GLOBALLY CONVERGENT CODERIVATIVE-BASED NEWTONIAN ALGORITHMS IN NONSMOOTH OPTIMIZATION

BORIS MORDUKHOVICH

Wayne State University, USA

Lecture 4 at the **Forum on Developments and Origins of Operations Research**

Shenzhen, China, November 2021

ORGANIZERS: **Operations Research Society of China & Southern University of Science and Technology**

NEWTONIAN METHODS FOR SMOOTH FUNCTIONS

First consider the unconstrained optimization problem

$$\text{minimize } \varphi(x) \text{ subject to } x \in I\!R^n$$

with $\mathcal{C}^2$-smooth objective function φ. The classical Newton method exhibits the local convergence with a quadratic rate provided that $\nabla^2\varphi(\bar{x})$ is positive-definite. To achieve the global convergence, various line search procedures are used

$$x^{k+1} := x^k + \tau_k d^k \text{ with } -\nabla\varphi(x^k) = H_k d^k$$

where H_k is an appropriate approximation of the Hessian $\nabla^2\varphi(\bar{x})$ for quasi-Newton methods. The Levenberg-Marquardt method

$$H_k := \nabla^2\varphi(x^k) + \mu_k I \text{ with } \mu_k := c\|\nabla\varphi(x^k)\|$$

works when $\nabla^2\varphi(x^k)$ is merely positive-semidefinite.

MAJOR GOALS

Replacing the Hessian $\nabla^2\varphi$ by its coderivative-based generalized Hessian (second-order subdifferential) $\partial^2\varphi$, pursue the following:

• Design and justify the globally convergent generalized damped Newton method with the backtracking line search for unconstrained problems of $\mathcal{C}^{1,1}$ optimization.

• Design and justify the globally convergent Levenberg-Marquardt method with the backtracking line search for unconstrained problems of $\mathcal{C}^{1,1}$ optimization.

• Using forward-backward envelopes, extend both coderivative-based generalized Newton methods to problems of convex composite optimization encompassing problems with constraints.

• Solving Lasso problems by the developed generalized Newton algorithms with numerical experiments and comparison with other first-order and second-order algorithms of optimization.

NORMALS, CODERIVATIVES, SUBGRADIENTS

See [M06,M18,RW98] for more details and references. The (limiting) **normal cone** to $\Omega \subset \mathbb{R}^n$ at $\bar{x} \in \Omega$ from

$$N_\Omega(\bar{x}) := \Big\{ v \;\Big|\; \exists\, x_k \to \bar{x},\ \alpha_k \ge 0,\ w_k \in \Pi_\Omega(x_k),\ \alpha(x_k - w_k) \to v \Big\}$$

where Π_Ω stands for the Euclidean projection. The **coderivative** of $F\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ at $(\bar{x},\bar{y}) \in \operatorname{gph} F$

$$D^*F(\bar{x},\bar{y})(v) := \Big\{ u \in \mathbb{R}^n \;\Big|\; (u,-v) \in N_{\operatorname{gph} F}(\bar{x},\bar{y}) \Big\}, \quad v \in \mathbb{R}^m.$$

When $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is $\mathcal{C}^1$-smooth, then

$$D^*F(\bar{x})(v) = \Big\{ \nabla F(\bar{x})^* v \Big\}, \quad v \in \mathbb{R}^m,$$

via the adjoint/transpose Jacobian matrix. The (first-order) **subdifferential** of $\varphi\colon \mathbb{R}^n := (-\infty,\infty]$ at $\bar{x} \in \operatorname{dom}\varphi$ [M76]

$$\partial\varphi(\bar{x}) := \Big\{ v \in \mathbb{R}^n \;\Big|\; (v,-1) \in N_{\operatorname{epi}\varphi}\big(\bar{x},\varphi(\bar{x})\big) \Big\}.$$

Despite their nonconvexity these constructions enjoy full calculus based on the variational/extremal principles of variational analysis.

