

Arch and scaffold:

How Einstein found his field equations

Michel Janssen and Jürgen Renn

In his later years, Einstein often claimed that he had obtained the field equations of general relativity by choosing the mathematically most natural candidate. His writings during the period in which he developed general relativity tell a different story.

This month marks the centenary of the Einstein field equations, the capstone on the general theory of relativity and the highlight of Albert Einstein's scientific career.¹ The equations, which relate space-time curvature to the energy and momentum of matter, made their first appearance in a four-page paper submitted on 25 November 1915 to the Prussian Academy of Sciences in Berlin and reprinted in *The Collected Papers of Albert Einstein (CPAE)*,² volume 6, document 21. How did Einstein, shown in figure 1, arrive at those equations? He later insisted that the gravitational equations “could *only* be found by a purely formal principle (general covariance).”³ Such statements mainly served to justify his strategy in the search for a unified field theory during the second half of his career. As a description of how he found the field equations of general relativity, they are highly misleading.

The 25 November paper was the last in a series of short communications submitted to the Berlin Academy on four consecutive Thursdays that month (CPAE 6; 21, 22, 24, 25). In the first paper, Einstein replaced the field equations that he had published in 1913 with equations that retain their form under a much broader class of coordinate transformations (see figure 2). In the second, a highly speculative hy-

pothesis he adopted about the nature of matter allowed him to change those equations to equations that are generally covariant—that is, retain their form under arbitrary coordinate transformations. In the fourth, he achieved the same end by changing the field equations of the first paper in a different and more convincing way, as shown in figure 3. In the third, based on the field equations of the second paper but unaffected by the modification of the fourth, he accounted for the 43 seconds of arc per century missing in the Newtonian account of the perihelion motion of Mercury.

In the first November paper, Einstein made it sound as if he had gone from the old to the new field equations by leveling one cathedral and building a new one on its ruins in a completely different style. The former was built according to principles of physics; the latter, Einstein wanted his readers to believe, according to principles of mathematics. “It is a real triumph of the method of the general differential calculus,” he rhapsodized—even before his theory was generally covariant (CPAE 6; 21). Careful examination of the November papers and his correspondence at the time suggests a different architectural metaphor.⁴ Einstein used the elaborate framework he had built around the field equations of 1913 as a scaffold on which he carefully placed the arch stones of the Einstein field equations.

Covariance lost

As the box on page 31 shows, Einstein had already considered the equations of his first November



Michel Janssen is a professor in the Program in the History of Science, Technology, and Medicine at the University of Minnesota in Minneapolis. **Jürgen Renn** is a director at the Max Planck Institute for the History of Science in Berlin.

Grossmann

$$\bar{T}_{il} = \sum_{\alpha\beta} \frac{\partial \{i\alpha\}}{\partial x_l} \frac{\partial \{l\beta\}}{\partial x_\alpha} - \frac{\partial \{i\alpha\}}{\partial x_\alpha} \frac{\partial \{l\beta\}}{\partial x_l} + \{i\alpha\} \{l\beta\} - \{i\beta\} \{l\alpha\}$$

Wenn G ein Skalar ist, dann $\frac{\partial g_{ik}}{\partial x_l} = T_{il}$ Tensor 2. Ranges.

$$\bar{T}_{il} = \left(\frac{\partial \{i\alpha\}}{\partial x_l} - \sum_{\alpha} \{i\alpha\} \frac{\partial \{l\beta\}}{\partial x_\alpha} \right) - \left(\sum_{\alpha} \{i\alpha\} \frac{\partial \{l\beta\}}{\partial x_\alpha} - \{i\alpha\} \{l\beta\} \right)$$

Tensor 2. Ranges

Vermutlicher Gravitations-tensor c_{il}

Weitere Umformung des Gravitations-tensors

$$\frac{\partial \{i\alpha\}}{\partial x_\alpha} = \frac{1}{2} \frac{\partial}{\partial x_\alpha} \left(g_{\alpha\alpha} \frac{\partial g_{il}}{\partial x_\alpha} + \frac{\partial g_{il}}{\partial x_\alpha} - \frac{\partial g_{il}}{\partial x_\alpha} \right)$$

Wir setzen voraus $\sum_{\alpha} \frac{\partial g_{\alpha\alpha}}{\partial x_\alpha} = 0$, dann ist dies gleich

$$-\sum_{\alpha} g_{\alpha\alpha} \frac{\partial^2 g_{il}}{\partial x_\alpha^2} + \sum_{\alpha} \left(\frac{\partial g_{\alpha\alpha}}{\partial x_\alpha} \frac{\partial g_{il}}{\partial x_\alpha} + \frac{\partial g_{\alpha\alpha}}{\partial x_l} \frac{\partial g_{il}}{\partial x_\alpha} \right)$$

