These were the problem assignments for the 18.065 course in 2019.

This file contains selected solutions by Elizabeth Chang-Davidson.

Part I, Section 5, Problem 7

Part I, Section 6, Problems 1, 16, 19

Part I, Section 7, Problems 23, 24, 26, 28

Part I, Section 8, Problems 4, 7, 8, 23

Part I, Section 9, Problems 1, 2, 10

Part I, Section 11, Problems 1, 3

Part II, Section 2, Problems 2, 3, 5, 9, 10, 11, 12, 22

Part II, Section 4, Problems 2, 3, 4, 6

Part VI, Section 4, Problem 6

Part VII, Section 1, Problems 8, 9, 11, 15, 16

18.065 Pset # 2 Elizabeth Chang-Davidson

I.5 7 Four eigenvectors of $P = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$ are $x_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$, $x_2 = \begin{pmatrix} 1 \\ i \\ i^2 \\ i^3 \end{pmatrix}$, $x_3 = \begin{pmatrix} 1 \\ i^2 \\ i^4 \\ i^6 \end{pmatrix}$, $x_4 = \begin{pmatrix} 1 \\ i^3 \\ i^6 \\ i^9 \end{pmatrix}$.

The corresponding eigenvalues are $\lambda_1 = 1$, $\lambda_2 = i$, $\lambda_3 = i^2 = -1$, $\lambda_4 = i^3 = -i$.

$$Q = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & i^2 & 1 & -1 \\ 1 & i^3 & -1 & i \end{pmatrix}$$

$$\overline{Q}^T Q = \frac{1}{4} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} = I$$

I.6 12 Find three eigenvalues and eigenvectors of A.

$$A = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} \begin{pmatrix} 2 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 1 & 2 \\ 4 & 2 & 4 \\ 2 & 1 & 2 \end{pmatrix}$$

$$x_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \lambda_1 = 6, x_2 = \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix}, \lambda_2 = 0, x_3 = \begin{pmatrix} 1 \\ -2 \\ 0 \end{pmatrix}, \lambda_2 = 0$$

I.6 16 Suppose $A = X\Lambda X^{-1}$. What is the eigenvalue matrix for $A + 2I$? What is the eigenvector matrix?

$$A + 2I = X\Lambda X^{-1} + X(2I)X^{-1} = X(2I + \Lambda)X^{-1}$$

Therefore, the eigenvalue matrix is $\Lambda + 2I$ and the eigenvector matrix is still X.

I.6 19 If the eigenvalues of A are (2, 2, 5) then A is positive definite, and therefore:

(a) invertible? Yes, none of the eigenvalues are zero.

(b) diagonalizable? Not necessarily, two of the eigenvalues are the same.

(c) not diagonalizable? Not necessarily, distinct eigenvalues is sufficient but not necessary to prove diagonalizable.

I.7 23 Suppose C is positive definite and A has independent columns. Apply the energy test to $x^T A^T C A x$ to show that $S = A^T C A$ is positive definite.

If $x^T A^T C A x$ is positive for all x, then S is positive definite. We can rewrite this in the following way:

$$x^T A^T C A x = (Ax)^T C (Ax)$$

For $y = Ax$,

$$x^T A^T C A x = y^T C y$$

which is positive for all y, since C is positive definite. Since A has independent columns, it has full rank, and is therefore invertible. For any x, we can find a y such that $Ax = y$ by taking $x = A^{-1}y$. Therefore, $A^T C A$ is positive definite.

I.7 24 For $F_1(x,y) = \frac{1}{4}x^4 + x^2y + y^2$ and $F_2(x,y) = x^3 + xy - x$ find the second derivative matrix H_2. Test for minimum, find the saddle point of F_2.

$$H_2 = \begin{pmatrix} 6x & 1 \\ 1 & 0 \end{pmatrix}$$
$$(6x - \lambda)(-\lambda) - 1 = 0$$
$$\lambda^2 - 6x\lambda - 1 = 0$$
$$\lambda = 3x \pm \sqrt{9x^2 + 1}$$

For H_2 to be positive definite, we need $\lambda = 3x - \sqrt{9x^2+1} > 0$, but this is never true because

$$\sqrt{9x^2+1} > \sqrt{9x^2} = 3x$$

so this function fails the test for a minimum.

The saddle point is where

$$\frac{\partial F}{\partial x} = \frac{\partial F}{\partial y} = \frac{\partial F}{\partial z} = 0$$

. For F_2, this is

$$3x^2 + y - 1 = x = 0 = 0$$

which gives us $(x,y) = (0,1)$ as the saddle point.

I.7 26 Without multiplying, find:

(a) the determinant of S: multiply determinants. This gives 1*10*1 = 10.

