**Matrices, Mappings and Partial Inverses**
**by**
**A.B. Novikoff and D.N. Seppala-Holtzman**

***Dedicated to R.J. Bumcrot on the Occasion of his Retirement***

## Section 0: Introduction

This paper has been inspired by and is intended to illustrate three methodological principles. First, the duality aspect of linear algebra: all spaces and all maps between them occur in dual pairs. Furthermore, if a space is given a choice of basis, its dual space acquires a dual basis. Second, mapping diagram techniques: These are indispensable for following the composition of maps, which, unlike their matrix formulations, present themselves in a basis-free way. Diagrams are also notationally useful for exhibiting characteristics like the injectivity and surjectivity of maps. Third, Euclidean structures, usually given to a space, V, by assigning it an inner product, can be also specified by a suitable map, $E: V \longrightarrow V^*$, from V to its dual. Such maps, E, will occupy an important place in our mapping diagrams. To emphasize the distinction between Euclidean vector spaces and those without a specified Euclidean structure, we call the latter affine vector spaces. These preliminary matters will be discussed in sections 1, 3 and 4, respectively. Section 2 is an introduction to right inverses to maps.

Our main results are contained in sections 5 - 7. The primary result (see section 5) is: Given a map, T, from a Euclidean vector space, V, onto an affine space, W, there is an induced Euclidean structure on W. If W is already Euclidean, this generally results in W having two competing Euclidean structures. We exploit this induced Euclidean structure in order to construct a right inverse, R, of the given map, T. In section 5, we explore the coordinate form for this Euclidean inner product when the map T has been given in matrix form, i.e. when V and W have been given coordinate systems.

This right inverse, R, constructed in a basis-free way as a mapping, is shown to reduce (when coordinates have been introduced in both V and W) to a well-known matrix formula:

$$B^t (B\, B^t)^{-1}.$$

This formula assumes that B is a matrix representing a surjective (onto) map. It is clearly a right inverse to the matrix, B. Our purpose here is to give basis-free meaning to the mappings: $B^t$ and $(B\, B^t)^{-1}$and explain why the product $B\, B^t$ should be expected to appear.

To describe in full detail the relation between the basis-free construction of R and its above coordinate representation requires that we explore the distinction between the dual of a matrix (considered as a mapping) and its transpose. The precise relation between dual and transpose (see 5.8) seems not to be generally available in the literature.

This construction of the right inverse, R, of the given map, T, can be modified (see section 6) to cover the case where we abandon the assumption that T is surjective. In this case, however, we must require that the range space, W, have its own Euclidean structure since the one induced by T is only defined on a proper subspace of W. There is a known, geometrically interpretable right inverse, when V and W are both Euclidean, called the pseudo-inverse or the Moore-Penrose inverse. Our construction of R is the mapping version of this inverse. It uses both the given Euclidean structure on W and the induced one (see section 5). The Moore-Penrose theorem is merely a characterization of right inverses and is given entirely in terms of matrices. By contrast, we present a construction in terms of mappings. There is no natural matrix formula for this inverse, independent of bases.

Finally, in section 7, we sketch the application of the above construction of R to elementary circuit theory. In this case, the map, T, is familiar (to topologists) as the boundary operator (of a 1-complex). The matrix of T with respect to the standard basis is also familiar as the incidence matrix of the associated directed graph. That is, both the mapping and the matrix representation of it are familiar objects, but not necessarily to the same mathematical communities.

**Section 1: Duality**

We summarize briefly those facts about duality for n-dimensional real vector spaces that will be needed below. Readers should feel free to skim over this section.

The Dual Space:

Given a finite dimensional real vector space, V, consider the set of linear mappings $V \longrightarrow \mathbb{R}$ (sometimes called linear functionals). These form a new vector space, V*, called the dual space of V. Its elements will be denoted by v*. Any $v^* \in V^*$ is completely determined by its values on any basis, $\{e_1 \ldots e_n\}$ of V. The constraints of being a linear map allow the definition of v* to be extended to all of V in only one way. We call this "extending by linearity."

Pairings:

For any $v \in V$ and $v^* \in V^*$, we use the notation $< v^* , v >$ to indicate the evaluation of the functional, v*, on the vector, v. If V is $\mathbb{R}^n$ (column vectors) then V* is $\mathbb{R}_n$ (row vectors) and the pairing $< v^* , v >$ is given by matrix multiplication:

$$(1.0) \qquad (y_1\ y_2 \ldots y_n)\begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{pmatrix}$$

This is the prototype of all duality pairings.

Bases:

If $\{e_1 \,.....\, e_n\}$ is a basis for V, there corresponds a unique basis $\{e_1^* \,.....\, e_n^*\}$ for V* such that:

$$< e_i^* , e_j > = \delta_{ij} \text{, the Kronecker delta, for } i ,j = 1, 2, \dots , n.$$

Each $e_i^*$ is defined by these equations for i fixed and j varying from 1 to n. Extending by linearity yields a functional on all of V. The basis, $\{e_1^* \,.....\, e_n^*\}$, is said to be dual to $\{e_1 \,.....\, e_n\}$ and the two bases are called bi-orthogonal. An immediate consequence is:

$$\dim (V) = \dim (V^*)$$

Subspaces and Their Annihilators:

If $V_1$ is a (proper) subspace of V, then it determines a subspace, $V_1°$, of V* called the **annihilator** of $V_1$. This consists of all $v^* \in V^*$ which vanish identically on $V_1$. Furthermore:

$$\dim (V_1) + \dim (V_1°) = \dim (V) = \dim (V^*)$$

The dual space, $V_1^*$, of the subspace, $V_1$, is the quotient space, $V^*/ V_1°$, consisting of equivalence classes, [v*], where $v_1^*$ is equivalent to $v_2^*$ if they agree on $V_1$, that is $v_1^* - v_2^* \in V_1°$. Thus, subspaces and quotient spaces are dual constructs.

