independent unknowns of the problem as we please, we may often succeed in carrying through the treatment of a problem by transformation into free functions; for an unknown number may be chosen quite arbitrarily in all its n coefficients, and each of the following unknowns looses, as a function of the observations, only an arbitrary coefficient in comparison to the preceding one; even the n^{th} unknown can still get an arbitrary factor. Altogether are $\frac{1}{2}n(n+1)$ of the n^2 coefficients of these transformations arbitrary.

But if the problem does not admit of any solution through a transformation into free functions, the mean errors for the several unknowns, no matter how many there may be, can be computed only in such a way that each of the sought numbers are directly expressed as a linear function of the observations. The same holds good also when the laws of errors of the observations are not typical, and we are to examine how it is with λ_3 and the higher half-invariants in the laws of errors of the sought functions.

Still greater importance, nay a privileged position as the only legitimate proceeding, gets the transformation into a complete set of free functions in the over-determined problems, which are rejected as self-contradictory in exact mathematics. When we have a collection of observations whose number is greater than the number of the independent unknowns of the problem, then the question will be to determine laws of actual errors from the standpoint of the observations. We must mediate between the observations that contradict one another, in order to determine their mean numbers, and the discrepancies themselves must be employed to determine their mean deviations, etc. But as we have not to do with repetitions, the discrepancies conceal themselves behind the changes of the circumstances and require transformations for their detection. All the functions of the observations which, as the problem is over-determined, have theoretically necessary values, as, for instance, the sum of the angles of a plane triangle, must be selected for special use. Besides, those of the unknowns of the problem, to the determination of which the theory does not contribute, must come forth by the transformation by which the problem is to be solved.

As we shall see in the following chapters on Adjustment, it becomes of essential moment here that we transform into a system of free functions. The transformation begins with mutually free observations, and must not itself introduce any bond, because the transformed functions in various ways must come forth as observations which determine laws of actual errors.

X. ADJUSTMENT.

§ 48. Pursuing the plan indicated in § 5 we now proceed to treat the determination of laws of errors in some of the cases of observations made under varying or different essential circumstances. But here we must be content with very small results. The general problem will hardly ever be solved. The necessary equations must be taken from the totality of the hypotheses or theories which express all the terms of each law of error — say their half-invariants — as functions of the varying or wholly different circumstances of the observations. Without great regret, however, the multiplicity of these *theoretical equations* can be reduced considerably, if we suppose all the laws of errors to be exclusively of the typical form.

For each observation we need then only two theoretical equations, one representing its presumptive mean value $\lambda_1(o_i)$, the other the square of its mean error $\lambda_2(o_i)$, as functions of the essential circumstances. But the theoretical equations will generally contain other unknown quantities, the arbitrary constants of the theory, and these must be eliminated or determined together with the laws of errors. The complexity is still great enough to require a further reduction.

We must, preliminarily at all events, suppose the mean errors to be given directly by theory, or at least their mutual ratios, the weights. If not, the problems require a solution by the indirect proceeding. Hypothetical assumptions concerning the $\lambda_2(o_i)$ are used in the first approximation and checked and corrected by special operations which, as far as possible, we shall try to expose beside the several solutions, using for brevity the word "criticism" for these and other operations connected with them.

But even if we confine our theoretical equations to the presumptive means $\lambda_1(o_i)$ and the arbitrary unknown quantities of the theory, the solutions will only be possible if we further suppose the theoretical equations to be linear or reducible to this form. Moreover, it will generally be necessary to regard as exactly given many quantities really found by observation, on the supposition only that the corresponding mean errors will be small enough to render such irregularity inoffensive.

In the solution of such problems we must rely on the found propositions about functions of observations with exactly given coefficients. In the theoretical equations of each problem sets of such functions will present themselves, some functions appearing as given, others as required. The observations, as independent variables of these functions, are, now the given observed values o_i, now the presumptive means $\lambda_1(o_i)$; the latter are, for instance, among the unknown quantities required for the exact satisfaction of the theoretical equations.

