Quantitative information on the strong convergence of algorithms via proof mining

Bruno Dinis and Pedro Pinto

CMAFcIO - University of Lisbon

bmdinis@fc.ul.pt

CCC 2020 - Faro
August 31, 2020

Overview

Outline

Proof mining

Framework

A theorem by Yao-Noor

A theorem by Wang-Cui

Krasnosel'skiĭ-Mann iteration with Tikhonov regularization terms

Proof mining

Proof mining program $\rightarrow$ analyses of mathematical proofs with the help of proof theoretic techniques, including functional interpretations, in search of concrete new information: effective bounds, algorithms, weakening of premisses, ...

Proof mining

Proof mining program $\rightarrow$ analyses of mathematical proofs with the help of proof theoretic techniques, including functional interpretations, in search of concrete new information: effective bounds, algorithms, weakening of premisses, ...

The underlying theoretical tools:

Proof mining

Proof mining program $\rightarrow$ analyses of mathematical proofs with the help of proof theoretic techniques, including functional interpretations, in search of concrete new information: effective bounds, algorithms, weakening of premisses, ...

The underlying theoretical tools:

- Ensure that we are always able to extract information for the corresponding quantitative versions

Proof mining

Proof mining program $\rightarrow$ analyses of mathematical proofs with the help of proof theoretic techniques, including functional interpretations, in search of concrete new information: effective bounds, algorithms, weakening of premisses, ...

The underlying theoretical tools:

- Ensure that we are always able to extract information for the corresponding quantitative versions
- Help navigate the original proof

Proof mining

Proof mining program $\to$ analyses of mathematical proofs with the help of proof theoretic techniques, including functional interpretations, in search of concrete new information: effective bounds, algorithms, weakening of premisses, ...

The underlying theoretical tools:

- Ensure that we are always able to extract information for the corresponding quantitative versions
- Help navigate the original proof
- Allow to avoid certain non-essential principles

Proof mining

Proof mining program → analyses of mathematical proofs with the help of proof theoretic techniques, including functional interpretations, in search of concrete new information: effective bounds, algorithms, weakening of premisses, ...

The underlying theoretical tools:

- Ensure that we are always able to extract information for the corresponding quantitative versions
- Help navigate the original proof
- Allow to avoid certain non-essential principles
- Allow to obtain explicit bounds

Proof mining with the BFI

We use the Bounded Functional Interpretation (BFI) and its characteristic principles, enriched with a new base type for elements of the space and the (universal) axioms for the Hilbert space.

Proof mining with the BFI

We use the Bounded Functional Interpretation (BFI) and its characteristic principles, enriched with a new base type for elements of the space and the (universal) axioms for the Hilbert space.

Proof mining with the BFI

We use the Bounded Functional Interpretation (BFI) and its characteristic principles, enriched with a new base type for elements of the space and the (universal) axioms for the Hilbert space.

- Usually proof mining disregards precise witnesses, caring only for bounds on them

Proof mining with the BFI

We use the Bounded Functional Interpretation (BFI) and its characteristic principles, enriched with a new base type for elements of the space and the (universal) axioms for the Hilbert space.

- Usually proof mining disregards precise witnesses, caring only for bounds on them
- Completely new translation of formulas

Proof mining with the BFI

We use the Bounded Functional Interpretation (BFI) and its characteristic principles, enriched with a new base type for elements of the space and the (universal) axioms for the Hilbert space.

- Usually proof mining disregards precise witnesses, caring only for bounds on them
- Completely new translation of formulas
- Independence on bounded parameters is made explicit (via the interpretation itself)

Outline

Framework

Let H be a real Hilbert space with inner product $\langle\cdot\rangle$ and norm $\|\cdot\|$ and let $T : H \to 2^H$ be an operator in H.
T is monotone if

$$(x, y), (x', y') \in \Gamma(T) \Rightarrow \langle x - x', y - y'\rangle \geq 0.$$

A monotone operator T is maximal monotone if $\Gamma(T)$ is not properly contained in the graph of any other monotone operator on H.

