Math for Computer Graphics
-Singular Value Decomposition

Advanced Computer Graphics

CS/Beihang University

吴壮志

zzwu@buaa.edu.cn

2020-09-28

我们希望读者能同意我们如下见解：即如果除了**微积分**和**微分方程**之外，还有什么数学领域是数学科学的基础的话，那就是**数字线性代数**。

----Lioyd N. Trefethen, David Bau III《数字线性代数》前言

Introduction

- The Singular Value Decomposition (SVD) is a topic rarely reached in undergraduate linear algebra courses and often skipped over in graduate courses.

- Consequently relatively few mathematicians are familiar with what M.I.T. Professor Gilbert Strang calls "absolutely a high point of linear algebra."

Matrix Action-Two different ways to compute a matrix times a point

- Column Way

$$\left(\begin{array}{c|c} 1 & 3 \\ -3 & 2 \end{array}\right)\begin{pmatrix} 3 \\ 1 \end{pmatrix} = 3\begin{pmatrix} 1 \\ -3 \end{pmatrix} + 1\begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ -9 \end{pmatrix} + \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 6 \\ -7 \end{pmatrix}$$

- Row Way

$$\left(\begin{array}{cc} 1 & 3 \\ \hline -3 & 2 \end{array}\right)\begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} 1 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 1 \end{pmatrix} \\ \begin{pmatrix} -3 \\ 2 \end{pmatrix} \cdot \begin{pmatrix} 3 \\ 1 \end{pmatrix} \end{pmatrix} = \begin{pmatrix} 6 \\ -7 \end{pmatrix}$$

Perpframes, Aligner/Hangers/Stretchers

- Perpframes（正交标架）
  - 2D: In 2D you need to perpendicular unit vectors to form a perpframe, $v_1 = \begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix} and \quad v_2 = \begin{pmatrix} -\sin\alpha \\ \cos\alpha \end{pmatrix}$
  - 3D: 3D perpframe consists of three mutually perpendicular unit vectors. Coming up with a 3D perpframe is a little trickier.
  - Lots of folks call perpframes by the name "orthonormal basis"（单位正交基）.

Perpframes, Aligner/Hangers/Stretchers

- Aligners
  - You get an aligner matrix by loading the vectors from a perpframe into the rows of the matrix

$$Aligner = \begin{pmatrix} \vec{v}_1 \\ \hline \vec{v}_2 \end{pmatrix}$$

  - The aligner matrix gets its name since it aligns the perpframe to the xy-axis.

$$\begin{pmatrix} \vec{v}_1 \\ \hline \vec{v}_2 \end{pmatrix}(\vec{v}_1) = \begin{pmatrix} \vec{v}_1 \bullet \vec{v}_1 \\ \hline \vec{v}_2 \bullet \vec{v}_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \begin{pmatrix} \vec{v}_1 \\ \hline \vec{v}_2 \end{pmatrix}(\vec{v}_2) = \begin{pmatrix} \vec{v}_1 \bullet \vec{v}_2 \\ \hline \vec{v}_2 \bullet \vec{v}_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Perpframes, Aligner/Hangers/Stretchers

- Hangers
  - You get a hanger matrix by loading the vectors from a perpframe into the columns of the matrix.

$$Hanger = \begin{pmatrix} \vec{v}_1 & \vec{v}_2 \end{pmatrix}$$

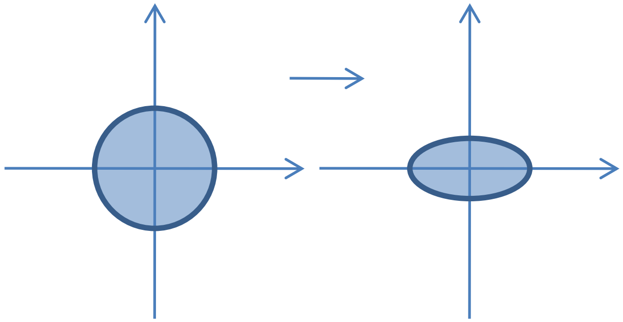
  - The hanger matrix gets its name since it hangs the xy-axis onto the perpframe.

$$\begin{pmatrix} \vec{v}_1 & \vec{v}_2 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \vec{v}_1 \qquad \begin{pmatrix} \vec{v}_1 & \vec{v}_2 \end{pmatrix}\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \vec{v}_2$$

Perpframes, Aligner/Hangers/Stretchers

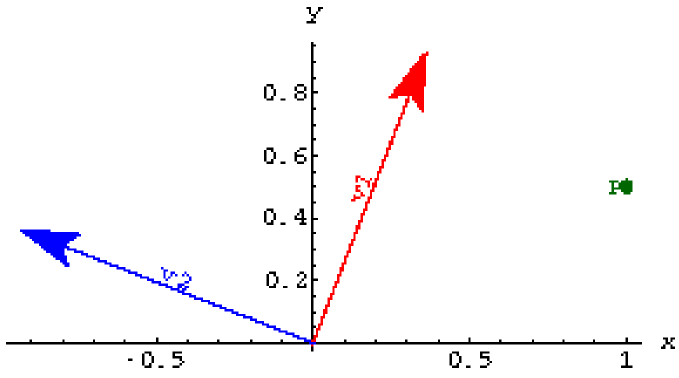
- Stretchers
  - The diagonal matrix $\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ stretches in the x direction by a factor of "a" and in the y direction by a factor of "b". You can verify this by hand using the column way to multiply a matrix times a vector:

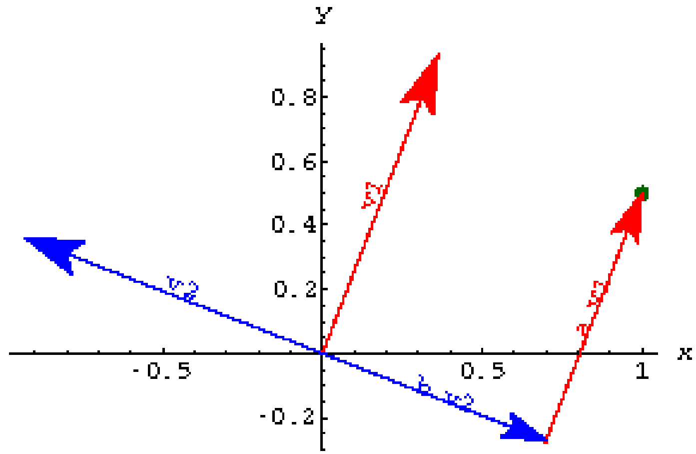
$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = x\begin{pmatrix} a \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ b \end{pmatrix} = \begin{pmatrix} ax \\ by \end{pmatrix}$$

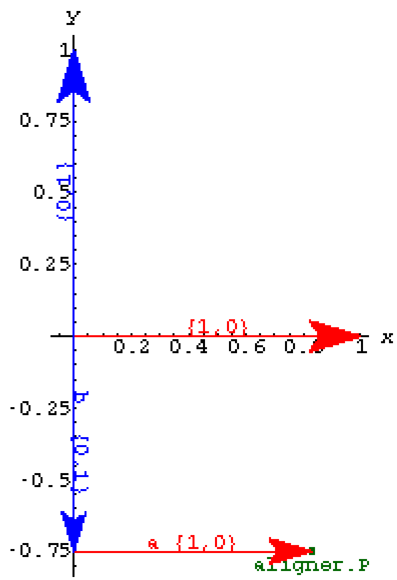


Coordinates

- **From standard coordinates to perpframe coordinates**
  - Find the numbers a and b which are the perpframe coordinates of the point P: $\vec{p} = a\vec{v}_1 + b\vec{v}_2$
  - You can get the numbers a and b by hitting the point P with the aligner matrix $Aligner = \begin{pmatrix} \vec{v}_1 \\ \hline \vec{v}_2 \end{pmatrix}$
  - The location of the rotated point aligner.P relative to the xy-axis is the same as the location of the original point P relative to the perpframe.


