Lec 3: Mathematical Economics

Inner Product Spaces

Sugata Bag

Delhi School of Economics

22nd August 2013

Definitions

Let V be a real or complex vector space over a field F of scalars (think $\Re^n$ over $\Re$ or C^n over C; but spaces of real-valued or complex valued functions are other important examples).

Definition

A function from $f : V \times V \to F$ (for every ordered pair (x,y) of vectors, we denote $f(x, y)$ by (x, y)) is an **inner product** on V if
(i) $(x, y) = \overline{(y, x)}$ (Conjugate symmetry)
(ii) $(\alpha x + \beta y, z) = \alpha(x, z) + \beta(y, z), \ \forall \alpha, \beta \in F, \ \forall x, y, z \in V$ (Linearity)
(iii) $(x, x) \geq 0$, with '$=$' iff $x = 0$. (Nonnegativity)

Remark (1). Recall that if $a + ib$ and $c + id$ are 2 complex numbers then their sum is $(a + c) + i(b + d)$, their product is $(a + ib)(c + id) = (ac - bd) + i(ad + bc)$ (since $i^2 = -1$), their quotient $\frac{(a+ib)}{(c+id)} = \frac{a+ib}{c+id}\frac{c-id}{c-id} = \frac{(ac+bd)+i(bc-ad)}{c^2+d^2}$.
A complex number $(a + ib)$ can be viewed as the ordered pair (a, b) on the (complex) plane, so $(a^2 + b^2)^{1/2}$ is its distance from the origin

Definitions

The **Conjugate** of $(a+ib)$ is $(a-ib)$, denoted $\overline{a+ib}$, and their sum and product are real numbers.

Remark (2). Since there's no difference between a real number and its conjugate, if V is a real space then $\overline{(y,x)} = (y,x)$ and condition (i) in the definition is just symmetry.

Remark (3). Let $y = (i, 2-i, 3)^T$ be a (column) vector in C^3. Take its transpose and then take the conjugate. $\overline{y^T} = (-i, 2+i, 3)$. We get the same result if we first take the conjugate vector and then take its transpose. The 'conjugate transpose' of a vector y is denoted y^*. For real spaces, y^* is just y^T.

Lemma

Let V be a real or complex n-dimensional v.s. Then the (scalar-valued) function defined by $(x,y) = y^*x = \sum_{k=1}^n x_k \bar{y}_k$ for every (ordered) pair of vectors x, y is an ***inner product*** *on V.*

Just check that this function satisfies conditions (i) - (iii) in definition.

Definitions

For example, $\overline{(y,x)} = \overline{\sum_1^n y_k \bar{x_k}} = \sum_1^n \overline{y_k \bar{x_k}} = \sum_1^n \bar{y_k} x_k = (x,y)$. So conjugate symmetry holds. Note that if V is a real space, the definition of inner product reduces to the familiar dot product.

Remark (4). Notice that $(x,y) = y^* x$, where x is taken as a column vector, and y^* is a row vector (transpose of a column vector).

Remark (5). The benefit of conjugation is that for all $x \in V$, (x,x) is a real number. Indeed,

$$(x,x) = x^* x = (a_1 - ib_1, \ldots, a_n - ib_n)(a_1 + ib_1, \ldots, a_n + ib_n)^T$$
$$= \sum_{k=1}^n (a_k^2 + b_k^2).$$

This helps us to define, for each $x \in V$, its *norm* (or distance from the origin) as a real number.

Definition

Let V be a real or complex v.s. The **Euclidean Norm** of a vector $x \in V$, denoted $||x||$ is the real number $(x,x)^{1/2}$.

We say that our inner product **induces** the above norm.

Definitions

If $x = (x_1, ..., x_n)$, where $x_k = a_k + ib_k$, we write $||x|| = (\sum_1^n |x_k|^2)^{1/2}$, where $|x_k|^2 = (a_k^2 + b_k^2)$. For real spaces, (x, x) is the usual Euclidean distance $(\sum_1^n x_k^2)^{1/2}$.