Tensor $\{i\alpha\} \{l\beta\} = g_{\alpha\alpha} g_{\beta\beta} \left(\frac{\partial g_{il}}{\partial x_\alpha} - \frac{\partial g_{il}}{\partial x_\alpha} + \frac{\partial g_{il}}{\partial x_\alpha} \right) \left(\frac{\partial g_{\alpha\beta}}{\partial x_\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x_\beta} + \frac{\partial g_{\alpha\beta}}{\partial x_l} \right)$

$$= -g_{\alpha\alpha} g_{\beta\beta} \left(\frac{\partial g_{il}}{\partial x_\alpha} - \frac{\partial g_{il}}{\partial x_\alpha} \right) \left(\frac{\partial g_{\alpha\beta}}{\partial x_\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x_\beta} \right) + g_{\alpha\alpha} g_{\beta\beta} \frac{\partial g_{il}}{\partial x_l} \frac{\partial g_{\alpha\beta}}{\partial x_\alpha}$$

$\alpha \neq \beta$
 $\alpha \neq \beta$

oder $-\frac{\partial g_{il}}{\partial x_l} \frac{\partial g_{\alpha\beta}}{\partial x_\alpha}$

Hiernach

$$-\bar{T}_{il} = \sum_{\alpha} \left(g_{\alpha\alpha} \frac{\partial^2 g_{il}}{\partial x_\alpha^2} - g_{\alpha\alpha} g_{\beta\beta} \left(\frac{\partial g_{il}}{\partial x_\alpha} - \frac{\partial g_{il}}{\partial x_\alpha} \right) \left(\frac{\partial g_{\alpha\beta}}{\partial x_\alpha} - \frac{\partial g_{\alpha\beta}}{\partial x_\beta} \right) \right)$$

$$+ \sum_{\alpha} \left(\frac{\partial g_{\alpha\alpha}}{\partial x_l} \left[\frac{\partial g_{il}}{\partial x_\alpha} \right] + \frac{\partial g_{\alpha\alpha}}{\partial x_l} \left[\frac{\partial g_{il}}{\partial x_\alpha} \right] \right) + \sum_{\alpha} \frac{1}{4} \frac{\partial g_{\alpha\alpha}}{\partial x_l} \frac{\partial g_{\alpha\alpha}}{\partial x_l}$$

- 44 -

General relativity under construction

The so-called Zürich notebook contains Albert Einstein's research notes of 1912–13 as, with the help of Marcel Grossmann, he began looking for field equations for his new theory of gravity.⁶ At the top of this particular page, 22R, Einstein wrote down the generally covariant Ricci tensor T_{il} under the heading "Grossmann," presumably because Grossmann had suggested it to him. "If [the determinant of the metric] G is a scalar" (Wenn G ein Skalar ist), Einstein noted, the first half of T_{il} transforms as a tensor. Coordinate transformations satisfying that determinant condition are now called unimodular transformations. If the first half transforms as a tensor under unimodular transformations, the second half must too, since their sum transforms as a tensor under arbitrary transformations. Underneath the second half of the second equation, Einstein wrote, "probable gravitation tensor" (Vermutlicher Gravitationstensor).

Thus in 1912–13 Einstein had already considered the field equations he proposed in the first of the four papers he submitted to the Prussian Academy of Sciences in Berlin in November 1915 (see figure 2). Near the middle of the page, Einstein imposed the condition $\Sigma_{\alpha} \partial g_{\alpha\alpha} / \partial x_\alpha = 0$, in modern notation $\partial_\mu g^{\mu\nu} = 0$. That is the same condition he imposed in the first November 1915 paper to show that those field equations reduce to their Newtonian counterpart in the case of weak static fields. (Courtesy of the Albert Einstein Archives, Hebrew University of Jerusalem.)

paper three years earlier in the course of his collaboration with mathematician Marcel Grossmann,⁵ shown in figure 4. The two of them had been classmates at what is now ETH Zürich and were reunited at their alma mater in July 1912. Notes of their collaboration have been preserved in the so-called Zürich notebook.⁶ In the notebook, as Einstein recalled in the introduction of his first November paper, they had given up the search for field equations based on the Riemann tensor "with a heavy heart." What had defeated them were problems with the physical interpretation of such equations. Eventually they adopted field equations specifically engineered to avoid those problems and published them in a joint paper in June 1913 (CPAE 4; 13). The theory and the field equations they presented are known, based on the title of that paper, as the *Entwurf* (outline or draft) theory and the *Entwurf* field equations.

Except for its field equations, the *Entwurf* theory has all the basic elements of the mathematical formalism of general relativity. Einstein nonetheless cautiously referred to it as a generalized theory of relativity—not a general theory—because of the limited covariance of the *Entwurf* field equations.

Part of Einstein's difficulty was that throughout the period from late 1912 to late 1916, he conflated general covariance and general relativity of motion. A map analogy illustrates the difference. The shortest route between two cities is a segment of the great circle connecting them. Suppose you have a map on which that route does not correspond to a straight line. By switching to a different map (probably one

that is not very practical) you can turn the route between the two cities into a straight line. That ability is the analogue of general covariance. Nonetheless, no matter what it looks like on any given map, the segment of the great circle remains the shortest route. Similarly, general covariance cannot make all trajectories in spacetime equivalent, and it is simply not the case that all states of motion are on the same footing.

