Lecture 14: Singular-value decomposition (class)

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

Recall:

$\boxed{\begin{array}{l} A \text{ is an isometry} \iff A \text{ preserves lengths and angles} \\ \Updownarrow \\ A\text{'s columns are orthonormal} \iff A^\dagger A = I \end{array}}$

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

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

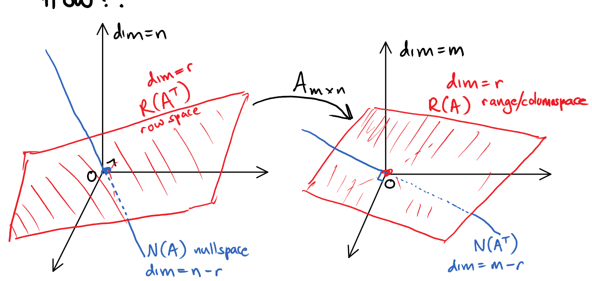
Properties

- $\|A\| \geq 0$, and $\|A\| = 0 \iff A = 0$
- $\|A\vec{x}\| \leq \|A\| \cdot \|\vec{x}\|$ (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 $\|UAV\| = \|A\|$ $\Rightarrow$ basis-independent
  (because unitaries don't change lengths).
- $\left\|\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}\right\| = \max\{\|A\|, \|B\|\}$
  eg, if A is a diagonal matrix, $\|A\| = \max_i |a_{ii}|$.
- If $\operatorname{rank}(A) = 1$, 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

Practical motivation: Many applications, including

* Solving linear equations $Ax = b$

What is the sensitivity, 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)

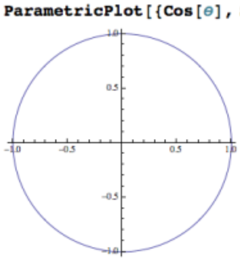
Informally: Any linear transformation can be split into:

- 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]=

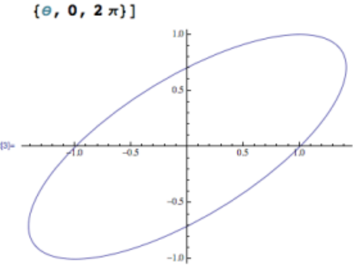


Choose a matrix, and plot its effect on the unit circle.
Observe that the result looks like an ellipse. (It is an ellipse!)

