9. Duality

This short chapter can be skipped without loss of continuity. Much of it can serve as a review of what has been covered so far. It owes much to the intriguing book [citGL].

Duality concerns mathematical concepts that come in pairs, that *complement* one another. Examples of interest in these notes include:

- $\subset$ and $\supset$;
- A *subset* S of T and its **complement**, $\backslash S := T \backslash S$;
- $\cap$ and $\cup$;
- $\forall$ and $\exists$;
- *1-1* and *onto*;
- *right* and *left* inverse;
- *bound* and *free*;
- *nullspace* and *range* of a linear map;
- an *invertible map* and its *inverse*;
- *column map* and *row map*;
- *synthesis* and *analysis*;
- a *basis* and its *inverse*;
- *columns* and *rows* of a matrix;
- a *matrix* and its (conjugate) *transpose*;
- a linear *subspace* and one of its *complement*s;
- dim and codim;
- the *vector space* X and its **dual**, $X' := L(X, \mathbb{F})$;
- the linear map $A \in L(X, Y)$ and its **dual**, $A' : Y' \to X' : \lambda \mapsto \lambda A$;
- a *norm* on the vector space X and the *dual norm* on X'.

Each such pair expresses a kind of *symmetry*. Such symmetry provides, with each result, also its 'dual', i.e., the result obtained by replacing one or more concepts appropriately by its complement. This leads to efficiency, both in the proving and in the remembering of results.

A classical example is that of *points* and *lines* in a geometry, and results concerning lines through points. E.g., *through every two distinct points there goes exactly one line*; its 'dual' statement is: *any two distinct lines have exactly one point in common.*

Another classical example is **DeMorgan's Law**, according to which any statement concerning the union, intersection and containment of subsets is true if and only if its 'dual' statement is true, i.e., the statement obtained by replacing each set by its complement and replacing $(\subset, \supset, \cap, \cup)$ by $(\supset, \subset, \cup, \cap)$, respectively. For example, the two 'distributive' laws

$$(R \cap S) \cup T = (R \cup T) \cap (S \cup T), \quad (R \cup S) \cap T = (R \cap T) \cup (S \cap T)$$

are 'dual' to each other. Again, having verified that *the intersection of a collection of sets is the largest set contained in all of them*, we have, by 'duality', also verified that *the union of a collection of sets is the smallest set containing all of them.*

Here are some specific examples concerning the material covered in these notes so far.

Let V, W be column maps. *If $V \subset W$ and W is 1-1, then so is V.* Its 'dual': *if $V \supset W$ and W is onto, then so is V.* This makes maximally 1-1 maps and a minimally onto maps particularly interesting as, by now, you know very well: *A column map is maximally 1-1 if and only if it is minimally onto if and only if it is a basis.*

Let $A \in \mathbb{F}^{m \times n}$. Then, *$A$ is 1-1(onto) if and only if A^{t} is onto(1-1).* In terms of the rows and columns of the matrix A and in more traditional terms, this says that the columns form a linearly independent (spanning) sequence if and only if the rows form a spanning (linearly independent) sequence. This is a special case of the result that $\operatorname{null} A = (\operatorname{ran} A^{\mathrm{t}})^{\perp}$, hence that $\dim \operatorname{null} A = \operatorname{codim} \operatorname{ran} A^{\mathrm{t}}$. By going from A to A^{t}, and from a subspace to its orthogonal complement, we obtain from these the 'dual' result that $\operatorname{ran} A = (\operatorname{null} A^{\mathrm{t}})^{\perp}$, hence that $\dim \operatorname{ran} A = \operatorname{codim} \operatorname{null} A^{\mathrm{t}}$.

Recall from (3.11) the factorization $A = A(:, \texttt{bound})\operatorname{rrref}(A)$. It supplies the corresponding factorization $A^{\mathrm{t}} = A^{\mathrm{t}}(:, \texttt{rbound})\operatorname{rrref}(A^{\mathrm{t}})$ with rbound the index sequence of bound columns of A^{t}, i.e. of bound *rows* of A. By combining these two factorizations, we get the more symmetric factorization

$$A = (\operatorname{rrref}(A^{\mathrm{t}}))^{\mathrm{t}} A(\texttt{rbound}, \texttt{bound}) \operatorname{rrref}(A),$$

which is called the **car**-factorization in [citGS].