In a sense, therefore, the formulation of generally covariant field equations in November 1915 was an empty victory. Despite its misleading name, however, general relativity was a powerful new theory of gravity, based on the equivalence principle, which in the mature form Einstein gave it in 1918 states that both the spacetime geometry and gravity should be represented by the metric tensor field (see reference 7 and CPAE 7; 4).

Coordinate restrictions

Why did Einstein reject the field equations of the first November paper when he and Grossmann first considered them in 1912–13? For a long time, the answer given by historians was that the two did not know about coordinate conditions. To compare field equations of broad covariance with field equations of limited covariance, such as the Poisson equation of Newtonian theory, the former need to be considered in a similarly restricted class of coordinate systems. That is done with the help of a coordinate condition consisting of four equations for the metric tensor field (hereafter, metric for short). Imposing a suitable coordinate condition, one eliminates various

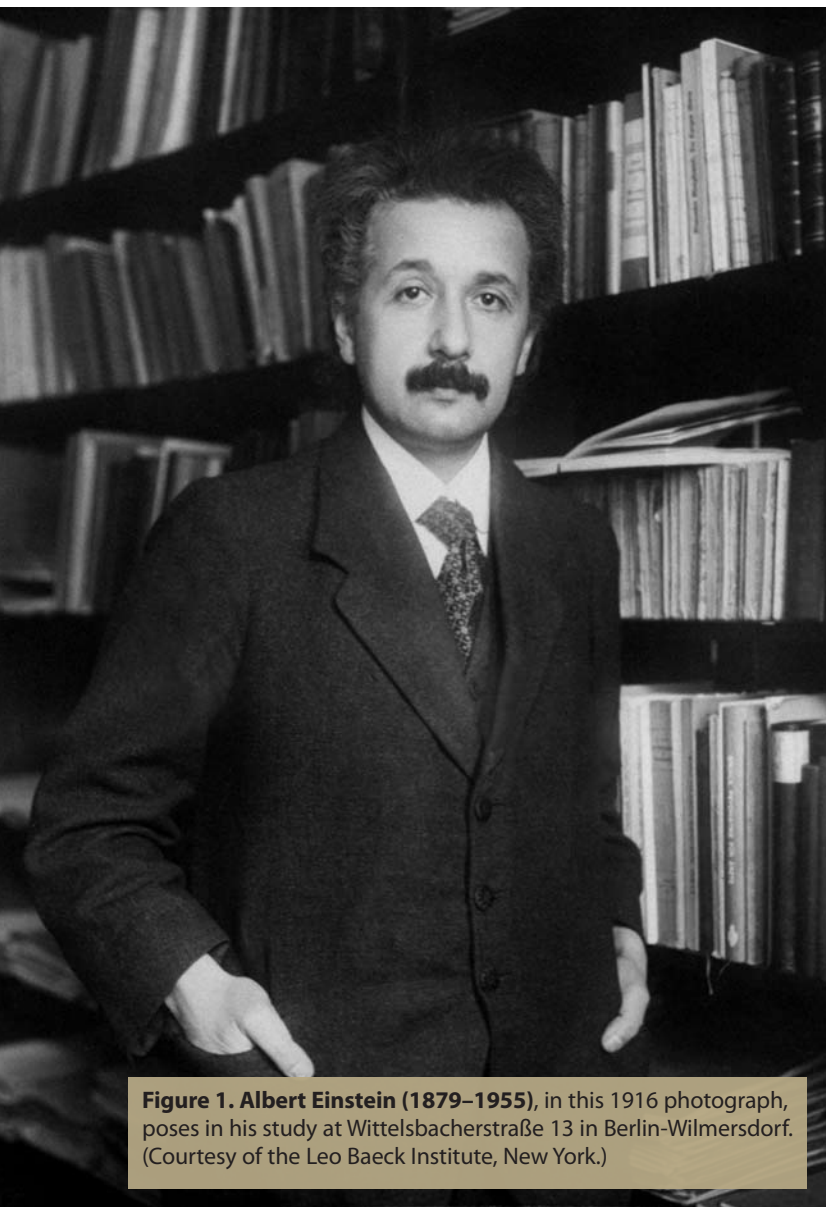


Figure 1. Albert Einstein (1879–1955), in this 1916 photograph, poses in his study at Wittelsbacherstraße 13 in Berlin-Wilmersdorf. (Courtesy of the Leo Baeck Institute, New York.)

terms from the broadly covariant field equations to leave only one term with second-order derivatives of the metric. One then shows that this term reduces to its Newtonian counterpart for weak static fields. In his well-known scientific biography of Einstein, Abraham Pais, for example, wrote that “Einstein still had to understand . . . that the choice of coordinates is a matter of convention without physical content.”⁸

Pais did not examine the Zürich notebook. In the entry reproduced in the box, Einstein imposed the condition that the four-dimensional divergence of the metric vanish. He used that same condition in his first November paper to eliminate unwanted terms with second-order derivatives of the metric from his field equations and to prove that the equations have the correct Newtonian limit. The Zürich notebook thus seems to provide incontrovertible evidence that Pais was wrong.

