The biggest Markov chain in the world

Randy's web-surfing behavior: From whatever page he's viewing, he selects one of the links uniformly at random and follows it.

Defines a Markov chain in which the states are web pages.

Idea: Suppose this Markov chain has a stationary distribution.

- Find the stationary distribution $\Rightarrow$ probabilities for all web pages.
- Use each web page's probability as a measure of the page's importance.
- When someone searches for "matrix book", which page to return? Among all pages with those terms, return the one with highest probability.

Advantages:

- Computation of stationary distribution is independent of search terms: can be done once and subsequently used for all searches.
- Potentially could use power method to compute stationary distribution.

Pitfalls: Maybe there are several, and how would you compute one?

Using Perron-Frobenius Theorem

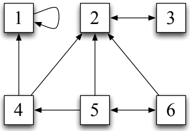
If can get from every state to every other state in one step, Perron-Frobenius Theorem ensures that there is only one stationary distribution....

and that the Markov chain converges to it ....

so can use power method to estimate it.
Pitfall: This isn't true for the web!

Workaround: Solve the problem with a hack: In each step, with probability 0.15, Randy just teleports to a web page chosen uniformly at random.

Mix of two distributions



Following random links:

$A_1 =$

	1	2	3	4	5	6
1	1			$\frac{1}{2}$		
2			1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{2}$
3		1				
4					$\frac{1}{3}$	
5						$\frac{1}{2}$
6					$\frac{1}{3}$	

Uniform distribution: transition matrix like

$A_2 =$

	1	2	3	4	5	6
1	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
2	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
3	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
4	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
5	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$
6	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Use a mix of the two: incidence matrix is
$A = 0.85 * A_1 + 0.15 * A_2$

To find the stationary distribution, use power method to estimate the eigenvector $\mathbf{v}$ corresponding to eigenvalue 1.

Adding those matrices? Multiplying them by a vector? Need a clever trick.

Clever approach to matrix-vector multiplication

$$A = 0.85 * A_1 + 0.15 * A_2$$

$$\begin{aligned} A\,\mathbf{v} &= (0.85 * A_1 + 0.15 * A_2)\mathbf{v} \\ &= 0.85 * (A_1\,\mathbf{v}) + 0.15 * (A_2\,\mathbf{v}) \end{aligned}$$

- Multiplying by A_1: use sparse matrix-vector multiplication you implemented in `Mat`
- Multiplying by A_2: Use the fact that

$$A_2 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \end{bmatrix}$$

Estimating an eigenvalue of *smallest* absolute value

We can (sometimes) use the power method to estimate the eigenvalue of largest absolute value.

What if we want the eigenvalue of *smallest* absolute value?

Lemma: Suppose M is an *invertible* endomorphic matrix. The eigenvalues of M^{-1} are the reciprocals of the eigenvalues of M.

Therefore a small eigenvalue of M corresponds to a large eigenvalue of M^{-1}.

But it's numerically a bad idea to compute M^{-1}. Fortunately, we don't need to!

The vector $\mathbf{w}$ such that $\mathbf{w} = M^{-1}\mathbf{v}$ is exactly the vector $\mathbf{w}$ that solves the equation $M\mathbf{x} = \mathbf{v}$.

```
def power_method(A, k):
    v = normalized random start vector
    for _ in range(k)
        w = M*v
        v = normalized(v)
    return v
```

```
def inverse_power_method(A, k):
    v = normalized random start vector
    for _ in range(k)
        w = solve(M, v)
        v = normalized(v)
    return v
```

Computing an eigenvalue: Shifting and inverse power method

You should be able to prove this:

Lemma:[Shifting Lemma] Let A be an endomorphic matrix and let μ be a number. Then λ is an eigenvalue of A if and only if $\lambda - \mu$ is an eigenvalue of $A - \mu\mathbb{1}$.

Idea of shifting: Suppose you have an estimate μ of some eigenvalue λ of matrix A.
You can test if estimate is perfect, i.e. if μ is an eigenvalue of A. Suppose not ....

If μ is close to λ then $A - \mu\mathbb{1}$ has an eigenvalue that is close to zero.
Idea: Use inverse power method on $(A - \mu\mathbb{1})$ to estimate smallest eigenvalue.

