Tame variational analysis

Dmitriy Drusvyatskiy
Mathematics, University of Washington

Joint work with
Daniilidis (Chile), Ioffe (Technion), and Lewis (Cornell)

May 19, 2015

Theme:

Semi-algebraic geometry is a powerful addition to the Variational Analysis toolkit.

Illustration

For closed, convex $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, the following are equivalent

Quadratic growth:

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

Error bound:

$$|x - \bar{x}| \leq \kappa \cdot \mathrm{dist}\big(0, \partial f(x)\big) \qquad \text{for } x \text{ near } \bar{x}.$$

Illustration

For closed, convex $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, the following are equivalent

Quadratic growth:

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

Error bound:

$$|x - \bar{x}| \leq \kappa \cdot \operatorname{dist}\big(0, \partial f(x)\big) \qquad \text{for } x \text{ near } \bar{x}.$$

(Proved by Kummer '91, Aragon-Geoffroy '08)

Illustration

For closed, convex $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, the following are equivalent

Quadratic growth:

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

Error bound:

$$|x - \bar{x}| \leq \kappa \cdot \operatorname{dist}\big(0, \partial f(x)\big) \qquad \text{for } x \text{ near } \bar{x}.$$

(Proved by Kummer '91, Aragon-Geoffroy '08)

Is this true generally?

Illustration

For closed, convex $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, the following are equivalent

Quadratic growth:

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

Error bound:

$$|x-\bar{x}| \leq \kappa \cdot \operatorname{dist}\big(0, \partial f(x)\big) \qquad \text{for } x \text{ near } \bar{x}.$$

(Proved by Kummer '91, Aragon-Geoffroy '08)

Is this true generally?

The implication $\Uparrow$ always holds (D-Mordukhovich-Nghia '14).

Illustration

For closed, convex $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, the following are equivalent

Quadratic growth:

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

Error bound:

$$|x - \bar{x}| \leq \kappa \cdot \operatorname{dist}\big(0, \partial f(x)\big) \qquad \text{for } x \text{ near } \bar{x}.$$

(Proved by Kummer '91, Aragon-Geoffroy '08)

Is this true generally?

The implication $\Uparrow$ always holds (D-Mordukhovich-Nghia '14).
But the converse $\Downarrow$ can easily fail.

Illustration

For closed, convex $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, the following are equivalent

Quadratic growth:

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

Error bound:

$$|x - \bar{x}| \leq \kappa \cdot \mathrm{dist}\big(0, \partial f(x)\big) \qquad \text{for } x \text{ near } \bar{x}.$$

(Proved by Kummer '91, Aragon-Geoffroy '08)

Is this true generally?

The implication $\Uparrow$ always holds (D-Mordukhovich-Nghia '14).
But the converse $\Downarrow$ can easily fail.

Theorem (D-Ioffe)

The equivalence holds at local minimizers of semi-algebraic f.

Illustration

For closed, convex $f\colon \mathbf{R}^n \to \overline{\mathbf{R}}$, the following are equivalent

Quadratic growth:

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

Error bound:

$$|x - \bar{x}| \leq \kappa \cdot \operatorname{dist}\big(0, \partial f(x)\big) \qquad \text{for } x \text{ near } \bar{x}.$$

(Proved by Kummer '91, Aragon-Geoffroy '08)

Is this true generally?

The implication $\Uparrow$ always holds (D-Mordukhovich-Nghia '14).
But the converse $\Downarrow$ can easily fail.

Theorem (D-Ioffe)

The equivalence holds at local minimizers of semi-algebraic f.

(More in Ioffe's talk tomorrow.)

Outline

- Basics of semi-algebraic geometry
- Consequences for Variational Analysis:
  - Subgradient descent
  - Sweeping process
  - Sard theorem
  - Size of subdifferential graphs
  - Approximation on singular domains

Semi-algebraic geometry

Semi-algebraic set: finite union of sets

$$\left\{ x : \quad \begin{array}{l} p_i(x) < 0 \text{ for } i \in I \\ p_j(x) = 0 \text{ for } j \in J \end{array} \right\}$$

where p_i, p_j are polynomials.