y
x
v1
v2
P
-0.5
0.5
1
0.2
0.4
0.6
0.8

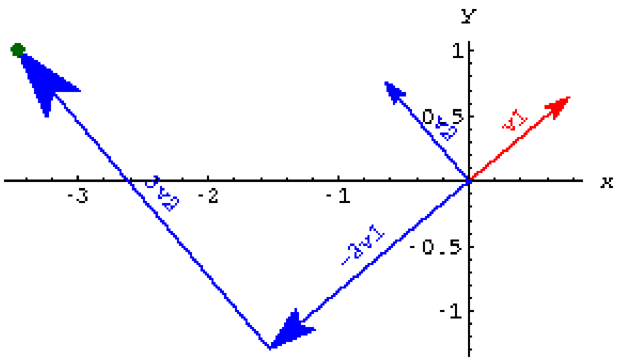

y
x
v1
v2
a v1
b v2
-0.5
0.5
1
-0.2
0.2
0.4
0.6
0.8


y
x
{1,0}
{0,1}
a {1,0}
b {0,1}
aligner.P
0.2
0.4
0.6
1
1
0.75
0.5
0.25
-0.25
-0.5
-0.75

Coordinates

- From perpframe coordinates to standard coordinates.
  - For the perpframe shown below, use a matrix hit to find the xy-coordinates of the point P that has perpframe coordinates {-2,3}.
  - The column method for matrix multiplication tells you that the standard xy-coordinates of the point P are given by

$$-2\vec{v}_1 + 3\vec{v}_2 = \begin{pmatrix} \vec{v}_1 & \vec{v}_2 \end{pmatrix} \begin{pmatrix} -2 \\ 3 \end{pmatrix}$$



Coordinates

- Summary
  - A hit with the aligner matrix takes standard coordinates to perpframe coordinates.
  - A hit with the hanger matrix takes perpframe coordinates to standard coordinates.

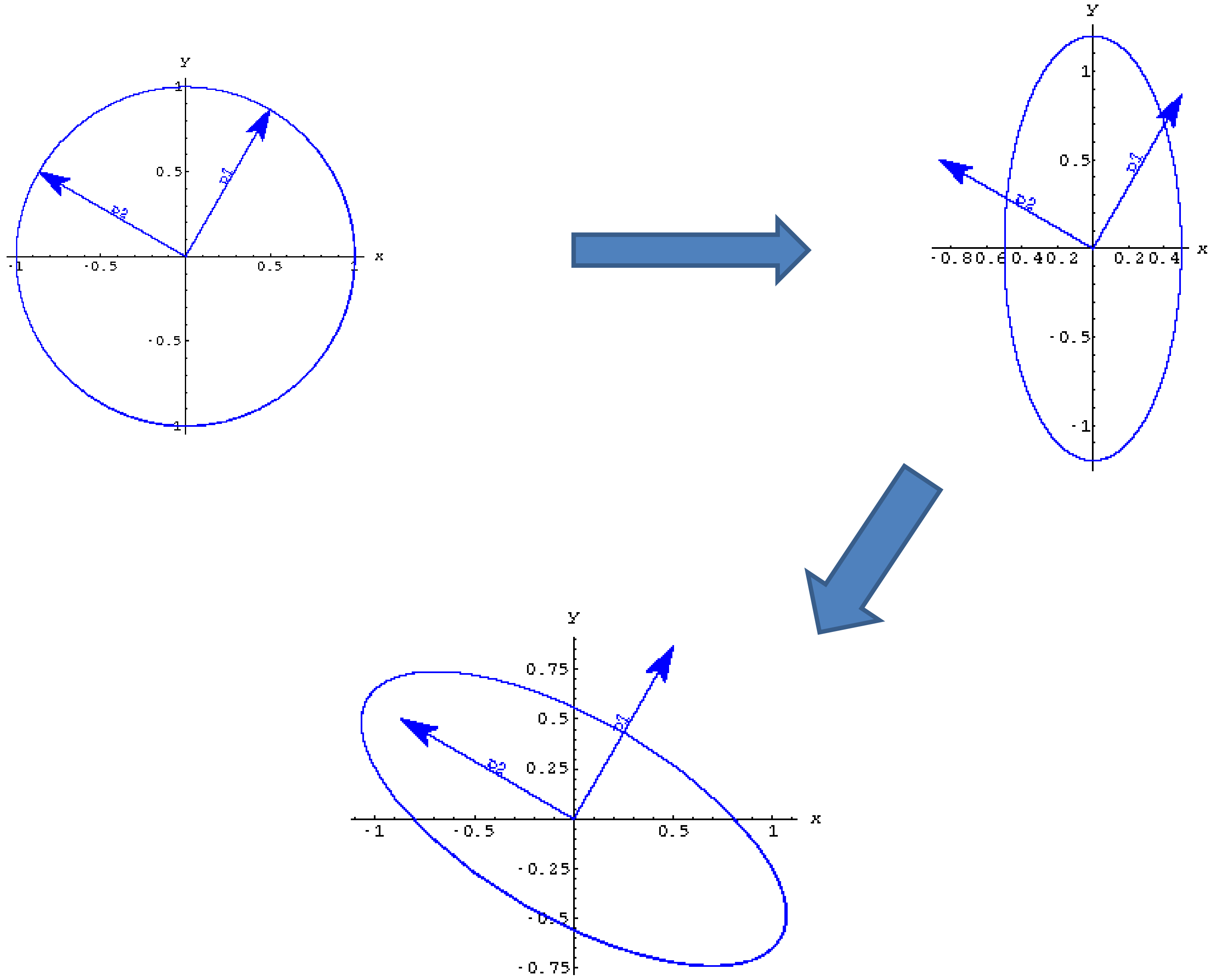
Projections

- Hitting with one matrix and then another

$$Hanger = \begin{pmatrix} \vec{v}_1 & \vec{v}_2 \end{pmatrix} = \begin{pmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{pmatrix} \quad v_1 = \begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix} and \quad v_2 = \begin{pmatrix} -\sin\alpha \\ \cos\alpha \end{pmatrix} \quad \alpha = \frac{\pi}{3}$$

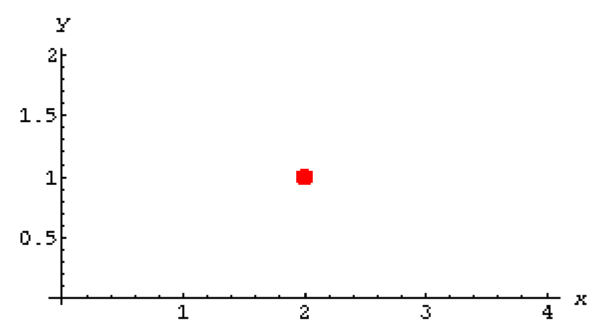
$$Stretcher = \begin{pmatrix} 0.5 & 0 \\ 0 & 1.2 \end{pmatrix}$$

$$M = HS$$


y
x
p1
p2
1
0.5
-0.5
-1
-0.5
0.5
-0.8
-0.6
-0.4
-0.2
0.2
0.4
0.75
0.25
-0.25
-0.75

