Published for SISSA by Springer

Received: July 1, 2017
Revised: October 6, 2017
Accepted: December 25, 2017
Published: January 17, 2018

# Entropy, extremality, euclidean variations, and the equations of motion

**Xi Dong**[a,b] **and Aitor Lewkowycz**[c]

[a] *School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540, U.S.A.*

[b] *Department of Physics, University of California, Santa Barbara, CA 93106, U.S.A.*

[c] *Stanford Institute for Theoretical Physics, Department of Physics, Stanford University, Stanford, CA 94305, U.S.A.*

*E-mail:* xidong@ucsb.edu, lewkow@stanford.edu

Abstract: We study the Euclidean gravitational path integral computing the Rényi entropy and analyze its behavior under small variations. We argue that, in Einstein gravity, the extremality condition can be understood from the variational principle at the level of the action, without having to solve explicitly the equations of motion. This set-up is then generalized to arbitrary theories of gravity, where we show that the respective entanglement entropy functional needs to be extremized. We also extend this result to all orders in Newton's constant $G_N$, providing a derivation of quantum extremality. Understanding quantum extremality for mixtures of states provides a generalization of the dual of the boundary modular Hamiltonian which is given by the bulk modular Hamiltonian plus the area operator, evaluated on the so-called modular extremal surface. This gives a bulk prescription for computing the relative entropies to all orders in $G_N$. We also comment on how these ideas can be used to derive an integrated version of the equations of motion, linearized around arbitrary states.

Keywords: AdS-CFT Correspondence, Gauge-gravity correspondence, Classical Theories of Gravity, Black Holes in String Theory

ArXiv ePrint: 1705.08453



https://doi.org/10.1007/JHEP01(2018)081

JHEP01(2018)081

## Contents



JHEP01(2018)081

## 1 Introduction and summary of results

Quantum entanglement has become a crucial aspect of understanding many physical systems including quantum gravity. A universal property of quantum gravity is that entropy satisfies an area law. This was first discovered for black holes [1–3], and more recently it was generalized in the context of AdS/CFT correspondence [4–6] by Ryu and Takayanagi [7, 8]. They gave an elegant prescription for the entanglement entropy of any spatial region $R$ in a holographic boundary theory in terms of the area of an extremal surface in the bulk spacetime:

$$S_R = \underset{X \sim R}{\text{ext}} \frac{A(X)}{4G_N}. \tag{1.1}$$

Here the entanglement entropy is defined in the boundary theory as the von Neumann entropy $S_R \equiv -\operatorname{Tr}\rho_R \log \rho_R$ of the reduced density matrix $\rho_R$, and is a measure of entanglement between the region $R$ and its complement $\overline{R}$. The constraint $X \sim R$ means that the Ryu-Takayanagi (RT) surface $X$ is homologous to the boundary region $R$, and $G_N$ denotes Newton's constant. This prescription for holographic entanglement entropy was derived from AdS/CFT in [9]. Furthermore, it is valid in general time-dependent cases [10, 11].

In general, the gravitational theory in the bulk is described at low energies in terms of Einstein gravity corrected by higher derivative interactions. These interactions generate higher derivative corrections to the RT formula (1.1). A prescription for these corrections was given in [12, 13] and has the form

$$A_{\text{gen}} = S_{\text{Wald}} + S_{\text{extrinsic}} \tag{1.2}$$

where the first term is the Wald entropy and the second consists of corrections from the extrinsic curvature of the RT surface. Since $A_{\text{gen}}$ is the full classical contribution to the gravitational entropy, we will refer to it as the "generalized area".[1] However, it has been an open question whether the extremization procedure in (1.1) works for general higher derivative gravity, using variations of the action. Our first result is that it does:

$$S_R = \underset{X\sim R}{\text{ext}}\, A_{\text{gen}}(X). \tag{1.3}$$

As a byproduct of this result, one can generalize the derivation of the integrated linearized equations of motion from the first law of entanglement [14–17] to arbitrary regions and states. This is done by defining the variation of the modular Hamiltonian using the replica trick and from the linearized equations of motion for an arbitrary state one should in principle be able to get the nonlinear equations of motion.

The RT prescription (1.1) and its higher derivative generalization (1.3) are valid in the large-$N$ limit of the boundary theory. Beyond the leading order in this limit, they would receive $1/N$ corrections from quantum effects in the bulk. A natural prescription for these quantum corrections is

$$S_R = \underset{X\sim R}{\text{ext}}\, S_{\text{gen}}(X), \quad S_{\text{gen}} \equiv \langle A_{\text{gen}} \rangle + S_{\text{bulk}}, \tag{1.4}$$

where the "generalized entropy" $S_{\text{gen}}$ is the sum of the expectation value of the generalized area $\langle A_{\text{gen}} \rangle$ and a bulk entanglement entropy $S_{\text{bulk}}$. The bulk entanglement entropy is defined with respect to the bulk spatial region between the RT surface $X$ and the boundary region $R$. The domain of dependence of this region defines the notation of the entanglement wedge [18–20]. It is worth noting that after extremization $X$ is known as the quantum extremal surface.

The prescription (1.4) agrees with the one-loop result of [21, 22] and was conjectured in [23] to hold for all loops. Our second result is to establish this from AdS/CFT to all orders in $1/N$.

[1] For Einstein gravity, $A_{\text{gen}} = \frac{A}{4G_N}$.



JHEP01(2018)081

Furthermore, entanglement entropy is not the only measure of quantum entanglement. To better understand the structure of entanglement, we also need the modular Hamiltonian

$$K_\rho \equiv -\log \rho \tag{1.5}$$

for a quantum state described by the density matrix $\rho$, as well as the relative entropy

$$S_{\rm rel}(\rho|\sigma) \equiv {\rm Tr}\left[\rho \log \rho - \rho \log \sigma\right] \tag{1.6}$$

which is a measure of distinguishability between an arbitrary state $\rho$ and a reference state $\sigma$. Our third result is

$$\langle K_{R,\sigma}\rangle_\rho = \operatorname*{ext}_{X\sim R}\left[\langle A^X_{\rm gen}\rangle_\rho + \langle K^X_{{\rm bulk},\sigma}\rangle_\rho\right] \tag{1.7}$$

where $K_{R,\sigma}$ is the modular Hamiltonian for the boundary region $R$ for the state $\sigma$, $A_{\rm gen}$ is viewed as an operator on the surface $X$ giving its generalized area, and $K_{\rm bulk}$ is the bulk modular Hamiltonian in the spatial region between $X$ and $R$. After extremization we call $X$ the "modular extremal surface" for the state $\sigma$.

Using the prescription (1.7) for the modular Hamiltonian, we find for the relative entropy

$$S_{\rm rel}(\rho|\sigma) = \langle A^{X_\sigma}_{\rm gen} + K^{X_\sigma}_{{\rm bulk},\sigma}\rangle_\rho - \langle A^{X_\rho}_{\rm gen} + K^{X_\rho}_{{\rm bulk},\rho}\rangle_\rho \tag{1.8}$$

where $X_\sigma$ and $X_\rho$ are modular extremal surfaces defined by (1.7) for the states $\sigma$ and $\rho$ respectively. Here we have dropped explicit references to the boundary region $R$ for brevity, and $\langle\cdots\rangle_\rho$ denotes the expectation value ${\rm Tr}\,(\rho\cdots)$ in the state $\rho$.

The results (1.7) and (1.8) agree with one-loop results of [24]. As we will show using AdS/CFT, they are valid to all orders in $1/N$. It is interesting to note from (1.8) that the boundary relative entropy is equal to the bulk relative entropy only at the one-loop order [24], and they generally differ at two loops or higher. This is because the two modular extremal surfaces $X_\sigma$ and $X_\rho$ differ by $O(G_N)$ in general.

Recently, the AdS/CFT dictionary has been clarified by viewing holography as a quantum error correcting code [25]. The relation between the bulk and boundary relative entropy was used in [26] to prove a theorem for reconstructing bulk operators in the entanglement wedge of $R$ in terms of boundary operators on $R$, and the one-loop result can be used to obtain an explicit large-$N$ reconstruction formula in terms of the modular flow [27]. As we will see, the all-loop result (1.8) can be used to extend the reconstruction theorem to all orders in $1/N$, at least for bulk operators at a fixed distance away from the RT surface, but it is not yet clear how to generalize the modular flow construction beyond one loop. A related issue is that the complementary recovery property discussed in [28] holds only at the one-loop order.

The outline of this paper is as follows. We begin in section 2 with a review of the classical statement of extremality and rephrase it in a way that can easily be generalized to arbitrary theories of gravity, using variations of the action. Section 3 is independent of the rest of the paper and uses the variational principle to derive the integrated equations of motion around an arbitrary background using the first law. In section 4, we generalize the classical discussion of section 2 by including quantum fields in the bulk theory, providing a



JHEP01(2018)081

derivation of quantum extremality. In section 5, we use quantum extremality for mixtures of states to write a formula for the bulk dual of the modular Hamiltonian to all orders in $G_N$. We conclude with some closing thoughts in the discussion.

## 2 Classical statement of extremality from variations

Let us start with a review of the replica trick applied to AdS/CFT. In the boundary theory, the von Neumann entropy may be determined by the $n \to 1$ limit of the Rényi entropy

$$S_n \equiv \frac{1}{1-n} \log \mathrm{Tr}\, \rho^n, \tag{2.1}$$

where $n$ is known as the Rényi index. When $n$ is an integer greater than 1, the Rényi entropy can be calculated from

$$S_n = \frac{1}{1-n} \log \frac{Z_n}{Z_1^n}, \tag{2.2}$$

where $Z_n$ is the partition function of the boundary theory on a manifold known as the $n$-fold branched cover. This partition function can be calculated via AdS/CFT. In the large-$N$ limit, we find the solution $M_n$ to the bulk equations of motion with the $n$-fold cover as the boundary condition and calculate its on-shell action $I_n$. Up to $1/N$ corrections, we have $\log Z_n = -I_n$. When there are more than one bulk solution, we choose the dominant one which has the smallest on-shell action.

The $n$-fold cover on the boundary enjoys a $\mathbb{Z}_n$ symmetry permuting the $n$ replicas cyclically. As in [9], we assume that the $\mathbb{Z}_n$ replica symmetry extends to the dominant bulk solution $M_n$. Let us take the quotient of the bulk solution $M_n$ by the $\mathbb{Z}_n$ replica symmetry. This quotient amounts to considering the action $\hat{I}_n = I_n/n$ which can be thought of as the on-shell action of the orbifold geometry $\hat{M}_n \equiv M_n/\mathbb{Z}_n$. The orbifold has a conical singularity at the $\mathbb{Z}_n$ fixed points. The derivative of the orbifold action with respect to $n$ is the modular entropy introduced in [29]:

$$\tilde{S}_n \equiv -n^2 \partial_n \left( \frac{1}{n} \log \mathrm{Tr}\, \rho^n \right) = n^2 \partial_n \hat{I}_n. \tag{2.3}$$

Since the orbifold geometry is seemingly singular, when doing variations one has to be careful with possible boundary terms at the tip of the cone. In other words, (2.3) reduces to a boundary term on the conical defect, and taking the $n \to 1$ limit we find the von Neumann entropy $S$ in terms of some geometric quantity $A_{\text{gen}}$ on a codimension-2 surface $X$.

The goal of this section is to show that for classical theories of gravity, the equations of motion close to $n \approx 1$ imply that the surface $X$ has to be extremal with respect to the entanglement entropy functional $A_{\text{gen}}$:

$$\delta_{\text{diff}} A_{\text{gen}} = 0 \tag{2.4}$$

where diff denotes to a diffeomorphism that would change the location of $X$ where the functional is evaluated.



JHEP01(2018)081

### 2.1 Double variations

If we vary the action around the solution $g_n$ to the equations of motion with an off-shell deformation $\delta g_n$ that preserves the conical deficit angle and vanishes on the asymptotic boundary, we have

$$\delta \hat{I}_n = \int_{\hat{M}_n} E_n \delta g_n + \int_{\partial \hat{M}_n} \Theta(g_n, \delta g_n)\Big|_{r=\epsilon} = 0 \tag{2.5}$$

where we have used the notation of [30]: $E_n \equiv \delta \hat{I}_n / \delta g$ denotes the equations of motion at integer $n$, and $\Theta(g_n, \delta g_n)$ is the boundary term at the tip of the cone, linear in $\delta g$ and obtained from integrating the Lagrangian by parts after a variation. The solution $g_n$ satisfies the equations of motion, leading to $E_n = 0$. The boundary term is evaluated on a regulated surface $r = \epsilon$ where $r$ is the radial distance from the tip of the cone, and we take the $\epsilon \to 0$ limit at the end of the calculation. The claim of (2.5) is that the boundary term vanishes in this limit.

For integer $n$, it is clear that (2.5) holds, since we can go to the parent space $M_n$ where there is no physical boundary at the $\mathbb{Z}_n$ fixed points.

In the next subsection, we will argue that (2.5) holds for general values of $n$. For now we will explore the consequences of this, saving the details for later. Since (2.5) is zero for any $n$, its derivative with respect to $n$ is also zero:

$$\partial_n \delta \hat{I}_n\big|_{n=1} = 0. \tag{2.6}$$

Note that this follows as long as the equations of motion are obeyed at $n \approx 1$.

We can take the two variations $\partial_n$ and $\delta$ in (2.6) in the opposite order, so that $\partial_n$ gives us the entanglement entropy functional $A_{\text{gen}}$ for a metric in the neighbourhood of the on-shell metric. Up until now we have kept the variation of the metric $\delta g_n$ arbitrary except for the boundary conditions of preserving the conical deficit angle and vanishing on the asymptotic boundary. Let us now choose $\delta g_n$ to become a diffeomorphism at $n = 1$.[2] If we consider the variations in the opposite order for a diffeomorphism at $n = 1$, we obtain

$$\begin{aligned} \partial_n \hat{I}_n|_{n=1} &= \lim_{\delta n \to 0} \frac{\hat{I}_{1+\delta n}[g_{1+\delta n}] - \hat{I}_1[g_1]}{\delta n} = A_{\text{gen}} \\ \left(\partial_n \hat{I}_n + \delta \partial_n \hat{I}_n\right)\Big|_{n=1} &= \lim_{\delta n \to 0} \frac{\hat{I}_{1+\delta n}[g_{1+\delta n} + \delta g_{1+\delta n}] - \hat{I}_1[g_1]}{\delta n} = A_{\text{gen}} + \delta_{\text{diff}} A_{\text{gen}} \end{aligned} \tag{2.7}$$

where $A_{\text{gen}}$ is defined from (2.7) and can be computed using the conical method of [12, 13, 31] or directly using the $n \to 1$ limit of the Wald entropy (see section 2.3). This discussion is independent of how one computes it. To derive the second line, we used that $g_1 + \delta g_1$ is a solution to the equations of motion at $n = 1$ and we can use the same entropy functional $A_{\text{gen}}$ evaluated on a slightly dislocated surface $X + \delta X$. In section 2.3, it will be clear how this works when one can take the $\partial_n$ variation inside the action.

