

# The Lagrangian Approach to General Relativity

*Nic Ford*

---

## 1 Introduction

---

This article is a supplement to another article on general relativity, which I wrote as part of my series on physics for mathematicians. We'll be discussing general relativity from the perspective of Lagrangian mechanics, a point of view I intentionally neglected in the main article.

The Lagrangian perspective is quite a bit more abstract, so I don't think it's the best approach to take when learning the subject for the first time. But if you have some sense of the shape of the theory in your head already, I think it can be a great way to clarify some of the conceptual underpinnings of the subject. In particular, while it's not in any way a prerequisite for the discussion we're about to have here, if you read and enjoyed "Electromagnetism as a Gauge Theory" from this series, my hope is that you'll like this one as well.

As far as prerequisites, it will be helpful to have some exposure to general relativity already — having read my article should be fine — and to understand how the Lagrangian approach to physics works in classical mechanics. We will be using the abstract index notation that was introduced in the main article, so if you don't know or don't remember how that works it might be useful to review it there.

This piece has two main goals. After a quick overview of how to extend the Lagrangian mechanics machinery to cover field theories in curved spacetime, we'll discuss how to express general relativity in the Lagrangian language — the Lagrangian which reproduces Einstein's equation is surprisingly concise, and it offers a useful perspective on the theory. We'll move from there to a discussion of an issue which I got somewhat stuck on when learning the theory: the relationship between the concept of energy-momentum as it appears in Einstein's equation and the concept as it appears in non-gravitational physics. I hope I can give an explanation that will prevent you from getting as confused about this point as I was.

In addition to the books cited in the main article, there are a couple of sources worth mentioning specifically for this supplement.

- The discussion on the relationship between the gravitational and inertial energy-momentum tensors is based on the first few pages of an article called "Canonical Pseudotensors, Sparling's Form, and Noether Currents" by László B. Szabados, which is mostly about the somewhat related topic of how one might define a quantity to represent the energy-momentum of the gravitational field itself.
- A very similar computation is also done in an article called "On the Energy-Momentum Tensor" by Ricardo E. Gamboa Saraví.

- Our discussion of the two energy-momentum tensors will be limited to the case of tensor fields coupled to gravity. The paper “Stress-Energy-Momentum Tensors and the Belinfante–Rosenfeld Formula” by Mark J. Gotay and Jerrold E. Marsden generalizes the procedure we’ll be discussing to cover essentially any situation in which diffeomorphisms of spacetime can be made to act on the fields.

I’m grateful to Jordan Watkins for helpful comments on an earlier draft, and many very helpful conversations on this topic.

---

## 2 Lagrangians in Curved Spacetime

---

We’ll start with a quick overview of how Lagrangian mechanics works for field theories in curved spacetime. While it is possible to set everything up in a very pleasingly coordinate-free way, I’m going to skip doing that here, since it doesn’t add enough to the discussion to be worth taking the time, opting instead for the weaker promise to write everything in such a way that it is usually not too difficult to translate each expression to its coordinate-free version. (I plan to return to the coordinate-free presentation in a future article in this series.)

### 2.1 Fields and Lagrangians

So, throughout this discussion, we’ll imagine that we’re working on a local coordinate patch  $U$  on a smooth oriented manifold  $M$  with coordinates  $x^1, \dots, x^n$ . The manifold  $M$  is meant to represent spacetime, so in most physical applications,  $n$  will be 4. Because we’re about to apply this to general relativity, we’ll assume  $M$  comes equipped with a metric  $g_{ab}$ , although for now we will simply take the metric to be fixed in advance rather than arising from Einstein’s equations.

The history of the physical system we’re modeling will be described by sections of some smooth fiber bundle  $\pi : E \rightarrow M$ . These sections are often called **fields**. A couple popular choices for  $E$  are  $M \times \mathbb{R}$  (giving us a **scalar field**) or the tangent bundle  $TM$  (giving us a **vector field**). The bundle  $E$  doesn’t necessarily have to be a vector bundle in general, although it will be for us. It’s also common for  $E$  to be a product of several different bundles, and in this case we think of each of the bundles in the product as corresponding to a different field.

For the purposes of this article, though, we’ll be specializing to the case of a single **tensor field**. This means, using the terminology and notation from the main article, that  $E$  is the  $(r, s)$ -tensor bundle  $T_s^r M$ . The coordinates on our patch  $U$  will then let us put coordinates on the fibers of  $E$  as well, which we could write  $y_{b_1 \dots b_s}^{a_1 \dots a_r}$ . For the sake of readability we’ll usually pack all of those indices into a “multi-index”  $A$ , writing an arbitrary coordinate simply as  $y^A$  and only expanding this out when absolutely necessary. (Despite appearing as a superscript, the multi-index  $A$  is meant to stand in for both the upper and lower indices of an  $(r, s)$ -tensor.)

As in classical mechanics, the laws of physics will be specified in terms of a **Lagrangian**, which is a real-valued function that depends on the values and derivatives of a section at a given point. In principle one could allow the Lagrangian to depend on the derivatives of a section up to  $k$ ’th order for some large  $k$ , but for simplicity we’ll mostly restrict to the case of first derivatives. In our present context, where our spacetime comes equipped with an arbitrary metric, these should be *covariant* derivatives, since this will make it easier to make sure the expression for the Lagrangian respects the symmetries of general relativity.

In other words, we'll think of our Lagrangian as a function of the form  $L(x^a, y^A, v_a^A)$ . (In an expression like this, think of  $x^a$  or  $y^A$  as standing in for all the possible coordinates of that form.) The  $x^a$ 's are coordinates on the base  $M$ , the  $y^A$ 's are coordinates in the fiber, and the  $v_a^A$ 's are coordinates which represent the derivatives of a section of  $E$ , in a way we'll now make precise.

