Mathematics Department Stanford University
Math 51H – Duals, adjoints/transposes and the spectral theorem

Suppose V is an n-dimensional vector space over a field F such as $\mathbb{R}$. When discussing determinants, we have shown that the vector space of m-linear maps, i.e. maps from $\times_{i=1}^k V$ (product of k copies of V) to F, linear in each slot, is n^k dimensional. Indeed, if $e_1, \ldots, e_n$ is a basis of V, such a map α is determined by its values $\alpha(e_{i_1}, \ldots, e_{i_k})$, $i_1, \ldots, i_k \in \{1, \ldots, n\}$, which values however can be specified freely. Concretely, in the case of $k = 1$, i.e. linear maps $\alpha : V \to \mathbb{R}$,

$$\alpha(\sum_{i=1}^n c_i e_i) = \sum_{i=1}^n c_i \alpha(e_i).$$

So consider the linear maps f_j specified by $f_j(e_i) = 0$ if $i \neq j$, $f_j(e_i) = 1$ if $i = j$ (commonly written as $f_j(e_i) = \delta_{ij}$ where δ_{ij} is the 'Kronecker delta', i.e. is $= 1$ if $i = j$, 0 otherwise), i.e.

$$f_j(\sum_i c_i e_i) = \sum_i c_i f_j(e_i) = c_j.$$

These form a basis for the set V^* of linear maps $V \to \mathbb{R}$ (called the **dual** of V) since any linear map α can be written as above

$$\alpha(\sum_{i=1}^n c_i e_i) = \sum_{i=1}^n c_i \alpha(e_i) = \sum_{i=1}^n (\sum_{j=1}^n f_i(e_j) c_j) \alpha(e_i) = \sum_{i=1}^n \alpha(e_i) f_i(\sum_{j=1}^n c_j e_j),$$

i.e.

$$\alpha = \sum_{i=1}^n \alpha(e_i) f_i.$$

A different way of looking at V^* is available when V is a vector space over $\mathbb{R}$ with an inner product. Namely consider the map $\iota : V \to V^*$ given by the following: for $x \in V$, $\iota(x) \in V^*$ is the map $V \to \mathbb{R}$ given by $(\iota(x))(v) = x \cdot v$. This map $\iota : V \to V^*$ is linear:

$$\iota(\alpha x + \beta y)(v) = (\alpha x + \beta y) \cdot v = \alpha(x \cdot v) + \beta(y \cdot v) = \alpha\iota(x)(v) + \beta\iota(y)(v).$$

(Note that this is a completely different statement from $\iota(x) : V \to \mathbb{R}$ being linear for each $x \in V$, since this is the statement that $\iota(x)(\lambda v + \mu w) = x \cdot (\lambda v + \mu w) = \lambda x \cdot v + \mu x \cdot w = \lambda\iota(x)(v) + \mu\iota(x)(w)$ — the two proofs are very similar, but the statements are quite different!) Moreover, ι is injective: if $\iota(x) = 0$ (i.e. is the zero map) then $\iota(x)(x) = 0$, but $\iota(x)x = x \cdot x = \|x\|^2$, so $x = 0$. Since V is n-dimensional, the rank-nullity theorem states that the range of ι is n-dimensional. But the latter is a subspace of V^* which is n-dimensional, so it is equal to all of V^*, so ι is onto. This shows that $\iota : V \to V^*$ is a bijection, i.e. an invertible map.

Now consider **adjoints/transposes**. Recall that if $A : V \to V$ is linear, V an inner product space, then there is a unique map $A^T : V \to V$ such that $Ax \cdot y = x \cdot A^T y$ for all $x, y \in V$; A^T is the adjoint or transpose of A. Indeed, recall that we have shown existence plus uniqueness within the collection of linear maps by computing A^T in an orthonormal basis. Here is a different way of showing this now, without assuming linearity for the uniqueness statement.

