Numerical Linear Algebra Homework

- 1st Week (pdf file)

Student ID and name : R10221026, Wei Wang

Exercise 1 Let $A \in \mathbf{R}^{m\times m}$ be nonsingular and let $a \in \mathbf{R}^m$. Let $\hat{A}$ be the rank-one update of the matrix A which differs from A by replacing the ith row of A with a^T, i.e.,

$$\hat{A} = A - e_i \left(e_i^T A - a^T \right)$$

where e_i is the ith column in the identity matrix I.

(a) Show that

$$\hat{A}^{-1} = A^{-1} \left(I + \frac{1}{a^T A^{-1} e_i} e_i \left(e_i^T - a^T A^{-1} \right) \right)$$

if $a^T A^{-1} e_i \neq 0$.

(b) Write and implement an algorithm which to given A^{-1}, a, e_i, and compute $\hat{A}^{-1}$ from the formula above with a checks if $a^T A^{-1} e_i \neq 0$ holds; a sample driver routine in Matlab is attached below.

(c) What is $\hat{A}$ if it differs from A by replacing the ith column of A with a and what is the formula for $\hat{A}^{-1}$ in this case ?

The sample routine driver for (b) given in the homework description : (note : I have made a modification to the name of the function `inverse_updata` to `inverse_update`).

```
function nla2022_hw1_1(m,i)
% A sample driver routine in Matlab for HW 1-1
% m: matrix dimension
% i: index for row, 1 \le i \le m
%
% generate random matrix A and vector a
A = rand(m);
a = rand(m,1);
%
% compute inverse of A
C = inv(A);
%
% define rank-one modification \hat{A} from A directly
Ahat = A;
Ahat(i,:) = a';
%
```

```
% compute inverse of \hat{A}
inv_Ahat_1 = inv(Ahat);
%
% compute inverse of \hat{A} using inverse-update formula
inv_Ahat_2 = inverse_updata(C,a,i);
%
% compute maximum norm error between inv_hatA_1 and inv_hatA_2
err = max(max(abs(inv_Ahat_2-inv_Ahat_1)))
end
```

Proof.

(a) We may prove by brute-force. Let us write the right hand side of the supposed formula for $\hat{A}^{-1}$ as B. We have :

$$\hat{A}B = \hat{A}A^{-1}AB = \left(I + e_i(a^TA^{-1} - e_i^T)\right)\left(I - \frac{1}{a^TA^{-1}e_i}e_i(a^TA^{-1} - e_i^T)\right)$$

To simplify expressions, write $C = a^TA^{-1} - e_i^T$. We get :

$$\hat{A}B = I + e_iC - \frac{1}{a^TA^{-1}e_i}e_iC - \frac{1}{a^TA^{-1}e_i}e_iCe_iC$$

As $Ce_i = a^TA^{-1}e_i - 1$, we get :

$$e_iCe_iC = (a^TA^{-1}e_i - 1)e_iC$$

which gives $\hat{A}B = I$. Note that $\hat{A}B = B\hat{A}$, as both $\hat{A}, B$ take the form $I + ke_iC$; this shows that $\hat{A}, B$ are inverses to each other, and are both invertible. Now the interesting question is the derivation of this formula; knowing this will gain insight to (c), although (c) may be deduced directly from the formula of updating rows.
The idea is the following algebraic fact : suppose given a matrix M with $M^2 = kM$ for some constant $k \neq -1$ (so M is almost idempotent), we have

$$(I + M)(I + xM) = (I + xM)(I + M) = I + (1 + (k + 1))xM$$

so by choosing $x = -(k+1)$, we get an explicit expression of the inverse of $I+M$ along with the fact that $I + M$ is invertible. Notice that by specializing to the case $I + M = A^{-1}\hat{A}$ (so that M is actually e_iC), one recovers the formula for $\hat{A}^{-1}$; note also that the condition $k \neq -1$ precisely corresponds to the condition $a^TA^{-1}e_i \neq 0$.

