Commun. Math. Phys. 69, 19–30 (1979)

Communications in Mathematical Physics

The Principle of Symmetric Criticality

Richard S. Palais*

Department of Mathematics, Brandeis University, Waltham, Massachusetts 02154, USA

Abstract. It is frequently explicitly or implicitly assumed that if a variational principle is invariant under some symmetry group **G**, then to test whether a symmetric field configuration φ is an extremal, it suffices to check the vanishing of the first variation of the action corresponding to variations $\varphi+\delta\varphi$ that are also symmetric. We show by example that this is not valid in complete generality (and in certain cases its meaning may not even be clear), and on the other hand prove some theorems which validate its use under fairly general circumstances (in particular if **G** is a group of Riemannian isometries, or if it is compact, or with some restrictions if it is semi-simple).

0. Introduction

What we call the Principle of Symmetric Criticality, abbreviated herein to "the Principle", states in brief that critical symmetric points are symmetric critical points.

In more detail, let **M** be a smooth (i.e., C^∞) manifold on which a group **G** acts by diffeomorphisms (a "smooth **G**-manifold") and let $f:\mathbf{M}\to\mathbf{R}$ be a smooth **G**-invariant function on **M** (that is, f is constant on the orbits of **G**). Then a *critical point* (of f) is a point p of **M** where df_p, the differential of f at p vanishes. And a *symmetric point* (of **M**) is an element of the set $\Sigma=\{p\in\mathbf{M}|gp=p \text{ for all } g\in\mathbf{G}\}$ of points fixed under the action of **G**. The Principle states that *in order for a symmetric point p to be a critical point it suffices that it be a critical point of $f|\Sigma$, the restriction of f to Σ;* in other words if the directional derivatives $df_p(X)$ vanish for all directions X at p tangent to Σ, then the Principle claims that directional derivatives in directions transverse to Σ also vanish. In particular, for example, an isolated point of Σ (where there are no directions tangent to Σ) should automatically be a critical point of f.

In this generality the Principle unfortunately is *not* valid, as we shall illustrate below by a number of examples. (We shall in particular give examples where Σ is not a submanifold of $\mathbf{M}$, so that the notion of a direction tangent to Σ is ill-defined and the Principle is ambiguous.)

Since the Principle is such a useful tool, both theoretically and practically for the location of critical points, it is pleasant to be able to report the good news that, despite the above, the Principle is both meaningful and valid in a reasonably broad context. We shall prove theorems that validate its use in most situations that seem to come up in mathematical physics. In particular we shall show that it is valid if $\mathbf{M}$ is Riemannian and $\mathbf{G}$ acts isometrically on $\mathbf{M}$ (*a fortiori* if $\mathbf{M}$ is a real or complex Hilbert space on which $\mathbf{G}$ acts by orthogonal or unitary transformations), quite generally if $\mathbf{G}$ is compact, and also if $\mathbf{G}$ is a semi-simple Lie group and $\mathbf{M}$ is a finite dimensional real analytic manifold on which $\mathbf{G}$ acts analytically.

As should be evident from the above, we use the term "manifold" in the modern extended sense of a possibly infinite dimensional manifold modelled on a Banach space [9], so that in particular $\mathbf{M}$ could be a manifold of sections of some smooth fiber bundle and f a Lagrangian "action" functional for a classical field theory (cf. [12], in particular Chaps. 13 and 19). A final notational point; in general we assume no topology on $\mathbf{G}$, however when $\mathbf{G}$ is explicitly assumed to be a Lie group, then the assumption that $\mathbf{M}$ is a smooth $\mathbf{G}$ manifold is to be understood to include the assumption that the map $(g,x)\mapsto gx$ is a smooth map of $\mathbf{G}\times\mathbf{M}$ into $\mathbf{M}$ (rather than merely smooth in x for each $g\in\mathbf{G}$).

The history of the Principle is somewhat obscure. Frequently there is an implicit appeal to it in a situation such as the following. One has a field theory with a rotationally invariant Lagrangian and looks for rotationally symmetric extremals. One therefore takes as an *Ansatz* that the field components are functions only of the length r of the radius vector x, and computes the variation of the integral of the Lagrangian with respect to infinitely close field configurations *that also satisfy* the *Ansatz*. Setting the latter to zero gives a system of ordinary differential equations for the field components, that frequently can be solved explicitly. (For a particularly easy sample of such a calculation see Example 1.2 below.) Of course these ordinary differential equations are *necessary* conditions for a rotationally invariant field configuration to be an extremum, but it is only the validity of the Principle in this case (because the rotation group is compact) that assures they are also sufficient conditions. An early and typical example of such an implicit use of the Principle will be found in Weyl's derivation of the Schwarzschild solution of the Einstein field equations (cf. [14, p. 252], or [13, p. 165]). A very explicit reference to the Principle, with a sketch of a proof, will be found in Coleman's well-known paper on "Classical Lumps and their Quantum Descendants" ([2, Appendix 4]). There are however a number of unstated hypotheses used in Coleman's argument, making it a little unclear just exactly what it does prove.