Computing an eigenvalue: Putting it together

Idea for algorithm:

- Shift matrix by estimate μ: $A - \mu\mathbb{1}$
- Use multiple iterations of inverse power method to estimate eigenvector for smallest eigenvalue of $A - \mu\mathbb{1}$
- Use new estimate for new shift.

Faster: Just use one iteration of inverse power method $\Rightarrow$ slightly better estimate $\Rightarrow$ use to get better shift.

```
def inverse_iteration(A, mu):
    I = identity matrix
    v = normalized random start vector
    for i in range(10):
        M = A - mu*I
        w = solve(M, v)
        v = normalized(w)
        mu = v*A*v
        if A*v == mu*v: break
    return mu, v
```

Computing an eigenvalue: Putting it together

Could repeatedly

- Shift matrix by estimate μ: $A - \mu\mathbb{1}$
- Use multiple iterations of inverse power method to estimate eigenvector for smallest eigenvalue of $A - \mu\mathbb{1}$
- Use new estimate for new shift.

Faster: Just use one iteration of inverse power method $\Rightarrow$ slightly better estimate $\Rightarrow$ use to get better shift.

```
def inverse_iteration(A, mu):
    I = identity matrix
    v = normalized random start vector
    for i in range(10):
        M = A - mu*I
        try:
            w = solve(M, v)
        except ZeroDivisionError:
            break
        v = normalized(w)
        mu = v*A*v
        test = A*v - mu*v
        if test*test < 1e-30: break
    return mu, v
```

```
def inverse_iteration(A, mu):
    I = identity matrix
    v = normalized random start vector
    for i in range(10):
        M = A - mu*I
        w = solve(M, v)
        v = normalized(w)
        mu = v*A*v
        if A*v == mu*v: break
    return mu, v
```

```
def inverse_iteration(A, mu):
    I = identity matrix
    v = normalized random start vector
    for i in range(10):
        M = A - mu*I
        try:
            w = solve(M, v)
        except ZeroDivisionError:
            break
        v = normalized(w)
        mu = v*A*v
        test = A*v - mu*v
        if test*test < 1e-30: break
    return mu, v
```

Limitations of eigenvalue analysis

We've seen:

- Every endomorphic matrix does have an eigenvalue ☺ but the eigenvalue might not be a real number ☹.
- Not every endomorphic matrix is diagonalizable ☹
- (Therefore) not every $n \times n$ matrix M has n linearly independent eigenvectors.

This is usually not a big problem since most endomorphic matrics are diagonalizable, and also there are methods of analysis that can be used even when not.

However, there is a class of matrices that arise often in applications for which everything is nice ....

Definition: Matrix A is *symmetric* if $A^T = A$.

Example: $\begin{bmatrix} 1 & 2 & -4 \\ 2 & 9 & 0 \\ -4 & 0 & 7 \end{bmatrix}$

Theorem: Let A be a symmetric matrix over $\mathbb{R}$. Then there is an orthogonal matrix Q and diagonal matrix Λ over $\mathbb{R}$ such that $Q^T A Q = \Lambda$

Eigenvalues for symmetric matrices

Theorem: Let A be a symmetric matrix over $\mathbb{R}$. Then there is an orthogonal matrix Q and diagonal matrix Λ over $\mathbb{R}$ such that $Q^T A Q = \Lambda$

For symmetric matrices, everything is nice:

- $Q \Lambda Q^T$ is a diagonalization of A, so A is diagonalizable!
- The columns of Q are eigenvectors.... Not only linearly independent but mutually orthogonal!
- Λ is over $\mathbb{R}$, so the eigenvalues of A are real!

See text for proof.

Eigenvalues for asymmetric matrices

For asymmetric matrices, eigenvalues might not even be real, and diagonalization need not exist. However, a “triangularization” always exists — called *Schur decomposition* 1

Theorem: Let A be an endomorphic matrix. There is an invertible matrix U and an upper triangular matrix T, both over the complex numbers, such that $A = UTU^{-1}$.

$$\begin{bmatrix} 1 & 2 & 3 \\ -2 & 0 & 2 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} -.127 & -.92 & .371 \\ -.762 & .33 & .557 \\ .635 & .212 & .743 \end{bmatrix} \begin{bmatrix} -2 & -2.97 & .849 \\ 0 & 0 & -2.54 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} -.127 & -.92 & .371 \\ -.762 & .33 & .557 \\ .635 & .212 & .743 \end{bmatrix}^{-1}$$

Recall that the diagonal elements of a triangular matrix are the eigenvalues.

