Lecture 16: Singular-value decomposition

Tuesday, October 20, 2015 9:30 AM

Admin: Homework 5 due Thursday.

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

Recall:

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

$\Updownarrow$

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

real, square isometry = Orthogonal

complex, square isometry = Unitary

} rows and columns orthonormal, $A^\dagger = A^{-1}$

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

Properties

- $\|A\| \geq 0$, and $\|A\| = 0 \Leftrightarrow A = 0$
- $\|A\vec{x}\| \leq \|A\| \cdot \|\vec{x}\|$ ($\|A\|$: matrix norm; $\|\vec{x}\|$: 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).

- $\left\| \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \right\| = \max\{\|A\|, \|B\|\}$

  e.g., if A is a diagonal matrix,
  $\|A\| = \max_i |a_{ii}|$.

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

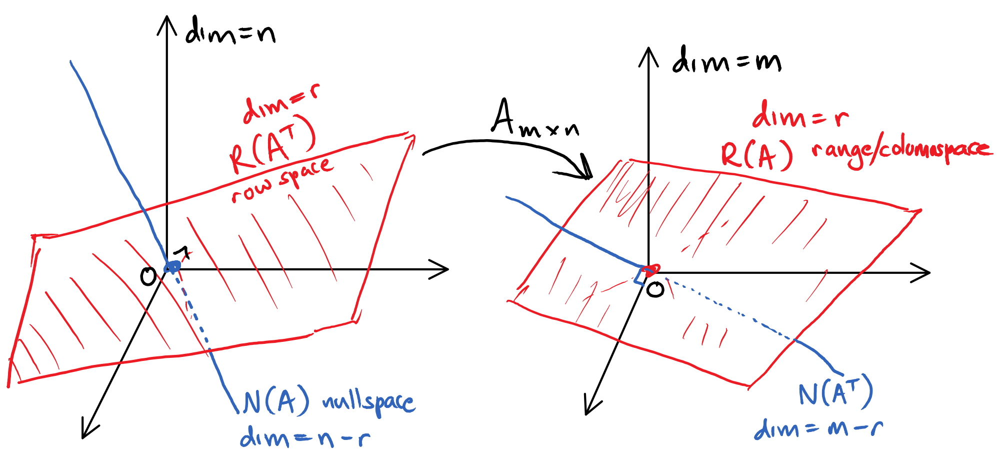
Today:

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 $Ax = b$ → What is the sensitivity, e.g., to numerical errors?

→ Find the shortest solution

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

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)

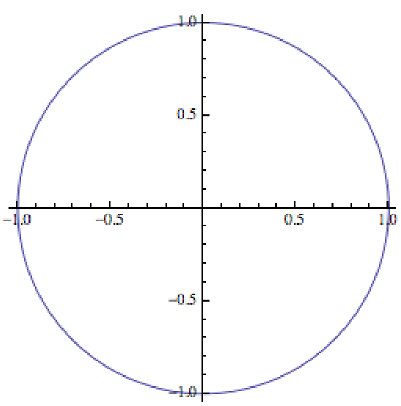
Informally: Any linear transformation can be split into:

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

Proof by example:

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

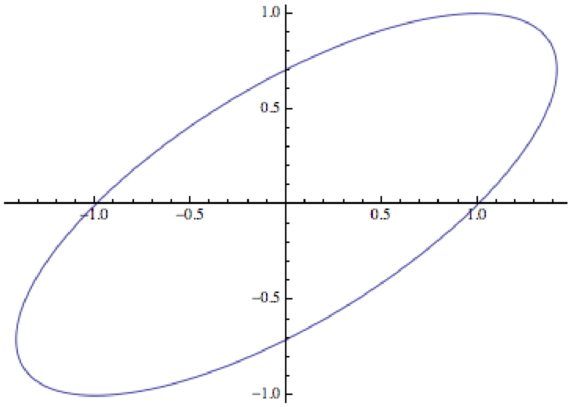
Out[1]=



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=

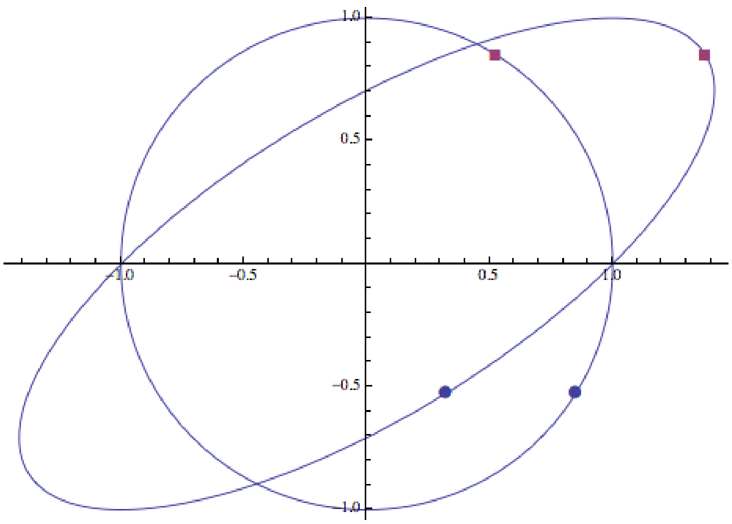
$$\begin{pmatrix} 1. & 1. \\ 0 & 1. \end{pmatrix}$$

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:

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

$$\|A\| = \max_{x:\|x\|=1} \|Ax\| \qquad \|A^T\| = \max_{y:\|y\|=1} \|yA\|$$

Key Lemma:

If $\|A\vec{x}\| = \|A\| \cdot \|\vec{x}\|$,

i.e., $\vec{x}$ is stretched maximally by A, achieving the norm, then

$$A^\dagger A\vec{x} = \|A\|^2 \vec{x}$$

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

Proof: For $y = Ax$, $A^\dagger y = \|A\|^2 x = \|A\| \cdot \|y\|$

$$\Rightarrow \|A^\dagger\| \geq \|A\|.$$

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

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

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

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 $\|Ax\| = \|A\| \cdot \|x\|$. ✓ $\square$

Proof of the lemma: Scale so $\|x\| = 1$.

Choose a basis for the domain (columns) of A so that the first basis element is x.

Choose a basis for the codomain (rows) so the first basis element is $y \doteq \dfrac{Ax}{\|Ax\|}$.

A's representation is these bases is

$$\begin{pmatrix} \|A\| & \\ 0 & \\ 0 & \text{stuff} \\ \vdots & \\ 0 & \end{pmatrix}$$

because $Ax = \|A\| y$. Then $A^\dagger y$ is the first row of A

$$\Rightarrow A^\dagger y = \|A\| x + b$$

where $b \perp x$ (b is everything else in the first row)

We claim that $b = 0$! (So $A^\dagger x = \|A\| x$, our goal.)

Because if $b \neq 0$, then spreading out between x and b would increase the norm, a contradiction of $\|Ax\| = \|A\|$.

$$\left\| A \frac{\|A\|x+b}{\sqrt{\|A\|^2+\|b\|^2}} \right\| \geq \left| y \cdot \left(A \frac{\|A\|x+b}{\sqrt{\|A\|^2+\|b\|^2}} \right) \right|$$

since $\|u\| \geq |v \cdot u|$ for all vectors v with $\|v\| = 1$

$$= \frac{\|Ax+b\|^2}{\sqrt{\|A\|^2+\|b\|^2}}$$

$$= \sqrt{\|A\|^2+\|b\|^2}$$

If $b \neq 0$, then we've found a unit vector that A stretches more than it stretches x, a contradiction. $\square$

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

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

Singular-Value Decomposition (SVD)

Informally: Any linear transformation can be split into:

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

Formally:

Theorem: Any $m \times n$ real matrix A can be written

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

where

$\lambda_1, \lambda_2, \dots, \lambda_{\min\{m,n\}} \geq 0$

$\{\vec{u}_i\}_{i=1,\dots,m}$ is an orthonormal basis for $\mathbb{R}^m$

$\{\vec{v}_i\}_{i=1,\dots,n}$ is an orthonormal basis for $\mathbb{R}^n$

Interpretation:

$$A\vec{v}_j = \sum_i \lambda_i u_i (\underbrace{v_i \cdot v_j}_{0 \text{ if } i \neq j}) = \lambda_j \vec{u}_j$$

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

Notation: $\{\lambda_j\}$ $\leftarrow$ singular values

largest singular value, $\max_j \lambda_j = \|A\|$

$\{\vec{v}_j\}$ $\leftarrow$ right singular vectors

$\{\vec{u}_j\}$ $\leftarrow$ left singular vectors

Matrix notation:

$$A = \underbrace{\begin{pmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_m \\ | & | & & | \end{pmatrix}}_{m \times m} \underbrace{\begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \lambda_3 & \\ 0 & & & \ddots \end{pmatrix}}_{m \times n} \begin{pmatrix} \text{—} & v_1^T & \text{—} \\ \text{—} & v_2^T & \text{—} \\ & \vdots & \\ \text{—} & v_n^T & \text{—} \end{pmatrix}$$

