Lecture 14: Singular-value decomposition

Admin: Reading: Meyer 5.12, Strang 6.3 } Singular-value decomposition

Recall:

A is an isometry $\Longleftrightarrow$ A preserves lengths and angles

$\Updownarrow$

A's columns are orthonormal $\Longleftrightarrow$ $A^\dagger A = I$

real, square isometry = Orthogonal
complex, square isometry = Unitary
} rows and columns orthonormal, $A^\dagger = A^{-1}$

Spectral norm $\|A\|$ = maximum stretch $= \max_{x \neq 0} \dfrac{\|Ax\|}{\|x\|}$

Properties

- $\|A\| \geq 0$, and $\|A\| = 0 \Leftrightarrow A = 0$
- $\|A\vec{x}\| \leq \|A\| \cdot \|\vec{x}\|$ (matrix norm; vector/matrix norm)
- $\|\alpha A\| = |\alpha| \cdot \|A\|$ for $\alpha \in \mathbb{C}$
- $\|AB\| \leq \|A\| \cdot \|B\|$

(the amount you can stretch an input by applying AB is at most the stretch from applying B times the stretch from applying A.)

- If U and V are unitary, $\|U\| = \|V\| = 1$ and $\|UAV\| = \|A\|$

- If $\mathcal{U}$ and V are unitary, $\|\mathcal{U}\| = \|V\| = 1$ and
  $\|\mathcal{U} A V\| = \|A\|$ $\Rightarrow$ basis-independent
  (because unitaries don't change lengths).

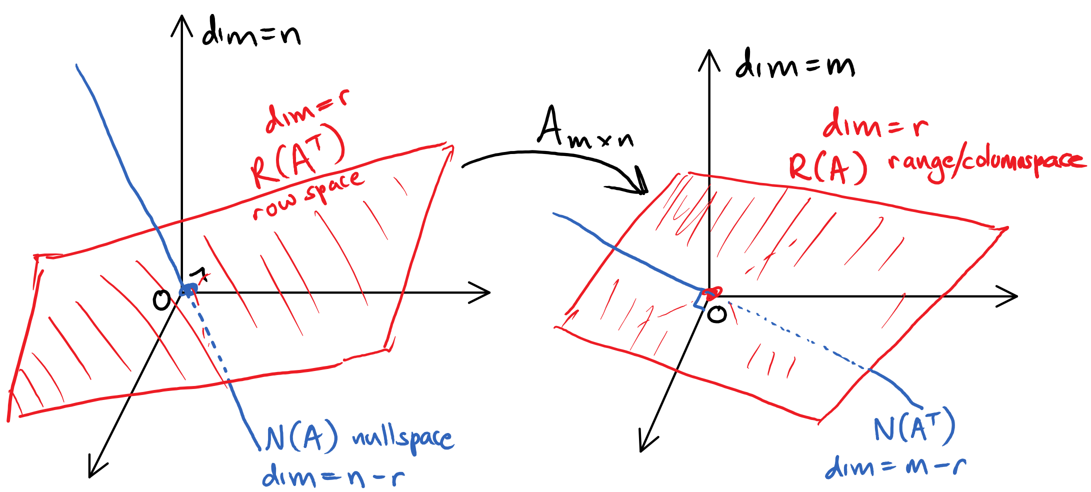
- $\left\|\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}\right\| = \max\{\|A\|, \|B\|\}$
  e.g., if A is a diagonal matrix,
  $\|A\| = \max_i |a_{ii}|$.

- If $\text{rank}(A) = 1$, with $A = \vec{u}\vec{v}^T$, $\|A\| = \|u\| \cdot \|v\|$.

Today:

SINGULAR-VALUE DECOMPOSITION (SVD)

Theoretical motivation:
Any linear transformation A maps points in the rowspace $R(A^T)$ to distinct points in the columnspace $R(A)$. [Rank-Nullity thm.]
How??



Practical motivation: Many applications, including

* Solving linear equations $Ax = b$ $\rightarrow$ What is the sensitivity, e.g., to numerical errors?

$Ax = b$ e.g., to numerical errors?

Find the shortest solution

When there is no solution, find x to minimize $\|Ax-b\|$

Least-squares regression analysis

* Rank minimization

Principal Component Analysis (PCA)

Data mining, clustering, recommendation systems, ...

