This was installed 10.22.16, do not edit here!

Section 6. Matrix proof of the Method of Lagrange Multipliers

In Section 2 it was shown that the Method of Lagrange Multipliers (Theorem 2) follows directly from Theorem 1, and really is just a restatement of Theorem 1 which we quote:

Theorem 1: A point **r** is a "stationary point for f(**r**) subject to constraints $a_i(\mathbf{r}) = 0$" $\Leftrightarrow$ *(1.24)*

1. $a_i(\mathbf{r}) = 0$ for i = 1,2...C (point **r** must satisfy all the constraints)
2. There exist C constants λ_i such that $\nabla f(\mathbf{r}) + \Sigma_{i=1}^{C} \lambda_i \nabla a_i(\mathbf{r}) = 0$.

Our proof of Theorem 1 was entirely geometric in nature. We talked about surfaces, normals to surfaces, tangent spaces, perp spaces, the intersection constraint surface, and so on. In this section, we shall rederive Theorem 1 using a certain "R matrix", though some geometry still appears. The reader should regard this simply as an alternative derivation of Theorem 1.

As in Section 1, a certain amount of background material is necessary to provide a framework for the proof. We define S = C+1 so the number of constraint functions is C = S-1, and we rename the constraint functions to be a,b,c....q instead of a_1, a_2, a_3....a_C. The purpose of this renaming is to avoid having double subscripts on a constraint function, one identifying it and one indicating a partial derivative.

6.1. The R matrix and its Rank

We shall operate in N dimensional Euclidean space E^N with coordinates $\mathbf{r} = (x_1, x_2 \ldots x_N)$.

We seek stationary points of a real function f(**r**) subject to S-1 constraints a(**r**) = 0, b(**r**) = 0, ... q(**r**) = 0 where we use q generically to represent the last constraint function. For example, if there are two constraints, then S = 3 and the two constraint equations are a(**r**) = 0 and b(**r**) = 0. So here is the problem:

find the stationary points of f(**r**) subject to constraints $\underset{1}{a(\mathbf{r})=0}, \underset{2}{b(\mathbf{r})=0}, \underset{3}{c(\mathbf{r})=0}, \ldots. \underset{S-1}{q(\mathbf{r})=0}$. (6.1.1)

Construct the following matrix of partial derivatives, where $f_i(\mathbf{r}) \equiv \partial_i f(\mathbf{r}) = \partial f(\mathbf{r})/\partial x_i$. That is, the subscript on f indicates which of the arguments $x_1, x_2, x_3 \ldots x_N$ the partial is with respect to. So here is that matrix, which we shall call R :

$$R = \begin{pmatrix} f_1 & f_2 & f_3 & \ldots & f_N \\ a_1 & a_2 & a_3 & \ldots & a_N \\ b_1 & b_2 & b_3 & \ldots & b_N \\ \ldots. & & & & \\ q_1 & q_2 & q_3 & \ldots & q_N \end{pmatrix} = \begin{pmatrix} \mathbf{f} \\ \mathbf{a} \\ \mathbf{b} \\ \ldots \\ \mathbf{q} \end{pmatrix} \qquad (6.1.2)$$

For example, $f_2 \equiv \partial f(x_1,x_2, x_3....x_N)/\partial x_2$.

On the far right we abbreviate each row of functions in **bold**, indicating a row vector like **f** .

The matrix has S rows and N columns. We shall only be interested in $N \geq S$ which is the same as saying $N \geq C+1$ or $N > C$. Recall from the discussion below (2.8) that if $N \leq C$, the problem is overconstrained and one cannot seek stationary points where df = 0. If N = S, the R matrix is square, otherwise for N > S it has more columns than rows, so it has a horizontal band shape.

The maximum possible **rank** of the above matrix is S since this is the smaller of the number of rows and columns (see below). S would then be the "full rank" of R.

Since all the functions are functions of **r**, we really have R(**r**). As the point **r** is varied, the elements of the matrix R(**r**) generally change.

If S = N, the matrix is square and we can talk about the determinant det(R). In this case there is only one SxS "submatrix" and it is the entire matrix R.

