Bringing the inner product out

B.Sury

Stat-Math Unit
Indian Statistical Institute
Bangalore

Workshop on linear algebra and analysis

University of Hyderabad

September 14-17, 2006

The focus of these notes is on the uses of inner product spaces in mathematics. There is more material here than what we would be able to discuss during the workshop; it has been added for the sake of completeness and with the hope that it could be useful in teaching. Let me add that the subject is quite geometric and, though we have not been able to draw pictures here, while lecturing one ought to draw pictures and show clearly what the various algebraic formulae really mean geometrically; the books [A] and [K] are good references. As an appendix, we have added a few simple applications of basic linear algebra to some other subjects and it may be worthwhile to motivate students by means of such examples.

Let us start with simple well-known observations. I recall them in order to naturally lead to the later discussions. It is generally seen that the more the entries of a matrix are zero, the easier it is. Diagonal (square) matrices have several zero entries and also commute among themselves; they are just as easy to work with as if they were numbers. So, if a matrix can be 'reduced' to a diagonal matrix in some natural algebraic fashion, it ought to be useful. One such natural reduction might be re-writing the matrix in terms of a changed basis; this amounts to conjugating the matrix by an invertible matrix. That is to say, given a matrix A, we look for an invertible matrix P so that PAP^{-1} is a diagonal matrix. The word 'conjugate' used here should not be confused with complex conjugates; in the old days, one used to call our 'conjugation' of matrices as 'similarity'. The modern notation here is consistent with group theory.

The key point about conjugating into a diagonal matrix is that there is at most one diagonal matrix (upto permuting the entries) to which A can be conjugate although there may be several conjugating matrices P. Often it is not necessary to actually find a P but only 'the' diagonalization of A. Here is a situation where it is useful to find a diagonalizing matrix P also.

If $dY/dt = AY$ is a system of ordinary differential equations to be solved, one tries to diagonalize A if possible. If $A = PDP^{-1}$ with D diagonal, then $Y = PX$ gives the new system $dX/dt = DX$. This is easy to solve, and one can get back Y as $Y = PX$.

The first simple observation is that not every complex matrix can be diagonalized; an example is $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. The proof follows basically from the easy observations that conjugate matrices have the same eigenvalues and the eigenvalues of an upper triangular (or similarly lower triangular) matrix are

its diagonal entries. Note that this also shows that a diagonalization of a matrix (if one exists at all) is unique (upto permuting the entries). So, if the above matrix were conjugate to a diagonal matrix, that diagonal matrix must be the zero matrix. But, evidently the only conjugate of the zero matrix is the zero itself and our matrix is not the zero matrix.

Notice also that if we start with a real matrix A which can be diagonalized, the diagonal matrix may only be a (non-real) complex matrix since the eigenvalues of a real matrix may be non-real. For example, the real matrix $\begin{pmatrix} Cos(\theta) & Sin(\theta) \\ -Sin(\theta) & Cos(\theta) \end{pmatrix}$ is conjugate to the diagonal matrix $\begin{pmatrix} exp(i\theta) & 0 \\ 0 & exp(-i\theta) \end{pmatrix}$ which is not real for $\theta \neq n\pi$.

The importance of eigenvalues in the above discussion and the fact that upper (or lower) triangular matrices advertise their eigenvalues clearly begs that next question as to whether each matrix can be conjugated to a (complex) triangular matrix. This is true in a stronger form as shown by Schur in the theorem below.

Before stating Schur's result, we remark on another use of knowing that matrices can be triangularised. Let us say, we need to compute the matrix obtainable as a polynomial expression $f(t)$ evaluated on A. If PAP^{-1} were a diagonal matrix with entries $a_1, \cdots, a_n$, then

$$Pf(A)P^{-1} = f(PAP^{-1}) = diag(f(a_1), \cdots, f(a_n)).$$

Even if PAP^{-1} were only a triangular matrix with diagonal entries $a_1, \cdots, a_n$, the eigenvalues of $f(A)$ are $f(a_1), \cdots, f(a_n)$. In fact, if we introduce a notion of size/norm for a matrix, then one can talk about convergent power series. A typical example is that for any matrix A, there is a well-defined matrix called its exponential $exp(A)$. The triangularisability shows that the eigenvalues of $exp(A)$ are e^{λ_i} for various eigenvalues λ_i of A. In particular, the determinant of $exp(A)$ is $\prod_i exp(\lambda_i) = exp(\sum_i \lambda_i) = e^{tr(A)}$, where $tr(A)$ is the trace of A. The exponential mapping on matrices plays a crucial role in Lie group theory. Later, we will see some density results follow easily because of triangularisation.

Theorem (Issai Schur) :
Every complex square matrix can be conjugated by means of a unitary matrix to an upper triangular matrix.

Recall that an $n \times n$ unitary matrix U is one for which $\overline{{}^tU} = U^{-1}$. A standard notation is to write B^* for $\overline{{}^tB}$. Thus, a unitary matrix satisfies $U^* = U^{-1}$. Note that $(BC)^* = C^*B^*$.

Proof.

We will postpone reading this proof until after we have properly introduced the inner product spaces.

If A is any $n \times n$ matrix and U is any invertible $n \times n$ matrix, then UAU^{-1} is the matrix (for the same linear transformation given by A) with respect to the new ordered basis $Ue_1, \cdots, Ue_n$, viz., the columns of U. Note that a matrix U is unitary if, and only if, its columns form an orthonormal basis of $\mathbf{C}^n$ with respect to the canonical inner product $x.y = \sum_{i=1}^n x_i\bar{y}_i$. In other words, the columns $C_1, \cdots, C_n \in \mathbf{C}^n$ satisfy $C_i.C_i = 1$ and $C_i.C_j = 0$ for $i \neq j$. Therefore, given A, the assertion of the theorem is equivalent to choosing an orthonormal basis $v_1, \cdots, v_n$ of $\mathbf{C}^n$ such that Tv_i is a linear combination of $v_1, \cdots v_i$; here, we have written T for the linear transformation corresponding to A. One advantage of phrasing the assertion in this form is one can easily apply induction on n. To start with, for $n = 1$, the assertion of the theorem is a tautology. So, assume $n > 1$ and that the assertion holds for $n - 1$. Choose an eigenvalue of T and let w_1 be a corresponding eigenvector. A non-zero multiple of w_1 is again an eigenvector and we take $v_1 = \frac{w_1}{\sqrt{w_1.w_1}}$ so that we have $v_1.v_1 = 1$ and $Tv_1 = \lambda_1 v_1$. If $M = \mathbf{C}v_1$ is the one-dimensional subspace generated by v_1, then $N = \{x \in \mathbf{C}^n : v_1.x = 0\}$ is the orthogonal complement to M; in particular, it is a subspace of $\mathbf{C}^n$ such that $M+N = \mathbf{C}^n$ and $M \cap N = (0)$. Let us call P_N the projection map

$$tv_1 + x \mapsto x$$

from $\mathbf{C}^n = M \oplus N \to N$. The transformation $x \mapsto P_N T(x)$ is linear and maps N into N. By the induction hypothesis, there is an orthonormal basis $v_2, \cdots, v_n$ of N such that $P_N T(v_i)$ is a linear combination of $v_2, \cdots, v_i$ for each $i \geq 2$. Since the projection map P_N has precisely the one-dimensional space generated by v_1 as its kernel, and since it is the identity map on N, this means that for $i \geq 2$, the vector $T(v_i)$ is a linear combination of $v_1, \cdots, v_i$. As $T(v_1)$ is a multiple of v_1, and, as evidently the vectors $v_1, v_2, \cdots, v_n$ form an orthonormal basis of $\mathbf{C}^n$, we are done. The result follows by induction.

