Gauge conditions and gauge transformations

L. V. Prokhorov
Institute of Physics, St. Petersburg State University, St. Petersburg

Fiz. Élem. Chastits At. Yadra **27**, 1399–1468 (September–October 1996)

The fundamental problems associated with elimination of the arbitrariness in gauge theories are the subject of this review. The role of the Dirac-extended gauge group and the importance of the requirement of only weak gauge invariance of physical quantities are stressed. It is shown that it is important to distinguish between dynamical and nondynamical gauges in quantum theories. The most popular gauges are analyzed, mainly for the example of electrodynamics, from the viewpoint of the role played by the physical degrees of freedom. The key role of the Fock gauge is studied. Very simple quantum-mechanical models illustrating some statements about systems of fields are studied in detail. The various types of gauge transformation and their physical interpretation are discussed, and the essential difference between local and global transformations is emphasized. [S1063-7796(96)00405-6]

1. INTRODUCTION

1.1. Gauge invariance

All the known field interactions—strong, electromagnetic, weak, and gravitational—possess the property of local gauge invariance. The principle of gauge invariance therefore is a fundamental principle of physics. It corresponds to the statement that the fundamental fields appearing in the Lagrangian admit transformations by arbitrary functions of the coordinates and time x which do not change the Lagrangian (the action). In the case of the electromagnetic field, described by the vector potential $A_\mu(x)$ (Ref. 1), these are the well known gradient transformations[1)]

$$A_\mu \to A_\mu + \partial_\mu \Lambda, \quad \psi \to e^{ie\Lambda}\psi \quad \left(\partial_\mu = \frac{\partial}{\partial x^\mu}\right), \tag{1.1}$$

where $\Lambda = \Lambda(x)$ is an arbitrary scalar function, $\psi = \psi(x)$ is a charged field, and e is the electric charge. In the case of Yang–Mills fields,[3–5] which are described by vector potentials A_μ taking values in the Lie algebra of some semisimple group G, these transformations have the form

$$A_\mu \to U A_\mu U^{-1} - \frac{1}{ig} U \partial_\mu U^{-1}, \quad \psi \to U\psi, \tag{1.2}$$

where $U = U(x)$ is an element of the non-Abelian gauge group and ψ is a "matter field." The gluon and quark fields transform in the same way [$G = SU(3)_c$, the strong interaction[6–8]], as do the bosonic and fermionic fields in the initial Lagrangian of electroweak interactions[9,10] [$G = SU(2) \times U(1)$]. In the theory of gravity[11,12] these are generally covariant transformations of the coordinates x^μ,

$$x' = x'(x), \tag{1.3}$$

and the corresponding transformations of the tensors $T^{\mu\cdots}$,

$$T'^{\mu\cdots} = \Lambda^\mu_\nu(x) \ldots T^{\nu\cdots}(x), \quad \Lambda^\mu_\nu(x) = \frac{\partial x'^\mu}{\partial x^\nu}. \tag{1.4}$$

The action of a relativistic particle[13,14]

$$S = -m \int \sqrt{\dot{x}^2}\, d\tau, \quad \dot{x}^\mu = dx^\mu/d\tau \tag{1.5}$$

[τ is the "invariant time," and the gauge transformation is $\tau \to \tau' = \tau'(\tau)$, $x'(\tau') = x(\tau)$] and the action of the Nambu–Goto string[13–19] and the superstring[13–16] also possess the property of gauge invariance in the above sense. For example, the Nambu–Goto action[17–19]

$$S = -\gamma \int \sqrt{(\dot{x}x')^2 - \dot{x}^2 x'^2}\, d\tau d\sigma, \quad \dot{x}^\mu = \frac{dx^\mu}{d\tau},$$

$$x'^\mu = \frac{dx^\mu}{d\sigma}, \quad \gamma = \text{const} \tag{1.6}$$

is invariant under reparametrization transformations

$$\tau \to f_1(\tau,\sigma), \quad \sigma \to f_2(\tau,\sigma), \tag{1.7}$$

because it is proportional to the area of the two-dimensional surface swept out by the string in spacetime.[2)]

For all interactions except the gravitational interaction, the dynamical variables are the vector potentials. The vector potential is treated as a connection in the principal bundle space[21,22] (the base is Minkowski space, and a fiber is the gauge group), and it transforms inhomogeneously under gauge transformations. In the case of gravity, the components of the metric tensor $g_{\mu\nu}(x)$ play the role of the dynamical variables. The connection in gravitational theory is compatible with a metric (the Christoffel symbols are expressed in terms of $g_{\mu\nu}$), so that the metric tensor itself plays a fundamental role. Incidentally (we shall not need to return to this point), we note that connections are often treated as unphysical objects, owing to their gauge noninvariance (in contrast to tensors, which transform homogeneously). This, of course, is a misunderstanding. Any quantity which changes (homogeneously or inhomogeneously) under gauge transformations contains unphysical components. The presence of the latter is what makes the transformation law nontrivial, i.e., makes the connection or tensors noninvariant. This is why the problem of unphysical degrees of freedom also exists in the theory of gravity.

1.2. Gauge conditions

Therefore, gauge theories contain unphysical degrees of freedom which ensure the relativistic invariance or the general covariance of the formalism. The law for their variation with time is not fixed by the equations of motion, which follow from the Hamilton variational principle, i.e., the solutions of the equations of motion contain a functional arbitrariness. At first glance, in classical physics there are at least three possible ways of dealing with the gauge variables: (1) fixing them; (2) reformulating the theory in terms of only invariant variables; (3) eliminating them from the Lagrangian, for example, by setting them equal to zero. It turns out that these possibilities are not equivalent. The third one radically changes the dynamics of the physical degrees of freedom (see Sec. 3), and the second one even in comparatively simple models can make the calculations unimaginably more complicated.[23] Therefore, only the first possibility is used in practice. The Lagrangian or the action do not explicitly indicate any preference for a time dependence of the unphysical degrees of freedom, so that the method of fixing them depends entirely on the whim of the author. Clearly, here there is wide scope for creative activity. Since the creation of electrodynamics and the theory of gravitation, many different gauge conditions (gauges) have been proposed to eliminate the arbitrariness. Even more appeared when quantum theory was developed, particularly in connection with the study of Yang–Mills fields.[3–5] Let us list some of the most useful ones.[3)]

1. The Lorenz gauge:[24]

$$\partial_\mu A_\mu = 0. \tag{1.8}$$

This is sometimes referred to as the Landau gauge.

2. The Coulomb or radiation gauge:

$$\text{div } =\mathbf{A}0. \tag{1.9}$$

This was first used by Maxwell,[1] and it would be appropriate to call it the Maxwell gauge.

3. The Weyl gauge:[25]

$$A_0 = 0. \tag{1.10}$$

This is the natural gauge following from the structure of the Lagrangian of the electromagnetic field (see Sec. 2). It is also (inappropriately) called the Hamiltonian gauge. In the English literature one finds the term temporal gauge. Jackiw has suggested naming the gauge (1.10) after Weyl.[26]

4. The Arnowitt–Fickler[27] or axial gauge:

$$A_3 = 0. \tag{1.11}$$

5. The light-cone gauge:

$$nA = 0, \quad n^2 = 0 \quad (nA \equiv n_\mu A_\mu \equiv g^{\mu\nu} n_\mu A_\nu). \tag{1.12}$$

The gauges (1.10)–(1.12) can be written as

$$nA = 0. \tag{1.13}$$

For $n^2 = n_0^2 - \mathbf{n}^2 = 1$ we have the Weyl gauge, and for $n^2 = -1$ we have the Arnowitt–Fickler gauge. Any gauge of the form (1.13) is sometimes referred to as an axial gauge.

6. The Fock gauge:[28–30]

$$x_\mu A_\mu(x) = 0, \tag{1.14}$$

or

$$(x - x_0)_\mu A_\mu(x) = 0. \tag{1.15}$$

This gauge is sometimes called the Fock–Schwinger gauge[31] (Schwinger was the first to notice its advantages).

Gauge conditions usually supplement the equations of motion, i.e., the law governing the variation of the unphysical variables is specified directly. However, it can be fixed in a different way, by changing the form of the original Lagrangian, i.e., by adding to it a term which spoils the gauge invariance and allows equations of motion to be written down for the unphysical variables. The following gauge-fixing terms $\mathscr{L}'$ are most commonly added to a Lagrangian $\mathscr{L}$:

7. The Heisenberg–Pauli gauge:[32,33]

$$\mathscr{L}' = -\frac{1}{2\alpha} (\partial_\mu A_\mu)^2, \tag{1.16}$$

where α is an arbitrary parameter. The case $\alpha = 1$ corresponds to the Feynman gauge, and the case $\alpha = 0$ to the Lorentz gauge. Gauges of the form (1.16) are sometimes referred to as gauges of the Fermi class,[34–37] although in his early studies Fermi did not explicitly add gauge-fixing terms to the Lagrangian.

8. The 't Hooft gauge:[38]

$$\mathscr{L}' = -\frac{1}{2\alpha} (\partial_\mu A_\mu - \alpha m \text{ Im } \varphi)^2, \tag{1.17}$$

where φ is a complex scalar field. The additional term $\mathscr{L}'$ is chosen from the condition that mixed terms $A_\mu \partial_\mu \varphi$ cancel in the effective Lagrangians describing the Higgs effect (Refs. 9, 10, and 39; see also Ref. 40). This is not the only gauge proposed by 't Hooft. The following one is useful when studying theories with vacuum rearrangement.[41] If φ is the Higgs field and $\chi = \arg \varphi$, we require that $\chi + \beta^{-2} \partial_\mu A_\mu = 0$, or

$$\mathscr{L}' = 2ea^2 \beta^2 \cos(\chi + \beta^{-2} \partial_\mu A_\mu). \tag{1.18}$$

Here β is a constant with the dimensions of mass, $a = \langle \varphi \rangle_0$, and e is the electric charge. In this gauge the fictitious fields separate out. For $\beta \to \infty$ the unitary gauge is obtained (unphysical fields do not propagate). Let us recall the Abelian gauge.[41,42] Its name comes from the fact that an Abelian subgroup of a non-Abelian group is singled out. For example, for the group $SU(2)$ the Abelian subgroup is $U(1)$, with the corresponding decomposition of the vector field $W_\mu^a \to (A_\mu, W_\mu^\pm)$, $W_\mu^3 = A_\mu$, $W_\mu^\pm = (W_\mu^1 \mp i W_\mu^2)/\sqrt{2}$, $a = 1,2,3$. The gauge-fixing term is chosen to be

$$\mathscr{L}' = -\frac{1}{2\alpha} (\partial_\mu A_\mu)^2 - \frac{1}{\beta} (D_\mu W_\mu^+)(D_\mu W_\mu^+)^*,$$

$$D_\mu = \partial_\mu - igA_\mu. \tag{1.19}$$

The first term in (1.19) violates the invariance under the Abelian subgroup; $\mathscr{L}'$ even violates the global $SU(2)$ symmetry, so that the renormalization constants of the fields A_μ and $W_\mu^\pm$ in this gauge turn out to be different.[42]

9. The 't Hooft–Veltman gauge:[43]

$$\mathscr{L}' = -\frac{1}{2} \left(\partial_\mu A_\mu - \frac{1}{2} \alpha e A_\mu^2 \right)^2 \tag{1.20}$$

(α is a parameter), which reveals the great breadth of possibilities offered by gauge freedom even in electrodynamics. In this gauge one can find all the features of non-Abelian theory: auxiliary anticommuting fields and nontrivial self-interaction of the electromagnetic field (3- and 4-photon vertices). Nevertheless, the theory is equivalent to the standard one. This has been checked in Ref. 43 in lowest-order perturbation theory.

10. The planar gauge:[31,44–46]

$$\mathscr{L}' = \frac{1}{2\alpha n^2} nA\Box nA, \quad \Box = -\partial_\mu^2, \quad n^2 \neq 0. \tag{1.21}$$

This is one of several gauges suggested by Kummer in Ref. 44. Among them, in addition to (1.8)–(1.13) listed above, is the gauge

$$\mathscr{L}' = -\frac{1}{2\alpha}[(n\partial)nA]^2. \tag{1.22}$$

The planar gauge of Ref. 46 [$n^2 \neq 0$, $\alpha \to 1$ in (1.21)] was a development of the gauge condition used by Lipatov[45] in studying deep-inelastic scattering processes [$n^2 \to 0$ in (1.21); see Sec. 4].

11. The background-field gauge[47,48] (see also the excellent review of Ref. 49):

$$\mathscr{L}' = -\frac{1}{2\alpha}(\partial_\mu q_\mu - igA_\mu q_\mu)^2. \tag{1.23}$$

Here A_μ and q_μ are the "background" (classical) and quantum (auxiliary) fields in the decomposition of the quantized field:

$$A_\mu^q = q_\mu + A_\mu. \tag{1.24}$$

This decomposition is used to study effective Lagrangians.[50–53] The essential feature is the postulate that there is a homogeneous transformation law for q_μ and an inhomogeneous one for A_μ, i.e., under gauge transformations the field q_μ transforms as a tensor and the field A_μ as a connection. The decomposition (1.24) is allowed because the sum of a tensor and a connection transforms as a connection. This trick allows the quantum theory of gauge fields to be constructed without the loss of explicit gauge invariance, in particular, it leads to a convenient computational scheme for gauge-invariant effective Lagrangians.[49] It is also effective in theories of gravitation[54] and supergravity.[55]

In principle, it is possible to construct new gauge conditions by combining known ones. The gauges obtained in this way are usually parameter-dependent. They can prove useful in particular problems. Let us give some examples of such parametric or interpolating gauges.

12. The class of "flow gauges" (Ref. 56):

$$A_0 = Z[\mathbf{A}] \tag{1.25}$$

where $Z[\mathbf{A}]$ is a functional. The authors refer to these gauges as "flow gauges" because the equation eliminating the arbitrariness,

$$\dot{U} + ig(A_0 U - UZ[\mathbf{A}^U]) = 0, \tag{1.26}$$

[which follows from the conditions $A_0^U = Z[\mathbf{A}^U]$, $A_\mu^U \equiv U^{-1}[A_\mu + (ig)^{-1}\partial_\mu]U$] "determines the flow known from nonlinear dynamics" (Ref. 56). A special case of (1.26) is the gauge

$$\beta A_0 = \operatorname{div}\mathbf{A}, \quad 0 < \beta < \infty. \tag{1.27}$$

For $\beta \to 0$ or $\beta \to \infty$ this condition is transformed into the Maxwell gauge (1.9) or the Weyl gauge (1.10), respectively. The class of gauges (1.25) contains the nonlocal gauges[56]

$$\alpha A_0 = (-\Delta)^{-1/2} \operatorname{div}\mathbf{A}, \quad \Delta = \boldsymbol{\partial}^2. \tag{1.28}$$

The parameter α is dimensionless [in contrast to β in (1.27), which has the dimensions of mass]. In the gauge with fixing term

$$\mathscr{L}' = -\frac{1}{2\alpha}(\beta A_0 - \boldsymbol{\partial}\mathbf{A})^2 \tag{1.29}$$

there are no infrared divergences (at least in the one-loop approximation[56]).

13. The parametric gauge specified by the fixing term[57,58]

$$\mathscr{L}' = -\frac{1}{6\beta^2}\partial A(\Box - \beta^2)\partial A. \tag{1.30}$$

For $\beta \to \infty$ the Fried–Yennie gauge[59] is obtained, and for $\beta \to 0$ one essentially obtains the Lorentz gauge.

The Fermi gauges (1.16) can also be classified as interpolating gauges, since special cases of them are the Feynman gauge ($\alpha = 1$), the Lorentz gauge ($\alpha = 0$), and the Fried–Yennie gauge ($\alpha = 3$). It is not hard to think up an intermediate gauge also for other conditions, for example,

$$\alpha\dot{A}_0 - \boldsymbol{\partial}\mathbf{A} = 0. \tag{1.31}$$

For $\alpha = 1$ and 0 this becomes the Lorentz gauge and the radiation gauge, respectively, and for $\alpha \to \infty$ it becomes the inhomogeneous Weyl gauge (1.10), $A_0 = f(x)$, $\partial_0 f = 0$.

The gauge conditions (1.8)–(1.23), (1.25), (1.27)–(1.31) have been written in the form natural for electrodynamics. Their generalization to non-Abelian theories amounts to the replacement $A_\mu \to A_\mu^a$ [and $\varphi \to \varphi^a$ in (1.17)], where $a = 1, \ldots, N$, $N = \dim X$ (X is the Lie algebra of the gauge group). Only the background-field gauge (1.23) needs more explanation; there it is necessary to make the further replacements $q_\mu \to q_\mu^a$, $A_\mu q_\mu \to if^{abc}A_\mu^b q_\mu^c$, where f^{abc} are the structure constants of the group. We should also mention the so-called contour gauges, which arise in working with P-exponentials and can be formally classified as nonlocal gauges.[60]

14. "Tensor" gauge conditions. In the examples given above the gauge freedom is eliminated by fixing the connection. In principle, it can be eliminated by fixing any quantity which transforms nontrivially, for example, a tensor. If the scalar field φ transforms as an "isovector" [gauge group $SO(n)$], then the $n-1$ unphysical variables are eliminated by the conditions

$$\varphi_2 = \varphi_3 = \ldots = \varphi_n = 0. \tag{1.32}$$

This example illustrates the statement that tensors cannot be considered to be more physical than connections: they both contain physical and unphysical components. The following important fact should be noted. The dimension of the group $SO(n)$ is $D=n(n-1)/2$, i.e., the number of gauge parameters is equal to D, whereas the rank-1 tensor φ has only n components, and $n<D$ for $n>2$. The paradox is simple to resolve: the stationary subgroup of the vector φ is $SO(n-1)$, i.e., $D'=(n-1)(n-2)/2$ generators of the group $SO(n)$ cancel the vector φ. There remain $D-D'=n-1$ parameters, which are also fixed. Therefore, when other tensor variables are present this gauge will be incomplete: $(n-1)(n-2)/2$ gauge parameters remain arbitrary.

15. The number of gauges used in the theory of gravitation is rather small. The most popular one is the De Donder–Lanczos–Fock gauge:[61–63]

$$\partial_\mu(\sqrt{-g}g^{\mu\nu})=0. \tag{1.33}$$

This condition is obviously the analog of the Lorentz condition (1.8). It determines the so-called "harmonic coordinates." The latter satisfy the d'Alembert equation

$$\Box x^\nu=0, \tag{1.34}$$

where $\Box\equiv-(-g)^{-1/2}\partial_\mu\sqrt{-g}g^{\mu\nu}\partial_\nu$. However, $\Box x^\nu=-(-g)^{-1/2}\partial_\mu(\sqrt{-g}g^{\mu\nu})$, i.e., the condition (1.34) ensures that the coordinates are harmonic. We note that also in the Lorentz condition (1.8) the remaining freedom is restricted by the d'Alembert equation $\Box\Lambda=0$ [see (1.1)]. This is not gauge freedom, because it is completely eliminated by specifying the initial conditions. In the relativistic theory of gravity the conditions (1.33) are raised to the status of field equations.[64]

16. There are also other gauges, for example, the light-cone gauge,[65,66] which is written as

$$n^\mu g_{\mu\nu}, \quad n^2=0. \tag{1.35}$$

The analog of the Weyl gauge (1.10) is the condition

$$g^{\mu 0}=0. \tag{1.36}$$

Dirac showed[67] that the remaining freedom associated with the choice of coordinates on the hypersurface $t=0$ is best eliminated by requiring that

$$\partial_i\tilde{g}^{ik}=0, \quad \tilde{g}^{ik}=g^{ik}(\det g_{ik})^{1/3}, \quad i,k=1,2,3. \tag{1.37}$$

In the Hamiltonian formalism Eq. (1.37) is understood in the weak sense, i.e., it is taken into account only after calculating the Poisson brackets.

The problem of eliminating the gauge freedom in string and superstring theories is in principle no different from that in field theories. For example, for the string with Nambu–Goto action (1.6) it is natural to use the orthonormal gauge[13]

$$\dot{x}^2-x'^2=0, \quad \dot{x}x'=0, \tag{1.38}$$

which implies transformation to conformal (orthonormal) coordinates on the surface swept out by the string. Other gauge conditions are also used, in particular, the light-cone gauge (see Refs. 13–16, and 31 for more details).

Special mention should be made of the BRST method.[68–70] Auxiliary (fictitious) anticommuting scalar fields are introduced into the formalism. The gauge-fixing term $\mathscr{L}'$ together with the Lagrangian of the fictitious fields $\mathscr{L}''$ is invariant under global supersymmetry transformations (see Sec. 2). In this approach the gauge (unphysical) degrees of freedom are the even elements of a Grassmann algebra. The explicit violation of local gauge invariance is formal in nature—the role of the unphysical degrees of freedom is taken over by the auxiliary Grassmann fields, so that the theory preserves all the properties following from the gauge invariance of the original Lagrangian. It is this formalism which is most convenient for deriving the Ward identities.[71] The method was generalized to the case of nonlinear gauges in Ref. 72.

The abundance of gauges suggests many questions.

What are the special features of any particular gauge?

Are there preferred gauges?

Are the arbitrary functions in Eqs. (1.1)–(1.3) really completely arbitrary?

Are there restrictions on the gauge-fixing terms $\mathscr{L}'$? If so, what are they?

How can the classical gauge conditions be carried over to quantum theory?

In order to analyze the large number of gauges, it is necessary to somehow classify them. The following criteria are usually used.

Relativistic invariance. The gauges (1.8), (1.14)–(1.20), (1.23), (1.30), (1.32), (1.33), and (1.38) can obviously be classified as Lorentz-invariant gauges, whereas in the others the relativistic invariance is explicitly violated. The Fock–Schwinger gauge violates invariance under translations $x\to x+a$.

Linearity. The conditions eliminating the gauge freedom can be linear or nonlinear in the gauge fields (A_μ) and the matter fields. For example, the functional on the right-hand side of (1.25) can be nonlinear.

Uniqueness. A gauge condition can have several or even an infinite number of solutions.[23,73–76]

The presence of ghosts (the appearance of fictitious fields in the formalism[77,78]). It becomes necessary to use unphysical anticommuting scalar fields for gauges of the type (1.8), (1.9), and (1.16) in theories with a non-Abelian gauge group,[77,78] and for the gauge (1.20) also in electrodynamics.

Locality. Local gauge conditions relate fields and their derivatives at a single point. Equation (1.28) is an example of a nonlocal gauge.

Renormalizability and unitarity. Gauges in which the theory has an explicitly renormalizable form (according to power counting) are referred to as R gauges. The gauges (1.16) and (1.17) are examples. Gauges in which the S matrix has an explicitly unitary form (unphysical fields do not propagate) are referred to as U gauges.[42] The gauge (1.17) for $\alpha\to\infty$ is an example (the vector-field propagator in this case does not fall off as the momentum tends to infinity).

Homogeneity. If, for example, instead of (1.8) we take

$$\partial_\mu A_\mu=f(x), \tag{1.39}$$

where f is a function (or field), this gauge is termed inhomogeneous.

