ANALYSIS OF MULTI-PARTICLE INTERACTIONS USING METRIC CONSTRUCTIONS IN PHASE SPACE

R.W. Hartung

CERN, Geneva, Switzerland

June 1968

(To be submitted to the
Journal of Computational Physics)

ANALYSIS IN PHASE SPACE

Richard W. Hartung

NP Division

CERN

1211 Genève 23

Switzerland

ABSTRACT

Simple geometric measures (lengths of lines, areas of triangles, etc., in an appropriate phase space) are proposed as tools in analysing multiparticle interactions. Several applications are proposed, including searches for maxima and minima in phase-space density, tests for reflection invariance and other symmetries, and experimental determination of the dimensionality of phase-space distributions.

I. INTRODUCTION

The experimental analysis of particle interactions (as observed, for example, in a bubble chamber) has traditionally been done using either a one-dimensional graph or histogram (such as cross section as a function of beam momentum or number of events as a function of some invariant mass) or the two-dimensional scatter diagram (of which the best examples are the Dalitz-Fabry plot and the Chew-Low plot) [1]. While these tools are quite adequate for simple processes such as elastic scattering where only four particles are involved, and while they have been remarkably useful in studying reactions of higher multiplicity, they are certainly in no sense adequate for a thorough analysis of the more complicated reactions.

The reason for this is quite simple: for a reaction involving n particles, neglecting the external polarizations, it requires $3n - 10$ scalar invariants to completely specify an event. Thus the pertinent phase space is a space of $3n - 10$ dimensions, and only if $n \leq 4$ can this be mapped into a one- or two-dimensional plot. (From time to time, people have tried to construct 3-dimensional plots, but the value of these as a real analytical tool is questionable.) Thus we are reduced to studying various projections of the distribution of events in a multi-dimensional space, perhaps comparing them with the projected phase-space volume. In deference to the true complexity of the problem it has become customary to present graphs subject to "cuts". That is, distributions are plotted for one or two variables but including only those events which satisfy certain inequalities with respect to one or more other variables. (By far the most common "cut", of course, is the fixing of the beam momentum, which reduces the number of dimensions in the phase space by one, if the target is stationary.)

While these techniques are not without merit, they are certainly crude. This will inevitably become apparent to all on the day when, if indeed it has not already occurred, a graduate thesis is submitted in which the number of graphs exceeds the number of events under consideration.

I present here another approach to the problem of how to achieve human comprehension of experimental data which unfortunately is available

in a space of many dimensions. The method is by no means a final solution; it may, however, be of some use in conjunction with traditional techniques. In the next section the measures to be used are defined and explained. Section III contains some proposed uses of these measures. The algebra needed for the least-squares solution to the length of a line appears in an Appendix.

II. GEOMETRIC MEASURES

It is useful to recall why $3n - 10$ quantities are required to specify an interaction of n particles of fixed masses. There are

$3n$ external momenta (with fixed mass, only 3 components of the 4 momentum for each particle are independent).

-4 equations of energy-momentum conservation

-3 parameters of an arbitrary Lorentz transformation

-3 parameters of an arbitrary rotation

$3n-10$ net independent quantities.

Thus we need to work in a phase space of $3n - 10$ dimensions.

Now suppose that we have experimental data for a certain reaction, say

$$K^- P \rightarrow \bar{K}^0 N \pi^+ \pi^- \qquad (1)$$

(for which $n = 6$). That is, there is a set of m events, for each of which the four vectors of each of the 6 particles are known. The four-vectors are expressed in the standard combination [2]:

$$p_\mu = (p_1, p_2, p_3, p_4) = (\underline{p}, iE) \qquad (2)$$

with both the momentum $\underline{p}$ and the energy E expressed in mass units (e.g. GeV). Then an "event" consists of a vector $\underline{V}$ of $4n = 24$ components, consisting of a string of the n particle vectors in a definite order. Consider two events, $\underline{V}_A$, and $\underline{V}_B$; let the difference vector be

