Confusion about what the Chain Rule says PhL 3.26.16

See summary at the end.

Motivation: In considering this example of a tall R matrix,

$x'^1 = x^1$
$x'^2 = x^2$
$x'^3 = (R^2 - (x^1)^2 - (x^2)^2)^{1/2}$ = height of dome at (x^1, x^2) // forward xform

$x^1 = x'^1$
$x^2 = x'^2$ // inverse xform

I find by explicit calculation that

```
R := matrix(3,2,R_);
```

$$R := \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -\frac{x_1}{\sqrt{4 - x_1^2 - x_2^2}} & -\frac{x_2}{\sqrt{4 - x_1^2 - x_2^2}} \end{bmatrix}$$

```
S := matrix(2,3,[1,0,0,0,1,0]);
```

$$S := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

```
evalm(R &* S);
```

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{x_1}{\sqrt{4 - x_1^2 - x_2^2}} & -\frac{x_2}{\sqrt{4 - x_1^2 - x_2^2}} & 0 \end{bmatrix}$$

```
evalm(S &* R);
```

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

When I write out the RS product I get

$$R^i{}_j S^j{}_k = (\partial x'^i / \partial x^j)(\partial x^j / \partial x'^k) = \frac{\partial x'^i}{\partial x^j}\frac{\partial x^j}{\partial x'^k} \neq \delta^i{}_k \qquad RS \neq 1$$

I am surprised, because I thought I had this theorem going

Theorem: The Chain Rule says $\frac{\partial x'^i}{\partial x^j}\frac{\partial x^j}{\partial x'^k} = \delta^i{}_k$

So how do I explain that "the chain rule fails"? Totally new to me. How do you prove the chain rule in the first place? Bucks do a very long proof starting on page 250 for the case that you have the same number of variables on each side. It is a non-trivial proof.

Wiki suggests that the chain rule should be valid in the non-square situation as well.

What exactly is the "chain rule" ?

The "chain rule" is a statement about the derivative of a concatenation of two functions for example

$\mathbf{f}(\mathbf{x}')$ $\qquad$ $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ $\qquad$ $\mathbf{f}(\mathbf{F}(\mathbf{x})) = (\mathbf{f} \circ \mathbf{F})(\mathbf{x})$

$$\partial f^{k}(\mathbf{x}')/\partial x^{i} = \Sigma_{j} \frac{\partial f^{k}(\mathbf{x}')}{\partial x'^{j}} \frac{\partial x'^{j}}{\partial x^{i}}$$

Now suppose you take this particular function

$\mathbf{f}(\mathbf{x}') = \mathbf{x}'$ $\qquad$ $f^{k}(\mathbf{x}') = x'^{k}$

Then the chain rule above says

$$\partial x'^{k}/\partial x^{i} = \Sigma_{j} \frac{\partial x'^{k}}{\partial x'^{j}} \frac{\partial x'^{j}}{\partial x^{i}} = \frac{\partial x'^{k}}{\partial x^{i}}$$ = a triviality, no δ is involved here

On the other hand consider (we are assuming that $\mathbf{x} = \mathbf{F}^{-1}(\mathbf{x}')$ exists _

$\mathbf{f}(\mathbf{x})$ $\qquad$ $\mathbf{x} = \mathbf{F}^{-1}(\mathbf{x}')$ $\qquad$ $\mathbf{f}(\mathbf{F}^{-1}(\mathbf{x}')) = (\mathbf{f} \circ \mathbf{F}^{-1})(\mathbf{x}')$

$$\partial f^{k}(\mathbf{x})/\partial x'^{i} = \Sigma_{j} \frac{\partial f^{k}(\mathbf{x})}{\partial x^{j}} \frac{\partial x^{j}}{\partial x'^{i}}$$

Now suppose you take this particular function

$\mathbf{f}(\mathbf{x}) = \mathbf{x}$ $\qquad$ $f^{k}(\mathbf{x}) = x^{k}$

Then the chain rule above says

$$\partial x^{k}/\partial x'^{i} = \Sigma_{j} \frac{\partial x^{k}}{\partial x^{j}} \frac{\partial x^{j}}{\partial x'^{i}}$$

or

$$\partial x^{k}/\partial x'^{i} = \frac{\partial x^{k}}{\partial x'^{i}}$$ = a triviality, no δ is involved here.