Corollary (spectral theorem for normal operators) :
A is unitarily conjugate to a diagonal matrix if, and only if, it is normal

(that is, A and A^* commute).
We will recall it again later in a different notation after discussing the notion of an inner product space.

Proof.
The proof is simple. Observe that a matrix B is normal if, and only if, UBU^* is normal for any unitary U :

$$(UBU^*)(UBU^*)^* = UBU^*(U^*)^*B^*U^*$$

$$= UBB^*U^* = UB^*BU^* = (UBU^*)^*(UBU^*).$$

As an upper triangular matrix is normal if, and only if, it is diagonal, the result follows.

To prove Schur's triangularisation on conjugation by unitary matrices, if we prove only triangularisation via conjugation by some arbitrary invertible matrix (by some other method, say), we can deduce Schur's theorem from that result using the Gram-Schmidt process which we will be recalling below.

Various applications of Schur's theorem :

(I) Define a notion of distance between matrices as $d(A, B) = |A-B|_2$ where

$$|M|_2 := (tr(M^*M))^{1/2} = (\sum_{i,j} |m_{ij}|^2)^{1/2}.$$

This makes it clear that $|M|_2 = |U^*MU|_2$ for any unitary matrix U. This notion makes the set $M(n)$ of all $n \times n$ matrices over $\mathbf{C}$ into a metric space. It is easy to see that the triangle inequality holds. Let us deduce now using Schur's theorem that the set of invertible matrices is 'dense' in $M(n)$. We start with any matrix A and show that for any $\epsilon > 0$, there is an invertible matrix B such that $|A - B|_2 < \epsilon$. Write $A = UTU^*$ where T is upper triangular and U is unitary. Replace those diagonal entries of T (= eigenvalues of A) which are zero by non-zero complex numbers which have small absolute values. Let the other entries of T remain as they are. We get an invertible, upper triangular matrix T_1 with $|T - T_1|_2$ as small as we want. Therefore, UT_1U^* is an invertible matrix as well. Moreover,

$$|A - UT_1U^*|_2 = |U(T - T_1)U^*|_2 = |T - T_1|_2.$$

This is as small as we please. Therefore,
invertible matrices are dense in all matrices.
Similarly, one can prove :
matrices with distinct eigenvalues are dense.

(II) The Cayley-Hamilton theorem asserts that for a polynomial $A \in M(n)$, if $\chi_A(T) := det(TI_n - A) = c_0 + c_1T + \cdots + T^n$ is its characteristic polynomial, then the matrix $\chi_A(A) := c_0I_n + c_1A + \cdots + c_{n-1}A^{n-1} + A^n$ is the zero matrix. Once again, this can be easily proved using Schur's theorem as we indicate now. If A is not invertible, get an invertible matrix B close to it and note that the coefficients of $\chi_A(T)$ and $\chi_B(T)$ are close. Also, then the matrices $\chi_A(A)$ and $\chi_B(B)$ are as close as we want. Thus, it suffices to prove the Cayley-Hamilton theorem for invertible matrices A. Also, one can take an invertible matrix with distinct eigenvalues which is close to A and it suffices to assume A has distinct eigenvalues. But then Schur's theorem allows us to diagonalize A by conjugation. Since $\chi_A(T)$ does not change under conjugation, it suffices to verify Cayley-Hamilton for diagonal matrices, where it is obvious !

We will now discuss the theory of inner product spaces and return again to the spectral theorem for normal operators, thereby understanding it better and putting it in perspective. The theme we wish to bring out is that the geometry of real/complex vector spaces is dramatically revealed by the notions of inner product and operators on them. One can prevail upon the inner product spaces to do a lot of work for us. This is especially so in infinite-dimensional spaces but their importance in finite-dimensional spaces cannot be undermined. We start with the definitions first.

Definition and examples of inner product spaces :
Let K stand for $\mathbf{R}$ or $\mathbf{C}$. If V is a vector space over K, an *inner product* on V is an association $V \times V \to K$,

$$(v, w) \mapsto < v, w >$$

which is additive in both variables, satisfies $< v, w >= \overline{< w, v >}$, $< v, v >$ is real and ≥ 0 with equality only for $v = 0$, and has the property $< av, w >= a < v, w >$ for all $a \in K$.
It is understood clearly that the map $a \mapsto \bar{a}$ is just the identity map when K is $\mathbf{R}$. Note that even for complex K, the numbers $< v, v >$ are real. The

above definition is modelled after the canonical inner product on K^n :

$$< (v_1, \cdots, v_n), (w_1, \cdots, w_n) >= \sum_{i=1}^{n} v_i \bar{w}_i.$$

Once we have chosen an inner product, we call the space an inner product space. Note that a subspace of an inner product space is again an inner product space under the same inner product. There are other natural interesting inner product spaces, one of which we recall now.
On the space $C([0,1], \mathbf{K})$ of all continuous K-valued functions, define

$$< f, g >= \int_0^1 f(t)\overline{g(t)}dt.$$

Of course, this vector space is infinite-dimensional but for a finite-dimensional example, one merely has to take the same definition on polynomial functions of degree bounded by a fixed N.
Also, one can give other inner products on the same space; for example, on $C([0,1], \mathbf{R})$, one could define

$$< f, g >= \int_0^1 t^2 f(t)g(t)dt.$$

We will consider only finite-dimensional inner product spaces most of the time unless we say otherwise explicitly.
Inner products allow us to define the notion of length and angle. If v, w are vectors in an inner product space $(V, <, >)$, then the *length of* v is defined to be $\sqrt{< v, v >}$; one writes it as $||v||$. If V is a real inner product space, one defines the angle θ between v and w by $< v, w >= ||v||||w||Cos(\theta)$. Thus, v, w are *orthogonal* to each other if $< v, w >= 0$. The familiar Pythagoras theorem which was also known in the Sulvasutras (and probably in other old civilizations as well) can be generalized to the statement :
If $v_1, \cdots, v_n$ *are pairwise orthogonal, then* $||v_1 + v_2 + \cdots v_n||^2 = \sum_{i=1}^{n} ||v_i||^2$.
Note as a consequence that such a set of pairwise orthogonal, non-zero vectors as above is linearly independent.

The following two simple properties are the most-used in mathematical situations where an inner product appears.