Singular-Value Decomposition (SVD)

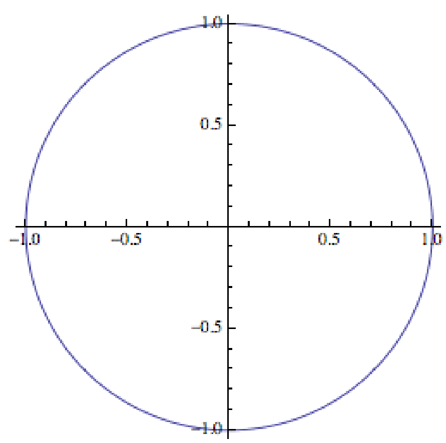
Informally: Any linear transformation can be split into:

- a rotation, followed by
- scaling vectors in or out

Proof by example:

```
In[1]:= ParametricPlot[{Cos[θ], Sin[θ]}, {θ, 0, 2 π}]
```

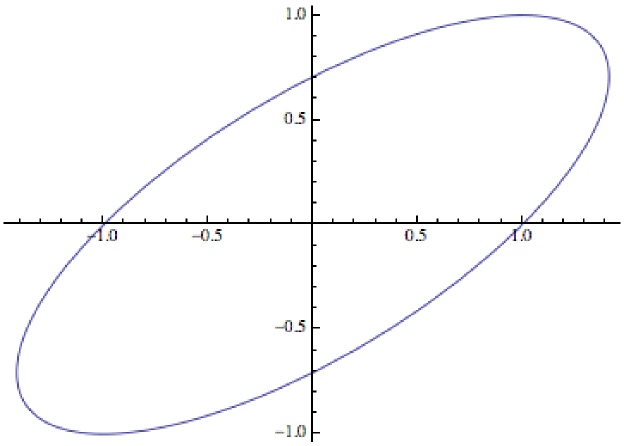
Out[1]=



Choose a matrix, and plot its effect on the unit circle.
Observe that the result looks like an ellipse. (It is an ellipse!)

In[2]:=
```
A = (1 1
     0 1);
ParametricPlot[A.{Cos[θ], Sin[θ]},
 {θ, 0, 2 π}]
```

Out[3]=



Now find the points that are shrunk or expanded the most, using calculus.
Take the derivative, with respect to θ, of
Norm[A.{Cos[θ], Sin[θ]}]^2, set it to zero, and solve.

In[4]:=
```
v = A.{Cos[θ], Sin[θ]}
D[v.v, θ] // Simplify
FindRoot[% == 0, {θ, 0}]
```

Out[4]= {Cos[θ] + Sin[θ], Sin[θ]}

Out[5]= 2 Cos[2 θ] + Sin[2 θ]

Out[6]= {θ → -0.553574}

Let θ_1 and θ_2 be the resulting angle, and the resulting angle plus $\frac{\pi}{2}$ (the angle of the perpendicular line).

In[7]:=
```
{θ1, θ2} = {-0.5535743588970453`,
    -0.5535743588970453` + π/2};
u1 = {Cos[θ1], Sin[θ1]};
u2 = {Cos[θ2], Sin[θ2]};

v1 = A.u1;
scale1 = Norm[v1];
v1 /= scale1;

v2 = A.u2;
scale2 = Norm[v2];
v2 /= scale2;
```

Now observe that the matrix we started with can be decomposed as a rotation/reflection taking u_1 to v_1 and u_2 to v_2, followed by scaling v_1 by scale$_1$ and v_2 by scale$_2$.
(The Chop[] commands cuts off small entries.)

In[16]:=
```
scale1 Transpose[{v1}].{u1} +
   scale2 Transpose[{v2}].{u2} // Chop //
 MatrixForm
```

