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

RTG2088

Outline I

Motivation

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

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 $k = 0, 1, 2, \ldots$

$$x_{k+1} = Tx_k := \left(\mathsf{prox}_{f_N,\lambda_N} \circ \cdots \circ \mathsf{prox}_{f_2,\lambda_2} \circ \mathsf{prox}_{f_1,\lambda_1}\right)(x_k)$$

Special case:

$$\inf_{x \in C} f(x) = \inf_{x \in G} \iota_C(x) + f(x)$$

Metric Projected Gradient

Choose $x_0 \in G$.
For $k = 0, 1, 2, \ldots$

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

(we interpret $\left((1-\beta)\,\mathsf{Id} \oplus \beta\,\mathsf{prox}_{f,\lambda}\right)$ as the direction of steepest descent of the Moreau-Yosida envelope of f with steplength β)

Main Result

Convergence of proximal algorithms in $\mathrm{CAT}(\kappa)$ 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) $T := \mathsf{prox}_{f_N,\lambda_N} \circ \mathsf{prox}_{f_{N-1},\lambda_{N-1}} \circ \cdots \circ \mathsf{prox}_{f_1,\lambda_1}$;

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

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

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

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

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

with constant μ, then the fixed point sequence initialized from any starting point close enough to Fix T is at least linearly convergent to a point in Fix T.

Main Result

Convergence of proximal algorithms in CAT(κ) spaces

Remarks:

- Convergence (without rates) of most 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(κ) is that prox mappings of convex functions are expansive and firmly nonexpansive $\not\Rightarrow$ nonexpansive
- The main innovation in extending to CAT(κ) spaces and nonconvex functions is pointwise almost α-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
  - M. Edelstein. Bull. Am. Math. Soc., 1972.
  - R. D. Nussbaum. J. Math. Anal. and Appl., 1972.
  - R. E. Bruck. Pacific J. Math., 1973.
  - R. E. Bruck and S. Reich. Houston J. Math., 1977.
  - J. B. Baillon, R. E. Bruck, and S. Reich. Houston J. Math., 1978.
  - K. Goebel and S. Reich. Uniform convexity, hyperbolic geometry, and nonexpansive mappings., 1984.
  - S. Reich and I. Shafrir. Proc. Am. Math. Soc., 1987.
  - S. Reich and I. Shafrir. Nonlinear Anal., Theory Methods Appl., 1990.
  - B. D. Rouhani. J. of Math. Anal. and Appl., 1990.
  - E. Kopecká and S. Reich. Nonlinear Anal., Theory Methods Appl., Ser. A, Theory Methods, 2009.
  - D. Ariza-Ruiz, L. Leuştean, and G. López-Acedo. Trans. Am. Math. Soc., 2014.
  - M. Bačák. SIAM J. Optim., 2014.
  - D. Ariza-Ruiz, G. López-Acedo, and A. Nicolae. J Optim Theory Appl, 2015.

A selection of important building blocks II

  - C. Izuchukwu, G. C. Ugwunnadi, O. T. Mewomo, A. R. Khan, and M. Abbas. Numer. Algorithms, 2019.
- Metric (sub)regularity
  - A. D. Ioffe. Math. Program., 2013.
  - A. L. Dontchev and R. T. Rockafellar. Implicit Functions and Solution Mappings, 2011/2014.
  - U. Kohlenbach, G. López-Acedo, and A. Nicolae., Isr. J. Math., 2019.
- Metric space theory
  - S. Ohta. Geom. Dedicata, 2007.
  - K. Kuwae. Calc. Var. Partial Differ. Equ., 2014.
  - K. Kuwae. Anal. Geom. Metr. Spaces, 2015.

