Introductory note to Zermelo's 1894 doctoral thesis, "Untersuchungen zur Variations-Rechnung", University of Berlin. In Ernst Zermelo - *Collected Works/Gesammelte Werke: Volume II/Band II - Calculus of Variations, Applied Mathematics, and Physics/Variationsrechnung*. Springer: Heidelberger Akademie der Wissenschaften, Heinz-Dieter Ebbinghaus et al. (Eds.), 2013, pp. 10-25.

Introductory note to 1894

Craig G. Fraser

In 1894 Zermelo published his doctoral dissertation from the University of Berlin, written under the direction of Hermann Amandus Schwarz and devoted to a study of Karl Weierstrass's methods in the calculus of variations. In an introductory note Zermelo stated that he had become familiar with the contents of Weierstrass's lectures in 1892 from the copy in the "Mathematischer Verein" in Berlin, as well as from a lecture given by Schwarz. Weierstrass had investigated the simplest case in which only the first derivatives of the variables appear in the variational integrand function. Zermelo's main goal was to extend Weierstrass's results on necessary and sufficient conditions involving the "excess" or E function (the material in the twentieth to the twenty-third lectures of Weierstrass's lectures as they were eventually published (*1927*)) to variational problems in parametric form in which the integrand contains derivatives of order higher than one.

1. Sufficient conditions before Weierstrass

A major goal of the calculus of variations in the nineteenth century was to identify conditions that ensure that a proposed solution to a given variational problem is a maximum or a minimum. Any such solution will have to satisfy the Euler differential equation and will also have to satisfy Legendre's condition. It was noticed that a function that satisfied these conditions turned out in certain instances not to be a genuine extremum. It was required to assemble a set of conditions that taken together are sufficient to ensure a maximum or a minimum. In researches of the late 1830s Carl Gustav Jacobi (*1837, 1838*) introduced some new ideas in this direction that became the basis for a very active program of research. Jacobi formulated a certain condition, known in the later subject as Jacobi's criterion, that must be satisfied by any solution to the problem. Jacobi's theory was also based on a new transformation of the second variation. The variational integrand was expressed in a form that enabled one to infer Legendre's condition for very general integrals. (The relevant history is presented in *Todhunter 1861, Goldstine 1980* and *Fraser 2003*.)

The primary object of interest here is an integral involving a single independent variable of the form $\int_a^b f(x, y, y', ..., y^{(n)})\, dx$, where the integrand function f is a function of x, y and the derivatives of y with respect to x up to order n. It is necessary to find the particular function $y = y(x)$ that maximizes or minimizes this integral. In the elementary case where $n = 1$, researches succeeded in providing a completely satisfactory theory. In *1857*

H.-D. Ebbinghaus, A. Kanamori (Eds.), *Ernst Zermelo - Collected Works/ Gesammelte Werke II*, DOI 10.1007/978-3-540-70856-8_2, Schriften der Mathematisch-naturwissenschaftlichen Klasse der Heidelberger Akademie der Wissenschaften 23,

Ludwig Otto Hesse showed in this case that if the Euler, Legendre and Jacobi conditions are satisfied then the resulting curve is indeed a maximum or a minimum. The more interesting and much more difficult case occurred when $n \geq 2$. Here it was found that constants appearing in functions required in the transformation of the second variation must satisfy certain conditions. It was necessary to show that it was possible to find a set of constants that worked in the general case. Essentially the problem was one of existence, of finding suitable mathematical objects that allowed the transformation to take place. (A succinct statement of the point in question is given in *Lindelöf and Moigno 1861*.)

The central question was resolved by the Leipzig mathematician Adolph Mayer in his *Habilitation* thesis *1866*, a work whose core content was presented by Mayer two years later in an article in Crelle's journal (*1868*). Mayer showed that if Jacobi's criterion held, then it was possible to carry out the desired transformation of the second variation. Assuming the validity of Legendre's criterion, one may infer that the given function satisfying the Euler equation is indeed a maximum or a minimum. Mayer presented his result in a very general setting, using a formulation of the variational problem that had been developed by Alfred Clebsch. Mayer's investigation showed both technical sophistication and a deep understanding of the theoretical issues at the foundation of the theory.

2. Weierstrass

Weierstrass's contributions to the calculus of variations were a product of his middle and late years. Although he began lecturing on the subject at the University of Berlin as early as 1865, his most significant results were presented in the summer lectures of 1879, when he was sixty-three years old. The edition which was eventually published in 1927 is based on these as well as a second set of lectures given in 1883. Although this delay in publication somewhat limited the dissemination of his ideas, he exerted considerable influence on contemporary variational research. Copies of his notes circulated privately and his results began to be disseminated in published form by other researchers beginning in the middle 1890s. The appearance of Zermelo's dissertation in 1895 was among the first publications of Weierstrass's ideas, developed in a more general setting than the one adopted by Weierstrass.

More than any other researcher Weierstrass established the critical outlook of the calculus of variations as a modern mathematical subject. In his lectures the distinction between necessary and sufficient conditions appears clearly for the first time. He carefully specified the continuity properties that must be satisfied by functions and their variations. In problems of constrained optimization he used theorems on implicit functions to ensure that the optimizing arc was embedded in a suitable family of comparison curves.

Traditionally researchers in the calculus of variations did not identify at the outset of their investigation the precise class of comparison arcs in a given variational problem. There was no prior logical conception concerning the nature of this class. However, the δ-process introduced by Lagrange required that both the comparison arc and its slope at each point differ by only a small amount from the value and slope of the solution curve. This condition was imposed by the nature of the variational process, which involved expanding the integrand function as a Taylor series and investigating the behavior of the second variation arising in this expansion. Isaac Todhunter (*1871*, 269) in an essay on what were known as "discontinuous" solutions seems to have been the first to explicitly call attention to this limitation on the class of comparison arcs:

> [...] if we assert that the relation [Euler equation] does give a minimum, we must bear in mind that this means a minimum with respect to admissible variations [...] our investigation is not applicable to such a variation as would be required in passing from the cycloid to the discontinuous figure: in such a passage p [$= \delta y'$] would not always be indefinitely small. Of course it might be possible to give some special investigation for such a case, but certainly the case is not included in the ordinary methods of the Calculus of Variations.

In his Berlin lectures Weierstrass developed an alternative to the traditional expansion methods that extended the variational theory to a larger class of comparison curves. The precise nature of these curves was still determined by the particular technical requirements of the new method, but the logical orientation of the subject had shifted. In earlier variational research the nature of the mathematical objects was determined implicitly by the variational process that was employed. By contrast, in Weierstrass there was a self-conscious and explicit focus on the objects being studied. His work involved a more intimate connection between the foundations of real analysis and the collection of concrete techniques and results that made up the variational theory.

It is necessary to call attention to one aspect of the style in which Weierstrass developed the theory. Traditionally researchers in the calculus of variations had adopted what is referred to as the ordinary or functional approach, in which the curve is expressed as $y = y(x)$ and the variational integrand (in the simplest case) takes the form $f(x, y, y')$. A distinctive aspect of Weierstrass's approach was his adoption of a parametric approach. The curve C is represented parametrically in the form $y = y(t)$ and $x = x(t)$. Here the variational integrand takes the form $I = \int_{t_0}^{t_1} f(x, y, x', y')\, dt$ where $x' = \frac{dx}{dt}$ and $y' = \frac{dy}{dt}$. In the parametric (or homogeneous) formulation of the variational problem it is necessary to impose conditions on the variational integrand in order to ensure that the problem is independent of the particular parameterization chosen. One must attend to these conditions in developing the theory.

Weierstrass used a parametric approach throughout his lectures on the calculus of variations. Researchers who adopted his new method tended to also use a parametric approach, and this was true in the case of Zermelo. However, not all researchers followed this practice. Although the parametric approach has certain advantages, particularly from a geometric viewpoint, its analytical development is less natural than the ordinary theory. During the years around 1900 when Weierstrass's ideas were becoming more widely known, researchers such as Oskar Bolza (*1909*), William Osgood (1901) and Emile Goursat (*1905*) went to some effort to reformulate his results in terms of the ordinary theory. In the large majority of the textbook literature of the past one hundred years the ordinary approach is taken as the standard formulation of the variational problem while the parametric theory is presented as a special subject.