GENERALIZED HESSIANS

The **second-order subdifferential**, or **generalized Hessian** of $\varphi\colon I\!R^n \to \overline{I\!R}$ at $\bar{x} \in \operatorname{dom}\varphi$ for $\bar{v} \in \partial\varphi(\bar{x})$ is defined as [M92]

$$\partial^2\varphi(\bar{x},\bar{v})(u) := \Big(D^*\partial\varphi\Big)(\bar{x},\bar{y})(u), \quad u \in I\!R^n.$$

If φ is $\mathcal{C}^2$-smooth around $\bar{x}$, then

$$\partial^2\varphi(\bar{x})(u) = \Big\{\nabla^2\varphi(\bar{x})u\Big\}, \quad u \in I\!R^n.$$

If φ of class $\mathcal{C}^{1,1}$($\mathcal{C}^1$ with Lipschitz gradient) around $\bar{x}$, then

$$\partial^2\varphi(\bar{x})(u) = \partial\Big\langle u, \nabla\varphi(\bar{x})\Big\rangle, \quad u \in I\!R^n.$$

It is realized that the generalized Hessian $\partial^2\varphi$ enjoys well-developed second-order calculus and can be viewed as an appropriate replacement of the Hessian $\nabla^2\varphi$ for nonsmooth problems. $\partial^2\varphi$ is fully computed in terms of the given data for broad classes of problems in optimization and variational analysis.

DAMPED NEWTON METHOD IN $\mathcal{C}^{1,1}$ OPTIMIZATION

Algorithm 1 Coderivative-based damped Newton algorithm for $\mathcal{C}^{1,1}$

Input: $x^0 \in I\!R^n$, $\sigma \in \left(0, \frac{1}{2}\right)$, $\beta \in (0,1)$

for $k = 0, 1, \ldots$ **do**
    If $\nabla\varphi(x^k) = 0$, stop; otherwise go to the next step
    Choose $d^k \in I\!R^n$ such that $-\nabla\varphi(x^k) \in \partial^2\varphi(x^k)(d^k)$
    Set $\tau_k = 1$.
    while $\varphi(x^k + \tau_k d^k) > \varphi(x^k) + \sigma\tau_k\langle d^k, \nabla\varphi(x^k)\rangle$ **do**
        set $\tau_k := \beta\tau_k$
    end while
    Set $x^{k+1} := x^k + \tau_k d^k$
end for

The main assumption for the well-posedness and global convergence

(PD) generalized Hessian $\partial^2\varphi$ is positive-definite on $I\!R^n$.

TILT STABILITY IN OPTIMIZATION

DEFINITION (Pol-Roc98) For $\varphi\colon I\!\!R^n \to \overline{I\!\!R}$, a point $\bar{x} \in \operatorname{dom}\varphi$ is a **tilt-stable local minimizer** with modulus ℓ if there is γ such that

$$M_\gamma\colon v \mapsto \text{argmin}\Big\{\varphi(x) - \langle v, x\rangle \;\Big|\; x \in I\!\!B_\gamma(\bar{x})\Big\}$$

is single-valued and Lipschitz continuous around the origin in $I\!\!R^n$ with constant ℓ and such that $M_\gamma(0) = \{\bar{x}\}$.

Theorem (Pol-Roc98) Let $\varphi\colon I\!\!R^n \to \overline{I\!\!R}$ is prox-regular and subdifferentially continuous [RW98] at $\bar{x}$ for $\bar{v} \in \partial\varphi(\bar{x})$ (this holds, in particular, for $\mathcal{C}^{1,1}$ and for convex functions). Then $\bar{x}$ is tilt stable local minimizer of φ for $\bar{v}$ **if and only if**

$$\partial^2\varphi(\bar{x},\bar{v}) \text{ is positive-definite.}$$

By now we have complete characterizations of tilt stability with precise formulas for computing the best modulus bounds for major classes problems in constrained optimization and optimal control.

WELL-POSEDNESS AND CONVERGENCE OF ALGORITHM 1

Theorem[KMPT21] Let $\varphi\colon I\!R^n \to I\!R$ be of class $\mathcal{C}^{1,1}$ under the fulfillment of (PD). Then whenever $\partial\varphi(x) \neq 0$ there is $d \neq 0$ with

$$-\nabla\varphi(x) \in \partial^2\varphi(x)(d) \quad \text{and} \quad \langle \varphi(x), d\rangle < 0.$$

Thus for each $\sigma \in (0,1)$ there exists $\delta > 0$ such that

$$\varphi(x+\tau d) \le \varphi(x) + \sigma\tau\langle \nabla\varphi(x), d\rangle \quad \text{whenever} \quad \tau \in (0,\delta).$$

Furthermore, for any starting point x^0, each limiting point $\bar{x}$ of the sequence of iterates $\{x^k\}$ is a tilt-stable local minimizer of φ satisfying the following conditions:

- The convergence rate of the sequence $\{\varphi(x^k)\}$ is at least Q-linear.
- The convergence rates of both sequences $\{x^k\}$ and $\{\|\nabla\varphi(x^k)\|\}$ are at least R-linear.