Out[16]//MatrixForm=

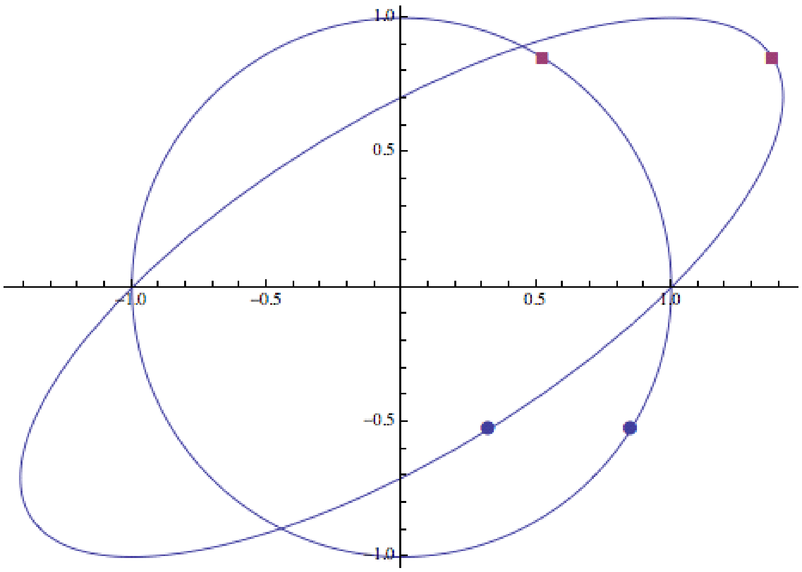
$$\begin{pmatrix} 1. & 1. \\ 0 & 1. \end{pmatrix}$$

We can plot the results to see visually that we have indeed identified the principal axes of the ellipse.

```
In[17]:= plot1 = ParametricPlot[A.{Cos[θ], Sin[θ]},
             {θ, 0, 2 π}];
         dots = ListPlot[{{A.u1}, {A.u2}},
             PlotMarkers → {Automatic, Medium}];

         plot2 = ParametricPlot[{Cos[θ], Sin[θ]},
             {θ, 0, 2 π}];
         dots2 = ListPlot[{{u1}, {u2}},
             PlotMarkers → {Automatic, Medium}];
         Show[plot1, dots, plot2, dots2]
```

Out[21]=



Mathematica's build-in SingularValueDecomposition command ([U, S, V] = svd(A) in Matlab/Octave) does all this automatically.

```
In[22]:= MatrixForm /@ SingularValueDecomposition[A] //
           FullSimplify // N
         {u2, u1}
         {scale2, scale1}
         {v2, v1}
```

Out[22]= $\left\{ \begin{pmatrix} 0.850651 & -0.525731 \\ 0.525731 & 0.850651 \end{pmatrix}, \begin{pmatrix} 1.61803 & 0. \\ 0. & 0.618034 \end{pmatrix}, \begin{pmatrix} 0.525731 & -0.850651 \\ 0.850651 & 0.525731 \end{pmatrix} \right\}$

Out[23]= {{0.525731, 0.850651}, {0.850651, -0.525731}}

Out[24]= {1.61803, 0.618034}

Out[25]= {{0.850651, 0.525731}, {0.525731, -0.850651}}

```
octave:1> A = [1 1; 0 1];
octave:2> [V, S, U] = svd(A)
V =

   0.85065  -0.52573
   0.52573   0.85065

S =

Diagonal Matrix

   1.61803         0
         0   0.61803

U =

   0.52573  -0.85065
   0.85065   0.52573

octave:3> V * S * U'
ans =

   1.0000e+00   1.0000e+00
   1.1102e-16   1.0000e+00
```

Before seeing why this works in general, we need one more fact about the spectral matrix norm:

Key Lemma: For any matrix A,

$$\boxed{\|A\| = \|A^T\|}$$

In fact, if $\|A\vec{u}\| = \|A\| \cdot \|\vec{u}\|$,
i.e., $\vec{u}$ is stretched maximally by A,
then $\vec{v} = A\vec{u}$ is stretched maximally by A^T:

$$\|A^T\vec{v}\| = \|A\| \cdot \|\vec{v}\|.$$

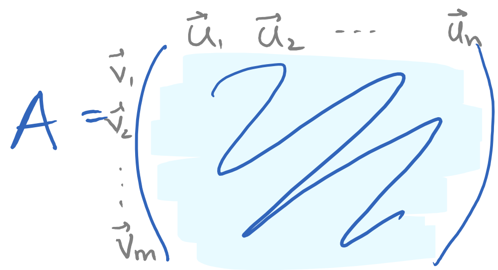
Proof:
Let $\vec{u}_1$ be any unit vector with $\|A\vec{u}_1\| = \|A\|$, and

$$\vec{v}_1 = \frac{A\vec{u}_1}{\|A\vec{u}_1\|} = \frac{A\vec{u}_1}{\|A\|}$$

$$\vec{v}_1 = \frac{A\vec{u}_1}{\|Au_1\|} = \frac{A\vec{u}_1}{\|A\|}$$

Arbitrarily extend $\vec{u}_1$ to an orthonormal basis for $\mathbb{R}^n$
" " $\vec{v}_1$ " " " " $\mathbb{R}^m$.