Semi-algebraic geometry

Semi-algebraic set: finite union of sets

$$\left\{ x : \quad \begin{array}{l} p_i(x) < 0 \text{ for } i \in I \\ p_j(x) = 0 \text{ for } j \in J \end{array} \right\}$$

where p_i, p_j are polynomials.

A mapping $F\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ is semi-algebraic if

$$\operatorname{gph} F = \{(x,y) : y \in F(x)\} \qquad \text{is semi-algebraic.}$$

Semi-algebraic geometry

Semi-algebraic set: finite union of sets

$$\left\{ x : \quad \begin{array}{l} p_i(x) < 0 \text{ for } i \in I \\ p_j(x) = 0 \text{ for } j \in J \end{array} \right\}$$

where p_i, p_j are polynomials.

A mapping $F\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ is semi-algebraic if

$$\operatorname{gph} F = \{(x, y) : y \in F(x)\} \qquad \text{is semi-algebraic.}$$

Robustness:

- Boolean operations preserve semi-algebraic sets.
- Linear mappings preserve semi-algebraic sets (Tarski-Seidenberg).

Semi-algebraic geometry

Semi-algebraic set: finite union of sets

$$\left\{ x : \quad \begin{array}{l} p_i(x) < 0 \text{ for } i \in I \\ p_j(x) = 0 \text{ for } j \in J \end{array} \right\}$$

where p_i, p_j are polynomials.

A mapping $F\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ is semi-algebraic if

$$\operatorname{gph} F = \{(x, y) : y \in F(x)\} \qquad \text{is semi-algebraic.}$$

Robustness:

- Boolean operations preserve semi-algebraic sets.
- Linear mappings preserve semi-algebraic sets (Tarski-Seidenberg).

Eg: Q semi-algebraic $\quad\Longrightarrow\quad \{x : \exists_y \ (x, y) \in Q\}$ semi-algebraic.

Semi-algebraic geometry

Semi-algebraic set: finite union of sets

$$\left\{ x : \quad \begin{array}{l} p_i(x) < 0 \text{ for } i \in I \\ p_j(x) = 0 \text{ for } j \in J \end{array} \right\}$$

where p_i, p_j are polynomials.

A mapping $F\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ is semi-algebraic if

$$\operatorname{gph} F = \{(x, y) : y \in F(x)\} \qquad \text{is semi-algebraic.}$$

Robustness:

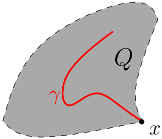
- Boolean operations preserve semi-algebraic sets.
- Linear mappings preserve semi-algebraic sets (Tarski-Seidenberg).

Eg: Q semi-algebraic $\implies$ $\{x : \exists_y \ (x, y) \in Q\}$ semi-algebraic.

Conclusion: ∂f, $|\nabla f|$, $\operatorname{sur} F$, $\operatorname{Lip} F$, $\ldots$ remain semi-algebraic.

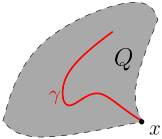
Basic properties

Curve selection: Given $x \in \operatorname{cl} Q$, there is an analytic curve γ with $\gamma(0) = x$ and $\gamma(0, \eta) \subset Q$. (Bruhat-Cartan '50, Milnor '68)



Basic properties

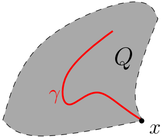
Curve selection: Given $x \in \operatorname{cl} Q$, there is an analytic curve γ with $\gamma(0) = x$ and $\gamma(0, \eta) \subset Q$. (Bruhat-Cartan '50, Milnor '68)



Semi-algebraic selection: Semialgebraic $F \colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ admit semi-algebraic selections $f \subset F$.

Basic properties

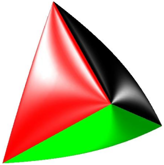
Curve selection: Given $x \in \operatorname{cl} Q$, there is an analytic curve γ with $\gamma(0) = x$ and $\gamma(0, \eta) \subset Q$. (Bruhat-Cartan '50, Milnor '68)



Semi-algebraic selection: Semialgebraic $F\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ admit semi-algebraic selections $f \subset F$.

