On linear convergence of fixed-point iterations and application to phase retrieval

Nguyen Hieu Thao

Delft Center for Systems and Control, TU Delft, The Netherlands

WoMBaT 2018, Deakin University, Melbourne, Australia

Motivation

- Application of optimization $\Longrightarrow$ convergence analysis.
- Two key ingredients:
  1. Regularity of individual functions/sets.
  2. Regularity of families of functions/sets.
- Analyze convergence $\Longleftrightarrow$ verify regularity properties.
- New characterizations of regularity $\Longrightarrow$ better understanding of convergence.

Research questions

- Let (x_k) be generated by $x_{k+1} \in Tx_k$, where $T : \mathbb{E} \rightrightarrows \mathbb{E}$.
- Goal: $x_k \to \tilde{x} \in \text{Fix } T$ with linear rate $c \in (0,1)$,

$$\|x_k - \tilde{x}\| \leq \gamma c^k \quad \forall k \in \mathbb{N}.$$

- Research questions:
  1. sufficient conditions?

T averaged $+ (I - \text{Id})$ metrically subregular $\Rightarrow$ linear convergence.

  2. necessary conditions?

linear convergence $\Rightarrow (I - \text{Id})$ metrically subregular.

  3. application to projection algorithms?

Linear convergence with Fejér monotonicity

Theorem (Bauschke-Combettes 2011)

Suppose that (x_k) is Fejér monotone w.r.t. S (convex),

$$\|x_{k+1} - x\| \leq \|x_k - x\| \quad \forall x \in S,\ \forall k \in \mathbb{N},$$

and linearly monotone w.r.t. S with rate $c \in [0,1)$,

$$\mathrm{dist}(x_{k+1}, S) \leq c\,\mathrm{dist}(x_k, S) \quad \forall k \in \mathbb{N}.$$

Then (x_k) converges linearly to some $\tilde{x} \in S$ with rate c.

Example of AP for a line intersecting a circle requires broader approach to linear convergence in nonconvex setting (next slide).

Why almost averaging operators?

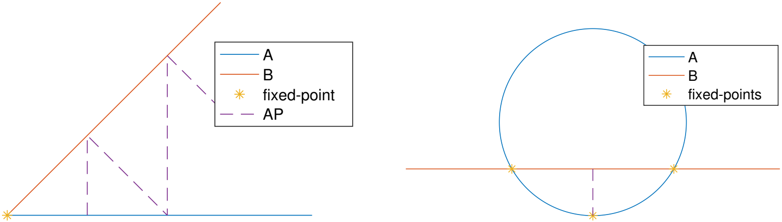


Figure: Convex vs nonconvex AP

- Left: convexity, Fejér monotonicity, averagedness.
- Right: none of those, though convergence!

$\Longrightarrow$ theory of almost averaging operators.

Pointwise almost averaged maps

Definition

$T : \mathbb{E} \rightrightarrows \mathbb{E}$ is pointwise almost averaged (p.a.a.) on U at $y \in U$ with violation ε and averaging constant α if

$$\left\|x^+ - y^+\right\|^2 \leq (1+\varepsilon)\left\|x-y\right\|^2 - \frac{1-\alpha}{\alpha}\left\|(x^+ - x) - (y^+ - y)\right\|^2,$$

for all $x \in U$, $x^+ \in Tx$ and $y^+ \in Ty$.
If the property holds for all $y \in U$, we say almost averaged on U.

For $\alpha = 1$ and $\alpha = 1/2$, one can talk about almost nonexpansive and almost firmly nonexpansive, respectively.

A criterion for linear convergence

Recall: a mapping $F : \mathbb{X} \rightrightarrows \mathbb{Y}$ is *metrically subregular* at $\bar{x}$ for $\bar{y} \in F(\bar{x})$ with constant $\kappa \geq 0$ if

$$\text{dist}\left(x, F^{-1}(\bar{y})\right) \leq \kappa \,\text{dist}\left(\bar{y}, F(x)\right) \quad \forall x \text{ near } \bar{x}.$$

