5 Vectorspace, coordinates with respect to a basis. Change of basis. Linear functions and their matrices

5.1 Linear maps from $\mathbf{R}^k$ to $\mathbf{R}^n$

See [D] Section 4.2. The important thing is that any linear map $f : \mathbf{R}^k \to \mathbf{R}^n$ can be expressed by a matrix multiplication. The matrix associated with f is

$$M_f = \begin{pmatrix} f(\mathbf{e}_1) & f(\mathbf{e}_2) & \dots & f(\mathbf{e}_k) \end{pmatrix}$$

i.e. the vectors $f(\mathbf{e}_i)$ form the columns of M_f, where $\mathbf{e}_1, \mathbf{e}_2, \dots$ is the standard basis in $\mathbf{R}^k$

5.2 Vectorspaces

So far we always played with vectors, subspaces etc. within $\mathbf{R}^n$. However, we did not really use that the vectors we considered actually sit in $\mathbf{R}^n$. All we used is that the vectors of $\mathbf{R}^n$ form a set where certain basic operations (namely additions and scalar multiplications) can be performed. Such sets are called vectorspaces. It will turn out that $\mathbf{R}^n$ actually plays an exceptional role among all vectorspaces, it can be used as a "canonical" vectorspace. In fact, after digesting the concept of general vectorspace and how to represent it with the use of the canonical vectorspace $\mathbf{R}^n$, we will see that everything we know about $\mathbf{R}^n$ can immediately be generalized to any vectorspace (of finite dimension) without any extra effort. So it will be worth going through the hassle of a few abstract definitions.

Definition 5.1 *A* **vectorspace** $\mathcal{V}$ *is a set of objects (called vectors, but you don't have to think of actual arrows), where an addition and a scalar multiplication is defined and these operations satisfy the natural properties:*

$$(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$$

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$$

$$\alpha(\mathbf{a} + \mathbf{b}) = \alpha\mathbf{a} + \alpha\mathbf{b}$$

$$(\alpha + \beta)\mathbf{a} = \alpha\mathbf{a} + \beta\mathbf{a}$$

$$\alpha(\beta\mathbf{a}) = (\alpha\beta)\mathbf{a}$$

$$1\mathbf{a} = \mathbf{a} \,.$$

Furthermore, we require that there exists a special vector, called nullvector, denoted by $\mathbf{0}$*, such that*

$$\mathbf{a} + \mathbf{0} = \mathbf{a}$$

for any vector $\mathbf{a}$*, and for any vector there exists another vector, denoted by* $-\mathbf{a}$ *such that*

$$\mathbf{a} + (-\mathbf{a}) = \mathbf{0}$$

REMARK 1.: All these properties ensure that you can follow the usual arithmetic rules. EXCEPT: you are not allowed to multiply two vectors and you are not allowed to divide two vectors. Scalar product and vector product are NOT included in the concept of a vectorspace, though every vectorspace can be equipped with a scalar product, but this is always an extra structure. Unlike in $\mathbf{R}^n$, where there is a “natural” scalar product, most vectorspace do not have a “natural candidate” for it.

REMARK 2.: You might find some of these requirements ridiculous and completely trivial (especially $1\mathbf{a} = \mathbf{a}$ looks funny). But the objects you are dealing with are abstract and a-priori they have no structure at all. You have to put all the rules in and of course you do it in the most “natural” and “convenient” way. But at some point you have to require even the trivialities. Equipping a set by operations abiding certain (even trivial) rules makes the set into a vectorspace. It is a bit like configurating a brand new computer; somewhere in the operating system there is a line which says that the key “A” on the keyboard means “A”.

Of course the most common vectorspaces are automatically "configurated" in the most natural manner, you even do not have to think about it. This convenience is due to the very cleverly developed mathematical formalism, exactly as a good operating system does elementary operations in a way which you think is "natural", hence you do not have to worry about it.

Just imagine that we replace the + sign by − sign in the definition of vector *additions*. Why not? We could say that from now on $\mathbf{a} - \mathbf{b}$ will mean what we used to think of as the *sum* of these vectors. But we keep the old convention when we talk about numbers, i.e. $\alpha + \beta$ is still the sum of these two numbers. There is nothing wrong about it: addition of vectors and addition of numbers are completely different concepts, their notations should not be related. But then the fourth property above would read

$$(\alpha + \beta)\mathbf{a} = \alpha\mathbf{a} - \beta\mathbf{a}$$

Horrible, isn't it? It is like swapping the meaning of the key "A" and "B" on the keyboard when it is uppercase, but keeping the original when it is lowercase. Of course any reasonable operating system would not do such a thing. But you have to tell the computer how to do it in a "reasonable" way. This is the secret of a convenient operating system and the secret of a good mathematical notation is similar.

EXAMPLES OF VECTORSPACES:

Example 1. The basic example is of course $\mathbf{R}^n$ with the usual vector addition and scalar multiplication

Example 2. Any linear subspace of $\mathbf{R}^n$ is a vectorspace itself. The definition of the subspace (see Definition 2.7) ensures that the usual vector addition and scalar multiplication among the vectors make sense within this subspace.

Example 3. Space of polynomials of degree at most n (we will denote it by $\mathcal{P}_n$). It is well known that the sum of two polynomials $p(x)$ and $q(x)$ of degree at most n, is again a polynomial of degree at most n. In short

$$p(x), q(x) \in \mathcal{P}_n \Longrightarrow p(x) + q(x) \in \mathcal{P}_n$$

and similarly

$$p(x) \in \mathcal{P}_n \Longrightarrow \alpha p(x) \in \mathcal{P}_n$$

for any $\alpha \in \mathbf{R}$ scalar. You can easily check that all properties listed in Definition 5.1 are satisfied (these are consequences of the same properties for numbers and functions). In this vectorspace, the "vectors" are actually polynomials (i.e. $x^2 - 3x + 1$ is one vector, $x^3 - 5x$ is another one etc.), hence they do not at all "look like" geometric vectors. But recall that the important thing is not how they look like but what operations you can perform with them.

REMARK: Polynomials of degree *exactly* n do NOT form a vectorspace. For, the zero polynomial is not in this space. Or, you can easily see that the sum of two polynomials of degree, say, $n = 2$, can be a polynomial of degree 1 (e.g. $p(x) = x^2 - 1$, $q(x) = -x^2 + 2x + 3$, then $p(x) + q(x) = 2x - 2$). It means that the addition operation leads out of the set.

Example 4. 3×2 matrices also form a vectorspace (you could replace 2 and 3 with any fixed numbers). The usual matrix addition and scalar multiplication will do. Multiplication of matrices is not involved in this definition, for this purpose you can even forget about it.

Example 5. All continuous functions on $[0, 1]$ also form a vectorspace, again with the "usual" operations (though *infinite dimensional*, which concept we have not defined rigorously). Similarly, all differentiable functions form a vectorspace etc. Any property (like continuity or differentiability) of functions which is preserved under addition and scalar multiplication actually defines a vectorspace (consisting of functions with this property). Hence there are much much more vectorspaces around us, than just the usual geometric spaces like lines, planes etc.

IMPORTANT OBSERVATION: Every concept and theorem discussed in Section 2 starting from Definition 2.1 is valid for any vectorspace not just for subspaces of $\mathbf{R}^n$. If you look at this section more carefully, we never used that the "vectors" are actual geometric "arrows", we used only the basic properties listed in Definition 5.1. In particular we know what it means that vectors are linearly (in)dependent in a vectorspace, we have the concept of subspace and the concept of basis.

A DELICATE REMARK: However, one ingredient is missing; Theorem 2.9 (existence of basis), and the definition of dimension (which relies on this theorem). Recall that in the proof of Theorem 2.9 we used EITHER the Gaussian elimination (to prove Theorem 3.3) OR we used that the vectorspace has a finite spanning set (See Delicate Remark after Lemma 2.12). Gaussian elimination is not available apriori if the vectors are abstract objects and not represented by n-tuples (in fact it is exactly the bases which allow us to "identify" every vectorspace with a subspace of $\mathbf{R}^n$ via coordinatization, see Definition 5.2 below). So the only way we know how to prove Theorem 2.9 in case of a general vectorspace is to start with a spanning set and reduce it to a basis according to part (ii) of Lemma 2.12. Hence what we can really prove for a general vectorspace S by copying Section 2 is that all definitions, theorems are valid for any general vectorspace S which has a finite spanning set. This will be the case in all examples in this course.

EXERCISE: (a) Write up a basis in the vectorspace $\mathcal{P}_3$, i.e. in the space of polynomials of degree at most 3. What is the dimension of $\mathcal{P}_3$? Write up the polynomial $x^3 - 4x^2 + 1$ as a linear combination of your basis elements.

(b) Write up a basis in the vectorspace of 2×3 matrices. What is the dimension of this space? Write up the matrix $\begin{pmatrix} 2 & 4 & -1 \\ 1 & 0 & 6 \end{pmatrix}$ as a linear combination of your basis elements.