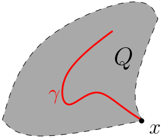
Stratification: Semialgebraic sets "stratify" into finitely many manifolds $\{\mathcal{M}_i\}$, and so have dimension.

(Whitney '65, Łojasiewicz '71)



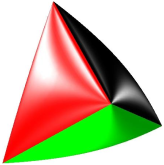
Basic properties

Curve selection: Given $x \in \operatorname{cl} Q$, there is an analytic curve γ with $\gamma(0) = x$ and $\gamma(0, \eta) \subset Q$. (Bruhat-Cartan '50, Milnor '68)



Semi-algebraic selection: Semialgebraic $F \colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ admit semi-algebraic selections $f \subset F$.

Stratification: Semialgebraic sets "stratify" into finitely many manifolds $\{\mathcal{M}_i\}$, and so have dimension. (Whitney '65, Łojasiewicz '71)



Łojasiewicz inequality: If f is semi-algebraic, then on compacta

$$\operatorname{dist}(x; f^{-1}(0)) \leq C|f(x)|^{\alpha}.$$

(Łojasiewicz '91, Kurdyka '98, Bolte-Daniilidis-Lewis '06)

What are consequences for Variational Analysis?

Subgradient systems

Theorem (D-Ioffe-Lewis)

f semi-algebraic, $\bar{x}$ not a local minimizer $\Longrightarrow$ there exists a nontrivial solution to

$$\boxed{\dot{x} \in -\partial f(x) \qquad \text{and} \qquad x(0) = \bar{x}.}$$

Subgradient systems

Theorem (D-Ioffe-Lewis)

f semi-algebraic, $\bar{x}$ not a local minimizer $\Longrightarrow$ there exists a nontrivial solution to

$$\boxed{\dot{x} \in -\partial f(x) \qquad \text{and} \qquad x(0) = \bar{x}.}$$

Moreover, then

$$\dot{x} \text{ is the shortest element of } -\partial f(x) \qquad a.e.$$

Subgradient systems

Theorem (D-Ioffe-Lewis)

f semi-algebraic, $\bar{x}$ not a local minimizer $\Longrightarrow$ there exists a nontrivial solution to

$$\boxed{\dot{x} \in -\partial f(x) \qquad \text{and} \qquad x(0) = \bar{x}.}$$

Moreover, then

$$\dot{x} \text{ is the shortest element of } -\partial f(x) \qquad \text{a.e.}$$

Theorem (Daniilidis-Bolte-Lewis)

If f is semi-algebraic, then any bounded subgradient curve has finite length.

Subgradient systems

Theorem (D-Ioffe-Lewis)

f semi-algebraic, $\bar{x}$ not a local minimizer $\Longrightarrow$ there exists a nontrivial solution to

$$\boxed{\dot{x} \in -\partial f(x) \qquad \textit{and} \qquad x(0) = \bar{x}.}$$

Moreover, then

$$\dot{x} \text{ is the shortest element of } -\partial f(x) \quad \text{a.e.}$$

Theorem (Daniilidis-Bolte-Lewis)

If f is semi-algebraic, then any bounded subgradient curve has finite length.

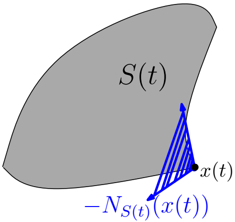
Many analogues for descent methods; e.g. proximal point, splitting, Gauss-Seidel, etc (Attouch, Bolte, Bot, Noll, Peypouquet, Soubeyran, Svaiter, ...)

Sweeping process

Sweeping process (Moreau '77):

$$\dot{x}(t) \in -N_{S(t)}(x(t))$$

with $S(t)$ a moving set.

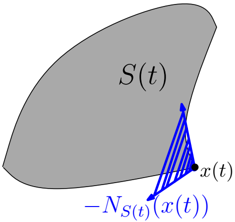


Sweeping process (Moreau '77):

$$\dot{x}(t) \in -N_{S(t)}(x(t))$$

with $S(t)$ a moving set.