In[2]:= $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$;

```
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=

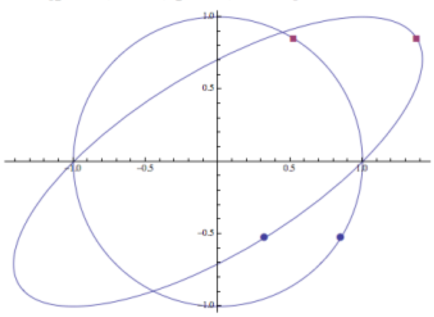
$$\begin{pmatrix} 1. & 1. \\ 0 & 1. \end{pmatrix}$$

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,

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

$$\left\| \underset{A}{\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 6 & 7 & 8 & 9 & 10 \end{pmatrix}} \right\| = \left\| \underset{A^T}{\begin{pmatrix} 1 & 6 \\ 2 & 7 \\ 3 & 8 \\ 4 & 9 \\ 5 & 10 \end{pmatrix}} \right\|$$

$\vec{x} = \begin{pmatrix} 1 \\ y \end{pmatrix}$

$\|A^T x\| = \|A^T\| \cdot \|x\|$

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

$G = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$

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

$$v_1 = \frac{A\vec{u}_1}{\|Au_1\|} = \frac{Au_1}{\|A\|}$$

Extend $\vec{u}_1$ to an orthonormal basis for the input
" $\vec{v}_1$ " " " " output

In these bases

$$[A] = \begin{array}{c|ccc} & u_1 & u_2 & u_3 \; \cdots \\ \hline v_1 & \|A\| & \multicolumn{2}{c}{\vec{b}^T} \\ v_2 & 0 & & \\ v_3 & 0 & \multicolumn{2}{c}{B} \\ \vdots & \vdots & & \end{array}$$

$Au_1 = \|A\| v_1$

Claim: $\vec{b} = \vec{0}$.

Pf: Let $\vec{x} = \alpha \vec{u}_1 + \beta \vec{b}$

$$\|A\vec{x}\|^2 = \left\| \left(\alpha\|A\| + \beta\|b\|^2\right)\vec{v}_1 + \beta \underset{\text{Span}\{v_2, \ldots\}}{\underset{\in}{B\vec{b}}} \right\|^2 = \left(\alpha\|A\| + \beta\|b\|^2\right)^2 + \beta^2\|Bb\|^2 \geq \left(\alpha\|A\| + \beta\|b\|^2\right)^2$$

$$\Rightarrow \|A\|^2 \geq \frac{\|A\vec{x}\|^2}{\|\vec{x}\|^2} \geq \frac{\left(\alpha\|A\| + \beta\|b\|^2\right)^2}{\alpha^2 + \beta^2\|b\|^2}$$

To get something nice, choose $\alpha = \|A\|$, $\beta = 1$

$$\Rightarrow \|A\|^2 \geq \frac{\left(\|A\|^2 + \|b\|^2\right)^2}{\|A\|^2 + \|b\|^2} = \|A\|^2 + \|b\|^2$$

$$\Rightarrow \|\vec b\| = 0 \Rightarrow \vec b = 0$$

$\Rightarrow$ In these bases

$$[A] = \begin{array}{c|c} & \begin{matrix} u_1 & u_2 & u_3 & \cdots \end{matrix} \\ \begin{matrix} v_1 \\ v_2 \\ v_3 \\ \vdots \end{matrix} & \begin{pmatrix} \|A\| & 0 \cdots 0 \\ 0 & \\ 0 & B \\ \vdots & \\ 0 & \end{pmatrix} \end{array} \qquad A^T = \begin{array}{c} v_1 \\ \begin{pmatrix} \|A\| & 0 \cdots 0 \\ 0 & \\ \vdots & B^T \\ 0 & \end{pmatrix} \end{array}$$

$\Rightarrow \|A^T\| \geq \|A\|$

applying this twice implies

$\|A\| \geq \|A^T\| \Rightarrow \|A\| = \|A^T\|$. $\square$

Now apply this proof recursively:

$$[A] = \begin{array}{c|c} & \begin{matrix} u_1 & u_2 \end{matrix} \\ \begin{matrix} v_1 \\ v_2 \\ 3 \\ \vdots \end{matrix} & \begin{pmatrix} \|A\| & 0 & \cdots & 0 \\ 0 & \|B\| & 000 & 0 \\ 0 & 0 & & \\ \vdots & \vdots & C & \\ 0 & 0 & & \end{pmatrix} \end{array} \xrightarrow[\text{to } C]{\text{apply}} \begin{array}{c|c} & \begin{matrix} u_1 & u_2 \end{matrix} \\ \begin{matrix} v_1 \\ v_2 \\ 3 \\ \vdots \end{matrix} & \begin{pmatrix} \|A\| & 0 & \cdots & 0 \\ 0 & \|B\| & 000 & 0 \\ 0 & 0 & \|C\| & \\ \vdots & \vdots & & D \\ 0 & 0 & & \end{pmatrix} \end{array} \longrightarrow \cdots \longrightarrow \begin{array}{c|c} & \begin{matrix} u_1 & u_2 & u_3 & \cdots \end{matrix} \\ \begin{matrix} v_1 \\ v_2 \\ v_3 \\ \vdots \end{matrix} & \begin{pmatrix} \|A\| = \sigma_1 & & 0 \\ & \|B\| = \sigma_2 & \\ & & \|C\| = \sigma_3 \\ 0 & & & \ddots \end{pmatrix} \end{array}$$