(b) the eigenvalues of S: the values along the diagonal of the central matrix. These are (2, 5).

(c) the eigenvectors of S: the columns of the outer matrices. This gives $x_1 = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$ and $x_2 = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$.

(d) a reason why S is symmetric positive definite: the eigenvalues are all positive, and the eigenvectors are orthogonal.

I.7 28 Suppose S is positive definite with eigenvalues $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_n$.

(a) What are the eigenvalues of $\lambda_1 I - S$? Is it positive semidefinite?
If we take the eigenvectors $v_1, v_2, \ldots v_n$ and multiple them by $\lambda_1 I - S$ we get

$$(\lambda_1 I - S)v_i = \lambda_1 v_i - Sv_i = \lambda_1 - \lambda_i v_i = (\lambda_1 - \lambda_i)v_i$$

Our new eigenvalues are then $0, \lambda_1 - \lambda_2, \lambda_1 - \lambda_3, \ldots \lambda_1 - \lambda_n$, and it is positive semidefinite since due to how we defined our λs, these are all greater than or equal to zero.

(b) How does it follow that $\lambda_1 x^T x \geq x^T S x$ for every x?
Since $\lambda_1 I - S$ is positive semidefinite, we have for all x

$$x^T(\lambda_1 I - S)x \geq 0$$
$$(\lambda_1 x^T I - x^T S)x \geq 0$$
$$\lambda_1 x^T x - x^T S x \geq 0$$
$$\lambda_1 x^T x \geq x^T S x$$

(c) Draw this conclusion: The maximum value of $x^T Sx/x^T x$ is λ_1.
Taking our previous equation and diving both sides by $x^T x$, we get (for all x)

$$\lambda_1 \geq x^T Sx/x^T x$$

We can set $x = v_1$ to get

$$(v_1^T(Sv_1))/(v_1^T v_1) = (v_1^T \lambda_1 v_1)/(v_1^T v_1) = \lambda_1$$

And since we have show that $x^T Sx/x^T x$ is no larger than λ_1 for all possible x, and that it attains λ_1 for at least one x, λ_1 must be the maximum value.

I.8 7 What is the norm $||A - \sigma_1 u_1 v_1^T||$ when that largest rank one piece of A is removed? What are all the singular values of this reduced matrix, and its rank?

Since we can write A as a sum of $\sigma_i u_i v_i^T$, the norm of A without the largest rank one piece is just the second largest σ_i. The singular values of this reduced matrix are the same but instead of σ_1 there's a 0, and its rank is rank$(A) - 1$.

I.8 8 Find the σ's and v's and u's, and verify that $A = \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 3 \\ 0 & 0 & 0 \end{pmatrix} = U\Sigma V^T$ such that the orthogonal matrices U and V are permutation matrices.

$A^T A$ has the eigenvalues 9, 4, 0. A therefore has the singular values $\sigma_1 = 3, \sigma_2 = 2, \sigma_3 = 0$. It has the eigenvectors correspondingly of $v_1 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, v_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. AA^T has the same eigenvalues but the eigenvectors are instead $u_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, u_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, u_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

Putting this all together,

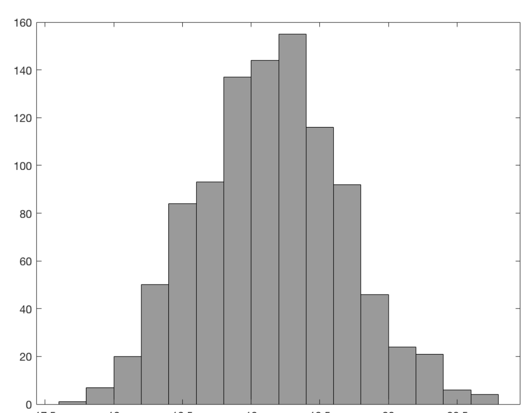
$$U = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \Sigma = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}, V = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

and we can see U and V are indeed permutation matrices.

Extra Here's my code (matlab) and the histogram:

```
n = 100;
largEig = ones(1, n);
for i = 1:1000
    A = randi([0 1],n,n);
    A = A - tril(A,-1) + triu(A,1)';
    zeros = (A == 0);
    B = A - zeros;
    largEig(i) = max(eig(B));
end

histogram(largEig)
```



18.065 Pset # 3 Elizabeth Chang-Davidson

I.8 4. A symmetric matrix $S = S^T$ has orthonormal eigenvectors v_1 to v_n. Then any vector x can be written as $c_1v_1 + \ldots + c_nv_n$. The Rayleigh quotient $R(x)$ can be written in the following way.

$$R(x) = \frac{x^TSx}{x^Tx} = \frac{\lambda_1 c_1^2 + \ldots + \lambda_n c_n^2}{c_1^2 + \ldots + c_n^2}$$

The maximum of $R(x)$ is λ_3 subject to what two conditions on x? We need $x^Tv_1 = 0$ and $x^Tv_2 = 0$, in order to rule out the two eigenvectors with larger eigenvalues. After those two conditions rule those cases out, the largest possible case left is where $x = v_3$ and the values achieved is λ_3.

