vector is discontinuous. We shall require that at such points the two tangent vectors

$$\left.\frac{\partial}{\partial t}\right|_- \quad \text{and} \quad \left.\frac{\partial}{\partial t}\right|_+ \quad \text{satisfy} \quad g\left(\left.\frac{\partial}{\partial t}\right|_-, \left.\frac{\partial}{\partial t}\right|_+\right) = -1,$$

that is, they point into the same half of the null cone.

Proposition 4.5.1

Let $\mathscr{U}$ be a convex normal coordinate neighbourhood about q. Then the points which can be reached from q by timelike (respectively non-spacelike) curves in $\mathscr{U}$ are those of the form $\exp_q(\mathbf{X})$, $\mathbf{X} \in T_q$ where $g(\mathbf{X}, \mathbf{X}) < 0$ (respectively $\leqslant 0$). (Here, and for the rest of this section, we consider the map exp to be restricted to the neighbourhood of the origin in T_q which is diffeomorphic to $\mathscr{U}$ under $\exp_q$.)

In other words, the null geodesics from q form the boundary of the region in $\mathscr{U}$ which can be reached from q by timelike or non-spacelike curves in $\mathscr{U}$. This is fairly obvious intuitively but because it is fundamental to the concept of causality we shall prove it rigorously. We first establish the following lemma:

Lemma 4.5.2

In $\mathscr{U}$ the timelike geodesics through q are orthogonal to the three-surfaces of constant σ ($\sigma < 0$) where the value of σ at $p \in \mathscr{U}$ is defined to be $g(\exp_q^{-1}p, \exp_q^{-1}p)$.

The proof is based on the fact that the vector representing the separation of points equal distances along neighbouring geodesics remains orthogonal to the geodesics if it is so initially. More precisely, let $\mathbf{X}(t)$ denote a curve in T_q, where $g(\mathbf{X}(t), \mathbf{X}(t)) = -1$. One must show that the corresponding curves $\lambda(t) = \exp_q(s_0\mathbf{X}(t))$ (s_0 constant) in $\mathscr{U}$, where defined, are orthogonal to the timelike geodesics $\gamma(s) = \exp_q(s\mathbf{X}(t_0))$ (t_0 constant). Thus in terms of the two-surface α defined by $x(s,t) = \exp_q(s\mathbf{X}(t))$, one must prove that

$$g\left(\left(\frac{\partial}{\partial s}\right)_\alpha, \left(\frac{\partial}{\partial t}\right)_\alpha\right) = 0$$

(see figure 11). Now

$$\frac{\partial}{\partial s} g\left(\frac{\partial}{\partial s}, \frac{\partial}{\partial t}\right) = g\left(\frac{\mathrm{D}}{\partial s}\frac{\partial}{\partial s}, \frac{\partial}{\partial t}\right) + g\left(\frac{\partial}{\partial s}, \frac{\mathrm{D}}{\partial s}\frac{\partial}{\partial t}\right).$$

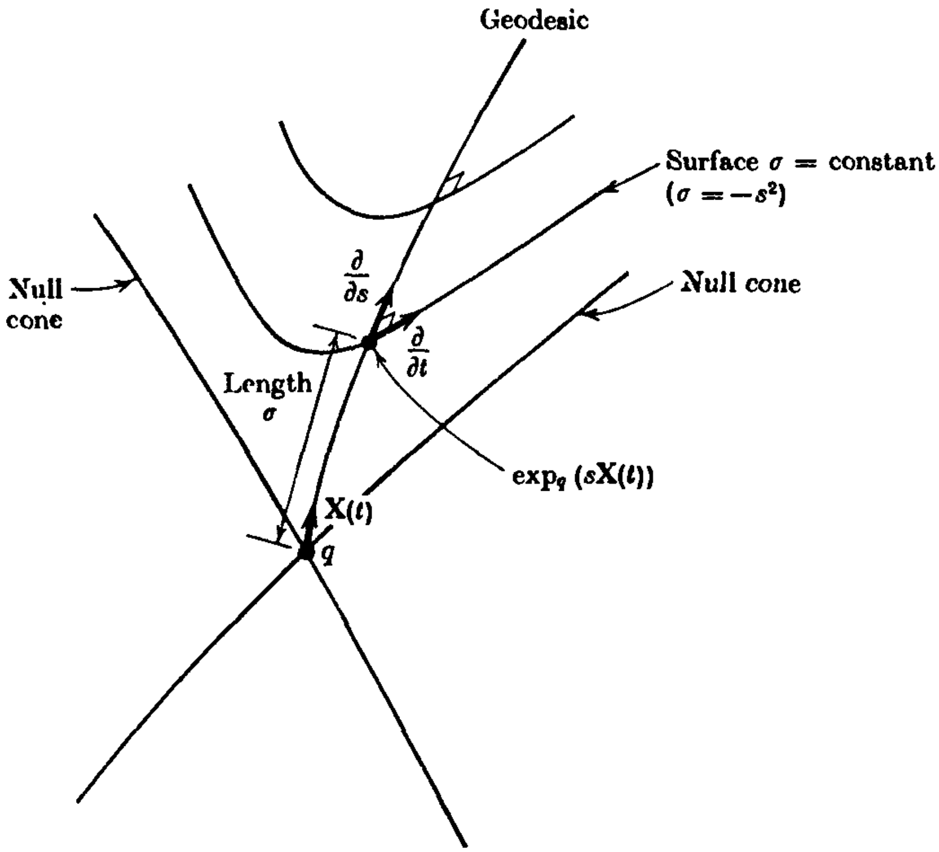

FIGURE 11. In a normal neighbourhood, surfaces at constant distance from q are orthogonal to the geodesics through q.

