CONVERGENCE OF PROXIMAL SPLITTING ALGORITHMS IN CAT(κ) SPACES AND BEYOND

Russell Luke

Universität Göttingen

Workshop on Nonlinear Functional Analysis and Its Applications in memory of Professor Ronald E. Bruck
Haifa, April 4-6, 2022
Outline I

Motivation

Metric Spaces

General Regularity Notions

Quantitative Convergence

Example: metric proximal mappings

References
Motivation

Let \((G, d)\) be a complete \text{CAT}(\kappa) space \((\kappa > 0)\), \(f_i : G \to G\) be proper lsc smooth functions for \(i = 1, 2, \ldots N\), and \(g : G \to \mathbb{R} \cap \{+\infty\}\) be proper, lower semicontinuous with \((\text{argmin } g) \cap (\bigcap_{j=1}^{N} \text{argmin } f_j) \neq \emptyset\).

Consider the problem

\[
\inf_{x \in G} g(x) + \sum_{i=1}^{N} f_i(x).
\]
Motivation

Proximal splitting

Given \( g := 0, f_1 \ldots, f_N \) and \( \lambda_i > 0 (i = 1, 2, \ldots, N) \). Choose \( x_0 \in G \).

For \( k = 0, 1, 2, \ldots \)

\[
x_{k+1} = T x_k := \left( \text{prox}_{f_N, \lambda_N} \circ \cdots \circ \text{prox}_{f_2, \lambda_2} \circ \text{prox}_{f_1, \lambda_1} \right) (x_k)
\]

Special case:

\[
\inf_{x \in C} f(x) = \inf_{x \in G} \nu_C(x) + f(x)
\]

Metric Projected Gradient

Choose \( x_0 \in G \).

For \( k = 0, 1, 2, \ldots \)

\[
x_{k+1} = T_{PG}(x_k) := P_C \left( (1 - \beta) \text{Id} \oplus \beta \text{prox}_{f, \lambda} \right) (x_k)
\]

(we interpret \( (1 - \beta) \text{Id} \oplus \beta \text{prox}_{f, \lambda} \) as the direction of steepest descent of the Moreau-Yosida envelope of \( f \) with steplength \( \beta \))
Main Result

Convergence of proximal algorithms in CAT(κ) spaces

Let $f$ and $g$ be proper, lsc and “nice”, and let $T : D \to D$ where $D \subset G$ denote one of the following:

(i) $T := \text{prox}_{f_N, \lambda_N} \circ \text{prox}_{f_{N-1}, \lambda_{N-1}} \circ \cdots \circ \text{prox}_{f_1, \lambda_1}$;

(ii) $T := \beta \text{prox}_{g, \lambda} \oplus (1 - \beta) \text{Id}$;

(iii) $T := \text{prox}_{f_1, \lambda_1} \circ (\beta \text{prox}_{g, \lambda_2} \oplus (1 - \beta) \text{Id})$;

(iv) $T := P_C \circ (\beta \text{prox}_{g, \lambda_1} \oplus (1 - \beta) \text{Id})$.

If $T$ satisfies $\text{Fix } T \neq \emptyset$ and is metrically subregular at $\text{Fix } T$

$$d(x, \text{Fix } T \cap D) \leq \mu d(x, Tx), \quad \forall x \in D \subset G,$$

with constant $\mu$, then the fixed point sequence initialized from any starting point close enough to $\text{Fix } T$ is at least linearly convergent to a point in $\text{Fix } T$. 
Main Result

Convergence of proximal algorithms in CAT($\kappa$) spaces

Remarks:

▶ Convergence (without rates) of most of these methods in the convex setting has been known on CAT(0) spaces at least since Bačák (2014).

