

- Background Linear Algebra -

Eugene M. Cliff

August 24, 1999

1 Introduction

These notes are intended to supplement the treatment in the Gill-Murray & Wright text [1]. While the discussion in the text is mostly adequate for our needs, it does present some unfortunate limitations in its outlook. We shall present some material to broaden the approach. Our presentation is heavily influenced by Luenberger [3]. There are a number of excellent texts for most of this material (*e.g.* [2]). Numerical aspects of linear algebra are in the recent book by Trefethen and Bau [4].

2 Vector Spaces

A vector space is a set of objects (the vectors), the real (or complex) field of scalars and two operations connecting them. The operations are: vector addition and multiplication by scalars. Associated with any two vectors is a (unique) third vector - their *sum*. Also, given a vector and a scalar there is a unique vector - the *product*. The operations must enjoy certain properties (vector addition is commutative and associative). These are generally stated in the form of certain axioms that the vector operations must satisfy (see [3], pp 11-12). Among these axioms is the requirement that the set contain a unique vector $\mathbf{0}$, the additive identity element (meaning that $\mathbf{v} + \mathbf{0} = \mathbf{v}$, for all elements).

The usual ‘vectors’ in \mathbf{R}^n are the most common example. Addition means add component-wise. Somewhat more abstractly, we also mention the space of continuous functions on the interval $[0, 1]$. Vector-addition and scalar multiplication are defined in the obvious way. To be rigorous one should

prove the ‘closure’ property, *e.g.* that the sum of two continuous functions is a continuous function.

2.1 Linear Combinations

Given a finite set of vectors, and a corresponding set of scalars we can form a ‘new’ vector by the operation

$$\mathbf{v}_{\text{new}} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_p \mathbf{v}_p$$

In this case we say that \mathbf{v}_{new} is a linear combination of the underlying vectors.

2.2 Linear Dependence

Strictly said, linear dependence is best thought of as a property of certain finite-sets. That is, given a finite set of vectors, say $\mathbf{S} \equiv \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$ we say that \mathbf{S} is a linearly dependent set iff there is a set of scalars (at least one such set), not all zero, such that the corresponding linear combination is the zero vector (*i.e.* $\sum_{i=1}^p \alpha_i \mathbf{v}_i = \mathbf{0}$). A set of vectors that is not linearly dependent is said to be linearly independent. Note that if \mathbf{S} is a linearly independent set and if $\sum_{i=1}^p \alpha_i \mathbf{v}_i = \mathbf{0}$, then necessarily all of the scalars (α_i) must be zero.

With this definition, any set that includes the zero-vector is linearly dependent. For another example of a linearly dependent set, consider the vector space \mathbf{R}^2 and the set of vectors given by

$$\mathbf{S} \equiv \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$$

In this case the scalars ($\alpha_1 = 1$, $\alpha_2 = 1$, $\alpha_3 = -1$) demonstrate the dependence.

2.3 Subspaces

It happens that certain subsets of vectors have the ‘closure’ property, meaning that addition or scalar multiplication of anything in this set produces another element of the set. Such sets are called subspaces - they are vector spaces in their own right. It’s clear that any subspace must contain the zero vector. We agree that the set consisting of only the zero vector is a subspace and

the whole space is also a subspace. We reserve the term *proper subspace* to mean a subspace that is not the whole space and not the ‘zero’ subspace. In the case of \mathbf{R}^2 (the plane), the proper subspaces can be interpreted as lines through the origin. In \mathbf{R}^3 the proper subspaces are all the lines through the origin and all the planes through the origin.

2.4 Addition of Subspaces

Since we can add vectors, it’s natural to define addition of subspaces. Given two subspaces \mathbf{U} and \mathbf{V} (each a subspace of the same underlying vector space), then by their sum we mean the set:

$$\mathbf{U} + \mathbf{V} \equiv \{\mathbf{x} \in \mathbf{X} \mid \mathbf{x} = \mathbf{u} + \mathbf{v}, \text{ for some } \mathbf{u} \in \mathbf{U} \text{ and some } \mathbf{v} \in \mathbf{V}\}$$

One can prove that this set is, in fact, a subspace.