Dimensionality. Gauge conditions are sometimes classified on the basis of their dimension. For example, the Lorentz gauge (1.8) has dimension 2 ($[\partial_\mu A_\mu]=M^2$), and the gauges (1.13) have dimension 1 ($[nA]=M$).

Algebraic gauges.[79] In this case the components of the gauge fields are related by algebraic conditions. Examples are (1.10)–(1.15), (1.32), (1.35), and (1.36).

Incomplete gauges. If the gauge condition completely eliminates the freedom, the gauge is termed complete; otherwise it is incomplete (see Secs. 3 and 4).

Dynamical and nondynamical gauges. This distinction is useful mainly in connection with the quantized theory. Gauges specified by algebraic conditions or (in the Lagrangian formalism) differential conditions of no higher than first order in time will be referred to as nondynamical gauges. Gauges specified by the inclusion in the Lagrangian of terms $\mathscr{L}'$ quadratic in the velocities of the unphysical components will be referred to as dynamical gauges. In the classical theory such a distinction has no particular meaning; for example, the Lorentz gauge (1.8) (nondynamical) and the Feynman gauge (1.16), $\alpha=1$, (dynamical) lead to the same equations of motion (Sec. 4). In the quantum theory nondynamical gauge conditions are conditions on the canonical variables. However, it is impossible to require that, for example, one of the canonical variables disappear. This would lead to violation of the canonical commutation relations. Dynamical gauges make it possible to avoid this problem.

1.3. Gauge transformations

Many questions also arise in connection with the problem of choosing the gauge functions. This is related to the problem of the gauge conditions, but is not identical to it. It is traditionally assumed that the function $\Lambda(x)$ in (1.1) is absolutely arbitrary. Of course, this is not so; it cannot be completely arbitrary. Gauge transformations cannot change the nature of the fields, so that it cannot be either a complex function (the field A_μ is real), or a 4-vector, or a generator of a Grassmann algebra, and so on. These are only the most obvious restrictions on the class of allowed transformations. The more subtle characteristics of the function $\Lambda(x)$ in (1.1) and the matrices $U(x)$ in (1.2) are also important for physics (this question is discussed in more detail in Sec. 5). The following classes of transformations are distinguished:

•local;
•global;
•-large;
•singular;
•supersymmetric.

Global transformations, in contrast to local ones, are characterized by group parameters which are independent of the coordinates and time.[4)] The transformations of the first two groups are classified as small. To perform them it is sufficient to specify the generators and parameters of the gauge transformations. Another concept has appeared, that of a "large transformation."[26] It is associated with the topology of the space and the fields. According to (1.2), the field $U\partial_\mu U^{-1}$ is a purely gauge field, i.e., the field-strength tensor corresponding to it is zero: $F_{\mu\nu}=0$. Therefore, when determining the law for the falloff of the fields at $|\mathbf{x}|\to 0$ we have no right to require that they vanish ($A_\mu\to 0$). We can only require that $F_{\mu\nu}\to 0$, i.e., $A_\mu\to U\partial_\mu U^{-1}$. The seemingly innocent weakening of the asymptotic condition has deep consequences. Let $U(x)\equiv U_x\to U_\infty$ for $|\mathbf{x}|\to\infty$. If this limit is independent of the direction of $\mathbf{x}$, this implies that we have actually gone from R^3 to S^3, i.e., we have compactified 3-dimensional Euclidean space by adding to it a point at infinity. However, $U(x)\in G$, i.e., $U(x)$ performs the mapping $S^3\to G$. It is known from topology that the set of such mappings splits up into topologically inequivalent classes characterized by topological numbers.[81] For G=SU(2) this number is[26]

$$n=-\frac{1}{48\pi^2}\int d^3x\ \mathrm{Tr}[U\partial_i U^{-1}U\partial_j U^{-1}U\partial_k U^{-1}]\varepsilon^{ijk}, \tag{1.40}$$

where ε^{ijk} is the unit antisymmetric tensor. Fields of different classes cannot be transformed into each other by transformations with continuous gauge parameters (see Sec. 5).

Large transformations are ones which take fields from one class into another (Ref. 26, p. 667), i.e., transformations U with $n\neq 0$. It is important that all the "vacuum fields" $A_\mu=U\partial_\mu U^{-1}$ of the different classes have zero tensor $F_{\mu\nu}$, i.e., they correspond to zero classical energy. The potential energy of the gauge field therefore has a countable set of minima separated by potential barriers.

Singular gauge transformations arise in the theory of the magnetic monopole. The Dirac monopole is a magnetic pole with a magnetic string attached to it.[82,83,40] This construction would be a monopole if the string were unobservable. In this case displacement of the string would correspond to an operation on an unphysical object and, as might be expected, would be accomplished by a gauge transformation. Such a "gauge" transformation actually can be imagined, but it is not described by a single-valued function Λ. If the contour around the initial undisplaced string does not encircle the string after displacement to its final position, the value of the function Λ after a circuit of the contour is different from its initial value. The string is a singularity line Λ, which explains the name of this type of transformation.[40,84] However, displacement of the string cannot be considered an unphysical operation, since the string itself cannot be an unphysical object—the flux of magnetic field through it is nonzero (see the introductory discussion in Ref. 83). Dropping the requirement that Λ be continuous leads to transformations which change the physics.

Supersymmetric gauge transformations. Right after the discovery of supersymmetry (Refs. 85–88; see also Refs. 89–92), supergravity[93,94] was proposed. It is the result of going from global supersymmetry to local supersymmetry. In this theory the gauge parameters are elements of a Grassmann algebra; the parameters of global supersymmetry transformations become functions of the coordinates. Even less is known about the physical nature of local supersymmetry than about gauge symmetry, although its importance for physics cannot be doubted.

Let us single out a few of the other problems associated with gauge invariance. A very important problem is that of the gauge-fixing term $\mathscr{L}'$. It is easy to find examples where $\mathscr{L}'$, even though it violates gauge invariance, does not lift the degeneracy of the Lagrangian (for example, $\mathscr{L}'=m^2A_\mu^2/2$), or even leads to inconsistent dynamics[95] ($\mathscr{L}'=x^\mu A_\mu$; see Sec. 4 for more details).

In addition, we should mention the *Dirac-extended group of gauge transformations*[96] (see Sec. 2). The point is that the set of unphysical degrees of freedom is not identical to the set of gauge parameters of the original Lagrangian. Analysis shows that in addition to primary constraints (the number of which is equal to the number of parameters of the gauge group), there are also secondary ones.[97,98] They indicate that an additional set of unphysical variables is present. The law for the time variation of the latter is usually completely determined by the law for the variation of the former, corresponding to the primary constraints. However, in principle, this law can be specified independently, i.e., the actual freedom in the theory is larger. A theory in which the number of gauge (arbitrary) parameters is equal to the number of all first-class constraints (both primary and secondary) is referred to as a theory with an extended gauge group.

Finally, the very concept of gauge invariance should be refined. The standard Lagrangian of gauge theory does not change under gauge transformations. In this case one speaks of strong gauge invariance. However, there exist physical objects whose change is proportional to the constraints, i.e., they are not changed only by the inclusion of the constraints (they are said to be "invariant on the constraints"). In this case there is *weak gauge invariance*. The Hamiltonian is an example of such an object (see Sec. 2).

By now it should be clear that the theory of gauge fields is quite complex and far from complete. There is an enormous literature devoted to the problems of the choice of gauge and the fine points of using various gauges. An entire book would be required to discuss them all. In the present review we discuss only some of the basic problems associated with the gauge invariance of a theory. Our attention is mainly focused on the role played by the physical and nonphysical degrees of freedom in a particular gauge choice. We only briefly touch upon the questions of gauge fixing in gravitation and string theory. We also do not discuss the problem of performing calculations in a given gauge—the Feynman gauge is the standard, and the corresponding technique is described in all texts on quantum field theory. Noncovariant gauges are discussed in detail in the thorough review of Ref. 31 and in Ref. 79.

In Sec. 2 we discuss the problem of physical and unphysical degrees of freedom, mainly in electrodynamics. We stress the importance of the concept of an extended gauge group introduced by Dirac and the concept of weak gauge invariance. In Sec. 3 we study simple models with a finite number of degrees of freedom illustrating some of the typical problems of gauge theories. In Sec. 4 we perform a comparative analysis of different gauges. The special role of the Fock gauge is discussed. Section 5 is devoted to problems related to gauge transformations. In Sec. 6 we briefly summarize our discussion. The Appendix contains material dealing with certain aspects of gauge invariance and gauge conditions: the background-field method, the role of the residual gauge group, historical remarks, and so on.

Notation. We use the metric $g_{\mu\nu}(+---)$; the components of a coordinate 4-vector x are denoted by $x^\mu=(\mathbf{x},t)$. As a rule, we shall use the word "coordinates" to refer to the space coordinates and the time. Greek indices take the values 0,1,2,3; Latin indices i,j,k,l, if not stated otherwise, take the values 1,2,3, while the indices a,b,c,d take the values $1,2,\ldots,\dim X$, where X is the Lie algebra of the gauge group G. We use abbreviated notation for the differentiation operator $\partial_\mu=\partial/\partial x^\mu$ and for the d'Alembertian $\Box\equiv-g^{\mu\nu}\partial_\mu\partial_\nu$ (or $\Box=-(1/\sqrt{-g})\,\partial_\mu\sqrt{-g}g^{\mu\nu}\partial_\nu$, where $g=\det g_{\mu\nu}$ in the case where $\partial_\rho g_{\mu\nu}\neq0$). Repeated indices of the same type are understood to be summed over with the appropriate metric tensor, for example, $q_\mu x_\mu=g^{\mu\nu}q_\mu x_\nu=q_\mu x^\mu\equiv qx$. The commutator or anticommutator is indicated by the sign on the square brackets: $[A,B]_\pm=AB\pm BA$. The Poisson brackets are defined by the condition $\{q,p\}=1$. Equations valid only when constraints are included ("equalities in the weak sense") are indicated by the symbol $\approx$. In path integrals we ignore the presence of the constant in the measure. A product of functions can be understood as an integration over coordinates: $JA\equiv\int dxJ_\mu(x)A_\mu(x)$, where $dx\equiv d^4x$.

2. GAUGE THEORIES: THE LAGRANGIAN AND HAMILTONIAN METHODS

A typical gauge-invariant Lagrangian looks like

$$\mathscr{L}=-\frac{1}{4}\operatorname{Tr}F_{\mu\nu}^2+\sum_f\bar{\psi}_f(i\hat{D}-m_f)\psi_f+\sum_f D_\mu\varphi_f(D_\mu\varphi_f)^*+\cdots,\tag{2.1}$$

where $F_{\mu\nu}=(ig)^{-1}[D_\mu,D_\nu]$, $D_\mu=\partial_\mu-igA_\mu$, $A_\mu=A_\mu^a\lambda^a$, and $\hat{D}=\gamma_\mu D_\mu$. The matrices λ^a forming the basis of the Lie algebra in the fundamental representation are normalized to unity, $\operatorname{Tr}\lambda^a\lambda^b=\delta^{ab}$, and the Dirac matrices γ_μ satisfy the standard condition $[\gamma_\mu,\gamma_\nu]_+=2g_{\mu\nu}$. In addition, ψ is a spinor field, $\bar{\psi}=\psi^*\gamma_0$, and φ is a complex scalar field; obviously, A_μ in the covariant derivative of φ is $A_\mu=A_\mu^aT^a$, where the T^a are the group generators in the representation realized by this field. The summation in (2.1) runs over the types of field f (flavors) if there are several of them; the dots stand for other allowed gauge-invariant terms, for example, a self-interaction of the scalar field and contributions of other fields. The Lagrangian (2.1) is invariant under the transformations (1.2). To explain the features of the dynamics of gauge systems, let us consider the simplest and most well studied theory, that of electrodynamics (Abelian gauge group).

2.1. Free electrodynamics

The corresponding Lagrangian is obtained from (2.1) by the replacement $\lambda^a\to1$ (the unit matrix):

$$\mathscr{L}=-\frac{1}{4}F_{\mu\nu}^2+\bar{\psi}(i\hat{D}-m)\psi+D_\mu\varphi(D_\mu\varphi)^*+\cdots.\tag{2.2}$$

Here $F_{\mu\nu}=\partial_\mu A_\nu-\partial_\nu A_\mu$ and $D_\mu=\partial_\mu-ieA_\mu$; for simplicity we consider only a single charged field. The dynamics of systems with a gauge group are determined already by the first term in (2.2).

The Lagrangian of the free electromagnetic field can be written as

$$\mathscr{L}_0=-\frac{1}{4}F^2_{\mu\nu}\equiv\frac{1}{2}\partial_\rho A_\mu T^{\rho\mu\sigma\nu}\partial_\sigma A_\nu, \qquad (2.3)$$

where

$$T^{\rho\mu\sigma\nu}=g^{\rho\nu}g^{\sigma\mu}-g^{\rho\sigma}g^{\mu\nu}. \qquad (2.4)$$

Using the tensor operator

$$K^{\mu\nu}=-T^{\rho\mu\sigma\nu}\partial_\rho\partial_\sigma, \qquad (2.5)$$

we rewrite (2.3) as

$$\mathscr{L}_0=\frac{1}{2}A_\mu K^{\mu\nu}A_\nu+\frac{1}{2}\partial_\rho(A_\mu T^{\rho\mu\sigma\nu}\partial_\sigma A_\nu). \qquad (2.6)$$

The last term in (2.6) is a total divergence and can be dropped. The features of the classical and quantum description of the electromagnetic field follow from Eqs. (2.3)–(2.6). The velocity dependence of $\mathscr{L}_0$ (the bilinear term) is determined by the matrix

$$T^{\mu\nu}=T^{0\mu0\nu}=\frac{\partial^2\mathscr{L}_0}{\partial\dot{A}_\mu\partial\dot{A}_\nu}=g^{0\mu}g^{0\nu}-g^{00}g^{\mu\nu}. \qquad (2.7)$$

This matrix is degenerate because, being diagonal, it has a zero along the diagonal: $T^{\mu\nu}=\mathrm{diag}(0,1,1,1)$. Lagrangians possessing this property are also called *degenerate* (or singular): $\mathscr{L}_0$ is not a quadratic function of the velocity $\dot{A}_0$, i.e., the coefficient of $\dot{A}_0^2$ in (2.3) is equal to zero. In fact, the Lagrangians (2.2) and (2.3) are completely independent of $\dot{A}_0$. The matrix K is also degenerate:

$$K_{\mu\nu}=-g_{\mu\nu}\Box-\partial_\mu\partial_{\mu\nu}, \quad \Box=-\partial^2_\mu, \qquad (2.8)$$

since $P^{(1)}_{\mu\nu}=-\Box^{-1}K_{\mu\nu}$ is a projection operator ($g_{\mu\nu}=P^{(1)}_{\mu\nu}+P^{(0)}_{\mu\nu}$, $P^{(i)}_{\mu\rho}P^{(j)}_{\rho\nu}=P^{(i)}_{\mu\nu}\delta^{ij}$, $i,j=0,1$). The most important consequences of these features are:

(a) in the classical theory: the problem of transforming to the Hamiltonian description of the system (insolubility of the equations $p_i=\partial L/\partial\dot{q}^i$ for $\dot{q}^i$ (Ref. 96);

(b) in the quantum theory: the problem of defining the propagator, which arises because in a nondegenerate theory the propagator is $\Delta_{\mu\nu}\sim K^{-1}_{\mu\nu}$.

2.2. The extended group of gauge transformations

The general theory of dynamical systems with constraints was constructed by Dirac and Bergmann,[96–98] although its basic elements for the case of electrodynamics are already contained in the papers of Heisenberg and Pauli.[32,33] The essence of the problem appears in going to the Hamiltonian formalism. From (2.3) we have

$$\pi^\mu=\frac{\partial\mathscr{L}}{\partial\dot{A}_\mu}=F^{\mu0}, \qquad (2.9)$$

i.e., the momentum canonically conjugate to A_0 is equal to zero:

$$\pi^0=0. \qquad (2.10)$$

In the Bergmann–Dirac terminology this is a primary constraint (following tradition, by "constraints" we mean both functions of the canonical variables and the conditions for them to vanish). If we ignore (2.10), we find the "Hamiltonian" in the standard manner:

$$H_0=\int d^3x\left[\frac{1}{2}(\mathbf{E}^2+\mathbf{H}^2)\boldsymbol{\partial}A_0\mathbf{E}+\pi^0\dot{A}_0\right],$$

$$E^k=\pi^k=F^{k0}, \quad \mathbf{H}^2=\frac{1}{2}F^2_{ik}. \qquad (2.11)$$

Strictly speaking, this is not a Hamiltonian, because H_0 depends on the velocity $\dot{A}_0$. However, assuming that $\dot{A}_0$ is an arbitrary function of time and taking into account (2.10), the functional (2.11) can be used to obtain the Hamiltonian equations of motion.[96] A check that (2.10) and (2.11) are consistent (that $\dot{\pi}^0=0$) leads to the condition[5)]

$$\dot{\pi}^0=\{\pi^0,H_0\}=\partial_k E^k=\mathrm{div}\,\mathbf{E}=0, \qquad (2.12)$$

where $\{,\}$ are the functional Poisson brackets:

$$\{A_\mu(\mathbf{x},t),\pi^\nu(\mathbf{y},t)\}=\delta^\nu_\mu\delta(\mathbf{x}-\mathbf{y}). \qquad (2.13)$$

We have thus obtained a secondary constraint, which obviously is in involution with the primary one ($\{\pi^0,\boldsymbol{\partial}\mathbf{E}\}=0$), i.e., π^0 and $\boldsymbol{\partial}\mathbf{E}$ are first-class constraints. There are no other constraints ($\{\boldsymbol{\partial}\mathbf{E},H_0\}=0$). According to the Dirac analysis,[96] it is possible to go to the total Hamiltonian H_{0T},

$$H_{0T}=\int d^3x\left[\frac{1}{2}(\mathbf{E}^2_\perp+\mathbf{H}^2)+u\pi^0-v\boldsymbol{\partial}\mathbf{E}\right], \quad \boldsymbol{\partial}\mathbf{E}_\perp=0, \qquad (2.14)$$

with arbitrary functions of time u,v. The Hamiltonian (2.14) shows that: (i) there are two first-class constraints in the theory, i.e., only two of the four components of A_μ are physical; (ii) the group of transformations not affecting the physical sector is characterized by two parameters u and v depending on the coordinates and the time. The constraints π^0 and $\boldsymbol{\partial}\mathbf{E}$ act as the generators of these transformations. We define

$$G=\int d^3x[u\pi^0-v\boldsymbol{\partial}\mathbf{E}]. \qquad (2.15)$$

Calculating the Poisson brackets of A_μ and G, we find

$$\delta A_0=\{A_0,G\}=u\equiv\delta_u A_0,$$
$$\delta\mathbf{A}=\{\mathbf{A},G\}=\partial v\equiv\boldsymbol{\delta}_v\mathbf{A}; \quad \delta_v A_0=\delta_u\mathbf{A}=0. \qquad (2.16)$$

We shall refer to the set of transformations given by (2.16) as the Dirac-extended group of gauge transformations of free electrodynamics. They exhaust the gauge freedom of the theory. The fact that the Lagrangian (2.3) is invariant only under the transformations (1.1) (a single arbitrary function) implies that the equations of motion in free electrodynamics are such that fixing one unphysical degree of freedom completely (or up to the initial conditions) fixes the other. The one-parameter group of transformations (1.1) is obtained from (2.15) and (2.16) by the reduction $u=\dot{v}$, $v\equiv\Lambda$; then Eq. (2.16) is written in the usual form:

$$\delta A_\mu=\partial_\mu\Lambda. \tag{2.17}$$

This can also be illustrated for the Lagrangian equations of motion:

$$\mu_\mu F_{\mu\nu}=0. \tag{2.18}$$

Expressing $F_{\mu\nu}$ in terms of the potentials, we obtain

$$\partial_k F_{k0}=\boldsymbol{\partial}(\dot{\mathbf{A}}-\boldsymbol{\partial}A_0)=0,$$

$$\partial_\mu F_{\mu i}=-[\Box\mathbf{A}+\boldsymbol{\partial}(\dot{A}_0-\boldsymbol{\partial}\mathbf{A})]_i=0. \tag{2.19}$$

Splitting $\mathbf{A}$ into longitudinal and transverse components

$$\mathbf{A}=\mathbf{A}_\perp+\mathbf{A}_\parallel,\quad \boldsymbol{\partial}\mathbf{A}_\perp=0,\quad \mathbf{A}_\parallel=\Delta^{-1}\boldsymbol{\partial}(\boldsymbol{\partial}\mathbf{A}) \tag{2.20}$$

(since $\boldsymbol{\partial}\mathbf{A}_\parallel=\boldsymbol{\partial}\mathbf{A}$), we write (2.19) in the form

$$\Delta A_0=\partial_0\boldsymbol{\partial}\mathbf{A}_\parallel,\quad \Box\mathbf{A}_\perp=\partial_0(\partial_0\mathbf{A}_\parallel-\boldsymbol{\partial}\mathbf{A}_0). \tag{2.21}$$

The first of these equations is the Lagrangian constraint [it contains only first derivatives with respect to time; cf. (2.12)]. Its solution makes the right-hand side of the second equation vanish. From this we conclude that (1) the transverse (physical) components of $\mathbf{A}$ satisfy the d'Alembert equation

$$\Box\mathbf{A}_\perp=0; \tag{2.22}$$

and (2) the longitudinal component of $\mathbf{A}$ is expressed in terms of A_0:

$$\mathbf{A}_\parallel=\int^t dt\,\boldsymbol{\partial}A_0. \tag{2.23}$$

Therefore, by specifying A_0 we thereby also fix the second unphysical variable $\mathbf{A}_\parallel$.

2.3. Weak gauge invariance

Usually, when speaking of gauge invariance in electrodynamics one means the invariance of some function or functional of the gauge fields and their derivatives under the gradient transformations (1.1). The Lagrangian (2.2) is an example. In the Hamiltonian formalism a dynamical quantity is gauge-invariant if its Poisson brackets with the generator (2.15) (for $u=\dot{v}$) are equal to zero. It is said that this quantity is invariant in the strong sense. Physically, this requirement is exaggerated. The physical sector of the theory is determined by the inclusion of all the constraints, so that the necessary condition for a dynamical quantity to belong to the physical sector is that it be weakly invariant under the extended gauge group: The quantity must be in involution with all the first-class constraints. This means that its Poisson brackets with the first-class constraints must vanish when the constraints are taken into account. Dropping this requirement (or strengthening it) leads to unacceptable consequences. For example, the Poisson brackets of the Hamiltonian (2.11) (with the replacement $\dot{A}_0\to u$) with the generator of gauge transformations (2.15) are

$$\{H_0,G\}=-\int d^3x\,v(x)\boldsymbol{\partial}\mathbf{E}, \tag{2.24}$$

i.e., the Hamiltonian (2.11) is not invariant even under the original group (1.1) [when $u=\dot{v}$ in (2.15)]. However, the Hamiltonian of the system is certainly a physical quantity. The Poisson brackets (2.24) are zero in the weak sense, i.e., when the constraint $\boldsymbol{\partial}\mathbf{E}=0$ is included.