$$\underline{V}_{AB} = \underline{V}_A - \underline{V}_B \qquad (3)$$

Since A and B are examples of reaction (1) their mass relations and the four conservation equations are certainly satisfied by $\underline{V}_A$ and by $\underline{V}_B$. Therefore the vectors $\underline{V}_A$, $\underline{V}_B$ and $\underline{V}_{AB}^-$ all exist on hypersurfaces of $3n-4$ dimensions embedded in the $4n$ dimensional Cartesian space. Holding event A fixed, consider possible rotations (i.e. space rotation and Lorentz translations) of event B and define

$$L^2(A,B) = \min_B \left[(\underline{V}_{AB}^-)^2 \right] \tag{4}$$

where the symbol $\min_B[\ ]$ denotes a minimization with respect to all possible rotations of event B (but see a generalization of this definition below). Note that $L^2(A,B)$ has the following properties:

1) It is the squared length of a vector in a phase space of $3n-10$ dimensions ($3n-4$ reduced by a 6 parameter minimization).

2) It is non-negative, since it is the sum of squared lengths of vectors which must be space like. Thus the minimum must exist; it is in fact unique (see Appendix).

3) $L^2(A,A) = 0$. Furthermore $L^2(A,A') = 0$, where A' consists of event A after any sequence of rotations and Lorentz transformations.

4) Thus $L^2(A,B) = 0$ only if each invariant of event A is identical to the corresponding invariant of event B. The converse is not true; see for example the parity conservation test in Section III.

5) $\bar{L}^2(\bar{A},\bar{B}) = \bar{L}^2(\bar{B},\bar{A})$.

Note that properties 2 through 4 hold only so long as corresponding particles in A and B have the same mass. It is forbidden, for present purposes, to lump, say, the π^+ and π^- together and call that a "particle" (unless, of course the "dipions" in ℓ events A and B have identical masses).

One further complication: the reaction (1) does not suffer from any ambiguity problems. Consider by contrast the reaction

$$PP \to PP\pi^+\pi^+\pi^-\pi^- \ . \tag{5}$$

Here each of the particles is indistinguishable from one other particle. Thus there are $2^4 = 16$ possible pairings that could be used in calculating $L^2(A,B)$. The properties derived above for $L^2(A,B)$ remain true only if we extend the definition of min[ ] to include selecting the smallest value with respect to possible permutations of indistinguishable particles.

In like manner for three events A, B, and C we can define the area of a triangle in phase space

$$T(A,B,C) = \min_B \left[\min_C [\sqrt{(\underline{V}_{AB}^-)^2 \ (\underline{V}_{AC}^-)^2} - \underline{V}_{AB}^- \cdot \underline{V}_{AC}^-] \right], \qquad (6)$$

the volume of a pyramid and so on. Note that to achieve the required minima with a geometric construction of ℓ dimensions, the number of pairing permutations required is expenential in $\ell - 1$. Thus the computational problems can be severe.

Finally there is a combinatorial problem. If m events exist, then there are $C_\ell^m = [m!/\ell!(m-\ell)!]$ independent ways in which to combine ℓ events to calculate a measure of dimension $\ell - 1$. Again the computational problems can be severe. Therefore in most of the applications only $\ell = 2$ will be considered.

III. APPLICATIONS

Most of the applications will involve a histogram of the number, N, of pairs (or triplets, etc.) versus the phase-space measure L^2 (or T, etc), on which it is assumed that all the C_ℓ^m combinations are plotted.

1) Detection of duplicate events. This is a trivial but remarkably effective application. Most analysis systems are susceptible to human errors (or chicanery) which might cause the same event to appear twice in the final set of "good" events. On a N,L^2 histogram this would lead to a distinct peak at $L^2 = 0$ (or nearby if it were different but similar measurements of the same event).

