Notes on Constructive Mathematics

by

PER MARTIN-LÖF

ALMQVIST & WIKSELL

STOCKHOLM

Uppsala 1970

The aim of this introduction is to explain briefly what sense we are giving to the word constructive.

All objects which we shall consider are to be *constructive objects* by which we mean finite configurations of signs. The signs, which can immediately be recognized as being equal or different, are treated as atoms that cannot be further decomposed. Constructive objects are to be considered as concrete objects, that is, in the very end as existing in time and space. For example, the formulas and proofs of a formal system are constructive objects in this sense, whereas meaningful assertions and informal proofs are not. Also, sets in the usual intuitive sense are certainly not constructive objects.

We accept the analysis of what it means to operate on constructive objects according to mechanical rules which was given by Post 1936 and Turing 1937*a*. On this analysis, the *rules* of computation as well as the *computations* themselves are again constructive objects, and it is decidable whether or not a computation is correct according to a given rule. More precisely, there is a decidable predicate $T(e, m, n)$ which expresses that n is a correct computation from the initial data m according to the rule e, in which case we can read off the result $U(n)$ of the computation n. Since constructive objects are simply configurations of signs, they may be coded into natural numbers. In particular, we may consider e, m and n to be natural numbers.

We shall say that e is *applicable* to m if there exists an n such that $T(e, m, n)$,

$$\bigvee n T(e, m, n).$$

It results immediately that there can be no rule which allows us to decide for every rule m whether or not it is applicable to itself. For then there would be a rule e which is applicable to m if and only if m is not applicable to itself,

$$\bigvee n T(e, m, n) \leftrightarrow - \bigvee n T(m, m, n).$$

In particular, e is applicable to itself if and only if e is not applicable to itself,

$$\bigvee n T(e, e, n) \leftrightarrow -\bigvee n T(e, e, n),$$

which is a contradiction.

In its final analysis, every theorem of constructive mathematics has the affirmative form that a constructive object with a certain property has been found. The crucial question is what properties we are going to consider as constructively meaningful.

The simplest case is that of a decidable property $P(n)$ which is simply a rule which allows us for every natural number n to compute a truth value. Applied to such properties, the classical propositional connectives $\wedge$ (and), $\vee$ (or), $-$ (not) and $\rightarrow$ (implies) have a clear computational meaning which is given by the ordinary truth tables and the laws of the classical propositional logic apply. For example, $P(n) \vee -P(n)$ is a decidable property which holds for all natural numbers n.

Classically, arbitrary logical formulas built up from decidable predicates by means of the propositional connectives and the quantifiers $\bigwedge$ (for all) and $\bigvee$ (there exists) applied to number variables are considered to be meaningful in terms of the classical notion of truth, and the laws of the classical first order predicate logic are valid under this interpretation. For example, the formula

$$\bigwedge e \bigvee m \bigwedge n (T(e, e, m) \vee -T(e, e, n))$$

is true classically.

Constructively, arbitrary formulas built up from decidable predicates by means of $\wedge$, $\vee$, $-$, $\rightarrow$, $\bigwedge$ and $\bigvee$ are not considered meaningful without further analysis, and only when such an analysis has been given is it meaningful to ask what laws of logic that are valid.

We shall begin by considering properties of the form Π^0_2,

$$\bigwedge m \bigvee n P(m, n),$$

where P is decidable, to be constructively meaningful. The forms Π^0_1 and Σ^0_1, that is, $\bigwedge n P(n)$ and $\bigvee n P(n)$, respectively, are special cases of this. The most typical example of a Π^0_2 property is that of a rule e being applicable to all numbers m,

$$\bigwedge m \bigvee n T(e, m, n).$$

When stating $\bigwedge m \bigvee n P(m, n)$ we mean that we have found a method which allows us, whenever a natural number m is given to us, to find a natural number n such that $P(m, n)$ is true. Note that the intended meaning of a Π^0_2 statement is the same under the classical interpretation. For if $\bigwedge m \bigvee n P(m, n)$ is true classically, then we actually have a method of finding for every m a natural number n such that $P(m, n)$ is true. We simply compute one after anohther the truth values of

$$P(m, 1), P(m, 2), \ldots$$

until we hit upon an n for which $P(m, n)$ is true. Thus, there is no difference between the classical and constructive interpretation of Π^0_2 statements as far as the intended meaning is concerned. The difference lies in the methods of proof that are accepted. For example, let $P(n)$ be the statement that n is the Gödel number of a proof in axiomatic set theory (or even second order arithmetic) whose endformula is $0 = 1$. Then P is a decidable predicate and the statement

$$\bigwedge n - P(n),$$

which expresses the consistency of the formal system, holds classically although we possess no constructive proof of it at present.