Cauchy-Schwarz inequality :
$| < v, w > | \leq ||v||\ ||w||$ *where equality holds exactly in the case when the vectors* v, w *are linearly dependent.*

Triangle inequality :
$||v + w|| \leq ||v|| + ||w||$ *where equality holds exactly in the case when one of the vectors is a non-negative multiple of the other.*

One calls an *orthonormal set*, a set of vectors $v_1, \cdots, v_n$ which are pairwise orthogonal and each of which has unit length. If the set is a vector space basis too, then it is called an orthonormal basis. The canonical basis $e_1, \cdots, e_n$ of K^n is an orthonormal basis. Just as one has in this space, $v = (a_1, \cdots, a_n) = \sum_{i=1}^n a_i e_i$ and $||v||^2 = \sum_{i=1}^n |a_i|^2$, one has in any inner product space :

If $v_1, \cdots, v_n$ *is an orthonormal basis, then for each* v*, we have* $v = \sum_{i=1}^n < v, v_i > v_i$ *and* $||v||^2 = \sum_{i=1}^n | < v, v_i > |^2$.

It would be well to know that every inner product space does have an orthonormal basis. Even better, it would be nice if one started with an arbitrary finite, linearly independent set S (in a possiblt infinite-dimensional inner product space) and produced an orthonormal set by some procedure which generates the same subspace which is generated by S. The following famous result showed how to do this by means of an algorithm.

Gram-Schmidt process :
Let $< V, <, >)$ *be a (possibly infinite-dimensional) inner product space. If* $\{v_1, v_2, \cdots, v_n\}$ *is any set of linearly independent vectors, the following algorithm produces an orthonormal set* $\{w_1, w_2, \cdots, w_n\}$ *of vectors such that the subspace generated by* $v_1, \cdots, v_r$ *equals that generated by* $w_1, \cdots, w_r$*, for all* $r = 1, \cdots, n$ *:*
$w_1 = \frac{v_1}{||v_1||}$, $w_{r+1} = \frac{v_{r+1} - \sum_{i=1}^r < v_r, w_i > w_i}{||v_{r+1} - \sum_{i=1}^r < v_r, w_i > w_i||}$.
In particular, when V *has finite dimension, then any basis can be reduced to an orthonormal basis.*

The Gram-Schmidt process actually shows that there is a homeomorphism

$$GL(n, \mathbf{R}) \equiv O(n) \times B$$

where $O(n)$ denotes the set of orthogonal matrices and B is the set of real

upper triangular invertible matrices having positive diagonal entries. In other words, the topology of $GL(n, \mathbf{R})$ is determined by the topology of a maximal compact subgroup since B is just homeomorphic to a Euclidean space. A similar theorem holds for many other groups of matrices.

Let us note :
A matrix is unitary (respectively, orthogonal) if and only if its columns/rows form an orthonormal basis of $\mathbf{C}^n$ *(respectively,* $\mathbf{R}^n$*).*

Corollary to Gram-Schmidt :
If T is a linear transformation on an inner product space $< V, <, >)$ and if T can be represented by an upper triangular matrix with respect to some ordered basis, then T can also be represented by an upper triangular matrix with respect to an orthonormal basis.
Proof.
Let $\{v_1, \cdots, v_n\}$ be an ordered basis with respect to which T is represented by an upper triangular matrix. By Gram-Schmidt algorithm, we get an orthonormal basis $\{w_1, \cdots, w_n\}$ such that, for every $r = 1, \cdots, n$, the subspace spanned by $v_1, \cdots, v_r$ is the same as that spanned by $w_1, \cdots, w_r$. As T transforms into itself the subspace spanned by $v_1, \cdots, v_r$ for each $r \leq n$, which is the same as that spanned by $w_1, \cdots, w_r$, the matrix of T with respect to the latter basis is upper triangular too.

Definition :
The *orthogonal complement* of a subset S of V is $S^\perp := \{v \in V :< s, v >= 0 \ \forall \ s \in S\}$. Note that $S^\perp$ is always a vector subspace of V even if S is just a set. Indeed, $S^\perp = W^\perp$, where W is the subspace of V spanned by S. The following simple result is extremely useful :

Proposition (orthogonal projection) :
Let $(V, <, >)$ be an inner product space (possibly infinite-dimensional). Then, for any finite-dimensional subspace W of V, we have $V = W \oplus W^\perp$.
One calls the canonical map from V to W given by the above decomposition as the orthogonal projection onto W.
Proof.
Let $\{w_1, \cdots, w_m\}$ be an orthonormal basis of W. Clearly, for any $v \in V$ we

have

$$v = \sum_{i=1}^{m} < v, w_i > w_i + (v - \sum_{i=1}^{m} < v, w_i > w_i).$$

Since the second summand is orthogonal to each w_i, it is in $W^\perp$. Hence $V = W + W^\perp$. If there is a vector $w \in W \cap W^\perp$, then $< w, w >= 0$ which gives $w = 0$.

Now is an appropriate point to read Schur's proof.

Corollary to proposition above :
Any subspace W of V satisfies $W \subseteq (W^\perp)^\perp$. Further, this is an equality when W has finite dimension.
Proof.
Let W be arbitrary, let $w \in W$. As w and $W^\perp$ are orthogonal, the vector w is in the orthogonal complement of $W^\perp$, which is the first assertion.
Conversely, assume W has finite dimension; the proposition is applicable. Let $w \in (W^\perp)^\perp$; writing $w = w_1 + w_2 \in W \oplus W^\perp$, we have $w - w_1 \in W \subseteq (W^\perp)^\perp$ whereas $w_2 \in W^\perp$. Thus, this must be the zero vector; so $w = w_1 \in W$.

Proposition (projection gives closest vector) :
If W is a subspace of $(V, <, >)$, then the vector in W closest to a given vector v of V is $P_W v$ where $P_W : V \to W$ is the projection onto W.
Proof.
Let $w \in W$ be arbitrary. Now,

$$||v - P_W v||^2 \leq ||v - P_W v||^2 + ||P_W v - w||^2 = ||v - w||^2$$

since $v - P_W v$ is orthogonal to W and, therefore, to $P_W v - w$.
Note that equality occurs then and only then, when the triangle inequality used is an equality, which gives $P_W v = w$. This justifies the word 'the' in the phrase 'the closest vector'.

One way to compute the projection map onto W is to take an orthonormal basis $\{w_1, \cdots, w_r\}$ of W; then for any $v \in V$, it is evident that

$$P_W v = \sum_{i=1}^{r} < v, w_i > e_i.$$

The proposition above is often used for approximation problems. We have used it in the analysis lectures to show mean square convergence of the Fourier

series of a continuous function. To see how the least squares approximation is the best in terms of having minimal variance (and for several other facts also), please see Meyer's book $[M]$ available online. Let us see another type of example now.