2) Detecting peaks in matrix elements. If the squared matrix element leading to the reaction under study is not constant throughout the phase-space volume, then it must have maxima and minima. It may be of great interest to know where these are. At the peaks of the squared matrix element the events will be clustered more closely together. Therefore, with

sufficient statistics, these peaks correspond to those events for which small values of, e.g., L^2 occur. More elaborately one could calculate $L^2(Z,A_i)$, i = 1, ..., m where Z is a test event which is allowed to sweep through phase space on a search pattern or extremum procedure to search for a maximum (or minimum) in event density.

3) Determining dimensionality of distributions. Normally the events would be distributed throughout the 3n - 10 (or 3n - 11 if beam and target momenta are fixed) dimensional phase space. There may, however, be a clustering of events into a space of fewer dimensions, either because the matrix element has a pole in the physical region, causing an abundance of events with values of a certain invariant (or invariants) in the pole region, or because of a pole outside the physical region, causing a condensation at the boundary of phase space. There are two ways to search for such a situation. Most obviously, the hypervolume of the elementary construction of dimension ℓ will be 0 if the ℓ events for which it is calculated lie in a rectangular hypervolume of dimension $< \ell$. Therefore it is useful to ask whether there is a significant number of combinations of ℓ = 3n - 11 (or 12) events giving a hypervolume near 0.

Alternatively the test can be made using only the two-dimensional measure L^2. Consider a N,L^2 histogram in which both axes are logarithmic. If the events were distributed uniformly in an unbounded phase space of ℓ dimensions, then the average distribution of the number $\int_0^L N(L)dL$ of events within a distance L, of any point would be

$$\int N(L)dL \propto L^{\ell} \qquad (7)$$

Thus the log N, log L^2 plot would be a straight line of slope $\ell/2 - 1$. The physical region of phase space is, of course, finite, and arbitrarily large values of L^2 are not possible, so the distribution must eventually curve back down from the straight line; nevertheless the initial slope (near $L^2 = 0$) should be $\ell/2 - 1$. By noting the initial slope on a log N, log L^2 plot one can get a direct measure of the dimensionality of the most significant clustering of events.

4) Experimental tests of symmetry principles. The phase-space metric construction can be used to make model-independent tests of various symmetries. A few examples follow.

Suppose that we wish to decide whether a given set of events is produced by a matrix element which is invariant under reflection of one momentum coordinate (parity conserving). We have merely to compare the distributions of

$$L^2(A_i,A_j) \text{ and } L^2(A_i,RA_j) \quad (i = 1, \ldots, m, \; j = 1, \ldots, i-1)$$

where RA_j is the event A_j with the sign of P_Z, for example, reversed. If the reaction is reflection-invariant, then these two distributions should differ only in a statistical way, because the event density at the point A_j will be on the average the same as at the point RA_j. But if the interaction does not conserve parity, then there must be a region in phase space where the density of real events A_i is higher than the density of reflected events RA_i. Then the real pairings $L^2(A_i,A_j)$ will tend to have more cases of small L^2 than the crossed pairings. Any significant shift of the distribution to larger values of L^2 is thus evidence of lack of reflection invariance.

Conversely if one wishes to assume (or even force) reflection invariance, the number of events used in other phases of this analysis can be doubled merely by using the set of 2 m events: A_i, RA_i. This remark applies also to other symmetry tests.

Some reactions might be suitable for a test of charge conjugation invariance. Consider the reaction

$$\bar{P}P \to \bar{P}P\pi^+\pi^-\pi^0 \qquad (8)$$

for example. Charge conjugation could be tested by comparing the distributions

$$L^2(A_i,A_j) \text{ and } L^2(A_i,CA_j)$$

where CA_j is the event A_j after the interchanges $\bar{P} \leftrightarrow P$ (twice) and $\pi^+ \leftrightarrow \pi^-$. The remarks in the parity discussion all apply.

