Generalized derivatives of the normal cone mapping to inequality systems and their applications

Helmut Gfrerer

Institute of Computational Mathematics
Johannes Kepler University Linz, Austria

Der Wissenschaftsfonds.

Introduction

Consider the following bilevel programming problem

$$(BLP) \qquad \min_{(x,y)\in\mathbb{R}^n\times\mathbb{R}^m} \quad f(x,y) \qquad \text{s.t.} \quad y \in S(x), \quad x \in C$$

where $S(x)$ denotes the set of solutions of the lower level problem

$$(P_x) \qquad \min_{y\in\mathbb{R}^m} \quad \varphi(x,y) \qquad \text{s.t.} \quad y \in \Gamma$$

- $f, \varphi : \mathbb{R}^{n\times m} \to \mathbb{R}$ are C^2
- $C \subset \mathbb{R}^n$ closed.
- $\Gamma := \{y \in \mathbb{R}^m \,|\, q_i(y) \leq 0, i = 1, \ldots, l\}$ with $q_i \in C^2$

First-order approach for (BLP)

- Condition $y \in S(x)$ is replaced by the first order optimality conditions.
- Karush-Kuhn-Tucker conditions: Under some Constraint qualification (CQ) there are Lagrange multipliers $\lambda \in \mathbb{R}^l$ such that

$$0 = \nabla_y \varphi(x,y) + \lambda^T \nabla q(y),\ \lambda \geq 0,\ q(y) \leq 0, \lambda^T q(y) = 0$$

(Complementarity conditions)
Disadvantage: Multiplier λ is introduced as additional variable and this will cause troubles if the multiplier is not unique

Example

Upper level objective: $\min_{x,y}(x_1 - 1)^2 + (x_2 - 1)^2$
Lower level problem:

$$
\begin{array}{lll}
(P_x) & \min\limits_{y} & x_1 y_1 + x_2 y_2 \\
 & \text{s.t.} & -y_2 \le 0 \\
 & & \frac{1}{2} y_1^2 - y_2 \le 0
\end{array}
$$

Unique solution is $\bar{x} = (1,1)$, $\bar{y} = (-1, \frac{1}{2})$. But for the formulation with complementarity conditions the point $\bar{x} = (0,1)$ $\bar{y} = (0,0)$, $\lambda = (1,0)$ also is a local minimizer.

Alternative to KKT condition

- Working with the generalized equation

$$0 \in \nabla_y \varphi(x, y) + \hat{N}_\Gamma(y)$$

and modern methods of variational analysis without introducing multipliers to the overall problem.

- $\hat{N}_\Gamma(y)$ denotes the regular normal cone to Γ at y.
- Hence we replace (BLP) by

$$(FOP) \qquad \min_{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m} \quad f(x, y)$$
$$\text{s.t.} \quad 0 \in \nabla_y \varphi(x, y) + \hat{N}_\Gamma(y)$$
$$x \in C$$

Combined first order and value function approach

If the lower level program is not convex, the first order approach may fail. In this case we can replace (BLP) by the equivalent problem

$$
\begin{array}{lll}
(VFP) & \displaystyle\min_{(x,y)\in\mathbb{R}^n\times\mathbb{R}^m} & f(x,y) \\
& \text{s.t.} & \varphi(x,y) \le V(x), \\
& & 0 \in \nabla_y\varphi(x,y) + \hat{N}_\Gamma(y), \\
& & x \in C,
\end{array}
$$

where

$$V(x) := \min\{\phi(x,y) \mid y \in \Gamma\}$$

denotes the optimal value function of the lower level problem.

Variational Geometry

Definition

Let $\Omega \subset \mathbb{R}^d$ be closed, $\bar{z} \in \Omega$

1. The (Bouligand-Severi) *tangent/contingent cone* to Ω at $\bar{z}$ is defined by

$$T_\Omega(\bar{z}) := \left\{ u \in \mathbb{R}^d \mid \exists\, t_k \downarrow 0,\ u_k \to u \ \text{ with } \ \bar{z} + t_k u_k \in \Omega \right\}.$$

2. The (Fréchet) *regular normal cone* to Ω at $\bar{z}$ can be equivalently defined by

$$\hat{N}_\Omega(\bar{z}) := \left\{ v^* \in \mathbb{R}^d \mid \limsup_{z \xrightarrow{\Omega} \bar{z}} \frac{\langle v^*, z - \bar{z} \rangle}{\|z - \bar{z}\|} \le 0 \right\} = (T_\Omega(\bar{z}))^\circ.$$

The regular normal cone is always convex.

