EECS 551 Homework 4 YUZHAN JIANG

P1:

1 (a) Proof: $vec(AXB^T) = (B \otimes A) vec(X)$ where $A \in F^{p\times m}$, $X \in F^{m\times n}$, $B \in C^{q\times n}$

By Kronecker Product definition:

$$B \otimes A = \begin{bmatrix} b_{11}A & \cdots & b_{1n}A \\ \vdots & & \vdots \\ b_{q1}A & \cdots & b_{qn}A \end{bmatrix}$$ is $qp \times mn$ block matrix

let $A = [a_1 \; a_2 \; a_3 \cdots a_m]$ $\quad X = [x_1 \cdots x_n]$ $\quad B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & \cdots & & b_{2n} \\ \vdots \\ b_{q1} & \cdots & & b_{qn} \end{bmatrix}$

$$B^T = \begin{bmatrix} b_{11}^* & \cdots b_{k1}^* & b_{q1}^* \\ b_{12}^* & \cdots b_{k2}^* & \vdots \\ \vdots & \vdots & \vdots \\ b_{1n}^* & \cdots b_{kn}^* & b_{qn}^* \end{bmatrix}$$

$$B^T = [b_1^* \; b_2^* \cdots b_q^*]$$

The kth column of AXB^T is:

$$(AXB^T)_{:,k} = AXb_k^T = A\sum_{i=1}^{n} x_i b_{k,i}^*$$

$$= [b_{k,1}^* A, \; b_{k,2}^* A \cdots b_{k,n}^* A] \underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{vec(x)}$$

$$= ([b_{k,1}^*, \; b_{k,2}^* \cdots b_{k,n}^*] \otimes A) \; vec(X)$$

$$= (\bar{b}_k \otimes A) \cdot vec\,x$$

$$\therefore vec(AXB^T) = \begin{bmatrix} (AXB^T)_{:,1} \\ (AXB^T)_{:,2} \\ \vdots \\ (AXB^T)_{:,q} \end{bmatrix} = \begin{bmatrix} (b_1 \otimes A) vec(X) \\ (b_2 \otimes A) vec(X) \\ \vdots \\ (b_q \otimes A) vec(X) \end{bmatrix} = \begin{bmatrix} b_k \otimes A \\ b_2 \otimes A \\ \vdots \\ b_q \otimes A \end{bmatrix} vec(X)$$

$$= (B \otimes A)\, vec(X)$$

$$\therefore \quad vec(AXB^T) = (B \otimes A)\, vec(X)$$

$$\begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}\begin{bmatrix} 2 \\ 2 \end{bmatrix} = 8$$

1.(b)

AXB^T where $A, X, B^T \in F^{N\times N}$

$$AX = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}\begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nn} \end{bmatrix}$$ $n\times n\times n = n^3$ scalar muliplications.

and $$(AX)B^T = \begin{bmatrix} ax_{11} & \cdots & ax_{1n} \\ \vdots & & \vdots \\ ax_{n1} & \cdots & ax_{nn} \end{bmatrix}\begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nn} \end{bmatrix}$$ also requires $n\times n\times n = n^3$ scalar muliplications

So there are total $2n^3$ operations

$B\otimes A$ requries $= (1\cdot n^2)\cdot n^2 = n^4$ muliplations,

$$(B\otimes A)\,vec(X) = \underset{n^2\times n^2}{\begin{bmatrix} b_1A \\ b_2A \\ \vdots \\ b_nA \end{bmatrix}}\underset{n^2\times 1}{\begin{bmatrix} X_{:,1} \\ X_{:,2} \\ \vdots \\ X_{:,n} \end{bmatrix}}$$ requires $n^2\cdot n^2 = n^4$ muliplications

So there are total $2n^4$ muliplications for $(B\otimes A)\,vec(X)$

Therefore, $vec(AXB^T)$ takes fewer muliplications

P2.