Framework

Let H be a real Hilbert space with inner product $\langle\cdot\rangle$ and norm $\|\cdot\|$ and let $T : H \to 2^H$ be an operator in H.
T is monotone if

$$(x, y), (x', y') \in \Gamma(T) \Rightarrow \langle x - x', y - y'\rangle \geq 0.$$

A monotone operator T is maximal monotone if $\Gamma(T)$ is not properly contained in the graph of any other monotone operator on H.
We denote by $\mathrm{zer}(T)$ the (nonempty) set of all zeros of T.

Framework

For $c > 0$, we use J_c to denote the resolvent of T, i.e. the single-valued function defined by

$$J_c = (I + cT)^{-1}.$$

Framework

For $c > 0$, we use J_c to denote the resolvent of T, i.e. the single-valued function defined by

$$J_c = (I + cT)^{-1}.$$

A mapping $f : K \to K$ is called nonexpansive if for all $x, y \in K$

$$\|f(x) - f(y)\| \leq \|x - y\|.$$

Framework

For $c > 0$, we use J_c to denote the resolvent of T, i.e. the single-valued function defined by

$$J_c = (I + cT)^{-1}.$$

A mapping $f : K \to K$ is called nonexpansive if for all $x, y \in K$

$$\|f(x) - f(y)\| \leq \|x - y\|.$$

The resolvent J_c is nonexpansive, and

$$\boxed{\mathrm{Fix}(J_c) = \mathrm{zer}(T)}$$

Framework

For $c > 0$, we use J_c to denote the resolvent of T, i.e. the single-valued function defined by

$$J_c = (I + cT)^{-1}.$$

A mapping $f : K \to K$ is called nonexpansive if for all $x, y \in K$

$$\|f(x) - f(y)\| \leq \|x - y\|.$$

The resolvent J_c is nonexpansive, and

$$\boxed{\mathrm{Fix}(J_c) = \mathrm{zer}(T)}$$

Notation

- *$B_D(p)$ denotes the closed ball of radius D, centered at p*

Why go PPA?

- Many problems can be formulated as finding a zero of maximal monotone operators.

Why go PPA?

- Many problems can be formulated as finding a zero of maximal monotone operators.
- PPA is a powerful and successful algorithm in finding a solution of maximal monotone operators.

Why go PPA?

- Many problems can be formulated as finding a zero of maximal monotone operators.
- PPA is a powerful and successful algorithm in finding a solution of maximal monotone operators.
- Starting from any initial guess $z_0 \in H$, the PPA generates a sequence which approximates weakly the solution.

Why go PPA?

- Many problems can be formulated as finding a zero of maximal monotone operators.
- PPA is a powerful and successful algorithm in finding a solution of maximal monotone operators.
- Starting from any initial guess $z_0 \in H$, the PPA generates a sequence which approximates weakly the solution.

How to modify the PPA so that strong convergence is guaranteed?

mPPA

Let H be a Hilbert space. The proximal point algorithm with multi-parameters is the following algorithm

$$\boxed{z_{n+1} = \lambda_n u + \gamma_n z_n + \delta_n J_{c_n}(z_n) + e_n, n \geq 0,} \qquad \text{(mPPA)}$$

where

- $u \in H$ is given,
- $c_n > 0$,
- $\lambda_n, \gamma_n, \delta_n \in (0, 1)$
- $\lambda_n + \gamma_n + \delta_n = 1, \forall n \geq 0$.

Outline

A theorem by Yao-Noor

$$\boxed{z_{n+1} = \lambda_n u + \gamma_n z_n + \delta_n J_{c_n}(z_n) + e_n}$$

Theorem

Let (z_n) *be generated by* (mPPA)*. Assume that*

1. $\lim_{n\to\infty} \lambda_n = 0$;
2. $\sum_{n=0}^{\infty} \lambda_n = \infty$;
3. $0 < \liminf_{n\to\infty} \gamma_n \leq \limsup_{n\to\infty} \gamma_n < 1$;
4. $c_n \geq c$, *where* c *is some positive constant;*
5. $c_{n+1} - c_n \to 0$;
6. $\sum_{n=1}^{\infty} \|e_n\| < \infty$.

Then (z_n) *converges strongly to a point* $z \in \mathrm{zer}(T)$ *which is the nearest point projection of* u *onto* $\mathrm{zer}(T)$.

