Introduction to Tensor Calculus

A.V.Smirnov

Contents

Preface

This material offers a short introduction to tensor calculus. It is directed toward students of continuum mechanics and engineers. The emphasis is made on tensor notation and invariant forms. A knowledge of calculus is assumed. A more complete coverage of tensor calculus can be found in [1, 2].

Nomenclature

$A \equiv B$	A is defined as B, or A is equivalent to B
$A_i B_i$	$\equiv \sum_i^3 A_i B_i$. Note: $A_i B_i = A_j B_j$
$\dot{A}$	partial derivative over time: $\frac{\partial A}{\partial t}$
$A_{,i}$	partial derivative over x_i: $\frac{\partial A}{\partial x_i}$
V	control volume
t	time
x_i	i-th component of a coordinate (i=0,1,2), or $x_i \equiv \{x, u, z\}$
RHS	Right-hand-side
LHS	Left-hand-side
PDE	Partial differential equation
..	Continued list of items

There are two aspects of tensors that are of practical and fundamental importance: *tensor notation* and *tensor invariance*. Tensor notation is of great practical importance, since it simplifies handling of complex equation systems. The idea of tensor invariance is of both practical and fundamental importance, since it provides a powerful apparatus to describe non-Euclidean spaces in general and curvilinear coordinate systems in particular.

A definition of a tensor is given in Section 1. Section 2 deals with an important class of Cartesian tensors, and describes the rules of tensor notation. Section 3 provides a brief introduction to general curvilinear coordinates, invariant forms and the rules of covariant differentiation.

1 Coordinates and Tensors

Consider a space of real numbers of dimension n, R^n, and a single real time, t. Continuum properties in this space can be described by arrays of different dimensions, m, such as scalars ($m = 0$), vectors ($m = 1$), matrices ($m = 2$), and general multi-dimensional arrays. In this space we shall introduce a *coordinate system*, $\{x^i\}_{i=1..n}$, as a way of assigning n real numbers[1] for every point of space There can be a variety of possible coordinate systems. A general *transformation rule* between the coordinate systems is

$$\tilde{x}^i = \tilde{x}^i(x^1..x^n) \tag{1}$$

Consider a small displacement dx^i. Then it can be transformed from coordinate system x^i to a new coordinate system $\tilde{x}^i$ using the partial differentiation rules applied to (1):

$$d\tilde{x}^i = \frac{\partial \tilde{x}^i}{\partial x^j} dx^j \tag{2}$$

This transformation rule[2] can be generalized to a set of vectors that we shall call *contravariant vectors:*

$$\tilde{A}^i = \frac{\partial \tilde{x}^i}{\partial x^j} A^j \tag{3}$$

[1] Super-indexes denote components of a vector ($i = 1..n$) and not the power exponent, for the reason explained later (Definition 1.1)

[2] The repeated indexes imply summation (See. Proposition 21)

That is, a contravariant vector is defined as a vector which transforms to a new coordinate system according to (3). We can also introduce the *transformation matrix* as:

$$a^i_j \equiv \frac{\partial \tilde{x}^i}{\partial x^j} \tag{4}$$

With which (3) can be rewritten as:

$$A^i = a^i_j A^j \tag{5}$$

Transformation rule (3) will not apply to all the vectors in our space. For example, a partial derivative $\partial/\partial x_i$ will transform as:

$$\frac{\partial}{\partial \tilde{x}^i} = \frac{\partial}{\partial \tilde{x}^i}\frac{\partial x^j}{\partial x^j} = \frac{\partial x^j}{\partial \tilde{x}^i}\frac{\partial}{\partial x^j} \tag{6}$$

that is, the transformation coefficients are the other way up compared to (2). Now we can generalize this transformation rule, so that each vector that transforms according to (6) will be called a *Covariant vector:*

$$\tilde{A}_i = \frac{\partial x^j}{\partial \tilde{x}^i} A_j \tag{7}$$

This provides the reason for using lower and upper indexes in a general tensor notation.

Definition 1.1 Tensor

Tensor of order m is a set of n^m numbers identified by m integer indexes. For example, a 3rd order tensor A can be denoted as A_{ijk} and an m-order tensor can be denoted as $A_{i_1..i_m}$. Each index of a tensor changes between 1 and n. For example, in a 3-dimensional space (n=3) a second order tensor will be represented by $3^2 = 9$ components.

Each index of a tensor should comply to one of the two transformation rules: (3) or (7). An index that complies to the rule (7) is called a **covariant index** *and is denoted as a sub-index, and an index complying to the transformation rule (3) is called a* **contravariant index** *and is denoted as a super-index.*

Each index of a tensor can be covariant or a contravariant, thus tensor A^k_{ij} is a 2-covariant, 1-contravariant tensor of third order.

From this relation and the independence of coordinates (9) it follows that $a^i_j b^j_k = b^i_j a^j_k = \delta_{ik}$, namely:

$$
\begin{aligned}
a^i_j b^j_k &= \frac{\partial \tilde{x}^i}{\partial x^j}\frac{\partial x^j}{\partial \tilde{x}^k} \\
&= \frac{\partial x^j}{\partial x^j}\frac{\partial \tilde{x}^i}{\partial \tilde{x}^k} = \frac{\partial \tilde{x}^i}{\partial \tilde{x}^k} = \delta_{ik}
\end{aligned}
\tag{13}
$$

2 Cartesian Tensors

Cartesian tensors are a sub-set of general tensors for which the transformation matrix (4) satisfies the following relation:

$$
a^k_i a^k_j = \frac{\partial \tilde{x}^k}{\partial x^i}\frac{\partial \tilde{x}^k}{\partial x^j} = \delta_{ij} \tag{14}
$$

For Cartesian tensors we have

$$
\frac{\partial \tilde{x}^i}{\partial x^k} = \frac{\partial x^k}{\partial \tilde{x}^i} \tag{15}
$$

(see Problem 4.3), which means that both (5) and (6) are transformed with the same matrix a^i_k. This in turn means that the difference between the covariant and contravariant indexes vanishes for the Cartesian tensors. Considering this we shall only use the sub-indexes whenever we deal with Cartesian tensors.

2.1 Tensor Notation

Tensor notation simplifies writing complex equations involving multi-dimensional objects. This notation is based on a set of *tensor rules.* The rules introduced in this section represent a complete set of rules for Cartesian tensors and will be extended in the case of general tensors (Sec.3). The importance of tensor rules is given by the following general remark:

Remark 2.1 **Tensor rules** *Tensor rules guarantee that if an expression follows these rules it represents a tensor according to Definition 1.1.*

Tensors are usually functions of space and time:

$$A_{i_1..i_m} = A_{i_1..i_m}(x^1..x^n, t)$$

which defines a tensor field, i.e. for every point x^i and time t there are a set of m^n nubers $A_{i_1..i_m}$.

Remark 1.2 Tensor character of coordinate vectors

Note, that the coordinates x^i are not tensors, since generally, they are not transformed as (5). Transformation law for the coordinates is actually given by (1). Nevertheless, we shall use the upper (contravariant) indexes for the coordinates.

Definition 1.3 Kronecker delta tensor

Second order delta tensor, δ_{ij} is defined as

$$\begin{aligned} i = j \;\Rightarrow\; \delta_{ij} = 1 \\ i \neq j \;\Rightarrow\; \delta_{ij} = 0 \end{aligned} \tag{8}$$

From this definition and since coordinates x^i are independent of each other it follows that:

$$\frac{\partial x^i}{\partial x^j} = \delta_{ij} \tag{9}$$

Corollary 1.4 Delta product

From the definition (1.3) and the summation convention (21), follows that

$$\delta_{ij} A_j = A_i \tag{10}$$

Assume that there exists the transformation inverse to (5), which we call b^i_j:

$$dx^i = b^i_j d\tilde{x}^j \tag{11}$$

Then by analogy to (4) b^i_j can be defined as:

$$b^i_j = \frac{\partial x^i}{\partial \tilde{x}^j} \tag{12}$$

Thus, following tensor rules, one can build tensor expressions that will preserve tensor properties of coordinate transformations (Definition 1.1) and coordinate invariance (Section 3).

Tensor rules are based on the following definitions and propositions.

Definition 2.2 Tensor terms

A tensor term is a product of tensors.

For example:

$$A_{ijk}B_{jk}C_{pq}E_qF_p \tag{16}$$

Definition 2.3 Tensor expression

Tensor expression is a sum of tensor terms. For example:

$$A_{ijk}B_{jk} + C_iD_{pq}E_qF_p \tag{17}$$

Generally the terms in the expression may come with plus or minus sign.

Proposition 2.4 Allowed operations

The only allowed algebraic operations in tensor expressions are the addition, subtraction and multiplication. Divisions are only allowed for constants, like $1/C$. If a tensor index appears in a denominator, such term should be redefined, so as not to have tensor indexes in a denominator. For example, $1/A_i$ should be redefined as: $B_i \equiv 1/A_i$.

Definition 2.5 Tensor equality

Tensor equality is an equality of two tensor expressions.

For example:

$$A_{ij}B_j = C_{ikp}D_kE_p + E_jC_{jki}B_k \tag{18}$$

Definition 2.6 **Free indexes**

A free index is any index that occurs only once in a tensor term. For example, index i is a free index in the term (16).

Proposition 2.7 **Free index restriction**

Every term in a tensor equality should have the same set of free indexes.

For example, if index i is a free index in any term of tensor equality, such as (18), it should be the free index in all other terms. For example

$$A_{ij}B_j = C_jD_j$$

is not a valid tensor equality since index i is a free index in the term on the RHS but not in the LHS.

Definition 2.8 **Rank of a term**

A rank of a tensor term is equal to the number of its free indexes.

For example, the rank of the term $A_{ijk}B_jC_k$ is equal to 1.

It follows from (2.7) that ranks of all the terms in a valid tensor expression should be the same. Note, that the difference between the order and the rank is that the order is equal to the number of indexes of a tensor, and the rank is equal to the number of free indexes in a tensor term.

Proposition 2.9 **Renaming of free indexes**

Any free index in a tensor expression can be named by any symbol as long as this symbol does not already occur in the tensor expression.

For example, the equality

$$A_{ij}B_j = C_iD_jE_j \tag{19}$$

is equivalent to

$$A_{kj}B_j = C_kD_jE_j \tag{20}$$

Here we replaced the free index i with k.

Definition 2.10 Dummy indexes

A dummy index is any index that occurs twice in a tensor term.

For example, indexes j, k, p, q in (16) are dummy indexes.

Proposition 2.11 Summation rule

Any dummy index implies summation, i.e.

$$A_i B_i = \sum_i^n A_i B_i \tag{21}$$

Proposition 2.12 Summation rule exception *If there should be no summation over the repeated indices, it can be indicated by enclosing such indices in parentheses.*

For example, expression:

$$C_{(i)} A_{(i)} B_j = D_{ij}$$

does not imply summation over i.

Corollary 2.13 Scalar product

A scalar product notation from vector algebra: $(A \cdot B)$ is expressed in tensor notation as $A_i B_i$.

The scalar product operation is also called a *contraction of indexes*.

Proposition 2.14 Dummy index restriction

No index can occur more than twice in any tensor term.

Remark 2.15 Repeated indexes

In case if an index occurs more than twice in a term this term should be redefined so as not to contain more than two occurrences of the same index. For example, term $A_{ik} B_{jk} C_k$ should be rewritten as $A_{ik} D_{jk}$, where D_{jk} is defined as $D_{jk} \equiv B_{j(k)} C_{(k)}$ with no summation over k in the last term.

Proposition 2.16 **Renaming of dummy indexes**

Any dummy index in a tensor term can be renamed to any symbol as long as this symbol does not already occur in this term.

For example, term A_iB_i is equivalent to A_jB_j, and so are terms $A_{ijk}B_jC_k$ and $A_{ipq}B_pC_q$.

Remark 2.17 **Renaming rules**

Note that while the dummy index renaming rule (2.16) is applied to each tensor term separately, the free index naming rule (2.9) should apply to the whole tensor expression. For example, the equality (19) above

$$A_{ij}B_j = C_iD_jE_j$$

can also be rewritten as

$$A_{kp}B_p = C_kD_jE_j \tag{22}$$

without changing its meaning.

(See Problem 4.1).

Definition 2.18 **Permutation tensor**

The components of a third order permutation tensor ε_{ijk} are defined to be equal to 0 when any index is equal to any other index; equal to 1 when the set of indexes can be obtained by cyclic permutation of 123; and -1 when the indexes can be obtained by cyclic permutation from 132. In a mathematical language it can be expressed as:

$$\begin{aligned} i = j \cup i = k \cup j = k \ \Rightarrow \ \varepsilon_{ijk} = 0 \\ ijk \in PG(123) \ \Rightarrow \ \varepsilon_{ijk} = 1 \\ ijk \in PG(132) \ \Rightarrow \ \varepsilon_{ijk} = -1 \end{aligned} \tag{23}$$

where $PG(abc)$ is a permutation group of a triple of indexes abc, i.e. $PG(abc) = \{abc, bca, cab\}$. For example, the permutation group of 123 will consist of three combinations: 123, 231 and 312, and the permutation group of 123 consists of 132, 321 and 213.

Corollary 2.19 Permutation of the permutation tensor indexes

From the definition of the permutation tensor it follows that the permutation of any of its two indexes changes its sign:

$$\varepsilon_{ijk} = -\varepsilon_{ikj} \tag{24}$$

A tensor with this property is called *skew-symmetric*.

Corollary 2.20 Vector product

A vector product (cross-product) of two vectors in vector notation is expressed as

$$\vec{A} = \vec{B} \times \vec{C} \tag{25}$$

which in tensor notation can be expressed as

$$A_i = \varepsilon_{ijk} B_j C_k \tag{26}$$

Remark 2.21 Cross product

Tensor expression (26) is more accurate than its vector counterpart (25), since it explicitly shows how to compute each component of a vector product.

Theorem 2.22 Symmetric identity

If A_{ij} is a symmetric tensor, then the following identity is true:

$$\varepsilon_{ijk} A_{jk} = 0 \tag{27}$$

Proof:

From the symmetry of A_{ij} we have:

$$\varepsilon_{ijk} A_{jk} = \varepsilon_{ijk} A_{kj} \tag{28}$$

Let's rename index j into k and k into j in the RHS of this expression, according to rule (2.16):

$$\varepsilon_{ijk} A_{kj} = \varepsilon_{ikj} A_{jk}$$

Using (24) we finally obtain:

$$\varepsilon_{ikj} A_{jk} = -\varepsilon_{ijk} A_{jk}$$

Comparing the RHS of this expression to the LHS of (28) we have:

$$\varepsilon_{ijk} A_{jk} = -\varepsilon_{ijk} A_{jk}$$

from which we conclude that (27) is true.

Theorem 2.23 Tensor identity

The following tensor identity is true:

$$\varepsilon_{ijk}\varepsilon_{ipq} = \delta_{jp}\delta_{kq} - \delta_{jq}\delta_{kp} \tag{29}$$

Proof

This identity can be proved by examining the components of equality (29) component-by-component.

