Index of linear operators

László Erdős

Oct 28, 2003

1 Notations, definitions

All maps are linear and all spaces are linear vectorspaces.

Let $M : X \longrightarrow U$ be a linear map. If both X, U are finite dimensional, then we know that

$$\text{dim}N_M + \text{dim}R_M = \text{dim}X$$

for example by the isomorphism theorem, $R_M \cong X/N_M$ and $\text{dim}X/Y = \text{dim}X - \text{dim}Y$.

However, it is typically not very informative in infinite dimensions, since this relation boils down to $\infty = \infty$.

Definition 1.1 *$Y \subset X$ subspace, then the codimension of Y is defined as*

$$codimY := dimX/Y$$

(it may be infinite).

Note that the relation $\text{codim}Y = \text{dim}X - \text{dim}Y$ is meaningless if it is of the form $\infty - \infty$. Still, $\text{codim}Y$ may exist and be finite.

If X, U are finite dimensional, then clearly

$$\text{dim}N_M - \text{codim}R_M = \text{dim}X - \text{dim}U \tag{1.1}$$

and this expression is independent of M. Based upon this, we define

Definition 1.2 *Let $M : X \longrightarrow U$ with $dimN_M, codimR_M < \infty$. Then the index of M is defined as*

$$indM := dimN_M - codimR_M$$

Note that the index may be finite even for maps between infinite dimensional spaces, when (1.1) is meaningless.

The index is the difference of two important quantities related to the map M. Recall that $\dim N_M$ expresses the degeneracy of the solutions to the equation $Mx = b$. When $\dim N_M \geq 1$, then the solution space is of the form $x = x_0 + N$, where x_0 a particular solution, i.e. the solution space can be described with $\dim N_M$ free parameters. The quantity $\mathrm{codim} R_M$ is equal to the number of linear constraints that b has to satisfy so that b be in the the range of M, i.e this is the number of conditions on b to ensure that $Mx = b$ has a solution.

Definition 1.3 *A linear map $G : X \longrightarrow U$ is called degenerate if $\dim R_G < \infty$.*

Think of them as "small" maps.

Check: that the degenerate maps form an ideal in the sense that their sums is degenerate and whenever they are composed with another linear map from either side (MG or GM), the result is also degenerate. Recall that multiplication of maps means composition: $ML = M \circ L$.

Definition 1.4 *The maps $M : X \longrightarrow U$ and $K : U \longrightarrow X$ are called pseudoinverses of each other if there exist degenerate maps G, G' (between the right spaces) s.t.*

$$LM = I + G$$

$$ML = I + G'$$

Check that pseudoinverse of a map M is not unique: if L is a pseudoinverse of M, so is $L + G$ for any degenerate map G.

Example: Let $X = U$ be the set of infinite sequences $\{x = (a_1, a_2, \ldots)\}$. Let L and R be the left and right shifts:

$$Lx := (a_2, a_3, \ldots)$$

$$Rx := (0, a_1, a_2, \ldots)$$

for $x(a_1, a_2, \ldots)$. Check that L and R are pseudoinverses of each other, $LR = I$ but $RL \neq I$ (such thing cannot happen in finite dimensions).

Theorem 1.5 *$M : X \longrightarrow U$ has pseudoinverse $\Longleftrightarrow$ $\dim N_M < \infty$ and $\mathrm{codim} R_M < \infty$.*

Proof. $\Longrightarrow$ direction. Let L be a pseudoinverse. Check that $LM = I + G$ implies $N_M \subset N_{I+G} \subset R_G$, hence $\dim N_M \leq \dim R_G < \infty$. Similarly, $ML = I + G'$ implies $R_M \supset R_{I+G'} \supset N_{G'}$, hence $\mathrm{codim} R_M \leq \mathrm{codim} N_{G'} = \dim R_{G'}$, the latter is from the isomorphism theorem.

