

FFTs in Graphics and Vision

Fast String Matching and Math Review

Fast Pattern Matching in Strings Knuth et al., 1977

Outline



Fast Substring Matching

Math Review

- Complex Numbers
- Vector Spaces
- Linear Operators



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Example:

$$T = CDB$$



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

- For each position in S:
 - Test if the next |T| letters in S match T

$$S=ACDBEFCDBE$$
 $T=CDB$



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

- For each position in S:
 - \circ Test if the next |T| letters in S match T

$$S$$
=ACDBEFCDBE T =CDB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

- For each position in S:
 - \circ Test if the next |T| letters in S match T

$$S=ACDB$$
EFCDBE $T=CDB$



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

- For each position in S:
 - \circ Test if the next |T| letters in S match T

$$S=ACDBEFCDBE$$
 $T=CDB$



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

- For each position in S:
 - \circ Test if the next |T| letters in S match T

$$S=ACDBEFCDBE$$
 $T=CDB$
 CDB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

- For each position in S:
 - \circ Test if the next |T| letters in S match T

$$S=ACDBEFCDBE$$
 $T=CDB$
 CDB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

- For each position in S:
 - \circ Test if the next |T| letters in S match T

$$S=ACDB$$
EFCDBE $T=CDB$ CDB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

- For each position in S:
 - \circ Test if the next |T| letters in S match T

$$S=ACDBEFCDBE$$
 $T=CDB$
 CDB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

- For each position in S:
 - \circ Test if the next |T| letters in S match T

$$S=ACDBEFCDBE$$
 $T=CDB$
 CDB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

- For each position in S:
 - \circ Test if the next |T| letters in S match T

$$S=ACDBEFCDBE$$
 $T=CDB$ $O(|S| \times |T|)$



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Brute Force:

- For each position in S:
 - Test if the next |T| letters in S match T

Can we do this more efficiently?



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Observation:

On a failed match, we don't have to compare all |T| letters in T:

$$S=ACDBEFCDBE$$
 $T=CDB$



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Observation:

On a failed match, we don't have to compare all |T| letters in T:

S=ACDBEFCDBE CDB T = CDB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Observation:

On a failed match, we don't have to compare all |T| letters in T:

S=ACDBEFCDBE T=CDB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Observation:

On a failed match, we don't have to compare all |T| letters in T:

S=ACDBEFCDBE CDB

T = CDB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Observation:

On a failed match, we don't have to compare all |T| letters in T:

S=ACDBEFCDBE CDB

T = CDB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Observation:

On a failed match, we don't have to compare all |T| letters in T:

S=ACDBEFCDBE CDB T = CDB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Observation:

On a failed match, we don't have to compare all |T| letters in T:

S=ACDBEFCDBE T=CDBCDB Comparisons: 1



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Observation:

On a failed match, we don't have to compare all |T| letters in T:

S=ACDBEFCDBE T=CDB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Observation:

On a failed match, we don't have to compare all |T| letters in T:

S=ACDBEFCDBE T=CDBCDB Comparisons: 1



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Observation:

What if the situation is more complex?

S = AAAAAAAAAAB

T = AAAB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Observation:

What if the situation is more complex?

S=AAAAAAAAB AAAB T = AAAB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Observation:

What if the situation is more complex?

S=AAAAAAAAB AAAB T = AAAB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Observation:

What if the situation is more complex?

S=AAAAAAAAB AAAB T = AAAB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Observation:

What if the situation is more complex?

S=AAAAAAAAB AAAB T = AAAB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Observation:

What if the situation is more complex?

S=AAAAAAAAB AAAB T = AAAB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Observation:

What if the situation is more complex?

S=AAAAAAAAB AAAB T = AAAB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Observation:

What if the situation is more complex?

S=AAAAAAAAAAAAB AAAB

T = AAAB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

On a failed match, we don't have to re-start the matching.

T = AAAB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

On a failed match, we don't have to re-start the matching.

$$T = AAAB$$



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

On a failed match, we don't have to re-start the matching.

T = AAAB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

On a failed match, we don't have to re-start the matching.

T = AAAB



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

On a failed match, we don't have to re-start the matching.



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

On a failed match, we don't have to re-start the matching.