If S < N, the matrix is wider than it is tall. In this case, we can talk about square submatrices within R which are of shape SxS. For example, for N = 6 and S = 4 we have three constraints and the R matrix is this,

$$R = \begin{matrix} f_1 & f_2 & f_3 & f_4 & f_5 & f_6 \\ a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ b_1 & b_2 & b_3 & b_4 & b_5 & b_6 \\ c_1 & c_2 & c_3 & c_4 & c_5 & c_6 \end{matrix} \qquad (6.1.3)$$

Here we have outlined two 4x4 submatrices in red and blue. The columns of a submatrix need not be contiguous, so the three green boxes indicate another 4x4 submatrix. In this example, the number of submatrices is given by the binomial factor (6,4) -- pick a committee of 4 columns from 6 candidates. So we have shown only 3 out of (6,4) = 6!/(4!2!) = 15 possible 4x4 submatrices. We can indicate a submatrix using this notation:

$$\text{red} = (f_1,f_2,f_3,f_4) \qquad \text{blue} = (f_2,f_3,f_4,f_5) \qquad \text{green} = (f_1,f_2,f_4,f_6) \ . \qquad (6.1.4)$$

That is to say, we abbreviate the submatrix by stating only the values in the first row of the submatrix. Since our main interest will be whether or not the determinants of these 4x4 submatrices vanish, the ordering of the columns within a submatrix is not important.

For the general R matrix shown in (6.1.2) with N columns and S rows, there are (N,S) possible submatrices of size SxS. A particular submatrix is then indicated by a sequence of S subscripted f values.

We now quote a few facts from linear algebra concerning the rank of a matrix:

Definitions: A **minor** of matrix A is the determinant of any square submatrix of A obtained by crossing out some rows and columns The **rank** r of an m x n matrix A is the dimension of the largest non-vanishing minor within A. [Shilov 1.92] Thus, $r \leq \min(m,n)$. (6.1.5)

Definitions: [Shilov 2.21] Consider an m x n matrix A. Let $\mathbf{r}_1$, $\mathbf{r}_2$, ...$\mathbf{r}_N$ be an arbitrary selection of N rows (row vectors) from the matrix A. This set of N rows is **linearly dependent** if one can find a set of constants k_i at least one of which is non-zero such that

$$k_1 \mathbf{r}_1 + k_2 \mathbf{r}_2 + ... + k_N \mathbf{r}_N = 0 \ . \qquad (6.1.6)$$

There could be several such equations, or there could be only one. This generally means that at least one row can be expressed as a linear combination of the other rows. A legal such equation would be $\mathbf{r}_2 = 0$ in which case $k_2 = 1$. One could then say that $\mathbf{r}_2$ is a linear combination of the other rows with coefficients all zero. In any event, if one or more rows are all zeros, a set of N rows are linearly dependent. If no such equation (6.1.6) exists, then the set of N rows is **linearly independent**. For example, the set of vectors $\mathbf{r}_1$ = (1,0) and $\mathbf{r}_2$ = (0,1) is linearly independent.

The above discussion also applies to columns. Take "row" → "column" and $\mathbf{r}_i \rightarrow \mathbf{c}_i$.

Facts: The rank r as defined in (6.1.5) has these properties for any m x n matrix A: (6.1.7)

(a) the rank r equals the number of linearly independent columns of A [Shilov 3.12]

(b) the rank r equals the number of linearly independent rows of A [Shilov 3.13]

In Appendix A below we provide our own proof of the claims of (6.1.7).

With the above as background, the following theorem is fundamental to our matrix proof of Theorem 1 :

Theorem 3. Point **r** is a "stationary point of f(**r**) subject to the constraints a=0,b=0...q=0" ⇔

$$\text{rank}[R(\mathbf{r})] < S. \qquad (6.1.8)$$

But rank[R(**r**)] < S ⇔ $\nabla f(\mathbf{r}) + \lambda_1 \nabla a(\mathbf{r}) + \lambda_2 \nabla b(\mathbf{r}) + ... + \lambda_{S-1} \nabla q(\mathbf{r}) = 0$ for some λ_i because the rows of R are the gradients of f,a,b...q and having rank(R) < S means these rows are linearly dependent by (6.1.7). Thus proving Theorem 3 proves Theorem 1.

According to the rank definition (6.1.5),

$$\text{rank}(R) < S \quad \Leftrightarrow \quad \text{all SxS submatrices must have zero determinant} \qquad (6.1.9)$$

If, for example, the red submatrix in Fig 1.3 had a non-zero determinant, then rank(R) = 4 = S.

The point of Theorem 3 is that the solution point **r** of the constrained stationary point problem must be a point where the R(**r**) matrix drops below full rank. We shall prove Theorem 3 in the ⇒ direction in Section 6.2 by showing that, if **r** is a stationary point, then all those SxS submatrices discussed above have zero determinant and thus rank(R) < S. If one is searching for candidate solutions **r** to the stationary point problem, one can restrict one's search to points **r** for which rank[R(**r**)] < S.