Projections

- The point on a line through {0,0} closest to P.
  - Here's a plot of the point $p = (2,1)$. What point on the x-axis is closest to the point ?
  - To get the closest point on the x-axis, hit with the xy-stretcher that squashes the y-coordinate but leaves the x-coordinate unchanged.The xy-stretcher you want is $S = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$.
  - The closed point is $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$

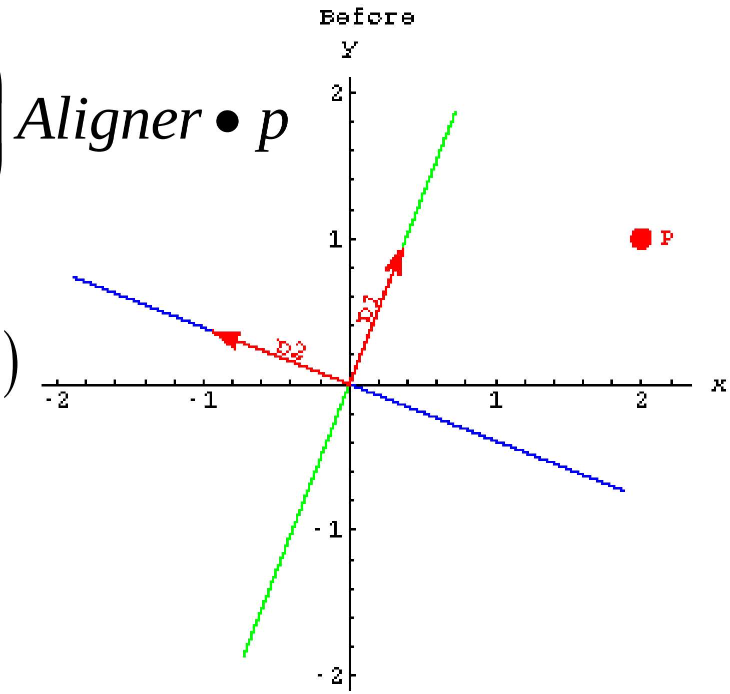


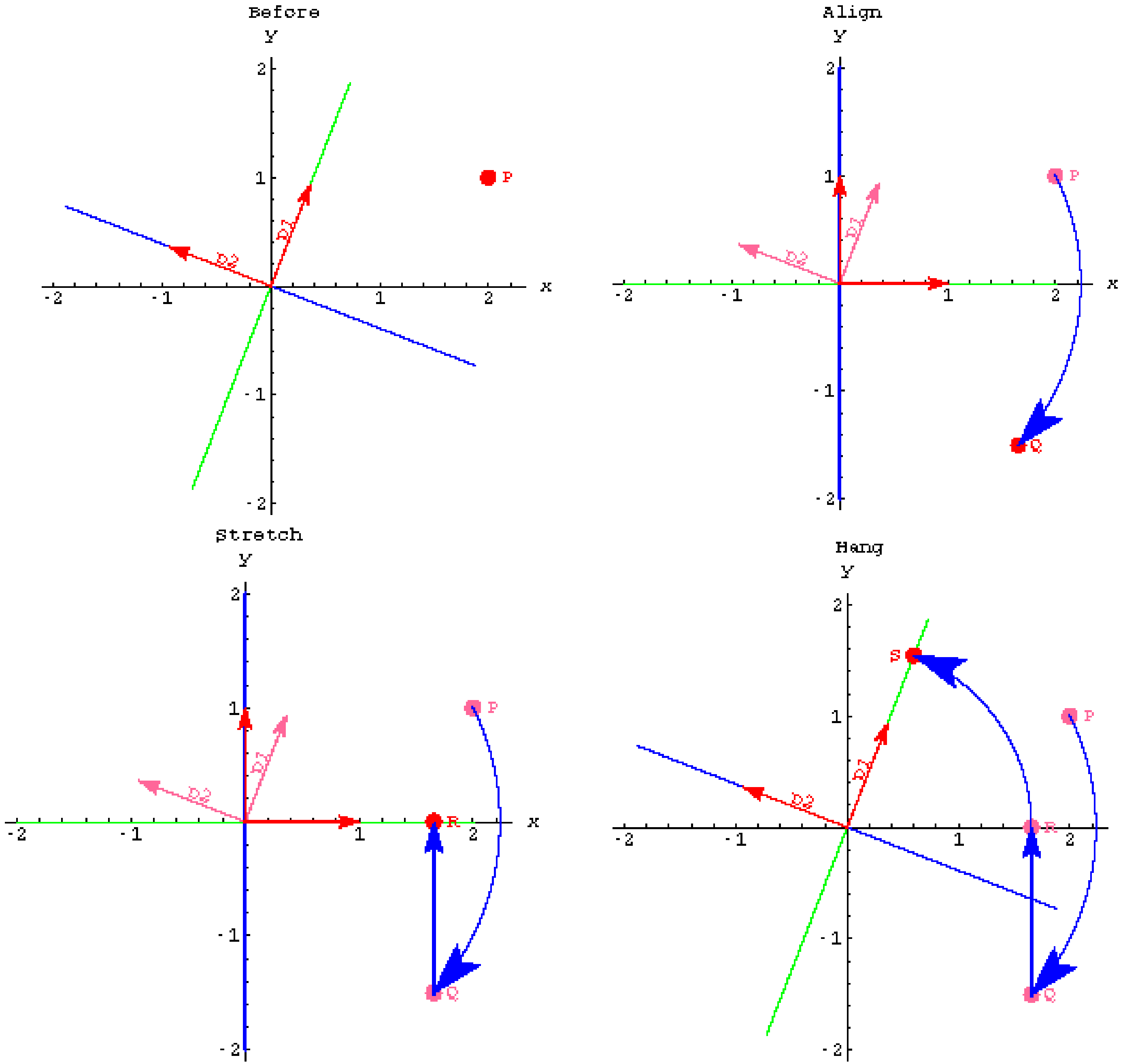
Projections

- Here's a plot of a perpframe and the point $p=(2,1)$ The plot also shows lines through $(0,0)$ parallel to p_1 *and* p_2. What point on the green line is closest to the point $p=(2,1)$?

- Answer: $Closed = Hanger \bullet \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} Aligner \bullet p$

$$Aligner = \begin{pmatrix} \vec{p}_1 \\ \hline \vec{p}_2 \end{pmatrix} \quad Hanger = \begin{pmatrix} \vec{p}_1 & \vec{p}_2 \end{pmatrix}$$





Before
Align
Stretch
Hang
P
Q
R
S
D1
D2
x
y

SVD

- Definition
  - The singular value decomposition for a matrix A writes A as a product

    A=(hanger)(stretcher)(aligner)

  - It's an amazing and useful fact that every m x n matrix has a singular value decomposition.

SVD

- Two-thirds Theorem
  - For an m*n matrix: $A: R^n \rightarrow R^m$ and any orthonormal basis $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n\}$ of R^n
  - Define $s_i = \| A\vec{a}_i \|$ and $h_i = \begin{cases} \vec{0} & if \quad s_i = 0 \\ \frac{1}{s_i} A\vec{a}_i & if \quad s_i \neq 0 \end{cases}$

$$A = \begin{pmatrix} \vec{h}_1 & \vec{h}_2 & \ldots & \vec{h}_n \end{pmatrix} \begin{pmatrix} s_1 & 0 & 0 & 0 \\ 0 & s_1 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & s_n \end{pmatrix} \begin{pmatrix} \vec{a}_1 \\ \vec{a}_1 \\ \vdots \\ \vec{a}_n \end{pmatrix}$$

  - you see that the right hand side of the equation sends to $\vec{a}_i$ to $s_i \vec{h}_i = A\vec{a}_i$

$$A = \begin{pmatrix} \vec{h}_1 & \vec{h}_2 & \dots & \vec{h}_n \end{pmatrix} \bullet Stretcher \bullet Aligner$$

- So you've got the stretcher and the aligner -- if $\begin{pmatrix} \vec{h}_1 & \vec{h}_2 & \dots & \vec{h}_n \end{pmatrix}$ were a hanger matrix then this would be a Singular Value Decomposition for *A*.
- For $\begin{pmatrix} \vec{h}_1 & \vec{h}_2 & \dots & \vec{h}_n \end{pmatrix}$ to be a hanger matrix requires that the columns $\vec{h}_i = \frac{1}{s_i} A\vec{a}_i$ be pairwise perpendicular.
- So one challenge to finding an SVD for *A* is to find an orthonormal basis of R^n ,$\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_n\}$ so that for all $i \neq j$, $A\vec{a}_i \bullet A\vec{a}_j = 0$.

