

# A way to introduce tensors in a linear algebra course

Pavel Grinfeld

## Abstract

The tensor notation is useful in several branches of science, including linear algebra, but it is difficult to teach since the learning curve is initially very steep. The problem of quadratic form minimization provides a welcome entry point. Having seen the algebra carried out without tensors, students have a clear picture of the operations denoted by the tensor symbols. Further, students will appreciate the conciseness and the structure of the tensor expressions.

## 1 Introduction

The tensor notation [1], [2] is a popular choice in several fields, including continuum mechanics, fluid dynamics and gravitation. In linear algebra, the tensor notation is also very effective as it reveals the covariant/contravariant nature of objects, emphasizes invariance, and offers mnemonic rules for the operations required to carry out a change of basis. The tensor notation offers an alternative view on a number of concepts and is a useful teaching tool.

However, while the tensor notation provides many benefits when teaching other topics, it is quite difficult to teach the tensor notation itself! The problem lies in the initial hump that the student needs to overcome. The emphasis on indices is very different from what the students are used to. The notation is very compressed ( $\delta_{ijk}^{ijk} = 6?$ ) and it takes time to learn to recognize what an expression is saying. The distinction between covariant and contravariant objects is a new concept for most linear algebra students. These factors combine to present the student with a formidable "barrier to entry". This is especially true for a linear algebra course where no more than a lecture or two could be devoted to this subject.

In my experience, many students welcome the tensor notation once they get over the initial hump. For a few of them, the language of tensors becomes a personal favorite. Tensors, therefore, deserve a chance!

## 2 Quadratic form minimization

In this short write-up, I would like to describe what in my teaching experience is a good place to present the tensor notation in a linear algebra course. I have

found that the best topic for doing so is quadratic form minimization. The problem is to minimize

$$f(x) = \frac{1}{2}x^T Ax - x^T b \quad (1)$$

with respect to a vector  $x$ , for a symmetric positive definite matrix  $A$  and an arbitrary vector  $b$ . The optimal  $x$  is given by the equation

$$Ax = b, \quad (2)$$

but the calculus derivation of this equation cannot be carried out without direct access to the entries of the matrix  $A$ . There is a short non-calculus derivation that introduces the new variable  $y = x - A^{-1}b$  and it demonstrates that  $y = 0$  is optimal but, in my opinion, that derivation is not a good way to present the subject since it does not show how the answer could be arrived if not known beforehand.

The calculus approach – which is the proper way to go, since the students need to know how to apply calculus to linear algebra problems – is typically an unpleasant experience. This is because by the time least squares are introduced in a linear algebra course, the students have just spent most of the semester learning to think of  $A$  as an integral object; yet calculus requires access to the individual elements of the matrix. It is also seen that the algebra that follows is quite cumbersome. It is so cumbersome, in fact, that I would argue that tensors are necessary in order to teach this topic effectively. Some textbooks use the "after some straightforward but lengthy algebra..." trick, but for most students, the algebra is by no means straightforward. The general  $N \times N$  case is almost impossible to present without some elements of tensor algebra. Indeed, consider the general  $N$ -dimensional quadratic form, written out explicitly:

$$f(x_1, \dots, x_n) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j - \sum_{i=1}^n x_i b_i \quad (3)$$

The next step in the calculus approach is to take a partial derivative with respect to  $x_k$ . (At this point most of the students will be confused why  $x_k$  and not  $x_i$  or  $x_j$ ). But what is  $\partial x_i / \partial x_k$ ? It is a question that is difficult to answer without at least some elements of tensor calculus.

The common approach is to avoid the treatment of the general case and to instead consider a  $2 \times 2$  or a  $3 \times 3$  example. It is certainly a good idea to start this way and for most of the students these examples will be sufficiently convincing. But these students would miss out on learning how to combine calculus with linear algebra. Nevertheless these short examples (actually, the  $3 \times 3$  example is not so short!) are extremely useful. In fact, I would suggest going even further and starting with a one dimensional example. Ordinary calculus can be used to show that  $f(x) = \frac{1}{2}ax^2 - bx$  leads to  $ax = b$ . This trivial example will remind the students that calculus is the appropriate tool for finding extremal values of a function and will also demonstrate to them what to expect as the eventual answer.

