

Matrix (Data) Approximation

Justin Wyss-Gallifent

July 21, 2021

9.1	Introduction	2
9.2	Geometric Inspiration for U	2
9.3	Algebraic Evidence	3
9.4	Summary and Matrix Comment	6
9.5	Data Modeling and Prediction	10
9.6	Centering Comment	11
9.7	Matlab	13
9.8	Exercises	13

9.1 Introduction

For almost all of our uses of the singular value decomposition we'll be thinking of an $m \times n$ matrix as made up of columns where each column represents something useful. For now just think about the matrix as containing n points in \mathbb{R}^m .

Thinking this way, it turns out that the SVD contains a lot of useful information about the points.

Before we proceed we should note that we will only need the matrices U and Σ but not V so typically we'll just write V^T in our SVD instead of actually writing V explicitly.

The general approach will be as follows:

- (I) Get some geometric inspiration for what the left-singular vectors (the columns of U) represent.
- (II) Get some algebraic evidence and figure out what the singular values (the entries in Σ) represent.
- (III) Put together a conclusion and solidify this with examples.

Before proceeding, two definitions:

Definition 9.1.0.1. The *variance* of a set of vectors equals the sum of the squares of the norms of the vectors.

Definition 9.1.0.2. The *norm* of a matrix, denoted $\|A\|$, equals the square root of the sum of the squares of the entries, meaning it equals the square root of the variance of the columns of the matrix.

It also follows that two matrices are close to one another (in every entry) if the norm of their difference is small.

9.2 Geometric Inspiration for U

Let's start off with an example:

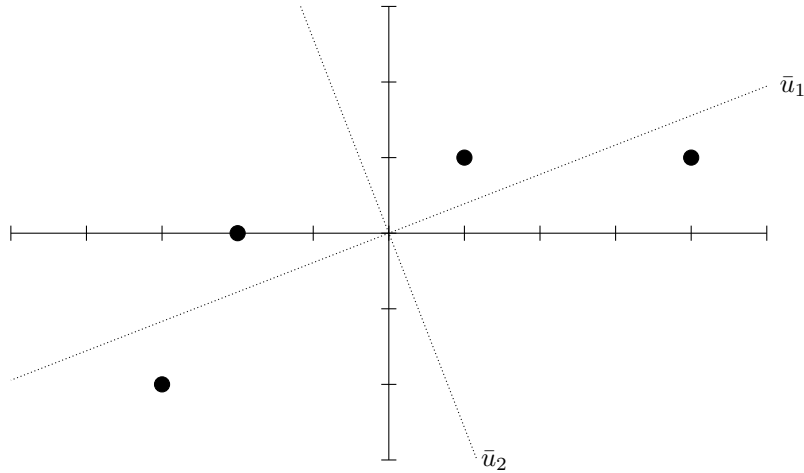
Example 9.1. Consider the matrix:

$$A = \begin{bmatrix} 4 & 1 & -2 & -3 \\ 1 & 1 & 0 & -2 \end{bmatrix}$$

This matrix has the following SVD:

$$A = \begin{bmatrix} -0.9320 & -0.3625 \\ -0.3625 & 0.9320 \end{bmatrix} \begin{bmatrix} 5.855 & 0 & 0 & 0 \\ 0 & 1.312 & 0 & 0 \end{bmatrix} V^T$$

If we plot the columns of A as points along with the lines generated by each of the columns of U (call these \bar{u}_1 and \bar{u}_2) we see:



What do we see? It looks like the data is really spread out in the \bar{u}_1 direction and not so spread out in the \bar{u}_2 direction.

What is this example suggesting? Well, it looks like:

- (a) Column \bar{u}_1 (associated to s_1 , the larger singular value) is indicating the direction in which the columns of A are most spread out.
- (b) Column \bar{u}_2 (associated to s_2 , the smaller singular value) the one associated to the second largest singular value, is indicating the direction perpendicular to \bar{u}_1 in which the columns of A are second most spread out.

It may seem reasonable to suggest that if there are more than two columns in U that they tell us how the columns of A are most spread out in perpendicular directions of decreasing importance.

