On abstract proximal point algorithms and related concepts

Andrei Sipoș

(part of this is jww Laurențiu Leuștean and Adriana Nicolae)

University of Bucharest
Institute of Mathematics of the Romanian Academy

September 17, 2021
IASM-BIRS Workshop 21w5156: New Frontiers in Proofs and Computation
Institute for Advanced Study in Mathematics at Zhejiang University
Hangzhou, Zhejiang, China (held online)

Functions on Hilbert spaces

I will start rather **abruptly**, around proximal mappings of convex functions.

Let H be a Hilbert space and $f : H \to (-\infty, \infty]$. Such a function is called **proper** if there is an $x \in H$ with $f(x) \neq \infty$, and **lower semicontinuous** (or **lsc**) if for any $x \in H$ and any $y < f(x)$ there is an $\varepsilon > 0$ such that for all z with $\|z - x\| < \varepsilon$ we have that $y < f(z)$.

If $\gamma > 0$, we denote by γf the function such that for any $x \in H$, $(\gamma f)(x) := \gamma \cdot f(x)$ and we remark that $Argmin(\gamma f) = Argmin(f)$.

Proximal mappings

If $f : H \to (-\infty, \infty]$ is proper, convex, lsc, we define its **proximal mapping** to be the map $Prox_f : H \to H$, defined, for any $x \in H$, by

$$Prox_f(x) := \text{argmin}_{y \in H} \left[f(y) + \frac{1}{2} \|x - y\|^2 \right].$$

This expression is well-defined (i.e. the minimizer of the bracket exists and is unique) and we have that $Fix(Prox_f) = Argmin(f)$.

We also remark that for any $\gamma > 0$ we have that $Fix(Prox_{\gamma f}) = Argmin(\gamma f) = Argmin(f)$ – the mapping $Prox_{\gamma f}$ is usually called the proximal mapping of f of **order** γ.

The proximal point algorithm (PPA)

Let H and f be as before, with $Argmin(f) \neq \emptyset$. What one wants to do is to find an element of $Argmin(f)$, i.e. a **minimizer** of f.

Let $(\gamma_n) \subseteq (0, \infty)$. We say that a sequence $(x_n) \subseteq H$ is **generated by the proximal point algorithm (PPA)** if for any $n \in \mathbb{N}$,

$$x_{n+1} = Prox_{\gamma_n f} x_n.$$

Given that the (unique) argmin in the definition has to be in turn computed at each step, the PPA is sometimes said not to be a proper algorithm, but a *pseudo-algorithm*.

Convergence

In order to obtain such a minimizer using the algorithm, one usually imposes a condition on the sequence (γ_n). One usually encounters the following conditions in the literature, listed here from the strongest to the weakest:

- $\lim_{n\to\infty} \gamma_n = \infty$;
- $\liminf_{n\to\infty} \gamma_n > 0$;
- $\sum_{n=0}^{\infty} \gamma_n^2 = \infty$;
- $\sum_{n=0}^{\infty} \gamma_n = \infty$.

In our case, the last condition is sufficient (Brézis and Lions, 1978): we have then that the sequence (x_n) **converges weakly** to an element of $Argmin(f)$, while $(f(x_n))$ tends nonincreasingly to the minimum of f. Strong convergence is usually not attainable (Güler, 1991).

Operators

We now move to a slightly different topic. We call a mapping $A : H \to 2^H$ an **operator** on H, which is uniquely determined by its **graph**,

$$\{(x, y) \in H \times H \mid y \in Ax\}.$$

We define the **zero set** of an operator A as

$$zer(A) := \{x \in H \mid 0 \in Ax\},$$

the inverse of an operator A as being the operator that has the graph

$$\{(x, y) \mid x \in Ay\},$$

and the sum of two operators A and B as being the operator that has the graph

$$\{(x, y + z) \mid y \in Ax,\ z \in Bx\}.$$

Operators

We also define the **identity operator** I as being the operator with the graph

$$\{(x, x) \mid x \in H\}$$

and for any operator A and any $\gamma > 0$ we define γA as being the operator with the graph

$$\{(x, \gamma y) \mid y \in Ax\},$$

remarking that $zer(\gamma A) = zer(A)$.

