

广义相对论

齐天博

tianboqi.github.io

目录

第一部分 数学基础	1
1 线性代数回顾	1
1.1 线性空间与对偶空间	1
1.1.1 线性空间与矢量	1
1.1.2 对偶空间与对偶矢量	1
1.2 张量	3
1.2.1 张量的定义	3
1.2.2 张量的运算	3
1.2.3 对称与反对称张量	5
1.3 度规	5
1.3.1 度规张量	5
1.3.2 指标升降	7
2 微分几何基础	9
2.1 流形和张量场	9
2.1.1 微分流形	9
2.1.2 流形的切空间和与余切空间	9
2.1.3 张量与张量场	11
2.1.4 张量密度*	12
2.2 流形上的度规、联络和曲率	12
2.2.1 流形上的度规	13
2.2.2 Levi-Civita 联络	15
2.2.3 测地线	18
2.2.4 黎曼曲率张量	20
2.3 李导数与 Killing 矢量场*	23
2.3.1 拉回与推前*	23
2.3.2 李导数*	25
2.3.3 Killing 矢量场*	26
第二部分 狭义相对论	29

3 洛伦兹变换	29
3.1 从伽利略变换到洛伦兹变换	29
3.1.1 伽利略变换	29
3.1.2 洛伦兹坐标变换	30
3.1.3 洛伦兹速度变换	31
3.1.4 时间的相对性	31
3.2 从四维时空看狭义相对论	32
3.2.1 闵氏时空	32
3.2.2 观测者与坐标系	33
3.2.3 从四维看洛伦兹变换	34
3.3 狭义相对论效应分析	37
3.3.1 时间与空间的相对性	37
3.3.2 尺缩效应与车库佯谬	38
3.3.3 钟慢效应与双生子佯谬	40
4 狭义相对论动力学	42
4.1 质点动力学	42
4.1.1 质能关系	42
4.1.2 4-矢量	43
4.1.3 瞬时观测者与标架	45
4.2 能量-动量张量	45
4.2.1 能量-动量张量的定义	45
4.2.2 能量-动量张量的性质	47
第三部分 广义相对论	49
5 广义相对论基础	49
5.1 弯曲时空	49
5.1.1 从平直时空到弯曲时空	49
5.1.2 弯曲时空中的物理定律	49
5.2 广义协变、等效原理与坐标系	50
5.2.1 广义协变与等效原理	50
5.2.2 弯曲时空中的坐标系	52
5.2.3 测地偏离方程	53
5.3 引力的几何描述	54
5.3.1 牛顿引力几何化	54
5.3.2 爱因斯坦场方程	55
5.3.3 爱因斯坦场方程的几点说明	56
5.3.4 真空场方程	56
6 线性近似与引力波	58
6.1 弱场近似	58
6.1.1 线性引力论	58

6.1.2	线性引力论的规范变换	59
6.1.3	牛顿近似	59
6.2	引力波	60
6.2.1	线性场方程的波动解	60
6.2.2	横向无迹规范	61
6.2.3	引力波的物理效应	61
7	史瓦西时空	63
7.1	史瓦西度规	63
7.1.1	稳态度规与静态度规	63
7.1.2	静态球对称度规	64
7.1.3	史瓦西外部解	65
7.2	史瓦西时空的物理定律	66
7.2.1	与牛顿引力论的对比	66
7.2.2	引力钟慢与引力红移	67
7.2.3	史瓦西时空的对称性与守恒量	68
7.2.4	史瓦西时空的测地线	69
8	黑洞	74
8.1	静态球对称黑洞	74
8.1.1	史瓦西时空的奇性	74
8.1.2	事件视界及其内部的时空结构	75
8.1.3	Kruskal 最大延拓	77
8.1.4	球对称恒星的坍缩与史瓦西黑洞	79
8.1.5	彭罗斯图	79
8.1.6	带电球对称黑洞*	80
8.2	旋转黑洞*	82
8.2.1	Kerr 度规*	82
8.2.2	Kerr 时空的结构*	83
8.2.3	Kerr 时空的最大延拓*	84
8.2.4	彭罗斯过程*	85
8.3	黑洞的一般性讨论	86
8.3.1	奇点与黑洞无毛定理	86
8.3.2	黑洞热力学	87
9	宇宙学	89
9.1	宇宙学的基本概念	89
9.1.1	宇宙学原理	89
9.1.2	宇宙的空间几何与 FLRW 度规	89
9.1.3	哈勃定律	91
9.2	宇宙动力学	92
9.2.1	Friedmann 方程	92
9.2.2	宇宙学常数与暗能量	93
9.2.3	尺度因子的演化	94

9.2.4	宇宙学标准模型	95
9.3	标准模型的修正	96
9.3.1	粒子视界疑难与平坦性疑难	96
9.3.2	暴涨模型	97
9.3.3	巧合性疑难与动力学暗能量	98

第一部分 数学基础

1 线性代数回顾

广义相对论是一个使用弯曲空间描述引力的理论，因此需要使用弯曲空间中的几何学。本章会回顾最基本的本科线性代数和微积分出发，从而为下一章介绍弯曲空间中的几何理论——微分几何打好基础。

下面，我们来简单回顾一下线性代数中的一些知识，这里主要回顾基变换下坐标和矩阵变换的知识，并使用新的记号来表示它们。

1.1 线性空间与对偶空间

1.1.1 线性空间与矢量

我们在《线性代数》一书中说过，线性空间是对于平直空间中的“箭头”的推广。我们在此处不再赘述线性空间的定义。假设我们有一个 n 维实线性空间 V ，其中有一个矢量 v 。我们可以给 V 取一组基 $\{e_\mu, \mu = 1, \dots, n\}$ ，则 v 可以写成基矢量的线性组合

$$v = \sum_{\mu=1}^n v^\mu e_\mu = v^\mu e_\mu$$

这样， v 可以被赋予 n 个坐标分量 v^μ 。注意我们在第二个等号处使用了 **爱因斯坦求和约定**，即同一个指标在上下标各出现一次时，默认对此指标求和，例如上式中 μ 在上下标各出现了一次，则省略掉对此指标的求和记号。本书后面将默认使用这种约定。

下面考虑对 V 取一组新的基 $\{e'_\nu\}$ ，那么 v 可以被赋予另一些坐标分量 v'^ν ，那么它和原来的坐标分量是什么关系呢？我们设坐标变换的矩阵为 P ，也就是 $[e'_1, \dots, e'_n] = [e_1, \dots, e_n]P$ 。那么线代的知识告诉我们，坐标变换应该满足

$$\begin{bmatrix} v'^1 \\ \vdots \\ v'^n \end{bmatrix} = P^{-1} \begin{bmatrix} v^1 \\ \vdots \\ v^n \end{bmatrix}$$

由于我们后面要遇到很高维的“矩阵”，难以用传统的矩阵记号表示，只能用分量的等式表示，因此我们从现在开始练习用分量书写等式。我们把 P^{-1} 的第 μ 行、第 ν 列的分量写作 $(P^{-1})^\mu_\nu$ ，那么上式可以写作

$$v'^\nu = (P^{-1})^\nu_\mu v^\mu$$

这种变换中使用的是基变换矩阵的逆矩阵，因此称为 **逆变**。矢量的分量在基变换下满足逆变规律，因此矢量也被称为**逆变矢量/矢量**。

注意，矢量是线性空间的元素，它是本质的、与基的选取无关的。我们给线性空间选取不同的基，一个矢量就可以有不同的坐标，但这个矢量本身并没有发生任何变化。这一点是非常重要的。

1.1.2 对偶空间与对偶矢量

在有了线性空间 V 后，它自然地就会诱导出一个新的线性空间。我们说过，线性空间 V 向 \mathbb{R} 的所有线性映射也组成一个线性空间 $\text{Hom}(V, \mathbb{R})$ 。我们把它称为 V 的**对偶空**

在本书的记号中，我们不再使用粗体字母表示矢量。同时，在广义相对论中我们习惯使用希腊字母作为维度的指标。

也正因如此，在这种记号下我们必须准确地地区分上下标，例如基矢量一定用下标区分，而坐标分量一定用上标区分。

我们后面会解释为什么把两个指标一个写在上一个写在下。另外注意我们这里 μ 和 ν 虽然写成了上下标，但左右仍然是有区分的，虽然在某些书中不会区分。

间, 记作 V^* , 其中的元素称为对偶矢量或者余矢量. 显然有 $\dim V = \dim V^*$. 对于 V 中的一组基 $\{e_\mu\}$, 我们可以很自然地诱导出 V^* 中的一组基 $\{\epsilon^\mu\}$, 称为对偶基, 满足

$$\epsilon^\mu(e_\nu) = \delta_\nu^\mu$$

也就是说第 μ 个对偶基作为一个线性映射, 它作用在 V 的第 ν 个基上, 当 $\mu = \nu$ 时得到 1, 否则得到 0. 顺带一说, 我们虽然是在 V 中选取了基 $\{e_\mu\}$, 但在上式中使用的却是 e_ν , 这是因为式子中的指标是“活的”, 它们只是独立地表示几个不同的维度的指标而已. 选定一组基以后, 我们可以任意选用字母来表示独立指标.

下面我们来看一看当 V 的基由 $\{e_\mu\}$ 变为 $\{e'_\nu\}$ 时会发生什么. 此时 V^* 的对偶基变为 $\{\epsilon'^\nu\}$. 很容易验算出, 新的对偶基与旧的对偶基之间满足

$$\begin{bmatrix} \epsilon'^1 \\ \vdots \\ \epsilon'^n \end{bmatrix} = P^{-1} \begin{bmatrix} \epsilon^1 \\ \vdots \\ \epsilon^n \end{bmatrix}$$

进而对于一个对偶矢量 $\omega \in V^*$, 它在对偶基下的坐标 $[\omega_1, \dots, \omega_n]$ 的变换满足

$$[\omega'_1, \dots, \omega'_n] = [\omega_1, \dots, \omega_n]P$$

或者写成分量形式

$$\omega'_\nu = P^\mu{}_\nu \omega_\mu$$

这种变换中使用的是基变换矩阵本身, 因此称为 **协变**. V^* 中的对偶矢量的分量在 V 的基变换下满足协变规律, 因此对偶矢量也被称为协变矢量.

在这里同样强调, 对偶矢量是对偶空间的元素, 它也是本质的、与基的选取无关的. 我们给线性空间选取不同的基, 一个对偶矢量就可以有不同的坐标, 但这个对偶矢量本身并没有发生任何变化. 这一点是非常重要的.

我们把矢量与对偶矢量的坐标分量变换规律总结如下:

定理 1.1: 矢量与对偶矢量的坐标变换

对于线性空间 V 和其对偶空间 V^* , 对 V 做基变换 $e'_\nu = P^\mu{}_\nu e_\mu$, 则矢量 $v \in V$ 的分量满足逆变规律

$$v'^\nu = (P^{-1})^\nu{}_\mu v^\mu$$

而对偶矢量 $\omega \in V^*$ 的分量满足协变规律

$$\omega'_\nu = P^\mu{}_\nu \omega_\mu$$

再次强调上下标的作用. 对于基矢量, 下标用来区分一个线性空间的基矢量, 上标用来区分对偶空间的基矢量.

我们把对偶矢量的坐标写成行矢量, 并且把其分量用下标表示, 这样对偶矢量 ω 作用在矢量 v 上 $\omega(v)$ 就可以写成行矢量与列矢量相乘 ωv , 或写作 $\omega_\mu v^\mu$.

注意是 V^* 中的元素对 V 的基变换是协变的, 而它对于 V^* 自身的基变换仍然是逆变的.

我们为什么要定义对偶空间并搞出一个协变规律来呢? 我们下面来看一个例子.

例 1.1. 我们设 V 是我们生活的三维空间的一个一维子空间, 我们以“米 (m)”作为单位, 也就是选择“移动 1m”作为基矢量. 现在我们找到一个位移矢量 x 是“移动 2m”, 那么它的坐标就是 2. 现在我们变为以“厘米 (cm)”为单位, 也就是做坐标变换 $P^\mu{}_\nu = 0.01\delta^\mu_\nu$, 那么 x 的坐标变为 200, 这就是逆变规律——与基矢量的变化方式相反.

现在我们考虑另外一种矢量——梯度矢量. 例如空间中存在着标量电势场 φ , 其

负梯度是一个矢量，表示电场强度 $E = -\nabla\varphi$ 。梯度表示标量场随着空间的变化程度。例如电场是每米变化 5V，用厘米作为单位后，电场则表示为每厘米变化 0.05V，也就是坐标缩小了 100 倍。这种变化与基矢量的变化方式相同，也就是说梯度矢量是协变矢量。

在本节的最后，我们再次强调，矢量和对偶矢量本质上是线性空间和对偶空间的元素，它们与基的选取无关。选取不同的基以后，它们就可以有不同的坐标，但它们本身不随基的变化而变化。

1.2 张量

1.2.1 张量的定义

我们上一节讲的矢量和对偶矢量都是张量的例子。我们下面来看一个稍稍复杂一些的例子，里面综合了协变和逆变规律。

注意带上下标的都是分量，也就是数，因此它们可以交换位置。

例 1.2. 考虑一个线性变换 $T: V \rightarrow V$ ，它在选定一组基后表示为一个矩阵 A 。若使用 P 进行基变换，则 T 的矩阵变为一个相似矩阵 $A' = P^{-1}AP$ 。或者用分量表示为

$$A'^{\alpha}_{\beta} = (P^{-1})^{\alpha}_{\mu} P^{\nu}_{\beta} A^{\mu}_{\nu}$$

也就是说， T 也是一个与坐标无关的东西，它的分量在基变换下会发生变化，其第一个指标满足逆变规律，第二个指标满足协变规律。

有了以上几个例子，我们来给出张量的定义。张量有许多种定义方式，我们在此以坐标变换的规律来定义张量。

定义 1.1: 张量

线性空间 V 上的一个 (p, q) -型张量 T 是一个与基无关的数学对象，它在给定基下表示为一个 $p + q$ 维数组，其中 p 个维度写作上指标，满足逆变规律， q 个维度写作下指标，满足协变规律，即在坐标变换 $e'_{\nu} = P^{\mu}_{\nu} e_{\mu}$ 下，其分量满足

$$T'^{\alpha_1 \dots \alpha_p}_{\beta_1 \dots \beta_q} = (P^{-1})^{\alpha_1}_{\mu_1} \dots (P^{-1})^{\alpha_p}_{\mu_p} \cdot P^{\nu_1}_{\beta_1} \dots P^{\nu_q}_{\beta_q} \cdot T^{\mu_1 \dots \mu_p}_{\nu_1 \dots \nu_q}$$

我们把 V 上所有的 (p, q) -型张量构成的集合记作 $\mathcal{T}_V(p, q)$ 。

标量可以看作是一个 $(0, 0)$ -型张量。

张量定义的核心在于它是一个与基的选取无关的量。它既然与基的选取无关，那么在选取不同的基的时候可能“看起来”是不一样的，也就是其分量是可以发生变化的。几何学的灵魂也在于此——我们可以通过建立坐标系来计算几何问题，但本质上描述几何的量应该与坐标系的选取无关。这就是定义张量的目的。

1.2.2 张量的运算

有了张量的定义，我们自然就希望定义它的运算。张量可以进行数乘，同一个线性空间上同形状的张量可以进行加减，这两种运算的定义都非常直观，也就是给定基后将分量对应进行数乘或相加即可。也容易理解，数乘或相加后的分量在坐标变换下仍然满足张量的变换规律，因此仍然是张量。

下面我们来定义张量的另外两种运算——张量积和缩并。

定义 1.2: 张量积

设 $T \in \mathcal{T}_V(p, q)$, $S \in \mathcal{T}_V(k, \ell)$, 那么它们的张量积 $T \otimes S \in \mathcal{T}_V(p+k, q+\ell)$, 且其分量满足

$$(T \otimes S)^{\alpha_1 \cdots \alpha_p \mu_1 \cdots \mu_k}_{\beta_1 \cdots \beta_q \nu_1 \cdots \nu_\ell} = T^{\alpha_1 \cdots \alpha_p}_{\beta_1 \cdots \beta_q} S^{\mu_1 \cdots \mu_k}_{\nu_1 \cdots \nu_\ell}$$

有时我们也会保持指标的顺序, 也就是把等号左侧的指标写作

$$(T \otimes S)^{\alpha \cdots \mu \cdots}_{\beta \cdots \nu \cdots}$$

这个定义非常直观, 张量积把协变和逆变的维度分别叠加起来, 且张量积的分量等于分量的积. 我们后面写张量积的时候, 一般直接会按照等式右侧这种分量相乘的方式书写. 下面我们来看张量积的一个简单例子

例 1.3. 对于两个矢量 $u, v \in V$, 它们的外积

$$uv^\top = \begin{bmatrix} u^1 v^1 & \cdots & u^1 v^n \\ \vdots & \ddots & \vdots \\ u^n v^1 & \cdots & u^n v^n \end{bmatrix}$$

这实际上就是 u 和 v 作为张量的张量积 $u \otimes v$, 因为其分量为 $(u \otimes v)^{\mu\nu} = u^\mu v^\nu$.

张量积用低阶张量构建高阶张量. 那么, 对于任意一个 (p, q) -阶张量 T , 我们希望用最低阶的张量——矢量和对偶矢量的张量积构建它. 考虑基矢量组 $\{e_\mu\}$ 和对偶基矢量组 $\{\epsilon^\nu\}$, 它们可以构建出一系列 (p, q) -型张量 $e_{\mu_1} \otimes \cdots \otimes e_{\mu_p} \otimes \epsilon^{\nu_1} \otimes \cdots \otimes \epsilon^{\nu_q}$. 注意这里所有的角标都可以从 1 取到 n , 这是一系列张量而非一个张量. 事实上, 这一系列张量构成了 $\mathcal{T}_V(p, q)$ 的一组基. 也就是说, 当我们说 T 的分量是 $T^{\mu_1 \cdots \mu_p}_{\nu_1 \cdots \nu_q}$ 时, 我们实际上是在说

$$T = T^{\mu_1 \cdots \mu_p}_{\nu_1 \cdots \nu_q} e_{\mu_1} \otimes \cdots \otimes e_{\mu_p} \otimes \epsilon^{\nu_1} \otimes \cdots \otimes \epsilon^{\nu_q}$$

当我们分解的是一个 $(1, 0)$ -型张量也就是矢量 v 时, 上式就退化成了我们熟悉的 $v = v^\mu e_\mu$. 同样, 对于一个 $(1, 1)$ -型张量也就是一个线性变换 A , 上式就成了 $A = A^\mu_\nu e_\mu \otimes \epsilon^\nu$.

下面我们来看张量的另一种运算. 我们先从矩阵引入. 对于一个线性变换 T 的矩阵 A , 它的迹 $\text{tr} A$ 定义为其对角线元素相加, 即 $\text{tr} A = A^\mu_\mu$. 可以证明, 这个数与基的选取无关, 即无论把 T 写成哪组基下的矩阵, 其迹都是相同的. 这就说明, 线性变换作为一个张量, 其迹仍然是一个张量. 这给了我们由张量运算出新的张量的方式.

定义 1.3: 张量的缩并

设 $T \in \mathcal{T}_V(p, q)$, 其第 i 个逆变指标与第 j 个协变指标的缩并是一个属于 $\mathcal{T}_V(p-1, q-1)$ 的张量, 且其分量为对这两个指标所有相等的情况取和, 即

$$T^{\mu_1 \cdots \mu_{i-1} \rho \mu_{i+1} \cdots \mu_k}_{\nu_1 \cdots \nu_{j-1} \rho \nu_{j+1} \cdots \nu_\ell}$$

张量积和缩并组合在一起, 可以表示出我们以前所有学过的矢量与矩阵的运算. 我们在此看一下.

例 1.4. 对于两个矢量 u, v , 其内积 $\langle u, v \rangle$ 可以用矩阵写作 $u^\top G v$, 我们可以用分量的语言写作

$$\langle u, v \rangle = G_{\mu\nu} u^\mu v^\nu$$

这实际上就是我们把 G, u, v 进行张量积产生 $G_{\alpha\beta} u^\alpha v^\beta$ 后再对 α 与 μ 、 β 与 ν 缩并而

成的。

例 1.5. 对于两个矩阵 A, B , 其乘积 $C = AB$ 可以用分量的语言写作

$$C^\mu{}_\nu = A^\mu{}_\rho B^\rho{}_\nu$$

这实际上是将 A, B 进行张量积得到 $A^\mu{}_\alpha B^\beta{}_\nu$ 后将 α 和 β 缩并而成的。

我们可以观察到, 张量分量的等式总是满足: **上下配对, 左右平衡**. 即根据爱因斯坦求和约定, 配对求和的指标必须在等号同侧一上一下. 而没有剩下的配对求和的指标则必须在等号两边平衡, 表示剩余这些指标的分量在两侧相等.

最后强调一下, 在张量等式中, $T^{\mu\nu}{}_\sigma$ 这种写法直接表示的是选定了一组基以后张量 T 的 ${}^{\mu\nu}{}_\sigma$ 位置的分量, 而并不是张量本身. 但这种张量等式往往本质上是想描述张量的数学关系, 而非他们的分量的数学关系. 只是在高阶张量中, 用分量写等式往往更简单、更清楚一些. 可以认为, 这种张量等式在所有坐标系下都成立. 只有在讨论具体的例子时, 我们可能会选定某个特定的坐标系, 这时我们会强调我们写出的是该特定坐标系下的分量等式, 而非张量等式.

注意, 在同一个式子中, 只要主字母相同就是同一个张量. 例如矩阵等式 $A = IA$ 可以写作

$$A^\mu{}_\nu = \delta^\mu{}_\rho A^\rho{}_\nu$$

其中 $A^\mu{}_\nu$ 和 $A^\rho{}_\nu$ 都表示矩阵 A .

在少数教科书中, 会使用英文字母表示张量等式, 并称之为抽象指标记号, 如 $T^{ab}{}_c$, 只在表示具体的分量时使用希腊字母. 但这种记号的使用并不广泛.

1.2.3 对称与反对称张量

在后面的学习中, 由于张量定义的对称性、或者时空本身的对称性, 我们会经常会讨论张量的对称性. 在此, 我们介绍一下张量对称性相关的知识.

张量的对称性来自非常简单而直观的定义. 一个张量 T 若满足 $T_{\dots\mu\dots\nu\dots} = T_{\dots\nu\dots\mu\dots}$, 则称 T 关于指标 μ 和 ν 是对称的. 若满足 $T_{\dots\mu\dots\nu\dots} = -T_{\dots\nu\dots\mu\dots}$, 则称 T 关于指标 μ 和 ν 是反对称的.

这种定义可以推广至多个指标. 对于张量 $T_{\mu_1\mu_2\dots\mu_n}$, 我们定义它的对称部分为 (一个新的张量)

$$T_{(\mu_1\mu_2\dots\mu_n)} = \frac{1}{n!} \sum (T \text{ 的指标的所有排列})$$

而其反对称部分定义为

$$T_{[\mu_1\mu_2\dots\mu_n]} = \frac{1}{n!} \sum (T \text{ 的指标的所有偶排列} - T \text{ 的指标的所有奇排列})$$

这样, 若 $T_{\mu_1\mu_2\dots\mu_n} = T_{(\mu_1\mu_2\dots\mu_n)}$, 则我们说 T 是对称的; 若 $T_{\mu_1\mu_2\dots\mu_n} = T_{[\mu_1\mu_2\dots\mu_n]}$, 则我们说 T 是反对称的.

我们上面以协变指标为例定义了张量的对称部分和反对称部分, 同样的定义也可以用于逆变指标. 但是注意, 对称与反对称只能在都是协变指标时或者都是逆变指标时讨论, 而不能纵跨二者. 另一点需要指出的是, 括号可以只括一部分指标, 同时还可以横跨多个进行张量积的指标. 例如, $T_{\mu[\nu}{}^\sigma S_{\alpha]\beta}$ 是只对 ν 和 α 两个指标进行反对称化, 而不涉及其他指标.

1.3 度规

1.3.1 度规张量

在线性代数中我们说过, 对于一个线性空间, 我们可以定义一个线性、正定、对称的二元函数 $\langle \cdot, \cdot \rangle$, 叫做内积. 由内积可以定义矢量的长度 $\|v\| = \sqrt{\langle v, v \rangle}$.

在相对论中，我们也需要衡量矢量的长度。但由于我们所在的时空的特征，我们需要对内积的正定性要求放松，改为“非退化的”。这样得到的“内积”，称为度规。

定义 1.4: 度规

对于线性空间 V ，**度规** 是一个附加结构，它是一个二元函数 $g: V \times V \rightarrow \mathbb{R}$ ，满足

1. 线性性: $g(au_1 + bu_2, v) = ag(u_1, v) + bg(u_2, v)$;
2. 对称性: $g(u, v) = g(v, u)$;
3. 非退化性: 若矢量 v 满足对任意 u 有 $g(u, v) = 0$ ，则有 $v = 0$ 。

定义了度规的线性空间称为广义欧氏空间。

显然，度规就是一个非退化的双线性型。在给定了一组基以后，度规可以写作一个矩阵 g ，使得 $g(u, v) = u^\top g v$ ，或者用分量的语言写作

$$g(u, v) = g_{\mu\nu} u^\mu v^\nu$$

根据角标的位置可知，度规相当于一个 $(0, 2)$ -型张量。的确，双线性型的矩阵在基变换 P 下满足合同变换 $g' = P^\top g P$ ，或者可以写成 $g'_{\alpha\beta} = P^\mu_\alpha P^\nu_\beta g_{\mu\nu}$ ，可知两个指标都满足协变规律。

度规的非退化性对应着该张量的矩阵的行列式非零，即 $\det g \neq 0$ 。Sylvester 惯性定理告诉我们，一个双线性型一定可以在一组基下写成对角元只有 $+1$ 、 -1 和 0 的对角矩阵。而非退化则要求这些对角元中不能有 0 ，也就是说只能有 $+1$ 和 -1 ，其中 $+1$ 的个数称为度规的正惯性指数， -1 的个数称为负惯性指数。

在相对论中，我们强调两种特殊的度规——欧式度规和闵氏度规。

例 1.6. 当对角化的度规的对角元全部为 $+1$ 时，称该度规为 **欧式度规**（欧几里得度规）或正定度规，配备了欧式度规的线性空间显然就是欧氏空间。欧式度规在标准正交基下写为

$$\delta_{\mu\nu} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}_{\mu\nu}$$

当对角元只有一个 -1 ，其余全部都是 $+1$ 时，称该度规为 **闵氏度规**（闵可夫斯基度规），配备了闵氏度规的线性空间为闵氏空间。闵氏度规在标准正交基下写为

$$\eta_{\mu\nu} = \begin{bmatrix} -1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}_{\mu\nu}$$

也有地方会把闵氏度规的分量写作

$$\text{diag}(+1, -1, -1, -1)$$

这两种写法甚至一样常见。

有了度规后我们就可以定义矢量的长度。我们以前通过 $\sqrt{\langle v, v \rangle}$ 定义长度，但现在 $g(v, v)$ 可以为负，因此我们对矢量长度的定义为

$$|v| = \sqrt{|g(v, v)|} = \sqrt{|g_{\mu\nu} v^\mu v^\nu|}$$

不过我们可以按照 $g(v, v)$ 的符号将 v 进行分类。不同类的比较是没有意义的。我们后面

会再提到这一点. 不过我们在此需要指出, 在闵氏度规下, $|v| = 0$ 并不意味着 v 是零矢量 (线性空间中的零元). 我们后面会对这些进行进一步讨论.

同样, 类比于线性代数里的欧氏空间, 我们也可以定义矢量之间的角度关系. 不过在非欧氏空间中, 我们往往不太关注一般的角度, 而只关心一种特殊的角度关系——垂直, 即正交: 我们定义两个矢量 u 和 v 正交当且仅当他们的内积为零, 即

$$g(u, v) = g_{\mu\nu} u^\mu v^\nu = 0$$

有了长度和正交关系的存在, 我们就可以定义我们中学时常用的“直角坐标系”了. 对于一组基 $\{e_\alpha\}$, 若它们的长度都为 1, 且两两正交, 即

$$g_{\mu\nu}(e_\alpha)^\mu(e_\alpha)^\nu = \pm 1, \quad g_{\mu\nu}(e_\alpha)^\mu(e_\beta)^\nu = 0, \quad (\alpha \neq \beta)$$

则称这组基 $\{e_\alpha\}$ 为一组 **标准正交基**. 一种等价说法是, 若度规 g 在某组基下化为规范型, 即对角元都为 ± 1 的对角矩阵, 则这组基为标准正交基. 标准正交基不仅可以简便计算, 在我们后面定义弯曲空间上的一些概念时也很有作用.

这里的指标有点复杂, e_α 表示第 α 个基向量, 它是一个向量; 而 $(e_\alpha)^\mu$ 表示它的第 μ 分量, 指标 μ 才是我们常用的指标的意义.

1.3.2 指标升降

有了度规以后, 对于任意一个矢量 v^μ , 我们都可以将它和一个对偶矢量 $g_{\mu\nu}v^\mu$ 一一对应. 因此我们把这个对偶矢量也记作 v , 但用下角标表示, 也就是 $v_\nu = g_{\mu\nu}v^\mu$. 由于这个映射是双射, 因此我们也一定可以将 v_μ 反向映射为一个矢量, 这也是通过一个张量来完成的, 并且是一个 $(2, 0)$ -型张量, 我们将它记作 $g^{\mu\nu}$, 并且也称为度规. 这样我们有逆映射 $v^\nu = g^{\mu\nu}v_\mu$. 这种操作称为 **指标升降**.

由于协变和逆变的两个度规互为逆映射, 它们显然满足

$$g^{\mu\sigma}g_{\sigma\nu} = \delta^\mu_\nu$$

例 1.7. 我们上面提到的欧式度规和闵氏度规的协变度规和逆变度规的分量相同, 即在标准正交基下, 有

$$\delta^{\mu\nu} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}_{\mu\nu}, \quad \eta^{\mu\nu} = \begin{bmatrix} -1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}_{\mu\nu}$$

对于一个任意阶张量, 我们都可以用度规对其中的指标进行升降. 由指标升降联系起来的张量被认为是同一个物理量的不同描述. 例如对于张量 $T^{\mu\nu}_\sigma$, 它的各种指标升降版本如 $T_{\alpha\beta\sigma} = g_{\alpha\mu}g_{\beta\nu}T^{\mu\nu}_\sigma$ 、 $T^\nu_{\alpha\sigma} = g_{\alpha\mu}T^{\mu\nu}_\sigma$ 等都使用字母 T 表示, 并被认为描述同一个物理量. 在指标平衡的情况下, 张量等式可以把指标随意上下移动, 究竟写哪个常常取决于习惯. 例如下面几个式子实际上是完全相同的张量等式.

$$R_{\mu\nu} = S^\sigma_\mu T_{\sigma\nu}, \quad R^{\mu\nu} = S_\sigma^{\mu\nu} T^{\sigma\nu}, \quad R^\mu_\nu = S^{\sigma\mu} T_{\sigma\nu}$$

指标升降的一个重要作用是使得我们可以缩并两个同为协变或同为逆变的指标, 例如 $T_{\mu\nu}$ 的缩并可以定义为将其一个指标升上去得到 $T^\sigma_\mu = g^{\sigma\nu}T_{\mu\nu}$ 以后再进行缩并, 最终得到

$$T^\mu_\mu = g^{\mu\nu}T_{\mu\nu}$$

在数学上, 如果我们可以在两个数学结构之间建立性质足够好的双射, 那么我们就说这两个数学结构是同构的, 并且我们可以将一一对应起来的元素视作等同, 因为有一个就有另一个.

当然我们也可以类似地缩并高阶张量中的两个指标. 注意, 我们定义的缩并 $T^\mu{}_\mu$ 和直接将 $T_{\mu\nu}$ 写成矩阵后对主对角线求和 (即 $\sum_\mu T_{\mu\mu}$) 常常是不同的, 这是因为把其中一个指标升上去时矩阵可能会变. 我们永远不会对张量进行类似于后者的计算, 这是因为它不再会得到一个张量.

在后面的章节中, 我们后面会见到指标升降的具体的例子.

2 微分几何基础

微分几何是在弯曲空间中的几何理论，是广义相对论的基本数学语言。我们下面来学习一些最基础的微分几何。

2.1 流形和张量场

2.1.1 微分流形

我们前面讲的线性代数所研究的数学结构是线性空间，而微分几何中我们研究的是弯曲的空间，这种空间称为 **微分流形**。不严谨地说，一个 n 维微分流形是满足以下两条性质的“空间”：

1. 局部像欧氏空间：对于任意一点 p ，一定存在至少一个包含 p 的开区域 U 可以被一个单射、连续的函数 ϕ 映射到 \mathbb{R}^n 上。这构成了 p 点附近的一个 **局域坐标系** (U, ϕ) ；
2. 光滑性：若两个坐标系 (U_i, ϕ_i) 和 (U_j, ϕ_j) 有重叠，则重叠处的坐标变换 $\phi_j \circ \phi_i^{-1}$ 必须光滑，这称为相容性。

这个定义可能有些抽象，下面让我们来看两个具体的例子。

例 2.1. 对于空间 \mathbb{R}^n ，整个这个空间可以作为一个坐标域，并用恒等映射 id 作为坐标函数。因此我们只需要一个坐标系 $(\mathbb{R}^n, \text{id})$ 。也就是说 \mathbb{R}^n 是一个 n 维微分流形。这种整个空间可以用一个坐标域覆盖的流形称为平凡流形。

例 2.2. 对于一个圆周 S^1 ，我们无法用一个坐标域去覆盖它，但我们可以选取多个坐标域。例如，我们可以设三个坐标域 U_1, U_2, U_3 为 S^1 的（开）上半圆、左半圆和下半圆，它们上的坐标函数就是将其中的点映为 $[0, 2\pi)$ 内的幅角，而坐标域 U_4 是右半圆，其坐标函数为 $[-\pi, \pi)$ 内的幅角。这样很容易发现 U_1, U_2, U_3 之间的坐标变换函数就是恒等函数，而 U_1 和 U_4 、 U_3 和 U_4 之间的坐标变换就是减去 π ，它们都是无穷次可微的。因此 S^1 是一个一维微分流形。

但强调一下，我们在此学习的微分几何是弯曲空间的“内蕴几何”。在上面的例子和示意图中，我们都是在一个更高维的空间中画出了我们的流形，例如在三维空间中画出了一个曲面。但是，实际上我们的一切定义和计算都是在这个流形内的，而并没有跳出流形进入这个高维空间。这一点在后面的曲率一节会有更多的体会。

2.1.2 流形的切空间和与余切空间

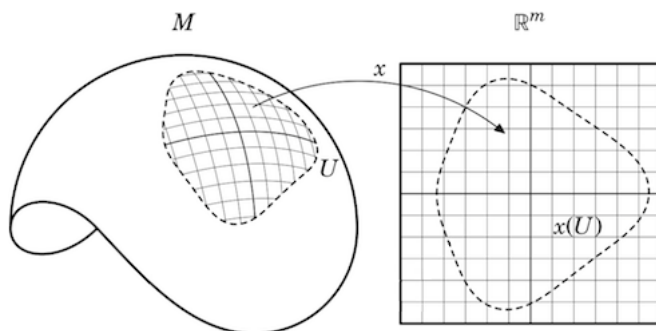
流形作为光滑的弯曲空间，自然在每一点都有一个与其相切的平直空间，称为 **切空间**，例如球面的切平面。流形 M 在点 p 的切空间记作 $T_p M$ 。切空间是一个线性空间，其中的矢量称为 **切矢量**，也就是与流形相切的矢量。

既然切空间是线性空间，我们自然就希望能写出它的一组基。我们在定义中说过，对于任意一点 p ，它一定有一个邻域 U 可以被一一、连续地映到 \mathbb{R}^n 上，从而给每个点赋予一个局部坐标 (x^1, \dots, x^n) 。而 \mathbb{R}^n 相当于一个 n 维网格，这就相当于我们可以在流形 M 上（在 U 内）“画出”这 n 维网格，如下图所示。

这里要求开区域是为了讨论连续性，没有学过拓扑学的读者可以不去理解而直接接受这个设定。

光滑的意思是无穷次可微 (C^∞)。第二点的主要目的是，一个函数（或张量场等）若在一个坐标系下光滑，则在另一坐标系下也光滑，这样这个流形的微分性质才是够好。

我们在这不给出切矢量的严格定义，我们可以暂且不以内蕴的方式想，而就以外在的方式理解切空间。



这样，对于切空间我们很自然地就可以选择那 n 条“沿着网格方向”的“单位矢量”作为基。出于某些复杂的原因，我们把沿着 x^μ 方向的基矢量记作 $\frac{\partial}{\partial x^\mu}$ 。这样，对于任意一个切矢量 $v \in T_p M$ ，我们可以将它分解为分量

注意，分母里的上标相当于下标。

$$v = v^\mu \frac{\partial}{\partial x^\mu}$$

例 2.3. 考虑流形 M 上的一条参数曲线 $C: \mathbb{R} \rightarrow M$ 。 C 的切矢量自然就是 M 的切矢量。我们不严谨地定义这条曲线在点 $p = C(t)$ 处的切矢量为

$$\frac{\partial}{\partial t} = \frac{\overrightarrow{C(t+dt)C(t)}}{dt}$$

注意 $C(t)$ 是 M 上的点，因此 $\overrightarrow{C(t+dt)C(t)}$ 的意思是参数从 t 处向前走了 dt 所画出的轨迹箭头。至于这个切矢量为什么记作 $\partial/\partial t$ ，我们在此也不做解释，不过可以看出这么表示下后面的展开公式很合理。

那么这条切矢量在基矢量下的展开式为

$$\frac{\partial}{\partial t} = \frac{dx^\mu}{dt} \frac{\partial}{\partial x^\mu}$$

这个展开式看起来有些奇怪，但它其实就是 $v = v^\mu e_\mu$ 而已。 dx^μ/dt 是坐标分量，相当于 v^μ ，而 $\partial/\partial x^\mu$ 是基矢量，相当于 e_μ 。

其中分量 dx^μ/dt 是指，将曲线 C 写成点的各个坐标分量关于 t 的函数，则 C 可以表示为 n 个分量函数 $x^\mu = C^\mu(t)$ 。对于这 n 个方程我们可以对 t 求导，即可得到 dx^μ/dt 。

例 2.4. 设流形 M 上有一个标量场 f 。 v 是 M 上的一个切矢量，则 f 沿着 v 的方向导数为 v 和 ∇f 的内积，即为

$$v^\mu \frac{\partial}{\partial x^\mu} f$$

而上式中 f 前面的那个整体就是 v ，因此我们可以把一个矢量 v 看作一个从标量场到标量场的映射， $v(f)$ 就是 f 沿着 v 的方向导数。这样矢量 v 就可以看作一个从标量场到标量场的函数。

由此，我们要介绍一个很重要的矢量运算，称为 **对易子** 或 **李括号**。对于矢量 u 和 v ，我们可以定义出一个新的矢量，记作 $[u, v]$ ，定义为

$$[u, v](f) = u(v(f)) - v(u(f))$$

可以证明我们定义出的 $[u, v]$ 是满足矢量的分量变换规律的，因此它的确是一个矢量。

既然切空间 $T_p M$ 是一个线性空间，它自然就有其对偶空间，我们将其称为 **余切空间**，记作 $T_p^* M$ 。而切空间由坐标系产生的这一组基 $\{\partial/\partial x^\mu\}$ 自然也可以诱导出的一组对偶基，同样由于某些复杂的原因，我们将其记为 dx^μ 。也就是说对于一个余切矢量 $\omega \in$

T_p^*M , 我们可以将其写作

$$\omega = \omega_\mu dx^\mu$$

例 2.5. 考虑 M 上的一个标量场, 即一个实值函数 $f: M \rightarrow \mathbb{R}$, 则 f 的“梯度” df 是 M 上的一个余切矢量, 它指向 f 增大最快的方向, 并且大小正比于 f 增大的速度. 它的分量展开式为

$$df = \frac{\partial f}{\partial x^\mu} dx^\mu$$

与上面的例子类似, $\partial f / \partial x^\mu$ 是指将 f 写成点的坐标向 \mathbb{R} 的函数, 则可以写成一个 n 元函数 $f(p) = f(x^1, \dots, x^n)$, 对这个函数求偏导即可得到 $\partial f / \partial x^\mu$.

下面我们来讨论切空间与余切空间中的坐标变换. 对于一个点 p , 它的切空间的基可以用一个局域坐标系 (U, ϕ) 诱导出 $\partial / \partial x^\mu$, 也可以用另一个局域坐标系 (U', ϕ') 诱导出 $\partial / \partial x'^\nu$. 那么二者之间满足什么变换关系呢? 我们知道, 坐标变换 $x^\mu \mapsto x'^\nu$ 在局部可以近似为一个线性变换, 其矩阵为 $J^\mu{}_\nu = \partial x^\mu / \partial x'^\nu$. 代入我们前面的基变换公式, 即可得到

注意系数 $\partial x'^\nu / \partial x^\mu$ 是在点 p 处的取值.

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

而切矢量 $v \in T_p$ 的分量的变换方式与基矢量相反 (逆变), 因此我们有

$$v'^\nu = \frac{\partial x'^\nu}{\partial x^\mu} v^\mu$$

对于余切矢量, 我们也可以类似地写出基变换公式

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

和分量变换公式

$$\omega'_\nu = \frac{\partial x^\mu}{\partial x'^\nu} \omega_\mu$$

2.1.3 张量与张量场

有了切矢量和余切矢量的基础, 我们就可以讨论流形上的张量了. 流形 M 上一点 p 处的张量即为切空间 $T_p M$ 上的一个张量, 或者说它是 $T_p M$ 和余切空间 $T_p^* M$ 共同“生出”的一个数学对象. 像我们之前所说的, 它是一个不随坐标系变化的东西, 但可以利用坐标系展开为分量. 由于我们把切空间的基写作 $\partial / \partial x^\mu$, 把余切空间的基写作 dx^μ , 因此 M 上的一个 (p, q) -阶张量的展开形式为

$$T = T^{\mu_1 \dots \mu_p}{}_{\nu_1 \dots \nu_q} \frac{\partial}{\partial x^{\mu_1}} \otimes \dots \otimes \frac{\partial}{\partial x^{\mu_p}} \otimes dx^{\nu_1} \otimes \dots \otimes dx^{\nu_q}$$

而其分量变换公式为

$$T'^{\alpha_1 \dots \alpha_p}{}_{\beta_1 \dots \beta_q} = \frac{\partial x'^{\alpha_1}}{\partial x^{\mu_1}} \dots \frac{\partial x'^{\alpha_p}}{\partial x^{\mu_p}} \frac{\partial x^{\nu_1}}{\partial x'^{\beta_1}} \dots \frac{\partial x^{\nu_q}}{\partial x'^{\beta_q}} T^{\mu_1 \dots \mu_p}{}_{\nu_1 \dots \nu_q}$$

最后, 我们后面要研究的大多都是流形上的 **张量场**. 流形上的每一点都有它的切空间和余切空间, 因此每一点都可以有一个自己的张量 (形状相同), 它们构成了流形上的一个张量场. 我们把流形 M 上的所有 (p, q) 型光滑张量场的集合记作 $\mathcal{F}_M(p, q)$. 张量场是我们以后要主要接触的数学对象.

流形上的所有点的切空间共同构成流形的切丛, 而所有点的余切空间共同构成余切丛.