9.3 Algebraic Evidence

If we look at the SVD of an $m \times n$ matrix A :

$$A = U\Sigma V^T$$

where the dimensions are $m \times m$, $m \times n$ (with k singular values s_1, \dots, s_k) and $n \times n$ respectively, then a relatively straightforward calculation shows that the columns of A are calculated by:

$$\begin{aligned}\bar{a}_1 &= s_1 v_{11} \bar{u}_1 + s_2 v_{12} \bar{u}_2 + \dots + s_k v_{1k} \bar{u}_k \\ \bar{a}_2 &= s_1 v_{21} \bar{u}_1 + s_2 v_{22} \bar{u}_2 + \dots + s_k v_{2k} \bar{u}_k \\ &\vdots \\ \bar{a}_n &= s_1 v_{n1} \bar{u}_1 + s_2 v_{n2} \bar{u}_2 + \dots + s_k v_{nk} \bar{u}_k\end{aligned}$$

First, this tells us that each column of A is built out of the columns in U , which makes sense since the columns of U form a basis for \mathbb{R}^m .

However since $s_1 \geq s_2 \geq \dots \geq s_k \geq 0$ this is suggesting exactly what we've seen geometrically, that \bar{u}_1 is most important, \bar{u}_2 next most and so on.

In addition this tells us that if we create Σ' by taking Σ and replacing some of the s_i by 0, and if we then recalculate $A' = U\Sigma'V^T$, the result will be the orthogonal projection of the columns of A onto the columns of U corresponding to the s_i which remain because essentially we are ignoring the directions associated to the s_i we eliminated.

To help clarify let's revisit our initial example:

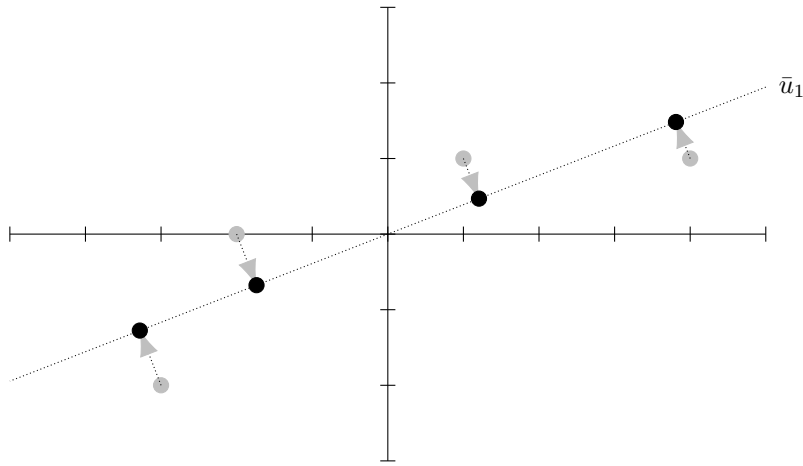
Example 9.2.

$$A = \begin{bmatrix} 4 & 1 & -2 & -3 \\ 1 & 1 & 0 & -2 \end{bmatrix} = \begin{bmatrix} -0.9320 & -0.3625 \\ -0.3625 & 0.9320 \end{bmatrix} \begin{bmatrix} 5.855 & 0 & 0 & 0 \\ 0 & 1.312 & 0 & 0 \end{bmatrix} V^T$$

If we set $s_2 = 0$ and recalculate the matrix product:

$$\begin{bmatrix} -0.9320 & -0.3625 \\ -0.3625 & 0.9320 \end{bmatrix} \begin{bmatrix} 5.855 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} V^T = \begin{bmatrix} 3.812 & 1.2060 & -1.7370 & -3.281 \\ 1.483 & 0.4693 & -0.6757 & -1.276 \end{bmatrix}$$