Corollary 2.24 Vector identity

Using the tensor identity (29) it is possible to prove the following important vector identity:

$$\vec{A} \times (\vec{B} \times \vec{C}) = \vec{B}(\vec{A} \cdot \vec{C}) - \vec{C}(\vec{A} \cdot \vec{B}) \tag{30}$$

See Problem 4.4.

2.2 Tensor Derivatives

For Cartesian tensors derivatives introduce the following notation.

Definition 2.25 Time derivative of a tensor

A partial derivative of a tensor over time is designated as

$$\dot{A} \equiv \frac{\partial A}{\partial t}$$

Definition 2.26 **Spatial derivative of a tensor**

A partial derivative of a tensor A over one or its spacial components is denoted as $A_{,i}$:

$$A_{,i} \equiv \frac{\partial A}{\partial x_i} \tag{31}$$

that is, the index of the spatial component that the derivation is done over is delimited by a comma (',') from other indexes. For example, $A_{ij,k}$ is a derivative of a second order tensor A_{ij}.

Definition 2.27 **Nabla**

Nabla operator acting on a tensor A is defined as

$$\nabla_i A \equiv A_{,i} \tag{32}$$

Even though the notation in (31) is sufficient to define the derivative, In some instances it is convenient to introduce the nabla operator as defined above.

Remark 2.28 **Tensor derivative**

In a more general context of non-Cartesian tensors the coordinate independent derivative will have a different form from (31). See the chapter on covariant differentiation in [1].

Remark 2.29 **Rank of a tensor derivative**

The derivative of a zero order tensor (scalar) as given by (31) forms a first order tensor (vector). Generally, the derivative of an m-order tensor forms an $m+1$ order tensor. However, if the derivation index is a dummy index, then the rank of the derivative will be lower than that of the original tensor. For example, the rank of the derivative $A_{ij,j}$ is one, since there is only one free index in this term.

Remark 2.30 **Gradient**

Expression (31) represents a gradient, which in a vector notation is ∇A:

$$\nabla A \;\longrightarrow\; A_{,i}$$

Corollary 2.31 Derivative of a coordinate

From (9) it follows that:

$$x_{i,j} = \delta_{ij} \tag{33}$$

In particular, the following identity is true:

$$x_{i,i} = x_{1,1} + x_{2,2} + x_{3,3} = 1 + 1 + 1 = 3 \tag{34}$$

Remark 2.32 Divergence operator

A divergence operator in a vector notation is represented in a tensor notation as $A_{i,i}$:

$$(\nabla \cdot \vec{A}) \longrightarrow A_{i,i}$$

Remark 2.33 Laplace operator

The Laplace operator in vector notation is represented in tensor notation as $A_{,ii}$:

$$\Delta A \longrightarrow A_{,ii}$$

Remark 2.34 Tensor notation

Examples (2.30), (2.32) and (2.33) clearly show that tensor notation is more concise and accurate than vector notation, since it explicitly shows how each component should be computed. It is also more general since it covers cases that don't have representation in vector notation, for example: $A_{ik,kj}$.

3 Curvilinear coordinates

In this section [3] we introduce the idea of *tensor invariance* and introduce the rules for constructing *invariant forms*.

[3] In this section we reinstall the difference between covariant and contravariant indexes.

3.1 Tensor invariance

The distance between the material points in a Cartesian coordinate system is computed as $dl^2 = dx_i dx_i$. The *metric tensor*, g_{ij} is introduced to generalize the notion of distance (39) to curvilinear coordinates.

Definition 3.1 Metric Tensor

The distance element in curvilinear coordinate system is computed as:

$$dl^2 = g_{ij} dx^i dx^j \tag{35}$$

where g_{ij} is called the metric tensor*.*

Thus, if we know the metric tensor in a given curvilinear coordinate system then the distance element is computed by (35). The metric tensor is defined as a tensor since we need to preserve the *invariance* of distance in different coordinate systems, that is, the distance should be independent of the coordinate system, thus:

$$dl^2 = g_{ij} dx^i dx^j = \tilde{g}_{ij} d\tilde{x}^i d\tilde{x}^j \tag{36}$$

The metric tensor is symmetric, which can be shown by rewriting (35) as follows:

$$g_{ij} dx^i dx^j = g_{ij} dx^j dx^i = g_{ji} dx^i dx^j$$

where we first swapped places of dx^i and dx^j, and then renamed index i into j and j into i. We can rewrite the equality above as:

$$g_{ij} dx^i dx^j - g_{ji} dx^i dx^j = (g_{ij} - g_{ji}) dx^i dx^j = 0$$

Since the equality above should hold for any $dx^i dx^j$, we get:

$$g_{ij} = g_{ji} \tag{37}$$

The metric tensor is also called the *fundamental tensor.* The inverse of the metric tensor is also called the *conjugate metric tensor*, g^{ij}, which satisfies the relation:

$$g^{ik}g_{kj} = \delta_{ij} \tag{38}$$

Let x^i be a Cartesian coordinate system, and $\tilde{x}^j$ - the new curvilinear coordinate system. Both systems are related by transformation rules (5) and (11). Then from (36) we get:

$$dl^2 = dx^i dx^i = \frac{\partial x^i}{\partial \tilde{x}^j} d\tilde{x}^j \frac{\partial x^i}{\partial \tilde{x}^k} d\tilde{x}^k = \frac{\partial x^i}{\partial \tilde{x}^j} \frac{\partial x^i}{\partial \tilde{x}^k} d\tilde{x}^j d\tilde{x}^k \tag{39}$$

When we transform from a Cartesian to curvilinear coordinates the metric tensor in curvilinear coordinate system, $\tilde{g}_{ij}$ can be determined by comparing relations (39) and (35):

$$\tilde{g}_{ij} = \frac{\partial x^k}{\partial \tilde{x}^i} \frac{\partial x^k}{\partial \tilde{x}^j} \tag{40}$$

Using (38) we can also find its inverse as:

$$\tilde{g}^{ij} = \frac{\partial \tilde{x}^i}{\partial x^k} \frac{\partial \tilde{x}^j}{\partial x^k} \tag{41}$$

Using these expression one can compute g_{ij} and g^{ij} in various curvilinear coordinate systems (see Problem 4.6).

Definition 3.2 Conjugate tensors

For each index of a tensor we introduce the conjugate tensor where this index is transfered to its counterpart (covariant/contravariant) using the relations:

$$A^i = g^{ij} A_j \tag{42}$$

$$A_i = g_{ij} A^j \tag{43}$$

Conjugate tensor is also called the *associate tensor*. Relations (42), (43) are also called as operations of *raising/lowering of indexes*.

Remark 3.3 Tensor invariance

Since the transformation rules defined by (1.1) have a simple multiplicative character, any tensor expression should retain it's original form under transformation into a new coordinate system. Thus if an expression is given in a tensor form it will be invariant *under coordinate transformations.*

Not all the expressions constructed from tensor terms in curvilinear coordinates will be tensors themselves. For example, if vectors A_i and B_i are tensors, then $A_i B_i$ is not generally a tensor[4]. However, if we consider the same operation on a contravariant tensor A^i and a covariant tenso B_i then the product will form an invariant:

$$\bar{A}^i \bar{B}_i = A^i B_i \tag{44}$$

Thus in curvilinear coordinates we have to refine the definition of the scalar product (Corollary 2.13) or the index contraction operation to make it invariant (Problem 4.12).

Definition 3.4 **Invariant Scalar Product**

The invariant form *of the scalar product between two covariant vectors A_i and B_i is $g^{ij} A_i B_j$. Similarly, the invariant form of a scalar product between two contravariant vectors A^i and B^i is $g_{ij} A^i B^j$, where g_{ij} is the metric tensor (40) and g^{ij} is its conjugate (38).*

Corollary 3.5 **Two forms of a scalar product**

According to (42), (43) the scalar product can be represented by two invariant forms: $A^i B_i$ and $A_i B^i$. It can be easily shown that these two forms have the same values (see Problem 4.12).

Corollary 3.6 **Rules of invariant expressions**

To build invariant tensor expressions we add two more rules to Cartesian tensor rules outlined in Section 2.1:

1. *Each free index should keep its vertical position in every term, i.e. if the index is covariant in one term it should be covariant in every other term, and vise versa.*

2. *Every pair of dummy indexes should be complementary, that is one should be covariant, and another contravariant.*

For example, a Cartesian formulation of a *momentum equation* for an incompressible viscous fluid is

$$\dot{u}_i + u_k u_{i,k} = -\frac{P_{,i}}{\rho} + \nu \tau_{ik,k}$$

[4] For Cartesian tensors any product of tensors will always be a tensor, but this is not so for general tensors

The invariant form of this equation is:

$$\dot{u}_i + u^k u_{i,k} = -\frac{P_{,i}}{\rho} + \nu \tau^k_{i,k} \tag{45}$$

where the rising of indexes was done using relation (42): $u^k = g^{kj} u_j$, and $\tau^k_i = g^{kj} \tau_{ij}$.

3.2 Covariant differentiation

A simple scalar value, S, is invariant under coordinate transformations. A partial derivative of an invariant is a first order covariant tensor (vector):

$$A^i = S_{,i} = \frac{\partial S}{\partial x^i}$$

However, a partial derivative of a tensor of the order one and greater is not generally an invariant under coordinate transformations of type (7) and (3).

In curvilinear coordinate system we should use more complex differentiation rules to preserve the invariance of the derivative. These rules are called the rules of *covariant differentiation* and they guarantee that the derivative itself is a tensor. According to these rules the derivatives for covariant and contravariant indices will be slightly different. They are expressed as follows:

$$A_{i,j} \equiv \frac{\partial A_i}{\partial x^j} - \left\{ {k \atop ij} \right\} A_k \tag{46}$$

$$A^i_{,j} \equiv \frac{\partial A^i}{\partial x^j} + \left\{ {i \atop kj} \right\} A^k \tag{47}$$

where the contstruct $\left\{ {k \atop ij} \right\}$ is defined as

$$\left\{ {k \atop ij} \right\} = \frac{1}{2} g^{kl} \left(\frac{\partial g_{il}}{\partial x^j} + \frac{\partial g_{jl}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^l} \right)$$

and is also known in tensor calculus as *Christoffel's symbol* of the second kind [1]. Tensor g^{ij} represents the inverse of the metric tensor g_{ij} (38). As can be seen differentiation of a single component of a vector will involve all other components of this vector.

In differentiating higher order tensors each index should be treated independently. Thus differentiating a second order tensor, A^{ij}, should be performed as:

$$A_{ij,k} = \frac{\partial A_{ij}}{\partial x^k} - \{^{m}_{ik}\}A_{mj} - \{^{m}_{jk}\}A_{im}$$

and as can be seen also involves all the components of this tensor. Likewise for the contravariant second order tensor A^{ij} we have:

$$A^{ij}_{,k} = \frac{\partial A^{ij}}{\partial x^k} + \{^{i}_{mk}\}A^{mj} + \{^{j}_{mk}\}A^{im} \quad (48)$$

And for a general n-covariant, m-contravariant tensor we have:

$$\begin{aligned} A^{j_1..j_m}_{i_1..i_n,p} &= \frac{\partial}{\partial x^p} A^{j_1..j_m}_{i_1..i_n,k} \\ &+ \{^{j_1}_{qp}\}A^{qj_2..j_m}_{i_1..i_n} + \cdots + \{^{j_m}_{qp}\}A^{j_1..j_{m-1}q}_{i_1..i_n} \\ &+ \{^{q}_{i_1p}\}A^{j_1..j_m}_{qi_2..i_n} + \cdots + \{^{q}_{i_np}\}A^{j_1..j_m}_{i_1..i_{n-1}q} \end{aligned} \quad (49)$$

Despite their seeming complexity, the relations of covariant differentiation can be easily implemented algorithmically and used in numerical solutions on arbitrary curved computational grids (Problem 4.8).

Remark 3.7 Rules of invariant expressions

As was pointed out in Corollary 3.6, the rules to build invariant expressions involve raising or lowering indexes (42), (43). However, since we did not introduce the notation for contravariant derivative, the only way to raise the index of a covariant derivative, say $A_{,i}$, it to use the relation (42) directly, that is: $g^{ij}A_{,j}$.

For example, we can re-formulate the momentum equation (45) in terms of contravariant free index i as:

$$\dot{u}^i + u^k u^i_{,k} = -\frac{g^{ik}P_{,k}}{\rho} + \nu\tau^{ik}_{,k} \quad (50)$$

where the index of the pressure term was raised by means of (42).

Using the invariance of the scalar product one can construct two important differential operators in curvilinear coordinates: *divergence* of a vector $div A \equiv A^i_{,i}$ (51) and *Laplacian*, $\Delta A \equiv g^{ik}A_{,ki}$ (55).

Definition 3.8 **Divergence**

Divergence of a vector is defined as $A^i_{,i}$*:*

$$divA \equiv A^i_{,i} \tag{51}$$

From this definition and the rule of covariant differentiation (47) we have:

$$A^i_{,i} = \frac{\partial A^i}{\partial x^i} + \{^{\,i}_{ki}\}A^k \tag{52}$$

this can be shown [2] to be equal to:

$$\begin{aligned} A^i_{,i} &= \frac{\partial A^i}{\partial x^i} + \left(\frac{1}{\sqrt{g}}\frac{\partial}{\partial x^i}\sqrt{g}\right)A^i \\ &= \frac{1}{\sqrt{g}}\frac{\partial}{\partial x^i}\left(\sqrt{g}A^i\right) \end{aligned} \tag{53}$$

where g is the determinant of the metric tensor g_{ij}.

The divergence of a covariant vector A_i is defined as a divergence of its conjugate contravariant tensor (42):

$$A^i_{,i} = g^{ij}A_{j,i} \tag{54}$$

Definition 3.9 **Laplacian**

A Laplace operator or a Laplacian of a scalar A is defined as

$$\Delta A \equiv g^{ik}A_{,ki} \tag{55}$$

The definitions (3.8), (3.9) of differential operators are invariant under coordinate transformations. They can be programmed using a symbolic manipulation packages and used to derive expressions in different curvilinear coordinate systems (Problem 4.9).

3.3 Orthogonal coordinates

3.3.1 Unit vectors and stretching factors

The coordinate system is *orthogonal* if the tangential vectors to coordinate lines are orthogonal at every point.