In the coordinate-free version of this story, the  $x^a$ 's,  $y^A$ 's, and  $v_a^A$ 's would be coordinates on the **first jet bundle** of  $E$ , written  $J^1E$ . (Because we've restricted ourselves to a coordinate patch, we don't actually need a global definition of this object, and I will in fact hold off on this until the promised future article. For now we will content ourselves with the local description we've already given.) Given any section  $\phi : M \rightarrow E$ , we can produce a section  $j^1\phi : M \rightarrow J^1E$  called the **first jet prolongation** of  $\phi$  by setting

$$j^1\phi(x) = (x, \phi^A(x), \nabla_a\phi^A(x)).$$

As a quick example, if we had  $M = \mathbb{R}^2$  with the standard coordinates  $x^1, x^2$  and  $E = T_0^2M$ , then the total space of  $E$  would be 6-dimensional, with coordinates  $x^1, x^2, y^{11}, y^{12}, y^{21}, y^{22}$ . The total space of the jet bundle  $J^1E$  would then be 14-dimensional: the bundle  $J^1E \rightarrow E$  has 8-dimensional fibers, with coordinates  $v_k^{ij}$  for all choices of  $i, j, k$  from  $\{1, 2\}$ , and the bundle  $J^1E \rightarrow M$  has 10-dimensional fibers.

We'll often have occasion to talk about covariant derivatives of partial derivatives of  $L$ , so it's worth taking a moment to be clear about what such expressions mean. Suppose for simplicity that we just have one  $(r, s)$ -tensor field so that  $E = T_s^r M$ . Given any section  $\phi(x) = (x, \phi^A(x))$ , I encourage you to convince yourself that

$$\frac{\partial L}{\partial y^A}(x, \phi^A(x), \nabla_a\phi^A(x))$$

should naturally be thought of as an  $(s, r)$ -tensor: the natural way to get a number out of it is to pair it with an  $(r, s)$ -tensor, thought of as a tangent vector at  $\phi(x)$  in the fiber of  $E$ . When we write expressions like  $\nabla_a(\partial L/\partial y^A)$ , we will therefore mean the covariant derivative of  $\partial L/\partial y^A$  thought of as an  $(s, r)$ -tensor. Similarly,

$$\frac{\partial L}{\partial v_a^A}(x, \phi^A(x), \nabla_a\phi^A(x))$$

is naturally an  $(s+1, r)$ -tensor.

In the physics literature, you will usually see  $\partial L/\partial\phi^A$  for what I'm calling  $\partial L/\partial y^A$ , and  $\partial L/\partial(\nabla_a\phi^A)$  for what I'm calling  $\partial L/\partial v_a^A$ . These expressions (especially the second one) are a common source of confusion for people learning this theory for the first time, so I'm deliberately avoiding them, but it's worth knowing the standard notation in case you encounter it elsewhere.

## 2.2 The Euler–Lagrange Equations

Just as in classical mechanics, specifying a Lagrangian determines the laws of physics that our fields have to satisfy via an **action principle**, in the following way. Given any compact region of spacetime  $D \subseteq M$  and any section  $\phi$ , we'll define the **action** to be the quantity

$$S_D[\phi] = \int_D L(j^1\phi(x))\omega_g,$$

where  $\omega_g$  is the volume form on  $M$  arising from our metric  $g_{ab}$ . We'll define a **variation of  $\phi$  on  $D$**  to be a one-parameter family of sections  $\phi_u$  such that  $\phi_0 = \phi$  and every  $\phi_u = \phi$  on  $\partial D$ .

We then declare that the physically realizable field histories  $\phi$  are the ones with the property that, for every compact region  $D$  and every variation  $\phi_u$  on  $D$ , we have

$$\left. \frac{d}{du} S_D[\phi_u] \right|_{u=0} = 0.$$

When we have a variation in mind, we will follow the common convention in physics of writing  $\delta$  for  $(d/du)|_{u=0}$  and omitting the subscript  $u$  on other expressions when we do so. For example,  $\delta\phi^A$  means  $(d/du)(\phi_u)^A|_{u=0}$ , and the condition on the action just described can be written “ $\delta S_D = 0$ .”

As in classical mechanics, it’s possible to rewrite this condition in a way which doesn’t refer to variations. Suppose we have a  $\phi$  which satisfies our condition. Then, for every region  $D$  and every variation  $\phi_u$  on  $D$ , we have

$$0 = \delta S_D = \int_D \left( \frac{\partial L}{\partial y^A} \delta\phi^A + \frac{\partial L}{\partial v_a^A} \delta(\nabla_a \phi^A) \right) \omega_g.$$

(The derivatives of  $L$  are still being evaluated at  $j^1\phi(x)$ .) The derivatives with respect to  $u$  and  $x^a$  commute, so  $\delta(\nabla_a \phi^A) = \nabla_a(\delta\phi^A)$ . We can therefore write

$$\begin{aligned} 0 &= \int_D \left( \frac{\partial L}{\partial y^A} \delta\phi^A + \frac{\partial L}{\partial v_a^A} \nabla_a(\delta\phi^A) \right) \omega_g \\ &= \int_D \left( \frac{\partial L}{\partial y^A} - \nabla_a \frac{\partial L}{\partial v_a^A} \right) \delta\phi^A \omega_g + \int_D \nabla_a \left( \frac{\partial L}{\partial v_a^A} \delta\phi^A \right) \omega_g. \end{aligned}$$

The second integral is a covariant divergence integrated against the volume form. By the covariant-derivative analogue of the divergence theorem, this is equal to an integral on the boundary  $\partial D$ , which vanishes because  $\delta\phi^A$  is zero there. We’re left with just the first integral. In order for this integral to vanish for all possible choices of region  $D$  and variation  $\delta\phi^A$ , we must have that

$$\text{EL}_A(\phi) := \frac{\partial L}{\partial y^A}(j^1\phi(x)) - \nabla_a \left( \frac{\partial L}{\partial v_a^A}(j^1\phi(x)) \right) = 0$$

for all  $A$ . These are called the **Euler–Lagrange equations**.