The proof of Theorem 3 in the ⇐ direction is fairly simple. If rank(R) < S, then the rows of R are linearly dependent by (6.1.7) and then $\nabla f(\mathbf{r}) + \lambda_1 \nabla a(\mathbf{r}) + \lambda_2 \nabla b(\mathbf{r}) + ... + \lambda_{S-1} \nabla q(\mathbf{r}) = 0$. This implies that

$$df(\mathbf{r}) = \nabla f(\mathbf{r}) \bullet d\mathbf{r} \ = \ -\lambda_1[\nabla a(\mathbf{r}) \bullet d\mathbf{r}] \ -\lambda_2 \ [\nabla b(\mathbf{r}) \bullet d\mathbf{r}] - ..... - \lambda_{S-1} \ [\nabla q(\mathbf{r}) \bullet d\mathbf{r}] \qquad (6.1.10)$$

for any d**r**. Restricting to d**r** which don't violate any of the constraints, we know that each term on the right vanishes. For example, $\nabla a(\mathbf{r}) \bullet d\mathbf{r} = 0$ because $\nabla a(\mathbf{r})$ is normal to the surface $a(\mathbf{r}) = 0$, see (1.2). Since all the terms on the right above then vanish, we get $df(\mathbf{r}) = 0$ and **r** is a stationary point of f by our definition (1.18) of a constrained stationary point.

Comment on the R matrix. Imagine a general transformation from x-space $\mathbf{x} = (x_1, x_2, ...x_N)$ to u-space with coordinates $\mathbf{u} = (u_1, u_2, ... u_S)$. One might write such a transformation as $\mathbf{u} = \mathbf{F}(\mathbf{x})$ where $F:E^N \rightarrow E^S$. Although this transformation in general is non-linear, in a tiny neighborhood of point **x** in x-space (and the corresponding point **u** in u-space), the transformation will be linear if certain conditions are met. That local linear relation is then described by $d\mathbf{u} = R d\mathbf{x}$ where R is a matrix (N columns, S rows) whose matrix elements are $R_{ij} = \partial u_i/\partial x_j$. If we think of

$$
\begin{array}{llll}
f = u_1(x_1, x_2, ...x_N) & R_{11} = \partial u_1/\partial x_1 = \partial f/\partial x_1 = f_1 & R_{12} = f_2 & \text{etc.} \\
a = u_2(x_1, x_2, ...x_N) & R_{21} = \partial u_2/\partial x_1 = \partial a/\partial x_1 = a_1 & R_{22} = a_2 & \text{etc} \\
b = u_3(x_1, x_2, ...x_N) & & & \\
... & & & \\
q = u_S(x_1, x_2, ...x_N) & & & \qquad (6.1.11)
\end{array}
$$

then the matrix shown in (6.1.2) is exactly the R matrix for this general transformation $\mathbf{u} = \mathbf{F}(\mathbf{x})$. If S < N, then the R matrix is not square, the relation $d\mathbf{u} = R d\mathbf{x}$ is a "projection" of a larger space into a smaller one, and the equation therefore cannot be inverted. This notation "R" is used extensively in our Tensor Analysis document, see in particular Chapter 2, though only invertible transformations $F:E^N \rightarrow E^N$ are considered there. Sometimes R is referred to as "the differential" of a transformation.

6.2. Proof of Theorem 3 ⇒

Theorem 3. Point **r** is a "stationary point of f(**r**) subject to the constraints a=0,b=0...q=0" ⇔
rank[R(**r**)] < S. *(6.1.8)*

(a) Proof for N = 6 and S = 4

The more general the proof is made, the less clear it becomes due to the notational baggage that must be added. Therefore we shall prove Theorem 3 for the specific case shown above in Fig (6.1.3) where N = 6 and S = 4 (3 constraints). The reader should have no trouble following each step for this sample case. When we reach the proof conclusion for the sample case, we start over again doing the general case and compare each step to the steps of this sample case. Although a lot of equations appear below (and a lot of column inches are consumed), everything is really quite simple.

So, for our sample case we start off with this R matrix, again with bolded row vectors on the right,

$$
R = \begin{array}{cccccc} f_1 & f_2 & f_3 & f_4 & f_5 & f_6 \\ a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ b_1 & b_2 & b_3 & b_4 & b_5 & b_6 \\ c_1 & c_2 & c_3 & c_4 & c_5 & c_6 \end{array} = \begin{array}{c} \mathbf{f} \\ \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{array} \qquad (6.2.1)
$$

We want then to find an stationary point of $f(x_1, x_2, x_3, x_4, x_5, x_6)$ subject to these constraints:

$$\begin{aligned} a(x_1, x_2, x_3, x_4, x_5, x_6) &= 0 \\ b(x_1, x_2, x_3, x_4, x_5, x_6) &= 0 \\ c(x_1, x_2, x_3, x_4, x_5, x_6) &= 0 \,. \end{aligned} \tag{6.2.2}$$

In practice, a person solving this problem might try to eliminate variables. For example, somehow with enough labor one should be able to "use up" the three constraint equations to eliminate three of the six variables x_i. There are of course (6,3) ways to select three variables for elimination.