Definition

Let $\Omega \subset \mathbb{R}^d$ be closed, $\bar{z} \in \Omega$, $u \in \mathbb{R}^d$.

1. The (Mordukhovich) *limiting normal cone* to Ω at $\bar{z}$ is defined by

$$N_\Omega(\bar{z}) := \left\{ v^* \in \mathbb{R}^d \,|\, \exists z_k \stackrel{\Omega}{\to} \bar{z},\ v_k^* \to v^* :\ v_k^* \in \hat{N}_\Omega(z_k) \right\}$$

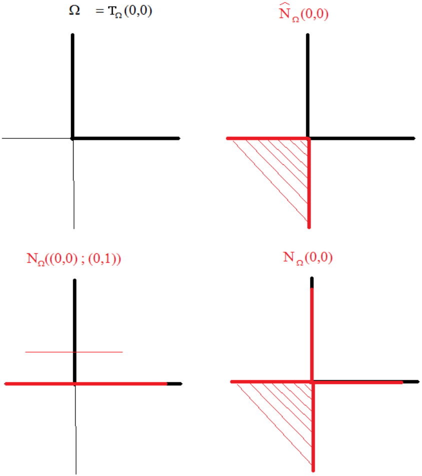
2. (Gfr2013) The *limiting normal cone* to Ω at $\bar{z}$ in direction u is defined by

$$N_\Omega(\bar{z}; u) := \left\{ v^* \in \mathbb{R}^d \,|\, \exists t_k \downarrow 0,\ u_k \to u,\ v_k^* \to v^* :\ v_k^* \in \hat{N}_\Omega(\bar{z} + t_k u_k) \right\}$$

- In general the limiting normal cone is nonconvex and $\hat{N}_\Omega(\bar{z}) \subset N_\Omega(\bar{z})$.
- If Ω is convex then $\hat{N}_\Omega(\bar{z}) = N_\Omega(\bar{z})$ coincides with the normal cone of convex analysis.
- $N_\Omega(\bar{z}) = N_\Omega(\bar{z}; 0)$.
- If $u \notin T_\Omega(\bar{z})$ then $N_\Omega(\bar{z}; u) = \emptyset$

Example

$$\Omega = \{(x_1, x_2) \in \mathbb{R}^2_+ \mid x_1 x_2 = 0\}$$



Generalized differentiation

Definition

Let $\Psi : \mathbb{R}^d \rightrightarrows \mathbb{R}^s$ be a multifunction,
$(\bar{z}, \bar{w}) \in \operatorname{gph} \Psi := \{(z, w) \mid w \in \Psi(z)\}$, $(u, v) \in \mathbb{R}^d \times \mathbb{R}^s$. The

1. *graphical derivative* $D\Psi(\bar{z}, \bar{w}) : \mathbb{R}^d \rightrightarrows \mathbb{R}^s$,
2. *regular coderivative* $\hat{D}^*\Psi(\bar{z}, \bar{w}) : \mathbb{R}^s \rightrightarrows \mathbb{R}^d$,
3. *limiting coderivative* $D^*\Psi(\bar{z}, \bar{w}) : \mathbb{R}^s \rightrightarrows \mathbb{R}^d$,
4. *limiting coderivative in direction* (u, v)
   $D^*\Psi((\bar{z}, \bar{w}); (u, v)) : \mathbb{R}^s \rightrightarrows \mathbb{R}^d$

are defined by

$$\operatorname{gph} D\Psi(\bar{z}, \bar{w}) = T_{\operatorname{gph} \Psi}(\bar{z}, \bar{w}),$$

$$\operatorname{gph} \hat{D}^*\Psi(\bar{z}, \bar{w}) = \{(w^*, z^*) \mid (z^*, -w^*) \in \hat{N}_{\operatorname{gph} \Psi}(\bar{z}, \bar{w})\},$$

$$\operatorname{gph} D^*\Psi(\bar{z}, \bar{w}) = \{(w^*, z^*) \mid (z^*, -w^*) \in N_{\operatorname{gph} \Psi}(\bar{z}, \bar{w})\},$$

$$\operatorname{gph} D^*\Psi((\bar{z}, \bar{w}); (u, v)) = \{(w^*, z^*) \mid (z^* - w^*) \in N_{\operatorname{gph} \Psi}((\bar{z}, \bar{w}); (u, v))\}.$$