We call an operator A to be **monotone** if for any (x, u), (y, v) in its graph,

$$\langle x - y, u - v \rangle \geq 0.$$

A monotone operator is called **maximally monotone** if its graph is maximal with respect to set inclusion among the graphs of monotone operators on H.

Resolvents

If A is a maximally monotone operator on H, we define its **resolvent** as

$$J_A := (I + A)^{-1}.$$

This operator is always single-valued and always has a full domain, so we may regard it as a mapping $J_A : H \to H$. The relevant property here is that $Fix(J_A) = zer(A)$.

We also remark that for any $\gamma > 0$ we have that $Fix(J_{\gamma A}) = zer(\gamma A) = zer(A)$ – the mapping $J_{\gamma A}$ is usually called the resolvent of A of **order** γ.

The proximal point algorithm for maximally monotone operators

Let H and A be as before, with $zer(A) \neq \emptyset$. What one wants to do is to find an element of $zer(A)$, i.e. a **zero** of A.

Let $(\gamma_n) \subseteq (0, \infty)$. We say that a sequence $(x_n) \subseteq H$ is **generated by the proximal point algorithm (PPA)** if for any $n \in \mathbb{N}$,

$$x_{n+1} = J_{\gamma_n A} x_n.$$

Here, it is necessary to impose the condition $\sum_{n=0}^{\infty} \gamma_n^2 = \infty$ in order to guarantee the weak convergence of the sequence to a zero of A.

Subdifferentials

Let $f : H \to (-\infty, \infty]$ be proper and convex. We define its **subdifferential** $\partial f : H \to 2^H$, for any $x \in H$, by

$$\partial f(x) := \{u \in H \mid \text{for all } y \in H,\ \langle y - x, u \rangle + f(x) \leq f(y)\}.$$

We remark that for any $\gamma > 0$, $\partial(\gamma f) = \gamma(\partial f)$.

If f is in addition lsc, then ∂f is a maximally monotone operator, and the crucial fact here is that

$$Prox_f = J_{\partial f},$$

so a proximal mapping is a special kind of resolvent and it may be called as such. We also remark that for any $\gamma > 0$ we have that

$$Prox_{\gamma f} = J_{\partial(\gamma f)} = J_{\gamma(\partial f)},$$

so the proximal point algorithm for maximally monotone operators encompasses the one using proximal mappings – this justifies using the same name.

Nonexpansive mappings

There is another kind of PPA which arises from nonexpansive mappings. If H is a Hilbert space, a mapping $T : H \to H$ is called **nonexpansive** if for all $x, y \in H$, $\|Tx - Ty\| \leq \|x - y\|$.

For $x \in H$ and $\gamma > 0$ we define

$$G_{T,x,\gamma} : H \to H, \quad G_{T,x,\gamma}(y) := \frac{1}{1+\gamma}x + \frac{\gamma}{1+\gamma}Ty.$$

It is easy to see that this mapping is Lipschitz with constant $\frac{\gamma}{1+\gamma} \in (0,1)$. Therefore it admits a unique fixed point, which we shall denote by $R_{T,\gamma}x$. We have therefore defined a mapping $R_{T,\gamma} : H \to H$, called the *resolvent of order* γ of T, which satisfies, for any $x \in H$,

$$R_{T,\gamma}x = \frac{1}{1+\gamma}x + \frac{\gamma}{1+\gamma}TR_{T,\gamma}x,$$

so $Fix(R_{T,\gamma}) = Fix(T)$.

The proximal point algorithm for nonexpansive mappings

Let H and T be as before, with $Fix(T) \neq \emptyset$. What one wants to do is to find an element of $Fix(T)$, i.e. a fixed point of T.

Let $(\gamma_n) \subseteq (0, \infty)$. We say that a sequence $(x_n) \subseteq H$ is **generated by the proximal point algorithm (PPA)** if for any $n \in \mathbb{N}$,

$$x_{n+1} = R_{T,\gamma_n} x_n.$$

Here, as before, it is necessary to impose the condition $\sum_{n=0}^{\infty} \gamma_n^2 = \infty$ in order to guarantee the weak convergence of the sequence to a fixed point of T.

Also a special case

But this kind of PPA may also be considered to be a special case of the one for maximally monotone operators. Let us see how.

