Stability of nonsmooth optimization problems

Tim Hoheisel (McGill)

Based on joint work with Aaron Berk (DeepRender), Simone Brugiapaglia (Concordia), Ebrahim Sarabi (Miami University), Nghia Tran (Oakland University)

ISMP 2024

Montréal, Québec, May 31, 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

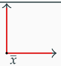
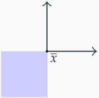
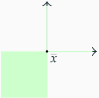
- (prox operator) $p := (\bar{x}, \lambda),\ h(p, x) := \frac{1}{2\lambda}\|x - \bar{x}\|^2$: $S(\bar{x}, \lambda) = P_\lambda \varphi(\bar{x})$.
- (unconstrained LASSO) $p := (A, b, \lambda),\ h(p, x) = \frac{1}{2\lambda}\|Ax - b\|^2,\ \varphi = \|\cdot\|_1$.

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

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

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

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

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.
- Graphical 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

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

- For more (subtle) conditions implying proto-differentiability, see, e.g., Hang and Sarabi (SIOPT 2024).

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 $\bar{x}$.

The workhorse (Dontchev/Rockafellar, Berk/Brugiapaglia/H.)

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

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

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

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

(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

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

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

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

then S is semismooth at $\bar{p}$.

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

The Mordukhovich criterion for regularized linear least-squares

Consider the regularized least-squares problem

$$\min_x \frac{1}{2}\|Ax-b\|^2 + \lambda g(x) \tag{3}$$

for $\lambda > 0$ and g closed, proper, convex.

Let $\bar{x}$ solve (3), i.e. $\bar{u} := \frac{1}{\lambda}A^T(b - A\bar{x}) \in \partial g(\bar{x})$, i.e.

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

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

$$-\frac{1}{\lambda}A^*Aw \in D^*(\partial g)(\bar{x}|\bar{u})(w). \tag{4}$$

By ‘positive semidefiniteness’ of $D^*(\partial g)(\bar{x}|\bar{u})$ we have

$$0 \le \langle w, -A^*Aw \rangle = -\|Aw\|^2 \quad \Longleftrightarrow \quad w \in \ker A$$

Inserting into (4) yields

$$0 \in D^*(\partial g)(\bar{x}|\bar{u})(w) \quad \overset{(\partial g)^{-1} = \partial g^*}{\Longleftrightarrow} \quad -w \in D^*(\partial g^*)(\bar{u}|\bar{x})(0).$$

Hence

$$\ker A \bigcap D^*(\partial g^*)(\bar{u}|\bar{x})(0) = \{0\} \quad \Longleftrightarrow \quad \text{Mordukhovich criterion holds} \tag{5}$$

Tangible conditions for the Mordukhovich criterion

Example Let $\bar{x}$ be a solution of the regularized linear least-squares problem

$$\min_x \frac{1}{2}\|Ax - b\|^2 + \lambda g(x), \tag{6}$$

i.e., $\bar{u} := \frac{1}{\lambda}A^*(b - A\bar{x}) \in \partial g(\bar{x})$.

- ($g^* \in C^{1,1}$) *If g^* has locally Lipschitz gradient*[1] *at $\bar{u}$, then*

$$D^*(\partial g^*)(\bar{u}|\bar{x})(0) \subset \partial^C(\nabla g^*)(\bar{u})^* 0 = \{0\}.$$

- (Polyhedral support) *Let $\mathcal{P} = \{x \mid \langle p_i, x\rangle \le \beta_i \ \forall i = 1, \dots, l\}$, and let $g = \sigma_{\mathcal{P}}$ be its support function. Then*

$$D^*(\partial g^*)(\bar{u}|\bar{x})(0) = D^* N_{\mathcal{P}}(\bar{u}|\bar{x})(0) = \operatorname{span}\{p_i \mid i : \langle p_i, \bar{u}\rangle = \beta_i\} = \operatorname{par}\partial g^*(\bar{u}).$$

We define the qualification condition

$$\operatorname{par}\partial g^*(\bar{u}) \cap \ker A = \{0\} \quad (\mathbf{R}).$$

Note: The condition **(R)** is (equivalent to) *generalized LICQ*[2] for the dual problem of (6)

$$\min_{y,t} \frac{\lambda}{2}\|y\|^2 - \langle b, y\rangle + t \quad \text{s.t.} \quad (A^*y, t) \in \operatorname{epi} g^*.$$

[1] See Goebel and Rockafellar (Journal of Convex Analysis, 2008) for a primal characterization.

