Matrices, Geometry&*Mathematica*

Authors: Bruce Carpenter, Bill Davis and Jerry Uhl ©2001

Producer: Bruce Carpenter

Publisher: Math Everywhere, Inc.

MGM.10 The Spectral Theorem for Symmetric Matrices and the Holy Grail of Matrix Theory

BASICS

B.1) The Spectral Theorem says that if A is a square symmetric matrix ($A = A^t$) hitting on hitdimD, then there is an orthonormal basis of hitdimD consisting of unit eigenvectors of A.

□B.1.a.i) The Spectral Theorem says that if A is a square symmetric matrix ($A = A^t$) hitting on hitdimD, then there is an orthonormal basis of hitdimD consisting of unit eigenvectors of A.

Talk a little bit about the spectral theorem.

□Answer:

The spectral theorem says that if you have symmetric matrix A (this means $A = A^t$), then there is an orthonormal basis of hitdimD consisting of unit eigenvectors of A.

Try it out on a random symmetric matrix A:

```
hitdim = Random[Integer, {3, 5}];
Clear[b, i, j];
b[i_, j_] := Random[Real, {-4, 4}]
B = Table[b[i, j], {i, 1, hitdim}, {j, 1, hitdim}];
A = Transpose[B] .B;
MatrixForm[A]
```

$$\begin{pmatrix} 9.69059 & 7.47987 & 6.06028 \\ 7.47987 & 19.0295 & -10.9031 \\ 6.06028 & -10.9031 & 31.8505 \end{pmatrix}$$

Let Mathematica calculate a linearly independent set of eigenvectors of A:

```
Clear[eigenvector, k];
eigenvector[k_] := Eigenvectors[A][[k]]
Table[eigenvector[k], {k, 1, hitdim}]
```

```
{{0.0618609, -0.475391, 0.877597},
 {-0.643887, -0.690849, -0.328843}, {0.762616, -0.544731, -0.348835}}
```

This set of eigenvectors of A forms an orthonormal basis of hitdimD as the following calculation confirms:

```
Table[Table[eigenvector[j].eigenvector[k], {k, 1, hitdim}],
 {j, 1, hitdim}]
```

```
{{1., 0, 0}, {0, 1., 0}, {0, 0, 1.}}
```

This output signals that if $k \neq j$, then eigenvector[k] is perpendicular to eigenvector[j] and

each of the calculated eigenvectors is a unit vector.

Try it for more random symmetric matrices:

```
hitdim = Random[Integer, {3, 10}];
Clear[b, i, j];
b[i_, j_] := Random[Real, {-4, 4}]
B = Table[b[i, j], {i, 1, hitdim}, {j, 1, hitdim}];
A = Transpose[B] .B;
MatrixForm[A]
Clear[eigenvector, k];
eigenvector[k_] := Eigenvectors[A][[k]]
Table[eigenvector[k], {k, 1, hitdim}];
Table[Table[eigenvector[j].eigenvector[k], {k, 1, hitdim}],
 {j, 1, hitdim}]
```

$$\begin{pmatrix} 43.0112 & 14.5039 & -12.9369 & 4.46889 & 15.6972 & -12.9376 & -16.0832 \\ 14.5039 & 50.7956 & 12.2548 & -25.1204 & 31.6864 & -11.9879 & 1.31127 \\ -12.9369 & 12.2548 & 28.8757 & -13.8104 & 12.4414 & -4.47383 & 11.048 \\ 4.46889 & -25.1204 & -13.8104 & 39.7176 & -3.68839 & 26.6258 & 0.945939 \\ 15.6972 & 31.6864 & 12.4414 & -3.68839 & 37.5907 & 4.65675 & 11.274 \\ -12.9376 & -11.9879 & -4.47383 & 26.6258 & 4.65675 & 37.0281 & 14.6097 \\ -16.0832 & 1.31127 & 11.048 & 0.945939 & 11.274 & 14.6097 & 26.9227 \end{pmatrix}$$

```
{{1., 0, 0, 0, 0, 0, 0}, {0, 1., 0, 0, 0, 0, 0},
 {0, 0, 1., 0, 0, 0, 0}, {0, 0, 0, 1., 0, 0, 0},
 {0, 0, 0, 0, 1., 0, 0}, {0, 0, 0, 0, 0, 1., 0}, {0, 0, 0, 0, 0, 0, 1.}}
```

Rerun as much as you want to.

□B.1.a.ii) Explanation of the Spectral Theorem

How about an explanation of where the Spectral Theorem comes from?

□Answer:

Hold your horses.

You will get a full explanation later in this lesson.

Learning how to use the Spectral Theorem is more important than knowing its proof.

B.2) The Spectral Theorem as the holy grail of matrix theory:

If A is any matrix, then $A^t.A$ is symmetric.

For any matrix A, the Spectral theorem gives you an orthonormal basis

$$\{X_1, X_2, X_3, X_4, ..., X_{hitdim}\}$$

of hitdimD consisting of unit eigenvectors of $A^t.A$.

You can get the alignerframe, the stretch factors and the hangerframe for A directly from these vectors.

□B.2.a.i) If A is any matrix, then $A^t.A$ is symmetric

Explain this:

If A is any matrix, then $A^t.A$ is symmetric.

□Answer:

Go with any matrix A and put $B = A^t.A$

Now note that:

$$B^t = (A^t.A)^t = A^t.(A^t)^t = A^t.A = B.$$

□B.2.a.ii) Using the Spectral Theorem to $A^t.A$ to unlock the secrets of SVD analysis of A

For any matrix A, the Spectral theorem gives you an orthonormal basis

$$\{X_1, X_2, X_3, X_4, ..., X_{hitdim}\}$$

of hitdimD consisting of **unit** eigenvectors of $A^t.A$.

Group and (if necessary relabel) these vectors into two groups:

Group I = $\{X_1, X_2, X_3, ..., X_p\}$ for which $A.X_j \neq \{0,0,...0\}$ for all j = 1,2,...,p.

Group 2 = $\{X_p, X_{p+2}, X_{p+3}, ..., X_{hitdim}\}$ for which $A.X_j = \{0,0,...\}$ for all j = p+1,p+2, ...,hitdimD.

Facts of the matter:

- An SVD alignerframe for A,
  {alignerframe[1],alignerframe[2],.alignerframe[3], . .,alignerframe[p]}

  is
  $$\{X_1, X_2, X_3, ..., X_p\}.$$
- The corresponding non-zero SVD stretch factors for A,
  {stretch[1],stretch[2],stretch[3],.....,stretch[p]},

  are
  $$\{ \| A.X_1 \|, \| A.X_2 \|, \| A.X_3 \|, ..., \| A.X_p \| \}.$$
- The corresponding SVD hangerframe for A,
  {hangerframe[1],hangerframe[2], . . .,hangerframe[p]},

  is
  $$\{ \tfrac{1}{\|A.X_1\|} A.X_1, \tfrac{1}{\|A.X_2\|} A.X_2, \tfrac{1}{\|A.X_3\|} A.X_3, . . ., \tfrac{1}{\|A.X_p\|} A.X_p \}.$$

Go with this matrix A:

```
hitdim = 8;
hangdim = 5;
A = Table[Random[Real, {-4, 4}], {i, 1, hangdim}, {j, 1, hitdim}];
MatrixForm[A]
```

$$\begin{pmatrix} -0.724982 & 2.75088 & -2.89746 & 2.26021 & -3.25357 & -0.689179 & -1.8510 \\ -3.13321 & -0.596952 & -3.28279 & 1.42886 & 1.75119 & -0.985254 & 3.5387 \\ -3.12608 & -0.366071 & -3.50024 & -2.05367 & 0.904197 & 1.90251 & -0.4997 \\ -2.37082 & 3.15163 & -1.6023 & -2.51006 & -3.11725 & -0.159192 & -3.751 \\ -3.98404 & -3.56224 & 3.53158 & -0.39405 & -1.73523 & 1.42301 & 3.9927 \end{pmatrix}$$

And use unit eigenvectors of $A^t.A$ to generate the full SVD of A.