The requirement of weak gauge invariance is consistent with the requirement of invariance under the extended gauge group also from the viewpoint of the Lagrangian formalism. In fact, the Lagrangian of the free theory (2.3) is not invariant in the strong sense under gauge transformations from the extended group (2.16). Since here

$$\delta F_{k0}=\partial_k(u-\dot{v}),\quad \delta F_{ik}=0, \tag{2.25}$$

we have

$$\delta\mathscr{L}_0=F_{k0}\delta F_{k0}=-(u-\dot{v})\boldsymbol{\partial}\mathbf{E}+\boldsymbol{\partial}[(u-\dot{v})\mathbf{E}]. \tag{2.26}$$

The last term is unimportant (it is a total divergence), so that the variation of the Lagrangian functional $L_0=\int d^3x\mathscr{L}_0$ in the transformations (2.16) vanishes only when the Lagrangian constraint (2.21) is included: $\delta L_0=0$ for $\boldsymbol{\partial}\mathbf{E}=\boldsymbol{\partial}(\dot{\mathbf{A}}-\boldsymbol{\partial}A_0)=0$.

The requirements discussed in the last two subsections form the basis of all gauge theories.

2.4. Electrodynamics with an interaction

The inclusion of an interaction changes only the secondary constraint, and, as before, the constraints are in involution:

$$G_1=\pi^0=0,\quad G_2=\boldsymbol{\partial}\mathbf{E}-j_0=0,\quad \{G_1,G_2\}=0 \tag{2.27}$$

(j_0 is the zeroth component of the current $j_\mu=-\partial\mathscr{L}/\partial A_\mu$), i.e., as before, the theory contains two unphysical degrees of freedom. One of these, as before, is A_0, and the other is the quantity canonically conjugate to G_2 (2.27). Since $\{\mathbf{A}_\parallel,G_2\}\neq 0$, $\mathbf{A}_\parallel$ is changed by gauge transformations, so that the longitudinal component of the field $\mathbf{A}$ is not a physical degree of freedom. This is manifested in the fact that it cannot propagate independently of the charges. However, the longitudinal component $\mathbf{A}_\parallel$ is easily detected: the static Coulomb field surrounding the charges is the excited longitudinal field;[100] in this sense the field $\mathbf{A}_\parallel$ is observable. A detailed analysis of this question can be found in Ref. 95. The point is that both $\mathbf{A}_\parallel$ and the phase of the charged field are linear combinations of the physical and unphysical degrees of freedom. The gauge-invariant quantity describing the Coulomb field is also such a combination. $\mathbf{A}_\parallel$ can be made to vanish by gauge transformations, but here the charged fields become nonlocal. The physical information about the Coulomb field is transferred to the phase of the charged fields.

The generator of the extended gauge group is written as

$$G=\int d^3x[u\pi^0-v(\boldsymbol{\partial}\mathbf{E}-j_0)], \tag{2.28}$$

and Eqs. (2.18) and (2.21) undergo obvious changes:

$$\partial_\mu F_{\mu\nu}=j_\nu, \quad (2.29)$$

$$\Delta A_0=\boldsymbol{\partial}\dot{\mathbf{A}}-j_0, \quad \Box\mathbf{A}_\perp=\partial_0(\partial_0\mathbf{A}_\parallel-\boldsymbol{\partial}A_0)-\mathbf{j}. \quad (2.30)$$

The first of these is a Lagrangian constraint (it contains only first derivatives of the bosonic field with respect to time). Using the decomposition $\mathbf{j}=\mathbf{j}_\perp+\mathbf{j}_\parallel$, $\boldsymbol{\partial}\mathbf{j}_\perp=0$, $\mathbf{j}_\parallel=\Delta^{-1}\boldsymbol{\partial}(\boldsymbol{\partial}\mathbf{j})$, and the first of Eqs. (2.30), we rewrite the second as

$$\Box\mathbf{A}_\perp=-\mathbf{j}_\perp+\Delta^{-1}\boldsymbol{\partial}(\partial_\mu j_\mu); \quad (2.31)$$

therefore, only transverse excitations propagate, because, owing to current conservation $\partial_\mu j_\mu=0$, from (2.31) it follows that

$$\Box\mathbf{A}_\perp=-\mathbf{j}_\perp. \quad (2.32)$$

As far as the longitudinal field $\mathbf{A}_\parallel$ is concerned, from Gauss's law (2.27) we find the following for the longitudinal component of the electric field strength: $\mathbf{E}_\parallel=\boldsymbol{\partial}\Delta^{-1}j_0$ ($\mathbf{E}=\mathbf{E}_\perp+\mathbf{E}_\parallel$, $\boldsymbol{\partial}\mathbf{E}_\perp=0$, curl $\mathbf{E}_\parallel=0$). Substituting this expression into the new Hamiltonian

$$H_T=\int d^3x\left[\frac{1}{2}(\mathbf{E}^2+\mathbf{H}^2)+\mathbf{A}\mathbf{j}\right]+G \quad (2.33)$$

and taking into account the equations

$$\mathbf{A}\mathbf{j}=\mathbf{A}_\perp\mathbf{j}_\perp+\mathbf{j}_\parallel\boldsymbol{\partial}(\Delta^{-1}\boldsymbol{\partial}\mathbf{A}), \quad \mathbf{E}_\parallel^2=(\Delta^{-1}\boldsymbol{\partial}j_0)^2, \quad (2.34)$$

we conclude that in the radiation gauge (1.9) the Hamiltonian (2.33) is written as

$$H_T=\int d^3x\left[\frac{1}{2}(\mathbf{E}_\perp^2+\mathbf{H}^2)-\frac{1}{2}j_0\Delta^{-1}j_0+\mathbf{j}_\perp\mathbf{A}_\perp\right]+G. \quad (2.35)$$

The longitudinal field has disappeared, but not without a trace: the second term in (2.35) is the contribution to the Hamiltonian from the Coulomb field [the kernel of the operator Δ^{-1} is $\Delta^{-1}(\mathbf{x},\mathbf{y})=-(4\pi|\mathbf{x}-\mathbf{y}|)^{-1}$]. The charged fields also change. Since $\mathbf{A}_\parallel=\boldsymbol{\partial}(\Delta^{-1}\boldsymbol{\partial}\mathbf{A})$, from the condition $\mathbf{A}_\parallel+\boldsymbol{\partial}\Lambda=0$ we find the Λ which makes $\mathbf{A}_\parallel$ vanish, i.e., which effects the passage to the Coulomb gauge: $\Lambda=-\Delta^{-1}\boldsymbol{\partial}\mathbf{A}$. According to (1.1), we have

$$\psi'=e^{-ie\Delta^{-1}\boldsymbol{\partial}\mathbf{A}}\psi; \quad \mathbf{A}_\parallel=\boldsymbol{\partial}(\Delta^{-1}\boldsymbol{\partial}\mathbf{A}). \quad (2.36)$$

The field ψ becomes nonlocal if the Poisson brackets with the canonical momentum $\pi^k=E^k$ are nonzero. The factor in front of ψ corresponds to the Coulomb field surrounding the charge (See Ref. 100, Sec. 80, and Ref. 95).

It is usually assumed (and not without reason) that a physical quantity, a variable, should be gauge-invariant, and that a quantity which changes under gauge transformations cannot have any relation to physics, because it could, for example, be transformed to zero. The case of the Coulomb field considered here contains the necessary subtleties. Quantities which are canonically conjugate to the first-class constraints are unphysical variables (degrees of freedom). Quantities which change under gauge transformations can carry physical information. A gauge transformation then corresponds to the transfer of this information to other degrees of freedom (for example, $\mathbf{A}_\parallel$; see also Sec. 4). Conversely, a gauge-invariant quantity may turn out to be unphysical (examples are $\boldsymbol{\partial}\mathbf{E}$ for free fields and $G_2=\boldsymbol{\partial}\mathbf{E}-j_0$ for interacting ones).

2.5. Non-Abelian theories

The general structure of the formalism does not change in going to a theory with a non-Abelian gauge group [the Lagrangian (2.1)]. Of course, now it is no longer possible to start with the free Lagrangian, setting $g=0$ in (2.1); it is gauge-noninvariant. However, the matrix

$$T_{ab}^{\mu\nu}=\frac{\partial^2\mathscr{L}_0}{\partial\dot{A}_\mu^a\partial\dot{A}_\nu^b}=T^{0\mu0\nu}\delta_{ab} \quad (2.37)$$

is degenerate, as before, i.e., all the problems inherent in electrodynamics are still present here. The momentum canonically conjugate to A_μ^a is

$$\pi_a^\mu=\frac{\partial\mathscr{L}}{\partial\dot{A}_\mu^a}=F_a^{\mu0}, \quad F_{\mu\nu}^a=\mathrm{Tr}(\lambda^aF_{\mu\nu}), \quad (2.38)$$

i.e., we have the primary constraints

$$G_1^a=\pi_0^a=0. \quad (2.39)$$

The first term in (2.1) corresponds to the Hamiltonian

$$H=\int d^3x\left[\frac{1}{2}(\mathbf{E}_a^2+\mathbf{H}_a^2)-\mathbf{E}^a\mathbf{D}^{ab}A_0^b+\dot{A}_0^a\pi_0^a\right],$$

$$\mathbf{H}_a^2=\frac{1}{2}(F_{ik}^a)^2=\frac{1}{2}\mathrm{Tr}(F_{ik}^2), \quad (2.40)$$

where

$$\mathbf{D}^{ab}=\delta^{ab}\boldsymbol{\partial}-igf^{abc}\mathbf{A}^c, \quad [\lambda^a,\lambda^b]=i\sqrt{2}f^{abc}\lambda^c. \quad (2.41)$$

We note that the last term in the Hamiltonian (2.40) can be written as $D_0^{ab}A_0^b\pi_0^a$. The secondary constraints

$$G_2^a=\{\pi_0^a,H\}=\mathbf{D}^{ab}\mathbf{E}^b=0 \quad (2.42)$$

are in involution,

$$\{G_2^a(\mathbf{x},t)G_2^b(\mathbf{y}t)\}=f^{abc}G_2^c(\mathbf{x},t)\delta(\mathbf{x}-\mathbf{y}), \quad (2.43)$$

i.e., their Poisson bracket is proportional to the constraints and vanishes with them. It is easily verified that $\{G_2,H\}\approx0$ and $\{G_1,G_2\}=0$, i.e., all the constraints have been found and they are in involution (first-class constraints). The total Hamiltonian and the generator of the extended group of gauge transformations are given by

$$H_T=\int d^3x\left[\frac{1}{2}(\mathbf{E}_a^2+\mathbf{H}_a^2)+u^aG_1^a-v^aG_2^a\right], \quad (2.44)$$

$$G=\int d^3x(u^aG_1^a-v^aG_2^a), \quad (2.45)$$

where $u^a(x)$ and $v^a(x)$ are arbitrary functions, so that the fields A_μ^a are changed by

$$\delta A_0^a=u^a, \quad \delta\mathbf{A}^a=\mathbf{D}^{ab}v^b. \quad (2.46)$$

If in (2.45) we take $u^a=D_0^{ab}v^b$, we arrive at the standard gauge transformations:

$$\delta A_\mu^a = D_\mu^{ab} v^b. \qquad (2.47)$$

Again as in electrodynamics, we confirm the fact that the Hamiltonian and Lagrangian are invariant only in the weak sense under the extended group of gauge transformations, and the total Hamiltonian H_T is weakly invariant also under the standard gauge group.

Whereas in electrodynamics it is not difficult to follow the fate of the physical and unphysical degrees of freedom, in theories with non-Abelian gauge symmetry this is no longer true. The obvious set of unphysical variables A_0^a (whose velocities do not appear in the Lagrangian) does not cause any trouble. However, the Lorentz-invariant isolation of all the unphysical variables is highly nontrivial.[101] This is in part a consequence of the fact, general for all gauge theories, that transformations from the spacetime symmetry group mix the physical and unphysical degrees of freedom, i.e., in a new reference frame the physical components can contain unphysical ones. This is why in perturbation theory transformations of the vector potential A_μ under the Lorentz group must be accompanied by gauge transformations.[102] This is mainly related to the noncommutativity of the gauge group. Attempts to isolate the physical variables after reformulating the theory in terms of only gauge-invariant quantities have also not met with success: in the simplest cases the formalism becomes much more complicated.[23,101,103] The only reasonable strategy is to combine the constraints with the equations of motion in the classical theory and require that they vanish on vectors from the physical Hilbert subspace in the quantum theory (see below).

2.6. Quantization

The general procedure for quantizing dynamical systems with constraints was worked out by Dirac (Ref. 96; see also Refs. 32–37). In the case of first-class constraints it reduces to the following recipe. (1) All the canonical variables entering into the Hamiltonian obey the canonical commutation relations

$$[A_\mu^\sigma(\mathbf{x},t), \pi_b^\nu(\mathbf{y},t)]_- = i\delta_\mu^\nu \delta_b^a \delta(\mathbf{x}-\mathbf{y}),$$

$$[\psi_\rho^+(\mathbf{x},t), \psi_\sigma(\mathbf{y},t)]_+ = \delta_\rho^\sigma \delta(\mathbf{x}-\mathbf{y}), \qquad (2.48)$$

where the letters ρ, σ include both spinor and other indices; (2) vectors from the physical Hilbert subspace Φ are fixed by requiring that all the first-class constraints vanish on them:

$$G_t^a \Phi = 0. \qquad (2.49)$$

This is the only reliable quantization recipe. Strictly speaking, all other recipes, for example, path-integral quantization, require proof of their equivalence to the canonical recipe.

We stress the fact that the rules applicable in the classical theory do not automatically carry over to the quantum theory. For example, it is impossible to require that the condition (1.8) be satisfied for operators—this would lead to violation of the commutation relations (2.48). The same is true of all other gauge conditions of this type. Here the most natural approach is to use dynamical gauges. In them the unphysical components of the fields are given very simple [or sometimes complicated; see (1.20)] dynamics, and they can be formally treated like the physical components. Their absence in the physical sector of the Hilbert space is ensured by the vanishing on physical vectors of the momenta canonically conjugate to them (the constraints). Similarly, in the quantum theory the constraints cannot be understood in the strong sense, i.e., as operator equations.[96] This would also lead to violation of the commutation relations (2.48).

2.7. The path-integral method

The path-integral method plays an important role in modern quantum field theory. It is popular for several reasons. First, it is the mathematical technique most suited to the problem; second, its use greatly simplifies the calculations (derivation of the Feynman rules, the Ward identities, and so on); finally, it serves as the basis for a unique, regular technique which is not based on smallness of the interaction constant, i.e., it provides a formalism which goes beyond perturbation theory (the semiclassical expansion). The generating functional for the Green function in this approach is given by the path integral[40]

$$Z[\mathcal{T}] = e^{iW[\mathcal{T}]} = \int d[A\bar{\psi}\psi] e^{i(S[Q] + \mathcal{T}Q)}. \qquad (2.50)$$

Here S is some action [in the case of gauge theories, it is the invariant action given, for example, by the Lagrangian (2.1), $S = \int dx \mathcal{L}$; in what follows we shall limit ourselves to one type of field], $\mathcal{J}Q \equiv \int dx (JA + \bar{\eta}\psi + \bar{\psi}\eta)$, A_μ^a is the gauge field, ψ and $\bar{\psi}$ are charged spinor fields, and $J_\mu^a(x)$, $\eta(x)$, and $\bar{\eta}(x)$ are classical sources (the last two belong to a Grassmann algebra). All indices (vector, spinor, group) not written out explicitly are understood to be summed over. Other fields may also be present in this integral. The generating functional $W[\mathcal{J}]$ given by (2.50) is the generating functional for the connected Green functions.[6)] Yet another important object is defined by a Legendre transformation:[51–53]

$$\Gamma[\phi] = W[\mathcal{T}] - \mathcal{T}\phi, \quad \frac{\delta W}{\delta \mathcal{T}} = \phi. \qquad (2.51)$$

It is assumed that the right-hand side of the first equation in (2.51) contains the solution of the second: $\mathcal{J} = \mathcal{J}[\phi]$. Then it is easily shown that

$$\frac{\delta \Gamma[\phi]}{\delta \phi} = -\mathcal{T}. \qquad (2.52)$$

The quantity $\Gamma[\phi]$ is the generating functional for 1-irreducible Green functions (pole diagrams have been eliminated). It gives all the vertices of the effective Lagrangian:[53] $\Gamma = S_{\text{eff}}$. The main property of the latter is that it reproduces the exact S matrix already in lowest-order perturbation theory (the exact probability amplitude of any process is given by the sum of tree graphs.[51,53]). In the operator formalism ϕ is the vacuum expectation value of Q (i.e., of A, ψ, and $\bar{\psi}$) in the presence of external currents.

As already noted, one of the main problems caused by gauge invariance is related to integration over purely gauge degrees of freedom. In calculating the connected Green functions (i.e., the functional derivatives of W with respect to the currents at zero value of the latter), the integration over gauge parameters extends to infinity (see Secs. 2.1. and 2.5.). The direct method of eliminating the gauge freedom is to insert inside the integral (2.50) the product of δ functions of all the group parameters ω, i.e., $\delta(\omega)$. The gauge freedom is usually eliminated by means of gauge conditions on the vector potentials A of the form $F(A)=0$ (Sec. 1). For infinitesimal transformations (2.47) we have

$$\delta(\omega)=\delta(F(A^{\omega}))D, \tag{2.53}$$

where $A_\mu^\omega=A_\mu+D_\mu\omega$, and the coefficient of the δ function is the functional determinant

$$D=\det\left(\frac{\partial F^a(A^\omega)}{\partial\omega^b}\right). \tag{2.54}$$

This is the Faddeev–Popov determinant.[71,78] For example, for the Lorentz gauge (1.8) we have

$$\left(\frac{\partial F}{\partial\omega}\right)^{ab}=\frac{\partial[\partial_\mu(A_\mu^a+D_\mu^{ac}\omega^c)]}{\partial\omega^b}=\partial_\mu D_\mu^{ab}. \tag{2.55}$$

We note that Eq. (2.53) is valid without any refinements only for small excitations of the fields A_μ and for small parameters ω. First, this is related to the fact that the equation $F(A^\omega)=0$ can have several solutions[73] (in fact, infinitely many[74,75]), and, second, to the fact that the expansion

$$\delta(f(x))=\sum_n\frac{\delta(x-x_n)}{|\det\partial f/\partial x|}, \tag{2.56}$$

where x_n are the zeros of the function f, involves the modulus of the determinant, whereas (2.54) involves the determinant [(2.56) contains multidimensional δ functions]. In practice, one proceeds as follows.

(1) One chooses an inhomogeneous gauge

$$F=\partial_\mu A_\mu-f=0, \tag{2.57}$$

where f is an arbitrary function.

(2) After substituting the δ function (2.53) into the integral (2.50), one integrates over f with Gaussian weight $P[f]=\text{const}\exp(-if^2/2\alpha)$. Here a term of the type (1.16) is added to the action (this is a gauge-fixing term; it is understood that all these assumptions remain true for any f).

(3) The determinant D is replaced by an integral over anticommuting scalar fields c^a, $\bar{c}^a$ (Refs. 71 and 78):

$$D=\int d[\bar{c},c]e^{-i\int dx\,\bar{c}F'c}, \tag{2.58}$$

i.e., the Lagrangian of the fictitious fields $\mathscr{L}''=\bar{c}F'c$, $F'=\partial F/\partial\omega$, is added to the original Lagrangian [see (2.55)].

After the gauge-fixing procedure the generating functional (2.50) takes the form

$$Z[\mathscr{T}]=\int d[A\bar{\psi}\psi\bar{c},c]e^{i(S+S'+s''+\mathscr{T}Q)}, \tag{2.59}$$

where the primed additions to the action correspond to the Lagrangians $\mathscr{L}'$ and $\mathscr{L}''$. The total Lagrangian in (2.59) $\mathscr{L}_T=\mathscr{L}+\mathscr{L}'+\mathscr{L}''$ is no longer invariant under gauge transformations, but it is invariant under global supersymmetry transformations.[68–70] For example, for the Lorentz gauge these transformations are

$$\delta A_\mu^a=\varepsilon D_\mu^{ad}c^d,\quad \delta\bar{c}^a=\frac{1}{\alpha}\,\varepsilon\partial_\mu A_\mu^a,$$

$$\delta c^a=-\frac{1}{2}\,\varepsilon g f^{adc}c^dc^c, \tag{2.60}$$

where ε is a coordinate-independent anticommuting parameter. Formally, the vector potentials transform in the standard way:

$$\delta A_\mu^a=D_\mu^{ab}\bar{\omega}^b,\quad \bar{\omega}^b(x)=\varepsilon c^b(x), \tag{2.61}$$

only now the gauge parameters $\bar{\omega}$ are even elements of the Grassmann algebra. It is remarkable that these transformations, which formally are global, do not weaken the restrictions imposed by local gauge invariance ($\bar{\omega}$ is a function of the coordinates).

3. MECHANICAL MODELS

3.1. Scalar electrodynamics in (0+1) spacetime

Field theory by itself is quite complicated, and therefore it is best to study the features of theories with local gauge symmetry using models. The simplest is scalar electrodynamics in $(0+1)$ spacetime,[103,104]

$$L=\frac{1}{2}(\dot{\mathbf{x}}-eyT\mathbf{x})^2-V(\mathbf{x}^2),\quad T=\begin{pmatrix}0&-1\\1&0\end{pmatrix}, \tag{3.1}$$

where the components of the vector $\mathbf{x}$ are the real and imaginary parts of the complex field φ (rather, $\varphi=(x_1+ix_2)/\sqrt{2}$), y corresponds to the zeroth component of the vector potential A_μ, and V is the potential energy. The Lagrangian (3.1) describes the motion of a particle of unit mass in the (x_1,x_2) plane, where x_1 and x_2 are the particle coordinates. In addition to these, there is yet another dynamical variable: y. The gauge transformations are given by

$$\delta\mathbf{x}=\omega eT\mathbf{x},\quad \delta y=\dot{\omega}=d\omega(t)/dt, \tag{3.2}$$

in which ω is an arbitrary infinitesimal function of time. The Lagrangian is independent of the velocity $\dot{y}$, so that there are constraints in the theory. The primary constraint is obvious: $\pi=\partial L/\partial\dot{y}=0$. Defining the momentum $\mathbf{p}=\partial L/\partial\dot{\mathbf{x}}$, we find the Hamiltonian

$$H=h+ey\mathbf{p}T\mathbf{x}+\dot{y}\pi,\quad h=\frac{\mathbf{p}^2}{2}+V(\mathbf{x}^2), \tag{3.3}$$

and from the condition $\dot{\pi}=\{\pi,H\}=0$ we find the secondary constraint

$$\sigma=e\mathbf{p}T\mathbf{x}=0. \tag{3.4}$$

The constraints are in involution, and there are no other constraints. Up to a coefficient, the constraint (3.4) is the generator of rotations in the (x_1,x_2) plane, so that the only physical variable is the invariant $r=|\mathbf{x}|$. There are two unphysical variables: y and $\theta=\arctan x_2/x_1$, $\{\theta,\sigma/e\}=1$. In spite of its simplicity, this model illustrates many of the important features of gauge theories.

