Gravitational Path Integrals

David Grabovsky

February 20, 2023

Abstract

This set of lecture notes is based on the first half of a quarter-long lecture course (PHYS 231C) given by Prof. Don Marolf at UCSB in the spring of 2021. It is intended as a pedagogical introduction to path integrals in QM and gravity; some aspects of classical GR; Euclidean techniques in quantum gravity; and the replica trick.

Contents

0 Introduction

Introduction to be written last.

Pre-requisites: classical mechanics (Lagrangian and Hamiltonian mechanics, and some ideas from Hamilton-Jacobi theory), quantum mechanics (path integrals to be reviewed), quantum statistical mechanics (in terms of density matrices), some QFT, and definitely GR (though many topics will be reviewed).

Conventions: natural units, but we keep G and sometimes use $\hbar$ as a heuristic. The natural numbers start at zero, and the natural logarithm is ln. Fourier transforms are unitary, i.e. they have factors of $\frac{1}{\sqrt{2\pi}}$. The spacetime dimension is d, so the spatial dimension is $d-1$. Bulk spacetime indices are early Latin, $abcd$, and boundary spacetime indices are middle Latin, $ijk\ell$. We occasionally use Greek indices $\alpha\beta\gamma\delta$ and $\mu\nu\rho\sigma$ for some calculations.

1 Path Integrals

1.1 The Canonical Path Integral

Our story begins with single-particle quantum mechanics in one dimension. The Hilbert space is $\mathcal{H} = L^2(\mathbb{R})$, and the system is governed by a self-adjoint Hamiltonian operator $\hat{H}$. We work in the Heisenberg picture, where observables $\hat{\mathcal{O}}$ evolve in time according to

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\mathcal{O}}(t) = i[\hat{H}, \hat{\mathcal{O}}]. \tag{1.1}$$

Now consider a one-parameter family of self-adjoint operators $\hat{q}(t)$ and $\hat{p}(t)$, the position and momentum, which have canonical equal-time commutator $[\hat{q}(t), \hat{p}(t)] = i$. The Hamiltonian is allowed to depend on both $\hat{q}$ and $\hat{p}$, as well as explicitly on time: $\hat{H} = \hat{H}(\hat{q}(t), \hat{p}(t), t)$. Note that for $t < t'$, the commutator $[\hat{q}(t), \hat{p}(t')]$ can only be obtained from the canonical one after solving the equation of motion: in other words, the commutator depends on $\hat{H}$.

The path integral. The operator $\hat{q}(t)$ has a complete set of eigenstates $\{|q, t\rangle\}$. These states are assumed to be orthonormal, in the sense that $\langle q', t|q, t\rangle = \delta(q - q')$ at equal times. This does not tell us what the overlap $K(q_i, t_i;\, q_f, t_f) \equiv \langle q_f, t_f|q_i, t_i\rangle$ is for $t_i < t_f$, but surely this quantity, the *propagator*, depends on how $\hat{H}$ evolves the system between times t_i and t_f. To construct the path integral that computes the propagator, we will proceed in four steps: (1) formally solve (1.1) in the case $\hat{\mathcal{O}}(t) = \hat{q}(t)$, and thereby relate the $\hat{q}$-eigenstates at times t_i and t_f; (2) slice up the interval $[t_i, t_f]$ into a large number N of infinitesimal time steps, and resolve the identity at each step; (3) resolve a tricky ordering ambiguity in the Hamiltonian; and (4) take the continuum limit to obtain the path integral.

Step 1. The formal solution to the "equation of motion" (1.1) is given by conjugating $\hat{\mathcal{O}}(t_i)$ by the unitary time evolution operator $U(t_i, t_f)$. This operator is constructed from the Hamiltonian by means of the *Dyson series*, which generalizes the familiar operator $e^{-i\hat{H}t}$ to the case of an arbitrary time-dependent Hamiltonian. For $\hat{\mathcal{O}} = \hat{q}$, we have

$$\begin{aligned}\frac{\mathrm{d}\hat{q}}{\mathrm{d}t} = i[\hat{H}(t), \hat{q}(t)] \implies \hat{q}(t_f) = U(t_i, t_f)\,\hat{q}(t_i)\,U(t_i, t_f)^{-1} = U(t_i, t_f)\,\hat{q}(t_i)\,U(t_f, t_i) = \\ = \mathcal{P}\exp\left[i\int_{t_i}^{t_f}\mathrm{d}t\,\hat{H}(t)\right]\hat{q}(t_i)\,\mathcal{P}\exp\left[i\int_{t_f}^{t_i}\mathrm{d}t\,\hat{H}(t)\right].\end{aligned} \tag{1.2}$$

Here $\mathcal{P}$ is the path ordering symbol. Notice that inverting U is tantamount to reversing the path, so as to evolve backwards in time. This solution allows us to show that the eigenstates $|q, t_i\rangle$ are related to $|q, t_f\rangle$ by time evolution. The phase of the state simply rotates:

$$|q, t_f\rangle = U(t_i, t_f)\,|q, t_i\rangle = \mathcal{P}\exp\left[i\int_{t_i}^{t_f}\mathrm{d}t\,\hat{H}(t)\right]|q, t_i\rangle\,. \tag{1.3}$$

This is straightforward to check by applying $\hat{q}(t_f)$ to the proposed form of $|q, t_f\rangle$:

$$\hat{q}(t_f)\,|q,t_f\rangle = U(t_i,t_f)\,\hat{q}(t_i)\,\cancel{U(t_i,t_f)^{-1}U(t_i,t_f)}\,|q,t_i\rangle = q\,U(t_i,t_f)\,|q,t_i\rangle = q\,|q,t_f\rangle\,. \quad \checkmark \tag{1.4}$$

Step 2. We embark on the journey of computing the propagator:

$$K(q_i,t_i;\,q_f,t_f) \equiv \langle q_f,t_f|q_i,t_i\rangle = \langle q_f,t_i|\,\mathcal{P}\exp\left[i\int_{t_f}^{t_i}\mathrm{d}t\,\hat{H}(t)\right]|q_i,t_i\rangle\,. \tag{1.5}$$

Here the path has been reversed because we are time-evolving the dual vector $\langle q_f, t_f|$. The path-ordered exponential is defined as the following product:

$$K(q_i,t_i;\,q_f,t_f) = \langle q_f,t_i|U(t_f,t_i)|q_f,t_i\rangle = \lim_{\Delta t\to 0}\langle q_f,t_i|e^{-i\hat{H}(t_i)\Delta t}\cdots e^{-i\hat{H}(t_f)\Delta t}|q_i,t_i\rangle\,. \tag{1.6}$$

The operator $U(t_f,t_i)$ appearing here is a string of infinitesimal time evolution operators proceeding from t_f to t_i in steps of size Δt. Thus the rightmost $\hat{H}$ is evaluated at t_f, the one before is evaluated at $t_f - \Delta t$, and so on until the leftmost $\hat{H}$ is evaluated at t_i. The minus signs above are responsible for undoing the backwards evolution of each step.

Next, we denote $q_N \equiv q_f$ and $q_0 \equiv q_i$, and insert $N-1$ position-space resolutions of the identity between each exponential factor, where $N = \frac{|t_f - t_i|}{\Delta t}$ is the number of time steps:

$$\begin{aligned} K(q_i,t_i;\,q_f,t_f) = \lim_{N\to\infty}\int \mathrm{d}q_1\cdots\mathrm{d}q_{N-1}\;\langle q_N,t|\,e^{-i\hat{H}(t)\Delta t}\,|q_{N-1},t\rangle\,\langle q_{N-1},t|\cdots \\ \cdots|q_1,t\rangle\,\langle q_1,t|\,e^{-i\hat{H}(t')\Delta t}\,|q_0,t\rangle\,. \end{aligned} \tag{1.7}$$

If we study just one of the factors above, we might try to work to first order in Δt:

$$K_1 \equiv \langle q_1,t|\,e^{-i\hat{H}(t_f)\Delta t}\,|q_0,t\rangle = \langle q_1,t|\left(1 - i\hat{H}(t_f)\Delta t + O(\Delta t^2)\right)|q_0,t\rangle\,. \tag{1.8}$$

But position-space matrix elements of $\hat{H}$ are difficult to calculate in general: it is much easier to find matrix elements of $\hat{H}$ sandwiched between position and momentum states. This motivates the insertion of yet another identity, this time in momentum space:

$$K_1 = \int \mathrm{d}p_1\,\langle q_1,t|p_1,t\rangle\,\langle p_1,t|\left(1 - i\hat{H}(t_f)\Delta t + O(\Delta t^2)\right)|q_0,t\rangle\,. \tag{1.9}$$

At this point, we will assume that $\hat{H}$ does not explicitly depend on time: $\hat{H}(t) = \hat{H}(t_f) = \hat{H}$. (In fact, this restriction can be loosened with some more work.)