### 2.3 Noether Currents

Another piece of the classical-mechanical Lagrangian story that generalizes nicely to this setting is the relationship between infinitesimal symmetries and conserved quantities from Noether’s Theorem.

There are a few ways to formalize these ingredients, and we’re going to go through just one. We’ll say that an **infinitesimal transformation** of sections of  $E$  is a smooth function  $w^A(x^a, y^A, v_a^A)$  which returns a vertical tangent vector at  $(x^a, y^A) \in E$ . (Recall that a tangent vector  $v$  at a point  $e \in E$  is *vertical* if  $\pi_* v = 0$  in  $T_{\pi(e)}M$ .) In other words,  $w^A$  takes in information about the value and derivative of a section at a point  $x$ , and it tells you a direction in which to push the value of the section at that same point. By starting with a section  $\phi$  and flowing along  $w^A$ , we can produce a **continuous transformation** of  $\phi$ , that is, a one-parameter group of sections  $\phi_u(x) = (x, (\phi_u)^A(x))$  such that

$$\frac{d}{du} (\phi_u)^A(x) = w^A(j^1\phi_u(x)).$$

To head off a confusion I had when encountering this for the first time, while it looks like the requirement that  $w^A$  be vertical would prevent us from considering things like spacetime translations, this is not actually the case. If, for example,  $X^a$  is a vector field on our spacetime  $M$ , then flowing sections along  $X^a$  is the continuous transformation that arises when we set  $w^A = \mathcal{L}_X \phi^A$  where  $\mathcal{L}$  denotes the Lie derivative. Lie derivatives of tensor fields only depend on the first derivatives of the field, so we can express the right side as a function of  $x^a$ ,  $y^A$ , and  $v_a^A$  as required.

In order to make Noether's Theorem work, we'll need a formal definition of an "infinitesimal symmetry" of a Lagrangian system: we want a condition we can impose on  $w^A$  which will imply that the corresponding continuous transformation takes solutions of the Euler–Lagrange equation to solutions. A natural first guess might be that we should call  $w^A$  an infinitesimal symmetry whenever the continuous symmetry preserves the Lagrangian, that is, whenever for any section  $\phi$  we have  $(d/du)L(j^1\phi_u(x)) = 0$  for all  $x$ , where  $\phi_u$  is the one-parameter group of sections we just defined.

In fact, this turns out to be too restrictive — for example, this won't be true for spatial translations in situations where that is a symmetry of the physical system. Luckily, as I encourage you to verify, it's enough if  $(d/du)L(j^1\phi_u(x))$  is a **total divergence**, that is, if there's a smooth function  $F^a(x^a, y^A, v_a^A)$  returning a tangent vector at  $x^a \in M$  such that

$$\frac{d}{du}L(j^1\phi_u(x)) = \nabla_a(F^a(j^1\phi_u(x))).$$

This is equivalent to requiring

$$\frac{\partial L}{\partial y^A}w^A + \frac{\partial L}{\partial v_a^A}(\nabla_a w^A) = \nabla_a F^a.$$

(Although we've stopped including it in the notation, all of these functions are still being evaluated at  $j^1\phi(x)$ !) When this happens, we'll call  $w^A$  an **infinitesimal symmetry** and the corresponding continuous transformation a **continuous symmetry**.

From this perspective, Noether's Theorem is quite simple to prove. Observe that, when  $w^A$  is an infinitesimal symmetry, we have

$$\begin{aligned} 0 &= \frac{\partial L}{\partial y^A}w^A + \frac{\partial L}{\partial v_a^A}(\nabla_a w^A) - \nabla_a F^a \\ &= \text{EL}_A(\phi)w^A + \nabla_a \left( \frac{\partial L}{\partial v_a^A}w^A - F^a \right). \end{aligned}$$

Inspired by this, let's define the **Noether current** corresponding to our symmetry  $w^I$  to be

$$J^a = F^a - \frac{\partial L}{\partial v_a^A}w^A.$$

What we've just shown is that, if you take a solution  $\phi$  of the Euler–Lagrange equations and plug in its values and derivatives as the  $y^A$ 's and  $v_i^A$ 's, then we'll have  $\nabla_a J^a = 0$ . In other words,  $J^a$  will be a *conserved current* in the sense discussed in the main article.

It's important to note here that this procedure does not actually uniquely define  $J^a$  — if  $\tilde{J}^a$  satisfies  $\nabla_a [J^a(j^1\phi(x)) - \tilde{J}^a(j^1\phi(x))] = 0$  for all sections  $\phi$ , then  $\tilde{J}^a$  will be an equally valid conserved current, and nothing about what we just did gives us a way to distinguish between them. (Indeed, the same ambiguity arises in the the definition of  $F^a$ .) When two currents  $J^a$  and

$\tilde{J}^a$  are related in this way, we'll call them **equivalent**, and for this reason, it is probably better to talk about *a* Noether current corresponding to a given symmetry rather than *the* Noether current. In particular, if you integrate  $J^a$  or  $\tilde{J}^a$  over a spacelike hypersurface to produce a conserved scalar quantity, the result will be the same. This will turn out to be important when we discuss the various ways to interpret the energy-momentum tensor.

