# NICHTGLATTE ANALYSIS UND OPTIMIERUNG

Vorlesungsskript, Wintersemester 2019/20

Christian Clason

Stand vom 30. Januar 2020

Fakultät für Mathematik
Universität Duisburg-Essen

# INHALTSVERZEICHNIS

# ÜBERBLICK

Die Optimierung beschäftigt sich mit Minimierungsproblemen der Form

$$\min_{x \in U} F(x)$$

für eine Funktion $F : U \to \mathbb{R}$. Konkret fragt man sich:

(i) Hat dieses Problem eine Lösung, d. h. existiert ein $\bar{x} \in U$ mit

$$F(\bar{x}) \le F(x) \qquad \text{für alle } x \in U?$$

(ii) Gibt es eine intrinsische Charakterisierung von $\bar{x}$, d. h. ohne Vergleich mit allen anderen $x \in U$?

(iii) Wie kann dieses $\bar{x}$ (effizient) berechnet werden?

Für $U \subset \mathbb{R}^n$ lässt sich das Vorgehen wie folgt skizzieren:

(i) Ist $U$ kompakt und ist $F$ stetig, so nimmt die Funktion $F$ nach dem Satz von Weierstraß ihr Minimum in $\bar{x} \in U$ an.

(ii) Ist $F$ differenzierbar, so gilt das *Fermat-Prinzip*

$$0 = F'(\bar{x}).$$

(iii) Ist $F$ stetig differenzierbar und $U$ offen, so kann man Gradientenverfahren verwenden: Wähle $x^0$ und setze für $k = 1, \ldots$

$$x^{k+1} = x^k - t_k F'(x^k)$$

mit geeigneter Schrittweite $t_k$, dann gilt $x^k \to \bar{x}$ für $k \to \infty$.

Ist $F$ sogar zweimal stetig differenzierbar, kann man das Newton-Verfahren für die Nullstelle von $F'$ anwenden: Wähle $x^0$ geeignet und setze für $k = 1, \ldots$

$$x^{k+1} = x^k - F''(x^k)^{-1} F'(x^k).$$

Nun gibt es zahlreiche praktisch relevante Funktionen, die *nicht* differenzierbar sind, wie etwa die Betrags- oder die Maximumsfunktion. In der nichtglatten Analysis sucht man daher nach verallgemeinerten Ableitungsbegriffen, die auch für solche Funktionen das obige Vorgehen erlauben und gleichzeitig genug Rechenregeln zulassen, um für eine große Klasse von Funktionen explizit angegeben werden können. In dieser Vorlesung konzentrieren wir uns auf die beiden Klassen

(i) der konvexen Funktionen und

(ii) der lokal Lipschitz-stetigen Funktionen,

die zusammen ein breites Spektrum an Anwendungen abdecken. Insbesondere bieten erstere eine Grundlage für verallgemeinerte Gradientenverfahren, letztere für verallgemeinerte Newton-Verfahren. Ziel ist dabei insbesondere, Funktionen zu behandeln von der Form

$$\min_{x\in C} \frac{1}{p}\|S(x)-z\|_Y^p + \frac{\alpha}{q}\|x\|_X^q$$

für eine konvexe Menge $C\subset X$, eine (eventuell nichtlineare aber differenzierbare) Abbildung $S: X\to Y$, $\alpha\geq 0$ und $p,q\in[1,\infty)$ (insbesondere $p=1$ und oder $q=1$). Solche Funktionen treten häufig auf als Regularisierung inverser und schlecht gestellter Probleme, in der mathematischen Bildverarbeitung, und in der optimalen Steuerung von Differentialgleichungen. Als weiterer Schwerpunkt betrachten wir die Optimierung daher in *unendlichdimensionalen* Funktionenräumen; die gesuchten Minimierer sind also Funktionen. Dadurch sind die vorgestellten Verfahren *diskretisierungsunabhängig*, können also auf jede (vernünftige) endlichdimensionale Approximation angewendet werden, ohne dass die Feinheit der Diskretisierung eine Rolle spielt.

Dieses Skriptum basiert vor allem auf den folgenden Werken:

[1] H. H. Bauschke & P. L. Combettes (2017), *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2. Aufl., CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, Springer, New York, doi: 10.1007/978-3-319-48311-5

[2] M. Brokate (2014), Konvexe Analysis und Evolutionsprobleme, Vorlesungsskript, Zentrum Mathematik, TU München, url: http://www-m6.ma.tum.de/~brokate/cev_ss14.pdf

[3] F. Clarke (2013), *Functional Analysis, Calculus of Variations and Optimal Control*, Springer, London, doi: 10.1007/978-1-4471-4820-3

[4] A. Schiela (2008), A simplified approach to semismooth Newton methods in function space, *SIAM J. Opt.* 19(3), 1417–1432, doi: 10.1137/060674375

[5] W. Schirotzek (2007), *Nonsmooth Analysis*, Universitext, Springer, Berlin, doi: 10.1007/978-3-540-71333-3

[6] M. Ulbrich (2011), *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*, Bd. 11, MOS-SIAM Series on Optimization, SIAM, Philadelphia, PA, doi: 10.1137/1.9781611970692

# Teil I

# GRUNDLAGEN

# 1 GRUNDLAGEN DER FUNKTIONALANALYSIS

In diesem Kapitel stellen wir die für diese Vorlesung wesentlichen Begriffe, Notationen und Resultate zusammen. Für Beweise wird auf die Standardliteratur verwiesen, z. B. auf [Werner 2011], sowie auf [Clason 2019].

## 1.1 NORMIERTE RÄUME

Im Folgenden bezeichne $X$ einen Vektorraum über dem Körper $\mathbb{K}$, wobei wir uns hier stets auf den Fall $\mathbb{K} = \mathbb{R}$ beschränken. Eine Abbildung $\|\cdot\| : X \to \mathbb{R}^+ := [0, \infty)$ heißt *Norm* (auf $X$), falls für alle $x \in X$ gilt

(i) $\|\lambda x\| = |\lambda|\|x\|$ für alle $\lambda \in \mathbb{K}$,

(ii) $\|x + y\| \leq \|x\| + \|y\|$ für alle $y \in X$,

(iii) $\|x\| = 0$ genau dann, wenn $x = 0 \in X$.