You may try to write up a basis in all continuous functions on $[0, 1]$, but it won't work. The reason is that there are "too many" of them. The truth is that these are infinite dimensional vectorspaces, i.e. they contain a set of infinitely many vectors which are linearly independent. The theory of these vectorspaces is well developed and play an enormously important role in a branch of higher analysis called *functional analysis*, but we will not deal with them any more. Here we restrict ourselves to finite dimensional vectorspaces.

5.3 Coordinates

Recall exercise (a) above and Theorem 2.13. Depending on what basis you chose, it may not be very convenient to write a polynomial in that basis. Of course if you smartly chose $\{1, x, x^2, x^3\}$ as a basis, then it was not hard to write up $x^3 - 4x^2 + 2$ as a linear combination of these basis vectors. This is because the "standard" way of presenting a polynomial uses exactly this basis. But the set $\{3, 2x - 1, 2x^2 + x, x^3 - 2x^2 + 1\}$ is also a basis in $\mathcal{P}_3$ (WHY (*)?). If you have to write up $x^3 - 4x^2 + 2$ as a linear combination of these basis vector, it is not immediate. Remember that similar problem was not trivial even in $\mathbf{R}^n$: solving a system of equations $A\mathbf{x} = \mathbf{b}$ can be interpreted as a question of writing the given vector $\mathbf{b}$ as a linear combination of other given vectors, namely the columns of A (recall Problem 4.9 from Section 4.3). However, we have already solved this problem for the case of the vectorspace $\mathbf{R}^n$, it would be nice to reduce our polynomial problem to this. It is indeed possible: from a right point of view every vectorspace $\mathcal{V}$ is "essentially" $\mathbf{R}^n$ (with n chosen to be the dimension of the original space). The idea is the following, which was already introduced in Theorem 2.13, but let us repeat it here for an arbitrary $\mathcal{V}$

Definition 5.2 *Fix a basis in your given vectorspace* $\mathcal{V}$ *(you can choose the "most convenient" one), let it be* $\mathbf{v_1}, \mathbf{v_2}, \ldots \mathbf{v}_n$ *(here of course* n *is the dimension of* $\mathcal{V}$*). We know (Theorem 2.13) that any vector* $\mathbf{u} \in \mathcal{V}$ *can be written as a linear combination of the basis vectors in a unique way:*

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n$$

The numbers $c_1, c_2, \ldots c_n$*, uniquely determined by the basis and* $\mathbf{u}$ *are called the* **coordinates** *of* $\mathbf{u}$ *in this basis.*

Now the key idea is to fix a basis $\mathbf{v_1}, \mathbf{v_2}, \ldots \mathbf{v}_n$ and to **identify** any vector $\mathbf{u}$ with the n-tuple of numbers $(c_1, c_2, \ldots c_n)$ (its coordinates in this basis). This n-tuple of numbers is a vector in $\mathbf{R}^n$, and it is denoted by

$$\Psi_{\{\mathbf{v}\}}(\mathbf{u}) = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} = \mathbf{c}$$

Notice that the notation involves the subscript $\mathbf{v}$ indicating that we are in the $\mathbf{v}$ basis (even more correct would have been to put $\{\mathbf{v_1}, \mathbf{v_2}, \ldots \mathbf{v}_n\}$ into the index, but that's too much...). We also fix the convention that boldface letters from the beginning of the alphabet $(\mathbf{a}, \mathbf{b}, \mathbf{c} \ldots)$ denote elements of $\mathbf{R}^n$, while letters from the end of the alphabet $(\mathbf{u}, \mathbf{v}, \mathbf{w})$ refer to vectors in $\mathcal{V}$.

Of course you can go backwards, and to any vector $\mathbf{c} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \in \mathbf{R}^n$ you can assign the vector

$$\Phi_{\{\mathbf{v}\}}(\mathbf{c}) := c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n \tag{5.1}$$

in $\mathcal{V}$. It is clear that the operations $\Psi_{\{\mathbf{v}\}}$ and $\Phi_{\{\mathbf{v}\}}$ are inverses of each other; the first one goes from the element of the abstract space $\mathcal{V}$ to its coordinates, i.e. to a concrete n-tuple of

$\mathbf{R}^n$, the other one goes backwards: it assigns an abstract element to given coordinates. The notation follows [HH] (Section 2.6), except that a new letter $\Psi_{\{\mathbf{v}\}}$ is introduced for the inverse map $\Phi^{-1}_{\{\mathbf{v}\}}$.

It is important to emphasize that the vector $\mathbf{u}$ and $\mathbf{c} = \Psi_{\{\mathbf{v}\}}(\mathbf{u})$ are not the same, in fact they are completely different objects ($\Psi_{\{\mathbf{v}\}}(\mathbf{u})$ is always an n-tuple of numbers, while $\mathbf{u}$ is an element of $\mathcal{V}$ and it can be a polynomial, a matrix or whatever, depending on $\mathcal{V}$). But for the purpose of additions and scalar multiplications (basic operations in linear algebra) they behave exactly the same way!!! We say that $\mathbf{c} = \Psi_{\{\mathbf{v}\}}(\mathbf{u})$ **represents u**. It means that whatever operation you do on the level of abstract elements in $\mathcal{V}$, their image (representation) will undergo the same operations in $\mathbf{R}^n$. For example

$$\Psi_{\{\mathbf{v}\}}(\mathbf{u}) + \Psi_{\{\mathbf{v}\}}(\mathbf{w}) = \Psi_{\{\mathbf{v}\}}(\mathbf{u} + \mathbf{w})$$

and similarly for scalar multiplication. This important relation is expressed by saying that the map is **linear**.

WARNING: The coordinates DO depend on the basis. The same vector can have different coordinates in different bases.

EXAMPLE: The set $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\} = \{1, x, x^2, x^3\}$ forms a basis in the vectorspace $\mathcal{P}_3$. The polynomial $p(x) = x^3 - 4x^2 + 2$ has a representation

$$p(x) = 2 \cdot 1 + 0 \cdot x + (-4)x^2 + 1 \cdot x^3$$

in this basis, i.e. the ordered set of numbers $(1, -4, 0, 2)$ correspond to $p(x)$ with respect to this basis. In other words

$$\Psi_{\{\mathbf{v}\}}\big(p(x)\big) = \begin{pmatrix} 2 \\ 0 \\ -4 \\ 1 \end{pmatrix}$$

The inverse operation reads

$$\Phi_{\{\mathbf{v}\}}\begin{pmatrix} 2 \\ 0 \\ -4 \\ 1 \end{pmatrix} = 2 \cdot 1 + 0 \cdot x + (-4)x^2 + 1 \cdot x^3 = p(x)$$

However, the set $\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4\} = \{1, 2x-1, 2x^2+x, x^3-2x^2+1\}$ is also a basis in the vectorspace $\mathcal{P}_3$ hence there must be numbers c_1, c_2, c_3, c_4 such that

$$p(x) = x^3 - 4x^2 + 2 = c_1 \cdot 1 + c_2(2x-1) + c_3(2x^2+x) + c_4(x^3 - 2x^2 + 1) \tag{5.2}$$

How to find such numbers? Compare coefficients, if two polynomials coincide, their coefficients must coincide as well (WHY?(*) Because $\{1, x, x^2, x^3\}$ is a basis). In other words, you have to solve the following system of equations:

$$\begin{array}{ccccccccc} c_1 & - & c_2 & & & & c_4 & = & 2 \\ & & 2c_2 & + & c_3 & & & = & 0 \\ & & & & 2c_3 & - & 2c_4 & = & -4 \\ & & & & & & c_4 & = & 1 \end{array} \tag{5.3}$$

The first equation is from equating the constant terms in (5.2), the second from equating the first order (x) terms etc. You easily get the solution $c_1 = \frac{3}{2}$, $c_2 = \frac{1}{2}$, $c_3 = -1$, $c_4 = 1$. Hence the representation is

$$x^3 - 4x^2 + 2 = \frac{3}{2} \cdot 1 + \frac{1}{2}(2x-1) + (-1)(2x^2+x) + 1 \cdot (x^3 - 2x^2 + 1)$$

and the SAME polynomial $p(x)$ correspond to the ordered set of numbers $(\frac{3}{2}, \frac{1}{2}, -1, 1)$. In other words

$$\Psi_{\{\mathbf{w}\}}\Big(p(x)\Big) = \begin{pmatrix} 3/2 \\ 1/2 \\ -1 \\ 1 \end{pmatrix}$$

REMARK: Notice that the system of equations for the c's was triangular. This is of course not necessary, it is due to the fact that the **w** basis is special: the first basis element has degree

0, the second has degree 1 etc. Of course you can have a basis in $\mathcal{P}_3$ such that every basis element is a cubic polynomial. In that case the system (5.3) would be a fully "rectangular" system of equations, and you have to use the row reduction to solve it.