(b) The following is the code for the function `inverse_update`. I have uploaded this file to COOL also.

```
function C_new = inverse_update(C,a,i)
    % <---Here the tolerance is set to 1e-18; this can be changed--->
    tol = 1e-18;
    % <---Define the vector e_i--->
    m = size(C,1);
    e_i = zeros(m,1);
    e_i(i,1) = 1;
    % <---Prevent division by small number--->
    val = a'*C*e_i;
    disp(val)
```

```
    if val < tol % <---Outputs error if intolerable--->
        error(strcat("Error in inverse_update formula : " + ...
            "division by small number." + ...
            " tolerance : ",num2str(tol)));
    end
    % <---If tolerable, update using the explicit formula--->
    C_new = C*(eye(m)+inv(val)*e_i*(e_i'-a'*C));
end
```

(c) There are two ways : we can derive this formula from scratch, or directly apply the results in (a). To update the i-th column to A to a is the same as updating the i-th row of A^{T} to a^{T}. Therefore we have :

$$\hat{\mathrm{A}}^{\mathrm{T}} = \mathrm{A}^{\mathrm{T}} - e_i(e_i^{\mathrm{T}}\mathrm{A}^{\mathrm{T}} - a^{\mathrm{T}})$$

$$\hat{\mathrm{A}}^{\mathrm{T}} = (\mathrm{A}^{\mathrm{T}})^{-1}e_i\left(\mathrm{I} + \frac{1}{a^{\mathrm{T}}(\mathrm{A}^{\mathrm{T}})^{-1}e_i}(e_i^{\mathrm{T}} - a^{\mathrm{T}}(\mathrm{A}^{\mathrm{T}})^{-1})\right)$$

and hence :

$$\hat{\mathrm{A}} = \mathrm{A} - (\mathrm{A}e_i - a)e_i^{\mathrm{T}}$$

$$\hat{\mathrm{A}}^{\mathrm{T}} = \left(\mathrm{I} + \frac{1}{a^{\mathrm{T}}(\mathrm{A}^{\mathrm{T}})^{-1}e_i}(e_i - \mathrm{A}^{-1}a)\right)e_i^{\mathrm{T}}\mathrm{A}^{-1}$$

Exercise 2 Suppose $x, y \in \mathbf{R}^m$ with $\|x\|_2 = \|y\|_2$ and $v = x - y \neq 0$.

(a) Show that

$$\mathrm{F}(v)x = y, \qquad \text{where } \mathrm{F}(v) = \mathrm{I} - 2\frac{vv^{\mathrm{T}}}{v^{\mathrm{T}}v}$$

(b) Suppose

$$x = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix}, \quad y = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix} = \|x\|_2 e_2.$$

Use to construct a unitary transformation $\mathrm{F}(v)$ such that $\mathrm{F}(v)x = y$.
In general, when $y = \|x\|_2 e_n$, we can find v so that $\mathrm{F}(v)x = \|x\|_2 e_n$ for any n.

Proof. We note that $\mathrm{F}(v)$ is linear.

(a) We have :

$$\mathrm{F}(v)x = x - 2\frac{v^{\mathrm{T}}x}{v^{\mathrm{T}}v}v$$

The vector $w := \frac{v^{\mathrm{T}}x}{v^{\mathrm{T}}v}v$ is x projected to v. The geometric proof of this formula is the diagram :

In this sense, $\mathrm{F}(v)$ is the linear transformation defined by reflection along the line spanned by $x + y$. Algebraically, write $x = \frac{1}{2}v + \frac{1}{2}(x + y)$. We only need to verify $\mathrm{F}(v)(x + y) = x + y$ (as $\mathrm{F}(v)v = -v$ and $y = -\frac{1}{2}v + \frac{1}{2}(x + y)$), which follows from :

$$v^{\mathrm{T}}(x + y) = \|x\|_2^2 - \|y\|_2^2 = 0$$

(b) We may use the formula given in (a). The fact that F is unitary (or equivalently orthogonal when the entries are all reals) follows from $F(v)^T = F(v)$ and $F(v)^2 = I$ (as $F(v)$ maps v to $-v$ and interchanges x, y). Explicitly, under the standard basis, $F(v)$ can be computed as follows :

$$v = x - y = (2, 2, 1)^T - (0, 3, 0)^T = (2, -1, 1)^T,$$

$$v^T v = 6, \ vv^T = \begin{bmatrix} 4 & -2 & 2 \\ -2 & 1 & 1 \\ 2 & -1 & 1 \end{bmatrix}$$

$$F(v) = I - \frac{2}{v^T v} vv^T = \frac{1}{3} \begin{bmatrix} -1 & 2 & -2 \\ 2 & -2 & -1 \\ -2 & 1 & 2 \end{bmatrix}$$

For the statement concerning the general situation, we note that when $x = y$, we may simply choose v to be any vector perpendicular to y. ■

Exercise 3 Consider the matrix

$$A = \begin{bmatrix} -2 & 11 \\ -10 & 5 \end{bmatrix}.$$

(a) Determine, on paper, a singular value decomposition (SVD) of A in the form $A = U\Sigma V^T$. The SVD is not unique, so find the one that has the minimal number of minus signs in U and V.

(b) Draw a careful, label picture of the unit ball in $\mathbf{R}^2$ and its image under A, together with the singular vectors, with the coordinates of their vertices marked.