But then why *did* Einstein reject the field equations of the first November paper in the Zürich notebook? In his groundbreaking study of the notebook,

John Norton suggested that it may have been because the four-divergence of the metric of Minkowski spacetime in rotating coordinates does not vanish.⁹ For ease of reference, we rephrase Norton’s suggestion using some shorthand introduced in *The Genesis of General Relativity*¹ (hereafter *Genesis*): Einstein rejected the November tensor (the R_{im} introduced in figure 2) because the rotation metric (the metric of Minkowski spacetime in rotating coordinates) does not satisfy the Hertz condition (the vanishing of the four-divergence of the metric). But why, Norton wondered, did Einstein see that as a problem? The analysis of the notebook in *Genesis* answers his question. It turns out that Pais was not entirely wrong.

Before November 1915 Einstein used coordinate conditions in a way that is fundamentally different from the way they are used today. Nowadays, it is well known that coordinate conditions are gauge conditions. They select one or more representatives from each equivalence class of metrics. Which ones are chosen is a matter of convenience: Different coordinate conditions are appropriate for different problems. By contrast, Einstein used coordinate conditions—not just in the Zürich notebook but throughout the reign of the Entwurf theory—in a one-size-fits-all fashion: The same coordinate condition suits all problems.

In *Genesis*, we introduced the term coordinate restriction to characterize that idiosyncratic use. Unlike coordinate conditions, coordinate restrictions are an integral part of the theory in which they are used. The fundamental field equations are no longer the equations of broad or general covariance; instead, they are the truncated equations obtained once various terms have been eliminated with the help of a coordinate restriction. An additional difference between coordinate conditions and coordinate restrictions is that Einstein expected the restrictions to do double duty. He imposed them to ensure both that the field equations have the right Newtonian limit and that they are compatible with energy-momentum conservation.

Einstein’s use of coordinate restrictions explains why it was a problem for him that the rotation metric does not satisfy the Hertz condition. He wanted the rotation metric to be a solution of the field equations in the absence of matter so that he could interpret the inertial forces in a rotating frame of reference as gravitational forces. The November tensor itself vanishes for the rotation metric. However, if the rotation metric does not satisfy the Hertz restriction, what is left of the November tensor after it has been truncated with the Hertz restriction does not vanish. By the time he wrote the first November paper, Einstein recognized that the truncation issue was not a problem. Right after using the Hertz condition to demonstrate that his field equations based on the November tensor have the correct Newtonian limit, he noted that they allow transformations to rotating coordinates; evidently, by November 1915 Einstein was using the Hertz condition as a coordinate condition in the modern sense.

In the Zürich notebook, Einstein studied the covariance of candidate field equations obtained by truncating the November tensor by investigating

the covariance of the coordinate restrictions used to do the truncating. In the case of the Hertz restriction, there seems to be nothing to investigate. The condition that the four-divergence of the metric must vanish is covariant only under linear transformations. The covariance, however, is broader if one allows what Einstein called “nonautonomous transformations” in a letter to Hendrik A. Lorentz (CPAE 5; 467) and then, in publications from 1914 (CPAE 6; 2, 9), “justified transformations between adapted coordinates” (adapted, that is, to the metric). For ordinary, or autonomous, transformations, the new coordinates are simply functions of the old ones; for nonautonomous transformations, they are functions of the old coordinates and of the metric in the old coordinates.

The notions of coordinate restrictions and nonautonomous transformations fell by the wayside in Einstein’s final theory.¹⁰ However, in both the Zürich notebook and his papers on the Entwurf theory, one of Einstein’s central concerns was to make sure that his coordinate restrictions were covariant under a broad enough class of nonautonomous transformations to implement a relativity principle for arbitrary motion, a goal that itself turned out to be illusory.⁷

The Entwurf theory

The pages of the Zürich notebook show Einstein eventually giving up on constructing candidate field equations by truncating the November tensor with various coordinate restrictions. Instead, relying on an elaborate analogy between gravitational and electromagnetic fields, he produced the Entwurf field equations. He made his peace with their limited covariance when in August 1913 he hit upon the so-called hole argument, which purported to show that generally covariant equations cannot uniquely determine the metric in a region of spacetime without matter—a hole.