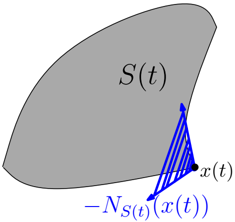
- The monotone case $S(t) = [f \leq t]$ is subgradient descent.



Sweeping process (Moreau '77):

$$\dot{x}(t) \in -N_{S(t)}(x(t))$$

with $S(t)$ a moving set.



- The monotone case $S(t) = [f \leq t]$ is subgradient descent.

Theorem (D-Daniilidis)

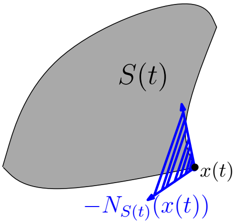
If $S\colon \mathbf{R} \to \mathbf{R}^n$ is semi-algebraic, then every bounded solution of the sweeping process has finite length.

Sweeping process (Moreau '77):

$$\dot{x}(t) \in -N_{S(t)}(x(t))$$

with $S(t)$ a moving set.

- The monotone case $S(t) = [f \leq t]$ is subgradient descent.



Theorem (D-Daniilidis)

If $S: \mathbf{R} \to \mathbf{R}^n$ is semi-algebraic, then every bounded solution of the sweeping process has finite length.

Key estimate:

$$|\dot{x}(t)| \leq \operatorname{Lip} S(t|x(t)) \leq \sup_{x \in S(t) \cap X} \operatorname{Lip} S(t|x)$$

and the upper-bound is integrable by the Łojasiewicz inequality.

Sard Theorem

Mapping $F\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ is metrically regular at $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ if

$$\frac{\operatorname{dist}(x, F^{-1}(y))}{\operatorname{dist}(y, F(x))} \quad \text{is bounded} \quad \text{near } (\bar{x}, \bar{y}).$$

If F is not metrically regular at $(\bar{x}, \bar{y})$, then $\bar{y}$ is a critical value.

Sard Theorem

Mapping $F\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ is metrically regular at $(\bar{x},\bar{y}) \in \operatorname{gph} F$ if

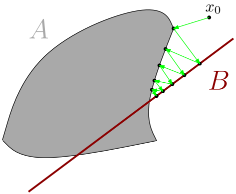
$$\frac{\operatorname{dist}(x, F^{-1}(y))}{\operatorname{dist}(y, F(x))} \quad \text{is bounded} \quad \text{near } (\bar{x},\bar{y}).$$

If F is not metrically regular at $(\bar{x},\bar{y})$, then $\bar{y}$ is a critical value.

Theorem (Ioffe '08)

Semi-algebraic $F\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ have almost no critical values.

- Justifies typical linear convergence of basic schemes, e.g. alternating projections (D-Ioffe-Lewis '13).



Sard Theorem

Mapping $F\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ is metrically regular at $(\bar{x}, \bar{y}) \in \operatorname{gph} F$ if

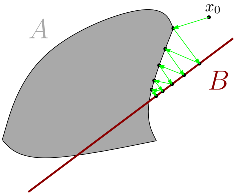
$$\frac{\operatorname{dist}(x, F^{-1}(y))}{\operatorname{dist}(y, F(x))} \quad \text{is bounded} \quad \text{near } (\bar{x}, \bar{y}).$$

If F is not metrically regular at $(\bar{x}, \bar{y})$, then $\bar{y}$ is a critical value.

Theorem (Ioffe '08)

Semi-algebraic $F\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^m$ have almost no critical values.

- Justifies typical linear convergence of basic schemes, e.g. alternating projections (D-Ioffe-Lewis '13).



Sard Theorem & "$\operatorname{gph} \partial f$ is thin"

$\Longrightarrow$ generic properties of semi-algebraic functions.

(cf. Lewis' talk)

Generic properties

Consider

$$\boxed{\min_{x} f(x) + h(G(x) + y) - \langle v, x \rangle}$$

where f, h, G are semi-algebraic and G is C^2-smooth.

