UNIVERSIDAD DEL CEMA
Buenos Aires
Argentina

Serie
DOCUMENTOS DE TRABAJO

Área: Matemática y Economía

Intuitive Mathematical Economics Series.
Linear Structures I
Linear Manifolds, Vector Spaces and Scalar Products

Sergio A. Pernice

Versión Mayo 2019
Nro. 689

www.cema.edu.ar/publicaciones/doc_trabajo.html
UCEMA: Av. Córdoba 374, C1054AAP Buenos Aires, Argentina
ISSN 1668-4575 (impreso), ISSN 1668-4583 (en línea)
Editor: Jorge M. Streb; asistente editorial: Valeria Dowding jae@cema.edu.ar

Intuitive Mathematical Economics Series

Linear Structures I

Linear Manifolds, Vector Spaces and Scalar Products

SERGIO A. PERNICE[1]

Universidad del CEMA
Av. Córdoba 374, Buenos Aires, 1054, Argentina

Abstract

Linear algebra is undoubtedly one of the most powerful structures of pure and applied mathematics. It combines extreme generality with deeply held spatial intuitions. In economics and data science, one would dare to say, it lies at the basis of most of the other mathematical techniques used. Yet, standard presentations of the subject tend to refrain from displaying the full extent of the deeply intuitive nature of the underlying structures, despite the fact that such intuitions are so useful when applying linear algebra, and when extending techniques to tackle nonlinear problems. This is the first paper of the "Intuitive Mathematical Economics Series", dedicated to presenting linear algebra's intuitive and general nature. In this case we present linear manifolds and vector spaces.

Keywords: Vector spaces, linear manifolds.

1 Introduction

Linear algebra is undoubtedly one of the most powerful structures of pure and applied mathematics. It combines extreme generality with deeply held spatial intuitions. In economics and data science, one would dare to say, it lies at the basis of most of the other mathematical techniques used. Yet, standard presentations of the subject tend to refrain from displaying the full extent of the deeply intuitive nature of the underlying structures, despite the fact that such intuitions are so useful when applying linear algebra, and when extending techniques to tackle nonlinear problems.

[1]sp@ucema.edu.ar

This is a β version of this paper. Any suggestion and/or pointing of errors by email are welcome.
The points of view of the author do not necessarily represent the position of Universidad del CEMA.

In the context of the “Intuitive Mathematical Economics Series", this first paper is dedicated to presenting linear manifolds, vector spaces and scalar products. We leave linear mappings for further works.

In line with the idea of presenting mathematical techniques displaying both the underlying intuitions and the powerful generality of the fundamental abstract ideas, there are many very useful videos on the Internet that I encourage the reader to explore. In particular, as a preview for this paper, I recommend watching the first two videos, [3B1B LA1] and [3B1B LA2], of the “Essence of linear algebra" series in the *3Blue1Brown*, Youtube channel, by Grant Sanderson.

The paper is organized as follows. In section 2 we present some general equilibrium linear models as motivators for linear structures. In section 3 we present the intuitions behind linear manifolds. We emphasize the distinction between the underlying linear manifold, modeled as a generalization of real space, and vector spaces, modeled as generalizations of displacements in real space. In section 4 we introduce vector spaces, first emphasizing the intuitions just mentioned, then their abstract generality, then standard definitions and theorems, and finally return to the intuitions behind them. In this section we fully display a general theme of this series: the utility of developing the ability to switch back and forth between an intuitive mode of thinking and a rational, abstract mode. It is in this back and forth that the full power of mathematical structures and their applications can be mastered.

Powerful as vector spaces are, they do not fully capture important aspects of our spatial intuitions such as distance and orthogonality. For that we need to endow linear spaces with a scalar product. This additional structure greatly enlarges the domain of applications of linear algebra. These are the contents of section 5. In section 6, we return, more formally, to the relationship between vector spaces and the underlying manifold, and the different ways to view linear equations that this relationship enables. Section 7 deals with the important problem of finding the point in a subspace closest to a given point in the full vector space, or its equivalent in the underlying manifold. We conclude this section by showing that the useful econometric techniques of least square linear regressions are nothing but an example of this general problem. Finally, in section 8 we conclude and briefly look ahead to the next paper on linear structures.

Notation: we use capital letters, like A, to refer to points in the linear space (or manifold), and also for collections of points such as subspaces. Small letters like a refer to numbers (or scalars), bold small letters like $\mathbf{v}$ refer to vectors, and bold capital letters $\mathbf{A}$ refer to matrices.

2 A prototypical example: general equilibrium in a linear market model

Suppose we have two goods, 1 and 2, that have linear demand and supply curves, and suppose that the two markets interact with each other:

$$\begin{aligned} q_1^d &= a_{10}^d + a_{11}^d p_1 + a_{12}^d p_2 \qquad (2.1) \\ q_1^s &= a_{10}^s + a_{11}^s p_1 + a_{12}^s p_2 \qquad (2.2) \\ q_2^d &= a_{20}^d + a_{21}^d p_1 + a_{22}^d p_2 \qquad (2.3) \\ q_2^s &= a_{20}^s + a_{21}^s p_1 + a_{22}^s p_2 \qquad (2.4) \end{aligned}$$

The upper index d refers to *demand*, and s to *supply*, p_i refers to the price of product i. For the moment we assume only two products. The coefficients a_{ij}^d (a_{ij}^s) characterize the demand (supply) curve: a_{i0}^d (a_{i0}^s) is the quantity demanded (supplied) of the good i when all prices are zero, and a_{ij}^d (a_{ij}^s) for $i, j \neq 0$ is the amount that the demand (supply) of good i increases (if positive) or decreases (if negative) for a unit increase in price of good j.

For the time being, there is no restriction on the possible values of these coefficients, but in every concrete case, the economics of the problem will naturally impose some restrictions.

Consider one numerical case:

$$\begin{aligned} q_1^d &= 10 - 1.0 \times p_1 + 0.5 \times p_2 \qquad (2.5) \\ q_1^s &= 0 + 1.0 \times p_1 + 0.0 \times p_2 \qquad (2.6) \\ q_2^d &= 20 + 0.7 \times p_1 - 2.0 \times p_2 \qquad (2.7) \\ q_2^s &= 0 + 0.0 \times p_1 + 1.5 \times p_2 \qquad (2.8) \end{aligned}$$

Let's interpret, for example, line (2.7): the quantity demanded of product 2 would be 20 units if $p_1 = p_2 = 0$, it will increase by 0.7 units for every unit increase in the price of product 1, and will decrease by 2.0 units for every unit increase in the price of product 2. Similarly for the other lines.

If the market is free, the prices will adjust so that

$$q_i^d = q_i^s \equiv q_i, \quad i = 1, 2. \qquad (2.9)$$

therefore the curves (2.1-2.4) imply equations in equilibrium that can be written in various ways.

On the one hand, we can simply equate the right hand side of (2.1) to the right hand side of (2.2), and similarly for (2.3) and (2.4), to find the equations:

$$\begin{aligned} a_{10}^d + a_{11}^d p_1 + a_{12}^d p_2 &= a_{10}^s + a_{11}^s p_1 + a_{12}^s p_2 \\ a_{20}^d + a_{21}^d p_1 + a_{22}^d p_2 &= a_{20}^s + a_{21}^s p_1 + a_{22}^s p_2 \end{aligned}$$

which can be reordered as

$$\left(a^d_{11} - a^s_{11}\right) p_1 + \left(a^d_{12} - a^s_{12}\right) p_2 = \left(a^s_{10} - a^d_{10}\right) \quad (2.10)$$

$$\left(a^d_{21} - a^s_{21}\right) p_1 + \left(a^d_{22} - a^s_{22}\right) p_2 = \left(a^s_{20} - a^d_{20}\right) \quad (2.11)$$

This is a two step process. In the first step, solving the 2 equations with 2 unknowns (2.10-2.11), one obtains the equilibrium prices. And in the second step, inserting these prices in (2.1) and (2.3), one obtains the equilibrium quantities.

For the particular numerical example (2.5-2.8), the system of equations (2.10-2.11), with a trivial rearrangement, becomes:

$$p_2 = 4p_1 - 20 \quad (2.12)$$

$$p_2 = 0.2p_1 + 5.71 \quad (2.13)$$

Alternatively, one can write 4 equations with 4 unknowns

$$\begin{aligned}
q_1 &= a^d_{10} + a^d_{11} p_1 + a^d_{12} p_2 \\
q_1 &= a^s_{10} + a^s_{11} p_1 + a^s_{12} p_2 \\
q_2 &= a^d_{20} + a^d_{21} p_1 + a^d_{22} p_2 \\
q_2 &= a^s_{20} + a^s_{21} p_1 + a^s_{22} p_2
\end{aligned}$$

which can be rewritten as

$$q_1 - a^d_{11} p_1 - a^d_{12} p_2 = a^d_{10} \quad (2.14)$$

$$q_1 - a^s_{11} p_1 - a^s_{12} p_2 = a^s_{10} \quad (2.15)$$

$$q_2 - a^d_{21} p_1 - a^d_{22} p_2 = a^d_{20} \quad (2.16)$$

$$q_2 - a^s_{21} p_1 - a^s_{22} p_2 = a^s_{20} \quad (2.17)$$

and solve it to find both, the prices and the quantities, simultaneously. This is a single step process, but it requires solving a system of 4 equations with 4 unknowns.

For the particular numerical example (2.5-2.8), the system of equations (2.14-2.17), becomes:

$$q_1 + 1.0 \times p_1 - 0.5 \times p_2 = 10.0 \quad (2.18)$$

$$q_1 - 1.0 \times p_1 - 0.0 \times p_2 = 0.0 \quad (2.19)$$

$$q_2 - 0.7 \times p_1 + 2.0 \times p_2 = 20.0 \quad (2.20)$$

$$q_2 - 0.0 \times p_1 - 1.5 \times p_2 = 0.0 \quad (2.21)$$

Whichever your preferred way of solving the equations, the solutions are $q_1 = 6.77$, $q_2 = 10.60$, $p_1 = 6.77$, $p_2 = 7.07$.

Systems (2.10-2.11), or (2.14-2.17) can be written in matrix form. In the first case the matrix form is:

$$\begin{pmatrix} a^d_{11} - a^s_{11} & a^d_{12} - a^s_{12} \\ a^d_{21} - a^s_{21} & a^d_{22} - a^s_{22} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} a^s_{10} - a^d_{10} \\ a^s_{20} - a^d_{20} \end{pmatrix} \quad (2.22)$$

which, for the particular numerical example (2.5-2.8) becomes:

$$\begin{pmatrix} -2.0 & 0.5 \\ 0.7 & -3.5 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} -10.0 \\ -20.0 \end{pmatrix} \tag{2.23}$$

equivalent to (2.12-2.13).

HW 2.1: Prove the equivalence.

The matrix form of (2.14-2.17) is:

$$\begin{pmatrix} 1 & 0 & -a^d_{11} & -a^d_{12} \\ 1 & 0 & -a^s_{11} & -a^s_{12} \\ 0 & 1 & -a^d_{21} & -a^d_{22} \\ 0 & 1 & -a^s_{21} & -a^s_{22} \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} a^d_{10} \\ a^s_{10} \\ a^d_{20} \\ a^s_{20} \end{pmatrix} \tag{2.24}$$

which, for the particular numerical example (2.5-2.8) becomes:

$$\begin{pmatrix} 1 & 0 & 1.0 & -0.5 \\ 1 & 0 & -1.0 & 0.0 \\ 0 & 1 & -0.7 & 2.0 \\ 0 & 1 & 0.0 & -1.5 \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \\ p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} 10.0 \\ 0.0 \\ 20.0 \\ 0.0 \end{pmatrix} \tag{2.25}$$

Forms (2.10-2.11) and (2.14-2.17) (or (2.12-2.13) and (2.18-2.21), if you prefer a more concrete example), evoke a different picture than forms (2.22) and (2.24) (or (2.23) and (2.25)).

Let us consider for simplicity and concreteness the two-dimensional case (2.12-2.13), and its corresponding matrix form (2.23). Look at equations (2.12) and (2.13); they correspond to straight lines in the "space of prices (p_1, p_2)", see Fig. 1. The task is to find the intersection of these two lines.

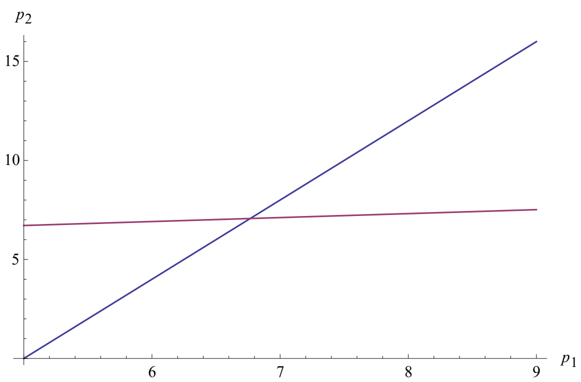


Figure 1: Picture evoked by equations (2.12-2.13): a pair of straight lines in the "space of prices (p_1, p_2)". Solving the equations implies finding the intersection between these two lines.

The picture evoked by the matrix-vector equation (2.23) is completely different: the matrix transforms vectors to vectors, solving the problem implies finding the particular vector of prices that is transformed by the matrix into the vector[2] $(-10, -20)^T$, see Fig. 2.

[2] The upper index "T", as in $(-10, -20)^T$, represents the "transpose" operation. It will defined in part 2 of this

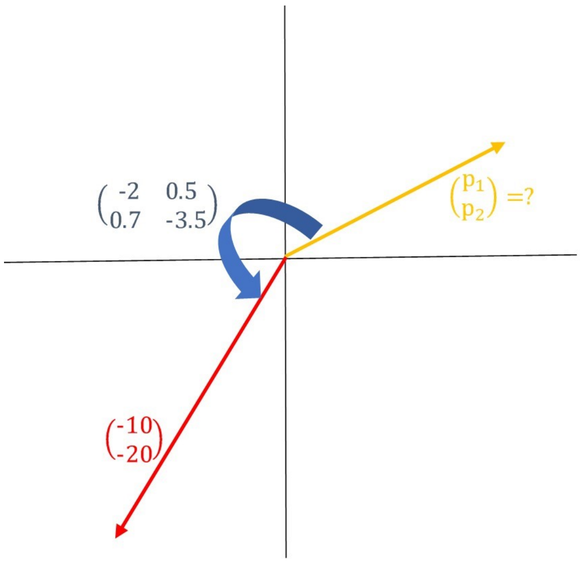


Figure 2: Picture evoked by equation (2.23) in matrix form: the matrix transforms the vector of prices into another vector. Solving the problem implies finding the particular vector of prices that is transformed by the matrix into the vector $(-10, -20)^T$.

These pictures are not limited to two-dimensional systems. As we will see, each line in equations (2.18-2.21) represents a linear "3-dimensional surface", or "3-linear surface", in the four-dimensional space of parameters (q_1, q_2, p_1, p_2). In general, except for degenerate cases, four 3-linear surfaces will intersect in a point, just like two 1-linear surfaces (or straight lines) will in general intersect in a point in 2 dimensions, and three 2-linear surfaces (or planes) will in general intersect in a point in 3 dimensions. More generally, except in degenerate cases, a number n of $(n-1)$-linear surfaces will intersect in a single point in an n-dimensional space.

Similarly, in general, except for degenerate cases, a 4×4 matrix like the one appearing in (2.25) transforms every 4-dimensional vector, or "4-vector" into another 4-vector in a one to one relationship, just as a 3×3 matrix in general transforms every "3-vector" into another 3-vector in a one to one relationship, and an $n \times n$ matrix in general transforms every "n-vector" into another n-vector in a one to one relationship.

If we have a market of n products with linear demand and supply curves, model (2.1-2.4) becomes, in a slightly more abstract notation:

$$q_i^d = a_{i0}^d + \sum_{j=1}^{n} a_{ij}^d p_j \tag{2.26}$$

$$q_i^s = a_{i0}^s + \sum_{j=1}^{n} a_{ij}^s p_j, \quad i = 1, \cdots, n \tag{2.27}$$

---

work. For the moment simply assume that $(a, b)^{\mathrm{T}} = \begin{pmatrix} a \\ b \end{pmatrix}$.

In equilibrium

$$q_i^d = q_i^s \equiv q_i, \quad i = 1, \cdots, n. \tag{2.28}$$

In line with the previous analysis, one can do two things. One possibility is to equate the right hand side of (2.26) with the right hand side of (2.27), leading to n equations with n unknown p_j's:

$$\sum_{j=1}^{n} \left(a_{ij}^d - a_{ij}^s\right) p_j = \left(a_{i0}^s - a_{i0}^d\right), \quad i = 1, \cdots, n. \tag{2.29}$$

For each i, (2.29) represents an "$(n-1)$-surface" in the n dimensional space of parameters $(p_1, \cdots, p_n)$. Since there are n of them, in general the solution will be a unique point in that space.

Or one can write it in the matrix-vector form, that in a more compact notation is

$$\mathbf{Ap} = \mathbf{a_0} \tag{2.30}$$

where $\mathbf{A}$ represents an $n \times n$ matrix with elements:

$$A_{ij} = a_{ij}^d - a_{ij}^s, \tag{2.31}$$

$\mathbf{p}$ represents an n-dimensional column vector whose elements are the unknown prices:

$$\text{element } j \text{ of } \mathbf{p} = p_j \tag{2.32}$$

and $\mathbf{a_0}$ represents an n-dimensional column vector whose elements are:

$$\text{element } i \text{ of } \mathbf{a_0} = a_{i0}^s - a_{i0}^d \tag{2.33}$$

Equation (2.30), like (2.22), evokes the picture represented in Fig. 2, but for n-dimensional vectors. It is the same idea: the problem is finding the unknown vector $\mathbf{p}$ that the matrix A transforms into $\mathbf{a_0}$.

Whatever form you prefer, the solution gives the n equilibrium prices p_j, and in a second step one finds the n equilibrium quantities q_j.

Alternatively one can write the $2n$ equations with $2n$ unknowns that generalizes equations (2.14-2.17) and represent $2n$ "$(2n-1)$-linear surfaces" in the $2n$ dimensional space of parameters $(q_1, \cdots, q_n, p_1, \cdots, p_n)$. Or, one can write the $2n \times 2n$ matrix-vector equation that generalizes (2.24), to find simultaneously the n equilibrium quantities and prices.

