

# NONSMOOTH ANALYSIS AND OPTIMIZATION

COMPACT COURSE, LOTHAR COLLATZ SCHOOL, MAY 2013

Christian Clason

May 14, 2013

Institute for Mathematics and Scientific Computing  
Karl-Franzens-Universität Graz

# CONTENTS

---

|     |  |    |
|-----|--|----|
| 1   | CONVEX ANALYSIS                                  | 3  |
| 1.1 | Convex subdifferentials                          | 3  |
| 1.2 | Convex conjugates                                | 5  |
| 1.3 | Fenchel duality                                  | 7  |
| 2   | SEMISMOOTH NEWTON METHODS                        | 9  |
| 2.1 | Semismooth Newton methods in finite dimensions   | 10 |
| 2.2 | Semismooth Newton methods in infinite dimensions | 11 |

# PREFACE

---

The task of finding a minimizer  $\bar{u}$  of a Fréchet differentiable functional  $J$  can often be reduced to solving the first order necessary optimality conditions  $J'(\bar{u}) = 0$ , which is sometimes referred to as *Fermat's principle*. If  $J$  is non-differentiable but convex, the convex subdifferential  $\partial J$  replaces the nonexistent Fréchet derivative, as it satisfies Fermat's principle and allows for a rich calculus – in particular Fenchel duality – that can be used to obtain explicit optimality conditions.

In the (twice) differentiable case, the numerical computation of a minimizer  $\bar{u}$  thus amounts to solving the (often nonlinear) equation  $J'(\bar{u}) = 0$ , for which Newton's method is an efficient choice. The convex analogue  $0 \in \partial J(\bar{u})$  can often be reformulated as a non-differentiable but locally Lipschitz continuous equation; here, the so-called Newton derivative is a suitable replacement for the Fréchet derivative, as it allows a similarly efficient *semismooth Newton* method.

The purpose of these notes is to give a brief, “bird's-eye”, overview over the concepts and most important results on these topics. The classical reference for convex analysis and duality is [Ekeland and Témam 1999], while [Attouch, Buttazzo, and Michaille 2006, Chapter 9] contains a readable and complete overview. A rigorous and extensive treatment can be found in the excellent textbook [Schrotzek 2007], which we follow here. For semismooth Newton methods, the reader is referred to the expositions in [Ito and Kunisch 2008; Ulbrich 2011; Schiela 2008].

## NOTATION

For a normed vector space  $V$ , we denote by  $V^*$  the topological dual of  $V$ . Note that this definition depends on the choice of the topology, specified via the *duality pairing*  $\langle \cdot, \cdot \rangle_{V, V^*}$  between  $V$  and  $V^*$  (i.e.,  $(V, V^*, \langle \cdot, \cdot \rangle_{V, V^*})$  is a *dual pair*; see, e.g., [Werner 2011, Chapter VIII.3]). The topological dual  $V^*$  is always a Banach space if equipped with the norm

$$\|v^*\|_{V^*} = \sup \{ \langle v, v^* \rangle_{V, V^*} : v \in V, \|v\|_V \leq 1 \}.$$

For (non-reflexive) Banach spaces, two different topologies are of particular relevance.

- (i) The *weak topology* corresponds to the duality pairing between a normed vector space  $V$  and the space  $V^*$  of all continuous linear forms on  $V$ , defined by

$$\langle v, v^* \rangle_{V, V^*} := v^*(v)$$

for all  $v \in V$  and  $v^* \in V^*$ . For example, the weak topological dual of  $L^1(\Omega)$  can be identified with  $L^\infty(\Omega)$ , with the duality pairing reducing to

$$\langle v, v^* \rangle_{L^1, L^\infty} = \int_{\Omega} v(x)v^*(x) dx,$$

see, e.g., [Brezis 2010, Theorem 4.14]. If not specified otherwise, the topological dual is to be understood with respect to the weak topology.

- (ii) If  $V^*$  is the weak topological dual of  $V$ , the duality pairing between  $V^*$  and  $V$  is defined by

$$\langle v^*, v \rangle_{V^*, V} := v^*(v)$$

for all  $v^* \in V^*$  and  $v \in V$ . This allows identifying the *weak- $\star$  topological dual* (or *predual*) of  $V^*$  with  $V$  (i.e, the weak- $\star$  dual of  $L^\infty(\Omega)$  is  $L^1(\Omega)$ ).

(For reflexive Banach spaces, of course, both notions coincide.)

For a linear operator  $A : X \rightarrow Y$  between the normed vector spaces  $X$  and  $Y$ , we call  $A^* : Y^* \rightarrow X^*$  the *adjoint operator* to  $A$  if

$$\langle x, A^*y^* \rangle_{X, X^*} = \langle Ax, y^* \rangle_{Y, Y^*}$$

for all  $x \in X$  and  $y^* \in Y^*$ . If the duality is taken with respect to the weak topology, this coincides again with the standard definition. On the other hand, if there exists  $B : Y \rightarrow X$  such that  $B^* = A$  with respect to the weak topology, we can identify the *weak- $\star$  adjoint* (or *preadjoint*)  $A^*$  of an operator  $A : X^* \rightarrow Y^*$  with  $B$ , since

$$\langle x^*, By \rangle_{X^*, X} = \langle Ax^*, y \rangle_{Y^*, Y}$$

for all  $x^* \in X^*$  and  $y \in Y$ .

