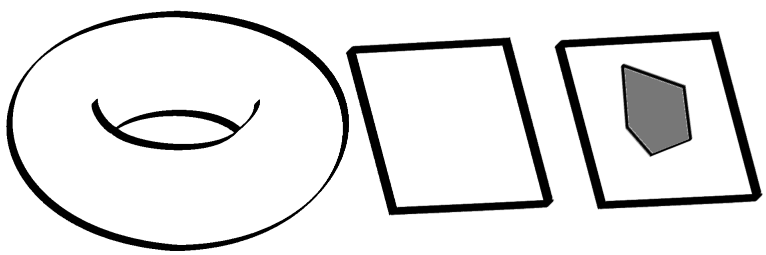
THE ATIYAH-GUILLEMIN-STERNBERG CONVEXITY THEOREM

WOLFGANG SCHMALTZ

Abstract. A 'moment map' is a way to generalize the definition of a Hamiltonian action of $\mathbb{R}$ on a symplectic manifold M. Associated to a Lie group G, a moment map μ is at the most basic level a map from M to $\mathfrak{g}^*$, the dual of the Lie algebra. In particular, a moment map then allows us to describe Hamiltonian actions of G on M. We present a proof, credited to Atiyah, Guillemin, and Sternberg, that investigates the properties of a Hamiltonian action of a torus Lie group, $\mathbb{T}^m$, and the properties of the associated moment map μ; in particular, we prove that image of the moment map $\mu(M) \subset \mathbb{R}^m$ must be convex.



Contents

Date: September 9, 2010.

1. Symplectic Manifolds

Definition 1.1. A **symplectic manifold** is a pair (M, ω), where M is a smooth manifold which possesses a closed, nondegenerate, skew-symmetric 2-form ω, called the **symplectic form**. We will often simply say that M is a symplectic manifold if the 2-form ω is understood.

The condition that ω is closed means that $d\omega = 0$, where d is the exterior derivative. That ω is nondegenerate means that at any point $p \in M$, if we let $X \in T_pM$ then if $\omega_p(X, Y) = 0$ for all $Y \in T_xM$ then we must have $X = 0$. Lastly, that ω is skew-symmetric means that at any point $p \in M$, we have $\omega_p(X, Y) = -\omega_p(Y, X)$ for all $X, Y \in T_pM$.

Furthermore, consideration of symplectic linear geometry of ω_p on T_pM, specifically, the fact that ω_p is nondegenerate and skew-symmetric means that the dimension of T_pM must be even. Therefore, the dimension of M is also even. We restate this as a proposition to note its importance:

Proposition 1.2. *If M is a symplectic manifold, then M is necessarily even dimensional.*

Definition 1.3. A **symplectomorphism** is a diffeomorphism from a symplectic manifold to itself which preserves the symplectic form. Explicitly, if M is a symplectic manifold, then $\psi \in \text{Diff}(M)$ is a symplectomorphism if $\psi^*\omega = \omega$. By the definition of the pullback, this means that at a point $p \in M$, and with vectors $X, Y \in T_pM$, we have

$$(\psi^*\omega)_p(X, Y) = \omega_{\psi(p)}(d\psi_p(X), d\psi_p(Y)) = \omega_p(X, Y)$$

The group (under composition) of symplectomorphisms of a symplectic manifold to itself is denoted as $\mathbf{Symp}(M, \omega)$.

Definition 1.4. A **symplectic submanifold** is a submanifold Y of a symplectic manifold (M, ω) such that at each point $p \in Y$, the restriction of ω_p to T_pY is symplectic, i.e., $\omega_p|_{T_pY \times T_pY}$ is nondegenerate (this restriction is automatically closed and skew-symmetric since ω is).

2. Almost Complex Structures

Definition 2.1. Let V be a vector space. A **complex structure** on V is a linear map $J : V \to V$ such that $J^2 = -\text{Id}$.

Definition 2.2. Let (V, ω) be a symplectic vector space. A complex structure J is called **compatible** if the map $g_J : V \times V \to \mathbb{R}$ defined by:

$$g_J(X, Y) = \omega(X, JY) \text{ for all } X, Y \in V$$

is a positive inner product on V.

Proposition 2.3. *Let (V, ω) be a symplectic vector space. Then there exists a compatible complex structure on V.*

Definition 2.4. Suppose that M is a smooth manifold. An **almost complex structure** on M is a smooth field of complex structures on the vector spaces of the tangent spaces. That is, at each point x in M we have a linear map $J_x : T_xM \to T_xM$ such that $J_x^2 = -\text{Id}$.

Definition 2.5. Suppose that (M, ω) is a symplectic manifold. An almost complex structure J on M is called **compatible** with ω if the 2-form g on TM defined by:

$$\begin{aligned} &g_x : T_xM \times T_xM \to \mathbb{R} \\ &g_x(X, Y) = \omega_x(X, J_x) \text{ for all } X, Y \in T_xM \end{aligned}$$

is a Riemannian metric on M. We call a triple (ω, g, J) where ω is a symplectic form, g is a Riemannian metric, and J is an almost complex structure a **compatible triple** when $g_x(\cdot, \cdot) = \omega_x(\cdot, J_x \cdot)$ for all $x \in M$.

Proposition 2.6. *Suppose that (M, ω) is a symplectic manifold, and g is a Riemannian metric on M. Then there exists an almost complex structure J on M which is compatible.*

Proposition 2.7. *Any symplectic manifold has compatible almost complex structures.*

Proposition 2.8. *Let (V, ω) be a symplectic vector space, and let (ω, g, J) be a compatible triple on V. A linear map $A : V \to V$ which preserves both the both the symplectic structure and the complex structure must be unitary, i.e., $A \in U(V)$.*

3. Symplectic and Hamiltonian Actions of $\mathbb{R}$

Definition 3.1. Let (M, ω) be a symplectic manifold. A **smooth symplectic action of $\mathbb{R}$ on** M is a group homomorphism $\psi : \mathbb{R} \to \mathrm{Symp}(M, \omega)$ such that the evaluation map $ev_\psi : M \times \mathbb{R} \to M$ given by $ev_\psi(p, t) = \psi_t(p)$ is smooth.

Definition 3.2. Let X be a vector field on a symplectic manifold (M, ω). Then we say the X is a **symplectic vector field** if the 1-form $i_X\omega$ is closed, that is, $di_X\omega = 0$.

For the next proposition, recall properties of the Lie derivative. Explicitly, given a tensor field τ and a smooth vector field X, we can let ψ_t be the flow of X, i.e., $\psi_0 = \mathrm{Id}$ and $\frac{d}{dt}\,\psi_t(p) = X(\psi_t(p))$. Then the Lie derivative of τ with respect to X is given by:

$$\mathcal{L}_X\tau = \frac{d}{dt}\bigg|_{t=0} \psi_t^*\tau.$$

We claim the following identities relating to the Lie derivative:

$$\begin{aligned} &(1) \text{The Cartan Magic Formula: } \mathcal{L}_X\tau = i_X d\tau + di_X\tau \\ &(2) \frac{d}{dt}\,\psi_t^*\tau = \psi_t^*\mathcal{L}_X\tau. \end{aligned}$$

Proposition 3.3. *Let (M, ω) be a compact, symplectic manifold. Let $\psi : \mathbb{R} \to$ Symp(M, ω) be a smooth symplectic action of $\mathbb{R}$. Then ψ generates a family of vector fields $\{X_t\}$ defined by:*

$$\frac{d}{dt}\,\psi_t = X_t \circ \psi_t.$$

Then X_t is a symplectic vector field for every $t \in \mathbb{R}$. Conversely, if $\{X_t\}$ is a time-dependent family of symplectic vector fields, then the flow of X_t determines a smooth family of diffeomorphisms $\{\psi_t\}$ satisfying:

$$\psi_0 = Id \text{ and } \frac{d}{dt}\,\psi_t = X_t \circ \psi_t.$$

Then $\{\psi_t\}$ is a smooth symplectic action $\psi : \mathbb{R} \to Symp(M,\omega)$. Thus, there is a one-to-one correspondence:

$$\{\textit{symplectic actions of } \mathbb{R} \textit{ on } M\} \leftrightarrow \{\textit{time-dependent symplectic vector fields on } M\}$$

Proof. Under either assumption, it is true that:

$$\frac{d}{dt}\psi_t^*\omega = \psi_t^*\mathcal{L}_{X_t}\omega = \psi_t^*(i_{X_t}\underbrace{d\omega}_{=0} + di_{X_t}\omega) = \psi_t^* di_X\omega$$

where $d\omega = 0$ since ω is closed. If ψ_t is a symplectomorphism for all $t \in \mathbb{R}$, then $\psi_t^*\omega = \omega$, hence $\frac{d}{dt}\psi_t^*\omega = 0$ and thus $\psi_t^* di_X\omega = 0$, which is only true if $di_{X_t}\omega = 0$, i.e., X_t is closed. Conversely, if X_t is a time-dependent family of vector fields, then X_t is closed for all $t \in \mathbb{R}$. Therefore $di_{X_t}\omega = 0$, hence $\frac{d}{dt}\psi_t^*\omega = 0$, and since $\psi_0 = \text{Id}$ so $\psi_0^*\omega = \omega$, we must have $\psi_t^*\omega = \omega$ and so $\psi : \mathbb{R} \to \text{Symp}(M,\omega)$ must be a smooth symplectic action. □

As a side note, given a complete vector field X, this proposition shows that the flow of X, $\{\exp tX : M \to M | t \in \mathbb{R}\}$ defined as the unique family of diffeomorphisms satisfying:

$$\exp tX\Big|_{t=0} = \text{Id and } \frac{d}{dt}\exp tX = X \circ \exp tX.$$

is a smooth symplectic action.

