Convex optimization in negatively curved geodesic spaces

Adrian Lewis

Joint work with A. Goodwin, G. Lopez, and A. Nicolae

Midwest Optimization at Waterloo November 2024

Outline: convex optimization without duality

Outline: convex optimization without duality

- Negatively curved geodesic spaces

Outline: convex optimization without duality

- Negatively curved geodesic spaces
  - Manifolds: hyperbolic space, $\mathbf{S}^n_{++}$, ...

Outline: convex optimization without duality

- Negatively curved geodesic spaces
  - Manifolds: hyperbolic space, $\mathbf{S}^n_{++}$, ...
  - Metric trees, BHV tree space, CAT(0) cubical complexes...

Outline: convex optimization without duality

- Negatively curved geodesic spaces
  - Manifolds: hyperbolic space, $\mathbf{S}^n_{++}$, ...
  - Metric trees, BHV tree space, CAT(0) cubical complexes...
- A fresh look at the subgradient method

Outline: convex optimization without duality

- Negatively curved geodesic spaces
  - Manifolds: hyperbolic space, $\mathbf{S}^n_{++}$, ...
  - Metric trees, BHV tree space, CAT(0) cubical complexes...
- A fresh look at the subgradient method
  - Halfspaces as horoballs

Outline: convex optimization without duality

- Negatively curved geodesic spaces
  - Manifolds: hyperbolic space, $\mathbf{S}^n_{++}$, ...
  - Metric trees, BHV tree space, CAT(0) cubical complexes...
- A fresh look at the subgradient method
  - Halfspaces as horoballs
  - Subgradients as constant-speed geodesic rays.

Outline: convex optimization without duality

- Negatively curved geodesic spaces
  - Manifolds: hyperbolic space, $\mathbf{S}^n_{++}$, ...
  - Metric trees, BHV tree space, CAT(0) cubical complexes...
- A fresh look at the subgradient method
  - Halfspaces as horoballs
  - Subgradients as constant-speed geodesic rays.
- Averaging:

Outline: convex optimization without duality

- Negatively curved geodesic spaces
  - Manifolds: hyperbolic space, $\mathbf{S}^n_{++}$, ...
  - Metric trees, BHV tree space, CAT(0) cubical complexes...
- A fresh look at the subgradient method
  - Halfspaces as horoballs
  - Subgradients as constant-speed geodesic rays.
- Averaging:
  - convex combinations versus...

Outline: convex optimization without duality

- Negatively curved geodesic spaces
  - Manifolds: hyperbolic space, $\mathbf{S}^n_{++}$, ...
  - Metric trees, BHV tree space, CAT(0) cubical complexes...
- A fresh look at the subgradient method
  - Halfspaces as horoballs
  - Subgradients as constant-speed geodesic rays.
- Averaging:
  - convex combinations versus...
  - ...weighted means, and how we recognize them.

Outline: convex optimization without duality

- Negatively curved geodesic spaces
  - Manifolds: hyperbolic space, $\mathbf{S}^n_{++}$, ...
  - Metric trees, BHV tree space, CAT(0) cubical complexes...
- A fresh look at the subgradient method
  - Halfspaces as horoballs
  - Subgradients as constant-speed geodesic rays.
- Averaging:
  - convex combinations versus...
  - ...weighted means, and how we recognize them.
- Cutting planes for optimization on CAT(0) cubical complexes

PART I

Negative curvature

Optimization over a **Hadamard** space **X**:

Optimization over a **Hadamard** space $\mathbf{X}$:

a complete metric space $(\mathbf{X}, d)$

Optimization over a **Hadamard** space **X**:

a complete metric space $(\mathbf{X}, d)$ where points connect via **geodesics**

Optimization over a **Hadamard** space $\mathbf{X}$:

a complete metric space $(\mathbf{X}, d)$ where points connect via **geodesics** (isometries from compact intervals into $\mathbf{X}$),

Optimization over a **Hadamard** space $\mathbf{X}$:

a complete metric space $(\mathbf{X}, d)$ where points connect via **geodesics** (isometries from compact intervals into $\mathbf{X}$), and **CAT(0)**,

Optimization over a **Hadamard** space **X**:

a complete metric space $(\mathbf{X}, d)$ where points connect via **geodesics** (isometries from compact intervals into **X**), and **CAT(0)**, meaning d_y^2 is **strongly convex** for all points $y \in X$:

Optimization over a **Hadamard** space $\mathbf{X}$:

a complete metric space $(\mathbf{X}, d)$ where points connect via **geodesics** (isometries from compact intervals into $\mathbf{X}$), and **CAT(0)**, meaning d_y^2 is **strongly convex** for all points $y \in X$:

$$\tau \mapsto d^2\big(g(\tau), y\big) - \tau^2 \qquad \text{convex for geodesics } g.$$

Optimization over a **Hadamard** space **X**:

a complete metric space $(\mathbf{X}, d)$ where points connect via **geodesics** (isometries from compact intervals into **X**), and **CAT(0)**, meaning d_y^2 is **strongly convex** for all points $y \in X$:

$$\tau \;\mapsto\; d^2\big(g(\tau), y\big) - \tau^2 \qquad \text{convex for geodesics } g.$$

Triangles in CAT(0) spaces are **skinny**.

Optimization over a **Hadamard** space $\mathbf{X}$:

a complete metric space $(\mathbf{X}, d)$ where points connect via **geodesics** (isometries from compact intervals into $\mathbf{X}$), and **CAT(0)**, meaning d_y^2 is **strongly convex** for all points $y \in X$:

$$\tau \;\mapsto\; d^2\big(g(\tau), y\big) - \tau^2 \qquad \text{convex for geodesics } g.$$

Triangles in CAT(0) spaces are **skinny**.

Example: **hyperbolic space**,

Optimization over a **Hadamard** space $\mathbf{X}$:

a complete metric space $(\mathbf{X}, d)$ where points connect via **geodesics** (isometries from compact intervals into $\mathbf{X}$), and **CAT(0)**, meaning d_y^2 is **strongly convex** for all points $y \in X$:

$$\tau \ \mapsto \ d^2\big(g(\tau), y\big) - \tau^2 \qquad \text{convex for geodesics } g.$$

Triangles in CAT(0) spaces are **skinny**.

Example: **hyperbolic space**,
the open unit ball in $\mathbf{R}^n$ with

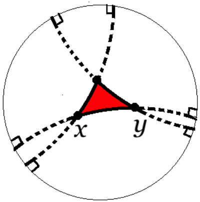
$$\cosh d(x,y) \ = \ 1 + 2\frac{|x-y|^2}{(1-|x|^2)(1-|y|^2)}.$$

Optimization over a **Hadamard** space **X**:

a complete metric space $(\mathbf{X}, d)$ where points connect via **geodesics** (isometries from compact intervals into **X**), and **CAT(0)**, meaning d_y^2 is **strongly convex** for all points $y \in X$:

$$\tau \;\mapsto\; d^2\big(g(\tau), y\big) - \tau^2 \qquad \text{convex for geodesics } g.$$

Triangles in CAT(0) spaces are **skinny**.



Example: **hyperbolic space**,
the open unit ball in $\mathbf{R}^n$ with

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

Modeling using Hadamard spaces

Modeling using Hadamard spaces

- Euclidean and Hilbert spaces: classical optimization

Modeling using Hadamard spaces

- Euclidean and Hilbert spaces: classical optimization
- Simply connected, negatively curved manifolds:

Modeling using Hadamard spaces

- Euclidean and Hilbert spaces: classical optimization
- Simply connected, negatively curved manifolds:
  - hyperbolic space: classify hierarchical data (De Sa. . . '18). . .

Modeling using Hadamard spaces

- Euclidean and Hilbert spaces: classical optimization
- Simply connected, negatively curved manifolds:
  - hyperbolic space: classify hierarchical data (De Sa... '18)...
  - positive definite matrices $\mathbf{S}^n_{++}$ with affine-invariant metric

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

Modeling using Hadamard spaces

- Euclidean and Hilbert spaces: classical optimization
- Simply connected, negatively curved manifolds:
  - hyperbolic space: classify hierarchical data (De Sa... '18)...
  - positive definite matrices $\mathbf{S}^n_{++}$ with affine-invariant metric

$$d(X, Y) = \left\| \log(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}) \right\|_{\text{Frob}}$$: matrix means (Bhatia...'12)

Modeling using Hadamard spaces

- Euclidean and Hilbert spaces: classical optimization
- Simply connected, negatively curved manifolds:
  - hyperbolic space: classify hierarchical data (De Sa. . . '18). . .
  - positive definite matrices $\mathbf{S}^n_{++}$ with affine-invariant metric

$$d(X, Y) = \left\| \log(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}) \right\|_{\text{Frob}}$$: matrix means (Bhatia. . . '12)

- Metric trees:

Modeling using Hadamard spaces

- Euclidean and Hilbert spaces: classical optimization
- Simply connected, negatively curved manifolds:
  - hyperbolic space: classify hierarchical data (De Sa... '18)...
  - positive definite matrices $\mathbf{S}^n_{++}$ with affine-invariant metric

$$d(X,Y) = \left\| \log(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}) \right\|_{\text{Frob}}$$: matrix means (Bhatia... '12)

- Metric trees: facility location

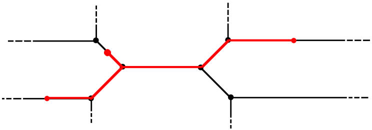
(Hansen... '87), ...

