JOURNAL OF DIFFERENTIAL EQUATIONS 30, 165–174 (1978)

General Relativity and Conjugate Ordinary Differential Equations*

FRANK J. TIPLER

Department of Mathematics, University of California, Berkeley, California 94720

Received July 7, 1977; revised November 4, 1977

I am convinced that the reinjection of more or less directly empirical ideas into mathematics is a necessary condition to conserve the freshness and vitality of the subject.

—JOHN VON NEUMANN

This paper will essentially be concerned with a very old problem in the theory of ordinary differential equations: what are sufficient conditions on the function $F(t)$ in the equation

$$\frac{d^2x}{dt^2} + F(t)x = 0 \tag{1}$$

such that (1) has a solution with at least two zeros in an interval I? However, the sufficient conditions derived here are unusual in one respect: the motivation for the theorems comes from the physics of the gravitational field—General Relativity—and not from mathematical aesthetics.

The notation and conventions of the equations and concepts from General Relativity theory will be the same as in [1]; see [2] for a more elementary introduction to General Relativity. Recall that the object of study in General Relativity is a *spacetime*, which is a four-dimensional boundaryless Hausdorff manifold with a non-degenerate Lorentz metric g. A *timelike vector* $\mathbf{V}$ is a vector satisfying $g(\mathbf{V}, \mathbf{V}) < 0$, a *null vector* satisfies $g(\mathbf{V}, \mathbf{V}) = 0$ and a *spacelike vector* satisfies $g(\mathbf{V}, \mathbf{V}) > 0$. Latin indices label space *and* time dimensions and run from 1 to 4; Greek indices label space dimensions only and run from 1 to 3.

Equation (1) is connected to General Relativity via the Jacobi equation

$$\frac{d^2Z^\alpha}{dt^2} = R_{\alpha a\beta b} V^a Z^\beta V^b, \tag{2}$$

* Work supported by the National Science Foundation under Grant MCS-76-21525.

0022-0396/78/0302-0165$02.00/0

which is defined along a timelike geodesic $\gamma(t)$. Z^α is a Jacobi field along $\gamma(t)$, t is the proper time along $\gamma(t)$, V^a is a unit tangent vector to $\gamma(t)$, and $R_{\alpha a\beta b}$ is the Riemann tensor. It can be shown ([3], [4]) that a vector Z^α of equation (2) will vanish at a given value t if and only if a corresponding solution x of equation (1) also vanishes at t, where

$$F(t) \equiv \tfrac{1}{3}(R_{ab}V^aV^b + 2\sigma^2) \tag{3}$$

$R_{ab} \equiv R^c{}_{acb}$ is the Ricci tensor and σ^2 is a non-negative function of t.

More generally, it can be shown that Z^α will vanish at two points t_1, t_2 in an interval I if and only if $x(t_1) = x(t_2) = 0$; that is, the points t_1, t_2 are *conjugate* along $\gamma(t)$ if these points are also zeros of x. This connection between (1) and the conjugate points of the Jacobi equation is well-known, and in fact it motivated the use of the word "conjugate" in connection with equation (1):

DEFINITION. Equation (1) will be said to be *conjugate* in an interval I if there exists a solution $x(t)$ with at least two zeros in I.

However, what is not so well-known is the fact that there is a connection between conjugate points along a timelike geodesic and the causal structure of spacetime. For example,

THEOREM (Avez–Hawking). *Let p and q be points in a globally hyperbolic set N such that there is a curve from p to q whose tangent vector is everywhere timelike. Then there exists a timelike geodesic from p to q of maximal length. Furthermore, a non-spacelike geodesic of maximal length between two points p, q in spacetime has no conjugate points between p and q.* (*For a proof of this theorem see* [1, *p.* 213]. *Recall that a globally hyperbolic set is, roughly speaking, a set whose properties are completely determined from Cauchy data on a certain spacelike hypersurface. See* [1, *p.* 206] *for more details.*)

Now Hawking and Penrose have shown [5] that under conditions which are thought to occur in the universe—conditions such as the absence of causality violation and the presence of a compact spacelike hypersurface—there is a timelike geodesic[1] which is entirely contained inside a globally hyperbolic set. Since by the above theorem this geodesic can have no conjugate points, the existence of this geodesic puts strong restrictions both on the Ricci tensor (because of equation (3)) and on the structure of the geodesic itself.

