# Geometric View of Classsical Physics

Johar M. Ashfaque

# Contents

# Part I
# Differential Geometry

## 1 Manifolds and vectors

### 1.1 Manifolds

An n-dimensional **manifold** is a topological space that can locally be identified with $\mathbf{R}^n$. That means it can be covered by open neighbourhoods which map one-to-one into open neighbourhoods of $\mathbf{R}^n$:

$$\begin{aligned} \varphi : M \supset U &\rightarrow V \subset \mathbf{R}^n, \\ p &\mapsto x^i := (x^1, x^2, \dots x^n). \end{aligned} \tag{1}$$

Such a map is called a **chart**, or a **coordinate system**. A collection of charts that covers the manifold is called an **atlas**. This definition becomes nontrivial only where two charts overlap. We then have a map from $\mathbf{R}^n$ to $\mathbf{R}^n$

$$\begin{aligned} \varphi_1 : p &\mapsto x_1^i, \\ \varphi_2 : p &\mapsto x_2^i, \\ \psi = \varphi_2 \circ \varphi_1^{-1} : \varphi_1(U_1 \cap U_2) &\rightarrow \varphi_2(U_1 \cap U_2) \\ x_1^i &\mapsto x_2^i. \end{aligned} \tag{2}$$

If all such **coordinate transformations** $\psi$ are $C^k$, we say the manifold is $C^k$. The type of manifold most used in physics is smooth ($C^\infty$).

### 1.2 Vectors

In going from $\mathbf{R}^n$ to an $n$-dimensional manifold $M$, we have kept the idea of coordinates (locally, each point $p$ has unique coordinates $x^i$), but we have lost the concept of subtracting two position vectors to get a direction vector: "$p-q$" is not defined, and $x^i - y^i$ makes sense as a vector only for Cartesian coordinates on $\mathbf{R}^n$.

We get vectors back as tangent vectors to curves. Let

$$\begin{aligned} c : I &\rightarrow M \\ t &\mapsto p(t) \end{aligned} \tag{3}$$

be a **curve** in $M$ with parameter $t$. In coordinates this is

$$\begin{aligned} \varphi \circ c : I &\rightarrow \mathbf{R}^n \\ t &\mapsto x^i(t). \end{aligned} \tag{4}$$

**Definition 1:** The **vector** tangent to the curve $c$ at the point $p = c(a)$ is

$$X^i = \left.\frac{dx^i}{dt}\right|_{t=a} \tag{5}$$

In a different chart $\varphi'$,

$$\begin{aligned} \varphi' \circ c : I &\rightarrow \mathbf{R}^n \\ t &\mapsto x'^i(t), \end{aligned}$$

the object

$$X'^i = \left.\frac{dx'^i}{dt}\right|_{t=a}$$

is a different $n$-tuple real numbers, but it should represent the same vector. We have

$$X'^i = \frac{dx'^i}{dt} = \sum_{j=1}^{n} \frac{\partial x'^i}{\partial x^j}\frac{dx^j}{dt} = \sum_{j=1}^{n} \frac{\partial x'^i}{\partial x^j} X^j. \tag{6}$$

From now on, we will use the **Einstein summation convention**, by which any pair of one index up and one index down (here $j$) is implicitly summed over without the need to write the sum explicitly.

This transformation leads us to an alternative but equivalent definition of a vector (now using the summation convention),

**Definition 2:** A vector is an $n$-tuple of real numbers $X^i$ that transforms under a coordinate transformation as

$$X'^i = \frac{\partial x'^i}{\partial x^j} X^j. \tag{7}$$

Yet another definition is often useful, and equivalent to the two above, namely of a vector as a directional derivative acting on functions on the manifold.

**Definition 3:** Let

$$\begin{aligned} f : M &\to \mathbf{R} \\ p &\mapsto f(p) \end{aligned} \tag{8}$$

be a scalar function on $M$. (A **scalar** is a number that does not change under a coordinate transformation.)

Let $X$ be the tangent vector at $p$ to the curve $c$. Then we can define $X$ as a map from any scalar function $f$ to a number $Xf$,

$$X : f \mapsto Xf = \left.\frac{d}{dt}\right|_{t=a} f\left(c(t)\right). \tag{9}$$

In local coordinates,

$$\varphi \circ c : t \mapsto x^i(t) \tag{10}$$

and

$$f \circ \varphi^{-1} : x^i \mapsto f(x^i). \tag{11}$$

Using the chain rule,

$$Xf = \frac{d}{dt} f\left(x^i(t)\right) = \frac{\partial f}{\partial x^i}\frac{dx^i}{dt} = X^i \frac{\partial f}{\partial x^i}. \tag{12}$$

Under a change of coordinates

$$Xf = X^i \frac{\partial f}{\partial x^i} = X^i \frac{\partial f}{\partial x'^j}\frac{\partial x'^j}{\partial x^i} = X'^j \frac{\partial f}{\partial x'^j}, \tag{13}$$

which is consistent with our second definition. We can also write this third definition as

$$X = X^i \frac{\partial}{\partial x^i} \tag{14}$$

by removing the function $f$, where once again summation is implied. We can then think of the $\partial/\partial x^i$ for $i = 1 \ldots n$ either as derivative operators, or simply as the **coordinate basis of vector fields** at a point. The $X^i$ are called the **components** of $X$ in this coordinate basis.

## 1.3 *Exercises*

1. Explain carefully how our definition of a manifold fixes its dimension.
2. How could one define a scalar function on $M$ to be $C^k$?
3. Show briefly that the three definitions of vectors obey the axioms of a vector space (closed under addition and multiplication by a number).
4. Give two local coordinate systems on the 2-sphere $S^2$. Give an example of one vector field in both coordinate systems.

# 2 Tensors

## 2.1 Definitions

We define the **tangent space** $T_pM$ to the manifold $M$ at the point $p$ as the vector space of all possible vectors at $p$. Define a **covector** $\omega$ at $p$ as a linear map on $T_pM$:

$$\begin{aligned} \omega : T_pM &\rightarrow \mathbf{R} \\ X &\mapsto \omega(X) \end{aligned} \tag{15}$$

Linearity means that $\omega(\alpha X + Y) = \alpha\omega(X) + Y$, and so for a given chart $\varphi$ there exist numbers $\omega_i$ such that

$$\omega(X) = \omega_i X^i. \tag{16}$$

(Note that index position matters: coordinate and vector indices are up, covector indices are down. $\partial/\partial x^i$ also has its index down. Only index pairs where one index is up and one is down are contracted under the Einstein summation convention.) In a given coordinate system, $\omega$ can also be expanded in the **coordinate basis of covector fields**

$$\omega = \omega_i dx^i. \tag{17}$$

(The connection of the notation $dx^i$ with the usual notation $\int dx$ will become clearer in Sec. 6.2.)

Under a change of coordinates

$$\omega(X) = \omega_i X^i = \omega_i \left( \frac{\partial x^i}{\partial x'^j} X'^j \right) = \left( \omega_i \frac{\partial x^i}{\partial x'^j} \right) X'^j \equiv \omega_j{}' X'^j, \tag{18}$$

so covectors transform as

$$\omega'_j = \frac{\partial x^i}{\partial x'^j} \omega_i \tag{19}$$

The space of all covectors at $p$ is again a vector space, and is by definition the **dual space** of $T_pM$. It is called $T_p^*M$, the **cotangent space** of $M$ at $p$.

Now we define a **tensor** of **rank** $(r, s)$ as a multilinear map on $r$ covectors and $s$ vectors,

$$\begin{aligned} T : (T_p^*M)^r \times (T_pM)^s &\rightarrow \mathbf{R} \\ \omega_1, \omega_2, \ldots, \omega_r, X_1, X_2, \ldots X_s &\mapsto T(\omega_1, \omega_2, \ldots, \omega_r, X_1, X_2, \ldots X_s) \end{aligned} \tag{20}$$

where

$$T(\ldots, \alpha\omega + \mu, \ldots) = \alpha T(\ldots, \omega, \ldots) + T(\ldots, \mu, \ldots), \tag{21}$$

and so for every "slot" of the tensor $T$. Note that the order of the slots matter.

From multilinearity it follows that there are numbers, called **tensor components** $T^{i\ldots j}{}_{k\ldots l}$ such that, for example the example of a $(2,1)$-tensor $T$,

$$T(\omega, \mu, X) = T^{ij}{}_k \, \omega_i \, \mu_j \, X^k, \tag{22}$$

and similarly for tensors of other rank.

Clearly tensor components transform as

$$T'^{ij\ldots}{}_{k\ldots} = \frac{\partial x'^i}{\partial x^l}\frac{\partial x'^j}{\partial x^m}\ldots\frac{\partial x^n}{\partial x'^k}\ldots T^{lm\ldots}{}_{k\ldots}. \tag{23}$$

Here the dots represent an additional factor on the right-hand side for each additional index on $T$: of the form $\partial x'/\partial x$ for up indices, and of the form $\partial x/\partial x'$ for down indices: it should be clear how all the indices match up. In particular, a vector is a $(1,0)$ tensor, a covector is a $(0,1)$ tensor, and a scalar is a $(0,0)$ tensor. Clearly, all tensors of the same rank form a vector space.

From the transformation properties of tensors, it follows that a $(1,1)$-tensor can also be seen as a map $T_pM \to T_pM$ or $T_p^*M \to T_p^*M$.

## 2.2 Contraction and abstract index notation

$\omega(X) = \omega_i X^i$ is the simplest example of the **contraction** of two tensors. The result in this case is a scalar. But we can always contract an up index on one tensor with a down index on another tensor to obtain a new tensor. The same thing can be done with more than one pair of indices. To save notation, we present an example rather than the general case. Take two tensors $S$ and $T$ whose components in a given chart are $S_{ij}$ and $T^{kl}$. Define a new $(1,1)$ tensor $S \cdot T$ by defining its components as

$$(S \cdot T)_i{}^j = S_{ik}T^{kj}, \tag{24}$$

again using the summation convention.

We still need to show that this really is a tensor, meaning that it transforms correctly. Transforming each tensor separately, we have

$$S_{ik}T^{kj} = \left(S'_{pq}\frac{\partial x'^p}{\partial x^i}\frac{\partial x'^q}{\partial x^k}\right)\left(T^{rs}\frac{\partial x^k}{\partial x'^r}\frac{\partial x^j}{\partial x'^s}\right). \tag{25}$$

From the chain rule for partial derivatives we have

$$\frac{\partial x'^q}{\partial x^k}\frac{\partial x^k}{\partial x'^r} = \frac{\partial x'^q}{\partial x'^r} = \delta^q{}_r, \tag{26}$$

meaning this is 1 if $q = r$ and 0 otherwise. Also

$$S'_{pq}\delta^q{}_r = S'_{pr} \tag{27}$$

and so

$$S_{ik}T^{kj} = \frac{\partial x'^p}{\partial x^i}\frac{\partial x^j}{\partial x'^s}\left(S'_{pr}T'^{rs}\right), \tag{28}$$

which is the correct transformation for a $(1,1)$ tensor.

$\omega \cdot X$ can only mean $\omega_i X^i$, but there are four ways of contracting $S$ and $T$ to obtain a $(1,1)$ tensor, so $S_{ik}T^{kj}$ is clearer notation than $S \cdot T$ because it tells us which index pair has been contracted. Often, the best notation to use in tensor equations is $S_{ik}T^{kj}$, but considering this not as a collection of tensor components in a specific coordinate system, but rather as a notation for the tensor as a whole. This is called the **abstract index notation**.

Wald consistently uses Latin indices for abstract tensors, and Greek indices for components in a particular basis. However, the use of $(i, j, k, \ldots)$ as indices is

so common that I will use $(i, j, k, l, \ldots)$ as both coordinate and abstract indices. However, if I want to be quite clear, which is meant, I will only ever use lower-case Greek letters $(\alpha, \beta, \ldots, \mu, \nu, \ldots)$ for coordinate indices, and I will only ever use lower-case Latin letters $(a, b, c, d, e, f)$ for abstract tensor indices.

Note that the order of the indices matters, so we should not write $T^i{}_j$ as $T^i_j$, which is ambiguous. Note also that an index that is summed over, also called a **dummy index**, can be renamed at will, so for example $T^i{}_i$ and $T^j{}_j$ are the same scalar.

Finally, if we put a tensor (for example a vector) at each point of the manifold, and its components are $C^k$ functions of the coordinates, then we have a $C^k$ **tensor field**.

## 2.3 The metric tensor

A tensor field that comes up particularly often in physics is the metric tensor. The physical, 3-dimensional space of Newtonian physics comes equipped with an **inner product** or **dot product** of vectors, which in Cartesian coordinates is given by

$$X \cdot Y := X^1 Y^1 + X^2 Y^2 + X^3 Y^3 \tag{29}$$

In particular, this gives us the **length** or **absolute value** of a vector

$$|X| := \sqrt{X \cdot X} \tag{30}$$

and the **angle** $\theta$ between two vectors by

$$\cos\theta := \frac{X \cdot Y}{|X|\,|Y|}. \tag{31}$$

We can define the inner product in geometric form as

$$X \cdot Y := X^a Y^b \gamma_{ab} \tag{32}$$

if we define the (0,2)-tensor $\gamma_{ab}$ by defining that its components in Cartesian coordinates are $\gamma_{ij} = \delta_{ij}$, that is, 1 for $i = j$ and 0 otherwise. It is clear that in Cartesian coordinates this is the definition (29), and that as $X^a$ and $Y^a$ transform as vectors and $\gamma_{ab}$ as a tensor, then their inner product transforms, correctly, as a scalar.

More generally any symmetric tensor $g_{ab} = g_{ba}$ that is invertible can be a **metric tensor**. Note that the metric $\gamma_{ab}$ of Euclidean space has the additional property that it is **positive definite**, meaning that $X^a X^b \gamma_{ab} > 0$ for $X^a \neq 0$, but we will soon see that the metric $\eta_{ab}$ of spacetime is not positive definite. (If we treat the metric tensor in any given coordinate system as a matrix, then it is positive definite if all its eigenvalues are positive, and invertible if all its eigenvalues are non-zero.)

For any metric $g_{ab}$, is useful to define the **inverse metric** $g^{ab}$ by defining

$$g^{ab} g_{bc} := \delta^a{}_c, \tag{33}$$

where the $(1,1)$-tensor $\delta_a{}^b$ represents the identity operator. This is to say that in *any* coordinate system, $g^{ij}$ written out as a (symmetric) matrix is the matrix inverse of $g_{ij}$. (See exercise).

If a metric is around, one often uses the metric to **raise or lower indices** on a given tensor, that is if a tensor $T^a$ is defined, then one would *implicitly* define the tensor $T_a$ as $g_{ab}T^b$, and so on, using the same stem (here $T$) for all index positions. Note that $g_{ac}g_{bd}g^{cd} = g_{ac}\delta_b{}^c = g_{ab}$, so this notation is consistent even for the metric tensor itself. Any metric tensor obeys the identity $g_a{}^b = \delta_a{}^b$.

## 2.4 *Exercises*

1. Show that if the components of a tensor are $\delta^i{}_j$ (meaning 1 if $i = j$, and 0 otherwise) in one coordinate system, they are $\delta^i{}_j$ in every coordinate system.

2. Explain in what sense a (1,1)-tensor $M^a{}_b$ can be considered as a map $TM \to TM$, or equivalently as a matrix. What are the components of the unit tensor?

# 3 Maps of manifolds

## 3.1 Motivation: derivatives of vector fields

We would like to define derivatives of tensor fields, for example of a vector field $X^i$. One could try the directional derivative of $X^i$ along the vector $Y$,

$$Y^j \frac{\partial X^i}{\partial x^j}, \tag{34}$$

but this object does not transform as a vector field. Or one could try

$$\frac{\partial X^i}{\partial x^j}, \tag{35}$$

but this does not transform as a $(1,1)$ tensor.

However, genuine tensors with similar properties can be defined, and they are the Lie derivative, the covariant derivative, and the exterior derivative, respectively. All have their applications in physics. All have some restrictions: the Lie derivative will require the derivatives $Y^j{}_{,i}$ of $Y^j$, so it is not quite a derivative at one point. The covariant derivative will require us to specify an additional geometric structure on the manifold, called a connection. Finally, the exterior derivative requires neither, but only works on totally antisymmetric tensors with all indices down, also called differential forms.

Behind the Lie derivative is the idea of a map from $M$ to $M$ and which depends differentiably on a parameter $t$, so that it "moves points along". We use this to define the Lie derivative along the tangent vector to a curve. It is clearer if we start with maps from $M$ to another manifold $N$, and only later set $N = M$.

## 3.2 Maps of manifolds

Let $M$ and $N$ be two manifolds, with dimensions $m$ and $n$ and

$$\begin{aligned} \phi : M &\rightarrow N \\ p &\mapsto \phi(p) \end{aligned} \tag{36}$$

a map from $M$ to $N$. In local charts on each manifold

$$\begin{aligned} \varphi : M &\rightarrow \mathbf{R}^m \\ p &\mapsto x^i \end{aligned} \tag{37}$$

and

$$\begin{aligned} \varphi' : N &\rightarrow \mathbf{R}^n \\ q &\mapsto y^\alpha \end{aligned} \tag{38}$$

For a function (scalar field) $f$ on $N$

$$\begin{aligned} f : N &\rightarrow \mathbf{R} \\ q &\mapsto f(q) \end{aligned} \tag{39}$$

we define its **pull-back** to M by

$$\begin{aligned} \phi_* f : M &\to \mathbf{R} \\ p &\mapsto (\phi_* f)(p) := f(\phi(p)) \end{aligned} \tag{40}$$

For a vector field $V$ on $M$ we can then define its **push-forward** to a vector field on $N$ by giving its action on any function $f$ on $N$, in the obvious way:

$$(\phi^* V) f := V(\phi_* f) \tag{41}$$

In local coordinates this equation is

$$(\phi^* V)^\alpha \frac{\partial f}{\partial y^\alpha} = V^i \frac{\partial}{\partial x^i} f(y(x)) = V^i \frac{\partial y^\alpha}{\partial x^i} \frac{\partial f}{\partial y^\alpha}, \tag{42}$$

and as this holds for any $f$, we must have

$$(\phi^* V)^\alpha = V^i \frac{\partial y^\alpha}{\partial x^i}. \tag{43}$$

Next we can define the pull-back of a covector from $N$ to $M$ through

$$(\phi_* \omega) V := \omega(\phi^* V). \tag{44}$$

In local coordinates

$$(\phi_* \omega)_i V^i := \omega_\alpha \left( \frac{\partial y^\alpha}{\partial x^i} V^i \right), \tag{45}$$

and so

$$(\phi_* \omega)_i = \frac{\partial y^\alpha}{\partial x^i} \omega_\alpha. \tag{46}$$

Now if $\phi$ is actually one-to-one (which implies that $m = n$) then $\partial x^i / \partial y^\alpha$ also exists (it is the matrix inverse of $\partial y^\alpha / \partial x^i$) and then we can define both the push-forward and the pull-back of any rank tensor, for example

$$(\phi_* T)^\alpha{}_\beta = \frac{\partial y^\alpha}{\partial x^i} \frac{\partial x^j}{\partial y^\beta} T^i{}_j, \tag{47}$$

and similarly for other cases – it is always obvious what Jacobian matrices must be used.

## 3.3 1-parameter families of maps

Let $\phi$ be a one-to-one map from $M$ to itself which depends on a parameter $t$ and is at least once differentiable in $t$,

$$\begin{aligned} \phi_t : M \times \mathbf{R} &\to M \\ (p, t) &\mapsto \phi_t(p) \end{aligned} \tag{48}$$

or in local coordinates,

$$(x^i, t) \mapsto \tilde{x}^i(x, t) \tag{49}$$

and with $\phi_0 = \mathrm{Id}$.

Then $\phi_t(p)$ for fixed $p$ is a curve in $M$ and

$$V = \left.\frac{\partial}{\partial t}\phi_t(p)\right|_{t=0} \tag{50}$$

is its tangent vector at $p$. In coordinates, this is

$$V^i = \left.\frac{\partial \tilde{x}^i}{\partial t}\right|_{t=0} \tag{51}$$

If $f$ is a scalar field on $M$, then

$$\left.\frac{d}{dt} f(\phi_t(p))\right|_{t=0} = V^i \frac{\partial f}{\partial x^i} = Vf. \tag{52}$$

## 3.4 Lie derivative of a function

We can now define the Lie derivative of the function (scalar field) $f$ along the vector field $V$ as

$$\mathcal{L}_V f := \lim_{t\to 0} \frac{1}{t} \left((\phi_t)_* f - f\right). \tag{53}$$

In coordinates we have

$$\begin{aligned} \tilde{x}^i(x,t) &= \tilde{x}^i(x,0) + \frac{\partial \tilde{x}^i}{\partial t}\, t + o(t) \\ &= x^i + V^i t + o(t) \end{aligned} \tag{54}$$

where the first line is just the definition of $\phi_t$ being at least once differentiable in $t$. Substituting the definition of $(\phi_t)_* f$, we have

$$\begin{aligned} \mathcal{L}_V f &= \lim_{t\to 0} \frac{1}{t} \left(f(\tilde{x}(x,t) - f(x)\right) \\ &= \lim_{t\to 0} \frac{1}{t} \left[f(x^i + V^i t + o(t)) - f(x^i)\right] \\ &= \lim_{t\to 0} \frac{1}{t} \left[f(x^i) + \frac{\partial f}{\partial x^i} V^i t + o(t) - f(x^i)\right] \\ &= V^i \frac{\partial f}{\partial x^i} = Vf, \end{aligned} \tag{55}$$

as we would expect.

## 3.5 *Exercises*

1. Show that (34) and (35) do not transform as tensors.
2. Derive (47).
3. Consider the composition of two 1-parameter family of maps of $M$ to itself, and calculate in coordinates

$$V = \left.\frac{\partial}{\partial t}\psi_t\left(\phi_t(p)\right)\right|_{t=0}. \tag{56}$$

# 4 The Lie derivative

## 4.1 Lie derivative of a vector field

The basic problem with defining the derivative of a vector field is that we need to take the difference of that vector field at two different points, but that this difference is not defined on a manifold. However, we can pull back the vector at $q = \phi_t(p)$ to a vector at $p$ and so subtract two vectors at $p$. Actually, this pull-back is the push-forward with $\phi_t^{-1}$. This gives rise to the following definition:

$$\mathcal{L}_V W := \lim_{t\to 0} \frac{1}{t}\left[(\phi_t^{-1})^* W(\phi_t(p)) - W(p)\right]. \tag{57}$$

Now $\phi_t^{-1}$ can be expanded in coordinates as (compare (54))

$$x^i = \tilde{x}^i - V^i t + o(t) = \delta^i{}_j \tilde{x}^j - V^i(\tilde{x}(x,t)) + o(t) \tag{58}$$

and so the matrix representing $(\phi_t^{-1})^*$ is

$$\frac{\partial x^i}{\partial \tilde{x}^j} = \delta^i{}_j - \frac{\partial V^i}{\partial \tilde{x}^j} t + o(t) = \delta^i{}_j - \frac{\partial V^i}{\partial x^j} t + o(t). \tag{59}$$

Substituting this last result, and the Taylor expansion of $W(\phi_t(p))$,

$$\begin{aligned}
(\mathcal{L}_V W)^i &= \lim_{t\to 0} \frac{1}{t}\left[\left(\delta^i{}_j - \frac{\partial V^i}{\partial x^j} t + o(t)\right)\left(W^j + \frac{\partial W^j}{\partial x^k} V^k t + o(t)\right) - W^i\right] \\
&= \lim_{t\to 0} \frac{1}{t}\left[W^i + \frac{\partial W^i}{\partial x^k} V^k t - \frac{\partial V^i}{\partial x^j} W^j t + o(t) - W^i\right] \\
&= V^j \frac{\partial W^i}{\partial x^j} - W^j \frac{\partial V^i}{\partial x^j}.
\end{aligned} \tag{60}$$

(Note we have relabelled a dummy index). The first term is the directional derivative (34) one would expect. The second term is unexpected, but only both terms together transform as a vector. The Lie derivative of a vector field is also written as the **commutator**, or **Lie bracket**

$$\mathcal{L}_V W = [V, W] \tag{61}$$

The Lie bracket of two vector fields is obviously antisymmetric, and it is easy to check that it obeys the **Jacobi identity**

$$[[A, B], C] + [[B, C], A] + [[C, A], B] = 0. \tag{62}$$

## 4.2 Lie derivative of a covector field

We define

$$\mathcal{L}_V \omega := \lim_{t\to 0} \frac{1}{t}\left[(\phi_t)_* \omega(\phi_t(p)) - \omega(p)\right]. \tag{63}$$

To work out what this is in coordinates, we can use the expressions we already have for the Lie derivative of a scalar and a vector field, and the product rule, which must apply for any derivative operator. From now on, we shall use the

comma to denote partial derivatives *in a given coordinate system* $x^i$, so that $f_{,i}$ will mean $\partial f/\partial x^i$, for any object $f$. We have

$$\begin{aligned}\mathcal{L}_V(\omega\cdot W) &= V^i\left(\omega_j W^j\right)_{,i} = V^i\omega_{j,i}W^j + V^i\omega_j W^j{}_{,i}\\ &= \omega\cdot(\mathcal{L}_V W) + (\mathcal{L}_V\omega)\cdot W\\ &= \omega_j\left(V^i W^j{}_{,i} - W^i V^j{}_{,i}\right) + (\mathcal{L}_V\omega)_i W^i. \end{aligned} \tag{64}$$

As this must hold for any $W^i$, we must have

$$(\mathcal{L}_V\omega)_i = V^j\omega_{i,j} + V^j{}_{,i}\omega_j. \tag{65}$$

Again, the first term is expected, but only both terms together transform as a covector.

## 4.3 Lie derivative of a tensor field of arbitrary rank

The product rule argument can be used on tensors of arbitrary rank, by contracting each slot with a vector or covector. We find

$$(\mathcal{L}_V T)^{i\ldots}{}_{j\ldots} = V^k T^{i\ldots}{}_{j\ldots,k} - V^i{}_{,k}T^{k\ldots}{}_{j\ldots} - \ldots + V^k{}_{,j}T^{i\ldots}{}_{k\ldots} + \ldots \tag{66}$$

where the first term is the expected one, and the dots indicate that one correction term is needed for each vector and covector index. As $\mathcal{L}_V T$ is always a tensor of the same type as $T$, we will no longer write the brackets on the left-hand side of (66).

Note that for the definition of the Lie derivative, $V$ needs to be a $C^1$ vector field, not just a vector at one point.

Note that if $V = \partial/\partial x^1$ in one particular coordinate system, then $V^i = \delta^i{}_1$ and so $\partial V^i/\partial x^j = 0$ in that coordinate system, and so, *in these coordinates only*,

$$\mathcal{L}_V T^{i\ldots}{}_{j\ldots} = \frac{\partial}{\partial x^1}T^{i\ldots}{}_{j\ldots} \tag{67}$$

This is often useful for calculations, as for a single given vector field $V$ we can locally (but not always globally) find coordinates $x^i$ such that $V = \partial/\partial x^1$.

## 4.4 Application: axisymmetry in fluid dynamics

Lie derivatives are closely related to symmetries. As an example, consider a fluid flow that is restricted to be axisymmetric. In spherical polar coordinates adapted to the symmetry, this just means that the density $\rho$ and velocity $v^\mu$ are independent of the angle $\varphi$, or $\rho = \rho(r,\theta)$ and $v^\mu = v^\mu(r,\theta)$ for $\mu = r,\theta,\varphi$. The vector

$$X = \frac{\partial}{\partial\varphi} \tag{68}$$

is said to be the **generator of a symmetry**, in the sense that

$$\mathcal{L}_X\rho = \frac{\partial\rho}{\partial\varphi} = 0, \tag{69}$$

$$\mathcal{L}_X v^\mu = \frac{\partial v^\mu}{\partial\varphi} = 0, \qquad \mu = r,\theta,\varphi \tag{70}$$

and $\rho$ and $v$ are said to be **invariant** under (the symmetry generated by) $X$.

