COT 5615 Math for Intelligent Systems Fall 2021 Homework #9

UFID: 96703101
Name: *Vyom Pathak*
Instructor: Professor Kejun Huang
Due Date: December 07, 2021

Problem 1

Solving regularized least-squares

Solution

We have $\hat{x} = (A^TA + \lambda I)^{-1}A^Tb$. using the kernel trick,

$$\hat{x} = (A^TA + \lambda I)^{-1}A^Tb = A^T(AA^T + \lambda I)^{-1}b$$

We can compute, $(AA^T + \lambda I)^{-1}b$ by computing the QR factorization of the (m+n)*m matrix. The other operations involve matrix-vector products and have order (at most) m*n flops, so we can use this method to compute $\hat{x}$ in around $2(m+n)*m^2$ flops.

Problem 2

Companion matrix

Solution

a The characteristic polynomial of a companion matrix can be found using the following equation: $det(xI - C)$. where C is the companion matrix.

$$\begin{aligned}
det(xI - C) &= det((x, -1, 0, \ldots, 0; 0, x, -1, 0, \ldots, 0; \ldots; 0, 0, \ldots, 1; c_0, c_1, c_2, \ldots, x + c_{n-1})) \\
&= x \cdot det((x, -1, 0, \ldots, 0; 0, x, -1, 0, \ldots 0; \ldots; 0, 0, \ldots, 1; c_1, c_2, \ldots, x + c_{n-1})) \\
&\quad + (-1)^{n+1} \cdot c_0 \cdot det(-1, 0, \ldots, 0; x, -1, 0, \ldots, 0; \ldots, 0, 0, \ldots, 0; 0, 0, \ldots, -1)
\end{aligned}$$

By induction we can replace the determinant on the left by $c_1 + c_2x + c_3x^2 + \ldots + c_{n-1}x^{n-2} + x^{n-1}$ and the right matrix's determinant is the product of its diagonals (since it's upper-triangular). The product of the diagonal is $(-1)^{n-1}$. Therefore, the determinant is $c_0 + c_1x + \ldots + c_{n-1}x^{n-1} + x^n$.

b Let's tackle the LHS first,

$$\begin{aligned}
VC &= (1, 1, \ldots, 1; \lambda_1, \lambda_2, \ldots, \lambda_n; \lambda_1^2, \ldots, \lambda_n^2; \ldots; \lambda_1^{n-1}, \lambda_2^{n-1} \ldots, \lambda_n^{n-1}) \cdot \\
&\quad (0, 1, \ldots, 0; 0, 0, 1, \ldots, 0; \ldots; 0, 0, 0, \ldots, 1; -c_0, -c_1, \ldots, -c_{n-1}) \\
&= (\lambda_1, \lambda_2, \ldots, \lambda_n; \lambda_1^2; \lambda_1^{n-1}, \lambda_2^{n-1} \ldots, \lambda_n^{n-1}; -(c_0 + c_1\lambda_1, \ldots, \lambda_1^{n-1}c_{n-1}), -(c_0 + c_1\lambda_2, \ldots, \lambda_2^{n-1}c_{n-1}), \ldots) \\
&= (\lambda_1, \lambda_2, \ldots, \lambda_n; \lambda_1^2; \lambda_1^{n-1}, \lambda_2^{n-1} \ldots, \lambda_n^{n-1}; \lambda_1^n, \lambda_2^n, \ldots, \lambda_n^n) \; [\because \lambda_i \textit{ are distinct root of the polynomial and} \\
&\quad \textit{we put the roots in the polynomial from part a and solve for it.}] \\
&= (\lambda_1, 0, 0, \ldots, 0; 0, \lambda_2, 0, \ldots, 0; \ldots; 0, 0, 0, \lambda_n) \cdot (1, 1, \ldots, 1; \lambda_1, \lambda_2, \ldots, \lambda_n; \lambda_1^2, \ldots, \lambda_n^2; \ldots; \lambda_1^{n-1}, \lambda_2^{n-1} \ldots, \lambda_n^{n-1}) \\
&= \wedge V = RHS
\end{aligned}$$

Problem 3

Circulant matrix

Solution

a We solve Fc as. follows:

$$Fc = \frac{1}{\sqrt{n}}(1,1,\ldots,1;1,\omega,\omega^2,\ldots,\omega^{n-1};\ldots,\omega^2,\omega^4,\ldots,\omega^{2(n-1)};\ldots;\ldots,\omega^{n-1},\omega^{2(n-1)},\ldots,\omega^{(n-1)(n-1)}).$$
$$(c_1,c_2,c_3,c_4,\ldots,c_n)$$
$$= (c_1+c_2+\ldots+c_n, c_1+\omega c_2+\ldots+\omega^{n-1}c_n,\ldots,c_1+\omega^{n-1}c_2+\ldots+\omega^{(n-1)(n-1)}c_n)$$

