Caristi's Theorem and Approximating Π^1_1 Comprehension

Henry Towsner

University of Pennsylvania

May 24, 2019

Joint work with David Fernández-Duque, Paul Shafer, and Keita Yokoyama

Definition

When $(\mathcal{X}, d)$ is a metric space, $T : \mathcal{X} \to \mathcal{X}$ is a *contraction* if there is a $k < 1$ so that, for all x, y, $d(T(x), T(y)) \leq k\, d(x, y)$ for all $x, y \in \mathcal{X}$.

Theorem (Banach)

When $(\mathcal{X}, d)$ is a complete metric space and $T : \mathcal{X} \to \mathcal{X}$ is a contraction, there is an $x_* \in \mathcal{X}$ so that $T(x_*) = x_*$.

Definition

When $(\mathcal{X}, d)$ is a metric space, $T : \mathcal{X} \to \mathcal{X}$ is a *contraction* if there is a $k < 1$ so that, for all x, y, $d(T(x), T(y)) \leq k\, d(x, y)$ for all $x, y \in \mathcal{X}$.

Theorem (Banach)

When $(\mathcal{X}, d)$ is a complete metric space and $T : \mathcal{X} \to \mathcal{X}$ is a contraction, there is an $x_* \in \mathcal{X}$ so that $T(x_*) = x_*$.

Theorem (Caristi)

When $(\mathcal{X}, d)$ is a complete metric space, $T : \mathcal{X} \to \mathcal{X}$, and $V : \mathcal{X} \to \mathbb{R}_{\geq 0}$ is a lower semicontinuous function such that, for all x, y, $d(x, T(x)) \leq V(x) - V(T(x))$, then there is an $x_* \in \mathcal{X}$ so that $T(x_*) = x_*$.

Theorem (Caristi)

When $(\mathcal{X}, d)$ *is a complete metric space,* $T : \mathcal{X} \to \mathcal{X}$*, and* $V : \mathcal{X} \to \mathbb{R}_{\geq 0}$ *is a lower semicontinuous function such that, for all* x, y, $d(x, T(x)) \leq V(x) - V(T(x))$*, then there is an* $x_* \in \mathcal{X}$ *so that* $T(x_*) = x_*$.

The function V is a potential function: $T(x)$ must have lower potential than x, and we can only move as far as the loss of potential allows.

Theorem (Caristi)

When $(\mathcal{X}, d)$ *is a complete metric space,* $T : \mathcal{X} \to \mathcal{X}$*, and* $V : \mathcal{X} \to \mathbb{R}_{\geq 0}$ *is a lower semicontinuous function such that, for all* x, y*,* $d(x, T(x)) \leq V(x) - V(T(x))$*, then there is an* $x_* \in \mathcal{X}$ *so that* $T(x_*) = x_*$*.*

The function V is a potential function: $T(x)$ must have lower potential than x, and we can only move as far as the loss of potential allows.

To obtain the Banach fixed point theorem, if T is a contraction, define $V(x) = \frac{1}{1-k} d(x, T(x))$. Then for any x,

$$\begin{aligned} V(x) - V(T(x)) &= \frac{1}{1-k}(d(x, T(x)) - d(T(x), T(T(x)))) \\ &\geq \frac{1}{1-k}(d(x, T(x)) - k\, d(x, T(x))) \\ &\geq d(x, T(x)), \end{aligned}$$

so V is a potential and Caristi's fixed point theorem applies.

Theorem (Caristi)

When $(\mathcal{X}, d)$ *is a complete metric space,* $T : \mathcal{X} \to \mathcal{X}$*, and* $V : \mathcal{X} \to \mathbb{R}_{\geq 0}$ *is a lower semicontinuous function such that, for all* x, y, $d(x, T(x)) \leq V(x) - V(T(x))$*, then there is an* $x_* \in \mathcal{X}$ *so that* $T(x_*) = x_*$*.*

Theorem (Caristi)

When $(\mathcal{X}, d)$ is a complete metric space, $T : \mathcal{X} \to \mathcal{X}$, and $V : \mathcal{X} \to \mathbb{R}_{\geq 0}$ is a lower semicontinuous function such that, for all x, y, $d(x, T(x)) \leq V(x) - V(T(x))$, then there is an $x_* \in \mathcal{X}$ so that $T(x_*) = x_*$.