# CONVEX ANALYSIS

---

Recall that a function  $f : V \rightarrow \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$  is called *convex* if

$$f(\lambda v_1 + (1 - \lambda)v_2) \leq \lambda f(v_1) + (1 - \lambda)f(v_2)$$

for all  $v_1, v_2 \in V$  and  $\lambda \in (0, 1)$ , and *proper* if  $f$  is not identically equal to  $+\infty$ . For example, the *indicator function*  $\delta_C$  of a nonempty, convex set  $C \subset V$  (i.e.,  $\lambda v_1 + (1 - \lambda)v_2 \in C$  for all  $v_1, v_2 \in C$  and  $\lambda \in (0, 1)$ ), defined by

$$\delta_C(v) := \begin{cases} 0 & \text{if } v \in C, \\ \infty & \text{otherwise,} \end{cases}$$

is convex and proper. Similarly, for any Banach space  $V$ , the norm  $\|\cdot\|_V$  is convex and proper (by virtue of the triangle inequality).

## 1.1 CONVEX SUBDIFFERENTIALS

Let  $f : V \rightarrow \overline{\mathbb{R}}$  be convex and proper, and let  $\bar{v} \in V$  with  $f(\bar{v}) < \infty$ . The set

$$(1.1) \quad \partial f(\bar{v}) := \{v^* \in V^* : \langle v - \bar{v}, v^* \rangle_{V, V^*} \leq f(v) - f(\bar{v}) \text{ for all } v \in V\}$$

is called *subdifferential* of  $f$  at  $\bar{v}$ . Every  $v^* \in \partial f(\bar{v})$  is called *subgradient* of  $f$  at  $\bar{v}$ . From the definition (1.1), we immediately obtain Fermat's principle for convex functions: The point  $\bar{v}$  is a minimizer of  $f$  if and only if  $f(\bar{v}) \leq f(v)$  for all  $v \in V$ , which is equivalent to  $0 \in \partial f(\bar{v})$ .

The usefulness of the convex subdifferential now lies in the fact that it can often be characterized explicitly. If  $f$  is convex, proper, and Gâteaux differentiable, then  $\partial f(v) = \{f'(v)\}$ ; see, e.g., [Schrotzek 2007, Proposition 4.1.8].

To give another example, we return to the indicator function of a convex set  $C$ . For  $\bar{v} \in C$ , we have

$$\begin{aligned} v^* \in \partial \delta_C(\bar{v}) &\Leftrightarrow \langle v - \bar{v}, v^* \rangle_{V, V^*} \leq \delta_C(v) && \text{for all } v \in V \\ &\Leftrightarrow \langle v - \bar{v}, v^* \rangle_{V, V^*} \leq 0 && \text{for all } v \in C, \end{aligned}$$

since the condition is trivially satisfied for all  $v \notin C$ . In other words, the subdifferential of the indicator function of a convex set is its normal cone. Of particular importance for us will be the case when the set  $C_\alpha$  for  $\alpha > 0$  is given by pointwise constraints,

$$C_\alpha = \{v \in C_0(\Omega) : -\alpha \leq v(x) \leq \alpha \text{ for all } x \in \Omega\},$$

where we can give a pointwise characterization of the subdifferential. By separate pointwise inspection of the

- *positive active set:*  $x \in \mathcal{A}^+ := \{x \in \Omega : \bar{v}(x) = \alpha\}$ ,
- *negative active set:*  $x \in \mathcal{A}^- := \{x \in \Omega : \bar{v}(x) = -\alpha\}$ ,
- *inactive set:*  $x \in \mathcal{J} := \{x \in \Omega : |\bar{v}(x)| < \alpha\}$ ,

we obtain the equivalent *complementarity conditions* for  $v^* \in \partial\delta_{C_\alpha}(\bar{v}) \subset \mathcal{M}(\Omega) = C_0(\Omega)^*$ :

$$v^*(\mathcal{A}^+) \geq 0, \quad v^*(\mathcal{A}^-) \leq 0, \quad v^*(\mathcal{J}) = 0.$$

The complementarity conditions can equivalently be expressed for any  $\gamma > 0$  as

$$(1.2) \quad v^* + \max(0, -v^* + \gamma(\bar{v} - \alpha)) + \min(0, -v^* + \gamma(\bar{v} + \alpha)) = 0,$$

where max and min are taken pointwise in  $\Omega$  (in the sense of Jordan decomposition of measures); this can again be seen by pointwise inspection.

Another relevant example is the subdifferential of the norm of  $V$ , which is given by

$$\partial(\|\cdot\|_V)(\bar{v}) = \begin{cases} \{v^* \in V^* : \langle \bar{v}, v^* \rangle_{V, V^*} = \|\bar{v}\|_V \text{ and } \|v^*\|_{V^*} = 1\} & \text{if } \bar{v} \neq 0, \\ \{v^* \in V^* : \|v^*\|_{V^*} = 1\} =: B_{V^*} & \text{if } \bar{v} = 0. \end{cases}$$

To see this, first let  $\bar{v} \neq 0$  and consider  $v^* \in \partial(\|\cdot\|_V)(\bar{v})$ . Then we have by inserting in turn  $0 \in V$  and  $2\bar{v} \in V$  into (1.1) that

$$\|\bar{v}\|_V \leq \langle \bar{v}, v^* \rangle_{V, V^*} = \langle 2\bar{v} - \bar{v}, v^* \rangle_{V, V^*} \leq \|2\bar{v}\|_V - \|\bar{v}\|_V = \|\bar{v}\|_V.$$