Double Dual and Identification:

The space, V*, has its own dual space, (V*)* = V**. Each element, $v \in V$, determines a linear functional on V*, namely a linear map, $v^{**}: V^* \longrightarrow \mathbb{R}$. It is given by:

(1.1) $$v^{**} (v^*) = < v^* , v >$$

Given this definition, it makes sense to denote v** by $< \bullet , v >$. Furthermore, this notation emphasizes the fact that V** and V are naturally isomorphic with the correspondence given by $v \leftrightarrow < \bullet , v >$. This permits an identification of V** with V, which is often made tacitly. Thus, when considering the pairing, $< v^{**} , v^* >$, which is defined on V** x V*, we will allow ourselves to write $< v , v^* >$. This, in turn, is equivalent to $< v^* , v >$, by virtue of (1.1). This results in a symmetry relation:

(1.2) $$< v , v^* > = < v^* , v >$$

where the left-hand expression denotes a pairing of V** and V* while the right-hand evaluation pairs an element from V* with one from V.

Because of this identification of $V^{**}$ with V, the roles of V and $V^*$ become interchangeable, each being the dual of the other. This results in yet another symmetry relation: not only is $V_1^\circ$ the annihilator of $V_1$, but $V_1$ is also the annihilator of $V_1^\circ$.

In section 7, we will exhibit a physical application of this material where the distinction between V and $V^*$ is emphasized because their elements are measured in different physical units. The edges and nodes of the underlying graph of an electric circuit form a basis of V and W, respectively, but the functions defined on them (which are at a higher level of abstraction) define their duals.

Maps:

Given a linear map

$$T: V \longrightarrow W$$

between vector spaces, there is an induced dual map,

$$T^*: W^* \longrightarrow V^*$$

determined by the identity:

(1.3) $$\langle T^*(w^*) , v \rangle = \langle w^* , T(v) \rangle$$

We shall call $T^*$ the dual of T, and not its adjoint, in order to avoid confusion with the notion of adjoint maps from a Euclidean space to itself. A prominent corollary of this definition of $T^*$ is that ker(T) and im($T^*$) are mutual annihilators. Indeed, by (1.3), if $T(v) = 0$, then $\langle T^*(w^*) , v \rangle = 0$ for all $w^* \in W^*$. So any $v \in \ker(T)$ is necessarily in the annihilator of im ($T^*$) and conversely. That is, we have:

**The Annihilator Relations**: $\ker(T) = \mathrm{im}(T^*)^\circ$ and $\ker(T^*) = \mathrm{im}(T)^\circ$

We will have occasion to make reference to this in section 7.

Injective and Surjective Maps: Their Construction as Dual Maps

If a map $T: V \longrightarrow W$ is not surjective ("onto"), then it is readily seen that its dual, $T^*$, is not injective ("one-to-one"). In this case, one can construct a modified version of T that is surjective:

(1.4) $$\{T\}: V \longrightarrow \mathrm{im}(T)$$

Here, we shrink the range space of T, W, to a proper subspace, im(T). Correspondingly, we can produce, via a method know as "quotienting out the kernel", a modified version of $T^*$ which is injective:

(1.5) $$[T^*] : W^* / \ker(T^*) \longrightarrow V^*$$

These two modifications are dual to one another, generalizing the previously noted dual character of subspaces and quotient spaces. We omit the details. (We note in passing that the "injectification" procedure given in (1.5) is a general method and is used in other contexts than our present one. That “shrinking the range” turns out to be its dual construction seems to be somewhat less well known.)

Self-Duality:

Finally, it is possible for a linear map $T : V \longrightarrow W$ to be identical to its own dual $T^*: W^* \longrightarrow V^*$. Clearly, a necessary condition for this is that W* to be identical to V and that V* be equal to W. Thus, this requires that T be a map from V to V*. We shall encounter, below, maps from V to V* (and their inverses from V* to V) which exhibit this property of being self-dual. (As above, we reserve the word, "self-adjoint", to refer to a related class of maps, called operators, defined from a Euclidean space to itself.) The condition that a map $T: V \longrightarrow V^*$ be self-dual is the identity:

$$(1.6) \qquad \langle T(v) , w \rangle = \langle v , T(w) \rangle \qquad \text{for all } v, w \in V.$$

This is an adaptation of (1.3) to the present special case.

We emphasize that this entire section is independent of any Euclidean structure that may happen to be defined on V or W or their duals. This is particularly true of our definitions of annihilator and self-duality.

**Section 2: Right Inverses of Surjective Maps**

Given a surjective map $T: V \longrightarrow W$, there will be an inverse map $T^{-1}:W \longrightarrow V$ if and only if ker(T) = 0. When this fails to be the case, there still exists a set inverse, the **preimage** of T defined as

$$T^{-1}(w) = \{ v \in V \mid T(v) = w \}$$

The assumed surjectivity of T assures that $T^{-1}(w)$ is never the empty set. It is well known that, in the surjective case, $T^{-1}(w)$ can be represented by $[v_0 + \ker(T)]$ where $v_0$ is any choice of element in $T^{-1}(w)$. If T is not assumed to be surjective, this is only true for w in im(T).

By a **right inverse** of T (sometimes called a pseudo-inverse or partial inverse) we mean a linear map

$$R: W \longrightarrow V$$

such that

$$T \circ R = id_W$$

This is equivalent to the statement that $R(w) \in T^{-1}(w)$, for all w in W, i.e. R "selects" a preimage of w from among the available candidates.

Another formulation of this notion is to say that im(R) is a subspace of V that bears a special relationship to ker(T), namely:

$$\text{im}(R) \cap \ker(T) = \{ 0 \}$$

so that T restricted to im(R) is non-singular. In words, im(R) is "complementary" or "transverse" to ker(T). Together they span V. Every element of V can be written uniquely as a sum of one element from ker(T) and one from im(R). An immediate consequence is that

$$\dim(\text{im}(R)) + \dim(\ker(T)) = \dim(V)$$

a numerical consequence of the transversality; in fact it is equivalent to it.