$$v_1 = \frac{A u_1}{\|A u_1\|} \qquad v_2 = \frac{A u_2}{\underbrace{\|A u_2\|}_{=\sigma_2}}$$

Singular-Value Decomposition (SVD)

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

$$A = \sum_{j=1}^{\min\{m,n\}} \sigma_j \vec v_j \vec u_j^T$$

where

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

$\vec u_1, \dots, \vec u_n$ is orthonormal $\leftarrow$ right singular vectors

$\vec v_1, \dots, \vec v_m$ is orthonormal $\leftarrow$ left sing. vectors

Notation:

Interpretation:

$$A\vec u_i = \sum_j \sigma_j \vec v_j (\underbrace{\vec u_j^T \vec u_i}_{u_j \cdot u_i = \delta_{ij} = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{o.w.} \end{cases}}) = \sigma_i \vec v_i$$

Problem with this proof: Transpose is basis-dependent

The above proof worked not with A and A^T

but with $\mathcal{U} A V$ and $(\mathcal{U} A V)^T$.

orthogonal basis-change matrices

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.

$$A_{m\times n} = \begin{array}{c|c} & \vec u_1 \\ \vec v_1 & \begin{pmatrix} \|A\| & 0 \cdots 0 \\ 0 & \\ \vdots & B \\ 0 & \end{pmatrix} \end{array}$$

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

$$A = \sum_{j=1}^{\min\{m,n\}} \sigma_j \vec{v}_j \vec{u}_j^T$$

where

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

$\vec{u}_1, \ldots, \vec{u}_n$ is orthonormal ← right singular vectors

$\vec{v}_1, \ldots, \vec{v}_m$ is orthonormal ← left sing. vectors

Notation:

Interpretation:

$$A\vec{u}_i = \sum_j \sigma_j \vec{v}_j (\vec{u}_j^T \vec{u}_i) = \sigma_i \vec{v}_i$$

$$u_j \cdot u_i = \delta_{ij} = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{o.w.} \end{cases}$$

Informally: Any linear transformation can be split into:

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

Matrix notation:

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

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

V, U^T: orthogonal; D: nonnegative diagonal

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.

$$A^T A = \begin{array}{c|cccc} & u_1 & u_2 & u_3 & \cdots \\ \hline u_1 & \sigma_1^2 & & & 0 \\ u_2 & & \sigma_2^2 & & \\ u_3 & & & \sigma_3^2 & \\ \vdots & 0 & & & \ddots \end{array} = \sum_j \sigma_j^2 u_j u_j^T$$

Applications:

$$A = \sum_j \sigma_j \vec{v}_j \vec{u}_j^T$$

$$AA^T = \sum_{j,k} \sigma_j \vec{v}_j \vec{u}_j^T \sigma_k \vec{u}_k \vec{v}_k^T = \sum_j \sigma_j^2 v_j v_j^T = \begin{array}{c|cccc} & v_1 & v_2 & v_3 & \\ \hline v_1 & \sigma_1^2 & & & \\ v_2 & & \sigma_2^2 & & \\ v_3 & & & \sigma_3^2 & \\ & & & & \ddots \end{array}$$

$$u_j \cdot u_k = \delta_{jk} = \begin{cases} 1 & j=k \\ 0 & j\neq k \end{cases}$$

$$\Rightarrow \|AA^T\| = \sigma_1^2 = \|A\|^2 = \|A\| \cdot \|A^T\|$$

Example:

$$AA^TAA^TAA^TA = \sum_j \sigma_j^7 v_j u_j^T$$

$$\| \cdot \| = \sigma_1^7 = \|A\|^7$$

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
$A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$:	$v_1 = e_1$	$\sigma_1 = 1$	$u_1 = e_1$
$\|A\| = 1$	$v_2 = -e_2$	$\sigma_2 = 1$	$u_2 = -e_2$
	$v_1 = u_1$	$\sigma_1 = 1$	$u_1 = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix}$
A :	$v_2 = u_2$	$\sigma_2 = 1$	$u_2 = \frac{1}{\ldots}(\ldots)$

$\|A\|=1$

	left sing. vectors	singular values	right sing. vectors
A :	$v_1 = u_1$	$\sigma_1 = 1$	$u_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$
	$v_2 = u_2$	$\sigma_2 = 1$	$u_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}$
