Basics of linear algebra in a nutshell

1 Vector spaces, base, coordinates

We start with a brief and intuitive summary of linear algebra principles, i. e. summary about vector spaces. See materials of basic course of linear algebra for more information. Vector space V (or linear space, it is the same) is a set of *vectors*. The vector $\vec{u} \in V$ is an abstract object which can be "scaled" by arbitrary scalar $s \in \mathbb{R}$ using scalar multiplication $s\,\vec{u} \in V$ and two vectors $\vec{u}, \vec{v} \in V$ can be "combined" using vector addition $\vec{u} + \vec{v} \in V$. These two operations must satisfy common (axiomatic) properties: commutative and associative law for addition, associative law for scalar multiplication, distributive laws and existence of the zero vector $\vec{o}$, which is the only one such vector unable to be scaled: $s\,\vec{o} = \vec{o}\ \forall s \in \mathbb{R}$. Vectors are abstract objects without any specific interpretation but there are many examples of vector spaces: vector space of functions, vector space of polynomials, vector space of matrices etc. We will deal with two specific vector spaces: geometrical vector space of free or bound vectors (see chapter 3 of [1]) and (of course) the $\mathbb{R}^n$ vector space of real n-tuples (written in columns). The scalar multiplication and vector addition in geometrical vector space V are performed "mechanically" (using rulers) as described in chapter 3 of [1]. On the other hand the scalar multiplication and vector addition in $\mathbb{R}^n$ are performed "numerically": $t\,[u_1, \ldots, u_n]^\top = [tu_1 \ldots, tu_n]^\top$ and $[u_1, \ldots, u_n]^\top + [v_1, \ldots, v_n]^\top = [u_1 + v_1, \ldots, u_n + v_n]^\top$.

One of the main results of the linear algebra says that if an abstract vector space V has a finite dimension n then there exist a mapping $V \to \mathbb{R}^n$ which is an isomorphism. It means that all operations executed with vectors in V can be executed with its representatives in $\mathbb{R}^n$ without loss of information. So the results computed numerically in $\mathbb{R}^n$ can be interpreted back in abstract linear space V. All results which are consequences of the two above mentioned operations in an abstract vector space have their corresponding results in $\mathbb{R}^n$. For example, we need not to deal with rulers and compasses in geometrical vector space in order to provide operations and features of geometrical vectors, we can transform these vectors to $\mathbb{R}^n$ and do these operations numerically (in computer, for example). Then the results can be interpreted back in the geometrical vector space, which serves for visualization of results. The maping $V \to \mathbb{R}^n$ mentioned here is well known: $\vec{u} \mapsto$ coordinates of the vector $\vec{u}$ w.r.t. a fixed chosen basis in V.

We remind what does mean basis in V and coordinates w.r.t. a basis. Let $\vec{u}_1, \ldots, \vec{u}_m$ be a collection of vectors from a vector space V. We are using the word "collection" (instead "set"), because the vectors here are ordered and maybe there are more instances of one vector. The vector space defines only two operations (scalar multiplication and vector addition), so when these operations are (repeatedly) applied to such collection of vectors, then the result cannot be nothing but *linear combination* of the vectors, i.e. the result is in the form $x_1\vec{u}_1 + \cdots + x_n\vec{u}_n$, where $x_i \in \mathbb{R}$ are scalars. All such results (with all possible scalars) fill a set $M \subset V$ o vectors called $\operatorname{span}(\vec{u}_1, \ldots, \vec{u}_n)$. We say that the collection of vectors $\vec{u}_1, \ldots, \vec{u}_m$ *generates* the set M. If it is possible to remove one vector from the collection without changing its span, then we call that such collection of vectors are *linearly dependent*. Otherwise, the collection of vectors is *linearly independent*. The maximal linear independent collection of vectors in V (it is the same as minimal collection of vectors which generate V) is called a *basis* of vector space V. It is possible to prove that all bases of the same vector space V have the same number of vectors. This number is called the *dimension* of V.

Let a basis $\beta = (\vec{b}_1, \ldots, \vec{b}_n)$ of vectors in a vector space V be given. For each $\vec{u} \in V$ there exists (because basis generates V) only one (because basis is linear independent) ordered set of scalars $x_i \in \mathbb{R}$ with the property

$$\vec{u} = x_1\vec{b}_1 + \cdots + x_n\vec{b}_n. \quad \text{This n-tuple of real numbers } x_i \text{ written in a column} \quad \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n \qquad (1)$$

is called *coordinates of $\vec{u}$ with respect to the basis β*. The notation of this n-tuple is $\vec{u}_\beta \in \mathbb{R}^n$. So the mapping $V \to \mathbb{R}^n$ mentioned above is realized by $\vec{u} \mapsto \vec{u}_\beta$. It is possible to prove that such mapping is an isomorphism.

2 Change of basis, matrices of such change

When a basis $\beta = (\vec{b}_1, \ldots, \vec{b}_n)$ in a vector space V is changed to another basis $\beta' = (\vec{b}'_1, \ldots, \vec{b}'_n)$ then the coordinates of a vector $\vec{u} \in V$ w.r.t. the basis β should be different than its coordinates w.r.t. the basis β', in brief $\vec{u}_\beta \neq \vec{u}_{\beta'}$. We will show how to find the matrix $\mathtt{A} \in \mathbb{R}^{n,n}$ which transforms $\vec{u}_\beta$ to $\vec{u}_{\beta'}$ using matrix product $\mathtt{A}\vec{u}_\beta = \vec{u}_{\beta'}$ for all $\vec{u} \in V$.