Outline I

Setting

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

$$(\forall t \in [0,1])(\forall x, y, z \in G)$$
$$d(z, (1-t)x \oplus ty)^p \leq (1-t)d(z,x)^p + td(z,y)^p - \tfrac{c}{2}t(1-t)d(x,y)^p$$

Examples

- L^p spaces
- CAT(0) spaces ($p = c = 2$)
- Hadamard spaces (complete CAT(0) spaces)
- Hilbert spaces (linear Hadamard spaces)
- CAT(κ) spaces: for all $\delta \in (0, \pi/(4\sqrt{\kappa}))$ the subspace $(\mathbb{B}_\delta(\overline{x}), d)$ is a 2-uniformly convex space with constant $c_\delta = 4\delta\sqrt{\kappa}\,\mathsf{tan}\,(\pi/2 - 2\delta\sqrt{\kappa})$ [Ohta, 2007]

Setting

Symmetric perpendicular spaces

Let (G, d) be a geodesic space and γ and η be two geodesics through p. Then γ is said to be perpendicular to η at point p denoted by $\gamma \perp_{\mathrm{p}} \eta$ if

$$\mathrm{d}(\mathrm{x}, \mathrm{p}) \leq \mathrm{d}(\mathrm{x}, \mathrm{y}) \quad \forall \mathrm{x} \in \gamma, \mathrm{y} \in \eta$$

A space is said to be symmetric perpendicular if for all geodesics γ and η with common point p we have

$$\gamma \perp_{\mathrm{p}} \eta \Leftrightarrow \eta \perp_{\mathrm{p}} \gamma$$

[Kuwae, 2014]

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

Outline I

Almost α-firmly nonexpansive mappings

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

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

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

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

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

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

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

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

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

Transport discrepancy

The transport discrepancy

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

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

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

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

$$\psi_T^{(2,2)}(x,y) \geq 0 \text{ on CAT(0) spaces}$$

and

$$\psi_T^{(2,2)}(x,y) = \|(Tx - x) - (Ty - y)\|^2 \qquad \text{on Hilbert spaces}$$

Calculus of aα-fne mappings [Lauster-L., 2021]

Compositions of pointwise aα-fne mappings

$T_0 : D \to G$ is pointwise aα-fne at y on D with constant α_0 and violation ϵ_0 and $T_1 : T_0(D) \to G$ is pointwise aα-fne at y on $T_0(D)$ with constant α_1 and violation ϵ_1. At $y \in \mathrm{Fix}\, T_0 \cap \mathrm{Fix}\, T_1$, the $\overline{T} = T_1 \circ T_0$ is pointwise aα-fne at y on D with violation $\overline{\epsilon} = \epsilon_0 + \epsilon_1 + \epsilon_0\epsilon_1$ and constant

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

Krasnoselsky-Mann relaxations

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

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

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

Outline I

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 $\mathrm{Fix}\, T \cap D \neq \emptyset$ and T is pointwise aα-fne at all $y \in \mathrm{Fix}\, T \cap D$ with the same constant $\overline{\alpha}$ and violation ϵ on D.
- $$\mathcal{T}_{\mathrm{Fix}\ T\cap D}(x) := \left| \left(\inf_{y \in \mathrm{Fix}\ T\cap D} \psi_T^{(p,c)}(x, y) \right)^{1/p} \right|$$

Necessary conditions for convergence rates

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

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

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 Φ is called metrically subregular at $y \in G_2$ with gauge μ on U relative to $\Lambda \subset G_1$ whenever

$$\forall x \in U \cap \Lambda, \quad \mathsf{dist}\left(x, \Phi^{-1}(y) \cap \Lambda\right) \leq \mu\left(\mathsf{dist}\left(y, \Phi(x)\right)\right).$$

Quantitative convergence, [Lauster-L., 2021]

- (G, d) is a complete p-uniformly convex space with constant c
- $D \subset G$ with (D, d)
- $T : D \to D$ with $\mathrm{Fix}\ T \cap D \neq \emptyset$ and T is pointwise aα-fne at all $y \in \mathrm{Fix}\ T \cap D$ with the same constant $\overline{\alpha}$ and violation ϵ on D.
- $$\mathcal{T}_{\mathrm{Fix}\ T\cap D}(x) := \left| \left(\inf_{y \in \mathrm{Fix}\ T \cap D} \psi_T^{(p,c)}(x, y) \right)^{1/p} \right|$$

Sufficient conditions for convergence rates

- T satisfies

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

with μ satisfying

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

.

Then $x^k \to \mathrm{Fix}\ T \cap D$ at least R-linearly convergent with rate $\gamma = \sqrt[p]{1 + \epsilon - \frac{1-\overline{\alpha}}{\overline{\alpha}\mu^p}}$ 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 Fix $T \cap D \neq \emptyset$ and T is pointwise aα-fne at all $y \in$ Fix $T \cap D$ with the same constant $\overline{\alpha}$ and violation ϵ on D.
- $$\mathcal{T}_{\text{Fix } T \cap D}(x) := \left| \left(\inf_{y \in \text{Fix } T \cap D} \psi_T^{(p,c)}(x, y) \right)^{1/p} \right|$$

Remarks

Convexity is nowhere assumed: this result covers ANY pointwise aα-fne mapping T. e.g. projections onto prox-regular sets, prox-mappings of prox-regular functions, etc.