2.1.4 张量密度*

考虑一个形如张量的量 $\mathfrak{T}^{\mu_1 \dots \mu_p}_{\nu_1 \dots \nu_q}$ ，它表示某个张量场 T 在给定坐标系 $\{x^\mu\}$ 后“一个微小坐标格”的“总张量”。即

$$\mathfrak{T}^{\mu_1 \dots \mu_p}_{\nu_1 \dots \nu_q} = V \left(\frac{\partial}{\partial x} \right) T^{\mu_1 \dots \mu_p}_{\nu_1 \dots \nu_q}$$

其中 $V(\partial/\partial x)$ 表示坐标系 $\{x^\mu\}$ 下一个微小“坐标网格”的“体积”，它可以写成带有符号的 $dx^1 dx^2 \dots dx^n$ 。根据微积分和线性代数所学，这一项在坐标系变换下满足

$$V \left(\frac{\partial}{\partial x'^{\nu}} \right) = \left(\det \frac{\partial x'}{\partial x} \right) V \left(\frac{\partial}{\partial x^\mu} \right)$$

其中 $\det \frac{\partial x'}{\partial x}$ 是坐标变换的雅可比行列式。这样， $\mathfrak{T}^{\mu_1 \dots \mu_p}_{\nu_1 \dots \nu_q}$ 在坐标系变换下的分量变换规律为

$$\mathfrak{T}'^{\alpha_1 \dots \alpha_p}_{\beta_1 \dots \beta_q} = \left(\det \frac{\partial x'}{\partial x} \right) \frac{\partial x'^{\alpha_1}}{\partial x^{\mu_1}} \dots \frac{\partial x'^{\alpha_p}}{\partial x^{\mu_p}} \frac{\partial x^{\nu_1}}{\partial x'^{\beta_1}} \dots \frac{\partial x^{\nu_q}}{\partial x'^{\beta_q}} \mathfrak{T}^{\mu_1 \dots \mu_p}_{\nu_1 \dots \nu_q}$$

因此，这种量的分量变换规律和张量相比，相差一个雅可比行列式。进一步推广，有些量可能正比于体积的多次方。由此我们就定义出了张量密度的概念。

这个“微小体积”严格来说叫体积形式，在后面的微分形式部分会学到。

注意本式中行列式和体积形式中的指标都是自己的，也就是说并没有爱因斯坦求和约定。

定义 2.1: 张量密度

张量密度 $\mathfrak{T}^{\mu_1 \dots \mu_p}_{\nu_1 \dots \nu_q}$ 是一种形如张量的量，它的分量变换满足

$$\mathfrak{T}'^{\alpha_1 \dots \alpha_p}_{\beta_1 \dots \beta_q} = \left(\det \frac{\partial x'}{\partial x} \right)^W \frac{\partial x'^{\alpha_1}}{\partial x^{\mu_1}} \dots \frac{\partial x'^{\alpha_p}}{\partial x^{\mu_p}} \frac{\partial x^{\nu_1}}{\partial x'^{\beta_1}} \dots \frac{\partial x^{\nu_q}}{\partial x'^{\beta_q}} \mathfrak{T}^{\mu_1 \dots \mu_p}_{\nu_1 \dots \nu_q}$$

其中 W 称为张量密度 $\mathfrak{T}^{\mu_1 \dots \mu_p}_{\nu_1 \dots \nu_q}$ 的权。

这样，我们上面举的正比于体积的例子即为一个权为 -1 的张量密度。我们再来看一个重要的例子。

例 2.6. 考虑 Levi-Civita 符号

$$\epsilon_{\mu_1 \mu_2 \dots \mu_n} = \begin{cases} +1, & \text{当 } (\mu_1 \mu_2 \dots \mu_n) \text{ 是偶排列} \\ -1, & \text{当 } (\mu_1 \mu_2 \dots \mu_n) \text{ 是奇排列} \\ 0, & \text{其他} \end{cases}$$

它显然不是一个张量，因为它的分量变换不满足张量的变换规律。可以证明，它是一个权为 1 的张量密度。如果我们把它乘上 $\sqrt{|g|}$ ，其中 g 为度规张量的行列式，即定义

$$\tilde{\epsilon}_{\mu_1 \mu_2 \dots \mu_n} = \sqrt{|g|} \epsilon_{\mu_1 \mu_2 \dots \mu_n}$$

则 $\tilde{\epsilon}_{\mu_1 \mu_2 \dots \mu_n}$ 成为了一个张量，称为 Levi-Civita 张量。

事实上，由于度规张量的行列式 g 是一个权为 -2 的标量密度，所以任意一个权为 W 的张量密度在乘上 $|g|^{W/2}$ 以后都可以得到一个张量。

2.2 流形上的度规、联络和曲率

我们说过流形就是一个弯曲的空间，但是我们目前定义的流形并不支持在上面做几何，因为上面既没有定义长度也没有定义角度。为了使流形上可以讨论长度和角度的概

念，我们需要在流形上引入一个附加结构，这个结构就是度规。同时，为了联系流形上相邻的两点，我们也需要引入另外一个附加结构，称为联络。由度规和联络，我们可以在流形上研究更加丰富的几何学。

2.2.1 流形上的度规

当流形上每一个切空间都有自己的度规张量时，我们就得到了流形的一个度规场 g 。我们说过，度规的作用是定义矢量的长度。对于流形上点 p 处的一个切矢量 v ，它的长度定义为 $|v| = \sqrt{g(v, v)} = \sqrt{g_{\mu\nu} v^\mu v^\nu}$ 。

现在考虑流形上的一条参数曲线 $C(t)$ ，它的切矢量表示为 v ，那么我们很自然地定义它的一段微元线长为

$$dl = |v|dt = \sqrt{g_{\mu\nu} v^\mu v^\nu} dt$$

而我们在例 2.3 中说过，切矢量的分量为 $v^\mu = dx^\mu/dt$ ，代入得到

$$dl = \sqrt{\left| g_{\mu\nu} \frac{dx^\mu}{dt} \frac{dx^\nu}{dt} \right|} dt = \sqrt{|g_{\mu\nu} dx^\mu dx^\nu|}$$

我们在此引入 **线元** ds^2 的概念，定义为

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu$$

线元用于表示流形上一段微小的线长，它实际上就是对度规的分量的另一种写法。不过注意，线元虽然写作 ds^2 ，但它并不是任何东西的平方，它是一个整体符号。而且显然在非正定度规下，线元是可以为负的。

最后，对于整个一段曲线的长度，它自然地被定义为

$$l = \int dl = \int \sqrt{|g_{\mu\nu} dx^\mu dx^\nu|} = \int \sqrt{|ds^2|}$$

我们来考虑两种度规场。当流形 M 的度规处处为欧式度规时，则此时的流形称为 **黎曼空间**，而该度规场称为黎曼度规。当度规处处为闵氏度规时，流形称为伪黎曼空间，在物理上称为 **洛伦兹时空**，简称时空，其上的度规场称为洛伦兹度规。下面我们来看一看这两种空间的最简单的例子，也就是平直空间的例子。

流形上每点的度规是欧式还是闵氏还是其他度规，取决于其正负惯性指数。另外，黎曼度规与欧式度规、洛伦兹度规与闵氏度规在属于上往往不做区分。

例 2.7. 对于流形 \mathbb{R}^n ，考虑其自然坐标 $\{x^\mu\}$ 。我们给 M 赋予一个欧式度规，这个度规在这组坐标下的矩阵处处为单位矩阵，即度规为 $\delta = \delta_{\mu\nu} dx^\mu \otimes dx^\nu$ 。此时的流形称为欧氏空间，是最简单的、平直的黎曼空间。

现在我们取另外一个坐标系，如果度规在这组坐标系下的分量也为 $\delta_{\mu\nu}$ ，也就是说基处处是正交归一的，那么我们称这个坐标系为笛卡尔坐标系。笛卡尔坐标系下的线元就是我们熟悉的线元

$$ds^2 = \delta_{\mu\nu} dx^\mu dx^\nu = \sum_{\mu=1}^n (dx^\mu)^2$$

笛卡尔坐标系之间的变换称为正交变换，也就是线性代数里学的正交矩阵，也就是保持线元长度不变的线性变换。线性变换 Q 是正交变换的充要条件是 $Q^\top I Q = I$ ，用分量方式写作

$$Q^\mu{}_\sigma Q^\nu{}_\rho \delta_{\mu\nu} = \delta_{\sigma\rho}$$

而 Q 下的坐标变换 $x' = Qx$ 写成分量形式即为

$$x'^{\mu} = Q^{\mu}_{\nu} x^{\nu}$$

例 2.8. 对于流形 \mathbb{R}^n 和其自然坐标 $\{x^{\mu}\}$ ，我们给它赋予一个处处相等的闵氏度规 $\eta = \eta_{\mu\nu} dx^{\mu} \otimes dx^{\nu}$ ，此时的流形称为闵氏时空，也就是平直的洛伦兹时空。洛伦兹时空常常也被记作 $\mathbb{R}^{1,n-1}$ 或 $\mathbb{R}^{1+(n-1)}$ 。

为了突出第一个维度的特殊性，我们把第一个维度的指标记作 0，它实际上在相对论中表示时间维度；其他指标记作 1 到 $n-1$ ，它们表示空间维度。如果我们选取另外一组坐标系，使得这组坐标系下的度规分量仍然为 $\eta_{\mu\nu}$ ，那么我们称这组坐标系为洛伦兹坐标系。洛伦兹坐标系下的线元为

$$\begin{aligned} ds^2 &= \eta_{\mu\nu} dx^{\mu} dx^{\nu} = -(dx^0)^2 + (dx^1)^2 + \cdots + (dx^{n-1})^2 \\ &= -dt^2 + (dx^1)^2 + \cdots + (dx^{n-1})^2 \end{aligned}$$

洛伦兹坐标系之间的变换称为 **洛伦兹变换**，也就是保持闵氏线元不变的线性变换，记作 Λ ，它就是狭义相对论中的时空变换。可以看出，洛伦兹变换相当于“闵氏时空中的正交变换”，它满足 $\Lambda^{\top} \eta \Lambda = \eta$ ，或用分量方式写作

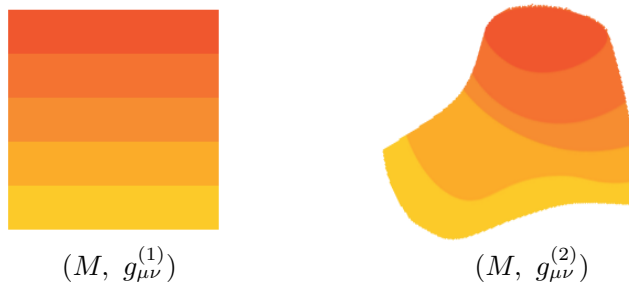
$$\Lambda^{\mu}_{\sigma} \Lambda^{\nu}_{\rho} \eta_{\mu\nu} = \eta_{\sigma\rho}$$

而洛伦兹坐标变换 $x' = \Lambda x$ 写成分量形式即为

$$x'^{\mu} = \Lambda^{\mu}_{\nu} x^{\nu}$$

度规是流形上的一种附加结构，也就是说它进一步给定了流形的一些几何特征。流形上只有拓扑结构和微分结构，前者告诉你流形是封闭的还是开放的、有几个洞等，后者让相邻点之间的“位置关系足够平滑”。然而，此时这个流形仍然是“软的”，也就是说相邻点之间的位置关系没有那么的确定，像一个橡胶膜，可以随意拉伸。而度规则一定程度上确定了点之间的位置关系。我们下面来具体来解释一下。

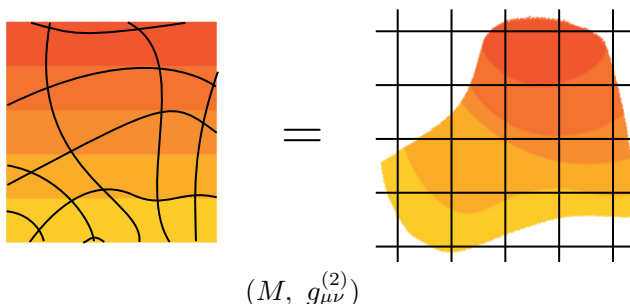
我们知道，度规确定了切矢量的长度和角度。而一个很短的切矢量可以被直观地认为成“指向临近的一个点”。因此，一点处的度规就相当于决定了附近的点相对于该点的“距离和角度”，从而使“软的”流形“硬了起来”。而度规场在流形的每一点都赋予了这样的结构，从而最终赋予流形一个“硬的”几何结构。例如，下图就显示了同一个微分流形在赋予了不同的度规后的几何结构。



我们还可以从另一个等价的角度来理解。某点的度规可以在该点的切空间中选择一些特殊的基矢量，即标准正交基。在流形上，这些作为基矢量的切矢量可以“延伸”成一个坐标轴，从而构成该点附近的一个“标准直角坐标系”，称为 **黎曼法坐标系**，又称为

这种由切矢量延伸成线的方式称为测地延伸，我们在 2.2.3. 中会看到对这种线的定义。

测地坐标系. 度规的存在, 就相当于在每一点附近都画出了一个坐标系, 把它认定为黎曼法坐标系, 也就是把画出的坐标轴“摆成标准直角坐标系”. 这样, 流形上也被赋予了硬的几何结构. 例如下图中, 我们在流形上画出了一个坐标系, 把它“摆正”, 就等价于赋予了上图中右图的度规.



$(M, g_{\mu\nu}^{(2)})$

2.2.2 Levi-Civita 联络

流形上有了硬的几何结构后, 我们就可以开始建立微积分了. 考虑我们在多元微积分里学过的 \mathbb{R}^n 上的微分算符 ∇ , 它可以用于计算标量场 f 的梯度 $\nabla_\mu f$ 或矢量场 v^μ 的散度 $\nabla_\mu v^\mu$. 其分量定义为 $\nabla_\mu = \partial/\partial x^\mu$. 在流形上, 我们可以将其推广: 在选定了一个坐标系 $\{x^\mu\}$ 后, 我们可以定义一个导数算符 $\partial_\mu: \mathcal{F}(p, q) \rightarrow \mathcal{F}(p, q+1)$, 使张量在该坐标系下的分量满足

$$\partial_\sigma T^{\mu\dots\nu\dots} = \frac{T^{\mu\dots\nu\dots}}{\partial x^\sigma}$$

我们称 ∂_μ 为 该坐标系下的偏导数算符. 偏导数算符是一个很自然的推广, 但它有一个很大的问题——它取决于我们选取的坐标系. 也就是说如果我们选择不同的坐标系, 那我们定义出的偏导数算符也是不同的. 然而我们想要研究的是流形上的几何性质, 几何性质应该是与坐标系无关的, 坐标系的作用只是方便计算. 那么如何定义一个与坐标系无关的导数算符呢? 考虑到我们以前微积分学习的 ∇ 的分量形式实际上是在标准直角坐标系下的分量, 我们在此也可以使用流形上的“直角坐标系”, 即黎曼法坐标系下定义一个导数算符, 这就是流形上的联络.

我们给导数算符协变指标是因为如果坐标轴伸长, 则移动单位长度后的函数值变化量也会相应变大, 即导数是协变的.

定义 2.2: Levi-Civita 联络

设 $(M, g_{\mu\nu})$ 是一个带度规的流形, 则每一点的黎曼法坐标系下的偏导数算符称为 $g_{\mu\nu}$ 诱导出的 **Levi-Civita 联络**, 简称联络, 记作 ∇_μ . 即对 $p \in M$, 有

$$(\nabla_\sigma T^{\mu\dots\nu\dots})|_p = \left(\frac{T^{\mu\dots\nu\dots}}{\partial x^\sigma} \right) \Big|_p$$

其中 $\{x^\mu\}$ 为 p 点的黎曼法坐标系.

这里不是很严谨, 因为黎曼法坐标系的建立需要用到联络. 这里这么定义是为了几何直观.

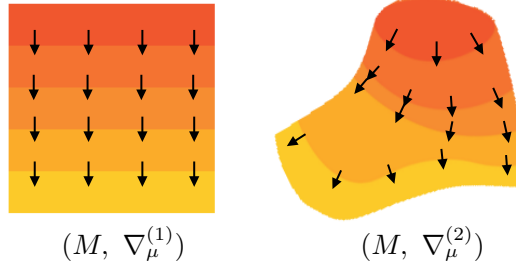
我们再来解释一下定义中“每一点”的意思——对于给定的一点 p , p 点的 Levi-Civita 联络是 p 点的黎曼法坐标系在 p 点的普通导数算符. 但不同点的黎曼法坐标系是可能不同的. 因此对于旁边的一点 q , 它可能在 p 点的黎曼法坐标系的范围内, 但 q 的联络不能是用 p 点建立的黎曼法坐标系导出的偏导数算符. 这点在我们后面讲曲率时是非常重要的.

我们对 Levi-Civita 联络的定义中虽然借用了坐标系, 但由于这个坐标系是被度规确定的, 而不是我们选择的, 我们定义出来的联络也是确定的. 因此, 与度规类似, 联

一个点处的黎曼法坐标系有多种选择方式, 但它们定义出的联络是相同的. 联络的唯一性称为黎曼几何基本定理.

络也是一个与坐标系无关的、体现流形几何性质的数学对象。

下面，我们用上一小节解释度规相同的例子，来看一下联络的几何意义。我们知道，度规是选择了流形上的一组坐标系——基矢量为标准正交基的坐标系——作为黎曼法坐标系。而联络作为度规诱导出的结构，自然承载着类似的信息。联络同样也是认定了流形上的一组特殊的坐标系，也就是度规确定的黎曼法坐标系，联络自身就是这个坐标系下的偏导数算符。为了表示出联络所体现出的几何结构的差异，我们把下图中的颜色看作一个标量函数 f 的热图。我们在下图中还画出了这个标量函数的梯度 $\nabla_\mu f$ 。可以发现，左图中这些对偶矢量是“平行的”，而右图中则不再平行。这体现了两个流形上联络所给出的几何结构的不同。



对于张量场 $T^{\mu\cdots\nu\cdots}$ ， $\nabla_\sigma T^{\mu\cdots\nu\cdots}$ 记录了它在流形上沿所有方向变化的信息。但有时，我们只需求它沿着给定矢量 v^σ 的变化大小，也就是其“方向导数”。在微积分中我们学过，函数沿一个矢量的方向导数就是其梯度与该矢量的内积。在这里也一样，我们称

许多书也会把协变导数记作 $\nabla_v T^{\mu\cdots\nu\cdots}$ 。

$$v^\sigma \nabla_\sigma T^{\mu\cdots\nu\cdots}$$

为 $T^{\mu\cdots\nu\cdots}$ 沿 v^σ 的协变导数。因此， ∇_μ 又称为协变导数算符。

Levi-Civita 联络与我们熟悉的导数运算类似，有一些很好的性质。同时，它作为由度规 $g_{\mu\nu}$ 诱导出的算符，自然与度规也有联系。我们把它的性质列举如下。这些性质都很容易理解。

定理 2.1: Levi-Civita 联络的性质

对于带有度规的流形 $(M, g_{\mu\nu})$ ， ∇_σ 是度规诱导出的 Levi-Civita 联络，有

1. 线性性：

$$\nabla_\lambda (a T^{\mu\cdots\nu\cdots} + b S^{\mu\cdots\nu\cdots}) = a \nabla_\lambda T^{\mu\cdots\nu\cdots} + b \nabla_\lambda S^{\mu\cdots\nu\cdots}$$

2. 莱布尼兹律：

$$\nabla_\lambda (T^{\mu\cdots\nu\cdots} S^{\sigma\cdots\rho\cdots}) = T^{\mu\cdots\nu\cdots} \nabla_\lambda S^{\sigma\cdots\rho\cdots} + S^{\sigma\cdots\rho\cdots} \nabla_\lambda T^{\mu\cdots\nu\cdots}$$

3. 与缩并满足交换律：

$$\nabla_\sigma (T^{\cdots\mu\cdots\cdots\mu\cdots}) = (\nabla_\sigma T)^{\cdots\mu\cdots\cdots\mu\cdots}$$

4. 与度规的相容性： $\nabla_\sigma g_{\mu\nu} = 0$

事实上，第 4 条性质是一个偏导数算符作为 Levi-Civita 联络的充要条件——如果某个坐标系下的偏导数算符 ∂_σ 满足 $\partial_\sigma g_{\mu\nu} = 0$ ，那么这个 ∂_σ 实际上就是该流形上度规诱导出的 Levi-Civita 联络 ∇_σ 。或者说，该坐标系就是一个黎曼法坐标系。

Levi-Civita 联络作为一个算符，它虽然并不是一个张量，但也与坐标系的选取无关。对于随坐标系不变的张量，我们在选取了一组基以后总可以把张量表示为一个随坐标系变化的数组。对于联络也相同，在选定了坐标系后，我们考虑它作用在基矢量上，通过指标平衡可以发现这个作用相当于一个 $(1, 2)$ -型张量，这个张量称为联络在该坐标系下的 Christoffel 符号。

有些地方也会把 Christoffel 符号称为联络。

定义 2.3: Christoffel 符号

对于一个带有联络的流形 (M, ∇_μ) ，在选定一个坐标系 $\{x^\mu\}$ 后，我们定义张量场 $\Gamma^\sigma_{\mu\nu}$ 满足

$$\nabla_\mu \frac{\partial}{\partial x^\nu} = \Gamma^\sigma_{\mu\nu} \frac{\partial}{\partial x^\sigma}$$

我们称 $\Gamma^\sigma_{\mu\nu}$ 为联络 ∇_μ 在该坐标系下的 Christoffel 符号。

在此说明很重要的一点：许多书都会强调 Christoffel 符号不是张量，它在坐标系变换下不满足张量的变换规律。这是因为换成另一个坐标系后，其基矢量 $\partial/\partial x^\mu$ 也发生了变化，因此由上式定义出来的 Christoffel 符号自然就是另一个不同的张量了。但也正因如此，我们也可以说 Christoffel 符号是一个与随坐标系变化的张量，也就是说每个坐标系都有属于自己的一个 Christoffel 符号。这两种理解实际上是等价的。

既然度规唯一地确定了 Levi-Civita 联络，而联络又在给定坐标系下可以唯一地写成 Christoffel 符号。那么在给定坐标系后，我们就一定可以通过度规的分量求出 Christoffel 符号。我们给出下面的计算公式。

定理 2.2: 由度规求 Christoffel 符号

对于带度规的流形 $(M, g_{\mu\nu})$ ，在选定一个坐标系 $\{x^\mu\}$ 后，设 ∂_μ 是该坐标系下的偏导数算符， $\Gamma^\sigma_{\mu\nu}$ 是该坐标系下的 Christoffel 符号，则有

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

介绍完 Levi-Civita 联络和 Christoffel 符号的数学定义，我们下面来直观地看一下两个简单的例子。

例 2.9. 对于欧氏空间 $(\mathbb{R}^n, \delta_{\mu\nu})$ ，其黎曼法坐标系就是我们熟悉的标准直角坐标系，因此其 Levi-Civita 联络就是标准直角坐标系下的偏导数算符。由于在欧氏空间中我们常常默认选择直角坐标系，因我们常常把仿射联络就写成偏导数算符 ∂_μ 。同理，闵氏时空 $(\mathbb{R}^n, \eta_{\mu\nu})$ 的黎曼法坐标系为洛伦兹坐标系，也是我们默认会选取的坐标系，因此其联络也会被写作 ∂_μ 。用上面的定理 2.2. 可以求得，上面两个空间在直角坐标系和洛伦兹坐标系下的 Christoffel 符号 $\Gamma^\sigma_{\mu\nu} = 0$ 。但是注意，如果我们换成了其他的坐标系，那么这两个空间的联络就不一定是该坐标系下的偏导数算符了，Christoffel 符号也不一定为零了。

有了 Christoffel 符号，我们就可以计算流形上的任何的张量场的导数。不过这种求导并不是 Christoffel 符号的直接作用。实际上很容易发现，一个坐标系的偏导数算符在该坐标系下的 Christoffel 符号为零，因此 Christoffel 符号表示的是联络与偏导数算符的差异，也就是 $\nabla_\sigma T^{\mu\cdots\nu\cdots}$ 相对于 $\partial_\sigma T^{\mu\cdots\nu\cdots}$ 的修正。我们下面给出用 Christoffel 符号求导的公式。

定理 2.3: 联络的作用与 Christoffel 符号的关系

设 (M, ∇_μ) 是一个带有联络的流形, 在选定坐标系 $\{x^\mu\}$ 后, $\Gamma^\sigma_{\mu\nu}$ 是该坐标系下的 Christoffel 符号, ∂_μ 是该坐标系下的偏导数算符. 则对切矢量场 v^μ 有

$$\nabla_\mu v^\nu = \partial_\mu v^\nu + \Gamma^\nu_{\mu\sigma} v^\sigma$$

对于余切矢量场 ω_μ 有

$$\nabla_\mu \omega_\nu = \partial_\mu \omega_\nu - \Gamma^\sigma_{\mu\nu} \omega_\sigma$$

而对于一个任意阶张量 $T^{\mu\dots\nu\dots}$, 我们只需要对每个逆变指标按照切矢量的方式求出 ∇_μ 与 ∂_μ 的差异项, 对每个协变指标按照余切矢量的方式求出差异项, 并按照符号相加即可.

例 2.10. 我们来举例解释下对任意张量的求导法则. 例如对于一个 $(1,2)$ -型张量场 $T^\alpha_{\beta\gamma}$, 联络作用在它身上, 首先会产生偏导数的一项 $\partial_\mu T^\alpha_{\beta\gamma}$. 在此之上需要加入由 Christoffel 符号修正的项. 我们先考虑其协变指标 α . 我们把这个指标完全按照矢量场的求导法则, 把它挪到 Γ 上, 并把原来的 α 的位置和 Γ 的第二个逆变指标缩并, 得到一项 $\Gamma^\alpha_{\mu\nu} T^\nu_{\beta\gamma}$. 同理, 对于两个协变指标, 每个指标都会产生一项负的修正. 对于其中的每个指标, 我们需要把它挪到 Γ 的第二个协变指标上, 并把原来的这个指标和 Γ 的逆变指标缩并. 最终的结果为

$$\nabla_\mu T^\alpha_{\beta\gamma} = \partial_\mu T^\alpha_{\beta\gamma} + \Gamma^\alpha_{\mu\nu} T^\nu_{\beta\gamma} - \Gamma^\nu_{\mu\beta} T^\alpha_{\nu\gamma} - \Gamma^\nu_{\mu\gamma} T^\alpha_{\beta\nu}$$

2.2.3 测地线

联络将流形上相邻点的切空间联系起来, 它一个重要的用途就是将张量在各个点之间平移. 对于一个矢量场 v^μ , 它沿着矢量 r^ν 的协变导数为 $r^\nu \nabla_\nu v^\mu$. 若该导数为零, 则表示矢量场 v^μ 沿着 r^ν 的方向平移 (无限小的距离) 后不变, 这就相当于定义了如何把一个矢量 v^μ 沿着给定方向 r^ν 平移. 注意, 矢量的平移一定要说清沿着哪条曲线平移, 因为在任意的流形上, 同一个矢量沿着不同曲线平移到同一个点可能得到不同的结果, 我们在下面的曲率这一小节中会专门讨论这一点.

有了平移的概念, 我们就可以定义流形上的一个非常重要的概念——测地线. 它是对直线概念的推广.

定义 2.4: 测地线

若流形 (M, ∇_μ) 上的一条参数曲线 $C(t)$ 的切矢量 T^μ 沿自身平移不变, 即

$$T^\nu \nabla_\nu T^\mu = 0$$

则称曲线 $C(t)$ 是流形上的一条 **测地线**.

测地线的几何意义非常直观——它的切矢量的方向不变, 自然就是指流形这种弯曲空间上的“直线”. 而方程 $T^\nu \nabla_\nu T^\mu = 0$ 称为 **测地线方程**. 在选定了坐标系后, 我们可以把协变导数用偏导数算符和 Christoffel 符号写出, 并把 $T^\mu = dx^\mu/dt$ 带入, 得到它的分量形式. 这个分量形式非常重要, 我们在此写出.

定理 2.4: 测地线方程

流形 (M, ∇_μ) 上的测地线方程在给定任意坐标系下的分量形式为

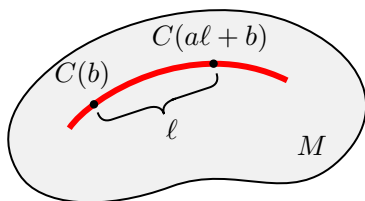
$$\frac{d^2 x^\sigma}{dt^2} + \Gamma^\sigma_{\mu\nu} \frac{dx^\mu}{dt} \frac{dx^\nu}{dt} = 0$$

例 2.11. 对于流形 M ，如果我们选择黎曼法坐标系，则联络 ∇_μ 退化为偏导数算符 ∂_μ ，进而有 $\Gamma^\sigma_{\mu\nu} = 0$ 。因此测地线方程退化为

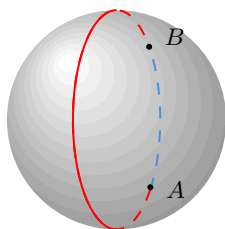
$$\frac{d^2 x^\mu}{dt^2} = 0$$

此方程的解为一次曲线 $x^\mu = a^\mu t + b^\mu$ ，即在黎曼法坐标系下测地线的形式就是我们熟悉的直线。这也说明，测地线方程的第二项是通过 Christoffel 符号描述在该坐标系下测地线对一次曲线的偏离。

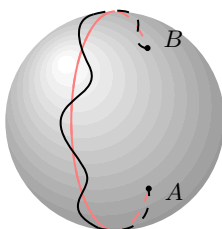
在此需要指出的是，我们的曲线是参数曲线。一条测地线做另一种参数化以后它的切矢量的长度就发生了变化，就不一定还是测地线了。但很容易发现，一条曲线如果可以成为测地线的话，使其成为测地线的参数（称为 **仿射参数**）一定是线长 ℓ 的一次函数（仿射变换） $a\ell + b$ ，如下图所示。



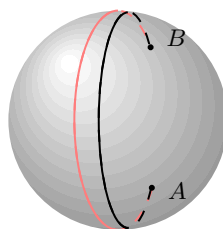
在流形上有了相容的度规和联络后，我们就可以讨论测地线的长度问题了。我们知道，在欧氏空间中，直线是连接两点最短的曲线。在一般的流形上，这一点并不一定成立。但是我们可以发现，在黎曼空间和洛伦兹时空中，测地线一定是使线长取极值的曲线。注意这个极值可以是极大值也可以是极小值。而且和我们熟悉的微积分不同，这个极值并不意味着它附近没有更长/更短的线了。以二维球面为例，过球面上两个点 A 和 B 的测地线是大圆弧，共有两条。下面我们考虑长的那一条。我们说这条线的线长取极大值。但是显然他附近有比它更长的线——只要走的时候左右绕一绕即可。同时，它附近也有比它更短的线——只要偏移一点，不绕到球面的正对侧即可。极值的严格定义是需要变分的知识的，我们在此不做具体讨论。



测地线



更长的线



更短的线

在上面的例子中，我们看到流形上两点之间可以有不只一条测地线，这使得测地线和平直空间中的直线的性质有所不同。但直线的有一个唯一性——一个矢量确定唯一一条直线——在测地线中仍然保持。这构成了我们以后讨论测地线的一个基础。我们在此把它列为一个重要的定理。

定理 2.5: 测地延伸的唯一性

给定流形上一点 p 和该点处的一个切矢量 v^μ ，存在唯一的测地线 $C(t)$ 使得其在 p 点的切矢量为 v^μ ，且有 $C(0) = p$ 。

2.2.4 黎曼曲率张量

在本节的最后，我们来学习与描述微分几何中直接用于描述流形弯曲的量——曲率。我们先直接给出黎曼曲率的数学定义，后面再对它的几何意义做出解释。

定义 2.5: Riemann 曲率张量

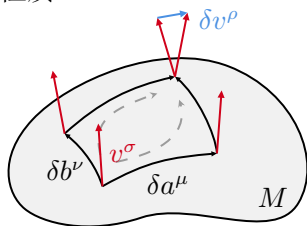
流形 (M, ∇_μ) 上的 **黎曼曲率张量** 场是一个 $(1, 3)$ -型张量场 $R^\rho_{\sigma\mu\nu}$ ，定义为：若 p 处的矢量 v^σ 先按照无限小矢量 δa^μ 再按 δb^ν 平移，与先按 δb^ν 再按 δa^μ 平移相比，二者的差异为 δv^ρ ，则 p 处的黎曼曲率张量 $R^\rho_{\sigma\mu\nu}$ 是一个描述它们之间线性关系的张量

$$\delta v^\rho = R^\rho_{\sigma\mu\nu} \delta a^\mu \delta b^\nu v^\sigma$$

也有一些地方会把它写作 $R_{\mu\nu\sigma}{}^\rho$ ，还有许多数学书会把它写作

$$R(\delta a, \delta b)v = \delta v$$

我们用下图直观地显示出以上定义中各个量的几何关系。回顾我们上一节提到，流形上的平移需要指出是沿着哪条线平移，这是因为沿着不同曲线的平移的结果是不一样的。而黎曼曲率正是描述曲面的这种性质。



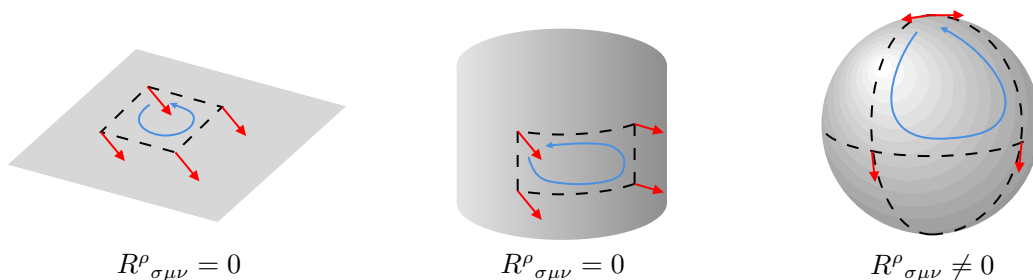
下面我们以几个平直与弯曲流形的例子作为对比，来讨论一下黎曼曲率张量的几何意义。不过注意，我们在本章的最开始就强调了，现代微分几何是流形的内蕴几何，也就是说我们只需要在流形本身内部研究其性质，而不需要把流形嵌入到一个更高维的空间中来看它的形状。但在本节，为了对曲率有一个直观的认识，我们还是以大家熟悉的嵌入在三维空间中的曲面为例，介绍黎曼曲率的几何意义。但注意这些曲面的曲率完全也可以以内蕴的方式体现。

从黎曼曲率张量的定义可以看出，如果流形上的任意一个切矢量沿着任意一个闭合路径平移，回到原处后与平移之前完全相等，则该流形的黎曼曲率张量为零，即该流形是平直的。反之，若平移后会发生变化，则该流形的曲率张量不为零，即该流形是弯曲的。

我们首先来看一个平面。在平面上，矢量的平移显然与路径无关，所以延闭合路径平移回原处后矢量不变，即平面是平直的。下面我们再来看圆柱面。圆柱面上的平移或许稍稍有些不容易想象，但如果我们把闭合曲线取成一个两边平行于长轴的矩形，那么也可以很容易发现切矢量平移回原处后不变。实际上，在柱面上的切矢量沿任何闭合曲线平移都不会变化，也就是说从黎曼曲率的角度看，柱面也是平直空间。

或许你已经开始疑惑柱面为什么是平直的了。但在解释这个之前，我们再来看看一个弯曲例子——球面。我们想象在球的北极处有一个切矢量。现在我们把这个切矢量沿着它自己的方向平移到赤道上，显然这个矢量现在指向南方。接下来，我们把它沿着赤道平移一段，这个过程中这个矢量显然一直保持向南。最后，我们把它沿着所在的经线平移回北

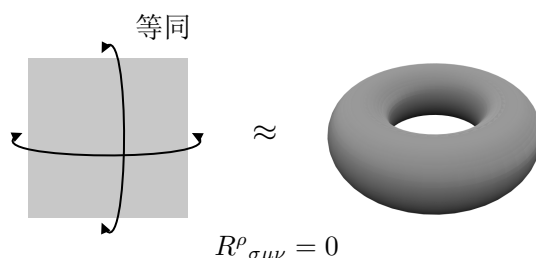
极。这时我们会发现，这个矢量的方向与原来不再相同了！也就是说，球面的确是一个弯曲的空间。



那么柱面的“弯曲”和球面的“弯曲”有什么区别呢？答案是，柱面可以展平，而球面不能。想象一个由纸做的曲面——纸可以弯曲，但不像橡皮膜一样可以被拉伸。那么柱面在剪开以后可以被铺平，但是球面却永远不能。这种“可以弯曲但不能拉伸”的变换可以被更严谨地描述为“不改变曲面上距离”的变换，因而在古典微分几何中被称为等距变换。那么可以说，黎曼曲率张量在等距变换下保持不变，或者说黎曼曲率张量不关心这个曲面在等距的情况下被弯成了什么样子，因为它不影响的曲面内蕴几何。想象有一个二维生物生活在我们的曲面上，它可以在曲面的局部做各种内蕴几何，那么黎曼曲率张量就是它能测量出的曲面弯曲的所有信息。它可以区分它是生活在平面还是球面上，但却无法区分它是生活在平面上还是柱面上。也就是说我们可以把它所生活的曲面进行等距弯曲，但是它永远无法感受到任何变化！这个例子说明了，黎曼曲率张量描述的是流形内蕴的弯曲特征。

显然它可以通过在柱面上走一圈的方式发现自己生活在柱面上而非平面上，但这属于全局拓扑特征，无法在局部发现。

为了进一步理解“内蕴”的含义，让我们来看这样一种流形：我们把一张平直的白纸的上下两端看作等同，左右两端看作等同。也就是说，如果有一个二维生物生活在这个空间上，那么它如果走出了上边界，它就会自动从下边界中走回纸面。很容易发现，这个流形是拓扑等价于环面的。但是，一张纸是无法通过在等距的情况下被弯折成环面的，所以我们无法把它画成一个嵌在三维空间里的二维环面。但这个流形在数学上又是确确实实可以存在的，而且它的黎曼曲率为零。生活在其上的二维生物，如果它不能绕着环跑一圈的话，它是无从得知自己生活的世界不是平面的。也就是说，这是一个平直的环境。同时，这个例子也说明，流形的弯曲作为一个局部性质，与其全局的拓扑结构是两种不同的性质。



当然，我们在这里只是用曲率举一个例子，来解释“内蕴”的含义。实际上我们学习的整个微分几何都是流形的内蕴几何。我们定义的度规、联络等都在等距变换下保持不变。这也很容易理解——等距变换保持距离不变，自然就保持度规不变。而 Levi-Civita 联络又由度规完全决定，因此联络也保持不变。实际上，我们未来学的一切概念都是在内蕴几何框架内的，它们都不受等距变换的影响。

黎曼曲率既然由流形上的矢量平移所定义，自然就被联络所唯一决定。而在给定坐标

系后，联络可以被写成 Christoffel 符号。那么在给定的坐标系下，黎曼曲率张量的分量该如何计算呢？我们在此给出以下公式。

定理 2.6: 用 Christoffel 符号计算黎曼曲率张量

对于带有联络的流形 (M, ∇_μ) ，在选定了一个坐标系后，设 ∂_μ 是该坐标系下的偏导数算符， $\Gamma^\sigma_{\mu\nu}$ 是该坐标系下的 Christoffel 符号，则黎曼曲率张量在该坐标系下的分量为

$$R^\rho_{\sigma\mu\nu} = \partial_\mu \Gamma^\rho_{\nu\sigma} - \partial_\nu \Gamma^\rho_{\mu\sigma} + \Gamma^\rho_{\mu\lambda} \Gamma^\lambda_{\nu\sigma} - \Gamma^\rho_{\nu\lambda} \Gamma^\lambda_{\mu\sigma}$$

对于 Christoffel 符号，我们可以通过选择黎曼法坐标系的方式使其化零。但如果时空不是平直的，那么黎曼曲率张量在任何坐标系下都不可能分量全为零。这是因为它涉及了 Christoffel 符号的导数。即使我们选择了某点的黎曼法坐标系，我们也只能保证 Christoffel 符号在该点处为零，但在附近的点却不一定为零。因此，我们得到的黎曼曲率张量的分量也不一定为零。这使得黎曼曲率张量是真正描述流形弯曲的量。

实际上，一个张量如果在某个坐标系下分量全为零，则该张量在任何坐标系下都只能分量全为零。

例 2.12. 欧氏空间和闵氏时空中，矢量的平移都与路径无关，因此它们的曲率都为零，都是平直空间。这也可以通过我们刚学习的其他方式来解释。在这两个空间中，我们熟悉的标准直角坐标系或洛伦兹坐标系对于任意一点都是黎曼法坐标系，因此我们可以在这个坐标系下求导。在这样的坐标系下，Christoffel 符号处处为零，用上面的公式也可以算出黎曼曲率张量为零。因此我们说欧氏空间和闵氏时空都是平直的。

黎曼曲率张量还有一种常用的表示方式是写成 (0,4)-型张量，即 $R_{\rho\sigma\mu\nu} = g_{\rho\lambda} R^\lambda_{\sigma\mu\nu}$ 。在这种写法下，我们可以证明 $R_{\rho\sigma\mu\nu}$ 具有高度的对称性。我们在这里不赘述大部分的对称性，但需要指出一个很重要的等式，它与广义相对论的方程密切相关。

定理 2.7: Bianchi 恒等式

设 (M, ∇_μ) 是一个带有联络的流形， $R_{\rho\sigma\mu\nu}$ 是其黎曼曲率张量，则有

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

最后，我们来学习两个由黎曼曲率导出的张量。它们也描述流形的弯曲性质，但描述的视角和黎曼曲率张量有所不同。

定义 2.6: Ricci 曲率张量与标量曲率

流形 (M, ∇_μ) 上的 **Ricci 曲率张量** 场 $R_{\mu\nu}$ 即为黎曼曲率的缩并，即

$$R_{\mu\nu} = R^\rho_{\mu\rho\nu}$$

而 **标量曲率** 又称为 **Ricci 标量** 场 R 则是 Ricci 张量的缩并，即

$$R = g^{\mu\nu} R_{\mu\nu}$$

黎曼曲率张量有多种缩并方式，但由于其对称性，导致只有这一种独立的结果。其他的缩并方式得到的要不是零张量，要不然是负的 Ricci 张量。

Ricci 曲率和标量曲率的几何意义并不那么明显，我们在此试图给出一个尽量几何的解释。黎曼曲率张量包含了流形弯曲的全部信息，而 Ricci 曲率则只包含了流形弯曲对于体积的影响。更精确一点地说，它表示了该点附近体积差异的低阶近似。在测地坐标系

下, 坐标为 x^μ 的点处的“微小体积”与平直空间的微小体积相比有

$$\frac{V_g}{V_{\text{flat}}} = 1 - \frac{1}{6} R_{\mu\nu} x^\mu x^\nu + o(|x|^3)$$

而标量曲率则进一步忽视掉了方向差异, 它比较的是该点处的微小体积和平直空间的差异. 若在该点处作半径为 r 的超球体, 则有

$$\frac{V_g}{V_{\text{flat}}} = 1 - \frac{R}{6(n+2)} r^2 + o(r^4)$$

其中 n 为流形的维数.

注意这几个式子不再是张量等式, 而是在测地坐标系下的分量等式.

2.3 李导数与 Killing 矢量场*

在微分几何这一章的最后, 我们来学习一个稍微进阶的内容——Killing 矢量场. 这一部分内容并不直接出现在广义相对论的方程中, 但是它仍然与广义相对论中时空的性质有着密切的联系.

2.3.1 拉回与推前*

在这一章, 我们开始考虑流形 M 上的变换 $\phi: M \rightarrow M$. 我们在这里考虑的变换都满足足够好的性质: (1) 变换 ϕ 是一个一一映射; (2) ϕ 和 ϕ^{-1} 都是光滑的. 满足这两条性质的变换 ϕ 称为 M 上的一个 **微分同胚**.

微分同胚 ϕ 把流形上的点 p 映为另外一个点 $\phi(p)$. 如果我们的流形上有张量场, 我们希望 p 点的张量可以被 ϕ 一起“带过去”, 或者把像点 $\phi(p)$ 的张量反向“寄回来”. 这样, 我们就可以把两个不同点的张量映到同一个点上, 使得我们可以做比较或者计算. 我们先从简单的低阶张量场开始, 定义 ϕ 诱导出的这种对张量场的映射.