What does the matrix for A look like in these bases?



① First column is $\begin{pmatrix} \|A\| \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ since $A\vec{u}_1 = \|A\|\,\vec{v}_1$.

$$A = \begin{array}{c|c} & \begin{array}{cccc} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_n \end{array} \\ \begin{array}{c} \vec{v}_1 \\ \vec{v}_2 \\ \vdots \\ \vdots \\ \vec{v}_m \end{array} & \begin{pmatrix} \|A\| & \text{---}\ \vec{b}^T\ \text{---} \\ 0 & \\ 0 & \boxed{} \\ \vdots & \\ 0 & \end{pmatrix} \end{array}$$

② Claim: $\vec{b} = \vec{0}$

Why? Spreading out is good!

For $\vec{x} = \begin{pmatrix} \|A\| \\ b \end{pmatrix}$ $\Rightarrow \|x\| = \sqrt{\|A\|^2 + \|b\|^2}$

$A\vec{x} = \vec{v}_1 \cdot (\|A\|^2 + \|b\|^2)$ $\Rightarrow \|A\vec{x}\| \geq \|A\|^2 + \|b\|^2$
$+ \cdots$

$$\Rightarrow \frac{\|A\vec{x}\|}{\|\vec{x}\|} \geq \frac{\|A\|^2 + \|b\|^2}{\sqrt{\|A\|^2 + \|b\|^2}} = \sqrt{\|A\|^2 + \|b\|^2}$$

but it is $\leq \|A\|$ $\Rightarrow \|b\| = 0$ $\Rightarrow \vec{b} = \vec{0}$ ✓

$$A = \begin{array}{c|c} & \begin{array}{cccc} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_n \end{array} \\ \begin{array}{c} \vec{v}_1 \\ \vec{v}_2 \\ \vdots \end{array} & \begin{pmatrix} \|A\| & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \boxed{} & \end{pmatrix} \end{array}$$

$$A = \begin{array}{c} \vec v_2 \\ \vdots \\ \vec v_m \end{array} \begin{pmatrix} 0 & \\ \vdots & \\ 0 & \end{pmatrix}$$

$$\Rightarrow \quad A^T \vec v_1 = \|A\| \cdot \vec u_1$$

$$\Rightarrow \quad \|A^T\| \geq \frac{\|A^T v_1\|}{\|v_1\|} = \|A\|$$

Repeating the argument with A and A^T switched gives $\|A\| \geq \|A^T\|$. Hence $\|A\| = \|A^T\|$. $\boxed{\checkmark}$

Problem with this proof: Transpose is basis-dependent

The above proof worked not with

$$A \qquad \text{and} \qquad A^T$$

but with

$$\underbrace{\mathcal{U} A V}_{\text{orthogonal basis-change matrices}} \qquad \text{and} \qquad (\mathcal{U} A V)^T.$$

It is okay since orthogonal matrices don't change norm:

$$\|\mathcal{U} A V\| = \|A\| \qquad \|V^T A^T \mathcal{U}^T\| = \|A^T\|.$$

Observe: We can repeat the argument.

$$\underset{m\times n}{A} = \begin{array}{c} \\ \vec v_1 \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{cccc} \vec u_1 & & & \end{array} \\ \begin{pmatrix} \|A\| & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & B & \\ 0 & & & \end{pmatrix} \end{array} \xrightarrow[\text{columns \& rows}]{\text{choose good bases for } B\text{'s}} \begin{array}{c} \\ \vec v_1 \\ \vec v_2 \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{ccccc} \vec u_1 & \vec u_2 & & & \end{array} \\ \begin{pmatrix} \|A\| & 0 & \cdots & & 0 \\ 0 & \|B\| & 0 & \cdots & 0 \\ \vdots & 0 & & & \\ & \vdots & & C & \\ 0 & 0 & & & \end{pmatrix} \end{array}$$

$\longrightarrow \cdots$ (Split off one dimension at a time)