As mentioned above, there are a few other choices one could make in the process of formalizing all this. One that you might see in the literature (including in the Gotay–Marsden paper mentioned in the introduction) is to remove the requirement that  $w^A$  be a vertical vector, that is, allowing our infinitesimal transformations to also point in spacetime directions. In order for the corresponding continuous transformation to still take sections of  $E$  to sections, we need to amend the action on sections a bit to drag each point of  $E$  back to the fiber where it belongs. This makes some symmetries a bit easier to write down, at the cost of making the formula for  $J^a$  a bit more complicated.

---

### 3 The Einstein–Hilbert Action

---

In general relativity, gravity is described in terms of the metric  $g_{ab}$  on our spacetime  $M$ , which we “promote” into a dynamical variable. That is, rather than think of the metric as simply a part of the specification of our background spacetime  $M$ , we treat it as one of the quantities that we are interested in solving for, just like the fields.

Given the framework we just laid out, this suggests that we need to include the metric in our Lagrangian somehow, and to require the derivative of the action to vanish under variations of both the metric and the fields. This requires a couple of small changes to our setup. First, in our previous discussion, the Lagrangian could depend on the metric indirectly via its dependence on the point  $x \in M$ . But now that the metric is a dynamical variable we're going to need to include it explicitly as a parameter. Second, we'll need to find a Lagrangian involving only the metric which we can add on to the Lagrangian for matter and which will produce Einstein's equation as one of the Euler–Lagrange equations. (If you read “Electromagnetism as a Gauge Theory,” it might be interesting to compare what we're about to do to the way electromagnetism was incorporated into the Lagrangian in that context.)

Our action will therefore take the form

$$S_D[\phi^A, g_{ab}] = (S_m)_D[\phi^A, g_{ab}] + (S_{EH})_D[g_{ab}].$$

The  $m$  stands for “matter,” and  $S_m$  is an action of the same type that we were considering in the previous section, except with any dependence on the metric now occurring through the metric's status as a parameter to the Lagrangian. The second term is called the **Einstein–Hilbert action**, and it takes a remarkably simple form: the action that reproduces Einstein's equation turns out to be

$$(S_{EH})_D[g_{ab}] = \frac{1}{16\pi G} \int_D R \omega_g,$$

where  $R$  is the scalar curvature of the metric and  $\omega_g$  is, as it was above, the volume form.

This action doesn't quite fit into our earlier framework for a couple of reasons. First, the curvature of a metric actually depends on its *second* derivatives, whereas the Lagrangians we considered earlier were only allowed to depend on first derivatives. We did this for simplicity, but there is actually no reason that one couldn't run exactly the same arguments with a Lagrangian that depends on derivatives up to any fixed order; the logic is essentially unchanged, but the

formulas for the Euler–Lagrange equations and the Noether current are a bit more complicated. (The Szabados paper mentioned in the introduction goes into this a bit.) Second, all covariant derivatives of the metric are zero, and so it would not make sense to use them as parameters. To address both of these concerns, we can think of the integrand in the Einstein–Hilbert action as a function of  $\partial_c g_{ab}$  and  $\partial_c \partial_d g_{ab}$ , the first and second partial derivatives of the metric in our chosen coordinate system.

By the linearity of partial derivatives, in order to check that the derivative of the action is zero under simultaneous variations of the fields and the metric, it's enough to consider variations of the fields and variations of the metric separately. When we vary just the fields, the variation of the Einstein–Hilbert action vanishes, and so we are back in the situation we considered in the previous section. The resulting Euler–Lagrange equations will therefore be exactly the same as the ones we would have gotten if we had thought of our system as living in a spacetime with a fixed metric.

Varying the metric is more interesting. We can't just throw everything into our formula for the Euler–Lagrange equation, because the dependence of the action on  $g_{ab}$  doesn't take the form we assumed when we derived it. In addition to the issues involving the derivatives of the metric that we just discussed, there is also the fact that the covariant derivatives  $\nabla_a \phi^A$  themselves depend on the metric, as does the volume form  $\omega_g$ .

To mimic the proof of the Euler–Lagrange equations in this new context, we would want to find some function  $\delta S / \delta g_{ab}$  depending on the values and derivatives of the fields and the metric such that, for any region  $D$  and any variation  $(g_u)_{ab}$  of the metric on  $D$ , we have

$$\delta S_D = \int_D \frac{\delta S}{\delta g_{ab}} (\delta g_{ab}) \omega_g.$$

(Recall that  $\delta$  refers to the derivative with respect to  $u$ .) We would then be able to conclude, just like in our proof of the Euler–Lagrange equations, that if  $\delta S_D = 0$  for all such variations, then  $\delta S / \delta g_{ab} = 0$ .

Such a  $\delta S / \delta g_{ab}$ , when it exists, is called the **functional derivative** of  $S_D$  with respect to  $g_{ab}$ . (The notation is meant to be reminiscent of the formula  $df = \sum_{i=1}^n (\partial f / \partial x_i) dx_i$  for a real-valued function  $f$  on  $\mathbb{R}^n$ . Note also that, using this notation, we could write the Euler–Lagrange expression from the last section simply as  $\text{EL}_A(\phi) = \delta S / \delta \phi^A$ .) Assuming that the functional derivatives of both the matter and Einstein–Hilbert actions exist and that we are able to compute them, we could write the resulting equations of motion in the form

$$\frac{\delta S_{EH}}{\delta g_{ab}} = - \frac{\delta S_m}{\delta g_{ab}}.$$

(If you read about this story elsewhere in the literature — including in Carroll's book — you should be aware that it's very common to adopt a different convention for functional derivatives, where they would instead say, for example,  $\delta S_D = \int_D \frac{\delta S}{\delta \phi} \delta \phi d^4 x$ . In other words, they use  $d^4 x$  in the position where we had the volume form  $\omega_g$ . This will make some formulas look a bit different.)

