Slope and geometry in variational mathematics

Dmitriy Drusvyatskiy
Combinatorics & Optimization, University of Waterloo

Joint work with
A. Daniilidis (Chile), A.D. Ioffe (Technion), A.S. Lewis (Cornell)

Oct 5, 2013

Fix a metric space $(\mathcal{X}, d)$ and a function $f\colon \mathcal{X} \to \overline{\mathbf{R}}$.

Fix a metric space $(\mathcal{X}, d)$ and a function $f\colon \mathcal{X} \to \overline{\mathbf{R}}$.

Slope: "Fastest instantaneous rate of decrease"

Fix a metric space $(\mathcal{X}, d)$ and a function $f\colon \mathcal{X} \to \overline{\mathbf{R}}$.

Slope: "Fastest instantaneous rate of decrease"

$$\boxed{|\nabla f|(\bar{x}) := \limsup_{x \to \bar{x}} \frac{f(\bar{x}) - f(x)}{d(\bar{x}, x)}}$$

Fix a metric space $(\mathcal{X}, d)$ and a function $f\colon \mathcal{X} \to \overline{\mathbf{R}}$.

Slope: "Fastest instantaneous rate of decrease"

$$\boxed{|\nabla f|(\bar{x}) := \limsup_{x \to \bar{x}} \frac{f(\bar{x}) - f(x)}{d(\bar{x}, x)}}$$

Limiting slope:

$$\boxed{\overline{|\nabla f|}(\bar{x}) := \liminf_{x \xrightarrow[f]{} \bar{x}} |\nabla f|(x)}$$

Fix a metric space $(\mathcal{X}, d)$ and a function $f\colon \mathcal{X} \to \overline{\mathbf{R}}$.

Slope: "Fastest instantaneous rate of decrease"

$$\boxed{|\nabla f|(\bar{x}) := \limsup_{x \to \bar{x}} \frac{f(\bar{x}) - f(x)}{d(\bar{x}, x)}}$$

Limiting slope:

$$\boxed{\overline{|\nabla f|}(\bar{x}) := \liminf_{x \xrightarrow[f]{} \bar{x}} |\nabla f|(x)}$$

Critical points:

$$\boxed{\bar{x} \text{ is critical for } f \quad \Longleftrightarrow \quad \overline{|\nabla f|}(\bar{x}) = 0}$$

Fix a metric space $(\mathcal{X}, d)$ and a function $f\colon \mathcal{X} \to \overline{\mathbf{R}}$.

Slope: "Fastest instantaneous rate of decrease"

$$\boxed{|\nabla f|(\bar{x}) := \limsup_{x \to \bar{x}} \frac{f(\bar{x}) - f(x)}{d(\bar{x}, x)}}$$

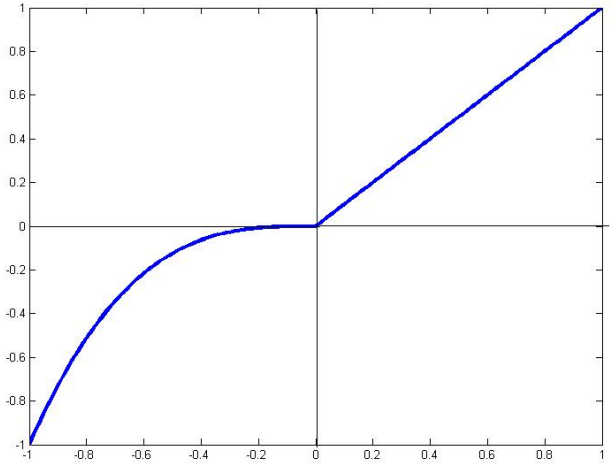
Limiting slope:

$$\boxed{\overline{|\nabla f|}(\bar{x}) := \liminf_{x \xrightarrow[f]{} \bar{x}} |\nabla f|(x)}$$

Critical points:

$$\boxed{\bar{x} \text{ is critical for } f \quad \Longleftrightarrow \quad \overline{|\nabla f|}(\bar{x}) = 0}$$

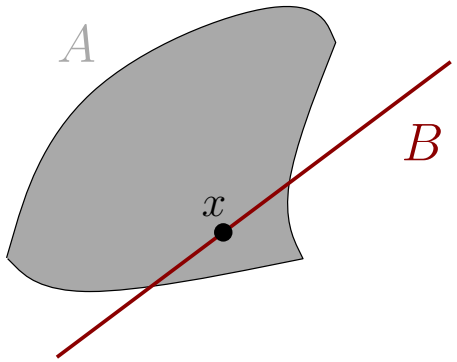
Eg:



Method of alternating projections

Common problem:

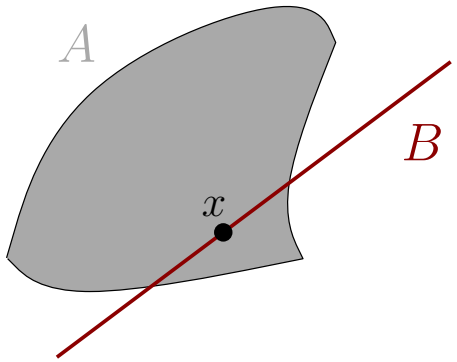
$$\boxed{\text{Given sets } A, B \subset \mathbf{R}^n, \text{ find some point } x \in A \cap B.}$$



Method of alternating projections

Common problem:

$$\boxed{\text{Given sets } A, B \subset \mathbf{R}^n, \text{ find some point } x \in A \cap B.}$$



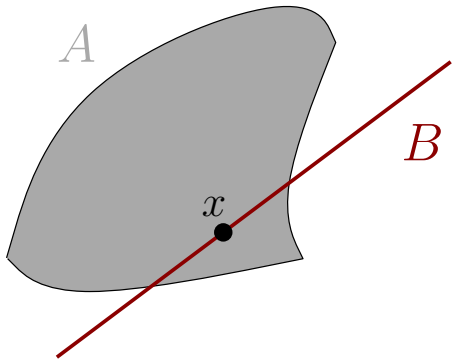
Distance and projection:

$$d_B(x) = \min_{y \in B} |x - y| \quad \text{and} \quad P_B(x) = \{\text{nearest points of } B \text{ to } x\}.$$

Method of alternating projections

Common problem:

$$\boxed{\text{Given sets } A, B \subset \mathbf{R}^n, \text{ find some point } x \in A \cap B.}$$



Distance and projection:

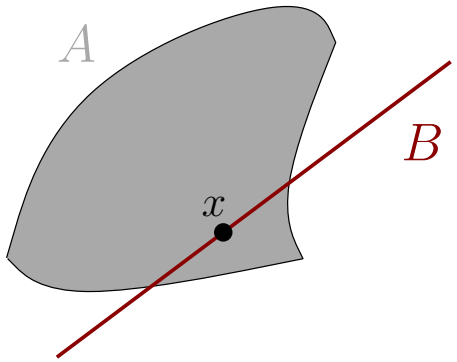
$$d_B(x) = \min_{y \in B} |x - y| \quad \text{and} \quad P_B(x) = \{\text{nearest points of } B \text{ to } x\}.$$

Finding points in P_A and P_B is often easy!

Method of alternating projections

Common problem:

$$\boxed{\text{Given sets } A, B \subset \mathbf{R}^n, \text{ find some point } x \in A \cap B.}$$



Distance and projection:

$$d_B(x) = \min_{y \in B} |x - y| \quad \text{and} \quad P_B(x) = \{\text{nearest points of } B \text{ to } x\}.$$

Finding points in P_A and P_B is often easy!

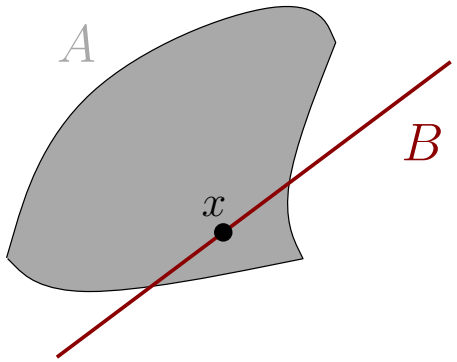
Eg 1 (simple example): Linear programming:

$$\{x : x \geq 0\} \bigcap \{x : Ax = b\}.$$

Method of alternating projections

Common problem:

$$\boxed{\text{Given sets } A, B \subset \mathbf{R}^n, \text{ find some point } x \in A \cap B.}$$



Distance and projection:

$$d_B(x) = \min_{y \in B} |x - y| \quad \text{and} \quad P_B(x) = \{\text{nearest points of } B \text{ to } x\}.$$

Finding points in P_A and P_B is often easy!

Eg 1 (simple example): Linear programming:

$$\{x : x \geq 0\} \bigcap \{x : Ax = b\}.$$

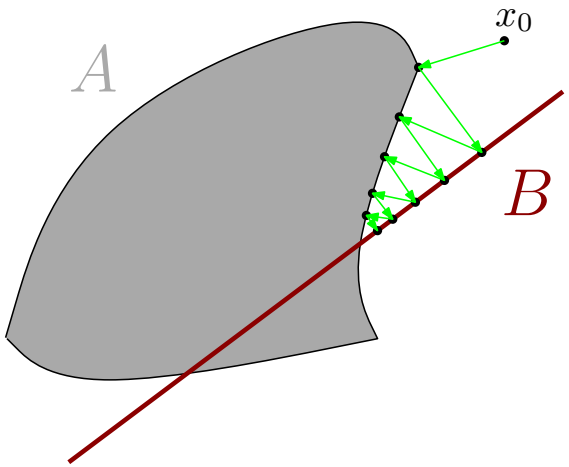
Eg 2 (more interesting): Low-order control:

$$\{X \succeq 0 : \operatorname{rank} X \leq r\} \bigcap \{X : \mathcal{A}(X) = b\}.$$

Method of alternating projections

Method of alternating projections (von Neumann '33):

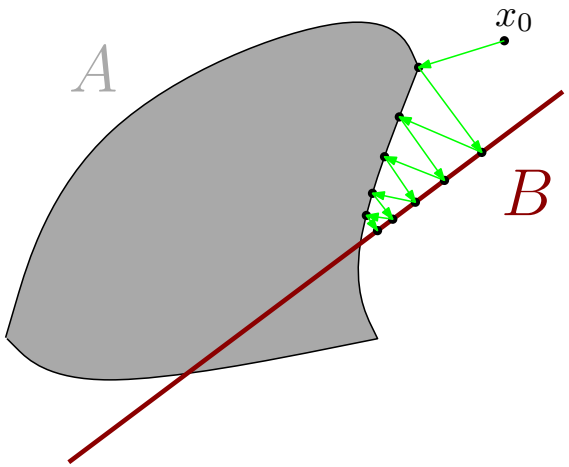
$$\boxed{\begin{array}{lcl} x_{k+1} & \in & P_B(x_k) \\ x_{k+2} & \in & P_A(x_{k+1}) \end{array}}$$



Method of alternating projections

Method of alternating projections (von Neumann '33):

$$\boxed{\begin{array}{lcl} x_{k+1} & \in & P_B(x_k) \\ x_{k+2} & \in & P_A(x_{k+1}) \end{array}}$$

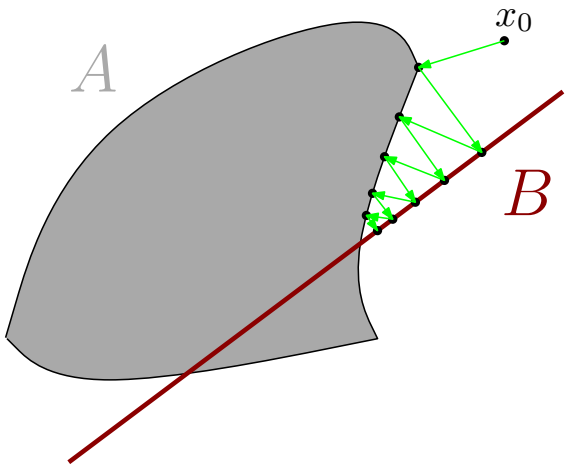


The "angle" between A and B drives the convergence!