定义 2.7: 标量场的拉回与矢量场的推前

设 M 是一个微分流形, 映射 $\phi: M \rightarrow M$ 是微分同胚, 则 ϕ 诱导出的 **拉回** 映射 ϕ^* 把 M 上的标量场 f 映为另一个标量场, 满足

$$(\phi^* f)|_p = f|_{\phi(p)}$$

而 ϕ 诱导出的 **推前** 映射 ϕ_* 把 M 上的矢量场 v 映为另一个矢量场, 满足对任意标量场 f 有

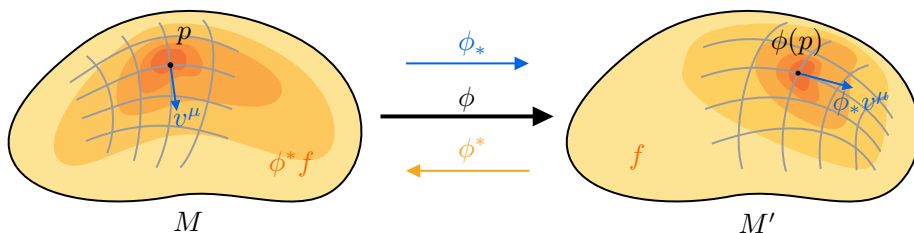
$$(\phi_* v)|_{\phi(p)}(f) = v|_p(\phi^* f)$$

$T|_p$ 表示张量场 T 在点 p 的张量.

这个定义看起来可能有些抽象, 但实际上非常直接, 无论从数学上还是从几何直观上. 为了方便理解, 我们可以把 ϕ 看作两个流形之间的映射 (即从 M 到一个和它一样的流形 M' 的映射). 现在 M' 上有一个标量场, 而 M 上没有. ϕ 对标量场 f 的拉回就是把 M' 上像点 $\phi(p)$ 的标量场值赋予 M 上的 p 点. 这样, 这个操作就是把 M' 上整个标量场的位置按照 ϕ 的逆方向拉回了 M 上.

下面我们来看矢量场的推前的定义. 回顾一下 $v(f)$ 的意思是 f 沿着 v 的方向导数. 推前映射 ϕ_* 把 M 上的矢量场 v 跟着 ϕ “拷贝”到了 M' 上, 使得在 M' 上 f 沿着 $\phi_* v$ 的方向导数与 M 上 $\phi^* f$ 沿着 v 的方向导数相同. 用方向导数来定义或许还是有些不直观, 但这可以解释为, 如果在 p 旁边有一点 q , 那么 M 上“从 p 指向 q ”的矢量会被映

射为 M' 上“从 $\phi(p)$ 指向 $\phi(q)$ ”的矢量。这样看，我们定义的推前映射也的确是很自然的。



有了矢量场的推前，我们就可以来定义我们在这章最为关心张量场的变换了——度规张量场的拉回。我们仍然是先给出数学定义。

定义 2.8: 度规张量的拉回

设 M 是一个微分流形，上面有度规场 $g_{\mu\nu}$ 。映射 $\phi: M \rightarrow M'$ 是微分同胚，则度规的拉回 ϕ^* 把 M 上的度规场映为另一个度规场，它对任意矢量场满足

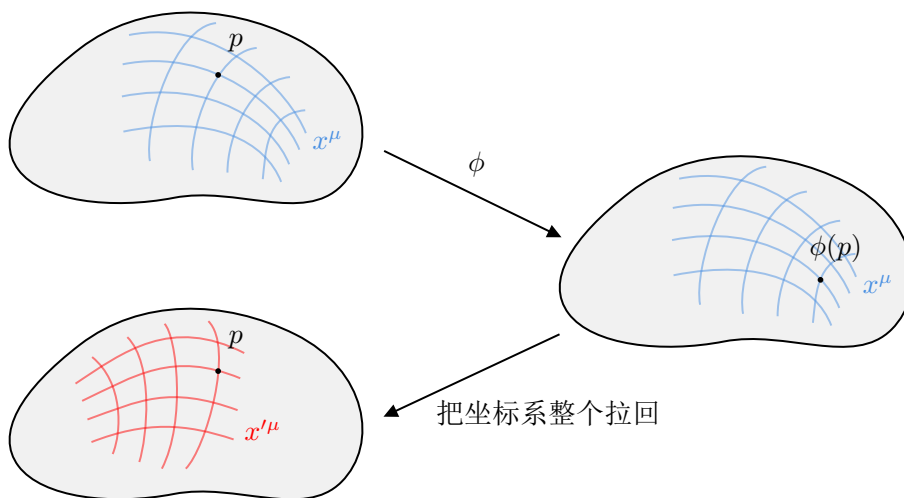
$$(\phi^*g)_{\mu\nu}u^\mu v^\nu = g_{\mu\nu}(\phi_*u)^\mu(\phi_*v)^\nu$$

其中等式左侧是在 p 点的值，等式右侧是在 $\phi(p)$ 点的值。

拉回和推前实际上可以被定义在任意一个张量场上，其定义思路也与这个定义相同。但其他张量场的拉回和推前对我们并不重要，我们在这里就不做讨论了。

我们在解释几何意义时仍然把 ϕ 看作 $M \rightarrow M'$ 的映射，其中 M' 上有度规场 $g_{\mu\nu}$ 。那么我们的定义式的意思是， $g_{\mu\nu}$ 被拉回到 M 上成为 $(\phi^*g)_{\mu\nu}$ ，它满足：在 M 上的两个矢量 u^μ 和 v^ν 用 $(\phi^*g)_{\mu\nu}$ 做内积，与把这两个矢量推前到 M' 上用 $g_{\mu\nu}$ 做内积，得到的结果相同。也就是说，度规张量的拉回配合着矢量场的推前，保持内积相同。这样看，这个定义也非常自然。

最后，我们回过头来再来看一看流形上的微分同胚。我们刚才一直都把微分同胚看作 M 上的点变换，即它把一个点映为另一个点。但微分同胚还可以有另一种理解方式：微分同胚把流形上的一个网格（坐标系） $\{x^\mu\}$ 映为另一个网格 $\{x'^\mu\}$ 。也就是说，在给定了一个坐标系以后，微分同胚会诱导出一个坐标变换 $x^\mu \mapsto x'^\mu$ 。



把坐标系“拉回来”和把点映过去，这两种看法下点与坐标系的位置关系应该是一样的，也就是说点的坐标发生的变化应该是一样的。因此，这个坐标变换就是把一个点的像

点的坐标作为自己的新坐标，即

$$x'^\mu(p) = x^\mu(\phi(p))$$

这可以进一步诱导出张量场的分量变换。对于某个点的张量，同样是像点的张量的坐标分量作为自己的新坐标分量。不过这个“像点的张量”实际上是该点张量的推前，而不是张量场在像点的取值（同样可以把 ϕ 理解为 $M \rightarrow M'$ 的映射，这样 M' 上本没有张量场，需要推前过去），即

$$T'^{\mu\cdots}{}_{\nu\cdots}|_p = (\phi_* T)^{\mu\cdots}{}_{\nu\cdots}|_{\phi(p)}$$

这两个坐标变换和分量变换在相对论中也有重要的作用。

2.3.2 李导数*

有了张量场的推前和拉回作为基础，我们就可以来学习流形上张量场的一种新的微分运算——李导数。

设流形 M 上有一个光滑矢量场 v ，我们现在考虑这样一个带有参数 t 的微分同胚： ϕ_t 把一个点顺着矢量场移动距离 t 。这称为 v 诱导出的单参微分同胚。通过这种同胚，我们在此直接给出李导数的定义。

定义 2.9: 张量场的 Lie 导数

设 M 是一个微分流形， v 是 M 上的一个光滑矢量场， ϕ_t 是其诱导出的单参微分同胚。 $T^{\mu\cdots}{}_{\nu\cdots}$ 是 M 上的一个光滑张量场，则 $T^{\mu\cdots}{}_{\nu\cdots}$ 沿着 v 的 **李导数** 定义为

$$\mathcal{L}_v T^{\mu\cdots}{}_{\nu\cdots} = \lim_{t \rightarrow 0} \frac{\phi_t^* T^{\mu\cdots}{}_{\nu\cdots} - T^{\mu\cdots}{}_{\nu\cdots}}{t}$$

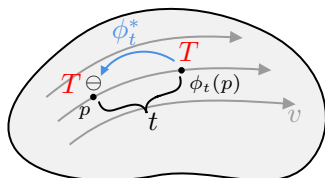
我们来解释一下这个定义的数学含义。我们想要定义张量场的“导数”，应该像微积分里学的一样，把张量场在该点的值与附近点的值相减，除以两点之间的参数差，再取极限，即

$$p \text{ 点处的“导数”} = \lim_{t \rightarrow 0} \frac{T|_{\phi_t(p)} - T|_p}{t}$$

但 $T|_{\phi_t(p)}$ 和 $T|_p$ 属于不同点的切空间，因此它们无法直接相减。那该怎么办呢？我们可以用刚刚定义的拉回映射，把 $T|_{\phi_t(p)}$ 拉回到 p 点，再进行计算。这就是我们的李导数

$$p \text{ 点处的李导数} = \lim_{t \rightarrow 0} \frac{\phi_t^*(T|_{\phi_t(p)}) - T|_p}{t}$$

我们用下图显示这个定义的几何直观。



那么李导数和我们前面学习到的协变导数有什么区别呢？我们注意到，在李导数的定义中，流形 M 上不需要联络等任何附加结构。而协变导数 $v^\sigma \nabla_\sigma T^{\mu\cdots}{}_{\nu\cdots}$ 需要指定联络结构 ∇_σ 。我们说过，联络实际上是在指定一个黎曼法坐标系，从而给出了流形的几何结构，让我们可以在上面用一个矢量求导。而在李导数中，我们虽然没有用联络指定出这样

为了简洁，我们在此略去张量的指标。

注意，协变导数中我们需要的是该点处的一个矢量，而李导数中我们需要的是一个矢量场。

的坐标系，但给定了一个完整的矢量场。这个矢量场无法给出一个坐标系，但可以给出其中一条“坐标轴”。有了这一条坐标轴，我们就可以沿着这一个方向进行求导，这就是李导数。

例 2.13. 对于流形 M 和上面的矢量场 v ，我们找到一个坐标系 $\{x^\mu\}$ ，使得第一坐标方向的切矢量场即为 v ，即

$$v = \frac{\partial}{\partial x^1}$$

则这个坐标系称为矢量场 v 的一个适配坐标系。在适配坐标系下，张量场 $T^{\mu\cdots}{}_{\nu\cdots}$ 沿着 v 的李导数 $\mathcal{L}_v T^{\mu\cdots}{}_{\nu\cdots}$ 为

$$\mathcal{L}_v T^{\mu\cdots}{}_{\nu\cdots} = \frac{\partial T^{\mu\cdots}{}_{\nu\cdots}}{\partial x^1}$$

或者说，若 ∂_μ 是 v 的适配坐标系下的偏导数算符，则有 $\mathcal{L}_v T^{\mu\cdots}{}_{\nu\cdots} = v^\sigma \partial_\sigma T^{\mu\cdots}{}_{\nu\cdots}$ 。

上式表明了我们可以通过适配坐标系的偏导数来计算李导数。但我们希望给出一种不依赖坐标系的方式计算李导数，这就需要我们前面熟悉的 Levi-Civita 联络。我们直接给出下面的结论。

定理 2.8: 李导数与协变导数

设 v^μ 是 M 上的一矢量场， ∇_μ 是任一联络，有

$$\mathcal{L}_v u^\mu = v^\nu \nabla_\nu u^\mu - u^\nu \nabla_\nu v^\mu$$

$$\mathcal{L}_v \omega_\mu = v^\nu \nabla_\nu \omega_\mu + \omega_\nu \nabla_\mu v^\nu$$

对于任意张量场，除协变导数项外，每个逆变指标都会产生一个负的与上式形式相同的项，每个下标都产生一个正的与下式形式相同的项。

2.3.3 Killing 矢量场*

在相对论中，我们关注时空的度规，因为度规决定时空的一切几何性质。我们也关注坐标变换，因为它涉及时间的时空坐标，而坐标变换又可以由微分同胚诱导出。因此，我们也应该关注度规与微分同胚的联系。

对于 M 上的一个微分同胚 ϕ ，如果它对度规场的拉回不改变度规场，即 $(\phi^* g)_{\mu\nu} = g_{\mu\nu}$ ，则称 ϕ 是一个 **等度规映射**。由这个概念，我们来定义一个描述空间度规对称性的量——Killing 矢量场。

定义 2.10: Killing 矢量场

设 $(M, g_{\mu\nu})$ 是一个带度规的流形， ξ^μ 是其上的一个光滑矢量场， ϕ_t 是 ξ^μ 诱导出的单参微分同胚。若 ϕ_t 是等度规映射，即 $(\phi_t^* g)_{\mu\nu} = g_{\mu\nu}$ ，亦即

$$\mathcal{L}_\xi g_{\mu\nu} = 0$$

则称 ξ^μ 是 $(M, g_{\mu\nu})$ 上的一个 **Killing 矢量场**。

既然定义 Killing 矢量场的流形已经带有度规了，它就会诱导出一个适配的 Levi-Civita 联络 ∇_μ 。我们可以利用这个联络把 Killing 矢量场定义式中的李导数展开，很容易得到

$\mathcal{L}_\xi g_{\mu\nu} = \nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu$. 于是我们就得到了 Killing 矢量场的一个充要条件, 称为 Killing 方程.

定理 2.9: Killing 方程

在流形 $(M, g_{\mu\nu})$ 上, 矢量场 ξ^μ 是 Killing 矢量场的充要条件是满足 Killing 方程

$$\nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu = 0$$

其中 ∇_μ 是 Levi-Civita 联络.

既然 Killing 矢量场描述了度规场的对称性, 我们自然就会关心一个流形上所有的 Killing 矢量场. 然而, Killing 方程是一个线性方程, 所以一个流形上的所有 Killing 矢量场构成一个线性空间. 几何直观上来说, 度规场顺着 Killing 矢量场移动后不变, 那么度规场先顺着一个 Killing 矢量场走一段、再顺着另一个 Killing 矢量场走一段, 最终也应该不变. 这也指向所有的 Killing 矢量场构成一个线性空间. 对于线性空间, 我们就不再考虑它包含的所有矢量, 而是关心它包含的所有线性独立的矢量, 也就是它的基. 求出所有的线性独立的 Killing 矢量场其实很有难度. 但我们对独立的 Killing 矢量场, 也就是线性空间的维数, 有下面的结论.

定理 2.10: Killing 矢量场的线性空间

流形 $(M, g_{\mu\nu})$ 上的所有 Killing 矢量场构成一个线性空间, 其维数不超过 $\frac{n(n+1)}{2}$, 其中 n 是流形 M 的维数.

我们上面说过, Killing 矢量场表示的是度规场的对称性. 显然每个线性独立的 Killing 矢量场, 也就是 Killing 矢量场构成的线性空间的每个基, 都应该代表一种独立的对称性. 我们来看两个例子. 这两个例子中的空间都有最多的 $n(n+1)/2$ 个独立的 Killing 矢量场, 因而称为 **最大对称空间**, 又称为 **常曲率空间**.

例 2.14. 考虑二维欧氏空间 $(\mathbb{R}^2, \delta_{\mu\nu})$, 其 Killing 矢量场所构成的线性空间的 **最大维数** 为 3. 选定笛卡尔坐标系 $\{x, y\}$, 显然它是 x 方向的基矢量场 $\frac{\partial}{\partial x}$ 的适配坐标系. 而在这个坐标系下, 线元

$$ds^2 = dx^2 + dy^2$$

与 x 坐标无关, 因此其对 x 偏导为零. 而这个偏导就是对矢量场 $\frac{\partial}{\partial x}$ 的李导数. 因此, $\frac{\partial}{\partial x}$ 就是该欧氏空间的一个 Killing 矢量场, 它代表度规沿着 x 方向平移不变. 同理, 我们也可以发现 $\frac{\partial}{\partial y}$ 也是一个 Killing 矢量场.

而如果使用选择极坐标系 $\{r, \varphi\}$, 则线元表示为

$$ds^2 = dr^2 + r^2 d\varphi^2$$

它与 φ 坐标无关, 因此 $\frac{\partial}{\partial \varphi} = -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y}$ 也是一个 Killing 矢量场, 它代表了度规的旋转对称性.

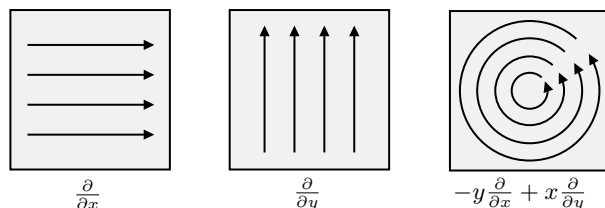
注意到线元其实表示的是度规.

注意组合参数不是常数, 所以它不是 $\partial/\partial x$ 和 $\partial/\partial y$ 的线性组合.

这样，我们就找全了 $(\mathbb{R}^2, \delta_{\mu\nu})$ 的三个线性独立的 Killing 矢量场

$$\left\{ \frac{\partial}{\partial x}, \frac{\partial}{\partial y}, -y \frac{\partial}{\partial x} + x \frac{\partial}{\partial y} \right\}$$

其几何直观如下



而对于一个 n 维欧氏空间 $(\mathbb{R}^n, \delta_{\mu\nu})$ ，其线性独立的 Killing 矢量场应包括 n 个表示平移对称性的矢量场和 $\frac{n(n-1)}{2}$ 个表示旋转对称性的矢量场，其形式与上面类似。

例 2.15. 再考虑二维闵氏时空 $(\mathbb{R}^{1,1}, \eta_{\mu\nu})$ 。在选定了洛伦兹坐标系 $\{t, x\}$ 后，其线元可以表示为

$$ds^2 = -dt^2 + dx^2$$

由类似的推理可知，两个坐标基矢量场 $\frac{\partial}{\partial t}$ 和 $\frac{\partial}{\partial x}$ 都是该闵氏时空的 Killing 矢量场。它们表示闵氏度规的时间与空间平移对称性。

而与欧氏空间的极坐标类似，在二维闵氏时空中我们可以定义下面的“双曲坐标” $\{\rho, \varphi\}$ 满足

$$x = \rho \cosh \varphi, \quad t = \rho \sinh \varphi$$

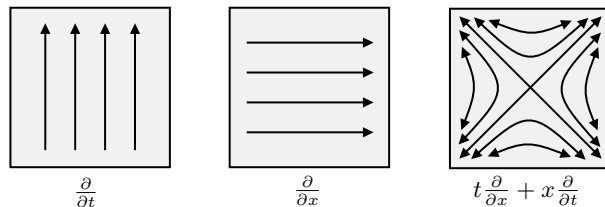
则在此坐标系下，线元可以表示为

$$ds^2 = d\rho^2 - \rho^2 d\varphi^2$$

进而可知 $\frac{\partial}{\partial \varphi} = t \frac{\partial}{\partial x} + x \frac{\partial}{\partial t}$ 也是一个独立的 Killing 矢量场。类似于欧氏空间的旋转，它表示的是闵氏度规的 **伪转动** 对称性。同样，我们就找到了 $(\mathbb{R}^{1,1}, \eta_{\mu\nu})$ 的三个线性独立的 Killing 矢量场

$$\left\{ \frac{\partial}{\partial t}, \frac{\partial}{\partial x}, t \frac{\partial}{\partial x} + x \frac{\partial}{\partial t} \right\}$$

其几何直观如下



而对于四维闵氏时空 $(\mathbb{R}^{1,3}, \eta_{\mu\nu})$ ，其线性独立的 Killing 矢量场应包括 4 个时间/空间平移、3 个空间旋转和 3 个伪转动。闵氏时空的度规也有最高的对称性。

第二部分 狭义相对论

3 洛伦兹变换

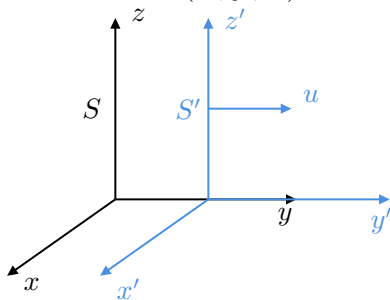
从本章起，我们开始从零学习广义相对论的前置理论——狭义相对论，理解狭义相对论的时空观与牛顿的时空观的差异。

3.1 从伽利略变换到洛伦兹变换

3.1.1 伽利略变换

千百年以来，生活经验告诉人们，时间和空间是绝对的，与任何事物无关。无论你是作者还是躺着、跑着还是跳着，时间的流逝都不会变快或者变慢，物体的体积也不会变大或者变小。牛顿本人曾说过：“绝对空间，就其本性来说，与任何外在的情况无关”，也说过“绝对的、真正的和数学的时间自身在流逝着，而且由于其本性而均匀地、与其他外界事物无关地流逝着”。牛顿的这种时空观被称为绝对时空观。

现在我们来定量描述一下这种绝对的时空观。假如我们有两个坐标系 S 和 S' ，其中 S' 相对 S 做速度方向沿 x 轴、大小为 u 的匀速直线运动。我们再假设在 $t = 0$ 时两系的原点重合。那么加入有一个事件 p ，它在 S 系看来发生的时间为 t ，空间坐标为 (x, y, z) ，在 S' 系看来发生的时间为 t' ，空间坐标为 (x', y', z') ，



则 p 在两系的时间与空间的关系称为 **伽利略（坐标）变换**，总结为下式。

$$\begin{aligned} t' &= t \\ x' &= x - ut \\ y' &= y \\ z' &= z \end{aligned}$$

在此基础上，我们可以用坐标除以时间，得到两个坐标系之间的速度变换，称为伽利略速度变换。

$$\begin{aligned} v'_x &= v_x - u \\ v'_y &= v_y \\ v'_z &= v_z \end{aligned}$$

于此直接相关的一个概念称为 **伽利略相对性原理**。它指出，力学定律在一切惯性参考系中具有相同的形式，任何力学实验都不能区分静止和匀速运动的惯性参考系。例如，在所有的惯性参考系中，牛顿第二定律 $\mathbf{F} = m\mathbf{a}$ 都成立，不会到某个惯性参考系中， \mathbf{F} 就不等于 $m\mathbf{a}$ 了。

伽利略变换一直都被人们默认是正确的，直到 19 世纪，它开始受到了挑战。

3.1.2 洛伦兹坐标变换

对伽利略变换的重要挑战来自电磁学. 十八世纪五十年代, 麦克斯韦发表了著名的麦克斯韦方程组, 统一了经典电磁理论. 人们发现, 通过麦克斯韦方程组可以解出真空中的光速

$$c = \frac{1}{\sqrt{\mu_0 \epsilon_0}}$$

这个解与参考系无关, 也就是说光速在任何参考系下都应该不变! 这显然与伽利略变换不相符. 为了解决这个矛盾, 一些人提出, 麦克斯韦方程组可能只在一个特殊的、绝对的参考系成立, 这个系称为以太系.

既然有一个绝对的以太系, 那么我们就应该能测出我们在以太系中的运动速度. 历史上, 最著名的一个实验称为迈克耳孙-莫雷干涉实验. 迈克耳孙和莫雷试图通过光的干涉来测定地球在以太系中运动的速度. 但实验的结果是——地球在以太系中的速度几乎为零! 既然地球是绕着太阳进行圆周运动的, 那地球的运动速度就不可能一直和以太系相同. 这说明, 不存在这样一种绝对的参考系! 也就是说, 光速真的在任何参考系下应该都是不变的!

这个结论非常令人匪夷所思, 许多物理学家都试图去解释它. 但爱因斯坦说, 既然物理世界是这样的, 那我就根据这条结论建立一套新的理论好了. 于是, 1905 年, 爱因斯坦发表了一篇革命性的论文《论动体的电动力学》, 提出了狭义相对论, 引发了一场物理学革命. 在这篇论文中, 爱因斯坦提出了以下两大基本假设:

1. 光速不变原理: 在所有的惯性系中, 真空光速都相等;
2. 爱因斯坦相对性原理: 物理定律在一切惯性参考系中具有相同的形式, 任何物理实验都不能区分静止和匀速运动的惯性参考系.

根据以上两条基本假设, 爱因斯坦推导出了一种全新的时间与空间坐标变换, 称为 **洛伦兹(坐标)变换**. 我们在此不进行洛伦兹变换的推导, 推导可以在任意一本相对论教科书中找到.

洛伦兹变换本身是洛伦兹提出的, 但他把这个变换解释为物质在以太中运动时自身的时间、空间的变化, 从而错失了提出狭义相对论的机会.

物理定律 3.1: 洛伦兹变换

若惯性坐标系 S' 相对于 S 以速度 u 沿 x 方向运动, 则事件 p 在两系的时间与空间坐标关系满足

$$\begin{aligned} t' &= \frac{t - ux/c^2}{\sqrt{1 - u^2/c^2}} \\ x' &= \frac{x - ut}{\sqrt{1 - u^2/c^2}} \\ y' &= y \\ z' &= z \end{aligned}$$

在相对论中, 我们经常引入两个记号

$$\beta = \frac{u}{c}, \quad \gamma = \frac{1}{\sqrt{1 - \beta^2}}$$

这样, 洛伦兹变换可以被写成

$$t' = \gamma \left(t - \frac{\beta}{c} x \right), \quad x' = \gamma (x - \beta ct), \quad y' = y, \quad z' = z$$

而其逆变换可以视作 S 系相对 S' 系以 $-u$ 的速度运动，于是可以立即得到逆变换

$$t = \gamma \left(t' + \frac{\beta}{c} x' \right), \quad x = \gamma(x' + \beta ct'), \quad y = y', \quad z = z'$$

在这里我们可以注意到，要想让 γ 有意义，也就是为实数，那么必须有 $u < c$ 。这就是我们所说的，在相对论下，任何物体的速度不能超过真空中的光速。当然，光的速度就是光速，因此光本身是不能作为参考系的。

同时我们可以发现，当 $u \ll c$ 时，有 $\gamma \approx 1$ ，可以发现洛伦兹变换退化为伽利略变换。因此，伽利略变换是洛伦兹变换在低速下的近似。

3.1.3 洛伦兹速度变换

有了时间和空间坐标的变换，自然就可以把它们相除，得到两个坐标系之间的速度变换。注意此处 v 代表物体运动的速度， u 代表坐标系 S' 相对 S 运动的速度。

物理定律 3.2: 洛伦兹速度变换

若惯性坐标系 S' 相对于 S 以速度 u 沿 x 方向运动，则同一物体在两系的速度关系满足

$$\begin{aligned} v'_x &= \frac{v_x - \beta c}{1 - \beta v_x/c} \\ v'_y &= \frac{v_y}{\gamma(1 - \beta v_x/c)} \\ v'_z &= \frac{v_z}{\gamma(1 - \beta v_x/c)} \end{aligned}$$

在这里需要指出，与洛伦兹坐标变换不同，在洛伦兹速度变换中，虽然坐标系的运动方向是沿着 x 轴的，但 y 和 z 方向的速度大小也会发生变化。这是因为两系的时间关系不再相同。

例 3.1. 考虑到光沿着 x 轴正向传播，其相对于 S 系的速度为 $v_x = c, v_y = v_z = 0$ 。则根据洛伦兹速度变换，它在 S' 系的速度为

$$\begin{aligned} v'_x &= \frac{c - u}{1 - cu/c^2} = c \\ v'_y &= v'_z = 0 \end{aligned}$$

因此光在 S' 系的速度仍然为 c ，这与光速不变原理相符。同样可以简单地推出，当 $v_x < c$ 时，也有 $v'_x < c$ 。因此洛伦兹速度变换会保持物体的运动速度小于光速。

同样，我们很容易发现，在 $u \ll c$ 时，洛伦兹速度变换退化为伽利略速度变换。

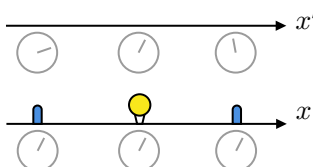
3.1.4 时间的相对性

洛伦兹变换与人们熟悉的伽利略变换的一个重要差异在于，时间不再是一个绝对的量，也就是 t 与 t' 不再恒等，而与坐标系的运动有关。我们考虑下面这样一个实验：在 x 轴上， $x = 0$ 处有一个光源， $x = \pm \ell$ 处各有一个接收器。在 $t = 0$ 时刻，光源发出一个光脉冲。显然，在 S 系看来，在 $t = \ell/c$ 时刻，两个接收器同时接收到光信号。但现在我们考虑 S' 系。在 S' 系看来，两个接收器都在以速度 u 向左运动，同时光信号也在以速度 c

向两侧运动。因此在 S' 系看来，左侧接收器应该在 $t' = \ell/(c - u)$ 时刻接收到光信号，而右侧接收器应该在 $t' = \ell/(c + u)$ 时刻接收到光信号。也就是说，两个接收器接收到光信号的事件在 S' 系看来不再是同时的了！这就是异地同时性的相对性。



由于这种时间的相对性，在相对论中所说的“坐标系”或者“参考系”，除了指我们熟悉的空间坐标系外，还需要在系里的每一个点配备一个钟表，这些钟表随着系一起运动。这样，在 S 系看来， S' 系的不同点的钟表应该指向不同的时间。

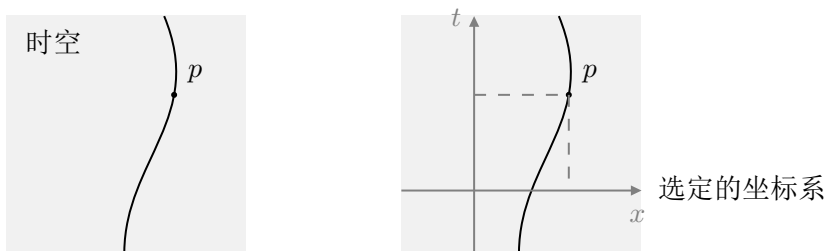


3.2 从四维时空看狭义相对论

上面我们学习了狭义相对论在时间和三维空间中的描述，这也是爱因斯坦最初提出相对论时的描述。这个描述的公式十分复杂。但在爱因斯坦发表狭义相对论后不久，数学家闵可夫斯基发现了一种更加简洁的描述方式。这就是相对论的四维描述。

3.2.1 闵氏时空

闵可夫斯基把时间和三维空间看作一个整体，称为 **四维时空**。事件是四维时空中的点，而一系列事件划过的一条时空曲线称为 **世界线**。我们在此强调一件非常重要的事情：正如线性代数中线性空间和矢量是绝对的、坐标是相对的一样，**相对论中时空是绝对的，事件、世界线是时空中的点和线，也是绝对的；只有在选定了时间和空间轴后，一个事件才有时间和空间坐标，因此事件发生的时间和空间位置是相对的。**可以说，这是相对论中“相对”一词的核心含义。



严格来说，闵氏时空是一个带有闵氏度规的流形 $(\mathbb{R}^{1,3}, \eta_{\mu\nu})$ 。在选定了洛伦兹坐标系后，其线元 $\eta_{\mu\nu} dx^\mu dx^\nu$ 可以写作

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2$$

根据线元的正负，我们可以把曲线（和矢量）分为三类：

- 当 $ds^2 < 0$ 时，曲线（矢量）称为 **类时的**；
- 当 $ds^2 = 0$ 时，曲线（矢量）称为 **类光的**；
- 当 $ds^2 > 0$ 时，曲线（矢量）称为 **类空的**。

很容易发现，当 $ds^2 = 0$ 时，曲线所代表的运动速率 $v = \sqrt{(dx^2 + dy^2 + dz^2)/dt^2} = 1$ 即为光速，因此在闵氏时空中，光的世界线是类光曲线。而对于非光的质点，其速度必须

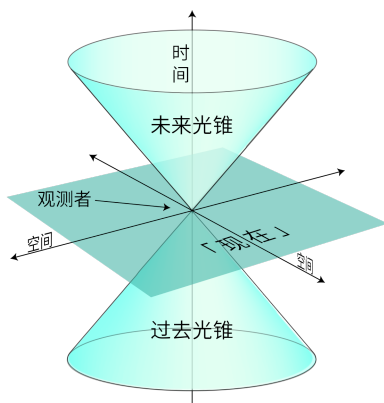
闵可夫斯基在刚开始提出四维表述时，使用的是复欧氏空间，其中的坐标为 (x, y, z, it) ，但这种表述后面已经被完全抛弃。

从本节以后，我们开始使用几何单位制，即令 $c = 1$ ，这样时间和空间就有相同的量纲。

小于光速。用四维语言说，就是质点的世界线为类时曲线。实际上，我们可以完全抛弃三位语言中“不能超光速”的说法，只改用四维语言的表述。我们后面会看到，在一些情况下我们定义出的三位速度是可以超过光速的，但质点仍然满足走类时曲线。

为了直观地画出这三类曲线，我们可以给四维时空选定一个洛伦兹坐标系，并把时间轴向上、空间轴水平，画出一个二维或三维的直角坐标系，以代表时空，这种图称为时空图。在时空图中，曲线所代表的运动速率 $v = \sqrt{(dx^2 + dy^2 + dz^2)/dt^2}$ 即为曲线的斜率的倒数。因此，类空曲线的斜率大于 1，类光曲线的斜率等于 1，而类时曲线的斜率小于 1。这样可以发现，在给定一个事件点 p 后，过 p 的所有类光曲线构成一个圆锥，称为**光锥**。类时曲线在光锥的内部，而类空曲线在光锥的外部。因此我们可以说，在时空中，质点的世界线只能在光锥内部，这也是“不能超光速”的另一种四维表述。

我们在例 2.7 中讲过，洛伦兹坐标系就是闵氏时空的标准直角坐标系。



在此再次强调：线元与度规是等价的，它表示一段微元线段的“平方线长”，与坐标系的选取毫无关系。因此，一个曲线（矢量）是类时的、类光的还是类空的也与坐标系的选取毫无关系。同理，光锥也与坐标系的选取毫无关系。在狭义相对论中画图时，我们会发现光锥总是围绕时间轴旋转对称的，但这并不意味着光锥与时间轴的选取有关。闵氏时空在这一点上有些反直观——由于光速不变原理，在任何惯性坐标系下，光锥与时间轴的关系总是这样的。

3.2.2 观测者与坐标系

闵氏时空中的事件和世界线是绝对的，但我们关心的往往是你相对论的世界中会观察到什么物理现象。这时，你叫做一个**观测者**。严谨地说，一个观测者就是一条类时世界线。如果你在做惯性运动，也就是说你的世界线是一条类时测地线（直线），那么你称为一个惯性观测者。

现在考虑一个质点，它的世界线的线元为 $ds^2 = -dt^2 + dx^2 + dy^2 + dz^2$ 。而在它自己本身看来，它自己并没有移动，因此有 $dx = dy = dz = 0$ 。因此世界线线元可以简化为 $ds^2 = -dt^2$ 。而由于类时曲线的线元为负，因此其线长为

这段推理并不严谨，但结论是正确的。

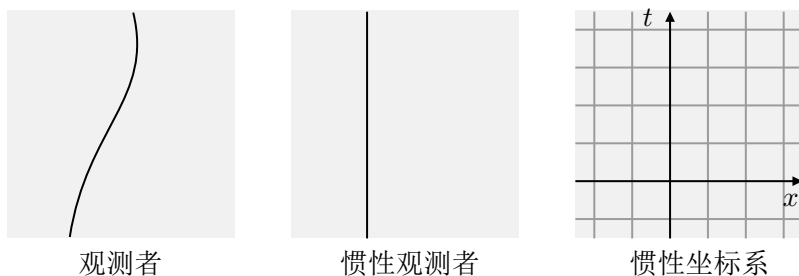
$$\int \sqrt{|ds^2|} = \int dt = t$$

因此，我们可以发现类时测地线的线长就是质点“自身感觉所经历的时间”，这称为**固有时**，记作 τ 。我们在前面讲过，流形上所说的曲线都是参数曲线。而从现在开始，我们一律用线长也就是固有时作为曲线的参数。直观地说，这样曲线 $C(\tau)$ 表示了粒子“在自己经历了时间 τ 后”所在的时空位置，这样可以为后面的计算带来极大的方便。

让我们再回到一个观测者的情况。由于观测者的世界线是对他来说只有时间流逝、而空间不变的曲线，因此可以说这就是这个观测者的时间轴。而垂直于这个时间轴的平面就

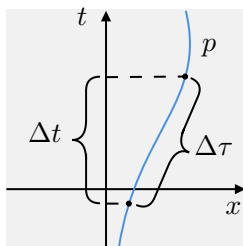
是观测者眼中的“三维空间”。如果我们在这三维空间中再建立直角坐标系，那么这个空间坐标系和时间轴就会共同构成一个**坐标系**。对于惯性观测者来说，它定义的坐标系称为**惯性坐标系**，实际上也就是我们之前提到的洛伦兹坐标系。

还有一个常见的概念叫参考系，它在不同人的口中的定义有所不同。



在给定了坐标系后，每个事件自然就会有一个时间坐标，称为**坐标时**，记作 t 。坐标时是指在这个坐标系看来，事件所发生的时间。在给定一个时间 t_0 后，坐标时 $t = t_0$ 定义了时空中的一个三维子空间，称为一个**等时面**。它就是该坐标系在 t_0 时刻所处的三维“空间”。

在此强调一下坐标时和固有时区别。固有时是世界线的线长，和坐标系的选择无关。而坐标时则是选定了坐标系后的时间分量，它取决于坐标系的选取。当然，两种时间的零点都可以任意选取，我们关心的常常是两个事件的时间差 Δt 和 $\Delta \tau$ 。



坐标时和固有时之间的关系很容易推出

$$\frac{dt}{d\tau} = \gamma = \frac{1}{\sqrt{1-v^2}}$$

其中 v 是粒子相对于坐标系的运动速度。

这里量纲看起来很奇怪，但这是因为我们使用了自然单位制。

3.2.3 从四维看洛伦兹变换

下面，我们用四维时空的语言来描述之前学习过的洛伦兹坐标变换。我们将以三种不同的视角来重新理解洛伦兹变换，每一种都比上一种更加进阶。

闵氏时空中的坐标系变换

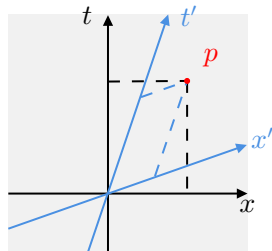
假设我们有一个惯性系 S ，它在时空中表现为一个洛伦兹坐标系。现在假设有一个惯性观测者，他相对于 S 系以速度 u 运动，那么他的世界线是一条直线 $x = ut$ 。那么这个观测者所确定的惯性系 S' 将以其世界线为 t' 轴，以垂直于它的平面作为 x' 轴。但注意，在闵氏时空中，我们需要以闵氏度规定义垂直。考虑一个沿着 t' 轴的矢量 $[ut, t]$ ，一个矢量 $[a, b]$ 若要与它垂直，则应有 $ut \cdot a - t \cdot b = 0$ ，从而得到 $b/a = u$ 。也就是说，与时间轴垂直的直线的斜率为 u 。总结起来，我们有两条新的坐标轴

由于这部分要画图，所以我们在二维闵氏空间中讨论。

$$t' \text{ 轴: } x = ut$$

$$x' \text{ 轴: } t = ux$$

可以发现，在闵氏时空中，互相垂直的两条直线是关于 $t = x$ 对称的，而 $t = x$ 正是光所走的类光直线。所以我们可以说，关于类光直线对称的两条直线是互相垂直的。



现在，对于一个在 S 系下坐标为 (x, t) 的事件，我们只需把它投影到两条新的坐标轴上，就可以得到新的时间和空间坐标。

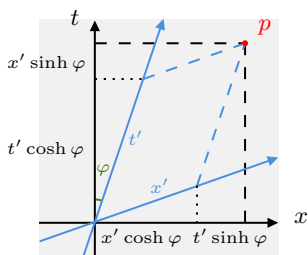
我们设两个时间轴之间的夹角为 φ ，则由几何关系可知

$$u = \tanh \varphi$$

进而有

$$\sinh \varphi = \gamma u, \quad \cosh \varphi = \gamma$$

用 φ 的双曲三角函数，我们可以很容易地得到以下关系



$$x = x' \cosh \varphi + t' \sinh \varphi$$

$$t = x' \sinh \varphi + t' \cosh \varphi$$

进而我们可以解出

$$t' = t \cosh \varphi - x \sinh \varphi$$

$$x' = -t \sinh \varphi + x \cosh \varphi$$

带入 u 与 $\sinh \varphi$ 和 $\cosh \varphi$ 的关系，我们就得到了我们熟悉的洛伦兹变换。

$$t' = \gamma(t - ux)$$

$$x' = \gamma(x - ut)$$

闵氏时空中的正交变换

我们之前讲过，洛伦兹变换是惯性坐标系之间的坐标变换。而惯性坐标系从闵氏时空的角度又称为洛伦兹坐标系，它是指使得度规分量满足

$$\eta = \begin{bmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}$$

的坐标系。而度规作为一个 $(0, 2)$ -型的协变张量，也就是一个二次型，若要通过基变换矩阵 Λ^{-1} 变为另一个坐标系，使得坐标变换为 $x' = \Lambda x$ ，则度规会变为 $(\Lambda^{-1})^\top \eta \Lambda^{-1}$ 。由于这个新坐标系仍然是洛伦兹坐标系，应有 $(\Lambda^{-1})^\top \eta \Lambda^{-1} = \eta$ ，也就是

$$\Lambda^\top \eta \Lambda = \eta$$

与三角函数类似，此处的双曲正切 \tanh 也代表“对边比邻边”，只不过这里的边长是闵氏长度。

与欧氏空间中的正交矩阵类似，这种矩阵 Λ 可以称为闵氏时空中的“正交矩阵”。显然，洛伦兹变换的叠加仍然是洛伦兹变换，因此所有满足上式的变换 Λ 构成一个群，称为 **洛伦兹群** 或者闵氏时空的 **正交群**，记作 $O(1, 3)$ 。洛伦兹群还有另一种等价的理解方式，就是保持时空间隔不变的变换构成的群，这些变换满足

$$|x| = |\Lambda(x)|$$

其中 $|x|$ 为闵氏度规下的长度。

洛伦兹群的元素很多，我们现在考虑其中特定的一类：满足 $\det \Lambda = 1$ ，也就是要求四个坐标轴的“手性”不变。同时，为了表示沿 x 轴移动的洛伦兹变换，我们还要求保持 y 和 z 坐标不变。类比于欧氏空间的正交矩阵，这样的矩阵 Λ 一定可以写成以下这种形式

$$\Lambda_x(\varphi) = \begin{bmatrix} \cosh \varphi & -\sinh \varphi & 0 & 0 \\ -\sinh \varphi & \cosh \varphi & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

对于一般的洛伦兹变换 Λ 一定有 $\det \Lambda = \pm 1$ 。

二维欧氏空间的满足行列式为 1 的正交矩阵可以写作

$$\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

类比于欧氏空间中行列式为 1 的正交矩阵所代表的转动，这样的变换称为 **伪转动**。按照和上面相同的推导，令 $\tanh \varphi = u$ ，我们就可以发现这个矩阵就代表了沿 x 轴运动的洛伦兹变换。

对于一个事件 (t, x, y, z) ，它通过洛伦兹变换 Λ 以后的时间坐标 t' 为

$$t' = \Lambda^0_0 t + \text{与 } t \text{ 无关的部分}$$

我们常常要求洛伦兹变换对时间不进行反演，即 $\Lambda^0_0 > 0$ 。同时，仍然要求 $\det \Lambda = 1$ ，则可要求任何轴都不进行反演。满足这两个条件的所有洛伦兹变换构成洛伦兹群的一个子群，称为 **特殊洛伦兹群** 或者闵氏时空的 **特殊正交群**，记作 $SO(1, 3)$ 。当我们提到洛伦兹变换或者洛伦兹群时，实际上一般默认指的就是这些特殊洛伦兹变换。在 $SO(1, 3)$ 中，最一般的洛伦兹伪转动矩阵为

$$\Lambda = \begin{bmatrix} \gamma & -\gamma v_x & -\gamma v_y & -\gamma v_z \\ -\gamma v_x & 1 + \frac{(\gamma - 1)v_x^2}{v^2} & \frac{(\gamma - 1)v_x v_y}{v^2} & \frac{(\gamma - 1)v_x v_z}{v^2} \\ -\gamma v_y & \frac{(\gamma - 1)v_y v_x}{v^2} & 1 + \frac{(\gamma - 1)v_y^2}{v^2} & \frac{(\gamma - 1)v_y v_z}{v^2} \\ -\gamma v_z & \frac{(\gamma - 1)v_z v_x}{v^2} & \frac{(\gamma - 1)v_z v_y}{v^2} & 1 + \frac{(\gamma - 1)v_z^2}{v^2} \end{bmatrix}$$

最后，洛伦兹变换作为一个线性变换，它是保持时空坐标原点不动的。当然，我们还可以改变原点的位置，同时也保持度规不变。这样的变换就包含了平移的部分。所有这样的变换也构成一个群，称为 **庞加莱群**。