9.1 Prove that, for any $A \in L(X, Y)$, $\operatorname{codim} \operatorname{null} A = \dim \operatorname{ran} A$.

9.2 In the list of pairs of complementary concepts, given at the beginning of this chapter, many of the pairs have been ordered so as to have the first term in each pair naturally correspond to the first term in any related pair.

For example, a right (left) inverse is necessarily 1-1 (onto).

Discover as many such correspondences as you can.

The dual of a vector space

The **dual of the vector space** X is, by definition, the vector space

$$X' := L(X, \mathbb{F})$$

of all linear maps into the underlying scalar field. Each such map is called a **linear functional** on X. (The term 'functional' is used to indicate a map, on a vector space, whose target is the underlying scalar field. Some books use the term 'form' instead.)

We have made much use of linear functionals, namely as the rows $\lambda_1, \ldots, \lambda_n$ of specific row maps (or data maps)

$$\Lambda^{\mathrm{t}} = [\lambda_1, \ldots, \lambda_n]^{\mathrm{t}} \in L(X, \mathbb{F}^n)$$

from the vector space X to n-dimensional coordinate space.

Example: If $X = \mathbb{F}^n$, then

$$X' = L(\mathbb{F}^n, \mathbb{F}) = \mathbb{F}^{1\times n} \sim \mathbb{F}^n,$$

and it has become standard to identify $(\mathbb{F}^n)'$ with $\mathbb{F}^n$ via

$$\mathbb{F}^n \to (\mathbb{F}^n)' : a \mapsto a^{\mathrm{t}}.$$

While this identification is often quite convenient, be aware that, strictly speaking, $\mathbb{F}^n$ and its dual are quite different objects. □

Here is a quick discussion of X' for an arbitrary finite-dimensional vector space, X. X being finite-dimensional, it has a basis, $V \in L(\mathbb{F}^n, X)$ say. Let

$$V^{-1} =: \Lambda^{\mathrm{t}} =: [\lambda_1, \ldots, \lambda_n]^{\mathrm{t}}$$

be its inverse. Each of its rows λ_i is a linear functional on X, hence

$$\Lambda := [\lambda_1, \ldots, \lambda_n]$$

is a column map into X'.

Λ is 1-1: Indeed, if $\Lambda a = 0$, then $\sum_i a(i)\lambda_i$ is the zero functional, hence, in particular, $\sum_i a(i)\lambda_i v_j = 0$ for all columns v_j of V. This implies that $0 = (\sum_i a(i)\lambda_i v_j : j = 1{:}n) = a^{\mathrm{t}}(\Lambda^{\mathrm{t}} V) = a^{\mathrm{t}} \operatorname{id}_n = a^{\mathrm{t}}$, hence $a = 0$.

It follows that $\dim \operatorname{ran} \Lambda = \dim \operatorname{dom} \Lambda = n$, hence we will know that Λ is also onto as soon as we know that the dimension of its target is $\le n$, i.e.,

$$\dim X' \le n.$$

For the proof of this inequality, observe that, for each $\lambda \in X'$, the composition λV is a linear map from $\mathbb{F}^n$ to $\mathbb{F}$, hence a 1-by-n matrix. Moreover, the resulting map

$$X' \to \mathbb{F}^{1\times n} \sim \mathbb{F}^n : \lambda \to \lambda V$$

is linear. It is also 1-1 since $\lambda V = 0$ implies that $\lambda = 0$ since V is invertible. Hence, indeed, $\dim X' \le n$.

(9.1) Proposition: For each basis V of the n-dimensional vector space X, the rows of its inverse, $V^{-1} =: \Lambda^{\mathrm{t}} =: [\lambda_1, \ldots, \lambda_n]^{\mathrm{t}}$, provide the columns for the basis $\Lambda = [\lambda_1, \ldots, \lambda_n]$ for X'. In particular, $\dim X' = \dim X$.