Consider three unit vectors, a^i, b^i, c^i, each directed along one of the coordinate axis (tangential unit vectors), that is:

$$a^i = \{a^1, 0, 0\} \tag{56}$$
$$b^i = \{0, b^2, 0\} \tag{57}$$
$$c^i = \{0, 0, c^3\} \tag{58}$$

The condition of orthogonality means that the scalar product between any two of these unit vectors should be zero. According to the definition of a scalar product (Definition 3.4) it should be written in form (44), that is, a scalar product between vectors a_i and b_i can be written as: $a^i b_i$ or $a_i b^i$. Let's use the first form for definiteness. Then, applying the operation of rising indexes (42), we can express the scalar product in contravariant components only:

$$\begin{aligned} 0 = a^i b_i = g_{ij} a^i b^j = \\ g_{11} a^1 0 + g_{12} a^1 b^2 + g_{13} 00 \\ g_{21} a^2 b^1 + g_{22} 0 b^2 + g_{23} 00 \\ g_{31} a^3 0 + g_{32} 0 b^2 + g_{33} 00 \\ = (g_{12} + g_{21}) a^1 b^2 = 2 g_{12} a^1 b^2 = 0 \end{aligned} \tag{59}$$

where we used the symmetry of g_{ij}, (37). Since vectors a^1 and b^2 were chosen to be non-zero, we have: $g_{12} = 0$. Applying the same reasoning for scalar products of other vectors, we conclude that the metric tensor has only diagonal components non-zero[5]:

$$g_{ij} = \delta_{ij} g_{(ii)} \tag{60}$$

Let's introduce stretching factors, h_i, as the square roots of these diagonal components of g_{ij}:

$$h_1 \equiv (g_{11})^{1/2}; \quad h_2 \equiv (g_{22})^{1/2}; \quad h_3 \equiv (g_{33})^{1/2}; \tag{61}$$

Now, consider the scalar product of each of the unit vectors (56)-(58) with itself. Since all vectors are unit, the scalar product of each with itself should be one:

[5] We use parenthesis to preclude summation (Proposition 2.12)

$$a^i a_i = b^i b_i = c^i c_i = 1$$

Or, expressed in contravariant components only the condition of unity is:

$$g_{ij} a^i a^j = g_{ij} b^i b^j = g_{ij} c^i c^j = 1$$

Now, consider the first term above and substitute the components of a from (56). The only non-zero term will be:

$$g_{11} a^1 a^1 = (h_1)^2 (a^1)^2 = 1$$

and consequently:

$$a^1 = \pm \frac{1}{h_1} \quad (62)$$

where the negative solution identifies a vector directed into the opposite direction, and we can neglect it for definiteness. Applying the same reasoning for each of the tree unit vectors a_i, b_i, c_i, we can rewrite (56), (57) and (58) as:

$$a^i = \{\frac{1}{h_1}, 0, 0\} \quad (63)$$

$$b^i = \{0, \frac{1}{h_2}, 0\} \quad (64)$$

$$c^i = \{0, 0, \frac{1}{h_3}\} \quad (65)$$

which means that the components of unit vectors in a curved space should be scaled with coefficients h_i. It follows from this that the expression for the element of length in curvilinear coordinates, (35), can be written as:

$$dl^2 = g_{ij} d\tilde{x}^i d\tilde{x}^j = h_i^2 (d\tilde{x}^i)^2 \quad (66)$$

Similarly, we introduce the h^i coefficients for the conjugate metric tensor (38):

$$g^{ij} = \delta_{ij} (h^{(i)})^2 \quad (67)$$

Combining the latter with (38), we obtain: $\delta_{ij} h_{(i)} h^{(i)} = \delta_{ij}$, from which it follows that

$$h_{(i)} = 1/h^{(i)} \tag{68}$$

3.3.2 Physical components of tensors

Consider a direction in space determined by a unit vector e_i. Then the *physical component* of a vector A_i in the direction e_i is given by a scalar product between A_i and e_i (Definition 3.4), namely:

$$A(e) = g^{ij} A_i e_j$$

According to Corollary 3.5 the above can also be rewritten as:

$$A(e) = A_i e^i = A^i e_i \tag{69}$$

Suppose the unit vector is directed along one of the axis: $e^i = \{e^1, 0, 0\}$. From (63) it follows that:

$$e^1 = 1/h_1$$

where h_1 is defined by (61). Thus according to (69) the physical component of vector A_i in direction 1 in orthogonal coordinate system is equal to:

$$A(1) = A_1/h_1$$

or, repeating the argument for other components, we have for the physical components of a covariant vector:

$$A_1/h_1,\ A_2/h_2,\ A_3/h_3 \tag{70}$$

Following the same reasoning, for the contravariant vector A^i, we have:

$$h_1 A^1,\ h_2 A^2\ h_3 A^3$$

General rules of covariant differentiation introduced in (Sec.3.2) simplify considerably in orthogonal coordinate systems. In particular, we can define the *nabla* operator by the physical components of a covariant vector composed of partial differentials:

$$\nabla_i = \frac{1}{h_{(i)}}\frac{\partial}{\partial x^i} \tag{71}$$

where the parentheses indicate that there's no summation with respect to index i.

In orthogonal coordinate system the general expressions for divergence (53) and Laplacian (55)) operators can be expressed in terms of stretching factors only [3]:

$$\begin{aligned} A^i_{,i} &= \frac{1}{H}\frac{\partial}{\partial x_i}\left(\frac{H}{h_{(i)}}A_i\right) \\ \Delta A &= \frac{1}{H}\frac{\partial}{\partial x_i}\left(\frac{H}{h_{(i)}}\frac{A}{\partial x_i}\right) \\ H &\equiv \prod_{i=1}^{n} h_i \end{aligned} \tag{72}$$

Important examples of orthogonal coordinate systems are spherical and cylindrical coordinate systems. Consider the example of a cylindrical coordinate system: $x_i = \{x_1, x_2, x_3\}$ and $\tilde{x}_i = \{r, \theta, l\}$:

$$\begin{aligned} x_1 &= r\cos\theta \\ x_2 &= r\sin\theta \\ x_3 &= l \end{aligned}$$

According to (40) only few components of the metric tensor will survive (Problem 4.5). Then we can compute nabla, divergence and Laplacian operators according to (71), (52) and (55), or using simplified relations (72)-(73):

$$\begin{aligned} \nabla &= \left(\frac{\partial}{\partial r}, \frac{1}{r}\frac{\partial}{\partial \theta}, \frac{\partial}{\partial z}\right) \\ divA &= \frac{\partial A_1}{\partial \tilde{x}^1} + \frac{1}{\tilde{x}^1}\frac{\partial A_2}{\partial \tilde{x}^2} + \frac{\partial A_3}{\partial \tilde{x}^3} + \frac{1}{\tilde{x}^1}A_1 \\ &= \frac{\partial A_r}{\partial r} + \frac{1}{r}\frac{\partial A_\theta}{\partial \theta} + \frac{\partial A_z}{\partial z} + \frac{1}{r}A_r \end{aligned}$$

Note, that instead of using the contravariant components as implied by the general definition of the divergence operator (51) we are using the covariant components as dictated by relation (70). The expression of the Laplacian becomes:

$$
\begin{aligned}
\Delta A &= \frac{\partial^2 A}{(\partial \tilde{x}_1)^2} + \frac{1}{\tilde{x}_1^2}\frac{\partial^2 A}{(\partial \tilde{x}_2)^2} + \frac{\partial^2 A}{(\partial \tilde{x}_3)^2} + \frac{1}{\tilde{x}_1}\frac{\partial A}{\partial \tilde{x}_1} \\
&= \frac{\partial^2 A}{(\partial r)^2} + \frac{1}{r^2}\frac{\partial^2 A}{(\partial \theta)^2} + \frac{\partial^2 A}{(\partial z)^2} + \frac{1}{r}\frac{\partial A}{\partial r}
\end{aligned}
$$

(see Problems 4.9,4.10).

The advantages of the tensor approach are that it can be used for any type of curvilinear coordinate transformations, not necessarily analytically defined, like cylindrical (85) or spherical. Another advantage is that the equations above can be easily produced automatically using symbolic manipulation packages, such as Mathematica (wolfram.com) (Problems 4.6,4.7,4.9). For further reading see [1, 2].

4 Problems

Problem 4.1 Check tensor expressions for consistency

Check if the following Cartesian tensor expressions violate tensor rules:

$$A_{ij}B_{jk} + B_{pq}C_qD_k = 0$$

$$E_{pqi}F_{kj}C_{pk} + B_{pj}D_{jq}G_q = F_{kp}$$

$$E_{ijk}A_jB_k - D_{ij}A_iB_j = F_{ij}G_{jk}H_{kj}$$

Problem 4.2 Construct tensor expression

Construct a valid Cartesian tensor expression, consisting of three terms, each including some of the four tensors: $A_{ijk}, B_{ij}, C_i, D_{ij}$. Term 1 should include tensors A, B, C only, term 2 tensor B, C, D and term 3 tensors C, D, A. The expression should have 2 free indexes, which should always come first among the indexes of a tensor. The free indexes should be at A and B in the first term, at B and C in the second term and C and D in the last term. How many different tensor expressions can be constructed?

Problem 4.3 Cartesian identity

Prove identity (15)

Problem 4.4 Vector identity

Using tensor identity (29):

$$\varepsilon_{ijk}\varepsilon_{ipq} = \delta_{jp}\delta_{kq} - \delta_{jq}\delta_{kp}$$

prove vector identity (30):

$$\vec{A} \times (\vec{B} \times \vec{C}) = \vec{B}(\vec{A} \cdot \vec{C}) - \vec{C}(\vec{A} \cdot \vec{B})$$

Problem 4.5 Metric tensor in cylindrical coordinates

Cylindrical coordinate system $y_i = \{r, \theta, l\}$ (85) is given by the following transformation rules to a Cartesian coordinate system, $x_i = \{x, y, z\}$:

$$x = r\cos\theta$$
$$y = r\sin\theta$$
$$z = l$$

Obtain the components of the metric tensor (40) g_{ij} and its inverse g^{ij} (38) in cylindrical coordinates.

Problem 4.6 Metric tensor in curvilinear coordinates

Using Mathematica Compute the metric tensor, g, (40) and its conjugate, $\hat{g}$, (38) in spherical coordinate system (r, ϕ, θ):

$$
\begin{aligned}
x &= r\sin\theta\cos\phi \\
y &= r\sin\theta\sin\phi \\
z &= r\cos\theta
\end{aligned}
\tag{73}
$$

Problem 4.7 Christoffel's symbols with Mathematica

Using the Mathematica package, write the routines for computing Christoffel's symbols.

Problem 4.8 Covariant differentiation with Mathematica

Using the Mathematica package, and the routines developed in Problem 4.7 write the routines for covariant differentiation of tensors up to second order.

Problem 4.9 Divergence of a vector in curvilinear coordinates

Using the Mathematica package and the solution of Problem 4.8, write the routines for computing divergence of a vector in curvilinear coordinates.

Problem 4.10 Laplacian in curvilinear coordinates

Using the Mathematica package and the solution of Problem 4.8, write the routines for computing the Laplacian in curvilinear coordinates.

Problem 4.11 Invariant expressions

Check if any of these tensor expressions are invariant, and correct them if not:

$$A_i B^i_{jk} C^j_{t,k} = D_t \tag{74}$$

$$A^{ij}_{jk} B_{ipq} C^{kq} - F_{kj} G^k_p H^j = H^k A^{jq}_{kj} C^{ti} B_{pit,q} \tag{75}$$

$$E^i B^i_{kp} + D^p_{kq} C^j_{jq} = D_{ki} G_{p,i} \tag{76}$$

Problem 4.12 Contraction invariance

Prove that $A^i B_i$ is an invariant and $A_i B_i$ is not.

A Solutions to problems

Problem 4.1: Check tensor expressions

Check if the following Cartesian tensor expressions violate tensor rules:

$$A_{ij}B_{jk} + B_{pq}C_qD_k = 0$$

Answer: term (1): ik = free, term (2): pk=free

$$E_{pqi}F_{kj}C_{pk} + B_{pj}D_{jq}G_q = F_{kp}$$

Answer: (1): ijq=free (2): p=free (3): kp=free

$$E_{ijk}A_jB_k - D_{ij}A_iB_j = F_{ij}G_{jk}H_{kj}$$

Answer: (1): i=free (2): none, (3): i=free, j = tripple occurrence

---

Problem 4.2: Construct tensor expression

Construct a valid Cartesian tensor expression, consisting of three terms, each including some of the four tensors: $A_{ijk}, B_{ij}, C_i, D_{ij}$. Term 1 should include tensors A, B, C only, term 2 tensor B, C, D and term 3 tensors C, D, A. The expression should have 2 free indexes, which should always come first among the indexes of a tensor. The free indexes should be at A and B in the first term, at B and C in the second term and C and D in the last term. How many different tensor expressions can be constructed?

Solution

One possibility is:

$$A_{ipk}B_{jk}C_p + B_{iq}C_pC_jD_{pq} + C_iD_{jp}A_{pqq} = 0$$

Since there are four locations for dummy indexes in each term, there could be three different combinations of dummies in each term. Thus, the total number of different expression is $3^3 = 27$

Problem 4.3: Cartesian identity

Prove identity (15).

Proof

Integrating (5) in the case of constant transformation marix coefficients, we have:

$$\tilde{x}^i = a^i_k x^k + b^i \tag{77}$$

where the transformation matrix is given by (4):

$$a^i_k \equiv \frac{\partial \tilde{x}^i}{\partial x^k} \tag{78}$$

By the definition of the Cartesian coordinates (79) we have:

$$a^k_i a^k_j = \frac{\partial \tilde{x}^k}{\partial x^i}\frac{\partial \tilde{x}^k}{\partial x^j} = \delta_{ij} \tag{79}$$

Let's multiply the transformation rule (77) by a^i_j. Then we get:

$$a^i_j \tilde{x}^i = a^i_j a^i_k x^k + a^i_j b^i = \delta_{jk} x^k + a^i_j b^i = x^j + a^i_j b^i$$

Differentiation this over $\tilde{x}^i$, we have:

$$a^i_j = \frac{\partial x^j}{\partial \tilde{x}^i}$$

Now rename index j into k:

$$a^i_k = \frac{\partial x^k}{\partial \tilde{x}^i}$$

Comparing this with (78), we have

$$\frac{\partial \tilde{x}^i}{\partial x^j} = \frac{\partial x^j}{\partial \tilde{x}^i}$$

which proves (15).

Problem 4.4: Tensor identity

Using the tensor identity:

$$\varepsilon_{ijk}\varepsilon_{ipq} = \delta_{jp}\delta_{kq} - \delta_{jq}\delta_{kp} \tag{80}$$

prove the vector identity (30):

$$\vec{A} \times (\vec{B} \times \vec{C}) = \vec{B}(\vec{A} \cdot \vec{C}) - \vec{C}(\vec{A} \cdot \vec{B}) \tag{81}$$

Proof

Applying (26) twice to the RHS of (81), we have:

$$\begin{aligned} &\vec{A} \times (\vec{B} \times \vec{C}) \\ &= \varepsilon_{ijk} A_j \varepsilon_{kpq} B_p C_q \\ &= \varepsilon_{ijk}\varepsilon_{kpq} A_j B_p C_q \end{aligned}$$

From (24) it follows that $\varepsilon_{ijk} = -\varepsilon_{ikj} = \varepsilon_{kij}$. Then we have:

$$\varepsilon_{ijk}\varepsilon_{kpq} A_j B_p C_q = \varepsilon_{kij}\varepsilon_{kpq} A_j B_p C_q \tag{82}$$

Now rename the dummy indexes: $k \to i, i \to j, j \to k$, so that the expression looks like one in (29):

$$\begin{aligned} &(\varepsilon_{ijk}\varepsilon_{ipq}) A_k B_p C_q \\ &= (\delta_{jp}\delta_{kq} - \delta_{jq}\delta_{kp}) A_k B_p C_q \\ &= \delta_{jp} B_p \delta_{kq} A_k C_q - \delta_{jq} C_q \delta_{kp} A_k B_p \end{aligned} \tag{83}$$

Using (10), and since $A_j = B_j$ is the same as $A_i = B_i$ the latter can be rewritten as:

$$= B_j A_q C_q - C_j A_p B_p \tag{84}$$

which is the same as

$$\vec{B}(\vec{A} \cdot \vec{C}) - \vec{C}(\vec{A} \cdot \vec{B})$$

Problem 4.5: Metric tensor in cylindrical coordinates.