Definition 3.4. Let (M,ω) be a symplectic manifold. Given any smooth function $H : M \to \mathbb{R}$ by the nondegeneracy of ω we can define a vector field X_H on M by:

$$i_{X_H}\omega = dH.$$

We then call H a **Hamiltonian function** and X_H a **Hamiltonian vector field**.

Note that since:

$$dH(X_H) = i_{X_H}\omega(X_H) = \omega(X_H, X_H) = 0$$

we conclude that the X_H is tangent to the level sets of H.

Definition 3.5. Since $di_{X_H}\omega = ddH = 0$, we automatically get that X_H is a symplectic vector field, and thus if M is compact, the flow ψ of X_H is a smooth symplectic action. We then say that ψ is a **Hamiltonian action of** $\mathbb{R}$.

4. Lie Groups

Definition 4.1. Recall that a **Lie group** is a group G which is also a smooth manifold, and where the operations of multiplication and inversion are smooth maps.

Definition 4.2. Let G be a Lie group. Given $g \in G$, we can define left multiplication by g as $L_g : G \to G$ given by $a \mapsto g \cdot a$. A vector field X on G is called **left-invariant** if $(L_g)_*X = X$ for every $g \in G$.

Proposition 4.3. *The set $\mathfrak{g}$ of all left-invariant vector fields on G, together with the Lie bracket $[\cdot,\cdot]$ is a Lie algebra, which we call the **Lie algebra of the Lie group** G.*

Proposition 4.4. *The map from $\mathfrak{g} \to T_eG$ given by $X \mapsto X_e$ (that is, it sends a left invariant vector field to its value at the identity e of G) is an isomorphism of vector spaces. In this way, we can identify $\mathfrak{g}$ with the vector space T_eG.*

Definition 4.5. The derivative at the identity of the map

$$\begin{aligned} \psi_g :& G \longrightarrow G \\ & g \mapsto g \cdot a \cdot g^{-1} \end{aligned}$$

gives an invertible linear map $\mathrm{Ad}_g : \mathfrak{g} \to \mathfrak{g}$ (where we have identified $\mathfrak{g}$ with T_eG). By varying g, we get an action of G on $\mathfrak{g}$, called the **adjoint action**, given by

$$\begin{aligned} \mathrm{Ad} :& G \to \mathrm{GL}(\mathfrak{g}) \\ & g \mapsto \mathrm{Ad}_g. \end{aligned}$$

Definition 4.6. Let $\mathfrak{g}^*$ be the dual vector space of $\mathfrak{g}$. We let $\langle \cdot, \cdot \rangle$ be the pairing of $\mathfrak{g}^*$ and $\mathfrak{g}$, that is,

$$\begin{aligned} \langle \cdot, \cdot \rangle :& \mathfrak{g}^* \times \mathfrak{g} \to \mathbb{R} \\ & (\xi, X) \mapsto \langle \xi, X \rangle = \xi(X). \end{aligned}$$

This allows us to define a map $\mathrm{Ad}_g^* : \mathfrak{g}^* \to \mathfrak{g}^*$: given $\xi \in \mathfrak{g}^*$ we define $\mathrm{Ad}_g^* \xi$ by $\langle \mathrm{Ad}_g^* \xi, X \rangle = \langle \xi, \mathrm{Ad}_{g^{-1}} X \rangle$ for any $X \in \mathfrak{g}$. By varying g, we get an action of G on $\mathfrak{g}^*$, called the **coadjoint action**, given by

$$\begin{aligned} \mathrm{Ad}^* :& G \to \mathrm{GL}(\mathfrak{g}^*) \\ & g \mapsto \mathrm{Ad}_g^*. \end{aligned}$$

If our Lie group G is abelian, it is easy to see that $\mathrm{Ad}_g = \mathrm{Id}$ on $\mathfrak{g}$ and $\mathrm{Ad}_g^* = \mathrm{Id}$ on $\mathfrak{g}^*$ for all $g \in G$. Since the Lie group $\mathbb{T}^m$ is abelian, we need not concern ourselves with the previous definitions; we state these properties solely so that we may formally define the moment map properly in the next section.

5. Moment Maps

Definition 5.1. Let (M, ω) be a symplectic manifold. A **smooth symplectic action of a Lie group** G is a group homomorphism $\psi : G \to \mathrm{Symp}(M, \omega)$ such that the evaluation map $ev_\psi : M \times G \to M$ given by $ev_\psi(p, g) = \psi_g(p)$ is smooth.

Definition 5.2. Given a vector $\xi \in \mathfrak{g}$ where $\mathfrak{g}$ is the Lie algebra of G, we define the **infinitesimal action of** ξ as the vector field X_ξ on M defined by:

$$X_\xi = \frac{d}{dt}\bigg|_{t=0} \psi_{\exp(t\xi)}.$$

We note that since $\mathbb{R} \to \mathrm{Symp}(M, \omega) : t \mapsto \psi_{\exp(t\xi)}$, we automatically get that X_ξ is a symplectic vector field.

Definition 5.3. Suppose that (M, ω) is a symplectic manifold, G is a Lie group, $\mathfrak{g}$ is the Lie algebra of G, $\mathfrak{g}^*$ is the dual vector space of $\mathfrak{g}$, and $\psi : G \to \mathrm{Symp}(M, \omega)$ is a symplectic action. Then we say that ψ is a **Hamiltonian action** if there exists a map

$$\mu : M \to \mathfrak{g}^*$$

which we call the **moment map**, and which satisfies:

(1) For each $\theta \in \mathfrak{g}$, we define $H_\theta : M \to \mathbb{R}$ by $H_\theta(p) = \langle \mu(p), \theta \rangle$. Then H_θ is the Hamiltonian function for the vector field X_θ:

$$dH_\theta = i_{X_\theta}\omega.$$

(2) μ is *equavariant* to the action of ψ of G on M and the coadjoint action Ad^* of G on $\mathfrak{g}^*$:

$$\mu \circ \psi_g = \mathrm{Ad}_g^* \circ \mu$$

for all $g \in G$.

In the case where G is abelian, then since $\mathrm{Ad}_g^* = \mathrm{Id}$ for all $g \in G$, the second condition simplifies to:

$$\mu \circ \psi_g = \mu.$$

6. Morse-Bott Functions

Definition 6.1. Let M be any compact Riemannian manifold. A smooth function $f : M \to \mathbb{R}$ is a **Morse-Bott function** if the critical set $\mathrm{Crit}(f) = \{x \in M | df(x) = 0\}$ decomposes into finitely many connected submanifolds of M, which we shall call the **critical manifolds**, and the tangent space of the critical set coincides with $\ker \nabla^2 f$. That is, for every $x \in \mathrm{Crit}(f)$,

$$T_x\mathrm{Crit}(f) = \ker \nabla^2 f(x)$$

Notice that the definition of a Morse function is a special case of a Morse-Bott function where the critical manifolds are all zero dimensional, and hence for any $x \in \mathrm{Crit}(f)$ we have $\ker \triangledown^2 f(x) = 0$, and therefore the Hessian is nondegenerate.

To make a bit more intuitive sense of this definition, it is useful to consider the following definition:

Definition 6.2. Let M be a compact Riemannian manifold, let $f : M \to M$ be a diffeomorphism, and let L be a f invariant subset of M. We say that L is a **normally hyperbolic invariant manifold** if for any point $x \in L$ the tangent space T_xM splits as a direct sum of three subbundles:

$$T_xM = T_xL \oplus E_x^+ \oplus E_x^-$$

where, with respect to some Riemannian metric on M:

(1) the restriction of df to E^+, called the **stable bundle**, is a contraction
(2) the restriction of df to E^-, called the **unstable bundle**, is an expansion
(3) the restriction of df to TL is relatively neutral.

In other words, there must exist constants $0 < \kappa < \delta^{-1} < 1$ and $0 < c$ such that:

(1) $df_xE_x^+ = E_{f(x)}^+$ and $df_xE_x^- = E_{f(x)}^-$ for all $x \in L$
(2) $\|df^n v\| \leq c\kappa^n \|v\|$ for all $v \in E^+$ and $n > 0$
(3) $\|df^{-n} v\| \leq c\kappa^n \|v\|$ for all $v \in E^-$ and $n > 0$
(4) $\|df^{-n} v\| \leq c\delta^n \|v\|$ for all $v \in TL$ and $n > 0$.

This definition allows us to make the following claim: if f is a Morse-Bott function then its critical manifolds are all normally hyperbolic invariant manifolds with respect to the negative gradient flow. More explicitly, the negative gradient flow is the family of diffeomorphisms $\phi_t : M \to M$ defined by $\frac{d}{dt}\phi_t = -\nabla f \circ \phi_t$ and $\phi_0 = \mathrm{id}$ for $t \in \mathbb{R}$. Then for any critical manifold C, and for any point $x \in C$, the tangent space T_xM decomposes as a direct sum:

$$T_xM = T_xC \oplus E_x^+ \oplus E_x^-$$

where E_x^+ is spanned by the positive eigenspaces and E_x^- is spanned by the negative eigenspaces of $\nabla^2 f(x)$. Additionally, since $\ker \nabla^2 f(x) = T_x C$, we see that $d\phi_t(x)$ is relatively neutral on $T_x C$, and $d\phi_t(x)$ is a contraction and an expansion on E_x^+ and E_x^-, respectively. Armed with this interpretation, we can construct the following definitions:

Definition 6.3. The set of points $x \in M$ whose trajectories $\phi_t(x)$ converge to some point in C as $t \to \infty$ form a manifold called the **stable manifold**, denoted $W^s(C)$. Additionally, for any point $x \in C$, $T_x W^s(C) = T_x C \oplus E_x^+$. Similarly, the set of points $x \in M$ whose trajectories $\phi_t(x)$ converge to some point in C as $t \to -\infty$ form a manifold called the **unstable manifold**, denoted $W^u(C)$. Additionally, for any point $x \in C$, $T_x W^u(C) = T_x C \oplus E_x^-$.