For the sake of exposition we will adopt the ordinary theory in explaining some of the basic ideas that underlie Weierstrass's theory. At the conclusion we indicate by way of comparison the parametric form in which Weierstrass originally presented his result. (In our account we have adopted the sensible notation used by Adolf Kneser to denote extremal and comparison curves: $y = y(x)$ is the comparison curve, while $\overline{y} = \overline{y}(x)$ is the extremal curve; both Weierstrass and Zermelo use the opposite convention.)

Suppose $y = y_0(x)$ is an arc C_0 on which the variational integral $I = \int_a^b f(x, y, y')\, dx$ is a minimum. (The case of a maximum is similar, with the inequalities reversed.) In the traditional formulation of the theory involving expansion methods and the second variation, the conditions that must be satisfied are all specified in terms of the function $y = y_0(x)$ and this function alone. Thus $y = y_0(x)$ will be a solution to the Euler equation and we must have $\dfrac{\partial^2 f(x, y_0(x), y_0'(x))}{\partial y'^2} \geq 0$ (Legendre's condition) and the Jacobi criterion must hold. One typically supposes that the family of comparison curves is of the form $C : y = y_0(x) + \epsilon\zeta(x)$. Here $\zeta(x)$ is any function subject to the usual continuity restrictions with $\zeta(a) = \zeta(b) = 0$. More generally we may have a family of comparison curves of the form $C : y = y(x, \epsilon)$ where $y(x, 0) = y_0(x)$ and $y(a, \epsilon) = y(b, \epsilon) = 0$. It is evident that the class of comparison curves is very extensive. Nevertheless, as Todhunter observed in 1870, it is also clear that there are some restrictions on this class. For example, if $y = y_0(x) + \epsilon\zeta(x)$ (ϵ small) then $y'(x) - y_0'(x) = \epsilon\zeta'(x)$ with similar relations for higher derivatives of x. It follows that the neighboring curve $y = y(x)$ differs by only a small amount from the optimizing curve $y = y_0(x)$ not just for corresponding values of y but for derivatives of y of all orders.

It turns out that it is possible for a variational integral to be a minimum for the function $y = y_0(x)$, considered with respect to a class of comparison curves of the type $y = y(x, \epsilon)$, but not be a minimum if we allow comparison curves whose slope differs by a finite amount from $y = y_0(x)$. The following simple example illustrates this situation. It is taken from Bolza's *Lectures on the calculus of variations* (*1904*, 74), one of the earliest expositions of the new theory developed by Weierstrass and extended by Zermelo. (The term "die neue Variationsrechnung" was sometimes used in reference to the Weierstrass theory.) We have $f(x, y, y') = y'^2 + y'^3$ defined on the interval $[0, 1]$.The variational integral is $I = \int_0^1 (y'^2 + y'^3)\, dx$. It follows that a solution to the Euler equation will be $y' = constant$. Since the solution must pass through the endpoints it follows that the hypothetical minimizing curve is simply $y = 0$. We have $\frac{\partial^2 f(x, y, y')}{\partial y'^2} = 2 \geq 0$, so the Legendre condition is satisfied. It is also the case that Jacobi's condition holds in this example. Hence the curve $y = 0$ minimizes the integral with respect to comparison arcs of the form $y = y(x, \epsilon)$. Consider now the comparison curve C consisting of two straight lines, the first joining the origin $(0, 0)$ to the point $(1 - p, q)$ and the second joining $(1 - p, q)$ to $(1, 0)$ (see Fig. 1). Here p is a number with $0 < p < 1$ and q is a small positive quantity. For this comparison curve we have

$$\Delta I = \frac{q^2}{p(1-p)}\left(1 + \frac{q}{1-p} - \frac{q}{p}\right).$$

We can make C lie within any neighborhood of $y = 0$ by making q sufficiently small. With q specified, it is clear that $\Delta I < 0$ for $p \ll q$. Hence I is not a minimum for the larger class of curves that includes the curve C.

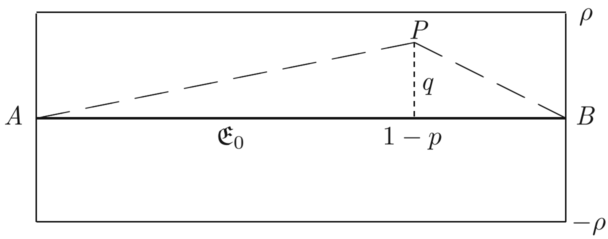

Fig. 1 (p. 74 of *Bolza 1904*)

In the terminology that was introduced by Kneser in *1900* and became standard, the traditional variational theory yields sufficient conditions for a *weak* extremum. Here each comparison curve is close to the minimizing curve at y and at all derivatives of y. By contrast, a solution will be a *strong* extremum if it is a minimum for the wider class of curves which are close to the solution curve but may have a slope that differs by a finite amount from the solution curve.

Consider again the problem of finding the curve $C_0 : y = y_0(x)$ that maximizes or minimizes $I = \int_a^b f(x, y, y')\, dx$. Suppose that the Euler and

Jacobi conditions hold for the arc C_0. We now enlarge the class of possible comparison curves to include ones whose slope differs by a finite amount from that of C. In order to establish that C_0 is a minimum with respect to this enlarged class of comparison curves it is necessary to formulate a condition that involves not just the function $y = y_0(x)$ but also the curves $C : y = y(x)$ in the comparison class. Perhaps the simplest approach would be simply to require that

$$f(x, y_0, y_0') \leq f(x, y, y') \ (a \leq x \leq b)$$

for all comparison curves $y = y(x)$. Imposing this condition is evidently not very informative, and indeed is only a restatement of the problem. Weierstrass succeeded in formulating a more meaningful condition involving a function E called the excess function. We are given the proposed minimizing arc $C_0 : y = y_0(x)$. We consider any comparison curve $C : y = y(x)$ and take any point on this curve with coordinates x and y. By assumption the point (x, y) on the comparison curve C is close to the point (x, y_0) on C_0. Let $y'(x)$ be the slope of the comparison curve at the given point. This quantity may vary by an arbitrary amount from $y_0'(x)$.

A solution to the Euler equation is called an extremal. We typically require that such a solution passes through the initial point. A key idea introduced by Weierstrass was to introduce a function $p(x, y)$—known as the slope function—defined as the slope of the extremal passing through the point (x, y) at this point. We now introduce the excess function $E(x, y, y', p)$ defined as

$$E(x, y, y', p) = f(x, y, y') - f(x, y, p) - (y' - p)\frac{\partial f}{\partial y'}(x, y, p)\,. \tag{1}$$

Consider a region or strip about the curve C_0 and suppose that for each point in this strip it is possible to determine an extremal joining the initial point and the given point. Such a set of solutions to the Euler equation is today called a field of extremals, a term introduced by Kneser. Consider the condition

$$E(x, y, y', p) \geq 0\,. \tag{2}$$

This condition is known as Weierstrass's condition. If this condition is satisfied for all comparison curves $y = y(x)$ on the interval then the value of I along C_0 is less than its value along any member of C of the comparison class. We may conclude in this case that $C_0 : y = y_0(x)$ is a strong minimum.

