Yizhou Liu liu773@wisc.edu

1. a) $Z = X^T = (U\Sigma V^T)^T = V\Sigma U^T$.

b). orthonormal basis: $U^{-1}\Sigma^{-1}V^{-T}$

2. a) for $y = Xw$, the soln. won't be unique if $\text{rank}\{X\} = \text{rank}\{X|y\}$ and $\text{rank}\{X\} < \dim\{w\}$.

b). $\hat{y} = (X^TX + \lambda I)^{-1}X^Ty$.

c) When we need a unique soln, $\text{rank}\{X\} = \text{rank}\{X|y\}$ & $\text{rank}\{X\} = \dim\{w\}$.

3. a) Since $X^TX = (U\Sigma V^T)^T(U\Sigma V^T)$

$= V\Sigma U^T U\Sigma V^T$.

$= V\Sigma^2 V^T$.

$\lambda I = V\lambda I V^T$.

then $(X^TX + \lambda I)^{-1}X^T = (V\Sigma^2V^T + V\lambda IV^T)^{-1}X^T$.

$= \sum_{i=1}^{p} G_i u_i v_i^T$

b). if X^TX is invertible, as $\lambda \to 0$.

we get. $X^+ = (X^TX + 0\lambda)^{-1}X^T$

$= (X^TX)^{-1}X^T$.

c). if X is square & invertible,

$X^+ = (X^TX)^{-1}X^T = X^{-1}X^{-T}X^T = X^{-1}$

d) if $r < p$, $\lambda > 0$.

$$(X^T X + \lambda I)^{-1} X^T = \sum_{i=1}^{p} \frac{\sigma_i^2}{\sigma_i^2 + \lambda} v_i u_i^T$$

e). if $r < p$; $X^+ = \lim_{\lambda \downarrow 0} (X^T X + \lambda I)^{-1} X^T$

$$= (X^T X)^{-1} X^T$$

$$= \cancel{e}\ X^{-1} X^{-T} X^T$$

$$= (U \Sigma U^T)^{-1} (U \Sigma V^T)^{-T} (U \Sigma V^T)^T$$

$$= V^{-T} \Sigma U^{-1} \cancel{V} (U \Sigma V^T)^{-T} (V \Sigma U^T)$$

$$= V \Sigma_r^{-1} U^T.$$

$$= \sum_{i=1}^{r} \frac{1}{\sigma_i} v_i u_i^T$$

4. a) Yes; No.

b). No. Since one-D subspace's center is not close to ϕ.

c). Yes.

Removing the mean make the data more close the origin center, thereby increasing the ~~c~~ PCA's desired effect