Some applications of implicit function theorems from variational analysis

Tim Hoheisel (McGill)

Joint work with Aaron Berk (McGill), Simone Brugiapaglia (Concordia), Michael Friedlander (UBC), Ariel Goodwin (Cornell)

Foundations of Computational Mathematics (FoCM) 2023
Paris, France, June 17, 2023

Motivation

Consider the optimization problem

$$\min_{x \in \mathbb{R}^n} h(p, x) + \varphi(x) \tag{1}$$

where

- $h : \mathbb{R}^p \times \mathbb{R}^n \to \mathbb{R}$ (locally) smooth and convex in x;
- $\varphi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ closed, proper, convex.

$$S(p) := \operatorname*{argmin}_{x \in \mathbb{R}^n} \{h(p, x) + \varphi(x)\} \quad \text{(solution map)}.$$

References: Bonnans/Shapiro (general NLP), Bolte et al. (monotone operators), Vaiter et al. (regularized LLS).

Examples

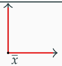
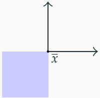
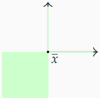
- (prox operator) $p := (\bar{x}, \lambda),\ h(p, x) := \frac{1}{2\lambda}\|x - \bar{x}\|^2$: $S(\bar{x}, \lambda) = P_\lambda \varphi(\bar{x})$.
- (unconstrained LASSO) $p := (A, b, \lambda),\ h(p, x) = \frac{1}{2\lambda}\|Ax - b\|^2,\ \varphi = \|\cdot\|_1$.

By convexity

$$S(p) = \{x \in \mathbb{R}^n \mid 0 \in \nabla_x h(x, p) + \partial\varphi(x)\}.$$

Tailor-made for the implicit function theorems of variational analysis based on graphical differentiation.

Variational analysis: normal cones and graphical differentiation

Name	Definition	Properties	Example
tangent cone	$T_A(\bar{x}) := \operatorname{Lim\,sup}_{t \downarrow 0} \frac{A - \bar{x}}{t}$	closed	
regular normal cone	$\hat{N}_A(\bar{x}) := T_A(\bar{x})^\circ$	closed, convex	
limiting normal cone	$N_A(\bar{x}) := \operatorname{Lim\,sup}_{x \to \bar{x}} \hat{N}_A(x)$	closed	

$S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, $(\bar{x}, \bar{y}) \in \operatorname{gph} S := \{(x, y) \mid y \in S(x)\}$.

- Graphical derivative (Aubin '81, Benko '21): $DS(\bar{x}|\bar{y}) : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ via

$$v \in DS(\bar{x}|\bar{y})(u) \;\; :\Longleftrightarrow \;\; (u, v) \in T_{\operatorname{gph} S}(\bar{x}, \bar{y}).$$

- Coderivative (Mordukhovich '80, Ioffe '84): $D^*S(\bar{x}|\bar{y}) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ via

$$v \in D^*S(\bar{x}|\bar{y})(u) \;\; :\Longleftrightarrow \;\; (v, -u) \in N_{\operatorname{gph} S}(\bar{x}, \bar{y}).$$

Variational analysis: proto-differentiability

Observe that graphical derivative of $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ at $(\bar{x}, \bar{u}) \in \operatorname{gph} S$ is (by definition)

$$DS(\bar{x} \mid \bar{u})(\bar{w}) = \operatorname*{Lim\,sup}_{t \downarrow 0,\, w \to \bar{w}} \frac{S(\bar{x} + tw) - \bar{u}}{t} \quad \forall\, \bar{w} \in \mathbb{R}^n. \tag{2}$$

Definition (Proto-differentiability (Rockafellar '89))

We call S is *proto-differentiable* at $(\bar{x}, \bar{u}) \in \operatorname{gph} S$ if the following hold:

$$\forall \bar{z} \in DS(\bar{x} \mid \bar{u})(\bar{w}),\ \{t_k\} \downarrow 0\ \exists \{w_k\} \to \bar{w},\ \{z_k\} \to \bar{z} :\ z_k \in \frac{S(\bar{x} + t_k w_k) - \bar{u}}{t_k}\ \forall k \in \mathbb{N}.$$

- Relates to semidifferentiability (Penot) which will yield directional differentiability for our purposes.
- Graphically regularity implies proto-differentiability.
- ∂f is proto-differentiable at $(\bar{x}, \bar{u})$, e.g., if $f = g \circ F$ is *fully amenable*, i.e., g PLQ and $F \in C^2$ such that

$$\ker F'(\bar{x})^* \cap N_{\operatorname{dom} g}(F(\bar{x})) = \{0\} \quad \text{(basic constraint qualification).}$$