Beispiel 1.1. (i) Auf $X = \mathbb{R}^N$ werden Normen definiert durch

$$\begin{aligned} \|x\|_p &= \left(\sum_{i=1}^N |x_i|^p\right)^{1/p} \qquad 1 \leq p < \infty, \\ \|x\|_\infty &= \max_{i=1,\dots,N} |x_i|. \end{aligned}$$

(ii) Auf $X = \ell^p$ (dem Raum der reellen Folgen, auf dem folgende Ausdrücke endlich sind) sind Normen definiert durch

$$\begin{aligned} \|x\|_p &= \left(\sum_{i=1}^\infty |x_i|^p\right)^{1/p} \qquad 1 \leq p < \infty, \\ \|x\|_\infty &= \sup_{i=1,\dots,\infty} |x_i|. \end{aligned}$$

(iii) Auf $X = L^p(\Omega)$ (dem Raum der messbaren reellen Funktionen auf dem Gebiet $\Omega \subset \mathbb{R}^n$, auf dem folgende Ausdrücke endlich sind) sind Normen definiert durch

$$\|u\|_p = \left(\int_\Omega |u(x)|^p\right)^{1/p} \qquad 1 \le p < \infty,$$
$$\|u\|_\infty = \operatorname*{ess\,sup}_{x\in\Omega} |u(x)|.$$

(iv) Auf $X = C(\overline{\Omega})$ (dem Raum der stetigen Funktionen auf $\overline{\Omega}$) ist eine Norm definiert durch

$$\|u\|_C = \sup_{x\in\overline{\Omega}} |u(x)|.$$

Eine analoge Norm ist auf $X = C_0(\Omega)$ (dem Raum der stetigen Funktionen auf $\Omega$ mit kompaktem Träger) definiert, wenn das Supremum nur über $x \in \Omega$ genommen wird.

Ist $\|\cdot\|$ eine Norm auf $X$, so bezeichnet man das Paar $(X, \|\cdot\|)$ als *normierten Raum*, und schreibt in diesem Fall oft $\|\cdot\|_X$. Ist die Norm kanonisch (etwa in Beispiel 1.1 (ii)–(iv)), so wird sie oft weggelassen.

Zwei Normen $\|\cdot\|_1, \|\cdot\|_2$ heißen *äquivalent* auf $X$, falls $c_1, c_2 > 0$ existieren mit

$$c_1\|x\|_2 \le \|x\|_1 \le c_2\|x\|_2 \qquad \text{für alle } x \in X.$$

Ist $X$ endlichdimensional, so sind alle Normen auf $X$ äquivalent. Die Konstanten $c_1, c_2$ hängen dann jedoch von der Dimension $N$ von $X$ ab; die Vermeidung solcher dimensionsabhängiger Konstanten ist einer der Gründe, warum wir Optimierung in unendlichdimensionalen Funktionenräumen betreiben wollen.

Sind $(X, \|\cdot\|_X)$ und $(Y, \|\cdot\|_Y)$ normierte Räume mit $X \subset Y$, so heißt $X$ *stetig eingebettet* in $Y$, geschrieben $X \hookrightarrow Y$, falls ein $C > 0$ existiert mit

$$\|x\|_Y \le C\|x\|_X \qquad \text{für alle } x \in X.$$

Wir betrachten nun Abbildungen zwischen normierten Räumen. Seien im Folgenden stets $(X, \|\cdot\|_X)$ und $(Y, \|\cdot\|_Y)$ normierte Räume, $U \subset X$, und $F : U \to Y$ eine Abbildung. Wir bezeichnen mit

- $\operatorname{dom} F := U$ den *Definitionsbereich* (englisch „domain“) von $F$;
- $\ker F := \{x \in U : F(x) = 0\}$ den *Kern* (englisch „kernel“ oder „null space“) von $F$;
- $\operatorname{ran} F := \{F(x) \in Y : x \in U\}$ das *Bild* (englisch „range“) von $F$;
- $\operatorname{graph} F := \{(x, y) \in X \times Y : y = F(x)\}$ den *Graph* von $F$.

Wir sagen, $F$ ist

- *stetig* in $x \in U$, wenn für alle $\varepsilon > 0$ ein $\delta > 0$ existiert mit
$$\|F(x) - F(z)\|_Y \leq \varepsilon \qquad \text{für alle } z \in U \text{ mit } \|x - z\|_X \leq \delta;$$
- *Lipschitz-stetig*, wenn ein $L > 0$ existiert (genannt *Lipschitz-Konstante*) mit
$$\|F(x_1) - F(x_2)\|_Y \leq L\|x_1 - x_2\|_X \qquad \text{für alle } x_1, x_2 \in U.$$
- *lokal Lipschitz-stetig* in $x \in U$, wenn ein $\delta > 0$ und ein $L = L(x, \delta) > 0$ existiert mit
$$\|F(x) - F(z)\|_Y \leq L\|x - z\|_X \qquad \text{für alle } z \in U \text{ mit } \|x - z\|_X \leq \delta. \tag{1.1}$$

Ist $T : X \to Y$ linear, so ist die Stetigkeit äquivalent zu der Bedingung, dass eine Konstante $C > 0$ existiert mit
$$\|Tx\|_Y \leq C\|x\|_X \qquad \text{für alle } x \in X.$$
Stetige lineare Abbildungen nennt man daher auch *beschränkt*; man spricht auch von einem beschränkten linearen *Operator*. Der Raum $L(X, Y)$ der beschränkten linearen Operatoren ist ein normierter Raum versehen mit der *Operatornorm*
$$\|T\|_{L(X,Y)} = \sup_{x \in X \setminus \{0\}} \frac{\|Tx\|_Y}{\|x\|_X} = \sup_{\|x\|_X = 1} \|Tx\|_Y = \sup_{\|x\|_X \leq 1} \|Tx\|_Y$$
(die gleich der kleinstmöglichen Konstante $C$ in der Definition der Stetigkeit ist). Ist $T \in L(X, Y)$ bijektiv, dann ist die Inverse $T^{-1} : Y \to X$ stetig genau dann, wenn ein $c > 0$ existiert mit
$$c\|x\|_X \leq \|Tx\|_Y \qquad \text{für alle } x \in X.$$
In diesem Fall ist $\|T^{-1}\|_{L(Y,X)} = c^{-1}$ für die größtmögliche Wahl von $c$.

## 1.2 STARKE UND SCHWACHE KONVERGENZ

Eine Norm vermittelt auf direkte Weise einen Konvergenzbegriff, die sogenannte *starke Konvergenz*: Eine Folge $\{x_n\}_{n\in\mathbb{N}} \subset X$ konvergiert (stark in $X$) gegen ein $x \in X$, geschrieben $x_n \to x$, wenn gilt
$$\lim_{n\to\infty} \|x_n - x\|_X = 0.$$

Eine Teilmenge $U \subset X$ nennen wir

- *abgeschlossen*, falls für jede konvergente Folge $\{x_n\}_{n\in\mathbb{N}} \subset U$ auch der Grenzwert $x \in U$ liegt;

- *kompakt*, falls jede Folge $\{x_n\}_{n\in\mathbb{N}} \subset U$ eine konvergente Teilfolge $\{x_{n_k}\}_{k\in\mathbb{N}}$ besitzt, deren Grenzwert $x \in U$ liegt.