Modeling using Hadamard spaces

- Euclidean and Hilbert spaces: classical optimization
- Simply connected, negatively curved manifolds:
  - hyperbolic space: classify hierarchical data (De Sa... '18)...
  - positive definite matrices $\mathbf{S}^n_{++}$ with affine-invariant metric

$$d(X,Y) = \left\| \log(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}) \right\|_{\text{Frob}} \; : \text{ matrix means (Bhatia...'12)}$$

- Metric trees: facility location (Hansen...'87), ...

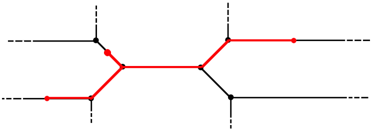


Modeling using Hadamard spaces

- Euclidean and Hilbert spaces: classical optimization
- Simply connected, negatively curved manifolds:
  - hyperbolic space: classify hierarchical data (De Sa. . . '18). . .
  - positive definite matrices $\mathbf{S}^n_{++}$ with affine-invariant metric

$$d(X, Y) = \left\| \log(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}) \right\|_{\mathrm{Frob}}$$: matrix means (Bhatia. . . '12)

- Metric trees: facility location (Hansen. . . '87), . . .



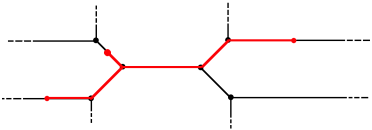
- More general cubical complexes,

Modeling using Hadamard spaces

- Euclidean and Hilbert spaces: classical optimization
- Simply connected, negatively curved manifolds:
  - hyperbolic space: classify hierarchical data (De Sa. . . '18). . .
  - positive definite matrices $\mathbf{S}^n_{++}$ with affine-invariant metric

$$d(X,Y) = \left\| \log(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}) \right\|_{\text{Frob}}$$: matrix means (Bhatia. . . '12)

- Metric trees: facility location (Hansen. . . '87), . . .



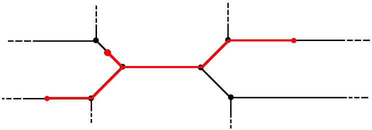
- More general cubical complexes, phylogenetic tree space (Billera-H-V '01)

Modeling using Hadamard spaces

- Euclidean and Hilbert spaces: classical optimization
- Simply connected, negatively curved manifolds:
  - hyperbolic space: classify hierarchical data (De Sa... '18)...
  - positive definite matrices $\mathbf{S}^n_{++}$ with affine-invariant metric

$$d(X, Y) = \left\| \log(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}) \right\|_{\text{Frob}}$$: matrix means (Bhatia...'12)

- Metric trees: facility location (Hansen...'87), ...



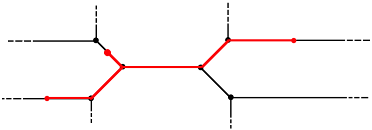
- More general cubical complexes, phylogenetic tree space (Billera-H-V '01) robot configurations (Ardila-Mantilla '20)

Modeling using Hadamard spaces

- Euclidean and Hilbert spaces: classical optimization
- Simply connected, negatively curved manifolds:
  - hyperbolic space: classify hierarchical data (De Sa... '18)...
  - positive definite matrices $\mathbf{S}^n_{++}$ with affine-invariant metric

$$d(X,Y) = \left\| \log(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}) \right\|_{\mathrm{Frob}}$$: matrix means (Bhatia...'12)

- Metric trees: facility location (Hansen...'87), ...



- More general cubical complexes,
phylogenetic tree space (Billera-H-V '01)
robot configurations (Ardila-Mantilla '20)

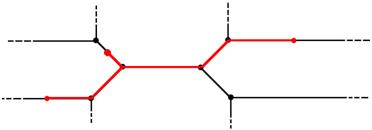
Assumption: **geodesics are tractable**

Modeling using Hadamard spaces

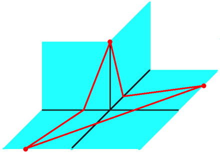
- Euclidean and Hilbert spaces: classical optimization
- Simply connected, negatively curved manifolds:
  - hyperbolic space: classify hierarchical data (De Sa... '18)...
  - positive definite matrices $\mathbf{S}^n_{++}$ with affine-invariant metric

$$d(X,Y) = \left\| \log(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}) \right\|_{\text{Frob}} \; : \text{ matrix means (Bhatia...'12)}$$

- Metric trees: facility location (Hansen...'87), ...



- More general cubical complexes, phylogenetic tree space (Billera-H-V '01) robot configurations (Ardila-Mantilla '20)



Assumption: **geodesics are tractable**

Convex optimization on Hadamard spaces

Example Find the **circumcenter** of a finite set $A \subset \mathbf{X}$:

Convex optimization on Hadamard spaces

Example Find the **circumcenter** of a finite set $A \subset \mathbf{X}$: minimize

$$(*) \qquad \phi_A(x) = \max_{a \in A} d(x, a).$$

Convex optimization on Hadamard spaces

Example Find the **circumcenter** of a finite set $A \subset \mathbf{X}$: minimize

$$(*) \qquad \phi_A(x) = \max_{a \in A} d(x, a).$$

Zhang-Sra '16 extends classical (sub)gradient methods to Hadamard *manifolds* $\mathbf{X}$,

Convex optimization on Hadamard spaces

Example Find the **circumcenter** of a finite set $A \subset \mathbf{X}$: minimize

$$(*) \qquad \phi_A(x) = \max_{a \in A} d(x, a).$$

Zhang-Sra '16 extends classical (sub)gradient methods to Hadamard *manifolds* $\mathbf{X}$, but with wrinkles...

Convex optimization on Hadamard spaces

Example Find the **circumcenter** of a finite set $A \subset \mathbf{X}$: minimize

$$(*) \qquad \phi_A(x) = \max_{a \in A} d(x,a).$$

Zhang-Sra '16 extends classical (sub)gradient methods to Hadamard *manifolds* $\mathbf{X}$, but with wrinkles...

- "subgradients" at $x \in \mathbf{X}$ are **tangents** in $T_x\mathbf{X}$,

Convex optimization on Hadamard spaces

Example Find the **circumcenter** of a finite set $A \subset \mathbf{X}$: minimize

$$(*) \qquad \phi_A(x) = \max_{a \in A} d(x,a).$$

Zhang-Sra '16 extends classical (sub)gradient methods to Hadamard *manifolds* $\mathbf{X}$, but with wrinkles...

- "subgradients" at $x \in \mathbf{X}$ are **tangents** in $T_x\mathbf{X}$, so ...
- the algorithm needs **exponentials** $\mathrm{Exp}_x \colon T_x\mathbf{X} \to \mathbf{X}$

Convex optimization on Hadamard spaces

Example Find the **circumcenter** of a finite set $A \subset \mathbf{X}$: minimize

$$(*) \qquad \phi_A(x) = \max_{a \in A} d(x, a).$$

Zhang-Sra '16 extends classical (sub)gradient methods to Hadamard *manifolds* $\mathbf{X}$, but with wrinkles. . .

- "subgradients" at $x \in \mathbf{X}$ are **tangents** in $T_x\mathbf{X}$, so . . .
- the algorithm needs **exponentials** $\mathrm{Exp}_x \colon T_x\mathbf{X} \to \mathbf{X}$ and . . .
- complexity analysis needs a **curvature lower bound** for $\mathbf{X}$.

Convex optimization on Hadamard spaces

Example Find the **circumcenter** of a finite set $A \subset \mathbf{X}$: minimize

$$(*) \qquad \phi_A(x) = \max_{a \in A} d(x,a).$$

Zhang-Sra '16 extends classical (sub)gradient methods to Hadamard *manifolds* $\mathbf{X}$, but with wrinkles...

- "subgradients" at $x \in \mathbf{X}$ are **tangents** in $T_x\mathbf{X}$, so ...
- the algorithm needs **exponentials** $\mathrm{Exp}_x \colon T_x\mathbf{X} \to \mathbf{X}$ and ...
- complexity analysis needs a **curvature lower bound** for $\mathbf{X}$.

Furthermore, beyond manifolds, only **proximal** methods are known (Bačák '13...):

Convex optimization on Hadamard spaces

Example Find the **circumcenter** of a finite set $A \subset \mathbf{X}$: minimize

$$(*) \qquad \phi_A(x) = \max_{a \in A} d(x, a).$$

Zhang-Sra '16 extends classical (sub)gradient methods to Hadamard *manifolds* $\mathbf{X}$, but with wrinkles...

- "subgradients" at $x \in \mathbf{X}$ are **tangents** in $T_x\mathbf{X}$, so ...
- the algorithm needs **exponentials** $\mathrm{Exp}_x \colon T_x\mathbf{X} \to \mathbf{X}$ and ...
- complexity analysis needs a **curvature lower bound** for $\mathbf{X}$.

Furthermore, beyond manifolds, only **proximal** methods are known (Bačák '13...): rarely implementable.

Convex optimization on Hadamard spaces

Example   Find the **circumcenter** of a finite set $A \subset \mathbf{X}$: minimize

$$(*) \qquad \phi_A(x) = \max_{a \in A} d(x,a).$$

Zhang-Sra '16 extends classical (sub)gradient methods to Hadamard *manifolds* $\mathbf{X}$, but with wrinkles...

