Part II: Doing It In 3-D.

11. Inversion. Compared with the modest inhabitants of the planar world, the typical 3×3-matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \tag{1}$$

looks like a monster. We shall avoid such explicit representations whenever we can, referring to such a matrix either by its name A or by its generic entry (a_{ij}). Addition and multiplication of such matrices is straight-forward; for the product we again "dot" the rows of the left factor with the columns of the right factor. Thus $C = AB$ means

$$c_{ij} = \sum_k a_{ik} b_{kj}. \tag{2}$$

Again, this is particularly simple if one of the factors is a diagonal matrix: just multiply rows or columns of the other factor by the relevant scalars; the identity matrix has no effect at all. Again, the only cause for amazement is the associativity of this multiplication. Except for the greater number of entries in the new matrices, this part of the story is an exact copy of the 2×2-case.

Finding A^{-1} (if it exists) is another matter. There *is* a formula based on determinants, generalizing what we did in lesson 2, but much more elaborate and impractical. By contrast, the elimination technique described in lesson 10 goes through without major changes.

You should have no trouble listing the various types of elementary 3×3-matrices: six kinds of $G_{ij}(s)$ having 1's in the main diagonal and s in (ij)-th place $(i \neq j)$, three kinds of $D_i(s)$ with zeroes off-diagonal and $s \neq 0$ in (ii)-th place. Again $P_{ij} = G_{ij}(1)G_{ji}(-1)G_{ij}(1)$ switches rows i and j and changes the sign on one of them. All of these are automatically invertible: given any elementary E, it is easy to write down an F (of the same type) such that $EF = FE = I$.

For any 3×3 matrix A, we can now play the elimination game just as in lesson 10, successively finding elementary matrices $E_1, \ldots, E_k$ (all of type $G_{ij}(\cdot)$ if desired) until we reach the stage where

$$E_k \cdots E_1 \cdot A = \begin{bmatrix} d_1 & a & b \\ 0 & d_2 & c \\ 0 & 0 & d_3 \end{bmatrix} = U \tag{3}$$

is *upper triangular*. At this point, two different things can happen:

(a) Suppose $d_1 d_2 d_3 \neq 0$. Then we can continue the elimination until we have $E_m \cdots E_1 \cdot A = I$ so that A^{-1} emerges as a product of elementary matrices.

(b) Suppose $d_1 d_2 d_3 = 0$. Then $UX = 0$ (and hence $AX = 0$) can be solved for some non-zero X. Indeed:
- if $d_1 = 0$, we can take $z = y = 0$ and $x = 1$;
- if $d_1 \neq 0$ but $d_2 = 0$, we can take $z = 0$, $y = 1$, and solve for x;
- if $d_1 \neq 0$ and $d_2 \neq 0$ but $d_3 = 0$, we can take $z = 1$ and solve for y and then for x.

Summing up: *we have described a process which either computes A^{-1} or solves $AX = 0$ with $X \neq 0$.*

Remarks: 1. In this style of computing A^{-1} it is convenient to start with the identity matrix and subject it to the same row-operations (i.e. multiply it by the same E_j) that are used to reduce A. By the time A is reduced to I, this former identity matrix will be built up to $A^{-1} = E_m \cdots E_1$.

2. Rewritten as $A = LU$, with $L = E_1^{-1} \cdots E_k^{-1}$, equation (3) is referred to as an *LU-factorization* of A, provided that the elementary factors of L are all of the type $G_{ij}(\cdot)$. Another popular procedure is the so-called *QR-factorization* in which the second factor is still upper triangular, while the first is orthogonal. Upper triangular matrices are obviously convenient for solving linear equations.

12. Determinants. The determinant of the matrix A displayed in lesson 11 is defined as

$$\begin{aligned}\det A = {} & a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ & - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}.\end{aligned}$$

A first glimmer of method in this madness occurs when we observe that its terms are of the form $\pm a_{1\alpha}a_{2\beta}a_{3\gamma}$, where (α, β, γ) runs over all permutations of the digits $(1, 2, 3)$, and where the sign is negative iff the term in question has exactly *one* factor a_{kk}. On interchanging row and column indices, these terms appear as $\pm a_{\alpha 1}a_{\beta 2}a_{\gamma 3}$, with the same rule of sign; thus, rearranging the factors in each term, we get the same expression as before, whence our first insight:

$$\det A^T = \det A.$$

The light grows brighter as we rebundle our terms in the following manner

$$\det A = a_{11} \det \begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix} - a_{12} \det \begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix} + a_{13} \det \begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}, \tag{1}$$

in which the entries of the first row are matched off with determinants created from the other rows. This is called "expansion by the first row" and has, of course, natural analogues for the second and third rows. Looking at such expansions, we can make some important observations.

Imagine that E is an elementary matrix affecting only rows 2 and 3. If $E = G_{ij}(s)$, left multiplication by E will not change the determinant, because it has no effect on the 2×2 subdeterminants shown in (1). On the other hand, if $E = D_i(s)$, each of these subdeterminants is multiplied by s, and so is $\det A$. In either case, we can say that $\det A$ is multiplied by $\det E$, since the latter is 1 or s respectively. Expanding by other rows, we get similar information for *any* E. Altogether, we see that

$$\det(EA) = \det E \cdot \det A, \tag{2}$$

for any elementary E. In particular, (a) *interchanging any two rows changes the sign of the determinant,* (b) *adding a multiple of one row to another row does not change the determinant,* and (c) *multiplying any one row by a scalar leads to a corresponding multiplication of the determinant.*

Therefore if we choose $E_1, \ldots, E_k$ of type $G_{ij}(\cdot)$ so as to make $E_k \cdots E_1 A = U$, we still have $\det U = \det A$. But if U is upper triangular as in lesson 11, it easy to see that $\det U = d_1 d_2 d_3$ (check!). This gives an effective way of computing $\det A$ by elimination. At the same time it shows that

$$A \textit{ is invertible if and only if } \det A \neq 0.$$

As in lesson 10, repeated application of (2) shows that

$$\det(BA) = \det B \cdot \det A,$$

for any invertible B. On the other hand, if B is not invertible then neither is BA (see?), and both sides of the equation are zero.