Fix $y \in V$, and consider the map $j_y : x \mapsto Ax \cdot y$. This is a linear map as $j_y(\alpha x + \beta z) = \alpha j_y(x) + \beta j_y(z)$, so it is an element of V^*. Thus, as $\iota : V \to V^*$ is bijective there exists a unique $w \in V$ such that $j_y = \iota(w)$, i.e. such that $j_y(x) = \iota(w)(x)$ for all $x \in V$. But then $Ax \cdot y = j_y(x) = x \cdot w$. So define $A^T : V \to V$ by $A^T y = w$, w being the unique element of V such that $j_y = \iota(w)$. Then A^T is well-defined, and $Ax \cdot y = x \cdot w = x \cdot Ay$. One also checks easily that A^T is linear: by the definition of A^T, $A^T(c_1 y_1 + c_2 y_2)$ is the unique element of V such that $Ax \cdot (c_1 y_1 + c_2 y_2) = x \cdot A^T(c_1 y_1 + c_2 y_2)$ for all $x \in V$; thus for all $x \in V$,

$$\begin{aligned} x \cdot A^T(c_1 y_1 + c_2 y_2) &= Ax \cdot (c_1 y_1 + c_2 y_2) = c_1 Ax \cdot y_1 + c_2 Ax \cdot y_2 \\ &= c_1 x \cdot A^T y_1 + c_2 x \cdot A^T y_2 = x \cdot (c_1 A^T y_1 + c_2 A^T y_2), \end{aligned}$$

where the second equality is from the linearity of the inner product in the second slot, the third from the definition of A^T, and the fourth again from the linearity of the inner product in the second slot. This is exactly the statement that

$$\iota(A^T(c_1y_1 + c_2y_2)) = \iota(c_1A^Ty_1 + c_2A^Ty_2),$$

which by the injectivity of ι means $A^T(c_1y_1 + c_2y_2) = c_1A^Ty_1 + c_2A^Ty_2$. Thus, $A^T : V \to V$ is linear, completing the construction of A^T without reference to an orthonormal basis.

To connect this with the notion of transposes of matrices, note that if $e_1, \ldots, e_n$ is an orthonormal basis of V then one can compute the matrix of any operator very easily. Recall that the matrix of $A : V \to V$ is given by $\{a_{ij}\}_{i,j=1}^n$ where $Ae_j = \sum_{i=1}^n a_{ij}e_i$. Since the e_i are orthonormal, we have

$$e_k \cdot Ae_j = \sum_{i=1}^{n} a_{ij}e_k \cdot e_i = a_{kj},$$

i.e. relabelling the indices,

$$a_{ij} = e_i \cdot Ae_j.$$

Correspondingly, the ij matrix entry of A^T is

$$e_i \cdot A^Te_j = Ae_i \cdot e_j = e_j \cdot Ae_i = a_{ji}.$$

Thus, *in any orthonormal basis*, the matrix of the transpose/adjoint of A is the transpose of the matrix of A. This, however, needs an orthonormal basis.

A very important class of linear maps are the **symmetric, or self-adjoint** ones. A linear map $A : V \to V$ is symmetric if $A = A^T$, i.e. if $Ax \cdot y = x \cdot Ay$ for all $x, y \in V$. Notice that by what we showed above, the matrix of a symmetric map *in any orthonormal basis* is symmetric.

Thus, orthonormal bases are the natural bases for vector spaces with inner products, i.e. inner product spaces. But how do they arise? One way is from any basis via Gram-Schmidt. But there are also more natural ways given a symmetric linear map $A : V \to V$. We first make a definition:

Definition 1 *If V a vector space over F, $I : V \to V$ the identity operator (so $Iv = v$ for all $v \in V$), $A : V \to V$ is linear, then $\lambda \in F$ is an* **eigenvalue** *of A if $N(A - \lambda I) \neq \{0\}$, i.e. if there exists $v \neq 0$ such that $(A - \lambda I)v = 0$, i.e. if there exists $v \neq 0$ such that $Av = \lambda v$.*

For an eigenvalue λ of A, we say that v is an **eigenvector** *of A if $Av = \lambda v$ (and $v \neq 0$, to follow the book).*

For an eigenvalue λ of A, **the λ-eigenspace** *is $N(A - \lambda I)$, i.e. is the subspace of V consisting of vectors v such that $Av = \lambda v$.*

Here 'eigen' is 'own' in German; it corresponds to these vectors being very well behaved under the action of A: they simply get stretched.

An important lemma relates orthogonality and eigenvectors for symmetric operators.

Lemma 1 *If $A : V \to V$ is symmetric, $Ax = \lambda x$, $Ay = \mu y$, $\lambda \neq \mu$ then $x \cdot y = 0$, i.e. x and y are orthogonal.*

Proof: $\lambda x \cdot y = Ax \cdot y = x \cdot Ay = \mu x \cdot y$, so $(\lambda - \mu)(x \cdot y) = 0$. Since $\lambda \neq \mu$, $x \cdot y = 0$. $\square$

Thus, for a symmetric operator eigenvectors with different eigenvalues are automatically orthogonal to each other.

