Slope and variational geometry in optimization

Dmitriy Drusvyatskiy
Mathematics, University of Washington

ISMP 2015

A.W. Tucker Prize session

AFOSR YIP

Fix a function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a point $\bar{x}$.

Fix a function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a point $\bar{x}$.

Slope: "Fastest instantaneous rate of decrease"

Fix a function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a point $\bar{x}$.

Slope: "Fastest instantaneous rate of decrease"

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

Fix a function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a point $\bar{x}$.

Slope: "Fastest instantaneous rate of decrease"

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

- If $f\colon \mathbf{R}^n \to \mathbf{R}$ is smooth, then $|\nabla f|(\bar{x}) = |\nabla f(\bar{x})|$.

Fix a function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a point $\bar{x}$.

Slope: "Fastest instantaneous rate of decrease"

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

- If $f\colon \mathbf{R}^n \to \mathbf{R}$ is smooth, then $|\nabla f|(\bar{x}) = |\nabla f(\bar{x})|$.

Critical points:

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

Fix a function $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ and a point $\bar{x}$.

Slope: "Fastest instantaneous rate of decrease"

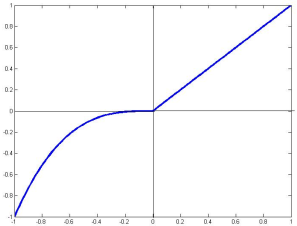
$$|\nabla f|(\bar{x}) := \limsup_{x \to \bar{x}} \frac{f(\bar{x}) - f(x)}{|\bar{x} - x|}$$

- If $f\colon \mathbf{R}^n \to \mathbf{R}$ is smooth, then $|\nabla f|(\bar{x}) = |\nabla f(\bar{x})|$.

Critical points:

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

Eg:

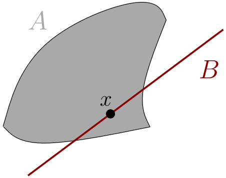


Origin is critical

Method of alternating projections

Common problem:

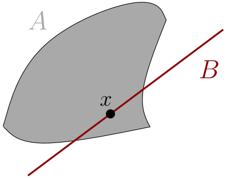
Given sets $A, B \subset \mathbf{R}^n$, find some point $x \in A \cap B$.



Method of alternating projections

Common problem:

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



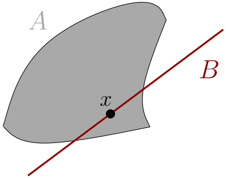
Distance and projection:

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

Method of alternating projections

Common problem:

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



Distance and projection:

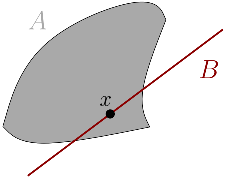
$$d_B(x) = \min_{y \in B} |x - y| \quad \text{and} \quad P_B(x) = \{\text{nearest points of } B \text{ to } x\}.$$

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!

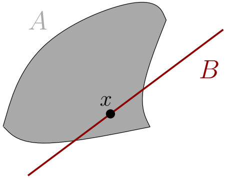
Eg 1 (Low-order recovery):

$$\{x : \operatorname{rank} x \le r\} \bigcap \{x : Ax = b\}.$$

Method of alternating projections

Common problem:

Given sets $A, B \subset \mathbf{R}^n$, 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 (Low-order recovery):

$$\{x : \operatorname{rank} x \leq r\} \bigcap \{x : Ax = b\}.$$

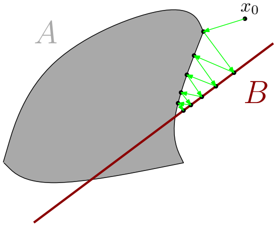
Eg 2 (Low-rank SDP):

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

Method of alternating projections

Method of alternating projections (von Neumann '33):

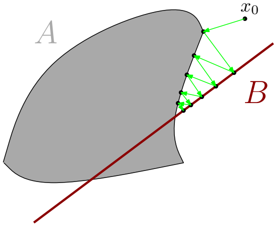
$$\boxed{\begin{array}{lll} x_{k+1} & \in & P_B(x_k) \\ x_{k+2} & \in & P_A(x_{k+1}) \end{array}}$$



Method of alternating projections

Method of alternating projections (von Neumann '33):

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

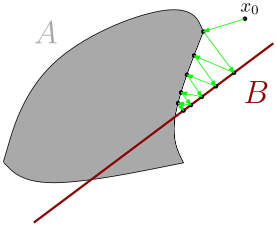


The "angle" between A and B drives the convergence!