For a long time, historians thought that Einstein simply mistook two different coordinate representations of the same metric for two physically different metrics and that what the hole argument showed was not indeterminism but the same lack of understanding of coordinate conditions that had made him forego general covariance in the first place.⁸ John Stachel, founding editor of CPAE, was the first to show that such an uncharitable reading of the argument is incompatible with Einstein’s most careful statement of it (CPAE 6; 9, page 1067). Its resolution, accordingly, involved more than the correction of an elementary error. Einstein said nothing about the hole argument in November 1915. Only when pressed on the issue afterwards did he come up with an escape from it.¹¹

In early 1914 Einstein and Grossmann produced a second paper together (CPAE 6; 2). The idea was to get a better handle on the covariance properties of the Entwurf equations by deriving them from a variational principle. After all, it is easier to study the covariance of the Lagrangian, which in the variational approach determines the field equations, than it is to ascertain the covariance of the many-component field equations themselves. The Lagrangian for the Entwurf field equations has the same structure as the Lagrangian for the source-free

$$G_{im} = \{il, lm\} = R_{im} + S_{im}$$

$$R_{im} = -\frac{\partial \{i^m_l\}}{\partial x_l} + \sum_{\rho} \left\{ \begin{matrix} il \\ \rho \end{matrix} \right\} \left\{ \begin{matrix} \rho m \\ l \end{matrix} \right\}$$

$$S_{im} = \frac{\partial \{i^l_l\}}{\partial x_m} - \left\{ \begin{matrix} im \\ \rho \end{matrix} \right\} \left\{ \begin{matrix} \rho l \\ l \end{matrix} \right\}.$$

Figure 2. In his paper of 4 November 1915, Einstein split up the Ricci tensor G_{im} , which encodes spacetime curvature. (The indices i, l, m , and ρ take on the values 1 through 4.) As his new field equations, he proposed $R_{im} = -\kappa T_{im}$, where T_{im} is the energy–momentum tensor for matter and κ is proportional to Newton’s gravitational constant. These are equations of broad but not yet general covariance.

Maxwell equations: It is essentially the square of the gravitational field, defined as minus the gradient of the metric. Since the metric plays the role of the gravitational potential in the theory, it was only natural to define the gravitational field as minus its gradient. Einstein and Grossmann found four conditions that, if imposed on the metric and its derivatives, not only ensure energy–momentum conservation but also determine the class of nonautonomous transformations under which the Entwurf Lagrangian retains its form. Einstein convinced himself that the class was broad enough to make all motion relative. He was wrong. The rotation metric, for instance, is not a matter-free solution of the Entwurf field equations.

By the time that second paper appeared, Einstein had moved from Zürich to Berlin. He reprised its variational approach in a lengthy, self-contained exposition on the Entwurf theory, which appeared

$$G_{im} = -\kappa \left(T_{im} - \frac{1}{2} g_{im} T \right),$$

$$\sum_{\rho\sigma} g^{\rho\sigma} T_{\rho\sigma} = \sum_{\sigma} T^{\sigma}_{\sigma} = T$$

Figure 3. The Einstein field equations first appear in Einstein’s 25 November 1915 paper. Here, G_{im} is the Ricci tensor; g_{im} , the metric tensor; and T_{im} , the energy–momentum tensor for matter. Three weeks earlier Einstein had proposed the field equations $R_{im} = -\kappa T_{im}$ (see figure 2), which retain their form under unimodular transformations. The following week, he had shown that, as long as the trace T of the energy–momentum tensor vanishes, those equations could be seen as generally covariant equations $G_{im} = -\kappa T_{im}$ expressed in unimodular coordinates, for which the determinant of the metric tensor is -1 and $G_{im} = R_{im}$. To guarantee the vanishing of T , Einstein had assumed that all matter could be reduced to electromagnetic and gravitational fields. The addition of the trace term in the 25 November communication obviated the need for that questionable assumption.

in November 1914 (*CPAE* 6; 9). As a newly minted member of the Berlin Academy, he dutifully submitted his review article for publication in its proceedings. The title, “The formal foundation of the general theory of relativity,” reflects Einstein’s increased confidence in the theory. In the article, Einstein initially left open how the Lagrangian depends on the metric and its first-order derivatives. Once again, he found that the same conditions that determine the covariance of the Lagrangian also ensure energy–momentum conservation. His result is a special case of a theorem, published in 1918 by Emmy Noether, that connects symmetries and conservation laws. But the way Einstein looked on his result in 1914 was that energy–momentum conservation requires that the covariance of the field equations be restricted. Einstein proceeded to argue that the restriction uniquely picks out the Lagrangian that generates the Entwurf field equations.

Einstein and Hilbert

In 1915 Einstein was invited to give the Wolfskehl Lectures in Göttingen. From 28 June to 5 July, he lectured on the Entwurf theory and captured the imagination of his host, the great mathematician David Hilbert. The only surviving notes of those lectures (*CPAE* 6; appendix B) do not mention the Entwurf field equations, but it is safe to assume that Einstein covered them along the lines of his 1914 review article. A few months later, as his letters reveal, his confidence in those equations crumbled. In September he checked once again whether the rotation metric is a matter-free solution and found to his dismay that it is not (*CPAE* 8; 123). In October he discovered that his uniqueness argument for the Entwurf Lagrangian was fallacious (*CPAE* 8; 129).

