A GENERAL RELATIVITY WORKBOOK

Thomas A. Moore

Pomona College

UNIVERSITY SCIENCE BOOKS
MILL VALLEY, CALIFORNIA

University Science Books
www.uscibooks.com

Production Manager: Paul Anagnostopoulos
Text Design: Yvonne Tsang
Cover Deisign: Genette Itoko McGrew
Manuscript Editor: Lee Young
Illustrator: Laurel Muller
Compositor: Cohographics
Printer & Binder: Victor Graphics

ISBN 978-1-891389-82-5

This book is printed on acid-free paper. ∞

Library of Congress Cataloging-in-Publication Data

Moore, Thomas A. (Thomas Andrew)
A general relativity workbook / Thomas A. Moore, Pomona College.
pages cm
Includes index.
ISBN 978-1-891389-82-5 (alk. paper)
1. General relativity (Physics) I. Title.
QC173.6.M66 2012
530.11—dc23
2012025909

Printed in North America
10 9 8 7 6 5 4 3 2 1

For Joyce, whose miraculous love always supports me and
allows me to take risks with life that I could not face alone,

and for Edwin Taylor, whose book with Wheeler set me on this path decades ago,
and whose gracious support and friendship has kept me going.

CONTENTS

PREFACE

Introductory Comments. General relativity is one of the greatest triumphs of the human mind. Together with quantum field theory, general relativity lies at the foundation of contemporary physics, and currently represents the most durable physical theory in existence, having survived nearly a century of development and increasingly rigorous testing without being contradicted or superseded. Long admired for its elegant beauty, general relativity has also (particularly in the past two decades) become an essential tool for working physicists. It provides the basis for understanding a huge variety of astrophysical phenomena ranging from active galactic nuclei, quasars, and pulsars to the formation, characteristics and destiny of the universe itself. It has driven the development of new experimental tools for testing the theory and for the detection of gravitational waves that represent one of the most lively and challenging areas of contemporary physics. Even engineers are starting to have to pay attention to general relativity: making the Global Positioning System function correctly requires careful attention to general relativistic effects.

In some ways, general relativity was so far ahead of its time that it took a long time for instrumentation and applications to catch up sufficiently to make it more than than an intellectual adventure for the curious. However, as general relativity has now moved firmly into the mainstream of contemporary physics with a wide and growing variety of applications, teaching general relativity to undergraduate physics majors has become both relevant and important, and the need for appropriate and up-to-date undergraduate-level textbooks has become urgent.

Audience. This textbook seeks to support a one-semester introduction to general relativity for junior and/or senior undergraduates. It assumes only that students have taken multivariable calculus and some intermediate Newtonian mechanics beyond a standard treatment of mechanics and electricity and magnetism at the introductory level (though students who have also taken linear algebra, differential equations, some electrodynamics and/or some special relativity will be able to move through the book more quickly and easily). This book has grown out of my experience teaching fourteen iterations of such an undergraduate course during my teaching career.

Those iterations have convinced me that undergraduates not only *can* develop a solid proficiency with the general relativity, but also that studying general relativity provides a superb introduction to the best practices of theoretical physics as well as a uniquely exciting and engaging introduction to ideas at the very frontier of physics, things that students rarely experience in other undergraduate courses.

Pedagogical Principles. Since students rarely see the tensor calculus used in general relativity in undergraduate mathematics courses, a course in general relativity must either teach this mathematics from scratch or seek to work around it (at some cost in coherence and depth of insight). In my experience, junior and senior undergraduates *can* master tensor calculus in an appropriately designed course, and that doing this is well worth the effort, as it provides the firm foundation needed for confidence and flexibility in confronting applications.

The pedagogical key for developing this mastery is for you (the student) to *personally own* the mathematics by working through most of the arguments and derivations *yourself*. Therefore, I have designed this textbook as a *workbook*. Each chapter opens with a concise core-concept presentation that helps you see the big picture without mathematical distraction. This presentation is keyed to subsequent "boxes" that I have designed to guide you in working through the supporting derivations as well as other details and applications whose direct presentation would obscure the core ideas. I have found this combination of overview and guided effort to be uniquely effective in building a practical understanding of the theory's core concepts and their mathematical foundations.

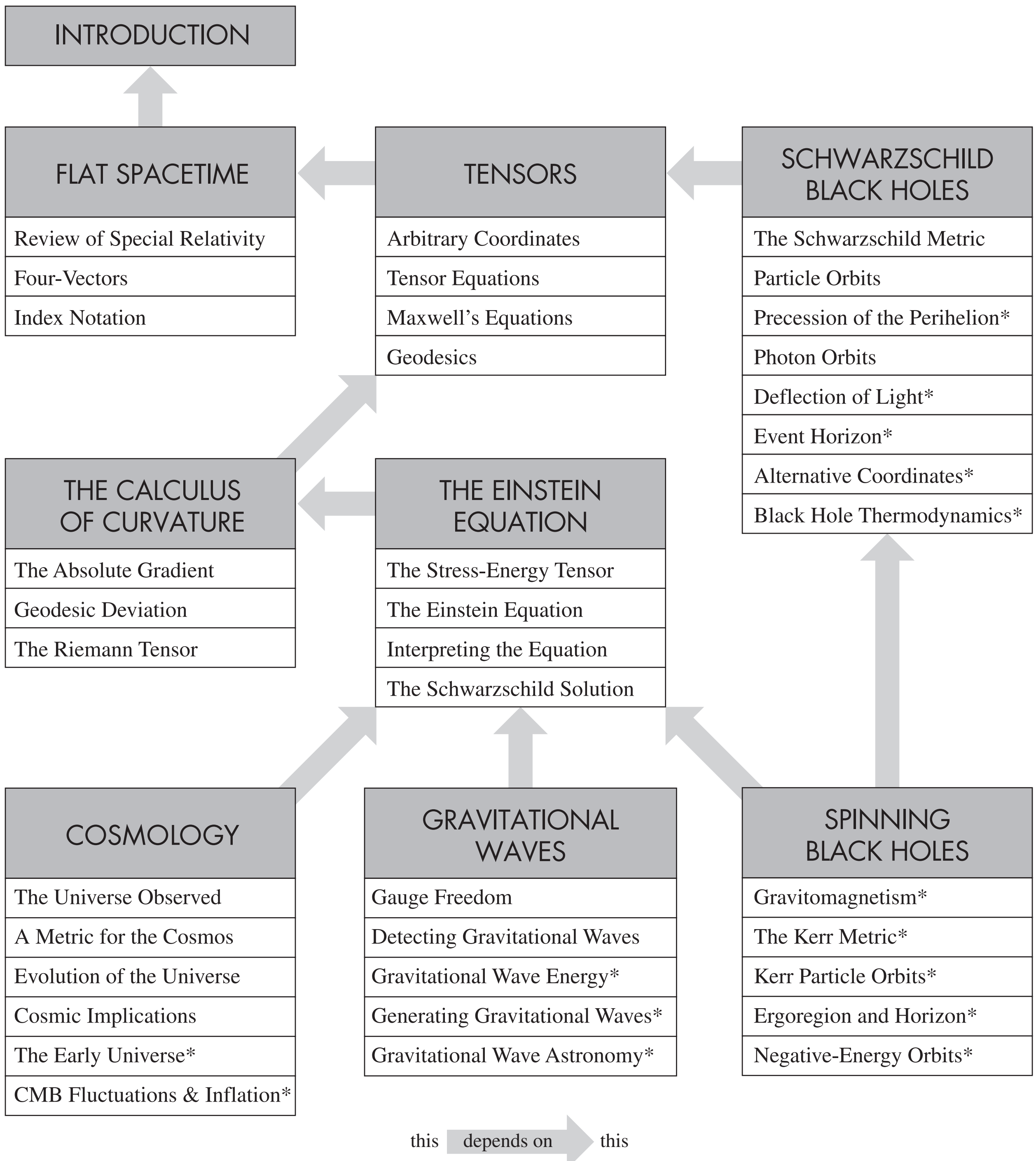
The overview-and-box design also helps keep you focused on the *physics* as opposed to the mathematics, underlining how the mathematics *supports* and *expresses* the physics. Other aspects of the textbook's design also support the principle that the physics should be foremost. I have ordered the topics so that the mathematics is presented not in one big lump but rather gradually and "as needed," thus allowing the physics to drive the presentation. For example, you will extensively practice using tensor notation by exploring real physical applications in flat space before learning about the geodesic equation that describes an object's motion in a curved spacetime. You will then spend a great deal of time exploring the physical implications of the geodesic equation in the particular curved spacetime surrounding a simple spherical object before learning the additional mathematical tools required to show *why* spacetime is curved in that particular way around a spherical object. Along the way, I use many "toy" examples in two-dimensional flat and curved spaces help develop your intuitive understanding of the physical meaning of the core ideas. The gradual development of the mathematics throughout the text also helps ensure that you have time to gain a firm footing for each step before continuing the climb.

The key to using this book successfully is working carefully through all the boxes in this book. Doing this will ultimately provide you with a range of experience and depth of understanding difficult to obtain any other way.

Chapter Dependencies. The chart that appears on each chapter's title page (and on the next page) shows how the major sections of the book depend on each other. For example, you can see from the chart that the **Introduction**, **Flat Space**, and **Tensors** sections (chapters 1 through 8) provide core material that every other section uses. After chapter 8, I strongly recommend going on to the **Schwarzschild Black Holes** section, because this will develop your understanding of how to work with curved spacetimes before having to wrestle with yet more math (and because black holes are fascinating applications of the theory). However, this is not essential; in a short course focused on cosmology, for example, one could go directly on to the **Calculus of Curvature**, **Einstein Equation**, and **Cosmology** sections. Note also that the final three sections (**Cosmology**, **Gravitational Waves**, and **Spinning Black Holes**) are completely independent of each other and can be explored in any order one might choose. However, all three of these sections require the **Calculus of Curvature** and **Einstein Equation** sections.

One also does not have to go all the way through the Schwarzschild section. The last three chapters (on black holes) are only necessary if you also plan to go through the last two chapters of the **Spinning Black Hole**s section (though it is hard to imagine why anyone would want to avoid learning about black holes!). One can easily omit the *Deflection of Light* chapter without loss of continuity. The *Precession of the Perihelion* chapter is necessary background for the *Deflection of Light* chapter, but you could omit both. The first two chapters are required for all of the other chapters in this section, and the fourth chapter on *Photon Orbits* presents a mathematical technique that is employed in certain homework problems throughout the rest of the book, but is only absolutely required for the *Deflection of Light* and the *Black Hole Thermodynamics* chapters.

In the **Cosmology** section, the first four chapters provide core material and should all be included if this section is to be explored at all. The last two chapters, however, are completely optional; you can omit either both or the last, as desired.



A chart showing the chapters of the book grouped in their major sections and how those sections depend on each other. Chapters marked with a * are optional, though later optional chapters typically depend on earlier such chapters.

While in principle it is possible to stop at after the first two chapters in the **Gravitational Wave** section, I think that a discussion of gravitational wave energy and generation is pretty important. I therefore recommend going through at least the first three chapters of this section if you want to explore gravitational waves at all.

One might reasonably elect to explore the *Gravitomagnetism* chapter alone in the **Spinning Black Holes** section, or stop after either the *Kerr Particle Orbits* chapter or the *Ergoregion and Horizon* chapter. However, the chapters in this section *do* need to be discussed in sequence; one cannot easily drop one from the middle.

The First Chapter. Please also note that the *first* chapter has a different structure than the others. After dealing with preliminaries, I usually end the first class session of the course I teach with a 40-minute interactive lecture. For the sake of completeness (and for later reference), I have provided in the first chapter what amounts to a polished transcript of that lecture. This chapter has no boxes because I don't expect my students to have read (or perhaps even own) the book before the first class. To help them track the lecture, I instead give them a two-sided handout that appears as the last two pages of the first chapter.

The Second Chapter. This chapter presents a very terse review of special relativity aimed primarily at students that have already encountered some relativity in a previous course. If you have not seen relativity before, you may find this chapter harder going. Even so, everything you need to know about special relativity for this book is presented there, and if you work through the chapter slowly, and do many of the homework problems, you should be fine. I have also included references to supplemental reading that you may find helpful.

Book Website. You can find a variety of other helpful information and supporting computer software on this textbook's website:

http://pages.pomona.edu/~tmoore/grw/

Please also feel free to email me suggestions, questions, and error notices: my email address is tmoore@pomona.edu.

Information for Instructors. So far in this preface, I have addressed issues of concern to all readers in language directed mostly to students. In the remainder, I want to specifically address issues of interest to instructors who are designing undergraduate courses around this book.

Course Pacing. I have designed the text so that (in my experience) *each chapter can generally be discussed in a single* (50-minute) *class session*, particularly if you use the format for class sessions I describe below. Your mileage may vary (for example, you may need to spend more time on chapter 2 if your students' background in special relativity is weak), but this general rule should help you appropriately pace the course.

You also have a lot flexibility in choosing which chapters to cover and which you might omit: there are at least twenty different chapter sequences that make sense. Be sure to examine thoroughly the section above on **Chapter Dependencies** above before designing a syllabus that omits chapters. However, I find that I can usually get through the entire book in one semester.

Let me emphasize again that the last three sections (**Cosmology**, G**ravitational Waves**, and **Spinning Black Holes**) are independent; you can present them in any order. One of my colleagues likes to end the course with cosmology, which he thinks provides an exciting climax. I have made that section first of the three precisely because I *also* think it is the most important. If I am working through the book sequentially and run out of time, I'd rather do so in the Spinning Black Holes section than omit any of the cosmology material! I also find that students have many other pressures and concerns near the end of the semester, so I tend to schedule material that I consider *less* crucial toward the end. But you can certainly choose what works best for you and your students, and you have lots of flexibility to do so.

How to Spend Class Time. The workbook format will push students to gain mastery *only* if your course design somehow rewards students for filling out the boxes. The last time I taught the course, I asked several students chosen at random each class session to hand me their books, which I subsequently graded for thoughtful *effort* in filling out the boxes since the last time they submitted their book, with special emphasis on the chapter discussed in class that day. Each student's average grade for these random samples counted about 13% of their course grade. I arranged things so that each student was called on about five to six times a semester.

One of my colleagues at a another institution uses a different approach that may be even better. After determining which box exercises seemed easy enough to skip discussion, he then assigns each remaining box exercise to a student in a strict rotation (including himself in the rotation). The student must present the solution in front of the class. This strongly motivates the students to come to class prepared without having to assign a formal grade for preparation, and also makes class time a bit more active than the way I did it. I intend to use this approach myself the next time that I teach the course.

You might find some other approach better than either of these for your students, but I consider it very important when designing a course based on this book to find *some* way of rewarding students for doing work in the boxes before class.

In either of the approaches outlined above, we spend much of the class period discussing the challenges students encountered in going through the boxes. Because students have at least *tried* to work out the boxes before class, they typically bring good questions to the table, questions that directly address the difficulties they are experiencing personally. We are therefore able to spend class time efficiently addressing students' *actual needs*. If we have time (and we often do), I often work some example problems in class, targeted toward either some interesting physics and/or preparing them better to do the homework. In my experience, this approach to using class time is much more effective and efficient than lecturing would be.

I also recommend that you (the instructor) work through all the boxes in an assigned chapter *yourself* before class. (I myself do this every time I offer the course, even though I have worked through all the boxes several times now!) This will help refresh your memory, help isolate any issues that you might need to resolve for yourself before class, and (most importantly) help you anticipate and appreciate the difficulties that students will have with the boxes.

I intentionally designed most of the boxes so that they ask students to prove something, as the primary goal of the boxes is to help students gain ownership of the concepts and derivations discussed in the text. The homework problems are usually much more open-ended, providing opportunities for students to extend the ideas presented in the text, explore physical applications, and even think about new topics. Some of the problems are also designed to provide a basis for class discussion of topics not covered in the main text.

Homework. I typically assign about two homework problems per chapter: this is enough to keep students pretty busy. Homework problems for this class can be pretty challenging, and even the best students may not get them right the first time. Homework-grading schemes that focus *only* on the final results can therefore make students anxious. However, one can devise grading schemes that (1) allow students to engage difficult problems without anxiety, (2) provide them with an opportunity for further learning, and (3) make grading easier for you or your TAs. The "Course Design" section of the book's website provides a link to a page that discusses a scheme for grading homework that I strongly recommend that you consider: it not only encourages students to tackle tough problems without fearing failure, but I can also guarantee that it will save you time grading!

Website for Instructors. I have set up a special, controlled-access website especially for instructors. If you are an instructor that has adopted the book, send me (1) your name, (2) your institution, and (3) how many students are in your course, and I will tell you how to access that site. This site includes complete problem solutions, box solutions, sample tests, and other information that instructors will find helpful but which should not be available to students in an uncontrolled way.

