Nonsmooth Implicit Differentiation for Machine Learning

Jérôme Bolte, Tam Le, Edouard Pauwels, and Antonio Silveti-Falls




Neurips Poster (2021)

Outline

1. Smooth implicit function theorem.
2. Nonsmooth implicit function theorem of Clarke.
3. Path differentiable nonsmooth implicit function theorem (with calculus).
4. Applications
5. What can go wrong?

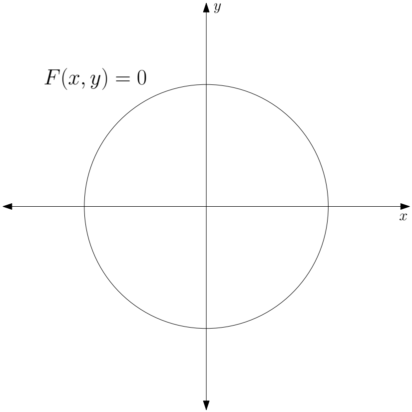
Implicit Functions

Consider the smooth function

$$F(x, y) = x^2 + y^2 - 1$$

and the equation

$$F(x, y) = 0.$$



Implicit Functions

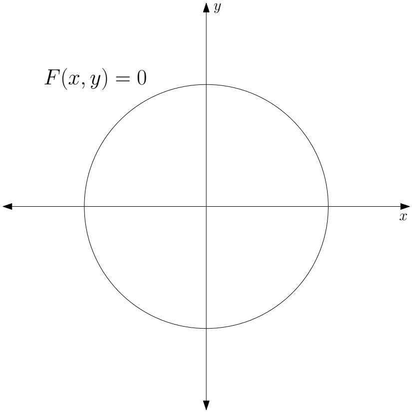
Consider the smooth function

$$F(x, y) = x^2 + y^2 - 1$$

and the equation

$$F(x, y) = 0.$$

- Can we find a function $y = G(x)$ so that $F(x, G(x)) = 0$?



Implicit Functions

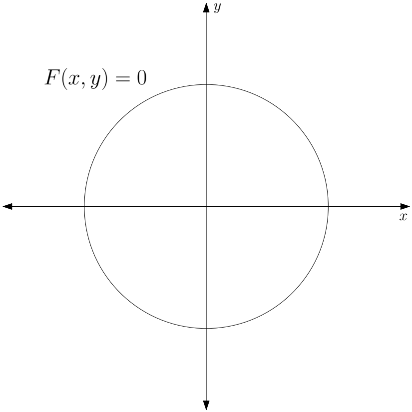
Consider the smooth function

$$F(x, y) = x^2 + y^2 - 1$$

and the equation

$$F(x, y) = 0.$$

- Can we find a function $y = G(x)$ so that $F(x, G(x)) = 0$?
- Can we compute the gradient of G?



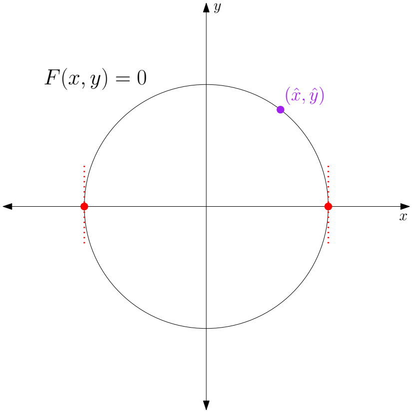
Implicit Functions

Consider the smooth function

$$F(x, y) = x^2 + y^2 - 1$$

and the equation

$$F(x, y) = 0.$$



- Can we find a function $y = G(x)$ so that $F(x, G(x)) = 0$?
- Can we compute the gradient of G?

Consider the smooth function

$$F(x, y) = x^2 + y^2 - 1$$

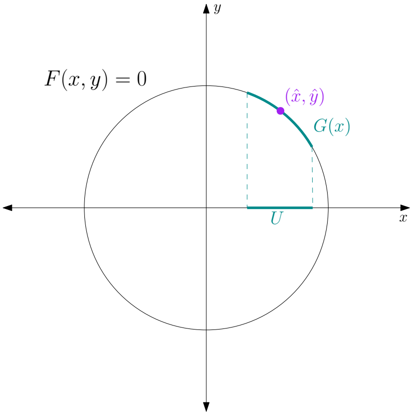
and the equation

$$F(x, y) = 0.$$

- Can we find a function $y = G(x)$ so that $F(x, G(x)) = 0$?