Consider a third case. In the above "other hand", suppose I give $\mathbf{f}(\mathbf{x})$ the name $\mathbf{x}'(\mathbf{x})$. Then

$$\partial x'^{k}(\mathbf{x})/\partial x'^{i} = \Sigma_{j} \frac{\partial x'^{k}(\mathbf{x})}{\partial x^{j}} \frac{\partial x^{j}}{\partial x'^{i}}$$

But just giving f(x) this artificial name does not mean $\partial x'^{k}(\mathbf{x})/\partial x'^{i} = \delta^{k}{}_{i}$.

So where am I getting the idea that you should have

$$\frac{\partial x'^i}{\partial x^j}\frac{\partial x^j}{\partial x'^k} = \delta^i{}_k \quad ??$$

OK, this is not an obvious fact. However, it is true in certain cases.

The way I write this in tensor doc is

$$\Sigma_k \frac{\partial x'_i}{\partial x_k}\frac{\partial x_k}{\partial x'_j} = \frac{\partial x'_i}{\partial x'_j} = \delta_{i,j}$$

I refer to this as "the chain rule" but I think that is a mistake.

Case 1: (tensor doc) If R is a square matrix of full rank, then its inverse S exists, and RS = SR = 1. In this case, you may conclude that

$$\frac{\partial x'^i}{\partial x^j}\frac{\partial x^j}{\partial x'^k} = \delta^i{}_k \qquad \text{or} \qquad R^i{}_j S^j{}_k = \delta^i{}_k$$

$$\frac{\partial x^i}{\partial x'^j}\frac{\partial x'^j}{\partial x^k} = \delta^i{}_k \qquad \text{or} \qquad S^i{}_j R^j{}_k = \delta^i{}_k$$

Saying that $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ is invertible means that $\mathbf{x}' = R\mathbf{x}$ is invertible at the microscale and so SR=RS=1.

Case 2: (**tall R matrix**) If R is full rank,

- it is possible to find a left inverse S such that SR = 1
- this S is generally not unique
- one S which works is $S = (R^T R)^{-1} R^T$. We know $(R^T R)^{-1}$ exists if R is full rank, see elsewhere.
- it is not possible to find a right inverse S such that RS = 1

The following is not obvious to me: Suppose you take a sample tall R matrix problem like that above and you compute S as I have done.

Question 1: Do you know that this S is a left inverse for R? (it was in my example)

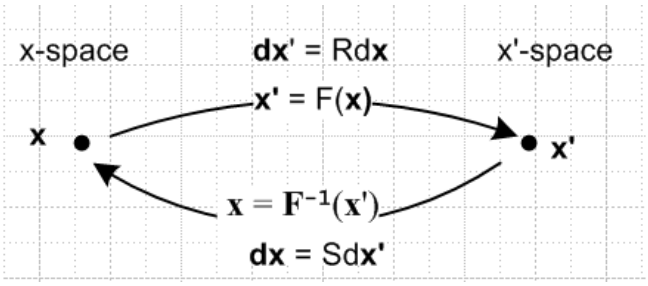
Suppose the answer is yes. You could then conclude that:

$$\Sigma_{j=1}^{m} \frac{\partial x^i}{\partial x'^j}\frac{\partial x'^j}{\partial x^k} = \delta^i{}_k \qquad \text{or} \qquad S^i{}_j R^j{}_k = \delta^i{}_k \quad \text{// always true}$$

$$\Sigma_{j=1}^{n} \frac{\partial x'^i}{\partial x^j}\frac{\partial x^j}{\partial x'^k} = \delta^i{}_k \qquad \text{or} \qquad R^i{}_j S^j{}_k = \delta^i{}_k \quad \text{// never true}$$

In this scenario, it is the equation whose sum has the larger number of terms that is true.