[2] We do not put additional constraints on $\delta g_n$ away from $n = 1$ except for the boundary conditions. In general $\delta g_n$ will be off-shell at finite $n - 1$ because the conical boundary condition essentially fixes the on-shell solution as $g_n$.



JHEP01(2018)081

Taking the difference of the two equations in (2.7) and compare it with (2.6) we get

$$\delta_{\text{diff}} A_{\text{gen}} = 0. \tag{2.8}$$

In other words, the entanglement entropy functional should be stationary with respect to shifts in the surface. This argument uses the equations of motion linearized in $n-1$ which is the same condition that led to extremality in [9]. However, the advantage of our method here is that by considering variations of the action, we do not have to evaluate the equations of motion explicitly.

We expect this to be true for an arbitrary theory of gravity. In the next subsections we discuss the subtleties that lie in these cases.

### 2.2 Boundary terms and the $n \to 1$ limit

In the previous discussion, we used the equations of motion at integer $n$ and at the same time deformed the metric off-shell (at finite $n-1$). However, since we want to do two variations of the action, we want to be able to define $\partial_n I(g_n)$ for an slightly off-shell metric, $g_n + \delta g_n$. We want to restrict to "regular" $\delta g_n$: deformations of the metric which give a finite contribution to the action and do not change the strength of the conical singularity. This constraints the variation and allows for a well defined action for the deformed off-shell geometry.

We would first like to show that $\delta \hat{I}_n = 0$ for all $n$. We can first consider Einstein gravity, where we get

$$\delta \hat{I}_n = \int_{\partial \hat{M}_n} \Theta_{\text{Einstein}}(g_n, \delta g_n)\Big|_{r=\epsilon} = \int_{\partial \hat{M}_n} \sqrt{g_n}(\nabla^b \delta g_{rb} - g_n^{bc}\nabla_r \delta g_{bc})\Big|_{r=\epsilon} \overset{\text{regular } \delta g_n}{=} 0. \tag{2.9}$$

Because $\sqrt{g_n} \propto r = \epsilon$, it is clear that only if $\delta g_n$ diverges approaching the tip one can get a non-zero answer.

More generally, for an arbitrary higher derivative theory, we have [30]:

$$\delta \hat{I}_n = \int_{\partial \hat{M}_n} \Theta(g_n, \delta g_n)\Big|_{r=\epsilon} = \int_{\partial \hat{M}_n} r E_{rbcd}\nabla_c \delta g_{bd}\Big|_{r=\epsilon} \tag{2.10}$$

where $E_{rbcd}$ would be the equations of motion for $R_{abcd}$, viewed as an independent field. For example for $f(\text{Riemann})$, $E_{rbcd} = \frac{\partial \mathcal{L}}{\partial R_{rbcd}}$.

It is clear for Einstein gravity that a regular variation of the metric cannot give a finite contribution to the boundary term. However, while (2.10) $= 0$ at integer $n$, we would also like to argue that this is true for $1 < n < 2$. The regularity condition for the variation requires the boundary term (2.10) to be finite if not zero. This is because there are no divergent terms at $n = 1$ and we are choosing the $\delta g_n$ to keep the variation finite for $n > 1$. However, the most general metric compatible with replica symmetry will be an expansion with positive powers of $r^{n-1}$ and integer powers of $r$ (see next section). Given that we are working at integer $n$ until the very end, $\epsilon^{n-1} \to 0$, which implies that there cannot be a finite term. This implies that (2.10) is zero.



JHEP01(2018)081

### 2.3 Variational approach for the gravitational entropy

While $\partial_n \hat{I}_n|_{n=1}$ in (2.7) can be computed explicitly using squashed cones, that approach requires being careful with several subtleties that arise in the $n \to 1$ limit and there is currently no complete formula for an arbitrary theory of gravity. In this subsection, we are going to propose an equivalent but perhaps clearer approach than (2.7), where we think of $\partial_n$ as a variation inside the action.

Close to the conical singularity, the metric near $n \approx 1$ will schematically look like (we refer the reader to [12] for more details):

$$\begin{aligned} g_n &= dr^2 + \frac{r^2}{n^2} d\tau^2 + (\gamma_n + K_n r^n e^{i\tau} + \cdots) dy^2 + \cdots \equiv g_{n;0} + r^{2(n-1)} g_{n;1} + \cdots \\ \gamma_n &= \gamma_{n;0} + \gamma_{n;1} r^{2(n-1)} + \gamma_{n;2} r^{4(n-1)} + \cdots, \quad K_n = K_{n;0} + K_{n;1} r^{2(n-1)} + \cdots \end{aligned} \tag{2.11}$$

where $\cdots$ denotes terms which are higher order in $r$, and $\tau$ has period $2\pi$. The $n = 1$ metric "splits":[3] it is determined by the sum of different terms at $n > 1$

$$g_{n=1} = g_{1;0} + g_{1;1} + g_{1;2} + \cdots . \tag{2.12}$$

This was seen as a problem for the squashed cone approach in [32] (see also [33, 34]): in order to determine the "splitting" one has to solve the most divergent part of the equations of motion, which could be problematic because in order to determine the form of $A_{\text{gen}}$ explicitly one needs the equations of motion at $n \sim 1$.

We would like to understand if we can treat $\partial_n g_n$ outside the $r = \epsilon$ tube as a small variation inside the action integral. This is not true at $n = 1$: the metric might include terms $g_n \propto \epsilon^{2(n-1)}$, which give $\partial_n g_n \propto \epsilon^{2(n-1)} \log \epsilon$, which is not small as $n \to 1$ (at fixed but small $\epsilon$). However, we can avoid this issue by working at $n > 1$. In this case, we expect that $\partial_n g_n$ is a small variation[4] and thus we can apply (2.10) for $\partial_n g_n$. All the contribution from $\partial_n g_n$ comes from the $g_{\tau\tau}$ component in (2.11). This gives the Wald entropy at finite (but non-integer) $n - 1$:

$$\widetilde{S}_n = \partial_n \hat{I}_n = n^{-2} S_{\text{Wald}}(g_{n;0}) + \int_{\hat{M}_n} E_n \partial_n g. \tag{2.13}$$

This formula is valid for non-integer $n$ and it is a finite $n - 1$, off-shell version of (2.7).[5] In order to avoid contradictions, it is important that the $n \to 1$ limit of the Wald entropy at finite $n - 1$ is *not* the Wald entropy at the $n = 1$ solution. The reason is that the Wald entropy at finite $n - 1$ is written in terms of the $g_{n,0}$ fields in (2.11), while at $n = 1$ one only have access to the sum over them (2.12). We expect the equations of motion close to $n = 1$ to determine $g_{1;0}$ in terms of $g_{n=1}$.

[3] One way to define $g_{n;0}, g_{n;1}, \cdots$ is to require that they contain only integer powers of $r$.

[4] This is true as long as there are no terms in the metric that go like $g \propto \epsilon^{f(n)}$ where $f(n)$ vanishes for any $1 < n < n_c$, in which case $\partial_n g$ is a small variation in a finite neighbourhood around $n = 1$ (not including $n = 1$ itself). We think that this is a very reasonable assumption and we have not been able to find any counterexample.

[5] We expect this formula to hold for any $n$ as long as $\partial_n g_n$ is a small variation. If for some reason, the metric splits at some $n_c$, we would define $\widetilde{S}_{n_c} = \lim_{n \to n_c} S_{\text{Wald}}(g_{n;0})$, as we will do in (2.14).



JHEP01(2018)081

By carefully taking the limit, one gets the generalized area:

$$\widetilde{S}_1 = A_{\text{gen}}[g_{n=1}] = \lim_{n\to 1^+} S_{\text{Wald}}(g_{n;0}) \tag{2.14}$$

where we used the $n = 1$ equations of motion. Note that this approach was used before for Einstein gravity in [9, 29]: because of the simplicity of this theory, one can evaluate (2.14) directly at $n = 1$ without worrying about subtleties in the limit.

For readers familiar with the squashed cone approach to higher-derivative entanglement entropy [12, 13, 31], (2.14) might look surprising, because $A_{\text{gen}}$ has a contribution from the Wald entropy at $n = 1$, but it also has an "anomalous" contribution which depends on the extrinsic curvature [12]. The anomalous contribution depends on the details of how the metric splits. In our case, $S_{\text{Wald}}$ is explicitly defined in terms of the Lagrangian and the $g_{1;0}$ metric. In this way, our approach gives an explicit formula for the holographic entanglement entropy: the Wald entropy of the split metric $g_{1;0}$. However, to determine its form in terms of $n = 1$ quantities, one has to solve the most divergent part of the equations of motion.

One should be able to show explicitly how (2.14) relates the squashed cones contribution and the Wald entropy. For Lovelock theories, it is easy to see how this works: $A_{\text{gen}}$ is just given by the Wald entropy in terms of induced Riemann tensor, which is the $n \to 1$ limit of the projected Riemann tensor on the surface. For higher derivative theories which have fewer derivatives than Lovelock, such as the one considered in [35], we do not have an "anomalous" contribution to $A_{\text{gen}}$ and there are no subtleties in taking the $n \to 1$ limit. In appendix A, we consider a set of two-dimensional examples which we believe capture (2.14) more generally.

In our discussion, we have always focused on families of metrics (not necessarily on-shell) which keep the action finite. It is often the case that in the $r \approx 0$ expansion, the most general form for the metric gives rise to an infinite action. In other words, there are some divergent terms in the equations of motion which give a divergent contribution to the gravitational action, while other metric contributions with divergent equations of motion have a finite action (for example, changes in the location of the surface). We will always work with metrics which have a finite action, which is equivalent to imposing the most divergent part of the equations of motion. Even if this class of metrics will depend on the Lagrangian, it is rather universal: it will not depend on the location of the conical singularity. In this way, by requiring the action to be finite, we expect that one can understand the relation between $g_{1;0}$ and $g_{n=1}$, which would determine $S_{\text{Wald}}$ explicitly in terms of $n = 1$ quantities.

## 3 The first law of entanglement and equations of motion

This section is a side product of the previous section. It is independent of the rest of the paper and it will not be mentioned again until the discussion. In the previous sections, we have explained how, in classical gravity, the commutativity of the double variation $\partial_n, \delta_{\text{diff}}$ implies the extremality of the entangling functional. We can also use this framework to consider more general variations which do not vanish at the boundary. In holography, it is



JHEP01(2018)081

natural to consider turning on a small source. This framework naturally allow us to derive the integrated equations of motion by assuming that the entanglement entropy is given by the area thus generalizing [16, 17].

The idea is that, from the field theory perspective, we can think of the second variation commuting as the first law [14, 15]: $[\delta, \partial_n]\frac{\log \mathrm{Tr}\rho^n}{n}|_{n=1} = \delta S - \partial_n \mathrm{Tr}\delta\rho\rho^{n-1} = \delta S - \delta K$. We would like to understand if we can recover this from the bulk point of view.

In order to do this, we want to be in the same setup as [17]. Consider a deformation of the density matrix which changes the one point function of the stress tensor by a small amount, $\delta\langle T_{\mu\nu}\rangle \ll 1$, which is achieved by turning on the respective source, the boundary metric. If we add a term $\lambda \int d^dx \delta g^{\mu\nu}_{\mathrm{bdy}} T_{\mu\nu}$ to the Lagrangian, then the stress tensor will get an expectation value linear in $\lambda$ (to first order in the deformation). In the original geometry, we expect the same change in the action by computing the variation of the action:

$$\delta_\lambda I = \int_M E\delta_\lambda g + \lambda \int_{M_\infty} d^dx \langle T^{BY}_{\mu\nu}\rangle \delta g^{\mu\nu}_{\mathrm{bdy}} \,. \tag{3.1}$$

The variation of the action will be given by the equations of motion plus a boundary term, the usual integral of the Brown-York stress tensor. This boundary term will vanish if the expectation value of the stress tensor is zero.

Now, if we repeat the same for the Rényi entropies, we obtain:

$$\delta_\lambda \hat{I}_n = \int_{M_n} E_n \delta_\lambda g_n + \lambda \int_{M_\infty} \langle T^{BY}_{\mu\nu}\rangle_n \delta g^{\mu\nu}_{\mathrm{bdy}} \,. \tag{3.2}$$

We can analytically continue this expression in $n$, take its $n$ derivative, and express it in terms of boundary quantities using the standard dictionary $\langle T^{BY}\rangle = \langle T\rangle$:

$$\partial_n \delta_\lambda \hat{I}_n|_{n=1} = \lambda \int d^dx \langle K T_{\mu\nu}\rangle \delta g^{\mu\nu}_{\mathrm{bdy}} + \partial_n \int_{M_n} E_n \delta_\lambda g_n|_{n=1} = \delta_\lambda \langle K\rangle + \partial_n \int_{M_n} E_n \delta_\lambda g_n|_{n=1} \,. \tag{3.3}$$

This formula for the variation of the boundary Hamiltonian from analytically continuing the one point function at integer $n$ was discussed previously in [36, 37]. Note that in the case where the modular Hamiltonian is local, the right-hand side (r.h.s.) will be given by $\int_R d\Sigma^\mu \xi^\nu \delta\langle T_{\mu\nu}\rangle$ and this can be understood from the left-hand side (l.h.s.) because $\delta\langle T_{\mu\nu}\rangle = \int d^dx \langle T_{\mu\nu} T_{\alpha\beta}\rangle \delta g^{\alpha\beta}_{\mathrm{bdy}}$. So we are in exactly the same setup as [17].

We can try to understand the variations in the opposite order:

$$\partial_n \hat{I}_n|_{n=1} = \int E\partial_n g + A_{\mathrm{gen}} \to \delta_\lambda \partial_n \hat{I}_n|_{n=1} = \delta_\lambda \int E\partial_n g + \delta_\lambda A_{\mathrm{gen}} \tag{3.4}$$

where we have not yet used any equation of motion.

In this way, given that the variations commute with each other, we obtain:

$$[\partial_n, \delta_\lambda]\hat{I}_n|_{n=1} = \delta_\lambda\langle K\rangle - \delta_\lambda A_{\mathrm{gen}} - \int \delta_\lambda E \partial_n g|_{n=1} + \int \partial_n E_n|_{n=1}\delta_\lambda g \,. \tag{3.5}$$

We have derived this equation by assuming that there is some action, but this equation should be a true equation independently of how we derive it. Note that, to derive it, we did not need to use the background equations of motion since they cancel in the double variation.



