

物理学中的“规范”及其历史起源

施郁

(复旦大学物理学系, 上海 200433)

摘要: 本文对物理学中的规范理论作了通俗介绍, 然后讨论物理学中“规范”一词的起源. 规范本来是尺子或者测度标准的意思. 经典电磁学中使用“规范”一词是在外尔 1928 ~ 1929 年提出电磁场中量子带电粒子的规范原理之后. 如果不是因为外尔曾经于 1918 ~ 1919 年作过失败的尝试, 试图从时空中平行移动导致的尺度变化来导出电磁矢量势, 那么后来人们就不会将“规范”一词用于电磁矢量势. 本文也梳理薛定谔、福克和伦敦的贡献.

关键词: 规范; 电磁学; 量子力学; 杨-米尔斯理论

中图分类号: O 4 **文献标识码:** A **文章编号:** 1000-0712(2017)09-0001-05

【DOI】10.16854/j.cnki.1000-0712.2017.09.001

1 经典电磁学中的“规范”

从目前物理学的学习过程来说, “规范”一词首先在经典电磁学或者经典电动力学中出现.

电场强度 E 和磁感应强度 B 都可以用另外两个物理量, 即矢量势 A 和标量势 ϕ 表达:

$$E = - (1/c) \partial A / \partial t - \text{grad} \phi \quad (1)$$

$$B = \text{curl} A \quad (2)$$

这里 c 是光速, $\partial A / \partial t$ 是 A 对时间的导数, 即随时间的变化率. grad 和 curl 分别代表梯度和旋度.

现在介绍上面出现的数学符号. 黑体符号代表矢量, 如 E, B, A , 是指一个既有方向, 又有大小的物理量. 非黑体符号代表标量, 如 ϕ , 只有大小, 没有方向. 这些物理量都是空间和时间的函数. 标量函数的梯度 (grad) 是一个矢量, 它在任意方向的投影就是标量函数在这个方向的变化率. 矢量函数的旋度 (curl) 也是一个矢量, 它在任意方向的投影是矢量函数沿着围绕这个方向的一个无穷小闭合路径的积分, 再除以所包围的小面积.

矢量势 A 有 3 个空间分量, 可以和标量势 ϕ 统一看成一个 4 维矢量势, 有 4 个时空分量, 写作

$$(A^0, A^1, A^2, A^3)$$

其中 $A^0 = \phi$ 是时间分量. 这类似于时间和空间坐标可以写作 4 维时空坐标

$$(x^0, x^1, x^2, x^3)$$

其中 $x^0 = ct$.

在经典电磁学里 A 不是可直接观测量, ϕ 只有相对差可以测量, 它们都有选择自由度. 可以看出, 如果由任意标量函数 χ 定义:

$$A' = A + \text{grad} \chi \quad (3)$$

$$\phi' = \phi - (1/c) \partial \chi / \partial t \quad (4)$$

那么用 A 和 ϕ 取代式 (1) 和式 (2) 中的 A 和 ϕ , 直接可观测测量电场强度 E 和磁感应强度 B 保持不变.

等式 (3) 和式 (4) 被称作规范变换 (gauge transformation). 在此规范变换下, E 和 B 的不变性叫作规范不变 (gauge invariance), 也就是说在规范变换下的不变性. 这两个名词是物理学的标准名词, 被普遍使用. “规范变换”一词甚至被借用到广义相对论中.

1851 年, 汤姆孙 (William Thomson), 即开尔文勋爵 (Lord Kelvin) 首先通过式 (2) 定义了矢量势 A . 1856 年, 麦克斯韦 (James Clerk Maxwell) 提出, 法拉第 (Michael Faraday) 发现的感生电场由式 (1) 的第 1 项给出, 而且确认 A 就是法拉第所说的电紧张密度 (electrotonic density)^[1]. 1880 年代, 亥维赛 (Oliver Heaviside) 和赫兹 (Heinrich Hertz) 发现, 麦克斯韦方程可以不涉及 A , 完全由电场强度和磁场强度表达, 这就是我们目前所用的麦克斯韦方程^[1].