EXERCISE: Find the representation of the polynomial $p(x) = x^3 - 4x^2 + 2$ in the base $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\} = \{x^3 - 1, -x^3 + x + 1, x^3 - 2x^2, 2x^3 + x - 2\}$

ANSWER:

$$p(x) = 2(x^3 - 1) + 1 \cdot (-x^3 + x + 2) + 2(x^3 - 2x^2) + (-1)(2x^3 + x - 2)$$

i.e. the coefficient vector in this basis is

$$\Psi_{\{\mathbf{u}\}}\Big(p(x)\Big) = \begin{pmatrix} 2 \\ 1 \\ 2 \\ -1 \end{pmatrix}$$

5.4 General linear maps and their matrices

The maps defined above $\Phi_{\{\mathbf{v}\}} : \mathbf{R}^n \to \mathcal{V}$ and $\Psi_{\{\mathbf{v}\}} : \mathcal{V} \to \mathbf{R}^n$ respect the basic addition and scalar multiplication in these two vectorspaces. This gives rise to the definition:

Definition 5.3 *A transformation (map) $T : \mathcal{V} \to \mathcal{W}$ between two vectorspaces $\mathcal{V}, \mathcal{W}$ is called* **linear** *if it satisfies*

$$T(\alpha \mathbf{v}_1 + \beta \mathbf{v}_2) = \alpha T(\mathbf{v}_1) + \beta T(\mathbf{v}_2)$$

for all numbers $\alpha, \beta \in \mathbf{R}$ and vectors $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{V}$.

Linear maps can be expressed by a matrix after some work. The issue is that matrices act on n-tuples of numbers, while the elements of a vectorspace can be any object (like polynomials). Hence beforehand, one has to choose a basis in both spaces in order to represent an abstract element of the vectorspace by its coordinates and let the matrix act on the n-tuple of coordinates instead of an abstract object.

Let $\{\mathbf{v}_1, \mathbf{v}_2 \dots \mathbf{v}_k\}$ be a fixed basis in $\mathcal{V}$ and $\{\mathbf{w}_1, \mathbf{w}_2 \dots \mathbf{w}_m\}$ be a fixed basis in $\mathcal{W}$ (notice that the dimensions of these two spaces can be different). Given a vector $\mathbf{u} \in \mathcal{V}$, first we "translate" it into its coordinates, i.e. we consider $\Psi_{\{\mathbf{v}\}}(\mathbf{u}) \in \mathbf{R}^k$. We can also "translate" the image $T\mathbf{u}$ into coordinates, but this vector is in $\mathcal{W}$, hence we have to use the basis there to get an n-tuple of numbers, i.e. $\Psi_{\{\mathbf{w}\}}(T\mathbf{u}) \in \mathbf{R}^m$. The transformation between the k-tuple $\Psi_{\{\mathbf{v}\}}(\mathbf{u})$ and the m-tuple $\Psi_{\{\mathbf{w}\}}(T\mathbf{u})$ is linear (since every map, T, Ψ, Φ are so) hence by Section 5.1 there is a matrix, denoted by ${}_{\mathbf{w}}M^T_{\mathbf{v}}$ such that

$$\Psi_{\{\mathbf{w}\}}(T\mathbf{u}) = \Big[{}_{\mathbf{w}}M^T_{\mathbf{v}}\Big]\Psi_{\{\mathbf{v}\}}(\mathbf{u}) \tag{5.4}$$

or, after inverting Ψ

$$T\mathbf{u} = \Phi_{\{\mathbf{w}\}}\Bigg(\Big[{}_{\mathbf{w}}M^T_{\mathbf{v}}\Big]\Psi_{\{\mathbf{v}\}}(\mathbf{u})\Bigg)$$

Here an apology is in order for the notation... To facilitate it we used square brackets around the matrix ${}_{\mathbf{w}}M^T_{\mathbf{v}}$ to clearly separate it (with its indices) from other terms. But it is important to indicate that the coordinate maps $\Psi_{\{\mathbf{v}\}}, \Psi_{\{\mathbf{w}\}}$ etc. and the transformation matrix depend on the bases.

How to find this matrix ${}_{\mathbf{w}}M^T_{\mathbf{v}}$? Use the relation (5.4) for cleverly chosen $\mathbf{u}$. For example, if $\mathbf{u} = \mathbf{v}_1$, then

$$\Psi_{\{\mathbf{v}\}}(\mathbf{v}_1) = \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

by definition of Ψ. Similarly for any $j = 1, 2, \dots k$

$$\Psi_{\{\mathbf{v}\}}(\mathbf{v}_j) = \mathbf{e}_j$$

where $\mathbf{e}_j$ is the j-th standard basis vector of $\mathbf{R}^k$ (see (2.3)). Since ${}_{\mathbf{w}}M^T_{\mathbf{v}}\Psi_{\{\mathbf{v}\}}(\mathbf{v}_j) = {}_{\mathbf{w}}M^T_{\mathbf{v}}\mathbf{e}_j$ is the j-th column of the matrix ${}_{\mathbf{w}}M^T_{\mathbf{v}}$, we obtain the following result:

Theorem 5.4 *The matrix* ${}_{\mathbf{w}}M^T_{\mathbf{v}}$ *of the linear map* $T : \mathcal{V} \to \mathcal{W}$ *in the fixed bases* $\{\mathbf{v}_1, \mathbf{v}_2 \ldots \mathbf{v}_k\}$ *and* $\{\mathbf{w}_1, \mathbf{w}_2 \ldots \mathbf{w}_m\}$ *is the following:*

$$ {}_{\mathbf{w}}M^T_{\mathbf{v}} = \Big[\Psi_{\{\mathbf{w}\}}(T\mathbf{v}_1) \;\; \Psi_{\{\mathbf{w}\}}(T\mathbf{v}_1) \;\; \ldots \Psi_{\{\mathbf{w}\}}(T\mathbf{v}_k) \Big] \tag{5.5} $$

i.e. it is an $m \times k$ *matrix whose* j*-th column is* $\Psi_{\{\mathbf{w}\}}(T\mathbf{v}_j)$*, i.e. the coordinates (in the* $\mathbf{w}$ *system) of the image of the* j*-th base vector* $\mathbf{v}_j$*.*

EXAMPLE: The differentiation D is a linear map on the set of polynomials (WHY? (*)). In particulat it maps the space $\mathcal{V} = \mathcal{P}_3$ into $\mathcal{W} = \mathcal{P}_2$.

(a) Write up its matrix in the usual basis, i.e. with $\{\mathbf{v}_j\} := \{1, x, x^2, x^3\}$ and $\{\mathbf{w}_j\} := \{1, x, x^2\}$.

(b) Write up its matrix in the basis $\{\mathbf{v}\} := \{2, x+1, x^2 - x, x^3 - x - 1\}$ and $\{\mathbf{w}\} := \{2, x-1, x^2 + x\}$ (First CHECK (*) that these are really bases).

SOLUTION: (a) Clearly $D\mathbf{v}_1 = 0$, $D\mathbf{v}_2 = 1 = \mathbf{w}_1$, $D\mathbf{v}_3 = 2x = 2\mathbf{w}_2$ and $D\mathbf{v}_4 = 3x^2 = 3\mathbf{w}_3$. Hence

$$ \Psi_{\{\mathbf{w}\}}(D\mathbf{v}_1) = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \qquad \Psi_{\{\mathbf{w}\}}(D\mathbf{v}_2) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} $$

$$ \Psi_{\{\mathbf{w}\}}(D\mathbf{v}_3) = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} \qquad \Psi_{\{\mathbf{w}\}}(D\mathbf{v}_4) = \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix} $$

and

$$ {}_{\mathbf{w}}M^T_{\mathbf{v}} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix} $$

(b) Again, we compute the derivatives of the first basis vectors: $D\mathbf{v}_1 = 0$, $D\mathbf{v}_2 = 1$, $D\mathbf{v}_3 = 2x - 1$, $D\mathbf{v}_4 = 3x^2 - 1$. Now we have to take their $\Psi_{\{\mathbf{w}\}}$ image, i.e. we have to write them in the $\{\mathbf{w}\}$ basis. It gives $D\mathbf{v}_1 = 0$, $D\mathbf{v}_2 = \frac{1}{2}\mathbf{w}_1$, $D\mathbf{v}_3 = \frac{1}{2} \cdot 2 + 2(x-1) = \frac{1}{2}\mathbf{w}_1 + 2\mathbf{w}_2$, $D\mathbf{v}_4 = -2 \cdot 2 - 3(x-1) + 3(x^2+x) = -2\mathbf{w}_1 - 3\mathbf{w}_2 + 3\mathbf{w}_3$. Hence

$$ {}_{\mathbf{w}}M^T_{\mathbf{v}} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & -2 \\ 0 & 0 & 2 & -3 \\ 0 & 0 & 0 & 3 \end{pmatrix} $$

5.5 Change of basis

Suppose we have two bases $\left\{\mathbf{v}_1, \mathbf{v_2}, \dots \mathbf{v_n}\right\}$ and $\left\{\mathbf{w}_1, \mathbf{w_2}, \dots \mathbf{w_n}\right\}$ in the same vectorspace $\mathcal{V}$. Let $\mathbf{u}$ be an arbitrary vector in $\mathcal{V}$. It has coordinates in both basis

$$ \mathbf{u} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \dots + a_n\mathbf{v}_n $$

and

$$ \mathbf{u} = b_1\mathbf{w}_1 + b_2\mathbf{w}_2 + \dots + b_n\mathbf{w}_n $$

The coordinate n-tuples (vectors in $\mathbf{R}^n$) $\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \Psi_{\{\mathbf{v}\}}(\mathbf{u})$ and $\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \Psi_{\{\mathbf{w}\}}(\mathbf{u})$ are completely different, but they are related as they determine the same vector $\mathbf{u}$. How to go from one to the other?

