Quantitative results on the method of averaged projections

Andrei Sipoș

Technische Universität Darmstadt
Institute of Mathematics of the Romanian Academy

July 27, 2018
Logic Colloquium 2018
Udine, Italy

The general problem

In one of the typical problems of nonlinear analysis one has:

- a space X
- a map $T : X \to X$

and wants to find an element of $Fix(T)$, i.e. a fixed point of T.

The space X is usually a linear space (Hilbert, Banach,...), but recently there has been a renewed focus in nonlinear spaces.

The general problem

In one of the typical problems of nonlinear analysis one has:

- a space X
- a map $T : X \to X$

and wants to find an element of $Fix(T)$, i.e. a fixed point of T.

The space X is usually a linear space (Hilbert, Banach,...), but recently there has been a renewed focus in nonlinear spaces.

The space

Let (X, d) be a metric space. We say that:

- a **geodesic** in X is a mapping $\gamma : [0,1] \to X$ such that for any $t, t' \in [0,1]$ we have that

$$d(\gamma(t), \gamma(t')) = |t - t'| d(\gamma(0), \gamma(1))$$

- X is **geodesic** if any two points of it are joined by a geodesic
- X is **CAT(0)** if it is geodesic and for any geodesic $\gamma : [0,1] \to X$ and for any $z \in X$ and $t \in [0,1]$ we have that

$$d^2(z, \gamma(t)) \leq (1-t)d^2(z, \gamma(0)) + td^2(z, \gamma(1)) - t(1-t)d^2(\gamma(0), \gamma(1))$$

Intuition: curvature at most 0. Also: CAT(0) spaces are *uniquely* geodesic, so denote $\gamma(t)$ by $(1-t)\gamma(0) + t\gamma(1)$.

The space

Let (X, d) be a metric space. We say that:

- a **geodesic** in X is a mapping $\gamma : [0, 1] \to X$ such that for any $t, t' \in [0, 1]$ we have that

$$d(\gamma(t), \gamma(t')) = |t - t'| d(\gamma(0), \gamma(1))$$

- X is **geodesic** if any two points of it are joined by a geodesic
- X is **CAT(0)** if it is geodesic and for any geodesic $\gamma : [0, 1] \to X$ and for any $z \in X$ and $t \in [0, 1]$ we have that

$$d^2(z, \gamma(t)) \leq (1-t) d^2(z, \gamma(0)) + t d^2(z, \gamma(1)) - t(1-t) d^2(\gamma(0), \gamma(1))$$

Intuition: curvature at most 0. Also: CAT(0) spaces are *uniquely* geodesic, so denote $\gamma(t)$ by $(1 - t)\gamma(0) + t\gamma(1)$.

The space

Let (X, d) be a metric space. We say that:

- a **geodesic** in X is a mapping $\gamma : [0,1] \to X$ such that for any $t, t' \in [0,1]$ we have that

$$d(\gamma(t), \gamma(t')) = |t - t'| d(\gamma(0), \gamma(1))$$

- X is **geodesic** if any two points of it are joined by a geodesic
- X is **CAT(0)** if it is geodesic and for any geodesic $\gamma : [0,1] \to X$ and for any $z \in X$ and $t \in [0,1]$ we have that

$$d^2(z, \gamma(t)) \leq (1-t) d^2(z, \gamma(0)) + t d^2(z, \gamma(1)) - t(1-t) d^2(\gamma(0), \gamma(1))$$

Intuition: curvature at most 0. Also: CAT(0) spaces are *uniquely* geodesic, so denote $\gamma(t)$ by $(1-t)\gamma(0) + t\gamma(1)$.

The space

Let (X, d) be a metric space. We say that:

- a **geodesic** in X is a mapping $\gamma : [0, 1] \to X$ such that for any $t, t' \in [0, 1]$ we have that

$$d(\gamma(t), \gamma(t')) = |t - t'| d(\gamma(0), \gamma(1))$$

- X is **geodesic** if any two points of it are joined by a geodesic
- X is **CAT(0)** if it is geodesic and for any geodesic $\gamma : [0, 1] \to X$ and for any $z \in X$ and $t \in [0, 1]$ we have that

$$d^2(z, \gamma(t)) \leq (1-t)d^2(z, \gamma(0)) + td^2(z, \gamma(1)) - t(1-t)d^2(\gamma(0), \gamma(1))$$

Intuition: curvature at most 0. Also: CAT(0) spaces are *uniquely* geodesic, so denote $\gamma(t)$ by $(1 - t)\gamma(0) + t\gamma(1)$.