HW 2.2: Write down explicitly the n products generalization of equations (2.14-2.17) and equation (2.24).

The two pictures could not be more different. But of course they have to be equivalent, since they give the same result. There are many ways of understanding this equivalence. In this paper we will focus on the picture described in Fig. 1. Various ways of understanding this picture will be presented. Along the way we will acquire many concepts very useful for economists, some of them not emphasized in standard textbooks. In paper II of this work, we will focus on the picture described in Fig. 2 and the equivalence between the two pictures.

3 The Manifold

The picture evoked by figures like 1, or its extension to higher dimensions, tends to be more familiar for economics students than the picture evoked by figures like 2. Fig. 1 is a graphical representation of the prices p_1, p_2, and it is common practice to call it a graphical representation in the "*space* of prices", or, more generally, in the "*space* of variables" of the problem. However, we will use the word *space* exclusively to refer to *vector spaces*, so it is convenient to use a different word, *manifold*, for the space of variables of our problem.

The word manifold is used in geometry to refer to "smooth" surfaces that are *locally* Euclidean. For example, a small enough patch of the Earth's curved surface can be considered as (part of) a flat two dimensional plane. In this paper we will be working with spaces that are also globally Euclidean (flat), so we will not be concerned for the moment with subtle definitions.

The objective of this section is to elevate the status of the manifold of variables to something more important than the simple graphical representation of the *space* of variables. This is a first necessary step towards a more intuitive understanding of linear structures.

The variables of economic problems may have restrictions in the values they can take. For example, if the variables are quantities of a given product, they usually are nonnegative numbers. However, we will assume in this paper that they can have any value from $-\infty$ to ∞. In fact, in finance, it is common practice to talk about negative quantities: to have, say, "-300 stocks of a company", corresponds to having a *short* position on that company, or having sold those stocks without previously having them.

Correspondingly, the price of a short position in a given stock is negative: when someone enters into a short position she receives money, i.e. her cost is negative, she sells the underlying product (without having it) therefore she receives, rather than pays, money. In that sense the price of a short position is negative. Of course, eventually, she has to cover the short position.

In any case, if the problem demands it, we will eventually learn to impose constraints on the values of variables. For example, the constraint that such and such variable should be non-negative. For the moment, however, let us assume that our variables can go from $-\infty$ to $+\infty$.

Consider, to ease the visualization, only two variables, say two prices, p_1 and p_2, and assume, as explained above, that they can be positive or negative. Let us compare the *manifold* of prices (p_1, p_2), understood as a simple visualization tool for economic variables, with a *real* two dimensional plane, say a piece of paper, or a smooth piece of land. Let us assume that the piece of land is big enough for the borders to be irrelevant when considering translations, and small enough to ignore the spherical nature of earth's surface. That is, let us consider what most people would call a "real plane". What are the differences between the real plane and the price manifold?

A first difference we can point out is that, while in the plane all points are equivalent, i.e. the plane is *homogeneous*, in the manifold of prices there is a *natural origin*, or special point: the point $p_1 = 0, p_2 = 0$, corresponding to *free* goods, see Fig. 3. In the real plane we can, and will, choose an arbitrary point and call it the "origin". But it is an *arbitrary* choice. In the manifold of

(a) Real space: no special points (homogeneous). (b) Manifold of prices: $(0, 0)$ is a special point.

Figure 3: Comparison between the homogeneous real space and the manifold of prices.

prices, the $p_1 = 0$, $p_2 = 0$ point is the *natural* origin.

A second difference is even deeper. We intuitively understand that the points of a real plane exist independently of how we name them. Naming them with a pair of real numbers is, in a sense, arbitrary[3]. On the other hand, in the case of prices, what really exists is the numbers, namely, the prices themselves. The assignments of points to these prices (the manifold), is simply a visualization device.

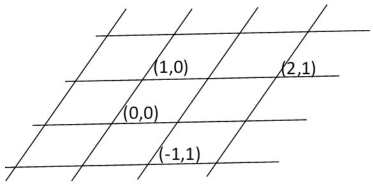
It is worth to pause and ponder about the above paragraph. One needs to adjust to the idea that the manifold exists independently from the numbers we use to describe them, like the real space exists completely independently from any Cartesian coordinate one can use to localize objects. Moreover, at least as an intellectual exercise, one should give *ontological primacy* to the manifold[4], and consider the numbers just as arbitrary names for its points. If one does that, then a vast reservoir of extremely useful, deeply held spatial intuitions becomes almost self-evident.

Giving ontological primacy to manifolds over numbers is not natural when learning math coming from economics. In economics what exists is the values of the variables of our models, the manifold being for most economics students just an ad-hoc visualization device. But still, I encourage economics students to think about the act of conceding ontological primacy to the manifolds over numbers simply as part of a “technology" to increase even more your mathematical visualization capacity.

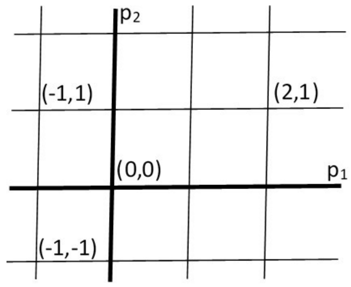
Think about it, when people try to develop technologies such as augmented reality, they go into the details of our visualization system to seamlessly merge the perception of real objects with the *augmented* parts. Similarly, it may help economists to use “devices" that enhance their ability to see the mathematical objects used in their models. Although our brains’ ability to develop intuitions in different areas of knowledge is well documented, not all intuitions are “created equal". For evolutionary and acquired reasons, few intuitions are as deeply held as our spatial

[3]On second thoughts, the assignments of a pair of real numbers to points of a plane, or, even simpler, the relationship of real numbers and points in a line, should produce awe: how can two ontologically different things, namely, points of a line and real numbers, share properties so deeply as to blur the difference between them? In fact, mathematicians use the real numbers to *define* the line. In this paper, however, to make more concrete the notion of manifold, I suggest the reader to think about a *real* plane, or the space of our experience.

[4]Giving ontological primacy to the manifold simply means thinking about the *existence* of the manifold as completely independent from the existence of numbers, exactly as one intuitively thinks about the real space of our experience. And thinking about the numbers as simple names for the points of the manifold.



(a) The ontological primacy of the manifold makes it obvious the arbitrariness of the assignments of numbers to points, i.e., the election of a coordinate system.



(b) Manifold of prices: “natural" assignments of points to numbers.

Figure 4: Comparison between real space and the manifold of prices: 2.

intuitions. Therefore, think of developing the ability to switch in your mind from the ontological primacy of the variables of your model towards manifolds and vice versa as one step in the development of “math visualization technology".

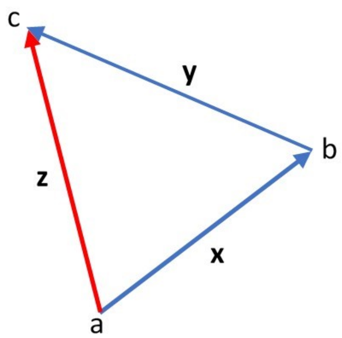
Let us consider again a plane. Assuming that the plane exists independently of how we name its points makes it obvious that the actual assignment of numbers to points of the plane, that is, the choice of a *coordinate system*, can be done in arbitrary ways, as the one in Fig 4a. In fact, if the manifold is the real thing, the assignment of numbers to points doesn't even have to be linear. However, we leave nonlinear assignments for a different work.

On the contrary, if we think of prices as the only real thing and relegate the manifold to just a basic visualization device, there is no such arbitrariness, and we will simply visualize the price manifold in the standard way of two perpendicular axes as in Fig 4b.

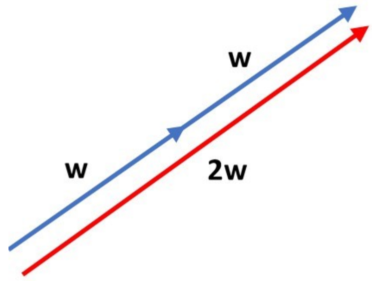
They are like two “mathematical visualization technologies", but one is version 1.0 and the other 2.0. The 2.0 version provides vastly more freedom to adapt the technology to your problem. As we will see when studying linear mappings of vector spaces, this freedom to select the most convenient “coordinate system" is very useful to visualize what matrices really are.

4 Vector Spaces

In the previous section we developed the ability to think of manifolds as existing independently of the values of the variables of our economic problem, just as we think of real space as independent of any coordinate system. But focus again on the real space: we infer most of its properties through *displacements*. Let us explore this notion of inferring the properties of space through displacements in an intuitive way first, and abstractly later.



(a) Different displacements to go from a to c.



(b) A sequence of two identical displacements is equivalent to a displacement in the same direction but twice as large.

Figure 5: Properties of displacements.

4.1 Intuition behind vectors

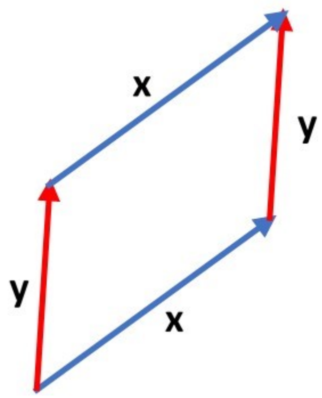
Let us choose an arbitrary point as both the origin in the manifold and the starting point of our displacements. If we characterize displacements only by their starting and final points, without concern for the intermediate points, it is clear that there is a one-to-one relationship between the points in the manifold and the displacements that start at the origin and end in the chosen point. We can represent such displacements as straight arrows that start at the origin and end at the chosen point. Let us see how most of our intuitions about space are derived from displacements like these.

The space of our everyday experience is 3 dimensional, although most of the time we move on a flat two-dimensional surface. We have an intuitive notion of dimension, at least from 1 to 3, associated to displacements in space: we intuitively understand that if we move in one direction, no matter how far we go, we will never reach an object that is in another direction.

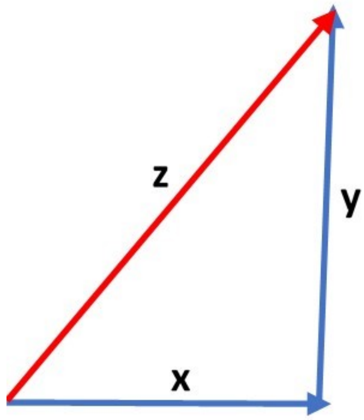
We also have an intuitive notion of distance. Our 3-D visual system allows us to perceive if one object is closer than another. And if the two objects are at similar distances, a compass will allow us to know which object is closer.

We also intuitively understand that, if the objective is going from a to c (see Fig. 5a), doing first a displacement $\mathbf{x}$, and after that a displacement $\mathbf{y}$, does the same as the direct displacement $\mathbf{z}$. We know that if we actually move in a plane with no obstacles, first by $\mathbf{x}$, and then by $\mathbf{y}$, the whole path takes more effort than displacement $\mathbf{z}$. But if we only care about the starting and final points, $\mathbf{x}$ followed by $\mathbf{y}$ represents the same displacement as $\mathbf{z}$. Therefore, we could say "$\mathbf{z} = \mathbf{x}$ followed by $\mathbf{y}$".

Doing first a displacement $\mathbf{w}$, and after that another displacement $\mathbf{w}$, with identical direction and length, would be equivalent to a net displacement twice as large and in the same direction (see Fig. 5b). And doing first a displacement $\mathbf{w}$, and after that another displacement of identical length and *opposite* direction, one would return to the starting point. It is natural to call $-\mathbf{w}$ the second displacement.



(a) Commutativity of displacements.



(b) Pythagoras theorem for perpendicular displacements.

Figure 6: Properties of displacements.

Perhaps not so intuitive is the validity of the *commutative* property of displacements: if moving on a plane surface with no obstacles, doing first a displacement **x** and then a displacement **y**, you reach the same point than if you do first a displacement **y** and then a displacement **x**, see Fig. 6a.

The commutativity of displacements may not be as intuitively obvious as the other properties. But take two identical blue sticks and another two identical red sticks, not necessarily of the same length as the blue sticks, and join them in the extremes like in Fig. 6a, allowing the angle between two adjacent sticks to freely change. Now it is immediately obvious that, no matter what the angle between adjacent sticks is, the sequence red-blue will lead you from one extreme to the opposite extreme, exactly as the sequence blue-red.

HW 4.1: By drawing, convince yourself of the validity of the *associative* property of displacements: doing a displacements **d**, that is the resultant of doing first a displacement **a** and then a displacement **b**, and after **d** doing a displacement **c**, is equivalent to doing a displacement **a** followed by a displacement **e**, that is the resultant of doing first a displacement **b** and then a displacement **c**.

Finally, although it is at first not obvious intuitively, we learn in primary school Pythagoras' theorem. In terms of displacements, if one displacement **x** is followed by a *perpendicular* displacement **y**, then the square of the length of the resulting displacement **z** equals the square of the length of **x** plus the square of the length of **y**, see Fig. 6b.

Notice that the above properties go from completely intuitive, to mildly intuitive, to not intuitive (at least at first sight). In the completely intuitive camp we include, for example, a basic dimensionality notion, composition of displacements (doing one displacement first and then another) and the notion that two different sequences of displacements can lead from the same initial point to the same point (Fig. 5a), so that if one only cares about initial and final points, then the two sequences of displacements are equal. It is also completely intuitive to compare the length of two displacements starting from the same point and both pointing in the same direction.

In the mildly intuitive camp, we include for example the commutativity and associativity of displacements.

Lastly, in the “not intuitive at first sight" camp, we include the notion of distance traveled when the sequence includes displacements in different directions. In particular, Pythagoras’ theorem.

Before we formalize all this, I suggest another mental “ontological exercise". As an economics or data science student, you probably tend to think of vectors as numbers ordered in columns (or rows), and visualize them in a Cartesian coordinate representation. But here we have introduced vectors as *displacements*. We have made that choice for many reasons, a very important one is that, if you think of vectors as displacements, they obviously have an existence independent of the numbers you use to name them. So, as we did with the manifold, another ingredient of our visualization technology 2.0 is to think of vectors (or displacements) as ontologically independent from its numerical (Cartesian or otherwise) representation.

Finally, I would like to summarize that we have defined two categories of objects: manifolds, as an extension of the intuitive notion of space, and vectors, as an extension of the notion of displacements in that space. If we select in the manifold an arbitrary point O as our origin, for any other point A in the manifold, if we only care about starting and final points, we can find one and only one vector (or displacement) $\mathbf{a}$ that starts at the origin O and ends at A. Moreover, this one-to-one relationship can also include the origin itself if we decide to include, among our set of vectors, the null vector $\mathbf{0}$, that starts and ends at the origin.

4.2 Abstract vector spaces

Before presenting the abstract definition of vector spaces, let us give the intuition behind its defining operations. The “sum" between vectors, $\mathbf{z} = \mathbf{x} + \mathbf{y}$, is an operation that returns another vector, and amounts to the concatenation of the two vectors $\mathbf{x}$ and $\mathbf{y}$. It coincides with the intuition of $\mathbf{z}$ as a displacement equal to doing first a displacement $\mathbf{x}$ followed by a displacement $\mathbf{y}$ in Fig. 5a, if one only cares about initial and final points.

The “product" between a real number a and a vector $\mathbf{w}$, $\mathbf{r} = a\,\mathbf{w}$, intuitively scales the length of the vector $\mathbf{w}$ by a factor of $|a|$, keeping the same direction as $\mathbf{w}$ if $a > 0$, and the opposite direction if $a < 0$, see Fig. 5b. Since this operation *scales* the vectors, we will sometimes refer to real numbers like a as “scalars."

The “visualization technology 2.0" that I am trying to present to you works best when one is able to switch at will between an intuitive mode of thinking and a rational, abstract mode. While reading the definition of vector spaces and the axioms of their operations, I suggest to switch to your abstract mode of thinking, suppressing for a moment the intuitions about vectors presented above. Think of them instead, for a moment, as abstract objects and operations defined only by the axioms.

Definition of vector space: a vector space V is a collection of vectors where two operations are defined. A *sum*, with two vectors as input and a vector as output, and a *product*, with a real number and a vector as input and a vector as output[5]. These operations satisfy the following axioms.

Axioms of the sum between vectors:

- Associativity: $(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c})$.
- Commutativity: $\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}$.
- Identity element: there exist a vector $\mathbf{z} \in V$ such that for any other vector $\mathbf{a}$, $\mathbf{a} + \mathbf{z} = \mathbf{a}$. It is customary to call to this vector the zero vector: $\mathbf{0}$.
- Inverse element: for every vector $\mathbf{a} \in V$, there is another vector $\mathbf{d} \in V$ such that $\mathbf{a} + \mathbf{d} = 0$. It is customary to call $-\mathbf{a}$ to the inverse vector of $\mathbf{a}$.

Axioms of the product between a real number and a vector:

- Compatibility of scalar multiplication with number multiplication: $a(b\mathbf{v}) = (ab)\mathbf{v}$.
- Identity element of scalar multiplication: $1\mathbf{v} = \mathbf{v}$, where 1 is the real number one.
- Distributivity of scalar multiplication with respect to vector addition: $a(\mathbf{v} + \mathbf{w}) = a\mathbf{v} + a\mathbf{w}$.
- Distributivity of scalar multiplication with respect to real numbers addition: $(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$.

Abstract mathematical structures, and the interplay between abstractions and intuitions, are incredibly useful for practical problems. Sometimes it is possible to map a problem A into another problem B, only to find that, at the abstract level, they are the same problem. So, a proven property of problem B automatically becomes a proven property of problem A. The mapping of problem A into problem B is specially useful when problem B is very visual, like ideal displacements in 1, 2 and 3 dimensional Euclidean spaces are.

Another way in which abstractions are useful is that sometimes they allow us to naturally create new mathematical objects, like spaces with arbitrary dimensions, that later turn out to be of practical use in unexpected ways. Moreover, since these new mathematical objects share many of the abstract properties with the original one, if the original one is very visual, again, like ideal displacements in 1, 2 and 3 dimensional Euclidean spaces are, then, with a little practice you start developing intuition in many dimensions too.

