Linear Algebra

(a crash course)

Dmytro Velychko

Theoretical Neuroscience lab
Philipps-Universität Marburg

Last edited: October 19, 2017

Some background

Why do we need mathematics?

Why do we need mathematics?

Mathematics - (Greek: $\mu\alpha\theta\eta\mu\alpha$) - knowledge.
Accumulates highly abstract formal knowledge about very abstract categories like quantity, structure, space, change, relation...

Why do we need mathematics?

Mathematics - (Greek: $\mu\alpha\theta\eta\mu\alpha$) - knowledge.
Accumulates highly abstract formal knowledge about very abstract categories like quantity, structure, space, change, relation...

What if you can't tackle you problem with maths?

Why do we need mathematics?

Mathematics - (Greek: $\mu\alpha\theta\eta\mu\alpha$) - knowledge.
Accumulates highly abstract formal knowledge about very abstract categories like quantity, structure, space, change, relation...

What if you can't tackle you problem with maths?

- maybe you are using maths a bit wrong
- maybe you are lucky enough to find a domain worth describing with new maths

A brief history of maths

Ancient Greek mathematics:

- arithmetic, geometry
- only natural and rational numbers

A brief history of maths

Ancient Greek mathematics:

- arithmetic, geometry
- only natural and rational numbers

Later was heavily influenced by research in physics - Newton, Leibnitz, Euler + many others:

- infinitesimal calculus, analysis
- differential equations
- variational calculus
- complex domain
- ...

A brief history of maths

Ancient Greek mathematics:

- arithmetic, geometry
- only natural and rational numbers

Later was heavily influenced by research in physics - Newton, Leibnitz, Euler + many others:

- infinitesimal calculus, analysis
- differential equations
- variational calculus
- complex domain
- ...

Evariste Galois - foundations of abstract algebra:

- rings, fields, lattices, groups, many sorts of algebras...
- linear algebra: vector spaces, linear transformations, tensors...
- ...

A brief history of maths

Ancient Greek mathematics:

- arithmetic, geometry
- only natural and rational numbers

Later was heavily influenced by research in physics - Newton, Leibnitz, Euler + many others:

- infinitesimal calculus, analysis
- differential equations
- variational calculus
- complex domain
- ...

Evariste Galois - foundations of abstract algebra:

- rings, fields, lattices, groups, many sorts of algebras...
- linear algebra: vector spaces, linear transformations, tensors...
- ...

Functional analysis, topology, differential geometry, theory of information, theory of algorithms ...

Foundations of maths

Naive set theory by Georg Cantor (predicate logic + set-defining predicate)

Foundations of maths

Naive set theory by Georg Cantor (predicate logic + set-defining predicate)

But there were many problems with it. Russell's paradox:
Let R be the set of all sets that are not members of themselves. Should R contain itself?

Foundations of maths

Naive set theory by Georg Cantor (predicate logic + set-defining predicate)

But there were many problems with it. Russell's paradox:
Let R be the set of all sets that are not members of themselves. Should R contain itself?

Set-theoretic foundation for mathematics: Zermelo-Fraenkel set theory

- set
- mapping (function)

Foundations of maths

Naive set theory by Georg Cantor (predicate logic + set-defining predicate)

But there were many problems with it. Russell's paradox:
Let R be the set of all sets that are not members of themselves. Should R contain itself?

Set-theoretic foundation for mathematics: Zermelo-Fraenkel set theory

- set
- mapping (function)

Constructive foundation for mathematics: Type theory

- term, type
- mapping (morphism)

Univalent foundations for mathematics: Homotopy type theory, V. Voevodsky, 2011

Essential basic elements

Linear Algebra operates with:

- sets (with special properties)
- functional mappings (with special properties)

Some historical motivation for Linear Algebra

But first - some historical motivation for LA

Historically linear algebra emerged from studies of systems of linear equations, methods of solving them and providing conditions when they have one, many or no solutions. Let's consider a very simple system of equations which share two unknown variables x_1 and x_2:

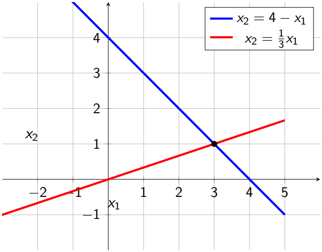
$$\begin{cases} x_1 + x_2 = 4 \\ x_1 - 3x_2 = 0 \end{cases} \tag{1}$$

If unknown variables form a polynomial of order 1 $\rightarrow$ **linear equations**

Row view. Lines on a plane

Each separate equation of the system (5) defines a line. By rearranging the terms one can get the following line equations:

$$\begin{cases} x_2 = 4 - x_1 \\ x_2 = \frac{1}{3}x_1 \end{cases} \tag{2}$$



The intersection point satisfies both equations, as it belongs to both lines, yielding the solution of the system $(x_1 = 3, x_2 = 1)$.

Column view. Vectors in 2D space

Here a naive definition of vector must be introduced.
Defintion: vector is an algebraic object which is described by a set of ordered numbers and two operations: multiplication by a scalar and summation of two vectors. If vectors $\vec{v}$ and $\vec{w}$ both has n elements:

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \qquad \vec{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \tag{3}$$

$$a\vec{v} = \begin{bmatrix} av_1 \\ av_2 \\ \vdots \\ av_n \end{bmatrix} \qquad \vec{v} + \vec{w} = \begin{bmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{bmatrix} \tag{4}$$

Column view. Vectors in 2D space

$$\begin{cases} x_1 + x_2 = 4 \\ x_1 - 3x_2 = 0 \end{cases} \tag{5}$$

Let's rewrite the system (5) in the vector form. For this we have to define 3 vectors, which are column-coefficients in the system of equations:

$$\vec{a_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \vec{a_2} = \begin{bmatrix} 1 \\ -3 \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} 4 \\ 0 \end{bmatrix} \tag{6}$$