JHEP01(2018)081

This gives a gravitational entanglement first law, in a very similar to Wald's first law [38]. In both cases one derives the first law by varying the Lagrangian. In Wald's case, the first law is a consequence of having a Killing vector: the conservation of diffeomorphism current relates the difference between the area in the extremal surface and the energy at infinity with the gravitational constraints, integrated in a Cauchy slice in the entanglement wedge. In our case, we do a $\partial_n$ variation, which is less symmetric and we obtain that the two boundary terms differ by a codimension 0 integral. In this way, under the assumptions that the entanglement entropy is given by the generalized area and that the background equations of motion are satisfied close to $n = 1$, we have derived the following equation:

$$\delta_\lambda S - \delta_\lambda K = \int \delta_\lambda E \tilde{g} \tag{3.6}$$

with $\tilde{g} = \partial_n g$, but the equation is true even if we do not know what $\tilde{g}$ is. In this case, $\delta E$ is integrated over the whole manifold. Since we have less symmetries that in Rindler (where there is a Killing vector), the integral is higher dimensional, but it does not seem possible to do better from the first law.

From the assumptions that the background metric to satisfy the background equations of motion at leading order in $n-1$, the standard bulk-boundary dictionary and that the entropy is given by the area, we have deduced that $\delta S = \delta K \iff \int \delta E \tilde{g} = 0$. Since this is true for an arbitrary entangling surface, this probably implies $\delta E = 0$ everywhere. In principle, the linearized equations around an arbitrary background could be integrated to give the nonlinear equations of motion. However, given that the leading order in $(n-1)$ background equation of motion is a necessary assumption for this discussion, one might need to assume the background equations of motion for all $n$ to derive the nonlinear Einstein equations.[6]

Note also that this expression for the modular Hamiltonian is compatible with [24]. In fact, for Einstein gravity, we can think of $\delta A = \int_{RT} \gamma^{\alpha\beta} \delta g_{\alpha\beta}$ and express $\delta g = \int_{M_\infty} dx G(X,x) T(x)$. This gives an expression for $\delta K$ from which we can read $\langle K T_{\mu\nu} \rangle$ in holographic theories (similar comments were made in [14, 39]). The reason why this is only true given the equations of motion is because in order to write the metric operator in terms of the boundary fields one imposes the linearized equations of motion for the graviton. The good thing about the euclidean prescription described above is that it provides a bulk definition for the modular Hamiltonian which is independent of the area through the asymptotic one point functions at $n \sim 1$.

## 4 Quantum corrections to entanglement entropy

In the presence of quantum corrections, we will have a path integral in the replicated space $M_n$. The presence of quantum corrections will modify the equations of motion to all orders in $G_N$, we are going to denote the backreacted background metric by $g_{cl,n}$ and will expand it in $G_N$: $g_{cl,n} = g^{(0)}_{cl,n} + G_N g^{(1)}_{cl,n} + \cdots$.[7] As in [9], we assume that the background metric $g_{cl,n}$ is $\mathbb{Z}_n$ symmetric.

[6] This is because the equivalent of the first law for the modular entropies would give us the linearized equation of motion at arbitrary $n$ from which the nonlinear equation can be obtained.

[7] By the label classical, we mean that it is not a fluctuating but rather a background field. $g_{cl,n}$ will contain $G_N$ corrections due to the backreaction of the quantum fields.



JHEP01(2018)081

We are going to define the "orbifolded" partition function by dividing by $n$:

$$-\log Z_n = I_{\rm grav}(g_{cl,n}) - \log Z_n^{\rm matter}(g_{cl,n}), \quad \hat{I}_n[g_{cl,n}] = -\frac{1}{n}\log Z\,. \tag{4.1}$$

Let us review the discussion of [21], where they describe how to think about $\log Z_n$, $\partial_n \hat{I}_n$ at non-integer $n$. In the previous classical discussion, because of the $\mathbb{Z}_n$ symmetry of the background, the calculation of the action only needed the metric in the quotient space, however the quantum partition function is only defined in the parent space.[8] We can exploit the $\mathbb{Z}_n$ symmetry of the background metric, to write the partition function as:

$$\log Z_n = \log {\rm Tr}\rho_n^n \tag{4.2}$$

where the gravitational density matrix $\rho_n$ is defined by the boundary condition that the background metric $g_{cl,n}$ has a conical singularity of strength $1/n$. By taking $n$ powers of this seemingly singular density matrix, one ends up with a geometry which does not have a conical singularity. Given that $\rho_n$ is defined for arbitrary $n$, one can analytically continue (4.2) to real $n$: it is just the $n$-th power of $\rho_n$. In this way, we can express the derivative of $\hat{I}_n$ as the sum of the derivatives with respect to the lower and upper arguments of ${\rm Tr}\rho_n^n$:

$$\partial_n \hat{I}_n = -\partial_{\delta n}\log {\rm Tr}\rho_{n+\delta n}^n - \partial_{\delta n}\frac{1}{n+\delta n}\log {\rm Tr}\rho_n^{n+\delta n} = n^{-2}\langle S_{\rm Wald}\rangle_n + \tilde{S}_{n,{\rm bulk}}\,. \tag{4.3}$$

These first term is obtained by taking a derivative with respect to the background metric inside the path integral and using the expectation value of the equations of motion as in [21] (but to all orders in $G_N$). To exploit the semiclassical part of the problem (which allowed us to use the $\rho_n$ notation), where we have a well defined background metric, one needs to work perturbatively in $G_N$ around a given saddle $g_{cl,n}^{(0)}$. This discussion only makes sense in the $G_N$ expansion. This formula is formally true for arbitrary $n$, however to get the corrections to the background metric $g_{cl,n}^{(k>0)}$ one needs to analytically continue the expectation value of the stress tensor, $\langle T\rangle_n$, to non-integer $n$.

We can take the $n \to 1$ limit:

$$S = \lim_{n\to 1^+}\left(n^{-2}\langle S_{\rm Wald}\rangle_n + \tilde{S}_{n,{\rm bulk}}\right) = \langle A_{\rm gen}\rangle + S_{\rm bulk} = S_{\rm gen}\,. \tag{4.4}$$

To one loop, this is the same as [21]. The notation is a little different. There, $\langle A_{\rm gen}\rangle$ was explicitly separated into two terms: one coming from the generalized area evaluated in the background metric $g_{cl}$ (which was denoted $\frac{\delta A}{4G_N}$) and a contribution coming from matter fields which couple with derivatives of the metric, $\langle S_{\text{wald-like}}\rangle$. This last term is easily illustrated with a scalar field with a term $\int R\phi^2$, where $S_{\text{wald-like}} = \int_{RT}\phi^2$. In this original notation, the expectation value of the area due to graviton fluctuations should be thought as included in $S_{\text{wald-like}}$.

This procedure is in principle well defined to all orders in $G_N$: $\log Z_n$ is a completely standard partition function, although equation (4.3) requires introducing a $r = \epsilon$ artificial

[8] This is just the statement that the background metric is a one point function, which is $\mathbb{Z}_n$ invariant, while higher correlators need the whole parent space.



JHEP01(2018)081

boundary in our gravitational background. This "brick wall" partition function has been discussed in detail in [40, 41].

More concretely, at integer $n$, the partition function is well defined and nothing special happens at the $\mathbb{Z}_n$ symmetric fixed point. In order to take the $n$ derivative, it is convenient to define the partition function with a boundary at $r = \epsilon$. We want to do this in a way that we recover the original partition function when $\epsilon \to 0$. This is achieved by choosing a set of boundary conditions for the quantum fields at $r = \epsilon$ and then integrating independently over all possible boundary conditions. This integration is often referred to as summing over edge modes [40, 41], there they write the partition in a smooth black hole background for abelian gauge fields in terms of the partition function in a brick wall geometry summed over all possible electric fluxes across the boundary. Of course, after setting up these boundary conditions to define the partition function in the presence of a boundary, the entropy ($n$ derivative) will also have the same boundary conditions and edge modes. We can think of these edge modes as the center variables of [42]. We expect this story to generalize straightforwardly to gravity, see [24] for a discussion about gauge invariant boundary conditions for free gravitons.

### 4.1 Variations

In order to take variations with respect to the background metric, we have to define our partition function slightly off-shell. We can do this by adding a background stress tensor which couples with the metric operator: $\int dx^d \sqrt{g} T^{\text{bkg}}_{\mu\nu} h^{\mu\nu}$, with $h^{\mu\nu} = g^{\mu\nu} - g^{\mu\nu}_{cl,n}$, the background subtracted metric, it is hopefully clear from context that $h, g$ denote operators while $g_{cl}$ is a c-number. This term in the Lagrangian naturally splits the metric operators into the background metric, $g_{cl,n}$ and background subtracted fluctuation, which we will denote by $h$.[9] Derivatives with respect to the background stress tensor generate then background subtracted metric correlations. The role of the background stress tensor is to turn on-shell an arbitrary background metric[10] which allows us to think of the partition function as a function of the background metric.

At integer $n$, we will consider the variation of $\hat{I}$ with respect to the background metric:

$$\delta \hat{I}_n|_{T_{\text{bkg}}=0} = \int_{\hat{M}_n} (E_n(g_{cl,n}) + \langle T \rangle_n) \delta g + \int_{\partial \hat{M}_n} \Theta(g_{cl,n}, \delta g)|_{r=\epsilon} = 0 \tag{4.5}$$

where we used the quantum corrected equations of motion and the results from the previous section. Since this equation is valid for arbitrary $n$, its $n$ derivative will be zero. The boundary term appears when $g_{cl,n}$ has to be integrated by parts and it should be thought as including an expectation value with respect to the fluctuating fields, but we omitted it to simplify the notation.

[9] To each order in $G_N$, we can think of the Einstein equation as simply the tadpole equation for the metric operator: $g_{cl,n} = \langle g \rangle_n$.

[10] We can think of $T^{\text{bkg}} = T^{\text{bkg}}[g_{cl}]$, since the equations of motion (tadpole equations) are $E(g_{cl}) - G_N \langle T \rangle = T^{\text{bkg}}$ and the l.h.s. defines $T^{\text{bkg}}[g_{cl}]$. Equivalently, we can do Legendre transformation and obtain the effective action, which is a function of the off-shell background metric.



JHEP01(2018)081

By turning a background stress tensor, we can also take variation of (4.3)

$$\delta\partial_n \hat{I}_n|_{T^{\rm bkg}=0} = n^{-2}\delta\langle S_{\rm Wald}\rangle_n + \delta\tilde{S}_{n,\rm bulk} + \int_{\hat{M}_n} (\delta E(g_{cl,n}) + \delta\langle T\rangle_n)\partial_n g\,. \tag{4.6}$$

As our variation would be off-shell at integer $n$, the last term will not cancel. However, if we consider a variation which is on-shell close at $n = 1$, a diffeomorphism, the variation of last term will be zero, so, asking for $\delta\partial_n \hat{I}_n = \partial_n \delta\hat{I}_n = 0$ implies that

$$\delta_{\rm diff}(\langle A_{\rm gen}\rangle + S_{\rm bulk}) = \delta_{\rm diff} S_{\rm gen} = 0\,. \tag{4.7}$$

This is the quantum extremality condition of [23]. To leading order in $G_N$, we will later show explicitly that this is true using the equations of motion at $n \sim 1$, but this approach is valid to higher orders in $G_N$. An example with finite backreaction would be that of the Polyakov action (see appendix B), but this example might be too simple, since its effective action is local.

#### $G_N$ perturbation theory, the stress tensor and gravitons

The previous discussion applied order by order in $G_N$ and here we will be more a little bit more explicit about how it is defined.

The Einstein equation is an operator equation, which means that:

$$\langle E(g)\rangle = G_N \langle T_{\rm matter}(g,\phi)\rangle\,. \tag{4.8}$$

We can expand the Einstein tensor in terms of $g = g_{cl} + h$ in $G_N$ and, to each order, we can basically think of the gravitons $h$ as interacting matter with an their effective stress tensor determined by the expectation value of the Einstein tensor, expanded around with $E(\langle g\rangle)$. In this way, we can write the $O(G_N^k)$ term in the previous equation as:

$$E_n^{\rm lin}(g_{cl}^{(k)}) + E(g_{cl}^{(j<k)})|_{O(G_N^k)} = [G_N\langle T_{\rm matter}(g_{cl},h,\phi)\rangle + \langle T_{\rm grav}(h,g_{cl})\rangle]_{O(G_N^k)} = G_N\langle T\rangle|_{O(G_N^k)} \tag{4.9}$$

where the first term in the l.h.s. is the linearized Einstein tensor and this equation determines $g_{cl,n}^{(k)}$ in terms of expectation values and $g_{cl,n}^{(j<k)}$ and can be thought as a tadpole contribution to $g_{cl,n}^{(k)}$. Note that $T_{\rm grav}$ is defined order by order in $G_N$ by expanding $\langle E_n(g)\rangle$. We schematically denote the r.h.s. as $\langle T\rangle$.

We can think of the equations of motion as a background field expansion of the action order by order in $G_N$ and consider the variation of the (effective) action with respect to the background metric. If we think about gravitons order by order, they are basically the same as complicated matter with an effective stress tensor determine by the previous equation.

### 4.2 The definition of quantum extremal surfaces

In the previous sections, we derived the quantum extremality condition. In this section, we will explore the quantum extremality equations. Note that, in order to have a non-trivial quantum extremal surface, there has to be some asymmetry between the inside and outside



JHEP01(2018)081

region, and, for the symmetric case of a sphere in the vacuum, there will not be corrections to the extremal surface.