If $T : H \to H$ is nonexpansive, we may define $A_T := I - T$, which is a (single-valued actually) maximally monotone operator, and one can see that $Fix(T) = zer(A_T)$. One then has, for any $\gamma > 0$, that

$$R_{T,\gamma} = J_{\gamma A_T}$$

and that

$$Fix(T) = Fix(R_{T,\gamma}) = Fix(J_{\gamma A_T}) = zer(A_T).$$

Nonlinear spaces

Recently there has been a renewed focus in proving these kinds of results in *nonlinear* spaces. What are those?

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 results of Bačák

Here, we do not necessarily have at hand the unifying framework of maximally monotone operators (or do we?), but the proximal point algorithms for proper convex lsc functions and for nonexpansive mappings may be defined analogously.

The fact that the algorithms still work has been proven by Bačák in 2013 in the setting of complete CAT(0) spaces, also known as **Hadamard spaces** – we briefly note that a Hilbert space is already complete as a metric space, and, moreover, it is a special case of a Hadamard space – and with weak convergence replaced by a notion of Δ-**convergence**, which is equivalent to it when the space is Hilbert.

Firm nonexpansiveness

We may ask ourselves: how are all these results proven? We repeat here the observation of Eckstein (1989), cited in the abstract, that the arguments used to prove the convergence of this algorithm "hinge primarily on the firmly nonexpansive properties of the resolvents". Of course, at that time, he was speaking about Hilbert spaces, so what are firmly nonexpansive mappings in Hilbert spaces?

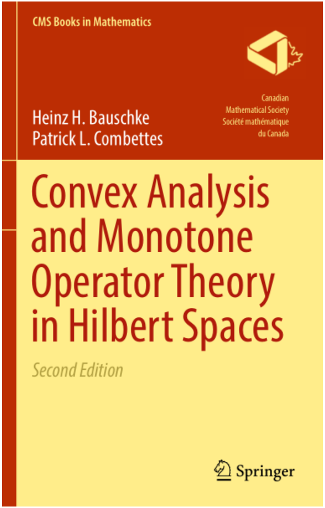
Definition

Let H be a Hilbert space and $T : H \to H$. The following are equivalent:

- for any $x, y \in H$ and any $t \in [0,1]$ we have that
$$\|Tx - Ty\| \leq \|(1-t)(x-y) + t(Tx - Ty)\|;$$
- for any $x, y \in H$, $\|Tx - Ty\|^2 \leq \langle Tx - Ty, x - y\rangle$

In this case, T is called **firmly nonexpansive**.

The proof of that equivalence, along with almost all the facts necessary for understanding convex optimization in Hilbert spaces, may be found in the following book:



Quasilinearization

How do we deal with inner products in the nonlinear setting? Well, in 2008, Berg and Nikolaev proved that in **any** metric space (X, d), the function $\langle \cdot, \cdot \rangle : X^2 \times X^2 \to \mathbb{R}$, defined, for any x, y, u, $v \in X$, by

$$\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle := \frac{1}{2}(d^2(x, v) + d^2(y, u) - d^2(x, u) - d^2(y, v))$$

(where an ordered pair of points $(p, z) \in X^2$ is denoted by $\overrightarrow{pz}$), called the *quasi-linearization function*, is the unique one such that, for any x, y, u, v, $w \in X$, we have that:

- $\langle \overrightarrow{xy}, \overrightarrow{xy} \rangle = d^2(x, y)$;
- $\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle = \langle \overrightarrow{uv}, \overrightarrow{xy} \rangle$;
- $\langle \overrightarrow{yx}, \overrightarrow{uv} \rangle = -\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle$;
- $\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle + \langle \overrightarrow{xy}, \overrightarrow{vw} \rangle = \langle \overrightarrow{xy}, \overrightarrow{uw} \rangle$.

The inner product notation is justified by the fact that if X is a Hilbert space, for any x, y, u, $v \in X$,

$$\langle \overrightarrow{xy}, \overrightarrow{uv} \rangle = \langle x - y, u - v \rangle = \langle y - x, v - u \rangle.$$

We then have **two** generalizations of firm nonexpansiveness to CAT(0) spaces, corresponding to the two equivalent conditions in the Hilbert setting definition. Let X be a CAT(0) space and $T : X \to X$.