□Answer:

The Spectral theorem gives you an orthonormal basis

$$\{X_1, X_2, X_3, X_4, ..., X_{hitdim}\}$$

of hitdimD consisting of **unit** eigenvectors of $A^t.A$.

Here they are:

```
Clear[eigenvector, k];
eigenvector[k_] := Eigenvectors[Transpose[A].A][[k]];
ColumnForm[Table[eigenvector[k], {k, 1, hitdim}]]
```

```
{-0.0612676, -0.551821, 0.396205, 0.0435278, 0.32543, 0.101044, 0.61628
{-0.469148, -0.0729769, 0.147408, 0.0647837, -0.41008, 0.00515767, 0.23
{0.637273, 0.0261766, 0.673269, 0.125147, -0.166337, -0.111545, -0.1771
{-0.259035, -0.119924, 0.366288, -0.67536, -0.216003, 0.355622, -0.331
{-0.0540848, 0.239526, 0.106126, -0.5373, 0.525231, -0.522554, 0.077939
{0.355869, -0.633514, -0.410821, -0.341361, -0.295814, -0.292757, -0.0
{0.0982053, -0.00604828, -0.0714084, -0.233943, -0.609436, -0.544568,
{-0.467733, 0.360164, 0.418477, 0.455668, 0.400308, 0.147468, -0.284538
```

Group and (if necessary relabel) these vectors into two groups:

Group I = $\{X_1, X_2, X_3, \ldots, X_p\}$ for which $A.X_j \neq \{0,0,\ldots0\}$ for all $j = 1,2,\ldots,p$.

Group 2 = $\{X_p, X_{p+2}, X_{p+3}, \ldots, X_{\text{hitdim}}\}$ for which $A.X_j = \{0,0,\ldots\}$
for all $j = p+1, p+2, \ldots$,hitdimD.

Find out what p is:

```
ColumnForm[Table[A.eigenvector[k], {k, 1, hitdim}]]
```

```
{-5.47382, 1.7407, -0.183325, -6.25027, 4.98049}
{3.33588, 1.9465, -2.3549, 3.048, 6.7244}
{-0.297714, -4.62073, -5.75089, -0.939461, -0.102899}
{-2.41243, -3.39999, 2.40301, 2.57408, 1.85573}
{-1.28488, 0.913754, -0.836527, 1.07846, -0.408885}
{0, 0, 0, 0, 0}
{0, 0, 0, 0, 0}
{0, 0, 0, 0, 0}
```

So:

```
p = 5
```

```
5
```

- An SVD alignerframe for A,

  {alignerframe[1],alignerframe[2],.alignerframe[3], . . .,alignerframe[p]}

is

$\{X_1, X_2, X_3, \ldots, X_p\}$:

```
p = 5;
alignerframe = Table[eigenvector[k], {k, 1, p}]
```

```
{{-0.0612676, -0.551821, 0.396205,
  0.0435278, 0.32543, 0.101044, 0.616283, 0.192219},
 {-0.469148, -0.0729769, 0.147408, 0.0647837, -0.41008,
  0.00515767, 0.230833, -0.726066}, {0.637273, 0.0261766,
  0.673269, 0.125147, -0.166337, -0.111545, -0.177119, -0.229706},
 {-0.259035, -0.119924, 0.366288, -0.67536, -0.216003, 0.355622,
  -0.331494, 0.212669}, {-0.0540848, 0.239526, 0.106126,
  -0.5373, 0.525231, -0.522554, 0.0779393, -0.291106}}
```

- The corresponding SVD stretch factors for A,

  {stretch[1],stretch[2],stretch[3],....,stretch[p]},

are

$\{ \| A.X_1 \|, \| A.X_2 \|, \| A.X_3 \|, \ldots, \| A.X_p \| \}$:

```
stretches = Table[Norm[A.eigenvector[k]], {k, 1, p}]
```

```
{9.84365, 8.65854, 7.4435, 5.76401, 2.12506}
```

- The corresponding SVD hangerframe for A,

  {hangerframe[1],hangerframe[2], . . .,hangerframe[p]},

is

$\left\{ \frac{1}{\|A.X_1\|} A.X_1, \frac{1}{\|A.X_2\|} A.X_2, \frac{1}{\|A.X_3\|} A.X_3, \ldots, \frac{1}{\|A.X_p\|} A.X_p \right\}$.

```
hangerframe =
 Table[1/Norm[A.eigenvector[k]] A.eigenvector[k], {k, 1, p}]
```

```
{{-0.556076, 0.176835, -0.0186236, -0.634955, 0.505959},
 {0.38527, 0.224807, -0.271974, 0.352022, 0.776621},
 {-0.0399965, -0.620773, -0.772606, -0.126212, -0.013824},
 {-0.418533, -0.589865, 0.416898, 0.446579, 0.321951},
 {-0.60463, 0.429989, -0.393648, 0.507495, -0.192411}}
```

Check it out:

```
aligner = alignerframe;
MatrixForm[aligner]
```

```
-0.0612676  -0.551821   0.396205  0.0435278   0.32543    0.101044    0.6
 -0.469148  -0.0729769  0.147408  0.0647837  -0.41008    0.00515767  0.2
  0.637273   0.0261766  0.673269  0.125147   -0.166337  -0.111545   -0.
 -0.259035  -0.119924   0.366288 -0.67536    -0.216003   0.355622   -0.
-0.0540848   0.239526   0.106126 -0.5373      0.525231  -0.522554    0.0
```

```
stretcher = DiagonalMatrix[stretches];
MatrixForm[stretcher]
```

```
9.84365   0         0        0         0
0         8.65854   0        0         0
0         0         7.4435   0         0
0         0         0        5.76401   0
0         0         0        0         2.12506
```

```
hanger = Transpose[hangerframe];
MatrixForm[hanger]
```

```
 -0.556076    0.38527   -0.0399965  -0.418533  -0.60463
  0.176835    0.224807  -0.620773   -0.589865   0.429989
 -0.0186236  -0.271974  -0.772606    0.416898  -0.393648
 -0.634955    0.352022  -0.126212    0.446579   0.507495
  0.505959    0.776621  -0.013824    0.321951  -0.192411
```

```
MatrixForm[hanger.stretcher.aligner]
```

```
-0.724982   2.75088   -2.89746   2.26021  -3.25357  -0.689179  -1.8510
 -3.13321  -0.596952  -3.28279   1.42886   1.75119  -0.985254   3.5387
 -3.12608  -0.366071  -3.50024  -2.05367   0.904197  1.90251   -0.4997
 -2.37082   3.15163   -1.6023   -2.51006  -3.11725  -0.159192  -3.751
 -3.98404  -3.56224    3.53158  -0.39405  -1.73523   1.42301    3.9927
```

```
MatrixForm[A]
```

```
-0.724982   2.75088   -2.89746   2.26021  -3.25357  -0.689179  -1.8510
 -3.13321  -0.596952  -3.28279   1.42886   1.75119  -0.985254   3.5387
 -3.12608  -0.366071  -3.50024  -2.05367   0.904197  1.90251   -0.4997
 -2.37082   3.15163   -1.6023   -2.51006  -3.11725  -0.159192  -3.751
 -3.98404  -3.56224    3.53158  -0.39405  -1.73523   1.42301    3.9927
```

On the money.

Big time math happens again.