$B = \begin{pmatrix}1 & 0\\0 & i\end{pmatrix}$:	$v_1 = e_1$	$\sigma_1 = 1$	$u_1 = e_1 = \begin{pmatrix}1\\0\end{pmatrix}$
	$v_2 = e_2$	$\sigma_2 = 1$	$u_2 = -i e_2$
$C = n \times n$ unitary: $\|C\| = 1$, square isometry	$v_1 = Cu_1$, $v_2 = Cu_2$, $v_3 = Cu_3$, $\vdots$	$\sigma_1 = \sigma_2 = \cdots = 1$	any orthonormal basis $u_1, u_2, u_3, \ldots$

	left sing. vectors	singular values	right sing. vectors
$D = m \times n$ isometry, $m \geq n$	$v_1 = Du_1$, $v_2 = Du_2$, $v_3 = Du_3$, $\vdots$, $v_n = Du_n$; $v_{n+1}, \ldots, v_m$ } extend $v_1, \ldots, v_n$ to an orthonormal basis	$\sigma_1 = \sigma_2 = \cdots = \sigma_n = 1$	any orthonormal basis $u_1, u_2, u_3, \ldots u_n$

$\|E\| = \sigma_1 = 6$

$$E = \begin{pmatrix} 0 & 5 & 0 \\ -6 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

left sing. vectors	sing. values	right singular vectors
$-e_2$	$\sigma_1 = 6$	$\vec{e}_1$
e_1	$\sigma_2 = 5$	e_2
e_3	$\sigma_3 = 3$	e_3

$F = \begin{pmatrix} 1 & \varepsilon \\ \varepsilon & 1 \end{pmatrix}$ $\varepsilon > 0$

$\|F\| = 1 + \varepsilon$

left	sing. values	right
$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$	$\sigma_1 = 1 + \varepsilon$	$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$
$\begin{cases} \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix} & \text{if } 0 \leq \varepsilon \leq 1 \\ \frac{1}{\sqrt{2}}\begin{pmatrix}-1\\1\end{pmatrix} & \text{if } 1 \leq \varepsilon \end{cases}$	$\sigma_2 = \|1 - \varepsilon\|$	$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}$

$\left[\frac{1}{\sqrt{2}} F \begin{pmatrix}1\\-1\end{pmatrix} = \frac{(1-\varepsilon)}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}\right]$

$\begin{pmatrix} 1 & \varepsilon \\ \varepsilon & 1 \end{pmatrix}$ $\varepsilon < 0$

left	sing. values	right
$\pm \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}$	$\sigma_1 = 1 + \|\varepsilon\|$	$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}$
$\pm \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$	$\sigma_2 = \|1 - \|\varepsilon\|\|$	$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$

$C = \vec{u}\vec{v}^T$ i.e. any rank-one matrix

$\frac{u}{\|u\|}$	$\|C\| = \|u\| \cdot \|v\|$	$\frac{v}{\|v\|}$

$D = \begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} = \sqrt{2} \underbrace{\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}}_{\text{unitary}}$

left sing. vectors	singular values	right sing. vectors
$\frac{1}{\sqrt{2}} D\vec{e}_1$	$\sqrt{2}$	$\vec{e}_1$
$\frac{1}{\sqrt{2}} D\vec{e}_2$	$\sqrt{2}$	$\vec{e}_2$

$\|E\| = \max\{1+|0| = 2, 1+\cancel{2 = 3}, 6\} = 6$

$$E = \begin{pmatrix} 1 & 0 & 0 & i & 0 \\ 0 & 1 & 0 & 0 & 2 \\ i & 0 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 & 1 \\ 0 & 0 & 6 & 0 & 0 \end{pmatrix}$$

left sing. vectors	singular values	right sing. vectors
$\vec{e}_5$	6	$\vec{e}_3$
$\frac{1}{\sqrt{2}}(\vec{e}_2 + \vec{e}_4)$	$2+1 = 3$	$\frac{1}{\sqrt{2}}(\vec{e}_2 + \vec{e}_5)$
$\frac{e_1 + ie_3}{\sqrt{2}}$	$\sigma_3 = \sqrt{2}$	e_1
$\frac{ie_1 + e_3}{\sqrt{2}}$	$\sigma_4 = \sqrt{2}$	e_4
$\frac{-1}{\sqrt{2}}(e_2 - e_5)$	$2 - 1 = 1$	$\frac{1}{\sqrt{2}}(e_2 - e_5)$
$\vec{e}_1$	$\sqrt{6}$	$\frac{1}{\sqrt{6}}\begin{pmatrix} \vdots \end{pmatrix}$

$\frac{-1}{\sqrt{2}}(e_2-e_5)$ $\quad 2-1=1.$ $\quad \sqrt{2}(e_2\ e_5)$

$$F=\begin{pmatrix}1&1&1&1&1&1\\1&-1&2&-2&3&-3\\1&0&-2&0&1&0\end{pmatrix}_{3\times 6}$$

$\vec{e}_1$	$\sqrt{6}$	$\frac{1}{\sqrt{6}}\begin{pmatrix}1\\ \vdots\\ 1\end{pmatrix}$