To discuss the further properties of $\det A$, it is convenient to think of A as a list of its rows R_1, R_2, R_3 and to write $\det(R_1, R_2, R_3)$ accordingly. A slick reformulation of (1) is now obtained by introducing the *cross product* $V \times W$ of two vectors, which for $V = R_2$, $W = R_3$ has exactly the three 2×2 determinants shown in (1) as its components (the middle one with a minus sign). The full message then reads

$$\det(R_1, R_2, R_3) = R_1 \bullet R_2 \times R_3 = -R_2 \bullet R_1 \times R_3 = R_3 \bullet R_1 \times R_2. \tag{3}$$

This can be compactified further by using the matrix $A^\star = [R_2 \times R_3, -R_1 \times R_3, R_1 \times R_2]$, which has assorted cross products as columns. It gives us the single equation $AA^\star = (\det A)I$, from which we could construct A^{-1} as in lesson 2, without resorting to elimination. Though flashy, this method is computationally impractical and even contains theoretical puzzles (e.g. is $A^\star \neq 0$, and is $A^\star A = AA^\star$?)

13. Nullity and Rank. Some matrices are more singular than others. To assess the degree of singularity — called "nullity" — of a matrix A, we look at the size of its *null-space* $\mathcal{N}(A)$, i.e. the set of all column vectors X such that $AX = 0$. At one extreme we have the invertible matrices, which kill nothing (so $\mathcal{N}(A) = 0$); at the other, we have the zero matrix, which kills everything. More interesting than these extremes are the singular matrices $A \neq 0$. We shall see that their nullity can be either 1 or 2.

The best way to size up $\mathcal{N}(A)$ is to picture it in a spatial (three-dimensional) coordinate system. If X has coordinates x, y, z, the matrix equation $AX = 0$ means exactly the same as the 3 scalar equations

$$\begin{aligned} a_{11}x + a_{12}y + a_{13}z &= 0 \\ a_{21}x + a_{22}y + a_{23}z &= 0 \\ a_{31}x + a_{32}y + a_{33}z &= 0. \end{aligned} \tag{1}$$

Unless all its coefficients are zero, each such equation determines a plane through the origin. (For instance, the first equation describes all X perpendicular to the first row of A.) Two such planes either coincide or intersect along a line; a third plane either contains this line or intersects it in the origin. Of course, $\mathcal{N}(A)$ is just this intersection, since it consists of the vectors X obeying all 3 equations. Therefore it can be

a) the single point 0 b) a line through 0
c) a plane through 0 d) the entire space,

i.e., it can have dimension 0, 1, 2, or 3. This number is known as the *nullity* of A.

Another way of viewing this hierarchy is through the notions of *dependence* and *rank*, applied to the columns of A.

In general, any set of vectors $V_1, \ldots, V_m$ is said to be dependent, if the equation $x_1V_1 + \cdots + x_mV_m = 0$ can be solved for $x_1, \ldots, x_m$ without setting $x_i = 0$ throughout. This is just an even-handed way of asserting that at least one of the V_i can be expressed in terms of the others (see?). When applied to the case of a pair V_1, V_2, it says that one is a scalar multiple of the other. Geometrically they lie on the same line through 0 (are "parallel"). For three vectors V_1, V_2, V_3, dependence requires that

$$xV_1 + yV_2 + zV_3 = 0 \tag{2}$$

can be solved non-trivially, i.e. that the matrix having these vectors as columns be singular. Geometrically, as any two vectors lie in a single plane through 0, it means that the third vector (being obtainable from the other two by scaling and summing) fits into that same plane, so the three vectors are "coplanar".

The *rank* of a matrix A is the maximum number of independent vectors to be found among its columns. With 3×3-matrices, it ranges from 0 for the zero matrix to 3 for an invertible one. A singular matrix $A \neq 0$ must therefore have rank 1 or 2. Rank is related to nullity: the bigger the one, the smaller the other.

Labeling the columns of A by V_1, V_2, V_3, suppose that A has rank 1. Then any *two* of these vectors are already dependent, and (2) can be solved non-trivially with $z = 0$, again with $y = 0$, and once more with $x = 0$ (see?). These solutions cannot possibly be multiples of a single one, hence A must have nullity = 2.

Now assume that A has rank 2. Then two of the columns (say V_1, V_2) are independent, and hence $z \neq 0$ in every non-trivial solution of (2). Thus $\mathcal{N}(A)$ meets the coordinate plane $\{z = 0\}$ in only one point, and therefore cannot itself be a plane. Hence A has nullity = 1.

