Referee Report

A comparison of two constructions for dynamical corrections to Wald entropy

Synopsis

The manuscript compares two notions of entropy for a small dynamical perturbation of a stationary black hole in a general diffeomorphism-invariant theory of gravity. The first is the Wall entropy, obtained by organizing the vv field equation in Gaussian null coordinates according to boost weight. Equations (2.5)–(2.12) isolate terms containing one positive-boost-weight factor and express them as two horizon derivatives of an entropy density, together with a spatial entropy current and terms containing at least two positive-boost-weight factors. This rearrangement is off shell and treats zero- and negative-boost-weight factors without an amplitude expansion, although its second-law interpretation is linearized about stationarity.

The second construction is the improved Noether-charge entropy of Hollands, Wald, and Zhang. Equations (2.13)–(2.21) first vary the covariant phase-space identity around a stationary background with Killing generator $\xi = v\partial_v - r\partial_r$. A horizon quantity $\mathcal{B}^r$ is chosen so that its first variation reproduces the presymplectic potential. The resulting entropy S_{dyn} contains explicit affine-parameter dependence. By comparing the two rearrangements after integration over a compact horizon cross-section, the authors recover, through first order in the perturbation, the known relation

$$S_{\text{dyn}} = (1 - v\partial_v)S_{\text{Wall}}.$$

The proposed new result is an inverse construction. Section 3 regards this relation as an ordinary differential equation for S_{Wall} and then seeks a local solution without integrating along the generator. Negative boost-weight background factors are written as $A_{(-m)} = v^m A^{[m]}$, and both entropies are expanded into products of these factors and positive-boost-weight perturbations. Equations (3.5)–(3.10) give a recursive prescription, with $\partial_v S_{\text{dyn}}|_{v=0} = 0$ in equation (3.8) presented as the sole consistency condition. Equations (3.11)–(3.22) relate this condition to a Noether-charge/presymplectic-potential identity used in the earlier entropy current construction.

Section 4 tests the discussion for a Riemann-squared Lagrangian. The authors calculate Q^{rv}, Q^{ri}, and $\mathcal{B}^r$, recover the known spatial entropy current, and obtain equations (4.10) and (4.11), which exhibit the relation between the two entropies. They then run the proposed inverse algorithm explicitly. The appendices state the boost-weight notation, check a presymplectic-potential identity for the three curvature-squared invariants, discuss the physical-process first law, give calculational details, and use a toy model to illustrate the ambiguity of nonlinear extensions.

Assessment, novelty, and recommendation

The question is worthwhile. The manuscript usefully emphasizes that the boost-weight and amplitude expansions are not identical bookkeeping schemes, and the distinction between a nonlinear geometric representative and the order to which it has an entropy interpretation deserves a clear treatment. The covariant phase-space identities in Section 2 are standard and the broad structure of the curvature-squared example is consistent with the cited entropy-current calculations. The toy model also makes the possible nonlinear mismatch intuitive.

The basic differential relation, however, is already a principal result of Hollands–Wald–Zhang, and its explicit verification for general f(Riemann) theories was given by Visser–Yan, as the manuscript itself notes. The Riemann-squared example is therefore mainly a check. The genuine novelty must reside in the claimed local inverse algorithm and in a precise comparison of the two approximation schemes. At present the former is not established with sufficient rigor. In addition, the displayed

ODE solution in equation (3.1) contains a definite sign error, the argument suppresses important integration and ambiguity data, and the paper alternates between an integrated entropy and a local entropy density/current. These issues affect the central claim rather than only its presentation.

I therefore recommend **major revision before reconsideration**. If the inverse construction is reformulated as a precise theorem, proved under clearly stated hypotheses, and distinguished sharply from the results already in the literature, the work could be suitable for Physical Review D or JHEP. The following points are essential.

Major comments

1. **The integrated solution in equation (3.1) has the wrong sign.**

   Starting from

   $$S_{\rm dyn} = S_{\rm Wall} - v\partial_v S_{\rm Wall},$$

   one obtains

   $$\partial_v\left(\frac{S_{\rm Wall}}{v}\right) = -\frac{S_{\rm dyn}}{v^2}.$$

   Thus the first line of equation (3.1) is correct, but integration by parts gives

   $$S_{\rm Wall}(v) = S_{\rm dyn}(v) - v\int^v du\,\frac{\partial_u S_{\rm dyn}(u)}{u} + Cv,$$

   not the expression with a plus sign printed in the second line. Direct substitution of the printed result yields $S_{\rm Wall} - vS'_{\rm Wall} = S_{\rm dyn} - 2vS'_{\rm dyn}$ rather than $S_{\rm dyn}$. For example, if $S_{\rm dyn} = S_0 + av^2$, the required solution is $S_{\rm Wall} = S_0 - av^2 + Cv$, whereas the printed formula gives $S_0 + 3av^2 + Cv$.