As these equations are tensor equations, we can express them in arbitrary coordinates. From the chain rule

$$X^i = \frac{\partial x^i}{\partial \varphi} \tag{71}$$

and from our formulae above

$$\begin{aligned} \mathcal{L}_X \rho &= X^i \rho_{,i} = 0, & (72) \\ \mathcal{L}_X v^i &= X^j v^i{}_{,j} - v^j X^i{}_{,j} = 0 & (73) \end{aligned}$$

In Cartesian coordinates $(x, y, z) = (r\cos\theta\cos\varphi, r\cos\theta\sin\varphi, r\sin\theta)$, we have

$$X = -y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y} \tag{74}$$

and hence

$$\begin{aligned} (\mathcal{L}_X v)^x &= -y v^x{}_{,x} + x v^x{}_{,y} + v^y = \frac{\partial}{\partial \varphi} v^x + v^y = 0, & (75) \\ (\mathcal{L}_X v)^y &= -y v^y{}_{,x} + x v^y{}_{,y} - v^x = \frac{\partial}{\partial \varphi} v^y - v^x = 0, & (76) \\ (\mathcal{L}_X v)^z &= -y v^z{}_{,x} + x v^x{}_{,z} = \frac{\partial}{\partial \varphi} v^z = 0. & (77) \end{aligned}$$

The partial derivative terms are expected, but the lower order terms are not. They can be thought of as the rotation of the basis vectors in the $x$ and $y$ direction as we rotate around the $z$ axis.

## 4.5 *Exercises*

1. (Schutz 3.1a) Show that, acting on scalars and vector fields, the commutator of two Lie derivatives is the Lie derivative along the commutator of vector fields, or
$$[\mathcal{L}_V, \mathcal{L}_W] = \mathcal{L}_{[V,W]}. \tag{78}$$

2. (Schutz 3.2b) Assume that in a coordinate basis the Lie derivative is given by (65). Now consider an arbitrary basis (not a coordinate basis) $e_\alpha$ of vector fields, such that $U = U^\alpha e_\alpha$, where summation over the basis index $\alpha$ is implied, but $\alpha$ is not a tensor index. In fact, in abstract index notation we could write $U^a = U^\alpha e^a_\alpha$ which makes clear that $U^\alpha$ is a scalar and $e^a_\alpha$ is a vector. However, it is less messy to suppress the abstract indices, and instead we shall write $U^\alpha$ and interpret this as a derivative operator as in Definition 3 of a vector field, see (14). Show that in an arbitrary basis
$$(\mathcal{L}_V U)^\alpha = V^\beta e_\beta U^\alpha - U^\beta e_\beta V^\alpha + V^\beta U^\gamma \left(\mathcal{L}_{e_\beta} e_\gamma\right)^\alpha \tag{79}$$
Hint: use the product rule on $U = U^\alpha e_\alpha$, where summation over the basis index $\alpha$ is implied, $U^\alpha$ is a scalar and $e_\beta$ is a vector field.

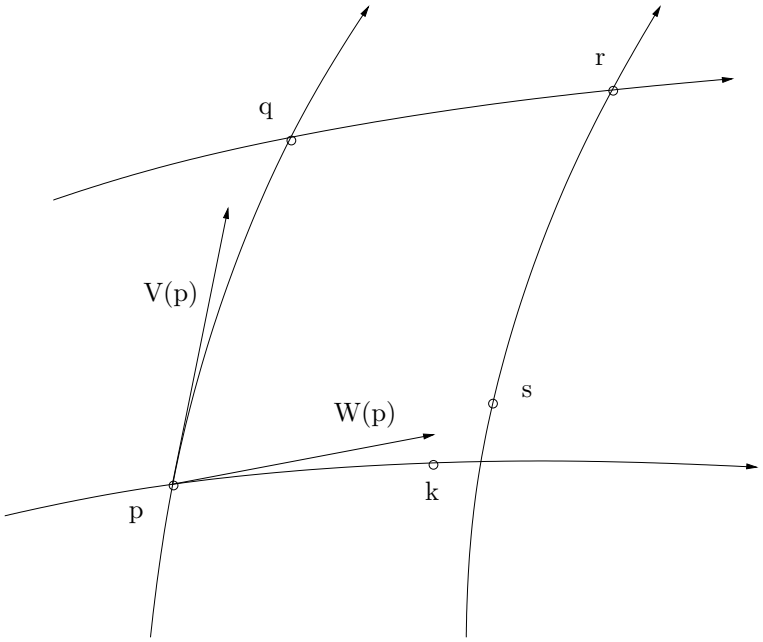


Figure 1: Graphical illustration of $\mathcal{L}_V W$. The left curve is $\phi_t(p)$ with $\phi_0(p) = p$ and $d\phi_t(p)/dt|_{t=0} = V(p)$. The bottom curve is $\psi_t(p)$, with $\psi_0(p) = p$ and and $d\psi_t(p)/dt|_{t=0} = W(p)$. Now consider some finite $t > 0$. Then $q = \phi_t(p)$ for this $t$ and $r = \psi_t(q) = \psi_t(\phi_t(p))$, and finally $s = (\phi_t)^{-1}(q) = (\phi_t)^{-1}(\psi_t(\phi_t(p)))$. Going the other way around the diagram, $k = \psi_t(p)$. The difference between $k$ and $s$ is proportional to $t^2$ and represents $\mathcal{L}_V W$. From this construction, $\mathcal{L}_V W = -\mathcal{L}_W V$ is obvious.

3. (Schutz 3.5) a) Show that if $V$ and $W$ are linear combinations (not necessarily with constant coefficients) of $m$ vector fields that commute with each other, then $[V, W]$ is a linear combination of the same $m$ vector fields (use the previous question).

   b) Prove the same result when the $m$ vector fields have Lie brackets which are nonvanishing linear combinations of the $m$ vector fields.

4. (Schutz 3.7) The vector fields $l_x$, $l_y$ and $l_z$ generate rotations about these $x$ $y$ and $z$ axes and obey $[l_x, l_y] = l_z$ and cyclic. Define $L^2 := \mathcal{L}_{l_x}^2 + \mathcal{L}_{l_y}^2 + \mathcal{L}_{l_z}^2$.

   a) Show that $\mathcal{L}_{l_z}$ and $L^2$ commute.

   b) Show for a scalar $f$ that

$$L^2 f = \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial f}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2 f}{\partial\varphi^2} \tag{80}$$

5. (Schutz 3.8) Show that if $T$ is invariant under $V$ and $W$, it is invariant under $aV + bW$ where $a$ and $b$ are constants.

# 5 The covariant derivative

## 5.1 Definition

The Lie derivative gives us a notion of the derivative of a tensor along a vector field. However, $\mathcal{L}_X T$ contains derivatives of $X$ as well as of $T$. In other words, it depends on the vector *field* $X$ and not just on a vector $X$ at one point. The covariant derivative does not have this disadvantage, but we will see that it requires an additional (non-unique) structure on the manifold called a connection.

We define the **covariant derivative** $\nabla_X Y$ of a vector field $Y$ along the vector $X$ with the following properties:

1. $\nabla_X$ is a derivative operator when acting on $Y$ (linearity in $Y$, product rule);

2. $\nabla_X Y$ is also linear in $X$ (no derivatives of $X^b$), meaning that there exists a $(1,1)$ tensor $\nabla_b Y^a$ such that

$$(\nabla_X Y)^a = X^b \nabla_b Y^a. \tag{81}$$

Note $\nabla_b Y^a$ is a single $(1,1)$-tensor, not a product! It is sometimes also written as $Y^a{}_{;b}$ (using a semicolon instead of a comma).

Now $Y^i{}_{,j}$ does not transform as a tensor. But we can demand that

$$\nabla_i Y^j := Y^j{}_{,i} + \Gamma^j{}_{ik} Y^k \tag{82}$$

does, by giving the **Christoffel symbols** $\Gamma^i{}_{jk}$ the correct transformation rule.

$$\begin{aligned}
\nabla_i Y^j &= Y^j{}_{,i} + \Gamma^j{}_{ik} Y^k \\
&= \frac{\partial \tilde{x}^p}{\partial x^i}\frac{\partial x^j}{\partial \tilde{x}^q}\left(\tilde{Y}^q{}_{,p} + \tilde{\Gamma}^q{}_{pr}\tilde{Y}^r\right) \\
&= \frac{\partial \tilde{x}^p}{\partial x^i}\frac{\partial x^j}{\partial \tilde{x}^q}\left[\left(\frac{\partial \tilde{x}^q}{\partial x^l}Y^l\right)_{,m}\frac{\partial x^m}{\partial \tilde{x}^p} + \tilde{\Gamma}^q{}_{pr}\tilde{Y}^r\right] \\
&= \frac{\partial \tilde{x}^p}{\partial x^i}\frac{\partial x^j}{\partial \tilde{x}^q}\left[\left(\frac{\partial \tilde{x}^q}{\partial x^l}Y^l{}_{,m} + \frac{\partial^2 \tilde{x}^q}{\partial x^l \partial l^m}Y^l\right)\frac{\partial x^m}{\partial \tilde{x}^p} + \tilde{\Gamma}^q{}_{pr}\frac{\partial \tilde{x}^r}{\partial x^k}Y^k\right] \\
&= Y^j{}_{,i} + \left(\frac{\partial x^j}{\partial \tilde{x}^q}\frac{\partial \tilde{x}^r}{\partial x^k}\frac{\partial \tilde{x}^p}{\partial x^i}\tilde{\Gamma}^q{}_{pr} + \frac{\partial x^j}{\partial \tilde{x}^q}\frac{\partial^2 \tilde{x}^q}{\partial x^k \partial x^i}\right)Y^k
\end{aligned} \tag{83}$$

and so we must have

$$\Gamma^j{}_{ik} = \frac{\partial x^j}{\partial \tilde{x}^q}\frac{\partial \tilde{x}^r}{\partial x^k}\frac{\partial \tilde{x}^p}{\partial x^i}\tilde{\Gamma}^q{}_{pr} + \frac{\partial x^j}{\partial \tilde{x}^q}\frac{\partial^2 \tilde{x}^q}{\partial x^k \partial x^i} \tag{84}$$

Note the first term looks as if the Christoffel symbol was a $(1,2)$-tensor, but the second term is inhomogeneous!

We now *define*

$$\nabla_X f := X^i f_{,i} \tag{85}$$

or equivalently

$$\nabla_a f := f_{,a} \tag{86}$$

(recall that the partial derivative of a scalar does transform like a covector) and we define $\nabla$ to obey the product rule

$$\nabla_X (TS) = (\nabla_X T)S + T(\nabla_X S), \tag{87}$$

where $TS$ stands for any two tensors, perhaps contracted (neither free indices nor dummy indices are shown). From these two rules, applied to $f = T^i S_i$ one finds

$$\nabla_i \omega_j = \omega_{j,i} - \Gamma^k{}_{ij} \omega_k, \tag{88}$$

and hence the general rule

$$\nabla_i T^{j\ldots}{}_{k\ldots} = T^{j\ldots}{}_{k\ldots,i} + \Gamma^j{}_{il} T^{l\ldots}{}_{k\ldots} + \ldots - \Gamma^l{}_{ik} T^{j\ldots}{}_{l\ldots} - \ldots \tag{89}$$

The dots represent one Christoffel term with a plus sign for each extra up index on $T$, and one Christoffel term with a minus sign for each extra down index on $T$.

We have generalised the partial derivative from scalars to arbitrary tensors in a geometric way, but at the price of introducing an additional geometric structure on the manifold, the **affine connection**. Although the connection is not a tensor, it is a geometric object which is specified by giving its $n^3$ components $\Gamma^i{}_{jk}$ in one coordinate system; they are given in any other coordinate system by the transformation rule (84).

Contrast $\nabla_X$ with the Lie derivative $\mathcal{L}_X$: the Lie derivative also defines a derivative of any tensor field along a vector field $X$, and without introducing a new structure, but at the price that it is not linear in $X$, i.e. the Lie derivative along $X$ contains derivatives of $X$ itself.

## 5.2 Curvature

Unlike partial derivatives, two covariant derivatives acting on a tensor do not in general commute. However, one finds that the commutator of two covariant derivatives is not a derivative operator as one would expect but a tensor, the **curvature tensor**. It is defined by

$$(\nabla_a \nabla_b - \nabla_b \nabla_a) \omega_c = R_{abc}{}^d \omega_d. \tag{90}$$

Note no derivatives of $\omega_d$ appear on the right-hand side, and as the left-hand side is a tensor, so is $R_{abc}{}^d$. Its components in a coordinate system can be calculated from this definition, in terms of the connection coefficients and their first partial derivatives. (See exercise).

For any connection, the identity

$$R_{[abc]}{}^d = 0, \tag{91}$$

holds for its Riemann tensor, and also

$$\nabla_{[e} R_{ab]c}{}^d = 0, \tag{92}$$

the **Bianchi identity**. (The proof is by straightforward calculation but long.) Here, the square brackets around a group of tensor indices denote that the tensor has been **antisymmetrised** in these indices, meaning that

$$X_{[ab]} := \frac{1}{2!} \left( X_{ab} - X_{ba} \right), \tag{93}$$

$$X_{[abc]} := \frac{1}{3!} \Big( X_{abc} + X_{bca} + X_{cab} - X_{bac} - X_{acb} - X_{cba} \Big), \tag{94}$$

and so on. Generally $p!$ terms are needed to antisymmetrise over $p$ indices. Similarly, round brackets around a group of tensor indices denote that these indices have been **symmetrised** – same formula, but with plus signs only.

A connection (or covariant derivative) with ${R_{abc}}^d = 0$ is called **flat**. For example, if $\nabla_\mu = \partial/\partial x^\mu$ in some coordinate system, then the curvature tensor vanishes in that coordinate system because the Christoffels vanish in that coordinate system, and so this connection is flat.

## 5.3 The metric-covariant derivative

If the manifold $M$ is equipped with a metric, there is a preferred covariant derivative, the **metric-covariant derivative** (sometimes just called "the" covariant derivative), which is defined by the properties

$$ {\Gamma^i}_{jk} = {\Gamma^i}_{kj} \tag{95} $$

(the connection is **symmetric**, or **torsion-free**, see exercise), and

$$ \nabla_a g_{bc} = 0 \tag{96} $$

It can be shown (see exercise) that its connection coefficients are then

$$ {\Gamma^i}_{jk} = \frac{1}{2} g^{il} \left( -g_{jk,l} + g_{kl,j} + g_{jl,k} \right). \tag{97} $$

A prime example is the physical space $\mathbf{R}^3$ of Newtonian physics. We have seen that in Cartesian coordinates the metric $\gamma_{ab}$ of Newtonian space is $\gamma_{ij} = \delta_{ij}$, and (97) then gives us ${\Gamma^i}_{jk} = 0$ in Cartesian coordinates. This means that in Cartesian coordinates the covariant derivative is simply the partial derivative, or $\nabla_i = \partial/\partial x^i$. As $\delta_{ij}$ is constant, $\nabla_i \gamma_{jk} = 0$ is clearly obeyed in Cartesian coordinates. It is however true in all coordinates, even if the ${\Gamma^i}_{jk}$ do not vanish (see exercise).

The curvature tensor of a metric connection is also called the **Riemann tensor**, and it contains information about the geometry, or shape, given to the manifold by that metric. A metric with vanishing Riemann tensor is again called **flat**, for example the Euclidean metric $\gamma_{ab}$ on $\mathbf{R}^n$ is flat. The Riemann tensor associated with a metric-covariant derivative, with indices suitably lowered with $g_{ab}$, obeys the identity

$$ R_{abcd} = R_{cdab}. \tag{98} $$

## 5.4 *Exercises*

1. Derive (89), for the special cases of a $(1,1)$-tensor (or for the general case if you are good with notation).

2. Show that in (66) the partial derivative can be replaced by any covariant derivative, because the connection coefficients cancel out.

3. Let $\tilde{x}^i$ represent the Cartesian coordinates $(x, y, z)$, and let $x^i$ represent the spherical polar coordinates $(r, \theta, \varphi)$. Assume that in Cartesian coordinates $\tilde{\gamma}_{ij} = \delta_{ij}$. Use (23) to calculate $\gamma_{ij}$ and hence $\gamma^{ij}$ in spherical polar coordinates. Calculate the ${\Gamma^i}_{jk}$ in spherical polar coordinates using (97).

4. Assume that in Cartesian coordinates $\tilde{\Gamma}^i_{jk} = 0$, and use (84) to calculate the ${\Gamma^i}_{jk}$ in spherical polar coordinates. You should get the same answer as before.

5. Use $g^{ij}$ and ${\Gamma^i}_{jk}$ from the previous questions to write down the three components of Euler's equation

$$\dot{v}^i + v^j \nabla_j v^i = -\frac{1}{\rho}\nabla^i p \tag{99}$$

in spherical polar coordinates. (The time derivative $\partial/\partial t$ denoted by an overdot does not change under a change of spatial coordinates.)

6. Assume that ${\Gamma^i}_{jk} = 0$ is symmetric in $j$ and $k$ (no torsion). Define $\Gamma_{ijk} := g_{il}{\Gamma^l}_{jk}$ as a shorthand notation. Write out $\nabla_i g_{jk} = 0$ in full using the connection coefficients $\Gamma$. By permutating indices, also write down $\nabla_j g_{ik} = 0$ and $\nabla_k g_{ij} = 0$. Add these three equations (perhaps with a minus sign) to solve for $\Gamma_{ijk}$. Raise an index to obtain ${\Gamma^i}_{jk}$.

7. Calculate, in coordinates, the components of the **torsion**

$$\nabla_U V - \nabla_V U - [U, V]. \tag{100}$$

Your answer should be proportional to $U$ and $V$ (undifferentiated) and an object for you to define, the **torsion tensor**, which is made from the connection coefficients.

8. Show that the difference between two Christoffel symbols transforms as a tensor. Hence the difference of two connections is a tensor.

9. From the definition (90), by expanding $\nabla_i \nabla_j \omega_k$, substituting (88) and then (89), assuming no torsion, show that the components of the Riemann tensor are given by

$${R_{ijk}}^l = {\Gamma^l}_{ik,j} + {\Gamma^m}_{ik}{\Gamma^l}_{mj} - (i \leftrightarrow j). \tag{101}$$

# 6 Differential forms

## 6.1 Differentiation

A differential form is simply a totally antisymmetric tensor field with indices down. A **$p$-form** has $p$ such indices, and in this section we use the notation

$$\omega_{a_1\ldots a_p} = \omega_{[a_1\ldots a_p]}. \tag{102}$$

Obviously, on an $n$-dimensional manifold $p$-forms can exist only for $0 \le p \le n$. (A 0-form is a scalar field, and a 1-form is a covector field.)

Next we define the **exterior product** or **wedge product**, which maps a $p$-form and a $q$-form into a $(p+q)$-form, as

$$(\omega \wedge \mu)_{a_1\ldots a_p b_1\ldots b_q} := \frac{(p+q)!}{p!q!}\omega_{[a_1\ldots a_p}\mu_{b_1\ldots b_q]}. \tag{103}$$

This implies $\omega \wedge \mu = (-1)^{pq}\mu \wedge \omega$. The factorials on the right-hand side of this definition make the exterior product associative, $(\alpha\wedge\beta)\wedge\gamma = \alpha\wedge(\beta\wedge\gamma)$, for any number of indices. Because a differential form has a unique index structure, it is customary to suppress all tensor indices in any tensor equations that contains only differential forms.

The **exterior derivative** $d$ is a derivative operator that maps a $p$-form into a $(p+1)$-form, defined by

$$(d\omega)_{ba_1\ldots a_p} := (p+1)\nabla_{[b}\omega_{a_1\ldots a_p]}. \tag{104}$$

The factor $(p+1)$ on the right hand side makes the product rule come out right for the exterior derivative of an exterior product, that is

$$d(\alpha \wedge \beta) = d\alpha \wedge \beta + \alpha \wedge d\beta. \tag{105}$$

The whole point of introducing differential forms is that this definition of the exterior derivative is actually **independent of** $\nabla$, that is, all connection coefficients cancel out of the right-hand side (exercise). In particular, we can use the partial derivative $\partial$ for $\nabla$ on the right-hand side, whatever coordinate system we use.

Finally, we note that

$$d(d\omega) = 0 \tag{106}$$

for any $p$-form $\omega$. This follows from the fact that partial derivatives commute (exercise). If there is an $n$-form $\phi$ such that $\omega = d\phi$, then the $(n+1)$-form $\omega$ is called **exact**. Any form $\omega$ that obeys $d\omega = 0$ is called **closed**. Every exact form is closed, and closed forms are exact in any sufficiently small neighbourhood of any given point on the manifold but not necessarily globally. (Look up **cohomology** in any textbook to find out more.).

## 6.2 Integration

Differential forms allow us to put integration on manifolds and submanifolds on an elegant footing. What you know about integration on lines and surfaces in $\mathbf{R}^3$, and Gauss' law and Stoke's law will be recovered as special cases, although this is left to the exercises.

Let $N$ be an $n$-dimensional manifold and let the $n$-form $\alpha$ on $N$ be defined as

$$\alpha = a\, dx^1 \wedge dx^2 \wedge dx^3 \wedge \ldots \wedge dx^n \tag{107}$$

where $x^i$, $i = 1, \ldots n$ is a local coordinate system. Equivalently we can write

$$\alpha_{\mu_1 \ldots \mu_n} = a[\mu_1 \ldots \mu_n] \tag{108}$$

where we use the shorthand notation

$$[\mu_1 \ldots \mu_n] := n!\, \delta^1{}_{[a_1} \delta^2{}_{a_2} \ldots \delta^n{}_{a_n]}, \tag{109}$$

so that $[123] = 1$, $[213] = -1$, etc. Any $n$-form on an $n$-dimensional manifold has only one algebraically independent component. (Here we have called it $a$. Note that while $\alpha_{a_1 \ldots a_n}$, or $\alpha$ for short, is a tensor, $[\mu_1 \ldots \mu_n]$ is not a tensor, and $a$ is not a scalar.)

We now define the **integral of $\alpha$ over $N$** as

$$\int_N \alpha := \int_{U=\phi(N)} a\, dx^1\, dx^2 \ldots dx^n \tag{110}$$

where $\phi : N \to \mathbf{R}^n$ is a chart and the integral on the right-hand side is just a multiple integral over a subset $U$ of $\mathbf{R}^n$ – I assume you already know how to evaluate multiple integrals over subsets of $R^n$. (For (110) to hold, the manifold $N$ must be *orientable*, a complication we need not worry about here.) This formula explains in hindsight why the notation $dx^i$ is used for the coordinate basis of 1-forms (covectors).

If we now change from the coordinates $x^i$ to a new set of coordinates $\tilde{x}^i$, the standard tensor transformation law applied to the tensor $\alpha$ gives

$$\int_N \alpha := \int_{\tilde{U}=\tilde{\phi}(N)} \tilde{a}\, d\tilde{x}^1\, d\tilde{x}^2 \ldots d\tilde{x}^n, \tag{111}$$

where

$$\tilde{a} = \left|\frac{\partial x}{\partial \tilde{x}}\right| a := \det\left(\frac{\partial x^i}{\partial \tilde{x}^j}\right) a. \tag{112}$$

So applying the usual tensor transformation law to a totally antisymmetric tensor of maximal rank gives rise to the Jacobian determinant $|\partial x/\partial \tilde{x}|$ which you may already have seen in multiple integrals (see exercise).

To integrate over a line or a surface, we define the concept of an $n$-dimensional **submanifold** $N$ of an $m$-dimensional manifold $M$, with $0 < n < m$ as follows: a neighbourhood of $N$ in $M$ is covered by coordinate charts $x^i$, $i = 1, \ldots m$, such that $N$ is given by $x^{n+1} = x^{n+2} = \ldots = x^m = 0$ and $x^i$ for $i = 1, \ldots n$ serve as coordinates on $N$. Then we can again use (110), except that now $N$ is a submanifold of some higher-dimensional manifold $M$.

This is precisely what one does in practice. For example, to integrate over a sphere, one introduces spherical polar coordinates such that the sphere is given by $r = R$, and the integration is over $\theta$ and $\varphi$.

## 6.3 Stokes' theorem

We can now state the wonderful **Stokes' theorem** in its differential forms version:

$$\int_M d\alpha = \int_{\partial M} \alpha \tag{113}$$

Here $M$ is an $m$-dimensional manifold (it does not matter if it is itself a submanifold of something bigger), $\partial M$ is its boundary, which means it is an $m-1$ dimensional manifold, and $\alpha$ is an $(m-1)$-form. ($\partial M$ can be empty, for example if $M$ is a sphere, in which case the integral is zero.)

Note that a consequence of applying Stokes' law twice is

$$0 \equiv \int_M dd\alpha = \int_{\partial M} d\alpha = \int_{\partial\partial M} \alpha \equiv 0 \tag{114}$$

where the first equality holds because of the identity $dd\alpha \equiv 0$ (every exact differential form is closed) and the last equality holds because of the identity $\partial\partial M \equiv \emptyset$ (the boundary of a boundary is the empty set). We find that the integral of an exact form over a boundary vanishes.

## 6.4 The volume form

You are familiar with the integration of scalar fields (functions) over regions of $\mathbf{R}^n$, or perhaps over hypersurfaces. How does this relate to the integration of an $n$-form over an $n$-dimensional manifold?

To integrate a scalar field over an $n$-dimensional manifold, one needs to multiply it with a **volume form**, which is just an $n$-form. In Newtonian 3-dimensional space $\mathbf{R}^3$ in Cartesian coordinates $x^i$, the natural volume form is given by $\epsilon = dx^1 \wedge dx^2 \wedge dx^3$, or equivalently $\epsilon_{123} = -\epsilon_{213} = \ldots = 1$, or $a = 1$ in (107). Then, from (107) and (110), $\int f\epsilon$ is just the usual $\int f\, dx^1\, dx^2\, dx^3$. Then $\tilde{a} = a|\partial x/\partial\tilde{x}|$ tells us how to evaluate the integral in arbitrary coordinates.

If a metric $g$ is already defined on the manifold, then we also have a metric-covariant derivative $\nabla$, and it turns out that the requirement

$$\nabla_b \epsilon_{a_1\ldots a_n} = 0 \tag{115}$$

uniquely defines a volume form $\epsilon$ up to an overall constant factor. To fix that remaining factor, we also impose the normalisation

$$\epsilon_{a_1\ldots a_n}\epsilon^{a_1\ldots a_n} = \pm n! \tag{116}$$

where the indices have been raised with the metric. (It follows from (115) and $\nabla_a g^{bc} = 0$ that the right-hand side of (116) is a constant in the first place.)

The sign $\pm$ in (116) is $+1$ if $g_{\mu\nu}$ has an even number of negative eigenvalues in any coordinate system (for example zero, for a positive definite metric), and $-1$ if the number of eigenvalues is odd (for example, 1 out of 4 in the metric used in relativity, see later). This is most easily seen by going to coordinates where $g_{\mu\nu}$, and therefore also its inverse $g^{\mu\nu}$, is diagonal with either $+1$ or $-1$ in each place on the diagonal. But (116) is a tensor equation, and so holds in every coordinate system. (The list of the signs of the eigenvalues, for example $(-+++)$ in spacetime, is called the **signature of the metric**.)