SUPERLINEAR GLOBAL CONVERGENCE OF ALGORITHM 1

Definition [Gfrerer-Outrata21] A mapping $F\colon I\!R^n \rightrightarrows I\!R^m$ is **semismooth***
at $(\bar{x},\bar{y}) \in \mathrm{gph}\, F$ if

$$\langle u^*, u\rangle = \langle v^*, v\rangle \ \text{ for all } \ (v^*, u^*) \in \mathrm{gph}\, D^*F\Big((\bar{x},\bar{y});(u,v)\Big).$$

For single-valued and locally Lipschitzian mappings, this reduces to the semismooth property if F is directionally differentiable.

Theorem [KMPT21] In the setting of the previous theorem, suppose that $\nabla\varphi(\bar{x})$ is semismooth* at $\bar{x}$. Then $\{x^k\}$ Q-superlinearly converges to $\bar{x}$ provided that either $\nabla\varphi$ is directionally differentiable at $\bar{x}$, or $\sigma \in (0, 1/(2\ell\kappa))$, where κ is a modulus of tilt stability of $\bar{x}$ and ℓ is a Lipschitz constant of $\nabla\varphi$ around $\bar{x}$. Moreover, in this case the sequence $\{\varphi(x^k)\}$ converges Q-superlinearly to $\varphi(\bar{x})$, and the sequence $\{\nabla\varphi(x^k)\}$ convergesQ-superlinearly to 0 as $k \to \infty$.

LEVENBERG-MARQUARDT METHOD IN $\mathcal{C}^{1,1}$ OPTIMIZATION

The (PD) assumption is now replaced by

(PSD) generalized Hessian $\partial^2\varphi$ is positive-semidefinite on $I\!R^n$.

Algorithm 2 Levenberg-Marquardt algorithm for $\mathcal{C}^{1,1}$ functions

Input: $x^0 \in I\!R^n$, $c > 0$, $\sigma \in \left(0, \frac{1}{2}\right)$, $\beta \in (0, 1)$

1: **for** $k = 0, 1, \ldots$ **do**

2: If $\nabla\varphi(x^k) = 0$, stop; else let $\mu_k := c\|\nabla\varphi(x^k)\|$ and go to Step 3

3: Choose $d^k \in I\!R^n$ such that $-\nabla\varphi(x^k) \in \partial^2\varphi(x^k)(d^k) + \mu_k d^k$

4: Set $\tau_k = 1$

5: **while** $\varphi(x^k + \tau_k d^k) > \varphi(x^k) + \sigma\tau_k\langle\nabla\varphi(x^k), d^k\rangle$ **do**

6: set $\tau_k := \beta\tau_k$

7: **end while**

8: Set $x^{k+1} := x^k + \tau_k d^k$

9: **end for**

WELL-POSEDNESS AND CONVERGENCE OF ALGORITHM 2

Theorem[KMPT21] Let φ be of class $\mathcal{C}^{1,1}$ under the fulfillment of (PSD). If $\partial\varphi(x) \neq 0$ and $\varepsilon > 0$, then there is $d \neq 0$ with

$$-\nabla\varphi(x) \in \partial^2\varphi(x)(d) + \varepsilon d \quad \text{and} \quad \langle \varphi(x), d\rangle < 0.$$

Thus for each $\sigma \in (0,1)$ there exists $\delta > 0$ such that

$$\varphi(x + \tau d) \leq \varphi(x) + \sigma\tau\langle\nabla\varphi(x), d\rangle \quad \text{whenever} \quad \tau \in (0, \delta).$$

Furthermore, any starting point x^0 produces iterates $\{x^k\}$ such that the sequence of values $\{\varphi(x^k)\}$ is monotonically decreasing and all the limiting points of $\{x^k\}$ satisfy the stationary condition.

METRIC REGULARITY

DEFINITION [M93,RW98] A mapping $F\colon I\!R^n \rightrightarrows I\!R^m$ is **metrically regular** around $(\bar{x},\bar{y}) \in \operatorname{gph} F$ if there exist $\mu > 0$ and neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that

$$\operatorname{dist}\Big(x; F^{-1}(y)\Big) \le \mu \operatorname{dist}\Big(y; F(x)\Big) \quad \text{for all} \quad (x,y) \in U \times V,$$

where $F^{-1}(y) := \{x \in I\!R^n \mid y \in F(x)\}$.