  Existence

- Can we compute the gradient of G?



Implicit Functions

Consider the smooth function

$$F(x, y) = x^2 + y^2 - 1$$

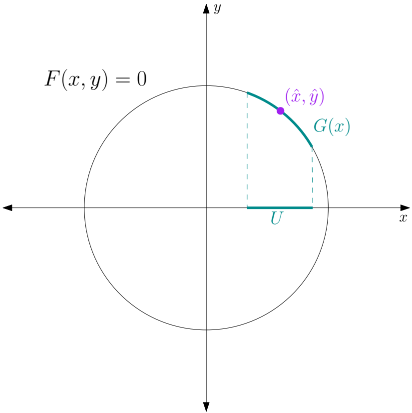
and the equation

$$F(x, y) = 0.$$

- Can we find a function $y = G(x)$ so that $F(x, G(x)) = 0$?

  Existence

- Can we compute the gradient of G?



Locally Lipschitz Functions and the Clarke Subdifferential

Definition (Locally Lipschitz-continuous)

A function $F : \mathbb{R}^n \to \mathbb{R}^m$ is locally Lipschitz-continuous if, $\forall x \in \mathbb{R}^n$, $\exists$ a neighborhood $U \subset \mathbb{R}^n$ of x and $c > 0$ such that, $\forall y, z \in U$,

$$\|F(z) - F(y)\| \leq c \|z - y\| .$$

Locally Lipschitz Functions and the Clarke Subdifferential

Definition (Locally Lipschitz-continuous)

A function $F : \mathbb{R}^n \to \mathbb{R}^m$ is locally Lipschitz-continuous if, $\forall x \in \mathbb{R}^n$, $\exists$ a neighborhood $U \subset \mathbb{R}^n$ of x and $c > 0$ such that, $\forall y, z \in U$,

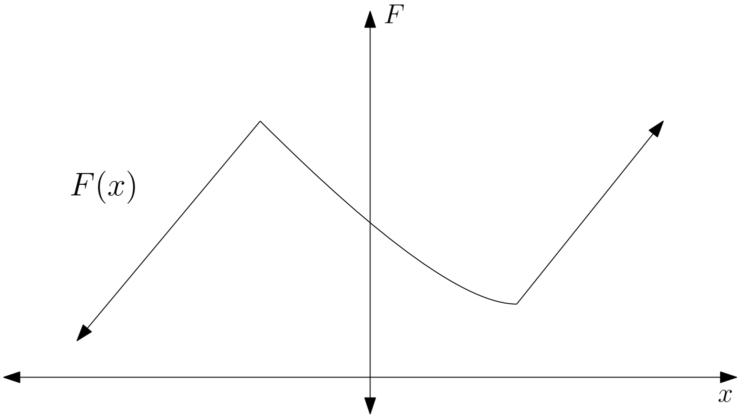
$$\|F(z) - F(y)\| \leq c \|z - y\| .$$

Definition (Clarke subdifferential (1983))

Given a locally Lipschitz function $F : \mathbb{R}^n \to \mathbb{R}^m$, the Clarke subdifferential at a point $x \in \mathbb{R}^n$ is