The computation of  $\delta S_{EH} / \delta g_{ab}$  is fairly straightforward but pretty long, so we won't go through it here. It can be found in most textbooks on the subject, and in particular in Appendix E of Wald and Section 4.3 of Carroll. The result is

$$\frac{\delta S_{EH}}{\delta g_{ab}} = - \frac{1}{16\pi G} (R^{ab} - \frac{1}{2} R g^{ab}).$$

This is (apart from the factor out front and the raised indices) the quantity that appears on the left side of Einstein's equation, which is a promising sign! In particular, we can at least conclude from this that, in the absence of matter, we do indeed recover the vacuum version of Einstein's equation. The extent to which we can claim to recover Einstein's equation in the *presence* of matter depends on whether we can conclude that the energy-momentum tensor of the matter is given by

$$T^{ab} = 2 \frac{\delta S_m}{\delta g_{ab}}.$$

In most textbook presentations of this part of the story, including the two I just cited, they essentially stop here and declare victory, stating that the quantity on the right side of this equation *is* the energy-momentum tensor and that we have therefore successfully reproduced Einstein's equation from a Lagrangian.

In my opinion, this leaves a fairly big question unanswered: why should we believe that  $2\delta S_m/\delta g_{ab}$  has anything to do with energy-momentum as we understood it before we knew about Einstein's equation? This is the question we'll take on in the next section.

---

## 4

 The Two Energy-Momentum Tensors
 

---

We can state the central issue as follows. In general relativity, energy-momentum plays two quite different roles, and it will be useful to have some terminology available to distinguish them from each other. Let's call it **inertial energy-momentum** when it plays the role it plays in non-gravitational physics. The inertial energy-momentum of a massive particle, for instance, is its mass times its 4-velocity. And we'll call it **gravitational energy-momentum** when it serves as the source for gravity. In other words, gravitational energy-momentum is the quantity that goes on the right side of Einstein's equation.

One of the central claims of general relativity is, of course, that these two quantities are identical; this is one of the things that makes gravity unique among the fundamental forces of nature. The argument from the last section certainly provides a justification for identifying  $2\delta S_m/\delta g_{ab}$  with gravitational energy-momentum, but if we're going to claim that our Lagrangian story has reproduced general relativity, it would be nice if we could find an argument for also identifying this quantity with inertial energy-momentum. Specifically, we have a preexisting characterization of inertial energy-momentum from our knowledge of Lagrangian mechanics in non-gravitational physics: it's the Noether current corresponding to spacetime translations. What does that Noether current have to do with  $2\delta S_m/\delta g_{ab}$ ?

It is worth saying at the outset that there is a weaker claim we could make with much less effort than we're about to expend. The natural generalization of a spacetime translation to curved spacetime is the flow along a vector field. As we'll discuss in more detail momentarily, flowing the fields along a vector field isn't a symmetry in general, but it is when we flow along a *Killing vector field*, that is, when the flows are isometries. In this case, we can quite easily extract a conserved current from the covariant conservation law  $\nabla_a T^{ab} = 0$ , which follows from Einstein's equation: the vector field  $K^a$  is Killing if and only if  $\nabla_a K_b = -\nabla_b K_a$ , and this, combined with the symmetry of  $T^{ab}$ , gives us that  $\nabla_a (K_b T^{ab}) = 0$ .

This does give us a way to turn our symmetry into a current that is conserved, but in my view it's not a fully satisfying answer to the question at hand. When I learned this for the first time, I wanted to see directly how the quantity that arises from applying the Noether machine to spacetime translations is related to the quantity that shows up on the right side of Einstein's



equation, and the fact we just established doesn't get us there on its own. Our goal in this section is to fill in this gap.

## 4.1 Gravitational Energy-Momentum

In order to perform this comparison, we'll need to find explicit formulas for the two energy-momentum tensors. We'll start with the gravitational energy-momentum tensor, which will require introducing a bit of notation that I'm borrowing from the Szabados paper mentioned in the introduction. There are many formulas in differential geometry, usually involving various sorts of derivatives, which involve a sum over all possible ways of replacing one of the indices on the tensor being differentiated with a new index. For example, given a vector field  $X^a$ , the Lie derivative  $\mathcal{L}_X T^{a_1 \dots a_r}_{b_1 \dots b_s}$  of an  $(r, s)$ -tensor can be written

$$\begin{aligned} \mathcal{L}_X T^{a_1 \dots a_r}_{b_1 \dots b_s} = & X^c \nabla_c T^{a_1 \dots a_r}_{b_1 \dots b_s} - (\nabla_c X^{a_1}) T^{c a_2 \dots a_r}_{b_1 \dots b_s} \\ & - \dots - (\nabla_c X^{a_r}) T^{a_1 \dots a_{r-1} c}_{b_1 \dots b_s} \\ & + (\nabla_{b_1} X^c) T^{a_1 \dots a_r}_{c b_2 \dots b_s} \\ & + \dots + (\nabla_{b_s} X^c) T^{a_1 \dots a_r}_{b_1 \dots b_{s-1} c}. \end{aligned}$$