In our framework, we will always have a well defined background metric $g_{cl}$ and interacting gravitons on top of it. We can think of the location of the entangling surface in similar terms: $X = X_0 + G_N X_1 + \cdots$, $X$ denotes the location of the surface to all orders.[11] For Einstein gravity (it generalizes trivially to higher derivatives but we are going to focus on Einstein for simplicity), the leading term corresponds to the location of the extremal surface

$$\frac{1}{4G_N}\mathcal{K}^I_{X_0}(g^{(0)}_{cl}; y) = 0 + O(G_N^0)\,, \tag{4.10}$$

where $\mathcal{K}$ is the extrinsic curvature of the surface at $X_0$ and it depends on the position on the RT surface $y$ and in the background metric, since it is codimension 2 surface, there are two normal directions which we denote by $I$. To leading order in $G_N$, we can write an equation for the quantum extremal surface using the results of [36, 43]. One can use perturbation theory to understand how the entropy changes by a small change in the subregion. As in the previous discussion, we are going to denote by $r = \epsilon$ the tubular region close to the entangling surface. Using their work, one can show that to first order in $G_N$:

$$\frac{1}{4G_N}\mathcal{K}^I_{X_0+X_1}(g_{cl}; y)\big|_{O(G_N^0)} = -\delta_{X^I} S_{\text{bulk}}(X_0) = -2\pi \lim_{\epsilon\to 0} \epsilon \langle T^{Ir}(r=\epsilon; y) K_0\rangle\,. \tag{4.11}$$

This is a linear equation for $X_1$, determined in terms of quantities evaluated at $X_0$ (the classical extremal surface) which are well defined. $T$ is the r.h.s. of (4.9) and it is evaluated $\epsilon$ away from the entangling surface. The finite contribution to the variation of the entanglement entropy comes from a divergent contribution of $\langle TK\rangle$. In general terms, we expect this object to diverge when the stress tensor approaches the boundary of the region and the leading divergence goes like $\frac{1}{\epsilon^{d-2}}$. All the contributions that give a divergent variation of the entropy will correspond to the renormalization of the gravitational couplings, and should disappear after adding the proper counterterms. So, only the divergent contribution $\langle TK\rangle \propto \frac{1}{\epsilon}$ will contribute. If the background has a Killing vector, this correlator will not have an odd divergent term. The higher orders can be obtained from solving the exact equation

$$\frac{1}{4G_N}\mathcal{K}^I_X(g_{cl}; y) = -2\pi \lim_{\epsilon\to 0} \epsilon \langle T^{Ir}(r=\epsilon; y) K^X\rangle \tag{4.12}$$

where $K^X$ is the modular Hamiltonian of the bulk surface $X$. This equation can be expanded in $X$ order by order in $G_N$. Of course, $\mathcal{K}$ should be thought as an expectation value and (4.12) as a tadpole equation for $X$, for example to $O(G_N^0)$, we can think of adding an extra term in the r.h.s. $-\langle \mathcal{K}_{X_0}(h; y)\rangle$.

To leading order in $G_N$, we can also see how one would obtain the quantum extremality condition from the equations of motion around $n \sim 1$. The extremality of the area in RT is obtained by expanding the equations of motion near $n \sim 1, r \sim 0$ [9]. Schematically:

$$E_n = 0 \times (n-1)^0 + (n-1)\frac{\mathcal{K}^I(y)}{r} + O((n-1)r^0) + O((n-1)^2) \to \mathcal{K}^I(y) = 0\,, \tag{4.13}$$

[11] Note that there are no $G_N^{1/2}$ contributions since the entanglement entropy from gravitons is $O(G_N^0)$.



JHEP01(2018)081

that is, extremality is derived from regularity of the metric close to the $\mathbb{Z}_n$ symmetric fixed point. In the presence of quantum matter, we will have:

$$E_n - G_N \langle T \rangle_n = 0 \times (n-1)^0 + (n-1) \left( \frac{\mathcal{K}(y)}{r} + 8\pi G_N \langle TK \rangle \right) + \cdots . \tag{4.14}$$

It is now clear that if there is a $1/r$ divergent term in $\langle TK \rangle$, regularity of the metric close to the $\mathbb{Z}_n$ symmetric fixed point will shift the surface to the quantum extremal. It is also clear from this equation that the stress tensor that appears in (4.12) is just the r.h.s. of Einstein equations.

### Subtleties with gravitons

It might not be completely clear how to evaluate the entanglement entropy in the quantum extremal surface for gravitons, or whether it is well defined (see [24] for a set of boundary conditions that works for extremal surfaces). We certainly expect $\log Z_n$ to be well defined to all orders in $G_N$ and $g_{cl,n}$ should also be well defined in the $G_N$ expansion. Upon the inclusion of a boundary and summing over the proper edge modes, we expect that (4.4), (4.12) makes sense order by order in $G_N$.

Of course, in order to make this more concrete one should understand better the entanglement entropy of gravitons. For free gravitons, we expect that one can apply the ideas of [40, 41] together with [24] to compute the entanglement entropy. Then, we expect that the interacting graviton can be treated in the same way, by considering the interaction in entanglement perturbation theory [36, 44, 45]. In the same way, we expect that the deformation of the surface away from extremality can be understood in similar terms. More explicitly, as long as the displacement is small, we will schematically have

$$\begin{aligned} S_{\text{bulk}} = \sum_m \int_{RT} dy_1 \int ds_1 \cdots \int_{RT} dy_m \int ds_m \delta X(y_1) \cdots \delta X(y_m) \times \\ \times \langle T_s(X_0, y_1) \cdots T_{s_m}(X_0, y_m) f(K_0) \rangle, \end{aligned} \tag{4.15}$$

with $T_s = e^{iK_0 s} T e^{-iK_0 s}$, the modular evolved stress tensor. That is, the bulk entanglement entropy in a neighboring surface will be a correlator of (modular evolved) stress tensors and some function of the modular Hamiltonian $K_0$ integrated several time over the extremal surface. So, in principle, we might only need the modular Hamiltonian in extremal surface to obtain the entanglement entropy in other surfaces. In this expression, part of the $G_N$ will come from $\delta X$, part from changing the background metric and part from the correlator: stress tensors and $K_0$, for example to $O(G_N)$ we will have $S_{\text{bulk}} = S_{\text{bulk,free}}(X_0) + S_{\text{bulk},G_N}(X_0) + \int_{RT} dy \delta X^I \langle T^{Ir}(X_0, y) K_0 \rangle + \int dx \delta g_{ab} \langle T^{ab}(x) K_0 \rangle$.

Alternatively, we could just define this graviton entanglement entropy in terms of the boundary replica trick. We expect the partition function in this smooth manifold to be perfectly well defined.

Note that quantum extremality relates the contributions from $\delta X$ of $S_{\text{bulk}}$ to the contribution from the area. We will discuss this more explicitly in the next section.



JHEP01(2018)081

### 4.3 Quantum extremality and mixtures

Up to here, we have discussed quantum extremality in terms of partition functions $\mathrm{Tr}\rho^n$ which have a well defined path integral preparation and correspond to a unique classical saddle. We would like to understand how to extend the previous methods to mixtures of states:

$$Z_n[\rho+\sigma] = \mathrm{Tr}(\rho+\sigma)^n = \sum_{A_k=\{\sigma,\rho\}} \mathrm{Tr}(A_1 A_2 A_3 \cdots A_n)\,. \tag{4.16}$$

Even if $Z_{\rho+\sigma,n}$ cannot be prepared in the Euclidean path integral, each of the terms in the r.h.s. of (4.16) can, so we can think of (4.16) as a sum of path integrals. So, $Z_{\rho+\sigma,n}$ is in principle well defined for integer $n$: we have an asymptotic circle with perimeter $2\pi n$ which is divided into $n$ slices and set boundary conditions in each of the slice determined by a configuration in the r.h.s. of (4.16). Because this definition is an $n-$dependent sum of path integrals, it seems hard to analytically continue in $n$.

At this point, it is useful to make a remark about mixtures of path integrals in general. In the effective action formalism that we described before, whenever we have a mixture of states, we want to fix the same background source across the different states. Since there is only one background source, there is only one corresponding classical value for the field, that is, one tadpole equation. Consider the example of the linear mixture of two density matrices: $\frac{1}{2}(\rho+\sigma)$. In the presence of the same background tensor $\frac{\delta}{\delta T^{\mathrm{bkg}}}(Z_\rho + Z_\sigma) = \langle g\rangle_\rho + \langle g\rangle_\sigma = 2g_{cl,\rho+\sigma}$, we can Legendre transform by adding the term $-\int dx^d T^{\mathrm{bkg}} g_{cl,\rho+\sigma}$ to the path integrals. This means that $\frac{\delta}{\delta g_{cl,\rho+\sigma}} Z = 0$ will give the sum of the equations of motion, the tadpole condition will be $\langle g - g_{cl,\rho+\sigma}\rangle_{\rho+\sigma} = 0$ which is not explicitly linear in $\rho+\sigma$ because it is expanded around a background. If $Z_\rho$ and $Z_\sigma$ share the same saddle to leading order in the saddle point expansion,[12] $g^{(0)}_{cl,\rho} = g^{(0)}_{cl,\sigma}$, then we can understand this formalism as adding a quantum mixture of states to the classical geometry and solving the sum of the equations of motion $E(g_{cl,\rho+\sigma}) = \langle T^{g_{cl,\rho+\sigma}}\rangle_{\rho+\sigma}$, which we can now compute in $G_N$ perturbation theory.[13] Note that even if the Einstein equations $\langle E(g)\rangle = \langle T_{\mathrm{matter}}\rangle$ are linear in the mixture, the expectation value of the tadpole is background dependent. This makes the linearity of Einstein equations hard to see if we write them around $g_{cl,\rho+\sigma}$, however it is clear that $g_{cl,\rho+\sigma} = \frac{1}{2}(g_{cl,\rho} + g_{cl,\sigma})$ (yet this is clear because $\langle g\rangle_\psi = g_{cl,\psi}$).

The previous discussion gives a prescription to extend our result to mixtures of states that have the same $O(G_N^0)$ value of the metric: $g^{(0)}_{cl,\rho} = g^{(0)}_{cl,\sigma}$. These two states share, to leading order in $G_N$, the same ($\mathbb{Z}_n$ symmetric) saddle for $Z_n$. We can think of the sum of path integrals in terms of a mixture of quantum states in the $g^{(0)}_{cl}$ geometry, satisfying the

[12] If $g^{(0)}_{cl,\rho} \neq g^{(0)}_{cl,\sigma}$, $g_{cl,\rho+\sigma}$ is not a saddle. While $g_{\rho+\sigma}$ appears when coupling of the two path integrals through a background stress tensor, it does not have a clear semiclassical interpretation and we will not be considering this situation.

[13] This gives a well-defined procedure to compute the partition functions. If the two states are macroscopically distinguishable, the gravitons $h = g - g_{cl,\rho+\sigma}$ would not have a well-controlled one-loop partition function. However, for $h = g - g^{(0)}_{cl} - G_N g^{(k>0)}_{cl,\rho+\sigma}$, this graviton is only slightly off-shell with respect to the path integral of $\rho$ or $\sigma$, so the difference is small and it has a well-defined partition function. Alternatively, we can compute these partition functions with respect to their on-shell background first and use linearity $g_{cl,\rho+\sigma} = g_{cl,\rho} + g_{cl,\sigma}$.



JHEP01(2018)081

equations of motion:

$$E(g_{cl,(\rho+\sigma)^n}) = \langle T^{g_{cl,(\rho+\sigma)^n}} \rangle_{(\rho+\sigma)^n} \tag{4.17}$$

where we think of the r.h.s. as a sum over partition functions and the superscript denote that we are expanding the gravitons around the $g_{cl,(\rho+\sigma)^n}$ background. It is key that we phrase the problem in terms of a unique geometry and not a mixture of them since this will allows us to analytically continue in $n$. To do this, we note that $g_{cl,(\rho+\sigma)^n}$ is $\mathbb{Z}_n$ symmetric, which allows us to think of $\mathrm{Tr}(\rho+\sigma)^n$ in terms of taking the $n$-th power of $(\rho+\sigma)_n$, where the subscript $n$ denotes that it has the metric determined by (4.17). Upon analytic continuation of the r.h.s. of (4.17), the previous gives a prescription to compute $Z_n(\rho+\sigma) = \mathrm{Tr}[(\rho+\sigma)_n]^n$ for non integer $n$. Given this, the discussion from the previous section follows and we get quantum extremality for mixtures:

$$\frac{1}{4G_N}\mathcal{K}^I_X(g_{cl,\rho+\sigma}; y) = -2\pi \lim_{\epsilon\to 0} \epsilon \langle T^{Ir}(r=\epsilon, y) K^X_{\rho+\sigma} \rangle_{\rho+\sigma} \,. \tag{4.18}$$

It is clear that we want to think of the $n=1$ solution as given by a unique geometry, $g_{cl,\rho+\sigma}$ where quantum states can be entangled. Note that the fact that at integer $n$ we have complicated sums of partition functions makes the quantum extremal surface nonlinear in the state, since it depends on the modular Hamiltonian of the mixture, ie $X_{\rho+\sigma} \neq X_\rho + X_\sigma$ because $g_{cl,(\rho+\sigma)^n} \neq g_{cl,\rho^n} + g_{cl,\sigma^n}$.

## 5 Modular extremality

A simple consequence of quantum extremality for mixtures is that we can compute the expectation value of modular Hamiltonians for states close to each other (same $g^{(0)}_{cl}$). The modular Hamiltonian is just the log of the density matrix:

$$\langle K_\sigma \rangle_\rho = -\langle \log \sigma \rangle_\rho = -\partial_n \langle \sigma^{n-1} \rho \rangle \,. \tag{5.1}$$

We can get this from a mixture $\sigma + \lambda\rho$, since $\partial_\lambda \mathrm{Tr}(\sigma+\lambda\rho)^n|_{\lambda=0} = \mathrm{Tr}\rho\sigma^{n-1}$.

In this way, if we combine this with quantum extremality for mixtures,[14] we get a formula for the dual to the modular Hamiltonian:

$$\langle K_\sigma \rangle_\rho = \frac{\langle A^{X_\sigma} \rangle_\rho}{4G_N} + \langle K^{X_\sigma}_{\mathrm{bulk},\sigma} \rangle_\rho, \quad \frac{\delta}{\delta X_\sigma} \langle K_\sigma \rangle_\rho = 0 \,. \tag{5.2}$$

The boundary modular Hamiltonian is just given by the area plus the expectation value of the bulk modular Hamiltonian of $\sigma$ in the $\rho$ background. For simplicity of notation, we will illustrate this for Einstein gravity, but it generalizes trivially to higher derivatives. The surface where these terms are evaluated is determined by quantum extremality for the mixture, which implies that the sum of the two terms is extremized. We will call the $X_\sigma$

[14] For small perturbations, one can actually derive modular extremality for $\langle K_\sigma \rangle_{\sigma+\delta\sigma}$ in terms of quantum extremality plus the first law for $S(\sigma+\delta\sigma) = \langle K_{\sigma+\delta\sigma} \rangle_{\sigma+\delta\sigma}$.