Based upon the previous section, it is quite easy. Simply consider the case in Theorem 5.4 when $\mathcal{W} = \mathcal{V}$, i.e. the vectorspaces are the same, but the basis are kept different, and let $T = I$ be the identity map. We obtain

Theorem 5.5 *The matrix of the base change from* $\{\mathbf{v}\}$ *to* $\{\mathbf{w}\}$ *(i.e. the matrix that changes the coordinates of a vector in the base* $\{\mathbf{v}\}$ *into the coordinates of the same vector in the base* $\{\mathbf{w}\}$ *) is* ${}_{\mathbf{w}}M^I_{\mathbf{v}}$*, i.e.*

$$ \Psi_{\{\mathbf{w}\}}(\mathbf{u}) = \left[{}_{\mathbf{w}}M^I_{\mathbf{v}}\right]\Psi_{\{\mathbf{v}\}}(\mathbf{u}) $$

In other words the matrix ${}_{\mathbf{w}}M^I_{\mathbf{v}}$ *(identity matrix in these bases) gives the matrix that changes the coordinates. By (5.5) we have that the columns of* ${}_{\mathbf{w}}M^I_{\mathbf{v}}$ *are the coefficients of the basis* *vectors* $\mathbf{v}_1, \mathbf{v}_2, \ldots$ *written in the* $\{\mathbf{w}\}$ *basis. Sometimes we omit the* I *superscript from the notation.*

Problem 5.6 *What is change of the basis matrix from the basis* $\{\mathbf{v}_j\} = \{2, x+1, x^2-2x, x^3-x-1\}$ *to* $\{\mathbf{w}_j\} = \{1, x, x^2, x^3\}$ *in* $\mathcal{P}_3$*? Use it to rewrite the polynomial* $p(x) = 3\cdot 2 + 2(x+1) - (x^2-2x) + 5(x^3-x-1)$ *(written in the* $\{\mathbf{v}\}$ *basis) into the* $\{\mathbf{w}\}$ *basis.*

Problem 5.7 *What is change of basis matrix from the basis* $\{\mathbf{w}_j\} = \{1, x, x^2, x^3\}$ *to the basis* $\{\mathbf{v}_j\} = \{2, x+1, x^2-2x, x^3-x-1\}$*? Use this matrix to rewrite the polynomial* $q(x) = 2 - 3x + 4x^2 + x^3$ *(written in the* $\{\mathbf{w}\}$ *basis) into the* $\{\mathbf{v}\}$ *basis.*

SOLUTION to Problem 5.6. We write the $\{\mathbf{v}\}$ basis vectors in terms of the $\{\mathbf{w}\}$ basis. Since the elements of $\{\mathbf{v}\}$ are already given in the $\{\mathbf{w}\}$ basis , this is just the coefficients of the $\{\mathbf{v}\}$ basis elements. The matrix is

$$_{\mathbf{w}}M_{\mathbf{v}} = \begin{pmatrix} 2 & 1 & 0 & -1 \\ 0 & 1 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The given polynomial $p(x)$ correspond to the vector

$$\Psi_{\{\mathbf{v}\}}\big(p(x)\big) = \begin{pmatrix} 3 \\ 2 \\ -1 \\ 5 \end{pmatrix}$$

in the $\mathbf{v}$ basis. The same polynomial in the $\mathbf{w}$ basis has coordinates

$$\Psi_{\{\mathbf{w}\}}\big(p(x)\big) = \begin{pmatrix} 2 & 1 & 0 & -1 \\ 0 & 1 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 3 \\ 2 \\ -1 \\ 5 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \\ -1 \\ 5 \end{pmatrix}$$

i.e.

$$p(x) = 3 + (-1)x + (-1)x^2 + 5x^3$$

Of course you could have just multiplied out the given form of the polynomial $p(x) = 3 \cdot 2 + 2(x+1) - (x^2 - 2x) + 5(x^3 - x - 1)$. Notice that the matrix multiplication above does exactly the same: it collects the terms with the same x-degree.

SOLUTION to Problem 5.7. We have to write the $\{\mathbf{w}\}$ basis in terms of $\{\mathbf{v}\}$. This is slightly harder, and in general it requires solving a system of linear equations (like (5.3)). The result is (CHECK (*)): $\mathbf{w}_1 = \frac{1}{2}\mathbf{v}_1$, $\mathbf{w}_2 = -\frac{1}{2}\mathbf{v}_1 + \mathbf{v}_2$, $\mathbf{w}_3 = -\mathbf{v}_1 + 2\mathbf{v}_2 + \mathbf{v}_3$ and $\mathbf{w}_4 = \mathbf{v}_4 + \mathbf{v}_2$. Hence

$${}_{\mathbf{v}}M_{\mathbf{w}} = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & -1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The coordinates of $q(x) = 2 - 3x + 4x^2 + x^3$ in the $\mathbf{w}$ basis is

$$\Psi_{\{\mathbf{w}\}}\Big(q(x)\Big) = \begin{pmatrix} 2 \\ -3 \\ 4 \\ 1 \end{pmatrix}$$

hence its coordinates in the $\mathbf{v}$ basis are given

$$\Psi_{\{\mathbf{v}\}}\Big(q(x)\Big) = \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} & -1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ -3 \\ 4 \\ 1 \end{pmatrix} = \begin{pmatrix} -\frac{3}{2} \\ 6 \\ 4 \\ 1 \end{pmatrix}$$

i.e

$$q(x) = -\frac{3}{2} \cdot 2 + 6(x+1) + 4(x^2 - 2x) + 1(x^3 - x - 1).$$

REMARK: Notice that the matrices ${}_{\mathbf{w}}M_{\mathbf{v}}$ and ${}_{\mathbf{v}}M_{\mathbf{w}}$ are inverses of each other:

$$\Big[{}_{\mathbf{w}}M_{\mathbf{v}}\Big]\;\Big[{}_{\mathbf{v}}M_{\mathbf{w}}\Big] = I$$

and in fact this is true in general:

Theorem 5.8 *Given two bases, $\{\mathbf{v}_j\}$, $\{\mathbf{w_j}\}$ in a vectorspace, the two change of basis matrices are inverses of each other:*

$$\left[{}_{\mathbf{w}}M_{\mathbf{v}}\right]\left[{}_{\mathbf{v}}M_{\mathbf{w}}\right] = I$$

In particular, the change of basis matrix is invertible.

Proof: This statement is natural if you think about what these matrices are doing. The matrix ${}_{\mathbf{w}}M_{\mathbf{v}}$ rewrites the coordinates from $\{\mathbf{v}\}$ to $\{\mathbf{w}\}$ basis, while ${}_{\mathbf{v}}M_{\mathbf{w}}$ rewrites the coordinates from $\{\mathbf{w}\}$ to $\{\mathbf{v}\}$. Recall that matrix multiplication is composition, hence the net effect of the product is the identity transformation (nothing changes).

It is also clear that if you have three bases in the same space, $\{\mathbf{v}_j\}$, $\{\mathbf{w_j}\}$ and $\{\mathbf{u}_j\}$, then the change of basis matrix from $\{\mathbf{v}\}$ basis to $\{\mathbf{u}\}$ basis is just the product of the change of basis matrices from $\{\mathbf{v}\}$ to $\{\mathbf{w}\}$ and from $\{\mathbf{w}\}$ to $\{\mathbf{u}\}$. The order of the multiplication is important:

$${}_{\mathbf{u}}M_{\mathbf{v}} = \left[{}_{\mathbf{u}}M_{\mathbf{w}}\right]\left[{}_{\mathbf{w}}M_{\mathbf{v}}\right]$$

The following is an even more general statement, which easily follows from similar considerations. The proof is not hard, but we omit it.