We'll need to talk about a few of these formulas in what follows, and so it will be very helpful to have a concise way to write these index games. Recall that  $\delta_b^a$  is the  $(1, 1)$ -tensor corresponding to the identity map; in other words, in coordinates, it's equal to 1 when  $a = b$  and 0 otherwise. If we then define

$$\begin{aligned} \Delta_f^{e a_1 \dots a_r d_1 \dots d_s}_{b_1 \dots b_s c_1 \dots c_r} = & (\delta_{c_1}^e \delta_f^{a_1} \delta_{c_2}^{a_2} \dots \delta_{c_r}^{a_r} + \dots + \delta_{c_1}^{a_1} \dots \delta_{c_{r-1}}^{a_{r-1}} \delta_{c_r}^e \delta_f^{a_r}) \delta_{b_1}^{d_1} \dots \delta_{b_s}^{d_s} \\ & - \delta_{c_1}^{a_1} \dots \delta_{c_r}^{a_r} (\delta_{b_1}^e \delta_f^{d_1} \delta_{b_2}^{d_2} \dots \delta_{b_s}^{d_s} + \dots + \delta_{b_1}^{d_1} \dots \delta_{b_{s-1}}^{d_{s-1}} \delta_{b_s}^e \delta_f^{d_s}), \end{aligned}$$

we can pack our expression for the Lie derivative into just two terms:

$$\mathcal{L}_X T^{a_1 \dots a_r}_{b_1 \dots b_s} = X^c \nabla_c T^{a_1 \dots a_r}_{b_1 \dots b_s} - (\nabla_e X^f) \Delta_f^{e a_1 \dots a_r d_1 \dots d_s}_{b_1 \dots b_s c_1 \dots c_r} T^{c_1 \dots c_r}_{d_1 \dots d_s}.$$

(Note that the definition of  $\Delta$  depends on the values of  $r$  and  $s$ , not just on the two lists of  $r + s + 1$  indices attached to  $\Delta$ ; a different way of splitting  $r + s$  into two nonnegative integers will result in a change of sign on some of the terms. Also, note that the above formula suggests that, if  $r = s = 0$ , suggests we should set  $\Delta_f^e = 0$ .)

This is definitely an improvement, but still kind of a notational handful. We can cut down on the indices considerably by reintroducing the "multi-index" notation we were using before. We'll adopt the further convention that, if a superscript  $A$  stands for the indices that belong on an  $(r, s)$ -tensor, then a subscript  $A$  stands for the indices of an  $(s, r)$ -tensor. Using this, we can write the previous formula as

$$\mathcal{L}_X T^A = X^c \nabla_c T^A - (\nabla_e X^f) \Delta_{fB}^{eA} T^B.$$

There are a lot of facts about derivatives of tensors that can similarly be written in this way. For example, we have

$$\begin{aligned} \nabla_a T^A &= \partial_a T^A + \Gamma_{ac}^b \Delta_{bB}^{cA} T^B \\ [\nabla_a, \nabla_b] T^A &= R^d{}_{cab} \Delta_{dB}^{cA} T^B \\ [\mathcal{L}_X, \nabla_a] T^A &= (R^c{}_{bda} X^d + \nabla_a \nabla_b X^c) \Delta_{cB}^{bA} T^B. \end{aligned}$$

Let's write  $T_{\text{grav}}^{ab} = 2\delta S_m / \delta g_{ab}$  for our gravitational energy-momentum tensor. If you pick an arbitrary variation of the metric and consider

$$\delta S_m = \int_D \delta(L_m \omega_g) = \int_D \left[ L_m \delta \omega_g + \left( \frac{\partial L_m}{\partial v_a^A} \delta(\nabla_a \phi^A) + \frac{\partial L_m}{\partial g_{ab}} \delta g_{ab} \right) \omega_g \right],$$

it's a nice exercise (also done, after some translation between our notations, on p. 10 of the Saraví paper) to show that this produces the formula

$$T_{\text{grav}}^{ab} = 2 \frac{\partial L_m}{\partial g_{ab}} + L_m g^{ab} + \frac{1}{2} \nabla_c (\Sigma^{abc} + \Sigma^{bac} - \Sigma^{acb} - \Sigma^{bca} - \Sigma^{cab} - \Sigma^{cba}),$$

where

$$\Sigma^{ij}_k = \frac{\partial L_m}{\partial v_i^A} \Delta_{kB}^{jA} \phi^B.$$

(If you try to prove this, remember that any integral of a total divergence is zero, and that any expression of the form  $T^{ab} \delta g_{ab}$  can also be written  $\frac{1}{2}(T^{ab} + T^{ba}) \delta g_{ab}$  because of the symmetry of  $g_{ab}$ .)

## 4.2 Inertial Energy-Momentum

Let's now turn to the inertial energy-momentum tensor. As mentioned above, in non-gravitational physics on flat spacetime, we have a clear picture of what inertial energy-momentum ought to mean: it's the Noether current corresponding to spacetime translations. In curved spacetime, of course, there is no reason to think that spacetime translation is any sort of symmetry at all; the metric is different at different points, so simply translating the fields across spacetime will certainly affect the value of the Lagrangian.

You might instead consider translating the fields *and* the metric. Unfortunately, while we won't go over this in detail here, Noether's Theorem turns out to be vacuous in this case. (This is probably to be expected — as we discussed in the main article, the covariant conservation law  $\nabla_a T^{ab}$  does not imply the existence of any conserved currents in general.) The main exception is, as we mentioned above, when our metric has a Killing vector field. In this case, we do get a symmetry by flowing along this vector field, and we can apply the Noether machine to get a nontrivial conserved current.

So for the moment let's suppose  $K^a$  is a Killing vector field. In order to ensure that flowing the fields along  $K^a$  is actually a symmetry, we'll also assume going forward that that our Lagrangian doesn't depend directly on the point  $x \in M$ .

We'll start by writing an expression for the derivative of the matter Lagrangian in the direction of  $K^a$ . Regardless of whether or not  $K^a$  is Killing, we have

$$K^a \nabla_a L_m = \frac{\partial L_m}{\partial y^A} \mathcal{L}_K \phi^A + \frac{\partial L_m}{\partial v_a^A} \mathcal{L}_K (\nabla_a \phi^A) + \frac{\partial L_m}{\partial g_{ab}} \mathcal{L}_K g_{ab}.$$

At first glance, this doesn't look like it fits into our Noether current story. First, for our definition of "infinitesimal symmetry," the expression multiplying  $\partial L_m / \partial v_a^A$  has to be  $\nabla_a$  of the expression multiplying  $\partial L_m / \partial y^A$ , which seems not to be the case here. Second, we need the left side to be a total divergence, which it seems not to be. And third, we have the extra term involving  $g_{ab}$ .