JHEP01(2018)081

surface *modular extremal*. The variation can be carried out [45]:[15]

$$\frac{1}{4G_N}\mathcal{K}^I_{X_\sigma(\rho)}(g_{cl},y) = -\frac{\delta}{\delta X^I_\sigma}\langle K^{X_\sigma}_{\text{bulk},\sigma}\rangle_\rho$$
$$= \lim_{\epsilon\to 0}\epsilon\left[\langle :T^{Ir}(r=\epsilon;y):_\rho K^{X_\sigma}_{\text{bulk},\sigma}\rangle_\rho + \int_{-\infty}^{\infty}\frac{ds}{4\sinh^2(s/2+i\varepsilon)}\langle :T^{Ir}_s(r=\epsilon;y):_\sigma\rangle_\rho\right] \quad (5.3)$$

where $:T:_\rho = T - \langle T\rangle_\rho$ and $:T_s:_\sigma \equiv \exp(iK^{X_\sigma}_{\text{bulk},\sigma}s):T:_\sigma \exp(-iK^{X_\sigma}_{\text{bulk},\sigma}s)$. As we discussed before, one should also add an expectation value of the extrinsic curvature for the gravitons in the r.h.s. but we omitted it for simplicity. The finite contribution arises from a $1/\epsilon$ divergence in the first term, as in quantum extremality, and the second term can in principle get finite contributions from the $s$ integral (for a local modular Hamiltonian there are contributions from $s\sim -\log\epsilon$ that make this finite). We can think of the first term of the variation of $\rho$ and the second term the variation of $\log\sigma$. If $\rho=\sigma$, then the second term does not contribute, since it is proportional to the one point function of the stress tensor and we recover quantum extremality for a single state.

To leading order in $G_N$, $X_\sigma(\rho)$ is just the classical extremal surface and this is the bulk expression for the modular Hamiltonian discussed in [24]. In that paper, it was also discussed what the dual of the relative entropy is to leading order in $G_N$ and our result generalizes it to higher orders:

$$S_{\text{rel}}(\rho|\sigma) \equiv \langle K_\sigma\rangle_\rho - \langle K_\rho\rangle_\rho = \langle\frac{A^{X_\sigma}}{4G_N} - \frac{A^{X_\rho}}{4G_N} + K^{X_\sigma}_{\text{bulk},\sigma} - K^{X_\rho}_{\text{bulk},\rho}\rangle_\rho\,. \quad (5.4)$$

Given that the surfaces where the modular Hamiltonians are evaluated are different, the relative entropy does not have a simple description. Its difference from the bulk relative entropy can be understood as coming from the difference in areas localized $O(G_N)$ away from the classical extremal surface. From our point of view, the object which has a natural bulk description is the modular Hamiltonians, since it has a well defined path integral.

### 5.1 A linear mapping of surfaces

From the point of view of the path integral at integer $n$, $\text{Tr}\rho\sigma^{n-1}$, it is clear that our expression should be linear in $\rho$. At $n=1$, this is the statement that we should be thinking of the position of the modular extremal surface $X_\sigma(\rho)$ as linear function of the state $\rho$.

In this way, given a state $\sigma$ and its quantum extremal surface $X_\sigma(\sigma)$, we can think of $X_\sigma(\rho)$ as a mapping from the quantum extremal surface in the $\sigma$ background to a surface in the $\rho$ background (this is similar to [26], where some unspecified mapping was proposed).

The $G_N$ corrections generalize the extremal area operator appearing in [24] to the $\sigma$-dependent modular area operator: $A^{X_\sigma}$ depends on the modular Hamiltonian of $\sigma$. Since our equation can be understood as the expectation value of an operator in the state $\rho$, we can write is as an operator equation:

$$K_\sigma = \frac{A^{X_\sigma}}{4G_N} + K^{X_\sigma}_{\text{bulk},\sigma}\,. \quad (5.5)$$

[15] Note the first term in this expression is only discussed in their appendix since in they focus on the full modular Hamiltonian, where such a term is not present.



JHEP01(2018)081

**Linearity and state dependent divergences**

In principle, one could worry about the fact that (5.3) is not linear in $\rho$ because of

$$\lim_{\epsilon\to 0}\epsilon\left[\langle T^{Ir}(r=\epsilon,y)K_{\text{bulk},\sigma}\rangle_\rho-\langle T^{Ir}(r=\epsilon,y)\rangle_\rho\langle K_{\text{bulk},\sigma}\rangle_\rho\right]. \tag{5.6}$$

Note however that in the second term, the divergent contribution has to come from $K_{\text{bulk},\sigma}$, since there is nothing special happening at $r=\epsilon$ in the original state. Now, if this divergent contribution from $K_{\text{bulk},\sigma}$ was state independent, $\langle K^X_{\text{bulk},\sigma}\rangle_\rho=\frac{c(X)}{\epsilon}$ and thus we recover a linear expression.[16] $\langle K_{\text{bulk},\sigma}\rangle_\rho$ could in principle have state dependent divergences. State dependent divergences in the entropy were studied in [46], and they look like $\langle\int_{\partial R}\mathcal{O}\rangle_\rho$, which using the first law they can be mapped to a contribution to the modular Hamiltonian $\int_{\partial R}\mathcal{O}$ [24], which will lead to state dependent divergences in the modular Hamiltonian. However, because our contribution to the entropy includes $\langle A_{\text{gen}}\rangle$, $S_{\text{gen}}$ will not have these divergences. In other words, $K_{\text{bulk},\sigma}+A_{\text{gen}}$ does not have state dependent bulk divergences and we can just shift the possible term from $K_{\text{bulk},\sigma}$ to $A_{\text{gen}}$ in a way that none of the terms will have state dependent divergences and we get a clearly linear (5.6).

### 5.2 Modular extremality and the $G_N$ expansion

One should think of the $G_N$ expansion of $\dfrac{A^{X_\sigma(\rho)}}{4G_N}+\langle K^{X_\sigma}_{\text{bulk},\sigma}\rangle$ as expanding around the classical extremal surface $\mathcal{K}^I_{X_{\text{ext}}}(g_{cl},\rho)=0$. In terms of $\delta X_\sigma(\rho)\equiv X_\sigma-X_{\text{ext}}$, we can expand the area:

$$A^{X_\sigma(\rho)}=A^{X_{\text{ext}}(\rho)}+\int dy f^{(1)}(g_{cl},y)\delta X_\sigma(y)^2+\int dy f^{(2)}(g_{cl},y)\delta X_\sigma(y)^3+\cdots. \tag{5.7}$$

And also the bulk modular Hamiltonian:

$$\langle K^{X_\sigma}_{\text{bulk},\sigma}\rangle_\rho=\langle K^{X_{\text{ext}}}_{\text{bulk},\sigma}\rangle_\rho+\int dy F^{(1)}_{K_\sigma,\rho}[y]\delta X_\sigma(y)+\int dydy' F^{(2)}_{K_\sigma,\rho}[y,y']\delta X_\sigma(y)\delta X_\sigma(y')+\cdots \tag{5.8}$$

where the $F$'s are determined using modular perturbation theory, for example $F^{(1)}[y]$ is the r.h.s. of (5.3). Modular extremality relates the two terms, schematically: $2f^{(1)}(y)\delta X_\sigma+3f^{(2)}(y)\delta X_\sigma(y)^2+\cdots=-G_N(F^{(1)}[y]+2\int dy' F^{(2)}[y,y']\delta X_\sigma[y']+\cdots)$, but there are no miraculous cancellations because the terms which are the same order in $G_N$ in the area and bulk modular Hamiltonian have different powers of $\delta X_\sigma$. Modular extremality does simplify the expression for the boundary modular Hamiltonian since it can be expressed purely in terms of only $\delta X_\sigma$ and $f$ (provided that we know $\delta X_\sigma$). Of course, this expansion also applies for quantum extremal surfaces (when $\rho=\sigma$).

From this expansion, one could require the relative entropy to be given by the bulk relative entropy of some surface, which is neither the modular nor quantum extremal surface. We could set up an equation

$$S_{\text{rel}}(\rho|\sigma)\equiv\langle K_\sigma\rangle_\rho-\langle K_\rho\rangle_\rho=\langle A^{X_\sigma}-A^{X_\rho}+K^{X_\sigma}_{\text{bulk},\sigma}-K^{X_\rho}_{\text{bulk},\rho}\rangle_\rho=\langle K^{X_S(\rho,\sigma)}_{\text{bulk},\sigma}-K^{X_S(\rho,\sigma)}_{\text{bulk},\rho}\rangle_\rho \tag{5.9}$$

[16] By state independent we mean in the smaller Hilbert space of bulk low energy excitations. This contribution would depend on the semiclassical background, since the RT surface and the metric will be different.



JHEP01(2018)081

which should be solved order by order in $G_N$ by expanding the r.h.s. using (5.8) and solving for $X_S(\rho, \sigma)$. While it is clear that this can be done to leading order, we are not completely sure if there it has a solution to all orders. If that is true, it might be helpful to think about the interpretation of modular extremality: it relates variations of the area with variations of the modular Hamiltonian and this can be used to write the relative entropy as the bulk relative entropy in some $X_S$ surface. However, even if it is the case, it is clear that $X_S$ will be complicated and nonlinear in $\rho, \sigma$.

### 5.3 $\langle K_{\text{bulk},\sigma}\rangle_\rho$ and local modular Hamiltonians

At this point, even if we have a formal definition for this modular extremal surfaces, it would be nice to understand better what the different terms mean.

To compute $\langle K_{\text{bulk},\sigma}\rangle_\rho$, in gravity in $G_N$ perturbation theory, we have to account for three facts: the surface changes, the background metric changes, the quantum state changes. Only the latter is present in usual field theories. As we discussed before, the fact that the surface changes can be understood in terms of entanglement perturbation theory (and can be combined with the change in the area), and we are going to ignore this dependence in this section. Given that the background metric changes, we should think of the change in the state as a combination of a change in the matter fields plus a shift in the metric due to backreaction. We could deal with by deforming the path integral inserting an operator that changes the metric and this would give us a deformed modular Hamiltonian, as for shape deformations. However, given that the theory is gravitational there seems to be a more natural way to do it: we should think of the bulk modular Hamiltonian in terms of the $G_N$ expansion, to leading order it will be quadratic on the fields and then interactions will be present at higher orders. Backreaction is easily introduces by just shifting the tadpole $g_{cl}$ which appears in the modular Hamiltonian, that is $K_{\text{bulk},\sigma}[g_{cl,\rho}, h_\rho] \equiv K_{\text{bulk},\sigma}[g_{cl,\sigma}, h_\sigma - (g_\rho - g_\sigma)]$. This is just a shift of the variables, but the different expressions are useful when evaluated in the respective $g_{cl}$ state.

As an example, we can consider $K_\sigma$, the modular Hamiltonian of a sphere $R$ in the vacuum and $\rho$ some state which varies by an $O(1)$ expectation value of the boundary stress tensor. In this case, the modular Hamiltonian is local:

$$\langle K_\sigma\rangle_\rho = \langle \int_R \xi.T\rangle_\rho \,. \tag{5.10}$$

When we have local modular Hamiltonian, we can use Wald's version of Gauss' law [30, 47] (see also [24]):

$$\langle \int_R \xi.T\rangle_{\rho=\sigma+\delta\rho} - \langle \int_R \xi.T\rangle_\sigma = \mathcal{E}_\infty(\delta g) = \int_{\Sigma_S} \xi^t_{\text{bulk}} E_{tt,\text{lin}}(\delta g) + A^S_{\text{lin}}(\delta g) \tag{5.11}$$

where $S$ is an arbitrary gauge-invariant surface that is well defined for the original and the perturbed state (for example by picking a gauge where the surface stays at the same position). $\Sigma_S$ is the spacelike surface between the boundary region $R$ and the surface $S$.

Now, we can use (5.11) to integrate in $\langle K_\sigma\rangle_\rho$ for $\rho$ perturbatively close to $\sigma$, to all orders in perturbation theory. The reason is simple, if we write $g_\rho = g_\sigma + \sum_k \lambda^k \delta^k g$, we



JHEP01(2018)081

have that $\langle \int_R \xi.T\rangle_\rho - \langle \int_R \xi.T\rangle_\sigma = \sum_k \lambda^k \mathcal{E}_\infty(\delta^k g)$ is linear in the metric (and $\rho$) and we can use the gravitational Gauss' law for each term individually.

Now, $E_{tt,\text{lin}}(\delta^k g)$ is nothing but the tadpole of equation (4.9) (technically, (4.9) referred to the $G_N$ expansion, but it of course applies to any other perturbative expansion) which we can morally think of as the stress tensor to that order. So, we can write the previous formula as:

$$\langle K_\sigma\rangle_\rho - \langle K_\sigma\rangle_\sigma = \sum_k \lambda^k \int_{\Sigma_S} \xi.T^{(k)}_{\text{grav}} + A^S_{\text{lin}}(\delta^k g)\,. \tag{5.12}$$

We expect that this can be used to write $K_{\text{bdy}} = K^S_{\text{bulk}} + A^S$ for an arbitrary gauge invariant surface $S$, but this requires a careful analysis of boundary terms which we will not pursue further.[17] This means that modular extremality is not very helpful for local modular Hamiltonians. As the surface $S$, the most natural candidates are classical extremal or modular extremal surfaces, but one could choose any other families of gauge invariant surfaces. It is clear from this discussion that we should think of the change in background in $K_{\text{bulk},\sigma}$ as simply shifting the tadpole from $g_\sigma$ to $g_\rho$.

Now, we would to connect the previous story with that of [48]. We can think of their setup in our terms as $\rho$ being a bulk coherent state on top of $\sigma$ with a semiclassical amplitude, schematically $|\Psi_\rho\rangle = e^{i\sqrt{\lambda/G_N}a^\dagger}|\Psi_\sigma\rangle$, with $a^\dagger$ the graviton creation operator. We can to work in the limit where the amplitude is large (so that the state is classical) but the states only change the metric perturbatively in $\lambda$. Since $g_\rho, g_\sigma$ correspond to the same saddle, we can apply our discussion. In this limit, even if in the entanglement entropy the area changes to order $G_N^{-1}$, the bulk entanglement entropy stays $O(G_N^0)$, so we do not need quantum extremality. It is less clear if the modular extremal surface changes for coherent states, but we do not need it because of (5.12). We can instead consider the simpler case when $S$ is the extremal surface. In this case, since the bulk entanglement entropy is $O(G_N^0)$, but the bulk modular Hamiltonian is $O(G_N^{-1})$, we deduce that:

$$S_{\text{rel}}(\rho|\sigma) = \Delta\langle K^{S_{\text{ext}}}_{\text{bulk}}\rangle + O(G_N^0) \tag{5.13}$$

where we used our expectation that $K_{\text{bdy}} = K^S_{\text{bulk}} + A^S$ and for $S$ being the extremal surface the areas cancel in the relative entropy, they would not cancel for modular extremal surfaces. In this way, it is very suggestive to think of the Hamiltonian of [48] as the bulk modular Hamiltonian in the entanglement wedge, in which case the positivity of relative entropy would be a consequence of the positivity of the bulk relative entropy. Again, modular extremality does not seem important in their case because in this symmetric situation, one can choose an arbitrary gauge invariant surface where to integrate the boundary modular Hamiltonian. Of course, to make full connection between (5.12), modular extremality and [48] more precise, one should understand better how the boundary terms and $E_{\text{lin}}$ combine to give the bulk modular Hamiltonian to all orders.