Method of alternating projections

Method of alternating projections (von Neumann '33):

$$\boxed{\begin{array}{lcl} x_{k+1} & \in & P_B(x_k) \\ x_{k+2} & \in & P_A(x_{k+1}) \end{array}}$$



The "angle" between A and B drives the convergence!

Quantifying the angle:

$$\psi(x,y) := \begin{cases} |x-y| & \text{if } x \in A, y \in B \\ +\infty & \text{otherwise} \end{cases}$$

Comparison of $|\nabla\psi_y|(x)$ and $|\nabla\psi_x|(y)$ quantifies the angle!

Outline

- **When is slope an adequate tool?**
  - Semi-algebraic case
  - Slope & error bounds

- **Applications:**
  - Alternating projections & transversality
  - Steepest descent curves
  - Active sets & generic properties.

Subdifferentials (an interlude)

Subdifferential:

$$v \in \partial f(\bar{x}) \quad \Longleftrightarrow \quad \bar{x} \text{ is critical for } f - \langle v, \cdot \rangle$$

Subdifferentials (an interlude)

Subdifferential:

$$v \in \partial f(\bar{x}) \quad \Longleftrightarrow \quad \bar{x} \text{ is critical for } f - \langle v, \cdot \rangle$$

Surprising consequence (Ioffe '00):

$$\boxed{\overline{|\nabla f|}(x) = \operatorname{dist}(0, \partial f(x)).}$$

Subdifferentials (an interlude)

Subdifferential:

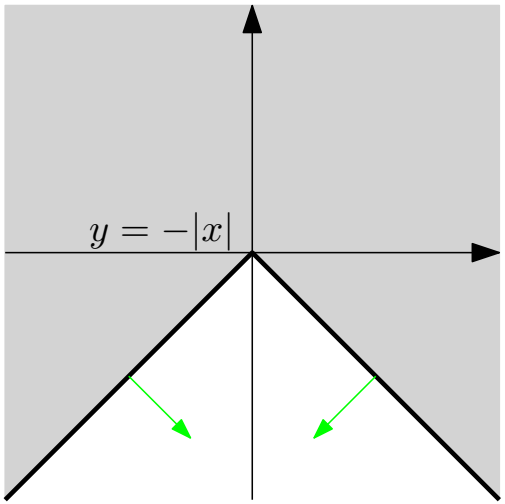
$$v \in \partial f(\bar{x}) \quad \Longleftrightarrow \quad \bar{x} \text{ is critical for } f - \langle v, \cdot \rangle$$

Surprising consequence (Ioffe '00):

$$\boxed{\overline{|\nabla f|}(x) = \operatorname{dist}(0, \partial f(x)).}$$

Example:

$$\partial(-|\cdot|)(x) = \begin{cases} 1 & \text{if } x < 0 \\ \{-1, 1\} & \text{if } x = 0 \\ -1 & \text{if } x > 0 \end{cases}$$



Subdifferentials (an interlude)

Subdifferential:

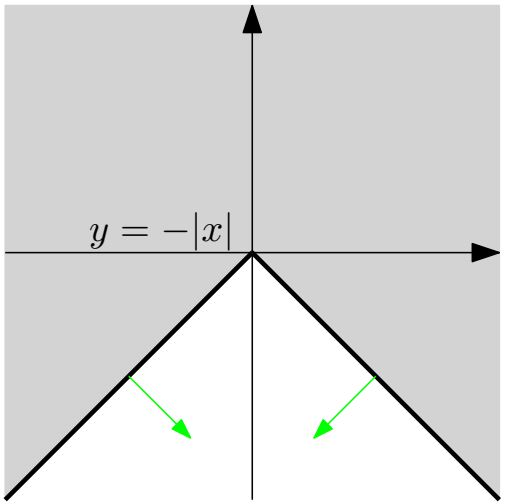
$$v \in \partial f(\bar{x}) \quad \Longleftrightarrow \quad \bar{x} \text{ is critical for } f - \langle v, \cdot \rangle$$

Surprising consequence (Ioffe '00):

$$\boxed{\overline{|\nabla f|}(x) = \operatorname{dist}(0, \partial f(x)).}$$

Example:

$$\partial(-|\cdot|)(x) = \begin{cases} 1 & \text{if } x < 0 \\ \{-1, 1\} & \text{if } x = 0 \\ -1 & \text{if } x > 0 \end{cases}$$



Subdifferential graph:

$$\boxed{\operatorname{gph} \partial f := \{(x, v) \in \mathbf{R}^n \times \mathbf{R}^n : v \in \partial f(x)\}}$$

Subdifferentials

Are these notions adequate?

Subdifferentials

Are these notions adequate?

Pathology: $\operatorname{gph} \partial f$ can be very large!

Subdifferentials

Are these notions adequate?

Pathology: $\operatorname{gph} \partial f$ can be very large!

What do we expect the size of $\operatorname{gph} \partial f$ to be?

- If $f \colon \mathbf{R}^n \to \mathbf{R}$ is smooth, then $\operatorname{gph} \partial f$ is n-dimensional smooth manifold.
- If $f \colon \mathbf{R}^n \to \mathbf{R}$ is convex, then $\operatorname{gph} \partial f$ is n-dimensional Lipschitz manifold (Minty '62).

Subdifferentials

Are these notions adequate?

Pathology: $\operatorname{gph} \partial f$ can be very large!

What do we expect the size of $\operatorname{gph} \partial f$ to be?

- If $f\colon \mathbf{R}^n \to \mathbf{R}$ is smooth, then $\operatorname{gph} \partial f$ is n-dimensional smooth manifold.
- If $f\colon \mathbf{R}^n \to \mathbf{R}$ is convex, then $\operatorname{gph} \partial f$ is n-dimensional Lipschitz manifold (Minty '62).

Multiple authors (Rockafellar, Borwein, Wang,...):
There are functions $f\colon \mathbf{R}^n \to \mathbf{R}$ with "$2n$-dimensional" $\operatorname{gph} \partial f$.

Semi-algebraic geometry

$Q \subset \mathbf{R}^n$ is semi-algebraic if is a finite union of

solution sets to finitely many polynomial inequalities.

Semi-algebraic geometry

$Q \subset \mathbf{R}^n$ is semi-algebraic if is a finite union of

solution sets to finitely many polynomial inequalities.

Eg: semi-definite representable sets (Nesterov-Nemirovskii)

Semi-algebraic geometry

$Q \subset \mathbf{R}^n$ is semi-algebraic if is a finite union of

solution sets to finitely many polynomial inequalities.

Eg: semi-definite representable sets (Nesterov-Nemirovskii)

Semi-algebraicity is robust (Tarski-Seidenberg theorem).
Eg:
f semi-algebraic $\quad\Longrightarrow\quad$ $\operatorname{gph} \partial f$ and $|\nabla f|$ are semi-algebraic.

Semi-algebraic geometry

$Q \subset \mathbf{R}^n$ is semi-algebraic if is a finite union of

solution sets to finitely many polynomial inequalities.

Eg: semi-definite representable sets (Nesterov-Nemirovskii)

Semi-algebraicity is robust (Tarski-Seidenberg theorem).
Eg:
f semi-algebraic $\quad\Longrightarrow\quad$ $\operatorname{gph} \partial f$ and $|\nabla f|$ are semi-algebraic.

Semi-algebraic Q "stratify" into finitely many manifolds $\{\mathcal{M}_i\}$.
Dimension:

$$\dim Q := \max_{i=1,\ldots,k} \{\dim \mathcal{M}_i\}.$$

Semi-algebraic geometry

Theorem (D-Ioffe-Lewis)

For semi-algebraic $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, *we have*

$$\dim \operatorname{gph} \partial f = n,$$

even locally around any point in $\operatorname{gph} \partial f$.

Semi-algebraic geometry

Theorem (D-Ioffe-Lewis)

For semi-algebraic $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, *we have*

$$\dim \operatorname{gph} \partial f = n,$$

even locally around any point in $\operatorname{gph} \partial f$.

Conclusion: Criticality is meaningful for concrete variational problems!

Semi-algebraic geometry

Theorem (D-Ioffe-Lewis)

For semi-algebraic $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, *we have*

$$\dim \operatorname{gph} \partial f = n,$$

even locally around any point in $\operatorname{gph} \partial f$.

Conclusion: Criticality is meaningful for concrete variational problems!

Semi-algebraic f have only finitely many critical values (cf. Sard's Theorem)

Semi-algebraic geometry

Theorem (D-Ioffe-Lewis)

For semi-algebraic $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, *we have*

$$\dim \operatorname{gph} \partial f = n,$$

even locally around any point in $\operatorname{gph} \partial f$.

Conclusion: Criticality is meaningful for concrete variational problems!

Semi-algebraic f have only finitely many critical values (cf. Sard's Theorem) $\Rightarrow$ intervals (a, b) of non-critical values.

Semi-algebraic geometry

Theorem (D-Ioffe-Lewis)

For semi-algebraic $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, *we have*

$$\dim \operatorname{gph} \partial f = n,$$

even locally around any point in $\operatorname{gph} \partial f$.

Conclusion: Criticality is meaningful for concrete variational problems!

Semi-algebraic f have only finitely many critical values (cf. Sard's Theorem) $\Rightarrow$ intervals (a, b) of non-critical values.

What can we learn from non-criticality?

Slope & error bounds

Common problem: Estimate

$$\operatorname{dist}(x, [f \leq r]) \quad \text{(difficult)}.$$

Slope & error bounds

Common problem: Estimate

$$\operatorname{dist}(x, [f \leq r]) \quad \text{(difficult)}.$$

"The residual":

$$f(x) - r \qquad \text{(easy)}.$$

Slope & error bounds

Common problem: Estimate

$$\operatorname{dist}(x, [f \le r]) \quad \text{(difficult)}.$$

"The residual":

$$f(x) - r \qquad \text{(easy)}.$$

Desirable quality: Exists κ with

$$\operatorname{dist}(x, [f \le r]) \le \kappa(f(x) - r).$$

Slope & error bounds

Common problem: Estimate

$$\operatorname{dist}(x, [f \leq r]) \quad \text{(difficult)}.$$

"The residual":

$$f(x) - r \qquad \text{(easy)}.$$

Desirable quality: Exists κ with

$$\operatorname{dist}(x, [f \leq r]) \leq \kappa(f(x) - r).$$

Restrict $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ to a "slice" $f^{-1}(a, b)$.

Slope & error bounds

Common problem: Estimate

$$\operatorname{dist}(x, [f \le r]) \quad \text{(difficult)}.$$

"The residual":

$$f(x) - r \qquad \text{(easy)}.$$

Desirable quality: Exists κ with

$$\operatorname{dist}(x, [f \le r]) \le \kappa(f(x) - r).$$

Restrict $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ to a "slice" $f^{-1}(a, b)$.

Lemma (Error bound)

The following are equivalent.

Non-criticality:

$$|\nabla f| \ge \frac{1}{\kappa}.$$

Error-bound:

$$\operatorname{dist}(x, [f \le r]) \le \kappa(f(x) - r), \quad \text{when } r \in (a, f(x)).$$

Slope & error bounds

Common problem: Estimate

$$\operatorname{dist}(x, [f \le r]) \quad \text{(difficult)}.$$

"The residual":

$$f(x) - r \qquad \text{(easy)}.$$

Desirable quality: Exists κ with

$$\operatorname{dist}(x, [f \le r]) \le \kappa(f(x) - r).$$

Restrict $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ to a "slice" $f^{-1}(a, b)$.