Luckily, all of these issues are solved when  $K^a$  is Killing. By the definition of a Killing vector field, we have  $\mathcal{L}_K g_{ab} = 0$ , so the last term vanishes. Furthermore,  $\mathcal{L}_K$  commutes with  $\nabla_a$

whenever  $K^a$  is Killing. (One quick way to see this is to use the definition of the Lie derivative in terms of flows; since the flows are isometries, they preserve covariant derivatives.) We therefore can write

$$\begin{aligned} K^a \nabla_a L_m &= \frac{\partial L_m}{\partial y^A} \mathcal{L}_K \phi^A + \frac{\partial L_m}{\partial v_a^A} \nabla_a (\mathcal{L}_K \phi^A) \\ &= \text{EL}_A(\phi) \mathcal{L}_K \phi^A + \nabla_a \left( \frac{\partial L_m}{\partial v_a^A} \mathcal{L}_K \phi^A \right). \end{aligned}$$

Using the fact that  $K^a$  is Killing once again to conclude that  $\nabla_a K^a = 0$ , we see that

$$\nabla_a \left( K^a L_m - \frac{\partial L_m}{\partial v_a^A} \mathcal{L}_K \phi^A \right) = 0,$$

when  $\phi^A$  satisfies the Euler–Lagrange equations. In other words,

$$J^a = K^a L_m - \frac{\partial L_m}{\partial v_a^A} \mathcal{L}_K \phi^A$$

is a Noether current for to our symmetry.

What if, as will almost always be the case, there is no Killing vector field? It still seems like we ought to be able to talk about inertial energy-momentum, at least in an approximate sense. If I'm at some point  $x$ , and I never travel far enough away from  $x$  to notice the curvature of spacetime, I could produce an inertial energy-momentum tensor from my (technically erroneous, but approximately correct) belief that spacetime is flat near  $x$ , and it still makes sense to ask about the relationship between this quantity and the source for gravity.

One way to formalize this concept of “the coordinates you would use if you were ignorant of gravity” is through the use of *Riemannian normal coordinates*. Recall that, around every point of spacetime, we can find a coordinate patch such that, at our chosen point,  $g_{ab}$  is equal to the Minkowski metric and the covariant derivatives are given by ordinary partial derivatives. In this coordinate system, the coordinate translation vector fields are not Killing, but they are close; if we restrict to points whose coordinates differ from our chosen point by at most  $\epsilon$ , then the error — that is, the difference between  $\nabla_a J^a$  and zero — is on the order of  $\epsilon |R_{ab}| + O(\epsilon^2)$ .

### 4.3 The Comparison

With explicit expressions for our two energy-momentum tensors in hand, we're almost ready to see how they are related.

To facilitate this comparison, we'd like to get our expression for the Noether current into a form which contains a  $(2, 0)$ -tensor that we might somehow relate directly to  $T_{\text{grav}}^{ab}$ . I encourage you to verify that we can write

$$J^a = K^b \Theta^a_b + (\nabla_b K_c) \Sigma^{abc},$$

where

$$\Theta^a_b = \delta_b^a L_m - \frac{\partial L_m}{\partial v_a^A} \nabla_b \phi^A$$

is the so-called **canonical energy-momentum tensor**.

It's instructive to take a small detour to see what this expression looks like in the case of translations and rotations in Minkowski space. For translations,  $K^b$  will be a *constant* vector, so the second term in  $J^a$  vanishes, leaving us with just

$$J^a = K^b \Theta^a{}_b.$$

In particular, if we had started by considering a field theory in flat spacetime and considered only spacetime translations,  $\Theta^a{}_b$  is the quantity we might have naturally singled out as “the” energy-momentum tensor, so that for any vector  $V^b$ , the quantity  $V^b \Theta^a{}_b$  could be thought of as linear energy-momentum current density in the  $V^b$  direction.

I encourage you to convince yourself that rotation in the plane spanned by two constant spacelike vectors  $V^a$  and  $W^a$  is generated by the vector field  $K^b = x^c (V_c W^b - W_c V^b)$ , where  $x^c$  is the  $\mathbb{R}^4$ -valued function on spacetime which outputs the spacetime coordinates of each point, and that the resulting current is

$$J^a = x^c (V_c W^b - W_c V^b) \Theta^a{}_b + V_b W_c (\Sigma^{abc} - \Sigma^{acb}).$$

Using the interpretation of  $V^b \Theta^a{}_b$  we just discussed, the first term is naturally interpreted as **orbital angular momentum**, that is, the contribution to angular momentum arising from linear motion parallel to the plane of rotation and orthogonal to  $x^c$ . The second term would then be an “intrinsic” contribution to angular momentum, that is, one that doesn't arise from linear motion but from motion “internal” to the space in which the fields take their values. (Notice that, for a scalar field,  $\Sigma^{abc} = 0$ , so this contribution will vanish.) Physicists often call this **spin angular momentum**, and we'll adopt this terminology when we discuss this quantity a bit more in just a moment.