Eine Abbildung $F : X \to Y$ ist genau dann stetig, wenn aus $x_n \to x$ auch $F(x_n) \to F(x)$ folgt, und *abgeschlossen*, wenn für $x_n \to x$ und $F(x_n) \to y$ folgt, dass $F(x) = y$ ist (also graph $F$ eine abgeschlossene Menge ist).

Weiterhin definieren wir für späteren Gebrauch für $x \in X$ und $r > 0$

- die *offene Kugel* $O_r(x) := \{z \in X : \|x - z\|_X < r\}$ und
- die *abgeschlossene Kugel* $K_r(x) := \{z \in X : \|x - z\|_X \leq r\}$.

Die abgeschlossene Kugel um 0 mit Radius 1 bezeichnet man auch als *Einheitskugel* $B_X$ (englisch „unit ball"). Eine Menge $U \subset X$ heißt

- *offen*, falls für alle $x \in U$ ein $r > 0$ existiert mit $O_r(x) \subset U$ (d. h. alle $x \in U$ *innere Punkte* von $U$ sind, deren Menge wir als *Inneres* $U^o$ bezeichnen);
- *beschränkt*, falls sie in einer abgeschlossenen Kugel $K_r(0)$ für ein $r > 0$ enthalten ist;
- *konvex*, falls für $x, y \in U$ auch $\lambda x + (1 - \lambda)y \in U$ für alle $\lambda \in [0, 1]$ gilt.

In normierten Räumen gilt, dass das Komplement einer offenen Menge abgeschlossen ist und umgekehrt (d. h. die abgeschlossenen Mengen im Sinne der Topologie sind genau die (Folgen-)abgeschlossenen Mengen im Sinne unserer Definition). Sowohl offene als auch abgeschlossene Kugeln sind wegen der Norm-Axiome konvex.

Ein normierter Raum $X$ heißt *vollständig*, wenn jede Cauchy-Folge in $X$ konvergiert; man nennt $X$ dann auch *Banachraum*. Alle Räume in Beispiel 1.1 sind Banachräume. Ebenso ist $L(X, Y)$, versehen mit der Operatornorm, ein Banachraum, wenn $Y$ ein Banachraum ist. Für konvexe Teilmengen von Banachräumen gilt folgende nützliche Eigenschaft, die auf dem Satz von Baire beruht.

**Lemma 1.2.** *Sei $X$ ein Banachraum und $U \subset X$ abgeschlossen und konvex. Dann gilt*

$$U^o = \{x \in U : \textit{für alle } h \in X \textit{ existiert } \delta > 0 \textit{ mit } x + th \in U \textit{ für alle } t \in [0, \delta]\}.$$

Die Menge auf der rechten Seite wird auch als *algebraisches Inneres* oder englisch „*core*" bezeichnet, weshalb Lemma 1.2 auch manchmal „core-int-Lemma" genannt wird.

Von wesentlicher Bedeutung wird für uns der Spezialfall $Y = \mathbb{R}$ sein, das heißt der Raum $L(X, \mathbb{R})$ der *linearen stetigen Funktionale* auf $X$. In diesem Fall bezeichnet man $X^* := L(X, \mathbb{R})$ als *Dualraum* von $X$. Ist $x^* \in X^*$, so schreibt man auch

$$\langle x^*, x\rangle_X := x^*(x) \in \mathbb{R}.$$

Diese *duale Paarung* soll andeuten, dass man auch $x$ auf $x^*$ wirkend auffassen kann, was später wichtig sein wird. Aus der Definition der Operatornorm folgt sofort, dass

$$\langle x^*, x\rangle_X \leq \|x^*\|_{X^*}\|x\|_X \qquad \text{für alle } x \in X, x^* \in X^*. \tag{1.2}$$

In vielen Fällen kann der Dualraum eines Banachraums mit einem bekannten Banachraum identifiziert werden.

Beispiel 1.3. (i) $(\mathbb{R}^N, \|\cdot\|_p)^* \cong (\mathbb{R}^N, \|\cdot\|_q)$ mit $p^{-1}+q^{-1} = 1$, wobei $0^{-1} = \infty$ und $\infty^{-1} = 0$ gesetzt wird. Die duale Paarung ist gegeben durch

$$\langle x^*, x\rangle_p = \sum_{i=1}^N x_i^* x_i.$$

(ii) $(\ell^p)^* \cong (\ell^q)$ für $1 < p < \infty$. Die duale Paarung ist gegeben durch

$$\langle x^*, x\rangle_p = \sum_{i=1}^\infty x_i^* x_i.$$

Darüber hinaus ist $(\ell^1)^* = \ell^\infty$, aber $(\ell^\infty)^*$ ist selber kein Folgenraum.

(iii) Ebenso ist $L^p(\Omega)^* \cong L^q(\Omega)$ für $1 < p < \infty$. Die duale Paarung ist gegeben durch

$$\langle u^*, u\rangle_p = \int_\Omega u^*(x)u(x)\,dx.$$

Es gilt auch $L^1(\Omega)^* \cong L^\infty(\Omega)$, aber $L^\infty(\Omega)^*$ ist selber kein Funktionenraum.

(iv) $C_0(\Omega)^* \cong \mathcal{M}(\Omega)$, dem Raum der *Radon-Maße*; er enthält unter anderem das Lebesgue-Maß, aber auch Dirac-Maße $\delta_x$ für $x \in \Omega$, definiert durch $\delta_x(u) = u(x)$ für $u \in C_0(\Omega)$. Die duale Paarung ist gegeben durch

$$\langle u^*, u\rangle_C = \int_\Omega u(x)\,du^*.$$

Ein zentrales Resultat über Dualräume ist der Satz von Hahn–Banach, auf dem viele der folgenden Aussagen beruhen. Es gibt von ihm eine algebraische und eine geometrische Version.