Variational analysis: directional normal cone and semismoothness*

Directional normal cone of A at $\bar{x}$ in direction $\bar{u}$:

$$N_A(\bar{x};\bar{u}) := \operatorname*{Lim\,sup}_{u\to\bar{u},\ t\downarrow 0} \hat{N}_A(\bar{x}+tu).$$

- $N(\bar{x};\bar{u}) = \emptyset$ if $\bar{u} \notin T_A(\bar{x})$;
- $N(\bar{x};\bar{u}) \subset N_A(\bar{x})$ for all $u \in \mathbb{R}^n$.

Semismoothness* (Gfrerer et al.):

i) $A \subset \mathbb{R}^n$ *semismooth** at $\bar{x} \in A$ $\quad :\Longleftrightarrow \quad \langle x^*, u\rangle = 0 \quad \forall u \in \mathbb{R}^n,\ x^* \in N_A(\bar{x};u)$.

ii) $S : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *semismooth** at $(\bar{x},\bar{y}) \in \operatorname{gph} S$ $\quad :\Longleftrightarrow \quad$ $\operatorname{gph} S$ semismooth* at $(\bar{x},\bar{y})$.

(Gfrerer and Outrata '19): For $F : D \subset \mathbb{R}^n \to \mathbb{R}^m$ locally Lipschitz at $\bar{x} \in \operatorname{int} D$, the following are equivalent:

- F semismooth (in the sense of Qi and Sun) at $\bar{x}$.
- F semismooth* and directionally differentiable at $\bar{x}$.

Let $f:\mathbb{R}^d\times\mathbb{R}^n\to\mathbb{R}^n$ be continuously differentiable at $(\bar p,\bar x)$ such that $f(\bar p,\cdot)$ is monotone, let $F:\mathbb{R}^n\rightrightarrows\mathbb{R}^n$ be monotone at $(\bar x,-f(\bar p,\bar x))$. Define $S:\mathbb{R}^d\rightrightarrows\mathbb{R}^n$ by

$$S(p)=\{x\in\mathbb{R}^n \mid 0\in f(p,x)+F(x)\},\quad \forall p\in\mathbb{R}^d.$$

The following hold if $(\bar p,\bar x)\in\operatorname{gph} S$ is such that

$$\ker\left(D_xf(\bar p,\bar x)^*+D^*F(\bar x|-f(\bar p,\bar x)\right)=\{0\}\quad \text{(Mordukhovich criterion).}$$

(a) S is locally Lipschitz at $\bar p$ with modulus

$$L\le\limsup_{p\to\bar p}\max_{\|q\|\le 1}\inf_{w\in DS(p)(q)}\|w\|.$$

(b) If F is *proto-differentiable* at $(\bar x,-f(\bar p,\bar x))$, S is directionally differentiable at $\bar p$ with locally Lipschitz directional derivative (for $G(p,x):=f(p,x)+F(x)$) given by

$$S'(\bar p;q)=\{w\in\mathbb{R}^n \mid 0\in DG(\bar p,\bar x|0)(q,w)\}\quad\forall q\in\mathbb{R}^d.$$

(b) If F is semismooth* and the following implication is satisfied:

$$\left.\begin{array}{rcl}-(v,w)&\in&N_{\operatorname{gph}F}(\bar x,-f(\bar p,\bar x)),\\ 0&=&D_pf(\bar p,\bar x)^*w,\\ v&=&D_xf(\bar p,\bar x)^*w\end{array}\right\}\implies (v,w)=(0,0),$$

then S is semismooth at $\bar p$.

(c) If $S'(\bar p;\cdot)$ is linear, then S is differentiable at $\bar p$.

Application: unconstrained LASSO (constraint qualifications)

The unconstrained LASSO[1] for $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, \lambda > 0$ reads

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|Ax - b\|^2 + \lambda \|x\|_1. \tag{3}$$

For a solution $\bar{x}$ of (**??**) define:

- $I := \{i \in \{1, \ldots, n\} \mid \bar{x}_i \neq 0\}$ (support);
- $J := \left\{i \in \{1, \ldots, n\} \;\middle|\; |A_i^T(b - A\bar{x})| = \lambda\right\}$ (equicorrelation set).

Note: $\boxed{I \subset J.}$

<u>Qualification conditions</u>

- (Intermediate) $\ker A_J = \{0\}$;
- (Strong) $I = J$ and $\ker A_I = \{0\}$.