收稿日期: 2017-04-29

基金项目: 国家自然科学基金面上项目 (11374060) 资助

作者简介: 施郁 (1968—), 江苏扬州人, 复旦大学物理学系教授, 主要从事理论物理和物理学史的研究工作.

2 量子力学中的规范

简单地说,量子力学是描述微观粒子的一套运算规则.在量子力学中,中心概念是量子态.描述粒子在空间中的运动的量子态通常用复数波函数 ψ 表示,它是空间和时间的函数,它的模的平方代表概率密度.而可观测物理量,比如动量、角动量,都表示成作用在波函数上的算符.比如动量算符 p 等于 $-i\hbar \text{grad}$,能量算符 H 等于 $i\hbar \partial/\partial t$,其中 $\hbar = h/2\pi$, h 是普朗克(Max Planck)常量.与 A 和 ϕ 类似 p 和 H/c 也可以统一看成一个4维动量算符,其中 p 是空间分量, H/c 是时间分量.

对于与电磁场耦合的电荷为 q 的粒子,动量算符 p 和能量算符 H 作如下改变:

$$p \rightarrow p - qA/c \quad (5)$$

$$H \rightarrow H + q\phi \quad (6)$$

甚至可以用 gauge(规范化)作为动词指称这两个改变.

在量子力学中,非相对论性粒子的波函数服从薛定谔方程,相对论性粒子服从克莱因-戈登(Klein-Gordon)方程或者狄拉克方程.在作变换式(3)和式(4)的时候,波函数也要相应地由原来的 ψ 做一个相位变换

$$\psi' = \exp(iq\chi/\hbar c)\psi \quad (7)$$

从而波函数服从的运动方程保持不变,这就是规范不变性(gauge invariance),即规范变换下的不变性. χ 是时空坐标的函数,因此规范变换是定域的.如果 χ 是常数,那么相位变换式(7)下的不变性是整体对称.量子力学与相对论相结合后,或者在多粒子系统的量子场论中,上面的波函数 ψ 要理解为量子场.

从式(5)和式(6)可以得到,电磁场给其中的带电粒子的波函数带来一个由电磁势沿路径的积分所给出的相位因子,即

$$\psi \rightarrow \exp(iqS/\hbar c)\psi \quad (8)$$

其中

$$S = \int A^1 dx^1 + \int A^2 dx^2 + \int A^3 dx^3 - c \int \phi dt$$

就是电磁势沿路径的积分.1959年,阿哈罗诺夫(Yakir Aharonov)和玻姆(David Bohm)发现这个相位因子有观测效应,即使在场强 E 和 B 消失的区域^[2].这确立了 A 和 ϕ 的物理实在性.

量子理论框架下的规范原理是外尔(Hermann Weyl)在1928~1929年确立的^[3-5].规范原理使得作为电磁场源的电荷的守恒成为规范不变性的后

果,而且时空中每个点上都可以有定域规范变换.如果量子粒子不与电磁场耦合,那么就没有规范不变性,也就是说,与电磁场的耦合保证了规范不变性.为了有规范不变性,电磁场必须存在.因此人们说,电磁场是规范场.

规范场本身的量子化导致它对应的规范粒子,后者是前者的量子.电磁场的量子是光子,没有质量.

3 杨-米尔斯理论

1950年代,各种奇异粒子的发现层出不穷,如何确定它们的相互作用成为一个重要问题.在此物理驱动下,1954年,杨振宁和米尔斯(Robert Mills)将外尔关于电磁场的规范理论推广为非阿贝尔(Non-Abelian)规范理论,也叫杨-米尔斯理论^[6,7].杨振宁和米尔斯强调,作为时空坐标的函数,规范变换是定域的.

