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, Commun. Math. Phys. **340**, 253 (2015); arXiv:1501.02522

and S. Green, S. Hollands, A. Ishibashi, and R.M. Wald, arXiv:1512.02644)

Stability of Black Holes and Black Branes

Black holes in general relativity in 4-dimensional spacetimes are believed to be the end products of gravitational collapse. Kerr black holes are the unique stationary black hole solutions in 4-dimensions. It is considerable physical and astrophysical importance to determine if Kerr black holes are stable.

Black holes in higher dimensional spacetimes are interesting playgrounds for various ideas in general relativity and in string theory. A wide variety of black hole solutions occur in higher dimensions, and it is of interest to determine their stability. It is also of interest to consider the stability of "black brane" solutions, which in vacuum general relativity with vanishing cosmological constant are simply $(D+p)$-dimensional 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.

In this work, we will define a quantity, $\mathcal{E}$, called the *canonical energy*, for a perturbation γ_{ab} of a black hole or black brane and show that positivity of $\mathcal{E}$ is necessary and sufficient for linear stability to axisymmetric perturbations in the following senses: (i) If $\mathcal{E}$ is non-negative for all perturbations, then one has mode stability, i.e., there do not exist exponentially growing perturbations. (ii) If $\mathcal{E}$ can be made negative for a perturbation γ_{ab}, then γ_{ab} cannot approach a stationary perturbation at late times; furthermore, if γ_{ab} is of the form $\pounds_t \gamma'_{ab}$, then γ_{ab} must grow exponentially with time.

These results are much weaker than one would like to prove, and our techniques, by themselves, are probably not capable of establishing much stronger results. Thus, our work is intended as a supplement to techniques presently being applied to Kerr stability, not as an improvement/replacement of them. Aside from its general applicability, the main strength of the work is that we can also show that positivity of $\mathcal{E}$ is equivalent to thermodynamic stability and is also equivalent to the satisfaction of a local Penrose inequality. This also will allow us to give an extremely simple sufficient criterion for the instability of black branes.

We restrict consideration here to asymptotically flat black holes in vacuum general relativity in D-spacetime dimensions, as well as the corresponding black branes. However, our techniques and many of our results generalize straightforwardly to include matter fields and other asymptotic conditions.

Thermodynamic Stability

Consider a *finite* system with a large number of degrees of freedom, with a time translation invariant dynamics. The energy, E, and some finite number of other "state parameters" X_i will be conserved under dynamical evolution but we assume that the remaining degrees of freedom will be "effectively ergodic." The entropy, S, of any state is the logarithm of the number of states that "macroscopically look like" the given state. By definition, a thermal equilibrium state is an extremum of S at fixed (E, X_i). For thermal equilibrium states, the change in entropy, S, under a perturbation depends only on the change in the state parameters, so perturbations of thermal equilibrium states satisfy the first law of thermodynamics,

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

where $Y_i = (\partial E/\partial X_i)_S$. Note that this relation holds even if the perturbations are not to other thermal equilibrium states.

A thermal equilibrium state will be locally *thermodynamically stable* if S is a local maximum at fixed (E, X_i), i.e., if $\delta^2 S < 0$ for all variations that keep (E, X_i) fixed to first and second order. In view of the first law (and assuming $T > 0$), this is equivalent the condition

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

for all variations for which (E, X_i) are kept fixed only to first order.

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 “intensive”). The condition for thermodynamic stability 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, for the system to be thermodynamically unstable, the above equation must hold for any first order variation. In particular, the system 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 stablity of a black hole (i.e., a finite system) is that for all perturbations for which $\delta M = \delta J_i = 0$, we have

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

We will show that this criterion is equivalent to positivity of canonical energy, $\mathcal{E}$, and thus, for axisymmetric perturbations, 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 $J_1, \dots, J_N$. 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 fundamental stability criterion $\mathcal{E} \geq 0$ 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.

Variational Formulas

Lagrangian for vacuum general relativity:

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

First variation:

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

with

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

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:

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

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:

$$\begin{aligned}\omega(g;\delta g,\pounds_X g) &= X\cdot[E(g)\cdot\delta g]+X\cdot\delta C\\ &\quad +d\,[\delta Q_X(g)-X\cdot\theta(g;\delta g)]\end{aligned}$$

Hamilton's equations of motion: H_X is said a Hamiltonian for the dynamics generated by X iff the equations of motion for g are equivalent to the relation

$$\delta H_X=\int_\Sigma \omega(g;\delta g,\pounds_X g)$$

holding for all perturbations, δg of 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

$$K^a = t^a + \sum \Omega_i \phi_i^a$$

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

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

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 such that at (M_0, J_{0A}) 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

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

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.