The reader may be amused to devise tests for other symmetries. I mention only one more, an example of a "two experiment" type. Suppose there are M events of the reaction

$$K^+P \to K^+P\pi^+\pi^- \tag{9}$$

and M′ events of the reaction

$$K^-P \to K^-P\pi^+\pi^- \ . \tag{10}$$

To try to separate the parts of the reaction which do and do not depend on the strangeness quantum number it might be useful to look at all three distributions

$$L^2(A_i,A_j) \quad L^2(A_i,A'_j) \text{ and } L^2(A'_i,A'_j)$$

where A_i is an event of reaction (9) and A'_j is one of reaction (10).

5) Tests of complete theoretical models. The ultimate goal in studying particle interactions is to develop a theoretical model which describes the experimental situation. Frequently these models are only "partial models" in the sense that they are intended to explain only some features of a particular reaction (such as the "low-momentum transfer events", or a particular angular distribution which occurs). These partial models are popular not only because they are often easier for the theorist to construct, but also because the experimentalist can more easily test the limited model.

A possible answer to the experimental problem is to use a "two experiment" test where one of the experiments consists of a complete theoretical model. To be specific, suppose that a theoretical model exists for reaction (9). Then by standard Monte Carlo techniques one can generate a set of fake events F_i, which (if the model is correct) will be indistinguishable from the real events A_i. Then the distributions

$$L^2(A_i,F_j) \quad \text{and} \quad [L^2(A_i,A_j) \quad \text{or} \quad L^2(F_i,F_j)]$$

should differ only statistically.

Acknowledgements

Professors Bincer, Ebel, Goebel, Good, Olsson, Reeder, Sandweiss and Thompson have provided useful ideas in discussion. This work was supported in part by the U.S. Atomic Energy Commission and by the American-Swiss Foundation for Scientific Exchange. It is a pleasure to thank the NP Division of CERN, in particular Professors Preiswerk and Cocconi, for their hospitality.

REFERENCES

1. Humphrey and Rosenfeld, Annual Reviews of Nuclear Science, 13, 103 (1963).

2. J.J. Sakurai, Advanced Quantum Mechanics, Addison-Wesley Publishing Co., Reading, Mass. (1964).

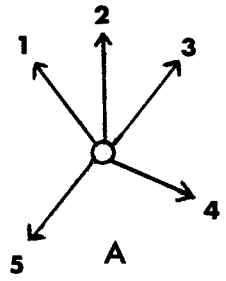
3. For other approaches to this problem see

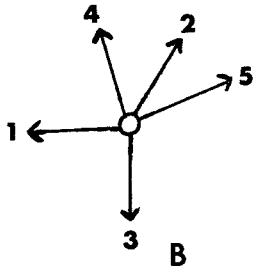
Racah, Nuovo Cimento (Supplemento), 14, 75.

Rühl, Lectures on the Lorentz Group and Harmonic Analysis, CERN Academic Training Program (1967-1968).

APPENDIX

The problem is to perform the rotation (rotation and Lorentz translation) which minimizes the squared length of the "difference vector" $\underline{V}^{-}_{AB}$. [It may be visualized by considering the "porcupine problem", which is the non-relativistic case. Imagine two rigid porcupines, A and B





with equal numbers of labelled quills. Suppose that rubber bands (of equilibrium length 0) connect the ends of like-numbered quills. Suppose further that except for the rubber bands the porcupines do not interact in any way; they "slide through" each other freely. Then let the system come to equilibrium. This will be the desired solution.]

No explicit use is made of the fact that each event satisfies conservation equations; indeed the analysis could be done even if they did not. Thus any sign convention can be used for the 4 momenta, either maintaining or suppressing the distinction between "incoming" and "outgoing" particles, provided only that the same convention is used in both events.

The solution is an iterative least square process:

1) From the current values of the 4 vector calculate the transformation which would minimize $(\underline{V}^{-})^2$ if the transformation were linear.

2) Perform the transformation (exactly, not in linear approximation) to obtain new 4 vectors .

3) If convergence is complete, calculate $(\underline{V}^{-})^2$, otherwise repeat from step 1.