In later mathematics this result would be proved using something called the Hilbert invariant integral, introduced by David Hilbert in 1900 to simplify Weierstrass's method. Here we present Weierstrass's original idea, which was also adopted by Zermelo. The minimizing arc is given as $C_0 : (x, y_0(x))$ while

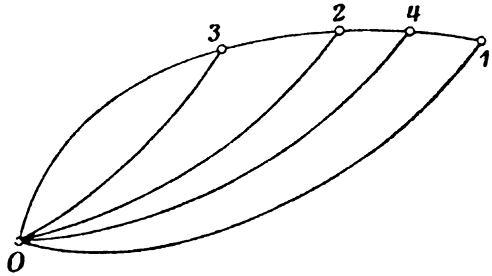

Fig. 2 (p. 219 of *Weierstrass 1927*)

the neighboring curve C is given as $C:(x,y(x))$. In Fig. 2 C_0 is the bottom arc **01**, while the neighboring curve C is the arc **0321**. We suppose that C lies in a narrow region about C_0. This region or "field" has the following property: for any point within it, there is a unique solution curve to the Euler equation—an extremal—passing through the initial point **0** and the given point. We designate extremal curves using the functional notation $\overline{y}=\overline{y}(x)$. Let (x,y) be any point on C; in Fig. 1 this point is labelled **2**. Consider the extremal curve $(x,\overline{y}(x))$ passing from **0** through **2**; in Fig. 1 it is the curve **02**. We introduce the function $\phi(x)$ defined as

$$\phi(x)=\int_{\mathbf{0}}^{\mathbf{2}} f(x,\overline{y},\overline{y}')\,dx+\int_{\mathbf{2}}^{\mathbf{1}} f(x,y,y')\,dx\,. \tag{3}$$

This integral is taken along the extremal arc from **0** to **2** and then along the comparison arc from **2** to **1**. Evidently we have $\phi(0)=\int_{\mathbf{0}}^{\mathbf{1}} f(x,y,y')\,dx$ and $\phi(1)=\int_{\mathbf{0}}^{\mathbf{1}} f(x,y_0,y_0')\,dx$. The statement that $\int_{\mathbf{0}}^{\mathbf{1}} f(x,y_0,y_0')\,dx$ is a minimum is equivalent to the inequality $\phi(0)\geq\phi(1)$, which would follow if we are able to show that $\phi(x)$ is a decreasing function of x. To do this we calculate the derivative of $\phi(x)$ and show that it is negative.

Let the integrals $\overline{I}_{02}$ and I_{21} be defined as

$$\overline{I}_{02}=\int_{\mathbf{0}}^{\mathbf{2}} f(x,\overline{y},\overline{y}')\,dx,\ \ I_{21}=\int_{\mathbf{2}}^{\mathbf{1}} f(x,y,y')\,dx\,. \tag{4}$$

From the standard formula for the variation of the integral when the endpoint is allowed to vary in both the x and y directions we have

$$\delta\overline{I}_{02}=\frac{\partial f}{\partial\overline{y}'}(x,\overline{y},\overline{y}')\delta y+\Big(f(x,\overline{y},\overline{y}')-\overline{y}'\frac{\partial f}{\partial\overline{y}'}(x,\overline{y},\overline{y}')\Big)\delta x\,. \tag{5}$$

We now let $\delta x=dx$ and $\delta y=dy$. Note that at the point **2** we have $\overline{y}=y$ and $\overline{y}'=p(x,y)$, where p is the slope function for the given field. Hence (5) becomes

$$\frac{d\overline{I}_{02}}{dx} = \frac{\partial f}{\partial y'}(x,y,p)y' + f(x,y,p) - p\frac{\partial f}{\partial y'}(x,y,p) \qquad (6)$$
$$= \frac{\partial f}{\partial y'}(x,y,p)(y'-p) + f(x,y,p)\,.$$

The derivative of I_{21} is given immediately as

$$\frac{dI_{21}}{dx} = -f(x,y,y')\,. \qquad (7)$$

Thus we have

$$\phi'(x) = \frac{d\overline{I}_{02}}{dx} + \frac{dI_{21}}{dx}$$
$$= -\left(f(x,y,y') - f(x,y,p) - (y'-p)\frac{\partial f}{\partial y'}(x,y,p)\right),$$

or

$$\phi'(x) = -E(x,y,y',p)\,. \qquad (8)$$

If $E(x,y,y',p) \geq 0$ on the interval then $\phi'(x)$ is negative and $\phi(x)$ is a decreasing function, which is what was required to be proved.

It is apparent that $\phi(0) - \phi(1) = \int_{\mathbf{0}}^{\mathbf{1}} E(x,y,y',p)\,dx$. But $\phi(0) - \phi(1)$ is equal to ΔI, the variation of the integral with respect to the comparison arc. Hence we have

$$\Delta I = \int_{\mathbf{0}}^{\mathbf{1}} E(x,y,y',p)\,dx\,. \qquad (9)$$

(9) is known in the modern literature as Weierstrass's theorem, although it does not appear explicitly in Weierstrass's lectures. From (9) it is clear why the function E is called the excess function, since the excess of the variational integral I in going from $y = y_0(x)$ to $y = y(x)$ is the integral of E over the given interval.

In Weierstrass's original parametric approach the variational integrand takes the form $\int_{t_0}^{t_1} F(x,y,x',y')\,dt$. The minimizing arc is given as C_0 : $(x_0(t), y_0(t))$ while the neighboring curve C is given as $C : (x(t), y(t))$. In Fig. 2 C_0 is the bottom arc **01**, while the neighboring curve C is the arc **0321**. We suppose that C lies in a narrow region about C_0. This region is supposed to be a "field" in the sense defined above: for any point within it there is a unique solution curve to the Euler equation passing through the initial point and the given point. Let $(x(t), y(t)$ be any point on C; in Fig. 2 this point is labelled **2**. Consider the extremal passing through this point; in Fig. 2 it is the curve **02**. Let $p(t), q(t)$ be the coordinate slope functions of the extremal at (x, y). The excess function E in parametric form is defined as

$$E(x,y,p,q,x',y') = F(x,y,x',y') - x'\frac{\partial F}{\partial p}(x,y,p,q) - y'\frac{\partial F}{\partial q}(x,y,p,q)\,. \quad (10)$$

If for each such comparison curve C we have

$$E(x,y,p,q,x',y') \geq 0 \quad (11)$$

for all values of x, y, x', y' then we may conclude that C_0 minimizes the integral $\int_{t_0}^{t_1} F(x,y,x',y')\,dt$.

In the ordinary calculus a condition that $y = y(x)$ be a minimum at $x = a$ is that $\frac{dy}{dx}(x = a) = 0$. No reference is made in this condition to neighboring values of a. Similarly, in the case of weak extrema in the calculus of variations the conditions are formulated solely in terms of the curve $C : y = y_0(x)$ and do not involve any reference to neighboring curves or functions. By contrast, Weierstrass's condition involves the comparison curve as well as the field function $p(x, y)$ defined in a neighborhood of C. It should be noted that while it is true that Weierstrass has obtained a stronger result, this is possible because the condition that must be satisfied is more restrictive; the stronger result is achieved at a higher price.

3. Zermelo's dissertation

It was inevitable that Zermelo's readership would be restricted because he was extending a mathematical theory that itself had not been published and that would have been familiar only to a fairly small group of researchers either at German universities or who had studied there. Furthermore, Weierstrass's parametric approach was not widely used in the calculus of variations; even investigators such as Ludwig Scheeffer (*1885*), Georg Erdmann (*1877*) and Edmund Husserl (*1882*) who were in a general sense part of the Weierstrass "school" and were influenced by his ideas employed the ordinary formulation of the variational problem in their researches of the 1870s and 1880s.

Zermelo's dissertation was also written in a rather formal manner, with very limited exposition of basic ideas and principles, and was excessively concerned with procedural matters, detailed formulations required in the general theory, and points of rigor. In matters of style, in its propensity for strenuous formal development, his approach bore similarities with the work of such earlier researchers as Hesse. Zermelo's work displayed as well a new element in variational research, a tendency to want to develop the subject from a larger viewpoint and to present the results as an instance of some more general and yet to be precisely specified subject. This tendency was manifested in his study of homogeneity properties of functions in the first chapter, as well as in his classification of families of curves in the second chapter.

Zermelo's dissertation would have been of primary interest to a reader who was either already familiar with Weierstrass's lectures or who was very motivated to learn about his new methods. For such a reader, the work would have been a valuable piece of mathematical research and exposition. In a fully detailed and methodical manner Zermelo developed a general theory, showing fully the non-trivial considerations that are involved in extending Weierstrass's methods to the problem $n > 1$. The first chapter was devoted to the homogeneity relations that must be satisfied in the parametric theory; the second to necessary conditions; the third to the excess function; and the final fourth chapter to sufficiency conditions involving the excess function.

