APM120, quiz #2, 2018. **Solutions**
Applied Linear Algebra and Big Data
Last updated: Friday 5th April, 2019, 17:59

your name: _____, HUID: _____

Read these instructions carefully: Please solve all problems, **deriving, calculating and showing explicitly all stages of your solution and explaining each step of each question**. The number of points for each question is noted below, the total number of points is 115 and the final score is min(100, your points). Use the space below and attach additional pages if needed. Use a non-programmable calculator to convert your answers to decimal number format, carrying out calculations to four significant digits. Whenever relevant, use the numerical checks provided to verify your solution. Limit your essay responses to no more than the specified number of words, longer responses would be truncated.

Start time: 7:00, end time: 9:00.

Good luck!

1. (34 pts) Consider $A\mathbf{x} = \mathbf{b}$ for A and $\mathbf{b}$ given below.

   (a) Do you expect to find a solution that satisfies all of these equations? Explain. If no solution exists, what would you need to assume to find a value of $\mathbf{x}$ that can be helpful in applications? Write down an expression for the general solution for $\mathbf{x}$ in such a case.

   (b) Solve for $\mathbf{x}$ and for the value of the residual $\mathbf{r} = A\mathbf{x} - \mathbf{b}$. *Numerical check:* $r_3 = 0.32$.

   (c) In no more than 120 words: How is the norm of a general matrix defined, and how is it calculated? How is the condition number of a general matrix A calculated? What is the condition number used for in the context of linear equations?

   (d) Estimate the condition number. Suppose the specified relative error in the right hand side $\mathbf{b}$ is specified to be 15%, and that the solution $\mathbf{x}$ is needed with a relative error of no more than 20%. Estimate effective rank of A in that case. Explain in no more than 40 words without solving explicitly: How would you solve for $\mathbf{x}$ in this case?

```
A=[ 2   1
    2  -1
    1  -2 ];
b=[ 4
    0
   -4 ];
[U,Sigma,V]=svd(A);
Sigma=[ 3.16228  0
        0        2.23607
        0        0 ];
```

Solution:

(a) more equations than unknowns, so likely no solution. Need to define a desired solution as that that minimizes the residual $\|r\|^2 = \|Ax - b\|^2$. Solution was shown in class to be given by $x = (A^T A)^{-1} A^T b$

(b) solving:

```
ATA=A'*A;
ATAinverse=inv(ATA);
ATAinverseAT=ATAinverse*A';
x=ATAinverse*A'*b;
r=A*x-b;
my_fprintf_array(ATA);my_fprintf_array(ATAinverse);my_fprintf_array(ATAinverseA
my_fprintf_array(x);my_fprintf_array(r);
ATA=[ 9  -2
      -2  6 ];
```

```
ATAinverse=[ 0.12  0.04
             0.04  0.18 ];
ATAinverseAT=[ 0.28  0.2  0.04
               0.26  -0.1  -0.32 ];
x=[ 0.96
    2.32 ];
r=[ 0.24
    -0.4
    0.32 ];
```

(c) the norm of A is $\max(\|Ax\|/\|x\|)$ over all non zero vectors x. It is equal to the largest singular value of A. The condition number gives the worse possible ratio of the relative error in the rhs to that in the solution: $\frac{\|\delta b\|}{\|b\|} = cond \frac{\|\delta x\|}{\|x\|}$.

(d) given sigma above, the condition number is 3.16228/2.23607=1.4142. A 15% error in rhs will be amplified to 1.4142*15=21.2130% which is unacceptable for this problem. We therefore need to treat A as rank deficient with rank=1 and solve using $b = A^{\dagger}b$ using a single nonzero singular value in $\Sigma^{\dagger}$

