 1 , each $x_i\ 2\ S_k$ belongs to the set $S_k^1 \quad O$. Thus, with probability at least 1 , each $x_i\ 2\ S$ belongs to $O$. In other words, with probability *at most* , each $x_i\ 2\ S$ belongs to $\overline{O}$. Based on the Chernoff bound, we have, with a probability at least 1 ,

$$j\overline{O}j \quad 2\mathrm{E}\left[j\overline{O}j\right] + 2\ln\frac{1}{\ } \quad 2\ jSj + 2\ln\frac{1}{\ }: \tag{15}$$

Since $S_k^1 \quad S_k$, we have $\widehat{S}_k\ n\ S_k \quad \widehat{S}_k\ n\ S_k^1 \quad \overline{O}$. Therefore, with a probability at least 1 , we have

$$j\widehat{S}_k\ n\ S_kj \quad 2\ jSj + 2\ln\frac{1}{\ }; 8k\ 2\ [K]: \tag{16}$$

Combining (10), (14) and (16), we have, with probability at least 1 3

$$kA_kk_\infty \quad 2\ _0\frac{2\ jSj + 2\ln\frac{1}{\epsilon}}{\frac{2}{9}\ _kjSj} = \frac{18\ _0}{_k}\left(\ + \frac{1}{jSj}\ln\frac{1}{\ }\right) = O(\ _0) + O\left(\frac{_0}{jSj}\right); 8k\ 2\ [K]: \tag{17}$$

### 2.4. Bound for $kB_kk_\infty$

Notice that $f\mathbf{g}_i : \mathrm{x}_i\ 2\ \widehat{S}_kg$, determined by the estimated centers $\widehat{\mathbf{c}}_1; : : : ; \widehat{\mathbf{c}}_K$, is a specific subset of $f\mathbf{g}_i : \mathrm{x}_i\ 2\ Sg$. Although $\mathbf{g}_i$ is drawn from the Gaussian distribution $N(0;\ ^2I)$, the distribution of elements in $f\mathbf{g}_i : \mathrm{x}_i\ 2\ \widehat{S}_kg$ is unknown. As a result, we cannot direct apply concentration inequality of Gaussian random vectors to bound $kB_kk_\infty$. Let $U_1\ 2\ \mathbb{R}^{d\times K}$ be a matrix whose columns are basis vectors of the subspace spanned by $\widehat{\mathbf{c}}_1; : : : ; \widehat{\mathbf{c}}_K$, and $U_2\ 2\ \mathbb{R}^{d\times(d-K)}$ be a matrix whose columns are basis vectors of the complementary subspace. We then divide each $\mathbf{g}_i$ as

$$\mathbf{g}_i = \mathbf{g}_i^{\|} + \mathbf{g}_i^{\perp};$$

where $\mathbf{g}_i^{\|} = U_1U_1^\top\mathbf{g}_i$, and $\mathbf{g}_i^{\perp} = U_2U_2^\top\mathbf{g}_i$.

First, we upper bound $kB_kk_\infty$ as

$$kB_kk_\infty \quad \underbrace{\left\|\frac{1}{j\widehat{S}_kj}\sum_{\mathbf{x}_i\in\widehat{\mathcal{S}}_k}\mathbf{g}_i^{\perp}\right\|_\infty}_{\widehat{B}_k^1} + \underbrace{\frac{j\widehat{S}_knS_k^1j}{j\widehat{S}_kj}\left\|\frac{1}{j\widehat{S}_knS_k^1j}\sum_{\mathbf{x}_i\in\widehat{\mathcal{S}}_k\setminus\mathcal{S}_k^1}\mathbf{g}_i^{\|}\right\|_\infty}_{\widehat{B}_k^2} + \underbrace{\frac{jS_k^1j}{j\widehat{S}_kj}\left\|\frac{1}{jS_k^1j}\sum_{\mathbf{x}_i\in\mathcal{S}_k^1}\mathbf{g}_i^{\|}\right\|_\infty}_{\widehat{B}_k^3}: \tag{18}$$

In the following, we discuss how to bound each term in the right hand side of (18).