Consider now an arbitrary prenex formula

$$\bigwedge m_1 \bigvee n_1 \ldots \bigwedge m_j \bigvee n_j P(m_1, n_1, \ldots, m_j, n_j)$$

with P decidable. We shall say that such a formula is *constructively valid* if we have found rules $f_1, f_2, \ldots, f_j$ such that f_i is applicable to all i tuples of natural numbers, $i = 1, \ldots, j$, and

$$\bigwedge m_1 \ldots m_j P(m_1, f_1(m_1), \ldots, m_j, f_j(m_1, \ldots, m_j)).$$

Written out in full this means that we have found rules $f_1, f_2, \ldots, f_j$ such that

$$\bigwedge m_1 \ldots m_j \bigvee n_1 \ldots n_j (T(f_1, m_1, n_1) \wedge \ldots$$
$$\wedge T(f_j, (m_1, \ldots, m_j), n_j) \wedge P(m_1, U(n_1), \ldots, m_j, U(n_j))).$$

This property is considered constructively meaningful since it is of the form Π^0_2.

The formula

$$\bigwedge e \bigvee m \bigwedge n (T(e, e, m) \vee -T(e, e, n))$$

is not constructively valid although it is classically true. For suppose that f is a rule which is applicable to all natural numbers and satisfies

$$\wedge en(T(e, e, f(e)) \vee -T(e, e, n)).$$

Then

$$T(e, e, f(e)) \leftrightarrow \vee n T(e, e, n)$$

so that we would have a decision procedure for the predicate $\vee n T(e, e, n)$ which is impossible. This example shows that the laws of classical logic (in particular, the law of the excluded middle) may lead to conclusions which are not constructively valid.

The first two chapters are based on uncritically accepting Π^0_2 statements as constructively meaningful, and the rudiment of logic that is involved may be understood in terms of the notion of constructive validity.

We have seen that the classical notion of truth differs from the notion of constructive validity. Nevertheless, it turns out that the notion of arithmetical truth may be constructively understood in a quite different way by means of the *no counterexample interpretation* (due to Herbrand 1930 in the case of first order predicate logic and extended to number theory by Kreisel 1951 who introduced the terminology). Consider again the prenex formula

$$\wedge m_1 \vee n_1 \dots \wedge m_j \vee n_j P(m_1, n_1, \dots, m_j, n_j)$$

which we denote by A. Then

$$-A \leftrightarrow \vee m_1 \wedge n_1 \dots \vee m_j \wedge n_j - P(f_1, n_1, \dots, m_j, n_j)$$

$$\leftrightarrow \vee f_1 \dots f_j \wedge n_1 \dots n_j - P(f_1, n_1, \dots, f_j(n_1, \dots, n_{j-1}), n_j)$$

where f_i ranges over all number theoretic functions of $i-1$ arguments, $i = 1, \dots, j$. Consequently,

$$A \leftrightarrow \wedge f_1 \dots f_j \vee n_1 \dots n_j P(f_1, n_1, \dots, f_j(n_1, \dots, n_{j-1}), n_j),$$

the latter formula being called the no counterexample interpretation of A. By coding $f_1, \dots, f_j$ into a single infinite sequence of natural numbers and $n_1, \dots, n_j$ into a single natural number we see that the no counterexample interpretation is of the form Π^1_1,

$$\wedge m_1 m_2 \dots m_n \dots \vee n P(m_1 m_2 \dots m_n),$$

P being a decidable property of finite sequences of natural numbers. The most typical example of a Π^1_1 property is that of being an ordinal of the constructive second number class.

We now define what it means for $m_1 m_2 \dots m_n$ to be *barred* by P. Firstly, if $P(m_1 m_2 \dots m_n)$ is true, then $m_1 m_2 \dots m_n$ is barred by P. Secondly, if $m_1 m_2 \dots m_n m$ is barred by P for all natural numbers m, then $m_1 m_2 \dots m_n$ is barred by P. Transfinite inductive definitions of this kind will be considered constructively meaningful, and we shall say that a Π^1_1 statement of the above form holds if the empty sequence $\square$ is barred by P.