HW 4.2: Consider ordered pairs of numbers and dispose them in a two-component column. a) Prove that all the axioms above are satisfied by the “natural" sum and product by scalars. b)

[5] A vector space can be defined over real numbers, over complex numbers, rational numbers, etc. In general, it can be defined over any *field*. However, most of our applications will involve real numbers, so the definition above is enough for our purposes.

Draw the arrows corresponding to two vectors, **v** and **w**, and the arrows corresponding to $\mathbf{v} + \mathbf{w}$ and to $1.5\mathbf{v}$ and convince yourself that these operations do exactly what we expect if we think of the arrows as displacements.

HW 4.3: Do the same for 3-component column vectors. By hand, or with your preferred software, draw in 3D the corresponding arrows. Convince yourself, now in 3-D, that these operations do exactly what we expect, if we think of the arrows as displacements[6].

HW 4.4: Prove that the set of column vectors of n real numbers, with the natural sum between vectors and product by a scalar, form a vector space (i.e., they satisfy the axioms).

Definition: We will call the vector space of n ordered real numbers R^n.

In the HWs above we insist in viewing vectors as displacements. However, it is natural to visualize, say, a two-component column vector, as the corresponding point in the plane itself, each component being the Cartesian coordinate of the corresponding point with respect to two axes perpendicular to each other. Similarly for a three-component column vector, as a point in space, etc. However, this would correspond to "visualization technology 1.0", not the 2.0 version we are presenting here. For this, it is very important to distinguish between the points of the manifold and the vector that starts at the origin and ends at the selected point. *We will associate n ordered numbers organized in columns (or in raws) with vectors, or, if you prefer, with displacements on the manifold,* ***not*** *with the points of the manifold themselves.*

Of course, once we have arbitrarily selected a point in the manifold and call it the *origin*, as already mentioned, there is a one-to-one relationship between points in the manifold and vectors. Moreover, for the sake of brevity, we will sometimes be sloppy and talk about "point **q**" instead of the more correct "point Q with associated vector **q**." But for the "visualization technology 2.0" to work properly, it is very important to keep in mind the distinction between points of the manifold and vectors of the vector space despite this on-to-one relationship.

4.3 Linear Independence, bases, linear subspaces, affine subspaces, etc.

In this section we present standard definitions and theorems on vector spaces. While reading this section, I recommend the student to activate the abstract mode of thinking for definitions and theorems (in particular, try to prove the theorems abstractly), and switch to the intuitive mode for the HWs that are not proofs of theorems.

Linear combination: given the vectors $\mathbf{v}^i \in V$, $i = 1, \cdots, n$, we call a linear combination of

[6] In all the HWs in which you have to graphically represent 3-D in a 2-D paper, the graph should be clear. If you find it difficult to do these graphs by hand at the beginning, do them with your preferred software. But you should learn to visualize 3-D graphs in your mind and in the plane. If you learn to visualize in 3-D, many extensions to n-D, helped by the mathematical machinery we are developing, will be trivial. Visualizing in 2-D only, is not enough to make the leap to n-D.

these vectors to an expression like:

$$\mathbf{y} = \sum_{i=1}^{n} a_i \mathbf{v}^i \tag{4.1}$$

In the above expression, different upper indices i, as in $\mathbf{v}^i$, correspond to different vectors, while different lower indices i, as in a_i, correspond to different scalars.

HW 4.5: Consider two ordered pairs of two numbers as vectors with the natural sum and product by scalars. a) Construct a linear combination of these two vectors and draw it to see its geometric meaning. b) Do the same for two ordered pairs of three numbers.

Linear dependence: a set of vectors $\mathbf{v}^i \in V$, $i = 1, \cdots, r$ is *linearly dependent*, if there is a set of scalars a_i, $i = 1, \cdots, r$, not all of them zero, such that

$$\sum_{i=1}^{r} a_i \mathbf{v}^i = \mathbf{0} \tag{4.2}$$

If the only way to make the sum above equal to zero is with all $a_i = 0$, then the vectors $\mathbf{v}^i$ are *linearly independent*.

The intuition here is that, if there are scalars a_i, $i = 1, \cdots, r$, not all of them zero such that (4.2) holds, suppose, for example, that $a_1 \neq 0$, then

$$\mathbf{v}^1 = \sum_{i=2}^{r} \frac{a_i}{a_1} \mathbf{v}^i \tag{4.3}$$

i.e., at least one of them is a linear combination of the others.

HW 4.6: a) Write two linearly dependent two-component column vectors. How do you draw them in a plane? b) Do the same for three linearly dependent three-component column vectors (two of them must be linearly independent, none of them should be a scaled version of one of the other two).

Span of a set of vectors: given a set of vectors $W = \{\mathbf{w}^i \in V, i = 1, \cdots, n\}$, the set of all vectors of the form

$$\mathbf{v} = \sum_{i=1}^{n} a_i \mathbf{w}^i \tag{4.4}$$

for arbitrary a_i, is called the span of W.

HW 4.7: a) Consider the span of a single two-component column vector. What does it correspond to graphically? b) Do the same for two linearly independent two-component column vectors. c) Do the same for 1 and 2 linearly independent three-component column vectors. d) Write 3 linearly *dependent* three-component column vectors such that two of them are linearly independent and none of them is a scaled version of one of the other two. Consider now the span of these 3 vectors, what does it correspond to geometrically? e) Do the same for 3 linearly *independent* three-component column vectors.

Basis of V: If a set of vectors W span V, then W is a *basis* of V.

HW 4.8: a) Write two vectors forming a basis of the vector space of two-component column vectors. b) Do the same for a basis of 3 vectors of the vector space of two-component column vectors. c) Do the same for a basis of 3 vectors of the vector space of three-component column vectors. d) Do the same for a basis of 5 vectors of the vector space of three-component column vectors.

Linearly independent basis of V: If a set of linearly independent vectors W span V, then W is a linearly independent basis of V.

Theorem 4.1 *If $W = \{\mathbf{w}^i \in V,\ i = 1, \cdots, n\}$ is a linearly independent basis of V, then every vector $\boldsymbol{v} \in V$ can be written as a linear combination of the $\mathbf{w}^i$s in a unique way.*

HW 4.9: Prove the theorem.

HW 4.10: Prove that R^n, for every $n > 0$, has a linearly independent basis.

HW 4.11: Prove that any linearly independent basis of R^n, has exactly n components.

Dimension of a vector space: The number of vectors of *any* linearly independent basis of a vector space V is the same. This number is by definition the *dimension* of V.

Theorem 4.2 *If a set of n vectors is a basis of V, then any set of $n + 1$ vectors is linearly dependent.*

HW 4.12: Prove the theorem.

Dimension of the manifold associated to a vector space V: The dimension of V is also the dimension of the manifold corresponding to V.

Theorem 4.3 *If V has dimension n, then a linearly independent set of n vectors in V is a basis of V.*

HW 4.13: Prove the theorem.

Linear subspaces of a vector space V: If a vector space V has dimension n, then the span W of any set of vectors in V is a subspace of V with dimension $\leq n$.

Corollary 1: The vector $\mathbf{0}$ belongs to any subspace of V.

Through the one-to-one relationship between manifolds and vector spaces, a linear subspace corresponds to a linear manifold of smaller dimension than the manifold in which it is embedded that passes through the origin.

Affine subspaces of a vector space V: The set S of all vectors of the form

$$\mathbf{r} = \mathbf{a} + \mathbf{w} \tag{4.5}$$

where $\mathbf{w} \in W$, a subspace of V, and $\mathbf{a} \neq 0$ is a fixed vector in V, is an *affine* subspace of V.

Geometrically, an affine subspace corresponds to a linear manifold, of smaller dimension than the manifold in which it is embedded, that does not (in general) contain the origin.

4.4 Back to intuitive thinking

Enough definitions, theorems, and purely abstract thinking for the moment. Let us return to the intuitions of subsection 4.1, trying to find the mathematical expression for these intuitions. We go from the completely intuitive, to the mildly intuitive, to the non-intuitive.

Completely intuitive properties:

1. **Basic dimensionality notion:** we mentioned that we intuitively understand that “if we move in one direction, no matter how far we go, will never reach an object that is in another direction."

   Let us model this mathematically with two-component column vectors: chose two linearly independent two-component column vectors $\mathbf{x}^1$ and $\mathbf{x}^2$. Consider the span of $\mathbf{x}^1$. It represents “moving in one direction (the direction of $\mathbf{x}^1$) as far we wish". $\mathbf{x}^2$, that is linearly independent of $\mathbf{x}^1$, represents the position of “an object that is in another direction". The fact that $\mathbf{x}^2$ is not in the span of $\mathbf{x}^1$ (because it is linearly independent), means that “if we move in one direction (the direction of $\mathbf{x}^1$), no matter how far we go, will never reach an object (located at $\mathbf{x}^2$) that is in another direction".

   We see how vector spaces capture mathematically exactly our intuitive notion of dimensionality, or independent directions. Do the corresponding graph.

   HW 4.14: Explain the same intuition with two linearly independent three-component column vectors, $\mathbf{x}^1$ and $\mathbf{x}^2$, whose span represents our possible displacements, say, on the surface of the earth. However, “no matter how far we go", will never reach an object located at $\mathbf{x}^3$ if $\mathbf{x}^3$ is linearly independent of the other two (say, it is ten meters high).

HW 4.15: Construct a similar story with n-component column vectors.

2. Composition of displacements: with column vectors and the corresponding graphs, we already showed that the sum of vectors correspond to compositions of displacements in 2-D and 3-D.

3. Length comparison between two vectors, $\mathbf{x}^1$ and $\mathbf{x}^2$, that start from the same point and both point in the same direction.

   HW 4.16: How do know which one is longer mathematically? Explain.

The mildly intuitive properties of associativity and commutativity have already been analyzed.

Non-intuitive at first sight properties:

HW 4.17: Select two linearly independent two-column vectors, spend 5 to 10 minutes (no more than that) trying to prove abstractly (this means, no graphs allowed, and *using only operations defined in the axioms*), whether these two vectors are **orthogonal** or not. Please do not turn the page until you try to solve this problem in the allotted time.

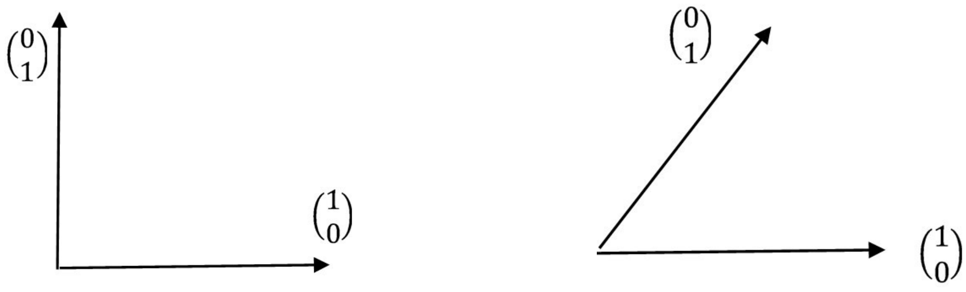


Figure 7: The axioms of vector spaces do not have a way to conceptualize whether two vectors are orthogonal. The graphical difference between 7a and 7b does not have a mathematical counterpart yet.

It turns out that you can't prove it...

> “...But wait!...", you are probably thinking, “...why can't I say, for example, that $\mathbf{x}^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{x}^2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ are orthogonal? As a matter of fact, we have been using their orthogonality all along in the graphical representations of the previous HWs! What is wrong with drawing them as two perpendicular arrows in a standard Cartesian coordinate system like Fig. 7a? Isn't there an obvious angle of 90° between them? how come they are not orthogonal?..."

Pause and ponder about the meaning of “proving something using only operations defined in the axioms". *Within the given axioms you can't even define what orthogonality means*. The closest thing you can prove is the weaker notion of linear independence between two vectors, which, as we have seen, is enough to define dimensionality. But it is not enough to define orthogonality.

This is an important lesson, it teaches us that if we really want to build and enhanced mathematical visualization system that draws on our deeply held spatial intuitions, for every one of these visual intuitions there has to be a corresponding mathematical conceptualization. Only if we have this tight relationship between intuitions and mathematical concepts, are we going to be able to extend these intuitions safely to spaces beyond our experience.

As it turns out, we still don't have the mathematical concept associated with such an important property of real space as the notion of orthogonality. This, of course, implies that we don't yet have the mathematical concept associated to the corresponding intuition.

To gauge the importance associated to a lack of mathematical concepts for orthogonality, note that a crucial ingredient of our spatial intuitions is the notion of *distance* in 2-D and 3-D. A mathematical codification of a notion of distance that mimics the properties of *real* distance between objects should be consistent with what we know about distance since primary school: in particular, Pythagoras' theorem[7]. But a prerequisite to safely apply that theorem is that two

[7]In 1-D this is not required, so a 1-D real vector space inherits the notion of distance of the real numbers.

sides of the triangle should be orthogonal to each other. So we see that in order to have a notion of distance that mimics the distance in real space we need first a notion of orthogonality![8]

This means that if we want to endow vector spaces and the underlying manifold with a notion of distance that mimics the real distance, we should extend our axioms to be able to tell, using only operations defined in the axioms, whether two vectors are orthogonal or not. We deal with this in the next section.

5 "Dot" (or scalar) product vector spaces

As we have seen, vector spaces do not have enough structure to define orthogonality. For that we need to incorporate an additional structure.

In addition to a sum, a linear operation that inputs two vectors and returns another vector, and a product, an operation that inputs a real number and a vector and returns another vector, we define now a "dot", or "scalar", product. It is an operation that inputs two vectors and returns a real number, or scalar, for which we use the notation "$\mathbf{x} \cdot \mathbf{y}$", and satisfies the following axioms.

Axioms of the dot, or scalar, product $\mathbf{x} \cdot \mathbf{y}$:

- Symmetry:
$$\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$$

- Linearity in the second argument:
$$\begin{aligned} \mathbf{x} \cdot (a\,\mathbf{y}) &= a\,(\mathbf{x} \cdot \mathbf{y}) \\ \mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) &= \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z} \end{aligned}$$

- Positive definite:
$$\begin{aligned} \mathbf{x} \cdot \mathbf{x} &\geq 0 \\ \mathbf{x} \cdot \mathbf{x} &= 0 \quad \Leftrightarrow \quad \mathbf{x} = \mathbf{0} \end{aligned}$$

HW 5.1: Prove that these axioms imply linearity in the first argument as well.

Since the dot product is linear in both the first and the second argument, it is said to be "bi-linear".

We could continue abstractly, but I think it is best to show a realization of this dot product for column vectors and eventually generalize the notation.

[8] Metric spaces (in which a distance is defined, but not yet a dot product) are in fact more general than inner product spaces. However, for the moment my intention is simply to define a distance that mimics the distance of real space, so I prefer not to open this discussion here.

HW 5.2: Prove that, if we have two n components column vectors, $\mathbf{x}$, with components x_i, and $\mathbf{y}$, with components y_i, then, the operation

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i \tag{5.1}$$

satisfies the axioms of the scalar product.

Now we can *define* orthogonality:

Definition of orthogonality: In a scalar product space V, two non-zero vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be *orthogonal* to each other if and only if $\mathbf{x} \cdot \mathbf{y} = 0$.

With this notion, we can finally distinguish between (7a) and (7b).

Let us use a slightly more abstract notation that will allow us to appreciate the generality of the concepts defined here. Call $\hat{\mathbf{e}}^i$ to the n-component column vector whose elements are all zero except the i^{th} component that is equal to 1.

HW 5.3: Prove that

$$\hat{\mathbf{e}}^i \cdot \hat{\mathbf{e}}^j = \delta^{ij} \tag{5.2}$$

where the symbol δ^{ij} is defined by

$$\delta^{ij} = \begin{cases} 1, & \text{if} \quad i = j \\ 0, & \text{if} \quad i \neq j \end{cases} \tag{5.3}$$

and is called the “Kronecker delta".

Any column vector can be written as

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^{n} x_i \, \hat{\mathbf{e}}^i \tag{5.4}$$

and the bi-linearity of the scalar product allows us to compute it as

$$\begin{aligned} \mathbf{x} \cdot \mathbf{y} &= \left(\sum_{i=1}^{n} x_i \, \hat{\mathbf{e}}^i \right) \cdot \left(\sum_{j=1}^{n} y_j \, \hat{\mathbf{e}}^j \right) && (5.5) \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_j \, \hat{\mathbf{e}}^i \cdot \hat{\mathbf{e}}^j \\ &= \sum_{i=1}^{n} \sum_{j=1}^{n} x_i y_j \, \delta^{ij} \\ &= \sum_{i=1}^{n} x_i y_i && (5.6) \end{aligned}$$

(5.6) is identical to (5.1).

HW 5.4: Justify each step from (5.5) to (5.6).

Definition of orthogonal and orthonormal basis:

Orthogonal basis: A basis $\{\mathbf{x}^i\}$, $i = 1, \cdots, n$, of an n-dimensional scalar product space V, is *orthogonal*, if

$$\mathbf{x}^i \cdot \mathbf{x}^j = 0 \quad \text{for } i \neq j \tag{5.7}$$

and different from zero when $i = j$.

Orthonormal basis: A basis $\{\hat{\mathbf{e}}^i\}$, $i = 1, \cdots, n$, of an n dimensional scalar product space V, is *orthonormal*, if

$$\hat{\mathbf{e}}^i \cdot \hat{\mathbf{e}}^j = \delta^{ij} \quad \text{for } i, j = 1, \cdots, n \tag{5.8}$$

Given an orthogonal basis $\{\mathbf{x}^i\}$, $i = 1, \cdots, n$, one can construct the corresponding orthonormal basis $\{\hat{\mathbf{e}}^i\}$, $i = 1, \cdots, n$ as:

$$\hat{\mathbf{e}}^i = \frac{\mathbf{x}^i}{(\mathbf{x}^i \cdot \mathbf{x}^i)^{1/2}} \quad \text{for } i = 1, \cdots, n \tag{5.9}$$

HW 5.5: Prove that the basis $\{\hat{\mathbf{e}}^i\}$, $i = 1, \cdots, n$, defined in (5.9) is orthonomal if the basis $\{\mathbf{x}^i\}$, $i = 1, \cdots, n$ is orthogonal.