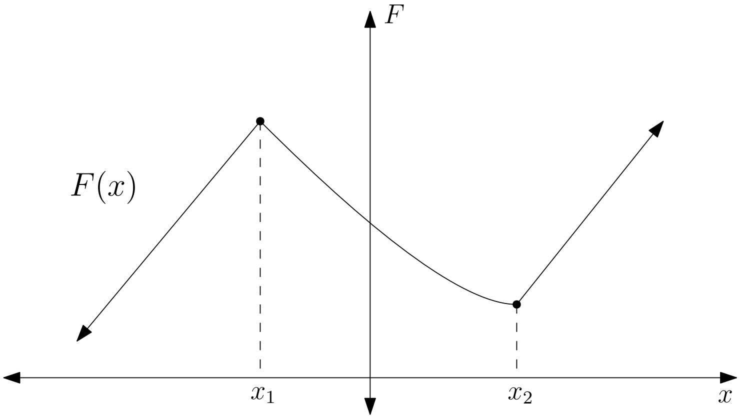
$$\partial^c F(x) = \mathrm{conv}\left(\left\{\lim_{k\to\infty} J_F(x_k) : x_k \in \mathrm{diff}_F \text{ and } x_k \to x\right\}\right).$$

Locally Lipschitz Functions and the Clarke Subdifferential



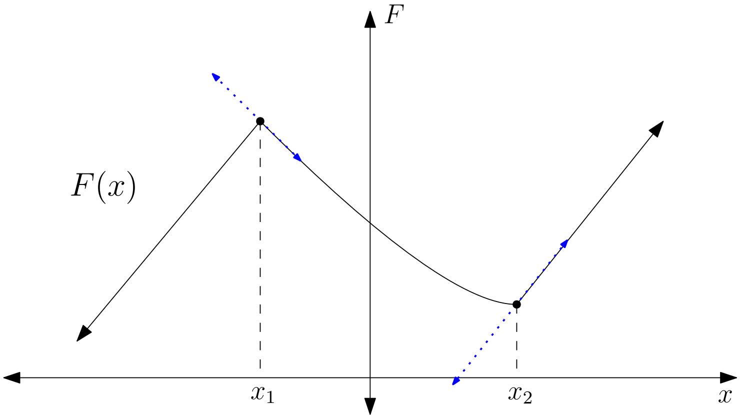
$$\partial^c F(x) = \operatorname{conv}\left(\left\{\lim_{k\to\infty} J_F(x_k) : x_k \in \mathrm{diff}_F \text{ and } x_k \to x\right\}\right)$$

Locally Lipschitz Functions and the Clarke Subdifferential



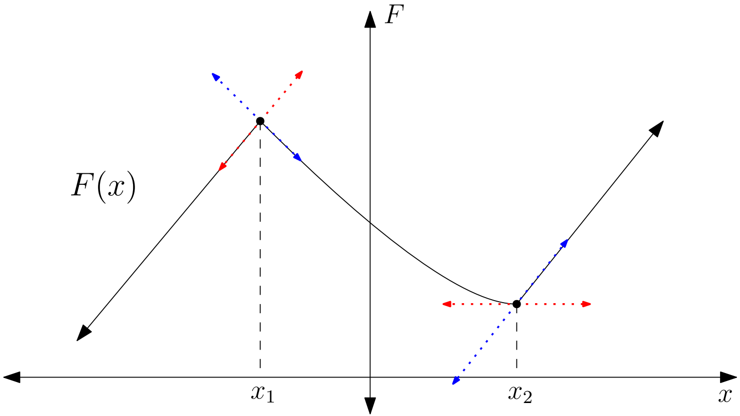
$$\partial^c F(x) = \mathrm{conv}\left(\left\{\lim_{k\to\infty} J_F(x_k) : x_k \in \mathrm{diff}_F \text{ and } x_k \to x\right\}\right)$$

Locally Lipschitz Functions and the Clarke Subdifferential



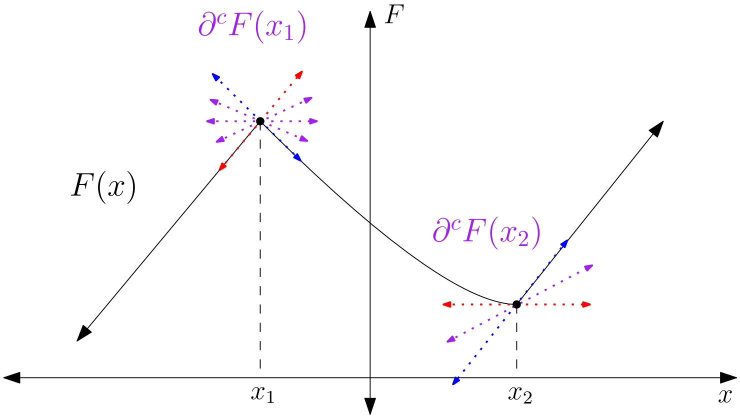
$$\partial^c F(x) = \mathrm{conv}\left(\left\{\lim_{k\to\infty} J_F(x_k) : x_k \in \mathrm{diff}_F \text{ and } x_k \to x\right\}\right)$$