Satz 1.4 (Hahn–Banach, algebraisch). *Sei $X$ ein normierter Raum. Zu jedem $x \in X$ existiert ein $x^* \in X^*$ mit*

$$\|x^*\|_{X^*} = 1 \qquad \text{und} \qquad \langle x^*, x\rangle_X = \|x\|_X.$$

Satz 1.5 (Hahn–Banach, geometrisch). *Seien $X$ ein normierter Raum und $A, B \subset X$ konvex, nichtleer und disjunkt.*

*(i) Ist $A$ offen, dann existiert ein $x^* \in X^*$ und ein $\lambda \in \mathbb{R}$ mit*

$$\langle x^*, x_1\rangle_X < \lambda \leq \langle x^*, x_2\rangle_X \qquad \text{für alle } x_1 \in A, x_2 \in B. \tag{1.3}$$

*(ii) Ist $A$ abgeschlossen und $B$ kompakt, dann existiert ein $x^* \in X^*$ und ein $\lambda \in \mathbb{R}$ mit*

$$\langle x^*, x_1\rangle_X \leq \lambda < \langle x^*, x_2\rangle_X \qquad \text{für alle } x_1 \in A, x_2 \in B. \tag{1.4}$$

Insbesondere die geometrische Version – auch als *Trennungssatz* bekannt – ist von zentraler Bedeutung für die konvexe Analysis; wir werden sie in der folgenden Variante benötigen, die als *Satz von Eidelheit* bekannt ist.

Folgerung 1.6. *Seien $X$ ein normierter Raum und $A, B \subset X$ konvex und nichtleer. Ist die Menge $A^o$ der inneren Punkte von $A$ nichtleer und disjunkt zu $B$, dann existiert ein $x^* \in X^* \setminus \{0\}$ und ein $\lambda \in \mathbb{R}$ mit*

$$\langle x^*, x_1\rangle_X \leq \lambda \leq \langle x^*, x_2\rangle_X \qquad \text{für alle } x_1 \in A, x_2 \in B. \tag{1.5}$$

*Beweis.* Satz 1.5 (i) liefert die Existenz von $x^*$ und $\lambda$, so dass die Aussage gilt für alle $x_1 \in A^o$ (sogar mit strikter Ungleichung, woraus auch $x^* \neq 0$ folgt). Es bleibt also zu zeigen, dass $\langle x^*, x\rangle_X \leq \lambda$ auch für $x \in A \setminus A^o$ gilt. Da $A^o$ nichtleer ist, existiert ein $x_0 \in A^o$, d. h. es existiert $r > 0$ mit $O_r(x_0) \subset A$. Aus der Konvexität von $A$ folgt dann, dass für alle $t \in [0, 1]$ und $\tilde{x} \in O_r(x_0)$ auch $t\tilde{x} + (1 - t)x \in A$ ist. Damit ist

$$tO_r(x_0) + (1 - t)x = O_{tr}(tx_0 + (1 - t)x) \subset A, \tag{1.6}$$

und insbesondere gilt $x(t) := tx_0 + (1 - t)x \in A^o$ für alle $t \in (0, 1)$.

Wir finden also eine Folge $\{x_n\}_{n\in\mathbb{N}} \subset A^o$ (zum Beispiel $x_n = x(n^{-1})$) mit $x_n \to x$. Aus der Stetigkeit von $x^* \in X = L(X, \mathbb{R})$ folgt mit Grenzübergang dann

$$\langle x^*, x\rangle_X = \lim_{n\to\infty} \langle x^*, x_n\rangle_X \leq \lambda. \qquad \square$$

Ein normierter Raum wird also in gewisser Weise durch seinen Dualraum charakterisiert. Als direkte Folgerung von Satz 1.4 erhalten wir, dass die Norm in einem Banachraum als Operatornorm dargestellt werden kann.

Folgerung 1.7. *Sei $X$ ein Banachraum. Dann gilt für alle $x \in X$*

$$\|x\|_X = \sup_{\|x^*\|_{X^*}\leq 1} |\langle x^*, x\rangle_X|,$$

*und das Supremum wird angenommen.*

Ein $x \in X$ können wir also auch als lineares und wegen (1.2) stetiges Funktional auf $X^*$ auffassen, also als Element im *Bidualraum* $X^{**} := (X^*)^*$. Die Einbettung $X \hookrightarrow X^{**}$ wird dabei vermittelt durch die *kanonische Injektion*

$$J : X \to X^{**}, \qquad \langle Jx, x^* \rangle_{X^*} := \langle x^*, x \rangle_X \quad \text{für alle } x^* \in X^*.$$

Offensichtlich ist $J$ linear; aus Satz 1.4 folgt weiterhin $\|Jx\|_{X^{**}} = \|x\|_X$. Ist die kanonische Injektion surjektiv – können wir also $X^{**}$ mit $X$ identifizieren – so nennt man $X$ *reflexiv*. Endlichdimensionale Räume sind reflexiv, sowie Beispiel 1.1 (ii) und (iii) für $1 < p < \infty$, nicht aber $\ell^1, \ell^\infty$, und $L^1(\Omega), L^\infty(\Omega)$ sowie $C(\overline{\Omega})$.

Durch die duale Paarung werden weitere Konvergenzbegriffe erzeugt: die *schwache* Konvergenz auf $X$ sowie die *schwach-∗* Konvergenz auf $X^*$.

(i) Eine Folge $\{x_n\}_{n\in\mathbb{N}} \subset X$ konvergiert schwach (in $X$) gegen $x \in X$, geschrieben $x_n \rightharpoonup x$, falls
$$\langle x^*, x_n \rangle_X \to \langle x^*, x \rangle_X \qquad \text{für alle } x^* \in X^*.$$

(ii) Eine Folge $\{x_n^*\}_{n\in\mathbb{N}} \subset X^*$ konvergiert schwach-∗ (in $X^*$) gegen $x^* \in X^*$, geschrieben $x_n^* \rightharpoonup^* x^*$, falls
$$\langle x_n^*, x \rangle_X \to \langle x^*, x \rangle_X \qquad \text{für alle } x \in X.$$

Die schwache Konvergenz verallgemeinert den Begriff der komponentenweisen Konvergenz in $\mathbb{R}^n$, der – wie aus dem Beweis des Satzes von Heine–Borel ersichtlich ist – im Kontext der Kompaktheit der wesentliche ist. Aus starker Konvergenz folgt schwache Konvergenz; ebenso folgt aus Konvergenz in der Operatornorm (auch *punktweise Konvergenz* genannt) die schwach-∗ Konvergenz. Ist $X$ reflexiv, so stimmen schwache und schwach-∗ Konvergenz (beide in $X = X^{**}$!) überein. In endlichdimensionalen Räumen stimmen alle Konvergenzbegriffe überein.