Killing 矢量场诱导的坐标变换

在 2.3.3 节中我们学到过，表示流形上度规的对称性的量是 Killing 矢量场。对于平直闵氏时空 $(\mathbb{R}^{1,3}, \eta_{\mu\nu})$ ，在给定了洛伦兹坐标系后，我们在例 2.15 中推出过它的 Killing 矢量场的基矢量场应包括

- 形如 $\frac{\partial}{\partial t}, \frac{\partial}{\partial x}$ 的矢量场，表示时间/空间平移对称性，共 4 个；

- 形如 $-y\frac{\partial}{\partial x} + x\frac{\partial}{\partial y}$ 的矢量场，表示空间旋转对称性，共 3 个；
- 形如 $t\frac{\partial}{\partial x} + x\frac{\partial}{\partial t}$ 的矢量场，表示伪转动对称性，共 3 个。

现在我们考虑第三类伪转动的 Killing 矢量场 $t\frac{\partial}{\partial x} + x\frac{\partial}{\partial t}$ 。由于这个矢量场与 y 和 z 坐标无关，我们可以忽视掉这两个坐标。设 ϕ_φ 表示沿着该矢量场移动参数 φ 的微分同胚。可以求出，这样的微分同胚在洛伦兹坐标系下可以写成

$$\phi_\varphi : (t, x) \mapsto (x \sinh \varphi + t \cosh \varphi, x \cosh \varphi + t \sinh \varphi)$$

因此由它诱导出的坐标变换为

$$t' = x \sinh \varphi + t \cosh \varphi$$

$$x' = x \cosh \varphi + t \sinh \varphi$$

进而同样可以令 $\tanh \varphi = u$ 从而得到我们熟悉的洛伦兹变换。因此，从这个角度来看，洛伦兹变换就是闵氏时空的等度规变换诱导出的坐标变换。实际上，这里的微分同胚 ϕ_φ 就是上面的正交矩阵 $\Lambda_x(\varphi)$ 。

3.3 狭义相对论效应分析

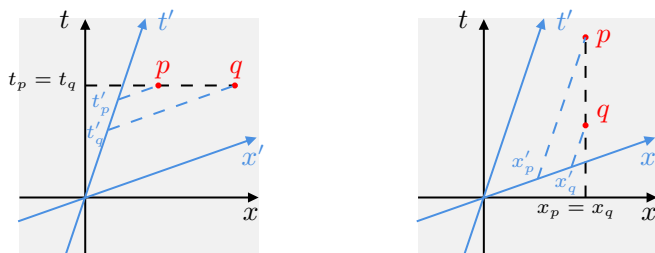
有了四维语言，让我们用它来分析一下相对论中一些耳熟能详的效应。与简单地用洛伦兹变换公式相比，这种分析更为直观也更为优雅。

3.3.1 时间与空间的相对性

与伽利略的时空观相比，狭义相对论的时空观的一大特点就是时间和空间不再是绝对的，而是相对的。进而，同时和同地的概念也是相对的。也就是说，在一个参考系同时（或同地）的两个事件，在另一个参考系下未必同时（或同地）。

我们在 3.1.4 节已经给出了一个同时的相对性的例子。当时，我们是用的三维语言 + 光速不变原理给出的解释，但这个解释或许还有些令人费解。事实上，在做题时，许多人也会不经意间忽略掉一些这种现象，导致解出错误的答案。但这个现象从四维时空的角度看则几乎是显然的，而且一定不会错。

考虑两个事件 p 和 q ，如果它们在惯性系 S 看来是同时的，这意味着它们在 S 系有相同的 t 坐标。但是和 S 相比，另一个惯性系 S' 的坐标轴是“倾斜的”。那么自然，两个事件投影到 t' 轴的坐标就不一定还相等了。因此在 S' 系看来，两个事件就不再同时了。空间的相对性也是类似。这个推理还说明，如果两个事件是同时同地发生的，则在任何参考系看来它们都是同时同地发生的，这是因为它们代表了同一个时空点。



根据同样的推理也可以发现，两个事件的时间先后关系在不同的参考系下可能相反。但有时，两个事件的时间顺序必须是绝对的，比如两个事件之间由因果关系的时候。

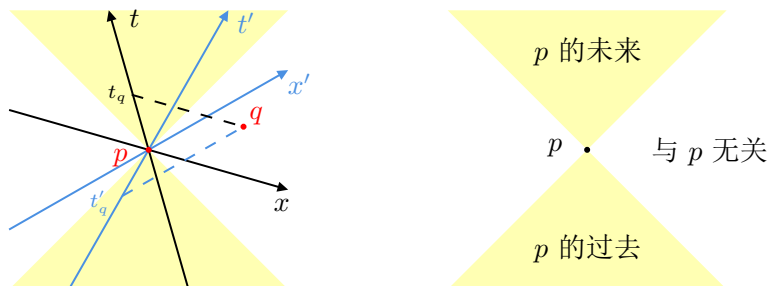
这种情况下，如果时间的先后是相对的，则会违背因果律，出现“你出生早于你爸爸出生”这种“果早于因”的悖论。因此，我们有必要来具体讨论一下两个事件的时间关系。

考虑两个事件 p 和 q ，我们以 p 为原点建立两个不同的惯性坐标系 S 和 S' 。作图可以发现，当 x 轴和 x' 轴位于 q 点的两侧时，两个参考系所观察到的 q 的时间一正一负，也就是说两个参考系一个认为 p 在先，一个认为 q 在先。然而，惯性坐标系的坐标轴的选取是有限制的：时间轴和空间轴必须对称地位于类光直线的两侧，并且时间轴必须类时，空间轴必须类空。因此，我们发现，当 q 和 p 之间是类时联系的时候，也就是说 q 位于 p 的光锥之内时， x 轴在变化时便无法跨越 q 点，因此 p 与 q 的时间先后是绝对的。反之，如果 q 和 p 之间是类空联系，即 q 位于 p 的光锥之外，则空间轴的选取可以在 q 的两侧，因此此时 p 和 q 之间的时间先后关系在此时是依赖坐标系的选取的。

因此，对于一个事件 p ，我们可以根据它把时空分为三个部分：

- p 的过去光锥：代表 p 的绝对过去，其中的事件可以是 p 的因；
- p 的未来光锥：代表 p 的绝对未来，其中的事件可以是 p 的果；
- p 的光锥之外：既不属于 p 的过去也不属于 p 的未来，其中的事件必与 p 无因果关系。在不同坐标系下看来与 p 的时间关系可先可后。

这也与我们熟悉的“信息的传播速度不超过光速”相一致。如果两个事件之间有因果关系，则“因事件”的信息必须可以传递至“果事件”，因此二者之间必须为类时联系。若两者为类空联系，则只有超光速才可以把信息在二者之间传递，而这是相对论所不允许的。



3.3.2 尺缩效应与车库佯谬

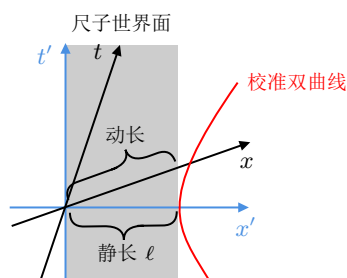
在相对论中，“空间相对性”的一个著名的效应就是**尺缩效应**，它常常被表述为：运动的物体在沿着运动方向上的长度会收缩。例如，一个静止时长度为 ℓ 的尺子，若沿着尺子的方向以速度 v 运动，则其长度会收缩为 $\gamma^{-1}\ell = \ell\sqrt{1-v^2}$ 。

不过，这个描述在参考系的问题上稍微有些模棱两可。我们在此给出一个更加没有歧义的描述：一个尺子，沿着 x 轴方向摆放。若它静止于 S 系时，在 S 系测得的长度为 ℓ 。那么它如果相对于 S 系以速度 v 沿着 x 轴运动，则在 S 系下测得的长度会变为 $\gamma^{-1}\ell$ 。

这个效应用洛伦兹变换公式可以很容易地解释，但我们在此用四维几何语言对它进行进一步阐释。考虑一个相对于尺子静止的参考系 S' ，尺子在时空中会划过一个在 x' 轴长度为静长 ℓ 、沿 t' 轴无限延伸的“世界面”。而对于相对于 S' 系运动的 S 系，它的时间轴和空间轴是倾斜与尺子世界面的。那么考虑 S' 系的一个等时面，例如 $t' = 0$ 的面，它就是 x' 轴。那么 x' 轴截尺子世界面的长度就应该是 S' 系观察到的尺子长度，即图中的动长。动长看似比静长要长，但注意在闵氏时空中，类空直线的长度定义为 $l = \sqrt{-t^2 + x^2}$ 。动长由于 t 不为零，所以长度实际上是更短的！这样我们就解释了尺缩效应——实际上尺

子并没有任何变化，只是 S 系和 S' 系观察到了尺子世界面的不同截线，或者说，它们观察到的不是同一把“空间尺子”！

请尝试把 S 画成直角，看看这样是否方便比较。



为了方便进行几何比较，我们在此引入一个概念——相对于某个给定点 o ，所有距 o 等距离的点，即满足下式的点

$$-\Delta t^2 + \Delta x^2 + \dots = l^2$$

定义了一条双曲线（一个双曲面），称为 **校准双曲线**（双曲面）。校准双曲线可以帮助我们很方便地比较两个不同长度。我们在图中画出了以静长 l 为距离的校准双曲线，这样可以轻易看出动长是小于静长的。

若要定量地计算尺缩效应，则可以根据上图的几何关系求出动长

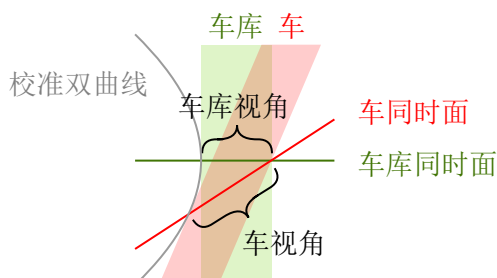
$$\text{动长} = \sqrt{l^2 - t'^2}$$

其中 t' 为 x 轴与尺子右端世界线的交点的 t' 坐标。而 x 轴在 S' 中表达式为 $t' = vx'$ ，考虑该点即有 $t' = vl$ 。那么我们把上式从根号里提出一个 l ，即可得到

$$\text{动长} = l \sqrt{1^2 - \left(\frac{t'}{l}\right)^2} = l \sqrt{1 - v^2} = \gamma^{-1} l$$

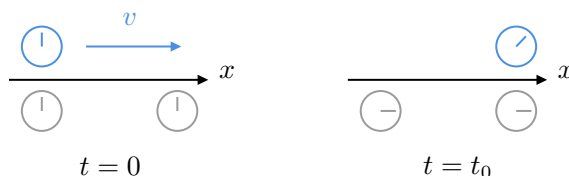
与尺缩效应相关的一个有趣的“悖论”叫做车库佯谬：有一辆车和一个车位，二者的长度相等。那么车开进车位时，车会认为，车库相对于它运动，因此长度收缩，车库无法装下车。而车库则认为，汽车相对于它运动，长度收缩，因此车库装下车绰绰有余。这个“悖论”该如何解释呢？用洛伦兹变换公式当然是可以的，但你需要非常清楚的头脑，不被任何同时或者同地的相对性绕晕。但如果用四维几何语言，则可以轻松解释这个现象。

下面画出了车和车库的世界面。考虑车右端恰好位于车库右端的事件，过该点作两个参考系的同时面，则可以发现上面车与车库的看法都正确，这是因为它们观察到的不是同一个“空间车”和“空间车库”。或者说，在车库看来，车左侧进库晚于车右侧出库；而在车看来，车左侧进库早于车右侧出库。既然这两个事件不是绝对同时的，那“能否装下”自然就也是相对的了。注意，作图时，为了满足车与车库静长相等，我们以该点为中心作校准双曲线，则车同时面与车左侧世界线的交点、和车库同时面与车库左侧世界线的交点，这两点应该过同一条校准双曲线。

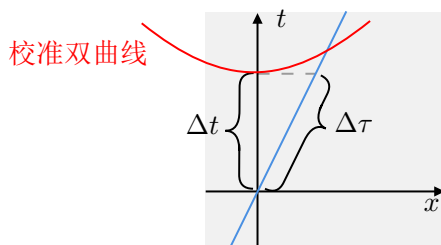


3.3.3 钟慢效应与双生子佯谬

在相对论中，时间的测量也与参考系有关，这称为 **钟慢效应**：运动的钟表走的会比静止的钟表慢。在三维语言中，我们常常把它画成以下示意图。一个钟表相对于 S 系运动，若在 $t = 0$ 时它与 S 系的钟表对时，时间相同，则在 $t = t_0$ 时刻再次对时，时间会不再相同。运动的钟表的时间读数会小于 S 系的钟表。注意，我们强调过，由于非同时同地事件的时间与空间相对性，我们的观测必须是同时同地的直接观测，也就是说，在两次对时的时刻， S 系都要有一个同地的钟表与它对时。



钟慢效应实际上体现的就是固有时和坐标时的差别。在 S 系所观察到的运动钟表走过的时间是 S 系的坐标时 Δt ，而钟表自身走过的时间则是其固有时 $\Delta \tau$ 。我们说过，固有时就是类时曲线的线长。这样，我们的问题就又化为了闵氏时空比较曲线长度的问题。



在这里也类似， $\Delta \tau$ 看似大于 Δt ，但由于这是闵氏几何，我们可以画出校准双曲线

$$-\Delta t^2 + \Delta x^2 + \dots = \Delta t_0^2$$

则可以发现， $\Delta \tau$ 实际上是大于 Δt 的。若要定量计算，我们可以列出

$$-\Delta \tau^2 = -\Delta t^2 + \Delta x^2$$

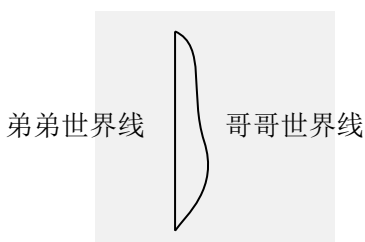
而 $\Delta x = v\Delta t$ ，因此我们可以得到

$$\Delta \tau = \gamma^{-1} \Delta t$$

即以速度 v 运动的钟表的走时率是静止的钟表的 γ^{-1} 倍。

与钟慢效应相关的一个有趣的“悖论”是双生子佯谬：一对双胞胎，其中哥哥去星际航行以后回到了地球，而弟弟则一直生活在地球上。那么在弟弟看来，哥哥一直在运动，因此经历的时间短，会更年轻；而由于运动是相对的，在哥哥看来，是弟弟一直在运动，因此弟弟会更年轻。

这个问题在四维表述中也可以轻易解决。哥哥和弟弟经历的时间是他们各自世界线的线长。弟弟由于没有经过加速运动，其世界线是一条直线。而哥哥去星际航行后返回地球，一定经历了加速运动，因此世界线是一条曲线。我们只需比较两条线的线长。



与上一节相比，这里的校准双曲线与原点是类时联系的。

注意类时的时空间隔是负的，因此线长的平方前有个负号。

我们现在讨论狭义相对论，不考虑地球的引力。

有了刚才对钟慢效应的分析作为铺垫，现在应该能直接看出弟弟的世界线更长，因此哥哥会更年轻。或者定量来说，我们以弟弟的视角建立惯性坐标系，则类时曲线的长度为

$$\Delta\tau = \int \sqrt{dt^2 - dx^2 - dy^2 - dz^2}$$

哥哥和弟弟的积分起止点相同，但哥哥在空间分量有变化，因此积分式更小，即哥哥所经历的固有时更短。

那么上面用钟慢效应认为“在每个人看来，另一个人都更年轻”的说法错在了哪里呢？首先，由于所有惯性系平权，因此在钟慢效应的场景中，运动的钟表较慢的确是相对的，也就是说，如果运动的钟表的参考系中有另外一个钟表可以和 S 系的同一个钟表对时，那么可以发现 S 系的钟表的确更慢。

但双生子佯谬和钟慢效应场景的区别在于，双生子佯谬中兄弟二人在地球重聚时对比时间，也就相当于钟慢效应中运动的钟表和 S 系的同一个钟表两次对时。为了使他们可以两次对时，其中一个人必须作非惯性运动，这样二人之间就产生了绝对的差异——世界线在时空中是否是直线是绝对的。或者从另一个角度说，双生子佯谬中比较的是两个固有时，而非固有时和坐标时。而固有时作为线长，是绝对的，因此二人之间会有绝对的年龄差异。

4 狭义相对论动力学

在上一章，我们只讨论了狭义相对论的运动学。而在本章，我们将开始讨论其中的动力学——在狭义相对论中，力如何影响运动。

4.1 质点动力学

在牛顿力学中，质点动力学的核心方程是 $\mathbf{F} = m\mathbf{a}$ 。显然这个式子不满足洛伦兹协变，而且也不是用四维语言表述的。而这一节，我们就来学习狭义相对论中的质点动力学。

4.1.1 质能关系

当我们从牛顿力学踏入相对论力学时，牛顿力学的定律——比如牛顿第二定律——不得不被抛弃。但有一些其他定律，例如守恒定律，我们还是希望保留的。在质点动力学中，一个重要的守恒定律就是动量守恒。但很可惜，由于洛伦兹速度变换不是简单的速度叠加关系，因此牛顿意义下的动量守恒也不再成立了。但爱因斯坦发现，如果我们把动量重新定义成

$$\mathbf{p} = \gamma m_0 \mathbf{v}$$

其中 m_0 为物体静止时的质量，那么动量守恒定律就在洛伦兹变换下成立了！因此，爱因斯坦提出，物体在运动时，其质量会增加为

$$m = \gamma m_0$$

这个质量称为 **相对论质量**，而 m_0 则称为 **静质量**。

这一质量关系式可以推出相对论中一个非常重要的（或许在物理界外最著名的）结论。现在让我们把 γ 展开成速度的函数，并暂时回归国际单位制，重新使用光速 c 。那么上式可以化为

$$m^2 v^2 = (m^2 - m_0^2) c^2$$

而等式左侧就是 \mathbf{p}^2 。我们把它取微分得到

$$\mathbf{p} \cdot d\mathbf{p} = mc^2 dm$$

下面我们来看一个力 \mathbf{F} 做功的结果。在相对论中，力和动量的关系仍然是 $\mathbf{F} = d\mathbf{p}/dt$ 。因此有

$$\mathbf{F} \cdot d\mathbf{r} = \frac{d\mathbf{p}}{dt} \cdot d\mathbf{r} = \mathbf{v} \cdot d\mathbf{p} = \frac{\mathbf{p} \cdot d\mathbf{p}}{m}$$

与上式对比，我们发现力做的功，也就是质点动能的变化量为 $c^2 dm$ 。考虑物体从静止加速，则力做的总功的量即为质点的动能，积分可得

$$E_k = mc^2 - m_0 c^2$$

那么如何理解这个式子呢？爱因斯坦提出， $m_0 c^2$ 是质点静止时所“自带的”能量，称为**静能**，而物体运动时其能量会增加为 $mc^2 = \gamma m_0 c^2$ ，称为**总能**。这样，质量（无论是静质量还是总质量）和能量（对应的静能或者总能）之间的关系为

$$E = mc^2$$

相对论质量的概念现在已被普遍抛弃，其原因简单来说就是它并不是一个物体的特征量，而是随着参考系变化的。

这就是著名的 **质能方程**。它告诉我们，质量和能量只差一个常数倍，可以被看作是同一个物理量。因此，在相对论中，我们只选取一个就可以了。我们一般选择静质量，用 m 表示，和总能量，用 E 表示。而相对论质量和静能的概念不再使用。注意，在这样的定义中，质量是物质的固有属性，与参考系无关；而能量则与参考系有关。这样，我们定义的质量和能量的关系就是

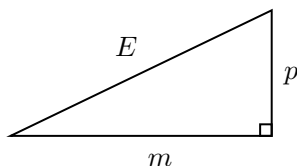
$$E = \gamma mc^2$$

注意此处的 m 的意义已经和之前不同了，它实际上是我们前面说的 m_0 。回到几何单位制，我们就有 $E = \gamma m$ 。这个式子我们以后会经常用到。

在上式中，我们用因子 γ 描述质点的运动。我们若将上式平方，并将 γ^2 展开，很容易得到另一个用动量 p 描述的等价的关系式

$$E^2 = (mc^2)^2 + (pc)^2$$

或者在几何单位制下写为更简洁的 $E^2 = m^2 + p^2$ ，这称为能量-动量关系式。这个关系式也表明，物质的能量是由两部分构成的——物质的静能和运动含有的能量。不过我们这里使用的是动量而非动能，所以二者并非简单的相加关系。能量-动量关系式常常使用下面的“爱因斯坦三角”来帮助记忆。



4.1.2 4-矢量

在牛顿力学中，描述质点动力学的重要物理量包括速度 \mathbf{v} 、动量 \mathbf{p} 、加速度 \mathbf{a} 和力 \mathbf{F} 等。为了在四维的闵氏时空中建立动力学，我们在本节把这些矢量概念迁移到四维，称为 4-矢量。

首先我们来看速度。质点的三维速度的定义为 $v^i = dx^i/dt$ 。类似地，我们可以定义 4-速度。

定义 4.1: 4-速度

质点的 **4-速度** U^μ 定义为其世界线的切矢量，即

$$U^\mu = \frac{dx^\mu}{d\tau}$$

我们用 i, j 等英文字母表示空间分量（也包括整个四维矢量的三个空间分量），仍然用 μ, ν 等字母表示时空分量。

注意我们默认使用线长参数 τ ，否则 4-速度不自动归一化。

注意，这是世界线对于固有时的导数，也就是对线长参数的导数，而非对于坐标时的导数。这个定义完全是张量定义，不涉及任何坐标系，因此 4-速度是一个张量。选定了任意一个坐标系后，4-速度的分量可以写为

$$U^\mu = (\gamma, \gamma u_x, \gamma u_y, \gamma u_z)^\mu$$

其中 u_x, u_y, u_z 是质点在该坐标系下的空间速度。由它的分量计算可知，4-速度是自动归一化的，即 $U^\mu U_\mu = \eta_{\mu\nu} U^\mu U^\nu = -1$ 。这也给了钟慢效应另一个直观理解——质点的空间速度越大，它的“时间方向上的速度”就越小，自然坐标时的增加就会变慢。

有了 4-速度，我们自然就可以定义另一个重要的物理量——动量——的四维矢量。它的定义也和三维定义非常类似。

定义 4.2: 4-动量

质点的 **4-动量** P^μ 定义为

$$P^\mu = mU^\mu$$

注意 m 是静质量。

由于静质量 m 也是一个张量，因此 4-动量也自然是一个张量。在给定坐标系后，它的分量可以写为

与坐标系无关的标量，也就是“标量张量”，也被称为洛伦兹标量。

$$\begin{aligned} P^\mu &= (\gamma m, \gamma m u_x, \gamma m u_y, \gamma m u_z)^\mu \\ &= (E, p_x, p_y, p_z)^\mu \end{aligned}$$

这也常常被写作 $P = (E, \mathbf{p})$ 。这说明，总能 E 相当于质点“沿着时间方向的动量分量”！而且重要的是，我们并不是硬把能量和动量凑成一个四维矢量，它是很自然地定义出来的。而且 4-动量是一个真正的矢量，也就是说它的分量是满足逆变规律的。这更加说明，能量和动量是天生的有机整体。我们后面会看到很多分量形式与此类似的四维张量。

接下来，我们来看加速度的四维矢量。三维空间中，加速度的定义为 $a^i = du^i/dt$ 。在闵氏时空中，我们也可以类似地定义

$$A^\mu = \frac{dU^\mu}{d\tau}$$

这种定义在平直的闵氏时空上的狭义相对论中是可行的，但在后面的弯曲时空上的广义相对论中则不可行，因为在任意流形上并不存在这种导数的定义。为了后面的广义相对论打下基础，我们在此使用微分几何中的导数概念来定义 4-加速度。在闵氏时空中，这种定义方式实际上和上式是等价的。

定义 4.3: 4-加速度

质点的 **4-加速度** A^μ 定义为其 4-速度沿着世界线的协变导数，即

$$A^\mu = U^\nu \partial_\nu U^\mu$$

注意 ∂_μ 是洛伦兹坐标系的偏导数算符，也就是闵氏时空的 Levi-Civita 联络。因此这里的定义中，用 ∂_μ 求导后仍然是一个张量，4-加速度也就是一个张量。

4-加速度的分量表达有些复杂，我们在这里不进行讨论。但 4-加速有两条很直观的性质：首先，质点的 4-加速和 4-速度一定正交，即 $A^\mu U_\mu = 0$ 。这可以直观地理解为，质点的 4-速度大小永远不变，因此只能有法向加速。第二，进行惯性运动的质点的 4-加速为零，这也与三维的运动学类似。

最后，以同样的方式，我们也可以用协变导数来定义质点所受的 4-力。

定义 4.4: 4-力

质点所受的 **4-力** 定义为其 4-动量沿着世界线的协变导数，即

$$F^\mu = U^\nu \partial_\nu P^\mu$$

显然，若质点运动时的静质量 m 不变，则有 $F^\mu = mA^\mu$ ，这也和我们熟悉的牛顿第二定律类似。但注意，我们这里是以动量的变化率作为力的定义，而非一个物理定律。在每个具体物理场景中，给出了力的表达式后，带入本式就可以给出运动方程。例如，在电磁学中，洛伦兹 4-力的表达式为 $F^\mu = qF^\mu{}_\nu U^\nu$ 。因此带电粒子在磁场中的世界线可以描述为 $qF^\mu{}_\nu U^\nu = mU^\nu \partial_\nu U^\mu$ 。

实际上，牛顿第二定律中的力并没有良好定义，因此它更应作为力的定义而非定律。

在给定了坐标系后，4-力的分量可以写为

$$F^\mu = (\gamma \mathbf{F} \cdot \mathbf{u}, \gamma F_x, \gamma F_y, \gamma F_z)^\mu$$

其中 \mathbf{F} 为在该坐标系下的三维力， F_x, F_y, F_z 是其分量， \mathbf{u} 为质点在该坐标系下的三维速度。也就是说，4-力的第一个分量是力的功率，整个矢量也被放缩了 γ 倍。

4.1.3 瞬时观测者与标架

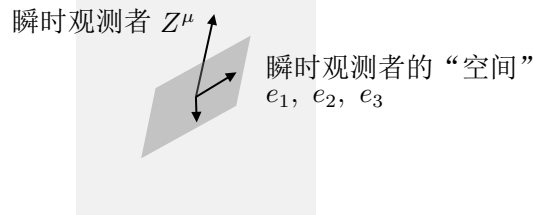
在我们上面的讨论中，每当写出分量时，我们都是取了时空中的一个坐标系。但实际上，我们在许多情况下只是在描述一个点的矢量，我们只需要这个点的切空间的一组基就行了，而不需要流形上的一个坐标系。

考虑若需要观察时空中 p 点的矢量，我们只需要 p 点处的一个观测者即可。而观测者在 p 点的运动状态被其 4-速度描述。因此，我们称这种矢量为一个**瞬时观测者**。也就是说，与观测者（一整条世界线）相比，瞬时观测者只考虑它在某一个时刻的运动状态。瞬时观测者的一个很重要的性质是，瞬时观测者 Z^μ 和所观察的粒子的 4-速度 U^μ 的关系为

$$Z^\mu U_\mu = -\gamma$$

其中 $\gamma = 1/\sqrt{1-u^2}$ 为粒子相对于观测者的 γ 。这个性质在后面会经常用到。

对于瞬时观测者 Z^μ ，这个矢量自然是可以作为该点切空间的“时间轴” e_0 ，再加入与其正交的三个“空间轴” e_i 后，我们就选定了 p 点切空间的一组基。这组基叫做该点的瞬时观测者的一个四维**标架**。



例如，我们上面介绍的 4-速度、4-动量的分量都可以仅通过一个瞬时观测者的标架表示出来，而无需通过一个惯性坐标系。然而，4-加速度和 4-力由于涉及了 p 点附近的矢量，因此无法通过瞬时观测者表示成分量，而需要一整个惯性观测者，也就相当于一个惯性坐标系。

4.2 能量-动量张量

4.2.1 能量-动量张量的定义

在牛顿力学中，最简单的引力源可以抽象为质点。但相对论中，时空中不存在单个的“质点”，物质至少也是以一个一维的世界线的形式存在的。同时，相对论统一了质量和能量，而许多能量则是以连续的场的形式存在的——例如电磁场。因此，在相对论中，我们必须开始研究时空中的连续介质，也就是物质/能量场。

闵氏时空中的质点在某瞬间的动力学性质完全被 4-动量所描述，它表示了质点在时空中的运动方向和“动量”大小。但对于连续介质，4-动量的描述就不再完整了。这是因为，物质之间可以有相互作用，因此物质场中的“一小块物质”不仅可以平移，还可以伸缩、剪切、旋转等。这时，我们需要一个二阶张量来描述物质的动力学性质，这就是能量-动量张量。我们在此先给出定义，再讨论其具体意义。

定义 4.5: 能量-动量张量

物质场的 **能量-动量张量**，简称 **能-动张量**，是一个 $(0,2)$ -型张量场 $T^{\mu\nu}$ ，满足对于任意瞬时观测者 Z^μ ，

$$W^\mu = -T^{\mu\nu} Z_\nu$$

为该瞬时观测者所测得的该点的 4-动量密度，即 $W^\mu = (\mu, w_x, w_y, w_z)^\mu$ ，其中 μ 为该观测者测得的（总）能量密度， w^i 为该观测者测得的动量密度。

能量-动量张量是一个有些难以理解的概念。为了理解这个定义，我们先来看一个在相对论中非常重要的物质场的例子。

例 4.1. 若对于时空中的每一点，该点的物质在统计上可以用一个 4-速度场 U^μ 描述，则这种物质场被称为 **尘埃**。相对尘埃静止的观测者，即 4-速度也为 U^μ 的观测者，称为其 **共动观测者**。设共动观测者测得的尘埃质量密度为 $\rho = dm/d^3x$ ，则尘埃的能动张量为

$$T^{\mu\nu} = \rho U^\mu U^\nu$$

若考虑一些 4-速度不同的尘埃成分的叠加，其能动张量就是各个尘埃成分的能动张量的叠加。此时，总的能动张量就不一定能够写成 $\rho U^\mu U^\nu$ 的形式，因此无法用一个 4-速度来描述。

式中 U^μ 为归一化的 4-速度。光子的能动张量无法如此计算，因为其 4-速度的长度永远为零，无法归一化。

我们以尘埃的能动张量为例，理解一下我们对于能动张量的定义。对于瞬时观测者 Z^ν ，该观测者所测得的 4-动量密度为

$$W^\mu = -\rho U^\mu U^\nu Z_\nu$$

根据 $U^\nu Z_\nu = -\gamma$ ，我们有

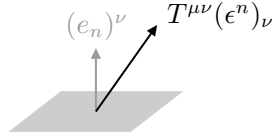
$$W^\mu = \gamma \rho U^\mu$$

定义中的负号就来自这里。

这比尘埃的真正 4-动量密度多了一个 γ 因子。这是由于尺缩效应，在观测者看来，尘埃被“压缩”了，其密度在观测者看来变为了 $\gamma\rho$ 。因此这里的 W^μ 就是该观测者所测得的 4-动量密度。我们可以发现这个 4-动量密度并不是描述物质场本质的一个矢量，它是随着观测者变化的。这也可以解释我们为什么必须使用能动张量而非 4-动量密度来描述物质场的动力学。

看过了能动张量的张量定义，我们下面来看一下在给定了标架后能动张量的分量是什么。现在考虑一个标架，对于其中的第 n 个归一化基矢量 $(e_n)^\nu$ ， $T^{\mu\nu}(\epsilon^n)_\nu$ 就是 $T^{\mu\nu}$ 在这个标架下的分量的第 n 列。这和我们定义中的 $T^{\mu\nu} Z_\nu$ 类似，在这里 $T^{\mu\nu}(\epsilon^n)_\nu$ 可以解释为： $(e_n)^\nu$ 决定了一个垂直于它的小平面， $T^{\mu\nu}(\epsilon^n)_\nu$ 代表了物质场穿过这个小平面的动量流密度。

注意 $(e_n)^\nu$ 是第 n 个基矢量， $(\epsilon^n)_\nu$ 是其对偶基，这里 n 不是矢量的指标。在 $n=0$ 情况下与其降指标后的对偶矢量 $\eta_{\mu\nu}(e_n)^\mu$ 不同。



总结起来，能动张量的分量 $T^{\mu\nu}$ 表示物质场穿过垂直于第 ν 个基矢量的小平面的动量流密度的 μ 分量。与前面所说的 4-动量的时间分量是能量相似，这里在指标为零时的分量也不完全是在空间意义上的动量流密度。我们把各个分量的意义列举如下，注意这些分量都是确定标架的观测者测得的。

- T^{00} 为能量密度；
- T^{i0} 为能流密度的 i 分量；
- T^{0j} 为动量密度的 j 分量；
- T^{ij} 为穿过垂直于第 j 个基矢量的小平面的动量流密度的 i 分量。

另外，能动张量的空间部分，即 T^{ij} 其实就是流体力学和弹性力学中的应力张量。也就是说，能动张量就是有了时间部分的应力张量，其中其对角元称为正应力，也就是压强；非对角元称为切应力。因此，能动张量也被称为 **应力-能量张量**，有时也简称为应力张量。下图总结了能动张量的各个分量的意义。

$$T^{\mu\nu} = \begin{array}{c} \begin{array}{cc} \text{能量} & \text{动量} \end{array} \\ \begin{array}{c} \begin{array}{cccc} T^{00} & T^{01} & T^{02} & T^{03} \\ T^{10} & T^{11} & T^{12} & T^{13} \\ T^{20} & T^{21} & T^{22} & T^{23} \\ T^{30} & T^{31} & T^{32} & T^{33} \end{array} \end{array} \begin{array}{c} \mu\nu \\ \text{密度} \\ \text{流密度} \end{array} \end{array}$$

$\underbrace{\begin{array}{cc} \text{切应力} & \text{正应力 (压强)} \end{array}}_{\text{应力}}$

有了这一点基础，让我们来看一类稍稍复杂的物质场的能动张量。

例 4.2. 一类很重要的物质场称为 **理想流体**，其能动张量定义为

$$T^{\mu\nu} = \rho U^\mu U^\nu + p(U^\mu U^\nu + \eta_{\mu\nu})$$

其中 ρ 为共动观测者测得的质量密度， p 为共动观测者测得的压强。很容易发现，理想流体的能动张量在共动坐标系下的分量为

$$T^{\mu\nu} = \begin{bmatrix} \rho & & & \\ & p & & \\ & & p & \\ & & & p \end{bmatrix}^{\mu\nu}$$

可以发现，理想流体在共动观测者看来只有压强而没有切应力。它实际上对应没有黏性的流体。

4.2.2 能量-动量张量的性质

在例 4.1 中，我们看到，尘埃的能动张量为 $\rho U^\mu U^\nu$ ，它显然是一个对称张量。事实上，任意物质场的能动张量一定是对称的。这就是能动张量的第一个重要性质——对称性。

定理 4.1: 能动张量的对称性

能-动张量是对称张量, 即

$$T^{\mu\nu} = T^{\nu\mu}$$

回到我们上面学到的能动张量的分量的物理意义, 能动张量对称意味着三维应力张量对称, 也意味着能流密度 T^{i0} 和动量密度 T^{0i} 相等. 前者的直观性稍有些难以解释, 但后者的意义非常直观. 动量就是能量 (质量) 的流动, 因此实际上我们上面所说的能流密度和动量密度是同一个东西.

对于我们熟悉的能量、动量和角动量, 在三维中我们有对应的能量守恒、动量守恒和角动量守恒. 这自然也应该体现在四维的能动张量上. 我们先给出下面的结论, 再进行解释.

定理 4.2: 能动张量的散度

对于封闭的物质场有 $\partial_\mu T^{\mu\nu} = 0$. 这体现了任意惯性系下的能量守恒、动量守恒与角动量守恒.

我们以能量守恒为例, 来看一下 $\partial_\mu T^{\mu\nu} = 0$ 为何体现了这个守恒律. 取了任意瞬时观测者 Z^μ 后, 该观测者所测得的 4-动量密度 W^μ 应满足

$$\partial_\mu W^\mu = \partial_\mu (-T^{\mu\nu} Z_\nu) = -Z_\nu \partial_\mu T^{\mu\nu} - T^{\mu\nu} \partial_\mu Z_\nu$$

显然这两项都为零, 所以我们有

$$\partial_\mu W^\mu = 0$$

若我们把时间分量和空间分量分开来写, 则上式变为

$$-\frac{\partial}{\partial t} \mu = \nabla \cdot \mathbf{w}$$

其中 ∇ 是三维的导数算符. 这就是三维语言中的 **连续性方程**. 方程的左侧代表该点处的能量密度降低的速率, 而方程右侧为动量密度 (能流密度) 的散度, 代表了能量从该点流出的速率. 二者相等, 自然意味着能量守恒.

从 $\partial_\mu T^{\mu\nu} = 0$ 还能推出动量守恒和角动量守恒, 但是推导更加复杂一些, 我们在此就不赘述了.

第三部分 广义相对论

5 广义相对论基础

有了狭义相对论的基础，再结合描述弯曲流形的微分几何，我们从这一章开始正式进入广义相对论。

5.1 弯曲时空

5.1.1 从平直时空到弯曲时空

狭义相对论把绝对的伽利略时空观革新为了相对的洛伦兹时空观。它并不是一条物理定律，而是一整个理论体系，几乎所有经典的物理理论，从力学到电磁学，都可以建立在其中——但有一个例外，这就是引力。人们曾试图像电磁理论一样建立与狭义相对论相容的引力理论，但都以失败告终。

转折发生在 1907 年。爱因斯坦意识到了引力的一个不同寻常的性质：任何物体在只受引力作用时的运动轨迹都是相同的。具体来说，以牛顿引力理论为例，质点在引力场中的运动方程为

$$m \frac{d^2 \mathbf{r}}{dt^2} = -G \frac{Mm}{|r|^3} \mathbf{r}$$

我们把等式两侧的质量 m 消去，即可得到质点的运动方程，可以发现它与物质本身的质量、电荷等性质无关。这是我们高中就学过的知识。但爱因斯坦敏锐地发现了一个问题：等式左边的质量，也就是牛顿第二定律 $\mathbf{F} = m\mathbf{a}$ 里的质量 m ，表示的是物质的惯性大小，我们称之为 **惯性质量**。而等式右侧的质量则是引力的荷，它决定了引力的大小，我们称之为 **引力质量**。惯性质量和引力质量的物理意义完全不同，毫无理由相等，但却完全相等。正因为如此，上式中的两个质量才能相消，使得运动方程与质量无关。

惯性质量和引力质量精确地相等，这不应该是一种巧合，而应该蕴含着深刻的物理学原理。因此，爱因斯坦指出，引力和惯性可能是同一种东西。也就是说，物质只受到引力时做的非匀速直线运动，实际上是可能是惯性运动！那么为什么惯性运动会是加速、曲线的运动呢？爱因斯坦指出，在引力场存在时，时空可能会被弯曲。物质在弯曲的时空中做惯性运动时，轨迹应画出一条测地线。而测地线并不一定是局部坐标系中的直线（一次曲线），这使得物质在引力场中看起来做曲线运动。也就是说，引力可能并不是一种力，而是时空弯曲的几何效应！根据这一思想，爱因斯坦在 1915 年终于发表了其定量描述，广义相对论就此诞生。

总结起来，广义相对论可以被精辟地定性描述为下面这句话。

物质告诉时空如何弯曲，时空告诉物质如何运动。

— J. A. Wheeler

5.1.2 弯曲时空中的物理定律

在学习爱因斯坦场方程之前，我们先来讨论一下上面 Wheeler 的话的后半句——时空告诉物质如何运动。这部分并不涉及任何引力理论，它是狭义相对论在弯曲时空中的自然推广。我们说过，狭义相对论的背景时空是四维闵氏时空，它是一个平直的流形。我们已经在这种情况下建立了完整的经典物理学理论（引力除外）。在狭义相对论的理论中，我

电磁学和狭义相对论自然是相容的，因此狭义相对论中已经建立了力学和电磁学。

们涉及了平直闵氏时空的一些几何量，例如度规、导数算符等。而从平直流形推广到弯曲流形，这些几何量会发生变化。因此，我们需要对这些几何量进行改动，替换成弯曲流形中相应的几何量。

我们学过，描述流形的最基本的几何量是度规。在狭义相对论的平直闵氏时空中，其度规为处处相等的闵氏度规 $\eta_{\mu\nu}$ 。而弯曲时空则应用对应的弯曲的度规来描述。而我们的时空的物理意义要求在没有引力——即不弯曲——的情况下退化为平直的闵氏时空，因此时空的度规需要有洛伦兹符号差，即每点的度规在对角化后等价于 $\eta_{\mu\nu}$ 。

度规由 $\eta_{\mu\nu}$ 替换为该时空的度规 $g_{\mu\nu}$

而另外一个常出现在式子中的几何量就是 Levi-Civita 联络。我们在微分几何中学过，它完全由时空上的度规所决定。在狭义相对论中，闵氏度规 $\eta_{\mu\nu}$ 诱导出的 Levi-Civita 联络就是洛伦兹坐标系下的偏导数算符 ∂_μ 。而在弯曲时空中，我们需要将其相应替换成相应的联络。也就是

联络由 ∂_μ 替换为度规 $g_{\mu\nu}$ 诱导出的 Levi-Civita 联络 ∇_μ

我们下面来看几个具体的例子。

例 5.1. 我们在狭义相对论中定义了质点的动力学量。例如，质点的 4-速度为其世界线的切矢量 $U^\mu = \frac{dx^\mu}{d\tau}$ 。这个定义中不涉及任何关于背景时空的几何量，因此可以不加改动直接移植到弯曲时空中。而 4-加速度的定义为 4-速度的协变导数 $A^\mu = U^\nu \partial_\nu U^\mu$ 。在弯曲时空中， ∂_ν 应被替换为 ∇_ν ，因此 4-加速度的定义变为

$$A^\mu = U^\nu \nabla_\nu U^\mu$$

例 5.2. 再举一个例子。我们学过能动张量的协变散度 $\partial_\mu T^{\mu\nu} = 0$ ，这意味着能量守恒、动量守恒和角动量守恒。而在弯曲时空中，此式自然地变为

$$\nabla_\mu T^{\mu\nu} = 0$$

以此方式，我们可以把狭义相对论中所有的物理定律移植到弯曲时空中。这涉及了所有的经典的动力学和电磁学，只有引力除外——引力是体现在时空弯曲里的。也就是说，现在我们已经完全在理论上建立了“时空告诉物质如何运动”的物理理论——我们已经可以在弯曲时空中讨论除引力以外的所有经典物理定律了。

在此需要指出的是，有的时候我们从狭义相对论的不同式子向广义相对论移植后，会得到不相容的结果。这是因为在狭义相对论中，式子之间相容可能依赖于平直时空的一些特殊性质，例如联络的对易性，而这在一般的弯曲时空中是不成立的。在此时，我们应该移植过来的是更加基础的式子，而非经过了一些推导的“二级结论”。

5.2 广义协变、等效原理与坐标系

在定量寻找引力场方程之前，我们先跟随爱因斯坦的心路历程，讨论两个启发爱因斯坦提出广义相对论的两个物理学原理。

5.2.1 广义协变与等效原理

在之前的狭义相对论中，我们曾讲过洛伦兹相对性原理：物理定律在一切惯性参考系中具有相同的形式，即一切惯性系平权。在弯曲时空的广义相对论中，爱因斯坦将其推广

为广义相对性原理，又称为 **广义协变性原理**：物理定律在一切参考系中具有相同的形式，即一切参考系平权，无论该参考系是否为惯性系。有时这个原理也称为微分同胚协变性，即物理定律的形式在任何微分同胚下不变。