1. Examples

1.1. Let $\mathbf{X}$ and $\mathbf{Y}$ be Banach spaces, $\mathbf{M}=\mathbf{X}\times\mathbf{Y}$, and let $\mathbf{G}$ be the two element group $\{E,T\}$, where $T:\mathbf{M}\to\mathbf{M}$ is the linear involution defined by the reflection $T(x,y)=(x,-y)$, so that the set Σ of symmetric points is just the $\mathbf{X}$-axis $\{(x,0)|x\in\mathbf{X}\}$. Let

$f:\mathbf{M}\to\mathbf{R}$ be any smooth function, even in y, i.e. $f(x,-y)=f(x,y)$. Then the chain rule gives immediately that the partial differential of f with respect to y satisfies $(d_2f)(x,-y)=-(d_2f)(x,y)$ and taking $y=0$ gives $d_2f=0$ at all points $(x,0)\in\Sigma$. It follows that if $\sigma=(x,0)\in\Sigma$ is a critical point of $f|\Sigma$ (i.e., if $d_1f_\sigma=0$), then $df_\sigma=d_1f_\sigma+d_2f_\sigma=0$, so σ is a critical point of f, verifying the Principle in this case. Take in particular $\mathbf{X}=\mathbf{Y}=\mathbf{R}$, so that $\mathbf{M}$ is the usual x,y-plane, and let $f(x,y)=\pm(x^2-y^2)$. Then $f|\Sigma$ has only one critical point, the origin, which depending on the choice of sign is either an absolute maximum or absolute minimum. However, the origin is clearly a saddle point of f. (Thus the Principle only locates extrema of f from extrema of $f|\Sigma$, not local maxima or local minima of f from those of $f|\Sigma$.)

1.2. Our second example is meant to illustrate a typical application of the Principle in mathematical physics. We shall use it to find all harmonic functions in the region $\mathbf{A}=\{x\in\mathbf{R}^n|r_1\leqq x\leqq r_2\}$ that are radially symmetric, i.e. invariant under the rotation group $\mathbf{SO}(n)$. Let $C^2(\mathbf{A})$ denote the Banach space of all C^2 real valued functions on $\mathbf{A}$. Then $\mathbf{SO}(n)$ is linearly represented in $C^2(\mathbf{A})$ by $gU(x)=U(g^{-1}x)$. Since the orbits of $\mathbf{SO}(n)$ on $\mathbf{A}$ are just the spheres concentric with the origin, a function $U(x)$ in $C^2(\mathbf{A})$ is a symmetric point if and only if it has the form $\tilde{U}(\|x\|)$ for some $\tilde{U}\in C^2([r_1,r_2])$. We note that the gradient of such a U is given by

$$(\nabla U)(x)=\tilde{U}'(\|x\|)\left(\frac{x}{\|x\|}\right).$$

Let $\mathbf{M}$ denote the submanifold of $C^2(\mathbf{A})$ consisting of all U assuming two arbitrary but fixed values c_1 and c_2 on the boundary spheres $\|x\|=r_1$ and $\|x\|=r_2$ respectively, so that Σ, the set of symmetric points of $\mathbf{M}$ under the action of $\mathbf{SO}(n)$ is identified with

$$\{\tilde{U}\in C^2([r_1,r_2])|\tilde{U}(r_i)=c_i\}.$$

Let $f:\mathbf{M}\to\mathbf{R}$ denote the Dirichlet functional:

$$\int_A \|\nabla U(x)\|^2dx.$$

By Dirichlet's principle a harmonic function in $\mathbf{A}$ having boundary values c_i on $\|x\|=r_i$ is just a critical point of f. Now if $U(x)=\tilde{U}(\|x\|)$, then clearly

$$f(U)=\Omega\int_{r_1}^{r_2}\tilde{U}'(r)^2r^{n-1}dr,$$

where Ω is the volume of the unit sphere in $\mathbf{R}^n$. The Euler-Lagrange equation satisfied by the extrema $\tilde{U}$ of this latter functional on Σ is just $(2\tilde{U}'(r)r^{n-1})'=0$ or $r\tilde{U}''+(n-1)\tilde{U}'=0$, which is easily seen to have the general solution $\tilde{U}(r)=a+br^{2-n}$ [except $\tilde{U}(r)=a+b\log(r)$ if $n=2$].

1.3. There are a number of important and natural ways to generalize the above example. We can replace $\mathbf{SO}(n)$ by a more general compact Lie group $\mathbf{G}$, and $\mathbf{A}$ by a more general $\mathbf{G}$-manifold $\mathbf{X}$. We can replace the Banach space $\mathbf{M}=C^2(\mathbf{A})$ by a more general Banach manifold of maps of $\mathbf{X}$ into a manifold $\mathbf{Y}$, or by a manifold of trivially on $\mathbf{Y}$ as well as on $\mathbf{X}$ [in the fiber bundle case one should then assume that the projection $\Pi:\mathbf{Y}\to\mathbf{X}$ is equivariant, i.e., satisfies $\Pi(gy)=g\Pi(y)$]. The action of $\mathbf{G}$ on $\mathbf{M}$ is now defined by $(gU)(x)=g(U(g^{-1}x))$. Note that the symmetric elements U of $\mathbf{M}$ are no longer the *invariant* maps, i.e., those satisfying $U(gx)=U(x)$, but rather the *equivariant* ones, satisfying $U(gx)=gU(x)$ (of course when the action of $\mathbf{G}$ on $\mathbf{M}$ is trivial these are the same). Finally we can replace f by a more general $\mathbf{G}$-invariant variational functional. When all these generalizations are made the Principle reduces to the following:

Theorem. *Let $\mathbf{G}$ be a compact Lie group, $\mathbf{X}$ a smooth $\mathbf{G}$-manifold, $\Pi:\mathbf{Y}\to\mathbf{X}$ a smooth $\mathbf{G}$-fiber bundle over $\mathbf{X}$, and $\mathbf{M}$ a Banach manifold of sections of $\mathbf{Y}$. Let $\mathbf{G}$ act on $\mathbf{M}$ by $(gU)(x)=g(U(g^{-1}x))$ and let $f:\mathbf{M}\to\mathbf{R}$ be a smooth $\mathbf{G}$-invariant function on $\mathbf{M}$. Then the set Σ of $\mathbf{G}$-equivariant sections in $\mathbf{M}$ is a smooth submanifold of $\mathbf{M}$, and if $U\in\Sigma$ is a critical point of $f|\Sigma$ then U is in fact a critical point of f.*

If $\mathbf{X}$ and $\mathbf{Y}$ are Riemannian manifolds and the actions of $\mathbf{G}$ are isometric, then for f we could take the "volume" functional or else the "energy" functional of Fuller [4] whose critical points are the harmonic maps of $\mathbf{X}$ into $\mathbf{Y}$ (for which see also [3] and [10]). A particularly interesting choice for $\mathbf{X}$ is one of the form $\mathbf{G/H}$, i.e. one on which $\mathbf{G}$ acts transitively. In this case Σ will be finite dimensional and often even compact, although $\mathbf{M}$ itself is infinite dimensional, so the Principle leads to existence theorems for extremals for appropriate functionals f. In the case of the volume functional we recover by an entirely different method a well-known and beautiful theorem of Hsiang [7].

In the above generality I believe this application of the Principle is new and should prove useful in pure differential geometry as well as in mathematical physics. We shall report on it more fully elsewhere.

2. The Riemannian Case

In this section we assume that $\mathbf{M}$ is a Riemannian manifold. That is $\mathbf{M}$ is locally diffeomorphic to a (finite or infinite dimensional) Hilbert space and there is an inner product $\langle\,,\rangle_x$ on each tangent space TM_x, that varies smoothly with x. For details see [9] or [11]. We also assume that each $g\in\mathbf{G}$ is an isometry of $\mathbf{M}$, meaning that $Dg_x:\mathbf{TM}_x\to\mathbf{TM}_{gx}$ preserves inner products. As a special case $\mathbf{M}$ could be a Hilbert space (real or complex) and the action of $\mathbf{G}$ an orthogonal or unitary representation.

A first important consequence of the Riemannian hypothesis is the existence of the geodesic spray or equivalently of the exponential map $\exp:\mathbf{O}\to\mathbf{M}$. Here $\mathbf{O}$ is a neighborhood of the zero section of the tangent bundle $\mathbf{TM}$ and exp is characterized by the property that for $v\in\mathbf{TM}_p$ and $t\in\mathbf{R}$ near zero, $\{t\to\exp(tv)\}$ is the unique geodesic of $\mathbf{M}$ parameterized proportionately to arc length emanating from p with tangent vector v. An obvious but crucial property of this map is that for any isometry $\phi:\mathbf{M}\to\mathbf{M}$ (and in particular for elements of $\mathbf{G}$),

$$\exp(D\phi(v))=\phi(\exp(v)),$$

which is just to say that an isometry maps geodesics to geodesics. Now exp maps a neighborhood of zero in the Hilbert space $\mathbf{TM}_p$ diffeomorphically onto a neighborhood of p in $\mathbf{M}$ (cf. [9, IV, §4]) and so may be regarded as a chart or coordinate system at p; these of course are just the classical "geodesic normal coordinates" at p. If $p \in \Sigma$ then the fact that $\exp \circ Dg_p = g \circ \exp$ shows immediately that in geodesic normal coordinates at p the action of $\mathbf{G}$ is just an orthogonal linear representation. In particular Σ intersects the domain of this coordinate system in a linear subspace of $\mathbf{TM}_p$, namely $\{v \in \mathbf{TM}_p | Dg_p(v) = v \text{ for all } g \in \mathbf{G}\}$, so that clearly Σ is a smooth submanifold of $\mathbf{M}$. In fact we see an important extra detail: If $v \in \mathbf{TM}_p$ is left fixed by all Dg_p then the geodesic starting from p with tangent v is left pointwise fixed by all $g \in \mathbf{G}$, and hence lies in Σ; it follows that Σ is totally geodesic in $\mathbf{M}$ and that v is tangent to Σ at p.

A second important consequence of the Riemannian structure of $\mathbf{M}$ is that if $f: \mathbf{M} \to \mathbf{R}$ is a smooth function then there is an associated gradient vector field ∇f on $\mathbf{M}$ related to the differential df of f by $df_x(v) = \langle v, \nabla f_x \rangle_x$ for all $v \in \mathbf{TM}_x$. In particular $\nabla f_x = 0$ if and only if x is a critical point of f, and more generally if x is in a smooth submanifold $\mathbf{N}$ of $\mathbf{M}$, then x is a critical point of $f|\mathbf{N}$ if and only if ∇f_x is orthogonal to $\mathbf{TN}_x$. If f is $\mathbf{G}$-invariant, then for any g in $\mathbf{G}$ we have $f \circ g = f$ so by the chain rule $df_x = d(f \circ g)_x = df_{gx} \circ Dg_x$, and since Dg_x maps $\mathbf{TM}_x$ isometrically onto $\mathbf{TM}_{gx}$ we see from the characterization of ∇f that $Dg(\nabla f_x) = \nabla f_{gx}$. In particular if $x \in \Sigma$ then we have $Dg(\nabla f_x) = \nabla f_x$ for all $g \in \mathbf{G}$. We now have all the ingredients for a geometrically satisfying proof of the Principle in this case.