When I teach this subject I typically go through the trouble of demonstrating the one dimensional, two dimensional, and three dimensional cases. If I may give you a recommendation: practice the  $3 \times 3$  case beforehand – it may be a whole lot more writing than you might expect!

Having seen these three relatively simple examples, the students are ready for the general  $N \times N$  case. They already know what the answer will be so the purpose of the whole presentation is to demonstrate how to combine calculus with linear algebra – and a good opportunity to introduce tensors.

### 3 Introduction to tensors

I decided to write this section as if I am addressing my students. But I will omit most of the details.

The starting point is the general  $N$ -dimensional quadratic form

$$f(x_1, \dots, x_n) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j - \sum_{i=1}^n x_i b_i \quad (4)$$

We begin by taking several steps towards condensing the notation. We first write the  $x$ 's with upper indices ( $x^i, x^j$ ) rather than lower indices. The choice of whether to use upper indices (superscripts) or lower indices (subscripts) has great algebraic significance, but it is not important here. Second, we adopt the Einstein notation that *summation over an index is implied if it appears once as a subscript and once as a superscript* [3]. Therefore, no summation signs are ever used for operations such as matrix multiplication, calculating traces and computing determinants. Finally, we write  $f(x)$  in place of  $f(x_1, \dots, x_n)$  since access to individual arguments will not be required on the left hand side. We arrive at the following compact expression:

$$f(x) = \frac{1}{2} A_{ij} x^i x^j - b_i x^i. \quad (5)$$

This is probably your first experience with a scientific contribution by Einstein. It may not *seem* like a very important one, but that is not so. Einstein himself valued this invention of his. Years after introducing this notation, he wrote in a letter to a friend [4]: "I have made a great discovery in mathematics; I have suppressed the summation sign every time that the summation must be made over an index which occurs twice..." Despite the whimsical tone of the letter, Einstein is obviously quite pleased with his discovery.

Getting back to the problem at hand, note that the above expression, the order of the multiplicative term no longer matters. This is because we are now multiplying together numbers rather than more complicated objects such as matrices, for which the order does matter. There is also no operation of the transpose. These two essential aspects of matrix algebra are automatically "enforced" by virtue of representing matrix multiplications with indices.

Introduce the Kronecker symbol  $\delta_j^i$ :

$$\delta_j^i = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

This symbol is probably familiar to you, but this is probably the first time you are seeing it with a superscript and a subscript.

**Exercise 1** Show that  $\delta_j^i = N$ , where  $N$  is the dimension of the linear space.

**Exercise 2** Show that  $\delta_j^i \delta_k^j = \delta_k^i$ .

**Exercise 3** Show that  $\delta_j^i \delta_i^j = N$ .

**Exercise 4** Show that  $A_{ij} \delta_k^j = A_{ik}$ .

Use the Kronecker delta symbol to express the fact that the partial derivative of a function  $g(x) = x^i$  with respect to  $x^j$  is 1 when  $i = j$  (the two variables are the same) and 0 otherwise (the two variables are different):

$$\frac{\partial x^i}{\partial x^j} = \delta_j^i$$

Now, the derivation of  $Ax = b$  can be carried out as follows:

$$\frac{\partial f(x)}{\partial x_k} = \frac{1}{2} \frac{\partial (A_{ij} x^i x^j)}{\partial x_k} - \frac{\partial (b_i x^i)}{\partial x_k} \quad (6a)$$

$$\text{(Product Rule)} = \frac{1}{2} \left( A_{ij} \delta_k^i x^j + A_{ij} x^i \delta_k^j \right) - \delta_k^i b_i \quad (6b)$$