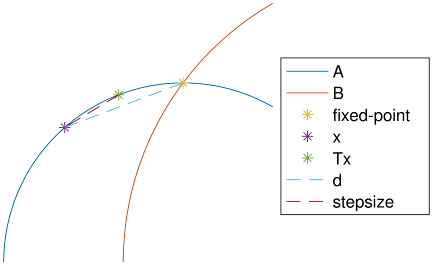
Theorem

Let $T : \mathbb{E} \rightrightarrows \mathbb{E}$ with $\text{Fix } T$ closed and nonempty. Suppose

1. *T is pointwise almost averaged on $\text{Fix } T + \delta\mathbb{B}$ at all $y \in \text{Fix } T$ with violation ε and averaging constant α,*
2. *the mapping $F := T - \text{Id}$ is metrically subregular at every point $x \in \text{Fix } T$ for 0 with constant $\kappa < \sqrt{\frac{1-\alpha}{\varepsilon\alpha}}$.*

Then the iteration $x_{k+1} \in T(x_k)$ with x_0 close to $\text{Fix } T$ converges linearly to a fixed point of T with rate $c \leq \sqrt{1 + \varepsilon - \frac{1-\alpha}{\alpha\kappa^2}} < 1$.

A criterion for linear convergence



Theorem

1. T *is* *pointwise almost averaging.*
2. $T - \mathrm{Id}$ *is* *metrically subregular:* $d \leq \gamma\, \mathrm{stepsize}$.

$$\Longrightarrow \mathrm{dist}\,(Tx, \mathrm{Fix}\ T) \leq c\, \mathrm{dist}\,(x, \mathrm{Fix}\ T)\,.$$

A way to linear convergence (for necessary conditions)

Let $T : \mathbb{E} \rightrightarrows \mathbb{E}$ with Fix T closed and nonempty.

Proposition

Suppose that T is pointwise almost averaged on $(\text{Fix } T + d_0\mathbb{B})$ at all point $y \in \text{Fix } T$, where $d_0 := \text{dist}(x_0, \text{Fix } T)$, and $x_{k+1} \in Tx_k$ is linearly monotone w.r.t. Fix T with rate $c \in [0, 1)$. Then (x_k) converges linearly to a fixed point of T with rate c.

Recall: (x_k) is linearly monotone w.r.t. Fix T with rate $c \in [0, 1)$ if

$$\text{dist}(x_{k+1}, \text{Fix } T) \leq c\, \text{dist}(x_k, \text{Fix } T) \quad \forall k \in \mathbb{N}.$$

Metric subregularity is necessary for linear monotonicity

Let $T : \mathbb{E} \rightrightarrows \mathbb{E}$ with $\mathrm{Fix}\ T$ closed and nonempty.

Theorem

If for all x_0 close to $\mathrm{Fix}\ T$, all iteration $x_{k+1} \in Tx_k$ is linearly monotone with respect to $\mathrm{Fix}\ T$ with rate $c \in (0,1)$, then the mapping $F := T - \mathrm{Id}$ is metrically subregular at every fixed point of T for 0 with constant $\kappa \leq \frac{1}{1-c}$.

Corollary

Under the assumption of pointwise almost averagedness on T,

$$\text{linear monotonicity} \iff \text{metric subregularity}.$$

Metric subregularity vs subtransversality (general)

- Metric subregularity of F at $(\bar{x}, \bar{y})$

$$\operatorname{dist}\left(x, F^{-1}(\bar{y})\right) \leq \kappa \operatorname{dist}\left(\bar{y}, F(x)\right) \quad \forall x \text{ near } \bar{x}.$$

- $\{A, B\}$ is *subtransversal* at $\bar{x} \in A \cap B$ with constant κ if

$$\operatorname{dist}\left(x, A \cap B\right) \leq \kappa \max\left\{\operatorname{dist}\left(x, A\right), \operatorname{dist}(x, B)\right\} \quad \forall x \text{ near } \bar{x}.$$

1. Given $\{A, B\}$, construct $F(x) = (A - x) \times (B - x)$.
2. Given $\{A, B\}$, construct $F(x, y) = \{x - y\}$ if $(x, y) \in A \times B$ and $F(x, y) = \emptyset$ otherwise.
3. Given F and $(\bar{x}, \bar{y}) \in \operatorname{gph} F$, construct $A = \operatorname{gph} F$ and $B = X \times \{\bar{y}\}$.

Fact

Metric subregularity of F $\Longleftrightarrow$ subtransversality of $\{A, B\}$.