Theorem. *Let* $\mathbf{G}$ *be a group of isometries of a Riemannian manifold* $\mathbf{M}$ *and let* $f: \mathbf{M} \to \mathbf{R}$ *be a* C^1 *function invariant under* $\mathbf{G}$. *Then the set* Σ *of stationary points of* $\mathbf{M}$ under the action of $\mathbf{G}$ *is a totally geodesic smooth submanifold of* $\mathbf{M}$, *and if* $p \in \Sigma$ *is a critical point of* $f|\Sigma$ *then* p *is in fact a critical point of* f.

Proof. By assumption ∇f_p is orthogonal to $\mathbf{T}\Sigma_p$ and hence it will suffice to show that also $\nabla f_p \in \mathbf{T}\Sigma_p$. But as we have just seen $Dg(\nabla f_p) = \nabla f_p$ for all $g \in \mathbf{G}$ and as remarked above this implies that the geodesic emanating from p in the direction ∇f_p is pointwise fixed under all g in $\mathbf{G}$, and hence lies inside Σ so that its tangent at p, ∇f_p, is in $\mathbf{T}\Sigma$. □

We now look at the other side of the coin and see how badly things can go wrong if *no* restrictive assumptions are made.

3. Counterexamples

3.1. Let Σ be a non-empty closed subset of a connected, compact, C^∞ manifold $\mathbf{M}$. It is well known that there exists a C^∞ real valued function g on $\mathbf{M}$ vanishing precisely on Σ and a C^∞ vector field Y on $\mathbf{M}$ vanishing at only one point, which we can take in Σ. Then $X = gY$ is a C^∞ vector field on $\mathbf{M}$ whose set of zeroes is exactly Σ, so that the flow $\{\phi_t\}$ generated by X is a C^∞ action of the real line $\mathbf{R}$ on $\mathbf{M}$ whose set of symmetric points is exactly Σ. Now Σ can be very far from manifold-like; for example it could be a Cantor set. In particular the tangent space to Σ at one of its points p will in general not have a well defined meaning, so that if

$f:\mathbf{M}\to\mathbf{R}$ is a smooth invariant function on $\mathbf{M}$, then the meaning of $f|\Sigma$ having a critical point at p becomes unclear, and hence the meaning of the Principle also becomes unclear.

A second, somewhat more formal definition of the Principle restores meaningfulness in such cases. Namely at $p\in\Sigma$ there is a natural linear representation of $\mathbf{G}$ on $\mathbf{TM}_p$, given by $g\mapsto Dg_p$. We simply define $\mathbf{T}\Sigma_p$ to mean

$$\{v\in\mathbf{TM}_p|Dg_p(v)=v \text{ for all } g\in\mathbf{G}\}$$

and define "$f|\Sigma$ has a critical point at p" to mean that df_p vanishes on this linear subspace of $\mathbf{TM}_p$. However as we shall see in the next example, when Σ *is* a smooth submanifold of $\mathbf{M}$ this formal tangent space to Σ at p can be strictly larger than the actual tangent space to Σ at p and hence leads to quite a different interpretation of the Principle. In the next section we shall introduce a condition (linearizability of the action of G at all points of Σ) that will avoid this ambiguity.

3.2. Let $\{\phi_t\}$ be the action of $\mathbf{R}$ on $\mathbf{M}=\mathbf{R}^2$ generated by the vector field $p^k\partial/\partial q$ where $k=1$ or 2. We note that these are Hamiltonian vector fields with corresponding Hamiltonian function $p^{(k+1)}/(k+1)$. The corresponding flow $\{\phi_t\}$ on $\mathbf{M}$ is $\phi_t(q,p)=(q+p^kt,p)$ from which we see that Σ is the q-axis. A smooth function $f:\mathbf{R}^2\to\mathbf{R}$ is invariant if and only if it is of the form $f(q,p)=h(p)$, and such an f has a critical point at $x=(q,0)\in\Sigma$ if and only if $h'(0)=0$. Now if $k=1$ then the flow is linear, so $(D\phi_t)_{(q,p)}=\phi_t$ for all (q,p), and the two interpretations of the Principle discussed in 3.1 coincide. Moreover $f(p,q)=p$ is a clear counter-example to the Principle, since although it has no critical points in $\mathbf{R}^2$, restricted to Σ it is identically zero, and hence *every* point of Σ is a critical point of $f|\Sigma$. In case $k=2$, $(D\phi_t)_x(v)=v$ for all $t\in\mathbf{R}$, $v\in\mathbf{R}^2$, and $x=(q,0)\in\Sigma$. Thus the formal tangent space to Σ at x introduced in 3.1 is two-dimensional, even though Σ is a one-dimensional manifold! Nevertheless this heroically illogical definition of the tangent space to Σ at x does have the virtue of making the Principle correct in this instance, whereas for the customary interpretation $f(p,q)=p$ is a counter-example, just as it was for *both* interpretations when we took $k=1$.