Let $[x_1, \ldots, x_n]^\top \in \mathbb{R}^n$ be a column of scalars and $\beta = (\vec{b}_1, \ldots, \vec{b}_n)$ be a basis of V. We introduce a very compact notation $(\beta)\,\vec{x}$ which denotes the linear combination of vectors $\vec{b}_i$ with coefficients x_i. The notation respects common rules of matrix product: one-row matrix (β) of *vectors* is multiplied by one-column matrix $[x_1, \ldots, x_n]$ of *scalars*. In more detail:

$$(\beta)\,\vec{x} = [\vec{b}_1, \ldots, \vec{b}_n] \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1\vec{b}_1 + \cdots + x_n\vec{b}_n.$$

Using this compact notation, we can say that $\vec{x} \in \mathbb{R}^n$ are coordinates of $\vec{u} \in V$ w.r.t. the basis β if and only if $\vec{u} = (\beta)\vec{x}$.

Let $\mathtt{A} \in \mathbb{R}^{n,n}$ be a matrix of scalars, $\mathtt{A} = [\vec{a}_1, \ldots, \vec{a}_n]$. It means that $\vec{a}_i$ is the i-th colum of matrix $\mathtt{A}$. Moreover, let $\beta = (\vec{b}_1, \ldots, \vec{b}_n)$ be a basis of V. Then a collection of vectors $\vec{u}_1, \ldots, \vec{u}_n$ such that $\vec{u}_i = (\beta)\vec{a}_i$, is denoted by $(\beta)\mathtt{A}$. This notation respects common rules of matrix multiplication: one-row matrix of vectors by $n \times n$ matrix of scalars gives one-row matrix of vectors.

Let $\beta = (\vec{b}_1, \ldots, \vec{b}_n)$ and $\beta' = (\vec{b}'_1, \ldots, \vec{b}'_n)$ be two bases of a linear space V. Then the *the matrix* $\mathtt{A}$ *transforming coordinates from basis* β *to* β' is defined by

$$(\vec{b}_1, \ldots, \vec{b}_n) = (\vec{b}'_1, \ldots, \vec{b}'_n)\mathtt{A} \tag{2}$$

or in a more compact form by $(\beta) = (\beta')\mathtt{A}$. The following is true for such a matrix $\mathtt{A}$:

(i) Matrix $\mathtt{A}$ exists for given bases β and β'. This matrix is unique.
(ii) The columns $\vec{a}_i$ of the matrix $\mathtt{A}$ include coordinates of $\vec{b}_i$ w.r.t. the basis β'.
(iii) The matrix $\mathtt{A} \in \mathbb{R}^{n,n}$ and it is a regular matrix.
(iv) $\vec{u}_{\beta'} = \mathtt{A}\vec{u}_\beta$

Note that the name of the matrix $\mathtt{A}$ is derived from the property (iv): for known coordinates of a vector $\vec{u} \in V$ w.r.t. the basis β it is possible to calculate coordinates of $\vec{u}$ w.r.t. the basis β' using the matrix $\mathtt{A}$. If a reverse transformation from β' to β is needed, then inverse matrix $\mathtt{A}^{-1}$ must be used. The property (ii) is direct consequence of equation (2). (ii) $\Rightarrow$ (i) because the coordinates exist and they are unique. The regularity in (iii) can be proven using argument that both collections of vectors in equation (2) are linearly independent. The proof the (iv) follows. Let $\vec{u} \in V$. Then $\vec{u} = (\beta)\vec{u}_\beta$ and $\vec{u} = (\beta')\vec{u}_{\beta'}$. From the associativite law of matrix multiplication we see that

$$(\beta')\vec{u}_{\beta'} = \vec{u} = (\beta)\vec{u}_\beta = ((\beta')\mathtt{A})\,\vec{u}_\beta = (\beta')(\mathtt{A}\vec{u}_\beta)$$

and we see the coordinates of vector $\vec{u}$ w.r.t. basis β' on both sides of the equation.

It is possible to do coordinate transformation via more bases in a given order of such bases. If we known matrices which do particular coordinate transformations, then their matrix product is the matrix of the composite coordinate transformation. More precisely, let there be bases β, γ, δ in a linear space V. Let there be matrices $\mathtt{A}$ and $\mathtt{B}$ with properties $\mathtt{A}\vec{u}_\delta = \vec{u}_\gamma$ and $\mathtt{B}\vec{u}_\gamma = \vec{u}_\beta$. Then the matrix product $\mathtt{BA}$ has the property $(\mathtt{BA})\vec{u}_\delta = \vec{u}_\beta$ and it is matrix transforming coordinates from basis δ to β. This is a simple consequence of associative law of matrix product: $(\mathtt{BA})\vec{u}_\delta = \mathtt{B}(\mathtt{A}\vec{u}_\delta) = \mathtt{B}\vec{u}_\gamma = \vec{u}_\beta$.

3 Matrix of linear transformation

Let $\beta = (\vec{b}_1, \ldots, \vec{b}_n)$ and $T : V \to V$ be a linear transformation. We define *matrix* $\mathtt{A}$ *of the linear transformation* T *w.r.t. the basis* β by equation

$$\big(T(\vec{b}_1), \ldots, T(\vec{b}_n)\big) = (\vec{b}_1, \ldots, \vec{b}_n)\mathtt{A}. \tag{3}$$

The following is true for such matrix $\mathtt{A}$ w.r.t. the basis β:

(i) The matrix $\mathtt{A}$ exists and it is unique for every given transformation T and basis β.
(ii) The columns $\vec{a}_i$ of the matrix $\mathtt{A}$ consist of the coordinates of $T(\vec{b}_i)$ w.r.t. the basis β.
(iii) $\mathtt{A} \in \mathbb{R}^{n,n}$.
(iv) The linear transformation T, defined by (3), uniquely exists for every given matrix $\mathtt{A}$ and basis β.
(v) $T(\vec{u})_\beta = \mathtt{A}\,\vec{u}_\beta$.