Metric subregularity vs subtransversality (for AP)

- Metric subregularity of F at $(\bar{x}, \bar{y})$

$$\operatorname{dist}\left(x, F^{-1}(\bar{y})\right) \leq \kappa \operatorname{dist}\left(\bar{y}, F(x)\right) \quad \forall x \text{ near } \bar{x}.$$

- $\{A, B\}$ is *subtransversal* at $\bar{x} \in A \cap B$ with constant κ if

$$\operatorname{dist}\left(x, A \cap B\right) \leq \kappa \max\left\{\operatorname{dist}\left(x, A\right), \operatorname{dist}(x, B)\right\} \;\; \forall x \text{ near } \bar{x}.$$

Fact

Metric subregularity of $P_A P_B - \mathrm{Id} \iff$ *subtransversality of* $\{A, B\}$.

Convexity-like properties yield almost everagedness

- Convexity of A and B yields averagedness of $P_A P_B$.
- Convexity-like properties of A and B yield (pointwise) almost averagedness of $P_A P_B$.

Convex alternating projections

Theorem (Subtransversality $\iff$ linear convergence)

- *(Bauschke-Borwein 1996) If $\{A, B\}$ is subtransversal at $\bar{x} \in A \cap B$ with constant $\kappa > 0$, then any AP iteration (x_k) with x_0 close to $\bar{x}$ is converges linearly to a point in $A \cap B$ with rate $c \leq 1 - 1/\kappa^2$.*
- *If for any x_0 close to $\bar{x}$, the AP iteration converges linearly to some point in $A \cap B$ with rate $c \in [0, 1)$, then $\{A, B\}$ is subtransversal at $\bar{x}$ with constant $\kappa \leq \frac{3-c}{1-c}$.*

- There is the global version of this result.
- linear monotonicity $\Leftrightarrow$ linear convergence $\Leftrightarrow$ subtransversality.

How were necessary conditions results obtained?

Theorem (ideas for necessary conditions)

- *(Drusvyatskiy-Ioffe-Lewis 2015, Theorem 6.2) The property called intrinsic transversality implies subtransversality.*
- *A finer characterization of subtransversality is presented:* $\{A, B\}$ *is subtransversal at* $\bar{x}$ *with constant* κ *if*

$$\text{dist}\,(x, A \cap B) \leq \kappa\, \text{dist}(x, B) \ \ \forall x \in A \text{ near } \bar{x}.$$

How about nonconvex alternating projections?

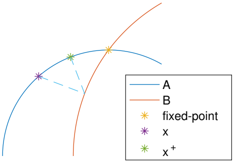
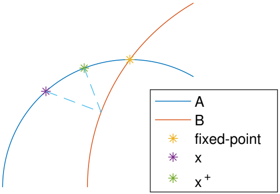


Figure: Convex and consistent AP.

1. Subtransversality is not sufficient.
2. Sufficient conditions: two approaches $\Longrightarrow$ a unified criterion.
3. Necessary conditions $\Longrightarrow$ conjecture on subtransversality.

Sufficient conditions



Two approaches:

1. P.a.a. and subtransversality (Hesse-Luke 2013).
2. Make use of transversality properties (Lewis, Luke, Malick, Bauschke, Phan, Wang, Drusvyatskiy, Ioffe, Lewis, Noll, Rondepierre). The finest criterion is established in Noll-Rondepierre 2015.

A unified criterion

Two approaches:

1. P.a.a. and subtransversality (Hesse-Luke 2013) $\Longrightarrow$ linear monotonicity $\Longrightarrow$ linear convergence.
2. Make use of transversality properties (Noll-Rondepierre 2015) $\Longrightarrow$ linear extendibility $\Longrightarrow$ linear convergence.

$$\Longrightarrow \text{ a unified criterion: } \begin{cases} \text{convexity-like property of one set} \\ \text{metric subregularity of } P_A P_B - \mathrm{Id}\,. \end{cases}$$

Linear monotonicity $\Longrightarrow$ subtransversality

- (x_k) is linearly monotone w.r.t. $A \cap B$ with rate c if

$$\text{dist}(x_{k+1}, A \cap B) \leq c\, \text{dist}(x_k, A \cap B) \quad \forall k \in \mathbb{N}.$$

- $\{A, B\}$ is *subtransversal* at $\bar{x} \in A \cap B$ with constant κ if

$$\text{dist}\,(x, A \cap B) \leq \kappa \max \{\text{dist}\,(x, A)\,, \text{dist}(x, B)\} \quad \forall x \text{ near } \bar{x}.$$

Theorem

Suppose that, for any x_0 close to $\bar{x}$, every AP iteration (x_k) is linearly monotone w.r.t. $A \cap B$ at rate $c \in [0, 1)$. Then $\{A, B\}$ is subtransversal at $\bar{x}$ with constant $\kappa \leq \frac{5-c}{1-c}$.

