Dynamic and Thermodynamic Stability of Black Holes and Black Branes

Robert M. Wald

with Stefan Hollands

arXiv:1201.0463

Commun. Math. Phys. **321**, 629 (2013)

(see also K. Prabhu and R.M. Wald, arXiv:1501.02522)

Stability

It is of considerable interest to determine the linear stablity of black holes in (D-dimensional) general relativity. We will consider vacuum general relativity without a cosmological constant here, but our methods apply to general theories of gravity. It is also of interest to determine the linear stability of the corresponding black branes in $(D+p)$-dimensions, i.e., spacetimes with metric of the form

$$d\tilde{s}^2_{D+p} = ds^2_D + \sum_{i=1}^{p} dz_i^2 \,,$$

where ds^2_D is a black hole metric. By *linear stability* we mean that linearized perturbations remain uniformly bounded in time (in some gauge). A weaker criterion is *mode stability*, the non-existence of (non-pure-gauge) solutions with time dependence $\exp(\alpha t)$, with $\alpha > 0$. (Mode stability normally is proven by finding a positive definite conserved norm for perturbations.) However, even in the very simplest cases—such as the Schwarzschild black hole (Regge-Wheeler, Zerilli) and the Schwarzschild black string (Gregory-Laflamme)—it is quite nontrivial to carry out the decoupling of equations and the fixing of gauge needed to even begin the analysis of stability directly from the equations of motion.

Thermodynamic Stability

Consider a (finite) system whose thermal equilibrium states are characterized by energy, E, and other state parameters X_i. Perturbations to the system will satisfy the first law of thermodynamics,

$$\delta E = T \delta S + \sum_i Y_i \delta X_i \, ,$$

where $Y_i = (\partial E / \partial X_i)_S$ (and this relation holds even if the perturbations are not to other thermal equilibrium states). In particular, a thermal equilibrium state is an extremum of entropy, S, at fixed values of the state parameters (E, X_i). The system will be *thermodynamically unstable* if S fails to be a local maximum, i.e., if one can find a variation that keeps (E, X_i) fixed to first and second order and is such that $\delta^2 S > 0$. In view of the first law, this is equivalent to finding a perturbation that keeps (E, X_i) fixed only to first order and satisfies

$$\delta^2 E - T\delta^2 S - \sum_i Y_i \delta^2 X_i < 0$$

where $Y_i = (\partial E/\partial X_i)_S$.

Now consider a homogeneous (and hence infinite) system, whose thermodynamic states are characterized by (E, X_i), where these quantities now denote the amount of energy and other state parameters "per unit volume" (so these quantities are now assumed to be "extensive"). The condition for thermodynamic instability remains the same, but now there is no need to require that (E, X_i) be fixed to first order because energy and other extensive variables can be "borrowed" from one part of the system and given to another. Thus, the system will be thermodynamically unstable if the above equation holds for any first order variation. In particular, it will be thermodynamically unstable if the Hessian matrix

$$\mathbf{H}_S = \begin{pmatrix} \frac{\partial^2 S}{\partial E^2} & \frac{\partial^2 S}{\partial X_i \partial E} \\ \frac{\partial^2 S}{\partial E \partial X_i} & \frac{\partial^2 S}{\partial X_i \partial X_j} \end{pmatrix} .$$

admit a positive eigenvalue. If this happens, then one can increase total entropy by exchanging E and/or X_i

between different parts of the system. For the case of E, this corresponds to having a negative heat capacity. In particular, a homogeneous system with a negative heat capacity must be thermodynamically unstable, but this need not be the case for a finite system.

Stability of Black Holes and Black Branes

Black holes and black branes are thermodynamic systems, with

$$\begin{aligned} E &\leftrightarrow M \\ S &\leftrightarrow \frac{A}{4} \\ X_i &\leftrightarrow J_i, Q_i \end{aligned}$$

Thus, in the vacuum case ($Q_i = 0$), the analog of the criterion for thermodynamic instablity of a black hole (i.e., a finite system) is that there exist a perturbation for which $\delta M = \delta J_i = 0$ for which

$$\delta^2 M - \frac{\kappa}{8\pi}\delta^2 A - \sum_i \Omega_i \delta^2 J_i < 0 \,.$$

We will show that this criterion is necessary and sufficient for dynamical stability of a black hole.