The map

Let (X, d) be a CAT(0) space. We call a map $T : X \to X$ to be **firmly nonexpansive** if for any $x, y \in X$ and any $t \in [0, 1]$ we have that

$$d(Tx, Ty) \leq d((1-t)x + tTx, (1-t)y + tTy).$$

- important in convex optimization, as primary examples include:
  - projections onto closed, convex, nonempty subsets
  - resolvents (of nonexpansive mappings, of convex lsc functions)
- introduced in a nonlinear context by Ariza-Ruiz/Leuștean/López-Acedo (TAMS 2014)
- they satisfy the slightly weaker property (P_2) (though equivalent to f.n.e. in Hilbert spaces): for all $x, y \in X$,

$$2d^2(Tx, Ty) \leq d^2(x, Ty) + d^2(y, Tx) - d^2(x, Tx) - d^2(y, Ty)$$

- in particular, even (P_2) implies nonexpansiveness: for any $x, y \in X$, $d(Tx, Ty) \leq d(x, y)$

The map

Let (X, d) be a CAT(0) space. We call a map $T : X \to X$ to be **firmly nonexpansive** if for any $x, y \in X$ and any $t \in [0, 1]$ we have that

$$d(Tx, Ty) \leq d((1-t)x + tTx, (1-t)y + tTy).$$

- important in convex optimization, as primary examples include:
  - projections onto closed, convex, nonempty subsets
  - resolvents (of nonexpansive mappings, of convex lsc functions)
- introduced in a nonlinear context by Ariza-Ruiz/Leuștean/López-Acedo (TAMS 2014)
- they satisfy the slightly weaker property (P_2) (though equivalent to f.n.e. in Hilbert spaces): for all $x, y \in X$,

  $$2d^2(Tx, Ty) \leq d^2(x, Ty) + d^2(y, Tx) - d^2(x, Tx) - d^2(y, Ty)$$
- in particular, even (P_2) implies nonexpansiveness: for any $x, y \in X$, $d(Tx, Ty) \leq d(x, y)$

The iteration

A primary way of obtaining fixed points is the *Picard iteration*: let $x \in X$ be arbitrary and set for any n, $x_n := T^n x$.

- best-known example is its use in the Banach fixed point theorem (for k-contractions), where one has strong convergence to the unique fixed point
- here, we will only have weaker forms of convergence, but most importantly **asymptotic regularity**:

$$\lim_{n \to \infty} d(x_n, Tx_n) = 0.$$

Intuition:

- convergence: "close to a fixed point"
- asymptotic regularity: "close to *being* a fixed point" (the iteration is then an **approximate fixed point sequence**)

The iteration

A primary way of obtaining fixed points is the *Picard iteration*: let $x \in X$ be arbitrary and set for any n, $x_n := T^n x$.

- best-known example is its use in the Banach fixed point theorem (for k-contractions), where one has strong convergence to the unique fixed point
- here, we will only have weaker forms of convergence, but most importantly **asymptotic regularity**:

$$\lim_{n \to \infty} d(x_n, Tx_n) = 0.$$

Intuition:

- convergence: "close to a fixed point"
- asymptotic regularity: "close to *being* a fixed point" (the iteration is then an **approximate fixed point sequence**)

The iteration

A primary way of obtaining fixed points is the *Picard iteration*: let $x \in X$ be arbitrary and set for any n, $x_n := T^n x$.

- best-known example is its use in the Banach fixed point theorem (for k-contractions), where one has strong convergence to the unique fixed point
- here, we will only have weaker forms of convergence, but most importantly **asymptotic regularity**:

$$\lim_{n \to \infty} d(x_n, Tx_n) = 0.$$

Intuition:

- convergence: "close to a fixed point"
- asymptotic regularity: "close to *being* a fixed point" (the iteration is then an **approximate fixed point sequence**)

The logical form

Asymptotic regularity can be logically expressed as:

$$\forall k \; \exists N \; \forall n \geq N \; d(x_n, Tx_n) < \frac{1}{k+1}.$$

In our future examples, the sequence $(d(x_n, Tx_n))$ will be nonincreasing, so we can simplify it as:

$$\forall k \; \exists N \; d(x_N, Tx_N) < \frac{1}{k+1},$$

with the same N. Since this is a Π_2 statement, it is amenable to the techniques of proof mining.