$\vec{e}_2$	$\sqrt{\ }$	$\frac{1}{\sqrt{\ }}\begin{pmatrix}-1\\2\\-2\\3\\-3\end{pmatrix}$
$\vec{e}_3$	$\sqrt{\ }$	$\frac{1}{\sqrt{\ }}\begin{pmatrix}1\\0\\-2\\0\\1\\0\end{pmatrix}$

$$G=\begin{pmatrix}1&1&1&1&1&1\\1&2&3&-1&-2&-3\\1&1&-1&1&1&-1\end{pmatrix}$$

?

Matlab example:

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

V =
      v1          v2
    0.8507     -0.5257
    0.5257      0.8507

D =
          σ1 = ||A||
    1.6180         0
         0    0.6180 = σ2

U =
      u1          u2
    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?

$$A=\sum_j \sigma_j \vec{v}_j \vec{u}_j^T$$

SVD and the rank-nullity theorem

Observe:

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

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

Corollary: (Rank-Nullity Theorem)

$$\left.\begin{aligned}\dim R(A)&=\dim R(A^T)\\ \dim N(A)+\dim R(A^T)&=n\\ \dim N(A^T)+\dim R(A)&=m\end{aligned}\right\}\text{ follows immediately from SVD!}$$

- A^{-1} exists $\Longleftrightarrow$

  $A^{-1}=$

$\Rightarrow \|A^{-1}\| =$

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

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

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

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 ($\|A\| = \|A^T\|$)

Corollary: $\|A^\dagger A\| = \|A\|^2 = \|A A^\dagger\|$.

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 $\|Ax\| = \|A\| \cdot \|x\|$. ✓ □

Corollary: $\|A^T\| = \|A\|$, even for complex matrices, since $\|A^\dagger\| = \|A^T\|$ — the difference between $A^\dagger$ and A^T 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 $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$

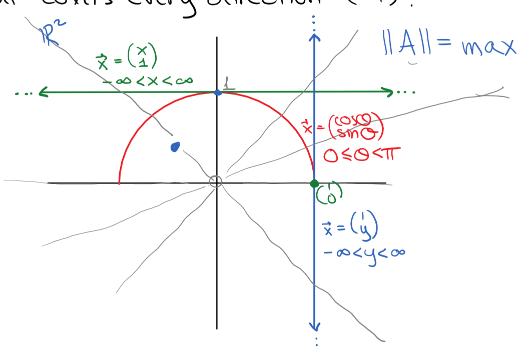
$\Rightarrow \|A^n\| = \sqrt{2} \ll \sqrt{2}^n$ ✓

Different parameterizations

Using $\|A\| = \max\limits_{\vec{x} : \|x\|=1} \|A\vec{x}\|$, we need to optimize over length-one vectors —— except $\|A(-\vec{x})\| = \|A\vec{x}\|$.

$\Rightarrow$ for an $m\times 2$ matrix, $\|A\| = \max\limits_{\theta\in[0,\pi)} \left\|A\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}\right\|$.

Using $\|A\| = \max\limits_{\vec{x}\neq\vec{0}} \dfrac{\|A\vec{x}\|}{\|\vec{x}\|} = \dfrac{\|Ab\vec{x}\|}{\|b\vec{x}\|}$ for any $b\neq 0$, we can use any parameterization that covers every direction (± 1).



$$\|A\| = \max\left\{ \max_{-\infty<x<\infty} \frac{\left\|A\binom{x}{1}\right\|}{\left\|\binom{x}{1}\right\|},\ \left\|A\binom{1}{0}\right\| \right\}$$

covers the horizontal direction

$$A = \begin{pmatrix} 0 & B \\ C & 0 \end{pmatrix}$$

$$\Rightarrow \|A\| = \max\{\|B\|, \|C\|\}$$

$$A = \begin{pmatrix} 1 & 3 \\ 2 & 4 \\ 3 & -\frac{1}{3} \end{pmatrix} \qquad \sigma_1 \ge \sigma_2$$

(columns $= v_1$, $= v_2$)

$$= v_1 e_1^T + v_2 e_2^T$$

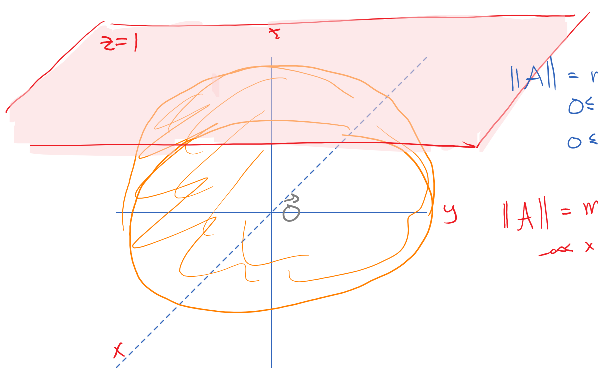
$$= \underbrace{\|v_1\|}_{=\sigma_2} \frac{v_1}{\|v_1\|} e_1^T + \underbrace{\|v_2\|}_{=\sigma_1} \frac{v_2}{\|v_2\|} e_2^T$$

left singular vectors

right sing. vectors

$$\Rightarrow \|A\| = \sigma_1 = \|v_2\|$$

$= \sqrt{\ }$ [illegible]



$$\|A\| = \max_{\substack{0\le\theta\le 2\pi \\ 0\le\phi\le\frac{\pi}{2}}} \left\| A\begin{pmatrix}\cos\theta\sin\phi \\ \sin\theta\sin\phi \\ \cos\phi\end{pmatrix} \right\|$$

$$\|A\| = \max_{-\infty < x,y < \infty} \frac{\left\|A\begin{pmatrix}x\\y\\1\end{pmatrix}\right\|}{\left\|\begin{pmatrix}x\\y\\1\end{pmatrix}\right\|}$$