Theorem 5.9 *Consider three vectorspaces $\mathcal{V}, \mathcal{W}, \mathcal{U}$ and three bases in them, $\{\mathbf{v}_j\}$, $\{\mathbf{w_j}\}$ and $\{\mathbf{u}_j\}$ (the dimensions do not have to be equal) Let $A : \mathcal{V} \to \mathcal{W}$ and $B : \mathcal{W} \to \mathcal{U}$ be linear maps. The matrix of A in the given bases is ${}_{\mathbf{w}}M^A_{\mathbf{v}}$, the matrix of B in the gives bases is ${}_{\mathbf{u}}M^B_{\mathbf{w}}$. One can consider the composition map $B \circ A$ (this is the meaningful order!) from $\mathcal{V}$ to $\mathcal{U}$. Then its matrix is given as the product of the first two matrices:*

$${}_{\mathbf{u}}M^{B\circ A}_{\mathbf{u}} = \left[{}_{\mathbf{u}}M^B_{\mathbf{w}}\right]\left[{}_{\mathbf{w}}M^A_{\mathbf{v}}\right]$$

5.6 Change of the matrix of a linear map as the basis changes

As we discussed in Section 5.4, any linear transformation can be represented by a matrix multiplication. The matrix depends on the choice of basis. Suppose we know the matrix of a linear map in one basis, and we wish to rewrite it in another basis. From the results of the previous section, it is very easy to solve this problem:

Theorem 5.10 *Let $\mathcal{V}$ and $\mathcal{W}$ be two vectorspaces and let $T : \mathcal{V} \to \mathcal{W}$ be a linear map. Let us given two bases in $\mathcal{V}$:*

$$\{\mathbf{v}_1, \mathbf{v}_2, \dots \mathbf{v}_n\} \qquad \textit{and} \qquad \{\widetilde{\mathbf{v}}_1, \widetilde{\mathbf{v}}_2, \dots \widetilde{\mathbf{v}}_n\}$$

and two bases in $\mathcal{W}$

$$\{\mathbf{w}_1, \mathbf{w}_2, \dots \mathbf{w}_n\} \qquad \textit{and} \qquad \{\widetilde{\mathbf{w}}_1, \widetilde{\mathbf{w}}_2, \dots \widetilde{\mathbf{w}}_n\} \ .$$

As usual, the matrix of the linear transformation T using the $\{\mathbf{v}\}$ basis in $\mathcal{V}$ and the $\{\mathbf{w}\}$ basis in $\mathcal{W}$ is ${}_{\mathbf{w}}M^T_{\mathbf{v}}$, and the matrix of the same transformation using the $\{\widetilde{\mathbf{v}}\}$ and $\{\widetilde{\mathbf{w}}\}$ bases is ${}_{\widetilde{\mathbf{w}}}M^T_{\widetilde{\mathbf{v}}}$. Then

$${}_{\widetilde{\mathbf{w}}}M^T_{\widetilde{\mathbf{v}}} = \left[{}_{\widetilde{\mathbf{w}}}M_{\mathbf{w}}\right] \left[{}_{\mathbf{w}}M^T_{\mathbf{v}}\right] \left[{}_{\mathbf{v}}M_{\widetilde{\mathbf{v}}}\right]$$

In other words the matrix in the new basis can be obtained by " sandwiching" the matrix in the old basis between the change of basis matrices.

Proof: Easily follows from Theorem 5.9. □

Notice that the notation is setup in such a way that you even cannot make a mistake: the indices must "match" as you multiply the matrices.

5.7 Change of basis in $\mathbf{R}^n$

In this section we apply the general theorems of the previous sections to the most important case, when the vectorspace $\mathcal{V}$ is $\mathbf{R}^n$ itself. Of course $\mathbf{R}^n$ has a very natural and convenient basis (2.3) but it is not always sufficient to use only this basis. The main reasons for this is that a simple linear transformation in $\mathbf{R}^n$ may look complicated in the standard basis. We will see, for example, that basic rigid motions (e.g. rotation around an axis) have very different matrices in various bases. If you have to operate with such a linear transformation, it could be worth rewriting all vectors etc. in the problem into a more convenient basis, where the matrix of the transformation is simple.

This very important idea stands behind the eigenvectors (among many other applications). We will review them later in Section 7, but you should remember, that if a square matrix has a full set of eigenvectors, then the matrix is diagonal in the basis of these eigenvectors. Diagonal matrices are very easy to deal with (recall Example II. from Section 1.2).

Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_n\}$ be a basis in $\mathbf{R}^n$. Each basis vector is an n-tuple of numbers, so the identification maps $\Phi_{\{\mathbf{v}\}}$ and $\Psi_{\{\mathbf{v}\}}$ are linear maps on $\mathbf{R}^n$ hence, they are actually matrices. We can easily find them:

Theorem 5.11 *Form the matrix* $V = [\mathbf{v}_1 \ \mathbf{v}_2 \ \ldots \mathbf{v}_n]$ *from the basis vectors as its column vectors. Then*

$$\Phi_{\{\mathbf{v}\}} = V, \qquad \Psi_{\{\mathbf{v}\}} = \Phi^{-1}_{\{\mathbf{v}\}} = V^{-1}$$

i.e. the identification maps are just multiplications with the matrices V *and* V^{-1}*.*

REMARK: This theorem implicitly contains the fact that V is invertible. But this is clear from Theorem 4.16 since V is a square matrix with linearly independent columns.

Proof of Theorem 5.11: From the definition of Φ (see (5.1)) clearly $\Phi_{\{\mathbf{v}\}}(\mathbf{e}_1) = \mathbf{v}_1$, $\Phi_{\{\mathbf{v}\}}(\mathbf{e}_2) = \mathbf{v}_2$, etc. Hence $\Phi_{\{\mathbf{v}\}}$ can be given as a multiplication by a matrix whose first column is $\mathbf{v}_1$, second column is $\mathbf{v}_2$, etc. (see Section 5.1). This is exactly the V matrix. Since the map $\Psi_{\{\mathbf{v}\}}$ is the inverse map of $\Phi_{\{\mathbf{v}\}}$ by definition, clearly it is given by the multiplication of the inverse matrix V^{-1}. □

Problem 5.12 *Let*

$$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix} \right\}$$

Write the vector $\mathbf{u} = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$ *in this basis.*

(Notice that this is almost like Problem 4.9, except here you know that the representation exists and unique.)

SOLUTION: We are looking for the coefficients c_1, c_2, c_3 in the representation

$$\mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3$$

i.e. we need to find the vector $\mathbf{c} = \Psi_{\{\mathbf{v}\}}(\mathbf{u})$. Form the matrix

$$V = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \\ 1 & 1 & 2 \end{pmatrix}$$

from these vectors. From Theorem 5.11 we know that $\Psi_{\{\mathbf{v}\}}$ is just the multiplication with the inverse matrix. Compute V^{-1}, we get

$$V^{-1} = \begin{pmatrix} 2 & -1 & 1 \\ -4 & 3 & -2 \\ 1 & -1 & 1 \end{pmatrix}$$

and

$$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = V^{-1}\mathbf{u} = \begin{pmatrix} 2 & -1 & 1 \\ -4 & 3 & -2 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = \begin{pmatrix} 3 \\ -7 \\ 2 \end{pmatrix}$$

Hence the representation is

$$\begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} = 3\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} - 7\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} + 2\begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix}$$

Now suppose that we have another basis $\{\mathbf{w}_1, \mathbf{w}_2, \ldots \mathbf{w}_n\}$ in $\mathbf{R}^n$. How to rewrite a vector $\mathbf{u}$ given in the $\{\mathbf{v}\}$ basis into the $\{\mathbf{w}\}$ basis? This is just a special case of Theorem 5.5.

Theorem 5.13 *Let* $\{\mathbf{v}_1, \mathbf{v}_2, \dots \mathbf{v}_n\}$ *and* $\{\mathbf{w}_1, \mathbf{w}_2, \dots \mathbf{w}_n\}$ *be two bases in* $\mathbf{R}^n$*. Form the matrices* $V = [\mathbf{v}_1\ \mathbf{v}_2\ \dots \mathbf{v}_n]$ *and* $W = [\mathbf{w}_1\ \mathbf{w}_2\ \dots \mathbf{w}_n]$ *from these vectors (as column vectors). Then the coordinates of any vector in the* $\{\mathbf{w}\}$ *basis can be obtained by multiplying the coordinate n-tuple of the same vector in the* $\{\mathbf{v}\}$ *basis by the matrix*

$$_{\mathbf{w}}M_{\mathbf{v}} = W^{-1}V \ .$$

In other words, if a vector $\mathbf{u}$ *is written in the two bases as*

$$\mathbf{u} = \sum_{i=1}^{n} a_i \mathbf{v}_i \qquad \textit{and} \qquad \mathbf{u} = \sum_{i=1}^{n} b_i \mathbf{w}_i \tag{5.6}$$

then the relation between the vectors $\mathbf{a}, \mathbf{b}$ *formed from these coordinates is*

$$\mathbf{b} = W^{-1}V\mathbf{a}$$

REMARK: The matrix $W^{-1}V$ can also be described as follows: its j-th column contains the coordinates of $\mathbf{v}_j$ with respect to the $\{\mathbf{w}\}$ basis. To see this, just choose $a_j = 1$ and all other $a_i = 0$, then from (5.6) $v_j = \sum_{i=1}^{n} b_i \mathbf{w}_i$, where $\mathbf{b} = W^{-1}V\mathbf{a}$ is clearly the j-th column of $W^{-1}V$.