Conversely, given any subspace $V_1$ of V, complementary to ker(T), an associated right inverse, R, is readily described:

$$R(w) = V_1 \cap T^{-1}(w)$$

It follows from the above that:

(2.1) there are as many right inverses of T as there are complementary subspaces to ker(T) and they determine each other

and

(2.2) we can anticipate that if V is endowed with a Euclidean structure (i.e. a Euclidean inner product or any of its equivalents; see section 4), then there is automatically associated a **distinguished** right inverse. This is obtained by choosing the complementary subspace, $V_1$, of ker(T) to be its **orthogonal complement**, $\ker(T)^{\perp}$. (The converse is inexact: choosing a complement $V_1$ to ker(T) does not determine a Euclidean structure on V since many different inner products can have $V_1$ and ker(T) as orthocomplements.) This remark is further exploited in section 6.

It is worth remarking that despite our reference in (2.2) to orthocomplements in V, no reference is made, as yet, to any Euclidean structure in the range space, W. We shall show, however, that a particular Euclidean structure is always induced on W from one given on V. It will play a major role in the construction of a corresponding right inverse, R.

## Section 3: Diagram Language

It is convenient, in order to visualize the composition of linear maps, to introduce the language of "exact sequences." If

$$W \xleftarrow{\;T\;} V \xleftarrow{\;S\;} U$$

is a sequence of linear maps between vector spaces, we say that this sequence is **exact at V** if

$$\mathrm{im}(S) = \ker(T)$$

In other words, the sequence is exact at V if the two distinguished subspaces of V, im(S) and ker(T), happen to coincide. Of course, this implies that the composite $T \circ S$ is the zero map, but the condition of exactness is far more demanding. (Aside: for topologists, it is the failure of exactness that is really interesting. Not so for us.)

Two special (almost degenerate) cases of this language come to mind immediately:

(3.1) To say that $0 \xleftarrow{\;Z\;} W \xleftarrow{\;T\;} V$

is exact at W means that T is **surjective** or **onto**. (Here, Z is the unique map that maps the entire space W onto the trivial zero space. In future, we shall unclutter our diagrams by omitting its name.) Similarly,

(3.2) To say that $0 \longrightarrow W^* \xrightarrow{\;T^*\;} V^*$

is exact at W* means that T* is **injective** (or **one-to-one** or **non-singular**).

It is readily verified that either of the two statements (3.1) and (3.2) above imply the other, so that the dual of an injective map is surjective and vice versa. This is, indeed, a special case of the more general, but equally elementary fact, that the two assertions

$$W \xleftarrow{\;T\;} V \xleftarrow{\;S\;} U \quad \text{is exact}$$

and

$$W^* \xrightarrow{\;T^*\;} V^* \xrightarrow{\;S^*\;} U^* \quad \text{is exact}$$

imply each other: **Exactness is preserved by duality.**

(For some reason, this elementary theorem is seldom stated in linear algebra texts although every step in its proof usually is.)

Our main interest is in the pair of dual exact sequence:

$$0 \longrightarrow W^* \xrightarrow{T^*} V^* \qquad \text{(injective)}$$

$$0 \longleftarrow W \xleftarrow[T]{} V \qquad \text{(surjective)}$$

We remark, that because of the perfect symmetry between these two statements, we can if we choose, interchange which we consider to be the dual and which is the original and which of the two we place on top and which on bottom. Indeed, it does not matter if one starts with an injection or a surjection since the one induces the other.

We conclude this section with a classic case of a pair of dual sequences of the type (3.1) and (3.2), above. Here, the injective map is the inclusion map of a subspace $V_1$ into its superspace, V. The corresponding dual surjective map is the natural projection of the dual space V* onto a suitable quotient space. In more detail, we have:

(3.3) $$V \xleftarrow{\;incl\;} V_1 \longleftarrow 0$$ (injection of subspace via inclusion map)

This is the diagram associated to an injective map so its dual should assert a projection:

(3.4) $$0 \longleftarrow V_1^* \longleftarrow V^*$$

Recalling from section 1 that:

$$V^*/\, V_1^\circ = V_1^*$$

we may consider the dual, (3.4), of the injection, (3.3), to be

(3.5) $$0 \longleftarrow V^*/\, V_1^\circ \xleftarrow{\;proj\;} V^*$$

Here the map of V* onto $V^*/\, V_1^\circ$ is the natural projection of a space onto a quotient space.

The dual pair, (3.3) and (3.5), allows us to reiterate the observation that subspaces and quotient spaces are dual objects and to add that inclusions and projections are dual concepts.

**Section 4: Euclidean Structures**

A vector space, V, is said to be Euclidean if it is endowed with an inner product:

(4.0) $(\bullet , \bullet ): V \times V \longrightarrow \mathbb{R}$

with the three properties of being symmetric, bilinear and positive definite. This definition, entirely in terms of V, is equivalent to another one, emphasizing the symmetry between V and V*, which we shall make use of later.

We observe, first, that the expression $(\bullet , v)$ defines a linear functional on V and that every functional can be represented uniquely in this way by a suitable choice of $v \in V$. Thus, the inner product in V establishes a linear map:

(4.1) $E: V \longrightarrow V^*$

which satisfies

(4.2) $< E(v) , w > = ( v , w ).$

It is readily established that this map, E, is ( i ) non-singular and ( ii )self-dual (see 1.6) in the following sense:

(4.3) $< E(v) , w > = ( v , w ) = ( w , v ) = < E(w) , v > = < v , E(w) >$

for any pair of elements v and w in V. This last equality is in (4.3) the formal symmetry of the pairing $< \bullet , \bullet >$ discussed in (1.2), above. Finally, E has property ( iii ) that $< E(v) , v >$ is greater than 0 unless $v = 0$. In the standard example, $V = \mathbb{R}^n$, the reader will recognize E as the transpose from columns to rows and will use the pairing given in (1.0) to interpret (4.2) and (4.3).