   This equation motivates the entire inverse problem and must be corrected. The associated regularity statement should also distinguish smoothness from analyticity. The condition $S'_{\rm dyn}(0) = 0$ removes the logarithm when $S'_{\rm dyn}/v$ is regular, but it neither proves analyticity nor fixes the homogeneous term Cv. The authors must specify the lower limit, the needed regularity class, and the physical or covariance condition that removes the homogeneous solution.

2. **The coefficient-matching step in equations (3.5)–(3.10) is not presently a proof of a local inverse.**

   Immediately after equation (3.6), the manuscript correctly observes that the purported coefficients of the explicit powers of v remain functions of v, because the perturbation is arbitrary. It then concludes that each coefficient must vanish separately. That conclusion does not follow from a power-series argument: an identity $\sum_m v^m F_m(v) = 0$ does not imply $F_m(v) = 0$ when the F_m have arbitrary v dependence. What may rescue the construction is a stronger identity in the jet variables of an arbitrary perturbation, with independent tensor monomials grouped by derivative order and boost weight. Such an argument is not supplied.

   A revised proof should introduce a sum over all independent tensor structures at fixed m, rather than the single symbolic product $S^{[m]}T_{(m)}$, and show which perturbation jets are independent after Bianchi identities, integration by parts, and residual gauge constraints. It must also justify the assertion following equation (3.10) that every highest-weight term has a local ∂_v primitive. Merely having $N+1$ unpaired ∂_v derivatives is not by itself a proof when covariant derivatives, contractions, commutators, and several local monomials are present. The recursion should be demonstrated for a general finite-derivative term, including the possible kernel at every integration step.

   The relevant hypotheses should be part of the result: finite derivative order, locality and polynomiality assumptions, smooth bifurcate Killing horizon, a perturbation and gauge that preserve the Gaussian null setup, compact cross-sections (if spatial divergences are discarded), and equality only through first order in amplitude. Without this work, equation (3.8) is a necessary regularity condition but has not been shown to be sufficient for the advertised algebraic construction.

3. **The paper does not yet distinguish a local density/current from an integrated entropy functional.**

   The comparison in equations (2.23)–(2.26) and the entire algorithm in Section 3 are performed after integration over Σ_v. In particular, equation (3.11) sets the integral of $\nabla_i \mathcal{J}^i$ to zero. This loses precisely the spatial-divergence information that distinguishes local entropy densities and entropy currents. From such data one can at most reconstruct an integrated functional modulo horizon divergences; one cannot uniquely recover a pointwise S_{Wall} density or the spatial current $\mathcal{J}^i$.

   The abstract and Sections 3–4 repeatedly call the output a "local expression." The authors should either carry out the inverse construction before spatial integration, retaining the Q^{ri} and $\mathcal{J}^i$ terms, or weaken the claim to locality in the affine parameter for an integrated entropy, modulo spatial total derivatives. They should also state what happens for noncompact cross-sections or cross-sections with boundary. The fact that Section 4 can import the known Riemann-squared current from the full Q^{ri} calculation does not show that the integrated recursive algorithm can reconstruct it.

4. **Integration constants and JKM ambiguities require a systematic treatment.**

   Equation (2.25) establishes only that

   $$\partial_v \left[S_{\text{dyn}} - (1 - v\partial_v) S_{\text{Wall}} \right] = O(\epsilon^2).$$

   The statement that the two entropies agree on stationary solutions does not, by itself, fix a v-independent term for an arbitrary off-shell dynamical perturbation. One needs a boundary condition on the same perturbation, for example at a regular bifurcation surface or in a stationary era, and a proof that the two representatives agree there. The same issue reappears as the Cv homogeneous solution of equation (3.1).

   Moreover, Q, Θ, $\mathcal{B}$, and the entropy density admit the usual covariant-phase-space and JKM improvements. The manuscript should show how equations (3.11)–(3.22) and the recursive inverse transform under these ambiguities, which representative is held fixed, and in what sense the answer is unique. This is especially important because the cited Hollands–Wald–Zhang entropy is unambiguous only to the perturbative order in question, while Appendix E deliberately exploits freedom in its nonlinear completion. A clear equivalence class of outputs would substantially strengthen the result.