Quantitative version of Yao-Noor's theorem I

Let (z_n) be generated by mPPA. Assume that there exist $a, b, c, d \in \mathbb{N}$ and $s \in \mathrm{zer}(T)$ and monotone functions $\ell, L, \Delta, \Gamma, E$ such that

(i) $\forall k \in \mathbb{N} \forall n \geq \ell(k) \left(\lambda_n \leq \frac{1}{k+1}\right)$;

(ii) $\forall k \in \mathbb{N} \left(\sum_{i=1}^{L(k)} \lambda_i \geq k\right)$;

(iii) $\forall n \geq a \left(\frac{1}{a+1} \leq \gamma_n \leq 1 - \frac{1}{a+1}\right)$;

(iv) $\forall n \in \mathbb{N} \left(c_n \geq \frac{1}{c+1}\right)$;

(v) $\forall n \in \mathbb{N} \left(\frac{1}{\Delta(n)+1} \leq \eta_n \leq 1 - \frac{1}{\Delta(n)+1}\right), \eta_n \in \{\lambda_n, \gamma_n, \delta_n\}$;

(vi) $\forall k \in \mathbb{N} \forall n \geq \Gamma(k) \left(|c_{n+1} - c_n| \leq \frac{1}{k+1}\right)$;

(vii) $\forall k \in \mathbb{N} \forall n \in \mathbb{N} \left(\sum_{i=E(k)+1}^{E(k)+n} \|e_i\| \leq \frac{1}{k+1}\right)$.

Quantitative version of Yao-Noor's theorem II

Theorem

Under the assumptions (i)-(vii), *for every* $k \in \mathbb{N}$ *and every* $f : \mathbb{N} \to \mathbb{N}$

$$\exists n \leq \phi(k, f) \forall i, j \in [n, n + f(n)] \left(\|z_i - z_j\| \leq \frac{1}{k+1} \right),$$

where $\phi(k, f)$ *is a recursive bound explicitly given in the proof.*

Remarks

- The original proof of the theorem uses explicitly the $\limsup$ of a certain sequence. From a logical point of view this amounts to using arithmetical comprehension which requires bar recursors in order to be interpreted.

Remarks

- The original proof of the theorem uses explicitly the $\limsup$ of a certain sequence. From a logical point of view this amounts to using arithmetical comprehension which requires bar recursors in order to be interpreted.
- We were able to eliminate this dependency by using the fact that the sequence in question is bounded and using a rational approximation to the $\limsup$ instead.

Remarks

- The original proof of the theorem uses explicitly the $\limsup$ of a certain sequence. From a logical point of view this amounts to using arithmetical comprehension which requires bar recursors in order to be interpreted.
- We were able to eliminate this dependency by using the fact that the sequence in question is bounded and using a rational approximation to the $\limsup$ instead.
- We were also able to eliminate an argument of weak compactness.

Remarks

- The original proof of the theorem uses explicitly the $\limsup$ of a certain sequence. From a logical point of view this amounts to using arithmetical comprehension which requires bar recursors in order to be interpreted.
- We were able to eliminate this dependency by using the fact that the sequence in question is bounded and using a rational approximation to the $\limsup$ instead.
- We were also able to eliminate an argument of weak compactness.
- Moreover countable choice (in the projection argument) was eliminated due to a previous observation by Kohlenbach.

Remarks

- The original proof of the theorem uses explicitly the $\limsup$ of a certain sequence. From a logical point of view this amounts to using arithmetical comprehension which requires bar recursors in order to be interpreted.
- We were able to eliminate this dependency by using the fact that the sequence in question is bounded and using a rational approximation to the $\limsup$ instead.
- We were also able to eliminate an argument of weak compactness.
- Moreover countable choice (in the projection argument) was eliminated due to a previous observation by Kohlenbach.
- The extracted bound is uniform on the parameters (depends only on $a, b, c, d, s, \ell, L, \Delta, \Gamma, E$).

