Gradient flows and proximal point methods in metric spaces

Miklós Pálfia

Corvinus University of Budapest and University of Szeged

September 8, 2022

miklos.palfia@uni-corvinus.hu

Introduction

Let (X, d) be a complete metric space. Consider a lower semi-continuous (lsc) function $\phi : X \longrightarrow (-\infty, \infty]$ such that

$$D(\phi) := X \setminus \phi^{-1}(\infty) \neq \emptyset.$$

If X is Riemannian, a gradient curve $\xi : [0, \infty) \longrightarrow X$ of ϕ with initial condition $\xi(0) := x_0 \in D(\phi)$ is a solution of

$$\dot{\xi} = -\nabla\phi(\xi).$$

We are interested in constructing gradient curves or finding minimizers of ϕ. Classically, the first is related to the Crandall-Liggett theory of contraction semigroups in Banach spaces generated by monotone nonlinear operators. Secondly, discrete approximations of gradient curves leads us to optimization techniques, such as proximal point methods. All can be treated in a unified manner as instances of (contractive) evolution systems in Banach spaces.

Given $x \in X$ and $\tau > 0$, the *Moreau–Yosida approximation* is $\phi_\tau(x) := \inf_{z \in X} \left\{ \phi(z) + \frac{d^2(x,z)}{2\tau} \right\}$ and set

$$J_\tau^\phi(x) := \left\{ z \in X \,\middle|\, \phi(z) + \frac{d^2(x,z)}{2\tau} = \phi_\tau(x) \right\}.$$

For $x \in D(\phi)$ and $z \in J_\tau^\phi(x)$ we have $d^2(x,z) \le 2\tau\{\phi(x) - \phi(z)\}$.

Assumption

(1) (*coercivity*) There exists $\tau_*(\phi) \in (0, \infty]$ such that $\phi_\tau(x) > -\infty$ and $J_\tau^\phi(x) \neq \emptyset$ for all $x \in X$ and $\tau \in (0, \tau_*(\phi))$.

(2) (*compactness*) For any $Q \in \mathbb{R}$, bounded subsets of the sub-level set $\{x \in X \,|\, \phi(x) \le Q\}$ are relatively compact in X.

Remark

If $\operatorname{diam} X < \infty$ *and* (2) *holds, then the lsc of* ϕ *implies that every sub-level set* $\{x \in X \,|\, \phi(x) \le Q\}$ *is (empty or) compact. Thus* ϕ *is bounded below and we can take* $\tau_*(\phi) = \infty$.

To construct discrete approximations of gradient curves of ϕ, we consider a *partition* of the interval $[0, \infty)$:

$$\mathscr{P}_{\boldsymbol{\tau}} = \{0 = t_{\boldsymbol{\tau}}^0 < t_{\boldsymbol{\tau}}^1 < \cdots\}, \qquad \lim_{k\to\infty} t_{\boldsymbol{\tau}}^k = \infty,$$

and set

$$\tau_k := t_{\boldsymbol{\tau}}^k - t_{\boldsymbol{\tau}}^{k-1} \quad \text{for } k \in \mathbb{N}, \qquad |\boldsymbol{\tau}| := \sup_{k\in\mathbb{N}} \tau_k.$$

We will always assume $|\boldsymbol{\tau}| < \tau_*(\phi)$. Given an initial point $x_0 \in D(\phi)$,

$x_{\boldsymbol{\tau}}^0 := x_0$ and recursively choose arbitrary $x_{\boldsymbol{\tau}}^k \in J_{\tau_k}^{\phi}(x_{\boldsymbol{\tau}}^{k-1})$ for each $k \in \mathbb{N}$.

We call $\{x_{\boldsymbol{\tau}}^k\}_{k\in\mathbb{N}}$ a *discrete solution* of the variational scheme associated with the partition $\mathscr{P}_{\boldsymbol{\tau}}$, which is thought of as a *discrete-time gradient curve* for the potential function ϕ.

Convergence of discrete solutions

Let $\phi : (-\infty, \infty] \longrightarrow X$ be λ-*convex* for some $\lambda \in \mathbb{R}$ in the sense that

$$\phi\big(\gamma(t)\big) \leq (1-t)\phi(x) + t\phi(y) - \frac{\lambda}{2}(1-t)td^2(x, y)$$

for any $x, y \in D(\phi)$ and some minimal geodesic $\gamma : [0, 1] \longrightarrow X$ from x to y.