Step 3. The problem with defining $\hat{H} = H(\hat{q}(t), \hat{p}(t))$ as a "function" of the $\hat{q}$ and $\hat{p}$ operators is that there are ordering ambiguities due to noncommutativity. For instance, $\hat{q}\hat{p} = \hat{p}\hat{q} + i \neq \hat{p}\hat{q}$. We will choose to define $\hat{H}$ so that all $\hat{p}$'s are moved to the left and all

$\hat{q}$'s are moved to the right. This convention is called *Weyl ordering*, and we define the *Weyl symbol* $H_W(q,p)$ to be the following possibly complex-valued c-number function:

$$H_W(q,p) = \frac{\langle p,t|\hat{H}|q,t\rangle}{\langle p,t|q,t\rangle}. \tag{1.10}$$

We will actually take H_W to mean the real part of of the above, since the imaginary part is of order $\hbar$ and vanishes in the semiclassical limit. So in fact $H_W(q,p)$ is real.

The numerator of H_W is exactly what appears in the matrix element above:

$$K_1 = \int \mathrm{d}p_1 \, \langle q_1,t|p_1,t\rangle \, \langle p_1,t|q_0,t\rangle \Big(1 - iH_W(q_1,p_1)\Delta t + O(\Delta t^2)\Big). \tag{1.11}$$

Now, we can evaluate the integrand. The inner product wave functions $\langle q_1,t|p_1,t\rangle$ and $\langle p_1,t|q_0,t\rangle$ are plane waves, and (to first order) the $1 - iH_W\Delta t$ term is also an exponential:

$$\begin{aligned} K_1 &= \frac{1}{2\pi}\int \mathrm{d}p_1 \, e^{ip_1q_1}e^{-ip_1q_0}e^{-iH_W(q_1,p_1)\Delta t} = \frac{1}{2\pi}\int \mathrm{d}p_1 \exp\Big(i\big[p(q_1-q_0) - H_W(q_1,p_1)\Delta t\big]\Big) = \\ &= \frac{1}{2\pi}\int \mathrm{d}p_1 \exp\Big(i\big[p_1\dot{q}_1 - H_W(q_1,p_1)\big]\Delta t\Big). \end{aligned} \tag{1.12}$$

The factor of $\frac{1}{2\pi}$ comes from the $\frac{1}{\sqrt{2\pi}}$ normalization on each of the two plane waves. We have multiplied and divided by Δt and introduced $\dot{q}_1 \equiv \frac{q_1-q_0}{\Delta t}$. We recognize the term in brackets as the contribution to the (canonical) action on a small time interval $[t_0,t_1]$ of length Δt.

Step 4. We can now go back to the full propagator (1.7) and take the continuum limit $\Delta t \longrightarrow 0$ (or $N \longrightarrow \infty$). the result is a product of expressions like the one above:

$$\begin{aligned} K(q_i,t_i;\,q_f,t_f) &= \lim_{N\to\infty}\int \mathrm{d}q_1\cdots q_{N-1}\, K_1\cdots K_N = \\ &= \lim_{N\to\infty}\left(\frac{1}{2\pi}\right)^N \int \mathrm{d}q_1\cdots\mathrm{d}q_{N-1}\,\mathrm{d}p_1\cdots\mathrm{d}p_N \exp\left(i\sum_{k=1}^{N}\big[p_k\dot{q}_k - H_W(q_k,p_k)\big]\Delta t\right) \equiv \\ &\equiv \int\limits_{q(t_i)=q_i}^{q(t_f)=q_f} \mathcal{D}q\,\mathcal{D}p\,\exp\left(i\int_{t_i}^{t_f}\mathrm{d}t\,\big[p\dot{q} - H_W(q,p)\big]\right) = \int\limits_{q(t_i)=q_i}^{q(t_f)=q_f} \mathcal{D}q\,\mathcal{D}p\,e^{iS_{\mathrm{L}}[q,p]}. \end{aligned} \tag{1.13}$$

The infinite-dimensional measure $\mathcal{D}q\,\mathcal{D}p$ on the set of phase-space trajectories is a formal expression, and is not rigorously defined. It instructs us to sum over "all" paths taken by a particle between positions q_i and q_f during the time interval $[t_i,t_f]$. The contribution of a given path to the propagator is weighted by the *Lorentzian action*

$$S_{\mathrm{L}}[q,p] = \int_{t_i}^{t_f}\mathrm{d}t\,\Big(p\dot{q} - H_W(q,p)\Big). \tag{1.14}$$

1.2 Comments and Extensions

The measure. In many cases, $S_{\rm L}$ is quadratic in the momenta. In this case the path integral over p is Gaussian and can be performed exactly. This is not always true: more generally, if one actually manages to perform the path integral over momenta exactly, one will be left with a "measure factor" $\mu[q]$. To be precise, we write

$$K(q_i,t_i;\,q_f,t_f) = \int_{q(t_i)=q_i}^{q(t_f)=q_f} \mathcal{D}q\,\mu[q]\,e^{iS_{\rm L}[q,\dot q]}, \qquad \dot q = \frac{\partial H_W}{\partial p}. \tag{1.15}$$

One often says that "accounting for the measure factor restores unitarity." However, the measure can safely be ignored (i.e. treated as $\mu = 1$) in the semiclassical limit because its nontrivial dependence on q is always of order $\hbar$. So we will sometimes ignore it.

Boundary conditions. Observe that the action $S_{\rm L}$ produced by our derivation has the correct form to yield a well-posed variational problem in which the initial and final positions are fixed. Fixing q_i and q_f also means that the measure $\mathcal{D}p$ includes one more p integral than $\mathcal{D}q$ has q integrals. (In this sense, the momenta live "between" positions.) We could instead have derived a path integral expression for $\langle p_f,t_f|p_i,t_i\rangle$. Such a path integral would have one more integral over q than over p, and its action would be modified by a boundary term. This boundary term is responsible for well-posing a variational problem in which the initial momenta are fixed. We can also interpret the boundary term as part of a generating function that enacts a canonical transformation that rotates the q's into the p's.

General states. The discussion above admits a nice generalization. The boundary terms above discussed can be regarded as coming from the wave functions of *arbitrary* states $|\psi\rangle$ and $|\psi'\rangle$. The path integral expression for their overlap is written

$$\begin{aligned}\langle\psi'|\psi\rangle &= \int \mathrm{d}q_i\,\mathrm{d}q_f\,\langle\psi'|q_f,t_f\rangle\,\langle q_f,t_f|q_i,t_i\rangle\,\langle q_i,t_i|\psi\rangle = \int \mathrm{d}q_i\,\mathrm{d}q_f\,\psi'(q_f,t_f)^*\psi(q_i,t_i)\,\langle q_f,t_i|q_i,t_i\rangle = \\ &= \int \mathcal{D}q\,\mathcal{D}p\,\psi'(q_f)^*\psi(q_i)\,e^{iS_{\rm L}} = \int \mathcal{D}q\,\mathcal{D}p\,\exp\Big[i\big(S_{\rm L} - \ln\psi'(q_f)^* - \ln\psi(q_i)\big)\Big]. \end{aligned}\tag{1.16}$$

Here we absorbed the two position integrals $\mathrm{d}q$ and $\mathrm{d}q'$ into the path integral measure $\mathcal{D}q$. This unfortunate ambiguity in the notation is often left for readers to resolve. In the last expression, the $\ln\psi$ terms are not integrated $\mathrm{d}t$, but are rather evaluated at the endpoints. They are boundary terms in $S_{\rm L}$ and, as above, they generate canonical transformations. More poetically, if $|\psi\rangle$ is somewhere "between" $|q\rangle$ and $|p\rangle$, then the $\ln\psi$ terms in $S_{\rm L}$ provide a suitable "rotation" of the variational problem to a basis that includes $|\psi\rangle$.

Correlators. Propagators are not the only quantities that path integrals can compute. They can also handle matrix elements of operators, or *correlators*. Fix instants in time $t_i < t_1 < t_2 < \cdots < t_n < t_f$ and consider operators $\mathcal{O}_1(t_1), ..., \mathcal{O}_n(t_n)$. Then by time-slicing and inserting complete sets of states between each operator, we arrive at

$$\langle q_f, t_f | \mathcal{O}_n(t_n) \cdots \mathcal{O}_1(t_1) | q_i, t_i \rangle = \int_{q(t_i)=q_i}^{q(t_f)=q_f} \mathcal{D}q\, \mathcal{D}p\, \mathcal{O}_{W,n}(t_n) \cdots \mathcal{O}_{W,1}(t_1)\, e^{iS_{\rm L}}, \tag{1.17}$$

where $\mathcal{O}_{W,j}$ is the Weyl symbol of $\mathcal{O}_j$. We describe this situation by saying that we path-integrate from t_i to t_f and insert the $\mathcal{O}_j$ at times t_j. Note that the ordering prescription above only gives the correct expression for the correlator if the $\mathcal{O}_j$ insertions are time-ordered.