在杨-米尔斯理论的最初形式中,波函数 ψ 被推广为有两个分量的波函数.基本粒子的量子态是两种基本内部状态的叠加状态,这两个分量波函数就是叠加系数.数学上可以将这两个分量波函数一起写成一个2行1列的矩阵.相应的规范变换(准确来说,应该叫相位变化)被推广为一个2行2列的矩阵变换,也就是说这个2行2列的矩阵乘以原来的二分量波函数,得到变换以后的新波函数.

做一个类比.一个平面上从坐标原点出发的矢量可以用两个坐标表示,这两个坐标可以写成一个2行1列的坐标矩阵.矢量绕着原点的转动就可以用一个2行2列的变换矩阵来表示.一个2行2列的变换矩阵乘以一个2行1列的坐标矩阵,就给出一个2行1列的新坐标矩阵,它的第一行等于变换矩阵第一行两个数与坐标矩阵的两个数分别相乘再相加,它的第二行等于变换矩阵第二行两个数与坐标矩阵的两个数分别相乘再相加.

连续作两次变换就是两个变换矩阵相乘.一般来说,矩阵相乘与顺序有关.数学上,与顺序无关的情况叫作阿贝尔(Abelian),与顺序有关的情况叫作非阿贝尔.所以杨-米尔斯理论又叫非阿贝尔规范理论.

在杨-米尔斯理论中,对于二分量波函数,引进与某个规范场的耦合,从而将规范场从电磁场推广到非阿贝尔规范场,保证定域规范不变性以及某种荷的守恒.

粒子与规范场的耦合与式(5)和式(6)类似.但

是,因为粒子波函数是一个2行1列的矩阵,规范势的每个时空分量是一个2行2列的矩阵.杨振宁和米尔斯还给出了由规范势决定规范场场强的公式,以及粒子波函数与规范场的运动方程.

在电磁规范理论中,运动方程和电磁场场强都是规范不变的.而在杨-米尔斯理论中,规范场场强的每个时空分量都是一个2行2列的矩阵,它们在规范变换下是有变化的,不是规范不变的,不过,变换与波函数的变换相对应,有规范协变性(gauge covariance).但是运动方程是规范不变的.非阿贝尔规范势的每个时空分量由内部状态空间中若干独立分量组合而成.独立分量的个数由规范对称性的数学性质决定.这导致规范粒子之间也有相互作用,这是与电磁场的一大区别.

1954年的时候,具体来说,杨振宁和米尔斯把定域规范场的思想用在强相互作用关于质子与中子的对称性,即同位旋守恒,就是把质子和中子当作同种粒子的两个内部态.当时对于这种情况下规范粒子质量问题没有很好的结论,而且后来人们了解到同位旋守恒只是近似的.

但是杨-米尔斯规范理论为确定基本粒子的相互作用提供了一个基本原理^[8-10].本来,基本粒子之间的相互作用形式无法确定,杨-米尔斯规范理论规定了相互作用必须是什么样的.

后来杨-米尔斯规范理论成为描述弱电相互作用和强相互作用的基本理论框架,分别借助自发对称破缺和渐进自由,导致粒子物理的标准模型.

温伯格(Steven Weinberg)、格拉肖(Sheldon Glashow)和萨拉姆(Abdus Salam)等人各自在1960年代的工作最终确立了弱电理论.弱电理论的最后形式建立在杨-米尔斯规范理论和自发对称破缺机制的基础上.其中有两种规范场,一个是2行2列矩阵,在内部空间中有3个独立分量 W^1 、 W^2 、 W^3 ,另一个规范场像电磁场那样在内部空间只有1个分量 B . W^1 和 W^2 的线性组合给出 W^+ 和 W^- , W^3 和 B 的线性组合给出 Z^0 和电磁场. W^+ 、 W^- 和 Z^0 的规范粒子质量通过自发对称破缺获得,即所谓的安德森-布劳特-恩格莱特-希格斯(Anderson-Brout-Englert-Higgs)机制.电磁场的规范粒子光子无质量.1983年, W^+ 、 W^- 和 Z^0 在CERN被鲁比亚(Carlo Rubbia)等人观测到.