Remark (6). Notation and Terminology. Treil uses (x, y) for inner product; since this is also notation for ordered pair, perhaps more preferable is $< x, y >$. But let's be consistent with our textbook.

Remark (7). A vector space V along with some inner product defined on it is called an **inner product space** or a **pre-Hilbert space**. Once the inner product is used to induce a norm, we can work with notions of convergence of sequences etc. A pre-Hilbert space that is "complete" (w.r.t. the induced norm) is called a **Hilbert space**. (A pre-Hilbert space is complete if every Cauchy sequence in it converges to a vector in the space itself).

Definition

Orthogonality: 2 vectors $u, v \in V$ are called **orthogonal** (denoted by $u \perp v$) if $(u, v) = 0$ i.e. inner product is zero.

Properties of Inner Product

1. $(x, \alpha y + \beta z) = \bar{\alpha}(x, y) + \bar{\beta}(x, z), \forall \alpha, \beta \in F, x, y, z \in V.$
Proof. $(\alpha y + \beta z)^* x = \bar{\alpha} y^* x + \bar{\beta} z^* x.$

2. Let $x \in V$. Then $(x, y) = 0, \forall y \in V$ iff $x = 0$.
Proof. If $x = 0$, then $(x, y) \equiv y^* x = 0$. If $x \neq 0$, then there exists y, namely $y = x$, s.t. $(x, y) = (x, x) > 0$.

3. Let $x, y \in V$. Then $(x, z) = (y, z) \forall z \in V$ iff $x = y$.
Proof. By (2), $((x - y), z) = 0 \forall z$ iff $(x - y) = 0$ or $x = y$. But $((x - y, z) = 0$, by linearity of the inner product, means $(x, z) - (y, z) = 0$.

4. Suppose $A, B : X \to Y$ are LTs or matrices that satisfy $(Ax, y) = (Bx, y)$, for all $x \in X$ and all $y \in Y$. Then $A = B$.
Proof. For fixed x, (3) implies $Ax = Bx$. Since we get this no matter what x we fix, the mappings A, B are identical.

Properties of Inner Product ... contd...

Theorem

Property (5) *(Cauchy-Schwarz Inequality).*
$|(x,y)| \leq ||x|| \; ||y||, \forall x, y \in V.$

Proof.

Let $x, y \in V$ and $t \in F$. Then

$$0 \leq ||x - ty||^2 = (x - ty, x - ty) = (x, x - ty) - (ty, x - ty)$$
$$= (x, x) - \bar{t}(x, y) - t(y, x) + \bar{t}(ty, y)$$
$$= ||x||^2 - \bar{t}(x, y) - t(y, x) + |t|^2 \; ||y||^2$$

Now, at $t = (x, y)/||y||^2$, (which minimizes the quadratic in t in the real-valued case), thus the expression equals to -

$$||x||^2 - \frac{\overline{(x,y)}(x,y)}{||y||^2} - \frac{\overline{(y,x)}(y,x)}{||y||^2} + \frac{|(x,y)|^2||y||^2}{(||y||^2)^2}$$

So at $t = (x, y)/||y||^2$, $\quad 0 \leq ||x||^2 - \frac{|(x,y)|^2}{||y||^2}$. *QED*.

** Think of the Cauchy-Schwarz inequality obtaining from the familiar cosine rule, since $-1 \leq \cos(\theta) \leq 1$.**

□

Properties of Inner Product ... contd...

Property (6). **Triangle Inequality** $||x+y|| \leq ||x|| + ||y||, \ \forall x, y \in V$.

Proof.

$||x+y||^2 = (x+y, x+y) = (x,x) + (x,y) + (y,x) + (y,y)$. By Cauchy's inequality, both (x,y) and (y,x) must be $\leq ||x|| \ ||y||$. So the RHS is $\leq ||x||^2 + 2||x|| \ ||y|| + ||y||^2$. QED. □

We can define a norm more generally than the Euclidean norm.

Definition

A **norm** on a vector space V is a function that associates, with each $v \in V$, a real number $||v||$ s.t.