这个原理初一听可能令人摸不着头脑——惯性系和非惯性系怎么会平权呢，毕竟非惯性系下物质会受到惯性力的作用。为了解释广义协变性的意思，我们在此先给出另一个显然等价的定义：**物理定律与参考系无关**。也就是说，我们常见的动力学量，如力、动量等，只能以张量的形式出现在物理定律中。而引力只能以协变的几何量出现，包括度规 $g_{\mu\nu}$ 及其衍生量，如 Levi-Civita 联络 ∇_μ 等。任何不满足协变的量，例如某个特定的基矢量、偏导数算符、Christoffel 符号等都不允许出现在方程中，因为它们依赖于特定的坐标系，这会使得这个坐标系与其他坐标系不再平权。例如，在弯曲时空中，质点运动定律可以写成

$$F^\mu = mU^\nu \nabla_\nu U^\mu$$

其中所有的物理量都以张量描述，而引力体现在联络 ∇_ν 中。那么为了解释“物理定律在一切惯性参考系中具有相同的形式”，我们把它在给定局部坐标系 $\{x^\mu\}$ 后写成下面的分量形式，当然此时的分量形式是会出现偏导数和 Christoffel 符号的。

$$F^\mu = \gamma^2 m \frac{d^2 x^\mu}{dt^2} + \gamma^2 m \Gamma^\mu_{\nu\sigma} \frac{dx^\nu}{dt} \frac{dx^\sigma}{dt}$$

此处的两个 γ 来自两个从 $d\tau$ 向 dt 的转换，也可以看作一个来自 4-力与三维力的倍数关系，一个来自相对论质量。

其中第一项就是我们熟悉的相对论性的“牛顿第二定律”，而后一项则表示引力的影响。因此，引力是体现在几何量在该坐标系下的表示，如度规的分量 $g_{\mu\nu}$ 和联络在该坐标下的 Christoffel 符号 $\Gamma^\sigma_{\mu\nu}$ 中的。而广义协变性要求这些项在任何坐标系下都可以写成这种形式，自然对于这些量在坐标变换下的分量变换提出了约束，也就是对引力场的动力学提出了约束。

我们在前面讨论过，启发爱因斯坦提出广义相对论的另一个假设是惯性质量和引力质量的等价性，这使得物质在引力场中沿相同的轨迹运动。现在我们来换一个角度看这个问题。我们知道，在牛顿动力学中，物体在非惯性参考系中可以被看作受到一个假想的惯性力 $\mathbf{F}_I = -m\mathbf{a}$ 的作用。而由于引力质量和惯性质量的等价性，它与一个强度为 $-\mathbf{a}$ 的引力场所产生的引力完全等效。因此，任何力学实验都无法区分该参考系是在做非惯性运动，还是处于引力场中。在狭义相对论中亦是如此。这称为 **弱等效原理**。而爱因斯坦把它推广为“一切物理实验都无法区分该参考系是在做非惯性运动，还是处于引力场中”，这个推广的等效原理称为 **爱因斯坦等效原理**。

爱因斯坦等效原理是所有关于广义相对论的科普都会提到的原理。它看似十分简单，但若不仔细思考却容易造成概念上的混淆。我们知道，引力来自时空的弯曲，而时空的弯曲由黎曼曲率张量描述。这是关于背景时空的绝对的性质，与参考系无关。也就是说，非惯性参考系并不会导致时空的弯曲。那么二者为什么会等价呢？正因如此，物理学家们对爱因斯坦等效原理褒贬不一。一些物理学家认为等效原理的定义并不清楚，或对广义相对论并没有实质作用，甚至一些人认为它在一些情况下是错误的。我们在此不去评价爱因斯坦等效原理的地位，但是我们需要解释一下二者的等价性到底是什么意思。

我们同样以局域坐标系的质点运动定律的例子，来解释等效原理中参考系的加速度和引力的等效性。在式子

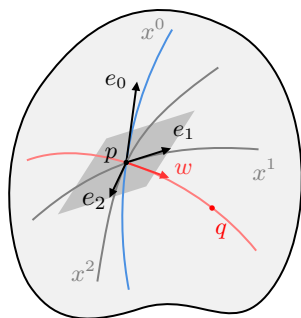
$$F^\mu = \gamma^2 m \frac{d^2 x^\mu}{dt^2} + \gamma^2 m \Gamma^\mu_{\nu\sigma} \frac{dx^\nu}{dt} \frac{dx^\sigma}{dt}$$

中，注意到多出来的“引力”项完全来自于该参考系下的 Christoffel 符号。而这一项可能来自时空的弯曲，也就是真正的引力，也可能来自参考系的对惯性系的偏离——同样的时空在不同参考系下的 Christoffel 符号是可以不同的，哪怕平直时空在非惯性参考系下也可以有非零的 Christoffel 符号，我们可以把这种来源的“引力”项称为“表观引力”。二者的本质是不同的，但既然它们最终只通过 Christoffel 符号体现在该参考系的物理定律中，那么自然就在该参考系中无法分辨。

5.2.2 弯曲时空中的坐标系

相对论中描述物理定律的自然语言是四维语言。但我们常常会使用坐标系把物理定律写成分量的形式，以得到和我们的生活经验一致的、分离的时间与空间。在狭义相对论中，我们一般使用惯性坐标系，因为这是平直时空中最简单的坐标系。但在弯曲时空中，并不存在如此简单的惯性坐标系。不仅如此，广义协变原理也要求我们可以在任意的参考系中描述物理定律。因此，我们在这里需要讨论弯曲时空中的一般的坐标系。

考虑任意一个观测者，其世界线为 $G(\tau)$ ，且该观测者携带一个正交归一的 4-标架 $\{e_\mu\}$ ，则该观测者的空间标架基矢量向周围的测地延伸可以形成三个空间坐标轴，再加上世界线本身作为时间轴，这个观测者会在其世界线周围建立一个局域坐标系，称为该观测者的固有坐标系。



固有坐标系使得我们让一个观测者可以描述他周围所发生的事件。注意，固有坐标系是一个局域坐标系，它的坐标域不一定能延伸到整个时空。这既和时空的结构有关，也和该观测者的测地线有关。但我们经常只用固有坐标系来描述距离世界线很近的事件，因为在小范围内度规分量近似与观测者世界线上相同，因此有近似相同的物理定律。

由于我们选取的标架是正交归一的，因此度规在坐标系原点的值 $g_{\mu\nu}$ 就是该坐标系的闵氏度规 $\eta_{\mu\nu}$ 。然而，Christoffel 符号和曲率含有度规的微分，因而未必为零。所以在任意观测者的固有坐标系下，物理定律的分量形式可能会有这些几何量的分量出现。等效原理告诉我们，这相当于这个坐标系中有引力的存在。我们在上一小节已经通过质点运动定律的例子对此给出了解释。

在狭义相对论中，我们主要讨论惯性系中的物理定律。在广义相对论中，我们也希望能有一种与狭义相对论中惯性系类似的坐标系，最好性质也一样好。弯曲时空中的“惯性观测者”自然就是（除引力外）不受力的观测者，或者说世界线为测地线的观测者。由于这样的观测者只可能受引力，因此也称为自由下落观测者。

除了要求世界线为测地线外，我们还需要要求这个观测者所带的标架没有发生空间转动。向量的转动可以由狭义相对论中的转动定律推广而来。在弯曲时空中的“无转动”即为下面的 F-W 平动。

但其实这里任意观测者的固有坐标系也可以用在狭义相对论中

定理 5.1: Fermi-Walker 平动

若世界线 $x^\mu(\tau)$ 上有法向的空间矢量场 v^μ ，则 v^μ 沿世界线运动无空间转动的充要条件为

$$U^\nu \nabla_\nu v^\mu + (A^\mu U^\nu - A^\nu U^\mu) v_\nu = 0$$

其中 U^μ 为世界线的 4-速度， A^μ 为其 4-加速度。此时我们称矢量场 v^μ 沿该世界线 **Fermi-Walker 平动**，简称 F-W 平动。

等式左侧称为 v^μ 沿世界线的 F-W 导数。

这样，我们就知道了观测者运动时，其上的标架该如何“无转动地”一起移动。此时的观测者称为 **自由下落无自转观测者**。这种观测者是在弯曲时空中我们可以定义的最接近惯性观测者的观测者，它所定义的坐标系自然就是最接近狭义相对论中惯性系的坐标系。我们称自由下落无自转观测者的固有坐标系为 **局部惯性系**，局部惯性系就是我们前面讲过的黎曼法坐标系。

在局部惯性系中，该观测者测地线上的物理定律的分量形式等价于狭义相对论中惯性系描述的物理定律，这是上面所讲的等效原理所指出的。我们可以发现，在局部惯性系下，观测者世界线上不仅度规的分量 $g_{\mu\nu} = \eta_{\mu\nu}$ ，还有有联络 ∇_μ 就是局部惯性系下的偏导数算符 ∂_μ ，或者说有 Christoffel 符号 $\Gamma^\sigma_{\mu\nu} = 0$ 。因此，在用局部惯性系写物理定律的分量形式时，我们无需进行上一节所讨论的度规和联络的替换，而可以保留和狭义相对论相同的形式。亦或者说，在局部惯性系下，观测者世界线上是观察不到引力的存在的，这也就是等效原理。需要注意的是，这个性质只在观测者的测地线上严格成立。不过有时我们也可以认为它在测地线附近近似成立。

5.2.3 测地偏离方程

我们讲到，在局部惯性系中，在观测者的世界线上，物理定律与狭义相对论中惯性系的物理定律相同，因此该观测者无法通过任何物理实验区分他是否处于引力场中。在世界线附近，我们也说这种性质近似成立。但实际上，由于周围时空几何的微小差异，在世界线附近不同位置的物质所受的“引力”实际上还是有细微的差别的。这种差异的结果将会使一个有大小的物体上的各个点受力不同，从而将其进行拉伸。这种拉伸所产生的力被称为 **潮汐力**。

在广义相对论中引力不再是力，但我们这里仍然使用力来作为一个类比。

几何上，不同的点受到引力不同的本质是它们的测地线不会保持平行，而是会弯曲而相互靠近或远离。于是我们可以通过描述相邻测地线的弯曲差异——也就是相邻测地线之间距离的二阶导数，来描述潮汐力。这个方程称为测地偏离方程，是广义相对论的一个重要方程。

定理 5.2: 测地偏离方程

设 $\gamma_s(\tau)$ 是一个单参类时测地线族， $U^\mu = \frac{\partial x^\mu}{\partial \tau}$ 为测地线的切矢量， $X^\mu = \frac{\partial x^\mu}{\partial s}$ 为相邻测地线之间的偏离矢量。定义潮汐加速度为偏移矢量沿测地线的二阶协变导数 $A^\mu = U^\beta \nabla_\beta (U^\alpha \nabla_\alpha X^\mu)$ ，则有

$$A^\rho = R^\rho_{\sigma\mu\nu} U^\sigma U^\mu X^\nu$$

其中 $R^\rho_{\sigma\mu\nu}$ 为黎曼曲率张量。

测地偏离方程中直接出现了黎曼曲率张量。而“潮汐力”，即相邻点的引力之差，可以视作 mA^μ 。我们之前只是定性地讨论了引力应该是时空的弯曲，而测地偏离方程则直接将引力与时空的曲率联系在了一起。这为我们后面用曲率描述引力提供了直接的数学支持。

5.3 引力的几何描述

在上面一节中，我们讨论了在弯曲时空中描述物理定律的基础。本节我们来讨论广义相对论的另一部分，即 Wheeler 所说的前半句，“物质告诉时空如何弯曲”。

5.3.1 牛顿引力几何化

广义相对论中使用几何描述引力。为了找到广义相对论的引力场方程，我们先从我们熟悉的牛顿引力论开始，尝试把它改造成一个几何理论，看看它对我们有什么启发。我们在高中学了牛顿万有引力公式

$$F = G \frac{Mm}{r^2}$$

下面我们使用几何单位制，即约定牛顿引力常量 $G = 1$ 。同时，为了方便后续与广义相对论比较，我们可以仿照静电场的写法，将上式变形为

$$\nabla^2 \Phi = 4\pi\rho$$

其中 Φ 称为引力势，它满足引力加速度 $\mathbf{a} = \nabla\Phi$ ， ρ 为质量密度， ∇^2 为三维的 Laplace 算符。等式的右侧描述了物质的分布，而等式的左侧则描述引力场的性质。然而，这里的引力场是使用引力势描述的，而非用时空的弯曲描述。为了将这个式子与时空几何联系在一起，让我们考虑一个在引力场作用下自由下落的质点，其运动由测地线方程描述

$$\frac{d^2 x^\mu}{d\tau^2} = -\Gamma^\mu_{\nu\sigma} \frac{dx^\nu}{d\tau} \frac{dx^\sigma}{d\tau}$$

在牛顿的低速近似下，4-速度的时间分量远大于空间分量，因此我们可以在第二项中只保留时间分量，变为

$$\frac{d^2 x^\mu}{d\tau^2} = -\Gamma^\mu_{00} \left(\frac{dx^0}{d\tau} \right)^2$$

注意到 x^0 就是 t ，于是我们可以将等式两侧乘上 $(d\tau/dt)^2$ ，得到

$$\frac{d^2 x^\mu}{dt^2} = -\Gamma^\mu_{00}$$

回忆起我们可以从度规出发计算 Christoffel 符号。计算上面的 Christoffel 符号分量可得 $\Gamma^\mu_{00} = -\frac{1}{2}g^{\mu\nu}\partial_\nu g_{00}$ 。且在牛顿的弱场近似下有 $g_{\mu\nu} \approx \eta_{\mu\nu}$ ，带入得到

$$\frac{d^2 x^\mu}{dt^2} = \frac{1}{2}\eta^{\mu\nu}\partial_\nu g_{00}$$

考虑物质的分布不随时间变化，则牛顿引力下的引力场也是不随时间变化的，即 $\partial_0 g_{00} = 0$ 。于是上式中 $\mu = 0$ 的项消去，只保留空间指标。而这样 ν 也只能为空间指标，否则 $\eta^{\mu\nu}$ 分量为零。这样我们只剩下空间部分。而注意到只保留空间分量的 η^{ij} 就是 δ^{ij} ，这样上式化简为

$$\frac{d^2 x^i}{dt^2} = \frac{1}{2}\partial_i g_{00}$$

此时可以发现等式左侧就是我们熟悉的加速度 a^i ，它应该是引力势 Φ 的三维梯度。而等式右侧就是 $\frac{1}{2}g_{00}$ 的梯度，也就是说这就是引力势。而同时又注意到能量密度 ρ 就是 T_{00} 。于是我们就得到了几何化的牛顿引力场方程

$$\nabla^2 g_{00} = 8\pi T_{00}$$

或许你会觉得这只是数学上的奇技淫巧，然而这种几何化的引力场方程将会引领我们找到真正的相对论性的引力方程。

5.3.2 爱因斯坦场方程

引力是由物质场的分布决定的。在牛顿动力学中，物质场的分布由其质量的分布描述。而在相对论中，对于物质场的完整描述需要使用能动张量。同时，完整的几何表述需要我们要描述时空的度规。因此，我们需要把牛顿引力论升级为由能动张量和几何量构成的张量方程。一个很直接的想法就是把度规和能动张量全部从 00 分量升级为完整的张量，即 $\nabla^\sigma \nabla_\sigma g_{\mu\nu} = 8\pi T_{\mu\nu}$ 。然而，由于度规在其相容的导数作用下为零，进而等式左侧必为零，因此这个方程显然无法成立。于是，我们需要对这个等式里的张量放松要求——一个很自然的想法是，等式左侧不一定要是度规的二阶梯度，而可以是某个二阶微分的几何量（当然也可以包括一阶微分和度规本身），即

$$\text{某个二阶微分的几何量} = \kappa T_{\mu\nu}$$

其中 κ 是比例常数。可以发现，我们之前学过的曲率张量就是二阶微分的几何量。同时，由于指标的对应性，我们需要找一个二阶张量。于是 Ricci 张量 $R_{\mu\nu}$ 就是一个很好的候选。同时，能动张量是对称的，而 Ricci 张量也是对称的。因此这或许会构成一个合理的方程！事实上，爱因斯坦在 1915 年的确发表了 $R_{\mu\nu} = \kappa T_{\mu\nu}$ 作为第一版的引力场方程。然而人们不久以后发现这个方程也有问题。能动张量的协变散度 $\nabla^\mu T_{\mu\nu} = 0$ ，但 Ricci 张量并不一定。如果我们强行要求 $\nabla^\mu R_{\mu\nu} = 0$ ，则会推出物理上不可接受的结果——在非相对论极限下可以推出物质场的能量密度处处相等！因此这也不是一个合理的引力场方程。

虽然上面两次尝试都没有找到合理的引力场方程，但我们在试错中已经找到了等式左侧的几何量应满足的性质——它应该某个二阶微分的几何量，且是对称的、协变散度为零的二阶张量。爱因斯坦在几个月后终于找到了满足这个性质的一个张量，称为 **爱因斯坦张量**，定义为

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$$

爱因斯坦张量是一个对称张量，且满足 $\nabla^\mu G_{\mu\nu} = 0$ ，这称为缩并的 Bianchi 恒等式。因此它可以作为等式左侧的几何项，使其正比于物质场的能动张量。这样我们就得到了一个可行的描述引力场的方程，称为爱因斯坦场方程。

物理定律 5.1: 爱因斯坦场方程 (EFE)

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = \kappa T_{\mu\nu}$$

其中 κ 称为爱因斯坦引力常量，我们在下一章中会看到 $\kappa = 8\pi$ 。另外，可以看出，我们并不是严格地推导出了引力场方程，而是通过与牛顿引力论的类比猜出了，或者说构造出了一个引力场方程。实际上，符合这些要求的引力场方程并不止爱因斯坦场方程这一

我们常常要求引力势在无穷远处为零，此时引力势为

$$\Phi = -\frac{1}{2}(1 + g_{00})$$

∇^2 在四维中的对应算符称为 d'Alembert 算符，定义为

$$\square = \nabla^\mu \nabla_\mu$$

回忆起我们在 2.2.5 节学过，黎曼曲率张量的作用相当于联络的对易子

$\nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu$ ，或者从另一个角度说，黎曼曲率张量的分量是 Christoffel 符号的一阶微分的量，而 Christoffel 符号则是度规的一阶微分的量，总共有二阶微分。

在标准单位制中

$$\kappa = \frac{8\pi G}{c^2}$$

其中 G 为牛顿引力常量。

个，自然也就无法唯一地推导出爱因斯坦场方程。但爱因斯坦场方程是其中最为简洁的一个，而且它经受住了实验的检验，因此它目前是一个非常成功的引力理论。

不过也有一些更复杂的方程目前并没有被实验推翻，它们还是广义相对论的竞争对手。

5.3.3 爱因斯坦场方程的几点说明

爱因斯坦场方程是关于时空几何的方程，也就是说我们需要解出来的张量是度规 $g_{\mu\nu}$ ，而非曲率，因为度规完整地刻画了时空的几何结构。注意到 $R_{\mu\nu}$ 是 $g_{\mu\nu}$ 的二阶微分的函数，因此整个场方程实际上是一个二阶非线性偏微分方程组。但爱因斯坦场方程的求解与一般的方程求解略有区别，我们在此给出一定的说明。

首先，爱因斯坦场方程描述物质分布引起的时空弯曲。所以看似正常思路是给定描述物质分布的 $T_{\mu\nu}$ ，我们就可以求解度规 $g_{\mu\nu}$ 。但注意到时空几何是由度规控制的，而 $T_{\mu\nu}$ 作为时空中的物理量，是无法独立于度规存在的。因此，我们常常无法“给定 $T_{\mu\nu}$ ”。例如，考虑最简单的物质分布——尘埃，其能动张量为 $T_{\mu\nu} = \rho U_\mu U_\nu$ ，而 U_μ 是由 4-速度 U^μ 降指标得到的，这需要度规。即使我们把整个方程升指标，变为 $G^{\mu\nu} = 8\pi T^{\mu\nu}$ ，此时我们虽然无需对 4-速度降指标，但它仍然应该满足归一化条件 $g_{\mu\nu} U^\mu U^\nu = -1$ ，而这也需要度规。因此在没有度规时，我们往往无法给出能动张量。在求解爱因斯坦场方程时，我们往往是给定一些束缚条件后，将 $g_{\mu\nu}$ 与 $T_{\mu\nu}$ 一同求解出来。

第二，爱因斯坦场方程是无法唯一确定度规 $g_{\mu\nu}$ 的。注意到 $g_{\mu\nu}$ 是一个四维对称张量，有 10 个自由变量，因此我们需要 10 个独立的方程来确定它。方程中的 $G_{\mu\nu}$ 和 $T_{\mu\nu}$ 也都是四维对称张量，看似也有 10 个自由变量。但 $G_{\mu\nu}$ 实际上满足缩并的 Bianchi 恒等式 $\nabla^\mu G_{\mu\nu} = 0$ 。这对 $G_{\mu\nu}$ 给出了 4 个束缚，因此它只有 6 个自由变量。也就是说，我们实际上只有 6 个独立的方程，这是无法唯一确定 10 个独立分量的 $g_{\mu\nu}$ 的。这也可以从另一个角度理解——若两个度规场之间只相差一个坐标变换，那么它们是等价的，因为坐标变换不影响时空的曲率。而坐标变换有 4 个自由度，因此我们在选择度规时也有 4 个额外的自由度。这种保持爱因斯坦场方程的解的变换称为 **规范变换**，而爱因斯坦场方程的这种性质称为 **规范自由性**。

我们这里以坐标分量变换解释的 $g_{\mu\nu}$ 的变换，但这其实也可以看作时空上的一个微分同胚，二者是被动观点和主动观点的差别。

从坐标变换的角度理解也给了我们一种解决这个问题的思路。既然坐标系的选择是随意的，那么我们就对坐标系进行束缚，这种束缚称为 **坐标条件**。一种常见的坐标条件称为高斯坐标条件，它要求时间轴是归一的，且时空是正交的，即

$$g_{00} = -1, \quad g_{0i} = 0$$

另一种常用的坐标条件称为调和坐标条件，它令坐标系满足

$$\square x^\mu = 0$$

可以发现两个坐标条件有 4 个分量方程，也就是提供了额外的 4 个束缚方程，使得度规可以唯一地被确定。调和坐标条件常常也被写作分量形式 $g^{\mu\nu} \Gamma^\sigma_{\mu\nu} = 0$ 。

5.3.4 真空场方程

引力场是由物质场产生的。但在真正考虑引力问题是，我们经常考虑的是物质场的外部——例如对于一个天体，我们常常考虑的是天体外部的引力场，而天体外部的绝大部分空间实际上是真空的。因此，我们想知道在没有物质场的时空区域，时空的几何性质会是什么样子的。事实上，这个问题有一个更为简单的方程。下面我们来推导一下。

我们首先补充一点基本知识. 考虑由 Ricci 张量 $R_{\mu\nu}$ 得到爱因斯坦张量 $G_{\mu\nu}$ 的变换, 这实际上是一种很常见的张量变换, 我们称之为 **迹反转**. 也就是说, 对于一个四维时空中的一个张量 $S_{\mu\nu}$, 我们定义其迹反转张量 $\bar{S}_{\mu\nu}$ 为

$$\bar{S}_{\mu\nu} = S_{\mu\nu} - \frac{1}{2}Sg_{\mu\nu}$$

很容易发现 $\bar{S}_{\mu\nu}$ 的缩并为 $\bar{S} = S - \frac{1}{2}SD = -S$, 其中 $D = 4$ 是时空的维数. 也就是说 $\bar{S}_{\mu\nu}$ 的缩并 (也就是迹) 与 $S_{\mu\nu}$ 的缩并相反, 因此将这个过程称为迹反转.

注意度规的缩并
 $g^{\mu\nu}g_{\mu\nu} = \delta^\mu_\mu = D$.

迹反转的一个重要的性质是两次迹反转会得到原来的张量, 即

$$\bar{\bar{S}}_{\mu\nu} = S_{\mu\nu}$$

因此可以说, 迹反转的张量 $\bar{S}_{\mu\nu}$ 携带了与 $S_{\mu\nu}$ 相同的信息. 若要描述张量 $S_{\mu\nu}$ 所满足的方程, 我们可以等价地描述 $\bar{S}_{\mu\nu}$ 所满足的方程. 我们下面就来看一看这样做的效果.

有了迹反转的定义, 下面我们来推导爱因斯坦场方程的一个等价形式. 我们将爱因斯坦场方程的两侧都进行迹反转, 左侧会回到 Ricci 张量, 而右侧就是迹反转的能动张量, 即

$$R_{\mu\nu} = \kappa \left(T_{\mu\nu} - \frac{1}{2}Tg_{\mu\nu} \right)$$

通过同样迹反转方式, 可以将上式推回爱因斯坦场方程, 因此上式是爱因斯坦场方程的等价形式, 称为迹反转的爱因斯坦场方程. 有时使用这个方程会更加方便.

利用迹反转的爱因斯坦场方程, 我们就可以很容易地发现, 在 $T_{\mu\nu} = 0$ 的真空中该方程化为

$$R_{\mu\nu} = 0$$

即真空中 Ricci 张量为零. 这个等式也称为 **真空爱因斯坦场方程**. 注意真空爱因斯坦场方程适用于任何没有物质场的时空区域, 而非要求整个宇宙没有物质的存在. 由于旁边可能有物质场的存在, 所以时空并不一定是平直的, 即黎曼曲率张量 $R^\rho_{\sigma\mu\nu}$ 并不一定为零. 但我们有 Ricci 曲率 $R_{\mu\nu} = 0$, 此时时空称为 Ricci 平直时空.

6 线性近似与引力波

我们已经看到，描述引力场的爱因斯坦场方程是一个高度非线性的方程，它的求解往往十分困难。然而，我们知道，在变量的变化足够小时，任何表达式都可以近似为线性的。在引力场较弱时，爱因斯坦场方程也可以化为一个线性方程。本章我们来看一下弱场近似下引力场的性质。

6.1 弱场近似

6.1.1 线性引力论

在讨论引力问题时，我们常常会计算一些非极端相对论条件下的引力场，例如一个质量不太大的天体周围、或距离天体较远处的时空。在这些情况下，引力场在相对论意义下是比较弱的。此时，时空相对于平直的闵氏时空只有较小的偏离。因此，我们可以将度规进行分解

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$$

其中 $h_{\mu\nu}$ 是一个小量。这样，我们可以把背景时空看作闵氏时空，而把 $h_{\mu\nu}$ 看作其上的代表引力的动力学量。我们说可以把时空看作闵氏时空，是指我们可以把度规近似为闵氏度规 $\eta_{\mu\nu}$ ，用它求内积或进行指标升降；同时我们可以在时空上选择洛伦兹坐标系，将联络近似为洛伦兹坐标系的偏导数算符 ∂_μ 。在这些量中， $h_{\mu\nu}$ 的影响都是小量。但在一些涉及时空弯曲的量中，若整个量都是小量，则我们需要保留最低阶部分，例如 Christoffel 符号近似为

$$\Gamma^\sigma_{\mu\nu} = \frac{1}{2}\eta^{\sigma\rho}(\partial_\mu h_{\nu\rho} + \partial_\nu h_{\mu\rho} - \partial_\rho h_{\mu\nu})$$

由 Christoffel 符号，我们可以进一步求出曲率。黎曼曲率张量有 $\partial\Gamma + \Gamma\Gamma$ 的形式。我们可以忽略二阶部分，得到一阶近似下的黎曼曲率张量，称为线性黎曼曲率张量

$$R_{\rho\sigma\mu\nu} = \frac{1}{2}(\partial_\mu\partial_\sigma h_{\rho\nu} + \partial_\nu\partial_\rho h_{\sigma\mu} - \partial_\nu\partial_\sigma h_{\rho\mu} - \partial_\mu\partial_\rho h_{\sigma\nu})$$

由线性黎曼张量，我们可以进一步求出线性爱因斯坦张量

$$G_{\mu\nu} = \frac{1}{2}(\partial^\sigma\partial_\mu h_{\nu\sigma} + \partial^\sigma\partial_\nu h_{\mu\sigma} - \partial_\mu\partial_\nu h - \square h_{\mu\nu} - \eta_{\mu\nu}\partial^\sigma\partial^\rho h_{\sigma\rho} + \eta_{\mu\nu}\square h)$$

其中 $\square = \partial^\mu\partial_\mu$ 为洛伦兹坐标系下的 d'Alembert 算符， $h = \eta^{\mu\nu}h_{\mu\nu}$ 为 $h_{\mu\nu}$ 的缩并。线性爱因斯坦张量虽然是一种近似，但它仍然保持了我们希望 $G_{\mu\nu}$ 所有的性质，即它是个对称张量，且其散度 $\partial^\mu G_{\mu\nu} = 0$ 。令线性爱因斯坦张量等于 $8\pi T_{\mu\nu}$ 即可得到 **线性爱因斯坦场方程**，它成为了 $h_{\mu\nu}$ 的线性方程。

不过这个方程还是有些复杂，我们可以把它进行简化。我们定义迹反转的度规微扰张量

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h$$

由于 $\bar{h}_{\mu\nu}$ 与 $h_{\mu\nu}$ 是线性关系，因此爱因斯坦张量也是 $\bar{h}_{\mu\nu}$ 的线性函数。此时爱因斯坦张量可以简化为四项

$$G_{\mu\nu} = \frac{1}{2}(\partial^\sigma\partial_\mu\bar{h}_{\nu\sigma} + \partial^\sigma\partial_\nu\bar{h}_{\mu\sigma} - \square\bar{h}_{\mu\nu} - \eta_{\mu\nu}\partial^\sigma\partial^\rho\bar{h}_{\sigma\rho})$$

我们马上会看到，通过另一种变换，我们可以进一步简化线性爱因斯坦场方程。

指标升降中只有 $g^{\mu\nu}$ 是例外，它仍然定义为 $g_{\mu\nu}$ 的逆，而非 $g_{\mu\nu}$ 用 $\eta^{\mu\nu}$ 升指标后的张量。

6.1.2 线性引力论的规范变换

我们知道，爱因斯坦场方程具有规范自由性。这一点在线性引力论中仍然存在。我们来具体推导一下线性引力论中规范变换的形式。对于一个平直的时空，我们可以有很多种方式选择近似的洛伦兹坐标系。如果我们引入这些洛伦兹坐标系之间的一个坐标变换

$$x^\mu \rightarrow x^\mu + \xi^\mu$$

其中 ξ^μ 是任意一个“变化缓慢”的矢量场，即 $\partial_\nu \xi^\mu$ 是小量（至少与 $h_{\mu\nu}$ 同阶小）。这样才可以保证新的坐标系仍然近似为一个洛伦兹坐标系。在这个坐标变换下，我们可以推出 $h_{\mu\nu}$ 的分量变换为

$$h_{\mu\nu} \rightarrow h_{\mu\nu} - \partial_\mu \xi_\nu - \partial_\nu \xi_\mu$$

即新的坐标系下的 $h_{\mu\nu}$ 分量等于老的坐标系下的 $h_{\mu\nu} - \partial_\mu \xi_\nu - \partial_\nu \xi_\mu$ 。这称为线性引力论中的规范变换。可以看到我们上面对 ξ^μ 的要求使得 $h_{\mu\nu}$ 在变换后仍然为小量，因而仍然可以用线性近似来处理它。

规范变换是时空上的坐标系变换，而时空本身并没有发生变化。所以在规范变换下，线性黎曼曲率张量不发生变化，进而线性爱因斯坦张量也不变。因此，规范变换可以保持线性爱因斯坦场方程的解。这种规范自由性使得我们可以选择一个较为方便的坐标系来处理问题。可以证明，在规范变换下，我们一定可以变换到一个调和坐标系，即满足 $\square x^\mu = 0$ 。此时很容易推出

$$\partial^\mu \bar{h}_{\mu\nu} = 0$$

这称为 **洛伦兹规范** 或 **调和规范**。因此爱因斯坦张量只保留一项 $G_{\mu\nu} = -\frac{1}{2}\square \bar{h}_{\mu\nu}$ ，此时线性爱因斯坦场方程进一步被简化为

$$\square \bar{h}_{\mu\nu} = -16\pi T_{\mu\nu}$$

这就是线性爱因斯坦场方程的最简形式。

6.1.3 牛顿近似

在牛顿近似，即弱场、低速近似下，爱因斯坦场方程应该退化为牛顿万有引力定律。在 5.3.1 中，我们把牛顿引力论写成了几何形式。这与爱因斯坦场方程的形式稍有类似，但等式左侧的几何项仍然与爱因斯坦张量的分量有别。因此我们并没有把二者严格的联系起来。在本节，我们更加完整地推导一下爱因斯坦场方程的牛顿近似，并给出爱因斯坦场引力常量 $\kappa = 8\pi$ 的来源。

牛顿近似首先是弱场近似，因此我们从线性爱因斯坦场方程出发。在此我们假装不知道 κ 的值，此时最简的爱因斯坦场方程为

$$\square \bar{h}_{\mu\nu} = -2\kappa T_{\mu\nu}$$

在非相对论的低速近似下，由于物质场接近与坐标系共动，且物质场的压强在相对论意义下很小，因此能动张量 $T_{\mu\nu}$ 只保留一个分量 $T_{00} = \rho$ ，其他分量均为零。我们首先看为零的这部分分量。我们希望无穷远处引力场为零，即 $h_{\mu\nu} = 0$ ，其中 μ, ν 不同时为 0。此时很容易发现也有 $\bar{h}_{\mu\nu} = 0$ ，而满足这个条件 $\square \bar{h}_{\mu\nu} = 0$ 的解只有 $\bar{h}_{\mu\nu} = 0$ 。因此我们发现 $\bar{h}_{\mu\nu}$ 也仅有 \bar{h}_{00} 这一个分量非零。

$\square x^\mu = 0$ 中的 \square 需要用 Christoffel 符号做修正，因为 $\partial^\nu \partial_\nu x^\mu$ 一定为零，所以需要保留小量。

注意我们的背景时空不能变，所以 $h_{\mu\nu}$ 作为张量没有变的实际上是 ∂_μ 。

我们现在可以只看非零的这个分量方程 $\square \bar{h}_{00} = -2\kappa\rho$ 。在低速近似下，度规近似不随时间变化，因此 d'Alembert 算符的第一项可以省去，退化为三维空间的 Laplace 算符 ∇^2 。所以我们有

$$\nabla^2 \bar{h}_{00} = -2\kappa\rho$$

而我们在 5.3.1. 中推导过， $g_{00} = -1 - 2\Phi$ ，由此可知 $h_{00} = -2\Phi$ 。而由于 $\bar{h}_{\mu\nu}$ 只有一个 \bar{h}_{00} 分量非零，因此可知它的缩并 $\bar{h} = \eta^{00}\bar{h}_{00} = -\bar{h}_{00}$ 。由 $\bar{h}_{\mu\nu}$ 的定义我们可以解出 $\bar{h}_{00} = 2h_{00} = -4\Phi$ 。于是上式意味着

$$-4\nabla^2 \Phi = -2\kappa\rho$$

于是我们得到了与牛顿引力的泊松方程 $\nabla^2 \Phi = 4\pi\rho$ 形式相同的方程。二者对比即可得到 $\kappa = 8\pi$ 。

6.2 引力波

本节来介绍广义相对论的一个重要预言——引力波。为了简化问题，我们将在弱场的线性近似下讨论引力波，而在强场下不使用线性近似计算引力波至今仍然是一件很难的事情。

6.2.1 线性场方程的波动解

下面我们用线性爱因斯坦场方程来计算弱场下的引力波。我们在此只讨论最简单的情况，即引力波在真空中的传播。注意到真空线性爱因斯坦场方程 $\square \bar{h}_{\mu\nu} = 0$ 如果写成三维形式就是我们熟悉的波动方程

$$\frac{\partial}{\partial t} \bar{h}_{\mu\nu} = \nabla^2 \bar{h}_{\mu\nu}$$

由这个波动方程的系数已经能看出波速是光速了，不过我们后面再更直接地推导一下。

因此这个方程有波动解。我们在此考虑最简单的平面单色波解，而复杂的波动可以由平面单色波叠加而来。平面单色波解可以写作

$$\bar{h}_{\mu\nu} = A_{\mu\nu} e^{ik_\sigma x^\sigma}$$

其中 $A_{\mu\nu}$ 为“振幅张量”， k^σ 为 4-波矢量。注意为了方便，这个解写成了复数形式，但度规实际上是实张量，因此我们可以只取它的实部。而波矢量也不是任意一个矢量。在单色波中，振幅是定值，因此有 $\partial_\sigma \bar{h}_{\mu\nu} = ik_\sigma \bar{h}_{\mu\nu}$ ，所以

$$0 = \partial^\sigma \partial_\sigma \bar{h}_{\mu\nu} = -k^\sigma k_\sigma \bar{h}_{\mu\nu}$$

因而有 $k^\sigma k_\sigma = 0$ ，即 k^σ 是一个类光矢量。这说明引力波是以光速传播的。而利用调和规范条件

$$0 = \partial^\mu \bar{h}_{\mu\nu} = ik^\mu A_{\mu\nu} e^{ik_\sigma x^\sigma}$$

所以有 $k^\mu A_{\mu\nu} = 0$ ，即 $A_{\mu\nu}$ 和波矢量 k^σ 在某种意义上是“垂直”的。因此我们说，引力波是横波。此处横波的具体意义我们会在后面进行简单的讨论。

注意我们这里是在弱场近似下讨论的引力波。对于真正的爱因斯坦场方程，其波动解就不是简单的正弦波了，也不满足叠加原理。但我们得到的引力波是横波、且以光速传播结论都是在一般情况下也成立的。

6.2.2 横向无迹规范

在上面的讨论中, 我们要求 $\bar{h}_{\mu\nu}$ 满足调和规范. 但调和规范在线性引力论中仍然无法唯一地确定度规. 很容易发现, 如果矢量场 ξ^μ 满足 $\square \xi^\mu = 0$, 则以它进行规范变换后的 $\bar{h}_{\mu\nu}$ 仍然满足调和规范. 因此, 我们可以选择再次选择额外的束缚, 来进一步简化问题. 可以证明, 在调和规范的前提下, 我们总可以进一步进行规范变换, 使得平面单色引力波的振幅 $A_{\mu\nu}$ 满足

$$A = 0 \quad \text{且} \quad A_{0\mu} = 0$$

这称为 **横向无迹规范** (TT 规范) 或 **辐射规范**. TT 规范可以简化 $A_{\mu\nu}$ 的分量形式, 同时还有一个重要的性质就是

$$h_{\mu\nu} = \bar{h}_{\mu\nu}$$

下面, 我们不妨设引力波的传播方向沿着 x 轴, 即 $k^\mu = (\omega, \omega, 0, 0)^\mu$. 由 $k^\mu A_{\mu\nu} = 0$ 可知 $A_{1\mu} = 0$. 因此 $A_{\mu\nu}$ 的分量可以被写作

$$A_{\mu\nu} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & A_{22} & A_{23} \\ 0 & 0 & A_{23} & -A_{22} \end{bmatrix}_{\mu\nu}$$

可以发现度规的波动分量的确是垂直于波的传播方向的, 而且只有这两个分量是有物理意义的自由度.

6.2.3 引力波的物理效应

我们上面讨论了引力波中时空度规的动力学, 但度规的波动会造成怎样的可观测的物理效应呢? 为了得到与坐标系无关的效应, 我们考虑粒子之间的位置关系的变化, 这被测地偏离方程描述. 对于一个粒子, 其潮汐加速度为 $U^\beta \nabla_\beta (U^\alpha \nabla_\alpha X^\mu) = R^\mu_{\sigma\rho\nu} U^\sigma U^\rho X^\nu$. 在低速近似下, U^μ 的时间分量占主导, 则有

$$U^\beta \nabla_\beta (U^\alpha \nabla_\alpha X^\mu) = R^\mu_{00\nu} U^0 U^0 X^\nu$$

而对于黎曼张量, 我们可以带入线性近似的黎曼张量的公式, 并且在 TT 规范下将 $h_{\mu 0}$ 的项消去, 得到

$$R^\mu_{00\nu} = \frac{1}{2} \partial_0 \partial_0 h^\mu_{\nu} = \frac{1}{2} \frac{\partial^2}{\partial t^2} h^\mu_{\nu}$$

同时, 低速近似下 $t \approx \tau$, 所以 X^μ 的协变导数也可以近似为对 t 的偏导数. 于是测地偏离方程变为

$$\frac{\partial^2}{\partial t^2} X^\mu = \frac{1}{2} X^\nu \frac{\partial^2}{\partial t^2} h^\mu_{\nu}$$

现在我们讨论一下这种情况下 X^μ 的各个分量的变化. 同样设引力波沿着 x 轴传播, 则由上一小节的讨论可知, 其振幅张量的分量可以写作

$$A_{\mu\nu} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & A_+ & A_\times \\ 0 & 0 & A_\times & -A_+ \end{bmatrix}_{\mu\nu}$$

让我们来分别讨论 A_+ 和 A_\times 的效应. 若只有 A_+ 项, 则测地偏离方程写成分量形式为

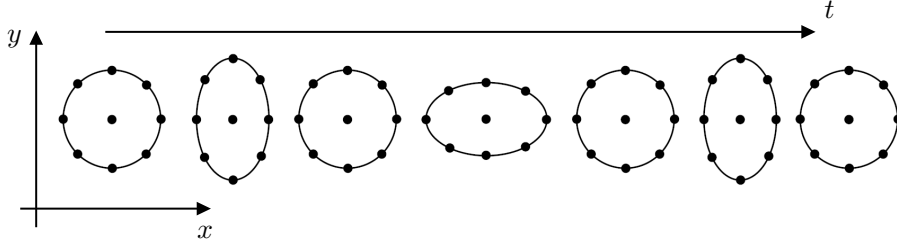
$$\begin{aligned}\frac{\partial^2}{\partial t^2} X^0 &= 0, & \frac{\partial^2}{\partial t^2} X^1 &= 0, \\ \frac{\partial^2}{\partial t^2} X^2 &= \frac{1}{2} X^2 \frac{\partial^2}{\partial t^2} (A_+ e^{ik_\sigma x^\sigma}), \\ \frac{\partial^2}{\partial t^2} X^3 &= -\frac{1}{2} X^3 \frac{\partial^2}{\partial t^2} (A_+ e^{ik_\sigma x^\sigma})\end{aligned}$$

现在为了只考虑引力波的效应, 我们设我们考虑的两个质点之间没有相对初速度, 也就是说偏移矢量的导数 $\partial X^\mu / \partial t$ 只能在 1 附近波动. 考虑到 $A_+ \ll 1$, 上面方程的解为

X^0 和 X^1 的解是精确的,
 X^2 和 X^3 是一阶近似.

$$\begin{aligned}X^0 &= X^0(0), & X^1 &= X^1(0), \\ X^2 &= \left(1 + \frac{1}{2} A_+ e^{ik_\sigma x^\sigma}\right) X^2(0), \\ X^3 &= \left(1 - \frac{1}{2} A_+ e^{ik_\sigma x^\sigma}\right) X^3(0)\end{aligned}$$

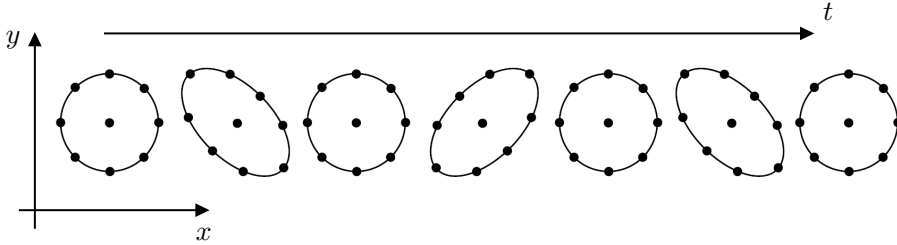
可以发现, 在 t 轴和平行于引力波传播方向的 x 轴上, 两个粒子之间的距离不会变化; 而在垂直于引力波传播方向的 y 轴和 z 轴上, 两个粒子之间的距离会有微小的振荡. 这就是引力波是横波的意义. 下图显示了这种情况下一个圆环的振荡方式.



而若只考虑 A_\times 的效应, 由同样的方法可以解得, X^0 和 X^1 分量仍然不变, 而 X^2 和 X^3 分量满足

$$\begin{aligned}X^2 &= X^2(0) + \frac{1}{2} A_\times e^{ik_\sigma x^\sigma} X^3(0), \\ X^3 &= X^3(0) + \frac{1}{2} A_\times e^{ik_\sigma x^\sigma} X^2(0)\end{aligned}$$

其振荡方式如下图所示



可以发现, 引力波的物理效应体现在潮汐力的波动上. 其振幅的分量 A_+ 和 A_\times 分别控制潮汐力沿 “+” 和 “ \times ” 两个方向的振荡模式, 我们称这是引力波的两种 **极化** 模式或 **偏振** 模式.