The property (ii) is a direct consequence of equation (3). (ii) $\Rightarrow$ (i) because the coordinates exist and they are unique. For proving (iv) we need to use the fact that if the values of a linear transformation T are known on the basic vectors then T is defined uniquely for each vector $\vec{u} \in V$. The consequence of (i) and (iv) is that for a given basis β there exist one to one mapping between linear transformations and their matrices. We omit the proof of (v). It is only technical and similar to the proof of the property $\vec{u}_{\beta'} = \mathtt{A}\,\vec{u}_\beta$.

We can conclude that for a matrix $\mathtt{A}$ transforming coordinates from basis β to β', there holds $(\beta) = (\beta')\mathtt{A}$ and it is also matrix $\mathtt{A}$ of a linear transformation T which is defined by $T(\vec{b}'_i) = \vec{b}_i$. Important note: the linear transformation T derived from matrix $\mathtt{A}$ works in reverse direction (from β' to β) than the transformation of coordinates using the same matrix $\mathtt{A}$.

We can write a short summary. Let β and β' be two bases in a vector space V. The following properties are equivalent.

- $\mathtt{A}$ transforms coordinates from β to β' by $\vec{u}_{\beta'} = \mathtt{A}\vec{u}_\beta$.
- $(\beta) = (\beta')\mathtt{A}$.
- $\mathtt{A}$ contains coordinates of $\vec{b}_i$ w.r.t. the basis β' in its columns $\vec{a}_i$.
- $\mathtt{A}$ is the matrix w.r.t. the basis β' of a linear transformation T with given values $T(\vec{b}'_i) = \vec{b}_i$.

Note that it is possible to construct composite linear transformations $T_2 \circ T_1 : V \to V$ defined by the rule $(T_2 \circ T_1)(\vec{u}) = T_2\big(T_1(\vec{u})\big)$. If $\mathtt{A}_2$ and $\mathtt{A}_1$ are corresponding matrices of T_2 and T_1 w.r.t. a basis β and $(T_2 \circ T_1)(\vec{u}) = \vec{v}$, then $\mathtt{A}_2\mathtt{A}_1\vec{u}_\beta = \vec{v}_\beta$. Roughly speaking: composite transformations are represented by matrix product of the corresponding matrices.

There are elementary transformations in geometrical vector spaces: rotation, scaling, slanting, reflection. It can be proven that a general linear transformation is a composition of the mentioned elementary transformations. Each elementary transformation has its very specific matrix. A general linear transformation can be represented by a matrix product of such specific matrices. We can *imagine* or *visualize* such transformations in geometrical vector space and we can *compute* them in the vector space $\mathbb{R}^n$ of their coordinates using matrices of such transformation.

4 Dot product

It is impossible to measure lengths of vectors and angles between them only by scalar multiplication and vector addition. Sizes and angles are very natural quantities in geometrical vector spaces. This is the reason why the dot product of two vectors as a new operation in vector space, is introduced.

We define the dot product in geometrical vector space V using geometrical tools only (rulers with a scale and protractor). Main features will be shown. We keep in geometrical vector space during such thinking. Then we show how to generalize this idea for abstract vector spaces where does not exist ruler nor protractor. Finally, we will show how the dot product can be *computed* in $\mathbb{R}^n$ and what is the relationship between dot product in geometrical vector space and dot product in $\mathbb{R}^n$.

The size of each vector $\vec{u} \in V$ in a geometrical vector space can be measured using a ruler with scale. We denote such size by $\|\vec{u}\|$. Note that for nonzero vectors, it is a positive number and there is the natural property of scalar multiplication: $\|s\,\vec{u}\| = |s|\,\|\vec{u}\|$. Really, scalar multiplication with $s > 0$ "scales" sizes of vectors. When $s < 0$ then scalar multiplication "scales" the vector and "reverses" its orientation.

Let $\vec{u}, \vec{v} \in V$ be two nonzero vectors in a geometrical vector space. Denote φ the angle between then (measured by a protractor) and $\|\vec{u}\|, \|\vec{v}\|$ their sizes. Then, we define the dot product of these vectors $\vec{u} \cdot \vec{v}$ as a real number:

$$\vec{u} \cdot \vec{v} = \|\vec{u}\|\,\|\vec{v}\|\,\cos\varphi \qquad (4)$$

If $\vec{u}$ or $\vec{v}$ is the zero vector, then $\vec{u} \cdot \vec{v}$ is defined as zero.

The dot product can be defined without any need of $\cos\varphi$ calculation using only geometrical tools: do orthogonal projection of the vector $\vec{u}$ to $\operatorname{span}(\vec{v})$ (i.e. to the line generated by $\vec{v}$). More exactly the projection determines the abscissa from origin to the projection point and we need to use a real number r which is the length of the abscissa when $\varphi \leq 90°$ and r is minus its length when $\varphi > 90°$. Then $\vec{u}\cdot\vec{v} = r\,\|\vec{v}\|$.