One can show that in coordinates

$$\epsilon = \sqrt{|g|}\; dx^1 \wedge dx^2 \wedge \ldots \wedge dx^n, \tag{117}$$

where $|g|$ is the determinant of the matrix of metric coefficients $g_{ij}$. This is consistent, as under a change of coordinates from $x^i$ to $\tilde{x}^i$,

$$\tilde{g}_{pq} = \frac{\partial x^i}{\partial \tilde{x}^p}\frac{\partial x^j}{\partial \tilde{x}^q} g_{ij} \tag{118}$$

and hence, from $|gJJ| = |g||J|^2$,

$$|\tilde{g}| = |g| \left|\frac{\partial x}{\partial \tilde{x}}\right|^2. \tag{119}$$

(A warning on notation: we reserve the symbol $\epsilon$ for the volume form, which is a tensor. Elsewhere, $\epsilon_{\mu_1\ldots\mu_n}$ is sometimes used for the symbol $[\mu_1 \ldots \mu_n]$. The two are the same only in Euclidean space in Cartesian coordinates.)

The volume form defines the **Hodge dual**, which maps a $p$-form $\omega$ to an $n-p$ form $\star\omega$ as

$$\star\omega_{a_1\ldots a_{n-p}} := \frac{1}{p!}\epsilon_{a_1\ldots a_{n-p}}{}^{b_1\ldots b_p}\omega_{b_1\ldots b_p} \tag{120}$$

The factorial is there to make $\star(\star\omega) = \pm\omega$, with the inevitable sign as above in (116).

## 6.5 Gauss' theorem

An important special case of Stokes' theorem is **Gauss' theorem**. In Stokes' theorem, let $X$ be a vector field on an $n$-dimensional manifold $M$, and let

$$\alpha = (-1)^{n-1} \star X, \tag{121}$$

or in tensor notation,

$$\alpha_{a_1\ldots a_{n-1}} = (-1)^{n-1}\epsilon_{a_1\ldots a_{n-1}b}X^b = X^b\epsilon_{ba_1\ldots a_{n-1}}. \tag{122}$$

From this we find (exercise) that

$$(d\alpha)_{ca_1\ldots a_{n-1}} = n\nabla_{[c}\left(\epsilon_{a_1\ldots a_{n-1}]b}X^b\right) \tag{123}$$

As $d\alpha$ is an $n$-form on an $n$-dimensional manifold, it must be proportional to the volume form. In fact one can show (exercise) that

$$d\alpha = (\nabla_b X^b)\epsilon. \tag{124}$$

Given an $(n-1)$-dimensional hypersurface $S$ with unit normal vector $n^a$, define an $(n-1)$ form $\tilde{\epsilon}$ by

$$\tilde{\epsilon}_{a_1\ldots a_{n-1}} := \epsilon_{a_1\ldots a_{n-1}b}n^b \tag{125}$$

Defining $h_{ab} := g_{ab} - n_a n_b$, as before, it is easy to show (exercise) that

$$h_{a_1}{}^b\tilde{\epsilon}_{ba_2\ldots a_{n-1}} = \tilde{\epsilon}_{a_1\ldots a_{n-1}}, \tag{126}$$

so that $\tilde{\epsilon}$ is restricted to the hypersurface. One can also show (exercise) that

$$\tilde{\epsilon}^{a_1 \ldots a_{n-1}} \tilde{\epsilon}_{a_1 \ldots a_{n-1}} = \pm (n-1)! \tag{127}$$

so $\tilde{\epsilon}$ is normalised correctly to be the volume form on $N$. From this it follows that in coordinates in which $N$ is given by $x^n = 0$,

$$\tilde{\epsilon} = \sqrt{|h|}\; dx^1 \wedge dx^2 \wedge \ldots \wedge dx^{n-1}. \tag{128}$$

where $|h|$ is the determinant of $h_{ij}$ in these coordinates, deleting the $n$th row and column, which are all zero.

Finally, one can show (exercise) that

$$h_{a_1}{}^{c_1} \ldots h_{a_{n-1}}{}^{c_{n-1}} \epsilon_{c_1 \ldots c_{n-1} b} X^b = \tilde{\epsilon}_{a_1 \ldots a_{n-1}}\, n_b X^b \tag{129}$$

which means that when restricted to $S$,

$$\alpha|_S = \tilde{\epsilon} n_b X^b. \tag{130}$$

If now $S$ is chosen to be $\partial M$, we have Gauss' theorem

$$\int_M \nabla_b X^b \epsilon = \int_{\partial M} n_b X^b \tilde{\epsilon}. \tag{131}$$

In the notation of vector calculus, which may be familiar to you, this is

$$\int_M \vec{\nabla} \cdot \vec{X}\, dV = \int_{\partial M} \vec{X} \cdot d\vec{S}. \tag{132}$$

## 6.6 *Exercises*

1. Derive (117). You will need to use the fact that the determinant of a matrix is a totally antilinear map on that matrix, namely for a matrix $M_{ij}$

   $$\det M = [\mu_1 \ldots \mu_n] M_{\mu_1 1} \ldots M_{\mu_n n} \tag{133}$$

   and hence

   $$\det M = \frac{1}{n!} [\mu_1 \ldots \mu_n][\nu_1 \ldots \nu_n] M_{\mu_1 \nu_1} \ldots M_{\mu_n \nu_n} \tag{134}$$

2. Show that $d\omega$ is independent of $\nabla$ in (104).

3. Show that $dd\omega = 0$.

4. Fill in the missing details in the derivation of Gauss' law: Derive a) (123), b) (124), c) (126), d) (127), e) (129).

5. Derive the 3-dimensional version Stokes' law

   $$\int_M (\vec{\nabla} \times \vec{X}) \cdot d\vec{S} = \int_{\partial M} \vec{X} \cdot d\vec{s}, \tag{135}$$

   where $M$ is a 2-dimensional surface with boundary in 3-dimensional Euclidean space, so that its boundary $\partial M$ is a closed curve, from the general Stokes' law (113). The **curl** is geometrically defined as

   $$\left(\vec{\nabla} \times \vec{X}\right)^a := \epsilon^{ab}{}_c \nabla_b X^c. \tag{136}$$

   Hint: in this case $\alpha$ is the 1-form $\alpha_a = g_{ab} X^b$. Explain why this is the forms equivalent of $\vec{X} \cdot d\vec{s}$, then calculate $d\alpha$ and interpret.

6. Prove the identity (91).

# Part II
# Special relativity and Electromagnetism

## 7 Kinematics of special relativity

### 7.1 Observers in motion

Einstein's moving train is a good example of two observers in relative motion at constant speed. We shall draw a **spacetime diagram** containing both space and time axes. For now it is sufficient to consider movement in only one dimension. A line in the spacetime diagram that shows the position of an object at all times is called the **worldline** of that object. Observer $S$ defines the railway track to be fixed, and uses the sleepers on the track to label points $x$ in space. In other words, each world line of a sleeper is a line of constant coordinate $x$. Observer $S'$ defines the train to be fixed, and uses railway cars to label points $x'$ in space. This means that the lines of constant $x$ (in particular the $t$ axis) are not parallel to the lines of constant $x'$ (in particular the $t'$ axis).

$S$ uses coordinates $(x, t)$ such that $x$ is distance as measured by a tape measure, $t$ is time as measured by a clock, and $dx/dt = v$ is constant for a particle in the absence of forces. The same is true for $S'$. Such coordinates are called **inertial coordinates**, and they are familiar to you from Newtonian physics. From now we use units where 1 second is equivalent to $\simeq 3 \times 10^8$ meters, so that the speed of light is $c = 1$ in those units. For simplicity we shift the origin of coordinates so that at $x = 0$ and $x' = 0$ coincide at $t = 0$.

$S'$ has constant speed $\beta$ in the positive $x$ direction with respect to $S$. By this we mean that in inertial coordinates the world line $x' = 0$ is given by $x = \beta t$. In Newtonian physics, in inertial coordinates, this also implies that any world line $x' = \text{const}$ is given by $x = x' + \beta t$, or inversely

$$x' = x - \beta t. \tag{137}$$

In Newtonian physics, both observers agree on what is simultaneous, and so share the same absolute Newtonian time, or

$$t' = t. \tag{138}$$

(137,138) form what is called a **Galileo transformation**. Consider a bullet fired from a gun by $S'$, for simplicity at $x' = t' = 0$. The bullet has speed $v'$ relative to $S'$, and so has the trajectory $x'(t') = v't'$. From (137,138), $x(t) = (v + \beta)t$, and so seen from $S'$, the bullet has speed

$$v = v' + \beta, \tag{139}$$

as we would expect. The same is true for the speed of a water wave in water that is stationary with respect to $S'$. This is called the **Galilean addition of velocities**.

If light behaved either like a bullet or like a water wave, one would expect the speed of light to depend on the observer in this way. However, this is not

true. In fact, Galilean velocity addition is wrong (it is only an approximation for speeds much smaller than the speed of light), and instead the following experimental fact holds:

**Invariance of speed of light:** The speed of light is measured to be the same by any two observers moving at constant relative velocity $\beta$, and in particular, is the same going in any direction.

## 7.2 Lorentz transformations

Fig. 2 illustrates the following. $S$ sees that a flash of light leaves $x = x_A$ at $t = 0$, reaches $x = x_B$ at $t = T$, and (because the speed of light must be the same going both ways) after reflection it gets back to $x = x_A$ at $t = 2T$.

As a physical fact, but in contradiction of Newtonian physics, light also travels with the same speed going both ways for $S'$. This means that the same flash of light leaves $x' = x'_C$ at $t' = 0$, reaches $x' = x'_D$ at $t' = T'$, and after reflection it gets back to $x' = x'_C$ at $t' = 2T'$.

And that is all we need to derive the kinematics of special relativity!

It is easy to see that (137,138) are incompatible with the invariance of the speed of light light. However, if we just assume that $x' = \text{const}$ implies $dx/dt = \beta$, we can generalise (137) slightly to

$$x' = \gamma(x - \beta t) \tag{140}$$

for some constant $\gamma > 0$ that depends only on $\beta$. Because left and right are on an equal footing, $\gamma$ should be an even function of $\beta$, and clearly $\gamma(0) = 1$.

Return now to Fig. 2. Consider the perpendicular triangle $OPQ$. By construction, the distances $OR$ and $RQ$ are equal. Hence $R$ is the mid-point of a rectangle with $OQ$ as its diagonal. But this also means that $OR$ and $RP$ are equal, so $OPR$ is an equilateral triangle. $OPS$ is the reflection of $OPR$ through $OP$. This argument shows that the angle between the $x$ and $x'$ axes is the same as between the $t$ and $t'$ axes.

In algebraic terms, this means that

$$t' = \gamma(t - \beta x) \tag{141}$$

with the *same* $\gamma$ as in (140). (141) is counterintuitive, because it implies the **relativity of simultaneity:** $t = \text{const}$ is not the same as $t' = \text{const}$. There is no longer a preferred universal Newtonian time.

From the point of view of $S'$, $S$ has speed $-\beta$ and so we also have

$$x = \gamma(x' + \beta t'), \tag{142}$$

$$t = \gamma(t' + \beta x'), \tag{143}$$

again with the same $\gamma$. We can use any three of these equations to solve for $\gamma$, for example, using the first three,

$$x' + \beta t' = \gamma(1 - \beta^2)x = \gamma^{-1}x, \tag{144}$$

and find

$$\gamma = (1 - \beta^2)^{-1/2}. \tag{145}$$

(140,141), with $\gamma$ given by (145), form what is called a Lorentz transformation. (142,143) is its inverse. Coordinates in which these transformations hold are

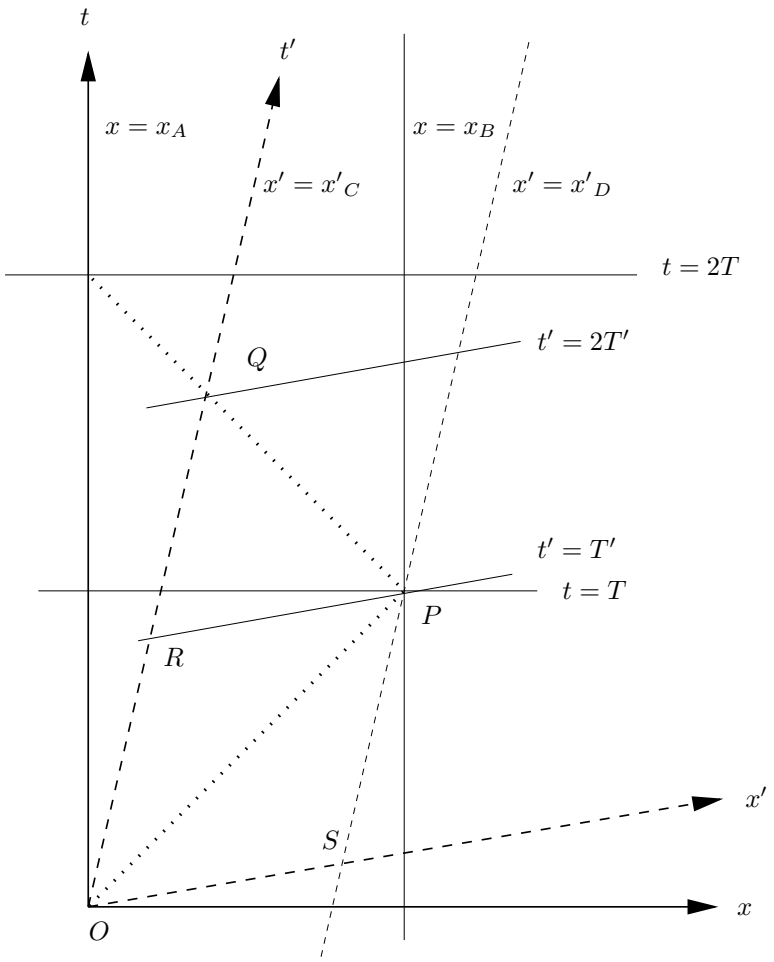


Figure 2: Graphical derivation of (141) from (140) and the assumptions that a light ray (dotted lines drawn at 45 degrees) travels with the same speed left and right in both $S$ and $S'$. The thin vertical lines are the world lines $x = x_A$ and $x = x_B$ of two observers at rest in $S$. The thin lines angled slightly to the right are the world lines $x' = x'_C$ and $x' = x'_D$ of two observers at rest in $S'$. (For simplicity, we have set $x_A = x'_C = 0$.) The horizontal lines are moments of time $t$ in $S$, and the lines angled slightly up are lines of constant time $t'$ in $S'$.

called **Minkowski coordinates**. Without proof (it is straightforward), we state that, with relative velocity in the $x$-direction, in 3 dimensions,

$$y' = y, \tag{146}$$
$$z' = z. \tag{147}$$

## 7.3 Consequences of the Lorentz transformation

And now we can derive the basic kinematic effects of (special) relativistic kinematics. This is only a sketch, and we do not have time for the generalisation to movement in 3 dimensions.

**Time dilation:** Place a clock on the world line $x' = 0$. The time it measures is $t'$. But

$$x' = 0 \quad \Rightarrow \quad t = \gamma t' > t', \tag{148}$$

so $t' < t$ and from the point of view of $S$ the moving clock at $x' = 0$ ticks more slowly. But we treated $S$ and $S'$ symmetrically, so it is also true that $S'$ sees the clock at $x = 0$ tick more slowly by the same factor. Each observer has the right to say that moving clocks tick more slowly.

**Space contraction:** Consider the moment of time $t = 0$. Now

$$t = 0 \quad \Rightarrow \quad x' = \gamma x > x, \tag{149}$$

and so a stick that is seen to have length $x'$ in $S'$ has length $x < x'$ as seen by $S$. Each observer has the right to say that moving objects contract in the direction of motion.

**Invariance of the speed of light:** We have really only used that light has the same speed going left and right. Say light has speed one in $S$. Then

$$-(\Delta t)^2 + (\Delta x)^2 = 0 \quad \Leftrightarrow \quad -(\Delta t')^2 + (\Delta x')^2 = 0, \tag{150}$$

and so it also has speed one in $S'$.

**Invariance of proper time:** In fact,

$$-(\Delta t)^2 + (\Delta x)^2 = -(\Delta t')^2 + (\Delta x')^2 =: -(\Delta \tau)^2. \tag{151}$$

If we apply these to the worldline $x' = 0$, then $\Delta\tau = \Delta t'$, and we see that $\tau$ is time elapsing for a clock at rest in $S'$. But we also see that $-(\Delta\tau)^2$ is given by $-(\Delta t)^2 + (\Delta x)^2$ in an arbitrary frame, and is invariant under changes of frame, just like the velocity of light. It is called the **proper time** of that observer.

**Relativistic velocity addition:** Consider at first motion with speed $v_\parallel$ parallel to the relative motion (with speed $\beta$ of the two observers. Substituting $x' = v'_\parallel t'$ into (140,141), we find $x(1 + v'_\parallel \beta) = t(v'_\parallel + \beta)$, and so

$$v_\parallel = \frac{v'_\parallel + \beta}{1 + v'_\parallel \beta} \tag{152}$$

Note that $v'_\parallel = \pm 1$ implies $v_\parallel = \pm 1$, as we put in by construction, and that

$$v_\parallel = (v'_\parallel + \beta)\left[1 + O(v'_\parallel \beta)\right], \tag{153}$$

so in the limit in which both $v'_\parallel$ and $\beta$ are small we recover the Galilean velocity addition (139).

Similarly, the component of the velocity perpendicular to the relative motion transform as

$$v_\perp = \frac{v'_\perp}{\gamma(1-\beta v'_\parallel)}. \tag{154}$$

Again $v_\perp^2 + v_\parallel^2 = 1$ if and only if ${v'_\perp}^2 + {v'_\parallel}^2 = 1$, and $v_perp \simeq v'_\perp$ in the low-velocity limit.

## 7.4 *Exercises*

1. A car fits precisely into a garage so that with the door closed there is no space left. Now the car drives into the garage at speed, and as soon as its back is inside, somebody closes the door. The driver says: The garage was moving with respect to me, and so had shrunk; the car hit the back wall before the door could be closed. The doorman says: the car was moving, and so had shrunk; I could close the door and there was space left. Resolve this paradox.

# 8 Spacetime

## 8.1 Newtonian spacetime

We have derived the basics of special relativity in inertial coordinates. We shall now restate it in geometric form, in terms of the geometry of spacetime. For contrast, we first revisit the geometric structure of Newtonian physics.

Newtonian physics already takes place in spacetime, in the sense that the PDEs of Newtonian physics such as the Euler equation have 4 coordinates, for example $(t, x, y, z)$. The movement of a point particle is described by a **world line** $(x, y, z) = (x(t), y(t), z(t))$.

However, time is clearly treated specially: we may change to other *spatial* coordinates, for example $(t, r, \theta, \varphi)$, but if we mixed up $t$ with the spatial coordinates, the equations would only become unnecessarily complicated.

Geometrically, **Newtonian spacetime** is a therefore a 4-dimensional manifold which has a preferred **foliation**, namely the surfaces of constant $t$. These **hypersurfaces** define what we mean by **simultaneous** in Newtonian physics. Furthermore, each Newtonian time surface is equipped with a flat Euclidean **3-dimensional metric of space** $\gamma_{ab}$ which defines angles and distances. This gives rise to a metric-covariant derivative $\nabla_a$, which appears for example in the fluid equations (see later). Note that the foliation, and the metric $\gamma_{ab}$ are geometric objects independently of the choice of spatial coordinates $x^i$. There are, however, preferred spatial coordinates $x^i$ called **Cartesian coordinates** in which $\gamma_{ij} = \delta_{ij}$ and $\nabla_i = \partial/\partial x^i$. They are not unique: A coordinate transformation that corresponds to a rigid rotation or translation leaves this form of the metric invariant, and is called a **Euclidean transformation**.

A set of inertial coordinates $(t, x^i)$ defines a **fibration** of Newtonian spacetime by curves of constant spatial coordinates $x^i$, whose tangent vector at each point is therefore $\partial/\partial t$. Note that the fibration as a geometric object is independent of coordinate transformations $x^i \to \tilde{x}^i(x)$ which do not involve $t$. By moving along one of these curves we identify points at different $t$, thus defining "the same point in space, at different times". However, these fibrations are not unique because observers moving at constant relative speed are equally valid. Their inertial coordinate systems are related by the Galileo transformations (137,138) on spacetime, combined with the Euclidean transformations of space. Fig. 3 summarises these concepts.

## 8.2 The spacetime of special and general relativity

The key difference between Newtonian physics (wrong) and relativistic physics (correct, as far as we know now) is the geometric structure of spacetime. From a geometric point, the relativistic structure is actually simpler. Spacetime is still a 4-dimensional manifold, and we can still call our coordinates $(t, x, y, z)$. However, there is no preferred foliation. Instead, there is a new preferred geometric structure on spacetime: this consists of all the curves in spacetime called that correspond to possible **light rays**: these are called **null curves**. We shall also see that there is a new preferred metric, but it is a **4-dimensional metric of spacetime**, and it is not Euclidean.

Generalising (151) to 3D, we have

$$-(\Delta\tau)^2 = -(\Delta t)^2 + (\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2. \tag{155}$$

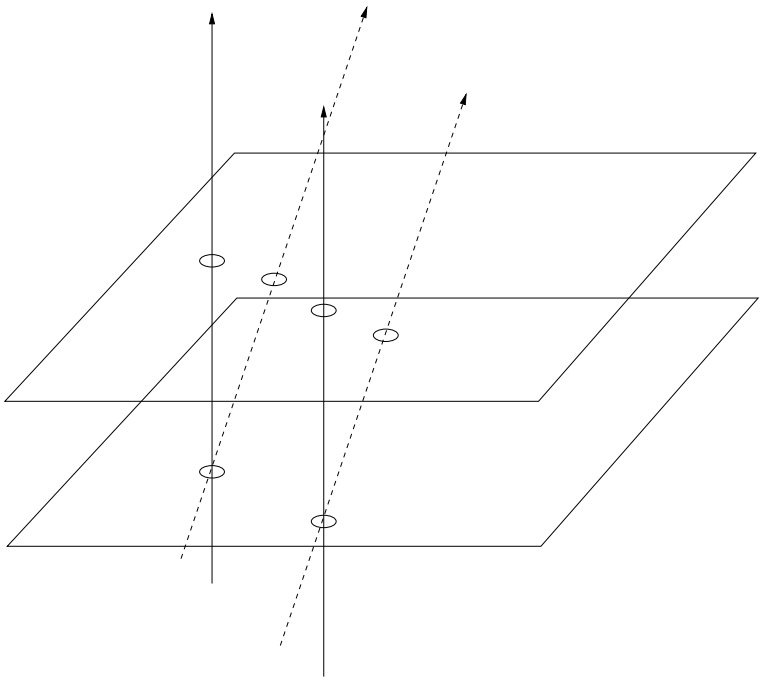

Figure 3: The geometric structure of Newtonian spacetime: through each point there is a unique hypersurface (given by $t = \text{const}$ in inertial coordinates) with a 3-dimensional flat, Euclidean metric $\gamma_{ab}$, and family of curves (with tangent vector $\partial/\partial t$ in inertial coordinates), two of which are shown (continuous and dashed lines).

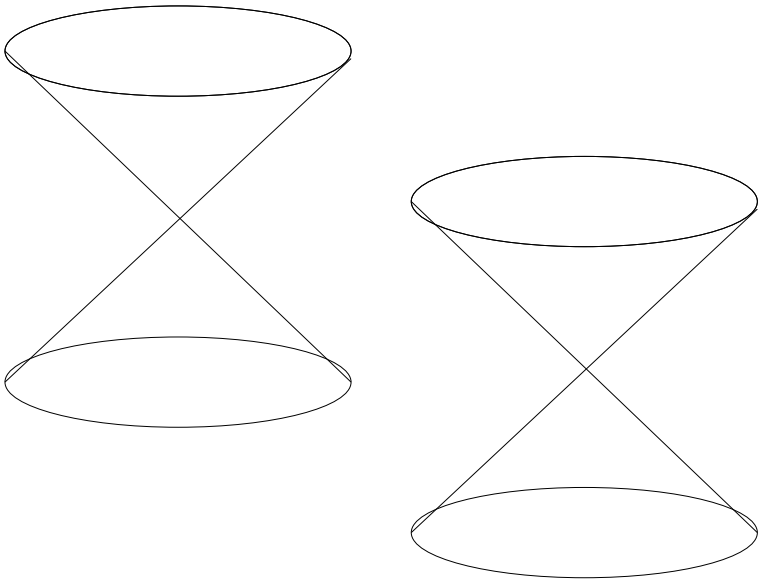

Figure 4: The geometric structure of Minkowski spacetime: through each point there is a lightcone formed from all null vectors $v^a$. It is given by the 4-dimensional, Lorentzian, metric $\eta_{ab}$, where null vectors obey $v^a v^b \eta_{ab} = 0$.

We can write this as

$$-\Delta\tau^2 = \eta_{\mu\nu}\,\Delta x^\mu\,\Delta x^\nu \tag{156}$$

where

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \tag{157}$$

and the Einstein summation convention is used. From now on I will use the convention that $(i, j, k, \ldots)$ are coordinate indices in 3-dimensional space, and $(\kappa, \lambda, \mu, \nu, \ldots)$ coordinate indices in 4-dimensional spacetime. By contrast $(a, b, c, \ldots)$ will be abstract tensor indices in spacetime.

But $\Delta\tau$ can be measured by clocks, and so must be independent of coordinates on the spacetime. Therefore $\eta_{ab}$ must be a tensor, the **Minkowski metric** tensor. It is symmetric, but not positive definite. For a vector $v^a$, $v^a v^b \eta_{ab} < 0$ if $v^a$ is the tangent to a possible world line of a physical object (**timelike**), $> 0$ if $v^a$ is tangent to a possible physical object at one moment of time (**spacelike**) and $= 0$ if $v^a$ is tangent to a possible light ray (**null**).

For a curved timelike world line parameterised by $x^\mu(s)$ , we have to write

$$\tau = \int_s \left(\eta_{\mu\nu}\frac{dx^\mu}{ds}\frac{dx^\nu}{ds}\right)^{1/2} ds, \tag{158}$$

and this is often written in shorthand as the **line element**

$$ds^2 = -d\tau^2 = g_{\mu\nu}\,dx^\mu\,dx^\nu. \tag{159}$$

## 8.3 The Lorentz group

The preferred coordinate systems in which the Lorentz metric takes the form (157) are called **Minkowski coordinates**. As a tensor, $\eta_{ab}$ transforms under coordinate transformations as

$$\tilde{\eta}_{\alpha\beta} = \frac{\partial x^\mu}{\partial \tilde{x}^\alpha}\frac{\partial x^\nu}{\partial \tilde{x}^\beta}\eta_{\mu\nu}. \tag{160}$$

The condition that a coordinate transformation with $\partial x^\mu/\partial\tilde{x}^\alpha \equiv L^\mu{}_\alpha$ does not change the simple Minkowski form of the metric, that is $\tilde{\eta}_{\alpha\beta} = \eta_{\alpha\beta}$, is

$$L^\mu{}_\alpha L^\nu{}_\beta \eta_{\mu\nu} = \eta_{\alpha\beta}, \tag{161}$$

or in matrix notation

$$L^t \eta L = \eta. \tag{162}$$

These coordinate transformations are called the Lorentz transformations, and the matrices $L$ form the **Lorentz group**. It comprises Euclidean rotations that only involve $(x, y, z)$, as well as the particular Lorentz transformations called **boosts** that we have studied above and which mix $x$ (or $y$ or $z$) and $t$. The full group of coordinate transformations that leaves (157) invariant is called the **Poincaré group**. It consists of the Lorentz group, combined with translations in space of time (which have $L = 1$).