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

On a failed match, we don't have to re-start the matching.



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

On a failed match, we don't have to re-start the matching.



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

On a failed match, we don't have to re-start the matching.



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

On a failed match, we don't have to re-start the matching.



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

On a failed match, we don't have to re-start the matching.



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

On a failed match, we don't have to re-start the matching.



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

On a failed match, we don't have to re-start the matching.



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

On a failed match, we don't have to re-start the matching.



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

On a failed match, we don't have to re-start the matching.



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

On a failed match, we don't have to re-start the matching.



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

On a failed match, we don't have to re-start the matching.

$$S = AAAAAAAAAAB$$

$$T = AAAB$$

$$O(|S| + |T|)$$



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

On a failed match, we don't have to re-start the matching.

The key is to know where in T we have to start comparing.



Challenge:

Given strings S and T, find all occurrences of T as a substring of S.

Knuth et al. (1977):

The size of the shift on a mismatch is determined by the repetitions in T, is independent of S, and can be computed in O(|T|) time.

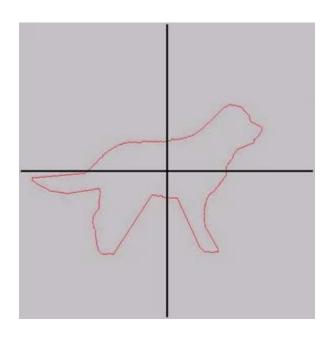
For more details, see:

Fast Pattern Matching in Strings.



Recall:

Our goal is to perform registration and symmetry detection on the circle.

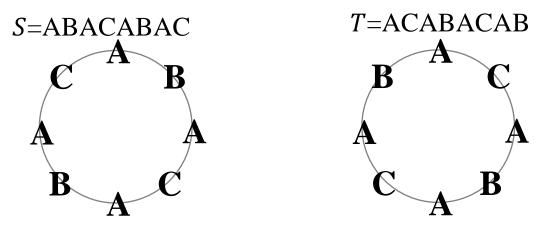




Applications:

If we think of a string as a signal on a circle:

We can test if signal T is a rotation of S
 by testing if T is a substring of SS



SS=ABACABACABACABAC

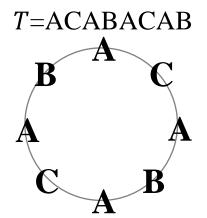


Applications:

If we think of a string as a signal on a circle:

We can test if signal T is a rotation of S
 by testing if T is a substring of SS





SS=ABACABACABACABAC ACABACAB

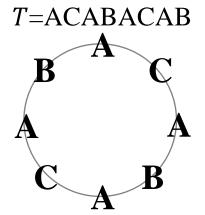


Applications:

If we think of a string as a signal on a circle:

We can test if signal T is a rotation of S
 by testing if T is a substring of SS





SS=ABACABACABACABAC ACABACAB

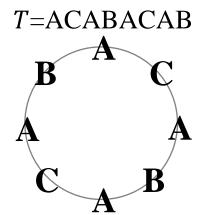


Applications:

If we think of a string as a signal on a circle:

We can test if signal T is a rotation of S
 by testing if T is a substring of SS





SS=ABACABACABACABAC ACABACABACABACAB

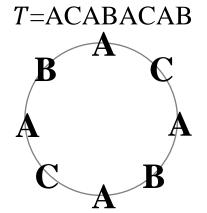


Applications:

If we think of a string as a signal on a circle:

We can test if signal T is a rotation of S
 by testing if T is a substring of SS





SS=ABACABACABACABAC ACABACAB

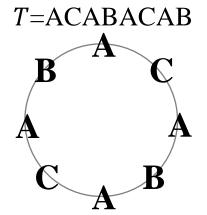


Applications:

If we think of a string as a signal on a circle:

We can test if signal T is a rotation of S
 by testing if T is a substring of SS





SS=ABACABACABACABAC ACABACABACABACABACAB

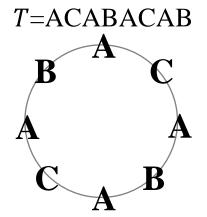


Applications:

If we think of a string as a signal on a circle:

We can test if signal T is a rotation of S
 by testing if T is a substring of SS





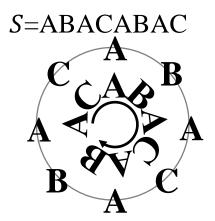
SS=ABACABACABACABAC ACABACAB

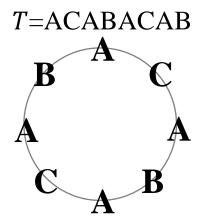


Applications:

If we think of a string as a signal on a circle:

We can test if signal T is a rotation of S
 by testing if T is a substring of SS





SS=ABACABACABACABACABACABACABACAB

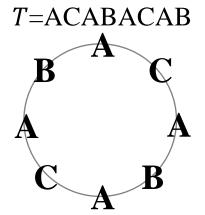


Applications:

If we think of a string as a signal on a circle:

We can test if signal T is a rotation of S
 by testing if T is a substring of SS





SS=ABACABACABACABAC ACABACAB

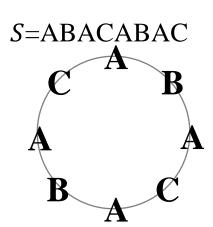


Applications:

If we think of a string as a signal on a circle:

- We can test if signal T is a rotation of S
 by testing if T is a substring of SS
- We can test if S has rotational symmetry by testing if S is a substring of SS

SS=ABACABACABACABAC
S=ABACABAC

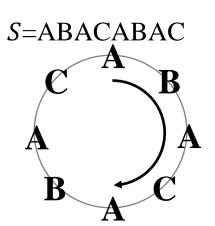




Applications:

If we think of a string as a signal on a circle:

- We can test if signal T is a rotation of S
 by testing if T is a substring of SS
- We can test if S has rotational symmetry by testing if S is a substring of SS



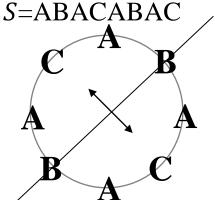


Applications:

If we think of a string as a signal on a circle:

- We can test if signal T is a rotation of S
 by testing if T is a substring of SS
- We can test if S has rotational symmetry by testing if S is a substring of SS
- We can test if S has reflective symmetry by testing if S is a substring of $(SS)^{\mathsf{T}}$ S=

 $(SS)^T$ =CABACABACABACABACABACABACABAC





- ✓ A fast (linear time) algorithm for performing pattern detection on discrete signals
- Can only tell us if there is a perfect match
 - For real-world data, we need a continuous measure of similarity
- Only works for signals on a circle (or a line)
 - Hard to generalize to signals on other domains

Outline



Fast Substring Matching

Math Review

- Complex Numbers
- Vector Spaces
- Linear Operators



A complex number $c \in \mathbb{C}$ is any number that can be written as:

$$c = a + ib$$

with $a, b \in \mathbb{R}$ and i a square root of -1:

$$i^2 = -1$$



Given two complex numbers

$$c_1 = a_1 + ib_1$$
 and $c_2 = a_2 + ib_2$

The sum of the numbers is:

$$c_1 + c_2 = (a_1 + a_2) + i(b_1 + b_2)$$

• The product of the numbers is:

$$c_1 \cdot c_2 = (a_1 + ib_1) \cdot (a_2 + ib_2)$$

$$= a_1 \cdot a_2 + ib_1 \cdot ib_2 + a_1 \cdot ib_2 + ib_1 \cdot a_2$$

$$= (a_1 \cdot a_2 - b_1 \cdot b_2) + i(a_1 \cdot b_2 + b_1 \cdot a_2)$$



Given a complex numbers:

$$c = a + ib$$

The negation of the number is:

$$-c = -a - ib$$

The conjugate of the number is:

$$\bar{c} = a - ib$$

The reciprocal of the number is:

$$\frac{1}{c} = \frac{1}{c} \cdot \frac{\bar{c}}{\bar{c}} = \frac{a - ib}{a^2 + b^2}$$



Why do we care?



Why do we care?

Fundamental Theorem of Algebra

Given any polynomial:

$$p(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n$$

there always exists a complex number $c \in \mathbb{C}$ s.t.:

$$p(c) = 0$$

Vector Spaces



A (real/complex) vector space V is a <u>set</u> of elements $v \in V$, with:

- An <u>addition</u> operator "+", and
- A scaling operator "."

