From now on, unless otherwise stated, the scalar field for each vector space we deal with will be $\mathbb{C}$, the complex field. And all the vector spaces will be finite dimensional.

Maschke's Theorem. *Let G be a finite group, V a G-module and U a G-submodule of V. Then there is a submodule W of V such that $V = U \oplus W$.*

In accordance with the definition of "irreducible" given in Lecture 6, a module is said to be *reducible* if it has a nonzero proper submodule. A module is said to be *completely reducible* if for every submodule there is a complementary submodule. Maschke's Theorem says that every module for a finite group $|G|$ (over a field such that $|G| \neq 0$) is completely reducible, and so the theorem is also known as the Theorem of Complete Reducibility.

If you have worked completely through the exercises in Tutorial 1 then you have already proved Maschke's Theorem. That proof goes as follows.

The first step is to note that it is possible to define an inner product on the space V. That is, there exists a function $(u, v) \mapsto u \cdot v$ from $V \times V$ to $\mathbb{C}$ such that

(i) $u \cdot (\lambda v + \mu w) = \lambda(u \cdot v) + \mu(u \cdot w)$ for all u, v, $w \in V$ and λ, $\mu \in \mathbb{C}$,

(ii) $u \cdot v = \overline{v \cdot u}$ for all u, $v \in V$ (where the overline indicates complex conjugation),

(iii) $u \cdot u$ is real and positive for all nonzero $u \in V$ (and is 0 if $v = 0$).

Indeed, if v_1, v_2, $\ldots$, v_n is any basis of V then there exists an inner product such that $v_i \cdot v_j = \delta_{ij}$ for all i and j. Explicitly, if $u = \sum_i \lambda_i v_i$ and $v = \sum_i \mu_i v_i$ then $u \cdot v = \sum_i \overline{\lambda_i} \mu_i$.

Fix an inner product $(u, v) \mapsto u \cdot v$ on V, and define another function $V \times V \to \mathbb{C}$ by the formula $u * v = \sum_{x \in G} xu \cdot xv$. It is easy to show that properties (i), (ii) and (iii) above are satisfied, so that $*$ is also an inner product. Moreover, it is G-invariant, in the sense that $gu * gv = u * v$ for all u, $v \in V$ and all $g \in G$, since

$$gu * gv = \sum_{x \in G} x(gu) \cdot x(gv) = \sum_{x \in G} (xg)u \cdot (xg)v = \sum_{y \in G} yu \cdot yv = u * v$$

(where we have used the fact that as x runs through all the elements of G then so too does $y = xg$, as $x \mapsto xg$ is a bijection $G \to G$). We now define W to be the orthogonal complement of U relative to this new inner product: $W = U^{\perp} = \{\, v \in V \mid u * v = 0 \text{ for all } u \in U \,\}$. Then $V = U \oplus W$. (It is a general property of inner product spaces that if U is any subspace then the whole space is the direct sum $U \oplus U^{\perp}$.) We only have to show that W is a G-submodule of V, and since we already know that it is a subspace we only have to show that if $w \in W$ and $g \in G$ then $gw \in W$. But this is easy: if $u \in U$ then $u * gw = g^{-1}u * w = 0$ since $w \in U^{\perp}$ and $g^{-1}u \in U$ (since U is a G-module); hence $gw \in U^{\perp}$ (since $u * gw = 0$ for all $u \in U$).

The key idea in this proof is to create a G-invariant object—in this case an inner product—by summing the G-transforms of an arbitrary object. This is a theme that will recur at several points in this course. Often as well as summing over G we divide by $|G|$, so that the process can be regarded as an averaging of the effects of the elements of G. We will now give a second proof of Maschke's Theorem; this proof can be applied unchanged if the complex field is replaced by any field in which $|G| \neq 0$. The key averaging idea remains.

Given the G-module V and submodule U, choose an arbitrary vector subspace Z of V that is complementary to U. Thus as a vector space $V = U \oplus Z$, but this will not generally be a G-module decomposition. Now for each $z \in Z$ and $g \in G$ the element $gz \in V$ can be split uniquely into a component in U and a component in Z; so we can write

$$gz = \tau_g z + \sigma_g z \tag{1}$$

where $\tau_g \colon Z \to U$ and $\sigma_g \colon Z \to Z$ are linear maps. Note that taking $g = 1$ gives $gz = z$; hence $\tau_1 z = 0$ and $\sigma_1 z = z$.