Further,$\wedge F$ can be written as follows:

$$\wedge F = \frac{1}{\sqrt{n}}(c_1+c_2+c_3+\ldots+c_n, c_1+\omega c_2+\ldots+\omega^{n-1}c_n,\ldots,c_1+\omega^{n-1}c_2+\ldots+\omega^{(n-1)(n-1)}c_n$$
$$;\ldots,\ldots,\ldots,\ldots;c_1+c_2+c_3+\ldots+c_n,\omega^{n-1}[c_1+\omega c_2+\ldots+\omega^{n-1}c_n],\ldots,$$
$$\omega^{(n-1)(n-1)}c_1+\omega^{(n-1)}c_2+\ldots+\omega^{(n-1)(n-1)}c_n)$$

We know that $\Omega = e^{\frac{2\pi i}{n}} \implies \omega\Omega = 1$ We use this to get $F^{-1} = F^*$ as follows:

$$F^{-1} = F^* = \frac{1}{\sqrt{n}}(1,1,\ldots,1;1,\Omega,\Omega^2,\ldots,\Omega^{n-1};\ldots,\Omega^2,\Omega^4,\ldots,\Omega^{2(n-1)};\ldots;\ldots,\Omega^{n-1},\Omega^{2(n-1)},\ldots,\Omega^{(n-1)(n-1)})$$

Thus, we use F^* and $\wedge F$ to calculate $F^* \wedge F$ and after applying the indentity of the nth root of unity and using the fact that sum of n roots of unity will be 0, which gives the final result as follows:

$$\begin{aligned} F^* \wedge F &= \frac{1}{n}(nc_1,nc_2,\ldots,nc_n;nc_2,nc_1,\ldots,nc_{n-1};\ldots;nc_n,nc_{n-1},\ldots,nc_1) \\ &= C \end{aligned}$$

Hence, $C = F^* \wedge F$.

b The solution of $Cx = b$ can be shown as follows:

$$\begin{aligned} Cx &= b \\ x &= (C^TC)^{-1}C^Tb \\ C^T &= (F^* \wedge F)^T = F^T \wedge^T (F^*)^T = F \wedge F^* \\ x &= (F \wedge F^* F^* \wedge F)^{-1} F \wedge F^{-1} b \\ x &= F^* \wedge^{-1} F F \wedge^{-1} F^* F \wedge F^* b \\ x &= F^* \wedge^{-1} Fb \end{aligned}$$

As we the matricies F and F^* are DFT and IDFT respectively, we can perform this operation using FFT in $nlog(n)$ time. Thus, the time complexity is $nlog(n)$.

Problem 4

General least-squares

Solution

a Let's start with tackling LHS:

$$\begin{aligned}
A^T A\hat{x} &= (V\Sigma^T U^T)(U\Sigma V^T)\hat{x} \\
&= (V\Sigma^T \Sigma V^T)(V\Sigma^{-1}U^T b) \\
&= (V\Sigma^T \Sigma\Sigma^{-1}U^T b) \\
&= (V\Sigma^T U^T b) \\
&= (U\Sigma V^T)^T b \\
&= A^T b \\
&= RHS
\end{aligned}$$

b Verification of $\hat{x} = V\Sigma^{-1}U^T b$ can be shown as follows using $\|A\hat{x} - b\|^2 < \|Ax - b\|^2$:

$$\begin{aligned}
\|Ax - b\|^2 &= \|(Ax - A\hat{x}) + (A\hat{x} - b)\|^2 \\
&= \|(Ax - A\hat{x})\|^2 + \|(A\hat{x} - b)\|^2 + 2(Ax - A\hat{x})^T(A\hat{x} - b)
\end{aligned}$$

For the last term, we have:

$$\begin{aligned}
2(Ax - A\hat{x})^T(A\hat{x} - b) &= 2(x - \hat{x})^T A^T(A\hat{x} - b) \\
&= 2(x - \hat{x})^T(A^T A\hat{x} - A^T b) \\
&= 2(x - \hat{x})^T(A\hat{x} - A^T b)\ [\because\ part\ \mathbf{(a)}] \\
&= 0
\end{aligned}$$

Thus using this in the above equation we get:

$$\|Ax - b\|^2 = \|(Ax - A\hat{x})\|^2 + \|(A\hat{x} - b)\|^2$$

This implies that $\|A\hat{x} - b\|^2 \leq \|Ax - b\|^2$ which means that $\hat{x}$ minimizes $\|Ax - b\|^2$. Now, the equality doesn;t hold true because A has linearly independent columns and thus $x = \hat{x}$ is the only solution. Thus, we have $\|A\hat{x} - b\|^2 < \|Ax - b\|^2$.