In fact, Hawking and Penrose show that this geodesic must be incomplete. They accomplish this by assuming first that all known forms of matter cause

[1] More precisely, they show that a *non-spacelike* geodesic must be entirely contained in a globally hyperbolic set. It can be shown that if this geodesic is null and complete, then there exists a solution $x(t)$ to (1) with a positive definite function $F(t)$ similar to (3) such that $x(t)$ has at most one zero in $(-\infty, +\infty)$. See Ref. [1] for details.

the Ricci tensor to satisfy the *timelike convergence condition*, which says that

$$R_{ab}K^aK^b \geqslant 0 \tag{4}$$

at every point of the manifold for all timelike vectors K^a. Equations (3) and (4) imply $F(t) \geqslant 0$ along any timelike geodesic. Then the incompleteness of the geodesic follows from the Avez–Hawking Theorem, the assumption that $R_{ab}K^aK^b > 0$ at least once in every timelike geodesic's history, and

THEOREM 1 (Hawking–Penrose). *Let $F(t)$ be continuous and $F(t) \geqslant 0$ on $(-\infty, +\infty)$. If $F(t_1) > 0$ for some point $t_1 \in (-\infty, +\infty)$, then* (1) *is conjugate on $(-\infty, +\infty)$.*

The proofs of Theorem 1 *which have appeared in the literature* [1, *p.* 98; 5, *p.* 541] *are based on the Riccati equation associated with* (1) *and are quite complicated. I shall give a simple proof of Theorem* 1.

Proof (construction). The solution $x(t)$ defined by

$$x(t_1) > 0, \qquad \left.\frac{dx}{dt}\right|_{t=t_1} = 0$$

has at least two zeros in $(-\infty, +\infty)$. To see this, note that $F(t_1) > 0$ implies

$$\int_{t_1}^{t_2} F(t)\,dt > 0, \qquad \int_{t_0}^{t_1} F(t)\,dt > 0$$

where $t_0 < t_1 < t_2$, since $F(t) \geqslant 0$ and $F(t)$ is continuous. From (1), this gives for some t_0, t_2

$$\left.\frac{dx}{dt}\right|_{t=t_2} = -\int_{t_1}^{t_2} xF(t)\,dt + \left.\frac{dx}{dx}\right|_{t=t_1} < 0 \tag{5}$$

and

$$\left.\frac{dx}{dt}\right|_{t=t_0} = \int_{t_0}^{t_1} xF(t)\,dt + \left.\frac{dx}{dt}\right|_{t=t_1} > 0 \tag{6}$$

The inequality (5) and $d^2x/dt^2 \leqslant 0$ tell us that $x(t)$ must have a zero in $(t_1, +\infty)$, and (6) tells us that it must have a zero in $(-\infty, t_1)$. ∎

The preceding theorem asserts conjugacy under the assumptions $F(t) \geqslant 0$ for all t and $F(t_1) > 0$ for some t_1. The former assumption is rather dubious from the physical point of view, for there are types of matter for which $R_{ab}K^aK^b < 0$ at certain points [1, p. 95; 6]. It is possible, however, to prove that a pair of conjugate points exists on any complete timelike geodesic $\gamma(t)$ provided only that $R_{ab}K^aK^b$ is greater than zero *on the average* along thegeodesic, where K^a is the unit tangent vector to γ.

THEOREM 2. *Let $F(t)$ be continuous on $(-\infty, +\infty)$. If*

$$\int_{-\infty}^{+\infty} F(t)\, dt > 0 \tag{7}$$

then (1) *is conjugate on* $(-\infty, +\infty)$.