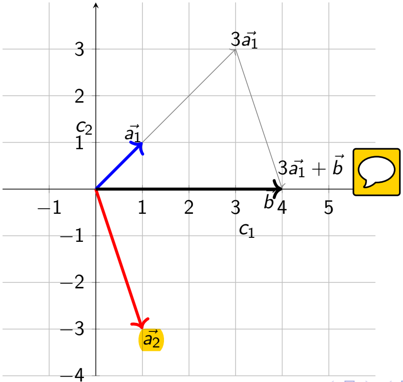
It is easy to see that (5) is equivalent to the vector equation

$$\vec{a_1}x_1 + \vec{a_2}x_2 = \vec{b} \tag{7}$$

Column view. Vectors in 2D space

$$\vec{a_1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \vec{a_2} = \begin{bmatrix} 1 \\ -3 \end{bmatrix} \qquad \vec{b} = \begin{bmatrix} 4 \\ 0 \end{bmatrix} \tag{8}$$

$$\vec{a_1}x_1 + \vec{a_2}x_2 = \vec{b} \tag{9}$$



Algebraic view. System of equations per se

Manipulate the system by

- multiplying an equation by a non-zero number
- adding one equation to another

Separate equations are labelled as E_1 and E_2:

$$\begin{cases} x_1 + x_2 = 4 & (E_1) \\ x_1 - 3x_2 = 0 & (E_2) \end{cases} \tag{10}$$

Subtract E_1 from E_2 and use this difference instead of E_2:

$$E_2 - E_1 \to E_2 : \qquad \begin{cases} x_1 + x_2 = 4 & (E_1) \\ - 4x_2 = -4 & (E_2) \end{cases} \tag{11}$$

Multiply E_2 by $-1/4$:

$$-\frac{1}{4}E_2 \to E_2 : \qquad \begin{cases} x_1 + x_2 = 4 & (E_1) \\ \phantom{x_1 + {}} x_2 = 1 & (E_2) \end{cases} \tag{12}$$

Substitute x_2 into E_1:

$$x_2 \to 1 : \qquad \begin{cases} x_1 \quad + 1 = 4 & (E_1) \\ \quad x_2 \qquad = 1 & (E_2) \end{cases} \tag{13}$$

Add $-1 = -1$ to E_1

$$E_1 + (-1 = -1) \to E_1 : \qquad \begin{cases} x_1 \quad = 3 & (E_1) \\ \quad x_2 = 1 & (E_2) \end{cases} \tag{14}$$

Solving systems of linear equations. Gaussian elimination

Consider a system of 3 equations:

$$\begin{cases} 2x_1 - 4x_2 + x_3 = -3 & (E_1) \\ 3x_1 - 6x_2 + 2x_3 = -3 & (E_2) \\ -3x_1 + x_2 - 2x_3 = -7 & (E_3) \end{cases} \tag{15}$$

For the sake of simplicity we will operate only with the coefficients in a compact form (augmented matrix, where the coefficients of the left hand side form a matrix, and the right hand side augments it with a vector):

$$\left[\begin{array}{rrr|r} 2 & -4 & 1 & -3 \\ 3 & -6 & 2 & -3 \\ -3 & 1 & -2 & -7 \end{array}\right] \tag{16}$$

Solving systems of linear equations. Gaussian elimination

It is also useful to keep in mind the following properties:

- interchanging lines (equations) does not change the solution
- interchanging columns does not change the solution (but in this case one has to track the indexes permutation to recover the solution)

Gaussian elimination

Gaussian elimination consists of 2 parts:

- transforming the system into a so-called row echelon (or upper triangular) form, where all coefficients below the diagonal are zeros
- sequential replacement of unknown variables by the found values

Gaussian elimination

On the first step we want to get the row echelon form:

$$\left[\begin{array}{ccc|c} 2 & -4 & 1 & -3 \\ 3 & -6 & 2 & -3 \\ -3 & 1 & -2 & -7 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} c_{1,1} & c_{1,2} & c_{1,3} & g_1 \\ 0 & c_{2,2} & c_{2,3} & g_2 \\ 0 & 0 & c_{3,3} & g_3 \end{array}\right] \tag{17}$$

Data: augmented matrix C with N rows
Result: row echelon form
for *i in 1 . . . N* **do**
    for *j in i+1 . . . N* **do**
        $row_j \leftarrow row_j - \frac{c_{j,i}}{c_{i,i}} * row_i$;
    end
end
return C

Gaussian elimination

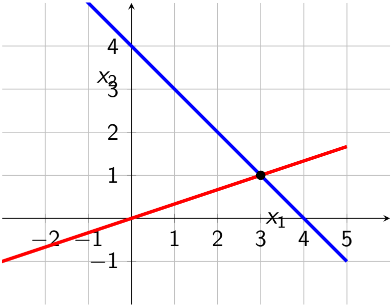
Then get zeroes everywhere but the diagonal of ones.
Solution in the right-side part:

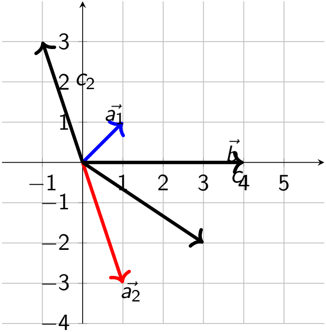
$$\left[\begin{array}{ccc|c} c_{1,1} & c_{1,2} & c_{1,3} & g_1 \\ 0 & c_{2,2} & c_{2,3} & g_2 \\ 0 & 0 & c_{3,3} & g_3 \end{array}\right] \rightarrow \left[\begin{array}{ccc|c} 1 & 0 & 0 & h_1 \\ 0 & 1 & 0 & h_2 \\ 0 & 0 & 1 & h_3 \end{array}\right] \tag{18}$$

Data: row echelon form matrix C
Result: solution of the system
for *i in N . . . 1* **do**
    $row_i \leftarrow \frac{row_i}{c_{i,i}}$;
    for *j in i-1 . . . 1* **do**
        $row_j \leftarrow row_j - c_{j,i} * row_i$;
    end
end
return right column of C

Number of solutions

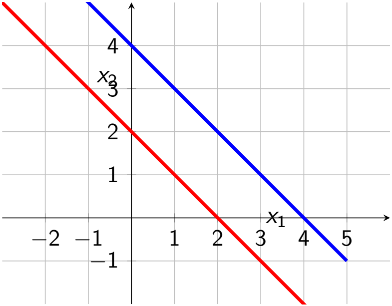
One solution. Lines intersect, vectors are not collinear:

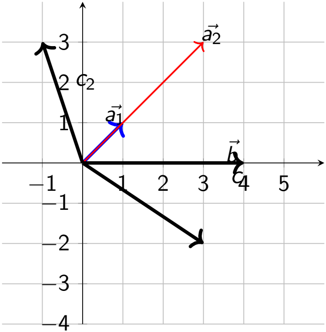




Number of solutions

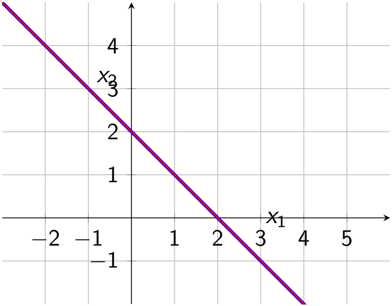
No solution. Lines are parallel, some vectors except the goal vector are collinear:

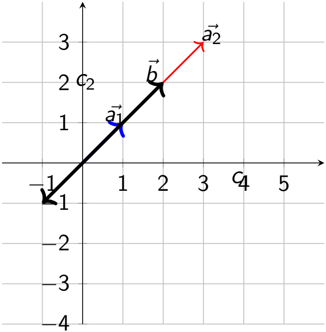




Number of solutions

Infinite number of solutions. Lines overlap, goal vector is collinear with base vectors:





Going abstract:
Linear transformations

linear transformations - core objects of LA

Linear transformations (as any other transformations) map objects some spaces to other spaces. Linear maps and the spaces they operate on have some fundamental properties, which make them interesting and relatively easy to operate with.

Definition

Linear map is a functional mapping $f : V \to W$ which satisfies the following two properties:

$$f(\mathbf{x} + \mathbf{y}) = f(\mathbf{x}) + f(\mathbf{y}) \tag{19}$$

$$f(\alpha \mathbf{x}) = \alpha f(\mathbf{x}) \tag{20}$$

where $\mathbf{x}, \mathbf{y} \in V, \alpha \in \mathbb{R}$.

These properties are additivity w.r.t. the sum operation, and homogenity of degree 1 (scaling by real numbers) respectively.

Vector spaces - core objects of LA

From this definition we may deduce that the spaces V and W must contain special elements with some operations (sum and multiplication by a scalar) defined on them. If $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in V$, then the following properties should hold:

$$\mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3) = (\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 \quad \text{associativity} \quad (21)$$

$$\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1 \quad \text{commutativity} \quad (22)$$

$$\alpha(\mathbf{v}_1 + \mathbf{v}_2) = \alpha\mathbf{v}_1 + \alpha\mathbf{v}_2 \quad \text{distributivity} \quad (23)$$

$$(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v} \quad \text{distributivity} \quad (24)$$

$$\exists\mathbf{0} : \mathbf{v} + \mathbf{0} = \mathbf{v} \quad \text{zero element} \quad (25)$$

$$1 * \mathbf{v} = \mathbf{v} \quad \text{identity w.r.t *1} \quad (26)$$

$$\forall\mathbf{v} : \exists\mathbf{w} \in V : \mathbf{v} + \mathbf{w} = \mathbf{0}; \mathbf{w} = -1 * \mathbf{v} \quad \text{inverse element} \quad (27)$$

Such spaces are called **vector spaces**, and elements of such spaces - **vectors**.

Vector spaces - core objects of LA

Some examples of vector spaces:

- velocities
- real numbers $\mathbb{R}$
- polynomials
- matrices
- rotations in $\mathbb{R}^2$
- space of linear transformations
- functions $\mathbb{R}^N \to \mathbb{R}^M$

One may easily check whether the properties (21) - (27) hold for such spaces.

Vector spaces - core objects of LA

Also here are some examples of sets, which are not vector spaces:

- integers (in general, integer multiplied by a real number may not belong to the set of integers)
- positive real numbers $\mathbb{R}^+$ (multiplied by a negative real $\notin \mathbb{R}^+$)
- probability distributions (as they must be normed to 1)
- colors (there are no negative colors)
- rotations in $\mathbb{R}^3$ (not commutative)

From the property (20) of linear transformations one may derive that any linear transformation, applied to a zero vector in V, should yield zero vector in W. Thus, a mapping $f_1(x) : x \to ax$, where $a, x \in \mathbb{R}$, is a linear map (it is clear, that $0 \to 0$). Although, $f_2(x) : x \to ax + b$, where $a, b, x \in \mathbb{R}$ is not a linear map (because $0 \to b$). However, it is possible to extend the mapping f_2 and make it linear.

Vectors

Next, we will introduce **vectors in** $\mathbb{R}^N$, which are basically tuples of N real numbers: $\mathbf{v} = [v_1, \ldots, v_N]$. To make this set a vector space, it must be augmented with * and + operations, which satisfy the requirements (21) … (27):

$$a\mathbf{v} = [av_1, \ldots, av_N] \tag{28}$$

$$\mathbf{v}_1 + \mathbf{v}_2 = [v_{1,1} + v_{2,1}, \ldots, v_{1,N} + v_{2,N}] \tag{29}$$

Notice, that these definitions of operations are equal to (4).

Linear combinations

Definition

Linear combination of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_N \in V$ *by a vector* $\mathbf{a} \in \mathbb{R}^N$ *is a transformation* $T_{\mathbf{v}} : \mathbb{R}^N \to V$*:*

$$T_{\mathbf{v}}(\mathbf{a}) = T_{\mathbf{v}}([a_1, \ldots, a_N]) \tag{30}$$

$$= a_1\mathbf{v}_1 + \ldots + a_N\mathbf{v}_N \tag{31}$$

$$= \sum_{i=1}^{N} a_i\mathbf{v}_i \tag{32}$$

(*Exercise*: prove that linear combination is a linear transformation).

Spanning set

Definition

Spanning set *of some set of vectors* $\mathbf{v}_1, \ldots, \mathbf{v}_N \in V$ *is the set of all vectors* $\mathbf{s}_i$*, which can be constructed by a linear combination of* $\mathbf{v}_1, \ldots, \mathbf{v}_N$*:*

$$span(\mathbf{v}_1, \ldots, \mathbf{v}_N) = \{\mathbf{s}_i | \exists \mathbf{a} : T_{\mathbf{v}}(\mathbf{a}) = \mathbf{s}_i\} \tag{33}$$

Basis

Definition

Basis *is any minimal set of vectors which spanning set is* V.

