A Glossary for Elementary Linear Algebra

NOTE: TEXT THAT IS WRAPPED IN /* ... */ REFERS TO FACTS THAT YOU WILL NOT UNDERSTAND UNTIL AFTER YOU ENCOUNTER THE LISTED CONCEPT THE FIRST TIME. I WANT YOU TO KEEP THE GLOSSARY AND SO I MAKE IT USEFUL FOR LATER BACK REFERENCE AND REPETITION. I DON'T USE /* ... */ FOR MATERIAL THAT IS ONLY A WEEK AHEAD.
EXAMPLE: /* some wisdom about eigenvalues */ end of M251

Linear Equation: An equation of the form $a_1x_1 + a_2x_2 + \ldots + a_nx_n = b$ where the a_j and b stand for given quantities (e.g., numbers) and the x_j are variables (unknowns). Occurrence of terms like x_i^2, x_1x_2, $\cos x_j$, $1/x_j$ would make the equation nonlinear.

Matrix: A rectangular array of numbers (or symbols standing for numbers). An $m \times n$ matrix is a matrix with m rows and n columns. We usually denote matrices with capital letters like A, and their entries with the corresponding lowercase letters: a_{ij} is the entry in the i^{th} row and j^{th} column of matrix A. Synonym: $a_{ij} = (A)_{ij}$. The fi**R**st index refers to the **R**ow, the seco**N**d index refers to the colum**N**.

Row and column vectors: A row vector is a $1 \times n$ matrix (a matrix with only one row), a column vector is a $m \times 1$ matrix (a matrix with only one column).

Square Matrix: A matrix with as many rows as columns. The **(main) diagonal** of a square matrix A (size $n \times n$) consists of the elements $a_{11}, a_{22}, \ldots, a_{nn}$; i.e., *those*: $\begin{bmatrix} \diagdown \end{bmatrix}$. The other 'diagonal' $\begin{bmatrix} \diagup \end{bmatrix}$ plays no role in matrix theory and doesn't deserve a name of its own. So diagonal always refers to main diagonal, unless specified otherwise explicitly.

Trace of a square matrix: The sum of its diagonal entries. The trace becomes more important in more advanced treatments of Lin'Alg'. In M251, it justly looks a bit boring. The following facts are relevant about the trace:
$\operatorname{tr}(A + B) = \operatorname{tr}A + \operatorname{tr}B$.
$\operatorname{tr}(cA) = c \operatorname{tr}A$.
$\operatorname{tr}(AB) = \operatorname{tr}(BA)$ (even though AB my not be equal to BA).
/* Important fact: The trace of a matrix equals the sum of the eigenvalues of this matrix (counting multiplicities, and accepting complex eigenvalues) */ end of M251.

Augmented matrix of a system of linear equations: For the system

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &=& b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &=& b_2 \\ \vdots \qquad\quad \vdots \qquad\qquad\quad \vdots \qquad\quad \vdots & & \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &=& b_m \end{array}$$

the augmented matrix is

$$\left[\begin{array}{ccccc} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{array}\right]$$

Why 'augmented'? Without the word 'augmented', the **coefficient matrix** of a system of linear equations would only denote the matrix (a_{ij}) containing only the coefficients from the left hand side. 'Augmented' refers to the fact that the b_i have been joined to the coefficient matrix.

Elementary row operations: These are the following:
(1) multiply a row by a non-zero number,
(2) add some multiple of one row to any other row,
(3) exchange rows.
Their main uses:
• Solving systems of linear equations by Gauss or Gauss Jordan elimination, applied to the augmented matrix of the system
• /* Manipulating square matrices to calculate their determinant */Ch. 2

Gauss(ian) elimination: achieves *row echelon form* of the augmented matrix of a system of linear equations by means of elementary row operations. The system so obtained can then be solved by back substitution.
See book how it works in detail; no point copying it here

Gauss-Jordan elimination: goes beyond Gauss elimination, and it achieves *reduced row echelon form* of the augmented matrix of a system of linear equations by means of elementary row operations. The system so obtained is is in a form where the solutions can be read off immediately.
See book how it works in detail; no point copying it here

Row echelon form: A matrix is in row echelon form, if
• The first nonzero entry in each row is 1 (if there is a nonzero entry in this row at all). This entry is called a **leading 1**.
• Rows further down have their leading 1's further right, and rows that are all 0's are at the bottom.

Example:
$$\begin{bmatrix} 0 & 1 & * & * & * & * & * & * & * \\ 0 & 0 & 1 & * & * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 1 & * & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

where $*$ may represent any number.

Reduced row echelon form: Row echelon form, and also satisfying:
• In a column containing a leading 1, all other entries are 0.

Example:
$$\begin{bmatrix} 0 & 1 & 0 & * & * & 0 & 0 & * & * \\ 0 & 0 & 1 & * & * & 0 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & * & * \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

where $*$ may represent any number.

Transpose of a matrix: If A is an $m \times n$ matrix then its transpose, denoted by A^T, is the $n \times m$ matrix arising from A by changing columns into rows and vice versa: $(A^T)_{ij} = (A)_{ji}$; example:

$$\text{If } A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 7 & 9 & 2 \\ -1 & 8 & \pi & 5 \end{bmatrix} \quad \text{then} \quad A^T = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 7 & 8 \\ 3 & 9 & \pi \\ 4 & 2 & 5 \end{bmatrix}$$

What is the transpose of a matrix good for? The deeper reason cannot become transparent until later: /* It has something to do with dual spaces */400 level linear algebra But you get good glimpses earlier: /* The gradient of a function $\mathbb{R}^n \to \mathbb{R}$ can be defined as the transpose of the Jacobi matrix of this function */M241. /* Transposing a column vector to a row vector and vice versa is a way to relate the standard scalar product between vectors to the matrix product */Ch. 4. /* Matrices that represent rotations have the property that their transpose equals their inverse. */Ch. 6

Symmetric matrices: A matrix A is called symmetric, if $A^T = A$. In particular, symmetric matrices are square. /* The tensor of inertia in physics is a symmetric matrix */theoretical mechanics course. /* Symmetric matrices have real eigenvalues */Ch. 7. /* In vector calculus, the Hessian of a (C^2) function (matrix of its second derivatives) is symmetric */M241.

Matrix algebra: Matrices of *equal* size can be added or subtracted by adding or subtracting corresponding entries. Matrices can be multiplied by a number by multiplying each of their entry with that number. In formulas: $(A+B)_{ij} = (A)_{ij} + (B)_{ij}$, $(A-B)_{ij} = (A)_{ij} - (B)_{ij}$, $(c \cdot A)_{ij} = c \cdot (A)_{ij}$. These operations obey the same rules as the corresponding operations on numbers, eg., $A + B = B + A$, $(A + B) + C = A + (B + C)$, $c(A + B) = cA + cB$. Addition and subtraction of matrices of different size is NOT defined.
/* Matrices of given size form a commutative group under addition */advanced algebra

The multiplication of two matrices is defined if and only if the first factor has as many columns as the second factor has rows. If A is an $m \times r$ matrix and B an $r \times n$ matrix, then AB is an $m \times n$ matrix, with the entries: $(AB)_{ij} = (A)_{i1}(B)_{1j} + (A)_{i2}(B)_{2j} + \ldots + (A)_{ir}(B)_{rj}$. The following rules hold, whenever the operations are defined:

$$(A+B)C = AC+BC\,, \quad C(A+B) = CA+CB\,, \quad (AB)C = A(BC)\,, \quad (AB)^T = B^T A^T$$

In general, it is NOT true that AB equals BA. If AB is defined, then BA may or may not be defined. If BA is also defined, it may or may not be equal to AB. In the (exceptional) case when AB happens to be equal to BA, we say that A and B commute. See the introductory notes for the geometrical interpretation of matrix multiplication.

Zero matrices: A zero matrix is one whose entries are all zero. It plays the same role with respect to matrix addition as the number 0 plays with respect to the addition of numbers. We denote it by 0, or if the size of the zero matrix is relevant, $0_{m \times n}$.

Identity matrices: An identity matrix (aka unit matrix) is a square matrix whose diagonal entries are all 1, and whose off-diagonal entries are all 0. We denote the $n \times n$ identity matrix by I_n. Ex:

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The identity matrix plays the same role for matrix multiplication as the number 1 plays in the multiplication of numbers: $AI_n = I_m A = A$ for any $m \times n$ matrix A.

Diagonal matrix: A square matrix that has non-zero entries only on the diagonal. Nothing subtle about the concept. — If the coefficient matrix of a system of linear equations is diagonal, the equations are uncoupled. /* The eigenvalue problem is equivalent to finding a digaonal matrix similar to a given matrix */Ch. 7.

Triangular matrix: A square matrix A is called upper triangular, if $a_{ij} = 0$ for $i > j$; it is called lower triangular, if $a_{ij} = 0$ for $i < j$. Triangular refers to either upper or lower triangular. Ex:

An upper triangular 3×3 matrix:
$$\begin{bmatrix} * & * & * \\ 0 & * & * \\ 0 & 0 & * \end{bmatrix}$$

A lower triangular 4×4 matrix:
$$\begin{bmatrix} * & 0 & 0 & 0 \\ * & * & 0 & 0 \\ * & * & * & 0 \\ * & * & * & * \end{bmatrix}$$

The $*$ may represent any numbers (zero or non-zero). How do you call a matrix that is upper triangular and lower triangular at the same time?

Linear homogeneous equation: If in a linear equation $a_1x_1+a_2x_2+\ldots+a_nx_n = b$, the given b is 0, the equation is called homogeneous. A system of homogeneous linear equation always has $x_1 = \ldots = x_n = 0$ as a solution (there may be other solutions as well). The solution $x_1 = \ldots = 0$ is called the trivial solution.

The word "homogeneous" comes from the Greek and means terms of the same kind. Here, the same kind is "coefficient times variable". A nonvanishing term on the right hand side would be of a different kind, since it doesn't contain a variable. (This reasoning doesn't apply to a zero on the rhs, because $0 = 0 \cdot x$, if you like.) The word "trivial" is of Latin origin. *trivium* referred to a fork in the road. So, "trivial" originally means "so common you can pick it up at every street corner". The zero solution is called trivial because you don't need to work to find it; you just pick it up for free.

/* In ODEs, a homogeneous linear ODE is also a linear ODE where the term that doesn't get multiplied with the unknown function or a derivative thereof vanishes; the same usage of the word applies in this context as well. */M231

Inverse matrix: An inverse of a square matrix A is a matrix B satisfying $AB = BA = I$. Basic facts: Not every square matrix has an inverse, but if it has an inverse, the inverse is unique (i.e., there is only one inverse), and it is then denoted by A^{-1}. Actually, only one of the conditions $AB = I$ and $BA = I$ is sufficient to ascertain that B is A^{-1}; you don't need both. /* A square matrix A has an inverse

if and only if its determinant is not 0. Except for 2×2, and maybe 3×3 matrices, this result is not very useful for practical purposes. */Ch. 2

If A and B are invertible, then so is AB. It holds $(AB)^{-1} = B^{-1}A^{-1}$. Note that taking the inverse requires reversing the order of the factors! – If A is invertible, then so is A^T, and $(A^T)^{-1} = (A^{-1})^T$. The shorthand A^{-T} is occasionally used for $(A^T)^{-1} = (A^{-1})^T$, but if you use it, make sure it doesn't seduce into wrongly thinking the ^{T} denoted a power.

invertible; singular: A square matrix is called invertible if it has an inverse. Otherwise, it is called singular.

Calculating the inverse of a matrix: Generally, as follows: Given an $n \times n$ matrix A, to solve $AB = I$ for B means to solve simultaneously n systems $A\boldsymbol{b}_j = \boldsymbol{e}_j$, where $\boldsymbol{b}_j$ and $\boldsymbol{e}_j$ denote the j^{th} column of B and I respectively. Apply Gauss–Jordan elimination to the "multiply augmented" matrix $[A|I]$ to obtain reduced row echelon form. If this form is $[I|B]$, then B is the inverse of A. If instead the reduced row echelon form contains a row beginning with n zeros (or as soon as a row beginning with n zeros shows up in the algorithm), you can conclude that A is not invertible.

$$\left[\begin{array}{ccc|ccc} 2 & -1 & 3 & 1 & 0 & 0 \\ \frac{1}{2} & 1 & -3 & 0 & 1 & 0 \\ 1 & 2 & -5 & 0 & 0 & 1 \end{array}\right] \xrightarrow{\text{Gauss–Jordan}} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{2}{5} & \frac{2}{5} & 0 \\ 0 & 1 & 0 & -\frac{1}{5} & -\frac{26}{5} & 3 \\ 0 & 0 & 1 & 0 & -2 & 1 \end{array}\right]$$

For a 2×2 matrix, the following formula is convenient (/* it's related to the determinant criterion */Ch. 2):

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

provided $ad - bc \neq 0$. If $ad - bc = 0$, the matrix is not invertible.

Elementary matrices: They are a theoretical device for the following purpose: Any elementary row operation on a matrix A corresponds to multiplying A from the left with an appropriate elementary matrix; namely (compare with the list of elementary row operations above):

To multiply the i^{th} row of an $m \times n$ matrix A with a nonzero number a is the same as calculating EA where ..... $E =$ and a is at position (i, i)

$$\begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & \ddots & & & \\ & & & a & & \\ & & & & 1 & \\ & & & & & \ddots \end{bmatrix}$$

To add a times the i^{th} row of an $m \times n$ matrix A to the j^{th} row is the same as calculating EA where ..... $E =$ and a is at position (j, i)

$$\begin{bmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & a & & \ddots & & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix}$$

To exchange the i^{th} and the j^{th} row of an $m \times n$ matrix A is the same as calculating EA where ..... $E =$ $\begin{bmatrix} 1 & & & & & \\ & 0 & & 1 & & \\ & & 1 & & & \\ & & & 1 & & \\ & 1 & & 0 & & \\ & & & & & 1 \end{bmatrix}$

and the off-diagonal ones are at positions (j,i) and (i,j)

This way, if you can reduce a square matrix A to an identity matrix by means of a sequence of elementary row operations, then A is the product of the elementary matrices corresponding to these row operations.

Elementary matrices are invertible, and the inverse of an elementary matrix is again an elementary matrix. (In other words: what an elementary row operation does, can also be undone by an elementary row operation.) Samples illustrating the inverse of elementary matrices are:

$$\begin{bmatrix} 1 & & \\ & 1 & \\ & & a \end{bmatrix}^{-1} = \begin{bmatrix} 1 & & \\ & 1 & \\ & & a^{-1} \end{bmatrix}, \qquad \begin{bmatrix} 1 & & \\ & 1 & \\ & a & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & & \\ & 1 & \\ & -a & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & & \\ & 0 & 1 \\ & 1 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & & \\ & 0 & 1 \\ & 1 & 0 \end{bmatrix}$$

Properties of systems of linear equations with a square coefficient matrix: (i.e., with as many equations as unknowns) — The following are equivalent for a square $(n \times n)$ matrix A and linear systems $A\boldsymbol{x} = \boldsymbol{b}$ with n-vectors $\boldsymbol{b}$:
(a) The coefficient matrix A is invertible
(b) The homogeneous equation $A\boldsymbol{x} = \boldsymbol{0}$ has *only* the trivial solution $\boldsymbol{x} = \boldsymbol{0}$
(c) The reduced row echelon form of A is the identity matrix I_n
(d) A can be written as a product of elementary matrices
(e) The inhomogeneous equation $A\boldsymbol{x} = \boldsymbol{b}$ has a solution for *every* $\boldsymbol{b}$
(f) For every $\boldsymbol{b}$, the inhomogeneous equation $A\boldsymbol{x} = \boldsymbol{b}$ has *exactly one* solution
(g) For some $\boldsymbol{b}$, the inhomogeneous equation $A\boldsymbol{x} = \boldsymbol{b}$ has *exactly one* solution.