Locally Lipschitz Functions and the Clarke Subdifferential



$$\partial^c F(x) = \operatorname{conv}\left(\left\{\lim_{k\to\infty} J_F(x_k) : x_k \in \operatorname{diff}_F \text{ and } x_k \to x\right\}\right)$$

Locally Lipschitz Functions and the Clarke Subdifferential



$$\partial^c F(x) = \mathrm{conv}\left(\left\{\lim_{k\to\infty} J_F(x_k) : x_k \in \mathrm{diff}_F \text{ and } x_k \to x\right\}\right)$$

Nonsmooth Implicit Function Theorem of Clarke

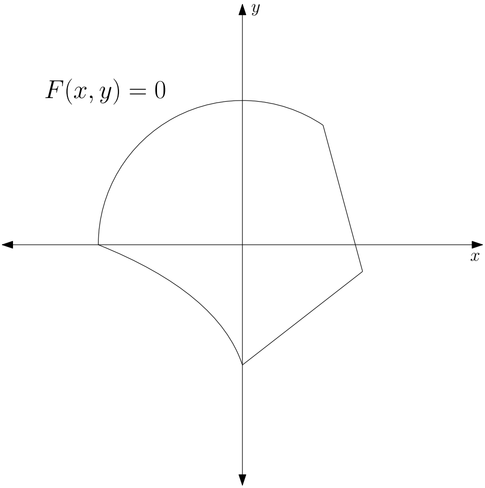
Theorem (Clarke 1983)

Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ *be locally Lipschitz and* $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ *such that*

$$F(\hat{x}, \hat{y}) = 0.$$

If, $\forall [A\ B] \in \partial^c F(\hat{x}, \hat{y})$, B *is invertible, then* $\exists U \subset \mathbb{R}^n$ *a neighborhood of* $\hat{x}$ *and a locally Lipschitz function* $G(x)$ *so that*

$$F(x, G(x)) = 0 \qquad \forall x \in U.$$



Nonsmooth Implicit Function Theorem of Clarke

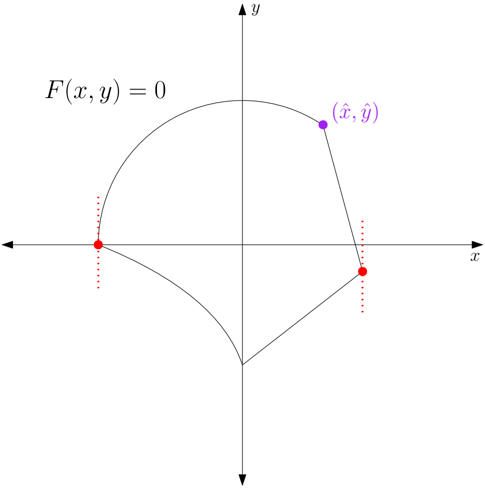
Theorem (Clarke 1983)

Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be locally Lipschitz and $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ such that

$$F(\hat{x}, \hat{y}) = 0.$$

If, $\forall [A\ B] \in \partial^c F(\hat{x}, \hat{y})$, B is invertible, then $\exists U \subset \mathbb{R}^n$ a neighborhood of $\hat{x}$ and a locally Lipschitz function $G(x)$ so that