Linear dependence

Definition

Vectors $\mathbf{v}_1, \ldots, \mathbf{v}_N \in V$ *are* **linearly dependent** *if*

$$\sum_{i=1}^{N} a_i \mathbf{v}_i = 0 \tag{34}$$

has a non-trivial (not $\mathbf{a} = (0)$*) solution.*

Null-space

Definition

Null-space *of transformation* $T : V \to W$ *is the set of all vectors* $\mathbf{v} \in V$*, which are mapped to* $\mathbf{0} \in W$*:*

$$N(T) = \{\mathbf{v} | T(\mathbf{v}) = \mathbf{0}\} \tag{35}$$

Trivial null-space is $\{\mathbf{0}\}$.

Definition

Nullity *of a transformation* T *is the dimensionality of its null-space:*

$$nullity(T) = dim(N(T)) \tag{36}$$

Rank

Definition

Range *of a transformation* $T : V \rightarrow W$ *is the set of all possible vectors in* W*, formed by the transformation:*

$$R(T) = \{\mathbf{w} | \mathbf{w} = T(\mathbf{v}), \mathbf{v} \in V\} \tag{37}$$

Definition

Rank *of a transformation* T *is the dimensionality of its range:*

$$rank(T) = dim(R(T)) \tag{38}$$

For any linear transformation T:

$$\forall T : V \rightarrow W : nullity(T) + rank(T) = dim(V) \tag{39}$$

Ordered basis

Definition

Ordered basis *is a tuple* $(\mathbf{b}_1, \ldots, \mathbf{b}_N)$ *of basis vectors* $\{\mathbf{b}_1, \ldots, \mathbf{b}_N\}$*, where the basis vectors are ordered (numbered).*

Every vector $\mathbf{v} \in V$ may be uniquely represented in an ordered basis:

$$\mathbf{v} = a_1 \mathbf{v}_1 + \ldots + a_N \mathbf{v}_N \tag{40}$$

and thus defined by the $\mathbf{a} = [a_1, \ldots, a_N]$ in the ordered basis β:

$$[\mathbf{v}]^\beta = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \tag{41}$$

Ordered basis

If we define some ordered basis vectors β for V and γ for W, then the vectors of the transformation $T : V \rightarrow W$ may be written as:

$$[\mathbf{v}]^{\beta} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \qquad [T(\mathbf{v})]^{\gamma} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} \tag{42}$$

Matrix as a representation of linear transformations

Linear transformations in basis vectors

Writing the transformation T explicitly in basis vectors $\beta = (\mathbf{v}_1 \dots \mathbf{v}_N), \mathbf{v}_i \in V$ and $\gamma = (\mathbf{w}_1 \dots \mathbf{w}_N), \mathbf{w}_i \in W$ due to the properties of linear transformations:

$$T(\mathbf{v}) = T(x_1\mathbf{v}_1) + \dots + T(x_N\mathbf{v}_N) \tag{43}$$

$$= x_1 T(\mathbf{v}_1) + \dots + x_N T(\mathbf{v}_N) \tag{44}$$

$$= y_1\mathbf{w}_1 + \dots + y_N\mathbf{w}_N \tag{45}$$

Transformed basis vectors of V represented on basis vectors in W:

$$T(\mathbf{v}_1) = a_{1,1}\mathbf{w}_1 + \dots + a_{M,1}\mathbf{w}_M \tag{46}$$

$$\vdots \tag{47}$$

$$T(\mathbf{v}_N) = a_{1,N}\mathbf{w}_1 + \dots + a_{M,N}\mathbf{w}_M \tag{48}$$

$a_{i,j}$ are scalars, which depend on T, β, γ.

Linear transformations in basis vectors

Substitute (46)-(48) into (43):

$$T(\mathbf{v}) = x_1(a_{1,1}\mathbf{w}_1 + \ldots + a_{M,1}\mathbf{w}_M) + \tag{49}$$

$$\vdots \tag{50}$$

$$x_N(a_{1,N}\mathbf{w}_1 + \ldots + a_{M,N}\mathbf{w}_M) \tag{51}$$

Switching back to the representation in the ordered basis γ, vector $\mathbf{y}$ may be expressed as:

$$y_1 = a_{1,1}x_1 + \ldots + a_{1,N}x_N \tag{52}$$

$$\vdots \tag{53}$$

$$y_M = a_{M,1}x_1 + \ldots + a_{M,N}x_N \tag{54}$$

(Here we may notice similarities with systems of linear equations.)

Linear transformations in basis vectors. Matrix

Let's define the set of $a_{i,j}$ from (52) as a **matrix** A:

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,M} \\ \vdots & \ddots & \vdots \\ a_{M,1} & \cdots & a_{M,N} \end{bmatrix} \tag{55}$$

If we define the operation of multiplication of the matrix A by the vector $\mathbf{x}$ as in (52) ... (54), then the transformation T in the bases β and γ may be written as:

$$\mathbf{y} = A\mathbf{x} \tag{56}$$

Thus, any linear transformation may be expressed as a matrix-by-vector multiplication in some ordered bases.

Matrix operations

Based on the (52) - (56) we can define some matrix operations and derive the methods (rules) using which these operations can be performed.