- "subgradients" at $x \in \mathbf{X}$ are **tangents** in $T_x\mathbf{X}$, so ...
- the algorithm needs **exponentials** $\mathrm{Exp}_x \colon T_x\mathbf{X} \to \mathbf{X}$ and ...
- complexity analysis needs a **curvature lower bound** for $\mathbf{X}$.

Furthermore, beyond manifolds, only **proximal** methods are known (Bačák '13...): rarely implementable.

Question   Are circumcenters in Hadamard space computable?

PART II

The subgradient method: a fresh look

Review: the subgradient method in Euclidean space

Review: the subgradient method in Euclidean space

Seek to minimize convex objective $\phi\colon \mathbf{R}^n \to \mathbf{R}$.

Review: the subgradient method in Euclidean space

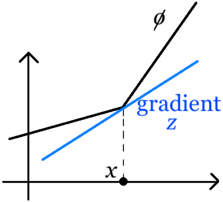
Seek to minimize convex objective $\phi\colon \mathbf{R}^n \to \mathbf{R}$.
At current point $x \in \mathbf{R}^n$,
$\partial\phi(x)$ consists of **subgradients** $z \in \mathbf{R}^n$:

$$x \quad \text{minimizes} \quad \phi - \langle z, \cdot \rangle \ .$$

Review: the subgradient method in Euclidean space

Seek to minimize convex objective $\phi\colon \mathbf{R}^n \to \mathbf{R}$.
At current point $x \in \mathbf{R}^n$,
$\partial\phi(x)$ consists of **subgradients** $z \in \mathbf{R}^n$:

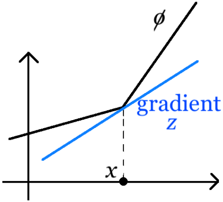
$$x \quad \text{minimizes} \quad \phi - \langle z, \cdot \rangle \ .$$



Review: the subgradient method in Euclidean space

Seek to minimize convex objective $\phi: \mathbf{R}^n \to \mathbf{R}$.
At current point $x \in \mathbf{R}^n$,
$\partial\phi(x)$ consists of **subgradients** $z \in \mathbf{R}^n$:

$$x \quad \text{minimizes} \quad \phi - \langle z, \cdot \rangle \ .$$

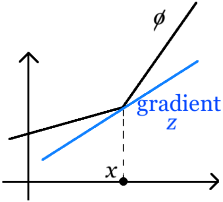


Move distance ϵ along **ray from x in direction z**: $\quad x \leftarrow x - \epsilon \frac{z}{|z|}$.

Review: the subgradient method in Euclidean space

Seek to minimize convex objective $\phi\colon \mathbf{R}^n \to \mathbf{R}$.
At current point $x \in \mathbf{R}^n$,
$\partial\phi(x)$ consists of **subgradients** $z \in \mathbf{R}^n$:

$$x \quad \text{minimizes} \quad \phi - \langle z, \cdot \rangle \ .$$

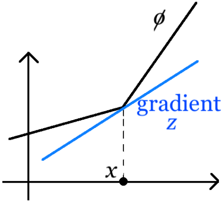


Move distance ϵ along **ray from x in direction z**: $\quad x \leftarrow x - \epsilon \frac{z}{|z|}$.
Iterating n times ensures objective excess $\phi(x) - \min \phi \ = \ O\left(\frac{1}{\sqrt{n}}\right)$.

Review: the subgradient method in Euclidean space

Seek to minimize convex objective $\phi\colon \mathbf{R}^n \to \mathbf{R}$.
At current point $x \in \mathbf{R}^n$,
$\partial\phi(x)$ consists of **subgradients** $z \in \mathbf{R}^n$:

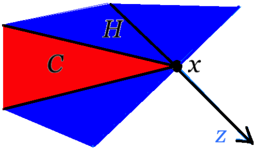
$$x \quad \text{minimizes} \quad \phi - \langle z, \cdot \rangle \ .$$



Move distance ϵ along **ray from x in direction z**: $\quad x \ \leftarrow \ x - \epsilon \frac{z}{|z|}$.

Iterating n times ensures objective excess $\phi(x) - \min \phi \ = \ O\Big(\frac{1}{\sqrt{n}}\Big)$.

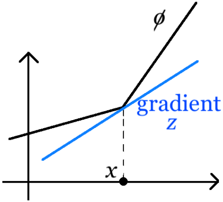
Underlying geometry:
level set C where $\phi \le \phi(x)$, and
halfspace H where $\langle z, \cdot \rangle \le \langle z, x \rangle$
satisfy $C \subset H$.



Review: the subgradient method in Euclidean space

Seek to minimize convex objective $\phi: \mathbf{R}^n \to \mathbf{R}$.
At current point $x \in \mathbf{R}^n$,
$\partial\phi(x)$ consists of **subgradients** $z \in \mathbf{R}^n$:

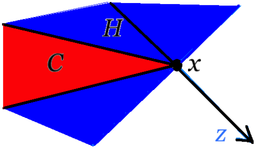
$$x \quad \text{minimizes} \quad \phi - \langle z, \cdot \rangle \ .$$



Move distance ϵ along **ray from x in direction z**: $\quad x \leftarrow x - \epsilon \frac{z}{|z|}$.

Iterating n times ensures objective excess $\phi(x) - \min \phi \; = \; O\Big(\frac{1}{\sqrt{n}}\Big)$.

Underlying geometry:
level set C where $\phi \le \phi(x)$, and
halfspace H where $\langle z, \cdot \rangle \le \langle z, x \rangle$
satisfy $C \subset H$.



Subgradients are **dual** objects. Any analogue in Hadamard space?

Rays, Busemann functions, and horoballs

Rays, Busemann functions, and horoballs

Rays from x are isometries $g\colon \mathbf{R}_+ \to \mathbf{X}$
with $g(0) = x$,

Rays, Busemann functions, and horoballs

Rays from x are isometries $g\colon \mathbf{R}_+ \to \mathbf{X}$ with $g(0) = x$, and correspond to **Busemann functions** (1955)

$$b_g(y) = \lim_{\tau \to +\infty} \big(d(g(\tau), y) - \tau \big)$$

Rays, Busemann functions, and horoballs

Rays from x are isometries $g\colon \mathbf{R}_+ \to \mathbf{X}$
with $g(0) = x$, and correspond to
Busemann functions (1955)

$$b_g(y) = \lim_{\tau \to +\infty} \Big(d(g(\tau), y) - \tau \Big)$$

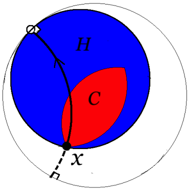
and **horoballs** H where $b_g \leq 0$.

Rays, Busemann functions, and horoballs

Rays from x are isometries $g\colon \mathbf{R}_+ \to \mathbf{X}$ with $g(0) = x$, and correspond to **Busemann functions** (1955)

$$b_g(y) = \lim_{\tau \to +\infty} \left(d(g(\tau), y) - \tau \right)$$

and **horoballs** H where $b_g \leq 0$.

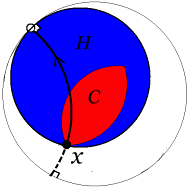


Rays, Busemann functions, and horoballs

Rays from x are isometries $g\colon \mathbf{R}_+ \to \mathbf{X}$ with $g(0) = x$, and correspond to **Busemann functions** (1955)

$$b_g(y) = \lim_{\tau \to +\infty} \Big(d(g(\tau), y) - \tau \Big)$$

and **horoballs** H where $b_g \le 0$.



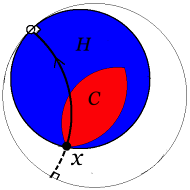
The ray **supports** objective $\phi\colon \mathbf{X} \to \mathbf{R}$ at x if H contains the **level set** C where $\phi \le \phi(x)$.

Rays, Busemann functions, and horoballs

Rays from x are isometries $g\colon \mathbf{R}_+ \to \mathbf{X}$ with $g(0) = x$, and correspond to **Busemann functions** (1955)

$$b_g(y) = \lim_{\tau \to +\infty} \big(d(g(\tau), y) - \tau \big)$$

and **horoballs** H where $b_g \leq 0$.

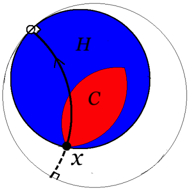


The ray **supports** objective $\phi\colon \mathbf{X} \to \mathbf{R}$ at x if H contains the **level set** C where $\phi \leq \phi(x)$.

Example: Euclidean space

Rays, Busemann functions, and horoballs

Rays from x are isometries $g\colon \mathbf{R}_+ \to \mathbf{X}$ with $g(0) = x$, and correspond to **Busemann functions** (1955)



$$b_g(y) = \lim_{\tau \to +\infty} \left(d(g(\tau), y) - \tau \right)$$

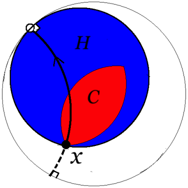
and **horoballs** H where $b_g \leq 0$.

The ray **supports** objective $\phi\colon \mathbf{X} \to \mathbf{R}$ at x if H contains the **level set** C where $\phi \leq \phi(x)$.

Example: Euclidean space   Rays are **halflines**,

Rays, Busemann functions, and horoballs

Rays from x are isometries $g\colon \mathbf{R}_+ \to \mathbf{X}$ with $g(0) = x$, and correspond to **Busemann functions** (1955)



$$b_g(y) = \lim_{\tau \to +\infty} \big(d(g(\tau), y) - \tau \big)$$

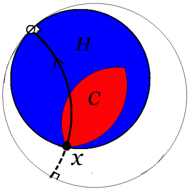
and **horoballs** H where $b_g \le 0$.