The Lorentz group is an example of a **Lie group**, a group that is also a manifold. Consider a one-parameter family of elements $L(\lambda)$ of the group (i.e. a curve in the group manifold), with

$$L(0) = I, \qquad \left.\frac{d}{d\lambda}\right|_{\lambda=0} L =: X \tag{163}$$

$X$ is called the **generator** of this one-parameter family (and is a tangent vector of the group manifold at the point $I$). We then have

$$0 = \left.\frac{d}{d\lambda}\right|_0 \left(L^T \eta L\right) = X^t \eta + \eta X, \tag{164}$$

or in tensor notation

$$X^\mu{}_\alpha \eta_{\mu\nu} + \eta_{\alpha\beta} X^\beta{}_\nu = 0. \tag{165}$$

The $X$ obeying this relation form a 6-dimensional vector space, the **Lie algebra** of the Lorentz group. The 6 vectors spanning it can be thought of as an infinitesimal rotation about the $x$-axis, an infinitesimal boost along the $x$-axis, and similarly for $y$ and $z$.

## 8.4 Killing vectors of the Minkowski metric

The generators of the Lorentz group are related to Killing vectors of the Minkowski metric – both express symmetries of **Minkowski spacetime**. A **Killing vector** of any given metric $g_{ab}$ is a vector field $\xi^a$ such that

$$\mathcal{L}_\xi g_{ab} = 0. \tag{166}$$

This expresses the idea that the metric $g_{ab}$ is "the same" along the integral curves of $\xi^a$ (the curves whose tangent vector is $\xi^a$).

We use the standard formula for the Lie derivative of a $(0,2)$-tensor $\eta_{ab}$ along a vector field $\xi^a$,

$$\mathcal{L}_\xi \eta_{\mu\nu} = \xi^\lambda \eta_{\mu\nu,\lambda} + \xi^\lambda{}_{,\mu} \eta_{\lambda\nu} + \xi^\lambda{}_{,\nu} \eta_{\lambda\mu} \tag{167}$$

In Minkowski coordinates $\eta_{\mu\nu}$ is constant (given by (157), and we define $\xi_\mu \equiv \eta_{\mu\nu}\xi^\nu$, that is, we have used the metric to move indices. Then

$$\mathcal{L}_\xi \eta_{\mu\nu} = \xi_{\mu,\nu} + \xi_{\nu,\mu}, \tag{168}$$

So we need to solve

$$\xi_{\mu,\nu} + \xi_{\nu,\mu} = 0. \tag{169}$$

It follows that

$$\xi_{\mu,\mu} = 0 \quad \text{(no summation over } \mu\text{)} \tag{170}$$

and hence

$$\left(\xi_{\mu,\nu} + \xi_{\nu,\mu}\right)_{,\mu} = \xi_{\nu,\mu\mu} = 0 \quad \text{(no summation over } \mu\text{)}. \tag{171}$$

The general solution of this system of differential equations is clearly

$$\xi_\mu(x) = c_{\mu\nu} x^\nu + d_\mu, \tag{172}$$

where $c_{\mu\nu} = -c_{\nu\mu}$ and $d_\mu$ are constants.

Raising the index on $\xi_\mu$ again, and calling the coordinates $x^\mu = (t, x^i)$ with $x^0 = t$, we have

$$\begin{aligned} \xi \;=\; & -d_0 \frac{\partial}{\partial t} \qquad \text{(time translation)} \\ & +d_i \frac{\partial}{\partial x^i} \qquad \text{(space translations)} \\ & +c_{ij} \frac{1}{2} \left( x^j \frac{\partial}{\partial x^i} - x^i \frac{\partial}{\partial x^j} \right) \qquad \text{(rotations)} \\ & +c_{i0} \left( t \frac{\partial}{\partial x^i} + x^i \frac{\partial}{\partial t} \right) \qquad (\textbf{boosts}) \end{aligned} \tag{173}$$

(Note the signs! They come from raising the index on $\xi_\mu$ to get $\xi = \xi^\mu \partial/\partial x^\mu$.) These 10 vectors are equivalent to the generators of the Poincaré group (exercise).

Finally, note that $\xi^a$ and $\eta_{ab}$ are tensors, so although we have derived the kinematics of special relativity in Minkowski coordinates, we can now express them in arbitrary coordinates.

## 8.5 *Exercises*

1. Spell out the details of the relationship between the generators of the Lorentz groups and the boost and rotation Killing vectors of Minkowski spacetime.

2. (Blandford and Thorne, Exercise 1.12a)

   Show that if $n^i$ is a 3-dimensional unit vector (with respect to the Euclidean metric $\delta_{ij}$), with $\beta$ and $\gamma$ defined as above, the following is a Lorentz transformation:

   $$L^0{}_0 = \gamma, \quad L^0{}_j = L^j{}_0 = \beta\gamma n^j, \quad L^j{}_k = L^k{}_j = (\gamma - 1) n^j n^k + \delta^{jk}. \tag{174}$$

   (Here, spatial indices are raised and lowered with $\gamma_{ij} = \delta_{ij}$.)

# 9 Dynamics in special relativity

## 9.1 4-velocity

To obtain the dynamics of general relativity, we take 3-dimensional vectors and tensors from Newtonian physics such as velocity or force, and generalise them to 4-dimensional ones.

Consider the world line of a point particle parameterised by its proper time, $x^\mu = x^\mu(\tau)$. Because $\tau$ is a scalar, the **4-velocity**

$$u^\mu = \frac{dx^\mu}{d\tau} \tag{175}$$

is in fact a tangent vector $u^a$ on Minkowski space. (One could consider Minkowski space itself as a 4-dimensional vector space, which is isomorphic to its tangent space at each point, but it is clearer if we only consider it as a manifold, with a tangent at each point. That view is also the one that will generate to general relativity later.) At any point on the world line we can align Minkowski coordinates $x^\mu = (t, x^i)$ so that instantaneously (meaning, at that point on the world line) $u^\mu = (1, 0, 0, 0)$. But this means that $u^\mu u^\nu \eta_{\mu\nu} = -1$. But this is a scalar, and so at every point on the worldline we must have

$$u^a u_a = -1, \tag{176}$$

where we have used the metric $\eta_{ab}$ to move the index on $u_a \equiv \eta_{ab} u^b$.

In another Minkowski coordinate system, one which is not aligned with the world line, we have

$$-1 = \eta_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = \left(\frac{dt}{d\tau}\right)^2 \eta_{\mu\nu} \frac{dx^\mu}{dt} \frac{dx^\nu}{dt} = \left(\frac{dt}{d\tau}\right)^2 (-1 + v^2), \tag{177}$$

where we have defined $v^i = dx^i/dt$ and $v^2 = v^i v^j \gamma_{ij}$ as in Newtonian physics. This means that

$$\frac{dt}{d\tau} = \gamma, \tag{178}$$

where the **Lorentz factor** $\gamma$ is as defined in (145), and so in general Minkowski coordinates

$$u^\mu = \gamma(1, v^i). \tag{179}$$

## 9.2 4-momentum and 4-force

Now we shall try to generalise Newton's first law $dp^i/dt = f^i$ to a 4-dimensional vector equality

$$\frac{dp^a}{d\tau} = f^a, \tag{180}$$

where the vector $p^a$ is called the **4-momentum** and the vector $f^a$ is called the **4-force**. To make sense of this we *postulate* that

$$p^a p_a =: -m^2 = \text{const} \tag{181}$$

From this postulate it follows that

$$\frac{1}{2} \frac{d}{d\tau} (p^a p_a) = p_a f^a = 0. \tag{182}$$

For a point particle of **rest mass** $m > 0$ (the mass measured in the frame where it is instantaneously at rest, i.e. our usual definition of mass), we can then consistently define

$$p^a = mu^a \qquad \Rightarrow \qquad m\frac{du^a}{d\tau} = f^a. \tag{183}$$

There are many experimental situations where it is useful to think of light or electromagnetic radiation as made up of particles. These particles, for example **photons**, move on world lines which are null, so that $u^a u_a = 0$. In this sense, photons are **massless**. For a massless particle (180) holds, but (183) does not. For a massless particle $u^a \propto p^a$, and both are null, but $u^a$ is now only defined up to an arbitrary factor, because proper time is zero along a null worldline.

(183) in Minkowski coordinates becomes

$$m\gamma\frac{du^\mu}{dt} = m\gamma\frac{d}{dt}\left[\gamma(1, v^i)\right] = (f^0, f^i). \tag{184}$$

But for $|v| \ll 1$, $\gamma = (1 - v^2)^{-1/2} \simeq 1 + v^2/2$, and so to leading order in $v$

$$m\frac{dv^i}{dt} = f^i + O(v^2), \tag{185}$$

$$\frac{d}{dt}\left(\frac{1}{2}mv^2\right) = f^0 + O(v^2) \tag{186}$$

Therefore, in the **Newtonian limit** $f^i$ is the usual force, and $f^0$ is the rate of change of energy. (Note that the Newtonian limit applies in any frame that is close enough to the instantaneous rest frame of the particle.)

The components of $p^\mu$ in Minkowski coordinates are

$$E := p^0 = m\gamma, \qquad p^i = m\gamma v^i, \tag{187}$$

the energy and 3-momentum respectively in the frame associated with those coordinates.

## 9.3 3+1 split

To make this **3+1 split** of $p^a$ and other 4-tensors more precise, we shall write it in geometric form, independently of the use of Minkowski coordinates. We shall use a method that is often used to obtain geometric expressions. We first obtain an equality between two tensors in some preferred coordinate system, but if we have taken care that both sides of the equations are actually components of tensors in that coordinate system, then we immediately have a geometric, or tensorial, equation.

Each Minkowski coordinate system corresponds to a **frame**, or family of observers. In some Minkowski coordinate system $(t, x^i)$, all observers at rest with respect to that frame have the same 4-velocity $w^\mu = (1, 0, 0, 0)$. We shall call $w^a$ the 4-velocity of this frame, or of any observer in that frame. Define

$$h_{ab} \equiv \eta_{ab} + w_a w_b \tag{188}$$

As always, we move indices implicitly with $\eta_{ab}$. It is easy to verify that

$$h_a{}^b h_b{}^c = h_a{}^c \quad \text{and} \quad h_a{}^b w_b = 0. \tag{189}$$

The first equation, which in matrix notation, or operator notation, can be written as $h^2 = h$ ($h$ is **idempotent**) means that $h_a{}^b$ (with one index up and one down) is a **projection operator** acting on the tangent space of spacetime.

The second equation tells us that it annihilates $w^a$, and so projects any 4-vector into the **surfaces of simultaneity** of $w^a$. Geometrically, they are the 3-surfaces with volume element

$$\epsilon^w_{abc} := \epsilon_{abcd} w^d. \tag{190}$$

They are normal to $w^a$ in the sense that $w^a \epsilon_{abc} = 0$. In those Minkowski coordinates in which $w^\mu = (1, 0, 0, 0)$ they are the surfaces $t = \text{const}$, and $\epsilon^w_{123} = \epsilon_{1230} w^0 = 1$.

$h_{ab}$ (with both indices down) is the **spatial metric** in those surfaces of simultaneity. In the preferred Minkowski coordinates its components are

$$h_{\mu\nu} = \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} = h^{\mu\nu} = h_\mu{}^\nu. \tag{191}$$

Therefore $h_{ab}$ is the 4-dimensional incarnation of the 3-dimensional metric $\gamma_{ab}$ we have used before, which has components

$$\gamma_{ij} = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} = \gamma^{ij} = \gamma_i{}^j \tag{192}$$

in the same coordinates.

We can redefine the Lorentz factor $\gamma$ geometrically as

$$\gamma := -w_a u^a =: (1 - v^2)^{-1/2}, \tag{193}$$

where $v$ is the magnitude of the **relative velocity** between $u^a$ and $w^a$. We can geometrically define the **energy** of a particle as

$$E_w := -w_a p^a. \tag{194}$$

The suffix $w$ here means *as measured by an observer with 4-velocity* $w^a$. For a **massive** particle,

$$E = m\gamma = m(1 - v^2)^{-1/2} = m + \frac{1}{2} m v^2 + O(v^4) \tag{195}$$

The first term (which, in ordinary units is $mc^2$) is the **rest energy**, while the remainder is the kinetic energy. We define

$$p^b_w := h_a{}^b p^a \tag{196}$$

as the **3-momentum** of the particle. This is formally a 4-vector, but as $p^b_w w_b \equiv 0$ it is purely **spatial with respect to** $w^b$, and has only 3 components. (In Minkowski coordinates where $w^\mu = (1, 0, 0, 0)$, $p^\mu_w = (0, p^i) = (0, m\gamma v^i)$.

## 9.4 Example: the electromagnetic force

As an example of force in special relativity, we consider the force a given electromagnetic field exerts on a charged point particle. (We take the electromagnetic field as given. The Maxwell equations will be covered in the next lecture.) The electromagnetic field is described by an antisymmetric rank-2 tensor

$$F_{ab} = -F_{ba} \tag{197}$$

which therefore has 6 independent components. (As always, we can raise or lower indices with $\eta_{ab}$ and $\eta^{ab}$.) The electromagnetic force on a point particle of charge $q$ is

$$f_a = qF_{ab}u^b. \tag{198}$$

(As $F_{ab}$ is antisymmetric, the order of indices matters!) Note that here the antisymmetry of $F_{ab}$ implies that $f_a u^a = 0$ and hence $f_a p^a = 0$ as required.

In Minkowski coordinates, the electric and magnetic fields are related to $F_{\mu\nu}$ by

$$F^{0i} = E^i, \qquad F^{ij} = \epsilon^{ij}{}_k B^k. \tag{199}$$

We find

$$m\frac{dp^\mu}{d\tau} = m\gamma\frac{d}{dt}(E, p^i) = qF^{\mu\nu}u_\nu, \tag{200}$$

or in components,

$$m\gamma\frac{dE}{dt} = qF^{0i}u_i = qE^i\gamma v_i, \tag{201}$$

$$m\gamma\frac{dp^i}{dt} = q\left(F^{i0}u_0 + F^{ij}u_j\right) = q(-E^i)(-\gamma) + \left(\epsilon^{ij}{}_k B^k\right)(\gamma v_j), \tag{202}$$

(We have moved the spatial indices with $\gamma_{ij} = \delta_{ij}$.) In the index-free vector calculus notation, this is

$$m\frac{dE}{dt} = q\vec{E}\cdot\vec{v}, \tag{203}$$

$$m\frac{d\vec{p}}{dt} = q(\vec{E} + \vec{v}\times\vec{B}). \tag{204}$$

The right-hand side of (204) is called the **Lorentz force**. The right-hand side of (203) is this force dotted into the velocity $\vec{v}$, the standard expression for power.

We can rewrite the split of $F_{ab}$ into the electric and magnetic field in tensor form as

$$E^a_w := F^{ab}w_b, \tag{205}$$

$$B^a_w := \frac{1}{2}\epsilon^{abcd}w_b F_{cd}. \tag{206}$$

$E^a_w$ and $B^a_w$ are again purely spatial vectors with respect to $w^a$. $F^{ab}$ can be reconstructed from $E$ and $B$ in any frame as

$$F^{ab} = -w^a E^b_w + w^b E^a_w - \epsilon^{ab}{}_{cd}w^c B^d_w. \tag{207}$$

## 9.5 *Exercises*

1. Check that in Minkowski coordinates (205,206) gives (199). Careful with signs on raising and lowering indices, or interchanging indices on $F_{\mu\nu}$.
2. Check that (207) is the inverse of (205,206).

# 10 Maxwell's equations

## 10.1 Tensor notation

Maxwell's equations determine the electromagnetic field given the charges and currents, initial data and boundary conditions. They can be written in 4-dimensional tensor form as

$$\nabla_a F^{ab} = -4\pi j^b, \tag{208}$$
$$\nabla_a \star F^{ab} = 0, \tag{209}$$

where $\star F_{ab}$ is the Hodge dual of $F_{ab}$ as defined in (120) (for $n = 4$ and $p = n - p = 2$), that is

$$\star F_{ab} := \frac{1}{2!}\epsilon_{ab}{}^{cd} F_{cd}. \tag{210}$$

Here $\eta_{ab}$ is the Minkowski metric, which is also used to move indices implicitly, $\nabla_a$ is the covariant derivative compatible with $\eta_{ab}$, and $\epsilon_{abcd}$ is the volume form compatible with both. As by definition $\nabla_e \epsilon_{abcd} \equiv 0$, (209) can also be written as

$$\epsilon^{abcd} \nabla_b F_{cd} = 0. \tag{211}$$

(If you worry about units: we work in Gauss units, with the additional assumption that $c = 1$).

$j^a$ is the **current 4-vector**, which splits into a charge density and a 3-dimensional current density with respect to an observer with 4-velocity $w^a$ as

$$\rho_w := -w_a j^a, \qquad j^b_w := h_a{}^b j^a. \tag{212}$$

In Minkowski coordinates aligned with that observer, $j^0 = \rho$ and $j^i$ is the 3-current. Note that

$$\nabla_b j^b = \frac{1}{4\pi} \nabla_a \nabla_b F^{ab} = 0 \tag{213}$$

because $F^{ab}$ is antisymmetric. In Minkowski coordinates,

$$\frac{\partial \rho}{\partial t} + \frac{\partial j^i}{\partial x^i} = 0, \tag{214}$$

which is the law of charge conservation.

It can be shown that in Minkowski coordinates (208,209) are equivalent to, in index-free vector calculus notation,

$$\vec{\nabla} \cdot \vec{E} = 4\pi\rho, \tag{215}$$
$$\vec{\nabla} \times \vec{B} = \frac{\partial \vec{E}}{\partial t} + 4\pi \vec{j}, \tag{216}$$
$$\vec{\nabla} \cdot \vec{B} = 0, \tag{217}$$
$$\vec{\nabla} \times \vec{E} = -\frac{\partial \vec{B}}{\partial t}, \tag{218}$$

where $\vec{\nabla}$, $\cdot$ and $\times$ have their usual meaning in Euclidean space.

## 10.2 Forms notation

Because $F_{ab}$ is totally antisymmetric, it is a 2-form. We can write it in index-free forms notation as

$$F := F_{\mu\nu} dx^\mu \otimes dx^\nu \tag{219}$$

Then

$$\epsilon^{abcd}\nabla_b F_{cd} = 0 \quad \Leftrightarrow \quad \nabla_{[a}F_{bc]} = 0 \quad \Leftrightarrow \quad dF = 0 \tag{220}$$

becomes an equation between differential forms, which no longer requires a particular covariant derivative.

Furthermore, using the definitions (210) and (104), we have

$$\begin{aligned}(\star d \star F)_a &= \frac{1}{3!}\epsilon_a{}^{bcd}(2+1)\nabla_{[b}\left(\frac{1}{2!}\epsilon_{cd]}{}^{ef}F_{ef}\right) \\ &= \frac{1}{4}\epsilon_{ab}{}^{cd}\epsilon^{ef}{}_{cd}\nabla^b F_{ef} = -\delta_{[a}{}^e\delta_{b]}{}^f\nabla^b F_{ef} \\ &= -\nabla^b F_{ab} = 4\pi j^a. \end{aligned} \tag{221}$$

Therefore, we can write the Maxwell equations in forms notation as

$$dF = 0, \tag{222}$$

$$\star d \star F = 4\pi j, \tag{223}$$

where $j$ is forms notation for the 1-form $j_a$.

The equality in the second line of (221) is an example of a product of tensors $\epsilon$, all of which can be worked out by the same two-step procedure as follows. The volume form $\epsilon$ only knows about the metric $g_{ab}$ (here $\eta_{ab}$), and so the product can only be constructed from $g_{ab}$, $g^{ab}$ and $g_a{}^b = \delta_a{}^b$. Furthermore, $\epsilon$ is totally antisymmetric. This fixes any product of only $\epsilon$ tensors already up to a constant factor $k$. In our example,

$$\epsilon_{ab}{}^{cd}\epsilon^{ef}{}_{cd} = k\eta_{[a}{}^e\eta_{b]}{}^f = k\delta_{[a}{}^e\delta_{b]}{}^f = \frac{k}{2}\left(\delta_a{}^e\delta_b{}^f - \delta_b{}^e\delta_a{}^f\right). \tag{224}$$

To work out $k$, we simply work out a suitable contraction of all indices which does not vanish, and compare with (116). Note that the signature of the metric $g_{ab}$ comes in at this stage. In our example, with negative signature,

$$-4!\epsilon_{ab}{}^{cd}\epsilon^{ab}{}_{cd} = \frac{k}{2}\left(\delta_a{}^a\delta_b{}^b - \delta_b{}^a\delta_a{}^b\right) = \frac{k}{2}(4\cdot 4 - 4) = -6k, \tag{225}$$

and so $k = -4$.

## 10.3 *Exercises*

1. Derive (215-218) by inserting (207) into (208,209) and contracting with $w^a$ and $h_a{}^b$.

# 11 Solutions of Maxwell's equations

## 11.1 Vector potential

It is common to solve the homogeneous Maxwell equations, (209) or $dF = 0$, by making the ansatz

$$F_{ab} = \nabla_a A_b - \nabla_b A_a, \tag{226}$$

or $F = dA$ in forms notation, for the **4-vector potential** $A_a$. (Strictly speaking, it is a 1-form rather than a vector field, but we can raise the index with the spacetime metric $g^{ab}$). From this it follows that $dF = ddA = 0$,

However, in this ansatz $A_a$ is not unique. In forms language, $d(A + d\varphi) = dA + ddφ = dA = F$, and so instead of $A$ we can equally well use $A' = A + d\varphi$ for an arbitrary scalar field $\varphi$. This non-uniqueness is called **gauge freedom**. The physical electromagnetic field $F_{ab}$, which can be measured experimentally from the force it exerts on charges, is unchanged under such a **gauge transformation**.

With $A_\mu \equiv (\phi, A_i)$ in Minkowski coordinates, we have, in index-free vector calculus notation

$$\vec{B} = \vec{\nabla} \times \vec{A}, \qquad \vec{E} = \vec{\nabla}\phi - \frac{\partial \vec{A}}{\partial t}, \tag{227}$$

and it is easy to see that this ansatz guarantees the homogeneous Maxwell equations in 3+1 form. The gauge transformation in 3+1 form is

$$\vec{A}' = \vec{A} + \vec{\nabla}\varphi, \qquad \phi' = \phi + \frac{\partial \varphi}{\partial t}, \tag{228}$$

and it is easy to see that this leaves (227) invariant.

## 11.2 Lorentz gauge

Expressed in terms of the vector potential, the inhomogeneous Maxwell equation (209) becomes

$$\nabla_a F^{ab} = \nabla_a \nabla^a A^b - \nabla^a \nabla^b A_a = \nabla_a \nabla^a A^b - \nabla^b \nabla^a A_a - R^{ab}{}_a{}^c A_c = -4\pi j^b, \tag{229}$$

where we have used (90) to commute the second set of covariant derivatives. In in Minkowski spacetime, the Riemann tensor term vanishes, and we assume this for the remainder of this section. The first term is a **wave operator**, as one can see by writing it in Minkowski coordinates:

$$\nabla_\mu \nabla^\mu A^\nu = \left(-\partial_0^2 + \partial_i \partial^i\right) A^\mu, \tag{230}$$

and we know how to solve a wave equation, but the second term confuses this picture.

We try to make a gauge transformation to a new gauge $A'_a$ where

$$\nabla_a A'^a = \nabla_a A^a + \nabla_a \nabla^a \varphi = 0. \tag{231}$$

This is a wave equation for $\varphi$ with a source term, which always has a solution. Therefore, we can always choose a gauge, called **Lorentz gauge**, such that $\nabla_a A^a = 0$. Then the Maxwell equations reduce to a wave equation for $A^a$,

$$\nabla_a \nabla^a A^b = -4\pi j^b \tag{232}$$

We note that (231) does not determine $\varphi$ completely. As any solution of an inhomogeneous PDE, the solution depends on the boundary conditions, and we can add a solution of the *homogeneous* PDE, here $\nabla_a \nabla^a \varphi = 0$, with suitable homogeneous boundary conditions. This additional term is called a complementary function in PDE language, and here represents the **residual gauge freedom** still left after imposing Lorentz gauge.

## 11.3 The initial value problem

We now carry out a 3+1 split to see if the Maxwell equations can be solved as an initial value problem. (Here we treat the sources $\rho$ and $j^i$ as given functions of $t$ and $x^i$.) This will also give us another viewpoint on gauge freedom and gauge choice. For simplicity, we carry out the 3+1 split in Minkowski coordinates; it could also be done geometrically using $w^a$ and $h_a{}^b$.

The inhomogeneous Maxwell equation, expressed in terms of the gauge potential, in Minkowski coordinates is

$$\partial_\mu \left( \partial^\mu A^\nu - \partial^\nu A^\mu \right) = -4\pi j^\nu \tag{233}$$

Setting $\nu = 0$ and $\nu = k$, we obtain respectively

$$\partial_i \partial^i A^0 + \partial_i (\partial_t A^i) = -4\pi\rho, \tag{234}$$

$$(-\partial_t^2 + \partial_i \partial^i) A^k + \partial^k (\partial_t A^0 + \partial_i A^i) = -4\pi j^k, \tag{235}$$

where we have written $\partial_0 = \partial_t$ and $\partial^0 = -\partial_t$. We want to understand how these equations can be solved by integrating forward in time, starting from data at one moment in time, say $t = 0$.

We note first that the highest time derivatives in (235) are second time derivatives of the $A^k$: (235) are in fact three wave equations for the $A^k$. This means that the initial data to be specified at $t = 0$ must include both $A^k$ and $\partial_t A^k$.

As we already have a wave equation for the $A^k$, (234) cannot be an evolution equation for $\partial_i A^i$ but instead must be interpreted as a Poisson equation for $A^0$. (It determines $A^0$ for given $\partial_t A^i$ and $\rho$, and suitable boundary conditions, for example the value of $A^0$ at the boundary or at infinity.)

But then how can $\partial_t A^0$ appear in (235)? At $t = 0$, our free data are $A^i$ and $\partial_t A^i$, and these also determine $A^0$ at $t = 0$. But $\partial_t A^0$ is *not* determined, and in fact is free, not only at $t = 0$ but at every $t$. This is not in conflict with the fact that (234) determines $A^0$ at every $t$: intuitively, the arbitrary **gauge choice** of $\partial_t A^0$ at every $t$ steers the part of $A^i$ that is gauge-dependent just so that the correct $A^0$ comes out of (234). More precisely, (234) and (235) are **compatible** (see exercise).

## 11.4 Plane waves

The Maxwell equations in flat spacetime, in Lorentz gauge, in the absence of charge are

$$\partial_\nu A^\nu = 0, \qquad \partial_\mu \partial^\mu A^\nu = 0. \tag{236}$$

A simple but important family of solutions are the **plane wave solutions**. The ansatz

$$A_\mu(x) = \tilde{A}^\mu e^{i k_\nu x^\nu}, \tag{237}$$

where $\tilde{A}^\mu$ and $k_\nu$ are real constants, gives

$$\partial_\mu A_\lambda = ik_\mu A_\lambda, \qquad \partial_\mu \partial_\nu A_\lambda = -k_\mu k_\nu A_\lambda, \tag{238}$$

and the Lorentz gauge and wave equation respectively give

$$k_\nu \tilde{A}^\nu = 0, \qquad k_\mu k^\mu = 0. \tag{239}$$

## 11.5 *Exercises*

1. From (234,235), show that if Lorentz gauge is imposed at $t = 0$ by setting $\partial_t A^0 = -\partial_i A^i$, it is actually obeyed at all times. (Take the time derivative of this condition and show it vanishes subject to (234,235) and itself being obeyed.)