Conclusion: There exist orthonormal bases

$\vec u_1, \dots, \vec u_n$ and $\vec v_1, \dots, \vec v_m$

such that

$$A \vec u_1 = \|A\| \vec v_1$$
$$A \vec u_2 = \|B\| \vec v_2$$

$$A\vec{u}_3 = \|C\|\vec{v}_3$$
$$\vdots$$

Singular-Value Decomposition (SVD)

Theorem: Any matrix $A_{m\times n}$ can be written

$$A = \sum_{i=1}^{\min\{m,n\}} \sigma_i \vec{v}_i \vec{u}_i^T$$

where

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min\{m,n\}} \geq 0$ ← singular values, $\sigma_1 = \|A\|$

$\vec{u}_1, \dots, \vec{u}_n$ orthonormal basis ← right singular vectors

$\vec{v}_1, \dots, \vec{v}_m$ orthonormal basis ← left sing. vectors

Notation:

Interpretation:

$$A\vec{u}_j = \sum_i \sigma_i v_i (\cancel{u_i \cdot u_j}) = \sigma_j \vec{v}_j$$

(0 if $i \neq j$)

$\Rightarrow$ A "rotates" $\vec{u}_j$ into $\vec{v}_j$, and scales it by $\sigma_j \geq 0$

Informally: Any linear transformation can be split into:
- a rotation, followed by
- scaling vectors in or out

Matrix notation:

left sing. vectors, singular values, right sing. vectors

$$A = \underbrace{\begin{pmatrix} | & | & & | \\ \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_m \\ | & | & & | \end{pmatrix}_{m\times m}}_{= V} \underbrace{\begin{pmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \sigma_3 & \\ 0 & & & \ddots \end{pmatrix}_{m\times n}}_{= D} \underbrace{\begin{pmatrix} \text{---} & \vec{u}_1^T & \text{---} \\ \text{---} & \vec{u}_2^T & \text{---} \\ & \vdots & \\ \text{---} & \vec{u}_n^T & \text{---} \end{pmatrix}}_{= U^T}$$

$$A = V \cdot D \cdot U^T$$

nonnegative diagonal (D); orthogonal (V, U^T)

Remark: If A is a complex matrix, everything still works, except U and V are unitary. The singular values in

D are still non-negative real numbers.

Examples of the SVD:

Note: We usually compute the SVD with a computer. Simple examples can be done by hand. Later, we'll see how to get the SVD of A from the spectral decomposition of A^TA (or AA^T).

Exercise: Compute, by hand, the SVDs of

	left sing. vectors	singular values	right sing. vectors
a) $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$	$\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\-1\end{pmatrix}$	1, 1	$\begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}$
b) $\begin{pmatrix} 0 & 5 & 0 \\ -6 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix}$	$-\vec{e}_2, \vec{e}_1, \vec{e}_3$	6, 5, 3	$\vec{e}_1, \vec{e}_2, \vec{e}_3$

$A\vec{e}_1 = 6(-\vec{e}_2)$, $A\vec{e}_2 = 5\vec{e}_1$, $A\vec{e}_3 = 3\vec{e}_3$

c) $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$	$\frac{1}{\sqrt2}\begin{pmatrix}1\\1\end{pmatrix}, \frac{1}{\sqrt2}\begin{pmatrix}-1\\1\end{pmatrix}$	3, 1	$\frac{1}{\sqrt2}\begin{pmatrix}1\\1\end{pmatrix}, \frac{1}{\sqrt2}\begin{pmatrix}1\\-1\end{pmatrix}$

Answer: What is $\|A\| = \sigma_1$?

$\left\|A\begin{pmatrix}x_1\\x_2\end{pmatrix}\right\|^2 = \left\|\begin{pmatrix}x_1+2x_2\\2x_1+x_2\end{pmatrix}\right\|^2$. To maximize this, with $x_1^2+x_2^2=1$, by symmetry we should choose $x_1 = x_2$, so $x_1 = x_2 = \frac{1}{\sqrt2}$.