It should be proven only by geometrical tools and geometrical arguments, that for $\vec{u}\cdot\vec{v}$ there holds true

(i) $\vec{u}\cdot\vec{v} = \vec{v}\cdot\vec{u}$ (dot product is commutative),
(ii) $(r\vec{u}+s\vec{v})\cdot\vec{w} = r\,(\vec{u}\cdot\vec{w}) + s\,(\vec{v}\cdot\vec{w})$, and the same is true in the 2$^{\text{nd}}$ argument (dot product is a bi-linear form),
(iii) $\vec{u}\cdot\vec{u} \geq 0$ and $\vec{u}\cdot\vec{u} = 0$ only if $\vec{u} = \vec{o}$ (dot product is positively defined).

These properties (i)–(iii) are axioms of a general dot product. More exactly, these axioms should be used as the definition of a dot product in any abstract vector space (with real scalars) where for example rulers and protractors cannot be used. Let V be such an abstract vector space and let $\vec{u}\cdot\vec{v}$ be a commutative bi-linear positively defined form. Then we define $\|\vec{u}\| = \sqrt{\vec{u}\cdot\vec{v}}$ (the norm induced by the given dot product) and for two nonzero vectors $\vec{u}$, $\vec{v}$ there exists φ such that $\vec{u}\cdot\vec{v} = \|\vec{u}\|\,\|\vec{v}\|\,\cos\varphi$. So, we can measure sizes of abstract vectors and angles between them using a given commutative bi-linear positively defined form.

The equation (4) says that nonzero vectors $\vec{u},\vec{v}$ from a geometrical vector space are orthogonal if and only if $\vec{u}\cdot\vec{v} = 0$. It suggests the following generalization of "orthogonality" for any vector space V with arbitrary dot product defined by (i)–(iii). Let $\vec{u},\vec{v} \in V$ be two vectors. They are *orthogonal* if $\vec{u}\cdot\vec{v} = 0$. Note that zero vector is orthogonal to every vector $\vec{u} \in V$. Two vectors $\vec{u},\vec{v}$ are *orthonormal* if $\|\vec{u}\| = \|\vec{v}\| = 1$ and they are orthogonal. A collection of vectors in the vector space V is called *orthonormal* (or *orthogonal*) if each pair of vectors from this collection are orthonormal (or orthogonal).

The scale-size property "$\|r\,\vec{u}\| = |r|\,\|\vec{u}\|$", Schwartz inequality "$|\vec{u}\cdot\vec{v}| \leq \|\vec{u}\|\,\|\vec{v}\|$", triangle inequality "$\|\vec{u}+\vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$", Pythagorean Theorem "$\|\vec{u}\|^2 + \|\vec{v}\|^2 = \|\vec{u}+\vec{v}\|^2$ if $\vec{u}\cdot\vec{v} = 0$" and Cosine rule "$\|\vec{u}-\vec{v}\|^2 = \|\vec{u}\|^2 + \|\vec{v}\|^2 - 2(\vec{u}\cdot\vec{v})$" can be derived only from properties (i)–(iii). These Theorems have very natural geometrical meaning but they are applicable for any abstract vector space.

When $\vec{x},\vec{y} \in \mathbb{R}^n$, then we can define the *standard dot product* in $\mathbb{R}^n$ by

$$\vec{x}\cdot\vec{y} = \vec{x}^\top\vec{y} = [x_1,\ldots,x_n]\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = x_1y_1 + \cdots + x_ny_n. \tag{5}$$

It is obvious that properties (i)–(iii) are true for the standard dot product in $\mathbb{R}^n$, so we are authorized to call it "dot product". There exists a norm $\|\vec{x}\|_2$ induced by the standard dot product in $\mathbb{R}^n$: $\|\vec{x}\|_2 = \sqrt{\vec{x}^\top\vec{x}} = \sqrt{x_1^2 + \cdots + x_n^2}$. Another non-standard dot products in $\mathbb{R}^n$ exist, i.e. there hold (i)–(iii) true for them, but we need not to deal with them here.

The relationship between a dot product in arbitrary vector space V and the standard dot product in $\mathbb{R}^n$ can be formulated as follows. Let β be an orthonormal basis in a vector space V with a dot product, $\vec{u},\vec{v} \in V$. Then $\vec{u}\cdot\vec{v} = \vec{u}_\beta^\top\vec{v}_\beta$. In other words: the dot product $\vec{u}\cdot\vec{v}$ can be computed using coordinates of given vectors w.r.t. the othonormal basis β and using standard dot product in $\mathbb{R}^n$. The proof is straightforward: Let $\beta = (\vec{b}_1,\ldots,\vec{b}_n)$, $\vec{u}_\beta = [x_1,\ldots,x_n]^\top$ and $\vec{v}_\beta = [y_1,\ldots,y_n]^\top$. Then

$$\begin{aligned}\vec{u}\cdot\vec{v} &= (x_1\vec{b}_1 + \cdots + x_n\vec{b}_n)\cdot(y_1\vec{b}_1 + \cdots + y_n\vec{b}_n) = \\ &= x_1y_1(\vec{b}_1\cdot\vec{b}_1) + x_1y_2(\vec{b}_1\cdot\vec{b}_2) + \cdots + x_ny_n(\vec{b}_n\cdot\vec{b}_n) = x_1y_1 + \cdots + x_ny_n = \vec{u}_\beta^\top\vec{v}_\beta.\end{aligned}$$

The property (ii) was used and the fact, that the basis β is orthonormal w.r.t. the dot product in V, so $\vec{b}_i\cdot\vec{b}_i = 1$ and $\vec{b}_i\cdot\vec{b}_j = 0$ for $i \neq j$. The consequence of this Theorem is $\|\vec{u}\| = \|\vec{u}_\beta\|_2$.