Note that an eigenvalue can occur more than once on the diagonal. We say, e.g. that 12 is an eigenvalue with multiplicity two.

Eigenvalues for asymmetric matrices

For asymmetric matrices, eigenvalues might not even be real, and diagonalization need not exist. However, a "triangularization" always exists — called *Schur decomposition* 2

Theorem: Let A be an endomorphic matrix. There is an invertible matrix U and an upper triangular matrix T, both over the complex numbers, such that $A = UTU^{-1}$.

$$\begin{bmatrix} 27 & 48 & 81 \\ -6 & 0 & 0 \\ 1 & 0 & 3 \end{bmatrix} = \begin{bmatrix} .89 & -.454 & .0355 \\ -.445 & -.849 & .284 \\ .0989 & .268 & .958 \end{bmatrix} \begin{bmatrix} 12 & -29 & 82.4 \\ 0 & 12 & -4.9 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} .89 & -.454 & .0355 \\ -.445 & -.849 & .284 \\ .0989 & .268 & .958 \end{bmatrix}^{-}$$

Recall that the diagonal elements of a triangular matrix are the eigenvalues.

Note that an eigenvalue can occur more than once on the diagonal. We say, e.g. that 12 is an eigenvalue with multiplicity two.

"Positive definite," "Positive semi-definite", and "Determinant"

Let A be an $n \times n$ matrix. Linear function $f(\mathbf{x}) = A\mathbf{x}$ maps an n-dimensional "cube" to an n-dimensional parallelpiped.

$$\text{“cube”} = \{[x_1, \ldots, x_n] \ : \ 0 \leq x_i \leq 1 \text{ for } i = 1, \ldots, n\}$$

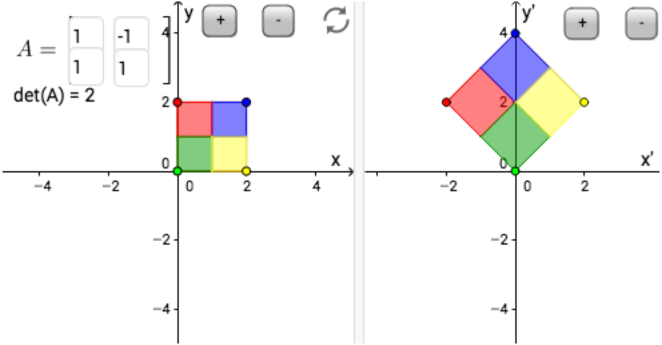
The n-dimensional volume of the input cube is 1.

The *determinant* of A ($\det A$) measures the volume of the output parallelpiped.

Example: $A = \begin{bmatrix} 2 & & \\ & 3 & \\ & & 4 \end{bmatrix}$ turns a $1 \times 1 \times 1$ cube into a $2 \times 3 \times 4$ box.

Volume of box is $2 \cdot 3 \cdot 4$. Determinant of A is 24.

Square to square



Signed volume

A can "flip" a square, in which case the determinant of A is negative.