HW 5.6: Write an orthonormal basis in the space of column vectors of dimension 5.

Now we can define the magnitude of a vector, and the length in the underlying manifold.

Definition of magnitude, or norm, of a vector: The *magnitude* of a vector $\mathbf{x}$, for which we will use the symbol $||\mathbf{x}||$, is defined by:

$$||\mathbf{x}|| = (\mathbf{x} \cdot \mathbf{x})^{1/2} \tag{5.10}$$

If $\mathbf{x}$ is written in terms of an orthonormal basis like (5.8),

$$\mathbf{x} = \sum_{i=1}^{n} x_i \, \hat{\mathbf{e}}^i \tag{5.11}$$

the square of the norm is

$$||\mathbf{x}||^2 = \mathbf{x} \cdot \mathbf{x} = \sum_{i=1}^{n} x_i^2 \tag{5.12}$$

this is just the extension to n dimensions of Pythagoras' theorem. We can appreciate how, when orthogonality is defined through the scalar product, Pythagoras' theorem, which requires orthogonality, emerges naturally.

See Fig. 8 for the particular case of equation (5.12) in 3 dimensions. Note that the red arrow is

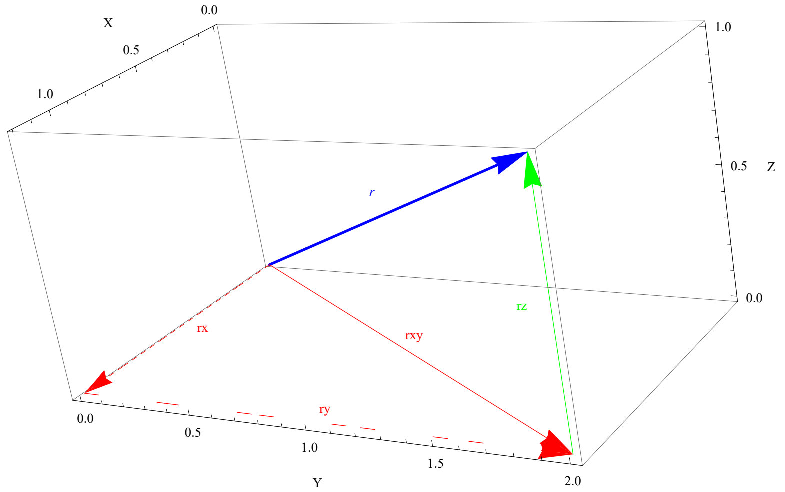


Figure 8: Pythagoras in 3-D: $||\mathbf{r}||^2 = ||\mathbf{r_{xy}}||^2 + ||\mathbf{r_z}||^2$. But, $||\mathbf{r_{xy}}||^2 = ||\mathbf{r_x}||^2 + ||\mathbf{r_y}||^2$. Therefore: $||\mathbf{r}||^2 = ||\mathbf{r_x}||^2 + ||\mathbf{r_y}||^2 + ||\mathbf{r_z}||^2$.

the projection of the blue arrow on the (x, y) plane, and the green arrow is the projection of the blue arrow on the z axis, orthogonal to the (x, y) plane. In the two-dimensional plane determined by these three arrows, the standard two-dimensional Pythagoras' theorem works as usual. But if we want to decompose the red arrow into its x and y components, applying again Pythagoras' theorem to the red arrow we finally arrive at the expression $||\mathbf{x}||^2 = \sum_{i=1}^{3} x_i^2$ corresponding to (5.12). Exactly the same happens in n dimensions.

For any two vectors $\mathbf{x}^1$ and $\mathbf{x}^2$ in a scalar product vector space V, the norm satisfies two important inequalities:

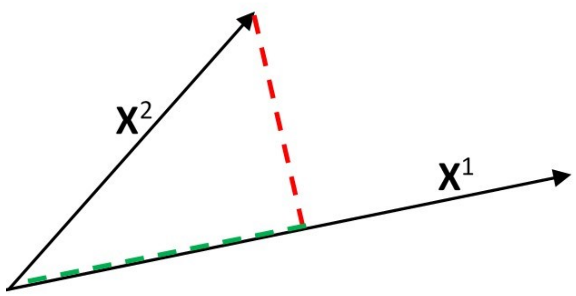


Figure 9: The green dashed line is the projection of $\mathbf{x}^2$ in the direction of $\mathbf{x}^1$, and the red dashed line is the projection orthogonal to $\mathbf{x}^1$.

Cauchy–Schwarz inequality: $\left(\mathbf{x}^1 \cdot \mathbf{x}^2\right)^2 \leq \left(\mathbf{x}^1 \cdot \mathbf{x}^1\right)\left(\mathbf{x}^2 \cdot \mathbf{x}^2\right)$, or, taking the square root on both sides:

$$|\mathbf{x}^1 \cdot \mathbf{x}^2| \leq ||\mathbf{x}^1||\,||\mathbf{x}^2|| \tag{5.13}$$

Proof: Consider the vector

$$\mathbf{x}^{2\perp 1} = \mathbf{x}^2 - \left(\frac{\mathbf{x}^2 \cdot \mathbf{x}^1}{\mathbf{x}^1 \cdot \mathbf{x}^1}\right)\mathbf{x}^1 \tag{5.14}$$

$\mathbf{x}^{2\perp 1}$ is the component of $\mathbf{x}^2$ orthogonal to $\mathbf{x}^1$, the red dashed line in Fig. 9. To see this note that $\mathbf{x}^{2\perp 1} \cdot \mathbf{x}^1 = 0$ and

$$\mathbf{x}^2 = \mathbf{x}^{2\perp 1} + \left(\frac{\mathbf{x}^2 \cdot \mathbf{x}^1}{\mathbf{x}^1 \cdot \mathbf{x}^1}\right)\mathbf{x}^1 \tag{5.15}$$

Then,

$$\begin{aligned} ||\mathbf{x}^2||^2 &= \mathbf{x}^2 \cdot \mathbf{x}^2 \\ &= \left(\mathbf{x}^{2\perp 1} + \left(\frac{\mathbf{x}^2 \cdot \mathbf{x}^1}{\mathbf{x}^1 \cdot \mathbf{x}^1}\right)\mathbf{x}^1\right) \cdot \left(\mathbf{x}^{2\perp 1} + \left(\frac{\mathbf{x}^2 \cdot \mathbf{x}^1}{\mathbf{x}^1 \cdot \mathbf{x}^1}\right)\mathbf{x}^1\right) \\ &= ||\mathbf{x}^{2\perp 1}||^2 + \left(\frac{\mathbf{x}^2 \cdot \mathbf{x}^1}{\mathbf{x}^1 \cdot \mathbf{x}^1}\right)^2 \mathbf{x}^1 \cdot \mathbf{x}^1 \\ &\geq \frac{\left(\mathbf{x}^2 \cdot \mathbf{x}^1\right)^2}{\mathbf{x}^1 \cdot \mathbf{x}^1} = \frac{\left(\mathbf{x}^2 \cdot \mathbf{x}^1\right)^2}{||\mathbf{x}^1||^2} \end{aligned} \tag{5.16}$$

Multiplying both sides by $||\mathbf{x}^1||^2$ and taking the square root we arrive at (5.13).

The proof makes clear that the equality holds when $||\mathbf{x}^{2\perp 1}||^2 = 0$, i.e., when $\mathbf{x}^2$ does not have a component orthogonal to $\mathbf{x}^1$. If $\mathbf{x}^2$ does not have a component orthogonal to $\mathbf{x}^1$, it means that $\mathbf{x}^2$ is linearly dependent on $\mathbf{x}^1$ (or either of them are zero, a particular case of linear dependence).

HW 5.7: Justify each step in the proof (5.16).

Triangle inequality: $||\mathbf{x}^1 + \mathbf{x}^2|| \leq ||\mathbf{x}^1|| + ||\mathbf{x}^2||$.

This inequality quantifies the qualitative idea described in Fig. 5a, that the length along path "$\mathbf{x}$ then $\mathbf{y}$" (equal to $||\mathbf{x}|| + ||\mathbf{y}||$) is longer than the length of the direct path $\mathbf{z}$ (equal to $||\mathbf{x} + \mathbf{y}||$). The only exception being if $\mathbf{y}$ points in the same direction than $\mathbf{x}$, as in Fig. 5b. Let us prove the inequality.

Proof:

$$\begin{aligned} ||\mathbf{x}^1 + \mathbf{x}^2||^2 &= \left(\mathbf{x}^1 + \mathbf{x}^2\right) \cdot \left(\mathbf{x}^1 + \mathbf{x}^2\right) \\ &= ||\mathbf{x}^1||^2 + ||\mathbf{x}^2||^2 + 2\,\mathbf{x}^1 \cdot \mathbf{x}^2 \\ &\leq ||\mathbf{x}^1||^2 + ||\mathbf{x}^2||^2 + 2\,|\mathbf{x}^1 \cdot \mathbf{x}^2| \qquad (\mathbf{x}^1 \cdot \mathbf{x}^2 \leq |\mathbf{x}^1 \cdot \mathbf{x}^2|) \\ &\leq ||\mathbf{x}^1||^2 + ||\mathbf{x}^2||^2 + 2\,||\mathbf{x}^1||\,||\mathbf{x}^2|| \qquad (\text{by Cauchy˘Schwarz inequality}) \\ &= \left(||\mathbf{x}^1|| + ||\mathbf{x}^2||\right)^2 \end{aligned} \tag{5.17}$$

Taking the square root we finally prove the inequality.

Note that equality only holds if $|\mathbf{x}^1 \cdot \mathbf{x}^2| = \|\mathbf{x}^1\| \, \|\mathbf{x}^2\|$ and $\mathbf{x}^1 \cdot \mathbf{x}^2 = |\mathbf{x}^1 \cdot \mathbf{x}^2|$ are true. As we saw in the proof of the Cauchy–Schwarz inequality, the first condition only holds if $\mathbf{x}^1 = a\,\mathbf{x}^2$. And the second condition holds if, in addition, $a \geq 0$. This means that the equality holds in situations like in Fig. 5b, where the triangle of Fig. 5a becomes a straight path that never changes direction. This makes sense, if the triangle becomes a straight path that never changes direction, there is no difference between the "$\mathbf{x}^1$ then $\mathbf{x}^2$" path and the direct one $\mathbf{x}^1 + \mathbf{x}^2$. But note that we have proved much more: as long as a scalar product satisfying the appropriate axioms is defined, this deeply rooted intuition is valid also in vector spaces of *any* dimension.

Definition of distance between two points in the underlying manifold: Consider the points X and Y in a manifold in which an arbitrary origin has been chosen. We have pointed out many times that there is a one to one relation between points in the manifold and vectors. Assume that the vectors $\mathbf{x}$ and $\mathbf{y}$ are, respectively, associated to these points. Then, the distance between X and Y is:

$$\mathrm{dist}\,(X, Y) \equiv \|\mathbf{x} - \mathbf{y}\| \tag{5.18}$$

HW 5.8: a) Interpret in 2-D the geometrical meaning of the vector $\mathbf{x} - \mathbf{y}$. b) Convince yourself, and write the corresponding explanation, that the definition (5.18), that requires the definition of magnitude of a displacement (5.10), corresponds to our intuitive notion of distance in real space. c) Do a) and b) for 3-D.

We have now all the ingredients we need to extend our "visualization technology" to any dimension. Before we apply it to practical problems, let us summarize what was done until now:

1. The notion of vector space captures the idea of the set of all possible displacements in an underlying manifold.
2. The notion of linear independence corresponds to the notion of translations in different directions.
3. The notion of span of a set of vectors captures the idea of all possible displacements generated by concatenating and scaling the displacements included in the set. This naturally leads to subspaces of a vector space.
4. If the span of a set of vectors is the whole space, then the set includes displacements in all possible directions. In that case the set is called a basis of the space.
5. If the basis is linearly independent, intuitively, the number of vectors in the basis is counting the maximum number of independent directions in the space, which defines its dimension. This number is basis independent and therefore is an intrinsic property of the vector space.
6. Through the one-to-one relationship between displacements and points of the underlying manifold, this number also defines the dimension of the underlying manifold.

7. However, this is not enough to define orthogonality and distance. A scalar product is necessary for that.

8. The endowment of a vector space with a scalar product structure is equivalent to selecting a linearly independent basis, establishing by decree that the vectors of this basis are orthonormal, and establishing that the scalar product is bilinear.

9. Once the vector space is endowed with a scalar product, the norm of a vector can be defined satisfying Pythagoras' theorem.

10. The norm of vectors allows us to define a distance in the underlying manifold. This is fitting; distance is ultimately measured by displacements of an arbitrarily defined unit of length in real space.

11. The necessity of a notion of orthogonality to define a distance between two points mimicking the real distance in dimension greater than one, ultimately comes from the requirement of the generalized validity of Pythagoras' theorem.

12. One way to see the relevance of this requirement for the distance between two points in *any* dimension, even though the original Pythagorean theorem is valid for *two*-dimensional right triangles, is that the vectors corresponding to these two points span a two-dimensional subspace, and in this subspace the theorem applies.

5.1 Geometric meaning of the scalar product

Let us remember the definition of the cosine and sine functions. If we have a circle of radius r, the radius making an angle θ with the x-axis has a projection on the x-axis equal to $r\cos\theta$, and a projection on the y-axis equal to $r\sin\theta$, see Fig. 10.

Let us write this vectorially: if we name $\mathbf{r}$ the vector corresponding to the radius r making an angle θ with the x-axis, we have:

$$\mathbf{r} = ||\mathbf{r}||\cos\theta\,\hat{\mathbf{x}} + ||\mathbf{r}||\sin\theta\,\hat{\mathbf{y}} \tag{5.19}$$

where $\hat{\mathbf{x}}$ is the unit vector in the direction of the positive x-axis and $\hat{\mathbf{y}}$ is the unit vector in the direction of the positive y-axis. Of course $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are orthogonal to each other. Therefore $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ form an orthonormal basis on the plane:

$$||\hat{\mathbf{x}}|| = 1 \tag{5.20}$$
$$||\hat{\mathbf{y}}|| = 1 \tag{5.21}$$
$$\hat{\mathbf{x}}\cdot\hat{\mathbf{y}} = 0 \tag{5.22}$$

From (5.19), (5.20), (5.21) and (5.22) we get:

$$\mathbf{r}\cdot\hat{\mathbf{x}} = ||\mathbf{r}||\cos\theta \tag{5.23}$$
$$\mathbf{r}\cdot\hat{\mathbf{y}} = ||\mathbf{r}||\sin\theta \tag{5.24}$$

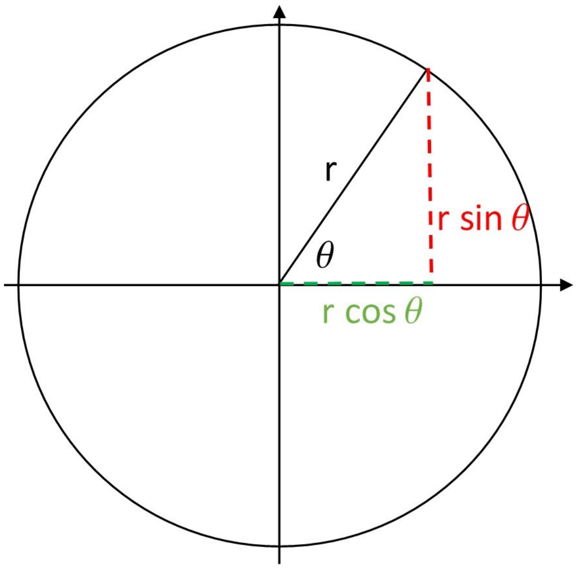


Figure 10: Definition of the functions $\cos\theta$ and $\sin\theta$.

i.e., the scalar product of $\mathbf{r}$ with the unit vector $\hat{\mathbf{x}}$ is the projection of $\mathbf{r}$ in the direction of $\hat{\mathbf{x}}$. This projection *defines* the cosine function. Similarly, the scalar product of $\mathbf{r}$ with the unit vector $\hat{\mathbf{y}}$ is the projection of $\mathbf{r}$ in the direction of $\hat{\mathbf{y}}$. This projection *defines* the sine function.

If, instead of calculating the scalar product of $\mathbf{r}$ with $\hat{\mathbf{x}}$ we do it with a vector proportional to $\hat{\mathbf{x}}$, say, $\mathbf{z} = ||\mathbf{z}||\,\hat{\mathbf{x}}$, the bi-linearity of the scalar product implies

$$\mathbf{r}\cdot\mathbf{z} = \mathbf{r}\cdot(||\mathbf{z}||\,\hat{\mathbf{x}}) = ||\mathbf{z}||\,(\mathbf{r}\cdot\hat{\mathbf{x}}) = ||\mathbf{z}||\,||\mathbf{r}||\cos\theta \tag{5.25}$$

In words, the scalar product between a vector $\mathbf{r}$*, of modulus* $||\mathbf{r}||$*, and a vector* $\mathbf{z}$ *of modulus* $||\mathbf{z}||$*, is the product of their modules times the cosine of the angle between them.* Note that, since $|\cos\theta| \leq 1$, taking the absolute value on both sides of (5.25) we get,

$$|\mathbf{r}\cdot\mathbf{z}| \leq ||\mathbf{r}||\,||\mathbf{z}|| \tag{5.26}$$

the equality holding only when $\mathbf{r}$, $\mathbf{z}$, or θ are zero, or if $\theta = \pi$. If $\theta = 0$ or π, then $\mathbf{r} = a\,\mathbf{z}$, i.e., they are linearly dependent. This is an instance of the Cauchy–Schwarz inequality (5.13).

Before we immerse ourselves in n dimensions, note that in equation (5.24), the fact that the sine function appears instead of the cosine, is just because θ is the angle with the x-axis, not with the y-axis. The angle between $\mathbf{r}$ and $\hat{\mathbf{y}}$ is $\alpha = \pi/2 - \theta$, and as you remember from high school, $\sin(\theta) = \cos(\pi/2 - \theta) = \cos(\alpha)$. So again, $\mathbf{r}\cdot\hat{\mathbf{y}}$ is the product of the modules times the cosine of the angle between the vectors.