Two other very natural results can be simply proved for linear spaces V with dot product:

- If a collection of non-zero vectors in V is orthogonal then it is linearly independent.
- Let β is orthonormal basis in V and $\vec{u} \in V$. Then the i-th coordinate of $\vec{u}$ w.r.t. β is $\vec{u}\cdot\vec{b}_i$.

5 Vector product

Let us look an interesting mapping from $L \times L$ to L, the *vector product* or *cross product*. The set L must be a linear space of dimension 3 or (specially) it is $\mathbb{R}^3$. We show two points of view to this mapping.

§**1 Vector product from geometrical point of view.** Let L be a linear space, $\dim L = 3$. The vector product of given linear independent vectors $\vec{u}$, $\vec{v}$ is a vector $\vec{w}$ with features: (i) $\vec{w}$ is orthogonal to the plane generated by $\vec{u}, \vec{v}$, (ii) the length of $\vec{w}$ is equal to the area size of the parallelogram given by the vectors $\vec{u}, \vec{v}$ and (iii) the vectors $\vec{u}, \vec{v}, \vec{w}$ create the right-handed basis. We write $\vec{w} = \vec{u} \times \vec{v}$. When the given vectors $\vec{u}, \vec{v}$ are linear dependent then $\vec{u} \times \vec{v} = \vec{o}$.

The direct consequence of such geometrical definition is:

$$(\alpha\vec{u}) \times \vec{v} = \vec{u} \times (\alpha\vec{v}) = \alpha(\vec{u} \times \vec{v}) \quad \forall \alpha \in \mathbb{R},$$

$$\vec{u} \times \vec{v} = -\vec{v} \times \vec{u}.$$

§**2 Vector product from numerical point of view.** Let $\vec{x} = [x_1, x_2, x_3]^\top$ and $\vec{y} = [y_1, y_2, y_3]^\top$ are given vectors from $\mathbb{R}^3$. Then vector product of these vectors is

$$\vec{x} \times \vec{y} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \times \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_2y_3 - x_3y_2 \\ -x_1y_3 + x_3y_1 \\ x_1y_2 - x_2y_1 \end{bmatrix} = \begin{bmatrix} \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} \\ -\begin{vmatrix} x_1 & y_1 \\ x_3 & y_3 \end{vmatrix} \\ \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} \end{bmatrix} \tag{6}$$

We can derive this equation from following arguments: The vector product $\vec{z} = \vec{x} \times \vec{y}$ must be orthogonal to $\vec{x}$ and $\vec{y}$. So scalar products $\vec{x}^\top \vec{z}$ and $\vec{y}^\top \vec{z}$ must be equal to zero. This yields the system of linear equations

$$\begin{bmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix} \vec{z} = \vec{o}$$

When $\vec{x}$ and $\vec{y}$ are linearly independent then one non-zero solution of the linear system above is the cross product defined by (6), so $\vec{z}$ satisfies the condition (i) from §1. In order to prove conditions (ii) and (iii) from §1 we need to calculate the determinant

$$V = |\vec{x}\ \vec{y}\ \vec{z}| = \begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{vmatrix} = z_1 \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} - z_2 \begin{vmatrix} x_1 & y_1 \\ x_3 & y_3 \end{vmatrix} + z_3 \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} = z_1^2 + z_2^2 + z_3^2 = \|z\|^2$$

We used the Laplace expansion of the determinant along third column. Because V is positive the $\vec{x}, \vec{y}, \vec{z}$ create right-handed basis (if $\vec{x}, \vec{y}$ are linearly independent), so (iii) from §1 is satisfied. V is the volume of the parallelepiped given by $\vec{x}, \vec{y}, \vec{z}$ and this volume can be also calculated as $S\,h$ where S is the area size of the parallelogram given by $\vec{x}, \vec{y}$ and h is the height of the parallelepiped. Because $h = \|z\|$ and $S\,h = V = \|z\|^2$, we see that $\|z\| = S$ and the condition (ii) from §1 is proved. Moreower, if $\vec{x}, \vec{y}$ are linearly dependent then $V = 0$, so $\|z\|^2 = 0$ and z must be null vector.

§**3 Determinant and vector product.** The equation $|\vec{x}\ \vec{y}\ \vec{s}| = (\vec{x} \times \vec{y})^\top\, \vec{s}$ is true because:

$$\begin{vmatrix} x_1 & y_1 & s_1 \\ x_2 & y_2 & s_2 \\ x_3 & y_3 & s_3 \end{vmatrix} = s_1 \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3 \end{vmatrix} - s_2 \begin{vmatrix} x_1 & y_1 \\ x_3 & y_3 \end{vmatrix} + s_3 \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} = (\vec{x} \times \vec{y})^\top\, \vec{s}.$$

§**4 Vector product under the change of basis.** Let us we study the behavior of the vector product under the change of basis. Let us we have two bases β, β' in $\mathbb{R}^3$ (or in general linear space L with dimension 3) and two vectors $\vec{x}$ and $\vec{y}$ in this space. Let $\mathtt{A}$ be a matrix transforming coordinates from basis β to basis β'. It means that $\vec{x}_{\beta'} = \mathtt{A}\vec{x}_\beta$ and $\vec{y}_{\beta'} = \mathtt{A}\vec{y}_\beta$. Because of §3 the following is true for arbitrary vector $\vec{s} \in \mathbb{R}^3$

$$(\vec{x}_{\beta'} \times \vec{y}_{\beta'})^\top \vec{s} =$$
$$|\,\vec{x}_{\beta'} \quad \vec{y}_{\beta'} \quad \vec{s}\,| = |\,\mathtt{A}\vec{x}_\beta \quad \mathtt{A}\vec{y}_\beta \quad \vec{s}\,| =$$
$$|\,\mathtt{A}\,[\vec{x}_\beta \quad \vec{y}_\beta \quad \mathtt{A}^{-1}\vec{s}]\,| = |\mathtt{A}|\,|\,\vec{x}_\beta \quad \vec{y}_\beta \quad \mathtt{A}^{-1}\vec{s}\,| = |\mathtt{A}|\,(\vec{x}_\beta \times \vec{y}_\beta)^\top (\mathtt{A}^{-1}\vec{s}) =$$
$$|\mathtt{A}|\,(\mathtt{A}^{-T}(\vec{x}_\beta \times \vec{y}_\beta))^\top \vec{s}.$$

