Chapter 11 Explanation: Addendum: Inductive Covers and Compactness

Alexander Grabanski - ajg137@case.edu

July 8, 2017

Compactness is generally pretty useful in elementary real analysis – for example, a continuous function from a compact subset of a metric space to the reals is guaranteed to attain a maximum within that set under the extreme value theorem, which is a useful property in applications to continuous optimization problems. Given the classical closure properties of the collection of compact sets (closed under images of continuous maps and infinite products by Tychonoff's Theorem), and the wide variety spaces we can build from the interval, it would be practical to define compactness in HoTT and prove that the closed interval is compact.

The book considers three different classical notions of compactness. The first is metric compactness, which says that the (metric) space is Cauchy-complete and "totally-bounded", which means that the space is coverable by a finite number of balls of fixed size for every possible size. The second is Bolzano-Weierstrass compactness, which says that every sequence has a convergent subsequence. The last is Heine-Borel compactness, which states that every open cover of the set has a finite subcover, which is particularly suitable when we only have the guarantee of a topology (but not necessarily a metric).

0.1 The Bolzano-Weierstrass property destroys realizability

Bolzano-Weierstrass compactness, however, has a difficulty that should be immediately apparent from the interpretation of constructive mathematics through realizability models. Suppose that we formalize Turing machines + input pairs in HoTT under some type

$$Turing : \mathcal{U}$$

with an operation

$$hasHaltedWithin : Turing \to \mathbb{N} \to \mathbf{2}$$

which determines if a given Turing machine and input has halted within a certain number of steps, returning $1_{\mathbf{2}}$ if so, and $0_{\mathbf{2}}$ if not. Then, for any element $t : Turing$, consider the sequence

$$x_n :\equiv rec_2(\mathbb{R}_C, hasHaltedWithin(t, n), 0_{\mathbb{R}}, 1_{\mathbb{R}})$$

Notice then that if we use $Bolzano - Weierstrass - Property$ on x to obtain a convergent subsequence s, we could then immediately ask for a value in s within $1/2$ of the limit, which would immediately tell us (since x takes on discrete values, and once a Turing machine has halted, it stays halted) whether t ever halts. Consequently, we could construct a function

$$HaltingOracle : Turing \to \mathbf{2}$$

which tells us whether or not a given (machine, input) pair halts, but that means that when we use realizability to extract an algorithm from $HaltingOracle$, we have solved the halting problem, which is an impossibility. So no realizability model is a model of any version of HoTT where $[0, 1]$ is Bolzano-Weierstrass compact.

Since we'd really like constructive HoTT to have a computational interpretation, leaving non-computable functions to Classical HoTT, the Bolzano-Weierstrass property is no good in this setting.

1 Metric Compactness

1.1 Basic Definitions

Recall that a metric space (M, d) is some set M together with an metric $d : M \times M \to \mathbb{R}$ which is non-negative, symmetric, satisfies the triangle inequality, and is such that $d(x, y) = 0$ if and only if $x = y$. Then, in analogy with what we did for real numbers, we can define Cauchy approximations in M by:

$$CauchyApx :\equiv \sum_{x:\mathbb{Q}_+ \to M} isCauchyApx(x)$$

$$isCauchyApx(x) :\equiv \prod_{\epsilon,\delta:\mathbb{Q}_+} d(x_\epsilon, x_\delta) < \epsilon + \delta$$

As usual, we will refer to the first projection of $CauchyApx$ as being a "Cauchy approximation". For both of the above, (M, d) is taken to be an implicit first argument.

We can also define a *limit* $L : M$ of a Cauchy approximation x by essentially demanding that x_ϵ is always at or within a distance of ϵ of the limit

$$isLimit(L, x) :\equiv \prod_{\epsilon,\delta:\mathbb{Q}_+} d(x_\epsilon, L) < \epsilon + \delta$$

We could have phrased this as

$$isLimit(L, x) :\equiv \prod_{\epsilon:\mathbb{Q}_+} d(x_\epsilon, L) \leq \epsilon$$

but the first formulation more clearly shows the connection with $isCauchyApx(x)$, because it essentially states that L is "as good" as any δ-close approximation x_δ.

Cauchy-completeness then becomes:

$$isCauchyComplete((M, d) :\equiv \prod_{x:CauchyApx((M,d))} \sum_{L:M} isLimit(L, pr_1(x)))$$

Note that the existence of the limit is not propositionally truncated. This will be useful in the later proof of an analogue of the Extreme Value Theorem, because the proof can be read as a recipe to construct a certain sequence, and the Cauchy-completeness will yield a "recipe" to extract the maximum out of that sequence.

