Subgradient Methods on Nonpositively Curved Geodesic Spaces

Ariel Goodwin

Cornell University

Joint work with A.S. Lewis, G. López-Acedo, and A. Nicolae

SIAM OP26

Convex optimization: the bare minimum

Claim

Geodesics + **nonpositive curvature** enable convex optimization that

- gives **new ways to think about old tools**
- is **useful and implementable**

Convex optimization: the bare minimum

Claim

Geodesics + **nonpositive curvature** enable convex optimization that

- gives **new ways to think about old tools**
- is **useful and implementable**

This talk

Do practical subgradient methods extend to nonpositively curved geodesic spaces?

Convex optimization: the bare minimum

Claim

Geodesics + **nonpositive curvature** enable convex optimization that

- gives **new ways to think about old tools**
- is **useful and implementable**

This talk

Do practical subgradient methods extend to nonpositively curved geodesic spaces?

Answer: Yes (**G**LLN26a, **G**LLN26b)

A convenient class of geodesic spaces

Geodesic: an isometry γ from $[a, b] \subseteq \mathbb{R}$ into a metric space (X, d)

A convenient class of geodesic spaces

Geodesic: an isometry γ from $[a, b] \subseteq \mathbb{R}$ into a metric space (X, d)

Hadamard space: a complete **geodesic space** (X, d)

A convenient class of geodesic spaces

Geodesic: an isometry γ from $[a, b] \subseteq \mathbb{R}$ into a metric space (X, d)

Hadamard space: a complete **geodesic space** (X, d) that is **CAT(0)**,

A convenient class of geodesic spaces

Geodesic: an isometry γ from $[a, b] \subseteq \mathbb{R}$ into a metric space (X, d)

Hadamard space: a complete **geodesic space** (X, d) that is **CAT(0)**, meaning $d(\cdot, y)^2$ is strongly convex for all $y \in X$, i.e.

$$t \mapsto d(\gamma(t), y)^2 - t^2 \text{ is convex for any geodesic } \gamma$$

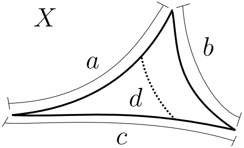
A convenient class of geodesic spaces

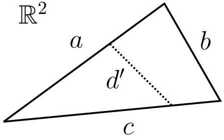
Geodesic: an isometry γ from $[a, b] \subseteq \mathbb{R}$ into a metric space (X, d)

Hadamard space: a complete **geodesic space** (X, d) that is **CAT(0)**, meaning $d(\cdot, y)^2$ is strongly convex for all $y \in X$, i.e.

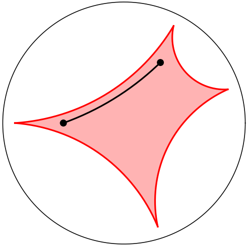
$$t \mapsto d(\gamma(t), y)^2 - t^2 \text{ is convex for any geodesic } \gamma$$

- $f : X \to \mathbb{R}$ is convex $\equiv$ $f \circ \gamma$ is convex for any geodesic γ
- Triangles in CAT(0) spaces are **skinnier** than in Euclidean space





Smooth examples: Hadamard manifolds



Hyperbolic space: open ball with metric

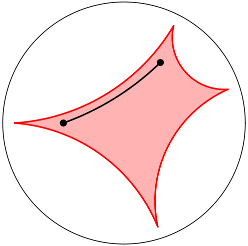
$$d(x, y) = \cosh^{-1}\left(1 + \frac{2\|x-y\|^2}{(1-\|x\|^2)(1-\|y\|^2)}\right)$$

Applications: Hierarchical data embedding (De Sa...'18), large-margin classification (Cho...'18)

Smooth examples: Hadamard manifolds

Hyperbolic space: open ball with metric

$$d(x, y) = \cosh^{-1}\left(1 + \frac{2\|x - y\|^2}{(1 - \|x\|^2)(1 - \|y\|^2)}\right)$$