In the case of \mathbf{R}^2 , if we take any two *distinct* proper subspaces, their sum will be the whole space. In the case of \mathbf{R}^3 , if we take any two distinct planes their sum will be the whole space. However, there is a difference in these two examples. In the former case the representation $\mathbf{x} = \mathbf{u} + \mathbf{v}$ is unique; that is, given any $\mathbf{x} \in \mathbf{R}^2$ there is a unique pair $\mathbf{u} \in \mathbf{U}$ and $\mathbf{v} \in \mathbf{V}$. In the latter case, there is some ‘overlap’ and the representation is not (necessarily) unique. It is worthwhile to distinguish these cases; when the representation is unique we speak of the *direct* sum and write $\mathbf{U} \oplus \mathbf{V}$.

2.5 Span of a Set

Given a finite set of vectors we can consider the set we generate by taking all possible linear combinations of vectors from the set. This is called the *span* of the set. It requires proof, but it should be ‘clear’ that the span of a set is a subspace. This is a pretty common way to generate subspaces.

2.6 Basis and Dimension

The idea of generating subspaces by the spanning procedure can be sharpened. If the beginning set is linearly dependent then, it’s clear that at least one member can be ‘generated’ as a linear combination of the others. Thus, if we throw out this member then the span of the reduced set will be the same as the span of the original set. We can continue this process of throwing out

members until we get down to a linearly independent set. If we throw out any more vectors the span of the thus-reduced set will not be the same as the span of the original. A spanning set that is linearly independent is said to be a *basis* for the subspace.

It turns out that the number of elements in a basis is a property of the subspace - this is the *dimension* of the subspace. Again, it requires proof to show that if we start with two sets with the same span, this process of throwing out results in linearly independent sets with the same number of elements.

Our description characterizes a basis as a minimal spanning set. If we throw out any more elements then the span of the resulting set is decreased. It's sometimes useful to think of starting with a small set and adding vectors - to produce the desired spanning set. In this view if we add a vector to a basis and keep the same spanning set, it's clear that the new set is not linearly independent. Hence one can also think of a basis as the largest linearly independent set (with a given span).

Any basis has a naturally associated subspace; namely the space spanned by the basis. While the basis is not unique (a given subspace has many bases) the number of elements in any such basis is always the same. As noted above, this number is the dimension of the subspace. When adding subspaces we have $\dim[(\mathbf{U} + \mathbf{V})] \leq \dim[\mathbf{U}] + \dim[\mathbf{V}]$, with equality iff the sum is direct.

2.7 A Vector and Its Representation

Once we have a basis then any vector can be uniquely described by providing the scalar components in its representation. That is, if

$$\mathbf{v}_{\text{sample}} = \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 + \cdots + \alpha_n \mathbf{v}_n$$

then the n-tuple of scalars $(\alpha_1, \alpha_2, \dots, \alpha_n)$ represent the vector $\mathbf{v}_{\text{sample}}$ in terms of the given basis.

Perhaps the canonical example of a n -dimensional space is \mathbf{R}^n , wherein the vectors are these n-tuples of scalars. In other examples the basis is fixed so that we can blur this distinction and think of the collection of scalars as the vector. In most cases this causes no harm, but we should be aware of it. To make this more concrete, consider the space of continuous functions on the unit interval, and the subspace spanned by the set of vectors $\mathbf{S} = \{\cos(\cdot), \sin(\cdot)\}$.

Since the two vectors are linearly independent, the set \mathbf{S} is a basis for this two-dimensional subspace. Now however, there should be a clear distinction between the pair (α, β) and the function $f(t) = \alpha \cos(t) + \beta \sin(t)$. In fact, we derive a lot of benefit from the fact that one can *blur* the distinction between $f(\cdot)$ and the pair (α, β) .

3 Transformations and Functionals

One of the things we do with spaces (or their subspaces) is to define certain related maps or functions. Suppose \mathbf{X} and \mathbf{Y} are two vector spaces, then a rule that associates a unique element of \mathbf{Y} to every element of \mathbf{X} is a *transformation* or a *map*. In some settings this idea must be generalized so that the rule is defined for only some proper subset $\mathbf{D} \subset \mathbf{X}$. We write this as

$$T : \mathbf{D} \subset \mathbf{X} \mapsto \mathbf{Y}$$

A special case of particular interest occurs when the image space \mathbf{Y} is the scalar field; such a transformation is called a *functional*.

