A constructive topological proof of van der Waerden's theorem

Thierry Coquand
Chalmers University

Revised version, December 1993

Introduction

Since the work of Gödel on constructible sets [11], it has been known that any arithmetical statement provable with the axiom of choice is provable without using this axiom. Intuitively, this means that, from any proof of a "concrete statement" (for instance, a purely existential statement) that uses Zorn's lemma, a more usual mathematical formulation of the axiom of choice, it is possible to extract a proof that does not use this axiom. It does not seem, however, that this method has been used in extracting constructive informations from given concrete examples of a complex proof of an arithmetical statement.

Stimulated by results of de Bruijn, van der Meiden [3], of Martin-Löf [14], and of Johnstone [13], we started in [5] a program of constructivisation of some infinitary proofs in combinatorics (Ramsey's theorem, Higman's lemma) by applying localic methodology. Although the extent to which this yields new results is not yet known, the ideas of point-free topology seem to provide intuitions into finding constructive versions of results which classically use the axiom of choice.

We give here a further illustration of this method by presenting a constructive version of the proof of van der Waerden's theorem on arithmetical progression which uses some elements of topological dynamics [8]. This last non-constructive proof is based on a simple topological lemma, which is a direct corollary of Zorn's lemma (namely, the existence of minimal dynamical system). We formulate a point-free constructive version of this lemma. As could be expected from other examples [5, 6], this localic version has a direct (constructive) proof. This may be seen as the first step, in a simple concrete case, of a point-free theory of dynamical systems. From this result, the finite version of van der Waerden's theorem can be derived along the line of the usual topological proof. Contrary to the classical argument however, there is no need to appeal to non-constructive principles in this derivation.

After a first version of this paper was written, we became aware of the work [2, 15], which, in the quite different field of functional analysis, illustrates this common idea that localic methods can be used to find sharper reformulations of basic non-constructive results, which become then constructively valid. Our work suggests that these methods, beside the purely mathematical advantage of solving problems concerned with equivariance or continuity in parameters [15], may be interesting also proof-theoretically in providing an elegant framework for extracting computational informations from given mathematical arguments.

Our treatment of the topological proof of van der Waerden's theorem is apparently different from the one presented in Girard's book on proof theory [10]. We have not tried though to compare in detail the two arguments, because the main emphasis is somewhat different. The main points of our paper are the formulation of a point-free version of a minimal property, whose ordinary version is proved via Zorn's lemma, and the observation that this point-free version has a direct inductive proof[1]. For the topological proof, we have been following the presentation of Graham, Rothschild and Spencer [12].

1 Infinitary Version of The Minimal Property

1.1 General Notations

A **block** is a finite sequence of 0s and 1s. We use the notation $A, B, C, \ldots$ for blocks, and write AB for the concatenation of two blocks A and B. If $A = b_1 \ldots b_p$ then p is called the **size** of the block A. We say that A is a **subblock** of B if, and only if, B can be written B_0AB_1, where B_0 or B_1 may be empty. If both B_0 and B_1 are empty, then A is B itself, otherwise we say that A is a **strict** subblock of B. If B_0 is empty, we say that A is an **initial** subblock of B, and if B_1 is empty, we say that A is a **final** subblock of B. We say that A **avoids** B if, and only if, B is not a subblock of A. These relations are decidable.

A **condition** is a finite sequence of integers. We write $\sigma, \tau, \ldots$ for conditions. We say that τ is a **direct extension** of σ if, and only if, τ is of the form σn, where n is an integer.

We say that a block A **satisfies** the condition $\sigma = n_1 \ldots n_p$ if, and only if, for any $i \leq p$, and any subblocks B and C of the block A, if B is of size i and C of size $\geq n_i$, then the block B is a subblock of C.

We let X be the product space $\{0,1\}^{\mathbf{Z}}$ of the discrete space $\{0,1\}$. An element of X is called a **colouring**. By Tychonoff's theorem, the space X of colourings is compact. This fact is crucial in the non-constructive topological proof of van der Waerden's theorem.

Notice that in a point-free way, the fact that X is compact is direct: this is a syntactical version of the compactness theorem for propositional logic, which has a direct inductive proof. We shall verify below that a systematic point-free presentation allows for direct inductive proofs, and avoids altogether the use of non-constructive principles.

For any positive integer N, let $WP(N)$ be the open subset of colourings $\alpha \in X$ such that there exist k, l which satisfy $\alpha(k) = \alpha(k+l) = \ldots = \alpha(k+(N-1)l)$. The infinitary version of van der Waerden's theorem can be expressed as the equality $WP(N) = X$.

