A. Proof of Lemma 1

Lemma 1 *The characteristic functions of the variable z's distributions in* $Z^{(i)}$ *,* $\forall\, i \in \{1, 2\}$*, are*

$$\Phi_{Z^{(1)}}(t) = \frac{\sigma_x}{r^{(1)}} \left(1 + t^2\right)^{-\frac{1}{2}} \sum_{i=1}^{r^{(1)}} H(i) \tag{A.1}$$

and

$$\Phi_{Z^{(2)}}(t) = \frac{\sigma_x}{r^{(1)}} \left(1 + t^2\right)^{-\frac{1}{2}} \sum_{i=1}^{r^{(1)}} H\left(\left\lceil \frac{i}{MK} \right\rceil \times MK\right), \tag{A.2}$$

where $H(k) = \frac{k(MN-k)}{(MN)^3 \cdot \left[f\left(F^{-1}\left(\frac{k}{MN}\right)\right)\right]^2}$*, f and F are probability density function and cumulative distribution function of* $W^{(0)}$*'s Gaussian distribution,* F^{-1} *denotes the quantile function associated with F. where K is the number of balance range per row,* $r^{(1)} = MK \cdot r^{(2)}$ *is the total number of pruned elements.*

Proof: To prove that, we consider each element w in $W^{(0)}$ as a random sample drawn from the the continuous distribution $\mathcal{N}(0, \sigma_w^2)$. Thus the kth smallest element in $Abs\left(W^{(0)}\right)$ is exactly the kth order statistic. In large sample size case(Mosteller 2006), the distribution of kth order statistic is

$$\xi_{k:num} \sim AN\left(F^{-1}(p), \frac{p(1-p)}{num \cdot \left[f\left(F^{-1}(p)\right)\right]^2}\right) \tag{A.3}$$

where num is the random sample times, $p = k/num$, f and F are probability density function and cumulative distribution function of Gaussian distribution in our case.

Since we prune the elements by magnitude order, the pruned element i.e. the element in $dW^{(i)}$ is satisfied with the distribution of order statistic. By considering those elements as different Gaussian distributions with same weights, we find that both the distributions of $dW^{(1)}$ and $dW^{(2)}$ are mixture of Gaussian distribution:

$$f_{dW}(x\,; \mu_1, ..., \mu_r, \sigma_1, ..., \sigma_r) = \frac{1}{r} \sum_{i=1}^{r} g(x\,; \mu_i, \sigma_i) \tag{A.4}$$

where $g(x\,; \mu, \sigma)$ denotes a probability density function of Gaussian distribution.

At last, We can calculate the product of dw and x. We provide a detailed description of the calculation on the Random Sparsity case (i.e., the denote $i = 1$) as an example.

$$z = dw \cdot x \tag{A.5}$$

$$\begin{aligned} f_{Z^{(1)}}(z) &= \int_{-\infty}^{+\infty} f_{dW^{(1)}}\left(\frac{z}{x}\,; \mu_1, ..., \mu_r, \sigma_1, ..., \sigma_r\right) \cdot \\ &\quad f_x(x\,; \mu_x, \sigma_x)\, dx \\ &= \int_{-\infty}^{+\infty} \frac{1}{r^{(1)}} \sum_{i=1}^{r^{(1)}} g\left(\frac{z}{x}\,; \mu_i, \sigma_i\right) g(x\,; \mu_x, \sigma_x)\, dx \\ &= \frac{1}{r^{(1)}} \sum_{i=1}^{r^{(1)}} \int_{-\infty}^{+\infty} g\left(\frac{z}{x}\,; \mu_i, \sigma_i\right) g(x\,; \mu_x, \sigma_x)\, dx \end{aligned} \tag{A.6}$$

where $f_x(x\,; \mu_x, \sigma_x)$ is the probability density function of each element in input X. Notice $\mu_x = 0$, a simplify can be done:

$$\begin{aligned} f_{Z^{(1)}}(z) &= \frac{1}{r^{(1)}} \sum_{i=1}^{r^{(1)}} \int_{-\infty}^{+\infty} g\left(\frac{z}{x}\,; \mu_i, \sigma_i\right) g(x\,; 0, \sigma_x)\, dx \\ &= \frac{1}{r^{(1)}} \sum_{i=1}^{r^{(1)}} \int_{-\infty}^{+\infty} g\left(\frac{z}{x}\,; 0, \sigma_i\right) g(x\,; 0, \sigma_x)\, dx + \\ &\quad \frac{1}{r^{(1)}} \sum_{i=1}^{r^{(1)}} \int_{-\infty}^{+\infty} \mu_i g(x\,; 0, \sigma_x)\, dx \\ &= \frac{1}{r^{(1)}} \sum_{i=1}^{r^{(1)}} \int_{-\infty}^{+\infty} g\left(\frac{z}{x}\,; 0, \sigma_i\right) g(x\,; 0, \sigma_x)\, dx \end{aligned} \tag{A.7}$$