I.8 23. Show that an m by n matrix of rank r has $r(m+n-r)$ free parameters in its SVD: $A = U\Sigma V^T = (m \times r)(r \times r)(r \times n)$. Why do r orthonormal vectors u_1 to u_r have $(m-1)+(m-2)+\ldots+(m-r)$ parameters?

U has $m-1$ choices for the first column, because it has to be normalized. There are $m-2$ for the second column, because has to be normalized and orthogonal to the first one. There are $m-3$ for the third, because it has to be normalized and orthogonal to the first two. This pattern continues until you get $(m-1)+(m-2)+\ldots+(m-r)$ for the rth column. This adds up to $mr - \frac{1}{2}r(r+1)$ choices overall for U.

By an identical train of logic on rows instead of columns, V has $nr - \frac{1}{2}r(r+1)$ choices.

Σ has r possible choices down the diagonal.

Putting this all together, we get

$$\left(mr - \frac{1}{2}r(r+1)\right) + \left(r\right) + \left(nr - \frac{1}{2}r(r+1)\right) = mr + nr + r - r^2 - r = r(m+n+r)$$

I.9 1. What are the singular values (in descending order) of $A - A_k$? Omit any zeros. We know we can write A as

$$A = \sigma_1 u_1 v_1^T + \ldots + \sigma_r u_r v_r^T$$

and that we can write A_k as

$$A_k = \sigma_1 u_1 v_1^T + \ldots + \sigma_k u_k v_k^T$$

so when we take the difference we get

$$A - A_k = \sigma_{k+1} u_{k+1} v_{k+1}^T + \ldots + \sigma_r u_r v_r^T$$

which means that the singular values are $\sigma_{k+1} \ldots \sigma_r$.

I.9 2. Find a closest rank-1 approximation to these matrices (L^2 or Frobenius norm).

The closest rank-1 approximation of a matrix is $A_1 = \sigma_1 u_1 v_1^T$. Therefore, we have the following approximations:

$$A = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$A_1 = 3 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$A = \begin{pmatrix} 0 & 3 \\ 2 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$A_1 = 3 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 3 \\ 0 & 0 \end{pmatrix}$$

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}^T$$

$$A_1 = \frac{3}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \end{pmatrix} = \frac{3}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

I.9 10. If A is a 2×2 matrix with $\sigma_1 \geq \sigma_2 > 0$, find $||A^{-1}||_2$ and $||A^{-1}||_F^2$.

A^{-1} will have the same eigenvectors, and the eigenvalues will be $\frac{1}{\lambda_1}$ and $\frac{1}{\lambda_2}$, the second of which will be larger, producing σs of $\frac{1}{\sigma_2}$ and $\frac{1}{\sigma_1}$ in descending order.

$$||A^{-1}||_2 = \frac{1}{\sigma_2}$$

$$||A^{-1}||_F^2 = \frac{1}{\sigma_2}^2 + \frac{1}{\sigma_1}^2$$

I.11 1. Show directly this fact about vector norms: $||v||_2^2 \leq ||v||_1 ||v||_\infty$

$$||v||_2^2 = |v_1|^2 + |v_2|^2 + \ldots + |v_n|^2$$

$$||v||_1 = |v_1| + |v_2| + \ldots + |v_n|$$

$$||v||_\infty = \max(|v_1|, |v_2|, \ldots, |v_n|)$$

Let us designate v_k to be $||v||_\infty$. We then have $|v_i| \leq |v_k|$ for all i. Therefore we have $|v_i|^2 \leq |v_i| * |v_k|$ for all i.

$$||v||_1 ||v||_\infty = |v_1||v_k| + |v_2||v_k| + \ldots + |v_n||v_k|$$

Looking at each term of $||v||_2^2$ and $||v||_1 ||v||_\infty$, we can see that by our inequality above, every term of $||v||_2^2$ is less than or equal to the corresponding term in $||v||_1 ||v||_\infty$, so we have proved our inequality overall.

I.11 3. Show that always $||v||_2 \leq \sqrt{n}||v||_\infty$. Also prove $||v||_1 \leq \sqrt{n}||v||_2$ by choosing a suitable vector w and applying the Cauchy-Swartz inequality.