We will see now, helped with the machinery we have built, that we can extend all this to n dimensions, and in the process finally understand the geometric meaning of the scalar product.

Consider two vectors, $\mathbf{x}^1$ and $\mathbf{x}^2$, in an n-dimensional scalar product space V, see Fig. 9. $\mathbf{x}^1$ and $\mathbf{x}^2$ span a two-dimensional subspace of V. Think of the axis generated by the span of $\mathbf{x}^1$ as our previous x-axis. Since $\mathbf{x}^1$ and $\mathbf{x}^2$ define a unique plane in V, on that plane everything above still holds:

$$\mathbf{x}^1 \cdot \mathbf{x}^2 = ||\mathbf{x}^1||\,||\mathbf{x}^2||\cos\theta \tag{5.27}$$

Equation (5.27) *defines* the cosine of the angle θ between $\mathbf{x}^1$ and $\mathbf{x}^2$. The Cauchy–Schwarz inequality (5.13) ensures that $|\cos\theta| \le 1$.

Ok, that was perhaps a little too fast. Let us go slower. The vector

$$\hat{\mathbf{e}}^1 \equiv \frac{\mathbf{x}^1}{||\mathbf{x}^1||} \tag{5.28}$$

is a unit vector pointing in the direction of $\mathbf{x}^1$, see equation (5.9). Multiplying both sides by the modulus we have

$$\mathbf{x}^1 = ||\mathbf{x}^1||\,\hat{\mathbf{e}}^1 \tag{5.29}$$

which is an expression for $\mathbf{x}^1$ that neatly separates the modulus $||\mathbf{x}^1||$, and the direction $\hat{\mathbf{e}}^1$.

The vector

$$\mathbf{Proj}_{\mathbf{x}^1}(\mathbf{x}^2) = \mathbf{Proj}_{\hat{\mathbf{e}}^1}(\mathbf{x}^2) \equiv (\hat{\mathbf{e}}^1 \cdot \mathbf{x}^2)\,\hat{\mathbf{e}}^1 = \left(\frac{\mathbf{x}^1 \cdot \mathbf{x}^2}{\mathbf{x}^1 \cdot \mathbf{x}^1}\right)\mathbf{x}^1 \tag{5.30}$$

is, by definition, the *orthogonal projection* of $\mathbf{x}^2$ in the direction of $\mathbf{x}^1$ (or of $\hat{\mathbf{e}}^1$).

HW 5.9: Prove the last equality in (5.30).

What gives us the right to define the *projection* of a vector in the direction of another, as in (5.30)? We already discussed this when we proved the Cauchy–Schwarz inequality (5.13), but let us do it again in this more geometric context.

The word “orthogonal projection" is charged with geometric meaning, as in Fig. 11, where the

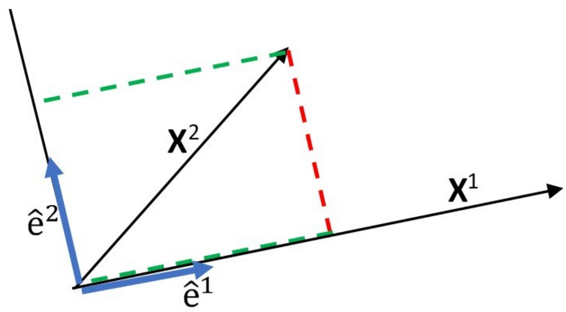


Figure 11: Orthogonal projection of $\mathbf{x}^2$ in the direction of $\mathbf{x}^1$, and in the orthogonal direction.

dashed green line is the orthogonal projection of $\mathbf{x}^2$ in the direction of $\hat{\mathbf{e}}^1$ (or $\mathbf{x}^1$), and the dashed red line is the orthogonal projection in the direction of $\hat{\mathbf{e}}^2$. $\hat{\mathbf{e}}^1$ and $\hat{\mathbf{e}}^2$ are themselves orthogonal to each other. Therefore $\mathbf{x}^2$, its projection in the direction of $\hat{\mathbf{e}}^1$, and its projection in the direction of $\hat{\mathbf{e}}^2$ form a right triangle where Pythagoras' theorem should apply. Does our definition, in an abstract n-dimensional space, satisfy all these properties? Only if it does do we have the right to call an expression like (5.30) an orthogonal projection.

Consider the following equation:

$$\mathbf{x}^2 = \mathbf{Proj}_{\hat{\mathbf{e}}^1}(\mathbf{x}^2) + \mathbf{Proj}_{\hat{\mathbf{e}}^2}(\mathbf{x}^2) \tag{5.31}$$

(5.31) *defines* $\mathbf{Proj}_{\hat{\mathbf{e}}^2}(\mathbf{x}^2)$, with $\mathbf{Proj}_{\hat{\mathbf{e}}^1}(\mathbf{x}^2)$ previously defined in (5.30). If all of these makes sense,

$$\mathbf{Proj}_{\hat{\mathbf{e}}^2}(\mathbf{x}^2) = \mathbf{x}^2 - \mathbf{Proj}_{\hat{\mathbf{e}}^1}(\mathbf{x}^2) \tag{5.32}$$

should be orthogonal to $\mathbf{x}^1$. Is it? Yes, here is the proof:

$$\begin{aligned}
\mathbf{Proj}_{\hat{\mathbf{e}}^2}(\mathbf{x}^2)\cdot\mathbf{x}^1 &= \left(\mathbf{x}^2 - \mathbf{Proj}_{\hat{\mathbf{e}}^1}(\mathbf{x}^2)\right)\cdot\mathbf{x}^1 \\
&= \left(\mathbf{x}^2 - \left(\frac{\mathbf{x}^1\cdot\mathbf{x}^2}{\mathbf{x}^1\cdot\mathbf{x}^1}\right)\mathbf{x}^1\right)\cdot\mathbf{x}^1 \\
&= \mathbf{x}^2\cdot\mathbf{x}^1 - \left(\frac{\mathbf{x}^1\cdot\mathbf{x}^2}{\mathbf{x}^1\cdot\mathbf{x}^1}\right)\left(\mathbf{x}^1\cdot\mathbf{x}^1\right) \\
&= \mathbf{x}^2\cdot\mathbf{x}^1 - \mathbf{x}^1\cdot\mathbf{x}^2 \\
&= 0
\end{aligned} \tag{5.33}$$

HW 5.10: Justify each step in the proof (5.33).

So the two projections in (5.31) are orthogonal to each other. This was the first condition to rightly call $\mathbf{Proj}_{\hat{\mathbf{e}}^1}$ in (5.30) a projection.

The second is the validity of Pythagoras' theorem. Does it hold for these abstract orthogonal projections in our n-dimensional context? Yes, here is the proof:

$$\begin{aligned}
\|\mathbf{x}^2\|^2 &= \mathbf{x}^2\cdot\mathbf{x}^2 \\
&= \left(\mathbf{Proj}_{\hat{\mathbf{e}}^1}(\mathbf{x}^2) + \mathbf{Proj}_{\hat{\mathbf{e}}^2}(\mathbf{x}^2)\right)\cdot\left(\mathbf{Proj}_{\hat{\mathbf{e}}^2}(\mathbf{x}^2) + \mathbf{Proj}_{\hat{\mathbf{e}}^2}(\mathbf{x}^2)\right) \\
&= \quad \mathbf{Proj}_{\hat{\mathbf{e}}^1}(\mathbf{x}^2)\cdot\mathbf{Proj}_{\hat{\mathbf{e}}^1}(\mathbf{x}^2) + \mathbf{Proj}_{\hat{\mathbf{e}}^1}(\mathbf{x}^2)\cdot\mathbf{Proj}_{\hat{\mathbf{e}}^2}(\mathbf{x}^2) \\
&\quad + \mathbf{Proj}_{\hat{\mathbf{e}}^2}(\mathbf{x}^2)\cdot\mathbf{Proj}_{\hat{\mathbf{e}}^1}(\mathbf{x}^2) + \mathbf{Proj}_{\hat{\mathbf{e}}^2}(\mathbf{x}^2)\cdot\mathbf{Proj}_{\hat{\mathbf{e}}^2}(\mathbf{x}^2) \\
&= \|\mathbf{Proj}_{\hat{\mathbf{e}}^1}(\mathbf{x}^2)\|^2 + \|\mathbf{Proj}_{\hat{\mathbf{e}}^2}(\mathbf{x}^2)\|^2
\end{aligned} \tag{5.34}$$

HW 5.11: Justify each step in the proof (5.34).

Orthogonal projections and Pythagoras' theorem are all we need to genuinely define the cosine function. So we are justified in saying that the scalar product between two vectors, $\mathbf{x}^1$ and $\mathbf{x}^2$, is equal to the product of the respective modules times the cosine of the angle between them:

$$\mathbf{x}^1\cdot\mathbf{x}^2 = \|\mathbf{x}^1\|\,\|\mathbf{x}^2\|\cos\theta \tag{5.35}$$

This is the geometric meaning of the scalar product between two vectors of a scalar product vector space V in *any* dimension. When $\mathbf{x}^1\cdot\mathbf{x}^2 = 0$ and none of them are $\mathbf{0}$, $\cos\theta = 0$ and the angle is 90° (or 270°), as it should for orthogonal vectors.

Note that the scalar product between two vectors only depends on the angle between them. It is independent of the direction of each of these vectors with respect to any coordinate system. In this sense, it is an intrinsic property of the relationship between two vectors. In the vector space all directions are equivalent, only *relations* between directions have meaning.

Remember that we mentioned that the underlying manifold is *homogeneous*, since there was no special point, see Fig. 3. Then we insisted on the idea that the manifold exists independently of any numerical representation. Now we are faced with another important property that the manifold inherits from its relationship with the corresponding vector space: since all directions are equivalent, it is said to be *isotropic*.

We don't experience the real 3-D space of our experience as isotropic; there is a clear distinction between the two independent directions on the surface of the earth and the vertical direction. The whole Aristotelian model of the world rested on this difference. We had to wait until the scientific revolution to learn that space was isotropic. But as we will see, as a concept to handle data, the isotropy of the vector space, and of the underlying manifold, is incredibly useful. And for our post-scientific-revolution intuition, it is not hard to imagine an isotropic manifold; it is just the extension of how we naturally think about a plane.

5.2 Gram-Schmidt process

The Gram-Schmidt process consist of transforming a set $J = \{\mathbf{x}^i\}$, $i = 1, \cdots, n$, of n linearly independent vectors, into another set $K = \{\mathbf{v}^i\}$ of n orthogonal vectors spanning the same subspace as J. And by a trivial extension, into the set $E = \{\hat{\mathbf{e}}^i\}$ of n orthonormal vectors spanning the same subspace as J. After reading the previous section, the process should be easy to grasp. It consists of the following:

1. Define $\mathbf{v}^1$ as equal to $\mathbf{x}^1$:
$$\mathbf{v}^1 \equiv \mathbf{x}^1 \tag{5.36}$$
$\hat{\mathbf{e}}^1$ is the unit norm vector in the direction of $\mathbf{v}^1$:
$$\hat{\mathbf{e}}^1 = \frac{\mathbf{v}^1}{||\mathbf{v}^1||} \tag{5.37}$$

2. Define $\mathbf{v}^2$ as:
$$\mathbf{v}^2 = \mathbf{x}^2 - \mathbf{Proj}_{\mathbf{v}^1}(\mathbf{x}^2) = \mathbf{x}^2 - \left(\frac{\mathbf{v}^1 \cdot \mathbf{x}^2}{\mathbf{v}^1 \cdot \mathbf{v}^1}\right) \mathbf{v}^1 \tag{5.38}$$
As we saw in (5.33), $\mathbf{v}^2$ is orthogonal to $\mathbf{v}^1$ (and therefore to $\mathbf{x}^1$), and $\mathbf{v}^1$ and $\mathbf{v}^2$ span the same subspace as $\mathbf{x}^1$ and $\mathbf{x}^2$. The unit norm vector in the direction of $\mathbf{v}^2$ is:
$$\hat{\mathbf{e}}^2 = \frac{\mathbf{v}^2}{||\mathbf{v}^2||} \tag{5.39}$$

3. Define $\mathbf{v}^3$ as:

$$\mathbf{v}^3 = \mathbf{x}^3 - \mathbf{Proj}_{\mathbf{v}^1}(\mathbf{x}^3) - \mathbf{Proj}_{\mathbf{v}^2}(\mathbf{x}^3) = \mathbf{x}^3 - \left(\frac{\mathbf{v}^1 \cdot \mathbf{x}^3}{\mathbf{v}^1 \cdot \mathbf{v}^1}\right)\mathbf{v}^1 - \left(\frac{\mathbf{v}^2 \cdot \mathbf{x}^3}{\mathbf{v}^2 \cdot \mathbf{v}^2}\right)\mathbf{v}^2 \tag{5.40}$$

$\mathbf{v}^3$ is orthogonal to $\mathbf{v}^1$ and $\mathbf{v}^2$, and $\mathbf{v}^1$, $\mathbf{v}^2$ and $\mathbf{v}^3$ span the same subspace as $\mathbf{x}^1$, $\mathbf{x}^2$ and $\mathbf{x}^3$. The unit norm vector in the direction of $\mathbf{v}^3$ is:

$$\hat{\mathbf{e}}^3 = \frac{\mathbf{v}^3}{||\mathbf{v}^3||} \tag{5.41}$$

4. In general, assuming that this process has been done for $i = 1, \cdots, j-1$, define $\mathbf{v}^j$ as:

$$\mathbf{v}^j = \mathbf{x}^j - \sum_{i=1}^{j-1} \mathbf{Proj}_{\mathbf{v}^i}(\mathbf{x}^j) = \mathbf{x}^j - \sum_{i=1}^{j-1}\left(\frac{\mathbf{v}^i \cdot \mathbf{x}^j}{\mathbf{v}^i \cdot \mathbf{v}^i}\right)\mathbf{v}^i \tag{5.42}$$

$\mathbf{v}^j$ is orthogonal to $\mathbf{v}^i$, for $i = 1, \cdots, j-1$. The set $\{\mathbf{v}^i\}$, with $i = 1, \cdots, j$, span the same subspace as $\{\mathbf{x}^i\}$, $i = 1, \cdots, j$. The unit norm vector in the direction of $\mathbf{v}^j$ is:

$$\hat{\mathbf{e}}^j = \frac{\mathbf{v}^j}{||\mathbf{v}^j||} \tag{5.43}$$

HW 5.12: a) Prove that $\mathbf{v}^3$, as defined in (5.40), is orthogonal to $\mathbf{v}^1$and $\mathbf{v}^2$. b) Prove that $\mathbf{v}^1$, $\mathbf{v}^2$ and $\mathbf{v}^3$ span the same subspace as $\mathbf{x}^1$, $\mathbf{x}^2$ and $\mathbf{x}^3$.

HW 5.13: a) Prove that $\mathbf{v}^j$, as defined in (5.42), is orthogonal to $\mathbf{v}^i$, for $i = 1, \cdots, j-1$. b) Prove that the set $\{\mathbf{v}^i\}$, with $i = 1, \cdots, j$, span the same subspace as $\{\mathbf{x}^i\}$.

HW 5.14: In R^2 suppose that you have the basis $\{\mathbf{x}^1, \mathbf{x}^2\}$ given by $\mathbf{x}^1 = \hat{\mathbf{e}}^1 + \hat{\mathbf{e}}^2$, $\mathbf{x}^2 = \hat{\mathbf{e}}^1$, where $\hat{\mathbf{e}}^1$ and $\hat{\mathbf{e}}^2$ are orthonormal. a) Draw $\mathbf{x}^1$ and $\mathbf{x}^2$ in a plane with the horizontal axis pointing in the direction of $\hat{\mathbf{e}}^1$ and the vertical axis pointing in the direction of $\hat{\mathbf{e}}^2$. b) Define $\mathbf{v}^1 = \mathbf{x}^1$ (see (5.36)) and compute $\mathbf{v}^2$ as in (5.38). Verify that $\mathbf{v}^2$ is orthogonal to $\mathbf{v}^1$. c) Draw $\mathbf{v}^1$ and $\mathbf{v}^2$ in plane with the horizontal axis pointing in the direction of $\hat{\mathbf{e}}^1$ and the vertical axis pointing in the direction of $\hat{\mathbf{e}}^2$. d) Normalize $\mathbf{v}^1$ and $\mathbf{v}^2$ to obtain $\hat{\mathbf{e}}^1$ and $\hat{\mathbf{e}}^2$. e) Assuming that the column vector representation of $\hat{\mathbf{e}}^1 = (1, 0)^\mathrm{T}$ and $\hat{\mathbf{e}}^2 = (0, 1)^\mathrm{T}$, give the column vector representation of $\mathbf{v}^1$ and $\mathbf{v}^2$ and of $\hat{\mathbf{e}}^1$ and $\hat{\mathbf{e}}^2$.

HW 5.15: Do the same in R^3. Assume that you have the base $\{\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3\}$ given by $\mathbf{x}^1 = \hat{\mathbf{e}}^1 + \hat{\mathbf{e}}^2 + \hat{\mathbf{e}}^3$, $\mathbf{x}^2 = \hat{\mathbf{e}}^1$, and $\mathbf{x}^3 = \hat{\mathbf{e}}^2$, where $\hat{\mathbf{e}}^1$, $\hat{\mathbf{e}}^2$ and $\hat{\mathbf{e}}^3$ are orthonormal. Assume first $\mathbf{v}^1 = \mathbf{x}^1$. Do the same assuming $\mathbf{v}^1 = \mathbf{x}^2$. Explain why you obtain two different bases. Prove that these two different basis span the same vector space.