We say that a finite block A is a **subblock** of $\alpha \in X$ if, and only if, there exist n, p such that A is $\alpha(n+1)\ldots\alpha(n+p)$. A colouring β is said to be a **subcolouring** of another colouring α if, and only if, any subblock of β is a subblock of α. This defines a preorder (that is, a reflexive, transitive relation) on the set X.

We define the map $T : X \to X$ by $T(\alpha)(n) = \alpha(n+1)$. This is clearly a bijective homeomorphism, and the pair (X, T) is a basic example of a **dynamical system**, that is a compact space X together with a bijective homeomorphism.

For $\alpha \in X$, let $\bar{\alpha}$ be the topological closure of the set $\{T^n(\alpha) \mid n \in \mathbf{Z}\}$. It is clear that β is a subcolouring of α if, and only if, β belongs to $\bar{\alpha}$ if, and only if, $\bar{\beta}$ is a subset of $\bar{\alpha}$.

[1] See the survey article [1] for a presentation of the theory of inductive definitions.

1.2 The Minimal Property

This proposition follows from a standard, non-constructive, use of Zorn's lemma.

Proposition (Minimal Property): For any $\alpha \in X$, there exists a subcolouring β of α which is minimal.

Proof: The set of non empty closed subsets of $\bar{\alpha}$ ordered by containment is such that any chain is dominated, by compactness. By Zorn's lemma, it contains a maximal element, which is clearly of the form $\bar{\beta}$, and β is then a minimal subcolouring of α. $\square$

Using this fact allows for an elegant derivation of van der Waerden's theorem. By the minimal property, in order to show that $WP(N)$ is a covering, it is enough to show that any minimal colouring is in $WP(N)$. We refer to [12] for the proof that indeed $WP(N)$ covers the subspace of minimal colourings.

This existence property however is highly problematic from a constructive point of view. In [12], it is shown (non constructively) that a colouring α is minimal if, and only if, there exists a fixed infinite sequence $n_1, n_2, n_3, \dots$ such that for any p, any subblock of α satisfies the condition $n_1 \dots n_p$. This property suggests the point-free version of the minimal property presented in the next section[2].

2 A point-free version of the Minimal Property

From now on, all our definitions, and proofs, will be constructive.

We describe in a point-free way the dynamical system (X, T), that is the topological space X with the group action $\mathbf{Z} \times X \to X$ defined by $(n, x) \longmapsto T^n(x)$.

Let us say that a set U of blocks is **monotone** if, and only if, it is upward closed in the subblock ordering, i.e. we have $B \in U$ whenever $A \in U$ and A is a subblock of B. We say that U is a **covering** if, and only if, there exists M such that U contains any block of size $\geq M$.

A typical example of a monotone set is the set $W(N)$ of blocks $A = b_1 \dots b_p$ such that there exist k, l such that $b_k = b_{k+l} = b_{k+2l} = \dots = b_{k+(N-1)l}$, where N is a fixed integer. This can be seen as the point-free version of the open subset $WP(N)$. The theorem of van der Waerden on arithmetical progression can be formulated as the fact that $W(N)$ is a covering.

More generally, let $W'(N, l)$ be the monotone set of blocks that have a subblock of the form $BA_1B \dots BA_{N-1}B$, where B is a block of size l and $A_1, \dots, A_{N-1}$ have all the same size. Notice that $W'(N, 1)$ coincides with $W(N)$. The topological proof of van der Waerden's theorem establishes that $W'(N, l)$ is a covering.

A set of conditions is said to be **monotone** if, and only if, it contains all direct extensions of an element whenever it contains this element. A monotone set of conditions is said to be **open** if, and only if, it contains a condition whenever it contains all its direct extensions. This terminology follows that of the monograph [14].

Given a set U of blocks, we associate to it a monotone set U^* of conditions: $\sigma \in U^*$ if, and only if, there exists M such that, if A satisfies σ and is of size $\geq M$, then $A \in U$. To say that U is a covering is thus equivalent to saying that the empty condition is in U^*.

[2] This equivalent formulation is called the Bounded Gaps condition in [12]. It follows from the minimal property by an usual compactness argument.

The following proposition is a point-free version of the minimal property.

Proposition (Minimal Property): if U is a monotone set of blocks, then U^* is an open set of conditions.

Proof: Let U be a monotone set of blocks and let $\sigma = n_1 \dots n_p$ be a condition such that all its direct extensions are in U^*. We show that $\sigma \in U^*$.