2. Show that (234) is equivalent to $\nabla \cdot E = 4\pi\rho$. Show that if this is obeyed initially, its time derivative vanishes automatically subject to (235).

3. Write out a plane wave in the $x^1$ direction, with $k_\mu = (k_0, k_1, 0, 0)$, in components, and show that it has two polarisations (in the $x^2$ and $x^3$ directions).

# Part III
# Thermodynamics

## 12 The first law

### 12.1 The state manifold

Heat is the energy contained in the unordered motion of the atoms in a macroscopic body. An area of physics called **statistical mechanics** uses this to derive the laws of **thermodynamics**, which operates with concepts like heat, energy and temperature, without ever mentioning atoms. In this course, we restrict ourselves to thermodynamics, which is much easier to teach if one uses the concepts of an abstract manifold and of 1-forms.

In many energy transfers, heat can be neglected to a good approximation. We then deal with a purely **mechanical system**. As an example of such a system, energy can be put into a coiled metal spring by stretching it, and that energy can be recovered at will later by relaxing it. Energy can be stored in a volume of gas by compressing it, and released by relaxing it. Energy can be stored in a car battery by running a current through it, and that energy can be recovered by running an opposite current through it. The **internal energy** $E$ of a mechanical system depends only on one or more **macroscopic variables** $x^i$,

$$E = W(x^i) \tag{240}$$

It is useful to think of this geometrically, as the definition of a scalar function $E$ on the **state manifold** of the system, considering the $x^i$ as coordinates. Each point on the manifold defines a **macroscopic state**.

The exterior derivative of $E$ is an exact 1-form on the state manifold,

$$dE = f_i dx^i, \qquad f_i = \frac{\partial W}{\partial x^i}, \tag{241}$$

where once again we use the Einstein summation convention. The $x^i$ are also called **generalised positions** (in our examples: length of the spring, volume of the gas, charge of the battery) and the $f^i$ **generalised forces** (tension of the spring, pressure of the gas, voltage of the battery). Note we have

$$0 \equiv ddE = df_i \wedge dx^i, \tag{242}$$

or in tensor notation $f_{i,j} - f_{j,i} = 0$.

### 12.2 Heat

Not all energy is mechanical, described by $W(x^i)$. We can put heat into a heat engine (for example a piston steam engine), and get out some waste heat and some mechanical energy. More heat has gone in than has come out, but after one cycle the engine is in the same state as before, and so, unlike "internal energy", there can be no variable "internal heat" to describe the state of the engine However, putting in or taking out heat clearly changes the internal

energy, and so for a thermodynamic system one formulates the **first law of thermodynamics:**

$$dE = \delta Q + f_i dx^i. \tag{243}$$

When heat is involved, the state manifold $M$ is not $n$-dimensional but $(n+1)$-dimensional. A possible set of local coordinates is $(E, x^i)$, where $E$ is no longer a function of the $x^i$, but independent. Note that $\delta Q$ is not exact: there exists no **state function** (scalar on the state manifold) such that $\delta Q = dQ$. Because $dE$ is closed and $\delta Q$ is not, neither is $f_i dx^i$, in contrast to the purely mechanical case. In other words, $f_{i,j} - f_{j,i} \neq 0$.

In elementary texts, Eq. (243) is stated as a relation between "infinitesimal changes". More rigorously, it is an equality between 1-forms on the state manifold $M$, which can be integrated over a curve $\gamma$ in the state manifold that represents a finite change of the system. The change of energy is

$$\Delta E = E_f - E_i = \int_\gamma dE = \int_\gamma \delta Q + \int_\gamma f_i dx^i. \tag{244}$$

Because $dE$ is closed, $\Delta E$ just depends on the initial and final state (endpoints of $\gamma$), but because $\delta Q$ is not exact, both $\int_\gamma \delta Q$ and $\int_\gamma f_i dx^i$ depend on the shape of $\gamma$ and not just its endpoints.

## 12.3 Temperature and entropy

Although $\delta Q$ is not exact, for any thermodynamic system there are two state variables, the **temperature** $T$ and the **entropy** $S$, such that

$$\delta Q = T dS. \tag{245}$$

It cannot be stressed enough that this is a key law of physics, not a mathematical definition. Mathematics tells us that for given 1-form $\delta Q$ we can always find scalars $T$ and $S$ only if $M$ is 1 or 2-dimensional, but not for 3 or more dimensions, so this law restricts $\delta Q$ for any system more complicated than a volume of gas in a cylinder. We will relate $T$ to the temperature already familiar from everyday experience below in Sec. 14.1.

So we can rewrite the first law in its final form as

$$dE = T dS + f_i dx^i. \tag{246}$$

$T$ and $S$ are uniquely defined (up to some trivial rescalings) by (246), and can therefore be measured implicitly through the generalised positions $x^i$ and forces $f_i$ with which we control the system, and the energy $E$ that goes into or comes out of the system. ($T$ can also be measured directly with a thermometer, but there is no instrument that measures $S$ directly.)

## 12.4 Thermodynamical equilibrium

What is behind (245)? At the microscopic level, a mole of gas (22.4 litres at room pressure and temperature) in a container consists of $N \simeq 6 \cdot 10^{23}$ atoms or molecules, each of which has 3 position and 3 momentum degrees of freedom. At the *microscopic* level, the state manifold (called phase space) is therefore $6N$-dimensional. However, even if we put all the molecules in one corner of

our container, after a very short time they will have spread throughout the container. (Proving this rigorously is subtle and difficult.)

This is an example of **thermodynamical equilibrium**: There is often a *macroscopic* level on which a system of many molecules can be described accurately by a very small number of variables, namely the mechanical variables, plus one more, which one can take to be $E$ (or $T$, or $S$). The state manifold is the manifold of equilibrium states, and curves on it correspond to changes of state in which equilibrium is preserved.

*When* thermodynamical equilibrium is a good approximation depends on the internal dynamics of the system, for example the forces between molecules. In general thermodynamical equilibrium is maintained if the system is large enough and all macroscopic variables $x^i$ are changed slowly enough – this may mean very small and very rapidly in practice. Whenever a system evolves from one thermodynamic equilibrium state into another, we need to know nothing at all about the internal dynamics to make predictions – this is the great strength of thermodynamics.

## 12.5 Examples

For a fixed (large) number of gas molecules in a container it turns out that two variables are enough, for example its volume $V$ and internal energy $E$. However, there are many other choices of coordinates on this 2-dimensional state manifold. Instead of $V$ we could use $p$ and instead of $E$ we could use $T$ or $S$. However, it would be less natural to use the pair $p$ and $V$ because they cannot be controlled separately in an experiment: to control the pressure, one needs to put the gas in a cylinder, with a weight on the piston, and let the volume adjust itself. The same holds for $x^i$ and $f_i$ (for the same $i$), or $T$ and $S$ – such pairs are called **conjugate variables**.

The first law for a fixed number of moles of gas is

$$dE = TdS - pdV. \tag{247}$$

A very similar example is a rubber band of length $l$, with an elastic force $f$, for which the first law is

$$dE = TdS + fdl. \tag{248}$$

If we consider situations in which the number $n$ of moles of gas in the container can changed at will, then the state manifold is 3-dimensional and the first law is

$$dE = TdS - pdV + \mu dn. \tag{249}$$

$\mu$ is the **chemical potential**, and from (249) is the increase of the internal energy, per mole of gas, as more gas (more atoms) is put into the fixed volume $V$ at fixed temperature $T$.

## 12.6 Intensive and extensive quantities

Some thermodynamical quantities add up if we put two subsystems together to make a new system, for example $E$, $V$, $n$, $l$ (adding the two rubber bands end to end); they are called **extensive**. Other quantities do not change if we put two identical subsystems together, for example $T$, $p$, $\mu$, but also quantities like $E/n$, $E/V$, $V/n$, $l/n$ etc.; these are called **intensive**.

A function $f(x^i)$ that obeys

$$f(\lambda x^i) = \lambda f(x^i) \tag{250}$$

for any $\lambda > 0$ is called homogeneous of order 1 in its arguments $x^i$. By taking a derivative of this identity with respect to $\lambda$ and setting $\lambda = 1$, we find

$$f = x^i \frac{\partial f}{\partial x^i} \tag{251}$$

(summation implied).

Now given what we have said about extensive quantities, we have $E(\lambda S, \lambda V, \lambda n) = \lambda E(S, V, n)$, and so

$$E = TS - pV + \mu n, \tag{252}$$

the **Euler equation** (for this particular system). By taking the exterior derivative of this equation, and subtracting (249), we find the **Gibbs-Duhem equation**

$$0 = SdT - Vdp + nd\mu. \tag{253}$$

## 12.7 *Exercises*

1. (This purely mathematical identity will come handy in the next Section). Consider three scalar fields $a$, $b$ and $c$ on a 2-dimensional manifold. If we use $b$ and $c$ as coordinates on the manifold, we can write

$$da = \left.\frac{\partial b}{\partial b}\right|_c db + \left.\frac{\partial a}{\partial c}\right|_b dc \tag{254}$$

   Rewrite $da \wedge db$ in coordinates $(b, c)$. Take the result and rewrite in coordinates $(c, a)$. Take that result and rewrite in coordinates $(a, b)$. Hence, derive

$$\left.\frac{\partial a}{\partial b}\right|_c \left.\frac{\partial b}{\partial c}\right|_a \left.\frac{\partial c}{\partial a}\right|_b = -1. \tag{255}$$

2. Does a similar result exist for manifolds of higher dimension?

# 13 Applications of the first law

## 13.1 Heat engines

One of the key results of early 19th century thermodynamics, long before it received its statistical mechanics foundation, was that a heat engine can never convert all the heat that goes in into mechanical energy. In an internal combustion engine, air is sucked in, mixed with petrol, and heated by combustion, then does work by expanding, and then expelled. A toy model for this process is a *fixed* quantity of gas (always the same molecules) which is heated in a cylinder, then expands, is then cooled, and contracts.

As $n$ is fixed, the first law for this toy model is

$$dE = TdS - pdV. \tag{256}$$

However, we could also consider a more complicated machine by considering the more abstract

$$dE = TdS + f_i dx^i. \tag{257}$$

As the machine goes through one cycle $\gamma$ in its two-dimensional state space, its internal energy changes, but at the end of the cycle must be the same as before (or the cycle could not be repeated endlessly). This means that

$$0 = \int_\gamma dE = \int_\gamma TdS + \int_\gamma f_i dx^i. \tag{258}$$

Now the first integral on the right-hand side is heat, and the second is work. We are interested in the total work that the machine does on the environment

$$W_{\text{tot}} = -\int_\gamma f_i dx^i = \int_\gamma TdS \tag{259}$$

compared to the amount of heat that goes into the machine from the environment,

$$Q_{\text{in}} = \int_{\gamma_1} TdS, \tag{260}$$

where $\gamma_1$ is that part of the cycle where heat goes in, and $\gamma_2$ the part where heat goes out, so that $\gamma = \gamma_1 \cup \gamma_2$. (Note that the second equality in (259) is (258). Note also that (258) only for a closed cycle $\gamma$ – we could not express $Q_i n$ in terms of work.) The **efficiency** of the engine is defined as

$$\eta := \frac{W_{\text{tot}}}{Q_{\text{in}}} = \frac{\int_{\gamma_1 \cup \gamma_2} TdS}{\int_{\gamma_1} TdS}. \tag{261}$$

Note that we have eliminated the mechanical variables, and have defined the efficiency purely in terms of heat fluxes. For the toy model of a fixed amount of gas, we can use $T$ and $S$ themselves as coordinates on the 2-dimensional state manifold. For a higher-dimensional state manifold, consider the projection of $\gamma$ into the $TS$ plane. In either case, $\int_{\gamma_1} TdS$ is just the area between the curve $\gamma_1$ and the $S$-axis. Similarly, $W_{\text{tot}}$ is the area between the curves $\gamma_1$ and $\gamma_2$. (This is elementary, but can also be seen as an application of Stokes' theorem, with the area measured by $dT \wedge dS$.) Fig. 5 illustrates this.

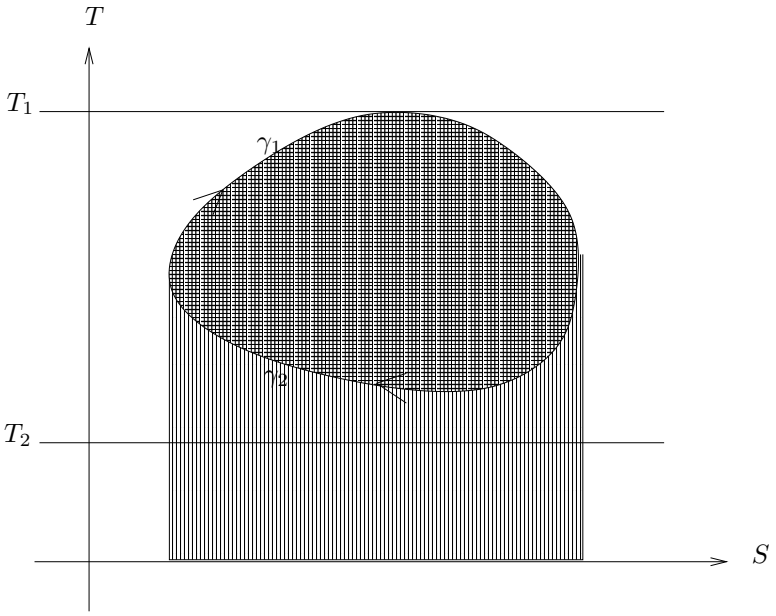


Figure 5: Plot of a heat engine cycle using coordinates $T$ and $S$ on the state space. The vertically hatched area represents $Q_{\text{in}}$, while the overlapping horizontally hatched area represents $W_{\text{tot}}$. $\eta$ is increased by pushing $\gamma_1$ or $\gamma_2$ down.

There is a maximum temperature $T_1$ at which heat can go into the machine (the temperature of the boiler in a steam engine, or the temperature at which the petrol-air mixture burns), and a minimum temperature $T_2$ at which it can leave (the ambient temperature).

Fig. 6 now illustrates the following geometrical fact: The largest efficiency is obtained if $\gamma$ is a rectangle in the $TS$-plane. This **Carnot cycle** consists of two **adiabatic** (constant entropy, no heat flux) curves ($\gamma_B$ and $\gamma_D$ in Fig. 6), and two **isothermal** (constant temperature) curves ($\gamma_A$ and $\gamma_C$). We find the **Carnot efficiency**

$$\eta_{\max} = \frac{T_1\Delta S - T_2\Delta S}{T_1\Delta S} = \frac{T_1 - T_2}{T_1}. \tag{262}$$

As we have eliminated the mechanical variables, this limit is independent of the specific construction of the machine - our derivation only assumes that heat goes in at $T \leq T_1$, heat goes out at $T \geq T_2$, and the machine returns to its original state after one cycle. It is a law of physics that no machine can transform heat into mechanical work (or electricity, or chemical potential energy) than a Carnot machine.

To make the Carnot cycle a bit more concrete, consider the toy model of a fixed amount of gas in a cylinder. Over $\gamma_A$ the gas is at temperature $T_1$, and is allowed to expand against resistance by the piston, thus doing work on the environment. The necessary energy is withdrawn from the environment in the form of heat. Over $\gamma_B$ the gas is allowed to expand further, while thermally insulated, doing more work while its internal energy falls. In this process, its temperature is allowed to fall from $T_1$ to $T_2$. Over $\gamma_C$, the gas is compressed

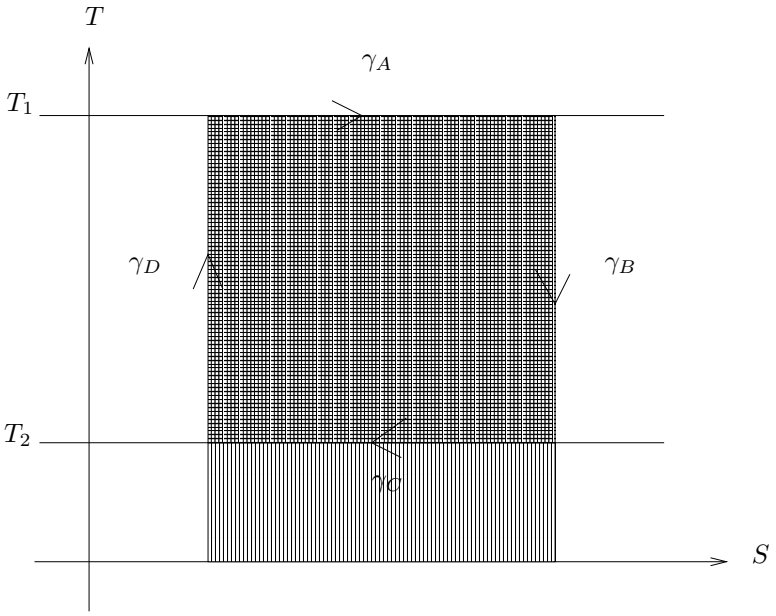


Figure 6: Plot of a Carnot cycle using coordinates $T$ and $S$ on the state space. The two vertical branches of $\gamma$ do not contribute to $\int TdS$, and so $\gamma_1$ is as high as possible and $\gamma_2$ is as low as possible.

while keeping its temperature fixed at $T_2$. Some of the work done to it by the environment flows into the environment as heat. Over $\gamma_D$, the gas is again thermally insulated, and is compressed to its starting volume, which causes its temperature to rise from $T_2$ back to $T_1$.

## 13.2 Thermodynamical identities

(246) is equivalent to the partial derivative identities

$$\left.\frac{\partial E}{\partial S}\right|_V = T, \qquad \left.\frac{\partial E}{\partial V}\right|_S = -p, \tag{263}$$

What is kept constant in a partial derivative is usually implied by the choice of coordinate system, but in thermodynamics one changes coordinate systems frequently, so it has become standard to write it explicitly.

If we rewrite the first law as

$$dS = \frac{1}{T}dE + \frac{p}{T}dV, \tag{264}$$

we find instead that

$$\left.\frac{\partial S}{\partial E}\right|_V = \frac{1}{T}, \qquad \left.\frac{\partial S}{\partial V}\right|_E = \frac{p}{T}. \tag{265}$$

We can obtain more identities by changing coordinates on the state manifold. In (247), $dS$ and $dV$ are elementary 1-forms if we use $S$ and $V$ as coordinates, but they become composite if we use, for example, $T$ and $V$ as coordinates:

$$dE \quad = \quad TdS - pdV$$

$$
\begin{aligned}
&= T\left(\left.\frac{\partial S}{\partial T}\right|_V dT + \left.\frac{\partial S}{\partial V}\right|_T dV\right) - p dV \\
&= T\left.\frac{\partial S}{\partial T}\right|_V dT + \left(\left.\frac{\partial S}{\partial V}\right|_T - p\right) dV, \qquad (266)
\end{aligned}
$$

which gives us

$$
T\left.\frac{\partial S}{\partial T}\right|_V = \left.\frac{\partial E}{\partial T}\right|_V, \qquad \left.\frac{\partial S}{\partial V}\right|_T - p = \left.\frac{\partial E}{\partial V}\right|_T. \qquad (267)
$$

The number of such identities is confusing, but with a bit of practice it is easy to only memorise the first law (246) and derive any identities as needed.

## 13.3 *Exercises*

1. (Chandler Exercise 1.2) An equation of state for a rubber band is either

$$
S = L_0\gamma(\theta E/L_0)^{1/2} - L_0\gamma\left[\frac{1}{2}\left(\frac{L}{L_0}\right)^2 + \frac{L}{L_0} - \frac{3}{2}\right], \quad L_0 = nl_0, \qquad (268)
$$

   or

$$
S = L_0\gamma e^{\theta nE/L_0} - L_0\gamma\left[\frac{1}{2}\left(\frac{L}{L_0}\right)^2 + \frac{L}{L_0} - \frac{3}{2}\right], \quad L_0 = nl_0, \qquad (269)
$$

   where $\gamma$, $l_0$ and $\theta$ are constants, $L$ is the length of the rubber band, $n$ the number of moles in it, and $E$ its internal energy. Which of the two possibilities is acceptable? For the acceptable choice determine the tension $f$ as a function of $T$ and $L/n$.

2. (Chandler Exercise 1.4) Consider two pieces of rubber band each obeying the (correct) equation of state from the previous question. The first piece initially has temperature $T_1$, length per mole $l_1$ and mole number $n_1$, and the second piece has $T_2$, $l_2$ and $n_2$. Assume they are each held at constant length, but are allowed to exchange heat. Calculate the final energies and temperatures.

3. (Chandler Exercise 1.12) Consider a rubber band of length $L$ held at tension $f$. The first law is

$$
dE = TdS + fdL + \mu dn. \qquad (270)
$$

   Derive the Gibbs-Duhem equation.

4. (Chandler Exercise 1.13) Suppose the equation of state for a rubber band is

$$
E = \theta S^2 L/n^2. \qquad (271)
$$

   Calculate $\mu(T, L/n)$. Then verify the Gibbs-Duhem equation for your result.

5. (Chandler Exercise 1.14) Show that for a gas with single type of molecules

$$
\left.\frac{\partial \mu}{\partial v}\right|_T = v\left.\frac{\partial p}{\partial v}\right|_T, \qquad (272)
$$

   where $v = V/n$ is the volume per mole. (You could start from the Gibbs-Duhem equation, per mole).

# 14 The second law

## 14.1 Heat flow statement of the second law

Consider two systems which are both separately in thermodynamical equilibrium, and bring them into contact. The classic textbook example is a container divided by an internal wall into two halves, each of which contains some gas, with the entire container and the dividing wall insulated so that no energy can be transmitted in the form of heat. Initially the system is described by $(E_1, V_1, n_1)$ and $(E_2, V_2, n_2)$. Remove the dividing wall and wait for a new thermodynamic equilibrium to be established. This final state will be described by some $(E, V, n)$ for the combined system. From the fact that the total volume, energy and particle number cannot have changed, we have

$$E_1 + E_2 = E, \tag{273}$$
$$V_1 + V_2 = V, \tag{274}$$
$$n_1 + n_2 = n. \tag{275}$$

The **second law of thermodynamics** for this system can be stated as

$$S_1 + S_2 \leq S, \tag{276}$$

that is, the entropy cannot decrease in the transition from one equilibrium state to another. In fact, the entropy is constant if thermal equilibrium is maintained throughout the transition, and increases otherwise. In our example, the transition is typically out of equilibrium – the gas will rush from container to the other, and this transition is not completely described by $(E, V, n)$.

This formulation of the second law allows us to recover the usual meaning of temperature. Consider the evolution of the two systems which are initially completely separate, and which are then brought into thermal contact so that energy in the form of heat can flow between them while the new equilibrium is being established. (This is different from the previous example in that the two system remain distinguishable, and no particles are exchanged). The first law for each of the subsystems, and the fact that the total energy is conserved, are then

$$dE_1 = T_1 dS_1, \tag{277}$$
$$dE_2 = T_2 dS_2, \tag{278}$$
$$dE_1 = -dE_2. \tag{279}$$

The second law says that

$$\int_\gamma (dS_1 + dS_2) = \int_\gamma \left( \frac{1}{T_1} - \frac{1}{T_2} \right) dE_1 \geq 0 \tag{280}$$

where $\gamma$ is any curve in the joint state space of systems 1 and 2 that describes a physically allowed change. We conclude that

$$\int_\gamma dE_1 > 0 \quad \Leftrightarrow \quad T_1 < T_2, \tag{281}$$

so that energy (in the form of heat) flows from the hotter to the colder system. Conversely, in equilibrium both subsystems must have the same temperature.

We can extend the previous example to allow for an exchange of particles or of volume, writing

$$dS_1 = \frac{1}{T_1} dE_1 + \frac{p_1}{T_1} dV_1 - \frac{\mu_1}{T_1} dn_1 \tag{282}$$

and similarly for system 2. Going through the same mathematical steps, we conclude that, for $T_1 = T_2$ and $\mu_1 = \mu_2$, and with total volume fixed, the subsystem with the higher pressure expands (as one would expect intuitively), and that for $T_1 = T_2$ and $p_1 = p_2$, and with total particle number fixed, particles diffuse to the subsystem with the smaller $\mu$ (which explains the name chemical potential).

## 14.2 Variational statement of the second law

We can think of the container of gas with an internal partition as a single system characterised by $(E, V, n)$, but with an **internal constraint** which does not allow it to reach full thermodynamic equilibrium. From this point of view, the second law can be restated as

$$S(E, V, n) \geq S(E, V, n; \text{internal constraint}) \tag{283}$$

for any internal constraint whatever. (The equality applies when the internal constraint is absent, or might as well not be there.) In words: the **entropy is maximal** in equilibrium, for given $(E, v, n)$. Note that a spontaneous fluctuation (most gas molecules suddenly find themselves in the right half of the container) can also be modelled by an internal constraint.

This formulation of the second law is interesting because it makes contact with our idea, from mechanics, that some **energy is minimal** in equilibrium. Let $y$ be a variable that parameterises the internal constraint. For example, in a cylinder of gas divided by a movable piston into two halves, let $y = 0$ denote the equilibrium position it finds itself in (this will be where the pressure is the same on both sides), and let $y \neq 0$ denote any deviation from this equilibrium. In geometric terms, we consider the manifold of equilibrium states as a submanifold embedded in a larger manifold of non-equilibrium states, and defined by $y = 0$.

We Taylor-expand the total entropy $S$ around equilibrium $(E_0, V_0, n_0; y = 0)$ in all the variables to leading order:

$$S(E, V, n; y) \simeq S_0 + \frac{1}{T}(E - E_0) + \frac{p}{T}(V - V_0) - \frac{\mu}{T}(n - n_0) - by^2. \tag{284}$$

Here the leading order for all variables is the linear order, and the necessary partial derivatives are given by the first law, except for $b$: by assumption, $y = 0$ is a local maximum of the entropy, and so the leading order is generically $-by^2$ for some $b > 0$.

To this order, we can solve for $E$ and find

$$E(S, V, n; y) \simeq E_0 + T(S - S_0) - p(V - V_0) + \mu(n - n_0) + bTy^2. \tag{285}$$

As $bT > 0$, we have shown that

$$E(S, V, n) \geq E(S, V, n; \text{internal constraint}). \tag{286}$$

In equilibrium, the energy is indeed minimal for given $(S, V, n)$. Fixing the volume and particle number has a clear experimental meaning; fixing the entropy

in this context means not allowing heat flow in or out of the system. This could be done by putting the system in a thermos, or simply by making any changes rapidly enough that the heat flow through the walls of the system is negligible.

Sometimes it is more natural to control the temperature than the entropy. Consider the **free energy** $F$ defined by

$$F := E - TS. \tag{287}$$

Substituting into (285), noting that how $T = T_0$ because we keep $T$ fixed, and defining the shorthand $F_0 := E_0 - T_0 S_0$, we have

$$\begin{aligned} F &\simeq E_0 - T_0 S_0 - p(V - V_0) + \mu(n - n_0) + bTy^2 \\ &= F_0 - S(T - T_0) - p(V - V_0) + \mu(n - n_0) + bTy^2. \end{aligned} \tag{288}$$

We have shown that

$$F(T, V, n) \geq F(T, V, n, \text{internal constraint}). \tag{289}$$

So if we control temperature rather than heat flow, the relevant **thermodynamic potential** is $F$. What is going on in the second line of (288) can be written more precisely in forms notation:

$$dE = TdS - pdV + \mu dn \quad \Leftrightarrow \quad dF = -SdT - pdV + \mu dn \tag{290}$$

for a gas, or

$$dE = TdS + f_i dx^i \quad \Leftrightarrow \quad dF = -SdT + f_i dx^i \tag{291}$$

for a general system. The transformation of $E$ into $F$ is an example of a **Legendre transformation**.

