Theoretical and Mathematical Physics

Franco Strocchi

Symmetry Breaking

Third Edition

Springer

Symmetry Breaking

Theoretical and Mathematical Physics

This series, founded in 1975 and formerly entitled (until 2005) *Texts and Monographs in Physics* (TMP), publishes high-level monographs in theoretical and mathematical physics. The change of title to *Theoretical and Mathematical Physics* (TMP) signals that the series is a suitable publication platform for both the mathematical and the theoretical physicist. The wider scope of the series is reflected by the composition of the editorial board, comprising both physicists and mathematicians.
The books, written in a didactic style and containing a certain amount of elementary background material, bridge the gap between advanced textbooks and research monographs. They can thus serve as a basis for advanced studies, not only for lectures and seminars at graduate level, but also for scientists entering a field of research.

More information about this series at http://www.springer.com/series/720

Franco Strocchi

Symmetry Breaking

Third Edition

Springer

Franco Strocchi
Department of Physics
Pisa University
Pisa, Italy

ISSN 1864-5879 ISSN 1864-5887 (electronic)
Theoretical and Mathematical Physics
ISBN 978-3-662-62164-6 ISBN 978-3-662-62166-0 (eBook)
https://doi.org/10.1007/978-3-662-62166-0

Originally published as volume 643 in the Springer series Lecture Notes in Physics

This Springer imprint is published by the registered company Springer-Verlag GmbH, DE part of Springer Nature.
The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

Preface to the Third Edition

The main motivation for having this third edition is the addition of an Appendix which deepens and clarifies some of the concepts introduced and developed in the core text. In the discussion of the breaking of continuous symmetries, the crucial point of the dynamical delocalization of the relevant variables does not seem to be sufficiently emphasized in the literature. Its impact on the local generation of a continuous symmetry, at the basis of Goldstone theorem, has been discussed in Part II, Sect. 25.2, but, in our opinion, this argument deserves more consideration, since the Coulomb delocalization is the critical one and, as such, affects most many-body systems and gauge theories of elementary particles. The main consequence is the key role of the boundary conditions which give rise to volume effects, in contrast with the standard case described by (essentially) local dynamical variables. The relevant physical result is the occurrence of energy gaps associated to the spontaneous breaking of continuous symmetries, in contrast with the Goldstone theorem. Examples are the Anderson model of superconductivity (close to the molecular field approximation of the Heisenberg model), the electron gas in uniform background (with the plasmon energy gap derived by the breaking of the Galilei group), the Higgs mechanism in the Coulomb gauge.

Other sections of the Appendix deal with the consequences and breaking of global gauge symmetries. Invariance under a local gauge symmetry is critically reviewed versus the validity of a local Gauss law on the physical states. This is argued to be the relevant property responsible for important physical properties like the Higgs mechanism in electroweak interactions, the topological group and the chiral symmetry breaking in quantum chromodynamics.

In line with the approach adopted in the previous editions, attention is paid to the main ideas and to the questions of principle leaving to the interested reader the task and duty of further developing the technical mastership.

Lastly, this new edition benefits from the correction of misprints and some refinements of the main text.

Pisa, Italy — Franco Strocchi

Preface to Previous Edition

The main motivation for such lecture notes is the importance of the concept and mechanism of spontaneous symmetry breaking in modern theoretical physics and the relevance of a textbook exposition at the graduate student level beyond the oversimplified (non-rigorous) treatments, often confined to specific models. One of the main points is to emphasize that the radical loss of symmetric behaviour requires both the existence of non-symmetric ground states and the infinite extension of the system.

The first Part on SYMMETRY BREAKING IN CLASSICAL SYSTEMS is devoted to the mathematical understanding of spontaneous symmetry breaking on the basis of classical field theory. The main points, which do not seem to appear in textbooks, are the following.

i) **Existence of disjoint Hilbert space sectors**, stable under time evolution, in the set of solutions of the classical (non-linear) field equations. They are the strict analogs of the different phases of statistical mechanical systems and/or of the inequivalent representations of local field algebras in quantum field theory (QFT). As in QFT, such structures rely on the concepts of locality (or localization) and stability, (see Chap. 5), with emphasis on the physical motivations of the mathematical concepts; such structures have the physical meaning of *disjoint physical worlds*, disjoint *phases* etc. which can be associated to a given non-linear field equation. The result of Theorem 5.2 may be regarded as a generalization of the criterium of stability to infinite dimensional systems and it links such stability to elliptic problems in $\mathbf{R}^n$ with non-trivial boundary conditions at infinity (Appendix 10.5).
ii) Such structures allow to reconcile the classical **Noether theorem** with spontaneous **symmetry breaking**, through a discussion of a mechanism which accounts for (and explains) the breaking of the symmetry group (of the equations of motion), in a given Hilbert space sector $\mathcal{H}$, down to the subgroup which leaves $\mathcal{H}$ stable (Theorem 7.2).

iii) The classical counterpart of the **Goldstone theorem** is proved in Chap. 9, which improves and partly corrects the heuristic perturbative arguments of the literature.

The presentation emphasizes the general ideas (implemented in explicit examples) without indulging on the technical details, but also without derogating from the mathematical soundness of the statements.

The second Part on “SYMMETRY BREAKING IN QUANTUM SYSTEMS” tries to offer a presentation of the subject, which should be more mathematically sounded and convincing than the popular accounts, but not too technical. The first chapters are devoted to the general structures which arise in the quantum description of infinitely extended systems with emphasis on the physical basis of *locality, asymptotic abelianess* and *cluster property* and their mutual relations, leading to a characterization of the **pure phases**.

