Quantitative results for differences of maximal monotone operators

Nicholas Pischke

Department of Mathematics
TU Darmstadt

New Frontiers in Proofs and Computation
IASM, Hangzhou, China
September 17, 2021

Topics from (Non-)Convex Optimization

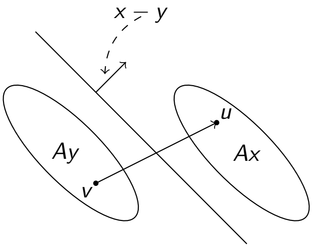
Maximal monotone operators

Throughout: X is a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$.

$A : X \to 2^X$ is monotone if

$$\forall (x, u), (y, v) \in \operatorname{gra} A \left(\langle x - y, u - v \rangle \geq 0 \right)$$

where $\operatorname{gra} A$ is the graph of A.



A is maximally monotone if it additionally is not strictly contained in any monotone operator.

Resolvents

The main tool for studying monotone operators: the resolvent

$$J_\gamma^A = (Id + \gamma A)^{-1} \quad \text{(for } \gamma > 0\text{)}.$$

For a monotone operator: single-valued and firmly nonexpansive (on its domain).

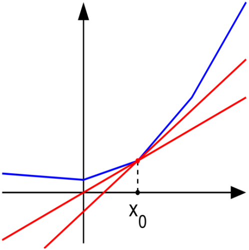
$\rightsquigarrow$ A is maximal if, and only if J_γ^A is total.

Key point for zeros: $\mathrm{zer} A = \mathrm{fix} J_\gamma^A$ for any $\gamma > 0$!

Convex functions and subgradients

The prime example: recall the *subgradient* of a convex function $\varphi : \mathbb{R}^N \to \mathbb{R}$!

$$\partial\varphi(x) := \{z \in \mathbb{R}^N \mid \varphi(y) \geq \varphi(x) + \langle y - x, z\rangle \text{ for all } y \in \mathbb{R}^N\}.$$



$\rightsquigarrow$ $\partial\varphi(x)$ is maximally monotone and

$$0 \in \partial\varphi(x) \text{ iff } x \text{ minimizes } \varphi.$$

Differences of monotone operators

We are interested in algorithms for the following problem:

Problem

Let T, S be maximally monotone operators. Find an x with

$$0 \in (T - S)(x)$$

where $(T - S)(x) = \{y - z \mid y \in Tx \text{ and } z \in Sx\}$.

Again a prime example: DC programming, i.e. finding critical points of $f - g$ for f, g convex.

A particular algorithm

For this, given initial data x_0, consider the sequence

$$x_{n+1} := J^S_{\mu_n}(x_n + \mu_n T_{\lambda_n} x_n)$$

with parameters μ_n, λ_n where

$$T_{\lambda_n}(x) = \frac{x - J^T_{\lambda_n} x}{\lambda_n}$$

is the Yosida approximate of T.

$\rightsquigarrow$ defined by A. Moudafi in 2015.

A particular algorithm

Intuition:

1. $0 \in (T - S)(x)$ if, and only if $Tx \cap Sx \neq \emptyset$.
2. Note $T_\lambda x \to y \in Tx$ for $\lambda \to 0$.
3. Move to the regularized problem

$$\text{find } x_\lambda \in X \text{ with } T_\lambda(x_\lambda) \in S(x_\lambda).$$

4. Equivalent to

$$\text{find } x_\lambda \in X \text{ with } x_\lambda = J^S_\mu(x_\lambda + \mu T_\lambda(x_\lambda)).$$

5. Akin to the Proximal Point Algorithm:

$$x_{n+1} := J^S_{\mu_n}(x_n + \mu_n T_{\lambda_n} x_n).$$

A particular algorithm

Theorem (Moudafi (2015))

Let T, S be maximally monotone on a finite dimensional Hilbert space X such that $\mathrm{zer}(T - S) \neq \emptyset$, $\mathrm{Dom} S \subseteq \mathrm{Dom}\, T$ and T is bounded on bounded sets , i.e.

$$T(B_r(0)) = \bigcup_{x \in B_r(0)} Tx \text{ is bounded for any } r > 0,$$

as well as

1. $\lim_{n\to\infty} \lambda_n = 0$,
2. $\sum_{n=0}^{\infty} \frac{\mu_n}{\lambda_n} < \infty$,
3. $\lim_{n\to\infty} \|x_n - x_{n+1}\| / \mu_n = 0$.