The ray **supports** objective $\phi\colon \mathbf{X} \to \mathbf{R}$ at x if H contains the **level set** C where $\phi \le \phi(x)$.

Example: Euclidean space   Rays are **halflines**, Busemann functions are **affine**,

Rays, Busemann functions, and horoballs

Rays from x are isometries $g\colon \mathbf{R}_+ \to \mathbf{X}$
with $g(0) = x$, and correspond to
Busemann functions (1955)



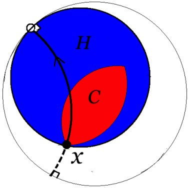
$$b_g(y) = \lim_{\tau \to +\infty} \big(d(g(\tau), y) - \tau \big)$$

and **horoballs** H where $b_g \le 0$.

The ray **supports** objective $\phi\colon \mathbf{X} \to \mathbf{R}$ at x
if H contains the **level set** C where $\phi \le \phi(x)$.

Example: Euclidean space   Rays are **halflines**,
Busemann functions are **affine**, horoballs are **halfspaces**,

Rays, Busemann functions, and horoballs



Rays from x are isometries $g\colon \mathbf{R}_+ \to \mathbf{X}$ with $g(0) = x$, and correspond to **Busemann functions** (1955)

$$b_g(y) = \lim_{\tau \to +\infty} \big(d(g(\tau), y) - \tau \big)$$

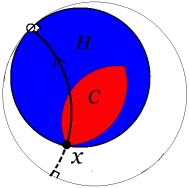
and **horoballs** H where $b_g \leq 0$.

The ray **supports** objective $\phi\colon \mathbf{X} \to \mathbf{R}$ at x if H contains the **level set** C where $\phi \leq \phi(x)$.

Example: Euclidean space Rays are **halflines**, Busemann functions are **affine**, horoballs are **halfspaces**, **subgradients** $z \in \partial\phi(x)$ give supporting rays $g(\tau) = x - \tau \frac{z}{|z|}$.

Rays, Busemann functions, and horoballs

Rays from x are isometries $g\colon \mathbf{R}_+ \to \mathbf{X}$ with $g(0) = x$, and correspond to **Busemann functions** (1955)



$$b_g(y) = \lim_{\tau \to +\infty} \big(d(g(\tau), y) - \tau \big)$$

and **horoballs** H where $b_g \le 0$.

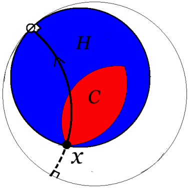
The ray **supports** objective $\phi\colon \mathbf{X} \to \mathbf{R}$ at x if H contains the **level set** C where $\phi \le \phi(x)$.

Example: Euclidean space Rays are **halflines**, Busemann functions are **affine**, horoballs are **halfspaces**, **subgradients** $z \in \partial\phi(x)$ give supporting rays $g(\tau) = x - \tau \frac{z}{|z|}$.

Example: circumenter $\phi(x) = \max_{a \in A} d(x, a)$

Rays, Busemann functions, and horoballs

Rays from x are isometries $g\colon \mathbf{R}_+ \to \mathbf{X}$ with $g(0) = x$, and correspond to **Busemann functions** (1955)



$$b_g(y) = \lim_{\tau\to+\infty} \big(d(g(\tau), y) - \tau\big)$$

and **horoballs** H where $b_g \leq 0$.

The ray **supports** objective $\phi\colon \mathbf{X} \to \mathbf{R}$ at x if H contains the **level set** C where $\phi \leq \phi(x)$.

Example: Euclidean space Rays are **halflines**, Busemann functions are **affine**, horoballs are **halfspaces**, **subgradients** $z \in \partial\phi(x)$ give supporting rays $g(\tau) = x - \tau\frac{z}{|z|}$.

Example: circumenter $\phi(x) = \max_{a\in A} d(x, a)$ is supported by rays from x through any maximizing point a.

A subgradient-style algorithm in Hadamard space

A subgradient-style algorithm in Hadamard space

Theorem (Complexity of supporting ray pursuit)

A subgradient-style algorithm in Hadamard space

Theorem (Complexity of supporting ray pursuit)

For L-Lipschitz objective ϕ

A subgradient-style algorithm in Hadamard space

Theorem (Complexity of supporting ray pursuit)

For L-Lipschitz objective ϕ *attaining its minimum,*

A subgradient-style algorithm in Hadamard space

Theorem (Complexity of supporting ray pursuit)

For L-Lipschitz objective ϕ attaining its minimum, initial point x_0 and any upper diameter bound D for $\{x : \phi(x) \leq \phi(x_0)\}$,

A subgradient-style algorithm in Hadamard space

Theorem (Complexity of supporting ray pursuit)

For L-Lipschitz objective ϕ *attaining its minimum, initial point* x_0 *and any upper diameter bound D for* $\{x : \phi(x) \leq \phi(x_0)\}$, *n successive steps of constant size* $\frac{D}{\sqrt{n}}$ *along supporting rays*

A subgradient-style algorithm in Hadamard space

Theorem (Complexity of supporting ray pursuit)

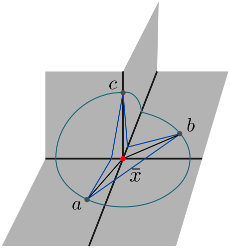
For L-Lipschitz objective ϕ attaining its minimum, initial point x_0 and any upper diameter bound D for $\{x : \phi(x) \leq \phi(x_0)\}$, n successive steps of constant size $\frac{D}{\sqrt{n}}$ along supporting rays guarantees average objective excess less than $\frac{LD}{\sqrt{n}}$.

A subgradient-style algorithm in Hadamard space

Theorem (Complexity of supporting ray pursuit)

For L-Lipschitz objective ϕ attaining its minimum, initial point x_0 and any upper diameter bound D for $\{x : \phi(x) \le \phi(x_0)\}$, n successive steps of constant size $\frac{D}{\sqrt{n}}$ along supporting rays guarantees average objective excess less than $\frac{LD}{\sqrt{n}}$.

Example: circumcenter of three points in a cubical complex

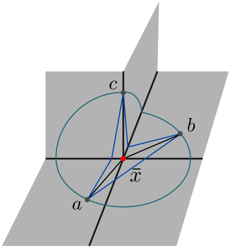


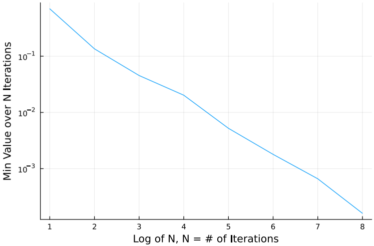
A subgradient-style algorithm in Hadamard space

Theorem (Complexity of supporting ray pursuit)

For L-Lipschitz objective ϕ attaining its minimum, initial point x_0 and any upper diameter bound D for $\{x : \phi(x) \le \phi(x_0)\}$, n successive steps of constant size $\frac{D}{\sqrt{n}}$ along supporting rays guarantees average objective excess less than $\frac{LD}{\sqrt{n}}$.

Example: circumcenter of three points in a cubical complex





But finding supporting rays for composite objectives is hard, so...

Incremental subgradient methods

Incremental subgradient methods

Euclidean space supports **subgradient calculus**:

Incremental subgradient methods

Euclidean space supports **subgradient calculus**: to minimize

$$\phi = \sum_{j=1}^{k} \phi_j$$

Incremental subgradient methods

Euclidean space supports **subgradient calculus**: to minimize

$$\phi = \sum_{j=1}^{k} \phi_j \qquad \text{we can use} \qquad \partial\phi(x) = \sum_j \partial\phi_j(x).$$

Incremental subgradient methods

Euclidean space supports **subgradient calculus**: to minimize

$$\phi = \sum_{j=1}^{k} \phi_j \qquad \text{we can use} \qquad \partial\phi(x) = \sum_j \partial\phi_j(x).$$

In Hadamard space, **subgradients** $(g, s) \in \partial\phi(x)$ consist of:

- **direction** of ray g from x with Busemann function b_g

Incremental subgradient methods

Euclidean space supports **subgradient calculus**: to minimize

$$\phi = \sum_{j=1}^{k} \phi_j \qquad \text{we can use} \qquad \partial\phi(x) = \sum_j \partial\phi_j(x).$$

In Hadamard space, **subgradients** $(g, s) \in \partial\phi(x)$ consist of:

- **direction** of ray g from x with Busemann function b_g
- **magnitude** (or "speed") $s \geq 0$

Incremental subgradient methods

Euclidean space supports **subgradient calculus**: to minimize

$$\phi = \sum_{j=1}^{k} \phi_j \qquad \text{we can use} \qquad \partial\phi(x) = \sum_j \partial\phi_j(x).$$

In Hadamard space, **subgradients** $(g, s) \in \partial\phi(x)$ consist of:

- **direction** of ray g from x with Busemann function b_g
- **magnitude** (or "speed") $s \geq 0$ such that x minimizes $\phi - sb_g$.

Incremental subgradient methods

Euclidean space supports **subgradient calculus**: to minimize

$$\phi = \sum_{j=1}^{k} \phi_j \qquad \text{we can use} \qquad \partial\phi(x) = \sum_j \partial\phi_j(x).$$

In Hadamard space, **subgradients** $(g, s) \in \partial\phi(x)$ consist of:

- **direction** of ray g from x with Busemann function b_g
- **magnitude** (or "speed") $s \geq 0$ such that x minimizes $\phi - sb_g$.

