An application of proof mining to the proximal point algorithm in CAT(0) spaces

Laurențiu Leuștean

Dedicated to the memory of Professor Solomon Marcus (1925-2016)

University of Bucharest

August 27, 2018

(joint work with Andrei Sipoș)

Proof mining

Input: Ineffective proof P of C

Goal: Additional information on C:

- effective bounds, algorithms
- weak dependency or independence from certain parameters,
- generalizations of proofs: weakening of premises

Kohlenbach, Applied Proof Theory: Proof Interpretations and their Use in Mathematics, Springer Monographs in Mathematics (2008).

- inspired by Kreisel's program of unwinding of proofs (50's)
- name of proof mining suggested by Dana Scott

Proof mining

Input: Ineffective proof P of C

Goal: Additional information on C:

- effective bounds, algorithms
- weak dependency or independence from certain parameters,
- generalizations of proofs: weakening of premises

Terence Tao's proposal of hard analysis (finitary, quantitative arguments), instead of soft analysis (infinitary, qualitative arguments)

Tao, Soft analysis, hard analysis, and the finite convergence principle, 2007: "*There are rigorous results from proof theory which can allow one to automatically convert certain types of qualitative arguments into quantitative ones.*"

Proof interpretations

- interpret formulas A occurring in the proof: $A \mapsto A^I$
- the interpretation C^I of the conclusion contains the additional information
- construct by recursion on P a new proof P^I of C^I
- our approach is based on novel extensions of Gödel's functional interpretation

Monotone functional interpretation - Kohlenbach (90's)

- systematically transforms any statement in the proof into a new version with explicit uniform bounds

Logical metatheorems

General metatheorems: "If a sentence has a certain logical form and can be proved in a theory $\mathcal{T}$, then the following strengthened version holds: ...".

- guarantee the extractability of uniform bounds for $\forall\exists$-sentences
- can be used as a black box: infer new uniform existence results without any proof analysis
- extract an explicit uniform bound
- new mathematical proof that uses NO logical tools

Applications: Rates of metastability

Assume that (x_n) is a sequence in a metric space (X, d).

$$(x_n) \text{ Cauchy} \Leftrightarrow \forall k \in \mathbb{N}\, \exists N \in \mathbb{N}\, \forall p \in \mathbb{N} \left(d(x_{N+p}, x_N) < \frac{1}{k+1} \right)$$

- $\forall\exists\forall$-sentence

Applications: Rates of metastability

Assume that (x_n) is a sequence in a metric space (X, d).

$$(x_n) \text{ Cauchy} \Leftrightarrow \forall k \in \mathbb{N}\, \exists N \in \mathbb{N}\, \forall p \in \mathbb{N} \left(d(x_{N+p}, x_N) < \frac{1}{k+1} \right)$$

- $\forall\exists\forall$-sentence
- Specker (1949): computable sequences in $[0,1] \cap \mathbb{Q}$ with NO computable bound on $\exists N$, i.e. with NO computable Cauchy modulus.

Applications: Rates of metastability

Assume that (x_n) is a sequence in a metric space (X, d).

$$(x_n) \text{ Cauchy} \Leftrightarrow \forall k \in \mathbb{N}\, \exists N \in \mathbb{N}\, \forall p \in \mathbb{N} \left(d(x_{N+p}, x_N) < \frac{1}{k+1} \right)$$

- $\forall\exists\forall$-sentence
- Specker (1949): computable sequences in $[0,1] \cap \mathbb{Q}$ with NO computable bound on $\exists N$, i.e. with NO computable Cauchy modulus.

IDEA: Herbrand normal form

$$(x_n) \text{ Cauchy} \Leftrightarrow \forall k \in \mathbb{N}\, \exists N \in \mathbb{N}\, \forall p \in \mathbb{N} \left(d(x_{N+p}, x_N) < \tfrac{1}{k+1} \right)$$

$$\Updownarrow$$

$$\forall k \in \mathbb{N}\, \forall g : \mathbb{N} \to \mathbb{N}\, \exists N\, \forall i, j \in [N, N + g(N)] \left(d(x_i, x_j) < \tfrac{1}{k+1} \right)$$

Applications: Rates of metastability

Metatheorems guarantee that we can extract $\Phi(k, g, \ldots)$ s.t.

$$(\star) \quad \forall k \in \mathbb{N}\, \forall g : \mathbb{N} \to \mathbb{N}\, \exists N \leq \Phi\, \forall i, j \in [N, N+g(N)] \left(d(x_i, x_j) < \frac{1}{k+1} \right)$$

$(\star)$ = monotone functional interpretation of the negative translation of Cauchyness

= Kreisel's no-counterexample interpretation of the Cauchy property

Applications: Rates of metastability

Metatheorems guarantee that we can extract $\Phi(k, g, \ldots)$ s.t.