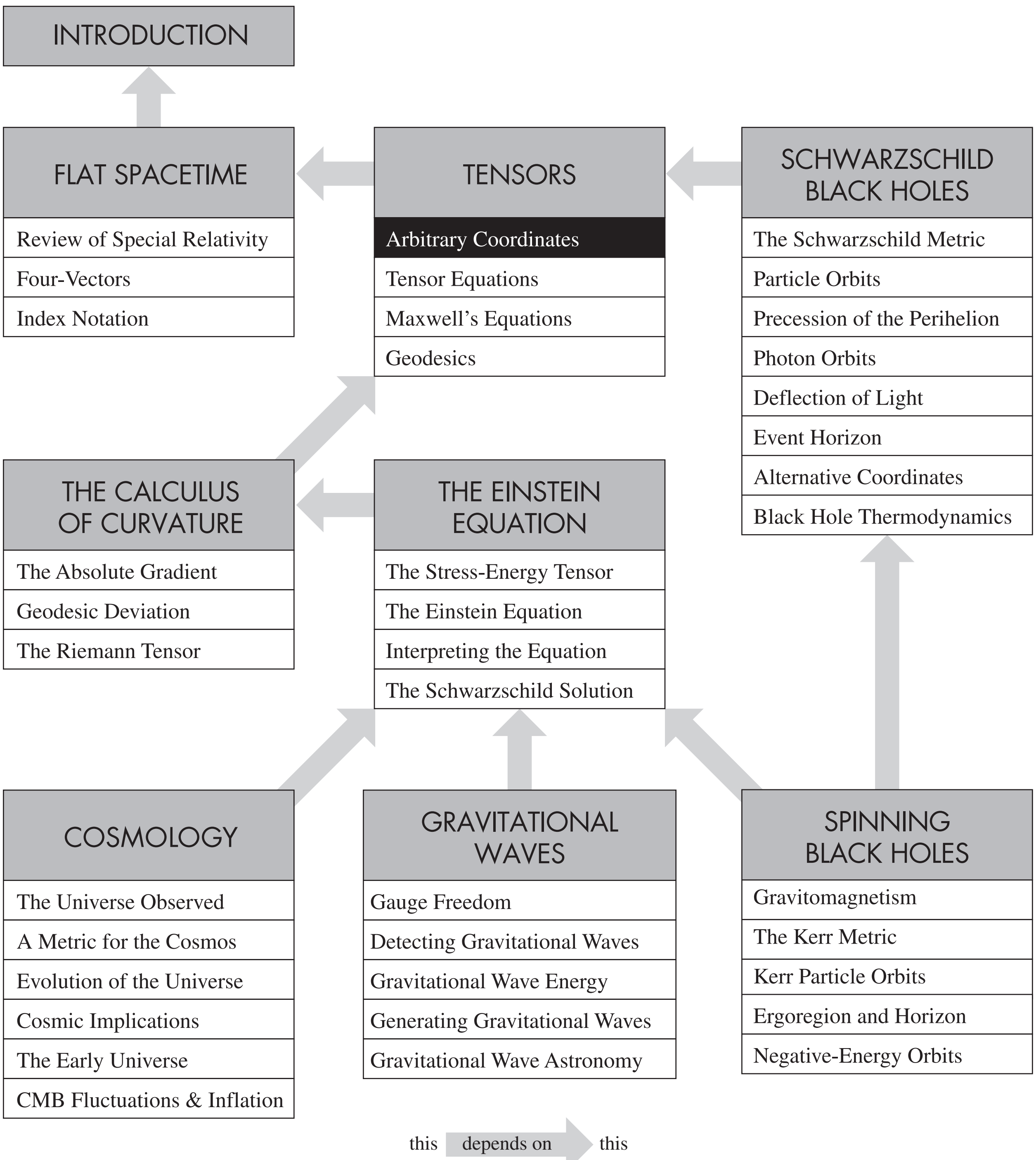
I also welcome emails if you have questions, error notices, or other comments.

Appreciation. I am grateful to many people have helped bring this text to fruition. First, let me thank the students in my Physics 160 class (and particularly Nathan Reed and Ian Frank) for documenting errors in early versions and offering feedback. The overview/box idea grew out of conversations with Dayton Jones (more than thirty years ago). A fruitful correspondence with James Hartle helped me formulate my goals for this book, and his excellent text both taught and inspired me. I am grateful to Edwin Taylor, who has challenged me and broadened my perspective since I was a high school student, for personal support and insight. Thanks also to Tom Baumgarte and Ben Sugerman for trying initial versions of the book in their classes and offering thoughtful feedback. I am grateful to Tom Helliwell, Nandor Bokor, Nelson Christensen and his students (Tom Callister, Ross Cawthon, Andrew Chael, Micah Koller, Dustin Anderson, and David Miller, who all sent me individual reviews), Tom Baumgarte, Tom Carroll, Bryan van der Ende, and an unknown reviewer for reading a nearly final draft and offering a number of valuable suggestions and error corrections. Needless to say, any remaining errors are my own. I want to thank the developers of MathMagic (my equation-editing software) for extraordinary attention and help beyond the call of duty when I encountered various problems. Hilda Dinolfo and Christine Maynard were very helpful in printing copies for various early readers. I am very grateful to Sergio Picozzi and John Mallinckrodt for carefully reviewing the final draft and being willing to write such nice endorsements for the back cover. I want to thank the book's production team, (Lee Young, Richard Camp, Yvonne Tsang, Genette Itako McGrew, and especially Laurel Muller and Paul Anagnostopoulos) for their excellent work, care, and extraordinary patience in dealing with a difficult book (and a sometimes difficult author). I also want to thank Jane Ellis, my managing editor at University Science Books, for her support, enthusiasm, and hard work in bringing this book to print, and for being willing to take a risk on a book that was a bit out of the ordinary.

Finally, let me thank my wife Joyce, whose unfailing support for my writing habit is loving and gracious beyond the call of duty. I am very grateful to all!

Thomas A. Moore
Claremont, CA
July 11, 2012

5. ARBITRARY COORDINATES



Introduction to Arbitrary Coordinates. As we saw at the beginning of the course, general relativity tells us that gravity results from curved spacetime. We have seen in the past few chapters how to describe the flat spacetime of special relativity using cartesian spatial coordinates and a time coordinate defined by synchronized clocks in an inertial frame. But in curved spacetimes, we cannot use cartesian coordinates. Moreover, since our eventual goal is to calculate *how* matter curves the spacetime around it, we often do not know the spacetime's geometry *a priori*, and therefore do not know what kind of coordinate system to use.

Our goal in the next few chapters is to develop mathematical techniques for writing physical equations in a way that is completely independent of the coordinate system we actually end up using. This will generalize the principle of relativity: we will learn how to express the laws of physics in a way that is not only independent of our choice of inertial reference frame, but in fact entirely independent of our choice of coordinates!

A **coordinate system** is ultimately simply some kind of organized scheme for attaching numbers (**coordinates**) to points in space and/or events in spacetime. The clock-lattice scheme that we considered in chapter 2 is one way, but by no means the only way, to attach coordinates to events. The *only* assumptions that we will make here about our coordinate systems are that (1) our space is not so horribly curved that we cannot treat a sufficiently small patch of it as if it were flat, and (2) our coordinates vary smoothly so that neighboring points have nearly the same coordinates.

To make things simple and easy to visualize, we will in this chapter be primarily working with arbitrary coordinates in a flat two-dimensional (2D) space. However, the *methods* we develop for handling arbitrary coordinates will end up working just as well for curved spaces in any number of dimensions.

No matter how we construct our coordinate system, the distance ds between two infinitesimally separated points is a coordinate-independent quantity, because we can measure it directly with a ruler without having to define a coordinate system at all. The fundamental way that we connect arbitrary coordinates to physical reality is by specifying how the distance between two infinitesimally-separated points depends on their coordinate separations. A **cartesian** x, y coordinate system is one in which the distance ds between two infinitesimally-separated points is given by $ds^2 = dx^2 + dy^2$ everywhere in the 2D plane. A **curvilinear** coordinate system is any non-cartesian coordinate system where this simple Pythagorean relationship is not true. How can we connect the coordinate-independent distance between two points with their coordinate separations in such a case?

Definition of a Coordinate Basis. Consider arbitrary coordinates u, w for a 2D space. When using index notation, we will interpret dx^u as being equivalent to du, and dx^w as being equivalent to dw, and we will assume that Greek indices have two possible values u and w. (In the last chapter, in the context of cartesian coordinates in flat spacetime, I stated that indices could represent either t, x, y, or z, but when we use arbitrary coordinates, the indices represent whatever the index names might be.) I will also represent 2D vectors with the same bold-face notation as we used for four-vectors in the previous chapters. This will keep the notation from changing when we generalize to 4D spacetimes.

Now, no matter how our u, w coordinate system is defined, at each point $\mathcal{P}$ in the space, we can define a pair of basis vectors $\boldsymbol{e}_u, \boldsymbol{e}_w$ such that

1. $\boldsymbol{e}_u$ points tangent to the w = constant curve toward increasing u.
2. $\boldsymbol{e}_w$ points tangent to the u = constant curve toward increasing w.
3. The lengths of $\boldsymbol{e}_u, \boldsymbol{e}_w$ are defined so that the displacement vector $d\boldsymbol{s}$ between the point $\mathcal{P}$ at coordinates u, w and any infinitesimally separated neighboring point $\mathcal{P}$ at coordinates $u + du$, $w + dw$ can be written

$$d\boldsymbol{s} = du\,\boldsymbol{e}_u + dw\,\boldsymbol{e}_w = dx^\mu\,\boldsymbol{e}_\mu \tag{5.1}$$

Figure 5.1 illustrates how these basis vectors are defined.

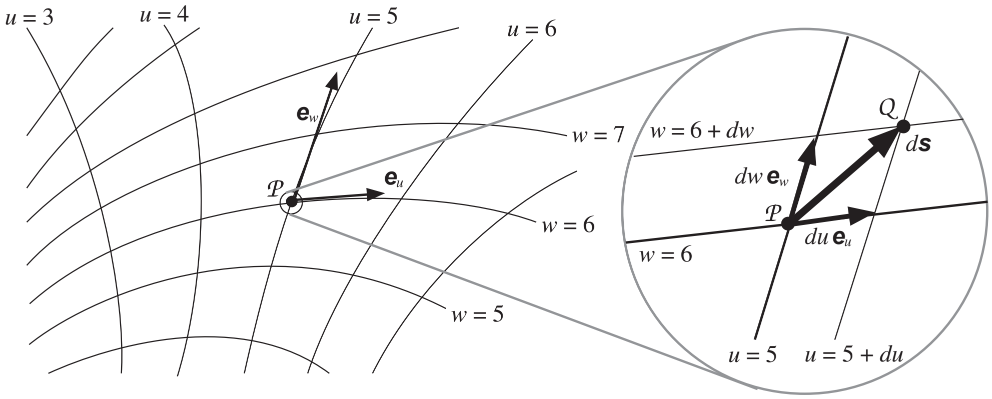


FIG. 5.1 This drawing shows an arbitrary coordinate system, a point $\mathcal{P}$, the basis vectors $\boldsymbol{e}_u$ and $\boldsymbol{e}_w$ at that point, and a close-up view of how we describe an infinitesimal displacement $d\boldsymbol{s}$ as a sum of the basis vectors multiplied by the corresponding changes in the coordinate values. Ensuring that $dw\,\boldsymbol{e}_w$ and $du\,\boldsymbol{e}_u$ add up to the actual displacement $d\boldsymbol{s}$ defines the lengths of $\boldsymbol{e}_u$ and $\boldsymbol{e}_w$.

If we define basis vectors this way, then du and dw become the components of $d\boldsymbol{s}$ in that basis, and we call the set of basis vectors $\boldsymbol{e}_u, \boldsymbol{e}_w$ a **coordinate basis**. We do not *have* to define the basis vectors this way, but it proves *very* convenient, as we will see. A coordinate basis is generally *different* than the cartesian coordinate basis vectors $\boldsymbol{e}_x, \boldsymbol{e}_y$ (more commonly written **i**, **j** or $\hat{x}, \hat{y}$) in that (1) $\boldsymbol{e}_u \cdot \boldsymbol{e}_w$ may be nonzero, (2) $\boldsymbol{e}_u$ and $\boldsymbol{e}_w$ may not have unit length, and (3) $\boldsymbol{e}_u$ and $\boldsymbol{e}_w$ may change in magnitude and/or direction as one moves from point to point.

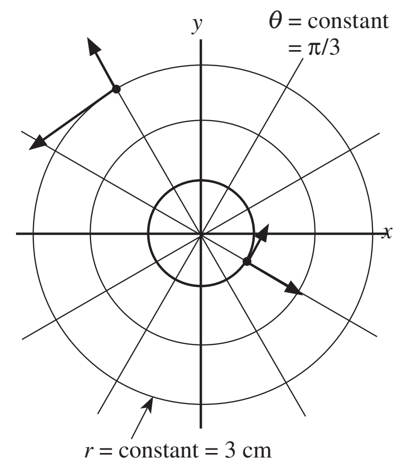


FIG. 5.2 Coordinate basis vectors for a polar coordinate system.

An Example. Consider r and θ coordinates for a flat 2D space. The coordinate basis for this coordinate system consists of the vectors $\boldsymbol{e}_r$ and $\boldsymbol{e}_\theta$, whose directions vary from point to point ($\boldsymbol{e}_r$ pointing radially away from the origin, and $\boldsymbol{e}_\theta$ perpendicular to it) and whose magnitudes are given by $\text{mag}(\boldsymbol{e}_r) = 1$ and $\text{mag}(\boldsymbol{e}_\theta) = r$ (see figure 5.2). This definition ensures that we can write $d\boldsymbol{s} = dr\,\boldsymbol{e}_r + d\theta\,\boldsymbol{e}_\theta$. If we were to use conventional polar-coordinate *unit* vectors $\boldsymbol{e}_{\hat{r}}$ and $\boldsymbol{e}_{\hat{\theta}}$, which both have unit magnitude by definition, we would have to write $d\boldsymbol{s} = dr\boldsymbol{e}_{\hat{r}} + rd\theta\,\boldsymbol{e}_{\hat{\theta}}$ instead. Therefore, the conventional polar-coordinate basis vectors $\boldsymbol{e}_{\hat{r}}$ and $\boldsymbol{e}_{\hat{\theta}}$ do *not* comprise a "coordinate basis." This example is discussed more fully in box 5.1.

General Vectors. Once we have established a coordinate basis, then we can *define* the components A^u, A^w of an arbitrary vector $\boldsymbol{A}$ at the point $\mathcal{P}$ so that

$$\boldsymbol{A} \equiv A^\mu \boldsymbol{e}_\mu = A^u \boldsymbol{e}_u + A^w \boldsymbol{e}_w \tag{5.2}$$

The Metric Tensor. The scalar product of $d\boldsymbol{s}$ with itself is the square of the physical distance between the endpoints of $d\boldsymbol{s}$:

$$\begin{aligned} ds^2 &= d\boldsymbol{s} \cdot d\boldsymbol{s} = (du\,\boldsymbol{e}_u + dw\,\boldsymbol{e}_w) \cdot (du\,\boldsymbol{e}_u + dw\,\boldsymbol{e}_w) \\ &= du^2\,\boldsymbol{e}_u \cdot \boldsymbol{e}_u + du\,dw\,\boldsymbol{e}_u \cdot \boldsymbol{e}_w + dw\,du\,\boldsymbol{e}_w \cdot \boldsymbol{e}_u + dw^2\,\boldsymbol{e}_w \cdot \boldsymbol{e}_w \\ &= dx^\alpha\,dx^\beta\,\boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta \equiv g_{\alpha\beta}\,dx^\alpha\,dx^\beta \end{aligned} \tag{5.3}$$

The set of four components $g_{\alpha\beta} \equiv \boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta$ comprises the **metric tensor** for our 2D coordinate basis. The equation above therefore specifies the relationship between the coordinate separations and the physical distance between two points, and represents a generalization of the Pythagorean theorem for our arbitrary coordinate system. Note that $g_{\alpha\beta}$ is generally a function of position in space.

This is the generalization of the metric tensor $\eta_{\alpha\beta}$ introduced in the last section. In *flat* spacetime, we can always find a cartesian coordinate basis where the basis vectors are mutually orthogonal ($\boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta = 0$ when $\alpha \neq \beta$) and have unit magnitude ($\boldsymbol{e}_\alpha \cdot \boldsymbol{e}_\beta = \pm 1$ when $\alpha = \beta$, with the negative value indicating a time coordinate) at all events. This is *not* generally possible in a curved spacetime.

Transformation of Coordinates. Now consider a general coordinate transformation in two dimensions between coordinates u, w and new coordinates $p(u, w)$ and $q(u, w)$. The chain rule for partial derivatives implies that infinitesimal changes in the new coordinates are related to changes in the old coordinates as follows:

$$dp = \frac{\partial p}{\partial u} du + \frac{\partial p}{\partial w} dw \quad \text{and} \quad dq = \frac{\partial q}{\partial u} du + \frac{\partial q}{\partial w} dw \tag{5.4}$$

If we consider the p, q coordinate system the primed coordinates and u, w the unprimed coordinates, then we can write this compactly in index notation as follows:

$$dx'^{\mu} = \frac{\partial x'^{\mu}}{\partial x^{\nu}} dx^{\nu} \tag{5.5}$$

(with an implicit sum over the ν index, since we will consider the superscript in the *denominator of a derivative* to be equivalent to a subscript).

But if we use a coordinate basis in both coordinate systems, then dp and dq are the *actual components* of the infinitesimal displacement $d\boldsymbol{s}$ in the primed system, and du and dw play the same role in the unprimed system. Since by definition the components of an arbitrary vector $\boldsymbol{A}$ must transform in the same way that the components of the displacement do, the components of $\boldsymbol{A}$ must transform as follows:

$$A'^{\mu} = \frac{\partial x'^{\mu}}{\partial x^{\nu}} A^{\nu} \tag{5.6}$$

This is the *general* transformation law for the components of a vector when we are using a coordinate basis. The simplicity of this transformation law is precisely why coordinate bases are so useful. From this point on, we will *assume* that we will use coordinate bases unless we explicitly state otherwise.