Criteria of spontaneous symmetry breaking are discussed in Chap. 20 along the lines of Wightman lectures at Coral Gables and their effectiveness and differences are explicitly worked ot and checked in the Ising model. The Bogoliubov strategy is shown to provide a simple rigorous control of spontaneous symmetry breaking in the free Bose gas as a possible alternative to Cannon and Bratelli-Robinson treatment.

The **Goldstone theorem** is critically discussed in Chap. 25, especially for non-relativistic systems or more generally for systems with long range delocalization. Such analysis, which does not seem to appear in textbooks, provides a non-perturbative explanation of *symmetry breaking with energy gap* in **non-relativistic Coulomb systems** and in the **Higgs phenomenon** and in our opinion puts in a more convincing and rigorous perspective the analogies proposed by Anderson. The Swieca conjecture about the role of the potential fall off is checked by a perturbative expansion in time. Such an expansion also supports the condition of integrability of the charge density commutators, which seems to be overlooked in the standard treatments and plays a crucial role for the energy spectrum of the Goldstone bosons. As a result of such an explicit analysis, the *critical decay of the potential* for allowing “massive” Goldstone bosons turns out to be that of the Coulomb potential, rather than the one power faster decay predicted by Swieca condition.

The *non-zero temperature* version of the *Goldstone theorem,* discussed in Chap. 26, corrects some wrong conclusions of the literature. An extension of the Goldstone theorem to non-symmetric Hamiltonians is discussed in Chap. 28 with the derivation of non-trivial (non-perturbative) information on the energy gap of the modified Goldstone spectrum.

The symmetry breaking in gauge theories, in particular the **Higgs phenomenon** which is at the basis of the standard model of elementary particles, is analyzed in Chap. 29. The problems of the perturbative explanation of the evasion of the Goldstone theorem are pointed out and a non-perturbative account is presented. In the local renormalizable gauges, the absence of physical Goldstone bosons follows from the Gauss law constraint or subsidiary condition on the physical states. In the

Coulomb gauge, the full Higgs phenomenon is explained by the failure of relative locality between the current and the Higgs field, by exactly the same mechanism discussed in Chap. 25 for the non-relativistic Coulomb systems; in particular, the Goldstone spectrum is shown to be given by the Fourier spectrum of the two point function of the vector boson field, which cannot have a $\delta(k^2)$ contribution, since otherwise the symmetry would not be broken.

The chapters marked with a * can be skipped in a first reading.

The second edition differs from the first by the correction of some misprints, by an improved discussion of some relevant points and by a significantly expanded and more detailed discussion of symmetry breaking in gauge theories.

Contents

Part I
SYMMETRY BREAKING IN CLASSICAL SYSTEMS

Introduction

These notes essentially reproduce lectures given at the International School for Advanced Studies (Trieste) and at the Scuola Normale Superiore (Pisa) on various occasions. The scope of the short series of lectures, typically a fraction of a one-semester course, was to explain on general grounds, also to mathematicians, the phenomenon of Spontaneous Symmetry Breaking (SSB), a mechanism which seems at the basis of most of the recent developments in theoretical physics (from Statistical Mechanics to Many-Body theory and to Elementary Particle theory).

Besides its extraordinary success, the idea of SSB also deserves being discussed because of its innovative philosophical content, and in our opinion it should be part of the background knowledge for mathematical and theoretical physics students, especially those who are interested in questions of principle and in general mathematical structures.

By the general wisdom of Classical Mechanics, codified in the classical Noether theorem, one learns that the symmetries of the Hamiltonian or of the Lagrangian are automatically symmetries of the physical system described by it, which does not mean that the (equilibrium) solutions are symmetric, but rather that the symmetry transformation commutes with time evolution and hence is a symmetry of the physical behaviour of the system. This belief therefore precludes the possibility of describing systems with different dynamical properties in terms of the same Hamiltonian. The realization that this obstruction does not *a priori* exist and that one may unify the description of apparently different systems in terms of a single Hamiltonian and account for the different behaviours by the mechanism of SSB, is a real revolution in the way of thinking in terms of symmetries and corresponding properties of physical systems. It is, in fact, non-trivial to understand how the conclusions of the Noether theorem can be evaded and how a symmetry of the dynamics cannot be realized as a mapping of the physical configurations of the system, which commutes with the time evolution.

The standard folklore explanations of SSB, which one often finds in the literature, is partly misleading, because it does not emphasize the crucial ingredient underlying the phenomenon, namely the role of infinite degrees of freedom. Despite the many popular accounts, the phenomenon of SSB is deep and subtle and it is not without reasons that it has been fully understood only in recent times. The standard cheap explanation identifies the phenomenon with the existence of a degenerate ground (or equilibrium) state, unstable under the symmetry operation, (*ground state asymmetry*), a feature often present even in simple mechanical models (as, for example, a free particle on a plane, each point of which defines a ground state unstable under translations), but which is usually not accompanied by a non-symmetric behaviour.