Example: Euclidean subgradient $z \neq 0$ gives Hadamard subgradient

$$(g, |z|) \in \partial\phi(x) \qquad \text{for the ray} \qquad g(\tau) = x - \tau\frac{z}{|z|}.$$

Incremental subgradient methods

Euclidean space supports **subgradient calculus**: to minimize

$$\phi = \sum_{j=1}^{k} \phi_j \qquad \text{we can use} \qquad \partial\phi(x) = \sum_j \partial\phi_j(x).$$

In Hadamard space, **subgradients** $(g, s) \in \partial\phi(x)$ consist of:

- **direction** of ray g from x with Busemann function b_g
- **magnitude** (or "speed") $s \geq 0$ such that x minimizes $\phi - sb_g$.

Example: Euclidean subgradient $z \neq 0$ gives Hadamard subgradient

$$(g, |z|) \in \partial\phi(x) \qquad \text{for the ray} \qquad g(\tau) = x - \tau\frac{z}{|z|}.$$

No calculus. . . .

Incremental subgradient methods

Euclidean space supports **subgradient calculus**: to minimize

$$\phi = \sum_{j=1}^{k} \phi_j \qquad \text{we can use} \qquad \partial\phi(x) = \sum_j \partial\phi_j(x).$$

In Hadamard space, **subgradients** $(g, s) \in \partial\phi(x)$ consist of:

- **direction** of ray g from x with Busemann function b_g
- **magnitude** (or "speed") $s \geq 0$ such that x minimizes $\phi - sb_g$.

Example: Euclidean subgradient $z \neq 0$ gives Hadamard subgradient

$$(g, |z|) \in \partial\phi(x) \qquad \text{for the ray} \qquad g(\tau) = x - \tau\frac{z}{|z|}.$$

No calculus. . . . However, following (Bertsekas-Nedic '01). . .

Incremental subgradient methods

Euclidean space supports **subgradient calculus**: to minimize

$$\phi = \sum_{j=1}^{k} \phi_j \qquad \text{we can use} \qquad \partial\phi(x) = \sum_j \partial\phi_j(x).$$

In Hadamard space, **subgradients** $(g, s) \in \partial\phi(x)$ consist of:

- **direction** of ray g from x with Busemann function b_g
- **magnitude** (or "speed") $s \geq 0$ such that x minimizes $\phi - sb_g$.

Example: Euclidean subgradient $z \neq 0$ gives Hadamard subgradient

$$(g, |z|) \in \partial\phi(x) \qquad \text{for the ray} \qquad g(\tau) = x - \tau\frac{z}{|z|}.$$

No calculus.... However, following (Bertsekas-Nedic '01)...

Theorem *For suitable $\delta > 0$, repeating for cycle $n = 1 \ldots, N$*

Incremental subgradient methods

Euclidean space supports **subgradient calculus**: to minimize

$$\phi = \sum_{j=1}^{k} \phi_j \qquad \text{we can use} \qquad \partial\phi(x) = \sum_j \partial\phi_j(x).$$

In Hadamard space, **subgradients** $(g, s) \in \partial\phi(x)$ consist of:

- **direction** of ray g from x with Busemann function b_g
- **magnitude** (or "speed") $s \geq 0$ such that x minimizes $\phi - sb_g$.

Example: Euclidean subgradient $z \neq 0$ gives Hadamard subgradient

$$(g, |z|) \in \partial\phi(x) \qquad \text{for the ray} \qquad g(\tau) = x - \tau\frac{z}{|z|}.$$

No calculus. . . . However, following (Bertsekas-Nedic '01). . .

Theorem *For suitable $\delta > 0$, repeating for cycle $n = 1 \ldots, N$*

$$\textbf{for } j = 1 \ldots k \quad \textit{find} \quad (g, s) \in \partial\phi_j(x) \quad \textit{and update} \quad x \leftarrow g\Big(\frac{\delta s}{\sqrt{n}}\Big)$$

Incremental subgradient methods

Euclidean space supports **subgradient calculus**: to minimize

$$\phi = \sum_{j=1}^{k} \phi_j \qquad \text{we can use} \qquad \partial\phi(x) = \sum_j \partial\phi_j(x).$$

In Hadamard space, **subgradients** $(g, s) \in \partial\phi(x)$ consist of:

- **direction** of ray g from x with Busemann function b_g
- **magnitude** (or "speed") $s \geq 0$ such that x minimizes $\phi - sb_g$.

Example: Euclidean subgradient $z \neq 0$ gives Hadamard subgradient

$$(g, |z|) \in \partial\phi(x) \qquad \text{for the ray} \qquad g(\tau) = x - \tau\frac{z}{|z|}.$$

No calculus. . . . However, following (Bertsekas-Nedic '01). . .

Theorem *For suitable $\delta > 0$, repeating for cycle $n = 1\ldots, N$*

$$\textbf{for } j = 1\ldots k \quad \textit{find} \quad (g, s) \in \partial\phi_j(x) \quad \textit{and update} \quad x \leftarrow g\Big(\frac{\delta s}{\sqrt{n}}\Big)$$

ensures objective excess $O(\frac{1}{\sqrt{N}})$.

Example in tree space (Billera-Holmes-Vogtmann '01)

Example in tree space (Billera-Holmes-Vogtmann '01)

BHV tree space consists of binary trees with weighted internal edges,

Example in tree space (Billera-Holmes-Vogtmann '01)

BHV tree space consists of binary trees with weighted internal edges, living in a CAT(0) orthant complex

Example in tree space (Billera-Holmes-Vogtmann '01)

BHV tree space consists of binary trees with weighted internal edges, living in a CAT(0) orthant complex with polytime computable geodesics (Owen-Provan '11)

Example in tree space (Billera-Holmes-Vogtmann '01)

BHV tree space consists of binary trees with weighted internal edges, living in a CAT(0) orthant complex with polytime computable geodesics (Owen-Provan '11) implemented in `SturmMean`.

Example in tree space (Billera-Holmes-Vogtmann '01)

BHV tree space consists of binary trees with weighted internal edges, living in a CAT(0) orthant complex with polytime computable geodesics (Owen-Provan '11) implemented in `SturmMean`.

Medians of finite sets A in Hadamard spaces $\mathbf{X}$

Example in tree space (Billera-Holmes-Vogtmann '01)

BHV tree space consists of binary trees with weighted internal edges, living in a CAT(0) orthant complex with polytime computable geodesics (Owen-Provan '11) implemented in `SturmMean`.

Medians of finite sets A in Hadamard spaces $\mathbf{X}$ minimize $\sum_{a \in A} d_a$.

Example in tree space (Billera-Holmes-Vogtmann '01)

BHV tree space consists of binary trees with weighted internal edges, living in a CAT(0) orthant complex with polytime computable geodesics (Owen-Provan '11) implemented in `SturmMean`.

Medians of finite sets A in Hadamard spaces $\mathbf{X}$ minimize $\sum_{a \in A} d_a$. To run incremental subgradient method

Example in tree space (Billera-Holmes-Vogtmann '01)

BHV tree space consists of binary trees with weighted internal edges, living in a CAT(0) orthant complex with polytime computable geodesics (Owen-Provan '11) implemented in `SturmMean`.

Medians of finite sets A in Hadamard spaces $\mathbf{X}$ minimize $\sum_{a \in A} d_a$. To run incremental subgradient method note rays g from x through $a \in A$ satisfy $(g, 1) \in \partial d_a(x)$.

Example in tree space (Billera-Holmes-Vogtmann '01)

BHV tree space consists of binary trees with weighted internal edges, living in a CAT(0) orthant complex with polytime computable geodesics (Owen-Provan '11) implemented in `SturmMean`.

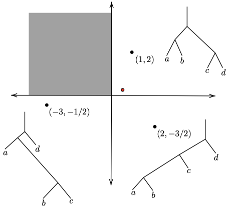
Medians of finite sets A in Hadamard spaces $\mathbf{X}$ minimize $\sum_{a \in A} d_a$. To run incremental subgradient method note rays g from x through $a \in A$ satisfy $(g, 1) \in \partial d_a(x)$.

Example (Owen-Miller-Provan '15)

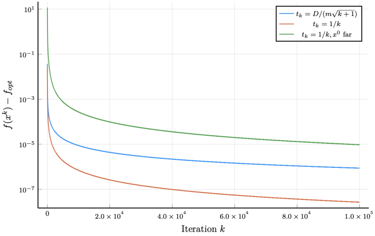
A median of three trees...

Example in tree space (Billera-Holmes-Vogtmann '01)

BHV tree space consists of binary trees with weighted internal edges, living in a CAT(0) orthant complex with polytime computable geodesics (Owen-Provan '11) implemented in `SturmMean`.



Medians of finite sets A in Hadamard spaces $\mathbf{X}$ minimize $\sum_{a \in A} d_a$. To run incremental subgradient method note rays g from x through $a \in A$ satisfy $(g, 1) \in \partial d_a(x)$.

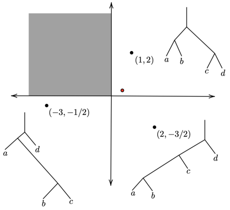


Example (Owen-Miller-Provan '15)

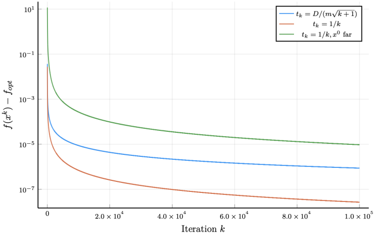
A median of three trees...