It follows from the argument above that the reverse transformation of vector components from the primed to the unprimed system is simply

$$A^{\mu} = \frac{\partial x^{\mu}}{\partial x'^{\nu}} A'^{\nu} \tag{5.7}$$

An Important Identity. Basic partial differential calculus implies that

$$\frac{\partial x'^{\mu}}{\partial x^{\alpha}} \frac{\partial x^{\alpha}}{\partial x'^{\nu}} = \delta^{\mu}{}_{\nu} \tag{5.8}$$

If we write this out explicitly for our p, q and u, w coordinate systems, this says that

$$\frac{\partial p}{\partial u}\frac{\partial u}{\partial p} + \frac{\partial p}{\partial w}\frac{\partial w}{\partial p} = \frac{dp}{dp} = 1, \quad \frac{\partial p}{\partial u}\frac{\partial u}{\partial q} + \frac{\partial p}{\partial w}\frac{\partial w}{\partial q} = \frac{dp}{dq} = 0, \tag{5.9a}$$

$$\frac{\partial q}{\partial u}\frac{\partial u}{\partial p} + \frac{\partial q}{\partial w}\frac{\partial w}{\partial p} = \frac{dq}{dp} = 0, \quad \frac{\partial q}{\partial u}\frac{\partial u}{\partial q} + \frac{\partial q}{\partial w}\frac{\partial w}{\partial q} = \frac{dq}{dq} = 1 \tag{5.9b}$$

We will find this identity *very* useful in what follows.

The Transformation of the Metric Tensor. The fact that the magnitude of $d\boldsymbol{s}$ is a frame-independent quantity by definition directly implies that

$$ds^2 = g'_{\mu\nu}\, dx'^{\mu}\, dx'^{\nu} = g_{\alpha\beta}\, dx^{\alpha}\, dx^{\beta} \tag{5.10}$$

This directly implies (see box 5.2) that

$$g'_{\mu\nu} = \frac{\partial x^{\alpha}}{\partial x'^{\mu}} \frac{\partial x^{\beta}}{\partial x'^{\nu}} g_{\alpha\beta} \quad \text{and} \quad g_{\alpha\beta} = \frac{\partial x'^{\mu}}{\partial x^{\alpha}} \frac{\partial x'^{\nu}}{\partial x^{\beta}} g'_{\mu\nu} \tag{5.11}$$

(Note again that we assume there to be implicit sums over the α and β indices in the first equation and the μ and ν indices in the second.)

Equations 5.11 provides a handy way to find the components of the metric tensor in a new coordinate system if you know how the new coordinates depend on the old coordinates and you know the metric tensor in latter system. Box 5.3 illustrates such a calculation for a simple 2D coordinate system.

Coordinate Transformations in Flat Spacetime. The cartesian-like coordinates that we use for an inertial reference frame in the flat spacetimes of special relativity are an example of a coordinate basis. The Lorentz transformation equations in fact represent a special case of the general transformation rule given above. To see this, consider the full coordinate transformations between two inertial frames in standard orientation and where the primed frame is moving with speed β in the $+x$ direction relative to the unprimed frame are

$$\begin{aligned} t' &= \gamma(t - \beta x) \\ x' &= \gamma(-\beta t + x) \\ y' &= y \\ z' &= z \end{aligned} \quad \text{and} \quad \begin{aligned} t &= \gamma(t' + \beta x') \\ x &= \gamma(\beta t' + x') \\ y &= y' \\ z &= z' \end{aligned} \tag{5.12}$$

By taking the partial derivatives of these equations, you can easily show that

$$\frac{\partial x'^{\mu}}{\partial x^{\nu}} = \Lambda^{\mu}{}_{\nu} = \begin{bmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.13a}$$

$$\frac{\partial x^{\mu}}{\partial x'^{\nu}} = (\Lambda^{-1})^{\mu}{}_{\nu} = \begin{bmatrix} \gamma & \gamma\beta & 0 & 0 \\ \gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{5.13b}$$

(See box 5.4.) Because the transformation equations 5.12 are linear, the partial derivatives in equations 5.13 are *constant*, which is *not* generally the case for arbitrary coordinate transformations.

Equations 5.13 imply that for cartesian-like coordinates in flat spacetime, the general transformation law for the components of an arbitrary four-vector $\boldsymbol{A}$ given in equation 5.6 is the same as the Lorentz transformation law we saw earlier:

$$A'^{\mu} = \frac{\partial x'^{\mu}}{\partial x^{\nu}} A^{\nu} = \Lambda^{\mu}{}_{\nu} A^{\nu} \tag{5.14}$$

(compare with equation 4.5). We also know that Lorentz transformations obey the identity specified in equation 5.8,

$$(\Lambda^{-1})^{\mu}{}_{\alpha} \Lambda^{\alpha}{}_{\nu} = \frac{\partial x'^{\mu}}{\partial x^{\alpha}} \frac{\partial x^{\alpha}}{\partial x'^{\nu}} = \delta^{\mu}{}_{\nu} \tag{5.15}$$

(compare with equation 4.12), and that the transformation law for the metric tensor of flat spacetime is

$$\eta'_{\alpha\beta} = \frac{\partial x^{\mu}}{\partial x'^{\alpha}} \frac{\partial x^{\nu}}{\partial x'^{\beta}} \eta_{\mu\nu} = \eta_{\mu\nu} (\Lambda^{-1})^{\mu}{}_{\alpha} (\Lambda^{-1})^{\nu}{}_{\beta} = \eta_{\alpha\beta} \tag{5.16}$$

according to equation 4.19. This means that the components of the metric tensor for flat spacetime have *the same numerical value* in all cartesian-like coordinate systems connected by Lorentz transformations. You can check that equation 5.16 is true component by component (see box 5.5).

The Metric for a Spherical Surface. In curved spaces and spacetimes, we are stuck with curvilinear coordinates. As an example, box 5.6 discusses the curvilinear θ, ϕ (latitude-longitude) coordinate system for the 2D surface of a sphere of radius R. We see there that the metric for this space in this coordinate system is

$$g_{\mu\nu} = \begin{bmatrix} g_{\theta\theta} & g_{\theta\phi} \\ g_{\phi\theta} & g_{\phi\phi} \end{bmatrix} = \begin{bmatrix} R^2 & 0 \\ 0 & R^2 \sin^2\theta \end{bmatrix} \tag{5.17}$$

This result will be very valuable to us later as a simple example of a curved space and when we seek to construct metrics for spherically symmetric spacetimes.

BOX 5.1 The Polar Coordinate Basis

Consider ordinary polar coordinates r and θ (see figure 5.3). Note that the distance between two points with the same r coordinate but separated by an infinitesimal step $d\theta$ in θ is $r\,d\theta$ (by the definition of angle). So there are (at least) two ways to define a basis vector for the θ direction (which we define to be tangent to the r = constant curve): (1) we could define a basis vector $\mathbf{e}_{\hat\theta}$ with a unit magnitude, in which case the differential displacement vector for the step we are considering would be $d\mathbf{s} = r\,d\theta\,\mathbf{e}_{\hat\theta}$, or (2) we can define a basis vector ds with magnitude r, so that we can write $d\mathbf{s} = d\theta\,\mathbf{e}_\theta$. In each case, the magnitude of the displacement will be $r\,d\theta$, but in the second case, the coordinate change $d\theta$ itself becomes the component of $d\mathbf{s}$, which is convenient. This latter choice is the one that defines the "coordinate basis" vector for the θ direction in polar coordinates.

The length of an infinitesimal step dr in the r direction (tangent to the θ = constant curve) is simply dr, so if we define $\mathbf{e}_r$ to have unit magnitude, we have $d\mathbf{s} = dr\,\mathbf{e}_r$ for such a step. Here, the basis vector with unit length is (in this case) the appropriate choice for a "coordinate basis" vector in the r direction.

Once we have established these basis vectors, we can write the components of an arbitrary infinitesimal displacement in any direction as

$$d\mathbf{s} = dr\,\mathbf{e}_r + d\theta\,\mathbf{e}_\theta \tag{5.18}$$

Note carefully that this equation does not apply to finite displacements, but only displacements small enough so that the basis vectors $\mathbf{e}_r$ and $\mathbf{e}_\theta$ do not change significantly over the distance spanned by the displacement. (See the exercise below.)

Note that by the nature of polar coordinates, basis vectors that point tangent to the θ = constant and r = constant curves are perpendicular to each other at all points, but $\mathbf{e}_r$ (for example) does not point in the same direction at one point as it does at another, as illustrated in figure 5.3. The metric for the polar coordinate basis is

$$g_{\mu\nu} \equiv \mathbf{e}_\mu \cdot \mathbf{e}_\nu = \begin{bmatrix} \mathbf{e}_r \cdot \mathbf{e}_r & \mathbf{e}_r \cdot \mathbf{e}_\theta \\ \mathbf{e}_\theta \cdot \mathbf{e}_r & \mathbf{e}_\theta \cdot \mathbf{e}_\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix} \tag{5.19}$$

Important note: We can always specify the components of a metric tensor either by listing them in a matrix (as above) or by writing out the metric equation. For example, if we compare the abstract and concrete versions of the metric equation

$$ds^2 = g_{\mu\nu}\,dx^\mu\,dx^\nu = dr^2 + r^2\,d\theta^2 \tag{5.20}$$

we can immediately infer that $g_{rr} = 1$, $g_{\theta\theta} = r^2$, and $g_{r\theta} = g_{\theta r} = 0$ (because terms involving $dr\,d\theta$ and $d\theta\,dr$ do not appear). The latter approach is often very convenient.

Exercise 5.1.1. By drawing on the diagram below, show that the displacement $\Delta\mathbf{s}$ between the two points with coordinates of r = 1 cm, $\theta = 0°$ and r = 2 cm, $\theta = 90°$ is *not* accurately given by equation 5.18 (because it is not infinitesimal).

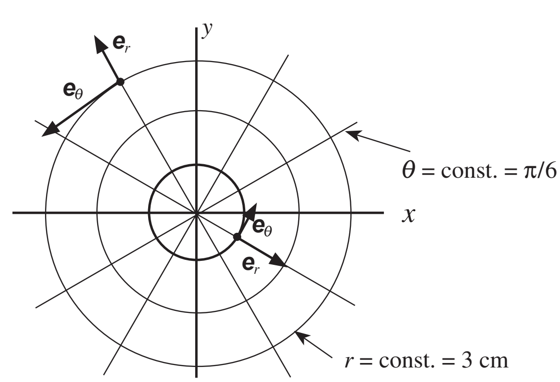


FIG. 5.3. A diagram that displays the r = constant and θ = constant curves for polar coordinates and the polar coordinate basis vectors at selected points.

BOX 5.2 Proof of the Metric Transformation Law

One can prove equation 5.11 as follows. Start with equation 5.10, repeated here for convenience:

$$g'_{\mu\nu}\, dx'^{\mu}\, dx'^{\nu} = g_{\alpha\beta}\, dx^{\alpha}\, dx^{\beta} \tag{5.10r}$$

Use the inverse transformation law for the components of an infinitesimal displacement (equation 5.7) to rewrite the above as

$$g'_{\mu\nu}\, dx'^{\mu}\, dx'^{\nu} = g_{\alpha\beta} \frac{\partial x^{\alpha}}{\partial x'^{\mu}} \frac{\partial x^{\beta}}{\partial x'^{\nu}}\, dx'^{\mu}\, dx'^{\nu} \tag{5.21}$$

Now you can follow the mode of argument used in box 4.6 to show that

$$g'_{\mu\nu} = g_{\alpha\beta} \frac{\partial x^{\alpha}}{\partial x'^{\mu}} \frac{\partial x^{\beta}}{\partial x'^{\nu}} \tag{5.22}$$

This is equation 5.11.

Exercise 5.2.1. Fill in the gap between equation 5.21 and 5.22. Note that because of the implicit sums in equation 5.21, this is more complicated than saying "divide both sides by $dx'^{\mu} dx'^{\nu}$"!

BOX 5.3 A 2D Example: Parabolic Coordinates

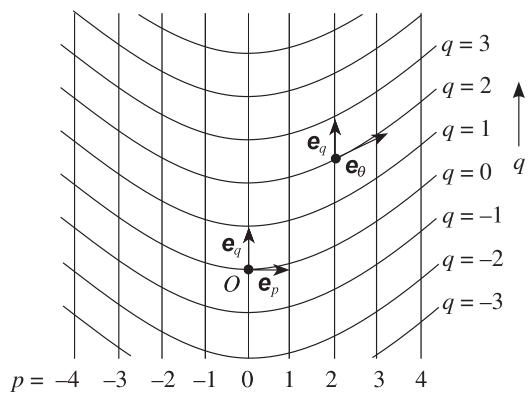


FIG. 5.4. A diagram that displays the p = constant and q = constant curves for parabolic coordinates and the parabolic coordinate basis vectors at selected points.

Consider the parabolic coordinate system p, q shown in figure 5.4. The transformation functions from ordinary cartesian coordinates x, y to these coordinates are

$$p(x,y) = x \quad \text{and} \quad q(x,y) = y - cx^2 \tag{5.23}$$

where c is a constant. The inverse transformation functions are

$$x(p,q) = p \quad \text{and} \quad y(p,q) = cp^2 + q \tag{5.24}$$

Exercise 5.3.1. Show that equations 5.24 are the correct inverse transformations.

Exercise 5.3.2. Calculate all eight partial derivatives $\partial x'^{\mu}/\partial x^{\nu}$ and $\partial x^{\mu}/\partial x'^{\nu}$.

The metric equation for the cartesian coordinates x, y is $ds^2 = dx^2 + dy^2$, so the metric tensor for these coordinates must be

$$g_{\alpha\beta} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{5.25}$$

You can use equation 5.11 to show that the metric for the p, q system is

$$g'_{\mu\nu} = \begin{bmatrix} 1+4c^2p^2 & 2cp \\ 2cp & 1 \end{bmatrix} \tag{5.26}$$

BOX 5.3 (continued) A 2D Example: Parabolic Coordinates

For example, if we choose the coordinate indices $\mu = p$, $\nu = q$, we see that

$$g_{pq} = \frac{\partial x^\alpha}{\partial p}\frac{\partial x^\beta}{\partial q} g_{\alpha\beta} = \frac{\partial x}{\partial p}\frac{\partial x}{\partial q} g_{xx} + \frac{\partial x}{\partial p}\frac{\partial y}{\partial q} g_{xy} + \frac{\partial y}{\partial p}\frac{\partial x}{\partial q} g_{yx} + \frac{\partial y}{\partial p}\frac{\partial y}{\partial q} g_{yy}$$

$$= 1 \cdot 0 \cdot 1 + 0 + 0 + 2cp \cdot 1 = 2cp \qquad (5.27)$$

Exercise 5.3.3. Use the same technique to verify the other components of equation 5.26. Does the fact that this metric has off-diagonal components make sense?

Exercise 5.3.4. Let a vector $\boldsymbol{A}$ have p, q components $A^p = 1, A^q = 0$.

a) Find this vector's components in the x, y coordinate system.
b) Do these components make sense? (*Hint:* Sketch $\mathbf{e}_p, \mathbf{e}_q$ at a typical point.)
c) Show that $A^2 = \boldsymbol{A} \cdot \boldsymbol{A}$ of this vector has the same value in both systems.

BOX 5.4 The LTEs as an Example General Transformation

Notice that for the Lorentz transformation

$$\frac{\partial x'^t}{\partial x^t} = \frac{\partial t'}{\partial t} = \frac{\partial}{\partial t}\gamma(t - \beta x) = \gamma + 0 = \gamma = \Lambda^t{}_t \tag{5.28}$$

Exercise 5.4.1. Similarly, check that $\partial x'^\mu / \partial x^\nu = \Lambda^\mu{}_\nu$ when $\mu = x$ and $\nu = t$, and when $\mu = \nu = y$.

BOX 5.5 The Metric Transformation Law in Flat Space

Let's check equation 5.16 for $\alpha = \beta = t$.

$$\begin{aligned}\eta'_{tt} &= (\Lambda^{-1})^\mu{}_t(\Lambda^{-1})^\nu{}_t\eta_{\mu\nu} \\ &= (\Lambda^{-1})^t{}_t(\Lambda^{-1})^\nu{}_t\eta_{t\nu} + (\Lambda^{-1})^x{}_t(\Lambda^{-1})^\nu{}_t\eta_{x\nu} \\ &\quad + (\Lambda^{-1})^y{}_t(\Lambda^{-1})^\nu{}_t\eta_{y\nu} + (\Lambda^{-1})^z{}_t(\Lambda^{-1})^\nu{}_t\eta_{z\nu}\end{aligned} \tag{5.29}$$

Now, $\eta_{t\nu}$ is only nonzero when $\nu = t$, $\eta_{x\nu}$ only when $\nu = x$, and so on. Moreover $(\Lambda^{-1})^y{}_t = (\Lambda^{-1})^z{}_t = 0$. Therefore,

$$\begin{aligned}\eta'_{tt} &= (\Lambda^{-1})^t{}_t(\Lambda^{-1})^t{}_t\eta_{tt} + (\Lambda^{-1})^x{}_t(\Lambda^{-1})^x{}_t\eta_{xx} \\ &= \gamma^2(-1) + (-\gamma\beta)^2(+1) = \frac{-1+\beta^2}{1-\beta^2} = -1 = \eta_{tt}\end{aligned} \tag{5.30}$$

The other components are analogous.