3.3. Let $\mathbf{S}^1$ denote the unit circle in the complex plane $\mathbf{C}=\mathbf{R}^2$ and define an action of $\mathbf{SL}(2,\mathbf{R})$ on $\mathbf{S}^1$ by defining $g\in\mathbf{SL}(2,\mathbf{R})$ acting on $e^{i\theta}$ to be $(ge^{i\theta})/\|ge^{i\theta}\|$. If we identify $\mathbf{SO}(2)\subseteq\mathbf{SL}(2,\mathbf{R})$ with $\mathbf{S}^1$ acting on itself by translation, then by the uniqueness of Haar measure, $d\theta$ is, up to a constant multiple, the only $\mathbf{SO}(2)$ invariant measure on $\mathbf{S}^1$. And since $d\theta$ is clearly *not* $\mathbf{SL}(2,\mathbf{R})$ invariant it follows that there is *no* $\mathbf{SL}(2,\mathbf{R})$ invariant measure $h(e^{i\theta})d\theta$ on $\mathbf{S}^1$ except $h=0$.

Now let $\mathbf{H}$ denote the Hilbert space $L^2(\mathbf{S}^1,\mathbf{R})$ with respect to the measure $d\theta$, so that

$$\langle f,h\rangle=\int_0^{2\pi} f(e^{i\theta})h(e^{i\theta})d\theta\,.$$

Define a continuous linear (but not orthogonal) representation A of $\mathbf{SL}(2,\mathbf{R})$ in $\mathbf{H}$ by

$$A_gf(e^{i\theta})=f(g^{-1}e^{i\theta})$$

and let A^t be the contragredient representation, i.e., $A_g^t = A_{g^{-1}}^*$, where * denotes Hilbert space adjoint, so that

$$\langle A_g f, h\rangle = \langle f, A_{g^{-1}}^t h\rangle .$$

Now regard **H** as a Banach **SL**(2, **R**)-space with A^t as the action. We claim that the linear functional $l:\mathbf{H}\to\mathbf{R}$ defined by:

$$l(h) = \langle 1, h\rangle = \int_0^{2\pi} h(e^{i\theta})d\theta$$

is invariant. Indeed since clearly $A_g 1 = 1$ we have:

$$l(A_g^t h) = \langle 1, A_g^t h\rangle = \langle A_{g^{-1}} 1, h\rangle = \langle 1, h\rangle = l(h).$$

Since $l \neq 0$ it follows that l has no critical points in **H**. To show that this gives a violation of the Principle it will suffice to see that the set Σ of symmetric points of **H** consists of the origin alone, for then clearly $l|\Sigma$ has the origin as a critical point. Suppose then that $h \in \Sigma$, or equivalently that $A_g^t h = h$ for all g in **SL**(2, **R**). Then for all f in **H** we have:

$$\langle A_{g^{-1}} f, h\rangle = \langle f, A_g^t h\rangle = \langle f, h\rangle$$

or equivalently

$$\int_0^{2\pi} f(ge^{i\theta})h(e^{i\theta})d\theta = \int_0^{2\pi} f(e^{i\theta})h(e^{i\theta})d\theta$$

which shows that $h(e^{i\theta})d\theta$ is a **SL**(2, **R**)-invariant measure on $\mathbf{S}^1$, so that as remarked above h must equal zero, as was to be shown. The point of this example is that it shows the Principle may fail to be true even for a continuous linear representation of a semi-simple Lie group in a Hilbert space. We shall see below however that for a finite dimensional representation of a semi-simple Lie group, or more generally for an analytic action of a semi-simple Lie group on a finite dimensional real analytic manifold, the Principle is valid.

4. Linearizability

Let p be any symmetric point of a smooth **G**-manifold **M**. Associated to p we have a representation of **G** in $\mathbf{TM}_p$, namely $\{g \mapsto Dg_p\}$, which we call the *linearization* of the action of **G** at p. As we saw in Example 3.2 this may have very little relation to the local properties of the action of **G** on **M** near p, which can be quite non-trivial even though the linearization is the trivial representation. We shall say that **G** (more properly, the action of **G**) is linearizable at p if in a suitable coordinate system about p it looks linear. To be precise, recall that a coordinate system, or chart, at p is a diffeomorphism ϕ of an open set $\mathcal{O}$ of **M** containing p onto an open set $\phi(\mathcal{O})$ in a Banach space **V**, and mapping p to the origin. We say that **G** is linearizable about p if there exists such a ϕ that for each $g \in \mathbf{G}$ the map

$$\phi \circ g \circ \phi^{-1} : \phi(\mathcal{O}) \to \mathbf{V}$$

is the restriction to $\phi(\mathcal{O})$ of a linear map $\tilde{g}:\mathbf{V}\to\mathbf{V}$. (We also say that ϕ linearizes $\mathbf{G}$ about p.) Clearly $g\mapsto\tilde{g}$ is a linear representation of $\mathbf{G}$ in $\mathbf{V}$, and by the chain rule $D\phi_p:\mathbf{TM}_p\to\mathbf{V}$ sets up an equivalence of this representation with the linearization of $\mathbf{G}$ at p. Since the maps $\tilde{g}$ are linear it follows that

$$\mathbf{W}=\{v\in\mathbf{V}|\tilde{g}v=v \text{ for all } g\in\mathbf{G}\}$$

is a closed linear subspace of $\mathbf{V}$, and since ϕ clearly maps $\mathcal{O}\cap\Sigma$ onto $\phi(\mathcal{O})\cap\mathbf{W}$, it follows that Σ is locally a smooth submanifold of $\mathbf{M}$ at p. Thus we have:

4.1. Proposition. *If* $\mathbf{M}$ *is a smooth* $\mathbf{G}$*-manifold such that the action of* $\mathbf{G}$ *is linearizable about each symmetric point, then the set* Σ *of symmetric points is a smooth submanifold of* $\mathbf{M}$.