A typical application to polynomial approximation :
Let us try to find the real polynomial of degree ≤ 5 which is closest to $Sin(x)$ on the interval $[-\pi, \pi]$ where the inner product on $C([-\pi, \pi], \mathbf{R})$ is $< f, g >:= \int_{-\pi}^{\pi} f(x)g(x)dx$. Therefore, we let $V = C([-\pi, \pi], \mathbf{R})$, W the subspace of polynomials. Note that V is infinite-dimensional but W has dimension 6 and the proposition on closest vectors, is applicable. To compute the projection map onto W, we need an orthonormal basis of W. For this, we start with the basis $\{1, x, x^2, x^3, x^4, x^5\}$ and apply the Gram-Schmidt process. After some computation, it turns out that the closest polynomial in W to $Sin(x)$ is

$$0.98786x - 0.155271x^3 + 0.00564312x^5$$

where the coefficients are written approximately after replacing π etc. approximately.
It is interesting to compare this approximation with the Taylor polynomial approximation $x - \frac{x^3}{6} + \frac{x^5}{120}$. It turns out that the former is much better than this one when $|x| > 2$.

Finite-dimensional Riesz representation theorem.
Let $(V, <, >)$ be a finite-dimensional inner product space. Let $T : V \to K$ be linear (T is called a linear functional). Then, there exists a unique $v_0 \in V$ such that $T(v) =< v, v_0 >$. In other words, each linear functional is represented by a vector.
Proof.
Start with an orthonormal basis $\{v_1, \cdots, v_n\}$ of V. Any v can be written as $\sum_{i=1}^{n} < v, v_i > v_i$; so

$$T(v) = \sum_{i=1}^{n} < v, v_i > T(v_i) = \sum_{i=1}^{n} < v, \overline{T(v_i)}v_i > .$$

Clearly, $v_0 := \sum_{i=1}^{n} \overline{T(v_i)}v_i$ does the job.
To prove uniqueness of v_0, note that if there are two such vectors v_0, w then $< v, v_0 - w >= 0$ for all $v \in V$. Applying this to $v = v_0 - w$ yields $v_0 = w$.

We remark that the Riesz representation theorem above can fail for infinite-dimensional inner product spaces. For instance, on $C([0,1],\mathbf{R})$, for any fixed point $x_0 \in [0,1]$, the linear functional *evaluation at* x_0 : $f \mapsto f(x_0)$ is not representable by a vector as above.

Definition :
If V, W are two finite-dimensional inner product spaces, and $T \in Hom(V,W)$ a linear transformation, the *adjoint* of T is defined to be the linear transformation $T^* \in Hom(W,V)$ so that

$$< Tv, w >=< v, T^*w > \ \forall \ \ v \in V, w \in W \cdots\cdots (A)$$

The existence of a map from W to V with this latter property is guaranteed by the previous theorem applied to the linear functional

$$T_w : V \to K \ , \ v \mapsto < Tv, w > .$$

That T^* so defined is linear is easily seen from the properties of the inner products. It should be noted that the two sides of (A) involve inner products from two different spaces although we have not explicitly indicated.
The adjoint also has the following properties verified by first principles :
$(T^*)^* = T$, $(S+T)^* = S^* + T^*$, $(cT)^* = \bar{c}T^*$, $(ST)^* = T^*S^*$.
$KerT^* = (ImT)^{\perp}$, $ImT^* = (KerT)^{\perp}$.

Lemma (matrix of adjoint) :
Let $\{v_1, \cdots, v_n\}$ *and* $\{w_1, \cdots, w_m\}$ *be ordered orthonormal bases of* V *and* W *respectively. If* $T \in Hom(V,W)$ *is represented by a matrix* A *with respect to these ordered bases, then the adjoint* T^* *is represented by the matrix* $A^* = \bar{A}^t$.
Proof.
Now, since $Tv_i = \sum_j < Tv_i, w_j > w_j$, the matrix A is given by $a_{ji} =< Tv_i, w_j >$. Similarly, the matrix B of T^* is $b_{ij} = \sum_j < T^*w_j, v_i >$. But

$$a_{ji} =< Tv_i, w_j >=< v_i, T^*w_j >= \overline{< T^*w_j, v_i >} = \overline{b_{ij}}.$$

So $\bar{A}^t = B$.

On infinite-dimensional inner product spaces, adjoints may not exist. For instance, on $C([0,1],\mathbf{R})$, the endomorphism $f(t) \mapsto tf(t)$ does not have an adjoint.

Definitions :
For inner product spaces V, W, and a linear transformation $T \in Hom(V, W)$, the *norm of* T - denoted by $||T||$ - is defined as

$$||T|| = Sup\{||Tv|| : ||v|| = 1\}.$$

This is always finite as the unit ball is compact and linear transformations are continuous.
One calls $T \in Hom(V, W)$ an *isometry* into W, if $< Tv_1, Tv_2 >=< v_1, v_2 >$ for arbitrary $v_1, v_2 \in V$. If T is also an isomorphism onto W, then one calls it an *isometric isomorphism*. Note that T is an isometry if, and only if, it preserves norms - for the 'if' part, just apply to the vectors of the form $v_1 + v_2$.

Proposition (norm-preserving implies unitary) :
For V, W are inner product spaces, a transformation $T \in Hom(V, W)$ is norm-preserving if, and only if, it is unitary (that is, $T^*T = I_V$).
Proof.
If $T^*T = I$, then clearly T preserves norms because

$$||Tv||^2 =< Tv, Tv >=< v, T^*Tv >=< v, v >= ||v||^2.$$

Conversely, if T preserves norms, then as we observed, it is an isometry. Therefore, for every $v_1, v_2 \in V$, we get

$$< v_1, v_2 >=< Tv_1, Tv_2 >=< v_1, T^*Tv_2 >,$$

which gives $< v_1, (I - T^*T)v_2 >= 0$. Taking v_2 arbitrary and v_1 to be the vector $(I - T^*T)v_2$, we have $T^*T = I$.

Remark:
Note that a matrix $A \in M_n(\mathbf{C})$ is unitary if, and only if, $||Av|| = ||v||$ for all v; that is, columns of A form an orthonormal basis of $\mathbf{C}^n$ for the canonical inner product.

Definitions and observations :
If T is a linear transformation from V to itself, then it is said to be *self-adjoint* if $T = T^*$. In terms of a fixed choice of ordered basis, the corresponding matrix A of T must satisfy $A = A^*$.
More generally, one calls T *normal*, if $TT^* = T^*T$. Note that self-adjointness

implies normality.
It can be checked that :

Lemma :
(a) *Eigenvalues of a self-adjoint T are all real; eigenvectors corresponding to distinct eigenvalues are orthogonal.*
(b) *T is normal if and only if $||Tv|| = ||T^*v||$ for all v.*
Proof.
(a) If $Tv = \lambda v$ with $v \neq 0$ and $T = T^*$, then

$$< Tv, v >= \lambda < v, v >$$

and

$$< v, Tv >=< v, \lambda v >= \overline{\lambda} < v, v > .$$

Selfadjointness of T implies these are equal; that is, λ is real.
Also, if $Tv_1 = \lambda_1 v_1$ and $Tv_2 = \lambda_2 v_2$ with $v_1, v_2 \neq 0$ and $\lambda_1 \neq \lambda_2$, then

$$\lambda_1 < v_1, v_2 >=< Tv_1, v_2 >=< v_1, Tv_2 >= \lambda_2 < v_1, v_2 >$$

which gives $< v_1, v_2 >= 0$.
(b) We have

$$||Tv||^2 =< Tv, Tv >=< v, T^*Tv >$$

and

$$< v, TT^*v >=< T^*v, T^*v >= ||T^*v||^2$$

where we have used the fact that $(T^*)^* = T$. This proves (b).