Cylindrical coordinate system $\tilde{x}^i = \{r, \theta, l\}$ (85) is given by the following transformation rules to a Cartesian coordinate system, $x^i = \{x, y, z\}$:

$$\begin{aligned} x &= r\cos\theta \\ y &= r\sin\theta \\ z &= l \end{aligned}$$

Obtain the components of the metric tensor (40) g_{ij} and its inverse g^{ij} (38) in cylindrical coordinates.

Solution:

First compute the derivatives of $x^i = \{x, y, z\}$ with respect to $\tilde{x}^i = \{r, \theta, l\}$:

$$\begin{aligned} \frac{\partial x^1}{\partial \tilde{x}^1} &= \frac{\partial x}{\partial r} \equiv x_r = \cos\theta \\ \frac{\partial x^2}{\partial \tilde{x}^1} &= \frac{\partial y}{\partial r} \equiv y_r = \sin\theta \\ \frac{\partial x^1}{\partial \tilde{x}^2} &= \frac{\partial x}{\partial \theta} \equiv x_\theta = -r\sin\theta \\ \frac{\partial x^2}{\partial \tilde{x}^2} &= \frac{\partial y}{\partial \theta} \equiv y_\theta = r\cos\theta \\ \frac{\partial x^3}{\partial \tilde{x}^3} &= \frac{\partial z}{\partial z} \equiv z_l = 1 \end{aligned} \tag{85}$$

Then the components of the metric tensor are:

$$\begin{aligned} g_{rr} &= x_r x_r + y_r y_r = 1 \\ g_{\theta\theta} &= x_\theta x_\theta + y_\theta y_\theta = r^2 \\ g_{zz} &= 1 \\ g^{rr} &= 1 \\ g^{\theta\theta} &= \frac{1}{r^2} \\ g^{zz} &= 1 \end{aligned}$$

Problem 4.6: Metric tensor in curvilinear coordinates

Using Mathematica, write a procedure to compute metric tensor in curvilinear coordinate system, and use it to obtain the components of metric tensor, g, (40) and its conjugate, $\hat{g}$, (38) in spherical coordinate system (r, ϕ, θ):

$$\begin{aligned} x &= r\sin\theta\cos\phi \\ y &= r\sin\theta\sin\phi \\ z &= r\cos\theta \end{aligned} \tag{86}$$

Solution with Mathematica

```
NX = 3

(* Curvilinear cooridnate system *)
Y = Array[,NX] (* Spherical coordinate system *)
Y[[1]] = r;    (* radius *)
Y[[2]] = th;   (* angle theta *)
Y[[3]] = phi;  (* angle phi *)

(* Cartesian coordinate system *)
X = Array[,NX]
X[[1]] = r Sin[th] Cos[phi];
X[[2]] = r Sin[th] Sin[phi];
X[[3]] = r Cos[th];

(* Compute the Jacobian: dXi/dYj *)
J = Array[,{NX,NX}]
Do[
   J [[i,j]] = D[X[[i]],Y[[j]]],
   {j,1,NX},{i,1,NX}
]

(* Covariant Metric tensor *)
g  = Array[,{NX,NX}] (* covariant *)
Do[
        g [[i,j]] = Sum[J[[k,i]] J[[k,j]],{k,NX}],
        {j,1,NX},{i,1,NX}
```

```
];
g=Simplify[g]

(* Contravariant metric tensor *)
g1 =Array[,{NX,NX}]
g1=Inverse[g]
```

With the result:

$$g = \{\{1,0,0\},\{0,r^2,0\},\{0,0,r^2\sin(\theta)^2\}\}$$
$$\hat{g} = \{\{1,0,0\},\{0,r^{-2},0\},\{0,0,\frac{\csc(\theta)^2}{r^2}\}\}$$

Problem 4.7: Christoffel's symbols with Mathematica

Using the Mathematica package, write the routines to compute Christoffel's symbols

Solution

```
(************* File g.m *************

  The metric tensor
  and Christoffel symbols

*************************************)
DIM = 3
(*
        The metric tensor
*)
g  = Array[,{DIM,DIM}] (* covariant *)
g1 =Array[,{DIM,DIM}] (* contravariant *)
Do[
        g [[i,j]] = 0;
        g1[[i,j]] = 0
        ,
        {j,1,DIM},{i,1,DIM}
```

```
]
(*
 Cylindrical coordinates
*)
Z=Array[,DIM]
Z[[1]] = r
Z[[2]] = th
Z[[3]] = z
g [[1,1]] = 1
g [[2,2]] = r^2
g [[3,3]] = 1
g1[[1,1]] = 1
g1[[2,2]] = 1/r^2
g1[[3,3]] = 1
(*
Christoffel symbols of the first and second type
*)
Cr1 = Array[,{DIM,DIM,DIM}]
Cr2 = Array[,{DIM,DIM,DIM}]
Do[
        Cr1[[i,j,k]] = 1/2
        (
           D[ g [[i,k]], Z[[j]] ]
         + D[ g [[j,k]], Z[[i]] ]
         - D[ g [[i,j]], Z[[k]] ]
        ),
        {k,DIM},{j,DIM},{i,DIM}
]
Do[
        Cr2[[l,i,j]] =
                Sum[
                        g1[[l,k]] Cr1[[i,j,k]],
                        {k,DIM}
                ],
        {j,DIM},{i,DIM},{l,DIM}
]
```

Problem 4.8: Covariant differentiation with Mathematica

Using the Mathematica package, write the routines for covariant differentiation of tensors up to second order.

solution

```
(************* File D.m ******************

      Rules of covariant differentiation

*****************************************)
(*
        B.Spain
        Tensor Calculus, 1965
        Eq.(22.2)
*)
D1[N_,A_,k_,X_,j_]:=
(*
   Computes covariant derivative
   of a mixed tensor of second order
   with index k - covariant (upper)
*)
Module[
   {i,s},
   s = Sum[Cr2[[k,i,j]] A[[i]],{i,N}];
   D[A[[k]],X[[j]]] + s
]
Dl1[N_,A_,l_,X_,t_]:=
(*
   Computes covariant derivative
   of a mixed tensor of second order
   with index l - covariant (lower)
*)
Module[
   {s,r},
   s =Sum[Cr2[[r,l,t]] A[[r]],{r,N}];
   D[A[[l]],X[[t]]] - s
]
D1l1[N_,A_,m_,l_,X_,t_]:=
(*
   Computes covariant derivative
   of a mixed tensor of second order
   with index m - contravariant (upper) and
```

```
   index l - covariant (lower)
*)
Module[
   {s1,s2,r},
   s1 =Sum[Cr2[[m,r,t]] A[[r,l]],{r,N}];
   s2 =Sum[Cr2[[r,l,t]] A[[m,r]],{r,N}];
   D[A[[m,l]],X[[t]]] + s1 - s2
]
D2[N_,A_,i_,j_,X_,n_]:=
(*
   Computes covariant derivative
   of second order tensor with
   both m and l contravariant (upper)
   indexes
   B.Spain
   Tensor Calculus, 1965
   Eq.(23.3)
*)
Module[
   {s1,s2,k},
   s1 =Sum[Cr2[[i,k,n]] A[[k,j]],{k,N}];
   s2 =Sum[Cr2[[j,k,n]] A[[i,k]],{k,N}];
   D[A[[i,j]],X[[n]]] + s1 + s2
]
D2l1[N_,A_,i_,j_,k_,X_,n_]:=
(*
   Computes covariant derivative
   of third order tensor with
   i and j contravariant (upper)
   and k contravariant (lower)
   indexes
   B.Spain
   Tensor Calculus, 1965
   Eq.(23.3)
*)
Module[
   {s1,s2,s3,m},
   s1 =Sum[Cr2[[i,m,n]] A[[m,j,k]],{m,N}];
   s2 =Sum[Cr2[[j,m,n]] A[[i,m,k]],{m,N}];
   s3 =Sum[Cr2[[m,k,n]] A[[i,j,m]],{m,N}];
   D[A[[i,j,k]],X[[n]]] + s1 + s2 - s3
```

```
]
D4l1[N_,A_,i1_,i2_,i3_,i4_,i5,X_,i6_]:=
(*
   Computes covariant derivative
   of 5 order tensor with
   4 first indexes contravariant (upper)
   and the last one contravariant (lower)
   B.Spain
   Tensor Calculus, 1965
   Eq.(23.3)
*)
Module[
   {k,s1,s2,s3,s4,s5},
   s1= Sum[Cr2[[i1,k,n]] A[[k,i2,i3,i4,i5]],{k,N}];
   s2= Sum[Cr2[[i2,k,n]] A[[i1,k,i3,i4,i5]],{k,N}];
   s3= Sum[Cr2[[i3,k,n]] A[[i1,i2,k,i4,i5]],{k,N}];
   s4= Sum[Cr2[[i4,k,n]] A[[i1,i2,i3,k,i5]],{k,N}];
   s5=-Sum[Cr2[[k,i5,n]] A[[i1,i2,i3,i4,k]],{k,N}];
   D[A[[i1,i2,i3,i4,i5]],X[[i6]]]+s1+s2+s3+s4+s5
]
```

Problem 4.9: Divergence of a vector in curvilinear coordinates

Using the Mathematica package and the solution of Problem 4.8, write the routines for computing divergence of a vector in curvilinear coordinates.

Solution

Using the algorithms of covariant differentiation developed in Problem 4.8 we have:

```
<<"./g.m" (* The g-tensor and Christoffel symbols *)
<<"./D.m" (* Rules of covariant differentiation *)

(* The original coordinates: *)
NX = DIM
X = Array[,NX]

(* Variables: *)
NV = DIM
```

```
U = Array[,NV]

(* New coordinate system *)
Y = Array[,NX]
Y[[1]] = r;
Y[[2]] = th;
Y[[3]] = z;
X[[1]] = r Cos[th];
X[[2]] = r Sin[th];
X[[3]] = z;

(* Compute the Jacobian *)
J = Array[,{DIM,DIM}]
Do[
   J [[i,j]] = D[X[[i]],Y[[j]]],
   {j,1,DIM},{i,1,DIM}
]
J1=Simplify[Inverse[J]]

(* Derivatives of a vector *)

V0 = Array[,NX]
V0[[1]] = Vr[r,th,z];
V0[[2]] = Vt[r,th,z];
V0[[3]] = Vz[r,th,z];

(*
   Rescaling for physical
  (dimensionally correct) coordinates
(\cite[5.102-5.110]{SyScTC69})
*)
V = Array[,NX]
Do[
   V[[i]] = PowerExpand[V0[[i]]/g[[i,i]]^(1/2)],
   {i,1,NX}
]

(*
   Transform vectors
   as first order contravariant tensors
*)
```

```
U = Array[,NX]
SetAttributes[RV1,HoldAll]
RV1[NX,V,U]
(*
   Compute first covariant derivatives
   of vectors
*)
DV = Array[,{NX,NX}];
Do[
   DV[[i,j]] = D1[NX,V,i,Y,j],
   {j,1,NX},{i,1,NX}
]
(* Divergence *)
div=0
Do[
   div=div+DV[[i,i]],
   {i,NX}
]
div0 = div/.th->0
```

Problem 4.10: Laplacian in curvilinear coordinates

Using the Mathematica package, write the routines for computing Laplacian in curvilinear coordinates.

solution

Using the algorithms of covariant differentiation developed in Problem 4.8 we have:

```
<<"./g.m" (* The g-tensor and Christoffel symbols *)
<<"./D.m" (* Rules of covariant differentiation *)

(* The original coordinates: *)
NX = DIM
X = Array[,NX]

(* Variables: *)
NV = DIM
U = Array[,NV]
```

```
(* New coordinate system *)
Y = Array[,NX]
Y[[1]] = r;
Y[[2]] = th;
Y[[3]] = z;
X[[1]] = r Cos[th];
X[[2]] = r Sin[th];
X[[3]] = z;

(* Compute the Jacobian *)
J = Array[,{DIM,DIM}]
Do[
   J [[i,j]] = D[X[[i]],Y[[j]]],
   {j,1,DIM},{i,1,DIM}
]
J1=Simplify[Inverse[J]]

(* Derivative of a scalar *)

DP = Array[,NX];
Do[
   DP[[i]] = D[p[r,th,z],Y[[i]]],
   {i,1,NX}
]
DDP = Array[,{NX,NX}];
Do[
   DDP[[i,j]] = Dl1[NX,DP,i,Y,j],
   {i,1,NX},{j,1,NX}
]
DDQ = Array[,{NX,NX}];
Do[
   DDQ[[i,j]] = Sum[DDP[[k,l]] J1[[k,i]] J1[[l,j]],{k,NX},{l,NX}],
   {i,1,NX},{j,1,NX}
]

(* Laplacian *)
(***  lap=lap+Sum[g[[i,j]]*Dl1[NX,DS,j,Y,i],{i,1,NX},{j,1,NX}],*)
lap=Sum[DDQ[[i,i]],{i,NX}]
lap0=lap/.th->0
```

Problem 4.11: Invariant expressions

Check if any of these tensor expressions are invariant, and correct them if not:

$$A_i B^i_{jk} C^j_{t,k} = D_t \tag{87}$$

$$A^{ij}_{jk} B_{ipq} C^{kq} - F_{kj} G^k_p H^j = H^k A^{jq}_{kj} C^{ti} B_{pit,q} \tag{88}$$

$$E^i B^i_{kp} + D^p_{kq} C^j_{jq} = D_{ki} G_{p,i} \tag{89}$$

Answers:

A corrected form of (87) is:

$$A_i B^{ik}_j C^j_{t,k} = D_t$$

Equality (89) requires no corrections. A corrected form of (89) is:

$$E_i B^i_{kp} + D_{pkq} C^{jq}_j = D^i_k G_{p,i}$$

Since there are two combinations for an invariant combination of dummy indexes (Corollary 3.5), there can be several different invariant expressions.

Problem 4.12: Contraction invariance

Prove that $A^i B_i = A_i B^i$, and both are invariant, while $A_i B_i$ is not.