Coderivative/Mordukhovich criterion: If a set-valued mapping $F\colon I\!R^n \rightrightarrows I\!R^m$ is of closed-graph around $(\bar{x},\bar{y})$, then its metric regularity around this point is equivalent to

$$D^*F(\bar{x},\bar{y})(0) = \{0\}.$$

RATES OF CONVERGENCE FOR ALGORITHM 2

THEOREM [KMPT21] Let $\bar{x}$ be a limiting point of the sequence of iterates in Algorithm 2. In addition to (PSD), suppose that $\nabla\varphi$ is metrically regular around this point. Then $\bar{x}$ is a tilt-stable local minimizer of φ, and Algorithm 2 converges to $\bar{x}$ with the convergence rates as follows:

• The sequence $\{\varphi(x^k)\}$ converges to $\varphi(\bar{x})$ at least Q-linearly.

• The sequences $\{x^k\}$ and $\{\nabla\varphi(x^k)\}$ converge at least R-linearly to $\bar{x}$ and 0, respectively.

• The convergence rates of $\{x^k\}$, $\{\varphi(x^k)\}$, $\{\nabla\varphi(x^k)\}$ are at least Q-superlinear if $\nabla\varphi$ is semismooth* at $\bar{x}$ and either one of the following two conditions holds:

(a) $\nabla\varphi$ is directionally differentiable at $\bar{x}$,

(b) $\sigma \in (0, 1/(2\ell\kappa))$, where $\kappa > 0$ and $\ell > 0$ are moduli of metric regularity and Lipschitz continuity of $\nabla\varphi$ around $\bar{x}$, respectively.

PROBLEMS OF CONVEX COMPOSITE OPTIMIZATION

Consider the class of optimization problems

$$\text{minimize } \varphi(x) := f(x) + g(x), \quad x \in I\!R^n,$$

where $f\colon I\!R^n \to I\!R$ is convex and smooth, while the regularizer $g\colon I\!R^n \to \overline{I\!R}$ is convex and extended-real-valued. This class encompasses problems of constrained optimization. For each $\gamma > 0$ consider the **proximal mapping** of the regularizer g by

$$\text{Prox}_{\gamma g}(x) := \underset{y \in I\!R^n}{\text{argmin}} \left\{ g(y) + \frac{1}{2\gamma}\|y - x\|^2 \right\}$$

and define [PB13] the **forward-backward envelope** (**FBE**) of φ

$$\varphi_\gamma(x) := \inf_{y \in I\!R^n} \left\{ f(x) + \langle \nabla f(x), y - x \rangle + g(y) + \frac{1}{2\gamma}\|y - x\|^2 \right\}.$$

If f is $\mathcal{C}^2$-smooth with the Lipschitz continuous ∇f, then

$$\nabla \varphi_\gamma(x) = \gamma^{-1}\Big(I - \gamma \nabla^2 f(x)\Big)\Big(x - \text{Prox}_{\gamma g}(x - \gamma \nabla f(x))\Big).$$

Algorithm 3 Coderivative-based damped Newton algorithm for convex composite optimization with $f(x) := \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle + \alpha$

Input: $x^0 \in I\!R^n$, $\gamma > 0$ such that $B := I - \gamma A \succ 0$, $\sigma \in \left(0, \frac{1}{2}\right)$, $\beta \in (0, 1)$, and φ_γ is FBE

1: **for** $k = 0, 1, \ldots$ **do**

2: If $\nabla\varphi_\gamma(x^k) \neq 0$, set $u^k := x^k - \gamma(Ax^k + b)$, $v^k := \text{Prox}_{\gamma g}(u^k)$

3: Find d^k as $-\frac{1}{\gamma}(x^k - v^k) - Ad^k \in \partial^2 g\left(v^k, \frac{1}{\gamma}(u^k - v^k)\right)(x^k - v^k + d^k)$

4: Set $\tau_k = 1$

5: **while** $\varphi_\gamma(x^k + \tau_k d^k) > \varphi_\gamma(x^k) + \sigma\tau_k\langle\nabla\varphi_\gamma(x^k), d^k\rangle$ **do**