1. Phase space. The physical phase space of the model is a cone unfoldable into a half-plane,[104] i.e., the presence of the unphysical degree of freedom y decisively affects such a fundamental characteristic of the Hamiltonian system as the phase space.

2. Elimination of unphysical variables in the Lagrangian. Setting $y=0$ in (3.1), we obtain a completely different model: a nongauge theory with two physical degrees of freedom. It contains no information about the secondary constraints.

3. Residual discrete gauge group. Fixing the gauge by the condition

$$x_2=0 \tag{3.5}$$

(the tensor gauge), we find that it is incomplete: there is a residual discrete gauge group Z_2 (Refs. 23 and 76), the nontrivial element of which is the transformation

$$x_1\to -x_1. \tag{3.6}$$

This statement is easily generalized to models with any reductive gauge group.[23,76] The residual group in this case is the Weyl group[105,106] (the symmetry group of the root diagram, a subgroup of the gauge group).

4. The dynamical gauge $\mathscr{L}'=-\dot{y}^2/2\alpha$. The addition of $\mathscr{L}'$ to the Lagrangian (3.1) not only specifies the dynamics of the unphysical variables, but, in general, changes the dynamics of the physical ones. In fact, the Hamiltonian now becomes

$$H=\frac{1}{2}\left(p_r^2+\frac{p_\theta^2}{r^2}\right)+V(r^2)-\frac{\alpha}{2}\,\pi^2+eyp_\theta, \tag{3.7}$$

where r and θ are polar coordinates ($r=|\mathbf{x}|$), p_r and p_θ are the canonically conjugate momenta, and $\sigma=ep_\theta$. However, θ is a cyclic variable, i.e., p_θ is a constant of the motion specified by the initial conditions. It is easy to check that for $p_\theta\neq 0$ the equations of motion of the physical variables r and p_r change [owing to the second term in the parentheses in (3.7)]. Therefore, in order that the unphysical sector not affect the physical one, the initial conditions for the unphysical variables in the dynamical gauge should be consistent with the constraints [in this case, $\pi(0)=p_\theta(0)=0$].

5. The gauge $\mathscr{L}'=-m^2y^2/2$. Sometimes $\mathscr{L}'=-(nA)^2/2\alpha$ is used as the gauge-fixing term. Although this term violates gauge invariance, it does not lift the degeneracy—the Lagrangian remains degenerate. The changes in the dynamics which occur are best explained by using the model (3.1) with the additional term $\mathscr{L}'=-m^2y^2/2$. We immediately stress the fact that this is not a dynamical gauge—the velocity $\dot{y}$ also does not enter into the new Lagrangian, i.e., as before, $\pi=\partial\mathscr{L}/\dot{y}=0$ (the primary constraint). The Hamiltonian has the form

$$H=h+\frac{1}{2}\,m^2y^2+ey\mathbf{p}T\mathbf{x}+u\pi, \tag{3.8}$$

where u is an arbitrary function of time. From the condition $\dot{\pi}=\{\pi,H\}=0$ we find the secondary constraint

$$\dot{\pi}=\tilde{\sigma}=-(m^2y+e\mathbf{p}T\mathbf{x})=0 \tag{3.9}$$

The constraints π and $\tilde{\sigma}$ do not occur in involution

$$\{\pi,\tilde{\sigma}\}=-m^2, \tag{3.10}$$

i.e., these are second-class constraints. According to Dirac,[96] in this case it is necessary to change the Poisson brackets to Dirac brackets. The self-consistency condition $\{\pi,\tilde{\sigma},\}=-m^2u=0$ now is a condition on the Lagrange multiplier $u(t)$. Eliminating the variables y and π from H, we obtain

$$H=\frac{1}{2}\left(p_r^2+\frac{p_\theta^2}{r^2}\right)+V(r^2)-\frac{e^2}{2m^2}\,p_\theta^2, \tag{3.11}$$

i.e., as in the case (3.7), p_θ is a constant of the motion. The inappropriate choice of initial conditions ($p_\theta\neq 0$) changes the dynamics in the physical sector. The only value of p_θ consistent with the condition (3.4) is $p_\theta(0)=0$.

6. Quantization. The model (3.1) allows the simplest illustration of the thesis that in theories with first-class constraints the operations of elimination of unphysical variables and quantization, in general, do not commute.[107] In fact, the unphysical momenta in (3.3) are π and p_θ. Setting them equal to zero, we obtain

$$H'_{ph}=\frac{1}{2}\,p_r^2+V(r^2). \tag{3.12}$$

The Hamiltonian H'_{ph} describes the one-dimensional motion of a particle in the field of a potential V. The quantization recipe is standard: $r,p_r\to\hat{r},\hat{p}_r$, $[\hat{r},\hat{p}_r]=i$, i.e.,

$$\hat{H}'_{ph}=-\frac{1}{2}\,\partial_r^2+V. \tag{3.13}$$

On the other hand, quantizing the model up to the exclusion of the unphysical variables, we have

$$\hat{H}_{ph}=-\frac{1}{2}\left[\partial_r^2+\frac{1}{r}\,\partial_r+\frac{1}{r^2}\,\partial_\theta^2\right]+V-iey\partial_\theta-iu\partial_y. \tag{3.14}$$

The physical sector is specified by the conditions $\hat{\pi}\Phi=\hat{\sigma}\Phi=0$, i.e., $\partial_y\Phi=\partial_\theta\Phi=0$, and on vectors from the physical Hilbert space the Hamiltonian (3.14) takes the form

$$\hat{H}_{ph}=-\frac{1}{2}\left(\partial_r^2+\frac{1}{r}\,\partial_r\right)+V. \tag{3.15}$$

The Hamiltonians (3.13) and (3.15) are not identical. This must be taken into account when substituting the δ functions from the eliminated variables into the path integral (see Sec. 4).

7. The inconsistent gauge $\mathscr{L}'=yt$. Although it violates gauge invariance, this gauge-fixing term leads to inconsistent dynamics. We have $\pi=\partial L/\partial\dot{y}=0$,

$$H=h+ey\mathbf{p}T\mathbf{x}+u\pi-yt. \tag{3.16}$$

The self-consistency condition $\dot{\pi}=\{\pi,H\}\equiv\Sigma=0$ gives

$$\Sigma=-e\mathbf{p}T\mathbf{x}+t. \tag{3.17}$$

Since Σ explicitly involves the time, the consistency condition (3.17) with dynamics specified by the Hamiltonian (3.16) must now be formulated as

$$\frac{d\Sigma}{dt}=\frac{\partial\Sigma}{\partial t}+\{\Sigma,H\}=0. \tag{3.18}$$

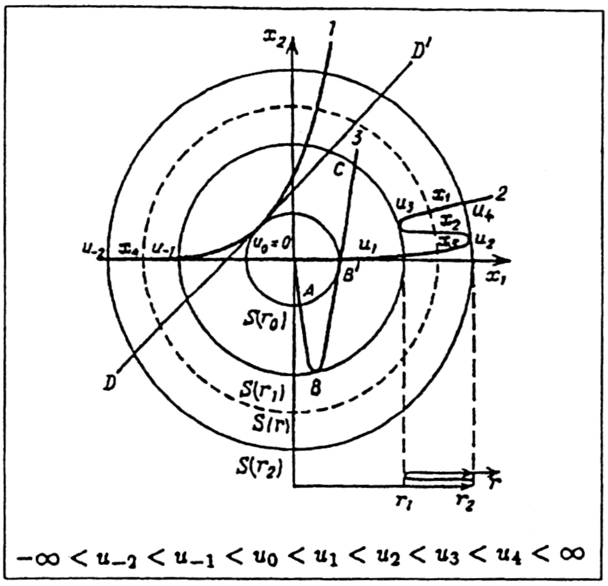

FIG. 1.

However, $\{\Sigma,H\}=0$, i.e., (3.18) reduces to the absurd requirement $1=0$.

8. Nonunique (incomplete) gauges. The model (3.1) helps to explain the "copy problem" (Ref. 73). When the conditions (1.8) and (1.9) were applied to non-Abelian theories, it was found that they do not completely eliminate the gauge freedom.[73–75] At one time a fundamental meaning was attributed to this. In fact, the problem is not a physical one—it originates not in the features of the dynamics of the physical sector of Yang–Mills fields, but in the choice of gauge; its origin thus lies completely outside physics. This is already suggested by the fact that there exist unique gauges, for example, the Weyl gauge (1.10), the Fock gauge (1.14), axial gauges, and so on. The copy problem can arise even in electrodynamics. To explain the situation we turn to Fig. 1 from Ref. 23. The gauge orbits in this figure (sets of points connected by a gauge transformation) are circles. The gauge condition $f(x_1,x_2)=0$, where f is some function, specifies a line in the plane. Since points lying on the same orbit are physically indistinguishable, one point of each orbit is sufficient for describing the evolution of the physical degrees of freedom. This is what the gauge condition does. However, it may happen that for an inappropriate choice of f the curve specified by the gauge condition will intersect the orbits two or more times. For example, the line $f=x_2=0$ intersects all the orbits twice. This is an incomplete gauge; there is a residual freedom specified by the discrete gauge group Z_2 with elements $(1,P)$, $Px_1=-x_1$. Curve 2 intersects some orbits four times, and curve 1 does not intersect any orbits near zero. This is an insufficient (incomplete) gauge—the freedom of the variables x_1, x_2 for $\sqrt{x_1^2+x_2^2}<r_0$ is not eliminated (see Ref. 23 for more detail).

Thus, when applying a gauge condition it is necessary to check that it is consistent and does not change the dynamics of the physical variables.

3.2. A model with a gauge group of translations

Another instructive model is that corresponding to the Lagrangian

$$L=\frac{1}{2}(\dot{\mathbf{x}}-eyT\mathbf{x})+\frac{1}{2}[(\dot{x}_3-y)^2+(\dot{x}_4-y)^2]$$
$$-V(\mathbf{x}^2,x_3-x_4), \tag{3.19}$$

where $\mathbf{x}=(x_1,x_2)$. Its invariance group is specified by the infinitesimal transformations

$$\delta\mathbf{x}=\omega eT\mathbf{x},\quad \delta y=\dot{\omega},\quad \delta x_3=\delta x_4=\omega,\quad \omega\to 0. \tag{3.20}$$

The distinguishing feature of the model is the fact that for the variable $2y_+=x_3+x_4$ the gauge transformations reduce to translations, where $2y_-=x_3-x_4$ is a gauge invariant. The primary constraint is $\pi=\partial L/\partial\dot{y}=0$; the Hamiltonian is

$$H=\frac{\mathbf{p}^2}{2}+\frac{1}{2}(p_3^2+p_4^2))+V(\mathbf{x}^2,x_3-x_4)$$
$$+y(p_3+p_4+e\mathbf{p}T\mathbf{x})+\dot{y}\pi, \tag{3.21}$$

i.e., the secondary constraint is

$$\sigma=\{\pi,H\}=-(p_3+p_4+e\mathbf{p}T\mathbf{x})=0. \tag{3.22}$$

The model can be used to explain the physical consequences of eliminating some degree of freedom. In the model (3.1) the variables x_1 and x_2 are transformed into each other by a gauge transformation, so that it makes no difference which variable we eliminate. In this model the condition (3.22) relates three momenta (p_3, p_4, and p_θ). The questions are, which of these should be eliminated, and are the alternatives (physically) equivalent?

It is easy to construct the invariant variables. They are

$$|\mathbf{x}|,\quad \mathbf{z}_+=e^{-y_+eT}\mathbf{x}\quad \mathbf{z}_{3,4}=e^{-y_{4,3}eT}\mathbf{x},$$

$$y_-=\frac{x_3-x_4}{2},\quad y_{3,4}=x_{3,4}-e^{-1}\arctan\frac{x_2}{x_1}. \tag{3.23}$$

The variables $\mathbf{z}_{3,4,+}$ are analogous to the nonlocal field ψ' (2.36). The exponential describes the Coulomb field of a charged particle, i.e., a quantum of this field corresponds to a charged particle plus the electrostatic field surrounding it. The invariance of the variables (3.23) is obvious—it follows from the transformation laws

$$\mathbf{x}'=e^{eT\Lambda}\mathbf{x},\quad x_{3,4}'=x_{3,4}+\Lambda, \tag{3.24}$$

where Λ is an arbitrary function of time. Guided by the analogy with electrodynamics, we conclude that the vector $\mathbf{x}$ can be associated with three variables: x_3, x_4, or y_+. The corresponding sets of invariant variables are

$$\{y^3,y_4,|\mathbf{x}|\},\quad \{y_4,\mathbf{z}_3\},\quad \{y_3,\mathbf{z}_4\},\quad \{y_-,\mathbf{z}_+\}. \tag{3.25}$$

These possibilities are *a priori* equivalent. However, in the real world for some reason only one of them has been realized. We see from (3.25) that the "physical particles" described by the variables $\mathbf{z}_{3,4,+}$ possess the "accompanying fields" $\exp(-y_{4,3,+}eT)$, and they are all different. It is instructive to write out the eliminated variables and the corresponding Hamiltonians:

$$\theta \quad \frac{1}{2}\left[p_r^2+\frac{1}{2}\frac{(p_3+p_4)^2}{e^2r^2}+p_3^2+p_4^2\right]+V(r^2,x_3-x_4)$$

$$x_{3,4} \quad \frac{1}{2}\left[\mathbf{p}^2+p_{4,3}^2+(p_{4,3}+ep_\theta)^2\right]+V(\mathbf{x}^2,\mp y_{4,3})$$

$$x_3+x_4 \quad \frac{1}{2}\left[\mathbf{p}^2+\frac{1}{2}(p_3-p_4)^2+\frac{e^2p_\theta^2}{2}\right]+V(\mathbf{x}^2,2y_-). \tag{3.26}$$

It is not difficult to go to the quantum description. Vectors from the physical Hilbert subspace are isolated by requiring that the constraints vanish on them. The Hamiltonian operators are obtained from (3.26) by changing to canonical operators [except in the first, where the operator $(i/r)\hat{p}_r$ must be added to p_r^2], and the physical wave functions will depend on the corresponding sets of physical variables: r, y_3, y_4; $\mathbf{z}_{3,4,+}$, $y_{3,4,-}$. Therefore, in this model all the possibilities are equivalent. In the real world the choice is made by the physics. An example is the case of photons in a superconductor above and below the critical point,[40] when either the longitudinal component $\mathbf{A}_\parallel$ is associated with the charged field, forming the Coulomb factor [see (2.36)], or the phase of the charged field together with the component $\mathbf{A}_\parallel$ form a massive longitudinal vector field.

Remark. Obviously, the model is easily generalized to any semisimple group G. Let us consider the Lagrangian

$$L=\tfrac{1}{2}\mathrm{Tr}(\dot{x}-yx)^2+\frac{1}{2}\sum_{a,k}(\dot{z}_k^a-y^a)^2$$
$$+\sum_{i>k}V(\mathrm{Tr}x^2,(z_i^a-z_k^a)^2), \tag{3.27}$$

where $\mathrm{Tr}x^2=\Sigma x_a^2$, $\mathrm{Tr}y^2=\Sigma y_a^2$, and $\mathrm{Tr}z_k^2=\Sigma(z_k^a)^2$, $a=1,2,\ldots,\dim G$, i.e., x, y, and z are elements of the Lie algebra of the group G in the fundamental representation, and k are natural numbers, $0<k<\infty$. It is invariant under infinitesimal transformations:

$$\delta x=\omega x,\quad \delta z_k=\omega,\quad \delta y=\dot{\omega},\quad \omega\to 0. \tag{3.28}$$

The model (3.19) with $e=0$ has been studied by Burnel.[108,109]

The problem of non-normalizability of physical states. One of the problems of quantum electrodynamics (QED) was that of the non-normalizability of physical states. Although a logically faultless formulation of the foundations of QED can be found already in the fundamental studies of Heisenberg and Pauli (Refs. 32 and 33; see also Refs. 34–37), it was this problem that led to the studies by Gupta and Bleuler[110–112] postulating the existence of unphysical states (vectors with zero norm) in the physical sector. The model that we are considering provides an explanation of this. Let us consider the special case

$$L=\frac{1}{2}\dot{\mathbf{x}}^2-V(\mathbf{x})+\frac{1}{2}(\dot{z}-y)^2. \tag{3.29}$$

The variables are $\mathbf{x}=(x_1,x_2)$, y, and z. The variable y is unphysical, because L is independent of the velocity $\dot{y}$, $\pi=\partial L/\partial\dot{y}=0$. The Hamiltonian is ($p_z=\partial L/\partial\dot{z}=\dot{z}-y$)

$$H=h+\frac{1}{2}p_z^2+yp_z; \tag{3.30}$$

from the self-consistency condition we find $\dot{\pi}=\{\pi,H\}=-p_z=0$, i.e., z is also an unphysical variable. In the quantum theory all the variables entering into the Lagrangian (3.29) become operators. According to (2.49), the first-class constraints must vanish on physical vectors Φ:

$$\hat{\pi}\Phi=-i\frac{d\Phi}{dy}=0,\quad \hat{p}_z\Phi=-i\frac{d\Phi}{dz}=0, \tag{3.31}$$

i.e., all the physical wave functions are independent of y and z. However, when normalizing the physical states it is necessary to integrate over all the variables $\mathbf{x}$, y, and z. Since each of the latter two varies along the entire real axis, the physical wave functions turn out to be non-normalizable:

$$\int d^2x\,dy\,dz|\Phi(\mathbf{x})|^2=\infty. \tag{3.32}$$

This non-normalizability clearly has nothing to do with physical processes in the (x_1,x_2) plane, so that it makes no sense to require that Φ be normalizable in the y, z subspace. In this case the problem is simple to solve: these variables must be ignored. The wave functions do not depend on them, so that it is also not necessary to integrate over them in the normalization conditions. After the Hamiltonian structure of the theory is understood they can be forgotten. However, we stress the fact that the situation is this simple only in this model. Gauge field theories possess Lorentz invariance, which leads to mixing of the physical and unphysical components of the vector potentials, so that it is necessary to be more clever.

Gauges of invariant form. For the model of (3.19) let us consider the gauges [see (3.24); $\theta=e\Lambda$]

$$x_2'=x_2\cos\theta-x_1\sin\theta=0 \tag{3.33}$$

and

$$x_4'=x_4+\Lambda=0. \tag{3.34}$$

For the variables x_1' and x_3' we have

$$x_1'=x_1\cos\theta+x_2\sin\theta,\quad x_3'=x_3+\Lambda. \tag{3.35}$$

Substituting into the first of these the solution of (3.33), $\theta=\arctan x_2/x_1$, and into the second the solution of (3.34), $\Lambda=-x_4$, we find

$$x_1'=x_1\cos\arctan\frac{x_2}{x_1}+x_2\sin\arctan\frac{x_2}{x_1}=\sqrt{x_1^2+x_2^2}=|\mathbf{x}|, \tag{3.36}$$

$$x_3'=x_3-x_4=2y_-. \tag{3.37}$$

However, the variables $|\mathbf{x}|$ and y_- are gauge-invariant. We have already encountered a similar situation in electrodynamics: the variable ψ' (2.36) was obtained by going to the Coulomb gauge (see also Ref. 23).

4. GAUGE CONDITIONS

4.1. Nondynamical gauges

The Weyl ($A_0=0$), *Maxwell* ($\boldsymbol{\partial}\mathbf{A}=0$), *and Lorentz* ($\partial_\mu A_\mu=0$) *gauges.*

Let us analyze the role of the physical degrees of freedom for a particular gauge choice. We shall not consider all gauges. Some of them are rarely used, and the situation regarding others is not completely clear. The Weyl gauge (1.10) is natural from the viewpoint of the physical content of the theory. The velocity $\dot{A}_0$ does not appear in the Lagrangian (2.2), so that the variable A_0 is certainly unphysical. This is also true for non-Abelian theories: the Yang–Mills Lagrangian is also independent of $\dot{A}_0^a$. The total Hamiltonian (2.14) [or (2.33), taking into account (2.28)] gives $\dot{A}_0=u(x)$, where u is an arbitrary function, i.e., A_0 is an arbitrary function and can be set equal to zero. Furthermore, in the free theory $\boldsymbol{\partial}\mathbf{E}=0$, so that $\mathbf{A}_\parallel$ is also an unphysical variable and can also be set equal to zero: div $\mathbf{A}=0$, i.e., the field $\mathbf{A}$ is transverse. If the theory were free, the analysis would end here if Lorentz transformations did not mix the physical and unphysical components. But help comes from the requirement of weak gauge invariance of the physical sector under the extended group of gauge transformations (2.16): a suitable choice of the functions u and v makes it possible for A_0 and $\mathbf{A}_\parallel$ to vanish in any reference frame. The relativistic invariance and locality of the theory are thus obtained (see also Refs. 102 and 113). This is the reason why unphysical degrees of freedom were allowed in the formalism.

The situation is somewhat changed when an interaction is switched on. Now the secondary constraint is $\sigma=\boldsymbol{\partial}\mathbf{E}-j_0=0$. This condition states that it is meaningless to consider the electric charge and the longitudinal electric field separately—it is sufficient to consider only one. Usually the charge (charged field) is chosen. It can be shown that here along with $A_0=0$ it is impossible to require that $\mathbf{A}_\parallel=0$ (i.e., div $\mathbf{A}=0$), since in this case we have $j_0=\boldsymbol{\partial}\mathbf{E}=\boldsymbol{\partial}(\mathbf{A}-\boldsymbol{\partial}A_0)=0$ (Ref. 114). Actually, a similar problem also occurs in the free theory: if u and v are arbitrary, then, in general, $\delta\boldsymbol{\partial}\mathbf{E}=\Delta(u-\dot{v})\neq 0$, i.e., the secondary constraint is violated. See the end of this subsection for a discussion of these questions.

Now let us make the gauge transformation $A_\mu\to A'_\mu$ (1.1), requiring that $\boldsymbol{\partial}\mathbf{A}'=0$, i.e., $\boldsymbol{\partial}\mathbf{A}+\Delta\Lambda=0$. We find

$$\Lambda=-\Delta^{-1}\boldsymbol{\partial}\mathbf{A}. \tag{4.1}$$

Clearly, Λ now contains information about the longitudinal electromagnetic field $\mathbf{A}_\parallel$. However, $A'_0=A_0+\dot{\Lambda}$, i.e., this information is transferred to the zeroth component of A'_μ. Setting $A_0=0$ and writing $A'_0=\phi$, we find $\phi=-\Delta^{-1}\boldsymbol{\partial}\dot{\mathbf{A}}=-\Delta^{-1}\boldsymbol{\partial}\mathbf{E}$. We have actually gone from the Weyl gauge to the radiation gauge. Using Gauss's law, we obtain the usual expression for the Coulomb potential:

$$\phi=-\Delta^{-1}j_0. \tag{4.2}$$

The charged field ψ also is changed: it acquires the factor $\exp(-ie\Delta^{-1}\boldsymbol{\partial}\mathbf{A})$ corresponding to the Coulomb field. This calculation demonstrates how an unphysical variable ($A'_0=\phi$) can contain physical information: information is transferred to the unphysical component by the gauge transformation. This explains why A_0 is usually associated with the Coulomb potential.