Proof

Using the operation of rising/lowering indexes (42), (43), we have

$$A^i B_i = g^{ij} A_j g_{ik} B^k = g^{ij} g_{ik} A_j B^k = \delta_{jk} A_j B^k = A_j B^j$$

which proves that both forms have the same values. If we now consider the first form then:

$$\bar{A}^i \bar{B}_i = \frac{\partial \bar{x}_i}{\partial x_j} A^j \frac{\partial x_k}{\partial \bar{x}_i} B_k = \delta_{jk} A^j B_k = A^j B_j = A^i B_i$$

which proves the point.

Consider now A_iB_i:

$$\bar{A}_i\bar{B}_i = \frac{\partial x_j}{\partial \bar{x}_i}A_j\frac{\partial x_k}{\partial \bar{x}_i}B_k$$

which can not be reduced further and, therefore is not invariant, since it has a different form from the LHS.

References

[1] Barry Spain. *Tensor Calculus*. Oliver and Boyd, 1965.

[2] J.L. Synge and A. Schild. *Tensor Calculus*. Dover Publications, 1969.

[3] P. Morse and H. Feshbach. *Methods of Theoretical Physics*. McGraw-Hill, New York, 1953.

Index

A No-Nonsense Introduction to General Relativity

Sean M. Carroll
Enrico Fermi Institute and Department of Physics,
University of Chicago, Chicago, IL, 60637
carroll@theory.uchicago.edu

1 Introduction

General relativity (GR) is the most beautiful physical theory ever invented. Nevertheless, it has a reputation of being extremely difficult, primarily for two reasons: tensors are everywhere, and spacetime is curved. These two facts force GR people to use a different language than everyone else, which makes the theory somewhat inaccessible. Nevertheless, it is possible to grasp the basics of the theory, even if you're not Einstein (and who is?).

GR can be summed up in two statements: 1) Spacetime is a curved pseudo-Riemannian manifold with a metric of signature $(-+++)$. 2) The relationship between matter and the curvature of spacetime is contained in the equation

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi G T_{\mu\nu} \ . \tag{1}$$

However, these statements are incomprehensible unless you sling the lingo. So that's what we shall start doing. Note, however, that this introduction is a very pragmatic affair, intended to give you some immediate feel for the language of GR. It does not substitute for a deep understanding – that takes more work!

Administrative notes: physicists love to set constants to unity, and it's a difficult habit to break once you start. I will *not* set Newton's constant $G = 1$. However, it's ridiculous not to set the speed of light $c = 1$, so I'll do that. For further reference, recommended texts include *A First Course in General Relativity* by Bernard Schutz, at an undergrad level; and graduate texts *General Relativity* by Wald, *Gravitation and Cosmology* by Weinberg, *Gravitation* by Misner, Thorne, and Wheeler, and *Introducing Einstein's Relativity* by D'Inverno. Of course best of all would be to rush to `<http://pancake.uchicago.edu/~carroll/notes/>`, where you will find about one semester's worth of free GR notes, of which this introduction is essentially an abridgment.

2 Special Relativity

Special relativity (SR) stems from considering the speed of light to be invariant in all reference frames. This naturally leads to a view in which space and time are joined together to form spacetime; the conversion factor from time units to space units is c (which equals 1, right? couldn't be simpler). The coordinates of spacetime may be chosen to be

$$\begin{aligned} x^0 &\equiv ct = t \\ x^1 &\equiv x \\ x^2 &\equiv y \\ x^3 &\equiv z. \end{aligned} \tag{2}$$

These are **Cartesian coordinates**. Note a few things: these indices are *superscripts*, not exponents. The indices go from zero to three; the collection of all four coordinates is denoted x^μ. Spacetime indices are always in Greek; occasionally we will use Latin indices if we mean only the spatial components, e.g. $i = 1, 2, 3$.

The stage on which SR is played out is a specific four dimensional manifold, known as **Minkowski spacetime** (or sometimes "Minkowski space"). The x^μ are coordinates on this manifold. The elements of spacetime are known as **events**; an event is specified by giving its location in both space and time. Vectors in spacetime are always fixed at an event; there is no such thing as a "free vector" that can move from place to place. Since Minkowski space is four dimensional, these are generally known as **four-vectors**, and written in components as V^μ, or abstractly as just V.

We also have the **metric** on Minkowski space, $\eta_{\mu\nu}$. The metric gives us a way of taking the norm of a vector, or the dot product of two vectors. Written as a matrix, the Minkowski metric is

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} . \tag{3}$$

Then the dot product of two vectors is defined to be

$$A \cdot B \equiv \eta_{\mu\nu} A^\mu B^\nu = -A^0 B^0 + A^1 B^1 + A^2 B^2 + A^3 B^3 . \tag{4}$$

(We always use the **summation convention**, in which identical upper and lower indices are implicitly summed over all their possible values.) This is especially useful for taking the infinitesimal (distance)^2 between two points, also known as the **spacetime interval**:

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu \tag{5}$$

$$= -dt^2 + dx^2 + dy^2 + dz^2 . \tag{6}$$

In fact, an equation of the form (6) is often called "the metric." The metric contains all of the information about the geometry of the manifold. The Minkowski metric is of course just the spacetime generalization of the ordinary inner product on flat Euclidean space, which we can think of in components as the Kronecker delta, δ_{ij}. We say that the Minkowski metric has **signature** $(-+++)$, sometimes called "Lorentzian," as opposed to the Euclidian signature with all plus signs. (The overall sign of the metric is a matter of convention, and many texts use $(+---)$.)

Notice that for a particle with fixed spatial coordinates x^i, the interval elapsed as it moves forward in time is negative, $ds^2 = -dt^2 < 0$. This leads us to define the **proper time** τ via

$$d\tau^2 \equiv -ds^2 . \tag{7}$$

The proper time elapsed along a trajectory through spacetime will be the actual time measured by an observer on that trajectory. Some other observer, as we know, will measure a different time.

Some verbiage: a vector V^μ with negative norm, $V \cdot V < 0$, is known as **timelike**. If the norm is zero, the vector is **null**, and if it's positive, the vector is **spacelike**. Likewise, trajectories with negative ds^2 (note – *not* proper time!) are called timelike, etc. These concepts lead naturally to the concept of a **spacetime diagram**, with which you are presumably familiar. The set of null trajectories leading into and out of an event constitute a **light cone**, terminology which becomes transparent in the context of a spacetime diagram such as Figure 1.

A path through spacetime is specified by giving the four spacetime coordinates as a function of some parameter, $x^\mu(\lambda)$. A path is characterized as timelike/null/spacelike when its tangent vector $dx^\mu/d\lambda$ is timelike/null/spacelike. For timelike paths the most convenient parameter to use is the proper time τ, which we can compute along an arbitrary timelike path via

$$\tau = \int \sqrt{-ds^2} = \int \sqrt{-\eta_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}}\, d\lambda \ . \tag{8}$$

The corresponding tangent vector $U^\mu = dx^\mu/d\tau$ is called the **four-velocity**, and is automatically normalized:

$$\eta_{\mu\nu}U^\mu U^\nu = -1 \ , \tag{9}$$

as you can check.

A related vector is the **momentum four-vector**, defined by

$$p^\mu = mU^\mu \ , \tag{10}$$

where m is the mass of the particle. The mass is a fixed quantity independent of inertial frame, what you may be used to thinking of as the "rest mass." The **energy** of a particle is simply p^0, the timelike component of its momentum vector. In the particle's rest frame we have $p^0 = m$; recalling that we have set $c = 1$, we find that we have found the famous equation $E = mc^2$. In a moving frame we can find the components of p^μ by performing a Lorentz transformation; for a particle moving with three-velocity $v = dx/dt$ along the x axis we have

$$p^\mu = (\gamma m, v\gamma m, 0, 0) \ , \tag{11}$$

where $\gamma = 1/\sqrt{1-v^2}$. For small v, this gives $p^0 = m + \frac{1}{2}mv^2$ (what we usually think of as rest energy plus kinetic energy) and $p^1 = mv$ (what we usually think of as Newtonian momentum).

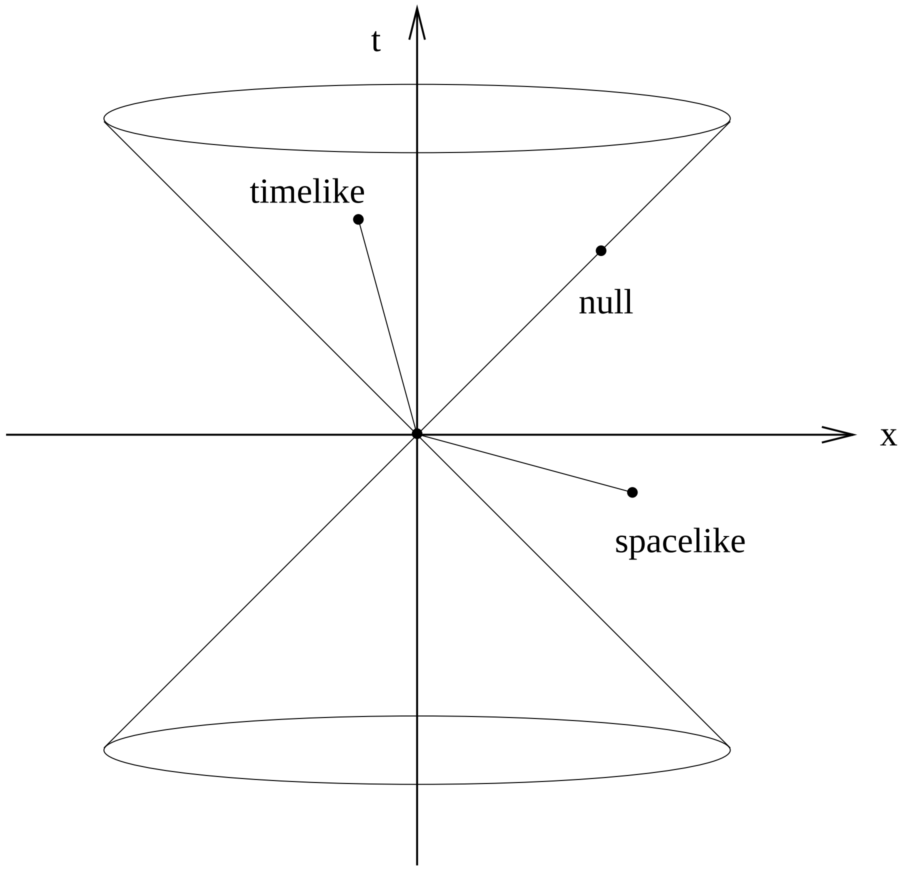

Figure 1: A lightcone, portrayed on a spacetime diagram. Points which are spacelike-, null-, and timelike-separated from the origin are indicated.

3 Tensors

The transition from flat to curved spacetime means that we will eventually be unable to use Cartesian coordinates; in fact, some rather complicated coordinate systems become necessary. Therefore, for our own good, we want to make all of our equations **coordinate invariant** – i.e., if the equation holds in one coordinate system, it will hold in any. It also turns out that many of the quantities that we use in GR will be **tensors**. Tensors may be thought of as objects like vectors, except with possibly more indices, which transform under a change of coordinates $x^\mu \to x^{\mu'}$ according to the following rule, the **tensor transformation law**:

$$S^{\mu'}{}_{\nu'\rho'} = \frac{\partial x^{\mu'}}{\partial x^\mu}\frac{\partial x^\nu}{\partial x^{\nu'}}\frac{\partial x^\rho}{\partial x^{\rho'}} S^\mu{}_{\nu\rho} \ . \tag{12}$$

Note that the unprimed indices on the right are dummy indices, which are summed over. The pattern in (12) is pretty easy to remember, if you think of "conservation of indices": the upper and lower *free* indices (not summed over) on each side of an equation must be the same. This holds true for any equation, not just the tensor transformation law. Remember also that upper indices can only be summed with lower indices; if you have two upper or lower indices that are the same, you goofed. Since there are in general no preferred coordinate systems in GR, it behooves us to cast all of our equations in tensor form, because *if an equation between two tensors holds in one coordinate system, it holds in all coordinate systems.*

Tensors are not very complicated; they're just generalizations of vectors. (Note that scalars qualify as tensors with no indices, and vectors are tensors with one upper index; a tensor with two indices can be though of as a matrix.) However, there is an entire language associated with them which you must learn. If a tensor has n upper and m lower indices, it is called a $(n,\ m)$ tensor. The upper indices are called **contravariant** indices, and the lower ones are **covariant**; but everyone just says "upper" and "lower," and so should you. Tensors of type $(n,\ m)$ can be **contracted** to form a tensor of type $(n-1,\ m-1)$ by summing over one upper and one lower index:

$$S^\mu = T^{\mu\lambda}{}_\lambda \ . \tag{13}$$

The contraction of a two-index tensor is often called the **trace**. (Which makes sense if you think about it.)

If a tensor is the same when we interchange two indices,

$$S_{\cdots\alpha\beta\cdots} = S_{\cdots\beta\alpha\cdots} \ , \tag{14}$$

it is said to be **symmetric** in those two indices; if it changes sign,

$$S_{\cdots\alpha\beta\cdots} = -S_{\cdots\beta\alpha\cdots} \ , \tag{15}$$

we call it **antisymmetric**. A tensor can be symmetric or antisymmetric in many indices at once. We can also take a tensor with no particular symmetry properties in some set of indices and pick out the symmetric/antisymmetric piece by taking appropriate linear combinations; this procedure of symmetrization or antisymmetrization is denoted by putting parentheses or square brackets around the relevant indices:

$$\begin{aligned} T_{(\mu_1\mu_2\cdots\mu_n)} &= \frac{1}{n!}\left(T_{\mu_1\mu_2\cdots\mu_n} + \text{sum over permutations of } \mu_1\cdots\mu_n\right) \\ T_{[\mu_1\mu_2\cdots\mu_n]} &= \frac{1}{n!}\left(T_{\mu_1\mu_2\cdots\mu_n} + \text{alternating sum over permutations of } \mu_1\cdots\mu_n\right). \end{aligned} \tag{16}$$

By "alternating sum" we mean that permutations which are the result of an odd number of exchanges are given a minus sign, thus:

$$T_{[\mu\nu\rho]\sigma} = \frac{1}{6}\left(T_{\mu\nu\rho\sigma} - T_{\mu\rho\nu\sigma} + T_{\rho\mu\nu\sigma} - T_{\nu\mu\rho\sigma} + T_{\nu\rho\mu\sigma} - T_{\rho\nu\mu\sigma}\right) . \tag{17}$$

The most important tensor in GR is the metric $g_{\mu\nu}$, a generalization (to arbitrary coordinates and geometries) of the Minkowski metric $\eta_{\mu\nu}$. Although $\eta_{\mu\nu}$ is just a special case of $g_{\mu\nu}$, we denote it by a different symbol to emphasize the importance of moving from flat to curved space. The metric is a symmetric two-index tensor. An important fact is that it is always possible to find coordinates such that, at one specified point p, the components of the metric are precisely those of the Minkowski metric (3) and the first derivatives of the metric vanish. In other words, the metric will look flat at precisely that point; however, in general the second derivatives of $g_{\mu\nu}$ cannot be made to vanish, a manifestation of curvature.