(i) $||\alpha v|| = |\alpha| \cdot ||v||, \ \forall v \in V, \alpha \in F$ (Homogeneity)

(ii) $||u+v|| \leq ||u|| + ||v||, \ \forall u, v \in V$ (Triangle Inequality)

(iii) $||v|| \geq 0$, with '$=$' iff $v = 0$. (Orthogonal Bases)

Definitions

Recall that 2 vectors $u, v \in V$ are called **orthogonal** (denoted $u \perp v$) if $(u, v) = 0$. (The motivation is again the cosine rule, since $\cos(\pi/2) = 0$). **Note**, $(u, v) = v^* u = 0$, implies, taking conjugate transpose of both sides, $(v^* u)^* = u^* v = 0^* = 0$, i.e. $(v, u) = 0$. So $u \perp v$ *iff* $v \perp u$.

Definition

A vector v is orthogonal to a subspace E if $v \perp w, \ \forall w \in E$.

Lemma

Let span$(\{w_1, ..., w_r\}) = E$. *Then* $v \perp E$ *iff* $v \perp w_k, \ \forall k = 1, ..., r$.

Proof.

Suppose $v \perp w_k, \ \forall k = 1, ..., r$ and let $w \in E$. So, $w = \sum_1^r c_k w_k$ for some c_k's. So, $(w, v) = (\sum_1^r c_k w_k, v) = \sum_1^r c_k (w_k, v) = 0$. So $v \perp w, \forall w \in E$. Conversely v not $\perp w_i$, for some i, implies v is not orthogonal to E. □

Definitions contd...

Definition

A set of vectors $\{v_1, ..., v_n\}$ is orthogonal if every pair $v_i \perp v_j, \forall i, j \in \{1, ..., n\}, i \neq j$.

Lemma

A set of vectors $\{v_1, ..., v_n\}$ *is non-zero & orthogonal implies that* $\{v_1, ..., v_n\}$ *linearly independent.*

Proof.

Let $\sum c_i v_i = 0$. So, $(\sum c_i v_i, v_j) = 0$. Now, $(\sum c_i v_i, v_j) = \sum_{i \neq j} c_i (v_i, v_j) + c_j ||v_j||^2$, which due to orthogonality equals to $c_j ||v_j||^2$. Since $||v_j||^2 > 0$, this implies $c_j = 0$. Since this is true irrespective of choice of j, LI follows. □

The Representation

Suppose we represent a vector $v \in V$ in terms of a basis $\{v_1, ..., v_n\}$, as $v = \sum c_i v_i$. In order to find the **Representation** (i.e. the c_i's) we have to solve a system of linear equations, namely $Ac = v$, where $A = (\mathbf{v_1} \ \mathbf{v_2} \ ... \ \mathbf{v_n})$ is the matrix whose columns are the basis vectors, and $c = (c_1, ..., c_n)^T$.

However, if the basis is orthogonal, and $v = \sum c_i v_i$, we don't need to solve a linear system to find the c_i's. Instead, notice that for any $k = 1, ..., n$, $(v, v_k) = (\sum c_i v_i, v_k) = \sum_{i \neq k} c_i (v_i, v_k) + c_k ||v_k||^2$, which, due to orthogonality, equals to $c_k ||v_k||^2$. So, we have -

$c_k = \frac{(v, v_k)}{||v_k||^2}, \ \forall k = 1, ..., n$.

Orthogonal Projection Definition

Definition

Let $v \in V$ and E a subspace of V. The orthogonal projection of v onto E, denoted $P_E v$, is the vector $w \in E$ s.t. $(v - w) \perp E$.

This presumes an orthogonal projection always exists and is unique. Visually, we are dropping a perpendicular from v to a plane or hyperplane E. If $P_E v$ exists, it is unique.

Theorem

$P_E v$ minimises distance from v to E: $||v - P_E v|| \leq ||v - x||, \ \forall x \in E$, and $' =' \Rightarrow x = P_E v$.

Proof.

