This is the Title PhL 1.11.15

Note that page numbering is turned on in this template and view is 125%,
located in phil/roaming/microsoft/templates size about 219K.

Suppose matrix B has n rows and m columns. Consider this matrix equation where B acts to the left on vector **x** to create vector **y**.

$$\mathbf{y} = \mathbf{x}B \qquad (y_1\, y_2\, \ldots\ldots y_m) = (x_1, x_2, \ldots x_n)\begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \ldots \\ \mathbf{r}_n \end{pmatrix}$$

The vectors $\mathbf{r}_i$ are now the rows of matrix B. Suppose N of these rows are linearly independent, so the row rank of matrix B is then N. Then in analogy with *** we can write

$$(y_1\, y_2\, \ldots\ldots y_m) = (\, x_1\mathbf{r}_1 + x_2\mathbf{r}_2 + \ldots\ldots x_N\mathbf{r}_N\,) + \; (x_{N+1}\, r_{N+1} \ldots\ldots x_n\mathbf{r}_n) = (k_1\mathbf{r}_1 + k_2\mathbf{r}_2 + \ldots\ldots k_N\mathbf{r}_N\,)$$

We conclude this time that the dimension of the range equals N which is the row rank of matrix B.

Now suppose m = n so that matrix B is a square matrix. Suppose furthermore that $B = A^T$, so B is the transpose of matrix A which is then also square. In *** we showed that the dimension of the range of the transformation $\mathbf{y} = A\mathbf{x}$ was equal to the column rank of A. We can write this as $\mathbf{y}^T = \mathbf{x}^T A^T$ or $\mathbf{y} = \mathbf{x}A^T$ where y and x are row vectors. We just showed that the row rank of A^T is equal to the dimension of the range. So we then know that row rank (A^T) = column rank (A).

This method then tells us nothing at all!!!

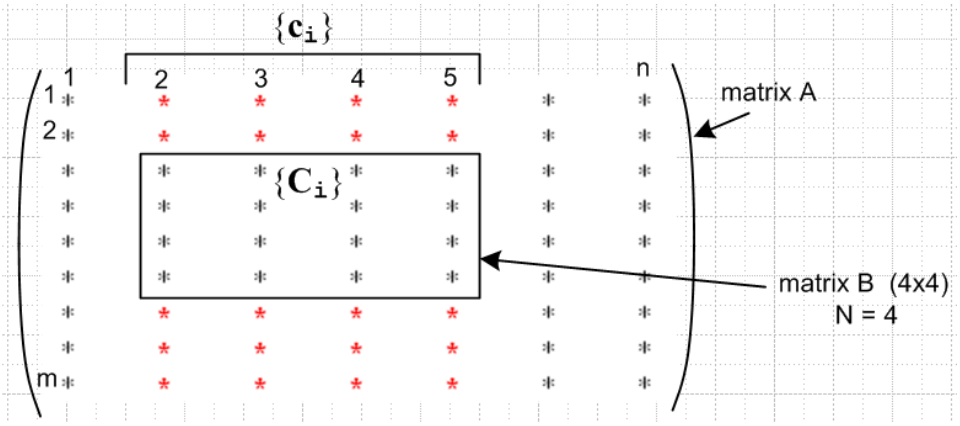
&**********

Theorem A.5. Suppose a set of N columns of matrix A are linearly independent. Assume each column has m components, so column i has components { $(\mathbf{c}_i)_1$, $(\mathbf{c}_i)_2$, ... $(\mathbf{c}_i)_m$}. Now form a set of N mini-columns by taking for each column the same subset of the full column components. Suppose these mini-columns then have J components where J < m. Consider then the JxJ matrix formed from these mini-columns. Then claim of the theorem is that within that JxJ matrix, the N mini-columns are linearly independent. THIS IS WRONG!

Theorem A.5. Consider a matrix with m rows and n columns. We can write A = ($\mathbf{c}_1$, $\mathbf{c}_2$....$\mathbf{c}_n$). Pick a subset {$\mathbf{c}_i$} of these columns which contains N column vectors where N ≤ n and N ≤ m. Then create N reduced-size column vectors {$\mathbf{C}_i$} by knocking out some set of components from all these {$\mathbf{c}_i$} such that the resulting $\mathbf{C}_i$ vectors have exactly N components. For example, we might remove rows 2, 3 and 5 of all the {$\mathbf{c}_i$} to create the {$\mathbf{C}_i$} Construct an NxN matrix B from these N $\mathbf{C}_i$ vectors. This matrix B is of course some NxN submatrix of the larger nxm matrix A. Here then is the theorem: If the {$\mathbf{C}_i$} form a set of linearly independent vectors within the NxN matrix B, then the set {$\mathbf{c}_i$} form a set of linearly independent vectors for the larger matrix A.