Even if spacetime is flat, the metric can still have nonvanishing derivatives if the coordinate system is non-Cartesian. For example, in spherical coordinates (on space) we have

$$\begin{aligned} t &= t \\ x &= r\sin\theta\cos\phi \\ y &= r\sin\theta\sin\phi \\ z &= r\cos\theta , \end{aligned} \tag{18}$$

which leads directly to

$$ds^2 = -dt^2 + dr^2 + r^2\,d\theta^2 + r^2\sin^2\theta\,d\phi^2 , \tag{19}$$

or

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2\sin^2\theta \end{pmatrix} . \tag{20}$$

Notice that, while we could use the tensor transformation law (12), it is often more straightforward to find new tensor components by simply plugging in our coordinate transformations to the differential expression (*e.g.* $dz = \cos\theta\,dr - r\sin\theta\,d\theta$).

Just as in Minkowski space, we use the metric to take dot products:

$$A \cdot B \equiv g_{\mu\nu} A^\mu B^\nu \ . \tag{21}$$

This suggests, as a shortcut notation, the concept of **lowering indices**; from any vector we can construct a $(0, 1)$ tensor defined by contraction with the metric:

$$A_\nu \equiv g_{\mu\nu} A^\mu \ , \tag{22}$$

so that the dot product becomes $g_{\mu\nu} A^\mu B^\nu = A_\nu B^\nu$. We also define the **inverse metric** $g^{\mu\nu}$ as the matrix inverse of the metric tensor:

$$g^{\mu\nu} g_{\nu\rho} = \delta^\mu_\rho \ , \tag{23}$$

where δ^μ_ρ is the (spacetime) Kronecker delta. (Convince yourself that this expression really does correspond to matrix multiplication.) Then we have the ability to raise indices:

$$A^\mu = g^{\mu\nu} A_\nu \ . \tag{24}$$

Note that raising an index on the metric yields the Kronecker delta, so we have

$$g^{\mu\nu} g_{\mu\nu} = \delta^\mu_\mu = 4 \ . \tag{25}$$

Despite the ubiquity of tensors, it is sometimes useful to consider non-tensorial objects. An important example is the determinant of the metric tensor,

$$g \equiv \det\left(g_{\mu\nu}\right) \ . \tag{26}$$

A straightforward calculation shows that under a coordinate transformation $x^\mu \to x^{\mu'}$, this doesn't transform by the tensor transformation law (under which it would have to be invariant, since it has no indices), but instead as

$$g \to \left[\det\left(\frac{\partial x^{\mu'}}{\partial x^\mu}\right)\right]^{-2} g \ . \tag{27}$$

The factor $\det(\partial x^{\mu'}/\partial x^\mu)$ is the Jacobian of the transformation. Objects with this kind of transformation law (involving powers of the Jacobian) are known as **tensor densities**; the determinant g is sometimes called a "scalar density." Another example of a density is the volume element $d^4x = dx^0 dx^1 dx^2 dx^3$:

$$d^4x \to \det\left(\frac{\partial x^{\mu'}}{\partial x^\mu}\right) d^4x \ . \tag{28}$$

To define an invariant volume element, we can therefore multiply d^4x by the square root of minus g, so that the Jacobian factors cancel out:

$$\sqrt{-g}\, d^4x \to \sqrt{-g}\, d^4x \ . \tag{29}$$

In Cartesian coordinates, for example, we have $\sqrt{-g}\, d^4x = dt\, dx\, dy\, dz$, while in polar coordinates this becomes $r^2 \sin\theta\, dt\, dr\, d\theta\, d\phi$. Thus, integrals of functions over spacetime are of the form $\int f(x^\mu)\sqrt{-g}\, d^4x$. ("Function," of course, is the same thing as "scalar.")

Another object which is unfortunately not a tensor is the partial derivative $\partial/\partial x^\mu$, often abbreviated to ∂_μ. Acting on a scalar, the partial derivative returns a perfectly respectable $(0,1)$ tensor; using the conventional chain rule we have

$$\partial_\mu \phi \to \partial_{\mu'}\phi = \frac{\partial x^\mu}{\partial x^{\mu'}}\partial_\mu \phi \ , \tag{30}$$

in agreement with the tensor transformation law. But on a vector V^μ, given that $V^\mu \to \frac{\partial x^{\mu'}}{\partial x^\mu} V^\mu$, we get

$$\begin{aligned} \partial_\mu V^\nu \to \partial_{\mu'} V^{\nu'} &= \left(\frac{\partial x^\mu}{\partial x^{\mu'}}\partial_\mu\right)\left(\frac{\partial x^{\nu'}}{\partial x^\nu} V^\nu\right) \\ &= \frac{\partial x^\mu}{\partial x^{\mu'}}\frac{\partial x^{\nu'}}{\partial x^\nu}(\partial_\mu V^\nu) + \frac{\partial x^\mu}{\partial x^{\mu'}}\frac{\partial^2 x^{\nu'}}{\partial x^\nu \partial x^\mu} V^\mu \ . \end{aligned} \tag{31}$$

The first term is what we want to see, but the second term ruins it. So we define a **covariant derivative** to be a partial derivative plus a correction that is linear in the original tensor:

$$\nabla_\mu V^\nu = \partial_\mu V^\nu + \Gamma^\nu_{\mu\lambda} V^\lambda \ . \tag{32}$$

Here, the symbol $\Gamma^\nu_{\mu\lambda}$ stands for a collection of numbers, called **connection coefficients**, with an appropriate non-tensorial transformation law chosen to cancel out the non-tensorial term in (31). Thus we need to have

$$\Gamma^{\nu'}_{\mu'\lambda'} = \frac{\partial x^\mu}{\partial x^{\mu'}}\frac{\partial x^\lambda}{\partial x^{\lambda'}}\frac{\partial x^{\nu'}}{\partial x^\nu}\Gamma^\nu_{\mu\lambda} - \frac{\partial x^\mu}{\partial x^{\mu'}}\frac{\partial x^\lambda}{\partial x^{\lambda'}}\frac{\partial^2 x^{\nu'}}{\partial x^\mu \partial x^\lambda} \ . \tag{33}$$

Then $\nabla_\mu V^\nu$ is guaranteed to transform like a tensor. The same kind of trick works to define covariant derivatives of tensors with lower indices; we simply introduce a minus sign and change the dummy index which is summed over:

$$\nabla_\mu \omega_\nu = \partial_\mu \omega_\nu - \Gamma^\lambda_{\mu\nu}\omega_\lambda \ . \tag{34}$$

If there are many indices, for each upper index you introduce a term with a single $+\Gamma$, and for each lower index a term with a single $-\Gamma$:

$$\begin{aligned} \nabla_\sigma T^{\mu_1\mu_2\cdots\mu_k}{}_{\nu_1\nu_2\cdots\nu_l} &= \partial_\sigma T^{\mu_1\mu_2\cdots\mu_k}{}_{\nu_1\nu_2\cdots\nu_l} \\ &\quad + \Gamma^{\mu_1}_{\sigma\lambda} T^{\lambda\mu_2\cdots\mu_k}{}_{\nu_1\nu_2\cdots\nu_l} + \Gamma^{\mu_2}_{\sigma\lambda} T^{\mu_1\lambda\cdots\mu_k}{}_{\nu_1\nu_2\cdots\nu_l} + \cdots \\ &\quad - \Gamma^\lambda_{\sigma\nu_1} T^{\mu_1\mu_2\cdots\mu_k}{}_{\lambda\nu_2\cdots\nu_l} - \Gamma^\lambda_{\sigma\nu_2} T^{\mu_1\mu_2\cdots\mu_k}{}_{\nu_1\lambda\cdots\nu_l} - \cdots \ . \end{aligned} \tag{35}$$

This is the general expression for the covariant derivative.

What are these mysterious connection coefficients? Fortunately they have a natural expression in terms of the metric and its derivatives:

$$\Gamma^\sigma_{\mu\nu} = \frac{1}{2} g^{\sigma\rho}(\partial_\mu g_{\nu\rho} + \partial_\nu g_{\rho\mu} - \partial_\rho g_{\mu\nu}) \ . \tag{36}$$

It is left up to you to check that the mess on the right really does have the desired transformation law. You can also verify that the connection coefficients are symmetric in their lower indices, $\Gamma^\sigma_{\mu\nu} = \Gamma^\sigma_{\nu\mu}$. These coefficients can be nonzero even in flat space, if we have non-Cartesian coordinates. In principle there can be other kinds of connection coefficients, but we won't worry about that here; the particular choice (36) are sometimes called **Christoffel symbols**, and are the ones we always use in GR. With these connection coefficients, we get the nice feature that the covariant derivative of the metric and its inverse are always zero, known as **metric compatibility**:

$$\nabla_\sigma g_{\mu\nu} = 0 \ , \qquad \nabla_\sigma g^{\mu\nu} = 0 \ . \tag{37}$$

So, given any metric $g_{\mu\nu}$, we proceed to calculate the connection coefficients so that we can take covariant derivatives. Many of the familiar equations of physics in flat space continue to hold true in curved space once we replace partial derivatives by covariant ones. For example, in special relativity the electric and magnetic vector fields $\vec{E}$ and $\vec{B}$ can be collected into a single two-index antisymmetric tensor $F_{\mu\nu}$:

$$F_{\mu\nu} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & B_z & -B_y \\ E_y & -B_z & 0 & B_x \\ E_z & B_y & -B_x & 0 \end{pmatrix} \ , \tag{38}$$

and the electric charge density ρ and current $\vec{J}$ into a four-vector J^μ:

$$J^\mu = (\rho, \ \vec{J}) \ . \tag{39}$$

In this notation, Maxwell's equations

$$\begin{aligned} \nabla \times \mathbf{B} - \partial_t \mathbf{E} &= 4\pi \mathbf{J} \\ \nabla \cdot \mathbf{E} &= 4\pi\rho \\ \nabla \times \mathbf{E} + \partial_t \mathbf{B} &= 0 \\ \nabla \cdot \mathbf{B} &= 0 \end{aligned} \tag{40}$$

shrink into two relations,

$$\begin{aligned} \partial_\mu F^{\nu\mu} &= 4\pi J^\nu \\ \partial_{[\mu} F_{\nu\lambda]} &= 0 \ . \end{aligned} \tag{41}$$

These are true in Minkowski space, but the generalization to a curved spacetime is immediate; just replace $\partial_\mu \to \nabla_\mu$:

$$\begin{aligned} \nabla_\mu F^{\nu\mu} &= 4\pi J^\nu \\ \nabla_{[\mu} F_{\nu\lambda]} &= 0 \ . \end{aligned} \tag{42}$$

These equations govern the behavior of electromagentic fields in general relativity.

4 Curvature

We have been loosely throwing around the idea of "curvature" without giving it a careful definition. The first step toward a better understanding begins with the notion of a **manifold**. Basically, a manifold is "a possibly curved space which, in small enough regions (infinitesimal, really), looks like flat space." You can think of the obvious example: the Earth looks flat because we only see a tiny part of it, even though it's round. A crucial feature of manifolds is that they have the same dimensionality everywhere; if you glue the end of a string to a plane, the result is not a manifold since it is partly one-dimensional and partly two-dimensional.

The most famous examples of manifolds are n-dimensional flat space $\mathbf{R}^n$ ("R" as in real, as in real numbers), and the n-dimensional sphere S^n. So, $\mathbf{R}^1$ is the real line, $\mathbf{R}^2$ is the plane, and so on. Meanwhile S^1 is a circle, S^2 is a sphere, etc. For future reference, the most popular coordinates on S^2 are the usual θ and ϕ angles. In these coordinates, the metric on S^2 (with radius $r = 1$) is

$$ds^2 = d\theta^2 + \sin^2\theta\, d\phi^2 \ . \tag{43}$$

The fact that manifolds may be curved makes life interesting, as you may imagine. However, most of the difficulties encountered in curved spaces are also encountered in flat space if you use non-Cartesian coordinates. The thing about curved space is, you can never use Cartesian coordinates, because they only describe flat spaces. So the machinery we developed for non-Cartesian coordinates will be crucial; in fact, we've done most of the work already.

It should come as no surprise that information about the curvature of a manifold is contained in the metric; the question is, how to extract it? You can't get it easily from the $\Gamma^\rho_{\mu\nu}$, for instance, since they can be zero or nonzero depending on the coordinate system (as we saw for flat space). For reasons we won't go into, the information about curvature is contained in a four-component tensor known as the **Riemann curvature tensor**. This supremely important object is given in terms of the Christoffel symbols by the formula

$$R^\sigma{}_{\mu\alpha\beta} \equiv \partial_\alpha \Gamma^\sigma_{\mu\beta} - \partial_\beta \Gamma^\sigma_{\mu\alpha} + \Gamma^\sigma_{\alpha\lambda}\Gamma^\lambda_{\mu\beta} - \Gamma^\sigma_{\beta\lambda}\Gamma^\lambda_{\mu\alpha} \ . \tag{44}$$

(The overall sign of this is a matter of convention, so check carefully when you read anybody else's papers. Note also that the Riemann tensor is constructed from non-tensorial elements — partial derivatives and Christoffel symbols — but they are carefully arranged so that the final result transforms as a tensor, as you can check.) This tensor has one nice property that a measure of curvature should have: *all of the components of $R^\sigma{}_{\mu\alpha\beta}$ vanish if and only if the space is flat.* Operationally, "flat" means that there exists a global coordinate system in which the metric components are everywhere constant.

There are two contractions of the Riemann tensor which are extremely useful: the **Ricci tensor** and the **Ricci scalar**. The Ricci tensor is given by

$$R_{\alpha\beta} = R^\lambda{}_{\alpha\lambda\beta} \ . \tag{45}$$

Although it may seem as if other independent contractions are possible (using other indices), the symmetries of $R^\sigma{}_{\mu\alpha\beta}$ (discussed below) make this the only independent contraction. The trace of the Ricci tensor yields the Ricci scalar:

$$R = R^\lambda{}_\lambda = g^{\mu\nu} R_{\mu\nu} \ . \tag{46}$$

This is another useful item.

Although the Riemann tensor has many indices, and therefore many components, using it is vastly simplified by the many symmetries it obeys. In fact, only 20 of the $4^4 = 256$ components of $R^\sigma{}_{\mu\alpha\beta}$ are independent. Here is a list of some of the useful properties obeyed by the Riemann tensor, which are most easily expressed in terms of the tensor with all indices lowered, $R_{\mu\nu\rho\sigma} = g_{\mu\lambda} R^\lambda{}_{\nu\rho\sigma}$:

$$\begin{aligned} R_{\mu\nu\rho\sigma} &= -R_{\mu\nu\sigma\rho} = -R_{\nu\mu\rho\sigma} \\ R_{\mu\nu\rho\sigma} &= R_{\rho\sigma\mu\nu} \\ R_{\mu\nu\rho\sigma} + R_{\mu\rho\sigma\nu} + R_{\mu\sigma\nu\rho} &= 0 \quad . \end{aligned} \tag{47}$$

These imply a symmetry of the Ricci tensor,

$$R_{\mu\nu} = R_{\nu\mu} \ . \tag{48}$$

In addition to these algebraic identities, the Riemann tensor obeys a differential identity:

$$\nabla_{[\lambda} R_{\mu\nu]\rho\sigma} = 0 \ . \tag{49}$$

This is sometimes known as the **Bianchi identity**. If we define a new tensor, the **Einstein tensor**, by

$$G_{\mu\nu} \equiv R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} \ , \tag{50}$$

then the Bianchi identity implies that the divergence of this tensor vanishes identically:

$$\nabla^\mu G_{\mu\nu} = 0 \ . \tag{51}$$

This is sometimes called the contracted Bianchi identity.