Konvergiert $x_n \to x$ und $x_n^* \rightharpoonup^* x^*$ oder $x_n \rightharpoonup x$ und $x_n^* \to x^*$, so gilt $\langle x_n^*, x_n \rangle_X \to \langle x^*, x \rangle_X$. Die duale Paarung aus schwach(-∗) konvergenten Folgen konvergiert aber in der Regel nicht!

Analog zur starken Konvergenz definiert man nun schwache(-∗) Stetigkeit und Abgeschlossenheit von Abbildungen sowie schwache(-∗) Abgeschlossenheit und Kompaktheit von Mengen. Letztere Eigenschaft wird für uns wesentlich sein; ihre Charakterisierung ist daher ein zentrales Resultat dieses Kapitels.

Satz 1.8 (Eberlein–Šmulyan). *Sei $X$ ein normierter Raum. Dann ist $B_X$ schwach kompakt genau dann, wenn $X$ reflexiv ist.*

In einem reflexiven Raum enthalten also insbesondere alle beschränkten Folgen eine schwach (aber im allgemeinen nicht stark) konvergente Teilfolge. Beachten Sie, dass schwache Abgeschlossenheit eine *stärkere* Forderung ist als Abgeschlossenheit. Für konvexe Mengen stimmen die Begriffe aber überein.

Lemma 1.9. *Eine konvexe Menge $U \subset X$ ist genau dann abgeschlossen, wenn sie schwach abgeschlossen ist.*

*Beweis.* Da eine konvergente Folge stets auch schwach konvergiert, ist jede schwach abgeschlossene Menge auch abgeschlossen. Sei daher $U \subset X$ konvex, abgeschlossen, und nichtleer (sonst ist nichts zu zeigen), und betrachte eine Folge $\{x_n\}_{n\in\mathbb{N}} \subset U$ mit $x_n \rightharpoonup x \in X$. Angenommen, $x \in X \setminus U$. Dann erfüllen die Mengen $U$ und $\{x\}$ die Bedingungen von Satz 1.5 (ii); wir finden daher ein $x^* \in X^*$ und ein $\lambda \in \mathbb{R}$ mit

$$\langle x^*, x_n\rangle_X \leq \lambda < \langle x^*, x\rangle_X \quad \text{für alle } n \in \mathbb{N}. \tag{1.7}$$

Grenzübergang $n \to \infty$ in der ersten Ungleichung ergibt dann aber den Widerspruch

$$\langle x^*, x\rangle_X < \langle x^*, x\rangle_X.$$ □

Ist $X$ nicht reflexiv (wie z. B. $L^\infty(\Omega)$), so müssen wir auf die schwach-∗ Konvergenz ausweichen.

Satz 1.10 (Banach–Alaoglu). *Ist der normierte Raum $X$ separabel (d. h. es gibt eine abzählbare dichte Teilmenge), so ist $B_{X^*}$ schwach-∗ kompakt.*

Nach dem Weierstraßschen Approximationssatz sind $C(\overline{\Omega})$ und $L^p(\Omega)$ für $1 \leq p < \infty$ separabel; auch $\ell^p$ ist separabel für $1 \leq p < \infty$. Also sind beschränkte und schwach-∗ abgeschlossene Kugeln in $\ell^\infty$, $L^\infty(\Omega)$ und $\mathcal{M}(\Omega)$ schwach-∗ kompakt; diese Räume sind aber selber nicht separabel. Beachten Sie aber, dass abgeschlossene konvexe Mengen in nichtreflexiven Räumen *nicht* schwach-∗ abgeschlossen sein müssen.

Schließlich werden wir noch den folgenden "schwach-∗"-Trennungssatz benötigen; der Beweis läuft analog zum Beweis von Satz 1.5 unter Verwendung der Tatsache, dass die linearen schwach-∗ stetigen Funktionale genau die Form $x^* \mapsto \langle x^*, x\rangle_X$ für ein $x \in X$ haben; siehe auch [Werner 2011, Satz VIII.2.11, Korollar VIII.3.4].

Satz 1.11. *Sei $A \subset X^*$ eine nichtleere, konvexe und schwach-∗ abgeschlossene Teilmenge und sei $x^* \in X^* \setminus A$. Dann existieren ein $x \in X$ und ein $\lambda \in \mathbb{R}$ mit*

$$\langle z^*, x\rangle_X \leq \lambda < \langle x^*, x\rangle_X \qquad \textit{für alle } z^* \in A.$$

Beachten Sie aber, dass abgeschlossene konvexe Mengen in nichtreflexiven Räumen *nicht* schwach-$*$ abgeschlossen sein müssen.

Da ein Dualraum den ursprünglichen Raum charakterisiert, ist dies auch der Fall für lineare Operatoren auf diesem Raum. Für $T \in L(X, Y)$ ist durch $T^* : Y^* \to X^*$,

$$\langle T^* y^*, x\rangle_X = \langle y^*, Tx\rangle_Y \qquad \text{für alle } x \in X,\, y^* \in Y^*$$

der *adjungierte Operator* $T^* \in L(Y^*, X^*)$ definiert. Es gilt stets $\|T^*\|_{L(Y^*,X^*)} = \|T\|_{L(X,Y)}$. Außerdem folgt aus der Stetigkeit von $T$, dass $T^*$ schwach-$*$ stetig (und $T$ natürlich schwach stetig) ist.

## 1.3 HILBERTRÄUME

Besonders weitgehende Dualitäts-Aussagen gelten in Hilberträumen. Eine Abbildung $(\cdot,\cdot) : X \times X \to \mathbb{R}$ auf dem Vektorraum $X$ über $\mathbb{R}$ heißt *Skalarprodukt*, falls gilt

(i) $(\alpha x + \beta y, z) = \alpha\,(x, z) + \beta\,(y, z)$ für alle $x, y, z \in X$ und $\alpha, \beta \in \mathbb{R}$;

(ii) $(x, y) = (y, x)$ für alle $x, y \in X$;

(iii) $(x, x) \geq 0$ für alle $x \in X$ mit Gleichheit genau dann, wenn $x = 0$.

Ein Banachraum mit Skalarprodukt $(X, (\cdot,\cdot)_X)$ wird *Hilbertraum* genannt; ist das Skalarprodukt kanonisch, lässt man es weg. Durch das Skalarprodukt wird eine Norm

$$\|x\|_X := \sqrt{(x, x)_X}$$

induziert, die der *Cauchy–Schwarz-Ungleichung* gehorcht:

$$(x, y)_X \leq \|x\|_X \|y\|_X.$$

Beispiel 1.3 (i–iii) für $p = 2 (= q)$ ist jeweils ein Hilbertraum, wobei das Skalarprodukt der dualen Paarung entspricht und die kanonischen Normen induziert.