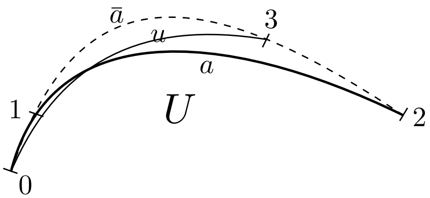

Fig. 3 (p. 90 of *Zermelo 1894*)

To indicate the basic idea behind Zermelo's development we describe how it plays out for the ordinary problem in the case where the variational integrand is a function of x, y, y' and y''. This was the setting in which Jacobi and so many other researchers set forth the theory. As before let $y = y_0(x)$ be the solution curve to the Euler equation joining **1** and **2**. In Fig. 3, from p. 90 of Zermelo's dissertation, this curve is denoted as a (note that the points are numbered slightly differently than in Weierstrass). We suppose there is a strip or region (what later was called a field) about a with the property that there is a unique extremal joining **0** (a point very close and to the left of **1** on a) and any given point **3** of the region. (In a technical refinement of Weierstrass's method, Zermelo takes the common starting point of the extremals to be **0** rather than **1** in order to simplify the analysis needed to establish the existence of the desired field. The value of the variational integral from **0** to **1** is taken to be negligible.) An arbitrary comparison curve 132: $y = y(x)$ is designated as $\overline{a}$. It is assumed that for each point **3** on $\overline{a}$ there is a unique extremal curve (a solution to the Euler equation) $\overline{y} = \overline{y}(x)$, joining **0** to **3**. This curve is designated as u in Fig. 3. Consider the function $\phi(x)$ given as

$$\phi(x) = \int_{\mathbf{0}}^{\mathbf{3}} f(x, \overline{y}, \overline{y}', \overline{y}'')\, dx + \int_{\mathbf{3}}^{\mathbf{2}} f(x, y, y', y'')\, dx\,. \tag{12}$$

Let the integrals in (12) be designated as $\overline{I}_{03}$ and I_{32}:

$$\overline{I}_{03} = \int_{\mathbf{0}}^{\mathbf{3}} f(x, \overline{y}, \overline{y}', \overline{y}'')\, dx,\ \ I_{32} = \int_{\mathbf{3}}^{\mathbf{2}} f(x, y, y', y'')\, dx\,. \tag{13}$$

Ludwig Otto Hesse showed in this case that if the Euler, Legendre and Jacobi conditions are satisfied then the resulting curve is indeed a maximum or a minimum. The more interesting and much more difficult case occurred when $n \geq 2$. Here it was found that constants appearing in functions required in the transformation of the second variation must satisfy certain conditions. It was necessary to show that it was possible to find a set of constants that worked in the general case. Essentially the problem was one of existence, of finding suitable mathematical objects that allowed the transformation to take place. (A succinct statement of the point in question is given in *Lindelöf and Moigno 1861*.)

The central question was resolved by the Leipzig mathematician Adolph Mayer in his *Habilitation* thesis *1866*, a work whose core content was presented by Mayer two years later in an article in Crelle's journal (*1868*). Mayer showed that if Jacobi's criterion held, then it was possible to carry out the desired transformation of the second variation. Assuming the validity of Legendre's criterion, one may infer that the given function satisfying the Euler equation is indeed a maximum or a minimum. Mayer presented his result in a very general setting, using a formulation of the variational problem that had been developed by Alfred Clebsch. Mayer's investigation showed both technical sophistication and a deep understanding of the theoretical issues at the foundation of the theory.

2. Weierstrass

Weierstrass's contributions to the calculus of variations were a product of his middle and late years. Although he began lecturing on the subject at the University of Berlin as early as 1865, his most significant results were presented in the summer lectures of 1879, when he was sixty-three years old. The edition which was eventually published in 1927 is based on these as well as a second set of lectures given in 1883. Although this delay in publication somewhat limited the dissemination of his ideas, he exerted considerable influence on contemporary variational research. Copies of his notes circulated privately and his results began to be disseminated in published form by other researchers beginning in the middle 1890s. The appearance of Zermelo's dissertation in 1895 was among the first publications of Weierstrass's ideas, developed in a more general setting than the one adopted by Weierstrass.

More than any other researcher Weierstrass established the critical outlook of the calculus of variations as a modern mathematical subject. In his lectures the distinction between necessary and sufficient conditions appears clearly for the first time. He carefully specified the continuity properties that must be satisfied by functions and their variations. In problems of constrained optimization he used theorems on implicit functions to ensure that the optimizing arc was embedded in a suitable family of comparison curves.

Traditionally researchers in the calculus of variations did not identify at the outset of their investigation the precise class of comparison arcs in a given variational problem. There was no prior logical conception concerning the nature of this class. However, the δ-process introduced by Lagrange required that both the comparison arc and its slope at each point differ by only a small amount from the value and slope of the solution curve. This condition was imposed by the nature of the variational process, which involved expanding the integrand function as a Taylor series and investigating the behavior of the second variation arising in this expansion. Isaac Todhunter (*1871*, 269) in an essay on what were known as "discontinuous" solutions seems to have been the first to explicitly call attention to this limitation on the class of comparison arcs:

> [...] if we assert that the relation [Euler equation] does give a minimum, we must bear in mind that this means a minimum with respect to admissible variations [...] our investigation is not applicable to such a variation as would be required in passing from the cycloid to the discontinuous figure: in such a passage p [$= \delta y'$] would not always be indefinitely small. Of course it might be possible to give some special investigation for such a case, but certainly the case is not included in the ordinary methods of the Calculus of Variations.

In his Berlin lectures Weierstrass developed an alternative to the traditional expansion methods that extended the variational theory to a larger class of comparison curves. The precise nature of these curves was still determined by the particular technical requirements of the new method, but the logical orientation of the subject had shifted. In earlier variational research the nature of the mathematical objects was determined implicitly by the variational process that was employed. By contrast, in Weierstrass there was a self-conscious and explicit focus on the objects being studied. His work involved a more intimate connection between the foundations of real analysis and the collection of concrete techniques and results that made up the variational theory.

It is necessary to call attention to one aspect of the style in which Weierstrass developed the theory. Traditionally researchers in the calculus of variations had adopted what is referred to as the ordinary or functional approach, in which the curve is expressed as $y = y(x)$ and the variational integrand (in the simplest case) takes the form $f(x, y, y')$. A distinctive aspect of Weierstrass's approach was his adoption of a parametric approach. The curve C is represented parametrically in the form $y = y(t)$ and $x = x(t)$. Here the variational integrand takes the form $I = \int_{t_0}^{t_1} f(x, y, x', y')\, dt$ where $x' = \dfrac{dx}{dt}$ and $y' = \dfrac{dy}{dt}$. In the parametric (or homogeneous) formulation of the variational problem it is necessary to impose conditions on the variational integrand in order to ensure that the problem is independent of the particular parameterization chosen. One must attend to these conditions in developing the theory.

Weierstrass used a parametric approach throughout his lectures on the calculus of variations. Researchers who adopted his new method tended to also use a parametric approach, and this was true in the case of Zermelo. However, not all researchers followed this practice. Although the parametric approach has certain advantages, particularly from a geometric viewpoint, its analytical development is less natural than the ordinary theory. During the years around 1900 when Weierstrass's ideas were becoming more widely known, researchers such as Oskar Bolza (*1909*), William Osgood (1901) and Emile Goursat (*1905*) went to some effort to reformulate his results in terms of the ordinary theory. In the large majority of the textbook literature of the past one hundred years the ordinary approach is taken as the standard formulation of the variational problem while the parametric theory is presented as a special subject.

For the sake of exposition we will adopt the ordinary theory in explaining some of the basic ideas that underlie Weierstrass's theory. At the conclusion we indicate by way of comparison the parametric form in which Weierstrass originally presented his result. (In our account we have adopted the sensible notation used by Adolf Kneser to denote extremal and comparison curves: $y = y(x)$ is the comparison curve, while $\overline{y} = \overline{y}(x)$ is the extremal curve; both Weierstrass and Zermelo use the opposite convention.)