Note. Theorem 2 is still true even if the integral in (7) does not converge, provided we regard the expression

$$\int_{-\infty}^{+\infty} F(t)\, dt > 0$$

as a shorthand notation for

$$\liminf_{\substack{t' \to -\infty \\ t'' \to +\infty}} \int_{t'}^{t''} F(t)\, dt > 0 \tag{8}$$

Proof of Theorem 2. Inequality (8) is equivalent to the existence of a number t_0 such that

$$\liminf_{t'' \to +\infty} \int_{t_0}^{t''} F\, dt > 0 \tag{9}$$

$$\liminf_{t' \to -\infty} \int_{t'}^{t_0} F\, dt > 0 \tag{10}$$

Let $x(t)$ be a solution of (1) satisfying the conditions

$$x(t_0) = 1, \qquad x'(t_0) = 0$$

We will now show that $x(t)$ must have a zero at some $t > t_0$. A similar argument will show that $x(t)$ will also have a zero at some $t < t_0$, and so (1) will be conjugate on $(-\infty, +\infty)$.

Suppose on the contrary that $x(t) > 0$ on $(t_0, +\infty)$. Let

$$z(t) = \frac{-x'(t)}{x(t)},$$

which gives

$$z'(t) - z^2(t) - F(t) = 0.$$

Thus for any $t \in [t_0, +\infty)$, we have

$$z(t) = + \int_{t_0}^{t} z^2(t)\, dt + \int_{t_0}^{t} F(t)\, dt \tag{11}$$

By the inequality (9), there is a number $t_1 > t_0$ and a number $c > 0$ for which

$$\int_{t_0}^{t} F(t)\, dt \geqslant c$$

for any $t \in (t_1, +\infty)$. Hence for any $t \in (t_1, +\infty)$

$$z(t) \geqslant \int_{t_0}^{t} z^2(t)\, dt + c > 0.$$

Let $R(t) = \int_{t_0}^{t} z^2(t)\, dt + c$, so that $R' = z^2 \geqslant [\int_{t_0}^{t} z^2(t)\, dt + c]^2 = R^2$. Integrating $R' \geqslant R^2$ gives $-1/R(t) \geqslant t$-const for all $t \in (t_1, +\infty)$. This is a contradiction, since $R(t) \geqslant c > 0$ in this interval. ∎

Equation (1) with $F(t) \equiv 0$ has the solution $x(t) \equiv 1$. This shows that (7) cannot be replaced with $\int_{-\infty}^{+\infty} F(t)\, dt = 0$.

As mentioned above, there will exist in many cases of physical interest a timelike geodesic $\gamma(t)$ of maximal length which is entirely contained in a globally hyperbolic set. By Theorems 1 and 2, this geodesic must be incomplete. We can also prove oscillation theorems which can be used to place restrictions on $R_{ab}K^aK^b$ along any maximal geodesic. For example, $R_{ab}K^aK^b$ is expected [1, pp. 256–261] to diverge along an incomplete timelike geodesic $t \to t_0$, where t_0 is the finite limit to the proper time. The following theorem places restrictions on the rate of divergence of $R_{ab}K^aK^b$ along $\gamma(t)$.

THEOREM 3. *Let $F(t)$ be continuous and positive in the finite interval (a, b). If either*

$$\liminf_{t \to b} [(t - b)^2 F(t)] > \tfrac{1}{4}$$

or

$$\liminf_{t \to a} [(a - t)^2 F(t)] > \tfrac{1}{4}$$

then (1) *is oscillatory on* (a, b).

Theorem 3 has the following obvious corollary:

COROLLARY 1. *Let $W(t) = F'(t)[F(t)]^{-3/2}$. If $\lim_{t \to b} W(t) = 0$ and $\lim_{t \to b} F(t) = \infty$, then* (1) *is oscillatory on the finite interval* (a, b).

Theorem 3 and its corollary are of course finite interval versions of well-known oscillation sufficiency conditions on $(1, +\infty)$. In fact, Theorem 3 was proved many years ago by Leighton [7, p. 45], and the proof of the corollary is essentially the same as the proof of its infinite interval analogue [8, p. 472]. I include these results in this paper in order to point out their application to physics.