What is said here provisionally about the problems that will be treated in the following, can be illustrated by the simplest case (discussed above) of n repetitions of the same observation, resulting in the observed values $o_1, \ldots o_n$. If we here write the theoretical equations without introducing any unnecessary unknown quantities, they will show the forms $0 = \lambda_1(o_i) - \lambda_1(o_k)$ or, generally, $0 = \lambda_1[a(o_i - o_k)]$. But these equations are evidently not sufficient for the determination of any $\lambda_1(o_i)$, which they only give if another $\lambda_1(o_k)$ is found beforehand. The sought common mean cannot be formed by the introduction of the observed values into any function $[u(o_i - o_k)]$, these erroneous values of the functions being useful only to check $\lambda_2(o_i)$ by our criticism. But we must remember what we know about free functions: that the whole system of these functions $[u(o_i - o_k)]$ is only a partial system, with $n-1$ differences $o_i - o_k$ as representatives. The only n^{th} functions which can be free of this partial system, must evidently be proportional to the sum $o_1 + \dots + o_n$, and by this we find the sought determination by

$$\lambda_1(o_i) = \frac{1}{n}(o_1 + \dots + o_n),$$

the presumptive mean being equal to the actual mean of the observed values.

If we thus consider a general series of unbound observations, $o_1, \dots o_n$, it is of the greatest importance to notice first that two sorts of special cases may occur, in which our problem may be solved immediately. It may be that the theoretical equations concerning the observations leave some of the observations, for instance o_1, quite untouched; it may be also that the theory fully determines certain others of the observations, for instance o_n.

In the former case, that is when none of all the theories in any way concern the observation o_1, it is evident that the observed value o_1 must be approved unconditionally. Even though this observation does not represent any mean value found by repetitions, but stands quite isolated, it must be accepted as the mean $\lambda_1(o_1)$ in its law of presumptive errors, and the corresponding square of the mean error $\lambda_2(o_1)$ must then be taken, unchanged, from the assumed investigations of the method of observation.

If, in the latter case, o_n is an observation which directly concerns a quantity that can be determined theoretically (for instance the sum of the angles of a rectilinear triangle), then it is, as such, quite superfluous as long as the theory is maintained, and then it must in all further computations be replaced by the theoretically given value; and in the same way $\lambda_2(o_n)$ must be replaced by zero, as the square of the mean error on the errorless theoretical value.

The only possible meaning of such superfluous observations must be to test the correctness of the theory for approbation or rejection (a third result is impossible when we are dealing with any real theory or hypothesis), or to be used in the criticism.

In such a test it must be assumed that the theoretical value corresponding to o_n, which we will call u_n, is identical with the mean value in the law of presumptive errors for o_n, consequently, that $u_n = \lambda_1(o_n)$, and the condition of an affirmative result must be obtained from the square of the deviation, $(o_n - u_n)^2$ in comparison with $\lambda_2(o_n)$. The equation $(o_n - u_n)^2 = \lambda_2(o_n)$ need not be exactly satisfied, but the approximation must at any rate be so close that we may expect to find $\lambda_2(o_n)$ coming out as the mean of numerous observed values of $(o_n - u_n)^2$. Compare § 34.

§ 44. If then all the observations $o_1 \ldots o_n$ fall under one or the other of these two cases, the matter is simple enough. But generally the observations o_i will be connected by theoretical equations of condition which, separately, are insufficient for the determination of the single ones. Then the question is whether we can transform the series of observations in such a way that a clear separation between the two opposite relations to the theory can be made, so that some of the transformed functions of the observations, which must be mutually free in order to be treated as unbound observations, become quite independent of the theory, while the rest are entirely dependent on it. This can be done, and the computation with observations in consequence of these principles, is what we mean by the word "adjustment".

For as every theory can be fully expressed by a certain number, $n - m$, of theoretical equations which give the exact values of the same number of mutually independent linear functions, and as we are able, as we have seen, from every observation or linear function of the observations, in one single way, to separate a function which is free and independent of these just named theoretically given functions, and which must thus enter into another system, represented by m functions, this system must include all those functions of the observations which are independent of the theory and cannot be determined by it. Each of the thus mutually separated systems can be imagined to be represented, the theoretical system by $n-m$, the non-theoretical or empirical system by m mutually free functions, which together represent all observations and all linear functions of the same, and which may be looked upon as a complete, transformed system of free functions, consequently as unbound observations. The two systems can be separated in a single way only, although the representation of each partial system, by free functions, can occur in many ways.

