Chapter 5: Differentiation [pp 221-295] 1
5.1 Transformations [221] 1
5.2 Linear Transformations [229] 2
5.3 Differential of a function [238] 3
5.4 Differentiation of Composite Functions [250] 8
5.5 Differentials of transformations [263] 16
5.6 Inverses of function of one variable [270] 18
5.7 Inverses of transformations [274] 19
5.8 Implicit Function Theorems [283] 21
5.9 Functional Dependence [287] 26

Chapter 5: Differentiation **[pp 221-295]**

5.1 Transformations **[221]**

This section is like the opening of my tensor doc. Bucks avoid saying $f : E^n \to E^m$ for some reason. Maybe that gets confused with $f : D \to R$. The words domain, range, transformation = mapping, image, inverse image all appear. Tensor doc starts with (5-3) as a set of real functions. Concatenation of mappings is called product or composite, they don't use the word concatenate. We then get the first three theorems of this chapter.

Continuity of f is defined as usual with | | being the E^m norm, though they fail to name that norm.

Theorem 1: [226] If $f : E^n \to E^m$ is continuous, then the inverse image of an open set S is an open set. If domain is restricted to D, then that inverse set is open relative to D.

Theorem 2: [226] If $f : E^n \to E^m$ is continuous, then D connected $\Rightarrow$ R is also connected, and same for any subset of D. That is to say S = connected $\Rightarrow$ f(S) = connected.

Theorem 3: [227]If $f : E^n \to E^m$ is continuous, then D compact $\Rightarrow$ f(D) = compact.

Three proofs are indicated, some involving Heine-Borel as you would expect, compact = closed + bounded, and "closed" is what will trigger Heine-Borel concerning open sets.

They use T to indicate a transformation, here I use f, and in tensor doc I write it $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ except there I use m = n only. Bucks do not bold T.

So this section is a very easy intro section for me on the subject of $f : E^n \to E^m$. Notice the importance in all theorems of f being *continuous*.

5.2 Linear Transformations **[229]**

In tensor doc, I have a small separate section on *linear* transformations, where F and R are the same thing. This seems to be the only type of transformation of interest to the Bucks, but they do allow m ≠ n for dimension on the two spaces. The issue of interest is the rank of my matrix R.

I am more familiar with the dimensionality of the nullspace, but as noted below rank = n - dim(N_A). If the nullspace has dim(N_A) = 1, then as you consider $\mathbf{y} = \mathbf{F}(\mathbf{x})$, a whole line (axis) in the domain maps into a point in the range, and in general the range will be a surface of dimension n-1. So the rank of the transformation matrix IS the dimension in the range of the entire mapped E

But the Bucks consider f: $E^n \rightarrow E^m$. How do things work in this case? I think you have to consider this in three separate cases: (rank cannot be larger than smallest dimension of matrix)

Case 1: n = m.

dim(image) = m - N	N = 0,1,2..m	N = dimension of nullspace = nullity
rank ≤ m	suspect: dim(image) = rank = m or less	

Case 2: n < m. If nullity N = 0, then seems to me range is surface of dimension n within E^m . So the max dimensionality of the image is that n of the domain:

dim(image) = n - N	N = 0,1,2..m	N = dimension of nullspace = nullity
rank ≤ n	suspect: dim(image) = rank = n or less	

Once you have N = n, everything maps into the origin so no point increasing N further.

Case 3: n > m. In this case you are already "projecting" from the get-go. You have a nullity of n-m right off the bat. In this case I suspect

dim(image) = m - N_{ex}	N_{extra} = 0,1,2..m	N = n-m+N_{extra}
rank ≤ m	suspect: dim(image) = rank = m or less	

Suppose n = 7 and m = 4. Then start off with N = 3 and dim(image) = 4-3 = 1. So N = 4 is as high as you can go here.

Preview Theorem 5: n = m = 3, it is case 1, no problem.

Preview Theorem 6: f: $E^n \rightarrow E^m$. Claims if rank = m, then all of E^m. But this rules out my Case 2 where you could not have rank = m. So they are really only talking Case 1 or 3. So they assume n ≥ m.

This is all old-hat for me. Linear transformation implies a matrix they call A = [a_{ij}], so Y = AX. I am reminded of how weak the "world" was on matrix notation and facts at the time of this book's first edition 1956, so authors avoid simple matrix notation. If det(A) ≠ 0 we are one-to-one and (locally) invertible etc.

Notion of **rank** is mentioned with examples, by no Sylvester, no nullity. Concatenation of matrices etc. We are still talking n x m here, so have to conform.

Theorem 4: [229] (**Form of Linear Xform**) Just states the *form* for a general *linear* transformation f : $E^n \rightarrow E^m$. Coefficients are real. Transformation is manifestly continuous since all component functions are. The have $\mathbf{y} = \mathbf{y}(\mathbf{x}) = A\mathbf{x}$.

Recall now from Stak Ch 2 notes: [is this for general n and m?]

$$R_A = (N_{A^*})^{\perp} \qquad \Leftrightarrow \qquad \text{rank} = r = \dim(R_A) = n - \dim(N_A)$$

So the **nullspaces** eats into the dimensionality of the range.

Theorem 5 [233] has n = 3 and shows how dim(range) = rank of the matrix A for r = 3,2,1,0. In each case, the range is a linear space of dimension r, a hyperplane through the origin.

Theorem 6 [233] (**Rank Theorem**) restates the above for general f: $E^n \rightarrow E^m$, general rank r. Here r = m is full rank and the range of linear f is all of E^m which one can think of as a plane of dimension m. For r < m, the range r is a plane of dimension r. These planes all contain the origin since 0→0 no matter what r is. [see notes above, so this theorem cannot have r = m unless n ≥ m. But OK.]

Bucks do not mention nullspace at this time.

Theorem 7 [236] shows that there exists a constant B such that if Y = AX, then |Y| < B|X| for all X.

In Stak for transformation T one writes ||T|| = max(| Y(X)|/|X|) over all X as the norm of the transformation. Just having a finite norm gives Theorem 7, and I guess ||T|| is the smallest value of B.

OK, enough. The tentative B is given by $\Sigma_{ij} |a_{ij}|^2$. For f : $E^n \rightarrow E^m$ you can often find a smaller upper bound that B, but for f : $E^n \rightarrow E^1$ this B is in fact the smallest bound.

5.3 Differential of a function **[238]**

Notation conversion:

Bucks	Tensor Doc
$\mathbf{D}_\beta(p_0)$	$\hat{\beta} \bullet \nabla f(\mathbf{p}_0) \equiv \nabla_\beta f$
$\mathbf{D}_\alpha\mathbf{D}_\beta(p_0)$	$\hat{\alpha} \bullet \nabla [\ \hat{\beta} \bullet \nabla f(\mathbf{p}_0)] \equiv \nabla_\alpha\nabla_\beta f$
Δp small but finite	$\Delta\mathbf{x}$ small but finite
dT	R (but might not be square in Buck)
df	R but with only one row
$\Delta f(\Delta p)$	$f(\mathbf{p} + \Delta\mathbf{p}) - f(\mathbf{p})$, $\Delta\mathbf{p}$ = finite
$df(\Delta p)$	$\nabla f \bullet \Delta\mathbf{p} = R\Delta\mathbf{p}$ when R is one row

This section uses a very strange notation. $f : E^n \rightarrow E^1$ so we are talking about a scalar function f only. For me, a directional derivative is $\hat{\beta} \bullet \nabla f(\mathbf{x}) = \nabla_\beta f(\mathbf{x})$ at some point x, and Bucks write this as $(\mathbf{D}_\beta f)(\mathbf{x})$ but their $\mathbf{x}$ is $\mathbf{p}_0$. Their β is a unit vector, no hat shown by them.

Page 240 gives lots of notations for partial derivatives = directional derivative along the axes.

Comment: In wiki, C^1 means "continuously differentiable" which in turn means differentiable everywhere in domain of interest *and* the derivative itself is continuous. The V function is continuous but is not differentiable at the origin. If you ignore the origin, the V derivative function is discontinuous so V violates the definition on two counts. The unit step function is discontinuous at the origin, but derivative seems continuous at the origin (made 0 everywhere), so violates the first count only. I address the C^n subject in Section 3.3 notes, page 8. In general, if a function is C^m then it is also C^k for all $k < m$. Bucks notation is $C^{(m)}$ with parens.

Then they define "the **differential** of a function" as $L \equiv [\partial_x f, \partial_y f, \partial_z f]$ in square brackets, which I call ∇f . For me, the "differential" is $df = d\mathbf{r} \bullet \nabla f$ and for Bucks this appears as $df = L(\Delta p)$ by which they really mean $df = L \bullet (\mathbf{\Delta} p) = \nabla f \bullet \mathbf{\Delta} p$. They talk about terms of different orders in $|\Delta p|$ as I would. We are on page 238, but Bucks don't define gradient until page 397.

Note Added: Their "differential" is more important than I at first thought. First, here are a few Buck examples:

p 243 w = f(x,y) "differential of f" = df = $[\partial_x w, \partial_y w]$

p 263 (u,v,w) = **T**(x,y,z) dT = $[\ \partial_x u \ \ \partial_y u \ \ \partial_z u\ ]$ for first row

"You get a row for each variable on the left, and a column for each variable on the right"

Now in tensor doc I would write $\mathbf{x}' = F(\mathbf{x})$ and $R_{ik} = (\partial x'_i / \partial x_k)$. Notice that

$R = [\ \partial_x u \ \ \partial_y u \ \ \partial_z u\ ]$ for first row

So for a function like (u,v,w) = **T**(x,y,z) we have: this is the connection: **dT = R.**

p 275 $J_{buck} = \det(dT)$ // = det(R) for (u,v,w) = **T**(x,y,z)

However, later we get T defined in the inverse manner, so now dT is different

p 304 (x,y,z) = **T**(u,v,w) dT = $[\ \partial_u x \ \ \partial_v x \ \ \partial_w x\ ]$ for first row

In this case, $T = F^{-1}$ of tensor doc, so here we would say dT = S.

Go back to tensor doc where I have this situation

$R_{ik}(\mathbf{x}) = (\partial x'_i / \partial x_k)$ is called the Jacobian matrix for the transformation $\mathbf{x}' = \mathbf{F}(\mathbf{x})$

$S_{ik}(\mathbf{x}') = (\partial x_i/\partial x'_k)$ is then the Jacobian matrix of the inverse transformation $\mathbf{x} = \mathbf{F}^{-1}(\mathbf{x}')$.

I showed in today's tensor doc edit that $\mathbf{x} = \mathbf{F}^{-1}(\mathbf{x}')$ is always the transformation that we end up looking at for curvilinear coordinates and it is thus the matrix S that is of interest and J = det(S). In tensor doc language, J is the Jacobian of the inverse transformation F^{-1} , but in Buck this is just called f so they don't have to think about inverse functions,

$\mathbf{x}' = \mathbf{F}(\mathbf{x})$ $\quad R_{ik}(\mathbf{x}) \equiv (\partial x'_i/\partial x_k)$ = "Jacobian matrix for $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ "

$$= \begin{matrix} \partial_1 x'_1 & \partial_2 x'_1 & \ldots\ldots \partial_m x'_1 \\ \partial_1 x'_2 & \partial_2 x'_2 & \ldots\ldots \partial_m x'_2 \\ \ldots\ldots & & \\ \partial_1 x'_n & \partial_2 x'_n & \ldots\ldots \partial_m x'_n \end{matrix}$$

$x' = f(x,y,z)$ $\quad R_{ik}(\mathbf{x}) \equiv (\partial x'_i/\partial x_k)$ = "Jacobian matrix for $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ "

$$= [\partial_x f \quad \partial_y f \quad \partial_z f\,] \quad = \text{"the differential for } x' = f(\mathbf{x})\text{"}$$

Buck differential dT = tensor doc R matrix of x,y,z are the independent variables

The important difference is that tensor doc always has m = n.