Sum of two matrices of transformations $T_A, T_B : V \to W$ is defined as a matrix of a transformation $T_S : V \to W$, which results in summation in the vector space W

$$S\mathbf{v} = (A + B)\mathbf{v} := A\mathbf{v} + B\mathbf{v} \tag{57}$$

$$S = A + B = \begin{bmatrix} a_{1,1} + b_{1,1} & \cdots & a_{1,M} + b_{1,M} \\ \vdots & \ddots & \vdots \\ a_{M,1} + b_{M,1} & \cdots & a_{M,N} + b_{M,N} \end{bmatrix} \tag{58}$$

$$s_{i,j} = a_{i,j} + b_{i,j} \tag{59}$$

Matrix operations

Scalar multiplication scales the resulting vector in W:

$$(\alpha A)\mathbf{v} := \alpha(A\mathbf{v}) \tag{60}$$

$$S = \alpha A = \begin{bmatrix} \alpha a_{1,1} & \cdots & \alpha a_{1,M} \\ \vdots & \ddots & \vdots \\ \alpha a_{M,1} & \cdots & \alpha a_{M,N} \end{bmatrix} \tag{61}$$

$$s_{i,j} = \alpha a_{i,j} \tag{62}$$

Matrix operations

Matrix multiplication results in sequential transformations (composition). If $T_A : V \to W$, $T_B : W \to Q$, then:

$$T_S = T_B \circ T_A : V \to Q \tag{63}$$

$$S\mathbf{v} = (BA)\mathbf{v} := B(A\mathbf{v}) \tag{64}$$

$$S = BA = \begin{bmatrix} s_{1,1} & \cdots & s_{1,M} \\ \vdots & \ddots & \vdots \\ s_{M,1} & \cdots & s_{M,N} \end{bmatrix} \tag{65}$$

$$s_{i,j} = \sum_{k=1}^{N} b_{i,k} a_{k,j} \tag{66}$$

Here "$\circ$" is the composition operator. Dimensions of the matrices must be compatible: the number of columns in B must be equal to the number of rows in A.

Matrix multiplication is also associative, but not commutative

$$(T_C \circ (T_B \circ T_A))\mathbf{v} = ((T_C \circ T_B) \circ T_A)\mathbf{v} \tag{67}$$

$$T_C \circ (T_B \circ T_A) = (T_C \circ T_B) \circ T_A \tag{68}$$

$$C(BA) = (CB)A = CBA \tag{69}$$

$$BA \neq AB \tag{70}$$

Identity matrix

Special matrix, which doesn't change the transformation, is called identity matrix I. It is a square matrix with ones on the main diagonal and zeros everywhere else:

$$IA = A \tag{71}$$

$$AI = A \tag{72}$$

$$I = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} \tag{73}$$

$$\tag{74}$$

Inverse transformation

For some transformations $T : V \to V$ we may define a transformation T^{-1}, which is the inverse of T:

$$T^{-1} : \forall \mathbf{v} \in V : T^{-1}T\mathbf{v} = \mathbf{v} \quad (75)$$

$$T^{-1}T = I \quad (76)$$

$$(77)$$

We can also derive that:

$$T^{-1}TT^{-1} = IT^{-1} = T^{-1}I \Rightarrow \quad (78)$$

$$TT^{-1} = I = T^{-1}T \quad (79)$$

Matrix inverse

From this, the inverse of the matrix product is:

$$
\begin{aligned}
BA &= C && (80)\\
(BA)^{-1} &= C^{-1} && (81)\\
BA(BA)^{-1} &= BAC^{-1} && (82)\\
I &= BAC^{-1} && (83)\\
B^{-1} &= AC^{-1} && (84)\\
A^{-1}B^{-1} &= C^{-1} && (85)\\
\Rightarrow & && (86)\\
(BA)^{-1} &= A^{-1}B^{-1} && (87)
\end{aligned}
$$

Matrix inverse

Not every square matrix has in inverse. Matrices which have an inverse are called **invertible**. For an invertible $N \times N$ matrix A the following equivalent conditions hold:

- linear transformation T_A is one-to-one
- the rank of A is N
- the null space of A is 0
- the columns/rows of A are linearly independent
- the columns/rows of A span $\mathbb{R}^N$
- the columns/rows of A from a basis in $\mathbb{R}^N$

Elementary row operation and matrix inverse

Here we will define some special transformations (operations), which may be used to find the matrix inverse. All these operations are invertible (full-rank), thus do not reduce the rank of the matrix being applied to. For the sake of simpliccity, assume we have a 3×3 matrix (extending it to $N \times N$ must be trivial)

Row multiplication

To multiply the i-th row of some matrix A by $\alpha \neq 0$, we may take the identity matrix and set the i-th element at the diagonal to α. Lets construct such transformation for the 3×3 matrix, which multiplies the 2-nd row by α:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}, \qquad E_{mult} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{88}$$

$$E_{mult} A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ \alpha a_{2,1} & \alpha a_{2,2} & \alpha a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix} \tag{89}$$

Row exchange

To exchange the i-th and j-th rows of some matrix A, we start from the identity matrix and swap its i-th and j-th columns. For a 3×3 matrix, which exchanges the 2-nd and the 3-rd rows:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}, \qquad E_{exch} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \tag{90}$$

$$E_{exch}A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix} \tag{91}$$

Row addition

To add the i-th row multiplied by α to the j-th, we again start with the identity matrix set the (i, j)-th element to α. For a 3×3 matrix, which adds $\alpha \times$ 2-nd row to the the 3-rd:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}, \qquad E_{add} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \alpha & 1 \end{bmatrix}, \tag{92}$$

$$E_{add}A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ \alpha a_{2,1} + a_{3,1} & \alpha a_{2,2} + a_{3,2} & \alpha a_{2,3} + a_{3,3} \end{bmatrix} \tag{93}$$

Inverting a matrix with elementary operations

If some $N \times N$ matrix A can be transformed to the identity matrix by a chain of elementary matrix operations:

$$E_m \dots E_2 E_1 A = I \tag{94}$$

then A is invertible, and the inverse is:

$$E_m \dots E_2 E_1 I = A^{-1} \tag{95}$$

These elementary matrix operations were used in the Gaussian elimination algorithm!

Matrix determinant

Matrix determinant

Lets take a 2×2 matrix, and try to invert it with the elementary row operations.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & b/a \\ c & d \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & b/a \\ 0 & d - c(b/a) \end{bmatrix} \tag{96}$$

In order the matrix A to be invertible, $d - c(b/a)$ must be not 0:

$$d - c(b/a) \neq 0 \Rightarrow ad - bc \neq 0 \tag{97}$$

$$det(A) := ad - bc \tag{98}$$

We can assign a special number $det(A)$ - determinant - for every matrix, which would indicate, that the matrix is invertible if this number is non-zero ($det(A) = 0 \Leftrightarrow rank(A) < N$).