SVD

- Theorem: if A is a m*n matrix, then there is an orthonormal basis of R^n , $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n\}$ so that for all $i \neq j$, $A\vec{a}_i \bullet A\vec{a}_j = 0$

- Proof 1: Here's one way to get such a perpframe.
  - Start with a m x n matrix *A*.
  - Step 1: Let be $\vec{a}_1$ a unit vector maximizing $|A\vec{v}|^2$
  - Step 2: Let S_2 be the space perpendicular to $span\{\vec{a}_1\}$
    Let $\vec{a}_2$ be a unit vector in S_2 maximizing $|A\vec{v}|^2$
  - Step 3: Let S_3 be the space perpendicular to $span\{\vec{a}_1, \vec{a}_2\}$
    Let $\vec{a}_3$ be a unit vector in S_3 maximizing $|A\vec{v}|^2$
    .
    .
    .
    Step n: Let S_n be the space perpendicular to $span\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_{n-1}\}$
    Let $\vec{a}_n$ be a unit vector in S_n maximizing $|A\vec{v}|^2$.

SVD

- Proof2: Based on the spectral theorem

This proof is slick IF YOU'VE ALREADY SEEN THE SPECTRAL THEOREM.

If you haven't seen the spectral theorem, then skip this proof.

Given $\mathrm{A} : \mathbb{R}^n \to \mathbb{R}^m$ and an orthonormal basis $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n\}$ of $\mathbb{R}^n$,

$A\vec{a}_i \cdot A\vec{a}_j = 0 \text{ for } i \neq j$

iff $\vec{a}_i \cdot (A^t A\vec{a}_j) = 0 \text{ for } i \neq j$

iff $A^t A\vec{a}_j = \lambda \vec{a}_j$

iff $\{\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_n\}$ are all eigenvectors of $A^t A$.

Conclusion: The desired basis is guaranteed by spectral theorem since $A^t A$ is symmetric.

SVD

Theorem: Every matrix has a singular value decomposition.

The theorem above almost gives you the SVD for any matrix.

The only problem is that although the columns of the "hanger" matrix are pairwise perpendicular, they might not form a basis for $\mathbb{R}^m$.

For example, suppose for a 5x4 matrix $A:\mathbb{R}^4 \to \mathbb{R}^5$ the procedure outlined above gives you:

$$A = (\vec{h}_1 \,|\, \vec{h}_2 \,|\, \vec{0} \,|\, \vec{0}) \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \vec{a}_1 \\ \hline \vec{a}_2 \\ \hline \vec{a}_3 \\ \hline \vec{a}_4 \end{pmatrix}.$$

To complete the decomposition, let $\{\vec{h}_3, \vec{h}_4, \vec{h}_5\}$ be an orthonormal basis for the three dimensional subspace of $\mathbb{R}^5$ perpendicular to $\mathrm{span}\{\vec{h}_1, \vec{h}_2\}$.

Then write

$$A = (\vec{h}_1 \,|\, \vec{h}_2 \,|\, \vec{0} \,|\, \vec{0}) \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \vec{a}_1 \\ \hline \vec{a}_2 \\ \hline \vec{a}_3 \\ \hline \vec{a}_4 \end{pmatrix}$$

$$\overset{(1)}{=} (\vec{h}_1 \,|\, \vec{h}_2 \,|\, \vec{h}_3 \,|\, \vec{h}_4 \,|\, \vec{h}_5) \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \vec{a}_1 \\ \hline \vec{a}_2 \\ \hline \vec{a}_3 \\ \hline \vec{a}_4 \end{pmatrix}$$

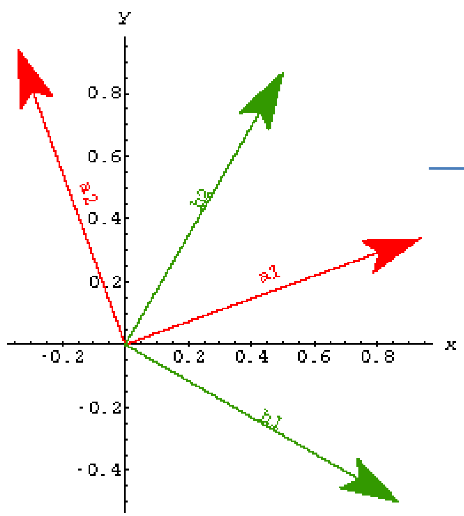
SVD

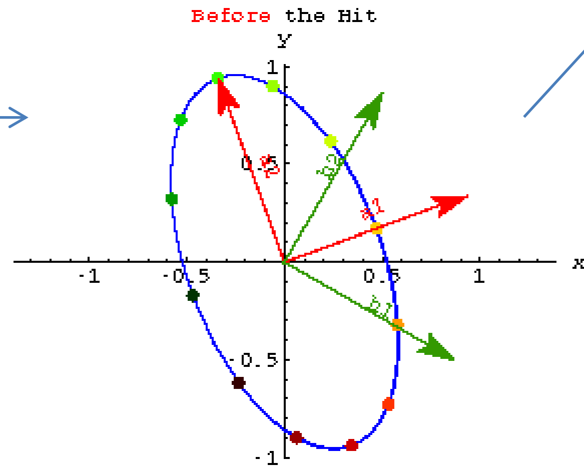
- Comments:
  - The diagonal entries of the stretcher matrix are called the "singular values of A".
  - An extra row of zeros has been added to the stretcher matrix to produce the dimensions required for the multiplication. If A is m x n with m>n , then rows will be deleted.
  - In either case, the dimensions of the stretcher matrix will always match the dimensions of A.