Rephrasing normal spectral theorem :
We saw earlier that Schur's unitary triangularisation theorem implied the spectral theorem for normal matrices; this can be rephrased as :
Given a normal operator on a finite-dimensional complex inner product space, there is an orthonormal basis of eigenvectors.

Caveat on real inner product spaces
We stress that the above spectral theorem for normal operators is for a complex inner product space. It is *not true* in real inner product spaces. Over a complex inner product space, normality is equivalent to the existence of an orthonormal basis consisting of eigenvectors. Over a real inner product

space, it turns out (the proof is on similar lines to the complex case) that self-adjointness is equivalent to the existence of an orthonormal basis consisting of eigenvectors. In real inner product spaces, there are normal operators which are not self-adjoint. A typical example in a 2-dimensional inner product space is one we introduced right in the beginning; namely, $\begin{pmatrix} Cos(\theta) & Sin(\theta) \\ -Sin(\theta) & Cos(\theta) \end{pmatrix}$ where $Sin(\theta) \neq 0$.

Here are two interesting consequences of the normal spectral theorem rephrased above :

Lemma :

(a) *T is normal if, and only if, $T^* = f(T)$ for some polynomial f.*

(b) *For any T with eigenvalues $\lambda_1, \cdots, \lambda_n$, we have*

$$\sum_{i=1}^{n} |\lambda_i|^2 \leq \sum_{i,j=1}^{n} |a_{ij}|^2$$

with equality if and only if T is normal.

Proof.

(a) If T^* is a polynomial in T, it evidently commutes with T.

Conversely, assuming that T s normal, take an orthonormal basis $\{v_1, \cdots, v_n\}$ consisting of eigenvectors for T. Let λ_i be the eigenvalue corresponding to v_i. It is evident that the polynomial f which interpolates the values $\overline{\lambda_i}$ at the points λ_i satisfies

$$f(T)v_i = f(\lambda_i)v_i = \overline{\lambda_i}v_i = T^*(v_i).$$

Thus, $T^* = f(T)$.

(b) Note that if $T = (a_{ij})_{i,j}$, then $Te_i = \sum_{j=1}^{n} a_{ji}e_j$ implies

$$< Te_i, Te_i >= \sum_j |a_{ji}|^2.$$

Hence $\sum_{i=1}^{n} ||Te_i||^2 = \sum_{i,j=1}^{n} |a_{ij}|^2$.

If U is unitary with $U^*TU = S$, which is upper triangular, then

$$\sum_i ||Se_i||^2 = \sum_i ||U^*TUe_i||^2 = \sum_i ||Te_i||^2.$$

The right side is $\sum_{i,j} |a_{ij}|^2$ whereas the left side is $\geq \sum_i |\lambda_i|^2$ as the diagonal entries of S are the λ_i's.

Finally, note that equality is precisely when S is diagonal in which case T is normal.

Locating zeroes of polynomials :
We would like to briefly discuss a very good application of the theory of inner product spaces to locating zeroes of a polynomial. Recall that one connection between polynomials and matrices arises while finding the rational canonical form of a matrix A - the basic idea is to regard V as a module over the polynomial ring $\mathbf{C}[X]$ by means of $f(X)v := f(A)v$. The fact that $\mathbf{C}[X]$ is a Euclidean domain allows a very nice structure theorem for finitely generated modules. Thus, polynomials and matrices are intimately related. Let us see how this can be exploited for zero location.

Let us first ask the interesting question as to what the best way is, to view the dual space of the complex vector space $\mathbf{C}[X]$. This is given by the following lemma. Recall that $\mathbf{C}[[X]]$ denotes the set of formal power series in X and its quotient field is $\mathbf{C}((X))$, the set of formal sums of the form $\sum_{n=-r}^{\infty} a_n X^n$ for some $r \in \mathbf{Z}$. Note that the field $\mathbf{C}(X)$ can be thought of as a subfield of the field of *truncated Laurent series* $\mathbf{C}((X^{-1})) := \{\sum_{n=-\infty}^{r} a_n X^n : a_n \in \mathbf{C}, r \in \mathbf{Z}\}$, which is the quotient field of $\mathbf{C}((X^{-1}))$.

Lemma (dual of $\mathbf{C}(X)$) :
The dual space of $\mathbf{C}(X)$ *is* $X^{-1}\mathbf{C}[[X^{-1}]]$.
Proof.
For any $g = \sum_{n=-\infty}^{r} b_n X^n \in X^{-1}\mathbf{C}[[X^{-1}]]$, we have a linear functional on $\mathbf{C}[X]$:

$$\sum_{n=0}^{k} a_n X^n \mapsto \sum_{n=-\infty}^{\infty} a_n b_{-n-1}.$$

Conversely, let T be a any linear functional on $\mathbf{C}[X]$. For each $i \geq 0$, this induces linear functionals on $\mathbf{C}$ as :

$$T_i : a \mapsto T(aX^i).$$

We may consider T_i as complex numbers by the identification via the canonical inner product. If we take $g(X) = \sum_{n=0}^{\infty} T_i X^{-i-1} \in X^{-1}\mathbf{C}[[X^{-1}]]$, then we have

$$T(\sum_{n=0}^{k} a_n X^n) = \sum_{n=-\infty}^{\infty} a_n T_n.$$

Definition :
Since $\mathbf{C}((X^{-1})) = \mathbf{C}[X] \oplus X^{-1}\mathbf{C}[[X^{-1}]]$, the second projection

$$p_- : \sum_{n=-\infty}^{r} a_n X^n \mapsto \sum_{n=-\infty}^{-1} a_n X^n$$

is a well-defined linear transformation. For any $g \in \mathbf{C}((X^{-1}))$, the *Hankel operator* is defined as the map

$$H_g : \mathbf{C}[X] \to X^{-1}\mathbf{C}[[X^{-1}]]; f \mapsto p_-(gf).$$

We do not go into details here but the Hankel operators have several nice properties one of which we mention. For instance, Kronecker had proved that *H_g has finite rank if and only if g is a rational function.* We finish by stating a result on zeroes of polynomials. We do not give the proof here but it is not difficult and one can refer to Fuhrmann's book [F] referred to at the end.

Theorem (zero location) :
Let $p(X)$ be any polynomial in $\mathbf{C}[X]$. If $g = \frac{p'}{p}$, then the number of distinct roots of $p(X)$ equals the rank of the Hankel matrix H_g.

A dramatic application :
Though, by this time, no reader would need convincing that inner product spaces play key roles in many places, we state here a rather stunning 'avataar' of it.