We see that $(\vec{x}_{\beta'} \times \vec{y}_{\beta'})^\top \vec{s} = |\mathtt{A}|\,(\mathtt{A}^{-T}(\vec{x}_\beta \times \vec{y}_\beta))^\top \vec{s}$ for all $\vec{s} \in \mathbb{R}^3$ (especially for $\vec{s} = [100]$ or $[010]$ or $[001]$). This implies that $\vec{x}_{\beta'} \times \vec{y}_{\beta'} = |\mathtt{A}|\ \mathtt{A}^{-T}(\vec{x}_\beta \times \vec{y}_\beta)$. We can write it as result:

$$\vec{x}_{\beta'} \times \vec{y}_{\beta'} = \frac{\mathtt{A}^{-T}}{|\mathtt{A}^{-T}|}(\vec{x}_\beta \times \vec{y}_\beta)$$

because $|\mathtt{A}^{-1}| = 1/|\mathtt{A}|$ and $|\mathtt{A}^\top| = |\mathtt{A}|$.

6 Change of coordinate systems in an affine space

In all the text above, we were working with "geometrical vector space" and it was possible to imagine it as vector space of bound vectors which is equivalent to a vector space of points, where points are the end-points of associate bound vectors. But there is a problem: the very natural transformation "shifting of all points by desired vector" cannot be represented by a linear transformation. So, we need a more advanced mathematical model.

Chapter 3 of [3] introduces the affine space, which is a pair of sets (V, X), where V is vector space of free vectors (with scalar multiplication and vector addition) and X is a set of points. The vectors from V and points from X are interconnected by new two operations "PLUS" and "MINUS":

- Operation *point* $P \in X$ *PLUS vector* $\vec{v} \in V$ gives a *point* $\in X$ which is an end-point of a representative of the vector $\vec{v}$ if such representative starts from point P.
- Operation *point* $P \in X$ *MINUS point* $Q \in X$ gives a *vector* $\in V$ which representative starts in P and ends in Q.

All operations: scalar multiplication, vector addition on V and PLUS, MINUS operations on V, X can be defined using pure geometrical tools. It is shown in chapter 3 of [1] in detail. We will note the PLUS and MINUS operations by the symbols $+$ and $-$. Points from X will be denoted by uppercase letters $(P, Q, R, \ldots)$ and vectors as usual: $\vec{u}, \vec{v}, \vec{w}, \ldots$ When $r \in \mathbb{R}$ then legal operations are: $(r\vec{u}) \in V$, $(\vec{u} + \vec{v}) \in V$, $(P + r\vec{u}) \in X$, $(P - Q) \in V$. Illegal operations are (for example): rP, $P + Q$.

In the following text, we introduce extended coordinates of vectors from V and points from X w.r.t. a coordinate system. *The coordinate system* is pair: a basis $\beta = (\vec{b}_1, \ldots, \vec{b}_n)$ of V and a point O from X. The coordinate system is denoted by (β, O). The point O is called *origin* of the coordinate system. The *radius vector* of a point $P \in X$ w.r.t. coordinate system (β, O) is the vector $P - O$. The coordinates of $\vec{v} \in V$ w.r.t. (β, O) are denoted $\vec{v}_{\beta,O}$ and the coordinates of $P \in X$ w.r.t. (β, O) are denoted by $P_{\beta,O}$. They both are (n+1)tuples from $\mathbb{R}^{n+1}$ defined by

$$\vec{v}_{\beta,O} = \begin{bmatrix} \vec{v}_\beta \\ 0 \end{bmatrix}, \qquad P_{\beta,O} = \begin{bmatrix} (P - O)_\beta \\ 1 \end{bmatrix}.$$

Note that coordinates of a vector w.r.t. the system (β, O) are coordinates w.r.t. the basis β and "0" is added. The coordinates of a point are coordinates of its radius vector w.r.t. the basis β and "1" is added. This last coordinate (0 or 1) is *type-info*, because we are able to detect from this coordinate if it is a vector or a point. Sometimes, the type-info coordinate is not mentioned when we are talking about coordinates of vectors or points and the type of the object is known. But we will show, that type-info plays very important role during calculation of coordinates by matrices.

There exists an "isomoprhism" between objects from V, X and their coordinates from $\mathbb{R}^{n+1}$ which keeps all operations: scalar multiplication, vector addition in V and PLUS, MINUS operations. The last two mentioned operations are mapped to common $+$ and $-$ in $\mathbb{R}^{n+1}$. This "isomorphism" maps (of course) vectors and points to their coordinates w.r.t. fixed chosen coordinate system. This is very flexible tool for computing with points and free vectors.

Note that the type-info coordinate allows to do legal operations with points and vectors (and their coordinates) but illegal operations like rP, $P+Q$ yields to "out of range" from the values $\{0,1\}$.

