N cheers for determinants ($N \gg 0$)

A purpose of this note is to present a gentle introduction to the theory of determinants, with the stress on the adjective "gentle", this in contrast to what I experienced in my student days, when my (otherwise beloved) professor J. Haantjes just pulled a few big rabbits out of his hat.

* * *

Homogeneous coordinates

The purpose of this section is to give a motivated introduction of the concept of "homogeneous coordinates", but let me introduce some notation and terminology first.

A *boolean expression* is one whose evaluation leads to either *true* or *false*. Examples of boolean expressions that yield *true* are

$$3 = 3 \quad , \quad 4 = 4 \quad , \quad 4 \geq 4 \quad , \quad 4 \geq 3 \quad , \quad 4 > 3$$

(and, of course, *true* itself); examples of boolean expressions that yield *false* are

$$3=4,\ 4\neq 4,\ 4<4,\ 4<3,\ 4\leq 3,\ \underline{false}.$$

And finally we have boolean expressions in which variables occur, such as

$$x<x+1\ ,\ p\leq q,\ a\cdot r+b=0\ ,\ a\cdot x+b\cdot y=0\ ;$$

whether such an expression yields <u>true</u> or <u>false</u> depends in general on the values that are substituted for (or "given to") the variables.

We can turn a boolean expression into an <u>equation</u> by denoting one or more of its variables into its <u>unknowns</u>. When necessary we list the unknowns in front of the boolean expression, with a colon at the end of that list. <u>Roots</u> of the equation are values for the unknown that "satisfy the equation", i.e. are such that the boolean expression yields <u>true</u>.

For instance, if the unknown x is of type "natural number" - i.e. its value is constrained to the nonnegative integers - then

$x:\ x\leq 3$ has 4 roots, viz. 0, 1, 2 and 3

$x:\ x>x+1$ has no roots at all, and

$x{:}\ x > 11$ has an unbounded number of roots, viz. 12, 13, 14, 15, ...

We now focus for a while our attention on equations of the form $r{:}\ a\cdot r + b = 0$; in contrast to r , its unknown, we call a, b its coefficients. Choosing specific values, we can ask ourselves: do we consider

$$r{:}\ 3\cdot r + 7 = 0 \quad \text{and} \quad s{:}\ 3\cdot s + 7 = 0$$

two different equations? It will become clear that the specific identifier used to denote the unknown is relatively so unimportant that we regard the above as two different renderings of the same equation. It is the coefficients that matter, as they express the constraint on the unknown:

$$r{:}\ 3\cdot r + 7 = 0 \quad \text{and} \quad r{:}\ 4\cdot r - 5 = 0$$

are most definitely two different equations.

Whether we consider

$$r{:}\ 3\cdot r + 7 = 0 \quad \text{and} \quad r{:}\ 6\cdot r + 14 = 0$$

to be the same equation is, at first sight, on the border. It will transpire that, again, it is most convenient not to distinguish between the two. We adopt the general rule that equivalent boolean expressions give rise to the same equation, and for the special case of linear equations —of which $r: a \cdot r + b = 0$ is an example— this means that a linear equation does not change when all its coefficients are multiplied by the same factor, provided that factor <u>differs from zero</u>!

That last constraint reflects the fact that we certainly can <u>not</u> consider

$$r: 3 \cdot r + 7 = 0 \quad \text{and} \quad r: 0 \cdot r + 0 = 0$$

being the same equation: the left one has only the root $-2\frac{1}{3}$, while for the right one <u>any</u> value is a root, i.e. the two boolean expressions are not equivalent. The last equation could in fact be rendered by