□B.2.a.iii) Explanations

For any matrix A, the Spectral theorem gives you an orthonormal basis

$\{X_1, X_2, X_3, X_4, \ldots, X_{\text{hitdim}}\}$

of hitdimD consisting of **unit** eigenvectors of $A^t.A$.

Group and (if necessary relabel) these vectors into two groups:

Group I = $\{X_1, X_2, X_3, \ldots, X_p\}$ for which $A.X_j \neq \{0,0,\ldots0\}$ for all $j = 1,2,\ldots,p$.

Group 2 = $\{X_p, X_{p+2}, X_{p+3}, \ldots, X_{\text{hitdim}}\}$ for which $A.X_j = \{0,0,\ldots\}$
for all $j = p+1, p+2, \ldots$,hitdimD.

Facts of the matter:

- An SVD alignerframe for A,

  {alignerframe[1],alignerframe[2],.alignerframe[3], . . .,alignerframe[p]}

is

$\{X_1, X_2, X_3, \ldots, X_p\}$.

- The corresponding SVD stretch factors for A,

  {stretch[1],stretch[2],stretch[3],....,stretch[p]},

are

$\{ \| A.X_1 \|, \| A.X_2 \|, \| A.X_3 \|, \ldots, \| A.X_p \| \}$.

- The corresponding SVD hangerframe for A,

  {hangerframe[1],hangerframe[2], . . .,hangerframe[p]},

is

$\left\{ \frac{1}{\|A.X_1\|} A.X_1, \frac{1}{\|A.X_2\|} A.X_2, \frac{1}{\|A.X_3\|} A.X_3, \ldots, \frac{1}{\|A.X_p\|} A.X_p \right\}$.

- And when you go with these assignments, you are fully guaranteed that

$$A.X = \sum_{k=1}^{p} \text{stretch}[k]\,(X.\text{alignerframe}[k])\ \text{hangerframe}[k]$$

for any X in hitdimD

Explain why all this works.

□ Answer:

This is a little bit of mathematical bean-counting.

- Why the advertised aligner fame $\{X_1, X_2, X_3, \ldots, X_p\}$ is a perpendicular frame.

For an explanation, click on the right

Reason:The Spectral theorem gives you an orthonormal basis

$\{X_1, X_2, X_3, X_4, \ldots, X_{\text{hitdim}}\}$ is an orthonormal basis (perpendicular frame)

of hitdimD consisting of **unit** eigenvectors of $A^t.A$.

So $\{X_1, X_2, X_3, \ldots, X_p\}$ is a perpendicular frame

- Why the advertised stretch factors

  stretch[k] = $\| A.X_k \|$ for $k = 1,2, \ldots, p$

are all positive.

For an explanation, click on the right

Reason:

stretch[k] = $\| A.X_k \|$ for $k = 1,2, \ldots, p$ are all positive

because

$A.X_k \neq \{0,0, \ldots, 0\}$ for $k = 1,2, \ldots, p$

- Why the vectors in the advertised hanger frame

  $\left\{ \frac{1}{\|A.X_1\|} A.X_1, \frac{1}{\|A.X_2\|} A.X_2, \frac{1}{\|A.X_3\|} A.X_3, \ldots, \frac{1}{\|A.X_p\|} A.X_p \right\}$.

are all unit vectors.

For an explanation, click on the right

Go with any k with $k = 1,2, \ldots, p$, and look at

and look at

$$\left(\frac{1}{||A.X_k||}A.X_k\right).\left(\frac{1}{||A.X_k||}A.X_k\right)$$

$$= \frac{1}{||A.X_k||^2}(A.\ X_k).(A.X_k)$$

$$= \frac{1}{||A.X_k||^2}\ ||\ A.X_k\ \ ||^2$$

$$= 1.$$

▪ Why the vectors in the advertised hanger frame

$$\{\frac{1}{||A.X_1||}A.X_1,\ \frac{1}{||A.X_2||}A.X_2,\frac{1}{||A.X_3||}A.X_3,. . .,\frac{1}{||A.X_p||}A.X_p\}.$$

are orthonormal (perpendicular to each other).

For an explanation, click on the right

Go with any k with k = 1,2,. . .,p, and any j with j = 1,2,. . .,p and j ≠ p

and look at

$$\left(\frac{1}{||A.X_k||}A.X_k\right).\left(\frac{1}{||A.X_j||}A.X_j\right)$$

$$= \frac{1}{||A.X_k||\ ||A.X_j|||}(A.\ X_k).(A.X_j)$$

$$= \frac{1}{||A.X_k||\ ||A.X_j|||}(\ X_k).(A^tA.X_j)$$

This comes from the transpose manipulation which says that
$(A.X).(Y) = (X).(A^tY)$
This manipulation will be explained in a Basic below..

$$= \frac{1}{||A.X_k||\ ||A.X_j|||}(\ X_k).(\text{eigenvalue}\ X_j)$$

X_j is an eigenvector of A^tA; so
$(A^tA.X_j) = \text{eigenvalue}\ X_j$

$$= \frac{\text{eigenvalue}}{||A.X_k||\ ||A.X_j||}(\ X_k).(X_j)$$

$$= \frac{\text{eigenvalue}}{||A.X_k||\ ||A.X_j||}\ 0\ = 0.$$

because X_j is perpendicular to X_k

▪ Why when you go with

{alignerframe[1],alignerframe[2],.alignerframe[3], . .,alignerframe[p]}

$$= \{X_1,X_2,X_3,....,X_p\}.$$

and

{stretch[1],stretch[2],stretch[3],....,stretch[p]}

$$= \{\ ||\ A.X_1\ ||,\ ||\ A.X_2\ ||,\ ||\ A.X_3\ ||,....,\ ||\ A.X_p\ ||\}.$$

$$X = \sum_{k=1}^{\text{hitdim}}(X.X_k)\,X_k$$

Hit both sides with A to get

$$A.X = \sum_{k=1}^{\text{hitdim}}(X.X_k)\,A.X_k$$

$$= \sum_{k=1}^{p}(X.X_k)\,A.X_k$$

Reason: $A.X_k$ = {0,0,0,. . .,0} for k = p+1,p+2,....,hitdim.

$$= \sum_{k=1}^{p}(X.\text{alignerframe}[k])\ A.\text{alignerframe}[k]$$

Reason: X_k = alignerframe[k] for k = 1,2,...,p.

$$= \sum_{k=1}^{p}(X.\text{alignerframe}[k])\ \text{stretch}[k]\ \text{hangerframe}[k]$$

Reason: A. alignerframe[k] = stretch[k] hangerframe[k] for k = 1,2,...,p.

$$= \sum_{k=1}^{p}\text{stretch}[k]\ (X.\text{alignerframe}[k])\ \text{hangerframe}[k]$$

B.3) The Spectral Theorem as the holy grail of Matrix theory:

If A is a matrix (symmetric or not symmetric) that hits on hitdimD and if $\{X_1, X_2, X_3, X_4, ..., X_{hitdim}\}$ is an orthonormal basis of unit eigenvectors of $A^t.A$ (as guaranteed by the Spectral Theorem), then you can read off

→ an orthonormal basis of the column space R[A] of A ,

→ an orthonormal basis of the null space N[A] of A,

→ an orthonormal basis of the row space R[A^t] of A,

and

→ the construction of the PseudoInverse of A

In this problem, you get to see how the Spectral Theorem reveals the secrets behind any matrix.
Here are the relevant terms:

□ **Column space R[A] of a matrix A**

Given a matrix A, the column space R[A] is the subspace of hangdimD consisting of all possible hits with A.