We remark that the compactness (2) in the Assumption implies the coercivity in this case; we even have $\tau_*(\phi) = \infty$ if $\lambda \geq 0$).

Fix an initial point $x_0 \in D(\phi)$. Take a sequence of partitions $\{\mathscr{P}_{\boldsymbol{\tau}_i}\}_{i\in\mathbb{N}}$ such that $\lim_{i\to\infty} |\boldsymbol{\tau}_i| = 0$ and associated discrete solutions $\{x^k_{\boldsymbol{\tau}_i}\}_{k\in\mathbb{N}}$ with $x^0_{\boldsymbol{\tau}_i} = x_0$. Under Assumption (2), by the compactness argument, a subsequence of the interpolated curves

$$\bar{\boldsymbol{x}}_{\boldsymbol{\tau}_i}(0) := x_0, \qquad \bar{\boldsymbol{x}}_{\boldsymbol{\tau}_i}(t) := x^k_{\boldsymbol{\tau}_i} \quad \text{for } t \in (t^{k-1}_{\boldsymbol{\tau}_i}, t^k_{\boldsymbol{\tau}_i}]$$

converges to a curve $\xi : [0, \infty) \longrightarrow D(\phi)$ point-wise in $t \in [0, \infty)$.

In general, under the coercivity and λ-convexity of ϕ (but without the compactness), if a curve ξ is obtained as above (called a *generalized minimizing movement*), then it is locally Lipschitz on $(0, \infty)$ and satisfies $\lim_{t \downarrow 0} \xi(t) = x_0$ as well as the *energy dissipation identity*:

$$\phi\big(\xi(T)\big) = \phi\big(\xi(S)\big) - \frac{1}{2} \int_S^T \{|\dot{\xi}|^2 + |\nabla \phi|^2(\xi)\} \, dt.$$

Here

$$|\dot{\xi}|(t) := \lim_{s \to t} \frac{d(\xi(s), \xi(t))}{|t - s|}$$

is the *metric speed* existing at almost all t, and

$$|\nabla \phi|(x) := \limsup_{y \to x} \frac{\max\{\phi(x) - \phi(y), 0\}}{d(x, y)}$$

is the (descending) *local slope*. We remark that $|\nabla \phi|$ is lower semi-continuous and $\lim_{i \to \infty} \phi(\bar{\boldsymbol{x}}_{\boldsymbol{\tau}_i}(t)) = \phi(\xi(t))$ for all $t \geq 0$.

CAT(1)-spaces

Given three points $x, y, z \in X$ with $d(x,y) + d(y,z) + d(z,x) < 2\pi$, we can take corresponding points $\tilde{x}, \tilde{y}, \tilde{z}$ in the 2-dimensional unit sphere $\mathbb{S}^2$ such that

$$d_{\mathbb{S}^2}(\tilde{x}, \tilde{y}) = d(x,y), \qquad d_{\mathbb{S}^2}(\tilde{y}, \tilde{z}) = d(y,z), \qquad d_{\mathbb{S}^2}(\tilde{z}, \tilde{x}) = d(z,x).$$

We call $\triangle\tilde{x}\tilde{y}\tilde{z}$ a *comparison triangle* of $\triangle xyz$ in $\mathbb{S}^2$.

Definition (CAT(1)-spaces)

A geodesic metric space (X, d) is called a $\mathrm{CAT}(1)$-*space* if, for any $x, y, z \in X$ with $d(x,y) + d(y,z) + d(z,x) < 2\pi$ and any minimal geodesic $\gamma : [0,1] \longrightarrow X$ from y to z, we have

$$d\bigl(x, \gamma(t)\bigr) \le d_{\mathbb{S}^2}\bigl(\tilde{x}, \tilde{\gamma}(t)\bigr)$$

at all $t \in [0,1]$, where $\triangle\tilde{x}\tilde{y}\tilde{z} \subset \mathbb{S}^2$ is a comparison triangle of $\triangle xyz$ and $\tilde{\gamma} : [0,1] \longrightarrow \mathbb{S}^2$ is the minimal geodesic from $\tilde{y}$ to $\tilde{z}$.

Lemma (Semi-convexity of distance functions)

Let (X, d) be a $\mathrm{CAT}(1)$-space and take $R \in (0, \pi)$. Then there exists $K = K(R) \in \mathbb{R}$ such that the squared distance function $d^2(x, \cdot)$ is K-convex on the open R-ball $B(x, R)$ for all $x \in X$.

