Appendix A and B Notes (Sjamaar) PhL 7.12.15

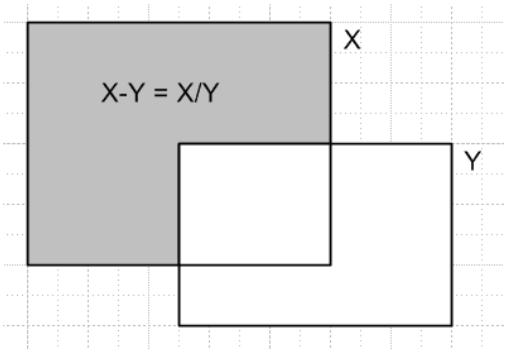
I read these appendices after finishing Chapter 4.

Appendix A: Set Theory

A.1 Glossary

This is a great concise summary of all the key definitions and facts. Some items are new to me.

X - Y = all elements in set X minus those in set Y (the complement of Y in X). Wiki says this notation is ambiguous since it might mean all differences of the form x-y, so they use X/Y = X-Y (X not in Y)



Cartesian product: X x Y Examples

S^1 x [0,1] = circle x (line segment) = cylinder shell S_1 = circle

What this means is your coordinates are (r,θ) x z

S^1 x S^1 = circle x circle = torus shell (r,θ) x (R,φ)

R x R = the plane

When Sja writes f: X→ Y, he means this: the mapping is defined for all x in X, so X is the **domain**, no confusion about that. The set of all f(x) in Y is called the **range** = **image**. Since Y is typically larger than the range = image, it has a different name. Y is the **codomain** or **target** space. Sja talks about image, but does not use the word range. If X and Y are reals, f is a **function**, otherwise it is a **mapping**, just a convention in math writing.

A is subset of X: f(A) = **image** of A under map f

B is subset of Y: $f^{-1}(B)$ = **preimage** of B under map f (f may not have an inverse!)

$f^{-1}(c)$ = the preimage in X of a point c in Y. It implies f(**x**) = c and the solution of this equation for the locus of **x** I usually call a level set or level curve. Sja calls it "**the level set of f at c**". Remember that the sets may be discrete so curves may not exist, so he is right. Also called a **the fiber of f at c**.

$f \circ g$ = composition, I am already happy with this.

injective = same as one-to-one $f : X \to Y$

Note that the image of f(X) might be less than all of Y.

If f(X) = all of Y so that f(X) = Y, then f is **surjective** which is the same as **onto**.

If f:X→Y is both injective and surjective, it is called **bijective**.

Theorem: f = bijective $\Leftrightarrow$ $f^{-1}f(x) = x$ and $f(f^{-1}(y) = y$ for all x in X and all y in Y

$f^{-1} \circ f = 1$ and $f \circ f^{-1} = 1$ (**two-sided inverse**)

Consider now f: X → R where X is a finite set (such as 1,2...N) and R = reals.

Consider: $\Sigma_{\texttt{all x in X}} f(x)$ = a real number.

X is called an **index set**. Image items I might write as f_n for n = x in X. So above is $\Sigma_n f_n$.

Example 1: X = {1,2.....N} sum of interest = $\Sigma_{i=1}^{N} f_i = \Sigma_{i=1}^{N} f(i)$

Example 2: X = { (i,j) where i in X, j in X} sum = $\Sigma_{i,j} f_{i,j}$

Example 3: X = { (i,j) where i in X, j in X and $i \leq j$} sum = $\Sigma_{i<j} f_{i,j}$

These sums are what we see when we talk about a general differential form!
The last two examples are a multi-index and an ordered multi-index.

Example 3 where N = 3. The dots in the "tableau" show a graph of the index set (i,j) with $i \leq j$.
The numbers by the dots indicate a sample function f(x) = f(i,j) = i + j. Exercise (A.2) asks reader to consider to compute the sum $S(N) \equiv \Sigma_{0 \leq i \leq j \leq N} (i+j)$ for arbitrary N using induction.