If the system is a liquid in a test tube rather than a gas in a box, it may be more natural to control the pressure (say atmospheric pressure) than the volume. The relevant thermodynamic potential is the **enthalpy** $H$,

$$H := E + pV, \quad \Rightarrow dH = TdS + Vdp + \mu dn, \tag{292}$$

which is minimal at fixed $(S, p, n)$. Finally, the **Gibbs free energy** $G$

$$G := E - TS + pV, \quad \Rightarrow dG = -SdT + Vdp + \mu dn, \tag{293}$$

is minimal in equilibrium at fixed $(T, p, n)$.

## 14.3 Stability criteria

Consider a system in thermal equilibrium, arbitrarily divided into two subsystems 1 and 2. We know that at fixed $(S, V, n)$, $E$ is minimal. Consider an entropy flux $\Delta S_1 = -\Delta_2$ between the two subsystems. The resulting change $\Delta E$ of total energy to first order

$$\left.\frac{\partial E_1}{\partial S}\right|_{V,n} \Delta S_1 + \left.\frac{\partial E_2}{\partial S}\right|_{V,n} \Delta S_2 = (T_1 - T_2)\Delta S_1 \tag{294}$$

must vanish because $E$ is extremal, and we find again that $T_1 = T_2$ in equilibrium. To second order, $\Delta E$ is

$$\frac{1}{2}\left.\frac{\partial^2 E_1}{\partial S^2}\right|_{V,n}(\Delta S_1)^2 + \frac{1}{2}\left.\frac{\partial^2 E_1}{\partial S^2}\right|_{V,n}(\Delta S_2)^2 = \frac{1}{2}\left(\left.\frac{\partial T_1}{\partial S}\right|_{V,n} + \left.\frac{\partial T_2}{\partial S}\right|_{V,n}\right)(\Delta S_1)^2, \tag{295}$$

and because $E$ is minimal, this must be positive. Therefore we must have

$$\left.\frac{\partial T}{\partial S}\right|_{V,n} > 0 \tag{296}$$

for each subsystem, or for the system as a whole.

Take the first law for a gas, (249), and rewrite it in coordinates $(T, V, n)$ instead of $(S, V, n)$. This means that we have to expand $dS$, as $S$ is no longer a coordinate but just a scalar. We obtain

$$dE = T\left(\left.\frac{\partial S}{\partial T}\right|_{V,n} dT + \left.\frac{\partial S}{\partial V}\right|_{T,n} dV + \left.\frac{\partial S}{\partial n}\right|_{T,V} dn\right) - pdV + \mu dn. \tag{297}$$

From this we can in the usual way read off the partial derivatives of $E$ in coordinates $(T, V, n)$. In particular, we have

$$C_V \equiv \left.\frac{\partial E}{\partial T}\right|_{V,n} = T\left.\frac{\partial S}{\partial T}\right|_{V,n}. \tag{298}$$

Here $C_V$ is the **heat capacity at constant volume** (and constant particle number).

(An identity like this one is sometimes derived by "dividing through" a relation between differential forms in order to obtain a partial derivative, but beware: you must be clear what you are holding constant. The approach given here is safer.)

But $T > 0$, and so together with (296) we have

$$C_V > 0. \tag{299}$$

This inequality means that a system in thermal equilibrium gets hotter as heat flows into it. If that was not the case, heat could flow into it indefinitely, and the system would be unstable. This is an example of an **thermodynamical stability criterion**.

There are many more thermodynamic stability criteria. From the fact that $H$ is minimal at fixed $(S, p, n)$ we find that the **isobaric heat capacity**

$$C_p \equiv \left.\frac{\partial E}{\partial T}\right|_{p,n} > 0. \tag{300}$$

From the fact that $F$ is minimal at fixed $(T, V, n)$ we find that the **isothermal compressibility**

$$K_T \equiv -\frac{1}{V}\left.\frac{\partial V}{\partial p}\right|_{T,n} > 0, \tag{301}$$

and from the fact that $E$ is minimal at fixed $(S, V, n)$ we find that the **adiabatic compressibility**

$$K_S \equiv -\frac{1}{V}\left.\frac{\partial V}{\partial p}\right|_{S,n} > 0. \tag{302}$$

## 14.4 *Exercises*

1. (Chandler Exercise 1.15) For a single gas, construct Legendre transformations of the $S$ which are natural functions of $(1/T, V, n)$ and of $(1/T, V, \mu/T)$. Show that

$$\left.\frac{\partial E}{\partial \beta}\right|_{\beta\mu,V} = -\left.\frac{\partial E}{\partial n}\right|_{\beta,V}\left.\frac{\partial n}{\partial \beta\mu}\right|_{\beta,V}\left.\frac{\partial \beta\mu}{\partial \beta}\right|_{n,V} + \left.\frac{\partial E}{\partial \beta}\right|_{n,V}. \tag{303}$$

2. (Chandler Exercise 1.16) Suppose a rubber band has the equation of state

$$l = \theta f/T. \tag{304}$$

   Compute

$$\left.\frac{\partial c_l}{\partial l}\right|_T \tag{305}$$

   where $c_l$ is the constant length heat capacity per unit mass.

3. (Chandler Exercise 1.17) Hold a rubber band in a heat bath at $T$. Suddenly increase the tension from $f$ to $f + \Delta f$. After a new thermal equilibrium has been reached, compute $\Delta S$. Assume the equation of state of the previous exercise.

4. (Chandler Exercise 2.20)

   a) It is easily verified that a rubber band heats up when it is stretched adiabatically. Given this fact, determine whether it will expand or contract when cooled at constant tension.

   b) The same amount of heat flows into two identical rubber bands, one at constant tension and one at constant length. Which has the larger increase in temperature?

# Part IV
# Fluids

## 15 Perfect fluids and conservation laws

### 15.1 Mass conservation

Consider a fluid with mass density (mass per volume) $\rho$ and velocity vector field $v^i$. Consider a volume $V$ that is fixed in space. The mass inside this volume at any one time is $\int_v \rho\, dV$, and it changes in time as

$$\frac{d}{dt}\int_V \rho\, dV = -\int_{\partial V} \rho v^i\, dS_i, \tag{306}$$

where $\partial V$ is the surface that bounds $V$ and $\rho v^i$ is **mass flux** (mass per time and area). Using the fact that $V$ is independent of time by assumption, and Gauss' law, this is

$$\int_V \frac{\partial \rho}{\partial t}\, dV = -\int_V \nabla_i(\rho v^i)\, dV. \tag{307}$$

As this holds for any volume $V$, we must have that

$$\frac{\partial \rho}{\partial t} = -\nabla_i(\rho v^i) \tag{308}$$

These are examples of a **conservation law** in its integral and differential form. The general form of a conservation law is

$$\frac{d}{dt}\int_V u\, dV = -\int_{\partial V} F^i(u)\, dS_i, \tag{309}$$

and if the **fluxes** $F^i$ on $\partial V$ vanish, for example because $V$ is all of space, or because the boundary $\partial V$ is a box that does not allow any fluxes, then the **conserved quantity**

$$Q = \int_V u\, dV \tag{310}$$

is independent of time.

### 15.2 The convective derivative

We can write (308) as

$$\frac{d}{dt}\rho = -\rho \nabla_i v^i \tag{311}$$

where the **convective derivative** is defined as

$$\frac{d}{dt} := \frac{\partial}{\partial t} + v^i \nabla_i. \tag{312}$$

(It will be clear from the context if $d/dt$ denotes an ordinary derivative, as in (306), or the convective derivative.) The convective derivative can be thought of as a total derivative: it gives the rate of change of any quantity from the point of view of an observer moving with the fluid, taking into account that the quantity changes with time at fixed position, and that the observer moves through space, where the quantity takes different values.

## 15.3 The Euler equation

The conservation of momentum can be written as

$$\frac{d}{dt}\int_V \rho v^i \, dV = -\int_{\partial V} \left[(\rho v^i)v^j + p\gamma^{ij}\right] \, dS_j + \int_V f^i \, dV. \tag{313}$$

Here $\rho v^i$ is the **momentum density** and $\rho v^i v^j$ is the **momentum flux**: compare the mass density and mass flux. $p$ is the **pressure**. It exerts the force $p\gamma^{ij} n_j |dS|$ on a surface element $dS_j \equiv n_j |dS|$. $f^i$ denotes all other forces, for example gravity or an electromagnetic force. The differential form of this conservation law is

$$\frac{\partial}{\partial t}(\rho v^i) = -\nabla_j \left(\rho v^i v^j + p\gamma^{ij}\right) + f^i. \tag{314}$$

Expanding and using $\nabla_i \gamma^{jk} = 0$, and using the overdot to denote $\partial/\partial t$, we have

$$\rho \dot{v}^i + \dot{\rho} v^i = -v^i v^j \nabla_j \rho - \rho v^i \nabla_j v^i - \rho v^i \nabla_j v^j - \nabla^i p + f^i. \tag{315}$$

Removing $\dot{\rho}$ using mass conservation, and putting together a convective derivative, we find

$$\rho \frac{d}{dt} v^i = -\nabla^i p + f^i \tag{316}$$

This is seen to be Newton's first law, mass times acceleration = force.

The Euler equation and mass conservation give us equations for $\dot{\rho}$ and $\dot{v}$. We also need an **equation of state**, which in its simplest form is $p = p(\rho)$. We can then specify the initial state of the fluid in terms of the density $\rho$ and velocity $v^i$ at each point in space, at $t = 0$, and use these equations to evolve the state in time.

## 15.4 Energy conservation and the first law of thermodynamics

The energy conservation law in integral form is

$$\frac{d}{dt}\int_V \left(\frac{1}{2}\rho v^2 + \rho u\right) dV = -\int_{\partial V} \left[\left(\frac{1}{2}\rho v^2 + \rho u\right) v^i + p v^i\right] \, dS_i + \int_V f^i v_i dV. \tag{317}$$

Here $\rho v^2/2$ is the density (per unit volume) of kinetic energy, and $\rho u$ is the density of internal energy of the fluid, with the definition that $u$ is the **internal energy per unit mass**. The first term in the square brackets is the passive flux of total energy as it is advected along by the fluid. The second term in the square brackets is work done by the fluid against the pressure force. The volume integral on the right-hand side is the work done by the other forces.

Now, in Newtonian mechanics, energy conservation follows from the equations of motion, and this is true here as well. The differential form of (317) can be shown to be an identity using the differential forms of mass conservation and the Euler equation. The details will be left as an exercise, except for one detail that we need to look at here: the internal energy $u$ can be changed by mechanical work done when the fluid is compressed, and so we will need a version of the first law of thermodynamics.

Consider a portion of the fluid (of fixed mass $m$) as it moves along. We assume that each fluid element on its own is in thermodynamical equilibrium. If the fluid element is chosen too small (say 100 molecules) it does not constitute a thermodynamical system. If it is chosen too large (say 10 cm, when considering flow in a jet engine), it will not have uniform density, pressure, etc. However, there is often a large range of scales where the approximation of a fluid element in thermal equilibrium applies. The first law for a fluid is

$$dE = TdS - pdV + \sum_{\alpha=1}^{n} \mu_\alpha dn_\alpha \tag{318}$$

where $n_\alpha$ are the mole numbers of different chemical components. We now define

$$s \equiv \frac{S}{m}, \quad \rho \equiv \frac{m}{V}, \quad x_\alpha \equiv \frac{n_\alpha}{m}, \quad u \equiv \frac{E}{m}, \tag{319}$$

where $s$ is the **entropy per mass**, $\rho$ is the mass density, $x_\alpha$ is the **mole fraction** of particle species $\alpha$, and $u$ is internal energy per mass. The first law is then

$$du = Tds + \frac{p}{\rho^2} d\rho + \sum_{\alpha=1}^{n} \mu_\alpha dx_\alpha. \tag{320}$$

Now as the fluid element evolves along a curve in space, its state moves along a curve in the state manifold. Therefore

$$\frac{du}{dt} = T\frac{ds}{dt} + \frac{p}{\rho^2}\frac{d\rho}{dt} + \sum_{\alpha=1}^{n} \mu_\alpha \frac{dx_\alpha}{dt}, \tag{321}$$

where $d/dt$ is the convective derivative.

Neglecting diffusion and chemical reactions, the mix of ingredients in the fluid element will not change, and so

$$\frac{d}{dt} x_\alpha = 0. \tag{322}$$

Similarly, if we can neglect heat conduction, chemical reactions (which produce or absorb heat) and shocks (see Sec. 18.3 below), the entropy is constant in the fluid element, or

$$\frac{d}{dt} s = 0. \tag{323}$$

These two approximations, together with the approximation of local thermal equilibrium, are called the **perfect fluid approximation**. Imposing them leaves us with

$$\frac{du}{dt} = \frac{p}{\rho^2}\frac{d\rho}{dt}, \tag{324}$$

which is needed when deriving the energy conservation law from mass conservation and the Euler equation.

## 15.5 *Exercises*

1. Write the energy conservation law (317) in its differential form, and show that it reduces to an identity if the mass conservation law, the Euler equation, and the first law of thermodynamics are substituted.

# 16 Vorticity and Bernoulli's principle

## 16.1 Cross product and curl

We define the inner product, cross product and curl in 3 dimensions as

$$\vec{u}\cdot\vec{v} = u_a v^a \quad (\vec{u}\times\vec{v})^a = \epsilon^a{}_{bc} u^a v^b, \quad (\vec{\nabla}\times\vec{v})^a = \epsilon^a{}_{bc}\nabla^b v^c, \tag{325}$$

Here the indices are implicitly raised or lowered with the 3-dimensional metric $\gamma_{ab}$ and its inverse $\gamma^{ab}$ and $\nabla_a$ is the metric-covariant derivative obeying $\nabla_a\gamma_{bc} = 0$. $\epsilon_{abc}$ is the volume form compatible with $\nabla_a$ and $\gamma_{ab}$. Using the contraction of two $\epsilon$ tensors, one can derive

$$\vec{u}\times(\vec{v}\times\vec{w}) = (\vec{u}\cdot\vec{w})\vec{v} - (\vec{u}\cdot\vec{v})\vec{w}. \tag{326}$$

Note the order of the two cross products matters. Replacing $\vec{u}$ with $\vec{\nabla}$, keeping in mind that $\vec{\nabla}$ acts only on $\vec{w}$ we find that

$$[\vec{u}\times(\vec{\nabla}\times\vec{w})]^a = u^b\nabla^a w_b - u^b\nabla_b w^a. \tag{327}$$

Replacing both $\vec{u}$ and $\vec{w}$ with $v$ in this last expression we obtain

$$\vec{v}\times(\vec{\nabla}\times\vec{v}) = \frac{1}{2}\vec{\nabla}v^2 - \vec{v}\cdot\vec{\nabla}\vec{v}. \tag{328}$$

These definitions and equations are all valid in arbitrary coordinates, and do not require $\gamma_{ab}$ to be flat. However, in Cartesian coordinates $x^i$, the flat metric $\gamma_{ij} = \delta_{ij}$, $\nabla_i = \partial/\partial x^i$, and $\epsilon_{123} = 1$. In the remainder of this section we use index-free notation as it is easier to read.

## 16.2 Bernoulli's principle

The Euler equation in index-free notation is

$$\dot{\vec{v}} + \left(\vec{v}\cdot\vec{\nabla}\right)\vec{v} = -\frac{1}{\rho}\vec{\nabla}p - \vec{\nabla}\phi \tag{329}$$

where $\phi$ is the gravitational potential and $-\vec{\nabla}\phi$ is the gravitational field (force/mass). If we define the **vorticity**

$$\omega \equiv \vec{\nabla}\times\vec{v}, \tag{330}$$

we can rewrite the Euler equation, using (328), as

$$\dot{\vec{v}} + \omega\times\vec{v} + \vec{\nabla}\left(\frac{1}{2}v^2 + \phi\right) + \frac{1}{\rho}\vec{\nabla}p = 0. \tag{331}$$

This equation has several useful applications.

**Steady flow** This means that $\dot{\vec{v}} = 0$. If we take the dot product of (331) with $\vec{v}$, and let $\dot{\vec{v}} = 0$, we obtain

$$\vec{v}\cdot\vec{\nabla}\left(\frac{1}{2}v^2 + \phi\right) + \frac{1}{\rho}\vec{v}\cdot\vec{\nabla}p = 0. \tag{332}$$

Recall that the enthalpy $H = E + PV$, and define the **enthalpy per mass**

$$h \equiv \frac{H}{m} = u + \frac{P}{\rho} \tag{333}$$

From (321),

$$dh = Tds + \frac{1}{\rho}dp + \sum_{\alpha=1}^{n} \mu_\alpha dx_\alpha. \tag{334}$$

Considering the change as a fluid element moves, in the ideal fluid approximation, this gives

$$\frac{dh}{dt} = \frac{1}{\rho}\frac{dp}{dt}. \tag{335}$$

In steady flow the convective derivatives reduce to

$$\vec{v} \cdot \vec{\nabla} h = \frac{1}{\rho}\vec{v} \cdot \vec{\nabla} p, \tag{336}$$

which gives us

$$\vec{v} \cdot \vec{\nabla} \left( \frac{1}{2}v^2 + \phi + h \right) = 0. \tag{337}$$

This is easily integrated to

$$\frac{1}{2}v^2 + \phi + h = \text{constant along streamlines}, \tag{338}$$

the **Bernoulli streamline theorem**.

**Irrotational flow of an isentropic fluid** This means that $\omega = 0$ and $s$ the same everywhere. We also assume that $x_\alpha$ are everywhere the same.

$\vec{\nabla} \times \vec{v} = 0$ implies that there exists locally a **velocity potential** $\psi$ with $\vec{v} = \vec{\nabla}\psi$. (In forms notation, $dv = 0$ implies $v = d\psi$ locally.) Instead of (331) we have

$$\vec{\nabla}\dot{\psi} + \vec{\nabla} \left( \frac{1}{2}v^2 + \phi \right) + \frac{1}{\rho}\vec{\nabla} p = 0. \tag{339}$$

Now we interpret (334) not as a change with time along a fluid worldline but as the change with space, that is, a gradient:

$$\vec{\nabla} h = T\vec{\nabla} s + \frac{1}{\rho}\vec{\nabla} p + \sum_{\alpha=1}^{n} \mu_\alpha \vec{\nabla} x_\alpha. \tag{340}$$

With $\vec{\nabla} s = \vec{\nabla} x_\alpha$ by assumption, we have

$$\vec{\nabla} h = \frac{1}{\rho}\vec{\nabla} p \tag{341}$$

and so

$$\vec{\nabla} \left( \dot{\psi} + \frac{1}{2}v^2 + \phi + h \right) = 0, \tag{342}$$

which integrates to

$$\dot{\psi} + \frac{1}{2}v^2 + \phi + h = \text{a function of } t \text{ only}. \tag{343}$$

**Irrotational steady flow of an isentropic fluid** This is particularly useful if the flow is *steady* as well as irrotational and isentropic, as we then have

$$\frac{1}{2}v^2 + \phi + h = \text{constant everywhere.} \tag{344}$$

A classic application is the **Pitot tube**, which is a tube sticking straight ahead out of the nose of an aircraft, such that it measures and compares two pressures: at the tip of the tube the air does not move relative to the tube, while at the side of the tube the air moves by with the speed of the airplane. Given that $\phi$ is the same in both places and $\rho$ is approximately constant in this flow, one finds the velocity from the pressure difference as

$$\frac{1}{2}v^2 = \Delta h = \frac{\Delta p}{\rho}. \tag{345}$$

## 16.3 The vorticity equation

We start again from (331), and use (340) to eliminate $(1/\rho)\vec{\nabla}p$:

$$\dot{\vec{v}} + \vec{\omega} \times \vec{v} + \vec{\nabla}\left(\frac{1}{2}v^2 + \phi + h\right) - T\vec{\nabla}s + \sum_\alpha \mu_\alpha \vec{\nabla}x_\alpha = 0. \tag{346}$$

Taking the curl of this equation we have

$$\dot{\vec{\omega}} + \vec{\nabla} \times (\vec{\omega} \times \vec{v}) - \vec{\nabla}T \times \vec{\nabla}s + \sum_\alpha \vec{\nabla}\mu_\alpha \times \vec{\nabla}x_\alpha = 0. \tag{347}$$

One can often assume that $s$ and $x_\alpha$ are constant everywhere. For example, in the flow round an aircraft $s$ and $x_\alpha$ are constant in the air in front of the aircraft, and they do not change in the flow either (when there are no shocks). In this case we have the **vorticity equation**

$$\dot{\vec{\omega}} + \vec{\nabla} \times (\vec{\omega} \times \vec{v}) = 0. \tag{348}$$

We can expand $\vec{\nabla} \times (\vec{\omega} \times \vec{v})$ using (326), keeping in mind that $\vec{\nabla}$ now acts on both $\vec{\omega}$ and $\vec{v}$ with the product rule. We therefore get four terms, one of which is $-\vec{v}\left(\vec{\nabla} \cdot \vec{\omega}\right)$, which vanishes identically. The remaining terms are

$$\dot{\vec{\omega}} + \vec{\omega}\left(\vec{\nabla} \cdot \vec{v}\right) + \left(\vec{v} \cdot \vec{\nabla}\right)\vec{\omega} - \left(\vec{\omega} \cdot \vec{\nabla}\right)\vec{v} = 0. \tag{349}$$

Using the mass conservation equation $\vec{\nabla} \cdot \vec{v} = -(1/\rho)d\rho/dt$, we can write this as

$$\frac{d}{dt}\vec{\omega} - \frac{1}{\rho}\frac{d}{dt}\rho - \left(\vec{\omega} \cdot \vec{\nabla}\right)\vec{v} = 0, \tag{350}$$

or, dividing by $\rho$,

$$\frac{\partial}{\partial t}\left(\frac{\vec{\omega}}{\rho}\right) + \left(\vec{v} \cdot \vec{\nabla}\right)\left(\frac{\vec{\omega}}{\rho}\right) - \left(\frac{\vec{\omega}}{\rho} \cdot \vec{\nabla}\right)\vec{v} = 0. \tag{351}$$

In the last two terms we spot a Lie derivative, and so we can write

$$\left(\frac{\partial}{\partial t} + \mathcal{L}_{\vec{v}}\right)\left(\frac{\vec{\omega}}{\rho}\right) = 0. \tag{352}$$

# 17 Weak solutions of conservation laws

## 17.1 Burgers' equation

Solutions to the perfect fluid equations beginning from initial data which are smooth (infinitely often differentiable) can develop diverging gradients and discontinuities in finite time. Once $\vec{\nabla}\rho$ (and hence $\vec{\nabla}p$) or $\vec{\nabla}v$ are no longer defined, the solution can no longer be evolved using the differential form of the equations. Nevertheless, such solutions are physical, and can be evolved further as **weak solutions** using the integral form of the equations.

A key nonlinearity in the fluid equations that leads to discontinuities is

$$\dot{v} + v \cdot \vec{\nabla} v = \dots . \tag{353}$$

Its effects can be more easily studied in **Burgers' equation**,

$$\dot{u} + uu' = 0, \tag{354}$$

a toy model in only one variable and one spatial dimension for the Euler equation. (Here $' = \partial/\partial x$.)

This equation can be solved graphically using a simple application of the **method of characteristics**. Introduce notation $x^i := (t, x)$ and $\nabla_i u := (\dot{u}, u')$. From (354) we can also write $\nabla_i u = (-u, 1)u'$. Consider a curve $x^i(s)$. For $u$ to be constant along that curve,

$$\frac{du}{ds} = \frac{dx^i}{ds}\nabla_i u = 0 \tag{355}$$

we must have

$$\frac{dx^i}{ds} \propto (1, u). \tag{356}$$

Such a curve is called a **characteristic.** If we choose $dx^i/ds = (1, u)$, and taking into account that by construction $u = u_0$ is constant on the curve, we can write the curve as

$$x(t) = x_0 + u_0 t \tag{357}$$

where $u(0, x_0) = u_0$.

Such a graphical solution for the case where $u(0, x)$ is a smooth monotonously increasing function of $x$ is given in Fig. 7, with $u_0 \to u_L$ as $x \to -\infty$, and $u_0 \to u_R$ as $x \to \infty$. This solution remains regular for all $t > 0$, and in fact the transition region between $u_L$ and $u_R$ becomes wider with time. Intuitively, a wave-like solution that moves into a region of higher propagation speed (here, larger $u$) becomes smoothed out because the front of the wave moves faster than the back.

Fig. 8 shows a graphical solution where $u_0(x)$ is a monotonously decreasing function. In this case characteristics from the transition region begin to intersect at some finite $t = t_1$. Intuitively, the back of the wave catches up with the front, until the wave breaks. At $t_1$ the transition region has steepened to an infinite gradient at one point. After $t_1$, the smooth solution obtained by the method of characteristics is formally multivalued, and in fact meaningless.

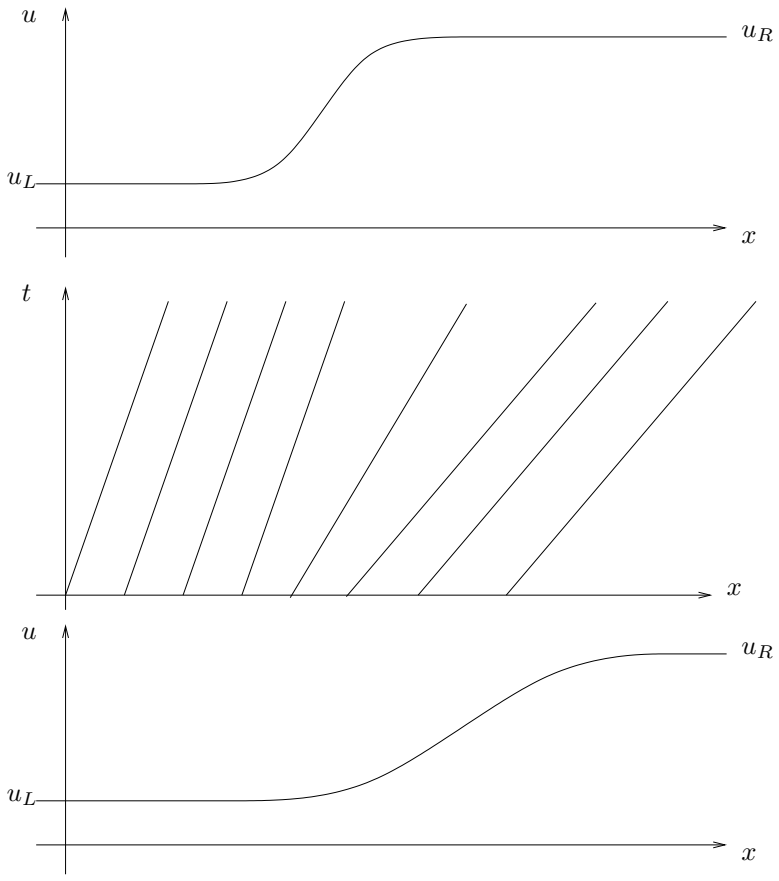


Figure 7: A wave in Burger's equation moving from low speed to high speed and becoming smoother. Shown are $u(x)$ at $t = 0$, the characteristics in the $xt$-plane, and $u(x)$ at a later time.