We can, however, reverse this reasoning by starting with a map, E, as in (4.1) with properties ( i ) - ( iii ) and then defining an inner product as in (4.0) by means of (4.2). This inner product is readily seen to be symmetric, bilinear and positive definite. The result will be to make V into a Euclidean space.

We, therefore, call a map, E, with properties ( i ) - ( iii ) a <u>Euclidean structure map</u> defined on V. Clearly $E^{-1}$ will have the same three properties. It follows that $E^{-1}$ defines an inner product on V*. It is given by:

(4.4) $( v^* , w^* ) = < E^{-1}( v^* ) , w^* > = < v^* , E^{-1}( w^* ) >$

If we let $v^* = E(v)$ and $w^* = E(w)$ in (4.4), we see (thanks to the self-duality of E) that:

(4.5) $( v^* , w^* ) = < E^{-1}(E(v)) , E(w) > = < v , E(w) > = < E(v) , w > = ( v , w )$

Thus, E and $E^{-1}$ are both <u>isometries</u> between V and V*. In particular, E maps an orthonormal basis in V onto a dual (<u>bi-orthogonal</u>) basis in V* which is, itself orthonormal. We can summarize the situation by saying that Euclidean structures occur in dual pairs.

Since the map, E, is determined entirely by its values, $< E(e_i) , e_j >$, when $\{e_1 \ldots e_n\}$ is a basis for V, E can be determined by declaring any one basis of V to be orthonormal.

Determining an inner product on V by selecting a basis to be orthonormal is standard, whereas stipulating the corresponding structure maps, E and $E^{-1}$ is less so. However, in our diagrams below, it is the structure maps that appear (always drawn vertically) while the choice of basis, if mentioned at all, is relegated to the accompanying text.

Finally, the maps, E and $E^{-1}$, permit us to relate as isometric images two similar, previously mentioned constructs associated with a given subspace $V_1$ of a Euclidean space, V, namely its annihilator, $V_1°$, and its orthocomplement $V_1^{\perp}$. That is:

(4.6) $$E : V_1^{\perp} \longrightarrow V_1° \qquad \text{and} \qquad E^{-1} : V_1° \longrightarrow V_1^{\perp}$$

Thus, E and $E^{-1}$ permit us to relate an affine concept (i.e. annihilator) and a corresponding Euclidean one (orthocomplement). In the special case that $V_1 = \ker(T)$, this yields the orthogonality relation:

$$\ker(T)^{\perp} = E^{-1}(\mathrm{im}(T^*))$$

which, in turn, gives rise to the following direct decomposition of V:

$$V = \ker(T) \oplus E^{-1}(\mathrm{im}(T^*)).$$

This, together with a corresponding decomposition of the Euclidean space, W, is the basis of references [8] and [9].

**Section 5: Constructing A Right Inverse Using E: The Concept of an Induced Euclidean Structure**

Let us incorporate our Euclidean structure map into our pair of dual exact sequences:

(5.1) $$\begin{array}{ccccc} 0 & \longrightarrow & W^* & \xrightarrow{T^*} & V^* \\ & & & & \downarrow E^{-1} \\ 0 & \longleftarrow & W & \underset{T}{\longleftarrow} & V \end{array}$$

For as long as possible, we omit any mention of a Euclidean structure on W. We observe that the map from W* to W indicated in the diagram:

$$(T\ E^{-1}\ T^*) : \quad W^* \longrightarrow W$$

is non-singular. In fact, for any non-zero $w^* \in W^*$,

$$< w^* , T\ E^{-1}\ T^* w^* > \ = \ <T^* w^*, E^{-1}\ T^* w^* >$$

is positive, since $E^{-1}$, like E, is positive definite. Thus, $T\ E^{-1}\ T^* w^*$ is non-zero and so the map $T\ E^{-1}\ T^*$ is non-singular. Since dim(W) = dim(W*), this shows that $(T\ E^{-1}\ T^*)$ has a bone-fide inverse:

$$(T\ E^{-1}\ T^*)^{-1} : W \longrightarrow W^*$$

Further, it is self-dual since $(T\ E^{-1}\ T^*)$ is, itself, self-dual. In short,

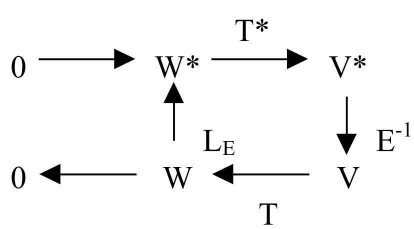
**Primary Theorem: $(T\ E^{-1}\ T^*)^{-1}$ is a Euclidean structure on W, induced by that on V.**

Or, more precisely, it is induced by the pair T and E. We shall call this map $L_E$ (because it is to the left of E in the diagram).

**Definition:** $\mathbf{L_E = (T\ E^{-1}\ T^*)^{-1}}$

Augmenting diagram (5.1) with this induced Euclidean structure map, we get the diagram:

(5.2)




A second, perhaps easier, way of constructing $L_E$ is to observe that W* inherits a positive definite inner product by virtue of being injected into V*. Therefore, by duality (section 1), so does W. The above $L_E$ is, in fact, the Euclidean structure map corresponding to this inherited inner product on W.

It follows (details omitted) that T* is an isometry onto its range and that T is an isometry onto W when restricted to $\ker(T)^{\perp}$. Both of these statements use the induced Euclidean structures on W* and W, respectively, to make them Euclidean spaces.