Consider the steps in detail:

1. Find a transformation. The least squares procedure for minimizing a vector starts by trying to make it be 0. The linearized (Newton's method) equations for this are

$$\underline{V}^{-} + \frac{\partial V^{-}}{\partial X_j} X_j = \underline{0} \qquad (11)$$

which are 4n equations to be solved for the 6 parameters, X_j, specifying the transformation. (Summation on repeated indices is assumed throughout.) Proceed by multiplying Eq. (11) by the transpose of the coefficient matrix to get a 6 × 6 linear system

$$\frac{\partial \underline{V}^-}{\partial X_i} \cdot \frac{\partial \underline{V}^-}{\partial X_j} \; X_j = - \frac{\partial \underline{V}^-}{\partial X_i} \cdot \underline{V} \; . \tag{12}$$

To calculate the quantities in Eq. (12), we must define the transformation parameters. Since at this stage linearity is assumed an infinitesmal transformation will suffice. Let a particle in event A have four momentum p_μ and in event B, q_μ. Although the actual rotation will be carried out keeping p_μ fixed, at this stage assume equal and opposite transformations

$$\begin{aligned} p'_\mu &= I^{-1}_{\mu\nu}(X)\; p_\nu \\ q'_\mu &= I_{\mu\nu}(x)\; q_\nu \end{aligned} \tag{13}$$

where I(X) is the infinitesimal transformation. The symmetric treatment has two desired consequences: succeeding equations are manifestly invariant to interchange of the events, and the (unknown) finite transformation sought is only half as big, thus making more realistic the linearity assumption.

A useful way to write the infinitesimal rotation is

$$I(X) = 1 + g_i X_i = \begin{pmatrix} 1 & -X_3 & X_2 & -iX_4 \\ X_3 & 1 & -X_1 & -iX_5 \\ -X_2 & X_1 & 1 & -iX_6 \\ iX_4 & iX_5 & iX_6 & 1 \end{pmatrix}, \quad I^{-1}(X) = I(-X) = \tilde{I}(X) \; . \tag{14}$$

The g_i are the 6 generators of the transformation group. In the limit $|X| \to 0$, I(X) is a pure rotation in 3 space of angle $\vartheta = \sqrt{X_1^2 + X_2^2 + X_3^2}$, plus a pure Lorentz translation of $\underline{\beta} = (X_4, X_5, X_6)$.

The right side of Eq. (12) becomes

$$C_i = - V^-_\mu \frac{\partial V^-_\mu}{\partial X_i} = V^-_\mu (g_i)_{\mu\nu} \; V^+_\nu = (\underline{Q} \times \underline{P}, F\underline{P} - E\underline{Q}) \tag{15}$$

where $\underline{V}^+$ is the sum vector analogous to the difference vector $\underline{V}^-$, and E and F are the energies associated with $\underline{P}$ and $\underline{Q}$ respectively. Likewise the matrix on the left of Eq. (12) is

$$M_{ij} = \frac{\partial V_\mu^-}{\partial X_i} \frac{\partial V_\mu^-}{\partial X_j} = (g_i)_{\mu\nu}(g_j)_{\mu\rho}\, V_\nu^+ V_\rho^+ \ . \tag{16}$$

Write M in terms of its 3×3 corners

$$M = \left(\begin{array}{c|c} UL & UR \\ \hline LL & LR \end{array}\right) \tag{17}$$

Then

$$(UL)_{ij} = (\underline{Q} + \underline{P})^2\, \delta_{ij} - (Q_i + P_i)(Q_j + P_j)$$

$$(LR)_{ij} = (E + F)^2\, \delta_{ij} - (Q_i + P_i)(Q_j + P_j)$$

$$UR = \begin{pmatrix} 0 & -(F+E)(Q_3+P_3) & (F+E)(Q_2+P_2) \\ (F+E)(Q_3+P_3) & 0 & -(F+E)(Q_1+P_1) \\ -(F+E)(Q_2+P_2) & (F+E)(Q_1+P_1) & 0 \end{pmatrix} \tag{18}$$

and, by the over-all symmetry of M, LL is the transpose of UR. δ_{ij} is the Kronecker delta ($\delta_{ij} = 1$ only if $i = j$, otherwise 0). In Eqs. (15) and (18) each component is to be summed over the n particle pairs.