Proof: From Theorem 5.5

$$\Psi_{\{\mathbf{w}\}}(\mathbf{u}) = \Big[{}_{\mathbf{w}}M_{\mathbf{v}}\Big]\ \Psi_{\{\mathbf{v}\}}(\mathbf{u})$$

hence combining it with Theorem 5.11 we have $W^{-1}\mathbf{u} = {}_{\mathbf{w}}M_{\mathbf{v}}V^{-1}\mathbf{u}$ for any $\mathbf{u}$, i.e. ${}_{\mathbf{w}}M_{\mathbf{v}} = W^{-1}V$. $\square$

Problem 5.14 *Let*

$$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} = \left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ 2 \end{pmatrix} \right\}$$

and

$$\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3\} = \left\{ \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix} \right\}$$

be two bases in $\mathbf{R}^n$. *Consider the vector*

$$\mathbf{u} = 2\mathbf{v}_1 + \mathbf{v}_2 - 3\mathbf{v}_3$$

written in the $\{\mathbf{v}\}$ *basis. Find the coordinates of this vector in the* $\{\mathbf{w}\}$ *basis.*

SOLUTION: The vector formed from the coordinates in the $\{\mathbf{w}\}$ basis is $\mathbf{c} = \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}$. Hence the coordinates in the $\{\mathbf{w}\}$ basis are given by the entries of the vector $W^{-1}V\mathbf{c}$, where as usual $V = [\mathbf{v}_1\ \mathbf{v}_2\ \ldots \mathbf{v}_n]$ and $W = [\mathbf{w}_1\ \mathbf{w}_2\ \ldots \mathbf{w}_n]$. Compute the inverse of W:

$$W^{-1} = \begin{pmatrix} 1 & -1 & 2 \\ 0 & 1 & -1 \\ -1 & 2 & -3 \end{pmatrix}$$

and compute the matrix of the change of basis:

$$W^{-1}V = \begin{pmatrix} 1 & -1 & 2 \\ 0 & 1 & -1 \\ -1 & 2 & -3 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \\ 1 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 3 \\ 0 & -1 & -4 \\ 0 & -1 & -5 \end{pmatrix}$$

Hence the new coordinates are given by

$$W^{-1}V\mathbf{c} = \begin{pmatrix} 1 & 1 & 3 \\ 0 & -1 & -4 \\ 0 & -1 & -5 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix} = \begin{pmatrix} -6 \\ 11 \\ 14 \end{pmatrix}$$

i.e.

$$\mathbf{u} = -6\mathbf{w}_1 + 11\mathbf{w}_2 + 14\mathbf{w}_3$$

REMARK 1. Notice that we have found the new coordinates even without ever writing up the $\mathbf{u}$ vector in the standard basis. Of course we could have done that, i.e. first we could have computed

$$\mathbf{u} = 2\mathbf{v}_1 + \mathbf{v}_2 - 3\mathbf{v}_3 = 2\begin{pmatrix}1\\2\\1\end{pmatrix} + \begin{pmatrix}0\\1\\1\end{pmatrix} - 3\begin{pmatrix}-1\\0\\2\end{pmatrix} = \begin{pmatrix}5\\5\\-3\end{pmatrix}$$

(by the way, this is clearly $V\mathbf{c}$), and then we could have used Problem 5.12 to find the $\{\mathbf{w}\}$-coordinates of $\mathbf{u}$ by computing $W^{-1}\begin{pmatrix}-5\\5\\-3\end{pmatrix}$. Notice that the formula $W^{-1}V\mathbf{c}$ does exactly the same in one step.

REMARK 2: Notice that the first basis-vectors in the two bases coincide $\mathbf{v}_1 = \mathbf{w}_1$. But this does NOT mean that the corresponding coordinates in the two representations are the same.

5.8 Change of the matrix of a linear map in $\mathbf{R}^n$ as the basis changes

Here we apply the result of Section 5.6 to the most important case of $\mathbf{R}^n$. The following is just a combination of Theorem 5.10 and Theorem 5.13

Theorem 5.15 *Let f be a linear map from $\mathbf{R}^k$ to $\mathbf{R}^n$. Let us given two bases in $\mathbf{R}^k$, namely $\{\mathbf{v}_1, \dots \mathbf{v}_k\}$ and $\{\widetilde{\mathbf{v}}_1, \dots \widetilde{\mathbf{v}}_k\}$ and two bases in $\mathbf{R}^n$, namely $\{\mathbf{w}_1, \dots \mathbf{w}_n\}$ and $\{\widetilde{\mathbf{w}}_1, \dots \widetilde{\mathbf{w}}_n\}$ Let V, $\widetilde{V}$, W and $\widetilde{W}$ be the matrices whose columns contain these vectors.*

Let the map f be described the matrix ${}_{\mathbf{w}}M^f_{\mathbf{v}}$ using the $\{\mathbf{v}\}$ and $\{\mathbf{w}\}$ bases and it is described by the matrix ${}_{\widetilde{\mathbf{w}}}M^f_{\widetilde{\mathbf{v}}}$ using the $\{\widetilde{\mathbf{v}}\}$ and $\{\widetilde{\mathbf{w}}\}$ bases. Then these two matrices are related by the following **change of bases formula***:*

$${}_{\widetilde{\mathbf{w}}}M^f_{\widetilde{\mathbf{v}}} = \left[\widetilde{W}^{-1}W\right]\left[{}_{\mathbf{w}}M^f_{\mathbf{v}}\right]\left[V^{-1}\widetilde{V}\right]$$

We also say that these two matrices are **conjugates** *of each other.*

This theorem has an obvious special case, when T maps $\mathbf{R}^n$ into itself, i.e. $n = k$. In this case $V = W$ and $\widetilde{V} = \widetilde{W}$ in the theorem above, and we get

Corollary 5.16 *Let f be a linear map from $\mathbf{R}^n$ into itself. Let us given two bases in $\mathbf{R}^n$, namely $\{\mathbf{v}_1, \ldots \mathbf{v}_k\}$ and $\{\tilde{\mathbf{v}}_1, \ldots \tilde{\mathbf{v}}_k\}$. Let V and $\widetilde{V}$ be the matrices whose columns contain these vectors.*

Let the map f be described the matrix ${}_{\mathbf{v}}M^f_{\mathbf{v}}$ using the $\{\mathbf{v}\}$ basis and it is described by the matrix ${}_{\tilde{\mathbf{v}}}M^f_{\tilde{\mathbf{v}}}$ using the $\{\tilde{\mathbf{v}}\}$ basis. Then these two matrices are related by

$$ {}_{\tilde{\mathbf{v}}}M^f_{\tilde{\mathbf{v}}} = \Big[\widetilde{V}^{-1}V\Big] \Big[{}_{\mathbf{v}}M^f_{\mathbf{v}}\Big] \Big[V^{-1}\widetilde{V}\Big] $$

One can also write it as

$$ {}_{\tilde{\mathbf{v}}}M^f_{\tilde{\mathbf{v}}} = \Big[\widetilde{V}^{-1}V\Big] \Big[{}_{\mathbf{v}}M^f_{\mathbf{v}}\Big] \Big[\widetilde{V}^{-1}V\Big]^{-1} $$

to indicate that the new matrix is obtained by "sandwiching" the old matrix between the change of the basis matrix $\widetilde{V}^{-1}V$ and its inverse. This is called **conjugation** *by the matrix $\widetilde{V}^{-1}V$.*

Another special case of Theorem 5.15 is when the $\{\tilde{\mathbf{v}}\}$ and $\{\tilde{\mathbf{w}}\}$ bases are the standard bases. In this case $\widetilde{V} = I$ and $\widetilde{W} = I$ are the identity matrices (of appropriate size), and we have

Corollary 5.17 *Let f be a linear map from $\mathbf{R}^k$ to $\mathbf{R}^n$. Let us given a basis $\{\mathbf{v}_1, \ldots \mathbf{v}_k\}$ in $\mathbf{R}^k$ and a basis $\{\mathbf{w}_1, \ldots \mathbf{w}_n\}$ in $\mathbf{R}^n$. As usual, let $V = \Big[\mathbf{v}_1\ \mathbf{v}_2 \ldots \mathbf{v}_k\Big]$ and $W = \Big[\mathbf{w}_1\ \mathbf{w}_2 \ldots \mathbf{w}_n\Big]$ be the matrices whose columns contain these vectors.*

The map f is described the matrix ${}_{\mathbf{w}}M^f_{\mathbf{v}}$ using the $\{\mathbf{v}\}$ and $\{\mathbf{w}\}$ bases and it is described by the matrix M^f in the standard basis. Then these two matrices are related by

$$ M^f = W \Big[{}_{\mathbf{w}}M^f_{\mathbf{v}}\Big] V^{-1} $$

MEMORIZE THIS THEOREM!!

Finally, we consider the most special case, when both spaces are the same and the tilde basis is the standard one.

Corollary 5.18 *Let f be a linear map from $\mathbf{R}^n$ to itself. Let us given a basis $\{\mathbf{v}_1, \dots \mathbf{v}_k\}$ in $\mathbf{R}^k$, and let V be the matrix whose columns contain these vectors.*

The map f is described the matrix ${}_{\mathbf{v}}M^f_{\mathbf{v}}$ using the $\{\mathbf{v}\}$ basis and it is described by the matrix M^f in the standard basis. Then these two matrices are related by

$$M^f = V \left[{}_{\mathbf{v}}M^f_{\mathbf{v}} \right] V^{-1}$$

MEMORIZE THIS THEOREM!

We introduce the following

Definition 5.19 *Two square matrices A, B of the same dimension are called* **similar** *if there exists an invertible matrix V such that*

$$A = VBV^{-1}$$

In other words, A and B can be viewed as the matrices of the same linear transformation with respect to different bases.

The Section 5.10 and Section 7 will contain applications of this theorem, but before that we discuss the special case of orthonormal bases.