$$r: \underline{true} \qquad ,$$

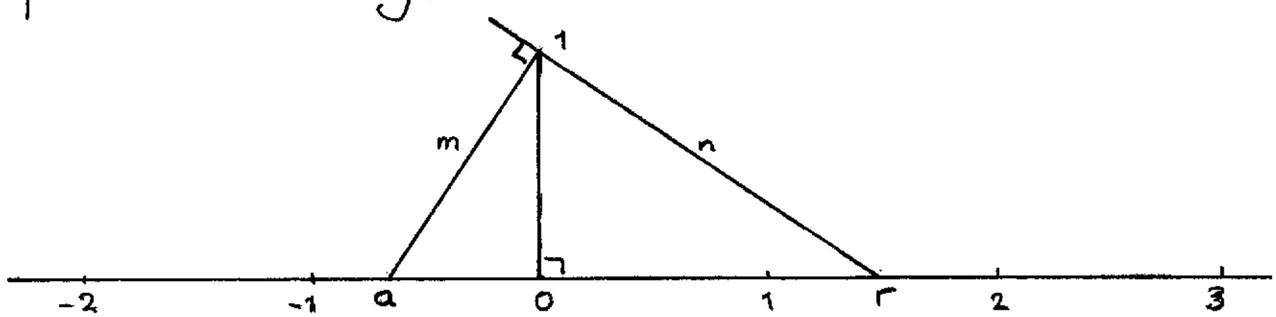
the equation that imposes no constraint at all. We prefer not to recognize it as a linear equation, in other words we adopt the convention that, in a linear equation, at least 1 of the coefficients differs from 0 .

But what if one of the coefficients equals zero? There is no problem if it is the second one: $1 \cdot r + 0 = 0$ is satisfied for $r = 0$ and for no other value of r; but there is a problem if the first coefficient equals zero; r: $0 \cdot r + 1 = 0$ has no solution because for any value of r, the left-hand side, being equal to 1, differs from 0.

For the pictorially inclined, we can show a graphical solver for the equation

$$r:\ a \cdot r + 1 = 0 \qquad ,$$

in which the real values a, r are plotted along the horizontal axis:



The graphical solver works as follows. From point a on the horizontal axis, draw the line m through point 1 on the vertical axis; line n, through that point and orthogonal to m, intersects the horizontal axis in point r.

It is clear that in the case $a = 0$, the construction fails (as it should): m coinciding with the vertical axis, line n, being horizontal, fails to intersect the horizontal axis.

The moral of the story is that, for the unknown r, the domain of the real numbers is just not rich enough: it fails one value. People have tried to solve this problem by simply extending the domain of the real numbers by one more value, usually called "infinity" and denoted by ∞, but with this extension you run into trouble because it destroys the structure of the domain of the real numbers: when this includes the extra value ∞, it is for instance no longer tenable that $r+1 \neq r$ holds for all real values r. We have to do something a bit more subtle.

Any real value r can be written as the quotient x/y of two real values x and y (with $y \neq 0$). Substituting in our original equation this quotient for the unknown r, and simplifying (by multiplying by y), yields the equation

$$x, y: \quad a \cdot x + b \cdot y = 0 \quad ;$$

please note that we stick to the rule that at least 1 of the coefficients a, b differs from 0. In a similar vein we require of a solution x, y that at least 1 of these two differs from 0. (We reject $x, y = 0, 0$ as a genuine solution because it would be trivial in the sense that it is totally independent of the coefficients a, b.)

The gain of the transition from r to the pair x, y is that now we have an equation that is perfectly solvable when the equation in r is not, viz. when $a = 0$: in this case, the equation in x, y is for instance solved by $x, y = 1, 0$.

The price we have paid, however, seems to be considerable. The equation in r was an equation in one unknown, and it determined its root uniquely, whereas the equation in x, y is indeterminate in the sense that it has infinitely many solutions: if x, y is a solution, so is $c \cdot x, c \cdot y$ for <u>any</u> c that differs from 0. We wish to consider such factor c irrelevant, we wish to regard the pairs x, y and $c \cdot x, c \cdot y$ as two <u>different</u> representations of the

<u>same</u> solution.