For this, we consider the proposition P_j that there exists M such that U contains any block that satisfies σ, is of size $\geq M$ and avoids at least j blocks of size $p+1$. Clearly P_0 is the property $\sigma \in U^*$ we want to show, and P_j holds for j big enough, taking M to be 0, because there are only a finite number of blocks of a given size. We show next that P_{j+1} implies P_j.

Let us assume P_{j+1}. We have then M such that U contains any block that satisfies σ, is of size $\geq M$ and avoids at least $j+1$ blocks of size $p+1$. The direct extension σM of σ is by hypothesis an element of U^*. Hence, there exists N such that U contains any block satisfying σM and of size $\geq N$. I now claim that U contains any block that satisfies σ, is of size $\geq N$ and avoids at least j blocks of size $p+1$, which establish the proposition.

Indeed, let A be a block satisfying σ, of size $\geq N$, and which avoids at least j blocks of size $p+1$.

If A satisfies σM, then $A \in U$ by choice of N.

Otherwise, A satisfies σ and not σM. In this case, there exists a subblock B of A of size $p+1$ and a subblock C of A of size $\geq M$ such that C avoids B. Since B is a subblock of A, the block B differs from all the subblock of size $p+1$ that A avoids. Hence, the subblock C of A avoids at least $j+1$ blocks of size $p+1$. By the choice of M, this implies $C \in U$ and hence $A \in U$ because U is monotone. $\square$

Let us define a **bar** as a monotone set of conditions V such that the empty condition belongs to any open set of conditions that contains V[3].

Corollary: If U is a monotone set of blocks and U^* contains a bar, then U is a covering.

This follows directly from the minimal property, and the definitions.

This result can be seen as a point-free version of the non-constructive fact that, in order to show that a monotone set U of blocks is a covering, it is enough to show that U contains a subblock of any minimal colouring.

In practice, this can be used in the same way than the usual minimal property. We are going to illustrate this point by showing how to derive some instances of van der Waerden's theorem, along the lines of its usual derivation from the classical minimal property [12].

3 Derivation of van der Waerden's theorem

Before presenting the full derivation of van der Waerden's theorem, we find it illuminating to show the derivation that $W'(3, l)$ is a covering for any fixed l, which illustrates the main points, and to give an example of how this proof can be instantiated. We recall that this means that any block that is "big enough" contains 3 identical subblocks of size l in arithmetical progression.

[3] Classically, to say that a monotone set of conditions V is a bar is equivalent to say that for any infinite sequences $n_1, n_2, n_3, \dots$ there exists p such that $n_1 \dots n_p \in V$, but intuitionistically, this equivalence is the content of Brouwer's bar theorem, discussed for instance in Dummett [7] and in our paper [4].

We shall not present a formal proof in the framework of inductive definitions, but instead try to emphasize the basic ideas behind this derivation[4].

Proposition: For any l, the monotone set of blocks $W'(3,l)$ is a covering.

Proof: By the point-free version of the minimal property, it is enough to show that $W'(3,l)^*$ is a bar.

Intuitively, we suppose given a "choice sequence" $n_1, n_2, \ldots$ of integers, and we want to show that eventually, we will find p, N such that $W'(3,l)$ contains any block of size $\geq N$ which satisfies $n_1 \ldots n_p$.

The construction proceeds then as follows. We choose N bigger than all elements of the sequence $v_1 = l,\ v_2 = 4n_{v_1}, \ldots, v_{2^l+1} = 4n_{v_{2^l}}$. Let A be any block of size N which satisfies the condition $n_1 \ldots n_{v_{2^l}}$. We are going to show that A contains three identical blocks of size l in arithmetical progression.

Remark first that if B is a subblock of A of size $\leq v_i$, for $i \leq 2^l$, and C is a subblock of A of size $\geq n_{v_i}$ then B is a subblock of C. Indeed, because A is of size N, B is a subblock of a subblock B' of A of size exactly v_i. Since A satisfies the condition $n_1 \ldots n_{v_{2^l}}$, it follows that B' is a subblock of C, and hence that B is a subblock of C.

We let B_1 be the initial subblock of A of size v_1. I claim that there exists an initial subblock B_2 of A of size $\leq v_2$ of the form $B_1C_1B_1D_1B_1'$, where C_1, D_1 have the same size, and B_1' is of the same size than B_1. Indeed, if we consider the first consecutive 4 subblock of A of size n_{v_1}, say $E = E_1E_2E_3E_4$, since A satisfies the condition $n_1 \ldots n_{v_1}$, the block E_2 contains a copy of B_1. The block E can thus be put in the form $B_1C_1B_1D_1B_1'E'$ where C_1, D_1 have the same size, and B_1' is of the same size than B_1.