Let $P_E v \equiv w$. Since $w, x \in E$ and E is a subspace, $(w - x) \in E$. Since $(v - w) \perp E$, $(v - w) \perp (w - x)$. Applying Pythagorean theorem to the right triangle Δvwx, $||v - x||^2 = ||v - w||^2 + ||w - x||^2 \geq ||v - w||^2$. Thus the equality holds iff $x = w$. □

Computing Orthogonal Projection

Proposition. Let $\{w_1, ..., w_r\}$ be an orthogonal basis in E. Then $P_E v = w = \sum_{k=1}^{r} \alpha_k w_k$, where $\alpha_k = \frac{(v, w_k)}{||w_k||^2}$, $\forall k = 1, ..., r$.

Proof. Write $P_E v \equiv w$. Since $w \in E$, $w = \sum_{k=1}^{r} c_k w_k$, where $c_k = \frac{(w, w_k)}{||w_k||^2}$. Finally, since $v = w + (v - w)$ and $(v - w) \perp w_k$, $\forall k = 1, ..., r$, we have
$(v, w_k) = (w + v - w, w_k) = (w, w_k) + (v - w, w_k) = (w, w_k)$.

Note. **(1)** Basically, the scalars in the LC for the projection w involve (v, w_k) rather than (w, w_k) because $v = w + (v - w)$, and $(v - w)$ has zero inner product with vectors in E, being orthogonal.

Note. **(2)** The projection P_E is a LT $(P_E : V \to V)$. Indeed, for vectors v, v' and scalars c_1, c_2, we have $P_E(c_1 v + c_2 v') = \sum_{k=1}^{r} \frac{(c_1 v + c_2 v', v_k)}{||v_k||^2} v_k$.
By linearity of inner product, this
$= c_1 \sum \frac{(v, v_k)}{||v_k||^2} v_k + c_2 \sum \frac{(v', v_k)}{||v_k||^2} v_k = c_1 P_E v + c_2 P_E v'$.

Orthogonal Complement

Note. (3) $P_E = \sum_{k=1}^{r} \frac{1}{||v_k||^2} w_k w_k^*$ (where $\{w_1, ..., w_r\}$ is an orthogonal basis in E). If V is n-dimensional, w_k is a $n \times 1$ vector, so that $w_k w_k^*$ is an $n \times n$ matrix. Indeed,

$P_E v = \sum \frac{(v, w_k)}{||w_k||^2} w_k = \sum \frac{1}{||w_k||^2} w_k (v, w_k) = \sum \frac{1}{||w_k||^2} w_k w_k^* v.$

The 2nd equality follows since (v, w_k) is a scalar.

Definition

Let V be a v.s. and E be a subspace of V. Then the **orthogonal complement** of E in V, denoted $E^\perp$ (pronounced, *E perp*) is the set of vectors **u** such that **u** is orthogonal to all vectors in E i.e.$\{u \in V | u \perp E\}$.

Note. (4) (i) $E^\perp$ is a subspace. (ii) $(E^\perp)^\perp = E$.

Note. (5) $v = P_E v + (v - P_E v)$ is the unique representation of v as the sum of 2 vectors, one lying in E (namely $P_E v$) and the other (namely $(v - P_E v)$) lying in $E^\perp$. Uniqueness results from the uniqueness of $P_E v$.

We write $V = E \oplus E^\perp$. In $\Re^2$, $(x_1, x_2) = (x_1, 0) + (0, x_2)$ is the simplest such representation, with E being the horizontal axis.

Orthogonal Complement - Example

Ex-1. Let $V = R^2$ and W be the subspace spanned by $(1, 2)$. Then $W^{\perp}$ is the set of vectors (a, b) with

$$(a, b) \cdot \mathbf{c}(1, 2) = 0 \text{ , (where } \mathbf{c} \neq 0 \text{ be some constant)}$$

or, $ac + 2bc = 0 \Rightarrow a + 2b = 0.$

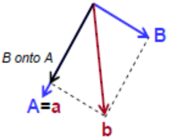
This is a 1 dimensional vector space spanned by (-2,1).

In the example above the orthogonal complement was a subspace. This will always be the case.

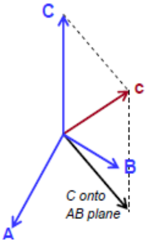
Ex-2. For more of an explanatory example on this please refer to page 2 of Lecture Notes 6.