Lemma (Error bound)

The following are equivalent.

Non-criticality:

$$|\nabla f| \ge \frac{1}{\kappa}.$$

Error-bound:

$$\operatorname{dist}(x, [f \le r]) \le \kappa(f(x) - r), \quad \text{when } r \in (a, f(x)).$$

- Observed by Azé-Corvellec '04, Ioffe '00.

Semi-algebraicity & error bounds

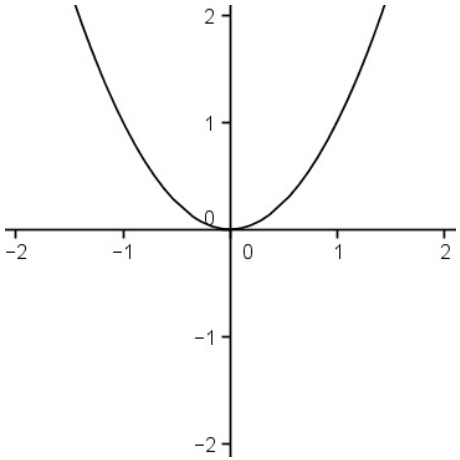


Figure: $f(x) = x^2$

Semi-algebraicity & error bounds

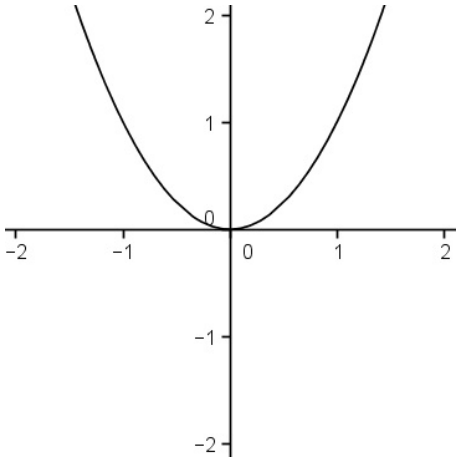


Figure: $f(x) = x^2$

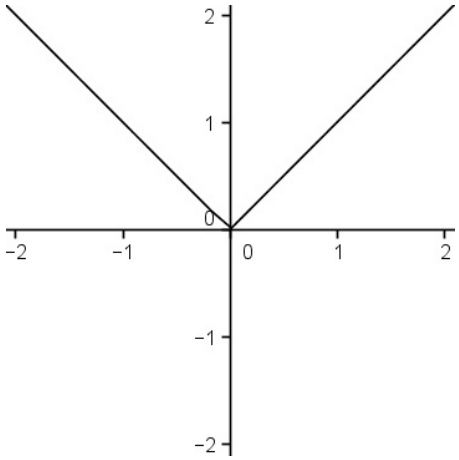


Figure: $\sqrt{f(x)} = |x|$

Semi-algebraicity & error bounds

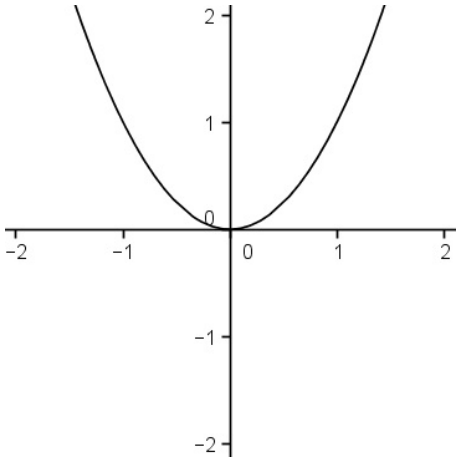


Figure: $f(x) = x^2$

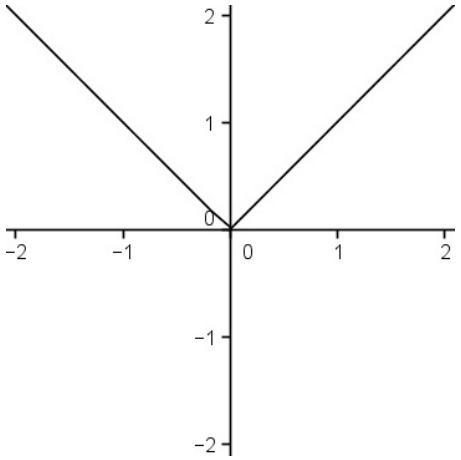


Figure: $\sqrt{f(x)} = |x|$

Desingularization: (Bolte-Daniilidis-Lewis '07)
For semi-algebraic f, there exists "nice" ϕ with

$$|\nabla(\phi \circ f)|(x) \geq 1 \quad \text{for } x \notin \operatorname{crit} f.$$

Semi-algebraicity & error bounds

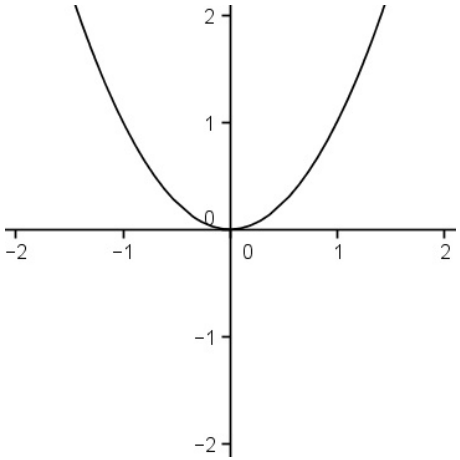


Figure: $f(x) = x^2$

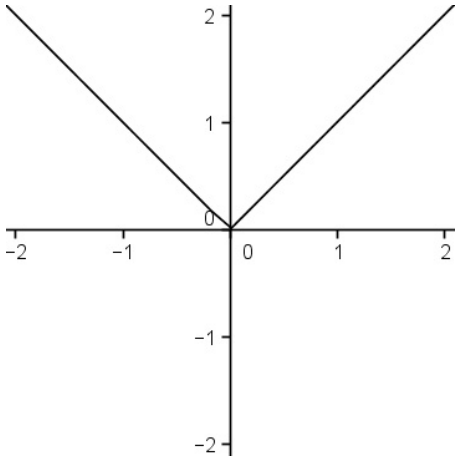


Figure: $\sqrt{f(x)} = |x|$

Desingularization: (Bolte-Daniilidis-Lewis '07)
For semi-algebraic f, there exists "nice" ϕ with

$$|\nabla(\phi \circ f)|(x) \geq 1 \quad \text{for } x \notin \operatorname{crit} f.$$

Error bounds always applicable for semi-algebraic functions!

Slope & Error bounds

Second-order behaviour (error bounds of the slope):

$$\boxed{x \mapsto \overline{|\nabla f|}(x)}$$

Slope & Error bounds

Second-order behaviour (error bounds of the slope):

$$\boxed{x \mapsto \overline{|\nabla f|}(x)}$$

Theorem (D-Ioffe)

For semi-algebraic f and a strict minimizer $\bar{x}$, the following are equivalent:

Quadratic growth:

$$f(x) \geq f(\bar{x}) + \alpha|x - \bar{x}|^2 \qquad \textit{for } x \textit{ near } \bar{x}.$$

Error-bound:

$$|x - \bar{x}| \leq \kappa \cdot d(0, \partial f(x)) \qquad \textit{for } x \textial{ near } \bar{x}.$$

Slope & Error bounds

Second-order behaviour (error bounds of the slope):

$$\boxed{x \mapsto \overline{|\nabla f|}(x)}$$

Theorem (D-Ioffe)

For semi-algebraic f and a strict minimizer $\bar{x}$, the following are equivalent:

Quadratic growth:

$$f(x) \geq f(\bar{x}) + \alpha|x - \bar{x}|^2 \qquad \textit{for } x \textit{ near } \bar{x}.$$

Error-bound:

$$|x - \bar{x}| \leq \kappa \cdot d(0, \partial f(x)) \qquad \textit{for } x \textit{ near } \bar{x}.$$

- Not true for general functions; **eg.** $f(x) = 2x^2 + \frac{1}{2}x^2 \sin(\frac{1}{x})$

Slope & Error bounds

Second-order behaviour (error bounds of the slope):

$$\boxed{x \mapsto \overline{|\nabla f|}(x)}$$

Theorem (D-Ioffe)

For semi-algebraic f and a strict minimizer $\bar{x}$, the following are equivalent:

Quadratic growth:

$$f(x) \geq f(\bar{x}) + \alpha|x - \bar{x}|^2 \qquad \textit{for } x \textit{ near } \bar{x}.$$

Error-bound:

$$|x - \bar{x}| \leq \kappa \cdot d(0, \partial f(x)) \qquad \textit{for } x \textit{ near } \bar{x}.$$

- Not true for general functions; **eg.** $f(x) = 2x^2 + \frac{1}{2}x^2 \sin(\frac{1}{x})$

Second-order growth/Regularity: Poliquin-Rockafellar '98, D-Nghia-Mordukhovich '13, Artacho-Geoffroy '08 '13

Alternating projections & transversality

Convergence of alternating projections

Transversality: $N_A(\bar{x}) \cap -N_B(\bar{x}) = \{0\}$ $(N_A = \partial\delta_A)$

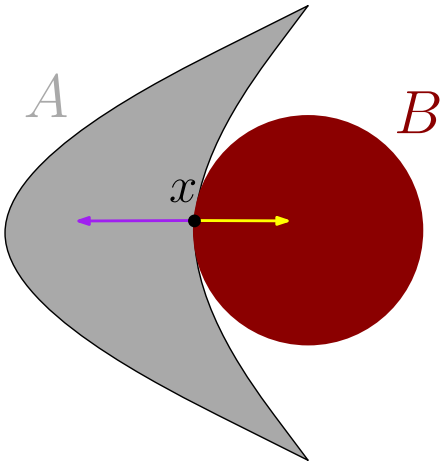


Figure: Not transverse

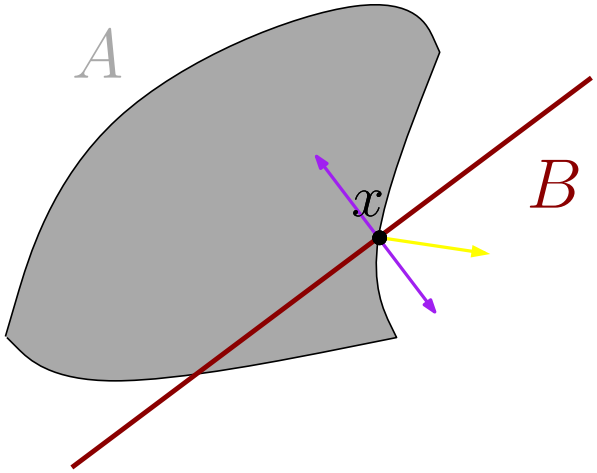


Figure: Transverse

Convergence of alternating projections

Transversality: $N_A(\bar{x}) \cap -N_B(\bar{x}) = \{0\}$ $(N_A = \partial\delta_A)$