Since he knew Hilbert was hot on his trail (see figure 5), Einstein may have rushed new field equations into print in early November. That haste helps explain why he had to correct them twice in the span of only three weeks. Hilbert presented his new field equations to the Göttingen Academy of Sciences on 20 November, five days before the Einstein field equations were presented in Berlin. It thus looks as if Hilbert, following a purely mathematical approach, beat Einstein to the punch. Einstein was annoyed that Hilbert was poaching on his preserves, but he quickly recognized the pettiness of that sentiment. Still, the episode goes a long way to explain both Einstein’s switch to a purely mathematical approach in his search for a unified field theory¹² and the selective amnesia that enabled him to believe that a mathematical approach had led him to the field equations of general relativity.¹³

Page proofs of Hilbert’s paper, which was not published until March 1916, surfaced in the late 1990s. An analysis of them¹⁴ shows that Einstein need not have worried about Hilbert in November 1915. In important ways, the theory presented in the page proofs is closer to the Entwurf theory than to general relativity in its final form. It has the same structure as the version of the Entwurf theory published in the review article of November 1914. It is true that Hilbert chose as his Lagrangian the Riemann curvature scalar, which is the gravitational part of the Lagrangian for the generally covariant



Figure 4. Marcel Grossmann (1878–1936) was a friend, classmate, and early collaborator of Einstein. (Portrait, from 1909, courtesy of ETH Zürich Library.)

Einstein field equations. At that point, however, Hilbert accepted Einstein’s hole argument against generally covariant field equations and imposed a coordinate restriction via a relation he suggestively called the energy theorem. So the actual field equations in the page proofs are not generally covariant. The restriction was dropped in the published version of the paper.

Replacing the Entwurf equations

The collapse of the uniqueness argument for the Entwurf Lagrangian in October 1915 was undoubtedly a setback for Einstein, but it also opened up new possibilities. It allowed him to consider different Lagrangians while keeping the general formalism of the 1914 review article intact. One option was to retain the Entwurf Lagrangian, modeled on the Lagrangian for the electromagnetic field, but to change the definition of the gravitational field. It had seemed only natural to define the gravitational field as minus the gradient of the gravitational potential, represented by the metric, just as the electrostatic field is minus the gradient of the electrostatic potential. However, the analogy between gravity and electromagnetism, on which Einstein relied so heavily in his work on the Entwurf theory, also illustrates that in general, the relation between field and potential is more complicated. In the first November paper, Einstein acknowledged that energy–momentum conservation had previously led him to view a single term with a gradient of the metric

as the natural expression of the components of the gravitational field, even though the formulae of the absolute differential calculus suggest the Christoffel symbols [a sum of three such terms]. This was a fateful prejudice. (*CPAE* 6; 21)

A few weeks later, Einstein wrote to Arnold Sommerfeld (*CPAE* 8; 153) that the redefinition of the gravitational field in terms of Christoffel symbols had been “the key to [the] solution.” If the new expression for the gravitational field is inserted into the Entwurf Lagrangian, the resulting field equations include the November tensor. If the formalism is restricted to unimodular transformations (a modest restriction requiring that the transformation determinant be equal to 1), the field equations are exactly the ones based on the November tensor found both in the Zürich notebook and in the paper of 4 November 1915.

In our account, then, Einstein found his way back to the equations of the first November paper through considerations of physics.¹³ The other possibility is suggested by Einstein’s claim in the introduction of the first November paper that the demise of the Entwurf equations had led him back to “the requirement of a broader covariance”: He returned to the November tensor because of its pedigree in the Riemann tensor.¹² In that account, which takes Einstein’s later recollections at face value and gives pride of place to considerations of mathematics, he only found the Lagrangian for the new field equations afterwards. There is no conclusive evidence to determine which came first, the redefinition of the gravitational field or the return to the Riemann tensor.

The important point, however, is that Einstein needed both to make sure his new field equations met all requirements. That the November tensor was connected to the Riemann tensor ensured that the field equations based on it had a broad enough covariance; that the field equations followed from a physically plausible Lagrangian ensured that they were compatible with energy–momentum conservation. With two routes to the same equations, Einstein had the luxury of a choice when he wrote his first November paper. Understandably, he chose the simpler mathematical route and followed the messier physical route only to handle energy–momentum conservation.