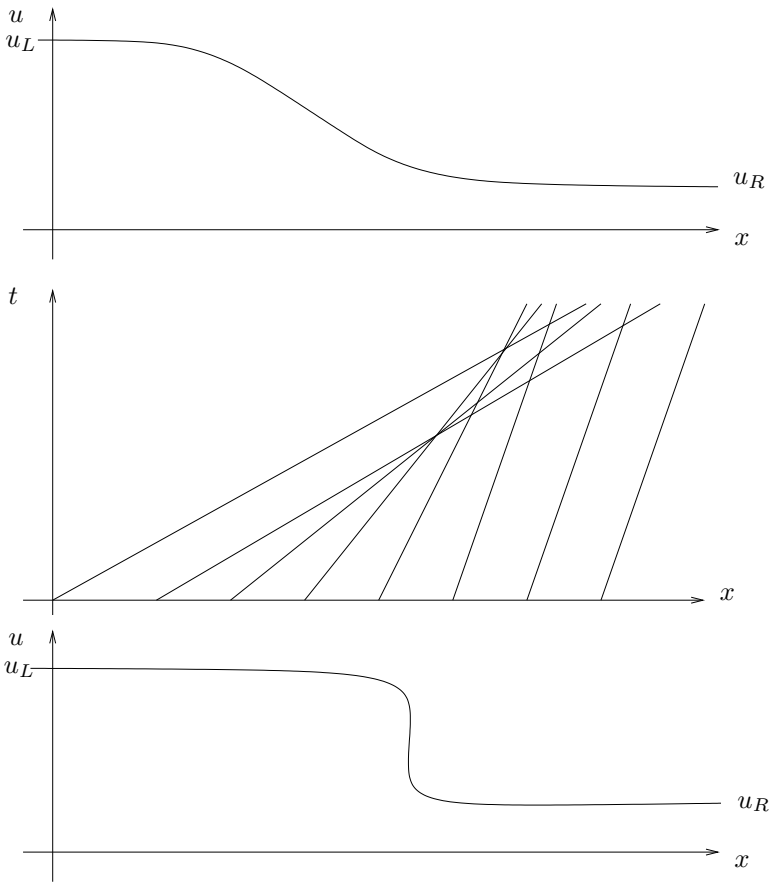


Figure 8: A wave moving from high into low speed and steepening to form a shock. Shown are $u(x)$ at $t = 0$, the characteristics in the $xt$-plane, and $u(x)$ at the moment when the characteristics begin to cross and the shock forms.

## 17.2 Riemann problem: shocks

To see what really happens, consider the **Riemann problem** for Burgers' equation. This is the particular set of initial data where $u = u_L$ for $-\infty < x < 0$, the "left state", and $u = u_R$ for $0 < x < \infty$, the "right state", with a discontinuity at $x = 0$. The reason for simplifying to constant left and right states is that compared to the discontinuity, finite slopes give rise to small effects. Therefore the Riemann problem captures the essential behaviour of a generic solution in a neighbourhood of a discontinuity.

The differential form of the equations, here $\dot{u} + uu' = 0$, does not have a solution with these initial data, as $u'$ is not defined at $x = 0$. We can however obtain a **weak solution** by writing the PDE as the **conservation law**

$$\dot{u} = \left(-\frac{1}{2}u^2\right)' \tag{358}$$

and using this in its integral form. The fixed volume $V$ in this 1-dimensional example is an interval $a < x < b$ that spans the discontinuity at $x = 0$. We have

$$\int_a^b \dot{u}\, dx = \int_a^b \left(-\frac{1}{2}u^2\right)' dx \tag{359}$$

and hence

$$\frac{d}{dt}\int_a^b u\, dx = \left[-\frac{1}{2}u^2\right]_a^b. \tag{360}$$

Note that these integrals are well defined even when $u'$ is not defined and $u$ is discontinuous. It is clear that sufficiently far to the left $u = u_L$ and sufficiently far to the right, $u = u_R$. We therefore try a solution of the form

$$u(x,t) = \begin{cases} u_L, & x < x_*(t) \\ u_R, & x > x_*(t) \end{cases} \tag{361}$$

We then have

$$\int_a^b u(x,t)\, dx = [x_*(t) - a]u_L + [b - x_*(t)]u_R, \tag{362}$$

and so

$$\frac{d}{dt}\int_a^b u\, dx = \frac{dx_*}{dt}(u_L - u_R) = \left[-\frac{1}{2}u^2\right]_a^b = -\frac{1}{2}(u_R^2 - u_L^2), \tag{363}$$

which gives

$$\frac{dx_*}{dt} = \frac{1}{2}(u_L + u_R) \tag{364}$$

This solution for $u_L > u_R$ is graphically represented in Fig. 9. The discontinuity at $x = x_*(t)$ is called a **shock**. Note that characteristics go into it from both sides.

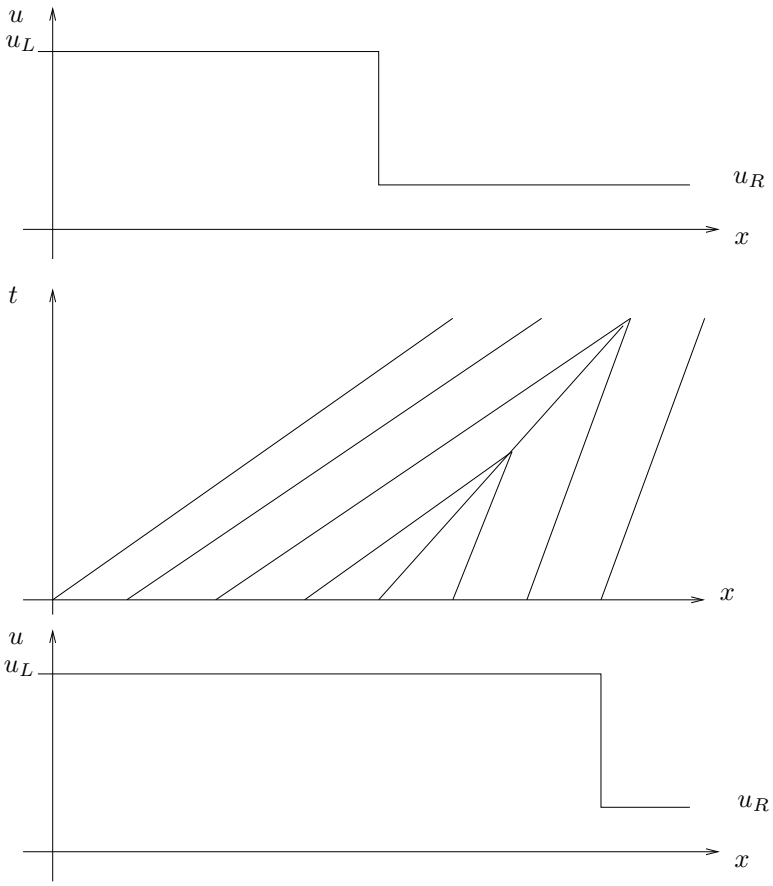


Figure 9: The shock solution to the Riemann problem for Burgers' equation with $u_L > u_R$.

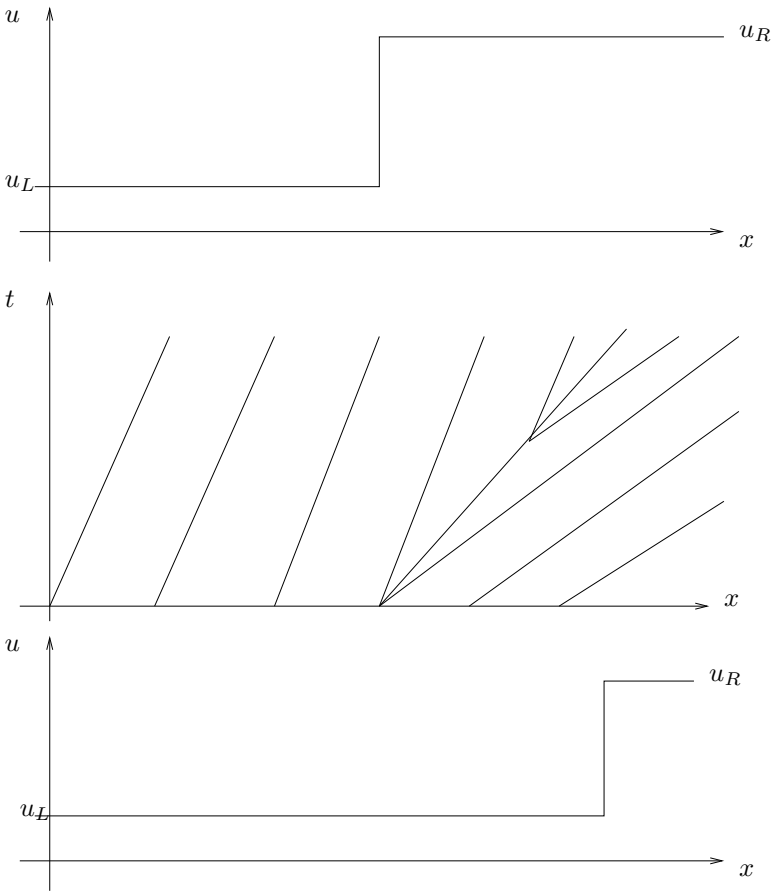


Figure 10: The shock solution to the Riemann problem for Burgers' equation with $u_L < u_R$. This solution is *not* physical. Note the characteristics emerging from the shock.

## 17.3 Rarefaction waves

Consider now the Riemann problem for Burgers' equations for $u_L > u_R$. A shock solution exists in this case, and is illustrated in Fig. 10. Now, however, characteristics are coming out of the shock. It is true, but we do not have the time to prove, that this always indicates that the shock is *unstable* against small perturbations, and therefore this solution is not physical. As a handwaving hint, note that small perturbations travel along characteristics (again, we do not have the time to prove this), which is harmless if they get annihilated in the shock, but dangerous if they come out of the shock.

Instead we try a **self-similar ansatz**

$$u(x,t) = \phi(z), \qquad z := \frac{x}{t} \tag{365}$$

Substituting gives

$$\dot{u} + uu' = \frac{1}{t}\phi'(z)\left(\phi(z) - z\right) = 0. \tag{366}$$

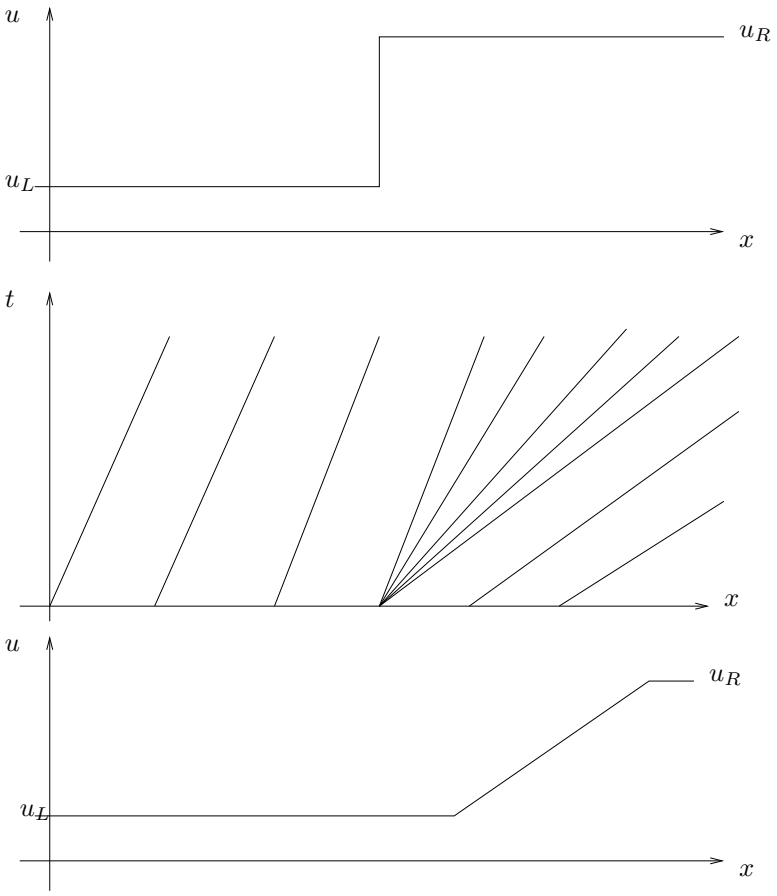


Figure 11: The rarefaction fan solution to the Riemann problem for Burgers' equation with $u_L < u_R$. This is the physical solution.

We now have to solve an ODE with boundary conditions,

$$\begin{aligned} \phi'(\phi - z) &= 0, && (367)\\ \phi(-\infty) &= u_L, && (368)\\ \phi(-\infty) &= u_L. && (369) \end{aligned}$$

The solution is

$$u(x,t) = \begin{cases} u_L, & z < u_L \\ z, & u_L < z < u_R \\ u_R, & z > u_R \end{cases} \qquad (370)$$

It is called a **rarefaction wave** (or rarefaction fan) and is illustrated in Fig. 11.

## 17.4 *Exercises*

1. Note that for $t > 0$ the rarefaction wave (370) is continuous, although not differentiable at the edges of the fan. We have therefore cheated slightly by solving the differential equation. Show that (370) is a solution of the

conservation law, and show that for the general 1-dimensional conservation law $\dot{u} + F(u)' = 0$ the fan obeys $F'(\phi) - z = 0$.

2. Explain, in words or by introducing suitable notation, why the solution to a general initial value problem with a discontinuity is approximated well by the solution to a Riemann problem in a sufficiently small neighbourhood of the discontinuity.

3. Explain why the lower order term $s$ in $\dot{u} + F(u)' = s(u)$ can be neglected in a sufficiently small neighbourhood of the discontinuity.

# 18 Weak solutions of the perfect fluid equations

## 18.1 Non-uniqueness of conservation law form

Introduce an arbitrary invertible function $f(u)$, and define $F(u)$ by

$$\frac{dF}{du} = u\frac{df}{du}. \tag{371}$$

With these, we can rewrite Burgers' equation as

$$\frac{df}{du}\dot{u} + \frac{dF}{du}u' = 0, \tag{372}$$

which gives

$$(f(u))^{\cdot} + (F(u))' = 0, \tag{373}$$

or in integral form

$$\frac{d}{dt}\int_a^b f(u)\,dx = -\left[F(u)\right]_a^b. \tag{374}$$

As long as $u'$ is defined everywhere, (373) is equivalent to (354), but the weak solutions depend on the choice of $f(u)$ because what is conserved now is the total amount of $f(u)$, not of $u$.

The calculation of the shock speed now gives

$$\frac{d}{dt}\left\{[x_*(t)-a]f(u_L) + [b-x_*(t)]f(u_R)\right\} = -\left[F(u_R)-F(u_L)\right], \tag{375}$$

and so

$$\frac{dx_*}{dt} = \frac{F(u_R)-F(u_L)}{f(u_R)-f(u_L)}, \tag{376}$$

which is the **Rankine-Hugoniot condition** for the **shock speed** $dx_*/dt$ for this Riemann problem.

We conclude that although all differential forms of an evolution PDE system are equivalent for differentiable solutions, different integral conservation law forms are inequivalent, depending on the choice of conserved variable $f(u)$.

## 18.2 More general Riemann shock problems

In fact, if we allow $f(u)$ and $F(u)$ in (374) to be independent of each other, we have the most general conservation law in one variable $u$ and in one space dimension $x$. $u$ is called the **primitive variable**, $f(u)$ the **conserved variable**, and $F(u)$ the **flux** of $f(u)$. The calculation of the shock speed is unchanged and gives (376). It is customary to redefine $u$ so that $f(u) \equiv u$, and write

$$\dot{u} + (F(u))' = 0. \tag{377}$$

If instead of being a single variable $u$ represents a vector of $n$ variables, but we are still in one space dimension, then the RH condition is again

$$\frac{dx_*}{dt}(u_R - u_L) = F(u_R) - F(u_L). \tag{378}$$

This is a system of $n$ independent algebraic equations which we cannot solve for a single shock speed.

However, if we make an ansatz of $m$ shocks, we must solve $mn$ RH conditions for $(m-1)n$ intermediate states as well as $m$ shock speeds. A necessary condition for this to be possible for generic left and right states is

$$nm \leq n(m-1) + m \quad \Leftrightarrow \quad m \geq n \tag{379}$$

so for $n$ variables we typically need $n$ shocks. This is necessary but not sufficient as the algebraic equations to be solved are nonlinear, and so may not have a solution.

In reality, solutions to the Riemann problem must be constructed from $n$ **waves**, which is the collective terms for shocks, rarefaction waves and contact discontinuities. Here, a **contact discontinuity** is a solution to a Riemann problem where characteristics are parallel to the discontinuity on both sides. A contact discontinuity arises for example in the Riemann problem for $\dot{u}+vu'=0$, with $v$ given.

## 18.3 Weak solutions of the perfect fluid equations

We are now ready to consider the perfect fluid equations. For simplicity we assume there is only a single component, so we do not have to include mole fractions $x_\alpha$, and that there are no external forces $f^i$. We have derived mass, momentum, and energy conservation above, and we can add an entropy conservation law that corresponds to $ds/dt = 0$:

$$\frac{d}{dt}\int_V \rho \, dV = -\int_{\partial V} \rho v^i \, dS_i, \tag{380}$$

$$\frac{d}{dt}\int_V \rho v^i \, dV = -\int_{\partial V} \left(\rho v^i v^j + p\gamma^{ij}\right) \, dS^j, \tag{381}$$

$$\frac{d}{dt}\int_V \rho \left(\frac{1}{2}v^2 + u\right) \, dV = -\int_{\partial V} \left(\frac{1}{2}\rho v^2 + h\right) v^i \, dS^i, \tag{382}$$

$$\frac{d}{dt}\int_V \rho s \, dV = -\int_{\partial V} \rho s v^i \, dS_i, \tag{383}$$

where $h = u + p/\rho$. We have seen (in an exercise) that the differential form of the energy equation is an identity when the mass, momentum and entropy equations hold in their differential forms. This means also that the four equations are compatible in their integral form as long as the solution is at least once differentiable everywhere.

Once a shock has formed, the perfect fluid approximation that every fluid element is locally in thermal equilibrium breaks down. To model the non-equilibrium collision processes would be extremely difficult, but it turns out that there is a wide class of solutions where this is not necessary. Rather, the solution can be approximated as piecewise smooth regions separated by shocks, rarefaction waves or contact discontinuities. Nevertheless it remains true that at a shock the gas is not in thermal equilibrium, and **shock heating** takes place. To model this, it is sufficient to solve a Riemann problem in a neighbourhood of the shock, based on (380-382), but dropping (383). Instead it is replaced by the second law of thermodynamics in the form of the **entropy condition**

$$\frac{d}{dt}\int_V \rho s \, dV + \int_{\partial V} \rho s v^i \, dS_i \geq 0. \tag{384}$$

It turns out that although there are 5 conserved variables, namely total mass $\int \rho \, dV$, 3 components of total momentum $\int \rho v^i \, dV$, and total energy $\int \rho(v^2/2 + u) \, dV$, the 2 velocity components parallel to the discontinuity decouple and we have to solve a Riemann problem in only $v_\perp$, $\rho$ and $u$. we have already seen that the Riemann problem is obtained from a generic discontinuity by neglecting the finite gradients in the left and right state. The motivation for this is that these gradients are negligible on the physical width (a few hundred molecules in reality, zero in our approximation) of the shock itself. Similarly, compared to the shock width, the curvature radius of the discontinuity as a 2-dimensional surface in 3-dimensional space is very large, and so the shock can locally be approximated as a plane.

We can then locally rotate and translate the coordinates so that the shock at $t = 0$ is given by the plane $x = 0$. Then the conservation of momentum in the $y$ and $z$ directions is automatic, and we only have conservation of mass, $x$-momentum, and energy. This means that we need a combination of three "waves" – shock, rarefaction wave, or contact discontinuity – to solve the Riemann problem. A **contact discontinuity** is a discontinuity in the mass density and the energy density, but with the velocity continuous and equal to the velocity of the discontinuity. As the RH conditions are a system of 9 nonlinear algebraic equations (3 waves, with 3 conservation laws at each), there are many possible solutions to the Riemann problem. Only one of them obeys (384); this is the physical solution.

The pressure is considered as a function of density and internal energy $p = p(\rho, u)$, the primitive variables are $\rho$, $u$ and $v^i$, and the conserved variables are $\rho$, $\rho v^i$, and $(v^2/2+u)\rho$. How this is done, in particular in numerical simulations, is a large research area in applied mathematics. Note that $p = p(\rho, u)$ is called a finite temperature perfect fluid equation of state, while $p = p(\rho)$ is called a **barotropic** or **cold** equation of state.

## 18.4 *Exercises*

1. Explain why, in a PDE in more than one space dimensions (say 3), any discontinuity can be approximated by a solution to the one-dimensional Riemann problem in a sufficiently small neighbourhood of the discontinuity. Hint: Approximate the discontinuity as a plane in spacetime, and align your spatial coordinate system with it. Treat derivatives in directions parallel to the discontinuity as lower-order terms.

2. Check that (383) is equivalent to $ds/dt$ for differentiable solutions.

# Part V
# General relativity

## 19 Matter and conservation laws in curved spacetime

### 19.1 The energy momentum tensor and the Einstein equations

General relativity is our current classical theory of gravity. The Newtonian idea of a gravitational force between masses is replaced by the idea that the metric of spacetime is not the flat Minkowski metric $\eta_{ab}$ but instead is a curved metric $g_{ab}$.

The **Einstein tensor** is defined from the Riemann tensor $R_{abc}{}^{d}$ of a metric $g_{ab}$ as

$$G_{ab} \equiv R_{acb}{}^{c} - \frac{1}{2} g_{ab} R_{cd}{}^{cd} \tag{385}$$

The Einstein equations are

$$G_{ab} = 8\pi G T_{ab} \tag{386}$$

where $G$ is Newton's constant, and $T_{ab}$ is called the **stress-energy tensor** and describes the matter. Both $G_{ab}$ and $T_{ab}$ are symmetric tensors. As the Riemann tensor is made from second derivatives of the metric, the Einstein equations are a set of nonlinear PDEs for the 10 metric coefficients $g_{\mu\nu}$, given $T_{\mu\nu}$.

It can be shown that the Bianchi identity (92) implies the **contracted Bianchi identity**

$$\nabla_a G^{ab} = 0. \tag{387}$$

The contracted Bianchi identity and the Einstein equations together imply that

$$\nabla_a T^{ab} = 0 \tag{388}$$

These equations will hold even in the limit where gravity can be neglected and $g_{ab} = \eta_{ab}$, and so say something about the matter, rather than gravity. They are called the **stress-energy conservation** laws.

### 19.2 The relativistic perfect fluid equations

To relate (388) to something we know, we now define the stress-energy tensor for a perfect fluid. The 3-velocity $v^i$ field is generalised to a 4-velocity field $u^a$ as in Sec. 9.1. The mass density of Newtonian physics is generalised to relativistic motion as the **rest mass density** $\rho$, that is the mass density measured by an observer that is moving with the fluid at that point in spacetime. Similarly $u$ and $p$ are the internal energy per rest mass and the pressure as seen by a comoving observer. The **total energy density in the rest frame** is

$$e := \rho(1 + u). \tag{389}$$

We also define the projector into the surfaces of simultaneity of $u^a$ as

$$\perp_a{}^{b} := g_a{}^{b} + u_a u^b, \tag{390}$$

with the properties that

$$\perp_a{}^b \perp_b{}^c = \perp_a{}^c \quad \text{and} \quad \perp_a{}^b u_b = 0. \tag{391}$$

(This is the same definition as (188) except that we use the preferred observer $u^a$ instead of the general observer $w^a$).

With these definitions, the stress-energy tensor of perfect fluid matter is

$$T^{ab} = eu^a u^b + p \perp^{ab} = (e+p)u^a u^b + pg^{ab}. \tag{392}$$

We define

$$\frac{d}{d\tau}\rho := u^a \nabla_a \rho, \qquad \frac{d}{d\tau}u^b := u^a \nabla_a u^b, \tag{393}$$

With $u^\mu = \gamma(1, v^i)$, this is the relativistic generalisation of the Newtonian convective derivative. We then have

$$\nabla_a T^{ab} = \frac{d}{d\tau}u^b(e+p) + (e+p)\frac{d}{d\tau}u^b + (e+p)u^b \nabla_a u^a + \nabla^b p = 0. \tag{394}$$

We can now split this 4-vector of equations into 3+1 equations with respect to $u^a$:

$$-u_b \nabla_a T^{ab} = \frac{d}{d\tau}e + (e+p)\nabla_a u^a = 0, \tag{395}$$

$$\perp^c{}_b \nabla_a T^{ab} = (e+p)\frac{d}{d\tau}u^c + \perp^c{}_b \nabla^b p = 0. \tag{396}$$

Note that $u_a u^a = -1$ implies that $u_a \frac{d}{d\tau}u^a = 0$, so $\frac{d}{d\tau}u^a$ is already a spatial vector. One can show that these equations reduce to the Newtonian energy and momentum conservation (Euler) equations.

The mass conservation law is *not* given by stress-energy conservation. It is

$$\nabla_a j^a = 0, \qquad j^a := \rho u^a \tag{397}$$

where $u^a$ and $\rho$ are the 4-velocity and rest mass density as before, and $j^a$ is called the **mass current**. Expanding, we have

$$\frac{d}{d\tau}\rho + \rho \nabla_a u^a = 0, \tag{398}$$

which is again reminiscent of its Newtonian counterpart.

## 19.3 Mass conservation in a curved spacetime

An interesting question to ask is if the relativistic equations of mass, energy and momentum conservation that we have just stated are in fact conservation laws. The answer is unambiguously yes only for the mass conservation law. A **conserved current** is any vector field $j^a$ that obeys

$$\nabla_a j^a = 0. \tag{399}$$

The mass current $j^a = \rho u^a$ is only one example. In coordinates, (399) is

$$j^\mu{}_{,\mu} + \Gamma^\nu{}_{\mu\nu} j^\nu = 0. \tag{400}$$

Now one can show from (97) that

$$\Gamma^{\nu}{}_{\mu\nu} = \frac{1}{2} g^{\kappa\lambda} g_{\kappa\lambda,\mu} = \frac{1}{2} \frac{|g|_{,\mu}}{|g|}, \tag{401}$$

and so

$$\frac{1}{\sqrt{|g|}} \left( \sqrt{|g|} j^{\mu} \right)_{,\mu} = 0, \tag{402}$$

or

$$\left( \sqrt{|g|} j^{0} \right)_{,0} + \left( \sqrt{|g|} j^{i} \right)_{,i} = 0. \tag{403}$$

This is a conservation law, where the conserved quantity is

$$Q = \int \sqrt{|g|} j^{0} \, dx^{1} \, dx^{2} \, dx^{3}, \tag{404}$$

evaluated on any surface of constant $t$.

We can get a more geometric picture by using Gauss' theorem (132) and using a spactime volume $V$ that is bounded by a **world tube** $W$, which we can think of as the coordinate surface $r = R$, and two spacelike hypersurfaces $\Sigma_0$ and $\sigma_1$, which we can think of as coordinate surfaces $t = t_0$ and $t = t_1$. This is shown in Fig. 12. Then

$$\int_{\Sigma_1} \vec{j} \cdot d\vec{S} + \int_{\Sigma_1} \vec{j} \cdot d\vec{S} + \int_{W} \vec{j} \cdot d\vec{S} = \int_{V_4} (\vec{\nabla} \cdot \vec{j}) \, dV = 0 \tag{405}$$

and taking into account that $dS$ points to the future on $\Sigma_1$ but to the past on $\Sigma_0$, we have that the change in the conserved $Q = \int_{\Sigma} j \cdot dS$ is given by the flux through the boundary $W$.

