Black Holes and Thermodynamics

II: Black Hole Entropy and Noether Charge

Robert M. Wald

Based mainly on V. Iyer and RMW, Phys. Rev. **D50**, 846 (1994)

Lagrangians and Hamiltonians in Classical Field Theory

Lagrangian and Hamiltonian formulations of field theories play a central role in their quantization. However, it had been my view that their role in classical field theory was not much more than that of a mnemonic device to remember the field equations. When I wrote my GR text, the discussion of the Lagrangian (Einstein-Hilbert) and Hamiltonian (ADM) formulations of general relativity was relegated to an appendix. My views have changed dramatically in the past 20 years: The existence of a Lagrangian or Hamiltonian provides important auxiliary structure to a classical field theory, which endows the theory with key properties.

Lagrangians and Hamiltonians in Particle Mechanics

Consider particle paths $q(t)$. If L is a function of $(q, \dot{q})$, then we have the identity

$$\delta L = [\frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}}]\delta q + \frac{d}{dt}[\frac{\partial L}{\partial \dot{q}}\delta q]$$

holding at each time t. L is a Lagrangian for the system if the equations of motion are

$$0 = E \equiv \frac{\partial L}{\partial q} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}}$$

The "boundary term"

$$\Theta(q, \dot{q}) \equiv \frac{\partial L}{\partial \dot{q}}\delta q = p\delta q$$

(with $p \equiv \partial L/\partial \dot{q}$) is usually discarded. However, by taking a second, antisymmetrized variation of Θ and evaluating at time t_0, we obtain the quantity

$$\begin{aligned} \Omega(q, \delta_1 q, \delta_2 q) &= [\delta_1 \Theta(q, \delta_2 q) - \delta_2 \Theta(q, \delta_1 q)]|_{t_0} \\ &= [\delta_1 p \delta_2 q - \delta_2 p \delta_1 q]|_{t_0} \end{aligned}$$

Then Ω is independent of t_0 provided that the varied paths $\delta_1 q(t)$ and $\delta_2 q(t)$ satisfy the linearized equations of motion about $q(t)$. Ω is highly degenerate on the infinite dimensional space of all paths $\mathcal{F}$, but if we factor $\mathcal{F}$ by the degeneracy subspaces of Ω, we obtain a finite dimensional *phase space* Γ on which Ω is non-degenerate. A *Hamiltonian*, H, is a function on Γ whose pullback to

$\mathcal{F}$ satisfies

$$\delta H = \Omega(q; \delta q, \dot{q})$$

for all δq provided that $q(t)$ satisfies the equations of motion. This is equivalent to saying that the equations of motion are

$$\dot{q} = \frac{\partial H}{\partial p} \qquad \dot{p} = -\frac{\partial H}{\partial q}$$

Lagrangians and Hamiltonians in Classical Field Theory

Let ϕ denote the collection of dynamical fields. The analog of $\mathcal{F}$ is the space of field configurations on spacetime. For an n-dimensional spacetime, a Lagrangian $\mathbf{L}$ is most naturally viewed as an n-form on spacetime that is a function of ϕ and finitely many of its derivatives. Variation of $\mathbf{L}$ yields

$$\delta \mathbf{L} = \mathbf{E}\delta\phi + d\mathbf{\Theta}$$

where $\mathbf{\Theta}$ is an $(n-1)$-form on spacetime, locally constructed from ϕ and $\delta\phi$. The equations of motion are then $\mathbf{E} = 0$. The symplectic current $\boldsymbol{\omega}$ is defined by

$$\boldsymbol{\omega}(\phi, \delta_1\phi, \delta_2\phi) = \delta_1\boldsymbol{\Theta}(\phi, \delta_2\phi) - \delta_2\boldsymbol{\Theta}(\phi, \delta_1\phi)$$

and Ω is then defined by

$$\Omega(\phi, \delta_1\phi, \delta_2\phi) = \int_{\mathcal{C}} \boldsymbol{\omega}(\phi, \delta_1\phi, \delta_2\phi)$$

where $\mathcal{C}$ is a Cauchy surface. Phase space is constructed by factoring field configuration space by the degeneracy subspaces of Ω, and a Hamiltonian, H_ξ, conjugate to a vector field ξ^a on spacetime is a function on phase space whose pullback to field configuration space satisfies