Because M is compact, its image $f(M) \subset \mathbb{R}$ must also be compact, and therefore has a minimum and maximum. Therefore, for any point $x \in M$, since f decreases along the trajectory $\phi_t(x)$ as $t \to \infty$, it follows that the trajectory must converge to some critical manifold C as $t \to \infty$. Thus:

$$M = \bigcup_C W^s(C)$$

By the same logic, for any point $x \in M$, since f increases along the trajectory $\phi_t(x)$ as $t \to -\infty$, it follows that the trajectory must converge to some critical manifold C as $t \to -\infty$. Thus:

$$M = \bigcup_C W^u(C)$$

And finally, we will need the following definitions:

Definition 6.4. The **index** of a critical manifold C is defined by:

$$n^-(C) = \dim W^u(C) - \dim C = \mathrm{codim} W^s(C).$$

Likewise, the **coindex** of a critical manifold C is defined by:

$$n^+(C) = \dim W^s(C) - \dim C = \mathrm{codim} W^u(C).$$

The Jordan-Brouwer Separation Theorem states that any compact hypersurface in $\mathbb{R}^n$ disconnects $\mathbb{R}^n$ into an 'inside' and an 'outside'. It is easy to see that this is *not* true for any embedded manifold of codimension not equal to 1: if M is a compact manifold embedded in $\mathbb{R}^n$, and $\mathrm{codim}(M) \neq 1$, then $\mathbb{R}^n - M$ is connected. Similarly, it is true that for *any* submanifold N of a compact manifold M with codimension greater than 1, the complement $M - N$ must be connected. Intuitively, if codim $\neq 1$, there is 'enough room to move around' to avoid being disconnected. The next lemma extends this basic intuition to a consideration of the level sets of a Morse-Bott function:

Lemma 6.5. *Suppose M is a compact connected manifold and $f : M \to \mathbb{R}$ is a Morse-Bott function such that for any of the critical manifolds C of f we have $n^\pm(C) \neq 1$. Then for every $c \in \mathbb{R}$ the level set $f^{-1}(c)$ is connected.*

Proof.

(1) There is exactly one connected critical manifold of index zero, and exactly one connected critical manifold of coindex zero

It is easy to see that there must be at *least* one critical manifold of index zero; if there were not, then $M = \bigcup_C W^s(C)$ would consist solely of a finite union of stable manifolds all of codimension greater than or equal to 2, which is impossible.

To see that there is *only* one such critical manifold, let C_0 be the union of all critical manifolds of index zero. Then the $M - W^s(C_0)$ consists of the stable manifolds of the other critical manifolds, and is therefore a union of submanifolds of codimension at least 2. It therefore follows by the previous discussion that $W^s(C_0)$ is connected, and therefore that C_0 is connected; for if C_0 were not connected, i.e., $C_0 = U \cup V$ and $U \cap V = \emptyset$, then we would have $W^s(C_0) = W^s(U) \cup W^s(V)$ and $W^s(U) \cap W^s(V) = \emptyset$, hence we would have $W^s(C_0)$ not connected, a contradiction. Similar reasoning shows that there is exactly one connected critical manifold of coindex zero.

Notice also, that if a critical manifold is a local minimum or maximum of f, then it must be of index zero or coindex zero, respectively. Since there is only one critical manifold of index zero, and one of coindex zero, we see that f has a unique local minimum (which is hence the minimum) and a unique local maximum (which is hence the maximum). Therefore, the critical manifold of index zero is where f attains its minimum, and the critical manifold of coindex zero is where f attains its maximum.

(2) $f^{-1}(c)$ is connected for every regular value $c \in \mathbb{R}$

Let $c_0 < c_1 < \ldots < c_N$ be the critical levels of f. Then $C_0 = f^{-1}(c_0)$ is the connected critical manifold of index zero, and $C_N = f^{-1}(c_N)$ is the connected critical manifold of coindex zero.

First, we prove that $f^{-1}(c)$ is connected for $c_0 < c < c_1$. To do this, take any two points $x_0, x_1 \in f^{-1}(c)$, and note that the trajectories $\phi_t(x_0)$ and $\phi_t(x_1)$ must converge to points $y_0, y_1 \in C_0$ as $t \to \infty$. Thus, we can join x_0 to x_1 by follow the flowlines of ϕ_t from x_0 to y_0, and x_1 to y_1, and then connect y_0 to y_1 in C_0, since C_0 is connected. We then only need notice that $\text{codim} C_0 = \dim M - \dim C_0 = \dim W^s(C_0) - \dim C_0 = n^+(C_0) \geq 2$, and thus consideration of dimensions and the Stability Theorem of transversality allows us to move our path slightly so it does not intersect C_0. From here, we can move the path up to the level of c via the gradient flow, leaving a path in $f^{-1}(c)$ from x_0 to x_1.

From here, we suppose by induction that $f^{-1}(c)$ is connected for regular values $c < c_k$. Suppose, then, that we have a regular value c with $c_k < c$. Take any two points $x_0, x_1 \in f^{-1}(c)$, and connect them via paths in $f^{-1}(c)$ to points in $W^s(C_0)$. From here we can connect these points in $W^s(C_0)$ to points in $f^{-1}(c_k - \epsilon)$ using the downward gradient flow. These resulting points can be joined together since by our inductive assumption, $f^{-1}(c_k - \epsilon)$ is connected. Again, by the Stability Theorem, we can move this path slightly so that it is transversal to all of the unstable manifolds. Since $\text{codim} W^u(C_i) \geq 2$ for all $i \neq N$, our path must lie entirely within $W^u(C_N)$. We can now use the flow to move this path back up to the level of $f^{-1}(c)$.

This proves, therefore, that $f^{-1}(c)$ is connected for every regular value $c \in \mathbb{R}$.

(3) $f^{-1}(c_j)$ is connected for the remaining critical values $0 < j < N$

Choose a regular value $c > c_j$ such that there are no critical values between c and c_j. Then we can define a continuous surjection by $\psi : f^{-1}(c) \to f^{-1}(c_j)$ defined by:

$$\psi(x) = \begin{cases} \lim_{t\to\infty} \phi_t(x) & \text{if } f(\phi_t(x)) > c_j \text{ for all } t > 0 \\ \psi_t(x) & \text{if } f(\phi_t(x)) = c_j \text{ for some } t. \end{cases}$$

The fact that f is Morse-Bott shows that ψ is surjective, and a consideration of limits of gradient flow lines shows that ψ is continuous. Therefore, we may conclude that $f^{-1}(c_j)$ is connected.

Taken as a whole, we see that the proof is complete. □

7. Precursors to Convexity

Lemma 7.1. *Suppose that (M, ω) is a compact connected symplectic manifold with a symplectic action of a compact group $G \to Symp(M, \omega) : \tau \mapsto \psi_\tau$. Then there exists an almost complex structure J on M which is compatible with ω and invariant under the action of G. By 'invariant under the action of G', we mean that $\psi_\tau^* J = J$ for every $\tau \in G$.*

Proof. Simply take any Riemannian metric g' and average (which we can do, since G is assumed to be compact) to obtain an invariant metric g: in other words:

$$g_p(X, Y) = \int_{\tau \in G} g'_p(d\psi_\tau X, d\psi_\tau Y) d\tau$$

for any vectors X, Y in any tangent space T_pM. Together with the symplectic form ω, this invariant g induces a compatible almost complex structure J. Thus, for any ψ_τ, we have:

$$\begin{array}{l} g_p(X, Y) = \omega_p(X, J_pY) = \psi_\tau^* \omega_p(X, J_pY) = \omega_{\psi_\tau(p)}(d\psi_\tau(p)X, d\psi_\tau(p)J_pY)) \\ \quad \| \\ \psi_\tau^* g_p(X, Y) = g_{\psi_\tau(p)}(d\psi_\tau(p)X, d\psi_\tau(p)Y) = \omega_{\psi_\tau(p)}(d\psi_\tau(p)X, J_{\psi_\tau(p)} d\psi_\tau(p)Y). \end{array}$$

for any vectors X, Y in any tangent space T_pM. By the nondegeneracy of ω, we must have $d\psi_\tau(p)J_pY = J_{\psi_\tau(p)} d\psi_\tau(p)Y$, i.e., $J_pY = (d\psi_\tau(p))^{-1} J_{\psi_\tau(p)} d\psi_\tau(p)Y = \psi_\tau^* J_pY$. Hence $\psi_\tau^* J = J$ as required. □

Proposition 7.2. *Let $H \subset G$ be a subgroup. Let **Fix**$(H) \subset M$ be the set of points of M fixed by every symplectomorphism in $Im(H) \subset Symp(M, \omega)$, that is*

$$Fix(H) = \bigcap_{h \in H} Fix(\psi_h).$$

Then $Fix(H)$ is a submanifold of M.

For the next lemma, we must recall the definition of the **exponential map** with respect to some chosen Riemannian metric g. Given a point $x \in M$, and a vector $\xi \in T_xM$ there is a unique geodesic γ (determined by g) satisfying $\gamma(0) = x$ with initial velocity $\gamma'(0) = \xi$. We can then define the exponential map $\exp_x : T_xM \to M$ by $\exp_x(\xi) = \gamma(1)$.