(c) What are the 1-, 2-, ∞-, and Frobenius norms of A ?

(d) Find A^{-1} not directly, but via the SVD.

(e) Find the eigenvalues λ_1, λ_2 of A.

(f) Verify that $\det(A) = \lambda_1 \lambda_2$ and $|\det(A)| = \sigma_1 \sigma_2$. *

Proof.

(a) We compute $A^T A$ and find the singular values :

$$A^T A = \begin{bmatrix} 104 & -72 \\ -72 & 146 \end{bmatrix}$$

The eigenvalues are $\lambda_1 = 200, \lambda_2 = 50$ so the singular values of A are $\sigma_1 = 10\sqrt{2}, \sigma_2 = 5\sqrt{2}$. Corresponding to the same eigenvalues, we may take the unit eigenvectors to be $v_1 = [3/5, -4/5]^T$, $v_2 = [4/5, 3/5]^T$. This gives us the description of Σ, V; we now compute U. Note that :

$$u_1 = \frac{1}{\sigma_1} A v_1, \quad u_2 = \frac{1}{\sigma_2} A v_2$$

This gives :

$$u_1 = \frac{1}{\sqrt{2}} [-1, -1]^T, u_2 = \frac{1}{\sqrt{2}} [1, -1]^T$$

We get :

$$A = U\Sigma V^T = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} \begin{bmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{bmatrix}$$

If we make the change $v_1 \mapsto -v_1$, $u_1 \mapsto -u_1$, we get a decomposition with less minus signs in U and V :

$$A = U\Sigma V^T = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 10\sqrt{2} & 0 \\ 0 & 5\sqrt{2} \end{bmatrix} \begin{bmatrix} -3/5 & 4/5 \\ 4/5 & 3/5 \end{bmatrix}$$

This is an SVD with the least minus signs.

(b) Geometrically, A maps v_i to $\sigma_i u_i$.

(c) By definition, the Frobenius norm of A is given by :

$$\|A\|_F = \left(\sum_{i,j=1}^{2} |a_{i,j}|^2 \right)^{1/2} = (4 + 121 + 100 + 25)^{1/2} = 5\sqrt{10}$$

The ∞-norm of A is by definition given by :

$$\|A\|_\infty = \max_{i=1,2} \sum_{j=1}^{2} |a_{ij}| = \max\{13, 15\} = 15$$

The 1-norm is given by :

$$\|A\|_1 = \max_{j=1,2} \sum_{i=1}^{2} |a_{ij}| = \max\{12, 16\} = 16$$

By (b), the 2-norm of A is the largest singular value of A, which is $10\sqrt{2}$.

(d) We have $A^{-1} = V\Sigma^{-1}U^T$, so we get :

$$A^{-1} = \left(\frac{1}{5} \begin{bmatrix} -3 & 4 \\ 4 & 3 \end{bmatrix} \right) \left(\frac{1}{10\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix} \right) \left(\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right) = \frac{1}{100} \begin{bmatrix} 5 & -11 \\ 10 & -2 \end{bmatrix}$$

(e) We have $\det(A - \lambda I) = (\lambda + 2)(\lambda - 5) + 110 = \lambda^2 - 3\lambda + 100$, so the eigenvalues are :

$$\frac{3 \pm \sqrt{-391}}{2}$$

(f) We have $\lambda_1\lambda_2 = (9 - (-391))/4$, while $\det(A)$ is obtained by setting $\lambda = 0$ in $\det(A - \lambda I)$, which gives $\det(A) = 2 * (-5) + 110 = 100$. On the other hand, we have $\sigma_1\sigma_2 = 10\sqrt{2} * 5\sqrt{2} = 100$. ■

Exercise 4 Given the statement:

If $A \in \mathbf{R}^{m\times m}$ has singular values $(\sigma_1, \sigma_2, \ldots, \sigma_m)$, then A^2 has singular values $\left(\sigma_1^2, \sigma_2^2, \ldots, \sigma_m^2\right)$.

(a) Find a class of matrices for which the statement is true.

(b) Show that the statement is not true in general. *

Proof. For (a), the class of symmetric matrices has this property, as if $A = U\Sigma V^T$, then $A^2 = AA^T = U\Sigma^2 V^T$. For (b), a counterexample can be given as follows: consider the matrix A given by

$$e_1 e_2^T = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

We have $A^2 = 0$, $A^TA = e_1e_1^T$, so the singular values of A are 0, 1, while those of A^2 are all 0. ■

Exercise 5 Let A be the matrix

$$\mathrm{A} = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}$$

(a) Find the spaces $\mathrm{C}(\mathrm{A}^\mathrm{T})$ and $\mathrm{N}(\mathrm{A})$.