R[A] is spanned by the vertical columns of A.

and

{hangerframe[1],hangerframe[2], . . .,hangerframe[p]},

$$= \{\frac{1}{||A.X_1||}A.X_1,\ \frac{1}{||A.X_2||}A.X_2,\frac{1}{||A.X_3||}A.X_3,. . .,\frac{1}{||A.X_p||}A.X_p\}.$$

you are guaranteed that

A.alignerframe[k] = stretch[k] hangerframe[k]

for all k = 1,2,3,. . .,p,

For an explanation, click on the right

Because alignerframe[k] = X_k, you can be sure that

$$A.\text{alignerframe}[k] = A.X_k$$

$$= \frac{||A.X_2||}{||A.X_2||}\ A.X_k$$

$$= ||A.X_k||\ \frac{1}{||A.X_k||}\ A.X_k$$

$$= \text{stretch}[k]\ \text{hangerframe}[k]$$

because stretch[k] = $||A.X_k||$ and hangerframe[k] = $\frac{1}{||A.X_k||}\ A.X_k$.

▪ Why when you go with

{alignerframe[1],alignerframe[2],.alignerframe[3], . .,alignerframe[p]}

$$= \{X_1,X_2,X_3,....,X_p\}.$$

and

{stretch[1],stretch[2],stretch[3],....,stretch[p]}

$$= \{\ ||\ A.X_1\ ||,\ ||\ A.X_2\ ||,\ ||\ A.X_3\ ||,....,\ ||\ A.X_p\ ||\}.$$

and

{hangerframe[1],hangerframe[2], . . .,hangerframe[p]},

$$= \{\frac{1}{||A.X_1||}A.X_1,\ \frac{1}{||A.X_2||}A.X_2,\frac{1}{||A.X_3||}A.X_3,. . .,\frac{1}{||A.X_p||}A.X_p\}.$$

you are guaranteed that for any X in hitdimD

$$A.X = \sum_{k=1}^{p}\text{stretch}[k]\ (X.\text{alignerframe}[k])\ \text{hangerframe}[k].$$

For an explanation, click on the right

The Spectral theorem gives you an orthonormal basis

$$\{X_1,X_2,X_3, X_4,....,X_{hitdim}\}$$

of hitdimD consisting of **unit** eigenvectors of $A^t.A$. Because this is a basis of hitdimD, it spans all of hitdim D, so you can resolve X into its perpendicular components in the directions of these basis vectors.

□ **Row space R[A^t] of a matrix A**

Given matrix A, the row space R[A^t] is the subspace of hitdimD consisting of all possible hits with A^t.

R[A^t] is spanned by the horizontal rows of A.

□ **Null space N[A] of a matrix A**

Given a matrix A, the null space N[A] is the subspace of hitdimD consisting of all X's with A.X = {0,0,...,0}

□ **The pseudo inverse of a matrix A**

Given a matrix A, and a vector Y in hangdimD, X = PseudoInverse[A].Y is the vector of smallest norm in hitdimD with the property that A.X is as close to Y as you can get by hitting with A.
If A is invertible, then PseudoInverse[A] = A^{-1}.

□ **B.3.a.i) Using an orthonormal basis$\{X_1,X_2,X_3, X_4,....,X_{hitdim}\}$ unit eigenvectors of $A^t.A$ to read off**

→ an orthonormal basis of the column space R[A] of A ,

→ an orthonormal basis of the null space N[A] of A,

→ an orthonormal basis of the row space R[A^t] of A,

and

→ the construction of the PseudoInverse of A.

Given a matrix A hitting on hitdimD and hanging in hangdim D, you already know that $A^t.A$ is a symmetric matrix which hits on hitdimD and hangs in hitdimD.

The Spectral theorem gives you an orthonormal basis $\{X_1,X_2,X_3, X_4,....,X_{hitdim}\}$ of hitdimD consisting of unit eigenvectors of $A^t.A$. Group and (if necessary relabel) these vectors into two groups.

Group I = $\{X_1,X_2,X_3,....,X_p\}$ for which $A.X_j$ ≠ {0,0,...0} for all j = 1,2,...,p.

Group 1 gives an SVD aligner frame for A.

Group 2 = $\{X_{p+1},X_{p+2},X_{p+3},....,X_{hitdim}\}$ for which $A.X_j$ = {0,0,...} for all j = p+1,p+2, ...,hitdimD.

The facts of the matter are:

▪ The rank of A is p.

▪ To get an orthonormal basis of the column space R[A], you just go with

$$\{\frac{A.X_1}{||A.X_1||},\ \frac{A.X_2}{||A.X_2||},\ \frac{A.X_3}{||A.X_3||},\ ...,\ \frac{A.X_p}{||A.X_p||}\}.$$

The upshot: SVD hangerframe for A is an orthonormal basis for R[A].

▪ To get an orthonormal basis of the row space R[A^t], you just go with $\{X_1,X_2,X_3,....,X_p\}$.

The upshot: SVD alignerframe (corresponding to non-zero stretch factors) for A is an orthonormal basis for $R[A^t]$.

- To get an orthonormal basis of N[A], you just go with $\{X_{p+1},X_{p+2},X_3,....,X_{hitdim}\}$.

These are the eigenvectors of $A^t.A$ that are hit into {0,0,...,0} by A.

- For any Y in hangdimD, you can calculate PseudoInverse[A].Y (or A^{-1}.Y if A is invertible) by going with

$$\text{PseudoInverse}[A].Y=\sum_{j=1}^{p}\frac{Y\bullet(A.X_j)}{\|A.X_j\|^2}X_j$$

Here's a matrix A.

```
A = Table[Random[Real, {-5, 5}], {i, 1, 4}, {j, 1, 6}];
MatrixForm[A]
```

$$\begin{pmatrix} -3.26145 & -2.76364 & 4.36629 & 3.33247 & 0.608307 & -0.141776 \\ -0.00900726 & 3.64478 & -1.42817 & 0.918688 & -3.00614 & 1.78236 \\ -2.5316 & -3.88232 & -3.31713 & 0.488843 & -2.55155 & -4.42952 \\ -2.73162 & -4.01859 & 4.61749 & -1.20829 & -2.7226 & -4.78397 \end{pmatrix}$$

Use the information above to come up with quick calculations of:

- The rank of A
- An orthonormal basis the row space of A and the dimension of the row space of A,
- An orthonormal basis of the null space of A and the dimension of the null space of A,
- An orthonormal basis of the column space of A and the dimension of the column space of A,

and

- PseudoInverse[A].Y for any Y in hangdimD.

Answer:

Look at A again:

```
MatrixForm[A]
```

$$\begin{pmatrix} -3.26145 & -2.76364 & 4.36629 & 3.33247 & 0.608307 & -0.141776 \\ -0.00900726 & 3.64478 & -1.42817 & 0.918688 & -3.00614 & 1.78236 \\ -2.5316 & -3.88232 & -3.31713 & 0.488843 & -2.55155 & -4.42952 \\ -2.73162 & -4.01859 & 4.61749 & -1.20829 & -2.7226 & -4.78397 \end{pmatrix}$$

This matrix hits on 6D and hangs in 4D.

```
hitdim = 6
hangdim = 4
```

6

4

Look at:an orthonormal basis $\{X_1,X_2,X_3, X_4,....,X_{hitdim}\}$ of hitdimD consisting of unit eigenvectors of $A^t.A$.

```
Clear[eigenvector, k];
eigenvector[k_] := Eigenvectors[Transpose[A].A][[k]];
ColumnForm[Table[eigenvector[k], {k, 1, hitdim}]]
```

```
{0.401995, 0.612286, -0.368661, -0.0418864, 0.183639, 0.540478}
{-0.0276326, 0.0650035, 0.812641, 0.201161, 0.378896, 0.388069}
{-0.38648, 0.365316, 0.0651125, 0.477568, -0.671771, 0.183275}
{0.259972, 0.339426, 0.430684, -0.641342, -0.430354, -0.187593}
{0.29876, 0.425713, 0.0879591, 0.462654, 0.21141, -0.680464}
{-0.787243, 0.221058, -0.10835, -0.485827, 0.254631, 0.137032}
```

See which are hit into {0,0,. . . ,0} by A:

```
Table[A.eigenvector[k], {k, 1, hitdim}]
```

```
{{-4.71741, 3.12734, -5.05498, -8.29592},
 {4.30454, -1.18595, -5.46544, 0.435466},
 {1.69203, 4.02682, 0.479843, 0.263456},
 {-2.27789, 0.989852, -1.78904, 2.75857}, {0, 0, 0, 0}, {0, 0, 0, 0}}
```

The first four are non-zero.