$\|A\| = \frac{1}{\sqrt2}\left\|A\begin{pmatrix}1\\1\end{pmatrix}\right\| = \frac{1}{\sqrt2}\left\|\begin{pmatrix}3\\3\end{pmatrix}\right\| = 3$

$\Rightarrow \sigma_1 = 3$, $\vec{u}_1 = \vec{v}_1 = \frac{1}{\sqrt2}\begin{pmatrix}1\\1\end{pmatrix}$

$\vec{u}_2$ has to be orthogonal to $\vec{u}_1$ $\Rightarrow$ either $\frac{1}{\sqrt2}\begin{pmatrix}1\\-1\end{pmatrix}$ or $\frac{1}{\sqrt2}\begin{pmatrix}-1\\1\end{pmatrix}$.

For $\vec{u}_2 = \frac{1}{\sqrt2}\begin{pmatrix}1\\-1\end{pmatrix}$,

$A\vec{u}_2 = \frac{1}{\sqrt2}\begin{pmatrix}-1\\1\end{pmatrix}$ $\Rightarrow \sigma_2 = 1$, $\vec{v}_2 = \frac{1}{\sqrt2}\begin{pmatrix}-1\\1\end{pmatrix}$

	left sing. vectors	singular values	right sing. vectors
d) $\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} = \sqrt2 \underbrace{\frac{1}{\sqrt2}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}}$	$\frac{1}{\sqrt2}\begin{pmatrix}1\\i\end{pmatrix}, \frac{1}{\sqrt2}\begin{pmatrix}i\\1\end{pmatrix}$	$\sqrt2, \sqrt2$	$\vec{e}_1, \vec{e}_2$

Answer: Notice that this matrix is unitary; it preserves lengths. Therefore all the singular values are 1, and any orthonormal basis works for right sing. vectors.

Then just choose the left sing. vectors to match up.

e) $\begin{pmatrix} 1 & 0 & 0 & i & 0 \\ 0 & 1 & 0 & 0 & 2 \\ i & 0 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 & 1 \\ 0 & 0 & 6 & 0 & 0 \end{pmatrix}$

f) $\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 2 & -2 & 3 & -3 \\ 1 & 0 & -2 & 0 & 1 & 0 \end{pmatrix}$

g) $\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & -1 & -2 & -3 \\ 1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix}$

Matlab example:

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

V =
```

$\vec{v}_1 \quad \vec{v}_2$

```
    0.8507   -0.5257
    0.5257    0.8507

D =
```

$\sigma_1 = \|A\|$

```
    1.6180         0
         0    0.6180
```

$= \sigma_2$

```
U =
```

$\vec{u}_1 \quad \vec{u}_2$

```
    0.5257   -0.8507
    0.8507    0.5257
```

```
>> V * D * U'

ans =

    1.0000    1.0000
    0.0000    1.0000
```

```
>> u1 = U(:,1);
>> v1 = V(:,1);
>> sigma1 = D(1,1);
>> A * u1

ans =

    1.3764
    0.8507

>> sigma1 * v1

ans =

    1.3764
    0.8507
```

Question: Is the SVD of a matrix unique?

Answer: No, never, since you can always multiply left and right singular vectors by -1.

But even beyond that, if some singular value is repeated, e.g., $\lambda_1 = \lambda_2$, then you can use any orthonormal basis for that 2D subspace.

The easiest example is the identity matrix I. All singular values are 1, and any orthonormal basis works.

SVD and the rank-nullity theorem

Observe:

- columnspace $R(A) = \text{Span}\{\vec{u}_i \mid \lambda_i > 0\}$ (left singular vectors)
  (since any output $A\vec{x}$ is in this span)
- rowspace $R(A^T) = \text{Span}\{\vec{v}_i \mid \lambda_i > 0\}$

$\Rightarrow$ Rank(A) = # of nonzero singular values

Corollary: (Rank-Nullity Theorem)

$$\dim R(A) = \dim R(A^T)$$
$$\dim N(A) + \dim R(A^T) = n$$
$$\dim N(A^T) + \dim R(A) = m$$

follows immediately from SVD!

- A^{-1} exists $\iff$ $m = n$ and all $\lambda_i > 0$

$$A^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i} \vec{v}_i \vec{u}_i^T$$

(singular vectors switched, singular values inverted)

since
$$\left(\sum_i \lambda_i u_i v_i^T\right)\left(\sum_j \frac{1}{\lambda_j} v_j u_j^T\right)$$
$$= \sum_{i,j} \frac{\lambda_i}{\lambda_j} u_i (v_i \cdot v_j) u_j^T$$
$$= \sum_i u_i u_i^T = I \checkmark$$

$$\Rightarrow \boxed{\|A^{-1}\| = \frac{1}{\min \lambda_i}}$$

Using singular values to determine numerically the rank of a matrix

Example:

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

but, generically, any perturbation of A will have rank 4.