7 史瓦西时空

我们之前提到过，爱因斯坦场方程的精确求解是很难的。然而，若时空有一些比较好的对称性，我们还是可以对其精确求解的。本章我们来介绍爱因斯坦场方程的最重要的一个解——史瓦西外部解。

7.1 史瓦西度规

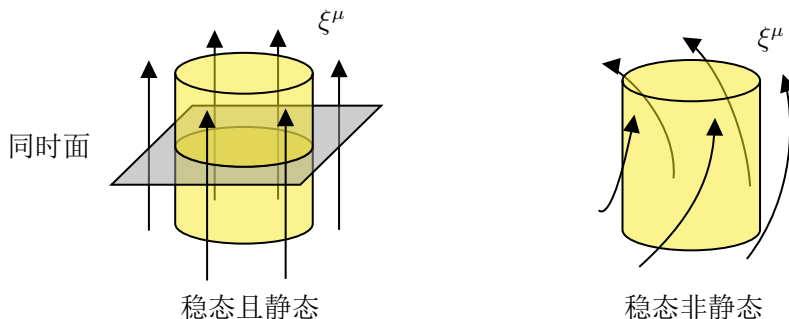
7.1.1 稳态度规与静态度规

在求解史瓦西度规前，我们有必要先来介绍两个重要的背景概念——度规的稳态性和静态性。

首先我们来看稳态性。一个 **稳态时空** 是指引力场不随时间变化的时空。不过这种定义有一个问题，那就是时空并不天生具有时间-空间分解。因此“不随时间变化”这个概念在没有定义出时间时是没有意义的，我们需要一个更严格的表述。首先，引力场不随时间变化即为度规具有时间平移对称性，而回顾一下我们在 2.3.3 小节中讲过，度规的对称性可以用 Killing 矢量场描述。另一方面，所谓的“时间对称性”也不是对任何参考系的时间——例如对于地球的引力场，在地球上“静止”的观测者觉得地球的引力场不变，但对于坐着火箭向上飞的观测者来说，地球的引力场是随着时间逐渐减小的。因此，只需要存在某种时间的定义，使得引力场具有时间平移对称性即可。那么我们就可以给出稳态时空的严格定义了：具有类时 Killing 矢量场的时空称为稳态时空。而这个类时 Killing 矢量场可以决定一族观测者，称为 **稳态观测者**。

稳态时空具有很好的时间对称性，但仍然不是时间对称性最好的时空。直观地说，其 Killing 矢量场不一定是“空间不动的”，而可能是“旋转的”。如果我们进一步要求该类时的 Killing 矢量场没有旋转成分，则该时空被称为 **静态时空**，对应的观测者称为 **静态观测者**。静态时空不仅具有时间平移对称性，还具有时间反演对称性。即若我们使用静态观察者的世界线作为时间轴，则时间反演变换 $t \mapsto -t$ 不会改变度规的形式。静态时空有一个重要的性质，那就是处处存在正交于整个 Killing 矢量场的超曲面，这可以视为在坐标时 t 相等下的全空间，称为这族静态观测者的 **同时面**。静态时空为我们提供了这样一个很好的时空分解。

为了更直观地理解稳态和静态时空的概念，我们用下面的图来表示这两种时空，其中黑色箭头表示 Killing 矢量场 ξ^μ 的积分曲线。



我们还可以直接从度规的数学形式对比稳态时空和静态时空。对于一个稳态时空，如果我们将稳态观测者的世界线作为时间轴，并任意地选取三个空间轴构成四维坐标系，则时空的线元 $ds^2 = g_{\mu\nu} dx^\mu dx^\nu$ 可以展开写成

$$ds^2 = g_{00} dt^2 + 2g_{0i} dt dx^i + g_{ij} dx^i dx^j$$

实际上就是选取 Killing 矢量场的适配坐标系。

而由于 $\partial/\partial t$ 是 Killing 矢量场，度规对其的李导数为零，因此上式中的所有度规分量都与 t 无关，而只能是空间分量的函数，这就是稳态时空的线元。

而对于静态时空，我们可以进一步在同时面内选择空间坐标，也就是说使得空间轴与时间轴处处垂直，则时空交叉分量全部为零，线元可以进一步简化为

$$ds^2 = g_{00}dt^2 + g_{ij}dx^i dx^j$$

其中的度规分量仍然都与 t 无关。

7.1.2 静态球对称度规

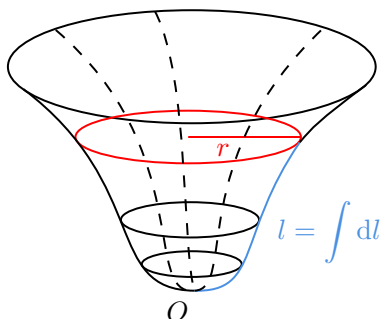
在牛顿引力论中，一个最简单的物理模型就是球对称、且物质分布为静态的天体外的引力场，它可以用来近似描述地球、太阳等各个恒星和行星所产生的引力场。这种模型中物质分布满足球对称性，进而时空度规也应满足球对称性。因此我们首先来推导一下球对称度规应该有什么样的形式。考虑在欧氏空间里的一个半径为 R 的球面，很容易推出欧式度规在其上的诱导度规的线元为

$$dl^2 = R^2(d\theta^2 + \sin^2\theta d\phi^2)$$

我们常常记 $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$ 为单位球面线元，则上式可以简化为 $dl^2 = R^2 d\Omega^2$ 。有了这个基础，接下来让我们进一步考虑三维球对称度规，我们使用球坐标系 $\{r, \theta, \phi\}$ 。注意到在任意一点， r 坐标的坐标基矢量 $\partial/\partial r$ 都是垂直于该点处的球面的，因此线元中 r 与另外两个坐标分量的交叉项应为 0。而在固定住 $r = R$ 时，度规应退化为上面的球面度规。因此三维球对称空间的线元一定形如

$$dl^2 = g_{11}(r)dr^2 + r^2 d\Omega^2$$

注意由于度规和坐标系同时具有的球对称性，我们的度规分量 g_{11} 只能是 r 的函数。当 $g_{11}(r) = 1$ 时，这就是我们熟悉的平直欧氏度规在球坐标系下线元。而当 $g_{11}(r)$ 不是 r 的常值函数时，这个空间就成为了一个“球对称弯曲”的空间。而在这种情况下，坐标 r 只能代表该处的球面在几何上“相当于多大的半径”，而不能再能代表该点与“球心”的距离。甚至所谓的“球心”可能在流形上可能有多个，或者根本不存在。我们用下图来直观地显示 r 的具体意义，并和与球心的距离 l 进行了对比。



这样，我们就推出了三维空间中球对称度规所应具有的形式。下面，我们就可以给这个空间添加上时间的维度，构造出四维球对称时空中的坐标系和度规。可以证明，如果静态球对称时空中仅存在一族静态观测者，则球对称的空间就是这族观测者的同时面。我们下面就假设只有一族静态观测者，那么我们可以将这族观测者的世界线作为时间轴，与同时面中的球坐标系共同构成四维的史瓦西坐标系 $\{t, r, \theta, \phi\}$ 。我们前面讲过，这种时空正

球坐标与笛卡尔坐标之间的关系为

$$x = r \sin \theta \cos \phi$$

$$y = r \sin \theta \sin \phi$$

$$z = r \cos \theta$$

再次强调，我们是用流形的内蕴几何描述弯曲。这里把流形画成高维空间中的曲面只是为了直观的显示流形的弯曲，而实际上这个外部的高维空间是不存在的，在生活在流形上的生物看来，底部的 O 点就是真正的球心。

注意这个正交的结论要求时空为静态，且静态观测者唯一。稳态但非静态的时空不具有这样的性质。

交的坐标系中线元一定可以写为 $ds^2 = g_{00}dt^2 + dl^2$ 的形式，也就是

$$ds^2 = g_{00}(r)dt^2 + g_{11}(r)dr^2 + r^2d\Omega^2$$

同样，由于球对称性， g_{00} 也一定只是 r 的函数。这就是四维球对称时空的线元所必须满足的形式。

7.1.3 史瓦西外部解

我们在上一节推出了静态球对称度规所必须满足的形式。然而，并不是所有满足这个形式的度规都是物理上所允许的。首先，我们要求度规具有洛伦兹符号差，且 t 为时间分量，那么必须有 $g_{00} < 0$ 且 $g_{11} > 0$ 。第二，这个度规必须满足爱因斯坦场方程。我们可以假设球对称天体的外部是真空，因此天体外部的时空应满足真空爱因斯坦场方程 $R_{\mu\nu} = 0$ 。带入静态球对称度规的形式，可以解得

$$-g_{00} = g_{11}^{-1} = 1 + \frac{C}{r}$$

其中 C 是一个常数。任意的 C 的取值都满足爱因斯坦场方程，并不是所有情况都有物理意义。为了使这个解代表物理上有意义的时空，我们还要求时空是 **渐进平直** 的，即当 $r \rightarrow +\infty$ 时度规应近似为闵氏度规。在此处做线性近似，则引力效应应退化为牛顿引力论。由此可以推出 $C = -2M$ ，其中 M 为天体的质量。于是我们就得到了最终的史瓦西外部解。

在标准单位制下

$$C = -\frac{2GM}{c^2}$$

定理 7.1: Schwarzschild 外部解

球对称静态时空的线元为

$$ds^2 = -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1}dr^2 + r^2d\Omega^2$$

对应的度规称为史瓦西度规。

在此有必要强调一下史瓦西外部解的适用范围。首先，史瓦西外部解描述的是天体外部的时空，也就是说 r 必须大于天体的半径。但不仅如此，我们还可以发现当 r 趋近于 $2M$ 时，史瓦西度规的 dt^2 的参数退化，而 dr^2 的参数发散。我们称 $r = 2M$ 这个特殊的半径为 **史瓦西半径**。因此我们同时还需要要求 $r > 2M$ 。不过，对于一般的天体来说，其半径常常是远大于其史瓦西半径的，因此我们史瓦西外部解在整个天体外部都适用。天体若半径小于 $2M$ 则会成为黑洞，我们后面会有专门的一章进行讨论。

标准单位制下史瓦西半径为

$$\frac{2GM}{c^2}$$

史瓦西度规仅有一个参数 M ，代表天体的质量。这意味着一个球对称天体的质量完全决定了其外部引力场，而其质量的分布并不会对此产生影响。例如对于我们在地球处感受到的来自太阳的引力场，无论太阳是当今的大小，还是变得像足球一样小，还是膨胀到了金星轨道处；无论太阳的质量是均匀分布的还是主要集中在球心或表面，只要总质量保持不变且质量是球对称分布的，那么我们在地球处感受到的引力场是不会变的。这一点与牛顿引力论的结论相同。

进一步讨论这一点，我们在求解史瓦西度规时，要求时空是静态球对称的。然而，人们后来发现对于静态的要求可以放开，这称为 **Birkhoff 定理**：真空爱因斯坦场方程的球对称解必为史瓦西解。这表明，哪怕物质分布不是静态的——例如一个天体可以坍缩、膨胀、甚至径向振荡——但只要保持球对称性，则其外部时空就一定保持静态，可以被史瓦西度规描述。

但这并不意味着球对称时空一定是静态的，我们后面会看到史瓦西外部解可以向 $r < 2M$ 进行延拓，而延拓后部分时空不再保持静态。只能说在 $r > 2M$ 处一定是静态的。

7.2 史瓦西时空的物理定律

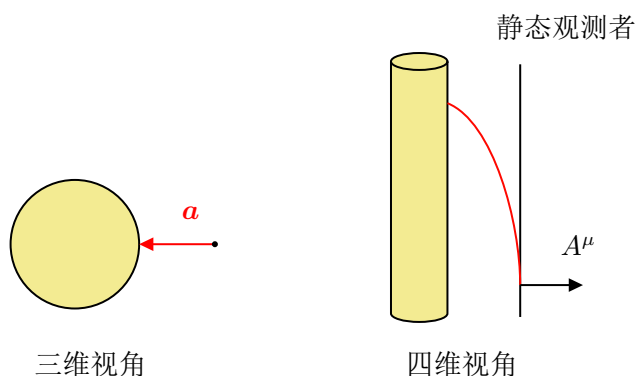
我们上一节介绍了球对称天体外部的时空结构。这一节，我们来看一下这种时空中的物理定律。

7.2.1 与牛顿引力论的对比

在牛顿引力理论中，引力场表现为引力加速度，即该点处自由落体的质点所具有的加速度

$$\mathbf{a} = \frac{M}{|r|^2} \hat{\mathbf{r}}$$

那么这个引力加速度在广义相对论中是怎样体现的呢？我们来对比一下牛顿引力论和广义相对论在这里的差别。在牛顿引力论中，引力加速度实际上是自由落体的质点相对于该处的静止观测者的加速度。然而广义相对论中对此的描述却几乎相反——从四维视角来看，所谓的“静止观测者”实际上是时空中的一个静态观察者，它具有非零的 4-加速度 A^μ 。而自由落体的质点在四维时空中划过一条测地线，因此它的 4-加速度反而为零。那么这两个加速度——自由落体质点的三维加速度和静态观测者的 4-加速度——有没有什么关系呢？



不过，三维和四维的矢量没法直接比较，我们需要把它们放在同一个线性空间里。既然自由落体质点的三维加速度 \mathbf{a} 是静态观测者所测得的加速度，那么我们自然可以把它嵌入到与静态观测者世界线正交的空间中，使它变成一个 4-矢量。由于静态观测者的 4-加速度 A^μ 也是垂直于其世界线的，且指向球心的反方向，因此两个加速度的方向相反。

下面我们来具体计算一下它们的大小关系。可以证明，对于一个稳态时空，其中的静态观测者的 4-加速度为

$$A^\mu = \frac{1}{2} \nabla^\mu \ln(-\xi^\nu \xi_\nu) = \frac{1}{2} \nabla^\mu \ln(-g_{00})$$

其中 ξ^ν 为类时 Killing 矢量场，不过为了计算方便我们可以取 $\xi^\nu = \partial x^\nu / \partial t$ 从而得到第二个等号。进而此处的 g_{00} 分量需要是稳态坐标系下的分量。对于史瓦西时空，有 $-g_{00} = 1 - 2M/r$ ，所以很容易得到

$$A^\mu = \frac{M}{r^2} \left(1 - \frac{2M}{r}\right)^{-1} \frac{\partial x^\mu}{\partial r}$$

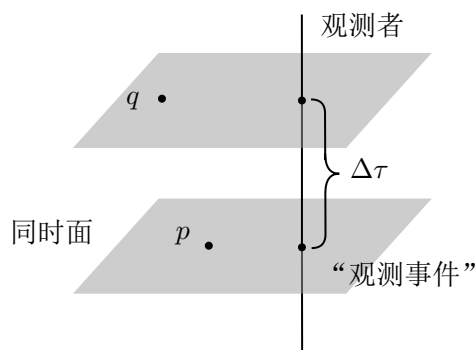
即 A^μ 垂直于时间轴且指向球心的反方向，其大小为 $\frac{M}{r^2} \left(1 - \frac{2M}{r}\right)^{-1}$ 。在牛顿近似下， r 远大于史瓦西半径 $2M$ ，所以括号里的部分近似为 1，进而 A^μ 的大小近似为 $\frac{M}{r^2}$ 。可以发现，这与牛顿引力论的引力加速度大小相同，方向相反。

我们以静态的史瓦西时空为例展示了三维的牛顿引力加速度和四维的静态观测者的 4-加速度的关系——在弱场近似下，二者大小相同，方向相反。实际上，对于任意一个稳态时空都有这样的关系。不过对于非稳态时空，引力场就无法简单地体现为一个引力加速度了，其引力效应就必须以四维视角来分析了。

7.2.2 引力钟慢与引力红移

在广义相对论中，有引力场时的时空是弯曲的，因此时间与观测者在引力场中的位置有关。在引力场中的观测者的时间走的会更慢一些，这称为 **引力钟慢效应**。我们下面来讨论一下这个现象。

在讨论钟慢效应之前，我们首先需要明确不同位置的时间是如何比较的。在狭义相对论的钟慢效应中，我们是让两个惯性观测者都建立了自己的惯性坐标系，从而比较了两个事件 p 和 q 在这两个坐标系下坐标时间间隔的差异。我们可以换一个角度去理解这个对时方式。对于一个观测者，我们可以假设它时时刻刻可以不受光速限制地观测它所在的同时面内的事件。也就是若事件 p 位于观测者的某个同时面内，则这个同时面与观测者世界线的交点就是该观测者“观测到事件 p ”的事件。而 p 和 q 在这个观测者看来的时间间隔就是从“观测到 p ”到“观测到 q ”之间的固有时。比较两个观测者看来的时间间隔差异，实际上就是比较他们的两次观测的固有时时间间隔的差异。



闵氏时空的这种对时方式可以很容易地推广到静态的弯曲时空中。我们考虑一个在三维意义下的“静止观测者”，也就是相对论语言中的静态观测者。任何静态观测者的同时面都是静态坐标系下的同时面。因此，我们只需讨论观测者穿过给定的两个同时面的固有时时间间隔即可。考虑坐标时 t 差异为 Δt 的两个同时面，静态观测者的固有时时间间隔显然为

$$\Delta\tau = \sqrt{-\xi^\mu \xi_\mu} \Delta t = \sqrt{-g_{00}} \Delta t$$

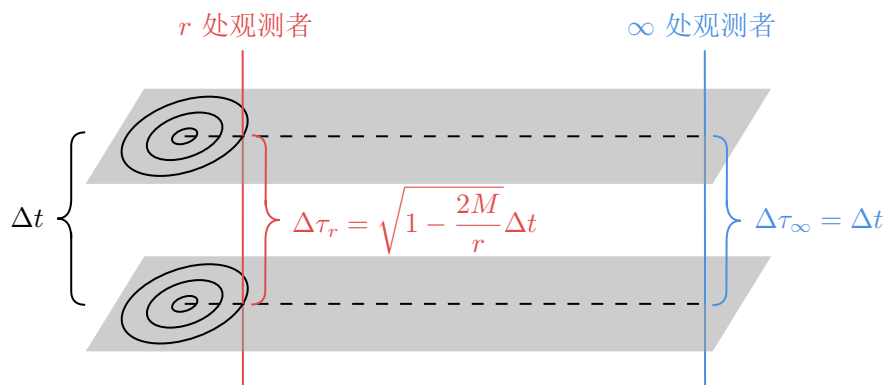
其中这里的 $\xi^\mu = \partial x^\mu / \partial t$ 为 t 坐标基矢量。在弯曲时空中不同处的度规不同，进而固有时也可以不同。在史瓦西时空下，对于径向坐标为 r 的静态观测者，他在两个同时面之间的固有时时间间隔为

$$\Delta\tau_r = \sqrt{1 - \frac{2M}{r}} \Delta t$$

可以发现这是 r 的函数，也就是说不同 r 坐标的观测者所经历的时间是不一样的。显然 r 越大， $\Delta\tau_r$ 就越大。即引力场越强，其时间流逝的就越慢，这就是引力钟慢效应。当 $r \rightarrow \infty$ 时，有 $\Delta\tau_r \rightarrow \Delta t$ 。因此我们可以说，史瓦西坐标系下的坐标时 t 实际上是无穷远处静态观测者的固有时。

我们可以用下图直观地表示史瓦西时空的引力钟慢效应。对于两张坐标时相差 Δt 的同时面，由于度规的差异，不同处的静态观测者经历的固有时不同。对于其他静态时空，

其原理也相同。不过对于非静态时空，我们不再有性质如此好的观测者和同时面，因此就无法这样简单地讨论对时问题了。



从引力钟慢我们还可以推出另外一个重要的现象，即 **引力红移**。考虑从引力场某处径向向外发射的一束光，其周期 T 在远处的观测者看来会由于引力钟慢效应而大于实际发射处的周期。而光的波长 $\lambda = cT$ 正比于周期，因此其波长的变化与时间的变化相同，有

$$\lambda \propto \sqrt{1 - \frac{2M}{r}}$$

注意此处的 r 是接收者的径向坐标。由于光向外传播时 r 增大，波长也相应增大，因此称为红移。当然，如果光是向引力场内传播的，实际上发生的是蓝移。引力红移可以从另一个视角解释为光逆着引力的方向传播，其能量降低。而光子的能量是正比于其频率的，因此频率降低，波长增加。另外还可以发现当发射源位置趋于史瓦西半径 $r = 2M$ 时红移趋向于无限大，因此此处 $r = 2M$ 处也被称为 **无限红移面**。

实际上无限红移面处发出的光线并不能向外传播，这种“无限红移”是一种极限的说法。

引力钟慢和引力红移效应都已经被实验所证实，它们是支持广义相对论的重要实验证据。

7.2.3 史瓦西时空的对称性与守恒量

守恒律在物理定律中占有非常重要的地位。在牛顿力学和狭义相对论中，我们有能量守恒、动量守恒、角动量守恒等许多个守恒律。这些守恒律的理论基础是著名的 **诺特定理**，它指出每一个对称性都对应着一个守恒量。时间平移对称性对应着能量守恒，空间平移对称性对应着动量守恒，而空间旋转对称性对应着角动量守恒。牛顿力学和狭义相对论的背景时空有这些良好的对称性，自然就有这些守恒律。然而在广义相对论中，背景时空可以是一个非常任意的流形，不一定具有这些对称性，因此这些守恒律也不再一定成立。

诺特定理要求连续参数的对称性，因此时间反演对称性不对应着守恒量。

然而，对于一些具有良好对称性的时空，我们还是可以找到这些对称性所对应的守恒律。这需要借助关于 Killing 矢量场的一个数学结论：Killing 矢量场与测地线的切矢量的内积沿测地线为常数。我们可以通过这种方式找到对应的守恒量。

首先考虑稳态时空的时间平移对称性，它由类时的 Killing 矢量场 $\xi^\mu = \partial x^\mu / \partial t$ 表示，由此可以写出它与测地线的切矢量 U^ν 的内积作为守恒量

$$E = -\xi_\mu U^\mu$$

这种时间平移对称性对应的守恒量我们将其视作能量。注意这里的能量是完全用质点世界线定义的，并不涉及质点的质量，因此它代表的是实际上是单位质量的能量。另外，由于这是两个类时矢量的内积，因此我们添加了个符号使之成为正值。

无质量粒子走类光测地线，不具有固有时参数，因此我们只能选择任意一个仿射参数，进而得到的结果也有任意性。但选定了一个仿射参数后，这个值在测地线上就一定是常数了。

下面我们来看一下它为什么代表这个量。在狭义相对论中，我们讨论了质能关系 $E = \gamma m$ ，其中能量是和观测者有关的。那么单位质量的能量即为 γ ，我们推导过瞬时观测者 Z^μ 会测得 $\gamma = -Z_\mu U^\mu$ 。我们这里的定义也类似， ξ^μ 是无穷远处的静态观测者，因此得到的内积可以看作在无穷远处的静态观测者测得的总能量。在有限远处，我们可以把这个概念定义为质点的总能量。

与 $\nabla_\mu T^{\mu\nu}$ 相比，这个能量守恒是全局的，而 $\nabla_\mu T^{\mu\nu}$ 代表的能量守恒是局部的。

下面我们来看一下史瓦西时空中的能量守恒。在史瓦西坐标系下我们很容易将定义式展开为

$$E = -g_{\mu\nu} \frac{\partial x^\mu}{\partial t} \frac{dx^\nu}{d\tau} = -g_{\mu\nu} \frac{\partial x^\mu}{\partial t} \frac{\partial x^\nu}{\partial t} \frac{dt}{d\tau} = \left(1 - \frac{2M}{r}\right) \frac{dt}{d\tau}$$

从这里也能看出 E 在无穷远处代表单位质量的能量。在无穷远处，上式趋于 $dt/d\tau$ ，其中 t 相当于无穷远处静态观测者的固有时，因此这个值就是静态观测者测得的 γ 。这个值在有限远处和当地的静态观测者测得的能量是不同的，对于有限远处的静态观测者 Z^μ ，由于 $Z^\mu = (1 - 2M/r)^{-1/2} \xi^\mu$ ，所以该静态观测者测得的质点能量 $\gamma = -Z_\mu U^\mu$ 是大于 $-\xi_\mu U^\mu$ 的。二者相差的部分可以视作质点在引力场中的引力势能。

史瓦西时空除了时间平移对称性外还有空间的球对称性。观察到史瓦西坐标系下的线元与 ϕ 坐标无关，因此 $\kappa^\mu = \partial x^\mu / \partial \phi$ 也是一个 Killing 矢量场，它代表了时空沿着 ϕ 坐标旋转的对称性，即绕 z 轴的旋转对称性。由这个 Killing 矢量场，我们可以写出它与测地线切矢量的内积作为第二个守恒量

三维球对称时空还有另外两个球对称 Killing 矢量场，不过我们这里只以这一个对称性为例讨论角动量守恒。

$$L = \kappa_\mu U^\mu$$

这个球对称性对应的守恒量自然就是质点的公转角动量。这点很容易理解，可以发现对于史瓦西时空，有

$$\begin{aligned} L &= g_{\mu\nu} \frac{\partial x^\mu}{\partial \phi} \frac{dx^\nu}{d\tau} = g_{\mu\nu} \frac{\partial x^\mu}{\partial \phi} \frac{\partial x^\nu}{\partial \phi} \frac{d\phi}{d\hat{t}} \frac{d\hat{t}}{d\tau} \\ &= g_{33} \omega \gamma = \gamma \omega r^2 \end{aligned}$$

我们默认三个空间坐标的顺序为 r, θ, ϕ ，所以 g_{33} 就是在 ϕ 方向上的分量。

其中 \hat{t} 是当地的静态观测者的时间， γ 是当地的静态观测者测得的 γ 。很容易发现这就是单位质量的质点由当地的静态观测者测得的质点公转角动量。因此，史瓦西时空中满足角动量守恒。

7.2.4 史瓦西时空的测地线

这一小节我们来求解史瓦西时空中的运动方程。我们知道，在没有外力作用时，粒子的世界线应画出一条测地线。我们早就在 2.3.3 小节中导出了测地线方程的分量形式，我们只需从度规求出 Christoffel 符号，再带入就可以得到测地线需要满足的方程了。然而，测地线方程是一个二阶微分方程组，同时它实际上由四个分量方程构成，而 Christoffel 符号使得这四个分量之间是耦合在一起的，难以直接求解。因此，我们需要一些数学技巧来简化测地线方程。

我们首先可以利用时空的球对称性，选取合适的坐标系，使得粒子在天体的赤道面 ($\theta = \pi/2$) 上运动，这样我们有 $d\theta = 0$ ，且 $\sin \theta = 1$ ，可以用于简化方程。接下来我们考虑粒子的 4-速度的大小，即测地线切矢量的长度 $\epsilon = -g_{\mu\nu} U^\mu U^\nu$ ，则对类时曲线有 $\epsilon = 1$ ，对类光曲线有 $\epsilon = 0$ 。我们将 $g_{\mu\nu} U^\mu U^\nu$ 按分量展开，考虑到赤道面上的简化，则有

$$-\epsilon = -\left(1 - \frac{2M}{r}\right) \left(\frac{dt}{d\tau}\right)^2 + \left(1 - \frac{2M}{r}\right)^{-1} \left(\frac{dr}{d\tau}\right)^2 + r^2 \left(\frac{d\phi}{d\tau}\right)^2$$

上式并没有要求曲线是测地线. 而对于测地线, 我们有两个守恒量守恒量 E 和 L , 将它们带入得到

$$-\epsilon = -\left(1 - \frac{2M}{r}\right)^{-1} E^2 + \left(1 - \frac{2M}{r}\right)^{-1} \left(\frac{dr}{d\tau}\right)^2 + \frac{L^2}{r^2}$$

上式只含有一个一元未知函数 $r(\tau)$, 且是它的一阶常微分方程, 理论上可以求解. 求解后带回 E 和 L 的表达式, 即可求出 $t(\tau)$ 和 $\phi(\tau)$, 即求出了整个测地线的方程.

现在我们去直接求解上式, 而是来定性地讨论一下上式会产生什么样的运动. 我们可以把上式重新整理成

$$\frac{1}{2} \left(\frac{dr}{d\tau}\right)^2 + V_{\text{eff}}(r) = \frac{1}{2} E^2$$

其中 $V_{\text{eff}}(r)$ 称为史瓦西时空中引力场的 **有效势**

$$V_{\text{eff}}(r) = \frac{1}{2} \epsilon - \epsilon \frac{M}{r} + \frac{L^2}{2r^2} - \frac{ML^2}{r^3}$$

这样, 粒子的径向运动可以看作一个总能量为 $E^2/2$ 的粒子在一维势场 $V_{\text{eff}}(r)$ 中的运动. 可以发现, 爱因斯坦有效势的第二项和第三项就是牛顿引力的有效势, 第一项是常数而可以忽略, 而最后一项则是广义相对论对于牛顿引力论的修正.

我们下面在例 7.1. 中给出一点牛顿引力的有效势的简单介绍.

例 7.1. 牛顿引力论的有效势能 牛顿引力论中粒子在球对称引力场中运动的能量守恒方程为

$$E = \frac{1}{2} m \dot{\mathbf{x}}^2 + V(r)$$

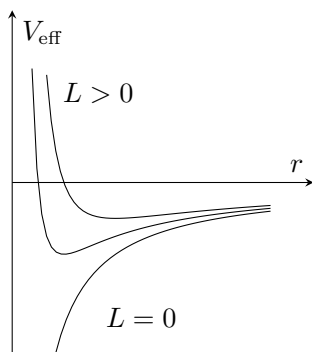
其中 $V(r)$ 为引力势能. 由于 $\dot{\mathbf{x}}^2 = \dot{r}^2 + r^2 \dot{\phi}^2$, 而角动量 $L = mr^2 \dot{\phi}$ 为守恒量, 所以上式可以写成

$$E = \frac{1}{2} m \dot{r}^2 + V(r) + \frac{L^2}{2mr^2}$$

那么我们可以定义 **有效势能**

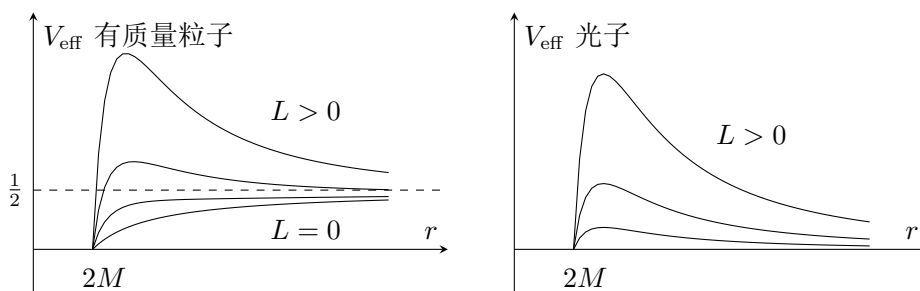
$$V_{\text{eff}}(r) = V(r) + \frac{L^2}{2mr^2}$$

那么粒子的径向运动可以看作在一个势能为 $V_{\text{eff}}(r)$ 的一维势场中的运动. 我们称多出来的一项 $L^2/2mr^2$ 为离心势能.



上图显示了在不同 L 情况下的牛顿有效势能. 可以发现当 $L = 0$ 时, 即质点瞄准引力场中心入射时, 会坠入引力场中心. 只要 $L > 0$, 有效势在 $r \rightarrow 0$ 时趋于无穷, 因此质点不会坠入引力场中心. 当总能量 $E < 0$ 时, 质点的径向位置 r 会被束缚在一个区间里, 这称为束缚运动; 而当 $E \geq 0$ 时, 质点会最终离开引力场到无穷远处, 这称为散射运动.

广义相对论的有效势相对于牛顿有效势的修正项为 r^{-3} 阶，所以在 r 较大时可以忽略，此时广义相对论的有效势退化为牛顿有效势。然而在 r 较小时，修正项占主导，此时广义相对论的有效势和牛顿有效势会有本质的区别。当然，我们的 r 不能任意小，它需要大于 $2M$ 。下图显示了有质量 ($\epsilon = 1$) 和无质量 ($\epsilon = 0$) 的粒子的广义相对论有效势。



可以发现，对于有质量的粒子，当 L 较大时，有效势在 $2M$ 右侧随着 r 的增大先达到极大值，再达到极小值，最终上升趋于 $1/2$ 。这两个极值点处可以有 r 不变的轨道，即圆周轨道。不过极大值处是一个不稳定平衡，而极小值处是一个稳定平衡。很容易解得两个极值点为

$$R_{\pm} = \frac{L^2 \pm L\sqrt{L^2 - 12M^2}}{2M}$$

可以发现这要求 $L \geq \sqrt{12}M$ ，这样进而可以得到 $R_+ \geq 6M$ 而 $R_- \leq 6M$ 。因此，在广义相对论中有质量的粒子只可能在 $r > 6M$ 处有稳定的圆周轨道，在其附近可以有非圆周的束缚轨道；而在 $6M$ 以内的轨道一定是不稳定的。而当 $L < \sqrt{12}M$ 时，系统没有极值，并且有效势随着 r 减小而单调减小，因此粒子最终一定会坠入天体内。

而对于没有质量的光子，广义相对论有效势只有一个极大值点而没有极小值点。也就是说光没有稳定的束缚轨道。因此光要么坠入天体，要么会重新远离天体，到达无穷远处。

稳定平衡是指在扰动下会自发复原的状态，其势能为极小值；而不稳定平衡的势能为极大值，在扰动下会愈发偏离原状态。

$r = 6M$ 处称为最内层稳定圆周轨道 (ISCO)，此处实际上是一个临界稳定的轨道。

我们上面只是定性分析了广义相对论中粒子的运动轨道的可能类型，但并没有真正解出粒子的轨道方程。实际上这个求解往往是很困难的。下面我们就来看两个非常重要的例子，求解其中的近似轨道方程。

例 7.2. 水星近日点的进动 牛顿引力论指出，行星绕太阳公转的束缚轨道一定为闭合椭圆（或正圆）。我们先来推导一下这个结论。牛顿引力论下质点的极坐标运动方程为

$$\begin{aligned} r^2 \dot{\phi} &= L \\ \ddot{r} - r \dot{\phi}^2 &= -\frac{M}{r^2} \end{aligned}$$

两式中 r 和 ϕ 都是 t 的函数。而我们希望得到质点的轨道，即 $r(\phi)$ 。若我们引入辅助变量 $u = r^{-1}$ ，则由上面两式可以得到

此式称为 Binet 方程。

$$\frac{d^2 u}{d\phi^2} + u = \frac{M}{L^2}$$

若我们选取初值 $\phi_0 = 0$ ，即初始位置的幅角为 0，则可以解得

$$u = \frac{1}{\ell}(1 + e \cos \phi)$$

其中 $\ell = L^2/M$ 确定了轨道的大小, $e = 1 + \frac{2EL^2}{mM^2}$ 为轨道的离心率 (E 为总机械能). 当 $e < 1$ 时轨道即为椭圆 ($e = 0$ 时为正圆).

然而广义相对论的预言略有不同. 下面我们只考虑类时测地线. 我们仍然使用变量 $u = r^{-1}$, 并把方程的 τ 消去, 可以得到

$$\left(\frac{du}{d\phi}\right)^2 + u^2 = \frac{E^2 - 1}{L^2} + \frac{2M}{L^2}u + 2Mu^3$$

将上式对 ϕ 求导可以得到

$$\frac{d^2u}{d\phi^2} + u = \frac{M}{L^2} + 3Mu^2$$

可以发现这个方程的右侧与牛顿引力论的方程相比加入了修正项 $3Mu^2$, 在 r 远大于史瓦西半径时这个修正项的影响较小, 因此可以近似回归牛顿定律. 带有修正项的轨道方程是没有解析解的, 但若我们真的要考虑修正项的影响, 可以在牛顿引力论的解的基础上进行近似. 若我们把上面牛顿引力论的解记作 $u_0(\phi)$, 则对上式我们可以考虑在牛顿解附近的微扰, 即考虑一阶近似 $u_1(\phi)$ 的方程

$$\frac{d^2u_1}{d\phi^2} + u_1 = \frac{M}{L^2} + 3Mu_0^2$$

此方程可以求得精确解

$$u_1 = u_0 + \alpha \left(1 + \frac{e^2}{2} - \frac{e^2 \cos 2\phi}{6} + e\phi \sin \phi \right)$$

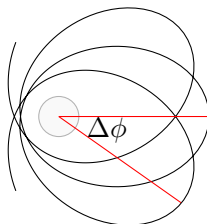
其中 $\alpha = 3M^2/L^2$. 可以发现这与牛顿解相比多出了一些修正项, 其中前两项为常数, 第三项以 π 为周期, 二者都不会影响轨道的周期性. 然而第三项不是周期函数, 会破坏轨道的周期性, 使轨道偏离闭合轨道. 我们可以忽略前两项, 而只保留第三项非周期项. 把 u_0 带入, 可得

$$\begin{aligned} u_1 &= \frac{1}{\ell} (1 + e(\cos \phi + \alpha \phi \sin \phi)) \\ &\approx \frac{1}{\ell} (1 + e \cos(\phi - \alpha \phi)) \end{aligned}$$

其中第二个等号是因为 α 是小量. 若我们考虑质点的近日点, 即 u_1 取最大值的点. 当 $\phi = 0$ 时显然是一个最大值, 而下一个最大值不难验证在一阶近似下为 $2\pi + 2\alpha\pi$. 可见质点每次的近日点都会偏离一个小角度, 其大小为

$$\Delta\phi \approx 2\alpha\pi = \frac{6M^2}{L^2}\pi$$

这称为近日点的进动. 下图画出了这种广义相对论语言的轨道并标出了进动角. 为了方便画图, 我们标出的是远日点的进动角, 但显然二者是等价的.