The Lorentz gauge is usually used in classical electrodynamics. Its advantages are its relativistic invariance and completeness (the remaining freedom reduces to that of the initial conditions; see Sec. 7.1.). According to (2.2), (2.6), and (2.8), the equations of motion

$$\Box A_\mu+\partial_\mu(\partial_\nu,A_\nu)=-j_\mu \tag{4.3}$$

in this gauge reduce to the Lorentz equations[24]

$$\Box A_\mu=-j_\mu. \tag{4.4}$$

The unphysical degree of freedom is eliminated from the theory in a relativistically invariant way. Making the decomposition

$$A_\mu=A_\mu^{\mathrm{tr}}+A_\mu^l,\quad \partial_\mu A_\mu^{\mathrm{tr}}=0, \tag{4.5}$$

we have $\partial_\mu A_\mu=\partial_\mu A_\mu^l$. Obviously, $A_\mu^l=-\Box^{-1}\partial_\mu(\partial_\nu A_\nu)$, i.e., A_μ^l is a purely gradient field and can be eliminated by a gauge transformation. It follows from (4.4) that it does not interact with the charged fields: $\Box A_\mu^l=0$, since $\partial_\mu j_\mu=0$. Therefore, it is neither emitted nor absorbed. The propagator of the electromagnetic field in this gauge

$$\Delta_{\mu\nu}^c=\frac{-i}{q^2+i0}\left(g_{\mu\nu}-\frac{q_\mu q_\nu}{q^2}\right) \tag{4.6}$$

has an additional pole which, however, does not cause any problem, owing to current conservation. In the static case ($\dot{A}_0^{\mathrm{tr}}=0$) the field A_0^{tr} satisfies the Poisson equation, the solution of which (4.2) is the Coulomb potential.

The changeover to the Lorentz gauge is accompanied by a change of the phase of the charged field: $\partial_\mu A'_\mu=\partial_\mu A_\mu-\Box\Lambda=0$, i.e., $\Lambda=\Box^{-1}\partial_\mu A_\mu$, $\psi\to\psi'=\exp(ie\Box^{-1}\partial_\mu A_\mu)\psi$. Thus, the field $\partial_\mu A_\mu$ present in the Lagrangian (2.2) is eliminated from the dynamics by the gauge condition. It is associated with the charged field ψ, i.e., its dynamics is now determined by that of ψ'. In the case of static fields ($\dot{A}_\mu=0$), the exponential in ψ' is transformed into the Coulomb exponential (2.36).

Now we are ready for a more detailed discussion about the simultaneous elimination of the fields A_0 and $\mathbf{A}_\parallel$ by means of transformations from the extended gauge group. This is closely related to another question: are the gauge parameters u and v completely arbitrary? This is equivalent to asking whether violation of first-class constraints is allowed (before going to the physical sector), in particular, by transformations from the extended gauge group, since $\delta(\boldsymbol{\partial}\mathbf{E}-j_0)=\Delta(u-\dot{v})\neq 0$. The Lagrangian contains all the components of A_μ, both physical and unphysical, and, as we have seen in Sec. 3, they are all important for the physics. It is known that the theory can be formulated in terms of only physical quantities which are invariant under transformations from the extended gauge group, namely, ψ' (2.36) and $\mathbf{A}_\perp$: $\delta_{u,v}\psi'=\delta_{u,v}\mathbf{A}_\perp=0$. The corresponding Hamiltonian of scalar electrodynamics is[95]

$$H_{ph}=\int d^3x\left\{\frac{1}{2}[\mathbf{E}_\perp^2+\mathbf{H}^2-j_0\Delta^{-2}j_0]+\frac{1}{2}\Pi^2\right.$$
$$\left.+\frac{1}{2}[(\boldsymbol{\partial}+e\mathbf{A}_\perp T)\mathbf{\Phi}']^2\right\},\qquad (4.7)$$

where the components of the two-dimensional vector $\mathbf{\Phi}$ are the real and imaginary parts of the complex scalar field φ: $\varphi=(\varphi_1+i\varphi_2)/\sqrt{2}$, Π is the canonical momentum of the scalar field, the matrix T is given in (3.1), and the primed field $\mathbf{\Phi}'$ differs from $\mathbf{\Phi}$ by the Coulomb exponential:

$$\Phi'(x)=\exp\left(-e\int\frac{d^3y}{4\pi}\frac{\boldsymbol{\partial}\mathbf{A}(y)}{|\mathbf{x}-\mathbf{y}|}T\right)\Phi(x).\qquad (4.8)$$

Thus, in the physical sector, i.e., when all the first-class constraints are taken into account, the theory is formulated in terms of only transverse fields:

$$A_0=\mathbf{A}_\parallel=0.\qquad (4.9)$$

Its distinguishing feature is the nonlocality of all the physical fields [Φ' is given by (4.9), ψ' by (2.36), and $\mathbf{A}_\perp=-\Delta^{-1}\mathrm{curl}\,\mathbf{H}$]. We have already taken into account the constraints in the Hamiltonian (4.7), so that despite the condition (4.9) it is impossible to speak of violation of Gauss's law, i.e., of a contradiction between the requirements (4.9) and the constraints. We stress the fact that the field $\mathbf{A}_\parallel$ vanishes only in this representation. From the viewpoint of the initial (local) variables Φ and A_μ, it has not vanished: the longitudinal field has gone into the exponential in (4.8). This implies that to follow the fate of $\mathbf{A}_\parallel$ it is necessary to solve the equations of motion for the field Φ'.

Thus, the unphysical degrees of freedom are the quantities canonically conjugate to all the first-class constraints (their Poisson brackets with all the other canonical variables must vanish). The unphysical variables can be fixed by means of gauge transformations: they can either be set equal to zero, or the information contained in the physical degrees of freedom can be transferred to them. Using the fact that the physical sector is independent of the unphysical variables, the latter can be changed arbitrarily. In particular, the generators of gauge transformations (the constraints) can be changed, which is what happens in, for example, the Feynman gauge: $\pi_0=\partial_\mu A_\mu\neq 0$. Here it is important only that $\pi_0=0$ in the physical sector. Using the full gauge freedom, i.e., the extended group of transformations, we act in the space which includes both the physical and the unphysical variables. Therefore, at this stage it is impossible to speak of the violation of any conditions (equations of motion, constraints)—it is only important that these operations do not affect the physical sector, to which it is possible to go at any time by taking into account the constraints.

Let us summarize: before going to the physical sector it is possible to use transformations from the extended gauge group with arbitrary parameters u and v.

The axial gauge $A_3=0$. Axial gauges are of particular interest because their use makes it unnecessary to resort to fictitious fields [the matrix in (2.54) is independent of the fields], so that the copy problem does not arise. When the third component of A_μ is fixed,

$$A_3'=A_3+\partial_3\Lambda=0,\qquad (4.10)$$

the information contained in it is transferred to the other components, because

$$\Lambda=\int_{x_3}A_3dx^3.\qquad (4.11)$$

The following facts should be noted. Not only the relativistic invariance, but also the isotropy of 3-dimensional space are violated explicitly. Furthermore, even for fields A_μ which fall off at infinity ($A_3\sim 1/x_3$, $|x_3|\to\infty$), the function Λ does not fall off: $\Lambda\sim\ln|x_3|$, $|x_3|\to\infty$. Obviously, the use of this gauge is correct only in problems with rapidly decreasing potentials. Finally, the parameters of the residual gauge group Λ', $\partial_3\Lambda'=0$, depend on the time. Therefore, the gauge freedom has not been completely eliminated—there remain unphysical degrees of freedom which can vary arbitrarily with time. Of course, $\Lambda'=\Lambda'(x_1,x_2,t)$ depends only on two coordinates, i.e., the set of unphysical variables is specified by two parameters and from the viewpoint of the original (three-dimensional) set has measure zero. In the extended gauge group the transformation (4.10) is generated by a secondary constraint, i.e., $\Lambda=v$.

The propagator of the electromagnetic field in the axial gauge[31]

$$\Delta_{\mu\nu}^c=\frac{-i}{q^2+i0}\left(g_{\mu\nu}-\frac{q_\mu n_\nu+n_\mu q_\nu}{qn}+\frac{q_\mu q_\nu}{(qn)^2}n^2\right)\qquad (4.12)$$

has a pole at $qn=0$, i.e., the fields propagating in the plane normal to n are completely correlated.

The causal function can look rather complicated when contrived gauges are used. There is a simple criterion which must be satisfied by the propagator of the electromagnetic field (it is natural for dynamical gauges, but useful also for nondynamical ones, except for the Weyl gauge): photon exchange by static sources must reproduce the Coulomb potential. Let us consider electrodynamics with static sources J_μ (Ref. 115):

$$S=-\int dx\left[\frac{1}{4}F_{\mu\nu}^2+A_\mu J_\mu\right].\qquad (4.13)$$

Substitution of this action into (2.50) gives

$$W[J]=\frac{i}{2}J_\mu\Delta_{\mu\nu}^cJ_\nu,\qquad (4.14)$$

where $\Delta_{\mu\nu}^c$ is the photon propagator. The inclusion of currents changes the energy of the "vacuum" (when currents are included, it becomes the ground state). In the case of static sources we have

$$\Delta E_{\rm vac}T=-W[J],\qquad (4.15)$$

where $\Delta E_{\rm vac}$ is the shift of the vacuum energy, and T is the time interval: $T=t_2-t_1\to\infty$. Since $J_\mu(x)\to J_0(x)$, $\dot{J}_0=0$, and in momentum space $\tilde{J}_0(q)=2\pi\delta(q_0)\tilde{J}_0(\mathbf{q})$, we find

$$\Delta E_{\rm vac}T=-\frac{i}{2}\int\frac{dq}{(2\pi)^4}\tilde{J}_0(q)\Delta_{00}^c(q)\tilde{J}_0(-q)$$

$$= \frac{2\pi\delta(0)}{2i}\int \frac{d^3q}{(2\pi)^3}\tilde{J}_0(\mathbf{q})\Delta^c_{00}(\mathbf{q})\tilde{J}_0(-\mathbf{q}), \tag{4.16}$$

i.e., owing to the replacement $2\pi\delta(0)\to T$,

$$\Delta E_{\text{vac}} = -\frac{i}{2}\int \frac{d^3q}{(2\pi)^3}\tilde{J}_0(\mathbf{q})\Delta^c_{00}(\mathbf{q})\tilde{J}_0(-\mathbf{q}). \tag{4.17}$$

On the other hand, going to the Hamiltonian in the physical sector (2.35), we obtain

$$\Delta E_{\text{vac}} = -\frac{1}{2}J_0\Delta^{-1}J_0 = \frac{1}{2}\int \frac{d^3q}{(2\pi)^3}\tilde{J}_0(\mathbf{q})\frac{1}{\mathbf{q}^2}\tilde{J}_0(-\mathbf{q}). \tag{4.18}$$

Comparing (4.17) and (4.18), we obtain the desired criterion:

$$-i\Delta^c_{00}(q)|_{q_0=0} = \frac{1}{\mathbf{q}^2}. \tag{4.19}$$

This example again demonstrates the specific features of the unphysical degrees of freedom. It would seem that by choosing the gauge $A_0=0$ in the case $\mathbf{J}=0$ we have eliminated the interaction in (4.13). However, no variables, even unphysical ones, can be set equal to zero in the Lagrangian. The role of A_0 is to generate a secondary constraint (Gauss's law), indicating the appearance of a static electric (i.e., longitudinal) field near a charge at rest. Longitudinal fields cannot propagate away from charges; in fact, one can say that they are confined. However, it is possible (for example, by choosing the Feynman gauge; see below) to make A_0 and $\mathbf{A}_\parallel$ satisfy the Lorentz equations of motion (4.4), with the net effect of the exchange of quanta of these fields being the Coulomb potential.

The Fock gauge. Fock arrived at the gauge (1.15) when studying the motion of a relativistic charged particle in an external electromagnetic field. The action

$$S = -\frac{m}{2}\int\left[\dot{x}^2 + 1 + \frac{e}{m}A_\mu\dot{x}^\mu\right]d\tau, \quad \dot{x} = \frac{dx}{d\tau}, \tag{4.20}$$

where τ is an invariant parameter, leads to the same equations of motion as the standard action with Lagrangian

$$L_0 = -m\sqrt{1-\mathbf{v}^2} - A_\mu J_\mu, \quad \mathbf{v} = \frac{d\mathbf{x}}{dt}, \tag{4.21}$$

if $\dot{x}^2=1$, i.e., if τ is the "proper time." The interaction of a classical particle with the current $J_\mu = e\,dx_\mu/dt$ in (4.21) is the result of going to massive fields in the Lagrangian (2.2): $m\to\infty$, $\bar{\psi}\gamma^\mu\psi\to dx^\mu/dt\,\delta(\mathbf{x}-\mathbf{x}(t))$.

The application of this idea to the relativistic electron made it possible to find a solution to the Dirac equation in the presence of an external field.[28,29] Performing the gauge transformation with the function

$$\Lambda = \int_x^{x_0} A_\mu dx^\mu, \tag{4.22}$$

we find that the new potentials $A' = A + \partial\Lambda$ satisfy the equation

$$(x-x_0)^\mu A'_\mu(x) = 0, \tag{4.23}$$

if in (4.22) the integration runs along the line joining the points x and x_0. This gauge is convenient in the case of massive particles or particles of high energy, when the emission of photons which are not too hard has little effect on the particle momentum, i.e., when the "infrared approximation" is applicable. This is why it is used in studying hard processes, in particular, in studying deep-inelastic scattering processes.[60,116] This also explains its relation to contour gauges.[60] The Fock gauge is remarkable because it allows the potentials to be expressed in terms of the field strengths:

$$A_\mu(x) = x_\nu\int_0^1 sF_{\mu\nu}(sx)ds. \tag{4.24}$$

The detailed proof of this can be found in Ref. 30. Equation (4.24) is valid also in non-Abelian theories. Incidentally, we note that the gauge invariance of the potentials in (4.24) is illusory. It is possible to add to A_μ the term $\partial_\mu\Lambda$ while preserving $F_{\mu\nu}$ but violating the gauge. Equation (4.24) is essentially a special case (the Fock gauge) where the noninvariant object takes an invariant form. See Sec. 7.3 regarding the relation between the representation (4.24) and exterior differential forms.

The relation between the Fock gauge and other gauges. The Fock gauge acts as a center, relating many physically important gauges.

The Maxwell gauge. It was shown in Refs. 99 and 117 that the exponential factor in (2.36) describing the Coulomb field is an infinite product of linear exponentials:

$$\prod_{i,j}^{N}\exp\left(-ie\int_{-\infty}^{x}A_\mu(y_{ij})dy^\mu_{ij}\right)\to e^{-ie\Delta^{-1}}, \quad N\to\infty. \tag{4.25}$$

The indices i,j label the area elements on the unit sphere surrounding the charge through which lines leaving the charge [the integration contours in (4.25)] pass. Clearly, in the case of a static source the Fock gauge (1.14) is equivalent to the Maxwell gauge (1.9), associating the longitudinal field $\mathbf{A}_\parallel$ with the matter field. Independently of the representation (4.25), this can be seen directly from Eq. (4.8). Integrating by parts in the argument of the exponential, we write it in the form

$$\exp\left(e\int\frac{d^3y}{4\pi}\frac{(\mathbf{x}-\mathbf{y})\mathbf{A}}{|\mathbf{x}-\mathbf{y}|^3}T\right). \tag{4.26}$$

Obviously, as in the case of the radiation gauge, in the problem with a fixed source the requirement (1.15) associates the longitudinal field $\mathbf{A}_\parallel$ with the charge.

Other gauges. The Fock gauge is closely related to gauges using linear exponentials. It is useful in problems involving line integrals of vector potentials, i.e., integrals of 1-forms. A typical example is the physics of infrared radiation. Infrared photons are generated by a classical current,[118,119] and they can be taken into account by solving the problem with the action (4.13). The motion of a point charge is associated with the current

$$J_\mu(p,x) = \frac{ep_\mu}{E_p}\delta\left(\mathbf{x}-\frac{\mathbf{p}}{E_p}t\right), \quad p_\mu = \frac{dx_\mu}{ds}m = \frac{dx_\mu}{dt}E_p, \tag{4.27}$$

where p_μ is the particle momentum, E_p is the particle energy, and $p^2=m^2$. A typical factor describing the field of the emitted photons in the Feynman gauge is given by the exponentials

$$\exp\left(i\int_0^\infty dt\int d^3xA_\mu^\Delta J_\mu\right)=\exp\left(ie\int_0^\infty dtA_\mu^\Delta(n_p^\mu t)n_p^\mu\right)$$
$$=\exp\left(ie\int_{n_p}A_\mu^\Delta(x)dx^\mu\right). \quad (4.28)$$

Here $n_p^\mu=p^\mu/E_p=dx^\mu/dt$; the integration in the last term in (4.28) runs along the line specified by the vector n_p. The superscript Δ on A_μ indicates that undetected photons in the energy range $\Delta=E-\varepsilon$ have been taken into account, where E is the resolution of the apparatus (only photons of energy $\omega_q>E$ are detected); ε^{-1} characterizes the minimum size of the region in which the emitting particle is located and which has total charge equal to zero (photons with wavelength $\lambda_q>\varepsilon^{-1}$ are not emitted). The exponentials (4.28) describe the infrared-photon field of the scattered particle; changing over to the field of the incident particle requires a change of sign of the argument of the exponential. These expressions are also valid when the energies of the emitted photons are not small. For them to be applicable it is only necessary that the emission of quanta has a negligible effect on the momentum of the charged particle, i.e., they are valid for $|\mathbf{p}|\to\infty$. This shows why the Fock gauge is related to the gauges used to describe hard processes. Since $n_p^2=m^2/E_p^2$, then $n_p^2\to 0$ for $E_p\to\infty$, and we arrive at the light-cone gauge. It is similar to the Lipatov gauge[45] [see Eq. (4.39) below]. The momenta of the photon (q) and of the proton (p) involved in the deep-inelastic scattering process are written as $q=q'+p'q^2/s'$, $p=p'+q'm^2/s'$, where $q'^2=p'^2=0$ and $s'=2p'q'$; the arbitrary momentum k is given by the sum $k=\alpha q'+\beta p'+k_\perp$, $q'k_\perp=p'k_\perp=0$. In this approach the propagator of a vector particle (gluon) with momentum k,

$$\Delta_{\mu\nu}^c=\frac{-i}{k^2+i0}\left(g_{\mu\nu}-\frac{q'_\mu k_\nu+k_\mu q'_\nu}{q'k}\right), \quad (4.29)$$

is identical to the propagator in the light-cone gauge [(4.12) for $n^2=0$]; the role of the vector n is played by the component q' of the incident photon ($q'^2=0$). In this gauge the contribution of graphs with bremsstrahlung quanta is neglected in the leading-log approximation. Equations (4.28) retain their meaning in non-Abelian theories if it is assumed that the fields take values from the Lie algebra of the gauge group and ordering of the operators along the integration path is introduced.

Thus, the Fock gauge is related to an entire group of physically important gauges having different physical interpretations. The essence of any gauge condition is the singling out of a field (degree of freedom) not contained in or eliminated from the equations of motion. In the case of the Maxwell gauge this is the longitudinal field $\mathbf{A}_\parallel$. It is associated with a charged matter field and cannot be emitted [the Coulomb exponential (2.36)], and therefore the dynamics of only charged fields and transverse fields $\mathbf{A}_\perp$, which propagate freely, are studied. The latter fields satisfy the condition $\boldsymbol{\partial}\mathbf{A}_\perp=0$; the gauge transformation $\mathbf{A}\to\mathbf{A}'=\mathbf{A}_\perp$ has essentially been made. The relation to the Fock gauge (1.14) arises from the possibility of representing the Coulomb factor in (2.36) by an infinite product of linear exponentials (4.25) and (4.26).

The gauges used in the physics of high-energy particles are also associated with line integrals of vector potentials. The semiclassical approximation is valid for $|\mathbf{p}|\to\infty$. Here the particle trajectories are approximated by lines, and the field of the undetected ("soft") quanta accompanying the particle is given by the exponentials (4.28). The Fock gauge thus arises: the "infrared fields" accompanying the particle are eliminated from the dynamics, i.e., the gauge transformation $A_\mu\to A'_\mu$, $(x-x_0)_\mu A'_\mu=0$ is actually accomplished. Now it is sufficient to follow the emitting particle. For $|\mathbf{p}|\to\infty$ the vector $n_p\sim x-x_0$ becomes light-like and the Fock gauge becomes the light-cone gauge.

Although both the Maxwell gauge and the light-cone gauge are based on the conditions (1.14) and (1.15), physically they are completely different. As we have shown, nondynamical gauge conditions single out the fields associated with charges and not explicitly involved in the equations of motion. In the case of the radiation gauge, this is the Coulomb field described by $\mathbf{A}_\parallel$. It is associated with charges and can propagate only with them, whereas the fields $\mathbf{A}_\perp$ are dynamical, i.e., they can be absorbed and emitted, and can propagate independently of charges. In the case of the axial gauges associated with the introduction of the factors (4.28), the field n_pA^Δ is the field of the undetected radiation, i.e., the hidden, passive degrees of freedom. The other components of A_μ actively manifest themselves and can be observed. Of course, the use of an improved detector would allow the observation of the fields that we have excluded from the field dynamics. This implies that the splitting of fields into passive and active components is arbitrary. Finally, another special feature of the gauges used to study real processes is that all the fields are excluded in the light-cone gauge (1.12), whereas according to (4.28) in hard processes only quanta with energy in the range Δ should be excluded.

In conclusion, we note that linear exponentials have a deep geometrical meaning. From the viewpoint of fiber bundles, the fields A_μ are connections, and the linear exponentials are parallel-transport operators.[22,121] Taking into consideration the fact that the Fock gauge also has a natural interpretation in the language of exterior differential forms (see Sec. 7.3), it becomes clear that its appearance is the result of deep penetration into the essence of gauge theories.

4.2. Dynamical gauges

The Feynman gauge $\mathscr{L}'=-(\partial_\mu A_\mu)^2/2$. First, let us show that the classical equations of motion in the Lorentz and Feynman gauges are the same. In fact, from the representation (2.6) and (2.8) it follows that after the four-dimensional divergence is subtracted, the Lagrangian of the free electromagnetic field $\mathscr{L}_0$ can be written as

$$\mathscr{L}_0=-\frac{1}{2}A_\mu\Box A_\mu+\frac{1}{2}(\partial_\mu A_\mu)^2. \quad (4.30)$$

The Heisenberg–Pauli–Feynman gauge fixing[32,33] amounts to adding to $\mathscr{L}_0$ the term $\mathscr{L}' = -(\partial_\mu A_\mu)^2/2$, so that in both gauges the second term in the Lagrangian (4.30) vanishes and we arrive at (4.4). The difference between them is that in the Lorentz gauge the field satisfies $\partial_\mu A_\mu = 0$, while according to (4.4) in the Feynman gauge it satisfies the d'Alembert equation $\Box \partial_\mu A_\mu = 0$.