In order to get a further result, we calculate the characteristic function $\Phi_{Z^{(1)}}(t)$.

$$\begin{aligned} \Phi_{Z^{(1)}}(t) &= \int_{-\infty}^{+\infty} e^{itz} f_Z(z)\, dz \\ &= \int_{-\infty}^{+\infty} e^{itz} \frac{1}{r^{(1)}} \sum_{i=1}^{r^{(1)}} \int_{-\infty}^{+\infty} g\left(\frac{z}{x}\,; 0, \sigma_i\right) \cdot \\ &\quad g(x\,; 0, \sigma_x)\, dx dz \\ &= \frac{\sigma_x}{r^{(1)}} \sum_{i=1}^{r^{(1)}} \sigma_i \left(1 + t^2\right)^{-\frac{1}{2}} \\ &= \frac{\sigma_x}{r^{(1)}} \left(1 + t^2\right)^{-\frac{1}{2}} \sum_{i=1}^{r^{(1)}} H(i) \end{aligned} \tag{A.8}$$

Similarly, the characteristic function $\Phi_{Z^{(2)}}(z)$ is

$$\begin{aligned} \Phi_{Z^{(2)}}(t) &= \frac{\sigma_x}{r^{(1)}} \left(1 + t^2\right)^{-\frac{1}{2}} MK \sum_{i=1}^{r^{(2)}} H(i) \\ &= \frac{\sigma_x}{r^{(1)}} \left(1 + t^2\right)^{-\frac{1}{2}} \sum_{i=1}^{r^{(1)}} H\left(\left\lceil \frac{i}{MK} \right\rceil \times MK\right) \end{aligned} \tag{A.9}$$

B. Proof of Theorem 1

Theorem 1 *The means of the variable z's distributions in* $Z^{(i)}$ *,* $\forall\, i \in \{1, 2\}$*, are*

$$Mean_{Z^{(1)}}(z) = Mean_{Z^{(2)}}(z) = 0. \tag{B.1}$$

The variances the variable z's distributions in $Z^{(i)}$ *,* $\forall\, i \in \{1, 2\}$*, are*

$$Var_{Z^{(1)}}(z) = \frac{\sigma_x}{r^{(1)}} \sum_{i=1}^{r^{(1)}} H(i) \tag{B.2}$$

and

$$Var_{Z^{(2)}}(z) = \frac{\sigma_x}{r^{(1)}} \sum_{i=1}^{r^{(1)}} H\left(\left\lceil \frac{i}{MK} \right\rceil \times MK\right). \quad \text{(B.3)}$$

Proof: Generally, the characteristic function has this property

$$E\left(X^k\right) = (-i)^k \, \Phi_Z^k(0) \quad \text{(B.4)}$$

Thus the mean values of z are

$$\begin{aligned} Mean_{Z^{(i)}}(z) &= E_{Z^{(i)}}(z) \\ &= (-i)\, \Phi'_{Z^{(i)}}(z)|_{z=0} \\ &= 0, \ \forall\, i \in \{1, 2\} \end{aligned} \quad \text{(B.5)}$$

Thus the variance values of z are

$$\begin{aligned} Var_{Z^{(i)}}(z) &= E_{Z^{(i)}}\left(z^2\right) - \left[E_{Z^{(i)}}(z)\right]^2 \\ &= (-i)^2\, \Phi''_{Z^{(i)}}(z)|_{z=0} \end{aligned} \quad \text{(B.6)}$$

Then substitute X into the above equation

$$Var_{Z^{(1)}}(z) = \frac{\sigma_x}{r^{(1)}} \sum_{i=1}^{r^{(1)}} H(i) \quad \text{(B.7)}$$

$$Var_{Z^{(2)}}(z) = \frac{\sigma_x}{r^{(1)}} \sum_{i=1}^{r^{(1)}} H\left(\left\lceil \frac{i}{MK} \right\rceil \times MK\right) \quad \text{(B.8)}$$