Superposition principle: If $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ solve the homogeneous equation $A\boldsymbol{x} = \boldsymbol{0}$, then so does any $\boldsymbol{x} = c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2$ with numbers c_1 and c_2. In particular, multiples of solutions to a homogeneous linear equation are also solutions.
If you have *some* solution $\boldsymbol{x}_*$ of the inhomogeneous equation $A\boldsymbol{x} = \boldsymbol{b}$, and know (one or several) solutions $\boldsymbol{x}_1, \ldots$ of the homogeneous equation $A\boldsymbol{x} = \boldsymbol{0}$, you can obtain more solutions of the inhomogeneous equation in the form $\boldsymbol{x}_* + c_1\boldsymbol{x}_1 + \ldots$. In this context the one solution $\boldsymbol{x}_*$ you started with is called a *particular* solution. (There is nothing particular about a particular solution as such; what gives it the adjective 'particular' is that, for whatever reason, it's the one you encountered first.)
If you have a particular solution $\boldsymbol{x}_*$ of the inhomogeneous equation $A\boldsymbol{x} = \boldsymbol{b}$, and know *all* solutions $\boldsymbol{x}_h$ of the homogeneous equation $A\boldsymbol{x} = \boldsymbol{0}$, you can obtain *all* solutions of the inhomogeneous equation in the form $\boldsymbol{x}_* + \boldsymbol{x}_h$. The word "general solution" is often used to denote the set of all solutions of a linear equation (homogeneous or inhomogeneous). `/*` This principle is of particular relevance, when applied

to linear differential equations. Be aware that it is a priciple based on linearity alone, it has nothing to do with ODEs as such; it can be applied to *linear* ODEs, because they are linear, not because they are ODEs. */*M231

An immediate consequence of the superposition principle is: An linear equation that has more than one solution, automatically has infinitely many. /* That assumes that there are infinitely many scalars (which in this class are real (or if you prefer: complex) numbers; in advanced linear algebra you may have different types of scalars, where this assumption becomes relevant */M457 maybe, or M455-456

Linear combinations; scalars: 'Scalar' is another word for 'number'; it is used with the focus: 'number, not vector or matrix' in mind. A linear combination of a (finite) set of vectors $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots \boldsymbol{x}_j$, is an expression of the form $c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \ldots c_j\boldsymbol{x}_j$, with any choice of numbers (scalars) $c_1, c_2, \ldots c_j$. /* The words are also used in a more general context, where vectors are abstract vectors, like, e.g., functions. The concept of linear combination of functions (that solve a linear ODE) is the most important instance for you */Ch. 5; M231
/* Borderline definition: the only linear combination of an empty set of vectors is the null vector $\mathbf{0}$ */

Products and inverses of special types of matrices: (see their definitions above) The product of upper triangular matrices (of the same size of course) is upper triangular. The product of lower triangular matrices (of the same size of course) is lower triangular. The inverse of an upper (lower) triangular matrix, is upper (lower) triangular, provided it exists at all. An upper (lower) triangular matrix is invertible, if and only if its diagonal entries are non-zero. (Don't even think of applying such a diagonal test to other matrices! You're doomed if you do.) Corresponding statements apply to diagonal matrices, which are upper and lower triangular at the same time.

The inverse of a symmetric matrix is symmetric, provided the inverse exists. However, the product of symmetric matrices (of the same size of course) *need not be* symmetric. (It is symmetric if and only if the matrices commute.) To understand this fact, remember that $(AB)^T = B^T A^T$, *not* $A^T B^T$.

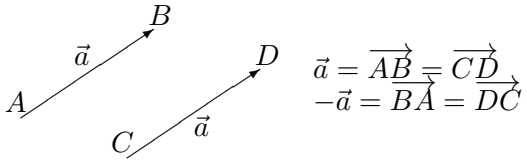
LU decomposition: If a square matrix A can be brought into row echelon form by means of elementary row operations *excluding* the interchange of rows, then it can be written as a product of a lower triangular matrix L and an upper triangular matrix U. This is called LU decomposition. In the LU decomposition of matrix A, the matrix U is what you get from A by means of Gauss elimination (without row exchanges), whereas the entries of the matrix L encode the row operations you used during the Gauss elimination.

The method (or a slightly more elaborate version that does allow for the exchange of rows) is the favorite general method for solving systems of linear equations with a square coefficient matrix on a computer. To solve a system $LU\boldsymbol{x} = \boldsymbol{b}$, given L, U, and $\boldsymbol{b}$, you first solve $L\boldsymbol{c} = \boldsymbol{b}$ for $\boldsymbol{c}$ (by simple forward substitution), then you solve $U\boldsymbol{x} = \boldsymbol{c}$, by backward substitution.

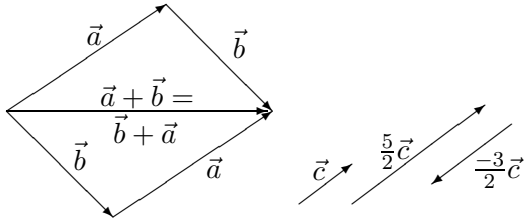
/* For large systems with a sparse coefficient matrix (i.e., one where most coefficients

are 0, more effective methodes are favored, because L and U may have all non-zero entries, even if A has many zero entries. — For instance if A is a 1000×1000 matrix of whose 1 million entries only 3000 are different from zero, you'd certainly prefer a method that only needs to calculate a few 1000 numbers, rather than the half million of entries in each of L and U. */numerical linear algebra course

Geometric vectors in the plane and in space: These can be viewed as directed line segments (arrows) between two points in the plane (or in space), but with the stipulation that translating such segments (without turning) results in the *same* vector. /* In a more formal and abstract way, one could define such a vector as an equivalence class of ordered pairs of points */e.g., M300. Notations for vectors are often boldface lowercase letters like $\boldsymbol{a}$, or lowercase letters with arrows on top, like $\vec{a}$. If points A, B are specified: $\overrightarrow{AB}$. The negative of a vector corresponds to the same segment, but with opposite direction. A vector $\overrightarrow{AA}$ is called zero vector, denoted $\vec{0}$ or $\mathbf{0}$.



Arithmetic of geometric vectors: Vectors can be added, and multiplied with (real) numbers. Numbers are called scalars when used in the context of vectors. Vectors are added as follows: To get $\vec{a} + \vec{b}$, move the tail of $\vec{b}$ at the tip of $\vec{a}$; then $\vec{a} + \vec{b}$ goes from the tail of $\vec{a}$ to the tip of $\vec{b}$, i.e., $\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$. For $k > 0$, k times a vector $\vec{a}$, denoted $k\vec{a}$ has the same direction as $\vec{a}$, but k times its length. For $k < 0$, $k\vec{a}$ has the opposite direction as $\vec{a}$, but $|k|$ times its length.



Geometric vectors expressed in (with respect to) coordinates: For points A, B in the plane, with coordinates $A(a_1, a_2)$ and $B(b_1, b_2)$, the geometric vector $\overrightarrow{AB}$ corresponds to the row vector (1×2 matrix) $[b_1 - a_1, b_2 - a_2]$. Then addition of geometric vectors corresponds to addition of their corresponding row vectors, and multiplication of geometric vectors with scalars (i.e., with numbers) corresponds to the multiplication of row vectors with scalars. The same can be done for geometric vectors in space; then the row vectors are 1×3-matrices.

It is likewise possible to get a correspondence with column vectors $\begin{bmatrix} b_1 - a_1 \\ b_2 - a_2 \end{bmatrix}$ instead of row vectors $[b_1 - a_1, b_2 - a_2]$, in the very same way. As long as vectors are not brought together with matrices, the choice between row and column vectors is arbitrary. For mere typographical reasons, row vectors are then preferred. However, as soon as we

will bring vectors together with matrices, a generally accepted convention requires that geometric vectors be represented by column vectors. The best way to abide by this convention and still harvest the typesetting convenience of rows is to write columns as the transpose of rows: $[a, b, c]^T$

Basic rules of vector arithmetic:

$\vec{a} + \vec{b} = \vec{b} + \vec{a}$	$(k + l)\vec{a} = k\vec{a} + l\vec{a}$
$(\vec{a} + \vec{b}) + \vec{c} = \vec{a} + (\vec{b} + \vec{c})$	$k(\vec{a} + \vec{b}) = k\vec{a} + k\vec{b}$
$\vec{a} + \vec{0} = \vec{a}$	$k(l\vec{a}) = (kl)\vec{a}$
$\vec{a} + (-\vec{a}) = \vec{0}$	$1\vec{a} = \vec{a}$

The list is not exhaustive; another rule is for instance $0\vec{a} = \vec{0}$. You won't have practical difficulties handling the rules of vector arithmetic. The list given here is sufficient in the sense that all other rules can be deduced (proved) as consequences of the given ones. `/*` These same rules will define the arithmetic with abstract vectors. Any set of objects for which an operation of addition and a multiplication with numbers is defined such that the above rules are valid, may be called a vector space (and the elements of that set will be called vectors). For instance, if you let the above arrowed symbols stand for functions, or $m \times n$ matrices, the same rules apply. The set of numbers (scalars) may then also be the set of complex, rather than real, numbers, even though this forfeits the geometric interpretation `*/`Ch. 5 — Note also the expected notation $\vec{a} - \vec{b} := \vec{a} + (-\vec{b})$.

Norm and dot product: For geometric vectors $\vec{a}$ and $\vec{b}$ in the plane or in space, the norm is defined as the length, and denoted by $\|\vec{a}\|$. In coordinates, if $\vec{a} = [a_1, a_2, a_3]^T$, then $\|\vec{a}\| := \sqrt{a_1^2 + a_2^2 + a_3^2}$ (for spatial vectors; in the plane, it's of course $\sqrt{a_1^2 + a_2^2}$). The dot product $\vec{a} \cdot \vec{b}$ is defined as the scalar $\|\vec{a}\| \, \|\vec{b}\| \cos\varphi$, where φ is the angle between the vectors $\vec{a}$ and $\vec{b}$. If one of the vectors is $\vec{0}$, the scalar product is defined to be 0, even though an angle φ between 0 and another vector is not meaningful. In coordinates, if $\vec{a} = [a_1, a_2, a_3]^T$ and $\vec{b} = [b_1, b_2, b_3]^T$, the dot product can be calculated as $\vec{a} \cdot \vec{b} = a_1b_1 + a_2b_2 + a_3b_3$ (for spatial vectors; only two terms occur for planar vectors). The dot product may also be called *scalar product*. The rules for the dot product are:

$\vec{a} \cdot \vec{b} = \vec{b} \cdot \vec{a}$
$\vec{a} \cdot (\vec{b} + \vec{c}) = \vec{a} \cdot \vec{b} + \vec{a} \cdot \vec{c}$
$\vec{a} \cdot (k\vec{b}) = k(\vec{a} \cdot \vec{b})$
$\vec{a} \cdot \vec{a} > 0$ unless $\vec{a} = \vec{0}$ (in which case it is 0)

The norm can be calculated from the dot product: $\|\vec{a}\| = \sqrt{\vec{a} \cdot \vec{a}}$. It is also possible to obtain the dot product in terms of the norm: $\vec{a} \cdot \vec{b} = \frac{1}{4}(\|\vec{a} + \vec{b}\|^2 - \|\vec{a} - \vec{b}\|^2)$.

Vectors, scalars, and the scalar product in physics: In Newton's times, equations for forces, velocities and accelerations would still have been written in terms of coordinate components, for lack of an appropriate vector notation. For instance the gravitation force by the sun (assumed to sit in the origin of the coordinate

system) on a planet is:

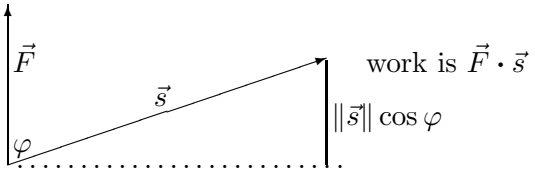
$$F_x = -GmMx/(x^2+y^2+z^2)^{3/2}$$
$$F_y = -GmMy/(x^2+y^2+z^2)^{3/2}$$
$$F_z = -GmMz/(x^2+y^2+z^2)^{3/2}$$

This is not only a lot to write, but it has certain disadvantages: The fact that the direction of the force goes towards the sun could be more succinctly visible than it is in these formulas. The coordinate system we are using has not been specified. It is a fiction of our mind for the purpose of doing calculation, but has no physical reality of its own. In spite of the fact that the position (represented by the triple of numbers (x, y, z)) and the force (represented by the triple of numbers (F_x, F_y, F_z)) are genuine physical quantities, each of the six coordinate quantities x, y, z, F_x, F_y, F_z has no physical reality, but depends on our artificial and arbitrary choice of a coordinate system. The geometric vectors $\vec{x}$ (whose coordinate representation happens to be $[x, y, z]^T$) and $\vec{F}$ (whose coordinate representation happens to be $[F_x, F_y, F_z]^T$) are the true objects of physics. Using them, you *may* introduce a coordinate system, but you are *not dependent on it.* The equation for the force now reads

$$\vec{F} = -\frac{GmM}{\|\vec{x}\|^2}\,\frac{\vec{x}}{\|\vec{x}\|}$$

Physics also explains the word "scalar" for number. The idea behind the word is that pure number quantities (which do not have a direction, unlike forces), e.g., temperature, or mass, can be represented on a *scale*, and that's where the name comes from.

The most relevant example for a scalar product in physics is *work*. Lifting a heavy object qualifies as work in physics, but moving it horizontally, does not (assuming there is no friction); nor does merely holding the object.



Vectors that are orthogonal to each other: Orthogonal means perpendicular in this context. Based on the formula $\vec{u} \cdot \vec{v} = \|\vec{u}\|\,\|\vec{v}\| \cos\varphi$, we see that two vectors $\vec{u}$ and $\vec{v}$ are orthogonal (perpendicular), if and only if $\vec{u} \cdot \vec{v} = 0$. Geometrically, this makes sense if neither of $\vec{u}$, $\vec{v}$ is the zero vector. By definition, we call $\vec{u}$ and $\vec{v}$ orthogonal, whenever $\vec{u} \cdot \vec{v} = 0$, even if this relation is due not to a right angle, but to the vanishing of one of the vectors: By definition, the zero vector is orthogonal to any vector. /* In the abstract theory of vector spaces, the same algebraic definition will apply, even in the absence of a geometrical interpretation as an angle */ Ch. 6 — /* When in calculus, or in the theory of Fourier series, the formula $\int_0^{2\pi} \sin mx \cos nx\, dx = 0$ and similar formulas are called orthogonality relations, then there is indeed a generalized version of this present definition of orthogonality behind the name. You'll see the analogy in chapter 6 */

Cross product of vectors; geometric definition: The cross product $\vec{u} \times \vec{v}$ (also called *vector product*) is defined *only* for geometric vectors *in space*, and it is again a vector. The length of $\vec{u} \times \vec{v}$ is $\|\vec{u}\| \, \|\vec{v}\| \sin \varphi$ where φ is the angle between $\vec{u}$ and $\vec{v}$. This is the area of the parallelogram spanned by $\vec{u}$ and $\vec{v}$. If $\vec{u}$ and $\vec{v}$ are parallel (i.e., $\varphi = 0$ or $\varphi = \pi$), then the length of $\vec{u} \times \vec{v}$ is 0 by this formula, and so $\vec{u} \times \vec{v} = \vec{0}$. Otherwise, the direction of $\vec{u} \times \vec{v}$ is perpendicular to the plane spanned by $\vec{u}$ and $\vec{v}$. These two conditions (a certain length and direction perpendicular to a certain plane) leave only two choices for $\vec{u} \times \vec{v}$, one of which is the negative of the other. The decision between the two possibilities is made as follows: Take the right hand with the thumb in hitchhiking position. Think of turning the first vector $\vec{u}$ towards the second vector $\vec{v}$. The four fingers should point in the direction in which you would turn $\vec{u}$. Then the thumb gives the direction of $\vec{u} \times \vec{v}$.