The Feynman gauge is essentially the standard. There are at least two reasons for this: (1) its explicit relativistic invariance, and (2) the possibility of following the fate of the physical degrees of freedom. In fact, the photon propagator in this gauge is proportional to the metric tensor:

$$\Delta^c_{\mu\nu} = \frac{-i}{q^2+i0} g_{\mu\nu}, \tag{4.31}$$

i.e., charged particles exchange not only transverse quanta $\mathbf{A}_\perp$, but also quanta of the fields A_0 and $\mathbf{A}_\parallel$. The net effect of the latter must be the Coulomb interaction. How do the physical and unphysical degrees of freedom coexist here? Feynman answered this question.[122–124] In the momentum representation the effect of exchange of A_μ quanta is proportional to M,

$$iM = \frac{j_\mu j_\mu}{q^2+i0}. \tag{4.32}$$

Taking into account the current conservation $q_\mu j_\mu = q_0 j_0 - \mathbf{q}\mathbf{j}_\parallel = 0$, we substitute into the sum $j_\mu j_\mu = j_0^2 - \mathbf{j}_\parallel^2 - \mathbf{j}_\perp^2$ the expression for the longitudinal component:

$$\mathbf{j}_\parallel = \frac{\mathbf{q}}{\mathbf{q}^2} q_0 j_0. \tag{4.33}$$

We obtain

$$iM = \frac{1}{q^2+i0}\left[j_0^2 - \frac{q_0^2}{\mathbf{q}^2} j_0^2 - \mathbf{j}_\perp^2 \right] = -\frac{j_0 j_0}{\mathbf{q}^2} - \frac{\mathbf{j}_\perp \mathbf{j}_\perp}{q^2+i0}. \tag{4.34}$$

The representation (4.34) has a clear physical interpretation: the exchange of quanta of the fields A_0 and $\mathbf{A}_\parallel$ gives the Coulomb interaction $-j_0^2/\mathbf{q}^2$, whereas the exchange of quanta of the transverse (gauge-invariant) components $A_\perp$ is determined by the magnetic characteristics of the particles. Knowing the propagator in the Feynman gauge, it is easy to obtain the propagators in the Lorentz and axial gauges. The starting point is the expression

$$A'_\mu = A_\mu + \partial_\mu \Lambda. \tag{4.35}$$

Going, for example, to the Lorentz gauge $\partial_\mu A'_\mu = 0$, we have $\Lambda = \Box^{-1} \partial_\mu A_\mu$, i.e., $A'_\mu = A_\mu + \Box^{-1} \partial_\mu \partial_\nu A_\nu$, and using the propagator A_μ given by (4.31), we obtain (4.6). In the case of axial gauges we have $nA' = 0$, i.e., $\Lambda = -(n\partial)^{-1} nA$ and $A'_\mu = A_\mu - \partial_\mu (n\partial)^{-1} nA$. Again using (4.31), we obtain the propagator (4.12) in axial gauges: the Weyl gauge ($n^2 = 1$), the light-cone gauge ($n^2 = 0$), and the Arnowitt–Fickler gauge ($n^2 = -1$). Finally, to go to the radiation gauge (1.9) it is necessary to take $\Lambda = -\Delta^{-1} \boldsymbol{\partial}\mathbf{A}$. Using the matrix T (2.7) and the operators $\Box_T = T^{\mu\nu} \partial_\mu \partial_\nu = \Delta$, $\partial^T_\mu = g_{\mu\nu} T^{\nu\rho} \partial_\rho$, $\partial \partial_T = \partial_T^2 = -\Box_T$, $\partial^T_\mu A_\mu = -\boldsymbol{\partial}\mathbf{A}$, we write A'_μ in the Maxwell gauge in invariant form: $A'_\mu = A_\mu + \partial_\mu \Box_T^{-1} \partial^T_\nu A_\nu$. The desired photon propagator has the form

$$\Delta^M_{\mu\nu} = \frac{-i}{q^2+i0}\left(g_{\mu\nu} - \frac{q_\mu q^T_\nu + q^T_\mu q_\nu}{q_T^2} + \frac{q_\mu q_\nu}{q_T^2} \right). \tag{4.36}$$

Owing to current conservation, the divergence in (4.35) does not change the probability amplitude of the process, i.e., in all these gauges the decomposition (4.34) and the criterion (4.19) will be valid. To make things completely clear, let us give all the details of the calculations for some single gauge. For example, in the Weyl gauge $\Delta^W_{00} = 0$, so that

$$iM^W = \frac{1}{q^2+i0}\left(-\mathbf{j}^2 + \frac{(\mathbf{qj})^2}{q_0^2} \right). \tag{4.37}$$

However, $\mathbf{j}^2 = \mathbf{j}_\parallel^2 + \mathbf{j}_\perp^2$, $(\mathbf{qj})^2 = \mathbf{q}^2 \mathbf{j}\|2$, and, taking into account (4.33), we arrive at the representation (4.34) for M^W. The corresponding calculations for other gauges differ only in inessential details.

Other gauges. Let us briefly discuss some of the other gauges involving the addition of gauge-noninvariant terms to the Lagrangian. The addition of $\mathscr{L}'$ (1.21) (the planar gauge) changes the matrix K (2.8) in the Lagrangian (2.6):

$$K_{\mu\nu} \to K'_{\mu\nu} = -\Box \left(g_{\mu\nu} + \frac{\partial_\mu \partial_\nu}{\Box} - \frac{n_\mu n_\nu}{\alpha n^2} \right). \tag{4.38}$$

The propagator of a vector particle has the form

$$\Delta^{Pl}_{\mu\nu} = \frac{-i}{q^2+i0}\left(g_{\mu\nu} - \frac{q_\mu n_\nu + n_\mu q_\nu}{qn} + \frac{q_\mu q_\nu}{(qn)^2}(1-\alpha) n^2 \right). \tag{4.39}$$

For $\alpha \to 1$ we obtain the planar gauge of Ref. 46 ($n^2 \neq 0$), and for $n^2 \to 0$ we obtain the Lipatov gauge[45] ($\Box nA = 1$). The latter is related to the light-cone gauge $nA = 0$ just as the Feynman gauge ($\Box \partial A = 0$) is related to the Lorentz gauge ($\partial A = 0$). The term "planar gauge" is used[46] because of the plane in which the vector $n = \alpha q' + \beta p'$, $n^2 \neq 0$, lies [see (4.29)]. Thus, the terminology associated with the gauge-fixing term (1.21) applies to various gauges.

The gauge (1.22) for $n^2 = 0$, 1 is a dynamical gauge ($\mathscr{L}'$ is quadratic in the velocities $\dot{A}_0$), and for $n^2 = -1$ the gauge-fixing term $\mathscr{L}'$ changes the dynamics of the component nA.

The background gauge (1.23) is important not only for obtaining general results, but also for specific calculations.[49] Both the choice of $\mathscr{L}'$ and the idea of splitting the connection into a tensor and a connection [the condition (1.24)] are important in this approach.

Sometimes the gauge-fixing term is taken to be[44]

$$\mathscr{L}' = -\frac{1}{2\alpha} (\boldsymbol{\partial}\mathbf{A})^2. \tag{4.40}$$

or

$$\mathscr{L}' = -\frac{\beta^2}{2} (nA)^2. \tag{4.41}$$

First, we note that these gauges cannot be classified as dynamical, as they do not contain the velocity $\dot{A}_0$. Furthermore, the addition of (4.40) to $\mathscr{L}$ is equivalent to the radiation gauge only for $\alpha \to 0$, and the addition of (4.41) (Ref. 31) amounts to giving a mass to the field nA and for $n^2 < 0$ changes the dynamics of the physical sector. The pseudoaxial gauge (4.41) reduces to the axial gauge for $\beta \to \infty$.

4.3. Quantization

The fundamental role of dynamical gauges is manifested in quantization. Nondynamical gauge conditions, for example, $A_0 = 0$, cannot be treated as operator equations. Constraints also cannot be treated as operator equations: the equation $\pi^0 = 0$ contradicts the commutation relations (2.48). Constraints must be understood in the weak sense, i.e., as conditions on the physical state vectors (2.49). However, then the Weyl gauge condition cannot be understood even in the weak sense (as a condition on the wave function), because this would contradict the uncertainty relation. Therefore, from the viewpoint of the canonical quantum theory the only possible way of eliminating the freedom is to use dynamical gauges. This is what is done in QED. New gauges started to appear when non-Abelian theories were studied, and also in connection with the use of the path-integral technique. The function $\delta(A_0)$ is substituted into the integral (2.50), and $\delta(A_0)\delta(\pi^0)$ is substituted into the path integral corresponding to the Hamiltonian.[71] Both of these methods of eliminating the freedom contradict the postulates of quantum mechanics. However, even in classical physics, as we saw in Sec. 3, the elimination of the variable y (i.e., A_0) from the Lagrangian ($y \to 0$) leads to the loss of the secondary constraint. It is possible to eliminate the unphysical variables by means of δ functions only after exposing all the constraints. As was shown in Sec. 3.2, after all the unphysical variables have been determined, they can in principle be set equal to zero or, more precisely, ignored. In other words, it is assumed that the wave functions do not depend on them, because they do not affect the physical sector. However, the situation is this simple only when the unphysical variables are not related to curvilinear coordinates. Otherwise, the rules for canonical quantization are inapplicable.[100] The correct quantization procedure presupposes transformation to Cartesian coordinates[100] or a special set of rules in path-integral quantization.[107] However, a relation between the unphysical variables in gauge theories and curvilinear coordinates is more the rule than the exception. As we have seen for the model (3.1), even in electrodynamics the unphysical variable θ is the angle in polar coordinates. The operations of quantization and elimination of the unphysical variables do not commute for this reason.[107] It is necessary to follow the rules for the quantization of systems with constraints formulated by Dirac.[96] In the case of the model (3.1) the inclusion of these features leads to the appearance of an additional term $(-1/2r)\partial_r$ in the physical Hamiltonian (3.15). This question has been analyzed for Yang–Mills theory in Refs. 125 and 126. It was shown that the inclusion of the curvilinear nature of the coordinates leads to modification of the physical Hamiltonian operator also in this case.

5. GAUGE TRANSFORMATIONS

5.1. Local and global gauge transformations

Although the fundamental difference between local and global gauge transformations has already been discussed in the literature,[63,127] so far this important problem has not yet found its way into textbooks and monographs (the subject of Ref. 63 is general relativity, and in Ref. 127 the discussion is very general). The lack of a clear understanding of the physics responsible for invariance under such transformations is a problem. Let us briefly discuss this topic.

1. Global gauge invariance. Formally, this amounts to invariance of the action (Lagrangian density) under transformations of the fields and coordinates with constant parameters ω ($\partial_\mu \omega = 0$). Conservation laws are the consequence of this invariance (Noether's first theorem[128]). This statement becomes trivial if we consider one-dimensional motion. In general, the Lagrangian L depends on coordinates and velocities, but the requirement of invariance under a displacement $x \to x + a$, $a = \text{const}$, eliminates the x dependence of L: $\partial L / \partial x = 0$. Physically, this means that all points on the axis are equivalent, i.e., the state and physical conditions in which the particle is located near, say, the origin are indistinguishable from the state and conditions near the point $x = a$. However, if nothing acts on the particle, its momentum is conserved. Formally, this follows from the equation of motion: $p = \partial L / \partial \dot{x}$, so that $dp/dt = 0$. In other words, p is a cyclic variable.[129] The situation does not change in the more general case—the statement about global symmetry is equivalent to the statement that there are cyclic variables on which the Lagrangian does not depend, and the momenta canonically conjugate to them, which are the generators of the given symmetry transformations, are conserved.

2. Local gauge invariance. As already mentioned, this amounts to the statement that the action (the Lagrangian, the Lagrangian density) is not changed by transformations of fields and coordinates with time-dependent parameters (in field theories the parameters usually depend also on the spatial coordinates **x**). The consequences of this symmetry are formulated in the second Noether theorem:[128] a local symmetry of the action leads to identities relating the Eulerians $\mathscr{E}$ and their derivatives with respect to coordinates if the transformation laws of the fields involve derivatives of parameters [as occurs, for example, in electrodynamics; see (1.1). Here $\mathscr{E} = \partial \mathscr{L} / \partial \varphi - \partial [\partial \mathscr{L} / \partial \partial_\mu \varphi] / \partial x^\mu$, where φ is a field; i.e., the equations $\mathscr{E} = 0$ are also the equations of motion]. In the case of one-dimensional motion, the requirement that the Lagrangian be invariant under the substitution

$$x \to x + a(t) \tag{5.1}$$

leads to the condition $L = 0$, which implies the absence of any physical process, i.e., the absence of motion. This is easy to understand when we recall that the function $a(t)$ is arbitrary. The equation of motion fixes the law for the time variation of x. The fact that the replacement (5.1) is possible implies that nothing is changed if at time t the coordinate x is replaced by $x+a(t)$. But then, using the arbitrariness of $a(t)$, it is possible, for example, to choose $a(t)=-x$. In this case the particle will always be located at the origin, independently of the choice of initial conditions, i.e., there is actually no motion. This implies that the variable x is unphysical (a purely gauge degree of freedom). The particle momentum is zero, and we have a constraint. In the case of a more complicated gauge group there will be more unphysical variables, but the essence of the problem remains unchanged.

The difference between local and global symmetries is that the first implies the presence of conservation laws, i.e., the presence of cyclic variables with trivial dynamics, and the second implies the absence of dynamics, i.e., complete arbitrariness of the variables appearing in the Lagrangian. Therefore, these are fundamentally different invariance properties of systems which physically are completely different.

We conclude by noting the following. Nothing prevents the inclusion of global transformations in the group of local transformations, whereas the extension of the group of global transformations to include local ones radically changes the physical system. In fact, in theories with a global symmetry the generalized momenta (generators of transformations) are conserved, and in theories with local symmetry they are equal to zero. However, at the same time this implies that they are conserved; they remain zero during the motion. Therefore, the inclusion of global transformations in the gauge group is natural; moreover, it would be unnatural to exclude them. Arguing somewhat differently, we can say that if arbitrary independent coordinate shifts are allowed at any instant of time, then shifts by the same arbitrary amount independent of time are a special case of this. We have dwelled on this point in some detail, because if this is indeed the case there are important physical consequences: first, the total electric charge of the Universe is equal to zero, and, second, there are superselection rules for electric charge (Refs. 99, 130, and 131).

5.2. Large gauge transformations

In the Introduction we noted that arbitrary parameters [the function Λ and the matrices U in (1.1)] cannot be completely arbitrary. The most obvious restriction is that they must not change the class of the field defined in the formulation of the theory (the geometrical and algebraic nature of a field cannot change, and so on; the exception is BRST transformations transforming vector potentials into even elements of the Grassmann algebra). However, in addition to these obvious restrictions there are more subtle ones concerning the continuity of fields, the spatial topology, and so on. For example, in going from Euclidean space R^3 to the three-sphere S^3 (i.e., compactifying space), we obtain a completely different theory. The matrices $U(x)$ in (1.2) realizing the representation of the gauge group G at each point in space map S^3 onto the group. It is known from topology[121] that the set of matrices $U(x)$ splits into classes (1.40), where the elements of different classes cannot be transformed into each other by a continuous transformation. This poses serious questions and has important physical consequences for the vacuum structure (the θ vacuum, the chirally noninvariant vacuum, and so on[26]). The nature of the problems which arise can best be seen for the example of electrodynamics in $(1+1)$ spacetime. However, before turning to this model let us explain why topological methods are applicable to field theories.

Topology studies continuous mappings of sets and the properties of these sets which do not change under such mappings. However, in classical field theory only continuous fields are allowed, since configurations of them containing discontinuities possess infinite energy. This is most easily seen from the Lagrangian (2.2). The scalar field φ (more precisely, one of its components) contributes to the potential energy an amount $\int d^3x(\nabla\varphi)^2/2$. Since the derivative of a discontinuous function has a δ-like singularity, the square of the δ function appears inside the integral, causing the energy integral to diverge. Therefore, only continuous field configurations are allowed in the theory, and this is why topological methods are so effective. In classical physics, field configurations separated by an infinite potential barrier cannot undergo transitions into each other, and it is sufficient to study only continuous fields of a single class. In quantum theory the question of transitions of one field configuration into others is answered by calculating the corresponding probabilities.

To illustrate some of the topological aspects of gauge transformations, let us turn to electrodynamics in two-dimensional spacetime with a compactified spatial axis (a cylinder with pseudo-Euclidean metric). In this case only the component F_{01} of the tensor $F_{\mu\nu}$ is nonzero. The property of continuity of the fields leads to their periodicity: $A(x+L,t)=A(x,t)$, $\psi(x+L,t)=\psi(x,t)$, where L is the circumference of the cylinder. The Lagrangian and Hamiltonian are obtained directly from (2.2) and (2.35) by going to two-dimensional space. The transformation law for ψ and F leads to continuity of the matrices $U(x,t)$ in (1.1); in the case of electrodynamics, $U(x,t)=\exp(ief(x,t))$ (in this model we use f instead of Λ). Since the transformed fields must also be periodic, we have

$$f(x+L,t)=f(x,t)+\frac{2\pi n}{e}. \tag{5.2}$$

For fixed charge e the parameter f varies in the range $0\leqslant f\leqslant 2\pi/e$, i.e., the gauge group is isomorphic to a circle. From the viewpoint of the theory of fiber bundles,[21,22] we have arrived at the following structure: with each point on the circle S^1, $x\in S^1$, there is associated a group $G=U(1)$, $U\in G$, i.e., a circle of radius $1/e$. In the end we obtain a torus (see Fig. 2). The exponential U [the function $f(x,t)$] at each instant of time is given by some line on the torus (a section). In this case the topological characteristic of the transformation U [the analog of (1.40)] is

$$n=\frac{i}{2\pi}\int_0^L dxU\partial U^{-1}=\frac{e}{2\pi}\int_0^L dx\partial f/\partial x$$
$$=\frac{e}{2\pi}[f(L,t)-f(0,t)]. \tag{5.3}$$

In the case $U=1$ ($f=0, 2\pi/e$) this is the circle S^1, the trivial mapping, $n=0$. Any other mapping whose "trajectory" on the torus can be deformed without discontinuities into the

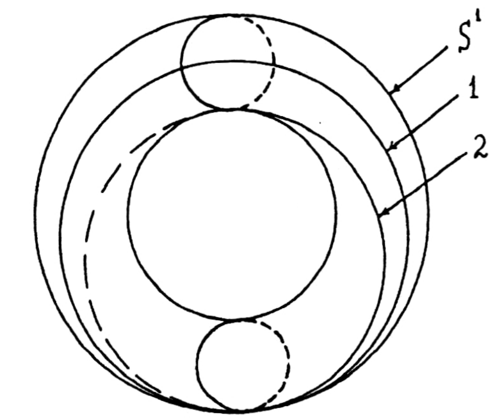

FIG. 2.

circle S^1 (line 1 in Fig. 2) belongs to this same class. As can be seen from (5.3), if $f(L,t)-f(0,t)=2\pi n/e$, i.e., if the function f is discontinuous at the point $x=0$ of the circle S^1, the corresponding transformation belongs to the class characterized by the number n. In Fig. 2 trajectory 2 gives an example of a transformation with $n=1$. We see from this figure that trajectory 1 cannot be deformed into trajectory 2 without discontinuities.

As was mentioned in the Introduction, the vacuum values of the fields can be taken as $A_\mu^{vac}=U\partial_\mu U^{-1}$, since for them $F_{\mu\nu}=0$. However, the matrices (exponentials) are characterized by the number n $[U\equiv U^{(1)}\to U^{(n)}=U^n]$, i.e., the connections also split up into classes. In this case they are distinguished by the Chern–Simons invariant

$$C=\int dxA_1. \tag{5.4}$$

Transformations $U^{(n)}$ with $n\neq 0$ are termed large.[26] From (1.1) and (5.3) we conclude that large transformations change the Chern–Simons index. The Pontryagin index in this model,

$$Q=\frac{1}{2}\int dxdt\varepsilon^{\mu\nu}F_{\mu\nu}=\int dxdt\partial_t A_1=C(\infty)-C(-\infty), \tag{5.5}$$

is the change of C in the motion. This example can be generalized to non-Abelian groups and spaces of large dimension.[26]

5.3. Singular gauge transformations

Singular transformations are best illustrated by the example of the Dirac monopole.[82,83,40] Since in electrodynamics the fundamental field is the vector potential A_μ and the divergence of the magnetic field $\mathbf{H}=\text{curl }\mathbf{A}$ is by definition equal to zero, the only possible way of constructing a magnetic monopole in the space R^3 is by providing the magnetic charge with a string containing the magnetic flux of the charge and ensuring that the condition div $\mathbf{H}=0$ is satisfied in all space. This construction would actually be a model of a magnetic charge if the string were unobservable.[82] The unobservability of the string opens up the following possibility: since now displacement of the string in space is not a

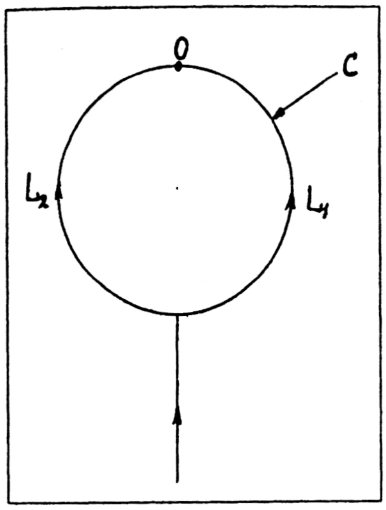

FIG. 3.

physical operation, it must reduce to a gauge transformation. We shall show that such a transformation actually exists.[84] Let L_1 in Fig. 3 be the initial location of the Dirac string, and L_2 be the final location (the magnetic field lines originating from the magnetic charge at the point O are not shown). The latter is obtained by adding to L_1 a closed contour $C(L_1+L_2)$ with magnetic field circulating in it, the direction of which is determined by its direction on the segment L_2. Let us construct the corresponding function Λ. We define $\Lambda=g\Omega(\mathbf{x})/4\pi$, where $\Omega(\mathbf{x})$ is the solid angle subtended by the contour C as seen from the point $\mathbf{x}$, and g is the magnetic charge. This function is independent of time, since the problem is stationary. The main features of $\Lambda(\mathbf{x})$ are: (1) it is multivalued—it changes by g in a circuit around the string forming the contour C; (2) it is singular—it is not defined on the contour C; (3) application of this transformation to $\mathbf{A}$ reduces to a displacement of the string, $L_1\to L_2$. The first two statements are obvious, and the proof of the last is elementary. Let $\mathbf{A}'=\mathbf{A}+\partial\Lambda$, where

$$\oint_{L_1} A_l dx^l=g. \tag{5.6}$$

However, by definition,

$$\oint_{L_1} \partial\Lambda d\mathbf{x}=\frac{g}{4\pi}\Delta\Lambda=-g, \tag{5.7}$$

i.e., the integral (5.6) of $\mathbf{A}'$ is zero. On the other hand, calculation of the analogous integral with contour enclosing the segment L_2 gives g. This implies displacement of the string. The passage to the general case, where Λ has arbitrary time dependence, is not difficult. Time dependence of Λ implies that the Dirac string can change its configuration with time in an arbitrary manner. This has no effect on the dynamics of the magnetic charge or on the physics if the string is unobservable. In reality, it cannot be unobservable—its magnetic flux is nonzero, it must gravitate, and so on. Here we see a clear difference between ordinary local gauge transformations and large or singular ones: the latter change the physical characteristics of the system (the topological numbers, the magnetic fluxes, and so on).