The Gram-Schmidt process shows the validity of the following theorems:

Theorem 5.1 *Every nonzero linear subspace of a dot product space V has an orthogonal basis.*

Theorem 5.2 *For every orthogonal collection of vectors $\mathbf{v}^i$, $i = 1, \cdots, j$, spanning a subspace W of the dot product state V, every vector $\mathbf{w} \in W$ has a (unique) expansion in terms of $\mathbf{v}^i$ as:*

$$\mathbf{w} = \sum_{i=1}^{j} \left(\frac{\mathbf{w} \cdot \mathbf{v}^i}{\mathbf{v}^i \cdot \mathbf{v}^i} \right) \mathbf{v}^i \tag{5.44}$$

If the collection of vectors $\hat{\mathbf{e}}^i$, $i = 1, \cdots, j$ is orthonormal, the denominator of (5.44) is 1, so:

$$\mathbf{w} = \sum_{i=1}^{j} \left(\mathbf{w} \cdot \hat{\mathbf{e}}^i \right) \hat{\mathbf{e}}^i \tag{5.45}$$

5.3 Orthogonal complements

Orthogonal complement: Given a subspace W of a scalar product space V, the set of all vectors orthogonal to W is the *orthogonal complement*, $W^\perp$ of W:

$$W^\perp = \{\mathbf{u} \in V, \text{ such that } \mathbf{u} \cdot \mathbf{w} = 0 \text{ for all } \mathbf{w} \in W\} \tag{5.46}$$

HW 5.16: Given the subspace spanned by $(2, 5)^{\mathrm{T}}$, find the orthogonal complement in R^2.

HW 5.17: Given the subspace spanned by $(-1, 1, -1)^{\mathrm{T}}$, find the orthogonal complement in R^3.

Theorem 5.3 *Given a vector $\mathbf{v}$ in a scalar product space V, and a subspace W of V, there is a unique way to express $\mathbf{v}$ as a sum*

$$\mathbf{v} = \mathbf{w} + \mathbf{u} \tag{5.47}$$

where $\mathbf{w} \in W$ and $\mathbf{u} \in W^{\mathrm{T}}$.

HW 5.18: Prove the theorem.

Theorem 5.4 *If W is a subspace of a scalar product space V of dimension n, and W^{T} is its orthogonal complement, then*

$$\dim W^{\mathrm{T}} = n - \dim W \tag{5.48}$$

HW 5.19: Prove the theorem.

Theorem 5.5 *If $\boldsymbol{x}^i$, $i = 1, \cdots, n$, is a basis of a scalar product space V, and $\boldsymbol{x}^i$, $i = 1, \cdots, m$, with $m < n$, is a basis of a subspace W, then, the Gram-Schmidt orthogonalization process starting with $\mathbf{v}^1 = \boldsymbol{x}^1$ generates an orthogonal basis of W given by $\boldsymbol{v}^i$, $i = 1, \cdots, m$, and an orthogonal basis of W^{T} given by $\boldsymbol{v}^i$, $i = m + 1, \cdots, n$.*

HW 5.20: Prove the theorem.

6 Manifold and vector space description of mathematical objects

As we have mentioned many times, once we choose an (arbitrary) origin in the manifold, there is a one-to-one relationship between vectors in the vector space and points in the manifold. But we have mostly been working with vectors. In this section we dig deeper into this relationship.

6.1 Coordinate systems and grids on the manifold

For simplicity of visualization, let us consider an Euclidean plane, and the associated vector space $V = \mathrm{R}^2$. Having chosen a point O in the plane as the origin, to every point G in the manifold corresponds a vector $\mathbf{g} \in V$.

Consider two linearly independent vectors, $\mathbf{x}^1$ and $\mathbf{x}^2$, that form a basis of V. Theorem 4.1 tells us that every vector $\mathbf{g} \in V$ can be written as a linear combination of the $\mathbf{x}^i$s in a unique way:

$$\mathbf{g} = x_1\mathbf{x}^1 + x_2\mathbf{x}^2 \tag{6.1}$$

Coordinate system: (x_1, x_2) are the *coordinates* of the point G of the manifold in the coordinate system determined by the origin O and the basis vectors $\mathbf{x}^1$ and $\mathbf{x}^2$.

We write the coordinates as an ordered pair of numbers in a row (in n dimensions it will be a row of n ordered numbers).

Let us use for this example the column representation of vectors:

$$\mathbf{x}^1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \quad \mathbf{x}^2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad \mathbf{g} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}, \tag{6.2}$$

Explicitly, $\mathbf{g}$ has components 3 and 4 in the orthogonal basis $\{(1, 0)^\mathrm{T}, (0, 1)^\mathrm{T}\}$. This means that in the coordinate system determined in the manifold by these orthogonal vectors, the coordinates of the point corresponding to the vector $(4, 3)^\mathrm{T}$ are $(4, 3)$. What are the coordinates of the *same point* in the coordinate system determined by $\{\mathbf{x}^1, \mathbf{x}^2\}$?

We have the following vectorial equation:

$$x_1\begin{pmatrix} 2 \\ 1 \end{pmatrix} + x_2\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2x_1 + x_2 \\ x_1 + 2x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \end{pmatrix} \tag{6.3}$$

This equation “lives" in the vector space V. Remember that you can always think of the vector space as the set of all possible displacements. Taking the scalar product of vectorial equation (6.3), first with $(1, 0)^\mathrm{T}$ and then with $(0, 1)^\mathrm{T}$, we get two equivalent equations determining the coordinates x_1 and x_2 in the coordinate system determined by $\{\mathbf{x}^1, \mathbf{x}^2\}$:

$$\begin{aligned} 2x_1 + x_2 &= 4 \qquad (6.4) \\ x_1 + 2x_2 &= 3 \qquad (6.5) \end{aligned}$$

Equations (6.4-6.5) “live" in the manifold. The solution is $x_1 = 5/3$, $x_2 = 2/3$. The graphical representation is given in Fig. 12.

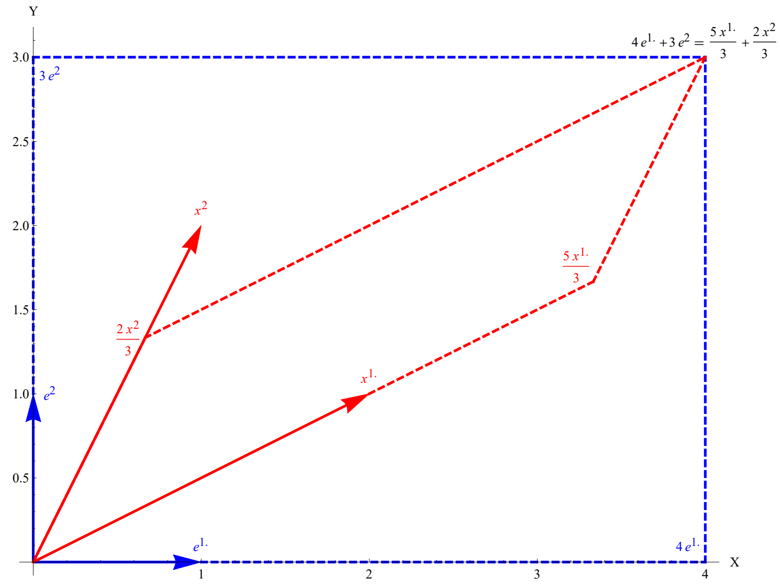


Figure 12: The same point in the manifold has different coordinates in coordinate systems determined by different bases. Note the importance of thinking about the manifold and its points as independent from the “names" we assign to these points.

One can visualize the coordinate systems as a grid in the manifold. Consider the affine subspaces of V given by

$$\mathbf{v}(n, \lambda) = n\,\mathbf{x}^2 + \lambda\,\mathbf{x}^1, \qquad \text{lines of the grid parallel to } \mathbf{x}^1 \tag{6.6}$$

$$\mathbf{w}(m, \beta) = m\,\mathbf{x}^1 + \beta\,\mathbf{x}^2, \qquad \text{lines of the grid parallel to } \mathbf{x}^2 \tag{6.7}$$

n and m can be any integer $(0, \pm 1, \pm 2, \cdots)$. For each value of n, λ runs from $-\infty$ to ∞ for a line of the grid parallel to $\mathbf{x}^1$. Similarly, for each value of m, β runs from $-\infty$ to ∞ for a line of the grid parallel to $\mathbf{x}^2$.

Grid: the points in the manifold corresponding to the affine subspaces (6.6) and (6.7) form the *grid* on the manifold associated to the coordinate system determined by the basis $\{\mathbf{x}^1, \mathbf{x}^2\}$.

In Fig 13 we can appreciate, in red, the grid on the plane corresponding to the basis $\{\mathbf{x}^1, \mathbf{x}^2\}$ given in equation (6.2), and in blue, the grid corresponding to the orthogonal basis $\{(1, 0)^{\mathrm{T}}, (0, 1)^{\mathrm{T}}\}$.

HW 6.1: In Fig 13, what are the coordinates in the coordinate system determined by $\{\mathbf{x}^1, \mathbf{x}^2\}$, of the points whose coordinates in the coordinate system determined by $\{(1, 0)^{\mathrm{T}}, (0, 1)^{\mathrm{T}}\}$ are: $(3, 3)$, $(-1, 4)$, and $(-2, -4)$?

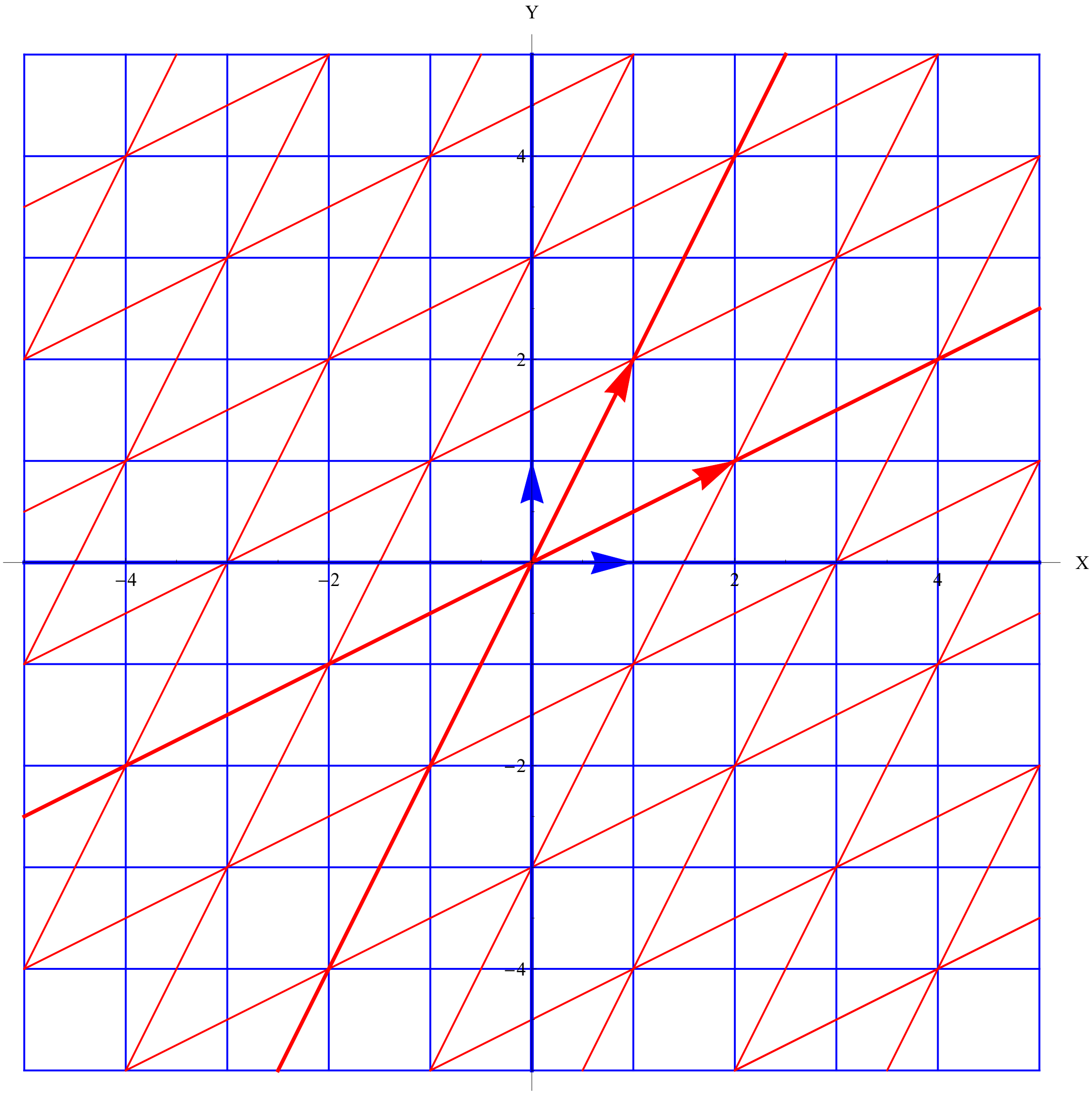


Figure 13: Coordinate systems in the manifold can be visualized as grids. The blue grid is generated by the orthogonal vectors $\hat{\mathbf{e}}^1 = (1,0)^T$ and $\hat{\mathbf{e}}^2 = (0,1)^T$. The red grid is generated by the non-orthogonal vectors $\mathbf{x}^1 = 2\hat{\mathbf{e}}^1 + \hat{\mathbf{e}}^2 = (2,1)^T$ and $\mathbf{x}^2 = \hat{\mathbf{e}}^1 + 2\hat{\mathbf{e}}^2 = (1,2)^T$.

HW 6.2: In Fig 13, what are the coordinates in the coordinate system determined by $\{(1,0)^T, (0,1)^T\}$, of the points whose coordinates in the coordinate system determined by $\{\mathbf{x}^1, \mathbf{x}^2\}$ are: $(2,-1)$, $(-2,2)$, and $(-1,-2)$?

Earlier we insisted in thinking about the manifold, and the corresponding vector space, as objects to which one should concede ontological primacy over the number representation. Now we are beginning to see the tip of the iceberg of the advantages of assigning such ontological primacy. A problem may be difficult in one coordinate system and simple in another. The manifold idea, and the freedom to choose the coordinate system we like, gives us the freedom to switch from one description to another while intuitively recognizing that it is still the same problem. This freedom will become far more important in part two, when we analyze linear mappings.

6.2 Equivalent descriptions of lines and surfaces

Equations like (6.6) or (6.7) are examples of vectorial representations of lines. The vectorial description of lines as the points in the manifold associated to the corresponding affine subspace extends to any dimension. One just has to select any point A in the line, with the corresponding vector $\mathbf{a}$, and a vector $\mathbf{d}$ pointing in the direction of the line. So the whole line is the collection of points corresponding to the vectors $\mathbf{r}(\lambda)$:

$$\mathbf{r}(\lambda) = \mathbf{a} + \lambda\,\mathbf{d}, \quad -\infty \leq \lambda \leq \infty \tag{6.8}$$

Note that if $\mathbf{a}^1$ and $\mathbf{a}^2$, both belong to the line[9], $\mathbf{r}(\lambda) = \mathbf{a}^1 + \lambda\,\mathbf{d}$ and $\mathbf{r}(\lambda) = \mathbf{a}^2 + \lambda\,\mathbf{d}$ represent the same line. It is just that a given value of λ corresponds to different points in the line in the two descriptions. For example, $\lambda = 0$ corresponds in the representation $\mathbf{r}(\lambda) = \mathbf{a}^1 + \lambda\,\mathbf{d}$, to $\mathbf{r}(0) = \mathbf{a}^1$, and in the representation $\mathbf{r}(\lambda) = \mathbf{a}^2 + \lambda\,\mathbf{d}$, to $\mathbf{r}(0) = \mathbf{a}^2$.

Similarly, if $\mathbf{d}$ points in the direction of the line, any other vector in the span of $\mathbf{d}$, like $\mathbf{f} = \gamma\,\mathbf{d}$, also points in the direction of the line. Therefore if $\mathbf{r}(\lambda) = \mathbf{a} + \lambda\,\mathbf{d}$ is a representation of the line, $\mathbf{r}(\lambda) = \mathbf{a} + (\lambda/\gamma)\,\gamma\mathbf{d} = \mathbf{a} + \delta\,\mathbf{f}$, with $\delta = \lambda/\gamma$ is also a representation of the same line.

If $\{\hat{\mathbf{e}}^i\}$, $i = 1, \cdots, n$, is an orthonormal basis of V, the vectors $\mathbf{a}$ and $\mathbf{d}$ in (6.8) can be expanded as $\mathbf{a} = \sum_{i=1}^n a_i\,\hat{\mathbf{e}}^i$ and $\mathbf{d} = \sum_{i=1}^n d_i\,\hat{\mathbf{e}}^i$. A generic vector $\mathbf{x} \in V$ can also be expanded as $\mathbf{x} = \sum_{i=1}^n x_i\,\hat{\mathbf{e}}^i$. With the vectors expanded in this base, equation (6.8) can be written as

$$\mathbf{x} = \sum_{i=1}^{n} x_i\hat{\mathbf{e}}^i = \mathbf{a} + \lambda\,\mathbf{d} = \sum_{i=1}^{n} a_i\hat{\mathbf{e}}^i + \lambda\sum_{i=1}^{n} d_i\hat{\mathbf{e}}^i = \sum_{i=1}^{n} (a_i + \lambda d_i)\,\hat{\mathbf{e}}^i \tag{6.9}$$

Taking the scalar product of (6.9) with $\hat{\mathbf{e}}^j$, for $j = 1, \cdots, n$, we obtain the n equations

$$x_j = a_j + \lambda d_j, \quad j = 1, \cdots, n \tag{6.10}$$

This is a *parametric* description of a line. While equation (6.9) is an equation in the vector space V, it is convenient to view (6.10) as an equation in the manifold, corresponding to the parametric coordinates of the line in the reference frame determined by the basis $\{\hat{\mathbf{e}}^i\}$.

From (6.10), $\lambda = (x_j - a_j)/d_j$, therefore, yet another form of equations determining the same line is

$$(x_1 - a_1)/d_1 = (x_2 - a_2)/d_2 = \cdots = (x_n - a_n)/d_n \tag{6.11}$$

[9]The *points* in the manifold A_1 and A_2 belong to the line, not the vectors $\mathbf{a}^1$ and $\mathbf{a}^2$. However, we have already mentioned at the end of section 4.2 that, for purposes of brevity, we will sometimes talk about "point $\mathbf{a}^i$" instead of the more correct "point A_i with associated vector $\mathbf{a}^i$".