In tensor doc, I think of x'_i as the curvilinear coordinates. But equations are always presented as an inverse function, such as

$x = r\cos\theta$ $\quad (x,y) = \mathbf{F}^{-1}(r,\theta)$ $\quad (r,\theta) = \mathbf{F}(x,y)$
$y = r\sin\theta$

I think I might then say from tensor doc,

$R_{ik}(\mathbf{x}) \equiv (\partial x'_i/\partial x_k)$ = "Jacobian matrix for $\mathbf{x}' = \mathbf{F}(\mathbf{x})$ "

$S_{ik}(\mathbf{x}') \equiv (\partial x_i/\partial x'_k)$ = "Jacobian matrix for $\mathbf{x} = \mathbf{F}^{-1}(\mathbf{x}')$ "

In the example above we find that [since ordering r,θ here]

$$S = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix} \qquad J = |\det(S)| = r$$

so in tensor doc it is the Jacobian matrix of the inverse transformation that I chose for J.

In Buck that notation would be this for polars

$x = r\cos\theta$ $\quad (x,y) = \mathbf{f}(r,\theta)$ $\quad // = \mathbf{F}^{-1}(r,\theta)$

$y = r\sin\theta$

So for them, we have

$S_{ik}(\mathbf{x}') \equiv (\partial x_i/\partial x'_k)$ = "Jacobian matrix for $\mathbf{x} = \mathbf{F}^{-1}(\mathbf{x}') = \mathbf{f}(\mathbf{x}')$ " = $[df]_{ik}$

So the Buck's "differential df for transformation f " is the same as my "S matrix for transformation F^{-1}".

The Bucks sensibly write $\mathbf{x} = \mathbf{f}(\mathbf{x}')$ and then they never have to worry about an inverse function. But in my curvilinear tensor doc, I want to start with the fundamental $\mathbf{x}' = \mathbf{F}(\mathbf{x})$. I just edited tensor doc to get this definition of J cleared up.

Warning: Bucks use R as a Remainder thing, different from tensor doc R matrix.

OK, below they are just keeping an "extra piece" which I normally ignore.

Theorem 8: [243] **(Approximation Theorem).** In the following, d**r** is differential, Δ**r** is just small. Then

$\Delta f(\Delta\mathbf{r}) = f(\mathbf{r} + \Delta\mathbf{r}) - f(\mathbf{r}) = df(\Delta\mathbf{r}) + R(\Delta\mathbf{r}) = \nabla f \bullet \Delta\mathbf{r} + R(\Delta\mathbf{r})$

Notice that $df(\Delta\mathbf{r}) \equiv \Delta f \bullet \Delta\mathbf{r}$. You can say Δf is a function of Δ**r**, and this is what that function is. The function R(Δ**r)** is not specified, only the fact that

$\lim_{\Delta\mathbf{r}\to 0} [\, R(\Delta\mathbf{r}) / |\Delta\mathbf{r}\,| \,] = 0$

which I interpret as meaning " $R(\Delta\mathbf{r})$ is of order $(\Delta r)^{\alpha}$ where $\alpha > 1$ ". I am used to the power series situation where you would have $R(\Delta\mathbf{r}) \sim |\Delta\mathbf{r}|^2$, but Bucks are being more general. So go back to

$\Delta f(\Delta\mathbf{r}) = df(\Delta\mathbf{r}) + R(\Delta\mathbf{r})$

In the limit $\Delta r \to dr$, this becomes

$df(d\mathbf{r}) = \nabla f \bullet d\mathbf{r} = \partial_x f\ dx + \partial_y f\ dy$

So R describes the variation from the exact linear form df(d**r**) = linear in d**r** . This is all the way you would treat a 1D derivative first with Δx then dx.

Prior to proving Theorem 8, Bucks prove a **Lemma** on page 244 which is a simple 2D extension of the 1D Mean Value Theorem of Differential Calculus of Chapter 2

$y(b) = y(a) + y'(c)(b-a)$

or

$f(p+\Delta p) = f(p) + f'(p')\,\Delta p$ 1D MVT

where p' lies with the interval between p and p+Δp. The generalization is this for f(x,y):

$f(\mathbf{p}+\Delta\mathbf{p}) = f(\mathbf{p}) + f_1(\mathbf{p}')\,\Delta x + f_2(\mathbf{p}'')\,\Delta y$ **2D MVT** // p 245 A

which they derive from the 1D MVT used twice. The location of p' and p" is shown in Fig 5-4 p 245.

I went through the proof of Theorem 8 and it is fine, and I note the use of a previous theorem which says continuous on a compact D is uniformly continuous on D, p 69 Theorem 6.

They point out that the above is defining a **tangent plane** at the point of interest on the plot of f over (x,y), see p 244 fig.

[244] Is a **Lemma** which is proven using the 1D MVT twice. It is just another way to write df which is linear in Δx and Δy. They then use this Lemma to prove theorem 8 above on page 245 top.

Theorem 9 [246] then says $(\mathbf{D}_\beta f) = df(\beta)$ by which they mean $(\mathbf{D}_\beta f) = \nabla_\beta f = \nabla f \bullet \boldsymbol{\beta} = df(\beta)$ so it is df going in the β direction a unit distance. Of course f has to be C' in order to have derivatives. But in general this is the notation:

$df(\boldsymbol{\beta}) \equiv \nabla f \bullet \boldsymbol{\beta}$ for any vector $\boldsymbol{\beta}$, not necessarily of unit length

A particular application of this notation is for $\boldsymbol{\beta} = d\mathbf{r}$, a small vector

$df(d\mathbf{r}) \equiv \nabla f \bullet d\mathbf{r} = \partial_x f\ dx + \partial_y f\ dy = df$

Thm 9 seems uninteresting to me. Just says that $df = \nabla f \bullet \boldsymbol{\beta}$ if you move in the β direction.

Theorem 10 [246] (**Local Maximum Theorem**) says that at a local interior **maximum**, $\partial_i f = 0$ for all variables. A proof is given. Seems pretty reasonable to me!

Theorem 11 [247] says that if df = 0 for all points in domain D, then f = constant. D must be open and connected. For disconnected you could have different constant in each piece. For not-open, funny things could happen at a boundary point.

Theorem 12 [247] says that if $\partial_z f = 0$ in a convex open domain D, then f is not a function of z and is the same for all values of z. Nice picture on page 248 shows why convex is needed, case is f(x,y).

The subject then changes from "partial derivatives" to "partial integrals".

Theorem 13 [248] **Four Corners Theorem**: is about a double integral over a finite rectangle. It shows that the double integral of $\partial_{xy} f$ gives a simple combination of the values of f at the four corners. Sort of a generalization of the 1D perfect differential idea. But now f must be C" to get that $\partial_{xy} f$ to exist.

Corollary: From the above fact, it follows that $\partial_{xy} f = \partial_{yx}$ on D, which must be an open set. [Later in these Ch 5 notes I prove this fact again because I forgot it was right here!]

5.4 Differentiation of Composite Functions [250]

Theorem 14 [250] **(Chain Rule)** applies to Fig 5-8 only! This figure is like a neural network in clean layers with nothing jumping ahead 2 or more layers. Each path segment spans one layer only. Each variable is a function only of variables one layer to the right. The two chain rules are (5-11).

I went through details of this proof, it is just brute force, step by step, and is all OK by me.
Embedded in the above is **Example 1** p 250.

Example 2 [252] is NOT such a simple neural network. As I show on page 252, this network has two paths which violate the above layered network rule. As a result, the "chain rule" is messed up and has extra terms. For example, in (5-13) the first and last terms come from the "illegal" paths b and a shown in my drawing.

Comments

(1) in this situation, the notation $\partial_{\mathbf{x}}w$ has two separate meanings!! Ambiguous!

(2) *you can remove all ambiguity* by using the $f_{\mathbf{n}}$ type notation as shown page 253 A. This gives a lot of support for this notation. For example, (5-13) becomes p 253 A.

(3) an alternative notation is to put a var and to its right show the variables which are held constant.

$$f_{\mathbf{1}} = w_{\mathbf{1}} = (\partial_{\mathbf{x}}w)|_{\mathbf{u,v}}$$

The numbered notation is a lot simpler!

(4) I think this relates to my long-standing confusion in thermodynamics, things like $(\partial S/dV)_{\mathbf{T}}$ and so on.
[see Example 8 below!]

(5) I don't remember ever seeing this kind of description anywhere! But doing thermo I must have done a little research on the topic, but that was decades ago, notes will be lost. So thank you Bucks!

Example 3. [253] This is so messy on the 2nd derivative, I just have to write it all out

$$\partial_{\mathbf{t}}w = F_{\mathbf{3}} + F_{\mathbf{1}}\,\partial_{\mathbf{t}}x + F_{\mathbf{2}}\,\partial_{\mathbf{t}}y \qquad w = F(x,y,t)$$

$$\partial_{\mathbf{t}}^{\mathbf{2}}w = \partial_{\mathbf{t}}\,(F_{\mathbf{3}} + F_{\mathbf{1}}\,\partial_{\mathbf{t}}x + F_{\mathbf{2}}\,\partial_{\mathbf{t}}y) = \qquad \text{// For example, } F_{\mathbf{3}}(x,y,t)$$

$$= \partial_{\mathbf{t}}F_{\mathbf{3}} + \partial_{\mathbf{t}}[F_{\mathbf{1}}\,\partial_{\mathbf{t}}x] + \partial_{\mathbf{t}}[F_{\mathbf{2}}\,\partial_{\mathbf{t}}y]$$

$$= \partial_{\mathbf{t}}F_{\mathbf{3}} + (\partial_{\mathbf{t}}F_{\mathbf{1}})(\partial_{\mathbf{t}}x) + F_{\mathbf{1}}(\partial_{\mathbf{t}}^{\mathbf{2}}x) + (\partial_{\mathbf{t}}F_{\mathbf{2}})(\partial_{\mathbf{t}}y) + F_{\mathbf{2}}(\partial_{\mathbf{t}}^{\mathbf{2}}y)$$

But now [$F_{\mathbf{12}}$ means $(F_{\mathbf{1}})_{\mathbf{2}}$ so 1 is done first]

$$\partial_t F_1(x,y,t) = F_{11}(\partial_t x) + F_{12}(\partial_t y) + F_{13}$$
$$\partial_t F_2(x,y,t) = F_{21}(\partial_t x) + F_{22}(\partial_t y) + F_{23}$$
$$\partial_t F_3(x,y,t) = F_{31}(\partial_t x) + F_{32}(\partial_t y) + F_{33}$$

Then insert these to get

$$\partial_t^2 w = \underset{a}{(\partial_t F_3)} + \underset{b}{(\partial_t F_1)(\partial_t x)} + \underset{c}{F_1(\partial_t^2 x)} + \underset{b}{(\partial_t F_2)(\partial_t y)} + \underset{c}{F_2(\partial_t^2 y)} \quad = 11 \text{ terms}$$

$$= [\underset{a1}{F_{31}(\partial_t x)} + \underset{a2}{F_{32}(\partial_t y)} + \underset{a3}{F_{33}}] + [\underset{b1}{F_{11}(\partial_t x)} + \underset{b2}{F_{12}(\partial_t y)} + \underset{b3}{F_{13}}](\partial_t x)$$

$$+ F_1(\partial_t^2 x) + [\underset{c1}{F_{21}(\partial_t x)} + \underset{c2}{F_{22}(\partial_t y)} + \underset{c3}{F_{23}}](\partial_t y) + F_2(\partial_t^2 y)$$

$$= F_{31}(\partial_t x) + F_{32}(\partial_t y) + F_{33} \qquad \text{a1,a2,a3}$$

$$+ F_{11}(\partial_t x)^2 + F_{12}(\partial_t y)(\partial_t x) + F_{13}(\partial_t x) \qquad \text{b1,b2,b3}$$

$$+ F_{21}(\partial_t x)(\partial_t y) + F_{22}(\partial_t y)^2 + F_{23}(\partial_t y) \qquad \text{c1,c2,c3}$$

$$+ F_1(\partial_t^2 x) + F_2(\partial_t^2 y)$$

This agrees finally with bottom page 253. What a mess.