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

$$
\begin{aligned}
&+2\ \pi^{ab}\pi^{dc}q_{ac}q_{bd} - \frac{1}{D-2}\ (\pi_c{}^c)^2 q_{ab}q^{ab} \\
&-\frac{1}{2(D-2)}\ (\pi_b{}^b)^2 (q_a{}^a)^2 - \frac{4}{D-2}\ \pi_c{}^c p^{ab} q_{ab} \\
&-\frac{2}{D-2}\ (\pi^{ab}q_{ab})^2 - \frac{4}{D-2}\ \pi_{ab} p_c{}^c q^{ab} \Bigg\}\Bigg) \\
&-\int_\Sigma N^a \Bigg(-2\ p^{bc} D_a q_{bc} + 4\ p^{cb} D_b q_{ac} + 2\ q_{ac} D_b p^{cb} \\
&-2\ \pi^{cb} q_{ad} D_b q_c{}^d + \pi^{cb} q_{ad} D^d q_{cb} \Bigg) \\
&+\kappa \int_B s^{\frac{1}{2}} \Bigg(\delta s_{ab} \delta s^{ab} - \frac{1}{2}\ \delta s_a{}^a \delta s_b{}^b \Bigg)
\end{aligned}
$$

Further Developments

One can naturally break-up the canonical energy into a *kinetic energy* (arising from the part of the perturbation that is odd under "$(t-\phi)$-reflection") and a *potential energy* (arising from the part of the perturbation that is even under "$(t-\phi)$-reflection"). Prabhu and I have proven that the kinetic energy is always positive (for any perturbation of any black hole or black brane). We were then able to prove that if the potential energy is negative for a perturbation of the form $\pounds_t \gamma'_{ab}$, then this perturbation must grow exponentially in time.

One can straightforwardly generalize our results to black holes with a negative cosmological constant in asymptotically AdS spacetimes. In this case, there is only one boundary through which there can be a canonical energy flux, so there is no need to restrict to axisymmetric perturbations. Green, Hollands, Ishibashi and I have proven that, in this context, all black holes that possess an ergoregion (i.e., a region where the horizon KVF becomes spacelike) are unstable.

Additional Further Development Recent Development: Comparison with DHR; Axisymmetric Stability of Kerr

Recently, Dafermos, Holzegel, and Rodnianski (DHR) have given a proof of stability and decay in Schwarzschild of gravitational perturbations. A key step in their argument is to construct a conserved, positive definite energy for metric perturbations γ. This energy is obtained by (i) constructing a (complex) Teukolsky (Weyl curvature) variable ψ, (ii) obtaining a (complex) Regge-Wheeler variable by taking 2 derivatives of ψ, and (iii) obtaining a conserved energy for this Regge-Wheeler variable. The resulting energy quantity is quadratic in 5th derivatives of the original metric perturbation.

The canonical energy of γ is quadratic in first derivatives of γ, so the DHR energy cannot be the canonical energy of γ. Is there a relationship between the DHR energy and canonical energy? Yes!

Hertz Potentials

Suppose one has a linear equation

$$E(\gamma) = 0$$

such as the linearized Einstein equation. Suppose one can apply a differential operator T to γ such that the resulting quantity $\psi = T\gamma$ is such that $E\gamma = 0 \Rightarrow O\psi = 0$ for some differential operator O. (In the case of interest, ψ is the Teukolsky variable and $O\psi = 0$ is the Teukolsky equation.) In this situation, there is an operator identity of the form

$$SE = OT$$

for some operator S. Taking (formal) adjoints, we obtain

$$E^\dagger S^\dagger = T^\dagger O^\dagger$$

But if $E^\dagger = E$ we thereby obtain the result that if χ satisfies $O^\dagger \chi = 0$, then $\gamma \equiv S^\dagger \chi$ satisfies $E\gamma = 0$. Thus, we can use solutions of $O^\dagger \chi = 0$ as "potentials" to generate solutions to the original equation of interest. In the case of Kerr $O^\dagger$ is just the Teukolsky operator for the opposite spin weight.

DHR Energy

The DHR energy is just the canonical energy of the new metric perturbation obtained by using the Teukolsky variable of the original metric perturbation as a Hertz potential.

Since the Hertz potentials and corresponding canonical energy construction can be generalized to Kerr, it should be possible to generalize the DHR proof of stability to axisymmetric perturbations of Kerr. However, although we thereby have an explicit expression for an energy quantity that should be positive, it does not appear straightforward to show that it is positive!

Main 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.