Finally Eq. (12) can be solved for the unknown vector X by standard techniques of linear algebra, thus completing the task of this section.

2. Perform the transformation. Now we have the 6 parameters X defining the desired transformation of the type (14); however the vector X was calculated assuming equal and opposite transformations on the two porcupines. In fact only the second of Eqs. (13) will be used; therefore X must be doubled. However, if the rotation angle $\vartheta = \sqrt{X_1^2 + X_2^2 + X_3^2}$ exceeds a quarter revolution the linear approximation of step 1 cannot have been very sensible. It is useful to scale down the vector X in this case rather than doubling it. In the rest of this section the symbol X denotes the solution of Eq. (12) after doubling or scaling as appropriate.

Since in general X is not small, the problem is to find a finite transformation T(X) which is the extension of I(X)

$$T(X) = \lim_{n\to\infty} \left[I\left(\frac{X}{n}\right)\right]^n = \lim_{n\to\infty} \left[1 + \frac{Z}{n}\right]^n = e^Z \qquad (19)$$

where $Z = I(X) - 1 = g_i X_i$.

[In fact, since the least square procedure is basically tolerant, it is possible to use an approximation to T(X), so long as the approximation is an exact (i.e. mass preserving) transformation and reduces to T(X) when |X| is small. An example of a useful approximation is

$$T'(X) = \mathcal{L}(X_4, X_5, X_6)\mathcal{R}(X_1, X_2, X_3) \qquad (20)$$

where $\mathcal{L}$ is a pure Lorentz transformation and $\mathcal{R}$ is a pure rotation. Of course, reversing the order in Eq. (20) would also work.]

The formula for T(X) does not seem to be given in many of the standard references, and is included here for completeness [3]. First define the building blocks:

$$Z = \begin{pmatrix} 0 & -X_3 & X_2 & -iX_4 \\ X_3 & 0 & -X_1 & -iX_5 \\ -X_2 & X_1 & 0 & -iX_6 \\ iX_4 & iX_5 & iX_1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & X_6 & -X_5 & -iX_1 \\ -X_6 & 0 & X_4 & -iX_2 \\ X_5 & -X_4 & 0 & -iX_3 \\ iX_1 & iX_2 & iX_3 & 0 \end{pmatrix}$$

$$Z^2 = \begin{pmatrix} X_4^2 - X_2^2 - X_3^2 & X_1X_2 + X_4X_5 & X_1X_3 + X_4X_6 & i(X_3X_5 - X_2X_6) \\ = Z_{12} & X_5^2 - X_1^2 - X_3^2 & X_2X_3 + X_5X_6 & i(X_1X_6 - X_3X_4) \\ = Z_{13} & = Z_{23} & X_6^2 - X_1^2 - X_2^2 & i(X_2X_4 - X_1X_5) \\ = Z_{14} & = Z_{24} & = Z_{34} & X_4^2 + X_5^2 + X_6^2 \end{pmatrix} \qquad (21)$$

$$S = \frac{\mathrm{Tr}Z^2}{2} = X_4^2 + X_5^2 + X_6^2 - (X_1^2 + X_2^2 + X_3^2) = \text{"}\beta^2 - \vartheta^2\text{"}$$

$$D = ZY = X_1X_4 + X_1X_5 + X_3X_6 = \text{"}\beta \cdot \vartheta\text{"}$$

Now since $Z^3 = SZ + DY$, the series expansion for $e^Z = 1 + Z + \frac{Z^2}{2} + \ldots$ can contain only the above matrices and functions of S and D. Therefore

$$T(Z) = e + fZ^2 + gZ + hY \tag{22}$$

where f, g, h, and e are functions only of S and D. With a modest amount of algebra it can be shown that

$$e = \frac{b^2 \cosh a + a^2 \cos b}{b^2 + a^2} \qquad b = \frac{\cosh a - \cos b}{b^2 + a^2}$$

$$g = \frac{a \sinh a + b \sin b}{b^2 + a^2} \qquad h = \frac{b \sinh a - a \sin b}{b^2 + a^2} \tag{23}$$

where

$$a = \sqrt{\frac{\sqrt{S^2 + 4D^2} + S}{2}} \qquad b = \sqrt{\frac{\sqrt{S^2 + 4D^2} - S}{2}}$$

The positive roots are to be taken in all cases.