Comment: Bucks have omitted this theorem: [well, it exists above as Corollary to Thm 13 p 248, see above, , but OK we can just do it again]

Theorem PL1 : (change order of partials theorem)

$$\partial_{ij} f = \partial_{ji} f$$

You see it in the example on page 241. In their notation, this would require a very long proof. Basically you just use the definition of the derivative twice to show this

$\partial_{xy} f(x,y) = \partial_y (\partial_x f(x,y))$ // using Bucks notation of which one is done first

$= \partial_y [f(x + \Delta x,y) - f(x,y) + R(\Delta x,0)]$

$= [f(x + \Delta x,y+\Delta y) - f(x,y+\Delta y) + R(\Delta x,\Delta y)] - [f(x + \Delta x,y) - f(x,y) + R(\Delta x,0)]$

$\rightarrow [f(x + dx,y+dy) - f(x,y+dy)] - [f(x + dx,y) - f(x,y)]$

$= f(x + dx,y+dy) - [f(x,y+dy)+ f(x + dx,y)] + f(x,y)$

Now do it in the other order

$\partial_{yx} f(x,y) = \partial_x (\partial_y f(x,y))$ // using Bucks notation of which one is done first

$= \partial_x [f(x,y+\Delta y) - f(x,y) + R(0,\Delta y)]$

$= [f(x+\Delta x,y+\Delta y) - f(x+\Delta x,y) + R(\Delta x,\Delta y)] - [f(x,y+\Delta y) - f(x,y) + R(0,\Delta y)]$

$\rightarrow$ $[f(x+dx,y+dy) - f(x+dx,y)] - [f(x,y+dy) - f(x,y)]$

$= x+dx,y+dy) - [f(x,y+dy)+ f(x + dx,y)] + f(x,y)$

and this is the same as doing it in the other order.

Now go back to our previous result

$$\begin{aligned}\partial_t^2 w &= F_{31} (\partial_t x) + F_{32} (\partial_t y) + F_{33} \\ &+ F_{11} (\partial_t x)^2 + F_{12} (\partial_t y) (\partial_t x) + F_{13} (\partial_t x) \\ &+ F_{21} (\partial_t x)(\partial_t y) + F_{22} (\partial_t y)^2 + F_{23}(\partial_t y) \\ &+ F_1(\partial_t^2 x) + F_2(\partial_t^2 y)\end{aligned}$$

$$\begin{aligned}&= 2 F_{31} (\partial_t x) + 2F_{32} (\partial_t y) + F_{33} \\ &+ F_{11} (\partial_t x)^2 + 2F_{12} (\partial_t y) (\partial_t x) \\ &+ F_{22} (\partial_t y)^2 \\ &+ F_1(\partial_t^2 x) + F_2(\partial_t^2 y)\end{aligned}$$

$$= F_{33} + 2 [F_{31} (\partial_t x) + F_{32} (\partial_t y)] + 2F_{12} (\partial_t y) (\partial_t x)$$

$$+ F_{11} (\partial_t x)^2 + F_{22} (\partial_t y)^2 + F_1(\partial_t^2 x) + F_2(\partial_t^2 y) \quad = 8 \text{ terms} \quad = \text{page 254 A}$$

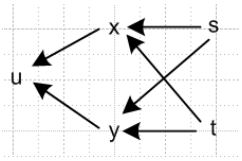
Example 4 [254] The next subject is how to determine partial derivatives if you are handed **a set of unsolved equations** such as Example 4 on page 254. Can write these generally as F(x,y,u,v) = 0 and G(x,y,u,v) = 0 (**Example 5**, p255) so we have four variables involved. A manual solution is provided on the bottom of page 254. But in the general case p 255 the solution is very straightforward and is an application of **Cramer's Rule** with solutions for $\partial_x u$ and $\partial_x v$ given as p 255A. For these two partials, y is treated as a constant so plays no role. But we could have held x constant and gotten similar equations for $\partial_y u$ and $\partial_y v$. The results can be expressed in **Jacobian notation** as shown bottom p 255, very good.

Pause to ponder what has been done in the above. You have two equations each in variables x,y,u,v, They have the general form F(x,y,u,v) = 0 and G(x,y,u,v) = 0. The variables x,y are assumed to be the independent ones, and u and v are dependent. If we instead had u = f(x,y,v) and v = v(x,y), we could compute something like $\partial u/\partial x$. We can do this as well in the F=0 and G=0 case, it just takes more work, and the answer is given for $\partial u/\partial x$ in the first line of A on page 255. I suspect this method is going to be important somewhere. A specific example of the general F=0 G=0 case is shown on page 254, and this

same result is recovered on page 256 using the general formula. The Cramer's rule is always going to produce some kind of ratio for the result.

Example 6 [256] . We are given the **wave equation** in x and y (instead of x and t). It is a PDE. Want to do a change of variables to s and t as shown and see what the PDE looks like. They turn the crank and the result is simply $\partial_{st}u = 0$ and footnote shows the general solution which makes perfect sense. So in this example we are converting a PDE from x,y to s,t. We are required to compute 1st and 2nd derivatives.

Pause to ponder: Forget the wave equation for a moment. Imagine we have some function u(x,y), then we can compute the object $\partial u/\partial x$ (for example) using the simple chain rule as shown in page 256 C.



Now go on to compute $\partial^2 u/\partial x^2$ and $\partial^2 u/\partial y^2$ and stick this into the 2D wave equation (y = time t), and what comes out is $\partial_{st}u = 0$ which says solution is A(s) + B(t) which is a(x+y) + b(x-y) which is the famous solution. I don't think this example fits into the mold of F = 0 and G = 0, it is a new idea where we have a differential equation involved. This certainly is an interesting way to solve this PDE!

Example 7 [257] . Convert 2D **Laplace equation** in x,y to polars r,θ. Again they turn the crank, lots of messy algebra, but out comes ∇^2_{2D} in polar coords in p 258A.

Pause to ponder: Instead of x,y functions of s and t, now x,y are functions of r and θ, polar coords. It is the same thing as Example 6 but more complicated functions. As before, we compute $\partial u/\partial x$ and then $\partial^2 u/\partial x^2$ and put into the Laplace equation, and the result (no surprise) is the Laplace equation in polar coordinates p 258 B.

Example 8. [258]. **The Thermodynamics Mystery**. We assume 4 variables (energy E, temp T, vol V, press p) and we have two relations as shown in (5-24), so now we really are doing an F=0 and G=0 type example. And we assume that when V,T are taken as the two independent variables, we have some operative equation (5-25). The question is: what does this same equation look like if you instead take p,T to be the two independent variables. Again, we are just converting a PDE from variables (V,T) to variables (p,T). The notion of "independent variables" arises, earlier these were always set to be (x,y). If a set is independent, then when you do ∂_a , all the others are held constant. Indep variables appear on the right end of a graph.

Pause to ponder: This is a great example from the thermo world. We start with variables V,T and we want to end up with variables p,T. We get to assume that E = f(p,T) and V = g(p,T). These are then our two equations which we could write E-f(p,T) = 0 and V-g(p,T) = 0 so like F = 0 and G = 0. The Cramer's Rule situation is then p 259 A and they solve these equations for $\partial E/\partial V$ and $\partial p/\partial T$ in terms of derivatives of the functions f and g. Note that $\partial E/\partial V$ and $\partial p/\partial T$ and the derivatives appearing in (5-25) which is our assumed starting thermo equation. The resulting thermo equation is completely different from the starting

one!! You don't just "replace" one of the derivatives. I would have liked to have this example many years ago when I studied thermo.

Pause. These really are excellent examples as my old note (date unknown) on p 257 says.

Theorem 15 [259] . (**Mean Value Theorem in N dimensions**)

The claim is this: $f(\mathbf{p_2}) - f(\mathbf{p_1}) = \nabla f(\mathbf{p}^*)\bullet(\mathbf{p_2} - \mathbf{p_1})$ $f: E^n \to E^1$

where $\mathbf{p}^*$ lies somewhere on the line segment joining $\mathbf{p_1}$ and $\mathbf{p_2}$.

First, compare this to the 1D "MVT of Diff Calc" from Section 2.7

$$y(b)- y(a) = y'(c)(b-a)$$

The generalization certainly seems plausible.

I will now do the proof in line:

$\mathbf{p_1} = (x,y)$ $\mathbf{p_2} = (x+\Delta x,y+\Delta y) = \mathbf{p_1} + (\Delta x,\Delta y) = \mathbf{p_1} + \mathbf{\Delta}$

$\mathbf{p_2} - \mathbf{p_1} = \mathbf{\Delta}$ $\mathbf{p}^* = \mathbf{p_1} + \lambda\mathbf{\Delta}$ $0 \le \lambda \le 1$ // candidate $\mathbf{p}^*$ I presume

Now define the following function of t

$F(t) \equiv f(\mathbf{p_1} + t\mathbf{\Delta})$ where f is the function f appearing in the claim above

Note in passing that,

$F(0) = f(\mathbf{p_1})$ $F(1) = f(\mathbf{p_2})$

Our 1D MVT may be applied to this function of one variable says

$y(b)- y(a) = y'(c)(b-a)$

or

$F(1) - F(0) = F'(t=\lambda)(1-0) = F'(\lambda)$ where $0 \le \lambda \le 1$

Now do the chain rule $\partial_t g(x + t\Delta x) = \partial_x g * \partial_t(x + t\Delta x) = \partial_x g\, \Delta x$

$\partial_t F(t) = \partial_t f(x+t\Delta x,y+t\Delta y) = f_1(x+t\Delta x,y+t\Delta y)\, \Delta x + f_2(x+t\Delta x,y+t\Delta y)\, \Delta y = \nabla f(\mathbf{p_1} + t\,\mathbf{\Delta}) \bullet \mathbf{\Delta}$

$\partial_t F(t)|^{\lambda} = \nabla f(\mathbf{p_1} + \lambda\,\mathbf{\Delta}) \bullet \mathbf{\Delta} = \nabla f(\mathbf{p}^*) \bullet \mathbf{\Delta} = F'(\lambda)$

Then we have shown that

$F(1) - F(0) = F'(\lambda)$

or

$f(\mathbf{p}_2) - f(\mathbf{p}_1) = \nabla f(\mathbf{p}^*) \bullet (\mathbf{p}_2 - \mathbf{p}_1)$

Although I wrote (x,y) in a few places, you can see that this proof is valid for points in E^n .

So our candidate form for $\mathbf{p}^*$ panned out, and λ is whatever the 1D MVT came up with!

Taylor's Theorem in 2D [Here I do it all in line]

[I think this all works for N dimensions in the obvious manner. This is a very good section!]

