1 The roots of science

There are no exercises from this chapter.

2 An ancient theorem and a modern question

2.1. Show that if Euclid's form of the parallel postulate holds, then Playfair's conclusion of uniqueness of parallels must follow.

Suppose this were not the case, i.e. with Euclid's parallel postulate holding, and given a point p not on a line A, we can find more than one line passing through p that is parallel to A.

Consider, without loss of generality, two such lines, B and C. In order for such a situation to occur, Euclid's fifth postulate stipulates that the two lines meet the transversal spanning A to p such that the angles on one side of the transversal add up to two right angles. The angle between the transversal and A remains fixed, imposing the requirement that $\angle A + \angle B = \pi$ and $\angle A + \angle C = \pi$. By the uniqueness of right angles—Euclid's fourth postulate—these two expressions can be equated, yielding $\angle B = \angle C$. This implies our two lines, assumed to be unique, are actually the same. Thus, by contradiction, Euclid's parallel postulate implies Playfair's.

2.2. Can you see a simple reason why?

The left side of $\pi - (\alpha+\beta+\gamma) = C\Delta$ is dimensionless, so $C\Delta$ must be as well. Since Δ is a quantity of area, it has units $[\text{length}]^2$, and so C must have units $[\text{length}]^{-2}$.

The given formula for distance appears dimensionless. To give it the correct units we must multiply it by some constant. C is a purely geometrical quantity, and so $C^{-1/2}$, having units of [length], is a natural choice. This explains why the given formula looks dimensionless: we have previously normalized C to 1.

2.3. See if you can prove that, according to this formula, if A, B, and C are three successive points on a hyperbolic straight line, then the hyperbolic distances AB, etc. satisfy $AB + BC = AC$. You may assume the general property of logarithms, $\log(ab) = \log(a) + \log(b)$ as described in §§5.2, 3.

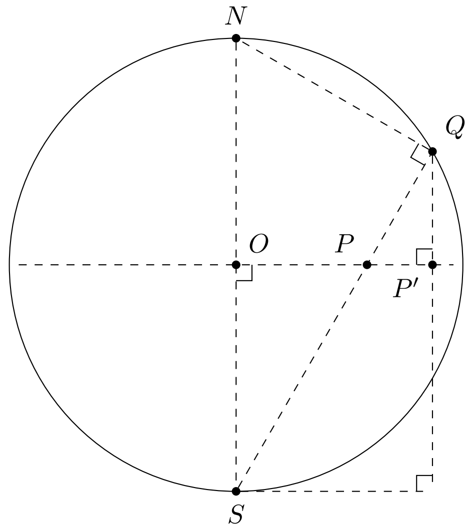
Given three points on a hyperbolic straight line, A, B, and C, we have

$$\ln\frac{QA\cdot PB}{QB\cdot PA} + \ln\frac{QB\cdot PC}{QC\cdot PB} = \ln\frac{QA\cdot PB}{QB\cdot PA}\cdot\frac{QB\cdot PC}{QC\cdot PB} = \ln\frac{QA\cdot PC}{QC\cdot PA}$$

or $AB + BC = AC$.

2.4. Show this. (*Hint*: You can use Beltrami's geometry, as illustrated in Fig. 2.17, if you wish.)

Imagine cutting the sphere of Beltrami's geometry with a plane coinciding with a conformal point P and its projective representation P'. Label other points of interest as follows.



We are interested in the ratio $\frac{OP'}{OP}$, which is equivalent to $\frac{SQ}{SP}$. From the above figure, we see that $\frac{SP}{R} = \frac{2R}{SQ}$, or $SQ = \frac{2R^2}{SP}$, which implies $\frac{OP'}{OP} = \frac{2R^2}{SP^2}$. By the Pythagorean theorem, $SP^2 = r_c^2 + R^2$, and so our scaling factor is given by $\frac{2R^2}{r_c^2+R^2}$.

2.5. Assuming these two stated properties of stereographic projection, the conformal representation of hyperbolic geometry being as stated in §2.4, show that Beltrami's hemispheric representation is conformal, with hyperbolic 'straight lines' as vertical semicircles.

Since we are projecting a conformal representation of hyperbolic geometry in a stereographic way, all angles remain the same, and circles—straight lines in hyperbolic geometry—do so as well.

To show that the projected circles are semicircles, notice that hyperbolic straight lines intersect the boundary of the disc at right angles. Therefore, by the conformal property of stereographic projection, the circles on the hemisphere must meet the hemisphere's boundary at right angles. This is only possible when such circles are semicircles.

2.6. Can you see how to prove these two properties? (*Hint*: Show, in the case of circles, that the cone of projection is intersected by two planes of exactly opposite tilt.)

2.7. See if you can do something similar, but with hyperbolic regular pentagons and squares.

2.8. Try to prove this spherical triangle formula, basically using only symmetry arguments and the fact that the total area of the sphere is $4\pi R^2$. *Hint*: Start with finding the area of a segment of a sphere bounded by two great circle arcs connecting a pair of antipodal points on the sphere; then cut and paste and use symmetry arguments. Keep Fig 2.20 in mind.

Consider the surface area bounded by two great circles connecting a pair of antipodal points. If the two circles make an angle α with each other, this area is given by $\alpha/2\pi \cdot 4\pi R^2$, or $\alpha/2 \cdot R^2$. If we only care about one hemisphere we can halve this formula: αR^2.

Look at Fig. 2.20. We can attempt to find the area of this triangle by summing the areas of hemispherical-spanning triangles with the given angles. But if we add $\alpha R^2 + \beta R^2 + \gamma R^2$, we are over-counting our triangle twice—not to mention the area extending beyond our triangle that is included in the sum. Rearrange our slices so that αR^2 is taken to be the area extending inward, while both βR^2 and γR^2 no longer include our triangle. From this, we see that the surface area we must subtract is given by πR^2, giving us an area of $\Delta = R^2(\alpha + \beta + \gamma - \pi)$.

3 Kinds of number in the physical world

3.1. Experiment with your pocket calculator (assuming you have '$\sqrt{}$' and 'x^{-1}' keys) to obtain these expansions to the accuracy available. Take $\pi = 3.141592653589793\ldots$ (*Hint*: Keep taking note of the integer part of each number, subtracting it off, and then forming the reciprocal of the remainder.)

Continued fractions can be written through the use of a simple recursive formula: subtract off the integer part of the number, take the reciprocal of the remainder, repeat. The numbers 'taken off' are the number in the continued fraction representation. In the case of π, these are 3, $(0.14159265)^{-1} = 7.06251348 \to 7$, $(0.06251348)^{-1} = 15.9965499 \to 15$, etc.

3.2. Assuming this eventual periodicity of these two continued-fraction expressions, show that the numbers they represent must e the quantities on the left. (*Hint*: Find a quadratic equation that must be satisfied by this quantity, and refer to Note 3.6.)

For any periodic continued fraction, we may represent it in terms of itself. Consider

$$1 + \cfrac{1}{2 + \frac{1}{2 + \frac{1}{2 + \cdots}}}.$$

Labeling this x, we may rewrite it as

$$x = 1 + \frac{1}{2 + (x-1)} = 1 + \frac{1}{1+x}.$$

This can be rearranged to $x + x^2 = 1 + x + 1$, or $x^2 = 2$. This has the solutions $x = \pm\sqrt{2}$. As our continued fraction is manifestly positive, we take the solution to be positive.

$7 - \sqrt{3}$ is more challenging to confirm. Focusing on the periodic portion of the continued fraction, which we label x, we see

$$x = \frac{1}{1 + \frac{1}{2+x}},$$

or $x^2 + 2x - 2 = 0$, which has the solutions $x = -1 \pm \sqrt{3}$. Again, we take the positive quantity to be the answer. When placed into the entire continued fraction, which we call y, this yields

$$y = 5 + \frac{1}{2 + \sqrt{3}},$$

or $(2+\sqrt{3})y - (11 + 5\sqrt{3}) = 0$. Solving for y gives us

$$y = \frac{11 + 5\sqrt{3}}{2 + \sqrt{3}} = (11 + 5\sqrt{3})(2 - \sqrt{3}) = 22 - 11\sqrt{3} + 10\sqrt{3} - 15 = 7 - \sqrt{3},$$

confirming the expression.

3.3. Can you see why this works?

In more compact notation, the claim given is that $(Ma > Nb) \wedge (Nd > Mc) \implies \frac{a}{b} > \frac{c}{d}$. We can put the conditions into fractional terms, with $Ma > Nb \to \frac{a}{b} > \frac{N}{M}$ and $Nd > Mc \to \frac{c}{d} < \frac{N}{M}$. By transitivity, these can be combined to become $\frac{a}{b} > \frac{c}{d}$. This can always be done because the rational numbers are dense, i.e. we can always find an M and an N such that the inequality signs we use are strict.

3.4. Can you see how to formulate these?

4 Magical complex numbers

4.1. Do this. (Alternatively, can you check it by multiplying both sides by $(c - id)$?)

Multiplying the left side by $c + id$ gives $a + ib$. Doing the same to the right reveals

$$\frac{(ac + bd + i(bc - ad))(c + id)}{c^2 + d^2} = \frac{ac^2 + ad^2 + ibc^2 + ibd^2}{c^2 + d^2} = a + ib.$$

4.2. Check this, the relevant rules being $w + z = z + w$, $w + (u + z) = (w + u) + z$, $wz = zw$, $w(uz) = (wu)z$, $w(u + z) = wu + wz$, $w + 0 = w$, $w1 = w$.

The complex field being isomorphic to the 2D plane, checking algebraic compatibility reduces to verifying the given rules independently on both axes. But each axis consists of a copy of $\mathbb{R}$, which we know satisfies these rules.

4.3. Check this.

In the case that this number is positive, we have

$$\begin{aligned}
&\Big(\sqrt{\frac{1}{2}\big(a + \sqrt{a^2 + b^2}\big)} + i\sqrt{\frac{1}{2}\big(- a + \sqrt{a^2 + b^2}\big)}\Big)^2 \\
=&\frac{1}{2}\big(a + \sqrt{a^2 + b^2}\big) - \frac{1}{2}\big(- a + \sqrt{a^2 + b^2}\big) + i\sqrt{\big(a + \sqrt{a^2 + b^2}\big)\big(- a + \sqrt{a^2 + b^2}\big)} \\
=&a + i\sqrt{-a^2 + a^2 + b^2} \\
=&a + ib.
\end{aligned}$$

The negative case is trivial, as the -1 factor squares to 1.

4.4. Can you see how to check this expression, in a formal algebraic way?

One way of finding such a solution is to label the sum as

$$S = 1 + x^2 + x^4 + x^6 + x^8 + \cdots$$

It is easy to see that $S - x^2 S = 1$, and this can be solved to find $S = (1 - x^2)^{-1}$.

4.5. Can you see an elementary reason for this simple relationship between the two series?

Returning to

$$1 + x^2 + x^4 + x^6 + x^8 + \cdots = (1 - x^2)^{-1}, ,$$

we see we can make the substitution $x \to ix$ to find

$$1 - x^2 + x^4 - x^6 + x^8 + \cdots = (1 + x^2)^{-1}.$$

4.6. Show this. (*Hint*: Show that no remainder survives if this polynomial is 'divided' by $z - b$ whenever $z = b$ solves the given equation.)

We are given that there will always exist a solution, say $z = b$, to any equation of the form

$$a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n = 0.$$

when $a_n \neq 0$. If we divide the left side by $z - b$, labeling the original sum S, the divisible polynomial P, and the remainder R, we see $S(z-b)^{-1} = P + R(z-b)^{-1}$, or $S = (z-b)P + R$. Because this quantity must equal 0 when $z = b$, $R = 0$ and we can write $S = (z-b)P$.[1] This can be repeated, with P taking the place of S, until we are left with a first-degree polynomial, at which point our process is complete.

5 Geometry of logarithms, powers, and roots

5.1. Examine the various possibilities.

In the case of addition, $w = 0$ or $w = cz$ creates a degenerate parallelogram. In the case of multiplication, $z = 0$ causes the original triangle to be dissimilar from the new one.

5.2. Do this.

If we consider w and z as vectors within $\mathbb{R}^2$ (which we may do given the isomorphic nature of $\mathbb{C}$ with $\mathbb{R}^2$), we may check the parallelogram rule by noting that $\vec{w} + \vec{z} - \vec{z} = \vec{w} - \vec{0}$, and hence the top and bottom are parallel and have the same length. Similarly, $w \vec{+} z - \vec{w} = \vec{z} - \vec{0}$, and so the sides are parallel and have the same length; $\vec{0}$, $\vec{w}$, $\vec{z}$, and $w \vec{+} z$ form a parallelogram.

To show that the triangles in Fig. 5.1b are similar, we need only show that both triangles have the same ratios between all their sides. An important fact used in doing so is that $|wz| = |w||z|$. We have

$$\frac{|wz|}{|z|} = \frac{|w||z|}{|z|} = |w|$$
$$\frac{|wz - z|}{|z|} = \frac{|z(w-1)|}{|z|} = \frac{|z||w-1|}{|z|} = |w-1|$$
$$\frac{|wz - z|}{|wz|} = \frac{|z(w-1)|}{|w||z|} = \frac{|z||w-1|}{|w||z|} = \frac{|w-1|}{|w|}$$

which shows that both triangles are similar.

5.3. Try to show this without detailed calculation, and without trigonometry. (*Hint*: This is a consequence of the 'distributive law' $w(z_1 + z_2) = wz_1 + wz_2$, which shows that the 'linear' structure of the complex plane is preserved, and $w(iz) = i(wz)$, which shows that rotation through a right angle is preserved; i.e. right angles are preserved.)

Consider multiplication of w by $z = a + ib$, $wz = w(a + ib)$. By the distributive and commutative laws, this becomes $wz = wa + i(wb)$. That is, w is stretched by a factor a and added to a rotation of itself, having been stretched by a factor b; complex multiplication consists of dilations and rotations. That the origin remains fixed is clear from the fact that $0z = 0a + i(0b) = 0$.

5.4. Spell this out.

When multiplying two complex numbers a and b, the resulting complex number c has a radial length that is the product of the radial lengths of a and b. Furthermore, the angle that c makes with respect to the real line is the sum of the angles that a and b make with respect to that same line.

5.5. Check this directly from the series. (*Hint*: The 'binomial theorem' for integer exponents asserts that the coefficient of $a^p b^q$ in $(a + b)^n$ is $n!/p!q!$.)

By multiplying out the first few terms of each power series and collecting like-powers, we find

$$
\begin{aligned}
e^a e^b &= \Big(\sum_{n=0}^{\infty} \frac{a^n}{n!}\Big)\Big(\sum_{m=0}^{\infty} \frac{b^m}{m!}\Big) \\
&= \Big(1 + a + \frac{a^2}{2} + \frac{a^3}{6} + \cdots\Big)\Big(1 + b + \frac{b^2}{2} + \frac{b^3}{6} + \cdots\Big) \\
&= 1 + (a+b) + \frac{1}{2}(a^2 + 2ab + b^2) + \frac{1}{6}(a^3 + 3a^2 b + 3ab^2 + b^3) + \cdots \\
&= \sum_{l=0}^{\infty} \frac{(a+b)^l}{l!}
\end{aligned}
$$

where the last line results from the parenthesized expressions in the previous line being the expansions of $(a+b)^l$ for various values of l.

5.6. Show from this that $z + \pi i$ is a logarithm of $-w$.

Identifying $e^{\pi i} = -1$, we may multiply both sides of $e^z = w$ to find $e^{z+\pi i} = -w$, showing that $z + \pi i$ is a logarithm of $-w$.

5.7. Check this.

Expanding the given expression gives

$$\cos(a+b) + i\sin(a+b) = (\cos a + i \sin a)(\cos b + i \sin b) = \cos a \cos b - \sin a \sin b + i(\sin a \cos b + \cos a \sin b)$$

Equating real and imaginary parts gives

$$
\begin{aligned}
\cos(a+b) &= \cos a \cos b - \sin a \sin b \\
\sin(a+b) &= \sin a \cos b + \cos a \sin b
\end{aligned}
$$

5.8. Do it.

Cubing $e^{i\theta}$ gives

$$
\begin{aligned}
(\cos\theta + i \sin\theta)^3 &= (\cos^2\theta + 2i\cos\theta\sin\theta - \sin^2\theta)(\cos\theta + i\sin\theta) \\
&= \cos^3\theta + 3i\sin\theta\cos^2\theta - 3\cos\theta\sin^2\theta - i\sin^3\theta
\end{aligned}
$$

Once again equating real and imaginary parts, we find

$$
\begin{aligned}
\cos 3\theta &= \cos^3\theta - 3\cos\theta\sin^2\theta \\
\sin 3\theta &= 3\sin\theta\cos^2\theta - \sin^3\theta
\end{aligned}
$$

5.9. Show this. How many ways? Also find all special cases.

5.10. Resolve this 'paradox': $e = e^{1+2\pi i}$, so $e = (e^{1+2\pi i})^{1+2\pi i} = e^{1+4\pi i-4\pi^2} = e^{1-4\pi^2}$.

5.11. Show this.

Consider adding $2\pi i/\log b$ to $\log_b w$, then raising b to the result. Using the change of base formula $\log_b e = \log e/\log b = 1/\log b$, the result becomes

$$b^{\log_b w+2\pi i \log_b e} = b^{\log_b w}(b^{\log_b e})^{2\pi i} = we^{2\pi i} = w.$$

Since this returns w when used to exponentiate b, it is a valid associated logarithm.

5.12. Why is this an allowable specification?

We can take $\log i = \frac{1}{2}\pi i$ for the simple reason that $e^{\frac{1}{2}\pi i} = i$.

5.13. Show why this works.

I believe Penrose meant that these other answers are arrived at by multiplying $e^{\frac{1}{2}\pi i}$ by $e^{2\pi n}$. This leaves the left hand side, given by i, untouched (as $e^{2\pi n} = 1$ for integer values of n). Raising this to a power of the form $4n + 1$ leaves it untouched for the reason that $i^{4n+1} = i^{4n}i = 1^n i = i$.

5.14. Spell this out.

The solutions to $z^n = w$ are given by $w^{1/n}$. Because n takes different depending on our chosen logarithm (i.e. $n = \log_z w$), cycling through these values gives us all possible answers.

5.15. Show this.

Taking the logarithm of both sides of the equation gives

$$\begin{aligned} \log(w^a)^b &= \log w^{ab} \\ b \log w^a &= ab \log w \end{aligned}$$

Being as we have fixed $\log w$, $\log w^a$ must equal $a \log w$ for the equation to hold.

6 Real-number calculus

6.1. Show this (ignored $x = 0$).

> Given a positive value for x, we have $\theta(x) = \frac{x+x}{2x} = 1$, while negating the input gives us $\theta(-x) = \frac{x-x}{-2x} = 0$; the formula given captures the step function.

6.2. Have a go at proving this if you have the background.

> This is not a rigorous proof, but one can see that the derivative of the non-trivial part of $h(x)$ consists of sums of terms of the form $\frac{C}{x^n}e^{-1/x}$. By itself, $\lim_{x\to 0} e^{-1/x} = 0$. When multiplied by a term of the form x^{-n}, the power series for this expression becomes
>
> $$\sum_{m=0}^{\infty} \frac{1}{m!}\Big(-\frac{1}{x^n}\Big)^m.$$
>
> That is, the power series is started at a higher power term and multiplied by some constant. But this does not change the limiting behavior of the expression, and so all terms of this form still tend to 0 as x goes to 0. This means that all derivatives of the non-trivial part of $h(x)$ maintain continuity with the trivial part, and thus $h(x)$ is C^∞. (We are assuming the smoothness of $h^{(n)}(x)$ holds away from the origin in the above.)

6.3. Show this, using rules given towards the end of §6.5.

> Consider expanding f in a power series,
>
> $$f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4 + \cdots .$$
>
> By setting $x = 0$, we see $a_0 = f(0)$. Differentiating both sides gives
>
> $$f'(x) = a_1 + 2a_2x + 3a_3x^2 + 4a_4x^3 + \cdots .$$
>
> and so $a_1 = f'(0)$. Continuing this pattern, we see $a_n = \frac{f^{(n)}(0)}{n!}$

6.4. Consider the 'one function' e^{-1/x^2}. Show that it is C^∞, but not analytic at the origin.

6.5. Using the power series of e^x given in 5.3, show that $\mathrm{d}e^x = e^x\mathrm{d}x$.

> Differentiating
>
> $$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$
>
> gives us
>
> $$\mathrm{d}e^x = \mathrm{d}x + x\mathrm{d}x + \frac{x^2}{2!}\mathrm{d}x + \frac{x^3}{3!}\mathrm{d}x = e^x\mathrm{d}x.$$

6.6. Establish this.

We may apply the Leibniz rule to x^n by splitting it into the product of x with x^{n-1},

$$\frac{\mathrm{d}(xx^{n-1})}{\mathrm{d}x} = x^{n-1} + x\frac{\mathrm{d}x^{n-1}}{\mathrm{d}x}.$$

By repeating this process (splitting x^{n-1} into xx^{n-2} and so forth), we find

$$\frac{\mathrm{d}x^n}{\mathrm{d}x} = x^{n-1} + x\Big(x^{n-2} + x\Big(x^{n-3} + x\Big(x + \cdots\Big)\Big)\Big) = nx^{n-1}.$$

6.7. Derive this.

Applying the chain rule in succession with the Leibniz rule yields

$$\mathrm{d}\big(f(x)[g(x)]^{-1}\big) = \mathrm{d}f(x)[g(x)]^{-1} + f(x)\cdot -[g(x)]^{-2}\mathrm{d}g(x) = \frac{\mathrm{d}f(x)g(x) - f(x)\mathrm{d}g(x)}{g(x)^2}.$$

6.8. Work out $\mathrm{d}y/\mathrm{d}x$ for $y = (1-x^2)^4$, $y = (1+x)/(1-x)$.

For the first, we find $\mathrm{d}y/\mathrm{d}x = 4(1-x^2)^3 \cdot -2x = -8x(1-x^2)^3$. For the second, we have $\mathrm{d}y/\mathrm{d}x = (1-x)^{-1} + (1+x)\cdot(1-x)^{-2} = \frac{2}{(1-x)^2}$.

6.9. With a constant, work out $\mathrm{d}(\log_a x)$, $\mathrm{d}(\log_x a)$, $\mathrm{d}(x^x)$.

We can work out the first two derivatives by changing the base of our logarithm. For the first, we have

$$\mathrm{d}(\log_a x) = \mathrm{d}\Big(\frac{\ln x}{\ln a}\Big) = \frac{\mathrm{d}x}{x\ln a}.$$

The second is found similarly,

$$\mathrm{d}(\log_x a) = \mathrm{d}\Big(\frac{\ln a}{\ln x}\Big) = -\frac{\ln a}{x(\ln x)^2}\mathrm{d}x = -\frac{\log_x a}{x\ln x}\mathrm{d}x.$$

The third can be derived as

$$\mathrm{d}(x^x) = \mathrm{d}(e^{x\ln x}) = e^{x\ln x}(1+\ln x)\mathrm{d}x.$$

6.10. For the first, see Exercise [6.5]; derive the second from $\mathrm{d}(e^{\log x})$; the third and fourth from $\mathrm{d}e^{ix}$, assuming that the complex quantities work like real ones; and derive the rest from the earlier ones, using $\mathrm{d}(\sin(\sin^{-1} x))$, etc., and noting that $\cos^2 x + \sin^2 x = 1$.

The calculation of the first of these is detailed in exercise 6.5. For the second, notice that $\mathrm{d}(x) = \mathrm{d}(e^{\ln x}) = \mathrm{d}x$, so by the chain rule $x\mathrm{d}\ln x = \mathrm{d}x$, i.e. $\mathrm{d}\ln x = \frac{\mathrm{d}x}{x}$.

The third and fourth can be derived simultaneously by noting that

$$\mathrm{d}e^{ix} = \mathrm{d}(\cos x + i\sin x) = i(\cos x + i\sin x)\mathrm{d}x = (-\sin x + i\cos x)\mathrm{d}x$$

. Matching the real and imaginary parts, we see $\mathrm{d}\cos x = -\sin x\mathrm{d}x$ and $\mathrm{d}\sin x = \cos x\mathrm{d}x$.

The fifth is found via

$$\mathrm{d}\tan x = \mathrm{d}\left(\frac{\sin x}{\cos x}\right) = 1 + \frac{\sin^2 x}{\cos^2 x} = 1 + \tan^2 x = \sec^2 x$$

.

The remaining three can be worked out by noting that $f(f^{-1}(x)) = x$, and so $\mathrm{d}(f(f^{-1}(x))) = f'(f^{-1}(x))(f^{-1})'(x)\mathrm{d}x = \mathrm{d}x$, implying $\mathrm{d}f^{-1}(x) = \frac{\mathrm{d}x}{f'(f^{-1}(x))}$. In the case of the sixth rule, $f(x) = \sin x$ and so $f'(x) = \cos x$. What does $\cos\sin^{-1} x$ equal? Remembering that sin can be interpreted as 'opposite' over 'hypotenuse', we see we can think of x as being the opposite side of the angle in question, setting the hypotenuse equal to 1. The adjacent side is then given by $\sqrt{1-x^2}$, and since the cos of this angle is 'adjacent' over 'hypotenuse', this is the value in the denominator of our expression for $\mathrm{d}\sin^{-1} x$. That is, $\mathrm{d}\sin^{-1} x = \frac{\mathrm{d}x}{\sqrt{1-x^2}}$.

The process to find the seventh rule is very similar, except now we must find $-\sin\cos^{-1} x$. This is the opposite of the value found above, and so $\mathrm{d}\cos^{-1} x = -\frac{\mathrm{d}x}{\sqrt{1-x^2}}$.

For the final rule, we need to find $\sec^2\tan^{-1} x$. Labelling the opposite side x and the adjacent one 1, we see $\sec\tan^{-1} x = \sqrt{1+x^2}$, and so the square of this value is $1 + x^2$. Plugging this into the denominator of our found relation gives $\mathrm{d}\tan^{-1} x = \frac{\mathrm{d}x}{1+x^2}$.

7 Complex-number calculus

7.1. Explain why $\oint z^n \mathrm{dz} = 0$ when n is an integer other than -1.

Integrating z^n when n is not -1 yields z^{n+1}. As this is a single-valued function, the contour integral of it results in $z_0^m - z_0^m$, which is 0.

7.2. Show this simply by substituting the Maclaurin series for $f(z)$ into the integral.

Doing as the question suggests yields

$$\frac{n!}{2\pi i}\oint \frac{f(0) + zf^{(1)}(0) + z^2 f^{(2)}(0)/2! + \cdots}{z^{n+1}}\mathrm{d}z.$$

Since each $f^{(n)}(0)$ is a constant, our integral becomes a linear superposition of terms of the form z^n. As explained in the answer to the previous question, these all evaluate to 0 except for the nth term, which becomes

$$\frac{n!}{2\pi i}\oint \frac{z^n f^{(n)}(0)}{z^{n+1} n!}\mathrm{d}z = \frac{f^{(n)}(0)}{2\pi i}\oint \frac{1}{z}\mathrm{d}z = f^{(n)}(0).$$

7.3. Show all this at least at the level of formal expressions; don't worry about the rigorous justification. *Hint*: Look at the origin-shifted Cauchy formula.

7.4. The function $f(z)$ is holomorphic everywhere on a closed contour Γ, and also within Γ except at a finite set of points where f has poles. Recall from §4.4 that a *pole* of order n at $z = \alpha$ occures where $f(z)$ is of the form $h(z)/(z-\alpha)^n$, where $h(z)$ is regular at α. Show that $\oint f(z)\mathrm{d}z = 2\pi i \times \{$sum of the residues at these poles$\}$, where the *residue* at the pole α is $h^{(n-1)}(\alpha)(n-1)!$.

We may divide our closed contour up into separate regions which surround the poles of our function. If we traverse each region in the opposite manner to its adjacent regions, the interior contours cancel and we are left with the original contour. This patchwork of regions can then be shrunk until it consists of separate patches surrounding each pole, being of it maintains its homology class. In this way, we may write

$$\oint_\Gamma f(z)\mathrm{d}z = \sum_k \oint_k \frac{h(z)}{(z-\alpha_k)^n}\mathrm{d}z.$$

where k denotes a different region surrounding a unique pole.

We recognize in the above a weighted Cauchy formula. Making this substitution gives us our answer

$$\oint_\Gamma f(z)\mathrm{d}z = \sum_k 2\pi i \frac{h^{(n-1)}(\alpha_k)}{(n-1)!}.$$

7.5. Show that $\int_0^\infty x^{-1}\sin x\mathrm{d}x = \frac{\pi}{2}$ by integrated $z^{-1}e^{iz}$ around a closed contour Γ consisting of two portions of the real axis, from $-R$ to $-\epsilon$ and from ϵ to R (with $R > \epsilon > 0$) and two connecting semi-circular arcs in the upper half-plane, of respective radii ϵ and R. Then let $\epsilon \to 0$ and $R \to \infty$.

Denote the small semi-circular arc by γ (traversed clockwise) and the large one by Γ (traversed anti-clockwise). Then we have

$$\oint \frac{e^{iz}}{z}\mathrm{d}z = \lim_{R\to\infty,\epsilon\to 0}\Big(\int_{-R}^{\epsilon}\frac{e^{iz}}{z}\mathrm{d}z + \int_\gamma \frac{e^{iz}}{z}\mathrm{d}z + \int_\epsilon^R \frac{e^{iz}}{z}\mathrm{d}z + \int_\Gamma \frac{e^{iz}}{z}\mathrm{d}z = 0\Big).$$

where the last equality comes from the fact that our integrand is analytic along this contour.

Focus on the last term. By Jordan's Lemma, this is 0, i.e. the very large imaginary values used in place of z completely damp the exponential. The first and third terms combine to give our sought after real-line integral. What about the third term? Rewrite it as

$$\int_\gamma \frac{e^{iz}-1}{z}\mathrm{d}z + \int_\gamma \frac{1}{z}\mathrm{d}z$$

The first term of this can be expanded in a power series as $\int_\gamma (1 + \frac{z}{2!} + \frac{z^2}{3!} + \cdots)\mathrm{d}z$. This has no poles, and so becomes 0 as ϵ shrinks to 0. The remaining part of this semi-circular contour integral is easily evaluated as

$$\int_\gamma \frac{1}{z}\mathrm{d}z = \ln \epsilon e^{i0} - \ln \epsilon e^{i\pi} = \ln\epsilon + i0 - \ln\epsilon - i\pi = -i\pi.$$

In order for our initial expansion to hold, we must have $\int_{-\infty}^{\infty} \frac{e^{iz}}{z} \mathrm{d}z = i\pi$. Taking the imaginary part of this and recognizing the resulting integrand as even, this implies

$$\int_0^\infty \frac{\sin x}{x} \mathrm{d}x = \frac{\pi}{2}.$$

7.6. Show that $1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \cdots = \frac{\pi^2}{6}$ by integrated $f(z) = z^{-2} \cot \pi z$ (see Note 5.1) around a large contour, say a square of side-length $2N + 1$ centered at the origin (N being a large integer), and then letting $N \to \infty$. (*Hint*: Use Exercise [7.4], finding the poles of $f(z)$ and their residues. Try to show why the integral of $f(z)$ around Γ approaches the limiting value 0 as $N \to \infty$.)

7.7. What is the power series, taken about the point p, for $f(z) = 1/z$?

Taking successive derivatives of $f(z) = z^{-1}$ yields $f^{(1)}(z) = -z^{-2}$, $f^{(2)}(z) = 2z^{-3}$, $f^{(3)}(z) = -3!z^{-4}$, and so on. Plugging this into the Taylor series for $f(z)$ gives

$$\sum_{k=0}^{\infty} \frac{f^{(k)}(p)}{k!}(z-p)^k = \sum_{k=0}^{\infty} \frac{(-1)^k k!}{p^{k+1} k!}(z-p)^k = \sum_{k=0}^{\infty} \frac{(-1)^k}{p^{k+1}}(z-p)^k.$$

7.8. Derive this series.

We can find this series by setting $p = 1$ in the series found in the answer to the previous question, then integrating (because $\int \frac{1}{z} \mathrm{d}z = \ln z$). Doing so gives us

$$\sum_{k=0}^{\infty} \frac{(-1)^k}{k+1}(z-1)^{k+1}.$$

8 Riemann surfaces and complex mappings

8.1. Explain why.

A finite number of sheets forming a Riemann surface is the result of a finite number of values assigned to an input of a multifunction. In the case of $z \to z^{m/n}$, consider the substitution $z = re^{i\theta}$. With this, we see that

$$z^{\frac{m}{n}} = re^{i\frac{m}{n}\theta} = re^{i(\frac{m}{n}\theta + 2\pi} = re^{i\frac{m}{n}(\theta + 2\pi n)}$$

That is, our multifunction returns to a previous value whenever the input encircles the origin n times.

8.2. Now try $(1 - z^4)^{1/2}$.

The four branch points of $(1-z^4)^{1/2}$ are the four roots of unity of z^4, namely 1, i, -1, and $-i$. As a reminder, these points are when our multifunction's values overlap. Being a square root function, there are two values associated with each input, with each the negative of the other. The points of overlap (branch points) are when $(1-z^4)^{1/2} = -(1-z^4)^{1/2}$, i.e. when our function evaluates to 0. Geometrically, our multifunction has two sheets in its Riemann surface, each adjoined to the other at the branch points.

We may stop here, noting the similarity with Fig. 8.2 in the text, but we may also be more explicit. Consider making two branch cuts in our function: from 1 to i and from -1 to $-i$. Our codomain is the extended complex plane, or Riemann sphere, with two cuts in it. Topologically, a sphere with two cuts is nothing but a cylinder. We can adjoin this to our second sheet to form a torus, just as we had down with $(1-z^3)^{1/2}$.

8.3. Can you see how this comes about? (*Hint*: Think of the Riemann sphere of the variable $w(=\log z)$; see §8.3.

This can be seen by constructing an injective mapping from the infinite sheets of the z-plane to the w-plane, and then from the w-plane to the Riemann sphere. The first of these mappings can be understood as a kind of change of coordinates from radial to polar. Consider $\log z = \log r + i\theta$. Fixing r and changing the angle traces out the a vertical line segment in the w-plane, with each winding number being separated from any other by 2π. On the other hand, fixing θ and changing r traces out a horizontal line the w-plane. In this way, every point in the infinite ramp of the z-plane maps to exactly one point in the w-plane. From here, simply stereographically project the w-plane to obtain the Riemann sphere.

8.4. Show this.

Consider an arbitrary circle in the complex plane, described by $z - a = re^{i\theta}$, or, equivalently, $(z-a)(\bar{z}-\bar{a}) = |r|^2$. We are considered with how this set changes under the mapping $w = z^{-1}$, or, equivalently, $w^{-1} = z$. Substituting in the second of these gives

$$\begin{aligned}
\left(\frac{1}{w} - a\right)\left(\frac{1}{\bar{w}} - \bar{a}\right) &= |r|^2 \\
(1-wa)(1-\bar{w}\bar{a}) &= |r|^2 w\bar{w} \\
w\bar{w}|a|^2 - wa - \bar{w}\bar{a} + 1 &= |r|^2 w\bar{w} \\
\frac{1}{|r|^2 - |a|^2} &= w\bar{w} + \frac{a}{|r|^2-|a|^2} w + \frac{\bar{a}}{|r|^2-|a|^2}\bar{w} \\
\alpha &= w\bar{w} + \alpha a w + \alpha\bar{a}\bar{w} \\
\alpha &= (w+\alpha\bar{a})(\bar{w}+\alpha a) - \alpha|a|^2 \\
|\gamma|^2 &= (w+b)(\bar{w}+\bar{b})
\end{aligned}$$

where $\alpha = (|r|^2 - |a|^2)^{-1}$, $b = \alpha\bar{a}$, and $\gamma = \alpha(1+|a|^2)$. This represents another set of points in the complex plain describing a circle, so $w = z^{-1}$ sends circles to circles. In the case of $|r|^2 = |a|^2$, circles in the z-plane get mapped to lines (circles of infinite radius) in the w-plane.

8.5. Verify that the sequence of transformations $z \to Az + B$, $z \to z^{-1}$, $z \to Cz + D$ indeed leads to a bilinear map.

Carrying out each transformation in turn gives a final result of $z \to \frac{C}{Az+B} + D$, which can be rearranged to give $\frac{(DA)z+(B+C)}{Az+B}$. This is a bilinear map.

8.6. Check that these two stereographic projections are related by $w = z^{-1}$.

8.7. Show this.

8.8. Show that these replacements give holomorphically equivalent spaces. Find all the special values of p where these equivalences lead to additional discrete symmetries of the Riemann surface.