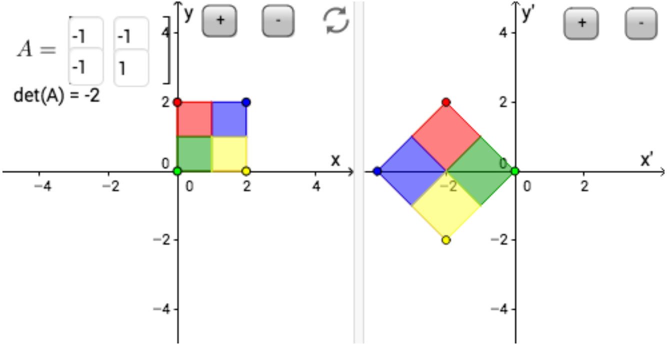
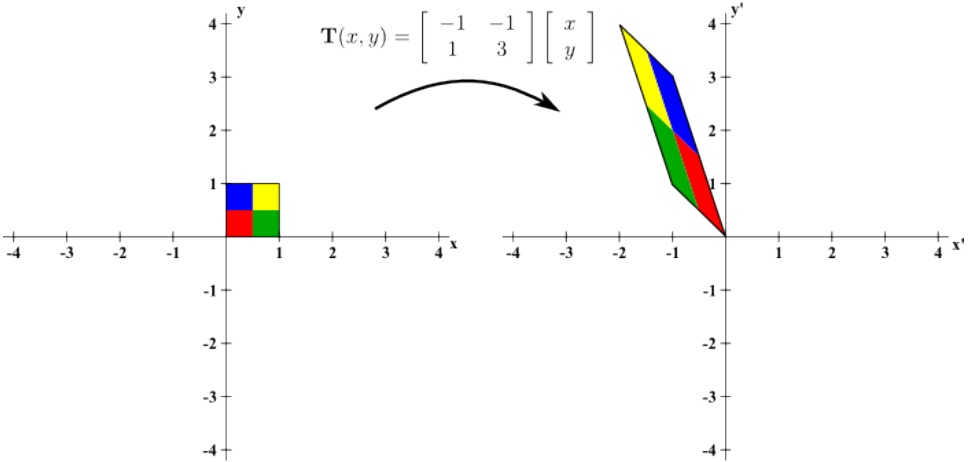


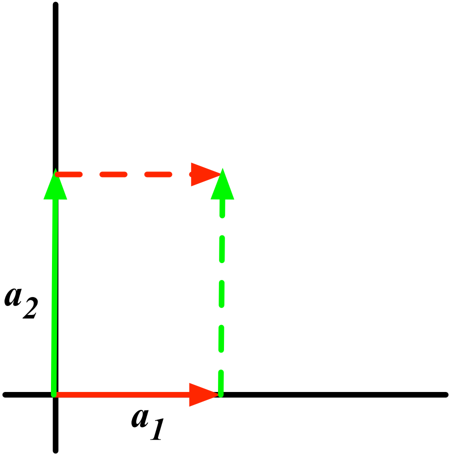
Image of square is a parallelogram



The area of parallelogram is 2, and flip occured, so determinant is -2.

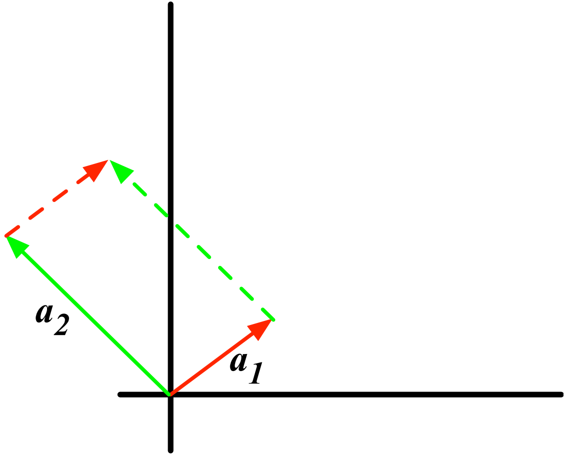
Special case: diagonal matrix

If A is diagonal, e.g. $A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$, image of square is a rectangle with area = product of diagonal elements.



Special case: orthogonal columns in dimension two

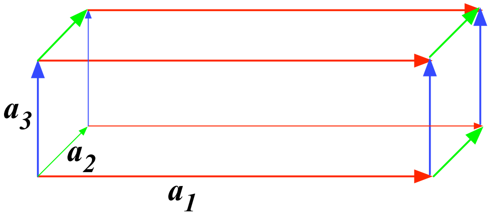
Let $A = \begin{bmatrix} \sqrt{2} & -\sqrt{9/2} \\ \sqrt{2} & \sqrt{9/2} \end{bmatrix}$. Then the columns of A are orthogonal, and their lengths are 2 and 3, so the area is again 6.