Similarly, we have for any  $v \in V$

$$\langle v, v^* \rangle_{V, V^*} = \langle (v + \bar{v}) - \bar{v}, v^* \rangle_{V, V^*} \leq \|v + \bar{v}\|_V - \|\bar{v}\|_V \leq \|v\|_V,$$

which implies that  $\|v^*\|_{V^*} \leq 1$ ; since  $\tilde{v} = \bar{v} / \|\bar{v}\|_V$  satisfies

$$\langle \tilde{v}, v^* \rangle_{V, V^*} = \|\bar{v}\|_V^{-1} \|\bar{v}\|_V = 1,$$

we conclude  $\|v^*\|_{V^*} = 1$ . Conversely, let  $v^* \in V^*$  satisfy  $\langle \bar{v}, v^* \rangle_{V, V^*} = \|\bar{v}\|_V$  and  $\|v^*\|_{V^*} = 1$ . Then, for any  $v \in V$  we have

$$\langle v - \bar{v}, v^* \rangle_{V, V^*} = \langle v, v^* \rangle_{V, V^*} - \langle \bar{v}, v^* \rangle_{V, V^*} \leq \|v\|_V - \|\bar{v}\|_V,$$

i.e.,  $v^* \in \partial(\|\cdot\|_V)(\bar{v})$ . If  $\bar{v} = 0$ , definition (1.1) reduces to

$$\langle v, v^* \rangle_{V, V^*} \leq \|v\|_V \quad \text{for all } v \in V,$$

which by the definition of the dual norm is equivalent to  $\|v^*\|_{V^*} \leq 1$ .

For  $V = L^1(\Omega)$ , we can use this definition and pointwise inspection to explicitly compute  $v^* \in L^\infty(\Omega)$  for given  $v$  to obtain

$$\partial(\|\cdot\|_{L^1})(v) = \text{sign}(v) := \begin{cases} 1 & \text{if } v(x) > 0, \\ -1 & \text{if } v(x) < 0, \\ t \in [-1, 1] & \text{if } v(x) = 0 \end{cases}$$

The convex subdifferential satisfies the following sum rule. Let  $f_1, f_2 : V \rightarrow \bar{\mathbb{R}}$  be convex and proper. If there exists a point  $\bar{v} \in V$  such that  $f_1(\bar{v}), f_2(\bar{v}) < \infty$  and  $f_2$  is continuous at  $\bar{v}$  (which allows application of separation theorems), then

$$(1.3) \quad \partial(f_1 + f_2)(v) = \partial f_1(v) + \partial f_2(v)$$

for all  $v \in V$  for which  $f_1$  and  $f_2$  are finite; see, e.g., [Schrotzek 2007, Proposition 4.5.1]. Similarly, if  $T : U \rightarrow V$  is a continuous linear operator,  $f : V \rightarrow \bar{\mathbb{R}}$  is convex, and there exists a  $u_0 \in U$  such that  $f$  is continuous at  $Tu_0$ , then  $g : U \rightarrow \bar{\mathbb{R}}, g(u) = f(Tu)$ , is convex and

$$(1.4) \quad \partial g(u) = T^* \partial f(Tu) = \{T^* v^* : v^* \in \partial f(Tu)\},$$

see [Clarke 2013, Theorem 4.13].

## 1.2 CONVEX CONJUGATES

One reason for the usefulness of convex subdifferentials in our context is their connection with the Legendre–Fenchel transform. For a function  $f : V \rightarrow \bar{\mathbb{R}}$ , the *Fenchel conjugate* (or *convex conjugate*) is defined as

$$(1.5) \quad f^* : V^* \rightarrow \bar{\mathbb{R}}, \quad f^*(v^*) = \sup_{v \in V} \langle v, v^* \rangle_{V, V^*} - f(v).$$

The convex conjugate is always convex and lower semicontinuous. If  $f$  is convex and proper, then  $f^*$  is proper as well; see, e.g., [Schrotzek 2007, Proposition 2.2.3]. We also introduce the *biconjugate* of  $f$ , defined as

$$f^{**} : V \rightarrow \bar{\mathbb{R}}, \quad f^{**}(v) = \sup_{v^* \in V^*} \langle v^*, v \rangle_{V^*, V} - f^*(v^*)$$

(i.e., if  $V^*$  is the weak dual of  $V$ , we take  $V$  as the weak- $\star$  dual of  $V^*$  (or vice versa) and set  $f^{**} = (f^*)^*$ ). If  $f$  is proper, the Fenchel–Moreau–Rockafellar theorem states that  $f^{**} = f$  if and only if  $f$  is convex and lower semicontinuous; see, e.g., [Schrotzek 2007, Theorem 2.2.4].

We give a few relevant examples.

- (i) Let  $V = L^2(\Omega)$  and  $f(v) = \frac{1}{2} \|v\|_{L^2}^2$ . We identify  $V^*$  with  $V$  (i.e., the duality pairing is the inner product in  $L^2(\Omega)$ ). Then, the function to be maximized in (1.5) is strictly concave and differentiable, so that the supremum is attained if and only if  $v^* = f'(v) = v$ . Inserting this into the definition and simplifying, we obtain

$$f^* : L^2(\Omega) \rightarrow \mathbb{R}, \quad f^*(v^*) = \frac{1}{2} \|v^*\|_{L^2}^2.$$

- (ii) Let  $V$  be a normed vector space and  $f(v) = \delta_{B_V}(v)$ . We take  $V^*$  as the weak (or weak- $\star$ ) dual of  $V$  and compute  $f^*(v^*)$  for  $v^* \in V^*$ :

$$\delta_{B_V}^*(v^*) = \sup_{v \in V} \langle v, v^* \rangle_{V, V^*} - \delta_{B_V}(v) = \sup_{\|v\|_V \leq 1} \langle v, v^* \rangle_{V, V^*} = \|v^*\|_{V^*}.$$

- (iii) Let  $V$  be as above,  $V^*$  its weak topological dual and  $f(v) = \|v\|_V$ . We compute  $f^*(v^*)$  for given  $v^* \in V^*$  by discerning two cases:

- a)  $\|v^*\|_{V^*} \leq 1$ . In this case,  $\langle v, v^* \rangle_{V, V^*} \leq \|v\|_V \|v^*\|_{V^*} \leq \|v\|_V$  for all  $v \in V$  and  $\langle 0, v^* \rangle_{V, V^*} = 0 = \|0\|_V$ . Hence,

$$f^*(v^*) = \sup_{v \in V} \langle v, v^* \rangle_{V, V^*} - \|v\|_V = 0.$$

- b)  $\|v^*\|_{V^*} > 1$ . Then by the definition of the dual norm, there exists a  $v_0 \in V$  with  $\langle v_0, v^* \rangle_{V, V^*} > \|v_0\|_V$ . Taking  $\rho \rightarrow \infty$  in

$$0 < \rho (\langle v_0, v^* \rangle_{V, V^*} - \|v_0\|_V) = \langle \rho v_0, v^* \rangle_{V, V^*} - \|\rho v_0\|_V \leq f^*(v^*)$$

yields  $f^*(v^*) = +\infty$ .