$$\text{(Exercise above)} = \frac{1}{2} (A_{kj} x^j + A_{ik} x^i) - b_k \quad (6c)$$

$$\text{(Rename } j \rightarrow i) = \frac{1}{2} (A_{ki} x^i + A_{ik} x^i) - b_k \quad (6d)$$

$$\text{(Since } A_{ik} = A_{ki}) = A_{ki} x^i - b_k = 0 \quad (6e)$$

Rewriting the resulting equation in the matrix notation we get

$$Ax = b.$$

Although the tensor derivation above is quite concise – and could be even more brief once you have mastered the technique – it may not be easy to accept for someone who is seeing the notation for the first time. It is important to distinguish between *dummy* indices, such as  $i$  and  $j$  above and *free* indices, such as  $k$ . Initially, I recommend to students that they mentally "unwrap" all expressions at each step to see the real arithmetic behind the mechanics.

## 4 Where to go from here

Another great illustration of the benefits of the tensor notation in the context of linear algebra is determinants. A great treatment can be found in McConnell [1]. It is another instance where unwieldy expressions can be made quite compact.

Of course the *best* subject for the tensor notation in linear algebra is *change of basis*. This topic illustrates the reason for introducing covariant and contravariant tensors (upper indices vs. lower indices) and presents the students with a fresh perspective on many operations in linear algebra. A demonstration of the dot product, the metric tensor, and the operation of lowering and raising indices I found to be an effective way of completing an introduction to the tensor notation.

As a final note, I would like to suggest a few exercises.

**Exercise 5** Consider a multivariable function  $f(x)$  of  $x^i$  and let each  $x^j$  be a multivariable function of  $y^j$ . Construct the composite function  $F(y)$  of  $y^i$  defined as (dropping the indices for function arguments, as always):

$$F(y) = f(x(y))$$

Show that the chain rule can be written in tensor notation as

$$\frac{\partial F}{\partial y^j} = \frac{\partial f}{\partial x^i} \frac{\partial x^i}{\partial y^j}$$

**Exercise 6** Apply the above formula to the function  $f(x) = (x^1)^2 + (x^2)^2 + (x^3)^2$  and  $x^1(y) = y^1 - y^2$ ,  $x^2(y) = y^2 - y^3$ , and  $x^3(y) = (y^1)^2$ .

**Exercise 7** Let  $x^i$  and  $y^i$  be two sets of  $N$  independent variables. Let the letters  $x$  and  $y$  also indicate the functional dependence between  $x^i$  and  $y^i$ :  $x^i \equiv x^i(y)$ ,  $y^i \equiv y^i(x)$ . (Note that the index is dropped when the variable is used as an argument to a multivariable function.) Show that

$$\frac{\partial x^i}{\partial y^k} \frac{\partial y^k}{\partial x^j} = \delta_j^i.$$

**Exercise 8** Show that

$$\frac{\partial x^i}{\partial y^k} \frac{\partial y^k}{\partial x^i} = N.$$

**Exercise 9** Demonstrate the above identity using  $x^1 = y^1 + y^2 + y^3$ ,  $x^2 = y^2 + y^3$ , and  $x^3 = y^3$  and  $y^1 = x^1 - x^2$ ,  $y^2 = x^2 - x^3$ , and  $y^3 = x^3$ .

**Exercise 10** Show that  $|Ax - b|$  is minimized by  $x$  that satisfies  $A^T Ax = A^T b$ .

## References

- [1] A.J. McConnell. Application of Tensor analysis. Dover Publications (1957)
- [2] T. Levi-Civita. The Absolute Differential Calculus: Calculus of Tensors. Dover Publications (August 26, 2005)
- [3] A. Einstein, "Die Grundlage der allgemeinen Relativitätstheorie." Ann. der Physik 49, 769-822, 1916.
- [4] Kollros 1956; Pais 1982, p. 216