Geometric Definition of Linear Transformations

Eduardo Magalhães

May 18, 2020

Abstract

This note motivates an alternative definition for linear transformations based on geometric and differential properties of linear maps. Let $V$ and $W$ be two vector spaces and let $T : V \to W$ be a function between $V$ and $W$. $T$ is a linear map if:

$$T(0) = 0$$

$$H(T) = 0$$

Where $H(T)$ is the Hessian tensor of the function $T$.

1 Introduction

Linear transformations are an essential idea, and the focus of the study of linear algebra. Let $V, W$ be vector spaces over a field $K$, and let $T : V \to W$ be a function between $V$ and $W$. Then $T$ is linear if:

$$T(v + w) = T(v) + T(w)$$

$$T(\alpha v) = \alpha T(v)$$

with $v, w \in V$ and $\alpha \in K$. This definition comes from the definition and from the notion of linearity. As a consequence of these 2 properties, comes 2 important results: for all linear transformations, the origin remains fixed in place (1) and, if we apply the transformation to any straight line, after the transformation, that line stays straight (in other words, there is no curvature) (2). If these conditions are not true for some transformation, then that transformation is not linear. This gives us another way to define linear transformations, not based on the definition of linearity, but based on the geometric and graphical behavior of these types of functions.
2 Content

Let $V, W$ be vector spaces over a field $K$, and let $T : V \rightarrow W$ be a function between $V$ and $W$. We can translate property (1) into mathematical notation quite easily: $T(0) = 0$, but to write property (2), we will use the hessian tensor. If we have a function $f : \mathbb{R} \rightarrow \mathbb{R}$ we normally use the second derivative to get an idea of the curvature of the function. If $f''(a) \neq 0$ we say that the function is curved at that point, and we even use the value of the second derivative to determine if a saddle point is a local maximum or minimum based on the curvature and shape of the function. So if $f''(a) = 0$, there is no curvature at that point. So we need that the “second derivative” of $T$ equals 0. We can express this “second derivative” of our function with the Hessian tensor which encapsulates all information about all second partial derivatives of the function and it’s basically a generalization of the second derivative.

Proposition 1. Let $V$ and $W$ be two vector spaces such that $\dim V = n$ and $\dim W = m$. If $T : V \rightarrow W$ is a function between $V$ and $W$ such that:

$$T(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix}$$

with $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

Then $T$ is a linear transformation only if:

$$T(0) = 0 \quad (1)$$

$$H(T) = 0 \quad (2)$$

Where $H(T)$ is the Hessian tensor of the function $T$.

3 Proof

We will start by proving that, if a function satisfies conditions (1) and (2), then it is a linear map, which we will call part 1. And then we will start from the regular definition of linear transformation and prove that conditions (1) and (2) are also satisfied. Thus, proving that the statements “a function $f$ is a linear transformation” and “conditions (1) and (2) are true” are equivalent.
Part 1

Let $T$ be a function between two vector spaces $V$ and $W$ with $\dim V = n$ and $\dim W = m$ such that:

$$T(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix}$$

with $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

Let’s suppose that $T$ satisfies conditions (1) and (2) and let’s show that if this happens, then $T$ is linear.

We have that:

$$H(T) = 0$$

So every component of the Hessian tensor has to be equal to 0. If we denote each component of the tensor as $H_{\gamma ij} = \partial_{x_i x_j} g_\gamma$:

$$H_{\gamma ij} = 0 \iff \partial_{x_i x_j} g_\gamma = 0, \quad \gamma = 1, ..., m$$

Now we need to find every $g_\gamma$ that satisfies this equation. Let’s first integrate both sides with respect to $x_j$:

$$\int \partial_{x_i x_j} g_\gamma \, dx_j = \int 0 \, dx_j \iff

\partial_{x_i} g_\gamma = \varphi_{i\gamma}(x\{x_j\})$$

where $x\{x_j\}$ denotes a column matrix with $n - 1$ lines that consists of matrix $x$ without the line that contains $x_j$.

Because $\varphi_{i\gamma}(x\{x_j\})$ never explicitly depends on $x_j$ then it’s partial derivative with respect to $x_j$ is 0.

So we have that:

$$\partial_{x_i} g_\gamma = \varphi_{i\gamma}(x\{x_j\})$$

Let’s fix $i$ and $\gamma$. This statement is true for every $j \in \{1, ..., n\}$. Basically the following needs to be true:
\[ \varphi_{i\gamma}(x\{x_1\}) = \varphi_{i\gamma}(x\{x_2\}) = \ldots = \varphi_{i\gamma}(x\{x_j\}) \]

Because all these functions are the same, for any value of \( j \), they all need to explicitly depend on the same variables. For example \( \varphi_{i\gamma}(x\{x_1\}) \) and \( \varphi_{i\gamma}(x\{x_2\}) \) are the same function (\( \varphi_{i\gamma} \)) so \( \varphi_{i\gamma}(x\{x_1\}) \) and \( \varphi_{i\gamma}(x\{x_2\}) \) need to explicitly depend on the same variables. Let \( j = 1 \), then \( \varphi_{i\gamma} \) can’t explicitly depend on \( x_1 \). If we let \( j = 2 \), then \( \varphi_{i\gamma} \) can’t explicitly depend on \( x_1 \) and \( x_2 \). If we continue this for every \( j \) and we continue to “sum” all this restrictions to the domain of \( \varphi_{i\gamma} \) we find out that \( \varphi_{i\gamma} \) can’t explicitly depend on any \( x_j \). So, because \( \varphi_{i\gamma} \) does not depend on no variables, \( \varphi_{i\gamma} \) must be a constant.