We conclude that  $f^* = \delta_{B_{V^*}}$ .

If we take the dual with respect to the weak- $\star$  topology between  $V^*$  and  $V$ , this result also follows directly from the Fenchel–Moreau–Rockafellar theorem and example (ii) by noting that

$$\begin{aligned} \delta_{B_V}(v) &= \delta_{B_V}^{**}(v) = \sup_{v^* \in V^*} \langle v, v^* \rangle_{V, V^*} - \delta_{B_V}^*(v^*) \\ &= \sup_{v^* \in V^*} \langle v^*, v \rangle_{V^*, V} - \|v^*\|_{V^*} = f^*(v) \end{aligned}$$

for all  $v \in V$ .

Furthermore, straightforward calculation yields the following useful transformation rules; see, e.g., [Ekeland and Témam 1999, page 17]. For  $f : V \rightarrow \overline{\mathbb{R}}$ , we have for all  $\alpha \in \mathbb{R}$  and  $a \in V$  that

$$\begin{aligned} (\alpha f(\cdot))^*(v^*) &= \alpha f^*(\alpha^{-1}v^*), \\ f(\cdot - a)^*(v^*) &= f^*(v^*) + \langle a, v^* \rangle_{V, V^*}. \end{aligned}$$



### 1.3 FENCHEL DUALITY

We now discuss the relation between the Fenchel conjugate and the subdifferential of convex functions. Let  $f$  be a proper and convex function. Then we immediately obtain from the definitions of the conjugate and the subdifferential that for all  $v \in V$  with  $f(v) < \infty$  and all  $v^* \in V^*$  the *Fenchel–Young inequality*

$$(1.6) \quad \langle v, v^* \rangle_{V, V^*} \leq f(v) + f^*(v^*),$$

is satisfied. If  $v^* \in \partial f(v)$ , we have for all  $v \in V$  that

$$\langle v, v^* \rangle_{V, V^*} - f(v) \leq \langle \tilde{v}, v^* \rangle_{V, V^*} - f(\tilde{v}),$$

Hence, taking the supremum on the left-hand side yields

$$(1.7) \quad f^*(v^*) = \sup_{v \in V} \langle v, v^* \rangle_{V, V^*} - f(v) \leq \langle \tilde{v}, v^* \rangle_{V, V^*} - f(\tilde{v}).$$

Hence, inserting arbitrary  $w^* \in V^*$  into (1.6) and subtracting (1.7) for  $v^* \in \partial f(v)$  yields

$$\langle v, w^* - v^* \rangle_{V, V^*} \leq (f(v) + f^*(w^*)) - (f(v) + f^*(v^*)) = f^*(w^*) - f^*(v^*)$$

for every  $w^* \in V^*$ , i.e.,  $v \in \partial f^*(v^*)$ . If  $f$  is in addition lower semicontinuous, we can apply the Fenchel–Moreau–Rockafellar theorem to also obtain the converse, and thus

$$(1.8) \quad v^* \in \partial f(v) \quad \Leftrightarrow \quad v \in \partial f^*(v^*),$$

see [Schiretzek 2007, Proposition 4.4.4]. This relation allows passing from a subdifferential that is difficult to characterize to one that is more tractable.

The *Fenchel duality theorem* combines in a particularly elegant way the relation (1.8), the sum rule (1.3), and the chain rule (1.4) to obtain existence of and optimality conditions for a solution to a convex optimization problem. Let  $V$  and  $Y$  be Banach spaces,  $f : V \rightarrow \overline{\mathbb{R}}$ ,  $g : Y \rightarrow \overline{\mathbb{R}}$  be convex, proper, lower semicontinuous functions and  $T : V \rightarrow Y$  be a continuous linear operator. If there exists a  $v_0 \in V$  such that  $f(v_0) < \infty$ ,  $g(Tv_0) < \infty$ , and  $g$  is continuous at  $Tv_0$ , then

$$(1.9) \quad \inf_{v \in V} f(v) + g(Tv) = \sup_{y^* \in Y^*} -f^*(T^*y^*) - g^*(-y^*),$$

and the optimization problem on the right hand side (referred to as the *dual problem*) has at least one solution; see, e.g., [Ekeland and Témam 1999, Theorem III.4.1]. (Existence of a solution to the problem on the left hand side – the *primal problem* – follows directly from the assumptions on  $f$ ,  $g$ , and  $\Lambda$  by standard arguments.) Furthermore, the equality in (1.9) is attained at  $(\tilde{v}, \tilde{y}^*)$  if and only if the *extremality relations*

$$\begin{cases} T^*\tilde{y}^* \in \partial f(\tilde{v}), \\ -\tilde{y}^* \in \partial g(T\tilde{v}), \end{cases}$$

hold; see, e.g., [Ekeland and Témam 1999, Proposition III.4.1]. Depending on the context, one or both of these relations can be reformulated in terms of  $f^*$  and  $g^*$  using the equivalence (1.8). The conditions and consequences of the Fenchel duality theorem should be compared with classical regular point conditions (e.g., [Maurer and Zowe 1979; Ito and Kunisch 2008]) for the existence of Lagrange multipliers in constrained optimization.