由于引力场越强，广义相对论与牛顿理论的偏离越大，因此在太阳系中距离太阳最近的水星的轨道进动应该是最显著的。带入水星的具体数据可得睡醒近日点的进动角为每世纪 43 秒。这已经被天文观测所证实。

例 7.3. 星光偏折 广义相对论的另一个重要预测是光线会受引力场影响而发生偏转。光子走类光测地线，其 $\epsilon = 0$ ，因此运动方程满足

实际上牛顿引力论也能得到星光偏折的结果，但其大小约为广义相对论预言的一半。

$$-\left(1 - \frac{2M}{r}\right)^{-1} E^2 + \left(1 - \frac{2M}{r}\right)^{-1} \left(\frac{dr}{d\tau}\right)^2 + \frac{L^2}{r^2} = 0$$

我们仍然引入变量 $u = r^{-1}$ ，则通过与上面类似的方式，上式可以化为

$$\frac{d^2 u}{d\phi^2} + u = 3Mu^2$$

若引力场不是特别强，即 $M \ll r$ ，那么我们可以用同样的近似方法求解。我们首先令 $M = 0$ ，那么同样若选取初值 $\phi_0 = 0$ ，即光从右向左入射，则方程的解为

$$u = \frac{1}{\ell} \sin \phi$$

很容易发现这其实就是极坐标下的直线方程，这与 $M = 0$ 情况下时空退化为平直时空相符，此时 ℓ 代表直线与原点的距离。而 ϕ 代表点的幅角。同样把这个解作为零级近似 u_0 ，带入得到一阶近似方程

$$\frac{d^2 u_1}{d\phi^2} + u_1 = \frac{3M}{\ell^2} \sin^2 \phi$$

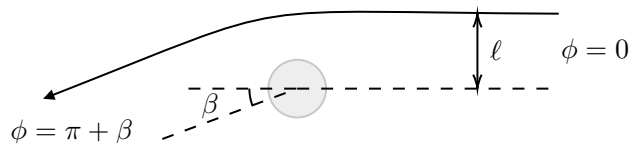
此方程有精确解

$$u_1 = u_0 + \frac{M}{\ell^2} (1 - \cos \phi)^2$$

当光距离原点无限远时，有 $u_1 = 0$ 。显然 $\phi = 0$ 是一个解，而另一个解应该在 π 附近。设它为 $\pi + \beta$ ，则可以近似解得

$$\beta \approx \frac{4M}{\ell}$$

而这个 β 就是光线经过引力场的偏折角。下图显示了这个几何关系。



水星近日点的进动和太阳对于星光的偏折都被天文观测证实，它们是支持广义相对论最早的实验证明。

8 黑洞

这一章，我们来学习广义相对论预言的最奇特的天体之一——黑洞。不过由于本章涉及的数学知识较深，很多地方我们只能浅尝辄止。

8.1 静态球对称黑洞

我们在上一章介绍了史瓦西度规，并讨论了史瓦西半径 $2M$ 外的时空结构和物理定律。我们提到，对于一般的天体来说，其半径远大于史瓦西半径，因此天体外的空间可以用史瓦西外部解描述。然而在一些极端情况下，天体的半径是可以小于史瓦西半径的——或者说其半径为零。这种天体称为 **黑洞**。对于黑洞，我们需要开始研究史瓦西半径以内的时空。

8.1.1 史瓦西时空的奇性

我们在上一章给出了史瓦西时空的线元在史瓦西坐标系下的表达式

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 d\Omega^2$$

这个线元表达式显然在 $r = 0$ 处和 $r = 2M$ 处是退化的，我们称这种位置为 **奇点**。这意味着我们线元表达式无法描述奇点处的时空。这使得史瓦西半径 $r = 2M$ 处的奇性给我们带来了困扰。这里的奇性导致这个线元只能用来分析完全处于史瓦西半径以外的物理现象，或者分析完全属于史瓦西半径以内的物理现象。但它不能用于分析涉及两者的现象，例如质点坠落跨过史瓦西半径的过程，因为这会涉及到史瓦西半径处的时空。

然而，线元表达式毕竟不是时空度规本身，它还与坐标系的选取有关。若坐标系选取得不好，只描述了时空的一部分，则“擦边在外”的时空点可能就成为奇点。这种奇点的来源并不是时空的性质差，而是坐标系的性质差，因此是“假奇点”，或称为 **坐标奇点**。如果我们换一个坐标系，那么坐标奇点的奇性可能会消失。而若这些点的曲率本身就发散，那么这种奇点才是“真奇点”，或者称为 **时空奇点**。时空奇点的奇性不能通过坐标变换来消除。

时空奇点实际上并不是时空流形上的点，因为度规需要在整个流形上良定义。时空奇点更接近于流形上的“洞”。

为了更直观地理解坐标奇点，我们来看一个例子。这个例子对后面对史瓦西半径处奇性的分析也很有帮助。

例 8.1. Rindler 坐标 考虑一个二维时空，并给定坐标系 $\{T, X\}$ 。度规在该坐标系下的线元表达式为

$$ds^2 = -X^2 dT^2 + dX^2$$

显然这个线元表达式在 $X = 0$ 处退化，即具有奇性。因此我们只能取 $X > 0$ 或 $X < 0$ 的区域单独讨论其时空结构，而不能同时讨论二者。不失一般性地，我们在这里讨论 $X > 0$ 的部分。但若我们做坐标变换

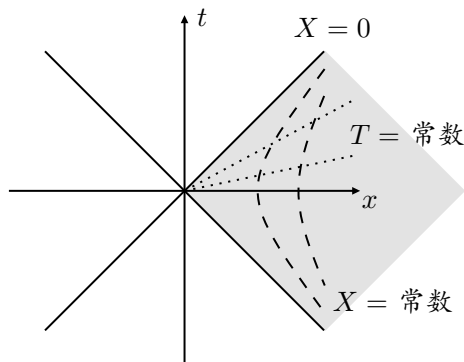
$$\begin{cases} t = X \sinh T \\ x = X \cosh T \end{cases}$$

则可以发现线元表达式变为 $ds^2 = -dt^2 + dx^2$ ，因此这个度规就是平直的闵氏度规。那么我们就可以很自然地将原来的时空 **延拓** 为整个闵氏时空。为了更直观地看到两个坐

标系之间的关系，我们找到逆变换

$$\begin{cases} T = \operatorname{arctanh}(t/x) \\ X = \sqrt{x^2 - t^2} \end{cases}$$

则可以发现每个 X 坐标线就是闵氏时空的一条双曲线，而每个 T 坐标线则是过原点的一条直线。注意到 $X = \sqrt{x^2 - t^2} > 0$ 且 $x = X \cosh T > 0$ ，因此原来的时空实际上只是闵氏时空里右 1/4 的子时空，其边界 $X = 0$ 是两条过原点倾斜 45° 的直线，如下图所示。



这样我们可以发现，最开始我们给出的线元的奇性实际上并不是时空本身的奇性，而是坐标系选取得不好造成的。我们称世界线为 X 坐标线的观测者为闵氏时空中的 **匀加速观测者**，因为他们具有大小不变的 4-加速度。Rindler 坐标系就是以这些匀加速观测者建立的坐标系。然而可以发现与这些观测者世界线垂直的“全空间”不经过 $X \leq 0$ 的部分，因此这些观测者“看不到” $X \leq 0$ 的时空。然而这是这些观测者的问题，而非时空的问题。

$X = 0$ 的两条曲线称为 Rindler 视界。

这个坐标变换和延拓的例子还告诉我们，延拓的区域可能会超过我们的预期。在这个例子中，我们本只想把 X 延拓到 $X \leq 0$ 的范围。在上图中，我们可以选择与原区域相隔一条 $X = 0$ 的区域作为延拓后 $X < 0$ 的区域，例如图中的上 1/4 或下 1/4。然而，完整的闵氏时空却包括对称的两个 $X > 0$ 和 $X < 0$ 的区域。这四个区域我们都可以给它们赋予 X 和 T 坐标，但它们的坐标变换是略有不同的，而且这样的结果是同一个 $\{T, X\}$ 坐标可以对应整个闵氏时空中的多个点。例如我们可以这样给它们赋予 T 和 X 坐标

$$\begin{aligned} \text{左和右: } T &= \operatorname{arctanh}(t/x) & X &= \sqrt{x^2 - t^2} \\ \text{上和下: } T &= \operatorname{arctanh}(x/t) & X &= -\sqrt{t^2 - x^2} \end{aligned}$$

8.1.2 事件视界及其内部的时空结构

既然史瓦西半径 $r = 2M$ 处对于史瓦西坐标系来说是一个坐标奇点，那么我们就应该用它描述涉及史瓦西半径处的物理现象。例如，我们考虑史瓦西半径外的径向光锥。我们令史瓦西线元 $ds^2 = 0$ ，同时由于我们考虑的是径向光锥，因此 $d\theta = d\phi = 0$ 。因此有

$$0 = -\left(1 - \frac{2M}{r}\right) dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2$$

因此径向光锥方程为

$$\frac{dr}{dt} = \pm \left(1 - \frac{2M}{r}\right)$$

可以发现 dr/dt 在 $r \rightarrow 2M$ 是趋于无穷的，即光锥只能无限接近史瓦西半径，但无法进入其内部（见后面的光锥图）。这说明任何物质都无法穿入史瓦西半径内。然而，我们使用的史瓦西坐标系只是无穷远处静态观测者的坐标系，因此这只能说明在无穷远处静态观测者看来，自由下落的物体无法在有限时间内穿入史瓦西半径内，只能无限接近。而对于下落的质点自己来说呢？让我们来计算一下。对于一个从 $r_1 > 2M$ 处自由径向下落的质点，它显然有角动量 $L = 0$ 。因此其径向运动方程满足

$$-1 = -\left(1 - \frac{2M}{r}\right)^{-1} E^2 + \left(1 - \frac{2M}{r}\right)^{-1} \left(\frac{dr}{d\tau}\right)^2$$

则可以解得

$$d\tau = \frac{dr}{\sqrt{E^2 - 1 + 2M/r}}$$

那么我们将 $d\tau$ 沿世界线积分即可得到质点的固有时。注意到 E 这个常数是不影响积分的收敛性的，因此我们可以令 $E = 1$ ，即质点在无穷远处为静止，这样 r 就可以取任意正实数。积分可以得到

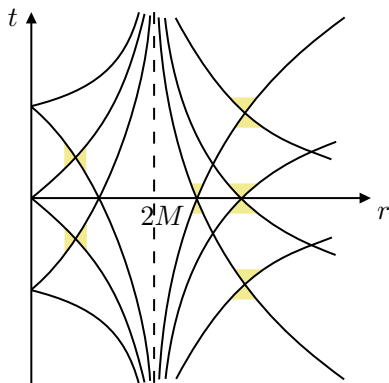
$$\Delta\tau = \int_{r_1}^{r_2} d\tau = \frac{4M}{3} \left(\frac{r}{2M}\right)^{3/2} \Big|_{r_1}^{r_2}$$

可以发现这个积分在 $r_1, r_2 > 0$ 时是收敛的，哪怕 $r_2 < 2M$ 。也就是说粒子经过有限时间后就会落入史瓦西半径内。这说明世界内外并不是分割开的两块时空，而是连续的。

我们暂且不考虑如何描述粒子穿过史瓦西半径时的物理过程，我们先来看一下史瓦西半径内的时空结构。由于史瓦西半径内仍然是真空的球对称时空，因此它也可以被史瓦西外部解描述

$$ds^2 = -\left(\frac{2M}{r} - 1\right)^{-1} dr^2 + \left(\frac{2M}{r} - 1\right) dt^2 + r^2 d\Omega^2$$

注意我们在这里稍改变了线元的写法，这是因为在 $r < 2M$ 处 r 变为了类时坐标， t 变为了类空坐标。所以这时 t 不再能视作时间坐标了，可以被视作时间坐标的变成了 r 。因此 Killing 矢量场 $\partial/\partial t$ 不再类时，时空失去了稳态性。如果要用几何方式体现这种时空结构，我们可以画出 $r < 2M$ 处的光锥，如下图所示。



可以发现在史瓦西半径内的光锥的朝向变为了径向，这体现了 r 坐标的类时性质。有物理意义的光锥还应区分过去和未来的方向。由于我们默认 t 坐标是指向未来的，因此在史瓦西半径外部的光锥的取向是确定的。然而在史瓦西半径内部，光锥不再沿着 t 的方向，而是沿着 r 的方向，那么它的因果结构就有两种情况。如果向内的方向是未来，向外的方向是过去，这意味着粒子在史瓦西半径内只能不断向内下落，这样的时空称为 **黑洞**。而若向内的方向是过去，向外的方向是未来，这意味着粒子只能向外运动，这种时空称为 **白洞**。

史瓦西半径内部的光锥曲线还说明了史瓦西半径处是一个“单向膜”，我们称之为 **事件视界**，简称视界。视界对于黑洞来说，物质和信息有进无出，或者说其内部的事件对外部无法产生任何影响。而对于白洞来说则相反。黑洞和白洞就是指视界包围的时空区域。注意视界内部的光锥决定了这个单向性是无法抵抗的，无论给质点提供多大的外力，质点的世界线都无法逃出光锥之外。这甚至决定了事件视界内部的质点无法保持自己的径向位置，它只能随着光锥向外或向内不可逆地运动。

注意事件视界和无限红移面不是同义词。对于史瓦西时空来说，事件视界和无限红移面的位置相同，但对于其他时空来说这并不一定。

8.1.3 Kruskal 最大延拓

我们前面在史瓦西坐标系下分析了视界内外的时空结构，然而 $r = 2M$ 坐标奇点的存在使得我们无法分析质点穿过视界的过程。下面我们来寻找一个新的坐标系，以消除 $r = 2M$ 处的坐标奇性。前面 Rindler 坐标系和匀加速观测者的例子可以给我们一点启示。在闵氏时空中，在匀加速观测者看来时空有一个视界。在史瓦西时空中，视界内部是无穷远处静态观测者无法观测到的区域。而静态观测者也有不变的 4-加速度，那么如果我们能找到一个坐标系，使得其中的光锥取向与闵氏时空类时，而静态观测者走双曲线，那么或许我们就能消除这个坐标奇点。这样的坐标系是存在的。如果我们把 $r > 2M$ 处的 t 和 r 坐标进行如下变换，而保持 θ 和 ϕ 坐标不变

$$\begin{cases} u = \left(1 - \frac{r}{2M}\right)^{1/2} e^{r/4M} \cosh\left(\frac{t}{4M}\right) \\ v = \left(1 - \frac{r}{2M}\right)^{1/2} e^{r/4M} \sinh\left(\frac{t}{4M}\right) \end{cases}$$

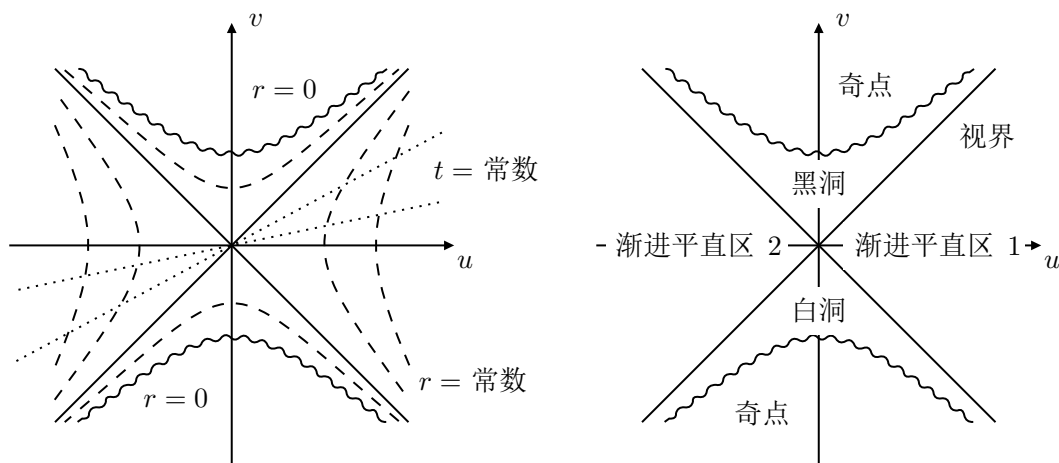
那么可以发现线元的表达式在新的坐标系变为

$$ds^2 = \frac{32M^3}{r} e^{-r/2M} (-dv^2 + du^2) + r^2 d\Omega^2$$

其中 r 可以表示为 u 和 v 的函数。这个新的坐标系 $\{u, v, \theta, \phi\}$ 称为 **Kruskal 坐标系**。可以发现线元表达式在 $r = 2M$ 处的奇性在 Kruskal 坐标系下消失了。因此我们可以放心地将时空延拓至 $r = 2M$ 的另一侧。不过 $r = 0$ 处的奇性仍然存在。实际上可以证明，这里的奇性是时空奇性，无法被坐标变换消除。因此不可能找到一个坐标系使得我们可以延拓至 $r = 0$ 的另一侧。这说明，在 Kruskal 坐标系下的延拓是史瓦西时空的最大延拓。

又称为 Kruskal-Szekeres 坐标系。

为了更直观地分析 Kruskal 坐标系下的史瓦西时空及其延拓，我们忽略掉球对称的 θ 和 ϕ 坐标，以 u 和 v 为坐标轴画出一张二维的图，如下图左图所示。此时图中的每个点代表一对确定的 t 和 r ，也就是一个时刻的一个球面。



很容易发现，我们本来的 $r > 2M$ 的在坐标变换下对应着 $u > 0, u^2 > v^2$ ，也就是图中的右 1/4 时空。视界 $r = 2M$ 对应着 $u^2 = v^2$ ，也就是过原点的两条 45° 的直线。这两条直线处线元不再退化，因此可以向另一侧延拓。延拓后的位置为了能与史瓦西坐标系对应，我们也应赋予 t 和 r 坐标。观察右 1/4 时空的坐标变换关系，可以发现每条等 r 线都是双曲线的一支

$$\left(\frac{r}{2M} - 1\right) e^{r/2M} = u^2 - v^2$$

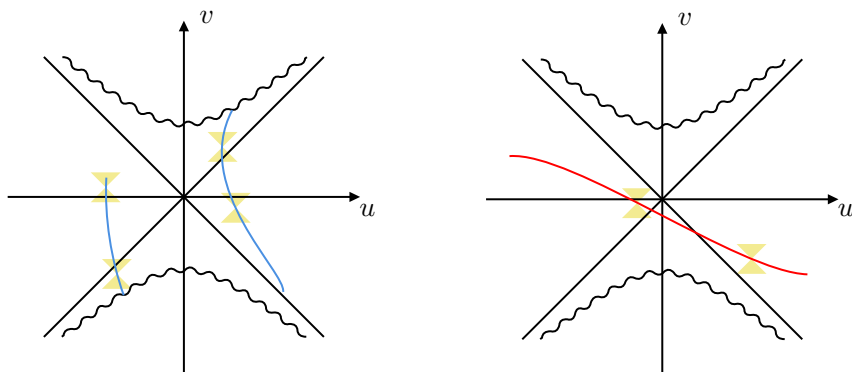
这个关系可以直接延拓至其他区域。而等 t 线是过原点的直线 $t = 4M \operatorname{arctanh}(v/u)$ ，这个关系可以直接延拓至左 1/4 时空。但在上和下 1/4 时空中，这个反双曲正切没有定义。因此我们需要对调 u 和 v ，也就是

$$t = \begin{cases} 4M \operatorname{arctanh}(v/u), & \text{上或下 1/4 时空} \\ 4M \operatorname{arctanh}(u/v), & \text{左或右 1/4 时空} \end{cases}$$

这样，在整个平面上我们都有了 t 和 r 坐标。很容易发现此时 $r = 0$ 对应着 $v^2 - u^2 = 1$ ，这是位于上下 1/4 时空的一对双曲线。奇点在图中一般用波浪线或锯齿线表示。由于此处是真奇点，时空不能延拓至另一侧。因此，最大延拓的时空包括这两条双曲线之间的整个区域。

下面我们来分析一下延拓后的时空的因果结构。为了得到 Kruskal 坐标系下的光锥方程，我们可以令 $ds^2 = 0$ 。此时我们也只考虑径向的光锥，即忽略掉 θ 和 ϕ 坐标。很容易得到光锥方程为 $du = \pm dv$ ，即两条 45° 的直线。由于在右 1/4 时空中向上 t 是增大的，因此未来光锥是向上的。这是 Kruskal 坐标系的一个重要优势——它的光锥形状与闵氏时空中惯性系相同。由光锥的取向很容易发现，上 1/4 时空内的未来光锥一定位于该区域内，且一定指向奇点。这说明该区域的质点无法逃出该区域，且一定会落入奇点（下面左图）。这部分实际上就是我们之前讨论过的黑洞。而在下 1/4 区情况正好相反，其未来光锥指向外部，因此质点一定会离开该区域，该区域就是白洞。而左右是两块对称的区域，它们对应于黑洞/白洞外部的渐进平直区，质点可以保持在其中，也可以进入黑洞，但无法进入白洞。

现在我们知道，整个时空由黑洞和白洞连接的两个渐进平直区构成，这种结构称为**虫洞**，或**爱因斯坦-罗森桥**。虫洞经常出现在科幻作品中，作为连接宇宙两处的“捷径”。然而，至少对于史瓦西虫洞来说，这并不可能实现。这是因为连接两个渐进平直区的世界线必然有类空部分（下面右图），而这是不被相对论所允许的。而试图穿越虫洞的类时曲线必然会落入奇点。若我们把 v 看作某种意义下的时间，那么虫洞可以被看作一个动态结构，它只在上下两个奇点之间是开放的，而在这之前或之后内部都是奇点。在质点穿过虫洞的过程中，虫洞就会闭合，因此质点最终只能落入奇点中。

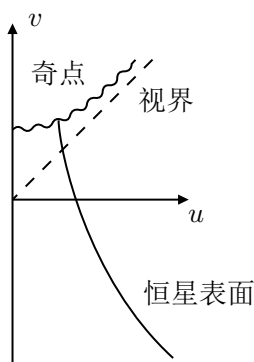


8.1.4 球对称恒星的坍缩与史瓦西黑洞

我们上面讨论了史瓦西真空解的最大延拓，其中出现了黑洞、白洞和虫洞的结构。许多物理学家对这种时空是否能真实存在表示怀疑，但一般大家都认为黑洞是可以真实存在的。计算表明，若晚期恒星的质量超过了一定上限，那么恒星内部的压力就无法抵抗引力，恒星的体积最终会坍缩至无限小，成为一个黑洞。

这个质量阈值称为 Tolman-Oppenheimer-Volkoff (TOV) 极限。

下面我们来分析一下理想状态下恒星的坍缩。我们假设恒星是完全球对称且没有自转的，那么 Birkhoff 定理告诉我们，它外部的时空可以用史瓦西度规描述，哪怕是在坍缩过程中。为了能描述坍缩的过程及其内部的时空，我们选择 Kruskal 坐标系。下图显示了坍缩这一物理过程的时空。注意在坍缩为奇点之前，恒星是有体积的，也就是说这一部分时空不是真空，也不能用史瓦西度规描述，那么也就阻挡住了史瓦西度规的延拓。由下图可以发现，恒星内部阻挡住了度规向白洞和另一个渐进平直区的延拓，但部分黑洞区没有被阻挡住。因此，我们有理由认为，黑洞是可以真实存在的。



一些人认为，在远处的观测者看来恒星是无法坍缩成黑洞的。因为初始状态下恒星的半径大于史瓦西半径，而在无穷远处的静态观测者看来，恒星表面无法在有限长的时间内穿过视界。但实际上即使这样，恒星表面发出的光也会因为靠近视界而导致红移越来越大，发出光的能量越来越低。因此在无穷远处的静态观测者看来，恒星的亮度会迅速降低，成为字面意义的“黑洞”。

8.1.5 彭罗斯图

我们前面已经看到，黑洞具有非常奇特的因果结构。本节来介绍讨论时空整体因果结构的一个重要的工具——彭罗斯图。

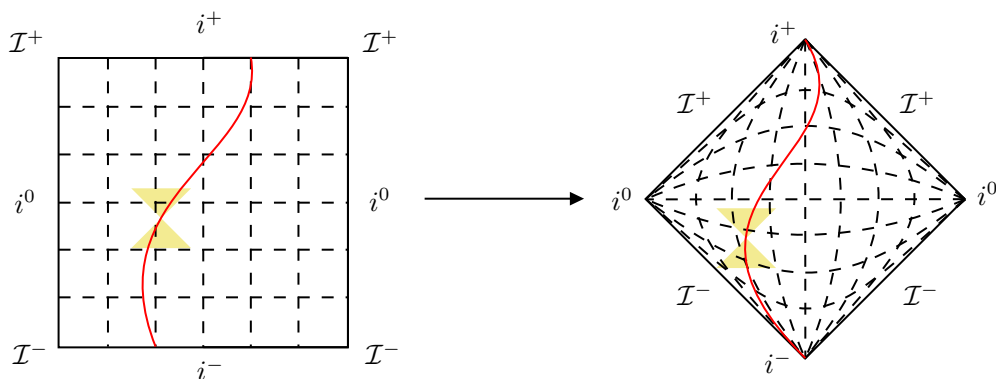
我们首先以平直的闵氏时空为例，介绍彭罗斯图的画法。彭罗斯图的目的是能用一个有限大小的图来体现无限大的时空中的因果结构。为了能在有限大小的图中画出整个闵氏时空，我们需要重点关注无穷远处的点。我们可以将时空中的无穷远分为下面几个类型

- 过去类时无穷远，记作 i^- ；
- 未来类时无穷远，记作 i^+ ；
- 过去类光无穷远，记作 \mathcal{I}^- 或 \mathcal{J}^- ；
- 未来类光无穷远，记作 \mathcal{I}^+ 或 \mathcal{J}^+ ；
- 类空无穷远，记作 i^0 。

出于一些数学上的原因，我们将这个无穷大的时空进行如下变换：我们将两种类时无穷远和一种类空无穷远分别压缩为一个点，并放在有限远的位置。这样，位于它们之间的两种类光无穷远成为了连接它们之间的直线。如右图所示。这个图称为 **彭罗斯图**，或 **共形图**。

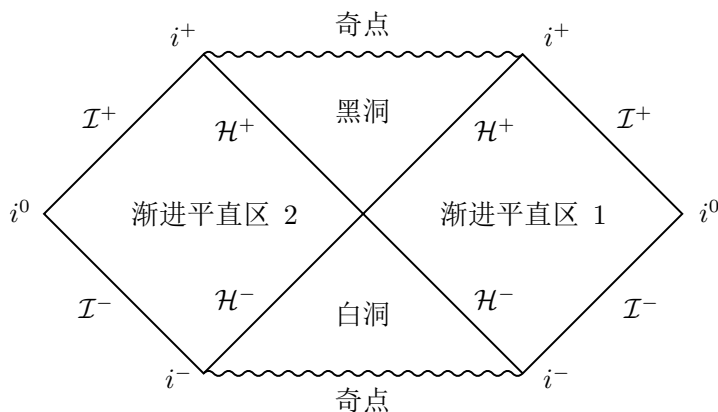
这个过程的数学原理叫做共形变换，它的重要性质是保持光锥。

也称为 Penrose-Carter 图。



由于我们将无穷远的点压缩的过程是关于光锥（即 45° 斜线）对称的，因此光锥本身的形状不会改变，彭罗斯图中的光锥仍然是两条 45° 斜线。而对于质点走的类时曲线，它除了必须位于光锥内部，还必须连接过去类时无穷远 i^- 和未来类时无穷远 i^+ ，如上图红线所示。这两条性质是彭罗斯图最重要的特点，它使得我们可以很容易地使用彭罗斯图来讨论时空的因果结构。

闵氏时空的因果结构是十分平凡的，没有过多讨论的内容。下面我们稍进一步，来看一下最大延拓的史瓦西时空的彭罗斯图。其中我们用 \mathcal{H} 表示视界（ \mathcal{H}^- 表示过去视界， \mathcal{H}^+ 表示未来视界）。



可以发现，最大延拓史瓦西时空的彭罗斯图与 Kruskal 坐标系下的图的结构十分类似，只是形状略有区别。如果在这个彭罗斯图上画出光锥，可以很容易地发现我们之前用 Kruskal 坐标系讨论过的因果结构，如黑洞和白洞视界的单向性等。我们在此不再赘述。彭罗斯图在后面分析其他黑洞时会有更重要的用武之地。

8.1.6 带电球对称黑洞*

在一些情况下，球对称的天体可能带电。此时天体外的时空含有电磁场，不再是真空。此时时空的结构被 **Reissner-Nordström 度规** 描述，它在史瓦西坐标系下的线元为

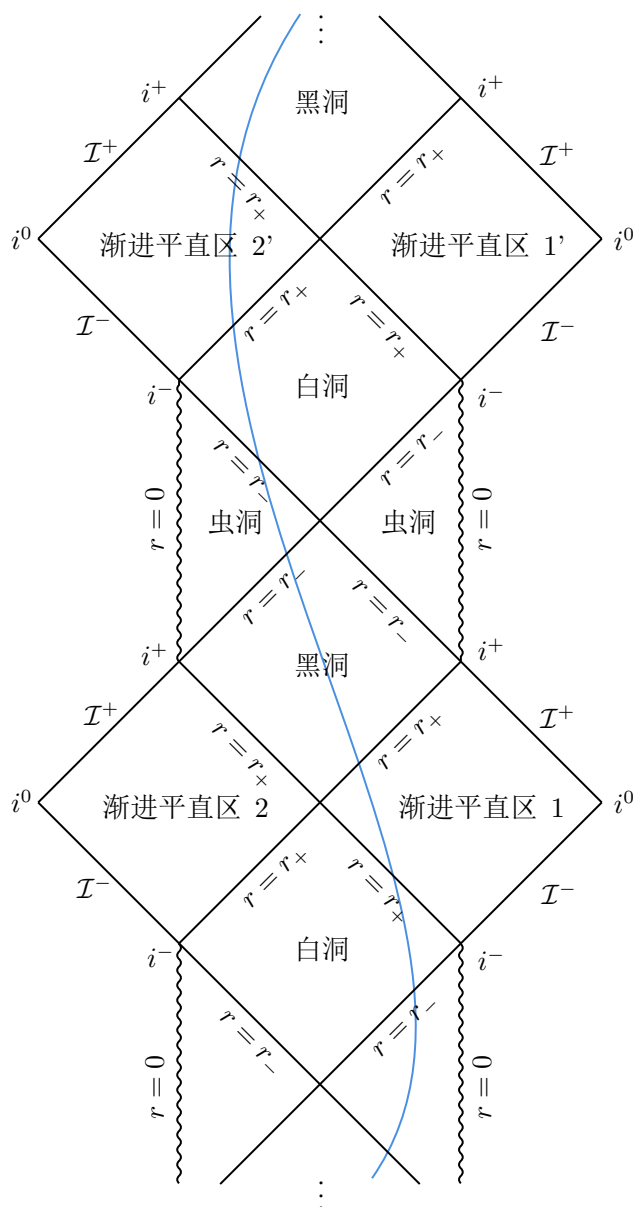
$$ds^2 = -\left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right) dt^2 + \left(1 - \frac{2M}{r} + \frac{Q^2}{r^2}\right)^{-1} dr^2 + r^2 d\Omega^2$$

很容易发现球对称的中心 $r = 0$ 为时空奇点，还有位于 $r_{\pm} = M \pm \sqrt{M^2 - Q^2}$ 的两个坐标奇点，分别称为内视界和外视界。两个视界都是只进不出的事件视界。这两个视界的存在使得 Reissner-Nordström 时空的延拓非常复杂。我们在这里不去具体计算它的延拓的表达式，而是直接给出其最大延拓时空的彭罗斯图，如下图所示。

许多地方也会把闵氏时空的彭罗斯图只画出右半边，这是闵氏时空只有一个 \mathcal{I}^- 、 \mathcal{I}^+ 和 i^0 。

一些书籍资料中会在度规中加入磁荷 P ，只需将式中的 Q^2 变为 $P^2 + Q^2$ 即可。

物理上一定有 $M \geq |Q|$ ，因为天体的质量需要大于电磁场的能量。 $M = |Q|$ 的 RN 黑洞称为极端 RN 黑洞，我们在此不讨论。

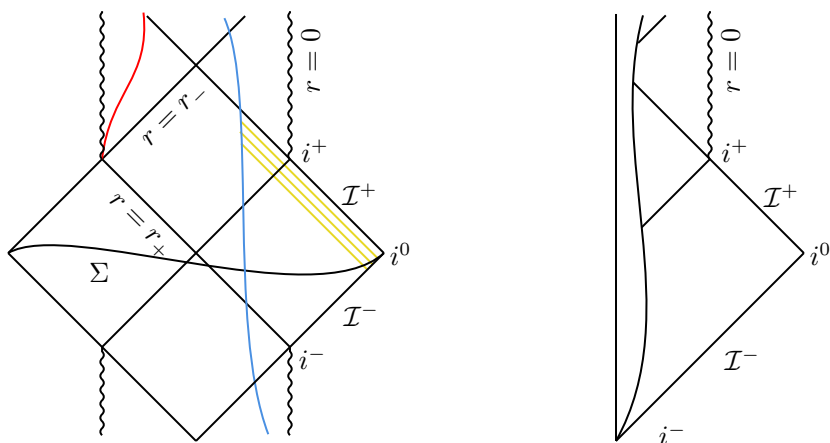


可以发现, Reissner-Nordström 最大延拓时空与史瓦西最大延拓时空有很大不同. 其 $r = 0$ 处的奇点是类时的, 这使得进入黑洞的粒子不必落入奇点. 而没有落入奇点的类时曲线不应终止在某处, 因此其彭罗斯图在纵向被无限重复. 进入黑洞的粒子可以从另一个重复单元的白洞穿出, 进入另一个渐近平直宇宙 (如上图中的蓝线所示). 事实上, 虽然我们可以画出通向奇点的类时曲线, 然而这类类时曲线的 4-加速度是发散的, 因此这个奇点在物理上是不可到达的.

Reissner-Nordström 时空的另一个不同的性质就是其物理可预言性发生了破坏. 在闵氏时空和史瓦西时空中, 若我们做一个连接两个 i^0 的超曲面, 则任何类时或类光曲线都一定会通过这个超曲面, 因此由这个超曲面的信息可以推知时空的整个过去和未来的一切事件. 然而在 Reissner-Nordström 时空中, 类时和类光曲线可以终止于奇点或无穷远处 (如下面左图中的红线所示), 而不必穿过我们做出的超曲面. 这些曲线一定都处于 $r = r_-$ 外部. 因此我们给 $r = r_-$ 处一个名字, 叫做 **柯西视界**. 在柯西视界之外的事件不再能由超面上的初值来预测, 例如从奇点或无穷远处可能突然冒出来一个什么东西, 我们完全无法预测, 因此柯西视界是物理可预言性的界限.

闵氏时空和史瓦西时空的这种性质称为 **整体双曲**.

与最大延拓的史瓦西时空相似，最大延拓的 Reissner–Nordström 时空也被认为不太可能真实存在。这不仅有与史瓦西时空类似的原因——坍缩天体内部不是真空——即使这样，很容易发现时空中仍然会有无穷个渐进平直宇宙（如下面右图）。同时还有另一个重要的原因，那就是柯西视界的存在。当一个观测者靠近柯西视界时，它会接收到来自后方渐进平直时空的“无穷远”处的信息（如下面左图的蓝线是观测者，黄线是无穷远处传来的信息）。换句话说，就是渐进平直时空中无限长时间的信息会被压缩至一瞬间传递给观测者，使观测者接受到无穷大的能量。这意味着时空的任何微小扰动都会在柯西视界处产生无穷大的扰动，使得时空结构与 Reissner–Nordström 时空产生极大的偏离。因此，最大延拓的 Reissner–Nordström 时空是不稳定的，无法真实存在。



8.2 旋转黑洞*

史瓦西度规描述了球对称天体外部的时空。然而，许多天体具有自转，不满足球对称性。本节我们来讨论有自转的天体和黑洞。

8.2.1 Kerr 度规*

描述旋转天体外部的度规称为 **Kerr 度规**，其推导较为复杂，我们在此直接给出其线元表达式

$$ds^2 = -dt^2 + \frac{\rho^2}{\Delta} dr^2 + \rho^2 d\theta^2 + (r^2 + a^2) \sin^2 \theta d\phi^2 + \frac{2Mr}{\rho^2} (a \sin^2 \theta d\phi - dt)^2$$

其中

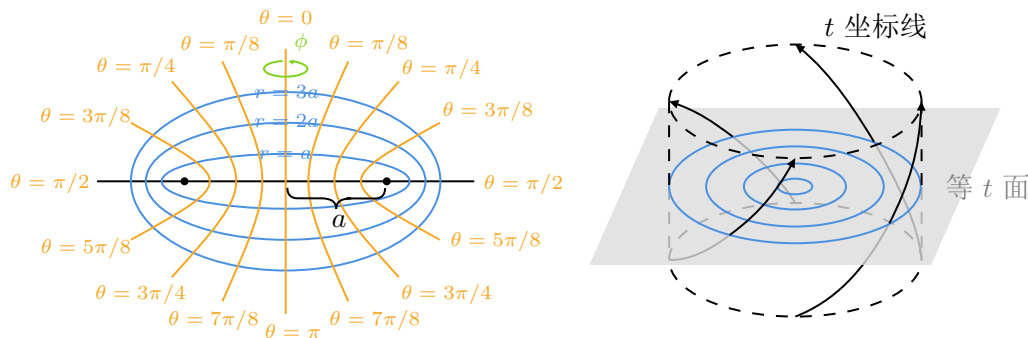
$$\Delta(r) = r^2 - 2Mr + a^2, \quad \rho^2(r, \theta) = r^2 + a^2 \cos^2 \theta$$

注意到 Kerr 度规多引入了一个参数 a ，它代表天体单位质量的角动量，即 $a = J/M$ 。当 $a = 0$ 时，Kerr 度规退化为史瓦西度规。

我们这里使用的坐标系 $\{t, r, \theta, \phi\}$ 称为 **Boyer-Lindquist 坐标系**，而非球对称的史瓦西坐标系。为了理解这个坐标系的几何意义，我们首先考虑 $M \rightarrow 0$ 的情况，此时度规应退化为闵氏度规。根据此时线元的形式，可以发现其空间坐标系是一种椭球坐标系（如下面左图所示）。 ϕ 坐标表示绕天体自转轴的角度，坐标系沿 ϕ 旋转对称。在给定 ϕ 坐标的截面下，等 r 线是一系列半焦距为 a 的椭圆，而等 θ 线为一系列半焦距为 a 的双曲线。因此在整个空间坐标系中，等 r 面是一系列椭球面，而等 θ 面是一系列单叶双曲面，它们都沿 ϕ 旋转对称。对于 $M \neq 0$ 的弯曲时空，我们很难给出每个坐标的具体含义。然而其空间坐标系仍然代表一种沿 ϕ 旋转对称的椭球坐标系。

将式中的 $2Mr$ 替换为 $2Mr - Q^2$ 即可得到有电荷的旋转天体的度规，称为 **Kerr-Newman 度规**，其定性性质与 Kerr 度规类似，因此我们这里只讨论没有电荷的情况。

下面我们再考虑时间坐标. 在 Kerr 度规中, 有时空交叉项 $-4Ma\rho^{-2}r\sin^2\theta d\theta d\phi$, 这意味着 t 轴不再垂直于空间坐标轴. 注意到度规中 t 只与 ϕ 坐标有交叉项, 而与 r 和 θ 坐标没有交叉项. 这说明时间轴是垂直于 r 轴和 θ 轴的, 但不垂直于 ϕ 轴. 很容易发现, 这意味着 t 坐标轴是沿着 ϕ 的方向旋转的 (如下面右图所示). 不过需要强调的是, 虽然我们直观上画出 t 坐标线是“旋转的”, 但 t 坐标线上 r, θ, ϕ 是不变的.



根据线元的形式, 很容易发现 $\partial/\partial t$ 和 $\partial/\partial\phi$ 是 Kerr 时空的两个 Killing 矢量场, 其中后者类空, 前者在 r 较大时类时. 然而, 在 Kerr 时空中, 我们无法找到正交于 $\partial/\partial t$ 的超曲面, 因此它是一个稳态时空而不是静态时空. 从几何直观地角度来说, 这对应着矢量 $\partial/\partial t$ 沿着 ϕ 方向的旋转. 我们后面会看到, 这体现了天体自转时“对时空的拖拽”.

我们后面会看到在 r 较小时 $\partial/\partial t$ 也类空, 这部分时空就连稳态时空都不是了.

8.2.2 Kerr 时空的结构*

下面我们来分析 Kerr 时空的奇点和视界. 使线元奇异的位置有 $\Delta = 0$ 或 $\rho = 0$, 我们首先来看后者. $\rho = 0$ 意味着

$$r = 0 \quad \text{且} \quad \theta = \pi/2$$

这在空间上是处于天体赤道面上、以天体旋转轴为中心的一个环, 称为 **奇环**. 这里是一个真奇点, 与静态球对称黑洞的中心处类似. 而线元发散的另一种情况 $\Delta = 0$ 对应着

我们在此考虑自转不是特别快的黑洞, 即 $M > a$.

$$r_{\pm} = M \pm \sqrt{M^2 - a^2}$$

这两个位置是 Boyer-Lindquist 坐标系的坐标奇点. 与前面讨论的 Reissner-Nordström 黑洞类似, 它们也是只进不出的事件视界, 分别称为内视界和外视界.

Kerr 时空的一个很重要的性质就是它的时空与中心天体一起发生了旋转, 这被称为 **参考系拖拽**. 若我们考虑沿着 ϕ 方向发出的光. 这束光的 4-速度为类光矢量. 由线元表达式很容易解得光的瞬时“角速度”为 $d\phi/dt = -g_{03}/g_{33} \pm \sqrt{(g_{03}/g_{33})^2 - g_{00}/g_{33}}$. 可以发现光锥的运动方向是向一侧偏的. 那么我们可以定义“时空的角速度” Ω 为光锥“中点”的角速度. 为了方便, 我们只考虑赤道面上的情况, 有

以 Ω 为角速度旋转的观测者称为 **局域非转动观测者** 或 **零角动量观测者** (ZAMO).

$$\Omega = -\frac{g_{03}}{g_{33}} = \frac{2Ma}{r^3 + a^2r + 2Ma^2}$$

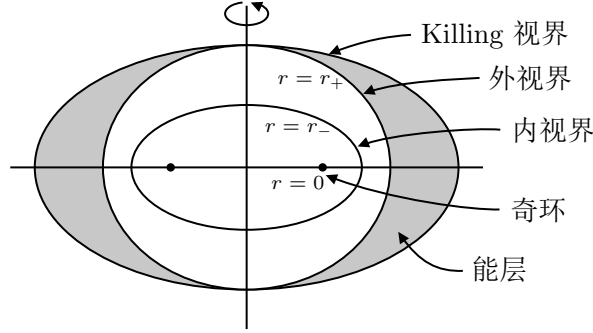
可以发现在 r 较大时, 时空的旋转“线速度” Ωr 随着 r 减小而增大. 那么当 r 小到一定程度时, 时空的旋转可能超过光速, 使得任何粒子无法保持静止不动, 而必须随着黑洞转动. 若我们考虑 Killing 矢量场 $\xi^\mu = \partial x^\mu / \partial t$ 的类时/类空性, 根据度规有 $\xi^\mu \xi_\mu = -\rho^{-2}(\Delta - \sin^2\theta)$. 可以发现 ξ^μ 在类时和类空的交界处为

$r = M - \sqrt{M^2 - a^2 \cos^2\theta}$ 处也是一个 Killing 视界, 在其内部 ξ^μ 又变为类时, 但我们先不考虑这部分.

$$r = M + \sqrt{M^2 - a^2 \cos^2\theta}$$

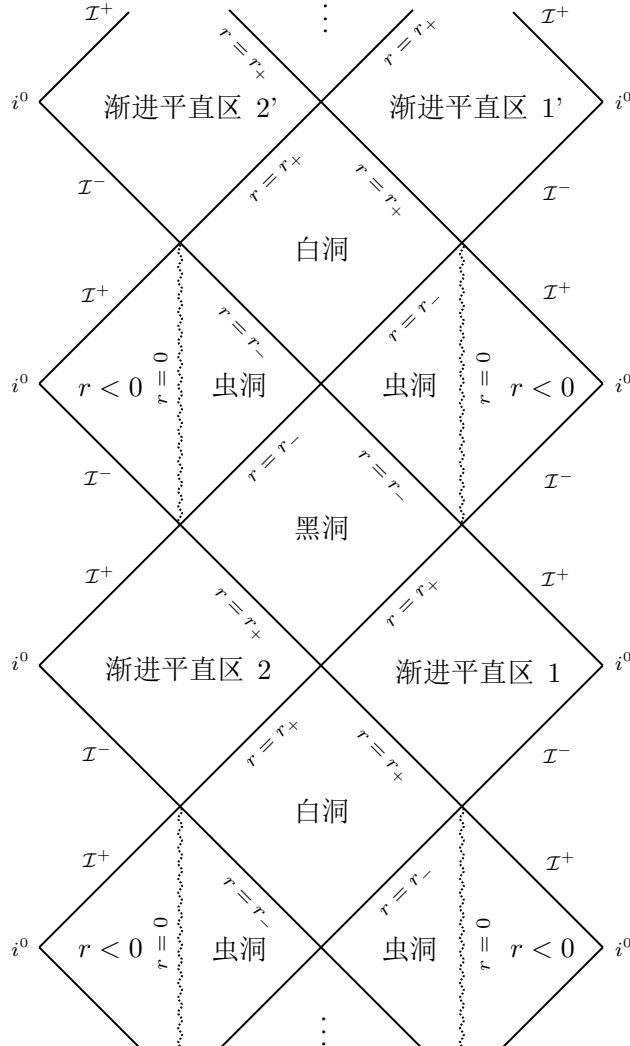
此处称为一个 **Killing 视界**，在 Killing 视界以外， ξ^μ 类时，也就是说它可以代表一族稳态观测者，其空间坐标不变，因而可以称为“静止观测者”；而在这个 Killing 视界以内， ξ^μ 类空，静止观测者不复存在，任何观测者一定会随着黑洞一起转动，因此这个 Killing 视界又称为 **稳态极限**。我们称稳态极限与外视界之间的部分为 **能层**，我们后面会看到这个名字的意义。

注意静止观测者不是静态观测者。同时，以固定角速度旋转的观测者也是稳态观测者，但不是静止观测者。



8.2.3 Kerr 时空的最大延拓*

对于 Kerr 时空，我们也不推导其最大延拓的表达式，而是直接给出其最大延拓时空的彭罗斯图。



首先需要指出，彭罗斯图是一个二维图，相比于真正的四维时空压缩掉了 θ 和 ϕ 两维。也就是说彭罗斯图的每个点代表了一个由 r 和 t 确定的二维超曲面。在前面讨论的球对称时空中，彭罗斯图中每点对应的超曲面是一个球面，上面每点的几何是相同的。然而在 Kerr 时空中，自转使得球对称性丧失，因此彭罗斯图上每点不再代表一个几何性质均一的球面代表的超曲面（实际上是一个椭球面）上的几何不再均一。

我们来看一下这个彭罗斯图给出的时空结构。注意到 Kerr 时空的最大延拓的彭罗斯图与 Reissner–Nordström 时空非常类似，都有无限重复的黑洞/白洞和渐进平直时空的单元。但不同的是，Kerr 时空在奇环的另一侧多出了 $r < 0$ 的渐进平直区。这是因为 Kerr 时空的奇环的位置不仅要求 $r = 0$ ，还要求 $\theta = \pi/2$ 。这意味着粒子可以从 $\theta \neq \pi/2$ 的位置穿越到 $r = 0$ 的另一侧，也就是说时空需要延拓至 $r < 0$ 。这也是图中的奇性 $r = 0$ 处画成虚线的原因。由于 $r < 0$ 时 Kerr 线元没有任何奇性，因此延拓后 $r < 0$ 的部分是一个简单的渐进平直时空，没有特殊的结构。

下面让我们讨论最大延拓 Kerr 时空的一个很奇特的性质。考虑一个在赤道面上的很小的闭合圆周，即保持 $\theta = \pi/2$ ， r 为一个绝对值很小的负的定值， t 也为某个定值，只有 ϕ 坐标变化。则这个圆周上的线元近似为

$$ds^2 \approx a^2 \left(1 + \frac{2M}{r} \right) d\phi^2$$

可以发现当 r 是一个绝对值足够小的负数时，可以有 $ds^2 < 0$ ，即这个闭合曲线可以是类时的。闭合类时曲线的存在意味着一个粒子可以在经历了一段时间以后回到了过去自己所在的某个时间和位置，遇到了过去的自己，这是物理上不能接受的。因此霍金提出，应该有某种物理机制可以阻止闭合类时曲线的产生，这称为 **时序保护猜想**，然而目前这种机制并没有被发现。

与最大延拓的史瓦西时空和 Reissner–Nordström 时空类似，最大延拓的 Kerr 时空也被认为在实际上不太可能存在。

8.2.4 彭罗斯过程*

转动的黑洞具有转动动能。彭罗斯指出，可以有一种物理过程从中提取转动动能，这称为 **彭罗斯过程**。而提取能量的位置是 Killing 视界和外视界之间的区域，就是这个区域被称为能层的原因。下面我们来分析一下这个过程。

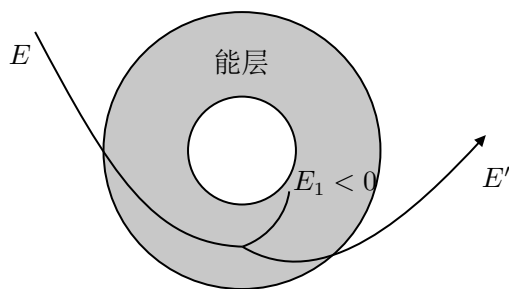
与史瓦西时空类似，我们可以通过 Killing 矢量场来构造测地运动中的守恒量。对于 Killing 矢量场 $\xi^\mu = \partial x^\mu / \partial t$ ，我们可以构造守恒的能量

$$E = -\xi_\mu P^\mu$$

注意我们此处定义的能量是真的能量，而非单位质量的能量。

其中 P^μ 为粒子的 4-动量，它是一个类时矢量。在能层之外， ξ^μ 是指向未来的类时矢量，因此 $E > 0$ 。然而在能层中， ξ^μ 变为类空，因此 E 可以为负值。我们称具有负的 E 的测地线为负能轨道。