描述夸克之间强相互作用的杨-米尔斯规范理论被创始人之一盖尔曼(Murray Gell-Mann)称作量子色动力学,其中的规范场是胶子场,与夸克的色自

由度耦合.色与弱电理论无关.夸克还有个自由度叫作味,与弱电理论相关,而与色动力学无关.每种味的夸克都有3种色,因此色波函数是3行1列的矩阵.胶子场是3行3列的矩阵,在色空间中有8个独立分量(可以说是8种胶子).带色的规范粒子胶子确实如杨-米尔斯理论原本所述,没有质量,但是通常与夸克一起被禁闭在强子中而不能被孤立出来.

4 为什么要用“规范”这个词

规范变换式(7)实际上是个相位变换,规范场是相位场.那么为什么要用“规范”这个词呢?是因为原先在经典电磁学里,式(3)和式(4)就被称作规范变换吗?

与直觉的推测相反,答案是否定的.

事实上,这是因为外尔1928~1929年的理论是他1918~1919年理论的修正,而后者所讨论的是名副其实的规范不变性.

1918~1919年,受爱因斯坦广义相对论的鼓舞,外尔试图用几何的方法导出电磁场,以便与引力统一.考虑在时空中的平行移动,广义相对论说时空的弯曲导致矢量方向有变化,而外尔猜想电磁势导致沿路径的积分给出一个时空变化因子.但是与后来1928~1929年理论不同,这个因子的指数上没有虚数单位 i ,因此不是相位因子,而是标度因子,确实是所谓的“规范”因子.“规范”本是尺子或者测度标准的意思.外尔1918~1919年的理论遭到了爱因斯坦的反对.

外尔1918~1919年的理论建立在3篇论文的基础上^[11-13].杨振宁注意到^[14],外尔在前两篇论文中,用的名词是 *masstab invarianz*,翻成英文是 *measure invariance*(测定不变性),在第三篇论文中,他用的名词是 *eich invarianz*,而 *eich invarianz* 在1921年被翻译为英文 *calibration invariance*.杰克孙(John David Jackson)和奥肯(Lev B. Okun)注意到^[15],英文 *gauge invariance* 后来首先出现在外尔本人1929年的英文文章中^[4].*Eich invarianz*, *calibration invariance* 或者 *gauge invariance* 就是中文的“规范不变性”.

1922年,薛定谔猜测可以在外尔的规范因子的指数里加上虚数单位 i ^[16].

1922年,卡鲁扎(Th. Kaluza)提出5维时空理论,第5维与4维时空之间的度规系数由电磁势给出.

1926年,薛定谔的4篇系列文章创立了波动力

学(与矩阵力学同为当时量子力学的两种形式)这4篇文章分别于1月27日、2月23日、5月10日、6月23日被杂志社收到^[17-20]。薛定谔的这4篇论文中都没有提到他1922年修改外尔理论的工作。但是在第4篇论文中,薛定谔指出在电磁场中,带电粒子的动量和能量算符必须如我们上面的式(5)和式(6)^[20]。

也是1926年,在薛定谔的波动力学工作带领下,克莱因(Oskar Klein)和福克(Vladimir Fock)分别独立讨论了卡鲁扎理论框架下的波动力学。福克的论文7月24日被杂志社收到,文中在得到波函数运动方程后,又指出在变换式(3)、(4)、(7)下的不变性,也就是我们现在所说的规范不变性^[21]。

当年12月10日左右,伦敦(Fritz London)写了一封幽默的信给薛定谔,提到后者1922年修改外尔规范因子的工作,并敦促他阐明与波动力学的联系^[22-23]。