Although Example 3.2 shows that Σ may well be a smooth submanifold of $\mathbf{M}$ without the action of $\mathbf{G}$ being linearizable about points of Σ, I know of no reasonably general hypotheses that imply the first property without also implying the second.

In Sect. 2 we saw that if $\mathbf{M}$ is Riemannian and $\mathbf{G}$ acts isometrically, then the action is linearized by geodesic coordinates at a symmetric point, and moreover that the Principle is valid in this case.

In Sect. 5 we shall see that if a compact Lie group $\mathbf{G}$ acts smoothly on a Banach manifold $\mathbf{M}$, then generalizing slightly an argument of Bochner the action of $\mathbf{G}$ can be linearized about any symmetric point of $\mathbf{M}$. In this case too we shall see that the Principle is valid.

In Sect. 6 we shall discuss a further case in which linearizability about each symmetric point is known; namely when $\mathbf{G}$ is a semi-simple Lie group acting real analytically on a real analytic manifold. Once again the Principle will easily be seen to be valid in this case.

However one should *not* jump to the conclusion that linearizability of the action at each symmetric point is itself a sufficient condition for the Principle to be valid. Looking back to Example 3.2 (with $k=1$) we have a case of a finite dimensional linear representation of R where the Principle fails. Even semi-simplicity of $\mathbf{G}$ together with linearizability at all symmetric points is not sufficient when $\mathbf{M}$ is infinite dimensional, for as we saw in Example 3.3 the simplest non-compact, semi-simple group, $\mathbf{SL}(2,R)$ has a (non-orthogonal) continuous linear Hilbert space representation for which the Principle fails.

In the remainder of this section we shall at least see how to develop a very simple test for validity of the Principle that works in the presence of linearizability.

Let $\mathbf{V}$ be a Banach $\mathbf{G}$-space, i.e., a Banach space in which $\mathbf{G}$ is represented linearly, and let $\mathbf{V}^*$ be the dual Banach space of continuous linear functionals on $\mathbf{V}$. Recall that $\mathbf{V}^*$ is also naturally a Banach $\mathbf{G}$-space, the action of $g\in\mathbf{G}$ on $l\in\mathbf{V}^*$ being given by $(gl)(v)=l(g^{-1}v)$, so that the subspace Σ_* of $\mathbf{V}^*$ consisting of symmetric points is exactly the set of linear functionals $l:\mathbf{V}\to\mathbf{R}$ invariant under the action of $\mathbf{G}$ on $\mathbf{V}$. If $\mathbf{W}$ is a subspace of $\mathbf{V}$ then as usual we denote by $\mathbf{W}^0$ its annihilator subspace in $\mathbf{V}^*$, consisting of all linear functionals vanishing identically on $\mathbf{W}$. Now if $l:\mathbf{V}\to\mathbf{R}$ is linear then it is *a fortiori* smooth and $dl_v=l$ for all v in $\mathbf{V}$. It follows that $l|\mathbf{W}$ has a critical point if and only if $l\in\mathbf{W}^0$ (in which case of course every point of $\mathbf{W}$ is a critical point of $l|\mathbf{W}$) and in particular l has no critical points unless $l=0$. Thus if we apply the Principle to an element $l\in\mathbf{V}^*$ it becomes the statement that if $l\in\Sigma_*$ and $l\in\Sigma^0$ then $l=0$, or more succinctly $\Sigma_*\cap\Sigma^0=0$. Clearly then this is a necessary condition condition for the validity of the Principle for arbitrary smooth invariant functions $f:\mathbf{V}\to\mathbf{R}$, and we now show that it is also sufficient.

4.2. Proposition. *If* $\mathbf{V}$ *is a Banach* $\mathbf{G}$*-space then the condition* $\Sigma_*\cap\Sigma^0=0$ *is both necessary and sufficient for the Principle to be valid for all smooth invariant functions* $f:\mathbf{V}\to\mathbf{R}$.

Proof. Let p be a critical point of $f|\Sigma$ and let $l=df_p$. We must show that $l=0$, and since $l|\Sigma=df_p|\Sigma=d(f|\Sigma)_p=0$, we have $l\in\Sigma^0$ so it will suffice to show $l\in\Sigma_*$. But since f is invariant, $f=f\circ g$ for any $g\in\mathbf{G}$, so by the chain rule $df_{gv}\circ g=df_v$ for all $v\in\mathbf{V}$. In particular taking $v=p\in\Sigma$ gives $g^{-1}l=l$ for all $g\in\mathbf{G}$, and hence $l\in\Sigma_*$. □

4.3. Definition. A Banach $\mathbf{G}$-space $\mathbf{V}$ will be called admissible if it satisfies the condition $\Sigma_*\cap\Sigma^0=0$. A smooth $\mathbf{G}$-manifold $\mathbf{M}$ will be called admissible if for each symmetric point p of $\mathbf{M}$ the action of $\mathbf{G}$ on $\mathbf{M}$ is linearizable about p and if the linearization of $\mathbf{G}$ at p is an admissible Banach $\mathbf{G}$-space.