#### 2.4.1. Upper bound of $\widehat{B}_k^1$

Following the property of Gaussian random vector, $\sum_{\mathbf{x}_i\in\widehat{\mathcal{S}}_k}U_2^\top\mathbf{g}_i=\left(\ \sqrt{j\widehat{S}_kj}\right)$ can be treated as a $(d\ \ K)$-dimensional Gaussian random vector. As a result, each element of $U_2\sum_{\mathbf{x}_i\in\widehat{\mathcal{S}}_k}U_2^\top\mathbf{g}_i=\left(\ \sqrt{j\widehat{S}_kj}\right)$ is a Gaussian random variable with variance smaller than 1. Based on the tail bound for the Gaussian distribution (Chang et al., 2011) provided in Appendix B and the union bound, with a probability at least 1 , we have

$$\left\|\sum_{\mathbf{x}_i\in\widehat{\mathcal{S}}_k}\mathbf{g}_i^{\perp}=\left(\ \sqrt{j\widehat{S}_kj}\right)\right\|_\infty = \left\|U_2\sum_{\mathbf{x}_i\in\widehat{\mathcal{S}}_k}U_2^\top\mathbf{g}_i=\left(\ \sqrt{j\widehat{S}_kj}\right)\right\|_\infty \quad \sqrt{2\ln\frac{Kd}{\ }}; 8k\ 2\ [K];$$

which implies

$$\widehat{B}_k^1 \quad \sqrt{\frac{2\ln\frac{Kd}{\epsilon}}{j\widehat{S}_kj}} \overset{(10),\,(14)}{\ } \sqrt{\frac{2\ln\frac{Kd}{\epsilon}}{2\ \ _kjSj=9}} = O\left(\ \sqrt{\frac{\ln d}{jSj}}\right); 8k\ 2\ [K]: \tag{19}$$

#### 2.4.2. Upper bound of $\widehat{B}_k^2$

First, we have

$$\left\|\frac{1}{j\widehat{S}_knS_k^1j}\sum_{\mathbf{x}_i\in\widehat{\mathcal{S}}_k\setminus\mathcal{S}_k^1}\mathbf{g}_i^{\|}\right\|_\infty = \left\|\frac{1}{j\widehat{S}_knS_k^1j}\sum_{\mathbf{x}_i\in\widehat{\mathcal{S}}_k\setminus\mathcal{S}_k^1}U_1U_1^\top\mathbf{g}_i\right\|_\infty \quad \left\|\frac{1}{j\widehat{S}_knS_k^1j}\sum_{\mathbf{x}_i\in\widehat{\mathcal{S}}_k\setminus\mathcal{S}_k^1}U_1^\top\mathbf{g}_i\right\| \tag{20}$$

Since $U_1^\top\mathbf{g}_i=$ can be treated as a $K$-dimensional Gaussian random vector, based on the tail bound for the $^2$ distribution (Laurent & Massart, 2000), we have with a probability at least 1 ,

$$kU_1^\top\mathbf{g}_ik \quad \left(p\overline{K} + \sqrt{2\log\frac{1}{\ }}\right)$$

Applying the union bound again, with a probability at least 1 , we have

$$\max_{1\le i\le|\mathcal{S}|} \left\|U_1^\top \mathbf{g}_i\right\| \qquad \left( {}^{p}\overline{K} + \sqrt{2\log\frac{jSj}{}} \right) \tag{21}$$

Combining (20) and (21), we have

$$\widehat{B}_k^2 \quad \frac{9}{\ _k} \left( \ + \frac{1}{jSj}\ln\frac{1}{\ } \right) \left( {}^{p}\overline{K} + \sqrt{2\log\frac{jSj}{}} \right) = O( \quad \sqrt{\ln jSj}) + O\left( \frac{\sqrt{\ln jSj}}{jSj} \right); 8k\ 2\ [K]: \tag{22}$$

### 2.4.3. Upper bound of $\widehat{B}_k^3$

First, we have

$$\left\| \frac{1}{jS_k^1 j} \sum_{\mathbf{x}_i\in\mathcal{S}_k^1} \mathbf{g}_i^{\|} \right\|_\infty = \left\| U_1 \frac{1}{jS_k^1 j} \sum_{\mathbf{x}_i\in\mathcal{S}_k^1} U_1^\top \mathbf{g}_i \right\|_\infty \quad \left\| \frac{1}{jS_k^1 j} \sum_{\mathbf{x}_i\in\mathcal{S}_k^1} U_1^\top \mathbf{g}_i \right\| := u_k \tag{23}$$

Recall the definition of $S_k^1$ in (13). Due to the fact that the domain is symmetric, we have $\mathrm{E}\left[U_1^\top \mathbf{g}_i\right] = 0$. Under the condition in (21), we can invoke the following lemma to bound $u_k$.