The two bases, Λ and V, are said to be **dual** or **bi-orthonormal** to signify that

$$\lambda_i v_j = \delta_{ij}, \qquad i, j = 1{:}n.$$

Here is the 'dual' claim.

(9.2)Proposition: Let X be an n-dimensional linear subspace of the vector space Y. Then, for each $\Lambda^{\mathrm{t}} \in L(Y, \mathbb{F}^n)$ that is 1-1 on X, there exists exactly one basis, V, for X that is dual or bi-orthonormal to Λ.

For every $\lambda \in Y'$, there exists exactly one $a \in \mathbb{F}^n$ so that

$$\lambda = \Lambda a \quad \text{on } X. \tag{9.3}$$

In particular, each $\lambda \in X'$ has a unique such **representation** Λa in $\operatorname{ran} \Lambda$.

Proof: Since $\dim X = \dim \operatorname{tar} \Lambda^{\mathrm{t}}$ and the restriction of $\Lambda^{\mathrm{t}} =: [\lambda_1, \ldots, \lambda_n]^{\mathrm{t}}$ to X is 1-1, it must be invertible, i.e., there exists exactly one basis V for X with $\Lambda^{\mathrm{t}} V = \mathrm{id}_n$, hence with Λ and V dual to each other.

In particular, $\Lambda := [\lambda_1, \ldots, \lambda_n]$ is a basis for its range. Let now $\lambda \in Y'$ and consider the equation

$$\Lambda ? = \lambda \qquad \text{on } X.$$

Since V is a basis for X, this equation is equivalent to the equation $(\Lambda ?)V = \lambda V$. Since

$$(\Lambda a)V = (\sum_i a(i)\lambda_i v_j : j = 1{:}n) = a^{\mathrm{t}}(\Lambda^{\mathrm{t}} V),$$

this equation, in turn, is equivalent to

$$?^{\mathrm{t}} \Lambda^{\mathrm{t}} V = \lambda V,$$

and, since $\Lambda^{\mathrm{t}} V = \mathrm{id}_n$, this has the unique solution $? = \lambda V = (\lambda v_j : j = 1{:}n)$. □

If X is not finite-dimensional, it may be harder to provide a complete description of its dual. In fact, in that case, one calls X' the **algebraic dual** and, for even some very common vector spaces, like $C[a\,.\,.\,b]$, there is no constructive description for its algebraic dual. If X is a normed vector space, one focuses attention instead on its **topological dual**. The topological dual consists of all a *continuous* linear functionals on X, and this goes beyond the level of these notes. Suffice it to say that, for any finite-dimensional normed vector space, the algebraic dual coincides with the topological dual.

The very definition of $0 \in L(X, \mathbb{F})$ ensures that $\lambda \in X'$ is 0 if and only if $\lambda x = 0$ for all $x \in X$. What about its dual statement: $x \in X$ *is* 0 *if and only if* $\lambda x = 0$ *for all* $\lambda \in X'$? For an arbitrary vector space, this turns out to require the Axiom of Choice. However, if X is a linear subspace of $\mathbb{F}^T$ for some set T, then, in particular,

$$\delta_t : X \to \mathbb{F} : x \mapsto x(t)$$

is a linear functional on X, hence the vanishing at x of all linear functionals in X' implies that, in particular, $x(t) = 0$ for all $t \in T$, hence $x = 0$.

(9.4) Fact: For any x in the vector space X, $x = 0$ if and only if $\lambda x = 0$ for all $\lambda \in X'$.

Proof: If X is finite-dimensional, then, by (9.1), the condition $\lambda x = 0$ for all $\lambda \in X'$ is equivalent, for any particular basis V for X with dual basis Λ for X', to having $b^{\mathrm{t}}\Lambda^{\mathrm{t}}Va = 0$ for all $b \in \mathbb{F}^n$ and for $x =: Va$. Since $\Lambda^{\mathrm{t}}V = \mathrm{id}_n$, it follows that $a = \Lambda^{\mathrm{t}}Va$ must be zero, hence $x = 0$. □

Finally, one often needs the following

(9.5) Fact: Every linear functional on some linear subspace of a vector space can be extended to a linear functional on the whole vector space.