In 2014, D. Ariza-Ruiz, L. Leuștean and G. López-Acedo stated that T is **firmly nonexpansive** if for any x, $y \in X$ and any $t \in [0, 1]$ we have that

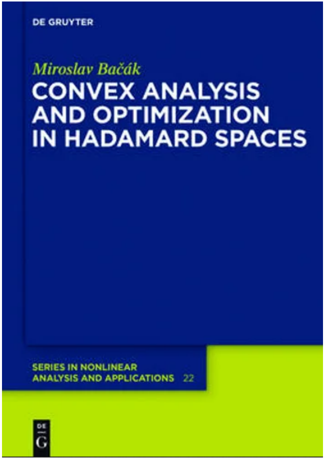
$$d(Tx, Ty) \leq d((1-t)x + tTx, (1-t)y + tTy),$$

while in 2015, D. Ariza-Ruiz, G. López-Acedo and A. Nicolae, stated that T satisfies **property** (P_2) if for any x, $y \in X$,

$$d^2(Tx, Ty) \leq \langle \overrightarrow{TxTy}, \overrightarrow{xy} \rangle.$$

We have that firm nonexpansiveness implies (P_2) but not vice-versa.

As one can notice, convex optimization in nonlinear spaces is a relatively emerging field, so we do not yet have a "bible" like in the Hilbert setting. The following book is a good start, though:



Abstract proof

Our goal is to obtain an **abstract** proof for the convergence of all these variants of the proximal point algorithm.

Still, even considering the remark before, it is not **just** firm nonexpansiveness by itself which is used in the corresponding proofs. We need more abstraction!

Quotes

“Good general theory does not search for the maximum generality, but for the right generality.”
– Saunders Mac Lane

“Formalizations [...] are sensitive to definitions. This makes it interesting. Don’t ask, ‘which is the right definition?’ Ask instead: what are the possible definitions? what are the relationships between them? in which contexts are they useful?”
– Jeremy Avigad

"To understand means to create again, to reproduce inside yourself the initial moment of the artwork."
– George Călinescu, Romanian literary critic

In our case, the initial moment was looking at the convergence proof in the Bauschke/Combettes book:

Proof. Set $(\forall n \in \mathbb{N})\ u_n = (x_n - x_{n+1})/\gamma_n$. Then $(\forall n \in \mathbb{N})\ u_n \in Ax_{n+1}$ and $x_{n+1} - x_{n+2} = \gamma_{n+1} u_{n+1}$. Hence, by monotonicity and Cauchy–Schwarz,

$$(\forall n \in \mathbb{N}) \quad \begin{aligned} 0 &\leqslant \langle x_{n+1} - x_{n+2} \mid u_n - u_{n+1} \rangle / \gamma_{n+1} \\ &= \langle u_{n+1} \mid u_n - u_{n+1} \rangle \\ &= \langle u_{n+1} \mid u_n \rangle - \|u_{n+1}\|^2 \\ &\leqslant \|u_{n+1}\|(\|u_n\| - \|u_{n+1}\|), \end{aligned} \tag{23.37}$$

which implies that $(\|u_n\|)_{n \in \mathbb{N}}$ converges.

This immediately leads to an abstract definition in Hilbert spaces!

If T and U are self-mappings of H and λ, $\mu > 0$, we say that T and U are (λ, μ)-*mutually firmly nonexpansive* if for all x, $y \in H$, one has that

$$\frac{1}{\mu}\langle Tx - Uy, y - Uy\rangle \leq \frac{1}{\lambda}\langle Tx - Uy, x - Tx\rangle.$$

If $(T_n)_{n\in\mathbb{N}}$ is a family of self-mappings of H and $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0, \infty)$, we say that (T_n) is *jointly firmly nonexpansive* with respect to (γ_n) if for all n, $m \in \mathbb{N}$, T_n and T_m are (γ_n, γ_m)-mutually firmly nonexpansive.

In addition, if $(T_\gamma)_{\gamma>0}$ is a family of self-mappings of H, we say that it is plainly *jointly firmly nonexpansive* if for all λ, $\mu > 0$, T_λ and T_μ are (λ, μ)-mutually firmly nonexpansive.