See Ref. 88 for a discussion of supersymmetric gauge transformations. Here again there are global and local transformations. This seems to be the only thing that can be said about them with complete reliability. The nature of supersymmetry remains unknown, and the theory of supersymmetric spaces is underdeveloped.[132]

6. CONCLUSION

Analysis of gauge systems shows that the presence of unphysical degrees of freedom in the Lagrangian has a decisive influence on the structure of the physical sector of the theory. In this sense they cannot be regarded as completely unphysical. Obviously, their meaning will only become clear in the future, when a unified theory of all gauge fields is constructed. In any case, the gauge principle on which all the fundamental interactions are based suggests that the construction of such a theory is possible. Let us conclude by briefly discussing certain topics mentioned briefly above.

The difference between dynamical and nondynamical gauges is of fundamental importance. Only the former admit systematic quantization in the original theory, when all the degrees of freedom present in the Lagrangian are taken into account. In the second case it is necessary to modify the formalism by eliminating the unphysical variables, which leads to complications in non-Abelian theories. Of course, dynamical gauges reduce to a change of the original Lagrangian, so that it is also necessary to modify the dynamical principle, although not so radically. The following two conditions must necessarily be satisfied when adding gauge-fixing terms to the Lagrangian:

- The modified theory must be consistent;
- The physical sector must not be touched.

The example of an inconsistent gauge in Sec. 3 shows that these conditions are not so far-fetched. In addition to these obvious conditions, there are others related to the requirements traditionally imposed on Lagrangians. They are:[95] relativistic invariance, locality, renormalizability, and absence of field derivatives of order higher than the first. These requirements are natural, particularly in perturbation theory. They lead practically uniquely to the Heisenberg–Pauli gauge-fixing term (1.16).

The importance of the extended group of gauge transformations introduced by Dirac[96] should be stressed. The generators of this group are all the first-class constraints (both primary and secondary). In accordance with the requirement of weak gauge invariance of physical quantities, it gives a natural generalization of the gauge group of the original Lagrangian. Consistency of the final picture means that analysis can be done within the Hamiltonian formalism.[96]

The special role of the Fock gauge (1.14) should be noted. Whereas the Feynman gauge is convenient for perturbative calculations, the Fock gauge comes into its own in the nonperturbative approach. In a remarkable manner it turns out to be the basis of many popular gauges. It reflects the deep connection of gauge theories to the geometry of fiber bundles and to the theory of exterior differential forms (Sec. 4 and 7).

In the case of compactified spaces, there is a theorem[133] which states that for theories with a compact non-Abelian Lie group in S^4 it is impossible to fix uniquely the gauge in an invariant manner.

An important tool for studying gauge theories is the background-field gauge,[47–49] or, more precisely, the idea of splitting the gauge field into a quantum field (a tensor) and a classical background (the connection), which not only allows the proof of general statements (for example, the gauge invariance of the effective action), but also simplifies the actual calculations.[49]

In gauge theories one is struck by the conjunction of locality and nonlocality. All theories are formulated in terms of local fields with local interactions. However, physical objects are nonlocal. An electron is meaningless without its surrounding electric field. Gauge theories are simply formulated in coordinate space, while in momentum space even the simplest gauge transformation of a charged field (1.1) becomes an integral transformation. This implies, in particular, that there is no gauge-invariant way of separating the regions of low and high energy. Let us write down Gauss's law in momentum space:

$$i\mathbf{qE}(q)=e\int \frac{d^4k}{(2\pi)^4}\,\psi^*(q-k)\psi(k). \tag{6.1}$$

Here the integration involves the entire momentum spectrum, and neglect of, for example, small-wavelength quanta violates the secondary constraint, i.e., spoils the gauge invariance of the theory.

A very important question is that of the range of values taken by the parameter Λ in (1.1). It can have either a finite range ($\Lambda\in[0,2\pi/e]$) or an infinite range [$\Lambda\in(-\infty,\infty)$]. Physically, these are fundamentally different possibilities. Going to global transformations and noting that they are generated by the electric charge,[131] we find that in the first case the spectrum of the charge operator is discrete (quantization of the charge), while in the second it is continuous. Since the electric charge is quantized, the range of values of the parameter Λ is bounded.

The Ward identities, along with the Noether identities, are obvious manifestations of the unphysical degrees of freedom. Being formally added to the theory, they remain arbitrary, which can be seen from the form of the Ward identities in quantum theory and of the Noether identities in classical theory. The former show that they are split off from the physical sector, and the latter show that there are no equations of motion for them (Sec. 7.5).

The features of gauge-fixing in stochastic quantization are discussed in Ref. 134.

7. APPENDIX

7.1. Residual gauge freedom

Nondynamical gauges usually do not completely eliminate the gauge freedom. As a rule, the theory still admits transformations from a smaller group which does not change the dynamics in the physical sector. For example, the Weyl gauge $A_0=0$ admits transformations of the vector potentials with time-independent parameters Λ: $\dot{\Lambda}=0$. Meanwhile, the fundamental difference between global and local gauge transformations is related to the time dependence of the transformation parameters. It is therefore ambiguous to speak of a residual freedom as a gauge freedom without indicating the type of the corresponding transformation (see, for example, Ref. 120, p. 77). Actually, the remaining symmetry with parameters $v=v(\mathbf{x})$ is not dynamical (i.e., local; see Sec. 7.5 below), and it must lead to an infinite number of conservation laws. Let us consider an example: $\mathscr{L}=\dot{\varphi}^2(\mathbf{x},t)/2$. The Lagrangian is invariant under the group of transformations $\varphi\rightarrow\varphi+v(\mathbf{x})$, and even though the group here is infinite-dimensional, while the fields transform locally (from the viewpoint of 3-dimensional space), the consequence is an infinite number of conserved quantities:

$$\frac{\partial\pi(\mathbf{x},t)}{\partial t}=0, \quad \pi=\frac{\partial\mathscr{L}}{\partial\dot{\varphi}}. \tag{7.1}$$

The question arises: why doesn't this happen in electrodynamics? The point is that here we are dealing with a strong (exact) local symmetry of the Lagrangian. Only the weak gauge invariance under the extended group of transformations is physically important. It is the secondary constraints which generate transformations with the parameters $v(\mathbf{x},t)$. Enlargement of the symmetry directly leads to $v(\mathbf{x})\rightarrow v(\mathbf{x},t)$. Therefore, new laws do not appear—there are the secondary constraints and the unphysical degrees of freedom corresponding to them. It can be stated that there exist an infinite number of conserved trivial quantities which can only take the value zero (see Sec. 5).

The situation changes somewhat in going to the Lorentz gauge (1.8). The residual gauge freedom is now related to functions satisfying the d'Alembert equation $\Box\Lambda=0$, which is also often interpreted as incomplete elimination of the gauge freedom. Actually, even though now the parameter Λ can also depend on time, the function Λ is not arbitrary—it satisfies the equation of motion, and its arbitrariness reduces to that of the initial conditions $(\Lambda_{t=0},\dot{\Lambda}_{t=0})$. Clearly, this is not gauge freedom (local gauge invariance). In the model of one-dimensional motion studied above (Sec. 5), the d'Alembert equation would reduce to a condition on the gauge parameter, $\ddot{a}=0$, i.e., $a(t)=a+bt$, $\dot{a}=\dot{b}=0$. Such an invariance does not restrict the dynamics of the variable x and does not make the dynamics meaningless, because the Newtonian equation $\ddot{a}=0$ is invariant under the transformation (5.1) with such a parameter a (this is the Galilean transformation). The same is true of the theory of gravity. The harmonic coordinates satisfy Eq. (1.34), $\Box x=0$. For suitable boundary conditions the freedom reduces to transformations from the Lorentz group,[63] in which the time also enters linearly.

7.2. The "breaking" of local symmetry

The fundamental difference between local and global gauge symmetries is clearly manifested in the study of systems with broken symmetry. By broken global symmetry we mean the presence of noninvariant solutions of the equations of motion; in quantum field theory the vacuum is noninvariant. However, in the case of a local gauge symmetry the physical sector must in any event be invariant under the extended group of gauge transformations. The vacuum state vector in any event cannot violate the local symmetry—this would contradict the initial postulates of the theory (gauge invariance of the Lagrangian). Regarding the specific mechanism of symmetry breaking—the appearance of a noninvariant vacuum expectation value of the Higgs field, which changes nontrivially under gauge transformations—there is the theorem of Elitzur (Ref. 135; see also Ref. 136), which states that such vacuum expectation values must be zero.

7.3. The Fock gauge and exterior differential forms

In Ref. 28 Fock not only formulated the gauge condition (1.14), but he also gave the solution:

$$A_\mu(x)=\int_0^1 sx^\nu F_{\mu\nu}(sx)ds. \tag{7.2}$$

The field A_μ (7.2) satisfies the gauge condition (1.14). In the theory of exterior differential forms[137] an important role is played by the Poincaré operator I. If ω is a differential m-form in the space R^n, we have the identity

$$\omega=d(I\omega)+I(d\omega), \tag{7.3}$$

where

$$I\omega\equiv\sum_{\{i\}}\sum_{r=1}^{m}(-1)^{r-1}\times\int_0^1 dt t^{m-1}\omega_{i_1\ldots i_m}(tx)x^{i_r}[dx^{i_1}\wedge\ldots\wedge dx^{i_m}]_r. \tag{7.4}$$

Here the symbol $\{i\}$ stands for summation over all i_k, $k=1,\ldots,m$, in the range $1\leqslant i_1<\ldots<i_m\leqslant n$, and the subscript r on the square brackets denotes the absence of the differential numbered i_r. If the form ω is closed ($d\omega=0$), it is exact ($\omega=d\Omega$). This is all valid for the star region $S\subset R^n$: if $x\in S$, then also $tx\in S$, $0\leqslant t\leqslant 1$, and it is certainly valid for R^n.

In electrodynamics $F_{\mu\nu}$ is a 2-form ($F=F_{\mu\nu}dx^\mu\wedge dx^\nu$); F is a closed form ($dF=0$; this is the homogeneous Maxwell equation $\partial_\mu\tilde{F}_{\mu\nu}=0$, $2\tilde{F}_{\mu\nu}=\varepsilon_{\mu\nu\rho\sigma}F_{\rho\sigma}$). Equation (7.2) is the result of applying the Poincaré operator I to the 2-form F, resulting in the 1-form $A=A_\mu dx^\mu$.

The relation of the Fock gauge to the Poincaré operator and to the theory of exterior differential forms shows that Fock deeply penetrated into the essence of the matter. In Ref. 28 this gauge arose naturally from physical considerations. Was Eq. (7.2) obtained independently, or did the Poincaré lemma play a role in its derivation?

7.4. The background-field method

The background-field method can be used to obtain effective Lagrangians by calculating only vacuum diagrams. The background field is a variable and by definition it is equal to the vacuum expectation value of the quantum field.

The method is very convenient not only for calculating effective Lagrangians, but also for proving their gauge invariance. We define

$$e^{i\widetilde{W}[\mathscr{J},\mathscr{F}]}\equiv\int DQ e^{i[S[Q+\mathscr{F}]+\mathscr{J}Q]}. \qquad (7.5)$$

As in Sec. 2, Q stands for all the quantum fields (A, $\bar{\psi}$, ψ,...), and $\mathscr{F}$ are the corresponding classical (external, background) fields, i.e., the substitution $Q\to Q+\mathscr{F}$ has been made in the action inside the integral (2.50). Now instead of (2.51) we have

$$\widetilde{\Gamma}[\tilde{\phi},\mathscr{F}]=\widetilde{W}[\mathscr{J},\mathscr{F}]-\mathscr{J}\tilde{\phi}, \quad \frac{\delta\widetilde{W}[\mathscr{J},\mathscr{F}]}{\delta\mathscr{J}}=\tilde{\phi}. \qquad (7.6)$$

The solution of the second equation is substituted for $\mathscr{J}$ in the first equation in (7.6). Making the change of variable $Q\to Q-\mathscr{F}$ in (7.5) and using the definition of W in (2.50), we have

$$\widetilde{W}[\mathscr{J},\mathscr{F}]=W[\mathscr{J}]-\mathscr{J}\mathscr{F}, \quad \tilde{\phi}=\phi-\mathscr{F}, \qquad (7.7)$$

i.e.,

$$\widetilde{\Gamma}[\tilde{\phi},\mathscr{F}]=W[\mathscr{J}]-\mathscr{J}\mathscr{F}-\mathscr{J}\tilde{\phi}=W[\mathscr{J}]-\mathscr{J}\phi=\Gamma[\phi]. \qquad (7.8)$$

Using the second equation in (7.7), we find

$$\widetilde{\Gamma}[0,\mathscr{F}]=\Gamma[\mathscr{F}]. \qquad (7.9)$$

This is the basic expression of the method: the effective action $\Gamma[\mathscr{F}]$ is equal to the action $\widetilde{\Gamma}[\tilde{\phi},\mathscr{F}]$ calculated for zero vacuum expectation values $\tilde{\phi}=0$, i.e., it is equal to the first term in the expansion of the functional $\widetilde{\Gamma}$ in powers of $\tilde{\phi}$. (We recall that $\widetilde{\Gamma}[\tilde{\phi},\mathscr{F}]$ is the generating functional of 1-irreducible Green functions on the background of external fields $\mathscr{F}$; for $\tilde{\phi}=0$ the functional $\widetilde{\Gamma}[0,\mathscr{F}]$ is the set of 1-irreducible vacuum diagrams on the background $\mathscr{F}$.)

In gauge theories the method is formulated as follows. As already pointed out in Sec. 1, the gauge field A_μ splits into a sum

$$A_\mu\to q_\mu+A_\mu, \qquad (7.10)$$

where q_μ is the quantum field integrated over in the path integral, and A_μ is the classical field. The background-field gauge (1.24) specific to this method is chosen:

$$\mathscr{L}'=-\frac{1}{2\alpha}(D_\mu q_\mu)^2, \quad D_\mu^{ab}=\delta^{ab}\partial_\mu-igf^{acb}A_\mu^c, \qquad (7.11)$$

i.e., $F=D_\mu q_\mu$. The field q_μ transforms as a tensor (homogeneously),

$$\delta q_\mu=igT\omega q_\mu \qquad (7.12)$$

($T\omega=T^a\omega^a$, where T^a are the group generators in the adjoint representation, and ω^a are the infinitesimal transformation parameters), and the field A_μ transforms as a connection (inhomogeneously),

$$\delta A_\mu=D_\mu\omega. \qquad (7.13)$$

According to the general equation (7.9), again in this case $\widetilde{\Gamma}[0,\mathscr{F}]=\Gamma[\mathscr{F}]$, so that to prove the gauge invariance of $\Gamma[\mathscr{F}]$ it is sufficient to prove the invariance of $\widetilde{\Gamma}[0,\mathscr{F}]$. However, this follows more or less obviously from the explicit expression for $\widetilde{W}$:

$$e^{i\widetilde{W}[\mathscr{J},\mathscr{F}]}=\int d[q,\bar{c},c,\bar{\psi},\psi]e^{i[S[Q+\mathscr{F}]-(1/2\alpha)F^2+\bar{c}F'c+\mathscr{J}Q]}. \qquad (7.14)$$

Here $Q=(q_\mu,\bar{\psi},\psi)$, $\mathscr{F}=(A_\mu,\bar{\zeta},\zeta)$, $\psi\to\psi+\zeta$, $F'=\delta F(A^\omega)/\delta\omega$, $\bar{c}$ and c are anticommuting scalar fields, and we can use the Lagrangian (2.1) as the action density. We assume that the external currents transform homogeneously:

$$\delta\mathscr{J}=gT_{\mathscr{J}}\omega\mathscr{J}. \qquad (7.15)$$

Here $T_{\mathscr{J}}$ are the group generators in the representation realized by the current $\mathscr{J}$. Since the sum $q_\mu+A_\mu$ transforms as a connection, while $\bar{\psi}$, ψ, $\bar{\zeta}$, ζ, and $\mathscr{J}$ transform as tensors, the argument of the exponential in (7.14) is gauge-invariant if the fictitious fields transform according to the adjoint representation. Therefore, $\widetilde{W}[\mathscr{J},\mathscr{F}]$ is also gauge-invariant, and, owing to (7.15), $\tilde{\phi}$ in (7.6) transforms as a tensor. However, then $\widetilde{\Gamma}[\tilde{\phi},\mathscr{F}]$ in (7.6) is gauge-invariant, and, owing to the tensor nature of $\tilde{\phi}$, $\widetilde{\Gamma}[0,\mathscr{F}]=\Gamma[\mathscr{F}]$ is also gauge-invariant. The field A_μ in the set $\mathscr{F}=(A_\mu,\bar{\zeta},\zeta)$ transforms inhomogeneously (as a connection), so that the effective action $\Gamma[\mathscr{F}]$ satisfies the identities

$$\frac{\delta\Gamma}{\delta\mathscr{J}}\delta\mathscr{J}\equiv\frac{\delta\Gamma}{\delta A_\mu}D_\mu\omega+\frac{\delta\Gamma}{\delta\psi}ig\hat{\omega}\psi-ig\bar{\psi}\hat{\omega}\frac{\delta\Gamma}{\delta\bar{\psi}}=0, \qquad (7.16)$$

where $\hat{\omega}\equiv T\omega$, and T are the generators in the representation realized by the matter fields.

We note that these conditions for the gauge invariance of $\Gamma[\mathscr{F}]$ are none other than the Noether identities[128] for the effective action $\Gamma\equiv S_{\text{eff}}$.

7.5. The conditions for local invariance of the effective action and the Noether identities

Before proving the statement made at the end of the previous subsection, let us briefly recall the derivation of the Noether identities. When studying the invariance properties of dynamical systems, one usually considers general transformations of the fields, including coordinate transformations:

$$\delta\varphi=\varphi'(x')-\varphi(x), \quad \delta x=x'-x \qquad (7.17)$$

(the tensor and other indices have been dropped). Introducing local variations of the fields[138]

$$\bar{\delta}\varphi=\varphi'(x)-\varphi(x), \qquad (7.18)$$

after some algebra we obtain the following conditions for invariance of the action:

$$\delta S \equiv \int dx\left[\mathscr{E}\bar{\delta}\varphi+\frac{\partial}{\partial x^{\mu}}\left(\frac{\partial \mathscr{L}}{\partial(\partial\varphi)}\,\bar{\delta}_{\varphi}+\mathscr{L}\delta x^{\mu}\right)\right]=0. \tag{7.19}$$

Here

$$\mathscr{E}\equiv\frac{\partial\mathscr{L}}{\partial\varphi}-\frac{\partial}{\partial x^{\mu}}\frac{\partial\mathscr{L}}{\partial(\partial_{\mu}\varphi)} \tag{7.20}$$

is the Eulerian. Equating it to zero gives the equations of motion. The general expression (7.20) allows all the most important relations of the Lagrangian formalism to be obtained:

•From the condition that S be stationary under variation of the fields vanishing on the boundary we obtain the equations of motion ($\mathscr{E}=0$).

•From the condition that S be invariant under global gauge transformations we obtain the conservation laws (the first Noether theorem;[128] here we mean quantities conserved in the motion, that is, when $\mathscr{E}=0$).

•From the condition that S be invariant under local gauge transformations we obtain the second Noether theorem[128]—identities relating the Eulerians $\mathscr{E}$ and their derivatives if the field transformations involve derivatives of parameters [as in (1.1)].

Writing the local variations of the fields as

$$\bar{\delta}_{\varphi}(x)=\sum_{k=1}^{N}\sum_{r=0}^{m}V_{k}^{\mu_1\ldots\mu_r}(\varphi,x)\partial_{\mu_t}\ldots\partial_{\mu_r}\omega_k(x),$$

$$N=\dim X, \tag{7.21}$$

we arrive at the following expression for the Noether identities:

$$\sum_{r=0}^{m}(-1)^{r}\partial_{\mu_1}\ldots\partial_{\mu_r}(\mathscr{E}V_{k}^{\mu_1\ldots\mu_r})\equiv 0. \tag{7.22}$$

Since $\mathscr{E}=\delta S/\bar{\delta}\varphi$ [or $\mathscr{E}=\delta S/\delta\mathscr{F}$ in the notation of (7.14)], the fact that these identities are the same Eq. (7.16) is obvious. The only difference is that the action density in S is a local Lagrangian, whereas its analog in S_{eff} is essentially nonlocal.[51–53]

7.6. The photon propagator in various gauges

Let us give the expressions for the photon propagator in some frequently encountered gauges. We restrict ourselves to writing down only the matrices for the general coefficient $1/i(q^2+i0)$.

(1) The class of Heisenberg–Pauli–Fermi gauges:

$$\mathscr{L}'=-\frac{1}{2\alpha}(\partial_{\mu}A_{\mu})^2;\quad g_{\mu\nu}+(\alpha-1)\frac{q_{\mu}q_{\nu}}{q^2}. \tag{7.23}$$

(2) The Lorentz gauge:

$$\partial_{\mu}A_{\mu}=0;\quad g_{\mu\nu}-\frac{q_{\mu}q_{\nu}}{q^2}. \tag{7.24}$$

(3) The Maxwell gauge:

$$\boldsymbol{\partial}\mathbf{A}=0;\quad g_{\mu\nu}-\frac{q_{\mu}q_{\nu}^{T}+q_{\mu}^{T}q_{\nu}}{q_T^2}+\frac{q_{\mu}q_{\nu}}{q_T^2}. \tag{7.25}$$

(4) Axial gauges:

$$nA=0;\quad g_{\mu\nu}-\frac{q_{\mu}n_{\nu}+n_{\mu}q_{\nu}}{qn}+\frac{q_{\mu}q_{\nu}}{(qn)^2}n^2. \tag{7.26}$$

(5) Pseudoaxial gauges:

$$\mathscr{L}'=-\frac{\beta^2}{2}(nA)^2;$$

$$g_{\mu\nu}-\frac{q_{\mu}n_{\nu}+n_{\mu}q_{\nu}}{qn}+\frac{q_{\mu}q_{\nu}}{(qn)^2}\left(n^2+\frac{q^2}{\beta^2}\right). \tag{7.27}$$

(6) The planar gauge:

$$\mathscr{L}'=-\frac{1}{2\alpha n^2}\,nA\Box nA;$$

$$g_{\mu\nu}-\frac{q_{\mu}n_{\nu}+n_{\mu}q_{\nu}}{qn}+\frac{q_{\mu}q_{\nu}}{(qn)^2}n^2(1-\alpha). \tag{7.28}$$

Depending on the parameter α, $\mathscr{L}'$ in (7.23) fixes the following gauges: $\alpha\to 0$ (Lorentz), $\alpha=1$ (Feynman), $\alpha=1/3$ (Fried–Yennie).