Look at the inner product space $\mathcal{H}$ consisting of all sequences $a := \{a_n\}$ of complex numbers which satisfy $\sum_{n=1}^{\infty} \frac{|a_n|^2}{n(n+1)} < \infty$. Here, we take

$$< a, b >= \sum_{n=1}^{\infty} \frac{a_n \overline{b_n}}{n(n+1)}.$$

All bounded sequences are in $\mathcal{H}$. For $k = 1, 2, 3 \cdots$ consider the special elements $a(k) \in \mathcal{H}$ given by $a(k)_n = \{\frac{n}{k}\}$, the fractional part of $\frac{n}{k}$. Then, Baez-Duarte (in a formulation due to B.Bagchi) has proved :

Theorem :
The following statements are equivalent :

(a) *The Riemann hypothesis.*
(b) *The constant sequence* $1, 1, 1, \cdots$ *is in the closure of the space spanned by the* $a(k)$*'s;* $k = 1, 2, \cdots$.
(c) *The set of finite linear combinations of the* $a(k)$*'s is dense in* $\mathcal{H}$.

An application to cryptography -the LLL-algorithm :

This 1982 method due to A.K.Lenstra, H.W.Lenstra Jr. and L.Lovasz broke new ground and has proved a most influential method for computations in number theory - especially in factorisation of polynomials over $\mathbf{Z}$ or even over number fields. In simple terms, this method starts with a basis of a lattice and reduces it to a basis which is nearly orthogonal and whose vectors are 'shorter' in a sense. This reduction is managed by the LLL-algorithm in polynomial time. In some sense, the LLL method unwarps a badly warped basis. Let us be more precise now.
The set-up is as follows. In the case of a real inner product space $(V, <, >)$, one has the related notion of a positive-definite quadratic form; this is the quadratic function $v \mapsto < v, v >$. Below, we consider pairs $(L.q)$ where L is a lattice of rank n (that is a subgroup of the additive group $\mathbf{R}^n$ which contains a basis and is isomorphic to $\mathbf{Z}^n$, and q is a positive-definite quadratic form on $\mathbf{R}^n$. One may define an equivalence $(L, q) \sim (L', q')$ if there is an abelian group isomorphism between the lattices which respects the forms. As a positive-definite quadratic form on $\mathbf{R}^n$ gives rise naturally to a positive-definite symmetric matrix, the equivalence above can be expressed in terms of matrices as follows.
The equivalence classes (L, q) correspond bijectively with the classes of positive-definite symmetric matrices Q, where $Q \sim Q'$ if $Q' = {}^tMQM$ for some $M \in GL(n, \mathbf{Z})$.
Let $\{v_1, \cdots, v_n\}$ be a basis of $\mathbf{R}^n$. Consider the lattice L with this as $\mathbf{Z}$-basis. One calls the positive real number $|det(v_1, \cdots, v_n)|$ given by the absolute value of the determinant of the matrix with v_i's as columns to be the discriminant of L and denotes it by $disc(L)$. Note that a change of $\mathbf{Z}$-basis does not affect the discriminant as the determinant inside can change by ± 1 only. Now, the Gram-Schmidt process produces an orthogonal (not necessarily orthonormal) basis of V in the usual way :
$w_1 = v_1, w_i = v_i - \sum_{j<i} \mu_{ij} w_j$ where $\mu_{ij} = \frac{<v_i, w_j>}{<w_j, w_j>}$.
One defines the $\mathbf{Z}$-basis $\{v_1, \cdots, v_n\}$ of L to be *LLL-reduced* if :

(i) $|\mu_{ij}| \leq \frac{1}{2}$ for all $i > j$, and
(ii) $|w_i + \mu_{i,i-1}w_{i-1}|^2 \geq \frac{3}{4}|w_{i-1}|^2$ for all $i > 1$.
In what follows, the constant $\frac{3}{4}$ in (ii) can be replaced by any $t \in (\frac{1}{4}, 1)$. Note that (ii) is equivalent to $|w_i|^2 \geq (\frac{3}{4} - \mu_{i,i-1}^2)|w_{i-1}|^2$ for $i > 1$ and that the vectors $w_i + \mu_{i,i-1}w_{i-1}$ and w_{i-1} are the projections of v_i and v_{i-1} respectively, on the orthogonal complement of $\sum_{j<i-1} \mathbf{R}v_j$.

Proposition.
Let $\{v_1, \cdots, v_n\}$ be an LLL-reduced basis of L. With w_i's defined as above, we have :
(a) $|v_j|^2 \leq 2^{i-1}|w_i|^2$ *for* $j \leq i$,
(b) $disc(L) \leq \prod_i |v_i| \leq 2^{n(n-1)/4} disc(L)$,
(c) $|v_1| \leq 2^{(n-1)/4} disc(L)^{1/n}$, *and*
(d) *For* $0 \neq x \in L$, $|v_1| \leq 2^{(n-1)/2} max(|x|$.
If the constant $\frac{3}{4}$ in (ii) is replaced by some $t \in (\frac{1}{4}, 1)$, then all the powers of 2 in the proposition are replaced by the same powers of the number $\frac{4}{4t-1}$. Also, the inequality $disc(L) \leq \prod_i |v_i|$ is true for any (not necessarily LLL-reduced) basis and is known as Hadamard's inequality. The proof of the proposition is simple.
The reduction of any basis to an LLL-reduced basis can be desribed by an algorithm whose running time is $O(n^6(log(C)^3)$, where C is a bound for all $|v_i|$. In practice, it is often seen to take even less time.
Further, if the Gram matrix of the inner products $< v_i, v_j >$ of a basis $\{v_i\}$ is integral, the algorithm can be given in such a way that all computations are done in $\mathbf{Z}$ itself (and not go to $\mathbf{Q}$ as may be the case for a general basis).
The LLL algorithm does not give the shortest vector (this is a notoriously difficult problem) but one reasonably close to it.
One can adopt the LLL-algorithm to compute the kernel and image of an integral matrix also but the algorithm has to be modified to deal with dependent vectors also.
Let us see how LLL comes into the picture when we wish to factorise monic integral polynomials.

The basic result on which the algorithm is based is the following :
Let p be a prime, $k \in \mathbf{N}$, $f \in \mathbf{Z}[X]$ of degree $n > 0$, $h \in \mathbf{Z}[X]$ monic such that h mod p is irreducible, h mod p^k divides f mod p^k and (h mod p)^2 does not divide f mod p. Then, there is an irreducible factor $h_0 \in \mathbf{Z}[X]$ of f determined uniquely upto sign such that h mod p divides h_0 mod p. Furthermore, a factor g of f in $\mathbf{Z}[X]$ is divisible by h_0 in $\mathbf{Z}[X]$ if, and only

if, g mod p^k is divisible by h mod p^k. In particular, h_0 mod p^k is divisible by h mod p^k.