$= \mathcal{U}$: $m \times m$ orthogonal matrix, columns are left singular vectors

$= D$: nonnegative diagonal matrix of singular values

$= V^T$: $n \times n$ orthogonal matrix, columns of V are right singular vectors

$$A = \mathcal{U} \cdot D \cdot V$$

Theorem: If A is a complex matrix, all the above still holds, except $\{\vec{u}_i\}$ and $\{\vec{v}_i\}$ are orthonormal bases for $\mathbb{C}^m$ and $\mathbb{C}^n$, instead of $\mathbb{R}^m$ and $\mathbb{R}^n$.

$\Rightarrow$ $\mathcal{U}$ and V are unitary instead of orthogonal (singular values are still non-negative reals)

Example:

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

$$\begin{array}{cc} u_1 & u_2 \\ \begin{pmatrix}0.85065\\ 0.52573\end{pmatrix} & \begin{pmatrix}-0.52573\\ 0.85065\end{pmatrix}\end{array}$$

```
D =

Diagonal Matrix
```

$\lambda_1 =$ 1.61803 0

0 λ_2 0.61803

singular values (norm = 1.61803)

V =

$$\begin{array}{cc} v_1 & v_2 \\ \begin{pmatrix}0.52573\\ 0.85065\end{pmatrix} & \begin{pmatrix}-0.85065\\ 0.52573\end{pmatrix}\end{array}$$

```
octave:3> U * D * V'
ans =

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

octave:4> v1 = V(:,1); u1 = U(:,1); lambda1 = D(1,1);
octave:5> A*v1 - lambda1 * u1
ans =

   0
   0
```

Example: What are the SVDs of

(a) $\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$ (b) $\begin{pmatrix}0 & 1 & 0\\ 1 & 0 & 0\end{pmatrix}$

(c) $\begin{pmatrix}1 & i\\ i & 1\end{pmatrix}$

PROOF OF THE SVD.

Key Lemma: If $\|A\vec{x}\| = \|A\| \cdot \|\vec{x}\|$,
then $A^\dagger A\vec{x} = \|A\|^2 \vec{x}$.

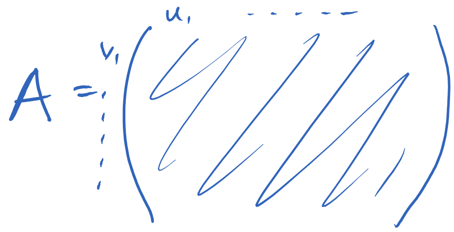
Goal: Find vectors $\vec{u}_1, \ldots, \vec{u}_n$ (orthonormal)
so $A\vec{u}_i = \lambda \vec{v}_i$ ($\vec{v}_i$s also orthonormal).

Easy!
Let $\vec{u}_1$ be any unit vector with $\|A\vec{u}_1\| = \|A\|$.

Let $\vec{v}_1 = A\vec{u}_1 / \|A\vec{u}_1\|$.

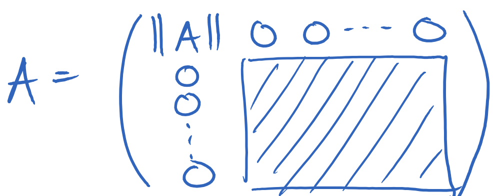
Arbitrarily extend $\vec{u}_1$ to a basis for $\mathbb{R}^n$
and " $\vec{v}_1$ " " " $\mathbb{R}^m$.

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



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

② First row is $\begin{pmatrix} \|A\| & 0 & \cdots & 0 \end{pmatrix}$ since $A^T\vec{v}_1 = \|A\| u_1$ (by the lemma).



now repeat the procedure here
(restricting domain to $u_1^\perp$, codomain to $v_1^\perp$)

$\square$

Remember this:

Theorem: Any $m \times n$ real matrix A can be written

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

where

$\lambda_1, \lambda_2, \ldots, \lambda_{\min\{m,n\}} \geq 0$

$\{\vec{u}_i\}_{i=1,\ldots,m}$ is an orthonormal basis for $\mathbb{R}^m$

$\{\vec{v}_i\}_{i=1,\ldots,n}$ is an orthonormal basis for $\mathbb{R}^n$

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.

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)

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

- A^{-1} exists $\iff$ $m = n$ 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

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

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

Using singular values to determine numerically the rank of a matrix

Example:

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

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

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

ans =
```

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).