$$\delta H_\xi = \Omega(\phi; \delta\phi, \mathcal{L}_\xi\phi)$$

Diffeomorphism Covariant Theories

A diffeomorphism covariant theory is one whose Lagrangian is constructed entirely from dynamical fields, i.e., there is no "background structure" in the theory apart from the manifold structure of spacetime. For a diffeomorphism covariant theory for which dynamical fields, ϕ, are a metric g_{ab} and tensor fields ψ, the Lagrangian takes the form

$$\mathbf{L} = \mathbf{L}\left(g_{ab}, R_{bcde}, ..., \nabla_{(a_1}...\nabla_{a_m)}R_{bcde}; \psi, ..., \nabla_{(a_1}...\nabla_{a_l)}\psi\right)$$

Noether Current and Noether Charge

For a diffeomorphism covariant theory, every vector field ξ^a on spacetime generates a local symmetry. We associate to each ξ^a and each field configuration, ϕ (*not* required, at this stage, to be a solution of the equations of motion), a Noether current $(n-1)$-form, $\mathbf{J}_\xi$, defined by

$$\mathbf{J}_\xi = \mathbf{\Theta}(\phi, \mathcal{L}_\xi \phi) - \xi \cdot \mathbf{L}$$

A simple calculation yields

$$d\mathbf{J}_\xi = -\mathbf{E}\mathcal{L}_\xi \phi$$

which shows $\mathbf{J}_\xi$ is closed (for all ξ^a) when the equations of motion are satisfied. It can then be shown that for all

ξ^a and all ϕ (not required to be a solution to the equations of motion), we can write $\mathbf{J}_\xi$ as

$$\mathbf{J}_\xi = \xi^a \mathbf{C}_a + d\mathbf{Q}_\xi$$

where $\mathbf{C}_a = 0$ are the constraint equations of the theory and $\mathbf{Q}_\xi$ is an $(n-2)$-form locally constructed out of the dynamical fields ϕ, the vector field ξ^a, and finitely many of their derivatives. It can be shown that $\mathbf{Q}_\xi$ can always be expressed in the form

$$\mathbf{Q}_\xi = \mathbf{W}_c(\phi)\xi^c + \mathbf{X}^{cd}(\phi)\nabla_{[c}\xi_{d]} + \mathbf{Y}(\phi, \mathcal{L}_\xi\phi) + d\mathbf{Z}(\phi, \xi)$$

Furthermore, there is some "gauge freedom" in the choice of $\mathbf{Q}_\xi$ arising from (i) the freedom to add an exact form to the Lagrangian, (ii) the freedom to add an exact form to $\mathbf{\Theta}$, and (iii) the freedom to add an exact form to $\mathbf{Q}_\xi$. Using this freedom, we may choose $\mathbf{Q}_\xi$ to take the form

$$\mathbf{Q}_\xi = \mathbf{W}_c(\phi)\xi^c + \mathbf{X}^{cd}(\phi)\nabla_{[c}\xi_{d]}$$

where

$$(\mathbf{X}^{cd})_{c_3 \dots c_n} = -E_R^{abcd}\boldsymbol{\epsilon}_{abc_3 \dots c_n}$$

where $E_R^{abcd} = 0$ are the equations of motion that would result from pretending that R_{abcd} were an independent dynamical field in the Lagrangian $\mathbf{L}$.

Hamiltonians

Let ϕ be any solution of the equations of motion, and let $\delta\phi$ be any variation of the dynamical fields (not necessarily satisfying the linearized equations of motion) about ϕ. Let ξ^a be an arbitrary, fixed vector field. We then have

$$\begin{aligned}
\delta \mathbf{J}_\xi &= \delta \mathbf{\Theta}(\phi, \mathcal{L}_\xi \phi) - \xi \cdot \delta \mathbf{L} \\
&= \delta \mathbf{\Theta}(\phi, \mathcal{L}_\xi \phi) - \xi \cdot d\mathbf{\Theta}(\phi, \delta\phi) \\
&= \delta \mathbf{\Theta}(\phi, \mathcal{L}_\xi \phi) - \mathcal{L}_\xi \mathbf{\Theta}(\phi, \delta\phi) + d(\xi \cdot \mathbf{\Theta}(\phi, \delta\phi))
\end{aligned}$$

On the other hand, we have

$$\delta \mathbf{\Theta}(\phi, \mathcal{L}_\xi \phi) - \mathcal{L}_\xi \mathbf{\Theta}(\phi, \delta\phi) = \boldsymbol{\omega}(\phi, \delta\phi, \mathcal{L}_\xi \phi)$$