Proof mining

Proof mining:

- an applied subfield of mathematical logic
- first suggested by G. Kreisel in the 1950s (under the name "proof unwinding"), then given maturity by U. Kohlenbach and his collaborators starting in the 1990s
- goals: to find explicit and uniform witnesses or bounds and to remove superfluous premises from concrete mathematical statements by analyzing their proofs
- tools used: primarily proof interpretations (modified realizability, negative translation, functional interpretation)
- e.g. in the example from before, one should find a **rate of asymptotic regularity**: an explicit formula for N in terms of the k and of (as few as possible of) the other parameters of the problem

Let us see what results have already been obtained for firmly nonexpansive mappings.

Proof mining

Proof mining:

- an applied subfield of mathematical logic
- first suggested by G. Kreisel in the 1950s (under the name "proof unwinding"), then given maturity by U. Kohlenbach and his collaborators starting in the 1990s
- goals: to find explicit and uniform witnesses or bounds and to remove superfluous premises from concrete mathematical statements by analyzing their proofs
- tools used: primarily proof interpretations (modified realizability, negative translation, functional interpretation)
- e.g. in the example from before, one should find a **rate of asymptotic regularity**: an explicit formula for N in terms of the k and of (as few as possible of) the other parameters of the problem

Let us see what results have already been obtained for firmly nonexpansive mappings.

Proof mining

Proof mining:

- an applied subfield of mathematical logic
- first suggested by G. Kreisel in the 1950s (under the name "proof unwinding"), then given maturity by U. Kohlenbach and his collaborators starting in the 1990s
- goals: to find explicit and uniform witnesses or bounds and to remove superfluous premises from concrete mathematical statements by analyzing their proofs
- tools used: primarily proof interpretations (modified realizability, negative translation, functional interpretation)
- e.g. in the example from before, one should find a **rate of asymptotic regularity**: an explicit formula for N in terms of the k and of (as few as possible of) the other parameters of the problem

Let us see what results have already been obtained for firmly nonexpansive mappings.

Proof mining

Proof mining:

- an applied subfield of mathematical logic
- first suggested by G. Kreisel in the 1950s (under the name "proof unwinding"), then given maturity by U. Kohlenbach and his collaborators starting in the 1990s
- goals: to find explicit and uniform witnesses or bounds and to remove superfluous premises from concrete mathematical statements by analyzing their proofs
- tools used: primarily proof interpretations (modified realizability, negative translation, functional interpretation)
- e.g. in the example from before, one should find a **rate of asymptotic regularity**: an explicit formula for N in terms of the k and of (as few as possible of) the other parameters of the problem

Let us see what results have already been obtained for firmly nonexpansive mappings.

The case of a single mapping

If X is a CAT(0) space and $T : X \to X$ is firmly nonexpansive, then:

- for any x, $(T^n x)$ is asymptotically regular by the previously mentioned paper of Ariza-Ruiz/Leuștean/López-Acedo
- rate of asymptotic regularity obtained in the same paper

Families of mappings

Consider now that we have a finite family $(T_i : X \to X)_{1 \le i \le n}$ of firmly nonexpansive mappings.
The problem is to find an element of

$$\bigcap_{i=1}^{n} Fix(T_i) \neq \emptyset$$

(also called a "convex feasibility" problem).

The usual solution is to set $T := T_n \circ ... \circ T_1$ and to iterate it (the "method of alternating projections"), since one can prove that under the nonemptiness assumption from above,

$$Fix(T) = \bigcap_{i=1}^{n} Fix(T_i).$$

Families of mappings

Consider now that we have a finite family $(T_i : X \to X)_{1 \le i \le n}$ of firmly nonexpansive mappings.
The problem is to find an element of

$$\bigcap_{i=1}^{n} Fix(T_i) \neq \emptyset$$

(also called a "convex feasibility" problem).

The usual solution is to set $T := T_n \circ ... \circ T_1$ and to iterate it (the "method of alternating projections"), since one can prove that under the nonemptiness assumption from above,

$$Fix(T) = \bigcap_{i=1}^{n} Fix(T_i).$$

Results in Hilbert spaces

The main issue is that the composition T may no longer be firmly nonexpansive. Still, in Hilbert spaces, it is **strongly nonexpansive**, and for the case $Fix(T) \neq \emptyset$ asymptotic regularity was proven under this hypothesis by Bruck/Reich in 1977.

Note that

$$\bigcap_{i=1}^{n} Fix(T_i)$$

is no longer required to be nonempty. We will call this situation a problem of "intermediate feasibility" for reasons that will be momentarily apparent.