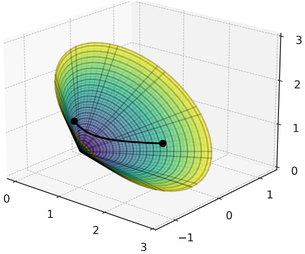


Applications: Hierarchical data embedding (De Sa...'18), large-margin classification (Cho...'18)

Positive definite cone with metric

$$d(X, Y) = \|\log(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}})\|_{\mathrm{Frob}}$$

Applications: Matrix means (Bhatia...'12), covariance estimation (Tyler '87)



Nonsmooth examples: CAT(0) cubical complexes

A complex X of Euclidean cubes and their faces

Nonsmooth examples: CAT(0) cubical complexes

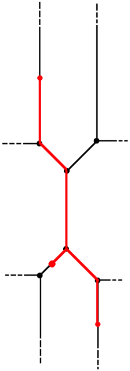
A complex X of Euclidean cubes and their faces, each pair of cubes sharing at most one face

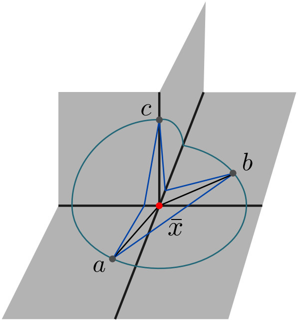
Nonsmooth examples: CAT(0) cubical complexes

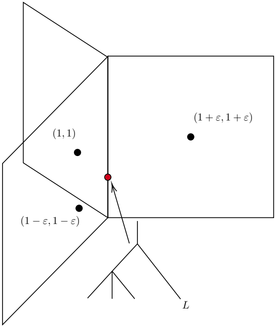
A complex X of Euclidean cubes and their faces, each pair of cubes sharing at most one face, is **CAT(0)** $\iff$ simply connected and **Gromov's link condition** holds

Nonsmooth examples: CAT(0) cubical complexes

A complex X of Euclidean cubes and their faces, each pair of cubes sharing at most one face, is **CAT(0)** $\iff$ simply connected and **Gromov's link condition** holds







Applications: Facility location (Hansen. . . '87), phylogenetics (Billera-Holmes-Vogtmann '01), robot configurations (Ardila-Mantilla '20)

What is a subgradient?

In Euclidean space $\mathbb{E}$, **subgradients** are dual objects: $v \in \partial f(x)$ means

$$f(y) \geq f(x) + \langle v, y - x \rangle \text{ for all } y \in \mathbb{E}$$

What is a subgradient?

In Euclidean space $\mathbb{E}$, **subgradients** are dual objects: $v \in \partial f(x)$ means

$$f(y) \geq f(x) + \langle v, y - x \rangle \text{ for all } y \in \mathbb{E}$$

Natural extension to Hadamard manifolds but ...

- Complexity depends on **lower curvature bound** (Zhang-Sra '16)
- **Acceleration** is obstructed (Criscitiello-Boumal '22)
- Generalization to **non-manifolds** is unclear

Lacking linear structure, we turn to primal tools: **distances and geodesics**

What optimization problems can we pose?

Given $A = \{a_1, \ldots, a_k\} \subseteq X$, determine a representative point:

1. **p-mean** of A: $\underset{x \in X}{\text{minimize}} \;\; \sum_{i=1}^{k} d(x, a_i)^p \;\; (p \geq 1)$
2. **minimal enclosing ball (MEB)** of A: $\underset{x \in X}{\text{minimize}} \;\; \max_{i=1,\ldots,k} \{d(x, a_i)\}$

What optimization problems can we pose?