9 Fourier decomposition and hyperfunctions

9.1. Show this.

We can replace the trigonometric functions in our expansion of $f(\chi)$ with complex exponentials by substituting the identities $\cos\omega\chi = (e^{i\omega\chi} + e^{-i\omega\chi})/2$ and $\sin\omega\chi = (ie^{-i\omega\chi} - ie^{i\omega\chi})/2$. This gives us

$$\begin{aligned} f(\chi) &= c + \frac{a_1}{2}e^{i\omega\chi} + \frac{a_1}{2}e^{-i\omega\chi} + i\frac{b_1}{2}e^{-i\omega\chi} - i\frac{b_1}{2}e^{i\omega\chi} + \cdots \\ &= \cdots + \left(\frac{a_1}{2} + i\frac{b_1}{2}\right)e^{-i\omega\chi} + c + \left(\frac{a_1}{2} - i\frac{b_1}{2}\right)e^{i\omega\chi} + \cdots \end{aligned}$$

Matching this with

$$f(\chi) = \cdots + \alpha_{-1}e^{-i\omega\chi} + \alpha_0 + \alpha_1 e^{i\omega\chi} + \cdots$$

gives us $\alpha_{-n} = \frac{a_n}{2} + i\frac{b_n}{2}$, $\alpha_0 = c$, and $\alpha_n = \frac{a_n}{2} - i\frac{b_n}{2}$. We can solve for a_n by adding the first and third equation together, yielding $\alpha_n + \alpha_{-n} = a_n$. Subtracting the first from the third gives us $\alpha_n - \alpha_{-n} = -ib_n$, which, upon multiplying by i, becomes $i\alpha_n - i\alpha_{-n} = b_n$.

9.2. Show that when F is analytic on the unit circle the coefficients α_n, and hence the a_n, b_n, and c, can be obtained by use of the formula $\alpha_n = (2\pi i)^{-1} \oint z^{-n-1}F(z)\mathrm{d}z$.

A simple substitution of the Laurent series of $F(z)$ gives

$$\begin{aligned} \alpha_n &= \frac{1}{2\pi i}\oint \frac{\sum_{k=-\infty}^{\infty} \alpha_k z^k}{z^{n+1}}\mathrm{d}z \\ &= \frac{1}{2\pi i}\oint \frac{\alpha_n}{z}\mathrm{d}z \\ &= \alpha_n \end{aligned}$$

where the second step is justified by the fact that $z^n = 0$ for $n \neq -1$ (see Exercise 7.1, this is the step that requires analyticity), and the third step is an application of the residue theorem.

9.3. Can you see why?

The positive-frequency part of $F(z)$ converges inside the unit circle (with the boundary being determined on a case-by-case basis using the Fourier series' coefficients), while the negative-frequency part of $F(z)$ converges outside the unit circle (with the same stipulation holding). But the inside of the unit circle extends holomorphically to the southern hemisphere of the Riemann sphere, while the outside maps to the northern hemisphere. Therefore, this split is determined by the map from $F(z)$ to the Riemann sphere.

9.4. Which are these mappings, explicitly?

9.5. Show that this gives the same t as above.

The given mapping takes the unit circle in $\mathbb{C}$ to $\mathbb{R}$. Substituting $z = e^{i\theta}$ in this expression shows us that

$$\begin{aligned} t &= \frac{e^{i\theta} - 1}{ie^{i\theta} + i} \\ &= \frac{e^{i\theta/2}}{e^{i\theta/2}} \frac{e^{i\theta/2} - e^{-i\theta/2}}{ie^{i\theta/2} + ie^{-i\theta/2}} \\ &= \frac{i \sin \frac{\theta}{2}}{i \cos \frac{\theta}{2}} \\ &= \tan \frac{\theta}{2} \end{aligned}$$

9.6. Show (in outline) how to obtain the expression for $g(p)$ in terms of $f(\chi)$ using a limiting form of the contour integral expresion $\alpha_n = (2\pi i)^{-1} \oint z^{-n-1} F(z) \mathrm{d}z$ of Exercise [9.2].

Keeping with the notation used in the chapter, we replace n with r, and note that $\omega = \frac{2\pi}{l}$, $l = 2\pi N$, $z = e^{i\omega\chi}$, and $p = \frac{r}{N}$. Working directly with the given expression, we see

$$\begin{aligned} \lim_{N\to\infty} \alpha_r &= \lim_{N\to\infty} \frac{1}{2\pi i} \oint \frac{F(z)}{z^{r+1}} \mathrm{d}z \\ &= \lim_{N\to\infty} \frac{1}{2\pi i} \int_{-\pi N}^{\pi N} \frac{f(\chi)}{e^{i\omega\chi(r+1)}} e^{i\omega\chi} i\omega \mathrm{d}\chi \\ &= \lim_{N\to\infty} \frac{1}{2\pi N} \int_{-\infty}^{\infty} \frac{f(\chi)}{e^{i\chi r/N}} \mathrm{d}\chi \\ &= \lim_{N\to\infty} \frac{1}{l} \int_{-\infty}^{\infty} f(\chi) e^{-i\chi p} \mathrm{d}\chi \end{aligned}$$

I am unsure of how to handle the denominator in the limit. Clearly, the infinite integral offsets the vanishing coefficient, but the exact nature of the cancellation is unknown to me.

9.7. Derive this expression.

If χ is to traverse arc length l in a clockwise manner, the circle in the complex plane representing such a mapping will be given by $\frac{l}{2\pi}e^{-i\frac{2\pi}{l}\chi}$. That is, we are scaling the radius by r using the relationship $2\pi r = l$, and scaling the frequency in such a way that moving χ from 0 to l traverses the entire circle. To obtain the needed expression, this must be offset along the negative imaginary axis by the circle's radius and χ must be advanced by $\frac{\pi}{2}$ (so that it begins at the origin). That is,

$$z = \frac{l}{2\pi}e^{-i(\omega\chi - \frac{\pi}{2})} - \frac{il}{2\pi} = \frac{il}{2\pi}(e^{-i\omega\chi} - 1).$$

9.8. Show this.

We can obtain the necessary expression by substituting Euler's identity,

$$\begin{aligned} s(\chi) &= \sum_{k=0}^{\infty} \frac{1}{2k+1} \sin((2k+1)\chi) \\ &= \sum_{k=0}^{\infty} \frac{1}{2k+1} \frac{e^{i(2k+1)\chi} - e^{-i(2k+1)\chi}}{2i} \end{aligned}$$

That is, $2is(\chi) = \cdots - \frac{1}{5}z^{-5} - \frac{1}{3}z^{-3} - z^{-1} + z + \frac{1}{3}z^3 + \frac{1}{5}z^5 + \cdots$, where $z = e^{i\chi}$.

9.9. Do this, by taking advantage of a power series expansion for $\log z$ taken about $z = 1$, given towards the end of §7.4

By looking at the Taylor series expansion of $\log z$ about $p = 1$,

$$\log z = \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k}(z-1)^k$$

we find

$$\begin{aligned} \log(1+z) &= \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} z^k \\ \log(1-z) &= \sum_{k=1}^{\infty} \frac{-1}{k} z^k \\ \log(1+z^{-1}) &= \sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} z^{-k} \\ \log(1-z^{-1}) &= \sum_{k=1}^{\infty} \frac{-1}{k} z^{-k} \end{aligned}$$

Subtracting the second from the first gives

$$\sum_{k=0}^{\infty} \frac{2z^{2k+1}}{2k+1}$$

while subtracting the fourth from the third gives

$$\sum_{k=0}^{\infty} \frac{2z^{-2k-1}}{2k+1}$$

. This gives us the expressions

$$\begin{aligned}
\frac{1}{2}\log(1+z) - \frac{1}{2}\log(1-z) = \frac{1}{2}\log\left(\frac{1+z}{1-z}\right) = \sum_{k=0}^{\infty} \frac{z^{2k+1}}{2k+1} \\
-\frac{1}{2}\log(1+z^{-1}) + \frac{1}{2}\log(1-z^{-1}) = -\frac{1}{2}\log\left(\frac{1+z^{-1}}{1-z^{-1}}\right) = \sum_{k=0}^{\infty} -\frac{z^{-2k-1}}{2k+1}
\end{aligned}$$

9.10. Show this (assuming that $|s(\chi)| < 3\pi/2$).

$$\begin{aligned}
S^- + S^+ &= \frac{1}{2}\log\left(\frac{1+z}{1-z}\right) - \frac{1}{2}\log\left(\frac{1+z^{-1}}{1-z^{-1}}\right) \\
&= \frac{1}{2}\Big[\log\left(\frac{1+e^{i\chi}}{1-e^{i\chi}}\right) - \log\left(\frac{1+e^{-i\chi}}{1-e^{-i\chi}}\right)\Big] \\
&= \frac{1}{2}\Big[\log\left(\frac{e^{i\chi/2}}{e^{i\chi/2}}\frac{e^{-i\chi/2}+e^{i\chi/2}}{e^{-i\chi/2}-e^{i\chi/2}}\right) - \log\left(\frac{e^{-i\chi/2}}{e^{-i\chi/2}}\frac{e^{i\chi/2}+e^{-i\chi/2}}{e^{-\chi/2}-e^{-i\chi/2}}\right)\Big] \\
&= \frac{1}{2}\Big[\log\left(\frac{2\cos\chi}{-2i\sin\chi}\right) - \log\left(\frac{2\cos\chi}{2i\sin\chi}\right)\Big] \\
&= \frac{1}{2}\log\left(\frac{2\cos\chi}{-2i\sin\chi}\frac{2i\sin\chi}{2\cos\chi}\right) \\
&= \frac{1}{2}\log(-1) \\
&= \frac{1}{2}\log(e^{\pm i\pi}) \\
&= \log(e^{\pm i\pi/2}) \\
&= \pm\frac{i\pi}{2}
\end{aligned}$$

So $2is(\chi) = \pm\frac{i\pi}{2}$, or $s(\chi) = \pm\frac{\pi}{4}$.

9.11. Show this.

Consider two paths from the starting points of our logarithms to the real axis, parameterized by $|r_-|e^{i\theta_-}$ and $|r_+|e^{i\theta_+}$. Here, θ_- starts at $\pi/2$ and θ_+ starts at $-\pi/2$. As these two approach the real axis, their angles approach 0 and their radii become equal, i.e. $|r_-| = |r_+| = |r|$. Substituting this in for t, we see

$$S^- + S^+ = -\frac{1}{2}\log|r| + \frac{1}{2}\log i + \frac{1}{2}\log|r| + \frac{1}{2}\log i = \log i = \frac{i\pi}{2}.$$

If we instead approach the negative real axis from these starting points, θ_- approach π while θ_+ approaches $-\pi$. As in the previous case, we have $|r_-| = |r_+| = |r|$. Putting this in our expression yields

$$S^- + S^+ = -\frac{1}{2}\log|r| - \frac{i\pi}{2} + \frac{1}{2}\log i + \frac{1}{2}\log|r| - \frac{i\pi}{2} + \frac{1}{2}\log i = -i\pi + \log i = -\frac{i\pi}{2}.$$

9.12. Why does 'holomorphic functions on $\mathbb{R} - \bar{\gamma}$, reduced modulo holomorphic functions on $\mathbb{R}$' become the definition of a hyperfunction that we had previously, when $\mathbb{R} - \bar{\gamma}$ splits into R^- and R^+.

9.13. There is a small subtlety here. Sort it out. *Hint*: Think carefully about the domains of definition.

9.14. Check the standard property of the delta function that $\int q(x)\delta(x)\mathrm{d}x = q(0)$, in the case when $q(x)$ is analytic.

Since $q(x)$ is analytic, we may multiply it by the δ function to obtain

$$\left(\!\left|\frac{q(z)}{2\pi iz}, \frac{q(z)}{2\pi iz}\right|\!\right).$$

This, being a hyperfunction, can be integrated as

$$\int q(x)\delta(x)\mathrm{d}x = \int_{L^+} \frac{q(z)}{2\pi iz}\mathrm{d}z - \int_{L^-} \frac{q(z)}{2\pi iz}\mathrm{d}z.$$

where L^+ is a contour just below the real line from $-\infty$ to ∞ and L^- is the same, but above the real line. Given that these integrals converge at $\pm\infty$, we may write them as a single integral and invoke the residue theorem to find

$$\int_C \frac{q(z)}{2\pi iz}\mathrm{d}z = q(0).$$

This shows that our δ function behaves as expected.

10 Surfaces

10.1. Explain why subtraction and divison can be constructed from these.

Subtraction may be reframed as addition of a negative number, that is $z - w = z + (-w)$. Division by w may be defined by identifying v in $wv = 1$, then multiplying z by v.

10.2. Derive both of these.

Replacing z with $x + iy$ yields (in the case of the first expression)

$$
\begin{aligned}
z^2 + \bar{z}^2 &= (x+iy)^2 + (x-iy)^2 \\
&= x^2 + i2xy - y^2 + x^2 - i2xy - y^2 \\
&= 2x^2 - 2y^2
\end{aligned}
$$

For the second expression, we have

$$
\begin{aligned}
z\bar{z} &= (x+iy)(x-iy) \\
&= x^2 - ixy + ixy + y^2 \\
&= x^2 + y^2
\end{aligned}
$$

10.3. Consider the real function $f(x,y) = xy(x^2+y^2)^{-N}$, in the respective cases $N = 2$, 1, and $\frac{1}{2}$. Show that in each case the function is differentiable (C^ω) with respect to x, for any fixed y-value (and that the same holds with the roles of x and y reversed). Nevertheless, f is not smooth as a function of the pair (x,y). Show this in the case $N = 2$ b demonstrating that the function is not even bounded in the neighbourhood of the origin $(0,0,)$ (i.e. it tkaes arbitrarily large values there), in the case $N = 1$ by demonstrating that the function though bounded is not actually continuous as a function of (x,y), and in the case of $N = \frac{1}{2}$ by showing that though the function is now continuous, it is not smooth aong the line $x = y$. (*Hint*: Examine the values of each function along straight ines through the origin in the (x,y)-plane.) Some readers may find it illuminating to use a suitable 3-dimensional graph-plotting computer facility, if this is available−though this is by not means necessary.

10.4. Prove that the mixed second derivatives $\partial^2 f/\partial y \partial x$ and $\partial^2 f/\partial x \partial y$ are always equal if $f(x,y)$ is a polynomial. (A *polynomial* in x and y is an expression built up from x, y, and constants by use of addition and multiplication only.)

We can represent f by $f = \sum_{m,n} c_{mn} x^m y^n$. Taking the partial derivatives of this gives

$$
\begin{aligned}
\frac{\partial f}{\partial x} &= \sum_{m,n} m c_{mn} x^{m-1} y^n \\
\frac{\partial f}{\partial y} &= \sum_{m,n} n c_{mn} x^m y^{n-1}
\end{aligned}
$$

Taking the opposite partial derivative of each of these expressions yields

$$
\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y} = \sum_{m,n} mn c_{mn} x^{m-1} y^{n-1}.
$$

10.5. Show that the mixed second derivatives of the function $f = xy(x^2-y^2)/(x^2+y^2)$ are unequal at the origin. Establish directly the lack of continuity in its second partial derivatives at the origin.

10.6. Find the form of $F(X, Y)$ explicitly when $f(x, y) = x^3 - y^3$ when $X = x - y$, $Y = xy$. *Hint*: What is $x^2 + xy + y^2$ in terms of X and Y; what does this have to do with f?

Our function can be rewritten as $f(x, y) = (x^2 + xy + y^2)(x - y)$. The last term is clearly X, while the first term is $X^2 + 3Y$. Putting this together, we find $F(X, Y) = X^3 + 3YX$.

10.7. Find A and B in terms of a and b; by analogy, write down a and b in terms of A and B.

We can represent $\frac{\partial}{\partial X}$ by $\frac{\partial x}{\partial X}\frac{\partial}{\partial x} + \frac{\partial y}{\partial X}\frac{\partial}{\partial y}$, and similarly for the other coordinate variables. Inserting these equivalent expressions into our different representatios of ξ gives us

$$\begin{aligned} A &= a\frac{\partial X}{\partial x} + b\frac{\partial X}{\partial y} \\ B &= a\frac{\partial Y}{\partial x} + b\frac{\partial Y}{\partial y} \\ a &= A\frac{\partial x}{\partial X} + B\frac{\partial x}{\partial Y} \\ b &= B\frac{\partial y}{\partial X} + B\frac{\partial y}{\partial Y} \end{aligned}$$

10.8. Derive this explicitly. *Hint*: You may use 'chain rule' expression for $\partial/\partial X$ and $\partial/\partial Y$ that are the exact analogies of the expression for $\partial/\partial x$ that was displayed earlier.

Working one variable at a time, we know $\frac{\partial}{\partial X} = \frac{\partial x}{\partial X}\frac{\partial}{\partial x} + \frac{\partial y}{\partial X}\frac{\partial}{\partial y}$. We also know that $Y = y + x$, or $y = Y - x = Y - X$. Putting these together gives us $\partial x/\partial X = 1$ and $\partial y/\partial X = -1$, so

$$\frac{\partial}{\partial X} = \frac{\partial}{\partial x} - \frac{\partial}{\partial y}.$$

Analogously, $\frac{\partial}{\partial Y} = \frac{\partial x}{\partial Y}\frac{\partial}{\partial x} + \frac{\partial y}{\partial Y}\frac{\partial}{\partial y}$. We have $\partial x/\partial Y = 0$ and $\partial y/\partial Y = 1$, and so

$$\frac{\partial}{\partial Y} = \frac{\partial}{\partial y}.$$

10.9. Show this explicitly using 'chain rule' expressions that we have seen earlier.

We know that $u = \partial\Phi/\partial x$ and $v = \partial\Phi/\partial y$. From this, it immediately follows that $\xi(\Phi) = a\frac{\partial\Phi}{\partial x} + b\frac{\partial\Phi}{\partial y} = au + bv$.

10.10. Explain this from three different points of view: (a) intuitively, from general principles (how could a $\bar{z}$ appear?), (b) using the geometry of holomorphic maps described in §8.2, and (c) explicitly, using the chain rule and the Cauchy-Riemann equations that we are about to come to.

10.11. Do this.

We can expand $\partial/\partial\bar{z}$ as $\frac{\partial x}{\partial \bar{z}}\frac{\partial}{\partial x}+\frac{\partial y}{\partial \bar{z}}\frac{\partial}{\partial y}$ Furthermore, we can express x and y in terms of z and $\bar{z}$: $x=(z+\bar{z})/2$ and $y=(z-\bar{z})/(2i)$. Quickly computing $\partial x/\partial\bar{z}=1/2$ and $\partial y/\partial\bar{z}=i/2$, we see

$$\frac{\partial\Phi}{\partial\bar{z}}=\frac{1}{2}\frac{\partial\Phi}{\partial x}+i\frac{1}{2}\frac{\partial\Phi}{\partial y}.$$

Setting this equal to 0 and multiplying through by 2 gives us

$$\frac{\partial\Phi}{\partial x}+i\frac{\partial\Phi}{\partial y}=0.$$

10.12. Give a more direct derivation of the Cauchy-Riemann equations, from the definition of a derivative.

Recalling that the Cauchy-Riemann equations are merely a consequence of a function holomorphicity, let us try to prove the latter fact. This boils down to showing that a given function is differentiable in the complex plane (as, if it appears to be C^1, it is immediately C^ω). We have

$$\begin{aligned}\frac{\mathrm{d}\Phi}{\mathrm{d}z}&=\lim_{\Delta z\to 0}\frac{\Phi(z+\Delta z)-\Phi(z)}{\Delta z}\\&=\lim_{\Delta z\to 0}\frac{\alpha(x+\Delta x,y+\Delta y)-\alpha(x,y)}{\Delta x+i\Delta y}+i\frac{\beta(x+\Delta x,y+\Delta y)-\beta(x,y)}{\Delta x+i\Delta y}\end{aligned}$$

Where we have split both Φ and z into their constituent real and imaginary parts. As differentiability in the complex plane must hold when approaching a point from any direction, let us consider two cases: $\Delta y=0$ and $\Delta x=0$.

$$\begin{aligned}\lim_{\Delta z\to 0}\frac{\alpha(x+\Delta x,y)-\alpha(x,y)}{\Delta x}+i\frac{\beta(x+\Delta x,y)-\beta(x,y)}{\Delta x}&=\frac{\partial\alpha}{\partial x}+i\frac{\partial\beta}{\partial x}\\\lim_{\Delta z\to 0}\frac{\alpha(x,y+\Delta y)-\alpha(x,y)}{i\Delta y}+i\frac{\beta(x,y+\Delta y)-\beta(x,y)}{i\Delta y}&=-i\frac{\partial\alpha}{\partial y}+\frac{\partial\beta}{\partial y}\end{aligned}$$

These must be equal to each other for our function to be holomorphic. Equating real and imaginary parts gives us

$$\frac{\partial\alpha}{\partial x}=\frac{\partial\beta}{\partial y}\qquad\frac{\partial\alpha}{\partial y}=-\frac{\partial\beta}{\partial x}.$$

10.13. Show this.

Assuming the Cauchy-Riemann conditions hold, we have

$$\nabla^2\alpha = \frac{\partial}{\partial x}\left(\frac{\partial\alpha}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{\partial\alpha}{\partial y}\right) = \frac{\partial}{\partial x}\left(\frac{\partial\beta}{\partial y}\right) - \frac{\partial}{\partial y}\left(\frac{\partial\beta}{\partial x}\right) = 0$$
$$\nabla^2\beta = \frac{\partial}{\partial x}\left(\frac{\partial\beta}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{\partial\beta}{\partial y}\right) = -\frac{\partial}{\partial x}\left(\frac{\partial\alpha}{\partial x}\right) + \frac{\partial}{\partial y}\left(\frac{\partial\alpha}{\partial x}\right) = 0$$

Where the last equality in both equations comes from the equality of mixed partial derivatives.

10.14. Show this.

Assuming α is harmonic (i.e. $\nabla^2\alpha = 0$), and defining $\beta = \int \partial\alpha/\partial x \mathrm{d}y$, we have

$$\begin{aligned}\frac{\partial\beta}{\partial x} &= \frac{\partial}{\partial x}\int \frac{\partial\alpha}{\partial x}\mathrm{d}y \\ &= \int \frac{\partial^2\alpha}{\partial x^2}\mathrm{d}y \\ &= -\int \frac{\partial^2\alpha}{\partial y^2}\mathrm{d}y \\ &= -\frac{\partial\alpha}{\partial y}\end{aligned}$$

and

$$\begin{aligned}\frac{\partial\beta}{\partial y} &= \frac{\partial}{\partial y}\int \frac{\partial\alpha}{\partial x}\mathrm{d}y \\ &= \frac{\partial\alpha}{\partial x}\end{aligned}$$

Where the first equation's third equality comes from α being harmonic, and the second equations last equality comes from integration being the inverse of differentiation.

10.15. Spell this out in the case $\Phi(u, v) = \theta(v)h(u)$, where the functions θ and h are defined as §§6.1, 3. THe kidney-shaped region must avoid the non-negative u-axis.

11 Hypercomplex numbers

11.1. Prove these directly from Hamilton's 'Brougham Bridge equations', assuming only the associative law $a(bc) = (ab)c$.

Starting from $\mathbf{ijk} = -1$, we can pre- and post-multiply both sides by k to obtain $\mathbf{kijk}^2 = -kk$, or $\mathbf{kij} = -1$. This can be done again with j to find $\mathbf{jki} = -1$. We can combine the first two of these expressions to find

$$\begin{aligned}(\mathbf{ijk})(\mathbf{kij}) &= 1\\ -\mathbf{ijij} &= 1\\ -\mathbf{ijij}^2\mathbf{i} &= \mathbf{ji}\\ -\mathbf{ij} &= \mathbf{ji}\end{aligned}$$

The other combinations of these results yield $-\mathbf{ik} = \mathbf{ki}$ and $-\mathbf{jk} = \mathbf{kj}$.

11.2. Express the sum and product of two general quaternions so that all these indeed hold.

The sum of two quaternions, $\mathbf{q}_1 = a_0 + a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}$ and $\mathbf{q}_2 = b_0 + b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k}$, can be defined as

$$\mathbf{q}_1 + \mathbf{q}_2 = (a_0 + b_0) + (a_1 + b_1)\mathbf{i} + (a_2 + b_2)\mathbf{j} + (a_3 + b_3)\mathbf{k}.$$

Multiplication is slightly trickier. After expanding everything, we can replace $\mathbf{ij}$ with $\mathbf{k}$, $\mathbf{jk}$ with $\mathbf{i}$, and $\mathbf{ki}$ with $\mathbf{j}$. Then we have

$$\mathbf{q}_1\mathbf{q}_2 = (a_0b_0 - a_1b_1 - a_2b_2 - a_3b_3) + (a_0b_1 + a_1b_0 + a_2b_3 - a_3b_2)\mathbf{i} + (a_0b_2 - a_1b_3 + a_2b_0 + a_3b_1)\mathbf{j} + (a_0b_3 + a_1b_2 - a_2b_1 + a_3b_0)\mathbf{k}.$$

11.3. Check that this definition of $\mathbf{q}^{-1}$ actually works.

In order to check this definition, we will consider a general quaternion of the form $\mathbf{q} = t + u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$. Its inverse is

$$\mathbf{q}^{-1} = \bar{\mathbf{q}}(\mathbf{q}\bar{\mathbf{q}})^{-1} = \frac{t - u\mathbf{i} - v\mathbf{j} - w\mathbf{k}}{t^2 + u^2 + v^2 + w^2}.$$

Expanding $\mathbf{q}\mathbf{q}^{-1}$ produces a lengthy, yet simple, expression. Upon replacing pairs of basis elements with single ones (e.g. replacing $\mathbf{ij}$ with $\mathbf{k}$), we find that all terms cancel with the exception of $t^2 - u^2\mathbf{i}^2 - v^2\mathbf{j}^2 - w^2\mathbf{k}^2$. Using Hamilton's identities, this equals the denominator, giving us $\mathbf{q}\mathbf{q}^{-1} = 1$.

11.4. Check this.

Once again denoting a general quaternion with $\mathbf{q} = t + u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$, we see that

$$\begin{aligned}\mathbf{iqi} &= t\mathbf{i}^2 + u\mathbf{i}^3 + v\mathbf{iji} + w\mathbf{iki}\\ &= -t - u\mathbf{i} + v\mathbf{ki} + w\mathbf{ij}\\ &= -t - u\mathbf{i} + v\mathbf{j} + w\mathbf{k}\end{aligned}$$

Similarly, $\mathbf{jqj} = -t + u\mathbf{i} - v\mathbf{j} + w\mathbf{k}$ and $\mathbf{kqk} = -t + u\mathbf{i} + v\mathbf{j} - w\mathbf{k}$. Summing all of these together with $\mathbf{q}$ gives

$$\mathbf{q} + \mathbf{iqi} + \mathbf{jqj} + \mathbf{kqk} = -2t + 2u\mathbf{i} + 2v\mathbf{j} + 2w\mathbf{k}.$$

Multiplying this quantity by $-\frac{1}{2}$ gives us $\bar{\mathbf{q}}$.

11.5. In Hamilton's original version of this construction, the 'dual' spherical triangle to this one is used, whose vertices are where the sphere meets the three axes of rotation involved in the problem. Give a direct demonstration of how this works (perhaps 'dualizing' the argument given in the text), the amounts of the rotations being represented as twice the *angles* of this dual triangle.

11.6. Find the geometrical nature of the transformation, in Euclidean 3-space, which is the composition of two reflections in planes that are not perpendicular.

11.7. Show this.

We must show that $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1$. For $\mathbf{i}^2$, we have

$$\gamma_2\gamma_3\gamma_2\gamma_3 = -\gamma_3\gamma_2^2\gamma_3 = \gamma_3^2 = -1.$$

Similarly, we find for $\mathbf{j}^2$ that

$$\gamma_3\gamma_1\gamma_3\gamma_1 = -\gamma_1\gamma_3^2\gamma_1 = \gamma_1^2 = -1.$$

And for $\mathbf{k}^2$,

$$\gamma_1\gamma_2\gamma_1\gamma_2 = -\gamma_2\gamma_1^2\gamma_2 = \gamma_2^2 = -1.$$

Finally, for $\mathbf{ijk}$, we see

$$\gamma_2\gamma_3\gamma_3\gamma_1\gamma_1\gamma_2 = \gamma_2(-1)(-1)\gamma_2 = \gamma_2^2 = -1.$$

11.8. Explain all this counting. *Hint*: Think of $(1+1)^n$.

In an n-dimensional space, the linearly independent elements are distinct products of the available γ's. Since there are n different individual γ's, each grouping of kth-order entities has $\binom{n}{k}$ elements.

11.9. Show this.

Writing our two vectors as $\mathbf{a} = a_1\eta_1 + \cdots + a_n\eta_n$ and $\mathbf{b} = b_1\eta_1 + \cdots + b_n\eta_n$, we see that their wedge product is

$$
\begin{aligned}
\mathbf{a} \wedge \mathbf{b} &= \Big(\sum_{i=1}^{n} a_i \eta_i\Big) \wedge \Big(\sum_{j=1}^{n} b_j \eta_j\Big) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i b_j \eta_i \wedge \eta_j \\
&= -\sum_{j=1}^{n} \sum_{i=1}^{n} b_j a_i \eta_j \wedge \eta_i \\
&= -\Big(\sum_{j=1}^{n} b_j \eta_j\Big) \wedge \Big(\sum_{i=1}^{n} a_i \eta_i\Big) \\
&= -\mathbf{b} \wedge \mathbf{a}
\end{aligned}
$$

11.10. Write out $\mathbf{a} \wedge \mathbf{b}$ fully in the case $n = 2$, to see how this comes about.

Carrying out the product, we see

$$
\begin{aligned}
\mathbf{a} \wedge \mathbf{b} &= (a_1\eta_1 + a_2\eta_2) \wedge (b_1\eta_1 + b_2\eta_2) \\
&= a_1 b_1 \eta_1 \wedge \eta_1 + a_1 b_2 \eta_1 \wedge \eta_2 + a_2 b_1 \eta_2 \wedge \eta_1 + a_2 b_2 \eta_2 \wedge \eta_2 \\
&= a_1 b_2 \eta_1 \wedge \eta_2 + a_2 b_1 \eta_2 \wedge \eta_1 \\
&= \frac{1}{2} a_1 b_2 \eta_1 \wedge \eta_2 + \frac{1}{2} a_2 b_1 \eta_2 \wedge \eta_1 + \frac{1}{2} a_1 b_2 \eta_1 \wedge \eta_2 + \frac{1}{2} a_2 b_1 \eta_2 \wedge \eta_1 \\
&= \frac{1}{2}(a_1 b_2 - a_2 b_1)\eta_1 \wedge \eta_2 + \frac{1}{2}(a_2 b_1 - a_1 b_2)\eta_2 \wedge \eta_1
\end{aligned}
$$

where in the third line we split each term a into $\frac{1}{2}(a + a)$. Put into this form, we see that each component of $\mathbf{a} \wedge \mathbf{b}$ is given by $a_{[p}b_{q]}$.

11.11. Write down this expression explicitly in the case of a wedge product of four vectors.

This is quite a tedious exercise, and so has been skipped. The result is the alternating sum of all permutations of four vector components divided by $4! = 24$.

11.12. Show that the wedge product remains unaltered if $\mathbf{a}$ is replaced by $\mathbf{a}$ added to any multiple of any of the other vectors involved in the wedge product.

The wedge product is zero if any vectors involved in the expression are the same. Therefore,

$$
\begin{aligned}
(\mathbf{a} + \mathbf{b}_k) \wedge \mathbf{b}_1 \wedge \cdots \wedge \mathbf{b}_k \wedge \cdots \wedge \mathbf{b}_n &= \mathbf{a} \wedge \mathbf{b}_1 \wedge \cdots \wedge \mathbf{b}_n + \mathbf{b}_k \wedge \mathbf{b}_1 \wedge \cdots \wedge \mathbf{b}_k \wedge \cdots \wedge \mathbf{b}_n \\
&= \mathbf{a} \wedge \mathbf{b}_1 \wedge \cdots \wedge \mathbf{b}_n + 0 \\
&= \mathbf{a} \wedge \mathbf{b}_1 \wedge \cdots \wedge \mathbf{b}_n
\end{aligned}
$$

11.13. Show this.

The wedge product of $\mathbf{P}$ and $\mathbf{Q}$ is given by

$$\sum\sum P_{a\ldots c}Q_{d\ldots f}\eta_1 \wedge \cdots \wedge \eta_p \wedge \xi_1 \wedge \cdots \wedge \xi_q.$$

The wedge product of $\mathbf{Q}$ and $\mathbf{P}$ can be reached from this expression by moving the elements of the form $\eta_{\mathbf{1}} \wedge \cdots \wedge \eta_p$ to the right of those of the form $\xi_1 \wedge \cdots \wedge \xi_q$. If either q or p are even, this introduces an even number of swaps, resulting in no change of sign. If, on the other hand, q and p are both odd, this results in an odd number of swaps, giving $\mathbf{Q} \wedge \mathbf{P} = -\mathbf{P} \wedge \mathbf{Q}$.

11.14. Deduce that $\mathbf{P} \wedge \mathbf{P} = 0$, if p is odd.

By the above argument, $\mathbf{P} \wedge \mathbf{P} = -\mathbf{P} \wedge \mathbf{P}$, implying this quantity equals 0.

12 Manifolds of n dimensions

12.1. Explain this dimension count more explicitly.

In space, we need three numbers to specify a location. For rotation, we need a further three: two for the orientation of the rotation axis, and one for the twist about this axis. All in all, this gives us six total dimensions.

12.2. Show how to do this, e.g. by appealing to the representation of R as given in Exercise [12.17].

See the answer to Exercise 12.17, where I have included this.

12.3. Show that 'dΦ', defined in this way, indeed satisfies the 'linearity' requirements of a covector, as specified above.

Using the defining feature of the exterior derivative,

$$\begin{aligned}
\mathrm{d}\Phi \cdot (\xi + \eta) &= (\xi + \eta)(\Phi) \\
&= (\xi^1 \frac{\partial}{\partial x^1} + \cdots + \xi^n \frac{\partial}{\partial x^n} + \eta^1 \frac{\partial}{\partial x^1} + \cdots + \eta^n \frac{\partial}{\partial x^n})(\Phi) \\
&= \xi^1 \frac{\partial \Phi}{\partial x^1} + \cdots + \xi^n \frac{\partial \Phi}{\partial x^n} + \eta^1 \frac{\partial \Phi}{\partial x^1} + \cdots + \eta^n \frac{\partial \Phi}{\partial x^n} \\
&= \xi(\Phi) + \eta(\Phi) \\
&= \mathrm{d}\Phi \cdot \xi + \mathrm{d}\Phi \cdot \eta
\end{aligned}$$

and

$$
\begin{aligned}
\mathrm{d}\Phi\cdot(\omega\xi) &= (\omega\xi)(\Phi)\\
&= (\omega\xi^1\frac{\partial}{\partial x^1}+\cdots+\omega\xi^n\frac{\partial}{\partial x^n})(\Phi)\\
&= \omega(\xi^1\frac{\partial\Phi}{\partial x^1}+\xi^n\frac{\partial\Phi}{\partial x^n})\\
&= \omega(\xi(\Phi))\\
&= \omega(\mathrm{d}\Phi\cdot\xi)
\end{aligned}
$$

12.4. Why?

Because $\mathrm{d}\Phi\cdot\xi=\xi(\Phi)=0$ occurs along surfaces of constant Φ.

12.5. For example, show that $\mathrm{d}x^2$ has components $(0,1,0,\cdots,0)$ and represents the tangent hyperplane elements to $x^2=$ constant.

For $\mathrm{d}x^n$ to be the dual of $\frac{\partial}{\partial x^n}$, it must obey $\mathrm{d}x^n\cdot\frac{\partial}{\partial x^k}=\delta^n_k$. In component form, this is easily captured as a row of all 0's with the exception of the nth spot, where a 1 is placed. Such a 'normal vector' can be visualized as defining a hyperplane orthogonal to it, wherin the components to x^n are constant.

12.6. Show, by use of the chain rule (see §10.3), that this expression for $\alpha\cdot\xi$ is consistent with $\mathrm{d}\Phi\cdot\xi=\xi(\Phi)$.

Noting that in the case of $\alpha=\mathrm{d}\Phi$, $\alpha_n=\frac{\partial\Phi}{\partial x^n}$, we see

$$
\begin{aligned}
\mathrm{d}\Phi\cdot\xi &= \xi(\Phi)\\
&= (\xi^1\frac{\partial}{\partial x^1}+\cdots+\xi^n\frac{\partial}{\partial x^n})(\Phi)\\
&= \xi^1\frac{\partial\Phi}{\partial x^1}+\cdots+\xi^n\frac{\partial\Phi}{\partial x^n}\\
&= \xi^1\mathrm{d}\Phi_1+\cdots+\xi^n\mathrm{d}\Phi_n
\end{aligned}
$$

12.7. Explain why this works.

All quantities may be decomposed into their symmetric and antisymmetric parts. If a quantity is equal to its antisymmetric part, it is entirely antisymmetric.

12.8. Justify the fact that $\psi\wedge\chi=\alpha\wedge\cdots\wedge\gamma\wedge\lambda\wedge\cdots\nu$ where $\psi=\alpha\wedge\cdots\wedge\gamma$, $\chi=\lambda\wedge\cdots\wedge\nu$.