Outline

A theorem by Wang-Cui

$$\boxed{z_{n+1} = \lambda_n u + \gamma_n z_n + \delta_n J_{c_n}(z_n) + e_n} \qquad \text{(mPPA)}$$

- (C_1) $\lim \lambda_n = 0$,
- (C_2) $\sum_{n=0}^{\infty} \lambda_n = \infty$,
- (C_3) $\liminf c_n > 0$,
- (C_4) $\liminf \delta_n > 0$,
- (C_5) $\sum_{n=1}^{\infty} \|e_n\| < \infty$ $\left(\text{or } \lim \frac{\|e_n\|}{\lambda_n} = 0\right)$.

A theorem by Wang-Cui

$$\boxed{z_{n+1} = \lambda_n u + \gamma_n z_n + \delta_n J_{c_n}(z_n) + e_n} \qquad \text{(mPPA)}$$

(C_1) $\lim \lambda_n = 0$,

(C_2) $\sum_{n=0}^{\infty} \lambda_n = \infty$,

(C_3) $\liminf c_n > 0$,

(C_4) $\liminf \delta_n > 0$,

(C_5) $\sum_{n=1}^{\infty} \|e_n\| < \infty$ $\left(\text{or } \lim \frac{\|e_n\|}{\lambda_n} = 0\right)$.

Theorem

Let (z_n) be generated by (mPPA). Assume that conditions $(C_1) - (C_5)$ hold. Then (z_n) converges strongly to a point $z \in \mathrm{zer}(T)$ (the nearest point projection of u onto $\mathrm{zer}(T)$).

Quantitative conditions

(Q_1) $\forall n \in \mathbb{N} \left(\lambda_n \geq \frac{1}{h(n)} \right)$

(Q_2) $\forall k \in \mathbb{N} \, \forall n \geq \ell(k) \left(\lambda_n \leq \frac{1}{k+1} \right)$

(Q_3) $\forall k \in \mathbb{N} \left(\sum_{i=1}^{L(k)} \lambda_i \geq k \right)$

(Q_4) $\forall n \in \mathbb{N} \left(\min\{c_n, \delta_n^2\} \geq \frac{1}{c} \right)$

(Q_{5a}) $\forall k \in \mathbb{N} \, \forall n \in \mathbb{N} \left(\sum_{i=E(k)+1}^{E(k)+n} \|e_i\| \leq \frac{1}{k+1} \right)$

(Q_{5b}) $\forall k \in \mathbb{N} \, \forall n \geq E(k) \left(\frac{\|e_n\|}{\lambda_n} \leq \frac{1}{k+1} \right)$

Quantitative version of Wang-Cui's theorem

Theorem

Let (z_n) be generated by (mPPA). Assume $(Q_1) - (Q_4)$ and either (Q_{5a}) or (Q_{5b}). Let $D \in \mathbb{N} \setminus \{0\}$ be such that $D \geq \max\{2\|u-p\|, \|z_0 - p\|\}$, for some $p \in \mathrm{zer}(T)$. Then for all $k \in \mathbb{N}$ and monotone function $f : \mathbb{N} \to \mathbb{N}$

$$\exists n \leq \mu(k,f)\, \exists z \in B_D(p)\, \forall i \in [n, f(n)] \left(s_i^z \leq \frac{1}{k+1} \right)$$

with $s_n^z := ||y_n - z||^2$ and where $\mu(k,f)$ is a recursive bound explicitly given in the proof.

Quantitative version of Wang-Cui's theorem

$$\boxed{z_{n+1} = \lambda_n u + \gamma_n z_n + \delta_n J_{c_n}(z_n) + e_n} \tag{mPPA}$$

$$\boxed{y_{n+1} = \lambda_n u + \gamma_n y_n + \delta_n J_{c_n}(y_n)} \tag{mPPA$_\mathrm{e}$}$$

Theorem

Let (z_n) *be generated by* (mPPA). *Assume* $(Q_1) - (Q_4)$ *and either* (Q_{5a}) *or* (Q_{5b}). *Let* $D \in \mathbb{N} \setminus \{0\}$ *be such that* $D \geq \max\{2\|u-p\|, \|z_0 - p\|\}$, *for some* $p \in \mathrm{zer}(T)$. *Then for all* $k \in \mathbb{N}$ *and monotone function* $f : \mathbb{N} \to \mathbb{N}$

$$\exists n \leq \mu(k,f)\, \exists z \in B_D(p)\, \forall i \in [n, f(n)] \left(s_i^z \leq \frac{1}{k+1} \right)$$

with $s_n^z := \|y_n - z\|^2$ *and where* $\mu(k,f)$ *is a recursive bound explicitly given in the proof.*