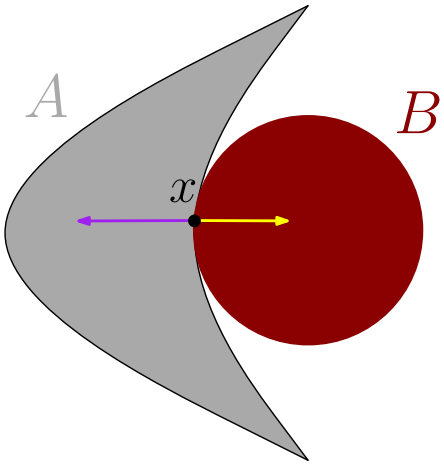


Figure: Not transverse

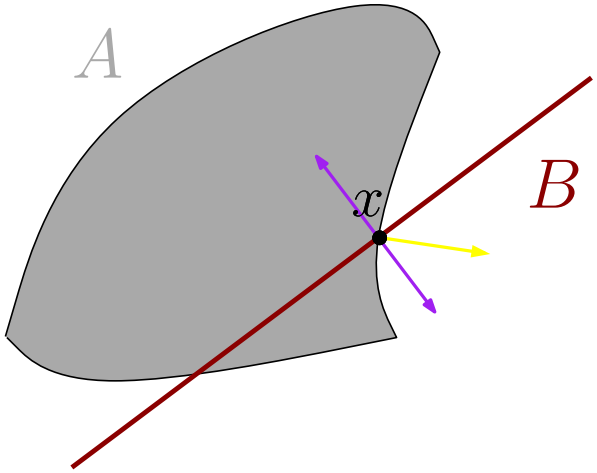


Figure: Transverse

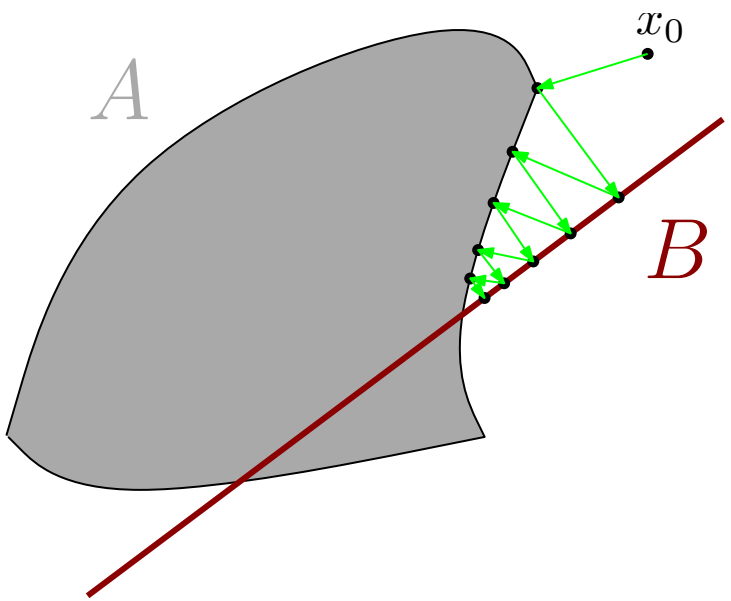
Local convergence (D-Ioffe-Lewis '13):

A and B transverse at $\bar{x}$ $\implies$ local **R**-linear convergence.

Transversality & error bounds

Coupling function:

$$\boxed{\psi(x, y) = \delta_A(x) + |x - y| + \delta_B(y).}$$



Transversality & error bounds

Coupling function:

$$\boxed{\psi(x,y) = \delta_A(x) + |x-y| + \delta_B(y).}$$

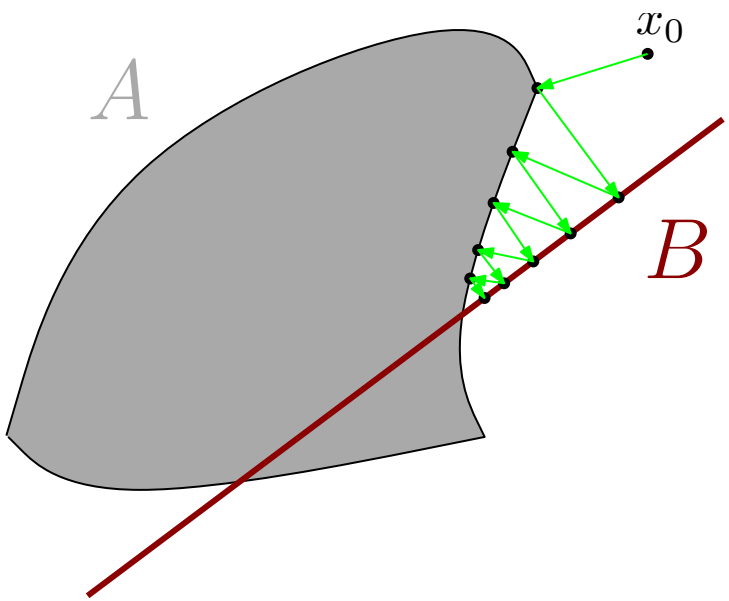
Error bound:

$$N_B(\bar{x}) \cap -N_A(\bar{x}) = \{0\}$$

$$\Downarrow$$

$$\boxed{\max\{|\nabla\psi_x|(y),\, |\nabla\psi_y|(x)\} \geq \kappa}$$

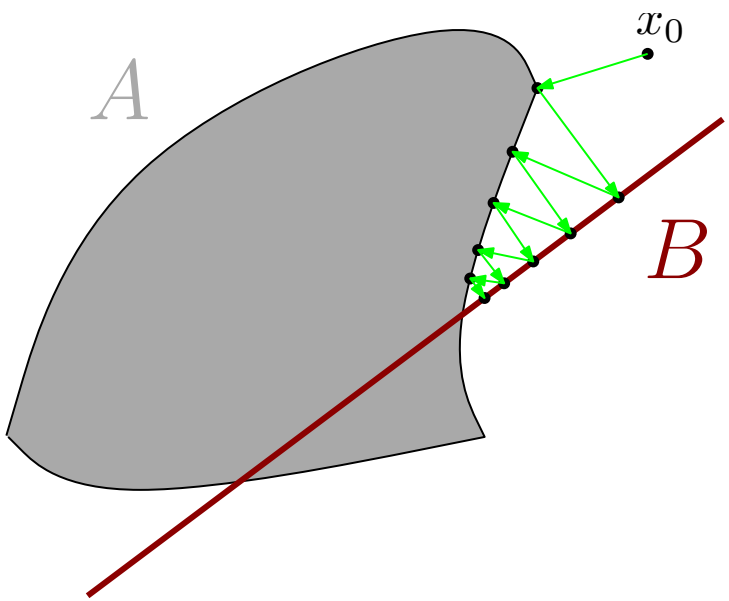
for $x \in A$ and $y \in B$, not in $A \cap B$.



Transversality & error bounds

Coupling function:

$$\boxed{\psi(x,y) = \delta_A(x) + |x-y| + \delta_B(y).}$$



Error bound:

$$N_B(\bar{x}) \cap -N_A(\bar{x}) = \{0\}$$

$$\Downarrow$$

$$\boxed{\max\{|\nabla\psi_x|(y),\, |\nabla\psi_y|(x)\} \geq \kappa}$$

for $x \in A$ and $y \in B$, not in $A \cap B$.

$$\Downarrow$$

Local linear convergence

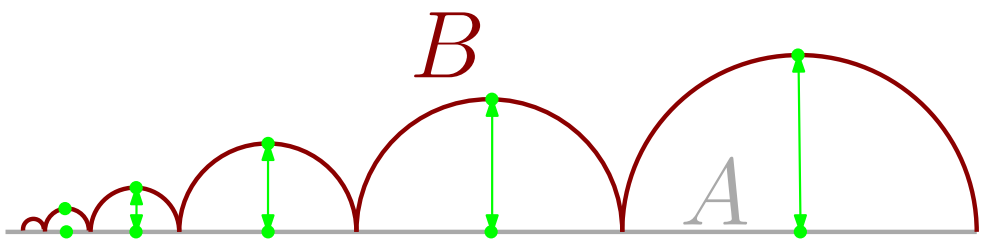
Convergence

- Transversality is necessary but not verifiable.

Convergence

- Transversality is necessary but not verifiable.

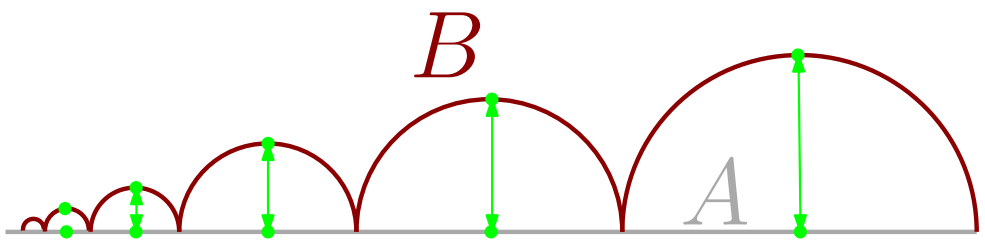
Eg:



Convergence

- Transversality is necessary but not verifiable.

Eg:

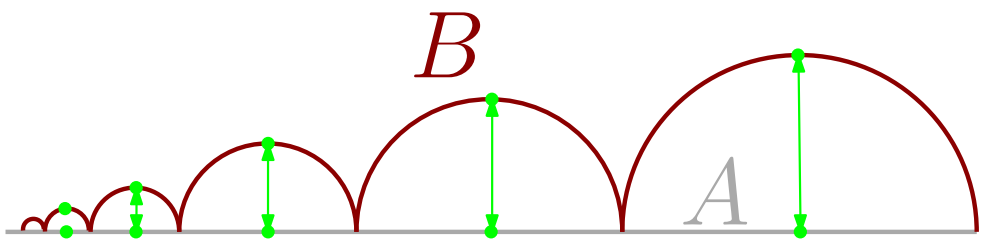


What about sublinear convergence?

Convergence

- Transversality is necessary but not verifiable.

Eg:



What about sublinear convergence?

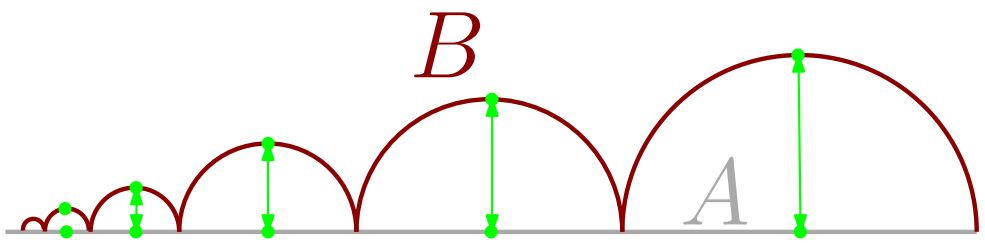
Theorem (D-Ioffe-Lewis)

A and B are semi-algebraic, $A \cap B$ is compact, x_0 near $A \cap B$ $\Longrightarrow$ alternating projections converge.

Convergence

- Transversality is necessary but not verifiable.

Eg:



What about sublinear convergence?

Theorem (D-Ioffe-Lewis)

A and B are semi-algebraic, $A \cap B$ is compact, x_0 near $A \cap B$ $\Longrightarrow$ alternating projections converge.

Generic transversality (D-Ioffe-Lewis):

If A and B are semi-algebraic, then

$A + a$ and $B + b$ are transverse for a.e. (a, b)

Open question

General paradigm:

$$\text{no convexity} \quad \Longrightarrow \quad \text{no global convergence.}$$

Open question

General paradigm:

$$\text{no convexity} \quad \Longrightarrow \quad \text{no global convergence.}$$

Variants of alternating projections work globally!

Open question

General paradigm:

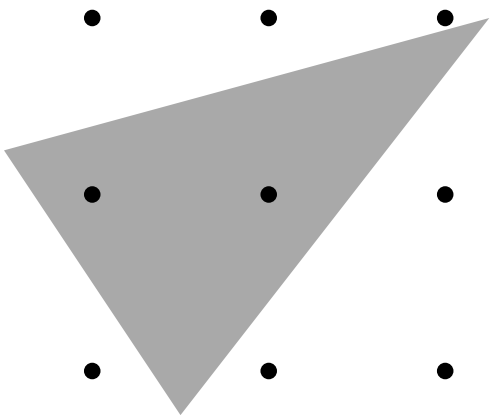
$$\text{no convexity} \quad \Longrightarrow \quad \text{no global convergence.}$$

Variants of alternating projections work globally!