Basically, there are only two things you have to know about curvature: the Riemann tensor, and geodesics. You now know the Riemann tensor – lets move on to geodesics.

Informally, a **geodesic** is "the shortest distance between two points." More formally, a geodesic is a curve which extremizes the length functional $\int ds$. That is, imagine a path parameterized by λ, i.e. $x^\mu(\lambda)$. The infinitesimal distance along this curve is given by

$$ds = \sqrt{\left| g_{\mu\nu} \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda} \right|}\, d\lambda \ . \tag{52}$$

So the entire length of the curve is just

$$L = \int ds \ . \tag{53}$$

To find a geodesic of a given geometry, we would do a calculus of variations manipulation of this object to find an extremum of L. Luckily, stronger souls than ourselves have come before and done this for us. The answer is that $x^\mu(\lambda)$ is a geodesic if it satisfies the famous **geodesic equation**:

$$\frac{d^2 x^\mu}{d\lambda^2} + \Gamma^\mu_{\rho\sigma} \frac{dx^\rho}{d\lambda} \frac{dx^\sigma}{d\lambda} = 0. \tag{54}$$

In fact this is only true if λ is an **affine parameter**, that is if it is related to the proper time via

$$\lambda = a\tau + b \ . \tag{55}$$

In practice, the proper time itself is almost always used as the affine parameter (for timelike geodesics, at least). In that case, the tangent vector is the four-velocity $U^\mu = dx^\mu/d\tau$, and the geodesic equation can be written

$$\frac{d}{d\tau} U^\mu + \Gamma^\mu_{\rho\sigma} U^\rho U^\sigma = 0 \ . \tag{56}$$

The physical reason why geodesics are so important is simply this: *in general relativity, test bodies move along geodesics.* If the bodies are massless, these geodesics will be null ($ds^2 = 0$), and if they are massive the geodesics will be timelike ($ds^2 < 0$). Note that when we were being formal we kept saying "extremum" rather than "minimum" length. That's because, for massive test particles, the geodesics on which they move are curves of *maximum* proper time. (In the famous "twin paradox", two twins take two different paths through flat spacetime, one staying at home [thus on a geodesic], and the other traveling off into space and back. The stay-at-home twin is older when they reunite, since geodesics maximize proper time.)

This is an appropriate place to talk about the philosophy of GR. In pre-GR days, Newtonian physics said "particles move along straight lines, until forces knock them off." Gravity was one force among many. Now, in GR, gravity is represented by the curvature of spacetime, *not* by a force. From the GR point of view, "particles move along geodesics, until forces knock them off." Gravity doesn't count as a force. If you consider the motion of particles under the influence of forces other than gravity, then they won't move along geodesics – you can still use (54) to describe their motions, but you have to add a force term to the right hand side. In that sense, the geodesic equation is something like the curved-space expression for $\mathbf{F} = m\mathbf{a} = 0$.

5 General Relativity

Moving from math to physics involves the introduction of dynamical equations which relate matter and energy to the curvature of spacetime. In GR, the "equation of motion" for the metric is the famous **Einstein equation**:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi G T_{\mu\nu} \ . \tag{57}$$

Notice that the left-hand side is the Einstein tensor $G_{\mu\nu}$ from (50). G is Newton's constant of gravitation (*not* the trace of $G_{\mu\nu}$). $T_{\mu\nu}$ is a symmetric two-index tensor called the **energy-momentum tensor**, or sometimes the stress-energy tensor. It encompasses all we need to know about the energy and momentum of matter fields, which act as a source for gravity. Thus, the left hand side of this equation measures the curvature of spacetime, and the right measures the energy and momentum contained in it. Truly glorious.

The components $T_{\mu\nu}$ of the energy-momentum tensor are "the flux of the $\mu^{\rm th}$ component of momentum in the $\nu^{\rm th}$ direction." This definition is perhaps not very useful. More concretely, we can consider a popular form of matter in the context of general relativity: a **perfect fluid**, defined to be a fluid which is isotropic in its rest frame. This means that the fluid has no viscosity or heat flow; as a result, it is specified entirely in terms of the rest-frame energy density ρ and rest-frame pressure p (isotropic, and thus equal in all directions). If use U^μ to stand for the four-velocity of a fluid element, the energy-momentum tensor takes the form

$$T_{\mu\nu} = (\rho + p)U_\mu U_\nu + p g_{\mu\nu} \ . \tag{58}$$

If we raise one index and use the normalization $g^{\mu\nu}U_\mu U_\nu = -1$, we get an even more understandable version:

$$T_\mu{}^\nu = \begin{pmatrix} -\rho & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix} \ . \tag{59}$$

If $T_{\mu\nu}$ encapsulates all we need to know about energy and momentum, it should be able to characterize the appropriate conservation laws. In fact these are formulated by saying that the covariant divergence of $T_{\mu\nu}$ vanishes:

$$\nabla^\mu T_{\mu\nu} = 0 \ . \tag{60}$$

Recall that the contracted Bianchi identity (51) guarantees that the divergence of the Einstein tensor vanishes identically. So Einstein's equation (57) guarantees energy-momentum conservation. Of course, this is a local relation; if we (for example) integrate the energy density ρ over a spacelike hypersurface, the corresponding quantity is not constant with time. In GR there is no global notion of energy conservation; (60) expresses local conservation, and the appearance of the covariant derivative allows this equation to account for the transfer of energy back and forth between matter and the gravitational field.

The exotic appearance of Einstein's equation should not obscure the fact that it a natural extension of Newtonian gravity. To see this, consider Poisson's equation for the Newtonian potential Φ:

$$\nabla^2\Phi = 4\pi G\rho \ , \tag{61}$$

where ρ is the matter density. On the left hand side of this we see a second-order differential operator acting on the gravitational potential Φ. This is proportional to the density of matter. Now, GR is a fully relativistic theory, so we would expect that the matter density should be replaced by the full energy-momentum tensor $T_{\mu\nu}$. To correspond to (61), this should be proportional to a 2-index tensor which is a second-order differential operator acting on the gravitational field, i.e. the metric. If you think about the definition of $G_{\mu\nu}$ in terms of $g_{\mu\nu}$, this is exactly what the Einstein tensor is. In fact, $G_{\mu\nu}$ is the only two-index tensor, second order in derivatives of the metric, for which the divergence vanishes.

So the GR equation is of the same essential form as the Newtonian one. We should ask for something more, however: namely, that Newtonian gravity is recovered in the appropriate limit, where the particles are moving slowly (with respect to the speed of light), the gravitational field is weak (can be considered a perturbation of flat space), and the field is also static (unchanging with time). We consider a metric which is almost Minkowski, but with a specific kind of small perturbation:

$$ds^2 = -(1+2\Phi)dt^2 + (1-2\Phi)d\vec{x}^2 \ , \tag{62}$$

where Φ is a function of the spatial coordinates x^i. If we plug this into the geodesic equation and solve for the conventional three-velocity (using that the particles are moving slowly), we obtain

$$\frac{d^2\vec{x}}{dt^2} = -\nabla\Phi \ , \tag{63}$$

where ∇ here represents the ordinary spatial divergence (not a covariant derivative). This is just the equation for a particle moving in a Newtonian gravitational potential Φ. Meanwhile, we calculate the 00 component of the left-hand side of Einstein's equation:

$$R_{00} - \frac{1}{2}Rg_{00} = 2\nabla^2\Phi \ . \tag{64}$$

The 00 component of the right-hand side (to first order in the small quantities Φ and ρ) is just

$$8\pi G T_{00} = 8\pi G\rho \ . \tag{65}$$

So the 00 component of Einstein's equation applied to the metric (62) yields

$$\nabla^2\Phi = 4\pi G\rho \ , \tag{66}$$

which is precisely the Poisson equation (61). Thus, in this limit GR does reduce to Newtonian gravity.

Although the full nonlinear Einstein equation (57) looks simple, in applications it is not. If you recall the definition of the Riemann tensor in terms of the Christoffel symbols, and the definition of those in terms of the metric, you realize that Einstein's equation for the metric are complicated indeed! It is also highly nonlinear, and correspondingly very difficult to solve. If we take the trace of (57), we obtain

$$-R = 8\pi G T \ . \tag{67}$$

Plugging this into (57), we can rewrite Einstein's equations as

$$R_{\mu\nu} = 8\pi G\left(T_{\mu\nu} - \frac{1}{2}Tg_{\mu\nu}\right) \ . \tag{68}$$

This form is useful when we consider the case when we are in the vacuum – no energy or momentum. In this case $T_{\mu\nu} = 0$ and (68) becomes Einstein's equation in vacuum:

$$R_{\mu\nu} = 0 \ . \tag{69}$$

This is somewhat easier to solve than the full equation.

One final word on Einstein's equation: it may be derived from a very simple Lagrangian, $\mathcal{L} = \sqrt{-g}R$ (plus appropriate terms for the matter fields). In other words, the action for GR is simply

$$S = \int d^4x\sqrt{-g}R \ , \tag{70}$$

an Einstein's equation comes from looking for extrema of this action with respect to variations of the metric $g_{\mu\nu}$. What could be more elegant?

6 Schwarzschild solution

In order to solve Einstein's equation we usually need to make some simplifying assumptions. For example, in many physical situations, we have spherical symmetry. If we want to solve for a metric $g_{\mu\nu}$, this fact is very helpful, because the most general spherically symmetric metric may be written (in spherical coordinates) as

$$ds^2 = -A(r,t)dt^2 + B(r,t)dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) \ , \tag{71}$$

where A and B are positive functions of (r,t), and you will recognize the metric on the sphere from (43). If we plug this into Einstein's equation, we will get a solution for a spherically symmetric matter distribution. To be even more restrictive, let's consider the equation in vacuum, (69). Then there is a unique solution:

$$ds^2 = -\left(1 - \frac{2Gm}{r}\right)dt^2 + \left(1 - \frac{2Gm}{r}\right)^{-1}dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2) \ . \tag{72}$$

This is the celebrated **Schwarzschild metric** solution to Einstein's equations. The parameter m, of course, measures the amount of mass inside the radius r under consideration. A remarkable fact is that *the Schwarzschild metric is the unique solution to Einstein's equation in vacuum with a spherically symmetric matter distribution.* This fact, known as **Birkhoff's theorem**, means that the matter can oscillate wildly, as long as it remains spherically symmetric, and the gravitational field outside will remain unchanged.

Philosophy point: the metric components in (72) blow up at $r = 0$ and $r = 2Gm$. Officially, any point at which the metric components become infinite, or exhibit some other pathological behavior, is known as a **singularity**. These beasts come in two types: "coordinate" singularities and "true" singularities. A coordinate singularity is simply a result of choosing bad coordinates; if we change coordinates we can remove the singularity. A true singularity is an actual pathology of the geometry, a point at which the manifold is ill-defined. In the Schwarzschild geometry, the point $r = 0$ is a real singularity, an unavoidable blowing-up. However, the point $r = 2Gm$ is merely a coordinate singularity. We can demonstrate this by making a transformation to what are known as **Kruskal coordinates**, defined by

$$\begin{aligned} u &= \left(\frac{r}{2Gm} - 1\right)^{1/2} e^{r/4Gm}\cosh(t/4Gm) \\ v &= \left(\frac{r}{2Gm} - 1\right)^{1/2} e^{r/4Gm}\sinh(t/4Gm). \end{aligned} \tag{73}$$

In these coordinates, the metric (72) takes the form

$$ds^2 = \frac{32(Gm)^3}{r}e^{-r/2Gm}(-dv^2 + du^2) + r^2(d\theta^2 + \sin^2\theta d\phi^2) \ , \tag{74}$$

where r is considered to be an implicit function of u and v defined by

$$u^2 - v^2 = e^{r/2Gm}\left(\frac{r}{2Gm} - 1\right) . \tag{75}$$

If we look at (74), we see that nothing blows up at $r = 2Gm$. The mere fact that we could choose coordinates in which this happens assures us that $r = 2Gm$ is a mere coordinate singularity.

The useful thing about the Schwarzschild solution is that it describes both mundane things like the solar system, and more exotic objects like black holes. To get a feel for it, let's look at how particles move in a Schwarzschild geometry. It turns out that we can cast the problem of a particle moving in the plane $\theta = \pi/2$ as a one-dimensional problem for the radial coordinate $r = r(\tau)$. In other words, the distance of a particle from the point $r = 0$ is a solution to the equation

$$\frac{1}{2}\left(\frac{dr}{d\tau}\right)^2 + V(r) = \frac{1}{2}E^2 . \tag{76}$$

This is just the equation of motion for a particle of unit mass and energy E in a one-dimensional potential $V(r)$. This potential, for the Schwarzschild geometry, is given by

$$V(r) = \frac{1}{2}\epsilon - \epsilon\frac{Gm}{r} + \frac{L^2}{2r^2} - \frac{GmL^2}{r^3} . \tag{77}$$

Here, L represents the angular momentum (per unit mass) of the particle, and ϵ is a constant equal to 0 for massless particles and +1 for massive particles. (Note that the proper time τ is zero for massless particles, so we use some other parameter λ in (76), but the equation itself looks the same). So, to find the orbits of particles in a Schwarzschild metric, just solve the motion of a particle in the potential given by (77). Note that the first term in (77) is a constant, the second term is exactly what we expect from Newtonian gravity, and the third term is just the contribution of the particle's angular momentum, which is also present in the Newtonian theory. Only the last term in (77) is a new addition from GR.

There are two important effects of this extra term. First, it acts as a small perturbation on any orbit – this is what leads to the precession of Mercury, for instance. Second, for r very small, the GR potential goes to $-\infty$; this means that a particle that approaches too close to $r = 0$ will fall into the center and never escape! Even though this is in the context of unaccelerated test particles, a similar statement holds true for particles with the ability to accelerate themselves all they like – see below. However, not to worry; for a star such as the Sun, for which the Schwarzschild metric only describes points outside the surface, you would run into the star long before you approached the point where you could not escape.

Nevertheless, we all know of the existence of more exotic objects: **black holes**. A black hole is a body in which all of the mass has collapsed gravitationally past the point of possible escape. This point of no return, given by the surface $r = 2Gm$, is known as the **event horizon**, and can be thought of as the "surface" of a black hole. Although it is impossible to go into much detail about the host of interesting properties of the event horizon, the basics are not difficult to grasp. From the point of view of an outside observer, a clock falling into a black hole will appear to move more and more slowly as it approaches the event horizon. In fact, the external observer will never see a test particle cross the surface $r = 2Gm$; they will just see the particle get closer and closer, and move more and more slowly.