Let us again designate v_k to be $||v||_\infty$. We then have that

$$||v||_2^2 = |v_1|^2 + |v_2|^2 + \ldots + |v_n|^2$$

$$n||v||_\infty^2 = n|v_k|^2$$

We can see that for each term in $||v||_2^2$, we have that $|v_i|^2 \leq |v_k|^2$ for all i. We have n such terms, so $||v||_2^2 \leq n||v||_\infty^2$, which gives us our equality when we take the square root of each side.

For the second half:

$$||v||_2^2 = |v_1|^2 + |v_2|^2 + \ldots + |v_n|^2 = \Big(\sum_{i=1}^{n} \frac{1}{n}\Big)\Big(\sum_{i=1}^{n} |v_i|^2\Big)$$

$$\frac{1}{n}||v||_1^2 = \frac{1}{n}(|v_1| + |v_2| + \ldots + |v_n|)^2 = \Big(\sum_{i=1}^{n} \frac{1}{\sqrt{n}}|v_i|\Big)^2$$

From Cauchy-Swartz, we know that

$$\Big(\sum_{i=1}^{n} \frac{1}{\sqrt{n}}|v_i|\Big)^2 \leq \Big(\sum_{i=1}^{n} \frac{1}{n}\Big)\Big(\sum_{i=1}^{n} |v_i|^2\Big)$$

$$\frac{1}{n}||v||_1^2 \leq ||v||_2^2$$

$$||v||_1^2 \leq n||v||_2^2$$

$$||v||_1 \leq \sqrt{n}||v||_2$$

A. Find the sample covariance matrix $S = \frac{AA^T}{3}$ and find the line through $(0,0,0)$ that is closest to the four columns (from the SVD of A).

$$S = \frac{1}{3}\begin{pmatrix} 2 & -1 & 1 \\ -1 & 10 & -3 \\ 1 & -3 & 2 \end{pmatrix}$$

The line through $(0,0,0)$ that is closest to the four columns of A is in the direction of u_1, which we find from the SVD to be $u_1 = \begin{pmatrix} -0.1371 & 0.9370 & -0.3213 \end{pmatrix}^T$. The equation for the line is then (in parametric and then nonparametric format):

$$(x, y, z) = t(-0.1371, 0.9370, -0.3213)$$

$$\frac{x}{-0.1371} = \frac{y}{0.9370} = \frac{z}{-0.3213}$$

B. Find the plane through $(0,0,0)$ that is closest to the four columns (from the SVD of A).

The plane through $(0,0,0)$ that is closest to the four columns is the plane containing the vectors $u_1 = \begin{pmatrix} -0.1371 & 0.9370 & -0.3213 \end{pmatrix}^T$ and $u_2 = \begin{pmatrix} -0.8716 & -0.2683 & -0.4103 \end{pmatrix}^T$, which has the normal vector:

$$n = \begin{pmatrix} -0.1371 & 0.9370 & -0.3213 \end{pmatrix} \times \begin{pmatrix} -0.8716 & -0.2683 & -0.4103 \end{pmatrix} = \begin{pmatrix} -0.4706 & 0.2238 & 0.8535 \end{pmatrix}$$

We then can form the equation for the plane:

$$-0.4706x + 0.2238y + 0.8535z = 0$$

C. Comparing with Fig I.16, what shapes (with rough sketches) show these three sets in 3D?

$$|v_1| + |v_2| + |v_3| \leq 1$$

This is a octahedron with points distance 1 away from the origin along the x, y, and z axes.

$$v_1^2 + v_2^2 + v_3^2 \leq 1$$

This is a sphere with radius 1 centered at the origin.

$$\max(|v_1|, |v_2|, |v_3|) \leq 1$$

This is a cube centered at the origin with face centers at the points of the octahedron.

D. If you blow up those 3 sets, where will they touch the plane $v_1 + 2v_2 + 5v_3 = 1$? Your 3 points will be the smallest solutions (in those 3 norms) to that linear equation.

The points are $(0, 0, 0.2)$, $(\frac{1}{30}, \frac{1}{15}, \frac{1}{6})$, and $(0.125, 0.125, 0.125)$ in the $||v||_1$, $||v||_2$, and $||v||_\infty$ norms respectively.

18.065 Pset # 4 Elizabeth Chang-Davidson

II.2 2. Why do A and A^+ have the same rank? If A is square, do A and A^+ have the same eigenvectors? What are the eigenvalues of A^+?

We know that

$$A^+ = V\Sigma^+U^T = \sum_{i=1}^{r} \frac{v_i u_i^T}{\sigma_i}$$

Each of the $v_i u_i^T$ is a rank one matrix, and there is one for each σ_i in A, so A^+ is the sum of r rank one matrices and therefore has rank r.