▶ The main problem in going from CAT(0) to CAT($\kappa$) is that prox mappings of convex functions are expansive and firmly nonexpansive ≠ nonexpansive

▶ The main innovation in extending to CAT($\kappa$) spaces and nonconvex functions is pointwise almost $\alpha$-firmly nonexpansive mappings and tricky facts about metric spaces - the symmetric perpendicular property

▶ The main tool for quantitative convergence estimates is metric subregularity
A selection of important building blocks I

- Convergence of fixed point iterations of (firmly) nonexpansive mappings
A selection of important building blocks II


- Metric (sub)regularity

- Metric space theory
Outline I

Motivation

Metric Spaces

General Regularity Notions

Quantitative Convergence

Example: metric proximal mappings

References
p-uniformly convex spaces

For \( p \in (1, \infty) \), a metric space \((G, d)\) is \( p \)-uniformly convex with constant \( c > 0 \) whenever it is a geodesic space, and

\[
(\forall t \in [0, 1])(\forall x, y, z \in G)
\]

\[
d(z, (1 - t)x \oplus ty)^p \leq (1 - t)d(z, x)^p + td(z, y)^p - \frac{c}{2} t(1 - t)d(x, y)^p
\]

Examples

- \( L^p \) spaces
- \( \text{CAT}(0) \) spaces (\( p = c = 2 \))
- Hadamard spaces (complete \( \text{CAT}(0) \) spaces)
- Hilbert spaces (linear Hadamard spaces)
- \( \text{CAT}(\kappa) \) spaces: for all \( \delta \in (0, \pi/(4\sqrt{\kappa}) \) the subspace \( (B_\delta(x), d) \) is a \( 2 \)-uniformly convex space with constant

\[
c_\delta = 4\delta \sqrt{\kappa} \tan \left( \frac{\pi}{2} - 2\delta \sqrt{\kappa} \right) \] [Ohta, 2007]
Symmetric perpendicular spaces

Let \((G, d)\) be a geodesic space and \(\gamma\) and \(\eta\) be two geodesics through \(p\). Then \(\gamma\) is said to be perpendicular to \(\eta\) at point \(p\) denoted by \(\gamma \perp_p \eta\) if

\[
d(x, p) \leq d(x, y) \quad \forall x \in \gamma, y \in \eta
\]

A space is said to be symmetric perpendicular if for all geodesics \(\gamma\) and \(\eta\) with common point \(p\) we have

\[
\gamma \perp_p \eta \iff \eta \perp_p \gamma
\]

[Kuwae, 2014]

Any CAT(\(\kappa\)) space \((G, d)\) with \(\text{diam}(G) < \frac{\pi}{2\sqrt{\kappa}}\) is symmetric perpendicular. In particular, CAT(0) spaces are automatically symmetric perpendicular.
Outline I

Motivation

Metric Spaces

General Regularity Notions

Quantitative Convergence

Example: metric proximal mappings

References
Almost $\alpha$-firmly nonexpansive mappings

[L.-Thao-Tam, 2018][Bërdëllima-Lauster-L., 2022]

$(G, d)$ is $p$-uniformly convex with constant $c$.  

$\blacktriangleright$ $T : G \to G$ is pointwise almost nonexpansive (ane) at $y \in D \subset G$ on $D$ with violation $\epsilon \geq 0$ whenever

$$\exists \epsilon \geq 0 : \quad d(Tx, Ty)^p \leq (1 + \epsilon)d(x, y)^p \quad \forall x \in D$$

$\blacktriangleright$ $T : G \to G$ is quasi strictly nonexpansive whenever

$$d(Tx, \bar{x}) < d(x, \bar{x}) \quad \forall x \in G \setminus \text{Fix } T, \forall \bar{x} \in \text{Fix } T$$

$\blacktriangleright$ $T : G \to G$ is pointwise almost $\alpha$-firmly nonexpansive (a$\alpha$-fne) at $y \in D \subset G$ on $D$ with violation $\epsilon > 0$ whenever

$$\exists \alpha \in (0, 1), \epsilon \geq 0 : \quad d(Tx, Ty)^p \leq (1 + \epsilon)d(x, y)^p - \frac{1-\alpha}{\alpha} \psi_T^{(p, c)}(x, y)$$

$$\psi_T^{(p, c)}(x, y) := \frac{c}{2} \left( d(Tx, x)^p + d(Ty, y)^p + d(Tx, Ty)^p + d(x, y)^p - d(Tx, y)^p - d(x, Ty)^p \right)$$
Transport discrepancy