Covariance regained

Despite the remarkable convergence of physical and mathematical considerations that led to the first November paper, more work remained to be done to make them mesh. The general variational formalism of the 1914 review article, restricted to unimodular transformations, gave four conditions on the metric. Einstein expected those conditions both to guarantee energy–momentum conservation and to determine the class of nonautonomous transformations under which the new field equations were covariant. He already knew, however, that the new field equations, based on the November tensor, were covariant under the much broader class of autonomous unimodular transformations. In the first

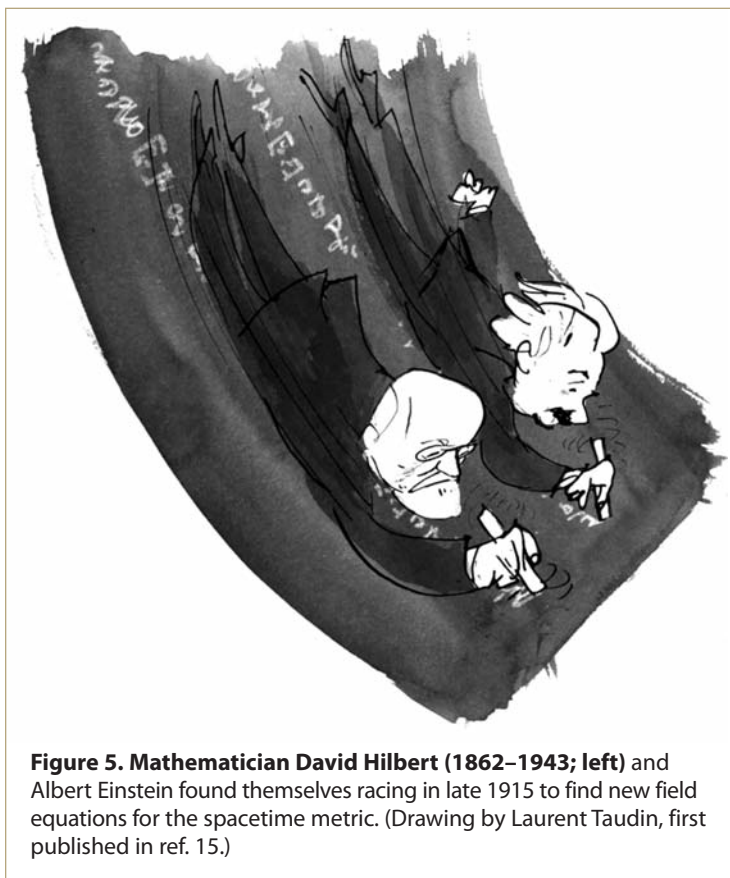


Figure 5. Mathematician David Hilbert (1862–1943; left) and Albert Einstein found themselves racing in late 1915 to find new field equations for the spacetime metric. (Drawing by Laurent Taudin, first published in ref. 15.)

November paper, he could not quite resolve that tension, but he did manage to alleviate it. He replaced the four conditions needed to secure energy–momentum conservation by just one on the determinant g of the metric.

The determinant condition, however, is an odd one. It implies that g can be a constant only if the trace T of the energy–momentum tensor vanishes. In particular, a nonvanishing T would rule out unimodular coordinates, in which $g = -1$. Unimodular coordinates are the natural ones to use in a theory covariant under unimodular transformations. Moreover, the Ricci tensor introduced in figure 2 reduces to the November tensor in unimodular coordinates, so the field equations of the first November paper could be seen as generally covariant equations expressed in special coordinates. Furthermore, one could then show that covariance (under unimodular transformations) guarantees energy–momentum conservation (in unimodular coordinates). Einstein thus had good reasons to seek a way around the prohibition against unimodular coordinates imposed by a nonvanishing T .

The field-equation amendments in the second and fourth November papers amount to two different ways of doing just that. At first, Einstein simply set $T = 0$. He justified that assumption by adopting the electromagnetic worldview, according to which all matter consists of electromagnetic fields governed by some nonlinear generalization of Maxwell’s equations. In that case, he argued, T probably vanishes, as it does for ordinary electromagnetic fields. It was a high price to pay for general covariance, and

Einstein quickly realized he overpaid. There was no need to set $T = 0$. If a term with T is added to the right-hand side of the field equations (figure 3), the determinant condition on g no longer blocks unimodular coordinates coexisting with a nonvanishing trace. Energy–momentum conservation provided a powerful independent argument for the addition of the trace term. It ensured that the energy–momentum of matter enters the field equations in the same way as the energy–momentum of the gravitational field. Einstein could be confident he got it right this time. He had found the field equations that will forever bear his name.

Arch and scaffold

The extant papers, notebooks, and letters from 1912–15 made it possible to reconstruct in considerable detail how Einstein used the Entwurf theory as a scaffold to build the arch of the Einstein field equations. The Entwurf theory contains expressions for and relations between the field equations, the four-force, the energy–momentum density, and the Lagrangian for the gravitational field, all of which mimic corresponding expressions and relations for the electromagnetic field in the four-dimensional formalism for electrodynamics in special relativity. It also features a tight connection between energy–momentum conservation and covariance. Those expressions and relations, encoded in the variational formalism of the 1914 review article, are what held the scaffold together.

The whole constellation survives intact when a basic building block is swapped out for a new one—that is, when the gravitational field is redefined from minus the gradient of the metric to minus the Christoffel symbols. The new building blocks, the stones of the arch, are kept in place the same way as the building blocks of the scaffold. Parts of the scaffold could be discarded, notably the ungainly concept of nonautonomous transformations. Parts were moved and reconfigured. Coordinate restrictions were turned into coordinate conditions. The relation between covariance and energy–momentum conservation was inverted from conservation restricting covariance to covariance guaranteeing conservation.