Caristi's proof is by transfinite induction: start with any x_0, define $x_{\alpha+1} = f(x_\alpha)$, and $x_\lambda = \lim_{\alpha<\lambda} x_\alpha$. (Since $\sum_{\alpha<\lambda} d(x_\alpha, x_{\alpha+1}) \leq V(x_0)$, this limit always exists.)

Lower semicontinuity of V ensures that the potential doesn't increase at limit steps. Since $V(x_\alpha)$ is decreasing, there must be some $\alpha < \omega_1$ so that $V(x_\alpha) = V(x_{\alpha+1})$, so this x_α is a fixed point.

Reverse mathematics is concerned with classifying the strength of theorems which can be expressed in the language of second order arithmetic (that is, with a sort for natural numbers and a sort for sets of natural numbers).

The "Big 5" theories of Reverse Mathematics, in increasing order of strength:

- RCA_0, recursive comprehension (roughly equivalent to primitive recursive mathematics),
- WKL_0, a weak compactness principle,
- ACA_0, arithmetic comprehension (the second order analog of Peano arithmetic),
- ATR_0, arithmetic transfinite recursion (related to predicative mathematics),
- Π^1_1-CA_0, Π^1_1 comprehension.

ATR_0 adds the axiom "every well-ordering has a jump hierarchy". (That is, whenever $\prec$ is a linear ordering, either it has an infinite descending sequence or it has a jump hierarchy.)

ATR_0 is usually the weakest theory in which we can work with transfinite inductions.

ATR_0 adds the axiom "every well-ordering has a jump hierarchy". (That is, whenever $\prec$ is a linear ordering, either it has an infinite descending sequence or it has a jump hierarchy.)

ATR_0 is usually the weakest theory in which we can work with transfinite inductions.

Π^1_1-CA_0 adds to ATR_0 the axiom saying that every Π^1_1 formula defines a set. In particular, Π^1_1-CA_0 is the theory needed to take an arbitrary linear ordering and guarantee that it's well-founded part exists as a set.

There's an obstacle to trying to formalize Caristi's Theorem in second order arithmetic.

Separable metric spaces can be encoded in a fairly natural way: natural numbers can encode a dense subset, and sets of natural numbers can encode Cauchy sequences converging to points.

Continuous and semicontinous functions can also be encoded by sets of natural numbers.

There's an obstacle to trying to formalize Caristi's Theorem in second order arithmetic.

Separable metric spaces can be encoded in a fairly natural way: natural numbers can encode a dense subset, and sets of natural numbers can encode Cauchy sequences converging to points.

Continuous and semicontinous functions can also be encoded by sets of natural numbers.

But arbitrary functions would be third-order objects, which don't belong in second order arithmetic.

Caristi's Theorem can be derived from Ekeland's variation principle:

Theorem (Ekeland)

When $(\mathcal{X}, d)$ is a complete metric space and $V : \mathcal{X} \to \mathbb{R}_{\geq 0}$ is a lower semicontinuous function, then V has a critical point x_* so that, for any y,

$$d(x_*, y) \leq V(x_*) - V(y)$$

implies $y = x_*$.

Caristi's Theorem can be derived from Ekeland's variation principle:

Theorem (Ekeland)

When $(\mathcal{X}, d)$ is a complete metric space and $V : \mathcal{X} \to \mathbb{R}_{\geq 0}$ is a lower semicontinuous function, then V has a critical point x_* so that, for any y,

$$d(x_*, y) \leq V(x_*) - V(y)$$

implies $y = x_*$.

The same transfinite induction: let x_0 be arbitrary, if x_α is not a critical point then there is an $x_{\alpha+1}$ with $d(x_\alpha, x_{\alpha+1}) \leq V(x_\alpha) - V(x_{\alpha+1})$.

Ekeland's variation principle does not involve arbitrary functions: it only requires lower semicontinuous functions, which we can encode in second order arithmetic.

Ekeland's variation principle does not involve arbitrary functions: it only requires lower semicontinuous functions, which we can encode in second order arithmetic.

Theorem (FD-S-Y)

The following are equivalent:

- Π^1_1-CA_0,
- *Ekeland's variation principle (*EVP*),*
- *Ekeland's variation principle for Baire space,*
- *Ekeland's variation principle for the closed unit ball of* $C([0,1])$*.*

Ekeland's variation principle does not involve arbitrary functions: it only requires lower semicontinuous functions, which we can encode in second order arithmetic.