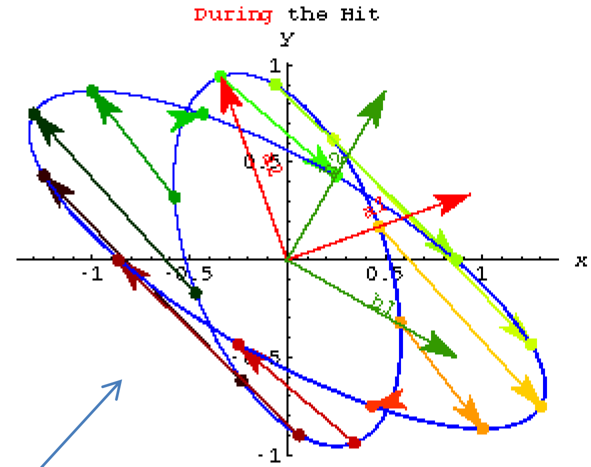
Subspaces

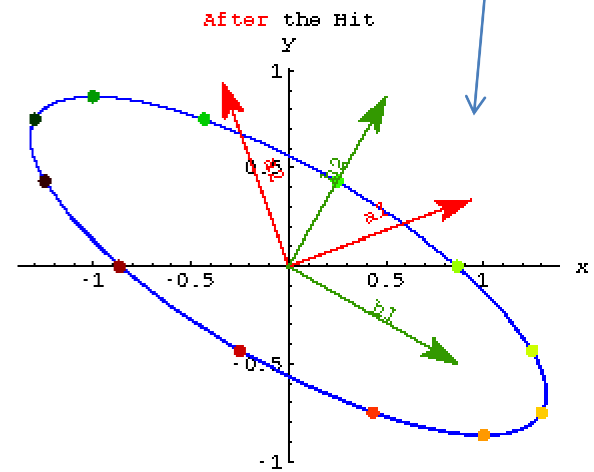
- *A=(hanger)(stretcher)(aligner).*
- For example

$$A = \begin{pmatrix} 2.35589 & 1.12352 \\ -1.55764 & -0.106131 \end{pmatrix} = (\vec{h}_1 \mid \vec{h}_2) \begin{pmatrix} 3 & 0 \\ 0 & 1/2 \end{pmatrix} \begin{pmatrix} \vec{a}_1 \\ \hline \vec{a}_2 \end{pmatrix}$$









- More generally *A,*
  - stretches vectors parallel to a1 by a factor of 3 and rotates them in the direction of h1,
  - stretches vectors parallel to a2 by a factor of ½ and rotates them in the direction of h2.

Subspaces

$$A = \left(\vec{h}_1 \,\middle|\, \vec{h}_2 \,\middle|\, \vec{h}_3 \,\middle|\, \vec{h}_4 \,\middle|\, \vec{h}_5 \right) \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \vec{a}_1 \\ \hline \vec{a}_2 \\ \hline \vec{a}_3 \\ \hline \vec{a}_4 \end{pmatrix}$$

Space	Basis
R^4	$\{\vec{a}_1 \quad \vec{a}_2 \quad \vec{a}_3 \quad \vec{a}_4\}$
R^5	$\{\vec{h}_1 \quad \vec{h}_2 \quad \vec{h}_3 \quad \vec{h}_4 \quad \vec{h}_5\}$
Column Space of A	$\{\vec{h}_1 \quad \vec{h}_2\}$
Row Space of A	$\{\vec{a}_1 \quad \vec{a}_2\}$
Null Space of A	$\{\vec{a}_3 \quad \vec{a}_4\}$
Null Space of A^T	$\{\vec{h}_3 \quad \vec{h}_4 \quad \vec{h}_5\}$

Subspaces

- **Orthonormal bases for fundamental subspaces:** Column Space of A

$$Col[A] = span\{\vec{h}_1, \vec{h}_2\}$$

$$Col[A] = \{Av \mid v \in R^4\}$$
$$= \{A(c_1\vec{a}_1 + c_2\vec{a}_2 + c_3\vec{a}_3 + c_4\vec{a}_4 \mid c_1, c_2, c_3, c_4 \in R\}$$
$$= \{c_1A\vec{a}_1 + c_2A\vec{a}_2 + c_3A\vec{a}_3 + c_4A\vec{a}_4 \mid c_1, c_2, c_3, c_4 \in R\}$$
$$= \{c_1A\vec{a}_1 + c_2A\vec{a}_2 + c_3A\vec{a}_3 + c_4A\vec{a}_4 \mid c_1, c_2, c_3, c_4 \in R\}$$
$$= \{c_1 3\vec{h}_1 + c_2 2\vec{h}_2 + c_3 0\vec{h}_3 + c_4 0\vec{h}_4 \mid c_1, c_2, c_3, c_4 \in R\}$$
$$= \{c_1 3\vec{h}_1 + c_2 2\vec{h}_2 \mid c_1, c_2 \in R\}$$

Subspaces

- **Orthonormal bases for fundamental subspaces:** Null Space of A(is the set of vector that A sends to the zero vector)

$$N[A] = span\{\vec{a}_3, \vec{a}_4\}$$

  – Suppose $x = c_1\vec{a}_1 + c_2\vec{a}_2 + c_3\vec{a}_3 + c_4\vec{a}_4 \in N[A]$

  – Then

$$\begin{aligned}
0 &= |\vec{0}|^2 = |A\vec{v}|^2 \\
&= |A(c_1\vec{a}_1 + c_2\vec{a}_2 + c_3\vec{a}_3 + c_4\vec{a}_4)|^2 \\
&= |c_1 3\vec{h}_1 + c_2 2\vec{h}_2 + c_3 0\vec{h}_3 + c_4 0\vec{h}_4|^2 \\
&= |c_1 3\vec{h}_1 + c_2 2\vec{h}_2|^2 \\
&= 9c_1^2 + 4c_2^2
\end{aligned}$$

– Now $9c_1^2 + 4c_2^2 = 0$ tell you that $c_1 = c_2 = 0$, so

$$x = c_1\vec{a}_1 + c_2\vec{a}_2 + c_3\vec{a}_3 + c_4\vec{a}_4 = c_3\vec{a}_3 + c_4\vec{a}_4$$

– This tell you that any vector in N[A] is a linear combination of a3 and a4, i.e. $N[A] \subseteq span\{a_3, a_4\}$

– On the other hand, $A(c_3\vec{a}_3 + c_4\vec{a}_4) = c_3 0\vec{h}_3 + c_4 0\vec{h}_4 = \vec{0}$

– so, $span\{a_3, a_4\} \subseteq N[A]$

Subspaces

- **Orthonormal bases for fundamental subspaces:**
  - Column Space of A^T(Row Space of A)

$$Col[A^T] = Row(A) = span\{\vec{a}_1, \vec{a}_2\}$$

  - Null Space of A^T(is the set of vector that A^T sends to the zero vector)

$$N[A^T] = span\{\vec{h}_3, \vec{h}_4, \vec{h}_5\}$$

$$A^T = \left(\left(\vec{h}_1 \mid \vec{h}_2 \mid \vec{h}_3 \mid \vec{h}_4 \mid \vec{h}_5 \right) \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \vec{a}_1 \\ \hline \vec{a}_2 \\ \hline \vec{a}_3 \\ \hline \vec{a}_4 \end{pmatrix} \right)^T = \begin{pmatrix} \vec{a}_1 \\ \hline \vec{a}_2 \\ \hline \vec{a}_3 \\ \hline \vec{a}_4 \end{pmatrix}^T \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}^T \left(\vec{h}_1 \mid \vec{h}_2 \mid \vec{h}_3 \mid \vec{h}_4 \mid \vec{h}_5 \right)^T$$

$$= \left(\vec{a}_1 \mid \vec{a}_2 \mid \vec{a}_3 \mid \vec{a}_4 \right) \begin{pmatrix} 3 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \vec{h}_1 \\ \hline \vec{h}_2 \\ \hline \vec{h}_3 \\ \hline \vec{h}_4 \\ \hline \vec{h}_5 \end{pmatrix}$$

- Now you have written as A^T

  (hanger) (stretcher) (aligner), i.e., you have an SVD of A^T.

- So, likewise the SVD of A

$$Col[A^T] = Row(A) = span\{\vec{a}_1, \vec{a}_2\}$$

$$N[A^T] = span\{\vec{h}_3, \vec{h}_4, \vec{h}_5\}$$

Linear Systems & Pseudo-Inverse

- How can you use the decomposition to solve the matrix equation Ax = y?

$$A = \left(\vec{h}_1 \,\middle|\, \vec{h}_2 \,\middle|\, \vec{h}_3 \,\middle|\, \vec{h}_4 \,\middle|\, \vec{h}_5 \right) \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \vec{a}_1 \\ \hline \vec{a}_2 \\ \hline \vec{a}_3 \\ \hline \vec{a}_4 \end{pmatrix}$$

Answer: The first thing you know is that no matter what x you use, $A\,x$ is always in the column space of A, Col[A].