[2] Or *partial constraint nondegeneracy*

Towards more general results

Proposition (Tran, H./Sarabi, H. '24) Let $\bar{x}$ be a solution of the regularized linear least-squares problem

$$\min_x \frac{1}{2}\|Ax - b\|^2 + \lambda g(x), \quad \lambda > 0 \tag{7}$$

with $\bar{u} = \frac{1}{\lambda}A^*(b - A\bar{x})$. Assume that g is in either of the following classes:

(i) (C^2-cone reducible conjugate) $\operatorname{epi} g^*$ is C^2*-cone reducible*[3]

(ii) (PLQ penalty) $g = \theta_{\mathcal{P},B}$ with

$$\theta_{\mathcal{P},B}(y) = \sup_{z\in\mathcal{P}} \left\{ \langle y, z\rangle - \frac{1}{2}\langle Bz, z\rangle \right\}, \quad B \succeq 0,\ \mathcal{P} \text{ polyhedron.}$$

Let $\bar{x}$ be a solution of (7) such that **(R)** holds. Then the solution map

$$(\hat{A}, \hat{b}, \hat{\lambda}) \mapsto \operatorname*{argmin}_x \frac{1}{2}\|\hat{A}x - \hat{b}\|^2 + \hat{\lambda} g(x)$$

is locally Lipschitz around (A, b, λ).

[3] See Bonnans/Shapiro (2000)

Application: unconstrained LASSO (constraint qualifications)

The unconstrained LASSO[4] for $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, \lambda > 0$ reads

$$\min_x \frac{1}{2}\|Ax - b\|^2 + \lambda\|x\|_1. \tag{8}$$

For a solution $\bar{x}$ of (8) define:

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

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

Qualification conditions

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

(Strong) $\implies$ (Intermediate) $\implies$ $\bar{x}$ is unique solution of (8)

[4] 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

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

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 (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: $(x_0)_j \overset{\text{iid}}{\sim} \mathcal{N}(m, m)$ $(j \in I)$.

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

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

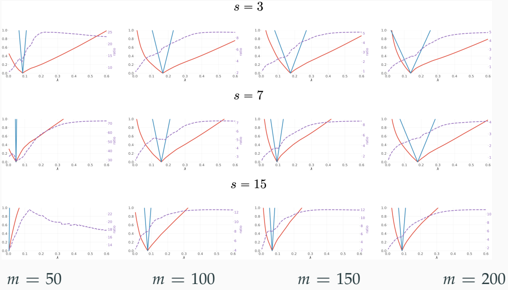


Figure 1: $\|x(\lambda) - \bar{x}\|$, $L|\lambda - \lambda^*|$, $\frac{L|\lambda - \lambda^*|}{\|x(\lambda) - \bar{x}\|}$.

References and Future directions

References

- T. Hoheisel and E. Sarabi: *Stability of regularized least-squares with PLQ regularizers.* Working paper, 2024.
- T. Hoheisel and N. Tran: *Lipschitz stability of least-squares problems with conjugate C^2-cone reducible regularizers.* Working paper, 2024.
- A. Berk, S. Brugiapaglia, and T. Hoheisel: *Square Root LASSO: well-posedness, Lipschitz stability and the tuning trade off.* SIAM Journal on Optimization, to appear.
- A. Berk, S. Brugiapaglia, and T. Hoheisel:*LASSO reloaded: a variational analysis perspective with applications to compressed sensing.* SIAM Journal on Mathematics of Data Science 5(4), 2023, pp. 1102–1129
- M.P. Friedlander, A. Goodwin, and T. Hoheisel:*From perspective maps to epigraphical projections.* Mathematics Operations of Research 48(2), 2023, pp. 1712–1740.

Future directions

- Expand *qualitative* analysis.
- Clarify the relation between proto-differentiability and semismoothness*.
- Explore implications in bilevel optimization.

Thanks for coming!