As it will be discussed in these lectures, the phenomenon of spontaneous symmetry breaking in the radical sense of non-symmetric behaviour is rather related to the fact that, for non-linear infinitely extended systems (therefore involving infinite degrees of freedom), the solutions of the dynamical problem generically fall into classes or “islands” or “phases”, each stable under time evolution and characterized by the same behaviour at infinity of the corresponding solutions. Since all physically realizable operations have an inevitable localization in space they cannot change such a behaviour at infinity and therefore starting from the configurations of a given island one cannot reach the configurations of a different island by physically realizable modifications. The different islands can then be interpreted as describing physically disjoint realizations or *different phases*, or *disjoint physical worlds* associated with the given dynamics.

The spontaneous breaking of a symmetry (of the dynamics) in a given phase or physical world can then be explained as the result of the instability of the given island under the symmetry operation. In fact, in this case one cannot realize the symmetry within the given island, namely one cannot operationally associate with each configuration the one obtained by the symmetry operation.

The existence of such structures is not obvious and in general it involves a mathematical control of the non-linear time evolution of systems with infinite degrees of freedom and the mathematical formalization of the concept of physical disjointness of different islands. For quantum systems, where the mathematical basis of SSB has mostly been discussed, the physical disjointness has been ascribed to the existence of inequivalent representations of the algebra of *local* observables.

The scope of Part I of these lectures is to discuss the general mechanism of SSB within the framework of classical dynamical systems, so that no specific knowledge of quantum mechanics of infinite systems is needed and the message may also be suitable for mathematical students. More specifically, the discussion will be based on the mathematical control of the non-linear evolution of classical fields, with *locally* square integrable initial data which may possibly have non-vanishing limits at infinity.

The mathematical formalization of physical disjointness relies on the constraint of essential localization in space of any physically realizable operation. One can in fact show that an island can be characterized by some bounded (locally “regular”) reference configuration, having the meaning of the “ground state”, and its H^1 perturbations. Each island is therefore isomorphic to a Hilbert space (*Hilbert space sector*).

The stability under time evolution is guaranteed by the condition that the reference configuration satisfies a generalized stationarity condition, i.e. it solves some elliptic problem. Such a condition is in particular satisfied by the time-independent solutions and *a fortiori* by the minima $\bar{\varphi}$ of the potential which define Hilbert space sectors $\mathcal{H}_{\bar{\varphi}}$ of the form $\{\bar{\varphi} + \chi,\, \chi \in H^1\}$. The existence of minima of the potential unstable under the symmetry therefore gives rise to phases or disjoint physical worlds in which the *symmetry* cannot be realized or, as one says, is *spontaneously broken*. This mechanism of symmetry breaking crucially involves both the asymmetry of the ground state *and* the infinite extension of the system, with no analog in the finite-dimensional case.

This phenomenon is deeply rooted in the non-linearity of the problem and the fact that infinite degrees of freedom are involved. A simple prototype is given by the non-linear wave equation for a Klein–Gordon field φ: $\mathbf{R}^s \to \mathbf{R}^n$, with "potential" $U(\varphi) = \lambda(\varphi^2 - a^2)^2$. The model displays some analogy with the mechanical model of a particle in $\mathbf{R}^n$ subject to the potential $U(q) = \lambda(q^2 - a^2)^2$, which can be regarded as the higher dimensional version of the one-dimensional double well potential. But the differences are substantial: in the infinite-dimensional case of the Klein–Gordon field, each point q has actually become infinite dimensional and, in fact, each absolute minimum $\bar{\varphi}$, with $|\bar{\varphi}| = a$ identifies the infinite set of configurations which have this point as asymptotic limit, namely the Hilbert space of configurations which are H^1 modifications of $\bar{\varphi}$. Whereas in the finite-dimensional case there is no physical obstruction or "barrier", which prevents the motion from one minimum to the other, in the infinite-dimensional case there is no physically realizable operation which leads from the Hilbert space sector defined by one minimum to that defined by another minimum, because this would require to change the asymptotic limit of the configurations and this is not possible by means of essentially localized operations, the only ones which are physically realizable. Pictorially, one could say that one cannot change the boundary conditions of the "universe" or of the (infinite volume) thermodynamical phase in which one is living.

The realization of the above structures allows to evade part of the conclusions of the standard textbook presentations of Noether's theorem and to account for spontaneous symmetry breaking; the point is that the standard presentations of the theorem do not consider the possibility of disjoint sectors unstable under the symmetry of the Hamiltonian and implicitly assume that the solutions vanish at infinity. In fact, one may prove that the *local conservation law*, $\partial^\mu j_\mu(x) = 0$, associated with a given symmetry of the Hamiltonian or of the Lagrangian, gives rise to a *global conservation law* or to a conserved "charge", which acts as the generator of the symmetry transformations for all the elements of a given Hilbert space sector $\mathcal{H}_{\bar{\varphi}}$, only if the symmetry leaves the sector invariant. Thus, only the stability subgroup of the given phase admits time-independent generators in that phase, given by the charges of the corresponding Noether currents.

Clearly, if G is the (concrete) group of transformations which commutes with the time evolution, the whole set of solutions of the non-linear dynamical problem can be classified in terms of irreducible representations (or multiplets) of G, but if G is spontaneously broken in a given island defined by the Hilbert space sector $\mathcal{H}_{\bar{\varphi}}$, the latter cannot be the carrier of a representation of the symmetry group G, and in particular the elements of $\mathcal{H}_{\bar{\varphi}}$ cannot be classified in terms of multiplets of G.