Method of alternating projections

Method of alternating projections (von Neumann '33):

$$
\begin{array}{|lll|}
\hline
x_{k+1} & \in & P_B(x_k) \\
x_{k+2} & \in & P_A(x_{k+1}) \\
\hline
\end{array}
$$



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

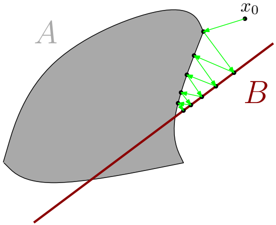
Coupling function:

$$\psi(x, y) := \delta_A(x) + |x - y| + \delta_B(y).$$

Method of alternating projections

Method of alternating projections (von Neumann '33):

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



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

Coupling function:

$$\psi(x, y) := \delta_A(x) + |x - y| + \delta_B(y).$$

If $|\nabla\psi| \geq \frac{1}{\kappa}$ near $A \cap B$, then get local linear convergence!

(D-Ioffe-Lewis '13)

Slope and error bounds

Common problem: Estimate

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

Slope and error bounds

Common problem: Estimate

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

“The residual”:

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

Slope and 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

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

Slope and 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" $f^{-1}(a, b)$.

Slope and 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:

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

Geometry of the slope

Fréchet subdifferential:

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

Geometry of the slope

Fréchet subdifferential:

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

Equivalently

$$v \in \hat{\partial} f(\bar{x}) \quad \Longleftrightarrow \quad f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle + o(|x - \bar{x}|).$$

Geometry of the slope

Fréchet subdifferential:

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

Equivalently

$$v \in \hat{\partial} f(\bar{x}) \quad \iff \quad f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle + o(|x - \bar{x}|).$$

Limiting slope and subdifferential:

$$\overline{|\nabla f|} := \operatorname{cl} |\nabla f| \qquad\qquad \partial f := \operatorname{cl} \hat{\partial} f.$$

Geometry of the slope

Fréchet subdifferential:

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

Equivalently

$$v \in \hat{\partial} f(\bar{x}) \quad \Longleftrightarrow \quad f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle + o(|x - \bar{x}|).$$

Limiting slope and subdifferential:

$$\overline{|\nabla f|} := \operatorname{cl} |\nabla f| \qquad\qquad \partial f := \operatorname{cl} \hat{\partial} f.$$

Relationship (Ioffe '00):

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

Geometry of the slope

Fréchet subdifferential:

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

Equivalently

$$v \in \hat{\partial} f(\bar{x}) \quad \Longleftrightarrow \quad f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle + o(|x - \bar{x}|).$$

Limiting slope and subdifferential:

$$\overline{|\nabla f|} := \mathrm{cl}\, |\nabla f| \qquad\qquad \partial f := \mathrm{cl}\, \hat{\partial} f.$$

Relationship (Ioffe '00):

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

Back to the coupling function

$$\psi(x, y) := \delta_A(x) + |x - y| + \delta_B(y).$$

Geometry of the slope

Fréchet subdifferential:

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

Equivalently

$$v \in \hat{\partial} f(\bar{x}) \quad \iff \quad f(x) \geq f(\bar{x}) + \langle v, x - \bar{x} \rangle + o(|x - \bar{x}|).$$

Limiting slope and subdifferential:

$$\overline{|\nabla f|} := \operatorname{cl} |\nabla f| \qquad\qquad \partial f := \operatorname{cl} \hat{\partial} f.$$

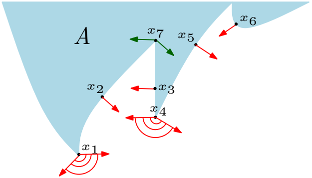
Relationship (Ioffe '00):

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

Back to the coupling function

$$\psi(x, y) := \delta_A(x) + |x - y| + \delta_B(y).$$

Normal cone $N_A(x) = \partial \delta_A(x)$ appears:



Transversality

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

Transversality

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

Eg:

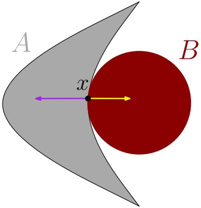


Figure: Not transverse

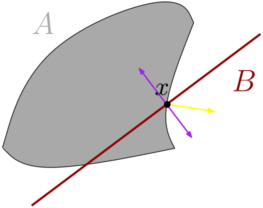


Figure: Transverse

Transversality

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

Eg:

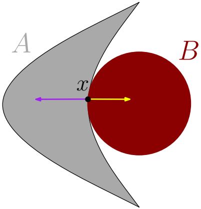


Figure: Not transverse

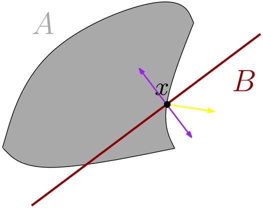


Figure: Transverse

Convergence of alternating projections (D-Ioffe-Lewis '13):

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

Transversality

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

Eg:

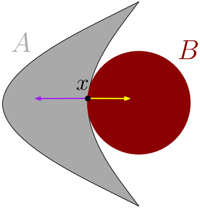


Figure: Not transverse

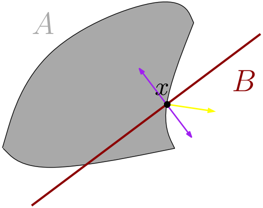


Figure: Transverse

Convergence of alternating projections (D-Ioffe-Lewis '13):

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

Extremal principle (Mordukhovich):

Transversality $\implies$ $(A + a) \cap B \neq \emptyset$ for small a.

Subdifferentials of eigenvalue functions

Eigenvalue functions

- $\mathbf{S}^n := \{n \times n \text{ symmetric matrices}\}$
- The eigenvalue map $\lambda(A) = (\lambda_1(A), \ldots, \lambda_n(A))$

Eigenvalue functions

- $\mathbf{S}^n := \{n \times n \text{ symmetric matrices}\}$
- The eigenvalue map $\lambda(A) = (\lambda_1(A), \dots, \lambda_n(A))$

Eigenvalue functions:

$$F = f \circ \lambda$$

where $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is permutation-invariant.

Eigenvalue functions

- $\mathbf{S}^n := \{n \times n \text{ symmetric matrices}\}$
- The eigenvalue map $\lambda(A) = (\lambda_1(A), \ldots, \lambda_n(A))$

Eigenvalue functions:

$$F = f \circ \lambda$$

where $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is permutation-invariant.

E.g. $\|X\|_* = |\lambda(X)|_1 \quad \text{and} \quad \operatorname{rank} X = \operatorname{rank} \lambda(X)$.

Eigenvalue functions

- $\mathbf{S}^n := \{n \times n \text{ symmetric matrices}\}$
- The eigenvalue map $\lambda(A) = (\lambda_1(A), \ldots, \lambda_n(A))$

Eigenvalue functions:

$$F = f \circ \lambda$$

where $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is permutation-invariant.

E.g. $\|X\|_* = |\lambda(X)|_1$ and $\operatorname{rank} X = \operatorname{rank} \lambda(X)$.

THE TRANSFER PRINCIPLE:

Variational properties of F and f are in 1-to-1 correspondence

Eigenvalue functions

- $\mathbf{S}^n := \{n \times n \text{ symmetric matrices}\}$
- The eigenvalue map $\lambda(A) = (\lambda_1(A), \ldots, \lambda_n(A))$

Eigenvalue functions:

$$F = f \circ \lambda$$

where $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is permutation-invariant.

E.g. $\|X\|_* = |\lambda(X)|_1 \quad$ and $\quad \operatorname{rank} X = \operatorname{rank} \lambda(X)$.

THE TRANSFER PRINCIPLE:

Variational properties of F and f are in 1-to-1 correspondence

Eg: Smoothness, convexity, prox-regularity, stratifications, faces, active manifolds

Eigenvalue functions

- $\mathbf{S}^n := \{n \times n \text{ symmetric matrices}\}$
- The eigenvalue map $\lambda(A) = (\lambda_1(A), \ldots, \lambda_n(A))$

Eigenvalue functions:

$$F = f \circ \lambda$$

where $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$ is permutation-invariant.

E.g. $\|X\|_* = |\lambda(X)|_1$ and $\operatorname{rank} X = \operatorname{rank} \lambda(X)$.

THE TRANSFER PRINCIPLE:

Variational properties of F and f are in 1-to-1 correspondence

Eg: Smoothness, convexity, prox-regularity, stratifications, faces, active manifolds

(Ball, Daniilidis, Davis, Lewis, Malick, Sendov, Šilhavý, Sylvester, von Neumann, ...)