On the other hand, black branes are homogeneous systems, so a sufficient condition for instability of a black brane is that the Hessian matrix

$$\mathbf{H}_A = \begin{pmatrix} \frac{\partial^2 A}{\partial M^2} & \frac{\partial^2 A}{\partial J_i \partial M} \\ \frac{\partial^2 A}{\partial M \partial J_i} & \frac{\partial^2 A}{\partial J_i \partial J_j} \end{pmatrix} \,.$$

admits a positive eigenvalue. It was conjectured by Gubser and Mitra that this condition is sufficient for black brane instability. We will prove the Gubser-Mitra conjecture.

As an application, the Schwarzschild black hole has negative heat capacity namely ($A = 16\pi M^2$, so $\partial^2 A/\partial M^2 > 0$). This does not imply that the Schwarzschild black hole is dynamically unstable (and, indeed, it is well known to be stable). However, this calculation does imply that the Schwarzschild black string is unstable!

Local Penrose Inequality

Suppose one has a family of stationary, axisymmetric black holes parametrized by M and angular momenta $J_1, \ldots, J_N$. Consider a one-parameter family $g_{ab}(\lambda)$ of axisymmetric spacetimes, with $g_{ab}(0)$ being a member of this family with surface gravity $\kappa > 0$. Consider initial data on a hypersurface Σ passing through the bifurcation surface B. By the linearized Raychauduri equation, to first order in λ, the event horizon coincides with the apparent horizon on Σ. They need not coincide to second order in λ, but since B is an extremal surface in the background spacetime, their areas must agree to second order. Let $\mathcal{A}$ denotes the area of the apparent horizon of the perturbed spacetime, $\bar{A}$ denotes the the event horizon area of the stationary black hole with the same mass and angular momentum as the perturbed spacetime. Suppose that to second order, we have

$$\delta^2 \mathcal{A} > \delta^2 \bar{A}$$

Since (i) the area of the event horizon can only increase with time (by cosmic censorship), (ii) the final mass of the black hole cannot be larger than the initial total mass (by positivity of Bondi flux), (iii) its final angular momenta must equal the initial angular momenta (by axisymmetry), and (iv) $\bar{A}(M, J_1, \ldots, J_N$ is an increasing function of M at fixed J_i (by the first law of black hole mechanics with $\kappa > 0$), it follows that there would be a contradiction if the perturbed black hole solution were to settle down to a stationary black hole in the family. This implies that satisfaction of this inequality implies instablity—although it does not imply stability if $\delta^2 \mathcal{A} \leq \delta^2 \bar{A}$ always holds.

Our more fundamental stability criterion implies that satisfaction of $\delta^2 \mathcal{A} \leq \delta^2 \bar{A}$ is necessary and sufficient for black hole stability with respect to axisymmetric perturbations.

Our Results

We consider perturbations γ of a static or stationary-axisymmetric black hole or black brane with bifurcate Killing horizon and consider the *canonical energy* of the perturbation defined by

$$\mathcal{E} = \int_\Sigma \omega(g; \gamma, \pounds_t \gamma)$$

where Σ extends from the bifurcation surface to infinity.

We prove that a sufficient condition for mode stability of a black hole (or black brane) with respect to axisymmetric perturbations is positivity of $\mathcal{E}$ on a Hilbert space, $\mathcal{V}$, of perturbations with vanishing perturbed mass, angular momentum, and linear momentum,

$\delta M = \delta J_i = \delta P_i = 0.$

We further show that if $\mathcal{E}$ is not positive on $\mathcal{V}$, then the black hole is unstable in the sense that there exist perturbations that cannot asymptotically approach a stationary black hole.

We will also show that

$$\mathcal{E} = \delta^2 M - \frac{\kappa}{8\pi}\delta^2 A - \sum_i \Omega_i \delta^2 J_i$$

Thus, dynamical stability with respect to axisymmetric perturbations holds for black holes if and only if they satisfy the analog of criterion for the thermodynamic stability of an inhomogeneous system: The area, A, is a maximum at fixed M and J_i.

Furthermore, if a black hole has a perturbation with $\mathcal{E} < 0$ with $\delta M \neq 0$ and/or $\delta J_i \neq 0$, we prove that there exists a sufficiently long wavelength perturbation of any corresponding black brane for which $\mathcal{E} < 0$ but $\delta M = \delta J_i = 0$. Thus, black branes are unstable if they satisfy the criterion for thermodynamic instability of a homogeneous system. This proves the Gubser-Mitra conjecture.

Thus, for example, the calculation that $\partial^2 A/\partial M^2 = 32\pi > 0$ for Schwarzschild tells one nothing about the stability of Schwarzschild black hole; the "change of mass" perturbation of Schwarzschild—for which $\mathcal{E} < 0$—does not "count" for testing stability because, obviously, $\delta M \neq 0$. However, this computation is sufficient to prove the instability of the Schwarzshild black string to sufficiently long wavelength perturbations.