Notational Puritanism requires that we can distinguish between the pair x, y and the solution represented by it. We could, for instance, use square brackets for the purpose, denoting by $[x, y]$ the solution represented by the pair x, y. The defining property of the square brackets is that for all x, y, c

$$c \neq 0 \;\Rightarrow\; [x, y] = [c \cdot x, c \cdot y]$$

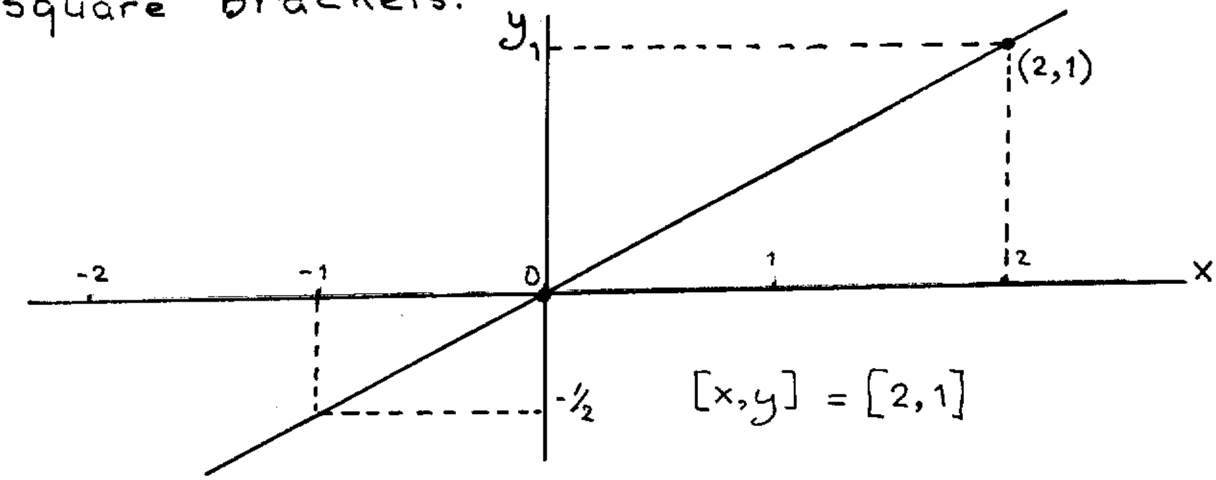
(and similarly for triples,

$$c \neq 0 \;\Rightarrow\; [x, y, z] = [c \cdot x, c \cdot y, c \cdot z] ,$$

etc.). Thus the square brackets indicate the insensitivity to multiplication of all elements of the enclosed list by c, a factor that differs from 0. It is the ratios that matter. The pair x, y is traditionally called "the homogeneous coordinates" of the solution, in the same way as we could denote the coefficients by $[a, b]$ and call them the homogeneous coordinates of the equation.

For the pictorially inclined we can now

give a graphical interpretation of our square brackets:



In the above picture we have illustrated the case $[x,y] = [2,1]$. In the general case we associate with the homogeneous coordinates $[x,y]$ the straight line through the point x,y and the origin $0,0$. Since these points differ —because $x,y = 0,0$ is ruled out for homogeneous coordinates— this line through the origin is uniquely determined by x,y , and this line through the origin can be chosen as the unique object representing $[x,y]$ because, by virtue of its construction, $c\cdot x, c\cdot y$ (for $c \neq 0$) would yield that same line.

In other words, in the transition from the (inhomogeneous) coordinate r to the (homogeneous) coordinates x,y , we have replaced the one-dimensional domain of all points on a line by the equally one-dimensional domain of all lines through a point. There is almost a one-to-one correspondence between the elements of these two domains: to the point r corresponds the line $[x, y]$ with $r = x/y$, and vice versa, but there is no point r corresponding to the line $[x, 0]$ (viz. the x-axis).

Remark Thanks to their elimination of the exception of the unsolvable equation, homogeneous coordinates have led to very elegant theories. The price to be paid for this elegance seems to be that in terms of real numbers homogeneous coordinates do not have a unique representation. (End of Remark.)