This signals that the rank of A is 4.:

An orthonormal basis of $R[A^t]$, the row space of A is $\{X_1,X_2,X_3, X_4\}$ ::

```
p = 4;
rowspacebasis = Table[eigenvector[k], {k, 1, p}]
```

```
{{0.401995, 0.612286, -0.368661, -0.0418864, 0.183639, 0.540478},
 {-0.0276326, 0.0650035, 0.812641, 0.201161, 0.378896, 0.388069},
 {-0.38648, 0.365316, 0.0651125, 0.477568, -0.671771, 0.183275},
 {0.259972, 0.339426, 0.430684, -0.641342, -0.430354, -0.187593}}
```

The dimension of the row space of A is:

```
p
```

4

An orthonormal basis of N[A], the null space of A is $\{X_5, X_6\}$:

```
nullspacebasis = Table[eigenvector[k], {k, p + 1, hitdim}];
ColumnForm[nullspacebasis]
```

```
{0.29876, 0.425713, 0.0879591, 0.462654, 0.21141, -0.680464}
{-0.787243, 0.221058, -0.10835, -0.485827, 0.254631, 0.137032}
```

The dimension of the null space of A is:

```
hitdim - p
```

2

An orthonormal basis {Y[1],Y[2],...,Y[p]} of R[A], the column space of A is:

```
Clear[Y, k];
```

$$Y[k\_] := \frac{A.eigenvector[k]}{\sqrt{(A.eigenvector[k]).(A.eigenvector[k])}}$$

```
colspacebasis = Table[Y[k] , {k, 1, p}]
```

```
{{-0.41958, 0.278154, -0.449604, -0.737861},
 {0.608776, -0.167724, -0.77296, 0.0615866},
 {0.384375, 0.914763, 0.109005, 0.0598486},
 {-0.552811, 0.240223, -0.434174, 0.669467}}
```

The dimension of the column space of A is:

```
p
```

4

PseudoInverse[A].Y for any Y = {y[1],y[2],...,y[hangdim]} in hangdimD is

$$\text{PseudoInverse}[A.Y]=\sum_{j=1}^{p}\frac{Y\bullet(A.X_j)}{\|A.X_j\|^2}X_j:$$

```
Clear[Y, y];
Y = Table[y[k], {k, 1, hangdim}];
```

$$\text{Expand}\left[\sum_{k=1}^{p}\left(\frac{Y.(A.eigenvector[k])}{(A.eigenvector[k]).(A.eigenvector[k])}\right)eigenvector[k]\right]$$

```
{-0.0860052 y[1] - 0.0545555 y[2] - 0.0500175 y[3] + 0.0103607 y[4],
 -0.0308916 y[1] + 0.109308 y[2] - 0.0583092 y[3] + 0.0204966 y[4],
 0.0316293 y[1] + 0.010242 y[2] - 0.117861 y[3] + 0.102131 y[4],
 0.146625 y[1] + 0.0560435 y[2] + 0.0590872 y[3] - 0.093205 y[4],
 0.0248473 y[1] - 0.169131 y[2] - 0.0200525 y[3] - 0.0878044 y[4],
 0.0544124 y[1] + 0.031315 y[2] - 0.0397311 y[3] - 0.0600767 y[4]}
```

Check:

```
PseudoInverse[A].Y
```

```
{-0.0860052 y[1] - 0.0545555 y[2] - 0.0500175 y[3] + 0.0103607 y[4],
 -0.0308916 y[1] + 0.109308 y[2] - 0.0583092 y[3] + 0.0204966 y[4],
 0.0316293 y[1] + 0.010242 y[2] - 0.117861 y[3] + 0.102131 y[4],
 0.146625 y[1] + 0.0560435 y[2] + 0.0590872 y[3] - 0.093205 y[4],
 0.0248473 y[1] - 0.169131 y[2] - 0.0200525 y[3] - 0.0878044 y[4],
 0.0544124 y[1] + 0.031315 y[2] - 0.0397311 y[3] - 0.0600767 y[4]}
```

Nailed it cold.

B.5.a.ii) Explanations

Explain where all this good stuff comes from.

Answer:

All of it comes from the Spectral theorem applied to a symmetric matrix

Given a matrix A hitting on hitdimD and hanging in hangdim D, you already know that $A^t.A$ is a symmetric matrix which hits on hitdimD and hangs in hitdimD.

The Spectral theorem gives you an orthonormal basis $\{X_1,X_2,X_3, X_4,....,X_{hitdim}\}$ of hitdimD consisting of unit eigenvectors of $A^t.A$. Group and (if necessary relabel) these vectors into two groups.

Group I = $\{X_1,X_2,X_3,....,X_p\}$ for which $A.X_j \neq \{0,0,...0\}$ for all j = 1,2,...,p.

These are also eigenvectors of $A^t.A$ whose corresponding eigenvalues are non -zero.

Group 2 = $\{X_{p+1},X_{p+2},X_{p+3},....,X_{hitdim}\}$ for which $A.X_j = \{0,0,...\}$ for all j = p+1,p+2, ...,hitdimD.

To get an orthonormal basis of column space of A, R[A], you just go with

$$\left\{\frac{A.X_1}{\|A.X_1\|}, \frac{A.X_2}{\|A.X_2\|}, \frac{A.X_3}{\|A.X_3\|}, ..., \frac{A.X_p}{\|A.X_p\|}\right\}.$$

The rank of A is p.

For a detailed explanation, click on the right.

The Spectral theorem gives you an orthonormal basis $\{X_1,X_2,X_3, X_4,....,X_{hitdim}\}$ of hitdimD consisting of unit eigenvectors of $A^t.A$.

$\{X_1,X_2,X_3,....,X_p\}$ consists of the unit eigenvectors for which $A.X_j \neq \{0,0,...\}$ for all j = 1,2,...,p.

Group 1 gives an SVD aligner frame for A. So

$$\left\{\frac{A.X_1}{\|A.X_1\|}, \frac{A.X_2}{\|A.X_2\|}, \frac{A.X_3}{\|A.X_3\|}, ..., \frac{A.X_p}{\|A.X_p\|}\right\}.$$

is an SVD hanger frame for A. This perpendicular frame automatically spans R[A].

This also reveals that the rank of A is p.

To get an orthonormal basis of the row space of A,

$R[A^t]$, you just go with $\{X_1, X_2, X_3, ..., X_p\}$.

The rank of A^t is also p and the dimension of the row space of A is p

For a detailed explanation, click on the right.

Note first:

The (dimension of $R[A^t]$) = (rank of A^t) = (rank of A) = dimension of R[A] = p.

Also note that

$\{X_1,X_2,X_3,....,X_p\}$ spans a p-dimensional subspace of hitdimD.

Now, if you can explain why each of the vectors in $\{X_1,X_2,X_3,....,X_p\}$ are also in $R[A^t]$ (the subspace consisting of all hits with A^t, you will have explained why $\{X_1,X_2,X_3,....,X_p\}$ is an orthonormal basis of $R[A^t]$.