Rules for the cross product:

(*a*) $\vec{u} \times \vec{v} = -\vec{v} \times \vec{u}$ Order is of essence! — In particular $\vec{u} \times \vec{u} = \vec{0}$

(*b*) Parentheses are of essence! It is NOT true that "$(\vec{u} \times \vec{v}) \times \vec{w} = \vec{u} \times (\vec{v} \times \vec{w})$" Therefore "$\vec{u} \times \vec{v} \times \vec{w}$" WITHOUT parentheses is MEANINGLESS

(*c*) $\vec{u} \times (\vec{v} + \vec{w}) = \vec{u} \times \vec{v} + \vec{u} \times \vec{w}$
$(\vec{v} + \vec{w}) \times \vec{u} = \vec{v} \times \vec{u} + \vec{w} \times \vec{u}$

(*d*) $\vec{u} \times (k\vec{v}) = k(\vec{u} \times \vec{v})$, $(k\vec{u}) \times \vec{v} = k(\vec{u} \times \vec{v})$

(a) is clear if you look at the right hand rule. To understand (b), try a counterexample: $\vec{u} = [1, 0, 0]^T$, $\vec{v} = \vec{w} = [0, 1, 0]^T$. Rule (c) can be seen geometrically in some special cases, but will become clearer from the component formulas below. You should be able to see rule (d) geometrically.

The cross product in terms of components: The formula for the cross product in a *right-handed* coordinate system is as follows:

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \times \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{bmatrix}$$

A coordinate system is called right-handed, if the right-hand rule applies as follows: you try to turn the first coordinate axis towards the second coordinate axis, and to display the turning direction by means of the right hand in hitchhike position. If the thumb points towards the third coordinate axis (not in the opposite direction), the coordinate system is right-handed; otherwise left-handed. It's not so easy to come up with the above formula, based on the geometric definition of $\vec{u} \times \vec{v}$, but it is straightforward to check that the vector $[u_2v_3 - u_3v_2 \, , \; u_3v_1 - u_1v_3 \, , \; u_1v_2 - u_2v_1]^T$ is indeed orthogonal both to $\vec{u}$ and $\vec{v}$, and has the correct length. Note that

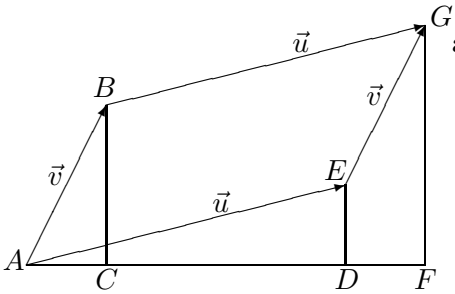
$$\|\vec{u} \times \vec{v}\| = \|\vec{u}\| \, \|\vec{v}\| \, |\sin \varphi| = \sqrt{\|\vec{u}\|^2 \|\vec{v}\|^2 - (\vec{u} \cdot \vec{v})^2}$$

The cross product in physics: /* Here are two occurrences of the cross product in physics: if an electrical particle with charge q (a scalar) moves with velocity $\vec{v}$ (a vector) through a magnetic field $\vec{B}$ (also a vector), it experiences a force $\vec{F}$

(called Lorentz force) that is perpendicular to the direction of motion, and also perpendicular to the magnetic field: $\vec{F} = q\vec{v} \times \vec{B}$.

If (e.g.) wind moves with velocity $\vec{v}$ on the earth, then due to the rotation of the earth, the wind undergoes a sidewards acceleration $\vec{a}$ according to the formula $\vec{a} = 2\vec{\omega} \times \vec{v}$. Here $\vec{\omega}$ is a vector pointing in the direction of the axis of rotation (namely form the south pole towards the north pole), but with length equal the angular velocity of the rotation (namely for the earth, $2\pi/24$ hours). This is called the Coriolis acceleration, caused by the Coriolis force, which applies to motion in rotating systems. You can experience the Coriolis force if you spin around like a figure skater on ice doing a pirouette, but with arms stretched out initially. If you then move your arms towards your body, you feel the force on the arms. For the same reason, long range air currents that would otherwise blow from the equator northwards, get deflected eastward. */For math, phys' and mech-engr majors

Relations among cross product, volume, area, and determinants: For vectors $[u_1, u_2]^T$ and $[v_1, v_2]^T$ in the plane, the (oriented) area of the parallelogram spanned by them is $u_1v_2 - u_2v_1$, as can be seen in the figure.



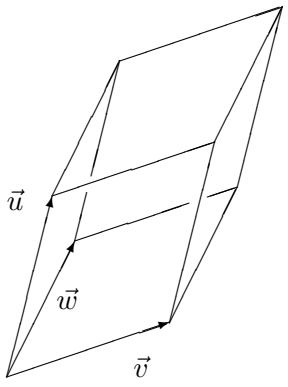
$$\begin{aligned}\text{area } AEGB &= \text{area } ACB + \text{area } CFGB \\ &\quad - \text{area } ADE - \text{area } DFGE \\ &= \frac{1}{2}v_1v_2 + u_1\frac{v_2 + (u_2 + v_2)}{2} \\ &\quad - \frac{1}{2}u_1u_2 - v_1\frac{u_2 + (u_2 + v_2)}{2} \\ &= u_1v_2 - u_2v_1\end{aligned}$$

This number $u_1v_2 - u_2v_1$ is exactly the third component of the cross product of the spatial vectors $\vec{u} = [u_1, u_2, 0]^T$ and $\vec{v} = [v_1, v_2, 0]^T$. The expression $u_1v_2 - u_2v_1$ is also written as

$$u_1v_2 - u_2v_1 =: \begin{vmatrix} u_1 & v_1 \\ u_2 & v_2 \end{vmatrix} =: \det \begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \end{bmatrix}$$

and is called the *determinant* of the 2×2 matrix $\begin{bmatrix} u_1 & v_1 \\ u_2 & v_2 \end{bmatrix}$. In other words, the determinant of a 2×2 matrix represents the oriented area of the parallelogram spanned by its column vectors.

The mixed product $\vec{u} \cdot (\vec{v} \times \vec{w})$, often called *scalar triple product*, of three spatial vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$ is the oriented volume of the parallelepiped spanned by these three vectors. (The crazy tongue-twister "parallelepiped" denotes the 3-dimensional analog of a parallelogram.)



the parallelepiped spanned by $\vec{u}$, $\vec{v}$, $\vec{w}$

This is because $\|\vec{v}\times\vec{w}\|$ is the base area, and $\|\vec{u}\|\cos\varphi$ (with φ the angle between $\vec{v}\times\vec{w}$ and $\vec{u}$) is just the height. We say *oriented* volume, because this number could be positive or negative, depending on the order in which $\vec{u}$, $\vec{v}$ and $\vec{w}$ are listed. The (standard) volume is the absolute value of the oriented volume.

In coordinates:

$$\vec{u}\cdot(\vec{v}\times\vec{w}) = \begin{bmatrix}u_1\\u_2\\u_3\end{bmatrix}\cdot\left(\begin{bmatrix}v_1\\v_2\\v_3\end{bmatrix}\times\begin{bmatrix}w_1\\w_2\\w_3\end{bmatrix}\right) = \begin{bmatrix}u_1\\u_2\\u_3\end{bmatrix}\cdot\begin{bmatrix}v_2w_3-v_3w_2\\v_3w_1-v_1w_3\\v_1w_2-v_2w_1\end{bmatrix} = u_1v_2w_3+u_2v_3w_1+$$

$$+u_3v_1w_2-u_1v_3w_2-u_2v_1w_3-u_3v_2w_1 =: \begin{vmatrix}u_1&v_1&w_1\\u_2&v_2&w_2\\u_3&v_3&w_3\end{vmatrix} =: \det\begin{bmatrix}u_1&v_1&w_1\\u_2&v_2&w_2\\u_3&v_3&w_3\end{bmatrix}$$

This latter expression is called the determinant of a 3×3 matrix, or, in short, a 3×3 determinant.

In general, to every $n\times n$ matrix A, we can assign a number $\det A$, called its determinant. For $n=2$ it represents an (oriented) area, for $n=3$, it represents an (oriented) volume. Below, it will be dicussed algebraically for general n (2, 3 or larger).

Note that the cross product can be written neatly in terms of determinants, if we use the abbreviations $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ for the unit coordinate vectors $[1,0,0]^T$, $[0,1,0]^T$ and $[0,0,1]^T$ respectively:

$$\begin{bmatrix}u_1\\u_2\\u_3\end{bmatrix}\times\begin{bmatrix}v_1\\v_2\\v_3\end{bmatrix} = \begin{vmatrix}\vec{e}_1&u_1&v_1\\\vec{e}_2&u_2&v_2\\\vec{e}_3&u_3&v_3\end{vmatrix} = \begin{vmatrix}\vec{e}_1&\vec{e}_2&\vec{e}_3\\u_1&u_2&u_3\\v_1&v_2&v_3\end{vmatrix} = \begin{bmatrix}\begin{vmatrix}u_2&v_2\\u_3&v_3\end{vmatrix}\\\begin{vmatrix}u_3&v_3\\u_1&v_1\end{vmatrix}\\\begin{vmatrix}u_1&v_1\\u_2&v_2\end{vmatrix}\end{bmatrix}$$

More formulas for cross and scalar triple products:
(not intended for memorization)

$$\vec{u}\cdot(\vec{v}\times\vec{w}) = \vec{v}\cdot(\vec{w}\times\vec{u}) = \vec{w}\cdot(\vec{u}\times\vec{v})$$
$$\vec{u}\times(\vec{v}\times\vec{w}) = \vec{v}\,(\vec{u}\cdot\vec{w}) - \vec{w}\,(\vec{u}\cdot\vec{v})$$
$$(\vec{u}\times\vec{v})\times\vec{w} = \vec{v}\,(\vec{u}\cdot\vec{w}) - \vec{u}\,(\vec{v}\cdot\vec{w})$$

The determinant of a square matrix: To every square matrix A, one can assign a number, $\det A$, called the determinant of A. An $n\times n$ determinant $\det A$ is a sum of the $n!$ terms $\pm a_{1j_1}a_{2j_2}\dots a_{nj_n}$, where $(j_1,j_2,\dots,j_n)$ can be any permutation of $(1,2,\dots,n)$, and the $+$ sign is to be chosen for so-called *even* permutations, the $-$ sign for so-called *odd* permutations:

Even and odd permutations: A permutation is called even, if it occurs an even number of times that a larger number precedes a smaller one. Otherwise the permutation is called odd. There is another way to characterize even and odd

permutations (and it gives the same result, although this is not obvious): Think of achieving the permutation by successive swapping of pairs of numbers. The permutation is even (resp. odd), if an even (resp. odd) number of swaps is required. Example: The permutation $(5, 3, 4, 6, 2, 1)$ is odd: there are 11 occurrences of a larger number preceding a smaller one, namely: $5 > 3$, $5 > 4$, $5 > 2$, $5 > 1$; $3 > 2$, $3 > 1$; $4 > 2$, $4 > 1$; $6 > 2$, $6 > 1$; $2 > 1$. You can obtain (534621) from (123456) by means of (e.g.) 5 swaps, e.g., (123456) $\to$ (523416) $\to$ (532416) $\to$ (534216) $\to$ (534261) $\to$ (534621). You may also do it with 7 or 9 swaps, but it won't be possible with an even number of swaps. — Calculating $n \times n$ determinants by means of adding $n!$ products (with $+$ or $-$ sign, depending on whether the permutation is even or odd), is outrageously tedious, except for $n = 2$ and $n = 3$, or rare cases where most factors vanish immediately.

More about determinants: The determinant satisfies the following properties, which can be used to calculate it:
• If you exchange two rows in a matrix, the determinant changes its sign.
• If you multiply a row by a number k, the determinant gets multiplied by k. If a row consists of zeros only, the determinant is 0.
• If you add a multiple of one row to another row, the determinant remains the same.

The same rules apply with "columns" instead of "rows". In principle, these rules permit to calculate determinants by means of row operations that transform the given matrix into reduced row echelon form. It is true, but far from obvious, that the result of the calculation does not depend on the precise row operations nor the order in which they are carried out.

They are geometrically intuitive if you have the area / volume interpretation for $n = 2$ or $n = 3$ in mind.

- The determinant of a triangular matrix is the product of its diagonal entries.
- $\det(A^T) = \det A$
- $\det(AB) = (\det A)\,(\det B)$
- $\det(kA) = k^n \det A$
- A matrix is invertible, if and only if its determinant is non-zero. Then, $\det(A^{-1}) = 1/\det A$
- $\det(A + B)$ is NOT $\det A + \det B$; no formula available for $\det(A + B)$
- If A, B, C differ only in a single row, say the r^{th} row, and if furthermore the r^{th} row of C is the sum of the r^{th} rows of A and B, then $\det C = \det A + \det B$

Minors and cofactors: The minor of an entry a_{ij} of a square $(n \times n)$ matrix A is the determinant of the $(n-1) \times (n-1)$ matrix obtained from A by deleting the column and the row containing a_{ij}. Let's call the minor as M_{ij}. The cofactor C_{ij} is defined as $C_{ij} = (-1)^{i+j} M_{ij}$. The cofactors play a role in the following formula for determinants (cofactor expansion along the i^{th} row, or along the j^{th} column):

$$\det A = a_{i1}C_{i1} + a_{i2}C_{i2} + \ldots + a_{in}C_{in}$$
$$\det A = a_{1j}C_{1j} + a_{2j}C_{2j} + \ldots + a_{nj}C_{nj}$$

/* One needs permutation groups: There is no point in trying to prove the basic results about determinants at the level of a first course in matrix algebra. The distinction between even and odd permutations can be naively defined by counting inversions, but it is a more sophisticated thing to understand why such a definition makes any sense. The key issue is that permutations form a *group*, as defined in abstract algebra; in order to prove results about determinants, one must first study this group. */ Abstract algebra

Calculational effectiveness: If you calculate an $n \times n$ determinant based on the definition, you have to add up $n!$ terms, each of which is a product of n factors. For $n = 10$, this would be 3.6 million additions and 32 million multiplications. The very same row operations that are used for an LU decomposition (possibly together with swapping some rows) can be achieved with roughly $n^3/3$ additions and $n^3/3$ multiplications (each about 330 for $n = 10$), so row or column operations are much more effective. You use them in a convenient way to create many zeros with little work.

Cofactor expansions alone would end up in the same $n!$ terms as a calculation by the original definition. They are useful in combination with row or column operations: First use row or column oerations to obtain a column or row that contains many zeros (typically only one, at most two nonzero entries), then use a cofactor expansion on that column / row.

The adjoint matrix: Given a square matrix A, the matrix whose entries are the cofactors C_{ij} is called the cofactor matrix. In itself, the cofactor matrix is of little importance; however its transpose is called the *adjoint matrix of* A, and written as $\operatorname{adj}(A)$; and the adjoint matrix does play a role: Namely, there is a "formula" for the inverse matrix:

$$A \text{ is invertible if and only if } \det A \neq 0, \text{ and then } A^{-1} = \frac{1}{\det A} \operatorname{adj}(A)$$

This formula is *not* useful for practical calculation of A^{-1}, except if $n = 2$ or $n = 3$. Too much calculational effort. However, it serves a theoretical purpose: For instance, it tells us, without actual calculation, that all entries of $A(x) = \begin{bmatrix} 1 & 2 & x \\ x & 2 & 1 \\ -1 & 1 & 4 \end{bmatrix}^{-1}$ are rational functions of x (and therefore depend continously on x as long as $A(x)$ is invertible).

Note however that for $n = 2$, the formula $A^{-1} = \operatorname{adj}(A)/\det A$ is very convenient, and it is nothing but the good old

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

for an invertible matrix.