3x3 matrix determinant

In (98) we derived the formula of determinant for 2×2 matrices. We can follow the same steps and find the determinant formula for a 3×3 matrix

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}, \qquad det(A) := aei - ceg + dhc - bdi + bfg - fha \tag{99}$$

NxN matrix determinant

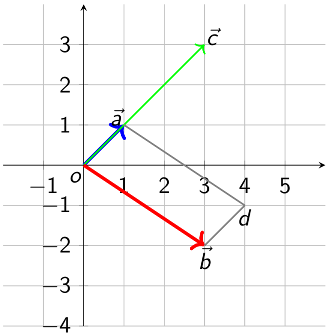
Derivation of the generic formula for the determinant is basically out of the scope of linear algebra, it involves some theory of permutations and functions, here we will just present the final formula:

$$A \in \mathbb{R}^{N \times N} \tag{100}$$

$$det(A) := \sum_{p \in perm(N)} (-1)^{\#transp(p)} \prod_{i=1}^{N} a_{i,p_i} \tag{101}$$

where $perm(N)$ is the set of all possible permutations of size n, $\#transp(p)$ is the number of transpositions (exchanges of 2 elements) in the permutation.

Geometric interpretation of determinant



Determinant is a signed measure of space occupied by a hyper-parallelogram formed by vector-columns of the matrix.
0 hyper-volume $\Leftrightarrow$ $det(A) = 0$ $\Leftrightarrow$ A is not full rank $\Leftrightarrow$ A is not invertible $\Leftrightarrow$ A columns (rows) are linearly dependant...

Diagonal matrix

Diagonal matrix

There is a special type of matrices, which are very easy to work with. These are diagonal matrices, which are comprised of zeros everywhere but the main diagonal:

$$A = diag(a_1, \dots, a_N) = \begin{bmatrix} a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_N \end{bmatrix} \tag{102}$$

Rank of a diagonal matrix is the number of non-zero elements on the main diagonal.

$$rank(diag(a_1, \dots, a_N)) = N - \sum_{i=1}^{N} \mathbb{1}_0(a_i) \tag{103}$$

Here $\mathbb{1}_{(\cdot)}(\cdot)$ is the indicator function.

Diagonal matrix

Determinant, inverse, product and exponential of such matrices are easy to calculate:

$$\det A = \det(diag(a_1, \ldots, a_N)) = \prod_{i=1}^{N} a_i \tag{104}$$

$$A^{-1} = diag(1/a_1, \ldots, 1/a_N) = \begin{bmatrix} 1/a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1/a_N \end{bmatrix} \tag{105}$$

$$A^p = diag(a_1^p, \ldots, a_N^p) = \begin{bmatrix} a_1^p & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_N^p \end{bmatrix} \tag{106}$$

$$e^A = diag(e^{a_1}, \ldots, e^{a_N}) = \begin{bmatrix} e^{a_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & e^{a_N} \end{bmatrix} \tag{107}$$

Basis change

Combining linear transformations in matrix form

Theorem: for bases α, β, γ in vector spaces V, W, Q and transformations $T : V \to W$, $S : W \to Q$:

$$[ST]_\alpha^\gamma = [S]_\beta^\gamma [T]_\alpha^\beta \tag{108}$$

Proof: apply the transformations to an arbitrary vector $\mathbf{v}$ *inV* and see that the results are equal

Invertible transformation

Theorem: $T : V \to W$, α -oredered basis in V, β -oredered basis in W, then $T : V \to W$ is invertible iff the matrix $[T]_\alpha^\beta$ is invertible:

$$([T]_\alpha^\beta)^{-1} = \left[T^{-1}\right]_\alpha^\beta \tag{109}$$

Corollary: matrix $[T]_\alpha^\beta$ is invertible iff transformation T is invertible.
Corollary: only square matrices are invertible.

Automorphism

Definition

Automorphism is a mapping from space V onto this very space V:

$$T : V \to V \tag{110}$$

Basis change

Basis change lemma $T : V \to V$ - automorphism, β, β' - bases in V

$$Q := [I_V]_{\beta}^{\beta'} \text{–change of basis matrix} \tag{111}$$
$$[T]_{\beta'}^{\beta'} = Q\,[T]_{\beta}^{\beta}\,Q^{-1} \tag{112}$$

Proof:

$$T = I_V T I_V \tag{113}$$

in bases β and β':

$$[T]_{\beta'}^{\beta'} = [I_V]_{\beta}^{\beta'}\,[T]_{\beta}^{\beta}\,[I_V]_{\beta'}^{\beta} \tag{114}$$
$$[T]_{\beta'}^{\beta'} = Q\,[T]_{\beta}^{\beta}\,Q^{-1} \tag{115}$$

Eigenvectors and eigenvalues

Eigenvectors and eigenvalues

Sometimes for some vectors $\mathbf{v} \in V$ linear transformation $T : V \to V$ acts as a simple scaling operator: $T\mathbf{v} = \lambda\mathbf{v}$. Such vectors $\{\mathbf{v}_i\}$ are called **eigenvectors**, and corresponding scaling factors $\{\lambda_i\}$ - **eigenvalues**.

$$T\mathbf{v} = \lambda\mathbf{v} \tag{116}$$

$$\lambda \in \mathbb{C} \tag{117}$$

Finding eigenvectors and eigenvalues

$$T\mathbf{v} = \lambda\mathbf{v} \tag{118}$$

$$T\mathbf{v} - \lambda\mathbf{v} = \mathbf{0} \tag{119}$$

$$(T - \lambda I)\mathbf{v} = \mathbf{0} \tag{120}$$

We have to find such $\{\lambda_i\}$ that $(T - \lambda I)$ has non-trivial (not 0) null space $\Rightarrow \det(T - \lambda I) = 0$.

$\det(T - \lambda I) = 0$ is the characteristic equation (polynomial). Having $\{\lambda_i\}$, it is easy to find the corresponding eigenvectors y solving the $(T - \lambda_i I)\mathbf{v}_i = \mathbf{0}$ system of equations w.r.t $\mathbf{v}_i$.

Eigenspace

Definition

Eigenspace *for* $\{\lambda_i\}$*:*

$$\{\mathbf{v} | T\mathbf{v} = \lambda_i \mathbf{v}\} \tag{121}$$

Eigendecomposition of a matrix

TITLE

Definition

$T : V \to V$ is diagonalizable if $\exists$ ordered basis β' : $[T]_{\beta'}^{\beta'}$ is diagonal.