- Integer programming:

$$\mathbb{Z}^n \cap \{x : Ax \leq b\}$$

(eg: sudoku, 3-SAT, 4 queens problem, etc ...)



Ongoing work Artacho, Borwein, Tam.

Steepest descent curves

Steepest descent curves

Bounded speed: A curve $x\colon [0, T] \to \mathcal{X}$ is 1-Lipschitz if

$$\operatorname{dist}(x(t), x(s)) \leq |t - s|.$$

Steepest descent curves

Bounded speed: A curve $x\colon [0, T] \to \mathcal{X}$ is 1-Lipschitz if

$$\operatorname{dist}(x(t), x(s)) \leq |t - s|.$$

$\boxed{\text{What are steepest descent curves?}}$

Steepest descent curves

Bounded speed: A curve $x\colon [0, T] \to \mathcal{X}$ is 1-Lipschitz if

$$\operatorname{dist}(x(t), x(s)) \leq |t - s|.$$

$$\boxed{\text{What are steepest descent curves?}}$$

Motivation: 1-Lipschitz curves $x\colon [0, T] \to \mathcal{X}$ satisfy

$$|\nabla(f \circ x)| \leq |\nabla f|(x).$$

Steepest descent curves

Bounded speed: A curve $x\colon [0, T] \to \mathcal{X}$ is 1-Lipschitz if

$$\operatorname{dist}(x(t), x(s)) \leq |t - s|.$$

$$\boxed{\text{What are steepest descent curves?}}$$

Motivation: 1-Lipschitz curves $x\colon [0, T] \to \mathcal{X}$ satisfy

$$|\nabla(f \circ x)| \leq |\nabla f|(x).$$

Definition (Near-steepest descent curves)

Curve $x\colon [0, T] \to \mathcal{X}$ is a near-steepest descent curve if

- x is 1-Lipschitz,
- $f \circ x$ is decreasing,
- $|\nabla(f \circ x)| \geq \overline{|\nabla f|}(x), \quad$ a.e. on $[0, T]$.

Example

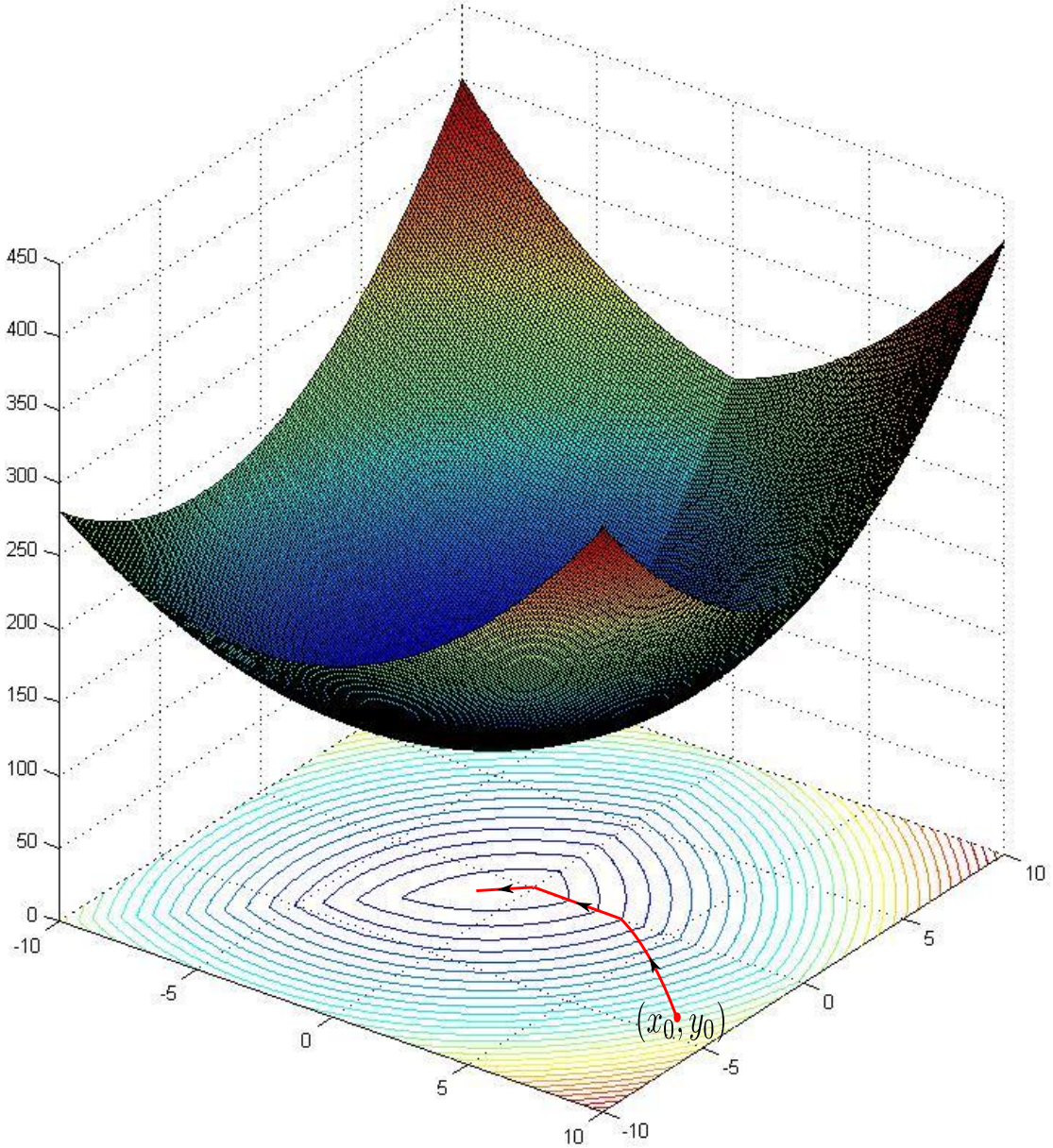


Figure: $f(x, y) = \max\{x + y, |x - y|\} + x(x + 1) + y(y + 1) + 100$

Subgradient dynamical systems

Theorem (D-Ioffe-Lewis)

For reasonable $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *and a curve* $x\colon [0, T] \to \mathbf{R}^n$ *the following are equivalent.*

1. x *is a near-steepest descent curve,*
2. $f \circ x$ *is decreasing and (after reparametrizing)*

$$\dot{x} \in -\partial f(x), \qquad a.e. \text{ on } [0, T].$$

Subgradient dynamical systems

Theorem (D-Ioffe-Lewis)

For reasonable $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *and a curve* $x\colon [0, T] \to \mathbf{R}^n$ *the following are equivalent.*

1. x *is a near-steepest descent curve,*
2. $f \circ x$ *is decreasing and (after reparametrizing)*

$$\dot{x} \in -\partial f(x), \qquad \textit{a.e. on } [0, T].$$

Remark: When 2. holds,

$$\dot{x} \text{ is the shortest element of } -\partial f(x), \qquad \text{a.e. on } [0, T].$$

Subgradient dynamical systems

Theorem (D-Ioffe-Lewis)

For reasonable $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ *and a curve* $x\colon [0, T] \to \mathbf{R}^n$ *the following are equivalent.*

1. x *is a near-steepest descent curve,*
2. $f \circ x$ *is decreasing and (after reparametrizing)*

$$\dot{x} \in -\partial f(x), \qquad \textit{a.e. on } [0, T].$$

Remark: When 2. holds,

$$\dot{x} \text{ is the shortest element of } -\partial f(x), \qquad \text{a.e. on } [0, T].$$

Reasonable conditions: f is smooth, convex, or semi-algebraic.

Existence

Theorem (Ambrosio et al. '05, De Giorgi '93)

For reasonable $f\colon \mathcal{X} \to \overline{\mathbf{R}}$, there exist near-steepest descent curves starting from any point.

Existence

Theorem (Ambrosio et al. '05, De Giorgi '93)

For reasonable $f\colon \mathcal{X} \to \overline{\mathbf{R}}$, there exist near-steepest descent curves starting from any point.

Proof ingredients:

- Moreau-Yosida approximation:

$$x_{k+1} = \operatorname*{argmin}_{x \in \mathcal{X}} \{f(x) + \frac{1}{2\tau} d^2(x, x_k)\}.$$

- Extraneous topologies $\Rightarrow$ existence of minimizers and convergence.

Existence

Theorem (Ambrosio et al. '05, De Giorgi '93)

For reasonable $f\colon \mathcal{X} \to \overline{\mathbf{R}}$, there exist near-steepest descent curves starting from any point.

Proof ingredients:

- Moreau-Yosida approximation:

$$x_{k+1} = \operatorname*{argmin}_{x \in \mathcal{X}} \{f(x) + \frac{1}{2\tau} d^2(x, x_k)\}.$$

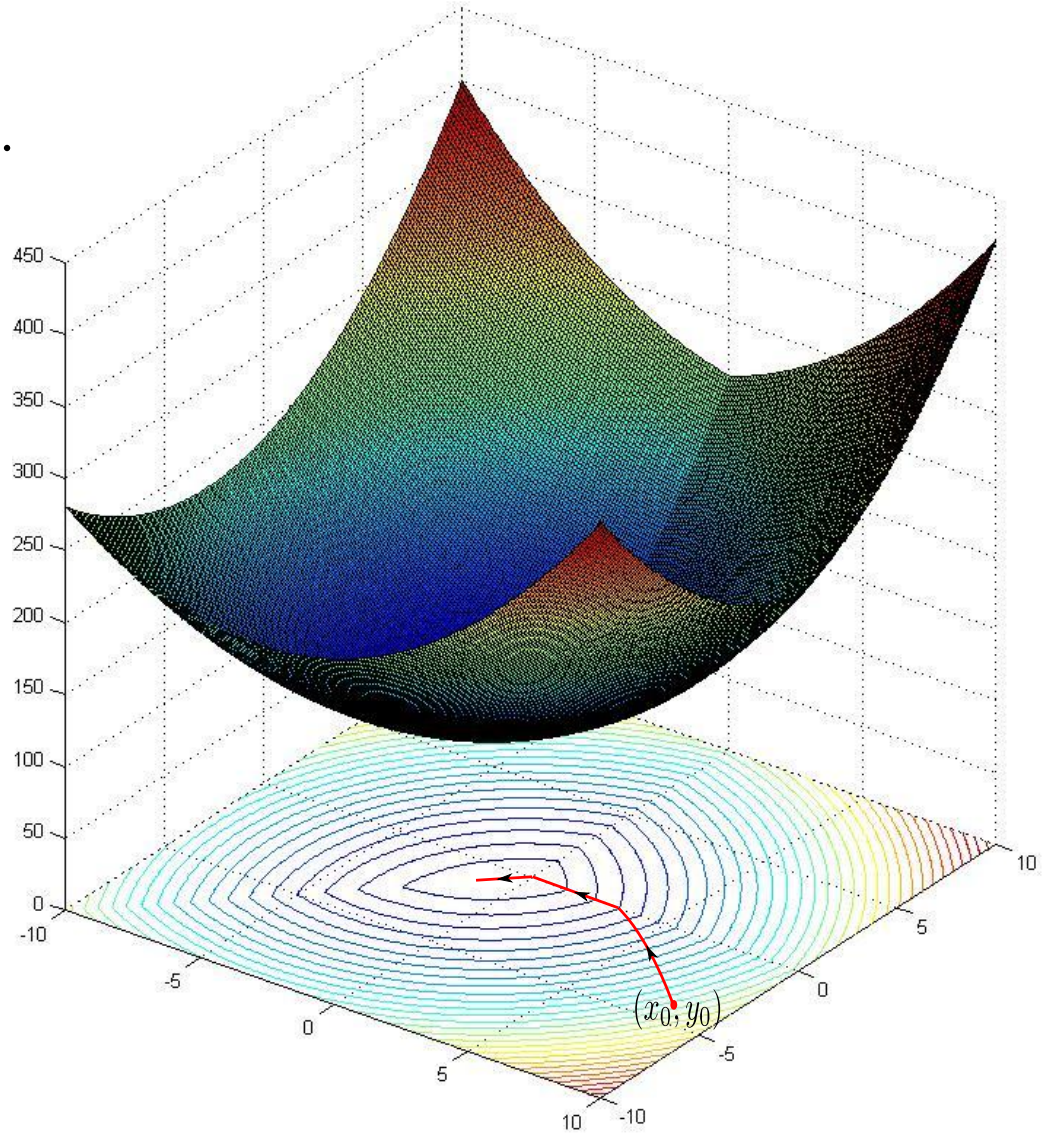
- Extraneous topologies $\Rightarrow$ existence of minimizers and convergence.