This gives a way of constructing an orthogonal set of vectors in V: suppose $\lambda_1, \ldots, \lambda_m$ are the eigenvalues of A. (Note that there can be at most n eigenvalues: for each there is a non-zero vector in the eigenspace, which are automatically orthogonal to each other, thus they are linearly independent, and there can be at most n linearly independent vectors in V.) Let $v_{j1}, \ldots, v_{jk_j}$ be an orthonormal basis of $N(A - \lambda_j I)$. Then putting these together, we get an orthonormal collection of vectors

$$v_{11}, \ldots, v_{1k_1}, v_{21}, \ldots, v_{2k_2}, \ldots, v_{m1}, \ldots, v_{mk_m};$$

these are orthonormal because eigenvectors in different eigenspaces are automatically orthogonal to each other. In particular, this is a linearly independent collection of vectors. On the other hand, a priori it is not clear whether they span V. However, this is guaranteed by the following theorem.

Theorem 1 (Spectral theorem.) *Suppose $A : V \to V$ is symmetric. Then V has an orthonormal basis consisting of eigenvectors of A.*

The key tool in proving this theorem is the following observation.

Definition 2 *An* **invariant subspace** *of $A : V \to V$ is a subspace W of V such that $A : W \to W$, i.e. such that for all $w \in W$, $Aw \in W$.*

Lemma 2 *Suppose that $A : V \to V$ is symmetric, and W is an invariant subspace of A. Then $W^\perp$ is also an invariant subspace of A.*

Proof: Suppose $x \in W^\perp$. Then for $w \in W$, $Ax \cdot w = x \cdot Aw = 0$ since $x \in W^\perp$ and $Aw \in W$. This shows $Ax \in W^\perp$, and thus the lemma. $\square$

Notice that any eigenspace of A is an invariant subspace of A: if, say, $x \in N(A - \lambda I)$, then $Ax = \lambda x \in N(A - \lambda I)$ as well just by virtue of $N(A - \lambda I)$ being a subspace of V. Moreover, if v is any eigenvector of A, then a similar argument gives that $\mathrm{Span}\{v\}$ is an invariant subspace of A. Thus, in both cases, the orthocomplement is also invariant for A.

The proof of the theorem now reduces to showing that any symmetric A on a space of dimension ≥ 1 has a single eigenvalue; if we find an eigenvalue λ, then taking v to be an eigenvector of A with eigenvalue λ, $\mathrm{Span}\{v\}$ is invariant for A, thus so is $\mathrm{Span}\{v\}^\perp$, and the latter has dimension $n - 1$, so by an inductive argument we may assume that $\mathrm{Span}\{v\}^\perp$ has an orthonormal basis consisting of eigenvectors for A; adding v to this set provides a desired basis of V consisting of eigenvectors of A.

In fact, one could avoid induction altogether: if $\lambda_1, \ldots, \lambda_m$ are the eigenvalues of V as above, let W be the sum of the corresponding eigenspaces:

$$W = \oplus_{j=1}^m N(A - \lambda_j I).$$

(Notice that the sum is direct since the summands are linearly independent, indeed orthogonal!) This is an invariant subspace of V since if $v_j \in N(A - \lambda_j I)$ then $A \sum_{j=1}^m v_j = \sum_{j=1}^m Av_j \in \oplus_{j=1}^m N(A - \lambda_j I)$. Thus, $W^\perp$ is also an invariant subspace of A, and A is a symmetric operator on it. If $W^\perp \neq \{0\}$, and if we have shown that any symmetric operator on any inner product space of dimension ≥ 1 has an eigenvalue, then $A|_{W^\perp}$ has an eigenvalue, λ, and a corresponding eigenvector $v \neq 0$, so $Av = \lambda v$, so λ is an eigenvalue of A and v is in the corresponding eigenspace, so $v \in W$, which is a contradiction with $v \in W^\perp$.

So the main point is to show the following lemma; once this is shown, the spectral theorem follows.