As is evident from the diagram, E and T have together created a map $L_E$ such that:

$$E^{-1}\ T^*\ L_E : W \longrightarrow V.$$

This map has an appropriate domain and range to be a candidate to be a right inverse for T. In diagrammatic terms, it is possible that the loop starting and ending at W is the identity map, $id_W$. If we take pains to write this conjecture out,

$$T ( E^{-1} \ T^* \ L_E ) = id_W$$

we see that it is equivalent to

$$(T \ E^{-1} \ T^*) L_E = id_W$$

which is true by the definition of $L_E$, above. Thus, we have arrived at our second major result:

**<u>Theorem 5.3:</u> The map $R = ( E^{-1} \ T^* \ L_E )$ is a right inverse of T.**

Furthermore, im(R), the complementary subspace to ker(T), turns out to be orthogonal to ker(T) with respect to the Euclidean structure of V. That is, $im(R) = ker(T)^{\perp}$, the orthocomplement of ker(T).

Since every element of V, and thus any element of $T^{-1}(w)$, is a (unique) sum of one element from ker(T) and one from $ker(T)^{\perp} = im(R)$, we know that $R(w) \in im(R)$ is, by the "Pythagorean Inequality," the "shortest" pre-image of w. (This inequality is the observation that the hypotenuse is longer than either leg of a right triangle.) This gives R a clear geometric interpretation. As such, we can write:

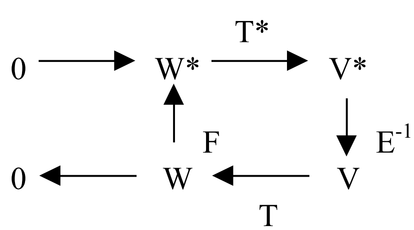
$$T^{-1}(w) = R(w) + ker(T)$$

where R(w) stands for the distinguished minimal length preimage of w.

<u>An Alternative Description of R:</u>

If we attempt to duplicate the above reasoning using an arbitrary Euclidean structure, F, on W, (a priori) unrelated to $L_E$ (the one induced by E), diagram (5.2) becomes:

(5.4)




An obvious candidate for R, the right inverse of T, is $(E^{-1} \ T^* \ F)$, as is seen from the diagram. (Here F is now playing the role that $L_E$ formerly occupied.) However, it can be shown that if $F \neq L_E$, then this is not a right inverse for T. Nevertheless, a right inverse can be constructed from it. If $(E^{-1}\, T^*\, F)$ is right-multiplied by a suitable invertible factor (which removes the factor F and replaces it by the factor $L_E$), then the product reduces to the previously constructed right inverse. The factor in question is:

$$F^{-1}\, L_E = (\, L_E^{-1}\ F\,)^{-1} = (T\ E^{-1}\, T^*\, F\,)^{-1}$$

We state as a theorem this alternative factorization of R:

**Theorem 5.5:** $(E^{-1}\, T^*\, F)\,(T\ E^{-1}\, T^*\, F\,)^{-1} = R$

Proof: The right-hand factor $(T\ E^{-1}\, T^*\, F\,)^{-1} = F^{-1}\,(T\ E^{-1}\, T^*)^{-1} = F^{-1}\, L_E$, from which the result is obvious.

■

The remarkable fact about the simple algebraic identity in Theorem 5.5 is that the product on the left is constructed using given Euclidean structures on both V and W (i.e. E and F, respectively). By contrast, the map, R, on the right, uses only the induced Euclidean structure $L_E$ on W, constructed from the given Euclidean structure, E, on V (and the fact that V "surjects" onto W).

Theorem 5.5 shows that the presence of F on the left is "illusory" (it is self-canceling) and that the resulting right inverse, R, depends exclusively on the Euclidean geometry of V, alone, as stated in theorem (5.3). Despite the fact that theorem (5.5) seems only an afterthought compared to the original, diagram-motivated formula, (5.3), it turns out that it is (5.5) that is the basis-free, or mapping, version of a well-known matrix formula. (See the matrix interpretation that follows).

Matrix Interpretation; A Low-Dimensional Example:

When T is realized as a matrix, B, the dual transformation, T*, also acquires a matrix formulation which we shall denote by B*. There is a slight but important distinction between this matrix, B*, and the transpose, $B^t$, of B.

To begin with, the realization of a linear transformation from V to W as a matrix requires that particular bases have been chosen in both V and W. Then, whether willed or not, corresponding Euclidean structures, E and F, will have been selected in each by declaring these bases to be orthonormal. B* and $B^t$, although not identical, relate to one another via E and F.

A simple example will illustrate this:

Let $B = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$ represent a surjection of $V = \mathbb{R}^3$ onto $W = \mathbb{R}^2$. This implies, in particular, that:

$e_1 = \begin{pmatrix}1\\0\\0\end{pmatrix}$ $e_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}$ and $e_3 = \begin{pmatrix}0\\0\\1\end{pmatrix}$ is a distinguished basis for $V = \mathbb{R}^3$ and

$f_1 = \begin{pmatrix}1\\0\end{pmatrix}$ and $f_2 = \begin{pmatrix}0\\1\end{pmatrix}$ is one for $W = \mathbb{R}^2$. The associated inner product in $V = \mathbb{R}^3$ is the one in which the basis, $\{ e_1, e_2, e_3 \}$ is orthonormal and the associated Euclidean structure map, E, sends $\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}$ to the linear functional defined by the row $(x_1 x_2 x_3)$. The pairing, $\langle v^*, v \rangle$ corresponds to matrix multiplication of a 1 x 3 matrix (a row vector) with a 3 x 1 matrix (a column vector) to produce a 1 x 1 matrix (a scalar). (See (1.0).) Similarly, $W = \mathbb{R}^2$ has a Euclidean structure, F, rendering $\{ f_1, f_2 \}$ orthonormal. Both E and F are defined by transposes (sending rows to columns) but defined on $\mathbb{R}^3$ and $\mathbb{R}^2$, respectively.