12.9. Show this explicitly, explaining how to treat the limits, for a definite integral $\int_a^b \alpha$.

Suppose we were to make a change of variables from x to X, where the latter is related to the former by $X = g(x)$. Then we have

$$\alpha = f(x)\mathrm{d}x \to \alpha = f(g^{-1}(X))\frac{1}{g'(g^{-1}(X))}\mathrm{d}X.$$

Where we have made use of the fact that $\mathrm{d}X = g'(x)\mathrm{d}x$ and $x = g^{-1}(X)$. Noting that $(g^{-1})'(X) = \frac{1}{g'(g^{-1}(X))}$, this becomes

$$\alpha = f(g^{-1}(X))(g^{-1})'(X)\mathrm{d}X.$$

This, however, is our original α. To see this, simply replace $g^{-1}(X)$ with x and note that $x'\mathrm{d}X = \mathrm{d}x$. Ergo, our expression remains unchanged under a change of variables.

If we were to place α within a definite integral and carry out the operation, the limits a and b would have been changed to $g(a)$ and $g(b)$.

12.10. Let $G = \int_{-\infty}^{\infty} e^{-x^2}\mathrm{d}x$. Explain why $G^2 = \int_{\mathbb{R}^2} e^{-(x^2+y^2)}\mathrm{d}x \wedge \mathrm{d}y$ and evaluate this by changing to polar coordinates (r, θ). (§5.1). Hence prove $G = \sqrt{\pi}$.

Squaring G gives

$$\left(\int_{-\infty}^{\infty} e^{-x^2}\mathrm{d}x\right)\left(\int_{-\infty}^{\infty} e^{-y^2}\mathrm{d}y\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)}\mathrm{d}x \wedge \mathrm{d}y$$

where we have moved the rearranged the integral signs by virtue of both integrands' analyticity and combined the one-forms through a wedge product to express a well-defined two dimensional integral.

In polar coordinates, $x = r\cos\theta$ and $y = r\sin\theta$; these are the equivalent of $g^{-1}(X)$ given in the last answer. Additionally, $\mathrm{d}x = \mathrm{d}r\cos\theta - r\sin\theta\mathrm{d}\theta$ and $\mathrm{d}y = \mathrm{d}r\sin\theta + r\cos\theta\mathrm{d}\theta$, so

$$\mathrm{d}x \wedge \mathrm{d}y = r\cos^2\theta\mathrm{d}r \wedge \mathrm{d}\theta - r\sin^2\mathrm{d}\theta \wedge \mathrm{d}r = r\mathrm{d}r \wedge \mathrm{d}\theta.$$

The limits for the integral over $\mathrm{d}r$ are from 0 to ∞, while the limits for the integral over θ are from 0 to 2π. We get

$$\begin{aligned} G^2 &= \int_0^{2\pi}\int_0^{\infty} re^{-r^2}\mathrm{d}r \wedge \mathrm{d}\theta \\ &= \frac{1}{2}\int_0^{2\pi}\int_0^{\infty} e^{-u}\mathrm{d}u \wedge \mathrm{d}\theta \\ &= \frac{1}{2}\int_0^{2\pi}\mathrm{d}\theta \\ &= \pi \end{aligned}$$

Which implies $G = \sqrt{\pi}$.

12.11. Using the above relations, show that $\mathrm{d}(A\mathrm{d}x + B\mathrm{d}y) = (\partial B/\partial x - \partial A/\partial y)\mathrm{d}x \wedge \mathrm{d}y$.

We have

$$\begin{aligned}
\mathrm{d}(A\mathrm{d}x + B\mathrm{d}y) &= \mathrm{d}(A\mathrm{d}x) + \mathrm{d}(B\mathrm{d}y) \\
&= \Big(\frac{\partial A}{\partial x}\mathrm{d}x + \frac{\partial A}{\partial y}\mathrm{d}y\Big) \wedge \mathrm{d}x + \Big(\frac{\partial B}{\partial x}\mathrm{d}x + \frac{\partial B}{\partial y}\mathrm{d}y\Big) \wedge \mathrm{d}y \\
&= \frac{\partial A}{\partial y}\mathrm{d}y \wedge \mathrm{d}x + \frac{\partial B}{\partial x}\mathrm{d}x \wedge \mathrm{d}y \\
&= \Big(\frac{\partial B}{\partial x} - \frac{\partial A}{\partial y}\Big)\mathrm{d}x \wedge \mathrm{d}y
\end{aligned}$$

12.12. Why?

Because, by the third axiom, $\mathrm{d}(\mathrm{d}\Phi) = 0$ for all p-forms Φ, including 0-forms.

12.13. Assuming the result of Exercise [12.10], prove the Poincare lemma for $p = 1$.

12.14. Show directly that all the 'axioms' for exterior derivative are satisfied by this coordinate definition.

12.15. Show that this coordinate definition gives the same quantity $\mathrm{d}\alpha$, whatever choice of coordinates is made, where the transformation of the components $\alpha_{r \cdots t}$ of a form is defined by the requirement that the form α itself be unaltered by coordinate change. *Hint*: Show that this transformation is identical with the passive transformation of $[0, p]$-valent tensor components, as given in §13.8.

12.16. Confirm the equivalence of all these conditions for simplicity; prove the sufficiency of $\alpha_{[rs}\alpha_{u]v} = 0$ in the case $p = 2$. (*Hint*: contract this expression with two vectors.)

12.17. By representing a rotation in ordinary 3-space as a vector pointing along the rotation axis of length equal to the angle of rotation, show that the topology of R can be described as a solid ball (of radius π) bounded by an ordinary sphere, where each point of the sphere is identified with its antipodal point. Give a direct argument to show why a closed loop representing a 2π-rotation cannot be continuously deformed to a point.

With rotations represented this way (as a vector in 3-space), restrict the available angle of rotation between $-\pi \leq \theta \leq \pi$. Clearly, the space of rotations is given by a solid ball of radius π (representing those rotations less than π) combined with its boundary. All points on this boundary correspond to rotations about an angle of π. But, as rotations about an angle of π are indistinguishable from rotations about an angle of $-\pi$, antipodal points on this boundary are to be identified with one another.

Now, consider a path in this space representing a 2π rotation. Starting at the origin, this is a line to the boundary, then from the antipodal point of *that* point back to the origin. Because of these identified antipodal points, such a line cannot be deformed to a point: any movement of one of the boundary points moves the other.

If, on the other hand, such a path is taken *twice*, then the path of one pair of antipodal points can be moved around the sphere to lie opposed to the other, cancelling them out and leaving a point.

13 Symmetry groups

13.1. Show that if we just assume $1a = a$ and $a^{-1}a = 1$ for all a, together with associativity $a(bc) = (ab)c$, then $a1 = a$ and $aa^{-1} = 1$ can be *deduced*. (*Hint*: Of course a is not the only element asserted to have an inverse.) Show why, on the other hand $a1 = a$, $a^{-1}a = 1$, and $a(bc) = (ab)c$ are insufficient.

To prove $aa^{-1} = 1$, consider

$$a^{-1}aa^{-1}a = 1 \cdot 1$$
$$a^{-1}(aa^{-1}a) = 1$$

where the second line implies $aa^{-1}a = a$, or, by associativity, $(aa^{-1})a = a$. By our first assumption, this implies $aa^{-1} = 1$, as $1a = a$. Using this we may now write $aa^{-1}a = a(a^{-1}a) = a1 = a$.

If we were to begin with $a1 = a$ instead of $1a = a$, we would not have been able to identify aa^{-1} with 1.

13.2. Explain why any vector space is an Abelian group—called an *additive* Abelian group—where the group 'multiplication' operation is the 'addition' operation of the vector space.

Clearly, vector spaces obey associativity under addition. The identity element is the zero vector, and the inverse of each vector is its negative. That vector spaces are Abelian follows from the commutative property of addition.

13.3. Verify these relations (bearing in mind that Ci stands for 'the operation $i\times$, followed by the operation C, etc.) (*Hint*: You can check the relations by just confirming their effects on 1 and i. Why?)

As Penrose points out, we can confirm these relations by checking their effects on 1 and i, as these form a complete basis for all complex numbers. We have

$$Ci1 = Ci = -i$$
$$(-i)C1 = (-i)1 = -i$$
$$Cii = C(-1) = -1$$
$$(-i)Ci = (-i)(-i) = -1$$

for the first operation. For the second operation, we see

$$C(-1)1 = C(-1) = -1$$
$$(-1)C1 = (-1)1 = -1$$
$$C(-1)i = C(-i) = i$$
$$(-1)Ci = (-1)(-i) = i$$

Likewise, the third operation yields

$$
\begin{aligned}
C(-i)1 &= C(-i) = i \\
iC1 &= i1 = i \\
C(-i)i &= C1 = 1 \\
iCi &= i(-i) = 1
\end{aligned}
$$

and the fourth operation gives

$$
\begin{aligned}
CC1 &= C1 = 1 \\
11 &= 1 \\
CCi &= C(-i) = i \\
1i &= i
\end{aligned}
$$

13.4. Show this.

The second equation is exactly the fourth multiplication rule, while the third equation is exactly the first (because $Ci = i^3C = i^2iC = (-1)iC = (-i)C$). To extract the second rule, multiply the third equation by i and substitute in the original equation,

$$Ci^2 = C(-1) = i^3Ci = i^3i^3C = i^4i^2C = 1(-1)C = (-1)C.$$

For the third, simply multiple the third equation by i twice from the right,

$$Ci^3 = C(-i) = i^3Ci^2 = i^6Ci = i^9C = iC.$$

13.5. Verify that all these in this paragraph are subgroups (and bear in mind Note 13.4).

Let us consider each set in turn. For $\{1, i, -1, -i\}$, we see the inclusion of the identity, 1, and the inclusion of all inverses: $(1)(1) = 1$, $i(-i) = 1$, $(-1)(-1) = 1$, and $(-i)(i) = 1$. Finally, this set is closed under the group operation,

	1	i	-1	$-i$
1	1	i	-1	$-i$
i	i	-1	$-i$	1
-1	-1	$-i$	1	i
$-i$	$-i$	1	i	-1

For the second, notice that it, too contains the identity. As for inverses, we have $(1)(1) = 1$, $(-1)(-1) = 1$, $CC = 1$, and $(-C)(-C) = 1$. This, too, is closed under the group operation,

	1	-1	C	$-C$
1	1	-1	C	$-C$
-1	-1	1	$-C$	C
C	C	$-C$	1	-1
$-C$	$-C$	C	-1	1

Finally, $\{1, -1\}$ clearly contains the identity, both elements' inverses (themselves), and is closed under multiplication.

13.6. Check these assertions, and find two more non-normal subgroups, showing that there are no further ones.

Maintaining the placement of the arrow in Fig 13.3 (a) can be done through either complex conjugation (as it lies on the real line) or the identity transformation. Keeping the arrow in Fig 13.3 (b) fixed amounts to applying either the identity transformation, or performing a $\pi/2$ rotation counter-clockwise, i.e. a multiplication by i, prior to taking the arrow's complex conjugate.

The two other non-normal subgroups are $\{1, iC\}$ and $\{1, -C\}$. There are no other possible subgroups containing just two members, and the other subgroups have already been found to be normal.

13.7. Show this. (*Hint*: which *sets* of rotations can be rotation-invariant?)

13.8. Verify this and show that the axioms fail if S is not normal.

Because $\mathcal{S}$ is a normal subgroup, h commutes with $\mathcal{S}$, allowing us to combine g with h before commuting it with $\mathcal{S}$ once again to arrive at the right-hand side. Furthermore, since $gh \in \mathcal{G}$, $(gh)\mathcal{S} = \mathcal{S}(gh)$, showing that our new group is closed under this operation.

The identity, in the case of a factor group, is the normal subgroup $\mathcal{S}$. To see this, consider the first axiom. Since $\mathcal{S}$ is closed under the group operation, any element in $\mathcal{S}$ times itself results in the same group. With this in mind, the axiom that fails for non-normal subgroups is the second one, as now

$$(\mathcal{S}a)(\mathcal{S}a^{-1}) \neq \mathcal{S}(aa^{-1})\mathcal{S} = \mathcal{S}.$$

13.9. Explain why the number of elements in $\mathcal{G}/\mathcal{S}$, for any finite subgroup $\mathcal{S}$ of $\mathcal{G}$, is the order of $\mathcal{G}$ divided by the order of $\mathcal{S}$.

In order for the factor group to 'contain' all the elements of our original (in a set theoretic sense), we must have a minimum of n/m elements, where m is the order of $\mathcal{S}$ and n is the order of $\mathcal{G}$. This is because, for each g that produces a new $\mathcal{S}g$, we get m more elements. However, we can also see that there can be no more than this number of elements in our factor group, as then there would be some two elements that are not distinct. Ergo, the order of our factor group is n/m.

13.10. Verify that $\mathcal{G} \times \mathcal{H}$ is a group, for any two groups $\mathcal{G}$ and $\mathcal{H}$, and that we can identify the factor group $(\mathcal{G} \times \mathcal{H})/\mathcal{G}$ with $\mathcal{H}$.

As the two groups are, in some sense, independent of one another, their pairing satisfies the group axioms by virtue of their own legality. The difference being that, now, our identity is given by (e_g, e_h) and the inverse of any pair, (g_1, h_1), takes the form (g_1^{-1}, h_1^{-1}).

If we take $(\mathcal{G} \times \mathcal{H})/\mathcal{G}$ to represent a group whose elements are given by $(\mathcal{G}g_1, \mathcal{G}h_1)$, then we can identify this with $\mathcal{H}$. This is because $\mathcal{G}g_1 = \mathcal{G}$ for all g_1, while $\mathcal{G}h_1$ preserves the identifying nature of h_1 for all h_1.

13.11. Show how this equation, giving the points of unit distance from O, follows from the Pythagorean theorem of §2.1.

> The distance to a point on S^2 is given by the Pythagorean theorem as
>
> $$c^2 + z^2 = 1$$
>
> where c is a line segment from the origin to a point in the unit circle and z is the height of the sphere's surface at that point. We can ascertain c from the Pythagorean theorem as well, as
>
> $$c^2 = x^2 + y^2.$$
>
> Substituting this into the first relation yields
>
> $$x^2 + y^2 + z^2 = 1.$$

13.12. Can you explain why? Just do this in the 2-dimensional case, for simplicity.

> A linear transformation takes the basis vectors of one space into those of another. Once the transformation is applied, we can decompose the new space's basis into our old basis, allowing us to write each new coordinate in terms of a linear combination of the old ones.

13.13. Show this explicitly in the 3-dimensional case.

> In the 3-dimensional case, we have
>
> $$x^a \to T^a{}_b x^b = T^a{}_1 x^1 + T^a{}_2 x^2 + T^a{}_3 x^3$$
>
> where a runs over all three dimensions. Here, the linear transformation T^a_b is identified with the set of coefficients needed to express the new coordinates in terms of the old ones.

13.14. Write this all out in full, explaining how this expresses $x^a \to T^a{}_b x^b$.

> This is essentially a more explicit version of the previous exercise. We have
>
> $$\begin{aligned} x^1 &\to T^1{}_a x^a = T^1{}_1 x^1 + T^1{}_2 x^2 + T^1{}_3 x^3 \\ x^2 &\to T^2{}_a x^a = T^2{}_1 x^1 + T^2{}_2 x^2 + T^2{}_3 x^3 \\ x^3 &\to T^3{}_a x^a = T^3{}_1 x^1 + T^3{}_2 x^2 + T^3{}_3 x^3 \end{aligned}$$
>
> The expression given in the book expresses vector-matrix multiplication, where each row of $\mathbf{x}$ is used to multiply the corresponding column of $\mathbf{T}$.

13.15. What is this relation between **R**, **S**, and **T**, written out explicitly in terms of the elements of a 3×3 square arrays of components. You may recognize this, the normal law for 'multiplication of matrices', if this is familiar to you.

> The component of **R** in its ith row and jth column can be explicitly calculated by summing the element-wise multiplication of the ith row of **S** with the jth row of **T**.

13.16. Verify.

> In the case of either $T^a{}_b\delta^b_c$ or $\delta^a_b T^b{}_c$, we sum over all possible indices of b. Since the resulting quantities are 0 except for the term with $b = c$ in the first expression and $b = a$ in the second, we are left with $T^a{}_c$.

13.17. Why? Show that this would happen, in particular, if the array of components has an entire column of 0s or two identical columns. Why does this also hold if there are two identical rows? *Hint*: For this last part, consider the determinant condition below.

> Geometrically, we can see why such a condition would correspond to a mapping of a space to another one of smaller dimension: if a vector is orthogonal to the space which we are mapping to, then its projection to that space, **T**, yields 0.
>
> If an array **T**'s ith column is filled with zeros, a vector **v** having null entries with the exception of its ith row will give $\mathbf{Tv} = 0$. If an array **R** has two identical columns, i and j, then a vector **w** having null entries everywhere except its ith and jth rows, which have values opposite to one another, gives $\mathbf{Rw} = 0$.

13.18. Show why, not using explicit expressions.

> If a transformation is non-singular, the only vector mapped to 0 will be the trivial case. This, in combination with the linearity of the transformation, ensures that every vector will be mapped to just one other, allowing us to invert the process.

13.19. Prove directly, using the diagrammatic relations given in Fib. 12.18, that this definition gives $\mathbf{TT}^{-1} = \mathbf{I} = \mathbf{T}^{-1}\mathbf{T}$.

13.20. Explain this, and give the full algebraic rules for rectangular matrices.

> In matrix multiplication, the elements of the ith column of the right matrix are used to weight a sum of the columns of the left matrix, resulting in a new column corresponding to the ith such entity in the new matrix. Therefore, the number of rows of the right matrix must match the number of columns in the left one. Explicitly, if the left matrix is given by **A** and has dimensions $a \times b$ and the right matrix is given by **B** and has dimensions $c \times d$, then for **AB** to be well-defined we require $b = c$.
>
> The standard associative and commutative laws hold for addition as long as **A** and **B** have the same dimensions,

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$
$$\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C}$$

For multiplication, $\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}$ provided, given the respective dimensions of $a \times b$, $c \times d$, and $e \times f$, $b = c$ and $d = e$.

The distributive law holds as well, with the obvious generalizations of the above properties.

13.21. Derive these from the expression of Fig. 13.8a.

The diagrams instruct us to compute the product of the complete anti-symmetrization of the matrix, then divide by $n!$. Working with explicit indices (where we are not using the Einstein summation convention), we have

$$\begin{aligned}
\frac{1}{2!}T^{[1}{}_{[1}T^{2]}{}_{2]} &= \frac{1}{2}(T^{1}{}_{[1}T^{2}{}_{2]} - T^{2}{}_{[1}T^{1}{}_{2]}) \\
&= \frac{1}{2}(T^{1}{}_{1}T^{2}{}_{2} - T^{1}{}_{2}T^{2}{}_{1} - T^{2}{}_{1}T^{1}{}_{2} + T^{2}{}_{2}T^{1}{}_{1}) \\
&= \frac{1}{2}(2T^{1}{}_{1}T^{2}{}_{2} - 2T^{1}{}_{2}T^{2}{}_{1}) \\
&= T^{1}{}_{1}T^{2}{}_{2} - T^{1}{}_{2}T^{2}{}_{1} \\
&= ad - bc
\end{aligned}$$

The case for the 3×3 matrix can be computed similarly. Because of the calculation's tediousness, it has been omitted here.

13.22. Show why these hold.

If this answer is to use Penrose's graphical notation, we see the first expression's validity by writing out the two anti-symmetric tensors and joining them along their anti-symmetrization line: this is clearly a product of Kronecker deltas with one of the indices anti-symmetrized. The factor of $n!$ is included for the case that all upper indices are contracted with their respective lower indices.

The second expression can be written as the product of a Kronecker delta with two contracted anti-symmetric tensors of rank $(n-1)$, yielding a factor of $(n-1)!$.

13.23. Show this.

Let us use abstract index notation, denoting $\mathbf{A} = A^{i}{}_{j}$ and $\mathbf{B} = B^{i}{}_{j}$. The trace of $\mathbf{C} = \mathbf{A} + \mathbf{B}$ is $C^{a}{}_{a} = A^{a}{}_{a} + B^{a}{}_{a}$, which is the sum of the traces of $\mathbf{A}$ and $\mathbf{B}$.

13.24. Show this.

Appealing to the given expression for the determinant, we see

$$\det(\mathbf{I}+\epsilon\mathbf{A}) = \frac{1}{n!}\epsilon^{ab\ldots d}(\delta^e_a + \varepsilon A^e{}_a)(\delta^f_b + \varepsilon A^f{}_b)\cdots(\delta^h_d + \varepsilon A^h{}_d)\epsilon_{ef\ldots h}$$

The product of all the delta functions will give us a term of $\varepsilon^{ab\ldots d}\varepsilon_{ab\ldots d}/n! = n!/n! = 1$. All other products (of which there will be n) will be composed of one A and $(n-1)$ delta functions, as all higher order of epsilon are taken to be 0. That is,

$$\begin{aligned}\det(\mathbf{I}+\epsilon\mathbf{A}) &= 1 + \frac{\varepsilon}{n!}\epsilon^{ab\ldots d}(A^e{}_a\delta^f_b\cdots\delta^h_d)\epsilon_{ef\ldots h} + \frac{\varepsilon}{n!}\epsilon^{ab\ldots d}(\delta^e_a A^f{}_b\cdots\delta^h_d)\epsilon_{ef\ldots h} + \cdots\\
&= 1 + \frac{\varepsilon}{n!}\epsilon^{ab\ldots d}\epsilon_{eb\ldots d}A^e{}_a + \frac{\varepsilon}{n!}\epsilon^{ab\ldots d}\epsilon_{af\ldots d}A^f{}_b + \cdots\\
&= 1 + \frac{\varepsilon}{n!}(n-1)!\delta^a_e A^e{}_a + \frac{\varepsilon}{n!}(n-1)!\delta^b_f A^f{}_b + \cdots\\
&= 1 + \frac{\varepsilon}{n}A^a{}_a + \frac{\varepsilon}{n}A^b{}_b + \cdots\\
&= 1 + \varepsilon A^a{}_a\\
&= 1 + \varepsilon\mathrm{trace}(\mathbf{A})\end{aligned}$$

13.25. Establish the expression for this. *Hint*: Use the 'canonical form' for a matrix in terms of its eigenvalues−as described in §13.5−assuming first that these eigenvalues are unequal (and see Exercise [13.27]). Then use a general argument to show that the equality of some eigenvalues cannot invalidate identities of this kind.

Simpler than the hint Penrose gives, consider

$$\begin{aligned}\det(e^{\mathbf{A}}) &= \det\left(\lim_{n\to\infty}\left(\mathbf{I}+\frac{\mathbf{A}}{n}\right)^n\right)\\
&= \lim_{n\to\infty}\det\left(\left(\mathbf{I}+\frac{\mathbf{A}}{n}\right)^n\right)\\
&= \lim_{n\to\infty}\left(\det\left(\mathbf{I}+\frac{\mathbf{A}}{n}\right)\right)^n\\
&= \lim_{n\to\infty}\left(1+\frac{\mathrm{trace}(\mathbf{A})}{n}\right)^n\\
&= e^{\mathrm{trace}(\mathbf{A})}\end{aligned}$$

where the third equality hold because the determinant of a product is the product of the constituent determinants, and the fourth equality holds from our previous result. This is not a proof, as we have neglected to address the swapping of the determinant with the limit in the first equality, but it does, as the exercise requests, establish the expression.

13.26. See if you can express the coefficients of this polynomial in diagrammatic form. Work them out for $n = 2$ and $n = 3$.

13.27. Show that $\det\mathbf{T} = \lambda_1\lambda_2\ldots\lambda_n$, $\mathrm{trace}\,\mathbf{T} = \lambda_1+\lambda_2+\cdots+\lambda_n$.

We know that

$$\det(\mathbf{T} - \lambda \mathbf{I}) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda)$$

Setting $\lambda = 0$ reduces this to $\det \mathbf{T} = \lambda_1 \lambda_2 \dots \lambda_n$. For the second equality, let us consider the effect of $e^{\mathbf{T}}\mathbf{v}$, where $\mathbf{T}v = \lambda \mathbf{v}$. We have

$$\begin{aligned} e^{\mathbf{T}}\mathbf{v} &= (\mathbf{I} + \mathbf{T} + \frac{\mathbf{T}}{2} + \cdots)\mathbf{v} \\ &= (1 + \lambda + \frac{\lambda}{2} + \cdots)\mathbf{v} \\ &= e^{\lambda}\mathbf{v} \end{aligned}$$

This shows that, if λ is an eigenvalue of $\mathbf{T}$, e^{λ} is an eigenvalue of $e^{\mathbf{T}}$. Combining this with our previous result and the answer to 13.25 gives

$$\begin{aligned} \det e^{\mathbf{T}} &= e^{\lambda_1} e^{\lambda_2} \cdots e^{\lambda_n} \\ &= e^{\lambda_1 + \lambda_2 + \cdots + \lambda_n} \\ &= e^{\operatorname{trace} \mathbf{T}} \end{aligned}$$

Taking the logarithm of the second and third lines gives

$$\lambda_1 + \lambda_2 + \cdots + \lambda_n = \operatorname{trace} \mathbf{T}$$

13.28. Show this.

13.29. Explain this notation.

The array given by

$$(\delta_j^1, \delta_j^2, \delta_j^3, \dots, \delta_j^n)$$

will have zeros in every spot except the jth, at which point it will have 1. This is exactly the set of components corresponding to the jth basis vector.

13.30. Why? What are the components of $\mathbf{e}_i$ in the $\mathbf{f}$ basis?

Applying $\mathbf{T}$ to an arbitrary basis vector $\mathbf{e}$ (using the form given in the previous exercise) yields

$$
\begin{aligned}
\mathbf{f}_j &= \mathbf{T}\mathbf{e}_j \\
&= T^i{}_j(\delta_i^1, \delta_i^2, \delta_i^3, \dots, \delta_i^n) \\
&= (T^i{}_j\delta_i^1, T^i{}_j\delta_i^2, T^i{}_j\delta_i^3, \dots, T^i{}_j\delta_i^n) \\
&= (T^1{}_j, T^2{}_j, T^3{}_j, \dots, T^n{}_j)
\end{aligned}
$$

By applying the inverse of $\mathbf{T}$ to our original transformation, we see $\mathbf{e}j = \mathbf{T}^{-1}\mathbf{f}_j$, and so

$$
\mathbf{e}_j = \left((T^{-1})^1{}_j, (T^{-1})^2{}_j, (T^{-1})^3{}_j, \dots, (T^{-1})^n{}_j\right)
$$

13.31. See if you can prove this. *Hint*: For each eigenvalue of multiplicity r, choose r linearly independent eigenvectors. Show that a linear relation between vectors of this entire collection leads to a contradiction when this relation is pre-multiplied by $\mathbf{T}$, successively.

13.32. Show this. *Hint*: Label each column of the representing matrix by a separate element of the finite group $\mathcal{G}$, and also label each row by the corresponding group element. Place a 1 in any position in the matrix for which a certain relation holds (find it!) between the element of $\mathcal{G}$ representing the row, that labelling the column, and the element of $\mathcal{G}$ that this particular matrix is representing. Place a 0 whenever this relation does not hold.

Consider an $n \times n$ matrix, taken to represent a group element. If we do as Penrose suggests, we can think of each matrix as encoding the multiplications rules for each element. To be explicit, consider an element g_i. The matrix representing this will have a 1 where the column labeled g_i intersects e. The other columns will describe how g_i acts on g_j, where exactly one 1 in each column will give the mapping between that column's label and the row's label.

13.33. Why is this expression just the *identity* group element when a and b commute.

Because, in that case, we have

$$
\begin{aligned}
aba^{-1}b^{-1} &= abb^{-1}a^{-1} \\
&= aea^{-1} \\
&= aa^{-1} \\
&= e
\end{aligned}
$$

where e denotes the group identity element.

13.34. Spell out this 'order ϵ^2' calculation.

Expanding out the last two terms in their power series, we see

$$\begin{aligned}(\mathbf{I}+\varepsilon\mathbf{A})^{-1}(\mathbf{I}+\varepsilon\mathbf{B})^{-1} &= (\mathbf{I}-\varepsilon\mathbf{A}+\varepsilon^2\mathbf{A}^2+\mathcal{O}(\varepsilon^3))(\mathbf{I}-\varepsilon\mathbf{B}+\varepsilon^2\mathbf{B}^2+\mathcal{O}(\varepsilon^3))\\ &= \mathbf{I}-\varepsilon\mathbf{A}-\varepsilon\mathbf{B}+\varepsilon^2\mathbf{A}^2+\varepsilon^2\mathbf{B}^2+\varepsilon^2\mathbf{AB}+\mathcal{O}(\varepsilon^3)\\ &= \mathbf{I}-\varepsilon(\mathbf{A}+\mathbf{B})+\varepsilon^2(\mathbf{A}^2+\mathbf{AB}+\mathbf{B}^2)+\mathcal{O}(\varepsilon^3)\end{aligned}$$

Now consider the first two terms,

$$(\mathbf{I}+\varepsilon\mathbf{A})(\mathbf{I}+\varepsilon\mathbf{B}) = \mathbf{I}+\varepsilon(\mathbf{A}+\mathbf{B})+\varepsilon^2\mathbf{AB}$$

If we then multiply these two together, we find

$$\begin{aligned}&\mathbf{I}-\varepsilon(\mathbf{A}+\mathbf{B})+\varepsilon^2(\mathbf{A}+\mathbf{AB}+\mathbf{B}^2)+\varepsilon(\mathbf{A}+\mathbf{B})-\varepsilon^2(\mathbf{A}+\mathbf{B})^2+\varepsilon^2\mathbf{AB}\\ =&\,\mathbf{I}+\varepsilon^2(\mathbf{A}^2+\mathbf{AB}+\mathbf{B}^2-\mathbf{A}^2-\mathbf{AB}-\mathbf{BA}-\mathbf{B}^2+\mathbf{AB})\\ =&\,\mathbf{I}+\varepsilon^2(\mathbf{AB}-\mathbf{BA})\end{aligned}$$

13.35. Show all this.

For distributivity, we easily find

$$\begin{aligned}[\mathbf{A}+\mathbf{B},\mathbf{C}] &= (\mathbf{A}+\mathbf{B})\mathbf{C}-\mathbf{C}(\mathbf{A}+\mathbf{B})\\ &= \mathbf{AC}+\mathbf{BC}-\mathbf{CA}-\mathbf{CB}\\ &= (\mathbf{AC}-\mathbf{CA})+(\mathbf{BC}-\mathbf{CB})\\ &= [\mathbf{A},\mathbf{C}]+[\mathbf{B},\mathbf{C}]\end{aligned}$$

and

$$\begin{aligned}[\lambda\mathbf{A},\mathbf{B}] &= \lambda\mathbf{AB}-\mathbf{B}\lambda\mathbf{A}\\ &= \lambda\mathbf{AB}-\lambda\mathbf{BA}\\ &= \lambda(\mathbf{AB}-\mathbf{BA})\\ &= \lambda[\mathbf{A},\mathbf{B}]\end{aligned}$$

The antisymmetric nature of the commutator is similarly proved,

$$\begin{aligned}[\mathbf{A},\mathbf{B}] &= \mathbf{AB}-\mathbf{BA}\\ &= -(\mathbf{BA}-\mathbf{AB})\\ &= -[\mathbf{B},\mathbf{A}]\end{aligned}$$

This property can then be used to prove distributivity in the second argument, as

$$\begin{aligned}[\mathbf{A}, \mathbf{C} + \mathbf{D}] &= -[\mathbf{C} + \mathbf{D}, \mathbf{A}] \\ &= -[\mathbf{C}, \mathbf{A}] - [\mathbf{D}, \mathbf{A}] \\ &= [\mathbf{A}, \mathbf{C}] + [\mathbf{A}, \mathbf{D}]\end{aligned}$$

and

$$\begin{aligned}[\mathbf{A}, \lambda\mathbf{B}] &= -[\lambda\mathbf{B}, \mathbf{A}] \\ &= -\lambda[\mathbf{B}, \mathbf{A}] \\ &= \lambda[\mathbf{A}, \mathbf{B}]\end{aligned}$$

The Jacobi identity is slightly more tedious,

$$\begin{aligned}&[\mathbf{A}, [\mathbf{B}, \mathbf{C}]] + [\mathbf{B}, [\mathbf{C}, \mathbf{A}]] + [\mathbf{C}, [\mathbf{A}, \mathbf{B}]] \\ =& [\mathbf{A}, \mathbf{BC}] - [\mathbf{A}, \mathbf{CB}] + [\mathbf{B}, \mathbf{CA}] - [\mathbf{B}, \mathbf{AC}] + [\mathbf{C}, \mathbf{AB}] - [\mathbf{C}, \mathbf{BA}] \\ =& \mathbf{ABC} - \mathbf{BCA} - \mathbf{ACB} + \mathbf{CBA} + \mathbf{BCA} - \mathbf{CAB} \\ &\quad - \mathbf{BAC} + \mathbf{ACB} + \mathbf{CAB} - \mathbf{ABC} - \mathbf{CBA} + \mathbf{BAC} \\ =& 0\end{aligned}$$

13.36. Show this.

To show antisymmetry, observe

$$[\mathbf{E}_\alpha, \mathbf{E}_\beta] = \gamma_{\alpha\beta}{}^{\chi}\mathbf{E}_\chi = -[\mathbf{E}_\beta, \mathbf{E}_\alpha] = -\gamma_{\beta\alpha}{}^{\chi}\mathbf{E}_\chi.$$

which implies $\gamma_{\alpha\beta}{}^{\chi} = -\gamma_{\beta\alpha}^{\chi}$. To show the constraints imposed by the Jacobi identity, consider

$$\begin{aligned}&[\mathbf{E}_\alpha, [\mathbf{E}_\beta, \mathbf{E}_\lambda]] + [\mathbf{E}_\beta, [\mathbf{E}_\lambda, \mathbf{E}_\alpha]] + [\mathbf{E}_\lambda, [\mathbf{E}_\alpha, \mathbf{E}_\beta]] \\ =& [\mathbf{E}_\alpha, \gamma_{\beta\lambda}{}^{\chi}\mathbf{E}_\chi] + [\mathbf{E}_\beta, \gamma_{\lambda\alpha}{}^{\chi}\mathbf{E}_\chi] + [\mathbf{E}_\lambda, \gamma_{\alpha\beta}{}^{\chi}\mathbf{E}_\chi] \\ =& \gamma_{\beta\lambda}{}^{\chi}[\mathbf{E}_\alpha, \mathbf{E}_\chi] + \gamma_{\lambda\alpha}{}^{\chi}[\mathbf{E}_\beta, \mathbf{E}_\chi] + \gamma_{\alpha\beta}{}^{\chi}[\mathbf{E}_\lambda, \mathbf{E}_\chi] \\ =& (\gamma_{\beta\lambda}{}^{\chi}\gamma_{\alpha\chi}{}^{\xi} + \gamma_{\lambda\alpha}{}^{\chi}\gamma_{\beta\chi}{}^{\xi} + \gamma_{\alpha\beta}{}^{\chi}\gamma_{\lambda\chi}{}^{\xi})\mathbf{E}_\xi \\ =& 0\end{aligned}$$

Swapping the indices on the inner commutators of the first line yields a similar expression,

$$(-\gamma_{\lambda\beta}{}^{\chi}\gamma_{\alpha\chi}{}^{\xi} - \gamma_{\alpha\lambda}{}^{\chi}\gamma_{\beta\chi}{}^{\xi} - \gamma_{\beta\alpha}{}^{\chi}\gamma_{\lambda\chi}{}^{\xi})\mathbf{E}_\xi = 0$$

adding these together gives

$$\gamma_{[\alpha\beta}{}^{\chi}\gamma_{\lambda]\chi}{}^{\xi} = 0.$$

13.37. Why?

The inverse of a matrix can be defined as the unique matrix that, when multiplied by the initial matrix (from either the left or right), gives the identity. In the case of a product of two matrices, we arrive at the identity via

$$(\mathbf{AB})(\mathbf{B}^{-1}\mathbf{A}^{-1}) = \mathbf{ABB}^{-1}\mathbf{A}^{-1} = \mathbf{AIA}^{-1} = \mathbf{AA}^{-1} = \mathbf{I}..$$

By the uniqueness of the inverse, $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

13.38. Why this number?

Because our tensor is a multilinear function acting on $p+q$ spaces of dimension n, giving an overall dimensionality of n^{p+q}.