Example in tree space (Billera-Holmes-Vogtmann '01)

BHV tree space consists of binary trees with weighted internal edges, living in a CAT(0) orthant complex with polytime computable geodesics (Owen-Provan '11) implemented in `SturmMean`.



Medians of finite sets A in Hadamard spaces $\mathbf{X}$ minimize $\sum_{a \in A} d_a$. To run incremental subgradient method note rays g from x through $a \in A$ satisfy $(g, 1) \in \partial d_a(x)$.



Example (Owen-Miller-Provan '15)

A median of three trees. . .

(See Ariel Goodwin's poster: "Incremental minimization. . .")

PART III

Weighted means in Hadamard spaces

Averaging finite sets

Averaging finite sets

Examples: matrix means (Bhatia. . . '12),

Averaging finite sets

Examples: matrix means (Bhatia. . . '12), computational geometry (Arnaudon. . . '05),

Averaging finite sets

Examples: matrix means (Bhatia. . . '12), computational geometry (Arnaudon. . . '05), phylogenetic trees (Billera. . . '01, Bačák '13). . .

Averaging finite sets

Examples: matrix means (Bhatia. . . '12), computational geometry (Arnaudon. . . '05), phylogenetic trees (Billera. . . '01, Bačák '13). . .

Question: given a finite set A, is a given point $\bar{x}$ an "average"?

Averaging finite sets

Examples: matrix means (Bhatia. . . '12), computational geometry (Arnaudon. . . '05), phylogenetic trees (Billera. . . '01, Bačák '13). . .

Question: given a finite set A, is a given point $\bar{x}$ an "average"?

Theorem (weighted averages versus convex combinations)

Averaging finite sets

Examples: matrix means (Bhatia. . . '12), computational geometry (Arnaudon. . . '05), phylogenetic trees (Billera. . . '01, Bačák '13). . .

Question: given a finite set A, is a given point $\bar{x}$ an "average"?

Theorem (weighted averages versus convex combinations)

A point $\bar{x}$ *minimizes* $\sum_a w_a d^2(\cdot, a)$ *for some weighting* $0 \neq w \geq 0$

Averaging finite sets

Examples: matrix means (Bhatia. . . '12), computational geometry (Arnaudon. . . '05), phylogenetic trees (Billera. . . '01, Bačák '13). . .

Question: given a finite set A, is a given point $\bar{x}$ an "average"?

Theorem (weighted averages versus convex combinations)

A point $\bar{x}$ minimizes $\sum_a w_a d^2(\cdot, a)$ for some weighting $0 \neq w \geq 0$ if and only if $\bar{x}$ minimizes **test function** $\max_a\{d(\cdot, a) - d(\bar{x}, a)\}$.

Averaging finite sets

Examples: matrix means (Bhatia. . . '12), computational geometry (Arnaudon. . . '05), phylogenetic trees (Billera. . . '01, Bačák '13). . .

Question: given a finite set A, is a given point $\bar{x}$ an "average"?

Theorem (weighted averages versus convex combinations)

A point $\bar{x}$ *minimizes* $\sum_a w_a d^2(\cdot, a)$ *for some weighting* $0 \neq w \geq 0$ *if and only if* $\bar{x}$ *minimizes* **test function** $\max_a \{d(\cdot, a) - d(\bar{x}, a)\}$.
In that case, $\bar{x} \in \operatorname{conv} A$

Averaging finite sets

Examples: matrix means (Bhatia. . . '12), computational geometry (Arnaudon. . . '05), phylogenetic trees (Billera. . . '01, Bačák '13). . .

Question: given a finite set A, is a given point $\bar{x}$ an "average"?

Theorem (weighted averages versus convex combinations)

A point $\bar{x}$ minimizes $\sum_a w_a d^2(\cdot, a)$ for some weighting $0 \neq w \geq 0$ if and only if $\bar{x}$ minimizes **test function** $\max_a\{d(\cdot, a) - d(\bar{x}, a)\}$.
In that case, $\bar{x} \in \operatorname{conv} A$ (but the converse fails outside $\mathbf{R}^n$).

Averaging finite sets

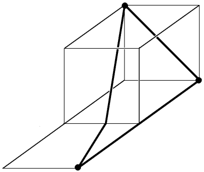
Examples: matrix means (Bhatia. . . '12), computational geometry (Arnaudon. . . '05), phylogenetic trees (Billera. . . '01, Bačák '13). . .

Question: given a finite set A, is a given point $\bar{x}$ an "average"?

Theorem (weighted averages versus convex combinations)

A point $\bar{x}$ minimizes $\sum_a w_a d^2(\cdot, a)$ for some weighting $0 \neq w \geq 0$ if and only if $\bar{x}$ minimizes **test function** $\max_a\{d(\cdot, a) - d(\bar{x}, a)\}$*. In that case, $\bar{x} \in \operatorname{conv} A$ (but the converse fails outside $\mathbf{R}^n$).*

Three points in a cubical complex



Averaging finite sets

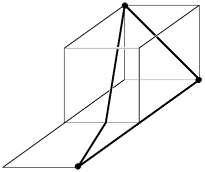
Examples: matrix means (Bhatia. . . '12), computational geometry (Arnaudon. . . '05), phylogenetic trees (Billera. . . '01, Bačák '13). . .

Question: given a finite set A, is a given point $\bar{x}$ an "average"?

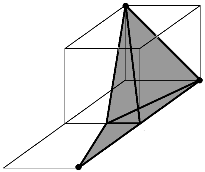
Theorem (weighted averages versus convex combinations)

A point $\bar{x}$ minimizes $\sum_a w_a d^2(\cdot, a)$ for some weighting $0 \neq w \geq 0$ if and only if $\bar{x}$ minimizes **test function** $\max_a\{d(\cdot, a) - d(\bar{x}, a)\}$*.*
In that case, $\bar{x} \in \operatorname{conv} A$ (but the converse fails outside $\mathbf{R}^n$).

Three points in a cubical complex



with 3-dimensional convex hull. . .



Averaging finite sets

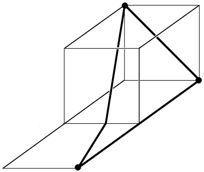
Examples: matrix means (Bhatia. . . '12), computational geometry (Arnaudon. . . '05), phylogenetic trees (Billera. . . '01, Bačák '13). . .

Question: given a finite set A, is a given point $\bar{x}$ an "average"?

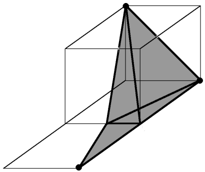
Theorem (weighted averages versus convex combinations)

A point $\bar{x}$ minimizes $\sum_a w_a d^2(\cdot, a)$ for some weighting $0 \neq w \geq 0$ if and only if $\bar{x}$ minimizes **test function** $\max_a\{d(\cdot, a) - d(\bar{x}, a)\}$. *In that case, $\bar{x} \in \operatorname{conv} A$ (but the converse fails outside $\mathbf{R}^n$).*

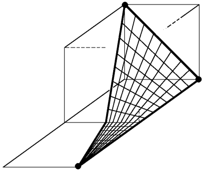
Three points in a cubical complex



with 3-dimensional convex hull. . .



. . . but 2-dimensional weighted mean set.



Recognizing weighted means computationally

Recognizing weighted means computationally

For $\bar{x} \in \mathbf{X}$ (locally compact Hadamard),

Recognizing weighted means computationally

For $\bar{x} \in \mathbf{X}$ (locally compact Hadamard), assume geodesics $[\bar{x}, a]$ extend to rays.

Recognizing weighted means computationally

For $\bar{x} \in \mathbf{X}$ (locally compact Hadamard), assume geodesics $[\bar{x}, a]$ extend to rays. Traversing at speed $d(\bar{x}, a)$ gives **tangents** $\underline{[\bar{x}, a]}$

Recognizing weighted means computationally

For $\bar{x} \in \mathbf{X}$ (locally compact Hadamard), assume geodesics $[\bar{x}, a]$ extend to rays. Traversing at speed $d(\bar{x}, a)$ gives **tangents** $\underline{[\bar{x}, a]}$ comprising "cone" $T_{\bar{x}}\mathbf{X}$.

Recognizing weighted means computationally

For $\bar{x} \in \mathbf{X}$ (locally compact Hadamard), assume geodesics $[\bar{x}, a]$ extend to rays. Traversing at speed $d(\bar{x}, a)$ gives **tangents** $\underline{[\bar{x}, a]}$ comprising "cone" $T_{\bar{x}}\mathbf{X}$. Suppose $T_{\bar{x}}\mathbf{X} \cong \mathbf{R}^n$.

Recognizing weighted means computationally

For $\bar{x} \in \mathbf{X}$ (locally compact Hadamard), assume geodesics $[\bar{x}, a]$ extend to rays. Traversing at speed $d(\bar{x}, a)$ gives **tangents** $\overline{[\bar{x}, a]}$ comprising "cone" $T_{\bar{x}}\mathbf{X}$. Suppose $T_{\bar{x}}\mathbf{X} \cong \mathbf{R}^n$. (Eg: $\mathbf{X}$ a manifold.)