3. Terminal Considerations. The result of steps 1 and 2 is a transformed version of the second porcupine which is a linearized solution to the least squares problem. The problem is non-linear, so it will frequently be necessary to return to step 1, using the approximate solution as the new starting point. The iterative procedure will grind to a halt when the right side of Eq. (12) becomes 0. Therefore it is wise to declare "convergence" and terminate the process when this vector becomes small.

So the process has converged, but is it a solution, namely the state with minimum $(\underline{V}^-)^2$? Consider the three requirements for a minimum:

i) $\frac{\partial(\underline{V}^-)^2}{\partial X_i} = 0$ This is obviously satisfied, since C is the gradient in question.

ii) $\frac{\partial^2(\underline{V}^-)^2}{\partial X_i \partial X_j} \geq 0$, i.e. this matrix should have no negative eigenvalue.

The second derivative matrix in question is just

$$\frac{\partial^2(\underline{V}^-)^2}{\partial X_i \partial X_j} = M_{ij} + 2\underline{V}^- \frac{\partial T(X)}{\partial X_i \partial X_j} \underline{V}^- . \tag{24}$$

While M_{ij} is a normal least square matrix and can have no negative eigenvalue, the complete second derivative matrix (24) certainly can have negative eigenvalues. [Since $(\underline{V}^-)^2$ must be periodic for pure

rotation about any axis there must be maxima as well as minima.] While the linearized least squares process can be in equilibrium at a non-minimum, it cannot be in stable equilibrium. Therefore only states where condition 2 is satisfied will normally arise. In any case the eigenvalues of Eq. (24) can be checked; if any are negative it is a simple matter to move away from the false solution and resume the iteration. With this (almost always superfluous) addition the process is guaranteed to produce at least a local minimum of $(\underline{V}^-)^2$.

iii) Minimum of all minima. Finally it must be shown that the minimum found is the absolute minimum rather than just one of many local minima. This is true, because there is only one local minimum. Proof: Suppose that p and q are the vectors corresponding to a local minimum. Then

$$(p_\mu - q_\mu) \frac{\partial q_\mu}{\partial X_i} = (p - q)\, g_i q = 0 \qquad \text{(for all i)} \tag{25}$$

where, for simplicity the derivatives are calculated assuming p fixed. Suppose another minimum exists with vector p″ and q″. There always exists a transformation T″ such that T″ q″ = g, so this second minimum also occurs at p′ = T″ p″ and q. Assume that p′ ≠ p, and let T be the transformation such that T p = p′. Then

$$(Tp - q)\, g_i q = 0 \qquad \text{(for all i)}. \tag{26}$$

Combining Eqs. (25) and (26)

$$(Tp - p)\, g_i q = 0\ . \tag{27}$$

Now consider

$$(T^2 p - q)\, g_i q = \Big[(T^2 p - Tp) + (Tp - q) \Big] g_i q = T(Tp - p) g_i q = 0\ . \tag{28}$$

Therefore, if the derivative vector is 0 at two distinct points, it has zeros at an infinite set of equally spaced points in X space. Hyperbolic functions do not have periodically spaced zeros; therefore the transformation T connecting the two presumed minima must be a pure rotation.

Rotation, of course, will lead to periodic zeros in the derivative: a rotation through 2π about any axis reproduces the initial state. So the only remaining possibility is that two local minima are related to each other by a rotation of angle $< 2\pi$. A brief look at the rotation matrix shows that this cannot happen.

Therefore any local minimum found is also the absolute minimum.