Linear extendability

Let (z_k) be the *joining sequence* of AP iteration $x_{k+1} \in P_A P_B x_k$,

$$z_{2k} = x_k \text{ and } z_{2k+1} = b_k,$$

where $b_k \in P_B x_k$ such that $x_{k+1} \in P_A b_k$ for all $k \in \mathbb{N}$.

Definition

(x_k) is linearly extendable with rate $c \in [0,1)$ if $\forall k = 1, 2, \ldots,$

$$\begin{aligned} \|z_{k+1} - z_k\| &\leq \|z_k - z_{k-1}\|, \\ \|z_{2k+2} - z_{2k+1}\| &\leq c\|z_{2k+1} - z_{2k}\|. \end{aligned}$$

Fact

Linear extendability $\Longrightarrow$ *linear convergence.*

Linear extendability $\Longrightarrow$ subtransversality

- (x_k) is linearly extendable with rate c if $\forall k = 1, 2, \ldots,$

$$\begin{aligned} \|z_{k+1} - z_k\| &\leq \|z_k - z_{k-1}\|, \\ \|z_{2k+2} - z_{2k+1}\| &\leq c\|z_{2k+1} - z_{2k}\|. \end{aligned}$$

- $\{A, B\}$ is *subtransversal* at $\bar{x} \in A \cap B$ with constant κ if

$$\operatorname{dist}(x, A \cap B) \leq \kappa \max\{\operatorname{dist}(x, A), \operatorname{dist}(x, B)\} \quad \forall x \text{ near } \bar{x}.$$

Theorem

Suppose that, for all x_0 close to $\bar{x}$, every AP iteration (x_k) is linearly extendable with rate $c \in [0, 1)$. Then $\{A, B\}$ is subtransversal at $\bar{x}$ with constant $\kappa \leq \frac{5-c}{1-c}$.

Is subtransversality necessary for linear convergence?

- All known criteria for *linear convergence* of AP follow:

  linear monotonicity/*linear extendability* $\Longrightarrow$ *linear convergence*.

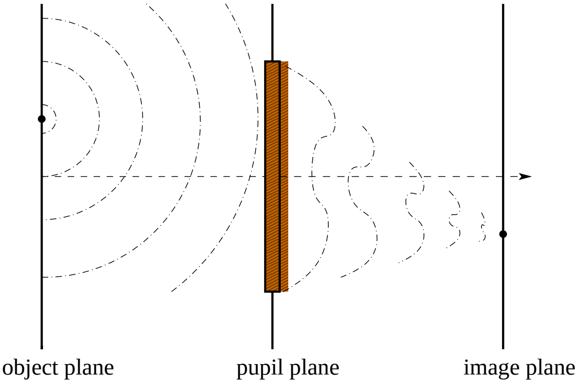
- *Linear monotonicity*/*linear extendability* $\Longrightarrow$ *subtransversality*.

Observation: *subtransversality* has appeared in all known criteria.

Conjecture

Subtransversality is necessary for linear convergence of AP.

Image formulation - the Fraunhofer diffraction model



$$I_k \propto |F \circ D_k(a.\exp(j\Phi))|^2 \propto |F\,(a.\exp(j(\Phi + \psi_k)))|^2\,.$$

Phase retrieval problem

Write $\hat{x} = a.\exp(j\Phi)$ and define the propagation matrix:

$$M = \frac{1}{\sqrt{m}} \begin{pmatrix} FD_1 \\ FD_2 \\ \cdots \\ FD_m \end{pmatrix}.$$

Phase retrieval is to solve:

$$|Mx|^2 = I + w, \quad x \in \mathbb{C}^n,$$

where $I = (I_1^T, I_2^T, \cdots, I_m^T)^T$ and $w \in \mathbb{R}^N$ represents noise.