2. (34 pts) Consider the following two data sets, one representing aerosol (smoke) emissions from two active volcanoes (X) and the second of surface air temperature over the northern hemisphere and over the southern hemisphere (Y) during $N = 5$ years (the time average has been subtracted off, and you do not need to normalize by the standard deviation in this case). Analyze the relation between the two by addressing the following points. Start by *carefully* considering all the following information:

```
X=[ -0.2266  1.3866 -1.3847 -0.1034  0.3282
    -0.1063  0.6736 -0.7061 -0.0660  0.2049];
Y=[ -2.5583 -0.0592  0.5291  1.0206  1.0678
     1.7288 -0.0264 -0.3987 -0.7281 -0.5757];
mean_X=mean(X');mean_Y=mean(Y');
mean_X=[ -0.0000  0.0000];
mean_Y=[  0.0000  0.0000];

N=length(X(1,:)); disp(N);
     5
F=[X;Y];C1=F*F'/N;
C1=[  0.8020  0.4020  0.0020  0.0020
      0.4020  0.2020  0.0020  0.0020
      0.0020  0.0020  1.8020 -1.1980
      0.0020  0.0020 -1.1980  0.8020];
[U1,S1,V1]=svd(F);
U1=[  0.0004 -0.8939 -0.1408 -0.4256
      0.0003 -0.4483  0.2728  0.8513
      0.8322 -0.0017  0.5277 -0.1703
     -0.5545 -0.0034  0.7921 -0.2554];
S1=[  3.6057  0.0000  0.0000  0.0000  0.0000
      0.0000  2.2401  0.0000  0.0000  0.0000
      0.0000  0.0000  0.1454  0.0000  0.0000
      0.0000  0.0000  0.0000  0.0000  0.0000];
V1=[ -0.8564  0.1110  0.1533 -0.1035  0.4692
     -0.0094 -0.6880 -0.4369  0.4337  0.3841
      0.1832  0.6941 -0.2356  0.5427  0.3669
      0.3475  0.0548 -0.2860 -0.6916  0.5622
      0.3350 -0.1719  0.8052  0.1682  0.4262];

C2=X*Y'/N;[U2,S2,V2]=svd(C2);
U2=[ -0.7071 -0.7071
     -0.7071  0.7071];
S2=[  0.0040  0.0000
      0.0000  0.0000];
V2=[ -0.7071  0.7071
```

```
-0.7071 -0.7071];
```

(a) Given the combined data matrix, $F = \binom{X}{Y}$, what does $C_1 = FF^T/N =$ `F*F'/N;` represent? Interpret all of its **diagonal** values and the elements on its **first row**.

(b) What does the matrix $C_2 = XY^T/N =$ `X*Y'/N;` represent? Write it based on the values C_1 (or calculate it explicitly from the data if you do not know how to do this), and Interpret each of its elements.

(c) what are the singular values (S1) and singular vectors (U1 and V1) of F used for? Calculate the fraction of variance explained by each of the principal components. Interpret the structure of all the singular U1 vectors.

(d) Interpret the singular values and singular vectors of C_2. Calculate the fraction of total covariance explained by each of the MCA modes.

(e) Based on your above results and the above data, discuss in no more than 40 words the amplitude of the co-varying modes vs the amplitude of the four variables themselves.

(f) In no more than 120 words, explain the difference between Maximum Covariance Analysis (MCA) and Multivariate Principal Component Analysis (multivariate PCA), and the advantages of each. Explain how the advantage of MCA is expressed in the above specific example.

(g) In no more than 120 words, explain the difference between PCA analysis based on the covariance matrix vs the one based directly on SVD of the data matrix. Why are the principal components guaranteed to be the same in both approaches? Discuss the advantages and disadvantages of each.

Solution:

(a) $C_1 = FF^T/N$ represents the covariance matrix of all variables. The diagonal elements indicate that the first Y variable has the strongest variability. The first X and second Y variables follow, and the second X variable has a weaker variability. The first row indicates that the first X variable is strongly correlated with the second X variable, but only weakly with the two Y variables.

(b) $C_2 = XY^T/N$ represents the covariance of the X and Y data sets. It is given by the block [0.0020, 0.0020; 0.0020, 0.0020] that appears in the top right part of C_1, as this block shows the correlation between X and Y variables rather than X with X or Y with Y variables. The values indicate that both X variables are correlated with both Y variables to the same degree. So there is a small component of these that varies together.

(c) The singular values of F can be used to calculate the fraction of variance explained by each of the principal components using $\sigma_i^2/(\sum_j \sigma_j^2)$. The singular U vectors are the PCs and the singular V vectors can be used to calculate the time series that

are given by ΣV^T. In this particular example, U1(:,1) represents the variability of the Y vectors that are anti-correlated to one another. U1(:,2) represents the variability of the X variables that are correlated with one another. U1(:,3) represents some unclear mix of variables, and because S1 has only 3 nonzero singular values, U1(:,4) does not matter.