Answer to question 1. I think we do know this. Consider



If you start with **x** and go over to **x'** on the right, you should be able to map that vector back to **x** as shown. Points on the manifold M in x'-space always have a unique inverse due to assumptions made about the mapping which I have downplayed. The mapping is 1-to-1 from a domain U in x-space to a domain V on the manifold M in x'-space, and that is why it has a unique inverse. For each mapping, I show both the full map, and the local linear map. Combining the two local maps you get

$$d\mathbf{x} = Sd\mathbf{x}' = S(Rd\mathbf{x}) = SR\, d\mathbf{x} = d\mathbf{x}.$$

This equation SR d**x** = d**x** is valid for every d**x** in x-space. In particular, this is valid for **x** = dx(1,0) and **x** = dx(0,1) so it is valid for all basis elements in x-space. Therefore, SR d**x** = d**x** $\Rightarrow$ SR = 1 as a matrix.

If you try it in the reverse order, you would be claiming that RS d**x'** = d**x'** , but this is only true for d**x'** which lie on the manifold M. Since this equation RS d**x'** = d**x'** is not valid for a full set of basis vectors in **x'**-space, you cannot conclude that RS = 1. In our example, there are at most two basis vectors lying on the manifold, but you need the equation to be true for all three basis vectors.

Therefore, if you find a matrix S from your calculation of S, you are saying d**x** = Sd**x'** for that S, and by the above argument, you have SR = 1 so yes, the S you compute must be a left inverse of R, as it was in my simple example.

Summary regarding S and R and the "chain rule identity" :

Case 1: (tensor doc) If R is a square matrix of full rank, then its inverse S exists, and RS = SR = 1.
In this case, you may conclude that

$$\frac{\partial x'^{i}}{\partial x^{j}}\frac{\partial x^{j}}{\partial x'^{k}} = \delta^{i}{}_{k} \qquad \text{or} \qquad R^{i}{}_{j}S^{j}{}_{k} = \delta^{i}{}_{k}$$

$$\frac{\partial x^{i}}{\partial x'^{j}}\frac{\partial x'^{j}}{\partial x^{k}} = \delta^{i}{}_{k} \qquad \text{or} \qquad S^{i}{}_{j}R^{j}{}_{k} = \delta^{i}{}_{k}$$

Saying that **x'** = **F**(**x**) is invertible means that d**x'** = Rd**x** is invertible at the microscale and so SR=RS=1.

Case 2: (**tall R matrix**) If R is full rank,

- it is possible to find a left inverse S such that SR = 1

- this S is generally not unique
- there is no trivial formula that I know of for computing all possible S
- one S which works is $S = (R^{T}R)^{-1}R^{T}$. We know $(R^{T}R)^{-1}$ exists if R is full rank, see elsewhere.
- if you "compute S" from $d\mathbf{x} = Sd\mathbf{x}'$ ($\mathbf{x} = \mathbf{F}^{-1}(\mathbf{x}')$), the S you get is such a left inverse and SR = 1.
- it is not possible to find a right inverse S such that RS = 1
- in this Case 2 we conclude that

$$\Sigma_{j=1}^{m} \frac{\partial x^{i}}{\partial x'^{j}} \frac{\partial x'^{j}}{\partial x^{k}} = \delta^{i}{}_{k} \quad \text{or} \quad S^{i}{}_{j}R^{j}{}_{k} = \delta^{i}{}_{k} \qquad \text{m terms in sum}$$

$$\Sigma_{j=1}^{n} \frac{\partial x'^{i}}{\partial x^{j}} \frac{\partial x^{j}}{\partial x'^{k}} \neq \delta^{i}{}_{k} \quad \text{or} \quad R^{i}{}_{j}S^{j}{}_{k} \neq \delta^{i}{}_{k} \qquad \text{n terms in sum}$$

Concerning the last non-equation, think of the manifold in x'-space. Consider $S^{j}{}_{k} = \frac{\partial x^{j}}{\partial x'^{k}}$. For this object to exist genuinely, you have to say how x^{j} varies in x-space as you vary x'^{k} in all directions in x'-space because all values of k are included in what I would call a chain rule sum. BUT no matter how you select coordinates in x'-space, there is at least one direction k for which you cannot vary x'^{k} without leaving the manifold surface, and this is simply an undefined concept and corresponds to nothing in x-space.