One might think of grouping together solutions corresponding to initial data of the form $\bar{\varphi} + g\chi$, $g \in G$, which might look like candidates for multiplets of G. As a matter of fact, such sets of initial data do correspond to representations of a group of transformations which is isomorphic to G, but which does *not* commute with the dynamics, and therefore the above form of the initial data does not extend to arbitrary times; thus the above identification of multiplets at the initial time is not stable under time evolution. As a matter of fact, the group of transformations which commute with the time evolution corresponds to $\bar{\varphi} + \chi \to g\bar{\varphi} + g\chi$, $g \in G$, which, however, does not leave $\mathcal{H}_{\bar{\varphi}}$ stable.

Within this approach, it is possible to prove a classical counterpart of the so-called Goldstone theorem, according to which there are massless modes (i.e. solutions of the free wave equation) associated to each broken generator. The theorem proved here provides a mathematically acceptable substitute of the heuristic arguments and improves the conclusions based on the quadratic approximation of the potential around an absolute minimum.

Explicit examples which illustrate how these ideas work in concrete models are discussed in Chap. 8.

The discussion of symmetry breaking in classical systems relies, with some additions, on papers written jointly with Cesare Parenti and Giorgio Velo, to whom I am greatly indebted (see the references at the relevant points). An attempt is made to reduce the mathematical details to the minimum required to make the arguments self-contained and also convincing for a mathematically minded reader. The required background technical knowledge is kept to a rather low level, in order that the lectures be accessible also to undergraduate students with a basic knowledge of Hilbert space structures.

Chapter 1
Symmetries of a Classical System

The realization of symmetries in physical systems has proven to be of help in the description of physical phenomena: it makes it possible to relate the behaviour of similar systems and therefore it leads to a great simplification of the mathematical description of Nature.

The simplest concept of symmetry occurs at the geometrical or kinematical level when the shape of an object or the configuration of a physical system is invariant or symmetric under geometric transformations like rotations, reflections, etc. At the dynamical level, a system is symmetric under a transformation of the coordinates or of the parameters which identify its configurations, if correspondingly its dynamical behaviour is symmetric in the sense that the action of the symmetry transformation and of time evolution commute.

To formalize the concept of dynamical symmetry, we first recall that the *description of a classical physical system* consists in

i) the identification of all its possible configurations $\{S_\gamma\}$, with γ running over an index set of coordinates or parameters which identify the configuration S_γ;
ii) the determination of their time evolution

$$\alpha^t : S_\gamma \to \alpha^t S_\gamma \equiv S_{\gamma(t)}. \tag{1.1}$$

A *symmetry* g of a physical system is a transformation of the coordinates (or of the parameters) γ, g: $\gamma \to g\gamma$, which

1) induces an invertible mapping of configurations

$$g : S_\gamma \to gS_\gamma \equiv S_{g\gamma} \tag{1.2}$$

F. Strocchi, *Symmetry Breaking*, Theoretical and Mathematical Physics,
https://doi.org/10.1007/978-3-662-62166-0_1

2) does not change the dynamical behaviour,[1] namely

$$\alpha^t g S_\gamma = \alpha^t S_{g\gamma} \equiv S_{(g\gamma)(t)} = S_{g\gamma(t)} = g\alpha^t S_\gamma. \tag{1.3}$$

The above condition states that the symmetry transformation commutes with time evolution. For classical canonical systems, this amounts to the invariance of the Hamiltonian under the symmetry g (*symmetric Hamiltonian*).

The realization of a symmetry which relates (the configurations of) two seemingly different systems clearly leads to a unification of their description. In particular, the solution of the dynamical problem for one configuration automatically gives the solution for the symmetry related configuration (see (1.3)).

Example 1.1, double well potential. Consider a particle moving on a line, subject to a double well potential, i.e. described by the following Hamiltonian

$$H = \tfrac{1}{2}p^2 + \tfrac{1}{4}\lambda(q^2 - a^2)^2, \tag{1.4}$$

with q, p the canonical coordinates which label the configurations of the particle. The reflection $g : q \to -q,\;\; p \to -p$ leaves the Hamiltonian invariant and is a symmetry of the system; obviously, it maps solutions (of the Hamilton equations) into solutions.

Now, consider the two classes $\Gamma_\pm$ of solutions corresponding to initial conditions in the neighbourhoods of the two absolute minima $q_0 = \pm a$, with $p_0 < \sqrt{\lambda}a^2/2$, respectively, and suppose that by some (artificial) *ansatz*, in the preparation of the initial configurations one cannot dispose of energies greater than $\lambda a^4/4$. This means that the two classes of solutions describe two disjoint realizations of the system, in the sense that *by fiat* no physically realizable operation allows to change a configuration from one class to the other. In this way, one gets a picture similar to the case of the thermodynamical phases, which are physically disjoint in the thermodynamical limit, but nevertheless described by the same Hamiltonian and related by a symmetry which is not implementable in each phase. Clearly, the existence of such a symmetry, even if devoid of physical operational meaning, provides a unified description of the two “phases”.

For a particle moving on a plane, the analog of the double well potential defines a Hamiltonian which is invariant under rotations around the axis (through the origin) orthogonal to the plane and one has a continuous group of symmetries. There is a continuous family of absolute minima lying on the circle $|\boldsymbol{q}_0|^2 = a^2$. Since such minima are not separated by any energy barrier, one cannot associate with them different systems by some artificial *ansatz* as above.