Given $A = \{a_1, \ldots, a_k\} \subseteq X$, determine a representative point:

1. **p-mean** of A: $\underset{x \in X}{\text{minimize}} \quad \sum_{i=1}^{k} d(x, a_i)^p \ (p \geq 1)$
2. **minimal enclosing ball (MEB)** of A: $\underset{x \in X}{\text{minimize}} \quad \max_{i=1,\ldots,k} \{d(x, a_i)\}$

In theory, (Bačák '13) by iterating the **proximal step**

$$x \leftarrow \underset{y}{\text{argmin}} \left\{ f(y) + \frac{1}{2\lambda} d(y, x)^2 \right\} =: \text{prox}_{\lambda f}(x) \quad (\lambda > 0)$$

for convex lsc f attaining minimum

Are proximal algorithms practical?

$$\textbf{Composite problem:} \quad \min_{x \in X} f(x) = \sum_{i=1}^{k} f_i(x)$$

Assume $f_i \colon X \to \mathbb{R}$ are convex, Lipschitz, and minimizer exists

Are proximal algorithms practical?

$$\textbf{Composite problem:} \quad \min_{x \in X} f(x) = \sum_{i=1}^{k} f_i(x)$$

Assume $f_i : X \to \mathbb{R}$ are convex, Lipschitz, and minimizer exists

Proximal Splitting Method (Bačák '14, Bertsekas '11 in $\mathbb{R}^N$)

Repeat for $n \geq 0$:

$$x \leftarrow \mathrm{prox}_{t_n f_{i(n)}}(x) \qquad (t_n > 0,\ i(n) \sim \mathrm{Unif}\,\{1, \ldots, k\})$$

Are proximal algorithms practical?

$$\textbf{Composite problem:} \quad \min_{x\in X} f(x) = \sum_{i=1}^{k} f_i(x)$$

Assume $f_i \colon X \to \mathbb{R}$ are convex, Lipschitz, and minimizer exists

Proximal Splitting Method (Bačák '14, Bertsekas '11 in $\mathbb{R}^N$)

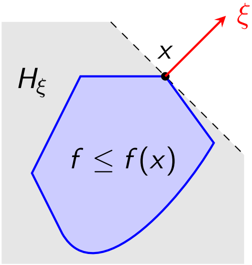
Repeat for $n \geq 0$:

$$x \leftarrow \operatorname{prox}_{t_n f_{i(n)}}(x) \qquad (t_n > 0,\ i(n) \sim \operatorname{Unif}\{1, \ldots, k\})$$

- $f_i = d(\cdot, a_i)^p$ has closed-form prox for $p = 1, 2$
- Proximal updates typically **intractable**
- **Complexity:** $t_j \sim \frac{1}{\sqrt{j}} \implies \mathbb{E}\left[\min_{j=1,\ldots,n} f(x_j)\right] - \min f = O\left(\frac{1}{\sqrt{n}}\right)$

A subgradient-like method in Hadamard space

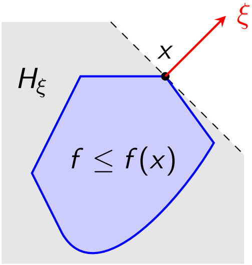
GLLN26a: **subgradient-like** method to minimize $f : X \to \mathbb{R}$



- $n \in \mathbb{N}$: move $t_n > 0$ in **direction** $-\xi_n$
- Based on level set geometry

A subgradient-like method in Hadamard space

GLLN26a: **subgradient-like** method to minimize $f : X \to \mathbb{R}$



- $n \in \mathbb{N}$: move $t_n > 0$ in **direction** $-\xi_n$
- Based on level set geometry
- **Curvature-independent complexity**:

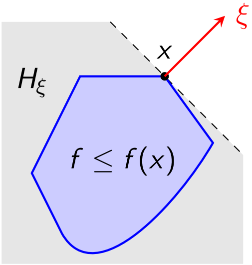
$$\min_{j=1,\ldots,n} f(x_j) - \min f = O\left(\frac{1}{\sqrt{n}}\right)$$

  f Lipschitz, min attained, $t_j \sim 1/\sqrt{j}$

- 1st algorithm for **MEB** in Hadamard space

A subgradient-like method in Hadamard space

GLLN26a: **subgradient-like** method to minimize $f : X \to \mathbb{R}$



- $n \in \mathbb{N}$: move $t_n > 0$ in **direction** $-\xi_n$
- Based on level set geometry
- **Curvature-independent complexity**:

$$\min_{j=1,\ldots,n} f(x_j) - \min f = O\left(\frac{1}{\sqrt{n}}\right)$$

f Lipschitz, min attained, $t_j \sim 1/\sqrt{j}$

- 1st algorithm for **MEB** in Hadamard space

Inadequate for ***p*-mean** problem — need a stronger **subgradient** notion

Busemann functions — a fresh take on convexity

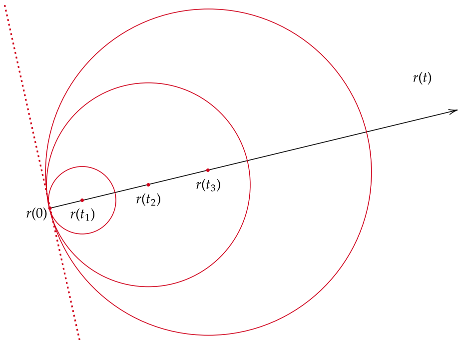
Given a (geodesic) ray $r\colon \mathbb{R}_+ \to X$, the **Busemann function** $b_r\colon X \to \mathbb{R}$,

$$b_r(x) := \lim_{t\to\infty} \left(d(x, r(t)) - t\right)$$

Busemann functions — a fresh take on convexity

Given a (geodesic) ray $r \colon \mathbb{R}_+ \to X$, the **Busemann function** $b_r \colon X \to \mathbb{R}$,

$$b_r(x) := \lim_{t \to \infty} (d(x, r(t)) - t)$$



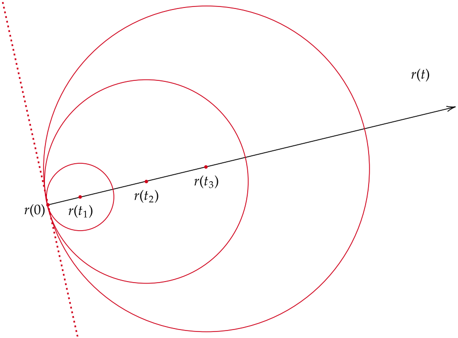
Euclidean: affine functions

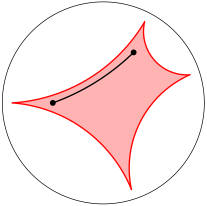
Level sets $\{b_r \leq c\}$ called **horoballs** — generalizes **halfspaces**

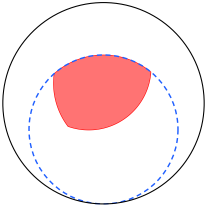
Busemann functions — a fresh take on convexity

Given a (geodesic) ray $r\colon \mathbb{R}_+ \to X$, the **Busemann function** $b_r\colon X \to \mathbb{R}$,

$$b_r(x) := \lim_{t\to\infty} (d(x, r(t)) - t)$$







Euclidean: affine functions

Level sets $\{b_r \leq c\}$ called **horoballs** — generalizes **halfspaces**

A new kind of subgradient

Characterization of (classical) subgradients: $v \in \partial f(x)$ iff

$$x \text{ minimizes } f(\cdot) - \langle v, \cdot \rangle$$

A new kind of subgradient

Characterization of (classical) subgradients: $v \in \partial f(x)$ iff

$$x \text{ minimizes } f(\cdot) - \langle v, \cdot \rangle$$

Two pieces of data:

- r is a ray issuing from x (**direction**)
- $s \in \mathbb{R}_+$ (**magnitude** or speed)