Finally, we prove that if one can find a perturbation of a black hole for which $\delta^2 \mathcal{A} > \delta^2 \bar{A}$, if and only if one can find a perturbation for which $\delta M = \delta J_i = \delta P_i = 0$ and $\mathcal{E} < 0$. This proves that satisfaction of the local Penrose inequality is equivalent to dynamical stability.

Variational Formulas

Lagrangian for vacuum general relativity:

$$L_{a_1 \ldots a_D} = \frac{1}{16\pi} R \, \epsilon_{a_1 \ldots a_D} \, .$$

First variation:

$$\delta L = E \cdot \delta g + d\theta \, ,$$

with

$$\theta_{a_1 \ldots a_{d-1}} = \frac{1}{16\pi} g^{ac} g^{bd} (\nabla_d \delta g_{bc} - \nabla_c \delta g_{bd}) \epsilon_{c a_1 \ldots a_{d-1}} \, .$$

Symplectic current ($(D-1)$-form):

$$\omega(g; \delta_1 g, \delta_2 g) = \delta_1 \theta(g; \delta_2 g) - \delta_2 \theta(g; \delta_1 g) \, .$$

Symplectic form:

$$\begin{aligned} W_\Sigma(g;\delta_1 g,\delta_2 g) &\equiv \int_\Sigma \omega(g;\delta_1 g,\delta_2 g) \\ &= -\frac{1}{32\pi}\int_\Sigma \left(\delta_1 h_{ab}\delta_2 p^{ab} - \delta_2 h_{ab}\delta_1 p^{ab}\right), \end{aligned}$$

with

$$p^{ab} \equiv h^{1/2}(K^{ab} - h^{ab}K)\,.$$

Noether current:

$$\begin{aligned} \mathcal{J}_X &\equiv \theta(g,\pounds_X g) - X\cdot L \\ &= X\cdot C + dQ_X\,. \end{aligned}$$

Fundamental variational identity:

$$\omega(g; \delta g, \pounds_X g) = X \cdot [E(g) \cdot \delta g] + X \cdot \delta C + d\,[\delta Q_X(g) - X \cdot \theta(g; \delta g)]$$

ADM conserved quantities:

$$\delta H_X = \int_\infty [\delta Q_X(g) - X \cdot \theta(g; \delta g)]$$

For a stationary black hole, choose X to be the horizon Killing field

$$K^a = t^a + \sum \Omega_A \phi_A^a$$

Integration of the fundamental identity yields the first law of black hole mechanics:

$$0 = \delta M - \sum_i \Omega_i \delta J_i - \frac{\kappa}{8\pi} \delta A \, .$$

Horizon Gauge Conditions

Consider stationary black holes with surface gravity $\kappa > 0$, so the event horizon is of "bifurcate type," with bifurcation surface B. Consider an arbitrary perturbation $\gamma = \delta g$. Gauge condition that ensures that the location of the horizon does not change to first order:

$$\delta\vartheta|_B = 0\,.$$

Additional gauge condition that we impose:

$$\delta\epsilon|_B = \frac{\delta A}{A}\epsilon\,.$$

Canonical Energy

Define the *canonical energy* of a perturbation $\gamma = \delta g$ by

$$\mathcal{E} \equiv W_\Sigma \left(g; \gamma, \pounds_t \gamma\right)$$

The second variation of our fundamental identity then yields (for axisymmetric perturbations)

$$\mathcal{E} = \delta^2 M - \sum_i \Omega_i \delta^2 J_i - \frac{\kappa}{8\pi}\delta^2 A\,.$$

More generally, can view the canonical energy as a bilinear form $\mathcal{E}(\gamma_1, \gamma_2) = W_\Sigma(g; \gamma_1, \pounds_t \gamma_2)$ on perturbations. $\mathcal{E}$ can be shown to satisfy the following properties:

- $\mathcal{E}$ is conserved, i.e., it takes the same value if evaluated on another Cauchy surface Σ' extending from infinity to B.
- $\mathcal{E}$ is symmetric, $\mathcal{E}(\gamma_1, \gamma_2) = \mathcal{E}(\gamma_2, \gamma_1)$
- When restricted to perturbations for which $\delta A = 0$ and $\delta P_i = 0$ (where P_i is the ADM linear momentum), $\mathcal{E}$ is gauge invariant.
- When restricted to the subspace, $\mathcal{V}$, of perturbations for which $\delta M = \delta J_i = \delta P_i = 0$ (and hence, by the first law of black hole mechanics $\delta A = 0$), we have $\mathcal{E}(\gamma', \gamma) = 0$ for all $\gamma' \in \mathcal{V}$ if and only if γ is a perturbation towards another stationary and

axisymmetric black hole.