/* Warning about the notion "adjoint matrix": There are two different concepts in mathematics that both carry the name "adjoint matrix", but have nothing to do with each other. In this course, we only need the one defined above. However,

you are very likely to encounter the other notion as well some time later. If you permit complex numbers as entries of matrices, then the conjugate transpose A^* of a matrix A (discussed in Ch. 10, which we will not cover) is sometimes also called adjoint. You will most likely encounter the concept of conjugate transpose some time, if you are a physics or math major. (Physicists have much use for A^*, but hardly any for $\operatorname{adj} A$, and are perfectly at ease with calling A^* "adjoint" of A. So when you encounter the notion outside this class, watch out what it means there. */

Cramer's rule: If A is an invertible $n \times n$ matrix, the solution of the linear system $A\boldsymbol{x} = \boldsymbol{b}$ is given by

$$x_1 = \frac{\det A_1}{\det A}, \quad x_2 = \frac{\det A_2}{\det A}, \ldots, \quad x_n = \frac{\det A_n}{\det A},$$

where you obtain the matrix A_i from the matrix A by replacing its i^{th} column with the right hand side $\boldsymbol{b}$. — Cramer's is *not* useful for solving linear systems in practice, except if $n = 2$ or $n = 3$. Too much calculational effort. It serves a theoretical purpose; similarly as for the formula $A^{-1} = \operatorname{adj}(A)/\det A$.

One other instance where Cramer's rule may be helpful in practice is, if you only are interested in solving for *one* variable in a system of equations, but do not care about the other variables.

n-dimensional space (n-space): $\mathbb{R}^n$ denotes the set of ordered n-tupels (i.e. pairs, triples, quadruples etc.) of real numbers: $(x_1, x_2, \ldots x_n)$ where each x_i is a real number. We can think of an element of $\mathbb{R}^2$ as a point in the plane, and of an element of $\mathbb{R}^3$ as a point in space. Equivalently, we may think of them as a vector in the plane or in space, where a point corresponds to its position vector: the position vector of a point is the vector pointing from the origin of the coordinate system to that point. The zero vector $\mathbf{0} = (0, 0, \ldots, 0)$ corresponds to the origin of the coordinate system. As anticipated already previously, we will soon need to write the vectors as *column* vectors (such as to multiply them with matrices); so we will write $[x_1, x_2, \ldots x_n]^T$ rather than $(x_1, x_2, \ldots x_n)$.

In the absence of an immediate geometric intuition, we will still call $\mathbb{R}^n$ for $n > 3$ the n-dimensional space, or n-space, and we will refer to its elements as points, and will use geometric notions from 2-space (the plane) or 3-space (our intuitive ambient space) also in $\mathbb{R}^n$ by analogy.

Vector arithmetic in n-space: Elements of $\mathbb{R}^n$ can be added and can be multiplied with real numbers; the definition of these operations is as expected:

$$[u_1, u_2, \ldots u_n]^T + [v_1, v_2, \ldots v_n]^T = [u_1 + v_1, u_2 + v_2, \ldots u_n + v_n]^T$$
$$c[u_1, u_2, \ldots u_n]^T = [cu_1, cu_2, \ldots cu_n]^T$$

The rules for this arithmetic are the very same we encountered in 2-space and 3-space (see page 9), and they will be same in abstract vector spaces (to be discussed

below), namely:

$\boldsymbol{u}+\boldsymbol{v}=\boldsymbol{v}+\boldsymbol{u}$	$(k+l)\boldsymbol{u}=k\boldsymbol{u}+l\boldsymbol{u}$
$(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}=\boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})$	$k(\boldsymbol{u}+\boldsymbol{v})=k\boldsymbol{u}+k\boldsymbol{v}$
$\boldsymbol{u}+\boldsymbol{0}=\boldsymbol{u}$	$k(l\boldsymbol{u})=(kl)\boldsymbol{u}$
$\boldsymbol{u}+(-\boldsymbol{u})=\boldsymbol{0}$	$1\boldsymbol{u}=\boldsymbol{u}$

where $-\boldsymbol{u}$ is shorthand for $(-1)\boldsymbol{u}$, and $\boldsymbol{u}-\boldsymbol{v}$ will be shorthand for $\boldsymbol{u}+(-\boldsymbol{v})$.

Norm and dot product in n-space: For $\boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^n$, we define the (euclidean) norm $\|\boldsymbol{u}\|:=\sqrt{u_1^2+u_2^2+\ldots+u_n^2}$, and the dot product $\boldsymbol{u}\cdot\boldsymbol{v}:=u_1v_1+u_2v_2+\ldots u_nv_n$. Clearly, the norm can be expressed in terms of the dot product: $\|\boldsymbol{u}\|=\sqrt{\boldsymbol{u}\cdot\boldsymbol{u}}$. Conversely (but not quite so obviously), the dot product can be expressed in terms of the norm as well:

$$\boldsymbol{u}\cdot\boldsymbol{v}=\frac{1}{4}\left(\|\boldsymbol{u}+\boldsymbol{v}\|^2-\|\boldsymbol{u}-\boldsymbol{v}\|^2\right)$$

This formula is sometimes known under the name of *polarization identity*, and you can immediately convince yourself that it is correct by writing the right hand side in terms of the dot product, and simplifying.

The rules for the dot product are are the same as in 2-space or 3-space (see page 9):

$\boldsymbol{u}\cdot\boldsymbol{v}=\boldsymbol{v}\cdot\boldsymbol{u}$
$\boldsymbol{u}\cdot(\boldsymbol{v}+\boldsymbol{w})=\boldsymbol{u}\cdot\boldsymbol{v}+\boldsymbol{u}\cdot\boldsymbol{w}$
$\boldsymbol{u}\cdot(k\boldsymbol{v})=k(\boldsymbol{u}\cdot\boldsymbol{v})$
$\boldsymbol{u}\cdot\boldsymbol{u}>0$ unless $\boldsymbol{u}=\boldsymbol{0}$ (in which case the dot product equals 0)

/* The concept of a norm is amenable to vast generalization; this is why we may call the particular norm used in this course the *euclidean* norm, to distinguish it from other examples of more general norms. However, as there is no other norm but the euclidean one in the course, we may also omit the adjective euclidean. */

Another important fact (in advanced calculus applications) is the inequality $|\boldsymbol{u}\cdot\boldsymbol{v}|\leq\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|$, which is known under the name Cauchy-Schwarz inequality in the western literature, and under the name (Cauchy-)Bunyakovsky inequality among Russian scholars. As an aside, the Cauchy-Schwarz inequality permits you to carry over the geometric intuition of an angle between vectors from 2-space and 3-space into n-space, by defining the angle between $\boldsymbol{u}$ and $\boldsymbol{v}$ to be $\arccos\Big(\boldsymbol{u}\cdot\boldsymbol{v}/(\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|)\Big)$.

Vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are called *orthogonal*, if their dot product is zero. For orthogonal vectors, the Pythagorean theorem $\|\boldsymbol{u}+\boldsymbol{v}\|^2=\|\boldsymbol{u}\|^2+\|\boldsymbol{v}\|^2$ applies.

Dot product and matrix product: For column vectors $\boldsymbol{u}$ and $\boldsymbol{v}$, it holds obviously: $\boldsymbol{u}\cdot\boldsymbol{v}=\boldsymbol{u}^T\boldsymbol{v}=\boldsymbol{v}^T\boldsymbol{u}$.

Distance and norm in n-dimensional space: /* The following properties of the norm are easy, but need to be stressed here, because they, too, will permit a vast generalization in more advanced treatments: */ Adv. calculus; functional analysis

(a) $\|\boldsymbol{u}\|\geq 0$ with $\|\boldsymbol{u}\|=0$ if and only if $\boldsymbol{u}=\boldsymbol{0}$
(b) $\|k\boldsymbol{u}\|=|k|\,\|\boldsymbol{u}\|$
(c) $\|\boldsymbol{u}+\boldsymbol{v}\|\leq\|\boldsymbol{u}\|+\|\boldsymbol{v}\|$

They spawn the concept of a distance between two points (or vectors) in n-space, which generalizes the euclidean distance in 2-space and 3-space: The distance between $\boldsymbol{u}$ and $\boldsymbol{v}$ is denoted by $d(\boldsymbol{u},\boldsymbol{v})$ and defined as $d(\boldsymbol{u},\boldsymbol{v}) := \|\boldsymbol{u}-\boldsymbol{v}\|$. Its properties are:

(a) $d(\boldsymbol{u},\boldsymbol{v}) \geq 0$ with $d(\boldsymbol{u},\boldsymbol{v}) = 0$ if and only if $\boldsymbol{u} = \boldsymbol{v}$
(b) $d(\boldsymbol{u},\boldsymbol{v}) = d(\boldsymbol{v},\boldsymbol{u})$
(c) $d(\boldsymbol{u},\boldsymbol{w}) \leq d(\boldsymbol{u},\boldsymbol{v}) + d(\boldsymbol{v},\boldsymbol{w})$

The properties (c) for norm and distance are also known under the name of *triangle inequality.*

/* The distance, too, plays no big role for linear algebra proper, but is introduced as a service for advanced calculus, for a modern language version of multivariable calculus, and for functional analysis. */

/* In contradistinction to more general norms, our euclidean norm obeys another formula: $\|\boldsymbol{u}+\boldsymbol{v}\|^2+\|\boldsymbol{u}-\boldsymbol{v}\|^2 = 2(\|\boldsymbol{u}\|^2+\|\boldsymbol{v}\|^2)$. As it turns out, it is this formula that is responsible for the fact that our norm can be wedded to a dot product according the polarization identity. */ functional analysis

$m \times n$ matrices and functions (linear transformations) $\mathbb{R}^n \to \mathbb{R}^m$: Every $m \times n$ matrix A represents a linear transformation $f : \mathbb{R}^n \to \mathbb{R}^m$, which is given by the formula $\boldsymbol{y} = f(\boldsymbol{x}) = A\boldsymbol{x}$. A function from $\mathbb{R}^n$ to $\mathbb{R}^m$ assigns to every vector $\boldsymbol{x} \in \mathbb{R}^n$ (viewed as "input" of the function) an "output" $\boldsymbol{y} \in \mathbb{R}^m$. To stress this notion that a function f converts an input from $\mathbb{R}^n$ into an output in $\mathbb{R}^m$, we use the arrow notation, i.e., we write $f : \mathbb{R}^n \to \mathbb{R}^m$; $\boldsymbol{x} \mapsto f(\boldsymbol{x})$. The standard arrow $(\to)$ goes between the *set* where the function takes input from and the *set* in which its output lies; the tailed arrow $(\mapsto)$ goes between the *individual* input vector and its corresponding output.

General remarks on geometric representation of functions $\mathbb{R}^n \to \mathbb{R}^m$: In single variable calculus, we represent a function $\mathbb{R} \to \mathbb{R}$ by its graph, which lies in $\mathbb{R}^2$. For a function $\mathbb{R}^2 \to \mathbb{R}^2$ this option is already unavailable, because the graph would have to be in $\mathbb{R}^4$, and we cannot draw $\mathbb{R}^4$. However we can understand functions $\mathbb{R}^2 \to \mathbb{R}^2$ geometrically by discussing what they do to certain input sets. In linear algebra, we restrict the discussion to *linear* functions $\boldsymbol{x} \mapsto A\boldsymbol{x}$, and we call them linear transformations. There are only a few interesting special cases to be discussed, and even the most general case can be described without too much difficulty:

Special linear transformations $\mathbb{R}^2 \to \mathbb{R}^2$:

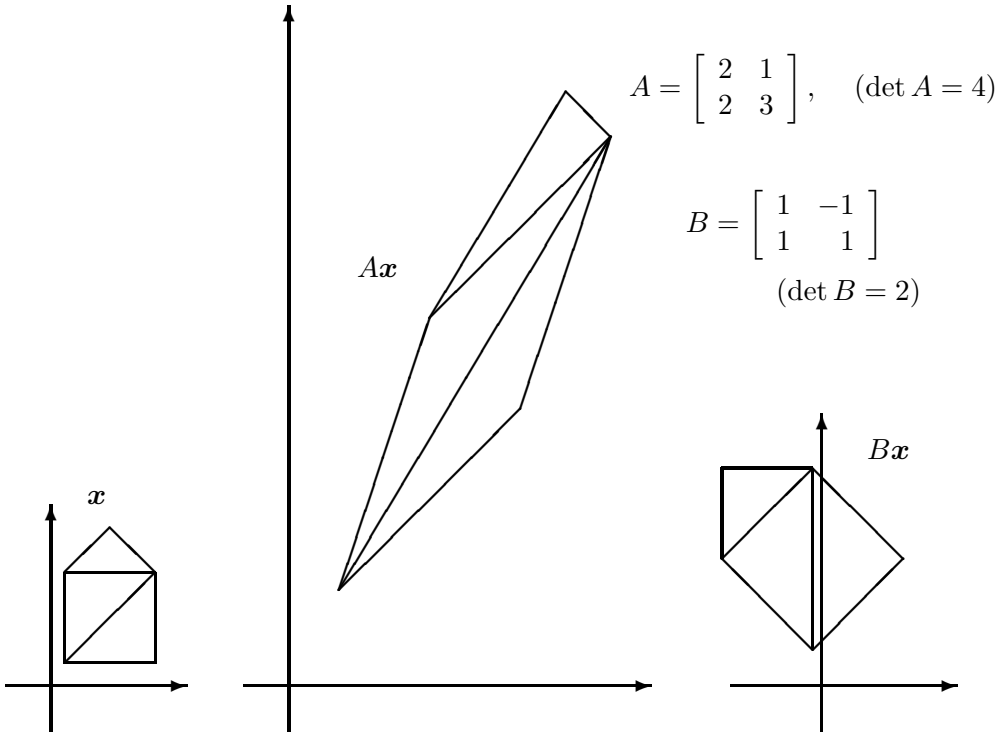
A **rotation** in the plane by the angle φ (counterclockwise) is described by the matrix $A = \begin{bmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{bmatrix}$. For any planar vector $\boldsymbol{x}$, you get $A\boldsymbol{x}$ from $\boldsymbol{x}$ by rotating it about the origin by an angle φ (counterclockwise).

A **reflection** with respect to the x_1 axis is described by the matrix $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Similarly, a reflection with respect to the x_2 axis is described by the matrix $A = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$. Reflections with respect to other lines will be discussed more generally below.

A **dilation** (to-scale magnification) is described by a scalar multiple of the identity matrix, like $A = \begin{bmatrix} a & 0 \\ 0 & a \end{bmatrix}$, where a is the magnification factor. If $a > 1$, we really have a magnification; if $0 < a < 1$ we have actually a reduction.

A **projection** onto (e.g.) the x_1 axis is described by the matrix $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. This projection leaves the first coordinate unchanged, but makes the second coordinate to zero. The image of the vector $\boldsymbol{x}$ under this projection is the "shadow" it casts on the x_1 axis, if light falls in vertically. Hence the name projection. Projections onto other lines (or planes in higher dimensions) will be discussed below.

A **general** linear transformation $\mathbb{R}^2 \to \mathbb{R}^2$ includes all of the above as special cases. The following pictures are representative of what it can do:



Rotations in n-space: Intuitively, a rotation should be a transformation $\boldsymbol{x} \mapsto Q\boldsymbol{x}$ with the following properties: If you take the unit coordinate vectors (also called standard basis vectors) $\boldsymbol{e}_1 = [1, 0, 0, \ldots, 0]^T$, $\boldsymbol{e}_2 = [0, 1, 0, \ldots, 0]^T$, $\ldots$, $\boldsymbol{e}_n = [0, 0, \ldots, 0, 1]^T$, then the rotated vectors $Q\boldsymbol{e}_1$, $Q\boldsymbol{e}_2$,$\ldots$, $Q\boldsymbol{e}_n$ should still be unit vectors and they should still be orthogonal. But these vectors are the columns of the matrix Q; therefore the condition for orthogonality and length one of these vectors can be written as $Q^TQ = I$. A matrix satisfying $Q^TQ = I$ is called an *orthogonal* matrix.

We call a matrix Q a rotation matrix, if $Q^TQ = I$ and $\det Q = 1$. (The determinant condition screens out reflections.)