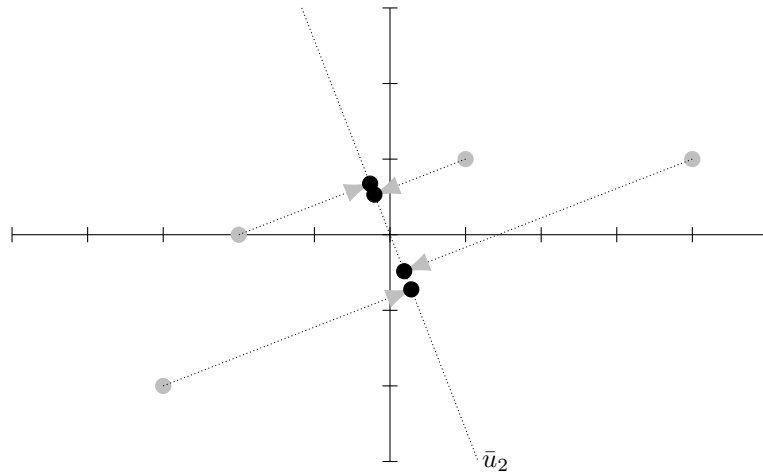
The result is the projection of the points onto \bar{u}_1 :



If we set $s_1 = 0$ and recalculate the matrix product:

$$\begin{bmatrix} -0.932 & -0.3625 \\ -0.3625 & 0.932 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1.312 & 0 & 0 \end{bmatrix} V^T = \begin{bmatrix} 0.1878 & -0.2064 & -0.2628 & 0.2815 \\ -0.4829 & 0.5307 & 0.6757 & -0.7236 \end{bmatrix}$$

The result is the projection of the points onto \bar{u}_2 :



Moreover the calculation from earlier also tells us the meaning of the s_i . If we take the square of the norm of each side, keeping in mind that the vectors in U and in V are orthonormal, we see:

$$\begin{aligned}
\|\bar{a}_1\|^2 &= s_1^2 v_{11}^2 + s_2^2 v_{12}^2 + \dots + s_k^2 v_{1k}^2 \\
\|\bar{a}_2\|^2 &= s_1^2 v_{21}^2 + s_2^2 v_{22}^2 + \dots + s_k^2 v_{2k}^2 \\
&\vdots \\
\|\bar{a}_n\|^2 &= s_1^2 v_{n1}^2 + s_2^2 v_{n2}^2 + \dots + s_k^2 v_{nk}^2
\end{aligned}$$

And then if we add these equations:

$$\|A\|^2 = \|\bar{a}_1\|^2 + \|\bar{a}_2\|^2 + \dots + \|\bar{a}_n\|^2 = s_1^2 + s_2^2 + \dots + s_k^2$$

This tells us that the total variation in all of the columns of A equals the sum of the squares of the singular values.

Moreover since each s_i corresponds to \bar{u}_i , this tells us that the total variation in all of the columns of A in the direction of \bar{u}_i equals s_i^2 and that the proportion of the total variation in all of the columns of A in the direction of \bar{u}_i equals

$$\frac{s_i^2}{s_1^2 + s_2^2 + \dots + s_k^2}$$

9.4 Summary and Matrix Comment

So where does this leave us? In summary so far, for an $m \times n$ matrix A when we find the singular value decomposition $A = U\Sigma V^T$ then:

- (a) The \bar{u}_i break down the directions in which the data is spread out.
- (b) The order of the directions $\bar{u}_1, \dots, \bar{u}_m$ gives us the order in which the variance of the data goes from largest to smallest
- (c) In the direction \bar{u}_i the variance of the data equals s_i^2 and the proportion of the total variance of the data equals $s_i^2 / (s_1^2 + s_2^2 + \dots + s_k^2)$.
- (d) If we create Σ' by taking Σ and replacing some of the s_i by 0, and if we then recalculate $A' = U\Sigma'V^T$, the result will be the orthogonal projection of the columns of A onto the columns of U corresponding to the s_i which remain.

As a consequence of the above:

- (a) We can simplify the data (meaning keeping only the large-scale trends with the most variance) by retaining only the large singular values, thereby keeping only the primary directions of variance in which the columns of A .

- (b) If we retain only s_1, \dots, s_j (with $j < k$) then the total variance in the data that is preserved can be calculated as:

$$\frac{s_1^2 + \dots + s_j^2}{s_1^2 + \dots + s_k^2}$$

- (c) If we wish to simplify the data while retaining a certain amount of variance we can choose how many singular values to preserve accordingly.
- (d) The columns in the matrix A' will generally be close to the columns in A , which translates to the matrix A' being close to A because $\|A - A'\|$ is small.