Eigenvectors of A are in the row space and the column space, since $Av_i = \lambda v_i$. We have $A^+Ax = x$ when x is in the row space, so

$$A^+(Av_i) = A^+(\lambda_i v_i)$$

$$A^+Av_i = v_i = \lambda_i A^+ v_i$$

$$\frac{1}{\lambda_i} v_i = A^+ v_i$$

Therefore, they have the same eigenvectors and the eigenvalues of A^+ are the inverse of the eigenvalues of A.

II.2 3. From A and A^+ show that A^+A is correct and that $(A^+A)^2 = A^+A =$ projection.

$$A = \sum \sigma_i u_i v_i^T$$

$$A^+ = \sum \frac{v_i u_i^T}{\sigma_i}$$

Because of orthonormality in u_i and v_i, we only need to worry about the terms with the same index when multiplying, since everything else goes to zero. Also, $u_i^T u_i = ||u_i|| = 1$.

$$A^+A = \sum \frac{v_i u_i^T}{\sigma_i} \sigma_i u_i v_i^T = \sum v_i u_i^T u_i v_i^T = \sum v_i v_i^T$$

Which matches what we are given. In the next half, we can use orthonormality again to get $v_i^T v_i = ||v|| = 1$.

$$(A^+A)^2 = \sum v_i v_i^T v_i v_i^T = \sum v_i v_i^T = A^+A$$

II.2 5. Suppose A has independent columns (rank $r = n$; nullspace = zero vector)

(a) Describe the m by n matrix Σ in $A = U\Sigma V^T$. How many nonzeros are there in Σ? Because A is full rank, there will be no zeros along the diagonal of Σ, but everywhere outside the diagonal will be zeros.

Σ is a m by n matrix, with rank $r = n$ and $m \geq n$. The first n rows of Σ form a diagonal matrix with nonzero entries along the diagonal. The last $m - n$ rows are all zeros.

(b) Show that $\Sigma^T\Sigma$ is invertible by finding its inverse. Let's define $m-n=p$ and also define Σ_r in the following way:

$$\Sigma_r = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \sigma_r \end{bmatrix}$$

$$\Sigma^T\Sigma = \begin{bmatrix} \Sigma_r & 0_{r\times p} \end{bmatrix} \begin{bmatrix} \Sigma_r \\ 0_{r\times p} \end{bmatrix} = \begin{bmatrix} \Sigma_r\Sigma_r + 0_{r\times p}0_{p\times r} \end{bmatrix} = \Sigma_r\Sigma_r = \Sigma_r^2$$

$$\Sigma_r^2 = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \sigma_r^2 \end{bmatrix}$$

The inverse of a square diagonal matrix with no zeros down the diagonal is just one over each diagonal entry.

$$(\Sigma_r^2)^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2^2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{\sigma_r^2} \end{bmatrix} = (\Sigma^T\Sigma)^{-1}$$

(c) Write down the n by m matrix $(\Sigma^T\Sigma)^{-1}\Sigma^T$ and identify it as Σ^+.

$$(\Sigma^T\Sigma)^{-1}\Sigma^T = \begin{bmatrix} (\Sigma_r^2)^{-1} \end{bmatrix} \begin{bmatrix} \Sigma_r & 0_{r\times p} \end{bmatrix} = \begin{bmatrix} (\Sigma_r^2)^{-1}\Sigma_r & 0_{r\times p} \end{bmatrix}$$

$$(\Sigma_r^2)^{-1}\Sigma_r = \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2^2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{\sigma_r^2} \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \sigma_r \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{\sigma_r} \end{bmatrix}$$

$(\Sigma^T\Sigma)^{-1}\Sigma^T$ is therefore equal to Σ^+, since this is the same result as if you had transposed and then inverted the nonzero diagonals.

(d) Substitute $A = U\Sigma V^T$ into $(A^TA)^{-1}A^T$ and identify that matrix as A^+.

$$\begin{aligned} (A^TA)^{-1}A^T &= ((U\Sigma V^T)^T(U\Sigma V^T))^{-1}(U\Sigma V^T)^T \\ &= ((V\Sigma^TU^T)(U\Sigma V^T))^{-1}(V\Sigma^TU) \\ &= (V\Sigma^T\Sigma V^T)^{-1}(V\Sigma^TU) \end{aligned}$$

Due to orthogonality, we know that

$$(V\Sigma^T\Sigma V^T)^{-1} = V(\Sigma^T\Sigma)^{-1}V^T$$

And so we have

$$
\begin{aligned}
(V\Sigma^T\Sigma V^T)^{-1}(V\Sigma^T U) &= V(\Sigma^T\Sigma)^{-1}V^T(V\Sigma^T U) \\
&= V(\Sigma^T\Sigma)^{-1}\Sigma^T U \\
&= V\Sigma^+ U
\end{aligned}
$$

Which is indeed A^+.