The dual space to V, V*, is, in this case, $\mathbb{R}_3$, the space of rows of length 3. Similarly, W* is $\mathbb{R}_2$, the space of rows of length 2. We incorporate this notation into the following (temporarily incomplete) diagram:

(5.6)

$$\begin{array}{ccccc}
0 & \longrightarrow & \mathbb{R}_2 & & \mathbb{R}_3 \\
 & & \uparrow F & & \downarrow E^{-1} \\
0 & \longleftarrow & \mathbb{R}^2 & \longleftarrow & \mathbb{R}^3 \\
 & & & B = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} &
\end{array}$$

Here, $E^{-1}$ is, again, the transpose, but now defined on rows in $\mathbb{R}_3$. Denoting E, $E^{-1}$ and F all by the same term, "transpose", is identifying so many distinct entities, despite their different domains and ranges, that, at least in this context, we shall no longer do it.

What is the dual (as defined in section 1) to the map represented by the matrix B, operating on the left of columns from $\mathbb{R}^3$? It is readily seen [3] to be the same matrix B acting, this time, on the right on rows from $\mathbb{R}_2$ (to produce rows in $\mathbb{R}_3$). Let us denote this matrix-defined map, B, but with its action now on the right, by B*. This matrix now completes (5.6) and yields the matrix equivalent to diagram (5.4):

(5.7)
$$\begin{array}{ccccc} & & & B^* & \\ 0 & \longrightarrow & \mathbb{R}_2 & \longrightarrow & \mathbb{R}_3 \\ & & \uparrow F & & \downarrow E^{-1} \\ 0 & \longleftarrow & \mathbb{R}^2 & \longleftarrow & \mathbb{R}^3 \end{array}$$

$$B = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$$

<u>The Transpose of a Matrix as a Composite of Three Maps:</u>

Let us now consider the 3 x 2 matrix transpose, $B^t$. This acts on the left of elements of $\mathbb{R}^2$ to produce elements of $\mathbb{R}^3$. Where does the transpose, $B^t$, fit into diagram (5.7)? The answer is that $B^t$ can now be identified as the only mapping present in the diagram that proceeds from $\mathbb{R}^2$ to $\mathbb{R}^3$. Explicitly:

(5.8)
$$B^t = E^{-1} B^* F$$

In words, the familiar $B^t$ is not the same as the dual, B*. Rather, it is "equivalent" to B*, differing from it only by being left and right multiplied by invertible maps as asserted in (5.8).

Put differently, when T is "identified" with a matrix B, bases are distinguished in V and W and Euclidean structures, E and F, automatically arise. Indeed, the presence of these Euclidean structures is all that is needed for the validity of (5.8). Neither injectivity nor surjectivity need be assumed. The basis-free T* is identified with the above B*, but not exactly with $B^t$. The familiar matrix formula:

(5.9)
$$B^t (B B^t)^{-1}$$

for a right inverse to a matrix, B, assumed surjective, is exactly the factorization of R asserted in theorem (5.5):

$$(E^{-1} T^* F)(T E^{-1} T^* F)^{-1} = B^t (B B^t)^{-1}$$

The self-canceling factor, F, in this product is exploited to make the “identification” of $B^t$ with B*, or more exactly, to state the formula (5.8) for $B^t$. Thus, (5.5) is the basis-free formulation of this matrix product, (5.9).

To sum up, the right inverse, R, to T that we have constructed in this section has been presented in various guises. These differ in appearance because different assumptions were made about the structures present in each context. In particular, we have seen:

$R = (E^{-1}\, T^*)\ (T\ E^{-1}\ T^*)^{-1}$ by (5.3) and the definition of $L_E$
$R = (E^{-1}\, T^*\, F)\ (T\ E^{-1}\ T^*\, F)^{-1}$ by (5.5) and
$R = B^t (B\, B^t)^{-1}$ by (5.9).

In the first case, W was endowed with the Euclidean structure, $L_E$, induced by E. In the second, W was given some arbitrary Euclidean structure, F. In the third case, V and W were $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, and the Euclidean structure maps, E and F, were the associated transpose maps (see section 4).

## Section 6: A Right Inverse in the Non-Surjective Case

We have, thus far, assumed that the given mapping, T, whose right inverse we seek, is surjective. We recall from section 1, (see equation (1.4)) that if $T: V \longrightarrow W_1$ is not surjective, it is nonetheless, intimately associated with a mapping which is:

$$\{T\}: V \longrightarrow W$$

where W is im(T), the image of the original map, T. We note that, apart from the shrinkage of the range space, from $W_1$ to W, the maps, T and {T}, are indistinguishable.

If the map, T, is represented by a matrix, B, and this matrix is row reduced, the resulting matrix will have a number of all-zero rows. Omitting these rows yields a matrix representation of {T}.

The map, {T} is surjective from its definition and has the same domain, V, as T. If V has a Euclidean structure map, E, and a corresponding inner product (as in section 5), then all of the constructions of that section, in particular those of $L_E$ and R, can be carried out with {T} replacing T and W replacing $W_1$, throughout. The assertion of theorem (5.3), that R is a right inverse of T, now becomes, in this setting:

$$\{T\} \circ R = id_W$$

Returning to consideration of the range, $W_1$, of the given T, we assume that we have been given a complementary subspace, U, to W = im(T). Thus, we have the direct decomposition:

$$W_1 = W + U$$

To this decomposition, one can associate the (affine) projection, 👌, of $W_1$ onto W along U:

$$\pi: W_1 \longrightarrow W$$

We can compose the projection π with the right inverse R of {T} to get a map:

$$R \circ \pi: W_1 \longrightarrow V$$

which, in view of its domain and range, is a candidate for being a partial inverse of T. In fact, $R\circ\pi$, although not a right inverse of the (original, non-surjective) T, does satisfy:

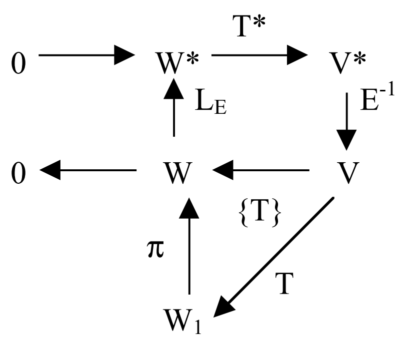
(6.1) $$\{T\}\circ(R\circ\pi) = (\{T\}\circ R)\circ\pi = id_W\circ\pi = \pi$$

This follows because R was created (as in section 5) as the right inverse to {T}. Now, by the construction of {T} and (6.1), we have:

(6.2) $$\pi = \{T\}\circ(R\circ\pi) = T\circ(R\circ\pi)$$

Note that $T\circ(R\circ\pi)$ is not the identity on $W_1$. It is an "affine projection" of $W_1$ onto W along U. Incorporating this into our previous diagram (5.2) and letting $L_E$ be restricted in its domain to W we get the slightly larger diagram that modifies (5.2) and allows us to consider non-surjective T as well as its surjectification, {T}:

(6.3)




It is at this point that a Euclidean structure, F, on $W_1$, whose introduction we have postponed as long as possible, finally plays an indispensable role. Suppose that T is a general (non-surjective) map between Euclidean spaces, V and $W_1$, with structure maps, E and F, respectively. Then, using F, we can define the subspace:

$$U = (\mathrm{im}(T))^{\perp} = (W)^{\perp}$$

to be the orthocomplement of W. The reader will recall that in (2.2) we similarly used a Euclidean structure to construct an orthocomplement. The resulting projection, $\pi_F$ , onto W along U is then an **orthogonal** projection defined using F. The map, $R\circ\pi_F$, in this construction is then called the Moore-Penrose inverse, or pseudo-inverse, of the original map, T. The associated diagram is the above (6.3) but with the orthogonal projection, $\pi_F$, in the role of $\pi$.

In purely geometric (i.e. Euclidean) terms, an element w of $W_1$ is orthogonally projected onto im(T) after which the (surjective) right inverse, R of section 5, maps the result onto the pre-image in V of minimal length.

There are, in general, many Moore-Penrose inverses to a given map

T: V $\longrightarrow$ $W_1$. This is because there are many complementary subspaces U to W = im(T) in $W_1$ and there are many subspaces of V which are complementary to ker(T). In addition, there are many Euclidean structures on V and $W_1$ in which these complementary subspaces are orthocomplements.

The merit of the original Moore-Penrose theorem was to characterize a right inverse in terms of a given matrix description of T, which, necessarily, has already determined bases and Euclidean structures on both the domain, V, and the range, $W_1$. By contrast, this section deals with a construction, not a characterization, and it shows how the diagram (5.2) generated in the surjective case can be augmented to become the diagram (6.3) for the general case.

## Section 7: An Application --- Electric Circuits

Thus far, we have emphasized the symmetry that characterizes the relationship between a space and its dual. Indeed, we have pointed out that either V or V* could have been looked upon as the "original space" and the other as the "induced one." In this section, we examine a specific, real application of the material presented above and observe that the two spaces, although dual, have very distinct roles. It was this case that was the starting point for our investigations.

An electric circuit (purely resistive for simplicity) can be represented by a connected, directed, weighted graph. The weight that is assigned to each edge signifies the resistance on that edge (or branch).

### The Spaces V and V*:

Given an electric circuit, one can associate to it a vector space, V, which consists of all formal linear combinations of the directed edges of its underlying graph. This space, V is defined independently of the weights on its edges. The individual edges, collectively, form a distinguished basis for V. The elements of V are called currents. V is called the current space or, sometimes, the branch space.

The dual space, V*, consists of real-valued, linear functions defined on V. In practice, they are often defined on the edges and then extended to all of V by linearity. The elements of V* are called voltages. Any externally imposed voltages (such as batteries) correspond to elements of V*. Similarly, externally imposed current sources correspond to elements of V. V* is called the voltage space.

In analogy with V, the branch space, we introduce $W_1$, the node space. This is the vector space consisting of all formal linear combinations of nodes, which are vertices of the underlying graph. (The notation, $W_1$, is chosen in conformity with the preceding section.)

### The Boundary Map and Its Dual:

If $A_i$ and $B_i$ are nodes of the graph and $e_i$ is the directed edge from $A_i$ to $B_i$, we define the function $\partial$ by

$$\partial(e_i) = B_i - A_i$$

This, in turn, induces a map

$$\partial_1: V \longrightarrow W_1$$

by extending linearly to all of V. This map, $\partial_1$, is called the boundary map. Its matrix is the incidence matrix of the graph.

The corresponding dual space, $W_1^*$, consists of linear functions defined on the nodes and then extended linearly to all of $W_1$. These functions are called "potentials" which we denote by **Φ**.

While $\partial_1$ is not surjective, in general, we can induce a surjective map, as in the previous section:

$$\partial: V \longrightarrow W$$

by shrinking the range space of $\partial_1$ to the subspace, W, which is the image of $\partial_1$. In the notation of the previous sections, $\partial = \{\partial_1\}$. (If the underlying graph is connected, W is of co-dimension 1 in $W_1$. This is a well-known result from graph theory.)

The dual, W*, of this subspace turns out to be the potentials of $W_1$ but now modulo an arbitrary constant, much like anti-derivatives, which they resemble in other ways. The potentials themselves fall into equivalence classes. We denote by [**Φ**] the equivalence class containing **Φ**. Some authors choose a specific representative of each class in the following way. They select a node, A, and choose that member of [**Φ**] which vanishes at A. This is called grounding the node, A.