[1]To simplify the discussion, here we do not consider the more general case in which the dynamics transform covariantly under g (like, e.g. in the case of Lorentz transformations). For a general discussion of symmetries and of their relevance in physics, see R.M.F. Houtappel, H. Van Dam and E.P. Wigner, Rev. Mod. Phys. **37**, 595 (1965).

Chapter 2
Spontaneous Symmetry Breaking

One of the most powerful ideas of modern theoretical physics is the mechanism of spontaneous symmetry breaking. It is at the basis of most of the recent achievements in the description of phase transitions in Statistical Mechanics as well as of collective phenomena in solid state physics. It has also made possible the unification of weak, electromagnetic and strong interactions in elementary particle physics. Philosophically, the idea is very deep and subtle (this is probably why its exploitation is a rather recent achievement) and the popular accounts do not fully do justice to it.

Roughly, spontaneous symmetry breaking is said to occur when a symmetry of the Hamiltonian, which governs the dynamics of a physical system, does not lead to a symmetric description of the physical properties of the system. At first sight, this may look almost paradoxical. From elementary courses on mechanical systems, one learns that the symmetries of a system are displayed by the symmetries of the Hamiltonian, which describes its time evolution; how can it then be that a symmetric Hamiltonian gives rise to an asymmetric physical description of a dynamical system?

The cheap standard explanation is that such a phenomenon is due to the existence of a non-symmetric absolute minimum or "ground state", but the mechanism must have a deeper explanation, since the symmetry of the Hamiltonian implies that an asymmetric stable point cannot occur by alone, (the action of the symmetry on it will produce another stable point). Now, the existence of a set of absolute minima related by a symmetry (or "degenerate ground states") does not imply a non-symmetric physical description. One actually gets a symmetric picture, if the correct correspondence is made between the configurations of the system (and their time evolutions), and such a correspondence is physically implementable if for any physically realizable configuration its transformed one is also realizable.

The way out of this argument is to envisage a mechanism by which, given a non-symmetric absolute minimum (or "ground" state) S_0, there are physical obstructions to reach its transformed one, $g\ S_0$, by means of physically realizable operations, so that effectively one gets confined to an asymmetric realization of the system. The

F. Strocchi, *Symmetry Breaking*, Theoretical and Mathematical Physics,
https://doi.org/10.1007/978-3-662-62166-0_2

purpose of the following discussion is to make such a rather vague and intuitive picture more precise.

For a classical finite-dimensional dynamical system, two configurations may be said to be related by *physically realizable operations* if there is no physical obstruction for operationally changing one into the other, e.g. if they are connected by a continuous path of configurations, all with finite energy. In this way, one gets a partition of the configurations into classes and given a configuration S, the set of configurations which can be reached from it, by means of physically realizable operations, will be called the *phase* Γ_S, or the "physical world", to which S belongs.

A *symmetry* g will be said to be *physically realized* (or *implementable* or *unbroken*), in the phase Γ, if it leaves Γ stable.

In the mechanical example of the double well potential discussed above, there is no natural and physically reasonable way of isolating the solutions in the neighbourhoods of the two minima, since an artificial limitation of the available energies looks rather unphysical. Actually, according to the above definitions, there is only one phase and the reflection symmetry is physically implementable or unbroken.

In order to further illustrate the above definitions, we consider a particle moving on a line, subject to a deformed double well potential, still invariant under the reflection $g : q \to -q$, with two absolute minima at $q_0 = \pm a$, but going to infinity as $q \to 0$.

Consider now two kinds of (one-dimensional) creatures, one living in the valley with bottom $q_0 = a$ and the other in the valley with bottom $q_0 = -a$. The infinite potential barrier prevents going from one valley to the other (tunnelling is impossible); then, e.g. the people living in the r.h.s. valley do not have access to the l.h.s. valley, neither by action on the initial conditions of the particle nor by time evolution. Thus, the operations which are physically realizable (by each of the two kinds of people) cannot make the transition from one valley to the other and the particle configurations get divided into two phases, labelled by the two minima Γ_a, Γ_{-a}, respectively.

The reflection symmetry is not physically realized in each of the two phases. As a matter of fact, even if the particle motion is described by a symmetric Hamiltonian, the particle physical world will look asymmetric to each kind of creatures: the *symmetry* is *spontaneously broken*.

The somewhat artificial example of spontaneous symmetry breaking discussed above is made possible by the infinite potential barrier between the two absolute minima. Clearly, such a mechanism is not available in the case of a continuous symmetry, since then the (absolute) minima are continuously related by the symmetry group and no potential barrier can occur between them (for a concrete example see the two-dimensional double well discussed above). Thus, for finite-dimensional classical dynamical systems, a continuous symmetry of the Hamiltonian is always unbroken (even if the ground state is degenerate and non-symmetric).

The often-quoted example of a particle in a two-dimensional double well potential is a somewhat misleading example of spontaneous breaking of continuous symmetry (it is also an incorrect example in one dimension, unless the potential is so deformed to produce an an infinite barrier between the two minima). Actually, most of the claimed simple mechanical examples of spontaneous symmetry breaking discussed in the literature are equally misleading.

Even if the existence of non-symmetric minima is a rather peculiar phenomenon which deserves special interest, it does not imply spontaneous symmetry breaking in the radical sense of its realization in elementary particle physics, many-body systems, statistical mechanics, etc., where a symmetry of the dynamics is not shared by physical realizations or disjoint phases of the system. This is a much deeper phenomenon than the mere existence of *non-symmetric minima*.