5.9 Orthonormal bases in $\mathbf{R}^n$ and orthogonal matrices

Let $\{\mathbf{q}_1, \mathbf{q}_2, \dots \mathbf{q}_n\}$ be a set of orthonormal vectors in $\mathbf{R}^n$ (we switched the notation from $\mathbf{v}$ to $\mathbf{q}$ because usually the letter q is used for orthogonal or orthonormal objects – this is why you see the letter Q in the QR-algorithm.) We know that these vectors form a basis in $\mathbf{R}^n$ (Lemma 2.6). These bases play a very important role in numerical methods because of their stability properties. They are almost always preferred over any other bases. Certainly the standard basis in $\mathbf{R}^n$ is orthonormal.

The key is the following

Lemma 5.20 *Let $Q = [\mathbf{q}_1\ \mathbf{q}_2\ \ldots \mathbf{q}_n]$ be the $n \times n$ matrix formed from orthonormal vectors. Then we have that*

$$Q^tQ = QQ^t = I \qquad \text{in other words} \qquad Q^{-1} = Q^t \tag{5.7}$$

Every matrix Q with this property has orthonormal columns. Such matrices are called **orthogonal matrices**.

Proof: By orthonormality, all the scalar products $\mathbf{q}_i^t \cdot \mathbf{q}_j$ (with $i \neq j$) are zero and $\mathbf{q}_i^t \cdot \mathbf{q}_i = 1$. This is exactly the same as the relation $Q^tQ = I$. Hence Q^t is the inverse of Q, and by the property of the inverse then it must be inverse "from the other side" as well, i.e. $QQ^t = I$ (see Remark after Definition 4.14). But this property translates into the fact that the *rows* of Q are also orthonormal! This is a highly nontrivial fact if you try to prove it from scratch!! $\square$.

IMPORTANT OBSERVATION: Notice that computing the inverse of an orthogonal matrix is very easy, just take the transpose. This is a big advantage, as in general computing inverse of an arbitrary matrix is hard, long and not very stable operation!

Notice that in Sections 5.7 and 5.8 all formulas contained the inverse of the matrix containing the basis elements. In all these formulas you can certainly replace the inverses by transposes if the basis is orthonormal. This is an ENORMOUS ADVANTAGE!! For example it is very easy to write a given vector in a given orthonormal bases (analogue of Problem 5.12).

Lemma 5.21 *Let $\{\mathbf{q}_1, \mathbf{q}_2, \ldots \mathbf{q}_n\}$ be an orthonormal basis in $\mathbf{R}^n$ and $\mathbf{u}$ is a given vector in $\mathbf{R}^n$. Then the i-th coordinate of $\mathbf{u}$ in this basis is $c_i = \mathbf{q}_i^t \cdot \mathbf{u}$, i.e. the base-decomposition of $\mathbf{u}$ is*

$$\mathbf{u} = (\mathbf{q}_1^t \cdot \mathbf{u})\mathbf{q}_1 + (\mathbf{q}_2^t \cdot \mathbf{u})\mathbf{q}_2 + \ldots + (\mathbf{q}_n^t \cdot \mathbf{u})\mathbf{q}_n \tag{5.8}$$

The square of the norm of $\mathbf{u}$ *is the sum of the coordinate squares:*

$$\|\mathbf{u}\|^2 = (\mathbf{q}_1^t \cdot \mathbf{u})^2 + (\mathbf{q}_2^t \cdot \mathbf{u})^2 + \ldots + (\mathbf{q}_n^t \cdot \mathbf{u})^2 \tag{5.9}$$

This is called **Parseval identity**.

First proof: From Theorem 5.11 it is clear that the coordinates are given by the map $\Psi_{\{\mathbf{q}\}}$, i.e. by multiplication with the matrix Q^{-1}. But $Q^{-1} = Q^t$, hence the entries of $Q^t\mathbf{u}$ are exactly the coordinates $c_i = \mathbf{q}_i^t \cdot \mathbf{u}$. Formula (5.9) is a simple calculation using that $\|\mathbf{u}\|^2 = \mathbf{u}^t \cdot \mathbf{u}$ and the orthonormality of the basis vectors. □

Second proof: We could prove this lemma directly. Since $\{\mathbf{q}\}$ is a basis, we know that there exists numbers c_i such that

$$\mathbf{u} = c_1\mathbf{q}_1 + c_2\mathbf{q}_2 + \ldots + c_n\mathbf{q}_n$$

How to find them? Just multiply this equation by $\mathbf{q}_1^t$. On the left side you get $\mathbf{q}_1^t \cdot \mathbf{u}$. On the right side, the first term gives $c_1\mathbf{q}_1^t \cdot \mathbf{q}_1 = c_1$ and all the other terms are zero by orthogonality. Hence $c_1 = \mathbf{q}_1^t \cdot \mathbf{u}$. You get all the other coordinates similarly. □

REMARK: The *Parseval identity* is valid in any vectorspace equipped with a scalar product. Recall that the concept of scalar product is not included in the definition of a vectorspace. However, the most important vectorspaces can be equipped with convenient scalar products.

Far the most important example is the scalar product defined on the vectorspace of continuous functions on an interval $[a, b]$; it is given as

$$\langle f, g\rangle = \int_a^b f(x)g(x)dx$$

Here we used a new notation $\langle f, g\rangle$ for the scalar product, since the notation $f \cdot g$ could be confused with the usual product of two functions. We again emphasize that the scalar product is a number. Once the scalar product is introduced, it makes sense to talk about "orthogonal" functions, and even orthogonal bases. Notice that (5.9) enormously simplifies calculations; one can compute the "size" of an abstract vector (i.e. function, in this case) as a simple expression of its coordinates, if the coordinates are with respect to an orthonormal basis. Similarly to $\mathbf{R}^n$, in the space of functions we also prefer to work in an orthonormal basis.

There are many orthogonal bases in the vectorspace of functions, however the most natural monomial base $\{1, x, x^2, \ldots\}$ is NOT orthogonal. There are several other choices, each of them is good for particular purposes. The most important is however the *Fourier basis* or *trigonometric* basis, consisting of functions $\{1, \sin x, \sin 2x, \ldots, \cos x, \cos 2x, \ldots\}$ in the space of functions defined on $[0, 2\pi]$. This basis plays a crucial role, among others, in signal analysis, noise filtering and in electrical engineering (an alternating current is most conveniently expressed in a basis of periodic functions). In the past decades more refined orthonormal bases have been developed, called *wavelets*, which are important, among others, in image compression and recognition.

Problem 5.22 *Let*

$$\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3\} = \left\{ \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}, \begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{-2}{\sqrt{6}} \end{pmatrix} \right\}$$

Check that this is an orthonormal basis. Then write the vector $\mathbf{u} = \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}$ *in this basis. Check (5.9) for the length of* $\mathbf{u}$.

SOLUTION: By computing the pairwise scalar products of these vectors, we easily see the orthonormality. Notice that in order to check orthogonality of different basis vectors you do not necessarily have to compute $\mathbf{q}_i^t \cdot \mathbf{q}_j$, any nonzero constant multiple will do. For example computing

$$\mathbf{q}_1^t \cdot \mathbf{q}_3 = \frac{1}{\sqrt{3}} \cdot \frac{1}{\sqrt{6}} + \frac{1}{\sqrt{3}} \cdot \frac{1}{\sqrt{6}} + \frac{1}{\sqrt{3}} \cdot \frac{-2}{\sqrt{6}} = 0$$

is more painful than checking that

$$\begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} = 1 \cdot 1 + 1 \cdot 1 + 1 \cdot (-2) = 0$$

But when you check the length of the vectors, $\mathbf{q}_i^t \cdot \mathbf{q}_i = 1$, then of course the "ugly" prefactors are important.

To give the representation of $\mathbf{u}$ in this basis, we compute

$$\mathbf{q}_1^t \cdot \mathbf{u} = -\frac{1}{\sqrt{3}}$$

$$\mathbf{q}_2^t \cdot \mathbf{u} = \frac{2}{\sqrt{2}} = \sqrt{2}$$

$$\mathbf{q}_3^t \cdot \mathbf{u} = \frac{2}{\sqrt{6}} = \frac{\sqrt{2}}{\sqrt{3}}$$

i.e.

$$\begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix} = -\frac{1}{\sqrt{3}} \begin{pmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix} + \sqrt{2} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} + \frac{\sqrt{2}}{\sqrt{3}} \begin{pmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{-2}{\sqrt{6}} \end{pmatrix}$$

Notice that we never had to invert a matrix (unlike in the analogous Problem 5.12).