So if $y \notin \mathrm{Col}[A]$, the equation won't have any solution.

If $y \notin \mathrm{Col}[A]$ you should solve $A\,x = \overline{y}$ where $\overline{y}$ is the vector in Col[A] closest to y.

The solution to $A\,x = \overline{y}$ is called the **least squares solution** to $A\,x = y$.

Here's what you do to solve the system:

Step 1: Find the vector $\overline{y}$ in Col[A] closest to y.

Since $\{h_1, h_2\}$ is an orthonormal basis for Col[A], it's easy to compute $\overline{y} = (h_1 \cdot y)\,h_1 + (h_2 \cdot y)\,h_2$.
(Note that if $y \in \mathrm{Col}[A]$, then $y = \overline{y}$.)

Step 2: Let $x = \dfrac{h_1 \cdot y}{3}\,a_1 + \dfrac{h_2 \cdot y}{2}\,a_2$.

That's it!

If the system has a solution, then you've just found it.

If the system has no solution, then you've done the best you can -- you've found the least squares solution.

Here's a check on your solution:

If you set $x = \frac{h_1 \cdot y}{3} a_1 + \frac{h_2 \cdot y}{2} a_2$, then

$$A x = A\left(\frac{h_1 \cdot y}{3} a_1 + \frac{h_2 \cdot y}{2} a_2\right)$$

$$= \frac{h_1 \cdot y}{3} A a_1 + \frac{h_2 \cdot y}{2} A a_2$$ (Linearity)

$$= \frac{h_1 \cdot y}{3} 3 h_1 + \frac{h_2 \cdot y}{2} 2 h_2$$ (Fundamental Equations)

$$= (h_1 \cdot y) h_1 + (h_2 \cdot y) h_2 = \overline{y}$$

Note that you can compute $x = \frac{h_1 \cdot y}{3} a_1 + \frac{h_2 \cdot y}{2} a_2$ using the matrix multiplication

$$x = (a_1 \mid a_2) \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} h_1 \\ \hline h_2 \end{pmatrix} y$$.

The matrix $A^{+} = (a_1 \mid a_2) \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} h_1 \\ \hline h_2 \end{pmatrix}$ is called the pseudo-inverse of A.

The least squares solution to $A\,\boldsymbol{x}=\boldsymbol{y}$ is $x = A^{+} y$.

Other least squares solutions have the form $v + A^{+} y$ where $v \in N[A]$.

Since the SVD gives you an orthonormal basis for N[A], you know any least squares solutions to $A\,\boldsymbol{x}=\boldsymbol{y}$ can be expressed as $x = A^{+} y + s a_3 + t a_4$

Reduced SVD

It's easiest to describe the pseudo-inverse in general terms by first defining the **reduced SVD** for A.

If $$A = (\vec{h}_1 \mid \vec{h}_2 \mid \vec{h}_3 \mid \vec{h}_4 \mid \vec{h}_5) \begin{pmatrix} 3 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \vec{a}_1 \\ \hline \vec{a}_2 \\ \hline \vec{a}_3 \\ \hline \vec{a}_4 \end{pmatrix}$$ is an SVD of A,

then $$A = (\vec{h}_1 \mid \vec{h}_2) \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \vec{a}_1 \\ \hline \vec{a}_2 \end{pmatrix}$$ is the reduced SVD for A.

You get the reduced SVD from the full SVD by keeping only

- the non-zero singular values in the stretcher matrix
- the columns of the hanger and rows of the aligner corresponding to non-zero singular values.

General Pseudo-Inverse

If you have the reduced SVD for an m x n matrix A :

$$A = \left(\vec{h}_1 \mid \cdots \mid \vec{h}_r \right) \begin{pmatrix} s_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & s_r \end{pmatrix} \begin{pmatrix} \vec{a}_1 \\ \hline \vdots \\ \hline \vec{a}_r \end{pmatrix},$$

then the pseudo-inverse of A is

$$A^{+} = \left(\vec{a}_1 \mid \cdots \mid \vec{a}_r \right) \begin{pmatrix} \frac{1}{s_1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{s_r} \end{pmatrix} \begin{pmatrix} \vec{h}_1 \\ \hline \vdots \\ \hline \vec{h}_r \end{pmatrix}.$$

The least squares solution to $A\,x = y$ is given by $x = A^{+}\,y$

The Inverse and the Absolute Value of the Determinant

Suppose A is a square matrix and an SVD is

$$A = \left(\vec{h}_1 \,\middle|\, \cdots \,\middle|\, \vec{h}_n\right) \begin{pmatrix} s_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & s_n \end{pmatrix} \begin{pmatrix} \vec{a}_1 \\ \hline \vdots \\ \hline \vec{a}_n \end{pmatrix}.$$

As it turns out the absolute value of the determinant of A, $\|\mathrm{Det}[A]\|$ equals the product of the singular values.

Thus, if $\mathrm{Det}[A] \neq 0$, then you know that all of the singular values are positive.

This tells you that A has rank n and so is invertible.

If all the singular values are positive, then you can compute

$$A^{-1} = \left(\left(\vec{h}_1 \,\middle|\, \cdots \,\middle|\, \vec{h}_n\right) \begin{pmatrix} s_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & s_n \end{pmatrix} \begin{pmatrix} \vec{a}_1 \\ \hline \vdots \\ \hline \vec{a}_n \end{pmatrix}\right)^{-1}$$

$$\overset{(1)}{=} \begin{pmatrix} \vec{a}_1 \\ \hline \vdots \\ \hline \vec{a}_n \end{pmatrix}^{-1} \begin{pmatrix} s_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & s_n \end{pmatrix}^{-1} \left(\vec{h}_1 \,\middle|\, \cdots \,\middle|\, \vec{h}_n\right)^{-1}$$

$$\overset{(2)}{=} \left(\vec{a}_1 \,\middle|\, \cdots \,\middle|\, \vec{a}_n\right) \begin{pmatrix} \frac{1}{s_n} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{s_n} \end{pmatrix} \begin{pmatrix} \vec{h}_1 \\ \hline \vdots \\ \hline \vec{h}_n \end{pmatrix}$$

$$\overset{(3)}{=} A^{+}.$$

(1) The inverse of a product is the product of the inverses in reverse order.

(2) The aligner matrix $\begin{pmatrix} \vec{a}_1 \\ \hline \vdots \\ \hline \vec{a}_n \end{pmatrix}$ takes the perpframe $\{a_1, \ldots, a_n\}$ to the standard basis.
The inverse of this procedure, taking the standard basis to the perpframe $\{a_1, \ldots, a_n\}$ is accomplished by the hanger matrix $\left(\vec{a}_1 \,\middle|\, \cdots \,\middle|\, \vec{a}_n\right)$.

A similar argument explains why $\left(\vec{h}_1 \mid \cdots \mid \vec{h}_n\right)^{-1} = \begin{pmatrix} \vec{h}_1 \\ \hline \vdots \\ \hline \vec{h}_n \end{pmatrix}$.

To undo the stretcher matrix $\begin{pmatrix} s_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & s_n \end{pmatrix}$ stretches the standard basis vectors by $s_1, \ldots, s_n$.

You can reverse this by stretching the standard basis vectors by $\frac{1}{s_1}, \ldots \frac{1}{s_n}$.