Der für uns wesentliche Punkt ist, dass der Dualraum eines Hilbertraums $X$ mit $X$ identifiziert werden kann.

**Satz 1.12 (Fréchet–Riesz).** *Sei $X$ ein Hilbertraum. Dann existiert zu jedem $x^* \in X^*$ genau ein $z_{x^*} \in X$ mit $\|x^*\|_{X^*} = \|z_{x^*}\|_X$ und*

$$\langle x^*, x\rangle_X = (x, z_{x^*})_X \qquad \textit{für alle } x \in X.$$

Man bezeichnet $z_{x^*}$ als *Riesz-Repräsentant* von $x^*$. Die (lineare) Abbildung $J_X : X^* \to X$, $x^* \mapsto z_{x^*}$ wird *Riesz-Isomorphismus* genannt. Mit ihrer Hilfe zeigt man zum Beispiel, dass jeder Hilbertraum reflexiv ist.

Satz 1.12 erlaubt, anstelle der dualen Paarung das Skalarprodukt zu verwenden. So konvergiert $\{x_n\}_{n\in\mathbb{N}}$ schwach gegen $x$ genau dann, wenn gilt

$$(x_n, z)_X \to (x, z)_X \qquad \text{für alle } z \in X. \tag{1.8}$$

Ähnliches gilt für lineare Operatoren auf Hilberträumen. Für Hilberträume $X, Y$ wird zu $T \in L(X, Y)$ der *Hilbertraum-adjungierte Operator* $T^\star \in L(Y, X)$ definiert durch

$$\left(T^\star y, x\right)_X = (Tx, y)_Y \qquad \text{für alle } x \in X, y \in Y.$$

Ist $T^\star = T$, so nennt man $T$ *selbstadjungiert*. Zwischen den beiden Definitionen einer Adjungierten besteht die Beziehung $T^\star = J_X T^* J_Y^{-1}$. Ist der Kontext klar, werden wir nicht in der Notation unterscheiden.

# 2 GRUNDLAGEN DER VARIATIONSRECHNUNG

Wir betrachten zuerst die Frage nach der Existenz von Minimierern eines (nichtlinearen) Funktionals $F : U \to \mathbb{R}$ für eine Teilmenge $U$ eines Banachraums $X$. Mit solchen Problemen beschäftigt sich die *Variationsrechnung*.

## 2.1 DIREKTE METHODE DER VARIATIONSRECHNUNG

Es ist hilfreich, die Beschränkung $\bar{x} \in U$ in das Funktional aufzunehmen, indem wir $F$ auf $X$ erweitern, dafür aber den Wert $\infty$ zulassen. Wir betrachten also

$$\bar{F} : X \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}, \qquad \bar{F}(x) = \begin{cases} F(x) & x \in U, \\ \infty & x \in X \setminus U. \end{cases}$$

Dabei wird $\overline{\mathbb{R}}$ mit der üblichen Arithmetik versehen, d. h. $t < \infty$ und $t + \infty = \infty$ für alle $t \in \mathbb{R}$; Subtraktion und Multiplikation von negativen Zahlen mit $\infty$ und insbesondere $F(x) = -\infty$ sind nicht zugelassen. Existiert überhaupt ein $x \in U$, so kann ein Minimierer $\bar{x}$ also nur in $U$ liegen.

Wir betrachten also in Folge Funktionale $F : X \to \overline{\mathbb{R}}$. Die Menge, auf der $F$ endlich ist, bezeichnet man als (*effektiven*) *Definitionsbereich*

$$\operatorname{dom} F := \{x \in X : F(x) < \infty\}.$$

Ist $\operatorname{dom} F \neq \emptyset$, so nennt man $F$ *eigentlich* (englisch: „proper").

Wir verallgemeinern nun den Satz von Weierstraß (jede reellwertige stetige Funktion auf kompakten Mengen nimmt ihr Minimum und Maximum an) auf Banachräume und insbesondere auf Funktionen der Form $\bar{F}$. Da wir nur an Minima interessiert sind, reicht dafür eine „einseitige" Stetigkeit: Man nennt $F$ *unterhalbstetig* in $x \in X$ (englisch: „lower semicontinuous"), falls gilt

$$F(x) \leq \liminf_{n\to\infty} F(x_n) \quad \text{für alle Folgen } \{x_n\}_{n\in\mathbb{N}} \subset X \text{ mit } x_n \to x.$$

Analog definiert man *schwach(-∗) unterhalbstetige* Funktionen über schwach(-∗) konvergente Folgen. Gilt schließlich für jede Folge $\{x_n\}_{n\in\mathbb{N}} \subset X$ mit $\|x_n\|_X \to \infty$ auch $F(x_n) \to \infty$, so heißt $F$ *koerziv*.

Damit haben wir alle Begriffe zur Hand, um das zentrale Resultat der Variationsrechnung zu beweisen.[1]

Satz 2.1. *Sei $X$ ein reflexiver Banachraum und $F : X \to \overline{\mathbb{R}}$ eigentlich, koerziv und schwach unterhalbstetig. Dann hat das Minimierungsproblem*

$$\min_{x\in X} F(x)$$

*eine Lösung $\bar{x} \in \operatorname{dom} F$.*

*Beweis.* Der Beweis kann in drei Schritte aufgeteilt werden.„Die Existenz eines Minimierers folgt aus Standard-Argumenten.“

(i) *Zeige, dass eine Minimalfolge existiert.*

Da $F$ eigentlich ist, ist $M := \inf_{x\in X} F(x) < \infty$ (wobei $M = -\infty$ noch nicht ausgeschlossen ist). Wir können also eine Folge $\{y_n\}_{n\in\mathbb{N}} \subset \operatorname{ran} F \setminus \{\infty\} \subset \mathbb{R}$ finden mit $y_n \to M$, d. h. es existiert eine Folge $\{x_n\}_{n\in\mathbb{N}} \subset X$ mit

$$F(x_n) \to M = \inf_{x\in X} F(x).$$

Eine solche Folge wird *Minimalfolge* genannt. Beachten Sie, dass wir aus der Konvergenz von $\{F(x_n)\}_{n\in\mathbb{N}}$ (noch) nicht auf die Konvergenz von $\{x_n\}_{n\in\mathbb{N}}$ schließen können.