Recognizing weighted means computationally

For $\bar{x} \in \mathbf{X}$ (locally compact Hadamard), assume geodesics $[\bar{x}, a]$ extend to rays. Traversing at speed $d(\bar{x}, a)$ gives **tangents** $\overline{[\bar{x}, a]}$ comprising "cone" $T_{\bar{x}}\mathbf{X}$. Suppose $T_{\bar{x}}\mathbf{X} \cong \mathbf{R}^n$. (Eg: $\mathbf{X}$ a manifold.)

The test function $\max_{a \in A}\{d(\cdot, a) - d(\bar{x}, a)\}$

Recognizing weighted means computationally

For $\bar{x} \in \mathbf{X}$ (locally compact Hadamard), assume geodesics $[\bar{x}, a]$ extend to rays. Traversing at speed $d(\bar{x}, a)$ gives **tangents** $\underline{[\bar{x}, a]}$ comprising "cone" $T_{\bar{x}}\mathbf{X}$. Suppose $T_{\bar{x}}\mathbf{X} \cong \mathbf{R}^n$. (Eg: $\mathbf{X}$ a manifold.)

The test function $\max_{a \in A}\{d(\cdot, a) - d(\bar{x}, a)\}$ then has **slope**

$$\mathrm{dist}\Big(0, \mathrm{conv}\big\{\underline{[\bar{x}, a]} : a \in A\big\}\Big),$$

Recognizing weighted means computationally

For $\bar{x} \in \mathbf{X}$ (locally compact Hadamard), assume geodesics $[\bar{x}, a]$ extend to rays. Traversing at speed $d(\bar{x}, a)$ gives **tangents** $\underline{[\bar{x}, a]}$ comprising "cone" $T_{\bar{x}}\mathbf{X}$. Suppose $T_{\bar{x}}\mathbf{X} \cong \mathbf{R}^n$. (Eg: $\mathbf{X}$ a manifold.)

The test function $\max_{a \in A}\{d(\cdot, a) - d(\bar{x}, a)\}$ then has **slope**

$$\mathrm{dist}\Big(0, \mathrm{conv}\big\{\underline{[\bar{x}, a]} : a \in A\big\}\Big),$$

quantifying whether $\bar{x}$ is a weighted mean.

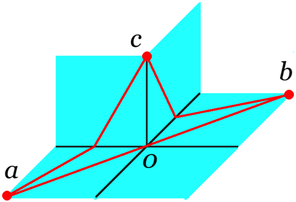
Recognizing weighted means computationally

For $\bar{x} \in \mathbf{X}$ (locally compact Hadamard), assume geodesics $[\bar{x}, a]$ extend to rays. Traversing at speed $d(\bar{x}, a)$ gives **tangents** $\underline{[\bar{x}, a]}$ comprising "cone" $T_{\bar{x}}\mathbf{X}$. Suppose $T_{\bar{x}}\mathbf{X} \cong \mathbf{R}^n$. (Eg: $\mathbf{X}$ a manifold.)

The test function $\max_{a \in A}\{d(\cdot, a) - d(\bar{x}, a)\}$ then has **slope**

$$\operatorname{dist}\Big(0, \operatorname{conv}\big\{\underline{[\bar{x}, a]} : a \in A\big\}\Big),$$

quantifying whether $\bar{x}$ is a weighted mean.



3 points in a cubical complex

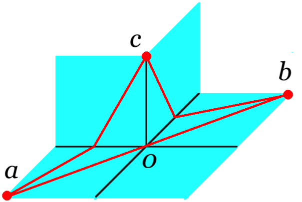
Recognizing weighted means computationally

For $\bar{x} \in \mathbf{X}$ (locally compact Hadamard), assume geodesics $[\bar{x}, a]$ extend to rays. Traversing at speed $d(\bar{x}, a)$ gives **tangents** $\underline{[\bar{x}, a]}$ comprising "cone" $T_{\bar{x}}\mathbf{X}$. Suppose $T_{\bar{x}}\mathbf{X} \cong \mathbf{R}^n$. (Eg: $\mathbf{X}$ a manifold.)

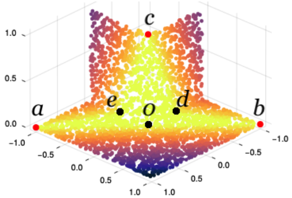
The test function $\max_{a \in A}\{d(\cdot, a) - d(\bar{x}, a)\}$ then has **slope**

$$\operatorname{dist}\Big(0, \operatorname{conv}\big\{\underline{[\bar{x}, a]} : a \in A\big\}\Big),$$

quantifying whether $\bar{x}$ is a weighted mean.



3 points in a cubical complex



Heat map of test function slope

PART IV

Convex optimization on CAT(0) cubical complexes

CAT(0) cubical complexes

CAT(0) cubical complexes

A **complex X** of **cells** — Euclidean **cubes** and their faces —

CAT(0) cubical complexes

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

CAT(0) cubical complexes

A **complex X** of **cells** — Euclidean **cubes** and their faces — each pair of cells sharing at most one face — is **CAT(0)** $\Leftrightarrow$ simply connected and **link condition** holds (Gromov '81).

CAT(0) cubical complexes

A **complex X** of **cells** — Euclidean **cubes** and their faces — each pair of cells sharing at most one face — is **CAT(0)** $\Leftrightarrow$ simply connected and **link condition** holds (Gromov '81).
Geodesics are **polynomial time** (Owen-Provan '11 . . . Hayashi '21)

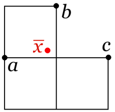
CAT(0) cubical complexes

A **complex** **X** of **cells** — Euclidean **cubes** and their faces — each pair of cells sharing at most one face — is **CAT(0)** $\Leftrightarrow$ simply connected and **link condition** holds (Gromov '81).
Geodesics are **polynomial time** (Owen-Provan '11 . . . Hayashi '21)

Example: Minimize over square complex

$$\phi(x) = d^2(x, a) + d^2(x, b) + d^2(x, c).$$

Cyclic proximal point (Bačák '13) is slow:

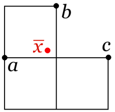


CAT(0) cubical complexes

A **complex X** of **cells** — Euclidean **cubes** and their faces — each pair of cells sharing at most one face — is **CAT(0)** $\Leftrightarrow$ simply connected and **link condition** holds (Gromov '81).
Geodesics are **polynomial time** (Owen-Provan '11 . . . Hayashi '21)

Example: Minimize over square complex

$$\phi(x) = d^2(x,a) + d^2(x,b) + d^2(x,c).$$



Cyclic proximal point (Bačák '13) is slow:

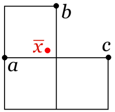
for $n = 1, 2, \ldots,$ $\quad x \in \frac{n}{n+1}x + \frac{1}{n+1}\{a, b, c\}$ $\quad$ **end(for)**

CAT(0) cubical complexes

A **complex X** of **cells** — Euclidean **cubes** and their faces — each pair of cells sharing at most one face — is **CAT(0)** $\Leftrightarrow$ simply connected and **link condition** holds (Gromov '81).
Geodesics are **polynomial time** (Owen-Provan '11 . . . Hayashi '21)

Example: Minimize over square complex

$$\phi(x) \;=\; d^2(x,a) + d^2(x,b) + d^2(x,c).$$



Cyclic proximal point (Bačák '13) is slow:

for $n = 1, 2, \ldots,$ $\quad x \in \frac{n}{n+1}x + \frac{1}{n+1}\{a, b, c\}$ $\quad$ **end(for)** $\quad$ Instead. . .

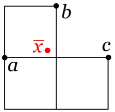
CAT(0) cubical complexes

A **complex** **X** of **cells** — Euclidean **cubes** and their faces — each pair of cells sharing at most one face — is **CAT(0)** $\Leftrightarrow$ simply connected and **link condition** holds (Gromov '81).
Geodesics are **polynomial time** (Owen-Provan '11 ... Hayashi '21)

Example: Minimize over square complex

$\phi(x) = d^2(x,a) + d^2(x,b) + d^2(x,c)$.

Cyclic proximal point (Bačák '13) is slow:



for $n = 1, 2, \ldots,$ $\quad x \in \frac{n}{n+1}x + \frac{1}{n+1}\{a, b, c\}$ **end(for)** Instead...

Algorithm: given current point x and list Ω of optimized cells

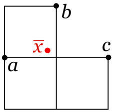
CAT(0) cubical complexes

A **complex X** of **cells** — Euclidean **cubes** and their faces — each pair of cells sharing at most one face — is **CAT(0)** $\Leftrightarrow$ simply connected and **link condition** holds (Gromov '81).
Geodesics are **polynomial time** (Owen-Provan '11 . . . Hayashi '21)

Example: Minimize over square complex

$$\phi(x) \ = \ d^2(x,a) + d^2(x,b) + d^2(x,c).$$

Cyclic proximal point (Bačák '13) is slow:



for $n = 1, 2, \ldots,$ $\quad x \in \frac{n}{n+1}x + \frac{1}{n+1}\{a, b, c\}$ **end(for)** Instead. . .

Algorithm: given current point x and list Ω of optimized cells
repeat

choose cell $P \notin \Omega$ containing x

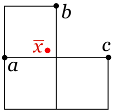
CAT(0) cubical complexes

A **complex** **X** of **cells** — Euclidean **cubes** and their faces — each pair of cells sharing at most one face — is **CAT(0)** $\Leftrightarrow$ simply connected and **link condition** holds (Gromov '81).
Geodesics are **polynomial time** (Owen-Provan '11 . . . Hayashi '21)

Example: Minimize over square complex

$$\phi(x) \;=\; d^2(x,a) + d^2(x,b) + d^2(x,c).$$

Cyclic proximal point (Bačák '13) is slow:



for $n = 1, 2, \ldots,$ $\quad x \in \frac{n}{n+1}x + \frac{1}{n+1}\{a, b, c\}$ **end(for)** Instead. . .