Special case: orthogonal columns in higher dimension

If $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$ where $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are mutually orthogonal then image of hypercube is a hyperrectangle

$$\{\alpha_1 \mathbf{a}_1 + \cdots + \alpha_n \mathbf{a}_n \ : 0 \leq \alpha_1, \ldots, \alpha_n \leq 1\}$$

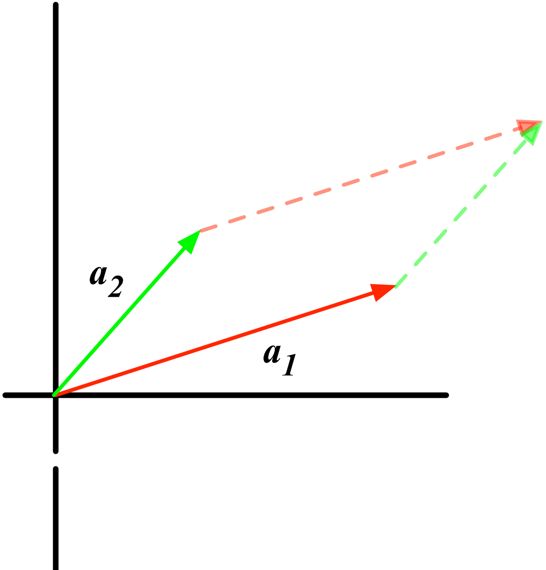


whose sides $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are mutually orthogonal, so volume is $\|\mathbf{a}_1\| \ldots \|\mathbf{a}_n\|$

so determinant is $\pm\|\mathbf{a}_1\| \ldots \|\mathbf{a}_n\|$.

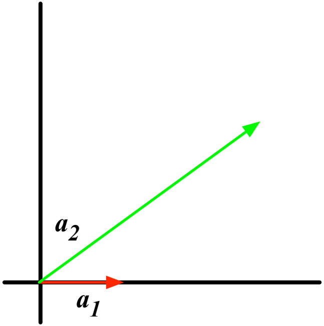
Non-orthogonal columns

If columns of A are non-orthogonal vectors $\mathbf{a}_1, \mathbf{a}_2$ then image of square is a parallelogram.



Example of non-orthogonal columns: triangular matrix

Columns of $\begin{bmatrix} 1 & 3 \\ 0 & 2 \end{bmatrix}$ are $\mathbf{a}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{a}_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$.



Lengths of orthogonal projections are the absolute values of the diagonal elements.

In fact, determinant is product of diagonal elements

Image of cube is parallelpiped

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & -1 \\ -3 & -1 & 2 \end{bmatrix}$$

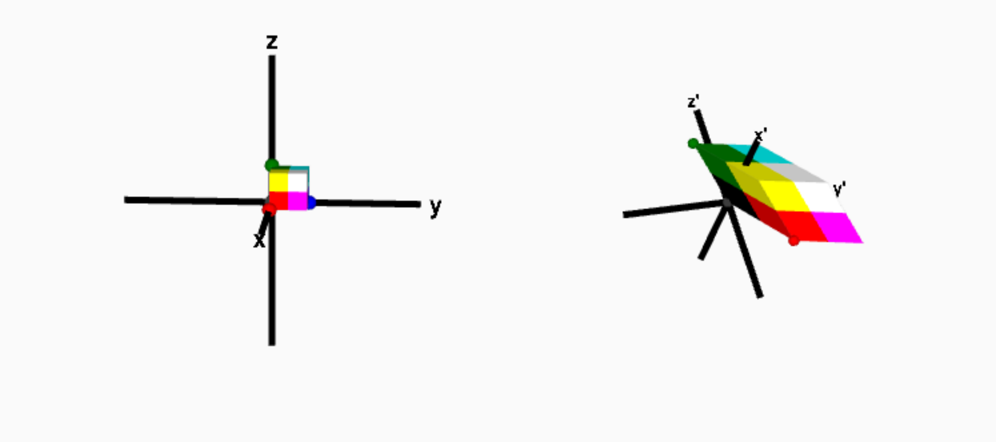


Image of a parallelpiped

If input is a parallelpiped instead of hypercube, determinant of A gives (signed) ratio

$$\frac{\text{volume of output}}{\text{volume of input}}$$

What is $\det AB$?