Reflections in n-space: In $\mathbb{R}^3$, a *unit* vector $\boldsymbol{u}$ ($\|\boldsymbol{u}\| = 1$) determines a plane that consists of those vectors (beginning in the origin) that are orthogonal to $\boldsymbol{u}$. In $\mathbb{R}^2$, they form a line instead, in $\mathbb{R}^n$ some 'dont-know-what-to-call-it' (which is actually called a hyperplane). Think of the plane (line, hyperplane) as a mirror.

Any vector $\boldsymbol{x}$ can be decomposed as $\boldsymbol{x} = \boldsymbol{x}_{||} + \boldsymbol{x}_{\perp}$, namely with a component $\boldsymbol{x}_{||}$ parallel to $\boldsymbol{u}$, and a component $\boldsymbol{x}_{\perp}$ orthogonal to $\boldsymbol{u}$. These components are $\boldsymbol{x}_{||} = (\boldsymbol{x} \cdot \boldsymbol{u})\boldsymbol{u}$ and $\boldsymbol{x}_{\perp} = \boldsymbol{x} - (\boldsymbol{x} \cdot \boldsymbol{u})\boldsymbol{u}$. The reflection in the plane (line, hyperplane) orthogonal to $\boldsymbol{u}$ leaves the orthogonal component $\boldsymbol{x}_{\perp}$ unchanged, but replaces the parallel component with its negative, i.e., it is represented by a matrix $R_{\boldsymbol{u}}$ such that $R_{\boldsymbol{u}}\boldsymbol{x} = \boldsymbol{x}_{\perp} - \boldsymbol{x}_{||} = (I - 2\boldsymbol{u}\boldsymbol{u}^T)\boldsymbol{x}$. An immediate consequence from $R_{\boldsymbol{u}} = I - 2\boldsymbol{u}\boldsymbol{u}^T$ is $R_{\boldsymbol{u}}^2 = I$ and $R_{\boldsymbol{u}}^T = R_{\boldsymbol{u}}$.

Projections in n-space: With the same notation as before, a projection onto the plane (line, hyperplane) orthogonal to $\boldsymbol{u}$ can be pictured as the shadow on that plane (line, hyperplane) when the light comes from direction $\boldsymbol{u}$). This projection leaves the orthogonal component unchanged, but sets the parallel component to 0. It is therefore represented by the matrix $P_{\boldsymbol{u}} = I - \boldsymbol{u}\boldsymbol{u}^T$. An immediate consequence is $P_{\boldsymbol{u}}^2 = P_{\boldsymbol{u}}$ and $P_{\boldsymbol{u}}^T = P_{\boldsymbol{u}}$.

A projection surface need not be orthogonal to the light rays. But the shadow of an object lying already on the projection surface is the object itself. (The shadow of a shadow is the shadow itself.) We generally would call a matrix P a projection matrix, if $P^2 = P$; we call it the matrix of an orthogonal projection, if $P^2 = P$ *and* $P^T = P$. More detailed discussions why this simple definition incorporates all the geometric ideas inherent in a projection go beyond this course.

Composition of transformations: If a transformation that maps $\boldsymbol{x}$ into $A\boldsymbol{x}$ ($\boldsymbol{x} \mapsto A\boldsymbol{x}$) is followed by another transformation that maps $\boldsymbol{y}$ into $B\boldsymbol{y}$ ($\boldsymbol{y} \mapsto B\boldsymbol{y}$), then the composition is $\boldsymbol{x} \mapsto A\boldsymbol{x} \mapsto BA\boldsymbol{x}$. The *composition* of transformations is therefore represented by the *matrix product*. Note that the transformation that is carried out *first* is the *last* (not first) factor in the matrix product. (Like functions in calculus: To find $\ln\cos\pi/6$, you *first* calculate the cosine, *then* the logarithm, not vice versa!)

Transformations and the determinant: Consider the unit coordinate vectors (also called standard basis vectors) $\boldsymbol{e}_1 = [1, 0, 0, \ldots, 0]^T$, $\boldsymbol{e}_2 = [0, 1, 0, \ldots, 0]^T$, $\ldots$, $\boldsymbol{e}_n = [0, 0, \ldots, 0, 1]^T$. Their images $A\boldsymbol{e}_1$, $A\boldsymbol{e}_2$,$\ldots$, $A\boldsymbol{e}_n$ under a transformation $\boldsymbol{x} \mapsto A\boldsymbol{x}$ are exactly the columns of A. These column vectors span a parallelogram (in 2-space) or a parallelepiped (in 3-space) whose area/volume is $\det A$, whereas the area of the original square (in 2-space) spanned by the vectors $\boldsymbol{e}_1, \boldsymbol{e}_2$, or the volume of the original cube (in 3-space) spanned by the vectors $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ was 1. The determinant of A gives the *volume magnification factor* of the transformation $\boldsymbol{x} \mapsto A\boldsymbol{x}$.

If you compose two transformations, each with its own volume magnification factor, then for the composite transformation, the volume magnification factors multiply. This is the reason why the previously encountered formula $\det(BA) = (\det B)(\det A)$ is true.

One-to-one transformations: A transformation is called one-to-one, if distinct points are mapped into distinct points. In other words, the transformation $\boldsymbol{x} \mapsto A\boldsymbol{x}$ is called one-to-one, if $A\boldsymbol{u} = A\boldsymbol{v}$ implies $\boldsymbol{u} = \boldsymbol{v}$. If A is a square matrix, then the transformation $\boldsymbol{x} \mapsto A\boldsymbol{x}$ is one-to-one, if and only if A is invertible.

Eigenvectors and eigenvalues: For certain linear transformations $\boldsymbol{x} \mapsto A\boldsymbol{x}$, there exist special vectors $\boldsymbol{v} \neq \boldsymbol{0}$ that get mapped into a scalar multiple of itself: $A\boldsymbol{v} = \lambda\boldsymbol{v}$ for a *scalar* λ. These vectors are called *eigenvectors* of the matrix A (or of the linear transformation $\boldsymbol{x} \mapsto A\boldsymbol{x}$), and the corresponding numbers λ are called *eigenvalues*.

The word "eigen" is German for "own" or "proper" and is one of the few instances of German loanwords in the English language (French for instance, translates the word completely: valeur propre = "proper value"). The choice of name indicates that eigenvectors and eigenvalues give some characteristic information about a transformation.

Eigenvectors and eigenvalues: important examples: For a rotation $\boldsymbol{x} \mapsto Q\boldsymbol{x}$ in 3-space, a vector $\boldsymbol{v}$ pointing along the axis of rotation will itself not change under the rotation: $Q\boldsymbol{v} = \boldsymbol{v}$. Such a vector is an eigenvector of Q with eigenvalue 1.

In contrast, for a rotation $\boldsymbol{x} \mapsto Q\boldsymbol{x}$ in 2-space, with $Q = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$ and with α *not* a multiple of π, no vector $\boldsymbol{x}$ has the property that its image $Q\boldsymbol{x}$ is a scalar multiple of $\boldsymbol{x}$. Therefore this matrix does not have eigenvectors nor eigenvalues.

/* When you reread this in a later semester, you may accept complex entries for vectors and matrices and work with complex numbers. Here we insist on geometric interpretation and therefore only accept real numbers. The claim that there are no eigenvectors and eigenvalues for a rotation matrix in 2-space depends on our acceptance of real numbers only. When you accept complex numbers and vectors later, you will find complex eigenvectors and eigenvalues for this matrix. */ M431, M453

For a reflection matrix $R_{\boldsymbol{u}} = I - 2\boldsymbol{u}\boldsymbol{u}^T$, the vector $\boldsymbol{u}$ gets reflected: $R_{\boldsymbol{u}}\boldsymbol{u} = -\boldsymbol{u}$. It is therefore an eigenvector with corresponding eigenvalue -1. So are all its multiples (like, e.g., $\frac{-7}{3}\boldsymbol{u}$). Vectors $\boldsymbol{v}$ that are orthogonal to $\boldsymbol{u}$ are unchanged under the reflection: $R_{\boldsymbol{u}}\boldsymbol{v} = \boldsymbol{v}$. These are therefore eigenvectors for eigenvalue 1. Vectors that are neither parallel nor orthogonal to $\boldsymbol{u}$ are not eigenvectors. — Remember that in the definition of $R_{\boldsymbol{u}}$, we assume $\boldsymbol{u}$ to have length one.

For a dilation kI, all vectors are eigenvectors with eigenvalue k.

For a projection matrix $P_{\boldsymbol{u}} = I - \boldsymbol{u}\boldsymbol{u}^T$, the vector $\boldsymbol{u}$ is an eigenvector with eigenvalue 0: $P_{\boldsymbol{u}}\boldsymbol{u} = \boldsymbol{0} = 0\boldsymbol{u}$. Vectors orthogonal to $\boldsymbol{u}$ are eigenvectors with eigenvalue 1. — Remember that in the definition of $P_{\boldsymbol{u}}$, we assume $\boldsymbol{u}$ to have length one.

Calculation of eigenvectors and eigenvalues: To find solutions $\boldsymbol{v} \neq \boldsymbol{0}$ to the equation $A\boldsymbol{v} = \lambda\boldsymbol{v}$ (and to find numbers λ for which nontrivial solutions $\boldsymbol{v}$ exist), you rewrite $A\boldsymbol{v} = \lambda\boldsymbol{v}$ as $(A - \lambda I)\boldsymbol{v} = \boldsymbol{0}$. Nontrivial solutions $\boldsymbol{v}$ exist, if and only if $\det(A - \lambda I) = 0$. This leads to an n^{th} degree polynomial equation in λ, and once you have its solutions λ_i (if any), you solve the linear system $(A - \lambda_i I)\boldsymbol{v} = \boldsymbol{0}$ to find eigenvectors for each of the eigenvalues λ_i.

Real vector spaces: A nonempty set V of objects is called a *(real) vector space*, if an operation + (called addition) and a multiplication with (real) numbers is defined for these objects in such a way that the usual rules of vector arithmetic encountered above and specifically relisted below apply to them. Namely these rules are:
(a) For any $\boldsymbol{u}$ and $\boldsymbol{v}$ in V, the sum $\boldsymbol{u}+\boldsymbol{v}$ is a well-defined object in V.
(b) $\boldsymbol{u}+\boldsymbol{v}=\boldsymbol{v}+\boldsymbol{u}$
(c) $(\boldsymbol{u}+\boldsymbol{v})+\boldsymbol{w}=\boldsymbol{u}+(\boldsymbol{v}+\boldsymbol{w})$
(d) There is an object $\mathbf{0}$ (called the zero vector) such that $\boldsymbol{u}+\mathbf{0}=\boldsymbol{u}$
(e) For each $\boldsymbol{u}$, there is an object $-\boldsymbol{u}$ (called the negative of $\boldsymbol{u}$) such that $\boldsymbol{u}+(-\boldsymbol{u})=\mathbf{0}$
(f) For any (real) number k and any object $\boldsymbol{u}$ in V, $k\boldsymbol{u}$ is a well-defined object in V.
(g) $k(\boldsymbol{u}+\boldsymbol{v})=k\boldsymbol{u}+k\boldsymbol{v}$
(h) $(k+l)\boldsymbol{u}=k\boldsymbol{u}+l\boldsymbol{u}$
(i) $k(l\boldsymbol{u})=(kl)\boldsymbol{u}$
(j) $1\boldsymbol{u}=\boldsymbol{u}$

Old example: $V=\mathbb{R}^n$.

A typical (new) example is the set $V=C^0[0,1]$ of continuous functions defined on the interval $[0,1]$, with real values. In calculus, you called such functions f and g rather than $\boldsymbol{u}$ and $\boldsymbol{v}$. But they do satisfy the above rules: What is the sum of the functions $f:x\mapsto x^2$ and $g:x\mapsto\sin x$ (in calculus, you'd have worded it a bit differently: "...sum of the functions f given by $f(x)=x^2$ and g given by $g(x)=\sin x$")? Well, it's the function $f+g:x\mapsto x^2+\sin x$, or in other words, $f+g$ is given by the formula $(f+g)(x):=f(x)+g(x)=x^2+\sin x$. Clearly $f+g=g+f$, and all the other rules hold as well. What should 7 times the function $f:x\mapsto x^2$ mean? It is the function called $7f:x\mapsto tf(x)=7x^2$. And again the rules listed above for multiplication with scalars apply.

With functions, you can do many more things than with vectors; for instance taking derivatives, composing them, multiplying them. When we call a collection of functions a vector space, we simply stress the particular operations with functions that are analog to operations with vectors, but at the same time show ourselves utterly disinterested in all the other things you can do with functions. When we call a set of functions a vector space we are essentially saying: all that we are going to do with the functions now depends only on addition and multiplication with numbers: We are zooming in on these features as the relevant ones. At another time, when we need to work with compositions or multiplication of functions, we would say "Yes, ok, this set of functions is a vector space all right, but that's not the issue here."

Counterexample: The set of all *nonnegative* continuous functions on the interval $[0,1]$ is *not* a vector space, because for f nonnegative, there is no *nonnegative* $(-f)$ such that $f+(-f)=0$.

Another example is $\mathbb{R}^{m\times n}$, the set of all $m\times n$ matrices. This set is a vector space, because all the rules (a)-(j) apply to addition of matrices and multiplication of matrices with numbers. Again, in calling $\mathbb{R}^{m\times n}$ a vector space, we technically stress that these rules do apply, but also pragmatically convey our disinterest (for the moment) in doing any other operations (like multiplying or transposing) with these matrices.

/* Abstract vector spaces: In the definition above, you can replace "real number" with "complex number" throughout and get the definition of a *complex vector space*. This is an important generalization for adv. calculus, analysis, differential equations, physics. Even more generally, if you know abstract algebra you can take any *field* instead of the set of real or complex numbers. This is relevant for abstract algebra, but irrelevant for adv. calculus, analysis, differential equations, physics. */

Subspace: A subset W of a vector space V is called a subspace of V (sometimes synonymously: a *linear subspace* of V), if for any $\boldsymbol{w}_1$ and $\boldsymbol{w}_2$ in W, their sum $\boldsymbol{w}_1 + \boldsymbol{w}_2$ is also in W, and also all multiples $k\boldsymbol{w}_1$ are in W. — In other words, a subset $W \subset V$ is a subspace, if W is a vector space in its own right, with the operations (addition and multiplication with scalars) that are already defined in the larger set V.

Linear combinations: Let $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ be a (finite) set of vectors in a vector space. A linear combination of / from S (or: a linear combination of $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\ldots$, $\boldsymbol{v}_n$) is any expression $k_1\boldsymbol{v}_1 + k_2\boldsymbol{v}_2 + \ldots + k_n\boldsymbol{v}_n$ with scalars $k_1, \ldots, k_n$. (Trivially, this includes sums with less than n terms, because the "unused" vectors can still be included formally with a coefficient 0.)

S need not be finite. But a linear combination from an infinite set $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots\}$ is still defined as a *finite* sum $k_1\boldsymbol{v}_1 + k_2\boldsymbol{v}_2 + \ldots + k_n\boldsymbol{v}_n$ involving as many (but only finitely many) vectors from S as you please. We don't have much need for this in this course, but linear combinations from infinite sets will be relevant for many applications, and they don't pose any extra difficulties at the level of this course.

Linear combinations from a set with only one vector, $S = \{\boldsymbol{v}\}$, are clearly just the multiples of $\boldsymbol{v}$. Linear combinations of the empty set $\{\}$ would be sums with no terms to sum up at all, and by definition such (empty) sums are the zero vector.

Span: The span of a set $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ of vectors from a vector space V is the set of all linear combinations of S. It forms a subspace of V and is denoted as $\operatorname{span} S = \operatorname{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$. Synonymous expressions are: the subspace (or vector space) spanned by S. The span of S is the smallest subspace of V containing S. In particular, $\operatorname{span}\{\} = \{\mathbf{0}\}$ (even though it may look weird when you see it the first time).