We define the *angle* between two geodesics γ and η emanating from $\gamma(0) = \eta(0) = x$ by $\angle_x(\gamma, \eta) := \lim_{s,t\downarrow 0} \angle \widetilde{\gamma(s)}\tilde{x}\widetilde{\eta(t)}$, where $\angle \widetilde{\gamma(s)}\tilde{x}\widetilde{\eta(t)}$ is the angle at $\tilde{x}$ of $\triangle \widetilde{\gamma(s)}\tilde{x}\widetilde{\eta(t)}$ in $\mathbb{S}^2$.

Theorem (First variation formula)

Let $\gamma : [0, 1] \longrightarrow X$ be a geodesic from x to z, and take $y \in X$ with $0 < d(x, y) < \pi$. Then we have

$$\lim_{s\downarrow 0} \frac{d(\gamma(s), y) - d(x, y)}{s} = -d(x, z) \cos \angle_x(\gamma, \eta),$$

where $\eta : [0, 1] \to X$ is the unique minimal geodesic from x to y.

Key lemma

Let (X, d) be a complete $\mathrm{CAT}(1)$-space and $\phi : X \longrightarrow (-\infty, \infty]$ satisfy the λ-convexity for some $\lambda \in \mathbb{R}$ and Assumption (1).

Lemma (Key lemma)

Let $x \in D(\phi)$ and $\tau \in (0, \min\{\pi^2/(2C), \tau_*(\phi)/8\})$ with $C = C(x, \tau_*(\phi), \phi(x), \tau_*(\phi)/8)$. Take $x_\tau \in J_\tau^\phi(x)$. Then we have, for any $y \in D(\phi) \cap B(x_\tau, R - d(x, x_\tau))$ with $R < \pi$ and for $K = K(R)$,

$$\begin{aligned} d^2(x_\tau, y) &\leq d^2(x, y) - \lambda\tau d^2(x_\tau, y) + 2\tau\{\phi(y) - \phi(x_\tau)\} - \frac{K}{2} d^2(x, x_\tau) \\ &\leq d^2(x, y) - \lambda\tau d^2(x_\tau, y) + 2\tau\{\phi(y) - \phi(x_\tau)\} \\ &\quad + \max\{0, -K\} \cdot \tau\{\phi(x) - \phi(x_\tau)\}. \end{aligned}$$

proof of the Key lemma

We have $d^2(x, x_\tau) \leq 2C\tau < \pi^2$ by an a priori lemma of Ambrosio-Gigli-Savaré and the choice of τ. Let $\gamma : [0,1] \longrightarrow X$ be the minimal geodesic from x_τ to y, and $\eta : [0,1] \longrightarrow X$ from x_τ to x. For any $s \in (0,1)$, by the definition of $J_\tau^\phi(x)$ and the λ-convexity of ϕ, we have

$$\begin{aligned}\phi(x_\tau) + \frac{d^2(x, x_\tau)}{2\tau} \leq & \phi(\gamma(s)) + \frac{d^2(x, \gamma(s))}{2\tau} \\ \leq & (1-s)\phi(x_\tau) + s\phi(y) - \frac{\lambda}{2}(1-s)sd^2(x_\tau, y) \\ & + \frac{d^2(x, \gamma(s))}{2\tau}.\end{aligned}$$

Hence

$$\phi(x_\tau) \le \phi(y) + \frac{1}{2\tau}\frac{d^2(x,\gamma(s)) - d^2(x,x_\tau)}{s} - \frac{\lambda}{2}(1-s)d^2(x_\tau,y).$$

Applying the first variation formula twice, we observe the *commutativity*:

$$\lim_{s\downarrow 0}\frac{d^2(x,\gamma(s)) - d^2(x,x_\tau)}{s} = \lim_{t\downarrow 0}\frac{d^2(\eta(t),y) - d^2(x_\tau,y)}{t},$$

since both sides equal $-2d(x_\tau,x)d(x_\tau,y)\cos\angle_{x_\tau}(\gamma,\eta)$.
Notice that η is contained in $B(y,R)$ by the choice of y. Thus it follows from the K-convexity of $d^2(\cdot,y)$ in $B(y,R)$ that

$$\begin{aligned}\lim_{t\downarrow 0}\frac{d^2(\eta(t),y) - d^2(x_\tau,y)}{t} &\le d^2(x,y) - d^2(x_\tau,y) - \frac{K}{2}d^2(x,x_\tau)\\ &\le d^2(x,y) - d^2(x_\tau,y) + \max\{0,-K\}\cdot\tau\{\phi(x) - \phi(x_\tau)\}.\end{aligned}$$