Definition

Matrix A is diagonalizable if its corresponding linear transformation is diagonalizable.

Eigenvector decomposition as diagonalization

Lemma: matrix A is diagonalizable iff it can be expressed as

$$A = QDQ^{-1} \tag{122}$$

Proof:
1) If A is diagonalizable, then by changing the basis:

$$A = [T_A]_\beta^\beta = Q\,[T_A]_{\beta'}^{\beta'}\,Q^{-1} = QDQ^{-1} \tag{123}$$

2) If $A = QDQ^{-1}$: lets take $D = diag(\lambda_1 \dots \lambda_N)$, $D\mathbf{e}_j = \lambda_j \mathbf{e}_j$ (here $\mathbf{e}_j$ are eigenvectors of D):

$$A(Q\mathbf{e}_j) = QDQ^{-1}Q\mathbf{e}_j = QD\mathbf{e}_j = Q\lambda_j\mathbf{e}_j = \lambda_j Q\mathbf{e}_j \tag{124}$$

Thus, $Q\mathbf{e}_j$ is an eigenvector of A.
Q is invertible, $\{\mathbf{e}_j\}$ is a basis $\Rightarrow \{Q\mathbf{e}_j\}$ is a basis also. This means, that A is diagonalizable.

Eigenvector decomposition

Corollary: What is $\mathbf{e}_j$? It is a vector of zeroes with 1 at the j-th position: $[0, \ldots, 0, 1, 0, \ldots, 0]$.
$Q\mathbf{e}_j$ - the j-th column of Q.
$A = QDQ^{-1}$ is the eigenvector decomposition of A
Matrix A is *diagonalizable* iff eigenvectors span the whole space V.

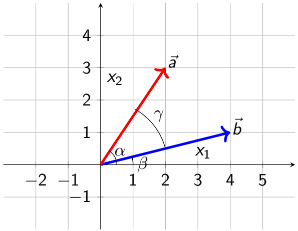
Dot product

Dot product

Dot product of two vectors $\mathbf{a}, \mathbf{b} \in V$ results in a scalar. It is possible to introduce the idea of *dot product* in several ways. One of them is that dot product can be viewed as a projection of one vector on another, scaled by the vectors lengths. Or a "measure" of non-orthogonality of two vectors. Another motivation is a "measure" of vectors similarity. It may be negative though, thus it does not satisfy the requirements of measure definition.
Here we will derive the dot product as a scaled cosine of the angle formed by the vectors.

Dot product

Let $\mathbf{a}, \mathbf{b} \in \mathbb{R}^2$, $\mathbf{e}_x, \mathbf{e}_x$ - orthonormal basis. The cosine of the angle formed by $\mathbf{a}, \mathbf{b}$:



$$cos(\gamma) = cos(\beta - \alpha) = cos(\beta)cos(\alpha) + sin(\beta)sin(\alpha) \tag{125}$$

$$= \frac{b_x}{||\mathbf{b}||}\frac{a_x}{||\mathbf{a}||} + \frac{b_y}{||\mathbf{b}||}\frac{a_y}{||\mathbf{a}||} = \frac{a_x b_x + a_y b_y}{||\mathbf{a}||||\mathbf{b}||} \tag{126}$$

Dot product

$$cos(\gamma) = \frac{a_x b_x + a_y b_y}{||\mathbf{a}||||\mathbf{b}||} \tag{127}$$

From this we can define the dot product:

$$\mathbf{a} \cdot \mathbf{b} := a_x b_x + a_y b_y = cos(\gamma)||\mathbf{a}||||\mathbf{b}|| \tag{128}$$

It is clear, that the dot product is equal to 0 iff the vectors are orthogonal.

Dot product

Dot product is linear w.r.t. scalar multiplication of vectors and commutative:

$$(\alpha \mathbf{a}) \cdot \mathbf{b} = \alpha(\mathbf{a} \cdot \mathbf{b}) \tag{129}$$

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a} \tag{130}$$

$$\tag{131}$$

General formula for dot product of two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^N$ in an orthonormal basis reads:

$$\mathbf{a} \cdot \mathbf{b} = \sum_{i=1}^{N} a_i b_i \tag{132}$$

Dot product

One may notice that matrix-vector product may be expressed as a vector comprised of dot products of vector-rows of the matrix with the vector. This may be extended to the matrix-matrix multiplication, where the resulting matrix is comprised of the corresponding vector-row by vector-column dot products.

Vector length

Vector length appears to be equal to square root of vector dot product with itself:

$$||\mathbf{a}|| = \sqrt{\sum_{i=1}^{N} a_i b_i} = \sqrt{\mathbf{a} \cdot \mathbf{a}} \tag{133}$$

Inner product space

Inner product is a more general concept than the dot product. Dot product is the representation of inner product in Euclidean space. Here we will just give the definition for the sake completeness of this short Linear Algebra overview:

Definition

Inner product space *is a vector space augmented with additional operation - inner product* $\langle \cdot, \cdot \rangle$ *- with the following properties:*

- *distributivity:* $\langle \mathbf{v} + \mathbf{v}', \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{v}', \mathbf{w} \rangle$
- *conjugate symmetry:* $\langle \mathbf{v}, \mathbf{w} \rangle = \overline{\langle \mathbf{v}, \mathbf{w} \rangle}$
- *linearity on the first argument:* $\langle \alpha \mathbf{v}, \mathbf{w} \rangle = \alpha \langle \mathbf{v}, \mathbf{w} \rangle, \alpha \in C$
- *positivity: if* $\mathbf{v}$ *is non-zero, then* $\langle \mathbf{v}, \mathbf{v} \rangle > 0$