Lemma 7.3. *Let $H \subset G$ be a subgroup. Let **Fix**$(H) \subset M$ be the set of points of M fixed by every symplectomorphism in $Im(H) \subset Symp(M, \omega)$, that is*

$$Fix(H) = \bigcap_{h \in H} Fix(\psi_h).$$

Then $Fix(H)$ is a symplectic *submanifold of M.*

Proof. Let $x \in \text{Fix}(H)$. For any $h \in H$, Lemma 7.1 proves that $d\psi_h(x) : T_xM \to T_xM$ (the differential of the symplectomorphism ψ_h) is a unitary action of G on the complex vector space (T_xM, ω, J_x). Given a vector $\xi \in T_xM$ there is a unique geodesic γ (determined by our previously determined invariant Riemannian metric g) with $\gamma(0) = x$ and initial velocity $\gamma'(0) = \xi$. Then $\exp_x(\xi) = \gamma(1)$. Since g is invariant under the action of G, it necessarily follows that since γ is a geodesic, so is $\psi_h \circ \gamma$. Thus we have a geodesic $\psi_h \circ \gamma$ with $\psi_h \circ \gamma(0) = \psi_h(x) = x$ and initial velocity $(\psi_h \circ \gamma)'(0) = d\psi_h(x) \circ \gamma'(0) = d\psi_h(x)\xi$. Hence, $\exp_x(d\psi_h(x)\xi) = \psi_h \circ \gamma(1) = \psi_h(\exp_x(\xi))$. Specifically:

$$\exp_x(d\psi_h(x)\xi) = \psi_h(\exp_x(\xi)).$$

Thus, there is a correspondence between the points fixed by ψ_h and the vectors fixed by $d\psi_h$. We can therefore concluded that:

$$T_x\mathrm{Fix}(H) = \bigcap_{h\in H} \ker(1 - d\psi_h(x))$$

Explicitly, if we take a vector $\xi \in T_x\mathrm{Fix}(H)$ we can restrict $\exp_x$ to $\exp_x|_{T_x\mathrm{Fix}(H)} : T_x\mathrm{Fix}(H) \to \mathrm{Fix}(H)$ to see that $\exp_x(\xi) \in \mathrm{Fix}(H)$. Thus, for any $h \in H$ we have $\psi_h(\exp_x(\xi)) = \exp_x(\xi) = \exp_x(d\psi_h(x)\xi)$, and provided ξ is small, $\exp_x$ is injective. Thus, ξ is fixed by $d\psi_h(x)$ for any $h \in H$ and so $\xi \in \bigcap_{h\in H} \ker(1 - d\psi_h(x))$. Alternatively, if we take a vector $\xi \in \bigcap_{h\in H} \ker(1 - d\psi_h(x))$, for any $h \in H$ we can get a geodesic $\gamma : [-1, 1] \to M$ by $\gamma(t) = \exp_x(d\psi_h(x)t\xi)$. Then, for any $h \in H$, $\gamma(t) = \exp_x(d\psi_h(x)t\xi) = \exp_x(t\xi) = \psi_h(\exp_x(t\xi))$. Hence, $\gamma[-1, 1] \subset \bigcap_{h\in H} \mathrm{Fix}(\psi_h) = \mathrm{Fix}(H)$, thus $\gamma'(0) = \xi \in T_x\mathrm{Fix}(H)$.

We can use this to prove that $\mathrm{Fix}(H)$ is a symplectic submanifold. Now, let $\xi \in T_x\mathrm{Fix}(H)$. Then for every $h \in H$, ξ is fixed by $d\psi_h(x)$. Since J_x is invariant under the action of G, we have $d\psi_h(x)J_x(\xi) = J_x d\psi_h(x)(\xi) = J_x(\xi)$. Thus, for every $h \in H$, $d\psi_h(x)J_x(\xi) = J_x(\xi)$ and so $J_x(\xi) \in \bigcap_{h\in H} \ker(1 - d\psi_h(x)) = T_x\mathrm{Fix}(H)$. Therefore, for every $x \in \mathrm{Fix}(H)$, $T_x\mathrm{Fix}(H)$ is a symplectic vector space, and we conclude that $\mathrm{Fix}(H)$ is a symplectic submanifold. □

Lemma 7.4. *Suppose that (M, ω) is a compact connected symplectic manifold with Hamiltonian torus action $\mathbb{T}^m \to Symp(M, \omega) : \theta \mapsto \psi_\theta$ with moment map $\mu : M \to \mathbb{R}^m$. For every $\theta \in \mathfrak{g}^* = \mathbb{R}^m$, let H_θ be the associated Hamiltonian function $H_\theta = \langle \mu, \theta \rangle : M \to \mathbb{R}$. Then the critical set of H_θ is equal to the set of points of M fixed by every symplectomorphism in $Im(T_\theta) \subset Symp(M, \omega)$, where $T_\theta = cl(\{t\theta + k | t \in \mathbb{R}, k \in \mathbb{Z}^m\}/\mathbb{Z}^m)$. In other words,*

$$Crit(H_\theta) = \bigcap_{\tau\in T_\theta} Fix(\psi_\tau).$$

Lastly, and most importantly, H_θ is a Morse-Bott function which has critical set $Crit(H_\theta)$ a symplectic submanifold, and critical manifolds which are both even dimensional, and of even index and coindex.

Proof. We get the vector field X_{H_θ} on M from solving $i_{X_{H_\theta}}\omega = dH_\theta$, and by properties of moment maps, we know that X_{H_θ} is equal to the vector field generated on M by the one-parameter subgroup $\{\exp(t\theta) | t \in \mathbb{R}\} \subset G$, and thus we also have:

$$\frac{d}{dt}\,\psi_{t\theta} = X_{H_\theta} \circ \psi_{t\theta}.$$

Suppose then, that $x \in \mathrm{Crit}(H_\theta)$. Then $dH_\theta(x) = 0$, and since $i_{X_{H_\theta}}\omega = dH_\theta$, we must have $X_{H_\theta}(x) = 0$. Thus, $\frac{d}{dt}\,\psi_{t\theta}(\psi_{t\theta}^{-1}(x)) = 0$, and since $\psi_0(x) = x$, we must have $\psi_t(x) = x$ for all $t \in \mathbb{R}$. It follows by continuity that x is fixed by the symplectomorphisms in the closure, as well. Thus, $x \in \bigcap_{\tau\in T_\theta} \mathrm{Fix}(\psi_\tau)$.

Alternatively, suppose that $x \in \bigcap_{\tau \in T_\theta} \mathrm{Fix}(\psi_\tau)$. Then $0 = \frac{d}{dt}\psi_{t\theta}(x) = X_{H_\theta} \circ \psi_{t\theta}(x) = X_{H_\theta}(x)$, and so $i_{X_{H_\theta}(x)}\omega_x = dH_\theta(x) = 0$. Thus, $x \in \mathrm{Crit}(H_\theta)$. Therefore, $\mathrm{Crit}(H_\theta) = \bigcap_{\tau \in T_\theta} \mathrm{Fix}(\psi_\tau)$ as claimed.

It then follows by Lemma 7.3 for the subgroup $T_\theta \subset \mathbb{T}^m$ that $\mathrm{Crit}(H_\theta)$ is a symplectic submanifold of M and therefore, has finitely many components. At any point $x \in M$, consider the Hessian $\nabla^2 H_\theta(x) : T_xM \to T_xM$. We claim that $dX_{H_\theta}(x) = -J_x\nabla^2 H_\theta(x)$ and therefore, $d\psi_{\exp(t\theta)}(x) = \exp(-tJ_x\nabla^2 H_\theta(x))$, and so we can conclude that the kernel of $\nabla^2 H_\theta(x)$ equals the fixed points of $d\psi_{\exp(t\theta)}(x)$. By continuity, we see that:

$$T_x\mathrm{Crit}(H_\theta) = \bigcap_{\tau \in T_\theta} \ker(\mathrm{Id} - d\psi_\tau(x)) = \ker \nabla^2 H_\theta(x).$$

This proves that H_θ is a Morse-Bott function. We now claim that since each $d\psi_{\exp(t\theta)}(x) = \exp(-tJ_x\nabla^2 H_\theta(x))$ is unitary, that $\nabla^2 H_\theta(x)$ commutes with J_x; therefore the eigenspaces of $\nabla^2 H_\theta(x)$ are invariant with J_x, and must therefore be even dimensional. Thus, we see that the critical manifolds of H_θ are even dimensional (since they are symplectic) and are of even index and coindex. □

Definition 7.5. We denote the components of the moment map $\mu : M \to \mathbb{R}^m$ as $\mu = (\mu_1, \ldots, \mu_m)$. We say that μ is **irreducible** if the 1-forms $d\mu_1, \ldots, d\mu_m$ are linearly independent, i.e., given a scalar $(\alpha_1, \ldots, \alpha_m) \in \mathbb{R}^m$, then

$$\alpha_1 d\mu_1(x)(\xi) + \ldots + \alpha_m d\mu_m(x)(\xi) = 0$$

at all points $x \in M$ and all vectors $\xi \in T_xM$ if and only if $\alpha_1 = \ldots = \alpha_m = 0$. We say the μ is **reducible** otherwise.

Definition 7.6. We say that a set of real numbers $\{\theta_i | 1 \leq i \leq s, \theta_i \in \mathbb{R}\}$ is **rationally dependent** if $\frac{\theta_i}{\theta_j}$ is rational for all nonzero $\theta_{i,j}$ with $1 \leq i, j \leq s$.