For starters, imagine that we have used up the 3 constraints to eliminate variables x_1, x_2 and x_3. Then we write

$$\begin{aligned} a(x_1, x_2, x_3, x_4, x_5, x_6) &= 0 \qquad\qquad x_1 = X^1(x_4, x_5, x_6) \\ b(x_1, x_2, x_3, x_4, x_5, x_6) &= 0 \quad \Rightarrow \quad x_2 = X^2(x_4, x_5, x_6) \\ c(x_1, x_2, x_3, x_4, x_5, x_6) &= 0 \qquad\qquad x_3 = X^3(x_4, x_5, x_6) \end{aligned} \tag{6.2.3}$$

where X^1, X^2 and X^3 are three resulting functions. This is a "theoretical" elimination, one does not actually have to do the process, one need only understand that in principle it could be done and the functions X^i could in principle be found. See Appendix C for an elaboration of this point.

Now rewrite the function f and the three constraint functions in this manner, substituting for example the function X^1 for x_1 everywhere it appears,

$$
\begin{aligned}
&f(X^1(x_4, x_5, x_6), X^2(x_4, x_5, x_6), X^3(x_4, x_5, x_6), x_4, x_5, x_6) && \equiv F(x_4, x_5, x_6) \\
&a(X^1(x_4, x_5, x_6), X^2(x_4, x_5, x_6), X^3(x_4, x_5, x_6), x_4, x_5, x_6) && \equiv A(x_4, x_5, x_6) \\
&b(X^1(x_4, x_5, x_6), X^2(x_4, x_5, x_6), X^3(x_4, x_5, x_6), x_4, x_5, x_6) && \equiv B(x_4, x_5, x_6) \\
&c(X^1(x_4, x_5, x_6), X^2(x_4, x_5, x_6), X^3(x_4, x_5, x_6), x_4, x_5, x_6) && \equiv C(x_4, x_5, x_6)\ . \qquad (6.2.4)
\end{aligned}
$$

In this way we have defined four new functions F,A,B,C each of the 3 variables shown (those that were not eliminated).

Next, compute the first partial derivatives of F using the chain rule. For example

$$
\frac{\partial F}{\partial x_4} = \frac{\partial f}{\partial x_1}\frac{\partial X^1}{\partial x_4} + \frac{\partial f}{\partial x_2}\frac{\partial X^2}{\partial x_4} + \frac{\partial f}{\partial x_3}\frac{\partial X^3}{\partial x_4} + \frac{\partial f}{\partial x_4}\ . \qquad (6.2.5a)
$$

In our compact notation where F_i means $\partial F/\partial x_i$ this can be restated as

$$
F_4 = f_1 X^1{}_4 + f_2 X^2{}_4 + f_3 X^3{}_4 + f_4\ . \qquad (6.2.5b)
$$

Doing this for each of the non-eliminated variables one gets,

$$
\begin{aligned}
F_4 &= f_1 X^1{}_4 + f_2 X^2{}_4 + f_3 X^3{}_4 + f_4 \\
F_5 &= f_1 X^1{}_5 + f_2 X^2{}_5 + f_3 X^3{}_5 + f_5 \\
F_6 &= f_1 X^1{}_6 + f_2 X^2{}_6 + f_3 X^3{}_6 + f_6\ . \qquad (6.2.6)
\end{aligned}
$$

Notice that the subscripts on the leftmost three f_i correspond to those of the eliminated variables i = 1,2,3.

Now, since we are seeking an extremal point of $F(x_4, x_5, x_6)$ with no constraints at all (the three initial constraints have all been absorbed in the process of eliminating three variables), we require that (x_4, x_5, x_6) be a "critical point" for the function F, which means we require that $F_4 = F_5 = F_6 = 0$. Then the above equations become,

$$
\begin{aligned}
f_1 X^1{}_4 + f_2 X^2{}_4 + f_3 X^3{}_4 + f_4 &= 0 \\
f_1 X^1{}_5 + f_2 X^2{}_5 + f_3 X^3{}_5 + f_5 &= 0 \\
f_1 X^1{}_6 + f_2 X^2{}_6 + f_3 X^3{}_6 + f_6 &= 0\ . \qquad (6.2.7)
\end{aligned}
$$