It is also well known that

$$\lim_{t \to t_0} \int_{1}^{t} F(t)\, dt = +\infty \tag{12}$$

implies that (1) is oscillatory on $[1, +\infty)$ if $t_0 = +\infty$. However, if t_0 is finite, then (1) need not be oscillatory on $[1, t_0)$. For the Euler equation

$$x'' + \frac{x}{4t^2} = 0 \tag{13}$$

has a solution $x = t^{1/2}$ which is positive in $(0, 1]$, and this implies that (13) is disconjugate in $(0, 1]$. But

$$\lim_{t\to 0} \int_t^1 \frac{dt}{4t^2} = +\infty$$

Note that any solution of (13) which is non-zero in $(0, 1]$ approaches zero as $t \to 0$. This illustrates the conclusions of

THEOREM 4. *Let $F(t)$ be continuous and $F(t) \geqslant 0$ on the finite interval $[a, b)$ and suppose*

$$\lim_{t\to b} \int_{t'}^{t} ds \int_{t'}^{s} F(s')\, ds' = +\infty \tag{14}$$

for any $t' \in [a, b)$. Then either (1) *is oscillatory on $[a, b)$ or else all solutions $x(t)$ satisfy* $\lim_{t\to b} x = 0$ *(or both).*

Proof. Suppose not. Then there exists a solution $x(t)$ such that $x(t) > 0$ in $[c, b)$ for some $c > a$ and $\lim_{t\to b} x(t) \geqslant d > 0$. Let $M = \inf[\inf_{c\leqslant t\leqslant b} x(t), d] > 0$. Then

$$x'(s) - x'(c) = -\int_c^s F(s')\, x(s')\, ds' \leqslant -\mathrm{M} \int_c^s F(s')\, ds',$$

which implies

$$\begin{aligned} x(t) &= x(c) + x'(c)(t - c) - \int_c^t ds \int_c^s F(s')\, x(s')\, ds' \\ &\leqslant x(c) + x'(c)(t - c) - M \int_c^t ds \int_c^s F(s')\, ds'. \end{aligned}$$

Since (14) holds, the R.H.S. of the above inequality will become negative as $t \to b$. This implies that $x(t)$ changes sign in $[c, b)$, contrary to assumption. ▌

If (1) is disconjugate on $[a, b)$, then the condition $\lim_{t\to b} x(t) = 0$ places a strong restriction on $F(t)$, as shown in the following:

THEOREM 5. *Suppose $F(t)$ is continuous and non-negative on the interval $[a, b)$. If* (1) *is disconjugate on this interval and for all solutions of* (1) *we have*

$$\lim_{t\to b} x(t) = 0$$

Then

$$\int_a^b F(t)\,dt = +\infty.$$

Proof. Suppose not. Then since $F(t) \geqslant 0$, the integral

$$\int_a^t F(t')\,dt' \tag{15}$$

is monotonically increasing. This means that (15) must converge to some positive number as $t \to b$. We have for any $c \in [a, b)$

$$x'(t) = x'(c) - \int_c^t F(t')x\,dt'.$$

If we choose the solution $x(c) = 0$, $x'(c) > 0$, then $x \geqslant 0$ on (c, b) and

$$\sup_{[c,b)} x \leqslant x'(c)(b - c)$$

This gives for any $t \in [c, b)$

$$\begin{aligned} x'(t) &\geqslant x'(c) - \left[x'(c)(b - c) \int_c^t F(t')\,dt'\right] \\ &= x'(c)\left[1 - (b - c) \int_c^t F(t')\,dt'\right] \end{aligned}$$

If $x(b)$ is to be zero, then the term in brackets must vanish for some $t \in [c, b)$. However, by choosing c to be sufficiently close to b we can prevent this if (15) converges. Thus (15) must diverge as $t \to b$. ∎

The preceding Theorem has been proved by Leighton [9, p. 262] under the assumption that $F(t) > 0$ near b.

We can use the following theorem to show that along a maximal geodesic which is complete, $R_{ab}K^aK^b$ must vanish in an average sense as $t \to +\infty$. (For more details of this application see [10].)