(a) Firstly, we find the eigenvalues of $B = A + xx'$

let $\det(B - \lambda I) = 0$

$\det(A + xx' - \lambda I) = 0$ where $A - \lambda I$ is invertiable

By the property,

$$\det(I - xy') = 1 - y'x$$

$$\begin{aligned}\det(A + xx' - \lambda I) &= \det[(A-\lambda I)(I + (A-\lambda I)^{-1}xx')] \\ &= \det[(A-\lambda I)]\det(I + (A-\lambda I)^{-1}xx') \\ &= \det[(A-\lambda I)](1 + x'(A-\lambda I)^{-1}x)\end{aligned}$$

Since $\det(A-\lambda I)$ cannot be 0, $(1 + x'(A-\lambda I)^{-1}x) = 0$

$$x'(A-\lambda I)^{-1}x = -1$$

let $x \in F^n$

$$x'(A-\lambda I)^{-1}x = [x_1 \ x_2 \cdots x_n]\begin{bmatrix} \frac{1}{\lambda - A_{11}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda - A_{22}} & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \frac{1}{\lambda - A_{nn}} \end{bmatrix}\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = -1$$

$\therefore \frac{x_1^2}{\lambda - A_{11}} + \frac{x_2^2}{\lambda - A_{22}} + \cdots + \frac{x_n^2}{\lambda - A_{nn}} = 1$ where $A_{11}, A_{22} \cdots A_{nn}$ are the eigenvalues of

(b) $A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ and $x = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$

$\therefore A_{11} = 1$ and $A_{22} = 2$, $A_{33} = 3$ are eigenvalues of A

$$\therefore \frac{1}{\lambda - 1} + \frac{1}{\lambda - 2} + \frac{1}{\lambda - 3} = 1$$

$$\frac{(\lambda-2)(\lambda-3) + (\lambda-1)(\lambda-3) + (\lambda-1)(\lambda-2)}{(\lambda-1)(\lambda-2)(\lambda-3)} = 0$$

$$\frac{2\lambda^2 - 9\lambda + 9 + \lambda^2 - 3\lambda + 2}{(\lambda-1)(\lambda-2)(\lambda-3)} = 0$$

$$\therefore \lambda^3 - 9\lambda^2 + 23\lambda - 17 = 0$$

$\therefore \lambda_1 = 1.32$, $\lambda_2 = 2.46$, $\lambda_3 = 5.21$ by using Julia's Polynomials.jl

P3:

q_1 and q_2 are two orthonormal vectors, and b some fixed vector in R^n

(a) let $A = [q_1 \quad q_2]$ and $x = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$

$Ax = \alpha q_1 + \beta q_2$

By formula, the linear least squares estimate that minimize $||Ax-b||_2$ where $x = A^T b$

SVD of A:

$$A = \sum_{i=1}^{2} 1 q_i e_i^T \quad \text{where } e_i \text{ is the unit vector}$$

$$\therefore A^T = \sum_{i=1}^{2} e_i q_i^T \cdot 1$$

Therefore, $\hat{x} = \sum_{i=1}^{2} e_i q_i^T b = \begin{bmatrix} e_1 q_1^T b \\ e_2 q_2^T b \end{bmatrix}$

(b) $r = b - \alpha q_1 - \beta q_2 = b - A\hat{x}$

$= b - AA^+ b$

$= b(I - AA^+)$

$A'r = A'(I - A \cdot A^+) b$

$= [A' - A'AA^+] b$

$= [A'AA^+ - A'AA^+] b$ since $A' = A'AA^+$

$= 0$

$A'(b - A\hat{x}) = 0$ implies that:

① $q_1' r$, r is orthogonal to q_1

② $q_2' r$, r is orthogonal to q_2

P4. $A \in F^{m \times n}$, $b \in F^m$ and $x \in F^n$.

A^+ is Moore-Penrose pseudo-inverse of A. By the def of orthogonal vectors,

$$
\begin{aligned}
\left((I - A^+A)x\right)' \cdot (A^+b) &= x'(I - A^+A)'(A^+b) \\
&= x'(I - (A^+A)')A^+b \\
&= x'(I - (A^+A))A^+b \qquad \text{(by properties of Moore-Penrose inverse)} \\
&= x'(A^+ - A^+AA^+)b \\
&= x'(A^+ - A^+)b \qquad \text{(By properties of Moore-Penrose inverse)} \\
&= 0
\end{aligned}
$$

Therefore, $(I - A^+A)x$ and A^+b are orthogonal vectors

P5.