Ordering and timefolds. More generally, we might wish to compute out-of-time-ordered correlators (OTOCs), where the restriction $t_1 < \cdots < t_n$ is lifted, by modifying our time-slicing. Suppose, for example, that we want to compute $\langle q_f, t_f | \mathcal{O}_3(t_3)\mathcal{O}_2(t_2)\mathcal{O}_1(t_1) | q_i, t_i \rangle$, where $t_i < t_3 < t_2 < t_1 < t_f$. To do so, we should start at t_i, evolve all the way up to t_1, insert $\mathcal{O}_1(t_1)$, then evolve *backwards* to t_2, insert $\mathcal{O}_2(t_2)$, evolve backwards again to t_3, insert $\mathcal{O}_3(t_3)$, and finally evolve forward to finish at t_f. We illustrate this situation in Fig. 1.

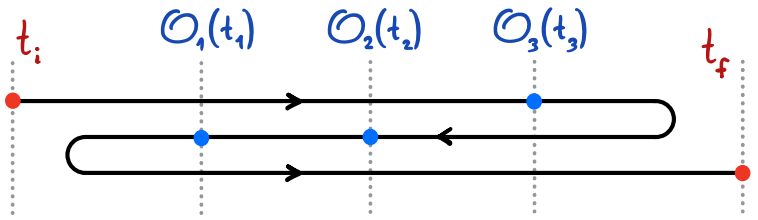

Figure 1: The path integral picture of an OTOC.

Writing down an explicit formula to describe this calculation would be extremely tedious; in practice, (1.17) is written down along with the picture above. The picture is equivalent to the following instructions: "construct a path integral of the form (1.14) on the interval $[t_i, t_1]$, evolving from one time step to the next using $e^{-i\hat{H}\Delta t}$; then insert $\mathcal{O}_1(t_1)$; then compute another path integral on $[t_2, t_1]$, evolving from one time step to the next using $e^{+i\hat{H}\Delta t}$; then insert $\mathcal{O}_2(t_2)$; etc." This procedure is often called *timefolding*—the terminology is a nod to orbifolds. The main idea is just that path integrals compute *path-ordered*, rather than time-ordered, correlators, and we are free to specify the path by evolving forwards or backwards using $e^{\pm i\hat{H}\Delta t}$, in order to hit all of the operator insertions in the right sequence.

Loops and traces. The object $\langle q_f, t | \mathcal{O}(t_1) | q_i, t \rangle$ is called an *in-in* or closed time path integral, because the initial and final times are equal, while (say) $t_1 > t$. It is computed by the path integral depicted in Fig. 2 (left). If we also set $q_i = q_f = q_0$ and integrate over all

q_0 as shown in Fig. 2 (right), the resulting path integral computes the trace of $\mathcal{O}(t_1)$:

$$\int \mathrm{d}q_0 \, \langle q_0, t|\mathcal{O}(t_1)|q_0, t\rangle = \mathrm{Tr}[\mathcal{O}(t_1)] = \oint \mathcal{D}q \, \mathcal{D}p \, \mathcal{O}_W(t_1) \, e^{iS_{\mathrm{L}}}, \tag{1.18}$$

where the loop around the integral indicates periodic boundary conditions $q(t_i) = q_0 = q(t_f)$. Notice that the trace depends only on t_1, and is independent of the time t at which we started! More generally, it is a general property of path integrals that as long as we do not hit any operator insertions, the path integral is invariant under arbitrary deformations of the time-slicing contour. This is "obvious" if you don't think too hard: going forward evolves by $e^{-i\hat{H}\Delta t}$, going backward evolves by $e^{+i\hat{H}\Delta t}$, and the two cancel each other out.

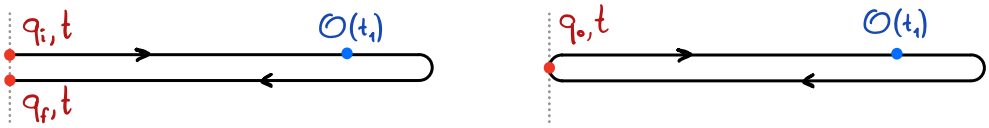

Figure 2: The closed time path integral (left) and the path integral for a trace (right).

Preparing states. Occasionally, we wish to calculate overlaps $\langle\psi'|\psi\rangle$ and correlators $\langle\psi'|\mathcal{O}|\psi\rangle$ in states that are *prepared* from some known vacuum or reference state $|0\rangle$. That is, there are (possibly complicated) "guiding" Hamiltonians H_g and H'_g for which

$$|\psi\rangle = \mathcal{P}\exp\left[i\int_{-\infty}^{t_i} \mathrm{d}t \, H_g(t)\right]|0\rangle\,, \qquad |\psi'\rangle = \mathcal{P}\exp\left[i\int_{+\infty}^{t_f} \mathrm{d}t \, H'_g(t)\right]|0\rangle\,. \tag{1.19}$$

Then to calculate $\langle\psi'|\psi\rangle = \langle 0|U(t_f, \infty)U(-\infty, t_i)|0\rangle$, we write down a path integral for $\langle 0|0\rangle$ on the entire real axis $t \in (-\infty, \infty)$, being careful to evolve using H_g on $(-\infty, t_i]$ and using H'_g on $[t_f, \infty)$. This situation is depicted in Fig. 3. In words, we often say that the path integral "prepares" $|\psi\rangle$ and $|\psi'\rangle$, or that these states are assembled from sources.

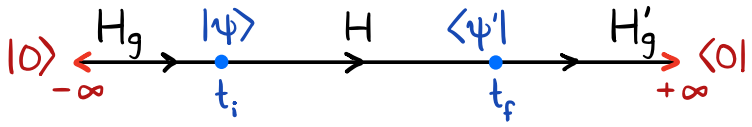

Figure 3: Preparing and evolving states using a path integral.

1.3 Euclidean Path Integrals

Euclidean time. Suppose that we want to study the object $\langle q_f, t_0|e^{-\beta H}|q_i, t_0\rangle$, where β is a positive real number and we assume for simplicity that H is time-independent. We can regard this as a special case of the in-in PI $\langle q_f, t_0|\mathcal{O}|q_i, t_0\rangle$ for $\mathcal{O} = e^{-\beta H}$. Alternatively, if we make the somewhat ad-hoc identification $\beta = i\Delta t$, the object above looks like an ordinary PI evolving in an imaginary time direction. Students of statistical mechanics will

recognize $e^{-\beta H}$ as proportional to the *thermal density matrix*, which leads to the mantra that "statistical mechanics is quantum mechanics in imaginary time."

In any case, let us use $\Delta t = -i\beta$ to write down a PI for this matrix element:

$$\langle q_f, t_0 | e^{-\beta H} | q_i, t_0 \rangle = \int_{q(t_0)=q_i}^{q(t_0)=q_f} \mathcal{D}q\, \mathcal{D}p\, \exp\left(i \int_{t_0}^{t_0 - i\beta} \mathrm{d}t \left[p\dot{q} - H_W(q,p) \right] \right). \tag{1.20}$$

Next, we define the *imaginary time* $\tau = i(t - t_0)$, so that the change in τ is $\Delta\tau = i\Delta t = \beta$. (The transformation $t \longrightarrow it$ is called a *Wick rotation*.) This allows us to rewrite the action by manipulating some factors of i. We have $\mathrm{d}\tau = i\,\mathrm{d}t$, so $\dot{q} = \frac{\mathrm{d}q}{\mathrm{d}t} = i\frac{\mathrm{d}q}{\mathrm{d}\tau}$. We then pull out another minus sign to define the *Euclidean action* S_E:

$$iS_\mathrm{L} = i \int_{t_0}^{t_0 - i\beta} \mathrm{d}t \left(p \frac{\mathrm{d}q}{\mathrm{d}t} - H_W \right) = -\int_0^\beta \mathrm{d}\tau \left(-ip \frac{\mathrm{d}q}{\mathrm{d}\tau} + H_W \right) \equiv -S_\mathrm{E}. \tag{1.21}$$

Henceforth using $\dot{q}$ to refer to $\frac{\mathrm{d}q}{\mathrm{d}\tau}$ for shorthand, we see that the PI becomes

$$\langle q_f, t_0 | e^{-\beta \hat{H}} | q_i, t_0 \rangle = \int_{q(0)=q_i}^{q(0)=q_f} \mathcal{D}q\, \mathcal{D}p\, \exp\left(-\int_0^\beta \mathrm{d}\tau \left[-ip\dot{q} + H_W \right] \right) = \int_{q(0)=q_i}^{q(0)=q_f} \mathcal{D}q\, \mathcal{D}p\, e^{-S_\mathrm{E}}. \tag{1.22}$$