We note that V and V*, the spaces of currents and voltages, respectively, have different physical units. As a result, any Euclidean structure map that we subsequently introduce:

$$E: V \longrightarrow V^*$$

can be expected to have a physical interpretation and, as such, be determined by information not solely contained in the incidence relations of the graph.

Returning to the map $\partial$, we consider its dual:

$$\partial^*: W^* \longrightarrow V^*$$

$\partial^*$ assigns to every equivalence class, $[\Phi]$, a linear functional, $\partial^*[\Phi]$, determined by its values on each of the edges of the graph. These are given by:

$$\partial^*[\Phi]\,(e_i) \;=\; \Phi(A_i) \;-\; \Phi(B_i)$$

(Classically, $\Phi$ is defined so that a factor of -1 is incorporated.) where $\partial(e_i) = B_i - A_i$. Thus, $\partial^*$ is well-defined on classes, $[\Phi]$, and assigns "voltage drops" to potential classes.

The subspace, $\mathrm{im}(\partial^*)$ of $V^*$, consists of those voltage assignments to edges that are "derivable from a potential." The condition:

$$v^* \in \mathrm{im}(\partial^*) \subset V^*$$

corresponds to saying that $v^* \in V^*$ satisfies Kirchhoff's Voltage Law: *There exists a function, $\Phi$, such that the voltage across each edge is given by the difference of $\Phi$ evaluated at the endpoints of that edge.* ($\Phi$, itself, is only determined up to an arbitrary constant.)

Analogously, the currents $v \in V$ that satisfy:

$$v \in \ker(\partial) \subset V$$

are said to satisfy Kirchhoff's Current Law: *The net flux of current at every node is zero; charge is conserved.*

By the annihilator relations developed in section 1, we must have that the two subspaces $\mathrm{im}(\partial^*)$ and $\ker(\partial)$ are mutual annihilators. In the theory of circuits, this is known as Tellegen's Theorem.

The Euclidean Structure:

Finally, a Euclidean structure map, $E: V \longrightarrow V^*$ is now defined using the assigned resistances across each edge. Let $r_i$ (>0) be the resistance on $e_i$. We define $E(e_i) \in V^*$ to be the function on the edges of V that satisfy:

$$< E(e_i)\,,\,(e_j) > \;=\; \delta_{ij}\, r_i$$

and then extend to all of V by linearity. Thus, $E(e_i)$ evaluated on some edge, $e_j$, is either zero, if $i \neq j$, or the resistance across edge i, if $i = j$. The case of equal indices is Ohm's Law for the edges, $e_1 \ldots e_n$. The inner product, $(e_i\,,\,e_j) \;=\; < E(e_i)\,,\,(e_j) >$ is the power dissipated by a unit current in branch $e_i$ as a result of a unit resistance in branch $e_j$. Tellegen's Theorem asserts that no power is dissipated if currents obeying Kirchhoff's Current Law passes through branches across which the voltages obey Kirchhoff's Voltage Law.

As foreseen above, the physics of the circuit (i.e. the resistances) is incorporated in the Euclidean structure maps, E (and the induced $L_E$), while $\partial$ and $\partial^*$ depend only upon the incidence relations in the graph underlying the circuit. Letting $\partial$ and $\partial^*$ place the roles of T and T*, respectively, in diagram (5.2), we see that the vertical maps, there, reflect the physics of the circuit while the horizontal maps depend only upon the underlying geometry.

The Fundamental Problem of Circuit Theory is to assume assigned current generators, $v \in V$ and voltage generators (batteries), $v^* \in V^*$ and to find corresponding elements $k \in V$ and $k^* \in V^*$ such that

$$v^* - E(v) = k^* - E(k) \tag{7.1}$$

The apparatus of V and V*, and the associated maps, $\partial$ and E, developed above, is the natural setting for the proof of the existence and uniqueness of k and k* which satisfy (7.1). The optimal setting involves embedding (5.2) in an even larger diagram. However, this diagram exploits the same three principles, duality, diagram chasing and Euclidean structure maps already introduced here. Within this larger diagram, the historical methods of Maxwell, which prove existence and uniqueness and provide a constructive algorithm for the solution of (7.1), become transparent and motivated. The observation that this diagram setting (essentially due to Hermann Weyl) uses the same mathematics as the Moore-Penrose inverse is our own. We resist the impulse to explore these matters further and refer the reader to reference [1].

We hope that this paper can serve as a bridge between the two main constituencies of linear algebra users: matrix lovers and diagram chasers.

**References**


1. Bamberg, P. and Sternberg, S., *A Course in Mathematics for Students of Physics*, Vol. 1 & 2, Cambridge University Press 1990
2. Ben-Israel, A. and Grenville, T.N.E., *Generalized Inverses - Theory and Applications*, John Wiley and Sons, 1974
3. Halmos, P., *Finite-Dimensional Vector Spaces* (2nd Ed.), Springer Verlag, 1986
4. Lax, P., *Linear Algebra*, John Wiley and Sons, 1996
5. Leon, S.J., *Linear Algebra with Applications*, Macmillan Pub. Co., 1980
6. Odell, P. L. and Piziak, R., Full Rank Factorization of Matrices, *Mathematics Magazine*, Vol. 72, No. 3, 1999, 193-201
7. Strang, G., *Linear Algebra and Its Applications* (3rd Ed.), Harcourt Brace Jovanovich, 1988
8. Strang, G., Graphs, Matrices and Subspaces, *College Math Journal*, Vol. 24, No. 1, 1993, 20-28
9. Thoo, J.B., A Picture is Worth a Thousand Words, *College Math Journal*, Vol. 29, No. 5, 1998, 408-411



A.B. Novikoff: Courant Institute, New York University novikoff@cims.nyu.edu
D.N. Seppala-Holtzman: St. Joseph's College holtzman@sjcny.edu