The first term on the right is zero as $\partial/\partial s$ is the unit tangent vector to the timelike geodesics from q. In the second term one has from the definition of the Lie derivative that

$$\frac{D}{\partial s}\frac{\partial}{\partial t} = \frac{D}{\partial t}\frac{\partial}{\partial s}.$$

Thus $$\frac{\partial}{\partial s} g\left(\frac{\partial}{\partial s}, \frac{\partial}{\partial t}\right) = g\left(\frac{\partial}{\partial s}, \frac{D}{\partial t}\frac{\partial}{\partial s}\right) = \frac{1}{2}\frac{\partial}{\partial t} g\left(\frac{\partial}{\partial s}, \frac{\partial}{\partial s}\right) = 0.$$

Therefore $g(\partial/\partial s, \partial/\partial t)$ is independent of s. But at $s = 0$, $(\partial/\partial t)_\alpha = 0$. Thus $g(\partial/\partial s, \partial/\partial t)$ is identically zero. □

Proof of proposition 4.5.1. Let C_q denote the set of all timelike vectors at q. These constitute the interior of a solid cone in T_q with vertex at the origin. Let $\gamma(t)$ be a timelike curve in $\mathcal{U}$ from q to p and let $\bar{\gamma}(t)$ be the piecewise C^2 curve in T_q defined by $\bar{\gamma}(t) = \exp_q^{-1}(\gamma(t))$. Then identifying the tangent space to T_q with T_q itself, one has

$$(\partial/\partial t)_\gamma|_q = (\partial/\partial t)_{\bar{\gamma}}|_q.$$

Therefore at q, $(\partial/\partial t)_{\bar{\gamma}}$ will be timelike. This shows that the curve $\bar{\gamma}(t)$ will enter the region C_q. But $\exp_q(C_q)$ is the region of $\mathscr{U}$ on which σ is negative and in which by the previous lemma the surfaces of constant σ are spacelike. Thus σ must monotonically decrease along $\gamma(t)$ since $(\partial/\partial t)_\gamma$ being timelike can never be tangent to the surfaces of constant σ and since at any non-differentiable point of $\gamma(t)$ the two tangent vectors point into the same half of the null cone. Therefore $p \in \exp_q(C_q)$ which completes the proof for timelike curves. To prove that a non-spacelike curve $\gamma(t)$ remains in $\exp_q(\bar{C}_q)$, one performs a small variation of $\gamma(t)$ which makes it into a timelike curve. Let $\mathbf{Y}$ be a vector field on T_q such that in $\mathscr{U}$ the induced vector field $\exp_{q*}(\mathbf{Y})$ is everywhere timelike and such that $g(\mathbf{Y}, (\partial/\partial t)_\gamma|_q) < 0$. For each $\epsilon \geqslant 0$ let $\beta(r, \epsilon)$ be the curve T_q starting at the origin such that the tangent vector $(\partial/\partial r)_\beta$ equals $(\partial/\partial t)_{\bar{\gamma}}|_{t=r} + \epsilon\mathbf{Y}|_{\beta(r,\epsilon)}$. Then $\beta(r, \epsilon)$ depends differentiably on r and ϵ. For each $\epsilon > 0$, $\exp_q(\beta(r, \epsilon))$ is a timelike curve in $\mathscr{U}$ and so is contained in $\exp_q(C_q)$. Thus the non-spacelike curve $\exp_q(\beta(r, 0)) = \gamma(r)$ is contained in $\overline{\exp_q(C_q)} = \exp_q(\bar{C}_q)$. □

Corollary

If $p \in \mathscr{U}$ can be reached from q by a non-spacelike curve but not by a timelike curve, then p lies on a null geodesic from q. □

The length of a non-spacelike curve $\gamma(t)$ from q to p is

$$L(\gamma, q, p) = \int_q^p \left[-g\left(\frac{\partial}{\partial t}, \frac{\partial}{\partial t}\right)\right]^{\frac{1}{2}} \mathrm{d}t,$$

where the integral is taken over the differentiable sections of the curve.

In a positive definite metric one may seek the shortest curve between two points but in a Lorentz metric there will not be any shortest curve as any curve can be deformed into a null curve which has zero length. However, in certain cases there will be a longest non-spacelike curve between two points or between a point and a spacelike three-surface. We deal first with the situation when the two points are close together. We shall then derive necessary conditions in the general case when the two points are not close. The sufficient condition in this case will be dealt with in §6.7.

Proposition 4.5.3

Let q and p lie in a convex normal neighbourhood $\mathscr{U}$. Then, if q and p can be joined by a non-spacelike curve in $\mathscr{U}$, the longest such curve is the unique non-spacelike geodesic curve in $\mathscr{U}$ from q to p. Moreover,