We're now finally ready to talk about how our two quantities are related. Following Szabados, we'll be aided by the following fact. Pick an arbitrary vector field  $X^a$ . Using the fact that  $\mathcal{L}_X g_{ab} = \nabla_a X_b + \nabla_b X_a$ , and still assuming that  $L_m$  doesn't depend directly on  $x$ , we can write

$$\begin{aligned} \nabla_a (L_m X^a) &= X^a \nabla_a L_m + \frac{1}{2} L_m g^{ab} \mathcal{L}_X g_{ab} \\ &= \text{EL}_A(\phi) \mathcal{L}_X \phi^A + \left( \frac{\partial L_m}{\partial g_{ab}} + \frac{1}{2} L_m g^{ab} \right) \mathcal{L}_X g_{ab} \\ &\quad + \nabla_a \left( \frac{\partial L_m}{\partial v_a^A} \mathcal{L}_X \phi^A \right) + \frac{\partial L_m}{\partial v_a^A} (\mathcal{L}_X (\nabla_a \phi^A) - \nabla_a (\mathcal{L}_X \phi^A)). \end{aligned}$$

After an honestly unreasonably long computation which I won't reproduce here (if you're interested, Saraví's paper goes through an essentially equivalent one) this equality can be digested into the following form:

$$\text{EL}_A(\phi) \mathcal{L}_X \phi^A + T_{\text{grav}}^{ab} (\nabla_a X_b) = \nabla_a (X^b \Theta^a{}_b + \tilde{\Sigma}^{abc} (\nabla_b X_c)),$$

where

$$\tilde{\Sigma}^{abc} = \frac{1}{2} (\Sigma^{abc} - \Sigma^{acb} + \Sigma^{bca} - \Sigma^{bac} + \Sigma^{cba} - \Sigma^{cab}).$$

Assume that the fields satisfy the Euler–Lagrange equations, so that the first term vanishes. Then, since the vector field  $X^a$  is arbitrary, the coefficients of  $X^c$ ,  $\nabla_b X^c$  and  $\nabla_a \nabla_b X^c$  on each side of this equation must all separately be equal, so we can conclude that

$$T_{\text{grav}}^{bc} = \Theta^{bc} + \nabla_a \tilde{\Sigma}^{abc}.$$

The expression on the right is called the **Belinfante–Rosenfeld formula** for the energy-momentum tensor. If we had started with the “canonical” energy-momentum tensor  $\Theta^{bc}$  — which we might have been led to do if we’d started by looking at spacetime translations in Minkowski space as above — then we can look at the second term as a “correction” to our energy-momentum tensor. Indeed,  $\Theta^{ab}$  is not especially close to being an adequate gravitational energy-momentum tensor; it isn’t even symmetric in general, and it will also often fail to be gauge-invariant in physical theories where that’s a relevant concern.

In Minkowski space, you can sort of picture the Belinfante–Rosenfeld formula as telling you that you need to include contributions from spin angular momentum in order to produce the quantity that can serve as a source for gravity, although that interpretation doesn’t generalize beyond that setting, because we needed to take advantage of the entire coordinate system on Minkowski space to identify  $\Theta^{ab}$  and  $\Sigma^{abc}$  with linear energy-momentum and spin respectively. In general,  $\Theta^{ab}$  and  $\tilde{\Sigma}^{abc}$  are not, by themselves, especially physically meaningful quantities, while  $T_{\text{grav}}^{ab}$  obviously is, and accordingly I think the perspective in which the Belinfante–Rosenfeld formula is seen as a way of correcting the deficiencies in  $\Theta^{ab}$  is a somewhat limited one.

Finally, let’s see what this tells us about our Noether current in the presence of a Killing field, which we had gotten into the form

$$J^a = K^b \Theta^a{}_b + (\nabla_b K_c) \Sigma^{abc}.$$

I encourage you to verify that, since  $\nabla_b K_c$  is antisymmetric (because  $K^c$  is Killing) this is equal to

$$J^a = K^b \Theta^a{}_b + (\nabla_b K_c) \tilde{\Sigma}^{abc},$$

and that the Belinfante–Rosenfeld formula then implies that

$$J^a = K_b T_{\text{grav}}^{ab} + \nabla_b (K_c \tilde{\Sigma}^{abc})$$

whenever the Euler–Lagrange equations are satisfied.

The tensor  $K_c \tilde{\Sigma}^{abc}$  is antisymmetric in  $a$  and  $b$ , and this implies that  $\nabla_a \nabla_b (K_c \tilde{\Sigma}^{abc}) = 0$ . (This fact, which is true of any antisymmetric tensor, is somewhat nonobvious but purely geometric, and it is worth working out on your own if it’s unfamiliar.) The currents  $J^a$  and  $K_b T_{\text{grav}}^{ab}$  therefore differ by a quantity that is divergence-free. In other words, in the language we used when we introduced Noether currents,  $J^a$  and  $K_b T_{\text{grav}}^{ab}$  are *equivalent* conserved currents for the symmetry arising from flowing along a Killing vector field.

This, finally, is the promised relationship between the gravitational and inertial energy-momentum tensors. They are not equal in general, just equivalent in the sense just described. (In a couple simple cases — most obviously for a scalar field, where  $\Sigma^{abc} = 0$  — we will have equality, but when this happens I think it should essentially be regarded as a coincidence.) Our  $J^a$  is the quantity that arises directly from Noether’s Theorem, but remember that Noether’s Theorem does not have the capacity to choose from among equivalent currents, and so we would be justified in saying that  $K_b T_{\text{grav}}^{ab}$  is a Noether current corresponding to our translation symmetry.

By contrast, the procedure that produced  $T_{\text{grav}}^{ab}$  *does* actually pick out that specific tensor once the Lagrangian has been specified. For this reason, it is best to think of this object, rather than anything involving  $\Theta^{ab}$  or  $\Sigma^{abc}$ , as “the” energy-momentum tensor whenever the distinction might matter: it is the quantity that serves as a source for gravity and also, in the presence of a Killing symmetry, it leads to one of the possible quantities that can be used as the corresponding Noether current.