The transport discrepancy

\[
\psi_{T}^{(p,c)}(x, y) := \\
\frac{c}{2} (d(Tx, x)^p + d(Ty, y)^p + d(Tx, Ty)^p + d(x, y)^p - d(Tx, y)^p - d(x, Ty)^p)
\]

is a central object. In general: for any \( x, y, u, v \in (G, d) \),

\[
\psi^{(p,c)}(x, y, u, v) := \\
\frac{c}{2} (d(u, x)^p + d(v, y)^p + d(u, v)^p + d(x, y)^p - d(u, y)^p - d(x, v)^p)
\]

For \( p = c = 2 \), \( \psi^{(p,c)} \geq 0 \) for all \( x, y, u, v \in (G, d) \).
In particular,

\[
\psi_{T}^{(2,2)}(x, y) \geq 0 \text{ on } \text{CAT}(0) \text{ spaces}
\]

and

\[
\psi_{T}^{(2,2)}(x, y) = \|(Tx - x) - (Ty - y)\|^2 \text{ on Hilbert spaces}
\]
Calculus of $a\alpha$-fne mappings [Lauster-L., 2021]

Compositions of pointwise $a\alpha$-fne mappings

$T_0 : D \rightarrow G$ is pointwise $a\alpha$-fne at $y$ on $D$ with constant $\alpha_0$ and violation $\epsilon_0$ and $T_1 : T_0(D) \rightarrow G$ is pointwise $a\alpha$-fne at $y$ on $T_0(D)$ with constant $\alpha_1$ and violation $\epsilon_1$. At $y \in \text{Fix } T_0 \cap \text{Fix } T_1$, the $\overline{T} = T_1 \circ T_0$ is pointwise $a\alpha$-fne at $y$ on $D$ with violation $\overline{\epsilon} = \epsilon_0 + \epsilon_1 + \epsilon_0 \epsilon_1$ and constant

$$\overline{\alpha} = \frac{\alpha_0 + \alpha_1 - 2\alpha_0 \alpha_1}{\frac{\epsilon}{2}(1 - \alpha_0 - \alpha_1 + \alpha_0 \alpha_1) + \alpha_0 + \alpha_1 - 2\alpha_0 \alpha_1}.$$  

(1)

Krasnoselsky-Mann relaxations

When $T$ is pointwise $ane$ at all $y \in \text{Fix } T$ with violation $\epsilon$, $T_\beta := \beta T \oplus (1 - \beta)\text{Id}$ is pointwise $a\alpha$-fne at all $y \in \text{Fix } T$ with constant

$$\alpha_\beta = \frac{\alpha_\beta^{p-1}}{\alpha_\beta^{p-1} - \alpha \beta + 1}$$

and violation $\epsilon_\beta := \epsilon \beta$. 
Outline I

Motivation

Metric Spaces

General Regularity Notions

Quantitative Convergence

Example: metric proximal mappings

References
Quantitative convergence, [Lauster-L., 2021]

- $(G, d)$ is a complete $p$-uniformly convex space with constant $c$
- $D \subset G$ with $(D, d)$
- $T: D \to D$ with $\text{Fix } T \cap D \neq \emptyset$ and $T$ is pointwise $\alpha$-fne at all $y \in \text{Fix } T \cap D$ with the same constant $\alpha$ and violation $\epsilon$ on $D$.

$$\mathcal{T}_{\text{Fix } T \cap D}(x) := \left( \inf_{y \in \text{Fix } T \cap D} \psi_{T}^{(p, c)}(x, y) \right)^{1/p}$$

Necessary conditions for convergence rates

- $(x^{k})_{k \in \mathbb{N}}$ defined by $x^{k+1} = Tx^{k}$ is linearly monotone relative to $\text{Fix } T \cap D$ with rate $\theta < 1$ for all $x^{0} \in D$.