Problem 5

QRSVD

Solution

a The following Julia code calculates SVD and compares it with the QRSVD method.

```julia
using LinearAlgebra
A = rand(6,3);
USV = svd(A);
U_org = USV.U;
S_org = USV.S;
```

```
    V_org = USV.V;
    A = rand(6,3);
    N = 1000
    q0, r0 = qr(A)
    r0[1,1] = abs(r0[1,1])
    r0[2,2] = abs(r0[2,2])
    r0[3,3] = abs(r0[3,3])
    rarr = []
    qarr = []
    append!(rarr, [r0])
    append!(qarr, [q0])
    for k in  1:N
        q,r = qr(rarr[k]')
        append!(rarr, [r])
        append!(qarr, [q])
    end

    z = zeros(3,3)
    U = qarr[1]
    V = [qarr[2] z; z z]
    for k in 3:N+1
        if k%2 == 1
            U *= [qarr[k] z; z z]
        else
            V *= [qarr[k] z; z z]
        end
    end
    S = rarr[N+1];

    println("S value using SVD: ");
    println(S_org);
    D = Diagonal(S);
    S = [D[1,1], D[2,2], D[3,3]];
    println("S value using QRSVD: ");
    println(S);
    println();
    println("V value by SVD function: ");
    println(V_org);
    V = V[1:3,1:3];
    println("V value by QRSVD function: ");
    println(V);
    println();
    println("U value by SVD function: ");
    println(U_org);
    U = U[:,1:3];
    println("U value by QRSVD function: ");
    println(U);
    println();
```

The output of the code is shown in figure 1:

b This can solved as follows:

$$Q_1(R_1R_0) = (Q_1R_1)R_0 = R_0^TR_0[\because\ \textit{It is given that } R_k^T = Q_{k+1}R_{k+1}]$$

Now, Q_1 is orthonormal and both R_1, R_0 are upper triangular matrices. Thus, $Q_1(R_1R_0)$ is the QR decomposition of $R_0^TR_0$.

```
S value using SVD:
[2.6720057339543923, 0.8016463946495971, 0.49217309001116744]
S value using QRSVD:
[1.865710086122108, 0.6245068051272025, 0.3502893029168897]

V value by SVD function:
[-0.6412184511271937 0.7400161485770496 -0.20301477231776238; -0.5515593213119188 -0.6284056750360019 -0.5485331554760222;
-0.5334990281228054 -0.2397548903103684 0.8111081182951416]
V value by QRSVD function:
[0.7623617022509794 0.4245479050388829 -0.4884298427288046; -0.5430898885043743 0.009265828703725082 -0.8396234378712512;
-0.35193466431327197 0.9053580621751431 0.23763158735567055]

U value by SVD function:
[-0.2936025293437453 0.3876290637426372 0.6298246833064968; -0.29697489116194814 0.5083933050493497 -0.0009750428859318164;
-0.5027940000109853 0.05852296897377848 -0.6162617901277659; -0.3288850836648491 -0.21451851885827908 -0.25998138947395116;
-0.4210126763990929 -0.7208736386410062 0.38419230269711574; -0.5360826339187785 0.1489228679713926 0.0913634242640073]
U value by QRSVD function:
[-0.436600636653543 0.872088673234636 -0.06641570821262836; -0.4921347705637104 -0.057865422817530786 0.6441313944554258;
-0.4834289190905406 -0.28832332849628595 -0.23455903204478895; -0.56779873048189 -0.3386660512859727 -0.41491003919040476;
-0.10432550379552213 -0.19424091173263638 0.5558146871677854; -0.01416060676108409 -0.02362987554901347 0.211152678821105]
```

Figure 1: Comparison of SVD values and QRSVD values

c part 1 can be shown as follows:

$$
\begin{aligned}
(R_1 R_0)Q_1 &= R_1((Q_1 R_1)^T)Q_1 \\
&= R_1 R_1^T Q_1^T Q_1 \qquad\qquad = R_1 R_1^T \ [\because Q_1 \ is \ orthonormal]
\end{aligned}
$$

part 2 can be shown as follows:

$$
\begin{aligned}
R_1 R_1^T &= (Q_2 R_2)^T (Q_2 R_2) \\
&= R_2^T Q_2^T Q_2 R_2 \\
&= R_2^T R_2 \ [\because Q_2 \ is \ orthonormal]
\end{aligned}
$$

part 3 $B_0 = R_0^T R_0 = Q_1 R_1 R_0 = Q_1(R_1 R_0)$
part 3 $B_1 = R_2^T R_2 = R_1 R_1^T = (R_1 R_0)Q_1$
Now, we know that $eigen(XY) = eigen(YX)$, thus
$eig(B_0) = eig(Q_1(R_1 R_0)) = eig((R_1 R_0)Q_1) = eig(B_1)$.

d Considering the relation between B_k and B_{k+1} as follows:

$$
\begin{aligned}
B_k &= R_{2k}^T R_{2k} \\
&= Q_{2k+1} R_{2k+1} R_{2k+1}^T Q_{2k+1}^T \\
&= Q_{2k+1}(R_{2k+2}^T Q_{2k+2}^T)(Q_{2k+2} R_{2k+2}) Q_{2k+1}^T \\
&= Q_{2k+1} B_{k+1} Q_{2k+1}^T
\end{aligned}
$$

Now, for $B_0 = Q_1 B_1 Q_1^T = Q_1 Q_3 B_2 Q_3^T Q_1^T = \ldots = Q_1 Q_3 \ldots Q_{2k-1} B_{K+1} Q_{2k-1}^T \ldots Q_3^T Q_1^T$. For each B_i if we solve them one by one, we can see that it is actually the QR iteration of the general matrix as mentioned above. Hence, $B_k = R_{2k}^T R_{2k}$ are the QR iterations that are applied on the matrix $B_0 = R_0^T R_0$.