Let $h \in G$ and apply h to both sides of Eq. (1), and compare the result with the equation obtained by replacing g by hg in Eq. (1):

$$\begin{aligned}\tau_{hg} z + \sigma_{hg} z = (hg)z &= h(gz) \\ &= h(\tau_g z + \sigma_g z) \\ &= h(\tau_g z) + h(\sigma_g z) \\ &= h(\tau_g z) + (\tau_h(\sigma_g z) + \sigma_h(\sigma_g z))\end{aligned} \tag{2}$$

where in the last step we have applied Eq. (1) with g replace by h and z by $\sigma_g z$. Since $\tau_g z \in U$ and U is a G-submodule it follows that $h(\tau_g z) \in U$, and so comparing the U and Z components of the first and last expressions in Eqq. (2) gives

$$\tau_{hg} z = h(\tau_g z) + \tau_h(\sigma_g z) \tag{3}$$

as well as $\sigma_{hg} z = \sigma_h(\sigma_g z)$. Note that taking $h = g^{-1}$ here yields that $\sigma_{(g^{-1})} = \sigma_g^{-1}$, since $\sigma_1 = \mathrm{id}$, and if we now replace z by $\sigma_g^{-1} z = \sigma_{hg}^{-1}(\sigma_h z)$ in Eq. (3) we obtain

$$\tau_{hg}(\sigma_{hg}^{-1}(\sigma_h z)) = h(\tau_g(\sigma_g^{-1} z)) + \tau_h z. \tag{4}$$

It is Eq. (4) to which we apply the averaging idea: summing it over all $g \in G$ gives

$$\sum_{g \in G} \tau_{hg}(\sigma_{hg}^{-1}(\sigma_h z)) = h\Big(\sum_{g \in G} \tau_g(\sigma_g^{-1} z)\Big) + |G| \tau_h z,$$

and now dividing by $|G|$ gives

$$\eta(\sigma_h z) = h(\eta z) + \tau_h z, \tag{5}$$

where we have defined $\eta \colon Z \to U$ by

$$\eta z = \frac{1}{|G|} \sum_{g \in G} \tau_g(\sigma_g^{-1} z).$$

(The crucial point is that if $h \in G$ is fixed then $\eta z' = \frac{1}{|G|} \sum_{g \in G} \tau_{hg}(\sigma_{hg}^{-1} z')$ for all $z' \in Z$, since hg runs through all elements of G as g does.)

Let $W = \{\, z + \eta z \mid z \in Z \,\}$, and let $w \in W$ be arbitrary. Choose $z \in Z$ such that $w = z + \eta z$. Then for all $h \in G$,

$$hw = hz + h(\eta z) = hz + (\eta(\sigma_h z) - \tau_h z)$$

(by Eq. (5)), and now using Eq. (1) we deduce that

$$hw = \sigma_h z + \eta(\sigma_h z) \in W.$$

Since this holds for all $w \in W$ and $h \in G$ we have shown that W is closed under the G-action. It is also a vector subspace of V since it is the image of the linear map $z \mapsto z + \eta z$ from Z to V. Thus W is a G submodule of V. If $v \in V$ then, for some $u \in U$ and $z \in Z$,

$$v = u + z = (u - \eta z) + (z + \eta z) \in U + W,$$

since $\eta z \in U$ and $z + \eta z \in W$. Furthermore, $U \cap W = \{0\}$, since if $u \in U$ and $u = z + \eta z$ for some $z \in Z$ then

$$z = u - \eta z \in U \cap Z = \{0\},$$

showing that $z = 0$ and hence $u = 0$. Thus the submodule W is complementary to U, as required.

As always in this subject, it is possible to rephrase proofs using matrices rather than linear transformations. Sometimes results become easier to understand this way, and sometimes harder. And often different people disagree about which is easier. So let us anyway go through the above proof in matrix terms. We start by choosing a basis $v_1, v_2, \ldots, v_n$ for the submodule U of V, and then extending this to a basis $v_1, v_2, \ldots, v_{n+m}$ of V. For $g \in G$ let Qg be the $(n+m) \times (n+m)$ matrix whose (i,j)-entry $Q_{ij}g$ is defined by

$$gv_j = \sum_{i=1}^{n+m} (Q_{ij}g)v_i.$$

That is, Qg is the matrix of the transformation $v \mapsto gv$ relative to our chosen basis. Observe that if $1 \le j \le n$ then $v_j \in U$ and so $gv_j \in U$, and it follows that gv_j is a linear combination of $v_1, v_2, \ldots, v_n$. Thus the coefficients $Q_{ij}g$ are zero for $n+1 \le i \le n+m$ and $1 \le j \le n$. So we have a block decomposition of the matrix Qg as