下面我们考虑这样一个物理过程：某个飞船从远处自由下落，落入黑洞的能层中。在能层中，飞船发射出一个物体，使得这个物体进入负能轨道，落入黑洞。那么根据能量守恒，飞船自身的能量将增加。还可以证明，可以找到一种将物体扔到负能轨道的方式，使得扔出物体后的飞船可以沿着测地线离开黑洞，回到远处。经过这个过程后，这个飞船最终的能量将高于初始值。这个能量来自黑洞的自转，因此这样实际上是在提取黑洞的转动动能。这就是彭罗斯过程。



8.3 黑洞的一般性讨论

我们上面以多种典型黑洞解为例，讨论了它们各自的重要性质。下面我们来看一下黑洞所满足的一般规律。

8.3.1 奇点与黑洞无毛定理

我们已经看到了，黑洞是广义相对论框架下预言的一种非常奇特的天体。黑洞有很多非常特殊的性质，但其中最为令人头疼之处是奇点的存在——广义相对论对奇点处的物理定律束手无策。例如前面讨论的 Reissner–Nordström 黑洞中，广义相对论无法阻止起始于奇点的世界线，也就是说在奇点处可以突然冒出来一个什么东西，广义相对论完全无法预言这件事。这令物理学家们很不能接受。因此一些物理学家曾经提出，奇点可能实际上是不存在的。对于我们的前面讨论过的黑洞，它们太过对称了，以至于在中心恰好产生了一个奇点。而真正的恒星是没有这么对称的，那么它们坍缩后或许就不会产生奇点。然而，彭罗斯和霍金证明了 **奇点定理**，它指出，在一些不苛刻的前提下，质量足够大的天体的坍缩一定会产生时空奇点。奇点定理不以对称性为条件，因此偏离对称性不能防止奇点的产生。所以，在广义相对论的框架下，物理学家们必须接受奇点的存在。

然而奇点定理只是经典广义相对论框架下的定理，一些科学家认为在考虑了量子效应后奇点可能不会产生。

我们前面看到了许多奇点的例子，这些奇点都被事件视界所包裹，形成了黑洞。在对于黑洞的讨论中，我们从最基本的史瓦西黑洞开始，逐步增加了电荷和旋转，最终得到了 Kerr–Newman 黑洞。但这些黑洞解都来自一些非常理想的假设——例如在史瓦西黑洞中，我们要求黑洞是完美的球对称的，在 Kerr 黑洞中，我们要求黑洞是完美的旋转对称的。那么真实的黑洞是否一定满足这么理想的假设呢？幸好，我们有 **黑洞无毛定理**：稳态、渐进平直的孤立黑洞解可以被三个参量完全确定：质量、电荷和角动量。所谓“无毛”，是指黑洞不可能具有复杂的细节结构，例如地球表面有凹凸不平的地形、不同类型的物质分布等。而对于一个黑洞，只要它是稳态的，那么这些细节就完全不可能存在。两个具有完全相同质量、电荷和角动量的稳态黑洞的一切性质都完全相同。因此，一个稳态黑洞——现实中恒星若坍缩成黑洞，那么这个黑洞会很快趋近于稳态——一定可以被我们讨论过的 Kerr–Newman 解描述。

黑洞无毛定理目前的证明中有很多限制条件，因此一些人认为它并没有被真正证明，因而称之为黑洞无毛猜想。

然而考虑了量子效应后，黑洞无毛就有很大争论了。

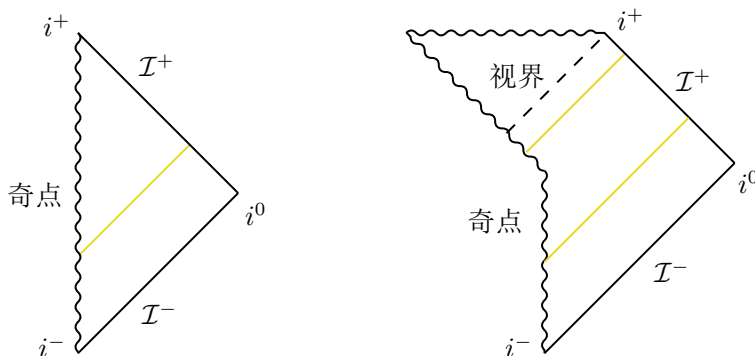
然而，奇点实际上并不一定被事件视界包裹。这种不被事件视界包裹的奇点称为 **裸奇点**。实际上，在我们前面学习的度规中就可以产生裸奇点。例如在 Reissner–Nordström 度规中，若我们要求 $M < |Q|$ ，或在 Kerr 度规中要求 $M < a$ ，则很容易发现内外两个视界都会消失，此时内部的奇点便成为了裸奇点。例如下面左图显示了超极端 Reissner–Nordström 时空的彭罗斯图。实际上，恒星若坍缩成为裸奇点而非黑洞，那么自然就无需满足黑洞无毛定理，此时的时空不一定被 Kerr–Newman 度规描述，其中的裸奇点也可以形形色色。例如下面右图就显示了另一种不同的具有裸奇点的时空。

这种时空称为 **超极端 RN 或 Kerr 时空**。

奇点的存在本身就令物理学家很头疼了，但若它被包裹在黑洞中也不算可以接受——就

算奇点处蹦出来一个什么东西，它也无法穿出事件视界，因而无法对外面的时空产生任何影响。然而裸奇点的存在则不同——裸奇点处蹦出来的东西可以跑到任意远处，对那里发生的事件产生影响，例如下面两图显示了从奇点处发出的类光曲线。为了摒除裸奇点存在的可能性，彭罗斯提出了 **宇宙监督假设**：物理上允许的物质场的坍缩无法产生裸奇点，这被霍金形象地称作“上帝憎恶裸奇点”。然而，宇宙监督假设目前既没有被证明也没有被证伪，因此仍然是一个猜想。

宇宙监督假设不阻止一直存在的奇点或原初奇点。



8.3.2 黑洞热力学

黑洞还有一些性质与传统的热力学系统非常相似，因此称为 **黑洞热力学**。黑洞热力学的意义远超过它在广义相对论和热力学中的意义本身，实际上它在量子引力论的建立中有着重要的作用。

我们下面以 Kerr 黑洞为例讨论黑洞的热力学问题。我们考虑 Kerr 黑洞的事件视界，它是一个椭球面，可以算出其面积为

$$A = 4\pi(r_+^2 + a^2) = 2M^2 + 2M\sqrt{M^2 - a^2}$$

将上式进行微分，并根据角动量 $J = Ma$ ，很容易得到

$$\delta M = \frac{\kappa}{8\pi} \delta A + \Omega_H \delta J$$

其中系数 $\Omega_H = \Omega|_{r=r_+}$ 为视界处时空转动的角速度，而 κ 称为黑洞的 **表面引力**，我们在此不列出其具体形式。

注意到上式与热力学第一定律 $dU = TdS + \delta W$ 的形式非常类似，因此称为 **黑洞的热力学第一定律**。对比两式可以发现，黑洞的质量 M 的地位类似于内能，表面引力 κ 类似于温度，而事件视界 A 类似于熵。

对于黑洞引入了内能、温度和熵的等价概念后，我们就可以讨论更多的热力学定律了。传统的热力学第二定律指出，系统的熵随时间是不减的，即 $\delta S \geq 0$ 。对应到黑洞上，这意味着黑洞的视界面积是不减的，即

$$\delta A \geq 0$$

这就是 **黑洞热力学第二定律**。例如在彭罗斯过程中，落入黑洞的是负能物质，因此黑洞的质量 M 是会降低的。然而即便如此，其视界面积仍然不会减小。另外，在热力学中，熵还是一个系统储存信息的量度。因此黑洞热力学第二定律也可以表述为：黑洞的信息储存在其视界上。

热力学实际上还有第三定律，它指出，热力学零度是无法达到的。对应到黑洞上就是黑洞的表面引力不能为零，这称为 **黑洞热力学第三定律**。实际上，我们在理论上可以构

注意对 Kerr 黑洞等具有内外两个事件视界的黑洞，当我们只说其视界时特指其外视界。

然而考虑了量子效应后，霍金辐射可以降低黑洞的视界面积。

造出表面引力为零的黑洞解，即为 **极端黑洞**，例如 $M = Q$ 的 Reissner–Nordström 黑洞或 $M = a$ 的 Kerr 黑洞。然而，这些极端黑洞具有裸奇点，不满足宇宙监督假设。因此，黑洞热力学第三定律实际上就是宇宙监督假设的体现。

9 宇宙学

在最后一章，我们来把广义相对论应用到我们所在的最大的时空——宇宙，来研究在整个宇宙尺度的物理规律。

我们这里只介绍与广义相对论直接相关的内容，宇宙的热历史、结构形成等内容我们不做讨论。

9.1 宇宙学的基本概念

我们前面的章节都是在纯理论地研究引力和各种时空的性质。而这一章的宇宙学则不同，它分析的是一个具体而实际的物理模型——我们生活的宇宙。那么第一步，我们先需要建立宇宙的基本模型，也就是说宇宙是什么样子的。在此之后，我们才能用各种数学和物理工具来分析这个模型的演化和性质。

9.1.1 宇宙学原理

首先，我们先来建立一个宇宙的定性模型。我们的宇宙模型应该能反映宇宙的特征，而不应拘泥于微观上的细节——我们不应在意具体的一个个天体，而应在意宇宙在大尺度上的性质，否则这个模型将难以计算。人们在大量天文观测的基础上，对宇宙在大尺度上的性质归结出了一个假设，称为 **宇宙学原理**：宇宙空间在大尺度上是均匀且各向同性的。我们后面讲的整个宇宙学模型都是基于宇宙学原理与广义相对论两块基石而衍生出来的，称为 **Friedmann-Lemaître-Robertson-Walker (FLRW) 宇宙模型**。

也称为 Friedmann-Robertson-Walker (FRW) 宇宙，或 Robertson-Walker (RW) 宇宙。

然而，从广义相对论的角度来看，宇宙是一个四维时空，它并不天生就有时间-空间的分解。因此，我们有必要以四维语言重新表述一下宇宙学原理：在宇宙时空中存在一种时空正交分解方式，使得每一个时刻 t 的全空间 Σ_t 是均匀且各向同性的。这样，我们在宇宙学中就有了默认的时间和空间的概念，我们称这个时间为 **宇宙时**。后面我们见到时间坐标 t 时，就默认是宇宙时。

我们这里定义的宇宙时代表了一族观测者，称为 **基本观测者** 或 **共动观测者**，而宇宙空间 Σ_t 就是基本观测者所观测到的全空间。基本观测者在平均意义上与宇宙中的物质一起运动，因此在它们看来宇宙是各向同性的，也就是说在它们看来宇宙中的物质不会整体向某一边运动，它们在宇宙中代表“静止”的观测者。当然，具体的每一个天体或者星系与基本观测者之间是可能有相对运动的，我们称天体的这种运动为 **本动**。也就是说，对于基本观测者来说，宇宙中的物质的平均本动速度为零。

9.1.2 宇宙的空间几何与 FLRW 度规

宇宙学原理只是对宇宙模型的一个定性假设。下面，我们需要根据宇宙学原理而建立宇宙的数学模型。宇宙学原理告诉我们，宇宙空间是各向同性的，因此宇宙空间的度规一定是一个三维球对称度规。在推导史瓦西度规时，我们已经说明了三维球对称度规一定可以写作

$$dl^2 = g_{11}(r)dr^2 + r^2 d\Omega^2$$

其中 $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$ 为单位球面度规。同时，我们还要求宇宙空间是均匀的，这可以表述为其标量曲率 R 处处相等。以这个为方程求解 g_{11} ，可以得到

$$g_{11}(r) = \frac{1}{1 - kr^2}$$

其中常量 $k \in \{-1, 0, 1\}$ 代表宇宙空间的曲率类型。这个度规看起来有些奇怪，下面我们就来分析一下不同的 k 分别对应着什么样的宇宙空间几何。

这里推导时取的都是宇宙学原理的必要条件，不过可以证明这个度规一定是满足宇宙学原理的。

- $k = 1$: 我们可以做坐标变换 $r = \sin \chi$, 则度规可以化为

$$dl^2 = d\chi^2 + \sin^2 \chi d\Omega$$

与二维球面度规类比, 很容易理解这就是三维球面度规. 此时宇宙称为 **闭合宇宙** 或 **de Sitter (dS) 宇宙**.

对比二维球面度规 $d\Omega^2 = d\theta^2 + \sin^2 \theta d\phi^2$ 可以发现二者的相似性.

- $k = 0$: 此时度规为

$$dl^2 = dr^2 + r^2 d\Omega$$

很容易发现这就是平直的欧式度规的球坐标形式. 此时宇宙称为 **平直宇宙**.

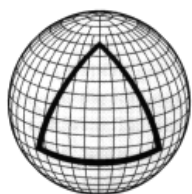
- $k = -1$: 我们可以做坐标变换 $r = \sinh \psi$, 则度规可以化为

$$dl^2 = d\psi^2 + \sinh^2 \psi d\Omega$$

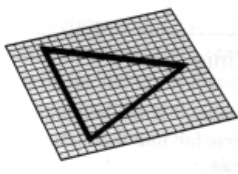
这实际上是一个三维双曲面的度规. 此时宇宙称为 **开放宇宙** 或 **反 de Sitter (AdS) 宇宙**.

我们的数学推导表明了, 满足宇宙学原理的宇宙空间的几何结构仅有可能是这三种之一. 这三种宇宙常常用下面的图来直观地表示. 不过这只是一通俗的表示方式而已, 不必深究.

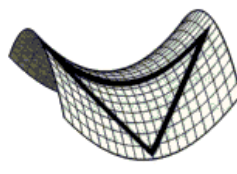
宇宙的闭合与开放只是一个名字, 它们并不代表宇宙的大小是有限或无限的, 这是因为度规作为局部性质, 无法决定整体的拓扑性质.



闭合宇宙

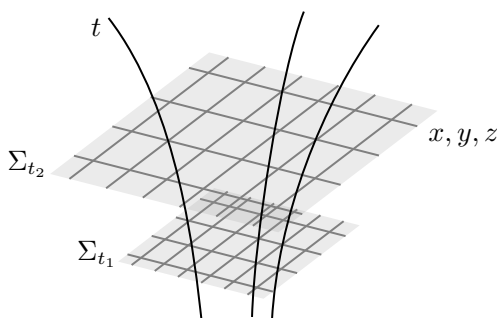


平直宇宙



开放宇宙

我们上面讨论的是宇宙空间 Σ_t 的度规. 下面, 我们只需再考虑进时间的维度, 即可得到宇宙作为四维时空的度规. 与史瓦西坐标系的思路相同, 我们可以把已有的三维球坐标系与宇宙时共同构成四维的坐标系, 称为 **共动坐标系**. 在共动坐标系中, 时间坐标线代表基本观测者, 即基本观测者的空间坐标是不变的; 而其等 t 面即为我们上面提到的均匀、各向同性的宇宙空间 Σ_t .



下面我们来写出 FLRW 宇宙的度规在共动坐标系下的线元. 首先, 我们知道时间轴, 即共动观测者的世界线是正交于宇宙空间的. 因此度规中不应有时空交叉项, 即它一定形如

$$ds^2 = g_{00} dt^2 + g_{ij} dx^i dx^j$$

由于宇宙对于共动观测者是均匀的, 因此 g_{00} 分量一定不能是空间坐标的函数, 只能是时间 t 的函数. 那么我们自然可以对时间轴进行伸缩, 使得 g_{00} 处处为 -1 . 这样, 我们的线元就一定可以写作 $ds^2 = -dt^2 + g_{ij} dx^i dx^j$.

下一步我们再考虑空间分量。我们上面推出的 dl^2 应为宇宙的等 t 面的空间线元，这实际上就是在说 $g_{ij}dx^i dx^j = dl^2$ 。不过，我们说宇宙的空间线元只能是上面推出的 dl^2 ，是说“形状”必须与它相同，但是其整体的大小是可以变化的。这样，我们可以引入 **尺度因子** a ，用于放缩空间度规部分。为了使它不改变宇宙的空间结构，它自然只能是时间的函数，即 $a(t)$ ，它用于描述宇宙的空间度规随时间的变化。这样，时空线元一定形如 $ds^2 = -dt^2 + a^2(t)dl^2$ 。于是我们就得到了宇宙学中的标准度规，称为 **FLRW 度规**。

定理 9.1: Friedmann-Lemaître-Robertson-Walker (FLRW) 度规

FLRW 宇宙学模型中，时空的度规称为 FLRW 度规，其线元可以写为

$$ds^2 = -dt^2 + a^2(t) \left(\frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right)$$

这样，我们就从定性的宇宙学原理开始，建立了一个精确的数学模型，FLRW 度规就是宇宙的数学模型的基石。

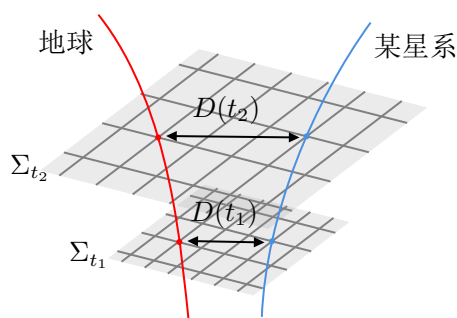
9.1.3 哈勃定律

1929 年，天文学家哈勃通过恒星光谱的红移发现了宇宙中所有的星系都在向远离地球的方向移动，且其速率（称为退行速率）正比于该星系距地球的距离，这称为 **哈勃定律**。我们记比例常数为 H_0 ，称为 **哈勃常数**，则有

$$v = H_0 D$$

其中 D 为星系距地球的距离， $v = \dot{D}$ 为退行速率。哈勃定律虽然是一个观测定律，但实际上已经蕴含在了 FLRW 度规中。当天体离我们足够远时，其本动速度相对于退行速度可以忽略。宇宙中的天体在忽略掉本动速度后，可以视作宇宙的基本观测者，也就是说他们的空间坐标是不变的。然而，宇宙的空间度规在正比于 $a(t)$ 地变化，因此天体之间的距离 D （称为物理距离）是正比于 $a(t)$ 的。下图显示了这种关系。

所有带上点的变量都代表对宇宙时求导。



进而我们可以得到哈勃常数的表达式

$$H(t) = \frac{\dot{D}(t)}{D(t)} = \frac{\dot{a}(t)}{a(t)}$$

由此我们可以发现，哈勃常数实际上并非常数，它是可以随着时间变化的，因此也称为 **哈勃参量**。 H_0 实际上是我们现在所处的时刻的哈勃参量。

我们后面一律用角标 0 表示当今值。

最后我们需要指出，哈勃定律表明，当某天体距离我们足够远时，它的退行速率是可以超过光速的，这个距离阈值称为哈勃半径。但这个现象与相对论中“物质运动不可超光速”并不矛盾。回想我们在 3.2.1. 小节中强调过，“物质不可超光速”实际上应该更加严

谨地用四维语言表述为“质点的世界线为类时曲线”，或者等价地表述为“质点的世界线只能在光锥内部”。而在上面的讨论中，星系的世界线的确确为类时曲线，因此并没有违反这一条原理。

9.2 宇宙动力学

我们在上一节给出了 FLRW 宇宙度规的形式，其中除了决定宇宙开闭的参数 k 外，还有尺度因子 $a(t)$ 没有被确定，而它描述了宇宙随时间的演化。本节我们将研究宇宙中的物质场如何决定宇宙的演化。

9.2.1 Friedmann 方程

宇宙作为一个时空，它的演化当然取决于其中的物质场，演化方程即为爱因斯坦场方程。我们如果假设宇宙的内容物是一种理想流体，其能动张量为

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

其中 ρ 为基本观测者测得的宇宙内容物的总能量密度， p 为基本观测者测得的压强。由宇宙学原理可知 ρ, p 和 U^μ 都应是空间均匀的，即它们都只能是 t 的函数。将这个能动张量和 FLRW 度规带入爱因斯坦场方程，可以得到时间和空间的分量方程。这个方程是 FLRW 宇宙模型的最基本方程，称为 Friedmann 方程。由于推导较为繁琐而没有技巧，我们直接给出结果。

我们这里用“内容物”代表宇宙中所有的东西，包括普通物质（简称物质）、辐射等。我们后面会讨论宇宙中的各个组分，到时候可以分别看到它们的确是一种理想流体。

定理 9.2: Friedmann 方程

在 FLRW 宇宙学模型中，尺度因子 $a(t)$ 的演化满足

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

有时 Friedmann 方程特指其中的第一个方程，第二个方程称为 Friedmann 第二方程。

Friedmann 方程描述了尺度因子随时间的演化，我们将在下一节具体计算这个问题。在本节，我们先从 Friedmann 方程开始讨论另外两个重要结论。

首先，从 Friedmann 第二方程很容易发现 $\ddot{a} < 0$ ，也就是宇宙的膨胀速率 \dot{a} 是在减小的。那么沿时间反向回溯回去，宇宙的收缩将越来越快，必定某个时间处有 $\dot{a} = 0$ ，此时宇宙空间缩为一个奇点。也就是说，宇宙是由一个点膨胀开产生的，这称为宇宙 **大爆炸**，对应的时空奇点称为大爆炸奇点。宇宙时在大爆炸奇点达到最小值，我们令此处为宇宙时的零点 $t = 0$ 。由于我们定义的宇宙时坐标在 $t = 0$ 处为奇点，因此宇宙时不能为负。“在大爆炸之前发生了什么”这种问题是没有意义的，因为宇宙时空不存在 $t < 0$ 的点。

后面我们会讲到，实验发现宇宙是加速膨胀的，所以此处的推理并不能用于我们的宇宙。然而许多其他的证据仍然支持大爆炸的存在。

既然 Friedmann 方程指出宇宙是有开端的，那么一个重要的问题是宇宙目前的年龄。如果我们做一个粗略的计算，假设宇宙始终以匀速膨胀，即 $a \propto t$ ，则很容易推出宇宙的现在的年龄为

$$t_0 = \frac{a_0}{\dot{a}_0} = H_0^{-1}$$

我们称 H_0^{-1} 为哈勃时间，其数值约为 1.44×10^{10} 年。由于宇宙的膨胀并不是匀速的，实际上我们后面会看到宇宙实际上是加速膨胀的，因此宇宙的年龄实际上略小于哈勃时间。现在宇宙的年龄的估计值被修正为约 1.38×10^{10} 年。

另外一个重要的问题是我们之前讨论过的宇宙的空间几何。我们知道， k 的符号决定了宇宙的空间几何，而 k 也出现在 Friedmann 方程中。对于临界的 $k = 0$ 的情况，可以发现 $H^2 = (\dot{a}/a)^2 = \frac{8}{3}\pi\rho$ ，从而解得此情况下的宇宙内容物总密度为

$$\rho_c = \frac{3H^2}{8\pi}$$

我们称这个密度为 **临界密度**。宇宙的空间几何取决于宇宙中真实的密度与临界密度之间的大小关系。我们进一步定义 **密度参数** 为宇宙中真实内容物的密度与临界密度的比值，即

$$\Omega = \frac{\rho}{\rho_c}$$

那么我们就可以根据密度来推知宇宙的空间几何：

- 若 $\rho < \rho_c$ ，即 $\Omega < 1$ ，则 $k = -1$ ，即宇宙为开放宇宙；
- 若 $\rho = \rho_c$ ，即 $\Omega = 1$ ，则 $k = 0$ ，即宇宙为平直宇宙；
- 若 $\rho > \rho_c$ ，即 $\Omega > 1$ ，则 $k = 1$ ，即宇宙为闭合宇宙。

天文观测显示，目前宇宙的密度参数 Ω_0 非常接近于 1，也就是说当今宇宙在空间上几乎是平直的。

9.2.2 宇宙学常数与暗能量

我们前面在 FLRW 宇宙模型中理论推导出了宇宙的膨胀速率 \dot{a} 是一直在降低的，也就是说尺度因子 a 不可能一直保持不变。然而在宇宙膨胀被天文观测证实之前，许多人认为宇宙应该是静态的，整体上不应随着时间变化。为了建立起静态的宇宙模型，爱因斯坦在 1917 年对他的引力场方程进行了修改，引入了所谓的宇宙学常数。

回顾我们在 3.5.2 小节中构造爱因斯坦场方程的过程。我们要求等式左侧的张量是一个对称、且协变散度为零的张量，从而构造出了爱因斯坦张量 $G_{\mu\nu}$ 。但是注意到度规张量也是一个对称张量，且有 $\nabla^\sigma g_{\mu\nu} = 0$ 。因此，我们给爱因斯坦张量加上 $g_{\mu\nu}$ 的任意常数倍，都可以满足要求。这样，爱因斯坦场方程可以拓展成

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi T_{\mu\nu}$$

其中 Λ 是一个常数，称为 **宇宙学常数**。这个新的爱因斯坦场方程虽然与原来不同，但我们前面的分析并没有白费。这是因为，这个方程可以被化为

$$G_{\mu\nu} = 8\pi \left(T_{\mu\nu} - \frac{\Lambda}{8\pi} g_{\mu\nu} \right)$$

从这个写法可以看出，宇宙学常数 Λ 的作用其实相当于在物质场中添加了一种能动张量为 $-\frac{\Lambda}{8\pi} g_{\mu\nu}$ 的“物质”。很容易得出其固有能量密度和压强为

$$\rho = -p = \frac{\Lambda}{8\pi}$$

可以发现这种物质场有正的能量密度，却有负的压强。也就是说，宇宙学常数等价于宇宙中存在一种随时间不变且在空间上均匀分布的这种奇特物质场。下面我们来分析一下这种物质场对于宇宙演化的影响。若我们只考虑宇宙学常数的贡献，将上式代入 Friedmann 第二方程，我们有

$$\ddot{a} = -\frac{4}{3}\pi a(\rho + 3p) = \frac{8}{3}\pi a\rho = \frac{\Lambda a}{3} > 0$$

可以发现，这种物质场的贡献可以使得宇宙加速膨胀。我们将这种能使得宇宙加速膨胀的物质称为 **暗能量**，也就是说宇宙学常数等价于均匀分布的一种暗能量。若宇宙中物质和暗能量的比例适当，使得总体上恰好有 $\rho + 3p = 0$ ，则可以有 $\ddot{a} = 0$ 。在 $\dot{a} = 0$ 的初值条件下即可有 \dot{a} 恒为零，即宇宙既不膨胀也不收缩，处于静态。这种模型称为爱因斯坦静态宇宙。但不久以后就发现，爱因斯坦静态宇宙并不是稳态的，也就是说一点微小的扰动就会让它开始无限膨胀或者收缩，因此这个解并没有实际的物理意义。后来天文观测又发现宇宙的确是在膨胀的，因此爱因斯坦又将宇宙学常数从方程中移除，并说引入宇宙学常数是他的“一辈子最大的错误”。但我们后面会看到，随着人们对宇宙的认识逐渐深入，人们发现我们的宇宙实际上存在一个正的宇宙学常数，因此宇宙学常数又卷土重来了。

9.2.3 尺度因子的演化

有了 Friedmann 方程，我们接下来只需要对其求解，就可以得到尺度因子 a 与 t 的函数关系，从而可以描述宇宙的演化了。但遗憾的是，Friedmann 方程的求解非常困难。因此，我们在此先具体一下求解宇宙中的各个组分对尺度因子的贡献，最后再定性讨论一下真实宇宙的演化。

辐射主导的宇宙

辐射是指光子等静质量为零的相对论性粒子构成的物质场，电磁学的结论告诉我们辐射的密度与压强满足 $\rho = 3p$ 。若我们从 Friedmann 方程中消去 \ddot{a} 和 k ，可以得到

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0$$

将 $\rho = 3p$ 带入上式，可以得到 $\dot{\rho}a + 4\rho\dot{a} = 0$ 。注意到等式左侧乘上 a^3 即为 ρa^4 对于 t 的导数，也就是这意味着

$$\rho \propto a^{-4}$$

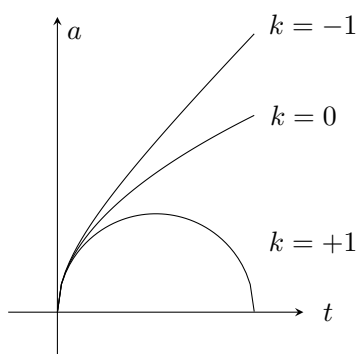
这是因为单位共动体元内粒子数不变，共动体元的体积正比于 a^3 ，同时每个粒子的能量会由于红移而反比于 a 。如果我们设 $C^2 = 8\pi\rho a^4/3$ ($C > 0$)，则 Friedmann 方程可以被写作

$$\dot{a}^2 = C^2 a^{-2} - k$$

我们要求 $t = 0$ 时 $a = 0$ ，以此为初值条件的解为

$$a(t) = \sqrt{2Ct - kt^2}$$

可以看到，这个解与参数 k 有关。下图显示了三种情况下 $a(t)$ 的曲线。可以发现当宇宙为封闭宇宙时，宇宙不仅来自一个奇点，最终也会坍缩为一个奇点，这称为大挤压。而对于开放宇宙，尺度因子 $a(t)$ 会没有上界地增大，即宇宙会无限地膨胀下去。



尘埃主导的宇宙

若宇宙中只存在静质量非零的各种物质，这些物质大部分构成了星系和星际尘埃等，其间的压强可以忽略，因此从能动张量来看这些物质可以看作尘埃。我们称这种宇宙模型为尘埃主导的宇宙。尘埃的压强 $p = 0$ 。用上面同样的方法，我们可以由 Friedmann 方程推出尘埃的密度满足

$$\rho \propto a^{-3}$$

这种情况更加好理解——既然共动体元的物理体积正比于 a^3 ，同时每个粒子的能量不变，那么密度自然就与体积成反比。同样，如果我们设 $C = 8\pi\rho a^3/3$ ，那么 Friedmann 方程可以被写作

$$\dot{a}^2 = Ca^{-1} - k$$

这个方程的解析解较为复杂，但其定性图像与上面尘埃宇宙的结果非常类似。同样，当宇宙为封闭宇宙时，宇宙最终也会发生大挤压。而对于平直和开放宇宙，宇宙会无限地膨胀下去。

真空主导的宇宙

若宇宙中既没有辐射也没有尘埃，此时还可以有的物质场即为暗能量。由于在有宇宙学常数时，暗能量可以看作时空“自带”的能动张量，因此有时也被视作真空的能动张量。由于宇宙学常数随时间是不变的，因此真空的能量密度不会随着 a 变化，或者可以写作

$$\rho \propto a^0$$

宇宙学常数所代表的暗能量满足 $\rho = -p$ 。在 ρ 是常数时求解 Friedmann 第二方程，可以很容易得到

$$a(t) = Ce^{Ht}$$

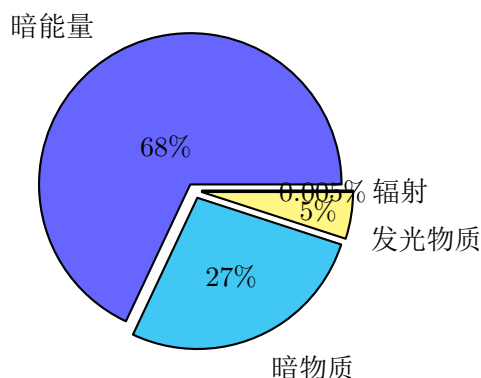
将这个解带入 Friedmann 方程中，可以发现 $k = 0$ 。也就是说，只有宇宙常数的宇宙模型一定是平直宇宙，而且这种宇宙不存在大爆炸奇点。这种模型并不能模拟真实的宇宙，但我们后面会看到它的作用。

这与量子场论中的真空能动张量并不相同。

以上我们讨论了在三种极端情况下宇宙的演化。我们真实的宇宙同时存在这三种，其演化更复杂些，但是我们可以根据宇宙中占主导的物质场来定性估计。根据各个物质场随着 a 的衰减关系可知，宇宙的极早期应该是由辐射主导的，随着其膨胀逐渐变为由尘埃主导。最终辐射和尘埃都会衰减至接近零，若有宇宙学常数的存在，宇宙最终会变为由真空主导。

9.2.4 宇宙学标准模型

爱因斯坦由于对宇宙膨胀的天文观测而放弃了宇宙学常数。但不久后，人们发现宇宙不仅是在膨胀，还是在加速膨胀的。为了建立加速膨胀的宇宙模型，宇宙学常数与暗能量又重新被加入到了方程中。同时，观测还发现宇宙中存在大量无法直接通过光学观测到的物质，但它们却对天体和宇宙的动力学有引力效应，因此被称为 **暗物质**。对于早期宇宙的粒子物理学的理论研究表明，这种暗物质不属于重子物质，而且是非相对论性的，称为 **冷暗物质**，简称 CDM。现代宇宙学的“标准模型”就是结合了暗能量（即宇宙学常数 Λ ）和冷暗物质的模型，称为 **Λ CDM 模型**。在这个模型中，现今宇宙中包括以下内容物。



我们前面说过，现今宇宙的密度参数 $\Omega_0 \approx 1$ ，这是指宇宙的所有内容物造的总密度参数，它们共同使得宇宙是近似平坦的。 Λ CDM 模型的具体参数（如各种物质的具体占比）在宇宙学的其他方向上（如早期宇宙的演化、结构形成等）有着重要的作用，我们不再赘述。

9.3 标准模型的修正

宇宙学标准模型成功地解释了许多宇宙学现象。然而，随着人们对宇宙的认识逐渐深入，人们发现仍然有一些观测到的宇宙学现象在标准模型的框架下无法解释。本节我们来介绍几个最为重要的疑难现象，并且简单地讨论如何对标准模型进行修正以解释这些疑难。

9.3.1 粒子视界疑难与平坦性疑难

我们首先来介绍宇宙学中最重要两个疑难——粒子视界疑难和平坦性疑难，它们共同造就了对标准模型最为重要的一个修正。

粒子视界疑难

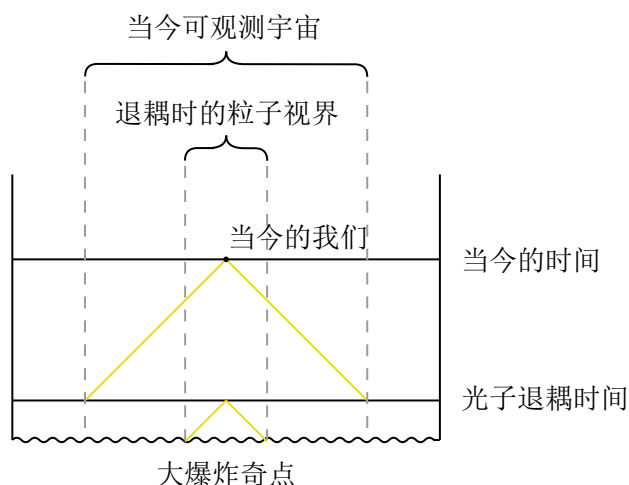
标准模型宇宙学的第一个重要的疑难称为粒子视界疑难。为了介绍这个疑难的具体内容，我们首先需要补充一些宇宙学基本知识。

我们知道，宇宙是由一个奇点膨胀而来的。在宇宙的极早期，其内容物的密度极高，因此粒子之间都很容易碰撞而发生相互作用，光子也不例外。此时，光子只要运动很短一段距离就会与其他粒子碰撞，因此难以在宇宙中穿行，或者说此时宇宙是“不透明的”。随着宇宙体积的逐渐增大，粒子的密度随之降低，物质之间的相互作用也逐渐减弱。在 $t \approx 4 \times 10^5$ 年时，粒子的密度降低到足够低，使得光子发生碰撞的概率已经极小了，此时光子可以几乎完全自由地在宇宙中穿行，或者说宇宙变得“透明”了。这个过程称为光子的 **退耦**。这些在退耦后自由穿行的光子保留到现在，并且随着宇宙的膨胀而不断红移，最终降低到约 3 K 的温度，构成了著名的 **宇宙微波背景辐射 (CMB)**。

天文观测显示，宇宙微波背景辐射在大体上是各向同性的，即在当今可观测宇宙的范围内，各个方向飞来的退耦光子的温度是相似的。这意味着在光子退耦的时候，当今可观测宇宙的范围内的物质场的温度是相同的。然而，在 FLRW 宇宙模型中，从大爆炸到光子退耦的这段时间内，如此大的范围内的物质是无法通过热传递而达到热平衡的。由于热传递的速度也不能超过光速，因此只有退耦时过去光锥内的物质（其边界称为 **粒子视界**）才可能来得及达到热平衡。而由下图可以发现，FLRW 宇宙模型中，退耦时的粒子视界的范围是远小于当今可观测宇宙的，计算可得二者相差了 10^{29} 倍！除非我们认为某种

注意与前面学习到的各种视界不同，粒子视界分隔的是粒子（或者说世界线），而非时空点。

力量使得宇宙在初始条件下就是热平衡的，否则这个现象将难以解释。这称为粒子视界疑难，又称为均匀性疑难。



此图是共形图，其时间轴不是线性的，这样使得我们不把空间的膨胀画出来的情况下，光锥仍然是 45° 的直线。

平坦性疑难

标准模型宇宙学的第二个重要的疑难称为平坦性疑难。我们在上一节讲过，密度参数 Ω 可以用来描述宇宙空间的弯曲，若 $\Omega = 1$ 则宇宙是平坦的。天文观测显示，我们的宇宙的密度参数是接近 1 的，也就是说宇宙空间几乎是平直的。然而这在标准模型中很难做到。为了描述密度参数的“非平坦程度”的演化，我们将 Friedmann 方程化为

$$1 - \frac{1}{\Omega} = \frac{3k}{8\pi\rho a^2}$$

可以看到，密度参数是会随着尺度因子 a 变化的。我们说过，早期宇宙是由辐射和尘埃所主导的。对于辐射有 $\rho \propto a^{-4}$ ，对于尘埃有 $\rho \propto a^{-3}$ 。在这两种情况下，若 $k \neq 0$ ，则等式的右侧分别正比于 a^2 和 a 。也就是说，宇宙的密度参数 Ω 若与 1 有差异，则这个差异会在宇宙膨胀的过程中被急剧地放大。因此，为了是当今宇宙达到我们测得的这种平直程度，早期宇宙的密度参数与 1 的差异只能在负数十个数量级内。除非我们认为宇宙的初值是被“人为”微调过的，或者是纯归为巧合，否则在 FLRW 宇宙模型中将难以解释。这称为平坦性疑难。

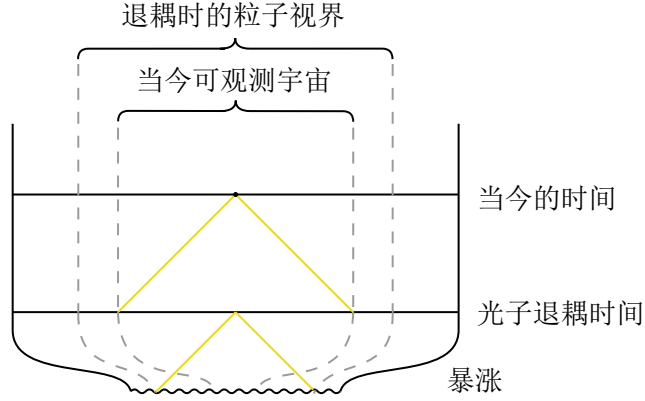
9.3.2 暴涨模型

为了解决粒子视界疑难与平坦性疑难，粒子物理学家 Guth 提出了著名的 **暴涨模型**。暴涨模型在几经修改后，已经被宇宙学家普遍认可，成为了宇宙学中“新标准模型”的一部分。

暴涨理论认为，在宇宙早期有特殊的一个阶段，在这个阶段起主导作用的是一种称为“暴涨场”的物质场，它与真空的性质类似，但处于高能态，具有正的能量和负的压强，且不随着宇宙的膨胀而改变其密度。因此，早期宇宙的演化类似于真空主导的宇宙，其尺度以指数进行膨胀。随着宇宙的膨胀，暴涨场的温度降低，出现了接近零能量的低能态。此时暴涨场将能量传递给其他物质场，而自身进入了低能量的态。于是真空成为了“真的真空”，即不再具有正能量和负压强，因而暴涨结束，宇宙恢复正常的膨胀模式。

暴涨的具体机制在这并不重要，重点是暴涨模型在宇宙演化早期引入了一个短暂的、极快速的膨胀时期。下面我们来看一下暴涨模型是如何成功解释粒子视界疑难和平坦性疑难的。

首先来看粒子视界疑难。在 FLRW 模型中，当今可观测宇宙的范围是远大于光子退耦时的粒子视界的，这与观测不符。但如果在光子退耦之前发生了暴涨，二者的关系就可以反过来。宇宙内容物可以在宇宙尺度非常小时通过热传递达到热平衡，再暴涨到非常大的尺度。这样，暴涨后的粒子视界可以大于当今可观测宇宙，如下图所示。这样，当今可观测宇宙就可以达到热平衡，从而使得宇宙微波背景辐射是各向同性的。



下面我们再来看一看平坦性疑难。我们前面推出了，决定了宇宙曲率的密度参数 Ω 的演化满足

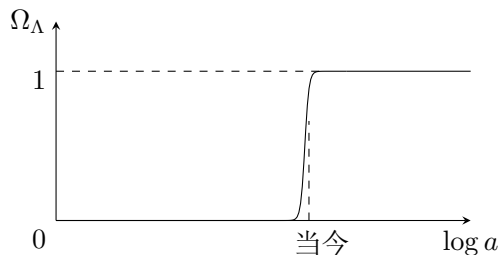
$$1 - \frac{1}{\Omega} \propto \frac{1}{\rho a^2}$$

对于辐射和尘埃， ρ 随着 a 以更高阶减小，导致上式随着 a 增大，进而使得 Ω 与 1 的差异在宇宙膨胀中被放大。然而，若引入了暴涨阶段，暴涨场的能量密度不随膨胀而变化，即 ρ 是一个常数。那么此时上式是正比于 a^{-2} 的。因此，暴涨阶段不但不会放大密度参数 Ω 与 1 的差异，反而会缩小这种差异。换句话说，暴涨过程会“抹平”宇宙。之后的膨胀过程虽然仍然会放大宇宙的不平直性，但由于暴涨过程的膨胀倍数极高，宇宙在暴涨结束时已经变得非常接近平直，因此后期的膨胀需要花费极长的时间才能将宇宙的不平直性重新放大。这样，暴涨理论也就解决了平直性疑难。

9.3.3 巧合性疑难与动力学暗能量

暴涨模型虽然解决了多个非常著名的宇宙学疑难，但还有一些疑难问题无法被暴涨模型解决。其中一个重要的疑难称为 **巧合性疑难**。

我们前面多次讨论了宇宙中的三大类内容物：物质（包括发光物质和暗物质）、辐射和宇宙学常数代表的暗能量。它们随着宇宙膨胀的密度变化是不同的：物质的密度 $\rho_M \propto a^{-3}$ ，辐射的密度 $\rho_R \propto a^{-4}$ ，而宇宙学常数代表的暗能量不随着宇宙膨胀而变化。这意味着随着宇宙的膨胀，物质和辐射的占比会越来越小，而暗能量的占比会越来越大，且暗能量的占比是以 a^3 至 a^4 的速率增大的。若进行计算，可以发现宇宙会在某一很短的时间段内迅速由物质主导变为暗能量主导，如下图所示。



然而，天文观测表明，当今宇宙中物质与暗能量的占比在同一数量级上。这说明我们当今恰好正处在这个极其短暂的转变时期中。这仍然似乎只能诉诸巧合，否则难以解释，因此称为巧合性疑难。

巧合性疑难的解决方式比较简单而直接：我们可以假设宇宙中有不由宇宙学常数代表的一类暗能量，它与其他种类的内容物类似，其密度也会随着宇宙的膨胀而降低。这样，暗能量的占比就不一定随着宇宙的膨胀而如此快地增大了。这种密度可以变化的暗能量我们称为 **动力学暗能量**。动力学暗能量的模型有很多种，其中著名的包括 quintessence 模型等，我们在这里就不具体介绍了。找到暗能量的真实身份不仅是一个宇宙学问题，更是一个基础的粒子物理学问题。目前我们还没有一个很好的答案。

本笔记的参考包括：

1. 梁灿彬《微分几何入门与广义相对论》及梁灿彬公开课视频
2. 陈斌《广义相对论》
3. *Lecture Notes on General Relativity and Spacetime and Geometry: An Introduction to General Relativity* by Sean Carroll
4. *A First Course in General Relativity* by Bernard Schutz
5. *Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity* by Steven Weinberg