We illustrate the above by applying Fenchel duality to derive optimality conditions for a classical optimal control problem with state constraints:

$$\left\{ \min_{y,u} \frac{1}{2} \|y - z\|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \|u\|_{L^2(\Omega)}^2 \text{ s. t. } Ay = u, \quad y(x) \geq 0 \text{ for all } x \in \Omega, \right.$$

where  $z \in L^2(\Omega)$  is given and the elliptic partial differential operator is such that the state  $y$  is continuous. Introducing the control-to-state mapping  $T : L^2(\Omega) \rightarrow C_0(\Omega)$ ,  $u \mapsto y$ , and the convex set  $B := \{v \in C_0(\Omega) : v(x) \geq 0 \text{ for all } x \in \Omega\}$ , we can formulate the reduced problem

$$\min_{u \in L^2(\Omega)} \frac{1}{2} \|Tu - z\|_{L^2(\Omega)}^2 + \frac{\alpha}{2} \|u\|_{L^2(\Omega)}^2 + \delta_B(Tu).$$

To apply Fenchel duality, we set

$$\begin{aligned} f : L^2(\Omega) &\rightarrow \mathbb{R}, & f(v) &= \frac{\alpha}{2} \|v\|_{L^2(\Omega)}^2, \\ g : C(\Omega) &\rightarrow \mathbb{R}, & g(v) &= \frac{1}{2} \|v - z\|_{L^2}^2 + \delta_B(v). \end{aligned}$$

Clearly,  $f$  and  $g$  are convex, proper (take  $v = 0$ ) and lower semicontinuous. We further assume that there exists a  $u_0 \in L^2(\Omega)$  such that  $Tu_0$  is in the interior of  $B$ , and hence  $\delta_B$  is continuous at  $Tu_0$  (this is the usual regular point condition). Then the Fenchel duality theorem yields the existence of a  $\bar{q} \in \mathcal{M}(\Omega)$  such that

$$\begin{cases} T^* \bar{q} \in \partial f(\bar{u}) &= \{\alpha \bar{u}\}, \\ -\bar{q} \in \partial g(T\bar{u}) &= \{T\bar{u} - z\} + \partial \delta_B(T\bar{u}), \end{cases}$$

where we have used that the sum rule holds due to the assumptions on  $u_0$ . Introducing now the optimal state  $\bar{y} = T\bar{u}$ , the adjoint state  $\bar{p} = -T^* \bar{q} \in L^2(\Omega)$  (i.e.,  $A^* \bar{p} = -\bar{q}$ ), and the subgradient  $\bar{\lambda} \in \partial \delta_B(T\bar{u}) \subset \mathcal{M}(\Omega)$ , we arrive at the usual optimality system

$$\begin{cases} A\bar{y} = \bar{u}, \\ A^* \bar{p} = \bar{y} - z + \bar{\lambda}, \\ \alpha \bar{u} = -\bar{p}, \\ \langle y - \bar{y}, \bar{\lambda} \rangle_{C_0, \mathcal{M}} \leq 0 \quad \text{for all } y \in B. \end{cases}$$

Control constraints can be treated in a similar way.

## SEMISMOOTH NEWTON METHODS

---

2

It remains to formulate a numerical method that can solve nonsmooth equations of the form (1.2) in an efficient manner. Just as the convex subdifferential proved to be suitable replacement for the Fréchet derivative in the context of optimality conditions, we need to consider a generalized derivative that can replace the Fréchet derivative in a Newton-type method and still allow superlinear convergence. In addition, it needs to provide a sufficiently rich calculus and the possibility for explicit characterization to be implementable in a numerical algorithm. These requirements lead to semismooth Newton methods.

To motivate the definitions, it will be instructive to first consider the convergence of an abstract generalized Newton method. Let Banach spaces  $U, V$ , a mapping  $F : U \rightarrow V$ , and  $u^* \in U$  with  $F(u^*) = 0$  be given. A generalized Newton method to compute an approximation of  $u^*$  can be described as follows:

- 1: Choose  $u^0 \in U$
- 2: **for**  $k = 0, 1, \dots$  **do**
- 3:     Choose an invertible linear operator  $M(x_k) \in \mathcal{L}(U, V)$
- 4:     Set  $u^{k+1} = u^k - M(u_k)^{-1}F(u^k)$

We can now ask ourselves when convergence of the iterates  $u^k \rightarrow u^*$  holds, and in particular when it is *superlinear*, i.e.,

$$(2.1) \quad \lim_{k \rightarrow \infty} \frac{\|u^{k+1} - u^*\|_U}{\|u^k - u^*\|_U} = 0.$$

Set  $d^k = u^k - u^*$ . Then we can use the definition of the Newton step and the fact that  $F(u^*) = 0$  to obtain

$$\begin{aligned} \|u^{k+1} - u^*\|_U &= \|u^k - M(u_k)^{-1}F(u^k) - u^*\|_U \\ &= \|M(u_k)^{-1} [F(u^k) - F(u^*) - M(u_k)(u^k - u^*)]\|_U \\ &= \|M(u_k)^{-1} [F(u^k) - F(u^*) - M(x_k)d^k]\|_U \\ &\leq \|M(u_k)^{-1}\|_{\mathcal{L}(V,U)} \|F(u^* + d^k) - F(u^*) - M(u^k)d^k\|_V \end{aligned}$$

Hence, (2.1) holds if both a

- *uniform regularity condition*: there exists a  $C > 0$  such that

$$\|M(\mathbf{u}_k)^{-1}\|_{\mathcal{L}(V,U)} \leq C$$

for all  $k$ , and an

- *approximation condition*:

$$\lim_{\|d^k\|_U \rightarrow 0} \frac{\|F(\mathbf{u}^* + d^k) - F(\mathbf{u}^*) - M(\mathbf{u}^* + d^k)d^k\|_V}{\|d^k\|_U} = 0,$$

hold. In this case, there exists a neighborhood  $N(\mathbf{u}^*)$  of  $\mathbf{u}^*$  such that

$$\|\mathbf{u}^{k+1} - \mathbf{u}^*\|_U < \frac{1}{2} \|\mathbf{u}^k - \mathbf{u}^*\|_U$$

for an  $\mathbf{u}^k \in N(\mathbf{u}^*)$ , which by induction implies  $d^k \rightarrow 0$  and hence the desired (local) superlinear convergence.