A.2 Topology of R^n

$B(\varepsilon,\mathbf{x})$ = closed ball radius ε around **x**

$B^{\circ}(\varepsilon,\mathbf{x})$ = open ball radius ε around **x** (every point has some space around it)

open set = not defined by Sja, but I think you can put an open ball around any point.

closed set = every sequence in the set converges to a point in the set

Example: the set [0,1) is neither closed nor open

Example: the set R is both closed and open (not many of these)

A is bounded set = $||\mathbf{x}|| < R$ for some fixed R for all x in A.

compact = closed and bounded (for R^n)

Theorem: integral of continuous function over compact set is always clean and finite.

Seems wrong. Consider integral of 1/x over [0,1] which diverges. I think he means

Theorem: integral of continuous *bounded* function over compact set is always clean and finite.

Maybe a continuous function has to be bounded? Yeah, that's right. The function 1/x is not continuous at x = 0, I am sure that Bucks show this. Yes, you cannot get $|f(x) - f(0)| < \varepsilon$ for any δ since $f(0) = \infty$ for 1/x. Restate: the function 1/x is not continuous at x= 0. More generally, an unbounded function is not continuous at any point where the function is infinite.

Appendix B: Calculus Review

B.1 The fundamental theorem of calculus

He states this three ways as shown. The function being integrated is called f(t). The first way is the statement of a definite integral from a to b, integral is called F where $f(x) = d_{\mathbf{x}}F = F'(x)$

$$\int_{\mathbf{a}}^{\mathbf{b}} f(t)dt = F(b)-F(a)$$

The second way says $d_{\mathbf{x}}(\int_{\mathbf{a}}^{\mathbf{x}} f(t)dt)) = f(x)$ and F does not appear. Derivation from first:

$$\int_{\mathbf{a}}^{\mathbf{x}} f(t)dt = F(x)-F(a) \qquad \Rightarrow \quad d_{\mathbf{x}}(\int_{\mathbf{a}}^{\mathbf{x}} f(t)dt)) = d_{\mathbf{x}}F(x) = f(x)$$

"the integral of a continuous function is differentiable" and "that derivative is the integrand f".
Easy to show that $\int_{\mathbf{a}}^{\mathbf{x}} f(t)dt)$ is continuous I think.

Third way is same as first way, but replace f(t) by F'(t)

$$\int_{\mathbf{a}}^{\mathbf{b}} F'(t)dt = F(b)-F(a)$$

or

$$\int_{\mathbf{a}}^{\mathbf{x}} F'(t)dt = F(x)-F(a)$$

or

$F(x) = F(a) + \int_a^x F'(t)dt$ F(t) arbitrary function, they call it g(t)

Expresses a function in terms of its derivative.

B.2 Derivatives of vector valued functions of vectors

Here (B.4) gives a definition of $\partial_j\varphi_i$ making use of the unit vector $\mathbf{e}_j$. A simple way to state this definition I think. We have $\varphi : R^n \to R^m$.

Warning: If you are discussing $\partial_j\varphi_i$ on an open subset U of R^n , this definition is fine because there is always room around a point to do the derivative. But if U is a closed set, the object $\partial_j\varphi_i$ might not be well-defined on boundary points of U. [this warning is stated bottom page 130]

The Jacobian matrix is this (generally not square)

$$J(x) = \begin{matrix} \partial_1\varphi_1 & \partial_2\varphi_1 & \dots & \partial_n\varphi_1 \\ \partial_1\varphi_2 & \partial_2\varphi_2 & \dots & \partial_n\varphi_2 \\ \dots & & & \\ \partial_1\varphi_m & \partial_2\varphi_m & \dots & \partial_n\varphi_m \end{matrix}$$

In tensor doc I have

$(\nabla \mathbf{A})_{ij} \equiv \partial_j A_i \equiv A_{i,j}$ // note index reversal (G.1.2)

so fair to say that

$J(x) = (\nabla\boldsymbol{\varphi})$ = a matrix // a notation Sjamaar does not use.

$J(x) = D\boldsymbol{\varphi}$ = a matrix // a notation Sjamaar does use.

Now what is this:

$J\mathbf{v} = (\nabla\boldsymbol{\varphi})\mathbf{v}$ = directional derivative along **v**.