Now what happens if we apply the 1D Taylor theorem to F(t) in the above example?? Buck's version of this Taylor theorem was presented in p 127 Corollary 2

Corollary 2. [127] This just write things out as // This is Taylor's Theorem

$f(x) = P(x; n, x_0) + R_n(x; x_0)$

$= \Sigma_{m=0}^{n} (1/m!) f^{(m)}(x_0) (x-x_0)^m + f^{(n+1)}(\tau) (x-x_0)^{n+1} / (n+1)!$ // $x_o < \tau < x$

truncated Taylor remainder part

So we apply this to F(t) to get

$F(t) = \Sigma_{m=0}^{n} (1/m!) F^{(m)}(t_0) (t-t_0)^m + F^{(n+1)}(\tau) (t-t_0)^{n+1} / (n+1)!$

Not sure what to do next. We know that $F(0) = f(\mathbf{p}_1)$, $F(1) = f(\mathbf{p}_2)$. What is t_0 here? Let's try $t_0 = 0$ so

$F(t) = \Sigma_{m=0}^{n} (1/m!) F^{(m)}(0) t^m + F^{(n+1)}(\tau) t^{n+1} / (n+1)!$ $0 < \tau < t$

Notice that

$F(1) = f(\mathbf{p}_2) = f(\mathbf{p}_1 + \mathbf{\Delta})$

so evaluating the above series at t = 1 gives

$f(\mathbf{p}_1 + \mathbf{\Delta}) = \Sigma_{m=0}^{n} (1/m!) F^{(m)}(0) + F^{(n+1)}(\tau) / (n+1)!$ $0 < \tau < 1$

This is going to be our 2D Taylor expansion! We already know that

$F^{(1)}(0) = f_1(\mathbf{p}_0) (x-x_0) + f_2(\mathbf{p}_0) (y-y_0)$ // $F^{(1)}(t) = \partial F(t)/\partial t$

This was just the chain rule which you can write like this:

$F^{(1)}(0) = \partial_t F = (\Delta x \, \partial_{\mathbf{x}} + \Delta y \partial_{\mathbf{y}}) f = \mathbf{\Delta} \bullet \nabla f$ // = U f Buck p 260A

How about the next one :

$$F^{(2)}(t) = \partial_t [f_1(\mathbf{p})\Delta x + f_2(\mathbf{p})\,\Delta y] = [\partial_t f_1(\mathbf{p}_0 + t\boldsymbol{\Delta})]\,\Delta x + [\partial_t f_2(\mathbf{p}_0 + t\boldsymbol{\Delta})]\,\Delta y \qquad \mathbf{p} \equiv \mathbf{p}_0 + t\boldsymbol{\Delta}$$

$$= [\,\partial_x f_1(\mathbf{p}_1 + t\,\boldsymbol{\Delta})\Delta x + \partial_y f_1(\mathbf{p}_1 + t\,\boldsymbol{\Delta})\Delta y]\,\Delta x$$

$$+ [\,\partial_x f_2(\mathbf{p}_1 + t\,\boldsymbol{\Delta})\Delta x + \partial_y f_2(\mathbf{p}_1 + t\,\boldsymbol{\Delta})\Delta y]\,\Delta y$$

Detail: $[\partial_t f_1(\mathbf{p}_0 + t\boldsymbol{\Delta})\,] = \partial f_1/\partial x * \partial x/\partial t + \partial f_1/\partial y * \partial y/\partial t = \partial_x f_1 * \Delta x + \partial_y f_1 * \Delta y$

Again we can rewrite this as

$$F^{(2)}(t) = \partial_t F^{(1)} = (\,x\,\partial_x + \Delta y\partial_y)\,F^{(1)} = (\,x\,\partial_x + \Delta y\partial_y)(\,x\,\partial_x + \Delta y\partial_y)\,f = (\,x\,\partial_x + \Delta y\partial_y)^2\,f$$

We can now see the general formula is going to be

$$F^{(m)}(t) = (\,\Delta x\,\partial_x + \Delta y\partial_y)^m\,f(\mathbf{p}) \qquad \mathbf{p} \equiv \mathbf{p}_0 + t\boldsymbol{\Delta}$$

$$F^{(m)}(0) = (\,\Delta x\,\partial_x + \Delta y\partial_y)^m\,f(\mathbf{p}_0) \qquad \mathbf{p} \equiv \mathbf{p}_0$$

Our double Taylor series is then

$$f(\mathbf{p}_1 + \boldsymbol{\Delta}) = \Sigma_{m=0}^{n}\,(1/m!)\,F^{(m)}(0) + F^{(n+1)}(\tau)\,/\,(n+1)! \qquad 0 < \tau < 1$$

$$= \Sigma_{m=0}^{n}\,(1/m!)\,(\Delta x\,\partial_x + \Delta y\partial_y)^m\,f(\mathbf{p}_0) + (\,\Delta x\,\partial_x + \Delta y\partial_y)^{n+1}\,f(\mathbf{p}_\tau)\,/\,(n+1)!$$

where $\mathbf{p}_\tau = \mathbf{p}_0 + \tau\boldsymbol{\Delta}$ **.** Maybe rewrite this as

$$f(\mathbf{p}_1 + \boldsymbol{\Delta}) = \Sigma_{m=0}^{n}\,(1/m!)\,(\Delta\mathbf{r} \bullet \nabla)^m\,f(\mathbf{p}_0) + (\Delta\mathbf{r} \bullet \nabla)^{n+1}\,f(\mathbf{p}_\tau)\,/\,(n+1)!$$

In the above form, I think the result is valid in N Cartesian dimensions! Later they refer to $\Delta\mathbf{r} \bullet \nabla$ as U.

Now consider [this step is more complicated for N = 3 dimensions or more, need trinomial expansion!]

$$(\Delta x\,\partial_x + \Delta y\partial_y)^m = \Sigma_{i=0}^{m}\,(m,i)\,(\Delta x\,\partial_x)^{m-i}(\Delta y\,\partial_y)^i$$

so that

$$[(\Delta x\,\partial_x + \Delta y\partial_y)^m\,f\,]\,|^{p0}$$

$$= \Sigma_{i=0}^{m}\,(\,m!\,/\,[i!\,(m-i)!\,])\,(x-x_0)^{m-i}\,(y-y_0)^i\,[\,\partial_x{}^{m-i}\,\partial_y{}^i\,f(\mathbf{p}_0)\,]$$

and then our expansion becomes

$$f(\mathbf{p}_1 + \mathbf{\Delta}) = \Sigma_{m=0}^{n} (1/m!)\ (\Delta x\, \partial_x + \Delta y \partial_y)^m f(\mathbf{p}_0) \ + \ (\Delta x\, \partial_x + \Delta y \partial_y)^{n+1} f(\mathbf{p}_\tau)\ / (n+1)!$$

$$= \Sigma_{m=0}^{n} (1/m!) \{ \Sigma_{i=0}^{m} (m! / [i!\ (m-i)!\ ])\ (x-x_0)^{m-i} (y-y_0)^i\ [\ \partial_x^{m-i}\, \partial_y^i\, f(\mathbf{p}_0)\ ]\}$$

$$+ \Sigma_{i=0}^{n+1} ((n+1)! / [i!\ (n+1-i)!\ ])\ (x-x_0)^{n+1-i} (y-y_0)^i\ [\ \partial_x^{n+1-i}\, \partial_y^i\, f(\mathbf{p}_\tau)\ ]\ / (n+1)!$$

In each term a pair of factorials cancels to give

$$f(\mathbf{p}_1 + \mathbf{\Delta}) = \Sigma_{m=0}^{n} \{ \quad \Sigma_{i=0}^{m} \frac{1}{i!\ (m-i)!}\ (x-x_0)^{m-i} (y-y_0)^i\ [\ \partial_x^{m-i}\, \partial_y^i\, f(\mathbf{p}_0)\ ]\ \ \}$$

$$+ \Sigma_{i=0}^{n+1} \frac{1}{i!\ (n+1-i)!}\ (x-x_0)^{n+1-i} (y-y_0)^i\ [\ \partial_x^{n+1-i}\, \partial_y^i\, f(\mathbf{p}_\tau)\ ]$$

This is my first shot at the final result, I will repair it post feeding. The leftover term has the name

$$R_n = \Sigma_{i=0}^{n+1} \frac{1}{i!\ (n+1-i)!}\ (x-x_0)^{n+1-i} (y-y_0)^i\ [\ \partial_x^{n+1-i}\, \partial_y^i\, f(\mathbf{p}_\tau)\ ]$$

We can write out some of these terms,

$$\begin{aligned}
R_n \ &= \frac{1}{0!\ (n+1)!}\ (x-x_0)^{n+1}\ [\ \partial_x^{n+1}\ f(\mathbf{p}_\tau)\ ] \\
&+ \ \frac{1}{1!\ (n)!}\ (x-x_0)^{n} (y-y_0)^1\ [\ \partial_x^{n}\, \partial_y\, f(\mathbf{p}_\tau)\ ] \\
&+ \ \frac{1}{2!\ (n-1)!}\ (x-x_0)^{n-1} (y-y_0)^2\ [\ \partial_x^{n-1}\, \partial_y^2\, f(\mathbf{p}_\tau)\ ] \\
&+ \ \frac{1}{3!\ (n-2)!}\ (x-x_0)^{n-2} (y-y_0)^3\ [\ \partial_x^{n-2}\, \partial_y^3\, f(\mathbf{p}_\tau)\ ] \\
&.... \\
&+ \ \frac{1}{0!\ (n+1)!}\ (y-y_0)^{n+1}\ [\ \partial_y^{n+1}\ f(\mathbf{p}_\tau)\ ]
\end{aligned}$$

Bucks in p 260B state only the first and last term, and I agree on these, so I think things are right.

Let's now expand the m^{th} term of the series. It is this

$$\Sigma_{i=0}^{m} \frac{1}{i!\ (m-i)!}\ (x-x_0)^{m-i} (y-y_0)^i\ [\ \partial_x^{m-i}\, \partial_y^i\, f(\mathbf{p}_0)\ ]$$

$$\begin{aligned}
&= \frac{1}{m!}\ \ (x-x_0)^{m}\ \partial_x^{m}\ f(\mathbf{p}_0) \\
&+ \ \frac{1}{1!\ (m-1)!}\ (x-x_0)^{m-1} (y-y_0)^1\ [\ \partial_x^{m-1}\, \partial_y^1\, f(\mathbf{p}_0)\ ] \\
&+ .... \\
\end{aligned}$$

$$+ \frac{1}{m!} \; (y-y_0)^m \; \partial_y{}^m \; f(\mathbf{p}_0)$$

and this agrees with Bucks.

Theorem 16 [260] (**2D Taylor's Theorem**) . I just did all this above.

I have therefore derived Theorem 16 in full detail. Going back now

$$f(\mathbf{p}_1 + \mathbf{\Delta}) = \Sigma_{m=0}^{n} (1/m!) \; (\Delta x \, \partial_{\mathbf{x}} + \Delta y \partial_{\mathbf{y}})^m \, f(\mathbf{p}_0) \; + \; (\, \Delta x \, \partial_{\mathbf{x}} + \Delta y \partial_{\mathbf{y}})^{n+1} \, f(\mathbf{p}_\tau) \; / \, (n+1)!$$

we can define $U \equiv (\Delta x \, \partial_{\mathbf{x}} + \Delta y \partial_{\mathbf{y}})$ to write this as

$$\begin{aligned} f(\mathbf{p}_1 + \mathbf{\Delta}) &= \overbrace{\Sigma_{m=0}^{n} (1/m!) \, U^m \, f(\mathbf{p}_0)}^{\text{finite sum}} \; + \; \overbrace{U^{n+1} \, f(\mathbf{p}_\tau) \; / \, (n+1)!}^{\text{remainder}} \\ &= f(\mathbf{p}_0) \; + U \, f(\mathbf{p}_0) + (1/2) \, U^2 \, f(\mathbf{p}_0) \; + ... + \; (1/n!) \, U^n \, f(\mathbf{p}_0) \; + \; U^{n+1} \, f(\mathbf{p}_\tau) \; / \, (n+1)! \end{aligned}$$

and this agrees with **p 261 A**, except they write $\mathbf{p}_\tau = \mathbf{p}^*$, same thing. Notice that the remainder last term has the same form as all the other terms except $\mathbf{p}_0$ is replaced by some MVT $\mathbf{p}^*$.