Proof: If X is a linear subspace of the finite-dimensional vector space Y, then there is a basis $[V, W]$ for Y with V a basis for X. If now $\lambda \in X'$, then there is a unique $\mu \in Y'$ with $\mu[V, W] = [\lambda V, 0]$, and it extends λ to all of Y.

If Y is not finite-dimensional, then it is, once again a job for the Axiom of Choice to aid in the proof. □

The dual of an inner product space

We introduced inner-product spaces as spaces with a ready supply of linear functionals. Specifically, the very definition of an inner product $\langle , \rangle$ on the vector space Y requires that, for each $y \in Y$, $y^{\mathrm{c}} := \langle \cdot, y \rangle$ be a linear functional on Y. This sets up a map

$$^{\mathrm{c}} : Y \to Y' : y \mapsto y^{\mathrm{c}}$$

from the inner product space to its dual. This map is additive. It is also homogeneous in case $\mathbb{F} = \mathbb{R}$. If $\mathbb{F} = \mathbb{C}$, then the map is **skew-homogeneous**, meaning that

$$(\alpha y)^{\mathrm{c}} = \overline{\alpha} y^{\mathrm{c}}, \qquad \alpha \in \mathbb{F},\ y \in Y.$$

Either way, this map is 1-1 if and only if its nullspace is trivial. But, since $y^{\mathrm{c}} = 0$ implies, in particular, that $y^{\mathrm{c}}y = 0$, the positive definiteness required of the inner product guarantees that then $y = 0$, hence the map $y \mapsto y^{\mathrm{c}}$ is 1-1.

If now $n := \dim Y < \infty$, then, by (9.1)Proposition, $\dim Y' = \dim Y = n$, hence, by the Dimension Formula, $y \mapsto y^c$ must also be onto. This proves

> **(9.6) Proposition:** If Y is a finite-dimensional inner product space, then every $\lambda \in Y'$ can be written in exactly one way as $\lambda = y^c$ for some $y \in Y$.
>
> We say in this case that y^c **represent**s λ.

If Y is not finite-dimensional, then the conclusion of this proposition still holds, provided we consider only the topological dual of Y and provided Y is 'complete', the very concept we declared beyond the scope of these notes when, earlier, we discussed the Hermitian (aka conjugate transpose) of a linear map between two inner product spaces.

The dual of a linear map

Any $A \in L(X, Y)$ induces in a natural way the linear map

$$A' : Y' \to X' : \lambda \mapsto \lambda A.$$

This map is called the **dual** to A.

If also $B \in L(Y, Z)$, then $BA \in L(X, Z)$ and, for every $\lambda \in Z'$, $\lambda(BA) = (\lambda B)A = A'(B'(\lambda))$, hence

$$(BA)' = A'B', \quad A \in L(X, Y),\ B \in L(Y, Z). \tag{9.7}$$

If both X and Y are coordinate spaces, hence A is a matrix, then, with the identification of a coordinate space with its dual, the dual of A coincides with its transpose i.e.,

$$A' = A^t, \qquad A \in \mathbb{F}^{m\times n} = L(\mathbb{F}^n, \mathbb{F}^m).$$

If $Y = \mathbb{F}^m$, hence A is a row map, $A = \Lambda^t = [\lambda_1, \ldots, \lambda_m]^t$ say, then, with the identification of $(\mathbb{F}^m)'$ with $\mathbb{F}^m$, $(\Lambda^t)'$ becomes the column map

$$(\Lambda^t)' = [\lambda_1, \ldots, \lambda_m] = \Lambda.$$

In this way, we now recognize a row map on X as the **pre-dual** of a column map into X'.