To summarize: *rank and nullity always add up to the total dimension (here = 3).*

Considerations of dependence and rank can also be applied to the *rows* of A, (i.e. the columns of A^T) and lead to the same connection with the nullity of A. If A^T has rank 1, all rows of A are scalar multiples of a single row; hence $\mathcal{N}(A)$, as the solution set of just *one* of the equations in (1), is a plane. On the other hand, if two (but not all three) rows are independent, only *one* of the equations in (1) can be eliminated, and $\mathcal{N}(A)$ is a line. Hence the nullity of A and the rank of A^T also add up to 3, and *the rank of A always equals rank of A^T.*

We note in passing that rank and nullity remain unchanged if we multiply A by an invertible matrix (left or right). In particular, similar matrices have the same rank.

14. Diagonalization. As in lesson 4, a matrix A is "diagonalizable" if it is similar to a diagonal matrix D, that is if

$$A = MDM^{-1}, \tag{1}$$

with a suitable matrix M. The aim of this lesson is to find out when and how this is possible.

Writing $M = \big[V_1, V_2, V_3\big]$ as a triple of columns, the similarity relation $AM = MD$ amounts to

$$A\big[V_1, V_2, V_3\big] = \big[\lambda_1 V_1, \lambda_2 V_2, \lambda_3 V_3\big];$$

in other words, V_1, V_2, and V_3 are *eigenvectors* of A, and they must be *independent* in order to make M non-singular. Thus A is diagonalizable iff it has three independent eigenvectors. (Needless to say, λ_1, λ_2, and λ_3 are the corresponding eigen*values*.)

As before, we use determinants to split the eigenrelation $AV = \lambda V$ into the two equations

$$\det(A - \lambda I) = 0 \qquad \text{and} \qquad (A - \lambda I)V = 0, \tag{2}$$

hoping to solve first the one for λ and then the other for V. It should be kept in mind that this clever procedure is quite impractical for larger matrices, because of the high cost of the first step. However, for 3×3-matrices A it still works reasonably well: the *characteristic polynomial* has the form

$$p_A(\lambda) = \det(\lambda I - A) = \lambda^3 + a\lambda^2 + b\lambda + c, \tag{3}$$

and *must* have at least one (real) root λ_1, because it is positive for $\lambda \to +\infty$ and negative for $\lambda \to -\infty$. (Incidentally, you should try computing this polynomial for the generic matrix (a_{ij}), and you will find that $a = -\mathrm{tr}\, A$, $b = \mathrm{tr}\, A^\star$, and $c = -\det A$.)

Good news: *every 3×3-matrix A has at least one (real) eigenvector V_1.*

But not quite good enough: for diagonalization, two special conditions must be met. Firstly, since similar matrices have equal characteristic polynomials, (1) would imply that $p_A(\lambda) = p_D(\lambda)$, i.e.

(i) *$p_A(\lambda)$ must factor completely into $(\lambda - \lambda_1)(\lambda - \lambda_2)(\lambda - \lambda_3)$.*

Secondly, since similar matrices have equal nullities, we can equate those of $(D - \lambda_k I)$ and $(A - \lambda_k I)$, and demand at least:

(ii) *the multiplicity of $(\lambda - \lambda_k)$ in the factorization must not exceed the nullity of $(A - \lambda_k I)$.*

Note that condition (ii) holds automatically if all three eigenvalues are different!

Now suppose that (i) and (ii) are satisfied. Is A then diagonalizable?

The case $p_A(\lambda) = (\lambda - \lambda_1)^3$ is easy: (ii) says that $A - \lambda_1 I$ has nullity $= 3$, hence $A = \lambda_1 I = D$. With that triviality out of the way, we now take λ_1 to be distinct from λ_2 and λ_3. Then if $\lambda_2 \neq \lambda_3$, *any* corresponding pair V_2, V_3 of eigenvectors is automatically independent (see?). On the other hand, if $\lambda_2 = \lambda_3$, condition (ii) ensures that there will still be a pair V_2, V_3 of independent eigenvectors. In either case we get the desired independent triple V_1, V_2, V_3, because of the following calculation (which is the heart of the matter). Any dependence relation $V_1 = yV_2 + zV_3$ would imply

$$\lambda_1 yV_2 + \lambda_1 zV_3 = \lambda_1 V_1 = AV_1 = y\lambda_2 V_2 + z\lambda_3 V_3.$$

This cannot be: subtracting the two extremes of this equation from each other, we would get the forbidden dependence relation $(\lambda_1 - \lambda_2)yV_2 + (\lambda_1 - \lambda_3)zV_3 = 0$ (explain the details!).

Remark: By itself, condition (i) makes A similar to a *triangular* matrix, as we shall point out in lesson 16. If complex numbers are allowed, it is always satisfied, and condition (ii) is the only criterion for diagonalizability.

15. Standard Forms. In practice it often happens that you know one eigenvalue (or eigenvector) without having analysed or even computed the characteristic polynomial. But you do have a foot in the door, and here is how you can pry it open in three simple steps, each involving a "conjugation" $A \mapsto M^{-1}AM$ by some invertible matrix M.

First, take $M = \big[V_1, V_2, V_3\big]$ with $V_1 \in \mathcal{N}(A - \lambda_1 I)$ but V_2 and V_3 not necessarily eigenvectors. For easy handling, let V_2 and V_3 be columns of I (if possible). Specifically, if $V_1 = (1, u, v)$, the conjugation by M takes the form

$$\begin{bmatrix} 1 & 0 & 0 \\ -u & 1 & 0 \\ -v & 0 & 1 \end{bmatrix} A \begin{bmatrix} 1 & 0 & 0 \\ u & 1 & 0 \\ v & 0 & 1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \alpha & \beta \\ 0 & a' & b' \\ 0 & c' & d' \end{bmatrix} = A_1. \tag{1}$$

This is a useful first step since it concentrates attention on the smaller matrix A' in the lower right corner. In particular, the formula $\det(xI - A) = (x - \lambda_1)\det(xI - A')$ gives convenient access to the characteristic polynomial of A.