Example

$J\mathbf{e}_n = (\nabla\boldsymbol{\varphi})\mathbf{e}_n$

$(J\mathbf{e}_n)_i = [(\nabla\boldsymbol{\varphi})\mathbf{e}_n]_i = (\nabla\boldsymbol{\varphi})_{ij}\,(\mathbf{e}_n)_j = \partial_j\varphi_i\,(\delta_{n,j}) = \partial_n\varphi_i = \partial\varphi_i/\partial x_n$

So this makes sense as the derivative along the n axis.

Example 1: $\boldsymbol{\varphi} : R^1 \to R^m$ (curve) $t \to \mathbf{x}$

$D\boldsymbol{\varphi} = [\ \partial_t\varphi_1, \partial_t\varphi_2 ....\partial_1\varphi_m]^T \quad = \partial_t\boldsymbol{\varphi}(\mathbf{x}(t))$

In this case, the image for some [a,b] is a curve in R^m [the mapping φ is the actual curve for Bucks]. So in this case $D\boldsymbol{\varphi} = \partial_t\varphi$ is the velocity. Bucks might write

$\boldsymbol{\gamma} : R^1 \rightarrow R^m$ (curve) $\qquad \mathbf{v} = \partial_t\boldsymbol{\gamma}$

Example 2: $\boldsymbol{\varphi} : R^n \rightarrow R^1 \qquad\qquad \mathbf{x} \rightarrow t$

$D\boldsymbol{\varphi} = \partial_1\varphi_1\ \partial_2\varphi_1 \qquad ... \qquad \partial_n\varphi_1 \quad = \quad (\nabla\varphi_1)^T$

So in this case the Jacobian matrix is a row vector equal to the gradient of your scalar function φ. The actual result is a row vector which we indicate with the T thing, since ∇f is normally a column vector.

Consider:

$J\mathbf{v} = (\nabla\boldsymbol{\varphi})\mathbf{v} = (\nabla\varphi_1)^T\mathbf{v} = \nabla\varphi_1 \bullet \mathbf{v}$ = a number

Sja comments that ∇φ points in the direction of max increase of φ, direction of steepest ascent. This is my usual argument due to $d\varphi = \nabla\varphi_1 \bullet d\mathbf{r}$.

C^1 = f is continuous and differentiable on U in R^n = "continuously differentiable"

One must mind the warning given above if U is a closed set! If open, less worry.

C^2 means that $\partial^2_{ij}f$ exists and is continuous in all of U.

C^r = obvious

C^∞= "smooth"

B.3 The Chain Rule

I will have to ponder this section, it views things differently than my usual view.

Suppose we have

$\mathbf{y} = \boldsymbol{\varphi}(\mathbf{x})$ $\qquad R^n \rightarrow R^m \rightarrow R^s$ (Sja uses R^k instead of R^s)

$\mathbf{z} = \boldsymbol{\psi}(\mathbf{y})$ $\qquad \mathbf{x}\ \ \varphi\ \ \mathbf{y}\ \ \psi\ \ \mathbf{z}$

U V W

Then one writes

$\mathbf{z} = \boldsymbol{\psi}(\boldsymbol{\varphi}(\mathbf{x})) = (\boldsymbol{\psi} \circ \boldsymbol{\varphi})(\mathbf{x})$ = a composition = $\boldsymbol{\theta}(\mathbf{x})$

Write a component of this latter

$z_i = \psi_i(\boldsymbol{\varphi}(\mathbf{x})) = (\boldsymbol{\psi} \circ \boldsymbol{\varphi})_i(\mathbf{x})$

I am happy to regard $z_i = z_i(\mathbf{x})$ as shown above. So

$[\partial z_i/\partial x_j] = \Sigma_{k=1}^{m} [\partial \psi_i(\mathbf{y})/\partial y_k] |^{\mathbf{y}=\boldsymbol{\varphi}(\mathbf{x})} * [\partial \varphi_k/\partial x_j]$ i = 1...s and j = 1.. n

This is completely clear to me. The question is how to write this in compressed notation.