A new kind of subgradient

Characterization of (classical) subgradients: $v \in \partial f(x)$ iff

$$x \text{ minimizes } f(\cdot) - \langle v, \cdot \rangle$$

Two pieces of data:

- r is a ray issuing from x (**direction**)
- $s \in \mathbb{R}_+$ (**magnitude** or speed)

Definition: Busemann Subdifferentiability (GLLN26b)

The pair (r, s) is a **Busemann subgradient** of $f \colon X \to \mathbb{R}$ at x if

$$x \text{ minimizes } f - sb_r$$

A new kind of subgradient

Characterization of (classical) subgradients: $v \in \partial f(x)$ iff

$$x \text{ minimizes } f(\cdot) - \langle v, \cdot \rangle$$

Two pieces of data:

- r is a ray issuing from x (**direction**)
- $s \in \mathbb{R}_+$ (**magnitude** or speed)

Definition: Busemann Subdifferentiability (GLLN26b)

The pair (r, s) is a **Busemann subgradient** of $f\colon X \to \mathbb{R}$ at x if

$$x \text{ minimizes } f - sb_r$$

- $\implies$ convexity (strictly stronger in general: $x \mapsto \mathrm{dist}_C(x)$)
- $\implies$ $\{f \leq \alpha\}$ is supported by **horoball** at boundary points

Other subgradients: Kakavandi-Amini '10, LLN '23, Criscitiello-Kim '25

Are there any Busemann subdifferentiable functions?

Are there any Busemann subdifferentiable functions?

GLLN26b gives examples and calculus for combining them:

Example	f(x)	Ray	Speed
Euclidean	Convex, $v \in \partial f(x)$	$r(t) = x - t\frac{v}{\|v\|}$	$\|v\|$
p-**Distance**	$d(x, a)^p$	r from x to a	$pd(x, a)^{p-1}$
Busemann	$b_{r'}(x)$	r from x with $r \sim r'$	1
Ball Distance	$\mathrm{dist}(x, B_\rho(a))$	r from x to a	$\mathbb{1}_{B_\rho(a)^c}(x)$
Horoball Distance	$\mathrm{dist}(x, \{b_{r'} \leq 0\})$	r from x with $r \sim r'$	$\mathbb{1}_{\{b_{r'} > 0\}}(x)$

Closed under finite max, composition with $\nearrow$ convex

Subgradient methods for the composite problem

$$\textbf{Goal:} \ \min_{x \in C} \sum_{i=1}^{k} f_i(x)$$

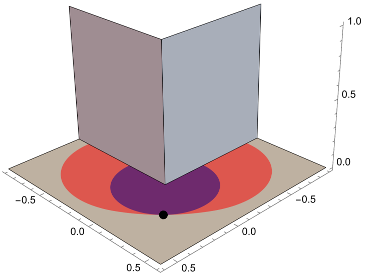
In $\mathbb{R}^n$ we have the **sum rule**: $\sum_{i=1}^{k} \partial f_i(x) \subseteq \partial f(x)$

Subgradient methods for the composite problem

$$\textbf{Goal:}\ \min_{x \in C} \sum_{i=1}^{k} f_i(x)$$

In $\mathbb{R}^n$ we have the **sum rule**: $\sum_{i=1}^k \partial f_i(x) \subseteq \partial f(x)$

No analogue in Hadamard space: $X =$ union of 5 quadrants in $\mathbb{R}^3$



$$f(x) = d^2(x, e_1) + d^2(x, e_2)$$

$\{\mathbf{f} \leq \mathbf{f}(\bar{\mathbf{x}})\}$ not supported by **horoball**

$\implies$ f is not Busemann subdifferentiable at $\bar{x}$

Need a different strategy to solve composite problems...