With that out of the way, note that we could easily express "metric compactness" given some kind of notion of a cover of a space by ϵ-sized balls. We'll call such a cover an ϵ-net, defined by:

$$\epsilon - net((M, d)) :\equiv \prod_{\epsilon:\mathbb{Q}_+} \sum_{L:List(M)} \prod_{y:M} Exists(L, x \mapsto (d(x, y) < \epsilon))$$

Where $Exists : \prod_{A:\mathcal{U}} List(A) \to (A \to Prop) \to Prop$ is defined by induction on $List$s as:

$$Exists(Nil, p) :\equiv \mathbf{0}$$

$$Exists(Cons(a, L), p) :\equiv Exists(L, p) \vee p(a)$$

It should then be clear that an element of $Exists(L, p)$ implies that there merely exists a $k : \mathbb{N}$ such that the kth element of L satisfies p, since we could prove the equivalence of the two through a simple $List$ induction (once we have enough of the theory of $List$ and $\mathbb{N}$ built up!)

This definition eliminates the notational abuse used in the book by following through on Remark 11.5.4. From here on out, though, we'll refer to list elements in particular positions using subscripts, and ellipses for simple iterations.

Then M is said to be totally bounded if

$$TotallyBounded((M, d)) :\equiv \prod_{\epsilon:\mathbb{Q}_+} \epsilon - net((M, d))$$

So we may define

$$MetricallyCompact((M, d)) :\equiv isCauchyComplete((M, d)) \times TotallyBounded((M, d))$$

1.2 All closed intervals are metrically compact

We can show that $[a, b]$ is metrically compact with the induced metric from $\mathbb{R}_C$. For Cauchy-completeness, note that back when we did the proof that Lipschitz functions on $\mathbb{Q}$ lift to Lipschitz functions on reals, we defined the lifting on limit points by:

$$\bar{f}(lim(x)) :\equiv lim(\lambda\epsilon.\bar{f}(x_{\epsilon/L}))$$

We also used a similar trick for the proof that a two-argument function that is a contraction map in each parameter individually lifts from rationals to the reals. Essentially, we defined the liftings so that "taking the limit" commutes with the lifted function. So, we can define

$$clamp(a, b, z) :\equiv max(a, min(b, z))$$

which takes z to the interval $[a, b]$, and since $clamp(a, b)$ is a composite of the kind of functions described above, each with Lipschitz constant 1, if x is a Cauchy approximation in $[a, b]$,

$$clamp(a, b)(lim(x)) = lim(\lambda\epsilon.clamp(a, b)(x_\epsilon)$$

but each x_ϵ is in $[a, b]$, so the clamping goes away, and this is

$$= lim(x)$$

so limits of Cauchy approximations stay in $[a, b]$.

1.3 Analogue of the Extreme Value Theorem

Unfortunately, we won't be able to prove the Extreme Value Theorem in its full generality, because we'll only get as far as showing that the image of uniformly continuous maps on totally bounded spaces have a supremum, and the classical way of moving from existence of a supremum to the existence of a maximum relies on the Bolzano-Weierstrass property, which we know is not valid. Nevertheless, we'll recover a suitable analogue.

A function $f : M \to \mathbb{R}$ is *uniformly continuous* if

$$uniformContin(f) :\equiv \prod_{\epsilon:\mathbb{Q}_+} \sum_{\delta:\mathbb{Q}_+} \prod_{x,y:M} d(x, y) < \delta \to |f(x) - f(y)| < \epsilon$$

Given a uniformly continuous function f, define its modulus of uniform continuity $modUnifContin(f)$ $\mathbb{Q}_+ \to \mathbb{Q}_+$ as the function taking ϵ to δ extracted from the witness to $uniformContin(f)$.

Suppose we have such an f, and (M, d) is totally bounded. Then, this means that (M, d) has an $\epsilon - net((M, d))$ for every ϵ, which we can view as increasingly-good finite ball-based approximations of M within ϵ-neighborhoods. First, note that if we have any ϵ-net with no points in it, M must be empty, and so it wouldn't make any sense to formulate the theorem to begin with. So what we could do is to consider a Cauchy approximation to the supremum of f constructed by taking the max over all points in an ϵ-net, and then take ϵ to zero. However, note that this awkwardly backwards, because the statement of uniform continuity lets us meet a standard of closeness the codomain by prescribing a standard in the domain. So instead, let $h :\equiv modUnifContin(f)$, and consider $h(\epsilon) - net$s.