$(D\boldsymbol{\varphi})_{kj} = (\nabla\boldsymbol{\varphi})_{kj} = \partial_j \varphi_k = [\partial \varphi_k(x)/\partial x_j]$ = last factor above

$(D\psi)_{ik} = (\nabla\boldsymbol{\psi})_{ik} = \partial_k \psi_i = [\partial \psi_i(x)/\partial y_k]$ = middle vector when taken at $\mathbf{y} = \boldsymbol{\varphi}(\mathbf{x})$

$(D\theta)_{ij} = (\nabla\boldsymbol{\theta})_{ij} = \partial_j \theta_i = [\partial \theta_i(x)/\partial x_j] = [\partial z_i(x)/\partial x_j]$ = left factor

So the chain rule then says

$(D\theta)_{ij}(\mathbf{x}) = \Sigma_{k=1}^{m} (D\psi)_{ik}(\mathbf{y}) |^{\mathbf{y}=\boldsymbol{\varphi}(\mathbf{x})} (D\boldsymbol{\varphi})_{kj}(\mathbf{x})$

or

$(D\theta)_{ij}(\mathbf{x}) = \Sigma_{k=1}^{m} (D\psi)_{ik}(\boldsymbol{\varphi}(\mathbf{x})) (D\boldsymbol{\varphi})_{kj}(\mathbf{x})$

or

$(D\theta)(\mathbf{x}) = (D\psi)(\boldsymbol{\varphi}(\mathbf{x})) (D\boldsymbol{\varphi})(\mathbf{x})$ matrix equation, generally non-square but conforming.

or

$(D(\boldsymbol{\psi} \circ \boldsymbol{\varphi}))(\mathbf{x}) = (D\psi)(\boldsymbol{\varphi}(\mathbf{x})) (D\boldsymbol{\varphi})(\mathbf{x})$ // agrees with p 131 A

So I guess I have never done this so carefully before. It is just matrix multiplication of the relevant R matrices, and nothing need be square, and we are not talking determinants at this point.

Once again, the component equation is this

$[\partial z_i/\partial x_j] = \Sigma_{k=1}^{m} [\partial \psi_i(\mathbf{y})/\partial y_k] |^{\mathbf{y}=\boldsymbol{\varphi}(\mathbf{x})} * [\partial \varphi_k/\partial x_j]$ i = 1...s and j = 1.. n

or

$[\partial(\boldsymbol{\psi} \circ \boldsymbol{\varphi})_i/\partial x_j] = \Sigma_{k=1}^{m} [\partial \psi_i(\mathbf{y})/\partial y_k] |^{\mathbf{y}=\boldsymbol{\varphi}(\mathbf{x})} * [\partial \varphi_k/\partial x_j]$ i = 1...s and j = 1.. n

Now you could remove the i subscript and write these two equations as

$[\partial \mathbf{z}/\partial x_j] = \Sigma_{k=1}^{m} [\partial \boldsymbol{\psi}(\mathbf{y})/\partial y_k] |^{\mathbf{y}=\boldsymbol{\varphi}(\mathbf{x})} * [\partial \varphi_k/\partial x_j]$ i = 1...s and j = 1.. n

or

$[\partial(\boldsymbol{\psi} \circ \boldsymbol{\varphi})/\partial x_j] = \Sigma_{k=1}^{m} [\partial \boldsymbol{\psi}(\mathbf{y})/\partial y_k] |^{\mathbf{y}=\boldsymbol{\varphi}(\mathbf{x})} * [\partial \varphi_k/\partial x_j]$ i = 1...s and j = 1.. n

If you leave off the $|^{\mathbf{y}=\boldsymbol{\varphi}(\mathbf{x})}$ you get

$[\partial(\boldsymbol{\psi} \circ \boldsymbol{\varphi})/\partial x_j] = \Sigma_{k=1}^{m} [\partial \boldsymbol{\psi}(\mathbf{y})/\partial y_k] * [\partial \varphi_k/\partial x_j]$ i = 1...s and j = 1.. n

which Sja refers to as "sloppy abbreviated notation". He does not use bold fonts despite his introductory section, so this adds a little to confusion. User must understand that $\mathbf{y} = \boldsymbol{\varphi}(\mathbf{x})$.

