CS 182 HW8.

1. (a). $x_1 = \cancel{A} Bu_0$

$x_2 = Ax_1 + Bu_1 = A^{0} B u_0 + Bu_1$

$\vdots$

$x_k = \sum_{j=0}^{k-1} A^{k-1-j} B u_j$

$y_k = Cx_k + Du_k = C \sum_{j=0}^{k-1} A^{k-1-j} B u_j + Du_k.$

Let $K_\ell = \begin{cases} D, & \ell=0 \\ CA^{\ell-1}B, & \ell \geq 1 \end{cases}$

$\therefore y_k = \sum_{\ell=0}^{k} K_\ell u_{k-\ell}$

(b). i. $x_{k+1} = 0.8x_k + u_k$, $y_k = 1.5x_k + \delta u_k$.

$K_0 = \delta$, $K_1 = 1.5 \cdot 1 = 1.5$, $K_2 = 1.5 \cdot 0.8 \cdot 1 = 1.2$

$K_3 = 1.5 \cdot 0.8^2 = 0.96$, $K_4 = 1.5 \cdot 0.8^3 = 0.768$.

ii. $K_0 = D = 0$, $K_1 = 1$, $K_2 = 0.7$, $K_3 = CA^2B = [1, 0] \begin{bmatrix} 0.51 & 0.13 \\ 0.26 & 0.38 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$= 0.51$

(c). naive: path: $y_0 \to y_1 \to \dots \to y_L$: $O(L^2)$

every y_k need $O(k)$ computation.

use FFT, parallelize computation., length: $O(\log L)$

$\therefore O(L \log L)$

(d). naive: $O(n^3 L)$, divide and conquer: $O(n^{\log_2 7} L)$

(e). $K_\ell = \begin{cases} D & \ell=0 \\ C\,\mathrm{diag}(\lambda_1^{\ell-1}, \dots \lambda_n^{\ell-1}) B, & \ell \geq 1 \end{cases}$

$\therefore O(nL)$

(f). $A = I_n + PP^T$,

$S_k = P^T A^k P$, $A^k B = B + PP^T A^{k-1} B = B + P \cdot \frac{P^T A^k B - P^T B}{S_k - \|P\|^2}$

$O(nL)$.

3/(a). (i). $XX^T = \begin{bmatrix} 18.5583 & 0.0015 \\ 0.0015 & 0.0010 \end{bmatrix}$

$\|W^{(\alpha)}\|_F^2 = 1^2+0^2+1^2+0^2 = 2.$

$\|W^{(\beta)}\|_F^2 = 1^2+0^2+0^2+0^2 = 1.$

α	0	2
β	0.001	1

(ii). $L_2(W^{(\alpha)}) = 2\lambda$, $L_2(W^{(\beta)}) = 0.001 + \lambda$

$\therefore 2\lambda > \lambda + 0.001$

$\therefore \lambda > 0.001$

(b). (i). $\hat{W} = U\begin{bmatrix} \frac{\sigma_1^2}{\sigma_1+\lambda} & & \\ & \ddots & \\ & & \frac{\sigma_m^2}{\sigma_m+\lambda} \end{bmatrix} U^T$

(ii).

(c). $g_i(\lambda) = \frac{\sigma_i^2}{\sigma_i + \lambda}$

1. preserve first 3 directions at at least 80%.

$g_i(\lambda) \geq 80\%$, $i = 1, 2, 3.$

$\frac{16}{16+\lambda} \geq 0.8 \Rightarrow \lambda \leq 4$

2. Attenuate remaining 5 directions to at most 50%.

$g_i(\lambda) \leq 0.5$

$g_4(\lambda) \leq 0.5 \Rightarrow \lambda \geq 1.$

$\therefore 1 \leq \lambda \leq 4.$

4/(a). $m = \frac{1}{n}\sum_{i=1}^{n} x_i$,

$m' = \frac{1}{n+1}\sum_{i=1}^{n+1} x_i$

$\therefore m' = \frac{nm + x_{n+1}}{n+1}$

(b). $A^T = [x_1, x_2 \cdots x_n]$, $y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$

$\therefore A^TA = \sum_{i=1}^{n} x_i x_i^T \Rightarrow A^TA + \lambda I = \lambda I + \sum_{i=1}^{n} x_i x_i^T$

$A^Ty = \sum_{i=1}^{n} y_i x_i$

compute:

(c). K^TV, K^TK : $O(nd^2)$ ~~compute~~ Invert $M(d\times d)$: $O(d^3)$

each n query : $q_j^T(M^{-1}p)$: $O(d^2)$ compute $M^{-1}p$ $O(d^3)$

Total: $O(nd^2)$.

(d). $W^* = (K^TK + \lambda I)^{-1} K^TV$

output $= q^T W^*$

λ, keys, values, queries are all differentiable.

(e). $O(d^2)$

(f).

(g). $o = q^T W^* = q^T (K^TK + \lambda I)^{-1} K^TV$

Let:

$\alpha = (K^TK + \lambda I)^{-1} K q$, $\alpha \in R^n$.

$o = \sum_{i=1}^{n} \alpha_i v_i$

$\therefore$ $w_i = \alpha_i = [(K^TK + \lambda I)^{-1} K^T q]_i$

$w = (K^TK + \lambda I)^{-1} K^T q$.

5. (a). Grok.

(b). Nobody

(c). 3-4 h