I thus dramatically arrive at the end of Section 5.4, Bucks had a lot to say.

In older notes, I have this same double Taylor expansion: [math/ integrals and series...]

$$f(x,y) = \Sigma_{n=0}^{\infty} \; \Sigma_{m=0}^{\infty} \; T_{nm}(x,y) \; = \; \Sigma_{n=0}^{\infty} \; \Sigma_{m=0}^{\infty} \; k_{nm} \; (x-a)^n \, (y-b)^m$$

$$= \Sigma_{n=0}^{\infty} \; \Sigma_{m=0}^{\infty} \; \frac{\{\partial^n x \, \partial^m y \, f\}^{ab}}{n! \; m!} \qquad (x-a)^n \, (y-b)^m$$

This is a slightly different organization of the double sum. Instead of zig-zag starting at the upper left corner, I do columns first then rows, or vice versa. Also, my derivation gives the infinite series with no statement about the remainder term if you stop somewhere, so Buck's result is more general than this old work of mine. I could surely extend the Buck work to f(x,y,z) and have a remainder for that too. Perhaps all you have to do is say $U \rightarrow (\Delta x \partial_{\mathbf{x}} + \Delta y \partial_{\mathbf{y}} + \Delta z \partial_{\mathbf{z}})$ but then $(\Delta x \partial_{\mathbf{x}} + \Delta y \partial_{\mathbf{y}} + \Delta z \partial_{\mathbf{z}})^m$ is a more complicated operator and you need some kind of triple binomial expansion deal. [yes!]

5.5 Differentials of transformations **[263]**

Bucks are now doing their version of my tensor doc. Here are the connections

Bucks	Me in tensor doc
$(dT)_{ij}$	R_{ij}
dT = matrix = "differential"	R = matrix = "Jacobi matrix"
x_i	x_i

y_i	x'_i
$\mathbf{T}(\mathbf{p})$ // but Bucks have no bolding	$\mathbf{F}(\mathbf{x})$

$$F_i(\mathbf{x} + d\mathbf{x}) \approx F_i(\mathbf{x}) + \Sigma_k(\partial F_i(\mathbf{x})/\partial x_k)\, dx_k$$
$$= F_i(\mathbf{x}) + \Sigma_k R_{ik}\, dx_k$$

Buck Theorem 17 [264]:

$\mathbf{T}(\mathbf{p}_0+\Delta\mathbf{p}) = \mathbf{T}(p_0) + dT\,(\Delta\mathbf{p}) + \mathbf{R}(\Delta p)$

$\mathbf{T}(\mathbf{p}_0+d\mathbf{p}) = \mathbf{T}(p_0) + dT\,(d\mathbf{p})$ $\quad\quad$ $\mathbf{F}(\mathbf{x} + d\mathbf{x}) = \mathbf{F}(\mathbf{x}) + R\, d\mathbf{x}$ // linearization idea

So Buck Theorem 17 just says the above line which is just the definition of the linearization. All fine.
[Makes me want to add EQ nums to tensor doc] [it is done!! took 41 days...]

Theorem 18 [265] just says you can concatenate two R matrices in the obvious way. I talk about this in tensor doc in Section 8.10 where I write $R = R_1R_2$ and $J = J_1J_2$.

Example 1: [266] They give an example where the individual linear transformations are not square. I get it and won't study this in detail. All the functions are given names as shown. They show that you get the same result whether you use the chain rule or the matrix mult rule. I drew the little simple neural net. The two transformations (non-linear each) are called S and T, I might call them $\mathbf{F}_1$ and $\mathbf{F}_2$. The R matrices conform in size in the non-square case.

Example 2: [267] Here the graph is not a simple neural net, but this is dealt with by using SRT as a triple combination as shown.

Theorem 19: [268] (**Mean Value Theorem for Transformation F: $\mathbf{R}^3 \rightarrow \mathbf{R}^3$**). In tensor doc language, I think this is what this theorem says [wrong!]

$\mathbf{F}(\mathbf{x}+\Delta\mathbf{x}) = \mathbf{F}(\mathbf{x}) + R(\mathbf{x}^*)\Delta\mathbf{x}$ $\quad\quad$ where $\mathbf{x}^*$ lies between $\mathbf{x}$ and $\mathbf{x} + \Delta x$

Restate this claimed theorem as

$\mathbf{F}(\mathbf{x}'') = \mathbf{F}(\mathbf{x}') + R(\mathbf{x}^*)\Delta\mathbf{x}$ $\quad\quad$ where $\mathbf{x}^*$ lies between $\mathbf{x}$ and $\mathbf{x}''$ and $\Delta\mathbf{x} = \mathbf{x}''-\mathbf{x}'$

$\mathbf{T}(\mathbf{p}'') = \mathbf{T}(\mathbf{p}') + R(\mathbf{p}^*)\Delta\mathbf{p}$ $\quad\quad$ where $\mathbf{x}^*$ lies between $\mathbf{x}$ and $\mathbf{x}''$ and $\Delta\mathbf{x} = \mathbf{x}''-\mathbf{x}'$

Well NO, it is not that simple! The thing that I show as $R(\mathbf{p}^*)$ is a more complicated object. It is the R matrix, *but* each row of the R matrix elements is evaluated at a different point unknown point in the range $\mathbf{x}$ to $\mathbf{x}+\Delta\mathbf{x}$! Instead of there being a single point p* that works, you end up with N points, one for each dimension, called $\mathbf{p}_1^*$, $\mathbf{p}_2^*$ and so on. It comes out this way because you end up applying the 1D MVT to each component equation,

Nevertheless, this theorem is interesting because as with all the MVT, it is exact for finite $\Delta\mathbf{x}$.

$\mathbf{F}(\mathbf{x}+\Delta\mathbf{x}) = \mathbf{F}(\mathbf{x}) + R(\mathbf{x}_1^*, \mathbf{x}_2^*, \mathbf{x}_3^*)\Delta\mathbf{x}$ $\quad\quad$ $\mathbf{F}$: $E^3 \rightarrow E^3$

Note: This theorem for f: $\mathbf{R}^2 \rightarrow \mathbf{R}^1$ we already encountered on page 244A. There the two unknown points are called p' and p". I did not realize until now that this is just a special case of the above 3D stated theorem, which I think is valid in N dimensions with N unknown points $\mathbf{p}_i$*.

5.6 Inverses of a function of one variable **[270]**

I am used to doing these things in complex variable theory, but Bucks are not allowed to talk about such things at this point, so no Riemann sheets, no branch points and so on. Just real function theory.

Example 1 [270] is a good prototype, $f(x) = x^2$ on all of R. Write $y = x^2$ so you would say $x = \sqrt{y}$ was the inverse, but of course $x = -\sqrt{y}$ is also a solution. Also, in $y = x^2$ y is ≥ 0. So the original function really has the form f : E → upper half plane of E = range ($y \geq 0$). Then $x = \pm\sqrt{y}$ has a domain $y \geq 0$.

Graphically what happens here? The graph of y = f(x) we well know. It is locus [x,f(x)] in E^2 which is points where y = f(x), called the graph of f(x). Suppose x = g(y) exists as *an* inverse. It's graph is [y,g(y)] which we call "the second graph" - the graph of the inverse function. Now, let (a,b) be a point on the first graph, so b = f(a). Then we know that a = g(b), so (b,a) must be a point on the second graph. So we have shown that if (a,b) lies on graph 1, then (b,a) must lie on graph 2. We can graphically create graph 2 by making the swap x↔y which is reflection in the y = x line. In this way, you see in Fig 5-12 that an inverse of the function $y = x^2$ must lie on the dotted line. This dotted line encompasses both the $y = \pm\sqrt{x}$ inverses.

Example 2 [271]. I had always "rotated" functions to do this by 90 degrees, but this reflection is what you should really be doing! The example of p 271 is better in this regard. Here a 90 degree rotation does NOT give you the locus of the inverse, only the reflection does.

The next point is that when you reflect a function y = f(x), you in general get something that is multi-valued such as points A,B,C in Fig 5-13, and in order to get "a function" you must pick a single-valued piece of the thing. So in this example, there are 3 inverse functions, each with a different range. They call these three inverse functions g_1 and g_2 and g_3. In complex we would say

$w = (3/2)z - (1/2)z^3$ = cubic equation

solutions = $z_i(w)$ for i = 1,2,3

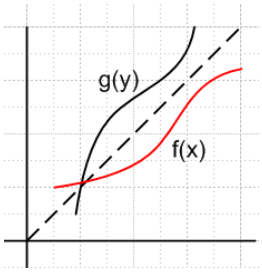
and you would find that each solution $z_i(w)$ was real only for a portion of the w axis.

Example 3 [272] is y = sin(x) which has an infinite number of inverses, and you usually pick the "principle inverse" (avoiding again the word branch).

Notice in the examples that the problem of multiple inverses only exists if the function has zero slope in the domain of interest. These will become infinite slope in the inverse plots, and those are points where different branch solutions meet!

Theorem 20 [272] If you have f(x) on a domain [a,b] where f(x) on which f'(x) $\neq 0$, and if range is [α,β], and if f(x) is C^1, then the inverse is unique *and* C^1 and can be found from the reflection method and you have then defined here f: D→R with f^{-1}: R→D which is one-to-one.

Here is my picture where the red f(x) has no points of f'(x) = 0. There is a unique inverse, and it is C^1.



In order even to talk about slope, you need f'(x) to exist and in addition this should be continuous for reasons I don't know (I could make counter example) and so f(x) has to be C^1. I skipped the long proof and surely it will require that f'(x) be continuous.

Interesting case: If $y = x^3$ on all of R (has zero slope only at x = 0), then reflect to get $x = y^{1/3}$ which is single-valued on all of R as well. How does Thm 20 apply to this? You seem to have a unique inverse despite the fact that f(x) has zero slope at x = 0. True, but the inverse has infinite slope at x = 0 and is therefore not C^1, so the theorem does in fact not apply! [OK, left part of $y = x^3$ really does reflect into the curve $y = x^{1/3}$. If x = -8, then $x^{1/3} = -2$. So this reflected curve $y = x^{1/3}$ has negative y values.]

5.7 Inverses of transformations **[274]**

Example 1 [274)] involves f:E^2→ E^2, domain and range are full plane. If $w = z^2$ in complex, we know that the full plane is mapped into the range plane twice so there are two Riemann sheets so this is a 1 to 2 mapping. The inverse is $z = \pm\sqrt{w}$ showing the two z values for each w value. If w = u+iv and z = x+iy, you find that $u = x^2-y^2$ and v = 2xy. This is their example (5-34) in the real world except u ↔ v for some reason. Any half plane in z will map into the full plane in w. With the domain restricted, then this mapping becomes one to one.

So how do we generalize the idea of f'(x) ≠ 0 in the domain to get to a 2D theorem about one to one and invertibility? We first define "the Jacobian" as on page 275. Unfortunately, their J is the inverse of my tensor doc Jacobian, but fine, not a big deal. Bucks J = det(R) = det(dT).