3.1 Linear Transformations

Linearity of maps is a natural idea in the vector space setting; it means that applying the transformation to a linear combination of vectors, produces the same result as applying the transformation to each component vector and then forming the linear combination. In symbols, we have

$$T\left(\sum_{i=1}^p \alpha_i \mathbf{u}_i\right) = \sum_{i=1}^p \alpha_i T(\mathbf{u}_i).$$

In the case when \mathbf{Y} is the (real or complex) field we speak of a *linear functional*.

3.2 Representation of a Linear Transformation

Suppose we are in the situation $T : \mathbf{X} \mapsto \mathbf{Y}$, that $\dim[\mathbf{X}] = n$, $\dim[\mathbf{Y}] = m$ and that we have bases, say \mathbf{U} , and \mathbf{V} , for \mathbf{X} and \mathbf{Y} , respectively. Then for any $\mathbf{x} \in \mathbf{X}$, we have $\mathbf{x} = \sum \alpha_j \mathbf{u}_j$, so that

$$T(\mathbf{x}) = T\left(\sum_j \alpha_j \mathbf{u}_j\right) = \sum_j \alpha_j T(\mathbf{u}_j).$$

Now for each $j \in \{1, 2, \dots, n\}$, $T(\mathbf{u}_j) \in \mathbf{Y}$ so that any such vector has a unique representation, say $T(\mathbf{u}_j) = \sum_{i=1, m} t_{ij} \mathbf{v}_i$. Indeed, we can write this in matrix terms as

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix} = \begin{bmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ t_{21} & t_{22} & \dots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ t_{m1} & t_{m2} & \dots & t_{mn} \end{bmatrix} \cdot \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix}.$$

The m -tuple of scalars $(\beta_1, \dots, \beta_m)$ represent the image of $T(\mathbf{x})$ in terms of the basis \mathbf{V} . In particular, the j -th column of the matrix represents the image of the vector $\mathbf{u}_j \in \mathbf{U} \subset \mathbf{X}$ in terms of the basis $\mathbf{V} \subset \mathbf{Y}$.

As an example, consider the earlier example of the space of continuous functions and the subspace \mathbf{X} spanned by the set of vectors $\mathbf{S} = \{\cos(\cdot), \sin(\cdot)\}$. In this setting consider the linear map defined by $T(x)(t) = \int_0^t x(s) ds$; from any function we get a new function - the indefinite integral. As a result of properties of the selected functions, it should be clear that T maps vectors in \mathbf{X} back into vectors in \mathbf{X} . In fact, in terms of the given basis the linear operator T is represented by the matrix

$$M = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

4 Norms and Inner Products

The vector space axioms provide an algebraic structure that is useful for many applications. For many purposes it is also necessary to have some way to say that two vectors are ‘close’. Our starting point for this is the idea of a *norm*. Fundamentally, a norm is a mapping (real-functional) that assigns to each vector a real number - the length of the vector. The most common notion is, of course, the Euclidean norm computed as the square-root of the sum of the squares of the components. Note that in this case the vector space is the set of such n -tuples \mathbf{R}^n . For the moment we shall call this a Euclidean space and denote it by \mathbf{E}^n .

Of course, not every functional will capture the properties needed to make a norm useful. In fact, the list of required axioms is short:

1. $\|\mathbf{v}\| \geq 0$ for all \mathbf{v} , and $\|\mathbf{v}\| = 0$ if, and only if $\mathbf{v} = \mathbf{0}$.

2. $\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$
3. $\|\alpha \mathbf{v}\| = |\alpha| \|\mathbf{v}\|$

A vector space and an associated norm are called a *normed space*. A commonly cited non-Euclidean example is $C[0, 1]$ - the space of real-valued continuous functions on the interval $[0, 1]$. The vector space structure has been noted above and the norm in this case is given by $\|\mathbf{v}\| \equiv \max_{0 \leq t \leq 1} |v(t)|$. The notation $C[0, 1]$ means the vector space with this norm. Note that we can define a different norm on the same vector space (*e.g.* $\|\mathbf{v}\| \equiv \int_0^1 |v(t)| dt$). The normed space with this integral norm is different from $C[0, 1]$.