Thus:

\[ \varphi_{i\gamma}(x\{x_j\}) = C_{i\gamma} \]

Giving us:

\[ \partial_{x_i} g_{\gamma} = C_{i\gamma} \]

Integrating both sides with respect to \( x_i \):

\[ \int \partial_{x_i} g_{\gamma} \, dx_i = \int C_{i\gamma} \, dx_i \iff \]

\[ g_{\gamma} = C_{i\gamma} x_i + \mu_{i\gamma}(x\{x_i\}) \]

Again, if we fix a value of \( \gamma \), this statement needs to be true for every \( i \in \{1, \ldots, n\} \).

\[ g_{\gamma} = C_{1\gamma} x_1 + \mu_{1\gamma}(x\{x_1\}) \]

\[ g_{\gamma} = C_{2\gamma} x_2 + \mu_{2\gamma}(x\{x_2\}) \]

\[ \vdots \]

\[ g_{\gamma} = C_{n\gamma} x_n + \mu_{n\gamma}(x\{x_n\}) \]

Let’s look first at what the term \( C_{i\gamma} x_i \) can tell us about the function \( g_{\gamma} \). Again, all these functions are the same, so they explicitly depend on the same variables. Just by looking at that first term, if we let \( i = 1 \), we see that \( g_{\gamma} \) does explicitly depend on \( x_1 \). To be more precise, \( g_{\gamma} \) has a term with \( C_{1\gamma} x_1 \). If we do the same with \( i = 2 \), because they are the same function, \( g_{\gamma} \) has to have a term with \( C_{1\gamma} x_1 + C_{2\gamma} x_2 \), and thus explicitly depending on \( x_1 \) and \( x_2 \). If we continue
this for all $i$ we get that $g_i$ need to have a term with $C_1 \gamma x_1 + C_2 \gamma x_2 + \ldots + C_n \gamma x_n$, so $g_i$ depends explicitly depend on every $x_i$.

In order for every $g_\gamma$ to have a term with $C_1 \gamma x_1 + C_2 \gamma x_2 + \ldots + C_n \gamma x_n$, then we need:

$$\mu_\gamma(x \setminus \{x_i\}) = A + \sum_{\rho \in \{1, \ldots, n\} \setminus \{i\}} C_\rho \gamma x_\rho$$

Where $A$ is an arbitrary constant.

Because that way:

$$g_\gamma = C_{i\gamma} x_i + \mu_\gamma(x \setminus \{x_i\}) \iff$$

$$g_\gamma = C_{i\gamma} x_i + A + \sum_{\rho \in \{1, \ldots, n\} \setminus \{i\}} C_\rho \gamma x_\rho \iff$$

$$g_\gamma = A + \sum_{\rho} C_\rho \gamma x_\rho$$

So we found an expression for every function $g_\gamma$. Now we need to find the value of that constant $A$.

We also know that $T$ satisfies condition (2), so:

$$T(0) = 0 \iff$$

$$g_\gamma(0) = 0 \iff$$

$$A = 0$$

So we have that:

$$g_\gamma = \sum_{\rho} C_\rho \gamma x_\rho$$

Thus:

$$T(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix} = \begin{pmatrix} \sum_{\rho} C_{\rho 1} x_\rho \\ \vdots \\ \sum_{\rho} C_{\rho m} x_\rho \end{pmatrix} = \begin{pmatrix} C_{11} & \cdots & C_{n1} \\ \vdots & \ddots & \vdots \\ C_{1m} & \cdots & C_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

So if a function satisfies conditions (1) and (2), then it is a linear map.
Part 2

Now we are going to start from a Linear transformation and show that every linear transformation satisfies conditions (1) and (2).

Let $T$ be a linear transformation between two vector spaces $V$ and $W$ with $\dim V = n$ and $\dim W = m$ such that:

$$T(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix}$$

with $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$

Let’s start by proving that $T(0) = 0$. Let $v \in V$:

$$T(0) = T(v - v) = T(v) + T(-v) = T(v) - T(v) = 0$$

Thus proving that any linear transformation satisfies conditions (1).

Now, for condition (2), let’s just simply compute the Hessian tensor of the linear transformation and show that it is equal to the zero tensor. If we denote each component of the Hessian tensor as $H_{\gamma ij} = \partial_{x_i x_j} g_{\gamma}$. If $T$ is a linear map then we can express $T$ with a matrix:

$$T(x) = \begin{pmatrix} C_{11} & \cdots & C_{n1} \\ \vdots & \ddots & \vdots \\ C_{1m} & \cdots & C_{nm} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

So we have that:

$$T(x) = \begin{pmatrix} \sum_{\rho} C_{\rho 1} x_{\rho} \\ \vdots \\ \sum_{\rho} C_{\rho m} x_{\rho} \end{pmatrix} = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix}$$

Thus:

$$g_{\gamma} = \sum_{\rho} C_{\rho \gamma} x_{\rho}$$

Now that we know $g_{\gamma}$ we can compute every component of the Hessian Tensor:

$$H_{\gamma ij} = \partial_{x_i x_j} \sum_{\rho} C_{\rho \gamma} x_{\rho} =$$
\[ \sum_{\rho} C_{\rho\lambda} \partial_{x_i x_j} x_{\rho} = \sum_{\rho} C_{\rho\lambda} \cdot 0 = 0 \]

So every component of the Hessian Tensor is zero, therefore \( \mathbf{H}(T) = 0 \). Thus proving that any linear transformation satisfies conditions (1) and (2).

Part 1 and part 2 together prove that \( T \) is a linear transformation only if conditions (1) and (2) satisfied, thus proving that this is a valid alternative definition of linear transformation.