You can use this format even if the first coordinate of the given eigenvector is 0: just take $W_1 = (1, u, v) = P_{1k}V_1$ for $k = 2$ or 3; then W_1 is an eigenvector of $P_{1k}AP_{1k}$. In other words, interchange the first and k-th rows *and* columns of A (think of this as the 0-th step), and then proceed as above, using W_1.

As a second step you may wish to use an invertible 2×2-matrix N and the formula

$$\begin{bmatrix} 1 & 0 \\ 0 & N^{-1} \end{bmatrix} \begin{bmatrix} \lambda_1 & (\alpha, \beta) \\ 0 & A' \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & N \end{bmatrix} = \begin{bmatrix} \lambda_1 & (\gamma, \delta) \\ 0 & B \end{bmatrix} = A_2 \tag{2}$$

for further simplification of the 2×2 matrix obtained in step one. Here $(\gamma, \delta) = (\alpha, \beta)N$, and $B = N^{-1}A'N$. Thus you can get the matrix B to be in one of the three standard forms (a) – (c) of lesson 5. If it happens that $B = \lambda_1 I$, the conjugation by N will leave it unchanged; then N can be chosen so as to make $(\gamma, \delta) = (1, 0)$, or $(0, 0)$.

The aim of the third step is to standardize the (γ, δ) — if possible make it zero — when $B \neq \lambda_1 I$. To this avail we try the formula

$$\begin{bmatrix} 1 & -X \\ 0 & I \end{bmatrix} \begin{bmatrix} \lambda_1 & (\gamma, \delta) \\ 0 & B \end{bmatrix} \begin{bmatrix} 1 & X \\ 0 & I \end{bmatrix} = \begin{bmatrix} \lambda_1 & (\sigma, \tau) \\ 0 & B \end{bmatrix} = A_3 \tag{3}$$

where $(\sigma, \tau) = (\gamma, \delta) - X(B - \lambda_1 I)$, and $X = (x, y)$ is to be chosen so as to make σ, τ as trivial as possible.

If $B - \lambda_1 I$ is invertible, it is obvious how to make $(\sigma, \tau) = (0, 0)$ or anything you please. This is the easiest and most common case. On the other hand, if λ_1 is an eigenvalue of B, there are several scenarios — but you can always get $\tau = 0$ and $\sigma = 1$ (if not 0) as follows.

Since $B \neq \lambda_1 I$ has emerged from step two in standard form, it is either $D(\lambda_1, \lambda_2)$, with $\lambda_1 \neq \lambda_2$, or it is $\lambda_1 I + E_{12}$. Hence the second column of $B - \lambda_1 I$ is non-zero, and the second coordinate of $X(B - \lambda_1 I)$ can be made $= \delta$, giving $\tau = 0$. But $\sigma = \gamma$ stays put, and if it is $\neq 0$, we conjugate by $D_1(\gamma)$ to make it $= 1$.

If, at this point, we have $\sigma = 1$ and B diagonal, we may perform a final conjugation by P_{13}, in order to wind up with one of the the two patterns displayed below.

Conclusion: For every 3×3-matrix A there is an invertible M such that $M^{-1}AM$ equals either

$$\begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{bmatrix} \qquad \textit{or} \qquad \begin{bmatrix} \lambda_1 & 0 \\ 0 & B \end{bmatrix}, \tag{4}$$

where λ_1 is a suitable eigenvalue, and B is one of the standard forms for 2×2-matrices.

16. Characteristic and Minimal Polynomials. As a by-product of the conclusion reached in the last lesson, we get some insight into the relation between A and its characteristic polynomial $p_A(\lambda)$. Remember that any two similar matrices have the *same* characteristic polynomial, because $\det(\lambda I - M^{-1}AM) = \det M^{-1}(\lambda I - A)M = \det(\lambda I - A)$. Therefore, the two cases shown in the "conclusion" will give $(\lambda - \lambda_1)^3$ and $(\lambda - \lambda_1)p_B(\lambda)$, respectively, as the characteristic polynomial of A.

In particular, this justifies the remark made at the end of lesson 14, that A is similar to an upper triangular matrix whenever $p_A(\lambda) = (\lambda-\lambda_1)(\lambda-\lambda_2)(\lambda-\lambda_3)$ splits into the product of three (not necessarily distinct) factors. In the first case, this is obvious; in the second, the splitting of $p_A(\lambda)$ forces the 2×2 matrix B to be of type (a) or (b), so that $p_B(\lambda)$ would have (real) roots.

Another consequence of the same conclusion is the Cayley-Hamilton relation $p_A(A) = 0$. Letting $A_0 = M^{-1}AM$ denote the standard form shown, we have to check that $p_A(A_0) = M^{-1}p_A(A)M = 0$. In the two cases, $p_A(A_0)$ evaluates as

$$(A_0 - \lambda_1 I)^3 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}^3 \qquad \text{and} \qquad p_A(A_0) = \begin{bmatrix} p_A(\lambda_1) & 0 \\ & \\ 0 & p_A(B) \end{bmatrix} \tag{1}$$

respectively. Both times the result is zero; in the second case, we see this using the 2×2 equality $p_A(B) = (B - \lambda_1 I)p_B(B) = 0$, derived from Cayley-Hamilton for B.