Optimization problems

Consider

$$(P) \quad \min_z f(z) \quad \text{subject to} \quad 0 \in \Psi(z)$$

where

- $f : \mathbb{R}^d \to \mathbb{R}$ continuously differentiable
- $\Psi : \mathbb{R}^d \rightrightarrows \mathbb{R}^s$ is a multifunction with closed graph

Then the basic optimality condition at a local minimizer $\bar{z}$ is

$$\nabla f(\bar{z})u \geq 0 \;\forall u \in T_{\psi^{-1}(0)}(\bar{z})$$

which can be equivalently written as

$$0 \in \nabla f(\bar{z}) + \hat{N}_{\psi^{-1}(0)}(\bar{z})$$

(B-stationarity: no feasible descent direction exists.)

Constraint qualifications

We want to know wether a small residual $\mathrm{d}(0, \Psi(z))$ at a point z means that z is near to the feasible region $\Psi^{-1}(0)$.

Definition

- Ψ is called metrically regular around $(\bar{z}, 0) \in \operatorname{gph} \Psi$ with modulus $\kappa > 0$, if there are neighborhoods U of $\bar{z}$ and V of 0 such that

$$\mathrm{d}(z, \Psi^{-1}(w)) \leq \kappa \mathrm{d}(w, \Psi(z)) \qquad \forall z \in U, w \in V.$$

- Ψ is called metrically subregular at $(\bar{z}, 0) \in \operatorname{gph} \Psi$ with modulus $\kappa > 0$, if there is a neighborhood U of $\bar{z}$ such that

$$\mathrm{d}(z, \Psi^{-1}(0)) \leq \kappa \mathrm{d}(0, \Psi(z)) \qquad \forall z \in U.$$

Note: If Ψ is metrically subregular at $(\bar{z}, 0)$ then

$$T_{\Psi^{-1}(0)}(\bar{z}) = \{u \mid 0 \in D\Psi(\bar{z}, 0)(u)\}$$

Characterizations of metric (sub)regularity

Note:

- Ψ is metrically regular around $(\bar{z}, 0) \Leftrightarrow \Psi^{-1}$ has the *Aubin property* (is *Lipschitz-like, pseudo-Lipschitz*) around $(0, \bar{z})$.
- Ψ is metrically subregular at $(\bar{z}, 0) \Leftrightarrow \Psi^{-1}$ is *calm* at $(0, \bar{z})$.

Theorem

Let $0 \in \psi(\bar{z})$*.*

1. *(Mordukhovich criterion)* Ψ *is metrically regular around* $(\bar{z}, 0)$ *if and only if*
$$0 \in D^*\Psi(\bar{z}, 0)(w^*) \;\Rightarrow\; w^* = 0$$
2. *(Gfr.2013, First order sufficient condition for metric subregularity (FOSCMS))* Ψ *is metrically subregular at* $(\bar{z}, 0)$ *if for every* $u \neq 0$ *with* $0 \in D\Psi(\bar{z}, 0)(u)$ *one has*
$$0 \in D^*\Psi((\bar{z}, 0), (u, 0))(w^*) \;\Rightarrow\; w^* = 0$$

Metric (sub)regularity of inequality systems

$$\Gamma = \{y \,|\, 0 \in M(y)\}, \quad M(y) := q(y) - \mathbb{R}^l_-$$

Notation:

$$\mathcal{I}(y) := \{i \,|\, q_i(y) = 0\},\ T^{\rm lin}_\Gamma(y) := \{v \,|\, \nabla q_i(y)v \le 0,\ i \in \mathcal{I}(y)\} \quad \forall y \in \Gamma$$

Theorem

1. *(Mordukhovich criterion) M is metrically regular around* $(\bar y, 0)$ *iff*

$$\nabla q(\bar y)^T\lambda = 0,\ \lambda \ge 0, q(\bar y)^T\lambda = 0 \ \Rightarrow\ \lambda = 0.$$

Moreover, this condition holds if and only if MFCQ is fulfilled at $\bar y$.