13.39. Show this.

Our tensor, acting as a multilinear map, acts on a collection of objects like so

$$Q_{a\ldots c}^{f\ldots h} x^a \cdots x^c x_f \cdots x_h.$$

Upon transforming our vector spaces and their duals as $x^a \to T^a{}_b x^b$ and $x_a \to S^b{}_a x_b$, this becomes

$$\hat{Q}_{a\ldots c}^{f\ldots h} T^a{}_{a'} x^{a'} \cdots T^c{}_{c'} x^{c'} S^{f'}{}_f x_{f'} \cdots S^{h'}{}_h x_{h'} = \hat{Q}_{a\ldots c}^{f\ldots h} T^a{}_{a'} \cdots T^c{}_{c'} S^{f'}{}_f \cdots S^{h'}{}_h x^{a'} \cdots x^{c'} x_{f'} \cdots x_{h'}.$$

where we have denoted the transformed $Q_{a\ldots c}^{f\ldots h}$ by $\hat{Q}_{a\ldots c}^{f\ldots h}$. Because everything has been contracted in our initial expression, this is a *geometric* quantity, i.e. it does not depend on coordinates. To satisfy this condition, our tensor Q must transform as the inverse as our vector spaces did,

$$Q_{a\ldots c}^{f\ldots h} \to S^{a'}{}_a \cdots S^{c'}{}_c T^f{}_{f'} \cdots T^h{}_{h'} Q_{a'\ldots c'}^{f'\ldots h'}.$$

13.40. Show this.

A tensor with two indices can be thought of (in component form) as a matrix. A symmetric matrix will be absolutely specified by its diagonal and all components to one side of it. The number of components to be found above the diagonal is given by $(n^2-n)/2$, or $n(n-1)/2$. This is because we are subtracting the number of diagonal components n from the total n^2 and dividing by two (so we're only counting one half of the matrix). The diagonal itself, as mentioned, gives n elements, and so a symmetric two-component tensor has

$$\frac{n(n-1)}{2} + n = \frac{n(n+1)}{2}.$$

components, corresponding to the same number of basis elements. Meanwhile, an antisymmetric matrix is specified by all the components to one side of the diagonal. (The diagonal itself contains only zeros, as these are the only number equal to their opposite.) Therefore, an antisymmetric two-component tensor has

$$\frac{n(n-1)}{2}.$$

basis elements.

13.41. Explain this.

Upon transforming a tensor, we are left with an object containing the same number of free indices as before. We can denote this new tensor by $\hat{Q}$, to which we can apply our usual symmetrization and anti-symmetrization schemes.

13.42. Show that the representation of $\frac{1}{1}$valent tensors is also reducible. *Hint*: Split any such tensor into a 'trace-free' part and a 'trace' part.

To split a tensor in such a way, we can remove its trace by

$$Q^a{}_b - \frac{1}{n}\delta^a_b Q^c{}_c.$$

This has this particular form because $Q^c{}_c$ is simply a number, and so we must multiply it by the identity to combine it with $Q^a{}_b$. The factor of $1/n$ is there for normalization purposes, as $\delta^a_a = n$, and so without it we would have $Q^a{}_a - \delta^a_a Q^c_c = (1-n)Q^c{}_c$.
We can simply add the trace back to obtain our original tensor, now split into two parts,

$$Q^a{}_b = (Q^a{}_b - \frac{1}{n}\delta^a_b Q^c{}_c) + \frac{1}{n}\delta^a_b Q^c{}_c.$$

13.43. Confirm this.

If every representative matrix has this form, then the components within the block matrix **A** only ever multiply other components in this block. The same is true for **B**. To see this explicitly, observe

$$\begin{pmatrix} \mathbf{A}_1 & \mathbf{C}_1 \\ \mathbf{O} & \mathbf{B}_1 \end{pmatrix} \begin{pmatrix} \mathbf{A}_2 & \mathbf{C}_2 \\ \mathbf{O} & \mathbf{B}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1\mathbf{A}_2 & \mathbf{A}_1\mathbf{C}_2 + \mathbf{C}_1\mathbf{B}_2 \\ \mathbf{O} & \mathbf{B}_1\mathbf{B}_2 \end{pmatrix}.$$

This implies that our group may be represented by either block matrix.

13.44. Why does $\kappa_{\alpha\beta} = \kappa_{\beta\alpha}$?

This is due to two facts: that the indicial forms of tensors commute with each other and dummy variables can be freely renamed. Explicitly, we have

$$\kappa_{\alpha\beta} = \gamma_{\alpha\zeta}{}^{\xi}\gamma_{\beta\xi}{}^{\zeta} = \gamma_{\beta\xi}{}^{\zeta}\gamma_{\alpha\zeta}{}^{\xi} = \gamma_{\beta\zeta}{}^{\xi}\gamma_{\alpha\xi}{}^{\zeta} = \kappa_{\beta\alpha}.$$

13.45. Use Note 13.18 to establish this.

Treating our tensor Q as a multilinear function, we have

$$\begin{aligned} Q(\hat{\mathbf{e}}^a, \ldots, \hat{\mathbf{e}}^c, \hat{\mathbf{e}}_p, \ldots, \hat{\mathbf{e}}_r) &= Q(t^a{}_d\mathbf{e}^d, \ldots, t^c{}_f\mathbf{e}^f, s^j{}_p\mathbf{e}_j, \ldots, s^l{}_r\mathbf{e}_l) \\ &= t^a{}_d \ldots t^c{}_f s^j{}_p \ldots s^l{}_r Q(\mathbf{e}^d, \ldots, \mathbf{e}^f, \mathbf{e}_j, \ldots, \mathbf{e}_l). \end{aligned}$$

13.46. Why?

Because δ^a_b is the identity transformation.

13.47. Why equivalent?

Because, by virtue of g_{ab} and g^{ab} being inverses of each other, the assumption that g_{ab} is symmetric allows us to show

$$\begin{aligned} g^{cd} &= g^{ac}g^{bd}g_{ab} \\ &= g^{ac}g^{bd}g_{ba} \\ &= g^{bd}g^{ac}g_{ba} \\ &= g^{ad}g^{bc}g_{ab} \\ &= g^{dc} \end{aligned}$$

where we have swapped the labels of a and d in the fourth line.

13.48. Can you confirm this characterization?

The key part of such an alternative characterization is that the given quantity must be greater than 0 for *every* vector we feed into it. Consider inserting the canonical basis of our space into this expression. This then yields

$$A_{ab}x^a x^b = \mathbf{e}_a \cdot \mathbf{e}_b = \delta^a_b.$$

This is the identity matrix, having a signature of all 1s.

13.49. Explain why.

13.50. Explain this. What is $\mathbf{T}^{-1}$ in the pseudo-orthogonal cases (defined in the next paragraph)?

Replacing the metric with the Kronecker delta gives

$$\delta_{ab}T^a{}_cT^b{}_d = \delta_{cd}.$$

Contracted indices are unaffected by an exchange in their position (upper or lower), and so we can rewrite this as

$$\delta_a^b T^a{}_c T_b{}^d = T^a{}_c T_a{}^d = T_a{}^d T^a{}_c = \delta_c^d.$$

This is the standard formula for matrix multiplication, with the first instance of T being the transpose of the second. That is,

$$\mathbf{T}^T\mathbf{T} = \mathbf{I}.$$

But this is exactly the relationship characterizing a matrix and its inverse, and so we have $\mathbf{T}^{-1} = \mathbf{T}^T$.

To see the form of $\mathbf{T}^{-1}$ in the pseudo-orthogonal case, it is easiest to work with matrix notation. Now, as opposed to the identity matrix, we will replace the metric in our initial relation with a diagonal matrix expressing the signature of the group, $\mathbf{I}_{p,q}$. Then we have

$$\mathbf{T}^T\mathbf{I}_{p,q}\mathbf{T} = \mathbf{I}_{p,q}.$$

From this, it is easy to see that $\mathbf{T}_{-1} = \mathbf{I}_{p,q}\mathbf{T}^T\mathbf{I}_{p,q}$.

13.51. Explain why this is equivalent to preserving the volume form $\varepsilon_{a\ldots c}$, i.e. $\varepsilon_{a\ldots c}T_p^a \ldots T^c{}_r = \varepsilon_{p\ldots r}$? Moreover, why is the preservation of its sign sufficient?

Returning to the expression for the determinant in terms of the volume form, we see that $\det\mathbf{T} = 1$ is equivalent to

$$\frac{1}{n!}\epsilon^{ab\ldots d}T^e{}_aT^f{}_b\cdots T^h{}_d\epsilon_{ef\ldots h} = 1.$$

Contracting the leftmost volume form with a (lower index) copy of itself yields

$$\frac{1}{n!}\epsilon^{ab\ldots d}\epsilon_{ab\ldots d}T^e{}_aT^f{}_b\cdots T^h{}_d\epsilon_{ef\ldots h} = T^e{}_aT^f{}_b\cdots T^h{}_d\epsilon_{ef\ldots h} = \epsilon_{ab\ldots d}$$

which shows such a restriction is equivalent to preserving the volume form. The preservation of the determinant's sign is significant because the determinant (and volume form) represents *oriented* volumes: a lack of change in a transformation's determinant ensures that no overall reflection has been made.

13.52. Why?

13.53. Verify these relations, explaining the notational consistency of $h^{ab'}$.

Given

$$\overline{v}_a = \overline{v}^{a'} h_{a'b}, \quad v_{a'} = h_{a'b} v^b,$$

We may multiply both sides by $h^{bc'}$ to obtain

$$\begin{aligned}
\overline{v}_a h^{bc'} &= \overline{v}^{a'} h_{a'b} h^{bc'} \\
&= \overline{v}^{a'} \delta^{c'}_{a'} \\
&= \overline{v}^{c'} \\
v_{a'} h^{a'c} &= h_{a'b} v^b h^{a'c} \\
&= \delta^c_b v^b \\
&= v^c
\end{aligned}$$

I am unsure what Penrose is referring to when he speaks of notational consistency.

13.54. Show this.

If we can reduce h to a series of 1s along its diagonal, we are effectively reducing it to the identity. Taking any vector $\mathbf{v}$ to be expressed with respect to such a basis, we have

$$\|v\| = \sqrt{\langle v, v \rangle} = \sqrt{\overline{v} \cdot v}.$$

Since any number times its complex conjugate returns a real, positive value, the square root of a sum of such numbers also will.

13.55. Show that these transformations are precisely those which preserve the Hermitian correspondence between vectors $\mathbf{v}$ and covectors $\mathbf{v}^*$, and that they are those which preserve $h_{ab'}$.

Treating our Hermitian form as a multilinear map, the preservation of such a form by a unitary transformation follows from its linearity in the second argument and its antilinearity in the first,

$$\mathbf{h}(\mathbf{Tx}, \mathbf{Ty}) = (\mathbf{T}^*\mathbf{T})\mathbf{h}(\mathbf{x}, \mathbf{y}) = \mathbf{h}(\mathbf{x}, \mathbf{y}).$$

Since the Hermitian form defines the correspondence between vectors and covectors, its in variance under unitary transformations extends to an in variance in the aforementioned correspondence.

13.56. Prove this.

A singular matrix is one which has a determinant of 0, or equivalently, at least one eigenvalue of 0. The eigenvalues of an anti symmetric matrix can be found by examining $\overline{\mathbf{x}}^T\mathbf{S}\mathbf{x}$, where $\mathbf{x}$ is a supposed age vector. Then we have

$$\overline{\mathbf{x}}^T\mathbf{S}\mathbf{x} = \overline{\mathbf{x}}^T\lambda\mathbf{x} = \lambda\overline{\mathbf{x}}^T\mathbf{x} = \lambda\|\mathbf{x}\|^2.$$

Taking the conjugate transpose of the leftmost side gives

$$(\overline{\mathbf{x}}^T\mathbf{S}\mathbf{x})^* = \overline{\mathbf{x}}^T\mathbf{S}^T\mathbf{x} = -\overline{\mathbf{x}}^T\mathbf{S}\mathbf{x} = -\lambda\|\mathbf{x}\|^2$$

where we are taking the matrix $\mathbf{S}$ to be real. Since these two expressions are the complex conjugates of each other, we see that $\overline{\lambda} = -\lambda$. This is possible only if λ is zero or purely imaginary.

The determinant of a real matrix, such as $\mathbf{S}$, is a real quantity. Said another way, the product of its eigenvalues must be real. The only way for this quantity to be nonzero is for our matrix $\mathbf{S}$ to contain an even number of purely imaginary eigenvalues. The product of an odd number of such eigenvalues would be imaginary and contradict the realness of the determinant, showing that such odd-dimensional matrices must have at least one eigenvalue of 0.

13.57. Find explicit descriptions of Sp(1) and Sp(1, 1) using this prescription. Can you see why the groups Sp(n, 0) are compact?

13.58. Show why these two different descriptions for the case $p = q = \frac{1}{2}n$ are equivalent.

13.59. Why are they the same?

They both describe rotation around a unit circle. The connection between the real, SO(2) representation and the complex, U(1) representation is Euler's formula,

$$e^{i\theta} = \cos\theta + i\sin\theta.$$

13.60. Explain where the equation $\mathbf{X}^T\mathbf{S} + \mathbf{S}\mathbf{X} = 0$ comes from and why $\mathbf{S}\mathbf{X} = (\mathbf{S}\mathbf{X})^T$. Why does trace $\mathbf{X}$ vanish? Give the Lie algebra explicitly. Why is it of this dimension?

The Lie algebra is made up of the generators of the representation of our group. That is, the matrices $\mathbf{X}$ such that $\mathbf{I}+\varepsilon\mathbf{X}$ satisfies the group conditions. In the case of the symplectic group, this condition is $\mathbf{T}^T\mathbf{S}\mathbf{T} = \mathbf{S}$.

Replacing $\mathbf{T}$ with $\mathbf{I} + \varepsilon\mathbf{X}$ and ignoring higher orders of epsilon, this relationship can be rewritten as

$$(\mathbf{I} + \varepsilon\mathbf{X})^T\mathbf{S}(\mathbf{I} + \varepsilon\mathbf{X}) = (\mathbf{I} + \varepsilon\mathbf{X}^T)\mathbf{S}(\mathbf{I} + \varepsilon\mathbf{X}) = \mathbf{S} + \varepsilon(\mathbf{X}^T\mathbf{S} + \mathbf{S}\mathbf{X}) = \mathbf{S},$$

which is equivalent to $\mathbf{X}^T\mathbf{S}+\mathbf{S}\mathbf{X} = 0$. Being as $\mathbf{S}$ is antisymmetric, we can replace $\mathbf{S}$ in the leftmost term with $-\mathbf{S}^T$ to get another equivalent condition,

$$-\mathbf{X}^T\mathbf{S}^T + \mathbf{S}\mathbf{X} = -(\mathbf{S}\mathbf{X})^T + \mathbf{S}\mathbf{X} = 0,$$

or $\mathbf{S}\mathbf{X} = (\mathbf{S}\mathbf{X})^T$. $\mathbf{X}$ satisfies these conditions when it is a symmetric matrix premultiplied by $\mathbf{S}^{-1}$.

The trace of $\mathbf{X}$ vanishes because $\mathbf{S}^{-1}$ is an antisymmetric matrix. This can be seen by considering $\mathbf{S}^{-1}\mathbf{S} = \mathbf{I}$. Taking the transpose of both sides leaves the rightmost one unaltered, while changing the left to

$$(\mathbf{S}^{-1}\mathbf{S})^T = \mathbf{S}^T(\mathbf{S}^{-1})^T = -\mathbf{S}(\mathbf{S}^{-1})^T.$$

In order for this to equal the identity, we must have $(\mathbf{S}^{-1})^T = -\mathbf{S}^{-1}$. Now, the trace of a matrix product is unaffected by the transpose operation and the order of the product, so

$$\operatorname{trace}\mathbf{X} = \operatorname{trace}\mathbf{S}^{-1}\mathbf{A} = \operatorname{trace}\mathbf{A}^T(\mathbf{S}^{-1})^T = -\operatorname{trace}\mathbf{A}\mathbf{S}^{-1}$$

which is only possible if the trace of $\mathbf{X}$ is zero.

To find the Lie algebra, we must express the generators in terms of a basis and compute the structure constants, $\gamma_{\alpha\beta}{}^{\chi}$. I am at a loss as to how to neatly do this in a completely general way, so we will focus on the simplest case: $n = 2$.

Let us put the antisymmetric form (and therefore its inverse) into a block diagonal one,

$$\mathbf{S}^{-1} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

For a basis of our symmetric matrix $\mathbf{A}$, we will choose

$$\mathbf{F}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad \mathbf{F}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \mathbf{F}_3 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Premultiplying each of these by the inverse of our symmetric form gives us our basis,

$$\mathbf{E}_1 = \mathbf{S}^{-1}\mathbf{F}_1 = \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}, \qquad \mathbf{E}_2 = \mathbf{S}^{-1}\mathbf{F}_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \qquad \mathbf{E}_3 = \mathbf{S}^{-1}\mathbf{F}_3 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

By simple matrix multiplication, we find

$$\begin{aligned} [\mathbf{E}_1, \mathbf{E}_2] &= -[\mathbf{E}_2, \mathbf{E}_1] = 2\mathbf{E}_1, \\ [\mathbf{E}_2, \mathbf{E}_3] &= -[\mathbf{E}_3, \mathbf{E}_2] = 2\mathbf{E}_3, \\ [\mathbf{E}_3, \mathbf{E}_1] &= -[\mathbf{E}_1, \mathbf{E}_3] = -\mathbf{E}_2. \end{aligned}$$

This implies the structure constants of

$$\begin{gathered} \gamma_{12}{}^{1} = \gamma_{23}{}^{3} = -\gamma_{21}{}^{1} = -\gamma_{32}{}^{3} = 2 \\ \gamma_{13}{}^{2} = -\gamma_{31}{}^{2} = 1 \end{gathered}$$

Going through this process also allowed us to see the dimensionality of the group: it is the number of basis elements, which is the number of degrees of freedom in a symmetric matrix. This is given by $\frac{1}{2}(n^2 - n) + n$, or $\frac{1}{2}n(n+1)$.

13.61. Describe these Lie algebras and obtain these dimensions.

13.62. Why, and what does this mean geometrically?

The infinitesimal elements of the orthogonal group are the matrices $\mathbf{T} = \mathbf{I} + \varepsilon\mathbf{X}$ that satisfy

$$\mathbf{T}^T\mathbf{g}\mathbf{T} = \mathbf{g}.$$

Taking the determinant of this expression reveals $(\det \mathbf{T})^2 = 1$. Since the determinant of $\mathbf{I} + \varepsilon\mathbf{X}$ cannot be -1 on account of it being infinitesimally close to $\mathbf{I}$, the previous relation implies $\det \mathbf{T} = 1$.

Geometrically, this means the elements of the orthogonal group do not deform volumes.

14 Calculus on manifolds

14.1. Let $[a, b; c, d]$ stand for the statement '$abcd$ form a parallelogram' (where a, b, d, and c are taken cyclicly, as in §5.1). Take as axioms (i) for any a, b, and c, there exists d such that $[a, b; c, d]$; (ii) if

$[a, b; c, d]$, then $[b, a; d, c]$ and $[a, c; b, d]$; (iii) if $[a, b; c, d]$ and $[a, b; e, f]$, then $[c, d; e, f]$. Show that, when any chosen point is singled out and labeled as the origin, this algebraic structure reduces to that of a 'vector space', but without the 'scalar multiplication' operation, as given in §11.1—that is to say, we get the rules of an additive Abelian group; see Exercise [13.2].

Given two points a and b—and a point o, thought of as the origin—the first axiom of the given operation allows us to find c such that $oabc$ forms a parallelogram: this is analagous to vector addition, where c is the element corresponding to $a + b$.

Commutativity of such an operation follows from the uniqueness guaranteed by the first axiom: the c found for o, a, and b is the same found for o, b, and a.

Associativity is slightly more involved. Consider the expression

$$(a + b) + c = a + (b + c).$$

The parenthesized term on the lefthand side is d, where d is the unique point closing o, a, and b, $[o, a; b, d]$. Similarly, the parenthesized term on the right side is e, where $[o, b; c, e]$. The entirety of the lefthand side then evaluates to f_1, where $[o, d; cf_1]$. Likewise, the righthand side evaluates to f_2, where $[o, a; e, f_2]$. We must now show that $f_1 = f_2$. By the second axiom, we may combine these final expression with their intermediate ones, such that taken together they imply

$$[a, b; c, f_1] \qquad [a, b; c, f_2].$$

By the uniqueness of the fourth point, $f_1 = f_2$.

The identity is given by the origin point. To show that this leaves group elements unchanged, we must show that e, given by $[o, a; o, d]$, is equivalent to a. To show this, we may combine two such expressions to obtain $[o, d; o, d]$ by the third axiom. Because the fourth point is unique, d *must* be a.

Finally, to show the existence of the inverse, we know that there is a unique b such that $[a, o; o, b]$ which implies $[o, a; b, o]$, i.e. $a + b = o$.

14.2. Can you see how to generalize this to the non-Abelian case?

14.3. See if you can confirm this assertion in the case of a spherical triangle (triangle on S^2 made up of great-circle arcs) where you may assume the Hariot's 1603 formula for the area of a spherical triangle given in §2.6.

Imagine a vector traversing such a triangle. To see how much it has rotated by the time it returns to its original position, we must add up the rotations made by our vector at each corner of the triangle. Taking these angles to be measured counter-clockwise from the great circles making up our path, we see that each angle contributes a rotation of $-(\pi - \theta_i)$, where the negative sign is a consequence of our vector rotating opposite to our path's rotation. Adding these together gives

$$-3\pi + \theta_1 + \theta_2 + \theta_3 = -3\pi + (V + \pi) = -2\pi + V,$$

where we have used Hariot's formula in the last two equalities. Since a rotation of 2π leaves our vector unchanged, the true change in angle is given by V.

14.4. Explain why unique. *Hint*: Consider the action of ∇ on $\alpha \cdot \xi$, etc.

Because, once we understand the effect of our connection on one-forms and vectors, we may extend it to arbitrary tensors through the product law. Consider fully contracting our tensor $\mathbf{T}$ with the necessary number of vectors and one-forms to produce a scalar. Then we have

$$\nabla_i(T^{a\ldots b}_{c\ldots d}\xi^c\cdots\xi^d\alpha_a\cdots\alpha_b) = \nabla_i(T^{a\cdots b}_{c\cdots d})\xi^c\cdots\xi^d\alpha_a\cdots\alpha_b + T^{a\ldots b}_{c\ldots d}\nabla_i(\xi^c\cdots\xi^d\alpha_a\cdots\alpha_b).$$

The term on the left is simply the exterior derivative of a scalar—that is, the gradient of a scalar—while the second term on the contains a covariant derivative of objects whose behaviors are known. To find the action of the connection on $\mathbf{T}$, we simply need to isolate it,

$$\nabla_i(T^{a\cdots b}_{c\cdots d})\xi^c\cdots\xi^d\alpha_a\cdots\alpha_b = \nabla_i(T^{a\ldots b}_{c\ldots d}\xi^c\cdots\xi^d\alpha_a\cdots\alpha_b) - T^{a\ldots b}_{c\ldots d}\nabla_i(\xi^c\cdots\xi^d\alpha_a\cdots\alpha_b).$$

14.5. See if you can show this, finding the expression explicitly. *Hints*: First look at the action of the difference between two connections on a vector field ξ, giving the answer in the index form $\zeta^c\Gamma^a_{bc}$; second show that this difference of connections acting on a covector α has the index form $-\alpha_c\Gamma^c_{ba}$; third, using the definition of a $\begin{bmatrix}p\\q\end{bmatrix}$-valent tensor $\mathbf{T}$ as a multilinear function of q vectors on p covectors (cf. $\S_{12}.8$), find the general index expression for the difference between the connections acting on $\mathbf{T}$.

Following the hints given by Penrose, let us look at the effect of two connections, ∇ and $\hat{\nabla}$. Focusing on one at a time, we see

$$\begin{aligned}[\nabla\xi]^a_b &= [\nabla\xi^c\partial_c]^a_b \\ &= [(\nabla\xi^c)\partial_c]^a_b + [\xi^c\nabla\partial_c]^a_b \\ &= [\mathrm{d}\xi^c\cdot\partial_c]^a_b + \xi^c[\nabla\partial_c]^a_b \\ &= [\partial_d\xi^c\mathrm{d}x^d\cdot\partial_c]^a_b + \xi^c[\nabla\partial_c]^a_b \\ &= \partial_b\xi^a + \xi^c\gamma^a_{bc},\end{aligned}$$

where we have identified $[\nabla\partial_c]^a_b = \gamma^a_{bc}$. Some comments are in order. In the third line, ∇ reduces to the exterior derivative because the components of ξ, given by ξ^c, are simply real numbers (see §12.3), and §14.3 says that all connections must correspond to the exterior derivative when acting on a scalar. In the last line, we have changed $[\partial_d\xi^c\mathrm{d}x^d\cdot\partial_c]^a_b$ to $\partial_b\xi^a$ because a and b are indexing the components of a geometric quantity: the basis elements of this quantity are $\mathrm{d}x^d\cdot\partial_c$, and its components are $\partial_d\xi^c$.

The difference between two connections is

$$[(\nabla-\hat{\nabla})\xi]^a_b = \xi^c(\gamma^a_{bc}-\hat{\gamma}^a_{bc}) = \xi^c\Gamma^a_{bc}.$$

In the case of covectors, we may use $\nabla_b(\alpha_a\xi^a) = (\nabla_b\alpha_a)\xi^a + \alpha_a\nabla_b\xi^a$. Explicitly, we have

$$\begin{aligned}(\nabla_b\alpha_a)\xi^a &= \partial_b(\alpha_a\xi^a) - \alpha_a(\partial_b\xi^a + \xi^c\gamma^a_{bc}) \\ &= (\partial_b\alpha_a)\xi^a + \alpha_a\partial_b\xi^a - \alpha_a\partial_b\xi^a - \alpha_a\xi^c\gamma^a_{bc} \\ &= (\partial_b\alpha_a)\xi^a - \alpha_a\xi^c\gamma^a_{bc}.\end{aligned}$$

Since ξ^a is arbitrary, this implies $[\nabla\alpha]_{ab} = \partial_b\alpha_a - \alpha_c\gamma^c_{ba}$. From this, we see that the difference between two connections when acting on a covector is

$$[(\nabla-\hat{\nabla})\alpha]_{ab} = -\alpha_c(\gamma^c_{ba}-\hat{\gamma}^c_{ba}) = -\alpha_c\Gamma^c_{ba}.$$

Using the extension of the connection to arbitrary tensors identified in the previous exercise, we find

$$[(\nabla - \hat{\nabla})\mathbf{T}]^{a\ldots b}_{ec\ldots d} = \partial_e T^{a\ldots b}_{c\ldots d} + T^{f\ldots b}_{c\ldots d}\Gamma^a_{ef} + \cdots + T^{a\ldots f}_{c\ldots d}\Gamma^c_{ef} - T^{a\ldots b}_{f\ldots d}\Gamma^f_{ec} - \cdots - T^{a\ldots b}_{c\ldots f}\Gamma^f_{ed},$$

where e is the abstract index corresponding to the connection.

14.6. As an application of this, take the two connections to be ∇ and the coordinate connection. Find a coordinate expression for the action of ∇ on any tensor, showing how to obtain the components Γ^a_{bc} explicitly from $\Gamma^a_{b1} = \nabla_b\delta^a_1, \ldots, \Gamma^a_{bn} = \nabla_b\delta^a_n$, i.e. in terms of the action of ∇ on each of the coordinate vectors $\delta^a_1, \ldots, \delta^a_n$. (Here a is a vector index, which may be thought of as an 'abstract index' in accordance with §12.8, so that 'δ^a_1' etc. indeed denote vectors and not simply sets of components, but n just denotes the dimension of the space. Note that the coordinate connection annihilates each of these coordinate vectors.)

From the previous problem, it can be immediately seen that substituting the coordinate connection in place of $\hat{\nabla}$ sets $\gamma^a_{bc} = \Gamma^a_{bc}$, and therefore $\Gamma^a_{bc} = [\nabla\partial_c]^a_b$.

14.7. Explain why the right-hand side must have this general form; find the components τ^a_{bc} in terms of Γ^a_{bc}. See Exercise [14.6].

Because Φ is a scalar, the covariant derivative reduces to $\nabla_a\Phi = \partial_a\Phi$. This, however, is a one-form, and so *its* covariant derivative is given by

$$\nabla_a\nabla_b\Phi = \nabla_a(\partial_b\Phi) = \partial_{ab}\Phi - \partial_c\Phi\Gamma^c_{ab}.$$

Swapping the indices and subtracting, we see

$$\begin{aligned}(\nabla_a\nabla_b - \nabla_b\nabla_a)\Phi &= \partial^2_{ab}\Phi - \partial_c\Phi\Gamma^c_{ab} - \partial^2_{ba}\Phi + \partial_c\Phi\Gamma^c_{ba} \\ &= (\Gamma^c_{ba} - \Gamma^c_{ab})\partial_c\Phi \\ &= (\Gamma^c_{ba} - \Gamma^c_{ab})\nabla_c\Phi \\ &= \tau_{ab}{}^c\nabla_c\Phi.\end{aligned}$$

In a way, torsion is a measure of the noncommutativity of the a and b indices of our Christoffel symbol.

14.8. Show what extra term is needed to make this expression consistent, when torsion is present.

From the rules derived above, we may write

$$\nabla_a(\nabla_b\xi^c) = \frac{\partial(\nabla_b\xi^c)}{\partial x^a} + \nabla_b\xi^d\Gamma^c_{ad} - \nabla_d\xi^c\Gamma^d_{ab}.$$

If a and b are symmetric, i.e. the torsion of the connection is zero, the last term will vanish upon permuting the two indices and subtracting. Therefore, the term *added* by torsion is

$$\tau_{ab}{}^d\nabla_d\xi^c,$$

giving a general expression

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)\xi^d = R_{abc}{}^d\xi^c + \tau_{ab}{}^c\nabla_c\xi^d.$$

14.9. What is the corresponding expression for $\nabla_a \nabla_b - \nabla_b \nabla_a$ acting on a covector? Derive the expression for a general tensor of valence $\begin{bmatrix} p \\ q \end{bmatrix}$.

Since we know the action of $\nabla_a \nabla_b - \nabla_b \nabla_a$ on a scalar and a vector, application of the product rule will yield its action on a covector. This product rule is given by

$$\begin{aligned}
(\nabla_a \nabla_b - \nabla_b \nabla_a)(\mathbf{T} \cdot \mathbf{Q}) &= (\nabla_a \nabla_b)(\mathbf{T} \cdot \mathbf{Q}) - (\nabla_b \nabla_a)(\mathbf{T} \cdot \mathbf{Q}) \\
&= \nabla_a[(\nabla_b \mathbf{T}) \cdot \mathbf{Q} + \mathbf{T} \cdot \nabla_b \mathbf{Q}] - \nabla_b[(\nabla_a \mathbf{T}) \cdot \mathbf{Q} + \mathbf{T} \cdot \nabla_a \mathbf{Q}] \\
&= \nabla_a[(\nabla_b \mathbf{T}) \cdot \mathbf{Q}] + \nabla_a[\mathbf{T} \cdot \nabla_b \mathbf{Q}] - \nabla_b[(\nabla_a \mathbf{T}) \cdot \mathbf{Q}] - \nabla_b[\mathbf{T} \cdot \nabla_a \mathbf{Q}] \\
&= (\nabla_a \nabla_b \mathbf{T}) \cdot \mathbf{Q} + \nabla_b \mathbf{T} \cdot \nabla_a \mathbf{Q} + \nabla_a \mathbf{T} \cdot \nabla_b \mathbf{Q} + \mathbf{T} \cdot \nabla_a \nabla_b \mathbf{Q} \\
&\quad - (\nabla_b \nabla_a \mathbf{T}) \cdot \mathbf{Q} - \nabla_a \mathbf{T} \cdot \nabla_b \mathbf{Q} - \nabla_b \mathbf{T} \cdot \nabla_a \mathbf{Q} - \mathbf{T} \cdot \nabla_b \nabla_a \mathbf{Q} \\
&= [(\nabla_a \nabla_b - \nabla_b \nabla_a)\mathbf{T}] \cdot \mathbf{Q} + \mathbf{T} \cdot [(\nabla_a \nabla_b - \nabla_b \nabla_a)\mathbf{Q}].
\end{aligned}$$

For a covector, then, we have

$$\begin{aligned}
[(\nabla_a \nabla_b - \nabla_b \nabla_a)\alpha_d]\xi^d &= (\nabla_a \nabla_b - \nabla_b \nabla_a)(\alpha_d \xi^d) - \alpha_d(\nabla_a \nabla_b - \nabla_b \nabla_a)\xi^d \\
&= \tau_{ab}{}^c \nabla_c(\alpha_d \xi^d) - \alpha_d(R_{abc}{}^d \xi^c + \tau_{ab}{}^c \nabla_c \xi^d) \\
&= \tau_{ab}{}^c(\nabla_c(\alpha_d \xi^d) - \alpha_d \nabla_c \xi^d) - \alpha_d R_{abc}{}^d \xi^c \\
&= \tau_{ab}{}^c(\nabla_c \alpha_d)\xi^d - \alpha_d R_{abc}{}^d \xi^c.
\end{aligned}$$

As ξ is arbitrary and we are looking at the torsion-free case, the action of this operator on a covector becomes $(\nabla_a \nabla_b - \nabla_b \nabla_a)\alpha_d = -\alpha_c R_{abd}{}^c$. Applying the product rule to an arbitrary tensor gives (in the torsion-free case) an action of

$$(\nabla_a \nabla_b - \nabla_b \nabla_a)T^{c\ldots d}_{e\ldots f} = R_{abg}{}^c T^{g\ldots d}_{e\ldots f} + \cdots + R_{abg}{}^d T^{c\ldots g}_{e\ldots f} - R_{abe}{}^g T^{c\ldots d}_{g\ldots f} - \cdots - R_{abf}{}^g T^{c\ldots d}_{e\ldots g}$$

14.10. First, explain the 'i.e'; then derive this from the equation defining $R_{abc}{}^d$, above, by expanding out $\nabla_{[a} \nabla_b(\xi^d \nabla_{d]} \Phi)$. (Diagrams can help.)

14.11. Derive this from the equation defining $R_{abc}{}^d$, above, by expanding out $\nabla_{[a} \nabla_b \nabla_{d]} \xi^e$ in two ways. (Diagrams can again help.)

14.12. Demonstrate the equivalence of all these conditions.

We have earlier identified $\nabla_{\mathbf{t}}$ with $t^a \nabla_a$, which shows the equivalence of the second and third conditions. As for the first: notice that u is a scalar, and so $\nabla_a u = \partial_a u$, but $t^a \partial_a u = (t^a \partial_a)u = \mathbf{t}(u)$, and so this expresses the same restriction as conditions two and three.

14.13. Show that if u and v are two affine parameters on γ, with respect to two different choices of $\mathbf{t}$, then $v = Au + B$, where A and B are constant along γ.

If both u and v describe affine parameters, they both obey

$$t^a \nabla_a v = 1 \qquad \hat{t}^a \nabla_a u = 1$$

where t^a and $\hat{t}^a$ are related by some scaling factor. If we call this scaling factor A, and recognize that constants are in the null space of the derivative operator, we find

$$\partial_a v = A\partial_a(u + C),$$

or $v = Au + AC = Au + B$.

14.14. Find this term.

14.15. Show it.

For the first property, we have

$$\begin{aligned}
\omega(\Phi + \Psi) &= \xi(\eta(\Phi + \Psi)) - \eta(\xi(\Phi + \Psi)) \\
&= \xi(\eta(\Phi) + \eta(\Psi)) - \eta(\xi(\Phi) + \xi(\Psi)) \\
&= \xi(\eta(\Phi)) + \xi(\eta(\Psi)) - \eta(\xi(\Phi)) - \eta(\xi(\Psi)) \\
&= \omega(\Phi) + \omega(\Psi).
\end{aligned}$$

For the second, we see

$$\begin{aligned}
\omega(\Phi\Psi) &= \xi(\eta(\Phi\Psi)) - \eta(\xi(\Phi\Psi)) \\
&= \xi(\Psi\eta(\Phi) + \Phi\eta(\Psi)) - \eta(\Psi\xi(\Phi) - \Phi\xi(\Psi)) \\
&= \xi(\Psi\eta(\Phi)) + \xi(\Phi\eta(\Psi)) - \eta(\Psi\xi(\Phi)) - \eta(\Phi\xi(\Psi)) \\
&= \eta(\Phi)\xi(\Psi) + \Psi\xi(\eta(\Phi)) + \eta(\Psi)\xi(\Phi) + \Phi\xi(\eta(\Psi)) - \xi(\Phi)\eta(\Psi) - \Psi\eta(\xi(\Phi)) - \xi(\Psi)\eta(\Phi) - \Phi\eta(\xi(\Psi)) \\
&= \Psi \cdot (\xi(\eta(\Phi)) - \eta(\xi(\Phi))) + \Phi \cdot (\xi(\eta(\Psi)) - \eta(\xi(\Psi))) \\
&= \Psi\omega(\Phi) + \Phi\omega(\Psi).
\end{aligned}$$

Finally, the third property can be seen through

$$\begin{aligned}
\omega(k) &= \xi(\eta(k)) - \eta(\xi(k)) \\
&= \xi(0) - \eta(0) \\
&= 0.
\end{aligned}$$

14.16. Do it.

This was shown in a previous exercise.