4.1 Inner Products

It turns out that there is a way to generalize the Euclidean case that will produce a special class of norms and normed spaces. An *inner-product* is a complex (possibly real) valued function that assigns to any *pair* of vectors a scalar value - their inner (or dot) product. We shall write this as: $\langle \mathbf{u}, \mathbf{v} \rangle$.

Again, we require certain axioms to end up with a useful concept

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$ (overbar means complex conjugate)
2. $\langle (\mathbf{u}_1 + \mathbf{u}_2), \mathbf{v} \rangle = \langle \mathbf{u}_1, \mathbf{v} \rangle + \langle \mathbf{u}_2, \mathbf{v} \rangle$
3. $\langle \alpha \mathbf{u}, \mathbf{v} \rangle = \alpha \langle \mathbf{u}, \mathbf{v} \rangle$
4. $\langle \mathbf{v}, \mathbf{v} \rangle \geq 0$ and $\langle \mathbf{v}, \mathbf{v} \rangle = 0$ iff $\mathbf{v} = \mathbf{0}$.

With these properties it turns out that $\|\mathbf{v}\| \equiv \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$ is a norm. The vector space \mathbf{X} , together with the inner-product define an *inner-product space* (sometimes called a *pre-Hilbert* space). Here again, the most common example is the Euclidean case:

$$\langle (\alpha_1, \alpha_2, \dots, \alpha_n), (\beta_1, \beta_2, \dots, \beta_n) \rangle = \sum_{i=1, n} \alpha_i \beta_i = \alpha^T \beta.$$

Note we have mixed in some matrix notation here. A common non-Euclidean example is the space of continuous functions along with the inner-product $\langle \mathbf{u}, \mathbf{v} \rangle \equiv \int_0^1 u(t) v(t) dt$.

4.2 Orthogonality

Just as in the Euclidean case we say that two vectors are *orthogonal* if the inner-product is zero. In this case one can easily show the Pythagorean Theorem holds: $\|(\mathbf{u} + \mathbf{v})\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

The idea of orthogonality of two vectors can be extended by saying that a vector is orthogonal to a set \mathbf{S} iff it is orthogonal to every element of the set. For given \mathbf{S} the collection of all vectors orthogonal to the given set is called the *orthogonal complement* of \mathbf{S} (we say *\mathbf{S} -perp* and write \mathbf{S}^\perp). In fact, one can prove that \mathbf{S}^\perp is a subspace (even if \mathbf{S} isn't). If we start with a set that is a subspace $\mathbf{S} \subset \mathbf{X}$, then we have a neat decomposition:

$$\mathbf{S} \oplus \mathbf{S}^\perp = \mathbf{X}.$$

This means that any vector in \mathbf{X} can be *uniquely* decomposed into a part that's in \mathbf{S} and a second part in its orthogonal complement. In the plane we picture a line through the origin (\mathbf{S}) and a second line at right-angles to it (\mathbf{S}^\perp). In three-dimensions we might have a plane for (\mathbf{S}) and a perpendicular line for the orthogonal complement. Finally, if \mathbf{S} is a subspace then $\mathbf{S}^{\perp\perp}$ (meaning $[\mathbf{S}^\perp]^\perp$) gets us back to \mathbf{S} .

4.3 Adjoints and Transposes

Suppose we have two *inner-product* spaces (\mathbf{X} and \mathbf{Y}) and a linear map between them $T : \mathbf{X} \mapsto \mathbf{Y}$. For a given $\mathbf{x} \in \mathbf{X}$ and $\mathbf{y} \in \mathbf{Y}$ we can compute the real-number $\langle T\mathbf{x}, \mathbf{y} \rangle_{\mathbf{Y}}$, the inner-product in the \mathbf{Y} -space.