Some people might write the above as

$$[\partial \boldsymbol{\psi}/\partial x_j] = \Sigma_{k=1}^{m} [\partial \boldsymbol{\psi}(\mathbf{y})/\partial y_k] * [\partial \varphi_k/\partial x_j] \qquad i = 1...s \quad \text{and } j = 1..\ n$$

where now the reader has to understand that $[\partial \boldsymbol{\psi}/\partial x_j]$ means $[\partial \boldsymbol{\psi}(\varphi(\mathbf{x}))/\partial x_j]$. The reason this notation is bad is that you might think $[\partial \boldsymbol{\psi}/\partial x_j]$ means $[\partial \boldsymbol{\psi}(\mathbf{x})/\partial x_j]$.

Pullback notation: $(\boldsymbol{\psi} \circ \boldsymbol{\varphi}) = \boldsymbol{\varphi}^* \boldsymbol{\psi}$ so then we have

$$[\partial(\boldsymbol{\varphi}^* \boldsymbol{\psi})/\partial x_j] = \Sigma_{k=1}^{m} [\partial \boldsymbol{\psi}(\mathbf{y})/\partial y_k] \,|^{\mathbf{y}=\boldsymbol{\varphi}(\mathbf{x})} * [\partial \varphi_k/\partial x_j] \qquad i = 1...s \quad \text{and } j = 1..\ n$$

$$(D(\boldsymbol{\varphi}^* \boldsymbol{\psi}))(\mathbf{x}) = (D\psi)(\boldsymbol{\varphi}(\mathbf{x}))\,(D\boldsymbol{\varphi})(\mathbf{x}) \qquad \text{// agrees with p 131 A}$$

and the first line is the famous (B.6) but my form is more general.

B.4 Implicit Function Theorem

Consider $\boldsymbol{\varphi}(\mathbf{u},\mathbf{v}) = 0$ where $\boldsymbol{\varphi}$ lies in R^m, $\mathbf{u}$ in R^n and $\mathbf{v}$ also in R^m.

Think of $(\mathbf{u},\mathbf{v})$ as being in R^{n+m}.

The idea is that $\boldsymbol{\varphi}(\mathbf{u},\mathbf{v}) = 0$ might define a function $\mathbf{v} = \mathbf{f}(\mathbf{u})$ or $\mathbf{u} = \mathbf{g}(\mathbf{v})$ and so these functions are "implicit" in the equation $\boldsymbol{\varphi}(\mathbf{u},\mathbf{v}) = \mathbf{0}$.

We are just combining $(\mathbf{u},\mathbf{v})$ into a single coordinate $\mathbf{r} = (\mathbf{u},\mathbf{v})$ in R^{n+m}. We can talk about

$$(\mathbf{D}_u \boldsymbol{\varphi})_{ij} = \partial \varphi_i/\partial u_j \qquad = \text{part of the total R matrix}$$

$$(\mathbf{D}_v \boldsymbol{\varphi})_{ij} = \partial \varphi_i/\partial v_j \qquad \text{// this matrix is m x m, so square, so likely invertible}$$

Now suppose there is a "solution implicit function" which is $\mathbf{v} = \mathbf{f}(\mathbf{u})$. Then we could talk about

$$(\mathbf{Df})_{ij} = \partial f_i/\partial u_j$$

Are these three Jacobian matrices related? It seems to me that

$$\boldsymbol{\varphi}(\mathbf{u},\mathbf{v}) = 0$$

so

$$\varphi_i(\mathbf{u},\mathbf{v}) = 0 \qquad \text{for all } \mathbf{u} \text{ and } \mathbf{v} \text{ in some range}$$