Outline I

Metric Proximal Mappings

$$\mathsf{prox}^{p}_{f,\lambda}(x) := \mathsf{argmin}_{\,y\in G}\, f(y) + \frac{1}{p\lambda^{p-1}} d(y,x)^{p} \quad (\lambda > 0).$$

Regularity of $\mathsf{prox}^{p}_{f,\lambda}$

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

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

$$d(\mathsf{prox}^{p}_{f,\lambda}(x), y) < d(x,y) \quad \forall x \in G, \forall y \in \mathsf{argmin}\, f$$

Metric Proximal Mappings

Regularity of $\mathsf{prox}^{p}_{f,\lambda}$

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

- [Lauster, L., 2021]

$$d(\mathsf{prox}^{p}_{f,\lambda}(x), \mathsf{prox}^{p}_{f,\lambda}(y))^{p} \leq (1+\epsilon)d(x,y)^{p} - \frac{1-\alpha_c}{\alpha_c}\psi^{(p,c)}_{\mathsf{prox}^{p}_{f,\lambda}}(x,y)$$
$$\forall x \in D, \forall y \in \mathsf{argmin}\ f$$

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 $\mathsf{prox}^{p}_{f,\lambda}$ [Lauster. L., 2021]

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

$$\begin{aligned} &\forall y \in \mathsf{argmin}\ f, \forall \epsilon > 0,\ \exists D_\epsilon(y) \subset G : \qquad \forall x \in D_\epsilon(y) \\ &d(\mathsf{prox}^{2}_{f,\lambda}(x), y)^2 \le (1+\epsilon) d(x,y)^2 - \psi^{(2,c)}_{\mathsf{prox}^{2}_{f,\lambda}}(x,y) \end{aligned}$$

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

$$\begin{aligned} &\forall y \in \mathsf{argmin}\ f, \forall \epsilon > 0,\ \exists D_\epsilon(y) \subset G : \qquad \forall x \in D_\epsilon(y) \\ &d(\mathsf{prox}^{2}_{f,\lambda}(x), y)^2 \le (1+\epsilon) d(x,y)^2 - \psi^{(2,c)}_{\mathsf{prox}^{2}_{f,\lambda}}(x,y) \\ &\text{and} \\ &d(\mathsf{prox}^{p}_{f,\lambda}(x), y) < d(x,y) \end{aligned}$$

(in general CAT(κ) 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 $\mathsf{argmin}\, f \cap \mathsf{argmin}\, g \neq \emptyset$, and $T : D \to D$ where $D \subset G$ denotes one of the following:

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

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

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

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

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

If

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

then $x^k \to \mathrm{Fix}\, T \cap D$ at least linearly with rate $\gamma = \sqrt{1 + \epsilon - \frac{1-\alpha}{\alpha\mu^2}} < 1$ whenever $x^0 \in D$ is close enough to $\mathrm{Fix}\, T$.

The asymptotic rate of convergence is $\overline{\gamma} = \sqrt{1 - \frac{1-\overline{\alpha}}{\overline{\alpha}\mu^2}} < 1$.

Main Result

Convergence of proximal algorithms in CAT(κ) spaces

Keys to the results:

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

Outline I

References I

A. Bërdëllima, F. Lauster, and D. R. Luke.
α-firmly nonexpansive operators on metric spaces.
J. Fixed Point Theory Appl., 24, 2022.
https://doi.org/10.1007/s11784-021-00919-4

F. Lauster, and D. R. Luke.
Convergence of proximal splitting algorithms in CAT(κ) spaces and beyond
Fixed Point Theory Algorithms Sci Eng 2021, 13 (2021).
https://doi.org/10.1186/s13663-021-00698-0

D. R. Luke, N. H. Thao, and M. K. Tam.
Quantitative convergence analysis of iterated expansive, set-valued mappings.
Math. Oper. Res., 43(4):1143–1176, 2018.