## 19.4 Energy and momentum conservation laws in curved spacetime

The mass conservation law is of the form

$$u_{,t} + F^{i}{}_{,i} = 0, \tag{406}$$

which is natural 3-dimensional generalisation of the 1-dimensional conservation law (373). By constrast, in general it is not possible to write (394) in coordinates as

$$(u^{\mu})_{,t} + (F^{\mu})^{i}{}_{,i} = 0. \tag{407}$$

To understand this failure geometrically, we can make a vector $P^a$ from the tensor $T^{ab}$ by contracting it with a reference vector field $k^a$, or

$$P^{a} := T^{ab} k_{b} \tag{408}$$

We immediately find that

$$\nabla_a P^a = \left( \nabla_a T^{ab} \right) k_b + T^{ab} \nabla_a k_b = T^{ab} \nabla_{(a} k_{b)}. \tag{409}$$

Now we can use $\nabla_a$ rather than a partial derivative in

$$\mathcal{L}_k g_{ab} = k^c \nabla_c g_{ab} + \nabla_a k^c g_{cb} + \nabla_b k^c g_{ac} = \nabla_a k_b + \nabla_b k_a = 2 \nabla_{(a} k_{b)}, \tag{410}$$

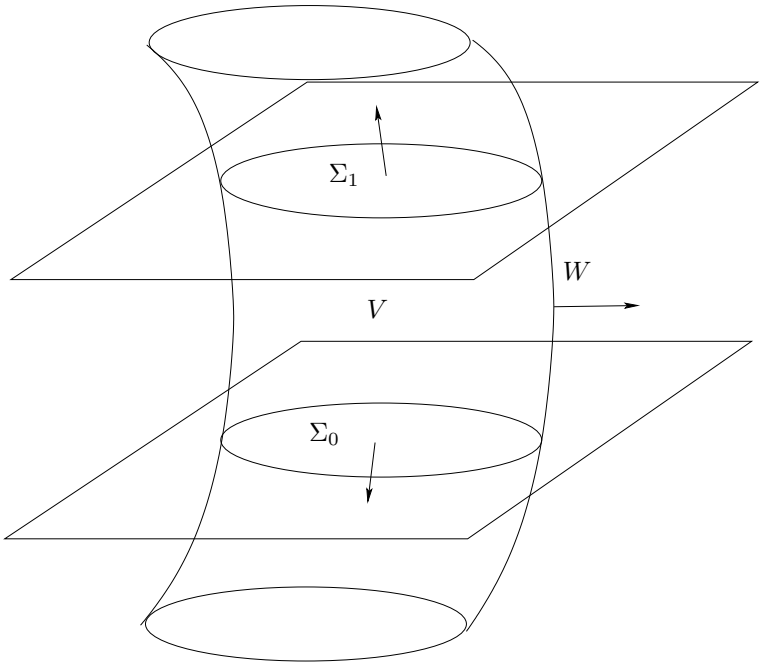


Figure 12: The 4-volume $V$ used to derive a relativistic conservation law in spacetime. The arrows denote the outward pointing normal vector, which is used in Gauss' law.

so $\nabla_{(a}k_{b)} = 0$ if and only if $k^a$ is a Killing vector of the metric $g_{ab}$. This means that we can obtain a conservation law if and only if the metric has a Killing vector.

Each Killing vector gives rise to one conservation law, and it can be shown that a 4-dimensional spacetime can have at most 10 of them. One spacetime that has this maximum number is Minkowski spacetime: In particular it has energy conservation because it is time-translation invariant, momentum conservation because of three spatial translation invariances, and angular momentum conservation because of rotation invariance. We derived these Killing vectors in Sec. 8.4.

To give a contrasting example, a spacetime that is only time-translation invariant and axisymmetric has conserved energy and angular momentum component in the direction of the symmetry axis.

## 19.5 *Exercises*

1. Show that the Newtonian energy and Euler equations are recovered from (395,396) in the limit $v \ll 1$, $u \ll 1$ and $p \ll \rho$ (or $v \ll c$, $\rho u \ll c^2\rho$ and $p \ll c^2\rho$ if we put $c^2$ back.)

# 20 The weak field limit of general relativity

## 20.1 The linearised Einstein equations

If gravity is weak, which is the case everywhere in the universe except near neutron stars and black holes, and for the universe as a whole, the curved spacetime metric $g_{ab}$ is approximately equal to the flat Minkowski metric $\eta_{ab}$, and we can introduce Minkowski-like coordinates in which the metric takes the form

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{411}$$

where $\eta_{\mu\nu}$ is given by (261) and all components of $h_{\mu\nu}$ obey $|h_{\mu\nu}| \ll 1$. The perturbation of the inverse metric is given to leading order by

$$g^{\mu\nu} \simeq \eta^{\mu\nu} - h^{\mu\nu}, \tag{412}$$

where the indices on $h$ are moved with $\eta$ here and in the following, because

$$g_{\mu\nu}g^{\nu\lambda} = \delta_\mu{}^\lambda \tag{413}$$

to all orders.

Expanding the Christoffel symbols to first order in this small **perturbation** of the metric, and taking into account that $\eta_{\mu\nu,\lambda} = 0$, we have to first order that

$$\Gamma^\lambda{}_{\alpha\beta} = \frac{1}{2}\eta^{\lambda\mu}\left(-h_{\alpha\beta,\mu} + h_{\alpha\mu,\beta} + h_{\beta\mu,\alpha}\right). \tag{414}$$

We can now construct the Riemann tensor (101) and hence the Einstein tensor (385), always to linear order in $h$. The result can be written more compactly if we define the **trace-reversed metric perturbation**

$$\bar{h}_{\alpha\beta} := h_{\alpha\beta} - \frac{1}{2}\eta_{\alpha\beta}h_\gamma{}^\gamma. \tag{415}$$

We then find that the **linearised Einstein tensor** is

$$G_{\alpha\beta} \simeq -\frac{1}{2}\bar{h}_{\alpha\beta,\gamma}{}^{,\gamma} + \bar{h}_{\gamma(\alpha,\beta)}{}^{,\gamma} - \frac{1}{2}\eta_{\alpha\beta}\bar{h}_{\gamma\delta}{}^{,\gamma\delta}. \tag{416}$$

## 20.2 Gauge freedom

The condition that $h$ is small does not fix the coordinates at all. Consider the coordinate transformation

$$\tilde{x}^\alpha := x^\alpha + \xi^\alpha, \tag{417}$$

where $\xi^\alpha$ is considered small, which leads to

$$\frac{\partial\tilde{x}^\alpha}{\partial x^\beta} = \delta^\alpha{}_\beta + \xi^\alpha{}_{,\beta} \tag{418}$$

and hence, to first order in $\xi^\alpha{}_{,\beta}$,

$$\frac{\partial x^\alpha}{\partial\tilde{x}^\beta} = \delta^\alpha{}_\beta - \xi^\alpha{}_{,\beta}. \tag{419}$$

This means that to first order in $\xi^{\alpha}{}_{,\beta}$,

$$\tilde{g}_{\alpha\beta} = \frac{\partial x^{\mu}}{\partial \tilde{x}^{\alpha}} \frac{\partial x^{\nu}}{\partial \tilde{x}^{\beta}} g_{\mu\nu} \simeq g_{\alpha\beta} - g_{\alpha\mu}\xi^{\mu}{}_{,\beta} - g_{\beta\mu}\xi^{\mu}{}_{,\alpha}. \tag{420}$$

This freedom of changing the value of the metric coefficients by changing the coordinate system, without changing the metric tensor itself, is called the **gauge freedom** of general relativity, in analogy with $\tilde{A}_{\alpha} = A_{\alpha} + \varphi_{,\alpha}$ in electromagnetism.

If we consider $h_{\alpha\beta}$ and $\xi^{\alpha}{}_{,\beta}$ as equally small, then to first order we have the **linear gauge transformation**

$$\tilde{h}_{\alpha\beta} = h_{\alpha\beta} - \xi_{\alpha,\beta} - \xi_{\beta,\alpha}. \tag{421}$$

It can be show that we can solve a PDE for $\xi^{\alpha}(x)$ so that we achieve the **Lorentz gauge**

$$\bar{h}^{,\beta}_{\alpha\beta} = 0, \tag{422}$$

in which the linearised Einstein tensor becomes

$$G_{\alpha\beta} \simeq -\frac{1}{2}\bar{h}^{,\gamma}_{\alpha\beta,\gamma}. \tag{423}$$

## 20.3 The Newtonian limit

We can now consider the **linearised Einstein equations**

$$-\frac{1}{2}\bar{h}_{\alpha\beta,\gamma}{}^{,\gamma} = 8\pi G T_{\alpha\beta}. \tag{424}$$

If we assume that $v \ll 1$ and hence $u^{\alpha} \simeq (1, v^{i})$ and $u \ll 1$ and $p \ll \rho$, as is the case for any matter less dense than a neutron star, we find that to leading order

$$T_{00} \simeq \rho, \quad T_{0i1} \simeq -\rho v_{i}, \quad T_{ij} \simeq \rho v_{i} v_{j}, \tag{425}$$

and so the linearised Einstein equations are dominated by

$$-\frac{1}{2}\bar{h}_{00,\gamma}{}^{,\gamma} = 8\pi G\rho, \tag{426}$$

with $\bar{h}_{0i}$ and $\bar{h}_{ij}$ much smaller than $\bar{h}_{00}$. Furthermore, as we assume that the sources move slowly, the solution of the inhomogeneous wave equation can be approximated by the solution of Laplace's equation, with the time derivatives neglected.

We therefore have

$$\bar{h}_{\alpha\beta} = \begin{pmatrix} -4\phi & & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{pmatrix}, \tag{427}$$

where the dimensionless **gravitational potential** $\phi$ is the solution of the **Poisson equation**

$$\Delta\phi = 4\pi G\rho. \tag{428}$$

(If we put the $c^2$ back, we have $\Delta\phi = 4\pi c^{-2} G\rho$, with $\phi$ dimensionless). Reversing the trace of $\bar{h}_{\alpha\beta}$, we obtain

$$h_{\alpha\beta} = \begin{pmatrix} -2\phi & & & \\ & 2\phi & & \\ & & 2\phi & \\ & & & 2\phi \end{pmatrix}. \tag{429}$$

To show that general relativity is a theory of gravity, we need to recover the gravitational force. For this, note that

$$\frac{d}{d\tau}u^\mu := u^\alpha \left(u^\mu{}_{,\alpha} + \Gamma^\mu{}_{\alpha\nu u^\nu}\right) \tag{430}$$

With

$$u^\alpha = \gamma(1, v^i) \simeq (1, v^i) \simeq (1, 0), \tag{431}$$

we have

$$\begin{aligned} \frac{d}{d\tau}u^i &\simeq \dot{v}^i + v^j v^i{}_{,j} + \Gamma^i{}_{00} \\ &\simeq \dot{v}^i + v^j v^i{}_{,j} + \phi^{,i} \end{aligned} \tag{432}$$

which is the Newtonian gravitational field.

Note that in general relativity the gravitational field is part of the convective derivative $u^a \nabla_a$, rather than an independent force. In fact, the exact equation of motion of a massive particle with 4-velocity $u^a$, in the absence of any non-gravitational forces is simply the **geodesic equation**

$$u^b \nabla_b u^a = 0. \tag{433}$$

A world line whose tangent vector $u^a$ obeys this equation is called a **geodesic**. A timelike geodesic can be shown to be the curve between two fixed spacetime whose proper time is longest. In this sense, it is the generalisation of a straight line to curved spacetime.

## 20.4 Gravitational waves

If we consider the linearised Einstein equations in vacuum, we have that $\bar{h}_{\alpha\beta}$ obeys the wave equation. Therefore, the Einstein equations admit solutions representing **gravitational waves**, which travel at the speed of light, like electromagnetic waves. Unfortunately, we do not have the time to pursue this very interesting topic (your lecturer's research area) further.

## 20.5 *Exercises*

1. From $G$, $c$, and the radius and mass of the sun, calculate the dimensionless gravitational potential $\phi$ just outside the sun (where it is largest), and check that the linear perturbation we have used in this section is justified. Calculate the gravitational acceleration (in $m/s^2$) just outside the sun.

# Part VI
# Appendix: Statistical mechanics

This material was removed from the lectures, for lack of time on the one hand, and because it is the least geometrical part of the course. It follows Chandler very closely.

## 21 Ensembles

### 21.1 Micro and macrostates

As we have said before, thermodynamics describes situations of "thermodynamic equilibrium", where a macroscopic system with $\sim 10^{23}$ degrees of freedom can be described effectively by a very small number of degrees of freedom $(E, X)$, where $X = \{x^i\}$ are mechanical degrees of freedom such as volume $V$ or mole number $n$. Fpr many purposes, a thermodynamical system behaves like a completely mechanical system with one extra degree of freedom, which we can think of as internal energy $E$, or entropy $S$, or temperature $T$: each of these describes an unordered internal motion of molecules that we do not need to model in detail in order to predict the macroscopic behaviour. In this section, we look at the microscopic foundation of thermodynamics.

A **microstate** is a complete specification of the system, for example the instantaneous position and momenta of all the $\sim 10^{23}$ molecules at some reference time $t = 0$. In classical mechanics there is a continuum of microstates because positions and momenta can take any real values, but an isolated quantum mechanical system of finite size has a *countable* set of quantum states. Even if we assume classical mechanics, it makes for simpler notation if assume that the microstates are discrete and can be counted off. We shall denote a particular microstate by $\nu$ and the value some macroscopic observable $O$ (such as the total energy) takes in state $\nu$ by $O_\nu$.

In thermodynamics we do not know what microstate the system is in. In fact, no real system can be isolated so perfectly that it will remain in that state. Statistical mechanics can only predict an **expectation value** $\langle O \rangle$, in the statistical sense, of an observable $O$. It will turn out that, for example, what we have called $E$ in thermodynamics is really $\langle E \rangle$, but that the difference hardly matters because the variance is tiny.

### 21.2 The microcanonical ensemble

We assume $\langle O \rangle$ is given by the usual formula in statistics as a sum over microstates

$$\langle O \rangle = \sum P_\nu O_\nu, \tag{434}$$

where $P_\nu$ is the probability to find the system in the microstate $\nu$. This is called the **statistical hypothesis** (in thermodynamics). Proving it is very subtle.

We also make the daring assumption that *all microstates for given* $(E, X)$ *are equally probable*. Normalising the total probability for given $E, X$ to one,

we have

$$P_\nu = \frac{1}{\Omega(E, X)}, \tag{435}$$

where $\Omega(E, X)$ is the **number of microstates** compatible with given $(E, X)$. Intuitively, this assumption makes sense: for example, in a gas in equilibrium in a container, a particular molecule is equally likely to be at any position at a given time.

We see that to make sense of $\Omega$, we need to have discrete, countable microstates. This is true in quantum mechanics. There, however, we face the new problem that for given $X$ (for example specifying the size of the container and the number of molecules in it) there is a discrete set $E_\nu$ of possible energies, and so $\Omega(E, X)$ is either one or zero. The answer is to define $\Omega(E, X)$ as the number of states between $E$ and $E + \Delta E$, where $\Delta E$ is some constant that is much smaller than $E$ but much larger than the distance between energy levels $E_\nu$. $\Delta E$ would always cancel out of results, and so we do not explicitly write it in the following.

Summing over all microstates that have a given $E$ and $X$, and assigning them all equal probability, defines the **microcanonical ensemble**. It allows us to calculate expectation values in thermodynamic equilibrium for a system whose energy is fixed, for example a system that is completely isolated.

## 21.3 Statistical definition of the entropy

We now define the **microcanonical entropy** $S$ as

$$S \equiv k \ln \Omega(E, X). \tag{436}$$

(The **Boltzmann constant** $k$ is just a historical relic. From the first law, the product $TS$ has dimensions of energy. The natural choice would be for $T$ to have dimension energy, and $S$ to be dimensionless, but in historic units, $kT$ has dimension energy.)

We note that $S$ is additive for two uncorrelated systems because uncorrelated probabilities multiply:

$$P_{12} = P_1 P_2 \quad \Rightarrow \quad \Omega_{12} = \Omega_1 \Omega_2 \quad \Rightarrow \quad S_{12} = S_1 + S_2. \tag{437}$$

So $S$ is an extensive quantity, as it should be. We also note that if, at fixed $E$ and $X$, we impose an internal constraint (for example, specifying how many particles are in the left half of the container), this rules out many states that would be accessible without the constraint. The microcanonical entropy is therefore maximal in the absence of internal constraints, or (283). But a quantity with these two properties must be the entropy (up to a constant factor and an additive constant).

## 21.4 Canonical ensemble

More often, one wants to consider a system whose temperature, rather than its energy, is fixed. To fix the temperature at some value $T$, we need to couple the system to a much larger **heat bath**. Heat then flows into or out of the system until the temperature of the system is that of the heat bath. In this situation,

the energy of the system is not fixed, but can fluctuate. This physical situation is represented by the **canonical ensemble**.

If we substitute the microcanonical entropy definition into the first law, we find

$$\left.\frac{\partial \ln \Omega}{\partial E}\right|_X = \frac{1}{k}\left.\frac{\partial S}{\partial E}\right|_X = \frac{1}{kT} =: \beta. \tag{438}$$

We can derive the canonical ensemble for the system alone from the *microcanonical* ensemble for the total system. Let $E_\nu$ be the energy of the system in its state $\nu$, $E_b$ the energy of the heat bath, and assume that the total energy $E_0 = E_\nu + E_b$ is fixed and cannot fluctuate. While the system is in the specific state $\nu$, the heat bath can be in any of the many different states that have total energy $E_b$. The number of these states is defined to be $\Omega_b(E_b)$. If we assume that they are all equally likely, we have

$$\begin{aligned} P_\nu &\propto \Omega_b(E_b) = \Omega_b(E_0 - E_\nu) = e^{\ln \Omega_b(E_0 - E_\nu)} \\ &\simeq e^{\ln \Omega_b(E_0) - \frac{\partial \ln \Omega_b}{\partial E} E_\nu} \propto e^{-\beta E_\nu}, \end{aligned} \tag{439}$$

and we have arrived at the famous **Boltzmann factor**. Clearly the approximation is good if $E_\nu \ll E_b$. The macrocanonical ensemble is the limit $E_b \to \infty$. If we now normalise so that $\sum P_\nu = 1$, then

$$P_\nu = Q^{-1} e^{-\beta E_\nu}, \qquad Q := \sum_\nu e^{-\beta E_\nu}. \tag{440}$$

$Q$ is called the **canonical partition function.** We can use it to express, for example, the expectation value of the energy of the system at temperature $T$:

$$\langle E \rangle = \sum P_\nu E_\nu = \frac{\sum E^\nu e^{-\beta E_\nu}}{\sum e^{-\beta E_\nu}} = -\left.\frac{\partial \ln Q}{\partial \beta}\right|_X. \tag{441}$$

We can similarly calculate the **expected fluctuation** $\delta E \equiv E - \langle E \rangle$ of $E$:

$$\begin{aligned} \langle (\delta E)^2 \rangle &= \langle (E - \langle E \rangle)^2 \rangle = \langle E^2 \rangle - 2\langle E \langle E \rangle \rangle + \langle \langle E \rangle^2 \rangle \\ &= < E^2 > - < E >^2 = \frac{\left.\frac{\partial^2 Q}{\partial \beta^2}\right|_X}{Q} - \left( \frac{\left.\frac{\partial Q}{\partial \beta}\right|_X}{Q} \right)^2 = \left.\frac{\partial^2 \ln Q}{\partial \beta^2}\right|_X \\ &= -\left.\frac{\partial \langle E \rangle}{\partial \beta}\right|_X = -\left.\frac{\partial \langle E \rangle}{\partial T}\right|_X \frac{dT}{d\beta} = kT^2 C_V \end{aligned} \tag{442}$$

Note that because both $E$ and $C_V$ are extensive quantities (do not confuse $C_V$ with $c_v$, the heat capacity per mole), $C_V \sim N$ and $< E > \sim N$ as the system increases in size, with $N$ the particle number, and so

$$\frac{\delta E}{\langle E \rangle} \sim \frac{1}{\sqrt{N}}. \tag{443}$$

In a sufficiently large system the fluctuations are so small that one can approximate $E$ by $\langle E \rangle$.

# 22 The Gibbs entropy

## 22.1 Generalised canonical ensembles

Instead of holding all the generalised positions $x^i$ (such as volume and particle number) fixed, one may want to hold one of the corresponding generalised forces (such as pressure and chemical potential) fixed instead, by connecting the system to a bath. From (436) and the first law we have

$$\left.\frac{\partial \ln \Omega}{\partial x^i}\right|_{E,Y} = -\beta f_i =: \xi_i \tag{444}$$

Here $Y$ stands for all $x^j$ with $j \neq i$. We can then generalise the calculation for the canonical ensemble to

$$P_\nu \propto \Omega_b(E - E_\nu, X - X_\nu) \simeq e^{\ln \Omega_b(E,X) - \beta E_\nu - \xi_i x^i{}_\nu}, \tag{445}$$

and so

$$P_\nu = \Xi^{-1} e^{-\beta E_\nu - \xi_i x^i_\nu}, \qquad \Xi \equiv \sum_\nu e^{-\beta E_\nu - \xi_i x^i_\nu}. \tag{446}$$

Exactly as for the canonical ensemble, we calculate

$$\langle E \rangle = \sum P_\nu E_\nu = -\left.\frac{\partial \ln \Xi}{\partial \beta}\right|_{\xi,Y}, \tag{447}$$

and similarly

$$\langle x^i \rangle = \sum P_\nu x^i_\nu = -\left.\frac{\partial \ln \Xi}{\partial \xi_i}\right|_{\beta,Y}. \tag{448}$$

For the expected fluctuation of $x^i$ we have

$$\langle (\delta x^i)^2 \rangle = \left.\frac{\partial^2 \Xi}{\partial \xi_i^2}\right|_{\beta,Y} = -\left.\frac{\partial \langle x^i \rangle}{\partial \xi_i}\right|_{\beta,Y} \tag{449}$$

An important example is the **grand canonical ensemble**, where the temperature and chemical potential are fixed by a bath, and the energy and particle number are allowed to fluctuate. In this case

$$\langle (\delta N)^2 \rangle = \frac{1}{\beta} \frac{\partial \langle N \rangle}{\partial \mu} \tag{450}$$

## 22.2 The Gibbs entropy

From (447,448) we have

$$d \ln \Xi = -\langle E \rangle d\beta - \langle x^i \rangle d\xi_i \tag{451}$$

For the canonical or any generalised canonical ensemble we define the **Gibbs entropy** as

$$\tilde{S} \equiv -k \sum P_\nu \ln P_\nu. \tag{452}$$

This gives

$$\begin{aligned} \tilde{S} &= -k \sum P_\nu (-\ln \Xi - \beta E_\nu - \xi_i x^i_\nu) \\ &= k \left( \ln \Xi + \beta \langle E \rangle + \xi_i \langle x^i \rangle \right) \end{aligned} \tag{453}$$

Taking its differential and using (451), we have

$$\begin{aligned} d\tilde{S} &= k\left(d\ln\Xi + \beta d\langle E\rangle + \langle E\rangle d\beta + \xi_i d\langle x^i\rangle + \langle x^i\rangle d\xi_i\right) \\ &= k\left(\beta d\langle E\rangle + \xi_i d\langle x^i\rangle\right) = \frac{1}{T}d\langle E\rangle - \frac{f_i}{T}d\langle x^i\rangle, \end{aligned} \tag{454}$$

or

$$d\langle E\rangle = Td\tilde{S} + f_i d\langle x^i\rangle. \tag{455}$$

Furthermore,

$$\begin{aligned} \tilde{S} &= -k\sum P_\nu \ln\frac{1}{\Omega(E_\nu, X_\nu)} \\ &= \sum_\nu P_\nu\left(-k\ln\Omega(E_\nu, X_\nu)\right) \\ &= \sum_\nu P_\nu S_\nu = \langle S\rangle. \end{aligned} \tag{456}$$

Therefore $\tilde{S}$ is the equivalent of the microcanonical $S$ in the generalised harmonic ensemble, where $E$ and $X$. and hence $S(E,X)$, are not fixed but take their thermodynamic equilibrium values.

## 22.3 A statistical derivation of the ideal gas law

We shall derive the equation of state of an ideal gas from first principles. Break a fixed volume under consideration into $m$ cells of equal size. The microstate of the system is then given by the set of numbers $\{n_i\}_{i=1}^m$, where $n_i$ is the number of particles in cell $i$. The total particle number in the volume is $N \equiv \sum n_i$. We now make a number of approximations, which define the ideal gas limit.

1. If we make the cells small enough, the $n_i = 0, 1$ to good approximation, and so we have $n_i^2 = n_i$.

2. We assume all cells are equally likely to be populated, so $\langle n_i\rangle = \langle n_1\rangle$.

3. We assume low density, so that $\langle n_i\rangle \ll 1$.

4. We assume there are no correlations between molecules, and hence between cells, so that $\langle n_i n_j\rangle = \langle n_i\rangle\langle n_j\rangle$ for $i \neq j$.

We do not assume $N$ to be fixed but allow it to fluctuate in contact with a bath of particles (outside the volume we consider). Then

$$\begin{aligned} \langle(\delta N)^2\rangle &= \langle N^2\rangle - \langle N\rangle^2 = \sum_{i,j=1}^m \left(\langle n_i n_j\rangle - \langle n_i\rangle\langle n_j\rangle\right) \\ &= \sum_i^m \left(\langle n_i^2\rangle - \langle n_i\rangle^2\right) = m\left(\langle n_1\rangle - \langle n_1\rangle^2\right) \\ &\simeq m\langle n_1\rangle = \langle N\rangle. \end{aligned} \tag{457}$$

Combining this with the expression (450) for $\langle(\delta N)^2$ in the grand canonical ensemble we have

$$\frac{1}{\beta}\left.\frac{\partial\langle N\rangle}{\partial\mu}\right|_\beta = \langle N\rangle \tag{458}$$

We define the density $\rho$ by $\langle N\rangle \equiv \rho V$, where $V$ is fixed, so, with $\beta \equiv kT$,

$$kT\left.\frac{\partial\rho}{\partial\mu}\right|_T = \rho. \tag{459}$$

The rest is a little calculation in thermodynamics. From the Gibbs-Duhem relation (253) for a gas

$$d\mu = \frac{S}{n}dT + \frac{1}{\rho}dp = \frac{S}{n}dT + \frac{1}{\rho}\left(\left.\frac{\partial p}{\partial T}\right|_{\rho} dT + \left.\frac{\partial p}{\partial \rho}\right|_{T} d\rho\right) \tag{460}$$

and so

$$\left.\frac{\partial \mu}{\partial \rho}\right|_{T} = \frac{1}{\rho}\left.\frac{\partial p}{\partial \rho}\right|_{T}. \tag{461}$$

Combining this with (459), we obtain

$$\left.\frac{\partial p}{\partial \rho}\right|_{T} = kT \quad \Rightarrow \quad p = kT\rho, \tag{462}$$

where the integration constant is fixed by the obvious requirement that the pressure vanishes when the density vanishes.

## 22.4 *Exercises*

1. Check from the definition (452) that the Gibbs entropy is extensive.