2. *(Gfr.2011, Second order sufficient condition for metric subregularity (SOSCMS)): If for every* $0 \ne v \in T^{\rm lin}_\Gamma(\bar y)$ *one has*

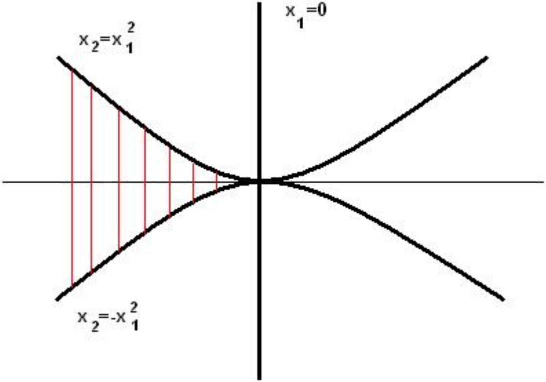
$$\nabla q(\bar y)^T\lambda = 0,\ \lambda \ge 0,\ q(\bar y)^T\lambda = 0,\ v^T\nabla^2(\lambda^T q)(\bar y)v \ge 0 \ \Rightarrow\ \lambda = 0,$$

then M is metrically subregular at $(\bar y, 0)$.

Example

$$\Gamma = \left\{ y \in \mathbb{R}^2 \,\middle|\, \begin{array}{l} -y_1^2 + y_2 \le 0 \\ -y_1^2 - y_2 \le 0 \\ y_1 \le 0 \end{array} \right\}$$

MFCQ is violated at $(0,0)$, but SOSCMS is fulfilled.



S-stationarity and M-stationarity

Recall that B-stationarity reads as

$$0 \in \nabla f(\bar{z}) + \hat{N}_{\psi^{-1}(0)}(\bar{z}).$$

Assumption: We have the representation

$$\psi^{-1}(0) = \{z \mid G(z) \in Q\}$$

where $G : \mathbb{R}^d \to \mathbb{R}^p$ smooth and $Q \subset \mathbb{R}^p$ closed.
Then we always have

$$\nabla G(\bar{z})^T \hat{N}_Q(G(\bar{z})) \subset \hat{N}_{\psi^{-1}(0)}(\bar{z})$$

and, if $G(\cdot) - Q$ is metrically subregular at $(\bar{z}, 0)$,

$$\hat{N}_{\psi^{-1}(0)}(\bar{z}) \subset \nabla G(\bar{z})^T N_{T_Q(G(\bar{z}))}(0).$$

Definition

1. $\bar{z}$ is called to be *S-stationary* if

$$0 \in \nabla f(\bar{z}) + \nabla G(\bar{z})^T \hat{N}_Q(G(\bar{z}))$$

2. $\bar{z}$ is called to be *M-stationary* if

$$0 \in \nabla f(\bar{z}) + \nabla G(\bar{z})^T N_{T_Q(G(\bar{z}))}(0)$$

- A S-stationary point is always B-stationary, but a local minimizer needs not to be S-stationary. An extra condition is required, e.g. that $\nabla G(\bar{z})$ is surjective.
- Under metric subregularity, a local minimizer is always M-stationary.
- A M-stationary point needs not to be B-stationary, i.e., a feasible descent direction can exist.

Theorem (Gfr./Outrata 2014)

Assume that $G(\bar{z}) \in Q$, $G(\cdot) - Q$ is metrically subregular at $(\bar{z}, 0)$. If there exists a subspace L such that

$$T_Q(G(\bar{z})) + L \subset T_Q(G(\bar{z}))$$

and

$$\nabla G(\bar{z})\mathbb{R}^d + L = \mathbb{R}^p$$

then

$$\hat{N}_{G^{-1}(Q)}(\bar{z}) = \nabla G(\bar{z})^T \hat{N}_Q(G(\bar{z}))$$

First order approach

$$(FOP) \qquad \min_{(x,y)\in\mathbb{R}^n\times\mathbb{R}^m} \quad f(x,y)$$

$$\text{s.t.} \quad 0 \in \nabla_y \varphi(x,y) + \hat{N}_\Gamma(y)$$

$$x \in C$$

Constraints can be written in the form $0 \in \Psi(x,y)$ respectively $G(x,y) \in Q$ with

$$\Psi(x,y) := \begin{pmatrix} \nabla_y \varphi(x,y) + \hat{N}_\Gamma(y) \\ x - C \end{pmatrix},$$

$$G(x,y) = \begin{pmatrix} (y, -\nabla_y \varphi(x,y)) \\ x \end{pmatrix}, \quad Q = \operatorname{gph} \hat{N}_\Gamma \times C$$