Then

$$d\varphi_{\mathbf{i}}(\mathbf{u},\mathbf{v}) = 0$$

which then says

$$\Sigma_{\mathbf{j}=1}^{\mathbf{n}}(\partial\varphi_{\mathbf{i}}/\partial u_{\mathbf{j}})\, du_{\mathbf{j}} + \Sigma_{\mathbf{k}=1}^{\mathbf{m}}(\partial\varphi_{\mathbf{i}}/\partial v_{\mathbf{k}})\, dv_{\mathbf{k}} = 0$$

or

$$\Sigma_{\mathbf{j}=1}^{\mathbf{n}}(\mathbf{D_u\varphi})_{\mathbf{ij}}\, du_{\mathbf{j}} + \Sigma_{\mathbf{k}=1}^{\mathbf{m}}(\mathbf{D_v\varphi})_{\mathbf{ik}}\, dv_{\mathbf{k}} = 0$$

or

$$(\mathbf{D_u\varphi})d\mathbf{u} + (\mathbf{D_v\varphi})d\mathbf{v} = 0$$

Back up and write

$$\Sigma_{\mathbf{j}=1}^{\mathbf{n}}(\mathbf{D_u\varphi})_{\mathbf{ij}}\, du_{\mathbf{j}} + \Sigma_{\mathbf{k}=1}^{\mathbf{m}}(\mathbf{D_v\varphi})_{\mathbf{ik}}\, dv_{\mathbf{k}} = 0$$

Apply $\partial/\partial u_{\mathbf{a}}$ to both sides

$$\Sigma_{\mathbf{j}=1}^{\mathbf{n}}(\mathbf{D_u\varphi})_{\mathbf{ij}}\, \delta_{\mathbf{j,a}} + \Sigma_{\mathbf{k}=1}^{\mathbf{m}}(\mathbf{D_v\varphi})_{\mathbf{ik}}\, \partial v_{\mathbf{k}}/\partial u_{\mathbf{a}} = 0$$

or

$$(\mathbf{D_u\varphi})_{\mathbf{ia}} + \Sigma_{\mathbf{k}=1}^{\mathbf{m}}(\mathbf{D_v\varphi})_{\mathbf{ik}}\, \partial v_{\mathbf{k}}/\partial u_{\mathbf{a}} = 0$$

or

$$(\mathbf{D_u\varphi})_{\mathbf{ia}} = -\ \Sigma_{\mathbf{k}=1}^{\mathbf{m}}(\mathbf{D_v\varphi})_{\mathbf{ik}}\, \partial v_{\mathbf{k}}/\partial u_{\mathbf{a}}$$ // the minus sign appears

Now use $\mathbf{v} = \mathbf{f}(\mathbf{u})$ so that

$$\partial v_{\mathbf{k}}/\partial u_{\mathbf{a}} = \partial f_{\mathbf{k}}(\mathbf{u})/\partial u_{\mathbf{a}} = \partial_{\mathbf{a}} f_{\mathbf{k}}(\mathbf{u}) = (Df)_{\mathbf{ka}}(\mathbf{u})$$

We then have from 2 lines above

$$(\mathbf{D_u\varphi})_{\mathbf{ia}} = -\ \Sigma_{\mathbf{k}=1}^{\mathbf{m}}(\mathbf{D_v\varphi})_{\mathbf{ik}}\, (Df)_{\mathbf{ka}}$$

or

$$(\mathbf{D_u\varphi})(\mathbf{u},\mathbf{v})|^{\mathbf{v=f(u)}} = -\ (\mathbf{D_v\varphi})(\mathbf{u},\mathbf{v})^{\mathbf{v=f(u)}}\ (\mathbf{Df})(\mathbf{u})$$ //matrix form, and show arguments

Notice that

$(\mathbf{D_u\varphi})(\mathbf{u},\mathbf{f(u)})$ is wrong because $\partial_{\mathbf{u}}\varphi_{\mathbf{i}}(\mathbf{u},\mathbf{f}(\mathbf{u}))$ picks up unintended terms from 2nd arg.
A bit of a semantic thing here

$\partial_{\mathbf{u}}[\ \varphi_{\mathbf{i}}(\mathbf{u},\mathbf{f}(\mathbf{u}))\ ]$ would pick up those extra terms