$$Qg = \begin{pmatrix} Rg & Tg \\ 0 & Sg \end{pmatrix} \qquad \text{for all } g \in G \tag{8}$$

where Rg and Sg are respectively $n \times n$ and $m \times m$ matrices, and Tg is $n \times m$. This can be viewed as the matrix version of reducibility; more precisely, a matrix representation of G is reducible if it is equivalent to a matrix representation having a block structure as in Eq. (8). If the subspace of V spanned by $v_{n+1}, v_{n+2}, \ldots, v_m$ were a G-submodule then gv_{n+j} would be a linear combination of $v_{n+1}, v_{n+2}, \ldots, v_m$, and the coefficients $Q_{ij}g$ would be zero for $i \le n$ and $j \ge n$; the matrix Tg would be 0 for all g. Thus a matrix representation is decomposable if it is equivalent to one of the form $g \mapsto \begin{pmatrix} Rg & 0 \\ 0 & Sg \end{pmatrix}$, and the matrix form of Maschke's Theorem is that a representation of the form given by Eq. (8) is equivalent to a representation of the same form with all the Tg's zero.

Since Rg has (i,j)-entry $Q_{ij}g$ for $i, j \in \{1, 2, \ldots, n\}$ we see that Rg is the matrix relative to $v_1, v_2, \ldots, v_n$ of the transformation $u \mapsto gu$ of U. Note also that $v_{n+1}+U, v_{n+2}+U, \ldots, v_{n+m}+U$ is a basis for the quotient module V/U, and since

$$g(v_{n+j} + U) = \Big(\sum_{i=1}^{n+m} (Q_{i,n+j}g)v_i\Big) + U = \sum_{i=1}^{m} (Q_{n+i,n+j}g)(v_{n+i} + U),$$

and $Q_{n+i,n+j}g$ is the (i,j)-entry of Sg, we see that Sg is the matrix of $v + U \mapsto g(v + U)$ relative to the above basis of V/U.

Since $g \mapsto Qg$ is a matrix representation of G, Eq. (8) gives

$$\begin{pmatrix} R(hg) & T(hg) \\ 0 & S(hg) \end{pmatrix} = Q(hg) = (Qg)(Qh) = \begin{pmatrix} Rh & Th \\ 0 & Sh \end{pmatrix} \begin{pmatrix} Rg & Tg \\ 0 & Sg \end{pmatrix}$$
$$= \begin{pmatrix} (Rh)(Rg) & (Rh)(Tg) + (Th)(Sg) \\ 0 & (Sh)(Sg) \end{pmatrix},$$

which confirms the formulas $R(hg) = (Rh)(Rg)$ and $S(hg) = (Sh)(Sg)$ (which we already knew since R and S are matrix versions of the representations of G on U and V/U), and also enables us to deduce that $T(hg) = (Rh)(Tg) + (Th)(Sg)$. Hence, on right multiplying by $(Sg)^{-1} = S(hg)^{-1}(Sh)$,

$$(T(hg)S(hg))^{-1}(Sh) = (Rh)((Tg)(Sg)^{-1}) + Th, \tag{9}$$

whch is the matrix analogue of Eq. (4). Averaging over $g \in G$ this gives

$$E(Sh) = (Rh)E + Th \qquad \text{for all } h \in G,$$

where we have defined $E = \frac{1}{|G|}\sum_{g\in G}(Tg)(Sg)^{-1}$. Hence we derive the following matrix equation:

$$\begin{pmatrix} Rh & Th \\ 0 & Sh \end{pmatrix} \begin{pmatrix} I & E \\ 0 & I \end{pmatrix} = \begin{pmatrix} I & E \\ 0 & I \end{pmatrix} \begin{pmatrix} Rh & 0 \\ 0 & Sh \end{pmatrix} \qquad \text{for all } h \in G.$$

Equivalently

$$\begin{pmatrix} I & E \\ 0 & I \end{pmatrix}^{-1} \begin{pmatrix} Rh & Th \\ 0 & Sh \end{pmatrix} \begin{pmatrix} I & E \\ 0 & I \end{pmatrix} = \begin{pmatrix} Rh & 0 \\ 0 & Sh \end{pmatrix} \qquad \text{for all } h \in G,$$

so that the representation Q is equivalent to the diagonal sum of the representations R and S, as required.

Lecture 8, 20/8/97

Having done Maschke's Theorem, let us proceed at once to the other main theorem of this course: Schur's Lemma. Calling it a lemma, as is traditional, belies its importance. But as befits a lemma, its proof is easy.