A simple example. Consider the Hamiltonian for a single particle moving in one dimension under the influence of a potential $V(q)$. The Hamiltonian is given by

$$\hat{H} = \frac{\hat{p}^2}{2m} + V(\hat{q}) \implies H_W = \frac{p^2}{2m} + V(q), \tag{1.23}$$

so our Euclidean PI takes the form

$$\langle q_f, t_0 | e^{-\beta \hat{H}} | q_i, t_0 \rangle = \int_{q(0)=q_i}^{q(0)=q_f} \mathcal{D}q\, \mathcal{D}p\, \exp\left(-\int_0^\beta \mathrm{d}\tau \left[-ip\dot{q} + \frac{p^2}{2m} + V(q) \right] \right). \tag{1.24}$$

The $\mathcal{D}p$ integral in (1.24) can be evaluated exactly because S_E is quadratic in p. The p integral evaluates to a q-independent number, so the result is a covariant (Lagrangian), rather than a canonical (Hamiltonian), PI over positions, of the form (1.15) with $\mu[q] = \mu$ a (formally infinite) constant. The covariant action $S_\mathrm{E}[q, \dot{q}]$ is obtained from the canonical one appearing in (1.24) by setting p to its equation of motion. This equation of motion is

$$\frac{\delta S_\mathrm{E}}{\delta p} = 0 = -i\dot{q} + \frac{\partial H_W}{\partial p} = -i\dot{q} + \frac{p}{m} \implies p = im\dot{q}. \tag{1.25}$$

The Hamiltonian $H_W = -\frac{m\dot{q}^2}{2} + V(q)$ looks like it has the wrong sign on its kinetic term, a

distinctive characteristic of Euclidean time. The covariant Euclidean action is

$$S_{\rm E}[q,\dot{q}] = \int_0^\beta {\rm d}\tau \left(m\dot{q}^2 - \frac{m\dot{q}^2}{2} + V(q) \right) = \int_0^\beta {\rm d}\tau \left(\frac{m\dot{q}^2}{2} + V(q) \right). \tag{1.26}$$

The integrand above looks distinctively energy-like. This leads to the continuation of our vague and grandiose mantra: "dynamics in real time is equivalent to statics in imaginary time." The important point here is that since the "energy functional" $S_{\rm E}[q,\dot{q}]$ is positive-definite, the quantity $e^{-S_{\rm E}}$ defines a convergent path integral.

Thermal correlators. Let us now take $q_i = q_f = q_0$ in the discussion above. As in (1.18), an extra integral over q_0 yields a path integral expression for the trace of $e^{-\beta\hat{H}}$, an object known as the *partition function* in statistical mechanics:

$$Z(\beta) = {\rm Tr}\left[e^{-\beta\hat{H}} \right] = \oint \mathcal{D}q\, \mathcal{D}p\, e^{-S_{\rm E}[q,p]}. \tag{1.27}$$

Earlier, we drew path integrals by indicating horizontal real-time evolution arrows; now, as shown in Fig. 4, vertical arrows denote evolution in imaginary time. One way to interpret the partition function is that the endpoints of the interval $[0,\beta]$ are identified, turning it into the *thermal circle* $S^1_\tau(\beta)$ of circumference β. Its compactness is one of the many reasons Euclidean path integrals are often easier to work with than their Lorentzian counterparts.

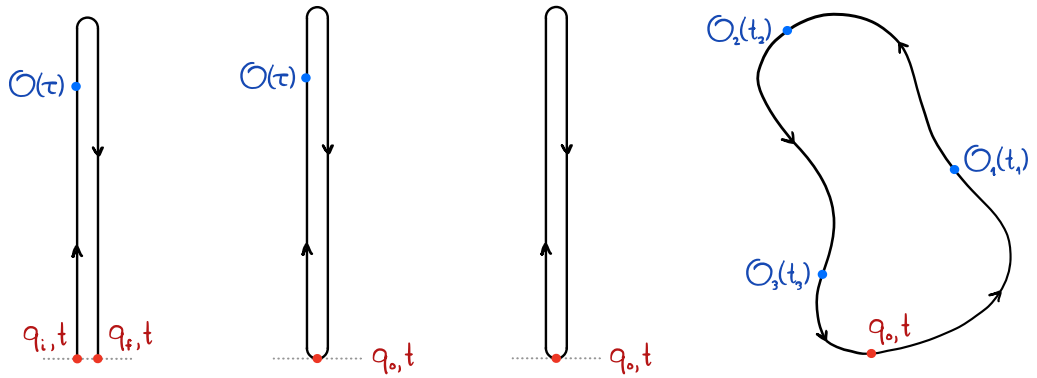

Figure 4: Thermal correlators. From left to right: the (q_i, q_f) matrix element of $\mathcal{O}$; the thermal expectation value of $\mathcal{O}$; the partition function; a generic three-point function.

Notice that if we take $\beta \longrightarrow \infty$, the integrand rapidly decays to zero. The slowest decay is provided by the smallest value taken on by $S_{\rm E}$; this corresponds to the smallest eigenvalue of H. Thus the limit $\beta \longrightarrow \infty$ corresponds to a projection of the thermal density matrix $e^{-\beta H}$ onto the ground state of the system, and the low-temperature properties of the system correspond to its ground-state dynamics. This happens irrespective of initial conditions, which is why we say that in equilibrium, any system settles into its ground state, and that the early-time fluctuations caused by the system's initial conditions are washed out.

Schwinger-Keldysh. The Euclidean PI and its attendant pictures strongly suggest that we should think of t as taking on complex values. This move allows us to draw arbitrary evolution contours in $\mathbb{C}$. So long as H is independent of time, it does not matter which contour we use to do PIs, so long as it passes through the right operator insertions in the right order. This partial analog of Cauchy's theorem from complex analysis proves extremely useful, and allows for the calculation of *thermal correlation functions* of the form

$$\langle \mathcal{O}(t_1) \cdots \mathcal{O}(t_n) \rangle_\beta \equiv \mathrm{Tr}\big[e^{-\beta H} \mathcal{O}(t_1) \cdots \mathcal{O}(t_n)\big], \qquad t_1, ..., t_n \in \mathbb{C} \tag{1.28}$$

by drawing a timefolded contour through $\mathbb{C}$ that passes through $t_1, ..., t_n$ in that order. These path integrals are called *Schwinger-Keldysh* PIs. But one must be careful: evolution with $e^{-\beta H}$ "always" makes sense because S_E is usually positive-definite, but by the same token evolution with $e^{+\beta H}$ "never" makes sense. One usually deals with this issue by ignoring it, or rather by trying hard never to decrease the imaginary part of t in constructing PIs.

1.4 Gravitational Path Integrals

Apologia. Having studied path integrals in some generality, we seek to apply them to quantum gravity. To put things optimistically, no one knows how to do this. Of course, being physicists, we will do so anyway. We will assume, for the most part, that our intuition from quantum mechanics carries over directly to quantum gravity, and that the resulting formal expressions are well-enough defined for us to do calculations with them. In this section, we will pause to take stock of the necessary generalizations and to "bemoan" some of the technical issues we will promptly leave behind afterwards.

The quantum gravity PI. We have considered the PI for a single particle moving in one dimension. But our formalism works equally well for any number of quantum particles moving on an arbitrary manifold. After this, the next step is to generalize to quantum field theory (QFT). This is usually done by considering a many-body system, often on a lattice, and taking the continuum limit of zero lattice spacing and infinitely many particles. Alternatively, one demotes position q from the status of an operator to the status of a label, and then defines fields as operator-valued functions of all of the spacetime parameters. In this sense, quantum mechanics is a field theory whose "fields," the operators $\hat{q}(t)$, live on a $(0+1)$–dimensional spacetime. After passing to field theory, one must grapple with the issue of gauge symmetry. All the while, one must ensure that the theory has a sensible Hamiltonian formulation; already at the level of gauge theories, this is complicated.

One is then finally ready for quantum gravity. The quantum field under scrutiny here is the spacetime metric $g_{\mu\nu}$, and one integrates e^{iS} over "all metrics" or "all universes" to compute a gravitational PI. Whereas ordinary QFT was an account of fields fluctuating on a fixed spacetime background, in quantum gravity it is spacetime itself that is dynamical.

The structure of gravity. The idea of integrating e^{iS} over all metrics begs two questions: (1) what, exactly, is S? and (2) if the metric g is the "position" variable here, then what is the momentum? We will discuss (1) in detail soon; the terse answer is that S is the Einstein–Hilbert action of general relativity with the Gibbons–Hawking–York boundary term. We will *not* discuss (2) in great detail. There are several approaches to a Hamiltonian formulation of GR, among them the ADM and first-order (Cartan) formalisms. Each framework comes with its own subtleties and difficulties, many of them related to gauge invariance.