Then $x^{k} \to \bar{x} \in \text{Fix } T \cap D$ linearly monotonically with rate $O(\theta^{k})$ whenever $x^{0} \in D$. Moreover, $\mathcal{T}_{\text{Fix } T \cap D}$ is metrically subregular for 0 relative to $D$ on $D$ with (linear) gauge $\mu = (\text{Id} - \theta)^{-1}$.
Metric Subregularity

**Metric subregularity on a set**

Let \((G_1, d_1)\) and \((G_2, d_2)\) be metric spaces and let \(\Phi : G_1 \rightrightarrows G_2\), \(U \subset G_1\). The mapping \(\Phi\) is called **metrically subregular** at \(y \in G_2\) with gauge \(\mu\) on \(U\) relative to \(\Lambda \subset G_1\) whenever

\[
\forall x \in U \cap \Lambda, \quad \text{dist} \ (x, \Phi^{-1}(y) \cap \Lambda) \leq \mu \left( \text{dist} \ (y, \Phi(x)) \right).
\]
Quantitative convergence, [Lauster-L., 2021]

- $(G, d)$ is a complete $p$-uniformly convex space with constant $c$
- $D \subset G$ with $(D, d)$
- $T : D \rightarrow D$ with $\text{Fix } T \cap D \neq \emptyset$ and $T$ is pointwise $\alpha$-fne at all $y \in \text{Fix } T \cap D$ with the same constant $\bar{\alpha}$ and violation $\epsilon$ on $D$.

$$T_{\text{Fix } T \cap D}(x) := \left( \inf_{y \in \text{Fix } T \cap D} \psi_T^{(p, c)}(x, y) \right)^{1/p}$$

Sufficient conditions for convergence rates

- $T$ satisfies

$$d(x, \text{Fix } T \cap D) \leq \mu(T_{\text{Fix } T \cap D}(x)) = \mu(d(x, Tx)), \quad \forall x \in D,$$

with $\mu$ satisfying

$$\sqrt[p]{\frac{1-\bar{\alpha}}{\bar{\alpha}(1+\epsilon)}} < \mu < \sqrt[p]{\frac{1-\bar{\alpha}}{\bar{\alpha}\epsilon}}.$$  

Then $x^k \rightarrow \text{Fix } T \cap D$ at least $R$-linearly convergent with rate

$$\gamma = \sqrt[p]{1 + \epsilon - \frac{1-\bar{\alpha}}{\bar{\alpha}\mu P}}$$

whenever $x^0 \in D$. 
Quantitative convergence, [Lauster-L., 2021]

- $(G, d)$ is a complete $p$-uniformly convex space with constant $c$.
- $D \subset G$ with $(D, d)$.
- $T : D \to D$ with $\text{Fix } T \cap D \neq \emptyset$ and $T$ is pointwise $a\alpha$-fne at all $y \in \text{Fix } T \cap D$ with the same constant $\bar{\alpha}$ and violation $\epsilon$ on $D$.

\[
\mathcal{T}_{\text{Fix } T \cap D}(x) := \left( \inf_{y \in \text{Fix } T \cap D} \psi_T^{(p, c)}(x, y) \right)^{1/p}
\]

Remarks
Convexity is nowhere assumed: this result covers ANY pointwise $a\alpha$-fne mapping $T$. e.g. projections onto prox-regular sets, prox-mappings of prox-regular functions, etc.
Outline I

Motivation

Metric Spaces

General Regularity Notions

Quantitative Convergence

Example: metric proximal mappings

References
Metric Proximal Mappings

\[
\text{prox}_{f, \lambda}^p(x) := \arg\min_{y \in G} f(y) + \frac{1}{p \lambda^{p-1}} d(y, x)^p \quad (\lambda > 0).
\]

Regularity of \( \text{prox}_{f, \lambda}^p \)

\((G, d)\) is a \( p \)-uniformly convex metric space with constant \( c \in (0, 2] \), and \( f : G \to \mathbb{R} \) is proper, convex and lsc. Then

- \( \text{prox}_{f, \lambda}^p \) is nonempty and single-valued [Izuchukwu et al, 2019]
- If \((G, d)\) is symmetric perpendicular, then [Lauster, L. 2021]

\[
d(\text{prox}_{f, \lambda}^p(x), y) < d(x, y) \quad \forall x \in G, \forall y \in \arg\min f
\]
Metric Proximal Mappings

Regularity of $\text{prox}^p_{f,\lambda}$

$(G, d)$ is a $p$-uniformly convex metric space with constant $c \in (0, 2]$, and $f : G \to \mathbb{R}$ is proper, convex and lsc. Then

\[ d(\text{prox}^p_{f,\lambda}(x), \text{prox}^p_{f,\lambda}(y))^p \leq (1 + \epsilon)d(x, y)^p - \frac{1 - \alpha_c}{\alpha_c} \psi^{(p,c)}_{\text{prox}^p_{f,\lambda}}(x, y) \]