THEOREM 6. *A sufficient condition for the conjugacy of* (1) *in the interval* $[t_0, +\infty)$ *is that there exist numbers* t_1, t_2 *with* $t_0 < t_1 < t_2$ *such that*

$$\frac{1}{t_1 - t_0} < \int_{t_1}^{t_2} F(t)\,dt \tag{16}$$

assuming that $F(t)$ *is continuous and* $F(t) \geqslant 0$ *in* $[t_0, +\infty)$.

Proof. There exists a solution $x(t)$ which has a zero at t_0. We will show that this solution has another zero in $(t_0, +\infty)$. For assume it does not. Then without loss of generality we can assume $x(t) > 0$ and $dx/dt \geqslant 0$ in $(t_0, +\infty)$, since if $dx/dt < 0$ at any point in $(t_0, +\infty)$, we would have a zero in $(t_0, +\infty)$ by the condition $F(t) \geqslant 0$. From (1) we obtain

$$\left.\frac{dx}{dt}\right|_{t=t_2} = \left.\frac{dx}{dt}\right|_{t=t_1} - \int_{t_1}^{t_2} x(t)F(t)\,dt$$

Since $dx/dt \geqslant 0$, we have $x(t) \geqslant x(t_1)$ for any $t > t_1$. Since $F(t) \geqslant 0$, we have

$$\frac{x(t_1)}{t_1 - t_0} \geqslant \left.\frac{dx}{dt}\right|_{t=t_1} \qquad \text{or} \qquad x(t_1) \geqslant (t_1 - t_0)\left.\frac{dx}{dt}\right|_{t=t_1}$$

Thus

$$\begin{aligned}\left.\frac{dx}{dt}\right|_{t=t_1} - \int_{t_1}^{t_2} x(t)F(t)\,dt &\leqslant \left.\frac{dx}{dt}\right|_{t=t_1} - \int_{t_1}^{t_2} x(t_1)F(t)\,dt \\ &\leqslant \left.\frac{dx}{dt}\right|_{t=t_1} - (t_1 - t_0)\left[\left.\frac{dx}{dt}\right|_{t=t_1}\right]\int_{t_1}^{t_2} F(t)\,dt \\ &= \left.\frac{dx}{dt}\right|_{t=t_1}\left[1 - (t_1 - t_0)\int_{t_1}^{t_2} F(t)\,dt\right]\end{aligned}$$

By hypothesis, the factor in brackets is negative. If $dx/dt\,|_{t=t_1} > 0$, then $dx/dt\,|_{t=t_2} < 0$, implying a zero of $x(t)$ in $(t_2, +\infty)$. If $dx/dt\,|_{t=t_1} = 0$, then $dx/dt\,|_{t=t_2} < 0$ since $\int_{t_1}^{t_2} x(t)F(t)\,dt > 0$ by assumption. In either case, $x(t)$ must have a zero in $(t_0, +\infty)$. This contradicts the assumption that $x(t)$ has no zeros in $(t_0, +\infty)$. Thus $x(t)$ must have at least one zero in $(t_0, +\infty)$, and so (1) is conjugate on $[t_0, +\infty)$. ∎

The notion of a point conjugate to a hypersurface is also useful in General Relativity [1, p. 273; 3, p. 13].

DEFINITION. A point p is said to be *conjugate* to a spacelike hypersurface S along a timelike geodesic $\gamma(t)$ which intersects S orthogonally if there exists along $\gamma(t)$ a function $x(t)$ with $x(p) = 0$, and in addition $x(t)$ satisfies equation (1) everywhere and the initial conditions

$$x(0) = 1, \qquad \left.\frac{dx}{dt}\right|_{t=0} = \chi^a{}_a$$

at the point $\gamma(0) = \gamma(t) \cap S$. $\chi^a{}_a$ is the contraction of the second fundamental form χ_{ab} of S [1, pp. 99–100; 3].