If  $F$  is continuously Fréchet differentiable, the approximation condition holds by definition for the Fréchet derivative  $M_k = F'(\mathbf{u}^k)$ , and we arrive at the classical Newton method. For nonsmooth  $F$ , we simply take a linear operator which satisfies the uniform regularity and approximation conditions. Naturally, the choice  $M_k \in \partial F(\mathbf{u}^k)$  for an appropriate generalized derivative suggests itself.

## 2.1 SEMISMOOTH NEWTON METHODS IN FINITE DIMENSIONS

If  $U$  and  $V$  are finite-dimensional, an appropriate choice is the Clarke subdifferential. Recall that by Rademacher's theorem, every Lipschitz function  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is differentiable almost everywhere; see, e.g. [Ziemer 1989, Theorem 2.2.1]. We can then define the *Clarke subdifferential* at  $x \in \mathbb{R}^n$  as

$$\partial_C F(x) = \text{co} \left\{ \lim_{n \rightarrow \infty} F'(x_n) : \{x_n\}_{n \in \mathbb{N}} \text{ with } x_n \rightarrow x, F \text{ differentiable at } x_n \right\},$$

where  $\text{co}$  denotes the convex hull. To use an element of the Clarke subdifferential as linear operator in our Newton method, we need to ensure in particular that the approximation condition holds. In fact, we will require a slightly stronger condition. A function  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is called *semismooth* at  $x \in \mathbb{R}^n$  if

- (i)  $F$  is Lipschitz continuous near  $x$ ,
- (ii)  $F$  is directionally differentiable at  $x$ ,
- (iii)  $\lim_{\|h\| \rightarrow 0} \sup_{M \in \partial_C F(x+h)} \frac{\|F(x+h) - F(x) - Mh\|}{\|h\|} = 0$ .

Note that we take the subgradient not in the linearization point but in a neighborhood, so we avoid evaluating  $\partial_C F$  at the points where  $F$  is not differentiable. This definition is equivalent to the original one of [Mifflin 1977] (for real-valued functions) and [Qi and Sun 1993] (for vector-valued functions); see [Ulbrich 2011, Proposition 2.7].

For a locally Lipschitz continuous function, this leads to the *semismooth Newton method*

- 1: Choose  $x^0 \in U$
- 2: **for**  $k = 0, 1, \dots$  **do**
- 3:     Choose  $M(x_k) \in \partial_C F(x^k)$
- 4:     Set  $x^{k+1} = x^k - M(x_k)^{-1} F x^k$

If  $f$  is semismooth at  $x^*$  with  $f(x^*) = 0$  and all  $M(x_k)$  satisfy the uniform regularity condition, this iteration converges (locally) superlinearly to  $x^*$ ; see, e.g., [Ulbrich 2011, Proposition 2.12]. A similar abstract framework for the superlinear convergence of Newton methods was proposed in [Kummer 1988].

We close this section with some relevant examples. Clearly, if  $F$  is continuously differentiable at  $x$ , then  $F$  is semismooth at  $x$  with  $\partial_C F(x) = \{F'(x)\}$ . This can be extended to continuous piecewise differentiable functions. Let  $F_1, \dots, F_N \in C^1(\mathbb{R}^n; \mathbb{R}^m)$  be given. A function  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is called *piecewise differentiable* if

$$F(x) \in \{F_1(x), \dots, F_N(x)\} \quad \text{for all } x \in \mathbb{R}^n.$$

Then,  $F$  is semismooth, and

$$\partial_C F(x) = \text{co}\{F'_i(x) : F(x) = F_i(x) \text{ and } x \in \text{cl int}\{y : F(y) = F_i(y)\}\};$$

see, e.g., [Ulbrich 2011, Proposition 2.26]. This means that we can differentiate piecewise, and where pieces overlap, take the convex hull of all possible values at  $x$  excluding those that are only attained on a null set containing  $x$ . As a concrete example, the function  $F : \mathbb{R} \rightarrow \mathbb{R}$ ,  $F(x) = \max(0, x)$  is semismooth, and

$$\partial_C F(x) = \begin{cases} \{0\} & \text{if } x < 0, \\ \{1\} & \text{if } x > 0, \\ [0, 1] & \text{if } x = 0. \end{cases}$$

Finally, a vector-valued function is semismooth if and only if all its component functions are semismooth; see [Ulbrich 2011, Proposition 2.10]. This implies semismoothness of (1.2) in finite dimensions.