Schur's Lemma (Version 1). *Let U and V be irreducible G-modules and $\phi: U \to V$ a G-homomorphism. Then ϕ is either a G-isomorphism or the zero map.*

Proof. By part of the First Isomorphism Theorem, $\ker\phi$ is a G-submodule of U. But U is irreducible, and so has no nontrivial proper submodules. Thus either $\ker\phi = U$ or $\ker\phi = \{0\}$. If $\ker\phi = U$ then ϕ is the zero map, which is one of the possibilities allowed in the statement of the theorem. In the alternative case ϕ is injective, since $\ker\phi = \{0\}$. Now another part of The First Isomorphism Theorem tells us that $\operatorname{im}\phi$ is a submodule of V, and so irreducibility of V tells us that $\operatorname{im}\phi$ is either $\{0\}$ or V. If it is $\{0\}$ then again ϕ must be the zero map, and since we have already excluded this case we deduce that $\operatorname{im}\phi = V$. So ϕ is surjective as well as injective, and thus it is an isomorphism, as required.* □

* It has just occurred to me that the definition of "irreducible" should have incorporated the assumption that an irreducible module has to be nonzero. We shall assume this henceforth.

So it would seem that Schur's Lemma is no big deal. Yet we will spend quite some time deriving consequences and reformulations of the result, many of which are quite striking. The truth is that the assumption of irreducibility is very strong. Thus irreducible modules are rather special objects, and, as we shall see, they have some striking properties.

First, we should derive the matrix form of the above statement:

Schur's Lemma (Version 2). *Let $R\colon G \to \mathrm{GL}(n,F)$ and $S\colon G \to \mathrm{GL}(m,F)$ be irreducible matrix representations of G, and let X be an $n \times m$ matrix which intertwines R and S. Then either $X = 0$ or X is invertible.*

This follows immediately from Version 1, in view of our discussion of G-homomorphisms and intertwining matrices in Lecture 5. Note that, of course, the case X invertible can only arise if $n = m$ (which in module terms says that if U and V are isomorphic they have the same dimension over F).

It should be noted that everything that has been said so far is totally general. Our assumption that $F = \mathbb{C}$ can be dispensed with, the group G does not have to be finite and the modules U and V do not have to be finite-dimensional over F. We have used only the First Isomorphism Theorem (and indeed only special cases of that) and the definition of irreducibility. For the next result, though, we do make use of the assumption that the field is $\mathbb{C}$.

Schur's Lemma (Version 3). *Let $R\colon G \to \mathrm{GL}(d,\mathbb{C})$ be an irreducible matrix representation of G, and suppose X is a $d \times d$ matrix such that $(Rg)X = X(Rg)$ for all $g \in G$. Then $X = \lambda I$ for some $\lambda \in \mathbb{C}$.*

Proof. Choose λ to be an eigenvalue of X. Because the ground field is $\mathbb{C}$, and every nonconstant polynomial with coefficients in $\mathbb{C}$ has a root in $\mathbb{C}$, the characteristic polynomial of X does have at least one root $\lambda \in \mathbb{C}$, and so a suitable λ certainly exists. By definition, $\det(X - \lambda I) = 0$, and the matrix $X - \lambda I$ is not invertible. Now for all $g \in G$,

$$(X - \lambda I)(Rg) = X(Rg) - \lambda(Rg) = (Rg)X - \lambda(Rg) = (Rg)(X - \lambda I),$$

since X commutes with every Rg. Thus $X - \lambda I$ commutes with every Rg, and by Version 2 above it follows that $X - \lambda I$ is invertible or zero. By the choice of λ it is not invertible; so $X = \lambda I$, as required. □

The module version of this statement is that if V is a finite-dimensional irreducible G-module over $\mathbb{C}$ and $\phi\colon V \to V$ is a G-homomorphism then ϕ is a scalar multiple of the identity map. The assumption here that V is finite-dimensional is necessary since infinite-dimensional vector spaces do admit linear operators which have no eigenvalues.

It will not have escaped the attention of the alert reader that, in our third form of Schur's Lemma, $\mathbb{C}$ could be replaced by any algebraically closed field.

Orthogonality relations

A little scepticism is a healthy thing, and it would be natural at this stage to be a little sceptical about the usefulness of Schur's Lemma. After all, it can only be applied if one has a G-homomorphism or intertwining matrix for a pair of irreducible representations. And homomorphisms are very special things: they may not be easy to find. Fortunately, the averaging argument used in the proof of Maschke's Theorem affords (for finite groups) a general method of constructing them.