14.17. Try to explain why there is torsion but no curvature.

14.18. Explain (at a formal level) why $e^{a\mathrm{d}/\mathrm{d}y} f(y) = f(y + a)$ when a is a constant.

As we saw in the previous chapter, the exponentiation of operators can be understood in terms of the Taylor series of the exponential function. That is,

$$e^{a\mathrm{d}/\mathrm{d}y} = 1 + a\frac{\mathrm{d}}{\mathrm{d}y} + \frac{a^2}{2}\frac{\mathrm{d}^2}{\mathrm{d}y^2} + \frac{a^3}{6}\frac{\mathrm{d}^3}{\mathrm{d}y^3} + \cdots.$$

The Taylor series of a function $f(y)$ about a point y_0 is given by

$$f(y) = f(y_0) + f'(y_0)(y-y_0) + \frac{1}{2}f''(y_0)(y-y_0)^2 + \frac{1}{6}f'''(y_0)(y-y_0)^3 + \cdots$$

If we expand the function $f(y+a)$ around y, then, we find

$$f(y+a) = f(y) + af'(y) + \frac{a^2}{2}f''(y) + \frac{a^3}{6}f'''(y) + \cdots$$

Returning to our exponentiated operator and applying it to a function $f(y)$, we see

$$\begin{aligned} e^{a\mathrm{d}/\mathrm{d}y}f(y) &= \Big(1 + a\frac{\mathrm{d}}{\mathrm{d}y} + \frac{a^2}{2}\frac{\mathrm{d}^2}{\mathrm{d}y^2} + \frac{a^3}{6}\frac{\mathrm{d}^3}{\mathrm{d}y^3} + \cdots\Big)f(y) \\ &= f(y) + af'(y) + \frac{a^2}{2}f''(y) + \frac{a^3}{6}f'''(y) + \cdots \\ &= f(y+a) \end{aligned}$$

14.19. Derive this formula for $\underset{\xi}{\pounds}\eta$.

The Lie derivative of one vector field with respect to another is the commutator of them. Applying the Lie derivative to a scalar shows us

$$\begin{aligned} (\underset{\xi}{\pounds}\eta)\Phi &= [\xi,\eta](\Phi) \\ &= \xi(\eta(\Phi)) - \eta(\xi(\Phi)) \\ &= \xi^a\nabla_a(\eta^b\nabla_b\Phi) - \eta^b\nabla_b(\xi^a\nabla_a\Phi) \\ &= \xi^a(\nabla_a\eta^b)(\nabla_b\Phi) + \xi^a\eta^b(\nabla_a\nabla_b\Phi) - \eta^b(\nabla_b\xi^a)(\nabla_a\Phi) - \eta^b\xi^a(\nabla_b\nabla_a\Phi) \\ &= \xi^a\nabla_a\eta^b\nabla_b\Phi - \eta^b\nabla_b\xi^a\nabla_a\Phi + \xi^a\eta^b(\nabla_a\nabla_b - \nabla_b\nabla_a)\Phi \\ &= \nabla_\xi\eta^b\partial_b\Phi - \nabla_\eta\xi^a\partial_a\Phi + \xi^a\eta^b(\nabla_a\nabla_b - \nabla_b\nabla_a)\Phi \\ &= \nabla_\xi\eta\Phi - \nabla_\eta\xi\Phi + \xi^a\eta^b(\nabla_a\nabla_b - \nabla_b\nabla_a)\Phi. \end{aligned}$$

Because Φ is arbitrary, we may remove it from both sides. Additionally, as we are considering a torsion-free connection, we may remove the last term in the last line. Thus,

$$\underset{\xi}{\pounds}\eta = \nabla_\xi\eta - \nabla_\eta\xi.$$

14.20. How does torsion modify the formula of Exercise [14.18]?

As we saw above, torsion modifies the Lie derivative by adding a term

$$\xi^a\eta^b(\nabla_a\nabla_b - \nabla_b\nabla_a) = \xi^a\eta^b\tau_{ab}{}^c\nabla_c.$$

14.21. Establish uniqueness, verifying above covector formula, and give explcitily the Lie derviative of a general tensor.

The uniquness of the Lie derivative follows from its associated product law. To confirm the covector formula, we may use the product law of a scalar to obtain

$$\begin{aligned}
(\underset{\xi}{\pounds}\alpha)_a\eta^a &= \underset{\xi}{\pounds}(\alpha_a\eta^a) - \alpha_a(\underset{\xi}{\pounds}\eta^a) \\
&= \xi(\alpha_a\eta^a) - \alpha_a(\xi^b\nabla_b\eta^a - \eta^b\nabla_b\xi^a) \\
&= \xi^b\partial_b(\alpha_a\eta^a) - \alpha_a\xi^b\nabla_b\eta^a + \alpha_a\eta^b\nabla_b\xi^a \\
&= \xi^b\nabla_b(\alpha_a\eta^a) - \alpha_a\xi^b\nabla_b\eta^a + \alpha_a\eta^b\nabla_b\xi^a \\
&= \xi^b\eta^a\nabla_b\alpha_a + \xi^b\alpha_a\nabla_b\eta^a - \alpha_a\xi^b\nabla_b\eta^a + \alpha_a\eta^b\nabla_b\xi^a \\
&= \xi^b\eta^a\nabla_b\alpha_a + \alpha_a\eta^b\nabla_b\xi^a.
\end{aligned}$$

As η was arbitrary, we may remove it from both sides of this equation (reindexing as appropriate) to find

$$(\underset{\xi}{\pounds}\alpha)_a = \xi^b\nabla_b\alpha_a + \alpha_b\nabla_a\xi^b.$$

In the mathematician's (non-indicial) notation, this becomes

$$\underset{\xi}{\pounds}\alpha = \nabla_\xi\alpha - \alpha\cdot\nabla\xi.$$

The rule for taking the Lie derivative of a general tensor can be easily generalized from this case and the one given by [14.19]—simply treat the upper and lower abstract indices appropriately. We have, then

$$(\underset{\xi}{\pounds}\mathbf{T})^{a\ldots b}_{c\ldots d} = \xi^e\nabla_e T^{a\ldots b}_{c\ldots d} - T^{e\ldots b}_{c\ldots d}\nabla_e\xi^a - \cdots - T^{a\ldots e}_{c\ldots d}\nabla_e\xi^b + T^{a\ldots b}_{e\ldots d}\nabla_c\xi^e + \cdots + T^{a\ldots b}_{c\ldots e}\nabla_d\xi^e.$$

14.22. Show how to find this second connection, taking the 'Γ' for the difference between the connections to be antisymmetric in its lower two indices. (See Exercise [14.5].)

14.23. Establish this and show how the presence of a torsion tensor τ modifies the expression.

14.24. Show this.

If we take the exterior derivative of such an expression, we obtain

$$\begin{aligned}
[\mathrm{d}(\mathrm{d}\alpha)]_{abc\ldots d} &= \nabla_{[a}\nabla_{[b}\alpha_{c\ldots d]]} \\
&= \nabla_{[a}\nabla_b\alpha_{c\ldots d]}.
\end{aligned}$$

This is completely antisymmetric, yet retains an intrinsic symmetry in its first two indices. The only entity which satisfies both symmetry and antisymmetry is 0, so $\mathrm{d}^2\alpha = 0$.

14.25. Demonstrate equivalence (if torsion vanishes) to the previous phyisicist's expression.

14.26. Derive the explicit component expression $\Gamma^a_{bc} = \frac{1}{2}g^{ad}(\partial g_{bd}/\partial x^c + \partial g_{cd}/\partial x^b - \partial g_{cb}/\partial x^d)$ for the connection quantities Γ^a_{bc} (Christoffel symbols). (See Exercise [14.6]).

Metric compatibility, in coordinates, can be expressed as

$$\nabla_c g_{ab} = \partial_c g_{ab} - g_{db}\Gamma^d_{ac} - g_{ad}\Gamma^d_{bc} = 0,$$

or $\partial_c g_{ab} = g_{db}\Gamma^d_{ac} + g_{ad}\Gamma^d_{bc}$. If we permute these indices and add the resulting expressions, we find (noting that we are considering the torsion-free case)

$$\partial_c g_{ab} + \partial_a g_{bc} + \partial_b g_{ca} = g_{db}\Gamma^d_{ac} + g_{ad}\Gamma^d_{bc} + g_{dc}\Gamma^d_{ba} + g_{bd}\Gamma^d_{ca} + g_{da}\Gamma^d_{cb} + g_{cd}\Gamma^d_{ab}$$

Notice that if we instead subtract one of these terms, we are left with one Christoffel symbol on the right-hand side,

$$\partial_c g_{ab} + \partial_a g_{bc} - \partial_b g_{ca} = 2g_{bd}\Gamma^d_{ac}.$$

If we multiply both sides by the inverse metric (reindexing as appropriate) and divide by two, we find

$$\Gamma^a_{bc} = \frac{1}{2}g^{ad}(\partial_c g_{bd} + \partial_b g_{dc} - \partial_d g_{bc}).$$

14.27. Derive the classical expression $R_{abc}{}^d = \partial\Gamma^d_{cb}/\partial x^a - \partial\Gamma^d_{ca}/\partial x^b + \Gamma^u_{cb}\Gamma^d_{ua} - \Gamma^u_{ca}\Gamma^d_{ub}$ for the curvature tensor in terms of Christoffel symbols. *Hint*: Use the definition in §14.4 of the curvature tensor, where ξ^d is each of the coordinate vectors $\delta^a_1, \ldots, \delta^a_n$, in turn. (As in Exercise [14.6], the quantities δ^a_1, δ^a_2, etc. are to be though of as actual individual vectors, where the upper index a may be viewed as an abstract index, in accordance with §12.8).

In Exercise 14.8, we expanded the expression for curvature to find

$$\nabla_a(\nabla_b\xi^c) = \frac{\partial(\nabla_b\xi^c)}{\partial x^a} + \nabla_b\xi^d\Gamma^c_{ad},$$

where the last term has been removed due to our current focus on the torsion-free case. If we expand out $\nabla_b\xi^c$, we find

$$\begin{aligned}\nabla_a(\nabla_b\xi^c) &= \partial_a\partial_b\xi^c + \partial_a(\xi^f\Gamma^c_{fb}) + (\partial_b\xi^d + \xi^f\Gamma^d_{fb})\Gamma^c_{ad} \\ &= \partial_a\partial_b\xi^c + (\partial_a\xi^f)\Gamma^c_{fb} + \xi^f\partial_a\Gamma^c_{fb} + (\partial_b\xi^d)\Gamma^c_{ad} + \xi^f\Gamma^d_{fb}\Gamma^c_{ad}\end{aligned}$$

Swapping a and b and subtracting yields

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)\xi^d = \xi^c\partial_a\Gamma^d_{cb} - \xi^c\partial_b\Gamma^d_{ca} + \xi^c\Gamma^f_{cb}\Gamma^d_{af} - \xi^c\Gamma^f_{ca}\Gamma^d_{bf} = R_{abc}{}^d\xi^c$$

Recognizing that ξ is arbitrary, we arrive at

$$R_{abc}{}^d = \partial_a\Gamma^d_{cb} - \partial_b\Gamma^d_{ca} + \Gamma^f_{cb}\Gamma^d_{fa} - \Gamma^f_{ca}\Gamma^d_{fb}.$$

14.28. Supply details for this entire argument.

14.29. Establish these relations, first deriving the antisymmetry in cd from $\nabla_{[a}\nabla_{b]}g_{cd} = 0$ and then using the two antisymmetries and Bianchi symmetry to obtain the interchange symmetry.

By compatibility with the metric $\nabla_{[a}\nabla_{b]}g_{cd} = 0$. We discovered the action of $\nabla_{[a}\nabla_{b]}$ on a general tensor in Exercise 14.9, and so we know

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)g_{cd} = -g_{ed}R_{abc}{}^e - g_{ce}R_{abd}{}^e$$

If we equate this to zero and use our metric to lower indices, we find

$$R_{abcd} = -R_{abdc}.$$

14.30. Verify that the symmetries allow only 20 independent components when $n = 4$.

14.31. Derive this equation.

Using the indicial form of the Lie derivative of a general tensor, we see that, if κ is a Killing vector field,

$$(\underset{\kappa}{\pounds}\mathbf{g})_{ab} = \kappa^c\nabla_c g_{ab} + g_{cb}\nabla_a\kappa^c + g_{ac}\nabla_b\kappa^c = 0.$$

By metric compatibility, the first term vanishes, while we can use the metric to lower the indices on the last two terms (which is valid by metric compatibility), giving us

$$\nabla_a\kappa_b + \nabla_b\kappa_a = \nabla_{(a}\kappa_{b)} = 0.$$

14.32. Verify this 'geometrically obvious' fact by direct calculation—and why is it 'obvious'?

If κ and ξ are Killing vector fields, then, by direct calculation,

$$\begin{aligned}
(\underset{[\kappa,\xi]}{\pounds}\mathbf{g})_{ab} &= [\kappa,\xi]^c\nabla_c g_{ab} + g_{cb}\nabla_a[\kappa,\xi]^c + g_{ac}\nabla_b[\kappa,\xi]^c \\
&= g_{cb}\nabla_a(\kappa\xi)^c - g_{cb}\nabla_a(\xi\kappa)^c + g_{ac}\nabla_b(\kappa\xi)^c - g_{ac}\nabla_b(\xi\kappa)^c \\
&= g_{cb}\nabla(\kappa\xi)^c + g_{ac}\nabla_b(\kappa\xi)^c - (g_{cb}\nabla_a(\xi\kappa)^c + g_{ac}\nabla_b(\xi\kappa)^c) \\
&= 0.
\end{aligned}$$

That is, $[\kappa,\xi]$ is also a Killing vector field. This is geometrically obvious because, if a metric is unchanged when dragged along two independent vector fields, then dragging it along one before dragging it along the other should leave it unchanged.

14.33. Explain why this can be written $\nabla_a S_{bc} + \nabla_b S_{ca} + \nabla_c S_{ab} = 0$, using any torsion-free connection ∇.

Expressing the exterior derviative in terms of the covariant derivative, this condition becomes

$$\begin{aligned}
\mathrm{d}\mathbf{S} &= \nabla_{[a}S_{bc]} \\
&= \frac{1}{6}(\nabla_a S_{bc} - \nabla_a S_{cb} + \nabla_c S_{ab} - \nabla_b S_{ac} + \nabla_b S_{ca} - \nabla_c S_{ba}) \\
&= \frac{1}{3}(\nabla_a S_{bc} + \nabla_b S_{ca} + \nabla_c S_{ab}),
\end{aligned}$$

i.e. $\nabla_a S_{bc} + \nabla_b S_{ca} + \nabla_c S_{ab} = 0$.

14.34. Demonstrate these relations, first establishing that $S^{a[b}\nabla_a S^{cd]} = 0$.

14.35. Explain why.

14.36. Explain why, in each case. *Hint*: Construct a coordinate system with $\xi = \partial/\partial x^1$; then take repeated Lie derivaties to construct a frame, etc.

15 Fibre bundles and gauge connections

15.1. Explain why the dimension of $\mathcal{M} \times \mathcal{V}$ is the sum of the dimensions of $\mathcal{M}$ and $\mathcal{V}$.

An element in the product space $\mathcal{M} \times \mathcal{V}$ is specified by the ordered pair (m, v). That is, we associate to every point in $\mathcal{M}$ a space $\mathcal{V}$. The dimension of our product space is, then, the dimension of $\mathcal{M}$ augmented by the dimension of $\mathcal{V}$, or $\dim \mathcal{M} + \dim \mathcal{V}$.

15.2. Explain this.

A structure preserving transformation is one such that

$$\mathbf{T}(x)\mathbf{T}(y) = \mathbf{T}(xy).$$

For the transformation $\mathbf{T}(x) = -x$, we have

$$\mathbf{T}(x)\mathbf{T}(y) = (-x)(-y) = xy \neq \mathbf{T}(xy) = -xy,$$

and so such a ‘flip’ of $\mathbb{R}$ is not a symmetry.

15.3. Spell this argument out, using the construction of B from two patches, as indicated above.

Suppose, as Penrose does, that we build our Möbius from two semi-circular bands—one having ‘vertical’ coordinates flipped in relation to the other. To smoothly glue such parts together, we must avoid discontinuities, and so our coordinate systems must match on the overlap in these two sections, but the only such coordinate that satisfies $x = -x$ is $x = 0$. Therefore, a smooth cross-section through the Möbius bundle must cross 0 if it is to pass through all patches.

15.4. Carry out this argument. Can you see how to do the S^{15} case?

15.5. Show this. *Hint*: Take the tangent vector to be $u\partial/\partial v - v\partial/\partial u + x\partial/\partial y - y\partial/\partial x$.

15.6. Why does every such spinor field take the value zero at at least one point of S^2?

Consider traversing such a field around a great circle of S^2. Being as this is a rotation of 2π, the spinor vectorspin we arrive at should be the negative of what we started with. But these are the same, and so must equal zero.

15.7. Explain this in detail.

15.8. Show that $\mathcal{B}'^{\mathbb{C}}$, interpreted as a real bundle over S^2 is indeed the same as $T(S^2)$. *Hint*: Re-examine Exercise [15.5].

15.9. Explain how to do this. *Hint*: Think of Cartesian coordinates (x, y, z). Take two at a time, with the canvas given by the third set to unity.

> To cover the whole of $\mathbb{P}^2$, our three planes must capture every single point. Points in a 2-dimensional projective space are represented as slopes of lines through the origin of a 3-dimensional vector space. We need to find 2 points on each line to calculate its slope, and so consider choosing as our three planes those Penrose suggests.
>
> Because the 3-dimensional vector space containing our projective space is Euclidean, any line through its origin will intersect these planes two times, except when such a line is exactly perpendicular to one plane. But, being as this occurs for only one such line per plane, it is uniquely specified by such a projection. Thus, all points in our projective space are captured on these three planes.

15.10. Explain why there are n independent ratios. Find $n+1$ sets of n ordinary coordinates (constructed from the zs), for $n+1$ different coordinate spatches, which together cover P^n.

> There are n independent ratios because we can normalize every value with respect to a particlar one. This is also the key to constructing the necessary coordinate sets. One possible answer is
>
> $$\begin{gathered} \{1, z^1, \ldots, z^n\} \\ \{z^0, 1, \ldots, z^n\} \\ \vdots \\ \{z^0, z^1, \ldots, 1\} \end{gathered}$$

15.11. Explain this geometry, showing that the bundle $\mathbb{R}^{n+1} - O$ over $\mathbb{RP}^n$ can be understood as the composition of the bundle $\mathbb{R}^{n+1} - O$ over S^n (the fibre, $\mathbb{R}^+$, being the positive reals) and of S^n as a twofold cover of $\mathbb{RP}^n$.

15.12. Check this.

> We are simply asked to check this, so
>
> $$\begin{aligned} \frac{\partial \Phi}{\partial z} &= \Big(\frac{\partial B}{\partial z} + \frac{\partial \overline{B}}{\partial z}\Big) e^{(B+\overline{B})} \\ &= A\Phi + \overline{\frac{\partial B}{\partial \overline{z}}}\Phi \\ &= A\Phi + \overline{\int \frac{\partial A}{\partial \overline{z}} \mathrm{d}z} \\ &= A\Phi, \end{aligned}$$
>
> where the second term becomes zero by virtue of A's holomorphicity.

15.13. Verify this formula.

Expanding each nabla yields

$$\begin{aligned}
(\nabla\overline{\nabla} - \overline{\nabla}\nabla)\Phi &= \Big(\frac{\partial}{\partial z} - A\Big)\overline{\Big(\frac{\partial}{\partial z} - A\Big)}\Phi - \overline{\Big(\frac{\partial}{\partial z} - A\Big)}\Big(\frac{\partial}{\partial z} - A\Big)\Phi \\
&= \Big(\frac{\partial}{\partial z} - A\Big)\Big(\frac{\partial}{\partial \overline{z}} - \overline{A}\Big)\Phi - \Big(\frac{\partial}{\partial \overline{z}} - \overline{A}\Big)\Big(\frac{\partial}{\partial z} - A)\Phi \\
&= \frac{\partial^2\Phi}{\partial z\partial\overline{z}} - \frac{\partial\overline{A}}{\partial z}\Phi - A\frac{\partial\Phi}{\partial\overline{z}} + A\overline{A}\Phi - \frac{\partial^2\Phi}{\partial\overline{z}\partial z} + \frac{\partial A}{\partial\overline{z}}\Phi + \overline{A}\frac{\partial\Phi}{\partial z} - \overline{A}A\Phi \\
&= \frac{\partial A}{\partial\overline{z}}\Phi - \frac{\partial\overline{A}}{\partial z}\Phi + \overline{A}\frac{\partial\Phi}{\partial z} - \overline{\overline{A}\frac{\partial\Phi}{\partial z}} \\
&= (\frac{\partial A}{\partial\overline{z}} - \frac{\partial\overline{A}}{\partial z})\Phi,
\end{aligned}$$

where the vanishing of the last two terms in the second to last line comes from the fact that Φ is real, and anything minus its complex conjugate is purely imaginary.

15.14. Confirm the assertions in this paragraph, finding the explicit value of k that gives this required factor 2.

16 The ladder of infinity

16.1. Show how these rules work, explaining why p has to be prime.

Addition and subtraction are quite straight-forward, being defined as

$$a_1 + a_2 \equiv b \mod p \iff a_1 + a_2 - b = kp, \quad k \in \mathbb{Z}$$

and

$$a_1 - a_2 \equiv b \mod p \iff a_1 - a_2 - b = kp, \quad k \in \mathbb{Z},$$

i.e. where b is the remainder of $a_1 + a_2$ (or $a_1 - a_2$) on division by p. The existence of such a remainder follows from the division theorem. The additive inverse of a_1 is simply $-a_1$, while the additive identity is 0—both rules arise due the definition of modulo arithmetic via the integers.

Multiplication is defined similarly, but here we see the need for p to be prime. We have

$$a_1 a_2 \equiv b \mod p \iff a_1 a_2 - b = kp, \quad k \in \mathbb{Z}.$$

The multiplicative identity is 1, but what about the multiplicative inverse? This is the number a^{-1} such that

$$aa^{-1} \equiv 1 \mod p \iff aa^{-1} - 1 = kp, \quad k \in \mathbb{Z}.$$

That is, aa^{-1} must be coprime to p. In each finite field (with p prime), there is only one element that, when multiplying a, returns such an element.

If p were composite, say $p = mn$, then elements m and n would not have a multiplicative inverse. This is because, when p is prime and $a \equiv b \mod p$, the smallest integer we may multiply a by to leave it unchanged is p—all others less than this are not divisible by p, and thus will modify b. If, on the other hand, $p = mn$, then element m and n may be multiplied by each other to return to themselves, thereby bypassing the other elements (including the multiplicative identity, which cannot exist because am will never be coprime to mn for any a).

16.2. Make complete addition and multiplication tables for $\mathbb{F}_4$ and check that the laws of algebra work (where we assume that $1+\omega+\omega^2=0$).

Assuming the multiplicative and additive identities are unaltered in this case, we have

$+$	0	1	ω	ω^2
0	0	1	ω	ω^2
1	1	0	ω^2	ω
ω	ω	ω^2	0	1
ω^2	ω^2	ω	1	0

$\times$	0	1	ω	ω^2
0	0	0	0	0
1	0	1	ω	ω^2
ω	0	ω	ω^2	1
ω^2	0	ω^2	1	ω

In the case of the addition table, a brief explanation of ω^2+1 is in order: if we assume $\omega^2=1+\omega$, then $\omega^2+\omega=1+\omega+\omega=1+0=1$.

16.3. Show this.

We can return to §15.6 to remind ourselves that a projective space $\mathbb{P}^n$ can be described by n independent ratios. We can express these as coordinates, just as in Exercise 15.10, taking care to ensure there is no overlap. Then, we simply count the number of possible coordinates (each mapping to a specific point).

Consider

$$A_1=\{1,0,\dots,0\}$$
$$A_2=\{z^0,1,\dots,0\}$$
$$\vdots$$
$$A_n=\{z^0,z^1,\dots,1\},$$

where a projective space $\mathbb{P}^n$ can be described in coordinates with the union of the first n such sets. Working with finite fields, each coordinate z^i is limited to a finite number of values q. Thus, for $\mathbb{P}^n(\mathbb{F}_q)$, there are

$$1+q+q^2+\cdots+q^n=\frac{q^{n+1}-1}{q-1}$$

different points.

16.4. Show how to construct *new* magic discs, in teh cases $q=3,5$ by starting at a particular marked point on one of the discs that I have given and then multiplying each of the angular distances from the other marked points by some fixed integer. Why does this work?

16.5. The finite field $\mathbb{F}_8$ has elements $0,1,\varepsilon,\varepsilon^2,\varepsilon^3,\varepsilon^4,\varepsilon^5,\varepsilon^6$, where $\varepsilon^7=1$ and $1+1=0$. Show that either (1) there is an identity of the form $\varepsilon^a+\varepsilon^b+\varepsilon^c=0$ whenever a, b, and c are numbers on the background circle of Fig. 16.1a which can line up with the three spots on the disc, or else (2) the same holds, but with ε^2 in place of ε (i.e. $\varepsilon^{3a}+\varepsilon^{3b}+\varepsilon^{3c}=0$).

16.6. Show that the 'associator' $a(bc)-(ab)c$ is antisymmetrical in a, b, c when these are generating elements, and deduce that this (whence also $a(ab)=a^2b$) holds for *all* elements. *Hint*: Make use of Fig. 16.3 and the full symmetry of the Fano plane.

16.7. See if you can provide such an explicit procedure, by finding some sort of systematic way of ordering all the fractions. You may find the result of Exercise [16.8] helpful.

Perhaps the simplest procedure is to imagine a table where the rows and columns are labeled by elements of $\mathbb{N}$ and each cell's value is given by the ratio of the corresponding column label to its row label. This construction lists all possible rational numbers.

To create a mapping between $\mathbb{N}$ and such a construction, imagine drawing a line from the origin of our table (the cell corresponding to $1:1$) up one cell, then diagonally down (skipping the original cell), then over one cell, then diagonally up, and so on. We can parameterize the line via the natural numbers, providing a mapping from $\mathbb{N}$ to $\mathbb{Q}$.

16.8. Show that the function $\frac{1}{2}((a+b)^2+3a+b)$ explicitly provides a 1-1 correspondence between the natural numbers and the pairs (a,b) of natural numbers.

16.9. Spell this out in detail.

Imagine we have three sets, A, B, and C, with respective cardinal numbers α, β, γ. The first two inequalities in the text imply that we can set up 1-to-1 correspondences between A and B and between B and C. Call the first of these T and the second S, so $T(A)=B$ and $S(B)=C$. The fact that we can write $\alpha\leq\gamma$ follows from our ability to string together successive mappings. That is, if A can be mapped to B in a bijective manner, then we can simply map all of *these* elements into C with S in a similarly bijective manner.

16.10. Prove this. Outline: there is a 1-1 map b taking A to some subset $bA(=B(A))$ of B, and a 1-1 map a taking B to some subset aB of A; consider the map of A to B which uses b to map A-aB to bA-baB and abA-$abaB$ to $babA$-$babaB$, etc. and which uses a^{-1} to map aB-abA to $B-bA$ and $abaB-ababA$ to $baB-babA$, etc., and sort out what to do with the rest of A and B.

16.11. Explain this.

Treating such a number as binary (with a decimal point to the far left) leads to ambiguities between numbers such as

$$1000000\cdots \qquad \text{and} \qquad 0111111\cdots,$$

as, in the limit, $1011111\cdots$ converges to $1100000\cdots$. To see this, we can convert such a number to our decimal system by expanding in base-2,

$$\begin{aligned} 0111111\cdots &= 0\cdot 2^{-1}+1\cdot 2^{-2}+1\cdot 2^{-3}+\cdots \\ &= \sum_{n=2}^{\infty} 2^{-n} \\ &= \frac{1}{2}, \end{aligned}$$

which equals $1000000\cdots = 1\cdot 2^{-1}=\frac{1}{2}$.

16.12. Exhibit one. *Hint*: Look at Fig. 9.8, for example.

> Commonly used is the arctan function, which takes the entirety of $\mathbb{R}$ to $(-\frac{\pi}{2}, \frac{\pi}{2})$.

16.13. Explain why this is essentially the same argument as the one I have given here, in the case $\alpha = \aleph_0$ for showing $\alpha < 2^\alpha$.

16.14. Show that this is what happens.

16.15. Give a rough description of how our algorithm might be performed and explain these particular values.

16.16. Show this.

16.17. Can you see why this is so? *Hint*: For an arbitrary Turing machine action of $\mathbf{T}$ applied to n, we can consider an effective Turing machine $\mathbf{Q}$ which has the property that $\mathbf{Q}(r) = 0$ if $\mathbf{T}$ applied to n has not stopped after r computational steps, and $\mathbf{Q}(r) = 1$ if it has. Take the modulo 2 sum of $\mathbf{Q}(n)$ with $\mathbf{T}_t(n)$ to get $\mathbf{T}_s(n)$.

16.18. See if you can establish this.

16.19. Explain why $(A^B)^C$ may be identified with $A^{B\times C}$, for sets A, B, C.

> Because, in the convention we are using, $(A^B)^C$ stands for the set of all mappings from C into the set of all mappings from B into A. This is associating each mapping from B into A with a copy of C, allowing us to view it as a mapping from the pair of elements in C and B into A. Applying this to the entire set of mappings yields $A^{B\times C}$.

17 Spacetime

17.1. Why?

> Because a bundle connection captures the idea of constancy, and Galilean relativity does not permit the notion of a constant point in space.

17.2. Explain the reason for this.

> Particles cannot jump discontinuously around the universe; their motion must be smooth and continuous. This motion translates to an unbroken line permeating each of the fibres of our space: a bundle cross-section.

17.3. Explain these three ways more thoroughly, showing why they all give the same structure.

> An affine structure naturally encodes the 'flatness' of Newtonian spacetime, though we could explicitly represent this via the ∞^6 family of lines permeating $\mathbb{E}^3$, where the dimension of such a family comes from the fact that, to specify three lines at each point, we need the coordinates of the point itself and three points through which our lines will run. Finally, the connection suggested—that without curvature or torsion—is simply the partial derivative operator. This encodes the notion of flatness, analogous to $\mathbb{R}^n$.

17.4. Try to write down an expression for this curvature, in terms of the connection ∇. What normalization condition on the tangent vectors is needed (if any)?

17.5. Find an explicit transformation of $\mathbf{x}$, as a function of t, that does this, for a given Newtonian gravitational field $\mathbf{F}(t)$ that is spatially constant at any one time, but temporally varying both in magnitude and direction.

In such a field, every object undergoes uniform acceleration $\mathbf{F}(t)/m$, where m is the object's mass. We can integrate this twice with respect to t to find the offset in position each object experiences,

$$\mathbf{x}_F = \frac{1}{m}\int_{-\infty}^{\tau}\int_{-\infty}^{\tau'} \mathbf{F}(\tau')\mathrm{d}\tau'.$$

Since this quantity is the same for every object, we may transform every $\mathbf{x} \to \mathbf{x} + \mathbf{x}_F$. Then we have

$$\mathbf{F}(t) = m\ddot{\mathbf{x}} \to m\ddot{\mathbf{x}} + m\ddot{\mathbf{x}}_F.$$

The new righthand side of this becomes

$$m\mathbf{a} + \mathbf{F}(t),$$

which, when combined with the lefthand side, gives

$$m\mathbf{a} = 0.$$

17.6. Derive these various properties, making clear by use of the $O(\ )$ notation, at what order these statements are intended to hold.

Consider a cross-section of a sphere of identical particles that divides it into two halves. (Such a cross-section would be a circle with the same radius r as our sphere.) Label the origin of this cross-section O. Let us imagine that the sphere of particles is attracted to a point A far below it at a distance R, and, furthermore, let us call an arbitrary point on the boundary of our cross-section P.

The force felt at the origin of the cross-section, in component terms (with the x-direction describing horizontal movement and the y-direction vertical movement), is given by

$$\mathbf{F}_O = \left(0, -\frac{GMm}{R^2}\right),$$

while the force felt at point P is given by

$$\mathbf{F}_P = \left(-\frac{GMm}{D^2}\sin\theta, -\frac{GMm}{D^2}\cos\theta\right),$$

where θ is the small angle $\angle OAP$ and D is the length of the line segment $\overline{AP}$. Using these forces we can find the evolution of the positions of our cross-section, which, by symmetry, is the same over the entire boundary of the sphere.

Since θ is hard to work with, let us express $\sin\theta$ and $\cos\theta$ in terms of D and ϕ (where ϕ is the angle $\angle AOP$ at the center of our cross-section). We have

$$\cos\theta = \frac{R - r\cos\phi}{D}$$
$$\sin\theta = \frac{r\sin\phi}{D}$$

Finally, let us expand out each constituent term in a Taylor series in $\varepsilon = \frac{r}{R}$, which satisfies $\varepsilon \ll 1$ by assumption. We have

$$\begin{aligned}
\frac{1}{D^2} &= \frac{1}{(r\sin\phi)^2 + (R - r\cos\phi)^2} \\
&= \frac{1}{R^2 - 2Rr\cos\phi + r^2} \\
&= \frac{1}{R^2(1 - 2\varepsilon\cos\phi + \varepsilon^2)} \\
&\approx \frac{1}{R^2} + \Big(\frac{2\cos\phi}{R^2}\Big)\varepsilon + \mathcal{O}(\varepsilon^2)
\end{aligned}$$

as well as

$$\begin{aligned}
\cos\theta &= \frac{R - r\cos\phi}{\sqrt{R^2 - 2Rr\cos\phi + r^2}} \\
&= \frac{1 - \varepsilon\cos\phi}{\sqrt{1 - 2\varepsilon\cos\phi + \varepsilon^2}} \\
&\approx 1 + \mathcal{O}(\varepsilon^2)
\end{aligned}$$

and

$$\begin{aligned}
\sin\theta &= \frac{r\sin\phi}{\sqrt{R^2 - 2Rr\cos\phi + r^2}} \\
&= \frac{\varepsilon\sin\phi}{\sqrt{1 - 2\varepsilon\cos\phi + \varepsilon^2}} \\
&\approx (\sin\phi)\varepsilon + \mathcal{O}(\varepsilon^2)
\end{aligned}$$

The force on the particles around the boundary of our sphere, by first approximation, is then

$$\mathbf{F}_P = \Big(-\frac{GMm\sin\phi}{R^2}\varepsilon + \mathcal{O}(\varepsilon^2),\ -\frac{GMm}{R^2} - \frac{2GMm\cos\phi}{R^2}\varepsilon + \mathcal{O}(\varepsilon^2)\Big).$$

We can then expand each particle's position in a Taylor series in time, giving

$$\mathbf{x}(t) = \mathbf{x}(0) + \dot{\mathbf{x}}(0)t + \frac{\ddot{\mathbf{x}}(0)}{2}t^2 + \mathcal{O}(t^3),$$

where t here is assumed to be small. The position of a particle on the boundary of our cross-section at time t (when such a particle starts at rest) is then

$$\mathbf{x}_P(t) = \Big(r\sin\phi - \frac{GM\sin\phi}{2R^2}\varepsilon t^2 + \mathcal{O}(\varepsilon^2, t^3),\ r\cos\phi - \big(\frac{GM}{2R^2} + \frac{GM\cos\phi}{R^2}\varepsilon\big)t^2 + \mathcal{O}(\varepsilon^2, t^3)\Big),$$

while the analogous position for center of the sphere is

$$\mathbf{x}_O(t) = \Big(0, -\frac{GM}{2R^2}t^2 + \mathcal{O}(t^3)\Big).$$

The difference $\mathbf{x}_P - \mathbf{x}_O$ tells us how the shape of the boundary of our cross-section varies with time, and is given by

$$\begin{aligned}
\mathbf{x}_P(t) - \mathbf{x}_O(t) &= \Big(r\sin\phi - \frac{GM\sin\phi}{2R^2}\varepsilon t^2 + \mathcal{O}(\varepsilon^2, t^3),\ r\cos\phi - \frac{GM\cos\phi}{R^2}\varepsilon t^2 + \mathcal{O}(\varepsilon^2, t^3)\Big) \\
&= \Big(\sin\phi\big(r - \frac{GMr}{2R^3}t^2\big) + \mathcal{O}(\varepsilon^2, t^3),\ \cos\phi\big(r - \frac{GMr}{R^3}t^2\big) + \mathcal{O}(\varepsilon^2, t^3)\Big)
\end{aligned}$$

where we have made the replacement $\varepsilon \to r/R$. This is the parametric equation of an ellipse. We can make this even clearer by defining

$$a = r - \frac{GMr}{2R^3}t^2 \qquad b = r - \frac{GMr}{R^3}t^2,$$

as then we have

$$\left(\frac{x(t)}{a}\right)^2 + \left(\frac{y(t)}{b}\right)^2 = 1,$$

which is the familiar implicit equation of an ellipse. The volume of the full three-dimensional ellipsoid is given by $\frac{4}{3}\pi a^2 b$, where a is the horizontal semi-axis ($\|\mathbf{x}(t)\|$ at $\phi = \pi/2$) and b is the vertical semi-axis ($\|\mathbf{x}(t)\|$ at $\phi = 0$), which, to first order, is

$$\begin{aligned}\frac{4}{3}\pi a^2 b &= \frac{4}{3}\pi\Big(r - \frac{GM}{2R^2}\varepsilon t^2\Big)^2\Big(r - \frac{GM}{R^2}\varepsilon t^2\Big) + \mathcal{O}(\varepsilon^2, t^3)\\ &= \frac{4}{3}\pi r^3 + \mathcal{O}(\varepsilon^2, t^3),\end{aligned}$$

that is, the volume of our ellipsoid is the same as the volume of our sphere.