(6.11) is sometimes referred to as the Cartesian equation of the line. One can express all the Cartesian coordinates in terms of one of them, say x_1: from (6.10), $\lambda = (x_j - a_j)/d_j$, therefore, yet another form of equations determining the same line is

$$x_j = a_j + \frac{d_j}{d_1}(x_1 - a_1), \quad j = 1, \cdots, n \tag{6.12}$$

Finally, in the particular case of $V = \mathrm{R}^2$, given an equation of the line like (6.8), there is a one-dimensional subspace of V orthogonal to the line. Any vector $\mathbf{n}$ on the orthogonal subspace must be orthogonal to $\mathbf{d}$, $\mathbf{d} \cdot \mathbf{n} = 0$. Therefore the scalar product of any vector $\mathbf{r}$ on the line and $\mathbf{n}$ should give the same result:

$$\mathbf{r} \cdot \mathbf{n} = \mathbf{a} \cdot \mathbf{n} + \lambda\, \mathbf{d} \cdot \mathbf{n} = \mathbf{a} \cdot \mathbf{n} \tag{6.13}$$

(6.13) is yet another description of the line in R^2. The line is the set of points $\mathbf{r}$ that have the same scalar product with a vector orthogonal to the line.

HW 6.3: Make a graph to visualize that any point in the line should have the same projection on a vector $\mathbf{n}$ orthogonal to the line. In particular, if the line passes through the origin, $\mathbf{r} \cdot \mathbf{n} = 0$.

Equation (6.8) is the vector equation for a line. Similarly, the two-dimensional affine subspace

$$\mathbf{r}(\lambda_1, \lambda_2) = \mathbf{a} + \lambda_1\, \mathbf{d}^1 + \lambda_2\, \mathbf{d}^2, \quad -\infty \leq \lambda_1, \lambda_2 \leq \infty \tag{6.14}$$

represents a two-dimensional linear surface in the manifold corresponding to the vector space V, where $\mathbf{a}$ is any point in the surface and $\mathbf{d}^1$ and $\mathbf{d}^2$ are two linearly independent vectors parallel to the surface.

Generalizing, the affine subspace

$$\mathbf{r}(\lambda_1, \cdots, \lambda_n) = \mathbf{a} + \sum_{i=1}^{n} \lambda_i\, \mathbf{d}^i, \quad -\infty \leq \lambda_i \leq \infty, \quad i = 1, \cdots, n \tag{6.15}$$

represents an n-dimensional linear hypersurface, where $\mathbf{a}$ is any point in the n-surface and $\{\mathbf{d}^i\}$ are n linearly independent vectors parallel to the linear n-surface.

It should be clear that if in (6.15) we replace the linearly independent vectors $\{\mathbf{d}^i\}$ by any other n linearly independent vectors $\{\mathbf{f}^i\}$ spanning the same subspace as $\{\mathbf{d}^i\}$, we are still describing the same n-dimensional linear surface.

HW 6.4: Find expressions equivalent to (6.15) in a form similar to and (6.10) and (6.12).

Consider an $(n-1)$-dimensional subspace W, of an n-dimensional vector space V, spanned by the linearly independent vectors $\mathbf{x}^i$, $i = 1, \cdots, n-1$. Through the Gram-Schmidt process described in section 5.2, one can construct an orthonormal basis of V, $\{\hat{\mathbf{e}}^i\}$, $i = 1, \cdots, n$, such that the first $n-1$ vectors constitute an orthonormal basis of W. Therefore any vector in W can be written as

$$\mathbf{r}(\lambda_1, \cdots, \lambda_{n-1}) = \sum_{i=1}^{n-1} \lambda_i\, \hat{\mathbf{e}}^i \tag{6.16}$$

And any $(n-1)$-dimensional affine subspace of V “parallel to W" can be written as

$$\mathbf{r}(\lambda_1, \cdots, \lambda_{n-1}) = \mathbf{a} + \sum_{i=1}^{n-1} \lambda_i \, \hat{\mathbf{e}}^i \tag{6.17}$$

The points in the manifold corresponding to this affine subspace form an $(n-1)$-dimensional hypersurface.

Another way of characterizing this $(n-1)$-dimensional hypersurface is obtained by taking the scalar product of (6.16) with $\hat{\mathbf{e}}^n$:

$$\mathbf{r} \cdot \hat{\mathbf{e}}^n = \mathbf{a} \cdot \hat{\mathbf{e}}^n \tag{6.18}$$

In words, an $(n-1)$-dimensional hypersurface in an n-dimensional manifold corresponds to the points that have the same projection on a vector $\hat{\mathbf{e}}^n$ orthogonal to the hypersurface. This is a generalization of (6.13).

For example, equation (2.12) corresponds to a line in the two-dimensional manifold (p_1, p_2) orthogonal to the vector $(4, -1)^{\mathrm{T}}$ and such that the scalar product of any vector $(p_1, p_2)^{\mathrm{T}}$ in the line with $(4, -1)^{\mathrm{T}}$ is equal to 20. Since the vector $(1, 4)^{\mathrm{T}}$ is orthogonal to $(4, -1)^{\mathrm{T}}$, and, say, the vector $(5, 0)^{\mathrm{T}}$ has a scalar product with $(4, -1)^{\mathrm{T}}$ equal to 20, a vectorial expression equivalent to (2.12) is, for example,

$$\begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} 5 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 4 \end{pmatrix} \tag{6.19}$$

HW 6.5: Find a vectorial expression equivalent to (2.13)

HW 6.6: Each one of the equations (2.18-2.21) correspond to a 3-hypersurface in the 4-dimensional manifold (q_1, q_2, p_1, p_2). Find a vectorial expressions equivalent to each one of these equations.

7 Projections and minimization of distance

7.1 Point in a line closest to a point Q external to it

Line embedded in R^2

The problem we want to solve is represented in Fig. 14. There is a one-dimensional subspace W of $V = \mathrm{R}^2$ spanned by a vector $\mathbf{d}$. The vectors $\mathbf{w} \in W$ can be written in the form

$$\mathbf{w} = \lambda \, \mathbf{d} \tag{7.1}$$

We want to find the point in W closest to a point Q corresponding to a vector $\mathbf{q}$. This is the same as minimizing the magnitude of the vector $\mathbf{r}$ such that:

$$\mathbf{w} + \mathbf{r} = \lambda \, \mathbf{d} + \mathbf{r} = \mathbf{q} \tag{7.2}$$

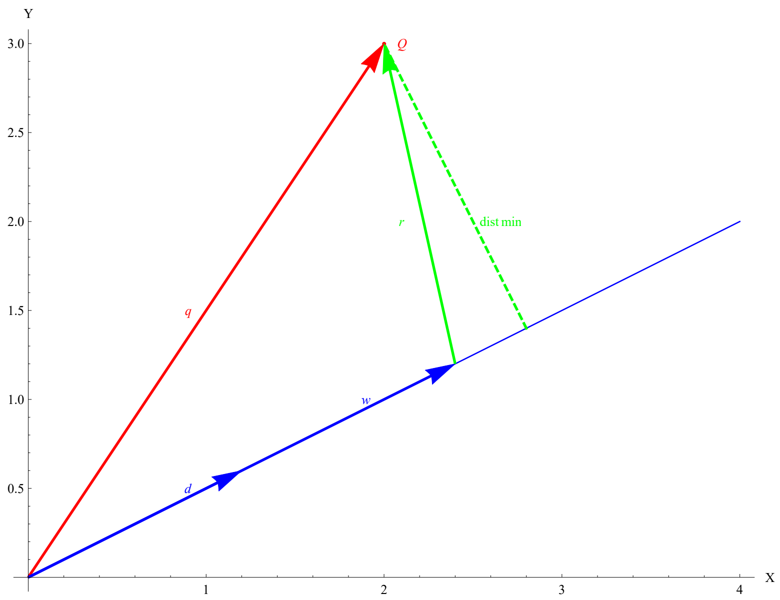


Figure 14: The closest point to Q in W corresponds to the orthogonal projection of $\mathbf{q}$ in the direction of $\mathbf{d}$.

One could write the expression for $||\mathbf{r}||$, or $||\mathbf{r}||^2$, and use calculus to minimize that magnitude. However, our spatial intuition, and the machinery we have built to exploit it, provides the answer almost effortlessly. A simple inspection at the Fig. 14 will convince you that the closest point corresponds to an $\mathbf{r}$ orthogonal to the line, or, what is the same, to $\mathbf{d}$:

$$\mathbf{r}_{\text{min}} \cdot \mathbf{d} = (\mathbf{q} - \lambda\, \mathbf{d}) \cdot \mathbf{d} = 0 \tag{7.3}$$

this equation determines the parameter λ corresponding to the point in W that minimizes the distance with Q:

$$\lambda = \frac{\mathbf{d} \cdot \mathbf{q}}{\mathbf{d} \cdot \mathbf{d}} \tag{7.4}$$

Therefore, the closest point to Q in W:

$$\mathbf{w} = \left(\frac{\mathbf{d} \cdot \mathbf{q}}{\mathbf{d} \cdot \mathbf{d}}\right)\mathbf{d} \tag{7.5}$$

In equation (5.30) we defined an expression like (7.5) as the *orthogonal projection* of $\mathbf{q}$ in the direction of $\mathbf{d}$. The solution to the problem of finding the point in the subspace W closest to a point Q is, therefore, the orthogonal projection of $\mathbf{q}$ in the direction of any vector $\mathbf{d}$ spanning W. $\mathbf{r}_{\text{min}}$ spans the orthogonal complement $W^{\perp}$ of W (see equation (5.46)):

$$\mathbf{r}_{\text{min}} = \mathbf{q} - \left(\frac{\mathbf{d} \cdot \mathbf{q}}{\mathbf{d} \cdot \mathbf{d}}\right)\mathbf{d} \tag{7.6}$$

and the actual minimal distance from Q to W is given by Pythagoras' theorem. Therefore the square of the length of a side of the triangle can be written in terms of the square of the hypotenuse and the square of the length of the other side:

$$\begin{aligned}
\min \operatorname{dist}^2(Q, W) &= \mathbf{r}_{\min} \cdot \mathbf{r}_{\min} \\
&= \left[\mathbf{q} - \left(\frac{\mathbf{d}\cdot\mathbf{q}}{\mathbf{d}\cdot\mathbf{d}}\right)\mathbf{d}\right] \cdot \left[\mathbf{q} - \left(\frac{\mathbf{d}\cdot\mathbf{q}}{\mathbf{d}\cdot\mathbf{d}}\right)\mathbf{d}\right] \\
&= \mathbf{q}\cdot\mathbf{q} - \frac{(\mathbf{d}\cdot\mathbf{q})^2}{\mathbf{d}\cdot\mathbf{d}} \qquad (7.7)
\end{aligned}$$

HW 7.1: Prove the last equality in (7.7).

It is important to note how both the closest point in W (7.5) and the actual minimal distance (7.7) scale when the vectors $\mathbf{q}$ and $\mathbf{d}$ are scaled to $\beta\mathbf{q}$ and $\gamma\mathbf{d}$ respectively:

$$\begin{aligned}
\text{when} \quad \mathbf{q}, \mathbf{d} &\to \beta\mathbf{q}, \gamma\mathbf{d} \\
\mathbf{w} &\to \left(\frac{\gamma\mathbf{d}\cdot\beta\mathbf{q}}{\gamma\mathbf{d}\cdot\gamma\mathbf{d}}\right)\gamma\mathbf{d} = \beta\mathbf{w} \qquad (7.8) \\
\min \operatorname{dist}(Q, W) &\to \left(\beta\mathbf{q}\cdot\beta\mathbf{q} - \frac{(\gamma\mathbf{d}\cdot\beta\mathbf{q})^2}{\gamma\mathbf{d}\cdot\gamma\mathbf{d}}\right)^{1/2} = \beta \min \operatorname{dist}(Q, W) \qquad (7.9)
\end{aligned}$$

As expected, they scale linearly with the scaling of the vector $\mathbf{q}$, and do not depend on the scaling of the vector $\mathbf{d}$. Since $\mathbf{d}$ is an arbitrarily chosen base of W, the solution should not depend on our arbitrary selection of a larger of a shorter vector.

Suppose now that the line corresponds to points in the manifold associated to an *affine* subspace W' of V. Instead of (7.1) we have

$$\mathbf{w}' = \mathbf{a} + \lambda\,\mathbf{d} \qquad (7.10)$$

for the vectors corresponding to the points in the line (see Fig. 15). The vector $\mathbf{r}$ is defined by the analog of equation (7.2)

$$\mathbf{w}' + \mathbf{r} = \mathbf{a} + \lambda\,\mathbf{d} + \mathbf{r} = \mathbf{q} \qquad (7.11)$$

or

$$\mathbf{r} = (\mathbf{q} - \mathbf{a}) - \lambda\,\mathbf{d} \qquad (7.12)$$

It is still the case, as in (7.3), that the closest point to Q is the one that makes $\mathbf{r}$ orthogonal to $\mathbf{d}$:

$$\mathbf{r}_{\min}\cdot\mathbf{d} = [(\mathbf{q} - \mathbf{a}) - \lambda\,\mathbf{d}]\cdot\mathbf{d} = 0 \qquad (7.13)$$

determining λ as

$$\lambda = \frac{\mathbf{d}\cdot(\mathbf{q} - \mathbf{a})}{\mathbf{d}\cdot\mathbf{d}} \qquad (7.14)$$

analogous to (7.4). This corresponds to the closest point to Q in W':

$$\mathbf{w}' = \mathbf{a} + \left(\frac{\mathbf{d}\cdot(\mathbf{q} - \mathbf{a})}{\mathbf{d}\cdot\mathbf{d}}\right)\mathbf{d} \qquad (7.15)$$

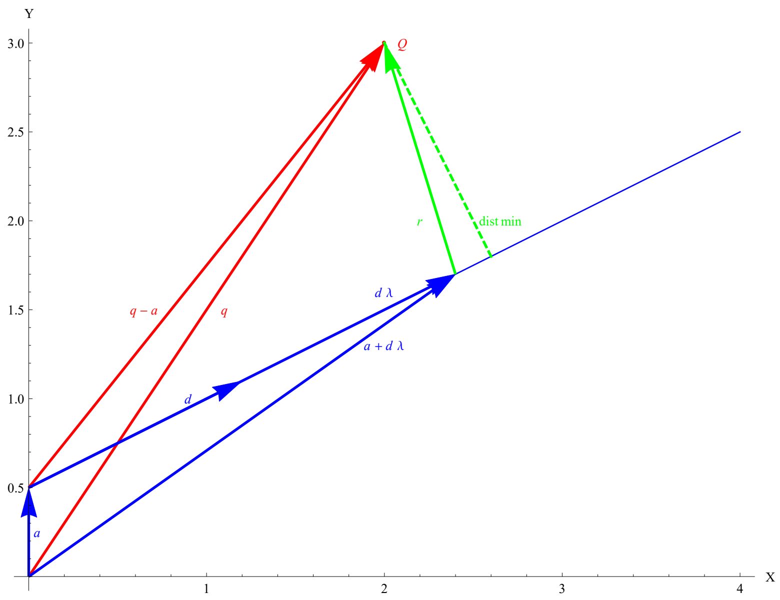


Figure 15: The closest point to Q in W corresponds to the orthogonal projection of $\mathbf{q}$ in the direction of $\mathbf{d}$.

analogous to (7.5).

Pause and look at Fig. 15 for a while: $\mathbf{q} - \mathbf{a}$ would be the vector corresponding to Q in a coordinate system with origin at $\mathbf{a}$, and $\lambda\,\mathbf{d} = \mathbf{w}' - \mathbf{a}$ would be the vector corresponding to the closest point in the same coordinate system. We see, then, that solutions (7.14) and (7.15) are identical to (7.4) and (7.5) but in a coordinate system shifted by $\mathbf{a}$.

This is another instance of the homogeneity of the underlying manifold. Solutions remain the same after translations because in the problems tackled so far there is nothing special about any point. Again, this is trivial once we give ontological primacy to the manifold and vector space rather than the numbers, with the corresponding geometric intuition that comes with it.

HW 7.2: write equations analogous to (7.6) and (7.7) for the case in which the line corresponds to an affine subspace W' of V.

Line embedded in R^n

By now, your geometric intuition should allow you to "see" that a one-dimensional line, and a point Q external to that line, determine a unique plane independently of the dimension of the space in which they are embedded[10]. The machinery we have built to exploit our intuition, for the most part, does not make explicit reference to the dimension of that space. Therefore, it should not be too surprising that equations (7.4), (7.5), (7.6) and (7.7), or (7.14-7.15), are the

[10] If the line passes through the origin, any vector in that line, and the vector $\mathbf{q}$ corresponding to the external point Q, form a basis of that plane. For the case in which the line does not passes through the origin, as we have seen, it is just the translation of this.

solution independently of the dimension of the embedding space!

7.2 Point in a plane closest to a point Q external to it

Plane embedded in R^3

Consider a two-dimensional subspace W of $V = \mathrm{R}^3$ spanned by the not necessarily orthogonal basis $\{\mathbf{x}^1, \mathbf{x}^2\}$. We want to find the point in W closest to a point Q corresponding to a vector $\mathbf{q}$.

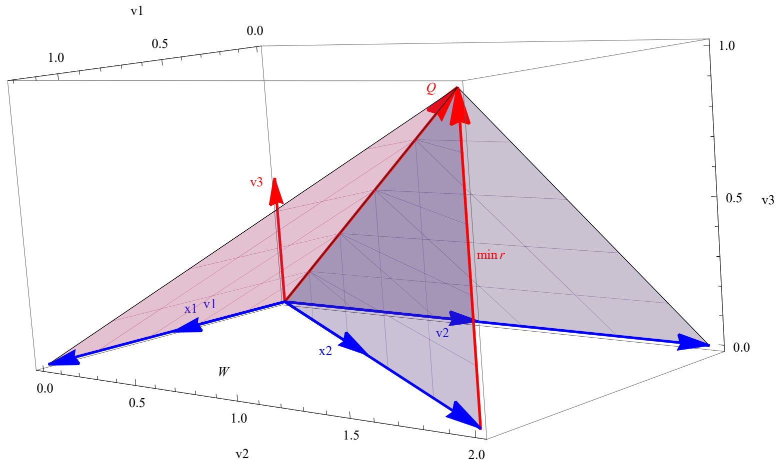


Figure 16: The closest point to Q in W corresponds to the orthogonal projection of $\mathbf{q}$ in W.