The Cayley-Hamilton relation suggests an entirely new way of approaching the polynomial $p_A(\lambda)$. Instead of struggling with determinants (a prohibitively onerous task in higher dimensions), we go fishing in the set of polynomials $f(\lambda)$ which *annihilate* A, i.e. which yield $f(A) = 0$. Such polynomials are not hard to find, as we shall see below. For the moment, let us stop to think exactly what we are looking for.

Given any two non-zero polynomials $f(\lambda)$ and $g(\lambda)$, we can carry out a long division of one by the other, getting a (partial) quotient $h(\lambda)$ and (possibly) leaving a remainder $r(\lambda)$, i.e.

$$f(\lambda) = g(\lambda)h(\lambda) + r(\lambda). \tag{2}$$

Obviously, $f(A) = g(A) = 0 \quad \Longrightarrow \quad r(A) = 0$. Now remember that the degree of the remainder $r(\lambda)$ is strictly *lower* than that of $g(\lambda)$. Hence, if $f(A) = g(A) = 0$ and if $g(\lambda)$ was chosen (non-zero) of lowest possible degree, there can be no remainder. In other words, *every* annihilating polynomial $f(\lambda)$ is then a multiple of $g(\lambda)$. If, moreover, we fix its leading coefficient at 1, the *minimal polynomial* $g(\lambda) = m_A(\lambda)$ of A is thereby uniquely determined.

In particular, $p_A(\lambda)$ is a multiple of $m_A(\lambda)$, but the two need not be equal: the characteristic polynomial is always of degree n (here $= 3$), while the minimal polynomial may have smaller degree. For instance, $p_I(\lambda) = (\lambda - 1)^3$ and $m_I(\lambda) = \lambda - 1$ (you should invent some less obvious examples). However, *both are equally suitable for tracking down eigenvalues*. Indeed, every root of $m_A(\lambda)$ is also a root of its multiple $p_A(\lambda)$ and therefore an eigenvalue. Conversely, $AV = \lambda_1 V$ implies $f(A)V = f(\lambda_1)V$ (for any polynomial) and hence $0 = m_A(\lambda_1)V$; for non-zero V, this says that the eigenvalue λ_1 must be a root of $m_A(\lambda)$.

To find the minimal polynomial, grab a vector $V \neq 0$ and form $V^{(k)} = A^k V$ for $k = 0, 1, 2, 3$. For instance, if V is the j-th column of I, then $V^{(k)}$ is simply the j-th column of A^k. Now solve the equation

$$x_0 V^{(0)} + x_1 V^{(1)} + x_2 V^{(2)} + x_3 V^{(3)} = 0, \tag{3}$$

taking care to pick $x_3 = 0$ or even $x_3 = x_2 = 0$ if this can be done without making the whole solution trivial. Then $g(\lambda) = x_0 + x_1\lambda + x_2\lambda^2 + x_3\lambda^3$ is of minimal degree with the property that $g(A)V = 0$. The reasoning around equation (2) shows that every $f(\lambda)$ such that $f(A)V = 0$ must be a multiple of $g(\lambda)$. In particular, this applies to $m_A(\lambda)$, and hence the roots of $g(\lambda)$ are eigenvalues. Most of the time you will find that $g(\lambda)$ already annihilates A and therefore (with $x_3 = 1$) equals $m_A(\lambda)$. If not, $m_A(\lambda)$ shows up as a common multiple of two or three such $g(\lambda)$ associated with different (independent) initial vectors V.

17. Symmetry. The most easily recognized symptom of diagonalizability is *symmetry*, i.e. $A^T = A$. It is equivalent to the property that

$$AV \bullet W = V \bullet AW, \tag{1}$$

for any pair of vectors V, W (cf. lesson 9). To see how this produces independent eigenvectors, let V_1 be a (real) eigenvector of A, and consider the *plane* $\Pi = V_1^\perp$ perpendicular to it. The equations

$$0 = V_1 \bullet W = \lambda_1 V_1 \bullet W = AV_1 \bullet W = V_1 \bullet AW$$

show that $W \in \Pi$ implies $AW \in \Pi$ (see?), so that A induces a linear transformation $T : \Pi \to \Pi$ on this plane.

Now we have a two-dimensional problem. If we put a rectangular coordinate system on Π, the transformation T will be given by some 2×2-matrix (cf. lesson 3), and since (1) ensures that $T(X) \bullet Y = X \bullet T(Y)$, this matrix is symmetric. According to lesson 9, it therefore has two perpendicular eigenvectors V_2 and V_3 in Π. Clearly these are also eigenvectors of A, which thus winds up with *three mutually perpendicular (real) eigenvectors.*

If we choose the eigenvectors of our symmetric A to have unit length and put $M = \big[V_1, V_2, V_3\big]$, we again get

$$A = MDM^{-1} = MDM^T, \tag{2}$$

with D diagonal and M orthogonal, as in lesson 9.

Slogan: *every symmetric matrix is orthogonally diagonalizable.* (Conversely, every matrix of the form MDM^T is obviously symmetric.)