Then (x_n) converges to a point $x^* \in \mathrm{zer}(T - S)$.

Proof Mining

Proof mining

Applied part of mathematical logic, going back to Georg Kreisel's work in the 50s.

Proof Interpretations (Dialectica, Negative Translation, etc.) are applied to theorems of ordinary mathematics to extract uniform bounds (or witnesses), by analyzing a concrete proof.
$\rightsquigarrow$ Metatheorems on Proof Mining.

In our context:

$$\text{convergence statements} \Rightarrow \text{rates of metastability.}$$

Metastability

Statement of Cauchyness for a sequence (x_n) in some metric space (X, d):

$$\forall k \in \mathbb{N} \exists N \in \mathbb{N} \forall n, m \geq N \left(d(x_m, x_n) < \frac{1}{k+1} \right).$$

Generally, one can not expect a bound on '$\exists N \in \mathbb{N}$'.

However, one can expect a bound on '$\exists N \in \mathbb{N}$' in the following statement:

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

Such a bound is called a *rate of metastability* (after Tao).
$\rightsquigarrow$ Guaranteed (in general contexts) by the Metatheorems on Proof Mining.

A Quantitative Analysis

(Quasi-)Fejér monotonicity

The main exploited property for the approximations (x_n) is quasi-Fejér monotonicity.

Let (X, d) be a metric space, $F \subseteq X$ nonempty and (x_n) be a sequence in X.

Definition

(x_n) is called

1. Fejér monotone w.r.t. F, if for all $n \in \mathbb{N}$, all $p \in F$:

$$d(x_{n+1}, p) \leq d(x_n, p).$$

2. quasi-Fejér monotone w.r.t. F, if for all $n \in \mathbb{N}$, all $p \in F$:

$$d(x_{n+1}, p) \leq d(x_n, p) + \varepsilon_n.$$

   Here: (ε_n) such that $\sum_n \varepsilon_n < \infty$.

(Quasi-)Fejér monotonicity

The main exploited property for the approximations (x_n) is quasi-Fejér monotonicity.

Let (X, d) be a metric space, $F \subseteq X$ nonempty and (x_n) be a sequence in X.

Definition

(x_n) is called

1. Fejér monotone w.r.t. F, if for all $n \in \mathbb{N}$, all $p \in F$:

$$H(d(x_{n+1}, p)) \leq G(d(x_n, p)).$$

2. quasi-Fejér monotone w.r.t. F, if for all $n \in \mathbb{N}$, all $p \in F$:

$$H(d(x_{n+1}, p)) \leq G(d(x_n, p)) + \varepsilon_n.$$

Here: (ε_n) such that $\sum_n \varepsilon_n < \infty$.

(Quasi-)Fejér monotonicity

Lemma

Let (X, d) be boundedly compact, (x_n) bounded and let $F \subseteq X$ be nonempty. Suppose that

1. *(x_n) is quasi-Fejér monotone w.r.t. F,*
2. *$F = \bigcap_{k \in \mathbb{N}} AF_k$ for closed sets $AF_k \subseteq X$ with $AF_k \supseteq AF_{k+1}$,*
3. *$\forall k, n \in \mathbb{N} \exists N \geq n \, (x_N \in AF_k)$ ($\liminf$-property),*

then (x_n) converges to a point $x \in F$.

By introducing quantitative versions of (1) and (3), Kohlenbach/Leuștean/Nicolae (CCM 2018) obtained rates of metastability for the convergence of (x_n).