Theorem (FD-S-Y)

The following are equivalent:

- Π^1_1-CA_0,
- *Ekeland's variation principle (*EVP*),*
- *Ekeland's variation principle for Baire space,*
- *Ekeland's variation principle for the closed unit ball of $C([0,1])$.*

FD-S-Y consider some variations (where the critical point needs to be chosen near a particular point, where $\mathcal{X}$ is compact, where V is continuous). Some combinations are equivalent to ACA_0 or WKL_0.

Why does EVP imply Π^1_1-CA$_0$?

Fix a collection of of orderings $\prec_n$ on $\mathbb{N}$. For any real $x \in \mathbb{N}^\mathbb{N}$, define

$V(x) = \sum\{2^{-n} \mid x_n$ is not an infinite descending sequence in $\prec_n\}$.

If $\prec_n$ is not well-founded but x_n is not an infinite descending sequence in $\prec_n$ then we can take y which is identical to x except for replacing the n-th column with an infinite descending sequence and verify that

$$d(x, y) \leq V(x) - V(y).$$

Since the proof of Caristi's Theorem from EVP is "easy", we can think of the bound on EVP as giving an upper bound on Caristi's Theorem: in Π^1_1-CA_0, we can prove Caristi's Theorem for any function we can get a handle on.

(And if we moved to a third order system extending Π^1_1-CA_0, we could prove Caristi's Theorem there without worrying about encoding.)

Since the proof of Caristi's Theorem from EVP is "easy", we can think of the bound on EVP as giving an upper bound on Caristi's Theorem: in Π^1_1-CA_0, we can prove Caristi's Theorem for any function we can get a handle on.

(And if we moved to a third order system extending Π^1_1-CA_0, we could prove Caristi's Theorem there without worrying about encoding.)

Question

Is EVP *necessary to prove Caristi's Theorem?*

If we want to talk about Caristi's Theorem in reverse math, we have to restrict it to "nice" functions that we can encode in second order arithmetic.

Theorem (FD-S-Y)

- CFP *for continuous functions implies* ACA_0,
- CFP *for Baire class* 1 *functions implies* ATR_0.

If we want to talk about Caristi's Theorem in reverse math, we have to restrict it to "nice" functions that we can encode in second order arithmetic.

Theorem (FD-S-Y)

- CFP *for continuous functions implies* ACA_0*,*
- CFP *for Baire class* 1 *functions implies* ATR_0*.*

For the second part: given a well-ordering $\prec$, we can construct a Baire class 1 function T with the property that, more or less, $\emptyset$, $T(\emptyset)$, $T(T(\emptyset))$, ... successively calculates the jump hierarchy along $\prec$.

One fairly large class of functions we could hope to work with while staying inside second-order arithmetic is the Borel functions:

Definition

When $\mathcal{X}$ is a complete separable metric space, a Borel code is a well-founded tree T of sequences whose leaves are labeled by open sets, encoding the Borel set

$$U(T) = \bigcap (\mathcal{X} \setminus U(T_n)).$$

A (code for) a Borel function from $\mathcal{X}$ to $\mathcal{Y}$ is a set F where, for each basic open set $B_r(a)$, $F_{(a,r)}$ is a Borel code, indicating that $\mathcal{F}(B_r(a)) \subseteq U(F_{(a,r)})$.

Since Ekeland's variation principle implies Caristi's Theorem for any function we can represent, Π^1_1-CA_0 implies Caristi's Theorem for Borel functions.

Since Ekeland's variation principle implies Caristi's Theorem for any function we can represent, Π^1_1-CA_0 implies Caristi's Theorem for Borel functions.

However Caristi's theorem for Borel functions is a Π^1_2 statement:

> *For every (encoding of a) complete metric space $(\mathcal{X}, d)$, every (code for a) lower semicontinuous V, and code for a Borel function T, either there is an infinite descending sequence through the code for T, or there is a fixed point.*

It is a standard fact that Π^1_1-CA_0 cannot be equivalent to a Π^1_2 statement.

Consider the structure of the proof of Caristi's Theorem from Ekeland's variational principle.