□

Remark

(a) Used before by Mayer, Ambrosio-Gigli-Savaré and Bačák is the direct application of the convexity of ϕ *and* $d^2(x,\cdot)$ *along* γ*, which implies in our setting*

$$\frac{K}{2}d^2(x_\tau,y) \le d^2(x,y) - \lambda\tau d^2(x_\tau,y) + 2\tau\{\phi(y)-\phi(x_\tau)\} - d^2(x,x_\tau).$$

This coincides with our estimate when $K=2$*. The commutativity was used to move the coefficient* $K/2$ *from* $d^2(x_\tau,y)$ *to* $d^2(x,x_\tau)$*.*
(b) The Riemannian nature of the space (i.e., the angle) is essential in the commutativity. In fact, on a Finsler manifold (M,F)*, commutativity (written using only the distance) implies*

$$g_v(v,w) = g_w(v,w) \qquad \text{for all } v,w \in T_xM\setminus\{0\},\ x\in M,$$

and the parallelogram identity on T_xM *and hence* F *is Riemannian.*

Applications to gradient flows

Our argument covers two cases. In both cases, (X, d) is complete, $\phi : X \longrightarrow (-\infty, \infty]$ is lower semi-continuous, λ-convex and $D(\phi) \neq \emptyset$.

Case (I)

(X, d) is a $\mathrm{CAT}(1)$-space.

Case (II)

(X, d) satisfies the commutativity and the K-convexity of the squared distance function, and ϕ satisfies the coercivity condition (Assumption (1)).

We stress that both $\lambda, K \in \mathbb{R}$ can be negative.

Interpolations

Given an initial point $x_0 \in D(\phi)$ and a partition $\mathscr{P}_{\boldsymbol{\tau}}$ with $|\boldsymbol{\tau}| < \tau_*(\phi)$, we fix a discrete solution $\{x_{\boldsymbol{\tau}}^k\}_{k\in\mathbb{N}}$. Let us also take a point $y \in X$. We interpolate the discrete data $x_{\boldsymbol{\tau}}^k$, $d(x_{\boldsymbol{\tau}}^k, y)$ and $\phi(x_{\boldsymbol{\tau}}^k)$ as follows:
For $t \in (t_{\boldsymbol{\tau}}^{k-1}, t_{\boldsymbol{\tau}}^k]$, $k \in \mathbb{N}$, define

$$\bar{\boldsymbol{x}}_{\boldsymbol{\tau}}(t) := x_{\boldsymbol{\tau}}^k \in J_{\tau_k}^{\phi}(x_{\boldsymbol{\tau}}^{k-1}) \quad (\bar{\boldsymbol{x}}_{\boldsymbol{\tau}}(0) := x_0),$$

$$\bar{\boldsymbol{d}}_{\boldsymbol{\tau}}(t;y) := \left\{ d^2(x_{\boldsymbol{\tau}}^{k-1}, y) + \frac{t - t_{\boldsymbol{\tau}}^{k-1}}{\tau_k} \{ d^2(x_{\boldsymbol{\tau}}^k, y) - d^2(x_{\boldsymbol{\tau}}^{k-1}, y) \} \right\}^{1/2},$$

$$\bar{\boldsymbol{\phi}}_{\boldsymbol{\tau}}(t) := \phi(x_{\boldsymbol{\tau}}^{k-1}) + \frac{t - t_{\boldsymbol{\tau}}^{k-1}}{\tau_k} \{ \phi(x_{\boldsymbol{\tau}}^k) - \phi(x_{\boldsymbol{\tau}}^{k-1}) \}.$$

Recall that $\tau_k = t_{\boldsymbol{\tau}}^k - t_{\boldsymbol{\tau}}^{k-1}$ and note that $\bar{\boldsymbol{\phi}}_{\boldsymbol{\tau}}$ is non-increasing.

