Report on the master thesis of Vojtěch Štěpančík
Formalization of Homotopy Pushouts in Homotopy Type Theory

Egbert Rijke

September 5, 2024

The thesis reports on the efforts towards the formalization of a new theorem due to David Wärn characterizing path spaces of homotopy pushouts, which is widely celebrated in the community of homotopy type theorists. Wärn's theorem asserts that the identity type of a homotopy pushout is the ∞-groupoid of "words" built from the gluing paths and their inverses, which is defined precisely as a sequential colimit of his "zigzag construction". The theorem could be seen as a generalization of the Van Kampen theorem, and has been used to settle questions that have been open since the conception of Homotopy Type Theory, such as the question of whether the suspension of a set is always a 1-groupoid. It should be noted here that even if the most common interpretation of types in homotopy type theory is that types are spaces, a major result of Shulman tells us that homotopy type theory is the internal language of arbitrary ∞-toposes. All results proven in homotopy type theory are therefore not only true for spaces, but indeed for arbitrary ∞-toposes. While it was known that suspensions of homotopy discrete spaces are 1-truncated, the generalization of this claim to arbitrary ∞-toposes was not known. Many developments in homotopy type theory lead to genuinely new results in this way, and are obtained via new methods that weren't available 15 years ago.

Even though David Wärn's theorem is widely considered to be a major step forward in the study of homotopy pushouts, we are unaware of any other attempts towards formalizing this result in a computer proof assistant. As far as we know, Štěpančík's project is the first such attempt. For this project, Štěpančík has developed extensive infrastructure about synthetic homotopy theory in the agda-unimath library. The agda-unimath library is a library of formalized univalent mathematics, containing a wide variety of mathematical subjects including number theory, algebra, combinatorics, (higher) category theory, and indeed synthetic homotopy theory. Although the agda-unimath project is fairly new, it stands as one of the largest libraries of univalent mathematics alongside Voevodsky's original Unimath library in Coq, and it is unique among libraries for homotopy type theory in that its scope encompasses such a wide range of mathematical subjects. Synthetic homotopy theory is an approach towards homotopy theory where types are interpreted as spaces, and the methods to study them are heavily inspired by algebraic topology. The thesis reports on three specific topics of formalization that fit within synthetic homotopy theory: homotopy pushouts, other homotopy colimits such as coequalizers and sequential colimits, and finaly Wärn's zigzag construction for the characterization of path spaces of homotopy pushouts.

Chapter 1 of the thesis outlines the basics of Homotopy Type Theory, and is clearly focused on those results that will be used later in the work. This chapter sets up notation, definitions, lemmas and theorems, which were already present in the agda-unimath library before the thesis project started. The outline is clear, concise, and well written, and touches on all the themes that are important in homotopy type theory and relevant for the pursuit of formalizing Wärn's theorem.

Minor comment: On page 6 in the second displayed equation, the codomain of the second operation should be $x = z$, not $x = y$.

In chapter 2 the main body of the work begins with the study of homotopy pushouts. The chapter begins with a pointed remark that while the agda-unimath library doesn't feature an implementation of homotopy pushouts as data-types and therefore lacks some computational properties, the programmer is in fact encouraged to write their code in a more modular way, which bypasses the perceived lack of computational properties almost entirely. This chapter on homotopy pushouts describes how they are specified, and then dives into the descent property, the flattening lemma, and identity systems on homotopy pushouts. While the library already contained some work on these topics, Štěpančík has developed much of the infrastructure and necessary lemmas surrounding them, such as descent data for families function types over pushouts, and descent data for families equivalence types over pushouts. The first milestone of his masters thesis project was the flattening lemma, which asserts that total spaces of families over pushouts are themselves pushouts. This technaically difficult result was formalized in an exceptionally beautiful way by Štěpančík. He demonstrated that when you write code with conceptual clarity, you can obtain technically difficult results in a clean and maintainable way. The level of clarity in his formalization of the flattening lemma was unmatched in prior formalizations of the same results in other libraries. Chapter 2 ends with a note on identity systems over pushouts, which describes in abstract terms what has to be done in order to formalize that the zigzag construction indeed charactierizes the identity type of pushouts.

Minor comment: In the displayed equation on page 23, the variable x should be x_0.

Chapter 3 is mainly about sequential colimits, but in order to develop the infrastructure for sequential colimits it is convenient to develop infrastructure for homotopy coequalizers first. In this chapter, Štěpančík describers his formalization of descent and the flattening lemma for both coequalizers and sequential colimits, extending the previous developments for pushouts. Again, it should be noted that some of these results, in particular the generalized flattening lemma for sequential colimits, are difficult to formalize in a proof assistant. Even though the result looks very plausible, when one tries to formalize them it becomes necessary to solve some difficult coherence problems, as was described in the work of Sojakova et al., which was referred to in the thesis. However, the zigzag construction involves sequential colimits, and therefore there was a necessity to formalize some basic results about them.

In Chapter 4 we finally get to the main topic of the thesis: the zigzag construction of Wärn. Štěpančík describes the formalization of the zigzag construction itself, and the partial proof of its correctness. Unfortunately, the project ran out of time before the final coherence problem of the proof of correctness of the zigzag construction could be completed. The thesis describes clearly how far the formalization got, and specifies precisely what coherence problem was left unsolved at the time of handing in the thesis. We should note here that this coherence problem was also not addressed in David Wärn's original proof. He stated instead that the correctness of his construction was "straightforward", which it was not. There has been private communication between David Wärn and Vojtěch Štěpančík, after which it became clear that the problem was in fact never addressed, and therefore one could argue that the proof of Wärn's theorem was incomplete. It is my understanding that Štěpančík now has a "pen-and-paper" proof of the missing coherence, and a plan for finishing the formalization.

Finally, I also wish remark on the exceptional quality of the content that Štěpančík has contributed to the agda-unimath library. While it is to be expected that students deliver "code that works", Štěpančík went beyond that base line and delivered sensibly refactored code with clear informal expositions of each of the constructions and proofs. Furthermore, he took the initiative for expository pages about formalization of results from the literature, in which we can precisely outline how each claim in a paper corresponds to a formalization entry, making the whole project more accessible and transparent.

In conclusion, it only remains to say that this thesis work is of truly exceptional quality, and that Štěpančík is fully deserving of the degree of Masters of Mathematics.