4.4. Theorem. *The Principle of Symmetric Criticality is valid for admissible smooth* $\mathbf{G}$*-manifolds.*

Proof. When the action of $\mathbf{G}$ is linearizable about a symmetric point p then locally the action looks like its linearization at p. Since the question of whether p is a critical point of a function is a local one, the theorem is immediate from 4.2 and 4.3. □

A hyperplane in a Banach space $\mathbf{V}$ is a closed linear subspace $\mathbf{H}$ of $\mathbf{V}$ of the form $l^0=\{v\in\mathbf{V}|l(v)=0\}$ for some non-zero $l\in\mathbf{V}^*$. If $l(v_0)=1$ then each element v of $\mathbf{V}$ can be written uniquely in the form $h+\alpha v_0$ for a scalar α and $h\in\mathbf{H}$ [namely $\alpha=l(v)$ and $h=v-l(v)v_0$] that is $\mathbf{V}$ is the direct sum of $\mathbf{H}$ and the one dimensional space $[v_0]$ spanned by v_0. Conversely if a closed subspace $\mathbf{H}$ of $\mathbf{V}$ has a complementary one dimensional subspace $[v_0]$ then $\mathbf{H}$ is a hyperplane; in fact $\mathbf{H}=l^0$ where $l(v)$ is defined by the condition $v-l(v)v_0\in\mathbf{H}$. The following is a simple but powerful test for admissibility of a Banach $\mathbf{G}$-space.

4.5. Theorem. *A Banach* $\mathbf{G}$*-space* $\mathbf{V}$ *is admissible provided that for each non-zero invariant linear functional* $l:\mathbf{V}\to\mathbf{R}$ *the invariant hyperplane* $\mathbf{H}=l^0$ *has an invariant complementary subspace.*

Proof. Assuming that $\Sigma_*\cap\Sigma^0$ contains a non-zero element l we shall derive a contradiction. Since $l\in\Sigma_*$, by assumption its null space $\mathbf{H}$ has an invariant linear complement, Γ. Now since $l\in\Sigma^0$ it follows that $\Sigma\subseteq\mathbf{H}$, hence Γ which is disjoint from $\mathbf{H}$ cannot be included in Σ. We will get our contradiction by showing that indeed $\Gamma\subseteq\Sigma$. For suppose $v_0\in\Gamma$. Since Γ is invariant $gv_0-v_0\in\Gamma$ for any $g\in\mathbf{G}$. But since l is invariant,

4.6. Corollary. *If* **V** *is completely reducible (i.e., every closed, invariant subspace has a complementary closed, invariant subspace) then* **V** *is admissible.*

5. The Compact Case

In this section **G** will be a compact Lie group with normalized Haar measure μ. If **V** is a Banach **G**-space and $v \in \mathbf{V}$ then the center of gravity of the orbit of v [that is $\int gv\,d\mu(g)$] will be denoted by $\bar{v}$ or by $A(v)$. The map $A:\mathbf{V}\to\mathbf{V}$, called "averaging over the group" has a number of well-known elementary but important properties. Firstly, it is trivial that $\bar{v}=v$ when v is in Σ, and by the invariance of μ it follows that in fact A is a continuous linear projection of **V** onto the subspace Σ. Secondly, if $\mathscr{C}$ is an invariant, closed, convex subset of **V** then $A(\mathscr{C})\subseteq\mathscr{C}$. Thirdly, if T is an equivariant linear map of **V** into another Banach **G**-space **W** then $TA=AT$ (i.e., if $Tv=w$ then $T\bar{v}=\bar{w}$) in particular if $l:\mathbf{V}\to\mathbf{R}$ is an invariant linear functional on **V** then $l(\bar{v})=l(v)$.

5.1. Theorem. *If* **G** *is a compact Lie group then every Banach* **G**-*space is admissible.*

Proof. Let $l:\mathbf{V}\to\mathbf{R}$ be an invariant non-zero linear functional and let $\mathscr{C}=\{v\in\mathbf{V}\,|\,l(v)=1\}$. Then $\mathscr{C}$ is an invariant, closed, convex subset of **V**, in fact a translate of the hyperplane $\mathbf{H}=\{v\in\mathbf{V}\,|\,l(v)=0\}$. Since $l\neq 0$, $\mathscr{C}$ is not empty and if we choose $v_0\in\mathscr{C}$, then $\bar{v}_0=A(v_0)\in\mathscr{C}\cap\Sigma$, so clearly $[\bar{v}_0]$ is an invariant linear complement for **H**, and 4.5 completes the proof. □

5.2. Lemma. *If* **M** *is any smooth* **G**-*manifold and p is a symmetric point of* **M**, *then p has arbitrarily small invariant neighborhoods. In fact if* $\mathcal{O}$ *is any neighborhood of p in* **M** *then*

$$\tilde{\mathcal{O}}=\bigcap_{g\in\mathbf{G}} g\mathcal{O}$$

is an invariant neighborhood of p included in $\mathcal{O}$.