(Quasi-)Fejér monotonicity

This could apply here: (x_n) is quasi-Fejér monotone w.r.t. $\mathrm{zer}(T-S)$.

Moudafi proves:

Lemma

For $x^* \in \mathrm{zer}(T-S)$ and $y^* \in T(x^*) \cap S(x^*)$:

$$\|x_{n+1} - x^*\| \leq \left(1 + \frac{\mu_n}{\lambda_n}\right) \|x_n - x^*\| + \mu_n(\|T^\circ x^*\| + \|y^*\|).$$

Here: $T^\circ x^*$ is the element of minimal norm in Tx^*.

$\rightsquigarrow$ need to assume a uniform bound on $\|T^\circ x^*\| + \|y^*\|$.

(Quasi-)Fejér monotonicity

Suitable AF_k for $\mathrm{zer}(T - S)$?

Write $\Gamma = \mathrm{zer}(T - S)$ and set

$$\Gamma_k := \left\{ x^* \mid \exists y^* \forall i \leq k : \left\| x^* - J^S_{\mu_i}(x^* + \mu_i y^*) \right\| \leq \frac{1}{k+1} \right\}.$$

$\rightsquigarrow$ Simplified! Needs tweaking for $\cap_k \Gamma_k = \Gamma$!

(Quasi-)Fejér monotonicity

What are these quantitative versions of quasi-Fejér monotonicity and the $\liminf$-property?

- uniform quasi-Fejér monotonicity.
- $\liminf$-bounds.

lim inf-bounds

Definition

A (monotone) bound $\Phi(k, n)$ on '$\exists N \in \mathbb{N}$' in

$$\forall k, n \in \mathbb{N} \exists N \geq n \, (x_N \in AF_k)$$

is called a $\liminf$-*bound*.

Lemma

For any n and any i, we have

$$\left\| x_n - J^S_{\mu_i}(x_n + \mu_i T_{\lambda_n} x_n) \right\| \leq \|x_n - x_{n+1}\| + |\mu_n - \mu_i| \frac{\|x_n - x_{n+1}\|}{\mu_n}.$$

$\rightsquigarrow$ to show $x_n \in \Gamma_k$ for given k:

1. $y^* = T_{\lambda_n} x_n$,
2. $\|x_n - x_{n+1}\| / \mu_n \to 0$ by assumption in Moudafi's result,
3. μ_n convergent and thus bounded.

lim inf-bounds

Lemma

Let $C \geq 1$ be an upper bound on both $\mathrm{diam}(\mu_n)$ and (μ_n). Further, let ϕ be monotone s.t.

$$\forall k, n \exists N \in [n; \phi(k,n)] \left(\|x_N - x_{N+1}\| / \mu_N < \frac{1}{k+1} \right).$$

Then the function

$$\Phi(k,n) = \phi\left(\lceil 2C(k+1) \rceil - 1, \max\{n,k\}\right)$$

is a $\liminf$*-bound for (x_n) w.r.t. Γ_k.*

Uniform quasi-Fejér monotonicity

Definition

(x_n) is called *uniformly* quasi-Fejér monotone w.r.t. F and (AF_k) if for all $r, n, m \in \mathbb{N}$:

$$\exists k \in \mathbb{N} \forall p \in AF_k \forall l \leq m$$
$$\left(d(x_{n+l}, p) < d(x_n, p) + \sum_{i=n}^{n+l-1} \varepsilon_i + \frac{1}{r+1} \right).$$

An upper bound (realizer) $\chi(n, m, r)$ on '$\exists k \in \mathbb{N}$' is called a *modulus for uniform quasi-Fejér monotonicity*.