Since any point in W can be written as $\mathbf{w} = \lambda_1\mathbf{x}_1 + \lambda_2\mathbf{x}_2$, the closest point will correspond to the values of λ_1 and λ_2 such that

$$\min_{\lambda_1,\lambda_2} \left\| \mathbf{q} - \left(\lambda_1\mathbf{x}^1 + \lambda_2\mathbf{x}^2\right) \right\| \tag{7.16}$$

Since the square of a quantity is an always increasing function of that quantity, the minimization (7.16) is equivalent to the minimization of its square:

$$\min_{\lambda_1,\lambda_2} \left[\mathbf{q} - \left(\lambda_1\mathbf{x}^1 + \lambda_2\mathbf{x}^2\right)\right] \cdot \left[\mathbf{q} - \left(\lambda_1\mathbf{x}^1 + \lambda_2\mathbf{x}^2\right)\right] \tag{7.17}$$

This is a calculus problem of minimization in two variables that can be fairly tedious to solve. But, as in the line problem, simple observation of Fig. 16, in which W corresponds to the horizontal plane, tells us that the solution is the orthogonal projection of $\mathbf{q}$ into W.

The problem is then how to compute this projection. The orthogonal projection into the non-orthogonal basis $\{\mathbf{x}^1, \mathbf{x}^2\}$ won't work: it is clear from Fig. 16 that the vector sum of the projection of $\mathbf{q}$ into $\mathbf{x}^1$ and the projection of $\mathbf{q}$ into $\mathbf{x}^2$ will not give the right answer, because the projection into $\mathbf{x}^2$ has a component in the direction of $\mathbf{x}^1$.

We need an orthogonal basis, like $\{\mathbf{v}^1, \mathbf{v}^2\}$ in Fig. 16, where the projection of $\mathbf{q}$ into any of these vectors has zero component into the other. The solution is then the following three step process:

1. From the non-orthogonal basis $\{\mathbf{x}^1, \mathbf{x}^2\}$ of W, find the orthogonal basis $\{\mathbf{v}^1, \mathbf{v}^2\}$. This can be done with the Gram-Smidth process.

2. Once $\{\mathbf{v}^1, \mathbf{v}^2\}$ has been found, simply project $\mathbf{q}$ into that basis:

$$\mathbf{w} = \sum_{i=1}^{2} \left(\frac{\mathbf{v}^i \cdot \mathbf{q}}{\mathbf{v}^i \cdot \mathbf{v}^i} \right) \mathbf{v}^i \tag{7.18}$$

   $\mathbf{w}$ is the vector corresponding to the point in W closest to Q.

3. The minimum distance is the magnitude of the vector $\mathbf{r}_{\text{min}}$, such that $\mathbf{w} + \mathbf{r}_{\text{min}} = \mathbf{q}$. It's square is:

$$\min \text{dist}^2 (Q, W) = \|\mathbf{r}_{\text{min}}\|^2 = \left\| \mathbf{q} - \sum_{i=1}^{2} \left(\frac{\mathbf{v}^i \cdot \mathbf{q}}{\mathbf{v}^i \cdot \mathbf{v}^i} \right) \mathbf{v}^i \right\|^2 = \mathbf{q} \cdot \mathbf{q} - \sum_{i=1}^{2} \frac{\left(\mathbf{v}^i \cdot \mathbf{q}\right)^2}{\mathbf{v}^i \cdot \mathbf{v}^i} \tag{7.19}$$

   $\mathbf{r}_{\text{min}} \in W^{\perp}$ is a side of the right triangle whose other side is $\mathbf{w}$ of (7.18) and has $\mathbf{q}$ as the hypothenuse. (7.19) is Pythagoras' theorem applied to this triangle.

HW 7.3: Finding the closest distance between two arbitrary lines in R^3, say $\mathbf{r}^1 = \mathbf{a}^1 + \lambda_1 \mathbf{d}^1$ and $\mathbf{r}^2 = \mathbf{a}^2 + \lambda_2 \mathbf{d}^2$, is a fairly complicated problem of calculus. a) Choose numerical vectors for $\mathbf{a}^i$ and $\mathbf{d}^i$, $i = 1, 2$, and do a 3-D graph of the two lines, and a vector that starts at a chosen point in line 1 and finishes at another chosen point in line 2. Appreciate in this graph that the problem is far from trivial. b) Solve the problem by mapping it into the problem of finding the closest point to a plane and applying the 3-steps solution. c) Pause and ponder about what you have just done: you mapped a complicated problem into a completely different problem. You knew the solution of this second problem, therefore you solved the first problem.

Plane embedded in R^n

If the two-dimensional surface is embedded in R^n, one would think that the problem gets far more complicated. But this is not the case with the machinery we have built. In fact, the 3-step process described above gives the solution in any dimension.

HW 7.4: Explain why.

7.3 Point in an r-dimensional subspace W closest to a point $Q \in V = \mathrm{R}^n$, $r < n$

The r-dimensional subspace is spanned by a linearly independent, not necessarily orthogonal basis $\{\mathbf{x}^i\}$, $i = 1, \cdots, r$. Any point in W can be written as $\mathbf{w} = \sum_{i=1}^{r} \lambda_i \mathbf{x}^i$; the closest point will correspond to the values of λ_i such that

$$\min_{\lambda_1, \cdots, \lambda_r} \left\| \mathbf{q} - \sum_{i=1}^{r} \lambda_i \mathbf{x}^i \right\| \tag{7.20}$$

The three-step process that worked for the projection of Q into the plane spanned by $\{\mathbf{x}^i\}$, $i = 1, 2$, also works, with a trivial extension, for the projection into an r dimensional subspace spanned by $\{\mathbf{x}^i\}$, $i = 1, \cdots, r$:

1. From the non-orthogonal basis $\{\mathbf{x}^1, \cdots, \mathbf{x}^r\}$ of W, find the orthogonal basis $\{\mathbf{v}^1, \cdots, \mathbf{v}^r\}$. This can be done with the Gram-Smidth process.

2. Once $\{\mathbf{v}^1, \cdots, \mathbf{v}^r\}$ has been found, simply project $\mathbf{q}$ into that basis:

$$\mathbf{w} = \sum_{i=1}^{r} \left(\frac{\mathbf{v}^i \cdot \mathbf{q}}{\mathbf{v}^i \cdot \mathbf{v}^i} \right) \mathbf{v}^i \tag{7.21}$$

   $\mathbf{w}$ is the vector corresponding to the point in W closest to Q.

3. The minimum distance is the magnitude of the vector $\mathbf{r}_{\min}$, such that $\mathbf{w} + \mathbf{r}_{\min} = \mathbf{q}$. The square of the minimum distance is:

$$\min \operatorname{dist}^2(Q, W) = \|\mathbf{r}_{\min}\|^2 = \left\| \mathbf{q} - \sum_{i=1}^{r} \left(\frac{\mathbf{v}^i \cdot \mathbf{q}}{\mathbf{v}^i \cdot \mathbf{v}^i} \right) \mathbf{v}^i \right\|^2 = \mathbf{q} \cdot \mathbf{q} - \sum_{i=1}^{r} \frac{\left(\mathbf{v}^i \cdot \mathbf{q}\right)^2}{\mathbf{v}^i \cdot \mathbf{v}^i} \tag{7.22}$$

   $\mathbf{r}_{\min} \in W^{\perp}$ is a side of the right triangle whose other side is $\mathbf{w}$ of (7.21) and has $\mathbf{q}$ as the hypothenuse. (7.22) is Pythagoras' theorem applied to this triangle.

Pause and ponder about what we have just achieved: almost effortlessly we have found the closest point in a subspace or r dimensions of a point that "lives" in n dimensions. n could be $1,000,000$, and r could be, for example, $299,891$[11], and the solution simply works. And we can confidently talk about Pythagoras' theorem, projections, etc. This is the power of a formalism that on the one hand exploits deeply ingrained spatial intuitions, and on the other it is sufficiently abstract so that these generalizations are possible.

7.4 Regressions and geometry

Least square regressions and its extensions rank among the most thoroughly used econometric techniques. Let us consider the simplest case of linear regression with one independent variable, x, and one dependent variable y. We typically have n pairs of empirical data (x_i, y_i), as in table 1.

[11] Just for fun, I chose the highest prime number smaller than 300.000.

Table 1: Regression data

X	Y
0.00	3.54
0.65	5.13
0.44	2.58
0.94	2.38
0.53	3.42
0.17	-1.23
0.09	-0.61
0.45	-2.07
0.95	3.34
0.67	-0.45

The objective is to find the “best" line $y = \lambda_1 + \lambda_2\, x$ that fits the data, see figure 17. In this figure,

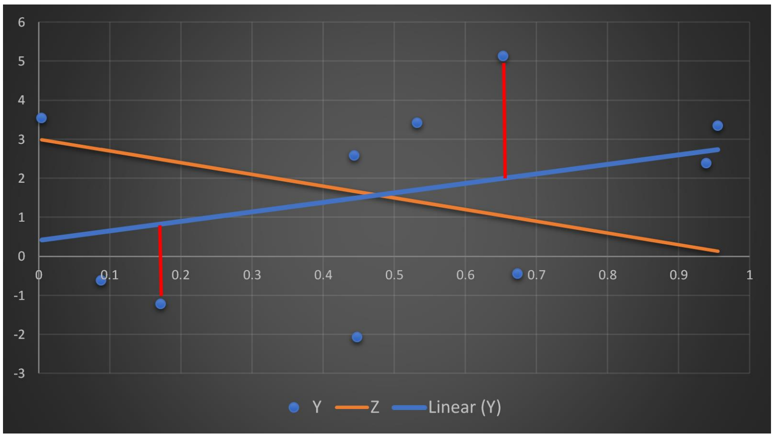


Figure 17: The blue dots are our data. The blue line is the “best" straight line for this data (compare it, for example, with the orange line Z). “Best" means the line that minimizes the sum of the squares of the “errors" (red bars).

every pair of data points in the table is represented as a blue dot. It is intuitively obvious that the blue line is a better fit to the data than, for example, the orange line Z. But what do we mean exactly by “better fit"?

We model the data with a straight line like $y = \lambda_1 + \lambda_2 x$. The y_i values of our data will be given by the value of the straight line plus “errors":

$$y_i = \lambda_1 + \lambda_2 x_i + \epsilon_i, \quad i = 1, \cdots, n \tag{7.23}$$

The problem is to find λ_1 and λ_2 such that the “errors" ϵ_i are somehow minimized. In figure 17, the errors of two points are marked with red bars. As can be seen, some errors are positive, and others are negative.

A simple criterion for “best" fit is to choose the parameters λ_1 and λ_2 in such a way that the sum of the *squares* of the “errors" is minimized, so that the sign of the errors is irrelevant:

$$\min_{\lambda_1,\lambda_2} S = \min_{\lambda_1,\lambda_2} \sum_{i=1}^{n} \epsilon_i^2 = \min_{\lambda_1,\lambda_2} \sum_{i=1}^{n} [y_i - (\lambda_1 + \lambda_2\, x_i)]^2 \tag{7.24}$$

It is not the purpose of this section to enter into the statistical fundamentals of the method; I just want to point out that there is a nice geometric interpretation of this problem that makes the solution much easier to visualize.

Consider a vector space V of dimension n, equal to the number of pairs of empirical data, and an orthonormal basis $\{\hat{\mathbf{e}}^i\}$ in V. Consider the following four vectors:

$$\mathbf{x} = \sum_{i=1}^{n} x_i\, \hat{\mathbf{e}}^i \tag{7.25}$$

$$\mathbf{1} = \sum_{i=1}^{n} 1\, \hat{\mathbf{e}}^i \tag{7.26}$$

$$\mathbf{w} = \lambda_1\, \mathbf{1} + \lambda_2\, \mathbf{x} = \sum_{i=1}^{n} (\lambda_1 + \lambda_2 x_i)\, \hat{\mathbf{e}}^i \tag{7.27}$$

$$\mathbf{y} = \sum_{i=1}^{n} y_i\, \hat{\mathbf{e}}^i \tag{7.28}$$

where the values of x_i and y_i are given by our data. For the particular example of Fig. 17, they are given in table 1.

The of sum of the squares of the errors (7.24) is nothing but the square of the modulus of the “error vector" $\epsilon \equiv \mathbf{y} - \mathbf{w}$:

$$\begin{aligned}
\epsilon \cdot \epsilon &= (\mathbf{y} - \mathbf{w}) \cdot (\mathbf{y} - \mathbf{w}) \\
&= (\mathbf{y} - \lambda_1\, \mathbf{1} - \lambda_2\, \mathbf{x}) \cdot (\mathbf{y} - \lambda_1\, \mathbf{1} - \lambda_2\, \mathbf{x}) \\
&= \left(\sum_{i=1}^{n} (y_i - \lambda_1 - \lambda_2 x_i)\, \hat{\mathbf{e}}^i \right) \cdot \left(\sum_{j=1}^{n} (y_j - \lambda_1 - \lambda_2 x_j)\, \hat{\mathbf{e}}^j \right) \\
&= \sum_{i,j=1}^{n} \left(y_i - \lambda_1 - \lambda_2 x_i\right) \left(y_j - \lambda_1 - \lambda_2 x_j\right) \hat{\mathbf{e}}^i \cdot \hat{\mathbf{e}}^j \\
&= \sum_{i=1}^{n} \left(y_i - \lambda_1 - \lambda_2 x_i\right)^2 \\
&= S
\end{aligned} \tag{7.29}$$

therefore, the problem of minimizing (7.24) is equivalent to minimizing (7.29):

$$\min_{\lambda_1,\lambda_2} S = \min_{\lambda_1,\lambda_2} \sum_{i=1}^{n} \left(y_i - \lambda_1 - \lambda_2 x_i\right)^2 = \min_{\lambda_1,\lambda_2} (\mathbf{y} - \lambda_1\, \mathbf{1} - \lambda_2\, \mathbf{x}) \cdot (\mathbf{y} - \lambda_1\, \mathbf{1} - \lambda_2\, \mathbf{x}) \tag{7.30}$$

but the right hand side of (7.30) is equivalent to the problem (7.17) with

$$\mathbf{q} \leftrightarrow \mathbf{y} \tag{7.31}$$

$$\mathbf{x}^1 \leftrightarrow \mathbf{1} \tag{7.32}$$

$$\mathbf{x}^2 \leftrightarrow \mathbf{x} \tag{7.33}$$

We conclude that if one has n pairs of data points (x_i, y_i), the least square linear regression is equivalent to finding in R^n the closest point to the plane spanned by the vectors $\mathbf{x}^1 = \mathbf{1}$ (7.26), and $\mathbf{x}^2 = \mathbf{x}$ (7.25). Equation (7.18) gives the vector $\mathbf{w}$ in this plane closest to $\mathbf{y}$.

It may seem surprising at first sight, that the seemingly two dimensional problem illustrated in Fig. 17 has a much deeper and fruitful interpretation in the space of n dimensions where these vectors "live", where n is the number of data points, in our example $n = 10$. This is just one example of how one can map a problem into another to find previously unimagined relationships and solutions.

The calculation of the vector $\mathbf{w}$ is a big step towards the solution of our regression problem. However, (7.18) expresses $\mathbf{w}$ in the basis $\{\mathbf{v}^i\}$, but we want λ_1 and λ_2 of equation (7.23). According to equation (7.27), these coefficients are the components of $\mathbf{w}$ in the basis $\{\mathbf{1}, \mathbf{x}\}$, not in the $\{\mathbf{v}^i\}$ basis! How do we find them?

One could express the basis vectors $\mathbf{I}$ and $\mathbf{x}$ in terms of the $\mathbf{v}^1$ and $\mathbf{v}^2$ by projecting them in this orthogonal basis. Once this is done, one could "invert" these and express $\mathbf{v}^1$ and $\mathbf{v}^2$ in terms of $\mathbf{I}$ and $\mathbf{x}$. Finally, inserting these expressions in equation (7.18) and rearranging, one would obtain the coefficients λ_1 and λ_2 that we are looking for.

However this procedure is tedious, especially when there are many explanatory (independent) variables. It also obscures the essential fact described above that a regression problem is simply a projection problem. The ultimate reason it is tedious is that working with nonorthogonal bases is a mess in general. In contrast, orthogonal basis are neat, because the projection on one vector has zero components in the projection on another vectors. This enormously simplifies our calculations.

Unfortunately, real world problems almost always present us with nonorthogonal vectors. This is one reason why it would be great if we could figure out a formalism to work with nonorthogonal bases as comfortably as we work with orthogonal ones. It turns out that such a formalism exists!

Moreover, as we will see, other regressions parameters such as the correlation coefficients, r^2 goodness tests, etc., also have neat geometrical interpretations that help us not to get lost when the complexity of the problem increases. We leave the mentioned formalism, these interpretations, and many others, for the next works in this series.

8 Conclusions

We have presented the fundamental building blocks of linear algebra: linear manifolds, vector spaces, scalar products, and many applications. These concepts lie at the basis of almost all the techniques used in mathematical economics, and the mathematics used in data analysis. However, the extremely important topics of linear mappings have not yet been covered. These will be the subject of the next two papers in this series. As we will see, with the help of the structures and intuitions presented here, linear mappings, and their many connected properties, will become far more transparent.

Acknowledgments

I want to thank Anthony Duncan, of the University of Pittsburgh, for helpful comments about content and editing.

References

[3B1B LA1] , Grant Sanderson, “Vectors, what even are they?", *Essence of linear algebra*, chapter 1, https://www.youtube.com/watch?v=fNk_zzaMoSs&t=160s.

[3B1B LA2] ,Grant Sanderson, “Linear combinations, span, and basis vectors", *Essence of linear algebra*, chapter 2, https://www.youtube.com/watch?v=k7RM-ot2NWY&list=PLZHQObOWTQDPD3MizzM2xVFitgF8hE_ab&index=3&t=0s.