(b) Consider the underdetermined linear system

$$\begin{bmatrix} 1 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 12 \end{bmatrix}.$$

Find the solution $x \in \mathbf{R}^3$ that minimizes $\|\mathrm{A}x\|_2$. *

Proof.

(a) For $\mathrm{C}(\mathrm{A}^\mathrm{T})$, it is the space spanned by the row vectors; it is the set of vectors of the form :

$$\{[a+b, b, -a+b]^\mathrm{T} : a, b \in \mathbb{R}^2\}$$

For $\mathrm{N}(\mathrm{A})$, it is the space spanned by the single vector $[1, -2, 1]^\mathrm{T}$ (more details : it contains this vector, while evidently $\mathrm{C}(\mathrm{A})$ is of dimension 2).

(b) We try to find a pseudo-inverse of A via SVD. In fact, we have :

$$\mathrm{AA}^\mathrm{T} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

so the singular values of A are $\sqrt{3}, \sqrt{2}, 0$; this determines Σ. To find V, consider the matrix

$$\mathrm{A}^\mathrm{T}\mathrm{A} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

We can pick the unit eigenvectors corresponding to 3, 2, 0 to be :

$$v_1 = \frac{1}{\sqrt{3}}[1,1,1]^\mathrm{T},\ v_2 = \frac{1}{\sqrt{2}}[1,0,-1]^\mathrm{T},\ v_3 = \frac{1}{\sqrt{6}}[1,-2,1]^\mathrm{T}$$

Correspondingly, we may compute U by $u_1 = \sigma_1^{-1}\mathrm{A}v_1$, $u_2 = \sigma_2^{-1}\mathrm{A}v_2$; we get :

$$u_1 = [0,1]^\mathrm{T},\ u_2 = [1,0]^\mathrm{T}$$

Therefore, the pseudo-inverse is then given by:

$$\mathrm{A}^+ = [v_1, v_2]\ \mathrm{diag}((\sigma_1^{-1}, \sigma_2^{-1}))[u_1, u_2]^\mathrm{T}$$

so such x is given by :

$$\mathrm{A}^+ b = [v_1, v_2]\ \mathrm{diag}((\sigma_1^{-1}, \sigma_2^{-1}))[u_1, u_2]^\mathrm{T}[4,12]^\mathrm{T} = [v_1, v_2]\ \mathrm{diag}((\sigma_1^{-1}, \sigma_2^{-1}))[12,4]^\mathrm{T}$$

$$= [v_1, v_2][4\sqrt{3}, 2\sqrt{2}]^\mathrm{T} = 4\sqrt{3}v_1 + 2\sqrt{2}v_2 = [4,4,4]^\mathrm{T} + [2,0,-2]^\mathrm{T} = [6,4,2]^\mathrm{T}$$ ■

Exercise 6 The routine `hello.m` on the NTUcool website constructs a 15×40 matrix A that is zero everywhere except for ones in the positions marked in the figure below. The upper-left point of the "H" is in the $(2, 2)$ entry, and the bottom-right point of the "O" is in the $(13, 39)$ entry. This figure was produced with the command `spy(A)`.

(a) Call Matlab's `svd` to compute singular values of A, and print the results. Plot these numbers using both `plot` and `semilogy`. What is the mathematically exact rank of A ? How does this show up in the computed singular values ?

(b) For each k from 1 to rank(A), construct the rank-k matrix A_k that is the best approximation to A in 2-norm. Visualize these matrices using the commands:

imagesc(Ak)
colormap(flipud(gray))

Submit print-outs of each of these images. $*$

Proof. I have included the Matlab code in my uploaded file `hw_1_6.m`, which contains the following content (note : to run this code, one has to supply the file `Hello_Mat.mat` which contains the output result of `hello.m`; I also uploaded this file `Hello_Mat.mat` to COOL) :

```
A = importdata("Hello_Mat.mat");
%<-- A peak at the singular values of S -->
v = svd(A);
display(v);
figure
plot(v);
figure;
semilogy(v); % The mathematical rank of A is 12
%<-- Get U, V for the SVD of A -->
figure;
[U, S, V] = svd(A);
%<-- Approximations of A -->
[x1, x2] = size(A);
A_k = zeros(x1, x2, 12); % 12 is the rank computed above
S_k = zeros(x1, x2);
str = 'Graph_for_A_';
for i = 1: 12
    S_k(i, i) = v(i);
    A_k(:, :, i) = U * S_k * V';
    % the following saves the results of the picture
    imagesc(A_k(:, :, i));
    colormap(flipud(gray));
    saveas(gcf, strcat(str, int2str(i), '.png'));
end
```