6: set $\tau_k := \beta\tau_k$

7: **end while**

8: Set $x^{k+1} := x^k + \tau_k d^k$

9: **end for**

TWICE EPI-DIFFERENTIABILITY

The second subderivative [RW98] of $\varphi\colon I\!R^n \to \overline{I\!R}$ at $\bar{x}$ for v, w is

$$d^2\varphi(\bar{x}, v)(w) := \liminf_{\tau\downarrow 0,\, u\to w} \Delta_\tau^2\varphi(\bar{x}, v)(u) \quad \text{where}$$
$$\Delta_\tau^2\varphi(\bar{x}, v)(u) := \frac{\varphi(\bar{x}+\tau u)-\varphi(\bar{x})-\tau\langle v,u\rangle}{\frac{1}{2}\tau^2}.$$

The function φ is twice epi-differentiable at $\bar{x}$ for v if for every w and $\tau_k \downarrow 0$ there exists a sequence $w^k \to w$ such that

$$\frac{\varphi(\bar{x}+\tau_k w^k) - \varphi(\bar{x}) - \tau_k\langle v, w^k\rangle}{\frac{1}{2}\tau_k^2} \to d^2\varphi(\bar{x}, v)(w).$$

A general and verifiable condition for twice epi-differentiability is provided by parabolic regularity, which covers a large territory in second-order variational analysis and optimization [MMS21].

SUPERLINEAR CONVERGENCE OF ALGORITHM 3

THEOREM [KMPT21] If A is positive-definite, then Algorithm 3 generates a sequence $\{x^k\}$ such that it globally R-linearly converges to $\bar{x}$, which a tilt-stable local minimizer of φ with modulus $\kappa := 1/\lambda_{\min(A)}$. Furthermore, the convergence rate of $\{x^k\}$ is at least Q-superlinear if ∂g is semismooth* at $(\bar{x}, \bar{v})$, where $\bar{v} := -A\bar{x} - b$, and if either one of two following conditions is satisfied:

- $\sigma \in (0, 1/(2LK))$, where $L := 2\left(1 - \gamma\lambda_{\min(A)}\right)/\gamma$ and $K := \kappa + \gamma\|B^{-1}\|$.
- g is twice epi-differentiable at $\bar{x}$ for $\bar{v}$.

LEVENBERG-MARQUARDT FOR CONVEX OPTIMIZATION

Algorithm 4 Coderivative-based Levenberg-Marquardt algorithm for convex composite optimization

Input: $x^0 \in \mathbb{R}^n$, $\gamma > 0$ such that $B := I - \gamma A \succ 0$, $\lambda > 0$, $\sigma \in \left(0, \frac{1}{2}\right)$, $\beta \in (0,1)$, and φ_γ is FBE

for $k = 0, 1, \ldots$ **do**

  Set $u^k := x^k - \gamma(Ax^k + b), v^k := \text{Prox}_{\gamma g}(u^k), \mu_k := \lambda \|\nabla \varphi_\gamma(x^k)\|$

  Set $d^k = Bz^k$, where z^k is from $-\frac{1}{\gamma}(x^k - v^k) - (\mu_k I + AB)z^k \in \partial^2 g\left(v^k, \frac{1}{\gamma}(u^k - v^k)\right)\left(x^k - v^k + (B + \gamma\mu_k I)z^k\right)$

  Set $\tau_k = 1$

  while $\varphi_\gamma(x^k + \tau_k d^k) > \varphi_\gamma(x^k) + \sigma\tau_k \langle \nabla \varphi_\gamma(x^k), d^k \rangle$ **do**

    set $\tau_k := \beta\tau_k$

  end while

  Set $x^{k+1} := x^k + \tau_k d^k$

end for

GLOBAL CONVERGENCE OF ALGORITHM 4

THEOREM [KMPT21] Let A be positive-semidefinite. Then:

- Any limiting point $\bar{x}$ of iterates $\{x^k\}$ of Algorithm 4 is an optimal solutions to φ.
- If $\partial\varphi$ is metrically regular ar $(\bar{x}, 0)$ with modulus $\kappa > 0$, then the sequence $\{x^k\}$ globally R-linearly converges to $\bar{x}$, and $\bar{x}$ is a tilt-stable local minimizer of φ with modulus κ.
- The rate of convergence of $\{x^k\}$ is at least Q-superlinear if ∂g is semismooth* at $(\bar{x}, \bar{v})$, where $\bar{v} := -A\bar{x} - b$, and if either one of following two conditions holds:

(a) $\sigma \in (0, 1/(2LK))$, where $L := 2\left(1 - \gamma\lambda_{\min(A)}\right)/\gamma$ and $K := \kappa + \gamma\|B^{-1}\|$.

(b) g is twice epi-differentiable at $\bar{x}$ for $\bar{v}$.