Despite Einstein's efforts to hide the scaffold, the arch unveiled in the first November paper still shows clear traces of it. Einstein wanted to display its mathematical splendor, but his structure was not quite up to code yet. To prevent a violation of energy–momentum conservation from bringing it down, he had to prop it up with some lumber from the Entwurf scaffold, a support beam cut to the size of unimodular transformations. Unfortunately, that support beam did not leave enough room for another critical piece of scaffolding, the restriction to unimodular coordinates. Only with that restriction in place could one fully appreciate the magnificence of the arch as a structure that looks the same from all angles even though during construction it could only be viewed from some. To make room for unimodular coordinates, Einstein brought in the heavy machinery of the electromagnetic worldview, then realized he did not need it. What he needed instead was an extra term on the right-hand side of the field

equations with the trace of the energy–momentum tensor. That trace term turned out to be the keystone of his arch.

That is how Einstein left the construction site in November 1915. During the next few months, he cleared away the debris of the first two November 1915 papers, which resulted in the first self-contained exposition of the new theory, published in May 1916 (CPAE 6; 30).¹⁵ In November 1916 he finally lifted the restriction to unimodular coordinates (CPAE 6; 41). The arch was self-supporting at last, a marvelous sight to behold for generations of physicists to come.

References

1. M. Janssen, J. Renn, *How Einstein Found His Field Equations*, Springer (forthcoming), based on J. Renn, ed., *The Genesis of General Relativity*, vols. 1–4, Springer (2007). An analysis of Einstein's path to his field equations written for a general audience will appear as J. Renn, *On the Shoulders of Giants and Dwarfs: Einstein's Unfinished Revolution*, Princeton U. Press (forthcoming).
2. All Einstein documents mentioned in this article can be found in J. Stachel et al., eds., *The Collected Papers of Albert Einstein*, Princeton U. Press (1987–). Hereafter we use the notation CPAE X; Y, with X the volume number and Y the document number. The Einstein Papers Project at Caltech has made all volumes freely available at <http://www.einstein.caltech.edu>. The Albert Einstein Archives at the Hebrew University of Jerusalem has likewise made most of its holdings freely available at <http://www.alberteinstein.info>.
3. A. Einstein to L. de Broglie, 8 February 1954, quoted and discussed in J. van Dongen, *Einstein's Unification*, Cambridge U. Press (2010), p. 2.
4. M. Janssen, in *Beyond the Meme: Articulating Dynamic Structures in Cultural Evolution*, A. C. Love, W. C. Wimsatt, eds., U. Minnesota Press (forthcoming).
5. T. Sauer, in *Proceedings of the Thirteenth Marcel Grossmann Meeting on General Relativity*, R. T. Jantzen, K. Rosquist, R. Ruffini, eds., World Scientific (2015), p. 456.
6. See J. Renn, ed., ref. 1, vols. 1–2, and CPAE 4; 10. High-resolution scans are available at the digital Einstein Archives (ref. 2; search for “Zürich notebook” in the archival database). Page numbers of those scans do not correspond to page numbers used here, which follow the notation (1L, 1R–43L, with L and R for left and right) of J. Renn, ed., ref. 1, vols. 1–2.
7. M. Janssen, in *The Cambridge Companion to Einstein*, M. Janssen, C. Lehner, eds., Cambridge U. Press (2014), p. 167.
8. A. Pais, “Subtle Is the Lord . . .”: *The Science and Life of Albert Einstein*, Oxford U. Press (1982), p. 222.
9. J. D. Norton, *Hist. Stud. Phys. Sci.* **14**, 253 (1984); see also D. Howard, J. Stachel, eds., *Einstein and the History of General Relativity*, Birkhäuser (1989), p. 119.
10. For later use of nonautonomous transformations, see P. Bergmann, A. Komar, *Int. J. Theor. Phys.* **5**, 15 (1972).
11. J. Renn, ed., ref. 1, vol. 2, p. 819; M. Janssen, ref. 7, p. 188; J. D. Norton, “The Hole Argument,” <http://plato.stanford.edu/entries/spacetime-holearg/>; J. Stachel, “The Hole Argument and Some Physical and Philosophical Implications,” <http://relativity.livingreviews.org/Articles/lrr-2014-1>.
12. J. D. Norton, *Stud. Hist. Phil. Mod. Phys.* **31**, 135 (2000).
13. J. van Dongen, ref. 3; J. Renn, ed., ref. 1, vol. 2, p. 839.
14. L. Corry, J. Renn, J. Stachel, *Science* **278**, 1270 (1997); J. Renn, ed., ref. 1, vol. 4, pp. 857–1015.
15. H. Gutfreund, J. Renn, *The Road to Relativity: The History and Meaning of Einstein's “The Foundation of General Relativity,”* Princeton U. Press (2015). ■