We have a Borel function T and a lower semicontinuous potential V. EVP gives us a critical point: a point x_* so that

$$\forall y \in \mathcal{X}(d(x_*, y) \leq V(x_*) - V(y) \rightarrow x_* = y).$$

Then, since $d(x_*, T(x_*)) \leq V(x_*) - V(T(x_*))$, we must have $x_* = T(x_*)$.

Consider the structure of the proof of Caristi's Theorem from Ekeland's variational principle.

We have a Borel function T and a lower semicontinuous potential V. EVP gives us a critical point: a point x_* so that

$$\forall y \in \mathcal{X}(d(x_*, y) \leq V(x_*) - V(y) \rightarrow x_* = y).$$

Then, since $d(x_*, T(x_*)) \leq V(x_*) - V(T(x_*))$, we must have $x_* = T(x_*)$.

But $T(x_*)$ isn't an arbitrary point! Since T has a Borel code of some height α, $T(x_*)$ can only be bounded more complicated than x_*.

So Caristi's Theorem for Borel functions follows from the following weakening of EVP:

Theorem

When $(\mathcal{X}, d)$ is a complete metric space and $V : \mathcal{X} \to \mathbb{R}_{\geq 0}$ is a lower semicontinuous function, then for any well-ordering α, V has a approximate critical point x_* so that, for any y which is Σ_α in $x_* \oplus V$,

$$d(x_*, y) \leq V(x_*) - V(y)$$

implies $y = x_*$.

So Caristi's Theorem for Borel functions follows from the following weakening of EVP:

Theorem

When $(\mathcal{X}, d)$ is a complete metric space and $V : \mathcal{X} \to \mathbb{R}_{\geq 0}$ is a lower semicontinuous function, then for any well-ordering α, V has a approximate critical point x_* so that, for any y which is Σ_α in $x_* \oplus V$,

$$d(x_*, y) \leq V(x_*) - V(y)$$

implies $y = x_*$.

Statements equivalent to Π^1_1-CA_0 are typically Π^1_3. Most of them have Π^1_2 "approximations" of this kind, and in most cases these are equivalent. The theory at this strength is called $TLPP_0$.

Theorem (FD-S-T-Y)

The following are equivalent:

- $TLPP_0$,
- *Caristi's Theorem for Baire functions,*
- *Caristi's Theorem for Borel functions.*

The defining axiom of Π^1_1-CA_0 is

> *for all parameters X and each formula ϕ, there is a set Y so that $n \in Y$ iff for every set Z, $\phi(n, X, Z)$ holds.*

The defining axiom of Π^1_1-CA_0 is

for all parameters X and each formula ϕ, there is a set Y so that $n \in Y$ iff for every set Z, $\phi(n, X, Z)$ holds.

$TLPP_0$ replaces this with a relativized form:

for all parameters X, each formula ϕ, and each ordering $\prec$, either there is an infinite descending sequence through $\prec$ or there is a set Y so that $n \in Y$ iff for every set Z which is $\Sigma_{\prec}$ in $X \oplus Y$, $\phi(n, X, Z)$ holds.

$TLPP_0$ stands for "transfinite leftmost path principle"; it comes from the following interpretation—one property equivalent to Π^1_1-CA_0 is:

> *every tree has a leftmost path: for every tree T, there is a path X such that there is no path Y to the left of X.*

$TLPP_0$ stands for "transfinite leftmost path principle"; it comes from the following interpretation—one property equivalent to Π^1_1-CA_0 is:

every tree has a leftmost path: for every tree T, there is a path X such that there is no path Y to the left of X.

The relativized version is:

for every tree T and every ordering $\prec$, either there is an infinite descending sequence through $\prec$ or there is an X so that there is no path Y which is to the left of X and is $\Sigma_\prec$ in $T \oplus X$.

The motivation for this theory comes from proof mining. Suppose we prove a Π_2 statement using a Π_3 statement in a basically constructive way:

$$\vdots$$

$$\forall x \exists y \forall z\, \phi(x, y, z)$$

$$\vdots$$

$$\forall u \exists v\, \psi(u, v)$$

The proof encodes information about a function f so that, for each u, $\psi(u, f(u))$ holds.

$$\vdots \qquad\qquad u$$

$$\forall x \exists y \forall z\, \phi(x, y, z)$$

$$\vdots$$

$$\forall u \exists v\, \psi(u, v)$$

The proof encodes information about a function f so that, for each u, $\psi(u, f(u))$ holds.

$$\vdots$$

$$\forall x \exists y \forall z\, \phi(x, y, z)$$

$$\vdots \uparrow$$

$$\forall u \exists v\, \psi(u, v)$$

$$u \Longrightarrow x$$

The proof encodes information about a function f so that, for each u, $\psi(u, f(u))$ holds.