To this end, take X_j with $1 \leq j \leq p$ and remember that X_j is and eigenvector of $A^t.A$ with corresponding eigenvalue λ_j .

This tells you (and everyone else) that

$$(A^t.A).X_j = \lambda_j X_j .$$

This is the same as saying

$$A^t.\left(\frac{A.X_j}{\lambda_j}\right) = X_j .$$

This tells you that to get X_j , all you have to do its to hit $\left(\frac{A.X_j}{\lambda_j}\right)$ with A^t.

This signals that X_j is in $R[A^t]$.

The upshot: The orthonormal set $\{X_1,X_2,X_3,....,X_p\}$ is in $R[A^t]$.

Explanation complete.

□ **To get an orthonormal basis of the null space of A, N[A], you just go with $\{X_{p+1},X_{p+2},....,X_{hitdim}\}$.**

The dimension of the null space of A is hitdim - p

For a detailed explanation, click on the right.

$\{X_1,X_2,X_3, X_4,....,X_{hitdim}\}$ is an orthonormal basis of hitdimD.

Any X in hitdimD is of this form:

$$X = \sum_{j=1}^{hitdim} (X \bullet X_j)\, X_j.$$

So

$$A.X = \sum_{j=1}^{hitdim} (X \bullet X_j)\, A.X_j$$

$$= \sum_{j=1}^{p} (X \bullet X_j)\,(A.X_j)$$

because $A.X_j$ = {0, 0, ..., 0} for j= p+1, p+2, ..., hitdim.

Now remember that

$$(A.X_i) \bullet (A.X_j) = 0 \text{ for } i \neq j$$

and calculate

$$(A.X) \bullet (A.X) = \sum_{j=1}^{p} \| A.X_j \|^2 \ (X \bullet X_j)^2.$$

So saying that A.X = {0, 0, ..., 0} is the same as saying

$$\sum_{j=1}^{p} \| A.X_j \|^2 \ (X \bullet X_j)^2 = 0.$$

Because each term in the sum is non negative, saying that A.X = {0, 0, ..., 0}is the same as saying that

$$\| A.X_j \|^2 \ (X \bullet X_j)^2 = 0 \text{ for all } j = 1, 2,...p.$$

But $\| A.X_j \|^2 > 0$ for each j = 1,2,...p.

So: Saying that A.X = {0, 0, ..., 0} is the same as saying that

$$X \bullet X_j = 0 \text{ for all } j = 1, 2,...p.$$

To recapitulate:

Starting with

$$X = \sum_{j=1}^{hitdim} (X \bullet X_j)\, X_j,$$

saying that A.X = {0, 0, ..., 0} is the same as saying that

$$X = \sum_{j=p+1}^{hitdim} (X \bullet X_j)\, X_j$$

and this is the same as saying that X is in the subspace of hitdimD spanned by

$$\{X_{p+1},X_{p+2},X_{p+3},....,X_{hitdim}\}.$$

This explains why $\{X_{p+1},X_{2\,p+2},X_{p+3},....,X_{hitdim}\}$ is an orthonormal basis of N[A].

□ **For any Y in hangdimD, you can calculate PseudoInverse[A] (or A^{-1}if A is invertible) by going with**

$$\textbf{PseudoInverse[A.Y]} = \sum_{j=1}^{p} \frac{Y \bullet (A.X_j)}{\|A.X_j\|^2}\, X_j$$

For a detailed explanation, click on the right.

$\{ \frac{A.X_1}{\|A.X_1\|}, \frac{A.X_2}{\|A.X_2\|}, \frac{A.X_3}{\|A.X_3\|}, ..., \frac{A.X_p}{\|A.X_p\|}\}$ is an orthonormal basis of R[A], the subspace of hangdimD consisting of all possible hits with A. Accordingly if Y is any vector in hangdimD, then the vector in R[A] closest to Y is

$$\text{R[A]closest} = \sum_{j=1}^{p} \left(Y \bullet \frac{A.X_j}{\|A.X_j\|}\right) \frac{A.X_j}{\|A.X_j\|} = \sum_{j=1}^{p} \frac{Y \bullet (A.X_j)}{\|A.X_j\|^2} (A.X_j)$$

To get PseudoInverse[A].Y, you just erase the red A to get

$$\text{Xapprox} = \sum_{j=1}^{p} \frac{Y \bullet (A.X_j)}{\|A.X_j\|^2} X_j .$$

Now when you hit Xapprox with A, you get

$$A.\text{Xapprox} = A. \sum_{j=1}^{p} \frac{Y \bullet (A.X_j)}{\|A.X_j\|^2} X_j = \sum_{j=1}^{p} \frac{Y \bullet (A.X_j)}{\|A.X_j\|^2} (A.X_j) = \text{R[A]closest}.$$

To see why Xapprox is the vector X in hitdimD of smallest norm satisfying

A.X = R[A]closest,

notice that if X is any other vector with

A.X = R[A]closest,

then X - Xapprox = Z is in N[A]

So

X = Xapprox+ Z with Z in N[A]

But:

-> Xapprox is in the span of

$\{X_1,X_2,X_3,....,X_p\}$

and

Z is the span of

$\{X_p,X_{p+2},....,X_{hitdim}\}$.

This guarantees that

$\text{Xapprox} \bullet Z = 0$

(because all the vectors in the span of $\{X_p,X_2,X_3,....,X_{hitdim}\}$ are perpendicular to all vectors in the span of $\{X_1,X_2,X_3,....,X_p\}$.

So

$$\| \text{Xapprox} + Z \|^2 = (\text{Xapprox} + Z) \bullet (\text{Xapprox} + Z)$$
$$= \text{Xapprox} \bullet \text{Xapprox} + 2\,(\text{Xapprox} \bullet Z) + Z \bullet Z$$
$$= \text{Xapprox} \bullet \text{Xapprox} + 2\,(0) + Z \bullet Z$$
$$= \| \text{Xapprox} \|^2 + \| Z \|^2 > \| \text{Xtest} \|^2$$

unless Z = {0,0,...,0}.

B.4) For the record: The transpose manipulation: $Y \bullet (A.X) = (A^t.Y) \bullet X$

□ **B.3.a) The transpose manipulation: $Y \bullet (A.X) = (A^t.Y) \bullet X$**

The transpose manipulation says that if a matrix A hits on hitdimD and hangs in hangdimD, you can be sure that

$Y \bullet (A.X) = (A^t.Y) \bullet X$.

for all hitdimD vectors X and for all hangdimD vectors Y.

Try it out for random vectors X and Y and random matrices A:

```
hitdim = Random[Integer, {2, 7}];
hangdim = Random[Integer, {2, 7}];
Clear[i, j];
A =
  Table[Random[Real, {-10, 10}], {i, 1, hangdim}, {j, 1, hitdim}];
MatrixForm[A];
X = Table[Random[Real, {-10, 10}], {j, 1, hitdim}];
Y = Table[Random[Real, {-10, 10}], {i, 1, hangdim}];
Y.(A.X)
(Transpose[A].Y).X
```

```
871.166
871.166
```

Rerun a couple of times.

Although not terribly interesting in its own right, the transpose manipulation

$Y \bullet (A.X) = (A^t.Y) \bullet X$.

lies at the heart of a lot of theory.

For the record, explain why the outcome

$Y \bullet (A.X) = (A^t.Y) \bullet X$

is always guaranteed.

□ **Answer:**

Agree that the arbitrary vector X in hitdimD is:

X = {x[1], x[2], . . ., x[hitdim]}.

Agree that COL[j] is the jth vertical column of A.