Proof is opaque and uses heavy machinery!

New proof idea: Error bound lemma

Equipartition: For $\eta > 0$,

$$f(x_0) - \eta = \tau_0 < \ldots < \tau_k = f(x_0).$$

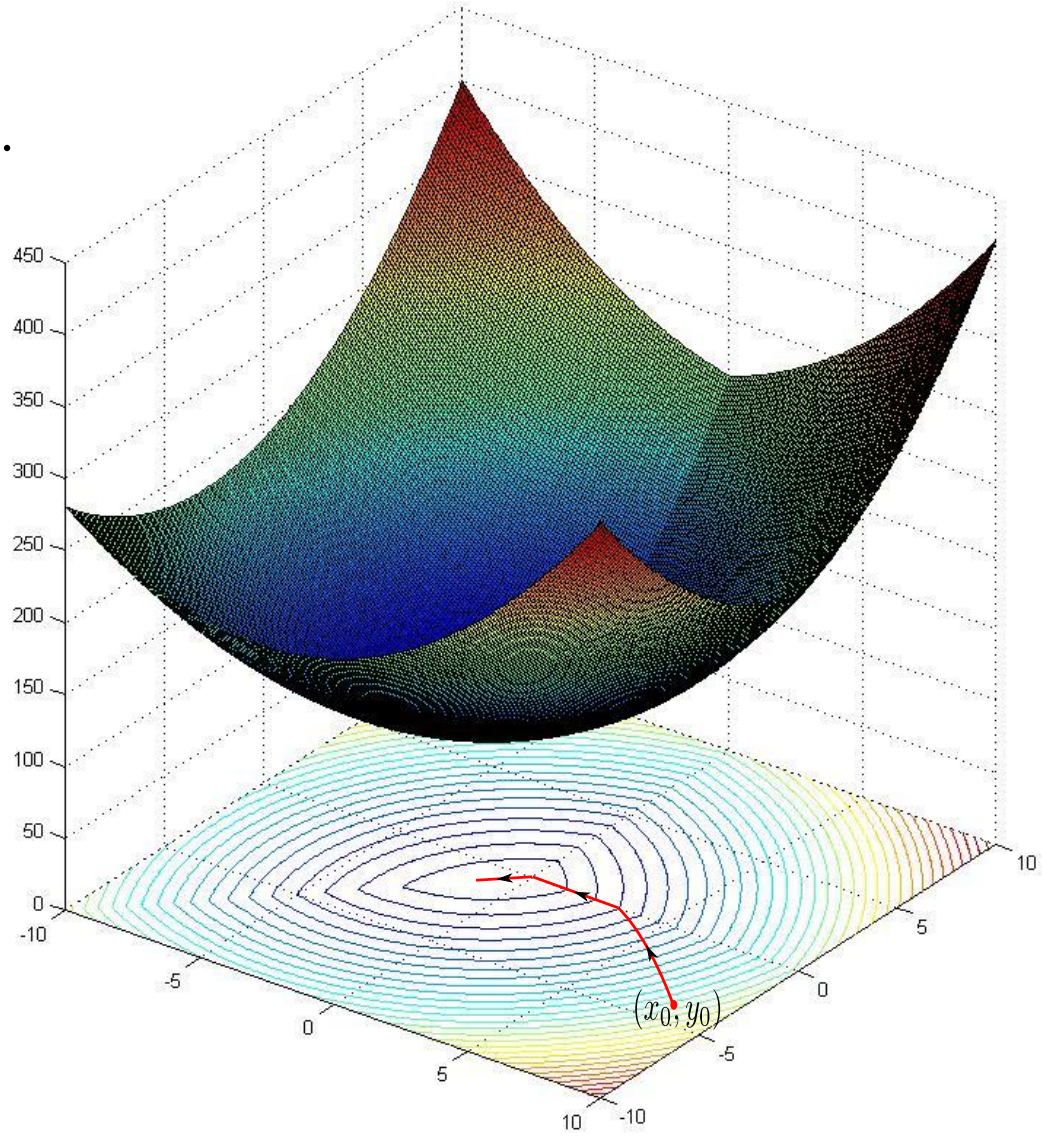


New proof idea: Error bound lemma

Equipartition: For $\eta > 0$,

$$f(x_0) - \eta = \tau_0 < \ldots < \tau_k = f(x_0).$$

Initialize: $j = 0$;
while $i \leq k$ **do**
  $x_{j+1} \leftarrow P_{[f \leq \tau_{j+1}]}(x_j)$;
end

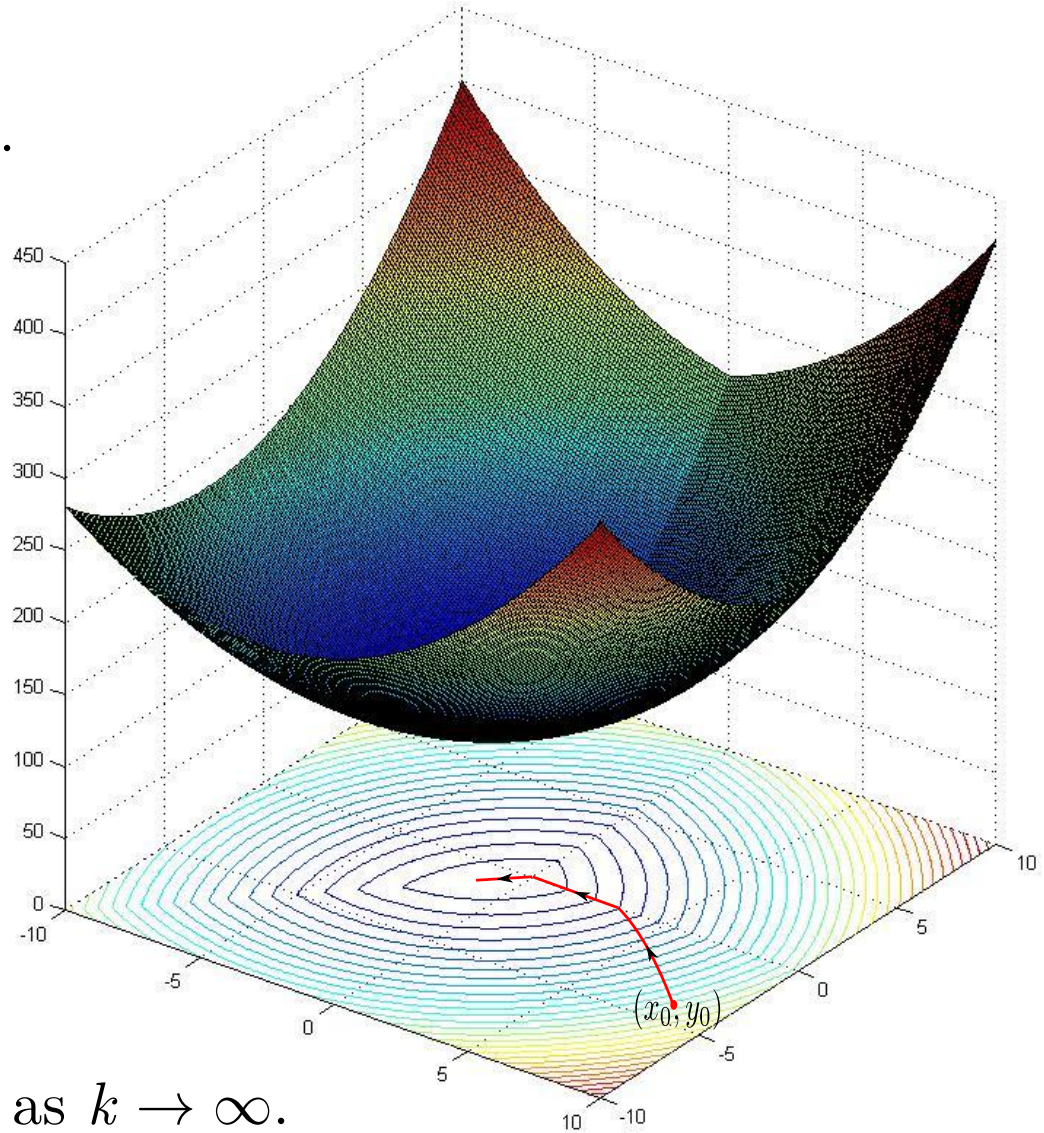


New proof idea: Error bound lemma

Equipartition: For $\eta > 0$,

$$f(x_0) - \eta = \tau_0 < \ldots < \tau_k = f(x_0).$$

Initialize: $j = 0$;
while $i \leq k$ **do**
  $x_{j+1} \leftarrow P_{[f \leq \tau_{j+1}]}(x_j)$;
end



Consider resulting trajectories as $k \to \infty$.

Semi-algebraic descent

Theorem (D-Ioffe-Lewis '13)

f semi-algebraic, $\bar{x}$ not a local minimizer $\Longrightarrow$

There exists a nontrivial descent curve starting from $\bar{x}$

Semi-algebraic descent

Theorem (D-Ioffe-Lewis '13)

f semi-algebraic, $\bar{x}$ not a local minimizer $\Longrightarrow$

There exists a nontrivial descent curve starting from $\bar{x}$

- Not true for $\mathbf{C}^\infty$ functions; **eg.** $f(x) = e^{-\frac{1}{|x|}} \sin(\frac{1}{x})$

Semi-algebraic descent

Theorem (D-Ioffe-Lewis '13)

f semi-algebraic, $\bar{x}$ not a local minimizer $\Longrightarrow$

There exists a nontrivial descent curve starting from $\bar{x}$

- Not true for $\mathbf{C}^\infty$ functions; **eg.** $f(x) = e^{-\frac{1}{|x|}} \sin(\frac{1}{x})$

Theorem (Daniilidis-Bolte-D-Ioffe-Lewis, Lojasiewicz)

If f is semi-algebraic, then any bounded near-steepest descent curve (with maximal domain) has bounded length and converges to a critical point.

Semi-algebraic descent

Theorem (D-Ioffe-Lewis '13)

f semi-algebraic, $\bar{x}$ not a local minimizer $\Longrightarrow$

There exists a nontrivial descent curve starting from $\bar{x}$

- Not true for $\mathbf{C}^\infty$ functions; **eg.** $f(x) = e^{-\frac{1}{|x|}} \sin(\frac{1}{x})$

Theorem (Daniilidis-Bolte-D-Ioffe-Lewis, Lojasiewicz)

If f is semi-algebraic, then any bounded near-steepest descent curve (with maximal domain) has bounded length and converges to a critical point.