Conclusion: $A^T A\hat{x} = A^T b$ leads to $A^+ = (A^T A)^{-1}A^T$, but only if A has rank n

II.2 9. Complete the Gram-Schmidt process in Problem 8 by computing $q_1 = a/||a||$ and $A_2 = b - (b^T q_1)q_1$ and $q_2 = A_2/||A_2||$ and factoring into QR.

$$q_1 = a/||a|| = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\1\end{bmatrix} = \begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}$$

$$A_2 = b - (b^T q_1)q_1 = \begin{bmatrix}4\\0\end{bmatrix} - (\begin{bmatrix}4 & 0\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix})\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix} = \begin{bmatrix}4\\0\end{bmatrix} - \frac{4}{\sqrt{2}}\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix} = \begin{bmatrix}4\\0\end{bmatrix} - \begin{bmatrix}2\\2\end{bmatrix} = \begin{bmatrix}2\\-2\end{bmatrix}$$

$$q_2 = A_2/||A_2|| = \begin{bmatrix}\frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}}\end{bmatrix}$$

$$
\begin{aligned}
\begin{bmatrix}1 & 4\\1 & 0\end{bmatrix} = \begin{bmatrix}q_1 & q_2\end{bmatrix}\begin{bmatrix}||a|| & 2\sqrt{2}\\ 0 & ||A_2||\end{bmatrix} &= \begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}\begin{bmatrix}\sqrt{2} & 2\sqrt{2}\\ 0 & 2\sqrt{2}\end{bmatrix} \\
Q &= \begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix} \\
R &= \begin{bmatrix}\sqrt{2} & 2\sqrt{2}\\ 0 & 2\sqrt{2}\end{bmatrix}
\end{aligned}
$$

II.2 10. If $A = QR$ then $A^T A = R^T R =$ lower triangular times upper triangular.

II.2 11. If $Q^T Q = I$ show that $Q^T = Q^+$. If $A = QR$ for invertible R, show that $QQ^T = AA^+$.

$$
\begin{aligned}
QQ^T Q &= Q * I = Q \\
Q^T QQ^T &= I * Q^T = Q^T \\
(QQ^T)^T &= QQ^T \\
(Q^T Q)^T &= I^T = I = Q^T Q
\end{aligned}
$$

Since Q^T satisfies the properties of Q^+ and Q^+ is unique, $Q^T = Q^+$.

Now for the second half. For A with independent columns, $A^+ = (A^TA)^{-1}A^T$. We can plug in $A = QR$ to this. This gives us

$$AA^+ = A(A^TA)^{-1}A^T = QR((QR)^T(QR))^{-1}(QR)^T = QR(R^TQ^TQR)^{-1}R^TQ^T$$

For invertible R and $Q^TQ = I$, this works out to:

$$= QR(R^TIR)^{-1}R^TQ^T = QRR^{-1}(R^T)^{-1}R^TQ^T = QQ^T$$

II.2 12. With $b = (0, 8, 8, 20)$ at $t = (0, 1, 3, 4)$, set up and solve the normal equations $A^TA\hat{x} = A^Tb$. For the best straight line in Figure II.3a, find its four height p_i and four errors e_i. What is the minimum squared error $E = e_1^2 + e_2^2 + e_3^2 + e_4^2$?

$$A^TA\hat{x} = \begin{bmatrix} 4 & 8 \\ 8 & 26 \end{bmatrix} \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 36 \\ 112 \end{bmatrix}$$

$$\hat{x} = \begin{bmatrix} \hat{C} \\ \hat{D} \end{bmatrix} = \begin{bmatrix} 4 & 8 \\ 8 & 26 \end{bmatrix}^{-1} \begin{bmatrix} 36 \\ 112 \end{bmatrix} = \frac{1}{20} \begin{bmatrix} 13 & -4 \\ -4 & 2 \end{bmatrix} \begin{bmatrix} 36 \\ 112 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

$p_i = \hat{C} + \hat{D}t_i$ so our p_is will be 1, 5, 10, 13. This makes our e_is be 1, 3, 2, 7 (signs not given) for a total minimum squared error of $E_1 = 63$.

II.2 22. The averages of the t_i and b_i are $\bar{t} = 2$ and $\bar{b} = 9$. Verify that $C + D\bar{t} = \bar{b}$. Explain!

$$C + D\bar{t} = 1 + 4 * 2 = 9 = \bar{b}$$

Intuitively, we are weighting all the errors equally, so it should go through the average point and rotate from there to find the best line. More quantitatively, our first equation from $A^TA\hat{x} = A^Tb$ gives us

$$m\hat{C} + \hat{D}\sum t_i = \sum b_i$$

If we divide by m all through, this gives us

$$\hat{C} + \hat{D}\frac{1}{m}\sum t_i = \frac{1}{m}\sum b_i$$

which we can see gives us this same relation.