It is *the idea of the adjustment*, by means of this transformation, to give the theory its due and the observations theirs, in such a way that every function of the theoretical system, and particularly the $n - m$ free representatives of the same, are exchanged, each with its theoretically given value, which, pursuant to the theory, is free of error. On the other hand, every function of the empiric system and, particularly, its m free representatives remain unchanged as the observations determine them. Every general function of the n observations $[do]$ and, particularly, the observations themselves are during the adjustment split into two univocally determined addenda: the theoretical function $[d'o]$, which should have a fixed value D', and the non-theoretical one $[d''o]$. The former $[d'o]$ is by the adjustment changed into D' and made errorless, the latter is not changed at all. The result of the adjustment, $D' + [d''o]$, is called the adjusted value of the function, and may be indicated as $[du]$, the adjusted values of the observations themselves being written $u_1 \ldots u_N$. The forms of the functions are not broken, as the distributive principle $f(x+y) = f(x)+f(y)$ holds good of every homogeneous linear function.

The determination of the adjusted values is analogous to the formation of the mean values of laws of errors by repetitions. For theoretically determined functions the adjusted value is the mean value on the very law of presumptive errors; for the functions that are free of the whole theory, we have the extreme opposite limiting case, mean values represented by an isolated, single observation. In general the adjusted values $[du]$ are analogous to actual mean values by a more or less numerous series of repetitions. For while $\lambda_2([do]) = \lambda_2[d'o]+\lambda_2[d''o]$, we have $\lambda_2[du] = \lambda_2(D')+\lambda_2[d''u] = \lambda_2[d''o]$, consequently smaller than $\lambda_2[do]$. The ratio $\frac{\lambda_2[do]}{\lambda_2[du]}$ is analogous to the number of the repetitions or the weight of the mean value.

§ 45. By *"criticism"* we mean the trial of the — hypothetical or theoretical — suppositions, which have been made in the adjustment, with respect to the mean errors of the observations; new determinations of the mean errors, analogous to the determinations by the square of the mean deviations, μ_2, will, eventually also fall under this. The basis of the criticism must be taken from a comparison of the observed and the adjusted values, for instance the differences $[do]-[du]$. According to the principle of § 34 we must expect the square of such a difference, on an average, to agree with the square of the corresponding mean error, $\lambda_2([do]-[du])$, but as $[do]-[du] = [d'o]-D'$, and $\lambda_2[d'o] = \lambda_2[do]-\lambda_2[du]$, we get

$$\lambda_2([do]-[du]) = \lambda_2[do]-\lambda_2[du], \tag{68}$$

which, by way of parenthesis, shows that the observed and the adjusted values of the same function or observation cannot in general be mutually free. We ought then to have

$$\frac{([do]-[du])^2}{\lambda_2[do]-\lambda_2[du]} = 1 \tag{69}$$

on the average; and for a sum of terms of this form we must expect the mean to approach the number of the terms, nota bene, if there are no bonds between the functions $[do]-[du]$; but in general such bonds will be present, produced by the adjustment or by the selection of the functions.

It is no help if we select the original and unbound observations themselves, and consequently form sums such as

$$\left[\frac{(o-u)^2}{\lambda_2(o)-\lambda_2(u)}\right],$$

for after the adjustment and its change of the mean errors, $u_1 \ldots u_n$ are not generally free functions such as $o_1 \ldots o_n$. Only one single choice is immediately safe, viz., to stick to the system of the mutually free functions which, in the adjustment, have themselves represented the observations: the $n-m$ theoretically given functions and the m which the adjustment determines by the observations. Only of these we know that they are free both before and after the adjustment. And as the differences of the last-mentioned m functions identically vanish, the criticism must be based upon the $n-m$ terms corresponding to the theoretically free functions $[ao]=A, \ldots [b'o]=B'$ of the series

$$\frac{([ao]-A)^2}{\lambda_2[ao]-\lambda_2[au]}+\cdots+\frac{([b'o]-B')^2}{\lambda_2[b'o]-\lambda_2[b'u]}=\frac{([ao]-A)^2}{[aa\lambda_2]}+\cdots+\frac{([b'o]-B')^2}{[b'b'\lambda_2]}, \quad (70)$$

the sum of which must be expected to be $=n-m$.