Suppose that two coordinate systems (β, O) and (β, O') are given in an affine space (V, X). *The matrix* $\mathtt{G} \in \mathbb{R}^{n+1,n+1}$ *transforming coordinates of vectors and points w.r.t. the coordinate system* (β, O) *to their coordinates w.r.t.* (β', O') is defined by

$$(\vec{b}_1, \ldots, \vec{b}_n, O) = (\vec{b}'_1, \ldots, \vec{b}'_n, O')\,\mathtt{G}$$

This is analogue definition to (2), so there are analogue properties:

(i) Matrix $\mathtt{G}$ exists for every given coordinate systems (β, O) and (β', O'). This matrix is unique.
(ii) First n columns $\vec{g}_i$ of the matrix $\mathtt{G}$ include coordinates of $\vec{b}_i$ w.r.t. the coordinate system (β', O'), the last column of $\mathtt{G}$ includes coordinates of the point O w.r.t. the coordinate system (β', O').
(iii) The matrix $\mathtt{G}$ is regular matrix.
(iv) $\mathtt{G}\,\vec{u}_{\beta,O} = \vec{u}_{\beta',O'}$, $\mathtt{G}\,P_{\beta,O} = P_{\beta',O'}$ for all vectors $\vec{u} \in V$ and for all points $P \in X$.

The property (iv) gives simple way to convert coordinates of points from one coordinate system to another even if the origin O of the coordinate system is shifted to another position O'. The coordinates of vectors are converted by the matrix $\mathtt{G}$ with the same results as converting using the matrix $\mathtt{A}$ from equation (2). The matrix $\mathtt{G}$ can be written in block form

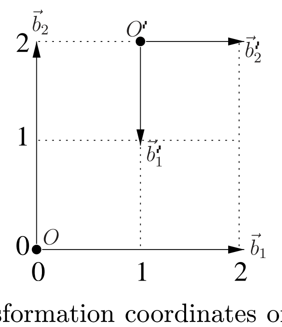
$$\mathtt{G} = \begin{bmatrix} \mathtt{A} & \vec{c} \\ \vec{o} & 1 \end{bmatrix}$$

where $\mathtt{A} \in \mathbb{R}^{n,n}$ is the matrix transforming coordinates from the basis β to the basis β', $\vec{c} \in \mathbb{R}^n$ includes coordinates of the radius vector $(O - O')$ w.r.t. the basis β', i.e. $\vec{c} = (O - O')_{\beta'}$ and finally $\vec{o} \in \mathbb{R}^{1,n}$ is a row with zeros. If we need to do the reverse coordinate transformation then we need to use the inverse of $\mathtt{G}$ which is in the form

$$\mathtt{G}^{-1} = \begin{bmatrix} \mathtt{A}^{-1} & -\mathtt{A}^{-1}\vec{c} \\ \vec{o} & 1 \end{bmatrix}.$$

It is possible to do coordinate transformation via more coordinate systems in given order of such systems. If we known matrices which do particular coordinate transformations, then their matrix product is the matrix of composite coordinate transformation. More precisely, let there be coordinate systems (β, O), (γ, P), (δ, Q) in an affine space (V, X). Let there be matrices $\mathtt{G}$ and $\mathtt{H}$ with properties $\mathtt{G}\,D_{\delta,Q} = D_{\gamma,P}$ and $\mathtt{H}\,D_{\gamma,P} = D_{\beta,O}$ for every point $D \in X$. Then the matrix product $\mathtt{HG}$ has the property $(\mathtt{HG})\,D_{\delta,Q} = D_{\beta,O}$ and it is a matrix transforming coordinates from coordinate system (δ, Q) to (β, O). This is a simple consequence of associative law of matrix product: $(\mathtt{HG})\,D_{\delta,Q} = \mathtt{H}(\mathtt{G}\,D_{\delta,Q}) = \mathtt{H}\,D_{\gamma,P} = D_{\beta,O}$.

Example (from alpha test, exercise 5). Two coordinate systems $(\vec{b}_1, \vec{b}_2, O)$ and $(\vec{b}'_1, \vec{b}'_2, O')$ are given on the figure below



We need to find the rules for transformation coordinates of arbitrary point with coordinates $[x, y]$ w.r.t. the system $(\vec{b}_1, \vec{b}_2, O)$ to its coordinates $[x', y']$ w.r.t. the system $(\vec{b}'_1, \vec{b}'_2, O')$. The transformation matrix is defined by:

$$(\vec{b}_1, \vec{b}_2, O) = (\vec{b}'_1, \vec{b}'_2, O')\,\mathtt{G}$$

and it includes the coordinates of vector $\vec{b}_1$ w.r.t. $(\vec{b}'_1, \vec{b}'_2, O')$ in first column, and the coordinates of vector $\vec{b}_2$ w.r.t. $(\vec{b}'_1, \vec{b}'_2, O')$ in second column (the last type-info element is 0 in both cases because they are vectors). $\mathtt{G}$ includes coordinates of the point O w.r.t. $(\vec{b}_1', \vec{b}_2', O')$ in the third column (here is 1 as the last element, because it is a point). So:

$$\mathtt{G} = \begin{bmatrix} 0 & -2 & 2 \\ 2 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & -2 & 2 \\ 2 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \qquad \begin{matrix} x' = -2y + 2 \\ y' = 2x - 1 \end{matrix}$$