(ii) *Zeige, dass die Minimalfolge eine konvergente Teilfolge besitzt.*

Wir zeigen zuerst, dass $\{x_n\}_{n\in\mathbb{N}}$ beschränkt ist. Angenommen, das ist nicht der Fall, d. h. $\|x_n\|_X \to \infty$ für $n \to \infty$. Aus der Koerzivität von $F$ folgt dann auch $F(x_n) \to \infty$, im Widerspruch zu $F(x_n) \to M < \infty$ nach Definition der Minimalfolge. Also ist $\{x_n\}_{n\in\mathbb{N}}$ beschränkt und enthält daher nach dem Satz von Eberlein–Šmulyan eine schwach konvergente Teilfolge $\{x_{n_k}\}_{k\in\mathbb{N}}$ mit Grenzwert $\bar{x} \in X$. Dieser Grenzwert ist Kandidat für einen Minimierer.

(iii) *Zeige, dass dieser Grenzwert ein Minimierer ist.*

Aus der Definition der Minimalfolge folgt, dass auch für die Teilfolge $F(x_{n_k}) \to M$ gilt. Mit der schwachen Unterhalbstetigkeit von $F$ und der Definition des Infimums erhalten wir daher

$$\inf_{x\in X} F(x) \le F(\bar{x}) \le \liminf_{k\to\infty} F(x_{n_k}) = M = \inf_{x\in X} F(x) < \infty.$$

[1]Diese Beweis-Strategie, bekannt unter dem Namen *direkte Methode der Variationsrechnung*, wird so häufig angewendet, dass in Forschungsarbeiten üblicherweise nur geschrieben wird: Das Grundprinzip geht auf Hilbert zurück; die hier verwendete Formulierung für unterhalbstetige Funktionen stammt von Leonida Tonelli (1885–1946), der damit die moderne Variationsrechnung nachhaltig geprägt hat.

Daraus folgt $\bar{x} \in \operatorname{dom} F$ sowie $\inf_{x \in X} F(x) = F(\bar{x}) > -\infty$ (da $F$ eigentlich ist). Das Infimum wird also in $\bar{x} \in \operatorname{dom} F$ angenommen, und damit ist $\bar{x}$ der gesuchte Minimierer. □

Ist $X$ nicht reflexiv, aber Dualraum eines separablen Banachraums, so zeigt man analog die Existenz von Minimierern schwach-$*$ unterhalbstetiger Funktionale mit dem Satz von Banach–Alaoglu.

Beachten Sie, wie im Beweis die zu verwendende Topologie auf $X$ durch Schritt (ii) und (iii) eingeschränkt wird: Schritt (ii) profitiert von einer groben Topologie (in der mehr Folgen konvergieren), Schritt (iii) von einer feinen (je weniger Folgen konvergieren, desto einfacher ist die $\liminf$-Bedingung zu erfüllen). Da wir in unseren Fällen nicht mehr als die Beschränktheit einer Minimalfolge erwarten können, können wir keine feinere als die schwache Topologie verwenden. Es bleibt daher die Frage, ob genügend (interessante) Funktionale schwach unterhalbstetig sind.

Ein erstes Beispiel sind beschränkte lineare Funktionale: Ist $x^* \in X^*$, so ist

$$F : X \to \mathbb{R}, \qquad x \mapsto \langle x^*, x \rangle_X,$$

schwach stetig (nach Definition der schwachen Konvergenz) und damit insbesondere schwach unterhalbstetig. Ein weiterer Vorzug der (schwachen) Unterhalbstetigkeit ist, dass sie unter bestimmten Operationen erhalten bleibt.

Lemma 2.2. *Seien $X, Y$ Banachräume und sei $F : X \to \overline{\mathbb{R}}$ schwach unterhalbstetig. Dann sind schwach unterhalbstetig*

*(i) $\alpha F$ für alle $\alpha \geq 0$;*

*(ii) $F + G$ für $G : X \to \overline{\mathbb{R}}$ schwach unterhalbstetig;*

*(iii) $\varphi \circ F$ für $\varphi : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ unterhalbstetig und monoton steigend;*

*(iv) $F \circ \Phi$ für $\Phi : Y \to X$ schwach stetig, d. h. aus $y_n \rightharpoonup y$ folgt $\Phi(y_n) \rightharpoonup \Phi(y)$;*

*(v) $x \mapsto \sup_{i \in I} F_i(x)$ mit $F_i : X \to \overline{\mathbb{R}}$ schwach unterhalbstetig für eine beliebige Menge $I$.*

Beachte, dass Aussage (v) für stetige Funktionen *nicht* gilt!

*Beweis.* Die Aussagen (i) und (ii) folgen direkt aus den Rechenregeln für den $\liminf$.

Aussage (iii) folgt aus der Monotonie und der schwachen Unterhalbstetigkeit von $\varphi$, denn für $x_n \rightharpoonup x$ gilt

$$\varphi(F(x)) \leq \varphi(\liminf_{n \in \mathbb{N}} F(x_n)) \leq \liminf_{n \in \mathbb{N}} \varphi(F(x_n)).$$

Aussage (iv) folgt direkt aus der schwachen Stetigkeit von $\Phi$: Gilt $y_n \rightharpoonup y$, so gilt auch $x_n := \Phi(y_n) \rightharpoonup \Phi(y) =: x$, und aus der Unterhalbstetigkeit von $F$ folgt

$$F(\Phi(y)) \leq \liminf_{n\to\infty} F(\Phi(y_n)).$$

Sei schließlich $\{x_n\}_{n\in\mathbb{N}}$ eine schwach konvergente Folge mit Grenzwert $x \in X$. Dann gilt nach Definition des Supremums

$$F_j(x) \leq \liminf_{n\to\infty} F_j(x_n) \leq \liminf_{n\to\infty} \sup_{i\in I} F_i(x_n) \qquad \text{für alle } j \in I. \tag{2.1}$$

Nehmen wir auf beiden Seiten das Supremum über alle $j \in I$, so folgt die Aussage (v). □

Folgerung 2.3. *Sei $X$ ein Banachraum. Dann ist $\|\cdot\|_X$ eigentlich, koerziv und schwach unterhalbstetig.*

*Beweis.* Koerzivität und $\operatorname{dom} \|\cdot\|_X = X$ folgen direkt aus der Definition; schwache Unterhalbstetigkeit folgt aus Lemma 2.2 (v) und Folgerung 1.7, denn

$$\|x\|_X = \sup_{\|x^*\|_{X^*}\leq 1} |\langle x^*, x\rangle_X|. \tag{2.2}$$

□

Ein weiteres häufig auftretendes Funktional ist die *Indikator-Funktion*[2] einer Menge $U \subset X$, definiert als

$$\delta_U(x) = \begin{cases} 0 & x \in U, \\ \infty & x \in X \setminus U. \end{cases}$$

Der Zweck dieser Definition ist natürlich, die Minimierung eines Funktionals $F : X \to \mathbb{R}$ unter der Nebenbedingung $x \in U$ auf die unbeschränkte Minimierung von $\bar{F} := F + \delta_U$ zurückführen. Für die Existenz von Minimierern ist daher das folgende Resultat wichtig.