$$(\star) \quad \forall k \in \mathbb{N}\, \forall g : \mathbb{N} \to \mathbb{N}\, \exists N \leq \Phi\, \forall i, j \in [N, N+g(N)] \left(d(x_i, x_j) < \frac{1}{k+1} \right)$$

$(\star)$ = monotone functional interpretation of the negative translation of Cauchyness
= Kreisel's no-counterexample interpretation of the Cauchy property

Popularized in the last years by Tao as the metastability of (x_n):

- Tao, Soft analysis, hard analysis, and the finite convergence principle (2007)
- Tao, Norm convergence of multiple ergodic averages for commuting transformations, Ergod. Th. Dynam. Syst. (2008)
- Walsh, Norm convergence of nilpotent ergodic averages, Ann. Math. (2012)

Proximal point algorithm in CAT(0) spaces

- The proximal point algorithm is a fundamental tool of convex optimization, going back to Martinet (1970), Rockafellar (1976) and Brézis and Lions (1978).

Let X be a CAT(0) space, $f : X \to (-\infty, +\infty]$ be a convex, lower semicontinuous (lsc) proper function.

- A minimizer of f is a point $x \in X$ s. t. $f(x) = \inf_{y \in X} f(y)$. We denote the set of minimizers of f by $Argmin(f)$.

- For $\gamma > 0$, the resolvent of f of order γ is the mapping

$$J_f^\gamma : X \to X, \quad J_f^\gamma(x) := \operatorname{argmin}_{y \in X} \left[\gamma f(y) + \frac{1}{2} d^2(x, y)\right].$$

- defined by Jost (1995)

- For any $\gamma > 0$, $Fix(J_f^\gamma) = Argmin(f)$.

Proximal point algorithm in CAT(0) spaces

Let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $(0,\infty)$.

The proximal point algorithm $(x_n)_{n\in\mathbb{N}}$ starting with $x\in X$ is defined as follows:

$$x_0 := x, \qquad x_{n+1} := J_f^{\gamma_n} x_n \text{ for all } n\in\mathbb{N}.$$

Theorem Bačák (2013)

Assume that X is a totally bounded CAT(0) space, $Argmin(f)\neq\emptyset$ and that $\sum_{n=0}^{\infty}\gamma_n=\infty$. Then (x_n) is Cauchy.
If, furthermore, X is complete, then (x_n) converges strongly to a a minimizer of f.

- The following quantitative version of Bačák's result is obtained by using proof mining methods.

Effective rates of metastability

Theorem

Let $b > 0$, $\alpha, \theta : \mathbb{N} \to \mathbb{N}$, $M : \mathbb{N} \to (0, \infty)$.
Then there exists $\Psi_{b,\theta,M,\alpha} : \mathbb{N} \times \mathbb{N}^{\mathbb{N}} \to \mathbb{N}$ such that for

- all totally bounded CAT(0) spaces with modulus of total boundedness α;
- all convex lsc proper mappings $f : X \to (-\infty, \infty]$ with $Argmin(f) \neq \emptyset$;
- all $x \in X$ such that $d(x, p) \leq b$ for some minimizer p of f;
- all sequences (γ_n) in $(0, \infty)$ such that $\sum_{n=0}^{\infty} \gamma_n = \infty$ with rate of divergence θ and $M(k) \geq \max_{0 \leq i \leq k} \gamma_i$ for all $k \in \mathbb{N}$,

$\Psi_{b,\theta,M,\alpha}$ is a rate of metastability for the proximal point algorithm (x_n) starting with x, i.e. for all $k \in \mathbb{N}$ and all $g : \mathbb{N} \to \mathbb{N}$,

$$\exists N \leq \Psi_{b,\theta,M,\alpha}(k, g) \forall i, j \in [N, N + g(N)] \left(d(x_i, x_j) \leq \frac{1}{k+1} \right).$$

Effective rates of metastability

$$
\begin{aligned}
\Psi_{b,\theta,M,\alpha}(k,g) &:= (\Psi_0)_{b,\theta,M}(\alpha(4k+3),k,g)\\
(\Psi_0)_{b,\theta,M}(0,k,g) &:= 0\\
(\Psi_0)_{b,\theta,M}(n+1,k,g) &:= \Phi_{b,\theta,M}\left(\chi_g^M((\Psi_0)_{b,\theta,M}(n,k,g),4k+3)\right)
\end{aligned}
$$

where

$$
\begin{aligned}
\chi_g^M(n,r) &:= \max_{i\le n}\max\{i+g(i)-1, g(i)(r+1)\}\\
\Phi_{b,\theta,M}(k) &:= \lceil b^2(k+1)^2\rceil + \beta_{b,\theta}(\lceil 2(k+1)^2 M(k)\rceil - 1)\\
\beta_{b,\theta}(k) &:= \theta^M(\lceil b^2(k+1)/2\rceil)+1\\
\theta^M(n) &:= \max_{i\le k}\theta(i).
\end{aligned}
$$