(d) The singular values (S2) show that there is a single co-variability mode between X and Y, with a single time series. The singular vector U2(:,1) shows that in this co-variability mode, the two X variables vary together. V2(:,1) shows that the Y variables also vary together in this co-variability mode. The fraction of total covariance explained by the first MCA mode is 100%.

(e) The small amplitude of $C2$ relative to the variance of both X and Y (given by C1(1:2,1:2) and C1(3:4,3:4), correspondingly) indicates that the co-variability mode has a smaller amplitude than the separate variability of X and Y.

(f) Maximum Covariance Analysis (MCA) provides information about the co-variability mode, while Multivariate Principal Component Analysis (MPCA) can also provide information about the variability of X and Y separately. MPCA often biases the structure of co-variability mode because the PCs must be orthogonal and the co-variability modes are not necessarily orthogonal to the variability of X and Y by themselves. In this particular example we know that the co-variability mode involves both X and Y variables varying together at the same amplitude. But the third MPCA mode U1(:,3) shows both positive and negative amplitudes, while the covariance terms of XY^T/N show that X and Y should co-vary, and we would therefore expect all elements of the MPCA mode describing the co-variability to have the same sign.

(g) The principal components are the U singular vectors of the data matrix F. This, based on the way the SVD is calculated, means that they are the eigenvectors of FF^T. In the covariance method, one calculates the eigenvectors of FF^T/N. Because of the normalization of the eigenvectors, these are guaranteed to be the same in both approaches. PCA analysis based on the covariance matrix has the advantage that the eigenvectors of a relatively small matrix need to be calculated, assuming that the data is composed of more time steps than variables (say stocks). PCA based on SVD does not require the construction of FF^T and therefore suffers less round-off errors.

3. (34 pts)

(a) In no more than 120 words: Describe how a gray-scale image can be compressed using SVD. How would you choose the right number of singular values to include? what is the compression ratio when including k modes for an n by m gray-scale image?

(b) Approximate the matrix A using one singular value and then using two singular values. Calculate the residual, original matrix minus the approximate one, for both approximations. Calculate the compression ratio for each of the two approximations. *Numerical check:* the middle value of the second singular V vector is `V(2,2)=0.967;`

```
A=[ 1  0  1
    1  0  1
    1  1  1 ];
[U,Sigma,V]=svd(A);
```

```
U=[ -0.542  -0.454  -0.707
    -0.542  -0.454   0.707
    -0.643   0.766  -0 ];
Sigma=[ 2.524  0      0
        0      0.792  0
        0      0      0 ];
```

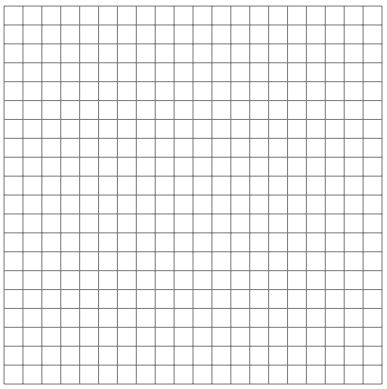
(c) Consider the transformation by the matrix B of the geometric shape whose corners are given by the columns of X. Derive the polar decomposition $B = SQ$ with Q being a rotation matrix. *Numerical check:* `U(:,2)=[0,1]';`

(d) Plot the original shape X as well as QX and BX. Explain the transformations represented by Q, S and B. *Hint:* a 2d rotation by an angle θ is given by the rotation matrix $P = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$.

```
X=[ -2  2  2  -2
     1  1  -1  -1 ];
B=[ 1.41  -1.41
    0.354  0.354 ];
[U,Sigma,V]=svd(B)
```

```
Sigma=[ 2  0
        0  0.5 ];
V=[ -0.707  0.707
    0.707  0.707 ];
```

polar decomposition



Solution:

(a) Writing the image as a sum over SVD modes, $A = \sum_i \mathbf{u}_i \sigma_i \mathbf{v}_i^T$, we can terminate the sum at any point to obtain a compressed image. Each SVD mode requires $n + m + 1$ for the u,v vectors and singular value. With k modes the compression ratio would be $k * (n + m + 1)/(n * m)$.

