APPENDIX I

Let $r_{i,j}$ and $\epsilon_{i,j}$ be the i,j-th entry of matrix $R_{k\times m}$ and R^TR, respectively. Each $r_{i,j}$ is independent and identically chosen from $N(0,\sigma_r)$. Now let us prove $E[\epsilon_{i,i}]=k\sigma_r^2$, $Var[\epsilon_{i,i}]=2k\sigma_r^4$, $\forall i$; and $E[\epsilon_{i,j}]=0$, $Var[\epsilon_{i,j}]=k\sigma_r^4$, $\forall i,j,i\neq j$.

Proof: Note that $\epsilon_{i,i}=\sum_{t=1}^k r_{t,i}^2$ and $\epsilon_{i,j}=\sum_{t=1}^k r_{t,i}r_{t,j}$, $i\neq j$, we have $E[\epsilon_{i,i}]=E[\sum_{t=1}^k r_{t,i}^2]=kE[r_{t,i}^2]=k\sigma_r^2$ and $E_{i\neq j}[\epsilon_{i,j}]=E[\sum_{t=1}^k r_{t,i}r_{t,j}]=\sum_{t=1}^k E[r_{t,i}r_{t,j}]=\sum_{t=1}^k E[r_{t,i}]E[r_{t,j}]=0$.

To obtain the variance of $\epsilon_{i,i}$, we first compute $E[\epsilon_{i,i}^2]=E[\sum_{t=1}^k r_{t,i}^4+\sum_{p\neq q,1\leq p,q\leq k} r_{p,i}^2 r_{q,i}^2]=kE[r_{t,i}^4]+k(k-1)E[r_{p,i}^2]E[r_{q,i}^2]=3k\sigma_r^4+k(k-1)\sigma_r^4=(2k+k^2)\sigma_r^4$. The second to the last equation in the above is based on the fact that $E[r_{t,j}^4]=3\sigma_r^4$ for random variable $r_{t,j}\sim N(0,\sigma_r)$[5]. Therefore, $Var[\epsilon_{i,i}]=E[\epsilon_{i,i}^2]-(E[\epsilon_{i,i}])^2=2k\sigma_r^4$. Similarly, $E_{i\neq j}[\epsilon_{i,j}^2]=E_{i\neq j}[\sum_{t=1}^k r_{t,i}^2 r_{t,j}^2+\sum_{p\neq q,1\leq p,q\leq k} r_{p,i}r_{p,j}r_{q,i}r_{q,j}]=kE[\sum_{t=1}^k r_{t,i}^2 r_{t,j}^2]+0=k\sigma_r^4$, hence, $Var_{i\neq j}[\epsilon_{i,j}]=k\sigma_r^4$. ■

APPENDIX II

Lemma 5.5: Let x, y be two data vectors in $\mathbb{R}^m$. Let R be a $k\times m$ dimensional random matrix. Each entry of the random matrix is independent and identically chosen from Gaussian distribution with mean zero variance σ_r^2. Further let

$$u=\frac{1}{\sqrt{k}\sigma_r}Rx, \qquad \text{and} \qquad v=\frac{1}{\sqrt{k}\sigma_r}Ry \text{ , then}$$

$$\begin{aligned} E[u^Tv-x^Ty] &= 0 \\ Var[u^Tv-x^Ty] &= \frac{1}{k}\left(\sum_i x_i^2\sum_i y_i^2+\left(\sum_i x_iy_i\right)^2\right) \end{aligned}$$

In particular, if both x and y are normalized to unity, $\sum_i x_i^2\sum_i y_i^2=1$ and $(\sum_i x_iy_i)^2\leq 1$. We have the upper bound of the variance as follows:

$$Var[u^Tv-x^Ty]\leq\frac{2}{k}$$

[5] http://mathworld.wolfram.com/NormalDistribution.html

Proof: Using Lemma 5.2, the expectation of projection distortion is

$$\begin{aligned} E[u^T v - x^T y] &= E[\frac{1}{k\sigma_r^2} x^T R^T R y - x^T y] \\ &= \frac{1}{k\sigma_r^2} x^T E[R^T R] y - x^T y \\ &= \frac{1}{k\sigma_r^2} k\sigma_r^2 x^T y - x^T y \\ &= 0 \end{aligned}$$

To compute the variance of the distortion, let us first express the inner product between the projected vectors as

$$\begin{aligned} u^T v &= \frac{1}{\sqrt{k}\sigma_r} x^T R^T \frac{1}{\sqrt{k}\sigma_r} R y \\ &= \frac{1}{k\sigma_r^2} x^T R^T R y \\ &= \frac{1}{k\sigma_r^2} (\sum_i x_i \epsilon_{i,i} y_i + \sum_{i \neq j} x_i \epsilon_{i,j} y_j) \\ &= \frac{1}{k\sigma_r^2} \sum_i x_i \epsilon_{i,i} y_i + \frac{1}{k\sigma_r^2} \sum_{i \neq j} x_i \epsilon_{i,j} y_j \end{aligned}$$

Denote $\frac{1}{k\sigma_r^2} \sum_i x_i \epsilon_{i,i} y_i$ as Φ and $\frac{1}{k\sigma_r^2} \sum_{i \neq j} x_i \epsilon_{i,j} y_j$ as Ψ. Then $Var[u^T v] = Var[\Phi] + Var[\Psi] + 2Cov[\Phi, \Psi]$.