Let

$$x_\epsilon :\equiv reduce(max, map(f, N_{h(\epsilon)}))$$

for $N_{h(\epsilon)}$ the list of points in the $h(\epsilon) - net$ from the witness to the total boundedness of M. This is a Cauchy approximation, since if we compare x_ϵ and x_δ, we can see that every point z in the list $N_{h(\delta)}$ is $h(\epsilon)$-close to *some* point y in the list $N_{h(\epsilon)}$, since the latter is a $h(\epsilon)$-net. So $d(y, z) < h(\epsilon)$, so by uniform continuity of f, $|f(y) - f(z)| < \epsilon$. But then, $f(z) < f(y) + \epsilon$, and since $f(y) \leq x_\epsilon$ by the very definition of $\leq$, we obtain $f(z) < x_\epsilon + \epsilon$, and so $x_\delta < x_\epsilon + \epsilon$. We could also have equally well reversed this argument, obtaining $x_\epsilon < x_\delta + \delta$, and so, putting the two together, we obtain $x_\delta - x_\epsilon < \epsilon(+\delta)$ and $x_\epsilon - x_\delta < \delta(+\epsilon)$, so $x_\epsilon \sim_{\epsilon+\delta} y_\delta$.

Now that we have a Cauchy approximation of maxima, passing to the limit should yield the supremum we want. Set $m \equiv lim(x)$. Then, we need to show that for any $x : M$,

$f(x) \leq m$. But since there's a function $\neg(x < y) \rightarrow (y \leq x)$ as part of $\mathbb{R}$ satisfying the ordered field axioms, we can suppose $m < f(x)$ and attempt to exhibit an element of $\mathbf{0}$. By the Archimedian principle on $\mathbb{R}$, there's some rational q such that $m < q < f(x)$, and so by the Archimedian principle again, we can obtain a rational ϵ between zero and $q - m$ such that $m + \epsilon < f(x)$. From there, note that the $h(\epsilon)$-net from x_ϵ contains a point y such that $f(y)$ is strictly ϵ-close to $f(x)$. Here, the book gets somewhat sloppy by assuming that $f(y) \leq m$ is obvious. We actually need to show that $max(f(y), m) = m$ to do this (boiling down to definitions), which requires another application of the Lipschitz-esque limit-switching on max to deal with m, and another use of the uniform continuity of f. Once that's done, we see that there's a contradiction, since $m < f(x) - \epsilon < f(y) \leq m$.

The above proof shows that m is an upper bound, but we also need to show that it's ϵ-close to some $f(u)$ for some $u : M$ to show that it's the supremum. Here, the book contains a repeated typo – it referred to "$f(x_{\epsilon/2})$", which is complete nonsense. I have no idea how the proof in the book is supposed to proceed in light of this error. Nevertheless, like the book, note that $|m - x_{\epsilon/2}| \leq (1/2)\epsilon$ since m is the limit of x.

I tracked down a page (p94) out of the Google preview of Bishop's *Constructive Analysis* which performs this proof, and it seems that the real way to do it is to note that any $h(\epsilon/2)$-net on M yields an $(\epsilon/2)$-net in $Im(f)$, which w.l.o.g. we could augment with information about the points in the domain that the points in $\mathbb{R}$ are images of. Then, note that $x_{\epsilon/2}$ is in $Im(f)$ (since finite subsets of reals are closed under max), and so there is some u such that $|f(u) - x_{\epsilon/2}| < (\epsilon/2)$ using the explicitly-constructed $(\epsilon/2)$-net. Consequently, by the triangle equality, we may obtain $|m - f(u)| < \epsilon$, as desired.

1.3.1 Discussion

While the above theorem is useful to some degree, it would be more widely applicable if we could at least weaken the demand for uniform continuity to a demand for continuity. One possible way to do that would be to demonstrate that any continuous function with a (metrically) compact co-domain is uniformly continuous. However, the simplest proof of this fact relies on the classical equivalence of metric compactness with Heine-Borel compactness.

2 Heine-Borel Compactness and Inductive Covers

For our purposes, it would be nice enough to show that closed intervals are Heine-Borel compact, since we could hope for this to yield a better analogue of the Extreme Value Theorem for functions between compact subsets of the reals. Define a *family of basic intervals* as a pair (I, F) such that $F : I \rightarrow \mathbb{Q} \times \mathbb{Q}$. Here, I is an index set, and F returns the endpoints of the (open) interval the family defines. To ease their construction, finite examples of such families are presented as lists $[(a_0, b_0), ...(a_n, b_n)]$, and are taken to be indexed by $Fin(n)$. Then, a *pointwise cover* of $[a, b]$ is a family of basic intervals (F, I) such that:

$$\forall(x : [a, b]). \quad \exists(i : I).pr_1(F(i)) < x < pr_2(F(i))$$

– that is, every point is merely contained in some interval in the family.

The notion of a pointwise cover is good enough classically to express Heine-Borel compactness – that's Theorem 11.5.11. However, constructively, we need something more structured. Instead of talking about individual points, it would be good to define an inductive type specifying rules for when something covers something else, so we would not only have a cover, but reasons for why it covers.

