Generic sensitivity analysis for semi-algebraic optimization

Adrian Lewis
(with D. Drusvyatskiy and A. Ioffe)

ORIE Cornell

Terry Fest

Limoges, May 2015

Outline

- Strong regularity and Sard's Theorem
- Semi-algebraic functions and thin subdifferentials
- Identifiability and the active set philosophy
- Example: low-rank matrix optimization via the nuclear norm
- Generic metric regularity and alternating projections

Inversion and strong regularity

Problem: Given a **set-valued mapping** $\Phi\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^n$, find a **solution** x with **data** $y \in \underbrace{\Phi(x)}_{\text{easy to compute}}$. Equivalently, $x \in \underbrace{\Phi^{-1}(y)}_{\text{hard}}$.

Strong regularity (Robinson '80) then means

$$\operatorname{graph}\Phi \;=\; \operatorname{graph}(G^{-1}) \quad \text{around } (x, y)$$

for some single-valued Lipschitz G.

Crucial for sensitivity, algorithms... (Dontchev-Rockafellar '14)

Example (Banach, 1922) Mappings

$$\Phi \;=\; \text{identity} \;+\; \text{single-valued contraction}$$

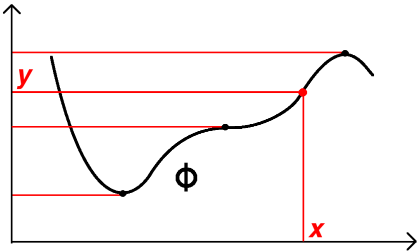
are strongly regular, and the iteration

$$x \leftarrow y + x - \Phi(x) \quad \text{converges to} \quad \Phi^{-1}(y).$$

Sard's Theorem (1942)

For **smooth** $\Phi\colon \mathbf{R}^n \to \mathbf{R}^n$, strongly regularity holds when $\nabla\Phi$ is invertible. (Inverse function theorem)

For **generic** y (almost all in Lebesgue measure), true at every $x \in \Phi^{-1}(y)$.



What if Φ is more general: nonsmooth or set-valued?

- Optimization: Φ a subdifferential.
- Variational inequalities: $\Phi =$ smooth map $+$ normal cone.

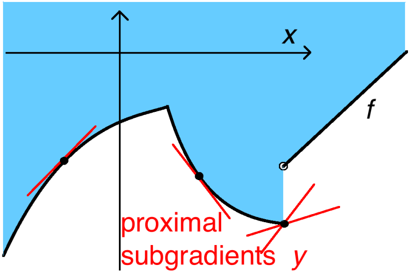
Structured: Saigal-Simon '73, Spingarn-Rockafellar '79, Alizadeh-Haeberly-Overton '97, Shapiro '97, Pataki-Tunçel '01.

Unstructured? Clearly Φ must have "*n*-dimensional" graph.

Subdifferentials and stationary points

Suppose $\Phi = \partial f$, for a function $f : \mathbf{R}^n \to \bar{\mathbf{R}}$, so $\Phi^{-1}(0)$ consists of stationary points.

As usual, $y \in \partial_P f(x)$ if $f(x+z) - f(x) \geq \langle y, z \rangle + O(|z|^2)$.



More stably, $y \in \partial f(x)$ means:

some $(x_r, y_r) \to (x, y)$ with $f(x_r) \to f(x)$ and $y_r \in \partial_P f(x_r)$.

In particular:

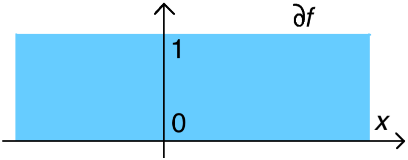
$$\partial f = \begin{cases} \nabla f & \text{if } f \text{ smooth} \\ \partial f & \text{if } f \text{ convex.} \end{cases}$$

Large subdifferentials

But many Lipschitz functions have subdifferentials with large graph.

Eg: Lipschitz $f : \mathbf{R} \to \mathbf{R}$ can have

$$\partial f(x) = [0, 1] \quad \text{for all } x.$$



(Benoist, Borwein-Girgensohn-Wang, 1998)

Subdifferentials of **convex** (or prox-regular) $f : \mathbf{R}^n \to \bar{\mathbf{R}}$ do have **thin** graphs:

$$\text{graph}\,\partial f \quad n\text{-dimensional}$$

as a Lipschitz manifold (Minty, 1962).

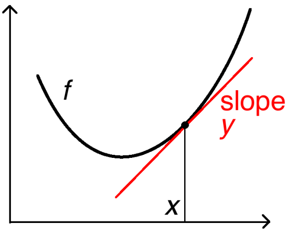
Regularity for convex minimization

For convex f,

$$(\partial f)^{-1}(y) \;=\; \text{argmin}\{f - \langle y, \cdot \rangle\}.$$

$(\partial f)^{-1}$ is **generically** single-valued and differentiable (Mignot, 1976)...

... **but not** Lipschitz, necessarily.



If $(f')^{-1}$ is the Lebesgue singular function, strong regularity of ∂f fails for **all** data y.

But what if f is more "concrete", or "tame" (Grothendieck)?