1.(a). $f(x) = \| Ax - b \|$

$\forall \alpha \in [0,1]$, $x_1, x_2 \in V$,

$$f(\alpha x_1 + (1-\alpha) x_2) \leq \alpha f(x_1) + (1-\alpha) f(x_2)$$

$$\| A(\alpha x_1 + (1-\alpha) x_2) - b \|_2 \leq \alpha \| A x_1 - b \|_2 + (1-\alpha) \| A x_2 - b \|_2$$ (This is what we need to prove)

$$\| A(\alpha x_1 + (1-\alpha) x_2) - b \|_2 = \| \alpha A x_1 + (1-\alpha) A x_2 - b \|_2$$
$$= \| \alpha A x_1 + (1-\alpha) A x_2 - b(\alpha + 1 - \alpha) \|_2.$$
$$= \| \alpha (A x_1 - b) + (1-\alpha)(A x_2 - b) \|_2$$
$$\leq \| \alpha (A x_1 - b) \|_2 + \| (1-\alpha)(A x_2 - b) \|_2$$ (By the triangle inequality)
$$= \alpha \| A x_1 - b \|_2 + (1-\alpha) \| A x_2 - b \|_2$$

Therefore, $f(x) = \| Ax - b \|$ is a convex function

(b) prove that $\sigma_1(X)$ is a Convex function of the elements of the $M \times N$ matrix X

Suppose there are two matrix $A, B \in F^{M \times N}$,

$$\sigma_1(\alpha A + (1-\alpha) B) \leq \alpha \sigma_1(A) + (1-\alpha) \sigma_1(B)$$ (this is what we need to prove)

$$\sigma_1(\alpha A + (1-\alpha) B) = \max_{\|u\|_2 = 1} \| (\alpha A + (1-\alpha) B) u \|_2$$ (By hint)

$$= \max_{\|u\|_2 = 1} \| \alpha A u + (1-\alpha) B u \|_2$$

$$\leq \max_{\|u\|_2 = 1} \| \alpha A u \| + \max_{\|u\|_2 = 1} \| (1-\alpha) B u \|_2$$ (By the triangle inequality)

$$= \alpha \max_{\|u\|_2 = 1} \| A u \| + (1-\alpha) \max_{\|u\|_2 = 1} \| B u \|_2$$

$$= \alpha \sigma_1(A) + (1-\alpha) \sigma_1(B)$$ (By hint)

$\therefore$ $\sigma_1(X)$ is a convex function

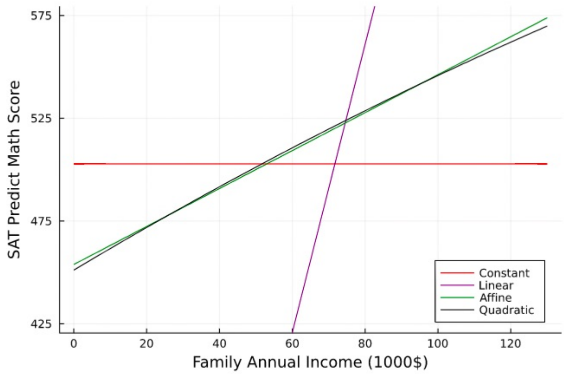
P6:

(a)

In [93]:

```julia
A, b,x, newmath1= constant_model(income, math)
A, b,x, newmath2= linear_model(income, math)
A, b,x, newmath3= affine_model(income, math)
A, b,x, newmath4= quadratic_model(income, math)
newincome = [0:1:130;]
plot(newincome, newmath1, label = "Constant", color = "red", ylim=[425,575], ytick=425:50:575,xlims=[0:130])
plot!(newincome, newmath2, label = "Linear", color = "purple")
plot!(newincome, newmath3, label = "Affine", color = "green")
plot!(newincome, newmath4, label = "Quadratic", color = "black")
plot!(xlabel = "Family Annual Income (1000\$)")
plot!(ylabel = "SAT Predict Math Score", legend = :bottomright)
```

Out[93]:

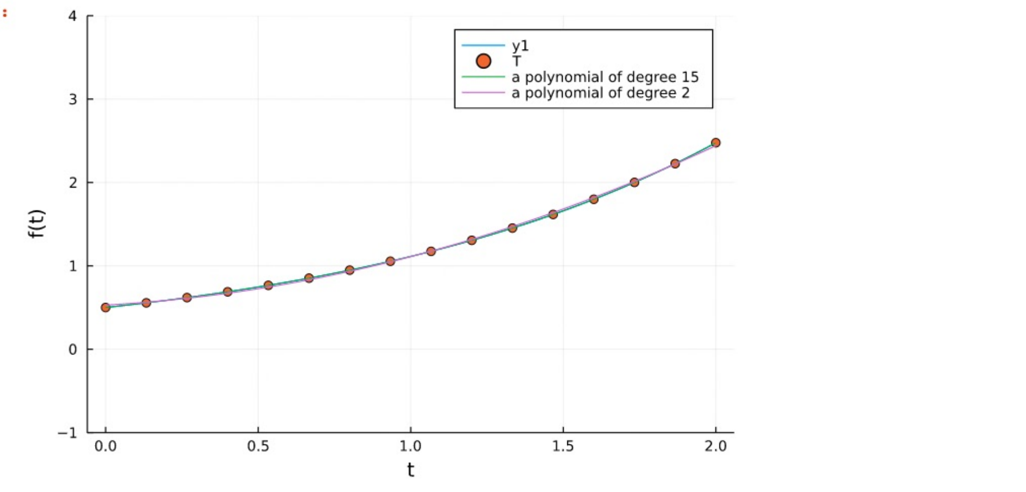


(b)

Constant Fit	Linear Fit	Affine Fit	Quadratic Fit
502.80000000000007	0	453.85977212971073	451.11444428602084
0	7.013164556962026	0.923400525854515	1.0584447335413572
0	0	0	-0.0011169658642307817

P7:

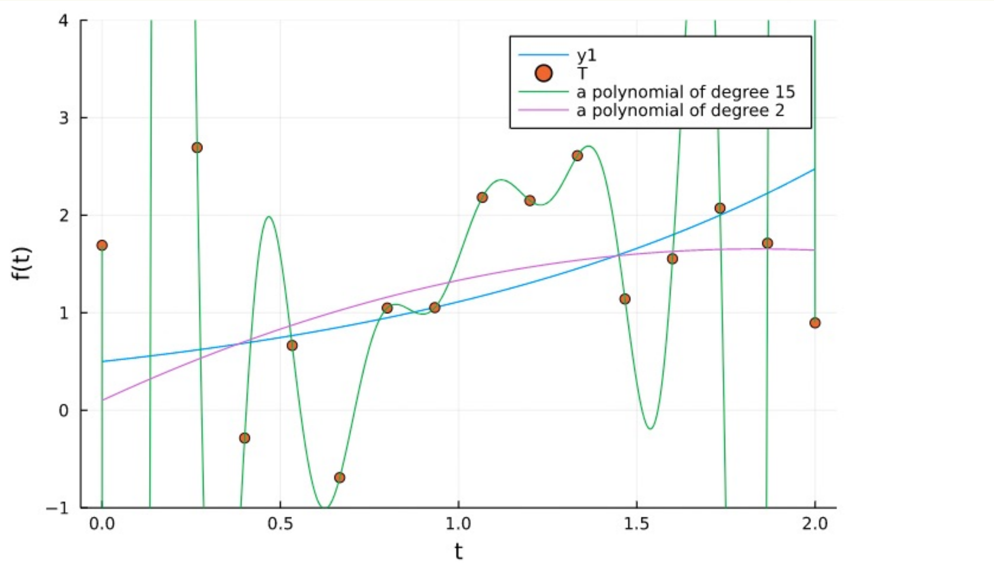
(a)




```
using LinearAlgebra
using Plots
using Random: seed!
seed!(3); e = randn(length(T))
```

As the plot shown, the coefficients of the least square error polynomial approximation of the data b for degree = 15 and degree = 2 are the same.

(b)



When noises are added, the least square error polynomial approximation pass through each sample data.

(c)

polynomial degree:	d = 2	d = 15
Residual norm $\|\|Ax(b) - b\|\|_2$ noiseless (a)	3502105199189893	0,0001622183167221819
Residual norm $\|\|Ax(y) - y\|\|_2$ noisy (b)	6.191069208724523	1047. 713599472052
Fitting error $\|\|Ax(b) - b\|\|_2$	4.641460505926311	4.641460505926311

(d)

Since $A_{15} = [A_2 \quad A_{3:15}]$. it implies that $R(A_2) \subseteq R(A_{15})$

$\therefore$ the residual norm for degree 2 is smaller than that degree of 15