17.7. Show that this tidal distortion is proportional to mr^{-3} where m is the mass of the gravitating body (regarded as a point) and r is its distance. The Sun and Moon display discs, at the Earth, of closely equal angular size, yet the Moon's tidal distortion on the Earth's oceans is about five times that due to the Sun. What does this tell us about their relative densities?

The proportionality of the tidal distortion to mr^{-3} was shown above, with the force explicitly proportional to

$$\mathbf{F} \propto \frac{GMm}{R^3}r.$$

In the case of the sun and moon, this is really

$$\mathbf{F}_{\earth\sun} \propto \frac{GM_{\sun}m_{\earth}}{R^3_{\earth\sun}}r_{\earth} \qquad \mathbf{F}_{\earth\leftmoon} \propto \frac{GM_{\leftmoon}m_{\earth}}{R^3_{\earth\leftmoon}}r_{\earth}.$$

We are given that the ratio of these forces is approximately $5:1$, and so, writing each mass in terms of density, we find

$$\frac{\rho_{\leftmoon}V_{\leftmoon}R^3_{\earth\sun}}{\rho_{\sun}V_{\sun}R^3_{\earth\leftmoon}} \approx 5.$$

Since the volume of both bodies is proportional to their radii cubed, we can rewrite this further as

$$\frac{\rho_{\leftmoon}r^3_{\leftmoon}R^3_{\earth\sun}}{\rho_{\sun}r^3_{\sun}R^3_{\earth\leftmoon}} \approx 5.$$

Now, we can imagine drawing a line from our viewpoint, through the center of the moon, and extending through the center of the sun. The right triangles formed by connecting (through the center of each body) this line to the one extending from our viewpoint and touching both bodies tangentially are similar, and we may equate ratios to find

$$\frac{r_{\leftmoon}}{R_{\leftmoon}} = \frac{r_{\sun}}{R_{\sun}}.$$

Putting this information in place of the previous ratio gives

$$\frac{\rho_{☾}}{\rho_{☼}} \approx 5.$$

17.8. Establish this result, assuming that all the mass is concentrated at the center of the sphere.

The rate of change of volume of a sphere is given by

$$\frac{\mathrm{d}V}{\mathrm{d}t} = 4\pi r^2 \frac{\mathrm{d}r}{\mathrm{d}t}.$$

In the case where the surrounding shell of particles starts at rest, the velocity $(\mathrm{d}r/\mathrm{d}t)$ of any particle particle an instant later in time is

$$\left.\frac{\mathrm{d}r(t)}{\mathrm{d}t}\right|_{t=\varepsilon} \approx \frac{GM}{r^2}\varepsilon + \mathcal{O}(\varepsilon^2).$$

Putting these together gives, to first order,

$$\left.\frac{\mathrm{d}V}{\mathrm{d}t}\right|_{t=\varepsilon} = 4\pi GM\varepsilon.$$

Since the rate that Penrose mentions is the initial acceleration of volume reduction, we may divide both sides by time (in this case ε) to arrive at

$$4\pi GM.$$

17.9. Show that this result is still true quite generally, no matter how large or what shape the surrounding shell of stationary particles is, and whatever the distribution of mass.

17.10. Explain why.

In the case of a 4-manifold with a Lorentzian signature, such an expression can be expanded to read

$$-v_0v^0 + v_1v^1 + v_2v^2 + v_3v^3 = 0.$$

Labeling v_iv^i as v_i^2, this becomes

$$v_1^2 + v_2^2 + v_3^2 = v_0^2,$$

which is the implicit equation of S^2—our expanding light 'cone.' Anticipating future sections, if we make the identification of $v_0 = ct$ (with v_1, v_2, and v_3 taking units of distance; we are dealing with a position vector), then such an equation describes the sphere of photons a time t after their emission from a point source.

18 Minowskian geometry

18.1. Find $\mathbf{C}$ explicitly for each of the three cases $\mathbb{E}^4$, $\mathbb{M}$, and $\tilde{\mathbb{M}}$. *Hint*: Think of how $\mathbf{C}$ is to act on ω, ξ, η, and ζ. It is not quite the standard operation of complex conjugation in the cases of $\mathbb{M}$ and $\tilde{\mathbb{M}}$.

Consider the case of normal complex conjugation: real numbers are characterized by $z = \overline{z}$, while purely imaginary ones satisfy $z = -\overline{z}$. To encode this 'realness' and 'imaginaryness,' we may define component-wise operations that make use of the above. Explicitly, we have

$$\begin{aligned}
\mathbf{C}_{\mathbb{E}^4} &: (\omega, \xi, \eta, \zeta) \to (\overline{\omega}, \overline{\xi}, \overline{\eta}, \overline{\zeta}) \\
\mathbf{C}_{\mathbb{M}} &: (\omega, \xi, \eta, \zeta) \to (\overline{\omega}, -\overline{\xi}, -\overline{\eta}, -\overline{\zeta}) \\
\mathbf{C}_{\tilde{\mathbb{M}}} &: (\omega, \xi, \eta, \zeta) \to (-\overline{\omega}, \overline{\xi}, \overline{\eta}, \overline{\zeta})
\end{aligned}$$

All of these operators satisfy $\mathbf{C}^2 = 1$ by virtue of their building blocks (ordinary complex conjugation). Furthermore, the requirement that a point $\mathbf{x}$ be real ($\mathbf{x} = \mathbf{C}\mathbf{x}$) allows these operators to select the necessary 'real' subspaces from $\mathbb{CE}^4$.

18.2. Can you see why?

18.3. Confirm it in this case examining the 4×4 Lie algebra matrices explicitly.

The Lie algebra matrices are those infinitesimal elements satisfying

$$\mathbf{A}^T \mathbf{g} \mathbf{A} = \mathbf{g},$$

i.e. those preserving the metric (and thus inner products). By writing $\mathbf{A} = \mathbf{I} + \varepsilon \mathbf{X}$, we find

$$(\mathbf{I}^T + \varepsilon \mathbf{X}^T)\mathbf{g}(\mathbf{I} + \varepsilon \mathbf{A}) = \mathbf{g} + \varepsilon(\mathbf{X}^T\mathbf{g} + \mathbf{g}\mathbf{X}) + \mathcal{O}(\varepsilon^2) = \mathbf{g},$$

or

$$\mathbf{X}^T\mathbf{g} = -\mathbf{g}\mathbf{X} = -(\mathbf{X}^T\mathbf{g})^T,$$

that is, $\mathbf{X}^T\mathbf{g}$ is antisymmetric. Antisymmetric matrices having the dimensions 4×4 are 6-dimensional, completing the exercise.

If we wish to continue on to find the explicit representation, we can write $\mathbf{X}^T\mathbf{g}$ in terms of the standard antisymmetric basis $\mathbf{E}_i$ (where there is exactly one 1 to the right of the diagonal and one -1 to the left), we have

$$\mathbf{X}^T\mathbf{g} = \sum_{i=1}^{6} a_i \mathbf{E}_i.$$

Postmultiplying by $\mathbf{g}$ (which is its own inverse) and taking the transpose gives us a basis for $\mathbf{X}$, whose elements are given by $\mathbf{E}_i^T\mathbf{g}$. Explicitly, we have

$$\mathbf{E}_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \mathbf{E}_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \mathbf{E}_2 = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{E}_4 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad \mathbf{E}_5 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \qquad \mathbf{E}_6 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}$$

18.4. Explain this action of the Poincare group a little more fully.

The generators of the Poincare group allow us to move any point within a (future directed) null-cone to any other point within the same cone: this is what is meant by the transitive action of the group on the bundle of future-timelike directions on $\mathbb{M}$.

Intuitively, this captures the loss of absolute simultaneity between separate observers.

18.5. (i) Under what circumstances is it possible for a 3-plane element η to contain its normal $\eta^{\perp}$, in $\mathbb{M}$. (ii) Show that there are two distinct families of 2-planes that are the orthogonal complements of themselves in $\mathbb{CE}^4$, but neither of these families survives in $\mathbb{M}$. (These so-called 'self-dual' and 'anti-self-dual' complex 2-planes will have considerable importance later; see §32.2 and §33.11.)

A hyperplane P can be described as follows: a vector ξ is contained within the hyperplane if it is orthogonal to a one-dimensional family of vectors, i.e.

$$g_{\alpha\beta}\eta^{\alpha}\xi^{\beta} = 0 \implies \xi \in P.$$

If a hyperplane is to contain its normal, then we must have

$$g_{\alpha\beta}\eta^{\alpha}\eta^{\beta} = 0.$$

This is satisfied when η is a null-vector, and thus our hyperplane is a light cone. An arbitrary 2-plane is specified by the set of vectors lying in the span of

$$a\eta + b\xi + \chi,$$

where a and b are arbitrary complex constants. For such a plane to be orthogonal to itself, we must have

$$(a\eta + b\xi + \chi) \cdot (c\eta + d\xi + \chi) = 0.$$

As a, b, c, and d are all arbitrary, we may choose convenient values to obtain the necessary conditions. From a combination of 0s and 1s, we find

$$\eta \cdot \eta = \xi \cdot \xi = \eta \cdot \xi + \eta \cdot \chi + \xi \cdot \chi = -\chi \cdot \chi = 0.$$

If we choose $\xi = \mathbf{0}$ so that our plane runs through the origin of $\mathbb{CE}^4$, then our conditions simplify to

$$\eta \cdot \eta = \xi \cdot \xi = \eta \cdot \xi = 0.$$

These can be satisfied by choosing

$$\eta = \begin{pmatrix} 1 & i & 0 & 0 \end{pmatrix} \qquad \xi = \begin{pmatrix} 0 & 0 & 1 & i \end{pmatrix}$$

or, alternatively,

$$\eta = \begin{pmatrix} 1 & -i & 0 & 0 \end{pmatrix} \qquad \xi = \begin{pmatrix} 0 & 0 & 1 & -i \end{pmatrix}.$$

In either case, the elements of our self-orthogonal 2-planes do not have the necessary 'realness' required by $\mathbb{M}$, i.e. they do not obey $\eta = \mathbf{C}\eta$ where $\mathbf{C}$ is defined in an earlier problem.

18.6. Show all this. *Hint*: It is handy to make use of coordinates x, y, and w, where $w = (t - z - 1/\lambda)\sqrt{\lambda} = (1 - t - z)/\sqrt{\lambda}$.

We will take Penrose's suggestion, as such a substitution encodes the planes we will use to take slices of our light cone. Rewriting our light cone as

$$(t-z)(t+z) - x^2 - y^2 = 0,$$

we can remove t and z by making the substitutions $t - z = (1 + \sqrt{\lambda}w)/\lambda$ and $t + z = (1 - \sqrt{\lambda}w)$. Doing so, we find the our slices are described by

$$\frac{1}{\lambda}(1+\sqrt{\lambda}w)(1-\sqrt{\lambda}w) - x^2 - y^2 = \frac{1}{\lambda} - w^2 - x^2 - y^2 = 0,$$

or, alternatively,

$$w^2 + x^2 + z^2 = \frac{1}{\lambda}.$$

This, of course, is the metric on a sphere of radius $1/\sqrt{\lambda}$.

These coordinates have a singularity at $\lambda = 0$. To show the parabolic nature of the metric in this case, we simply set $\lambda = 0$ in the equation for our plane, obtaining $z + t = 2$. Removing t, our metric becomes

$$x^2 + y^2 + 4z = 4,$$

which describes a paraboloid.

18.7. Show why the hyperbolic straight lines are repesented as straight in the 'Klein' case and by circles meeting the boundary orthogonally in the 'Poincare' case,, indicating, by use of a 'signature flip' why this second case is indeed conformal.

Hyperbolic straight lines, or geodesics, are given by intersections of $\mathcal{H}^+$ with a plane through the origin. The projective plane, then, naturally encodes these structures as straight lines—the projection is simply one plane interescted with another.

The conformal projection is considerably more complicated. To characterize geodesics, let us describe them first on the projective plane. All such structures are straight lines here, so let us rotate our coordinate system until the line of interest lies parallel to the x-axis, as then it takes the very simple form $t = 1$, $x = t\lambda$, and $y = tk$, where λ is a real number parameterizing our line and k is a constant denoting the distance our line is from the x-axis.

If we allow t to change values, we obtain a vector emenating from the origin, with $t = 1$ being our projective plane. This vector intersects the hyperboloid when its coordinates obey the equation

$$t^2 - x^2 - y^2 = t^2 - (t\lambda)^2 - (tk)^2 = 1.$$

Since the hyperboloid $\mathcal{H}^+$ exists only for positive t values, we can reduce this requirement to $t = 1/\sqrt{1-\lambda^2-k^2}$, the positive square root of t^2. So our hyperbolic geodesic is parameterized by the vector

$$\Big(\frac{1}{\sqrt{1-\lambda^2-k^2}}, \frac{\lambda}{\sqrt{1-\lambda^2-k^2}}, \frac{k}{\sqrt{1-\lambda^2-k^2}}\Big).$$

The easiest way to project this conformally is to imagine stretching our vector by 1 in the t direction and normalizing t (projecting to the plane $t = 1$, which is really the $t = 0$ plane of our original vector), i.e. $(t, t\lambda, tk) \to (1, t\lambda/(t+1), tk/(t+1))$. Carrying out this operation gives a projected geodesic vector of

$$\Big(1, \frac{\lambda}{1+\sqrt{1-\lambda^2-k^2}}, \frac{k}{1+\sqrt{1-\lambda^2-k^2}}\Big).$$

Graphing such a curve gives a clear circle shifted in the y direction (our last component). To explicitly illustrate this, consider $x^2 + (y-c)^2$, or

$$\begin{aligned}\frac{\lambda^2}{(1+\sqrt{1-\lambda^2-k^2})^2} + \left(\frac{k}{1+\sqrt{1-\lambda^2-k^2}} - c\right)^2 =& \frac{\lambda^2}{(1+\sqrt{1-\lambda^2-k^2})^2} + \frac{k^2}{(1+\sqrt{1-\lambda^2-k^2})^2} \\ &- \frac{2kc}{1+\sqrt{1-\lambda^2-k^2}} + c^2 \\ =& \frac{\lambda^2 + k^2 - 2kc(1+\sqrt{1-\lambda^2-k^2}) + c^2(1+\sqrt{1-\lambda^2-k^2})^2}{(1+\sqrt{1-\lambda^2-k^2})^2} \\ =& \frac{2(c^2-kc)(1+\sqrt{1-\lambda^2-k^2}) - (c^2-1)\lambda^2 - (c^2-1)k^2}{2(1+\sqrt{1-\lambda^2-k^2}) - \lambda^2 - k^2}.\end{aligned}$$

This is equal to a constant when $c^2 - kc = c^2 - 1$, or $c = 1/k$. Hence, hyperbolic geodesics on $\mathcal{H}^+$ get mapped to circles contained within the unit circle when projected in this way.

To see that these circles meet the unit circle at right angles, we must first find the condition for when two arbitrary circles meet at right angles. This is when their radii r_1, $r-2$ can be considered as two legs to a right triangle, and hence the distance d between their center can be found via $d^2 = r_1^2 + r_2^2$. In the case of our projected circle, it has a radius of $\sqrt{1/k^2 - 1}$, while our unit circle has a radius of 1. The two are separated by a distance $1/k$. Substituting these into our relationship between $d^2 = r_1^2 + r_2^2$ yields an identity,

$$\frac{1}{k^2} = 1^2 + \frac{1}{k^2} - 1 = \frac{1}{k^2},$$

showing that the circle does indeed meet the boundary of the unit circle orthogonally. That this projection is a conformal follows from the fact that a projection of the complex sphere is conformal, and taking a 'real' slice of such a sphere maintains this property.

18.8. Use a 'signature-flip' argument, to see why adding lengths in hyperbolic geometry should give rise to the addition formula being used here, namely $(u+v)c/(1+uv)$, for 'adding' the velocities uc and vc in the same spatial direction. Consider adding arc lengths around a circle or sphere, the 'velocity' corresponding to each arc length being the tangent of the angle it subtends at the centre.

If such a hyperbolic surface is taken to be the velocity space, then any velocity vector can be reached from another by a hyperbolic rotation. This is because such a surface is defined by the vectors satisfying $g_{ab}v^a v^b = c$ for some constant c.

Instead of jumping directly to adding angles within hyperbolic geometry and considering their tangents the velocities of interest, let us examine the same phenomenon on a circle. That is, what is the tangent of θ_w given that we know the tangents of θ_u and θ_v, and $\theta_w = \theta_u + \theta_v$. By simple trigonometric identities (which may most easily be derived by considering the rotation matrix in

$\mathbb{R}^2$), we have

$$\begin{aligned}
\tan\theta_w &= \tan(\theta_u+\theta_v)\\
&= \frac{\sin(\theta_u+\theta_v)}{\cos(\theta_u+\theta_v)}\\
&= \frac{\sin\theta_u\cos\theta_v+\cos\theta_u\sin\theta_v}{\cos\theta_u\cos\theta_v-\sin\theta_u\sin\theta_v}\\
&= \frac{\tan\theta_u+\tan\theta_v}{1-\tan\theta_u\tan\theta_v}.
\end{aligned}$$

Considering these tangents to be velocities, we have

$$w = \frac{u+v}{1-uv}.$$

We have been working in $\mathbb{CE}^4$ with our attention restricted to one direction. If we switch to Minkowski space, we must take all of our velocities to be imaginary (as they are 3-velocities, and these components are purely imaginary in our scheme). Then we have

$$iw = \frac{iu+iv}{1-i^2uv},$$

or

$$w = \frac{u+v}{1+uv}$$

where we have set $c=1$.

18.9. Justify this assertion; prove the equivalence of the above two displayed formulae.

For speeds $v \ll 1$, we may expand ρ in v to find

$$\begin{aligned}
\rho &= \frac{1}{2}\log\frac{1+v}{1-v}\\
&= \frac{1}{2}\log(1+v)-\frac{1}{2}\log(1-v)\\
&\approx \frac{1}{2}\Big(0+v+\mathcal{O}(v^2)\Big)-\frac{1}{2}\Big(0-v+\mathcal{O}(v^2)\Big)\\
&= v+\mathcal{O}(v^2)
\end{aligned}$$

To show the equivalence of the two given formulas, we may simply solve for v given ρ,

$$\begin{aligned}
\rho &= \frac{1}{2}\log\frac{1+v}{1-v}\\
e^\rho &= \sqrt{\frac{1+v}{1-v}}\\
(1-v)e^{2\rho} &= 1+v\\
v(1+e^{2\rho}) &= e^{2\rho}-1\\
ve^\rho(e^\rho+e^{-\rho}) &= e^\rho(e^\rho-e^{-\rho})\\
v &= \frac{e^\rho-e^{-\rho}}{e^\rho+e^{-\rho}}
\end{aligned}$$

18.10. Try to fill in the details of an ingenious argument ofr this, due to the highly original and influential Irish relativity theorist John L. Synge, which requires no calculation! THe argument proceeds roughly as follows. Consider the geometrical configuration consisting o fhte past light cone $\mathcal{C}$ of an event $\mathcal{O}$ and a (timelike) 3-plane P through O. Let Σ be the intersection of $\mathcal{C}$ and $\mathcal{P}$. Describe the 'history', as time progresses, of the respective spatial descriptions of $\mathcal{C}$, $\mathcal{P}$, and Σ, according to some particular Minkowskian reference frame. Explain why any observer at O see Σ as a circle and, moreover, that this geometrical constructon characterizes, in a frame-independent way, those bundles of rays that appear to an observer as a circle.

18.11. Derive this formula.

Geometrically, it is obvious that a great circle on our sphere will map to a line on the plane under stereographic projection. If this plane is to be identified with $\mathbb{C}$, this amounts to saying that the projection will take the form

$$\zeta = R(\theta)e^{i\phi}.$$

Furthermore, we must have that $R(\pi) = 0$, $R(\pi/2) = 1$, and $R(0) = \infty$. This is accomplished in a smooth manner by making the identification $R(\theta) = \cot \frac{1}{2}\theta$, and so

$$\zeta = e^{i\phi} \cot \frac{1}{2}\theta.$$

18.12. Try to derive this formula using the spacetime geometry ideas above.

Consider two observers moving away from each other with a velocity v. We will concentrate on the direction of motion and only consider one spatial dimension. Let us denote the coordinates of the observer at rest with t and x, and those of the observer in motion with t' and x'. We can find the intersection of a hyperbolic space-like surface with our two observer's time axes through the relation

$$t^2 - x^2 = t'^2 - x'^2$$

where we set $x = vt$ and $x' = 0$. This gives the relationship

$$t' = t\sqrt{1 - v^2}.$$

That is, an observer sees those in motion with (respect to themselves) as having slower clocks.

Now, we know that any change in velocity maintains the causal structure of spacetime, i.e. light cones are preserved. This means the lines described by

$$\frac{t}{x} = 1 \qquad \frac{t'}{x'} = 1$$

must coincide. We can substitute our found expression for t' into this and further express x' as a corrected x via $x' = \gamma(v)x$. This requirement becomes

$$\frac{t}{x} = 1 \iff \frac{t\sqrt{1 - v^2}}{\gamma(v)x} = \frac{\sqrt{1 - v^2}}{\gamma(v)} = 1.$$

Ergo, $\gamma(v) = \sqrt{1 - v^2}$, and so distances moving at a speed relative to an observer are squashed. Reinserting units of c, this becomes $x' = x\sqrt{1 - (v/c)^2}$.

18.13. Develop this argument in detail, to show why the FitzGerald-Lorentz flattening exactly compensates for the effect arising from the path-length difference. Show that for small angular diameter, the apparent effect is a rotation of the sphere, rather than a flattening.

18.14. Use conservation of energy and momentum to show that if a stationary billiard ball is hit by another of the same mass, then they emerge at right angles (assuming an elastic collision, so there is no conversion of kinetic energy to heat).

When all masses are the same, conservation of momentum reduces to

$$\mathbf{v}_1 = \mathbf{v}_2 + \mathbf{v}_3.$$

Since energy is proportional to the square of the speed, we are encouraged to take the squared magntiude of this expression,

$$|\mathbf{v}_1|^2 = |\mathbf{v}_2|^2 + 2\mathbf{v}_2 \cdot \mathbf{v}_3 + |\mathbf{v}_3|^2.$$

This is an elastic collision, and so

$$\frac{1}{2}m|\mathbf{v}_1|^2 = \frac{1}{2}m|\mathbf{v}_2|^2 + \frac{1}{2}m|\mathbf{v_3}|^2,$$

or $|\mathbf{v}_1|^2 = |\mathbf{v}_2|^2 + |\mathbf{v}_3|^2$. Taken together with the second equation, this implies $\mathbf{v}_2 \cdot \mathbf{v}_3 = 0$, which shows that the particles depart at a right angle to each other.

18.15. Show all this.

18.16. Why do spinning skaters pull in their arms to increase their rate of rotation?

In order for $\mathbf{M}$ to be conserved, a reduction in $\mathbf{x}$ necessarily implies an increase on $\mathbf{p}$. To see this explicitly, we can expand $\mathbf{M}$ as

$$M^{ab} = x^a p^b - x^b p^a,$$

where it is now clear that the momentum increases in a way consistent with rotation.

18.17. Show this. (N.B. The position vector of the mass centre is the sum of the quantities m$\mathbf{x}$ divided by the sum of the masses m.)

18.18. Show that the formula for the increases mass is $m(1 - v^2/c^2)^{-1/2}$, where v is the velocity of the particle in the second frame; see below.

We are given that the 4-momentum is the 4-velocity scaled by the rest mass μ. Let us rotate our coordinate system so that the velocity vector of interest points along one axis, i.e. $v^a = (1, v^1, 0, 0)$. Reexamining the velocity space $\mathcal{H}^+$ considered earlier, we see that v^0 is given by $\cosh\rho$ and v^1 is given by $\sinh\rho$. Furthermore, from the formula for rapidity, we know that $\sinh\rho = v_3\cosh\rho$, where v_3 is the magnitude of the 3-velocity. So our four-momentum is

$$p^a = \mu\cosh\rho(1, v_3, 0, 0)$$

From inspection, we immediately see that $\mu\cosh\rho = m$, as this coincides with $p^a = (E, -\mathbf{p})$. The hyperbolic cosine function can be expressed in terms of the 3-velocity by a careful examination of Fig. 18.11. Being as the 4-velocity is normalized to unity time, $\cosh\rho$ will be equal to this normalization factor.

We can find $\cosh\rho$ from the formula given in §18.4,

$$\begin{aligned}\cosh\rho &= \frac{1}{2}e^{\rho} + \frac{1}{2}e^{-\rho}\\ &= \frac{1}{2}\sqrt{\frac{1+v}{1-v}} + \frac{1}{2}\sqrt{\frac{1-v}{1+v}}\end{aligned}$$

and so

$$\begin{aligned}(\cosh\rho)^2 &= \frac{1}{4}\Big(\frac{1+v}{1-v}\Big) + \frac{1}{2}\sqrt{\frac{1+v}{1-v}\frac{1-v}{1+v}} + \frac{1}{4}\Big(\frac{1-v}{1+v}\Big)\\ &= \frac{1}{4}\Big(\frac{2+2v^2}{1-v^2}\Big) + \frac{1}{2}\\ &= \frac{1}{1-v^2}\end{aligned}$$

Putting this all together gives

$$m = \frac{\mu}{\sqrt{1-v^2}}.$$

18.19. Why?

The first equation, $\mathbf{p} = m\mathbf{v}$, is simply the definition of the 3-momentum. The second and third were found in the previous exercise.

18.20. Use the Taylor series of §6.4 to derive $(1+x)^{1/2} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 - \cdots$ Hence, obtain a power series expansion for the energy $E = [(c^2\mu)^2 + c^2\mathbf{p}^2]^{1/2}$ of a particle of rest-mass μ and 3-momentum $\mathbf{p}$. Show that the leading term is just Einstein's $E = mc^2$ applied to the rest energy μ, and that the next term is the Newtonian expression for kinetic energy. Write down the next two terms, so as to give better approximations to the full relativistic energy.

If we remove $c^2\mu$ from the radical, we obtain $E = c^2\mu\sqrt{1+(v/c)^2}$. When $v \ll c$, we may expand this in $(v/c)^2$ to find

$$\begin{aligned}E &= \mu c^2\Big(1 + \frac{1}{2}\frac{v^2}{c^2} - \frac{1}{8}\frac{v^4}{c^4} + \frac{1}{16}\frac{v^6}{c^6} + \mathcal{O}(\frac{v^8}{c^8})\Big)\\ &= \mu c^2 + \frac{1}{2}\mu v^2 - \frac{1}{8}\mu\frac{v^4}{c^2} + \frac{1}{16}\mu\frac{v^6}{c^4} + \mathcal{O}\Big(\frac{v^8}{c^8}\Big)\end{aligned}$$

18.21. Why?

On the particle's worldline, the only nonzero components of 4-momentum and 4-position are p^0 and x^0, respectively—to itself, the particle is at rest and at the origin of its coordinate system. So the only possible component of M^{ab} that might be nonzero is M^{00}, but this is zero by the antisymmetric nature of the angular momentum tensor.

18.22. Explain, in detail, in the relativistic case.

19 The classical fields of Maxwell and Einstein

19.1. Check both these statement.

The raised version of the Levi-Civita symbol is

$$\varepsilon^{0123} = g^{0a} g^{1b} g^{2c} g^{3d} \varepsilon_{abcd}$$

which, given that $g^{ab} = \mathrm{diag}(1, -1, -1, -1)$, reduces to $\varepsilon^{0123} = -\varepsilon_{0123}$. Given that $\varepsilon_{0123} = 1$, then, we have $\varepsilon^{0123} = -1$.

Since ϵ is defined by the normalization $\epsilon \cdot \varepsilon = n!$, the choice of $\epsilon^{abcd} = -\varepsilon^{abcd}$ makes sense, as then we have $\epsilon \cdot \varepsilon = \varepsilon^{abcd} \varepsilon_{abcd} = n!$.

19.2. Write these out fully, in terms of the electric and magnetic field components, showing how these equations provide a time-evoluton of the electric and magnetic fields, in terms of the operator $\partial/\partial t$

The exterior derivative of $\mathbf{F}$ is given by

$$\begin{aligned} \mathrm{d}\mathbf{F} &= \frac{\partial}{\partial x^{[a}} F_{bc]} \\ &= \frac{1}{6}\Big(\frac{\partial F_{bc}}{\partial x^a} - \frac{\partial F_{cb}}{\partial x^a} + \frac{\partial F_{ca}}{\partial x^b} - \frac{\partial F_{ac}}{\partial x^b} + \frac{\partial F_{ab}}{\partial x^c} - \frac{\partial F_{ba}}{\partial x^c}\Big). \end{aligned}$$

When we choose $a = 0$, $b = 1$ and $c = 2$, this becomes

$$\frac{1}{3}\Big(- \frac{\partial B_3}{\partial t} - \frac{\partial E_2}{\partial x} + \frac{\partial E_1}{\partial y}\Big) = 0,$$

or

$$\frac{\partial B_3}{\partial t} = \frac{\partial E_1}{\partial y} - \frac{\partial E_2}{\partial x}.$$

Likewise, $a = 0$, $b = 1$, and $c = 3$ gives

$$\frac{\partial B_2}{\partial t} = \frac{\partial E_3}{\partial x} - \frac{\partial E_1}{\partial z},$$

and $a = 0$, $b = 2$, and $c = 3$ results in

$$\frac{\partial B_1}{\partial t} = \frac{\partial E_2}{\partial z} - \frac{\partial E_3}{\partial y}.$$

Finally, $a = 1$, $b = 2$ and $c = 3$ gives

$$\frac{\partial B_1}{\partial x} + \frac{\partial B_2}{\partial y} + \frac{\partial B_3}{\partial z} = 0.$$

We can interpret the first three of these to be expressing Faraday's law, which states that a changing magnetic field is equal to the negative curl of the electric field. The final equation encodes the lack of magnetic monopoles (the divergence of the magnetic field is zero).

The exterior derivative of ${}^*\mathbf{F}$ is given by a similar expression,

$$\mathrm{d}^*\mathbf{F} = \frac{1}{6}\Big(\frac{\partial^* F_{bc}}{\partial x^a} - \frac{\partial^* F_{cb}}{\partial x^a} + \frac{\partial^* F_{ca}}{\partial x^b} - \frac{\partial^* F_{ac}}{\partial x^b} + \frac{\partial^* F_{ab}}{\partial x^c} - \frac{\partial^* F_{ba}}{\partial x^c}\Big) = \frac{4}{3}\pi^* J_{abc}.$$

Choosing $a = 0$, $b = 1$, and $c = 2$ yields

$$\frac{1}{3}\Big(- \frac{\partial E_3}{\partial t} + \frac{\partial B_2}{\partial x} - \frac{\partial B_1}{\partial y} \Big) = \frac{4}{3}\pi j_3,$$

or

$$\frac{\partial E_3}{\partial t} = \frac{\partial B_2}{\partial x} - \frac{\partial B_1}{\partial y} - 4\pi j_3.$$

Moving on, choosing $a = 0$, $b = 1$, and $c = 3$ gives us

$$\frac{\partial E_2}{\partial t} = \frac{\partial B_1}{\partial z} - \frac{\partial B_3}{\partial x} - 4\pi j_2,$$

while $a = 0$, $b = 2$, and $c = 3$ yields

$$\frac{\partial E_1}{\partial t} = \frac{\partial B_3}{\partial y} - \frac{\partial B_2}{\partial z} - 4\pi j_1.$$

Lastly, $a = 1$, $b = 2$, and $c = 3$ results in

$$\frac{\partial E_1}{\partial x} + \frac{\partial E_2}{\partial y} + \frac{\partial E_3}{\partial z} = 4\pi\rho.$$

We may interpret the first three of *these* equations as Ampere's law expressing the change in the electric field as the curl of the magnetic field minus the current density. The final equation is Gauss's law, relating the divergence of the electric field to the enclosed charge density.

19.3. Show the equivalence to the previous pair of equations.

To show the equivalence of the two sets of equations, we may make the replacement $\partial_a \to \nabla_a$, as this is the generalization of the above equations to the curvilinear case. The first equations equivalence is immediately apparent, as it becomes

$$\mathrm{d}\mathbf{F} = \nabla_{[a} F_{bc]} = 0.$$

For the second equation, we may make use of the index expression of the Hodge dual to obtain

$$\begin{aligned} \mathrm{d}^*\mathbf{F} &= \nabla^*_{[a} F_{bc]} \\ &= \frac{1}{2}\nabla_{[a}\varepsilon_{bc]de}F^{de} \\ &= \frac{1}{6}\Big(\nabla_a \varepsilon_{bcde}F^{de} + \nabla_b \varepsilon_{cade}F^{de} + \nabla_c \varepsilon_{abde}F^{de}\Big). \end{aligned}$$

Meanwhile, the righthand side is given by

$$\begin{aligned} \frac{4}{3}\pi^*\mathbf{J} &= \frac{4}{3}\pi^* J_{abc} \\ &= \frac{4}{3}\pi\varepsilon_{abcd}J^d. \end{aligned}$$

The combined expression has four possible values, set by the antisymmetry of a, b, and c. When $a = 0$, $b = 1$, and $c = 2$, we get

$$\frac{1}{6}\Big(\nabla_0(F^{03} - F^{30}) + \nabla_1(F^{13} - F^{31}) + \nabla_2(F^{23} - F^{32})\Big) = \frac{4}{3}\pi J^3,$$

or, by the antisymmetry of F^{ab},

$$\nabla_a F^{a3} = 4\pi J^3.$$

Doing this for the remaining three valid permutations gives

$$\begin{aligned}\nabla_a F^{a2} &= 4\pi J^2\\ \nabla_a F^{a1} &= 4\pi J^1\\ \nabla_a F^{a0} &= 4\pi J^0\end{aligned}$$

which is succinctly encapsulated by $\nabla_a F^{ab} = 4\pi J^b$.

19.4. Show that the two versions of this vanishing divergence are equivalent.

Following the same steps as above, we may expand the expression for the Hodge dual of the charge-current vector as

$$\begin{aligned}\mathrm{d}^*\mathbf{J} &= \nabla_{[a}{}^*J_{bcd]}\\ &= \frac{1}{4}\Big(\nabla_a\varepsilon_{bcde}J^e + \nabla_b\varepsilon_{dcae}J^e + \nabla_c\varepsilon_{bdae}J^e + \nabla_d\varepsilon_{bace}J^e\Big).\end{aligned}$$

Because both $\{a, b, c, d\}$ and $\{b, c, d, e\}$ are completely antisymmetric, a moment's thought reveals that the coordinate shared by ∇ will also by shared by $\mathbf{J}$, and therefore this expression becomes

$$\frac{1}{4}\nabla_a J^a = 0$$

which, of course, is equivalent to $\nabla_a J^a = 0$.

19.5. Show this, first demonstrating that dualizing twice yields minus the original quantity. Does this sign relate to the Lorentzian signature of spacetime? Explain.

The relationship between the raised and lowered volume element is, as given by Penrose, $\varepsilon^{abcd} = -\varepsilon_{abcd}$. Utilizing this, and from the definition of the Hodge dual, it is easy to see that dualizing twice (which introduces both the raised and lowered volume element) returns the negative of the original quantity. Explicitly, we have

$$\begin{aligned}{}^*({}^*\mathbf{F}) &= {}^*\Big(\frac{1}{2}\varepsilon_{abcd}F^{cd}\Big)\\ &= -\frac{1}{4}\varepsilon^{efab}\varepsilon_{abcd}F^{cd}\\ &= -\frac{1}{2}(\delta^e_c\delta^f_d - \delta^e_d\delta^f_c)F^{cd}\\ &= -\frac{1}{2}(F^{ef} - F^{fe})\\ &= -F^{ef}\\ &= -\mathbf{F},\end{aligned}$$

where the fourth to last line uses the expression for the partial contraction of two volume elements given in Fig. 12.18.

Knowing that dualizing a quantity twice returns negative the original, we see

$$\begin{aligned}
{}^*({}^+\mathbf{F}) &= \frac{1}{2}({}^*\mathbf{F} + i\mathbf{F}) \\
&= i\frac{1}{2}(\mathbf{F} - i{}^*\mathbf{F}) \\
&= i{}^+\mathbf{F}
\end{aligned}$$

and

$$\begin{aligned}
{}^*({}^-\mathbf{F}) &= \frac{1}{2}({}^*\mathbf{F} - i\mathbf{F}) \\
&= -i\frac{1}{2}(\mathbf{F} + i{}^*\mathbf{F}) \\
&= -i{}^-\mathbf{F}.
\end{aligned}$$

This sign does indeed relate to the Lorentzian signature of our metric, as the sign reversal of the antisymmetric symbol upon index juggling occurs for this reason.

19.6. Can you spell this out? What happens to the components of $\mathbf{F}$ and ${}^*\mathbf{F}$ in a general curvilinear coordinate system? Why are the Maxwell equations unaffected if expressed correctly?

19.7. Although correct, this argument has been given somewhat glibly. Spell out hte details more fully, in teh case when $\mathcal{R}$ is a spacetime 'cylinder' consisting of some bounded spatial region that is constant in time, for a fixed finite interval of the time coordinate t. Explain the different notions of 'flux of charge' involved, contrasting this for spacelike 'base' and 'top' of the cylinder with that for the timelike 'sides'.