SOLVING LASSO PROBLEMS

The basic Lasso problem appeared in statistic [T86] as

$$\text{minimize} \quad \varphi(x) := \frac{1}{2}\|Ax - b\|_2^2 + \mu\|x\|_1, \quad x \in \mathbb{R}^n,$$

where A is an $m \times n$ matrix and $\mu > 0$. All the parameters of Algorithms 3 (GDNM) and Algorithm 4 (GLMM) are computed entirely in terms of given data of the Lasso problem.

Numerical experiments are conducted for GDNM and GLMM by using random data with $\mu := 10^{-3}$ and compare with the performance of ADMM [BPCPE10], FISTA [BT09] and SSNAL [LST18].

The conducted experiments show that both GDNM and GLMM behave better (exhibiting the Q-superlinear convergence) than the other algorithms for $m \geq n$. It may not be the case for $m < n$ when GLMM behaves better than GDNM and often better than FISTA and ADMM but usually worse than SSNAL.

SOLVING LASSO ON RANDOM INSTANCES

Problem size and ID			iter					CPU time				
ID	m	n	SSNAL	FISTA	ADMM	GLMM	GDNM	SSNAL	FISTA	ADMM	GLMM	GDNM
1	400	800	25	37742	22873	1813	Error	0.45	145.52	10.89	45.62	Error
2	4000	8000	153	19173	19173	2499	Error	847.87	10000.00	2359.36	10000.00	Error
3	2000	2000	43	239701	12785	59	12	78.38	8138.94	158.12	11.07	2.24
4	4000	4000	246	73374	5970	59	218	1253.45	10000.00	320.81	48.16	178.91
5	2000	2000	22	3619	90501	394	292	18.11	123.38	1141.64	65.60	58.80
6	4000	4000	24	3629	103868	520	555	231.40	462.53	5166.16	369.27	474.74
7	800	400	4	430	10	6	3	0.14	0.86	0.02	0.11	0.08
8	8000	4000	13	487	11	7	3	18.80	117.92	3.67	8.46	4.39
9	800	400	11	245	426	31	7	0.18	0.53	0.12	0.23	0.11
10	8000	4000	11	238	411	72	9	8.37	59.18	32.17	56.37	8.88

REFERENCES

[BT09] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J. Imaging Sci. 2 (2009), 183–202.

[BPCPE10] S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learning 3 (2010), 1–122.

[GO21] H. Gfrerer and J. V. Outrata, On a semismooth* Newton method for solving generalized equations, SIAM J. Optim. 31 (2021), 489–517.

[KMPT21] P. D. Khanh, B. S. Mordukhovich, V. T. Phat and D. B. Tran, Globally convergent coderivative-based generalized Newton methods in nonsmooth optimization, arXiv:2109.02093.

[LST21] X. Li, D. Sun and K.-C. Toh, A highly efficient semismooth Newton augmented Lagrangian method for solving Lasso problems, SIAM J. Optim. 28 (2018), 433–458.

[MMS21] A. Mohammadi, B. S. Mordukhovich and M. E. Sarabi, Parabolic regularity in geometric variational analysis, Trans. Amer. Math. Soc. 374 (2021), 1711–1763.

[M76] B. S. Mordukhovich, Maximum principle in problems of time optimal control with nonsmooth constraints, J. Appl. Math. Mech. 40 (1976), 960–969.

[M92] B. S. Mordukhovich, Sensitivity analysis in nonsmooth optimization. In Theoretical Aspects of Industrial Design (D. A. Field and V. Komkov, eds.), pp. 32–46, SIAM Proc. Appl. Math. 58, Philadelphia, 1992.

[M93] B. S. Mordukhovich, Complete characterizations of openness, metric regularity, and Lipschitzian properties of multifunctions, Trans. Amer. Math. Soc. 340, 1–35 (1993)

[M06] B. S. Mordukhovich, Variational Analysis and Generalized Differentiation, I: Basic Theory, II: Applications, Springer, Berlin, 2006.

[M18] B. S. Mordukhovich, Variational Analysis and Applications, Springer, Cham, Switzerland, 2018.

[PB13] P. Patrinos and A. Bemporad, Proximal Newton methods for convex composite optimization, In: IEEE Conference on Decision and Control (2013), 2358–2363.

[PR98] R. A. Poliquin and R. T. Rockafellar, Tilt stability of a local minimum, SIAM J. Optim. 8 (1998), 287–299.

[RW98] R. R. Rockafellar and R. J-B. Wets, Variational Analysis, Springer, Berlin, 1998.