(Strong) $\implies$ (Intermediate) $\implies$ $\bar{x}$ is unique solution of (**??**)

[1] Santosa and Symes (1986), Tibshirani (1996)

Application: unconstrained LASSO (stability)

Apply the main theorem with $f(b,\lambda,x) := \frac{1}{\lambda}A^T(Ax-b), \quad F := \partial\|\cdot\|_1$ such that

$$S(b,\lambda) = \{x \mid 0 \in f(b,\lambda,x) + F(x)\} = \operatorname*{argmin}_{x\in\mathbb{R}^n}\left\{\frac{1}{2}\|Ax-b\|^2 + \lambda\|x\|_1\right\} \quad (\lambda > 0).$$

For $(\bar{b},\bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}_{++}$ let $\bar{x} \in S(\bar{b},\bar{\lambda})$. Then:

(a) If the intermediate condition holds, S is semismooth at $(\bar{b},\bar{\lambda})$ with Lipschitz modulus

$$L \le \frac{1}{\sigma_{\min}(A_J)^2}\left(\sigma_{\max}(A_J) + \left\|\frac{A_J^T(A\bar{x}-\bar{b})}{\bar{\lambda}}\right\|\right).$$

Moreover, the directional derivative $S'((\bar{b},\bar{\lambda});(\cdot,\cdot)) : \mathbb{R}^m \times \mathbb{R} \to \mathbb{R}^n$ is locally Lipschitz and given as follows: for $(q,\alpha) \in \mathbb{R}^m \times \mathbb{R}$ there exists an index set $K = K(q,\alpha)$ with $I \subseteq K \subseteq J$ such that

$$S'((\bar{b},\bar{\lambda});(q,\alpha)) = L_K\left((A_K^TA_K)^{-1}A_K^T\left(q + \frac{\alpha}{\bar{\lambda}}(A\bar{x}-\bar{b})\right), 0\right).$$

(b) If the strong assumptions holds, S is continuously differentiable at $(\bar{b},\bar{\lambda})$ with

$$DS(\bar{b},\bar{\lambda})(q,\alpha) = L_I\left((A_I^TA_I)^{-1}A_I^T\left(q + \frac{\alpha}{\bar{\lambda}}(A\bar{x}-\bar{b})\right), 0\right), \quad \forall (q,\alpha) \in \mathbb{R}^m \times \mathbb{R}.$$

In particular, S is locally Lipschitz with modulus given above with $I = J$.

Application: unconstrained LASSO (Mordukhovich criterion verified)

Let $\bar{x}$ solve the unconstrained LASSO, i.e.

$$0 \in \underbrace{\frac{1}{\lambda} A^T(A\bar{x} - b)}_{=f(b,\lambda,\bar{x})} + \underbrace{\partial\|\cdot\|_1(\bar{x})}_{F(\bar{x})}.$$

Assume that the intermediate assumption holds, i.e. (with $\bar{u} := \frac{1}{\lambda} A^T(b - A\bar{x}) \in \partial\|\cdot\|_1(\bar{x})$)

$$\ker A_J = \{0\} \quad \text{for} \quad J := \{\, i \in [1:n] \mid |\bar{u}_i| = 1 \,\}. \tag{4}$$

Let $0 \in D_x f(b, \lambda, \bar{x})^* w + D^* F(\bar{x}|\bar{u})(w) = \frac{1}{\lambda} A^T A w + D^*(\partial\|\cdot\|_1)(\bar{x}|\bar{u})(w)$, i.e.

$$-\frac{1}{\lambda} A^T A w \in D^*(\partial\|\cdot\|_1)(\bar{x}|\bar{u})(w). \tag{5}$$

By ‘positive semidefiniteness’ of $D^*(\partial\|\cdot\|_1)(\bar{x}|\bar{u})$ it follows that

$$w \in \ker A. \tag{6}$$

Therefore (**??**) implies

$$\begin{aligned} 0 \in D^*(\overbrace{\partial\|\cdot\|_1}^{=N_{\mathbb{B}_\infty}^{-1}})(\bar{x}|\bar{u})(w) \quad &\iff \quad w \in D^* N_{\mathbb{B}_\infty}(\bar{u}|\bar{x})(0) = \operatorname{span}\{e_i \mid i \in J\} \\ &\implies \quad w_{J^C} = 0 \\ &\overset{(\mathbf{??})}{\implies} \quad w \in \ker A_J \\ &\overset{(\mathbf{??})}{\implies} \quad w = 0. \end{aligned}$$

Application: unconstrained LASSO (tuning parameter sensitivity)