$\Longrightarrow$ direction. We need a

Lemma 1.6 *(Existence of a complement). Let $N \subset X$ subspace. Then there exists $Y \subset X$ subspace s.t.*

$$X = N \oplus Y$$

i.e. every $x \in X$ can be uniquely written as $x = n + y$ with some $n \in N$, $y \in Y$. The space Y is called a complement of N in X.

Remark: The complement is not unique. It is not the concept of orthogonal complement (which is unique), since there is no scalar product. OF course in scalar product spaces the orthogonal complement is a complement in the sense above.

Check: The map $P : x \to Px := n$ is well defined, linear and projects onto N (i.e. $R_P = N$ and $P^2 = P$). This is called the projection map onto N in the direction of Y. It depends on Y, sometimes we call it P_Y.

Check: There is a canonical isomorphism between $Y \cong X/N$.

Proof. We need the Zorn lemma that states the following:

Lemma 1.7 *Let $(S, \prec)$ be a partially ordered set such that every totally ordered subset $S' \subset S$ has an upper bound, i.e. there is $s \in S$ s.t. $t \prec s$ for any $t \in S'$. Then S has a maximal element, i.e. an element $\sigma \in S$ s.t. if $\sigma \prec t$ for some $t \in S$, then $\sigma = t$.*

Remark: (i) Read it carefully, make sure you understand it. It is really a compactness type statement, like Bolzano-Weierstrass theorem (a compact subset of $\mathbf{R}$ has a maximal element), except that the set is not totally ordered.

Also recall that maximal is not the same as maximum. Maximal element of a set is an element s.t. there is no bigger element. Maximum element is an element that is bigger than anything else in the set.

(ii) Zorn lemma is not really a lemma. It is an extra axiom that we add to the usual (Zermelo-Frenkel) axioms of set theory. It is equivalent to the Axiom of Choice (For any set E, there exists a map c from the power set $P(E)$ of E into E such that $c(B) \in B$ whenever $B \neq \emptyset$; in other words that we can select an element from every nonempty subset).

(iii) Zorn lemma is a convenient mathematical tool to prove many theorems in full generality. Some people do not like it because it inherently brings a nonconstructive flavor into mathematics. This is mostly a philosophical debate. The fact is, that every theorem in analysis can be proven without the Zorn lemma (sometimes with a bit more efforts), but the proof is valid only for constructive objects.

For example, if you want to prove the lemma above for an arbitrary vectorspace, then you need Zorn lemma, since the arbitrary vectorspace can be so big that constructive methods cannot cover it. But if you want to prove the lemma for any (more or less) explicitly given vectorspace X and its (more or less) explicitly given subspace N, then you can do it without Zorn lemma (but the proof may slightly depend on your vectorspace).

In practice one assumes Zorn lemma because the proofs are easier in this way, but one keeps in mind, that it is not really necessary.

We go back to the proof of the existence of the complement. Let

$$\mathcal{Y} := \Big\{ Y \subset X \ \text{ subspace } \ : Y \cap N = \{0\} \Big\}$$

be the collection of all subspaces of X that intersect N trivially.

Check that $\mathcal{Y}$ is partially ordered for the inclusion (i.e. $Y_1 \prec Y_2$ iff $Y_1 \subset Y_2$).

Check that every totally ordered subcollection of $\mathcal{Y}$, $\{Y_\alpha\}_{\alpha \in I}$ has an upper bound in $\mathcal{Y}$, namely the union $\cup_{\alpha \in I} Y_\alpha$.

By Zorn lemma, there exists a maximal element, $Y \in \mathcal{Y}$. Check that this is the good one. □

Now we can go back to the $\Longleftarrow$ direction of our theorem. We can find subspaces Y abd V s.t.

$$X = N_M \oplus Y, \qquad U = R_M \oplus V$$

and $\dim V < \infty$ since $\text{codim} R_M < \infty$.