Exercise 5.2.1. Check the cases where $\alpha = t$ and $\beta = x$, and where $\alpha = \beta = x$.

BOX 5.6 A Metric for a Sphere

Consider the 2D surface of a sphere of radius R. The most commonly used coordinate system for a spherical surface is a latitude-longitude system using angular coordinates θ and ϕ. As illustrated in figure 5.5, curves of constant longitude ϕ are great circles that intersect at both poles. The coordinate ϕ labels these curves by the angle each makes at the north pole with a longitude curve arbitrarily chosen to have $\phi = 0$ (in the case of the earth's surface, the great circle going through Greenwich, England). The curves of constant latitude are circles a constant distance from the pole (as measured along the sphere's surface). On the earth's surface, we conventionally label these circles by the angle θ that a line drawn from any point on the circle to the earth's center makes with the earth's equatorial plane (so that the equator has $\theta = 0$). However, in physics, we usually define θ to be the angle measured down from the north pole, so that $\theta = 0$ at the north pole, $\pi/2$ at the equator, and π at the south pole. I will use the physics definition throughout this book.

A nice feature of this coordinate system is that the curves of longitude and latitude are always perpendicular to each other. This means that $g_{\theta\phi} = g_{\phi\theta} = \boldsymbol{e}_\theta \cdot \boldsymbol{e}_\phi = 0$, i.e., the matrix for this coordinate system's metric is diagonal. We can determine the other metric components as follows. Consider first the infinitesimal displacement corresponding to an infinitesimal change in latitude $d\theta$ along a curve of constant longitude. Since that curve is a great circle, its radius is R, so the arclength along the sphere's surface subtended by the angle $d\theta$ is $R\,d\theta$. Similarly, since the diagram shows that a circle of latitude θ has a radius of $R\sin\theta$, the length of the infinitesimal displacement corresponding to an infinitesimal change $d\phi$ along a circle of constant latitude must have a length $R\sin\theta\,d\phi$. Because these displacements are perpendicular, and because in the infinitesimal limit, the patch of area spanned by these displacements is almost flat, we can use the Pythagorean theorem to determine the squared length of the displacement $d\boldsymbol{s}$ that is the sum of such displacements:

$$ds^2 = (R\,d\theta)^2 + (R\sin\theta\,d\phi)^2 = R^2 d\theta^2 + R^2\sin^2\theta\,d\phi^2 \tag{5.31}$$

Comparing this to the abstract form of the metric equation $ds^2 = g_{\mu\nu}\,dx^\mu\,dx^\nu$, we see that $g_{\theta\theta} = R^2$, $g_{\phi\phi} = R^2\sin^2\theta$, and $g_{\phi\theta} = g_{\theta\phi} = 0$.

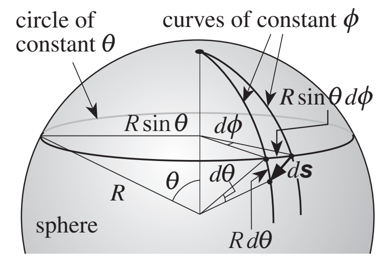


FIG. 5.5: A drawing of the surface of a sphere, showing curves of constant longitude ϕ and latitude θ, and an infinitesimal displacement $d\boldsymbol{s}$ comprised of infinitesimal steps in both the θ and ϕ directions.

Exercise 5.6.1. What would the metric components be if we were to measure θ up from the equator rather than down from the pole?

HOMEWORK PROBLEMS

P5.1 Consider the coordinate basis discussed in box 5.1 for polar coordinates r, θ.

a. Find the transformation equations that take one from 2D cartesian coordinates x, y to r, θ and vice versa.

b. Evaluate the partial derivatives of x and y with respect to r and θ and vice versa.

c. Find the metric for polar coordinates by transforming the metric from cartesian coordinates. Is your result consistent with equation 5.19?

P5.2 (Do problem P5.1 first.) Consider the coordinate basis discussed in box 5.1 for polar coordinates r, θ.

a. We know that an object in uniform circular motion has a constant radius, so it must have a velocity $\boldsymbol{v}$ such that $v^r = 0$. Use the polar-coordinate metric to show that if we assume this velocity has a constant (and coordinate-system-independent!) squared magnitude of $\boldsymbol{v} \cdot \boldsymbol{v} = v^2$, then we must have $v^\theta = \pm v/r$ (where the sign depends on which way the object moves around the circle).

b. Find this object's velocity components v^x and v^y in cartesian coordinates. Express your result as a function of r and θ. Use a sketch to show that these components do indeed describe a vector tangent to a circle of radius r, and also show that the squared magnitude of this vector in the cartesian system is indeed v^2.

P5.3 (Do problem 5.1 first.) Consider the coordinate basis discussed in box 5.1 for polar coordinates r, θ.

a. Consider an object moving at a constant speed v in the $+y$ direction of the cartesian coordinate system, so that $v^y = v$, $v^x = 0$. Find the components v^r and v^θ of this object's velocity in the polar coordinate system. Express your results both purely in terms of r and θ and purely in terms of x and y.

b. Imagine that the object starts at $x = b$, $y = 0$ at time $t = 0$. Its subsequent y position at later times t will therefore be simply $y = vt$. Use this to express both the object's r and θ position and its polar coordinate velocity components v^r and v^θ at all times $t > 0$ in terms of v, b, r, and t. Does your result make sense? (In particular, if you sketch the object's path, you should be able to see that its velocity will be mostly in the θ direction at early times, but mostly in the r direction at late times. Is this consistent with your mathematical expressions?)

P5.4 We can define "semilog" coordinates p, q for a flat 2D plane by the relations $p = x$ and $q = e^{by}$, where b is a constant. For the sake of argument, let $b = 0.40\ \text{cm}^{-1}$.

a. Sketch what the "curves" of constant p and constant q look like in a cartesian x, y coordinate system.

b. An object at $y = 2.0$ cm has an acceleration $\boldsymbol{a}$ whose coordinates in the cartesian coordinate system are $a^x = 0.2\ \text{cm/s}^2$ and $a^y = -0.5\ \text{cm/s}^2$. What are the components of $\boldsymbol{a}$ in the semilog system? (Be careful with units!)

c. What is the metric of the semilog coordinate system? Is this metric diagonal?

d. Show that $\boldsymbol{a}$ as defined in part b has the same magnitude in both the cartesian and semilog systems.

e. What is the length of the basis vector ∂x?

P5.5 We can define "sinusoidal" coordinates u, w on a flat 2D plane by the relations $u = x$ and $w = y - A\sin(bx)$, where A and b are constants. For the sake of concreteness, let $A = 1.0$ cm and $b = \pi/2\ \text{cm}^{-1}$.

a. Sketch what the "curves" of constant u and constant w look like in a cartesian x, y coordinate system.

b. What is the metric of the sinusoidal coordinate system? Is this metric diagonal?

c. Imagine that an object moves with constant velocity $\boldsymbol{v}$ such that $v^x = v$ and $v^y = 0$. Such an object's position will be $x = vt$ (assuming $x = 0$ at $t = 0$) and y = constant. Find the object's velocity components v^u and v^w in the u, w coordinate system. Express your results in terms of v, t, A, and b.

d. Show that the squared magnitude of $\boldsymbol{v}$ is still the constant v^2 in this coordinate system, even though the velocity component v^w is *not* constant in time. Explain *why* v^w is not constant, even though the vector $\boldsymbol{v}$ in abstract always points in the same direction and always has the same magnitude.

e. Argue therefore that dv^w/dt *cannot* be equal to the component a^w of the object's acceleration vector $\boldsymbol{a}$ in the u, w coordinate system. (*Hint:* Note that $a^x = a^y = 0$ in the cartesian system.) We will learn in a later chapter how to take derivatives *correctly* in an curvilinear coordinate system.

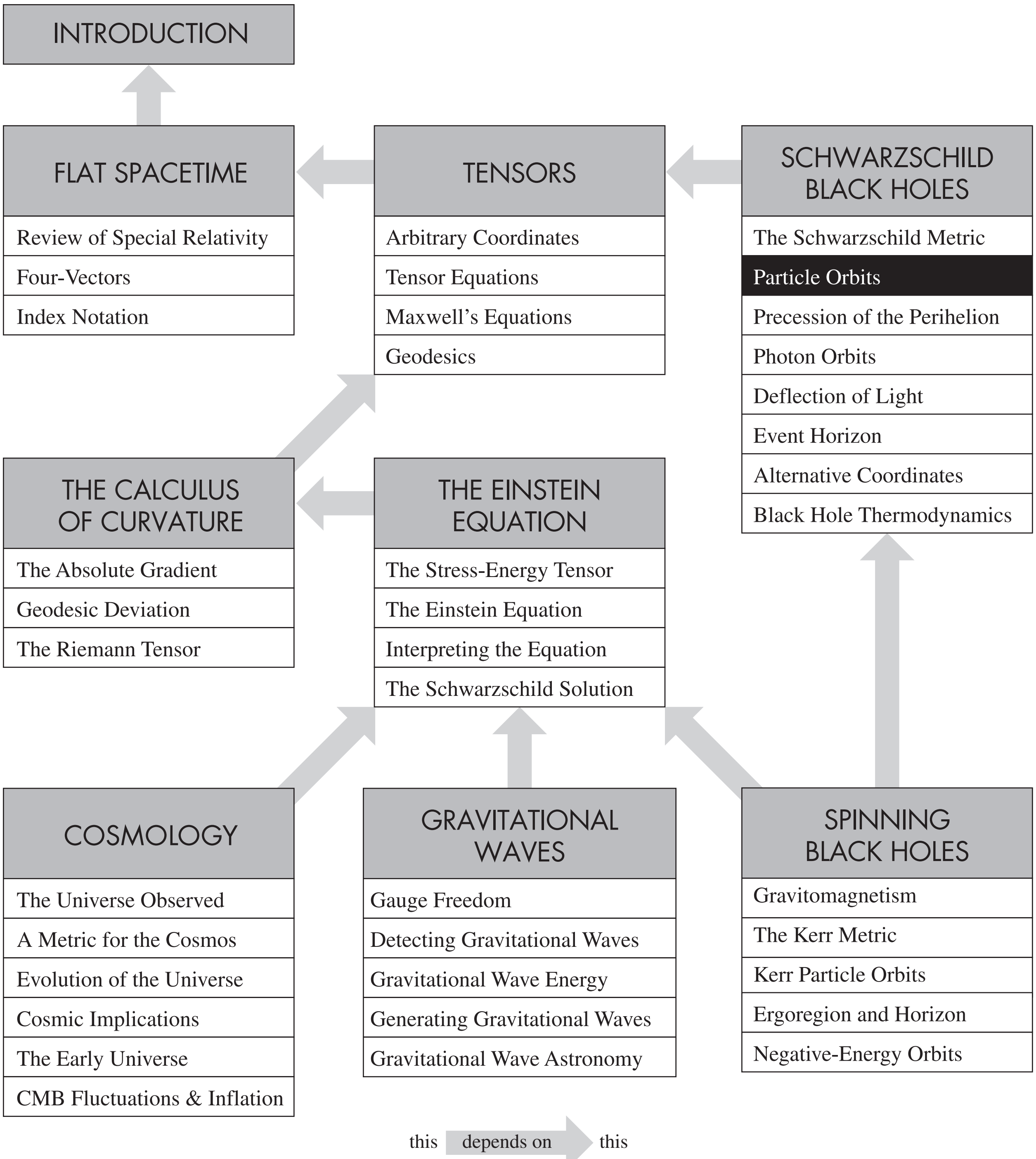
P5.6 Consider polar-coordinate-like "radial-longitude" coordinates r, ϕ for the 2D surface of a sphere of radius R, where r is the distance along the sphere's surface measured from the north pole and ϕ is an angular longitude coordinate measured around the pole. Note that curves of constant r and curves of constant ϕ are always perpendicular to each other everywhere on the sphere. Therefore (as we did in Box 5.1 for polar coordinates), by considering displacements on the sphere's surface that lie purely in the θ and ϕ directions, infer the components $g_{\mu\nu}$ of the metric for this coordinate system (assuming we use a coordinate basis). Express these components purely in terms of R and r. (We will later find a similar coordinate system helpful in describing the spatial geometry of the universe.)

P5.7 Consider the two-dimensional surface of a paraboloid defined by the relation $z = br^2$ (where b is some constant and $r^2 = x^2 + y^2$) in a 3D flat (Euclidean) space.

a. Sketch this surface in a 3D xyz plot.

b. Define coordinates r, ϕ for this surface, where the r coordinate of a point on the surface is defined as above and ϕ is an angle measured around the surface's axis of symmetry (the z axis), like a longitudinal coordinate on the earth. Determine the metric components $g_{\mu\nu}$ for these coordinates on the paraboloid's surface, assuming that we use a coordinate basis. (*Hint:* Note that a step toward larger r on the surface means not only moving *away* from the z axis in the 3D space but also moving *upward* to more positive z. What is the distance ds along the surface involved in a step of dr along a curve of constant ϕ?)

10. PARTICLE ORBITS



Introduction. In chapter 8, we discussed the general geodesic equation for arbitrary coordinates in arbitrary spacetimes. The geodesic equation, in the form that will be most useful in this chapter, reads

$$0 = \frac{d}{d\tau}\left(g_{\mu\nu}\frac{dx^\nu}{d\tau}\right) - \tfrac{1}{2}\partial_\mu g_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} \tag{10.1}$$

In this chapter, we will use the geodesic equation to explore the trajectories that particles with nonzero rest mass follow in Schwarzschild spacetime and develop a variety of tools for visualizing and modeling those trajectories.

The Schwarzschild Metric Tensor. The Schwarzschild metric equation is

$$ds^2 = -\left(1 - \frac{2GM}{r}\right)dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1}dr^2 + r^2 d\theta^2 + r^2\sin^2\theta\, d\phi^2 \tag{10.2}$$

Comparing this with the general metric equation $ds^2 = g_{\mu\nu}\,dx^\mu dx^\nu$, we can read off the nonzero components of the Schwarzschild metric tensor:

$$g_{tt} = -\left(1 - \frac{2GM}{r}\right), \quad g_{rr} = \left(1 - \frac{2GM}{r}\right)^{-1}, \quad g_{\theta\theta} = r^2, \quad g_{\phi\phi} = r^2\sin^2\theta \tag{10.3}$$

Conserved Quantities. The $\mu = t$ component of equation 10.1 tells us that

$$0 = \frac{d}{d\tau}\left(g_{t\nu}\frac{dx^\nu}{d\tau}\right) - \frac{1}{2}\partial_t g_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} \tag{10.4}$$

The Schwarzschild metric is both diagonal and time-independent, so this becomes

$$0 = \frac{d}{d\tau}\left(g_{tt}\frac{dt}{d\tau}\right) + 0 \quad\Rightarrow\quad \text{constant} = -g_{tt}\frac{dt}{d\tau} = \left(1 - \frac{2GM}{r}\right)\frac{dt}{d\tau} \equiv e \tag{10.5}$$

The quantity e is therefore conserved along all geodesic trajectories in Schwarzschild spacetime. We can interpret this quantity to be the *relativistic energy per unit mass that the object would have at infinity*, because if we substitute $r = \infty$ into equation 10.5, then e reduces to $dt/d\tau$. Since t is the time measured by a clock at infinity, this is the same as the object's four-velocity component u^t as measured by the observer at infinity, which in turn is p^t/m = relativistic energy per mass.

The $\mu = \phi$ component of equation 10.1 tells us that

$$0 = \frac{d}{d\tau}\left(g_{\phi\nu}\frac{dx^\nu}{d\tau}\right) - \frac{1}{2}\partial_\phi g_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} \tag{10.6}$$

The Schwarzschild metric is both diagonal and independent of ϕ, so this becomes

$$0 = \frac{d}{d\tau}\left(g_{\phi\phi}\frac{d\phi}{d\tau}\right) \quad\Rightarrow\quad \text{constant} = g_{\phi\phi}\frac{d\phi}{d\tau} = r^2\sin^2\theta\frac{d\phi}{d\tau} \equiv \ell \tag{10.7}$$

The quantity ℓ is therefore also conserved along all geodesic trajectories in Schwarzschild spacetime. For a trajectory on the equatorial plane, $\sin^2\theta = 1$, so this quantity reduces to being $\ell = r^2(d\phi/d\tau)$, which we can interpret as being relativistic angular momentum per unit mass [in Newtonian mechanics, $L/m = r^2\omega = r^2(d\phi/dt)$.]

As discussed in box 10.1, symmetry requires that each geodesic in Schwarzschild spacetime lies on a plane through the origin. We can, therefore, without loss of generality, choose our coordinates so that any given orbit of interest lies in the equatorial ($\theta = \pi/2$) plane. We will assume this in what follows.