The relevance of the distinction between *non-symmetric minima or ground states* and *spontaneous symmetry breaking* appears clear if one considers, e.g. a free particle on a line, where each configuration ($q_0 \in \mathbb{R},\ p_0 = 0$) is a minimum of the Hamiltonian and it is not stable under translations, but nevertheless one does not speak of symmetry breaking; in fact, according to our definition, there is only one phase stable under translations.

The two concepts of symmetry breaking coincide for infinitely *extended systems*, since in this case, as we shall see below, different ground states define different phases or disjoint worlds; therefore their asymmetry necessarily leads to symmetry breaking in the radical sense of a non-symmetric physical description (see Chap. 7 below).

Similar considerations apply to classical systems which exhibit bifurcation[2] for which, strictly speaking, one does not have spontaneous symmetry breaking as long as the multiple solutions are related by physically realizable operations. As we shall see later, the latter property may fail if one considers the infinite volume (or thermodynamical) limit, and in this way spontaneous symmetry breaking may occur.

[2]D.H. Sattinger, Spontaneous Symmetry Breaking: mathematical methods, applications and problems in the physical sciences, in *Applications of Non-Linear Analysis*, H. Amann et al. eds., Pitman 1981.

Chapter 3
Symmetries in Classical Field Theory

As the previous discussion indicates, it is impossible to realize the phenomenon of (spontaneous) breaking of a continuous symmetry in classical mechanical systems with a finite number of degrees of freedom described by canonical variables. We are thus led to consider infinite-dimensional systems, like classical fields.

Our main purpose is to recognize the existence of disjoint "phases", in the set of solutions of the classical field equations, with the interpretation of possible disjoint realizations of the system (Chap. 5). The phenomenon of spontaneous symmetry breaking in a given "phase" will then be explained by its instability under the symmetry transformation.

To simplify the discussion, we will focus our attention to the standard case of the non-linear equation

$$\Box\varphi + U'(\varphi) = 0, \tag{3.1}$$

where $\Box \equiv (\partial_t)^2 - \Delta,\ \varphi = \varphi(x,\ t), x \in \mathbb{R}^s,\ t \in \mathbb{R}$, is a field taking values in $\mathbb{R}^n$, (an n-component field), $U(\varphi)$ is the potential, which for the moment will be assumed to be sufficiently regular, and U' denotes its derivative.

Equation (3.1) can be derived by the stationarity of the following action integral

$$\mathcal{A}(\varphi, \dot{\varphi}) = \int d^s x\, dt \left[-\tfrac{1}{2}(\nabla\varphi)^2 + \tfrac{1}{2}\dot{\varphi}^2 - U(\varphi)\right].$$

A typical prototype is given by

$$U(\varphi) = \tfrac{1}{4}\lambda(\varphi^2 - a^2)^2 \tag{3.2}$$

which is the infinite-dimensional version of the double well potential discussed in Chap. 1.

F. Strocchi, *Symmetry Breaking*, Theoretical and Mathematical Physics,
https://doi.org/10.1007/978-3-662-62166-0_3

Quite generally, (3.1) occurs in the description of non-linear waves in many branches of physics like non-linear optics, plasma physics, hydrodynamics, elementary particle physics etc.[3] The above equation (3.1) will be used to illustrate general structures likely to be shared by a large class of non-linear hyperbolic equations.

The solution of the Cauchy problem for the (in general non-linear) equation (3.1), with given initial data

$$\varphi(x,\ t = 0) = \varphi_0(x), \quad \partial_t\varphi(x,\ t = 0) = \psi_0(x), \tag{3.3}$$

provides the corresponding classical field $\varphi(x,\ t)$ described by (3.1).

In analogy with the previous discussion of the finite-dimensional systems, a description of the system (3.1) consists in the identification of the class of initial conditions, for which the time evolution is well defined. Deferring the mathematical details, we will now denote by X the functional space within which the Cauchy problem is well posed, i.e. such that for any initial data

$$u_0 = \begin{pmatrix} \varphi_0 \\ \psi_0 \end{pmatrix} \in X \tag{3.4}$$

there is a unique solution $u(x,\ t)$ continuous in time (in the topology of X, see below) and belonging to X for any t, briefly $u(x,\ t) \in C^0(X,\ \mathbb{R})$.

Thus, X can be regarded as describing the initial configurations of the system (3.1) and it is stable under time evolution.[4]

In analogy with the finite-dimensional case, a *symmetry* of the system (3.1) is an invertible mapping T_g of X onto X, which commutes with the time evolution. To simplify the discussion, we will make the technical assumption that T_g is a continuous mapping (in the X topology) of the form

$$T_g \begin{pmatrix} \varphi(x) \\ \\ \psi(x) \end{pmatrix} = \begin{pmatrix} g(\varphi(x)) \\ \\ J_g(\varphi(x))\psi(x) \end{pmatrix}, \tag{3.5}$$

[3]See, e.g. G.B. Whitham, *Linear and Non-Linear Waves*, J. Wiley, New York 1974; R. Rajaraman, Phys. Rep. **21C**, 227 (1975); S. Coleman, *Aspects of Symmetry*, Cambridge Univ. Press 1985, Chap. 6.