$(\partial_{\mathbf{u}}\varphi_{\mathbf{i}})(\mathbf{u},\mathbf{f}(\mathbf{u}))$ does not pick up the extra terms
this means compute $(\partial_{\mathbf{u}}\varphi_{\mathbf{i}})$ and THEN evaluate it as shown

So *I will use the second notation* and restate

$$(\mathbf{D_u\varphi})(\mathbf{u},\mathbf{f(u)}) = -\ (\mathbf{D_v\varphi})(\mathbf{u},\mathbf{f(u)})\ (\mathbf{Df})(\mathbf{u})$$

Now I noted already that the $(\mathbf{D_v\varphi})$ matrix is mxm = square, so matrix $(\mathbf{D_v\varphi})^{-1}$ exists with suitable conditions, then we have

$$(\mathbf{D_v\varphi})^{-1}(\mathbf{u},\mathbf{f(u)})\,(\mathbf{D_u\varphi})(\mathbf{u},\mathbf{f(u)}) = -\ (\mathbf{Df})(\mathbf{u})$$

so final form is then

$$(\mathbf{Df})(\mathbf{u}) = -\,(\mathbf{D_v\varphi})^{-1}(\mathbf{u},\mathbf{f(u)})\ \ (\mathbf{D_u\varphi})(\mathbf{u},\mathbf{f(u)}) \qquad \text{// which is p 132 B}$$

Now what is all this business about **u** and **v** having to be close to $\mathbf{u_0}$ and $\mathbf{v_0}$? Well this same issue came up in my Buck notes (see Ch 5 meta). The question has to do with the invertibility of $(\mathbf{D_v\varphi})$. You have to be assured that $J = \det(\mathbf{D_v\varphi}) \neq 0$ = like det(S) in tensor doc. If you pick a point $(\mathbf{u_0},\mathbf{v_0})$ where this det ≠ 0, then you can reasonable imagine that there are at least small balls around $\mathbf{u_0}$ and $\mathbf{v_0}$ where det ≠ 0 and for **u** and **v** each in its small ball, we then have $\det(\mathbf{D_v\varphi}) \neq 0$. The invertibility of this thing is reasonable at a local level, but globally there is likely to be some combination of **u** and **v** where det = 0. In terms of the function **v** = **f**(**u**), that is like my **x**' = **F**(**x**) and we have d**x**' = Rd**x** and d**x** = Sd**x**' so the R matrix is invertible locally. OK, I like it.

Why is this result above called "the implicit function theorem" (**Dini's theorem** in Italy)? It is a statement about the R matrix of the implicit function **v** = **f**(**u**) which is inherent in φ(**u**,**v**) = 0 . I guess one point is that if the above expression is valid (meaning if you are near a point where $D_{\mathbf{v}}\varphi$ is invertible), then the implicit function **v** = **f**(**u**) exists AND is differentiable and I guess C^1.

Application: Suppose in **φ**(**u**,**v**) = 0 we have **φ**(**u**,**v**) = g(**v**) - **u** , a special case. Then our rule above says

$$(\mathbf{Df})(\mathbf{u}) = -\,(\mathbf{D_v\varphi})^{-1}(\mathbf{u},\mathbf{f(u)})\ (\mathbf{D_u\varphi})(\mathbf{u},\mathbf{f(u)})$$

and we have

$$(\mathbf{D_v\varphi}) = (\mathbf{D_v g}) \qquad \text{and} \qquad (\mathbf{D_u\varphi}) = -\,1$$

so it seems that we have

$$(\mathbf{Df})(\mathbf{u}) = -\,[(\mathbf{D_v g})]^{-1}\,(-1) \ = \ +\,(\mathbf{D_v g})^{-1}(\mathbf{u},\mathbf{f(u)})$$

But of course φ = 0 says that we have **g**(**v**) = **u** . Since we already were writing **v** = f(**u**) we can perhaps assume locally that things are invertible (as just commented on above), so then $f(u) = g^{-1}(u)$ and then

$$(\mathbf{Dg^{-1}})(\mathbf{u}) = \ (\mathbf{D_v g})^{-1}(\mathbf{u},\mathbf{g^{-1}(u)}) \qquad \text{// which is p 133 A}$$

This gives you the "differential" or R matrix of an inverse vector function in terms of the R matrix of the function. Item B then states this result for φ(u,v) = 0 in 1D. Again we have **u** near some $\mathbf{u_0}$ and **v** near $\mathbf{v_0}$.