$\det(B) \neq 0 \Rightarrow \{\mathbf{C_i}\}$ lin indep $\Rightarrow$ $\{\mathbf{c_i}\}$ lin indep

Before stating the trivial proof, here is a picture to clarify the setup for this theorem:



We happened to select for $\{\mathbf{c_i}\}$ the contiguous set of columns $\{\mathbf{c_2}, \mathbf{c_3}, \mathbf{c_4}, \mathbf{c_5}\}$. This causes the 4x4 matrix B to be contiguous in this picture. If the selected column set $\{\mathbf{c_i}\}$ were non-contiguous, one creates the set $\{\mathbf{C_i}\}$ and assembles them into a matrix B off to the side.

Proof of theorem: Suppose the four short columns $\mathbf{C_i}$ of matrix B are linearly independent. Suppose the set $\{\mathbf{c_i}\}$ were linearly *dependent*. Then one could write $\Sigma_{i=2}^{5}\, k_i \mathbf{c_i} = 0$. This is a set of m equations which contains as a subset the set of N equations $\Sigma_{i=2}^{5}\, k_i \mathbf{C_i} = 0$. But this then says that the $\mathbf{C_i}$ are linearly dependent which is a contradiction. Therefore the set $\{\mathbf{c_i}\}$ must be linearly independent within A, given that the set $\{\mathbf{C_i}\}$ is linearly independent within B. Extending the size of a set of linearly independent vectors cannot make the extended set become linearly dependent.

Theorem A.6. The setup for this theorem is the same as for the previous theorem. This time we imagine constructing all possible NxN matrices B for our selected set $\{\mathbf{c_i}\}$. We do this by knocking out different sets of rows to get the $\{\mathbf{C_i}\}$. If at least one of these B matrices has a non-zero determinant, then the $\{\mathbf{c_i}\}$ are linearly independent:

$\det(B) \neq 0$ for at least one B $\Rightarrow$ $\{\mathbf{c_i}\}$ are linearly independent

Proof: We use any B which has $\det(B) \neq 0$ in the Theorem A.5, since $\det(B) \neq 0$ implies that the $\{\mathbf{C_i}\}$ which make up matrix B are linearly independent.

Theorem A.7. This is the converse of Theorem A.6:

N columns $\{\mathbf{c_i}\}$ are linearly independent $\Rightarrow$ $\det(B) \neq 0$ for at least one NxN submatrix B

We shall prove the contrapositive which states:

det(B) = 0 for all B $\Rightarrow$ $\{\mathbf{c}_i\}$ are linearly dependent .

Proof: To make this proof simple, assume that the N linearly independent columns of A are the leftmost N columns, so the set $\{\mathbf{c}_i\} = \{\mathbf{c}_1, \mathbf{c}_2....\mathbf{c}_N\}$. Start off with the B matrix formed from the first N rows of the column set $\{\mathbf{c}_i\}$. Since det(B) = 0, the $\{\mathbf{C}_i\}$ which form this NxN B matrix are linearly dependent, so there are constants k_i such that $\Sigma_{i=1}^{N} k_i \mathbf{C}_i = 0$. In components this says $\Sigma_{i=1}^{N} k_i (\mathbf{c}_i)_j = 0$ for the first N rows j = 1,2.. N. Assume these constants are scaled so $k_1 = 1$. Consider $\Sigma_{i=1}^{N} k_i (\mathbf{c}_i)_j = 0$ just for the rows j = 2,3....N which is (N-1) rows. One then has N-1 equations in N-1 unknowns which are the k_i other than k_1. One can then solve for to get $k_i = f_i[(\mathbf{c}_a)_b]$, some functions of the matrix elements.

Now move the B matrix down one row. Since this new B matrix also has det(B) = 0, its columns are also linearly dependent, so there exists some set of constants k'_i such that $\Sigma_{i=1}^{N} k'_i (\mathbf{c}_i)_j = 0$ for the rows j = 2,3..N+1, again with $k'_1 = 1$. But consider $\Sigma_{i=1}^{N} k'_i (\mathbf{c}_i)_j = 0$ only for rows j = 2,3..N. These are exactly the same N-1 equations encountered with the previous B matrix, so the solution k'_i will be exactly the same, and therefore $k'_i = k_i$. Repeat this process until the B matrix reaches the bottom of the A matrix. At this point one has shown that in fact $\Sigma_{i=1}^{N} k_i (\mathbf{c}_i)_r = 0$ for all rows r of the A matrix and therefore $\Sigma_{i=1}^{N} k_i \mathbf{c}_i = 0$ and therefore the first N columns of matrix A are linear dependent, QED.

Comment: The above proof is no good. You cannot assume $k_1 = 1$ and $k_1' = 1$. The proof just completely collapses, several hours down the drain.