Thus, if we restrict to perturbations in the subspace, $\mathcal{V}'$, of perturbations in $\mathcal{V}$ modulo perturbations towards other stationary black holes, then $\mathcal{E}$ is a non-degenerate quadratic form. Consequently, on $\mathcal{V}'$, either (a) $\mathcal{E}$ is positive definite or (b) there is a $\psi \in \mathcal{V}'$ such that $\mathcal{E}(\psi) < 0$. If (a) holds, we have mode stability.

Flux Formulas

Let δN_{ab} denote the perturbed Bondi news tensor at null infinity, $\mathcal{I}^+$, and let $\delta\sigma_{ab}$ denote the perturbed shear on the horizon, $\mathcal{H}$. If the perturbed black hole were to "settle down" to another stationary black hole at late times, then $\delta N_{ab} \to 0$ and $\delta\sigma_{ab} \to 0$ at late times. We show that—for axisymmetric perturbations—the change in canonical energy would then be given by

$$\Delta\mathcal{E} = -\frac{1}{16\pi}\int_{\mathcal{I}} \delta\tilde{N}_{cd}\delta\tilde{N}^{cd} - \frac{1}{4\pi}\int_{\mathcal{H}} (K^a\nabla_a u)\,\delta\sigma_{cd}\delta\sigma^{cd} \;\leq 0\,.$$

Thus, $\mathcal{E}$ can only decrease. Therefore if one has a perturbation $\psi \in \mathcal{V}'$ such that $\mathcal{E}(\psi) < 0$, then ψ cannot "settle down" to a stationary solution at late times because $\mathcal{E} = 0$ for stationary perturbations with $\delta M = \delta J_i = \delta P_i = 0$. Thus, in case (b) we have instability in the sense that the perturbation cannot asymptotically approach a stationary perturbation.

Instability of Black Branes

Theorem: Suppose a family of black holes parametrized by (M, J_i) is thermodynamically unstable at (M_0, J_{0A}), i.e., there exists a perturbation within the black hole family for which $\mathcal{E} < 0$. Then, for any black brane corresponding to (M_0, J_{0A}) one can find a sufficiently long wavelength perturbation for which $\tilde{\mathcal{E}} < 0$ and $\delta\tilde{M} = \delta\tilde{J}_A = \delta\tilde{P}_i = \delta\tilde{A} = \delta\tilde{T}_i = 0$.

This result is proven by modifying the initial data for the perturbation to another black hole with $\mathcal{E} < 0$ by multiplying it by $\exp(ikz)$ and then re-adjusting it so that the modified data satisfies the constraints. The new data will automatically satisfy

$\delta\tilde{M} = \delta\tilde{J}_A = \delta\tilde{P}_i = \delta\tilde{A} = \delta\tilde{T}_i = 0$ because of the $\exp(ikz)$ factor. For sufficiently small k, it can be shown to satisfy $\tilde{\mathcal{E}} < 0$.

Equivalence to Local Penrose Inequality

Let $\bar{g}_{ab}(M, J_i)$ be a family of stationary, axisymmetric, and asymptotically flat black hole metrics on M. Let $g_{ab}(\lambda)$ be a one-parameter family of axisymmetric metrics such that $g_{ab}(0) = \bar{g}_{ab}(M_0, J_{0A})$. Let $M(\lambda), J_i(\lambda)$ denote the mass and angular momenta of $g_{ab}(\lambda)$ and let $\mathcal{A}(\lambda)$ denote the area of its apparent horizon. Let $\bar{g}_{ab}(\lambda) = \bar{g}_{ab}(M(\lambda), J_i(\lambda))$ denote the one-parameter family of stationary black holes with the same mass and angular momenta as $g_{ab}(\lambda)$.

Theorem: There exists a one-parameter family $g_{ab}(\lambda)$ for which

$$\mathcal{A}(\lambda) > \bar{\mathcal{A}}(\lambda)$$

to second order in λ if and only if there exists a perturbation γ'_{ab} of $\bar{g}_{ab}(M_0, J_{0A})$ with $\delta M = \delta J_i = \delta P_i = 0$ such that $\mathcal{E}(\gamma') < 0$.