5. **The perturbative scope and several physical claims must be made consistent.**

   Throughout Sections 2–4, equalities derived from equations (2.14)–(2.20) hold only to first order in the amplitude expansion, while the boost-weight representatives retain selected higher-order terms. This distinction is the paper's theme, but equations (1.5), (2.27)–(2.29), (3.11)–(3.22), and many steps in Section 4 sometimes appear as exact identities. Each occurrence should state whether it is an off-shell boost-weight identity, an identity of first variations, an equality modulo $O(\epsilon^2)$, or an on-shell statement. In particular, the total-derivative step in equation (4.24) uses stationary-background relations and is true only at the stated linearized order.

   The statement in Section 5 that "up to the linear order ... entropy does not change" contradicts equations (1.1)–(1.4). External matter with $\delta T_{vv} = O(\epsilon)$ produces an $O(\epsilon)$ entropy change. It is the gravitational-wave flux for a vacuum metric perturbation that typically begins at quadratic order. These cases must be separated. Likewise, $E_{vv} \sim -\partial_v^2 S_{\text{Wall}}$ plus the null energy condition implies monotonic increase only after imposing the future equilibrium condition $\partial_v S_{\text{Wall}} \to 0$; the null energy condition alone gives concavity. The maximum-entropy and teleological statements in

Sections 1, 2, and 5 should include this boundary condition and distinguish the local geometric density from the teleological choice of event horizon.

6. **The scope and novelty relative to the cited literature need a sharper statement.**

   The relation between the improved Noether charge and the Dong–Wall entropy is already established in Hollands–Wald–Zhang, and Visser–Yan explicitly checked it for general f(Riemann) theories in Gaussian null coordinates. The Riemann-squared result in equations (4.9)–(4.11) is a special case of that check as well as a reuse of the entropy current tabulated in earlier work by some of the present authors. The manuscript should say plainly which part of Section 4 is an independent calculation serving as a validation rather than a new physical result.

   The new content could be valuable if the authors prove that a specified class of improved Noether-charge entropies admits a unique affine-local preimage, modulo stated ambiguities. The current phrase "provided certain technical conditions are satisfied" is too vague for the main claim. Comparison with the expansion around equation (53) of Hollands–Wald–Zhang should be explicit, and the manuscript should explain what the present recursion adds to the already known general relation. It would also be helpful to state whether the argument covers Lagrangians with covariant derivatives of curvature, nonminimally coupled matter, and gauge fields, or only local pure-metric theories.

Minor comments

1. Equation (D.14) states $\xi_r = r$. For $\xi = v\partial_v - r\partial_r$ and the metric (2.1), the correct horizon-adapted covector component is $\xi_r = v$. Equation (D.16) and the calculation below equation (D.17) in fact use the latter value, so the displayed equation must be corrected and the remaining components rechecked for consistency.

2. The statement in the Introduction that Wald entropy is defined only on the bifurcation surface is too strong. On a stationary Killing horizon the Noether-charge expression can be transported to other stationary cross-sections; the real issue is its extension to a nonstationary slice and the associated ambiguity.

3. In equation (2.6), the same symbol k labels both the number of positive-weight factors and the summation in $\sum_k m_k$. The latter should be $\sum_{i=1}^{k} m_i$. The manuscript also alternates between $\mathfrak{T}^{\{1\}}$ and $\mathfrak{T}^{(1)}$, and between $S^{\text{zero}}_{\text{Wall}}$ and $S^{\text{zero}}_{\text{wall}}$.

4. The notation $\bar{K}_{ij}$ for radial extrinsic curvature is easily confused with the overbar denoting a stationary background. A different symbol, or an explicit reminder at the start of each calculation, would improve readability.

5. Equations (3.20)–(3.21) switch without explanation from square superscripts, such as $\alpha^{[0]}$, to parentheses. The notation should be made uniform. Similarly, equation (3.11) uses $\tilde{Q}_{rv}$ where the surrounding formulas use $\tilde{Q}^{rv}$.

6. The normalization and sign conventions for the pure Riemann-squared contribution should be stated, including whether an overall coupling and the Einstein-Hilbert term have simply been suppressed. This would make equations (4.9)–(4.11) easier to compare with the cited formulas.

7. The author/affiliation labels in the title block appear inconsistent: the second and third authors use labels for which no correspondingly labeled affiliations are declared. This should be fixed before submission.

8. There are numerous small language and typesetting issues, including “for for,” “atleast,” “upto,” inconsistent “equation/eqn.” usage, and doubled manual line breaks. A careful proofread would materially improve the presentation.

Recommendation

The manuscript contains a useful organizing idea, but its main new claim is not yet supported by a correct and complete derivation. I recommend major revision. Reconsideration would be appropriate after the sign error is fixed, the local inverse is formulated and proved with all hypotheses and ambiguities explicit, the integrated-versus-local scope is corrected, and the perturbative physical claims are made consistent.