With no justification, Bucks "conjecture" that maybe J is the thing that replaces f'(z), and maybe if J ≠ 0, then you get one-to-one. This is of course their motivation for defining J as they did. The Jacobian meaning is always tied to how you define your transformation. Here is what tensor doc says

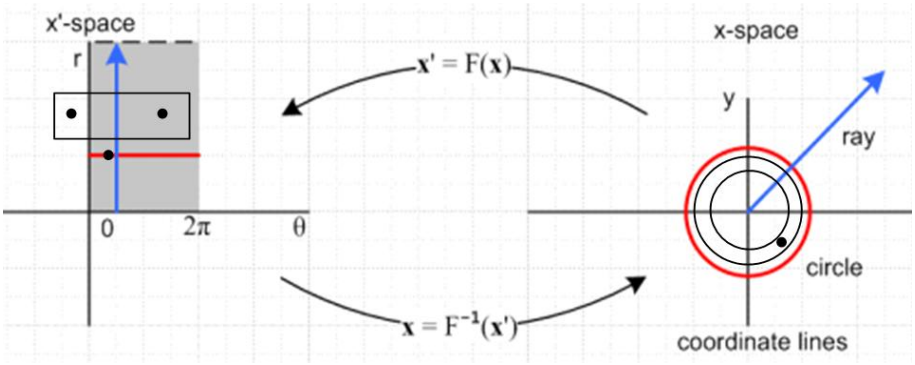
The Jacobian $J(\mathbf{x}')$ is defined as follows,

$$J(\mathbf{x}') \equiv \det(S(\mathbf{x}')) = \det(\partial x_i/\partial x'_k) = 1/\det(R(\mathbf{x}(\mathbf{x}')) = 1/\det(\partial x'_i/\partial x_k)$$

So if you take x' = curvilinear = r,θ,φ and your transformation of interest is z = rcosθ and so on, then my Jacobian comes out $r^2\sin\theta$. But in tensor doc this transformation is F^{-1} and not F.

So for now I just accept that $J_{Buck} = 1/J_{PL}$ for identifying y and x with x' and x.

Example 2 [275] I will translate to being x = rcosθ and y = rsinθ of polars. Here is my tensor doc picture



Bucks picture on page 276 has θ going vertically. Roughly their Fig p 276 is what I have added above, and the two points on the left map into the one point on the right because of the multiple sheets. They then rule out this idea by saying "locally one-to-one". I deal with the problem by restricting the domain to be the gray area which means selecting a sheet.

Locally one to one = locally univalent means that for any point in the mapping, there is a neighborhood around that point in which you have one-to-one. Just fine by me.

Theorem 21. [276] (**one-to-one theorem**) Let $T:E^n \rightarrow E^n$ and functions of T are C^1 on open set D. Then if $J \neq 0$ over D, the mapping is locally one to one.

The proof is straightforward and uses the earlier MVT with the 3 p* points.

The following Corollary does not do much for me right now so I ignore it.

Theorem 22. [277] If continuous T is one-to-one on closed bounded domain S, then the inverse transformation T^{-1} is also continuous, and of course maps set T(S) back into S.

There is no "*locally* one-to-one" here. So my tensor doc drawing above provides an example where S is the gray half-strip on the left, and T(S) is the entire plane on the right. In this case $T = F^{-1}$ which gets back to my Jacobian convention. There are still some "problem points" of course. Notice that S is closed but not bounded in my example, but we could cut off at any large R and then the theorem would apply.

Theorem 23 [278] If T is class C^1 on D and if $J \neq 0$ on all of D and if $D \rightarrow T(D)$ is one-to-one, *then*

(1) T^{-1} is also C^1

(2) $d(T^{-1}) = (dT)^{-1}$ // $S = R^{-1}$ in my language.

This theorem has a huge long proof filling 2 pages which involves their error thing R(Δp). I skip this proof since I am quite happy with the claims of the theorem.

Theorem 24 [279] If $T:E^n \to E^n$ is C^1 on open set D and if $J \neq 0$ on D, then T(D) is also an open set.

Usually we have the inverse of this idea, that continuous means open set in range back-maps into open set in the domain. But here we have it in the forward direction, with caveats as shown.

Question: How would you show f maps open to open? What does that mean? Any point q_0 in the range must have a ball around it that lies in the range. So back-map of said ball has to lie in the domain. This is what the proof sets out to show. The proof however is very long and a bit detailed, so as usual I skip the proof and accept the claim of the theorem.

Theorem 25 [281] (an existence theorem). If $T:E^n \to E^n$ is C^1 on open set D and if $J \neq 0$ on D, *then* :

For any p_0 in D you can find a ball N such that N → T(N) is one-to-one and T(N) is an open set, and you can find a workable T^{-1} inverse transformation, and $(dT^{-1})(dT) = 1$, for me this is SR = 1.

Example 1: They use the Cartesian to Polar transformation as an example. They compute dT^{-1} = my S and they show that dSdT = 1 which for them means $(dT^{-1})(dT) = 1$ and for me means SR = 1.

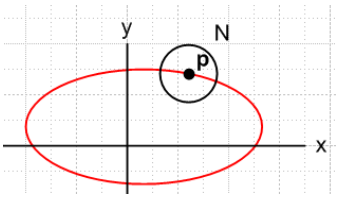
Perhaps Theorem 25 would apply if D is non-connected or some complicated sheet thing. You are still OK *locally* for an inverse. I guess this is the point they make in the page 276 polars picture. I can see that local inverse can handle a lot of fancy topology problems including sheets.

Final Claim [283] If T:D→D* is locally one-to-one, there are certain conditions which would make it be fully one-to-one. Two of those conditions are that D is compact and that D* is **simply connected**. But this is a Chapter 7 concept, so we have to wait. Vallee Poussin did whatever theorem this is in 1923 (Belgian).

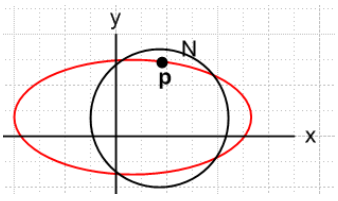
Comment: For many things, the Bucks provide the math rigor that I explicitly omit in tensor doc. It is nice to have a source where I can track down a real proof of something if I need it.

5.8 Implicit Function Theorems **[283]**

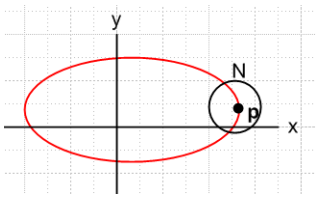
Consider F(x,y) = 0. In what sense do you know there is a solution to this thing of the form y = φ(x) ? This solution would be "implicit" in (implied by) F(x,y) = 0, and y = φ(x) would be called an implicit function. We know that F(x,y) = 0 describes some kind of curve in E^2. For example, if might be an ellipse as shown here (curve of course might not be closed, or might be a very strange spiral, etc.)



(a)



(b)



(c)

In (a) the red curve inside ball N is some assumed function $y = \varphi(x)$. We can see that this function "exists" for the N shown. For a different neighborhood N like that shown in (b), which encompasses two different pieces of the curve, the red locus is not a function since it is two valued. What we have here is really two different solutions (two different implied functions), and we don't want that. So the idea is that you can shrink the ball N such that it only contains one "piece" of the red curve, and then *within that* you have a well defined function. However, in (c) we try to do this at a point p which has $dy/dx = \infty$. In terms of $f(x,y) = 0$ we have $\partial_y f = 0$ which is to say we have $F_2 = 0$. As you vary y slightly, you stay on the curve $f = 0$ so $df = 0$ although $dy \neq 0$. The problem here is that no matter how much we shrink the ball N, the enclosed red curve is never a function because it is never single valued! So we have to restrict things so that $F_2(p) \neq 0$, and for some domain D of p, we need $F_2(p) \neq 0$ for the entire domain. In that case, for any p in D we have a well-defined implied solution function $y = \varphi(x)$ and of course $y_0 = \varphi(x_0)$ at the point p_0.

What does this look like in 3D? Think of the ellipse as an ellipsoid, and think of the vertical axis as z, the x axis as shown, and the y axis into the plane of paper. The ellipsoid is $F(x,y,z) = 0$. In this example, the ellipsoid surface represents two distinct functional surfaces over the x-y plane, each of them is a separate function really. In Fig (a) we put a 3D ball around point p and it encloses a little patch of ellipsoidal surface, and this patch is single valued and this an actual function, and we have $z = \varphi(x,y)$ describing this function, and of course $z_0 = \varphi(x_0,y_0)$. In (b) the N is too big, so no go. In (c) we have a point p on the surface F such that $\partial F/\partial z = 0$. You go "up" a little and F does not change. This says $F_3 = 0$. In this case, you cannot enclose a true function no matter how small you make the 3D ball. You always enclose two different patches which result in 2 solutions at least for some points on the curve. So in this case, you have to avoid places where $F_3 = 0$ for your function $F(x,y,z)$. If you then have a D which avoids such places, then inside a sufficiently small N ball you will have a well-defined single piece of surface and that is the function $z = \varphi(x,y)$.

Theorem 26: [284] Says what the above paragraph says! I think my geometric view is pretty clear and clean. The theorem assumes F is C^1 and the solution patch will also be C^1 within small N.

Details: The problem here is to start with $w = F(x,y,z)$ mapping $E^3 \rightarrow E^1$ and we know that $F(x_0,y_0,z_0) = 0$ and $F_3 \neq 0$ there. (point on ellipsoid, say). Then locally to this point p_0 we can solve to get $z = \varphi(x,y)$. This is the words that go with the video above. To deal with this problem, Bucks so this

$u = x$
$v = y$
$w = F(x,y,z)$

and thereby create a mapping $E^3 \rightarrow E^3$ taking $(x,y,z) \rightarrow (u,v,w)$. For this mapping, it turns out that $J = F_3$ see p 284. Since $J \neq 0$, a local inverse exists (earlier theorem 25)which means you know

$x = u$
$y = v$
$z = f(u,v,w) = f(x,y,w)$ // this is the 3x3 inverse

We can then say that the following is true for x,y,z *near* the point p_0 meaning for small values of w, since w = 0 puts you right at the point.

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

The magic claim then is that this is also true for w = 0 where we get

$$0 = F(x,y, f(x,y,0))$$

But the third argument is z, so we then claim that

$$z = f(x,,y,0) \equiv \varphi(x,y)$$

and thus we have found our "implicit function" φ(x,y).

Note: This just tells you that the solution exists, it does *not* tell you want the solution function is ! You have to go do that problem on your own somehow. Since things are linear near the point of interest, it should be a Cramer's Rule problem of some sort.

How about in 4D? We have F(x,y,z,t) = 0 and we want t = φ(x,y,z). We cannot visualize this but we know what the result will be. The function t = φ(x,y,z) exists and is unique if $F_4 \neq 0$ in our neighborhood.

Bucks give a proof of the 3D case only, but I presume the same proof would work for nD as well. Their proof for 3D creates a transformation T which has $J = F_3$. Then they get to make use of Theorem 25 which says N exists and you can compute T^{-1} as they show. They state on p 285 that in fact the same proof does in fact work for any nD case.