Proof: The first law of black hole mechanics implies $\mathcal{A}(\lambda) = \bar{\mathcal{A}}(\lambda)$ to first order in λ, so what counts are the second order variations. Since the families have the same mass and angular momenta, we have

$$\begin{aligned} \frac{\kappa}{8\pi}\left[\frac{d^2 A}{d\lambda^2}(0) - \frac{d^2 \bar{A}}{d\lambda^2}(0)\right] &= \mathcal{E}(\bar{\gamma},\bar{\gamma}) - \mathcal{E}(\gamma,\gamma) \\ &= -\mathcal{E}(\gamma',\gamma') + 2\mathcal{E}(\gamma',\bar{\gamma}) \\ &= -\mathcal{E}(\gamma',\gamma') \end{aligned}$$

where $\gamma' = \bar{\gamma} - \gamma$.

Are We Done with Linear Stability Theory for Black Holes?

Not quite:

- The formula for $\mathcal{E}$ is rather complicated, and the linearized initial data must satisfy the linearized constraints, so its not that easy to determine positivity of $\mathcal{E}$.
- There is a long way to go from positivity of $\mathcal{E}$ and (true) linear stability and instability.
- Only axisymmetric perturbations are treated.

And, of course, only linear stability is being analyzed.

$$
\begin{aligned}
\mathcal{E} \;=\; & \int_\Sigma N\Bigg(h^{\frac{1}{2}}\Bigg\{\frac{1}{2}\; R_{ab}(h)q_c{}^c q^{ab} - 2\; R_{ac}(h)q^{ab}q_b{}^c \\
& -\frac{1}{2}\; q^{ac}D_aD_cq_d{}^d - \frac{1}{2}\; q^{ac}D^bD_bq_{ac} + q^{ac}D^bD_aq_{cb} \\
& -\frac{3}{2}\; D_a(q^{bc}D^aq_{bc}) - \frac{3}{2}\; D_a(q^{ab}D_bq_c{}^c) + \frac{1}{2}\; D_a(q_d{}^dD^aq_c{}^c) \\
& +2\; D_a(q^a{}_cD_bq^{cb}) + D_a(q^b{}_cD_bq^{ac}) - \frac{1}{2}\; D^a(q_c{}^cD^bq_{ab})\Bigg\} \\
& +h^{-\frac{1}{2}}\Bigg\{2\; p_{ab}p^{ab} + \frac{1}{2}\; \pi_{ab}\pi^{ab}(q_a{}^a)^2 - \pi_{ab}p^{ab}q_c{}^c \\
& -3\; \pi^a{}_b\pi^{bc}q_d{}^dq_{ac} - \frac{2}{D-2}\; (p_a{}^a)^2 + \frac{3}{D-2}\; \pi_c{}^cp_b{}^bq_a{}^a \\
& +\frac{3}{D-2}\; \pi_d{}^d\pi^{ab}q_c{}^cq_{ab} + 8\; \pi^c{}_bq_{ac}p^{ab} + \pi_{cd}\pi^{cd}q_{ab}q^{ab}
\end{aligned}
$$

$$+2\ \pi^{ab}\pi^{dc}q_{ac}q_{bd} - \frac{1}{D-2}\ ({\pi_c}^c)^2 q_{ab}q^{ab}$$

$$-\frac{1}{2(D-2)}\ ({\pi_b}^b)^2({q_a}^a)^2 - \frac{4}{D-2}\ {\pi_c}^c p^{ab}q_{ab}$$

$$-\frac{2}{D-2}\ (\pi^{ab}q_{ab})^2 - \frac{4}{D-2}\ \pi_{ab}{p_c}^c q^{ab}\Bigg\}\Bigg)$$

$$-\int_\Sigma N^a\Bigg(-2\ p^{bc}D_a q_{bc} + 4\ p^{cb}D_b q_{ac} + 2\ q_{ac}D_b p^{cb}$$

$$-2\ \pi^{cb}q_{ad}D_b {q_c}^d + \pi^{cb}q_{ad}D^d q_{cb}\Bigg)$$

$$+\kappa\int_B s^{\frac{1}{2}}\left(\delta s_{ab}\delta s^{ab} - \frac{1}{2}\ \delta {s_a}^a \delta {s_b}^b\right)$$

Conclusion

Dynamical stability of a black hole is equivalent to its thermodynamic stability with respect to axisymmetric perturbations.

Thus, the remarkable relationship between the laws of black hole physics and the laws of thermodynamics extends to dynamical stability.