Again, S need not be a finite set; the linear combinations involved are still finite sums by definition. For instance, in the vector space $C^0(\mathbb{R})$ of all continuous real functions, the span of the set $S = \{1, x, x^2, x^3, \ldots\}$ is the subspace of all polynomials (of any degree). (The notation x^n for a *function* is sloppy, but common, and familiar to you. To put the record straight: I should have said, e.g., "The function f_3 given by $f_3(x) = x^3$")

Linearly dependent, linearly independent: A set $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ is called linearly independent, if the *only* way of writing the zero vector as a linear combination from S is with all coefficients equal to 0. Otherwise, the set is called linearly dependent. Ex: $\{[0,\ 1]^T, [1,\ 0]^T\}$ is linearly independent, because $k_1[0,\ 1]^T + k_2[1,\ 0]^T = [k_1,\ k_2]^T$ equals the zero vector *only* if $k_1 = k_2 = 0$. In contrast, $\{[0,\ 1]^T, [1,\ 0]^T, [2,\ 3]^T\}$ is linearly dependent, because $\mathbf{0}$ can be written

as a nontrivial linear combination. e.g.: $\mathbf{0} = 3\,[0,\ 1]^T + 2\,[1,\ 0]^T - [2,\ 3]^T$.

Basis: A basis of a vector space V is a linearly independent set S that spans V.

The following facts help to appreciate the meaning of a basis informally:
(1) If a set S spans V, then any larger set $\hat{S} \supset S$ also spans V (trivially), but a smaller set $\check{S} \subset S$ may or may not span V.
(2) If a set S is linearly independent, then any smaller set $\check{S} \subset S$ is still linearly independent, but a larger set $\hat{S} \supset S$ may or may not be linearly independent.
(3) Therefore a basis is a compromise: it must have enough vectors to span V, but no redundant ones, which would forfeit linear independence of the set S.

The following theorem sums this up more formally:
A *maximal* linearly independent set S in V is a basis. By "maximal" we mean it's impossible to join any other vector to S without losing linear independence. — A *minimal* spanning set S for V is a basis. By "minimal", we mean that you cannot discard any vector from the set S without losing the property that S spans V.

Dimension: A vector space is called *finite dimensional*, if it has a finite set as a basis. Otherwise the vector space is called infinite dimensional. (As an extreme example, the set $\{\mathbf{0}\}$ is a vector space, and $\{\}$ is the only basis for it.)
`/*` Does this mean that an infinite dimensional vector space has some *infinite* set as a basis? – Yes, this is indeed true. But it is not easy to understand *why* it is true, and in applications *outside* linear algebra (like, e.g., vector spaces of functions), it is also not a particularly useful fact, because you usually cannot get hold of one practically `*/`

There is (usually) a vast variety of choices of a basis for any given vector space. But each basis of a given vector space V has the same number of elements, and this number is called the dimension of V, in short $\dim V$.
`/*` It is also true that each basis of a given infinite dimensional vector space has "the same number" of elements, but this statement is much more subtle than it seems already and goes beyond this class. Nevertheless, `*/` infinite dimensional vector spaces are quite natural already at the level of this class; only we have no business worrying about their bases.

Theorems about bases, spanning, and linear independence: (the numbers refer to thm numbers in the textbook)
If S is lin. indep., and $\boldsymbol{v} \notin \operatorname{span} S$, then $S \cup \{\boldsymbol{v}\}$ is still lin. indep. (5.4.4a)
If $\boldsymbol{v} \in S$ and $\boldsymbol{v}$ is spanned by the remaining vectors of S (i.e., $\boldsymbol{v} \in \operatorname{span}(S \setminus \{\boldsymbol{v}\})$, then $S \setminus \{\boldsymbol{v}\}$ spans the same space as S. ($S \setminus \{\boldsymbol{v}\}$ means "S with the element $\boldsymbol{v}$ removed".) (5.4.4b)

For S a (finite) set of vectors in a (finite dimensional) vector space V, the following hold:
If S is lin. indep., but does not span V (and is therefore not a basis of V), then S can be enlarged to a basis of V. (5.4.6b)
If S spans V, but is not lin. indep. (and is therefore not a basis of V), then S can be reduced to a basis of V by removing appropriate vectors from S. (5.4.6a)

In an n-dimensional vector space,
a set with less than n vectors will never span V (a set with n or more vectors may or may not span V),
a set with more than n vectors will always be lin. dep. (a set with n or less vectors may or may not be lin. dep.),
a set with exactly n vectors is either lin. indep. *and* spanning, or else it is neither lin. indep. nor spanning. (5.4.5)

If W is a subspace of a finite dimensional vector space V, then $\dim W \leq \dim V$ with equality *only if* $W = V$. (5.4.7)

Standard basis: Every vector space (except $\{0\}$) has many bases. However, in some vector spaces, there is one basis on which everbody agrees that it is the simplest basis and the one that comes to mind first. We call such a basis the standard basis of that vector space. In particular, we agree that in $\mathbb{R}^n$, the basis

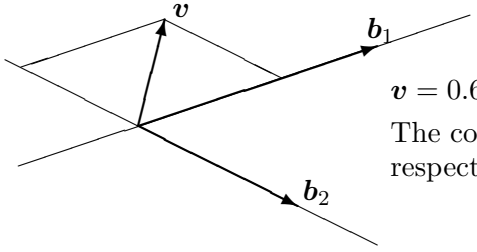
$$\left\{ \boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \boldsymbol{e}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \ldots, \boldsymbol{e}_{n-1} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{bmatrix}, \boldsymbol{e}_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \right\}$$

should be called the standard basis. In the vector space of polynomials of degree at most n, the basis

$$\{1, x, x^2, \ldots, x^n\}$$

is called the standard basis. — In contrast, in a vector space like $C^0[0,1]$, we are (in practice) not even able to explicitly write down a single basis, let alone identifying a particular basis as standard basis; this in spite of the fact that advanced abstract theory assures us that every vector space *has* a basis, even it eludes us to find one.

Coordinates: If V is a vector space and $S = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n\}$ a basis, then every vector $\boldsymbol{u}$ in V can be written in exactly one way as a linear combination of the basis vectors: $\boldsymbol{u} = k_1\boldsymbol{b}_1 + k_2\boldsymbol{b}_2 + \ldots + k_n\boldsymbol{b}_n$. The row $(k_1, k_2, \ldots, k_n)$ is called the *coordinate vector* of $\boldsymbol{u}$ with respect to the basis S, and the numbers k_i are called the coordinates. Clearly, for the vector space $\mathbb{R}^n$, if we choose the standard basis $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$, the coordinate vector of $\boldsymbol{u}$ with respect to the *standard* basis is $\boldsymbol{u}$ itself. General bases correspond to oblique coordinate systems.



$\boldsymbol{v} = 0.6\,\boldsymbol{b}_1 - 0.73\,\boldsymbol{b}_2$

The coordinate vector of $\boldsymbol{v}$ with respect to $\{\boldsymbol{b}_1, \boldsymbol{b}_2\}$ is $(0.6, -0.73)$.

Solution space: The set of those $\boldsymbol{x} \in \mathbb{R}^n$ that solve a *homogeneous* linear equation $A\boldsymbol{x} = \boldsymbol{0}$ (with an $m \times n$ matrix A) is a linear subspace of $\mathbb{R}^n$, called the *solution space*

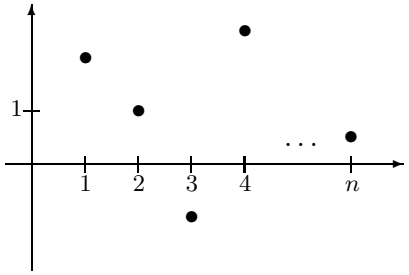
of the homogeneous linear equation. (Synonym: the null space of the matrix A)

Solution space of a linear ODE: /* The set of those $f \in C^2(\mathbb{R})$ that solve a *homogeneous* linear ordinary differential equation

$$y'' + p(x)y' + q(x)y = 0 \qquad (\ y = f(x)\ )$$

is a linear subspace of $C^2(\mathbb{R})$. The same applies, with obvious modifications, to linear ODEs of higher order than 2 –or of first order–, and even to systems of linear ODEs. In this context, a basis of the solution space is often called a fundamental set of solutions. This deviation in vocabulary is due to the simple fact that historically, the theory of linear ODEs is older than the unifying language of modern linear algebra. */M231

$\mathbb{R}^n$ as a vector space of functions: The main use of this view of $\mathbb{R}^n$ is so you can better appreciate the analogy that enticed us to put $\mathbb{R}^n$ and $C^0[a,b]$ into the same category "vector space". Instead of graphing vectors in $\mathbb{R}^2$ as arrows in the plane, vectors in $\mathbb{R}^3$ as arrows in space, and bailing out in despair for vectors in $\mathbb{R}^4$ etc., you can graph vectors in any $\mathbb{R}^n$ like a function defined only for $x = 1, 2, \ldots, n$, as can be seen in the following example:



The vector $[2,\ 1,\ -1,\ \frac{5}{2},\ \ldots,\ \frac{1}{2}]^T$ in $\mathbb{R}^n$ graphed as a function v with $v(1) = 2$, $v(2) = 1$, $v(3) = -1$, $v(4) = \frac{5}{2}$, ..., $v(n) = \frac{1}{2}$ (and $v(x)$ not def'd elsewhere).

Row space of a matrix: The rows of an $m \times n$ matrix are vectors in $\mathbb{R}^n$. Their span is a subspace of $\mathbb{R}^n$ (possibly all of $\mathbb{R}^n$), called the row space of this matrix.

Column space of a matrix: The columns of an $m \times n$ matrix are vectors in $\mathbb{R}^m$. Their span is a subspace of $\mathbb{R}^m$ (possibly all of $\mathbb{R}^m$), called the column space of this matrix. The column space of A is exactly the set of those $\boldsymbol{b}$ for which the linear system $A\boldsymbol{x} = \boldsymbol{b}$ has a solution $\boldsymbol{x}$. If this linear system has more than one solution, say $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$, then their difference $\boldsymbol{x}_1 - \boldsymbol{x}_2$ is in the null space of A:

Null space of a matrix: For an $m \times n$ matrix A, the solution space of the homogeneous linear equation $A\boldsymbol{x} = \boldsymbol{0}$ (a subspace of $\mathbb{R}^n$) is called the null space of A.

Superposition principle: (See already p. 6) Any linear combination of solutions of the homogeneous equation $A\boldsymbol{x} = \boldsymbol{0}$ is also a solution.

The difference of any two solutions $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ of an *in*homogeneous equation $A\boldsymbol{x} = \boldsymbol{b}$ solves the *homogeneous* equation $A\boldsymbol{x} = \boldsymbol{0}$. Therefore, if we know just a basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3, \ldots, \boldsymbol{v}_n\}$ of the solution space of the homogeneous equation $A\boldsymbol{x} = \boldsymbol{0}$ and one particular solution $\boldsymbol{x}_0$ of the inhomogeneous equation $A\boldsymbol{x} = \boldsymbol{b}$, then we obtain all solutions $\boldsymbol{x}$ of the inhomogeneous equation as all linear combinations $\boldsymbol{x} = \boldsymbol{x}_0 + c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + c_3\boldsymbol{v}_3 + \ldots + c_n\boldsymbol{v}_n$.

/* This principle is of course also valid for linear ordinary differential equations, with a function f playing the role of the vector $\boldsymbol{x}$, and an ODE like e.g. $y'' + p(x)y' + q(x)y = 0$, ($y = f(x)$) instead of $A\boldsymbol{x} = \boldsymbol{0}$, and an inhomogeneous ODE like $y'' + p(x)y' + q(x)y = b(x)$ instead of $A\boldsymbol{x} = \boldsymbol{b}$. In ODEs, the name *general solution* is used for the *set* of all solutions. */M231

Bases for row, null, and column space of a matrix: Elementary row operations do not change the row space of a matrix, nor its null space.

Therefore it is possible to obtain a basis of the row space of a matrix A by reducing it to row echelon form. The non-zero rows of the row echelon form constitute a basis of the row space of A. — It is also true that an appropriate subset of the rows of the original matrix A can serve as a basis for the row space; but except for cases of successful eyeballing, it is tedious to determine (without echelon form) which selection of rows qualifies as a basis.

As the solution space of a system of linear equations can be readily read off from the row echelon form, the same calculation determines the null space as well.

However, elementary row operations do change the column space. Nevertheless, they preserve one crucial property of columns: If a subset of columns like e.g. $\{\boldsymbol{c}_2, \boldsymbol{c}_4, \boldsymbol{c}_5\}$ of a matrix A is linearly independent, then the set of new columns $\{\boldsymbol{c}'_2, \boldsymbol{c}'_4, \boldsymbol{c}'_5\}$ arising from them under elementary row operations is also linearly independent. And if a subset of columns like e.g. $\{\boldsymbol{c}_2, \boldsymbol{c}_4, \boldsymbol{c}_5\}$ of a matrix A spans the column space of A, then corresponding set of new columns $\{\boldsymbol{c}'_2, \boldsymbol{c}'_4, \boldsymbol{c}'_5\}$ spans the column space of the new matrix. Therefore by selecting in the *original* matrix A those columns in which the leading 1's occur in the row echelon form, one obtains a basis of the column space of A.

Rank: From the practical method of calculating bases for the row and column space, it is clear that the dimension of the row space and the dimension of the column space of a matrix A are the same; namely they are the number of leading 1's in the row echelon form. This common number is called the *rank* of the matrix. Sometimes the fact that row space and column space have the same dimension is referred to by the phrase "row rank equals column rank".

Since for any matrix A, the row space of A is the column space of A^T (and vice versa), A and A^T have the same rank.

Nullity: The dimension of the null space of a matrix is called its *nullity*. Nullity is the ugliest word in the theory of linear algebra.

Dimension theorem: For any matrix A,

$$\text{rank}(A) + \text{nullity}(A) = \text{number of columns of } A$$

This is clear, because the rank is the number of leading 1's in row echelon form and the nullity is the number of columns without leading 1's in the row echelon form, and there can be at most a single leading 1 in each column.

Inner product: An inner product on a real vector space V is a function $\langle \cdot, \cdot \rangle$ that assigns to any two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in V a real number $\langle \boldsymbol{u}, \boldsymbol{v} \rangle$ in such a way that the following rules (familiar from the dot product in $\mathbb{R}^n$) hold: For all vectors $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w}$

and all scalars k,

$\langle \boldsymbol{u}, \boldsymbol{v} \rangle = \langle \boldsymbol{v}, \boldsymbol{u} \rangle$
$\langle \boldsymbol{u} + \boldsymbol{v}, \boldsymbol{w} \rangle = \langle \boldsymbol{u}, \boldsymbol{w} \rangle + \langle \boldsymbol{v}, \boldsymbol{w} \rangle$
$\langle k\boldsymbol{u}, \boldsymbol{v} \rangle = k\langle \boldsymbol{u}, \boldsymbol{v} \rangle$
$\langle \boldsymbol{u}, \boldsymbol{u} \rangle > 0$ unless $\boldsymbol{u} = \boldsymbol{0}$ (in which case the inner product equals 0)

A real vector space with an inner product is called a real inner product space.