Let's try to summarize with some detailed examples.

Example 9.3. Consider the matrix with SVD:

$$A = \begin{bmatrix} -1 & 1 & 2 \\ -1 & 2 & 2 \end{bmatrix} = \begin{bmatrix} -0.6287 & -0.7777 \\ -0.7777 & 0.6287 \end{bmatrix} \begin{bmatrix} 3.8287 & 0 & 0 \\ 0 & 0.5840 & 0 \end{bmatrix} V^T$$

The total variance in the data is:

$$3.8287^2 + 0.5840^2 = 15$$

The left-singular vector of A corresponding to $s_1 = 3.8287$ captures

$$\frac{3.8287^2}{3.8287^2 + 0.5840^2} = 0.9773 = 97.73\%$$

of the total variance in the data, meaning that the data is extremely spread out in that direction.

The left-singular vector of A corresponding to $s_2 = 0.5840$ captures

$$\frac{0.5840^2}{3.8287^2 + 0.5840^2} = 0.0227 = 2.27\%$$

of the total variance in the data, meaning that the data is not very spread out in that direction.

We can see that if we let Σ' be Σ but with the smaller singular value set to zero and recalculate:

$$\begin{aligned}
A' = U\Sigma'V^T &= \begin{bmatrix} -0.6287 & -0.7777 \\ -0.7777 & 0.6287 \end{bmatrix} \begin{bmatrix} 3.8287 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} V^T \\
&= \begin{bmatrix} -0.8841 & 1.3730 & 1.7683 \\ -1.0937 & 1.6984 & 2.1873 \end{bmatrix}
\end{aligned}$$

The result is almost A in the sense that the columns of A have been rebuilt using only the first building-block vector. Take a second to look at the values in A and A' . They're close to one another:

$$A = \begin{bmatrix} -1 & 1 & 2 \\ -1 & 2 & 2 \end{bmatrix} \approx \begin{bmatrix} -0.8841 & 1.3730 & 1.7683 \\ -1.0937 & 1.6984 & 2.1873 \end{bmatrix}$$

Example 9.4. Consider the matrix:

$$A = \begin{bmatrix} 1 & 2 & 4 & 3 \\ -1 & 0 & 2 & 5 \\ 0 & 1 & 5 & 5 \end{bmatrix}$$

The singular value decomposition $A = U\Sigma V^T$ has:

$$\begin{aligned}
A &= U\Sigma V^T \\
&= \begin{bmatrix} -0.5098 & 0.6299 & -0.586 \\ -0.4949 & -0.7719 & -0.3991 \\ -0.7037 & 0.08654 & 0.7052 \end{bmatrix} \begin{bmatrix} 10.1205 & 0 & 0 & 0 \\ 0 & 2.8502 & 0 & 0 \\ 0 & 0 & 0.6722 & 0 \end{bmatrix} V^T
\end{aligned}$$

The total variance in the data is:

$$10.1205^2 + 2.8502^2 + 0.6722^2 = 111$$

The variance captured by each of the three directions is:

- The left-singular vector of A corresponding to $s_1 = 10.12$ captures

$$\frac{10.1205^2}{10.1205^2 + 2.8502^2 + 0.6722^2} = 0.9227 = 92.27\%$$

of the total variance.

- The left-singular vector of A corresponding to $s_2 = 2.85$ captures

$$\frac{2.8502^2}{10.1205^2 + 2.8502^2 + 0.6722^2} = 0.0732 = 7.32\%$$

of the total variance.

- The left-singular vector of A corresponding to $s_3 = 0.6722$ captures

$$\frac{0.6722^2}{10.1205^2 + 2.8502^2 + 0.6722^2} = 0.0041 = 0.41\%$$

of the total variance.