Proposition 7.7. *If μ is reducible, then we can reduce it to an action of an $(m-1)$-torus. Specifically, there exists a Hamiltonian torus action $\mathbb{T}^{m-1} \to Symp(M, \omega)$: $\tau \mapsto \psi'_\tau$ with moment map $\mu' : M \to \mathbb{R}^{m-1}$ and an integer matrix $A \in \mathbb{Z}^{(m-1)\times m}$ such that, for $\theta \in \mathbb{T}^m$ and $x \in M$:*

$$\psi_\theta = \psi'_{A\theta} \text{ and } \mu(x) = A^T\mu'(x).$$

Proof. Note that we have $\mathfrak{g} = \mathbb{R}^m$ and $\mathfrak{g}^* = \mathbb{R}^m$, and that given $\theta = (\theta_1, \ldots, \theta_m) \in \mathfrak{g}$ and $\mu(p) = (\mu_1(p), \ldots, \mu_m(p)) \in \mathfrak{g}^*$, then $\langle \mu(p), \theta \rangle = \sum_{i=1}^m \theta_i \mu_i(p)$. Therefore, we can write the Hamiltonian action $H_\theta = \langle \mu, \theta \rangle$ as:

$$H_\theta = \sum_{i=0}^m \theta_i \mu_i.$$

Then we also have:

$$dH_\theta = \theta_1 d\mu_1 + \ldots + \theta_m d\mu_m.$$

By assumption μ is reducible, and therefore there must exist some nonzero $\theta = (\theta_1, \ldots, \theta_m) \in \mathbb{R}^m$ such that $dH_\theta(x)(\xi) = 0$ at all points $x \in M$ and all vectors $\xi \in T_xM$. It follows therefore, that $H_\theta : M \to \mathbb{R}$ is constant for this θ. Then we also note that $H_{t\theta} =$ constant and thus $dH_{t\theta} = 0$ for all $t \in \mathbb{R}$. Since $i_{X_{H_{t\theta}}}\omega = dH_{t\theta}$, we have $X_{H_{t\theta}} = 0$ and thus $\psi_{\exp(t\theta)} = \mathrm{Id}$ for all $t \in \mathbb{R}$. Lastly, note that $\exp : \mathbb{R}^m \to \mathbb{T}^m$ is the same as the natural projection $\pi : \mathbb{R}^m \to \mathbb{T}^m$.

Now, take a maximal rationally dependent subset $\{\theta_{i_1}, \ldots, \theta_{i_l} | 1 \leq i_1 < \ldots < i_l \leq m\} \subset \{\theta_1, \ldots, \theta_m\}$, and reorder indices so that this subset is in the first l spots. It is an exercise to show that the projection of the line $L_a = \{t(\theta_1, \ldots, \theta_l) | t \in \mathbb{R}\} \subset \mathbb{R}^l$ to the torus $\mathbb{T}^l$ via the natural projection $\pi : \mathbb{R}^l \to \mathbb{R}^l/\mathbb{Z}^l$ 'closes up', i.e., $\pi|_{L_a}$ is *not* surjective. If $l \neq 1$, then we can say additionally that $\pi(L_a)$ is *not* dense in $\mathbb{T}^l$. It is also an exercise to show that the projection of the line $L_b = \{t(\theta_{l+1}, \ldots, \theta_m) | t \in \mathbb{R}\} \subset \mathbb{R}^{m-l}$ to the torus $\mathbb{T}^{m-l}$ via the natural projection $\pi : \mathbb{R}^{m-l} \to \mathbb{T}^{m-l}$ is dense, i.e., $cl\{\pi(L_b)\} = \mathbb{T}^{m-l}$. If $m - l \neq 1$, then we can say additionally that $\pi(L_b)$ does *not* 'close up'.

Thus, we conclude that we can find a rationally dependent direction $\nu \in cl\{\exp(t\theta) | t \in \mathbb{R}\} = cl\{\pi(L)\} \subset \mathbb{T}^m$ (where $L = \{t\theta | t \in \mathbb{R}^m\}$). For the first l positions, take $\pi(\theta_1, \ldots, \theta_l)$ and for the last $m - l$ positions we can choose compatible values since we have all of $cl\{\pi(L_b)\} = \mathbb{T}^{m-l}$ to pick from; for instance, we could take θ_1 for all of the remaining positions. Since $\psi_\gamma = \mathrm{Id}$ for every $\gamma \in \{\exp(t\theta) | t \in \mathbb{R}\} = \pi(L)$, we deduce by continuity that since $\nu \in cl\{\exp(t\theta) | t \in \mathbb{R}\} = cl\{\pi(L)\}$ we must have $\psi_\nu = \mathrm{Id}$.

It is immediate that our previous observations about θ are true for ν as well (when we consider ν as an element of $\mathbb{R}^m$): in particular, $\psi_{\exp(t\nu)} = id$ for all $t \in \mathbb{R}$. Thus, we may quotient out the direction of ν: it is easy to show that $\mathbb{R}^m/L \cong \mathbb{R}^m = \mathbb{R}^m \cap \nu^\perp$, where $L = \{t\nu | t \in \mathbb{R}^m\}$ and $\nu^\perp$ is the unique plane in $\mathbb{R}^m$ normal to ν. However, we claim it is only because ν is rationally dependent that $\mathbb{R}^m \cap \nu^\perp/\mathbb{Z}^m \cong \mathbb{R}^{m-1}/\mathbb{Z}^{m-1}$ (this is because ν is rationally dependent, there must be some nonzero vector with integer components in $\mathbb{R}^m \cap \nu^\perp$, etc.). Then the matrix that takes $\mathbb{R}^m \cap \nu^\perp/\mathbb{Z}^m \subset \mathbb{R}^m/\mathbb{Z}^m$ to $\mathbb{R}^{m-1}/\mathbb{Z}^{m-1}$ is an integer matrix $A \in \mathbb{Z}^{(m-1)\times m}$. This is the required matrix. □

8. Convexity

Theorem 8.1. *(The Atiyah-Guillemin-Sternberg Convexity Theorem) Suppose that (M, ω) is a compact connected symplectic manifold with Hamiltonian torus action $\mathbb{T}^m \to Symp(M, \omega) : \theta \mapsto \psi_\theta$ with moment map $\mu : M \to \mathbb{R}^m$. Then the image of μ is a convex subset of $\mathbb{R}^m$. Specifically, the points of M fixed by every symplectomorphism in $Im(\mathbb{T}^m) \subset Symp(M, \omega)$ are a finite union of connected symplectic submanifolds $C_1, \ldots, C_N$, i.e.*

$$\bigcap_{\theta \in \mathbb{T}^m} Fix(\psi_\theta) = \bigcup_{j=1}^{N} C_j.$$

Furthermore, the image of any of these symplectic submanifolds is constant: $\mu(C_j) = \eta_j \in \mathbb{R}^m$. Lastly, the image of μ itself is given by the convex hull of these points:

$$\mu(M) = K(\eta_1, \ldots, \eta_N)$$

Proof.

(1) By induction over the dimension m of the torus, the preimage $\mu^{-1}(\eta) \subset M$ is connected for every regular value $\eta \in \mathbb{R}^m$

The base case $m = 1$ is almost immediate. We have $\mathbb{T}^m = S^1$ and thus $\mathfrak{g} = \mathbb{R}$ and $\mathfrak{g}^* = \mathbb{R}$, hence the moment map $\mu : M \to \mathbb{R}$ is simply a function. For any $\theta \in \mathfrak{g} = \mathbb{R}$, by Lemma 7.4 we know that H_θ must be Morse-Bott with critical manifolds of even index, and since $H_\theta = \theta \cdot \mu$, if we let $\theta = 1$ we see that μ is also Morse-Bott with critical manifolds of even index. Then by Lemma 6.5, the preimage $\mu^{-1}(\eta)$ must be connected for every $\eta \in \mathbb{R}$.

Suppose by our inductive hypothesis that the assertion is true for any Hamiltonian torus action $\mathbb{T}^{m-1} \to \mathrm{Symp}(M, \omega)$. Consider any Hamiltonian torus action $\mathbb{T}^m \to \mathrm{Symp}(M, \omega)$ with moment map $\mu : M \to \mathbb{R}^m$. If μ is reducible, then by Lemma 7.7 we have $\mu = A^T \circ \mu' : M \xrightarrow{\mu'} \mathbb{R}^{m-1} \xrightarrow{A^T} \mathbb{R}^m$, and any regular value $\eta \in \mathbb{R}^m$ of μ must also be a regular value of $A^T : \mathbb{R}^{m-1} \to \mathbb{R}^m$, and since we must have $(A^T)^{-1}(\eta) = \emptyset$, we also must have $\mu^{-1}(\eta) = \emptyset$. Thus, the preimage of any regular value of a reducible moment map is trivially connected. Therefore, let us assume that μ is irreducible.

If μ is irreducible, then:

$$\sum_{i=1}^{m} \alpha_i d\mu_i(x)(\xi) = 0$$

at all points $x \in M$ and all vectors $\xi \in T_x M$ if and only if $\alpha_1 = \ldots = \alpha_m = 0$, and since $H_\theta = \sum_{i=1}^m \theta_i \mu_i$, we also have:

$$dH_\theta(x) = \sum_{i=1}^{m} \theta_i d\mu_i(x) = 0$$

at all points $x \in M$ and all vectors $\xi \in T_x M$ if and only if $\theta_1 = \ldots = \theta_m = 0$. We conclude that $H_\theta : M \to \mathbb{R}$ is nonconstant for every nonzero vector $\theta \in \mathbb{R}^m$.