It is clear that a family $(T_\gamma)_{\gamma>0}$ is jointly firmly nonexpansive if and only if for every $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0, \infty)$, $(T_{\gamma_n})_{n\in\mathbb{N}}$ is jointly firmly nonexpansive with respect to (γ_n).

Examples

It is then shown that examples of jointly firmly nonexpansive families of mappings are furnished by the resolvent-type mappings we mentioned before – specifically, by:

- the family $(Prox_{\gamma f})_{\gamma>0}$, where f is a proper convex lsc function on H;
- the family $(R_{T,\gamma})_{\gamma>0}$, where T is a nonexpansive self-mapping of H;
- the family $(J_{\gamma A})_{\gamma>0}$, where A is a maximally monotone operator on H.

A first abstract proximal point algorithm

We thus obtain the following abstract PPA in Hilbert spaces.

Theorem (L. Leuștean, A. Nicolae, A. S., 2018)

Let H be a Hilbert space, $T_n : H \to H$ for every $n \in \mathbb{N}$ and (γ_n) be a sequence of positive real numbers satisfying $\sum_{n=0}^{\infty} \gamma_n^2 = \infty$. Assume that the family (T_n) is jointly firmly nonexpansive with respect to (γ_n) and that $F := \bigcap_{n\in\mathbb{N}} Fix(T_n) \neq \emptyset$. Let $(x_n) \subseteq H$ be such that for any n, $x_{n+1} = T_n x_n$. Then (x_n) converges weakly to a point in F.

Mutual firm nonexpansiveness in CAT(0) spaces

How would we define a notion of mutual firm nonexpansiveness in a CAT(0) space X? Well, we may first state a definition in the Ariza-Ruiz/Leuștean/López-Acedo style, as follows: if T and U are self-mappings of X and λ, $\mu > 0$, we say that T and U are (λ, μ)-*mutually firmly nonexpansive* if for all x, $y \in X$ and all α, $\beta \in [0, 1]$ such that $(1 - \alpha)\lambda = (1 - \beta)\mu$, one has that

$$d(Tx, Uy) \leq d((1 - \alpha)x + \alpha Tx, (1 - \beta)y + \beta Uy).$$

The corresponding "joint firm nonexpansiveness" notions are then defined analogously.

The motivation for the above definition arises from the so-called *resolvent identity*, that those resolvent-type mappings were already known to satisfy, namely that if $(T_\gamma)_{\gamma>0}$ was such a family, then for all $\gamma > 0$, $t \in [0, 1]$ and $x \in X$,

$$T_{(1-t)\gamma}((1 - t)x + tT_\gamma x) = T_\gamma x.$$

Let us now dwell on the precise connection between the two.

Resolvent identity implies joint firm nonexpansiveness

Proposition (L. Leuștean, A. Nicolae, A. S., 2018)

Let $(T_\gamma)_{\gamma>0}$ be a family of nonexpansive self-mappings of X that satisfy the resolvent identity. Then $(T_\gamma)_{\gamma>0}$ is jointly firmly nonexpansive.

Proof

Let λ, $\mu > 0$, x, $y \in X$ and α, $\beta \in [0,1]$ be such that $(1-\alpha)\lambda = (1-\beta)\mu =: \delta$. We get that

$$
\begin{aligned}
&d(T_\lambda x, T_\mu y) \\
&\quad = d(T_{(1-\alpha)\lambda}((1-\alpha)x + \alpha T_\lambda x), T_{(1-\beta)\mu}((1-\beta)y + \beta T_\mu y)) \\
&\quad = d(T_\delta((1-\alpha)x + \alpha T_\lambda x), T_\delta((1-\beta)y + \beta T_\mu y)) \\
&\quad \leq d((1-\alpha)x + \alpha T_\lambda x, (1-\beta)y + \beta T_\mu y).
\end{aligned}
$$

This was used in that paper to show that the examples shown before carry over to the CAT(0) setting.

Joint firm nonexpansiveness implies resolvent identity

Proposition (A. S., 2020)

Let $(T_\gamma)_{\gamma>0}$ be a jointly firmly nonexpansive family of self-mappings of X. Then all the mappings in the family are nonexpansive and the family satisfies the resolvent identity.