The Radial Equation of Motion. Since all of the Schwarzschild metric components depend on r, the r component of the geodesic equation is a bit complicated. This is another case where using $-1 = \boldsymbol{u}\cdot\boldsymbol{u} = g_{\mu\nu}u^\mu u^\nu$ really pays off. If you substitute in the metric components from equation 10.3 and use the results in equations 10.5 and 10.7, the result (see box 10.2) is that

$$\tfrac{1}{2}(e^2 - 1) = \frac{1}{2}\left(\frac{dr}{d\tau}\right)^2 - \frac{GM}{r} + \frac{\ell^2}{2r^2} - \frac{GM\ell^2}{r^3} \equiv \tilde{E} \tag{10.8}$$

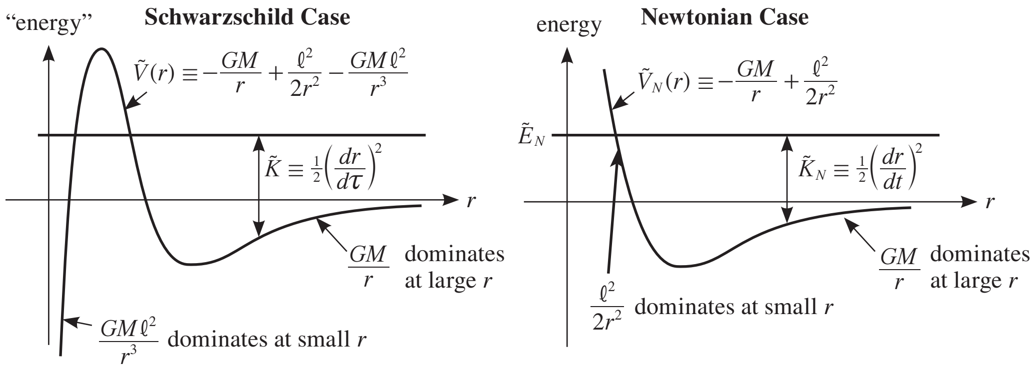


FIG. 10.1 These graphs show the effective potential-energy-per-unit-mass functions governing radial motion in the Schwarzschild case (left) and Newtonian case (right).

We can interpret this as being structurally equivalent to a conservation-of-energy equation, with $\tilde{K} \equiv \frac{1}{2}(dr/d\tau)^2$ serving as an effective "radial kinetic energy per unit mass," $\tilde{E}$ as an effective conserved "energy per unit mass," and

$$\tilde{V}(r) \equiv -\frac{GM}{r} + \frac{\ell^2}{2r^2} - \frac{GM\ell^2}{r^3} \tag{10.9}$$

as an effective "potential energy per unit mass" that depends only on r. With these definitions, equation 10.8 becomes simply $\tilde{E} = \tilde{K} + \tilde{V}(r)$.

Indeed, one can show (see box 10.3) that one can use conservation of Newtonian energy to derive an analogous equation for Newtonian orbits:

$$\frac{1}{2}\left(\frac{dr}{dt}\right)^2 + \tilde{V}_N \equiv \tilde{E}_N \quad \text{where} \quad \tilde{V}_N(r) \equiv -\frac{GM}{r} + \frac{\ell^2}{2r^2} \tag{10.10}$$

and $\tilde{E}_N$ is the actual Newtonian total energy per unit mass. The only formal difference between the two equations is that the Schwarzschild equation has an extra inverse-r^3 term in the $\tilde{V}(r)$ function. (But also note that in the Schwarzschild case, r is the *circumferential* radius, and e, not $\tilde{E}$, is the true relativistic energy per unit mass.)

Graphical Interpretation of the Possible Orbits. Figure 10.1 shows graphs of the effective potential-energy-per-unit-mass curves for the Schwarzschild and Newtonian cases. One can read these graphs to determine characteristics of the possible trajectories much the way that one interprets potential energy graphs for one-dimensional motion in ordinary mechanics. For both cases:

- $\tilde{K} = 0$ where $\tilde{E} = \tilde{V}(r)$: such points are "turning points" where outgoing radial motion becomes ingoing motion and vice versa.
- $\tilde{E} < 0$ orbits correspond to bound orbits (ellipses in the Newtonian case) that have maximal and minimal radial coordinates [at radii where $\tilde{E} = \tilde{V}(r)$].
- $\tilde{E} > 0$ orbits correspond to unbound orbits (hyperbolas in the Newtonian case).
- Radii where $d\tilde{V}/dr = 0$ are possible circular-orbit radii (see box 10.4).

However, there are special features of the curve for Schwarzschild spacetime:

- The effective potential energy goes to negative infinity as $r \to 0$ instead of going to positive infinity.
- For sufficiently high $\tilde{E}$, spiral orbits exist that go into $r = 0$.
- As long as $\ell > \sqrt{12}\,GM$ (see box 10.4 and box 10.6), we have an unstable circular orbit (at the radial coordinate where $\tilde{V}(r)$ is maximum) *and* a stable circular orbit (at the radial coordinate where $\tilde{V}(r)$ is minimum).

Circular Orbits. Indeed, by solving for the radii where $d\tilde{V}/dr = 0$, one can show (see box 10.4) that for a given value of ℓ, the radial coordinates r_c of possible circular orbits in Schwarzschild spacetime are

$$r_c = \frac{6GM}{1 \pm \sqrt{1 - 12(GM/\ell)^2}} \tag{10.11}$$

with the inner orbit being unstable and the other orbit being stable. The radius of the single (stable) Newtonian circular orbit is $r^2 = \ell^2/GM$. One can also prove (see box 10.5) that Kepler's third law

$$\Omega^2 = \frac{GM}{r_c^3} \quad \Rightarrow \quad T^2 = \left|\frac{2\pi}{\Omega}\right|^2 = \frac{4\pi^2}{GM} r_c^3 \tag{10.12}$$

still applies in Schwarzschild spacetime as long as we take T to be the period of the orbit as measured by an observer at infinity ($\Omega = d\phi/dt$ is the angular speed of the orbiting object as determined by an observer at infinity). However, it does not quite *mean* the same thing, as the Schwarzschild radial coordinate is not equal to the Newtonian radial coordinate.

Equations for Radial Acceleration. By taking the τ-derivative of both sides of equation 10.8, you can find the following equation for the radial acceleration of an object in Schwarzschild spacetime:

$$0 = \tfrac{1}{2} 2 \frac{d\cancel{r}}{\cancel{d\tau}}\left(\frac{d^2 r}{d\tau^2}\right) + \frac{GM}{r^2}\frac{d\cancel{r}}{\cancel{d\tau}} - \frac{\cancel{2}\ell^2}{\cancel{2}r^3}\frac{d\cancel{r}}{\cancel{d\tau}} + \frac{3GM\ell^2}{r^4}\frac{d\cancel{r}}{\cancel{d\tau}}$$

$$\Rightarrow \quad \frac{d^2 r}{d\tau^2} = -\frac{GM}{r^2} + \frac{\ell^2}{r^3} - \frac{3GM\ell^2}{r^4} \quad \text{(Schwarzschild)} \tag{10.13}$$

The corresponding Newtonian equation (found by taking t-derivative of both sides of equation 10.10) is

$$\frac{d^2 r}{dt^2} = -\frac{GM}{r^2} + \frac{\ell^2}{r^3} \qquad \text{(Newtonian)} \tag{10.14}$$

Note that again the Newtonian equation lacks the final term that appears in the Schwarzschild case. Note that for an object moving radially, $\ell \equiv r^2 d\phi/d\tau = 0$, so the radial acceleration is precisely $d^2 r/d\tau^2 = -GM/r^2$ in such a case.

These equations turn out to be more useful for constructing computer models of trajectories than equations 10.8 or 10.10.

Astrophysical Implications. For Schwarzschild spacetime, one can show (see box 10.6) that contrary to the predictions of Newtonian gravity (where stable circular orbits exist for all radii), there are *no* stable circular orbits with $r \leq 6GM$. Therefore, $r = 6GM$ represents the "innermost stable circular orbit" (ISCO) for anything orbiting a highly compact (non-rotating) object.

This has astrophysical relevance, because compact objects do exist whose outer radii are smaller than $6GM$. Neutron stars, which are created by catastrophic stellar core collapse during supernova, have masses typically on the order of 1.4 solar masses and radii of roughly 10 km $\approx 5GM$. The spacetime in the vacuum outside this radius will be Schwarzschild if the star is not rotating (we will study rotating objects later). A non-rotating black hole (which we will also discuss more later) is entirely Schwarzschild spacetime all the way down to $r = 0$. Astrophysicists have strong evidence that both neutron stars and black holes exist in the universe.

Part of the evidence comes from observations of point-like X-ray sources in our local group of galaxies. The graphs shown in figure 10.1 mean that particles with any significant angular momentum per unit mass ℓ falling toward a compact object will "bounce off" of the potential barrier in $\tilde{V}(r)$ before reaching the object. Eventually, such particles will organize themselves into a flat "accretion disk" around the object. Friction between particles in the disk radiates energy away, rapidly circularizing their orbits (note that stable circular orbits have the lowest energy for a given angular momentum, since they correspond to the bottom point of a valley in the "potential energy" graph). Since there is no easy way for particles to "radiate" angular momentum out of the disk, it might seem the disk's total angular momentum should be conserved, thus keeping the circularized orbits stable. However, it is now widely believed that magnetic interactions between charged particles create turbulence that allows a few particles to carry angular momentum outward, slowly decreasing the angular momentum of the rest. This allows most particles in the disk to slowly spiral inward. Once particles pass through the ISCO, there is no longer any barrier in $\tilde{V}(r)$ (see box 10.6), so particles rapidly fall into or accrete onto the object from there.

How much energy can be released by particles in the accretion disk? Consider a particle starting essentially at rest at $r = \infty$. According to equation 10.5, its relativistic energy per unit rest mass at infinity is $e = 1$ (meaning that it has its mass energy and nothing else). At the ISCO radius $r = 6GM$, equation 10.11 implies that the particle's angular momentum per unit mass ℓ in its last circular orbit is such that

$$12\left(\frac{GM}{\ell}\right)^2 = 1 \quad \Rightarrow \quad \ell = \sqrt{12}\,GM \tag{10.15}$$

If you plug this, $r = 6GM$, and $dr/d\tau = 0$ into the energy equation 10.8, you can show (see box 10.7) that the particle's final energy-per-unit-mass just as it drifts past the ISCO has decreased by about

$$\Delta e = \sqrt{\frac{8}{9}} - 1 \approx -0.057 \tag{10.16}$$

This means that in order to make it from infinity to the ISCO, the particle must radiate away energy equivalent to 5.7% of its rest mass. Even if the particle then falls into a black hole (where its remaining energy is entirely absorbed), the energy released just by the disk is *enormous*. For comparison, the fusion reaction used by most stars converts only about 0.7% of their hydrogen fuel's mass-energy into radiated energy (and nuclear fission is more than an order of magnitude *less* efficient).

One can get an order-of-magnitude estimate of how the inner disk's temperature depends on its luminosity as follows. The inner disk is where particle velocities are highest and the disk will be hottest. Let's crudely estimate that essentially all of the energy comes from the portion of the disk between its inner radius R and $2R$, that the disk's temperature T is constant over this region, that it radiates energy like an ideal black body, and that the rest of the disk emits nothing. The Stefan-Boltzmann law says that the luminosity L of a blackbody of area A at temperature T is

$$L \equiv \frac{\text{energy radiated}}{\text{time}} = \sigma A T^4 \tag{10.17}$$

where σ is the Stefan-Boltzmann constant = 5.67×10^{-8} W/(m^{2}K^4). Given the assumptions above, you can show (see problem P10.1) that the temperature of the disk will be of order of magnitude

$$T \sim \left(\frac{L}{L_\odot}\right)^{1/4}\left(\frac{M_\odot}{M}\right)^{1/2}(2 \times 10^6\,\text{K}) \tag{10.18}$$

where $L_\odot$ and $M_\odot$ are the sun's luminosity and mass, respectively. Note that although the temperature goes up with increased luminosity and decreases with increasing mass, T varies as small powers of these quantities, so for any stellar-sized source radiating a stellar-like energy, T will not be much different than 10^6 K. Since such temperatures are about 500 times higher than the sun's surface temperature, the typical wavelengths of light emitted will be about 500 times shorter, or on the order of magnitude of 1 nm, in the X-ray region (photon energy $\approx$ 1.2 keV).

In fact, we observe a number of point X-ray sources in our galaxy and in neighboring galaxies having luminosities up to $10^6\,L_\odot$. Emission from accretion disks around highly compact objects is the most reasonable explanation for such sources (as other energy sources would require an implausible rate of fuel consumption).

BOX 10.1 Schwarzschild Orbits Must Be Planar

A simple symmetry argument provides the proof that all Schwarzschild orbits must be planar. Consider an object whose initial velocity lies in the equatorial plane. Its subsequent trajectory *must also* lie in the equatorial plane, because in this spherically symmetric spacetime, one side of the equatorial plane is identical to the other, so there is no reason for a free object (whose motion is completely determined by the spacetime) to leave the plane and thus choose one side over the other. Also in a spherically symmetric spacetime, the equatorial plane is no different than any other plane going through the origin. No matter what an object's initial velocity might be, that velocity and the origin define a plane through the origin, so the object's trajectory will be confined to that plane. Therefore *all geodesic trajectories in Schwarzschild spacetime must be planar.*

You can show that this statement is consistent with the geodesic equation. The $\mu = \theta$ component of equation 10.1 tells us that

$$0 = \frac{d}{d\tau}\left(g_{\theta\nu}\frac{dx^\nu}{d\tau}\right) - \frac{1}{2}\partial_\theta g_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} \tag{10.19}$$

You can show that this component of the geodesic equation becomes

$$0 = r^2\frac{d^2\theta}{d\tau^2} + 2r\frac{dr}{d\tau}\frac{d\theta}{d\tau} - r^2\sin\theta\cos\theta\left(\frac{d\phi}{d\tau}\right)^2 \tag{10.20}$$

and that $\theta = \pi/2 =$ constant is a solution to this equation. This means that orbits in the equatorial plane are allowed by the geodesic equation. Moreover, the fact that the value of θ does not accelerate when $d\theta/d\tau = 0$ and $\theta = \pi/2$ also means that if the object's trajectory initially lies in the equatorial plane, it cannot not curve away from that plane. Again, since the equatorial plane is no different than any other plane through the origin, this proves (in a different way) that geodesic trajectories in Schwarzschild spacetime must be planar.

Exercise 10.1.1. Verify equation 10.20 and show that $\theta = \pi/2 =$ constant is a solution to that equation.

BOX 10.2 The Schwarzschild "Conservation of Energy" Equation

The Schwarzschild metric tensor is diagonal, so the implied sums in the equation $-1 = g_{\mu\nu}u^{\mu}u^{\nu}$ yield only four nonzero terms:

$$-1 = g_{tt}\left(\frac{dt}{d\tau}\right)^2 + g_{rr}\left(\frac{dr}{d\tau}\right)^2 + g_{\theta\theta}\left(\frac{d\theta}{d\tau}\right)^2 + g_{\phi\phi}\left(\frac{d\phi}{d\tau}\right)^2 \qquad (10.21)$$

If you substitute in the value of the metric components and use equations 10.5 and 10.7, you should be able to show that this can be written

$$1 = \left(1 - \frac{2GM}{r}\right)^{-1} e^2 - \left(1 - \frac{2GM}{r}\right)^{-1}\left(\frac{dr}{d\tau}\right)^2 - \frac{\ell^2}{r^2} \qquad (10.22)$$

for orbits in the equatorial plane ($\theta = \pi/2 =$ constant).

Exercise 10.2.1. Verify equation 10.22.

From equation 10.22, it is only a few steps to equation 10.8.

Exercise 10.2.2. Work out the steps between equation 10.22 and equation 10.8.

BOX 10.3 Deriving Conservation of Newtonian Energy for Orbits

Consider an object following a Newtonian orbit in the r, ϕ plane. We can write conservation of Newtonian energy for such an orbit in the following form:

$$E = \tfrac{1}{2}m(v_r^2 + v_\phi^2) - \frac{GMm}{r} = \tfrac{1}{2}m\left[\left(\frac{dr}{dt}\right)^2 + r^2\left(\frac{d\phi}{dt}\right)^2\right] - \frac{GMm}{r} \tag{10.23}$$

where E is the object's total Newtonian energy, m is its mass, M is the mass of the primary at the origin, v_r is the radial component of the object's velocity, and v_ϕ is the component perpendicular to the radial component in the direction in which ϕ increases. But notice that the object's angular momentum around the origin is

$$L = mrv_\phi = mr^2\frac{d\phi}{dt}, \quad \text{so} \quad \ell \equiv \frac{L}{m} = r^2\frac{d\phi}{dt} \tag{10.24}$$

Exercise 10.3.1. Show that these equations together imply equation 10.10, as long as we define $\tilde{E}_N \equiv E/m$.

BOX 10.4 The Radii of Circular Orbits

Note that if we take the τ-derivative of both sides of equation 10.8, we get

$$0 = \frac{d^2r}{d\tau^2}\frac{dr}{d\tau} + \frac{d\tilde{V}}{dr}\frac{dr}{d\tau} \quad \Rightarrow \quad \frac{d^2r}{d\tau^2} = -\frac{d\tilde{V}}{dr} \tag{10.25}$$

This means that points where $d\tilde{V}/dr = 0$ will be points where an object will experience no radial acceleration. If the object's effective energy $\tilde{E} = \tilde{V}(r)$ at such a point, then the object will have no radial velocity and no radial acceleration, so it will remain at constant r. Such a particle must have nonzero angular momentum (figure 10.1 makes it clear that if $\ell = 0$, then we will never have $d\tilde{V}/dr = 0$), so it will therefore follow a circular orbit around the origin.