/* Disclaimer: Whereas so far everything about real vector spaces could have been done for complex vector spaces exactly alike, the definition of the inner product requires modifications if you use complex numbers. This course does not cover the subject of inner products in complex vector spaces, so you needn't worry about this now; however if you major in mathematics, you may encounter complex inner product spaces; if you major in physics you ought to encounter them: they pervade quantum physics */

Norm: An inner product gives rise to a norm just as the dot product in $\mathbb{R}^n$ did: $\|\boldsymbol{u}\| := \sqrt{\langle \boldsymbol{u}, \boldsymbol{u} \rangle}$. The properties of a norm are given on page 17 already and are just repeated here for convenience:

(a) $\|\boldsymbol{u}\| \geq 0$ with $\|\boldsymbol{u}\| = 0$ if and only if $\boldsymbol{u} = \boldsymbol{0}$
(b) $\|k\boldsymbol{u}\| = |k| \, \|\boldsymbol{u}\|$
(c) $\|\boldsymbol{u} + \boldsymbol{v}\| \leq \|\boldsymbol{u}\| + \|\boldsymbol{v}\|$

/* Not every norm $\|\cdot\|$, as defined by these properties, originates as $\|\boldsymbol{u}\| := \sqrt{\langle \boldsymbol{u}, \boldsymbol{u} \rangle}$ from some inner product; but for those norms that do originate from an inner product, this inner product can be reconstructed in terms of the norm only, by the polarization formula $\boldsymbol{u} \cdot \boldsymbol{v} = \frac{1}{4} \left(\|\boldsymbol{u} + \boldsymbol{v}\|^2 - \|\boldsymbol{u} - \boldsymbol{v}\|^2 \right)$ */

Distance: Every norm gives rise to a distance $d(\boldsymbol{u}, \boldsymbol{v}) := \|\boldsymbol{u} - \boldsymbol{v}\|$ between two vectors. The characteristic properties of a distance have been listed on p. 18 already and will not be repeated here.

Examples of inner products:
(a) In $\mathbb{R}^n$, the dot product.
(b) Weighted dot products in $\mathbb{R}^n$, e.g. in $\mathbb{R}^4$: $\langle \boldsymbol{u}, \boldsymbol{v} \rangle := 2u_1v_1 + 5u_2v_2 + \frac{1}{2}u_3v^3 + 3u_4v_4$; any positive numbers can be chosen as coefficients.
(c) For an invertible $n \times n$ matrix A, $\langle \boldsymbol{u}, \boldsymbol{v} \rangle := (A\boldsymbol{u}) \cdot (A\boldsymbol{v}) = \boldsymbol{v}^T A^T A \boldsymbol{u}$ is an inner product.
(d) On $C^0[a, b]$, we have the inner product $\langle f, g \rangle := \int_a^b f(x)g(x) \, dx$
(e) On $\mathbb{R}^{m \times n}$, we use the trace to get an inner product $\langle A, B \rangle := \operatorname{tr}(A^T B)$. — Note that $\operatorname{tr}(AB)$ (without the transpose) is *not* an inner product. If $m \neq n$, the expression doesn't even make sense; if $m = n$, the expression does make sense, but $\operatorname{tr}(AA)$ may be negative.

The Cauchy Schwarz inequality: For any inner product, the following inequality holds (and is called the Cauchy Schwarz inequality):

$$|\langle \boldsymbol{u}, \boldsymbol{v} \rangle| \leq \|\boldsymbol{u}\| \, \|\boldsymbol{v}\|$$

It has a very simple but also quite tricky proof: Since $\langle t\boldsymbol{u}+\boldsymbol{v}, t\boldsymbol{u}+\boldsymbol{v}\rangle \geq 0$ for every number t, and since we can calculate $\langle t\boldsymbol{u}+\boldsymbol{v}, t\boldsymbol{u}+\boldsymbol{v}\rangle = t^2\langle \boldsymbol{u},\boldsymbol{u}\rangle + 2t\langle \boldsymbol{u},\boldsymbol{v}\rangle + \langle \boldsymbol{v},\boldsymbol{v}\rangle$, we have a quadratic polynomial in t that is never negative. Therefore, if you try to find zeros t_1, t_2 with the quadratic formula, the term under the square root in the quadratic formula must be ≤ 0. This latter condition is the Cauchy Schwarz inequality.

Orthogonality: Vectors in an inner product space are called orthogonal (to each other), if their inner product is 0. The zero vector $\mathbf{0}$ is orthogonal to every other vector. — Since there are many choices of an inner product on a vector space, one needs to specify: orthogonality *with respect to which inner product*: $[1,\, 2]^T$ and $[1,\, -1]^T$ are orthogonal with respect to the inner product $\langle \boldsymbol{u},\boldsymbol{v}\rangle := 2u_1v_1 + u_2v_2$, but they are not orthogonal with respect to the dot product.

You may omit to mention the inner product explicitly, (only) if it is clear from the context. In case of any doubt: state it.

Angles: Choosing an inner product in a vector space amounts to defining an angle measurement. The "angle" φ between vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in an inner product space is *defined* by the equality $\cos\varphi = \langle \boldsymbol{u},\boldsymbol{v}\rangle/(\|\boldsymbol{u}\|\,\|\boldsymbol{v}\|)$. The usual angle measurement in $\mathbb{R}^2$ or $\mathbb{R}^3$ is the one defined by the dot product.

Orthogonal complement: Given a linear subspace W in a vector space V, you can consider the collection of all those vectors $\boldsymbol{u}$ in V that are orthogonal to each vector in W. The collection of these vectors $\boldsymbol{u}$ forms a linear subspace of V itself; this linear subspace is called the *orthogonal complement* of W, in symbols $W^\perp$. For instance, in $\mathbb{R}^3$ (with the dot product), let W be the space spanned by $[0,\, 1,\, 2]^T$ and $[1,\, 1,\, 0]^T$. Then $W^\perp$ is the vector space spanned by $[-2,\, 2,\, -1]^T$.

A synonym for 'orthogonal complement' is *orthocomplement.*

Facts about orthogonal complements: Let V be an inner product space and W a subspace. Then
• $W \cap W^\perp = \{\mathbf{0}\}$
• $(W^\perp)^\perp = W$
• Every vector $\boldsymbol{u}$ in V can be written in exactly one way as a sum $\boldsymbol{u} = \boldsymbol{w}_1 + \boldsymbol{w}_2$ such that $\boldsymbol{w}_1$ is in W and $\boldsymbol{w}_2$ is in $W^\perp$.
• $\dim W + \dim W^\perp = \dim V$.

Fundamental spaces of a matrix, and orthogonality: The null space and the row space of an $m \times n$ matrix A are each other's orthogonal complement with respect to the dot product in $\mathbb{R}^n$.

The column space of an $m \times n$ matrix A and the null space of A^T are each other's orthogonal complement with respect to the dot product in $\mathbb{R}^m$.

This second statement (and its natural generalizations to other vector spaces than $\mathbb{R}^m$) is a very fundamental principle in more advanced treatments and applications of linear algebra. `/*` and of functional analysis `*/`

Orthogonal and orthonormal bases: A basis of an inner product space is called orthogonal, if any two vectors in this bases are orthogonal to each other (i.e., have

inner product 0). If moreover, each vector in the basis has length 1, then the basis is called orthonormal.

The relevance of such bases lies in the simplicity of practical calculations with them, as compared to non-orthogonal bases.

Projections: Let W be a subspace of an inner product space V, and $W^\perp$ its orthocomplement. How can we practically decompose a vector $\boldsymbol{v}$ into a sum of one vector in W (we'll call it $\mathrm{proj}_W \boldsymbol{v}$), and one vector in $W^\perp$ (we'll call that vector $\mathrm{proj}_{W^\perp} \boldsymbol{v}$)?

If W is given in terms of a basis $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m\}$, then $\mathrm{proj}_W \boldsymbol{v}$, being in W, must be a linear combination $k_1\boldsymbol{u}_1 + k_2\boldsymbol{u}_2 + \ldots + k_m\boldsymbol{u}_m$ of the basis vectors. The difference $\mathrm{proj}_{W^\perp} \boldsymbol{v} = \boldsymbol{v} - (k_1\boldsymbol{u}_1 + k_2\boldsymbol{u}_2 + \ldots + k_m\boldsymbol{u}_m)$ must be orthogonal to each of the basis vectors $\boldsymbol{u}_i$. This gives a linear system of m equations for the m unknowns $k_1, k_2, \ldots, k_m$:

$$\begin{array}{c} \langle \boldsymbol{u}_1, \boldsymbol{u}_1\rangle k_1 + \langle \boldsymbol{u}_1, \boldsymbol{u}_2\rangle k_2 + \ldots + \langle \boldsymbol{u}_1, \boldsymbol{u}_m\rangle k_m = \langle \boldsymbol{u}_1, \boldsymbol{v}\rangle \\ \langle \boldsymbol{u}_2, \boldsymbol{u}_1\rangle k_1 + \langle \boldsymbol{u}_2, \boldsymbol{u}_2\rangle k_2 + \ldots + \langle \boldsymbol{u}_2, \boldsymbol{u}_m\rangle k_m = \langle \boldsymbol{u}_2, \boldsymbol{v}\rangle \\ \vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots \\ \langle \boldsymbol{u}_m, \boldsymbol{u}_1\rangle k_1 + \langle \boldsymbol{u}_m, \boldsymbol{u}_2\rangle k_2 + \ldots + \langle \boldsymbol{u}_m, \boldsymbol{u}_m\rangle k_m = \langle \boldsymbol{u}_m, \boldsymbol{v}\rangle \end{array}$$

It can be shown that the coefficient matrix will be invertible if $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m\}$ is a basis of W. Better however, if the basis is orthogonal, the coefficient matrix is diagonal, and if the basis is orthonormal, the coefficient matrix is the identity matrix! Therefore, if $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m\}$ is an orthogonal basis (OGB) or an orthonormal basis (ONB) of W respectively

$$\begin{array}{ll} \mathrm{proj}_W \boldsymbol{v} = \dfrac{\langle \boldsymbol{v}, \boldsymbol{u}_1\rangle}{\langle \boldsymbol{u}_1, \boldsymbol{u}_1\rangle}\boldsymbol{u}_1 + \dfrac{\langle \boldsymbol{v}, \boldsymbol{u}_2\rangle}{\langle \boldsymbol{u}_2, \boldsymbol{u}_2\rangle}\boldsymbol{u}_2 + \ldots + \dfrac{\langle \boldsymbol{v}, \boldsymbol{u}_m\rangle}{\langle \boldsymbol{u}_m, \boldsymbol{u}_m\rangle}\boldsymbol{u}_m & \text{(OGB)} \\ \mathrm{proj}_W \boldsymbol{v} = \langle \boldsymbol{v}, \boldsymbol{u}_1\rangle\boldsymbol{u}_1 + \langle \boldsymbol{v}, \boldsymbol{u}_2\rangle\boldsymbol{u}_2 + \ldots + \langle \boldsymbol{v}, \boldsymbol{u}_m\rangle\boldsymbol{u}_m & \text{(ONB)} \\ \mathrm{proj}_{W^\perp} \boldsymbol{v} = \boldsymbol{v} - \mathrm{proj}_W \boldsymbol{v} & \end{array}$$

Gram–Schmidt orthogonalization procedure: This algorithm serves the purpose to obtain an orthogonal, or orthonormal, basis $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m\}$ from a given basis $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_m\}$:

First let $\boldsymbol{u}_1 := c_1\boldsymbol{v}_1$ with the number c_1 chosen such that $\boldsymbol{u}_1$ has the desired length ($\boldsymbol{u}_1 := \boldsymbol{v}_1/\|\boldsymbol{v}_1\|$, if an ONB is desired). Let W_1 be the vector space spanned by $\{\boldsymbol{u}_1\}$, equal the vector space spanned by $\{\boldsymbol{v}_1\}$.

Next decompose $\boldsymbol{v}_2 = \mathrm{proj}_{W_1} \boldsymbol{v}_2 + \mathrm{proj}_{W_1^\perp} \boldsymbol{v}_2$, according to the above projection formula; let $\boldsymbol{u}_2 = c_2 \mathrm{proj}_{W_1^\perp} \boldsymbol{v_2}$ with the number c_2 chosen such that $\boldsymbol{u}_2$ has the desired length. Let $W_2 = \mathrm{span}\{\boldsymbol{u}_1, \boldsymbol{u}_2\} = \mathrm{span}\{\boldsymbol{v}_1, \boldsymbol{v}_2\}$.

Decompose $\boldsymbol{v}_3 = \mathrm{proj}_{W_2} \boldsymbol{v}_3 + \mathrm{proj}_{W_2^\perp} \boldsymbol{v}_3$, and let $\boldsymbol{u}_3 = c_3 \mathrm{proj}_{W_2^\perp} \boldsymbol{v_3}$, etc.

This algorithm shows that we can always obtain an ONB in an inner product space. However, the algorithm is not recommended for practical large scale implementation on a computer, because the ever present roundoff errors in each step may accumulate worse than to the unavoidable extent.

QR decomposition: Every square matrix A can be written as the product $A = QR$, where Q is an orthogonal matrix and R is an upper triangular matrix. For invertible A, this can be seen as a consequence of the Gram–Schmidt algorithm: Take the columns of A, $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$, as the basis vectors of the column space W of A. Let $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m\}$, be the ONB obtained by Gram–Schmidt from $\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_m\}$. Then the QR decomposition is explicit in components as follows:

$$\begin{bmatrix} | & & | \\ \boldsymbol{v}_1 & \cdots & \boldsymbol{v}_m \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ \boldsymbol{u}_1 & \cdots & \boldsymbol{u}_m \\ | & & | \end{bmatrix} \begin{bmatrix} \langle \boldsymbol{v}_1, \boldsymbol{u}_1 \rangle & \langle \boldsymbol{v}_2, \boldsymbol{u}_1 \rangle & \ldots & \langle \boldsymbol{v}_m, \boldsymbol{u}_1 \rangle \\ 0 & \langle \boldsymbol{v}_2, \boldsymbol{u}_2 \rangle & \ldots & \langle \boldsymbol{v}_m, \boldsymbol{u}_2 \rangle \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \langle \boldsymbol{v}_m, \boldsymbol{u}_m \rangle \end{bmatrix}$$

Practical calculation of a QR decomposition is via algorithms different from Gram–Schmidt. /* Probably the most important usage of the QR decomposition is the so called QR algorithm for practical calculation of eigenvalues. */ Numerics or advanced lin' alg

Orthogonal matrices: (cf. rotations in n-space p. 19) A matrix Q is called orthogonal, if $Q^T Q = I$. Since this implies $Q^T = Q^{-1}$, it also follows $QQ^T = I$. A matrix is orthogonal if and only if its columns are mutually orthogonal and have each length 1. Equivalently: if its rows are mutually orthogonal and have each length 1. Orthogonal matrices have determinant $+1$ or -1. Inverses, transposes, and products of orthogonal matrices are orthogonal. A matrix Q is orthogonal if and only if the linear transformation $\boldsymbol{x} \mapsto Q\boldsymbol{x}$ preserves the euclidean norm (i.e., lengths); and also if and only if this transformation preserves the dot product (i.e., angles): $\|Q\boldsymbol{x}\| = \|\boldsymbol{x}\|$; $(Q\boldsymbol{x}) \cdot (Q\boldsymbol{y}) = \boldsymbol{x} \cdot \boldsymbol{y}$.

Eigenvectors and eigenvalues; update: see p. 21 for the basic facts. — One modification: We will now accept complex eigenvalues, i.e., complex solutions λ to the characteristic equation $\det(A - \lambda I) = 0$. If λ is a complex (but not real) number and A is a matrix with real entries, then nontrivial solutions of the linear system $A\boldsymbol{v} = \lambda\boldsymbol{v}$ are only possible if we also permit complex entries for $\boldsymbol{v}$. This poses no problem whatsoever, algebraically. To the contrary, any deeper discussion on eigenvalues gets simplified if we admit complex numbers. The only flipside is that the geometric intuition for vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ is no longer available.

Survey of complex vector spaces: (FYI only; not course or exam contents; but do read it, lest the last few sections appear half-baked and incoherent.)

When I say 'complex number' without extra qualification, this *includes* real numbers. For instance, $1 + i$, 3, $-5i$, $1/\pi$, $3 - \sqrt{2}i$ *all* are complex numbers, and two among them (3 and $1/\pi$) happen to be real numbers as well.

We defined matrices as rectangular arrays of real numbers, and defined addition and multiplication of matrices. Let us now accept also complex numbers as entries of matrices, retaining the very same definitions for addition and multiplication of matrices, the same procedures (row operations) for solving linear equations. The same rules of matrix algebra are still valid.