Results in CAT(0) spaces

The problem of intermediate feasibility was studied in CAT(0) spaces only for $n = 2$, but for mappings satisfying property (P_2):

- asymptotic regularity: Ariza-Ruiz/López-Acedo/Nicolae (JOTA 2015)
- an explicit rate: Kohlenbach/López-Acedo/Nicolae (Optimization 2017)

If one defines (where b is a bound on the distance between the initial point x and a given fixed point p):

$$k_b(\varepsilon) := \left\lceil \frac{2b}{\varepsilon} \right\rceil, \qquad \Phi_b(\varepsilon) := k_b(\varepsilon) \cdot \left\lceil \frac{2b(1 + 2^{k_b(\varepsilon)})}{\varepsilon} \right\rceil + 1,$$

then the rate (as N in terms of an ε) is given by $\Phi_b(\varepsilon)$.

Results in CAT(0) spaces

The problem of intermediate feasibility was studied in CAT(0) spaces only for $n = 2$, but for mappings satisfying property (P_2):

- asymptotic regularity: Ariza-Ruiz/López-Acedo/Nicolae (JOTA 2015)
- an explicit rate: Kohlenbach/López-Acedo/Nicolae (Optimization 2017)

If one defines (where b is a bound on the distance between the initial point x and a given fixed point p):

$$k_b(\varepsilon) := \left\lceil \frac{2b}{\varepsilon} \right\rceil, \qquad \Phi_b(\varepsilon) := k_b(\varepsilon) \cdot \left\lceil \frac{2b(1 + 2^{k_b(\varepsilon)})}{\varepsilon} \right\rceil + 1,$$

then the rate (as N in terms of an ε) is given by $\Phi_b(\varepsilon)$.

Most general case

When not even $Fix(T)$ is required to be nonempty (but the T_i's have approximate fixed points), we deal with a problem of "inconsistent feasibility". In Hilbert spaces:

- Bauschke (Proc. AMS 2003) proved asymptotic regularity for the case of T_i's being projections
- Bauschke/Martin-Marquez/Moffat/Wang 2012: asymptotic regularity for arbitrary firmly nonexpansive mappings
- Kohlenbach (FoCM, to appear): explicit polynomial rates for the above

One uses deep results of the theory of maximal monotone operators (Minty's theorem, Brézis-Haraux theorem). As a result, in CAT(0) spaces the problem is still open.

Averaged projections

An alternate way of solving these feasibility problems is the "method of averaged projections". Let $(\lambda_i)_{1\leq i\leq n} \subseteq (0,\infty)$ with

$$\sum_{i=1}^{n} \lambda_i = 1.$$

One then sets:

$$T := \sum_{i=1}^{n} \lambda_i T_i$$

(i.e. a convex combination) and iterates it.

The asymptotic regularity in Hilbert spaces for this operator is usually reduced to the result for the method of alternating projections by the following trick.

The trick

We put on H^n the following scalar product that makes it into a Hilbert space:

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle := \sum_{i=1}^{n} \lambda_i \langle x_i, y_i \rangle.$$

The diagonal of H^n, denoted by Δ_H, is, then, a subspace isometric to H. If we put Q to be the projection onto Δ_H and U to be the operator given by $U(x_1, \ldots, x_n) := (T_1 x_1, \ldots, T_n x_n)$, then one sees that $Q \circ U$ is an operator on Δ_H that is the pushforward by isometry of T.

This idea was used most recently by Bauschke/Martin-Marquez/Moffat/Wang to prove asymptotic regularity for the inconsistent feasibility problem corresponding to this convex combination case.

Our goal: to adapt it to CAT(0) spaces in order to study the intermediate feasibility problem for the same case (with $n = 2$).

Let (X, d) be a metric space and $\lambda \in (0, 1)$. We define $d_\lambda : X^2 \times X^2 \to \mathbb{R}_+$, for any $(x_1, x_2), (y_1, y_2) \in X^2$ by:

$$d_\lambda((x_1, x_2), (y_1, y_2)) := \sqrt{(1 - \lambda)d^2(x_1, y_1) + \lambda d^2(x_2, y_2)}.$$

Then:

- (X^2, d_λ) is a metric space
- if (X, d) is complete, geodesic or CAT(0), then (X^2, d_λ) is also complete, geodesic or CAT(0), respectively

Asymptotic regularity

Therefore, let $T_1, T_2 : X \to X$ be (P_2) mappings and set $T := (1-\lambda)T_1 + \lambda T_2$. Then, by carefully using the geodesic structure of CAT(0) spaces, one may define operators Q and U similar to the ones presented before and prove that they satisfy the required properties.