19.8. Spell out why this is just the electric flux.

The spacelike components of ${}^*\mathbf{F}$ are those of the electric field, so any integral for fixed time of this quantity will necessarily refer to the electric field. Hence, Gauss's law applied to this quantity relates the flux of the electric field to the charge enclosed by the spatial cross-section of spacetime under consideration.

19.9. Why can we add such a quantity?

Because $\mathrm{d}^2 = 0$, or, explicitly,

$$\begin{aligned}
\mathrm{d}(\mathbf{A} + \mathrm{d}\Theta) &= \mathrm{d}\mathbf{A} + \mathrm{d}^2\Theta \\
&= \mathrm{d}\mathbf{A}.
\end{aligned}$$

Θ must be a scalar to have a well-defined quantity: $\mathbf{A}$ is a one-form, and, if Θ is a scalar, $\mathrm{d}\Theta$ is also a one-form.

19.10. Show this. *Hint*: Have a look at §15.8.

This is easily seen by referring to the coordinate expression of the curvature tensor given a connection,

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)\xi^d = R_{abc}{}^d\xi^c.$$

We will modify the above in accordance with §15.8, where the last two indices refer to the directions in the fibre. Since our fibre is $U(1)$, these coordinates refer to just one direction, and so we will omit them. Given a connection of $\nabla_a = \partial/\partial x^a - ieA_a$, the above expression then becomes

$$\begin{aligned}(\nabla_a\nabla_b - \nabla_b\nabla_a)\psi &= \Big(\frac{\partial}{\partial x^a} - ieA_a\Big)\Big(\frac{\partial\psi}{\partial x^b} - ieA_b\psi\Big) - \Big(\frac{\partial}{\partial x^b} - ieA_b\Big)\Big(\frac{\partial\psi}{\partial x^a} - ieA_a\psi\Big)\\ &= \frac{\partial^2\psi}{\partial x^a\partial x^b} - ie\frac{\partial A_b}{\partial x^a}\psi - ieA_b\frac{\partial\psi}{\partial x^a} - ieA_a\frac{\partial\psi}{\partial x^b} - e^2A_aA_b\psi\\ &\quad - \frac{\partial^2\psi}{\partial x^b\partial x^a} + ie\frac{\partial A_a}{\partial x^b}\psi + ieA_a\frac{\partial\psi}{\partial x^b} + ieA_b\frac{\partial\psi}{\partial x^a} + e^2A_bA_a\psi\\ &= -ie\Big(\frac{\partial A_b}{\partial x^a} - \frac{\partial A_a}{\partial x^b}\Big)\psi\\ &= -ieF_{ab}\psi\end{aligned}$$

where the last line is the indicial expression for $\mathbf{F} = 2\mathrm{d}\mathbf{A}$.

19.11. Explain this.

Because $\mathrm{d}\mathbf{A}$ vanishes within $\mathcal{R}$, expanding or otherwise distorting the contour of integration does not change the resulting value. If it did, the region the loop would be entering would have some nonzero value. This is somewhat analagous to contour integration around poles in the complex plane, wherein the analyticity of the function outside the singularity would give zero if that pole were removed—it is the global nature of the function that introduces a non-zero value.

19.12. How do the individual components $T^a{}_b$ relate to T_{ab}, in a local Minkowskian frame, where the components g_{ab} have the diagonal form $(1, -1, -1, -1)$?

Upon raising the a index, $T^a{}_b = g^{ac}T_{cb}$, we see that the only surviving terms in the summation are those where $c = a$. When $a = 0$, these terms are exactly T_{ab}, while $a \neq 0$ resuts in $-T_{ab}$. Altogether, we have

$$\begin{aligned}T^0{}_b &= T_{0b},\\ T^1{}_b &= -T_{1b},\\ T^2{}_b &= -T_{2b},\\ T^3{}_b &= -T_{3b}.\end{aligned}$$

19.13. Show that this satisfies the conservation equation $\nabla^a T_{ab} = 0$ if $\mathbf{J} = 0$. Obtain the 00 component of this tensor, and recover Maxwell's original expression $(E^2 + B^2)/8\pi$ for the energy density of an electromagnetic field in terms of (E_1, E_2, E_3) and (B_1, B_2, B_3).

When $\mathbf{J} = 0$, ${}^*\mathbf{J} = 0$, and we have tight symmetry between both of Maxwell's equation in the language of differential forms, namely

$$\begin{aligned} \mathrm{d}\mathbf{F} &= 0, \\ \mathrm{d}^*\mathbf{F} &= 0. \end{aligned}$$

By this symmetry we can express the indicial versions of these equations in two ways,

$$\begin{aligned} \nabla_a F^{ab} &= 0, \\ \nabla_{[a} F_{bc]} &= 0, \\ \nabla_a {}^*F^{ab} &= 0, \\ \nabla_{[a} {}^*F_{bc]} &= 0. \end{aligned}$$

Using these identities, and contracting the claimed energy-momentum tensor with the raised covariant derivative, gives

$$\begin{aligned} \nabla^a\left(\frac{1}{8\pi}\left(F_{ac}F^c{}_b + {}^*F_{ac}{}^*F^c{}_b\right)\right) &= \frac{1}{8\pi}(F^c{}_b\nabla^a F_{ac} + F_{ac}\nabla^a F^c{}_b + {}^*F^c{}_b\nabla^{a*}F_{ac} + {}^*F_{ac}\nabla^{a*}F^c{}_b) \\ &= \frac{1}{8\pi}(F_{cb}\nabla_a F^{ac} + F^{ac}\nabla_a F_{cb} + {}^*F_{cb}\nabla_a{}^*F^{ac} + {}^*F^{ac}\nabla_a{}^*F_{cb}) \\ &= \frac{1}{8\pi}(F^{ac}\nabla_a F_{cb} + {}^*F^{ac}\nabla_a{}^*F_{cb}) \end{aligned}$$

where the third line has used the fact that both $\nabla_a F^{ab}$ and $\nabla_a{}^*F^{ab}$ are zero. The remaining terms are also zero. To see this, notice that both are antisymmetry in *ac* as well as *cb*, thereby being antisymmetric in *abc*. But the antisymmetric parts of $\nabla_a F_{bc}$ and $\nabla_a{}^*F_{bc}$ vanish, and so T_{ab} satisfies the conservation equation $\nabla^a T_{ab} = 0$.

19.14. Why? Why does this procedure specialize to the above $\nabla_a T^a{}_0 = 0$, etc.? Can you find an analogue of the continuous-field conservation law $\nabla^a(T_{ab}\kappa^b) = 0$, for a discrete system of particles where 4-momentum is conserved in collisons? *Hint*: Find a quantity, given the Killing vector κ^a, that is constant for each particle between collisions.

Applying the covariant derivative to L_a yields

$$\begin{aligned} \nabla^a L_a &= \nabla^a(T_{ab}\kappa^b) \\ &= \kappa^b\nabla^a T_{ab} + T_{ab}\nabla^a\kappa^b. \end{aligned}$$

When the energy-momentum tensor satisifes the conservation law $\nabla^a T_{ab} = 0$, the first term vanishes. Meanwhile, the second term vanishes because, by virtue of $\nabla^a\kappa^a$ being contracted with a symmetric tensor, it is totally symmetric, and the symmetric part of $\nabla^a\kappa^a$ is zero (see the indicial definition of a Killing vector).

I am unsure how to answer the second part of this question.

19.15. Why is R_{ab} symmetric?

This is simply because

$$\begin{aligned}
R_{ab} &= R_{acb}{}^{c} \\
&= R_{acbd}g^{dc} \\
&= R_{bdac}g^{dc} \\
&= R_{bda}{}^{d} \\
&= R_{ba}.
\end{aligned}$$

19.16. See if you can prove this using the Ricci identity and the properties of Lie derivative.

19.17. Show fully why we can 'lop off' all the t^as, explaining the role of the symmetry of the tensors.

Clearly, $t^a t^b$ is symmetric, and thus both $R_{ab}t^a t^b$ and $T_{ab}t^a t^b$ effectively symmetrize both the Ricci tensor and the energy momentum tensor, effectively stating

$$R_{(ab)} = T_{(ab)}.$$

Since R_{ab} and T_{ab} are manifestly symmetric, $R_{(ab)} = R_{ab}$ and $T_{(ab)} = T_{ab}$, allowing us to continue holding equality between both sides of the expression when $t^a t^b$ is removed.

19.18. Show this, using the diagrammatic notation, if you like.

Expanding the Ricci identity, we find

$$\begin{aligned}
\nabla_{[a}R_{bc]d}{}^{e} &= \frac{1}{6}(\nabla_a R_{bcd}{}^{e} - \nabla_a R_{cbd}{}^{e} + \nabla_b R_{cad}{}^{e} - \nabla_b R_{acd}{}^{e} + \nabla_c R_{abd}{}^{e} - \nabla_c R_{bad}{}^{e}), \\
&= \frac{1}{3}(\nabla_a R_{bcd}{}^{e} - \nabla_b R_{acd}{}^{e} + \nabla_c R_{abd}{}^{e}),
\end{aligned}$$

where the last line uses the antisymmetry of the first two indices of the Riemann tensor. If we contract e with c, we obtain

$$\begin{aligned}
\nabla_{[a}R_{bc]d}{}^{c} &= \frac{1}{3}(\nabla_a R_{bd} - \nabla_b R_{ad} + \nabla_c R_{abd}{}^{c}), \\
&= \frac{1}{3}(\nabla_a R_{bd} - \nabla_b R_{ad} + \nabla^c R_{bacd}),
\end{aligned}$$

where, now, the antisymmetry of both the first and last pairs of indices is utilized in the last line. Finally, raising a and contracting it with d gives

$$\begin{aligned}
g^{ad}\nabla_{[a}R_{bc]d}{}^{c} &= \frac{1}{3}(\nabla^d R_{bd} - \nabla_b R + \nabla^c R_{bc}), \\
&= \frac{1}{3}(2\nabla^a R_{ab} - \nabla_b R), \\
&= \frac{1}{6}\nabla^a(R_{ab} - \frac{1}{2}Rg_{ab}).
\end{aligned}$$

Since this equals zero, we may multiply through by 6 to arrive at

$$\nabla^a(R_{ab} - \frac{1}{2}Rg_{ab}) = 0.$$

19.19. Why?

We can contract the suggested field equation to obtain

$$R = -4\pi GT,$$

where $T = T^a_a$. This can be multipled by $\frac{1}{2}g_{ab}$ and subtracted from the original, uncontracted equation to get

$$R_{ab} - \frac{1}{2}Rg_{ab} = -4\pi GT_{ab} + 2\pi GTg_{ab}.$$

From here, we may take the covariant derivative of both sides (remembering that $\nabla^a T_{ab} = 0$):

$$\begin{aligned}\nabla^a(R_{ab} - \frac{1}{2}Rg_{ab}) &= -4\pi G\nabla^a T_{ab} + 2\pi G\nabla^a Tg_{ab},\\ 0 &= 2\pi G\nabla_b T,\end{aligned}$$

which implies that T^a_a must be constant—but this is clearly not the case in the real world.

19.20. Explain the coefficient $-8\pi G$, as compared with $-4\pi G$.

Following a similar process to that in the next exercise (which was completed first), we find (by considering the diagonal terms of both sides of the field equation) that

$$\begin{aligned}\frac{1}{2}(R_{00} + R_{11} + R_{22} + R_{33}) &= -8\pi GT_{00},\\ \frac{1}{2}(R_{00} + R_{11} - R_{22} - R_{33}) &= -8\pi GT_{11},\\ \frac{1}{2}(R_{00} - R_{11} + R_{22} - R_{33}) &= -8\pi GT_{22},\\ \frac{1}{2}(R_{00} - R_{11} - R_{22} + R_{33}) &= -8\pi GT_{33}.\end{aligned}$$

Adding all of these together yields

$$\frac{1}{2}R_{00} = -8\pi G(T_{00} + T_{11} + T_{22} + T_{33}).$$

Since the energy density (T_{00}) is expected to be much larger than the pressure (T_{11}, T_{22}, T_{33}) in the Newtonian limit, this reduces to

$$R_{00} = -4\pi GT_{00},$$

which is exactly the necessary relationship needed to reproduce the Newtonian volume-acceleration effects, as, in an observer's frame, (where $t^0 = 1$ and all other components are zero), we have

$$\begin{aligned}R_{ab}t^a t^b \delta V &= -4\pi GT_{ab}t^a t^b \delta V,\\ R_{00} &= -4\pi GT_{00}.\end{aligned}$$

19.21. Why?

Clearly, in a vaccuum, the metric is simply the Minkowski metric. Therefore, for all $a \neq b$, the field equation collapses to $R_{ab} = 0$. All that's left is to show that, given a vaccuum, every diagonal component of R_{ab} is also zero.

Recognizing that $R = R^a_a = R_{00} - R_{11} - R_{22} - R_{33}$, we may rewrite the vaccuum field equation as

$$R_{ab} - \frac{1}{2}(R_{00} - R_{11} - R_{22} - R_{33})g_{ab} = 0.$$

When $a = b = 0$, this simplifies to

$$R_{00} + R_{11} + R_{22} + R_{33} = 0.$$

When $a = b = 1$, this becomes

$$R_{00} + R_{11} - R_{22} - R_{33} = 0.$$

Likewise, for the remaining cases, we have

$$R_{00} - R_{11} + R_{22} - R_{33} = 0,$$
$$R_{00} - R_{11} - R_{22} + R_{33} = 0.$$

By adding together the permutations of the bottom three equations, we obtain

$$R_{00} = R_{11} = R_{22} = R_{33},$$

and thus the first equation is only satisifed when all four components are equal to zero. Hence we can write

$$R_{ab} = 0$$

in a vaccuum.

19.22. Why?

By contracting both sides of the field equation, we obtain

$$R - 2R = -8\pi GT,$$

or $R = 8\pi GT$ (where we have used the fact that ${g_a}^b = 4$ in a four-dimensional spacetime). Substituting this back into the field equation gives

$$R_{ab} - 4\pi GTg_{ab} = -8\pi GT_{ab},$$

which may be rearranged to read

$$R_{ab} = -8\pi G(T_{ab} - \frac{1}{2}Tg_{ab}).$$

From here, we may simply add the cosmological constant (being as it is just that: a constant) to the right side,

$$R_{ab} = -8\pi G(T_{ab} - \frac{1}{2}Tg_{ab}) + \Lambda g_{ab}.$$

19.23. Show that all the 'traces' of **C** vanish (e.g. ${C_{abc}}^a = 0$, etc.). Do this calculation in diagrammatic form, if you wish.

20 Lagrangians and Hamiltonians

20.1. Fill in the full details, completing the argument to obtain Galileo's parabolic motion for free fall under gravity.

Looking at the Euler-Lagrange equation for z, we find

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial z} &= -mg,\\ \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{z}} &= \frac{\mathrm{d}}{\mathrm{d}t}(m\dot{z}),\\ &= m\ddot{z}.\end{aligned}$$

Setting both of these equal to each other gives

$$m\ddot{z} = -mg.$$

If we wish to find the motion of the freely falling particle, we may simply divide by m and integrate both sides twice with respect to time, obtaining

$$\int_{-\infty}^{t}\int_{-\infty}^{t'} \ddot{z}\,\mathrm{d}t''\mathrm{d}t' = z(0) + v_z(0)t - \frac{1}{2}gt^2,$$

which is the arc of a parabola.

20.2. Why?

$\mathbf{S}$ satisfies $\mathrm{d}\mathbf{S} = 0$ by virtue of the fact that $\mathrm{d}^2 = 0$, as we have

$$\begin{aligned}\mathrm{d}\mathbf{S} &= \mathrm{d}(\mathrm{d}p_a \wedge \mathrm{d}q^a),\\ &= \mathrm{d}^2 p_a \wedge \mathrm{d}q^a - \mathrm{d}p_a \wedge \mathrm{d}^2 q^a,\\ &= 0.\end{aligned}$$

20.3. Do this explicitly. Use Hamilton's equations to obtian the Newtonian equations fo motion for a particel falling in a constant gravitational field.

From the Lagrangian given in problem 20.1, we may perform a Legendre transform to find the associated Hamiltonian. The conjugate momentum of our Lagrangian are given by

$$\begin{aligned}p_x &= \frac{\partial \mathcal{L}}{\partial \dot{x}} = m\dot{x},\\ p_y &= \frac{\partial \mathcal{L}}{\partial \dot{y}} = m\dot{y},\\ p_z &= \frac{\partial \mathcal{L}}{\partial \dot{z}} = m\dot{z}.\end{aligned}$$

Using this, we see our Hamiltonian is

$$\begin{aligned}\mathcal{H} &= \dot{q}^r \frac{\partial \mathcal{L}}{\partial \dot{q}^r} - \mathcal{L}, \\ &= m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - \Big(\frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - mgz\Big), \\ &= \frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) + mgz, \\ &= \frac{p_x^2 + p_y^2 + p_z^2}{2m} + mgz.\end{aligned}$$

Using Hamilton's equations along the z direction, we find

$$\begin{aligned}\frac{\mathrm{d}p_z}{\mathrm{d}t} &= -\frac{\partial \mathcal{H}}{\partial z} = -mg, \\ \frac{\mathrm{d}z}{\mathrm{d}t} &= \frac{\partial \mathcal{H}}{\partial p_z} = \frac{p_z}{m}.\end{aligned}$$

Taking the time derivative of the second equation and using the first gives

$$\ddot{z} = \frac{\dot{p_z}}{m} = -g,$$

which is the equation of a particle undergoing uniform acceleration. The solution to this equation is given in problem 20.1.

20.4. Confirm this, explaining why $\omega/2\pi$ is the frequency. Explain why the graph of this function still looks like a sine curve. Why is this the *general* solution?

It is trivial to confirm that the given q satisfies the differential equation,

$$\begin{aligned}\frac{\mathrm{d}^2 q}{\mathrm{d}t^2} &= \frac{\mathrm{d}^2}{\mathrm{d}t^2}\Big(a\cos\omega t + b\sin\omega t\Big), \\ &= \frac{\mathrm{d}}{\mathrm{d}t}\Big(-a\omega\sin\omega t + b\omega\cos\omega t\Big), \\ &= -a\omega^2\cos\omega t - b\omega^2\sin\omega t, \\ &= -\omega^2 q.\end{aligned}$$

That this is the general solution can be seen by inserting the inverse Fourier transform of both sides of the differential equation, obtaining

$$\frac{1}{\sqrt{2\pi}}\int Q(\Omega)(i\Omega)^2 e^{i\Omega t}\mathrm{d}\Omega = \frac{1}{\sqrt{2\pi}}\int Q(\Omega)(-\omega^2)e^{i\Omega t}\mathrm{d}\Omega.$$

This holds only when

$$(i\Omega)^2 + \omega^2 = 0$$

which is an algebraic equation with exactly two unique roots, $\Omega = \pm\omega$. Hence $Q(\Omega) = \sqrt{2\pi}\delta(\Omega \pm \omega)$, or $q(t) = e^{\pm i\omega t}$. Since the original differential equation is linear, any superposition of these solutions yields another solution. Using Euler's identity, we may make a change of basis from complex exponentials to trigonometric functions via

$$\cos \omega t = \frac{1}{2} e^{i\omega t} + \frac{1}{2} e^{-i\omega t},$$
$$\sin \omega t = \frac{1}{2j} e^{i\omega t} - \frac{1}{2} e^{-i\omega t},$$

which yields two linearly independent solutions that may be combined just as our original solutions could.

That $\omega/2\pi$ is frequency can be seen from the fact that q undergoes one complete oscillation when $\omega T = 2\pi$, where T is the period of oscillation. But $T = 1/f$, and so $f = \omega/2\pi$.

Finally, the fact that any combination of sine and cosine waves at the same frequency still looks like a sine curve can be explain as

$$\begin{aligned}
a \cos \omega t + b \sin \omega t &= \frac{a}{2}\left(e^{i\omega t} + e^{-i\omega t}\right) + \frac{b}{2j}\left(e^{i\omega t} - e^{-i\omega t}\right), \\
&= \frac{(a - jb)}{2} e^{i\omega t} + \frac{(a + jb)}{2} e^{-i\omega t}, \\
&= \frac{\sqrt{a^2 + b^2} e^{i \arctan(-b/a)}}{2} e^{i\omega t} + \frac{\sqrt{a^2 + b^2} e^{i \arctan(b/a)}}{2} e^{-i\omega t}, \\
&= \frac{\sqrt{a^2 + b^2}}{2}\left(e^{i(\omega t - \arctan(b/a))} + e^{-i(\omega t - \arctan(b/a))}\right), \\
&= \sqrt{a^2 + b^2} \cos\left(\omega t - \arctan \frac{b}{a}\right).
\end{aligned}$$

20.5. Show this, finding the *full* equation, (a) using the Lagrangian method, (b) using the Hamiltonian method, and (c) directly from Newton's laws. *Hint*: Show that $\mathcal{L} = \frac{1}{2} m h^2 \dot{q}^2 (h^2 - q^2)^{-1} + mg(h^2 - q^2)^{1/2}$. (Note that the Lagrangian and Hamiltonian methods do not gain us anything in this simple case; their power resides in treating more general situations.)

In such a simplistic situation, the Lagrangian is simply equal to the kinetic minus the potential energy. The kinetic energy is most easily seen by examining the angle θ the pendulum makes with the downward vertical. In this case, we have

$$T = \frac{1}{2} m v^2 = \frac{1}{2} m (\dot{\theta} h)^2 = \frac{1}{2} m h^2 \dot{\theta}^2.$$

For the potential energy, it is most convenient to take the zero line to be when the pendulum is exactly horizontal, in which case it becomes

$$V = -mgh \cos \theta,$$

where $h \cos \theta$ is simply the vertical coordinate of the pendulum bob.

Altogether, our Lagrangian is

$$\mathcal{L} = T - V = \frac{1}{2} m h^2 \dot{\theta}^2 + mgh \cos \theta.$$

This has the Euler-Lagrange equations

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot\theta} &= \frac{\mathrm{d}}{\mathrm{d}t}\Big(mh^2\dot\theta\Big),\\ &= mh^2\ddot\theta,\\ \frac{\partial \mathcal{L}}{\partial \theta} &= -mgh\sin\theta.\end{aligned}$$

Equating these gives

$$\ddot\theta = -\frac{g}{h}\sin\theta,$$

which, for small angles, is simply

$$\ddot\theta \approx -\frac{g}{h}\theta.$$

The Hamiltonian can be found in a straightforward manner by identifiying the total energy of the system, but instead we will take the route of performing a Legendre transform. Doing so yields

$$\begin{aligned}\mathcal{H} &= \dot\theta\frac{\partial \mathcal{L}}{\partial \dot\theta} - \mathcal{L},\\ &= mh^2\dot\theta^2 - \Big(\frac{1}{2}mh^2\dot\theta^2 + mgh\cos\theta\Big),\\ &= \frac{1}{2}mh^2\dot\theta^2 - mgh\cos\theta,\\ &= \frac{1}{2}\frac{p_\theta^2}{mh^2} - mgh\cos\theta,\end{aligned}$$

where p_θ is the conjugate momentum, given by $\partial\mathcal{L}/\partial\dot\theta = mh^2\dot\theta$. From this, Hamilton's equations give us

$$\dot p_\theta = \frac{\partial \mathcal{H}}{\partial \theta} = mgh\sin\theta, \qquad \dot\theta = -\frac{\partial \mathcal{H}}{\partial p_\theta} = -\frac{p_\theta}{mh^2}.$$

Taking the time derivative of $\dot\theta$ and substituting in our expression for $\dot p_\theta$, we find

$$\ddot\theta = -\frac{g}{h}\sin\theta \approx -\frac{g}{h}\theta,$$

just as we found from the Lagrangian formulation.

Using the standard Newtonian formulation, we recognize that there are two forces acting on our pendulum: gravity, directed downward, and tension, directed along the rod to which our mass is attached. Let us take θ to refer to the angle the pendulum makes with the downward vertical (just as before) and use polar coordinates where a positive movement in θ is taken to mean counterclockwise. Then we have

$$m\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = mg\cos\theta - T = 0 \qquad m\frac{\mathrm{d}^2(h\theta)}{\mathrm{d}t^2} = -mg\sin\theta.$$

where the angular coordinate is $h\theta$ by dimensional considerations. Clearly, this second equation is equivalent to the ones found above, giving

$$\ddot\theta = -\frac{g}{h}\sin\theta \approx -\frac{g}{h}\theta.$$

20.6. Why?

Since $\mathbf{F} = -\nabla U$, we see that U attaining a stationary point (in which its gradient is zero) necessarily implies that our system is in equilibrium (as this is when there are no external forces, $\mathbf{F} = 0$). The condition that this equilibrium be stable implies that a small disturbance about this point produces a restoring force, i.e. $\|\nabla U\| > 0$ at all points near q^a. But this is equivalent to the condition that our stationary point is a minimum.

20.7. Can you explain this all more fully? Can we have the linear terms if the equilibrium is unstable? Explain.

Let us examine the power series expansion of $\mathcal{H}$ about $(0,0)$,

$$\mathcal{H}(q,p) = \mathcal{H}(0,0) + \frac{\partial \mathcal{H}}{\partial q^a}\bigg|_{(0,0)} q^a + \frac{\partial \mathcal{H}}{\partial p_a}\bigg|_{(0,0)} p_a + \frac{1}{2}\frac{\partial^2 \mathcal{H}}{\partial q^a q^b}\bigg|_{(0,0)} q^a q^b + \frac{\partial^2 \mathcal{H}}{\partial q^a \partial p_b}\bigg|_{(0,0)} q^a p_b + \frac{1}{2}\frac{\partial^2 \mathcal{H}}{\partial p_a p_b}\bigg|_{(0,0)} p_a p_b + \cdots$$

Clearly, $\mathcal{H}(0,0)$ is simply a constant—the value of the total energy of the system at its equilibrium point. What about the first order terms? By virtue of $(0,0)$ being a point of equilibrium, these are 0—and so are our terms with mixed partial derivatives. That is, when a minimum occurs about $(0,0)$, our Hamiltonian appears as

$$\mathcal{H}(q,p) = \text{constant} + \frac{1}{2}\frac{\partial^2 \mathcal{H}}{\partial q^a q^b}\bigg|_{(0,0)} q^a q^b + \frac{1}{2}\frac{\partial^2 \mathcal{H}}{\partial p_a p_b}\bigg|_{(0,0)} p_a p_b + \cdots$$

Because our point is a *stable* equilibrium, the second derivatives of $\mathcal{H}$ with respect to momentum or position are necessarily positive, i.e. the matrices being contracted with $q^a q^b$ and $p_a p_b$ are positive definite, and so we arrive at the form Penrose gives

$$\mathcal{H}(q,p) = \text{constant} + \frac{1}{2}Q_{ab}q^a q^b + \frac{1}{2}P^{ab}p_a p_b + \cdots$$

Through the above arguments, it is clear that linear terms preclude any sort of equilibrium, stable or not. An unstable equilibrium instead implies the lack of positive-semi-definiteness of both Q_{ab} and P^{ab}.

20.8. See if you can prove this deduction. *Hint*: Show that the inverse of a positive-definite matrix is positive-definite.

Consider the geometric meaning of eigenvalues: the amount by which space is stretched in the direction of the corresponding eigenvector. Clearly, positive stretching in all directions followed by additional positive stretching leads to a total positive deformation, i.e. the product of two positive-definite matrices is again a positive-definite matrix.

20.9. See if you can carry out the foregoing analysis in the Lagrangian, rather than Hamiltonian, form.

The arguments given by Penrose for $q^a = 0 = p_a$ extend to the Lagrangian formalism, though we will instead refer to the velocity of the generalized coordinates instead of the conjugate momenta, so $q^a = \dot{q}^a = 0$. This gives the requirement that $\mathcal{L}$ be stationary, though it does *not* impose the requirement of a local minimum, instead requiring our Lagrangian to lie on a saddle point. This is because, with the interpretation of

$$\mathcal{L}(q,\dot{q}) = K(q,\dot{q}) - V(q,\dot{q}),$$

any change in q^a increases the value of V (which decreases $\mathcal{L}$) while any change in $\dot{q}^a$ increases the value of K (which increases $\mathcal{L}$).

Assuming $\mathcal{L}$ is analytic, we may expand it in a power series to find

$$\mathcal{L}(q,\dot{q}) = \mathcal{L}(0,0) + \frac{\partial \mathcal{L}}{\partial q^a}\Big|_{(0,0)} q^a + \frac{\partial \mathcal{L}}{\partial \dot{q}^a}\Big|_{(0,0)} \dot{q}^a + \frac{1}{2}\frac{\partial^2 \mathcal{L}}{\partial q^a q^b}\Big|_{(0,0)} q^a q^b + \frac{\partial^2 \mathcal{L}}{\partial q^a \dot{q}^b}\Big|_{(0,0)} q^a \dot{q}^b + \frac{1}{2}\frac{\partial^2 \mathcal{L}}{\partial \dot{q}^a \dot{q}^b}\Big|_{(0,0)} \dot{q}^a \dot{q}^b + \cdots$$

By the same arguments given in the previous problem, this reduces to

$$\mathcal{L} = \text{constant} + \frac{1}{2}\frac{\partial^2 \mathcal{L}}{\partial q^a q^b}\Big|_{(0,0)} q^a q^b + \frac{1}{2}\frac{\partial^2 \mathcal{L}}{\partial \dot{q}^a \dot{q}^b}\Big|_{(0,0)} \dot{q}^a \dot{q}^b + \cdots$$

With this form for $\mathcal{L}$, the Euler-Lagrange equations become

$$\frac{\partial \mathcal{L}}{\partial q^a \partial q^b}\Big|_{(0,0)} q^b = \frac{\partial \mathcal{L}}{\partial \dot{q}^a \partial \dot{q}^b}\Big|_{(0,0)} \ddot{q}^b.$$

As $\mathcal{L}$ is at a saddle point, we are unable—in this current form—to arrive at the result for a harmonic oscillator. However, recalling that K is really independent of q and V is independent of $\dot{q}$ (for conservative, and therefore all fundamental, forces), our equation reduces to

$$-\frac{\partial V}{\partial q^a \partial q^b}\Big|_{(0,0)} q^b = \frac{\partial K}{\partial \dot{q}^a \partial \dot{q}^b}\Big|_{(0,0)} \ddot{q}^b,$$

or, labeling the above matrices A and B,

$$-A_{ab} q^b = B_{ab} \ddot{q}^b.$$

Both of these matrices are necessarily positive-definite as a consequence of V and K attaining their minimal values at $(0,0)$. We may inverse B, then, to arrive at

$$\ddot{q}^c = -C^c{}_b q^b.$$

Since C is positive definite, we may now follow the argument given by Penrose to show that our system undergoes oscillations about its normal modes.

20.10. Desribe the system of eigenvectors in such degenerate cases.

In such cases, the necessary independent solutions for such modes are described by so-called generalized eigenvectors.

20.11. Prove this. (Recall from §13.7 that 'T' stands for 'transposed'.)

20.12. Describe this behaviour.

The basic equation of the harmonic oscillator, whereby ω^2 is an eigenvalue, is

$$\frac{\mathrm{d}^2\mathbf{q}}{\mathrm{d}t^2} + \omega^2\mathbf{q} = 0.$$

This has solutions of the form

$$\mathbf{q} = Ae^{-i\omega t} + Be^{i\omega t}.$$

When $\mathbf{W}$ is not a positive-definite matrix, some of its eigenvalues will be negative (or zero, in which case $\mathbf{q} = At + B$ trivially), producing solutions of the form

$$\mathbf{q} = Ae^{-\omega t} + Be^{\omega t}.$$

Such solutions diverge exponentially away from equilibrium.

20.13. Confirm that this expression for $\{\Phi, \Psi\}$ agrees with that of §14.8.

20.14. Show this.

20.15. Why?

Because

$$\{\mathcal{H}, \mathcal{H}\} = \frac{\partial \mathcal{H}}{\partial p_a}\frac{\partial \mathcal{H}}{\partial q^a} - \frac{\partial \mathcal{H}}{\partial q^a}\frac{\partial \mathcal{H}}{\partial p_a} = 0.$$

20.16. Explain why.

If the Lagrangian is free of some generalized 'position' coordinate q^r, then

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \mathcal{L}}{\partial \dot{q}^r} = \frac{\partial \mathcal{L}}{\partial q^r} = 0,$$

i.e. $p_r = \partial\mathcal{L}/\partial\dot{q}^r$ does not change with time.

20.17. Show this.

I believe Penrose is mistaken. By referring to the electromagnetic field tensor and its raised counterpart,

$$F_{ab} = \begin{pmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & -B_3 & B_2 \\ -E_2 & B_3 & 0 & -B_1 \\ -E_3 & -B_2 & B_1 & 0 \end{pmatrix} \qquad F^{ab} = \begin{pmatrix} 0 & -E_1 & -E_2 & -E_3 \\ E_1 & 0 & -B_3 & B_2 \\ E_2 & B_3 & 0 & -B_1 \\ E_3 & -B_2 & B_1 & 0 \end{pmatrix}$$

we can immediately write down

$$\begin{aligned} \frac{1}{4}F_{ab}F^{ab} &= \frac{1}{4}(-E_1^2 - E_2^2 - E_3^2 - E_1^2 + B_3^2 + B_2^2 - E_2^2 + B_3^2 + B_1^2 - E_3^2 + B_2^2 + B_1^2), \\ &= \frac{1}{2}(\mathbf{B}^2 - \mathbf{E}^2). \end{aligned}$$

In order for the Lagrangian to equal $(\mathbf{E}^2 - \mathbf{B}^2)/8$, it must be given by

$$\mathcal{L} = -\frac{1}{16} F_{\mu\nu} F^{\mu\nu}.$$

20.18. Show that this prescription is equivalent to that given in the main text.

21 The quantum particle

21.1. Show that $(1 + D^2)\cos x = 0$ and $(1 + D^2)\sin x = 0$ (referring to formulae in §6.5, if you need them).

Both cosine and sine are eigenfunctions of D^2 with eigenvaues of -1, and so

$$(1 + D^2)\cos x = (1 - 1)\cos x = 0,$$
$$(1 + D^2)\sin x = (1 - 1)\sin x = 0.$$

21.2. Taking note of Exercise [21.1], find the *general* solution of $(1 + D^2)y = x^5$, providing a proof that your solution is, in fact, the most general.

To find the most general solution of this equation, we must add to the particular solution the homogenous one, i.e. the set of solutions satisfying

$$(1 + D^2)y = 0.$$

This set is exactly that given in the solution to the above problem, yielding a general solution of

$$y = x^5 - 20x^3 + 120x + A\cos x + B\sin x.$$

21.3. See if you can explain why the procedure given in the text isses most of the solutions given in Exercise [21.2]. Can you suggest a modified general procedure which finds them all? *Hint*: To what extent does '$1 - D^2 + D^4 - D^6 + \dots$' really satisfy the requirements for an inverse to $1 + D^2$? Try acting on $(1 + D^2)\cos x$ with this infinite expression.

21.4. Why?

When $a \neq b$, we have $D_b x^a = x^a D_b$, i.e. x and D commute. This is patently obvious, though, as the partial derivative of an expression absent of the coordinate to differentiated is necessarily zero.

21.5. Solve this Schrodinger equation explicitly in the case of a particle of mass m in a constant Newtonian gravitational field: $V = mgz$. (Here z is the height above the Earth's surface and g is the downward gravitational acceleration.)

There is no reason for the solution to such an equation to vary in either the x or y coordinate, and so the Schrodinger equation with the relevant Hamiltonian collapses to

$$i\hbar \frac{\partial \psi}{\partial t} = -\frac{\hbar^2}{2m} \frac{\partial^2 \psi}{\partial z^2} + mgz\psi.$$

Let us suppose that our wavefunction is a product of two independent functions, $\psi(z;t) = \phi(z)\nu(t)$. Then our equation becomes

$$i\hbar \frac{1}{\nu} \frac{\mathrm{d}\nu}{\mathrm{d}t} = -\frac{\hbar^2}{2m} \frac{1}{\phi} \frac{\mathrm{d}^2 \phi}{\mathrm{d}z^2} + mgz.$$

Because ϕ and ν are independent, this equation can only be satisified if both sides equal the same, constant value E.

The time function ν is easily solved as $\nu = e^{-i\frac{E}{\hbar}t}$. Unfortunately, the solution of ϕ is not as nice. We are looking for a function ϕ that satisfies

$$-\frac{\hbar^2}{2m} \frac{\mathrm{d}^2 \phi}{\mathrm{d}z^2} + (mgz - E)\phi = 0,$$

or, equivalently,

$$\frac{\mathrm{d}^2 \phi}{\mathrm{d}z^2} - \frac{mgz - E}{\hbar^2/2m}\phi = 0.$$

Functions that satisfy such a relation are called Airy functions. If we assume ϕ goes to zero at infinity, ϕ is given by $\mathrm{Ai}(x)$, where such a function equals

$$\mathrm{Ai}(x) = \frac{1}{\pi} \int_0^\infty \cos\left(\frac{t^3}{3} + xt\right) \mathrm{d}t$$

and x is the variable multiplying ϕ in the rightmost term of the differential equation.