Orthogonalisation: The Gram Schmidt Procedure

- **The Gram-Schmidt orthogonalization procedure operates as follows**
  - Start with A=a
    - This gives the first direction
  - The second direction must be perpendicular to A
    - Start with B=b and subtract its projection along A
    - This leaves the perpendicular part, which is the orthogonal vector B



  - The third direction must be perpendicular to A and B
    - Start with C=c and subtract its projections along A and B



  - The resulting vectors {A,B,C} are orthogonal and span the same space as {a,b,c}

Application - Least Squares ... I

You have a model which explains the wage y of an individual as a function of k explanatory variables, such as years of schooling, gender etc.

Say, $y = \sum_j x_j \beta_j + \epsilon$, where $\epsilon, x_j's$ are random variables. This is the true model in the population but you do not know the β_j's.

One way to estimate them is as follows:

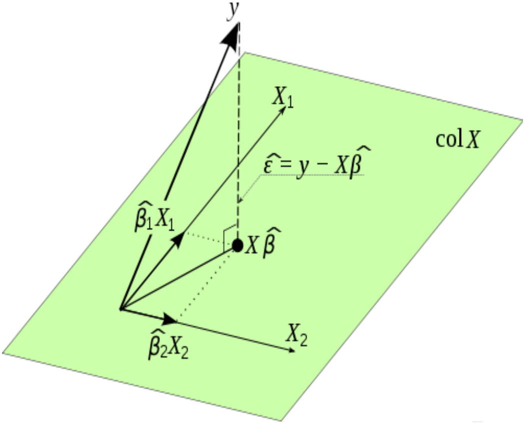
You collect data on say the wages of n individuals, y_i equals wage of individual i, as well as on k variables that may explain wage $(x_{i1}, ..., x_{ik}) \equiv x_i^T$, for individual i). Typically, $n >> k$. You want β_i's that minimise $\sum_{i=1}^n (y_i - x_i^T \hat{\beta})^2$. $\longleftarrow$: the sum of squares of errors.

That means, choose $\hat{\beta} = (\hat{\beta}_1, ..., \hat{\beta}_k)^T$ to minimise $||y - X\hat{\beta}||^2$, where $y = (y_1, ..., y_n)^T$, and X is the $n \times k$ matrix whose jth column contains observations x_{ij} for all individuals i. Since $||y - X\hat{\beta}||^2 \geq 0$, this is minimised iff $||y - X\hat{\beta}||$ is minimised.

If we can find $\hat{\beta}$ that solves the linear system $y = X\hat{\beta}$ we are done. But since $n >> k$, $Rank(X) \leq k < n$, and $y \in \Re^n$, it is 'unlikely' that $y \in R(X)$, the span of the columns of X. (Note $X : \Re^k \rightarrow \Re^n$).

Least Squares ... II

So the general problem is to find $\hat{\beta}$ s.t. $X\hat{\beta} = P_{R(X)}y$, i.e. s.t. $X\hat{\beta}$ is the orthogonal projection of y onto $R(X)$, as this minimizes the distance between y and $R(X)$. (Notice that for any β, $X\beta \in R(X)$; we want the $X\beta$ which is the closest point to y, from $R(X)$).



Least Squares ... III

If we know an orthogonal basis, (or construct one using say Gram-Schmidt method) in the subspace $R(X)$, it is easy to get $P_{R(X)}y$; then just solve the linear system $X\hat{\beta} = P_{R(X)}y$ for $\hat{\beta}$. But constructing an orthogonal basis takes more calculations than the following simpler method.

If $X\hat{\beta}$ is the orthogonal projection of y onto $R(X)$, then

$$(y - X\hat{\beta}) \perp x_j, \text{ for all columns } x_j,\ j = 1, ..., k \text{ of } X,$$

because $y - X\hat{\beta}$ is orthogonal to every vector in $R(X)$.