This is done with the stretcher matrix $\begin{pmatrix} \frac{1}{s_n} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{s_n} \end{pmatrix}$ which explains why

$$\begin{pmatrix} s_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & s_n \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{s_n} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \frac{1}{s_n} \end{pmatrix}.$$

(3) The definition of A^+.

Comment: You've just shown that if A is invertible, then $A^+ = A^{-1}$.

Matrix Norm and Rank One Decomposition

- Frobenius Norm
  - The singular values of the matrix $\begin{pmatrix} 3 & 4 & 4 \\ 4 & 1 & 3 \\ 2 & 1 & 4 \end{pmatrix}$ are $\{9.00403, 2.17671, 1.47965\}$
  - When you add up the squares of each entry of A you get

    $$3^2 + 4^2 + 4^2 + 4^2 + 1^2 + 3^2 + 2^2 + 1^2 + 4^2 = 88$$

  - When you add the squares A of the singular values you get

    $$9.00403^2 + 2.17671^2 + 1.47965^2 = 88$$

  - Accident?
  - In math there are no accidents!
  - The Frobenius norm of a matrix A, $|A|_F$, is defined as the square root of the sum of the squares of all its entries.
  - E.g. $\left\| \begin{pmatrix} 3 & 4 & 4 \\ 4 & 1 & 3 \\ 2 & 1 & 4 \end{pmatrix} \right\|_F = \sqrt{88}$.

Matrix Norm and Rank One Decomposition

- Theorem: If A has singular values $s_1, \ldots, s_n$

  Then $\|A\|_F^2 = \sum s_i^2$

Let $A = \text{(hanger)}\,\text{(stretcher)}\,\text{(aligner)}$ be an SVD of A.

First note that for any matrix $C = (c_1 \mid \cdots \mid c_n)$ given in terms of its columns,
$\|C\|_F^2 = \|c_1\|^2 + \cdots + \|c_n\|^2$.

Now,
$\|A\|_F^2 = \|\text{(hanger)}\,\text{(stretcher)}\,\text{(aligner)}\|_F^2$

$\overset{(1)}{=} \|\text{(stretcher)}\,\text{(aligner)}\|_F^2$

$= \|(\text{(stretcher)}\,\text{(aligner)})^T\|_F^2$

$= \|\text{(aligner)}^T\,\text{(stretcher)}^T\|_F^2$

$\overset{(2)}{=} \|\text{(stretcher)}^T\|_F^2 = \|\text{stretcher}\|_F^2$

(1) Let $(c_1 \mid \cdots \mid c_n)$ be the columns of (stretcher)(aligner).

Since the hanger matrix simply rotates the columns of $(c_1 \mid \cdots \mid c_n)$ without changing their lengths, the two sides are equal.

(2) The (aligner)^T matrix simply rotates the columns of (stretcher)^T without changing their lengths.

Matrix Norm and Rank One Decomposition

- Rank One Decomposition

$$A = \left(\vec{h}_1 \,\middle|\, \vec{h}_2 \,\middle|\, \vec{h}_3 \,\middle|\, \vec{h}_4 \,\middle|\, \vec{h}_5 \right) \begin{pmatrix} s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & s_3 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \vec{a}_1 \\ \hline \vec{a}_2 \\ \hline \vec{a}_3 \\ \hline \vec{a}_4 \end{pmatrix}$$

$$A = \left(\vec{h}_1 \,\middle|\, \vec{h}_2 \,\middle|\, \vec{h}_3 \right) \begin{pmatrix} s_1 & 0 & 0 \\ 0 & s_2 & 0 \\ 0 & 0 & s_3 \end{pmatrix} \begin{pmatrix} \vec{a}_1 \\ \hline \vec{a}_2 \\ \hline \vec{a}_3 \end{pmatrix}$$

$$A = s_1 \begin{pmatrix} | \\ h_1 \\ | \end{pmatrix} \begin{pmatrix} a_1 \end{pmatrix} + s_2 \begin{pmatrix} | \\ h_2 \\ | \end{pmatrix} \begin{pmatrix} a_2 \end{pmatrix} + s_3 \begin{pmatrix} | \\ h_3 \\ | \end{pmatrix} \begin{pmatrix} a_3 \end{pmatrix}$$

Matrix Norm and Rank One Decomposition

- Lower Rank Approximations

$$A = \begin{pmatrix} 17.942 & 4.06363 & -11.6675 \\ 26.9125 & 7.27646 & -18.6767 \\ 20.6602 & -27.9415 & 19.7945 \end{pmatrix}$$

$$B = \begin{pmatrix} 17.9444 & 4.06938 & -11.6619 \\ 26.911 & 7.27273 & -18.6804 \\ 20.6601 & -27.9416 & 19.7943 \end{pmatrix}$$

$$A = \left(\vec{h}_1 \mid \vec{h}_2 \mid \vec{h}_3\right) \begin{pmatrix} 40 & 0 & 0 \\ 0 & 40 & 0 \\ 0 & 0 & 0.01 \end{pmatrix} \begin{pmatrix} \vec{a}_1 \\ \hline \vec{a}_2 \\ \hline \vec{a}_3 \end{pmatrix}$$

$$B = \left(\vec{h}_1 \mid \vec{h}_2 \mid \vec{h}_3\right) \begin{pmatrix} 40 & 0 & 0 \\ 0 & 40 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \vec{a}_1 \\ \hline \vec{a}_2 \\ \hline \vec{a}_3 \end{pmatrix}$$

$$A - B = 0.01 \begin{pmatrix} | \\ h_3 \\ | \end{pmatrix} \left(a_3 \right)$$

$$\|A - B\|_F = 0.01$$

$$A = 40 \begin{pmatrix} | \\ h_1 \\ | \end{pmatrix} \left(a_1 \right) + 40 \begin{pmatrix} | \\ h_2 \\ | \end{pmatrix} \left(a_2 \right) + 0.01 \begin{pmatrix} | \\ h_3 \\ | \end{pmatrix} \left(a_3 \right)$$

$$B = 40 \begin{pmatrix} | \\ h_1 \\ | \end{pmatrix} \left(a_1 \right) + 40 \begin{pmatrix} | \\ h_2 \\ | \end{pmatrix} \left(a_2 \right) + 0 \begin{pmatrix} | \\ h_3 \\ | \end{pmatrix} \left(a_3 \right)$$

  - So you see that if A has a small singular value, then you can get a lower rank matrix B close to A by setting the small singular value to zero.
  - The next theorem says that this is the best way to find nearby neighbors of lower rank.

Matrix Norm and Rank One Decomposition

Theorem: Suppose A is an mxn matrix with singular values $s_1 \geq \cdots \geq s_n$ **.**

If A has rank r, then for any k<r, a *best* rank k approximation to A is

$$B = \sum_{i=1}^{k} s_i \vec{h}_i \vec{a}_i^T$$

For a proof of this theorem see page 414 of Steven J. Leon's excellent book, Linear Algebra with Applications, 5th Edition.

Comments:

Best means that $\|A - B\|_F$ is minimized.