Moreover if we wanted to ignore the last direction and project all the points onto the first two directions we simply take Σ , change the 0.6722 to 0 to get Σ' , and recalculate.

$$A' = U\Sigma'V^T = \begin{bmatrix} 0.8905 & 1.7265 & 4.2251 & 2.8669 \\ -1.0746 & -0.1863 & 2.1533 & 4.9094 \\ 0.1318 & 1.3291 & 4.7290 & 5.1602 \end{bmatrix}$$

The new matrix is said to preserve

$$\frac{10.1205^2 + 2.8502^2}{10.1205^2 + 2.8502^2 + 0.6722^2} = 0.9959 = 99.59\%$$

of the variance of the original matrix. Really compare the values in the A and A' . They are quite close!

$$A = \begin{bmatrix} 1 & 2 & 4 & 3 \\ -1 & 0 & 2 & 5 \\ 0 & 1 & 5 & 5 \end{bmatrix} \approx \begin{bmatrix} 0.8905 & 1.7265 & 4.2251 & 2.8669 \\ -1.0746 & -0.1863 & 2.1533 & 4.9094 \\ 0.1318 & 1.3291 & 4.7290 & 5.1602 \end{bmatrix}$$

Example 9.5. Consider the matrix:

$$A = \begin{bmatrix} 1 & 2 & 3 & 0 & 5 & 2 \\ 3 & 4 & -1 & 2 & 3 & 4 \\ 5 & 4 & 1 & 2 & 0 & 0 \\ 6 & 4 & 1 & 2 & 3 & -1 \\ 6 & 3 & 1 & 2 & 3 & 4 \end{bmatrix}$$

Suppose we wanted to simplify the columns of this matrix as much as possible while retaining 95% of the variance.

We calculate the singular value decomposition $A = U\Sigma V^T$ and find that the singular values are $\{15.0759, 5.8017, 4.2418, 2.1720, 1.5318\}$

The total variance in the data is:

$$15.0759^2 + 5.8017^2 + 4.2418^2 + 2.1720^2 + 1.5318^2 = 286$$

Noting that:

$$\frac{15.0759^2 + 5.8017^2}{15.0759^2 + 5.8017^2 + 4.2418^2 + 2.1720^2 + 1.5318^2} = 0.9124$$

and

$$\frac{15.0759^2 + 5.8017^2 + 4.2418^2}{15.0759^2 + 5.8017^2 + 4.2418^2 + 2.1720^2 + 1.5318^2} = 0.9753$$

We see that we will need to preserve the largest three singular values and we will retain 97.53% of the variance.

If Σ' is Σ but with the smallest two replaced by zero then our simpler matrix is:

$$A' = U\Sigma'V^T = \begin{bmatrix} 1.163 & 1.728 & 2.755 & -0.006 & 5.214 & 1.916 \\ 3.618 & 3.171 & -0.361 & 1.829 & 2.643 & 4.425 \\ 5.335 & 3.419 & 0.310 & 2.007 & 0.579 & -0.262 \\ 6.012 & 4.081 & 1.771 & 1.924 & 2.428 & -0.632 \\ 5.172 & 4.147 & 0.425 & 2.202 & 3.269 & 3.564 \end{bmatrix}$$

Notice that this matrix is fairly close to the original A .

9.5 Data Modeling and Prediction

Consider the following matrix where each column is considered a point in \mathbb{R}^3 :

$$A = \begin{bmatrix} 1 & 15 & 23 & 28 & 3 & 17 & 14 & 0 & 10 & 4 \\ 4 & 26 & 48 & 33 & 6 & 38 & 36 & 0 & 22 & 8 \\ 10 & 98 & 142 & 191 & 21 & 111 & 91 & 2 & 62 & 25 \end{bmatrix}$$

If we find the singular value decomposition of this we get:

$$A = \underbrace{\begin{bmatrix} -0.15 & 0.02 & 0.99 \\ -0.26 & 0.96 & -0.06 \\ -0.95 & -0.27 & -0.14 \end{bmatrix}}_U \underbrace{\begin{bmatrix} 317.49 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 24.88 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.26 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_\Sigma V^T$$

Notice that u_1 captures

$$\frac{317.49^2}{317.49^2 + 24.88^2 + 1.26^2} = 0.9938790675 = 99.38790675\%$$

of all variance. This means the data is essentially one dimensional. More specifically it means that the data is basically spanned by \bar{u}_1 and so all of the data more or less looks like

$$c_1 \bar{u}_1 = c_1 \begin{bmatrix} -0.1459 \\ -0.2604 \\ -0.9544 \end{bmatrix}$$

for various values of c_1 .