Some consequences

Corollary (Metastability for $(\mathrm{mPPA_e})$)

For all $k \in \mathbb{N}$ and monotone function $f : \mathbb{N} \to \mathbb{N}$,

$$\exists n \leq \mu(4(k+1)^2 - 1, f)\, \forall i, j \in [n, f(n)] \left(\|y_i - y_j\| \leq \frac{1}{k+1} \right).$$

Some consequences

Corollary (Metastability for $(\mathrm{mPPA_e})$)

For all $k \in \mathbb{N}$ and monotone function $f : \mathbb{N} \to \mathbb{N}$,

$$\exists n \leq \mu(4(k+1)^2 - 1, f)\, \forall i, j \in [n, f(n)] \left(\|y_i - y_j\| \leq \frac{1}{k+1} \right).$$

Corollary ("Metastable" version of Asymptotic regularity)

For all $k \in \mathbb{N}$ and monotone function $f : \mathbb{N} \to \mathbb{N}$, we have

(*i*) $\exists n \leq \widetilde{\mu}(k, f)\, \forall i \in [n, f(n)] \left(\|J_i(y_i) - y_i\| \leq \frac{1}{k+1} \right)$,

(*ii*) $\exists n \leq \widetilde{\mu}(2k+1, f)\, \forall i \in [n, f(n)] \left(\|J(y_i) - y_i\| \leq \frac{1}{k+1} \right)$,

where
$\widetilde{\mu}(k, f) := \max\{\mu(16c^2(k+1)^2 - 1, \check{f} + 1), \ell(4cD(k+1) - 1)\}$
and $\check{f}(m) := f(\max\{m, \ell(4cD(k+1) - 1)\})$.

Remarks

- We only required a weaker version of the metric projection principle, replacing the projection point by suitable approximations

Remarks

- We only required a weaker version of the metric projection principle, replacing the projection point by suitable approximations
- We were able to bypass the sequential weak compactness argument

Remarks

- We only required a weaker version of the metric projection principle, replacing the projection point by suitable approximations
- We were able to bypass the sequential weak compactness argument
- The original proof has a discussion by cases which, in our case, imposes a discussion by cases for each approximation to the projection point

Remarks

- We only required a weaker version of the metric projection principle, replacing the projection point by suitable approximations
- We were able to bypass the sequential weak compactness argument
- The original proof has a discussion by cases which, in our case, imposes a discussion by cases for each approximation to the projection point
- The conditions allow for $\gamma_n \equiv 0$, in which case (mPPA) reduces to (HPPA) and we can obtain a quantitative version (under the weakest conditions but with slightly more complicated bounds)

Outline

The T-KM iteration

$$\boxed{x_{n+1} = \beta_n x_n + \lambda_n \big(T(\beta_n x_n) - \beta_n x_n \big)} \qquad \text{(T-KM)}$$

where

- $x_0 \in H$
- $(\beta_n), (\lambda_n) \subset (0, 1]$
- $T : H \to H$ is a nonexpansive mapping

Boț, Csetnek and Meier obtained a strong convergence result for (T-KM) and for similar modifications of the forward-backward and the Douglas-Rachford algorithms.

Boț, Csetnek and Meier's Theorem

$$\boxed{x_{n+1} = \beta_n x_n + \lambda_n(T(\beta_n x_n) - \beta_n x_n)} \qquad \text{(T-KM)}$$

Theorem

Let $(\beta_n), (\lambda_n) \subset (0,1]$ *be real sequences satisfying:*

(i) $\lim \beta_n = 1$,

(ii) $\sum_{n\geq 0}(1-\beta_n) = \infty$,

(iii) $\sum_{n\geq 1} |\beta_n - \beta_{n-1}| < \infty$,

(iv) $\liminf \lambda_n > 0$,

(v) $\sum_{n\geq 1} |\lambda_n - \lambda_{n-1}| < \infty$.