We would like to find a way to compute h_0 efficiently. To do this, one (starts with f, p, k, h as above and) fixes some $m \geq l := deg(h)$ and considers the lattice L consisting of all integral polynomials of degree $\leq m$ which are, mod p^k, divisible by h mod p^k. This is a lattice in the vector space $\mathbf{R}+\mathbf{R}X+\cdots+\mathbf{R}X^m$ which we can think of as $\mathbf{R}^{m+1}$. Note that the Euclidean length provides the notion of the length of a polynomial. That is, $|\sum_{i=0}^{m} a_i X^i| = (\sum |a_i|^2)^{1/2}$.
Observe that L has a basis $\{p^k, p^k X, \cdots, p^k X^{l-1}, h, Xh, \cdots, X^{m-l}h\}$.
Note that h_0 itself belongs to L if, and only if, $deg(h_0) \leq m$. Now, it can be checked that an element $b \in L$ satisfying the condition $|b|^n < \frac{p^{kl}}{|f|^m}$ is divisible by h_0 in $\mathbf{Z}[X]$ (this requires proof).
In particular, such an element b gives a factor $GCD(b, f)$ of f of degree > 1.
We choose $\{b_1, \cdots, b_{m+1}\}$ to be an LLL-reduced basis.
The factorisation of integral polynomials (upto factorisation of natural numbers) can be done in polynomial time with the LLL-algorithm.

An appendix : Some basic applications outside of mathematics

It is clear that linear algebra is central to mathematics and has an all-pervading role in it. However, it is often not easy to motivate a student of science as to why one should learn it. In what follows, we briefly indicate how useful it is in various branches of science at even a basic level but applications go much deeper than what we have indicated here. We have followed some discussions by Joseph Khoury of University of Ottawa, Canada who has won awards for his expositions on linear algebra. I urge everyone who has access to internet to look at a webpage maintained by Khoury for many more applications as well as to see simulated pictures of them.

Applications to chemistry.

Here is a typical way basic linear algebra is applied to a simple problem in chemistry.
Suppose we want to produce a certain chemical compound using 3 different ingredients A, B, C. Usually, one needs to dissolve each of these substances separately in water in various concentrations and then mix them so as to allow

chemical interaction and thereby produce the compound. For example, let us say that a certain amount of A is dissolved in water to make a solution having 1.5 gms. per cc. Similarly, say a certain amount of B and of C are separately dissolved in water to yield respective concentrations of 3.6 gms. per cc. and 5.3 gms. per cc. Suppose the three solutions totally give 25.07 gms. of the compound. Now, if the proportions of A, B, C (without changing the volumes of their solutions) are changed to $2.5, 4.3, 2.4$ gms. per cc. respectively, then let us say 22.36 gms. of the chemical compound are produced. Finally, if the proportions of A, B, C are $2.7, 5.5, 3.2$ gms. per cc. respectively, then suppose 28.14 gms. of the chemical compound are produced. We want to find the individual volumes of the 3 solutions.
Denoting by x, y, z the volumes of solutions containing A, B, C, this data leads to a system of linear equations

$$1.5x + 3.6y + 5.3z = 25.07$$

$$2.5x + 4.3y + 2.4z = 22.36$$

$$2.7x + 5.5y + 3.2z = 28.14$$

One can solve them (by Gaussian elimination, for example) to get

$$x = 1.5, y = 3.1, z = 2.2$$

Another typical application to chemistry is in the balancing of chemical equations. The basic scientific principle behind is the law of conservation of mass; thus, in any chemical equation, the total number of atoms must match.
For example, look at the equation

$$xC_2H_6 + yO_2 \to zCO_2 + tH_2O.$$

This leads to the system of linear equations

$$2x = z, 6x = 2t, 2y = 2z + t$$

which has the general solution

$$y = 7x/2, z = 2x, t = 3x.$$

Being numbers of atoms, the solutions need to be natural numbers. One has the balanced equation

$$2C_2H_6 + 7O_2 \to 4CO_2 + 6H_2O.$$

It should be remarked that there are important applications of representing abstract finite groups by groups of matrices to chemistry; for example, the enumeration of isomers is made using Polya's theory, which is on representing finite groups by groups of symmetries.

Applications to Genetics.

The great human genome mapping project has become common knowledge now. Suppose, we are interested in the type of genetic inheritance known as autosomal - those governed by a single gene. Let us say there are 2 types of genes A and a, and each individual carries a pair of genes - called his genotype. Thus, the possible genotypes for each inheritable trait are one of the three : AA, Aa, aa. Let us look at a specific problem. Suppose, in a certain animal population, the eye-colour is governed by autosomal model. Let the genotypes AA and Aa have brown eyes, while aa has black eyes. The A gene is said to dominate the a gene and an animal is called dominant, hybrid or recessive according as to whether it has AA, Aa or aa genes. Note that this means the genotypes AA and Aa cannot be distinguished in terms of eye-colour. Assume that each offspring inherits a gene from each parent in a random manner. The experiment we discuss is of crossing an offspring with a dominant animal (that is one with genotype AA). After repeating the experiment many times, we would like to know the proportions of each genotype. If AA is crossed with AA the result has to be AA. But, if AA is crossed with Aa, each of the two possibilities AA and Aa occurs with probability $1/2$ while aa cannot occur. Finally, crossing aa with AA produces Aa and no other genotypes. We wish to find the proportions of genotypes in the n-th generation once the initial proportions are given. Suppose that the initial proportions of genotypes are given as follows :

$$AA : 1/3, Aa : 1/3, aa : 1/3.$$

We will write this as a column vector X_0. By the discussion above, it is clear that the first generation has the column vector $X_1 = TX_0$, where T is the transition matrix $\begin{pmatrix} 1 & 1/2 & 0 \\ 0 & 1/2 & 1 \\ 0 & 0 & 0 \end{pmatrix}$. Indeed, the columns of T give the effects of crossing with AA, Aa, aa respectively ad the entries are the probabilities of obtaining AA, Aa, aa respectively.
In this manner, the vector X_n for the n-th generation is $X_n = TX_{n-1}$ for all

$n \geq 1$. The first few are :

$$X_1 = \begin{pmatrix} 1/2 \\ 1/2 \\ 0 \end{pmatrix}, X_2 = \begin{pmatrix} 3/4 \\ 1/4 \\ 0 \end{pmatrix}, X_3 = \begin{pmatrix} 7/8 \\ 1/8 \\ 0 \end{pmatrix}, \cdots$$

Since we have crossed offsprings of each generation only with genotype AA, the genotype aa never appears; that is, the 3rd entry of each column is 0. It is easy to show that the vectors $X_n \to \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ as $n \to \infty$. In other words, all the animals in the population would have brown eyes in the long run. Note that this last vector is an eigenvector with eigenvalue 1 for the transition matrix.
If we had always crossed with Aa, we can see that probabilities of offsprings of the n-th generation having brown eyes will be positive and less than 1 and so will be the case for black eyes. The best way to study these problems is using Markov chains.