This concept of conjugacy to a hypersurface is used in General Relativity primarily to deduce restrictions on the size of globally hyperbolic sets. It can be shown [1, pp. 112, 206–217] that to each point q in a globally hyperbolic set containing the hypersurface S there is a timelike geodesic orthogonal to S, of maximal length from S to q, which does not contain any point conjugate to S between S and q. Thus if *all* the timelike geodesics which are orthogonal to S have points conjugate to S within a distance c from S, then all timelike curves must either have a length less than $2c$, or else leave the globally hyperbolic set about S after a proper time $2c$.

The following theorem can be used to place limits on the size of the globally hyperbolic set containing a spacelike hypersurface S provided it is known that $R_{ab}K^aK^b \geqslant 0$ everywhere, that $R_{ab}K^aK^b$ is positive in a suitable way near S, and that S is extremal. (An extremal hypersurface is one for which $\chi^a{}_a = 0$ at every point [2, pp. 539–540]. This class of hypersurfaces has been extensively studied by General Relativity theorists [11, 12].)

THEOREM 7. *Let $F(t) \geqslant 0$ be continuous on $[0, +\infty)$. Then the solution of* (1) *defined by*

$$x(0) = 1, \qquad \left.\frac{dx}{dt}\right|_{t=0} = 0$$

has a zero in the interval $[0, a + 1/b]$ provided

$$\int_0^a F(t)\, dt \geqslant b$$

Proof. Suppose not. Then we have $x > 0$ in this interval and

$$\left.\frac{dx}{dt}\right|_{t=a} = -\int_0^a F(t)\, x(t)\, dt \leqslant -x(a) \int_0^a F(t)\, dt = -x(a)b$$

Since $dx/dt \leqslant dx/dt\,|_{t=a}$ for all $t \geqslant a$ before the first zero of x, there must be a zero of x within a distance c of $t = a$, where c is defined by

$$\left.\frac{dx}{dt}\right|_{t=a} = \frac{-x(a)}{c}$$

Thus

$$c = -\frac{x(a)}{dx/dt\,|_{t=a}} \leqslant \frac{-x(a)}{-x(a)b} = \frac{1}{b}$$

Thus assuming $x > 0$ in the interval $[0, a + 1/b]$ implies a zero in this interval. This is a contradiction, so x must pass through zero in the interval $[0, a + 1/b]$. ∎

ACKNOWLEDGMENTS

It is a pleasure to thank Professor M. H. Protter, Professor J. B. McLeod, and an anonymous referee for very helpful comments on an earlier version of this paper.

References

1. S. W. Hawking and G. F. R. Ellis, "The Large Scale Structure of Spacetime," Cambridge Univ. Press, Cambridge, England, 1973.
2. Charles W. Misner, Kip S. Thorne, and John A. Wheeler, "Gravitation," Freeman, San Francisco, 1973.
3. Frank J. Tipler, Singularities in universes with negative cosmological constant, *Astrophys. J.* **209** (1976), 12–15.
4. Frank J. Tipler, "Causality Violation in General Relativity," Dissertation, University of Maryland, 1976.
5. S. W. Hawking and Roger Penrose, The singularities of gravitational collapse and cosmology, *Proc. Roy. Soc. London. Ser. A* **314** (1970), 529–548.
6. H. Epstein, V. Glaser, and A. Jaffe, Nonpositivity of the energy density in quantized field theories, *Nuovo Cimento* **36** (1965), 1016–1022.
7. Walter Leighton, On self-adjoint differential equations of second order, *J. London Math. Soc.* **27** (1952), 37–47.
8. Einar Hille, "Lectures on Ordinary Differential Equations," Addison–Wesley, Reading, Mass., 1969.
9. Walter Leighton, Principal quadratic functionals, *Trans. Amer. Math. Soc.* **67** (1949), 253–274.
10. Frank J. Tipler, Causality violation and singularities, *Ann. Physics* **108** (1977), 1–36.
11. James W. York, Jr., Role of conformal three-geometry in the dynamics of gravitation, *Phys. Rev. Lett.* **28** (1972), 1082–1085.
12. Yvonne Choquet-Bruhat, Sous-variétés maximales, ou à courbure constante, de variétés lorentziennes, *CR Acad. Sci. Paris, Ser. A* **280** (1975), 169–171.