Algorithm: given current point x and list Ω of optimized cells
repeat

choose cell $P \notin \Omega$ containing x

$x = \text{argmin}_P\, \phi$

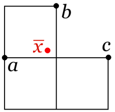
CAT(0) cubical complexes

A **complex** $\mathbf{X}$ of **cells** — Euclidean **cubes** and their faces — each pair of cells sharing at most one face — is **CAT(0)** $\Leftrightarrow$ simply connected and **link condition** holds (Gromov '81).
Geodesics are **polynomial time** (Owen-Provan '11 ... Hayashi '21)

Example: Minimize over square complex

$\phi(x) \; = \; d^2(x,a) + d^2(x,b) + d^2(x,c)$.

Cyclic proximal point (Bačák '13) is slow:



for $n = 1, 2, \ldots,$ $\quad x \in \frac{n}{n+1} x + \frac{1}{n+1} \{a, b, c\}$ $\quad$ **end(for)** $\quad$ Instead...

Algorithm: given current point x and list Ω of optimized cells

repeat

choose cell $P \notin \Omega$ containing x

$x = \operatorname{argmin}_P \phi$ $\qquad$ **How?**

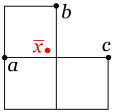
CAT(0) cubical complexes

A **complex** $\mathbf{X}$ of **cells** — Euclidean **cubes** and their faces — each pair of cells sharing at most one face — is **CAT(0)** $\Leftrightarrow$ simply connected and **link condition** holds (Gromov '81).
Geodesics are **polynomial time** (Owen-Provan '11 ... Hayashi '21)

Example: Minimize over square complex

$\phi(x) \;=\; d^2(x,a) + d^2(x,b) + d^2(x,c).$

Cyclic proximal point (Bačák '13) is slow:



for $n = 1, 2, \ldots,$ $\quad x \in \frac{n}{n+1}x + \frac{1}{n+1}\{a, b, c\}$ **end(for)** Instead...

Algorithm: given current point x and list Ω of optimized cells
repeat

choose cell $P \notin \Omega$ containing x
$x = \operatorname{argmin}_P \phi$ **How?**
update Ω

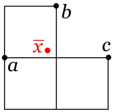
CAT(0) cubical complexes

A **complex** $\mathbf{X}$ of **cells** — Euclidean **cubes** and their faces — each pair of cells sharing at most one face — is **CAT(0)** $\Leftrightarrow$ simply connected and **link condition** holds (Gromov '81).
Geodesics are **polynomial time** (Owen-Provan '11 ... Hayashi '21)

Example: Minimize over square complex

$\phi(x) \;=\; d^2(x,a) + d^2(x,b) + d^2(x,c).$

Cyclic proximal point (Bačák '13) is slow:



for $n = 1, 2, \ldots,$ $\quad x \in \frac{n}{n+1}x + \frac{1}{n+1}\{a, b, c\}$ $\quad$ **end(for)** $\quad$ Instead...

Algorithm: given current point x and list Ω of optimized cells
repeat
  choose cell $P \notin \Omega$ containing x
  $x = \operatorname{argmin}_P \phi$ **How?**
  update Ω
until done

Cutting plane strategy for low-dimensional cells

Cutting plane strategy for low-dimensional cells

In CAT(0) cubical complex, consider points $a \neq x \in$ cell $P \subset \mathbf{R}^n$.

Aim: at x, find a (Euclidean) **subgradient** for the function

$$w \in \mathbf{R}^n \;\mapsto\; \begin{cases} d(w,a) & (w \in P) \\ +\infty & (w \notin P) \end{cases}$$

Cutting plane strategy for low-dimensional cells

In CAT(0) cubical complex, consider points $a \neq x \in$ cell $P \subset \mathbf{R}^n$.

Aim: at x, find a (Euclidean) **subgradient** for the function

$$w \in \mathbf{R}^n \;\mapsto\; \begin{cases} d(w,a) & (w \in P) \\ +\infty & (w \notin P) \end{cases}$$

Answer. Suppose geodesic $[x, a]$

Cutting plane strategy for low-dimensional cells

In CAT(0) cubical complex, consider points $a \neq x \in$ cell $P \subset \mathbf{R}^n$.

Aim: at x, find a (Euclidean) **subgradient** for the function

$$w \in \mathbf{R}^n \;\mapsto\; \begin{cases} d(w,a) & (w \in P) \\ +\infty & (w \notin P) \end{cases}$$

Answer. Suppose geodesic $[x,a]$
has first segment $[x,y]$

Cutting plane strategy for low-dimensional cells

In CAT(0) cubical complex, consider points $a \neq x \in$ cell $P \subset \mathbf{R}^n$.

Aim: at x, find a (Euclidean) **subgradient** for the function

$$w \in \mathbf{R}^n \mapsto \begin{cases} d(w,a) & (w \in P) \\ +\infty & (w \notin P) \end{cases}$$

Answer. Suppose geodesic $[x,a]$
has first segment $[x,y]$ in cell Q.

Cutting plane strategy for low-dimensional cells

In CAT(0) cubical complex, consider points $a \neq x \in$ cell $P \subset \mathbf{R}^n$.

Aim: at x, find a (Euclidean) **subgradient** for the function

$$w \in \mathbf{R}^n \mapsto \begin{cases} d(w,a) & (w \in P) \\ +\infty & (w \notin P) \end{cases}$$

Answer. Suppose geodesic $[x,a]$
has first segment $[x,y]$ in cell Q.
Project y onto its nearest point z
in the face F shared by P and Q.

Cutting plane strategy for low-dimensional cells

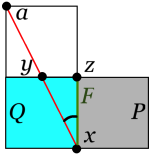
In CAT(0) cubical complex, consider points $a \neq x \in$ cell $P \subset \mathbf{R}^n$.

Aim: at x, find a (Euclidean) **subgradient** for the function

$$w \in \mathbf{R}^n \mapsto \begin{cases} d(w,a) & (w \in P) \\ +\infty & (w \notin P) \end{cases}$$

Answer. Suppose geodesic $[x,a]$ has first segment $[x,y]$ in cell Q. Project y onto its nearest point z in the face F shared by P and Q. Then, one such subgradient is

$$\frac{\cos(\angle yxz)}{|x-z|}(x-z)$$



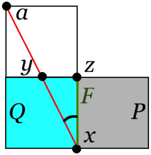
Cutting plane strategy for low-dimensional cells

In CAT(0) cubical complex, consider points $a \neq x \in$ cell $P \subset \mathbf{R}^n$.

Aim: at x, find a (Euclidean) **subgradient** for the function

$$w \in \mathbf{R}^n \;\mapsto\; \begin{cases} d(w,a) & (w \in P) \\ +\infty & (w \notin P) \end{cases}$$

Answer. Suppose geodesic $[x,a]$ has first segment $[x,y]$ in cell Q. Project y onto its nearest point z in the face F shared by P and Q. Then, one such subgradient is



$$\frac{\cos(\angle yxz)}{|x-z|}(x-z) \qquad \text{(interpreted as 0 if } z = x\text{).}$$

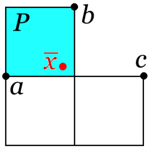
Example: means via cutting planes

Example: means via cutting planes

Minimize over cell P

$d^2(\cdot, a) + d^2(\cdot, b) + d^2(\cdot, c)$

using various cutting plane algorithms, compared with the cyclic proximal point method (Bačák '13).

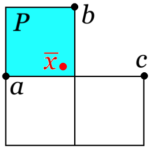


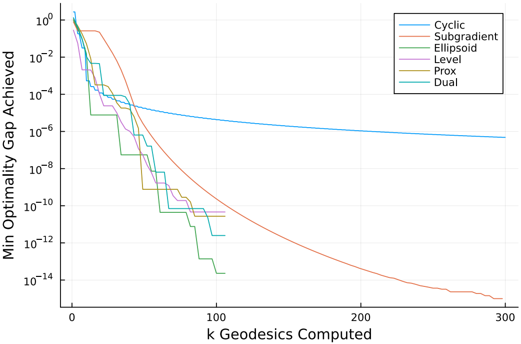
Example: means via cutting planes

Minimize over cell P

$d^2(\cdot, a) + d^2(\cdot, b) + d^2(\cdot, c)$

using various cutting plane algorithms, compared with the cyclic proximal point method (Bačák '13).





Summary

Summary

- Convex optimization in Hadamard spaces
  models a range of interesting applications.
- Even without linear or smooth structure, or dual spaces,
  classical convex optimization extends surprisingly well.

Summary

- Convex optimization in Hadamard spaces
  models a range of interesting applications.
- Even without linear or smooth structure, or dual spaces,
  classical convex optimization extends surprisingly well.

References

- Horoballs and the subgradient method
  `arXiv:2403.15749`
- Recognizing weighted means in geodesic spaces
  `arXiv:2406.03913`
- Convex optimization on CAT(0) cubical complexes
  `arXiv:2405.01968`
- Incremental minimization in spaces of nonpositive curvature
  `arielgoodwin.github.io/talks`