Theorem (Discrete evolution variational inequality)

Assuming $|\boldsymbol{\tau}| < \tau_*(\phi)$*, we have*

$$\frac{1}{2}\frac{d}{dt}\left[\bar{\boldsymbol{d}}^2_{\boldsymbol{\tau}}(t;y)\right] + \frac{\lambda}{2} d^2\big(\bar{\boldsymbol{x}}_{\boldsymbol{\tau}}(t), y\big) + \bar{\boldsymbol{\phi}}_{\boldsymbol{\tau}}(t) - \phi(y) \leq \mathscr{R}_{\boldsymbol{\tau},K}(t)$$

for almost all $t \in (0, T)$ *and all* $y \in D(\phi)$*, where for* $t \in (t_{\boldsymbol{\tau}}^{k-1}, t_{\boldsymbol{\tau}}^{k}]$

$$\mathscr{R}_{\boldsymbol{\tau},K}(t) := \left(\frac{t_{\boldsymbol{\tau}}^k - t}{\tau_k} + \frac{\max\{0, -K\}}{2}\right)\{\phi(x_{\boldsymbol{\tau}}^{k-1}) - \phi(x_{\boldsymbol{\tau}}^{k})\}.$$

Convergence of discrete solutions

Theorem (Unique limits of discrete solutions)

Fix an initial point $x_0 \in D(\phi)$ *and consider discrete solutions* $\{x_{\boldsymbol{\tau}_i}^k\}_{k\in\mathbb{N}}$ *with* $x_{\boldsymbol{\tau}_i}^k = x_0$ *associated with a sequence of partitions* $\{\mathscr{P}_{\boldsymbol{\tau}_i}\}_{i\in\mathbb{N}}$ *such that* $\lim_{i\to\infty} |\boldsymbol{\tau}_i| = 0$*. Then the interpolated curve* $\bar{\boldsymbol{x}}_{\boldsymbol{\tau}_i} : [0,\infty) \longrightarrow X$ *converges to a curve* $\xi : [0,\infty) \longrightarrow X$ *with* $\xi(0) = x_0$ *as* $i \to \infty$ *uniformly on each bounded interval* $[0,T]$*. In particular, the limit curve* ξ *is independent of the choice of the sequence of partitions nor discrete solutions.*

We can define the *gradient flow* operator

$$\mathcal{G} : [0,\infty) \times D(\phi) \longrightarrow D(\phi) \tag{4.1}$$

by $\mathcal{G}(t,x_0) := \xi(t)$, where $\xi : [0,\infty) \longrightarrow X$ is the unique gradient curve with $\xi(0) = x_0$. Then the semigroup property holds:

$$\mathcal{G}\big(t, \mathcal{G}(s,x_0)\big) = \mathcal{G}(s+t, x_0) \quad \text{for all } s,t \geq 0.$$

Contraction property

Theorem (Contraction property)

Take $x_0, y_0 \in D(\phi)$ *and put* $\xi(t) := \mathcal{G}(t, x_0)$ *and* $\zeta(t) := \mathcal{G}(t, y_0)$. *Then we have, for any* $t > 0$,

$$d\big(\xi(t), \zeta(t)\big) \leq e^{-\lambda t} d(x_0, y_0).$$

The contraction property allows us to take the continuous limit

$$\mathcal{G} : [0, \infty) \times \overline{D(\phi)} \longrightarrow \overline{D(\phi)}$$

of the gradient flow operator, which again enjoys the semigroup property as well as the contraction property.

Evolution variational inequality

Theorem (Evolution variational inequality)

Take $x_0 \in D(\phi)$ and put $\xi(t) := \mathcal{G}(t, x_0)$. Then we have

$$\limsup_{\varepsilon \downarrow 0} \frac{d^2(\xi(t+\varepsilon), y) - d^2(\xi(t), y)}{2\varepsilon} + \frac{\lambda}{2} d^2\big(\xi(t), y\big) + \phi\big(\xi(t)\big) \leq \phi(y)$$

for all $y \in D(\phi)$ and $t > 0$. In particular,

$$\frac{1}{2} \frac{d}{dt}\big[d^2\big(\xi(t), y\big)\big] + \frac{\lambda}{2} d^2\big(\xi(t), y\big) + \phi\big(\xi(t)\big) \leq \phi(y)$$

for all $y \in D(\phi)$ and almost all $t > 0$.

Stationary points and large time behavior of the flow

Theorem

A point $x_0 \in D(\phi)$ satisfies $|\nabla\phi|(x_0) = 0$ if and only if $\mathcal{G}(t, x_0) = x_0$ for all $t > 0$.

Theorem (Large time behavior)

Take $x_0 \in D(\phi)$, put $\xi(t) := \mathcal{G}(t, x_0)$ and assume $\lim_{t\to\infty} \phi(\xi(t)) > -\infty$. Then we have $\lim_{t\to\infty} |\nabla\phi|(\xi(t)) = 0$.