- Not true for $\mathbf{C}^\infty$ functions (Palis, de Melo)

Semi-algebraic descent

Theorem (D-Ioffe-Lewis '13)

f semi-algebraic, $\bar{x}$ not a local minimizer $\Longrightarrow$

There exists a nontrivial descent curve starting from $\bar{x}$

- Not true for $\mathbf{C}^\infty$ functions; **eg.** $f(x) = e^{-\frac{1}{|x|}} \sin(\frac{1}{x})$

Theorem (Daniilidis-Bolte-D-Ioffe-Lewis, Lojasiewicz)

If f is semi-algebraic, then any bounded near-steepest descent curve (with maximal domain) has bounded length and converges to a critical point.

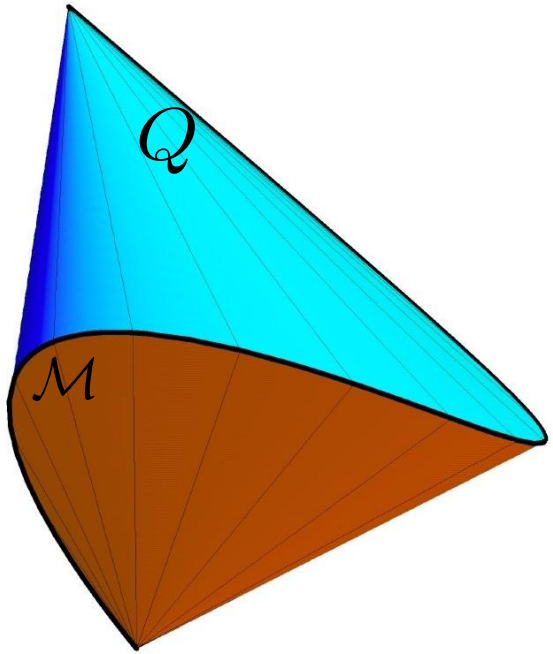
- Not true for $\mathbf{C}^\infty$ functions (Palis, de Melo)

One motivation: Algorithm complexity (Eg: Attouch et al.)

Active sets in optimization.

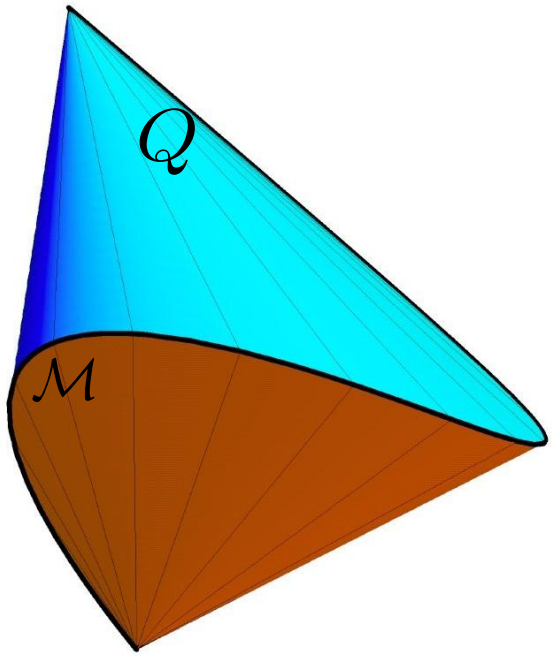
Active sets in optimization

Figure: Q is 4×4 Toeplitz spectrahedron



Active sets in optimization

Figure: Q is 4×4 Toeplitz spectrahedron



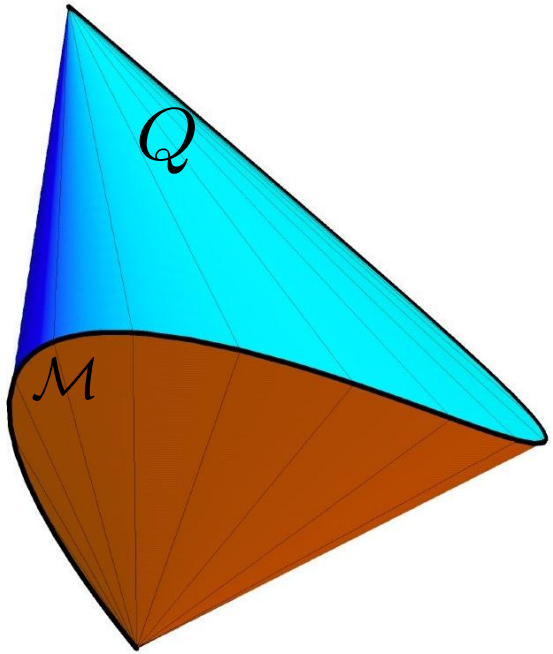
Definition (Partial Smoothness)

A set Q is partly smooth relative to $\mathcal{M} \subset Q$ if

1. **(Smoothness)** $\mathcal{M}$ is a smooth manifold,
2. **(Sharpness)** $N_{\mathcal{M}} = \operatorname{span} N_Q$ on $\mathcal{M}$,
3. **(Continuity)** N_Q varies continuously on $\mathcal{M}$.

Active sets in optimization

Figure: Q is 4×4 Toeplitz spectrahedron



Definition (Partial Smoothness)

A set Q is partly smooth relative to $\mathcal{M} \subset Q$ if

1. **(Smoothness)** $\mathcal{M}$ is a smooth manifold,
2. **(Sharpness)** $N_{\mathcal{M}} = \operatorname{span} N_Q$ on $\mathcal{M}$,
3. **(Continuity)** N_Q varies continuously on $\mathcal{M}$.

(Originates in Lewis '03)

Active sets in optimization

Partial smoothness has classical roots!

Eg: Smooth constraints

$$Q := \{x : g_i(x) \leq 0, \quad \text{for } i = 1, \ldots, m\}$$

where $g_1, \ldots, g_m$ are smooth.

Active sets in optimization

Partial smoothness has classical roots!

Eg: Smooth constraints

$$Q := \{x : g_i(x) \leq 0, \quad \text{for } i = 1, \ldots, m\}$$

where $g_1, \ldots, g_m$ are smooth.

Active indices:

$$I(x) = \{i : g_i(x) = 0\}$$

Active sets in optimization

Partial smoothness has classical roots!

Eg: Smooth constraints

$$Q := \{x : g_i(x) \leq 0, \quad \text{for } i = 1, \ldots, m\}$$

where $g_1, \ldots, g_m$ are smooth.

Active indices:

$$I(x) = \{i : g_i(x) = 0\}$$

Reasonable conditions $\Longrightarrow$

$\mathcal{M} := \{x : I(x) = I(\bar{x})\}$ is a partly smooth manifold near $\bar{x}$

Active sets in optimization

Partial smoothness has classical roots!

Eg: Smooth constraints

$$Q := \{x : g_i(x) \leq 0, \quad \text{for } i = 1, \ldots, m\}$$

where $g_1, \ldots, g_m$ are smooth.

Active indices:

$$I(x) = \{i : g_i(x) = 0\}$$

Reasonable conditions $\Longrightarrow$

$\mathcal{M} := \{x : I(x) = I(\bar{x})\}$ is a partly smooth manifold near $\bar{x}$

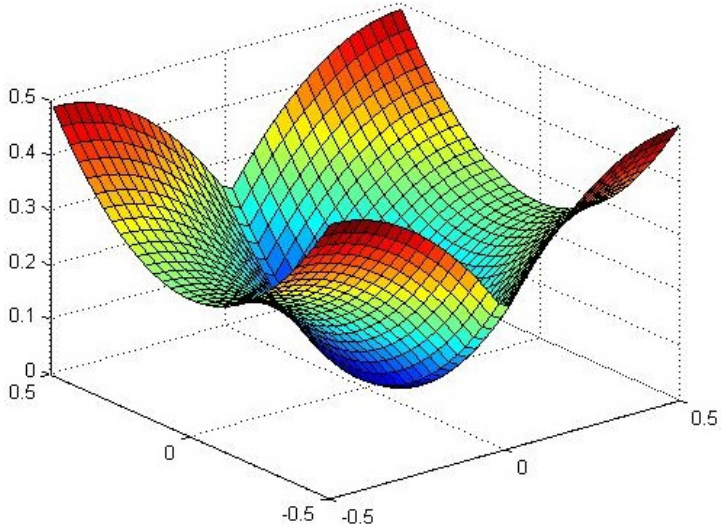


Figure: $Q = \{(x, y, z) : z \geq x(1 - x) + y^2,\ z \geq -x(1 + x) + y^2\}$

Active sets in optimization

Partial smoothness has classical roots!

Eg: Sum of perturbed norms

$$\min_x \; f(x) := \sum_{i=1}^{m} \|F_i(x)\|$$

where $F_1, \ldots, F_m$ are smooth.

Active sets in optimization

Partial smoothness has classical roots!

Eg: Sum of perturbed norms

$$\min_x \; f(x) := \sum_{i=1}^{m} \|F_i(x)\|$$

where $F_1, \ldots, F_m$ are smooth.

Active indices:

$$I(x) := \{i : F_i(x) = 0\}$$

Active sets in optimization

Partial smoothness has classical roots!

Eg: Sum of perturbed norms

$$\min_x \; f(x) := \sum_{i=1}^{m} \|F_i(x)\|$$

where $F_1, \ldots, F_m$ are smooth.

Active indices:

$$I(x) := \{i : F_i(x) = 0\}$$

Reasonable conditions $\Longrightarrow$ $\operatorname{epi} f$ is partly smooth relative to

$$\mathcal{M} := \{(x, f(x)) : I(x) = I(\bar{x})\}$$

Active sets in optimization

Partial smoothness has classical roots!

Eg: Sum of perturbed norms

$$\min_x \; f(x) := \sum_{i=1}^{m} \|F_i(x)\|$$

where $F_1, \ldots, F_m$ are smooth.

Active indices:

$$I(x) := \{i : F_i(x) = 0\}$$

Reasonable conditions $\Longrightarrow$ epi f is partly smooth relative to

$$\mathcal{M} := \{(x, f(x)) : I(x) = I(\bar{x})\}$$

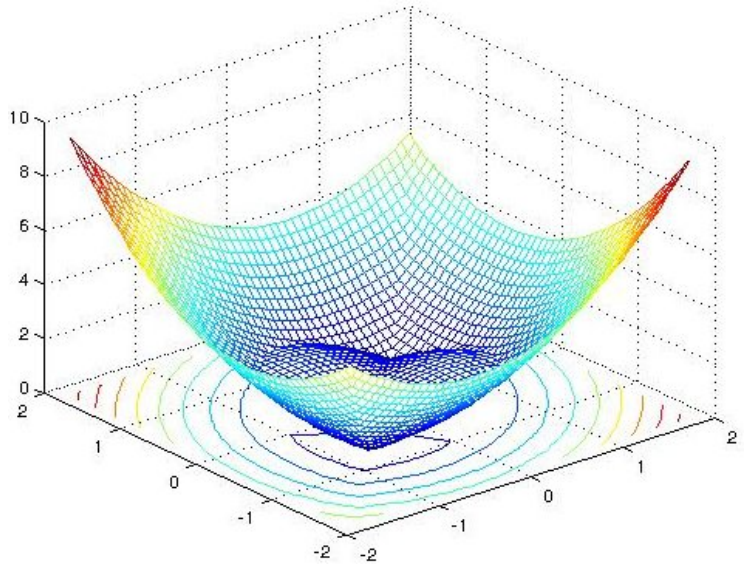


Figure: $f(x, y) := |x^2 + y^2 - 1| + |x - y|$

Active sets in optimization

Why do optimizers care?