Gauge issues I. Classical general relativity (GR) is a gauge theory. This means that our description of the theory has redundancies: there are configurations of the system that look like distinct states "on paper," but which are physically identical. We denote by $\mathcal{A}$ the (possibly redundant) set of configurations of the system, and by $\mathcal{G}$ the group of gauge transformations of the theory. Consider a particular state $\phi \in \mathcal{A}$: applying a transformation $g \in \mathcal{G}$ will produce a gauge-equivalent state $g\phi \in \mathcal{A}$, and the set of *all* configurations gauge-equivalent to ϕ is called its *gauge orbit*. Each gauge orbit is typically isomorphic to $\mathcal{G}$ itself, because each state $g\phi$ is usually distinct (in $\mathcal{A}$) from any other $g'\phi$. Moreover, the space $\mathcal{A}$ is foliated (i.e. ruled, in the sense of lined paper) by its gauge orbits. Each physical degree of freedom in a gauge theory is therefore labeled uniquely by the gauge orbit it lies on; accordingly, we write that the phase space of the theory is the quotient $\mathcal{P} = \mathcal{A}/\mathcal{G}$.

One might imagine coordinatizing $\mathcal{P}$ by gauge-invariant variables q_{inv} and p_{inv}; one might further imagine using the fact that the action is constant along the gauge orbits to construct a PI by integrating the quantity e^{iS} over $\mathcal{P}$. In practice, we do this by *gauge fixing*: we choose a representative configuration on each gauge orbit and label points on the resulting "slice" by the gauge orbits it passes through, as shown in Fig. 5 below.

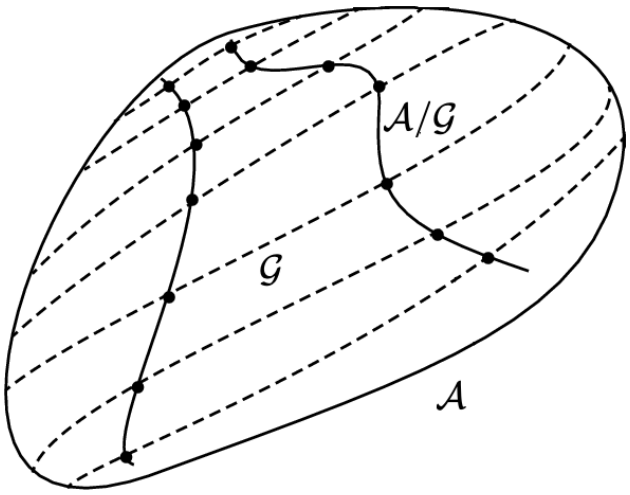

Figure 5: The space $\mathcal{A}$ of field configurations. Each gauge orbit (dotted) is isomorphic to the gauge group $\mathcal{G}$, and the gauge slices (solid) are isomorphic to the phase space $\mathcal{P} = \mathcal{A}/\mathcal{G}$.

Gauge issues II. In all but the simplest gauge theories, it is hard to find a global set of such coordinates. It is usually easier to integrate over *all* of $\mathcal{A}$. The fact that S is constant along gauge orbits manifests as a prefactor of the (infinite) volume of the gauge group:

$$\int \mathcal{D}q\,\mathcal{D}p\,e^{iS} = \int \mathcal{D}q_{\text{inv}}\,\mathcal{D}p_{\text{inv}} \int \mathcal{D}[\text{orbit}(q_{\text{inv}}, p_{\text{inv}})]\,e^{iS} = \text{vol}(\mathcal{G}) \int \mathcal{D}q_{\text{inv}}\,\mathcal{D}p_{\text{inv}} e^{iS}. \quad (1.29)$$

On one hand, what we really want to study is the naïve path integral over metrics and momenta, divided by this volume; for the most part, this is what we will do. Ignoring gauge symmetry like this sometimes has funny consequences: if two $q \neq q'$ lie on the same gauge orbit, then "obviously" $\langle q', t|q, t\rangle \neq 0$, even though the notation would suggest otherwise. On the other hand, extremely sophisticated technology exists for handling the true gauge-invariant path integral. All of these techniques stem from the Faddeev–Popov approach to gauge fixing and from BRST symmetry. Among them, the BFV, BV, and antifield formalisms are the most powerful; they are also extremely abstract. It should be noted that the infamous *Gribov ambiguity* often prevents one from being able to find global coordinates on $\mathcal{P}$, i.e. from fixing a gauge slice that intersects each orbit exactly once. This problem is especially manifest in GR, where the gauge group is the infinite-dimensional diffeomorphism group: every coordinate transformation is a gauge transformation.[1]

For the most part, we will be able to evade all of these issues by working semiclassically. This often amounts to the declaration that the gauge-invariant PI is exactly equal to e^{iS_0}, where S_0 is the value of the gravitational action evaluated at a stationary point, i.e. on a classical solution. In this way, we can simply assume that gauge fixing and all its discontents have come and gone; what remains after the dust clears is just classical GR.

Indefiniteness. The Euclidean gravitational action, which we will discuss soon, is *not* positive-definite! In fact, it is unbounded above and below. We offer two terse proofs. The first proof: the Einstein–Hilbert Lagrangian consists essentially of the scalar curvature R of a spacetime, and by choosing arbitrarily curved spacetimes like spheres or hyperbolic spaces, one can make R (and hence the gravitational action) arbitrarily positive or negative. The second—and more physical—proof: recall the Friedman equation from cosmology,

$$\left(\frac{\dot{a}}{a}\right)^2 = -\frac{k}{a^2} + \frac{8\pi G\rho}{3}. \quad (1.30)$$

Suppose the matter content of the universe is a spatially homogeneous, massless scalar field $\phi(t)$. The energy density ρ will then include (among other things) a kinetic term $\dot{\phi}^2$. Thus the Friedman equation reads $\dot{\phi}^2 - (\dot{a}/a)^2 + \cdots = 0$; but this says that the scale factor a can contribute a *negative* "kinetic" energy, and can make the total energy unbounded.

[1] Note that this statement goes beyond the usual platitude that every physical theory must be generally covariant, i.e. independent of one's choice of coordinates. What makes GR special in this regard is that the spacetime geometry itself is dynamical, rather than a background structure—more on this later.

The Gauss law. We saw in the previous section that the positive-definiteness of "sensible" Euclidean actions helps the operator $e^{-\beta\hat{H}}$ yield a convergent PI. So the situation here looks bad—really bad. But energy in GR is extremely subtle: the gravitational Hamiltonian is in fact positive-definite, but is only so *on shell*, i.e. when the equations of motion hold. Actually, one only needs to satisfy the gravitational *constraint equations*, which impose gauge invariance and are analogous to the Gauss law in electromagnetism.

The cosmological example above actually shows that the total energy (and thus the Hamiltonian) of the system is always constrained to be exactly zero. In fact, this is also what happens in electromagnetism, where Gauss's law forces the total charge in a closed universe to vanish. So the Euclidean action is unbounded, but the on-shell Hamiltonian is positive-definite. This suggests that a truly gauge invariant formulation of GR should produce a manifestly positive Euclidean action, and that somehow the process of dropping the constraints and losing gauge invariance is responsible for the unboundedness of S_{E}. (Another way to say this is that we should constrain, then quantize, rather than quantize, then constrain.) Two approaches to this problem, which is called the *conformal factor problem* in the literature, are typically taken: (1) work in the semiclassical approximation, which is already on shell and thus avoids these issues; or (2) Wick-rotate the part of the metric responsible for the scale factor to improve the convergence of the path integral.

The Euclidean action. The (Lorentzian) Einstein–Hilbert action for a spacetime (M, g) is the integral of its scalar curvature over the spacetime manifold:

$$S_{\mathrm{L}}[g] = \frac{1}{16\pi G}\int_M \mathrm{d}^d x\sqrt{-g}\,R, \tag{1.31}$$

One immediate problem with defining a corresponding Euclidean action is that arbitrary pseudo-Riemannian manifolds generically lack a global time coordinate to Wick-rotate, so it is unclear what the prescription $t \longrightarrow \tau = it$ means. We will return to this issue soon; for the moment let us assume that it all makes sense. The Wick-rotated spacetime is then a Riemannian manifold whose metric and scalar curvature are also denoted g and R.