$\vdots \uparrow\downarrow$

$\forall x \exists y \forall z\, \phi(x, y, z)$

$\vdots \uparrow$

$\forall u \exists v\, \psi(u, v)$

$u \Longrightarrow x$

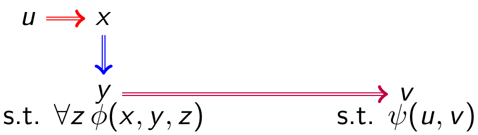
$\Downarrow$

y

s.t. $\forall z\, \phi(x, y, z)$

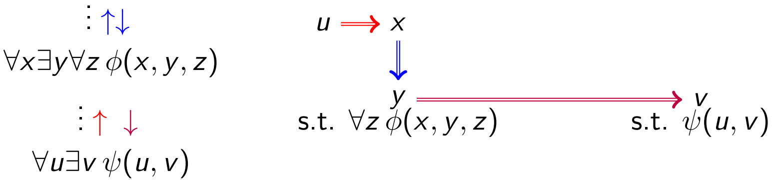
The proof encodes information about a function f so that, for each u, $\psi(u, f(u))$ holds.

$$\vdots \uparrow\downarrow$$
$$\forall x \exists y \forall z\, \phi(x, y, z)$$
$$\vdots \uparrow \ \downarrow$$
$$\forall u \exists v\, \psi(u, v)$$



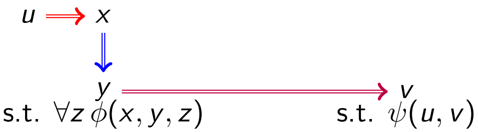
The horizontal lines are computable, but the vertical one typically isn't.

The proof encodes information about a function f so that, for each u, $\psi(u, f(u))$ holds.



The horizontal lines are computable, but the vertical one typically isn't.

In the process of verifying that $\psi(u, v)$ holds, I might use the fact that $\phi(x, y, z)$ holds for various values of z. But since the verification of $\psi(u, v)$ is computable, I only have time to check finitely many values of z.

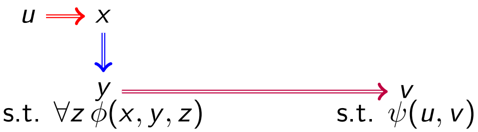


If we examine the arrow from y to v, we can identify which values of z we actually check $\phi(x, y, z)$ for.

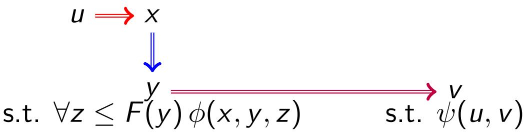
$$\begin{array}{ccc} u \Longrightarrow x & & \\ \Downarrow & & \\ y & \Longrightarrow & v \\ \text{s.t. } \forall z\, \phi(x,y,z) & & \text{s.t. } \psi(u,v) \end{array}$$

If we examine the arrow from y to v, we can identify which values of z we actually check $\phi(x, y, z)$ for. This possibly depends on y, so we get a function $F(y)$ so that

$$\begin{array}{ccc} u \Longrightarrow x & & \\ \Downarrow & & \\ y & \Longrightarrow & v \\ \text{s.t. } \forall z \leq F(y)\, \phi(x,y,z) & & \text{s.t. } \psi(u,v) \end{array}$$



If we examine the arrow from y to v, we can identify which values of z we actually check $\phi(x, y, z)$ for. This possibly depends on y, so we get a function $F(y)$ so that



So instead of working with a true witness y, it would suffice to be given a sufficiently good fake. (Proof mining depends on the idea that sufficiently good fake witnesses are computable.)

Consider what happens when we use Π^1_1-CA_0 exactly once:

$$\vdots$$

$$\forall X \exists Y \forall n (n \in Y \leftrightarrow \forall Z \phi(n, X, Z))$$

$$\vdots$$

$$\forall U \exists V \, \psi(U, V)$$

where ϕ, ψ are arithmetic and the $\vdots$ are arguments in ATR_0.