As is easily seen, the equivalence

$$\wedge m_1 m_2 \dots m_n \dots \vee n P(m_1 m_2 \dots m_n) \leftrightarrow \square \text{ is barred by } P$$

holds classically. Thus, the intended meaning of a Π^1_1 statement is the same under the constructive interpretation we have adopted as under the classical interpretation. The difference lies only in the methods of proof that are accepted. This is entirely analogous to the situation for Π^0_2 statements discussed earlier.

In intuitionistic mathematics both sides of the above equivalence are considered to be meaningful, the universal quantifier on the left being understood as ranging over all *choice sequences* of natural numbers. The fact that the equivalence holds intuitionistically is the content of Brouwer's *bar theorem*. We consider the arguments given by Brouwer 1927*b* in the proof of the bar theorem rather as an intuitive analysis justifying the definition we have adopted of what it means for a Π^1_1 statement to hold.

It results from what we have just said that as long as we restrict ourselves to Π^1_1 statements we can understand Brouwer's continuum in an intuitionistically equivalent way without using choice sequences. A completely different way of getting rid of the choice sequences is to interpret quantifiers over number theoretic functions as ranging over all rules which determine such functions, that is, over all numbers e such that $\wedge m \vee n T(e, m, n)$. This is the way chosen by Markov's school of constructivists as well as by Bishop 1967 in his recent book (which the author did not get access to until after this manuscript was finished). In Bishop's case, the pathologies born out of this conception are avoided by making suitable continuity assumptions.

The no counterexample interpretation together with our analysis of Π^1_1 statements allows us to understand constructively the classical notion of truth as applied to arithmetical formulas. Since this way of understanding arithmetical truth is classically equivalent to the classical one, it is no wonder that the laws of classical logic turn out to be valid under this interpretation.

Actually, once we agree to accept Π_1^1 statements as constructively meaningful, we can understand constructively much more than the notion of arithmetical truth. Large parts of analysis, including the theory of ordinals of the second number class, Borel sets and Lebesgue measure, can be treated using these abstractions. This is shown in the last two chapters. However, stronger methods are required for a constructive analysis of the Cantor-Bendixson theorem (this was shown by Kreisel 1959*a*) as well as for a treatment of the analytic and, even more so, projective sets.

More powerful tools, adequate at least for a treatment of the Cantor-Bendixson theorem, are obtained by introducing constructive higher number classes.

CHAPTER I

RECURSIVE FUNCTIONS

1. Constructive objects

We shall try to delimit the notion of a *constructive object*. The simplest examples of such objects are obtained by combining the *letters*, *signs* or *symbols* of a finite *alphabet* into *strings* or *words*. Taking in particular the alphabet whose only letter is the stroke I, we get the *natural numbers*

$$\Box \quad \mathrm{I} \quad \mathrm{II} \quad \mathrm{III} \ldots \mathrm{IIIIIII} \ldots$$

Here, and whenever there is a risk of confusion, the presence of the *empty word* is indicated by the symbol $\Box$. The *integers*

$$\ldots -\mathrm{II} \quad -\mathrm{I} \quad \Box \quad \mathrm{I} \quad \mathrm{II} \quad \mathrm{III} \ldots$$

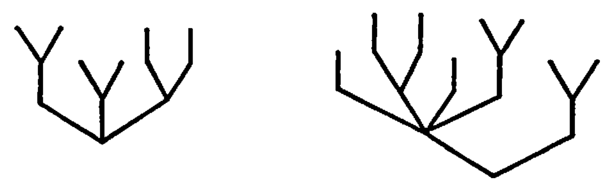
are words in the two letter alphabet I –. Adjoining the sign / we can construct the *rational numbers*

$$\mathrm{I/III} \quad -\mathrm{III/IIII} \quad -\mathrm{II/I} \ldots$$

Slightly more complicated examples are afforded by the formulas of an axiomatic theory such as first order arithmetic

$$\Lambda x -(x'=0) \quad \mathrm{V}x(x'+a=b) \rightarrow -(a=b) \ldots$$

Also, we can consider constructive objects which are not built up in a linear fashion, e.g. finite trees



matrices with integral elements