[17] Although naively only the linearized area operator appears, the r.h.s. of Einstein equations ($T_{\text{grav}}$) is the bulk modular Hamiltonian modulo boundary terms which turn the linearized area operator into the full area operator. One can see how this works to second order by carefully rewriting $T_{\text{grav}}$ as the canonical energy (bulk modular Hamiltonian) plus the quadratic area operator [24].



JHEP01(2018)081

More broadly, understanding if (classical) coherent states give an $O(1)$ shift to the position extremal surface when considering modular extremality seems interesting, since quantum extremal surfaces can only shift the entangling surface by $O(G_N)$. This might give a simpler classical setup to compute the dual of the modular Hamiltonian. For example, if we consider a coherent state of scalar fields, where $\langle\phi\rangle_\lambda = 0 + \sqrt{\lambda G_N^{-1}}\phi_{cl}$, we expect the modular extremal surface to shift by a classical $\delta X^I(X,y) \propto \lambda^2 \int \frac{ds}{\sinh^2(s/2+i\varepsilon)} T_s^{Ir}$ when computing $\langle K_{\rm bdy}(\lambda=0)\rangle_\lambda$ holographically. Of course, this is hard to do explicitly, because we have little control over modular Hamiltonians other than those which are local, where we can apply (5.12).

## 6 Discussion

In this paper, we have exploited the variational principle at the level of the replicated path integral to derive the extremality of the entangling functional of higher derivatives, quantum extremality and modular extremality. This is done by thinking about the Rényi entropies and taking the $n \to 1$ limit carefully. This gives closure to the approach of [9] which naturally gives the entanglement entropy functional but makes it hard to derive the extremality condition for general gravitational theories and higher orders in $G_N$. This variational framework is also useful to generalize relation between the equations of motion and the first law for general states.

We would like to close with some comments and future directions.

As a general note, across this paper, we have assumed that the bulk saddles have replica symmetry. It would be nice if one could relax this or justify it better (see [49, 50] for some discussion about this).

### Higher-derivative gravity

By working at integer $n > 1$ and then taking the $n \to 1$ limit in higher-derivative theories of gravity, we have discussed how one should in principle determine the splitting terms of [32]. These are determined by demanding that the gravitational action is finite. After fixing these terms, the only remaining freedom comes from changing the location of the surface, and this deformation keeps the action finite.

In appendix A, we have demonstrated in some nontrivial examples how the $n \to 1$ limit of the Wald entropy at $n > 1$ gives the gravitational entropy of [12, 13]. While our approach is strongly suggestive that this is true generally, it would be useful to work it out explicitly for more general examples.

### The equations of motion

About the equations of motion, it would be nice to understand better if by varying the regions that in consideration, one can derive the local equations of motion from the integrated equations of motion. Note that, in contrast with [17], the equations are integrated over one more dimension because of the lack of symmetry.

In order to derive the equations of motion from the first law of entanglement of [14, 15], one has to understand the modular Hamiltonian. In general it is complicated, yet its



JHEP01(2018)081

variations are well defined in terms of analytically continued one point functions in the replicated theory. We expect that, in the absence of a more explicit expression for the boundary Hamiltonian, the only way in which one can obtain the equations of motion from the first law is by using the replica trick via the procedure described in section 3.

Of course, there are other ways in which one could try to get the equations of motion from the RT formula. An alternative option pursued by [51, 52] is to show that the boundary expression for the relative entropy around the vacuum for a spherical region matches the expression for bulk relative entropy. The bulk and boundary relative entropies differ off-shell by an integral of the equations of motion and thus one can derive the backreacted equations of motion from the equality of these two quantities. More generally, one might be able to use similar ideas to the ones that we described combined with modular perturbation theory to generalize this approach to other surfaces and states.

**Entanglement entropy of gravitons**

We defined the entanglement entropy of gravitons by analytically continuing the finite $n-1$ partition function. Technically speaking, only $S_{\text{gen}}$ is well defined, since the separation into two terms is ambiguous: it depends on the details of how the boundary is inserted. This ambiguity is related with the choice of center of [42]. It would be nice to understand better the graviton entanglement entropy from a Hilbert space perspective, along the lines of [24, 40, 41]. It would be interesting to carry out the perturbation theory described in section 4.2 to define the entanglement entropy of gravitons beyond the extremal surfaces in $G_N$ perturbation theory.

**Local modular Hamiltonians and modular extremality**

We have also given an argument of how one can in principle think of the results of [48] in terms of bulk relative entropy. Of course, it would be nice to understand this more precisely, by being careful about the boundary terms in the graviton modular Hamiltonian to higher orders.

Modular extremality does not seem necessary when the modular Hamiltonian is local, since there we can just use Gauss' law to integrate in the energy at infinity. It seems hard yet very interesting to understand explicitly some examples of modular extremality for modular Hamiltonians which are non-local. In contrast with quantum extremality, we expect the modular extremal surface to be different from the extremal surface for deformations which are classical (coherent states).

**Modular flow and bulk reconstruction**

To leading order in $G_N$, the commutator between a properly dressed local operator at a point $Z$ in the entanglement wedge and the modular Hamiltonian is given by the commutator with the bulk modular Hamiltonian. This was used in [26] to show that one can reconstruct operators in the entanglement wedge in terms of the boundary subregion and more recently, it was used in [27] to derive a boundary expression of the bulk operators.

Furthermore, [26] showed that if $\rho_{\text{bulk}} = \sigma_{\text{bulk}} \to \rho = \sigma$, which is clearly true from (1.8), then one can also reconstruct operators deep inside the entanglement wedge. As has been



JHEP01(2018)081

argued recently [53], the analysis of [25, 26] is stable under $G_N$ perturbations and we expect that our discussion can help find the explicit bulk to boundary mapping in the presence of backreaction. Because of the previous, we do not expect the approach of [27] to break down when $G_N$ corrections are considered. To next order, it seems like the correction to the difference between modular flows is determined by the shift in the surface:

$$[K_\sigma, \Phi(Z)] = [K_{\text{bulk},\sigma}, \Phi(Z)] + G_N \int_{RT} dy [\delta X(y)^2, \Phi(Z)] \,. \tag{6.1}$$

We leave for future work understanding this contribution to the commutator, but we expect that by carefully understanding the previous one can generalize [27] to higher orders in $G_N$.


## Acknowledgments

We would like to thank Eric D'Hoker, Tom Faulkner, Daniel Harlow, Nima Lashkari, Juan Maldacena, Onkar Parrikar, Mukund Rangamani, Douglas Stanford, and Aron Wall for useful discussions. X.D. was supported in part by the National Science Foundation under Grant No. PHY-1606531, by the Department of Energy under Grant No. DE-SC0009988 and, by the Martin A. and Helen Chooljian Founders' Circle Membership at the Institute for Advanced Study. A.L. acknowledges support from the Simons Foundation through the It from Qubit collaboration, as well as the support of a Myhrvold-Havranek Innovative Thinking Fellowship. A.L. would also like to thank the Department of Physics and Astronomy at the University of Pennsylvania for hospitality during the development of this work.


## A Dilaton gravity with higher derivative interactions

In this appendix, we study the gravitational entropy in toy models of higher derivative gravity: 2d dilaton gravity coupled to matter fields with higher derivative interactions. These theories can arise from dimensional reduction of higher derivative gravity in more than two dimensions. We demonstrate how to solve the "splitting problem" and calculate the entropy functional $A_{\text{gen}}$ in these toy models. Furthermore, we verify (1.3) and (2.14) by showing directly from the equations of motion that the entropy is obtained by extremizing $A_{\text{gen}}$, and its extremal value agrees with the $n \to 1$ limit of the Wald entropy.

Throughout this appendix, we define $\epsilon \equiv n-1$ and adopt a complex coordinate system $(z, \bar{z})$ on $M_n$ such that the metric is in the conformal gauge

$$ds^2 = e^{2\psi(z,\bar{z})} dz d\bar{z} \tag{A.1}$$

and the origin is the $\mathbb{Z}_n$ fixed point. The $\mathbb{Z}_n$ symmetry acts as a discrete rotation $z \to z e^{2\pi i/n}$.

We will study solutions of the equations of motion for $\psi$, the dilaton $\phi$, and additional matter fields. At $n = 1$, these fields have regular Taylor expansions around $z = 0$. For example, we have

$$\phi(z,\bar{z})\Big|_{n=1} = \mathring{\phi} + \mathring{\phi}_z z + \mathring{\phi}_{\bar{z}} \bar{z} + \frac{1}{2}\mathring{\phi}_{zz} z^2 + \frac{1}{2}\mathring{\phi}_{\bar{z}\bar{z}} \bar{z}^2 + \mathring{\phi}_{z\bar{z}} z\bar{z} + \cdots \tag{A.2}$$

for the dilaton. Away from $n = 1$, such expansions become much more complicated. Near $n \approx 1$, we may generally expand the dilaton as

$$\phi(z,\bar{z}) = \phi_0 + \phi_1(z\bar{z})^\epsilon + \phi_2(z\bar{z})^{2\epsilon} + \cdots + \left\{z^{1+\epsilon}\left[\phi_{z,0} + \phi_{z,1}(z\bar{z})^\epsilon + \cdots\right] + \text{c.c.}\right\} \tag{A.3}$$
$$+ \left\{\frac{1}{2}z^{2(1+\epsilon)}\left[\phi_{zz,0} + \phi_{zz,1}(z\bar{z})^\epsilon + \cdots\right] + \text{c.c.}\right\} + z\bar{z}\left[\phi_{z\bar{z},0} + \phi_{z\bar{z},1}(z\bar{z})^\epsilon + \cdots\right] + \cdots$$

and similarly for other fields. Here c.c. denotes complex conjugate. As we go away from $n = 1$, each term in the expansion (A.2) "splits" into a Taylor expansion in $(z\bar{z})^\epsilon$. Continuity at $n = 1$ therefore requires the following matching conditions:

$$\mathring{\phi} = \phi_0 + \phi_1 + \phi_2 + \cdots, \tag{A.4}$$
$$\mathring{\phi}_\mu = \phi_{\mu,0} + \phi_{\mu,1} + \phi_{\mu,2} + \cdots, \tag{A.5}$$
$$\mathring{\phi}_{\mu\nu} = \phi_{\mu\nu,0} + \phi_{\mu\nu,1} + \phi_{\mu\nu,2} + \cdots, \tag{A.6}$$

and their higher-order analogues. Here $\mu = z, \bar{z}$, and we have only kept zeroth-order terms in $\epsilon$ in coefficients such as $\phi_m$ and $\phi_{\mu,m}$. Higher-order terms in $\epsilon$ are negligible for the purpose of calculating the von Neumann entropy in our examples.

The gravitational entropy $A_{\text{gen}}$ can be calculated as in [12], but the result would depend on how the $n = 1$ coefficients split into $n \neq 1$ coefficients in (A.4)–(A.6). On the other hand, $A_{\text{gen}}$ should depend only on the $n = 1$ solution (A.2) in order to be a useful entropy functional. This is the "splitting problem." As we will demonstrate explicitly below, the solution to this problem is that the equations of motion near $n \approx 1$ are sufficient to fix the split of coefficients in (A.4)–(A.6), at least to the extent of allowing us to write $A_{\text{gen}}$ in terms of the $n = 1$ coefficients appearing in (A.2).

### A.1 One matter field

Let us first consider the following theory of dilaton gravity coupled with a single scalar field $\sigma$ with higher derivative interaction:

$$I = -\frac{1}{2}\int d^2x\sqrt{g}\left[\phi R - (\nabla\sigma)^2 + \lambda\nabla_\mu\nabla_\nu\sigma\nabla^\mu\nabla^\nu\sigma\right]. \tag{A.7}$$

The equation of motion for the metric is

$$\frac{1}{\sqrt{g}}\frac{\delta I}{\delta g^{\mu\nu}} = g_{\mu\nu}\left[-\frac{1}{2}\nabla^2\phi - \frac{1}{4}(\nabla\sigma)^2 + \frac{\lambda}{4}\nabla_\rho\nabla_\sigma\sigma\nabla^\rho\nabla^\sigma\sigma\right] + \frac{1}{2}\nabla_\mu\nabla_\nu\phi + \frac{1}{2}\nabla_\mu\sigma\nabla_\nu\sigma$$
$$+ \frac{\lambda}{2}\left(\nabla_\mu\sigma\nabla^2\nabla_\nu\sigma + \nabla_\nu\sigma\nabla^2\nabla_\mu\sigma - \nabla^2\sigma\nabla_\mu\nabla_\nu\sigma - \nabla_\rho\sigma\nabla^\rho\nabla_\mu\nabla_\nu\sigma\right) = 0, \tag{A.8}$$

whereas the equations of motion for the dilaton $\phi$ and the scalar $\sigma$ are

$$-\frac{1}{\sqrt{g}}\frac{\delta I}{\delta\phi} = \frac{1}{2}R = 0, \tag{A.9}$$
$$-\frac{1}{\sqrt{g}}\frac{\delta I}{\delta\sigma} = \nabla^2\sigma + \lambda\nabla_\mu\nabla_\nu\nabla^\mu\nabla^\nu\sigma = 0. \tag{A.10}$$



JHEP01(2018)081

Using (A.9) we find a flat space with the conformal factor $\psi = 0$, greatly simplifying the other equations. If we want, we could get an AdS solution instead by replacing $R$ with $R + 2$ in (A.7); this leads to $\psi = -\log\left(1 - \frac{z\bar{z}}{4}\right)$ but our conclusion is largely unaffected.