Feasibility models

Define

$$\Omega_k := \left\{x \in \mathbb{C}^n : |FD_k(x)|^2 = I_k\right\} \quad (1 \leq k \leq m).$$

A feasibility model is:

$$\text{find} \quad x \in \bigcap_{k=0}^{m} \Omega_k, \tag{1}$$

where $\Omega_0 := \chi$ captures *a priori* constraint of the solutions.

Feasibility models (cont.)

Define

$$\Omega := \Omega_1 \times \Omega_2 \times \cdots \times \Omega_m,\; D := \{(x, x, \ldots, x) \in \mathbb{C}^{nm} \mid x \in \chi\}.$$

A feasibility model in the Cartesian product space is:

$$\text{find} \quad u \in \Omega \cap D. \tag{2}$$

The counterpart of (2) in the Fourier domain is:

$$\text{find} \quad y \in A \cap B, \tag{3}$$

where

$$A := M(\chi) \text{ and } B := \{y \in \mathbb{C}^N \mid |y|^2 = I\}.$$

Proposition

Let $\hat{x} \in \mathbb{C}^n$ and $\hat{y} = M\hat{x}$. Then

$$\hat{x} \textit{ solves } (1) \Leftrightarrow [\hat{x}]_m \textit{ solves } (2) \Leftrightarrow \hat{y} \textit{ solves } (3).$$

Calculation of projectors

Recall the feasibility model:

$$\text{find} \quad y \in A \cap B,$$

where

$$A := M(\chi) \text{ and } B := \{y \in \mathbb{C}^N \mid |y|^2 = I\}.$$

Two projectors:

- P_A - dependent on P_χ: $\quad P_A(y) = MP_\chi(M^* y).$
- P_B - rescale elementwise: $\quad P_B(y) = b \odot \frac{y}{|y|}.$

About convergence?

1. The sets are prox-regular $\Longleftarrow$ almost averagedness.
2. Randomly chosen phase diversity patterns (the only chosen input) lead to subtransversality almost surely.

As a result, linear convergence is almost surely.

References I

- Bauschke-Borwein. On projection algorithms for solving convex feasibility problems. *SIAM Rev.* **3**, 367–426 (1996)
- Bauschke-Combettes. Convex Analysis and Monotone Operator Theory in Hilbert Spaces. *Springer* (2011)
- Bauschke-Luke-Phan-Wang. Restricted normal cones and the method of alternating projections: Applications. Set-Valued Var. Anal. **21**, 475–501 (2013)
- Drusvyatskiy-Ioffe-Lewis. Transversality and alternating projections for nonconvex sets. *Found. Comput. Math.* **15**, 1637–1651 (2015)
- Hesse-Luke. Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems. *SIAM J. Optim.* **23**, 2397–2419 (2013)

References II

- Kruger-Luke-Th. Set regularities and feasibility problems. *Math. Program. B* **168**(1):1–33 (2018)
- Kruger-Luke-Th. About subtransversality of collections of sets. *Set-Valued Var. Anal.* **25**(4):701–729 (2017)
- Kruger-Th. Regularity of collections of sets and convergence of inexact alternating projections. *J. Convex Anal.* **23**(3):823–847 (2016)
- Lewis-Luke-Malick. Local linear convergence of alternating and averaged projections. *Found. Comput. Math.* **9**, 485–513 (2009)
- Lewis-Malick. Alternating projections on manifolds. *Math. Oper. Res.* **33**, 216–234 (2008)

References III

- Luke-Burke-Lyon, Optical wavefront reconstruction: theory and numerical methods. *SIAM Rev.* **44**, 169–224 (2002)
- Luke-Th-Tam. Quantitative convergence analysis of iterated expansive set-valued mappings. *Math. Oper. Res.*, to appear.
- Noll-Rondepierre. On local convergence of the method of alternating projections. *Found. Comput. Math.* **16**, 425–455 (2016)
- Pierra. Decomposition through formalization in a product space. *Math. Programming* **28**, 96–115 (1984)
- Th. Algorithms for Structured Nonconvex Optimization: Theory and Practice. PhD Thesis, Gottingen University (2017)

Thank you for your attention!