Active sets in optimization

$\boxed{\text{Why do optimizers care?}}$

- Many optimization algorithms identify $\mathcal{M}$ in finite time!

Eg: Gradient projection, Newton-like, proximal point.

Active sets in optimization

$\boxed{\text{Why do optimizers care?}}$

- Many optimization algorithms identify $\mathcal{M}$ in finite time!

Eg: Gradient projection, Newton-like, proximal point.

$\Longrightarrow$ Acceleration strategies!

Active sets in optimization

$\boxed{\text{Why do optimizers care?}}$

- Many optimization algorithms identify $\mathcal{M}$ in finite time!

Eg: Gradient projection, Newton-like, proximal point.

$\Longrightarrow$ Acceleration strategies!

Eg: Wright, Burke-Moré, Ferris, Flåm, ...

Active sets in optimization

$$\boxed{\text{Why do optimizers care?}}$$

- Many optimization algorithms identify $\mathcal{M}$ in finite time!

Eg: Gradient projection, Newton-like, proximal point.

$\Longrightarrow$ Acceleration strategies!

Eg: Wright, Burke-Moré, Ferris, Flåm, ...

Finite Identification: For $\bar{x} \in \mathcal{M}$ and $\bar{v} \in \operatorname{ri} N_Q(\bar{x})$, have

$$\boxed{\left.\begin{array}{c} x_i \to \bar{x},\, v_i \to \bar{v} \\ v_i \in N_Q(x_i) \end{array}\right\} \Longrightarrow x_i \in \mathcal{M} \text{ for all large } i}$$

Active sets in optimization

$$\boxed{\text{Why do optimizers care?}}$$

- Many optimization algorithms identify $\mathcal{M}$ in finite time!

Eg: Gradient projection, Newton-like, proximal point.

$\Longrightarrow$ Acceleration strategies!

Eg: Wright, Burke-Moré, Ferris, Flåm, ...

Finite Identification: For $\bar{x} \in \mathcal{M}$ and $\bar{v} \in \operatorname{ri} N_Q(\bar{x})$, have

$$\boxed{\left.\begin{array}{c} x_i \to \bar{x},\, v_i \to \bar{v} \\ v_i \in N_Q(x_i) \end{array}\right\} \Longrightarrow x_i \in \mathcal{M} \text{ for all large } i}$$

finite identification $\Longleftrightarrow$ partial smoothness (D-Lewis '13)

Active sets in optimization

$$\boxed{\text{Why do optimizers care?}}$$

- Many optimization algorithms identify $\mathcal{M}$ in finite time!

Eg: Gradient projection, Newton-like, proximal point.

$\Longrightarrow$ Acceleration strategies!

Eg: Wright, Burke-Moré, Ferris, Flåm, ...

Finite Identification: For $\bar{x} \in \mathcal{M}$ and $\bar{v} \in \operatorname{ri} N_Q(\bar{x})$, have

$$\boxed{\left.\begin{array}{c} x_i \to \bar{x},\, v_i \to \bar{v} \\ v_i \in N_Q(x_i) \end{array}\right\} \Longrightarrow x_i \in \mathcal{M} \text{ for all large } i}$$

finite identification $\Longleftrightarrow$ partial smoothness (D-Lewis '13)

Generic existence (D-Lewis): For semi-algebraic Q consider

$$\max\{\langle v, x\rangle : x \in Q\}$$

Active sets in optimization

$$\boxed{\text{Why do optimizers care?}}$$

- Many optimization algorithms identify $\mathcal{M}$ in finite time!

Eg: Gradient projection, Newton-like, proximal point.

$\Longrightarrow$ Acceleration strategies!

Eg: Wright, Burke-Moré, Ferris, Flåm, ...

Finite Identification: For $\bar{x} \in \mathcal{M}$ and $\bar{v} \in \operatorname{ri} N_Q(\bar{x})$, have

$$\boxed{\left.\begin{array}{c} x_i \to \bar{x},\, v_i \to \bar{v} \\ v_i \in N_Q(x_i) \end{array}\right\} \Longrightarrow x_i \in \mathcal{M} \text{ for all large } i}$$

finite identification $\Longleftrightarrow$ partial smoothness (D-Lewis '13)

Generic existence (D-Lewis): For semi-algebraic Q consider

$$\max\{\langle v, x\rangle : x \in Q\}$$

Then for "typical" v at any local minimizer: unique partly smooth manifold, strict complementarity, quadratic growth.

Active sets in optimization

$$\boxed{\text{Why do optimizers care?}}$$

- Many optimization algorithms identify $\mathcal{M}$ in finite time!

Eg: Gradient projection, Newton-like, proximal point.

$\Longrightarrow$ Acceleration strategies!

Eg: Wright, Burke-Moré, Ferris, Flåm, ...

Finite Identification: For $\bar{x} \in \mathcal{M}$ and $\bar{v} \in \operatorname{ri} N_Q(\bar{x})$, have

$$\boxed{\left.\begin{array}{c} x_i \to \bar{x},\, v_i \to \bar{v} \\ v_i \in N_Q(x_i) \end{array}\right\} \Longrightarrow x_i \in \mathcal{M} \text{ for all large } i}$$

finite identification $\Longleftrightarrow$ partial smoothness (D-Lewis '13)

Generic existence (D-Lewis): For semi-algebraic Q consider

$$\max\{\langle v, x\rangle : x \in Q\}$$

Then for "typical" v at any local minimizer: unique partly smooth manifold, strict complementarity, quadratic growth.

- cf. Spingarn-Rockafellar '79, Pataki-Tunçel '01,...

Summary

- Slope is an elegant tool.
- Variational analysis is especially effective for semi-algebraic function.
- Consequences for alternating projections, steepest descent, active sets.

Thank you.

Spectral sets

How to see this structure in eigenvalue optimization?

Spectral sets

How to see this structure in eigenvalue optimization?

Consider $\mathbf{S}^n := \{n \times n \text{ symmetric matrices}\}$ and the eigenvalue map

$$A \mapsto (\lambda_1(A), \ldots, \lambda_n(A))$$

where

$$\lambda_1(A) \leq \ldots \leq \lambda_n(A).$$

Spectral sets

How to see this structure in eigenvalue optimization?

Consider $\mathbf{S}^n := \{n \times n \text{ symmetric matrices}\}$ and the eigenvalue map

$$A \mapsto (\lambda_1(A), \ldots, \lambda_n(A))$$

where

$$\lambda_1(A) \leq \ldots \leq \lambda_n(A).$$

Spectral sets:

$$\lambda^{-1}(Q) = \{A \in \mathbf{S}^n : \lambda(A) \in Q\}$$

Spectral sets

How to see this structure in eigenvalue optimization?

Consider $\mathbf{S}^n := \{n \times n \text{ symmetric matrices}\}$ and the eigenvalue map

$$A \mapsto (\lambda_1(A), \ldots, \lambda_n(A))$$

where

$$\lambda_1(A) \leq \ldots \leq \lambda_n(A).$$

Spectral sets:

$$\lambda^{-1}(Q) = \{A \in \mathbf{S}^n : \lambda(A) \in Q\}$$

where Q is permutation-invariant.

Spectral sets

How to see this structure in eigenvalue optimization?

Consider $\mathbf{S}^n := \{n \times n \text{ symmetric matrices}\}$ and the eigenvalue map

$$A \mapsto (\lambda_1(A), \ldots, \lambda_n(A))$$

where

$$\lambda_1(A) \leq \ldots \leq \lambda_n(A).$$

Spectral sets:

$$\lambda^{-1}(Q) = \{A \in \mathbf{S}^n : \lambda(A) \in Q\}$$

where Q is permutation-invariant.

(e.g. $\{A : \lambda_n(A) \leq 1\}$ or $\{A : \operatorname{rank} A \leq r\}$)

Spectral sets

How to see this structure in eigenvalue optimization?

Consider $\mathbf{S}^n := \{n \times n \text{ symmetric matrices}\}$ and the eigenvalue map

$$A \mapsto (\lambda_1(A), \ldots, \lambda_n(A))$$

where

$$\lambda_1(A) \leq \ldots \leq \lambda_n(A).$$

Spectral sets:

$$\lambda^{-1}(Q) = \{A \in \mathbf{S}^n : \lambda(A) \in Q\}$$

where Q is permutation-invariant.

$$\text{(e.g. } \{A : \lambda_n(A) \leq 1\} \text{ or } \{A : \operatorname{rank} A \leq r\}\text{)}$$

- Partial smoothness of $\lambda^{-1}(Q)$ / partial smoothness of Q.

Spectral sets

Recognizing partial smoothness (Daniilidis-D-Lewis):

Q partly smooth at $\lambda(\bar{X})$ relative to $\mathcal{M}$

$\Longrightarrow \lambda^{-1}(Q)$ partly smooth at $\bar{X}$ relative to $\lambda^{-1}(\mathcal{M})$.

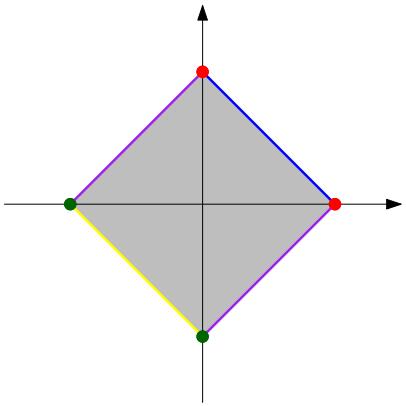
Spectral sets

Recognizing partial smoothness (Daniilidis-D-Lewis):

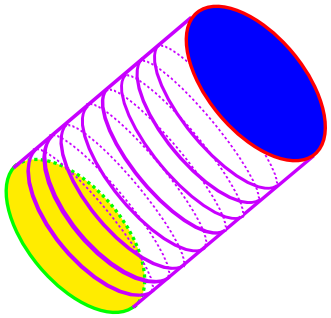
Q partly smooth at $\lambda(\bar{X})$ relative to $\mathcal{M}$

$\Longrightarrow \lambda^{-1}(Q)$ partly smooth at $\bar{X}$ relative to $\lambda^{-1}(\mathcal{M})$.

Eg:



$|x| + |y| \leq 1$



$|\lambda_1(X)| + |\lambda_2(X)| \leq 1$

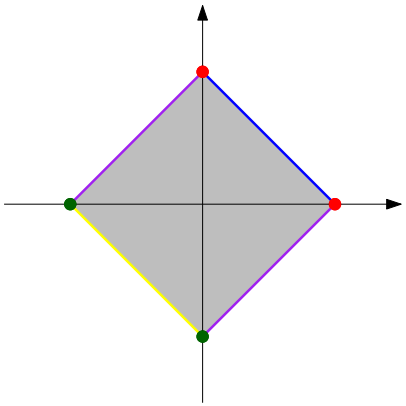
Spectral sets

Recognizing partial smoothness (Daniilidis-D-Lewis):

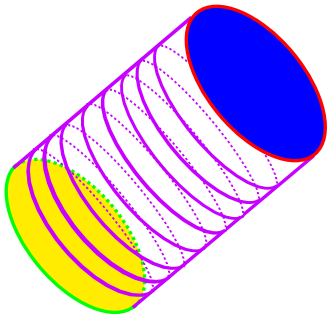
Q partly smooth at $\lambda(\bar{X})$ relative to $\mathcal{M}$

$\Longrightarrow \lambda^{-1}(Q)$ partly smooth at $\bar{X}$ relative to $\lambda^{-1}(\mathcal{M})$.

Eg:



$|x| + |y| \leq 1$



$|\lambda_1(X)| + |\lambda_2(X)| \leq 1$

Related work: Daniilidis-Malick-Sendov '11