## 2.2 SEMISMOOTH NEWTON METHODS IN INFINITE DIMENSIONS

In infinite dimensions, Rademacher's theorem is not available, and thus the construction above cannot be carried out. Instead of starting from Lipschitz continuous functions, we directly

demand the approximation condition to hold. We call  $F : U \rightarrow V$  *Newton differentiable* at  $u \in U$  if there exists a neighborhood  $N(u)$  and a mapping  $G : N(u) \rightarrow \mathcal{L}(U, V)$  with

$$(2.2) \quad \lim_{\|h\|_U \rightarrow 0} \frac{\|F(u+h) - F(u) - G(u+h)h\|_V}{\|h\|_U} = 0.$$

Any  $D_N F \in \{G(s) : s \in N(u)\}$  is then a *Newton derivative* at  $u$ . Note that Newton derivatives are in general not unique, and need not be elements of any generalized subdifferential. If  $F$  is Newton differentiable at  $u$  and

$$\lim_{t \rightarrow 0^+} G(u+th)h$$

exists uniformly in  $\|h\|_U = 1$ , then  $F$  is called *semismooth* at  $u$ . This approach to semismoothness in Banach spaces was proposed in [Hintermüller, Ito, and Kunisch 2002], based on the similar (but stronger) notion of slant differentiability introduced in [Chen, Nashed, and Qi 2000]. Related approaches to nonsmooth Newton methods in Banach spaces based on set-valued generalized derivatives were treated in [Kummer 2000] and [Ulbrich 2002]. The exposition here is adapted from [Ito and Kunisch 2008].

For Newton differentiable  $F$ , this definition leads to the semismooth Newton method

- 1: Choose  $u^0 \in U$
- 2: **for**  $k = 0, 1, \dots$  **do**
- 3:     Choose Newton derivative  $D_N F(u^k)$
- 4:     Set  $u^{k+1} = u^k - D_N F(u^k)^{-1} F(u^k)$

If  $F$  is Newton differentiable (in particular, if  $F$  is semismooth) at  $u^*$  with  $F(u^*) = 0$  and all  $D_N F(u) \in \{G(u) : u \in N(u^*)\}$  satisfy the uniform regularity condition  $\|D_N F(u)\|_{\mathcal{L}(V,U)} \leq C$ , this iteration converges (locally) superlinearly to  $u^*$ ; see, e.g., [Ito and Kunisch 2008, Theorem 8.16].

If we wish to apply a semismooth Newton method to a concrete function  $F$  such as the one in (1.2), we need to decide whether it is semismooth and give an explicit and computable Newton derivative. Clearly, if  $F$  is continuously Fréchet differentiable near  $u$ , then  $F$  is semismooth at  $u$ , and its Fréchet derivative  $F'(u)$  is a Newton derivative (albeit not the only one). However, this cannot be extended directly to “piecewise differentiable” functions such as the pointwise max operator acting on functions in  $L^p(\Omega)$ . It is instructive to consider a concrete example. Take  $F : L^p(\Omega) \rightarrow L^p(\Omega)$ ,  $F(u) = \max(0, u)$ . A candidate for its Newton derivative is defined by its action on  $h \in L^p(\Omega)$  as

$$[G(u)h](x) = \begin{cases} 0 & u(x) < 0 \\ h(x) & u(x) > 0 \\ \delta h(x) & u(x) = 0 \end{cases}$$

for almost all  $x \in \Omega$  and arbitrary  $\delta \in \mathbb{R}$ . (Since the Newton derivative coincides with the Fréchet derivative where  $F$  is continuously differentiable, we only have the freedom to choose

its value where  $u(x) = 0$ .) To show that the approximation condition (2.2) is violated at  $u(x) = -|x|$  on  $\Omega = (-1, 1)$  for any  $1 \leq p < \infty$ , we take the sequence

$$h_n(x) = \begin{cases} \frac{1}{n} & \text{if } |x| < \frac{1}{n}, \\ 0 & \text{otherwise,} \end{cases}$$

with  $\|h_n\|_{L^p}^p = \frac{2}{n^{p+1}}$ . Then, since  $[F(u)](x) = \max(0, -|x|) = 0$  almost everywhere, we have

$$[F(u + h_n) - F(u) - G(u + h_n)h_n](x) = \begin{cases} -|x| & \text{if } |x| < \frac{1}{n}, \\ 0 & \text{if } |x| > \frac{1}{n}, \\ -\frac{\delta}{n} & \text{if } |x| = \frac{1}{n} \end{cases}$$

and thus

$$\|F(u + h_n) - F(u) - G(u + h_n)h_n\|_{L^p}^p = \int_{-\frac{1}{n}}^{\frac{1}{n}} |x|^p dx = \frac{2}{p+1} \left(\frac{1}{n}\right)^{p+1}.$$

This implies

$$\lim_{n \rightarrow \infty} \frac{\|F(u + h_n) - F(u) - G(u + h_n)h_n\|_{L^p}}{\|h_n\|_{L^p}} = \left(\frac{1}{p+1}\right)^{\frac{1}{p}} \neq 0$$

and hence that  $F$  is not semismooth from  $L^p(\Omega)$  to  $L^p(\Omega)$ . A similar example can be constructed for  $p = \infty$ ; see, e.g., [Ito and Kunisch 2008, Example 8.14].

On the other hand, if we consider  $F : L^q(\Omega) \rightarrow L^p(\Omega)$  with  $q > p$ , the terms involving  $n^{-1}$  do not cancel and the approximation condition holds (at least for this choice of  $h_n$ ). In fact, for arbitrary  $h \in L^q(\Omega)$  one can use Hölder's inequality to create a term involving the Lebesgue measure of the support of the set where the “wrong” linearization is taken (i.e., where  $\max(u(x) + h(x)) \neq \max(u(x)) + G(u(x) + h(x))h(x)$ ), which can be shown to go to zero as  $h \rightarrow 0$ ; see [Hintermüller, Ito, and Kunisch 2002, Proposition 4.1]. Semismoothness in function spaces hence fundamentally requires a *norm gap*, which is why approximation may be necessary to apply a semismooth Newton method to equations of type (1.2).