The Euclidean gravitational action looks almost identical to the Lorentzian action, except for an unexpected minus sign:

$$S_{\mathrm{E}}[g] = -\frac{1}{16\pi G}\int_M \mathrm{d}^d x\sqrt{g}\,R. \tag{1.32}$$

This sign is just a consequence of carrying out the Wick rotation carefully and consistently imposing $iS_{\mathrm{L}} = -S_{\mathrm{E}}$. It can be seen most clearly in the spacetime $M = S^3 \times \mathbb{R}$, where $\mathbb{R}$ is the time axis. Here $R_{\mathrm{L}} = R_{\mathrm{E}} = R^{(3)}$ is the constant curvature of the 3-sphere. Upon

integrating over a finite time interval $[t_i, t_f]$ with $t = -i\tau \iff \mathrm{d}t = -i\,\mathrm{d}\tau$, we find

$$iS_{\mathrm{L}} = \frac{i}{16\pi G}\int_M \mathrm{d}^4x\,\sqrt{-g}\,R_{\mathrm{L}} = i\frac{\mathrm{vol}(S^3)}{16\pi G}R^{(3)}\int_{t_i}^{t_f}\mathrm{d}t = \cancel{(i)(-i)}\frac{\mathrm{vol}(S^3)}{16\pi G}R^{(3)}\int_{\tau_i}^{\tau_f}\mathrm{d}\tau =$$
$$= +\frac{1}{16\pi G}\int_M \mathrm{d}^4x\,\sqrt{g}\,R_{\mathrm{E}} \overset{!}{=} -S_{\mathrm{E}} \implies S_{\mathrm{E}} = -\frac{1}{16\pi G}\int_M \mathrm{d}^4x\,\sqrt{g}\,R. \tag{1.33}$$

Topology change. In classical GR, the best-behaved spacetimes (with respect to causality) are called *globally hyperbolic*. These manifolds have the topology $M = \Sigma \times \mathbb{R}$, where $\mathbb{R}$ is the time axis, and the *Cauchy surface* Σ retains the same topology at all times. It is this class of spacetimes for which the initial value problem in GR is well posed, for which the Hamiltonian formalism is well defined, and on which QFT in curved spacetime is most naturally formulated. In particular, globally hyperbolic spacetimes have a globally defined time coordinate, and this makes it more clear what a Wick rotation means.

A big question arises: should the gravitational PI integrate over spacetimes with arbitrary topology? In Lorentzian signature, where topology change is classically forbidden in the sense described above, the answer seems to be "no; restrict to globally hyperbolic manifolds." In Euclidean signature, however, a plethora of topologies exist which have finite action and satisfy the Einstein equations. The question of which spacetimes to include in the gravitational PI is still open; in the remainder of these notes, we will take the view that the gravitational PI *does* sum over all topologies. The consequences will be dramatic.

Pants. One justification for this choice comes from the fact that Hamilton-Jacobi theory (in the semiclassical approximation) can be used to show that Euclidean solutions in gravity are related to tunneling processes in the Lorentzian theory. This seems to indicate that there are topology-changing processes even in the Lorentzian theory. A cute illustration of this idea is based on the (Euclidean) pair of pants spacetime, shown in Fig. 6.

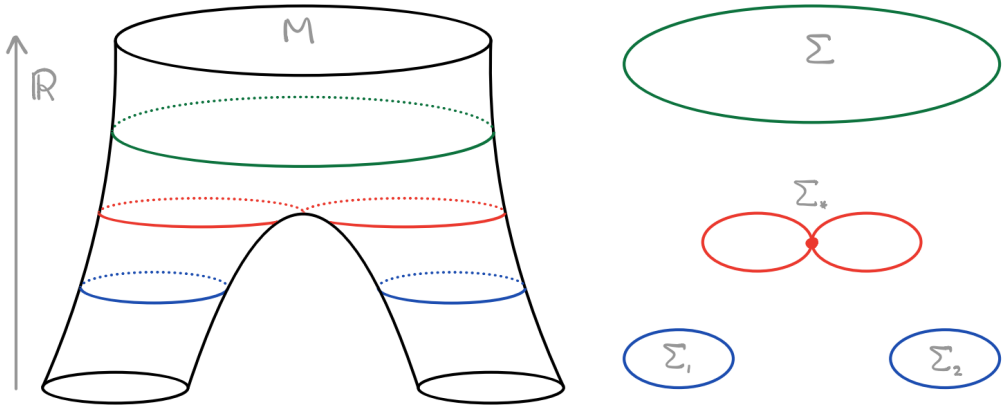

Figure 6: Pants.

We imagine the time arrow as running upwards and consider spatial slices of the surface.

There are initially two disjoint universes; as they coalesce into a larger univese, the spatial geometry is momentarily singular at the crossover point. Even though the Euclidean spacetime is perfectly smooth, it does not admit a smooth Lorentzian metric. Rather than discarding such metrics, however, we will grapple directly with singularities in the gravitational PI. We sometimes think of the lower half of the pair of pants as living in the configuration space of 2-universe states, and the upper half as living in the configuration space of 1-universe states. They are joined at their boundary by a singular configuration, which nevertheless should have finite action. This suggests that it can be tunneled through, and that state spaces of metrics with different topologies should be somehow stitched together at degenerate configurations.

In retrospect, this should not be too surprising. The PI measure is generically supported on singular configurations, paths (or geometries) which are not smooth, nor even C^1.

2 Gravity in a Box

2.1 The Gravitational Action

Motivation. After developing the canonical path integral in quantum mechanics in both Lorentzian and Euclidean signature, we discussed its generalization to the case of gravity. For the most part, our PI calculations will be semiclassical, which means that we will evaluate the gravitational action on solutions to Einstein's equations. It therefore behooves us to understand the classical action of GR in more detail. In this section, we will do so in Lorentzian signature; in the next section, we will pass to Euclidean signature.

The variational principle. In most GR courses, one is taught that the Einstein field equations of general relativity follow from the Einstein–Hilbert action (1.31). This is indeed the case, but the derivation involves throwing away certain boundary terms:

$$S_{\text{EH}}[g] = \frac{1}{16\pi G}\int_M \mathrm{d}^d x\sqrt{-g}\,R \implies \delta S_{\text{EH}} = \int_M \mathrm{d}^d x\sqrt{-g}\,(\text{EOM})_{ab}\delta g^{ab} + \left(\begin{smallmatrix}\text{boundary}\\ \text{terms}\end{smallmatrix}\right). \quad (2.1)$$

These terms do not affect local physics in the interior of M, so often the Einstein–Hilbert action is good enough. But we are interested in the actual numerical value of the action, so we must tread carefully to make sure we are using the correct action to begin with.

We say that an action has a *good variational principle* or yields a *well-posed variational problem* when its variation gives *only* the equations of motion, with no extra boundary terms. Such boundary terms may be interpreted as "unwanted" equations of motion on the boundary; moreover, the discussion of semiclassical PIs above applies only to actions with a good variational principle. The Einstein–Hilbert action does not satisfy this requirement, so we must add to it an additional boundary term that cancels the one above.

Gravity in a box. For pedagogical simplicity, we will take the spacetime (M, g) to be the "box" shown in Fig. 7 (left). We assume that M is bounded to the future and past by initial and final surfaces, and to the "sides" by a timelike (i.e. spatial) boundary manifold ∂M. Depending on which boundary conditions we choose to impose and which variational problem we wish to solve, the aforementioned boundary term in the action will live on either the spatial or past/future boundaries. For the standard Cauchy problem in GR, the boundary term lives *only* on the spatial boundary: this is a definition, rather than a claim to be proven. Below are two pieces of intuition that explain why we consider such problems.

First of all, we will soon see that this boundary prescription is equivalent to the claim that the Hamiltonian in the canonical formulation of GR is a boundary term; hence it generates time translations on ∂M. In other words, the Hamiltonian defines the gravitational energy of M by generating a canonical transformation that moves the future and past boundaries while preserving the spatial boundary. All of this befits a variational problem where the action is specified on the fixed spatial boundary, but not on the "movable" temporal boundaries. Yet another equivalent way to say this is that modifying the action at the spatial boundaries modifies the Hamiltonian and therefore changes the system under consideration.

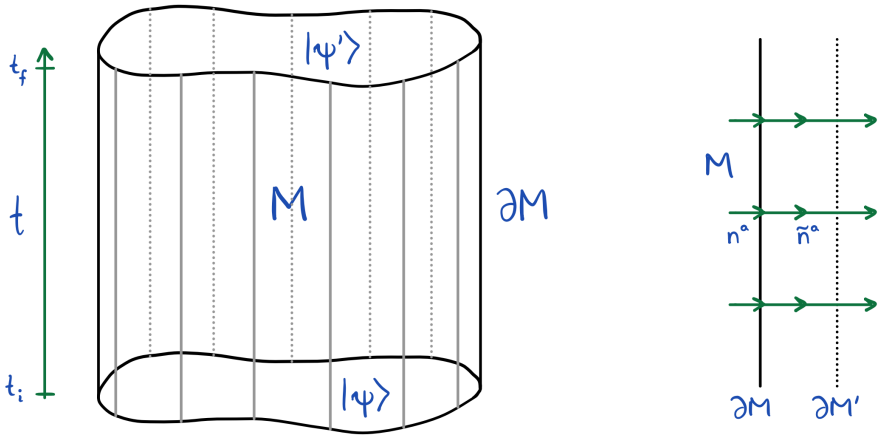

Figure 7: Left: gravity in a box. Right: the extrinsic curvature.