$$F(x, G(x)) = 0 \qquad \forall x \in U.$$



Nonsmooth Implicit Function Theorem of Clarke

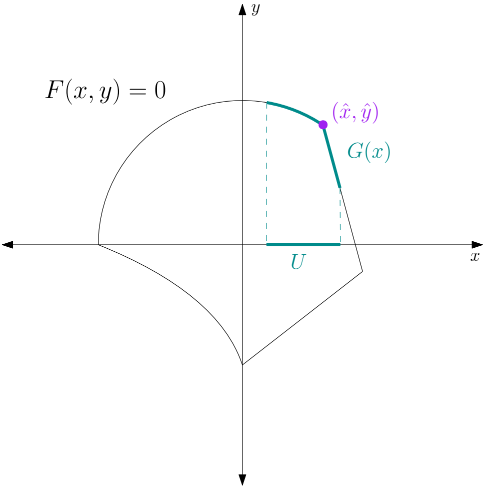
Theorem (Clarke 1983)

Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be locally Lipschitz and $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ such that

$$F(\hat{x}, \hat{y}) = 0.$$

If, $\forall [A\ B] \in \partial^c F(\hat{x}, \hat{y})$, B is invertible, then $\exists U \subset \mathbb{R}^n$ a neighborhood of $\hat{x}$ and a locally Lipschitz function $G(x)$ so that

$$F(x, G(x)) = 0 \qquad \forall x \in U.$$



Nonsmooth Implicit Function Theorem of Clarke

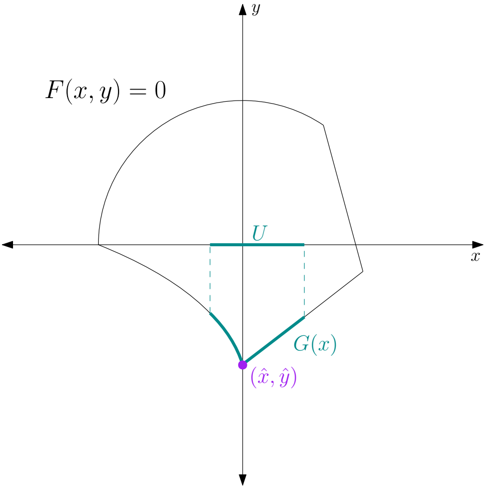
Theorem (Clarke 1983)

Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be locally Lipschitz and $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ such that

$$F(\hat{x}, \hat{y}) = 0.$$

If, $\forall [A\ B] \in \partial^c F(\hat{x}, \hat{y})$, B is invertible, then $\exists U \subset \mathbb{R}^n$ a neighborhood of $\hat{x}$ and a locally Lipschitz function $G(x)$ so that

$$F(x, G(x)) = 0 \qquad \forall x \in U.$$



Lack of Calculus

Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be locally Lipschitz and $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ such that Clarke's IFT holds with implicit function $G(x)$.

Lack of Calculus

Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be locally Lipschitz and $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ such that Clarke's IFT holds with implicit function $G(x)$.

Recall from smooth IFT: $J_G(x) = -B^{-1}A \quad [A\ B] = J_F(x, G(x))$.

Lack of Calculus

Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be locally Lipschitz and $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ such that Clarke's IFT holds with implicit function $G(x)$.

Recall from smooth IFT: $J_G(x) = -B^{-1}A \quad [A\ B] = J_F(x, G(x))$.

Question

Does it hold

$$\left\{-B^{-1}A : [A\ B] \in \partial^c F(\hat{x}, \hat{y})\right\} = \partial^c G(x) \quad ?$$

Lack of Calculus

Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be locally Lipschitz and $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ such that Clarke's IFT holds with implicit function $G(x)$.

Recall from smooth IFT: $J_G(x) = -B^{-1}A \quad [A\ B] = J_F(x, G(x))$.

Question

Does it hold

$$\left\{-B^{-1}A : [A\ B] \in \partial^c F(\hat{x}, \hat{y})\right\} = \partial^c G(x) \quad ?$$

No - need something beyond ∂^c.