Now, consider the set:

$$Z = \bigcup_{\theta \neq 0} \mathrm{Crit}(H_\theta).$$

By Lemma 7.4, we know $\mathrm{Crit}(H_\theta) = \bigcap_{\tau \in T_\theta} \mathrm{Fix}(\psi_\tau)$, as well as that $\mathrm{Crit}(H_\theta)$ is a set of even dimensional proper submanifolds. It is easy to see that the set of fixed points $\bigcap_{\tau \in T_\theta} \mathrm{Fix}(\psi_\tau)$ decreases as the subtorus $T_\theta \subset \mathbb{T}^m$ increases, hence it is sufficient to restrict our attention to 1-dimensional subtori. Explicitly, if the components of θ are not rationally dependent, then as we saw in Lemma 7.7 we can choose a rationally dependent direction $\nu \in T_\theta$. Then T_ν is a 1-dimensional subtorus, and since $T_\nu \subset T_\theta$, we must have $\bigcap_{\tau \in T_\theta} \mathrm{Fix}(\psi_\tau) \subset \bigcap_{\tau \in T_\nu} \mathrm{Fix}(\psi_\tau)$. If we let $R = \{\theta \in \mathbb{R}^m | \theta \neq 0$ and the components of θ are rationally dependent$\}$, it follows that:

$$Z \subset \bigcup_{\theta \in R} \mathrm{Crit}(H_\theta).$$

Notice that if $\tau = t\theta$ for some $t \in \mathbb{R}$, and nonzero $\tau, \theta \in \mathbb{R}^m$ (τ and θ have the same 'direction'), it follows that $T_\tau = T_\theta$ so that $\mathrm{Crit}(H_\tau) = \mathrm{Crit}(H_\theta)$, therefore it is only the 'direction' that matters. Since there are only countably many rationally dependent directions in $\mathbb{R}^m$, we know there are only countably many distinct critical sets $\mathrm{Crit}(H_\theta)$ in $\bigcup_{\theta \in R} \mathrm{Crit}(H_\theta)$. And since each critical set $\mathrm{Crit}(H_\theta)$ is a set of even dimensional proper submanifolds, we can conclude that Z is a countable union of proper submanifolds of M. Because Z is a countable union, an application of Baire's Category Theorem tells us that $M - Z$ must be dense in M. We lastly note that $M - Z$ is open; a point x is in $M - Z$ if and only if $dH_\theta(x) = \sum_{i=1}^m \theta_i d\mu_i(x) \neq 0$ for all $\theta \in \mathbb{R}^m$, i.e., if and only if the linear functionals $d\mu_1(x), \ldots, d\mu_m(x)$ are linearly independent. Since $d\mu_1, \ldots, d\mu_m$ must also be linearly independent in a neighborhood of x, it follows that $M - Z$ is open.

We can now show that the regular values of μ are dense in the image $\mu(M) \subset \mathbb{R}^m$. To do this, take any $\eta \in \mu(M)$ and any point $x \in \mu^{-1}(\eta) \subset M$. Since $M - Z$ is dense in M, we can approximate x by a sequence $\{x_i\} \in M - Z$. Then, at any x_i we have $d\mu_1(x_i), \ldots, d\mu_m(x_i)$ linearly independent, and therefore we know that μ takes a sufficiently small neighborhood of x_i to a neighborhood U of $\mu(x_i)$. By Sard's Theorem, we can find a regular value $\eta_i \in \mathbb{R}^m$ which is arbitrarily close to $\mu(x_i)$, i.e., $\eta_i \in U$ so that $\eta_i \in \mu(M)$. Thus we can find a regular value arbitrarily close to $\mu(x) = \eta$, and therefore we conclude the regular values of μ are dense in $\mu(M)$. By nearly identical reasoning, if we let $\lambda = (\mu_1, \ldots, \mu_{m-1}) : M \to \mathbb{R}^{m-1}$ be the reduced moment map, i.e., $\lambda(x) = (\mu_1(x), \ldots, \mu_{m-1}(x))$, then the set of all points $\eta \in \mu(M)$ such that $(\eta_1, \ldots, \eta_{m-1})$ is a regular value of λ is also dense in $\mu(M)$.

We now show that the submanifold $\mu^{-1}(\eta)$ is connected whenever $(\eta_1, \ldots, \eta_{m-1})$ is a regular value of λ. Notice that λ is a moment map for the reduced Hamiltonian torus action $\mathbb{T}^{m-1} \to \mathrm{Symp}(M, \omega) : (\theta_1, \ldots, \theta_{m-1}) \mapsto \psi_{(\theta_1, \ldots, \theta_{m-1}, 0)}$. Therefore, by our inductive hypothesis, if $\eta' \in \mathbb{R}^{m-1}$ is a regular value of λ, then $\lambda^{-1}(\eta') \subset M$ must be connected. In particular, for any $\eta = (\eta_1, \ldots, \eta_m) \in \mathbb{R}^m$, if $(\eta_1, \ldots, \eta_{m-1} \in \mathbb{R}^{m-1}$ is a regular value of λ, then $\lambda^{-1}(\eta_1, \ldots, \eta_{m-1}) \subset M$ is connected, i.e., the submanifold:

$$Q = \lambda^{-1}(\eta_1, \ldots, \eta_{m-1}) = \bigcap_{i=1}^{m-1} \mu_i^{-1}(\eta_i)$$

is connected. Note, also that if we let $\dim M = d$, then $\dim Q = k = \dim M - (m - 1) = d - (m - 1)$. Now, let us consider the restricted function:

$$\mu_m : Q \to \mathbb{R}.$$

We will briefly show that a point $x \in Q$ is critical for μ_m if and only if there exist $\theta_1, \ldots, \theta_{m-1} \in \mathbb{R}$ such that:

$$\sum_{i=1}^{m-1} \theta_i d\mu_i(x)(\xi) + d\mu_m(x)(\xi) = 0 \text{ for all } \xi \in T_x M.$$

If a point $x \in Q$ is critical for μ_m, then $d\mu_m(x)(\zeta) = 0$ for all $\zeta \in T_x Q$. For any vector $\zeta \in T_x Q$ we have $d\mu_i(x)(\zeta) = 0$ for $1 \leq i \leq m-1$, since μ_i is constant on Q. Thus, we need to find $\theta_1, \ldots, \theta_{m-1} \in \mathbb{R}$ such that $\sum_{i=1}^{m-1} \theta_i d\mu_i(x)(\xi) + d\mu_m(x)(\xi) = 0$ for all $\xi \in T_x M - T_x Q$. But, notice that $\dim(T_x M - T_x Q) = m - 1$, and therefore by considering $d\mu_1(x), \ldots, d\mu_m(x)$ as elements of the dual vector space of $T_x M - T_x Q$, we must have a linear dependence, which we can normalize so that $\sum_{i=1}^{m-1} \theta_i d\mu_i(x)(\xi) + d\mu_m(x)(\xi) = 0$ for all $\xi \in T_x M - T_x Q$ and hence also for all $\xi \in T_x M$. On the other hand, suppose for some point $x \in Q$ there exist $\theta_1, \ldots, \theta_{m-1} \in \mathbb{R}$ such that $\sum_{i=1}^{m-1} \theta_i d\mu_i(x)(\xi) + d\mu_m(x)(\xi) = 0$ for all $\xi \in T_x M$. As noted above, for any vector $\zeta \in T_x Q$ we have $d\mu_i(x)(\zeta) = 0$ for $1 \leq i \leq m-1$, and therefore, $0 = \sum_{i=1}^{m-1} \theta_i d\mu_i(x)(\zeta) + d\mu_m(x)(\zeta) = d\mu_m(x)(\zeta)$, i.e., $d\mu_m(x) = 0$ on $T_x Q$. Hence $x \in Q$ is critical for μ_m.

Therefore, x is also a critical point for the Hamiltonian function $H_\theta = \langle \mu, \theta \rangle : M \to \mathbb{R}$ where $\theta = (\theta_1, \ldots, \theta_{m-1}, 1)$. Thus, by Lemma 7.4, we know that H_θ is Morse-Bott with even dimensional critical manifolds of even index. Let $C \subset M$ be the critical manifold of H_θ which contains x. We wish to demonstrate that C

intersects Q transverally, i.e.:

$$T_xM = T_xQ + T_xC.$$

The easiest way to do this is in our case is to show that the dual vector space to $T_xQ + T_xC$ has the same dimension as T_xM. To do this, it is enough to find d linearly independent linear functionals ξ_i^* on $T_xQ + T_xC$. Since $\dim T_xQ = k$, we can pick a basis $e_1, \ldots, e_k \in T_xQ$ and a corresponding dual basis $e_1^*, \ldots, e_k^*$. Since $d\mu_1(x), \ldots, d\mu_{m-1}(x)$ vanish on T_xQ, if we can show that $d\mu_1(x), \ldots, d\mu_{m-1}(x)$ are linearly independent on T_xC, then it is easy to check that, given:

$$\alpha_1, \ldots, \alpha_k, \beta_1, \ldots, \beta_{m-1} \in \mathbb{R}$$

then:

$$\sum_{i=1}^{k} \alpha_i e_i^*(\xi) + \sum_{j=1}^{m-1} \beta_j d\mu_j(x)(\xi) = 0$$

for all $\xi \in T_xQ + T_xC$ if and only if $\alpha_1 = \ldots = \alpha_k = \beta_1 = \ldots = \beta_{m-1} = 0$. Since $k + (m-1) = d - (m-1) + (m-1) = \dim T_xM$, this would prove that $T_xQ + T_xC$ has the same dimension as T_xM, and therefore, $T_xM = T_xQ + T_xC$. Thus, all we have to do is prove that $d\mu_1(x), \ldots, d\mu_{m-1}(x)$ remain linearly independent on T_xC.