Corollary

Take $x_0 \in D(\phi)$, put $\xi(t) := \mathcal{G}(t, x_0)$ and assume that there is a sequence $\{t_n\}_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} t_n = \infty$ and $\{\xi(t_n)\}_{n\in\mathbb{N}}$ converges to a point $\bar{x}$. Then $\bar{x}$ is a stationary point of ϕ and $\lim_{t\to\infty} \phi(\xi(t)) = \phi(\bar{x})$.

A Trotter–Kato product formula

Assumption

Let (X, d) be a complete metric space in either Case (I) or Case (II), and assume additionally $D := \mathrm{diam}\, X < \infty$. For $i = 1, 2$, we consider lsc, λ_i-convex function $\phi_i : X \longrightarrow (-\infty, \infty]$ ($\lambda_i \in \mathbb{R}$) satisfying $D(\phi_1) \cap D(\phi_2) \neq \emptyset$ and the compactness (Assumption (2)).

Given $z_0 \in D(\phi) = D(\phi_1) \cap D(\phi_2)$ and a partition $\mathscr{P}_{\boldsymbol{\tau}}$, we consider the discrete variational schemes for ϕ_1 and ϕ_2 in turn, namely

$$z_{\boldsymbol{\tau}}^0 := z_0, \text{ choose arbitrary } \hat{z}_{\boldsymbol{\tau}}^k \in J_{\tau_k}^{\phi_1}(z_{\boldsymbol{\tau}}^{k-1}) \text{ and } z_{\boldsymbol{\tau}}^k \in J_{\tau_k}^{\phi_2}(\hat{z}_{\boldsymbol{\tau}}^k) \text{ for } k \in \mathbb{N}.$$

The *Trotter–Kato product formula* asserts that $\{z_{\boldsymbol{\tau}}^k\}_{k \geq 0}$ converges to the gradient curve of $\phi := \phi_1 + \phi_2$ emanating from z_0 in an appropriate sense.

Assumption

Given $z_0 \in D(\phi)$ and a partition $\mathscr{P}_{\boldsymbol{\tau}}$, set

$$\delta_{\boldsymbol{\tau}}^k(z_0) := \max\{0, \phi_2(\hat{z}_{\boldsymbol{\tau}}^k) - \phi_2(z_{\boldsymbol{\tau}}^{k-1}), \phi_1(z_{\boldsymbol{\tau}}^k) - \phi_1(\hat{z}_{\boldsymbol{\tau}}^k)\}$$

for $k \in \mathbb{N}$ by suppressing the dependence on the choice of $\{\hat{z}_{\boldsymbol{\tau}}^k, z_{\boldsymbol{\tau}}^k\}_{k\in\mathbb{N}}$. Assume that, for any ε, $T > 0$, there is $\Delta_\varepsilon^T(z_0) < \infty$ such that

$$\sum_{k=1}^{N} \delta_{\boldsymbol{\tau}}^k(z_0) \leq \Delta_\varepsilon^T(z_0)$$

for any $\mathscr{P}_{\boldsymbol{\tau}}$ with $|\boldsymbol{\tau}| < \varepsilon$, $N \in \mathbb{N}$ with $t_{\boldsymbol{\tau}}^N \leq T$, and for any solution $\{\hat{z}_{\boldsymbol{\tau}}^k, z_{\boldsymbol{\tau}}^k\}_{k\in\mathbb{N}}$. This in particular guarantees that $\hat{z}_{\boldsymbol{\tau}}^k \in D(\phi)$ and $z_{\boldsymbol{\tau}}^k \in D(\phi)$.

Introduce the interpolated curve $\bar{\boldsymbol{z}}_{\boldsymbol{\tau}}$:

$$\bar{\boldsymbol{z}}_{\boldsymbol{\tau}}(0) := z_0, \qquad \bar{\boldsymbol{z}}_{\boldsymbol{\tau}}(t) := z_{\boldsymbol{\tau}}^k \quad \text{for } t \in (t_{\boldsymbol{\tau}}^{k-1}, t_{\boldsymbol{\tau}}^k].$$

Theorem (A Trotter–Kato product formula)

Let the above assumptions be satisfied. Given $z_0 \in D(\phi)$, the curve $\bar{\boldsymbol{z}}_{\boldsymbol{\tau}}$ converges to the gradient curve $\xi := \mathcal{G}(\cdot, z_0)$ of ϕ (constructed in the previous section) as $|\boldsymbol{\tau}| \to 0$ uniformly on each bounded interval $[0, T]$.