So now suppose we know $y = 100$ and we wish to know x and z . We solve $-0.2604c_1 = 100$ to get $c_1 = -384.0246$ and then calculate

$$-384.0246 \bar{u}_1 \approx \begin{bmatrix} 56.0483 \\ 99.9895 \\ 366.5182 \end{bmatrix}$$

So that $x \approx 56.0483$ and $z \approx 366.5182$.

9.6 Centering Comment

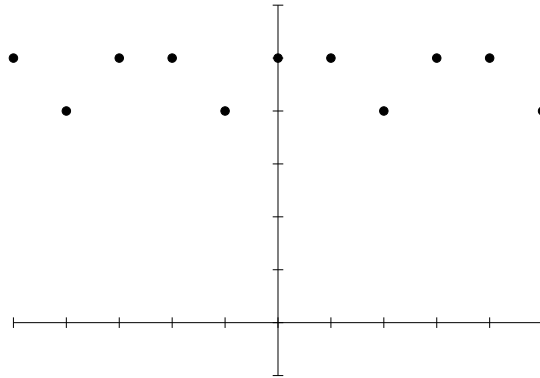
In the real world if data is analyzed this way usually the first step is to center it around the origin by subtracting the average of all columns of A from each column of A . This is because, if the data is not centered around the origin in a particularly nasty way, the vectors in U don't really make sense for the data.

We are completely ignoring that in order to avoid an extra step and to keep the values in A nice, as mostly the data we are using is already essentially around the origin in all the examples. There is one homework question which steps through the analysis with the centering process.

Example 9.6. The points in this matrix:

$$A = \begin{bmatrix} -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 \\ 5 & 4 & 5 & 5 & 4 & 5 & 5 & 4 & 5 & 5 & 4 \end{bmatrix}$$

When plotted look like this:



The singular value decomposition has:

$$A = \begin{bmatrix} -0.0155 & 1.0 \\ 1.0 & 0.0155 \end{bmatrix} \begin{bmatrix} 15.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 10.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} V^T$$

These vectors don't make intuitive sense at all for the data.

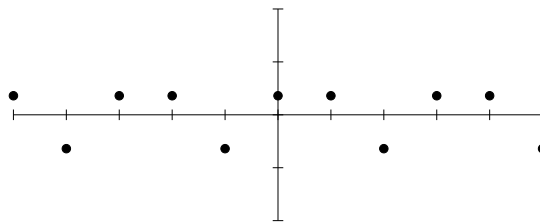
If we center the data, subtracting the average of all points

$$\begin{bmatrix} 0 \\ 4.6364 \end{bmatrix}$$

from each point, we get the centered data:

$$\begin{bmatrix} -5.0 & -4.0 & -3.0 & -2.0 & -1.0 & 0 & 1.0 & 2.0 & 3.0 & 4.0 & 5.0 \\ 0.364 & -0.636 & 0.364 & 0.364 & -0.636 & 0.364 & 0.364 & -0.636 & 0.364 & 0.364 & -0.636 \end{bmatrix}$$

Which, when plotted, looks like:



The singular value decomposition now has:

$$A = \begin{bmatrix} -1.0 & 0.0186 \\ 0.0186 & 1.0 \end{bmatrix} \begin{bmatrix} 10.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1.58 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} V^T$$

which makes much more sense if we think of these vectors centered at the center of the data.