So, $\quad x_j^*(y - X\hat{\beta}) = 0,\ \forall j = 1, ..., k.$

Stacking the rows x_j^*, we have

$$X^*_{k\times n}(y - X\hat{\beta}) = 0,$$

or $\quad X^*X\hat{\beta} = X^*y.$

This is a linear system called **the Normal Equations** and can be solved for $\hat{\beta}$.

Least Squares ... IV

If $(X^*X)_{k\times k}$ has full rank k (which happens iff X has full rank k), we also get $\hat{\beta} = (X^*X)^{-1}X^*y$. Note that in econometrics, the "data matrix" X is likely to have only real entries, so that the "star" or conjugate transpose above can be replaced with a "T" or transpose.

Before we close the section with a theorem on matrices like X^*X, we note that a $k \times k$ possibly complex matrix $A : C^k \to C^k$ (or $A : \Re^k \to \Re^k$ if A is real) is invertible iff $Rank(A) = k$ iff $Nullity(A) = 0$. For, to be invertible, A needs to be a bijection from $C^k \to C^k$.
$Rank(A) = k \Rightarrow R(A) = C^k$, so A is surjective. Moreover, if 2 vectors $v_1 \neq v_2$, then $Av_1 \neq Av_2$. Because $Av_1 = Av_2 \Rightarrow (v_1 - v_2) \in N(A)$. But $Rank(A) = k \Rightarrow N(A) = \{0\}$, so this can't be true. So, A is injective.
So, A is bijective.
We end by showing that X^*X has full rank k iff X does.

Least Squares ... V

Theorem

Let X be $n \times k$, $n >> k$. Then X^*X is invertible iff $Rank(X) = k$.

Proof.

X^*X is invertible iff $Nullity(X^*X) = 0$. $Rank(X) = k$ iff $Nullity(X) = 0$.
So we are done if we show $Nullity(X^*X) = 0$ iff $Nullity(X) = 0$.
We show the stronger result that $N(X^*X) = N(X)$. Indeed, let
$b \in N(X)$, *so* $Xb = 0$. So $X^*Xb = 0$. So $b \in N(X^*X)$.
Conversely, let $b \in N(X^*X)$. So, $X^*Xb = 0$. So, $(X^*Xb, b) = 0$ or
$b^*X^*Xb = 0$ or $(Xb, Xb) = 0$. So, $Xb = 0$ or $b \in N(X)$. □

Identity Property of Adjoint Matrix

Definitions

$(Ax, y) = (x, A^*y) \ \ \forall x, y \in \mathbb{C}^n$

Proof.

Now, to prove the main identity:
$(Ax, y) = \underbrace{y^*A}\, x = (A^*y)^*x = (x, Ay)$,
the first and the last equalities here follow from the definition of inner product in F^n, and the middle one follows from the fact that $(A^*y)^* = y^*(A^*)^* = y^*A$. We also have used here, $(AB)^* = B^*A^*$. □

Fact

Let us note that the adjoint operator is unique: if a matrix B satisfies $(Ax, y) = (x, By) \ \forall x, y$ *then* $B = A^*$*.*
Indeed, by the definition of A^* *we get* $(x, A^*y) = (x, By) \ \ \forall x$ *and therefore by [Corollary 1.5 Treil 118]* $A^*y = By$*. Since it is true for all* y*, the linear transformations, and therefore the matrices* A^* *and* B *coincide*

Relation between Range and Null Spaces

Of A and A^*

Consider an LT $A : V \to W$

Theorem

$RanA^* = (NullA)^\perp$.

Proof.

$NullA = \{z | Az = 0\}$. $Az = 0$ iff $z^* A^* = 0^*$ (taking conjugate transpose).
That is, iff $z^* a_i = 0^*$, or $a_i^* z = 0$, $\forall$ columns a_i of A^*.
That is, iff $z \perp a_i$ for all columns a_i of A^*, iff $z \perp RanA^*$.
So, $NullA = \{z | z \perp RanA^*\}$. (In the real matrix case, just think of $N(A)$ as all vectors z orthogonal to the rows of A and hence to the columns of A^T and hence to $R(A^T)$).
So, $NullA = (RanA^*)^\perp$. Taking orthogonal complements on both sides, $(NullA)^\perp = RanA^*$. □

As a corollary, interchanging A and A^*, we get $(NullA^*)^\perp = RanA$.