Suppose $y = y_0(x)$ is an arc C_0 on which the variational integral $I = \int_a^b f(x, y, y')\, dx$ is a minimum. (The case of a maximum is similar, with the inequalities reversed.) In the traditional formulation of the theory involving expansion methods and the second variation, the conditions that must be satisfied are all specified in terms of the function $y = y_0(x)$ and this function alone. Thus $y = y_0(x)$ will be a solution to the Euler equation and we must have $\dfrac{\partial^2 f(x, y_0(x), y_0'(x))}{\partial y'^2} \geq 0$ (Legendre's condition) and the Jacobi criterion must hold. One typically supposes that the family of comparison curves is of the form $C : y = y_0(x) + \epsilon\zeta(x)$. Here $\zeta(x)$ is any function subject to the usual continuity restrictions with $\zeta(a) = \zeta(b) = 0$. More generally we may have a family of comparison curves of the form $C : y = y(x, \epsilon)$ where $y(x, 0) = y_0(x)$ and $y(a, \epsilon) = y(b, \epsilon) = 0$. It is evident that the class of comparison curves is very extensive. Nevertheless, as Todhunter observed in 1870, it is also clear that there are some restrictions on this class. For example, if $y = y_0(x) + \epsilon\zeta(x)$ (ϵ small) then $y'(x) - y_0'(x) = \epsilon\zeta'(x)$ with similar relations for higher derivatives of x. It follows that the neighboring curve $y = y(x)$ differs by only a small amount from the optimizing curve $y = y_0(x)$ not just for corresponding values of y but for derivatives of y of all orders.

It turns out that it is possible for a variational integral to be a minimum for the function $y = y_0(x)$, considered with respect to a class of comparison curves of the type $y = y(x, \epsilon)$, but not be a minimum if we allow comparison curves whose slope differs by a finite amount from $y = y_0(x)$. The following simple example illustrates this situation. It is taken from Bolza's *Lectures on the calculus of variations* (*1904*, 74), one of the earliest expositions of the new theory developed by Weierstrass and extended by Zermelo. (The term "die neue Variationsrechnung" was sometimes used in reference to the Weierstrass theory.) We have $f(x, y, y') = y'^2 + y'^3$ defined on the interval $[0, 1]$.The variational integral is $I = \int_0^1 (y'^2 + y'^3)\, dx$. It follows that a solution to the Euler equation will be $y' = constant$. Since the solution must pass through the endpoints it follows that the hypothetical minimizing curve is simply $y = 0$. We have $\frac{\partial^2 f(x,y,y')}{\partial y'^2} = 2 \geq 0$, so the Legendre condition is satisfied. It is also the case that Jacobi's condition holds in this example. Hence the curve $y = 0$ minimizes the integral with respect to comparison arcs of the form $y = y(x, \epsilon)$. Consider now the comparison curve C consisting of two straight lines, the first joining the origin $(0, 0)$ to the point $(1 - p, q)$ and the second joining $(1 - p, q)$ to $(1, 0)$ (see Fig. 1). Here p is a number with $0 < p < 1$ and q is a small positive quantity. For this comparison curve we have

$$\Delta I = \frac{q^2}{p(1-p)}\left(1 + \frac{q}{1-p} - \frac{q}{p}\right).$$

We can make C lie within any neighborhood of $y = 0$ by making q sufficiently small. With q specified, it is clear that $\Delta I < 0$ for $p \ll q$. Hence I is not a minimum for the larger class of curves that includes the curve C.

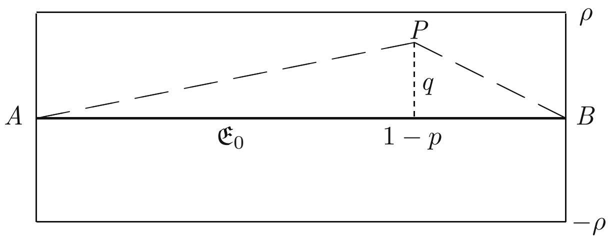

Fig. 1 (p. 74 of *Bolza 1904*)

In the terminology that was introduced by Kneser in *1900* and became standard, the traditional variational theory yields sufficient conditions for a *weak* extremum. Here each comparison curve is close to the minimizing curve at y and at all derivatives of y. By contrast, a solution will be a *strong* extremum if it is a minimum for the wider class of curves which are close to the solution curve but may have a slope that differs by a finite amount from the solution curve.

Consider again the problem of finding the curve $C_0 : y = y_0(x)$ that maximizes or minimizes $I = \int_a^b f(x, y, y')\, dx$. Suppose that the Euler and

Jacobi conditions hold for the arc C_0. We now enlarge the class of possible comparison curves to include ones whose slope differs by a finite amount from that of C. In order to establish that C_0 is a minimum with respect to this enlarged class of comparison curves it is necessary to formulate a condition that involves not just the function $y = y_0(x)$ but also the curves $C : y = y(x)$ in the comparison class. Perhaps the simplest approach would be simply to require that

$$f(x, y_0, y_0') \leq f(x, y, y') \ (a \leq x \leq b)$$

for all comparison curves $y = y(x)$. Imposing this condition is evidently not very informative, and indeed is only a restatement of the problem. Weierstrass succeeded in formulating a more meaningful condition involving a function E called the excess function. We are given the proposed minimizing arc $C_0 : y = y_0(x)$. We consider any comparison curve $C : y = y(x)$ and take any point on this curve with coordinates x and y. By assumption the point (x, y) on the comparison curve C is close to the point (x, y_0) on C_0. Let $y'(x)$ be the slope of the comparison curve at the given point. This quantity may vary by an arbitrary amount from $y_0'(x)$.

A solution to the Euler equation is called an extremal. We typically require that such a solution passes through the initial point. A key idea introduced by Weierstrass was to introduce a function $p(x, y)$—known as the slope function—defined as the slope of the extremal passing through the point (x, y) at this point. We now introduce the excess function $E(x, y, y', p)$ defined as

$$E(x, y, y', p) = f(x, y, y') - f(x, y, p) - (y' - p)\frac{\partial f}{\partial y'}(x, y, p)\,. \tag{1}$$

Consider a region or strip about the curve C_0 and suppose that for each point in this strip it is possible to determine an extremal joining the initial point and the given point. Such a set of solutions to the Euler equation is today called a field of extremals, a term introduced by Kneser. Consider the condition

$$E(x, y, y', p) \geq 0\,. \tag{2}$$

This condition is known as Weierstrass's condition. If this condition is satisfied for all comparison curves $y = y(x)$ on the interval then the value of I along C_0 is less than its value along any member of C of the comparison class. We may conclude in this case that $C_0 : y = y_0(x)$ is a strong minimum.

In later mathematics this result would be proved using something called the Hilbert invariant integral, introduced by David Hilbert in 1900 to simplify Weierstrass's method. Here we present Weierstrass's original idea, which was also adopted by Zermelo. The minimizing arc is given as $C_0 : (x, y_0(x))$ while

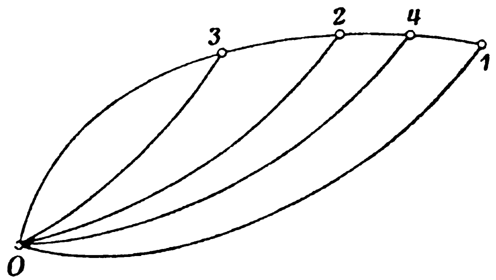

Fig. 2 (p. 219 of *Weierstrass 1927*)

the neighboring curve C is given as $C:(x,y(x))$. In Fig. 2 C_0 is the bottom arc **01**, while the neighboring curve C is the arc **0321**. We suppose that C lies in a narrow region about C_0. This region or "field" has the following property: for any point within it, there is a unique solution curve to the Euler equation—an extremal—passing through the initial point **0** and the given point. We designate extremal curves using the functional notation $\overline{y}=\overline{y}(x)$. Let (x,y) be any point on C; in Fig. 1 this point is labelled **2**. Consider the extremal curve $(x,\overline{y}(x))$ passing from **0** through **2**; in Fig. 1 it is the curve **02**. We introduce the function $\phi(x)$ defined as

$$\phi(x)=\int_{\mathbf{0}}^{\mathbf{2}} f(x,\overline{y},\overline{y}')\,dx+\int_{\mathbf{2}}^{\mathbf{1}} f(x,y,y')\,dx\,. \tag{3}$$

This integral is taken along the extremal arc from **0** to **2** and then along the comparison arc from **2** to **1**. Evidently we have $\phi(0)=\int_{\mathbf{0}}^{\mathbf{1}} f(x,y,y')\,dx$ and $\phi(1)=\int_{\mathbf{0}}^{\mathbf{1}} f(x,y_0,y_0')\,dx$. The statement that $\int_{\mathbf{0}}^{\mathbf{1}} f(x,y_0,y_0')\,dx$ is a minimum is equivalent to the inequality $\phi(0)\geq\phi(1)$, which would follow if we are able to show that $\phi(x)$ is a decreasing function of x. To do this we calculate the derivative of $\phi(x)$ and show that it is negative.