FOSCMS

Let $0 \in \Psi(\bar{x}, \bar{y})$, $\bar{y}^* := -\nabla_y \varphi(\bar{x}, \bar{y})$.
If there does not exist $(u, v) \neq (0, 0)$ and $w \neq 0$ satisfying

$$u \in T_C(\bar{x}),\ 0 \in \nabla^2_{xy}\varphi(\bar{x}, \bar{y})u + \nabla^2_{yy}\varphi(\bar{x}, \bar{y})v + D\hat{N}_\Gamma(\bar{y}, \bar{y}^*)(v),$$

$$-\nabla^2_{xy}\varphi(\bar{x}, \bar{y})^T w \in N_C(\bar{x}; u),$$

$$0 \in \nabla^2_{yy}\varphi(\bar{x}, \bar{y})w + D^*\hat{N}_\Gamma((\bar{y}, \bar{y}^*); (v, -\nabla^2_{xy}\varphi(\bar{x}, \bar{y})u - \nabla^2_{yy}\varphi(\bar{x}, \bar{y})v))(w),$$

then the multifunction Ψ is metrically subregular at $((\bar{x}, \bar{y}), 0))$.

In Gfr./Outrata 2014 explicit formulas for the graphical derivative and the regular coderivative of $\hat{N}_\Gamma$ under a weak constraint qualification for the lower level problem (SOSCMS) were given.

Notation: $\mathcal{L}(x,y,\lambda) = \varphi(x,y) + \lambda^T q(y)$

$$K(\bar{y},\bar{y}^*) = \{v \in T_\Gamma^{\rm lin}(\bar{y}) \mid \bar{y}^{*T} v = 0\}$$

$$\Lambda(\bar{y},\bar{y}^*) := \{\lambda \in N_{\mathbb{R}^l_-}(q(\bar{y})) \mid \nabla q(\bar{y})^T \lambda = \bar{y}^*\},$$

$$\Lambda(\bar{y},\bar{y}^*;v) := \mathop{\arg\max}_{\lambda \in \Lambda(\bar{y},\bar{y}^*)} v^T \nabla^2(\lambda^T q)(\bar{y}) v$$

Theorem (Gfr.2014)

Let $0 \in \Psi(\bar{x},\bar{y})$, assume that SOSCMS holds for $q(\cdot) - \mathbb{R}^l_-$ at $\bar{y}$ and assume that there does not exist (u,v,λ,μ,w) satisfying

$$(0,0) \neq (u,v) \in T_C(\bar{x}) \times K(\bar{y},\bar{y}^*),\ w \neq 0,\ -\nabla^2_{xy}\varphi(\bar{x},\bar{y})^T w \in N_C(\bar{x};u),$$
$$\lambda \in \Lambda(\bar{y},\bar{y}^*;v),\ \mu \in T_{N_{\mathbb{R}^l_-}(q(\bar{y}))}(\lambda),\ \mu^T \nabla q(\bar{y}) v = 0,$$
$$0 = \nabla^2_{xy}\varphi(\bar{x},\bar{y})u + \nabla^2_{yy}\mathcal{L}(\bar{x},\bar{y},\lambda)v + \nabla q(\bar{y})^T \mu,$$
$$\nabla q_i(\bar{y}) w = 0, \forall i : \lambda_i > 0 \vee \mu_i > 0$$
$$w^T \nabla^2_{yy}\mathcal{L}(\bar{x},\bar{y},\lambda) w \leq 0.$$

Then Ψ is metrically subregular at $((\bar{x},\bar{y}),0)$.