Two minor remarks about the now always solvable equation

$$x, y:\quad a \cdot x + b \cdot y = 0$$

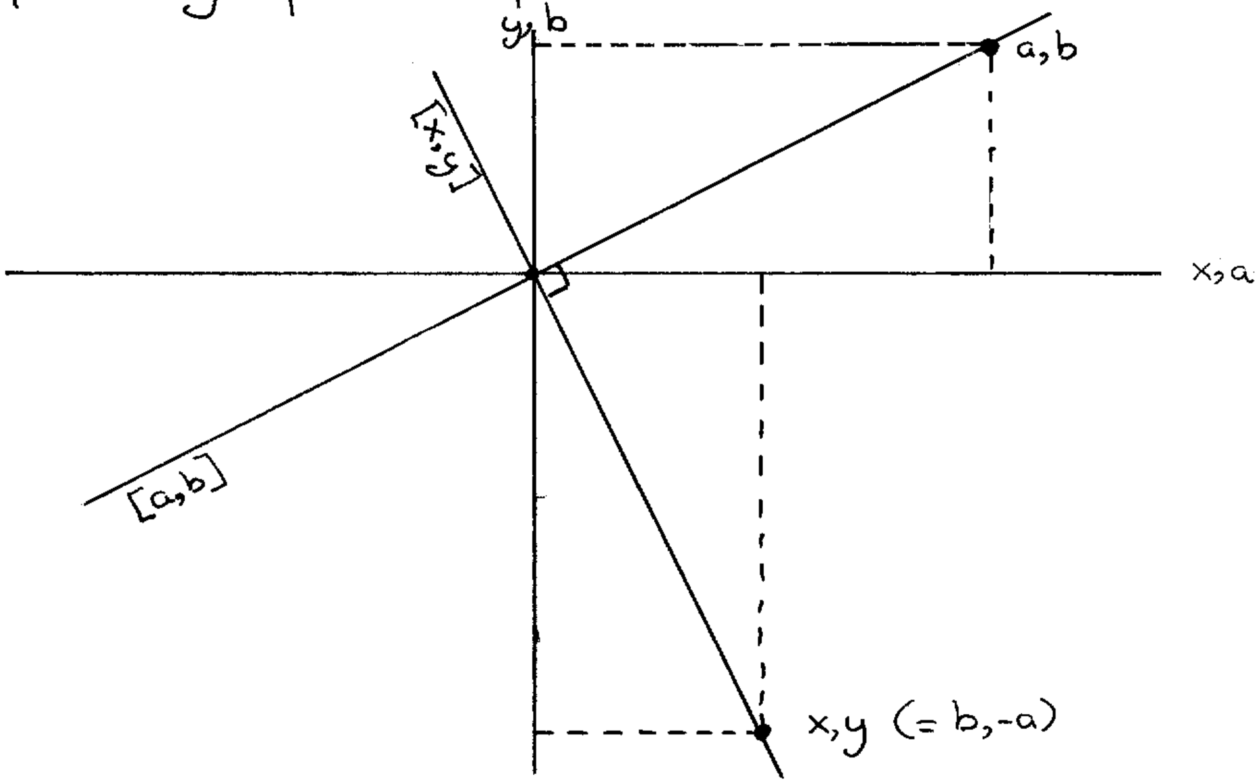
The first one is that if we wish to stress that we are interested in solving it in terms of homogeneous coordinates, we can write

$$[x,y]:\quad a\cdot x + b\cdot y = 0 \quad .$$

The difference is considerable: while the former equation in two variables is highly indeterminate, the latter has a unique solution, viz. $[b,-a]$. (Considerable as the difference is, it is usually not indicated as mathematicians have learned to live with invisible square brackets.)

The second remark is not about the solution, but about the equation. The equation is determined by its coefficients, but we have already seen that a linear equation "does not change when all its coefficients are multiplied by the same factor, provided that factor <u>differs from zero</u>!". So, it is not so much the pair of coefficients a, b that matter, it is the homogeneous coordinates $[a,b]$ that capture the essentials of the equation.

So, the solution characterized by $[x,y]$ solving the equation characterized by $[a,b]$ is in essence a relation between two entities characterized by homogeneous coordinates, which for the pictorially inclined is a relation between two lines through the origin. Plotting a,x horizontally and b,y vertically, and taking into account that x,y = b,-a solves the equation, we get the following picture for a graphical equation solver:



The line corresponding to the solution is orthogonal to the line corresponding to the equation, once more confirming that the homogeneous coordinates did eliminate the unsolvability.

(To be continued.)

prof. dr Edsger W. Dijkstra
Department of Computer Sciences
The University of Texas at Austin.
Austin, TX 78712-1188