Consider what happens when we use Π^1_1-CA$_0$ exactly once:

$$\vdots$$

$$\forall X \exists Y \forall n (n \in Y \leftrightarrow \forall Z \phi(n, X, Z))$$

$$\vdots$$

$$\forall U \exists V \, \psi(U, V)$$

where ϕ, ψ are arithmetic and the $\vdots$ are arguments in ATR$_0$.

$$U \Longrightarrow X$$

$$X \Downarrow Y$$

$$Y \Longrightarrow V$$

s.t. $n \in Y \leftrightarrow \forall Z \, \phi(n, X, Z)$ s.t. $\psi(U, V)$

In this case, the horizontal arrows end up being hyperarithmetic constructions.

$$U \Rightarrow X$$
$$\Downarrow$$
$$Y \Longrightarrow V$$

s.t. $n \in Y \leftrightarrow \forall Z\, \phi(n, X, Z)$ s.t. $\psi(U, V)$

Once again, in the process of verifying $\psi(U, V)$, we can only examine sets Z which we construct in ATR_0.

$$U \Longrightarrow X$$

$$X \Downarrow Y$$

$$Y \Longrightarrow V$$

s.t. $n \in Y \leftrightarrow \forall Z\, \phi(n, X, Z)$ s.t. $\psi(U, V)$

Once again, in the process of verifying $\psi(U, V)$, we can only examine sets Z which we construct in ATR_0.

So instead of obtaining a true witness Y, it suffices to obtain an approximate witness so that

$$n \in Y \leftrightarrow \forall Z \in \Sigma_\alpha^{X \oplus Y}\ \phi(n, X, Z).$$

$$U \Rightarrow X$$
$$\Downarrow$$
$$Y \Longrightarrow V$$

s.t. $n \in Y \leftrightarrow \forall Z\, \phi(n, X, Z)$ s.t. $\psi(U, V)$

Once again, in the process of verifying $\psi(U, V)$, we can only examine sets Z which we construct in ATR_0.

So instead of obtaining a true witness Y, it suffices to obtain an approximate witness so that

$$n \in Y \leftrightarrow \forall Z \in \Sigma_\alpha^{X \oplus Y}\ \phi(n, X, Z).$$

TLPP_0 is the theory that promises that these approximate witnesses exist. It captures the Π^1_2 consequences of *one* use of Π^1_1-CA_0.

In theory (but not, as far as I can tell, in practice), one could use Π^1_1-CA_0 repeatedly.

Question

What is a (natural-ish) Π^1_2 theory capturing all Π^1_2 consequences of Π^1_1-CA_0?

There should be a countable hierarchy extending $TLPP_0$ (capturing "n nested applications of Π^1_1-CA_0") which does this. But finding the right way to capture nested applications is non-trivial.

In theory (but not, as far as I can tell, in practice), one could use Π^1_1-CA_0 repeatedly.

Question

What is a (natural-ish) Π^1_2 theory capturing all Π^1_2 consequences of Π^1_1-CA_0?

There should be a countable hierarchy extending TLPP_0 (capturing "*n* nested applications of Π^1_1-CA_0") which does this. But finding the right way to capture nested applications is non-trivial.

Question

Can this be done at higher arity? For instance, is there a hierarchy of Π^1_3 theories approximating the Π^1_3 consequences of Π^1_2-CA_0?

In theory (but not, as far as I can tell, in practice), one could use Π^1_1-CA_0 repeatedly.

Question

What is a (natural-ish) Π^1_2 theory capturing all Π^1_2 consequences of Π^1_1-CA_0?

There should be a countable hierarchy extending $TLPP_0$ (capturing "n nested applications of Π^1_1-CA_0") which does this. But finding the right way to capture nested applications is non-trivial.

Question

Can this be done at higher arity? For instance, is there a hierarchy of Π^1_3 theories approximating the Π^1_3 consequences of Π^1_2-CA_0?

Question

When α is an actual ordinal, what can we say about the complexity of the sets Y of the form $n \in Y \leftrightarrow \forall Z \in \Sigma^{X \oplus Y}_{\alpha}\, \phi(n, X, Z)$?