Generic properties

Consider

$$\boxed{\min_x f(x) + h(G(x) + y) - \langle v, x \rangle}$$

where f, h, G are semi-algebraic and G is C^2-smooth.

Optimality conditions:

$$\begin{bmatrix} v \\ y \end{bmatrix} \in \begin{bmatrix} \nabla G(x)^* \lambda \\ -G(x) \end{bmatrix} + \left(\partial f \times (\partial h)^{-1} \right) (x, y).$$

Generic properties

Consider

$$\boxed{\min_x f(x) + h(G(x) + y) - \langle v, x \rangle}$$

where f, h, G are semi-algebraic and G is C^2-smooth.

Optimality conditions:

$$\begin{bmatrix} v \\ y \end{bmatrix} \in \begin{bmatrix} \nabla G(x)^* \lambda \\ -G(x) \end{bmatrix} + \left(\partial f \times (\partial h)^{-1} \right)(x, y).$$

Sard theorem & thinness $\implies$ generic properties:

- qualification conditions, strict complementarity, smooth dependance of (x, λ), existence of identifiable manifolds, second order sufficient conditions are necessary at local minimizers.

Generic properties

Consider

$$\boxed{\min_x f(x) + h(G(x) + y) - \langle v, x \rangle}$$

where f, h, G are semi-algebraic and G is C^2-smooth.

Optimality conditions:

$$\begin{bmatrix} v \\ y \end{bmatrix} \in \begin{bmatrix} \nabla G(x)^* \lambda \\ -G(x) \end{bmatrix} + \Big(\partial f \times (\partial h)^{-1}\Big)(x, y).$$

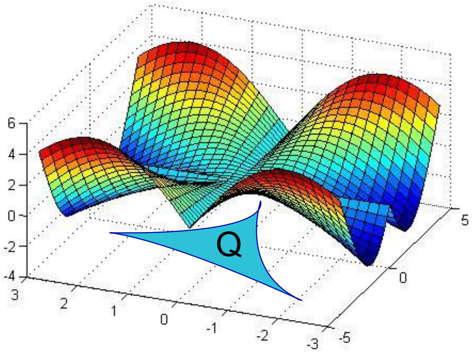
Sard theorem & thinness $\implies$ generic properties:

• qualification conditions, strict complementarity, smooth dependance of (x, λ), existence of identifiable manifolds, second order sufficient conditions are necessary at local minimizers.

Remark: Without semi-algebraicity, one needs geometric measure theory and not all properties above are generic.

Approximation of functions

Set-up: $$Q \hookrightarrow \mathbf{R}^n \xrightarrow{f} \mathbf{R}.$$



Assume Q is a disjoint union of manifolds

$$\boxed{Q = \mathcal{M}_1 \cup \mathcal{M}_2 \cup \ldots \cup \mathcal{M}_{k-1} \cup \mathcal{M}_k}$$

Approximation of functions

Motivation: Integration by parts

$$\int_Q g\Delta f \, dx$$

Approximation of functions

Motivation: Integration by parts

$$\int_Q g\Delta f\,dx = \int_{\partial Q} g\langle \nabla f, \hat{n}\rangle dS - \int_Q \langle \nabla g, \nabla f\rangle\,dx.$$

Approximation of functions

Motivation: Integration by parts

$$\int_Q g\Delta f\,dx = \int_{\partial Q} g\langle \nabla f, \hat{n}\rangle dS - \int_Q \langle \nabla g, \nabla f\rangle\,dx.$$

Goal: approximate f by a $\mathbf{C}^2-$smooth $\tilde{f}$ so that the $\nabla\tilde{f} \perp \hat{n}$.

Approximation of functions

Motivation: Integration by parts

$$\int_Q g\Delta f\,dx = \int_{\partial Q} g\langle \nabla f, \hat{n}\rangle dS - \int_Q \langle \nabla g, \nabla f\rangle\,dx.$$

Goal: approximate f by a $\mathbf{C}^2-$smooth $\tilde{f}$ so that the $\nabla \tilde{f} \perp \hat{n}$.