Let the integrals $\overline{I}_{02}$ and I_{21} be defined as

$$\overline{I}_{02}=\int_{\mathbf{0}}^{\mathbf{2}} f(x,\overline{y},\overline{y}')\,dx,\ \ I_{21}=\int_{\mathbf{2}}^{\mathbf{1}} f(x,y,y')\,dx\,. \tag{4}$$

From the standard formula for the variation of the integral when the endpoint is allowed to vary in both the x and y directions we have

$$\delta\overline{I}_{02}=\frac{\partial f}{\partial\overline{y}'}(x,\overline{y},\overline{y}')\delta y+\big(f(x,\overline{y},\overline{y}')-\overline{y}'\frac{\partial f}{\partial\overline{y}'}(x,\overline{y},\overline{y}')\big)\delta x\,. \tag{5}$$

We now let $\delta x=dx$ and $\delta y=dy$. Note that at the point **2** we have $\overline{y}=y$ and $\overline{y}'=p(x,y)$, where p is the slope function for the given field. Hence (5) becomes

$$\frac{d\overline{I}_{02}}{dx} = \frac{\partial f}{\partial y'}(x,y,p)y' + f(x,y,p) - p\frac{\partial f}{\partial y'}(x,y,p) \tag{6}$$
$$= \frac{\partial f}{\partial y'}(x,y,p)(y'-p) + f(x,y,p)\,.$$

The derivative of I_{21} is given immediately as

$$\frac{dI_{21}}{dx} = -f(x,y,y')\,. \tag{7}$$

Thus we have

$$\phi'(x) = \frac{d\overline{I}_{02}}{dx} + \frac{dI_{21}}{dx}$$
$$= -\left(f(x,y,y') - f(x,y,p) - (y'-p)\frac{\partial f}{\partial y'}(x,y,p)\right),$$

or

$$\phi'(x) = -E(x,y,y',p)\,. \tag{8}$$

If $E(x,y,y',p) \geq 0$ on the interval then $\phi'(x)$ is negative and $\phi(x)$ is a decreasing function, which is what was required to be proved.

It is apparent that $\phi(0) - \phi(1) = \int_{\mathbf{0}}^{\mathbf{1}} E(x,y,y',p)\,dx$. But $\phi(0) - \phi(1)$ is equal to ΔI, the variation of the integral with respect to the comparison arc. Hence we have

$$\Delta I = \int_{\mathbf{0}}^{\mathbf{1}} E(x,y,y',p)\,dx\,. \tag{9}$$

(9) is known in the modern literature as Weierstrass's theorem, although it does not appear explicitly in Weierstrass's lectures. From (9) it is clear why the function E is called the excess function, since the excess of the variational integral I in going from $y = y_0(x)$ to $y = y(x)$ is the integral of E over the given interval.

In Weierstrass's original parametric approach the variational integrand takes the form $\int_{t_0}^{t_1} F(x,y,x',y')\,dt$. The minimizing arc is given as C_0 : $(x_0(t), y_0(t))$ while the neighboring curve C is given as $C : (x(t), y(t))$. In Fig. 2 C_0 is the bottom arc **01**, while the neighboring curve C is the arc **0321**. We suppose that C lies in a narrow region about C_0. This region is supposed to be a "field" in the sense defined above: for any point within it there is a unique solution curve to the Euler equation passing through the initial point and the given point. Let $(x(t), y(t)$ be any point on C; in Fig. 2 this point is labelled **2**. Consider the extremal passing through this point; in Fig. 2 it is the curve **02**. Let $p(t), q(t)$ be the coordinate slope functions of the extremal at (x, y). The excess function E in parametric form is defined as

$$E(x, y, p, q, x', y') = F(x, y, x', y') - x' \frac{\partial F}{\partial p}(x, y, p, q) - y' \frac{\partial F}{\partial q}(x, y, p, q) . \quad (10)$$

If for each such comparison curve C we have

$$E(x, y, p, q, x', y') \geq 0 \quad (11)$$

for all values of x, y, x', y' then we may conclude that C_0 minimizes the integral $\int_{t_0}^{t_1} F(x, y, x', y')\, dt$.

In the ordinary calculus a condition that $y = y(x)$ be a minimum at $x = a$ is that $\frac{dy}{dx}(x = a) = 0$. No reference is made in this condition to neighboring values of a. Similarly, in the case of weak extrema in the calculus of variations the conditions are formulated solely in terms of the curve $C : y = y_0(x)$ and do not involve any reference to neighboring curves or functions. By contrast, Weierstrass's condition involves the comparison curve as well as the field function $p(x, y)$ defined in a neighborhood of C. It should be noted that while it is true that Weierstrass has obtained a stronger result, this is possible because the condition that must be satisfied is more restrictive; the stronger result is achieved at a higher price.

3. Zermelo's dissertation

It was inevitable that Zermelo's readership would be restricted because he was extending a mathematical theory that itself had not been published and that would have been familiar only to a fairly small group of researchers either at German universities or who had studied there. Furthermore, Weierstrass's parametric approach was not widely used in the calculus of variations; even investigators such as Ludwig Scheeffer (*1885*), Georg Erdmann (*1877*) and Edmund Husserl (*1882*) who were in a general sense part of the Weierstrass "school" and were influenced by his ideas employed the ordinary formulation of the variational problem in their researches of the 1870s and 1880s.

Zermelo's dissertation was also written in a rather formal manner, with very limited exposition of basic ideas and principles, and was excessively concerned with procedural matters, detailed formulations required in the general theory, and points of rigor. In matters of style, in its propensity for strenuous formal development, his approach bore similarities with the work of such earlier researchers as Hesse. Zermelo's work displayed as well a new element in variational research, a tendency to want to develop the subject from a larger viewpoint and to present the results as an instance of some more general and yet to be precisely specified subject. This tendency was manifested in his study of homogeneity properties of functions in the first chapter, as well as in his classification of families of curves in the second chapter.

Zermelo's dissertation would have been of primary interest to a reader who was either already familiar with Weierstrass's lectures or who was very motivated to learn about his new methods. For such a reader, the work would have been a valuable piece of mathematical research and exposition. In a fully detailed and methodical manner Zermelo developed a general theory, showing fully the non-trivial considerations that are involved in extending Weierstrass's methods to the problem $n > 1$. The first chapter was devoted to the homogeneity relations that must be satisfied in the parametric theory; the second to necessary conditions; the third to the excess function; and the final fourth chapter to sufficiency conditions involving the excess function.

