(1) $u^T X_i$ へ変換したデータの分散は、もとのデータ X の分散を用いて

$\mathrm{var}(u^T X_i) = u^T\ \mathrm{var}(X) u$ となる。

$$
\begin{aligned}
\mathrm{var}(u^T X_i) &= \frac{1}{N}\sum_{i=1}^{N}(u^T X_i - u^T \frac{1}{N}\sum_{j=1}^{N} X_j)(u^T X_i - u^T \frac{1}{N}\sum_{j=1}^{N} X_j)^T \\
&= u^T \frac{1}{N}\sum_{i=1}^{N}(X_i - \frac{1}{N}\sum_{j=1}^{N} X_j)(X_i - \frac{1}{N}\sum_{j=1}^{N} X_j)^T u \\
&= u^T V u
\end{aligned}
$$

よって $u_1 = \arg\max_{\|u\|=1} \frac{1}{N}\{\sum_{i=1}^{N} u^T (X_i - \frac{1}{N}\sum_{j=1}^{N} X_j)\}^2 = \arg\max_{\|u\|=1} u^T V u$

(2)b は直交行列で基底変換した 1 次独立な固有ベクトルである。よって N×N の正方行列

である $\tilde{K}$ は対角化可能であり、$b^T \tilde{K} b = diag\{\lambda_1, \lambda_2, ..., \lambda_n\}$ が成り立つ。よって与えられ

た最適化問題は固有値を求める問題となる。

(3) $Ax = \lambda x$ の固有値の定義より、$\tilde{K}x = \lambda x \to \tilde{K} = x\lambda x^T$

$x = U, \lambda = \Lambda$

U は直交行列で、Λ は固有値を対角成分とする対角行列である。

(4) $\min_a \sum_{i=1}^{N} \| Y_i - a^T X_i \|^2 + \lambda \| a \|^2$

最小値を求めるために $\| Y_i - a^T X_i \|^2 + \lambda \| a \|^2$ を偏微分したものを 0 と置く。

$$
\begin{aligned}
\frac{\partial}{\partial a} &= (Y - a^T X)^T (Y - a^T X) + \lambda \| a \|^2 \\
&= 2(X^T X a - X^T Y + \lambda a) = 0 \\
&= (X^T X - \lambda I_N) a = X^T Y
\end{aligned}
$$

$$
a = (X^T X - \lambda I_N)^{-1} X^T Y
$$