When matrices multiply, functions compose, so blow-ups in volume multiply:

Key Fact: $\det(AB) = \det(A)\det(B)$

Since det(identity matrix) is 1, $\det(A^{-1}) = 1/\det(A)$

Determinant and triangular matrices

Consider triangularization $A = UTU^{-1}$.

$$\begin{bmatrix} 27 & 48 & 81 \\ -6 & 0 & 0 \\ 1 & 0 & 3 \end{bmatrix} = \begin{bmatrix} .89 & -.454 & .0355 \\ -.445 & -.849 & .284 \\ .0989 & .268 & .958 \end{bmatrix} \begin{bmatrix} 12 & -29 & 82.4 \\ 0 & 12 & -4.9 \\ 0 & 0 & 6 \end{bmatrix} \begin{bmatrix} .89 & -.454 & .0355 \\ -.445 & -.849 & .284 \\ .0989 & .268 & .958 \end{bmatrix}^{-1}$$

Shows $\det A = \det T$

Thus $\det A$ is the product of eigenvalues (taking into account multiplicities)

Measure n-dimensional volume

For $n \times n$ matrix, must measure n-dimensional volume.

Consider $A = \begin{bmatrix} 3 & 2 & 1 \\ 5 & 4 & 1 \\ 10 & 9 & 1 \end{bmatrix}$

Cols $\mathbf{a}_1, \ldots, \mathbf{a}_3$ are linearly dependent, so

$$\{\alpha_1\,\mathbf{a}_1 + \alpha_2\,\mathbf{a}_2 + \alpha_3\,\mathbf{a}_3 \ : \ 0 \leq \alpha_1, \alpha_2, \alpha_3 \leq 1\}$$

is two-dimensional
so volume is zero.

Key Fact: If columns are linearly dependent then determinant is zero.

Multilinearity

Key Fact: The determinant of $n \times n$ matrix can be written as a sum of (many) terms, each a (signed) product of n entries of matrix.

- 2×2 matrix A: $A[1,1]\ A[22] - A[1,2]\ A[2,1]$
- 3×3 matrix A: $A[1,1]A[2,2]A[3,3] - A[1,1]A[2,3]A[32] - A[1,2]A[2,1]A[3,3] + A[1,2]A[2,3]A[3,1] + A[1,3]A[2,1]A[3,2] - A[1,3]A[2,2]A[3,1]$
- 4×4?

Number of terms is $n!$ so not a good way of computing determinants of big matrices!

Better algorithms use matrix factorizations.

Uses of determinants

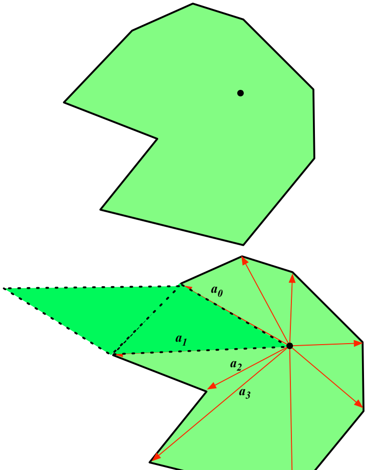
Mathematically useful but computationally not so much.

- Testing a matrix for invertibility? Good in theory but our other methods are better numerically.
- Arises in chain rule for multivariate calculus.
- Can be used to find eigenvalues—but in practice, other methods are better.

Area of polygon

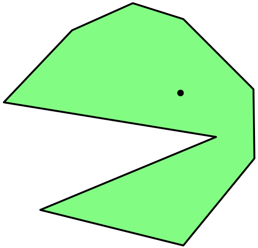
Polygon with vertices $\mathbf{a}_0, \mathbf{a}_1, \ldots, \mathbf{a}_{n-1}$. Break it into triangles:

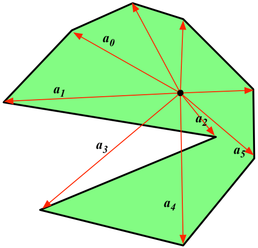
- triangle formed by origin with $\mathbf{a}_0$ and $\mathbf{a}_1$,
- with $\mathbf{a}_1$ and a_2,
- etc.



Area of polygon

What if polygon looks like this?





Method fails because triangles are not disjoin and don't lie within polygon.

Works if you use signed area.