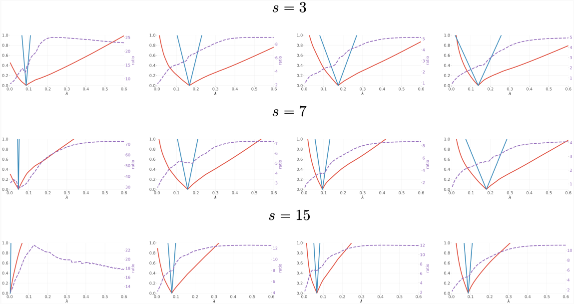
Suppose

$$b = Ax_0 + e :$$

- $n = 200$,
- $A_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, 1/m)$,
- $e_i \overset{\text{iid}}{\sim} \mathcal{N}(0, 0.01)$ and
- x_0 s-sparse: $(x_0)_j \overset{\text{iid}}{\sim} \mathcal{N}(m, m)$ $(j \in I)$.

- $x(\lambda) := \underset{x}{\operatorname{argmin}} \left\{ \frac{\|Ax - b\|^2}{2} + \lambda \|x\|_1 \right\}$,
- $\lambda^* := \inf \underset{\lambda > 0}{\operatorname{argmin}} \|x(\lambda) - x_0\|$,
- $\bar{x} := x(\lambda^*)$.

Under the strong assumption at $\bar{x}$, $x(\cdot)$ is locally Lipschitz with $L := \frac{\sqrt{|I|}}{\sigma_{\min}(A_I)^2}$.



$m = 50$ $m = 100$

Figure 1: $\|x(\lambda) - \bar{x}\|$, $L|\lambda - \lambda^*|$, $L|\lambda - \lambda^*|$ ...

Application: the proximal operator

For $f \in \Gamma_0 := \{f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\} \mid f$ lsc, proper, convex $\}$, the proximal operator is

$$P_\lambda f(x) = \underset{u \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \varphi(x) + \frac{1}{2\lambda} \|x - u\|^2 \right\} \quad \forall x \in \mathbb{R}^n,\ \lambda > 0.$$

Theorem

The following hold for $P_f : (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}_{++} \mapsto P_\lambda f(x)$:

(a) P_f *is locally Lipschitz at* $(\bar{x}, \bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}_{++}$.

(b) *If* ∂f *is proto-differentiable at* $\left(P_f(\bar{x}, \bar{\lambda}), \frac{\bar{x} - P_f(\bar{x}, \bar{\lambda})}{\bar{\lambda}}\right)$, *then* P_f *is directionally differentiable at* $(\bar{x}, \bar{\lambda})$ *with*

$$P_f'((\bar{x}, \bar{\lambda}); (d, \Delta)) = \left[\bar{\lambda} D(\partial f)\left(P_f(\bar{x}, \bar{\lambda}) \middle| \frac{\bar{x} - P_f(\bar{x}, \bar{\lambda})}{\bar{\lambda}}\right) + I\right]^{-1} \left(d - \frac{\Delta}{\bar{\lambda}}(\bar{x} - P_f(\bar{x}, \bar{\lambda}))\right).$$

(c) *If* ∂f *is proto-differentiable and semismooth* at* $\left(P_f(\bar{x}, \bar{\lambda}), \frac{\bar{x} - P_f(\bar{x}, \bar{\lambda})}{\bar{\lambda}}\right)$ *then* P_f *is semismooth at* $(\bar{x}, \bar{\lambda})$.

Note: $f \in C^2$ or f PLQ $\Longrightarrow \partial f$ proto-differentiable and semismooth*.

References

A. Berk, S. Brugiapaglia, and T. Hoheisel: *Square Root LASSO: well-posedness, Lipschitz stability and the tuning trade off.* arXiv, March 2023.

A. Berk, S. Brugiapaglia, and T. Hoheisel:*LASSO reloaded: a variational analysis perspective with applications to compressed sensing.* arXiv, May 2022.

M.P. Friedlander, A. Goodwin, and T. Hoheisel:*From perspective maps to epigraphical projections.* Mathematics Operations of Research, 2021.

Future directions

- Explore new techniques by Gfrerer and Outrata (*subspace containing derivatives*) for establishing strong metric regularity of hypomonotone operators (e.g., subdifferentials of weakly convex functions).
- Clarify the relation between proto-differentiability and semismoothness*.
- Apply the graphical derivative-based implicit function framework to, e.g.,:
  - regularized (linear) least-squares with PLQ regularizers;
  - nuclear norm regularized minimization.
- Explore implications in bilevel optimization.

Thanks for coming!