So the radii r_c of possible circular orbits correspond to values of the radial coordinate r where $d\tilde{V}/dr = 0$. Setting the r-derivative of equation 10.9 to zero yields

$$0 = +\frac{GM}{r_c^2} - \frac{\ell^2}{r_c^3} + \frac{3GM\ell^2}{r_c^4} \tag{10.26}$$

Exercise 10.4.1. Verify equation 10.26.

BOX 10.4 (continued) The Radii of Circular Orbits

The fastest way to equation 10.11 (which is the most useful form for the equation for r_c) is to define $u_c \equiv 1/r_c$. If you substitute this into equation 10.26, divide both sides by u_c^2, and solve the resulting quadratic equation, you will get equation 10.11, which (for the sake of convenience) I reproduce here:

$$r_c = \frac{6GM}{1 \pm \sqrt{1 - 12(GM/\ell)^2}} \tag{10.11r}$$

Exercise 10.4.2. Verify this.

Note that equation 10.11 implies that as long as the square root is real, there will be a circular orbit outside $r = 6GM$ (corresponding to the negative sign in the denominator) and one inside that radius (corresponding to the positive sign). From this equation, you can also find the smallest value for ℓ for which circular orbit solutions exist at all.

Exercise 10.4.3. What is this value of ℓ?

BOX 10.5 Kepler's Third Law

The component of the geodesic equation 10.1 with $\mu = r$ implies that for the diagonal Schwarzschild metric

$$0 = \frac{d}{d\tau}\left(g_{rr}\frac{dr}{d\tau}\right) - \frac{1}{2}\frac{\partial g_{\alpha\beta}}{\partial r}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} \qquad (10.27)$$

For a circular orbit ($dr/d\tau = 0$) in the equatorial plane ($d\theta/d\tau = 0$), you can show that this reduces to

$$0 = \frac{\partial g_{tt}}{\partial r}\left(\frac{dt}{d\tau}\right)^2 + \frac{\partial g_{\phi\phi}}{\partial r}\left(\frac{d\phi}{d\tau}\right)^2 \qquad (10.28)$$

Exercise 10.5.1. Verify this.

Define $\Omega \equiv d\phi/dt$. This is the angular speed that an observer at infinity (whose time is equal to the coordinate time t) will consider the orbiting particle to have. The orbital period T this observer measures is simply $T = |2\pi/\Omega|$. If we multiply equation 10.28 through by $(d\tau/dt)^2$ and apply the chain rule, we get

$$0 = \frac{\partial g_{tt}}{\partial r} + \frac{\partial g_{\phi\phi}}{\partial r}\left(\frac{d\phi}{d\tau}\frac{d\tau}{dt}\right)^2 = \frac{\partial g_{tt}}{\partial r} + \frac{\partial g_{\phi\phi}}{\partial r}\left(\frac{d\phi}{dt}\right)^2 = \frac{\partial g_{tt}}{\partial r} + \frac{\partial g_{\phi\phi}}{\partial r}\Omega^2 \qquad (10.29)$$

From this, you can prove that $\Omega^2 = GM/r^3$ and from that the other part of equation 10.12.

Exercise 10.5.2. Do this.

BOX 10.6 The Innermost Stable Circular Orbit (ISCO)

Stable circular-orbit radii correspond to local *minima* of the effective potential energy function $\tilde{V}(r)$ displayed in figure 10.1 (see if you can remember why). To determine whether an extremum is a minimum, we need to see whether $d^2\tilde{V}/dr^2$ is positive (i.e., the curve is concave up) at the extremum. Equation 10.26 implies that

$$\frac{d\tilde{V}}{dr} = \frac{GM}{r^2} - \frac{\ell^2}{r^3} + \frac{3GM\ell^2}{r^4} \tag{10.30}$$

$$\Rightarrow \quad \frac{d^2\tilde{V}}{dr^2} = -\frac{2GM}{r^3} + \frac{3\ell^2}{r^4} - \frac{12GM\ell^2}{r^5} \tag{10.31}$$

At a local *minimum*, we must have $d\tilde{V}/dr = 0$ and $d^2\tilde{V}/dr^2 > 0$.

Exercise 10.6.1. Set the first expression to 0, multiply it by $2/r$, and add it to the second to prove that minima can exist only for $r > 6GM$.

For the record, figure 10.2 shows a graph of $\tilde{V}(r)$ when $\ell/GM = \sqrt{12}$ (the value of ℓ that leads to a circular orbit of radius $r = 6GM$ according to equation 10.11). You can see that $r = 6GM$ in this case corresponds to an inflection point, not a minimum, and that there is no barrier preventing a particle with slightly more energy than the circular-orbit energy from falling in.

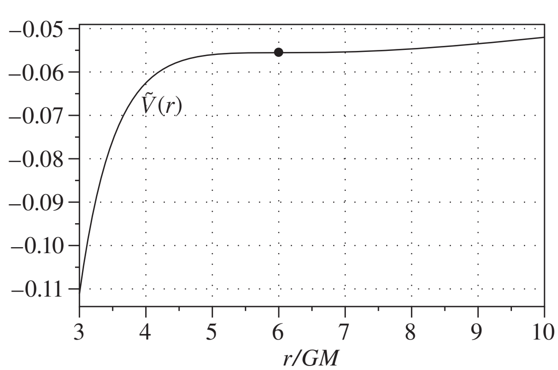


FIG. 10.2 A plot of $\tilde{V}(r)$ when $\ell = \sqrt{12}\,GM$ (the value of ℓ that an object has in the ISCO).

BOX 10.7 The Energy Radiated by an Inspiraling Particle

Exercise 10.7.1. Substitute $\ell/GM = \sqrt{12}$, $r = 6GM$, and $dr/d\tau = 0$ into equation 10.8 and solve for e to find a particle's energy per unit rest mass as measured at infinity when it is in the ISCO. Then subtract from the value of e for the particle when it *was* at infinity to find its *change* in e during the inspiral process. Note that it is e that is the physically relevant energy here, not $\tilde{E}$ (which was invented to make it easier to compare the Schwarzschild and Newtonian cases.)

HOMEWORK PROBLEMS

P10.1 Verify that equation 10.18 is correct. (*Hint:* Calculate the approximate area A of the disk between $R = 6GM$ and $2R$, substitute this into equation 10.17, multiply top and bottom by $L_\odot$ and $M_\odot$, and solve for T. Note that $GM_\odot = 1477$ m and $L_\odot = 3.9 \times 10^{26}$ W.)

P10.2 An object falls radially inward toward a black hole with mass M), starting at rest at infinity. How much time will a clock on the object register between the events of the object passing through the Schwarzschild radial coordinates $r = 10GM$ and $r = 2GM$? (*Hint:* Argue that an object released from rest at infinity will have $\tilde{E} = 0$, i.e., $e = 1$.)

P10.3 Two objects fall radially in from infinity, one having $e = 1$ and the other having $e = 2$. An observer at rest at $r = 6GM$ watches these objects pass. How much faster is the second object moving than the first object according to this observer? (*Hints:* Remember that the observer will measure each object's energy to be $E = -\boldsymbol{p} \cdot \boldsymbol{u}_{\text{obs}}$. Calculate this in Schwarzschild coordinates: you may find equation 9.20 helpful. Then one can infer the speed the observer will measure using $E = m/\sqrt{1 - v^2}$.) (Adapted from problem 9.7 in Hartle, *Gravity,* Addison Wesley, 2003.)

P10.4 **a.** Find a general expression for dr/dt for a geodesic in the equatorial plane as a function of r, GM, e, and ℓ. What does this equation say happens as r approaches $2GM$? (As we will see in chapter 14, this proves to be an artifact of the Schwarzschild coordinate system.)

b. Find expressions for both $dr/d\tau$ and dr/dt if we drop an object from rest in the equatorial plane at radial coordinate r_0. (*Hint:* You should be able to determine specific values of e and ℓ in this case.)

P10.5 Imagine we launch an object radially from $r = r_0$ with sufficient speed so that it comes to rest at $r = r_1 > r_0$ before falling back to $r = r_0$. Find an expression (in terms of GM, r_0, and r_1) for the proper time measured by the object during this trajectory. (*Hints:* Determine e in terms of r_1, and change variables to $u \equiv r/r_1$. Note that $u \leq 1$ for the entire trajectory. Feel free to look up a fairly nasty integral.)

P10.6 Imagine that an object in a stable circular orbit around a neutron star ($GM = 2.2$ km) has an angular momentum per unit mass of $\ell = 6GM = 13.2$ km.

a. Calculate the radius of the orbit.

b. Calculate the period of orbit as measured by a clock traveling with the object. Express your answer in milliseconds. (*Hint:* You can very easily calculate $d\phi/d\tau$, which is constant for the orbit.)

c. Calculate the period of the orbit as measured by an observer at infinity. Express your answer in milliseconds.

P10.7 **a.** Use equation 10.11 to show that for a circular orbit around a gravitating object of mass M,

$$\ell^2 = \frac{GMr_c^2}{(r_c - 3GM)} \tag{10.32}$$

for both signs of the original equation, where r_c is the circular orbit's radial coordinate. Note that this equation implies that for objects with nonzero mass, no circular orbits of any kind exist for $r \leq 3GM$.

b. Use this to show that the effective energy-per-unit-mass $\tilde{E}$ for a circular orbit as a function of r_c is

$$\tilde{E} = -\frac{GM}{2r_c}\left(1 - \frac{3GM}{r_c}\right)^{-1}\left[1 - \frac{4GM}{r_c}\right] \tag{10.33}$$

and compare to the Newtonian result $E/m = -GM/2r_c$.

c. Find e as a function of r_c alone and check that $e = \sqrt{8/9}$ when $r_c = 6GM$.

P10.8 Consider an object starting essentially at rest at infinity, but with an infinitesimal tangential velocity sufficient to give it an angular-momentum-per-unit-mass ℓ. Argue that if ℓ has the appropriate value, this particle can spiral in to an unstable circular orbit at $r_c = 4GM$, and find that appropriate value of ℓ in terms of GM. (*Hint:* Use the results of problem P10.7.)

P10.9 A spaceship is in a stable circular orbit at a Schwarzschild radial coordinate of $r = 10GM$ around a supermassive black hole whose mass is 10^6 solar masses.

a. What is this orbit's circumference in kilometers?

b. What is the effective energy per unit mass $\tilde{E}$ and angular momentum per unit mass ℓ for this object? (*Hint:* Use the results of problem P10.7.)

c. What is the period of the spaceship's orbit according to its own clock?

P10.10 Find the relation between $\omega \equiv d\phi/d\tau$ and r for a circular orbit. How does this compare to the relationship $\Omega^2 = GM/r^3$ found in box 10.5?

P10.11 Using the method displayed in box 10.4, calculate the expression that for Newtonian mechanics is analogous to equation 10.11. Also show that the Newtonian result is the large-ℓ, large-r limit of equation 10.11.

P10.12 Consider three observers, one in a spaceship in a circular orbit of radius r, one stationary at radius r, and one effectively at infinity. Calculate the period of the orbit measured by each observer as a function of r, and from that period, infer the speed at which each would consider the spacecraft to be moving if we define that speed to be the circumference of the orbit $2\pi r$ divided by the observer's time. Rank these speeds from smallest to greatest, and explain why this ranking makes sense physically. Are any of the speeds (so calculated) possibly greater than 1? If so, also explain how that is possible. (*Hint:* Equation 10.32 may be helpful.)

P10.13 In chapter 12 we will see that photons can orbit a Schwarzschild black hole at a constant radial coordinate of $r = 3GM$. Consider a photon in such an orbit.

a. The definition of the Schwarzschild r coordinate implies that if the photon moves through an angular displacement of $d\phi$ in a certain coordinate time dt, the physical distance the photon moves is $r\,d\phi$. Therefore, an observer at infinity (whose clock measures time dt) will conclude that the photon's speed is $r\,d\phi/dt$. Use the fact that $ds = 0$ along a photon worldline to show that its speed (so defined) is $V = 0.577$.

b. An observer at $r = 3GM$ observes this same photon orbit exactly once in a time T. Use the Schwarzschild metric to compute the time T this stationary observer's clock measures between the two events of the photon passing once and then passing a second time as a fraction of the coordinate time Δt between these events. Use this to calculate the photon's speed v as measured by that stationary observer.

c. Explain qualitatively and physically why v measured by the observer at $r = 3GM$ is not the same the value of V measured by the observer at $r \approx \infty$.

P10.14 A comet with mass m comes in from essentially rest at infinity but with sufficient angular momentum so that it approaches a black hole, loops partway around it, then recedes back to to infinity. Our goal in this problem is to determine the comet's speed as measured by a stationary observer at the comet's point of closest approach.

a. Argue that as r goes to infinity, $d\phi/d\tau$ must go to zero for any finite ℓ. Then use the metric equation and the definition of e to argue that $e \approx 1$ for an comet having $dr/d\tau \approx 0$ at large r, even if it has finite ℓ.

b. Show that at any radial coordinate r

$$\frac{d\phi}{d\tau} = \left(1 - \frac{2GM}{r}\right)\frac{\ell}{r^2}\frac{dt}{d\tau} \tag{10.34}$$

c. Write out the relation $-1 = \boldsymbol{u}\cdot\boldsymbol{u}$ at the point of closest approach and use equation 10.34 to show that this comet's four-momentum $\boldsymbol{p} \equiv m\boldsymbol{u}$ at its point of closest approach has the following time component in the Schwarzschild coordinate system:

$$p^t = \frac{m}{\sqrt{\left(1 - \frac{2GM}{R}\right) - \left(1 - \frac{2GM}{R}\right)^2 \frac{\ell^2}{R^2}}} \tag{10.35}$$

where R is the (unknown) radial coordinate of the comet's closest approach.

d. We have seen in previous contexts that an object's energy as measured by an observer moving with four-velocity $\boldsymbol{u}_{\text{obs}}$ will be $E_{\text{obs}} = -\boldsymbol{p}\cdot\boldsymbol{u}_{\text{obs}}$. Since dot products have the same value in every coordinate system, we can use Schwarzschild coordinates to calculate this dot product, but the result will still be the energy that the observer would measure. Find $\boldsymbol{u}_{\text{obs}}$ in Schwarzschild coordinates for a stationary observer at R, and evaluate E_{obs} for the comet in terms of ℓ, GM, and R.

e. In the observer's orthonormal frame $E = m/\sqrt{1 - v^2}$. Use this to evaluate the comet's speed v according to a stationary observer at R in terms of ℓ, GM, and R.

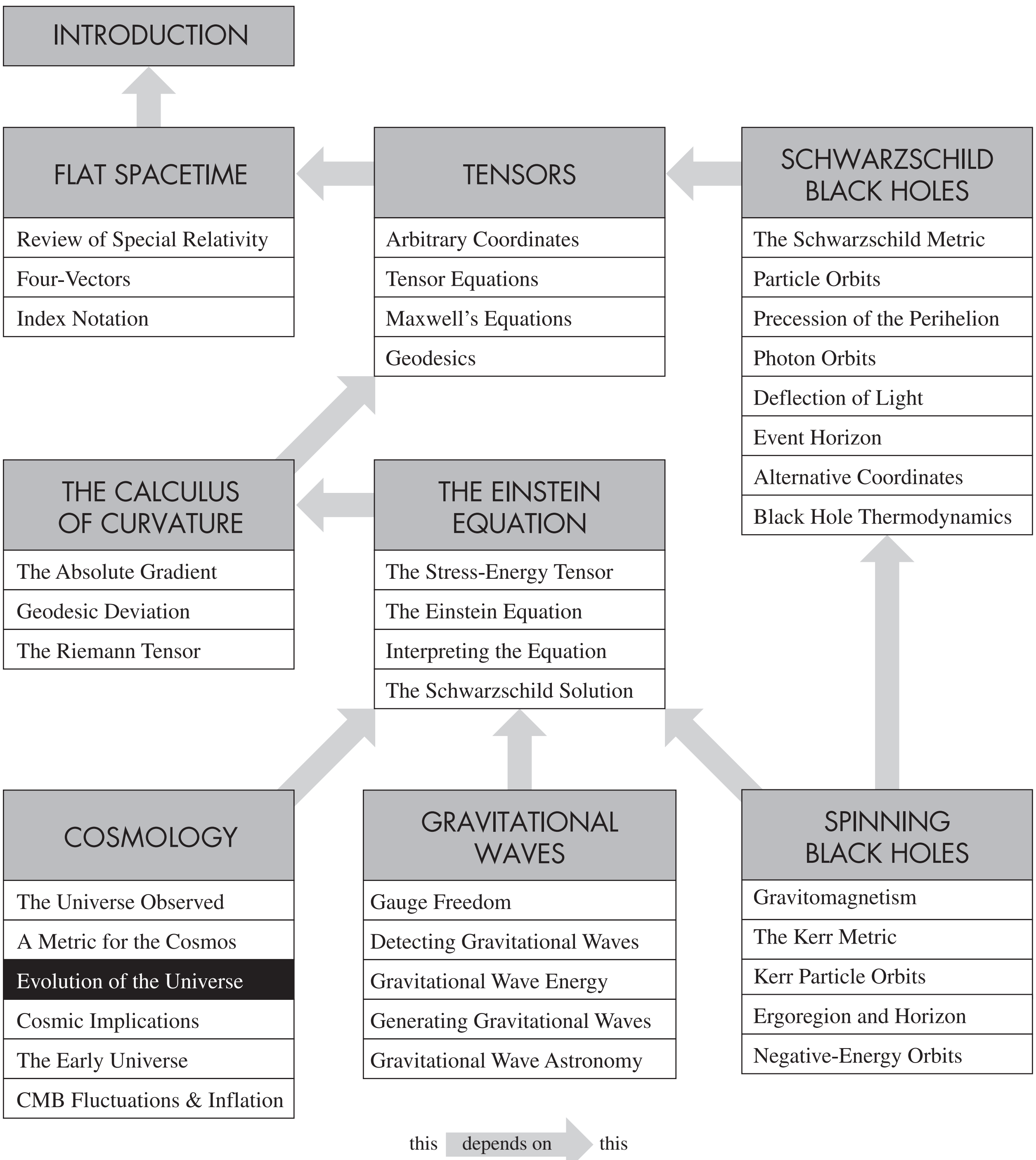
f. Use equation 10.8 to find the radial coordinate of closest approach R in terms of ℓ. Explain why there are two solutions, and argue which one you want. Also, show that in the large-ℓ limit, your desired solution approaches the result we would get if gravity were Newtonian, which is $\ell^2/2GM$. Is the point of closest approach closer or farther than the Newtonian result? Does this make sense? (*Hint:* Study figure 10.1 to help you answer the question about why there are two solutions and answer the last question.)