Proof. Trivially $\tilde{\mathcal{O}}$ is invariant and included in $\mathcal{O}$. By continuity of the map $(g,x)\mapsto gx$ of $\mathbf{G}\times\mathbf{M}$ into **M** and the fact that $gp=p$, for any $g\in\mathbf{G}$ there is a neighborhood N_g of g in **G** and a neighborhood U_g of p in **M** such that $N_gU_g\subseteq\mathcal{O}$. By compactness a finite number $N_{g_1},\dots,N_{g_n}$ of the N_g will cover **G**, and if U is the intersection of corresponding U_g then clearly U is a neighborhood of p that is included in $\tilde{\mathcal{O}}$. □

The following is a straightforward generalization to the Banach manifold setting of a classical theorem of Bochner (cf. [1]).

5.3. Theorem. *Let* **G** *be a compact Lie group acting* C^1 *on a Banach manifold* **M**. *Then the action of* **G** *on* **M** *can be linearized about any symmetric point p of* **M**.

Proof. By choosing a chart $\phi:\mathcal{O}\to\mathbf{V}$ at p and using 5.2 there is no loss of generality in assuming that p is the origin of a Banach space **V** and that **M** is a neighborhood of p in **V**. Let **W** be the Banach space of C^1 maps $\psi:\mathbf{M}\to\mathbf{V}$ such that $\|\psi\|<\infty$

where $\|\psi\|$ is defined to be the sum of the suprema of $\|\psi(v)\|$ and of $\|D\psi_v\|$ as v ranges over **M**. We define a continuous linear action of **G** on **W** by $g\psi = Dg_p \circ \psi \circ g^{-1}$. The convex subset $\mathscr{C}$ of **W** defined by

$$\mathscr{C} = \{\psi \in \mathbf{W} | \psi(p) = p \quad \text{and} \quad D\psi_p = \text{identity}\}$$

is invariant under **G** by an easy application of the chain rule. Note that by the inverse function theorem any element of $\mathscr{C}$ is a chart at p. If we take some element of $\mathscr{C}$, for example the inclusion of **M** into **V**, and average it over the group, we get an element ψ_0 of $\mathscr{C}$, and hence a chart at p, satisfying $g\psi_0 = \psi_0$ for all $g \in \mathbf{G}$. But the latter is equivalent to $Dg_p \circ \psi_0 \circ g^{-1} = \psi_0$, or to $Dg_p \circ \psi_0 = \psi_0 \circ g$, and hence ψ_0 linearizes the action of **G** about p. □

5.4. Theorem. *If* **G** *is a compact Lie group, then any smooth* **G***-manifold* **M** *is admissible and hence the Principle of Symmetric Criticality is valid for* **M**.

Proof. Immediate from 4.3, 5.1, 5.3, and 4.4. □

6. The Semi-Simple Case

Let **G** be a connected, semi-simple Lie group. It is a standard result (cf. [8]) that every finite dimensional linear representation space of **G** is completely reducible, hence by 4.6 admissible. In 1965 the author and Smale conjectured that if **M** were any smooth **G**-manifold then the action of **G** would be linearizable about each symmetric point of **M**. While in its full generality this question remains open, a partial answer has been provided by Hermann and by Guillemin and Sternberg who showed [5, 6] that the conjecture was valid provided **M** was a real analytic manifold on which **G** acted real analytically. Putting this together with 4.3 and 4.4 we have:

Theorem. *If* **G** *is a connected semi-simple Lie group then the Principle of Symmetric Criticality is valid for finite dimensional real analytic* **G***-manifolds.*

Acknowledgement. I would like to thank Prof. Coleman for reading and commenting on a draft of this paper.

References

1. Bochner, S.: Compact groups of differentiable transformations. Ann. Math. **46**, 372–381 (1945)
2. Coleman, S.: Classical lumps and their quantum descendants. Preprint, Harvard Physics Dept. (1975)
3. Eells, J., Lemaire, L.: A report on harmonic maps. Bull. London Math. Soc. **10**, 1–68 (1978)
4. Fuller, F.B.: Harmonic mappings. Proc. Nat. Acad. Sci. (U.S.A.) **40**, 110–116 (1954)
5. Guillemin, V., Sternberg, S.: Remarks on a paper of Hermann. Trans. Am. Math. Soc. **130**, 110–116 (1968)
6. Hermann, R.: The formal linearization of a semi-simple Lie algebra of vector fields about a singular point. Trans. Am. Math. Soc. **130**, 105–109 (1968)
7. Hsiang, W.Y.: On the compact, homogeneous, minimal submanifolds. Proc. Nat. Acad. Sci. (U.S.A.) **50**, 5–6 (1966)

8. Jacobson, N.: Lie algebras. New York: Interscience 1962
9. Lang, S.: Introduction to differentiable manifolds. New York: Interscience 1966
10. Misner, C.: Harmonic maps as models for physical theories. Phys. Rev. D (in press)
11. Palais, R.S.: Morse theory on Hilbert manifolds. Topology **2**, 299–340 (1963)
12. Palais, R.S.: Foundations of global nonlinear analysis. New York: Benjamin 1968
13. Pauli, W.: Theory of relativity. London: Pergamon Press 1958
14. Weyl, H.: Space-time-matter. New York: Dover 1951

Communicated by A. Jaffe

Received February 26, 1979