Nonsmooth convex optimization

Definition (Proximal Point Algorithm)

Let (X, d) be a complete Alexandrov space either with curvature bounded above or below by κ, and $G \subset X$ be a closed, geodesically convex set satisfying the following:

(1) In the upper curvature bound case, $\mathrm{diam}\, G < \pi/(2\sqrt{\kappa})$ if $\kappa > 0$;

(2) In the lower curvature bound case, $\dim X < \infty$, $\partial X = \emptyset$, and $\mathrm{diam}\, G < \infty$ if $\kappa < 0$.

Let $f_i : G \to (-\infty, \infty]$ be a convex, lower semi-continuous function for $i = 1, \ldots, n$. Set $f(x) := \sum_{i=1}^{n} f_i(x)$ and suppose that it is proper. Take $\lambda_k > 0$ s.t. $\sum_{k=0}^{\infty} \lambda_k = +\infty$ and also $\sum_{k=0}^{\infty} \lambda_k^2 < +\infty$. Given $x_0 \in G$ and for each $k \geq 0$ and $1 \leq i \leq n$, we set

$$x_{kn+i} := J_{\lambda_k}^{f_i}(x_{kn+i-1}).$$

Theorem

Let (X, d), $G \subset X$, $f = \sum_{i=1}^{n} f_i$ and $\{\lambda_k\}_{k\geq 0}$ be as above. Assume further that X is locally compact, f_i is L-Lipschitz for some $L \geq 1$ and all i, and that $\inf_G f$ is attained at some point. Then x_m converges to a minimizer of f in G as $m \to \infty$.

Proposition

Let (X, d), $G \subset X$, $f = \sum_{i=1}^{n} f_i$ be as above and further assume that f_i is L-Lipschitz, and that f is K-convex for some $K > 0$. Take $\lambda_k > 0$ with $\lambda_k K < 1$, $\lambda_k \to 0$ and $\sum_{k=0}^{\infty} \lambda_k = +\infty$, and consider the sequence $\{x_m\}_{m\geq 0}$ generated by the above. Then x_m converges to the unique minimizer $y \in G$ of f as $m \to \infty$.

An application: Sturm's law of large numbers

Theorem (Sturm 2002, Annals of Prob.)

Let (X, d) be a $CAT(0)$-space and let $\mathcal{P}^2(X)$ denote the set of all probability measures μ s.t. $\int_X d^2(x, a)d\mu(a) < \infty$. Let $a\#_t b$ denote the unique geodesic between $a, b \in X$. Then for $\mu \in \mathcal{P}^2(X)$

$$\Lambda(\mu) := \arg\min_{x \in X} \int_X d^2(x, a)d\mu(a)$$

exists and is unique. Moreover consider an i.i.d. sequence of random variables $\{Y_i\}_{i \in \mathbb{N}}$ with law μ and define

$$S_1 := Y_1,$$
$$S_{k+1} := S_k \#_{\frac{1}{k+1}} Y_{k+1}.$$

Then S_k converges to $\Lambda(\mu)$ almost surely, if $\mathrm{supp}(\mu)$ is bounded.

An application: Nodice theorem for the Karcher mean

A deterministic version of Sturm's law (cf. also Holbrook 2012):

Theorem (Lim-Pálfia 2014, Bull. LMS)

Let (X, d) be a $CAT(0)$-space and let $\mu := \sum_{i=0}^{n-1} \frac{1}{n}\delta_{a_i}$ with $a_i \in X$. Consider the deterministic sequence $\{S_k\}_{k\in\mathbb{N}}$ defined as the inductive sequence of geometric means

$$S_1 := a_0,$$
$$S_{k+1} := S_k \#_{\frac{1}{k+1}} a_{\overline{k}}$$

where $\overline{k} := k \mod (n)$. Then $S_k \to \Lambda(\mu)$ with rate $d(S_k, \Lambda(\mu)) = O(1/k)$.

The above along with Sturm's slln even generalizes to CAT(κ) spaces (Ohta-Pálfia 2015, Yokota 2018) and positive operators (Lim-Pálfia 2020, 2021).