Question: Suppose you have three linearly independent tall column vectors with m elements. Can you do the usual row operations without changing the fact that the columns are independent?

I think the answer is yes but I have no idea how to show it.

****

****

Conjectured Theorem/

Suppose you have N linearly *independent* column vectors each with m ≥ N components.
Theorem: you cannot have all NxN subdeterminants vanish!

Try simplest case. Suppose you have 2 independent column vectors. Could you have all 2x2 determinants vanish?

Suppose m = 3 so we then have 3 subdeterminants to think about.

$$\begin{pmatrix} c_1 & d_1 \\ c_2 & d_2 \\ c_3 & d_3 \end{pmatrix}$$

If all dets vanish, then we have

$c_1d_2 - c_2d_1 = 0$
$c_1d_3 - c_3d_1 = 0$
$c_2d_3 - c_3d_2 = 0$

Suppose **c** is known. Then above is then 3 equations in the three unknowns **d**. Assume some solution exists and we find

$d_1 = f(\mathbf{c})$
$d_2 = g(\mathbf{c})$
$d_3 = h(\mathbf{c})$

This does not seem to imply that **d** is a multiple of **c**.

Construct a counter example?

$$\begin{pmatrix} 1 & \sqrt{6} \\ \sqrt{6} & 6 \\ c_3 & d_3 \end{pmatrix}$$

Go back to

$$\begin{pmatrix} c_1 & d_1 \\ c_2 & d_2 \\ c_3 & d_3 \end{pmatrix}$$

Suppose all 2x2 dets are 0. Then row 1 is a multiple of row 3, so $\mathbf{r}_1 = \alpha\, \mathbf{r}_3$. Similarly, $\mathbf{r}_2 = \beta \mathbf{r}_3$. Then

$(c_1, d_1) = \alpha(c_3,d_3)$ and $(c_2, d_2) = \beta(c_3,d_3)$

The matrix is then

$$\begin{pmatrix} \alpha c_3 & \alpha d_3 \\ \beta c_3 & \beta d_3 \\ c_3 & d_3 \end{pmatrix}$$

If you multiply the first column by d_3/c_3 you get the second column, so the columns are then linearly dependent! So this sort of supports my conjectured theorem.

https://books.google.com/books?id=cMLCAgAAQBAJ&pg=PA149&lpg=PA149&dq=rank+subdeterminants&source=bl&ots=s_3u4dmqiY&sig=mIrrCo5qnsmH5vYBwpsB_CgGhpc&hl=en&sa=X&ei=lStqVZahMtPWoAS174OQCA&ved=0CDoQ6AEwBA#v=onepage&q=rank%20subdeterminants&f=false

3.9.3 THEOREM. *Any* d *rows (columns) of an* m × n *matrix* A *that contain a nonvanishing subdeterminant* |D| *of* A *of order* d *are linearly independent.*

The rows (columns) of D are linearly independent, since D is non-singular. Hence the d rows (columns) of A are linearly independent, for a linear relation among these rows (columns) of A is also a linear relation among the rows (columns) of D.

```
*   *   *   *   *   *   *
*   *   *   *   *   *   *
*   *   *   *   *   *   *
*   *   *   *   *   *   *
*   *   *   *   *   *   *
*   *   *   *   *   *   *
*   *   *   *   *   *   *
*   *   *   *   *   *   *
*   *   *   *   *   *   *
```

**********88

Theorem. (setup) Suppose matrix A has m rows and n columns, so there are n column vectors $\mathbf{c_i}$ each of which has m components. We then create n reduced-size column vectors $\mathbf{C_i}$ by knocking out some set of components from all the $\mathbf{c_i}$. For example, we might remove rows 2, 3 and 5 of all the $\mathbf{c_i}$ to create the $\mathbf{C_i}$. Suppose the $\mathbf{C_i}$ then have J < m components. We then create a matrix "B" having n columns and J rows using these reduced $\mathbf{C_i}$ column vectors.

(theorem) If the original columns $\mathbf{c_i}$ are linearly *dependent* within the full mxn matrix A, then the derived $\mathbf{C_i}$ columns will be linearly *dependent* with this Jxn smaller matrix B.

$\{\mathbf{c_i}\}$ lin dep in A $\quad\Rightarrow\quad$ $\{\mathbf{C_i}\}$ lin dep in B

Proof: If the $\mathbf{c_i}$ are linearly dependent, we can find k_i such that $\Sigma_{i=1}^{n} k_i\mathbf{c_i} = 0$. This is a set of m equations, one for each component of the $\mathbf{c_i}$. If we take the subset of these equations which corresponds to the $\mathbf{C_i}$, then certainly $\Sigma_{i=1}^{n} k_i\mathbf{C_i} = 0$. Thus, the $\mathbf{C_i}$ are linearly dependent within their smaller matrix.