(Adding two vectors gives a vector and scaling a vector by a real/complex number gives a vector.)

Vector Spaces (Formal Properties 1)



For all $u, v, w \in V$:

Associative addition:

$$(u+v)+w=u+(v+w)$$

Commutative addition:

$$u + v = v + u$$

Additive identity:

There exists a unique vector $0 \in V$ such that:

$$v + 0 = v$$

Additive inverse:

There exists a vector $(-v) \in V$ such that:

$$v + (-v) = 0$$

Vector Spaces (Formal Properties 2)



For all $u, v \in V$, and $\alpha, \beta \in \mathbb{R}/\mathbb{C}$:

Distributive over vector addition:

$$\alpha \cdot (u + v) = (\alpha \cdot u) + (\alpha \cdot v)$$

Distributive over scalar addition:

$$(\alpha + \beta) \cdot u = (\alpha \cdot u) + (\beta \cdot u)$$

Compatible scalar multiplication:

$$\alpha \cdot (\beta \cdot u) = (\alpha \cdot \beta) \cdot u$$

Scalar Identity:

$$1 \cdot u = u$$

Vector Spaces: Examples



Real Vector Spaces:

- The real / complex numbers
- \circ The space of n-dimensional arrays with real / complex entries
- The space of $m \times n$ matrices with real / complex entries
- The space of real / complex valued functions on a circle / line / plane / sphere / etc.

Complex Vector Spaces:

- The complex numbers
- \circ The space of n-dimensional arrays with complex entries
- \circ The space of $m \times n$ matrices with complex entries
- The space of complex valued functions on a circle / line / plane / sphere / etc.

Vector Spaces



Note:

With the exception of the last corollary of today's lecture, the following discussion is identical for real and complex vector spaces.



A basis of a vector-space V is a set $\{v^1, ..., v^n\} \subset V$ such that:

1. Any vector $v \in V$ can be expressed as:

$$v = a_1 \cdot v^1 + \dots + a_n \cdot v^n$$
 with $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$.

2. No basis vector v^i can be expressed as the linear sum of the other basis vectors.

The dimension of *V* is the number basis vectors and does not depend on the particular choice of basis.



Given an n-dimensional vector space V, if we have a basis $\{v^1, ..., v^n\}$, we can associate vectors in V with vectors in \mathbb{R}^n :

$$v \in V \leftrightarrow \mathbf{a} \in \mathbb{R}^n$$

by setting $\mathbf{a} = (a_1, ..., a_n)$ to be the coefficients of v with respect to the basis $\{v^1, ..., v^n\}$:

$$v = a_1 \cdot v^1 + \dots + a_n \cdot v^n$$



Many different bases can be used to represent the same vector space.

Example:

Consider the set of points in 2D Euclidean space.

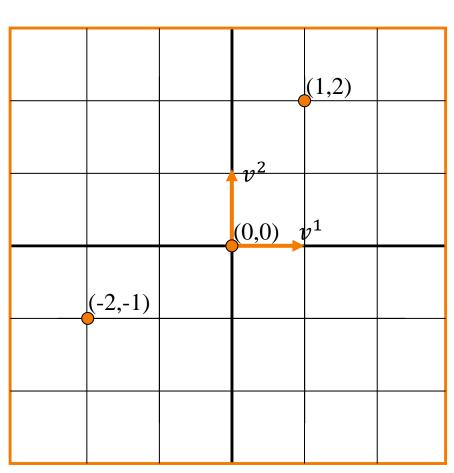


Many different bases can be used to represent the same vector space.

Example:

Consider the set of points in 2D Euclidean space.

We can represent each vector in terms of its (x, y)-coordinates.