Note that

$$\|A - B\|_F = \left\| \sum_{i=1}^{n} s_i \vec{h}_i \vec{a}_i^T - \sum_{i=1}^{k} s_i \vec{h}_i \vec{a}_i^T \right\|_F$$

$$\left(= \left\| \sum_{i=k+1}^{n} s_i \vec{h}_i \vec{a}_i^T \right\| \right)_F$$

$$= \sqrt{s_{k+1}^2 + \cdots s_n^2}\ .$$

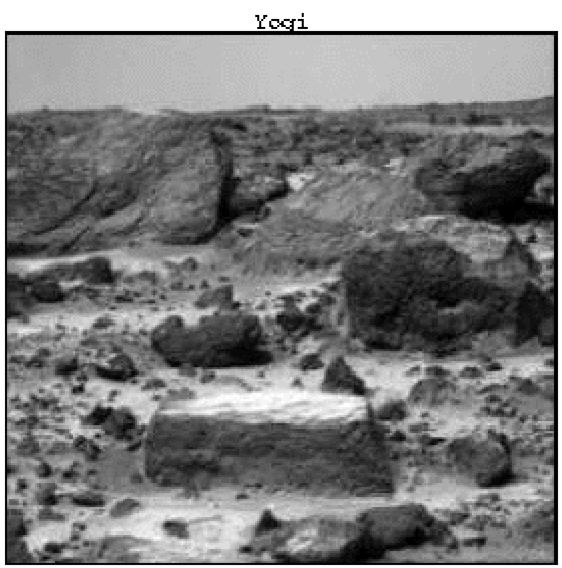
Image Compression

- From the really great book by David Kahaner, Cleve Moler and Stephen Nash - "Numerical Methods and Software", Prentice-Hall, 1989: "Suppose that a satellite in space is taking photographs of Jupiter to be sent back to earth.
- If each photograph were divided into 500 x 500 pixels, it would have to send 250,000 numbers to earth for each picture.

Image Compression

- This would take a great deal of time and would limit the number of photographs that could be transmitted.
- It [is sometimes] possible to approximate this matrix with a matrix which requires less storage."

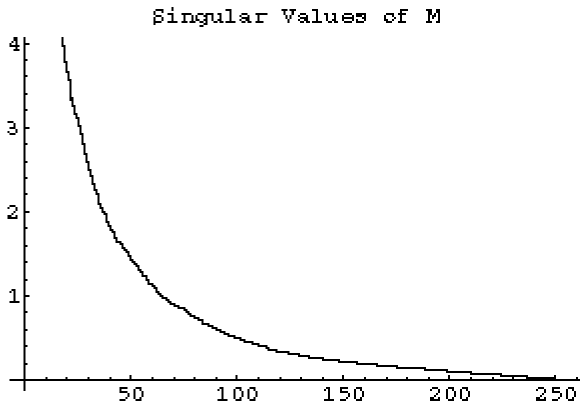
- **Yogi:** Here's a Martian image of a rock called "Yogi" sent back by the Sojourner rover. The image is stored as a 256 x 264 matrix *M* with entries between 0 and 1.
- The matrix *M* has rank 256.
- Here's a plot of the singular values for M.

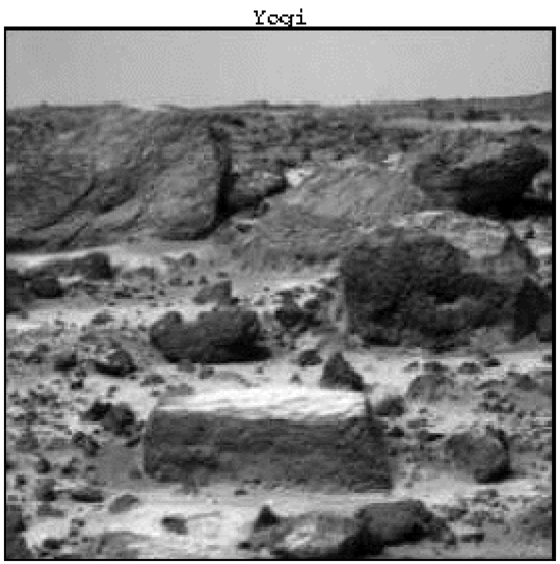


Since the singular values are decaying so rapidly, you can expect that there will be a good lower rank approximation to *M*.

That is, for a relatively small k, you should have $M = \sum_{i=1}^{256} s_i \vec{h}_i \vec{a}_i^T \approx \sum_{i=1}^{k} s_i \vec{h}_i \vec{a}_i^T$

Have a look at various lower rank approximations to *M.*





- The rank 36 approximation looks fairly good.
- What's the advantage of using a lower rank approximation for M?
  - To send the matrix M you need to send 256 x 264 = 67584 numbers.
  - To send the rank 36 approximation to M you need only send
    - the first 36 singular values,
    - the first 36 hanger vectors, each of which has 256 entries,
    - the first 36 aligner vectors, each of which has 264 entries.
    - So in total you need to send only 36(1+256+264)=18756 numbers(18756/67584=28%).

Thanks

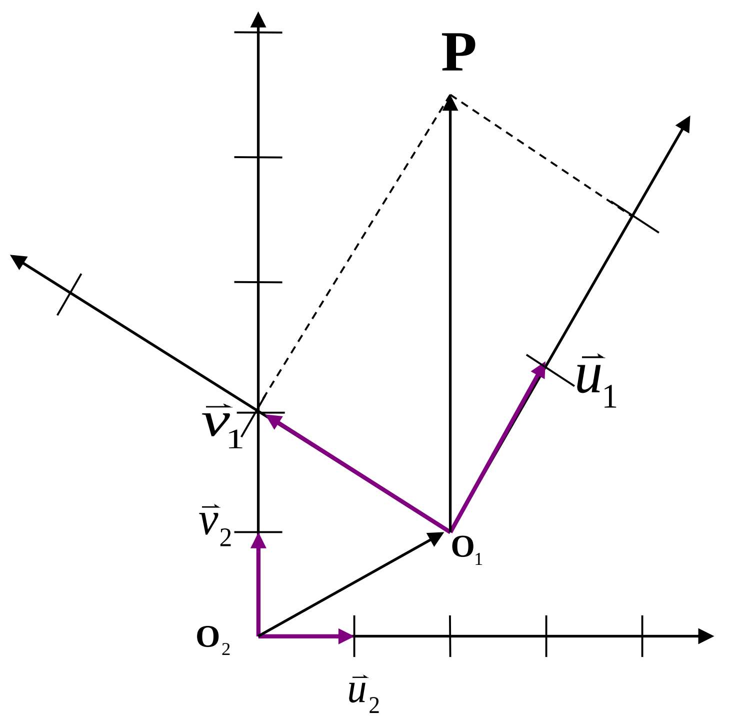
- this materials come from: http://www.uwlax.edu/faculty/will/svd/index.html

Reference

- Gilbert Stang Homepage: http://www-math.mit.edu/~gs/
- Golub, Gene H.; van Loan, Charles F. (1996), Matrix Computations, 3rd edition, Johns Hopkins University Press

Question1

- 已知： $\vec{u}_1 = 1\vec{u}_2 + 1.3\vec{v}_2 \quad \vec{v}_1 = -2\vec{u}_2 + 1\vec{v}_2$
- 问题：P点分别在坐标系(o_1, u_1, v_1) 和坐标系在坐标系(o_1, u_1, v_1) 下的坐标是多少？



$$\mathbf{P} = 2\vec{u}_1 + 1\vec{v}_1 + \mathbf{O}_1$$

$$\vec{u}_1 = 1\vec{u}_2 + 1.3\vec{v}_2$$

$$\vec{v}_1 = -2\vec{u}_2 + 1\vec{v}_2$$

$$\mathbf{O}_1 = 2\vec{u}_2 + 1\vec{v}_2 + \mathbf{O}_2$$

$$\begin{bmatrix} c_1' \\ c_2' \\ 1 \end{bmatrix} = \begin{bmatrix} e_{1,1} & e_{2,1} & e_{3,1} \\ e_{1,2} & e_{2,2} & e_{3,2} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & -2 & 2 \\ 1.3 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4.6 \\ 1 \end{bmatrix}$$