Proof

Let $\gamma > 0$. Take x, $y \in X$. To show that $d(T_\gamma x, T_\gamma y) \leq d(x,y)$, simply set in the joint firm nonexpansiveness condition $\lambda := \gamma$, $\mu := \gamma$, $\alpha := 0$ and $\beta := 0$.

To show that the resolvent identity holds, take $t \in [0,1]$ and $x \in X$. Set $y := (1-t)x + tT_\gamma x$, so one has to prove that $T_{(1-t)\gamma} y = T_\gamma x$. Then, if one sets $\lambda := \gamma$, $\mu := (1-t)\gamma$, $\alpha := t$ and $\beta := 0$, since then $(1-\alpha)\lambda = (1-\beta)\mu$, one gets that

$$d(T_\gamma x, T_{(1-t)\gamma} y) \leq d((1-t)x + tT_\gamma x, y) = 0,$$

so $T_{(1-t)\gamma} y = T_\gamma x$.

Mutually (P_2)

Of course, we may also define the corresponding Ariza-Ruiz/López-Acedo/Nicolae style (P_2) versions of the definition: if T and U are self-mappings of X and λ, $\mu > 0$, we say that T and U are (λ, μ)-*mutually* (P_2) if for all x, $y \in H$, one has that

$$\frac{1}{\mu}\langle \overrightarrow{TxUy}, \overrightarrow{yUy} \rangle \leq \frac{1}{\lambda}\langle \overrightarrow{TxUy}, \overrightarrow{xTx} \rangle.$$

Again, the corresponding "jointly (P_2)" notions are then defined analogously, and all those (P_2) notions generalize their firmly nonexpansive counterparts and coincide with them in the case where X is a Hilbert space.

The abstract proximal point algorithm

Moreover, the property (P_2) is sufficient to prove the Δ-convergence of the corresponding abstract proximal point algorithm.

Theorem (L. Leuştean, A. Nicolae, A. S., 2018)

Let X be a complete CAT(0) space, $T_n : X \to X$ for every $n \in \mathbb{N}$ and (γ_n) be a sequence of positive real numbers satisfying $\sum_{n=0}^{\infty} \gamma_n^2 = \infty$. Assume that the family (T_n) is jointly (P_2) with respect to (γ_n) and that $F := \bigcap_{n \in \mathbb{N}} Fix(T_n) \neq \emptyset$. Let $(x_n) \subseteq X$ be such that for any n, $x_{n+1} = T_n x_n$. Then (x_n) Δ-converges to a point in F.

Enter proof mining

Most of my research so far has taken place in the field of *proof mining*, an applied subfield of mathematical logic:

- first suggested by G. Kreisel in the 1950s (under the name "unwinding of proofs"), then given maturity by U. Kohlenbach and his school 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
- the adequacy of the tools to the goals is guaranteed by **general logical metatheorems**
- a survey of recent results may be found in Kohlenbach's contribution to ICM 2018
- also see my series of articles on the Proof Theory Blog, entitled **What proof mining is about**, for a short and hopefully accessible introduction

The uniform case

In the case where the resolvent mappings arise from a uniform object (e.g. uniformly monotone operators, uniformly convex functions), the mappings are **uniformly** (P_2), a condition which looks like

$$\langle \overrightarrow{TxTy}, \overrightarrow{yTy} \rangle \leq \langle \overrightarrow{TxTy}, \overrightarrow{xTx} \rangle - \varphi(\varepsilon).$$

and the corresponding optimizing point is unique.

In this case, using ideas by Kohlenbach '90 and Kohlenbach/Oliva '03, one may obtain a sufficiently constructive proof in order to get a **rate of convergence**.

We did this in 2018, but last year we found out that we may use a quantitative lemma of Kohlenbach/Powell '20 to get one with weaker restrictions. For example, it is enough to assume that

$$\sum_{n=0}^{\infty} \gamma_n = \infty.$$

The statement for the uniform case

Theorem (A. S., 2020)

Let $(T_n)_{n\in\mathbb{N}}$ be a family of self-mappings of X and $p \in \bigcap_{n\in\mathbb{N}} Fix(T_n)$. Let $(\gamma_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$ and $\theta : (0,\infty) \to \mathbb{N}$ and assume that for all $x > 0$,

$$\sum_{n=0}^{\theta(x)} \gamma_n \geq x.$$

Let $b > 0$ and denote by C the closed ball of center p and radius b. Let $\varphi : (0,\infty) \to (0,\infty)$ and assume that for all n, $T_n(C) \subseteq C$ and T_n is uniformly (P_2) on C with modulus $\gamma_n\varphi$. Let $(x_n) \subseteq C$ be such that for all n, $x_{n+1} = T_n x_n$.
Then for all $\varepsilon > 0$ and all $n \geq \theta\left(\frac{(b+1)^2}{\varphi(\varepsilon)}\right) + 1$, $d(x_n, p) \leq \varepsilon$.