After running the code, I got the following plots :

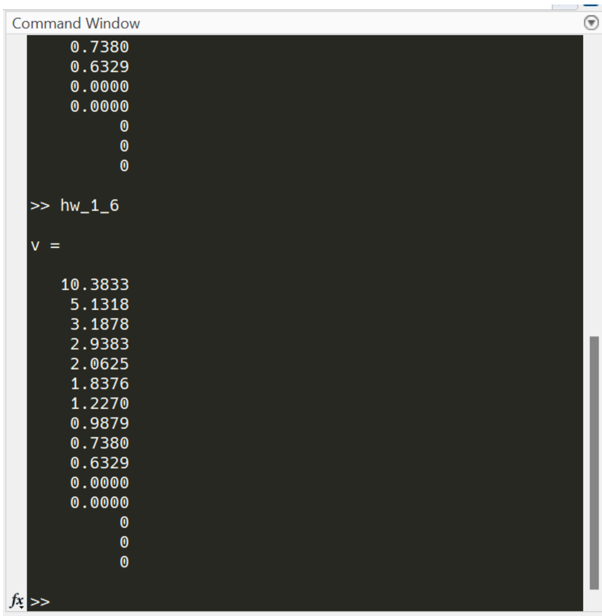

Figure 1: Screenshot of the output terminal (here `v` is the vector of singular values given by the Matlab function `svd`); the smallest 2 positive singular values all looked like `0.0000`, but after using `semilogy` (see the figure below), their values can be seen to be non-zero.

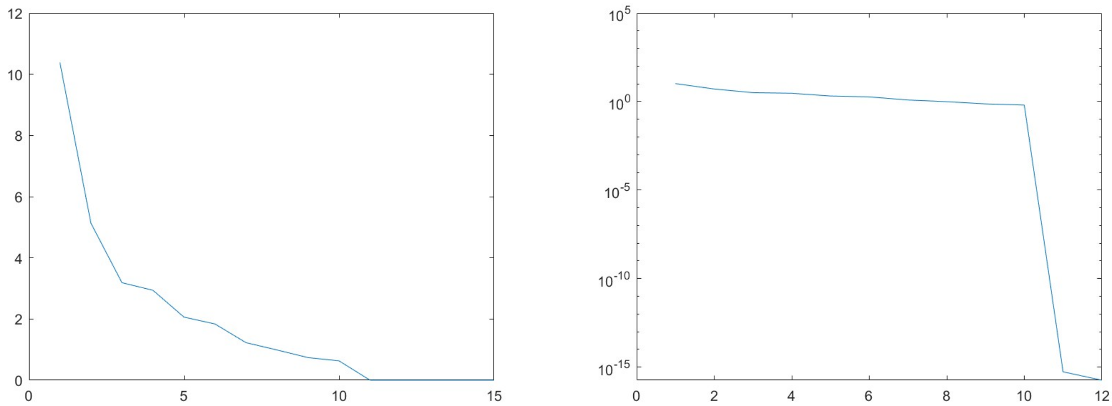

Figure 2: The plots given by `plot` and `semilogy`.

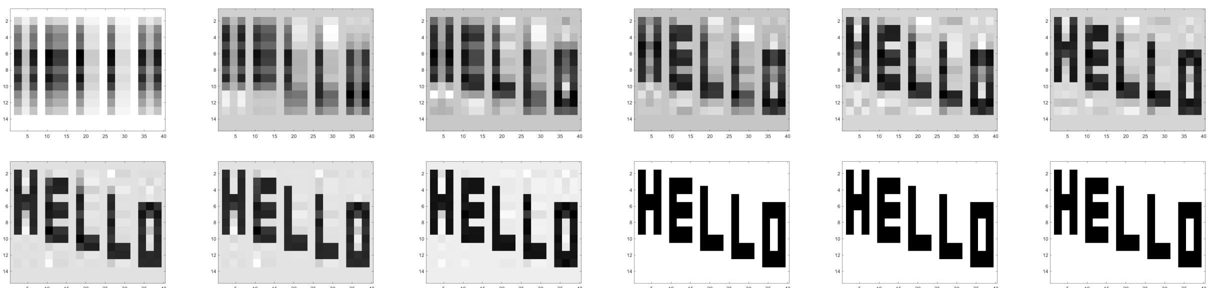

Figure 3: The visualization of the matrices A_1 to A_{12}, given by the Matlab commands `imagesc` and `colormap(flipud(gray))`. These photos are also submitted in the uploaded file with filename `hello_approximations.zip`.