Solving the other equations of motion near $n \approx 1$, we find

$$\sigma_{m>0} = 0, \qquad \sigma_{\mu,m>0} = 0, \qquad \sigma_{\mu\nu,m>0} = 0, \tag{A.11}$$

$$\phi_1 = 2\lambda\sigma_{z,0}\sigma_{\bar{z},0}, \qquad \phi_{m>1} = 0, \tag{A.12}$$

$$\phi_{z,0} = 2\lambda\sigma_{z,0}\sigma_{z\bar{z},0}, \qquad \phi_{z,1} = 2\lambda\sigma_{\bar{z},0}\sigma_{zz,0}, \qquad \phi_{z,m>1} = 0. \tag{A.13}$$

This holds for arbitrary $\lambda$ and constrains how the coefficients split in (A.4)–(A.6):

$$\sigma_0 = \mathring{\sigma}, \quad \sigma_{\mu,0} = \mathring{\sigma}_\mu, \quad \sigma_{\mu\nu,0} = \mathring{\sigma}_{\mu\nu}, \quad \phi_0 = \mathring{\phi} - 2\lambda\mathring{\sigma}_z\mathring{\sigma}_{\bar{z}}, \tag{A.14}$$

$$\mathring{\phi}_z = 2\lambda\left(\mathring{\sigma}_z\mathring{\sigma}_{z\bar{z}} + \mathring{\sigma}_{\bar{z}}\mathring{\sigma}_{zz}\right). \tag{A.15}$$

Let us make two comments before continuing. First, these relations are uniquely determined from a local analysis of the equations of motion near a small conical defect in the quotient space $\hat{M}_n$, and are universal in the sense that they do not depend on whatever boundary conditions we impose at the asymptotic boundary of spacetime. The reason for this is that these relations arise from setting to zero the most singular terms in the equations of motion expanded around $z = 0$. This is a good feature because the entropy functional $A_{\text{gen}}$ should only depend on local geometric quantities once we fix the gravitational action. Our second comment is that the split of $\mathring{\sigma}_z$ (and $\mathring{\sigma}_{\bar{z}}$) is over-constrained as shown in (A.15), but we will see that this is a feature not a bug.

The gravitational entropy can be easily calculated as in [12]:

$$A_{\text{gen}} = 2\pi(\phi_0 + \phi_1) - 4\pi\lambda\sigma_{z,0}\sigma_{\bar{z},0}. \tag{A.16}$$

As promised, this entropy functional can be rewritten[18] in terms of fields and their derivatives at $n = 1$:

$$A_{\text{gen}} = S_{\text{Wald}} + S_{\text{anomaly}}, \quad S_{\text{Wald}} = 2\pi\mathring{\phi}, \quad S_{\text{anomaly}} = -4\pi\lambda\mathring{\sigma}_z\mathring{\sigma}_{\bar{z}}. \tag{A.17}$$

Moreover, it agrees with the $n \to 1^+$ limit of the Wald entropy

$$\lim_{n\to 1^+} S_{\text{Wald}}(g_n) = 2\pi\phi_0 \tag{A.18}$$

which is identical to (A.17) after using (A.14). It is worth noting that in taking the above limit we need to calculate the Wald entropy at $n > 1$, and $\phi_1$ does not contribute to this. Therefore, the Wald entropy has a discontinuity of the amount $2\pi\phi_1$ at $n = 1$, which is precisely compensated by $S_{\text{anomaly}}$ in (A.17).

We satisfy the extremality condition $\partial_\mu A_{\text{gen}} = 0$ because it reduces to

$$\partial_z A_{\text{gen}} = \partial_z(2\pi\mathring{\phi} - 4\pi\lambda\mathring{\sigma}_z\mathring{\sigma}_{\bar{z}}) = 2\pi\left[\mathring{\phi}_z - 2\lambda\left(\mathring{\sigma}_z\mathring{\sigma}_{z\bar{z}} + \mathring{\sigma}_{\bar{z}}\mathring{\sigma}_{zz}\right)\right] \tag{A.19}$$

which vanishes due to the extra constraint (A.15).

[18] This rewriting only uses the most singular part of the equations of motion expanded around $z = 0$, and is valid even after a $\delta g_n$ variation as long as it is regular as defined in section 2.2.



JHEP01(2018)081

### A.2 Two matter fields

The previous example may seem too simple for experts, so let us now study a more complicated theory of dilaton gravity coupled with two scalar fields $\sigma$ and $\omega$ with higher derivative interaction:

$$I = -\frac{1}{2}\int d^2x\sqrt{g}\left[\phi R - (\nabla\sigma)^2 - (\nabla\omega)^2 + \lambda\omega\nabla_\mu\nabla_\nu\sigma\nabla^\mu\nabla^\nu\sigma\right]. \tag{A.20}$$

The equation of motion for the metric is

$$\begin{aligned}\frac{1}{\sqrt{g}}\frac{\delta I}{\delta g^{\mu\nu}} = g_{\mu\nu}&\left[-\frac{1}{2}\nabla^2\phi - \frac{1}{4}(\nabla\sigma)^2 - \frac{1}{4}(\nabla\omega)^2 + \frac{\lambda}{4}\omega\nabla_\rho\nabla_\sigma\sigma\nabla^\rho\nabla^\sigma\sigma\right]\\ &+\frac{1}{2}\nabla_\mu\nabla_\nu\phi + \frac{1}{2}\nabla_\mu\sigma\nabla_\nu\sigma + \frac{1}{2}\nabla_\mu\omega\nabla_\nu\omega\\ &+\frac{\lambda}{2}\Big\{\left[\nabla_\mu\sigma\nabla^\rho(\omega\nabla_\rho\nabla_\nu\sigma) + (\mu\leftrightarrow\nu)\right] - \nabla^\rho(\omega\nabla_\rho\sigma\nabla_\mu\nabla_\nu\sigma)\Big\} = 0,\end{aligned} \tag{A.21}$$

whereas the equations of motion for the dilaton $\phi$ and the other scalars $\sigma$, $\omega$ are

$$-\frac{1}{\sqrt{g}}\frac{\delta I}{\delta\phi} = \frac{1}{2}R = 0, \tag{A.22}$$

$$-\frac{1}{\sqrt{g}}\frac{\delta I}{\delta\sigma} = \nabla^2\sigma + \lambda\nabla_\mu\nabla_\nu(\omega\nabla^\mu\nabla^\nu\sigma) = 0, \tag{A.23}$$

$$-\frac{1}{\sqrt{g}}\frac{\delta I}{\delta\omega} = \nabla^2\omega + \frac{\lambda}{2}\nabla_\mu\nabla_\nu\sigma\nabla^\mu\nabla^\nu\sigma = 0. \tag{A.24}$$

Again we find a flat space with the conformal factor $\psi = 0$.

It is difficult to solve the other equations of motion for arbitrary $\lambda$, so we will work perturbatively in $\lambda$ and write the solution as

$$\phi = \phi^{(0)} + \lambda\phi^{(1)} + \lambda^2\phi^{(2)} + \cdots \tag{A.25}$$

with similar expansions for other fields.

At the zeroth order in $\lambda$, we find the familiar case of dilaton gravity without any higher derivative interaction:

$$\phi^{(0)}_{m>0} = 0, \quad \phi^{(0)}_{\mu,m\geq 0} = 0, \tag{A.26}$$

$$\sigma^{(0)}_{m>0} = 0, \quad \sigma^{(0)}_{\mu,m>0} = 0, \quad \sigma^{(0)}_{z\bar{z},m\geq 0} = 0, \tag{A.27}$$

$$\omega^{(0)}_{m>0} = 0, \quad \omega^{(0)}_{\mu,m>0} = 0, \quad \omega^{(0)}_{z\bar{z},m\geq 0} = 0. \tag{A.28}$$

At the linear order in $\lambda$, we find

$$\phi^{(1)}_1 = 2\omega^{(0)}_0\sigma^{(0)}_{z,0}\sigma^{(0)}_{\bar{z},0}, \qquad \phi^{(1)}_{m>1} = 0, \tag{A.29}$$

$$\phi^{(1)}_{z,0} = 0, \qquad \phi^{(1)}_{z,1} = 2\sigma^{(0)}_{\bar{z},0}\left[\omega^{(0)}_0\sigma^{(0)}_{zz,0} + \omega^{(0)}_{z,0}\sigma^{(0)}_{z,0}\right], \ \phi^{(1)}_{z,m>1} = 0, \tag{A.30}$$

$$\sigma^{(1)}_1 = -\left[\omega^{(0)}_{z,0}\sigma^{(0)}_{\bar{z},0} + \omega^{(0)}_{\bar{z},0}\sigma^{(0)}_{z,0}\right], \qquad \sigma^{(1)}_{m>1} = 0, \tag{A.31}$$

$$\sigma^{(1)}_{z,1} = -\left[\omega^{(0)}_{zz,0}\sigma^{(0)}_{\bar{z},0} + \sigma^{(0)}_{zz,0}\omega^{(0)}_{\bar{z},0}\right], \ \sigma^{(1)}_{z,m>1} = 0, \tag{A.32}$$

$$\omega^{(1)}_1 = -\sigma^{(0)}_{z,0}\sigma^{(0)}_{\bar{z},0}, \qquad \omega^{(1)}_{m>1} = 0, \qquad \omega^{(1)}_{z,1} = -\sigma^{(0)}_{\bar{z},0}\sigma^{(0)}_{zz,0}, \quad \omega^{(1)}_{z,m>1} = 0. \tag{A.33}$$



JHEP01(2018)081

At the second order in $\lambda$, all we need to find is

$$\phi_1^{(2)} = 2\omega_0^{(0)}\left[\sigma_{z,0}^{(0)}\sigma_{\bar{z},0}^{(1)} + \text{c.c.}\right] - 2\omega_0^{(1)}\omega_1^{(1)}, \tag{A.34}$$

$$\phi_2^{(2)} = \frac{1}{2}\left[\sigma_1^{(1)}\right]^2 - \frac{3}{2}\left[\omega_1^{(1)}\right]^2 + 2\omega_0^{(0)}\left[\sigma_{z,0}^{(0)}\sigma_{\bar{z},1}^{(1)} + \text{c.c.}\right], \quad \phi_{m>2}^{(2)} = 0\,. \tag{A.35}$$

From these results we can determine the gravitational entropy as in [12]. Let us find the contribution to $A_{\text{gen}}$ from each term in the action (A.20). We will work to second order in $\lambda$. The contribution of the $\phi R$ term is

$$2\pi\left[\phi_0^{(0)} + \lambda\left(\phi_0^{(1)} + \phi_1^{(1)}\right) + \lambda^2\left(\phi_0^{(2)} + \phi_1^{(2)} + \phi_2^{(2)}\right)\right]. \tag{A.36}$$

From the $(\nabla\sigma)^2$ term we get

$$\pi\lambda^2\left[\sigma_1^{(1)}\right]^2, \tag{A.37}$$

whereas the contribution of the $(\nabla\omega)^2$ term is

$$\pi\lambda^2\left[\omega_1^{(1)}\right]^2. \tag{A.38}$$

From the $\lambda\omega\nabla_\mu\nabla_\nu\sigma\nabla^\mu\nabla^\nu\sigma$ term we get the contribution

$$\begin{aligned}&-4\pi\lambda\left[\omega_0^{(0)} + \lambda\left(\omega_0^{(1)} + \frac{1}{2}\omega_1^{(1)}\right)\right]\left[\sigma_{z,0}^{(0)} + \lambda\left(\sigma_{z,0}^{(1)} + \sigma_{z,1}^{(1)}\right)\right]\left[\sigma_{\bar{z},0}^{(0)} + \lambda\left(\sigma_{\bar{z},0}^{(1)} + \sigma_{\bar{z},1}^{(1)}\right)\right]\\ &+ 2\pi\lambda^2\sigma_1^{(1)}\left[\omega_{z,0}^{(0)}\sigma_{\bar{z},0}^{(0)} + \omega_{\bar{z},0}^{(0)}\sigma_{z,0}^{(0)}\right].\end{aligned} \tag{A.39}$$

Combining these four contributions we get the gravitational entropy

$$A_{\text{gen}} = A_{\text{gen}}^{(0)} + \lambda A_{\text{gen}}^{(1)} + \lambda^2 A_{\text{gen}}^{(2)} + \cdots \tag{A.40}$$

where

$$A_{\text{gen}}^{(0)} = 2\pi\phi_0^{(0)}, \tag{A.41}$$

$$A_{\text{gen}}^{(1)} = 2\pi\left[\phi_0^{(1)} + \phi_1^{(1)}\right] - 4\pi\omega_0^{(0)}\sigma_{z,0}^{(0)}\sigma_{\bar{z},0}^{(0)}, \tag{A.42}$$

$$\begin{aligned}A_{\text{gen}}^{(2)} = {}& 2\pi\left[\phi_0^{(2)} + \phi_1^{(2)} + \phi_2^{(2)}\right] + 3\pi\left[\omega_1^{(1)}\right]^2 - \pi\left[\sigma_1^{(1)}\right]^2\\ &+ 4\pi\omega_0^{(1)}\omega_1^{(1)} - 4\pi\omega_0^{(0)}\left[\sigma_{z,0}^{(0)}\left(\sigma_{\bar{z},0}^{(1)} + \sigma_{\bar{z},1}^{(1)}\right) + \text{c.c.}\right].\end{aligned} \tag{A.43}$$

Again, this entropy function can be rewritten in terms of fields and their derivatives at $n = 1$. To second order in $\lambda$ we find

$$A_{\text{gen}} = 2\pi\mathring{\phi} - 4\pi\lambda\mathring{\omega}\mathring{\sigma}_z\mathring{\sigma}_{\bar{z}} - \pi\lambda^2\left[\mathring{\sigma}_z^2\mathring{\sigma}_{\bar{z}}^2 + (\mathring{\omega}_z\mathring{\sigma}_{\bar{z}} + \mathring{\omega}_{\bar{z}}\mathring{\sigma}_z)^2\right] + O(\lambda^3). \tag{A.44}$$

This also agrees with the $n \to 1^+$ limit of the Wald entropy

$$\lim_{n\to 1^+} S_{\text{Wald}}(g_n) = 2\pi\left[\phi_0^{(0)} + \lambda\phi_0^{(1)} + \lambda^2\phi_0^{(2)}\right] + O(\lambda^3) \tag{A.45}$$

which can easily be shown to be identical to (A.40).

It is worth noting that if we forgot about splitting and proceeded naïvely, we would miss the $\lambda^2$ term in (A.44). Therefore, this example shows that we cannot in general forget about splitting in calculating the gravitational entropy.

It is possible to check the extremality condition $\partial_\mu A_{\text{gen}} = 0$ by working out the relevant part of the $zz$ component of (A.21) to second order in $\lambda$.