Conservative Fields

Definition (Conservative field (Bolte-Pauwels 2019))

A set valued mapping $D_F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a conservative field (or conservative Jacobian) for $F : \mathbb{R}^n \to \mathbb{R}$ locally Lipschitz if:

Conservative Fields

Definition (Conservative field (Bolte-Pauwels 2019))

A set valued mapping $D_F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a conservative field (or conservative Jacobian) for $F : \mathbb{R}^n \to \mathbb{R}$ locally Lipschitz if:

1. For all $x \in \mathbb{R}^n$, $D_F(x)$ is nonempty (ideally convex!).

Conservative Fields

Definition (Conservative field (Bolte-Pauwels 2019))

A set valued mapping $D_F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a conservative field (or conservative Jacobian) for $F : \mathbb{R}^n \to \mathbb{R}$ locally Lipschitz if:

1. For all $x \in \mathbb{R}^n$, $D_F(x)$ is nonempty (ideally convex!).
2. D_F has a closed graph and is locally bounded.

Conservative Fields

Definition (Conservative field (Bolte-Pauwels 2019))

A set valued mapping $D_F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a conservative field (or conservative Jacobian) for $F : \mathbb{R}^n \to \mathbb{R}$ locally Lipschitz if:

1. For all $x \in \mathbb{R}^n$, $D_F(x)$ is nonempty (ideally convex!).
2. D_F has a closed graph and is locally bounded.
3. For any absolutely continuous curve $\gamma : [0,1] \to \mathbb{R}^n$,

$$\frac{d}{dt} F(\gamma(t)) = \langle u, \dot{\gamma}(t) \rangle \qquad \forall u \in D_F(\gamma(t))$$

   for almost all $t \in [0,1]$.

We call F path differentiable.

Path Differentiable Nonsmooth Implicit Function Theorem

Theorem

Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ *be path differentiable and* $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ *be such that*

$$F(\hat{x}, \hat{y}) = 0.$$

Path Differentiable Nonsmooth Implicit Function Theorem

Theorem

Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ *be path differentiable and* $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ *be such that*

$$F(\hat{x}, \hat{y}) = 0.$$

Assume $D_F(\hat{x}, \hat{y})$ *is convex and* $\forall [A \ B] \in D_F(\hat{x}, \hat{y})$, B *is invertible.*

Path Differentiable Nonsmooth Implicit Function Theorem

Theorem

Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be path differentiable and $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ be such that

$$F(\hat{x}, \hat{y}) = 0.$$

Assume $D_F(\hat{x}, \hat{y})$ is convex and $\forall [A \ B] \in D_F(\hat{x}, \hat{y})$, B is invertible.

Path Differentiable Nonsmooth Implicit Function Theorem

Theorem

Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be path differentiable and $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ be such that

$$F(\hat{x}, \hat{y}) = 0.$$

Assume $D_F(\hat{x}, \hat{y})$ is convex and $\forall [A\ B] \in D_F(\hat{x}, \hat{y})$, B is invertible.
Then $\exists U \subset \mathbb{R}^n$ a neighborhood of $\hat{x}$ and a path differentiable function G such that

$$\forall x \in U \qquad F(x, G(x)) = 0.$$

Path Differentiable Nonsmooth Implicit Function Theorem

Theorem

Let $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ be path differentiable and $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^m$ be such that

$$F(\hat{x}, \hat{y}) = 0.$$

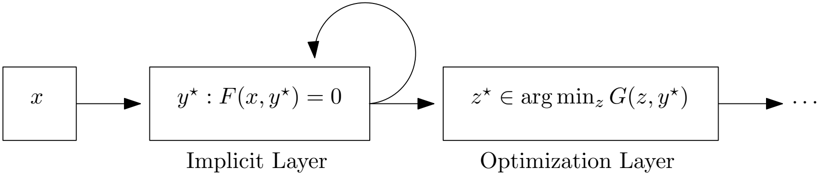
Assume $D_F(\hat{x}, \hat{y})$ is convex and $\forall [A\ B] \in D_F(\hat{x}, \hat{y})$, B is invertible.
Then $\exists U \subset \mathbb{R}^n$ a neighborhood of $\hat{x}$ and a path differentiable function G such that

$$\forall x \in U \qquad F(x, G(x)) = 0.$$

The conservative Jacobian of G is given by

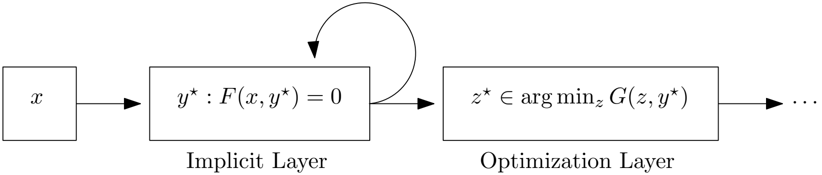
$$D_G(x) = \left\{ -B^{-1}A : [A\ B] \in D_F(x, G(x)) \right\}.$$