Of course we must not expect this equation to be strictly satisfied; according to the second equation (46) the square of the mean error on 1, as the expected value of each term of the series, ought to be put down $=2$; for the whole series, consequently, we can put down the expected value as $n-m \pm \sqrt{2(n-m)}$.

But now we can make use of the proposition (66) concerning the free functions. It offers us the advantage that we can base the criticism on the deviations of the several observations from their adjusted values, the latter, we know, being such a special set of values as may be compared to the observations like $v_1 \ldots v_n$ loc. cit.; $u_1 \ldots u_n$ are only distinguished from $v_1 \ldots v_n$ by giving the functions which are free of the theory the same values as the observations. We have consequently

$$\frac{([ao]-A)^2}{[aa\lambda_2]}+\cdots+\frac{([b'o]-B')^2}{[b'b'\lambda_2]}=\left[\frac{(o-u)^2}{\lambda_2(o)}\right]=n-m\pm\sqrt{n-m}. \quad (71)$$

If we compare the sum on the right side in this expression with the above mentioned $\left[\frac{(o-u)^2}{\lambda_2(o)-\lambda_2(u)}\right]$, which we dare not approve on account of the bonds produced by the adjustment, then there is no decided contradiction between putting down $\left[\frac{(o-u)^2}{\lambda_2(o)}\right]$ at the smaller value $n-m$ only, while $\left[\frac{(o-u)^2}{\lambda_2(o)-\lambda_2(u)}\right]$, by the diminution of the denominators, can get the value n; only we can get no certainty for it.

The ratios between the corresponding terms in these two sums of squares, consequently $\frac{\lambda_2(o)-\lambda_2(u)}{\lambda_2(o)}=1-\frac{\lambda_2(u)}{\lambda_2(o)}$, we call "scales", viz. *scales* for measuring the influence of the adjustment on the single observation. More generally we call

$$1-\frac{\lambda_2[du]}{\lambda_2[do]} \text{ the scale for the function } [do]. \quad (72)$$

If the scale for a function or observation has its greatest possible value, viz. 1, $\lambda_2[du]=0$. The theory has then entirely decided the result of the adjustment. But if the scale sinks to its lowest limit $=0$, we get just the reverse $\lambda_2[du]=\lambda_2[do]$, i. e. the theory has had no influence at all; the whole determination is based on the accidental value of the observation, and for observations in this case we get $\frac{(o-u)^2}{\lambda_2(o)-\lambda_2(u)} = \frac{0}{0}$. Even though the scale has a finite, but very small value it will be inadmissible to depend on the value of such a term becoming = 1. We understand now, therefore, the superiority of the sum of the squares $\left[\frac{(o-u)^2}{\lambda_2(o)}\right] = n-m$ to the sum of the squares $\left[\frac{(o-u)^2}{\lambda_2(o)-\lambda_2(u)}\right] = n$ as a bearer of the summary criticism.

We may also very well, on principle, sharpen the demand for adjustment on the part of the criticism, so that not only the whole sum of the squares $\left[\frac{(o-u)^2}{\lambda_2(o)}\right]$ must approach the value $n-m$, but also partial sums, extracted from the same, or even its several terms, must approach certain values. Only, they are not to be added up as numbers of units, but must be sums of the scales of the corresponding terms. So much we may trust to the sum of the squares $\left[\frac{(o-u)^2}{\lambda_2(o)-\lambda_2(u)}\right]$ that this principle, when judiciously applied, may be considered as fully justified.

The sum of the squares $\left[\frac{(o-u)^2}{\lambda_2(o)}\right]$ possesses an interesting property which all other authors have used as the basis of the adjustment, under the name of *"the method of the least squares"*. The above sum of the squares gets by the adjustment the least possible value that $\left[\frac{(o-v)^2}{\lambda_2(o)}\right]$ can get for values $v_1 \ldots v_n$ which satisfy the conditions of the theory. The proposition (66) concerning the free functions shows that the condition of this minimum is that $[c''o] = [c''u], \ldots [d'''o] = [d'''u]$ for all the free functions which are determined by the observations, consequently just by putting for each v the corresponding adjusted value u.