Now apply the *same process* to the three constraint functions A,B,C. First compute the derivatives, and then set the derivatives to zero. One does this for $A(x_4, x_5, x_6)$, for example, because the constraint condition a = 0 says that that $A(x_4, x_5, x_6) = 0$ = a constant, for all values of x_4, x_5, x_6 , so all first derivatives must vanish. Doing this for function A then gives these three equations,

$$
\begin{aligned}
a_1 X^1{}_4 + a_2 X^2{}_4 + a_3 X^3{}_4 + a_4 &= 0 \\
a_1 X^1{}_5 + a_2 X^2{}_5 + a_3 X^3{}_5 + a_5 &= 0 \\
a_1 X^1{}_6 + a_2 X^2{}_6 + a_3 X^3{}_6 + a_6 &= 0\ . \qquad (6.2.8)
\end{aligned}
$$

We have just taken the previous equations and replaced f→ a everywhere. Then do this for B and C. One ends up then with this set of 12 equations:

$$
\begin{aligned}
&f_1X^1_4 + f_2X^2_4 + f_3X^3_4 + f_4 = 0\\
&f_1X^1_5 + f_2X^2_5 + f_3X^3_5 + f_5 = 0\\
&f_1X^1_6 + f_2X^2_6 + f_3X^3_6 + f_6 = 0\\
&\\
&a_1X^1_4 + a_2X^2_4 + a_3X^3_4 + a_4 = 0\\
&a_1X^1_5 + a_2X^2_5 + a_3X^3_5 + a_5 = 0\\
&a_1X^1_6 + a_2X^2_6 + a_3X^3_6 + a_6 = 0\\
&\\
&b_1X^1_4 + b_2X^2_4 + b_3X^3_4 + b_4 = 0\\
&b_1X^1_5 + b_2X^2_5 + b_3X^3_5 + b_5 = 0\\
&b_1X^1_6 + b_2X^2_6 + b_3X^3_6 + b_6 = 0\\
&\\
&c_1X^1_4 + c_2X^2_4 + c_3X^3_4 + c_4 = 0\\
&c_1X^1_5 + c_2X^2_5 + c_3X^3_5 + c_5 = 0\\
&c_1X^1_6 + c_2X^2_6 + c_3X^3_6 + c_6 = 0\ .
\end{aligned}
\qquad (6.2.9)
$$

Here there are S = 4 equation groups, and each group has N-S+1 = 6-4+1 = 3 equations (this last 3 is the number of non-eliminated variables).

Next *reorder* these equations into three groups of four, for example taking the first equation in each group above to make the first new group below,

$$
\begin{aligned}
&f_1X^1_4 + f_2X^2_4 + f_3X^3_4 + f_4 = 0\\
&a_1X^1_4 + a_2X^2_4 + a_3X^3_4 + a_4 = 0\\
&b_1X^1_4 + b_2X^2_4 + b_3X^3_4 + b_4 = 0\\
&c_1X^1_4 + c_2X^2_4 + c_3X^3_4 + c_4 = 0\\
&\\
&f_1X^1_5 + f_2X^2_5 + f_3X^3_5 + f_5 = 0\\
&a_1X^1_5 + a_2X^2_5 + a_3X^3_5 + a_5 = 0\\
&b_1X^1_5 + b_2X^2_5 + b_3X^3_5 + b_5 = 0\\
&c_1X^1_5 + c_2X^2_5 + c_3X^3_5 + c_5 = 0\\
&\\
&f_1X^1_6 + f_2X^2_6 + f_3X^3_6 + f_6 = 0\\
&a_1X^1_6 + a_2X^2_6 + a_3X^3_6 + a_6 = 0\\
&b_1X^1_6 + b_2X^2_6 + b_3X^3_6 + b_6 = 0\\
&c_1X^1_6 + c_2X^2_6 + c_3X^3_6 + c_6 = 0\quad .
\end{aligned}
\qquad (6.2.10)
$$

Now there are N-S+1 = 3 equation groups, and each group has S = 4 equations.

Next, write each of these three equation sets as a matrix equation,

$$\begin{pmatrix}0\\0\\0\\0\end{pmatrix} = \begin{pmatrix}f_1 & f_2 & f_3 & f_4\\ a_1 & a_2 & a_3 & a_4\\ b_1 & b_2 & b_3 & b_4\\ c_1 & c_2 & c_3 & c_4\end{pmatrix}\begin{pmatrix}X^1{}_4\\ X^2{}_4\\ X^3{}_4\\ 1\end{pmatrix} = (f_1, f_2, f_3, f_4)\begin{pmatrix}X^1{}_4\\ X^2{}_4\\ X^3{}_4\\ 1\end{pmatrix}$$