(b)
```
A=[1,0,1;1,0,1;1,1,1];
[n,m]=size(A);
[U,Sigma,V]=svd(A);
U=round(1000*U)/1000; Sigma=round(1000*Sigma)/1000; my_fprintf_array(A);my_fpri
A=[ 1  0  1
    1  0  1
    1  1  1 ];
U=[ -0.542  -0.454  -0.707
    -0.542  -0.454  0.707
    -0.643  0.766  0 ];
Sigma=[ 2.524  0  0
        0  0.792  0
        0  0  0 ];
V=[ -0.683811  -0.180008  -0.707107
    -0.25457  0.967054  1.11022e-16
    -0.683811  -0.180008  0.707107 ];
k=1,Ak=U(:,1:k)*Sigma(1:k,1:k)*V(:,1:k)';R=A-Ak;
my_fprintf_array(Ak);my_fprintf_array(R);
compression_ratio=k*(n+m+1)/(n*m),
k =      1
Ak=[ 0.935459  0.348254  0.935459
     0.935459  0.348254  0.935459
     1.10978  0.41315  1.10978 ];
R=[ 0.0645415  -0.348254  0.0645415
    0.0645415  -0.348254  0.0645415
    -0.109778  0.58685  -0.109778 ];
compression_ratio =     0.7778
k=2,Ak=U(:,1:k)*Sigma(1:k,1:k)*V(:,1:k)';R=A-Ak;
my_fprintf_array(Ak);my_fprintf_array(R);
compression_ratio=k*(n+m+1)/(n*m),
k =     2
Ak=[ 1.00018  0.000531927  1.00018
     1.00018  0.000531927  1.00018
     1.00057  0.999835  1.00057 ];
R=[ -0.000183671  -0.000531927  -0.000183671
    -0.000183671  -0.000531927  -0.000183671
```

```
    -0.000572365  0.000165476  -0.000572365 ];
compression_ratio =    1.5556
```

(c) Polar decomposition: $B = U\Sigma V^T = (U\Sigma U^T)(UV^T) = SQ$

```
X=[ -2  2  2  -2
     1  1  -1  -1 ];
B=[ 1.41  -1.41
    0.354  0.354 ];
[U,Sigma,V]=svd(B);
my_fprintf_array(U);my_fprintf_array(Sigma);my_fprintf_array(V);
U=[ -1  -2.77556e-17
    -2.77556e-17  1 ];
Sigma=[ 1.99404  0
        0  0.500632 ];
V=[ -0.707107  0.707107
    0.707107  0.707107 ];
S=U*Sigma*U';Q=U*V';
my_fprintf_array(S);my_fprintf_array(Q);
S=[ 1.99404      4.14504e-17
    4.14504e-17  0.500632 ];
Q=[ 0.707107  -0.707107
    0.707107  0.707107 ];
```

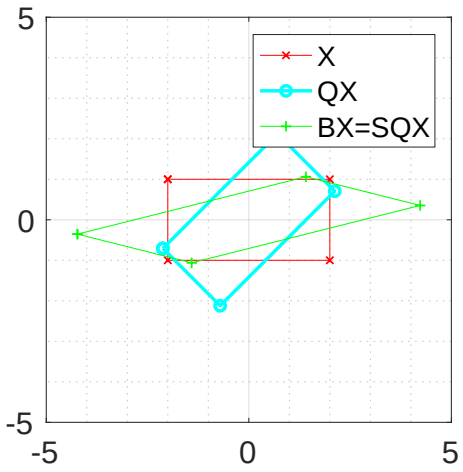
(d) transformations of X:

```
QX=Q*X;my_fprintf_array(QX);
BX=B*X;my_fprintf_array(BX);
QX=[ -2.12132  0.707107  2.12132  -0.707107
     -0.707107  2.12132  0.707107  -2.12132 ];
BX=[ -4.23  1.41  4.23  -1.41
     -0.354  1.062  0.354  -1.062 ];
```

plot:

```
figure(1);clf
set(0,'defaulttextfontsize',18); set(0,'defaultaxesfontsize',18);
plot([X(1,:),X(1,1)],[X(2,:),X(2,1)],'r-x')
hold on
plot([QX(1,:),QX(1,1)],[QX(2,:),QX(2,1)],'c-o','linewidth',2)
plot([BX(1,:),BX(1,1)],[BX(2,:),BX(2,1)],'g-+')
legend('X','QX','BX=SQX')
xlim([-5,5]);ylim([-5,5]);
axis square
grid on
grid minor
%% save as pdf:
```

```
set(gcf, 'PaperUnits', 'inches'); set(gcf, 'PaperSize', [5 4]);
set(gcf, 'PaperPosition', [0 0 5 4]); % [left, bottom, width, height];
saveas(gcf,sprintf('Figures/solution-polar_decomposition.pdf'));
```



Given that `acos(0.7071)*180/pi=45` we conclude that this transformation involves a 45 degree rotation. The stretching, based on S, is by a factor of 2 in the x direction and a compression by a factor of 2 in the y direction.