States. Second of all, recall the discussion of section 1.2, where it was shown that the boundary terms appearing in the action are not universal, but are rather motivated by the particular quantum states whose overlap we want to compute using a PI,

$$\langle\psi'|\psi\rangle = \int \mathcal{D}q\,\mathcal{D}p\,\exp\Big[i\big(S_{\rm L} - \ln\psi'(t_f)^* - \ln\psi(t_i)\big)\Big], \tag{2.2}$$

where—crucially—$|\psi\rangle$ and $|\psi'\rangle$ are specified on the past and future boundaries of M. So changing the action on the past and future boundaries is tantamount to modifying the *state* of the system, while changing the action on the spatial boundary is equivalent to modifying the system itself. We aim to fix the physical system and examine the evolution of its states, so we should fix the action on the spatial boundaries.

As a side comment, (2.2) shows that the true variational principle that defines the instantiation of the classical theory we care about is *not* $\delta S_{\rm L} = 0$, but rather

$$\delta(S_{\rm L} - \ln\psi'(t_f)^* - \ln\psi(t_i)) = 0 \iff \delta S_{\rm L} = \ln[\psi'(t_f)^*\psi(t_i)]. \tag{2.3}$$

In other words, an "action flux" at the past and future boundaries is allowed, so long as we understand it to represent a modification of the initial and final states of the system.

A look ahead. In what follows, we will make the discussion above more precise. We will begin by varying the Einstein–Hilbert action explicitly to better understand (2.1). Next, we discuss the choice of boundary conditions on ∂M that defines the variational problem we aim to solve. We will then define and meditate on the extrinsic curvature as the key tool for constructing the requisite boundary term. And finally, after a few more comments, we arrive at the celebrated Gibbons–Hawking–York boundary term in the gravitational action. For the impatient reader who wishes to skip §2.2, the end result is

$$S = S_{\rm EH} + S_{\rm GHY} = \frac{1}{16\pi G}\int_M \mathrm{d}^d x\sqrt{-g}\,R + \frac{1}{8\pi G}\int_{\partial M}\mathrm{d}^{d-1}x\sqrt{-h}\,K, \tag{2.4}$$

where h is the induced metric on ∂M and K is the extrinsic curvature of $\partial M \subset M$.

2.2 Boundary Terms and Conditions

Variation of the action. A calculation, found (e.g.) in Appendix E of Wald, shows that

$$\delta S_{\rm EH} = \frac{1}{16\pi G}\int_M \mathrm{d}^d x\sqrt{-g}\left(R_{ab} - \frac{1}{2}Rg_{ab}\right)\delta g^{ab} + \frac{1}{16\pi G}\int_{\partial M}\mathrm{d}^{d-1}x\sqrt{-h}\,n^a_{\partial M}v_a, \tag{2.5}$$

where h is the (Lorentzian) boundary metric on ∂M induced from the bulk metric g, $n_{\partial M}$ is the unit normal to ∂M, and v is the variation given by

$$v^a = \left(-g^{ab}g^{cd} + g^{bd}g^{ac}\right)\nabla_b(\delta g_{cd}). \tag{2.6}$$

At first sight, v looks like a complete mess. But on closer inspection, dimensional analysis almost completely fixes its form. The Ricci scalar contains two derivatives of the metric, so the variation and an integration by parts removes one: $R \sim \partial\partial g \implies \delta R \sim \nabla(\delta g) + \cancel{\delta(\nabla g)}$. The resulting object has three indices and must be contracted with the one-index unit normal, so it needs to be an object that carries four indices and has no metric derivatives. The only two options for such a term are exactly what appear in parentheses above. The only thing not fixed by this argument is the numerical prefactors at each term.

Dirichlet and Neumann. From the index contractions present in (2.5–2.6), we see that the first term of $n^a_{\partial M}v_a$ involves only derivatives of the metric in the normal direction to ∂M,

while the second term involves derivatives in both the normal and tangential directions.

In order to specify the variational problem to be solved by the full gravitational action, we need to choose the boundary conditions obeyed by the metric on ∂M. Two common choices are Dirichlet conditions, where h is held fixed, and Neumann conditions, where the normal derivatives of h are held fixed. We will take the Dirichlet case: we set $\delta h = 0$, and as a consequence all derivatives of δg tangent to ∂M must vanish. Therefore the boundary term we seek must have a variation which, subject to $\delta h = 0$, produces exactly the right normal-derivative contributions to cancel (2.6).

Explicitly, the full gravitational action must take the form

$$S_{\mathrm{L}} = S_{\mathrm{EH}} + S_{\partial} = \frac{1}{16\pi G}\int_M \mathrm{d}^d x\sqrt{-g}\,R + \frac{1}{16\pi G}\int_{\partial M}\mathrm{d}^{d-1}x\sqrt{-h}\,\mathcal{L}_{\partial}, \tag{2.7}$$

where $\mathcal{L}_{\partial}$ is a scalar on ∂M constructed only from the boundary metric h using at most one normal derivative. What can such a term look like?

Extrinsic curvature. While many objects familiar from GR, like the Riemann tensor, measure the curvature of a manifold intrinsically and without reference to an embedding in some larger ambient space, the *extrinsic curvature tensor* measures the curvature of a submanifold as embedded within a larger space. The extrinsic curvature is a naïve notion: a cylinder in 3D space, for example, has vanishing Riemann tensor, but is "obviously" not flat relative to the ambient $\mathbb{R}^3$, and indeed has positive extrinsic curvature.

The extrinsic curvature K of a codimension-1 submanifold (N, h) embedded in (M, g) is the trace of half the (outward-pointing) normal Lie derivative of the induced metric:

$$K_{ij} = \frac{1}{2}\mathcal{L}_n h_{ij}, \qquad K = K_{ij}h^{ij} \tag{2.8}$$

Let's unpack this definition in the case $N = \partial M$: see Fig. 7 (right) above. The unit normal n^a to ∂M is a vector field defined only on ∂M, but we can extend it to a vector field $\widetilde{n}^a$ in the bulk of M in an arbitrary (but smooth) way. Although it is not yet obvious, the following is actually independent of $\widetilde{n}^a$. Viewing $\widetilde{n}^a$ as the infinitesimal generator of a diffeomorphism, we imagine moving ∂M along the flow lines of $\widetilde{n}^a$: each point on ∂M moves by the same proper distance in the ñormal direction. We can also use $\widetilde{n}^a$ to transport the boundary metric h to the new location of ∂M. (This is called "pulling back the metric.") The pulled-back metric h' on the flowed boundary $\partial M'$ can then be compared to the restriction of the bulk metric g on the surface $\partial M'$. The difference between the two, evaluated infinitesimally close to ∂M, is the Lie derivative $\mathcal{L}_n h_{ij}$. In this way, n^a provides a measure of how "wrinkled" ∂M is, and these wrinkes are what the extrinsic curvature picks up.

Example: the sphere. Let $S^n(r) \subset \mathbb{R}^{n+1}$ be the n-sphere of radius r. The unit normal vector points radially outwards, $n^a = (\partial_r)^a$, and the metric induced on S^n is the usual round

metric $\mathrm{d}s^2 = h_{ij}\mathrm{d}x^i\mathrm{d}x^j = r^2\mathrm{d}\Omega_n^2$. Under an infinitesimal variation in the radial direction, i.e. extending n^a to $\widetilde{n}^a = (\partial_r)^a$ away from $S^n(r)$, we find

$$\mathrm{d}s^2 = r^2\mathrm{d}\Omega_n^2 \implies \delta\big(\mathrm{d}s^2\big) = 2r\,\mathrm{d}\Omega_n^2\,\delta r \implies \delta h_{ij} = 2r\,\delta r. \tag{2.9}$$

The normal Lie derivative and the extrinsic curvature are therefore

$$\mathcal{L}_n h_{ij} = \frac{\delta h_{ij}}{\delta r} = 2r\,\mathrm{d}\Omega_n^2 = \frac{2}{r}h_{ij} \implies K_{ij} = \frac{1}{r}h_{ij} \implies K = \frac{1}{r}h_{ij}h^{ij} = \frac{n}{r}. \tag{2.10}$$

Some useful tools. Here we collect some formulæ that we will not discuss in detail, but which prove useful and whose derivations can be found in textbooks like Wald's.

First, we define an object h_{ab}, with bulk indices in M and closely related to the boundary metric h_{ij}, by $h_{ab} = g_{ab} - n_a n_b$. Formally, h_{ab} is the pullback of the boundary metric from ∂M to M along the natural projection $p\colon M \longrightarrow \partial M$. Raising an index gives the matrix h_a^b of (the differential of) this projection. The main property of h_{ab} is that it annihilates the normal direction: $n^a h_{ab} = n_b - (n^a n_a)n_b = 0$. Conversely, if v^a and w^b are tangent to ∂M, then contraction with h_{ab} leaves them unchanged: $v^a h_{ab} w^b = v^a w_a = v^i h_{ij} w^j$. In other words, h_{ab} is something like a projection onto the directions tangent to ∂M.