So, we start defining an *inductive cover* as a relation between an open interval represented by its endpoints, and a family of basic intervals:

$$\triangleleft : (\mathbb{Q} \times \mathbb{Q}) \to (\sum_{I:\mathcal{U}} (I \to \mathbb{Q} \times \mathbb{Q})) \to Prop$$

It's clear that any interval in a family (I, F) should be covered by the family, so we have *reflexivity*:

$$F(i) \triangleleft (I, F)$$

It should also be the case that if a family (I, F) covers every interval in the family (J, G), and (J, G) covers (q, r), (I, F) should cover (q, r) by *transitivity*:

$$(q, r) \triangleleft (J, G) \wedge \forall(j : J).G(j) \triangleleft (I, F) \to (q, r) \triangleleft (I, F)$$

Analogously, we also need to deal with intervals that are subsets of others (which we could view as a "singleton covering"), which leads to *monotonicity*:

$$(q, r) \subseteq (s, t) \wedge (s, t) \triangleleft (I, F) \to (q, r) \triangleleft (I, F)$$

While reflexivity already is sufficient to handle limited cases of unions of intervals, we should also do something about their intersections. We should be able to relativize a cover (I, F) to some interval (s, t), and that is exactly the content of *localization*:

$$(q, r) \triangleleft (I, F) \to (q, r) \cap (s, t) \triangleleft (I, \lambda i.(F(i) \cap (s, t)))$$

All of these would be very general if we replaced open intervals with open sets, but we can leverage properties of open intervals in $\mathbb{R}$ to obtain two more properties. The first one says that overlapping intervals cover one big interval:

$$q < s < t < r \to (q, r) \triangleleft [(q, t), (r, s)]$$

and the second says that the infinite collection of open intervals contained in a fixed open interval covers that interval:

$$(q, r) \triangleleft (\{(s, t) : \mathbb{Q} \times \mathbb{Q} | q < s < t < r\}, \lambda u.u)$$

From these, we can prove that $[a, b]$ is Heine-Borel compact fairly quickly. Define "$[a, b]$ is covered by (I, F)" by the mere existence of an ϵ-enlarged open interval covered by (I, F). Then, the book proves a lemma (11.5.14) which states that for rationals $q < s < t < r$ where $(q, r) \triangleleft (I, F)$, (I, F) has a finite subcover covering (s, t). To use this, take (q, r) to be the ϵ-enlarged open interval, and (s, t) to be an $\epsilon/2$-enlarged open interval containing $[a, b]$. The finite cover of (s, t) from the lemma then immediately yields a finite cover of $[a, b]$ by definition.

The proof of the lemma is by induction on the structure of $S : (q, r) \triangleleft (I, F)$.

- If S was obtained by reflexivity , we know $(s,t) \subseteq (q,r)$, and so by monotonicity, (s,t) is covered by $[(q,r)]$, a finite subfamily of (I,F).
- If S was obtained by transitivity, applying the IH twice yields a finite subfamily covering (s,t) and finite subfamilies of the larger cover covering every interval in the other finite subfamily, which we can just concatenate together.
- If S was obtained by monotonicity, we can just apply the IH to the larger of the two intervals in the constructor, since it still contains (s,t).
- If S was obtained by localization of (J,G) to (a,b), we may simply take a finite subfamily of (J,G) localized to (a,b).
- The case of two overlapping intervals is direct, because the family is already finite, so monotonicity on $(s,t) \subseteq (q,r)$ suffices.
- For ”coverage from within”, (s,t) is an interval within the family, and so reflexivity suffices.

Theorem 11.5.6 shows that under classical assumptions, the notions of ”pointwise cover” and ”inductive cover” coincide, but that inductive covers are always pointwise covers (a fact with a rather boring proof).

Now, using the new-fangled notion of an inductive cover, it's possible to show that Heine-Borel compactness of a subset A of $\mathbb{R}$ with respect to inductive covers implies that any continuous function $f : A \to \mathbb{R}$ is merely uniformly continuous. Recall that continuity of f means:

$$\forall(u : \mathbb{R})\forall(\epsilon : \mathbb{Q}_+).\exists(\delta : \mathbb{Q}_+).(\forall v : \mathbb{R}).(u \sim_\delta v) \to (f(u) \sim_\epsilon f(v))$$

Then, using this, for any $x : A$, we can consider (under one big propositional truncation) x_ϵ to be the interval $(x-\delta, x+\delta)$ with δ, ϵ as in the definition of continuity. Then, if we could construct an infinite family of intervals, one x_ϵ, for every $x : A$, we could extract a finite subfamily, and take a minimum over all the δ to obtain a modulus of uniform continuity. However, at the time of writing I'm not sure if the constructors for inductive covers allow for this in a constructive setting. (In a classical setting, it's obvious that such an infinite family would create a pointwise cover, and hence an inductive one.) The best bet looks like the ”overlapping intervals” constructor somehow combined with transitivity, but I'll have to explore that one a bit more.