21.6. By transforming to the freely-falling frame with coordinates $X = x$, $Y = y$, $Z = z - \frac{1}{2}t^2 g$, $T = t$, show that the Schrodinger equation of Exericse [21.5] transforms to one without a gravitational field, with wavefunction $\Psi = e^{i(\frac{1}{6}mt^3g^2 + mtzg)}\psi$. What does this tell us about Einstein's principle of equivalence (see §17.4), as applied to quantum systems? (Take note of §21.9.)

21.7. Show that, if the quantum Hamiltonian $\mathcal{H}$ has a translation invariance, say being independent of the position variable x^3, then the corresponding momentum p_3 is conserved in the sense that the operator p_3 commutes with the time evolution $\partial/\partial t$. Explain, in the light fo the interpretations given later, why this commutation implies conservation.

From Schrodinger's equation, we know that

$$\hat{\mathcal{H}} = i\hbar \frac{\partial}{\partial t},$$

i.e. the Hamiltonian *is* the time-evolution operator. Simultaneously, we also know that

$$\hat{p}_a = -i\hbar \frac{\partial}{\partial q^a},$$

i.e. the operator for the ath momentum tells us how the quantity acted upon changes in the ath coordinate. But if $\mathcal{H}$ is independent of the ath coordinate, then $\hat{p}_a(\hat{\mathcal{H}}\psi) = \hat{\mathcal{H}}(\hat{p}_a\psi)$, i.e. p_a commutes with the Hamiltonian, or time-evolution, operator.

The commutativity of p_a and $\mathcal{H}$ allows us to equate both ordered products, and since $\hat{p}_a\mathcal{H} = 0$, so also does $\hat{\mathcal{H}}p_a = 0$. But this last equation says that the ath momentum does not change in time, i.e. it is conserved.

21.8. See if you can see why the requirements of special relativity enable Planck's $E = h\nu$ to be deduced from de Broglie's $p = h\lambda^{-1}$. (*Hint*: You may assume that the hyperplanes in $\mathbb{M}$ along which the wave takes a constant value are Lorentz-orthogonal to the particle's 4-velocity.)

From SR, we know that energy and momentum may be combined to form a four-vector

$$(E, \mathbf{p}) = (E, \hbar\mathbf{k}),$$

where we have used de Broglie's $p = h\lambda^{-1} = \hbar k$ in the second equality. We can then *define* the frequency as $E = \hbar\omega = h\nu$.

21.9. Why? Here linear dependence can involve continuous sums, namely integrals.

21.10. Why can I split it this way?

This is because

$$\begin{aligned} e^{-iP_a x^a/\hbar} &= e^{-i(Et - p_x x - p_y y - p_z z)/\hbar} \\ &= e^{-iEt/\hbar + i(p_x x + p_y y + p_z z)/\hbar} \\ &= e^{-iEt/\hbar + i\mathbf{P}\cdot\mathbf{x}/\hbar} \\ &= e^{-iEt/\hbar} e^{i\mathbf{P}\cdot\mathbf{x}/\hbar} \end{aligned}$$

21.11. Replacing the real number C in the above displayed expression by the complex number $C+iD$ (where C and D are real), find the frequency of the wave packet and the location of its peak.

Making the substitution and expanding yields

$$\begin{aligned} Ae^{-B^2(x-C)^2} &= Ae^{-B^2[x-(C+iD)]^2} \\ &= Ae^{-B^2[x^2-2x(C+iD)+(C+iD)^2]} \\ &= Ae^{-B^2[x^2-2xC-i2xD+C^2+i2CD-D^2]} \\ &= Ae^{-B^2(x^2-2xC+C^2-D^2)} e^{-i2B^2D(C-x)} \\ &= \left[Ae^{B^2D^2} e^{-B^2(x-C)^2}\right] e^{-i2B^2D(C-x)} \end{aligned}$$

We can immediately identify the frequency as $\omega = 2B^2D$. The peak occurs when the argument of the exponential dictating the amplitude of the wave is largest. This occurs at $x = C$.

21.12. Show that the probability of such double-spot appearances, according to such a picture, must be quite appreciable, whatever the law of probability of spot appearance in terms of wavefunction intensity might be. *Hint*: Divide the screen into two parts, with equal probability of spot appearance in each.

21.13. Check this from the hyperfunctional definition given in §9.7.

The definition of the Dirac delta function is

$$\delta(x) = \left(\!\!\left| \frac{1}{2\pi i z}, \frac{1}{2\pi i z} \right|\!\!\right).$$

By the properties of hyperfunctions,

$$x\delta(x) = \left(\!\!\left| \frac{x}{2\pi i z}, \frac{x}{2\pi i z} \right|\!\!\right),$$

which, when evaluated at 0, becomes

$$0 \cdot \delta(0) = \left(\!\!\left| 0, 0 \right|\!\!\right) = 0$$

21.14. Show that replacing ψ by $x^1\psi$ or by $i\hbar\partial\psi/\partial x^1$ corresponds, respectively, to replacing $\tilde{\psi}$ by $-i\hbar\partial\tilde{\psi}/\partial p_1$ or by $p_1\tilde{\psi}$. Show that replacing $\psi(x^a)$ by $\psi(x^a + C^a)$ corresponds to replacing ψ by $e^{-iC^a p_a/\hbar}\tilde{\psi}$ (where a ranges over 1, 2, 3).

These correspondences are easily seen when either representation is expressed as the Fourier transform of the other. For example,

$$\begin{aligned}
i\hbar\frac{\partial\tilde{\psi}}{\partial p_1} &= i\hbar\frac{\partial}{\partial p_1}\Big((2\pi)^{-3/2}\int_{\mathbb{E}^3}\psi(\mathbf{X})e^{-i\mathbf{p}\cdot\mathbf{X}/\hbar}\mathrm{d}^3\mathbf{X}\Big),\\
&= i\hbar(2\pi)^{-3/2}\int_{\mathbb{E}^3}\psi(\mathbf{X})\frac{\partial}{\partial p_1}\Big(e^{-i(p_1x^1+p_2x^2+p_3x^3)/\hbar}\Big)\mathrm{d}^3\mathbf{X},\\
&= i\hbar(2\pi)^{-3/2}\int_{\mathbb{E}^3}\psi(\mathbf{X})\Big(\frac{-ix^1}{\hbar}\Big)e^{-i(p_1x^1+p_2x^2+p_3x^3)/\hbar}\mathrm{d}^3\mathbf{X},\\
&= (2\pi)^{-3/2}\int_{\mathbb{E}^3}\Big(x^1\psi(\mathbf{X})\Big)e^{-i\mathbf{p}\cdot\mathbf{X}/\hbar}\mathrm{d}^3\mathbf{X},
\end{aligned}$$

which shows the identification

$$i\hbar\frac{\partial\tilde{\psi}}{\partial p_1} \leftrightarrow x^1\psi.$$

Meanwhile, we have

$$\begin{aligned}
-i\hbar\frac{\partial\psi}{\partial x^1} &= -i\hbar\frac{\partial}{\partial x^1}\Big((2\pi)^{-3/2}\int_{\mathbb{E}^3}\tilde{\psi}(\mathbf{p})e^{i\mathbf{p}\cdot\mathbf{X}/\hbar}\mathrm{d}^3\mathbf{p}\Big),\\
&= -i\hbar(2\pi)^{-3/2}\int_{\mathbb{E}^3}\tilde{\psi}(\mathbf{p})\frac{\partial}{\partial x^1}\Big(e^{i(p_1x^1+p_2x^2+p_3x^3)/\hbar}\Big)\mathrm{d}^3\mathbf{p},\\
&= -i\hbar(2\pi)^{-3/2}\int_{\mathbb{E}^3}\tilde{\psi}(\mathbf{p})\Big(\frac{ip_1}{\hbar}\Big)e^{i(p_1x^1+p_2x^2+p_3x^3)/\hbar}\mathrm{d}^3\mathbf{p},\\
&= (2\pi)^{-3/2}\int_{\mathbb{E}^3}\Big(p_1\tilde{\psi}(\mathbf{p})\Big)e^{i\mathbf{p}\cdot\mathbf{X}/\hbar}\mathrm{d}^3\mathbf{p},
\end{aligned}$$

which shows

$$-i\hbar\frac{\partial\psi}{\partial x^1} \leftrightarrow p_1\tilde{\psi}.$$

In both cases, I believe Penrose has mistakenly used the wrong sign.

For the last case, observe that

$$\begin{aligned}
e^{iC^a p_a/\hbar}\tilde{\psi} &= e^{iC^a p_a/\hbar}\Big((2\pi)^{-3/2}\int_{\mathbb{E}^3}\psi(x^a)e^{-ip_a x^a/\hbar}\mathrm{d}^3\mathbf{X}\Big),\\
&= (2\pi)^{-3/2}\int_{\mathbb{E}^3}\psi(x^a)e^{-ip_a x^a/\hbar}e^{ip_a C^a/\hbar}\mathrm{d}^3\mathbf{X},\\
&= (2\pi)^{-3/2}\int_{\mathbb{E}^3}\psi(x^a)e^{-ip_a (x^a - C^a)/\hbar}\mathrm{d}^3\mathbf{X},\\
&= (2\pi)^{-3/2}\int_{\mathbb{E}^3}\psi(\bar{x}^a + C^a)e^{-ip_a \bar{x}^a/\hbar}\mathrm{d}^3\bar{\mathbf{X}},
\end{aligned}$$

where we made the substitution $\bar{x}^a = x^a - C^a$ in the last line. This shows

$$e^{-iC^a p_a/\hbar}\tilde{\psi} \leftrightarrow \psi(x^a + C^a).$$

In all of the above cases, I believe Penrose has mistakenly used the wrong sign in each of the transforms' exponential arguments.

21.15. Use the results of Exercises [21.11], [21.13], and [21.14] to show that the Fourier transform of the wave packet $\psi = Ae^{-B^2(x-C)^2}e^{i\omega x}$ is $\tilde{\psi} = (Ae^{i\omega C}/B\sqrt{2})e^{-(p-\omega)^2/4B^2}e^{-iCp}$ (putting $\hbar = 1$ for convenience.)

21.16. Can you see how to justify the factor $-\frac{1}{2}\hbar\log 2$? (The half-life is the time at which the probability of decay has reached one half.)

22 Quantum algebra, geometry, and spin

22.1. Make it clear why the action of any Schrodinger evolution is linear, despite the fact that $\mathcal{H}$ may be a highly non-linear function of the ps and qs.

Given that the operator representing a Schrodinger evolution is identified with the partial derivative operator, its linearity is simply a consequence of the linearity of partial derivatives.

22.2. See if you can explain why the $\langle\phi|\psi\rangle$ integral converges whenever both $\langle\phi|\phi\rangle$ and $\langle\psi|\psi\rangle$ converge. *Hint*: Consider what is implied by the integral of $|\phi - \lambda\psi|^2$ being non-negative over any *finite* region of $\mathbb{E}^3$, deriving an inequality connecting the square modulus of the integral of $\bar{\phi}\psi$ with the product of the integral of $\bar{\phi}\phi$ with the integral of $\bar{\psi}\psi$. As an intermediate step, find conditions on complex numbers a, b, c, d that imply $a + \lambda b + \bar{\lambda}c + \bar{\lambda}\lambda d \geq 0$ for all λ.

The non-negativity of the integral of $|\phi - \lambda\psi|^2$ over any finite region Ω of $\mathbb{E}^3$ implies

$$\int_\Omega \bar{\phi}\phi\,\mathrm{d}^3x - \lambda\int_\Omega \bar{\phi}\psi\,\mathrm{d}^3x - \bar{\lambda}\int_\Omega \bar{\psi}\phi\,\mathrm{d}^3x + \lambda\bar{\lambda}\int_\Omega \bar{\psi}\psi\,\mathrm{d}^3x \geq 0.$$

We can rewrite this as

$$\int_\Omega \bar{\phi}\phi\,\mathrm{d}^3x - 2\mathrm{Re}\Big(\lambda\int_\Omega \bar{\phi}\psi\,\mathrm{d}^3x\Big) + |\lambda|^2\int_\Omega \bar{\psi}\psi\,\mathrm{d}^3x \geq 0.$$

If we express the middle term in polar form (with angle θ), we find

$$\begin{aligned}-2\mathrm{Re}\Big(\Big|\lambda\int_\Omega \bar{\phi}\psi\,\mathrm{d}^3x\Big|e^{i\theta}\Big), &= -2\Big|\lambda\int_\Omega \bar{\phi}\psi\,\mathrm{d}^3x\Big|\cos\theta \\ &\geq -2\Big|\lambda\int_\Omega \bar{\phi}\psi\,\mathrm{d}^3x\Big|,\end{aligned}$$

and hence

$$\int_\Omega \bar{\phi}\phi\,\mathrm{d}^3x - 2|\lambda|\Big|\int_\Omega \bar{\phi}\psi\,\mathrm{d}^3x\Big| + |\lambda|^2\int_\Omega \bar{\psi}\psi\,\mathrm{d}^3x \geq 0.$$

This is a quadratic equation in $|\lambda|$. In order for the inequality to hold true, the parabola formed must never dip below the x-axis, i.e. the discriminant of the above equation must be less than or equal to 0. In other words, we have

$$\Big|\int_\Omega \bar{\phi}\psi\,\mathrm{d}^3x\Big|^2 \leq \Big(\int_\Omega \bar{\phi}\phi\,\mathrm{d}^3x\Big)\Big(\int_\Omega \bar{\psi}\psi\,\mathrm{d}^3x\Big)$$

Since both integrals on the right are convergent, we may take the limit as Ω goes to $\mathbb{E}^3$ to show that $\langle\phi|\psi\rangle$ converges,

$$\Big|\int_{\mathbb{E}^3} \bar{\phi}\psi\,\mathrm{d}^3x\Big| \leq \Big|\int_{\mathbb{E}^3} \bar{\phi}\psi\,\mathrm{d}^3x\Big|^2 \leq \Big(\int_{\mathbb{E}^3} \bar{\phi}\phi\,\mathrm{d}^3x\Big)\Big(\int_{\mathbb{E}^3} \bar{\psi}\psi\,\mathrm{d}^3x\Big)$$

22.3. Following on from Exercise [22.2], show that the normalizable wavefunctions indeed constitute a vector space.

Take the standard function space definitions of addition and multiplication as pointwise operations. Clearly, all 'standard properties,' such as associativity and commutativity, are immediately satisfied. We need only check whether the space is closed. If two normalizable (but not necessarily normalized) wave functions ϕ and ψ are added together to produce σ, we may normalize the latter by dividing by $\sqrt{D}$, where

$$\begin{aligned}\|\sigma\| &= \int \bar{\sigma}\sigma\,\mathrm{d}^3x \\ &= \int (\bar{\phi}+\bar{\psi})(\phi+\psi)\,\mathrm{d}^3x \\ &= \int \bar{\phi}\phi + \bar{\phi}\psi + \bar{\psi}\phi + \bar{\psi}\psi\,\mathrm{d}^3x \\ &= \int \bar{\phi}\phi\,\mathrm{d}^3x + \int \bar{\phi}\psi\,\mathrm{d}^3x + \int \bar{\psi}\phi\,\mathrm{d}^3x + \int \bar{\psi}\psi\,\mathrm{d}^3x \\ &= A + B + \bar{B} + C \\ &= D\end{aligned}$$

Here, the finite nature of A and C was given by the normalizability of ϕ and ψ. Meanwhile, the finite nature of B and $\bar{B}$ was shown in the previous exercise.

Closure under scalar multiplication is also simple to address. If $\|\phi\| = A$, then, by the linearity of integration, $\|s\phi\| = |s|^2A$, and hence ϕ can be normalized by dividing by $s\sqrt{A}$.

22.4. Verify this, stating carefully which properties of integration are being used.

Going down the list, we see

$$\begin{aligned}
\langle\phi|\psi+\chi\rangle &= \int \bar{\phi}(\psi+\chi)\,\mathrm{d}^3x \\
&= \int \bar{\phi}\psi\,\mathrm{d}^3x + \int \bar{\phi}\chi\,\mathrm{d}^3x \\
&= \langle\phi|\psi\rangle + \langle\phi|\chi\rangle \\
\langle\phi|a\psi\rangle &= \int \bar{\phi}a\psi\,\mathrm{d}^3x \\
&= a\int \bar{\phi}\psi\,\mathrm{d}^3x \\
&= a\langle\phi|\psi\rangle \\
\langle\phi|\psi\rangle &= \int \bar{\phi}\psi\mathrm{d}^3x \\
&= \overline{\int \phi\bar{\psi}\,\mathrm{d}^3x} \\
&= \overline{\int \bar{\psi}\phi\,\mathrm{d}^3x} \\
&= \overline{\langle\psi|\phi\rangle}
\end{aligned}$$

where we have used the properties superposition, homogeneity, and commutativity with complex conjugation, respectively.

Finally, if $\psi \neq 0$, then

$$\begin{aligned}
\langle\psi|\psi\rangle &= \int \bar{\psi}\psi\,\mathrm{d}^3x \\
&= \int |\psi|^2\,\mathrm{d}^3x \\
&\geq 0
\end{aligned}$$

then

22.5. Show why.

By the third listed property, we see

$$\begin{aligned}
\langle\phi+\chi|\psi\rangle &= \overline{\langle\psi|\phi+\chi\rangle} \\
&= \overline{\langle\psi|\phi\rangle + \langle\psi|\chi\rangle} \\
&= \overline{\langle\psi|\phi\rangle} + \overline{\langle\psi|\chi\rangle} \\
&= \langle\phi|\psi\rangle + \langle\chi|\psi\rangle
\end{aligned}$$

and

$$
\begin{aligned}
\langle a\phi|\psi\rangle &= \overline{\langle\psi|a\phi\rangle} \\
&= \overline{a\langle\psi|\phi\rangle} \\
&= \bar{a}\langle\phi|\psi\rangle
\end{aligned}
$$

22.6. Show how $\langle\phi|\psi\rangle$ can be defined from the norm. *Hint*: Work out the norms of $\phi+\psi$ and $\phi+i\psi$.

Doing as Penrose suggests, we find

$$
\begin{aligned}
\|\phi+\psi\| &= \langle\phi+\psi|\phi+\psi\rangle \\
&= \langle\phi|\phi\rangle + \langle\phi|\psi\rangle + \langle\psi|\phi\rangle + \langle\psi|\psi\rangle \\
&= \|\phi\| + 2\mathrm{Re}(\langle\phi|\psi\rangle) + \|\psi\| \\
\|\phi+i\psi\| &= \langle\phi+i\psi|\phi+i\psi\rangle \\
&= \langle\phi|\phi\rangle + i\langle\phi|\psi\rangle - i\langle\psi|\phi\rangle + \langle\psi|\psi\rangle \\
&= \|\phi\| - 2\mathrm{Im}(\langle\phi|\psi\rangle) + \|\psi\|
\end{aligned}
$$

Therefore, we may define $\langle\phi|\psi\rangle$ as

$$
\langle\phi|\psi\rangle = \frac{1}{2}\Big(\|\phi+\psi\| - i\|\phi+i\psi\|\Big)
$$

22.7. Spell this argument out a little more fully. Can you explain why we should expect the Leibniz property to hold for a Hilbert-space scalar product?

Because the Hilbert-space scalar product of ϕ and ψ is defined in terms of the convergent integral

$$
\langle\phi|\psi\rangle = \int \bar{\phi}\psi \mathrm{d}^3x,
$$

we may use the Leibniz integral rule to find

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\langle\phi|\psi\rangle &= \frac{\mathrm{d}}{\mathrm{d}t}\int \bar{\phi}\psi \mathrm{d}^3x \\
&= \int \frac{\mathrm{d}}{\mathrm{d}t}\Big(\bar{\phi}\psi\Big)\mathrm{d}^3x \\
&= \int \frac{\mathrm{d}\bar{\phi}}{\mathrm{d}t}\psi + \bar{\phi}\frac{\mathrm{d}\psi}{\mathrm{d}t}\mathrm{d}^3x \\
&= \int \overline{\frac{\mathrm{d}\phi}{\mathrm{d}t}}\psi \mathrm{d}^3x + \int \bar{\phi}\frac{\mathrm{d}\psi}{\mathrm{d}t}\mathrm{d}^3x \\
&= \Big\langle \frac{\mathrm{d}}{\mathrm{d}t}\phi\Big|\psi\Big\rangle + \Big\langle\phi\Big|\frac{\mathrm{d}}{\mathrm{d}t}\psi\Big\rangle
\end{aligned}
$$

In the above, it was necessary to have the integral be convergent to take the derivative under the integral sign. We have kept the derivative as an ordinary derivative through the assumption that the coordinate parameters of ϕ and ψ do not depend on time.

22.8. Explain all this in detail.

To prove that observables maintain their eigenvalues, we examine the characteristic equation of an arbitrary, time-evolved observable,

$$\begin{aligned}\det(\mathbf{Q}_{\mathrm{H}} - \lambda\mathbf{I}) &= \det(\mathbf{U}_t^*\mathbf{Q}\mathbf{U}_t - \lambda\mathbf{I}) \\ &= \det(\mathbf{U}_t^*\mathbf{Q}\mathbf{U}_t - \lambda\mathbf{U}_t^*\mathbf{I}\mathbf{U}_t) \\ &= \det(\mathbf{U}_t^*[\mathbf{Q} - \lambda\mathbf{I}]\mathbf{U}_t) \\ &= \det(\mathbf{U}_t^*)\det(\mathbf{Q} - \lambda\mathbf{I})\det(\mathbf{U}_t) \\ &= \det(\mathbf{Q} - \lambda\mathbf{I})\end{aligned}$$

Since this is the same as the original observable's characteristic equation, the two share the same set of eigenvalues.

Meanwhile, all scalar products are held because

$$\begin{aligned}\langle\phi|_H\mathbf{Q}_H|\psi\rangle_H &= \langle\phi|\mathbf{U}_t\mathbf{U}_t^*\mathbf{Q}\mathbf{U}_t\mathbf{U}_t^*|\psi\rangle \\ &= \langle\phi|\mathbf{Q}|\psi\rangle\end{aligned}$$

22.9. See if you can confirm this.

We can easily verify this by first recalling that

$$\mathbf{U}_t = e^{\frac{i\mathcal{H}t}{\hbar}}.$$

Using the product rule gives

$$\begin{aligned}i\hbar\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{Q}_{\mathrm{H}} &= i\hbar\frac{\mathrm{d}}{\mathrm{d}t}\Big(\mathbf{U}_t^*\mathbf{Q}\mathbf{U}_t\Big) \\ &= i\hbar\frac{\mathrm{d}}{\mathrm{d}t}\Big(\mathbf{U}_t^*\Big)\mathbf{Q}\mathbf{U}_t + i\hbar\mathbf{U}_t^*\mathbf{Q}\frac{\mathrm{d}}{\mathrm{d}t}\Big(\mathbf{U}_t\Big) \\ &= i\hbar\mathbf{U}_t^*\Big(-\frac{i\mathcal{H}}{\hbar}\Big)\mathbf{Q}\mathbf{U}_t + i\hbar\mathbf{U}_t^*\mathbf{Q}\mathbf{U}_t\Big(\frac{i\mathcal{H}}{\hbar}\Big) \\ &= \mathcal{H}\mathbf{U}_t^*\mathbf{Q}\mathbf{U}_t - \mathbf{U}_t^*\mathbf{Q}\mathbf{U}_t\mathcal{H} \\ &= \mathcal{H}\mathbf{Q}_{\mathrm{H}} - \mathbf{Q}_{\mathrm{H}}\mathcal{H} \\ &= [\mathcal{H}, \mathbf{Q}_{\mathrm{H}}]\end{aligned}$$

In the above, we have made use of the fact that $\mathcal{H}$ commutes with itself, and hence with the matrix exponential of itself (the time-evolution operator).

22.10. Show that any eigenvalue of a Hermitian operator $\mathbf{Q}$ is indeed a *real* number.

Consider the action of $\mathbf{Q}$ on one of its normalized eigenvectors $|\lambda_i\rangle$. We know that

$$\begin{aligned}\mathbf{Q}|\lambda_i\rangle &= \lambda_i|\lambda_i\rangle \\ \langle\lambda_j|\mathbf{Q}^* &= \langle\lambda_j|\bar{\lambda}_j\end{aligned}$$

Taking the scalar product of both sides of the above yields

$$\langle\lambda_j|\mathbf{Q}^*\mathbf{Q}|\lambda_i\rangle = \bar{\lambda}_j\lambda_i\langle\lambda_j|\lambda_i\rangle$$

But, since $\mathbf{Q}^* = \mathbf{Q}$, we can apply $\mathbf{Q}^2$ to $|\lambda\rangle$ to obtain

$$\langle\lambda_j|\mathbf{Q}^*\mathbf{Q}|\lambda_i\rangle = \lambda_i^2\langle\lambda_j|\lambda_i\rangle$$

Subtracting one from the other and examining the case when $i = j$ gives

$$|\lambda_i|^2 = \lambda_i^2$$

This can only be true if $\lambda_i \in \mathbb{R}$.

22.11. See if you can prove this. *Hint*: By considering the expression $\langle\psi|(\mathbf{Q}^* - \bar{\lambda}\mathbf{I})(\mathbf{Q} - \lambda\mathbf{I})|\psi\rangle$, show first that if $\mathbf{Q}|\psi\rangle = \lambda|\psi\rangle$, then $\mathbf{Q}^*|\psi\rangle = \bar{\lambda}|\psi\rangle$.

Expanding out the above expression gives

$$\langle\psi|(\mathbf{Q}^* - \bar{\lambda}\mathbf{I})(\mathbf{Q} - \lambda\mathbf{I})|\psi\rangle = \langle\psi|\mathbf{Q}^*\mathbf{Q}|\psi\rangle - \lambda\langle\psi|\mathbf{Q}^*|\psi\rangle - \bar{\lambda}\langle\psi|\mathbf{Q}|\psi\rangle + |\lambda|^2\langle\psi|\psi\rangle$$

If λ is an eigenvalue of $\mathbf{Q}$, the lefthand side evaluates to zero. Meanwhile, the righthand side becomes

$$0 = \langle\psi|\mathbf{Q}\mathbf{Q}^*|\psi\rangle - \lambda\langle\psi|\mathbf{Q}^*|\psi\rangle - |\lambda|^2\langle\psi|\psi\rangle + |\lambda|^2\langle\psi|\psi\rangle = \langle\psi|\mathbf{Q}\mathbf{Q}^*|\psi\rangle - \lambda\langle\psi|\mathbf{Q}^*|\psi\rangle,$$

where we have made use of the fact that $\mathbf{Q}$ is normal in switching the order of operations of $\mathbf{Q}$ and $\mathbf{Q}^*$. Since

$$\mathbf{Q}^*|\psi\rangle = \omega|\psi\rangle \implies \langle\psi|\mathbf{Q} = \bar{\omega}\langle\psi|,$$

the previous equation implies

$$|\omega|^2\langle\psi|\psi\rangle = \lambda\omega\langle\psi|\psi\rangle$$

or simply

$$\bar{\omega}\omega = \lambda\omega$$

That is, $\omega = \bar{\lambda}$, and so

$$\mathbf{Q}|\psi\rangle = \lambda|\psi\rangle \implies \mathbf{Q}^*|\psi\rangle = \bar{\lambda}|\psi\rangle$$

Now consider the scalar product of

$$\langle\lambda_j|\mathbf{Q}\mathbf{Q}|\lambda_i\rangle$$

On the one hand, this is equivalent to operating on $|\lambda_i\rangle$ with $\mathbf{Q}^2$, which yields

$$\langle\lambda_j|\mathbf{Q}\mathbf{Q}|\lambda_i\rangle = \lambda_i^2\langle\lambda_j|\lambda_i\rangle$$

On the other, using $\langle\lambda_j|\mathbf{Q} = \lambda_j\langle\lambda_j|$ gives

$$\langle\lambda_j|\mathbf{Q}\mathbf{Q}|\lambda_i\rangle = \lambda_j\lambda_i\langle\lambda_j|\lambda_i\rangle.$$

Subtracting one from the other yields the condition

$$\lambda_i(\lambda_j - \lambda_i)\langle\lambda_j|\lambda_i\rangle = 0.$$

If all of the eigenvalues are distinct and nonzero, the only way for such a condition to hold when $i \neq j$ is when $\langle\lambda_j|\lambda_i\rangle = 0$, i.e. when the eigenvectors of $\mathbf{Q}$ are mutually orthogonal.

22.12. Show this, from the algebraic properties of $\langle\ |\ \rangle$ by methods used in Exercise [22.2].

In 22.2, we showed that
$$\langle\phi|\psi\rangle\langle\psi|\phi\rangle \le \langle\phi|\phi\rangle\langle\psi|\psi\rangle$$
from which it immediately follows that
$$\frac{\langle\phi|\psi\rangle\langle\psi|\phi\rangle}{\langle\phi|\phi\rangle\langle\psi|\psi\rangle} \le 1$$
If $\phi = C\psi$, we this becomes
$$\frac{\langle\phi|\psi\rangle\langle\psi|\phi\rangle}{\langle\phi|\phi\rangle\langle\psi|\psi\rangle} = \frac{\bar{C}\langle\psi|\psi\rangle C\langle\psi|\psi\rangle}{\bar{C}C\langle\psi|\psi\rangle\langle\psi|\psi\rangle} = 1.$$

22.13. Show that if an observable $\mathbf{Q}$ satisfies some polynomial equation, then every one of its eigenvalues satisfies the same equation.

Suppose we have
$$a_n\mathbf{Q}^n + a_{n-1}\mathbf{Q}^{n-1} + \cdots + a_1\mathbf{Q} + a_0 = 0.$$
Acting on this with $|q\rangle$, an eigenvector of $\mathbf{Q}$, reveals
$$(a_nq^n + a_{n-1}q^{n-1} + \cdots + a_1q + a_0)|q\rangle = 0.$$
Since $|q\rangle$ is not the zero vector, we must have
$$a_nq^n + a_{n-1}q^{n-1} + \cdots + a_1q + a_0 = 0$$

22.14. Show this.

Using $\mathbf{E}^* = \mathbf{E}$ and $\mathbf{E}^2 = \mathbf{E}$, we see
$$\begin{aligned}\langle\psi|\mathbf{E}^*(\mathbf{I}-\mathbf{E})|\psi\rangle &= \langle\psi|\mathbf{E}(\mathbf{I}-\mathbf{E})|\psi\rangle \\ &= \langle\psi|\mathbf{E}-\mathbf{E}^2|\psi\rangle \\ &= \langle\psi|\mathbf{E}-\mathbf{E}|\psi\rangle \\ &= \langle\psi|\mathbf{0}|\psi\rangle \\ &= 0\end{aligned}$$
i.e. $\mathbf{E}|\psi\rangle$ and $(\mathbf{I}-\mathbf{E})|\psi\rangle$ are orthogonal.

22.15. Why?

By the Pythagorean theorem, we have
$$\langle\psi|\psi\rangle = \langle\psi|(\mathbf{I}-\mathbf{E})^*(\mathbf{I}-\mathbf{E})|\psi\rangle + \langle\psi|\mathbf{E}^*\mathbf{E}|\psi\rangle$$
Therefore, probability of either projection (onto $\mathbf{E}$ or orthogonal to it) is described by the amount the norm of $|\psi\rangle$ is reduced in such a space.

22.16. Can you see a simple reason for this?

If we do as Penrose suggests and think of a photon spinning about its direction of motion, we can see that a 180° change in direction that leaves its direction of spin unaltered must necessarily flip the polarization of the photon.

22.17. Explain more fully why the correct answer is given by 'projection'.

As Penrose points out, the measuring device can only tell us yes or no. Because of this, we might expect to find a measurement of either one would simply indicate that the particle is within the associated eigenspace, but it is more precise than this. The original state has an effect on the final. Following Penrose's example, where $|\rho+\rangle$ and $|\rho-\rangle$ are in the no eigenspace and $|\tau+\rangle$ and $\tau-\rangle$ span the yes eigenspace, we find a measurement of makes the following mapping of states

$$
\begin{aligned}
|\tau+\rangle + |\rho+\rangle &\to |\rho+\rangle \\
|\tau+\rangle + |\rho-\rangle &\to |\rho-\rangle \\
|\tau-\rangle + |\rho+\rangle &\to |\rho+\rangle \\
|\tau-\rangle + |\rho-\rangle &\to |\rho-\rangle
\end{aligned}
$$

This is most appropriately captured by projection.

22.18. Use quaternions to check this.

22.19. Check this. Explain how their multiplication rules relate to those of quaternions.

We first compute all necessary products,

$$
\begin{aligned}
\mathbf{L}_1\mathbf{L}_2 &= \frac{\hbar^2}{4}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \frac{\hbar^2}{4}\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} = \frac{i\hbar}{2}\mathbf{L}_3 \\
\mathbf{L}_2\mathbf{L}_1 &= \frac{\hbar^2}{4}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{\hbar^2}{4}\begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} = -\frac{i\hbar}{2}\mathbf{L}_3 \\
\mathbf{L}_2\mathbf{L}_3 &= \frac{\hbar^2}{4}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \frac{\hbar^2}{4}\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix} = \frac{i\hbar}{2}\mathbf{L}_1 \\
\mathbf{L}_3\mathbf{L}_2 &= \frac{\hbar^2}{4}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} = \frac{\hbar^2}{4}\begin{pmatrix} 0 & -i \\ -i & 0 \end{pmatrix} = -\frac{i\hbar}{2}\mathbf{L}_1 \\
\mathbf{L}_3\mathbf{L}_1 &= \frac{\hbar^2}{4}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \frac{\hbar^2}{4}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \frac{i\hbar}{2}\mathbf{L}_2 \\
\mathbf{L}_3\mathbf{L}_1 &= \frac{\hbar^2}{4}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \frac{\hbar^2}{4}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = -\frac{i\hbar}{2}\mathbf{L}_2
\end{aligned}
$$

From the above, we can immediately see

$$
\begin{aligned}
\mathbf{L}_1\mathbf{L}_2 - \mathbf{L}_2\mathbf{L}_1 &= i\frac{\hbar}{2}\mathbf{L}_3 + i\frac{\hbar}{2}\mathbf{L}_3 = i\hbar\mathbf{L}_3 \\
\mathbf{L}_2\mathbf{L}_3 - \mathbf{L}_3\mathbf{L}_2 &= i\frac{\hbar}{2}\mathbf{L}_1 + i\frac{\hbar}{2}\mathbf{L}_1 = i\hbar\mathbf{L}_1 \\
\mathbf{L}_3\mathbf{L}_1 - \mathbf{L}_1\mathbf{L}_3 &= i\frac{\hbar}{2}\mathbf{L}_2 + i\frac{\hbar}{2}\mathbf{L}_2 = i\hbar\mathbf{L}_2
\end{aligned}
$$

From the products we computed at the beginning of this exercise, we see

$$\mathbf{L}_1\mathbf{L}_2\mathbf{L}_3 = \frac{i\hbar}{2}\mathbf{L}_3^2 = \frac{i\hbar^3}{8}\mathbf{I}$$

If we drop the factors of $\hbar/2$ from these matrices, there is a clear correspondence between their products and the multiplication rules for quaternions. The primary difference is the additional factor of i in the Pauli products.

22.20. Do this explicitly.

Let's pick (arbitrarily) the generator corresponding to the first Pauli matrix,

$$\mathcal{L}_1 = -\frac{i}{\hbar}\mathbf{L}_1 = -\frac{i}{2}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Noting that $\mathcal{L}_1^2 = -\mathbf{I}/4$, we can exponentiate this through an angle θ to find

$$\begin{aligned}
\exp(\theta\mathcal{L}_1) &= \mathbf{I} + \theta\mathcal{L}_1 + \frac{1}{2!}\theta^2\mathcal{L}_1^2 + \frac{1}{3!}\theta^3\mathcal{L}_1^3 + \frac{1}{4!}\theta^4\mathcal{L}_1^4 + \cdots \\
&= \left[\mathbf{I} - \frac{1}{2!}\frac{\theta^2}{4}\mathbf{I} + \frac{1}{4!}\frac{\theta^4}{16}\mathbf{I} + \cdots\right] + \left[\theta\mathcal{L}_1 - \frac{1}{3!}\frac{\theta^3}{4}\mathcal{L}_1 + \cdots\right] \\
&= \left[1 - \frac{1}{2!}\left(\frac{\theta}{2}\right)^2 + \frac{1}{4!}\left(\frac{\theta}{2}\right)^4 + \cdots\right]\mathbf{I} + \left[\frac{\theta}{2} - \frac{1}{3!}\left(\frac{\theta}{2}\right)^3 + \cdots\right]\mathcal{L}_1 \\
&= \cos\left(\frac{\theta}{2}\right)\mathbf{I} + \sin\left(\frac{\theta}{2}\right)\mathcal{L}_1
\end{aligned}$$

If $\theta = 2\pi$, this becomes

$$\exp(2\pi\mathcal{L}_1) = \cos(\pi)\mathbf{I} + \sin(\pi)\mathcal{L}_1 = -\mathbf{I}.$$

Since $\mathcal{L}_i^2 = -\mathbf{I}/4$ holds for each generator corresponding to a Pauli matrix, we would have arrived at the same result if we were to choose $\mathcal{L}_2$ or $\mathcal{L}_3$ instead of $\mathcal{L}_1$.