Thus, by applying the corresponding result for alternating projections, one can prove that if $Fix(T) \neq \emptyset$, then T is asymptotically regular with the same rate obtained by Kohlenbach/López-Acedo/Nicolae for compositions.

The case of projections

If the T_i's are projections, then one can obtain more intricate quantitative results.

Let A, B be two closed, convex, nonempty subsets of X and set

$$d(A, B) := \inf_{(x,y)\in A\times B} d(x, y).$$

We denote by $S_{A,B}$ the set of pairs (x^*, y^*) such that

$$d(x^*, y^*) = d(A, B)$$

(called "best approximation pairs").

In the composition case, one has the following result.

Theorem (Ariza-Ruiz/López-Acedo/Nicolae 2015)

Set $T := P_A \circ P_B$ and $q := d^2(A, B)$. Assume that there is a pair $(x^*, y^*) \in S_{A,B}$. Let $M, b > 0$.
Then, for any $x \in X$ with $d(x, x^*) \le M$ and $d^2(P_A P_B x, P_B x) \le b$, we have that

$$\forall \varepsilon > 0 \ \forall n \ge \left(\left\lfloor \frac{4M^2 b}{\varepsilon^2} \right\rfloor + 2 \right) \ d^2(T^n x, P_B T^n x) \le q + \varepsilon.$$

If we want to use the theorem above, we should link the set $S_{\Delta_X, A \times B}$ that one obtains using the trick with the set $S_{A,B}$ that is meaningful for the problem.

The qualitative result is the following.

Lemma (A.S.)

$S_{\Delta_X, A\times B} = \{(((1-\lambda)a+\lambda b, (1-\lambda)a+\lambda b), (a,b)) \mid (a,b) \in S_{A,B}\}.$

One must analyze here the forward inclusion "$\subseteq$", which proves that, for any w, a, $b \in X$, if for all x', y', $z' \in X$,

$$d_\lambda^2((w,w),(a,b)) \leq d_\lambda^2((x',x'),(y',z')),$$

then for all y, $z \in X$,

$$d(a,b) \leq d(y,z).$$

Proof analysis

One first notices that the final part of the antecedent can be rewritten as:

$$\forall \delta \; d_\lambda^2((w, w), (a, b)) \leq d^2((x', x'), (y', z')) + \delta$$

and the one of the consequent as

$$\forall \varepsilon \; d(a, b) \leq d(y, z) + \varepsilon.$$

A more interesting observation that the proof only uses two instances of the antecedent, namely:

- $x' := (1 - \lambda)a + \lambda b$, $y' := a$, $z' := b$
- $x' := (1 - \lambda)y + \lambda z$, $y' := y$, $z' := z$

Therefore, by pushing some universal quantifiers to the front and getting rid of others, we obtain two δ's for the two possible antecedents and then take their minimum.

The δ that does the job is

$$\delta := \frac{\varepsilon^2}{4} \cdot \lambda(1-\lambda).$$

Using it, we obtain the following.

Corollary (A.S.)

Put $T := (1-\lambda)P_A + \lambda P_B$ and set $r := d(A, B)$. Let $(x^*, y^*) \in S_{A,B}$ and $M, b > 0$. Set $u^* = (1-\lambda)x^* + \lambda y^*$. Then, for any $x \in X$ with $d(x, u^*) \leq M$ and $d^2(P_A x, P_B x) \leq b$, we have that

$$\forall \varepsilon > 0 \; \forall n \geq \left(\left\lfloor \frac{64 M^2 b}{\varepsilon^4 \lambda(1-\lambda)} \right\rfloor + 2 \right) \; d(P_A T^n x, P_B T^n x) \leq r + \varepsilon.$$

Δ-convergence

Related results not involving proof mining concern the way in which this iteration is convergent. We do not have strong convergence, but rather a generalization of the Hilbert space notion of weak convergence to nonlinear spaces, introduced by T. C. Lim in 1976 and called Δ-convergence. In the context of our method, we have the following.

Lemma (A.S.)

Let (x_n) be a sequence in X and $u \in X$. If $((x_n, x_n))$ Δ-converges to (u, u) in (X^2, d_λ), then (x_n) Δ-converges to u in X.

Using this, we can obtain Δ-convergence results for the method of averaged projections.

Thank you for your attention.