# Appendices

## A. Proofs

**Proposition 1.** *Let* $\mathbf{u} = \mathbf{W}\mathbf{x}$ *where* $\mathbf{x} \in \mathbb{R}^n$ *and* $\mathbf{W} \in \mathbb{R}^{m \times n}$ *such that* $\mathbb{E}_{\mathbf{x}}[\mathbf{x}] = \mathbf{0}$ *and* $\mathbb{E}_{\mathbf{x}}[\mathbf{x}\mathbf{x}^T] = \sigma^2 \mathbf{I}$ *(*$\mathbf{I}$ *is the identity matrix) . Then the covariance matrix of* $\mathbf{u}$ *is approximately canonical satisfying,*

$$\min_{\alpha} \|\mathbf{\Sigma} - \operatorname{diag}(\alpha)\|_F \leq \sigma^2 \sqrt{m(m-1)\mu^2 \sum_{i,j=1; i \neq j}^{m} \|\mathbf{W}_i\|_2^2 \|\mathbf{W}_j\|_2^2} \tag{15}$$

*where* $\mathbf{\Sigma} = \mathbb{E}_{\mathbf{u}}[(\mathbf{u} - \mathbb{E}_{\mathbf{u}}[\mathbf{u}])(\mathbf{u} - \mathbb{E}_{\mathbf{u}}[\mathbf{u}])^T]$ *is the covariance matrix of* $\mathbf{u}$*,* $\mu$ *is the coherence of the rows of* $\mathbf{W}$*,* $\alpha \in \mathbb{R}^m$ *is the closest approximation of the covariance matrix to a canonical ellipsoid and* $\operatorname{diag}(.)$ *diagonalizes a vector to a diagonal matrix. The corresponding optimal* $\alpha_i^* = \sigma^2 \|\mathbf{W}_i\|_2^2 \; \forall i \in \{1, \ldots, m\}$*.*

*Proof.* Notice that,

$$\mathbb{E}_{\mathbf{u}}[\mathbf{u}] = \mathbf{W}\mathbb{E}_{\mathbf{x}}[\mathbf{x}] = \mathbf{0} \tag{16}$$

On the other hand, the covariance of $\mathbf{u}$ is given by,

$$\begin{aligned} \mathbf{\Sigma} = \mathbb{E}_{\mathbf{u}}[(\mathbf{u} - \mathbb{E}_{\mathbf{u}}[\mathbf{u}])(\mathbf{u} - \mathbb{E}_{\mathbf{u}}[\mathbf{u}])^T] &= \mathbb{E}_{\mathbf{x}}[(\mathbf{W}\mathbf{x} - \mathbf{W}\mathbb{E}_{\mathbf{x}}[\mathbf{x}])(\mathbf{W}\mathbf{x} - \mathbf{W}\mathbb{E}_{\mathbf{x}}[\mathbf{x}])^T] \\ &= \mathbb{E}_{\mathbf{x}}[\mathbf{W}(\mathbf{x} - \mathbb{E}_{\mathbf{x}}[\mathbf{x}])(\mathbf{x} - \mathbb{E}_{\mathbf{x}}[\mathbf{x}])^T \mathbf{W}^T] \\ &= \mathbf{W}\mathbb{E}_{\mathbf{x}}[(\mathbf{x} - \mathbb{E}_{\mathbf{x}}[\mathbf{x}])(\mathbf{x} - \mathbb{E}_{\mathbf{x}}[\mathbf{x}])^T]\mathbf{W}^T \end{aligned} \tag{17}$$

Since $\mathbf{x}$ has spherical covariance, the off-diagonal elements of $\mathbb{E}_{\mathbf{x}}[(\mathbf{x} - \mathbb{E}_{\mathbf{x}}[\mathbf{x}])(\mathbf{x} - \mathbb{E}_{\mathbf{x}}[\mathbf{x}])^T]$ are zero and the diagonal elements are the variance of any individual unit, since all units are identical. Thus,

$$\mathbb{E}_{\mathbf{u}}[(\mathbf{u} - \mathbb{E}_{\mathbf{u}}[\mathbf{u}])(\mathbf{u} - \mathbb{E}_{\mathbf{u}}[\mathbf{u}])^T] = \sigma^2 \mathbf{W}\mathbf{W}^T \tag{18}$$