Subdifferential formula

Main tool (Lewis '99): $V \in \partial(f \circ \lambda)(X)$ exactly when

$$\left\{ \begin{array}{c} V = U(\operatorname{Diag} v)\,U^T \\ X = U(\operatorname{Diag} \lambda(X))\,U^T \end{array} \right\} \text{ with } U \in \mathbb{O}^n \text{ and } v \in \partial f(\lambda(X)).$$

Subdifferential formula

Main tool (Lewis '99): $V \in \partial(f \circ \lambda)(X)$ exactly when

$$\left\{ \begin{array}{c} V = U(\operatorname{Diag} v)\,U^T \\ X = U(\operatorname{Diag} \lambda(X))\,U^T \end{array} \right\} \text{ with } U \in \mathbb{O}^n \text{ and } v \in \partial f(\lambda(X)).$$

Convex proof (Lewis '96):

$$\boxed{(f \circ \lambda)^\star = f^\star \circ \lambda}$$

where

$$f^\star(x) = \sup_y \{\langle y, x\rangle - f(y)\} \qquad \text{is the Fenchel conjugate.}$$

Subdifferential formula

Main tool (Lewis '99): $V \in \partial(f \circ \lambda)(X)$ exactly when

$$\left\{ \begin{array}{c} V = U(\mathrm{Diag}\, v)\, U^T \\ X = U(\mathrm{Diag}\, \lambda(X))\, U^T \end{array} \right\} \text{ with } U \in \mathbb{O}^n \text{ and } v \in \partial f(\lambda(X)).$$

Convex proof (Lewis '96):

$$\boxed{(f \circ \lambda)^\star = f^\star \circ \lambda}$$

where

$$f^\star(x) = \sup_y \{\langle y, x\rangle - f(y)\} \qquad \text{is the Fenchel conjugate.}$$

Nonconvex proof (D-Kempton '14):

$$\boxed{(f \circ \lambda)_\alpha = f_\alpha \circ \lambda}$$

where

$$f_\alpha(x) = \inf_y \left\{ f(y) + \frac{1}{2\alpha}|y - x|^2 \right\} \qquad \text{is the Moreau envelope.}$$

Active manifolds

Active manifolds

Definition (Active manifolds)

Manifold $\mathcal{M} \subset \mathbf{R}^n$ is active for f at $\bar{x}$ provided $f\big|_{\mathcal{M}}$ is smooth and

$$\left.\begin{array}{c} x_i \to \bar{x}, v_i \to 0 \\ v_i \in \partial f(x_i) \end{array}\right\} \Longrightarrow x_i \in \mathcal{M} \text{ for all large } i$$

Active manifolds

Definition (Active manifolds)

Manifold $\mathcal{M} \subset \mathbf{R}^n$ is active for f at $\bar{x}$ provided $f\big|_{\mathcal{M}}$ is smooth and

$$\left.\begin{array}{c} x_i \to \bar{x},\, v_i \to 0 \\ v_i \in \partial f(x_i) \end{array}\right\} \implies x_i \in \mathcal{M} \text{ for all large } i$$

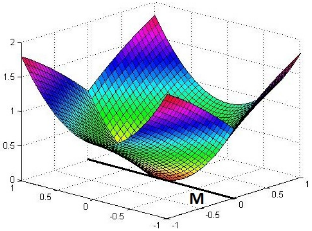


Figure: $f(x, y) = y^2 - x^2 + |x|$, $\qquad \mathcal{M} = \{0\} \times \mathbf{R}$

Active manifolds

Definition (Active manifolds)

Manifold $\mathcal{M} \subset \mathbf{R}^n$ is active for f at $\bar{x}$ provided $f\big|_{\mathcal{M}}$ is smooth and

$$\left.\begin{array}{c} x_i \to \bar{x},\, v_i \to 0 \\ v_i \in \partial f(x_i) \end{array}\right\} \Longrightarrow x_i \in \mathcal{M} \text{ for all large } i$$

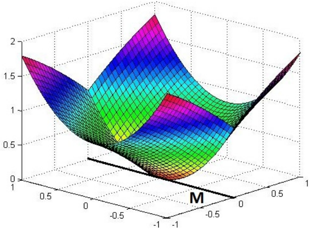


Figure: $f(x, y) = y^2 - x^2 + |x|, \qquad \mathcal{M} = \{0\} \times \mathbf{R}$

Introduced by Wright '93, Lewis '04.