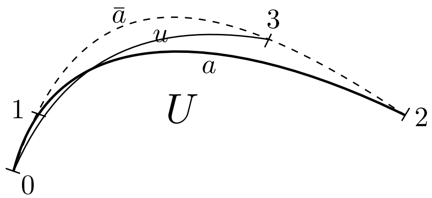

Fig. 3 (p. 90 of *Zermelo 1894*)

To indicate the basic idea behind Zermelo's development we describe how it plays out for the ordinary problem in the case where the variational integrand is a function of x, y, y' and y''. This was the setting in which Jacobi and so many other researchers set forth the theory. As before let $y = y_0(x)$ be the solution curve to the Euler equation joining **1** and **2**. In Fig. 3, from p. 90 of Zermelo's dissertation, this curve is denoted as a (note that the points are numbered slightly differently than in Weierstrass). We suppose there is a strip or region (what later was called a field) about a with the property that there is a unique extremal joining **0** (a point very close and to the left of **1** on a) and any given point **3** of the region. (In a technical refinement of Weierstrass's method, Zermelo takes the common starting point of the extremals to be **0** rather than **1** in order to simplify the analysis needed to establish the existence of the desired field. The value of the variational integral from **0** to **1** is taken to be negligible.) An arbitrary comparison curve 132: $y = y(x)$ is designated as $\overline{a}$. It is assumed that for each point **3** on $\overline{a}$ there is a unique extremal curve (a solution to the Euler equation) $\overline{y} = \overline{y}(x)$, joining **0** to **3**. This curve is designated as u in Fig. 3. Consider the function $\phi(x)$ given as

$$\phi(x) = \int_{\mathbf{0}}^{\mathbf{3}} f(x, \overline{y}, \overline{y}', \overline{y}'')\, dx + \int_{\mathbf{3}}^{\mathbf{2}} f(x, y, y', y'')\, dx\,. \tag{12}$$

Let the integrals in (12) be designated as $\overline{I}_{03}$ and I_{32}:

$$\overline{I}_{03} = \int_{\mathbf{0}}^{\mathbf{3}} f(x, \overline{y}, \overline{y}', \overline{y}'')\, dx,\ \ I_{32} = \int_{\mathbf{3}}^{\mathbf{2}} f(x, y, y', y'')\, dx\,. \tag{13}$$

$\overline{I}_{03}$ is evaluated along the extremal curve u from **0** to **3**, while I_{32} is evaluated along the comparison curve $\overline{a}$ from **3** to **2**. The key idea is to write $\delta\overline{I}_{03}$ using the variable endpoint formula applied to the case where there are second derivatives in the variational integrand. We have

$$\begin{aligned}\delta\overline{I}_{03} = {} & \frac{\partial f(x,\overline{y},\overline{y}',\overline{y}'')}{\partial\overline{y}'}\delta y + \frac{\partial f(x,\overline{y},\overline{y}',\overline{y}'')}{\partial\overline{y}''}\delta y' + \big(f(x,\overline{y},\overline{y}',\overline{y}'') \\ & -\overline{y}'\big(\frac{\partial f(x,\overline{y},\overline{y}',\overline{y}'')}{\partial\overline{y}'} - \frac{d}{dx}\frac{\partial f(x,\overline{y},\overline{y}',\overline{y}'')}{\partial\overline{y}''}\big) - \overline{y}''\frac{\partial f(x,\overline{y},\overline{y}',\overline{y}'')}{\partial\overline{y}''}\big)\delta x\,. \end{aligned} \tag{14}$$

We now let $\delta x = dx$ and $\delta y = dy$. Note that at the point **3** we have $\overline{y} = y$ and $\overline{y}' = p(x,y)$, $\overline{y}'' = q(x,y)$, where p and q are the field functions for the first and second derivatives of the extremal passing though **3**. With these designations (14) becomes

$$\begin{aligned}\frac{d\overline{I}_{03}}{dx} = {} & \frac{\partial f(x,y,p,q)}{\partial p}y' + \frac{\partial f(x,y,p,q)}{\partial q}y'' + f(x,y,p,q) \\ & -p\big(\frac{\partial f(x,y,p,q)}{\partial p} - \frac{d}{dx}\frac{\partial f(x,y,p,q)}{\partial q}\big) - q\frac{\partial f(x,y,p,q)}{\partial q}\,. \end{aligned} \tag{15}$$

We also have

$$\frac{dI_{32}}{dx} = -f(x,y,y',y'')\,. \tag{16}$$

Hence the derivative of $\phi(x)$ is

$$\phi'(x) = \int_{\mathbf{0}}^{\mathbf{3}} -E(x,y,y',y'',p,q)\,dx\,, \tag{17}$$

where

$$\begin{aligned}E(x,y,y',y'',p,q) = {} & f(x,y,y',y'') - f(x,y,p,q) \\ & -(y'-p)\big(\frac{\partial f}{\partial p} - \frac{d}{dx}\frac{\partial f}{\partial q}\big) - (y''-q)\frac{\partial f}{\partial q}\,. \end{aligned} \tag{18}$$

If $E(x,y,y',y'',p,q) \geq 0$ then it follows that

$$\phi'(x) \leq -1 \text{ and } \int_{\mathbf{0}}^{\mathbf{2}} f(x,y_0,y_0',y_0'')\,dx \leq \int_{\mathbf{0}}^{\mathbf{2}} f(x,y,y',y'')\,dx\,.$$

It is also apparent from (17) that $\phi(0)-\phi(1) = \Delta I$. Hence we have the following expression for the variation of the integral with respect to the comparison arc 032:

$$\Delta I = \int_{\mathbf{0}}^{\mathbf{2}} E(x,y,y',y'',p,q)\,dx\,. \tag{19}$$

(19) is Weierstrass's theorem and is the culminating result of Zermelo's treatise. It is stated on p. 79 of his dissertation in parametric form for the general case involving derivatives of order up to n. (It should be noted that the appellation "Weierstrass's theorem" was not used by Zermelo.) In the case $n = 1$ we have:

$$\Delta I = \int_{\sigma_1}^{\sigma_2} E(x, y, x', y', p, q)\, d\sigma\,, \tag{20}$$

where σ is the parameter and where the excess function in parametric form is given as

$$E = f(x, y, x', y') - f(x, y, p, q) - \frac{\partial f(x, y, p, q)}{\partial x'}(x' - p) - \frac{\partial f(x, y, p, q)}{\partial y'}(y' - q)\,. \tag{21}$$

In the traditional theory of sufficiency based on expansion methods it is necessary to ensure that there is no admissible δy which makes the second variation vanish. A point at which the second variation vanishes came to be called a conjugate point (a term coined by Weierstrass) and the problem has a solution only if there are no conjugate points on the interval. It is also necessary to show that there exist certain functions that allow one to transform the second variation to a suitable quadratic form. Mayer's achievement in his publications of the 1860s was to show in a very general setting that if there is no conjugate point on the interval then it is possible to produce the requisite functions needed in the transformation of the second variation. The basic problem here is one of mathematical existence. Zermelo following Weierstrass was confronted with a different kind of existence question. In order to carry out the derivation of equation (20) it is necessary to embed the extremal joining the endpoints in a field of extremals. Zermelo supplemented his presentation of (20) with an extended discussion of the existence of such a field and the conditions that are required for it. His approach was to write down an analytical condition stating that there is no conjugate point on the interval. From this condition it is shown that there is a strip or field about the given extremal joining the points A and B with the following property: for each point P in this region there is a unique extremal passing through it. If the variational integrand contains derivatives up to order n, then the extremal at P will have an nth order derivative at P that is a function of the values of $x, y', y'', ..., y^{(n-1)}$ there. Field-theoretic questions were an important part of Zermelo's theory and would become the focus of much further work in the calculus of variations.

4. Further discussion of Zermelo's theory

The variable endpoint formula (14) plays an essential role in the derivation of the condition $E(x, y, y', y'', p, q) \geq 0$. In (14) the increments δx, δy and $\delta y'$ are small increments in x, y and y'. It is immediately clear that the slope of any comparison curve may only differ from the slope of the actual solution curve by a small amount. Thus in the Weierstrassian theory the case $n = 2$ is essentially different from the case $n = 1$, where the slope of the comparison curve may differ by any finite amount from the slope of the solution curve. In fact, the restriction on the slope of the comparison curve in the case $n = 2$ is the same as in the case of weak extrema for $n = 1$! In the general case in which the integrand function f contains derivatives of y up to order n, the comparison curve must differ by only a small amount at its derivatives up to order $n-1$. Of course, the derivatives of order n and higher may take on any value, so it is clear that the class of comparison curves is still larger than in the case of weak extrema.

There are several aspects of Zermelo's theory that somewhat limited its influence on the later development of the calculus of variations. His formulation using a parametric approach seems to have stemmed from a desire to remain faithful to Weierstrass's original exposition. However, the parametric formulation really constitutes a special topic, valuable from a certain geometric viewpoint but much too awkward to form the primary basis of the subject. Another important event was Hilbert's introduction (*1900a, 1905*) of the invariant integral, giving rise to an essential tool that transformed the theory. As we show below, the use of the invariant integral simplified the derivation of Weierstrass's theorem and provided a tool that could be applied to more general problems.