Determinants of matrices with complex entries continue to obey the very same rules as we had for real entries.

The geometric vector spaces (arrows in the plane or in space) are models for the *real* vector spaces $\mathbb{R}^2$ and $\mathbb{R}^3$ respectively. We don't have such models for $\mathbb{C}^2$, $\mathbb{C}^3$, or any $\mathbb{C}^n$, which are the sets of column vectors with complex entries. But we can do without these geometric examples.

The transpose of a matrix with complex entries is still defined as it was in the case of real entries, and obeys the same rules; however it will be a much less significant concept, in contrast to its importance for matrices with real entries. — For a matrix A with complex entries, A^* is defined as the complex conjugate of A^T; i.e., in A^T replace every entry with its complex conjugate. In case you forgot, the complex conjugate of $a+bi$ is $a-bi$: change all the i's into $-i$'s to get the complex conjugate. The complex conjugate of a *real* number is that same real number, because there are no i's to change into $-i$. Therefore, for matrices with real entries A^* and A^T is the same. — A^* obeys similar calculational rules as A^T; for instance $(AB)^* = B^*A^*$. All the *useful* services of A^T will be performed by A^*, when entries are permitted to be complex.

Complex vector spaces function with the same rules as real vector spaces; only the scalars are now complex numbers. — However, norms and inner products must be defined differently, and we will not discuss here, how.

When admitting complex entries, symmetric matrices (those that satisfy $A^T = A$) and orthogonal matrices (those that satisfy $A^TA = I$) suffer the same fate of insignificance as the transpose does; however matrices satisfying $A^* = A$ or $A^*A = I$ respectively take over their role, and I refrain from introducing the names for them.

In conclusion, the only reason we have omitted complex numbers all the way till now is that we cater for students who are uneasy with them: Think you take a crash course in practical arithmetic with complex numbers until you feel perfectly comfortable with them, then I can fill all the hwk and exam problems (with the above-mentioned exceptions) with complex numbers instead of real numbers, and you'll do them with just as much ease or difficulty as you did them with real numbers.

As soon as you get serious about eigenvalues and eigenvectors, you *want* complex numbers; it is only because we do just a bit of eigenvalues and eigenvectors that the complex numbers show up only briefly.

Characteristic polynomial; How many eigenvalues? Given an $n \times n$ matrix A, an eigenvalue λ is a solution to the equation $\det(\lambda I - A) = 0$. Now $\det(\lambda I - A)$ is a polynomial in λ of degree n (called the *characteristic polynomial*) of A; therefore the equation $\det(\lambda I - A) = 0$ has *at most* n *real* solutions; if n is odd, it has at least one real solution (because of the intermediate value theorem from calculus), but if n is even, it may not have any real solutions.

If you admit complex solutions (see next entry), there will still be at most n of them, but always at least one. If you *count solutions with their multiplicity*, there will be *exactly* n *complex* solutions (some of which may be real, and they need not all be distinct).

Fundamental theorem of algebra: Every polynomial $\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0$ of degree n with any choice of coefficients $a_0, a_1, \ldots, a_{n-1}$ (real or complex)

can be written as a product of n linear factors:

$$\lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0 = (\lambda - \lambda_1)(\lambda - \lambda_2)\cdots(\lambda - \lambda_n)$$

for appropriate complex numbers $\lambda_1, \ldots, \lambda_n$, which need not all be distinct. If the coefficients $a_0, a_1, \ldots, a_{n-1}$ are real numbers, then the complex conjugate of each λ_i is also among the $\lambda_1, \ldots, \lambda_n$, and is repeated in the list just as many times as λ_i is.

This result is not concerned with the issue how to find the λ_i in practice. For quadratic polynomials, take the quadratic formula, for degree 3 and 4, there are similar (cubic and quartic) formulas that are however too unwieldy for practical use. If you encounter a polynomial of degree 3 or higher and are asked to find its zeros $\lambda_1, \ldots, \lambda_n$, you must resort to numerical methods, unless you can guess some zeros by eyeballing, and then factor them off by long division of polynomials.

Algebraic multiplicity of an eigenvalue: Let

$$\det(\lambda I - A) = \lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0$$

be the charactistic polynomial of the $n \times n$ matrix A. It can be factorized as given above, using the fundamental theorem of algebra. The zeros λ_1 of the characteristic polynomial are the eigenvalues of A, and the algebraic multiplicity of an eigenvalue λ_i is the number how often λ_i occurs in the list $\lambda_1, \ldots, \lambda_n$.

Eigenspaces; Geometric multiplicity of an eigenvalue: Suppose λ_i is an eigenvalue of the $n \times n$ matrix A. Then the space of all vectors $\boldsymbol{v}$ satisfying the linear system $(A - \lambda I)\boldsymbol{v} = \mathbf{0}$ is called the eigenspace of A for the eigenvalue λ_i. Its dimension is called the geometric multiplicity of the eigenvalue λ_i.

In this course, we only discuss the geometric multiplicity of *real* eigenvalues (and consequently look for eigenvectors $\boldsymbol{v} \in \mathbb{R}^n$, i.e., with real entries). If we were to discuss the geometric multiplicity of complex eigenvalues that are not real, we would have to admit eigenvectors $\boldsymbol{v}$ with complex entries and thus redo the whole chapter on real vector spaces to accomodate complex vector spaces.

Geometric vs. algebraic multiplicity: The algebraic multiplicity of an eigenvalue is at least as large as the geometric multiplicity. Depending on the matrix and on the individual eigenvalue, it may be equal or larger. Except for the knowledge that the geometric multiplicity of an eigenvalue is less or equal the algebraic multiplicity, information on the geometric multiplicity cannot be obtained from the characteristic polynomial; the full matrix is needed for such information.

A matrix is *diagonalizable* (see below) if and only if the geometric multiplicity of each eigenvalue is equal to its algebraic multiplicity.

Trace and determinant in the characteristic polynomial: Let

$$\det(\lambda I - A) = \lambda^n + a_{n-1}\lambda^{n-1} + \ldots + a_1\lambda + a_0$$

be the charactistic polynomial of the $n \times n$ matrix A. The coefficient a_0 equals $(-1)^n \det A$, and the coefficient a_{n-1} equals $-\operatorname{tr} A$. In other words, if you include complex eigenvalues and count them according to their algebraic multiplicity, the

sum of all eigenvalues of a matrix is its trace, and the product of all eigenvalues is the determinant.

In particular: for a 2×2 matrix, the characteristic polynomial is $\lambda^2-(\operatorname{tr} A)\lambda+(\det A)$.

Basis of eigenvectors? — The diagonalization problem: The problem: "Given an $n \times n$ matrix A, does there exist a basis of $\mathbb{R}^n$ consisting only of eigenvectors of A?" is of fundamental importance. We'll see soon that for *symmetric* matrices A, the answer is "yes", but for some non-symmetric matrices A, the answer can well be "no".

Clearly, unless we are willing to modify the question such as to refer to "a basis of $\mathbb{C}^n$" (which we don't do in this course), a positive answer requires that all eigenvalues are real. But even if all eigenvalues are real, we still may or may not have a basis of $\mathbb{R}^n$ consisting only of eigenvectors of A.

There is no answer to the above question that is both simple and complete. `/*` unless go beyond this course... `*/` However, the following statement gives a complete (but not simple) answer:

$\mathbb{R}^n$ has a basis consisting only of eigenvectors of a given $n \times n$ matrix A, *if and only if A is diagonalizable*. To say that A is diagonalizable means that there is an invertible matrix P such that $P^{-1}AP =: D$ is a diagonal matrix. In this case, the diagonal entries of the diagonal matrix $P^{-1}AP$ *are* the eigenvalues of A (and also the eigenvalues of $P^{-1}AP$). To understand this relation, rewrite $P^{-1}AP = D$ as $AP = PD$ and denote by $\boldsymbol{p}_1$, $\boldsymbol{p}_2$, ..., $\boldsymbol{p}_n$ the columns of P: The equation

$$\begin{bmatrix} \\ A \\ \\ \end{bmatrix} \begin{bmatrix} | & | & & | \\ \boldsymbol{p}_1 & \boldsymbol{p}_2 & \cdots & \boldsymbol{p}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \boldsymbol{p}_1 & \boldsymbol{p}_2 & \cdots & \boldsymbol{p}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n \end{bmatrix}$$

amounts to the equations

$$A\boldsymbol{p}_1 = d_1\boldsymbol{p}_1\,, \quad A\boldsymbol{p}_2 = d_2\boldsymbol{p}_2\,, \ldots, \quad A\boldsymbol{p}_n = d_n\boldsymbol{p}_n$$

So, if you have a basis of eigenvectors, then you put them as columns in a matrix P, and this matrix will be invertible, because its columns are linearly independent; next you put the corresponding eigenvalues on the diagonal of a diagonal matrix D: then it will indeed be true that $AP = PD$ and hence $P^{-1}AP = D$.

Conversely, if you have matrices P (invertible) and D (diagonal) satisfying $P^{-1}AP = D$, you can read off the eigenvalues from the diagonal of D, and the corresponding eigenvectors will be the columns of P. They will be a basis of $\mathbb{R}^n$, because they are linearly independent (as columns of an invertible matrix always are) and there are as many of them as the dimension of $\mathbb{R}^n$ (namely n).

`/*` **Normal matrices:** If you go complex all the way, a matrix A is known to be diagonalizable, if A and A^* commute ($AA^* = A^*A$). In that case A is called a normal matrix `*/`M453, probably

Calculating powers of a matrix: You can always calculate $A^2 = AA$, $A^3 = AAA$, etc. — That takes 2536 matrix multiplications to evaluate A^{2537}. You can get

along with much less if you are smart and write $2537 = 2^{11}+2^8+2^7+2^6+2^5+2^3+1$, and calculate A, A^2, $A^4 = A^2A^2$, $A^8 = A^4A^4$, A^{16}, etc. as intermediate steps.

But there is one very different method that is of theoretical interest, and sometimes of practical interest, too; it works best for diagonalizable matrices (even though it can be generalized to non-diagonalizable matrices, too), and it relies on the following simple observation: For a *diagonal* matrix D, powers can be calculated very easily: D^{2537} is obtained by taking the 2537^{th} power of the diagonal elements.

If A is diagonalizable, we can write $A = PDP^{-1}$ and conclude

$$A^{2537} = (PDP^{-1})(PDP^{-1})\cdots(PDP^{-1}) = PDD\cdots DP^{-1} = PD^{2537}P^{-1} .$$

So we need to solve an eigenvalue problem, take powers of the eigenvalues and do two matrix multiplications and one matrix inversion.

Examples:
Eigenvectors to *different* eigenvalues are always linearly independent. Therefore, if an $n \times n$ matrix has n distinct (real) eigenvalues, then it is diagonalizable (in terms of real matrices P, D).

$\begin{bmatrix} 3 & 1 \\ 0 & 3 \end{bmatrix}$ has a double eigenvalue 3 (i.e., the algebraic multiplicity of the eigenvalue 3 is two). However, the eigenspace is $\text{span}\{\begin{bmatrix} 1 \\ 0 \end{bmatrix}\}$, so the geometric multiplicity of the eigenvalue 3 is one. As there are not enough linearly independent eigenvectors to span $\mathbb{R}^2$, this matrix is not diagonalizable.

A 2-dim rotation matrix $\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$ has complex eigenvalues $\cos\alpha + i\sin\alpha$ and $\cos\alpha - i\sin\alpha$. The issue of diagonalizing this matrix does not arise, unless you admit complex entries in matrices. In that case, however, the matrix is (complex) diagonalizable:

$$\begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} = \begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \cos\alpha + i\sin\alpha & 0 \\ 0 & \cos\alpha - i\sin\alpha \end{bmatrix} \begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix}^{-1}$$

/* Plugging a matrix into its characteristic polynomial: If you plug a matrix into its characteristic polynomial, the expression evaluates to the zero matrix. This is generally true (we don't prove it), but for *diagonalizable* matrices you can see why it is true, when you write the matrix powers as $A^n = PD^nP^{-1}$. – Note: Plugging a matrix A into a polynomial like, e.g., $\lambda^3 - 2\lambda^2 + 5\lambda - 7$ means to calculate $A^3 - 2A^2 + 5A - 7I$; you would probably have been stuck with the number 7 instead of "7 times the identity matrix", if I hadn't told you this. */M453

AB and BA have the same eigenvalues: You know already that for square matrices of the same size, $\text{tr}(AB) = \text{tr}(BA)$ and $\det(AB) = \det(BA)$, even though AB and BA are usually not the same. However, AB and BA do have the same characteristic polynomial, and therefore the same eigenvalues with the same algebraic multiplicities.

Eigenvectors for different eigenvalues are linearly independent: If λ_1, λ_2, $\ldots$, λ_k are *different* eigenvalues of a matrix A and $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\ldots$, $\boldsymbol{v}_k$ are eigenvectors for these eigenvalues respectively, then $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ is linearly independent.

The eigenvalue problem for symmetric matrices: The following theorem holds: If A is a symmetric matrix ($n \times n$), then all its eigenvalues are real. Their algebraic multiplicity is equal to their geometric multiplicity. Eigenvectors to different eigenvalues are orthogonal to each other (with respect to the euclidean inner product, aka dot product). It is possible to choose an *orthonormal* basis of $\mathbb{R}^n$ consisting of eigenvectors of A.

In other words: A symmetric matrix A can be written as $A = QDQ^{-1} = QDQ^T$ with an orthogonal matrix Q and a real diagonal matrix D.

/* **Proof of diagonalization of symmetric matrices:** Is beyond this course, but so interesting that it *must* be here for reference purposes: You solve the following multivariable calculus problem for variables x_1, x_2, $\ldots$, x_n (entries of the vector $\boldsymbol{x}$), by means of Lagrange multipliers: Find the minimum of $\boldsymbol{x}^T A\boldsymbol{x}$ under the constraint $\boldsymbol{x}^T\boldsymbol{x} = 1$. If you do it with Lagrange multipliers, it turns out that the minimum is an eigenvalue of A, and indeed it will turn out to be the smallest eigenvalue. And the $\boldsymbol{x}_*$ that yields the minimum is a corresponding eigenvector.

Next you solve the following multivariable calculus problem: Find the minimum of $\boldsymbol{x}^T A\boldsymbol{x}$ under the *two* constraints $\boldsymbol{x}^T\boldsymbol{x} = 1$ and $\boldsymbol{x}^T\boldsymbol{x}_* = 0$. If you do it with Lagrange multipliers, it turns out that now the minimum is another eigenvalue of A, and it will turn out to be the second smallest, and the new $\boldsymbol{x}$ that yields the minimum is a corresponding eigenvector. Continuing with more and more constraints yields the next eigenvalues.

This is probably the most prominent and most useful example for Lagrange multipliers. If you are a strong student in M241, make sure you try this example. They may omit to assign or to cover it, because not all of your peers in that course may have had matrices. */M241

/* **A weird theorem about eigenvalues:** If all entries of a square matrix are positive numbers, then this matrix has (at least) one positive eigenvalue. To understand why this is the case, you need Brouwer's fixed point theorem, which is usually covered in topology, not before the 400 level */

/* **On practical calculation of eigenvalues:** For large matrices, it is *not* efficient to calculate the characteristic polynomial and then find its zeros. Rather, the basic idea for the QR algorithm is as follows: To find the eigenvalues of A, decompose A into a product of an orthogonal and an upper triangular matrix: $A = QR$. Now let $A_1 := RQ$. A_1 is different from A, but it has the same eigenvalues. Decompose A_1 similarly: $A_1 = Q_1R_1$. Calculate $A_2 := R_1Q_1$ and continue that way. Amazingly, A_n converges to a diagonal matrix as $n \to \infty$, and the diagonal entries of the limit will be the eigenvalues. A number of bells and whistles needs to be attached for speed of calculation, but basically this is the idea. For particular types of matrices, other, yet better, algorithms exist. */Numerical linear algebra

— THE END —