Take an arbitrary Y in hangdimD and start bean-counting calculations:

$$A.X = \sum_{j=1}^{hitdim} x[j]\,COL[j].$$

So:

$$Y.(A.X) = \sum_{j=1}^{hitdim} x[j]\ (Y.COL[j]).$$

On the other hand because the columns of A are the rows of A^t ,

$$A^t.Y = \{(COL[1].Y), (COL[2].Y), \ldots, (COL[hitdim].Y)\}.$$

So

$$(A^t.Y).X = \sum_{j=1}^{hitdim} (Y.COL[j])\,x[j]$$

And now because

$$\sum_{j=1}^{hitdim} (Y.COL[j])\,x[j] = \sum_{j=1}^{hitdim} x[j]\ (Y.COL[j]).,$$

the inescapable conclusion is

$$Y\bullet(A.X) = (A^t.Y)\bullet X.$$

Explanation complete.

And you're out of here.

B.5) For the record: An explanation the Spectral Theorem:

If A is a symmetric kD matrix, then there is a perpendicular frame (orthonormal basis)

$$\{X_1, X_2, X_3, X_4, ..., X_k\}$$

of real eigenvectors of A which spans all of kD

□B.5.a.i) Visualization

Here's a random symmetric 2D matrix A:

```
Clear[i, j];
B = Table[Random[Real, {-4, 4}], {i, 1, 2}, {j, 1, 2}];
A = 1/2 (B + Transpose[B]);
MatrixForm[A]
```

$$\begin{pmatrix} 1.63687 & 1.40568 \\ 1.40568 & 0.252874 \end{pmatrix}$$

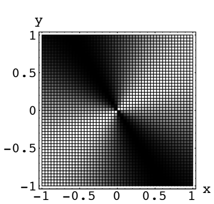
Make this function:

```
Clear[g, x, y];
g[x_, y_] = Expand[ ({x, y}.(A.{x, y})) / ({x, y}.{x, y}) ]
```

$$\frac{1.63687\,x^2}{x^2+y^2} + \frac{2.81137\,x\,y}{x^2+y^2} + \frac{0.252874\,y^2}{x^2+y^2}$$

Look at this graphic:

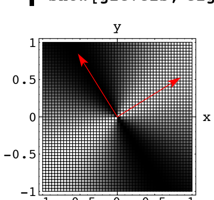
```
glevels = DensityPlot[Evaluate[g[x, y]], {x, -1, 1}, {y, -1, 1},
   Axes → True, AxesLabel → {"x", "y"}, PlotPoints → 50];
```



The lighter the square, the higher g[x,y] = $\frac{\{x,y\}.(A.\{x,y\})}{\{x,y\}.\{x,y\}}$ is.

Throw in the unit eigenvectors of A:

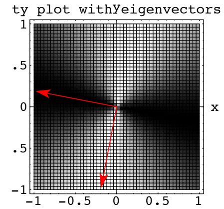
```
Clear[eigenvector];
{eigenvector[1], eigenvector[2]} = Eigenvectors[A];
eigenplot = Table[Arrow[eigenvector[k],
     Tail → {0, 0}, VectorColor → Red], {k, 1, 2}];
Show[glevels, eigenplot];
```



Here's another:

```
Clear[i, j];
B = Table[Random[Real, {-4, 4}], {i, 1, 2}, {j, 1, 2}];
A = 1/2 (B + Transpose[B]);
MatrixForm[A]
Clear[g, x, y];
g[x_, y_] = Expand[ ({x, y}.(A.{x, y})) / ({x, y}.{x, y}) ];
glevels = DensityPlot[Evaluate[g[x, y]],
   {x, -1, 1}, {y, -1, 1}, Axes → True, AxesLabel → {"x", "y"},
   PlotPoints → 50, DisplayFunction → Identity];
Clear[eigenvector];
{eigenvector[1], eigenvector[2]} = Eigenvectors[A];
eigenplot = Table[Arrow[eigenvector[k],
    Tail → {0, 0}, VectorColor → Red], {k, 1, 2}];
Show[glevels, eigenplot, PlotLabel ->
   "Density plot with eigenvectors of A",
  DisplayFunction → $DisplayFunction];
```

$$\begin{pmatrix} 1.01411 & 0.452616 \\ 0.452616 & 3.26855 \end{pmatrix}$$



The lighter the square, the higher g[x,y] = $\frac{\{x,y\}.(A.\{x,y\})}{\{x,y\}.\{x,y\}}$ is.

Rerun a couple of times.
From these graphics, you can see that one eigenvector of the symmetric matrix A points into the place where g[x,y] is biggest and the other eigenvector of A points into the place where g[x,y] is smallest.

What relevance does this have for the explanation of the Spectral Theorem?

□Answer:

These graphics set the strategy for the explanation of the Spectral Theorem.

For more, go on to the next part.

□B.5.a.ii) Explanation: The Spectral Theorem for symmetric matrices

The Spectral Theorem for symmetric matrices says this.
If A is a kD symmetric matrix, then you are guaranteed a perpendicular frame $\{X_1, X_2, X_3, X_4, ..., X_k\}$ of unit eigenvectors of A which spans all of kD .
Explain the reasoning behind the Spectral Theorem.

□Answer:

This full blown proof is adapted from Peter Lax's book "Linear Algebra" (Interscience, 1997).

Step 1: Coming up with an eigenvector X_1

Go with a kD symmetric matrix A and put for each kD vector X

$$g_1[X] = \frac{X\bullet(A.X)}{X\bullet X}.$$

Notice that

$$g_1[t\,X] = \frac{(t\,X)\bullet(A.(t\,X))}{(t\,X)\bullet(t\,X)} = \frac{t^2\,(X\bullet(A.X))}{t^2\,(X\bullet X)} = g[X]$$

so that $g_1[X]$ does not depend on the length of X.

The function $g_1[X]$ assumes a global maximum value on kD at a unit vector X_1.

You are going to see why X_1 is an eigenvector of A with corresponding eigenvalue $X_1 \bullet A.X_1$.

To this end, take any other kD vector Y and put

Put

$$f_1[t] = g_1[X_1 + t\,Y]$$

and notice that $t = 0$ maximizes $f_1[t]$ so that $f_1'[0] = 0$.

Now examine $f_1[t]$ noticing first that

$$f_1[t] = \frac{p_1[t]}{q_1[t]}$$

where

$$\begin{aligned} p_1[t] &= (X_1 + t\,Y)\bullet(A.(X_1 + t\,Y)) \\ &= X_1 \bullet A.X_1 + t\,X_1 \bullet (A.Y) + t\,Y\bullet(A.X_1) + t^2\,Y\bullet(A.Y) \\ &= X_1 \bullet A.X_1 + t\,X_1 \bullet (A.Y) + t\,(A^t.Y)\bullet X_1 + t^2\,Y\bullet(A.Y) \\ &= X_1 \bullet A.X_1 + t\,X_1 \bullet (A.Y) + t\,(A.Y)\bullet X_1 + t^2\,Y\bullet(A.Y) \\ &= X_1 \bullet A.X_1 + 2\,t\,Y\bullet(A.X_1) + t^2\,Y\bullet(A.Y) \end{aligned}$$

and

$$\begin{aligned} q_1[t] &= (X_1 + t\,Y)\bullet(X_1 + t\,Y) \\ &= X_1 \bullet X_1 + 2\,t\,Y\bullet X_1 + t^2\,Y\bullet Y \end{aligned}$$

The quotient rule from calculus says

$$f_1'[t] = \frac{q_1[t]\,p_1'[t] - p_1[t]\,q_1'[t]}{q[t]^2}.$$

At this stage you know $f_1'[0] = 0$ so

$$q_1[0]\,p_1'[0] - p_1[0]\,q_1'[0] = 0$$

.