1969年拉曼(V. V. Raman)和福曼(P. Forman)挖掘出这封伦敦致薛定谔的信后,人们发现,薛定谔1922年的工作对他创立波动力学确实起了作用,他1925年11月致爱因斯坦的一封信就提到德布罗意理论与外尔理论的联系^[23-24]。

伦敦在1926年12月给薛定谔写了信后,自己写了两篇文章,将波动力学与外尔1918~1919年的规范理论联系起来,提出电磁场带来波函数的相位因子,即我们上面的式(8)^[25-26]。

终于,外尔在他1928年的书《量子力学中的群论》^[3]和1929年的两篇文章中^[4-5],修改了他1918~1919年的理论,将标度因子改为相位因子,但是却沿用了原来的名词“eich(规范)”。1929年两篇文章在他访问美国普林斯顿时完成。第一篇是英文的,其中出现了英文名词 principle of gauge invariance。第二篇是第一篇的扩充,是德文的,其中还包含其他量子场论基本问题,比如二分量旋量理论。

外尔提出规范不变原理,将电磁场与物质(而非原先所说的电磁场与引力)联系在一起,为了规范不变原理的成立,电磁场作为物质波的必要伴侣而被推导出来,电荷守恒是规范不变性的后果。

泡利(Wolfgang Pauli)在他关于场论的著名综述文章中,介绍了外尔规范理论^[27-28]。学生时代的杨振宁正是通过泡利的综述对规范原理留下了深刻印象^[8]。

5 经典电磁学中“规范”一词的来源

1977年,杨振宁考证^[14]经典电磁学中的 E 和 B 在变换式(3)和式(4)下保持不变的性质原先似乎并没有一个专有名词,比如著名的Fopple - Abraham - Becker - Sauter电磁学教材有很多版,而直到1964年英文版,“gauge(规范)”一词才出现在这部经典电磁学教科书中。

我们核查了很多早期的经典电磁学的书籍,其中有变换式(3)和式(4),但是确实没有使用专有名词。

我们核查了很多早期的经典电磁学的书籍,其中有变换式(3)和式(4),但是确实没有使用专有名词。

在杨振宁的1977年文章之后,派斯(Abraham Pais)1986年明确断言^[29]，“规范”一词就是从外尔1919年的文章开始使用的。但是派斯没有给出这个断言的理由。我们猜测,他很可能就是从杨振宁文章中得到的信息。

我们认为,eich、gauge或者规范本来是尺子或者测度标准的意思,如果不是因为外尔提出平行移动时的尺度变换问题,没有道理用它来指电磁场的变换性质,所以,“规范”(以德文的形式)应该是从外尔1918~1919年的论文才开始用这个名词来表示电磁场的性质。而该理论当时被爱因斯坦反对,没有被人们接受,所以可以理解,“规范”当时也没有被他人用于经典电磁学。而外尔1928~1929年用量子力学修正的规范理论因为其正确性,被人们接受了,结果“规范”一词也走进了经典电磁学。

致谢:感谢杨振宁先生几年前告知笔者“规范”一词起源问题,以及相关的讨论。

参考文献:

- [1] Yang C N. The conceptual origins of Maxwell's equations and gauge theory [J]. Physics Today, 2014, 67 (11): 45-51.
- [2] Aharonov Y, Bohm D. Significance of electromagnetic potentials in the quantum theory [J]. Phys Rev, 1959, 115: 485-491.
- [3] Weyl H. Gruppentheorie und Quantenmechanik [M]. S. Hirzel, Leipzig, 1928.
- [4] Weyl H. Gravitation and the electron [J]. Proc Natl Acad Sci U S A, 1929, 15: 323-334.
- [5] Weyl H. Elektron und gravitation. I [J]. Z Phys, 1929, 56: 330-352.
- [6] Yang C N, Mills R. Isotopic spin conservation and a generalized gauge invariance [J]. Phys Rev, 1954, 95: 631-631.
- [7] Yang C N, Mills R. Conservation of isotopic spin and isotopic gauge invariance [J]. Phys Rev, 1954, 96: 191-195.