Now we get a little wierd. For the given T and fixed \mathbf{y} suppose we want to evaluate the inner-product for a variety of vectors $\mathbf{x} \in \mathbf{X}$. We complain that this is alot of work: first map the new \mathbf{x} to \mathbf{Y} by computing $T(\mathbf{x})$, and then compute the \mathbf{Y} -space inner-product. For this fixed T and \mathbf{y} is there an element in \mathbf{X} that will work with the \mathbf{X} -space inner-product ? We are asking for an element $\mathbf{x}^* \in \mathbf{X}$ so that

$$\langle \mathbf{x}, \mathbf{x}^* \rangle_{\mathbf{X}} = \langle T\mathbf{x}, \mathbf{y} \rangle_{\mathbf{Y}} \quad \forall \mathbf{x} \in \mathbf{X}.$$

In fact, this requirement defines a new linear operator $T^* : \mathbf{Y} \mapsto \mathbf{X}$. T^* called the *adjoint* of the operator T . In the real-Euclidean case, where T is represented by a matrix M the calculation looks like:

$$\begin{aligned} \langle T\mathbf{x}, \mathbf{y} \rangle_{\mathbf{Y}} &= \langle \mathbf{x}, \mathbf{x}^* \rangle_{\mathbf{X}} \\ (M\mathbf{x})^T \mathbf{y} &= \mathbf{x}^T (M^T \mathbf{y}). \end{aligned}$$

The matrix representation of T^* is simply the transpose of the matrix representation of T . If the scalars are complex we need to transpose the matrix and conjugate the entries.

5 Decomposing Linear Operators

It turns out that a linear map between two vector-spaces ($T : \mathbf{X} \mapsto \mathbf{Y}$), provides a ‘natural’ way to decompose the spaces. First consider the subset of points in \mathbf{Y} that can be produced as the T -image of something in \mathbf{X} . This may be all of \mathbf{Y} , but it need not be. This set is called the *range-space* of T , and, as the name implies, it is a subspace of \mathbf{Y} . The usual symbol is $\mathcal{R}(T)$.

Another set (this one in \mathbf{X}) is those points that get mapped under T to the zero vector in \mathbf{Y} . The set obviously includes the zero vector (in \mathbf{X}), but it can include a lot more. This set is called the *null-space* of T (written $\mathcal{N}(T)$) and is also a subspace (now of \mathbf{X}). Based on the orthogonal complement construction we clearly have

$$\begin{aligned} \mathcal{N}(T) \oplus \mathcal{N}(T)^\perp &= \mathbf{X} \\ \mathcal{R}(T) \oplus \mathcal{R}(T)^\perp &= \mathbf{Y}. \end{aligned}$$

5.1 Decomposition - Inner-Product Spaces

We now examine the decomposition ideas when we have (finite-dimensional) inner-product spaces. It turns out that in this case we have

$$\begin{aligned} [\mathcal{R}(T)]^\perp &= \mathcal{N}(T^*), \quad \text{and} \\ \mathcal{R}(T^*) &= [\mathcal{N}(T)]^\perp. \end{aligned}$$

Let’s consider what the first claim means in the Euclidean-case. We will abuse notation by using the term *matrix* to mean both the transformation and its representation. Let $M : \mathbf{E}^n \mapsto \mathbf{E}^m$. The j -th column of M is a vector (say) $\mathbf{m}^j \in \mathbf{E}^m$. The range space $\mathcal{R}(M)$ is the subspace of \mathbf{E}^m spanned by the set of these column vectors. The subspace $[\mathcal{R}(M)]^\perp \in \mathbf{E}^m$ is the set of vectors orthogonal to each and every such column vector, $\mathbf{m}^j \in \mathbf{E}^m$.

The matrix M^T has the \mathbf{m}^j for its row vectors - so that the action of any such row on vectors in $[\mathcal{R}(M)]^\perp$ will be zero. This means that any vector in $[\mathcal{R}(M)]^\perp$ is also in $\mathcal{N}(M^*)$. The argument can be reversed to show inclusion the other way — the two subspaces ($[\mathcal{R}(M)]^\perp$ and $\mathcal{N}(M^*)$) are the same.

References

- [1] Practical Optimization, Gill, P.E., Murray, W. and Wright, M.H., Academic Press, 1982
- [2] Finite-dimensional vector spaces, Halmos, Paul R., Van Nostrand, 1958, QA261 H33 1958
- [3] Optimization by vector space methods, Luenberger, David G., Wiley, 1969, QA402.5 L8
- [4] Numerical Linear Algebra, Trefethen, L.N., and Bau, D., SIAM, Philadelphia, QA184 T74, 1997