P10.15 As you may know from discussions of the so-called twin paradox, one can effectively travel to the future by getting into a spaceship and traveling to and back from some distant point at nearly the speed of light. However, if you have a local black hole, you can do this *much* less expensively as follows. Put yourself in an orbit with the correct $\tilde{E}$ and ℓ at your starting point at approximately infinite r so that (subsequently without using any fuel) you spiral into an unstable circular orbit near to the black hole, hang out there for a while, and then spiral back out to your starting place.

a. If you start essentially at rest at a very large radius, but give yourself just the right tiny bump in the tangential direction to give yourself the right ℓ, show that you can spiral in to an unstable circular orbit at $r = 4GM$ and hang out there for a while, before spiraling back out again. Calculate the correct value of ℓ. (*Hint:* See problem P10.7.) Also, for the portion of your trajectory where your orbit is approximately circular at $r = 4GM$, by what factor does your clock run slower than one at approximately infinite r?

b. You can improve this performance by giving yourself enough radially inward velocity at very large r to end up in an unstable circular orbit at a closer radius. Imagine that for the portion of your trajectory where your orbit is approximately circular, you want your clock to run 10 times more slowly than a clock at very large r. Calculate the value of ℓ that you would need and what your speed v at very large r needs to be to put yourself into the required orbit. You should find that the required v will be relativistic, but that traveling such a speed in flat spacetime would give you *much* less of a slowdown.

26. EVOLUTION OF THE UNIVERSE



Introduction. In the previous chapter, we developed a trial metric for an homogeneous and isotropic universe that had the form

$$ds^2 = -dt^2 + a^2\left[d\bar{r}^2 + q^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right] \tag{26.1}$$

where $\bar{r}$ is a radial coordinate that is comoving with the galactic "gas" as the universe expands, $a(t)$ is a unitless time-dependent quantity that specifies the scale of the universe, and $q(\bar{r})$ is either $R\sin(\bar{r}/R), \bar{r}$, or $R\sinh(\bar{r}/R)$. We have (arbitrarily) defined the fixed $\bar{r}$ coordinate of an object at rest with respect to the galactic "gas" to be equal to its radial distance from the origin at the present time, implying that the scale factor a has the value 1 at the present time. In the last chapter, we used the Einstein equation to determine the possible solutions for $q(\bar{r})$: our goal in this chapter is to use the Einstein equation to link these solutions to the total energy density of the universe and to solve for the time-dependent scale factor $a(t)$.

The Einstein Equation Revisited. If we substitute equations 25.5 into the Einstein equation components given by equations 25.12, we get

$$\frac{3\ddot{a}}{a} = R^t{}_t = 8\pi G(T^t{}_t - \tfrac{1}{2}\delta^t{}_t T) = -4\pi G(\rho_0 + 3p_0) + \Lambda \tag{26.2a}$$

$$\frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a^2} - \frac{2q''}{qa^2} = R^{\bar{r}}{}_{\bar{r}} = 8\pi G(T^{\bar{r}}{}_{\bar{r}} - \tfrac{1}{2}\delta^{\bar{r}}{}_{\bar{r}} T) = 4\pi G(\rho_0 - p_0) + \Lambda \tag{26.2b}$$

where $\dot{a} \equiv da/dt, \ddot{a} \equiv d^2a/dt^2, q'' \equiv d^2q/d\bar{r}^2$, and ρ_0 and p_0 are the energy density and pressure of the galactic "gas" in its own rest frame. But since we know from our previous work that $q(\bar{r}) = R\sin(\bar{r}/R), \bar{r}$, or $R\sinh(\bar{r}/R)$, we know that

$$\frac{2q''}{qa^2} = \begin{Bmatrix} -2/R^2a^2 & \text{for } q = R\sin(\bar{r}/R) \\ 0 & \text{for } q = \bar{r} \\ +2/R^2a^2 & \text{for } q = R\sinh(\bar{r}/R) \end{Bmatrix} \equiv \frac{2K}{a^2}, \quad \text{where } K \equiv \mp\frac{1}{R^2} \tag{26.3}$$

for the three cases, respectively. K is the constant of integration we encountered before (note that $K = 0$ corresponds to the curvature radius R going to infinity). If we substitute this back into equation 26.2*b*, we see that equations 26.2 become

$$\frac{3\ddot{a}}{a} = -4\pi G(\rho_0 + 3p_0) + \Lambda \tag{26.4a}$$

$$\frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a^2} - \frac{2K}{a^2} = 4\pi G(\rho_0 - p_0) + \Lambda \tag{26.4b}$$

(The θ-θ and ϕ-ϕ components of the Einstein equation are the same as equation 26.4*b*: see box 26.1). Note that the energy density ρ_0 and the pressure p_0 of the galactic "gas" depend on time, so equations 26.4 represent two equations in three unknown time-dependent quantities, so we do not yet have enough information to solve these equations.

Local Conservation of Energy. We can get some useful information that we need from the local conservation-of-energy law

$$0 = \nabla_\mu T^{t\mu} = \nabla_\mu T^{\mu t} \tag{26.5}$$

where the last step follows because the stress-energy tensor is symmetric. We can lower the t index and expand the absolute divergence to get

$$0 = \partial_\mu T^\mu{}_t + \Gamma^\mu_{\alpha\mu} T^\alpha{}_t - \Gamma^\beta_{t\mu} T^\mu{}_\beta \tag{26.6}$$

You can use the Diagonal Metric Worksheet to compute the necessary Christoffel symbols (or even compute them by hand; they are not that difficult). The number of Christoffel symbols you need is greatly reduced by the fact that $T^\mu{}_\nu$ is diagonal. The result (see box 26.2) is the simple relationship

$$\frac{d}{dt}(\rho_0 a^3) = -p_0\frac{d}{dt}(a^3) \tag{26.7}$$

This result has a simple physical interpretation. Remember that the coordinate $\bar{r}$ is comoving with the galactic "gas" in the expanding universe. The number of galaxies enclosed by a given coordinate $\bar{r}$ is therefore constant. But the metric implies that the physical radius of the volume enclosed by a given value of $\bar{r}$ is proportional to a, so the physical volume enclosed is proportional to a^3. Let's say that $V = Ba^3$, where B is the constant of proportionality. The total energy U in this volume is thus $U = \rho_0 Ba^3$. Equation 26.7 can therefore be written

$$\frac{dU}{dt} = -p_0 B \frac{d}{dt} a^3 \quad \Rightarrow \quad dU = -p_0 dV \tag{26.8}$$

This is the first law of thermodynamics for the universal "gas" inside the volume V: the change in the total energy inside the volume has to be equal to the work energy flowing into the volume. (Note that there can be no heat flow across the boundary because the universe is homogeneous, so all points in space at a given instant of cosmic time t have the same temperature.)

However, it should be noted that equation 26.5 is a consequence of the definition of the Einstein equation, so it does not actually tell us anything that is not already implicit in equations 26.4 (which also specify the implications of the Einstein equation). So while we will find equation 26.7 helpful in solving the field equations in a moment, it does not provide the missing information we need.

Equations of State. What we really need to complete the solution of equations 26.4 is how the pressure p_0 of the galactic "gas" depends on its density ρ_0. An equation that specifies $p(\rho)$ is called an **equation of state**.

In general, the "stuff" in the universe has three important components, non-relativistic *matter*, relativistic *radiation*, and *vacuum energy*. The **matter** component is (now) represented by galaxies and the dark matter that accompanies them. The measured random velocities of galaxies with respect to each other are on the order of 100 km/s, which (though large by human standards) is very small compared to the speed of light. The pressure of the universal "gas" whose "molecules" are galaxies will thus be negligible compared to its energy density, and we can accurately model this component as if it were pressureless dust. Equation 26.7 implies that

$$\frac{d}{dt}(\rho_m a^3) = 0 \quad \Rightarrow \quad \rho_m a^3 = \text{const.} = \rho_{m0} \quad \text{(for } \textbf{matter}\text{)} \tag{26.9a}$$

where ρ_m is the portion of the total energy density ρ_0 that is matter and ρ_{m0} is that density at the present time (note that $a = 1$ at the present time).

However, some of the "stuff" of the universe consists of photons, neutrinos, and other highly relativistic particles. As discussed in problem P20.4, the pressure p_r of a photon gas is related to its energy density ρ_r as follows: $p_r = \frac{1}{3}\rho_r$. If you plug this relationship back into equation 26.7, you can show (see box 26.3) that

$$\rho_r a^4 = \text{const.} = \rho_{r0} \quad \text{(for } \textbf{radiation}\text{)} \tag{26.9b}$$

where ρ_{r0} is the present density of radiation.

Now, a photon gas in thermal equilibrium with its surroundings at an absolute temperature T has an energy density that is proportional to T^4 (this is a consequence of the Stefan-Boltzmann law). This means that the effective temperature of the photon gas (and whatever is in thermal equilibrium with it) varies as follows:

$$Ta = \text{const.} \tag{26.10}$$

In other words, the temperature of any "radiation" component of the universe varies in *inverse* proportion to the universe's scale a.

Finally, there is the **vacuum energy**. As discussed in chapter 21, we can treat the cosmological constant term as if it were a type of energy that we can include on the stress-energy side of the Einstein equation. In what follows, it will help us to treat the density of this energy like that of other sources. According to equation 21.21, the effective stress-energy tensor for this vacuum energy term is

$$T^{\mu\nu} = -g^{\mu\nu}\frac{\Lambda}{8\pi G} \quad \Rightarrow \quad \rho_v \equiv T^{tt} = -g^{tt}\frac{\Lambda}{8\pi G} = +\frac{\Lambda}{8\pi G} \tag{26.11}$$

We see that the vacuum energy density is constant and so does *not* vary with a.

Note also that if we consider the *pressure* of the vacuum to be

$$p_v = (T^{\bar{r}}{}_{\bar{r}})_{\text{vac}} = -\frac{\Lambda}{8\pi G}\delta^{\bar{r}}{}_{\bar{r}} = -\frac{\Lambda}{8\pi G} \tag{26.12}$$

(see equation 25.9), then $\rho_v - p_v = \Lambda/4\pi G$, so the right side of equation 26.4*b* can be written as $4\pi G(\rho_{0,\text{tot}} - p_{0,\text{tot}})$, where $\rho_{0,\text{tot}}$ and $p_{0,\text{tot}}$ include contributions from matter, radiation, *and* vacuum. We see from that equation that the *difference* $\rho_{0,\text{tot}} - p_{0,\text{tot}}$ (by specifying K) uniquely determines the type and magnitude of spatial curvature here, just as that difference does in the weak-field limit.

The Friedman Equation. Now we are finally ready to finish our solution of the Einstein equation. If you add the negative of equation 26.4*a* to 3 times equation 26.4*b* (as discussed in box 26.4), you will find that the terms involving $\ddot{a}/a$ cancel on the right and the pressure terms cancel on the left, leaving the simpler equation

$$\dot{a}^2 - \frac{8\pi G}{3}(\rho_m + \rho_r + \rho_v)a^2 = K \tag{26.13}$$

where I have written $\rho_0 = \rho_m + \rho_r$ and used equation 26.11 to express the vacuum energy term as an energy density. This is the **Friedman equation** for the time-evolution of the universe.

There is some hidden time dependence in this equation, because the densities of matter and radiation depend on a and thus on time. We can use equations 26.9 to make this dependence explicit. Remember that we have defined the value of a to be unity at the present time. Using this notation and equations 26.9, we can write equation 26.13 in the form

$$\dot{a}^2 - \frac{8\pi G}{3}\left(\frac{\rho_{m0}}{a^3} + \frac{\rho_{r0}}{a^4} + \rho_v\right)a^2 = K \tag{26.14}$$

where ρ_{m0} and ρ_{r0} are the matter and energy densities at the present (ρ_v is constant in time, so $\rho_{v0} = \rho_v$). If we know the values of K, ρ_{m0}, ρ_{r0}, and ρ_v, we can in principle solve this differential equation for $a(t)$.

The Hubble Parameter. While we might be able to measure the densities of matter, radiation, and vacuum in the present universe, it is hard to see how we might determine K and a. There is a clever way to address this problem, though.

We saw earlier that the distance from the origin to a particular galaxy at a given fixed value of $\bar{r}$ is given by $d = a\bar{r}$. The rate at which the distance to that galaxy increases due to the expansion of the universe (as reflected by the increase in the value of the universal scale factor a) is thus $\dot{a}\bar{r}$. We can interpret this rate of increase of distance as a recessional velocity v. So at any instant of cosmological time t,

$$v \equiv \dot{a}\bar{r} = \frac{\dot{a}}{a}(a\bar{r}) = \frac{\dot{a}}{a}d = Hd \quad \text{where } H \equiv \frac{\dot{a}}{a} \tag{26.15}$$

H is therefore the Hubble "constant" at that t. Note that H is *not* generally constant with time. It only *appears* constant if we limit ourselves to observing the motion of relatively nearby galaxies, so that the difference in the cosmological t between light's departure from a galaxy and its detection on earth is tiny compared to the age of the universe. Therefore, I will call H the **Hubble parameter**.

However, we can measure H at the *present* time by examining the distances and apparent recessional velocity of relatively nearby galaxies. The present value of H is therefore $H_0 \equiv \dot{a}_0/a_0 = \dot{a}_0$ (since $a_0 \equiv 1$). I will call H_0 (the present value of the Hubble parameter H) the **Hubble constant**. If we divide both sides of equation 26.14 by $\dot{a}^2$ and evaluate it at the present time, we get (see box 26.5)

$$1 - \frac{8\pi G}{3H_0^2}(\rho_{m0} + \rho_{r0} + \rho_v) = \frac{K}{H_0^2} \tag{26.16}$$

The Critical Density. Now, $|K| = 1/R^2$, where R is the scale of the universe's spatial curvature and the sign of K determines the *type* of that curvature. We see here that both the value and sign depends on how the present total energy density of the universe compares to the value of $3H_0^2/8\pi G$. If $\rho_{\text{tot}} \equiv \rho_{m0} + \rho_{r0} + \rho_v > 3H_0^2/8\pi G$, then K is negative, meaning that the geometry of the spatial part of the universe is spherical, with radius $aR = a/|K|^{1/2}$. If $\rho_{\text{tot}} < 3H_0^2/8\pi G$, then K is positive, meaning that the universe's spatial geometry is like that of a saddle surface, with aR again specifying roughly the radial scale where this curvature becomes important. If $\rho_{\text{tot}} = 3H_0^2/8\pi G$, then K is zero, and the universe has a flat spatial geometry. We therefore define the **critical density** ρ_c for the universe at the present to be

$$\rho_c \equiv \frac{3H_0^2}{8\pi G} \tag{26.17}$$

and compare the present energy densities of matter, radiation, and the vacuum to this critical density by defining the unitless ratios

$$\Omega_m \equiv \frac{\rho_{m0}}{\rho_c}, \quad \Omega_r \equiv \frac{\rho_{r0}}{\rho_c}, \quad \Omega_v \equiv \frac{\rho_v}{\rho_c} \tag{26.18}$$

Then we can rewrite equation 26.16 in the form

$$1 - (\Omega_m + \Omega_r + \Omega_v) = \frac{K}{H_0^2} \equiv \Omega_k \tag{26.19}$$

Therefore, if we can measure the present density of matter, radiation, and vacuum energy, and we know the Hubble constant H_0, then we can determine the present value of the **curvature parameter** Ω_k. The sign of Ω_k determines the spatial curvature of the universe just as the sign of K does: if Ω_k is positive, the universe's spatial geometry is saddle-shaped. If Ω_k is negative, then the universe's spatial geometry is spherical. If $\Omega_k = 0$, then $K = 0$, and the universe's spatial geometry is flat.