**Lemma 3.** *(Lemma 2 from (Smale & Zhou, 2007)) Let* $H$ *be a Hilbert space and be a random variable on* $(Z; \ )$ *with values in* $H$*. Assume* $k\ k \quad M < 1$ *almost surely. Denote* ${}^2(\ ) = \mathrm{E}(k\ k^2)$*. Let* $fz_i g_{i=1}^m$ *be independent random drawers of . For any* $0 < \ < 1$*, with confidence* $1$ *,*

$$\left\| \frac{1}{m}\sum_{i=1}^m (\ _i \quad \mathrm{E}[\ _i]) \right\| \quad \frac{2M\ln(2= \ )}{m} + \sqrt{\frac{2\ {}^2(\ )\ln(2= \ )}{m}}$$

From Lemma 3 and the union bound, with a probability at least 1 , we have

$$u_k \quad \left( {}^{p}\overline{K} + \sqrt{2\log\frac{jSj}{}} \right) \left( \frac{2\ln(2K= \ )}{jS_k^1 j} + \sqrt{\frac{2\ln(2K= \ )}{jS_k^1 j}} \right); 8k\ 2\ [K]: \tag{24}$$

Combining (23) and (24), we have

$$\begin{aligned} \widehat{B}_k^3 &\quad \left( {}^{p}\overline{K} + \sqrt{2\log\frac{jSj}{}} \right) \left( \frac{2}{jS_k^1 j}\ln\frac{2K}{\ } + \sqrt{\frac{2}{jS_k^1 j}\ln\frac{2K^2}{\ }} \right) \\ &\overset{(10),(14),(5)}{\quad} \left( {}^{p}\overline{K} + \sqrt{2\log\frac{jSj}{}} \right) 2\sqrt{\frac{9}{\ _k jSj}\ln\frac{2K}{\ }} = O\left( \sqrt{\frac{\ln jSj}{jSj}} \right); 8k\ 2\ [K]: \end{aligned} \tag{25}$$

In summary, under the condition that (10), (14) and (15) are true, with a probability at least 1 3 ,

$$kB_k k_\infty \quad O( \quad \sqrt{\ln jSj}) + O\left( \frac{\sqrt{\ln jSj} + {}^{p}\overline{\ln d}}{\sqrt{jSj}} \right); 8k\ 2\ [K]: \tag{26}$$

## A. Chernoff Bound

**Theorem 2** (Multiplicative Chernoff Bound (Angluin & Valiant, 1979)). *Let* $X_1, X_2; : : : ; X_n$ *be independent binary random variables with* $\Pr[X_i = 1] = p_i$*. Denote* $S = \sum_{i=1}^n X_i$ *and* $\ = \mathrm{E}[S] = \sum_{i=1}^n p_i$*. We have*

$$\Pr[S \quad (1 \quad )\ ] \quad \exp\left( \frac{{}^2}{2} \right); \textit{ for } 0 < \ < 1;$$

$$\Pr[S \quad (1 + \ )\ ] \quad \exp\left( \frac{{}^2}{2 + \ } \right); \textit{ for } \ > 0:$$

Therefore,

$$\Pr\left[S \quad \left(1 \quad \sqrt{\frac{2}{\ }\ln\frac{1}{\ }}\right) \quad \right] \quad ;\ \text{for}\ \exp\left(\quad \frac{2}{\ }\right) < \quad < 1;$$

$$\Pr\left[S \quad 2 \quad + 2\ln\frac{1}{\ } \quad \left(1 + \frac{\ln\frac{1}{\delta} + \sqrt{2 \quad \ln\frac{1}{\delta}}}{\ }\right) \quad \right] \quad ;\ \text{for}\ 0 < \quad < 1:$$

## B. Tail bounds for the Gaussian distribution

**Theorem 3** (Chernoff-type upper bound for the $Q$-function (Chang et al., 2011)). *The $Q$-function defined as*

$$Q(x) = p\frac{1}{\overline{2}}\int_x^\infty \exp\left(\quad \frac{t^2}{2}\right) dt$$

*is the tail probability of the standard Gaussian distribution. When $x > 0$, we have*

$$Q(x) \quad \frac{1}{2}\exp\left(\quad \frac{x^2}{2}\right):$$

Let $X \quad N(0;1)$ be a Gaussian random variable. According to Theorem 3, we have

$$\Pr[jXj \quad ] \quad \exp\left(\quad \frac{^2}{2}\right); \text{or}$$

$$\Pr\left[jXj \quad \sqrt{2\ln\frac{1}{\ }}\right] \quad :$$

## References

Angluin, D. and Valiant, L.G. Fast probabilistic algorithms for hamiltonian circuits and matchings. *Journal of Computer and System Sciences*, 18(2):155–193, 1979.

Chang, Seok-Ho, Cosman, Pamela C., and Milstein, Laurence B. Chernoff-type bounds for the gaussian error function. *IEEE Transactions on Communications*, 59(11):2939–2944, 2011.

Laurent, B. and Massart, P. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, 28 (5):1302–1338, 2000.

Smale, Steve and Zhou, Ding-Xuan. Learning theory estimates via integral operators and their approximations. *Constructive Approximation*, 26(2):153–172, 2007.