Some of this magic even rubs off on *arbitrary* (not necessarily symmetric) A as follows. Since AA^T is obviously symmetric, we can find an orthogonal M and a diagonal D such that $D = M^T(AA^T)M = BB^T$, where $B = M^TAM$. So, for *any* A we have

$$A = MBM^{-1} = MBM^T \qquad \text{with} \qquad BB^T = D. \tag{3}$$

The second equation expresses the fact that the rows of B are mutually perpendicular (see?). We shall call such a matrix "row-rectified". Any diagonal matrix is row-rectified, and so is any orthogonal matrix. Equation (3) says that *any square matrix is orthogonally similar to a row-rectified matrix.*

This is closely related to the factorization of any A into symmetric and orthogonal parts, as at the end of lesson 9. First consider a row-rectified matrix B. Since the diagonal entries of $D = BB^T$ are the lengths of the rows of B, they are non-negative. Hence there is a diagonal matrix D_0 such that $D_0^2 = D$, and then

$$B = D_0Q \qquad \text{with} \qquad QQ^T = I, \tag{4}$$

i.e., Q orthogonal. In fact, Q is made up of the non-zero rows of B scaled to unit length, supplemented by suitable perpendicular unit vectors wherever some rows of B are 0 (since D_0 also has zeroes in those places, it does not matter how we choose these supplementary rows, as long as all are perpendicular.)

Taking (3) and (4) together, we conclude that any A can be written as

$$A = MD_0N = (MD_0M^T)(MN) = (MN)(N^TD_0N), \tag{5}$$

with M and $N = QM^T$ orthogonal and D_0 non-negative diagonal. The latter two formulas show A factored into a symmetric and an orthogonal matrix, in either order. This is called the "polar decomposition" of A. The first form $A = MD_0N$ is usually referred to as the "singular value decomposition".

18. Isometries. Recall that a square matrix A is *orthogonal* if $A^TA = I$, that is, if its columns are mutually perpendicular unit vectors. As in lesson 9, this is equvalent to saying that

$$AY \bullet AX = A^TAY \bullet X = Y \bullet X, \tag{1}$$

for any pair X, Y of vectors. This means that, as a linear transformation, A is an *isometry*, i.e. it preserves lengths and angles. We shall now survey such transformations.

First, two easy consequences of the definition:

a) $\lambda = \pm 1$, for any (real) eigenvalue, because A cannot change the length of the eigenvector.

b) If $X \in V^\perp$ is perpendicular to an eigenvector V, then so is $AX \in V^\perp$.

Grab a (real) eigenvector $V = V_1$ of A, and hold on to it for the rest of this discussion. Then (a) says that $AV = \pm V$, and (b) is even more revealing: A induces an orthogonal linear transformation $T : \Pi \to \Pi$ on the plane $\Pi = V^\perp$. As in lesson 9, T must be a reflection or a rotation, and accordingly A falls into two possible patterns.

A) If T is a reflection, it must have two perpendicular eigenvectors V_2, V_3 in $V^\perp$ with, say $\lambda_2 = 1, \lambda_3 = -1$. It follows that $A^2 = I$ (see?) and $A = A^{-1} = A^T$, so that A is also *symmetric*.

But then A is similar to a diagonal matrix D whose diagonal entries $\lambda_1, \lambda_2, \lambda_3$ (the eigenvalues of A) must each equal ± 1. Therefore the trace $\tau = \lambda_1 + \lambda_2 + \lambda_3$ of A can only have the values ± 3 or ± 1 (see?), and this can be used to distinguish the possible geometric patterns.

We pass over the trivial cases $A = \pm I$ (which are the only ones with $\tau = \pm 3$) and are left with two others:

i) $\tau = 1$, say $\lambda_1 = -1, \lambda_2 = \lambda_3 = 1$. Then A fixes everything in $V^\perp$ and reverses V and all its multiples. This is called *reflection through the plane* $V^\perp$ or *along the line* given by V. A simple sketch of the situation, and a look back at formula (5) of lesson 8, should help to convince you that, for every X,

$$AX = X - 2 \cdot \mathrm{proj}_V(X) = X - 2\frac{V \bullet X}{V \bullet V}V \tag{2}$$

in this case.

ii) $\tau = -1$, say $\lambda_1 = 1, \lambda_2 = \lambda_3 = -1$. Then A fixes V and reverses all vectors in $V^\perp$. This is called *reflection through the line* given by V or *rotation about* V through $180°$.

B) Now to the hard core. If A is *not* symmetric, it must act on $V^\perp$ as a rotation through a certain angle θ with $\sin\theta \neq 0$ (which can be computed from $W \bullet AW$, using any $W \in V^\perp$). If $AV = V$, A is called a *(proper) rotation about* the line given by V. If $AV = -V$, we call it an *improper rotation* about that line; it may be thought of as a rotation followed by a reflection along V.

Starting from geometric data, we can retrieve the matrix A by computing $A = (AM)M^{-1}$, where the columns of M are suitable non-coplanar vectors V_i whose images AV_i are known ($i = 1, 2, 3$). If possible, take them to be eigenvectors (symmetric case). If not, let V be an eigenvector and choose $V_2, V_3 \in V^\perp$ to be perpendicular to each other and of equal length. On these, A must have the effect of a rotation, i.e. $AV_2 = rV_2 + sV_3$, $AV_3 = -sV_2 + rV_3$, where $r = \cos\theta$, $s = \sin\theta$. In either case, we get $AM = MB$ with

$$B = M^{-1}AM = \begin{bmatrix} \pm 1 & 0 & 0 \\ 0 & \pm 1 & 0 \\ 0 & 0 & \pm 1 \end{bmatrix} \qquad \text{or} \qquad \begin{bmatrix} \pm 1 & 0 & 0 \\ 0 & r & -s \\ 0 & s & r \end{bmatrix}, \tag{3}$$

respectively.