Lemma 3 *If $\dim V \geq 1$, $A : V \to V$ symmetric then A has an eigenvalue on V.*

In order to motivate the proof, consider the Rayleigh quotient:

$$\mathcal{A}(x) = \frac{Ax \cdot x}{\|x\|^2},\ x \in V,\ x \neq 0.$$

Notice that this is unchanged if one replaces x by a non-zero multiple:

$$\mathcal{A}(tx) = \frac{A(tx) \cdot (tx)}{\|tx\|^2} = \frac{Ax \cdot x}{\|x\|^2} = \mathcal{A}(x),\ t \neq 0.$$

Let us now see how A behaves on a sum of eigenvectors $x = \sum_{j=1}^m v_j$, $v_j = N(A - \lambda_j I)$, λ_j distinct eigenvalues, so the v_j are automatically orthogonal to each other. Then (if not all v_j vanish)

$$\mathcal{A}(x) = \frac{\sum_{j=1}^m Av_j \cdot \sum_{i=1}^m v_i}{\sum_{j=1}^m v_j \cdot \sum_{i=1}^m v_i} = \frac{\sum_{j=1}^m \lambda_j v_j \cdot \sum_{i=1}^m v_i}{\sum_{j=1}^m \sum_{i=1}^m v_j \cdot v_i} = \frac{\sum_{j=1}^m \lambda_j \sum_{i=1}^m v_j \cdot v_i}{\sum_{j=1}^m v_j \cdot v_j} = \frac{\sum_{j=1}^m \lambda_j \|v_j\|^2}{\sum_{j=1}^m \|v_j\|^2}$$

In particular, if the λ_j are ordered: $\lambda_1 < \lambda_2 < \ldots < \lambda_m$ then

$$\lambda_1 = \frac{\sum_{j=1}^m \lambda_1 \|v_j\|^2}{\sum_{j=1}^m \|v_j\|^2} \leq \frac{\sum_{j=1}^m \lambda_j \|v_j\|^2}{\sum_{j=1}^m \|v_j\|^2} \leq \frac{\sum_{j=1}^m \lambda_m \|v_j\|^2}{\sum_{j=1}^m \|v_j\|^2} = \lambda_m.$$

Correspondingly, a reasonable idea is to look at the minimum of $\mathcal{A}$ to find λ_1; similarly, one could look for the maximum of $\mathcal{A}$ to find λ_m.

But this is easy: since $\mathcal{A}(x) = \mathcal{A}(\frac{x}{\|x\|})$, with $\frac{x}{\|x\|}$ a unit vector, simply consider the constrained minimization problem: find the minimum of $\mathcal{A}$ on the unit sphere $S = \{x :\ \|x\| = 1\}$. This is equivalent to finding the minimum of $f(x) = Ax \cdot x$ on S. But

$$f(x+h) = A(x+h) \cdot (x+h) = Ax \cdot x + Ah \cdot x + Ax \cdot h + Ah \cdot h = f(x) + 2Ax \cdot h + Ah \cdot h,$$

with $|Ah \cdot h| \leq C\|h\|^2$ for some C (for instance, $C = \|A\|$). Thus, $Df(x)h = 2Ax \cdot h$, i.e. $\nabla f(x) = 2Ax$. Similarly, with $g(x) = \|x\|^2$, $Dg(x)h = 2x \cdot h$, i.e. $\nabla g(x) = 2x$. Correspondingly, since critical points of f on S are exactly the points x at which $\nabla f(x) = \lambda \nabla g(x)$ for some $\lambda \in \mathbb{R}$, they are exactly the points x for which $Ax = \lambda x$, i.e. exactly the unit vectors which are eigenvectors of A. Since S is compact, f actually attains its maximum and minimum on S, and as we have just seen, these are necessarily attained at eigenvectors of A. Thus, A has an eigenvector and eigenvalue on V, proving the lemma, and thus the spectral theorem.

Note that this is actually more than just an existence proof: for any A, we actually found the smallest eigenvalue by minimizing the Rayleigh quotient (and the largest by maximizing it). One could proceed inductively to find all eigenvalues: once one has the smallest one, λ_1 with an eigenvector v_{11}, restrict A to $\mathrm{Span}\{v_{11}\}^\perp$, and find the smallest eigenvalue and an eigenvector there. The eigenvalue one finds is $\geq \lambda_1$; if it is $= \lambda_1$, one found another, orthogonal to the first, eigenvector v_{12} in the λ_1-eigenspace; otherwise one found the second smallest eigenvalue λ_2 and an eigenvector v_{21}. In the first case one now restricts A to $\mathrm{Span}\{v_{11}, v_{12}\}^\perp$, in the second to $\mathrm{Span}\{v_{11}, v_{21}\}^\perp$, and proceeds inductively to find all eigenvalues and an orthonormal basis of eigenvectors.