Semi-algebraic sets

Polynomial level sets in $\mathbf{R}^n$:

$$\{x : p(x) \leq 0\}.$$

Basic sets are finite intersections of these and their complements.

Finite unions of basic sets are called **semi-algebraic**.

A prevalent property, often easy to recognize,
since linear projection maps preserve it (Tarski-Seidenberg).

Semi-algebraic sets are finite unions of manifolds,
so have **dimension**.

We call n-dimensional subsets of $\mathbf{R}^n \times \mathbf{R}^n$ **thin**.

Generic regularity and stationarity

Following Sard... (Drusvyatskiy-Ioffe-L 2013–15)

Theorem Consider a semi-algebraic set-valued mapping $\Phi\colon \mathbf{R}^n \rightrightarrows \mathbf{R}^n$ with thin graph. For generic data y, strong regularity holds at every solution $x \in \Phi^{-1}(y)$.

Theorem The subdifferential of a semi-algebraic function has thin graph.

So, finding stationary points for any generically perturbed semi-algebraic function is well behaved.

For classical nonlinear programs, much more holds (Spingarn-Rockafellar '79): second-order sufficiency...

Can we extend?

Identifiability and "active set" philosophy

Many algorithms for minimizing functions f (maybe nonsmooth, high-dimensional, nonconvex) generate sequences satisfying

$$x_k \to \bar{x} \qquad y_k \to 0$$
$$f(x_k) \to f(\bar{x}) \qquad y_k \in \partial f(x_k)$$

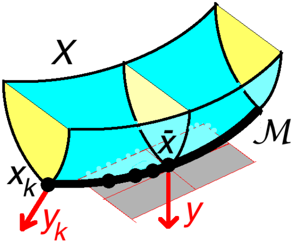
Example. Proximal point: $\rho(x_k - x_{k+1}) \in \partial f(x_{k+1})$.

A manifold $\mathcal{M}$ around $\bar{x}$ is **identifiable** (Wright 1993) when

- $f|_{\mathcal{M}}$ is $C^{(2)}$-smooth
- every such sequence (x_k) eventually lies in $\mathcal{M}$.

Then minimizing f reduces to minimizing the low-dimensional smooth function $f|_{\mathcal{M}}$.

Example $\quad f = \delta_X - \langle y, \cdot \rangle$:



Example: matrix nuclear norm regularization

Rank-constrained optimization (Candès-Recht, -Tao '09) relaxes to

$$\min_{X \in \mathbf{R}^{m \times n}} \left\{ g(X) + \|X\|_* \right\}$$

for smooth convex g and **nuclear norm** $\|\cdot\|_* = \sum_i \sigma_i$.

Optimal $\bar{X}$ and $\nabla g(\bar{X})$ have simultaneous SVD, singular values

$$\sigma_i\big(\nabla g(\bar{X})\big) \;\leq\; 1.$$

Equality holds if $\sigma_i(\bar{X}) > 0$. **Generically**, the converse holds, and $g + \|\cdot\|_*$ shows local smooth quadratic growth on the manifold

$$\{X : \operatorname{rank} X = \operatorname{rank} \bar{X}\}.$$

Huge examples (`Netflix, Yahoo-Music...`), $m \sim 10^6$, $n \sim 10^5$ but **low-rank** $\bar{X}$: solvable via smooth reduction (Hsieh-Olson '14).

Generic identifiability

Bolte-Daniilidis-L '11 (convex case) and Drusvyatskiy-Ioffe-L '14.

Consider any semi-algebraic closed function f_0. A generic linear perturbation $f = f_0 - \langle y, \cdot \rangle$ has a finite set of stationary points $x \in (\partial f)^{-1}(0)$, each satisfying:

- f is **prox-regular** at x for 0
- $0 \in \operatorname{ri} \partial_P f(x)$ (**strict complementarity**)
- f has the **identifiable manifold**

$$\mathcal{M} \;=\; \{z \text{ near } x : 0 \text{ near } \partial f(z)\}$$

- ∂f is **strongly regular** at x for 0
- **2nd-order sufficiency**... $f|_{\mathcal{M}}$ grows quadratically around x.

Metric regularity, transversality, and alternating projections

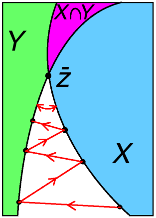
Strong regularity strengthens **metric regularity**:

$$(x, y) \mapsto \frac{d\big(x, \Phi^{-1}(y)\big)}{d\big(y, \Phi(x)\big)} \quad \text{locally bounded.}$$

Theorem (Ioffe '07). Any semi-algebraic closed Φ is metrically regular for generic data y at all solutions $x \in \Phi^{-1}(y)$.

Example. Given semi-algebraic closed sets $X, Y \subset \mathbf{R}^n$, under a generic perturbation w, the intersection of X and $Y - w$ is everywhere transversal.

Transversality (alone!) implies that alternating projections (von Neumann '33) converges linearly (Drusvyatskiy-Ioffe-L '13).



Summary

- Semi-algebraic generalized equations with thin graphs are strongly regular for generic data.
- Example: stationary points of semi-algebraic functions.
- Identifiable manifolds exist generically in semi-algebraic optimization, and the 2nd-order sufficient conditions hold.
- Generic transversality and alternating projections.