Remark: Since $\det(A^TA) = (\det A)^2 = 1$, we always have $\det A = \pm 1$. You can easily satisfy yourself that, among all the types described above, the ones having $\det A = 1$ are exactly the rotations ($0°$ or $180°$ included).

19. Volumes and Areas. For a handy geometric interpretation of determinants, we go back to the context of lesson 12. Letting vol(A) denote the volume of the parallelepiped spanned by the row-vectors of the matrix A, we shall now establish the miraculous formula

$$\text{vol}\,(A) = |\det A|. \tag{1}$$

First note that both sides are 0 if the rows R_1, R_2, R_3 of A are dependent (coplanar), so we can forget about that case. On the other hand, if they are *independent*, then $\det A \neq 0$ and A is a product of elementary matrices. Hence we need only show that

$$\text{vol}\,(EA) = |\det E| \cdot \text{vol}\,(A), \tag{2}$$

for elementary E and any A (see?). Indeed if E is one of the shears $G_{ij}(s)$, it does not change vol(A). For instance, $G_{12}(s)$ replaces R_1 by $R_1 + sR_2$, thus moving the tip of R_1 in a direction parallel to R_2; hence it does not change the area of the parallelogram spanned by these two — nor the volume of the parallelepiped, which equals this area times a "height" given by R_3. On the other hand, $E = D_i(s)$ obviously has the effect of multiplying vol(A) by $|s|$.

Since $\det A = \det A^T$, we get the same volume by working with the columns of A. As these are the images (under the transformation $X \mapsto AX$) of the unit coordinate vectors, $|\det A|$ is also the volume of the image of the unit cube. If you imagine space filled with tiny coordinate cubes, you can see that $|\det A|$ is the factor by which *any* volume is changed under this transformation.

Another way of looking at this is via the singular value decomposition $A = MD_0N$ where the orthogonal factors M and N have no effect in equation (1).

Tracking down the geometric meaning of the cross product $V \times W$ is even more fun. Let Y be a third vector, and consider the matrix $\big[Y, V, W\big]$ having these as columns, so that

$$\det\big[Y, V, W\big] = Y \bullet V \times W, \tag{3}$$

by lesson 12. Taking Y to be first V or W, and then a unit vector in the direction of $V \times W$, you should easily be able to arrive at the following conclusion:

$V \times W$ is a vector perpendicular to both V and W, and equal in length to the area of the parallelogram spanned by V and W.

The problem with this description is that there are two candidates (differing in sign) which fill the bill. We shall eliminate this ambiguity by deriving the same conclusion from the formula

$$P(V \times W) = PV \times PW, \tag{4}$$

valid for any rotation P. To verify (4) we recall that $\det\big[BY, BV, BW\big] = \det B \det\big[Y, V, W\big]$ for any square matrix B. With $B = P$ and $\det P = 1$, this yields the first of the following equalities:

$$PY \bullet PV \times PW = Y \bullet V \times W = PY \bullet P(V \times W),$$

the second one being due to the fact that P preserves the dot product. Since we may choose Y so that $X = PY$ is any desired vector (e.g. any row of I), this proves (4).

To identify $V \times W$, we choose P so that $PV = (x_1, 0, 0)$ and $PW = (x_2, y_2, 0)$, with $x_1, y_2 \geq 0$. In other words, we rotate V into alignment with the positive x-axis, and then rotate W into the x, y-plane with positive y-coordinate. If we now subject $V \times W$ to the same rotation, (4) says that we obtain $PV \times PW = (0, 0, x_1y_2)$. This not only fits the description given above (see?) but specifies a unique direction for the cross-product, namely the one corresponding to the positive z-axis.

This is often referred to as the "right hand rule", but the game works just as well with left hands if your original x, y, z-axes were oriented accordingly.

20. Orthogonalization. While still in the visible world of three dimensions, we shall describe another matrix manoeuvre with geometric overtones, namely the *Gram-Schmidt orthogonalization process*. Given $A = \big[V_1, V_2, V_3\big]$, considered as a triple of columns, we shall find a factorization

$$A = \big[V_1, V_2, V_3\big] = \big[W_1, W_2, W_3\big] \begin{bmatrix} 1 & t_{12} & t_{13} \\ 0 & 1 & t_{23} \\ 0 & 0 & 1 \end{bmatrix} = GS, \tag{1}$$

where $G = \big[W_1, W_2, W_3\big]$ is *column-rectified*, i.e. its columns are mutually orthogonal. We shall see that there is only one way to define these new columns; in other words, G is uniquely determined by A.

If we read (1) column by column, we get the following equations for each successive W_k:

$$W_1 = V_1, \quad W_2 = V_2 - t_{12}W_1, \quad W_3 = V_3 - t_{13}W_1 - t_{23}W_2. \tag{2}$$

Each time the new vector W_k is expressed in terms of the old V_k and the previously found W_i, with $i < k$. In every case, the orthogonality conditions $W_i \bullet W_k = 0$ obviously require that

$$W_i \bullet V_k = t_{ik} W_i \bullet W_i, \tag{3}$$

and this determines t_{ik} unless $W_i = 0$; but even then $t_{ik}W_i$ is clearly unique. Hence there is exactly one way of choosing each W_k. Looking back at lesson 8, we see that in fact

$$W_1 = V_1, \quad W_2 = V_2 - \mathrm{proj}_{W_1}(V_2), \quad W_3 = V_3 - \mathrm{proj}_{W_1}(V_3) - \mathrm{proj}_{W_2}(V_3), \tag{4}$$

which is a better way to remember this process.