Differentiate $p_1[t]$ and $q_1[t]$ with respect to t to get

$$p_1'[t] = 2\,Y\bullet(A.X_1) + 2\,t\,Y\bullet(A.Y)$$

$$q_1'[t] = 2\,Y\bullet X_1 + 2\,t\,Y\bullet Y$$

Read off:

$p_1[0] = X_1 \bullet (A.X_1)$, $\quad p_1'[0] = 2\,Y\bullet(A.X_1)$

$q_1[0] = X_1 \bullet X_1$ $\quad$ and $q_1'[0] = 2\,Y\bullet X_1$

Plug these into

$$q_1[0]\,p_1'[0] - p_1[0]\,q_1'[0] = 0$$

to get

$$(X_1 \bullet X_1)\,(2\,Y\bullet(A.X_1)) - (X_1 \bullet(A.X_1))\,(2\,Y\bullet X_1) = 0.$$

Because X_1 is a unit vector (so that $X_1 \bullet X_1 = 1$), this is the same as

$$(2\,Y\bullet(A.X_1)) - (X_1 \bullet(A.X_1))\,(2\,Y\bullet X_1) = 0.$$

Divide both sides by 2 to get

$$(Y\bullet(A.X_1)) - (X_1 \bullet(A.X_1))\,(Y\bullet X_1) = 0.$$

Put

$$e_1 = X_1 \bullet (A.X_1)$$

to get

$$(Y\bullet(A.X_1)) - e_1\,(Y\bullet X_1) = 0.$$

Factor out Y to get

$$Y\bullet((A.X_1) - e_1\,X_1) = 0.$$

Now remember that Y is any kD vector. This gives you the freedom to set

$$Y = (A.X_1) - e_1\,X_1.$$

to see that

$$\|A.X_1 - e_1\,X_1\|^2 = (A.X_1 - e_1\,X_1)\bullet(A.X_1 - e_1\,X_1) = 0.$$

This tells you and everyone else that

$$A.X_1 - e_1\,X_1 = \{0, 0, 0, \ldots 0\}.$$

This is the same as

$$A.X_1 = e_1\,X_1,$$

signalling that X_1 is an eigenvector of A with corresponding eigenvalue $e_1 = X_1 \bullet (A.X_1)$

Step 2: Coming up with a new eigenvector X_2 perpendicular to X_1

Stay with the same kD symmetric matrix A and put S_2 = perpendicular complement of the one dimensional subspace of kD spanned by the eigenvector X_1.

Put for each kD vector X in S_2

$$g_2[X] = \frac{X\bullet(A.X)}{X\bullet X}.$$

As in the part above, $g_2[X]$ does not depend on the length of X.

The function $g_2[X]$ assumes a global maximum value on S_2 at a unit vector X_2.

You are going to see why X_2 is an eigenvector of A and why X_2 is perpendicular to X_1.

Things to note:

1) X_2 is perpendicular to X_1.

Reason:

X_2 is in S_2 and everything is S_2 is perpendicular to X_1.

2) $A.X_2$ is perpendicular to X_1 so that $A.X_2$ is in S_2.

Reason:

$$\begin{aligned} X_1 \bullet (A.X_2) &= (A^t.X_1)\bullet X_2 = (A.X_1)\bullet X_2 \\ &= (e_1.X_1)\bullet X_2 = e_1\,(X_1 \bullet X_2) = 0 \end{aligned}$$

because X_2 is in S_2 and everything in S_2 is perpendicular to X_1.

Now for any vector Y in S_2, put

$$f_2[t] = g_2[X_2 + t\,Y]$$

and notice that $t = 0$ maximizes $f_2[t]$ so that $f_2'[0] = 0$.

Computations virtually identical to those in the first part reveal that for any Y in S_2

$$Y\bullet((A.X_2) - e_2\,X_2) = 0.$$

where $e_2 = X_2\,(A.X_2)$.

Now remember that Y is any vector in S_2. This gives you the freedom to set

$$Y = (A.X_2) - e_2\,X_2.$$

(because both $(A.X_2)$ and $e_2\,X_2$ are in S_2).

to see that

$$\|A.X_2 - e_2\,X_2\|^2 = (A.X_2 - e_2\,X_2)\bullet(A.X_2 - e_2\,X_2) = 0.$$

This tells you and everyone else that

$$A.X_2 - e_2\,X_2 = \{0, 0, 0, \ldots 0\}.$$

This is the same as

$$A.X_2 = e_2\,X_2,$$

signalling that X_2 is an eigenvector of A with corresponding eigenvalue $e_2 = X_2 \bullet (A.X_2)$

At this stage, you have two mutually perpendicular eigenvectors X_1 and X_2 of A.

Step 3: Coming up with a new eigenvector X_3 perpendicular to both X_1 and X_2

Stay with the same kD symmetric matrix A and put

S_3 = perpendicular complement of the two dimensional subspace of kD spanned by the eigenvectors X_1 and X_2.

Stay with the same kD symmetric matrix A and put for each kD vector X in S_3

$$g_3[X] = \frac{X\bullet(A.X)}{X\bullet X}.$$

As in the part above, $g_3[X]$ does not depend on the length of X.

The function $g_3[X]$ assumes a global maximum value on S_3 at a unitvector X_3.

You are going to see why X_3 is an eigenvector of A and why X_3 is perpendicular to both X_1 and X_2.

Things to note:

1) X_3 is perpendicular to both X_1 and X_2

Reason: X_3 is in S_3 and everything is S_3 is perpendicular to both X_1 and X_2.

□**2) $A.X_3$ is perpendicular to both X_1 and X_2 so that $A.X_3$ is in S_3.**

Reasons:

$X_1 \bullet (A.X_3) = (A^t.X_1) \bullet X_3 = (A.X_1) \bullet X_3$

$= (e_1.X_1) \bullet X_3 = e_1 (X_1 \bullet X_3) = 0$

because X_3 is in S_3 and everything in S_3 is perpendicular to X_1.

Also

$X_2 \bullet (A.X_3) = (A^t.X_2) \bullet X_3 = (A.X_2) \bullet X_3$

$= (e_2.X_2) \bullet X_3 = e_2 (X_2 \bullet X_3) = 0$

because X_3 is in S_3 and everything in S_3 is perpendicular to X_2.

This explains why $A.X_3$ is perpendicular to both X_1 and to X_2. And this puts $A.X_3$ into S_3

Now for any vector Y in S_3, put

$f_3[t] = g_3[X_3 + t\, Y]$

and notice that $t = 0$ maximizes $f_3[t]$ so that $f_3'[0] = 0$.

Computations virtually identical to those in the first part reveal that for any Y in S_3

$Y \bullet ((A.X_3) - e_3\ X_3) = 0.$

where $e_3 = X_3\ (A.X_3)$

Now remember that Y is any vector in S_2 . This gives you the freedom to set

$Y = (A.X_3) - e_3\ X_3 .$

(because both $(A.X_3)$ and $e_3\ X_3$ are in S_3).

to see that

$\|\ A.X_3 - e_3\ X_3\ \|^2 = (A.X_3 - e_3\ X_3) \bullet (A.X_3 - e_3\ X_3) = 0.$

This tells you and everyone else that

$A.X_3 - e_3\ X_3 = \{0, 0, 0, \ldots 0\}.$

This is the same as

$A.X_3 = e_3\ X_3 ,$

signalling that X_3 is an eigenvector of A with corresponding eigenvalue $e_3 = X_3 \bullet (A.X_3)$

At this stage, you have three mutually perpendicular eigenvectors X_1 , X_2 and X_3 of A.

□**Step 4 to Step k:**

To get the advertised orthonormal basis of eigenvectors of A, you continue the process started above until you run out of dimensions.

This results in a perpendicular frame

$\{X_1, X_2, X_3, X_4, \ldots, X_k\}$

of eigenvectors of A which spans all of kD - an orthonormal basis of kD consisting of eigenvectors of A.