9.7 Matlab

Throwing out a singular value and recalculating is easy too:

```
>> A=[
1 2 0
3 1 4
1 1 -1
0 1 2
]
A =
     1     2     0
     3     1     4
     1     1    -1
     0     1     2
>> [U,S,V] = svd(A)
U =
 -0.2079    0.7246   -0.3584   -0.5507
 -0.9171   -0.1139    0.3788   -0.0501
 -0.0103    0.6615    0.2666    0.7009
 -0.3400   -0.1560   -0.8106    0.4506
S =
 5.5200         0         0
         0  2.5538         0
         0         0  1.4170
         0         0         0
V =
 -0.5380    0.4090    0.7371
 -0.3049    0.7208   -0.6225
 -0.7859   -0.5596   -0.2631
>> SP=S;SP(3,3)=0;
>> U*SP*transpose(V)
ans =
 1.3743    1.6839   -0.1336
 2.6044    1.3341    4.1412
 0.7216    1.2351   -0.9006
 0.8466    0.2851    1.6979
```

9.8 Exercises

Exercise 9.1. Find the (vector) direction in which the following set of points have the most variance, then plot the points and the corresponding line.

$$(1, 2), (4, 3), (-1, -3), (-5, -8)$$

Exercise 9.2. Find the (vector) direction in which the following set of points have the most variance, then plot the points and the corresponding line.

$$(1, -1), (3, -2), (6, -4), (10, -4)$$

Exercise 9.3. For each of the following matrices find the SVD and identify the direction in which the column data is most spread out, as well as the variance and proportion of total variance in that direction.

$$(a) A = \begin{bmatrix} 5.0 & 6.0 & 5.0 & 5.0 & 6.0 \\ 6.0 & 5.0 & 5.0 & 6.0 & 5.0 \\ 5.0 & 6.0 & 5.0 & 6.0 & 5.0 \end{bmatrix}$$

$$(b) A = \begin{bmatrix} 0 & 1.0 & 1.0 & 2.0 & 2.0 \\ 1.0 & 2.0 & 2.0 & 0 & 1.0 \\ 7.0 & 8.0 & 8.0 & 8.0 & 9.0 \\ 8.0 & 9.0 & 9.0 & 8.0 & 7.0 \\ 7.0 & 7.0 & 7.0 & 9.0 & 7.0 \end{bmatrix}$$

$$(c) A = \begin{bmatrix} 4.0 & 5.0 & 0 & 0 & 0 \\ 5.0 & 5.0 & 0 & 0 & 0 \\ 0 & 1.0 & 7.0 & 8.0 & 8.0 \\ 0 & 0 & 8.0 & 7.0 & 7.0 \end{bmatrix}$$

$$(d) A = \begin{bmatrix} 2.0 & 3.0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.0 & 0 & 0 & 0 \\ 3.0 & 3.0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5.0 & 5.0 & 5.0 & 5.0 \end{bmatrix}$$

Exercise 9.4. Given the matrix

$$A = \begin{bmatrix} 2.0 & 3.0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.0 & 0 & 0 & 0 \\ 3.0 & 3.0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 5.0 & 5.0 & 5.0 & 5.0 \end{bmatrix}$$

- Find the singular value decomposition of A .
- Find an approximation A' to A preserving the largest two singular values only. What proportion of the total variance is preserved?

Exercise 9.5. Given the matrix

$$A = \begin{bmatrix} 1 & 2 & -1 & 0 & 5 \\ 0 & 1 & 1 & -1 & 6 \\ -1 & 0 & -2 & 1 & 4 \\ 2 & 2 & 1 & 2 & -5 \\ 2 & 1 & 0 & 0 & -5 \end{bmatrix}$$

- (a) Find the singular value decomposition of A .
- (b) Find an approximation A' to A preserving the largest three singular values only. What proportion of the total variance is preserved?

Exercise 9.6. Find an approximation to the following matrix which preserves at least 95% of the total variance using as few singular values as possible:

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

Exercise 9.7. Consider the points stored in the columns of this matrix:

$$\begin{bmatrix} 2.0 & -0.1 & 5.9 & 2.3 & 2.9 & -5.1 & 1.9 & 3.6 \\ 3.2 & 4.8 & -2.2 & 5.5 & 0.9 & -1.7 & 3.0 & 3.6 \\ 17.0 & 13.0 & 1.5 & 21.0 & 3.7 & -18.0 & 15.0 & 15.0 \end{bmatrix}$$

These points mostly lie close to a plane through the origin.