Mention should also be made of the central problem of concern to Zermelo. Although the variational problem with higher-order derivatives had been very prominent in the writings of Jacobi and his successors, it virtually disappeared from the textbook literature in the twentieth century. Instead one developed the theory for n dependent variables with variational integrands that contain only the first derivatives of the variables. The general variational integral is here $\int_a^b f(x, y_1, y_2, ..., y_n, y_1', y_2', ..., y_n')\, dx$. The investigation of sufficiency is carried out in this setting. The case of higher order derivatives is then treated as an optimization problem subject to constraint. The basic idea goes back to *Clebsch 1858a,b* and is illustrated by the problem of minimizing $\int_a^b f(x, y, y', y'')\, dx$. This problem can be reformulated as the problem of minimizing the integral $\int_a^b f(x, y_1, y_2, y_2')\, dx$ subject to the side constraint $y_1' - y_2 = 0$. Using the multiplier rule this problem is equivalent to minimizing the integral $\int_a^b \big(f(x, y_1, y_2, y_2') + \lambda(x)(y_1' - y_2)\big)\, dx$. The Euler equations for this problem are $\frac{\partial f}{\partial y_1} - \frac{d(\lambda(x))}{dx} = 0$ and $\frac{\partial f}{\partial y_2} - \lambda(x) - \frac{d(\frac{\partial f}{\partial y_2'})}{dx} = 0$. Noting that $y_1' - y_2 = 0$ we find that these two equations reduce to

$\frac{\partial f}{\partial y_1} - \frac{d}{dx}\frac{\partial f}{\partial y_1'} + \frac{d^2}{dx^2}\frac{\partial f}{\partial y_1''} = 0$, the Euler equation for $\int_a^b f(x,y,y',y'')\,dx$ with $y = y_1$. Sufficient conditions for the problem $\int_a^b f(x,y,y',y'')\,dx$ are in turn deduced from the general theory of sufficiency developed for the integral $\int_a^b f(x,y_1,y_2,...,y_n,y_1',y_2',...,y_n')\,dx$ and applied to the particular integral $\int_a^b \big(f(x,y_1,y_2,y_2') + \lambda(x)(y_1' - y_2)\big)\,dx$.

Despite these limitations, Zermelo's dissertation was important in bringing Weierstrass's ideas forward in published form and in developing the theory in new directions. It provided a source for the work of Kneser, Hilbert, Mayer (*1904*), Osgood (*1900/1901, 1901a*) and Bolza as well as the other researchers of the period. The work is cited no less than eight times in Bolza's *Lectures on the calculus of variations* (*1904*), on pp. 9, 35, 72, 76, 82, 119, 143, 174. Special note should be made of Bolza's discussion (p. 174) in Chapter V of transversals to sets of extremals, where attention is called to a result proved by Zermelo (p. 96 of his dissertation) concerning the envelope of a set of extremals.

In the history of the calculus of variations there are examples of researchers who began in this branch of mathematics and continued to make important contributions to it throughout their career. One might mention here such figures as Lagrange, Mayer and the American mathematician Gilbert Bliss. However, Zermelo belongs to another historical pattern of investigators who cut their teeth in the calculus of variations and then went on to prominence in very different fields of research. One could mention in addition to Zermelo (set theory) such figures as Charles Delaunay (celestial mechanics), Clebsch (algebraic geometry), Husserl (philosophy), and Herman Goldstine (computer science and numerical analysis).

5. Epilogue: Hilbert's invariant integral

Weierstrass's theorem in parametric form is given by (20). This statement is evidently relative to the particular parameterization chosen. Let us assume that we write the theorem in a form that is independent of any particular parameterization. One obvious way to do this would be to develop the theory in traditional ordinary form, using x as the independent variable and y as the dependent variable. In ordinary form Weierstrass's theorem is written:

$$\Delta I = \int_{x_1}^{x_2} E(x,y,y',p)\,dx \tag{22}$$

where

$$E = f(x,y,y') - f(x,y,p) - (y'-p)\frac{\partial f}{\partial y'}(x,y,p)\,. \tag{23}$$

We have

$$\Delta I = \int_{x_1}^{x_2} \big(f(x,y,y') - f(x,y_0,y_0')\big)\,dx\,, \tag{24}$$

where $y_0 = y_0(x)$ is the extremal joining the initial and final points. From (22), (23) and (24) it follows that

$$\begin{aligned}\int_{x_1}^{x_2} \big(f(x,y,y') - f(x,y_0,y_0')\big)\,dx \\ = \int_{x_1}^{x_2} \Big(f(x,y,y') - f(x,y,p) - (y'-p)\frac{\partial f}{\partial p}(x,y,p)\Big)\,dx\,,\end{aligned} \tag{25}$$

or

$$\int_{x_1}^{x_2} f(x,y_0,y_0')dx = \int_{x_1}^{x_2} \Big(f(x,y,p) + (y'-p)\frac{\partial f}{\partial p}(x,y,p)\Big)\,dx\,. \tag{26}$$

Because $y = y_0(x)$ is given, the quantity on the left side of (26) is constant. Hence from (26) we deduce that the integral

$$H = \int_{x_1}^{x_2} \Big(f(x,y,p) + (y'-p)\frac{\partial f}{\partial p}(x,y,p)\Big)\,dx$$

has the same value for all comparison curves $y = y(x)$: the integral H is invariant with respect to the path.

Hilbert did not discuss how he arrived at the idea of the invariant integral: in his account it is something that is introduced without any explanation. However, it is reasonable to suppose that he first came across the idea by simply writing down Weierstrass's theorem in ordinary form, and noticing as we did above that the integral H is invariant. It was then a simple matter to show directly that H is invariant. Using the Euler equation $\dfrac{\partial f}{\partial y} - \dfrac{d}{dx}\dfrac{\partial f(x,y,p)}{\partial p} = 0$ it is straightforward to prove that

$$\frac{\partial}{\partial y}\Big(f(x,y,p) - p\frac{\partial f}{\partial p}(x,y,p)\Big) = \frac{\partial}{\partial x}\Big(\frac{\partial f}{\partial p}(x,y,p)\Big)\,, \tag{27}$$

and so the condition for the integrability of the differential form $\big(f(x,y,p) - p\dfrac{\partial f}{\partial p}(x,y,p)\big)\,dx + \dfrac{\partial f}{\partial p}(x,y,p)\,dy$ is satisfied. Having established that H is invariant directly we can then use this fact to provide a new proof of Weierstrass's theorem, which is what Hilbert did. A significant advantage of Hilbert's approach is that a wider class of fields can be used in the sufficiency proof. In the theory of Weierstrass and Zermelo, the extremals of the field pass through a single point: in the case of Weierstrass this point is the initial point of the extremal, and in the case of Zermelo it is a point very close to the initial point (see *Bolza 1904*, 82, note 1). Such a field is said to be a central field. By contrast, the proof of Weierstrass's theorem using the invariant integral applies to any covering of a region surrounding the solution curve by a family of extremals in which only one extremal passes through each point of the region. The invariant integral can also be applied to more general variational problems, and is an important field-theoretic tool in the investigation of extrema. (For later literature related to this subject, see *Hadamard 1910, Bliss 1925* and *Bliss 1946*. The relevant history may be found in *Thiele 2007*.)

In the publication of his Paris address of 1900, where Hilbert first presented the idea of the invariant integral, he referred to Kneser's *Lehrbuch* but not to Zermelo. However, we know that he held Zermelo's work in the calculus of variations in high regard. In 1903 he recommended Zermelo for a position at the University of Breslau, writing[1] "Zermelo is a modern mathematician who combines versatility with depth in a rare way. He is an expert in the calculus of variations (and working on a comprehensive monograph about it). I regard the calculus of variations as a branch of mathematics which will belong to the most important ones in the future." Hilbert added that some years earlier, "Zermelo was my main mathematical company, and I have learnt a lot from him, for example, the Weierstrassian calculus of variations." Zermelo was not offered the position.

[1] In a letter to the hiring committee; cf. *Ebbinghaus 2007*, 35–36, 276–277.