$$\begin{pmatrix}0\\0\\0\\0\end{pmatrix} = \begin{pmatrix}f_1 & f_2 & f_3 & f_5\\ a_1 & a_2 & a_3 & a_5\\ b_1 & b_2 & b_3 & b_5\\ c_1 & c_2 & c_3 & c_5\end{pmatrix}\begin{pmatrix}X^1{}_5\\ X^2{}_5\\ X^3{}_5\\ 1\end{pmatrix} = (f_1, f_2, f_3, f_5)\begin{pmatrix}X^1{}_5\\ X^2{}_5\\ X^3{}_5\\ 1\end{pmatrix}$$

$$\begin{pmatrix}0\\0\\0\\0\end{pmatrix} = \begin{pmatrix}f_1 & f_2 & f_3 & f_6\\ a_1 & a_2 & a_3 & a_6\\ b_1 & b_2 & b_3 & b_6\\ c_1 & c_2 & c_3 & c_6\end{pmatrix}\begin{pmatrix}X^1{}_6\\ X^2{}_6\\ X^3{}_6\\ 1\end{pmatrix} = (f_1, f_2, f_3, f_6)\begin{pmatrix}X^1{}_6\\ X^2{}_6\\ X^3{}_6\\ 1\end{pmatrix} \quad . \qquad (6.2.11)$$

On the right we show our abbreviated notation for a matrix, just showing the first row.

Now we claim that the three matrices shown must have zero determinant! Suppose the first matrix equation had a non-zero determinant. It could then be inverted (see (B.4.8)) to give

$$\begin{pmatrix}X^1{}_4\\ X^2{}_4\\ X^3{}_4\\ 1\end{pmatrix} = (f_1, f_2, f_3, f_4)^{-1}\begin{pmatrix}0\\0\\0\\0\end{pmatrix} = \begin{pmatrix}0\\0\\0\\0\end{pmatrix} \quad . \qquad (6.2.12)$$

But this produces a blatant contradiction that 1 = 0 (not to mention the other rows), and therefore we must have

$$\det(f_1, f_2, f_3, f_4) = 0$$
$$\det(f_1, f_2, f_3, f_5) = 0$$
$$\det(f_1, f_2, f_3, f_6) = 0$$

or

$$\det(f_1, f_2, f_3, f_m) = 0 \qquad m \neq 1,2,3 \quad . \qquad (6.2.13)$$

Therefore, we have shown that three of the SxS = 4x4 submatrices of Fig (1.3) vanish. But there are (6,4) = 15 such submatrices, and in order to show that rank(R) < S, we have to show that *all* 15 SxS determinants vanish.

But that can be demonstrated by starting over several times and each time choosing to eliminate three different variables. Notice in the det = 0 equations above that the first three f_i values correspond to the three eliminated variables x_1, x_2 and x_3, Had we instead chosen to eliminate x_1, x_2 and x_4, we would have obtained,

$$\det(f_1, f_2, f_4, f_3) = 0$$
$$\det(f_1, f_2, f_4, f_5) = 0$$
$$\det(f_1, f_2, f_4, f_6) = 0$$

or

$$\det(f_1, f_2, f_4, f_m) = 0 \qquad m \neq 1,2,4 \ . \tag{6.2.14}$$

More generally, had we chosen to eliminate variables x_i, x_j and x_k ($i \neq j \neq k$), we would have obtained

$$\det(f_i, f_j, f_k, f_m) = 0 \qquad m \neq i,j,k \ . \tag{6.2.15}$$

As i,j,k range over all possible values 1,2,3,4,5,6, we clearly hit all possible 4x4 subdeterminants, and thus we have shown that they all vanish, and therefore rank(R) < S, QED.

Lest it be overlooked, one must keep in mind that $\mathbf{r} = (x_1, x_2, x_3, x_4, x_5, x_6)$ must be a stationary point of f(**r**) subject to the constraints. And then the conclusion is that, for such a solution point **r**, one has rank[R(**r**)] < S.

To summarize, one theoretically eliminates S-1 of the N variables to create the functions F,A,B,C of (6.2.4). One then looks for the normal (unconstrained) critical points of F where all partials vanish. Since the constraints all have the form A = 0, B = 0, C = 0, their partials vanish as well. Writing all these equations in matrix form (for various sets of eliminated variables) leads to the conclusion that all the SxS minors of matrix R vanish so that rank(R) < S.

(b) Proof for general S ≤ N

To generalize now to general N and S, we restate some of the previous equations (adding a prime to the equation number) using a more generalized notation. The reader will note how even our simple notation becomes rather unpleasant.