Uniform quasi-Fejér monotonicity

Definition

(x_n) is called *uniformly* quasi-Fejér monotone w.r.t. F and (AF_k) if for all $r, n, m \in \mathbb{N}$:

$$\exists k \in \mathbb{N} \forall p \in AF_k \forall l \leq m$$
$$\left(H(d(x_{n+l}, p)) < G(d(x_n, p)) + \sum_{i=n}^{n+l-1} \varepsilon_i + \frac{1}{r+1} \right).$$

An upper bound (realizer) $\chi(n, m, r)$ on '$\exists k \in \mathbb{N}$' is called a *modulus for uniform quasi-Fejér monotonicity*.

Uniform quasi-Fejér monotonicity

Lemma

Assume $\mathrm{Dom} S \subseteq \mathrm{Dom} T$*. Let* $M \geq \|T^\circ x^*\|$ *for any* $x^* \in \Gamma_0$*. Further, let* $A \geq \sum_{n=0}^{\infty} \frac{\mu_n}{\lambda_n}$ *and assume that* $\lim_{n\to\infty} \lambda_n = 0$*. Then* (x_n) *is uniformly quasi-Fejér monotone w.r.t.* Γ_k *with modulus* χ*, that is for all* $r, n, m \in \mathbb{N}$*:*

$$\forall x^* \in \Gamma_k \forall l \leq m \Bigg(\|x_{n+l} - x^*\| < e^A \|x_n - x^*\| + 2Me^A \sum_{i=n}^{n+l-1} \mu_i + \frac{1}{r+1} \Bigg)$$

where

$$k = \chi(r, n, m) = \max\{n + m \dot{-} 1, \lceil (r+1) \cdot m \cdot e^A \rceil\}.$$

A rate of metastability

Theorem

Let T, S be maximally monotone on a finite dimensional Hilbert space X with $\mathrm{Dom}S \subseteq \mathrm{Dom}T$. Let $M \geq \|T^\circ x^*\|$ for any $x^* \in \Gamma_0$ be non-zero. Further, let $A \geq \sum_{n=0}^{\infty} \frac{\mu_n}{\lambda_n}$ and assume that $\lim_{n\to\infty} \lambda_n = 0$. Let $C \geq 1$ be an upper bound on both $\mathrm{diam}(\mu_n)$ and (μ_n). Further, let ϕ be monotone s.t.

$$\forall k, n \exists N \in [n; \phi(k,n)] \left(\|x_N - x_{N+1}\| / \mu_N < \frac{1}{k+1} \right)$$

and such that it is monotone w.r.t. k and n.
Let $L \geq \mathrm{diam}(x_n)$ and let ξ be a Cauchy rate for $\sum_n \mu_n < \infty$. Then (x_n) is Cauchy with a rate of metastability

$$\Psi(k,g) = \Psi_0(P, k, g, \Phi, \chi, \tilde{\xi}).$$

A rate of metastability

Theorem (continued)

Here: Ψ_0 *defined by recursion via*

$$\begin{cases} \Psi_0(0,k,g,\Phi,\chi,\tilde{\xi}) = 0 \\ \Psi_0(n+1,k,g,\Phi,\chi,\tilde{\xi}) = \\ \qquad\qquad \Phi(\chi_g^M(\Psi_0(n,k,g,\Phi,\chi,\tilde{\xi}),8k+7,\tilde{\xi}(8k+7)) \end{cases}$$

with $P = \lceil 2\lceil 8e^A(k+1)\rceil\sqrt{d}L\rceil^d + 1$ *where* $d = \mathrm{dim}X$,
$\tilde{\xi}(n) = \xi(\lceil 2Me^A(n+1)\rceil - 1)$ *and*

$$\Phi(k,n) = \phi\left(\lceil 2C(k+1)\rceil - 1, \max\{n,k\}\right)$$

as well as

$$\chi(r,n,m) = \max\{n+m \dot{-} 1, \lceil (r+1)\cdot m\cdot e^A\rceil\},$$

$$\chi_g(n,k) = \chi(n,g(n),k), \quad \chi_g^M(n,k) = \max\{\chi_g(i,k) \mid i \leq n\}.$$

Generalizations

Error terms

We can generalize Moudafi's result: the finitary analysis suggest that it is possible to incorporate error terms into the sequence.