Lemma. *Let G be a finite group, and let R, S be matrix representations of G of degrees n and m respectively. If X is any $n \times m$ matrix then $Y = \sum_{g\in G}(Rg)X(S(g^{-1}))$ intertwines R and S.*

Proof. For all $h \in G$,

$$\begin{aligned}(Rh)Y = \sum_{g\in G}(Rh)(Rg)X(S(g^{-1})) &= \sum_{g\in G}(R(hg))X(S(g^{-1}))(S(h^{-1}))(Sh)\\ &= \Bigl(\sum_{g\in G}(R(hg))X(S(g^{-1}h^{-1})\Bigr)(Sh) = \Bigl(\sum_{k\in G}(Rk)X(S(k^{-1})\Bigr)(Sh) = Y(Sh)\end{aligned}$$

since $k = hg$ runs through all elements of G as g does. □

Suppose now that $R^{(1)}$, $R^{(2)}$, …, $R^{(s)}$ are irreducible matrix representations of the finite group G, of degrees $d_1, d_2, \ldots, d_s$ respectively. Assume furthermore that $R^{(k)}$ and $R^{(l)}$ are not equivalent if $k \neq l$. Write $R^{(k)}_{ij}g$ for the (i, j)-entry of $R^{(k)}g$.

Choose k and l arbitrarily from the set $\{1, 2, \ldots, s\}$, and for $1 \leq m \leq d_k$ and $1 \leq n \leq d_l$ let $X^{(k,l)}_{m,n}$ be the $d_k \times d_l$ matrix whose (t, u)-entry is 0 unless $t = m$ and $u = n$, in which case it is 1. In other words, the (t, u)-entry of $X^{(k,l)}_{m,n}$ is $\delta_{tm}\delta_{un}$. The lemma above tells us that the matrix

$$Y^{(k,l)}_{m,n} = \frac{1}{|G|}\sum_{g\in G}(R^{(k)}g)X^{(k,l)}_{m,n}(R^{(l)}(g^{-1}))$$

intertwines $R^{(k)}$ and $R^{(l)}$. Hence, by Schur's Lemma, $Y^{(k,l)}_{m,n}$ is zero if $k \neq l$ (since the two representations are inequivalent in this case), while $Y^{(k,k)}_{m,n}$ must be a scalar multiple of I. Thus the (p, q)-entry of $Y^{(k,l)}_{m,n}$ is $\lambda(k, m, n)\delta_{pq}\delta_{kl}$ for some $\lambda(k, m, n) \in \mathbb{C}$.

Computing the (p, q)-entry of $Y^{(k,l)}_{m,n}$ directly from the definition we find that

$$\begin{aligned}\lambda(k, m, n)\delta_{pq}\delta_{kl} &= \frac{1}{|G|}\sum_{g\in G}\Bigl(\sum_{t=1}^{d_k}\sum_{u=1}^{d_l}(R^{(k)}_{pt}g)\delta_{tm}\delta_{un}(R^{(l)}_{uq}(g^{-1}))\Bigr)\\ &= \frac{1}{|G|}\sum_{g\in G}(R^{(k)}_{pm}g)(R^{(l)}_{nq}(g^{-1})).\end{aligned}$$

Considering the (n, m)-entry of $Y^{(l,k)}_{q,p}$ yields by the same calculation (or by renaming the variables above) that

$$\begin{aligned}\lambda(l, q, p)\delta_{nm}\delta_{kl} &= \frac{1}{|G|}\sum_{g\in G}(R^{(l)}_{nq}g)(R^{(k)}_{pm}(g^{-1}))\\ &= \frac{1}{|G|}\sum_{g\in G}(R^{(l)}_{nq}(g^{-1}))(R^{(k)}_{pm}g)\end{aligned}$$

where in the last step we have simply changed the dummy variable of summation from g to g^{-1}. But the right hand sides of the last two displayed formulas are equal, and so we conclude that for all values of k, l, m, n, p and q,

$$\lambda(k, m, n)\delta_{pq}\delta_{kl} = \lambda(l, q, p)\delta_{nm}\delta_{kl}.$$

Putting $q = p$ and $l = k$ shows that $\lambda(k, m, n) = \lambda(k, p, p)\delta_{nm}$, and now putting $m = n$ shows that $\lambda(k, n, n) = \lambda(k, p, p) = \mu_k$ depends only k and not on p or n. So $\lambda(k, m, n) = \mu_k\delta_{nm}$.