Comp. Q. Create a random 6 by 10 matrix. (You can choose the definition of random) Find its SVD and its pseudoinverse.

A

6x10 double

	1	2	3	4	5	6	7	8	9	10
1	0.1239	0.5650	0.1057	0.9312	0.9844	0.1339	0.4671	0.3479	0.8985	0.2878
2	0.4904	0.6403	0.1420	0.7287	0.8589	0.0309	0.6482	0.4460	0.1182	0.4145
3	0.8530	0.4170	0.1665	0.7378	0.7856	0.9391	0.0252	0.0542	0.9884	0.4648
4	0.8739	0.2060	0.6210	0.0634	0.5134	0.3013	0.8422	0.1771	0.5400	0.7640
5	0.2703	0.9479	0.5737	0.8604	0.1776	0.2955	0.5590	0.6628	0.7069	0.8182
6	0.2085	0.0821	0.0521	0.9344	0.3986	0.3329	0.8541	0.3308	0.9995	0.1002
7										

U

6x6 double

	1	2	3	4	5	6
1	-0.4285	0.4671	-0.0432	0.0366	0.2681	-0.7233
2	-0.3663	-0.0076	-0.3694	0.0510	0.6946	0.4942
3	-0.4453	-0.0651	0.7820	-0.3820	0.0772	0.1844
4	-0.3642	-0.7691	0.0147	0.4484	-0.0326	-0.2711
5	-0.4535	-0.1054	-0.4956	-0.5533	-0.4804	0.0241
6	-0.3815	0.4182	0.0654	0.5856	-0.4558	0.3529

S

6x10 double

	1	2	3	4	5	6	7	8	9
1	4.0316	0	0	0	0	0	0	0	0
2	0	1.1980	0	0	0	0	0	0	0
3	0	0	1.1596	0	0	0	0	0	0
4	0	0	0	0.8614	0	0	0	0	0
5	0	0	0	0	0.8439	0	0	0	0
6	0	0	0	0	0	0.3825	0	0	0

V

10x10 double

	2	3	4	5	6	7	8	9	10
1	-0.5132	0.3217	0.0789	0.2208	0.4003	-0.2390	-0.1223	-0.3589	-0.3715
2	0.0065	-0.3417	-0.5690	0.1528	-0.0505	-0.1840	-0.3600	0.4161	-0.3296
3	-0.3997	-0.1713	-0.0709	-0.2130	-0.2920	0.6741	0.2411	-0.0845	-0.3568
4	0.5281	-0.0834	-0.1291	-0.0339	0.4077	0.3924	-0.1551	-0.3892	0.0954
5	0.1296	0.1726	0.1683	0.7554	-0.3579	0.1128	0.2293	0.1088	0.0988
6	-0.1023	0.5148	-0.2157	-0.2059	0.3516	0.1715	0.2634	0.5680	0.2130
7	-0.1151	-0.3869	0.7069	-0.1279	0.1926	0.0080	-0.1829	0.3714	0.0525
8	0.0733	-0.3808	-0.0916	-0.0802	0.1661	-0.3932	0.7719	-0.0980	-0.0704
9	0.2358	0.3566	0.1132	-0.4901	-0.4859	-0.3055	-0.0782	-0.0862	-0.1707
10	-0.4432	-0.1636	-0.2292	-0.0743	-0.1821	-0.1025	-0.1262	-0.2237	0.7246

Aplus

10x6 double

	1	2	3	4	5	6
1	-0.8656	0.6299	0.4541	0.1078	-0.2119	0.1693
2	0.1666	0.1627	0.0441	-0.2480	0.4543	-0.5048
3	0.3500	-0.4845	-0.2041	0.4480	0.2756	-0.3358
4	-0.5316	0.5542	0.2140	-0.6554	0.1664	0.5277
5	1.0077	0.1474	-0.0275	0.2649	-0.6039	-0.5335
6	-0.7758	0.1281	0.6225	-0.2621	0.0908	0.3024
7	-0.3702	0.3395	-0.4504	0.3352	-0.1564	0.6967
8	-0.2789	0.2826	-0.1248	-0.1957	0.2944	0.1579
9	0.8932	-1.0998	-0.0533	0.3150	0.0517	0.0373
10	0.1751	-0.2287	-0.0471	0.3213	0.3196	-0.4202