To begin with, we know that $d\mu_1(x), \ldots, d\mu_{m-1}(x)$ are linearly independent on all of T_xM since x is a regular point of $\lambda : M \to \mathbb{R}^{m-1}$. Then the Hamiltonian vector fields X_{μ_i} given by $d\mu_i = i_{X_{\mu_i}}\omega$ are also linear independent at x, i.e.:

$$\sum_{i=1}^{m-1} \alpha_i d\mu_i(x)(\xi) = \omega_x(\sum_{i=1}^{m-1} \alpha_i X_{\mu_i}(x), \xi) = 0$$

for all vectors $\xi \in T_xM$ if and only if $\alpha_i = 0$, $1 \leq i \leq m-1$, hence by the nondegeneracy of ω, $\sum_{i=1}^{m-1} \alpha_i X_{\mu_i}(x) = 0$ if and only if $\alpha_i = 0$, $1 \leq i \leq m-1$. Our next observation is that the vector fields X_{μ_i} must all lie tangent to C. Since μ is a moment map, and since $\mathbb{T}^m$ is abelian, we know:

$$\mu(\psi_g(p)) = \mu(p)$$

for all points $p \in M$, and therefore, $\mu_i(\psi_g(p)) = \mu_i(p)$. Since $\frac{d}{dt}\psi_{\exp(t\theta)} = X_{H_\theta} \circ \psi_{\exp(t\theta)}$, we have:

$$0 = \frac{d}{dt}\bigg|_{t=0} \mu_i(\psi_{\exp(t\theta)}) = d\mu_i(X_{H_\theta}).$$

Thus, we have:

$$\begin{aligned} 0 &= d\mu_i(X_{H_\theta}) = i_{X_{\mu_i}}\omega(X_{H_\theta}) \\ &= \omega(X_{\mu_i}), X_{H_\theta}) = -\omega(X_{H_\theta}, X_{\mu_i}) \\ &= i_{X_{H_\theta}}\omega(X_{\mu_i}) \\ &= -dH_\theta(X_{\mu_i}). \end{aligned}$$

Therefore, H_θ is constant on the level curves of μ_i and hence, the level curves of μ_i must preserve the critical manifold C. Therefore, we conclude that the Hamiltonian vector fields X_{μ_i} are tangent to C and thus:

$$X_{\mu_1}(x), \ldots, X_{\mu_{m-1}}(x) \in T_xC.$$

By Lemma 7.4 we know that the critical manifolds of H_θ are symplectic submanifolds, and therefore T_xC is a symplectic vector space. This means that ω_x is nondegenerate on T_xC. Thus, if we have $\alpha_1, \ldots, \alpha_{m-1} \in \mathbb{R}$ with not all zero, then there exists a vector $\xi \in T_xC$ such that:

$$0 \neq \omega_x(\sum_{i=1}^{m-1} \alpha_i X_{\mu_i}(x), \xi) = \sum_{i=1}^{m-1} \alpha_i i_{X_{\mu_i}(x)} \omega_x(\xi) = \sum_{i=1}^{m-1} \alpha_i d\mu_i(x)(\xi).$$

Hence we conclude that $d\mu_1(x), \ldots, d\mu_{m-1}(x)$ are linearly independent on T_xC.

The fact that $T_xM = T_xQ + T_xC$ means that $T_xC^\perp \subset T_xQ$. From this we notice that $\nabla^2 H_\theta(x)$ is nondegenerate on $T_xQ \cap T_xC^\perp$; i.e., T_xQ splits as $T_xQ = (T_xC \cap T_xQ) \oplus E_x^+ \oplus E_x^-$. In particular, this means that the restriction $H_\theta|_Q : Q \to \mathbb{R}$ is Morse-Bott with critical manifold $C \cap Q$. Furthermore, since the index of H_θ on M is $n^-(C) = \dim W^u(C) - \dim C = \dim E^-$ and the coindex is $n^+(C) = \dim W^s(C) - \dim C = \dim E^+$ are both even, we see that the index of $H_\theta|_Q$ on Q is $n^-(C \cap Q) = \dim E^-$ and the coindex is $n^+(C \cap Q) = \dim E^+$, and so they are also even. Lastly, we note that the difference between $\mu_m|_Q$ and $H_\theta|_Q$ is simply the constant $\sum_{i=1}^{m-1} \theta_i \eta_i$, and therefore these conclusions are true for $\mu_m|_Q$ as well, i.e., $\mu_m|_Q : Q \to \mathbb{R}$ is Mores-Bott and has critical manifolds of even index and coindex.

Therefore, by Lemma 6.5, we know the $\mu_m^{-1}(\eta_m) \subset Q$ is connected for every $\eta_m \in \mathbb{R}$. In other words, $\mu^{-1}(\eta) = Q \cap \mu_m^{-1}(\eta_m)$ is connected whenever $(\eta_1, \ldots, \eta_{m-1}) \in \mathbb{R}^{m-1}$ is a regular value of $\lambda = (\mu_1, \ldots, \mu_{m-1})$. As we argued before, the set of regular values of λ is dense in $\mu(M)$ and therefore, by a continuity argument we know that $\mu^{-1}(\eta)$ is connected for every regular value $\eta \in \mathbb{R}^m$.

(2) By induction over the dimension m of the torus, the image $\mu(M) \subset \mathbb{R}^m$ is convex

The base case $m = 1$ follows from the fact that that $\mu(M) \subset \mathbb{R}^m = \mathbb{R}$ is connected, and therefore, convex. Suppose by our inductive hypothesis that the assertion is true for any Hamiltonian torus action $\mathbb{T}^{m-1} \to \mathrm{Symp}(M, \omega)$. Consider any Hamiltonian torus action $\mathbb{T}^m \to \mathrm{Symp}(M, \omega)$ with moment map $\mu : M \to \mathbb{R}^m$. If μ is reducible, then by Lemma 7.7 we have $\mu = A^T \circ \mu' : M \xrightarrow{\mu'} \mathbb{R}^{m-1} \xrightarrow{A^T} \mathbb{R}^m$. By our inductive hypothesis, $\mu'(M)$ must be convex, and thus we know that $A^T(\mu'(M)) = \mu(M)$ must also be convex. Therefore, let us assume that μ is irreducible.

If we choose an injective integer matrix $A \in \mathbb{Z}^{m \times (m-1)}$, then we obtain a torus action given by:

$$\mathbb{T}^{m-1} \to \mathrm{Symp}(M, \omega) : \theta \mapsto \psi_{A\theta}$$

with moment map $\mu_A = A^T \mu : M \to \mathbb{R}^{m-1}$. The fact that μ is irreducible implies that μ_A must also be irreducible, and therefore the regular values of μ_A are dense in $\mu_A(M)$. We also know by the previous part that for any regular value $\eta \in \mathbb{R}^m$, we have $\mu_A^{-1}(\eta)$ connected. Fix a point $x_0 \in \mu_A^{-1}(\eta)$. It is easy to see that we can write the set $\mu_A^{-1}(\eta)$ as:

$$\mu_A^{-1}(\eta) = \{x \in M | \mu(x) - \mu(x_0) \in \ker A^T\}.$$

Since A is injective, the transpose $A^T : \mathbb{R}^m \to \mathbb{R}^{m-1}$ is surjective, and so we see that $\dim \ker A^T = 1$. Therefore, given $x_0, x_1 \in \mu_A^{-1}(\eta)$, we have a path $\gamma : [0,1] \to \mu_A^{-1}(\eta)$. Then the image $\mu(\gamma[0,1]) \subset \mathbb{R}^m$ must connected, and must be in the 1-dimensional $\ker A^T$, and therefore must be convex. Thus, we see that:

$$(1-t)\mu(x_0) + t\mu(x_1) \in \mu(M), \; 0 \leq t \leq 1.$$

We now modify this method slightly to show that given *any* two points $y_0, y_1 \in M$ then every convex combination $(1-t)\mu(y_0) + t\mu(y_1)$, $0 \le t \le 1$ is in $\mu(M)$. Start with any two points $y_0, y_1 \in M$. We can approximate these points arbitrarily closely by points $y_0', y_1' \in M$ such that $\mu(y_1') - \mu(y_0') \in \ker A^T$ for an injective integer matrix $A \in \mathbb{Z}^{m\times(m-1)}$. We can furthermore approximate these points by points $y_0'', y_1'' \in M$ such that $\eta = A^T\mu(y_0'') = A^T\mu(y_1'') \in \mathbb{R}^{m-1}$ is a regular value of $A^T\mu = \mu_A$, which shows that every convex combination $(1-t)\mu(y_0'') + t\mu(y_1'')$, $0 \le t \le 1$ is in $\mu(M)$. Thus, given any points $y_0, y_1 \in M$ we can approximate these points arbitrarily closely by points with images such that every convex combination of these images is contained in $\mu(M)$. It follows by continuity therefore that every convex combination $(1-t)\mu(y_0) + t\mu(y_1)$, $0 \le t \le 1$ is in $\mu(M)$.

This proves that the image $\mu(M)$ is convex.