An Equation of Motion for the Universe. If you divide both sides of equation 26.14 by $\dot{a}_0^2 = H_0^2$ and use equations 26.17 through 26.19, you can show (see box 26.6) that we can express the Friedman equation in the form

$$\left(\frac{1}{H_0}\frac{da}{dt}\right)^2 = \Omega_k + \frac{\Omega_m}{a} + \frac{\Omega_r}{a^2} + \Omega_v a^2 \tag{26.20}$$

We can in principle solve this equation for $a(t)$ at all times, thereby comparing the scale of the universe at any time as a fraction or multiple of the present scale. Taking the absolute value of equation 26.19 and using equation 26.3 allow us to determine the spatial comoving curvature scale R given H_0, Ω_m, Ω_r, and Ω_v:

$$R = \frac{1}{H_0\sqrt{|\Omega_k|}} = \frac{1}{H_0\sqrt{|1 - \Omega_m - \Omega_r - \Omega_v|}} \tag{26.21}$$

($K = 0$ corresponds to infinite curvature scale, meaning that space is flat.) From these last two equations we see that *the four parameters* H_0, Ω_m, Ω_r, and Ω_v *completely determine the evolution and spatial geometry of the universe.*

Equation 26.20 has the form of a one-dimensional conservation of energy equation where the $(1/H_0)^2(da/dt)^2$ term is the kinetic energy, the curvature parameter plays the role of the conserved total energy, and the remaining terms (when negated) play the role of an a-dependent potential energy. Interpreted this way, we see that in determining how a evolves with time t, matter density acts like a simple attractive gravitational force, radiation acts like the potential for an attractive $1/a^3$-dependent force, and the vacuum energy acts like a repulsive spring-like force. This means that you can predict the dynamical behavior of the universe by drawing an effective potential energy graph that expresses these ideas. You can practice this by working through box 26.7, where you will consider the possible behaviors for a matter-dominated universe ($\Omega_r \approx 0$, $\Omega_v \approx 0$) for values of Ω_m both greater than and less than 1.

BOX 26.1 The Other Components of the Einstein Equation

According to equation 25.5*c*,

$$R^{\theta}{}_{\theta} = R^{\phi}{}_{\phi} = 2\frac{\dot{a}^2}{a^2} + \frac{\ddot{a}}{a} + \frac{1}{a^2 q^2}\left[1 - (q')^2 - qq''\right] \tag{26.22}$$

Equation 25.12*b* tells us that the Einstein equation implies that

$$R^{\theta}{}_{\theta} = R^{\phi}{}_{\phi} = 4\pi G(\rho_0 - p_0) + \Lambda \tag{26.23}$$

One can see that that the θ-θ and ϕ-ϕ components of the Einstein equation yield only one distinct differential equation connecting the metric functions a and q with ρ_0, p_0 and Λ. Moreover, you can show that if $q = R\sin(\bar{r}/R), \bar{r}$, or $R\sinh(\bar{r}/R)$, then this differential equation becomes equivalent to equation 26.4*b*, repeated here for convenience:

$$\frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a^2} - \frac{2K}{a^2} = 4\pi G(\rho_0 - p_0) + \Lambda \tag{26.4br}$$

Exercise 26.1.1. Verify that the last statement is true.

BOX 26.2 Consequences of Local Energy/Momentum Conservation

According to equation 25.9, the universal stress-energy tensor is diagonal and has components $T^t{}_t = -\rho_0 - \Lambda/8\pi G, T^{\bar{r}}{}_{\bar{r}} = T^\theta{}_\theta = T^\phi{}_\phi = p_0 - \Lambda/8\pi G$. Using this and the Diagonal Metric Worksheet, you can prove that $d(\rho_0 a^3)/dt = -p_0 d(a^3)/dt$ (equation 26.7) follows from $0 = \partial_\mu T^\mu{}_t + \Gamma^\mu_{\alpha\mu} T^\alpha{}_t - \Gamma^\beta_{t\mu} T^\mu{}_\beta$ (equation 26.6).

Exercise 26.2.1. Verify this. (*Hint:* You should find that you need to calculate only Christoffel symbols of the form Γ^i_{tj}, where i and j are spatial indices. These are very easy to calculate from the definition of the Christoffel symbols if you don't want to bother with the Diagonal Metric Worksheet.)

BOX 26.3 Deriving the Density/Scale Relationship for Radiation

Exercise 26.3.1. Show that $d(\rho_0 a^3)/dt = -p_0 d(a^3)/dt$ (equation 26.7) and the relationship $p_r = \frac{1}{3}\rho_r$ implies $\rho_r a^4$ = constant (equation 26.9*b*).

BOX 26.4 Deriving the Friedman Equation

In equations 26.4 (repeated here), we saw that the Einstein equation becomes

$$\frac{3\ddot{a}}{a} = -4\pi G(\rho_0 + 3p_0) + \Lambda \qquad (26.4ar)$$

$$\frac{\ddot{a}}{a} + 2\frac{\dot{a}^2}{a^2} - \frac{2K}{a^2} = 4\pi G(\rho_0 - p_0) + \Lambda \qquad (26.4br)$$

Exercise 26.4.1. Add the negative of equation 26.4*a* to three times equation 26.4*b* to derive the Friedman equation $\dot{a}^2 - \frac{8}{3}\pi G(\rho_0 + \rho_v)a^2 = K$ (equation 26.13).

BOX 26.5 The Friedman Equation for the Present Time

Equation 26.14 (repeated here for convenience) says that

$$\dot{a}^2 - \frac{8\pi G}{3}\left(\frac{\rho_{m0}}{a^3} + \frac{\rho_{r0}}{a^4} + \rho_v\right)a^2 = K \qquad (26.14r)$$

If we divide both sides of this by $\dot{a}^2$ and evaluate at the present time using using $H_0 \equiv \dot{a}_0$ and $a_0 = 1$, we get equation 26.16 (repeated below).

$$1 - \frac{8\pi G}{3H_0^2}(\rho_{m0} + \rho_{r0} + \rho_v) = \frac{K}{H_0^2} \qquad (26.16r)$$

Exercise 26.5.1. Verify this.

BOX 26.6 Deriving the Friedman Equation in Terms of the Omegas

Equation 26.20 (repeated here for convenience) claims that

$$\left(\frac{1}{H_0}\frac{da}{dt}\right)^2 = \Omega_k + \frac{\Omega_m}{a} + \frac{\Omega_r}{a^2} + \Omega_v a^2 \qquad (26.20r)$$

Exercise 26.6.1. Show that if we divide both sides of equation 26.14 (see above) by $\dot{a}_0^2 = H_0^2$ and use $\rho_c \equiv 3H_0^2/8\pi G$, $\Omega_m \equiv \rho_{m0}/\rho_c$, $\Omega_r \equiv \rho_{r0}/\rho_c$, $\Omega_v \equiv \rho_v/\rho_c$, and $K/H_0^2 \equiv \Omega_k$ (equations 26.17–26.19), we get the above.

BOX 26.7 The Behavior of a Matter-Dominated Universe

Assume that the universe always has been dominated by matter ($\Omega_r \approx \Omega_v \approx 0$) throughout its history. Equation 26.20 then becomes

$$\left(\frac{1}{H_0}\frac{da}{dt}\right)^2 - \frac{\Omega_m}{a} = \Omega_k \tag{26.24}$$

Exercise 26.7.1. Interpreting this as a one-dimensional "conservation of energy" equation, plot a potential energy graph, and use the graph to qualitatively describe the evolution of a in the case where $\Omega_m > 1$ and when $\Omega_m < 1$. Also describe how the time evolution of the universe is connected with its spatial curvature in this case. (*Hint:* What does $\Omega_m > 1$ mean for the value of Ω_k? See equation 26.19.)

HOMEWORK PROBLEMS

P26.1 Use a "potential energy graph" approach to discuss the qualitative behavior of a radiation-dominated universe where $\Omega_m \approx \Omega_v \approx 0$. Describe the evolution of the universal scale a in the case where $\Omega_r > 1$ and when $\Omega_r < 1$.

P26.2 Use a "potential energy graph" approach to discuss the qualitative behavior of an empty vacuum-dominated universe where $\Omega_m \approx \Omega_r \approx 0$. Qualitatively describe the evolution of the universal scale a in the cases where $\Omega_v > 1$ and $\Omega_v < 1$. For what such universes will there be a Big Bang? (Empty vacuum-dominated universes are called **Lemaître universes**.)

P26.3 When Einstein first applied general relativity to the problem of cosmology in 1917, it was reasonable to consider the universe to be a homogeneous, isotropic, and *static* collection of stars similar to the stars near to the earth. Einstein could find a static solution to the Einstein equation $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi G T_{\mu\nu}$ if and only if he added the "cosmological constant" term $\Lambda g_{\mu\nu}$ to the left side of the equation. As we saw in chapter 21, we now consider this "cosmological constant" term to be instead a "vacuum energy" term that we add to the equation's *right* side.

a. Use a "potential energy graph" approach to argue that one can indeed find a static solution to the Friedman equation in the form given in 26.14 if ρ_{m0} and ρ_v are both nonzero, $\rho_{r0} \approx 0$, and K has exactly the right value.

b. Assuming that $a = 1$ at the present, express ρ_v and K in terms of ρ_{m0} assuming that the universe is static at present and ρ_{r0} is negligibly small.

c. A plausible mass density for the universe in Einstein's time might have been the approximate local density of stars near the earth, which is very roughly 0.05 solar masses per cubic parsec. Determine the value of ρ_{m0} corresponding to this value, and use your result from part b to determine the values of ρ_v and K.

d. Is your value for ρ_v comfortably smaller than the upper limit established by solar system measurements (see equation 21.36 in problem P21.1)?

e. You should have found K to be negative for this hypothetical static universe, implying that its spatial geometry is spherical. Assuming that it is also *topologically* spherical, find its radius in light-years, its total volume in cubic light-years, and the total mass of matter in solar masses. (*Hint:* Be sure to use the metric to find the volume, which you should find to be $V = 2\pi^2 R^3$.)

f. However, use the "potential energy diagram" to argue that this static universe is *unstable* (something that wasn't initially clear to Einstein).

As we saw in chapter 24, Lemaître and Hubble established in 1927 and 1929, respectively, that the universe was in fact expanding. In 1931, Einstein formally abandoned the cosmological constant term, later calling it "the biggest blunder of his life" according to George Gamow in his autobiography *My Worldline* (Viking Press, 1970, p. 44).

P26.4 Consider a model of the universe where there is no vacuum energy, only matter and radiation. Argue that the age of such a universe must be less than H_0^{-1} = 13.7 Gy. [*Hint:* Use a "potential energy graph" approach to determine qualitatively how a depends on time, and sketch a qualitative graph of $a(t)$. Note that H_0 is related to the present slope of such a graph.]

P26.5 Consider the case of an empty, vacuum-dominated universe where $\Omega_m \approx \Omega_r \approx 0$ and $\Omega_v = 1$. (Such a universe is called a **De Sitter universe**.)

a. Show that this universe expands exponentially:

$$a(t) = e^{+H_0(t-t_0)} \quad \text{or} \quad a(t) = e^{-H_0(t-t_0)} \tag{26.25}$$

b. Argue that for this to be consistent with an expanding universe, we must choose the first solution.

c. What is the age of the universe in this case?

d. Show that the Hubble parameter $H \equiv \dot{a}/a$ in such a universe happens to actually be constant in time.

P26.6 (Important!) Argue that the universe's present age is

$$t_0 = \frac{1}{H_0}\int_0^1 \frac{da}{\sqrt{\Omega_k + \Omega_r a^{-2} + \Omega_m a^{-1} + \Omega_v a^2}} \tag{26.26}$$

P26.7 Consider a universe whose metric is

$$ds^2 = -dt^2 + \left(\frac{t}{t_0}\right)(dx^2 + dy^2 + dz^2) \tag{26.27}$$

where t_0 is some constant.

a. Explain how we can interpret this metric as a special case of the general universal metric given by equation 26.1. What is $a(t)$ in this case? What is the age of the universe when $a = 1$?

b. Is the spatial geometry of this universe saddle-like, flat, or spherical? Explain your reasoning.

c. Is this universe radiation, matter, or vacuum dominated? Explain your reasoning.

P26.8 When I first learned cosmology in the 1970s, both the average matter density of the universe and the value of the Hubble constant H_0 were so poorly known that it was possible (though improbable) that $\Omega_m > 1$. This was also during the time that most physicists believed that $\Omega_v = 0$ (and, more correctly, that Ω_r is negligible). Let's consider the evolution of such a super-critical "matter-only" universe.

a. Argue that the spatial geometry of such a universe is spherical ("closed").

P26.8 (continued)

b. Equation 26.20 for the time evolution of the universe in this case (where $\Omega_r = \Omega_v = 0$) is still a nonlinear differential equation. To solve such an equation, one must use intelligent guessing, trickery, or both. Usually people use intelligent guessing to get the solution the first time and then invent clever tricks to find the solution more elegantly after it is known. We will use such a trick in this case. Let's define a "time parameter" ψ such that

$$(\Omega_m - 1)^{1/2} H_0 dt = a\, d\psi \qquad (26.28)$$

and $\psi = 0$ at $t = 0$. Show that if we consider a to be a function of of the new parameter ψ, we can re-express equation 26.20 in the form

$$\left(\frac{da}{d\psi}\right)^2 + a^2 = \frac{\Omega_m}{\Omega_m - 1} a \qquad (26.29)$$

c. Take the ψ-derivative of both sides of this expression to show that

$$\frac{d^2 a}{d\psi^2} + a = \frac{1}{2}\frac{\Omega_m}{\Omega_m - 1} \qquad (26.30)$$

d. This is the harmonic oscillator equation with a constant driving term. In a differential equations or mechanics course, you may have learned that the general solution to such an equation is the most general solution to the *homogeneous* equation

$$\frac{d^2 a}{d\psi^2} + a = 0 \qquad (26.31)$$

plus any *particular* solution to the *full* equation. Argue that $a = \frac{1}{2}\Omega_m / (\Omega_m - 1) = \text{constant} \equiv A$ is a solution to the full equation.

e. The solution to the homogeneous harmonic oscillator equation is $a = B \sin\psi + C \cos\psi$, where B and C are constants determined by initial conditions. Therefore, the general solution to equation 26.30 is

$$a(\psi) = A + B\sin\psi + C\cos\psi \qquad (26.32)$$

Argue that requiring that $a \to 0$ as $\psi \to 0$ puts no constraints on B but requires that $C = -A$.

f. Argue, however, that this solution will not satisfy the *original* relation in equation 26.29 as $\psi \to 0$ unless we also have $B = 0$.

g. Use equation 26.28 to determine t in terms of H_0, Ω_m and ψ. Your answer to this part and equation 26.32 with $B = 0$ and $A = -C = \frac{1}{2}\Omega_m / (\Omega_m - 1)$ provide a parametric solution for $a(t)$ in terms of the parameter ψ.

h. Argue that such a universe expands, reaches a maximum scale a, and then contracts to a Big Crunch. Is this consistent with the results of box 26.7?

i. If $H_0^{-1} = 13.9$ Gy, and $\Omega_m = 1.10$, how long after the Big Bang does the universe reach its maximum spatial size, and what is the radius of its spherical geometry at that point? How long does the universe last between the Big Bang and the Big Crunch? (*Hint:* See equation 26.21.)

j. Argue that a graph of $Ra(t)$ has the shape of a cycloid, i.e., the path of a point on the rim of a rolling wheel. (*Hint:* Look up "cycloid" online.)

P26.9 Consider a universe where $\Omega_v > 1$ and matter and energy densities are negligible ($\Omega_m \approx 0$ and $\Omega_r \approx 0$). It turns out that such a universe will never have a Big Bang singularity, but will have an instant of maximal (finite) density. Define that instant to be $t = 0$. Assume that observers in this universe at some time t_0 measure the Hubble constant to be $H_0 = (15 \text{ Gy})^{-1}$.

a. Show that for such a universe, $a = b\cosh(\omega t)$, where $b = \sqrt{(\Omega_v - 1)/\Omega_v}$ and $\omega = H_0\sqrt{\Omega_v}$. If this universe is expanding at time t_0, will it ever cease expanding? If so, at what time t?

b. Imagine that observers in this universe determine from observations of their cosmic microwave background that $\Omega_v = 2$. How old is their universe at time t_0?

c. Is the spatial geometry of this universe spherical, flat, or saddle-like?

d. What is the scale factor R of this universe (which is the scale over which the spatial curvature of the universe becomes evident)?

P26.10 We have expressed our equations of motion for the universal scale factor a in terms of the *constants* Ω_m, Ω_r, and Ω_v, which are ratios of the *current* mass, radiation, and vacuum energy densities (respectively) to the *current* critical density. However, a hypothetical observer at a different cosmic time t would determine these constants to have different values. Define

$$\Omega_m(t) = \frac{\rho_m(t)}{\rho_c(t)} = \frac{8\pi G}{3[H(t)]^2}\rho_m(t) \qquad (26.33)$$

where $\rho_m(t)$ is the density of matter at time t, $\rho_c(t)$ is the critical density at time t, and $H(t) \equiv \dot{a}/a$ is the Hubble constant at time t. We can define Ω_r and Ω_v similarly.

a. Show that we can write

$$\Omega_m(t) = \frac{\Omega_m}{a^3}\left(\frac{H_0}{H}\right)^2 \quad \text{and} \quad \Omega_v(t) = \Omega_v\left(\frac{H_0}{H}\right)^2 \qquad (26.34)$$

where Ω_m and Ω_v are the values we would measure.

b. Imagine a vacuum-dominated universe where the current value of $\Omega_v \approx 1$ and Ω_m and Ω_r are both $<< \Omega_v$. Argue that if the value of Ω_m is not *strictly* zero, then observers in the distant past would determine the value of $\Omega_m(t)$ to be greater than Ω_m, while $\Omega_v(t) = \Omega_v$ always. Argue therefore that the approximation that the universe is vacuum dominated must break down at some point sufficiently far in the past. (*Hint:* You can use the results of problem P26.5.)