Deep Learning with Implicit Layers



- Deep equilibrium networks [Shaojie Bai, J. Zico Kolter, Vladlen Koltun 2019], Monotone deep equlibirium networks [Ezra Winston, J. Zico Kolter 2020].
- Optimization layers (OptNET) [Brandon Amos, J. Zico Kolter 2017].

Deep Learning with Implicit Layers



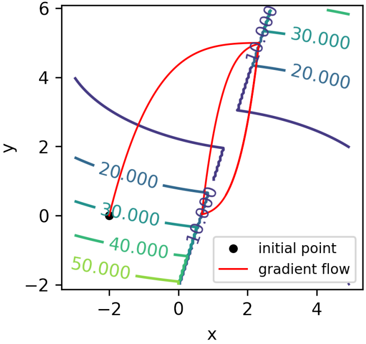
- Deep equilibrium networks [Shaojie Bai, J. Zico Kolter, Vladlen Koltun 2019], Monotone deep equlibirium networks [Ezra Winston, J. Zico Kolter 2020].
- Optimization layers (OptNET) [Brandon Amos, J. Zico Kolter 2017].
- Convergence guarantees for training.

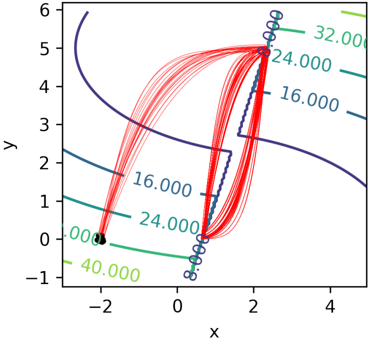
Pathological Examples - Cycles

Gradient descent type algorithm (using backprop) applied to:

$$\min_{x,y,s} \quad \ell(x,y,s) \stackrel{\text{def}}{=} (x - s_1)^2 + 4(y - s_2)^2$$

$$\text{s.t.} \quad s \in \arg\max\{(a+b)(-2x+y+2) : a \in [0,3], b \in [0,5]\}.$$





Pathological Examples - Lorenz Attractor

For $u \in \mathbb{R}^3$, define $L(u) \stackrel{\text{def}}{=} \left(10(y-x), x(28-z)-y, xy-\frac{8}{3}z\right)$.

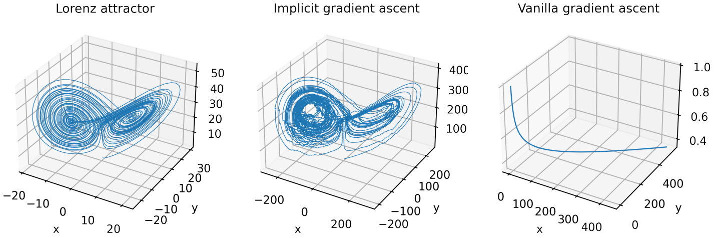
Implicit formulation

$$\max_{u \in \mathbb{R}^3} u^T z \quad \text{s.t.}$$

$$z \in \operatorname*{argmin}_{s \in \mathbb{R}^3} \|s - L(u)\|^4$$

Explicit (vanilla) formulation

$$\max_{u \in \mathbb{R}^3} u^T L(u)$$



Thanks for Listening

Thanks for listening.

Full paper available on arxiv: https://arxiv.org/abs/ 2106.04350

"Nonsmooth Implicit Differentiation for Machine Learning and Optimization" - Jérôme Bolte, Ngoc Tâm Lê, Edouard Pauwels, Antonio Silveti-Falls