4. (13 pts) Consider the following partial differential equation for a function $u(x,y)$, $-\nabla^2 u(x,y) \equiv -(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}) = f(x,y)$ over the unit square, where the rhs $f(x,y)$ is given, and the boundary condition is $u = 0$ on the boundaries of the domain. Suppose we defined a grid with equal spacing, (x_i, y_j) and wrote the equation as a linear matrix equation for a vector $U_k \equiv u_{ij} = u(x_i, y_j)$ as $AU = F$ where F is a vector representing $f(x,y)$. (1) Approximate the condition number of the matrix A representing the Laplacian operator $-\nabla^2$ for an $n \times m$ grid, for large n, m. (2) What rhs F and what noise $\delta F = \delta f(x,y)$ would lead to an equality in the relation $\frac{\|\delta U\|}{\|U\|} \leq cond \frac{\|\delta F\|}{\|F\|}$?

Hint: This question is not asking you to repeat a calculation from HW or class, but to think independently based on material covered in class and HW. You may make assumptions regarding whichever information you find needed yet unspecified.

Solution:

(This problem was copied from http://www4.ncsu.edu/∼zhilin/TEACHING/MA402/chapt5.pdf, section 5.2.1, which is part of the course MA402: Computational Mathematics: Models, Methods and Analysis; Fall Semester, 2013; Instructor: Dr. Zhilin Li; E-mail: zhilin@math.ncsu.edu.)

(1) As an approximation, we consider the continuous formulation of the problem in order to estimate the condition number, instead of the finite difference one. The Laplacian operator is self-adjoint (APM105!) and therefore its eigenvalues are all real. Its eigenfunctions (corresponding to the eigenvectors of A) are found as follows. Looking for ϕ and λ such that $-\nabla^2 \phi(x,y) = \lambda \phi(x,y)$, we find that there is a two-parameter family of such eigenfunctions $\phi_{ij}(x,y) = \sin(i\pi x)\sin(j\pi y)$, with $i = 1, \ldots, n$, $j = 1, \ldots, m$, which satisfy,

$$-\nabla^2 \sin(i\pi x)\sin(j\pi y) = ((i\pi)^2 + (j\pi)^2)\sin(i\pi x)\sin(j\pi y),$$

and therefore that the corresponding eigenvalue is $\lambda_{ij} = ((i\pi)^2 + (j\pi)^2)$. Because the matrix is symmetric, its condition number is the ratio of largest to smallest eigenvalues. On an $n \times m$ grid, the highest mode (largest eigenvalue) that is resolved is $\phi_{nm}(x,y)$ and the lowest mode (smallest eigenvalue) is $\phi_{11}(x,y)$. The ratio of these two eigenvalues is the condition number,

$$cond = \frac{(n\pi)^2 + (m\pi)^2)}{(1\pi)^2 + (1\pi)^2} \sim n^2 + m^2.$$

(2) For the equality to take place, we need $U = \phi_{nm}(x,y)$ and the noise $\delta F = \phi_{11}(x,y)$. In that case, the equation is $-\nabla^2 \phi_{nm} = F$ which implies $F = (n^2 + m^2)\phi_{nm}$ (ignoring π factors). The disturbed equation is $-\nabla^2(u + \delta u) = F + \delta F$, so that the equation for the error is $-\nabla^2 \delta u = \delta F$. Letting $\delta u = \phi_{11}$ we find $\delta F = \phi_{11}$. Note that all eigenfunctions are normalized, so that $\|\phi_{ij}\| = 1$. Therefore, $\|\delta F\|/\|F\| \sim (n^2 + m^2)^{-1}$

while $\|\delta u\|/\|u\| \sim 1$. This implies,

$$\frac{\|\delta U\|}{\|U\|} = (n^2 + m^2)\frac{\|\delta F\|}{\|F\|}$$

as desired.