Approximating curves

Finally, we have obtained the following result about the asymptotic behaviour at infinity, which subsumes a lot of classical results due to Minty, Halpern, Bruck, Jost, as well as a recent one due to Bačák and Reich.

Theorem (A. S., 2020)

Assume that X is complete. Let $(T_\gamma)_{\gamma>0}$ be a jointly firmly nonexpansive family of self-mappings of X. Put $F := \bigcap_{\gamma>0} Fix(T_\gamma)$. Let $x \in X$, $b > 0$, and $(\lambda_n)_{n\in\mathbb{N}} \subseteq (0,\infty)$ with $\lim_{n\to\infty} \lambda_n = \infty$ and assume that for all n, $d(x, T_{\lambda_n} x) \leq b$. Then $F \neq \emptyset$ and the curve $(T_\gamma x)_{\gamma>0}$ is continuous and converges to the unique point in F which is closest to x.

The Halpern PPA

Another, very fresh, development was to provide an abstract version of the so-called Halpern proximal point algorithm, where one fixes an "anchor point" u and a sequence of 'weights' $(\alpha_n) \subseteq (0,1)$ and one builds the iterative sequence as follows: for every $n \in \mathbb{N}$,

$$x_{n+1} := \alpha_n u + (1-\alpha_n) T_n x_n.$$

The advantage of this algorithm is that it is strongly convergent! We chose to base our approach to its abstraction on an argument of Aoyama and Toyoda (2017) – recently proof-mined by Kohlenbach (2020) – because it shows its strong convergence in quite general conditions on the weights and on the step-sizes.

Strong (quasi-)nonexpansiveness

The main difficulty to overcome was that their original argument relied on a property of resolvents called **strong nonexpansiveness**, which is 'somewhat artificial' (Kohlenbach 2016) to adapt to the metric context, where one usually has at most a property called uniform strong quasi-nonexpansiveness. We first showed that this property holds for (P_2) mappings, giving its corresponding 'SQNE-modulus':

Proposition

Let ε, $b > 0$, $z \in X$, $T : X \to X$ a (P_2) mapping and $p \in Fix(T)$. Assume that $d(z, p) \leq b$. Then, if

$$d(z, p) - d(Tz, p) < \varepsilon^2/(2b),$$

we have that $d(z, Tz) < \varepsilon$.

Quantitative quasiness

This turned out to be enough for the pure abstract strong convergence theorem, but we needed something more if we wanted to obtain a quantitative version of convergence (the details of which we shall not expound upon). Thus, we mined the previous lemma to obtain the following:

Proposition

Let ε, $b > 0$, z, $p \in X$ and $T : X \to X$ a (P_2) mapping. Assume that $d(z,p) \le b$ and $d(p, Tp) \le b$. Then, if

$$d(z,p) - d(Tz,p) \le \varepsilon^2/(15b)$$

and

$$d(p, Tp) \le \varepsilon^2/(15b),$$

we have that $d(z, Tz) \le \varepsilon$.

This stronger property we dubbed 'quantitative quasiness', and made the whole argument go through.

These results may be found in:

L. Leuștean, A. Nicolae, A. Sipoș, An abstract proximal point algorithm. *Journal of Global Optimization*, Volume 72, Issue 3, 553–577, 2018.

A. Sipoș, Revisiting jointly firmly nonexpansive families of mappings. arXiv:2006.02167 [math.OC], 2020. To appear in: *Optimization*.

A. Sipoș, Abstract strongly convergent variants of the proximal point algorithm. arXiv:2108.13994 [math.OC], 2021. Submitted.

And there are most probably many more optimization algorithms that can be fit into our framework!

Thank you for your attention.