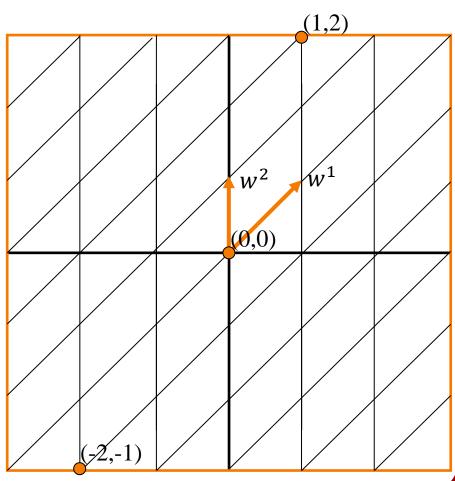


Many different bases can be used to represent the same vector space.

Example:

Consider the set of points in 2D Euclidean space.

Or we could use a different basis...



Linear Maps



Given vector spaces V and W, the map $L: V \to W$, is <u>linear</u> if for all $v, w \in V$ and $\alpha, \beta \in \mathbb{R}$:

$$L(\alpha v + \beta w) = \alpha L(v) + \beta L(w)$$

If it exists, the inverse of a linear map L is the map L^{-1} with the property that:

$$L^{-1}\big(L(v)\big) = v$$

Linear Maps



If $L: V \to W$, is a linear map:

The set of vectors:

$$K = \{v \in V | L(v) = 0\}$$

is a vector subspace called the kernel.

The set of vectors:

$$I = \{ w \in W | \exists v \in V \text{ s. t. } w = L(v) \}$$

is a vector subspace called the image.

Matrices



Given an n-dimensional vector space V with basis $\{v^1, ..., v^n\}$ and an m-dimensional vector space W with basis $\{w^1, ..., w^m\}$, a linear map $L: V \to W$ can be uniquely expressed by a $\underline{\text{matrix}} \ \mathbf{L} \in \mathbb{R}^{m \times n}$ s.t.:

$$b = La$$

whenever:

$$b_1 \cdot w^1 + \dots + b_m \cdot w^m = L(a_1 \cdot v^1 + \dots + a_n \cdot v^n)$$



Given an n-dimensional vector space V, and two bases $\{v^1, ..., v^n\}$ and $\{w^1, ..., w^n\}$, the change of basis matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$ is the matrix satisfying:

$$\mathbf{b} = \mathbf{B}\mathbf{a}$$

whenever:

$$b_1 \cdot w^1 + \dots + b_n \cdot w^n = a_1 \cdot v^1 + \dots + a_n \cdot v^n$$



Given:

- An n-dimensional vector space V,
- Two bases $\{v^1, \dots, v^n\}$ and $\{w^1, \dots, w^n\}$,
- A linear operator L represented by the matrix L in terms of the basis $\{v^1, \dots, v^n\}$.

The matrix representation for L in terms of the basis $\{w^1, \dots, w^n\}$ is given by:

 $\mathbf{B}\mathbf{L}\mathbf{B}^{-1}$



Why do we care?



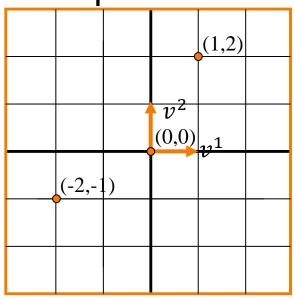
Why do we care?

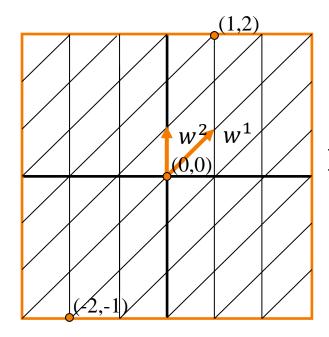
Choosing the appropriate basis can make it much easier to understand a linear operator.



Why do we care?

Example:





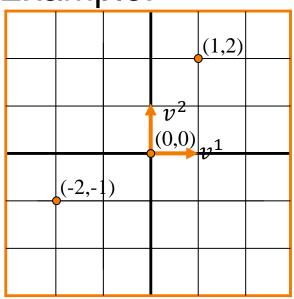
$$\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$
$$\mathbf{B}^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$$

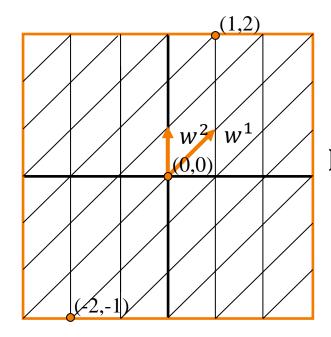
$$\mathbf{L} = \begin{pmatrix} 2 & 0 \\ -1 & 1 \end{pmatrix}$$



Why do we care?