• **Common**: appear from regularizers (l_∞, $l_{1,2}$, $\|\cdot\|_*$, TV), and are generic for semi-algebraic problems (D-Ioffe-Lewis '14).

Active manifolds

Definition (Active manifolds)

Manifold $\mathcal{M} \subset \mathbf{R}^n$ is active for f at $\bar{x}$ provided $f\big|_{\mathcal{M}}$ is smooth and

$$\left.\begin{array}{c} x_i \to \bar{x},\ v_i \to 0 \\ v_i \in \partial f(x_i) \end{array}\right\} \Longrightarrow x_i \in \mathcal{M} \text{ for all large } i$$

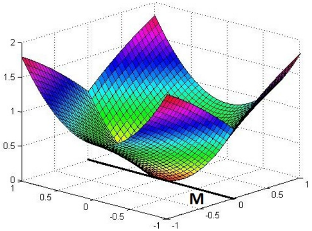


Figure: $f(x, y) = y^2 - x^2 + |x|,\qquad \mathcal{M} = \{0\} \times \mathbf{R}$

Introduced by Wright '93, Lewis '04.

- **Common**: appear from regularizers (l_∞, $l_{1,2}$, $\|\cdot\|_*$, TV), and are generic for semi-algebraic problems (D-Ioffe-Lewis '14).
- **Computation**: algorithm acceleration on $\mathcal{M}$. (Fadili at ISMP)

Active manifolds

Definition (Active manifolds)

Manifold $\mathcal{M} \subset \mathbf{R}^n$ is active for f at $\bar{x}$ provided $f|_{\mathcal{M}}$ is smooth and

$$\left.\begin{array}{c} x_i \to \bar{x},\, v_i \to 0 \\ v_i \in \partial f(x_i) \end{array}\right\} \Longrightarrow x_i \in \mathcal{M} \text{ for all large } i$$

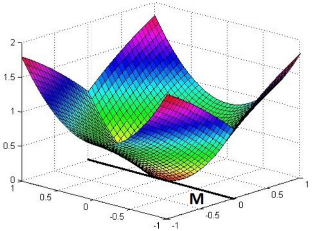


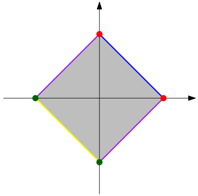
Figure: $f(x,y) = y^2 - x^2 + |x|$, $\quad \mathcal{M} = \{0\} \times \mathbf{R}$

Introduced by Wright '93, Lewis '04.

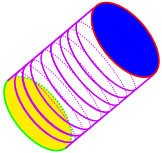
- **Common**: appear from regularizers (l_∞, $l_{1,2}$, $\|\cdot\|_*$, TV), and are generic for semi-algebraic problems (D-Ioffe-Lewis '14).
- **Computation**: algorithm acceleration on $\mathcal{M}$. (Fadili at ISMP)
- **Sensitivity**: only the restriction $f|_{\mathcal{M}}$ matters (D-Lewis '14).

Spectral active sets

Example:

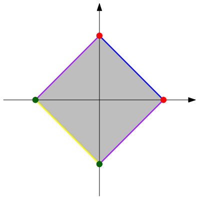


$|x| + |y| \leq 1$

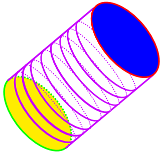


$|\lambda_1(X)| + |\lambda_2(X)| \leq 1$

Example:



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

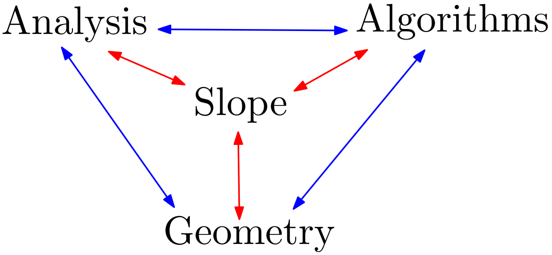


$|\lambda_1(X)| + |\lambda_2(X)| \leq 1$

Spectral active sets (Daniilidis-D-Lewis '14):

$\mathcal{M}$ active manifold at $\lambda(X)$ for f

$\iff \lambda^{-1}(\mathcal{M})$ active manifold at X for $f \circ \lambda$.


Analysis
Algorithms
Slope
Geometry

Thank you.