[4]For an extensive review on the mathematical problems of the non-linear wave equation see M. Reed, *Abstract non-linear wave equation*, Springer-Verlag, Heidelberg 1976. For the solution of the Cauchy problem for initial data not vanishing at infinity, a crucial ingredient for discussing spontaneous symmetry breaking, see C. Parenti, F. Strocchi and G. Velo, Phys. Lett. **59B**, 157 (1975); Ann. Scuola Norm. Sup. (Pisa), III, 443 (1976), hereafter referred as I. A simple account with some addition is given in F. Strocchi, in *Topics in Functional Analysis 1980-81*, Scuola Normale Superiore Pisa, 1982. For a beautiful review of the recent developments see W. Strauss, *Nonlinear Wave Equations*, Am. Math. Soc. 1989.

with g a diffeomorphism of $\mathbb{R}^n$ of class C^2 and J_φ the Jacobian matrix of g. Such symmetries are called *internal symmetries*, since they commute with space and time translations.[5]

Under general regularity assumptions on the potential, such that for infinitely differentiable initial data the corresponding solution of (3.1) is of class C^2 in the variables x and t, one gets a characterization of the internal symmetries of the system (3.1).

Theorem 3.1 [6]*Under the above assumption on U, any internal symmetry of the system (3.1) is characterized by a g which is an affine transformation*

$$g(z) = Az + a, \tag{3.6}$$

where $a, z \in \mathbb{R}^n$ and A is an $n \times n$ invertible matrix. Furthermore, the invariance of the action integral up to a scale factor requires

$$A^T A = \lambda \boldsymbol{1}, \tag{3.7}$$

with A^T the transpose of A and λ a suitable constant. $A,\ a,\ \lambda$, which depend on g, satisfy the following condition

$$U(Az + a) = \lambda U(z) + U(a). \tag{3.8}$$

Proof. The condition that $T_g \alpha^t\ u_0 = \alpha^t T_g u_0$ be a solution of (3.1), for any initial data u_0, implies[7]

$$\begin{aligned} 0 = \Box g_k(\varphi) + U_k'(g(\varphi)) = \\ = \frac{\partial^2 g_k}{\partial z_i \partial z_j}(\varphi)\partial^\mu \varphi_i \partial_\mu \varphi_j - \frac{\partial g_k}{\partial z_i}(\varphi) U_i'(\varphi) + U_k'(g(\varphi)). \end{aligned} \tag{3.9}$$

Choosing the initial data such that $\varphi_0(x) = \text{const} \equiv c,\ \psi_0(x) = 0$, for x in some region of $\mathbb{R}^s$, the first term of (3.9) vanishes there and one gets

$$- \frac{\partial g_k}{\partial z_i}(c) U_i'(c) + U_k'(g(c)) = 0. \tag{3.10}$$

Since c is arbitrary, the sum of the last two terms vanishes for any φ. Choosing now $\varphi_0(x) = c,\ \psi_0(x) = \text{const} = b,\ x \in V \subset \mathbb{R}^s$, one gets

---

[5]For the discussion of more general symmetries see C. Parenti, F. Strocchi and G. Velo, Comm. Math. Phys. **53**, 65 (1977), hereafter referred to as II; Phys. Lett. **62B**, 83 (1976).

[6]Ref. II (see footnote 4).

[7]We use the convention by which sum over dummy indices is understood; furthermore the relativistic notation is used: $\mu = 0, 1, 2, 3,\ \partial_0 = \partial/\partial t,\ \partial_i = \partial/\partial x^i,\ i = 1, 2, 3,\ \partial^\mu = g^{\mu\nu}\partial_\nu,\ g^{00} = 1 = -g^{ii},\ g^{\mu\nu} = 0$ if $\mu \neq \nu$.

$$\frac{\partial^2 g_k}{\partial z_i \partial z_j}(c) = 0, \ \forall c \in \mathbb{R}^n, \quad \text{i.e. } g(z) = Az + a.$$

Equation (3.9) then becomes

$$\frac{\partial}{\partial z_l} U(Az + a) = (A^T A)_{li} \frac{\partial}{\partial z_i} U(z).$$

The invariance of the action integral up to a scale factor requires $A^T A = \lambda \mathbf{1}$ and $U(Az + a) = \lambda U(z) +$ const; the normalization $U(0) = 0$ identifies the latter constant as $U(a)$.

Having characterized the possible symmetries of (3.1), we may now ask whether symmetry breaking can occur. For continuous groups this possibility seems to be in conflict with Noether's theorem.

Theorem 3.2 [8]*Let G be an N parameter Lie group of internal symmetries for the classical system (3.1), then there exist N conserved currents*

$$\partial^\mu J^a_\mu(x, \ t) = 0, \quad a = 1, \ldots, N \tag{3.11}$$

and N conserved quantities

$$Q^a(t) = \int d^s \, x \; J^a_0(x, \ t) = Q^a(0), \tag{3.12}$$

which are the generators of the corresponding one-parameter subgroups $\{g^a_\alpha, \alpha \in \mathbb{R}\}$ of symmetry transformations

$$\delta^a u \equiv d g^a_\alpha(u)/d\alpha|_{\alpha=0} = \{u, \ Q^a\}, \tag{3.13}$$

where the curly brackets denote the Poisson brackets.

For the proof, we refer to any standard textbook.[9,10]

One should stress that for (3.12) some regularity properties of the solution are needed, even if they are not spelt out in the standard accounts of the theorem. Actually, the deep physical question of spontaneous breaking requires a more refined analysis of the mathematical properties of the solutions and of their behaviour at infinity. As we shall see, the problem of existence of "islands" or phases, stable under time

[8]E. Noether, Nachr. d. Kgl. Ges. d. Wiss. Göttingen (1918), p. 235.