This is called "the inverse function theorem" and again if $(D_v g)$ avoids det=0, the inverse function is *differentiable*. If g^{-1} is differentiable, then you can probably integrate to show g^{-1} is continuous.

A simple example where $u = g(v) = v^2$ is considered, and we have to make sure we don't have $v_0 = 0$ because at that point there is no inverse. We have $g^{-1}(u) = \sqrt{u}$ which has a branch point at u = 0, and the quantity $[\ g^{-1}]'(u) = 1/[1\sqrt{u}\ ]$ is undefined at $v_0 = u_0 = 0$.

B.5 The substitution formula for integrals

The Jacobian Integration Rule is stated for the case $\mathbf{y} = \mathbf{p}(\mathbf{x})$ where it is required that $\mathbf{p}$ is invertible (bijective) meaning one-to-one, AND both $\mathbf{p}$ and $\mathbf{p}^{-1}$ are C^1. He writes [note $\mathbf{x} = \mathbf{p}^{-1}(\mathbf{y})$]

$$\int_V f(\mathbf{y})d\mathbf{y} = \int_U f(\mathbf{p}(\mathbf{x}))d\mathbf{x}\ |\ \det(\mathbf{Dp})(x)| \qquad \text{// page 133 C}$$

Now thing of $\mathbf{g}(\mathbf{v}) = \mathbf{u}$ as $\mathbf{p}(\mathbf{x}) = \mathbf{y}$. Our inverse function theorem above then reads

$$(\mathbf{Dp}^{-1})(\mathbf{y}) = (\mathbf{D_x p})^{-1}(\mathbf{y},\mathbf{p}^{-1}(\mathbf{y}))$$

We have to make sure that $\det(\mathbf{D_x p}) \neq 0$ in order that $\mathbf{p}^{-1}$ be differentiable. Also, the sets V and U have to be open sets.

As usual, Sja always states the 1D version, and those abs values ARE there!

I often apply this rule to sets that are not open and there are points where the function is not 1-to-1. I guess one would have to show that the problem areas have no contribution to the integral and so on.

I am just aimlessly playing around here. The claim being made in p 132 B is this

$$(\mathbf{Df}) = -\ (\mathbf{D_v \varphi})^{-1}\ (\mathbf{D_u \varphi})|^{v=f(u)}$$

That would seem to say that

$$(\mathbf{D_v\varphi})(\mathbf{Df}) = \; - (\mathbf{D_u\varphi})|^{\mathbf{v=f(u)}}$$

or

$$(\mathbf{D_v\varphi})_{\mathbf{ij}}(\mathbf{Df})_{\mathbf{jk}} = \; - (\mathbf{D_u\varphi})|^{\mathbf{v=f(u)}}{}_{\mathbf{ik}}$$

or

$$\Sigma_{\mathbf{j}}\, \partial\varphi_{\mathbf{i}}/\partial v_{\mathbf{j}} \; * \partial f_{\mathbf{j}}/\partial u_{\mathbf{k}} \; = \; - \; \partial\varphi_{\mathbf{i}}/\partial u_{\mathbf{k}}|^{\mathbf{v=f(u)}}$$

Well remember that $\mathbf{v} = \mathbf{f}(\mathbf{u})$ so $v_{\mathbf{j}} = f_{\mathbf{j}}$ and then this says

$$\Sigma_{\mathbf{j}}\, \partial\varphi_{\mathbf{i}}/\partial v_{\mathbf{j}} \; * \partial v_{\mathbf{j}}/\partial u_{\mathbf{k}} \; = \; - \; \partial\varphi_{\mathbf{i}}/\partial u_{\mathbf{k}}|^{\mathbf{v=f(u)}}$$

and this begins to look like a chain rule, but what about that minus sign? I think Bucks dealt with this already, I will take a look.