$\forall x \in D, \forall y \in \text{argmin } f$

where $\psi^{(p,c)}_T(x, y)$ is the transport discrepancy and

\[ \alpha_c = \frac{c(c-1)}{c(c-1)+2} \quad \text{and violation} \quad \epsilon_c = \frac{2-c}{c-1}. \]
Metric Proximal Mappings

Regularity of $\text{prox}^p_{f,\lambda}$ [Lauster. L., 2021]

If $(G, d)$ is a CAT($\kappa$) space, then $\text{prox}^2_{f,\lambda}$ satisfies

$$\forall y \in \arg\min f, \forall \epsilon > 0, \exists D_\epsilon(y) \subset G : \forall x \in D_\epsilon(y)$$

$$d(\text{prox}^2_{f,\lambda}(x), y)^2 \leq (1 + \epsilon)d(x, y)^2 - \psi^{(2, \kappa)}(x, y)$$

Put this together with Kuwae’s result to get that, on a CAT($\kappa$) space with small enough diameter, $\text{prox}^2_{f,\lambda}(x)$ is also quasi strictly nonexpansive [Bërdëllima-Lauster-L., 2022], that is:

$$\forall y \in \arg\min f, \forall \epsilon > 0, \exists D_\epsilon(y) \subset G : \forall x \in D_\epsilon(y)$$

$$d(\text{prox}^2_{f,\lambda}(x), y)^2 \leq (1 + \epsilon)d(x, y)^2 - \psi^{(2, \kappa)}(x, y)$$

and

$$d(\text{prox}^p_{f,\lambda}(x), y) < d(x, y)$$

(in general CAT($\kappa$) spaces the first inequality does not imply the second)
Main Result [Lauster-L., 2022]

Recall: \((G, d)\) is a complete CAT(\(\kappa\)) space with \(\kappa > 0\) \(f\) and \(g\) are proper, convex and lsc with \(\text{argmin } f \cap \text{argmin } g \neq \emptyset\), and \(T : D \to D\) where \(D \subset G\) denotes one of the following:

(i) \(T := \text{prox}_{f_N, \lambda_N} \circ \text{prox}_{f_{N-1}, \lambda_{N-1}} \circ \cdots \circ \text{prox}_{f_1, \lambda_1}\);

(ii) \(T := \beta \text{prox}_{g, \lambda} \oplus (1 - \beta) \text{Id}\);

(iii) \(T := \text{prox}_{f_1, \lambda_1} \circ (\beta \text{prox}_{g, \lambda_2} \oplus (1 - \beta) \text{Id})\);

(iv) \(T := P_C \circ (\beta \text{prox}_{g, \lambda_1} \oplus (1 - \beta) \text{Id})\),

Convergence of proximal algorithms in CAT(\(\kappa\)) spaces

If

\[
d(x, \text{Fix } T \cap D) \leq \mu d(x, Tx), \quad \forall x \in D,
\]

then \(x^k \to \text{Fix } T \cap D\) at least linearly with rate

\[
\gamma = \sqrt{1 + \epsilon - \frac{1-\alpha}{\alpha \mu^2}} < 1 \quad \text{whenever } x^0 \in D \text{ is close enough to } \text{Fix } T.
\]

The asymptotic rate of convergence is \(\bar{\gamma} = \sqrt{1 - \frac{1-\alpha}{\alpha \mu^2}} < 1\).
Main Result

Convergence of proximal algorithms in CAT(κ) spaces

Keys to the results:

- CAT(κ) spaces are symmetric perpendicular on small enough domains
- pointwise almost α-firmly nonexpansive mappings and
- metric subregularity
Outline I

Motivation

Metric Spaces

General Regularity Notions

Quantitative Convergence

Example: metric proximal mappings

References
\(\alpha\)-firmly nonexpansive operators on metric spaces.
https://doi.org/10.1007/s11784-021-00919-4

Convergence of proximal splitting algorithms in CAT(\(\kappa\)) spaces and beyond
https://doi.org/10.1186/s13663-021-00698-0

Quantitative convergence analysis of iterated expansive, set-valued mappings.