Continuing in this way, and using the remark above, we build an initial subblock B_{i+1} of A of size $\leq v_{i+1}$ of the form $B_iC_iB_iD_iB_i'$, where C_i, D_i have the same size, and B_i' is of the same size than B_i.

Let B_i'' be the final subblock of B_i of size l. Notice that, for any $i < j$, we can find a subblock of A of the form $B_i''CB_i''DB_j''$, where C and D have the same size.

By the box principle, since there are exactly 2^l blocks of size l, there exists $1 \leq i < j \leq 2^l+1$ such that $B_i'' = B_j''$. The block A contains then three copies of the subblock B_i'' in arithmetical progression. $\square$

It is interesting to instantiate this argument on the simple case $l = 1$, which proves that $W(3)$ is a covering. The instantiated argument gives that any block of size $\geq N = 12\ 4^{2^{12}}$, is in $W(3)$. This can be compared with the bound 325 given by the combinatorial proof of van der Waerden [12]. This difference of bounds is analysed in more details in the treatment of [10]. The instantiated argument runs as follows.

First we remark that any block of the form $BCBDB'$, where C and D have the same size, and B, B' are of size 12, is in $W(3)$. Indeed, if we write $B = b_1 \ldots b_{12}$, either we have $b_4 = b_5 = b_6$, in which case $B \in W(3)$, or we have $b_1 = b_i$ for some i, $4 \leq i \leq 6$. If we write $B' = b_1' \ldots b_{12}'$, two of the three elements b_1, b_{2i-1}, b_{2i-1}' are equal, and in each case, we get that $BCBDB' \in W(3)$.

The next step is to write $E_1E_2E_3E_4$ for the initial subblock of A of size N where each E_i is of size $N_1 = N/4$. If the subblock E_2 contains a copy of the initial subblock of A of size 12, we can

[4]The reference [4] contains all the basic tools for expressing this proof in the framework of inductive definitions. Alternatively, one can use Brouwer's bar theorem, as presented in [7].

apply the previous remark and conclude $A \in W(3)$. Otherwise, we have found a subblock A_1 of A of size N_1 which avoids the initial subblock of A of size 12. We can continue in this way, as long as we cannot apply the previous remark, building a sequence of blocks $A_0 = A, A_1, A_2, \dots$ of size $N, N_1 = N/4, N_2 = N_1/4, \dots$ where A_{i+1} is a subblock of A_i which avoids the initial subblock B_i of A_i of size 12. By construction A_i avoids all the blocks $B_0, \dots, B_{i-1}$, and hence the blocks B_i are all distinct. This implies that, at some point, we can apply the previous remark and conclude that $A \in W(3)$.

We present now one constructive derivation of van der Waerden's theorem, that is similar to its derivation from the classical minimal property presented in the reference [12].

Theorem (van der Waerden): For all n, the monotone set of blocks $W(n)$ is a covering.

Proof: We show that $W'(n,l)^*$ is a bar for all l, by induction on n.

Intuitively, this means that, given a "choice sequence" $n_1, n_2, \dots$ of integers and an integer l, we can find p, N such that $W'(n,l)$ contains any block of size $\geq N$ which satisfies $n_1 \dots n_p$. It is clear that this holds for the value $n = 1$. We assume that this holds for a given n, and prove that it holds for the value $n+1$.

We recall that a block contains n copies of a given block B in arithmetic progressions if this block contains a subblock of the form $BA_1B \dots BA_{n-1}B$, where $A_1, \dots, A_{n-1}$ are all of the same size. Notice that if a block contains n copies of B in arithmetic progressions, then it contains n copies of any subblock of B in arithmetic progressions. We say that a block **almost** contains n copies of a given block B in arithmetic progressions if this block contains a subblock of the form $BA_1B \dots BA_{n-1}B'$, where $A_1, \dots, A_{n-1}$ are all of the same size, and B' is of the same size than B.

We notice first that a much stronger property follows if $W'(n,l)$ holds for all l. Given a choice sequence $n_1, n_2 \dots$ of integers and an integer l, we can find p, N such that any block A of size $\geq N$ which satisfies $n_1 \dots n_p$ contains n copies of *any* subblock of A of size l in arithmetic progressions. Indeed, it is enough that $p \geq l$ and that N is such that any block A of size $\geq N$ which satisfies $n_1 \dots n_p$ contains n copies of *one* subblock of size n_l, since this subblock will then contain as a subblock any subblock of A of size l.

If p, N are chosen in such a way, any block A of size $\geq 2N$ which satisfies $n_1 \dots n_p$ almost contains $n+1$ copies of any subblock of A of size l in arithmetic progressions.