****

4. The Gradient Interpretation

Recall that when rank[R(**r**)] < S we can find Lagrange multipliers λ_i such that

$$\mathbf{f}(\mathbf{r}) + \lambda_1\mathbf{a}(\mathbf{r}) + \lambda_2\,\mathbf{b}(\mathbf{r}) + ... + \lambda_{S-1}\,\mathbf{q}(\mathbf{r}) = 0 \qquad (2.5)$$

where the bolded letters are the rows of the R matrix shown in (1.2). In derivative notation this equation reads

$$f_i(\mathbf{r}) + \lambda_1 a_i(\mathbf{r}) + \lambda_2\, b_i(\mathbf{r}) + ... + \lambda_{S-1}\, q_i(\mathbf{r}) = 0 \qquad i = 1,2...N \qquad (4.1)$$

where recall $f_i = \partial f/\partial x_i$. With the usual gradient operator ∇ this can be written

$$\nabla f(\mathbf{r}) + \lambda_1\nabla a(\mathbf{r}) + \lambda_2\,\nabla b(\mathbf{r}) + ... + \lambda_{S-1}\,\nabla q(\mathbf{r}) = 0 \qquad \nabla = (\partial_1,\partial_2...\partial_N) \qquad (4.2)$$

If there were no constraints, this equation would read

$$\nabla f(\mathbf{r}) = 0 \qquad \mathbf{r} = x_1, x_2.....x_N \qquad (4.3)$$

This equation says all the partial derivatives of f vanish at a point **r** which is a "critical point" for this very reason. The other terms in (4.2) show how this simple condition gets modified in the presence of constraints. We know that

$$df = \nabla f \bullet d\mathbf{r} \qquad d\mathbf{r} \text{ is a vector in N dimensional space}$$

If we are on a surface f(**r**) = constant (in N dimensional space) and we want to move a small distance d**r** while remaining on the surface, we need to have df = 0 for this small movement. That means d**r** is perpendicular to ∇f. Since this is true for any d**r** movement on the surface, we conclude that ∇f must be *normal* to the surface. Furthermore, if we move d**r** in the direction of this normal ∇f, we get a maximal |df|. Now if it happens that $\nabla f(\mathbf{r}) = 0$ for some particular point **r** on the surface f(**r**) = constant, then we find that df = 0 in all possible directions of movement.

Example: Assume N = 3 and we have f(x,y,z) = constant as a surface in E^3. We could write this z = F(x,y) if we wanted. I presume that ∇f is normal to this surface. Let's say the surface is a sphere. The normal to a sphere is never 0, so how can you talk about $\nabla f = 0$ at an extremum? Something is wrong here. I am off a dimension somewhere. Paradox!

****

****

For purposes of the theorem below, we shall use a *different notation* for the S-1 constraint functions:

$$\mathbf{g}(\mathbf{r}) \equiv \begin{pmatrix} g^{(1)}(\mathbf{r}) \\ g^{(2)}(\mathbf{r}) \\ \dots \\ g^{(S-1)}(\mathbf{r}) \end{pmatrix} \equiv \begin{pmatrix} a(\mathbf{r}) \\ b(\mathbf{r}) \\ \dots \\ q(\mathbf{r}) \end{pmatrix} \qquad \text{// column vector with S-1 elements}$$

The letter g is a traditional one for indicating constraint functions.

$$= f(\mathbf{r}) + \Sigma_{k=1}^{S-1} \lambda_k \, g^{(k)}(\mathbf{r})$$

$$= f(\mathbf{r}) + \boldsymbol{\lambda} \bullet \mathbf{g}(\mathbf{r})$$

$$= f_i + \Sigma_{k=1}^{S-1} \lambda_k \, [g^{(k)}]_i$$

$$\mathbf{f}(\mathbf{r}) + \lambda_1 \mathbf{a}(\mathbf{r}) + \lambda_2 \, \mathbf{b}(\mathbf{r}) + \dots\dots + \lambda_{S-1} \, \mathbf{q}(\mathbf{r}) = 0$$

$$\nabla f + \lambda_1 \nabla a + \dots \lambda_{S-1} \nabla q = 0$$

$$\nabla f + \lambda_1 \nabla g^{(1)} + \dots \lambda_{S-1} \nabla g^{(S-1)} = 0$$

Now use the vector sense indicated above to write

$$\nabla f + \lambda_1 \nabla g^{(1)} + \dots \lambda_{S-1} \nabla g^{(S-1)} = 0$$

or

$$\nabla f + \boldsymbol{\lambda} \bullet \nabla \mathbf{g}(\mathbf{r}) = 0$$

What exactly does this say! Web sites talk about it! Something to ponder a bit.

*********8