Note that G will be singular if and only if A is, since S is invertible and hence $X \neq 0 \iff SX \neq 0$.

Least Squares. Given a single column C, consider the two equations

$$\text{(i)} \quad AX = C, \qquad\qquad \text{(ii)} \quad A^TAX = A^TC, \tag{5}$$

to be solved for the unknown X. If A (and hence A^T) is invertible, they are of course equivalent. The second one, obtained from the first one by multiplication with A^T, is called the *normal* equation for (i). It has the advantage of being *always solvable* — even if (i) is not. In that case a solution of (ii) is moreover a *best possible approximation* to a solution of (i). To handle (ii) and to understand its meaning, it is very convenient to work with the factorization $A = GS$ described above.

Initially the normal equation (ii) is a mess: $S^TG^TGSX = S^TG^TC$. But S^T is invertible and can be cancelled, leaving $DSX = G^TC$, where D is the non-negative diagonal matrix G^TG. It all reduces to

$$DY = G^TC \qquad \text{and} \qquad SX = Y. \tag{6}$$

If we can solve the first of these for Y, the second one (for X) will be a cinch because of the special nature of S. But the first one can *always* be solved: wherever D may have a zero row, G^T is sure to have one too (see?). After all, the diagonal entries of D are just $|W_k|^2$, where $G = \big[W_1, W_2, W_3\big]$.

Having seen how to get a solution of (ii), you may wonder whether it has any relevance to the original problem (i). The answer is: yes, it is a "least squares" approximate solution of (i), in the sense that it makes the difference $|AX - C|$ as small as possible.

In fact, the normal equation says literally that $A^T(AX - C) = 0$, i.e. that $(AX - C) = 0$ is orthogonal to every column of A. But then it is also orthogonal to every vector of the form AZ (see?). Hence

$$|A(X + Z) - C|^2 = |AX - C|^2 + |AZ|^2, \tag{7}$$

by Pythagoras. This shows that, for *any* other solution candidate $X' = X + Z$, the difference $AX' - C$ is at least as bad as our $AX - C$.

E. The QR-Factorization. In your further study of matrices, you are likely to run into the so-called QR-factorization. It will probably remind you of the Gram-Schmidt process, and make you wonder what the difference is between the two techniques. The present page is intended to help you sort this out.

Suppose that $A = GS$ is the factorization described in lesson 20, with the unique column-rectified G, and the "unipotent" upper triangular S. Taking a cue from lesson 17, we write $G = QD_0$, where Q is an appropriate orthogonal matrix, and D_0 is non-negative diagonal. Q consists essentially of columns of G scaled to unit length (supplemented by others wherever G has a zero column). Hence

$$A = QD_0S \tag{1}$$

can also be expressed as $A = QR$, with $R = D_0S$ still upper triangular and having non-negative diagonal entries.

If A is non-singular, all the factors in (1) are unique, and the distinction between GS and QR is only a question of bracketting. However, in QR it is the *second* factor R which turns out to have an intrinsic meaning and is therefore uniquely determined (while the Q may be varied somewhat if A is singular).

To see this meaning, let X_1, X_2, X_3 be the columns of R. Since Q is an isometry, all lengths and angles between the X_k are exactly as those between the corresponding V_k. Moreover X_1 is aligned with the positive x-axis, X_2 with the y-positive half of the x, y-plane, and X_3 with the z-positive half of space. If you picture $R = \big[X_1, X_2, X_3\big]$ as a kind of tripod, you will find that its position is completely determined hereby. Thus R is derived from the original A by *isometrically realigning its columns with the coordinate axes* as much as possible.

Writing $P = Q^T$, this factorization takes the form

$$PA = \begin{bmatrix} x_1 & x_2 & x_3 \\ 0 & y_2 & y_3 \\ 0 & 0 & z_3 \end{bmatrix} = R, \tag{2}$$

reminiscent of the LU-factorization obtained by Gaussian elimination: in both instances the given matrix A is transformed into triangular shape by an easily invertible left multiplier. Just as in lesson 11, this multiplier (here P) can be built up by a step-wise procedure which "sweeps out" the subdiagonal portion of one column after another (here by reflections instead of by shears). It is based on the simple observation that *given two vectors V and X of equal length, the reflection along $X - V$ will transform V into X.*

Now take V_1 to be the first column of A, let $x_1 = |V_1|$ be its length, and put $X_1 = (x_1, 0, 0)$. From these data, it is easy to work out the reflection along $X_1 - V_1$ using the very explicit formula (2) of lesson 18. Since this is the first of several steps, we label this reflection H_1 and get

$$H_1A = \begin{bmatrix} x_1 & * & * \\ 0 & a' & b' \\ 0 & c' & d' \end{bmatrix} = \begin{bmatrix} x_1 & * \\ & \\ 0 & A' \end{bmatrix} \tag{3}$$

where the asterisks denote unspecified entries. Next, we leave the first row alone and repeat the same trick on the 2×2-matrix A'. This yields a second reflection H_2, and H_2H_1A is already upper triangular (see?) with the first two diagonal entries non-negative. If the third diagonal entry happens to be negative at this point, we apply the reflection $D_3(-1)$ to finish up.

This is an efficient and useful algorithm in practice, but not in school: the square roots occurring right from the start (in computing lengths) ruin all efforts at maintaining the "round numbers" essential for exams. If we ever need QR in this course, we shall compute it via GS.