Relation between Range and Null of A contd...

Consider an LT $A : V \to W$

Theorem

$NullA^* = (RanA)^\perp$.

Proof.

Let $w \in W$. Then
$w \in null(A^*) \Leftrightarrow A^*w = 0$
Now, $(v, A^*w) = 0$ for all $v \in V$ which is orth. comp. to N(A*)
$\Leftrightarrow (Av, w) = 0$ for all $v \in V$ (by Indentity property of Adjoint)
$\Leftrightarrow w \in (RanA)^\perp$.
Thus $NullA^* = (RanA)^\perp$ □

Note: If we take the orthogonal complement of both sides of the previous result, we then get $RanA = (Null(A^*))^\perp$.

A as a Composition

w.r.t. Fundamental Subspaces

Definition

A LT $T : V \longrightarrow W$ is an **isomorphism** if it is invertible.

If such an isomorphism (a LT from V to W) exists, it can be shown that V and W have the same dimension (say n). Looking at things in terms of some bases, they are both practically like $\Re^n$. For example if $\{v_1, ..., v_n\}$ is a basis in V, a vector $x \in V$ is expressed as the unique LC $x = \sum x_i v_i$. So the scalars $(x_1, ..., x_n)$ express x w.r.t. this basis, and this is exactly like looking at a vector in $\Re^n$. Similarly for vectors in W. The point is, V and W are practically like the same space with different names.
Now turn to a matrix A. Say $A : V \longrightarrow W$. From the result on the fundamental subspaces, we can decompose the action $v \mapsto Av$ in two: First, v is carried to $P_{N(A)^\perp} v$ in $N(A)^\perp$ or $R(A^*)$. Then, the linear transformation A restricted to the domain $R(A^*)$, i.e. $A|_{R(A^*)}$, carries

A as a Composition

$P_{N(A)^\perp}v$ to Av in $R(A)$. In other words, A is a composition of 2 LTs. Indeed, $A|_{R(A^*)}$ is an isomorphism between $R(A^*)$ and $R(A)$, (and can be represented by an $r \times r$ submatrix of A, where $r = Rank(A) = Rank(A^*)$). We have the

Theorem

A is a composition of two LTs: $A = A|_{R(A^*)} P_{R(A^*)}$.

Proof.

Since $V = N(A)^\perp \oplus N(A) = R(A^*) \oplus N(A)$, for any $v \in V$ we have the unique orthogonal decomposition $v = v_1 + v_2$, where $v_1 = P_{R(A^*)}v \in R(A^*)$ and $v_2 \in N(A)$. So, $Av = A(v_1 + v_2) = Av_1$. And Av_1 is just equal to $AP_{R(A^*)}v$. So for all $v \in V$, we have $Av = AP_{R(A^*)}v$. □

We end by showing that $A|_{R(A^*)}$ is an isomorphism from $R(A^*)$ to $R(A)$.

Fundamental Subspaces of A

Theorem

$A|_{R(A^*)} : R(A^*) \to R(A)$ *is an isomorphism.*

Proof.

A is a LT. It remains to show that it (restricted to the domain above) is a bijection.
(i) Let $v, v' \in R(A^*)$ and $Av = Av'$. So $(v - v') \in N(A)$, and so $v = v' + (v - v')$ is the unique decomposition of v w.r.t. $N(A)^{\perp} \oplus N(A)$. However, since $v \in N(A)^{\perp}$, its (unique) decomposition on $N(A)^{\perp} \oplus N(A)$ must be $v = v + 0$. so, $v' = v$, and $A|_{R(A^*)}$ is injective.
(ii) Let $w \in R(A)$. So $w = Av$ for some $v \in V$. Since v has a unique decomposition $v = v_1 + v_2$ where $v_1 \in N(A)^{\perp} \equiv R(A^*)$ and $v_2 \in N(A)$, $Av = Av_1$, so w has a preimage $v_1 \in R(A^*)$. So, $A_{R(A^*)}$ is surjective as well. □