18.065 Pset # 6 Elizabeth Chang-Davidson

II.4 2 *(for functions)* Given $a(x) > 0$ find $p(x) > 0$ by analogy with problem 1, so that

$$\int_0^1 p(x)dx = 1 \text{ and } \int_0^1 \frac{(a(x))^2}{p(x)}dx \text{ is a minimum}$$

$$L(p(x), \lambda) = \int_0^1 \frac{(a(x))^2}{p(x)}dx + \int_0^1 \lambda p(x)dx - \int_0^1 \lambda dx$$

$$L(p(x), \lambda) = \int_0^1 \Big[\frac{(a(x))^2}{p(x)}dx + \lambda p(x)dx - \lambda\Big]dx$$

$$\frac{\partial L}{\partial p(x)} = 0 = -\frac{(a(x))^2}{(p(x))^2} + \lambda$$

$$\frac{(a(x))^2}{(p(x))^2} = \lambda$$

$$p(x) = \frac{a(x)}{\sqrt{\lambda}} = \frac{a(x)}{C}$$

We take C such that the integral of $p(x)$ is 1:

$$C = \int_0^1 a(x)dx$$

And this minimizes our integral.

II.4 3 Prove that $n(a_1^2 + \ldots + a_n^2) \geq (a_1 + \ldots + a_n)^2$. This is problem 1 with $p_i = 1/n$. Back in Problem Set I.11 you proved that $||a||_1 \leq \sqrt{n}||a||_2$.

We take $p_i = \frac{1}{n}$ and so

$$V = \sum_{i=1}^n \frac{a_i^2}{p_i} = \sum_{i=1}^n n(a_i^2) = n\sum_{i=1}^n a_i^2$$

which is the left side of our equation. In problem 1, we saw that choosing p_i correctly gives the minimum value of V, which is $(\sum_{i=1}^n a_i)^2$. Since this is a minimum, the choice of any p_is must yield a value for V greater than or equal to that value, which gives us

$$n\sum_{i=1}^n a_i^2 \geq (\sum_{i=1}^n a_i)^2$$

which is the inequality we wanted.

II.4 4 If $M = 11^T$ is the $n \times n$ matrix of 1s, prove that $nI - M$ is positive semidefinite. Problem 3 was the energy test. For Problem 4, find the eigenvalues of $nI - M$.

The eigenvalues of $nI - M$ are the solutions to the equation $\det(nI - M - \lambda I) = 0$.

$$\det(nI - M - \lambda I) = \det((n - \lambda)I - M) = 0$$

We can see that $\lambda = n$ gives us just $\det(-M)$ and since this is the all ones matrix, this will be zero. The nullspace of M is going to be $n-1$-dimensional, which gives us $n-1$ eigenvectors with $\lambda = n$.

We can also see that every entry down the diagonal of $nI - M$ will be $n-1$, and all the other $n-1$ entries of each row will be -1, so the total sum of entries down each row will be zero. This gives us an eigenvector of the all ones vector, with an eigenvalue of 0.

In the end, all the eigenvalues are either n or 0, which makes $nI - M$ positive semidefinite.

II.4 6 The variance computed in equation 7 cannot be negative! Show this directly:

$$||AB||_F^2 \leq (\sum ||a_k||\,||b_k^T||)^2$$

By the Triangle Inequality ($||P+Q||_F \leq ||P||_F + ||Q||_F$):

$$||AB||_F^2 = \Big(||\sum_{k=1}^{n} a_k b_k^T||_F\Big)^2 \leq \Big(\sum_{k=1}^{n} ||a_k b_k^T||_F\Big)^2$$

By the definition of the Frobenius norm,

$$\Big(\sum_{k=1}^{n} ||a_k b_k^T||_F\Big)^2 = \Big(\sum_{k=1}^{n} ||a_k||_F ||b_k^T||_F\Big)^2 = \Big(\sum_{k=1}^{n} ||a_k||\,||b_k^T||\Big)^2$$

which is what we wanted to show

Computational Problem

Take a matrix A where A is 0 below the diagonal and 1 above and on the diagonal and is of order 1000 by 1000. Compare the actual SVD of A to the randomized SVD of A reduced to $Y = AG$: the first Gaussian random matrix G is 1000 by 10 and the second G is 1000 by 100.

After implementing this algorithm, I found that the Frobenius norm of the difference between the singular value matrix found by randomization and the actual singular value matrix was 3.0774×10^{-12} when using G 1000 by 10 and 2.9577×10^{-12} when using G 1000 by 100, showing that this is a very accurate method.

Martinsson's 4 steps are:

1. $Y = AG$
2. Factor $Y = QR$
3. Find the SVD of $Q^T A = UDV^T$
4. The approximate SVD of A is $(QU)DV^T$

Here is the code I used:

```
A = triu(ones(1000));
[U1,S1,V1] = svd(A);
G = normrnd(0,1,[1000,10]);
Y = A*G;
[Q,R] = qr(Y);
[U2,S2,V2] = svd(Q'*A);
Uapp = Q*U2;
Sapp = S2;
Vapp = V2;
err = norm(Sapp-S1,'fro');
```