§ 46. The carrying out of adjustments depends of course to a high degree on the form in which the theory is given. The theoretical equations will generally include some observations and, beside these, some unknown quantities, elements, in smaller number than those of the equations, which we just want to determine through the adjustment. This general form, however, is unpractical, and may also easily be transformed through the usual mathematical processes of elimination. We always go back to one or the other of two extreme forms which it is easy to handle: *either*, we assume that all the elements are eliminated, so that the theory is given as above assumed by $n-m$ linear equations of condition with theoretically given coefficients and values, ***adjustment by correlates;*** or, we manage to get an equation for each observation, consequently no equations of condition between several observations. This is easily attained by making the number of the elements as large ($= m$) as may be necessary: we may for instance give some values of observations the name of elements. This sort of adjustment is called ***adjustment by elements.*** We shall discuss these two forms in the following chapters XI and XII, first the adjustment by correlates whose rules it is easiest to deduce. In practice we prefer adjustment by correlates when m is nearly as large as n, adjustment by elements when m is small.

XI. ADJUSTMENT BY CORRELATES.

§ 47. We suppose we have ascertained that the whole theory is expressed in the equations $[au] = A, \ldots [cu] = C$, where the adjusted values u of the n observations are the only unknown quantities; we prefer in doubtful cases to have too many equations rather than too few, and occasionally a supernumerary equation to check the computation. The first thing the adjustment by correlates then requires is that the functions $[ao] \ldots [co]$, corresponding to these equations, are made free of one another by the schedule in § 42.

Let $[ao], \ldots [c''o]$ indicate the $n-m$ mutually free functions which we have got by this operation, and let us, beside these, imagine the system of free functions completed by m other arbitrarily selected functions, $[d'''o], \ldots [g^{\nu}o]$, representatives of the empiric functions; the adjustment is then principally made by introducing the theoretical values into this system of free functions. It is finally accomplished by transforming back from the free modified functions to the adjusted observations. For this inverse transformation, according to (62), the n equations are:

$$o_i = \left\{\frac{a_i}{[aa\lambda_2]}[ao] + \ldots + \frac{c''_i}{[c''c''\lambda_2]}[c''o] + \frac{d'''_i}{[d'''d'''\lambda_2]}[d'''o] + \ldots + \frac{g^{\nu}_i}{[g^{\nu}g^{\nu}\lambda_2]}[g^{\nu}o]\right\}\lambda_2(o_i) \quad (73)$$

and according to (35) (compare also (63))

$$\left.\begin{aligned} \lambda_2(o_i) &= \left\{\frac{a_i^2\lambda_2^2(o_i)}{[aa\lambda_2]^2}\cdot\lambda_2[ao] + \ldots\ldots\ldots\ldots + \frac{g^{\nu 2}_i\lambda_2^2(o_i)}{[g^{\nu}g^{\nu}\lambda_2]^2}\lambda_2[g^{\nu}o]\right\} \\ &= \left\{\frac{a_i^2}{[aa\lambda_2]} + \ldots + \frac{c''^2_i}{[c''c''\lambda_2]} + \frac{d'''^2_i}{[d'''d'''\lambda_2]} + \ldots + \frac{g^{\nu 2}_i}{[g^{\nu}g^{\nu}\lambda_2]}\right\}\lambda_2^2(o_i) \end{aligned}\right\} \quad (74)$$

As the adjustment influences only the $n-m$ first terms of each of these equations, we have, because $[au] = A, \ldots [c''u] = C''$, and $\lambda_2[au] = \ldots = \lambda_2[c''u] = 0$,

$$u_i = \left\{\frac{a_i}{[aa\lambda_2]}A + \ldots + \frac{c''_i}{[c''c''\lambda_2]}C'' + \frac{d'''_i}{[d'''d'''\lambda_2]}[d'''o] + \ldots + \frac{g^{\nu}_i}{[g^{\nu}g^{\nu}\lambda_2]}[g^{\nu}o]\right\}\lambda_2(o_i) \quad (75)$$

and

$$\lambda_2(u_i) = \left\{\frac{d'''^2_i}{[d'''d'''\lambda_2]} + \ldots + \frac{g^{\nu 2}_i}{[g^{\nu}g^{\nu}\lambda_2]}\right\}\lambda_2^2(o_i). \quad (76)$$