Applications to image compression :

More than the other applications, this discussion is likely to fire the imagination of the student. When a digital image is to be viewed from a computer somewhere, the following technique called the Haar wavelet transform proves very useful in compressing it for the sake of storage space. Wavelets are orthogonal bases in certain inner product spaces. Roughly speaking, here is the idea. The image is treated as an array of numbers i.e., as a matrix. Each image consists of a large number of pixels (picture elements). The matrix corresponding to a digital image assigns a non-negative integer to each pixel; the numbers essentially codify the shades of black. The JPEG compression technique divides an image into 8×8 blocks and assigns a matrix to each block. One can use linear algebra to maximise compression while still retaining enough detail. This is how it works.
Suppose $r = (420, 680448, 708, 1260, 1420, 1600, 1600)$ is a row of an 8×8 image matrix. The transformation we will do is in 3 steps

$$r \mapsto r_1 \mapsto r_2 \mapsto r_3$$

(for the 2^n-size matrices, there would be n steps).
Step I :

Divide the entries of r into 4 pairs $(420, 680), (448, 708), (1260, 1420), (1600, 1600)$.
Find the 4 averages - these are the first 4 entries of r_1 and are called approximation coefficients.
Subtract each average from the first entry of the pair - these are the last 4 entries of r_1 and are called detail coefficients.
Thus, in this case, $r_1 = (550, 578, 1340, 1600, -130, -130, -80, 0)$.
Note that $r_1 = rW_1$, where the columns of W_1 form an orthogonal basis of $\mathbf{R}^8$.

Step II :

Retain the last 4 detail coefficients of r_1 as they are. To get the first 4 entries of r_2, divide the first 4 entries of r_1 into 2 pairs and find their averages. Subtract each average from the first entry of the pair as before; we get the first 4 entries of r_2 in this manner. In our case above,

$$r_2 = (564, 1470, -14, -130, -130, -130, -80, 0).$$

Once again, there is an explicit invertible matrix W_2 such that $r_2 = r_1W_2$; its columns are again orthogonal.

Step III :

Here, retain the last 6 entries of r_2 as they are, for r_3 too. From the first 2 entries of r_2, take average and subtract from the first entry of the pair as before to get the first two entries of r_3. In the above case, we have

$$r_3 = (1017, -453, -14, -130, -130, -130, 80, 0).$$

We have W_3 with columns of orthogonal vectors such that $r_3 = r_2W_3$.

Once these 3 steps are completed, we have $W = W_1W_2W_3$ with $r_3 = rW$. The matrix W is :

$$W = \begin{pmatrix} 1/8 & 1/8 & 1/4 & 0 & 1/2 & 0 & 0 & 0 \\ 1/8 & 1/8 & 1/4 & 0 & -1/2 & 0 & 0 & 0 \\ 1/8 & 1/8 & -1/4 & 0 & 0 & 1/2 & 0 & 0 \\ 1/8 & 1/8 & -1/4 & 0 & 0 & -1/2 & 0 & 0 \\ 1/8 & -1/8 & 0 & 1/4 & 0 & 0 & 1/2 & 0 \\ 1/8 & -1/8 & 0 & 1/4 & 0 & 0 & -1/2 & 0 \\ 1/8 & -1/8 & 0 & -1/4 & 0 & 0 & 0 & 1/2 \\ 1/8 & -1/8 & 0 & -1/4 & 0 & 0 & 0 & -1/2 \end{pmatrix}.$$

For the Haar wavelet transform, one starts with each row of the image matrix A and gets AW.

Then, one performs the same operations on the columns of this new matrix AW to get $S := W^tAW$.
Thus, the compressed image is represented by the matrix W^tAW and one decompresses/retrieves the image as $A = (W^t)^{-1}SW^{-1}$.
The point of all this is that areas of A which contain numbers of nearly equal size end up as zeroes in S - thus S might be a 'sparse' matrix. Usually, a threshold $\epsilon > 0$ is fixed and entries of S which are less than ϵ are reset to zero. Every time we click an image, the source computer recalls S from memory. It sends the overall approximation coefficients and the larger detail coefficients. A little later, it sends the smaller detail coefficients. As our computer receives the information, it starts reconstructing the image progressively in more and more detail.
Normally, one normalizes all the columns of each of W_1, W_2, W_3 to get orthonormal bases of $\mathbf{R}^8$. Therefore,

$$A = (W^t)^{-1}SW^{-1} = WSW^t.$$

Thus, orthonormalization leads to a faster process of compression.

Applications to Image Processing :

In simple terms, suppose we photograph a face from 100 different angles and represent each of the 100 images by a vector v in a fixed large Euclidean space $\mathbf{R}^N$ whose entries are between 0 and 1. The aim is to 'recognise' or 'reconstruct' the face. If $v_1, \cdots, v_{100}$ are the vectors and $w_i = v_i - v_0$ where $v_0 = \frac{1}{100}\sum_{i=1}^{100} v_i$ is the average, one plots the points $w_1, \cdots, w_{100}$ in $\mathbf{R}^N$. This would be a hyper-ellipsoid; finding suitable axes for it, the ellipsoid looks like

$$\frac{x_1^2}{a_1^2} + \cdots + \frac{x_N^2}{a_N^2} = 1.$$

In this notation, each axis is an eigenvector (an 'eigenface'!) and the first 20, say (ordered from smallest eigenvalue onwards), may be enough to construct through linear combinations all the 100 images. The idea is to be able to make a reduction of consideration from a big space like $\mathbf{R}^{20,000}$ to a small dimension like 40, and this will be accomplished through projections. One needs also to formulate and solve the problem of how to plot those points and how to find the suitable axes for which the hyper-ellipsoid looks like the standard one. Let us formulate things mathematically now.

Start with the $100 \times 20,000$ *matrix* A *giving the date as* 100 *vectors in* $\mathbf{R}^{20,000}$.
Diagonalize ${}^tAA$, *say* $UD_1\ {}^tU = {}^t\ AA$, *where the columns of* U *are eigenvectors of* ${}^tAA$ *and* D_1 *is diagonal.*
Diagonalize $A\ {}^tA = VD_2\ {}^tV$, *where the columns of* V *are eigenvectors of* $A\ {}^tA$, *and* D_2 *is diagonal.*
Find D *with* $D\ {}^tD = D_2, {}^t\ DD = D_1$ *and* $A = VD\ {}^tU$.
The eigenfaces are the columns of U.

Applications to Economics :
The applications of linear algebra to economics and finance is manifold. For instance, an economic model invented by Leontief (a Nobel prize winner in 1973) amounts to a system of linear equations of the form $AX = X$ where one looks for a non-zero solution X with all entries non-negative. This is a closed economy system where no goods leave or enter the system. There is also a related model which leads to an inhomogeneous system $AX = X + d$. Here, the economy is open, that is, there is a certain outside demand (occurring as the column d above) which has to be met.

References :

[A] : Sheldon Axler, *Linear algebra done right*, Springer 2005.
[F] : Paul A.Fuhrmann, *A polynomial approach to linear algebra*, Universitext, Springer 1996.
[K] : S.Kumaresan, *Linear algebra : a geometric approach*, Eastern Economy Edition, Prentice-Hall 2000.
[M] : Carl Meyer, *Matrix analysis and applied linear algebra*, Available online.