We therefore obtain

$$\boldsymbol{\omega}(\phi, \delta\phi, \mathcal{L}_\xi \phi) = \delta \mathbf{J}_\xi - d(\xi \cdot \mathbf{\Theta})$$

Replacing $\mathbf{J}_\xi$ by $\xi^a \mathbf{C}_a + d\mathbf{Q}_\xi$ and integrating over a Cauchy surface $\mathcal{C}$, we obtain

$$\begin{aligned}
\Omega(\phi, \delta\phi, \mathcal{L}_\xi \phi) &= \int_{\mathcal{C}} [\xi^a \delta \mathbf{C}_a + \delta d \mathbf{Q}_\xi - d(\xi \cdot \mathbf{\Theta})] \\
&= \int_{\mathcal{C}} \xi^a \delta \mathbf{C}_a + \int_{\partial \mathcal{C}} [\delta Q_\xi - \xi \cdot \mathbf{\Theta})]
\end{aligned}$$

The $(n-1)$-form $\mathbf{\Theta}$ cannot be written as the variation of a quantity locally and covariantly constructed out of the dynamical fields (unless $\boldsymbol{\omega} = 0$). However, it is possible that for the class of spacetimes being considered, we can find a (not necessarily diffeomorphism covariant) $(n-1)$-form, $\mathbf{B}$, such that

$$\delta \int_{\partial \mathcal{C}} \xi \cdot \mathbf{B} = \int_{\partial \mathcal{C}} \xi \cdot \boldsymbol{\Theta}$$

A Hamiltonian for the dynamics generated by ξ^a exist on this class of spacetimes if and only if such a $\mathbf{B}$ exists. This Hamiltonian is then given by

$$H_\xi = \int_{\mathcal{C}} \xi^a \mathbf{C}_a + \int_{\partial \mathcal{C}} [\mathbf{Q}_\xi - \xi \cdot \mathbf{B}]$$

Note that "on shell", i.e., when the field equations are satisfied, we have $\mathbf{C}_a = 0$ so the Hamiltonian is purely a "surface term".

Energy and Angular Momentum

If a Hamiltonian conjugate to a time translation $\xi^a = t^a$ exists, we define the *energy*, $\mathcal{E}$ of a solution $\phi = (g_{ab}, \psi)$ by

$$\mathcal{E} \equiv H_t = \int_{\partial\mathcal{C}} (\mathbf{Q}_t - t \cdot \mathbf{B})$$

Similarly, if a Hamiltonian, H_φ, conjugate to a rotation $\xi^a = \varphi^a$ exists, we define the *angular momentum*, $\mathcal{J}$ of a solution by

$$\mathcal{J} \equiv -H_\varphi = -\int_{\partial\mathcal{C}} [\mathbf{Q}_\varphi - \varphi \cdot \mathbf{B}]$$

If φ^a is tangent to $\mathcal{C}$, the last term vanishes, and we obtain simply

$$\mathcal{J} = -\int_{\partial\mathcal{C}} \mathbf{Q}_{\varphi}$$

Energy and Angular Momentum in General Relativity:

ADM vs Komar

In general relativity in 4 dimensions, the Einstein-Hilbert Lagrangian is

$$\mathbf{L}_{abcd} = \frac{1}{16\pi} \boldsymbol{\epsilon}_{abcd} R$$

This yields the symplectic potential 3-form

$$\boldsymbol{\Theta}_{abc} = \boldsymbol{\epsilon}_{dabc} \frac{1}{16\pi} g^{de} g^{fh} \left(\nabla_f \delta g_{eh} - \nabla_e \delta g_{fh} \right).$$

The corresponding Noether current and Noether charge are

$$(\mathbf{J}_\xi)_{abc} = \frac{1}{8\pi} \boldsymbol{\epsilon}_{dabc} \nabla_e \left(\nabla^{[e} \xi^{d]} \right),$$

and

$$(\mathbf{Q}_\xi)_{ab} = -\frac{1}{16\pi}\boldsymbol{\epsilon}_{abcd}\nabla^c\xi^d.$$

For asymptotically flat spacetimes, the formula for angular momentum conjugate to an asymptotic rotation φ^a is

$$\mathcal{J} = \frac{1}{16\pi}\int_\infty \boldsymbol{\epsilon}_{abcd}\nabla^c\varphi^d$$