The above holds for any pointwise defined operator. If  $\psi : \mathbb{R} \rightarrow \mathbb{R}$  is semismooth, the corresponding *Nemytskii operator*  $\Psi : L^q(\Omega) \rightarrow L^p(\Omega)$ , defined pointwise almost everywhere as

$$[\Psi(u)](x) := \psi(u(x)),$$

is semismooth if and only if  $1 \leq p < q \leq \infty$ , and a Newton derivative of  $\Psi$  at  $x$ , acting on  $h$ , can be taken as

$$[D_N(\Psi(u))h](x) \in \partial_C(\psi(u(x)))h(x),$$

see, e.g., [Ito and Kunisch 2008, Example 8.12]. This connection was first investigated systematically in [Ulbrich 2002]; an alternative approach which parallels the theory of Fréchet differentiability is followed in [Schiela 2008]. In particular,  $F(u) = \max(0, u)$  is semismooth from  $L^q(\Omega)$  to  $L^p(\Omega)$  for any  $q > p$ , with Newton derivative

$$[D_N F(u)h](x) = \begin{cases} 0 & u(x) \leq 0 \\ h(x) & u(x) > 0. \end{cases}$$

This can be conveniently expressed with the help of the *characteristic function*  $\chi_{\mathcal{A}}$  of the *active set*  $\mathcal{A} := \{x \in \Omega : u(x) > 0\}$  (i.e., the function taking the value 1 at  $x \in \mathcal{A}$  and 0 otherwise) as  $D_N F(u) = \chi_{\mathcal{A}}$ .

There is a useful calculus for Newton derivatives. It is straightforward to verify that the sum of two semismooth functions  $F_1$  and  $F_2$  is semismooth, and

$$D_N(F_1 + F_2)(u) := D_N F_1(u) + D_N F_2(u)$$

is a Newton derivative for any choice of Newton derivatives  $D_N F_1$  and  $D_N F_2$ . We also have a chain rule: If  $F : U \rightarrow V$  is continuously Fréchet differentiable at  $u \in U$  and  $G : V \rightarrow Z$  is Newton differentiable at  $F(u)$ , then  $H := G \circ F$  is Newton differentiable at  $u$  with Newton derivative

$$D_N H(u + h) = D_N G(F(u + h))F'(u + h)$$

for any  $h \in U$  sufficiently small; see [Ito and Kunisch 2008, Lemma 8.15].



## BIBLIOGRAPHY

---

- Attouch, H., Buttazzo, G., and Michaille, G. (2006). *Variational Analysis in Sobolev and BV Spaces*. Vol. 6. MPS/SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
- Brezis, H. (2010). *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, New York.
- Chen, X., Nashed, Z., and Qi, L. (2000). *Smoothing methods and semismooth methods for nondifferentiable operator equations*. SIAM J. Numer. Anal. 38.4, pp. 1200–1216. DOI: [10.1137/S0036142999356719](https://doi.org/10.1137/S0036142999356719).
- Clarke, F. (2013). *Functional Analysis, Calculus of Variations and Optimal Control*. Springer, London. DOI: [10.1007/978-1-4471-4820-3](https://doi.org/10.1007/978-1-4471-4820-3).
- Ekeland, I. and Témam, R. (1999). *Convex Analysis and Variational Problems*. Classics Appl. Math. 28. SIAM, Philadelphia.
- Hintermüller, M., Ito, K., and Kunisch, K. (2002). *The primal-dual active set strategy as a semismooth Newton method*. SIAM J. Optim. 13.3, 865–888 (2003). DOI: [10.1137/S1052623401383558](https://doi.org/10.1137/S1052623401383558).
- Ito, K. and Kunisch, K. (2008). *Lagrange Multiplier Approach to Variational Problems and Applications*. Vol. 15. Advances in Design and Control. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
- Kummer, B. (1988). *Newton's method for non-differentiable functions*. Mathematical Research 45, pp. 114–125.
- Kummer, B. (2000). *Generalized Newton and NCP-methods: Convergence, regularity, actions*. Discuss. Math. Differ. Incl. Control Optim. 20.2, pp. 209–244. DOI: [10.7151/dmdico.1013](https://doi.org/10.7151/dmdico.1013).
- Maurer, H. and Zowe, J. (1979). *First and second order necessary and sufficient optimality conditions for infinite-dimensional programming problems*. Math. Programming 16.1, pp. 98–110. DOI: [10.1007/BF01582096](https://doi.org/10.1007/BF01582096).
- Mifflin, R. (1977). *Semismooth and semiconvex functions in constrained optimization*. SIAM J. Control Optimization 15.6, pp. 959–972. DOI: [10.1137/0315061](https://doi.org/10.1137/0315061).
- Qi, L. Q. and Sun, J. (1993). *A nonsmooth version of Newton's method*. Math. Programming 58.3, Ser. A, pp. 353–367. DOI: [10.1007/BF01581275](https://doi.org/10.1007/BF01581275).

- Schiela, A. (2008). *A simplified approach to semismooth Newton methods in function space*. SIAM J. on Opt. 19.3, pp. 1417–1432. DOI: [10.1137/060674375](https://doi.org/10.1137/060674375).
- Schiotzke, W. (2007). *Nonsmooth Analysis*. Universitext. Springer, Berlin.
- Ulbrich, M. (2002). *Semismooth Newton methods for operator equations in function spaces*. SIAM J. Optim. 13.3, 805–842 (2003). DOI: [10.1137/S1052623400371569](https://doi.org/10.1137/S1052623400371569).
- Ulbrich, M. (2011). *Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces*. Vol. 11. MOS-SIAM Series on Optimization. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. DOI: [10.1137/1.9781611970692](https://doi.org/10.1137/1.9781611970692).
- Werner, D. (2011). *Funktionalanalysis*. 7th ed. Springer-Verlag, Berlin.
- Ziemer, W. P. (1989). *Weakly differentiable functions*. Vol. 120. Graduate Texts in Mathematics. Sobolev spaces and functions of bounded variation. Springer-Verlag, New York.