```
>> A = diag([1 1 0 0]); rank(A + 10^-3 * randn(4,4))

ans =

     4
```

Intuitively, after a small perturbation, the new matrix's SVD will look like

$$\begin{pmatrix} 1\pm\varepsilon & & & \\ & 1\pm\varepsilon & & 0 \\ & & \varepsilon & \\ 0 & & & \varepsilon \end{pmatrix} \Rightarrow \text{rank is } 4$$

<u>Observe</u>: Small perturbations can increase the rank, but they <u>can't</u> decrease the rank — at least if they are small enough.

If $\|\text{perturbation}\| < \underset{\text{singular value}}{\text{smallest positive}}(A)$

then $\text{rank}(A + \text{perturbation}) \geq \text{rank}(A)$.

$\Rightarrow$ Generically, numerical matrices will have full rank.

To compute rank numerically, compute all the singular values, and then throw away the really small ones (below the threshold of numerical accuracy).

More corollaries of the key lemma ($\|A\| = \|A^T\|$)

<u>Corollary</u>: $\|A^\dagger A\| = \|A\|^2 = \|AA^\dagger\|$.

Proof: We showed before that $\|AB\| \leq \|A\| \cdot \|B\|$. Hence,

$$\|A^\dagger A\| \leq \|A^\dagger\| \cdot \|A\|$$
$$= \|A\|^2,$$

and by the lemma this norm is achieved by the same $\vec{x}$ that achieves $\|Ax\| = \|A\| \cdot \|x\|$. ✓ $\square$

<u>Corollary</u>: $\|A^T\| = \|A\|$, even for complex matrices,

$\|A^\dagger\| = \|A^T\|$

since $\|A^\dagger\| = \|A^T\|$ —— the difference between $A^\dagger$ and A' is just complex conjugation, which doesn't change any lengths.

Corollary: $\|\underbrace{A \cdot A^\dagger \cdot A \cdot A^\dagger \cdot \ldots \cdot A \cdot A^\dagger}_{m \text{ times}}\| = \|A\|^{2m}$

Corollary: If A is real and symmetric $(A = A^T)$
or complex and Hermitian $(A = A^\dagger)$,
then $\|A^m\| = \|A\|^m$.

Example: For $A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$, $\|A\| = \sqrt{2}$

but $A^n = A$

since $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$

$\Rightarrow \|A^n\| = \sqrt{2} \ll \sqrt{2}^n$ ✓

Corollary (of the proof of SVD):

$$\lambda_1 = \|A\| = \|Au_1\| = \max_{x:\|x\|=1} \|Ax\|$$
$$\rotatebox{90}{$\leq$}$$
$$\lambda_2 = \|Au_2\| = \max_{\substack{x:\|x\|=1 \\ x \perp u_1}} \|Ax\|$$
$$\rotatebox{90}{$\leq$}$$
$$\lambda_3 = \|Au_3\| = \max_{\substack{x:\|x\|=1 \\ x \perp u_1, x \perp u_2}} \|Ax\|$$
$$\vdots$$

Also, $$\lambda_2 = \max_{\substack{S \text{ a 2D} \\ \text{subspace of } \mathbb{R}^n}} \left(\min_{\substack{x \in S \\ \|x\|=1}} \|Ax\| \right)$$

$$\vdots$$

This variational characterization of the singular values is not very useful for calculations, but makes it easy to prove statements like

$$\left| k^{th} \text{ largest s.v.}(A) - k^{th} \text{ largest s.v}(A+E) \right| \leq \|E\|$$

Corollary: $\|A^+\| = \|A\|$.

Proof: For $y = Ax$, $A^+ y = \|A\|^2 x = \|A\| \cdot \|y\|$
$$\Rightarrow \|A^+\| \geq \|A\|.$$

The same inequality with $A \leftrightarrow A^+$ switched gives $\|A\| \geq \|A^+\|$. $\square$

$\Rightarrow A\bar{x}$ is stretched maximally by A^+, achieving the norm.