[9]See, e.g. H. Goldstein, *Classical Mechanics*, 2nd. ed., Addison-Wesley 1980; E. L. Hill, Rev. Mod. Phys. **23**, 253 (1951); N.N. Bogoljubov and D.V. Shirkov, *Introduction to the theory of quantized fields*, Interscience 1958, Sect. 2.5.

[10]For the representations of Lie groups and their generators in classical systems, see D.G. Currie, T.F. Jordan and E.C.G. Sudarshan, Rev. Mod. Phys. **35**, 350 (1963); E.C.G. Sudarshan and N. Mukunda, *Classical Dynamics*: *A Modern Perspective*, J. Wiley and Sons 1974.

evolution (playing the role of the valleys of the example discussed in Chap. 2) and characterized by a non-trivial behaviour at infinity of the corresponding solutions, will require a sort of *stability theory for the infinite-dimensional system* (3.1).

Chapter 4
General Properties of Solutions of Classical Field Equations

The first basic question is to identify the possible configurations of the systems (3.1), namely the set X of initial data for which the time evolution is well defined and which is mapped onto itself by time evolution. In the mathematical language, one has to find the functional space X for which the Cauchy problem is well posed. In order to see this, one has to give conditions on $U'(\varphi)$ and to specify the class of initial data or, equivalently, the class of solutions one is interested in. Here one faces an apparently technical mathematical problem, which has also deep physical connections.

In the pioneering work by Jörgens[11] and Segal,[12] the choice was made of considering those initial data (and, consequently, those solutions) for which the total "kinetic" energy is finite[13]

$$E_{kin} \equiv \tfrac{1}{2} \int [(\nabla\varphi)^2 + \varphi^2 + \psi^2] d^s x < \infty, \quad \psi = \dot{\varphi}. \tag{4.1}$$

From a physical point of view, condition (4.1) is unjustified and it automatically rules out very interesting cases, like the external field problem, the symmetry breaking solutions, the soliton-like solutions and, in general, all the solutions which do not decrease sufficiently fast at large distances to make the above integral (4.1) convergent. Actually, there is no physical reason why E_{kin} should be finite, since even the splitting of energy into a kinetic and a potential part is not free of ambiguities.

[11] K. Jörgens, Mat. Zeit. **77**, 291 (1961).

[12] I. Segal, Ann. Math. **78**, 339 (1963).

[13] Strictly speaking, the kinetic energy should not involve the term φ^2. Our abuse of language is based on the fact that the bilinear part of the total energy corresponds to what is usually called the "non-interacting" theory (whose treatment is generally considered as trivial or under control by an analysis in terms of normal modes). The remaining term in the total energy is usually considered as the true interaction potential.

F. Strocchi, *Symmetry Breaking*, Theoretical and Mathematical Physics,
https://doi.org/10.1007/978-3-662-62166-0_4

Therefore, we have to abandon condition (4.1) and we only require that the initial data are *locally* smooth in the sense that

$$\int_V [(\nabla\varphi)^2 + \varphi^2 + \psi^2] d^s x < \infty \tag{4.2}$$

for any bounded region V (*locally finite kinetic energy*).

As it is usual in the theory of second-order differential equations, one may write (3.1) in first order (or Hamiltonian) formalism, by grouping together the field $\varphi(t)$ and its time derivative $\psi(t) = \dot{\varphi}(t)$ in a two-component vector

$$u(t) = \begin{pmatrix} \varphi(t) \\ \psi(t) \end{pmatrix} \equiv \begin{pmatrix} u_1(t) \\ u_2(t) \end{pmatrix}.$$

Equation (3.1) can then be written in the form

$$\frac{du}{dt} = Ku + f(u), \tag{4.3}$$

with the initial condition

$$u(0) = u_0 = \begin{pmatrix} \varphi_0 \\ \psi_0 \end{pmatrix}, \tag{4.4}$$

where

$$K = \begin{pmatrix} 0 & 1 \\ \triangle & 0 \end{pmatrix}, \quad f(u) = \begin{pmatrix} 0 \\ -U'(\varphi) \end{pmatrix}. \tag{4.5}$$

One of the two components of (4.3) is actually the statement that $\psi = \dot{\varphi}$.

It is more convenient to rewrite (4.3) as an integral equation which incorporates the initial conditions. To this purpose, we introduce the one-parameter continuous group $W(t)$ generated by K and corresponding to the free wave equation (see Appendix 10.1)

$$W(0) = 1, \quad W(t+s) = W(t)\, W(s) \quad \forall t,\ s.$$

Then, the integral form of (4.3) is

$$u(t) = W(t)u_0 + \int_0^t W(t-s) f(u(s)) ds. \tag{4.6}$$

The main advantage of (4.6) is that, in contrast to (4.3), it does not involve derivatives of u and, as we will see, it is easier to give it a precise meaning.

In first-order formalism, the condition that the kinetic energy is locally finite reads: $u_1 = \varphi \in H^1_{loc}(\mathbb{R}^s)$, (i.e. $|\nabla\varphi|^2 + |\varphi|^2$ is a locally integrable function); $u_2 = \psi \in L^2_{loc}(\mathbb{R}^s)$. Thus, we assume the following *local regularity condition of the initial data*