If $X = \mathbb{F}^n$, hence A is a column map, $A = V = [v_1, \ldots, v_n]$ say, then, with the identification of $(\mathbb{F}^n)'$ with $\mathbb{F}^n$, V' becomes a row map on Y', namely the row map that associates $\lambda \in Y'$ with the n-vector $(\lambda v_j : j = 1{:}n)$. Its rows are the linear functionals

$$Y' \to \mathbb{F} : \lambda \mapsto \lambda v_j$$

on Y' 'induced' by the columns of V. Each of these rows is therefore a linear functional on Y', i.e., an element of $(Y')'$, the **bidual** of Y. Also *if, addition, V is 1-1, the V' is onto.* Indeed, in that case, V is a basis for its range, hence has an inverse, Λ^t say. Now, for arbitrary $b^t \in (\mathbb{F}^n)'$, $b^t = b^t(\Lambda^t V) = (\Lambda b)V$, with Λb a linear functional on $\operatorname{ran} V$. By (9.5)Fact, there is some $\lambda \in Y'$ that agrees with Λb on $\operatorname{ran} V$. In particular, $\lambda V = (\Lambda b)V = b^t$, showing that V' is, indeed onto.

Finally, if X and Y are arbitrary vector spaces but A is of finite rank, then, for any basis V for $\operatorname{ran} A$ with dual basis M, we have

$$A = V\mathrm{M}^t A =: V\Lambda^t,$$

and, by (???), this is a minimal factorization for A. It follows that

$$A' = \Lambda V',$$

and, since V is 1-1, hence V' is onto, and also Λ is 1-1, we conclude that Λ is a basis for $\operatorname{ran} A'$, hence $\Lambda V'$ is a minimal factorization for $\operatorname{ran} A'$.

In particular, $\operatorname{rank} A' = \operatorname{rank} A$. Also, if A is onto,

10. The powers of a linear map and its spectrum

If $\operatorname{tar} A = \operatorname{dom} A$, then we can form the powers

$$A^k := \underbrace{AA\cdots A}_{k \text{ factors}}$$

of A. Here are some examples that show the importance of understanding the powers of a linear map.

Examples

Fixed-point iteration: A standard method for solving a large linear system $A? = y$ (with $A \in \mathbb{F}^{n\times n}$) is to split the matrix A suitably as

$$A = M - N$$

with M 'easily invertible', and to generate the sequence $x_0, x_1, x_2, \ldots$ of approximate solutions by the **iteration**

$$(10.1) \qquad x_k := M^{-1}(Nx_{k-1} + y), \quad k = 1, 2, \ldots.$$

Assuming this iteration to converge, with $x := \lim_{k\to\infty} x_k$ its limit, it follows that

$$(10.2) \qquad x = M^{-1}(Nx + y),$$

hence that $Mx = Nx + y$, therefore finally that $Ax = (M - N)x = y$, i.e., the limit solves our original problem $A? = y$.

Let $\varepsilon_k := x - x_k$ be the **error** in our kth approximate solution. Then on subtracting the iteration equation (10.1) from the exact equation (10.2), we find that

$$\varepsilon_k = x - x_k = M^{-1}(Nx + b\ - (Nx_{k-1} + y)) = M^{-1}N\varepsilon_{k-1}.$$

Therefore, by induction,

$$\varepsilon_k = B^k\varepsilon_0, \quad \text{with } B := M^{-1}N$$

the **iteration map**. Since we presumably don't know the solution x, we have no way of choosing the **initial guess** x_0 in any special way. For convergence, we must therefore demand that

$$\lim_{k\to\infty} B^k z = 0 \quad \text{for all } z \in \mathbb{F}^n.$$

It turns out that this will happen if and only if all eigenvalues of B are less than 1 in absolute value.

random walk: Consider a random walk on a graph G. The specifics of such a random walk are given by a **stochastic** matrix M of order n, with n the number of vertices in the graph. This means that all the entries of M are nonnegative, and all the entries in each row add up to 1, i.e.,

$$M \ge 0, \qquad Me = e,$$

with e the vector with all entries equal to 1,

$$e := (1, 1, 1, \ldots, 1).$$

The entries of M are interpreted as probabilities: $M(i, j)$ gives the probability that, on finding ourselves at vertex i, we would proceed to vertex j. Thus, the probability that, after two steps, we would have gone from