It turns out that h_{ab} appears in a very practical formula for computing the (pullback of the) extrinsic curvature: $K_{ab} = h_a^c \nabla_c n_b$. This formula avoids the explicit use of Lie derivatives, and it makes manifest the claim that K_{ab} depends only on the normal vector to ∂M, and not on its extension into M. As with h_{ab} above, K_{ab} has bulk indices and is related to K_{ij} in the same tangential-projection manner as h_{ab} is to h_{ij}. It therefore satisfies $K = K_{ab}g^{ab} = K_{ab}h^{ab} = K_{ij}h^{ij}$. Finally, we note that the extrinsic curvature tensor K_{ij}, the Riemann tensor R^M_{abcd}, and the boundary Riemann tensor $R^{\partial M}_{ijk\ell}$ are not independent! They are related by the celebrated *Gauss-Codazzi equations*—but we will not need them here.

The Gibbons–Hawking–York term. The extrinsic curvature turns out to be essentially the boundary term we need. We define the *Gibbons–Hawking-York term* to be

$$S_{\mathrm{GHY}}[h] = \frac{1}{8\pi G}\int_{\partial M} \mathrm{d}^{d-1}x\sqrt{-h}\,K, \qquad S_{\mathrm{tot}}[g,h] = S_{\mathrm{EH}}[g] + S_{\mathrm{GHY}}[h]. \tag{2.11}$$

It can then be shown, by a somewhat tedious calculation, that

$$\delta S_{\mathrm{tot}} = \delta(S_{\mathrm{EH}} + S_{\mathrm{GHY}}) = \int_M \mathrm{d}^d x\sqrt{-g}\,(\mathrm{EOM})_{ab}\delta g^{ab} - \frac{1}{16\pi G}\int_{\partial M}\mathrm{d}^{d-1}x\sqrt{-h}\,\Pi^{ij}\delta h_{ij}, \tag{2.12}$$

where $\Pi^{ij} \equiv K^{ij} - Kh^{ij}$. If we take $S_{\mathrm{tot}} = S_{\mathrm{EH}} + S_{\mathrm{GHY}}$ to be the gravitational action, then the calculation above shows that its variation produces, aside from the equations of motion, a boundary term that vanishes upon imposing the Dirichlet boundary condition $\delta h_{ij} = 0$. The quantity Π^{ij} is sometimes called the *gravitational momentum*, and on spacelike surfaces, it

is the variable canonically conjugate to the metric (in the sense of Hamilton-Jacobi theory). As we shall see, it is also closely related to the Brown–York boundary stress tensor. So now we are done: by taking $S_{\rm tot} = S_{\rm EH} + S_{\rm GHY}$ and specifying $\delta h = 0$ on ∂M, we have set up a good variational principle for gravity in a box.

2.3 The Brown–York Stress Tensor

Background structures. In physics, we make a distinction between *dynamical objects*, like pendulum bobs or scalar fields, and *background structures*, like the fixed background metric of Newtonian mechanics, or a choice of boundary conditions. This distinction gives rise to a broad philosophy: if you vary the dynamical fields while keeping the background data fixed, you get the equations of motion; and if you vary the background data while ensuring that the equations of motion are satisfied, you get conservation laws and symmetries.

The two parts of this mantra can be expressed in the following set of hieroglyphics:

$$\delta S = \int_M (\text{EOM})_a \underbrace{\delta\phi^a}_{\text{bulk variation}} + \int_{\partial M} \left(\begin{smallmatrix}\text{boundary}\\ \text{terms}\end{smallmatrix}\right)_i \underbrace{\delta\phi^i_\partial}_{\text{bdy variation}} = 0 \implies \begin{cases} (\text{EOM})_a = 0, & \text{bdy fixed} \\ (\text{bdy})_i = 0, & \text{on shell} \end{cases} \tag{2.13}$$

Setting $\delta\phi^i_\partial = 0$ fixes the boundary data, kills the second term, and yields the dynamics. Setting $\delta\phi^a = 0$ sets the fields to their EOM, kills the first term, and yields the symmetries.

We'll see this in action shortly. But first, recall that in ordinary field theories with a fixed (but possibly curved) background metric, we define the stress tensor by varying the action with respect to that background structure: $T_{ab} = -\frac{2}{\sqrt{-g}}\frac{\delta S}{\delta g^{ab}}$. This is a much more concrete way to see that varying a boundary condition can tell us about the theory's symmetries.

The Brown–York stress tensor. Brown and York were motivated to do something similar for gravity in a box, where the boundary metric h on ∂M is a background structure and a boundary condition. They defined the *Brown–York boundary stress tensor* by

$$T^{\rm BY}_{ij} = -\frac{2}{\sqrt{-h}}\frac{\delta S_{\rm tot}}{\delta h^{ij}} \qquad (\text{lives on } \partial M). \tag{2.14}$$

At first, it's not clear what $\delta S_{\rm tot}/\delta h^{ij}$ should mean: S is defined on all of M, but h is defined only on ∂M. It looks like one needs to choose an extension of h into M, and to show that $T^{\rm BY}_{ij}$ is independent of the choice of extension. But in practice, the extension is not necessary due to (2.12), which shows that the *on-shell* variation of $S_{\rm tot}$ depends only on h_{ij}:

$$\delta S_{\rm tot} = -\frac{1}{16\pi G}\int_{\partial M} \mathrm{d}^{d-1}x\sqrt{-h}\,\Pi^{ij}\,\delta h_{ij} \tag{2.15}$$

BY and GHY. Let us compare this variation with Brown and York's definition, which can be rearranged to read $\delta S_{\rm tot} = -\frac{1}{2}\sqrt{-h}\,T^{\rm BY}_{ij}\delta h^{ij}$. Upon using the identity $\delta h_{ij} = -h_{ik}h_{\ell j}\delta h^{k\ell}$

to lower indices [intuition for the proof: $\frac{\mathrm{d}}{\mathrm{d}x}\frac{1}{x} = -\frac{1}{x^2}$], we arrive at the celebrated result:

$$T^{\mathrm{BY}}_{ij} = -\frac{1}{8\pi G}\Pi_{ij}. \tag{2.16}$$

The physical picture is that for pure gravity in a box, the only stress-energy in the theory arises from the shape of the box ∂M, is defined only on ∂M, and moreover the "energy" stored in ∂M depends only on the extrinsic curvature of the embedding of ∂M within M. Importantly, however, T^{BY}_{ij} contains normal derivatives of g_{ab} at ∂M, and is not determined by the boundary data h_{ij}; instead it depends on the particular solution.

Covariant conservation. An important property of T^{BY}_{ij} is that it is divergence-free, i.e. covariantly conserved. Since it is defined on ∂M, where the metric is fixed, the usual proof goes through. Nevertheless, we will recapitulate it to illustrate the main idea. Let ξ be any vector field on M that is tangent to ∂M at the boundary (i.e. its flow fixes ∂M). Under the local diffeomorphism it generates, the metric transforms as $g_{ab} \longrightarrow g_{ab} + \nabla_a\xi_b + \nabla_b\xi_a$. In particular, we also have $h_{ij} \longrightarrow h_{ij} + D_i\xi^\partial_j + D_j\xi^\partial_i$, where D is the covariant derivative on ∂M, and where ξ^∂ is the restriction of ξ to ∂M. Now consider an on-shell variation of the metric g_{ab} along ξ. Substituting $\delta h = D_i\xi^\partial_j + D_j\xi^\partial_i$ in (2.15), we find

$$\delta S_{\mathrm{tot}} = -\frac{1}{16\pi G}\int_{\partial M}\mathrm{d}^{d-1}x\sqrt{-h}\,\Pi^{ij}\big(D_i\xi^\partial_j + D_j\xi^\partial_i\big) = \frac{1}{8\pi G}\int_{\partial M}\mathrm{d}^{d-1}x\big(D_i\Pi^{ij}\big)\xi^\partial_j = 0, \tag{2.17}$$

where we have used the symmetry of Π^{ij} and integrated by parts. The integral above vanishes for *all* suitable ξ, so we conclude that $D_i\Pi^{ij} = 0$. Thus diffeomorphism invariance guarantees that the BY stress tensor is covariantly conserved, as it should be.

Conserved charges. If, in addition to being tangent to the boundary, ξ_∂ is also a Killing symmetry of ∂M, then the BY stress tensor gives rise to a conserved Noether charge $Q[\xi]$ that corresponds to the symmetry of evolution along the ξ_∂ direction. It is defined by

$$Q[\xi] = \int_\Sigma \mathrm{d}^{d-2}x\sqrt{-h_\Sigma}\,n^a\xi^b_\partial T^{\mathrm{BY}}_{ab}, \tag{2.18}$$

where Σ is a Cauchy surface in ∂M, h_Σ is the induced metric on Σ, and n is the future-directed unit normal to Σ. The charge $Q[\xi]$ is conserved in the sense that it does not depend on the choice of Σ—that is, it is a global characterization of ∂M. For example, when ξ_∂ points in the time direction, $Q[\xi]$ is the ADM energy of the spacetime.

(As an aside, the formula above gives the standard notion of energy; however, for all other charges (momentum, etc.), an extra minus sign is typically introduced by convention.)