Contrast this to what you would experience as a test observer actually thrown into a black hole. To you, time always seems to move at the same rate; since you and your wristwatch are in the same inertial frame, you never "feel time moving more slowly." Therefore, rather than taking an infinite amount of time to reach the event horizon, you zoom right past – doesn't take very long at all, actually. You then proceed directly to fall to $r = 0$, also in a very short time. *Once you pass* $r = 2Gm$*, you cannot help but hit* $r = 0$; it is as inevitable as moving forward in time. The literal truth of this statement can be seen by looking at the metric (72) and noticing that r becomes a *timelike* coordinate for $r < 2Gm$; therefore your voyage to the center of the black hole is literally moving forward in time! What's worse, we noted above that a geodesic (unaccelerated motion) maximized the proper time – this means that the more you struggle, the sooner you will get there. (Of course, you won't struggle, because you would have been ripped to shreds by tidal forces. The grisly death of an astrophysicist who enters a black hole is detailed in Misner, Thorne, and Wheeler, pp. 860-862.)

The spacetime diagram of a black hole in Kruskal coordinates (74) is shown in Figure 2. Shown is a slice through the entire spacetime, corresponding to angular coordinates $\theta = \pi/2$ and $\phi = 0$. There are two asymptotic regions, one at $u \to +\infty$ and the other at $u \to -\infty$; in both regions the metric looks approximately flat. The event horizon is the surface $r = 2Gm$, or equivalently $u = \pm v$. In this diagram all light cones are at $\pm 45°$. Inside the event horizon, where $r < 2Gm$, all timelike trajectories lead inevitably to the singularity at $r = 0$. It should be stressed that this diagram represents the "maximally extended" Schwarzschild solution — a complete solution to Einstein's equation in vacuum, but not an especially physically realistic one. In a realistic black hole, formed for instance from the collapse of a massive star, the vacuum equations do not tell the whole story, and there will not be two distinct asymptotic regions, only the one in which the star originally was located. (For that matter, timelike trajectories cannot travel between the two regions, so we could never tell whether another such region did exist.)

In the collapse to a black hole, all the information about the detailed nature of the collapsing object is lost: what it was made of, its shape, etc. The only information which is not wiped out is the amount of mass, angular momentum, and electric charge in the hole. This fact, the **no-hair theorem**, implies that the most general black-hole metric will be a function of these three numbers only. However, real-world black holes will prob-

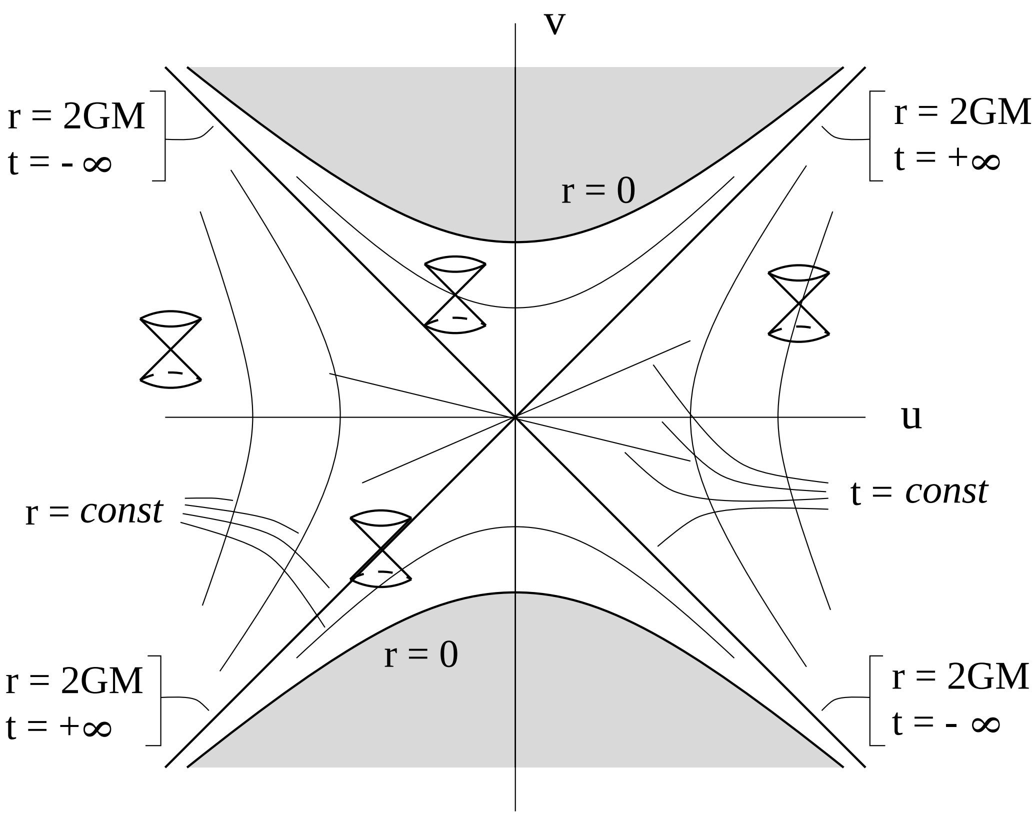

Figure 2: The Kruskal diagram — the Schwarzschild solution in Kruskal coordinates (74), where all light cones are at $\pm 45°$. The surface $r = 2Gm$ is the event horizon; inside the event horizon, all timelike paths hit the singularity at $r = 0$. The right- and left-hand side of the diagram represent distinct asymptotically flat regions of spacetime.

ably be electrically neutral, so we will not present the metric for a charged black hole (the **Reissner-Nordstrom metric**). Of considerable astrophysical interest are spinning black holes, described by the **Kerr metric**:

$$ds^2 = -\left[\frac{\Delta - \omega^2 \sin^2\theta}{\Sigma}\right] dt^2 - \left[\frac{4\omega m G r \sin^2\theta}{\Sigma}\right] dt d\phi + \frac{\Sigma}{\Delta} dr^2 + \Sigma d\theta^2 + \left[\frac{(r^2+\omega^2)^2 - \Delta\omega^2 \sin^2\theta}{\Sigma}\right] \sin^2\theta d\phi^2, \quad (78)$$

where

$$\Sigma \equiv r^2 + \omega^2 \cos^2\theta, \qquad \Delta \equiv r^2 + \omega^2 - 2Gmr\ , \quad (79)$$

and ω is the angular velocity of the body.

Finally, among the many additional possible things to mention, there's the **cosmic censorship conjecture**. Notice how the Schwarzschild singularity at $r = 0$ is hidden, in a sense – you can never get to it without crossing an horizon. It is conjectured that this is *always* true, in any solution to Einstein's equation. However, some numerical work seems to contradict this conjecture, at least in special cases.

7 Cosmology

Just as we were able to make great strides with the Schwarzschild metric on the assumption of sperical symmetry, we can make similar progress in cosmology by assuming that the Universe is homogeneous and isotropic. That is to say, we assume the existence of a "rest frame for the Universe," which defines a universal time coordinate, and singles out three-dimensional surfaces perpendicular to this time coordinate. (In the real Universe, this rest frame is the one in which galaxies are at rest and the microwave background is isotropic.) "Homogeneous" means that the curvature of any two points at a given time t is the same. "Isotropic" is trickier, but basically means that the universe looks the same in all directions. Thus, the surface of a cylinder is homogeneous (every point is the same) but not isotropic (looking along the long axis of the cylinder is a preferred direction); a cone is isotropic around its vertex, but not homogeneous.

These assumptions narrow down the choice of metrics to precisely three forms, all given by the **Robertson-Walker (RW) metric**:

$$ds^2 = -dt^2 + a^2(t)\left[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right]\ , \quad (80)$$

where the constant k can be -1, 0, or $+1$. The function $a(t)$ is known as the **scale factor** and tells us the relative sizes of the spatial surfaces. The above coordinates are called **comoving coordinates**, since a point which is at rest in the preferred frame of the universe will have $r,\ \theta,\ \phi = constant$. The $k = -1$ case is known as an **open** universe, in which the preferred three-surfaces are "three-hyperboloids" (saddles); $k = 0$ is a **flat** universe, in which the preferred three-surfaces are flat space; and $k = +1$ is a **closed** universe, in which the preferred three-surfaces are three-spheres. *Note that the terms "open," "closed," and "flat" refer to the spatial geometry of three-surfaces, not to whether the universe will eventually recollapse.* The volume of a closed universe is finite, while open and flat universes have infinite volume (or at least they can; there are also versions with finite volume, obtained from the infinite ones by performing discrete identifications).

There are other coordinate systems in which (8.1) is sometimes written. In particular, if we set $r = (\sin\psi,\ \psi,\ \sinh\psi)$ for $k = (+1,\ 0,\ -1)$ respectively, we obtain

$$ds^2 = -dt^2 + a^2(t)\left\{\begin{array}{c} d\psi^2 + \sin^2\psi(d\theta^2 + \sin^2\theta d\phi^2) \\ d\psi^2 + \quad \psi^2(d\theta^2 + \sin^2\theta d\phi^2) \\ d\psi^2 + \sinh^2\psi(d\theta^2 + \sin^2\theta d\phi^2) \end{array}\right\} \begin{array}{c} (k = +1) \\ (k = 0) \\ (k = -1) \end{array} \tag{81}$$

Further, the flat ($k = 0$) universe also may be written in almost-Cartesian coordinates:

$$\begin{aligned} ds^2 &= -dt^2 + a^2(t)(dx^2 + dy^2 + dz^2) \\ &= -a^2(\eta)(-d\eta^2 + dx^2 + dy^2 + dz^2). \end{aligned} \tag{82}$$

In this last expression, η is known as the **conformal time** and is defined by

$$\eta \equiv \int \frac{dt}{a(t)} \ . \tag{83}$$

The coordinates $(\eta,\ x,\ y,\ z)$ are often called "conformal coordinates."

Since the RW metric is the only possible homogeneous and isotropic metric, all we have to do is solve for the scale factor $a(t)$ by using Einstein's equation. If we use the vacuum equation (69), however, we find that the only solution is just Minkowski space. Therefore we have to introduce some energy and momentum to find anything interesting. Of course we shall choose a perfect fluid specified by energy density ρ and pressure p. In this case, Einstein's equation becomes two differential equations for $a(t)$, known as the **Friedmann equations**:

$$\begin{aligned} \left(\frac{\dot a}{a}\right)^2 &= \frac{8\pi G}{3}\rho - \frac{k}{a^2} \\ \frac{\ddot a}{a} &= -\frac{4\pi G}{3}(\rho + 3p) \ . \end{aligned} \tag{84}$$

Since the Friedmann equations govern the evolution of RW metrics, one often speaks of Friedman-Robertson-Walker (FRW) cosmology.

The expansion rate of the universe is measured by the **Hubble parameter**:

$$H \equiv \frac{\dot a}{a} \ , \tag{85}$$

and the change of this quantity with time is parameterized by the **deceleration parameter**:

$$q \equiv -\frac{\ddot{a}a}{\dot{a}^2} = -\left(1 + \frac{\dot{H}}{H^2}\right) . \tag{86}$$

The Friedmann equations can be solved once we choose an equation of state, but the solutions can get messy. It is easy, however, to write down the solutions for the $k = 0$ universes. If the equation of state is $p = 0$, the universe is **matter dominated**, and

$$a(t) \propto t^{2/3} . \tag{87}$$

In a matter dominated universe, the energy density decreases as the volume increases, so

$$\rho_{matter} \propto a^{-3} . \tag{88}$$

If $p = \frac{1}{3}\rho$, the universe is **radiation dominated**, and

$$a(t) \propto t^{1/2} . \tag{89}$$

In a radiation dominated universe, the number of photons decreases as the volume increases, and the energy of each photon redshifts and amount proportional to $a(t)$, so

$$\rho_{rad} \propto a^{-4} . \tag{90}$$

If $p = -\rho$, the universe is **vacuum dominated**, and

$$a(t) \propto e^{Ht} . \tag{91}$$

The vacuum dominated universe is also known as **de Sitter space**. In de Sitter space, the energy density is *constant*, as is the Hubble parameter, and they are related by

$$H = \sqrt{\frac{8\pi G\rho_{vac}}{3}} = constant . \tag{92}$$

Note that as $a \to 0$, ρ_{rad} grows the fastest; therefore, if we go back far enough in the history of the universe we should come to a radiation dominated phase. Similarly, ρ_{vac} stays constant as the universe expands; therefore, if ρ_{vac} is not zero, and the universe lasts long enough, we will eventually reach a vacuum-dominated phase.

Given that our Universe is presently expanding, we may ask whether it will continue to do so forever, or eventually begin to recontract. For energy sources with $p/\rho \geq 0$ (including both matter and radiation dominated universes), closed ($k = +1$) universes will eventually recontract, while open and flat universes will expand forever. When we let $p/\rho < 0$ things get messier; just keep in mind that spatially closed/open does not necessarily correspond to temporally finite/infinite.

The question of whether the Universe is open or closed can be answered observationally. In a flat universe, the density is equal to the **critical density**, given by

$$\rho_{crit} = \frac{3H^2}{8\pi G} \ . \tag{93}$$

Note that this changes with time; in the present Universe it's about 5×10^{-30} grams per cubic centimeter. The universe will be open if the density is less than this critical value, closed if it is greater. Therefore, it is useful to define a **density parameter** via

$$\Omega \equiv \frac{\rho}{\rho_{crit}} = \frac{8\pi G\rho}{3H^2} = 1 + \frac{k}{\dot{a}^2} \ , \tag{94}$$

a quantity which will generally change with time unless it equals unity. An open universe has $\Omega < 1$, a closed universe has $\Omega > 1$.

We mentioned in passing the **redshift** of photons in an expanding universe. In terms of the wavelength λ_1 of a photon emitted at time t_1, the wavelength λ_0 observed at a time t_0 is given by

$$\frac{\lambda_0}{\lambda_1} = \frac{a(t_0)}{a(t_1)} \ . \tag{95}$$

We therefore define the redshift z to be the fractional increase in wavelength

$$z \equiv \frac{\lambda_0 - \lambda_1}{\lambda_1} = \frac{a(t_0)}{a(t_1)} - 1 \ . \tag{96}$$

Keep in mind that this only measures the net expansion of the universe between times t_1 and t_0, not the relative speed of the emitting and observing objects, especially since the latter is not well-defined in GR. Nevertheless, it is common to speak as if the redshift is due to a Doppler shift induced by a relative velocity between the bodies; although nonsensical from a strict standpoint, it is an acceptable bit of sloppiness for small values of z. Then the Hubble constant relates the redshift to the distance s (measured along a spacelike hypersurface) between the observer and emitter:

$$z = H(t_0)s \ . \tag{97}$$

This, of course, is the linear relationship discovered by Hubble.