In (7.28) for $n^2\to 0$ we obtain the Lipatov gauge,[45] and for $\alpha\to 1$ we obtain the planar gauge,[46] in which $n^2\neq 0$.

The following expression is useful for obtaining the propagators. The matrix

$$K_{\mu\nu}^{-1}=g_{\mu\nu}+c(q_{\mu}n_{\nu}+n_{\mu}q_{\nu})+dq_{\mu}q_{\nu} \tag{7.29}$$

is the inverse of the matrix

$$K_{\mu\nu}=g_{\mu\nu}+aq_{\mu}q_{\nu}+bn_{\mu}n_{\nu} \tag{7.30}$$

for $a=-1/q^2$, $c=-1/qn$, and $d=(1+bn^2)/b(qn)^2$.

The gluon propagators differ from these by the factor δ^{ab}.

7.7. Some historical remarks

1. The gauge (1.8) first appeared in the 1867 study of L. V. Lorentz. At this time H. A. Lorentz was 14 years old. Maxwell[1] made the following comment about Ref. 24: "...Lorentz has derived...by adding a few terms, which do not affect any of the experimental results, a new system of equations... These derivations are similar to those of the present chapter [Chap. XX], but they have been obtained by a completely different method. The theory presented in this chapter was first published in ...1865." Incidentally, Lorentz not only wrote down the equations

$$\Box A_{\mu}=-j_{\mu}, \tag{7.31}$$

i.e., the equations of motion in the Lorentz gauge, but he also gave their retarded solutions, referring to his earlier study. Curiously, the studies of Maxwell are not mentioned at all in Ref. 24. The brief remarks about the life and scientific activity of Ludwig Lorentz given in Ref. 139 end with the statement: "Independently of J. Maxwell and not knowing his theory, he constructed (1867) the electromagnetic theory of light."

2. Although the commonly used names of the gauge (1.9) seem appropriate, they deviate from the tradition of naming gauges after their authors. Naming this gauge after Maxwell would be an act of historical justice.

3. The commonly used names of the gauge (1.10) are unsuitable, and the proposal of Jackiw[26] that this gauge be named after Weyl can only be greeted with approval.

4. The axial gauge first appeared in the study by Kummer[140] devoted to the quantization of the free electromagnetic field. However, here the requirement (1.13) was accompanied by the condition nq=const, where q is the photon momentum.

5. The Fock gauge is known by many names. It is often referred to as the Fock–Schwinger gauge, but it is also sometimes called the Schwinger–Fock gauge, and even the Schwinger gauge.[116] Schwinger, of course, knew of the authorship of Fock. In Ref. 141, using the Fock proper-time method, he refers to Ref. 28. Later this gauge was repeatedly rediscovered[142,143] (studies written after the appearance of Schwinger's book[30] in 1970, and even after its translation into Russian). In Ref. 144 it is called the "fixed-point gauge," and the authors of Ref. 143 call it the "Poincaré gauge" (which, however, has some resonance; see Sec. 7.3). In Ref. 31 the condition (1.15) is called the Fock–Schwinger gauge, and (1.14) is called the Poincaré gauge.

But this is not all. The Fock gauge was rediscovered even earlier in Refs. 145 and 146. In a subsequent series of studies (see the literature cited in Ref. 147) it is called the "multipolar gauge," and Refs. 143 and 144 are described as "pioneering." It should be noted that even publications of great importance by prominent physicists remain unacknowledged. And the constancy with which this gauge reappears in the pages of authoritative journals is a true indication of its importance. The Fock gauge is not used in standard perturbation-theory calculations, but it naturally arises in going beyond perturbation theory.

6. Fictitious scalar fields with anomalous sign for loops [an additional factor of (-1) for each loop] first appeared in a study by Feynman.[77]

7. The gauge invariance of the effective action Γ indicates that this quantity is fundamental. When $S_{\rm eff}$ is used, the exact S matrix is obtained in lowest-order perturbation theory, i.e., neglecting quantum corrections.[53] This fact can also be interpreted as the cancellation of all corrections to Γ from graphs with loops involving exact propagators and exact vertices. In Ref. 148 equations were written down for Γ which explicitly take into account this property of the effective vertices.

8. Fock was apparently the first to draw attention to the fundamental difference between local and global transformations in general relativity.[63] Physicists were rather restrained in their acceptance of these conclusions. However, Fock was decisively supported by Wigner,[127] who studied not only general relativity, but also electrodynamics. He referred to local and global gauge transformations as dynamical and geometrical transformations, respectively.

9. The model with a gauge group consisting only of translations [with trivial first term in the Lagrangian (3.19), $e=0$] was first proposed by Burnel.[108,109] The model (3.19) has been studied independently by the present author (1988; unpublished).

[1)]This term was suggested by Fock; see Ref. 2.

[2)]The dynamics of one-dimensional objects were first studied in Ref. 20.

[3)]Here it is not our goal to list all the gauge conditions which can be found in the literature. This is practically impossible and not even necessary—"natural selection" has eliminated the less useful ones.

[4)]The old names for global and local transformations are, respectively, gauge transformations of the first and second kind.[80]

[5)]See Ref. 99 regarding the surface terms which arise in integration by parts in the second term of (2.11).

[6)]As far as the author knows, this simple statement does not yet have any simple proof.

[1] J. C. Maxwell, *A Treatise on Electricity and Magnetism* (Dover, New York, 1954) [Russ. transl., Nauka, Moscow, 1989].

[2] W. Pauli, Helv. Phys. Acta **12**, 147 (1939) [Russ. transl. in *Trudy po kvantovoĭ teorii*, edited by Ya. A. Smorodinskiĭ (Nauka, Moscow, 1977)].

[3] O. Klein, in *Proc. of the Conf. on New Theories in Physics*, Warsaw, 1938, p. 66.

[4] C. N. Yang and R. L. Mills, Phys. Rev. **96**, 191 (1954).

[5] R. Shaw, Ph.D. Thesis, Cambridge University (1954).

[6] S. Weinberg, Phys. Rev. Lett. **31**, 494 (1973).

[7] J. Pati and A. Salam, Phys. Rev. D **8**, 1240 (1973).

[8] H. Fritzsch, M. Gell-Mann, and H. Leutwyler, Phys. Lett. **47B**, 365 (1973).

[9] S. Weinberg, Phys. Rev. Lett. **19**, 1264 (1967).

[10] A. Salam, in *Elementary Particle Theory*, edited by N. Svartholm (Almqvist and Wiksell, Stockholm,1968), p. 637.

[11] A. Einstein, Sitzungsber. Preuss. Akad. Wiss. **48**, 844 (1915).

[12] D. Hilbert, Nachr. Ges. Wiss. Göttingen **3**, 395 (1915).

[13] B. M. Barbashov and V. V. Nesterenko, *The Relativistic String Model and Hadron Physics* (World Scientific, Singapore, 1989) [Russ. original, Énergoatomizdat, Moscow, 1987].

[14] D. Lüst and S. Theisen, *Lectures on String Theory*, Lecture Notes in Physics, No. 346 (Springer-Verlag, Berlin, 1989).

[15] M. B. Green, J. H. Schwarz, and E. Witten, *Superstring Theory* (Cambridge University Press, Cambridge, 1987) [Russ. transl., Mir, Moscow, 1990].

[16] Yu. N. Kafiev, *Anomalies and String Theory* [in Russian] (Nauka, Novosibirsk, 1991).

[17] Y. Nambu, Lectures at the Copenhagen Symposium (1970).

[18] T. Goto, Prog. Theor. Phys. **46**, 1560 (1971).

[19] O. Hara, Prog. Theor. Phys. **46**, 1549 (1971).

[20] B. M. Barbashov and N. A. Chernikov, Zh. Éksp. Teor. Fiz. **50**, 1296 (1966) [Sov. Phys. JETP **23**, 861 (1966)]; **51**, 658 (1966) [**24**, 437 (1967)].

[21] M. Daniel and C. M. Viallet, Rev. Mod. Phys. **52**, 175 (1980).

[22] B. A. Dubrovin, A. T. Fomenko, and S. P. Novikov, *Modern Geometry—Methods and Applications*, Parts I–III (Springer-Verlag, New York, 1984, 1985, 1990) [Russ. original, Parts I–III, Nauka, Moscow, 1979].

[23] L. V. Prokhorov and S. V. Shabanov, Usp. Fiz. Nauk **161**, 13 (1991) [Sov. Phys. Usp. **34**, 108 (1991)].

[24] L. Lorentz, Phil. Mag. **34**, 287 (1867).

[25] H. Weyl, *Gruppentheorie und Quantenmechanik* (Leipzig, 1930).

[26] R. Jackiw, Rev. Mod. Phys. **52**, 661 (1980).

[27] R. Arnowitt and S. Fickler, Phys. Rev. **127**, 1821 (1962).

[28] V. Fock, Sov. Phys. **12**, 404 (1937) [Russ. transl. in Ref. 29].

[29] V. A. Fok, *Works on Quantum Field Theory* [in Russian] (Leningrad University Press, Leningrad, 1957).

[30] J. Schwinger, *Particles, Sources and Fields* (Addison-Wesley, Reading, Mass., 1970) [Russ. transl., Mir, Moscow, 1976].

[31] G. Leibbrandt, Rev. Mod. Phys. **59**, 1067 (1987).

[32] W. Heisenberg and W. Pauli, Z. Phys. (Leipzig) **56**, 1 (1929).

[33] W. Heisenberg and W. Pauli, Z. Phys. (Leipzig) **59**, 168 (1930).

[34] E. Fermi, Rend. Acad. Lincei **9**, 881 (1929).

[35] E. Fermi, Rend. Acad. Lincei **12**, 431 (1930).

[36] E. Fermi, Rev. Mod. Phys. **4**, 87 (1932).

[37] E. Fermi, *Scientific Works*, Vol. 1 (Russ. transl., Nauka, Moscow, 1971).

[38] G. 't Hooft, Nucl. Phys. **B35**, 167 (1971).

[39] P. W. Higgs, Phys. Lett. **12**, 132 (1964).

[40] L. H. Ryder, *Quantum Field Theory* (Cambridge University Press, Cambridge, 1985) [Russ. transl., Mir, Moscow, 1987].

[41] G. 't Hooft, Nucl. Phys. **B190**, 455 (1981).

[42] H. Min, T. Lee, and P. Y. Pac, Phys. Rev. D **32**, 440 (1985).

[43] G. 't Hooft and M. Veltman, Nucl. Phys. **B50**, 318 (1972).

[44] W. Kummer, Acta Phys. Austriaca, Suppl. **XV**, 423 (1976).

[45] L. N. Lipatov, Yad. Fiz. **20**, 181 (1974) [Sov. J. Nucl. Phys. **20**, 94 (1974)].
[46] Yu. L. Dokshitzer, D. I. Dyakonov, and A. I. Troyan, Phys. Rep. **58**, 269 (1980).
[47] B. S. DeWitt, Phys. Rev. **162**, 1195 (1967).
[48] B. S. DeWitt, *Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965) [Russ. transl., Nauka, Moscow, 1987].
[49] L. F. Abbott, Acta. Phys. Pol. B **13**, 33 (1982).
[50] G. Jona-Lasinio, Nuovo Cimento **34**, 1790 (1964).
[51] L. V. Prokhorov, Nuovo Cimento A **57**, 245 (1968).
[52] L. V. Prokhorov, Phys. Rev. **183**, 1515 (1969).
[53] N. A. Il'ichev and L. V. Prokhorov, Teor. Mat. Fiz. **6**, 305 (1971) [Theor. Math. Phys. (USSR)].
[54] G. 't Hooft, Nucl. Phys. **B62**, 444 (1973).
[55] S. Deser, J. Kay, and K. Stelle, Phys. Rev. Lett. **38**, 527 (1977).
[56] H. S. Chan and M. B. Halpern, Phys. Rev. D **33**, 540 (1985).
[57] F. Urrutia, Lett. Nuovo Cimento **5**, 788 (1972).
[58] S. G. Karshenboĭm, Yad. Fiz. **50**, 1374 (1989) [Sov. J. Nucl. Phys. **50**, 853 (1989)].
[59] H. M. Fried and D. R. Yennie, Phys. Rev. **112**, 1391 (1958).
[60] S. V. Ivanov, Fiz. Élem. Chastits At. Yadra **21**, 75 (1990) [Sov. J. Part. Nucl. **21**, 32 (1990)].
[61] T. De Donder, *La Gravitique Einsteinienne* (Gauthier-Villars, Paris, 1931).
[62] K. Lanczos, Phys. Z. **23**, 537 (1922).
[63] V. A. Fock, *The Theory of Space, Time and Gravitation* (Pergamon Press, Oxford, 1964) [Russ. original, Fizmatgiz, Moscow, 1961].
[64] A. A. Logunov and M. A. Mestvirishvili, *The Relativistic Theory of Gravitation* (Mir, Moscow, 1989) [Russ. original, Nauka, Moscow, 1989].
[65] J. Scherk and J. H. Schwarz, Gen. Rel. Grav. **6**, 537 (1975).
[66] M. Kaku, Nucl. Phys. **B91**, 99 (1975).
[67] P. A. M. Dirac, Phys. Rev. **114**, 924 (1959).
[68] C. Becchi, A. Rouet, and R. Stora, Phys. Lett. **52B**, 344 (1974).
[69] C. Becchi, A. Rouet, and R. Stora, Ann. Phys. (N.Y.) **98**, 287 (1976).
[70] I. V. Tyutin, Preprint No. 39, Lebedev Physics Institute, Moscow (1975) [in Russian].
[71] L. D. Faddeev and A. A. Slavnov, *Gauge Fields: Introduction to Quantum Theory* (Addison-Wesley, Redwood City, Calif., 1990) [Russ. original, 2nd ed., Nauka, Moscow, 1988].
[72] J. Zinn-Justin, Nucl. Phys. **B246**, 246 (1984).
[73] V. N. Gribov, Nucl. Phys. **B139**, 1 (1978).
[74] M. Yu. Logachev, Teor. Mat. Fiz. **70**, 412 (1987) [Theor. Math. Phys. (USSR)].
[75] M. A. Solov'ev, Teor. Mat. Fiz. **78**, 163 (1989) [Theor. Math. Phys. (USSR)].
[76] L. V. Prokhorov and S. V. Shabanov, Phys. Lett. **216B**, 341 (1989).
[77] R. P. Feynman, Acta Phys. Pol. **24**, 697 (1963).
[78] L. D. Faddeev and V. N. Popov, Phys. Lett. **25B**, 29 (1967).
[79] A. Bassetto, G. Nardelli, and R. Soldati, *Yang–Mills Theories in Algebraic Noncovariant Gauges* (World Scientific, Singapore, 1991).
[80] W. Pauli, Phys. Rev. **58**, 716 (1940).
[81] S. Coleman, in *The Whys of Subnuclear Physics*, edited by A. Zichichi (Plenum Press, New York, 1979).
[82] P. A. M. Dirac, Proc. R. Soc. London, Ser. A **133**, 60 (1931).
[83] *The Dirac Monopole (A Collection)* [Russ. transl., Mir, Moscow, 1970].
[84] P. Goddard and D. I. Olive, Rep. Prog. Phys. **41**, 1357 (1978).
[85] Yu. A. Gol'fand and E. P. Likhtman, Pis'ma Zh. Éksp. Teor. Fiz. **13**, 452 (1971) [JETP Lett. **13**, 323 (1971)].
[86] D. V. Volkov and V. P. Akulov, Pis'ma Zh. Éksp. Teor. Fiz. **16**, 621 (1972) [JETP Lett. **16**, 438 (1972)].
[87] J. Wess and B. Zumino, Phys. Lett. **49B**, 52 (1974).
[88] J. Wess and B. Zumino, Nucl. Phys. **B70**, 39 (1974).
[89] A. Neveu and J. H. Schwarz, Nucl. Phys. **B31**, 86 (1971).
[90] J. L. Gervais and B. Sakita, Nucl. Phys. **B34**, 632 (1971).
[91] B. Zumino, Preprint Th-1779, CERN, Geneva (1973).
[92] Y. Aharonov, A. Casher, and L. Susskind, Phys. Lett. **35B**, 512 (1971).
[93] D. Z. Freedman, P. van Nieuwenhuizen, and S. Ferrara, Phys. Rev. D **13**, 3214 (1976).
[94] S. Deser and B. Zumino, Phys. Lett. **62B**, 355 (1976).
[95] L. V. Prokhorov, Usp. Fiz. Nauk **154**, 299 (1988) [Sov. Phys. Usp. **31**, 151 (1988)].
[96] P. A. M. Dirac, *Lectures on Quantum Mechanics* (Yeshiva University, New York, 1964) [Russ. transl., Mir, Moscow, 1968].
[97] P. A. M. Dirac, Can. J. Math. **2**, 129 (1950).
[98] P. Bergmann, Rev. Mod. Phys. **33**, 510 (1961).
[99] L. V. Prokhorov, Fiz. Élem. Chastits At. Yadra **25**, 559 (1994) [Phys. J. Part. Nucl. **25**, 235 (1994)].
[100] P. A. M. Dirac, *The Principles of Quantum Mechanics*, 4th ed. (Clarendon Press, Oxford, 1958) [Russ. transl., Nauka, Moscow, 1960].
[101] J. Goldstone and R. Jackiw, Phys. Lett. **74B**, 81 (1978).
[102] S. Weinberg, Phys. Rev. **135**, B1049 (1964).
[103] L. V. Prokhorov and S. V. Shabanov, No. B1-2-93-312, Dep. publ. 11-08-93, JINR (1993) [in Russian].
[104] L. V. Prokhorov, Yad. Fiz. **35**, 229 (1982) [Sov. J. Nucl. Phys. **35**, 129 (1982)].
[105] D. P. Zhelobenko, *Compact Lie Groups and Their Representations* (American Mathematical Society, Providence, R.I., 1973) [Russ. original, Nauka, Moscow, 1970].
[106] O. Loos, *Symmetric Spaces* [Russ. transl., Mir, Moscow, 1985].
[107] L. V. Prokhorov, Fiz. Élem. Chastits At. Yadra **13**, 1094 (1982) [Sov. J. Part. Nucl. **13**, 456 (1982)].
[108] A. Burnel, Phys. Rev. D **26**, 442 (1982).
[109] A. Burnel, Phys. Rev. D **32**, 450 (1985).
[110] S. N. Gupta, Proc. Phys. Soc. Ser. A **63**, 681 (1950).
[111] K. Bleuler, Helv. Phys. Acta **23**, 567 (1950).
[112] S. N. Gupta, Can. J. Phys. **35**, 961 (1957).
[113] S. Weinberg, Phys. Rev. **134**, B1049 (1964).
[114] D. N. Gitman and I. V. Tyutin, *Canonical Quantization of Fields with Constraints* [in Russian] (Nauka, Moscow, 1986).
[115] L. V. Prokhorov, D. V. Fursaev, and S. V. Shabanov, Teor. Mat. Fiz. **97**, 373 (1993) [Theor. Math. Phys. (USSR)].
[116] A. V. Radyushkin, Fiz. Élem. Chastits At. Yadra **14**, 58 (1983) [Sov. J. Part. Nucl. **14**, 23 (1983)].
[117] L. V. Prokhorov, Vestn. Leningr. Univ. No. 18, 3 (1990) [in Russian].
[118] F. Bloch and A. Nordsieck, Phys. Rev. **52**, 54 (1937).
[119] T. Murota, Prog. Theor. Phys. **24**, 1109 (1960).
[120] S. V. Ivanov and G. P. Korchemsky, Phys. Lett. **154B**, 197 (1985).
[121] A. S. Shvarts, *Quantum Field Theory and Topology* [in Russian] (Nauka, Moscow, 1989).
[122] R. P. Feynman, Phys. Rev. **76**, 769 (1949).
[123] R. P. Feynman, in *Recent Developments in Quantum Electrodynamics* [in Russian] (IL, Moscow, 1954), p. 184.
[124] R. P. Feynman, *Theory of Fundamental Processes* (Benjamin, New York, 1962) [Russ. transl., Nauka, Moscow, 1978].
[125] N. H. Crist and T. D. Lee, Phys. Rev. D **22**, 939 (1980).
[126] Yu. P. Malyshev and L. V. Prokhorov, Vestn. Leningr. Univ. No. 18, 99 (1986) [in Russian].
[127] E. P. Wigner, *Symmetries and Reflections: Scientific Essays of Eugene P. Wigner* (MIT Press, Cambridge, Mass., 1970) [Russ. transl., Mir, Moscow, 1971].
[128] E. Noether, in *Variational Principles of Mechanics* [in Russian] (IL, Moscow, 1959).
[129] H. Goldstein, *Classical Mechanics*, 2nd ed. (Addison-Wesley, Reading, Mass., 1980) [Russ. transl. of 1st ed., Nauka, Moscow, 1975].
[130] L. V. Prokhorov, Lett. Math. Phys. **19**, 245 (1990).
[131] L. V. Prokhorov, Vestn. S. Petersburg. Gos. Univ. No. 18, 87 (1994) [in Russian].
[132] D. A. Leĭtes, *Theory of Supermanifolds* [in Russian] (Petrozavodsk, 1983).
[133] I. M. Singer, Commun. Math. Phys. **60**, 7 (1978).
[134] D. Zwanziger, Nucl. Phys. **B192**, 259 (1981).
[135] S. Elitzur, Phys. Rev. D **12**, 3978 (1975).
[136] L. V. Prokhorov and S. V. Shabanov, Vestn. Leningr. Univ. No. 4, 3 (1990) [in Russian].
[137] H. Cartan, *Differential Forms*, transl. from the French (Hermann, Paris, 1970) [Russ. transl., Mir, Moscow, 1971].
[138] N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields*, 3rd ed. (Wiley, New York, 1980) [Russ. original, 4th ed., Nauka, Moscow, 1984].
[139] Yu. A. Khramov, *Physics. A Biographical Handbook* [in Russian], 2nd ed. (Nauka, Moscow, 1983).
[140] W. Kummer, Acta. Phys. Austriaca **14**, 149 (1961).
[141] J. Schwinger, Phys. Rev. **82**, 664 (1951).
[142] V. A. Fateev, A. S. Schwarz, and Yu. S. Tyupkin, Preprint No. 155, Lebedev Phys. Inst. (1976).

[143] W. E. Brittin, W. R. Smythe, and W. Wyss, Am. J. Phys. **50**, 693 (1982).
[144] M. S. Dubovikov and A. V. Smilga, Nucl. Phys. **B185**, 109 (1981).
[145] E. A. Power and S. Zienau, Nuovo Cimento **6**, 7 (1957).
[146] E. A. Power and S. Zienau, Philos. Trans. R. Soc. London, Ser. A **251**, 427 (1959).
[147] K.-H. Yang, Ann. Phys. (N.Y.) **186**, 209 (1988).
[148] L. V. Prokhorov, Yad. Fiz. **16**, 854 (1972) [Sov. J. Nucl. Phys. **16**, 473 (1972)].

Translated by Patricia A. Millard