JHEP01(2018)081

## B Polyakov action

A toy model to understand these issues would be to consider 2d dilaton gravity in the presence of $m$ quantum scalar fields [54, 55]

$$I = \frac{1}{2\pi}\int dx^2\sqrt{g}\left[e^{-2\phi}(R+4(\partial\phi)^2+4\lambda^2)\right] - \frac{m\hbar}{96\pi}\int R\nabla^{-1}R\,. \tag{B.1}$$

In the limit of large $m$, one can analyze the theory at finite $N = m\hbar$. The second term can be thought of as $\int(\partial\eta)^2 - 2\eta R$, with $\nabla\eta = R$. This expression suggests that $S_{\rm Wald} = \frac{N}{12}\eta_0$. This might seem too quick, but [56] showed using the Noether charge methods that $S_{\rm Wald} = \frac{N}{12}\eta_0$, so that the total entropy is

$$S_{\rm total} = 2e^{-2\phi_0} + \frac{N}{12}\eta_0 = 2e^{-2\phi_0} + \frac{N}{6}\rho_0 \tag{B.2}$$

where $\eta_0$, which is non-local in general, was expressed in terms of the metric in conformal gauge, $ds^2 = e^{2\rho}dzd\bar{z}$. The quantum extremality condition would be $-4e^{-2\phi_0}\partial\phi_0 + \frac{N}{6}\partial\rho_0 = 0$.

The equations of motion are [55]:

$$0 = e^{-2\phi}(4\partial\rho\partial\phi + 2\partial^2\phi) - \frac{N}{12}(\partial\rho\partial\rho + \partial^2\rho) \tag{B.3}$$

and similarly for $\bar{\partial}$.

Now, the question is whether given some metric $\rho$, the equations of motion can be solved if one adds a small conical singularity $\delta_n\rho = (n-1)\log z\bar{z}$. If $N = 0$, then it was shown [9] that a $\delta_n\phi$ change cannot cancel the singularity of $\partial\delta\rho = \frac{(n-1)}{z}$, so one concludes that $\partial\phi = 0$.

In the presence of $N$, there will be two kind of terms linear in $\delta_n$:

$$\partial\delta_n\rho\left(4\partial\phi e^{-2\phi} - \frac{N}{6}\partial\rho\right) + \left[\delta_n(e^{-2\phi}\partial\phi)\partial\rho + 2\delta_n(e^{-2\phi}\partial^2\phi) - \frac{N}{12}\partial^2\delta_n\rho\right] = 0 \tag{B.4}$$

with $\delta_n\rho = (n-1)\log z\bar{z}$. If we consider $\phi = \rho = 0$, then the equation is solved by setting $\delta_n\phi = \frac{N}{24}\delta_n\rho$. For a non-trivial background, one can cancel the $\frac{n-1}{z^2}$ between brackets by picking an appropriate $\delta_n\phi$. This then results in the condition $(4\partial\phi e^{-2\phi} - \frac{N}{6}\partial\rho) = 0$ which is the quantum extremality condition.

Naively, it seems non trivial that one would get the quantum extremality condition because the gravitational action is non-local. However, in this particular case, after adding an extra field the action becomes local and thus the usual arguments apply.



JHEP01(2018)081

## References

[1] J.D. Bekenstein, *Black holes and entropy*, *Phys. Rev.* **D 7** (1973) 2333 [INSPIRE].

[2] J.M. Bardeen, B. Carter and S.W. Hawking, *The Four laws of black hole mechanics*, *Commun. Math. Phys.* **31** (1973) 161 [INSPIRE].

[3] S.W. Hawking, *Particle Creation by Black Holes*, *Commun. Math. Phys.* **43** (1975) 199 [*Erratum ibid.* **46** (1976) 206] [INSPIRE].

[4] J.M. Maldacena, *The large-N limit of superconformal field theories and supergravity*, *Int. J. Theor. Phys.* **38** (1999) 1113 [hep-th/9711200] [INSPIRE].

[5] S.S. Gubser, I.R. Klebanov and A.M. Polyakov, *Gauge theory correlators from noncritical string theory*, *Phys. Lett.* **B 428** (1998) 105 [hep-th/9802109] [INSPIRE].

[6] E. Witten, *Anti-de Sitter space and holography*, *Adv. Theor. Math. Phys.* **2** (1998) 253 [hep-th/9802150] [INSPIRE].

[7] S. Ryu and T. Takayanagi, *Holographic derivation of entanglement entropy from AdS/CFT*, *Phys. Rev. Lett.* **96** (2006) 181602 [hep-th/0603001] [INSPIRE].

[8] S. Ryu and T. Takayanagi, *Aspects of Holographic Entanglement Entropy*, *JHEP* **08** (2006) 045 [hep-th/0605073] [INSPIRE].

[9] A. Lewkowycz and J. Maldacena, *Generalized gravitational entropy*, *JHEP* **08** (2013) 090 [arXiv:1304.4926] [INSPIRE].

[10] V.E. Hubeny, M. Rangamani and T. Takayanagi, *A covariant holographic entanglement entropy proposal*, *JHEP* **07** (2007) 062 [arXiv:0705.0016] [INSPIRE].

[11] X. Dong, A. Lewkowycz and M. Rangamani, *Deriving covariant holographic entanglement*, *JHEP* **11** (2016) 028 [arXiv:1607.07506] [INSPIRE].

[12] X. Dong, *Holographic Entanglement Entropy for General Higher Derivative Gravity*, *JHEP* **01** (2014) 044 [arXiv:1310.5713] [INSPIRE].

[13] J. Camps, *Generalized entropy and higher derivative Gravity*, *JHEP* **03** (2014) 070 [arXiv:1310.6659] [INSPIRE].

[14] D.D. Blanco, H. Casini, L.-Y. Hung and R.C. Myers, *Relative Entropy and Holography*, *JHEP* **08** (2013) 060 [arXiv:1305.3182] [INSPIRE].

[15] G. Wong, I. Klich, L.A. Pando Zayas and D. Vaman, *Entanglement Temperature and Entanglement Entropy of Excited States*, *JHEP* **12** (2013) 020 [arXiv:1305.3291] [INSPIRE].

[16] N. Lashkari, M.B. McDermott and M. Van Raamsdonk, *Gravitational dynamics from entanglement 'thermodynamics'*, *JHEP* **04** (2014) 195 [arXiv:1308.3716] [INSPIRE].

[17] T. Faulkner, M. Guica, T. Hartman, R.C. Myers and M. Van Raamsdonk, *Gravitation from Entanglement in Holographic CFTs*, *JHEP* **03** (2014) 051 [arXiv:1312.7856] [INSPIRE].

[18] B. Czech, J.L. Karczmarek, F. Nogueira and M. Van Raamsdonk, *The Gravity Dual of a Density Matrix*, *Class. Quant. Grav.* **29** (2012) 155009 [arXiv:1204.1330] [INSPIRE].

[19] A.C. Wall, *Maximin Surfaces and the Strong Subadditivity of the Covariant Holographic Entanglement Entropy*, *Class. Quant. Grav.* **31** (2014) 225007 [arXiv:1211.3494] [INSPIRE].

JHEP01(2018)081

[20] M. Headrick, V.E. Hubeny, A. Lawrence and M. Rangamani, *Causality & holographic entanglement entropy*, *JHEP* **12** (2014) 162 [arXiv:1408.6300] [INSPIRE].

[21] T. Faulkner, A. Lewkowycz and J. Maldacena, *Quantum corrections to holographic entanglement entropy*, *JHEP* **11** (2013) 074 [arXiv:1307.2892] [INSPIRE].

[22] T. Barrella, X. Dong, S.A. Hartnoll and V.L. Martin, *Holographic entanglement beyond classical gravity*, *JHEP* **09** (2013) 109 [arXiv:1306.4682] [INSPIRE].

[23] N. Engelhardt and A.C. Wall, *Quantum Extremal Surfaces: Holographic Entanglement Entropy beyond the Classical Regime*, *JHEP* **01** (2015) 073 [arXiv:1408.3203] [INSPIRE].

[24] D.L. Jafferis, A. Lewkowycz, J. Maldacena and S.J. Suh, *Relative entropy equals bulk relative entropy*, *JHEP* **06** (2016) 004 [arXiv:1512.06431] [INSPIRE].

[25] A. Almheiri, X. Dong and D. Harlow, *Bulk Locality and Quantum Error Correction in AdS/CFT*, *JHEP* **04** (2015) 163 [arXiv:1411.7041] [INSPIRE].

[26] X. Dong, D. Harlow and A.C. Wall, *Reconstruction of Bulk Operators within the Entanglement Wedge in Gauge-Gravity Duality*, *Phys. Rev. Lett.* **117** (2016) 021601 [arXiv:1601.05416] [INSPIRE].

[27] T. Faulkner and A. Lewkowycz, *Bulk locality from modular flow*, *JHEP* **07** (2017) 151 [arXiv:1704.05464] [INSPIRE].

[28] D. Harlow, *The Ryu-Takayanagi Formula from Quantum Error Correction*, *Commun. Math. Phys.* **354** (2017) 865 [arXiv:1607.03901] [INSPIRE].

[29] X. Dong, *The Gravity Dual of Renyi Entropy*, *Nature Commun.* **7** (2016) 12472 [arXiv:1601.06788] [INSPIRE].

[30] V. Iyer and R.M. Wald, *Some properties of Noether charge and a proposal for dynamical black hole entropy*, *Phys. Rev.* **D 50** (1994) 846 [gr-qc/9403028] [INSPIRE].

[31] D.V. Fursaev, A. Patrushev and S.N. Solodukhin, *Distributional Geometry of Squashed Cones*, *Phys. Rev.* **D 88** (2013) 044054 [arXiv:1306.4000] [INSPIRE].

[32] R.-X. Miao and W.-z. Guo, *Holographic Entanglement Entropy for the Most General Higher Derivative Gravity*, *JHEP* **08** (2015) 031 [arXiv:1411.5579] [INSPIRE].

[33] R.-X. Miao, *Universal Terms of Entanglement Entropy for 6d CFTs*, *JHEP* **10** (2015) 049 [arXiv:1503.05538] [INSPIRE].

[34] J. Camps, *Gravity duals of boundary cones*, *JHEP* **09** (2016) 139 [arXiv:1605.08588] [INSPIRE].

[35] A. Castro, S. Detournay, N. Iqbal and E. Perlmutter, *Holographic entanglement entropy and gravitational anomalies*, *JHEP* **07** (2014) 114 [arXiv:1405.2792] [INSPIRE].

[36] V. Rosenhaus and M. Smolkin, *Entanglement Entropy: A Perturbative Calculation*, *JHEP* **12** (2014) 179 [arXiv:1403.3733] [INSPIRE].

[37] A. Lewkowycz and E. Perlmutter, *Universality in the geometric dependence of Renyi entropy*, *JHEP* **01** (2015) 080 [arXiv:1407.8171] [INSPIRE].

[38] V. Iyer and R.M. Wald, *A Comparison of Noether charge and Euclidean methods for computing the entropy of stationary black holes*, *Phys. Rev.* **D 52** (1995) 4430 [gr-qc/9503052] [INSPIRE].

[39] D.L. Jafferis and S.J. Suh, *The Gravity Duals of Modular Hamiltonians*, *JHEP* **09** (2016) 068 [arXiv:1412.8465] [INSPIRE].



JHEP01(2018)081

[40] W. Donnelly and A.C. Wall, *Entanglement entropy of electromagnetic edge modes*, *Phys. Rev. Lett.* **114** (2015) 111603 [arXiv:1412.1895] [INSPIRE].

[41] W. Donnelly and A.C. Wall, *Geometric entropy and edge modes of the electromagnetic field*, *Phys. Rev.* **D 94** (2016) 104053 [arXiv:1506.05792] [INSPIRE].

[42] H. Casini, M. Huerta and J.A. Rosabal, *Remarks on entanglement entropy for gauge fields*, *Phys. Rev.* **D 89** (2014) 085012 [arXiv:1312.1183] [INSPIRE].

[43] A. Allais and M. Mezei, *Some results on the shape dependence of entanglement and Rényi entropies*, *Phys. Rev.* **D 91** (2015) 046002 [arXiv:1407.7249] [INSPIRE].

[44] T. Faulkner, R.G. Leigh and O. Parrikar, *Shape Dependence of Entanglement Entropy in Conformal Field Theories*, *JHEP* **04** (2016) 088 [arXiv:1511.05179] [INSPIRE].

[45] T. Faulkner, R.G. Leigh, O. Parrikar and H. Wang, *Modular Hamiltonians for Deformed Half-Spaces and the Averaged Null Energy Condition*, *JHEP* **09** (2016) 038 [arXiv:1605.08072] [INSPIRE].

[46] D. Marolf and A.C. Wall, *State-Dependent Divergences in the Entanglement Entropy*, *JHEP* **10** (2016) 109 [arXiv:1607.01246] [INSPIRE].

[47] R.M. Wald, *Black hole entropy is the Noether charge*, *Phys. Rev.* **D 48** (1993) R3427 [gr-qc/9307038] [INSPIRE].

[48] N. Lashkari, J. Lin, H. Ooguri, B. Stoica and M. Van Raamsdonk, *Gravitational positive energy theorems from information inequalities*, *PTEP* **2016** (2016) 12C109 [arXiv:1605.01075] [INSPIRE].

[49] T. Faulkner, *The Entanglement Renyi Entropies of Disjoint Intervals in AdS/CFT*, arXiv:1303.7221 [INSPIRE].

[50] J. Camps and W.R. Kelly, *Generalized gravitational entropy without replica symmetry*, *JHEP* **03** (2015) 061 [arXiv:1412.4093] [INSPIRE].

[51] T. Faulkner, *Bulk Emergence and the RG Flow of Entanglement Entropy*, *JHEP* **05** (2015) 033 [arXiv:1412.5648] [INSPIRE].

[52] T. Faulkner, F.M. Haehl, E. Hijano, O. Parrikar, C. Rabideau and M. Van Raamsdonk, *Nonlinear Gravity from Entanglement in Conformal Field Theories*, *JHEP* **08** (2017) 057 [arXiv:1705.03026] [INSPIRE].

[53] J. Cotler, P. Hayden, G. Salton, B. Swingle and M. Walter, *Entanglement Wedge Reconstruction via Universal Recovery Channels*, arXiv:1704.05839 [INSPIRE].

[54] C.G. Callan Jr., S.B. Giddings, J.A. Harvey and A. Strominger, *Evanescent black holes*, *Phys. Rev.* **D 45** (1992) R1005 [hep-th/9111056] [INSPIRE].

[55] J.G. Russo, L. Susskind and L. Thorlacius, *The Endpoint of Hawking radiation*, *Phys. Rev.* **D 46** (1992) 3444 [hep-th/9206070] [INSPIRE].

[56] R.C. Myers, *Black hole entropy in two-dimensions*, *Phys. Rev.* **D 50** (1994) 6412 [hep-th/9405162] [INSPIRE].

JHEP01(2018)081