Subgradient algorithms based on splitting

Bertsekas-Nedic '01 propose **splitting** algorithms for $\min_{x \in C} \sum_{i=1}^{k} f_i(x)$

Euclidean Subgradient Splitting Method

Repeat for $n \geq 0$:

Choose $t_n > 0$, $i(n) \sim \text{Unif}\{1, \ldots, k\}$, $v^n \in \partial f_{i(n)}(x^n)$

$x^{n+1} \leftarrow \text{proj}_C(x^n - t_n v^n)$

Subgradient algorithms based on splitting

Bertsekas-Nedic '01 propose **splitting** algorithms for $\min_{x \in C} \sum_{i=1}^{k} f_i(x)$

Euclidean Subgradient Splitting Method

Repeat for $n \geq 0$:

Choose $t_n > 0$, $i(n) \sim \mathrm{Unif}\{1, \ldots, k\}$, $v^n \in \partial f_{i(n)}(x^n)$

$x^{n+1} \leftarrow \mathrm{proj}_C(x^n - t_n v^n)$

Busemann Subgradient Method (GLLN26b)

$x^{n+1} \leftarrow \mathrm{proj}_C(r_n(t_n s_n))$ where (r_n, s_n) Busemann subgrad of $f_{i(n)}$ at x^n

"Take a step proportional to speed s_n along the ray r_n"

Complexity results

Theorem (GLLN26b)

$$\text{Assume} \quad \min_{x \in C} f(x) = \sum_{i=1}^{k} f_i(x) \quad \text{is attained, where}$$

1. $C \subseteq X$ is closed, convex, and bounded;
2. Each $f_i \colon C \to \mathbb{R}$ is Lipschitz and **Busemann subdifferentiable**.

Complexity results

Theorem (GLLN26b)

$$\text{Assume} \quad \min_{x \in C} f(x) = \sum_{i=1}^{k} f_i(x) \quad \text{is attained, where}$$

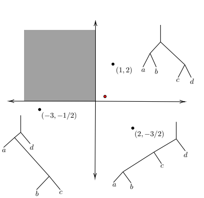
1. $C \subseteq X$ is closed, convex, and bounded;
2. Each $f_i \colon C \to \mathbb{R}$ is Lipschitz and **Busemann subdifferentiable**.

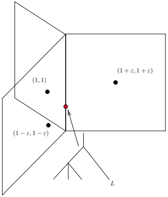
Then the **stochastic or incremental Busemann subgradient method** with $t_n \sim \frac{1}{\sqrt{n+1}}$ achieves

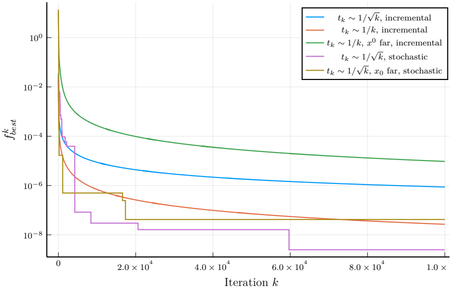
$$\mathbb{E}\left[\min_{j=1,\dots,n} f(x^j)\right] - \min_C f = O\left(\frac{1}{\sqrt{n}}\right).$$

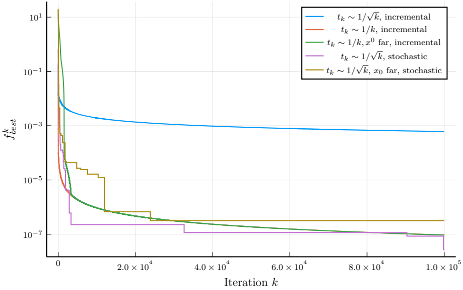
Matches **Euclidean rate**, does not require **curvature lower bound**.

Application: 1-means in tree space









GLLN26a: Circumcenters and mean sets in Hadamard space: horospherical subgradient methods. *Math. of OR*, 2026, to appear.

GLLN26b: Stochastic and incremental subgradient methods for convex optimization on Hadamard spaces. *Math. Prog.*, 2026.