(3) The points of M fixed by every symplectomorphism in Im$(\mathbb{T}^m) \subset$ Symp(M,ω) decomposes into a finite union of symplectic submanifolds $C_1, \dots, C_N$, and the moment map is constant on these symplectic submanifolds

Lemma 7.3 shows that $\mathrm{Fix}(\mathbb{T}^m) = \bigcap_{\theta\in\mathbb{T}^m} \mathrm{Fix}(\psi_\theta)$ is a symplectic submanifold of M, hence decomposes into a finite union of symplectic submanifolds $C_1, \dots, C_N$. Each component of the moment map is equal to a Hamiltonian function:

$$\mu_i = H_\theta \text{ where } \theta = (0, \dots, 0, \underbrace{1}_{i^{th}}, 0, \dots, 0).$$

And thus, by Lemma 7.4, we have $C_i \subset \mathrm{Crit}(H_\theta)$ for every $1 \le i \le N$ and every $\theta \in \mathbb{R}^m$, and therefore we conclude that the components of the moment map are critical and therefore constant on the symplectic submanifolds C_i, $1 \le i \le N$. Therefore:

$$\mu(C_i) = \eta_i \in \mathbb{R}^m \text{ for every } 1 \le i \le N.$$

(4) The image of μ is the convex hull of the points $\eta_j = \mu(C_j) \in \mathbb{R}^m$, $1 \le j \le N$

Since we have already proved that $\mu(M)$ is convex, it is certainly true that the convex hull K of $\eta_1, \dots, \eta_N$ is contained in $\mu(M)$, i.e.:

$$K = K(\eta_1, \dots, \eta_N) \subset \mu(M).$$

To see that they must be equal, let $\alpha \in \mathbb{R}^m - K$. We can choose a vector $\theta \in \mathbb{R}^m$ which has rationally independent components such that:

$$\langle \eta_i, \theta\rangle < \langle \alpha, \theta\rangle, \; 1 \le i \le N$$

(it is easy to see that such a vector θ exists by geometrical considerations). Since θ has rationally independent components, we see that $T_\theta = cl(\{t\theta + k | t \in \mathbb{R}, k \in \mathbb{Z}^m\}/\mathbb{Z}^m) = \mathbb{T}^m$. Therefore, by Lemma 7.4 we see that the critical set of the Hamiltonian function $H_\theta = \langle \mu, \theta\rangle : M \to \mathbb{R}$ equals $C_1 \cup \ldots \cup C_N$, i.e.:

$$\mathrm{Crit}(H_\theta) = \bigcap_{\tau\in\mathbb{T}^m} \mathrm{Fix}(\psi_\tau) = \mathrm{Fix}(\mathbb{T}^m) = C_1 \cup \ldots \cup C_N.$$

It necessarily follows that H_θ must achieve its maximum on one of these sets C_i, in other words, for all $p \in M$:

$$\langle \mu(p), \theta\rangle \le \sup_{\substack{x\in C_i \\ 1\le i\le N}} \langle \mu(x), \theta\rangle = \sup_{1\le i\le N} \langle \eta_i, \theta\rangle.$$

And therefore, since $\langle \eta_i, \theta \rangle < \langle \alpha, \theta \rangle$ for all $1 \leq i \leq N$, we conclude that, for all $p \in M$:

$$\langle \mu(p), \theta \rangle < \langle \alpha, \theta \rangle.$$

Thus, $\alpha \notin \mu(M)$. And therefore, $K(\eta_1, \ldots, \eta_N) = \mu(M)$ as was claimed. □

9. Examples

Example 9.1. We may the sphere S^2 as a symplectic manifold with symplectic form $\omega = d\theta \wedge dh$, where (θ, h) are cylindrical polar coordinates on S^2. We have a torus action of $\mathbb{T}^1 = S^1$ given by rotations about the vertical axis, i.e., $S^1 \to \mathrm{Symp}(S^2, \omega) : t \mapsto \psi_t$ where $\psi_t(\theta, h) = (\theta + 2\pi t, h)$. It is easy to check that this torus action is indeed, a Hamiltonian action $\mu : S^2 \to \mathbb{R}$ given by $\mu(\theta, h) = h$:

- For any $\tau \in \mathbb{R}$, considered as the Lie algebra of S^1, we have:

  $$H_\tau = \langle \mu, \tau \rangle = \tau \cdot \mu : S^2 \to \mathbb{R} \text{ given by } H_\tau(\theta, h) = \tau \cdot h.$$

  Since $dH_\tau = \tau dh = i_{X_{H_\tau}}(d\theta \wedge dh)$, we see that:

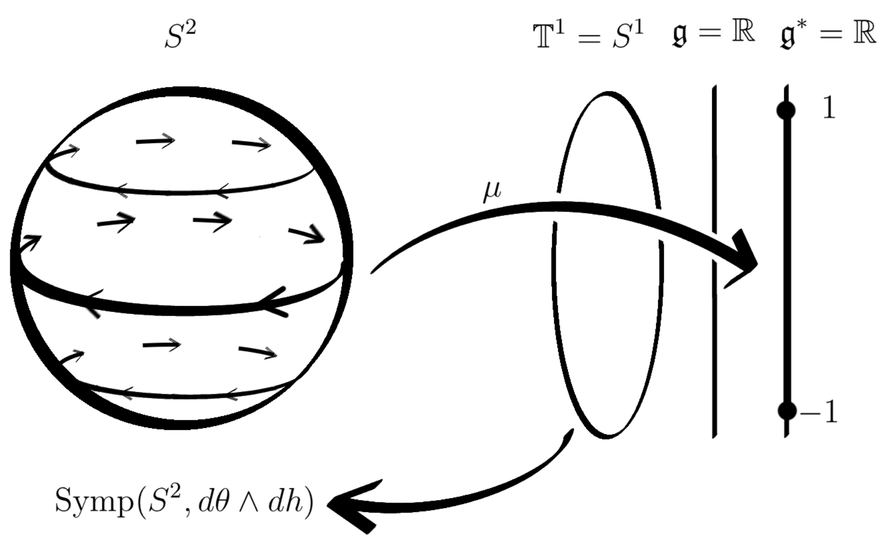
  $$X_{H_\tau} = \tau \frac{\partial}{\partial \theta}$$

  and, indeed:

  $$X_\tau = \left. \frac{d}{dt} \right|_{t=0} \psi_{\exp t\tau} = \tau \frac{\partial}{\partial \theta}.$$

  So $X_{H_\tau} = X_\tau$ as required.
- $\mu(\psi_t(\theta, h)) = \mu(\theta + 2\pi t, h) = h = \mu(\theta, h)$ as required.

We have $\mu(S^2) = [-1, 1] \subset \mathbb{R}$, so the image of the moment map is convex, as required.



Example 9.2. We may the complex projective space $\mathbb{CP}^n$ as a symplectic manifold with symplectic form given by the Fubini-Study form ω_{FS}. We have a torus action of $\mathbb{T}^n$:

$$\mathbb{T}^n \to \text{Symp}(\mathbb{CP}^n, \omega_{FS}) : (\theta_1, \dots, \theta_n) \mapsto \psi_{(\theta_1,\dots,\theta_n)}$$

where

$$\psi_{(\theta_1,\dots,\theta_n)}[z_0 : z_1 : \dots : z_n] = [z_0 : e^{-2\pi i\theta_1} z_1 : \dots : e^{-2\pi i\theta_n} z_n].$$

We claim that this torus action is, indeed, a Hamiltonian action with moment map $\mu : \mathbb{CP}^n \to \mathbb{R}^n$ given by:

$$\mu[z_0 : z_1 : \dots : z_n] = \pi \left(\frac{|z_1|^2}{\|z\|^2}, \dots, \frac{|z_n|^2}{\|z\|^2} \right) \in \mathbb{R}^n$$

where $\|z\|^2 = \sum_{i=0}^n |z_i|^2$.

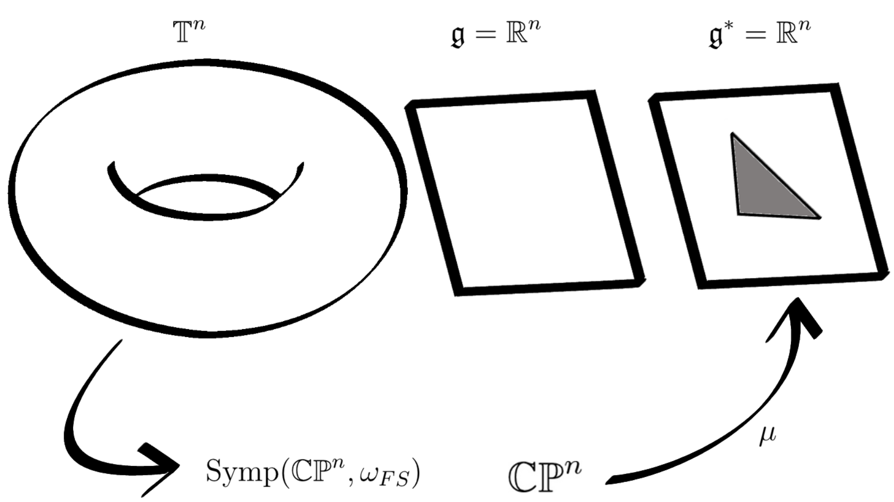
Thus, the image of μ is a simplex:

$$\Delta = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n | 0 \le \sum_{i=1}^n x_i \le \pi \right\}$$

and μ has $n+1$ isolated fixed points in $\mathbb{CP}^n$ at:

$$c_i = [0 : \dots : 0 : \underbrace{1}_{i^{th}} : 0 : \dots : 0],\ 0 \le i \le n$$

which get mapped by μ to the vertices of Δ. Thus, the convex hull $K(\mu(c_1), \dots, \mu(c_n)) = \mu(\mathbb{CP}^n) = \Delta$ as required.



Acknowledgments. The debt I owe to my fellow peers, my professors, and to other assorted members of the mathematics department of the University of Chicago is immeasurable. In particular, I would like to thank my mentor, Alex Wright, for helping me work through the proof and fully understand it. I would again like to thank my girlfriend, Anna, for her encouragement and for her occasional feigned interest in a topic so far removed from her own area of expertise. Lastly, I would like to thank Professor Leonid Polterovich for introducing me and inspiring me to investigate the mathematical universe of symplectic topology.

References

McDuff, Dusa Waddington, and Dietmar Salamon. Introduction to Symplectic Topology. Oxford: University, 2005. Print.

Silva, Ana Cannas Da. Lectures on Symplectic Geometry. Berlin: Springer, 2008. Print.