Consider the iterative scheme (T-KM) *with an arbitrary starting point* $x_0 \in H$ *and a nonexpansive mapping* $T : H \to H$ *such that* $\mathrm{Fix}\, T \neq \emptyset$ *. Then* (x_n) *converges strongly to* $\mathrm{proj}_{\mathrm{Fix}\, T}(0)$.

Quantitative conditions

(K_1) $\forall n \in \mathbb{N}(\beta_n \geq \frac{1}{h(n)})$

(K_2) $\forall k \in \mathbb{N}\, \forall n \geq b(k)(|1-\beta_n| \leq \frac{1}{k+1})$

(K_3) $\forall k \in \mathbb{N} \left(\sum_{i=1}^{D(k)}(1-\beta_i) \geq k\right)$

(K_4) $\forall k \in \mathbb{N}\, \forall n \in \mathbb{N} \left(\sum_{i=B(k)+1}^{B(k)+n} |\beta_i - \beta_{i-1}| \leq \frac{1}{k+1}\right)$

(K_5) $\forall n \in \mathbb{N}(\lambda_n \geq \frac{1}{\ell})$

(K_6) $\forall k \in \mathbb{N}\, \forall n \in \mathbb{N} \left(\sum_{i=L(k)+1}^{L(k)+n} |\lambda_i - \lambda_{i-1}| \leq \frac{1}{k+1}\right)$

Boț, Csetnek and Meier's Theorem (Quantitative version)

$$\boxed{x_{n+1} = \beta_n x_n + \lambda_n (T(\beta_n x_n) - \beta_n x_n)} \qquad \text{(T-KM)}$$

Theorem

Let (z_n) *be generated by* (T-KM). *Assume that the conditions* $(K_1) - (K_6)$ *hold. Then for all* $k \in \mathbb{N}$ *and monotone function* $f : \mathbb{N} \to \mathbb{N}$

$$\exists n \leq \mu(k, f) \forall i, j \in [n, f(n)] \left(\|x_i - x_j\| \leq \frac{1}{k+1} \right)$$

Remarks

- The proof of metastability does not require the projection argument nor sequential weak compactness

Remarks

- The proof of metastability does not require the projection argument nor sequential weak compactness
- In the presence of the projection point, one can follow the arguments of the proof and show that our theorems are indeed quantitative versions of the original theorems

References I

R. Boţ, E. Csetnek and, D. Meier *Inducing strong convergence into the asymptotic behaviour of proximal splitting algorithms in Hilbert spaces*, **Optim. Methods Softw. 34(3), 489–514 (2019).**

B. Dinis and P. Pinto, *Proof mining of the proximal point algorithm with multi-parameters*, **(submitted)** arXiv:1906.09129v2.

B. Dinis and P. Pinto, *Quantitative results on the multi-parameters proximal point algorithm*. **(To appear in Journal of Convex Analysis)**

B. Dinis and P. Pinto, *On the convergence of algorithms with Tikhonov regularization terms*. **Optimization Letters** (2020). https://doi.org/10.1007/s11590-020-01635-7

F. Ferreira and P. Engrácia, *Bounded functional interpretation with an abstract type*, **(submitted)**

F. Ferreira and P. Oliva, *Bounded functional interpretation*, **Annals of Pure and Applied Logic**, vol. 135 (2005), pp. 73–112.

References II

- F. Ferreira, L. Leustean and P. Pinto, *On the removal of weak compactness arguments in proof mining*, **Advances in Mathematics**, vol 354, 106728 (2019).
- U. Kohlenbach, **Applied Proof Theory: Proof Interpretations and their Use in Mathematics**, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2008.
- L. Leustean and P. Pinto, *Quantitive results on Halpern type proximal point algorithms*, **(submitted)**.
- F. Wang and H. Cui, *On the contraction-proximal point algorithms with multi-parameters*, **Journal of Global Optimization**, vol. 54 (2012), no. 3, pp. :485–491.
- Y. Yao and A. Noor, *On convergence criteria of generalized proximal point algorithms*, **Journal of Computational and Applied Mathematics**, vol. 217 (2008), no. 1, pp. 46–55.

Thank you!