We start off with this R matrix which has N columns and S ≤ N rows. The S-1 constraint functions are a,b,....q. Remember that f_2 means $\partial f/\partial x_2$.

$$R = \begin{matrix} f_1 & f_2 & f_3 & \dots & f_N \\ a_1 & a_2 & a_3 & \dots & a_N \\ b_1 & b_2 & b_3 & \dots & b_N \\ .... & & & & \\ q_1 & q_2 & q_3 & \dots & q_N \end{matrix} = \begin{matrix} \mathbf{f} \\ \mathbf{a} \\ \mathbf{b} \\ ... \\ \mathbf{q} \end{matrix} \tag{6.2.1}'$$

Here are the S-1 constraint equations,

$$a(x_1, x_2, \dots x_N) = 0 \qquad \text{// S-1 constraint equations}$$
$$b(x_1, x_2, \dots x_N) = 0$$
$$\dots$$
$$q(x_1, x_2, \dots x_N) = 0 \ . \tag{6.2.2}'$$

Eliminate (theoretically) the first S-1 variables $x_1, x_2 \ldots x_{S-1}$ using the S-1 constraint equations:

$$\begin{aligned} a(x_1, x_2, \ldots x_N) &= 0 && x_1 = X^1(x_S, x_{S+1} \ldots x_N) \\ b(x_1, x_2, \ldots x_N) &= 0 \quad \Rightarrow && x_2 = X^2(x_S, x_{S+1} \ldots x_N) \\ \ldots & && \ldots \\ q(x_1, x_2, \ldots x_N) &= 0 && x_{S-1} = X^{S-1}(x_S, x_{S+1} \ldots x_N)\ . \end{aligned} \qquad (6.2.3)'$$

Substitute to get the F,A,B...Q equations (a set of S equations),

$$\begin{aligned} &f(X^1(x_S, x_{S+1} \ldots x_N), X^2(x_S, x_{S+1} \ldots x_N), \ldots X^{S-1}(x_S, x_{S+1} \ldots x_N), x_S, x_{S+1} \ldots x_N) \equiv F(x_S, x_{S+1} \ldots x_N) \\ &a(X^1(x_S, x_{S+1} \ldots x_N), X^2(x_S, x_{S+1} \ldots x_N), \ldots X^{S-1}(x_S, x_{S+1} \ldots x_N), x_S, x_{S+1} \ldots x_N) \equiv A(x_S, x_{S+1} \ldots x_N) \\ &b(X^1(x_S, x_{S+1} \ldots x_N), X^2(x_S, x_{S+1} \ldots x_N), \ldots X^{S-1}(x_S, x_{S+1} \ldots x_N), x_S, x_{S+1} \ldots x_N) \equiv B(x_S, x_{S+1} \ldots x_N) \\ &\ldots \\ &q(X^1(x_S, x_{S+1} \ldots x_N), X^2(x_S, x_{S+1} \ldots x_N), \ldots X^{S-1}(x_S, x_{S+1} \ldots x_N), x_S, x_{S+1} \ldots x_N) \equiv Q(x_S, x_{S+1} \ldots x_N)\ . \end{aligned}$$

(6.2.4)'

Now take first derivatives as shown in (6.2.5) to get this generalized version of (6.2.6),

$$\begin{aligned} F_S &= f_1 X^1{}_S + f_2 X^2{}_S + f_3 X^3{}_S + \ldots + f_{S-1} X^{S-1}{}_S + f_S \\ F_{S+1} &= f_1 X^1{}_{S+1} + f_2 X^2{}_{S+1} + f_3 X^3{}_{S+1} + \ldots + f_{S-1} X^{S-1}{}_{S+1} + f_{S+1} \\ F_{S+2} &= f_1 X^1{}_{S+2} + f_2 X^2{}_{S+2} + f_3 X^3{}_{S+2} + \ldots + f_{S-1} X^{S-1}{}_{S+2} + f_{S+2} \\ &\ldots \\ F_N &= f_1 X^1{}_N + f_2 X^2{}_N + f_3 X^3{}_N + \ldots + f_{S-1} X^{S-1}{}_N + f_N\ . \end{aligned} \qquad \text{// N-S+1 equations} \qquad (6.2.6)'$$

Requiring $(x_S, x_{S+1} \ldots x_N)$ to be a "critical point", we set all the these first derivatives to 0 to get,

$$
\begin{aligned}
&f_1X^1{}_S + f_2X^2{}_S + f_3X^3{}_S + \ldots + f_{S-1}X^{S-1}{}_S + f_S = 0\\
&f_1X^1{}_{S+1} + f_2X^2{}_{S+1} + f_3X^3{}_{S+1} + \ldots + f_{S-1}X^{S-1}{}_{S+1} + f_{S+1} = 0\\
&f_1X^1{}_{S+2} + f_2X^2{}_{S+2} + f_3X^3{}_{S+2} + \ldots + f_{S-1}X^{S-1}{}_{S+2} + f_{S+2} = 0\\
&\ldots\\
&f_1X^1{}_N + f_2X^2{}_N + f_3X^3{}_N + \ldots + f_{S-1}X^{S-1}{}_N + f_N = 0\ .\quad \text{// N-S+1 equations} \qquad (6.2.7)'
\end{aligned}
$$