This agrees with the ADM expression, and when φ^a is a Killing vector field, it agrees with the Komar formula. For an asymptotic time translation t^a, a Hamiltonian, H_t, exists with

$$t^a\mathbf{B}_{abc} = -\frac{1}{16\pi}\tilde{\boldsymbol{\epsilon}}_{bc}\left((\partial_r g_{tt} - \partial_t g_{rt}) + r^k h^{ij}(\partial_i h_{kj} - \partial_k h_{ij})\right)$$

The corresponding Hamiltonian

$$H_t = \int_{\mathcal{C}} t^a \mathbf{C}_a + \frac{1}{16\pi} \int_\infty dS r^k h^{ij} (\partial_i h_{kj} - \partial_k h_{ij})$$

is precisely the ADM Hamiltonian, and the surface term is the ADM mass,

$$M_{\mathrm{ADM}} = \frac{1}{16\pi} \int_\infty dS r^k h^{ij} (\partial_i h_{kj} - \partial_k h_{ij})$$

By contrast, if t^a is a Killing field, the Komar expression

$$M_{\mathrm{Komar}} = -\frac{1}{8\pi} \int_\infty \boldsymbol{\epsilon}_{abcd} \nabla^c t^d$$

happens to give the correct (ADM) answer, but this is merely a fluke.

The First Law of Black Hole Mechanics

Return to a general, diffeomorphism covariant theory, and recall that for any solution ϕ, any $\delta\phi$ (not necessarily a solution of the linearized equations) and any ξ^a, we have

$$\Omega(\phi, \delta\phi, \mathcal{L}_\xi \phi) = \int_{\mathcal{C}} \xi^a \delta \mathbf{C}_a + \int_{\partial\mathcal{C}} [\delta Q_\xi - \xi \cdot \mathbf{\Theta})]$$

Now suppose that ϕ is a stationary black hole with a Killing horizon with bifurcation surface Σ. Let ξ^a denote the horizon Killing field, so that $\xi^a|_\Sigma = 0$ and

$$\xi^a = t^a + \Omega_H \varphi^a$$

Then $\mathcal{L}_\xi \phi = 0$. Let $\delta\phi$ satisfy the linearized equations, so $\delta \mathbf{C}_a = 0$. Let $\mathcal{C}$ be a hypersurface extending from Σ to infinity.

$$0 = \int_\infty [\delta Q_\xi - \xi \cdot \mathbf{\Theta})] - \int_\Sigma \delta Q_\xi$$

Thus, we obtain

$$\delta \int_\Sigma Q_\xi = \delta \mathcal{E} - \Omega_H \delta \mathcal{J}$$

Furthermore, from the formula for Q_ξ and the properties of Killing horizons, one can show that

$$\delta \int_\Sigma Q_\xi = \frac{\kappa}{2\pi} \delta S$$

where S is defined by

$$S = 2\pi \int_\Sigma \mathbf{X}^{cd} \epsilon_{cd}$$

where ϵ_{cd} denotes the binormal to Σ. Thus, we have shown that the first law of black hole mechanics

$$\frac{\kappa}{2\pi}\delta S = \delta\mathcal{E} - \Omega_H \delta\mathcal{J}$$

holds in an arbitrary diffeomorphism covariant theory of gravity, and we have obtained an explicit formula for black hole entropy S.

Black Holes and Thermodynamics

Stationary black hole $\leftrightarrow$ Body in thermal equilibrium

Just as bodies in thermal equilibrium are normally characterized by a small number of "state parameters" (such as E and V) a stationary black hole is uniquely characterized by M, J, Q.

0th Law

Black holes: The surface gravity, κ, is constant over the horizon of a stationary black hole.

Thermodynamics: The temperature, T, is constant over a body in thermal equilibrium.

1st Law

Black holes:

$$\delta M = \frac{1}{8\pi}\kappa\delta A + \Omega_H \delta J + \Phi_H \delta Q$$

Thermodynamics:

$$\delta E = T\delta S - P\delta V$$

2nd Law

Black holes:

$$\delta A \geq 0$$

Thermodynamics:

$$\delta S \geq 0$$

Analogous Quantities

$M \leftrightarrow E \leftarrow$ But M really is E!

$\frac{1}{2\pi}\kappa \leftrightarrow T$

$\frac{1}{4}A \leftrightarrow S$