Thus,

$$\begin{aligned} \|\mathbf{\Sigma} - \operatorname{diag}(\alpha)\|_F^2 &= \operatorname{tr}\left((\sigma^2 \mathbf{W}\mathbf{W}^T - \operatorname{diag}(\alpha))(\sigma^2 \mathbf{W}\mathbf{W}^T - \operatorname{diag}(\alpha))^T\right) \\ &= \operatorname{tr}\left(\sigma^4 \mathbf{W}\mathbf{W}^T\mathbf{W}\mathbf{W}^T + \operatorname{diag}(\alpha^2) - 2\sigma^2 \operatorname{diag}(\alpha)\mathbf{W}\mathbf{W}^T\right) \\ &= \sigma^4 \|\mathbf{W}\mathbf{W}^T\|_F^2 + \sum_{i=1}^{m} \left(\alpha_i^2 - 2\sigma^2 \alpha_i \|\mathbf{W}_i\|_2^2\right) \\ \leq \sigma^4 \sum_{i=1}^{m} \left(\|\mathbf{W}_i\|_2^4\right) + \sum_{i,j=1; i \neq j}^{m} & m(m-1)\mu^2 \|\mathbf{W}_i\|_2^2 \|\mathbf{W}_j\|_2^2 + \sum_{i=1}^{m} \left(\alpha_i^2 - 2\sigma^2 \alpha_i \|\mathbf{W}_i\|_2^2\right) \end{aligned} \tag{19}$$

$\alpha^2$ in the above equation denotes element-wise square of elements of $\alpha$. Finally minimizing w.r.t $\alpha_i \; \forall i \in \{1, \ldots, m\}$, leads to $\alpha_i^* = \sigma^2 \|\mathbf{W}_i\|_2^2$. Substituting this into equation 19, we get,

$$\|\mathbf{\Sigma} - \operatorname{diag}(\alpha)\|_F^2 \leq \sigma^4 \sum_{i,j=1; i \neq j}^{m} m(m-1)\mu^2 \|\mathbf{W}_i\|_2^2 \|\mathbf{W}_j\|_2^2 \tag{20}$$

□

**Remark 1.** *Let* $X \sim \mathcal{N}(0, 1)$ *and* $Y = \max(0, X)$*. Then* $\mathbb{E}[Y] = \frac{1}{\sqrt{2\pi}}$ *and* $\operatorname{var}(Y) = \frac{1}{2}\left(1 - \frac{1}{\pi}\right)$

*Proof.* For the definition of $X$ and $Y$, we have,

$$\mathbb{E}[Y] = \frac{1}{2}.0 + \frac{1}{2}\mathbb{E}[Z] = \frac{1}{2}\mathbb{E}[Z] \tag{21}$$

where $Z$ is sampled from a Half-Normal distribution such that $Z = |X|$; thus $\mathbb{E}[Z] = \sqrt{\frac{2}{\pi}}$ leading to the claimed result. In order to compute variance, notice that $\mathbb{E}[Y^2] = 0.5\mathbb{E}[Z^2]$. Then,

$$\text{var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = 0.5\mathbb{E}[Z^2] - \frac{1}{4}\mathbb{E}[Z]^2 = 0.5(\text{var}(Z) + \mathbb{E}[Z]^2) - \frac{1}{4}\mathbb{E}[Z]^2 \tag{22}$$

Substituting $\text{var}(Z) = 1 - \frac{2}{\pi}$ yields the claimed result. □

**Remark 2.** *Let $X \sim \mathcal{N}(0,1)$ and $Y = PReLU_a(X)$. Then $\mathbb{E}[Y] = (1-a)\frac{1}{\sqrt{2\pi}}$ and $\text{var}(Y) = \frac{1}{2}\left((1+a^2) - \frac{(1-a)^2}{\pi}\right)$*

*Proof.* For the definition of $X$ and $Y$, half the mass of $Y$ is concentrated on $\mathbb{R}^+$ with Half-Normal distribution, while the other half of the mass is concentrated on $\mathbb{R}^{-\text{sign}(a)}$with Half-Normal distribution scaled with $|a|$. Thus,

$$\mathbb{E}[Y] = -\frac{a}{2}\mathbb{E}[Z] + \frac{1}{2}\mathbb{E}[Z] = (1-a)\frac{1}{2}\mathbb{E}[Z] \tag{23}$$

where $Z$ is sampled from a Half-Normal distribution such that $Z = |X|$; thus $\mathbb{E}[Z] = \sqrt{\frac{2}{\pi}}$ leading to the claimed result. Similarly in order to compute variance, notice that $\mathbb{E}[Y^2] = 0.5\mathbb{E}[(aZ)^2] + 0.5\mathbb{E}[Z^2] = 0.5\mathbb{E}[Z^2](1+a^2)$. Then,

$$\begin{aligned}\text{var}(Y) &= \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = 0.5\mathbb{E}[Z^2](1+a^2) - (1-a)^2\frac{1}{4}\mathbb{E}[Z]^2 \\ &= 0.5(1+a^2)(\text{var}(Z) + \mathbb{E}[Z]^2) - (1-a)^2\frac{1}{4}\mathbb{E}[Z]^2\end{aligned} \tag{24}$$

Substituting $\text{var}(Z) = 1 - \frac{2}{\pi}$ yields the claimed result. □