S-stationarity conditions

Theorem (Gfr./Outrata 2014)

Let $(\bar{x}, \bar{y})$ be a local minimum for (FOP), assume that SOSCMS holds for $q(\cdot) - \mathbb{R}^l_-$ at $\bar{y}$ and that Ψ is metrically subregular at $((\bar{x}, \bar{y}), 0)$. If there is a subspace L with $T_C(\bar{x}) + L \subset T_C(\bar{x})$ and a multiplier $\lambda \in \Lambda(\bar{y}, \bar{y}^*)$ such that

$$\nabla^2_{xy}\varphi(\bar{x}, \bar{y})L + \mathrm{span}\,\{\nabla q_i(\bar{y}) \mid i \in I^+(\lambda)\} = \mathbb{R}^m,$$

then there is a multiplier w such that the S-stationarity conditions

$$\begin{aligned} 0 &\in \nabla_x f(\bar{x}, \bar{y}) + \nabla^2_{xy}\varphi(\bar{x}, \bar{y})^* w + \hat{N}_C(\bar{x}) \\ 0 &\in \nabla_y f(\bar{x}, \bar{y}) + \nabla^2_{yy}\varphi(\bar{x}, \bar{y})^* w + \hat{D}^* \hat{N}_\Gamma(\bar{y}, \bar{y}^*)(w) \end{aligned}$$

hold.

Value function approach

$$
\begin{array}{lll}
(VFP) & \min\limits_{(x,y)\in\mathbb{R}^n\times\mathbb{R}^m} & f(x,y) \\
& \text{s.t.} & \varphi(x,y) \leq V(x), \\
& & 0 \in \nabla_y \varphi(x,y) + \hat{N}_\Gamma(y), \\
& & x \in C,
\end{array}
$$

- The constraint $\varphi(x,y) \leq V(x)$ introduces some redundancy and the multifunction associated with the constraints is never metrically regular. However we can give sufficient conditions for metric subregularity.
- Then we can replace (VFP) by the problem

$$
\begin{array}{ll}
\min\limits_{(x,y)\in\mathbb{R}^n\times\mathbb{R}^m} & f(x,y) + \sigma(\varphi(x,y) - V(x)) \\
\text{s.t.} & 0 \in \nabla_y \varphi(x,y) + \hat{N}_\Gamma(y), \\
& x \in C,
\end{array}
$$

where the penalty parameter σ is chosen sufficiently large (but finite).

$$\hat{\Psi}(x,y) := \begin{pmatrix} \varphi(x,y) - V(x) - \mathbb{R}_- \\ \nabla_y \varphi(x,y) + \hat{N}_\Gamma(y) \\ x - C \end{pmatrix}$$

Theorem

Let $0 \in \hat{\Psi}(\bar{x},\bar{y})$, $\bar{y}^* := -\nabla_y \varphi(\bar{x},\bar{y})$, $C = \{x \mid h_i(x) \le 0, i = 1,\ldots,p\}$, *where* $h_i \in C^1$. *Assume that there is a compact set* $\Omega \subset \mathbb{R}^m$ *and a neighborhood* U *of* $\bar{x}$ *such that* $S(x) \subset \Omega \; \forall x \in U$ *and assume that* $q(\cdot) - \mathbb{R}^l_-$ *fulfills SOSCMS at* $\bar{y}$. *If there is a direction* $u \in \mathbb{R}^n$ *satisfying*

$$\nabla h_i(\bar{x})u < 0, \; i : h_i(\bar{x}) = 0,$$

$$\nabla_x \varphi(\bar{x},\bar{y})u < \nabla_x \varphi(\bar{x},y)u \; \forall y \in S(\bar{x}), y \neq \bar{y}$$

and for every critical direction $0 \neq v \in K(\bar{y},\bar{y}^*)$, *every extreme point* λ *of* $\Lambda(\bar{y},\bar{y}^*;v)$ *and every* $w \neq 0$ *with* $\nabla q_i(\bar{y})w = 0, \forall i : \lambda_i > 0$ *one has*

$$w^T \mathcal{L}(\bar{x},\bar{y},\lambda)w > 0,$$

then $\hat{\Psi}$ *is metrically subregular at* $((\bar{x},\bar{y}),0)$.

Some References

- H. GFRERER, *First order and second order characterizations of metric subregularity and calmness of constraint set mappings*, SIAM J. Optim., 21 (2011), pp. 1439–1474.
- H. GFRERER, *On directional metric regularity, subregularity and optimality conditions for nonsmooth mathematical programs*, Set-Valued Var. Anal., 21(2013), pp. 151–176.
- H. GFRERER, J. V. OUTRATA, *On computation of generalized derivatives of the normal-cone mapping and their applications*, 2014, `www.numa.uni-linz.ac.at`