Abstract law of large numbers

Let $G \subset X$ be a closed, geodesically convex set. We assume that (G, d) is separable. Consider the set of all lower semi-continuous, convex functions $f : G \to (-\infty, \infty]$ not identically $+\infty$, denoted by $F(G)$. For $K > 0$, we denote by $F_K(G)$ the subset of all lower semi-continuous, K-convex functions $f : G \to (-\infty, \infty]$ not identically $+\infty$.

Denote by $\mathfrak{P}(F_K(G))$ the set of all complete probability measures on $F_K(G)$ with σ-field generated by the topology of one-sided uniform convergence, such that $g(x) := \int_{F_K(G)} f(x) d\mu(f)$ is lsc $(-\infty, +\infty]$-valued K-convex and there exists $x \in G$ so that $g(x) < +\infty$.

Definition (Variance)

We define the *variance* of $\mu \in \mathfrak{P}(F_K(G))$ by

$$\mathrm{var}(\mu) := \inf_{x \in G} \int_{F_K(G)} f(x) d\mu(f).$$

A fixed $\mu \in \mathfrak{P}(F_K(G))$ can be viewed as the distribution of an $F_K(G)$-valued random variable. $\mathbb{E}\varphi := \int_{F_K(G)} \varphi(f) d\mu(f)$

Definition (Expectation)

Let $\mu \in \mathfrak{P}(F_K(G))$. We define the *expectation* of μ as

$$\mathbb{E}\mu := \arg\min_{x \in G} \int_{F_K(G)} f(x) d\mu(f),$$

which is indeed uniquely determined by the K-convexity of $g(x) = \int_{F_K(G)} f(x) d\mu(f)$.

The above is motivated by the definition of Sturm of the expectation as $\mathbb{E}\nu := \arg\min_{x\in G} \int_G d(x,a)^2 d\nu(a)$ of a probability measure ν supported over G.
Note that $g(\mathbb{E}\mu) = \mathrm{var}(\mu)$. Let L_x denote the evaluation operator at $x \in G$ defined as $L_x f := f(x)$. Clearly L_x is a linear functional on the cone $F_K(G)$.

Proposition (Variance inequality)

Let $\mu \in \mathfrak{P}(F_K(G))$. Then, for all $x \in G$, we have

$$d(x, \mathbb{E}\mu)^2 \leq \frac{2}{K}\mathbb{E}\left(L_x - L_{\mathbb{E}\mu}\right) = \frac{2}{K}\int_{F_K(G)} [f(x) - f(\mathbb{E}\mu)] d\mu(f).$$

Theorem (Law of large numbers)

Let (X,d) and $G \subset X$ be as above. Fix $\mu \in \mathfrak{P}(F_K(G))$ supported on L-Lipschitz functions and let $\{f_k\}_{k\geq 0}$ denote a sequence of i.i.d. random variables taking values in $F_K(G)$ with distribution μ. Take a positive sequence $\{\lambda_k\}_{k\geq 0}$ with $\lambda_k K < 1$, $\lambda_k \to 0$ and $\sum_{k=0}^{\infty} \lambda_k = +\infty$. Define the sequence $S_k \in G$ recursively as

$$S_{k+1} := J_{\lambda_k}^{f_k}(S_k), \quad k \geq 0,$$

with an arbitrary starting point $S_0 \in G$, assuming that $S_k \in G$ for all $k \geq 0$ in the lower curvature bound case. Then $S_k \to \mathbb{E}\mu$ almost surely.

- Y. Lim and M. Pálfia, *Strong law of large numbers for the L^1-Karcher mean*, Journal of Functional Analysis 279:7 (2020), 108672.
- Y. Lim and M. Pálfia, *Existence and uniqueness of the L^1-Karcher mean*, Adv. Math. 376 (2021), 107435.
- S. Ohta and M. Pálfia, *Gradient flows and a Trotter-Kato formula of semi-convex functions on CAT(1)-spaces*, Amer. J. Math. 139:4 (2017), pp. 937–965.
- S. Ohta and M. Pálfia, *Discrete-time gradient flows and law of large numbers in Alexandrov spaces*, Calc. Var. PDE 54:2 (2015), pp. 1591–1610.

L. Ambrosio, N. Gigli and G. Savaré, Gradient flows in metric spaces and in the space of probability measures. Second edition, Birkhäuser Verlag, Basel, 2008.

K.-T. Sturm, Nonlinear martingale theory for processes with values in metric spaces of nonpositive curvature, Ann. Probab. **30** (2002), pp. 1195–1222.

Thank you for your kind attention!