Like in the previous derivation, we iterate 2^l times this operation. In this way, given a choice sequence $n_1, n_2 \dots$ of integers and an integer l, we can find p, N such that any block A of size $\geq N$ which satisfies $n_1 \dots n_p$ contains a family (B_i) of subblocks of A, for $i \leq 2^l + 1$, such that B_1 is of size l and B_{i+1} is of the form $B_iA_1 \dots B_iA_nB'_i$ where $A_1, \dots, A_n$ are all of the same size, and B'_i is of the same size than B_i.

Let B''_i be the final subblock of B_i of size l. Notice that, for any $i < j$, we can find a subblock of A of the form $B''_iA_1B''_i \dots B''_iA_nB''_j$, where $A_1, \dots, A_n$ have the same size.

By the box principle, since there are exactly 2^l blocks of size l, there exists $1 \leq i < j \leq 2^l+1$ such that $B''_i = B''_j$. The block A contains then $n+1$ copies of the subblock B''_i in arithmetical progression. □

Conclusion

We list some questions naturally suggested by this work. The first one concerns the axiom of choice. How general is the method of point-free (or constructive) topology in extracting constructive proofs from classical arguments that use the axiom of choice? In particular, is it not possible in such a way to give an alternative proof of the fact that arithmetical statements provable with the axiom of choice are provable without using the axiom? It should be noted that, in all examples analysed so far, the point-free statement is provable intuitionistically within the framework of inductive definitions. Is there a natural mathematical statement that cannot be proved in this framework? Examples given in [2, 15] suggest some candidates, but we have not yet any conclusive result.

Since we get a constructive proof of van der Waerden's theorem, it will be natural to try to describe in more details the algorithm to which it corresponds. This may be connected to the algorithm we get by using the point-free version of Nash-Williams minimal bad sequence argument [5].

It should be interesting to develop in a point-free way other notions used in the theory of dynamical systems, maybe following the framework sketched in the reference [16], and to formulate, for instance, generally in this framework the existence of minimal dynamical systems. Finally, a point-free version of measure and ergodic theory should be also of some interest, and may allow a constructive interpretation of the Ergodic theorem of Furstenberg [9].

Acknowledgement

I want to thank Marc Bezem, Chris Mulvey and one anonymous referee for helpful comments on a first version of this paper.

References

[1] P. Aczel. (1977) An introduction to inductive definitions In Hanbook of mathematical logic, J. Barwise ed., North-Holland, 739-782

[2] B. Banaschewski, C.J. Mulvey. Stone-Cech Compactification of Locales II. Journal of Pure and Applied Algebra 33, p. 107-122, 1984

[3] N. de Bruijn and W. van der Meiden. (1968) Notes on Gelfand's theory. Indigationes 31, 467-474

[4] Th. Coquand. (1991) An Analysis of Ramsey's Theorem. to appear in Information and Computation

[5] Th. Coquand. (1991) Constructive Topology and Combinatorics. In Constructivity in Computer Science, LNCS 613, 159-164

[6] Th. Coquand. (1992) An Intuitionistic Proof of Tychonoff's theorem. Journal of Symbolic Logic, 57, 28-32

[7] M. Dummett. *Elements of Intuitionism.* Oxford University Press, 1977.

[8] H. Furstenberg and B. Weiss. (1978) Topological Dynamics and Combinatorial Number Theory. J. Anal. Math. 34, 61-85

[9] H. Furstenberg. (1977) Ergodic Behavior of Diagonal Measures and a Theorem of Szemeredi on Arithemtic Progressions. J. Anal. Math. 31, 204-256

[10] J.Y. Girard. *Proof-Theory and Logical Complexity.* Bibliopolis, 1987.

[11] K. Gödel. The consistency of the axiom of choice and of the generalized continuum hypothesis. Proceedings of the National Academy of Sciences, vol. 24, p. 556-557,1938

[12] R. Graham, B. Rothschild, and J. Spencer. *Ramsey Theory.* John Wiley & Sons, New York, 1980

[13] P. Johnstone. *Stone Spaces.* Cambridge Studies in Advanced Mathematics, 1981

[14] P. Martin-Löf. *Notes on Constructive Mathematics* Almqvist & Wiskell, Stockholm

[15] Ch. Mulvey, J.W. Pelletier. A Globalization of the Hahn-Banach Theorem Advances in Mathematics, Vol. 89, p. 1-59, September 1991

[16] G. Sambin. Intuitionistic Formal Spaces - a first communication. In Mathematical Logic and its applications, D. Skordev ed., Plenum, 187-204, 1987