A similar set of equations is obtained for a,b,c...q. Just replace f→a, then f→b and so on. We shall not write out all these equations as we did in (6.2.9). We have then S sets of equations, each set containing N-S+1 equations. As in the sample case, we then reorder the equations to get first this group,

$$
\begin{aligned}
&f_1X^1{}_S + f_2X^2{}_S + f_3X^3{}_S + \ldots + f_{S-1}X^{S-1}{}_S + f_S = 0\\
&a_1X^1{}_S + a_2X^2{}_S + a_3X^3{}_S + \ldots + a_{S-1}X^{S-1}{}_S + a_S = 0\\
&b_1X^1{}_S + b_2X^2{}_S + b_3X^3{}_S + \ldots + b_{S-1}X^{S-1}{}_S + b_S = 0\\
&\ldots.\\
&q_1X^1{}_S + q_2X^2{}_S + q_3X^3{}_S + \ldots + q_{S-1}X^{S-1}{}_S + q_S = 0\ .\qquad \text{// S equations (6.2.10)'}_S
\end{aligned}
$$

The next group has S→S+1 in all subscript positions. The last group has S→N, and here is that last group,

$$
\begin{aligned}
&f_1X^1{}_N + f_2X^2{}_N + f_3X^3{}_N + \ldots + f_{S-1}X^{S-1}{}_N + f_N = 0\\
&a_1X^1{}_N + a_2X^2{}_N + a_3X^3{}_N + \ldots + a_{S-1}X^{S-1}{}_N + a_N = 0\\
&b_1X^1{}_N + b_2X^2{}_N + b_3X^3{}_N + \ldots + b_{S-1}X^{S-1}{}_N + b_N = 0\\
&\ldots.\\
&q_1X^1{}_N + q_2X^2{}_N + q_3X^3{}_N + \ldots + q_{S-1}X^{S-1}{}_N + q_N = 0\ .\qquad \text{// S equations (6.2.10)'}_N
\end{aligned}
$$

So there are now N-S+1 groups of equations and each group has S equations each having S terms.

The next task is to write each set of S equations as a matrix equation. Here we do that just using the abbreviated notation for the matrix. There are then N-S+1 matrix equations,

$$
\begin{pmatrix} 0\\0\\\ldots\\0\\0 \end{pmatrix} = (f_1, f_2, f_3, \ldots f_{S-1}, f_S) \begin{pmatrix} X^1{}_S\\X^2{}_S\\\ldots\\X^{S-1}{}_S\\1 \end{pmatrix}
$$

$$
\begin{pmatrix} 0\\0\\\ldots\\0\\0 \end{pmatrix} = (f_1, f_2, f_3, \ldots f_{S-1}, f_{S+1}) \begin{pmatrix} X^1{}_{S+1}\\X^2{}_{S+1}\\\ldots\\X^{S-1}{}_{S+1}\\1 \end{pmatrix}
$$

●●●●

$$
\begin{pmatrix} 0\\0\\\ldots\\0\\0 \end{pmatrix} = (f_1, f_2, f_3, \ldots f_{S-1}, f_N) \begin{pmatrix} X^1{}_N\\X^2{}_N\\\ldots\\X^{S-1}{}_N\\1 \end{pmatrix}\ . \qquad (6.2.11)'
$$

We then argue as before that, to avoid the contradiction 0 = 1, the determinants of these SxS submatrices must be zero,

$$\det(f_1, f_2, f_3, \ldots f_{S-1}, f_S) = 0$$
$$\det(f_1, f_2, f_3, \ldots f_{S-1}, f_{S+1}) = 0$$
$$\ldots$$
$$\det(f_1, f_2, f_3, \ldots f_{S-1}, f_N) = 0$$

or

$$\det(f_1, f_2, f_3, \ldots f_{S-1}, f_m) = 0 \qquad m \neq 1,2,3\ldots S-1 \quad . \qquad (6.2.13)'$$

Finally, if we start off by eliminating the S-1 variables x_i, x_j, x_k.... x_r where $i \neq j \neq k \ldots \neq r$, we would find that

$$\det(f_i, f_j, f_k, \ldots f_r, f_m) = 0 \qquad m \neq i,j,k\ldots r \quad . \qquad (6.2.15)'$$

Since this includes any possible SxS subdeterminant of the R matrix, we have then shown that all (N,S) SxS subdeterminants vanish, and therefore rank[R(**r**)] < S when **r** is a stationary point of f(**r**) subject to the constraints.