Now let us compute $Cov[\Phi, \Psi]$.

$$Cov[\Phi, \Psi] = E[\Phi\Psi] - E[\Phi]E[\Psi]$$

Since $E[\epsilon_{i,j}] = 0 \; \forall i, j, i \neq j$, so $E[\Psi] = 0$. Hence,

$$\begin{aligned} Cov[\Phi, \Psi] &= E[\Phi\Psi] - 0 \\ &= \frac{1}{k^2\sigma_r^4} E[\sum_i x_i \epsilon_{i,i} y_i \times \sum_{p \neq q} x_p \epsilon_{p,q} y_q] \end{aligned}$$

It is straightforward to verify that $E[\epsilon_{i,i}\epsilon_{p,q}] = 0$ when $p \neq q$. So $Cov[\Phi, \Psi] = 0$.

The variance of Φ is

$$\begin{aligned}
Var[\Phi] &= Var[\frac{1}{k\sigma_r^2}\sum_i x_i\epsilon_{i,i}y_i] \\
&= \frac{1}{k^2\sigma_r^4}Var[\sum_i x_i\epsilon_{i,i}y_i] \\
&= \frac{1}{k^2\sigma_r^4}(E[(\sum_i x_i\epsilon_{i,i}y_i)^2] - (E[\sum_i x_i\epsilon_{i,i}y_i])^2) \\
&= \frac{1}{k^2\sigma_r^4}(E[\sum_i x_i^2\epsilon_{i,i}^2y_i^2 + \sum_{p\neq q} x_p y_p\epsilon_{p,p}x_q y_q\epsilon_{q,q}] \\
&\quad -(E[\sum_i x_i\epsilon_{i,i}y_i])^2)
\end{aligned}$$

Since $E[\epsilon_{i,i}] = k\sigma_r^2$, $E[\epsilon_{i,i}^2] = (2k+k^2)\sigma_r^4$ and $E[\epsilon_{p,p}\epsilon_{q,q}] = k^2\sigma_r^4$, we have

$$\begin{aligned}
Var[\Phi] &= \frac{1}{k^2\sigma_r^4}(2k+k^2)\sigma_r^4\sum_i x_i^2y_i^2 + \sum_{p\neq q} x_p y_p x_q y_q - (\sum_i x_i y_i)^2 \\
&= (\frac{2}{k}+1)\sum_i x_i^2y_i^2 + \sum_{p\neq q} x_p y_p x_q y_q - (\sum_i x_i y_i)^2
\end{aligned}$$

The variance of Ψ is

$$\begin{aligned}
Var[\Psi] &= \frac{1}{k^2\sigma_r^4}Var[\sum_{i\neq j} x_i\epsilon_{i,j}y_j] \\
&= \frac{1}{k^2\sigma_r^4}(E[(\sum_{i\neq j} x_i\epsilon_{i,j}y_j)^2] - (E[\sum_{i\neq j} x_i\epsilon_{i,j}y_j])^2) \\
&= \frac{1}{k^2\sigma_r^4}(E[(\sum_{i\neq j} x_i\epsilon_{i,j}y_j)^2] - 0 \\
&= \frac{1}{k^2\sigma_r^4}\sum_{i\neq j}\sum_{p\neq q} x_i y_j x_p y_q E[\epsilon_{i,j}\epsilon_{p,q}]
\end{aligned}$$

Since $E[\epsilon_{i,j}\epsilon_{p,q}] = 0$ unless $i = p$ and $j = q$, or $i = q$ and $j = p$. Therefore,

$$\begin{aligned}
Var[\Psi] &= \frac{1}{k^2\sigma_r^4}(\sum_{i\neq j} x_i^2y_j^2 + \sum_{i\neq j} x_i y_j x_j y_i)E_{i\neq j}[\epsilon_{i,j}^2] \\
&= \frac{1}{k^2\sigma_r^4}(\sum_i x_i^2\sum_{j\neq i} y_j^2 + \sum_i x_i y_i\sum_{j\neq i} x_j y_j)k\sigma_r^4 \\
&= \frac{1}{k}(\sum_i x_i^2\sum_i y_i^2 - \sum_i x_i^2y_i^2 + (\sum_i x_i y_i)^2 - \sum_i x_i^2y_i^2) \\
&= \frac{1}{k}(\sum_i x_i^2\sum_i y_i^2 + (\sum_i x_i y_i)^2 - 2\sum_i x_i^2y_i^2)
\end{aligned}$$

Thus,

$$
\begin{aligned}
Var[u^T v] &= Var[\Phi] + Var[\Psi] + 0 \\
&= (\frac{2}{k} + 1)\sum_i x_i^2 y_i^2 + \sum_{p \neq q} x_p y_p x_q y_q - (\sum_i x_i y_i)^2 \\
&\quad + \frac{1}{k}(\sum_i x_i^2 \sum_i y_i^2 + (\sum_i x_i y_i)^2 - 2\sum_i x_i^2 y_i^2) \\
&= \frac{1}{k}(\sum_i x_i^2 y_i^2 + (\sum_i x_i y_i)^2) + (\sum_i x_i^2 y_i^2 \\
&\quad + \sum_{p \neq q} x_p y_p x_q y_q - (\sum_i x_i y_i)^2) \\
&= \frac{1}{k}(\sum_i x_i^2 \sum_i y_i^2 + (\sum_i x_i y_i)^2)
\end{aligned}
$$

This gives the final result $Var[u^T v - x^T y] = \frac{1}{k}(\sum_i x_i^2 \sum_i y_i^2 + (\sum_i x_i y_i)^2)$. ■