Systems of linear differential equations. Matrix exponential

Linear differential equation

Lets recall the solution of a simple linear differential equation:

$$x'(t) = ax(t) \tag{134}$$

$$x(0) = x_0 \tag{135}$$

$$\frac{dx(t)}{dt} = ax(t) \tag{136}$$

$$\frac{dx(t)}{x(t)} = adt \tag{137}$$

$$\int_{x(0)}^{x(T)} \frac{dx(t)}{x(t)} = \int_0^T adt \tag{138}$$

$$\log(x(t))|_{x(0)}^{x(T)} = at|_0^T \tag{139}$$

$$\log(\frac{x(T)}{x(0)}) = aT \tag{140}$$

$$x(T) = e^{aT}x_0 \tag{141}$$

System of linear differential equations

Now consider a system of linear differential equations, where the state $\mathbf{x}(t)$ is a vector, A is an evolution matrix:

$$\mathbf{x}'(t) = A\mathbf{x}(t) \tag{142}$$

$$\mathbf{x}(0) = \mathbf{x}_0 \tag{143}$$

In general, the solution of such systems is the matrix exponential:

$$\mathbf{x}(t) = e^{At}\mathbf{x}_0 \tag{144}$$

Taylor series expansion

In Taylor series simple exponential function writes:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots = \sum_{n=0}^{\infty} \frac{x^n}{n!} \tag{145}$$

Correspondingly, matrix exponential is defined via its Taylor series as:

$$e^A = I + \frac{A}{1!} + \frac{A^2}{2!} + \ldots = \sum_{n=0}^{\infty} \frac{A^n}{n!} \tag{146}$$

TITLE

We may make an important observation from this definition. If matrix B is invertible, then:

$$e^{BAB^{-1}} = I + \frac{BAB^{-1}}{1!} + \frac{(BAB^{-1})^2}{2!} + \dots \tag{147}$$

$$= I + \frac{BAB^{-1}}{1!} + \frac{BA^2B^{-1}}{2!} + \dots \tag{148}$$

$$= B(I + \frac{A}{1!} + \frac{A^2}{2!} + \dots)B^{-1} \tag{149}$$

$$= Be^A B^{-1} \tag{150}$$

If a matrix is diagonal, the exponential boils down to a simple matrix of exponentials:

$$e^{diag(a_1,\dots,a_n)t} = diag(e^{a_1 t}, \dots, e^{a_n t}) \tag{151}$$

Matrix exponent in eigenbasis

Thus, we may use the previous observation to move the exponent inside of the product to the diagonal matrix:

$$
\begin{aligned}
exp(At) &= exp(QDQ^{-1}t) && (152)\\
&= Qe^{Dt}Q^{-1} && (153)\\
&= Qdiag(e^{\lambda_1 t}, \dots, e^{\lambda_n t})Q^{-1} && (154)
\end{aligned}
$$

Using this result, the solution of the original system of linear differential equations may be written as:

$$\mathbf{x}(t) = Qdiag(e^{\lambda_1 t}, \dots, e^{\lambda_n t})Q^{-1}\mathbf{x}_0 \tag{155}$$

where Q is the matrix of eigenvectors, $\lambda_1, \dots \lambda_n$ are the corresponding eigenvalues.

Matrix exponent in eigenbasis

$$\mathbf{x}(t) = Qdiag(e^{\lambda_1 t}, \ldots, e^{\lambda_n t})Q^{-1}\mathbf{x}_0 \tag{156}$$

Behaviour of such system is completely determined by its eigenvalues. For positive eigenvalues it is divergent, negative eigenvalues indicate that it converges to the stationary solution, with complex eigenvalues we get an oscillating system.

Random walk and graph eigenvalues

Random walk problem

Problem: *Wanderland is a simple country with towns and one-way roads. The people of Wanderland like to travel, but they don't really care where to go and make a random choice of the road to take. Unfortunately all hotels are run by a monopoly company "Boogle". Estimate the number of hotel rooms in each town to maximize the profit of the monopoly.*
This is an example of a random walk on a graph problem, where we are interested in a stationary distribution - the distribution, to which the system relaxes running infinite amount of time.

Let $\mathbf{x}_t$ be the current distribution of the travellers, $\mathbf{x}_t[i]$ is the proportion of traveller in the i-th town. Next, given the map of the country, we may construct a *transition matrix* A, where $A_{i,j}$ is the proportion of the direct roads from a town i to a town j to the total number of outgoing roads from the town j. Note that the sum of every column is 1. From this we get the next state as a simple matrix multiplication:

$$\mathbf{x}_{t+1} = A\mathbf{x}_t = A^{t+1}\mathbf{x}_0 \tag{157}$$

Stationary distribution condition

For the stationary distribution the following should hold:

$$\mathbf{x}^* = A\mathbf{x}^* \tag{158}$$

$$\begin{aligned}
\mathbf{x}^* &= \lim_{n\to\infty} A^n \mathbf{x}_0 && (159)\\
&= \lim_{n\to\infty} (QDQ^{-1})^n \mathbf{x}_0 && (160)\\
&= \lim_{n\to\infty} QD^nQ^{-1}\mathbf{x}_0 && (161)\\
&= Q\left(\lim_{n\to\infty} D^n\right) Q^{-1}\mathbf{x}_0 && (162)\\
&= Q diag(1, 0, \dots, 0) Q^{-1}\mathbf{x}_0 && (163)\\
&= \mathbf{v}_1 / sum(\mathbf{v}_1) && (164)
\end{aligned}$$

Thus, $\mathbf{x}^*$ is an eigenvector of the transition matrix A.

Positive-(semi)definite matrix

Positive-(semi)definite matrix

Definition

A square symmetric matrix Σ *is* **positive-definite** *iff*

$$\forall \mathbf{x} : \mathbf{x}^T \Sigma \mathbf{x} > 0 \tag{165}$$

A square symmetric matrix Σ *is* **positive-semidefinite** *iff*

$$\forall \mathbf{x} : \mathbf{x}^T \Sigma \mathbf{x} \geq 0 \tag{166}$$

Corollary: all eigenvalues of Σ must be positive:

$$\begin{aligned} \mathbf{x}^T \Sigma \mathbf{x} &= \mathbf{x}^T Q^{-1} D Q \mathbf{x} && (167) \\ &= \mathbf{x}^T Q^T D Q \mathbf{x} && (168) \\ &= (Q\mathbf{x})^T D (Q\mathbf{x}) > 0 && (169) \end{aligned}$$

where $D = diag(\lambda_1 \dots \lambda_N)$ - matrix of eigenvalues,
$\Sigma = Q^{-1} D Q$ - eigendecomposition of Σ.