- (a) Find the equation of this plane by finding the two directions with the most variance, finding a normal vector for the plane, then finding the equation.
- (b) Using this equation predict which z -value would correspond to $x = 10$ and $y = -3$.

Exercise 9.8. Consider the points stored in the columns of this matrix:

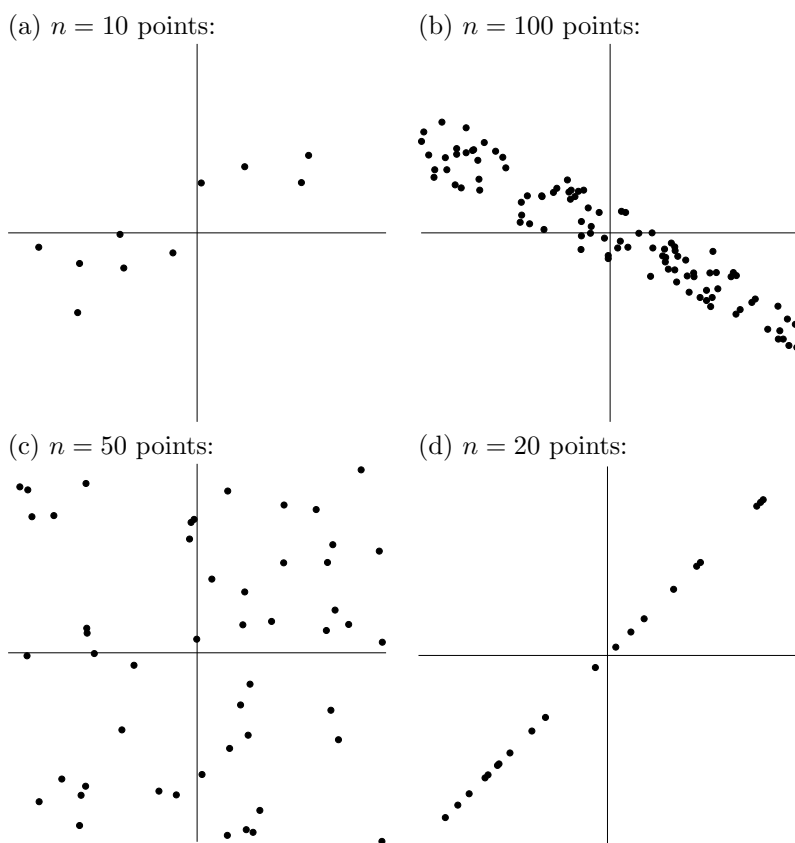
$$\begin{bmatrix} 1.9 & 0.9 & 2.6 & 6.1 & 3.4 & -0.5 & 4.5 & 6.0 \\ 5.3 & 11.0 & 3.9 & 7.0 & 8.2 & 9.2 & 12.0 & 2.6 \\ -0.5 & 1.6 & -1.6 & -13.0 & -10.0 & 3.3 & -18.0 & -8.8 \end{bmatrix}$$

These points mostly lie close to a plane but not through the origin.

- (a) Find the mean of all the points. This is basically where all the points are centered.
- (b) Center the points around the origin by subtracting this mean from each point.

- (c) Find the two directions with the most variance and use these to find a normal vector for the plane.
- (d) Find the equation of the plane using this normal vector and also the mean.
- (e) Using this equation predict x so that $(x, 1, -2)$ is on the plane, predict y so that $(0, y, 10)$ is on the plane, and predict z so that $(3, 4, z)$ is on the plane.

Exercise 9.9. Suppose n points are placed as columns in a $2 \times n$ matrix. If $A = U\Sigma V^T$ is the SVD and if the points are plotted below, say as much as you can about the matrices U and Σ



Exercise 9.10. Suppose a 100×100 matrix has 100 singular values whose squares add to 9512 and in decreasing order are $\{82.2, 27.6, 23.3, 19.1, 16.3, 8.5, 5.3, \dots\}$. How many singular values must be preserved in order to keep 90% of the data variance? How about 80%?

Exercise 9.11. Explain the mathematical process by which you would approx-

imate a 200×200 matrix while keeping 99% of the variance.