Now, what happens if we have more than one equation such as shown in (5-47) ? They use as an example two equations each of which maps $E^5 \rightarrow E^1$. Like F(x,y,z,u,v) = 0 and G(x,y,z,u,v) = 0 . You would like to eliminate two of the unknowns u and v and find u = φ(x,y,z) and v = ψ(x,y,z). I think basically we know we can do this in a tiny neighborhood N because in such a tiny N things are linear and we know it is a solvable problem, but we have to watch out for the singular situation. We are in effect changing variables from x,y,z,u,v to x,y,z,F,G and we know things will be OK as long as this big 5x5 $J \neq 0$, because that is what our transformation theory said above. But this 5x5 J is the same as the 2x2 J since 1's on diagonals in 3 places, so we suspect we will be OK as long as the 2D $J \neq 0$ as shown in steps A,B,C on p 285. If this $J \neq 0$, there is a local inverse of the transform as shown p 286 E.

Details: We want to solve these two equations for implicit functions u = φ(x,y,z) and v = ψ(x,y,z)

F(x,y,z,u,v) = 0	maps $E^5 \rightarrow E^1$
G(x,y,z,u,v) = 0	maps $E^5 \rightarrow E^1$

Construct a 5x5 transformation as shown in page 285 B. Compute J for this 5x5 and you find that $J = F_4G_5 - G_4F_5$ where these are derivatives as in C or as in D which shows this as a little F,G,u,v Jacobian.

Now as in this 5x5 case, if we avoid this thing being 0, we can write the inverse equations for t_1 through t_5 as shown in p 286 E. So we then have the following inverse which we know exists,

$x = t_1$
$y = t_2$
$z = t_3$
$u = f(t_1,t_2,t_3,t_4,t_5)$
$v = g(t_1,t_2,t_3,t_4,t_5)$

Now set $t_4 = t_5 = 0$ just as we set w = 0 in the previous 3x3 example and we are in a local region,

$u = f(x,y,z,0,0) \equiv \varphi(x,y,z)$
$v = g(x,y,z,0,0) \equiv \psi(x,y,z)$

So as long as we avoid places where the Jacobian F,G,u,v = 0, we have solved our equations

Theorem 27 [285] is thus proved. The domain D in E^5 is an open set, the functions G and F are C^1. The solution functions are the "implicit functions" implied by the original function set.

Comment: Here we had m = 2 equations in n = 5 unknowns, and we found some "implied" solutions for 2 of the variables. In general you might have m < n equations in n unknowns $F^{(i)} = 0$, and you could then solve for m of the variables in terms of the first n-m variables. Such as solution for those m variables is called an "implicit solution" of the original equation set.

The section closes with a simple **Example 3** of Theorem 27 with n = 4 and m = 2 so F(x,y,u,v) = 0 etc. They get an actual solution here, so I will try to do that too:

$x^2-yu = 0$ F = 0
$xy+uv = 0$ G = 0

Well, I can see from the first equation alone that $u = x^2/y$ and then the second says $xy + v\,x^2/y = 0$ and then we get $y + vx/y = 0$ and $y^2+vx = 0$ and $v = -y^2/x$. The Jacobian is

$$\partial(F,G)/\partial(u,v) = \begin{vmatrix} \partial_u F & \partial_u G \\ \partial_v F & \partial_v G \end{vmatrix} = \begin{vmatrix} -y & v \\ 0 & u \end{vmatrix} = -yu$$ /agrees

So solution must now have y = 0 and must now have x = 0 since that makes u = 0, so stay away from axes! And this is obvious from the solution anyway.

How can we generalize this analysis? In the last case above, we had 3 independent variables x,y,z and we had 2 dependent variables u and v, and we had 2 equations F = 0 and G = 0. To generalize, we could have I independent variables x_i, J dependent variables u_j, and J equations of the form $F_j = 0$. Write

$F_j(x_1, x_2.....x_I, u_1, u_2....u_J) = 0$ j = 1,2...J

We want to solve for these implicit functions,

$u_j = \varphi_j(x_1, x_2.....x_I)$ $\quad\quad$ $j = 1,2...J$

We treat this as a (J+I)x(J+1) problem and find that Jacobian J = Jacobian(F_i, u_j) which is JxJ. If we avoid points where J = 0, then we this initial problem

$t_i = x_i$ $\quad\quad$ $i = 1,2..I$
$t_{I+j} = F_j(x_i,u_j)$ $\quad\quad$ $j = 1.2..J$ $\quad\quad$ the transform

$x_i = t_i$ $\quad\quad$ $i = 1,2...I$
$u_j = f_j(t_1..t_I, t_{I+1}..t_{I+J})$ $\quad\quad$ $j = 1,2....J$ $\quad\quad$ the inverse transform

We then set the higher t's to zero to make the J equations of the form $F_j = 0$ be valid, We then get

$u_j = f_j(t_1..t_I, 0..0) = f_j(x_1..x_I, 0..0)$

and there are the solution "implicit functions".

In the local sense, one can regard our transformation as linear if $J \neq 0$ and then we have in that local region a set of J equations $F_j = 0$ which we want to solve for the J unknowns u_i in terms of the x_i.

Now how would tensor doc approach this problem? Well, there is a dimension I+J R matrix $R_{ij}(\mathbf{x}_0)$ at some point of interest $\mathbf{x}_0$ where Jacobian $\neq 0$. Lets try this idea:

$d\mathbf{x} = \{ x_i - x_i^0, u_j - u_j^0 \}$ $\quad$ has I+J components

$d\mathbf{x}' = \{t_i\}$ $\quad\quad$ has I+J components, these are all 0 when $\mathbf{x} = \mathbf{x}^0$ let's say

We are then given :

$t_i = x_i - x_i^0$ $\quad\quad$ $i = 1,2....I$
$t_{I+j} = F_j(x_i - x_i^0, u_j - u_j^0)$ $\quad$ $j = 1,2...J$
OR
$d\mathbf{x}' = R d\mathbf{x}$

Now our only task is to invert R to get S, and then we have

$d\mathbf{x} = S\, d\mathbf{x}'$
OR
$x_i - x_i^0 = t_i$ $\quad\quad$ $i = 1,2..I$
$u_j - u_j^0 = f_j(t_i)$

When you set the high $t_i = 0$, you are forcing the equations $F_j = 0$ to be 0. I think the Bucks have quietly assumed that the point of interest is $\mathbf{x}_0 = 0$ which simplifies things the way they present it. But the theory

as I just outlined it above to show existence and mechanically how you do it: first, compute R, which is the linearized transformation near x_0. Then invert that to get S, and then you have it.
ok to here

5.9 Functional Dependence **[287]**

Up to now we have been mainly interested in f: $E^n \rightarrow E^n$ where $J \neq 0$ which corresponds to a transformation having full rank n. That is to say, the linearized R will have rank n. But if J = 0 in some region, then that corresponds to R having rank < n. What are the implications of J = 0 in more detail?

Example 1 [288] . We have an f:$E^2 \rightarrow E^2$ which has J = 0, so rank(R) < 1. We can see that there is a relation between u and v of the form F(u,v) = 0, a sort of constraint. It is that $u^2 + v^2 = 1$. We started here with u = f(x,y) and v = g(x,y) but we noticed this relation. The presence of F(u,v) = 0 tells us that the functions f and g are "functionally dependent". You are saying that F(f,g) = 0 with this extra equation which you can write as g = h(f). More generally, suppose you have f:$E^n \rightarrow E^n$ and you have your various equations (n of them) which say $y_i = f_i(x_1, x_2 ..... x_n)$ defining a transformation. But then someone hands you a new equation which says $g(f_1, f_2 .... f_n) = 0$. This equation indicates that the functions of the set $\{f_i\}$ are "**functionally dependent**". If all functions were linear, they would be "**linearly dependent**". In out Example 1 we have $f_1, f_2 = u, v$ and $g = u^2+v^2-1$ and g = 0 says that $u^2+v^2 = 1$ which forces the entire mapping range to be a unit circle. One can write $f_2 = \pm\sqrt{f_1^2-1}$.

In the example, the dimensionality of the domain is 2, but of the range is 1. When J = 0 for D, this is always what happens, the range loses some dimensionality.

I think you can study things in terms of the linearized matrix R, and then that carries over to the full transformation (will be proved below). We know this for linear **F**(**x**) // N = nullspace

rank(S) = 0	=>	**F**(**x**) = 0 for all x	dim(N) = n	dim(range) = 0
rank(S) = 1	=>	**F**(**x**) = 0 for x in N	dim(N) = n-1	dim(range) = 1
rank(S) = 2	=>	**F**(**x**) = 0 for x in N	dim(N) = n-2	dim(range) = 2

For nonlinear F, if you study it close to some point where it is linear, you get these same results in that tiny ball world surrounding the point. In the larger world, instead of getting a line in the domain, you will get a curved line, and instead of getting a plane, you will get a curved surface, and so on. At each point you match the linear transformation. But the dimensionalities are the same! That is, a line and a curve both have dimensionality 1.

Here we consider f: $E^3 \rightarrow E^3$ which is non-linear but has some matrix R. Here are the choices:
[289]

	rank(R) = 3	=>	$J \neq 0$	range = all of E^3
Theorem 28:	rank(R) = 2	=>	J = 0	range = surface in E^3
Theorem 29:	rank(R) = 1	=>	J = 0	range = curve in E^3
	rank(R) = 0	=>	J = 0	range = point at the origin (I guess)

Think of (x,y,z)→(u,v,w).

In the rank 2 case above, a surface in the range space is something like w = w(u,v) just as we would write z = f(x,y) for a surface above the x-y plane. Again, it might be better as u = u(v,w), and in general it is going to be some h(u,v,w) = 0.

In the rank 1 case above, you get a *curve* in the range. I know that a curve can be expressed in parametric form as

u = u(t) and v = v(t) and w = w(t) // describes a curve in 3D space, parameter t

If you take t = u as parameter, this becomes

u = u and v = v(u) and w = w(u) // describes a curve in 3D space

and you ignore the first, so you have then a *pair* of equations. There are of course two other ways to take the pair of "constraints" since two other choices of the parameter.

Now, how do you actually prove the conjecture above about dimensionalities of the range being based on the rank of the linearized R matrix??

Proof of **rank 1 case:** [this is a proof of Theorem 29]

1. Assume $f_1 \neq 0$ in R is a non-zero 1x1 in R since rank = 1
2. Show that x = K(y,z,u) from implicit function theorem (since $f_1 \neq 0$ and x = coordinate 1)
3. Show that v = G(y,z,u) and w = H(y,z,u) [trivial]
4. Show that y is really a ghost in these two equations. Ie, show $\partial_y G = 0$ and $\partial_y H = 0$.This part makes explicit use of the fact that rank = 1 !
5. Since a ball I guess is convex, conclude from Thm 12 that y is ghost in G and H
6. Repeat step 3 with y→ z to show that z is also a ghost
7. End up with v = G(u) and w = H(u) and u = u and these are parametrics for a **curve**, QED.

Proof of **rank 2 case**: [this is a proof of Theorem 28]

1. Assume $\partial(f,g)/\partial(x,y)$ is a non-zero 2x2 in R since rank 2

2a.The equations u = f(x,y,z) and v = g(x,y,z) are same as A(u,x,y,z) = 0 and B(v,x,y,z) = 0

2b. Theorem 27 says you can solve these last two equations to get x = F(u,v,z) y = G(u,v,z) where only the remaining variables appear. [this is result A]

3. Shove these into w = h(x,y,z) to get w = H(u,v,z)
4. Show that $\partial_z w = 0$ and thus z is a ghost. To do this, first apply ∂_z to (5-49) to get 3 equations as shown in C, then solve these by Cramer's for $\partial_z w$ as in D . But J = 0 by hypothesis, *and* the denom $\partial(f,g)/\partial(x,y) \neq 0$ also by hypothesis since rank 2. Thus $\partial_z w = 0$ and z is a ghost.
5. Therefore w = H(u,v,z) = H(u,v), and this describes a **surface**, QED.