We consider

$$x_{n+1} := J^S_{\beta_n}(x_n + \alpha_n z_n + \beta_n T_{\mu_n} x_n).$$

with the additional condition

$$\sum_n \alpha_n \|z_n\| < \infty.$$

Not artificial: there are generalizations like this in the literature, e.g.

$$x_{n+1} := J^S_{\beta_n}(x_n + \alpha_n(x_n - x_{n-1}) + \beta_n T_{\mu_n} x_n)$$

which adds an *inertia term* emanated from the evolution equation of a heavy ball with friction system.

$\rightsquigarrow$ one can provide a similar analysis and obtain a corresponding rate of metastability.

Moduli of regularity

Due to results from recursion theory, we can not hope (in general) for rates of convergence.

However, adding certain assumptions allows for the construction of these. This has been explored under the name of moduli of regularity in a work by Kohlenbach/López-Acedo/Nicolae (IJM 2019).

Here:

Definition

ϕ is a modulus of regularity for $T - S$ if for all $\varepsilon > 0$ and all x:

$$|\mathrm{dist}(0, (T - S)(x))| < \phi(\varepsilon) \text{ implies } \mathrm{dist}(x, \mathrm{zer}(T - S)) < \varepsilon.$$

Moduli of regularity

Due to results from recursion theory, we can not hope (in general) for rates of convergence.

However, adding certain assumptions allows for the construction of these. This has been explored under the name of moduli of regularity in a work by Kohlenbach/López-Acedo/Nicolae (IJM 2019).

Here:

Definition

ϕ is a modulus of regularity for $T - S$ at z if for all $\varepsilon > 0$ and all $x \in B_r(z)$:

$$|\mathrm{dist}(0, (T - S)(x))| < \phi(\varepsilon) \text{ implies } \mathrm{dist}(x, \mathrm{zer}(T - S)) < \varepsilon.$$

Moduli of regularity

Theorem

Let T, S be maximally monotone on a finite dimensional Hilbert space X with $\mathrm{Dom} S \subseteq \mathrm{Dom}\, T$ and $\mathrm{zer}(T - S) \neq \emptyset$. Let $M \geq \| T^\circ x^* \|$ for any $x^* \in \Gamma_0$. Further, let $A \geq \sum_{n=0}^{\infty} \frac{\mu_n}{\lambda_n}$ and assume that $\lim_{n\to\infty} \lambda_n = 0$. Let $C \geq 1$ be an upper bound on both $\mathrm{diam}(\mu_n)$ and (μ_n). Further, let Φ be s.t.

$$\forall \varepsilon, n \exists N \in [n; \Phi(\varepsilon, n)] \left(\| x_N - x_{N+1} \| / \mu_N < \varepsilon \right)$$

and such that it is monotone w.r.t. ε and n. Let ξ be a Cauchy rate for $\sum_n \mu_n \leq d < \infty$.

Moduli of regularity

Theorem (continued)

Let $b \geq \|x_0 - z\|$ *for some* $z \in \mathrm{zer}(T - S)$ *and suppose* ϕ *is a modulus of regularity for* $T - S$ *at* $B_{e^A b + d}(z)$.

Then (x_n) *is Cauchy with Cauchy rate*

$$\forall \varepsilon > 0 \forall n, m \geq \theta(\varepsilon) := \Phi\left(\frac{\phi\left(\frac{\varepsilon}{4e^A}\right)}{2C}, \tilde{\xi}\left(\frac{\varepsilon}{4}\right)\right)(d(x_n, x_m) < \varepsilon)$$

where $\tilde{\xi}(n) = \xi(\lceil 2Me^A(n+1)\rceil - 1)$.

$\rightsquigarrow$ Can be generalized to the error terms.

At last ...

Thank You!