Check that the map M restricted to Y, $M|_Y Y \longrightarrow R_M$ is invertible. Define a map $K : U \longrightarrow X$ that is $(M|_Y)^{-1}$ on R_M and is 0 on V and is extended linearly to the whole $R_M \oplus V$. Check that this is a pseudoinverse of M. □

Theorem 1.8 *Let $M : X \longrightarrow U$, $L : U \longrightarrow W$. Suppose that M, L both have pseudoinverses. Then*

(i) LM has pseudoinverse

(ii)

$$ind(LM) = ind(L) + ind(M) \tag{1.2}$$

Proof. (i) Exercise (multiply the two pseudoinverses in the right order and get the pseudoinverse of the product)

(ii) We need a lemma

Lemma 1.9 *Consider the following exact sequence of finite dimensional vectorspaces*

$$0 \xrightarrow{T_0} V_1 \xrightarrow{T_1} V_2 \xrightarrow{T_2} \dots \xrightarrow{T_{n-2}} V_{n-1} \xrightarrow{T_{n-1}} 0$$

(recall, exactness means $R(T_j) = N(T_{j+1})$). Then

$$\sum (-1)^j \dim V_j = 0$$

Proof. By the dimension formula and the exactness

$$\dim V_j = \dim N_{T_j} + \dim R_{T_j} = \dim N_{T_j} + \dim N_{T_{j+1}}$$

Adding up these equalities for $j = 1, 2, \dots n-1$ with prefactor $(-1)^j$, using $\dim N_{T_0} = 0$, $\dim V_{n-1} = \dim N_{T_{n-1}}$, we obtain the claim. □

Armed with this, we construct the following exact sequence:

$$0 \longrightarrow N_M \xrightarrow{I_0} N_{LM} \xrightarrow{M} N_L \xrightarrow{Q} U/R_M \xrightarrow{L} W/R_{LM} \xrightarrow{E} W/R_L \longrightarrow 0$$

where I_0 is the natural embedding, Q is the factormap $U \mapsto U/R_M$ restricted to N_L and E is the grouping map of equivalence classes using $R_{LM} \subset R_L$.

Check that at every step this is an exact sequence.

Using the lemma, we obtain

$$\dim N_M - \dim N_{LM} + \dim N_L - \operatorname{codim} R_M + \operatorname{codim} R_{LM} - \operatorname{codim} R_L = 0$$

and the claim (1.2) follows. □.

The following theorem shows the stability of index under a "small" (i.e. degenerate) perturbation.

Theorem 1.10 *Let $M : X \longrightarrow U$ have a pseudoinverse and let $G : X \longrightarrow U$ be degenerate. Then $M + G$ has pseudoinverse and*

$$ind(M + G) = ind M$$

Proof. First we prove it for the special case $X = U$, $M = I$. Let $K = I + G$ and we need that $\operatorname{ind} K = 0$, since $ind I = 0$ trivially. Let $I_0 : N_G \to X$ be the embedding map. Check that I_0 has pseudoinverse (hint: construct a complement V to N_G in X. Since G is degenerate, $\operatorname{codim} N_G < \infty$, show that this implies that V is finite dimensional, then $I - P_V$ is a good pseudoinverse.) Also check that $I_0 = K I_0$, then apply the product rule

$$\operatorname{ind} I_0 = \operatorname{ind} K + \operatorname{ind} I_0$$

Since $\operatorname{ind} I_0$ is finite (it is $-\operatorname{codim} N_G$), we get the claim.

The general case easily follows from this. Let L be the pseudoinverse of M, i.e. $LM = I{+}G$. Then

$$\mathrm{ind}(L) + \mathrm{ind}(M) = \mathrm{ind}(LM) = \mathrm{ind}(I + G) = 0$$

by the product rule and special case of theorem that we checked above.

But L is also the pseudoinverse of $M + G$, so one gets similarly:

$$\mathrm{ind}(L) + \mathrm{ind}(M + G) = 0$$

and the claim follows from these last two equations. $\square$