Example:





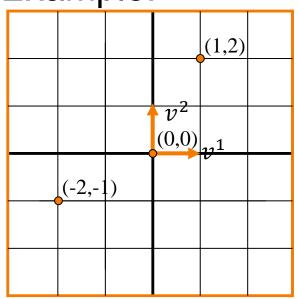
$$\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$
$$\mathbf{B}^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$$

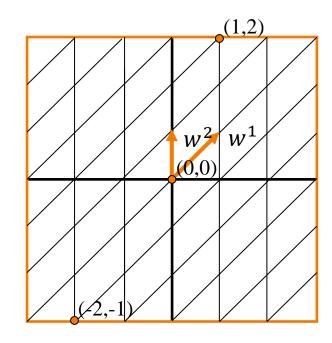
$$\mathbf{L} = \begin{pmatrix} 2 & 0 \\ -1 & 1 \end{pmatrix} \Longrightarrow \mathbf{B}\mathbf{L}\mathbf{B}^{-1}$$



Why do we care?

Example:





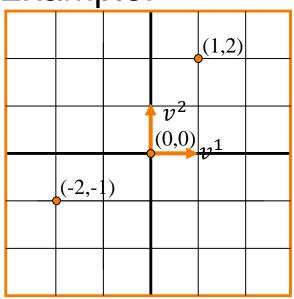
$$\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$
$$\mathbf{S}^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$$

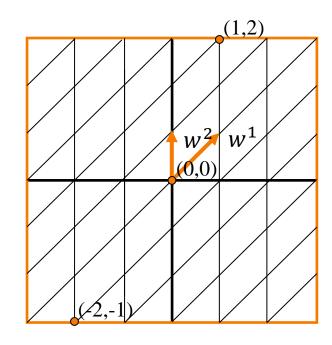
$$\mathbf{L} = \begin{pmatrix} 2 & 0 \\ -1 & 1 \end{pmatrix} \Longrightarrow \mathbf{B} \mathbf{L} \mathbf{B}^{-1} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$$



Why do we care?

Example:





$$\mathbf{B} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$
$$^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$$

$$\mathbf{L} = \begin{pmatrix} 2 & 0 \\ -1 & 1 \end{pmatrix} \Longrightarrow \mathbf{B} \mathbf{L} \mathbf{B}^{-1} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

In this basis the linear operator becomes anisotropic scaling!

Determinants



The <u>determinant</u> is a function that associates a scalar value to every linear map $L: V \to V$.

The determinant of $L: V \to V$ is the (signed) volume of the image of (any) unit cube in V.

Determinants



The determinant of a linear map $L: V \to V$ equals zero if and only if there exists $v \in V$, with $v \neq 0$, such that L(v) = 0.

Eigenvalues and Eigenvectors



The scalar $\lambda \in \mathbb{R}$ is an <u>eigenvalue</u> of a linear operator $L: V \to V$ if there exists $v \in V$ such that: $\lambda \cdot v = L(v)$.

In this case, v is an <u>eigenvector</u> of L.

Eigenvalues and Eigenvectors



If $L: V \to V$ has an eigenpair $(\lambda, v) \in \mathbb{R} \times V$, then: $0 = (L - \lambda \cdot \text{Id.})(v)$

⇒ The linear operator

$$(L - \lambda \cdot \mathrm{Id.}): V \to V$$

has zero determinant.

Characteristic Polynomials



If we treat λ as a variable, the determinant:

$$\chi_L(\lambda) = \det(L - \lambda \cdot \mathrm{Id.})$$

is a polynomial of degree n in λ .

This is the characteristic polynomial of L.

Characteristic Polynomials



The roots of the characteristic polynomial of L:

$$\chi_L(\lambda) = \det(L - \lambda \cdot \mathrm{Id.})$$

are precisely the eigenvalues of L.

Corollary:

If $L: V \to V$ is a linear map on a **complex** vector space, L always has at least one eigenvalue.