[8] Yang C N. Selected Papers 1945—1980 With Commentary

- [M]. W. H. Freeman and Company Publishers 1983.
- [9] 施郁. 物理学之美:杨振宁的13项重要科学贡献[J]. 物理, 2014, 43(1): 57-62.
- [10] Shi Y. Beauty and Physics: 13 important contributions of Chen Ning Yang [J]. Int J Mod Phys A, 2014, 29(17):1475001.
- [11] Weyl H. Gravitation and electricity [J]. Sitzber Preuss Akad Wiss, 1918: 465-480.
- [12] Weyl H. Pure infinitesimal geometry [J]. Math Z, 1918 2: 384-411.
- [13] Weyl H. A new expansion of the relativity theory [J]. Ann Phys, 1919 59:101-133.
- [14] Yang C N. Magnetic monopoles, fiber bundles, and gauge fields [J]. Annals of New York Academy of Sciences, 1977, 294(8): 86-97.
- [15] Jackson J D, Okun L B. Historical roots of gauge invariance [J]. Rev Mod Phys, 2001 73:663-680.
- [16] Schrödinger E. A new remarkable characteristic of the quantum channel of a single electron [J]. Z Phys, 1922, 12: 13-23.
- [17] Schrödinger E. Quantisation as an eigen value problem [J]. Ann Phys, 1926 79: 361-408.
- [18] Schrödinger E. Quantification of the eigen value problem [J]. Ann Phys, 1926 79: 489-527.
- [19] Schrödinger E. Quantification of the eigen value problem [J]. Ann Phys, 1926 80:437-490.
- [20] Schrödinger E. Quantization as a problem of eigenvalue [J]. Ann Phys, 1926 81:109-139.
- [21] Fock V. Invariable forms of waves and the movement equations for a loaded mass point [J]. Z Phys, 1926, 39: 226-232.
- [22] Raman V V, Forman P. Hist. Studies Phys. Sci., 1969, 1: 291-314.
- [23] Yang C N. Square root of minus one, complex phases and Erwin Schrödinger. In Kilmister C W. (ed.) Schrödinger Centenary Celebration of a Polymath [M]. Cambridge University Press: 1987.
- [24] Hanle P. Schrödinger-Einstein correspondence and the sources of wave mechanics [J]. Am J Phys, 1979 47: 644-648.
- [25] London F. Weyl's theory and quantum mechanics [J]. Naturwissenschaften, 1927, 15: 187-187.
- [26] London F. Quantum - mechanical interpretation of Weyl's theory [J]. Z Phys, 1927, 42: 375-389.
- [27] Pauli W. Handbuch der Physik, 2nd Ed., 1933, 24(1): 83.
- [28] Pauli W. Relativistic Field Theories of Elementary Particles [J]. Rev Mod Phys, 1941, 13: 203-232.
- [29] Pais A. Inward Bound [M]. Oxford University Press: 1986.

“Gauge” in physics and its historic origin

SHI Yu

(Department of Physics, Fudan University, Shanghai 200433, China)

Abstract: We make a popular introduction to the gauge theories in physics, and then we discuss the origin of the word “gauge” used in physics. Originally, gauge means ruler or calibration standard. The use of “gauge” in classical electromagnetism started after Hermann Weyl had presented the gauge theory of a quantum charged particle coupled with an electromagnetic field in 1928 ~ 1929. If Weyl had not made a failed attempt in 1918 ~ 1919, trying to deduce electromagnetic vector potential through the scale change induced by parallel transport in spacetime, there would not have been the use of the word “gauge” for electromagnetic vector potential. We also review the contributions of Schrödinger, Fock and London.

Key words: gauge; electromagnetism; quantum mechanics; Yang - Mills theory

大学物理

<http://dxwl.bnu.edu.cn>