The length of $\mathbf{u} = \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}$ is clearly $\sqrt{3} = \sqrt{1^2 + (-1)^2 + (-1)^2}$. The sum of the squares of the coordinates in the $\{\mathbf{q}\}$ basis is

$$\Big(-\frac{1}{\sqrt{3}}\Big)^2 + \big(\sqrt{2}\big)^2 + \Big(\frac{\sqrt{2}}{\sqrt{3}}\Big)^2 = 3$$

hence (5.9) checks out. □

Now we use this big advantage of orthonormal bases for the most important case, which is just a reformulation of Corollary 5.18:

Corollary 5.23 *Let f be a linear map from $\mathbf{R}^n$ to itself. Let us given an orthonormal basis $\{\mathbf{q}_1, \ldots \mathbf{q}_k\}$ in $\mathbf{R}^k$, and let Q be the orthogonal matrix whose columns contain these vectors.*

The map f is described the matrix ${}_{\mathbf{q}}M^f_{\mathbf{q}}$ using the $\{\mathbf{q}\}$ basis and it is described by the matrix M^f in the standard basis. Then these two matrices are related by

$$M^f = Q \left[{}_{\mathbf{q}}M^f_{\mathbf{q}} \right] Q^t$$

Recall that a matrix multiplication can be interpreted as a geometric transformation in the plane (if $n = 2$), in the space (if $n = 3$) and even in general n with some abstraction. How do the orthogonal matrices act?

Theorem 5.24 *Let Q be an $n \times n$ orthogonal matrix. Then the action of Q on $\mathbf{R}^n$ preserves the length of vectors and the angle between vectors. In other words, it leaves the geometry unchanged; there is no distortion.*

The converse statement is also true. Let Q be an $n \times n$ matrix with the propety that $\|Q\mathbf{v}\| = \|\mathbf{v}\|$ for any vector $\mathbf{v} \in \mathbf{R}^n$. Then Q is orthogonal.

Proof: Since both the length and the angle are determined by the scalar product, it is enough to show that Q leaves the scalar product unchanged. This means that if $\mathbf{x}$ and $\mathbf{y}$ are arbitrary vectors, then

$$\mathbf{x}^t \cdot \mathbf{y} = (Q\mathbf{x})^t \cdot Q\mathbf{y}$$

But this is straighforward, since $(Q\mathbf{x})^t \cdot Q\mathbf{y} = \mathbf{x}^t \cdot Q^t Q\mathbf{y} = \mathbf{x}^t \cdot \mathbf{y}$ by (5.7).

For the converse statement, notice that $\mathbf{q}_1 = Q\mathbf{e}_1$, $\mathbf{q}_2 = Q\mathbf{e}_2 \ldots \mathbf{q}_n = Q\mathbf{e}_n$ are the columns of Q. We know that $\|\mathbf{q}_j\| = \|Q\mathbf{e}_j\| = \|\mathbf{e}_j\| = 1$, hence the columns are normalized. Now compute $\|\mathbf{q}_i + \mathbf{q}_j\|^2$ in two different ways if $i \neq j$

$$\|\mathbf{q}_i + \mathbf{q}_j\|^2 = (\mathbf{q}_i + \mathbf{q}_j)^t \cdot (\mathbf{q}_i + \mathbf{q}_j) = 2 + \mathbf{q}_i^t \cdot \mathbf{q}_j \qquad (CHECK(*))$$

and

$$\|\mathbf{q}_i + \mathbf{q}_j\|^2 = \|Q(\mathbf{e}_i + \mathbf{e}_j)\|^2 = \|\mathbf{e}_i + \mathbf{e}_j\|^2 = 2 + \mathbf{e}_i^t \cdot \mathbf{e}_j = 2 \qquad (CHECK(*)!)$$

Hence we get that $\mathbf{q}_i^t \cdot \mathbf{q}_j = 0$, i.e. we proved that the columns of Q are orthonormal vectors, i.e. Q is an orthogonal matrix. $\square$

In particular, if $\{\mathbf{v}_1, \mathbf{v}_2, \dots \mathbf{v}_n\}$ is an orthonormal basis, then $\{Q\mathbf{v}_1, Q\mathbf{v}_2, \dots Q\mathbf{v}_n\}$ is also an orthonormal basis. Moreover, usual rotations and reflections in $n = 2, 3$ dimension are given by orthogonal matrices. In fact, the orthogonal matrices (or orthogonal transformations) should be viewed as the higher dimensional generalization of rotations and reflections.

You can consider an orthonormal set of vectors $\{\mathbf{q}_1, \mathbf{q}_2, \dots \mathbf{q}_k\}$ in $\mathbf{R}^n$ even if $k < n$ (clearly $k \le n$ must always be true. WHY?(*)). The matrix $Q = [\mathbf{q}_1\ \mathbf{q}_2\ \dots \mathbf{q}_k]$ is not a square matrix any more, but the orthogonality relation implies that $Q^tQ = I$ is a $k \times k$ identity matrix. In this case QQ^t is not the identity matrix (its rank is at most k, since its columns are linear combinations of $\{\mathbf{q}_1, \mathbf{q}_2, \dots \mathbf{q}_k\}$), but it is still an important matrix: this is the **matrix of orthogonal projection** from $\mathbf{R}^n$ to $\text{Span}\{\mathbf{q}_1, \mathbf{q}_2, \dots \mathbf{q}_k\} = R(Q)$. For more details see Section 6.

5.10 Rotation around arbitrary axis in $\mathbf{R}^3$

Recall that in Section 4.5 [D] the matrix of the rotation around the x-axis with angle θ is given as

$$M = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \tag{5.10}$$

(there was a typo in [D]; the upper left corner is 1 and not 0), and similar formulas are valid for rotations around the other two axes. The question is to write up the matrix of a general rotation.

Problem 5.25 *Write up the matrix of rotation around the vector* $\mathbf{v} = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$ *with angle* $\theta = \frac{\pi}{6}$. *Compute the image of the point* $P(1,1,2)$ *under this map.*

REMARK 1: If we do not specify the basis, then the matrix of a linear transformation in $\mathbf{R}^n$ is understood to be with respect to the standard basis.

REMARK 2: Recall that the rotation axis with a direction and a (possibly negative) angle uniquely determines a rotation. Remember the convention of right hand rule: grab the axis with your right hand so that the thumb points in the given direction. Then the rotation is given by a positive angle if it rotates in the direction where your fingers point.

SOLUTION: We will find a good orthonormal basis in which the rotation matrix is nice, then we will use Corollary 5.23 to go back to the standard basis. The good basis contains the rotation axis as one of its element. Hence we first normalize it

$$\mathbf{q}_1 = \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{1}{3}\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$$

Now we need two more orthonormal basis vectors in the plane orthogonal to $\mathbf{q}_1$. The fancy way of doing this is to use Gram-Schmidt procedure for the basis $\{\mathbf{q}_1, \mathbf{e}_2, \mathbf{e}_3\}$. But here one can just find a vector orthogonal to $\mathbf{q}_1$, say $\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$ (just to make sure that their scalar product is zero) and normalize it to become $\mathbf{q}_2$:

$$\mathbf{q}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix}$$

There is not much freedom in choosing the third vector, since it has to be orthogonal to both $\mathbf{q}_1, \mathbf{q}_2$, it has to be normalized, and it has to form a positively oriented triple with $\mathbf{q}_1, \mathbf{q}_2$ (this latter is not terribly important, but we have the convention about positively and negatively oriented rotations and the rotation matrix (5.10) complies with the right hand rule only in positively oriented bases). Hence it is

$$\mathbf{q}_3 = \frac{\mathbf{q}_1 \times \mathbf{q}_2}{\|\mathbf{q}_1 \times \mathbf{q}_2\|}$$

i.e. it is the unit vector in the direction of

$$\mathbf{q}_1 \times \mathbf{q}_2 = \begin{pmatrix} 4 \\ -1 \\ -1 \end{pmatrix}$$

hence, after normalization

$$\mathbf{q}_3 = \frac{1}{\sqrt{18}} \begin{pmatrix} 4 \\ -1 \\ -1 \end{pmatrix}$$

Again, we form the matrix Q from these vectors:

$$Q = \begin{pmatrix} \frac{1}{3} & 0 & \frac{4}{\sqrt{18}} \\ \frac{2}{3} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} \\ \frac{2}{3} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} \end{pmatrix}$$

(CHECK (*) that it is an orthogonal matrix!)

The key is that in the $\{\mathbf{q}\}$ basis our rotation matrix is nice, it has the form of (5.10), i.e.

$$_{\mathbf{q}}M_{\mathbf{q}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix}$$

using $\theta = \frac{\pi}{6}$. Hence by Corollary 5.23 the matrix of this rotation in the standard basis is

$$M = Q \left[_{\mathbf{q}}M_{\mathbf{q}} \right] Q^t = \begin{pmatrix} \frac{1}{3} & 0 & \frac{4}{\sqrt{18}} \\ \frac{2}{3} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} \\ \frac{2}{3} & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{3}}{2} & -\frac{1}{2} \\ 0 & \frac{1}{2} & \frac{\sqrt{3}}{2} \end{pmatrix} \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{4}{\sqrt{18}} & -\frac{1}{\sqrt{18}} & -\frac{1}{\sqrt{18}} \end{pmatrix}$$

$$= \begin{pmatrix} .8809 & -.3035 & .3631 \\ .3631 & .9255 & -.1071 \\ -.3035 & .2262 & .9255 \end{pmatrix} \qquad (5.11)$$

(notice that no inverse was taken!)

Finally, we compute the image of the point $P(1, 1, 2)$ as

$$\begin{pmatrix} .8809 & -.3035 & .3631 \\ .3631 & .9255 & -.1071 \\ -.3035 & .2262 & .9255 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1.303 \\ 1.074 \\ 1.773 \end{pmatrix}$$