Example (Image projection matrix). The matrix transforming coordinates of point from the coordinates w.r.t. a world coordinate system (δ, O) to coordinates w.r.t. a camera coordinate system (β, C) is known as an *image projection matrix*. This matrix $\mathtt{P}_\beta$ is defined by

$$(\vec{d}_1, \vec{d}_2, \vec{d}_3, O) = (\vec{b}_1, \vec{b}_2, \vec{b}_3, C)\, \mathtt{P}_\beta$$

but the last row of the matrix is omited because we know that the results are points and we don't have to calculate type-info of the results. So $\vec{P} \in \mathbb{R}^{3,4}$ and it is in the block form

$$\mathtt{P}_\beta = [\mathtt{A} \mid \vec{c}]$$

where $\mathtt{A}$ is a matrix transforming coordinates w.r.t. δ to coordinates w.r.t. β and $\vec{c} = (O - C)_\beta$. It is more common that we know coordinates of the point C w.r.t. (δ, O), i.e. $C_\delta = (C - O)_\delta$. We can calculate $\vec{c} = -\mathtt{A} C_\delta$ because $\mathtt{A} C_{\delta,O} = \mathtt{A}\,(C - O)_\delta = (C - O)_\beta = -(O - C)_\beta = -\vec{c}$.

Notice. Extended coordinates introduced here are also known as *homogeneous coordinates*. But the concept of homogeneous coordinates is slightly more general: points in an affine space with $\dim = n$ are represented by lines in a vector space with $\dim = n + 1$. For more information, see the section 9.2.5 in [1].

7 Affine transformations

Let (V, X) be an affine space with a coordinate system $(\beta, O) = (\vec{b}_1, \dots, \vec{b}_n)$. Suppose, we have a transformation $T : V \cup X \to V \cup X$, which transforms vectors the same way as a linear transformation $T' : V \to V$ does it and it transforms points $P \in X$ by the following manner: create radius vector $(P - O)$, transform it by T' and finally shift the resulting end-point by given fixed vector $\vec{s}$. More precisely: $T(\vec{u}) = T'(\vec{u})$ and $T(P) = O + T'(P - O) + \vec{s}$. Then T is called *affine transformation* and the transformation T' is called *associated linear transformation with T*. More lapidary: affine transformation is linear transformation plus shifting points.

The affine transformation T has its matrix $\mathtt{G} \in \mathbb{R}^{n+1,n+1}$ w.r.t. the coordinate system (β, O) defined by the equation:

$$\big(T(\vec{b}_1), \dots T(\vec{b}_n), T(O)\big) = (\vec{b}_1, \dots, \vec{b}_n, O)\, \mathtt{G} \tag{7}$$

This is analogue to the equation (3), thus the following properties about the matrix $\mathtt{G}$ are true:

(i) The matrix $\mathtt{G}$ exists and it is unique for every given transformation T and coordinate system (β, O).
(ii) First n columns $\vec{g}_i$ of the matrix $\mathtt{G}$ include the coordinates of $T(\vec{b}_i)$ w.r.t. the coordinate system (β, O). The last column includes the coordinates of the point $T(O)$ w.r.t. the coordinate system (β, O).
(iii) The affine transformation T is defined by the equation (7) uniquely and it exists for every given matrix $\mathtt{G}$ and given coordinate system (β, O).
(iv) $T(\vec{u})_{\beta,O} = \mathtt{G}\, \vec{u}_{\beta,O}$ for all $\vec{u} \in V$ and $T(P)_{\beta,O} = \mathtt{G}\, P_{\beta,O}$ for all $P \in X$.

The property (iv) gives a way to calculate coordinates of $T(\vec{v})$ and $T(P)$ when the coordinates of vector $\vec{v}$ or point P are known. The matrix $\mathtt{G}$ defined above can be written in block form like

$$\mathtt{G} = \begin{bmatrix} \mathtt{A} & \vec{c} \\ \vec{o} & 1 \end{bmatrix}$$

where $\mathtt{A}$ is the matrix of the associated linear transformation $T' : V \to V$ and $\vec{c} = (T(O) - O)_\beta$.

The matrix of a composite affine transformation can be calculated as matrix product of matrices of particular transformations. This si analogue of composite linear transformation and its matrix mentioned in section 3. If an affine transformation T has its matrix $\mathtt{G}$ and it is bijective then its inverse has the matrix $\mathtt{G}^{-1}$.

We can see that the matrix $\mathtt{G}$ transforming coordinates from coordinate system (β, O) to (β', O') holds $(\beta, O) = (\beta', O')\mathtt{G}$ and it is also matrix $\mathtt{G}$ of an affine transformation T which is defined by $T(\vec{b}'_i) = \vec{b}_i$, $T(O) = O'$. Important notice: the affine transformation T derived from such matrix $\mathtt{G}$ works in direction from (β', O') to (β, O) which is opposite to the transformation of coordinates using the same matrix $\mathtt{G}$. If we need to use the matrix of an inverse affine transformation, then $\mathtt{G}^{-1}$ must be used.

We can write a short summary. Let (β, O) and (β', O') be two coordinate systems in a vector space V. The following properties are equivalent.

- $\mathtt{G}$ transforms coordinates from (β, O) to (β', O) by $\mathtt{G}\,\vec{u}_{\beta,O} = \vec{u}_{\beta',O'}$ and $\mathtt{G}\,P_{\beta,O} = P_{\beta',O'}$.
- $(\beta, O) = (\beta', O')\mathtt{G}$.
- $\mathtt{G}$ includes coordinates of $\vec{b}_i$ w.r.t. (β', O') in its first n columns $\vec{a}_i$ and includes coordinates of $(O - O')$ w.r.t. (β', O') in its last column.
- $\mathtt{G}$ is matrix of an affine transformation T with given values $T(\vec{b}'_i) = \vec{b}_i, T(O) = O'$.

8 Reference

[1] *Tomas Pajdla*: Elements of Geometry for Computer Vision (text available in materials of “Geometry of Computer Vision and Computer Graphics” Course)