Theorem (D-Larsson)

Given a continuous $f\colon \mathbf{R}^n \to \mathbf{R}$ and any $\epsilon > 0$, there exists a $\mathbf{C}^1$-smooth $\tilde{f}$ satisfying

1. **Closeness:** $|\tilde{f}(x) - f(x)| < \epsilon$ *for all* $x \in \mathbf{R}^n$,
2. **Neumann Boundary condition:**

$$\boxed{x \in \mathcal{M}_i \quad \Longrightarrow \quad \nabla \tilde{f}(x) \in T_x\mathcal{M}_i}.$$

Approximation of functions

Motivation: Integration by parts

$$\int_Q g\Delta f\, dx = \int_{\partial Q} g\langle \nabla f, \hat{n}\rangle dS - \int_Q \langle \nabla g, \nabla f\rangle\, dx.$$

Goal: approximate f by a $\mathbf{C}^2-$smooth $\tilde{f}$ so that the $\nabla \tilde{f} \perp \hat{n}$.

Theorem (D-Larsson)

Given a continuous $f\colon \mathbf{R}^n \to \mathbf{R}$ *and any* $\epsilon > 0$, *there exists a* $\mathbf{C}^1$*-smooth* $\tilde{f}$ *satisfying*

1. **Closeness:** $|\tilde{f}(x) - f(x)| < \epsilon$ *for all* $x \in \mathbf{R}^n$,
2. **Neumann Boundary condition:**

$$\boxed{x \in \mathcal{M}_i \quad \Longrightarrow \quad \nabla \tilde{f}(x) \in T_x\mathcal{M}_i}.$$

provided $\{M_i\}$ *is a Whitney stratification of* Q.

Approximation of functions

Motivation: Integration by parts

$$\int_Q g\Delta f\,dx = \int_{\partial Q} g\langle \nabla f, \hat{n}\rangle dS - \int_Q \langle \nabla g, \nabla f\rangle\,dx.$$

Goal: approximate f by a $\mathbf{C}^2$−smooth $\tilde{f}$ so that the $\nabla \tilde{f} \perp \hat{n}$.

Theorem (D-Larsson)

Given a continuous $f\colon \mathbf{R}^n \to \mathbf{R}$ *and any* $\epsilon > 0$, *there exists a* $\mathbf{C}^1$*-smooth* $\tilde{f}$ *satisfying*

1. **Closeness:** $|\tilde{f}(x) - f(x)| < \epsilon$ *for all* $x \in \mathbf{R}^n$,
2. **Neumann Boundary condition:**

$$\boxed{x \in \mathcal{M}_i \quad \Longrightarrow \quad \nabla \tilde{f}(x) \in T_x\mathcal{M}_i}.$$

provided $\{M_i\}$ *is a Whitney stratification of* Q.

- For semi-algebraic Q, Whitney stratifications always exist!

Conclusion

- Semi-algebraic geometry is a powerful addition to the Variational Analysis toolkit.

- **Applications:** quadratic growth and error bounds, subgradient descent and the sweeping process, Sard theorem, and approximation on singular domains.

Thank you.

References

- **Generic minimizing behavior in semi-algebraic optimization**, D-Ioffe-Lewis, preprint `arXiv:1504.07694`, 2015.
- **Trajectory length of the tame sweeping process**, D-Daniilidis, preprint `arXiv:1412.8581`, 2015.
- **Quadratic growth and critical point stability of semi-algebraic functions**, D-Ioffe. To appear in *Math. Program.*, 2015.
- **Clarke subgradients for directionally Lipschitzian stratifiable functions**, *Math. Oper. Res.*, 40(2), 328-349, 2015.
- **Approximating functions on stratified sets**, D-Larsson, *Trans. Amer. Math. Soc.*, 367, 725-749, 2015.
- **Curves of descent**, D-Ioffe-Lewis, *SIAM J. Control and Optim.*, 53(1), 114-138, 2015.
- **Semi-algebraic functions have small subdifferentials**, D-Lewis, *Math. Program.*, 140(1): 5-29, 2013.

Available at www.math.washington.edu/~ddrusv