These are pretty fancy proofs IMHO! There is obviously some fully general case theorem you could attempt to prove mimicking the above proofs, but Bucks are not going to do that! I get the idea! It is just a warped version of what happens with the linear R.

All the above was for F: $E^3 \rightarrow E^3$, here we consider instead F: $E^3 \rightarrow E^2$. In this case, we know that 2x2 is the largest possible subdeterminant (for linearized R) so full rank here means rank = 2. So think of this situation as u = f(x,y,z) and v = g(x,y,z). Let's again look at the cases

rank(R) = 2	=>	$J \neq 0$	range = area in E^2
rank(R) = 1	=>	J = 0	range = curve in E^2
rank(R) = 0	=>	J = 0	range = point at the origin in E^2 (I guess)

Theorem 30. [292] says: [as usual, T is class C^1 and D = open set]

rank = 2: f and g functionally *independent*, open set D in $E^2 \rightarrow$ open set in E^2

rank = 1: f and g functionally *dependent*, open set D in $E^2 \rightarrow$ curve in E^2

This is the expected result for T: $E^3 \rightarrow E^2$. Short proofs given below.

Corollary (to Thm 30). [292]. If all the three 2x2 Jacobians shown are 0, then we know that the 2x3 linearized R matrix must have rank 1 (or 0), not rank 2. If rank 1, we showed in Thm 30 that the trace is a curve in E^2 and a curve can be represented parametrically as u = u(t) and v = v(t). Taking u as parameter this becomes a single equation v = v(u) or taking v we get u = u(v). If rank 0 then (u,v) = (0,0) but this could still be put in the same form such as v = v(u) where $v \equiv 0$ or u = u(v) where $u \equiv 0$. Presumably this rank 0 case is never of much interest.

Proof of Theorem 30. Add third variable w = 0 so have (x,y,z)→(u,v,w) and now $E^3 \rightarrow E^3$. The new 3x3 R matrix has the form shown top p 293 A which you see has rank 2 or rank 1. But this situation is handled already by Theorem 28 (for rank 2) and Theorem 29 (for rank 1). In the first case, we know we map to a surface in E^3 but with w = 0, such a surface is a 2D area and I guess it is an open set from earlier theorem. In the second case, Thm 29 says we map to a curve in E^3 and again this becomes a curve in E^2 if w = 0.

We now look at the reverse situation, that of F: $E^2 \rightarrow E^3$. Again, full rank here means rank = 2. Our equations are then x = f(u,v), y = g(u,v), z = h(u,v).

Theorem 31 [292] says: [as usual, T is class C^1 and D = open set]

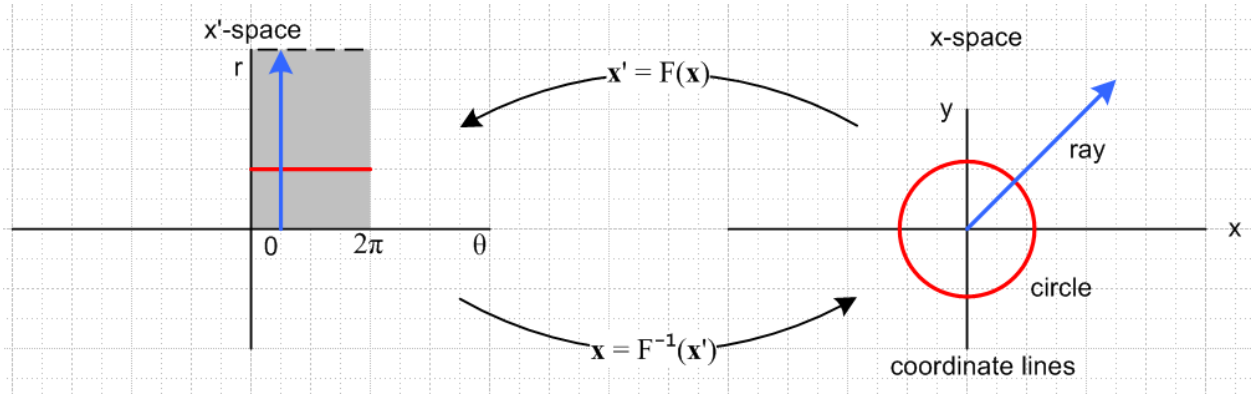
rank = 2: range for D is a surface in E^3

rank = 1: range for D is a curve in E^3

Proof of Theorem 31. Add dummy ghost variable w so now x = f(u,v,w), y = g(u,v,w), z = h(u,v,w). The new 3x3 R has the form shown p 293 B which you see has rank 2 or 1. But this situation is handled already by Theorem 28 (for rank 2) and Theorem 29 (for rank 1). In the first case, we know we map to a surface in E^3. In the second case, Thm 29 says we map to a curve in E^3.

Result 293 C: If the linearized matrix in B is rank 2, then at least one of the 2x2 sub Jacobians must be non-vanishing, and that then forces the sum shown in C to be positive, very simple.

Polar coordinates as an example:



J = -r the way I compute it in tensor doc. For me this is for one direction, and I guess for Bucks this would be for the other direction.

For a point out in the gray area including its left and right edges, we have $J \neq 0$ and so we have full rank and we have area → area. However, for a point on the r = 0 bottom segment, we have J = 0, so Bucks stuff applies (maps to point!). The R matrix is one of these

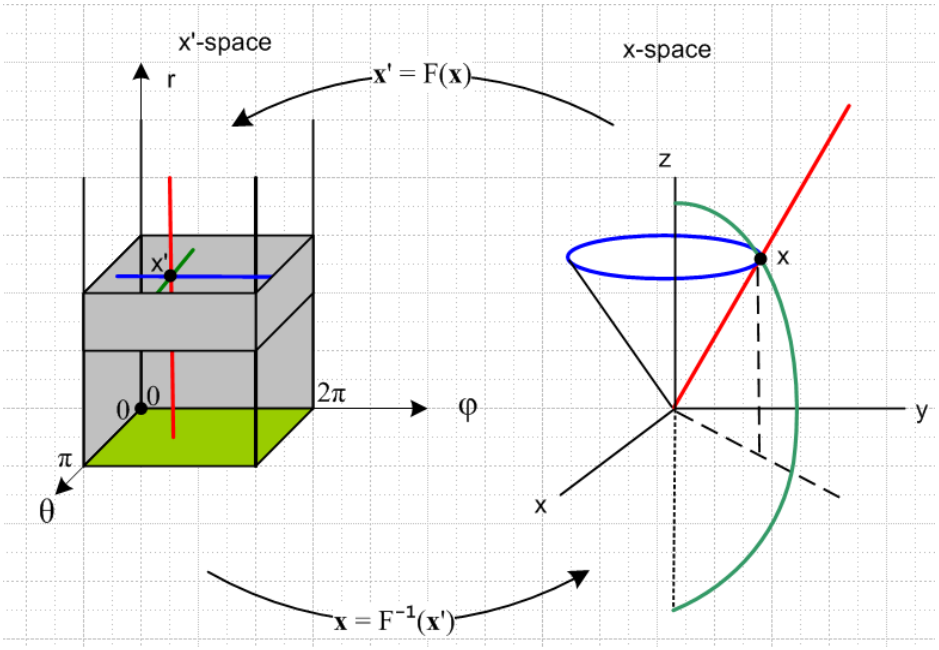
$$R = S^{-1} = \begin{pmatrix} -\sin\theta/r & \cos\theta/r \\ \cos(\theta) & \sin\theta \end{pmatrix} \qquad S = \begin{pmatrix} -r\sin\theta & \cos\theta \\ r\cos\theta & \sin\theta \end{pmatrix}$$

Assume it is S that applies, so that r = 0 means rank(S) = 1 (never 0). We know that the bottom segment with r = 0 all maps into the origin since x = rcosθ etc. So we have a loss of 1 dimension tick here! This would lie in the $E^2 \rightarrow E^2$ world which Bucks never directly treated, but the conclusion seems exactly right. What about the left and right edges of the gray? On these rank(S) = 2 = full so there is no loss of dimension. It just happens that both edges map to the same ray. This is the locally one-to-one issue, not a loss of dimensionality.

<u>Spherical coordinates as an example:</u>

$$S = \begin{pmatrix} \sin\theta\cos\varphi & r\cos\theta\cos\varphi & -r\sin\theta\sin\varphi \\ \sin\theta\sin\varphi & r\cos\theta\sin\varphi & r\sin\theta\cos\varphi \\ \cos\theta & -r\sin\theta & 0 \end{pmatrix} \qquad R = \begin{pmatrix} \sin\theta\cos\varphi & \sin\theta\sin\varphi & \cos\theta \\ \cos\theta\cos\varphi/r & \cos\theta\sin\varphi/r & -\sin\theta/r \\ -\sin\varphi/(r\sin\theta) & \cos\varphi/(r\sin\theta) & 0 \end{pmatrix}$$

$J = r^2\sin\theta$



In this case, for a general point inside the gray cube, $J \neq 0$ and volume → volume, no issues.

For the bottom green plate where r = 0, we do have J = 0. In this case we have

$$S = \begin{pmatrix} \sin\theta\cos\varphi & 0 & 0 \\ \sin\theta\sin\varphi & 0 & 0 \\ \cos\theta & 0 & 0 \end{pmatrix}$$

and this is rank 1. In our Buck analysis, in this case volume D → curve, so perhaps area D → point. We know that at r = 0 (green square) entire square maps to the origin, so yes, it is a point.

There is also a problem at θ = 0 where again J = 0 and we then have

$$S = \begin{pmatrix} 0 & r\cos\varphi & 0 \\ 0 & r\sin\varphi & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

This has rank 2 for general φ. Thus, we expect the back face of the cube (θ=0) to have vol → area, which then means area → curve. In fact, the back face maps onto the positive z axis, which is a curve!

Suppose φ = 0 as well? Then

$$S = \begin{pmatrix} 0 & r & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \text{rank } 1$$

Here we expect vol → curve and area → point and curve → ?? The locus here is the vertical left back edge of the cube at $\theta = 0$. As we go up in r for that segment, $\varphi = 0$ all the time, and we are going up the positive z axis. We seem to have curve → curve which seems wrong by Bucks. Ignore.

What about general $\varphi = 0$ which is the left face of the cube. Then

$$S = \begin{pmatrix} \sin\theta & r\cos\theta & 0 \\ 0 & 0 & r\sin\theta \\ \cos\theta & -r\sin\theta & 0 \end{pmatrix}$$

which is rank 3. Expect area → area. so this left face maps to the half disk in green so Bucks OK.

p is a **critical point for T** when rank(T(p)) < full rank. [293]

Example 1 [294] which is $E^2 \rightarrow E^2$ with $u = x^2$ and $v = y^2$. S is shown and is obvious. The entire E^2 plane maps into quadrant 1 as suggested by the figure shading. Obviously the mapping is 4 to 1 in general. On either axis, the mapping is 2 to 1, and at the origin the mapping is 1 to 1.

The second figure (b) is strange. They are trying to show that each of the four domain quadrants maps into the same quadrant 1. In the final picture there are four "leafs" of paper, and this is the general 4:1 situation. Each sheet is a different source point, and all four source points go to the same range point. I drew a line showing in general that a range point has 4 source points. If I move that line to the left edge I get the 2:1 situation. If I draw a line on the bottom edge, I guess there are still two sheets there. Origin line gives a single point. Well as they say, just a crude attempt to illustrate something.

In this example, all points on the two axes are critical points, including the origin. Now 2.22.15.
[I did not get back here until 5.15.15, so about 3 months has gone by!!!]