Lemma 2.4. *Sei $X$ ein Banachraum und $U \subset X$. Dann ist $\delta_U : X \to \overline{\mathbb{R}}$*

*(i) eigentlich, wenn $U$ nichtleer ist;*

*(ii) schwach unterhalbstetig, wenn $U$ konvex und abgeschlossen ist;*

*(iii) koerziv, wenn $U$ beschränkt ist.*

[2] nicht zu verwechseln mit der *charakteristischen* Funktion $\mathbb{1}_U$ mit $\mathbb{1}_U(x) = 1$ für $x \in U$ und 0 sonst!

*Beweis.* Aussage (i) ist klar. Für (ii) betrachte eine schwach konvergente Folge $\{x_n\}_{n\in\mathbb{N}} \subset X$ mit Grenzwert $x \in X$. Ist $x \in U$, dann ist wegen $\delta_U \geq 0$ natürlich

$$\delta_U(x) = 0 \leq \liminf_{n\to\infty} \delta_U(x_n).$$

Sei nun $x \notin U$. Da $U$ konvex und abgeschlossen und daher nach Lemma 1.9 auch schwach abgeschlossen ist, muss ein $N \in \mathbb{N}$ existieren mit $x_n \notin U$ für alle $n \geq N$ (sonst könnten wir – durch Übergang zu einer Teilfolge – eine Folge mit $x_n \rightharpoonup x \in U$ konstruieren, im Widerspruch zur Annahme). Also gilt $\delta_U(x_n) = \infty$ für alle $n \geq N$ und damit

$$\delta_U(x) = \infty = \liminf_{n\to\infty} \delta_U(x_n).$$

Für (iii) sei $U$ beschränkt, d. h. es gebe ein $M > 0$ mit $U \subset K_M(0)$. Gilt $\|x_n\|_X \to \infty$, so existiert ein $N \in \mathbb{N}$ mit $\|x_n\|_X > M$ für alle $n \geq N$, und damit $x_n \notin K_M(0) \supset U$ für alle $n \geq N$. Also gilt auch $\delta_U(x_n) \to \infty$. □

## 2.2 DIFFERENZIERBARKEIT IN BANACHRÄUMEN

Auch im Banachraum möchte man Minimierer intrinsisch mit Hilfe des Fermatschen Prinzip charakterisieren. Dafür übertragen wir den klassischen Ableitungsbegriff in den Banachraum. Dies geschieht in mehreren Schritten.

Seien $X, Y$ Banachräume, $F : X \to Y$ eine Abbildung, und $x, h \in X$.

- Existiert der einseitige Grenzwert

  $$F'(x; h) := \lim_{t\to 0^+} \frac{F(x + th) - F(x)}{t} \in Y,$$

  so nennen wir diesen *Richtungsableitung* in $x$ in Richtung $h$.

- Falls $F'(x; h)$ für alle $h \in X$ existiert und durch

  $$DF(x) : X \to Y, h \mapsto F'(x; h)$$

  ein linearer beschränkter(!) Operator definiert wird, so heißt $F$ *Gâteaux-differenzierbar* (in $x$) und $DF \in L(X, Y)$ *Gâteaux-Ableitung*.

- Gilt zusätzlich

  $$\lim_{\|h\|_X\to 0} \frac{\|F(x + h) - F(x) - DF(x)h\|_Y}{\|h\|_X} = 0,$$

  so heißt $F$ *Fréchet-differenzierbar* (in $x$) und $F'(x) := DF(x) \in L(X, Y)$ *Fréchet-Ableitung*.

- Ist die Abbildung $x \mapsto F'(x)$ stetig, so heißt $F$ *stetig differenzierbar*.

Der Unterschied zwischen Gâteaux- und Fréchet-Differenzierbarkeit liegt also im Approximationsfehler von $F$ in der Nähe von $x$ durch $F(x) + DF(x)h$: Während für Gâteaux-differenzierbare Funktionen dieser nur beschränkt durch $\|h\|_X$ – also linear in $\|h\|_X$ – sein muss, ist er für Fréchet-differenzierbare Funktionen sogar superlinear in $\|h\|_X$. (Für eine feste Richtung $h$ ist dies natürlich auch für Gâteaux-differenzierbare Funktionen der Fall; für Fréchet-differenzierbare Funktionen ist zusätzlich also Gleichmäßigkeit in $h$ gefordert.)

Ist $F$ Gâteaux-differenzierbar, kann man die Gâteaux-Ableitung berechnen via

$$DF(x)h = \left(\tfrac{d}{dt}F(x+th)\right)\Big|_{t=0}.$$

Offensichtlich sind lineare beschränkte Operatoren $F \in L(X, Y)$ überall Fréchet-differenzierbar mit Ableitung $F'(x) = F \in L(X, Y)$ für alle $x \in X$. Weitere Ableitungen erhält man durch die üblichen Rechenregeln, die genau wie in $\mathbb{R}^n$ gezeigt werden. Beispielhaft sei die folgende Kettenregel angegeben.

Satz 2.5. *Seien $X, Y, Z$ Banachräume und $F : X \to Y$ Fréchet-differenzierbar in $x \in X$ und $G : Y \to Z$ Fréchet-differenzierbar in $y := F(x) \in Y$. Dann ist $G \circ F$ Fréchet-differenzierbar in $x$ und*

$$(G \circ F)'(x) = G'(F(x)) \circ F'(x).$$

Eine analoge Regel für Gâteaux-Ableitungen gilt dagegen nicht!

Wir brauchen noch die folgende Variante des Mittelwertsatzes. Sei $[a, b] \subset \mathbb{R}$ ein beschränktes Intervall und $f : [a, b] \to X$ stetig. Dann ist das *Bochner-Integral* $\int_a^b f(t)\,dt \in X$ wohldefiniert und erfüllt nach Konstruktion

$$\left\langle x^*, \int_a^b f(t)\,dt \right\rangle_X = \int_a^b \langle x^*, f(t)\rangle_X\,dt \qquad \text{für alle } x^* \in X^* \tag{2.3}$$

sowie

$$\left\| \int_a^b f(t)\,dt \right\|_X \leq \int_a^b \|f(t)\|_X\,dt, \tag{2.4}$$

siehe z. B. [Růžička 2004, Folgerung 2.1.14].

Satz 2.6. *Sei $F : U \to Y$ Fréchet-differenzierbar, und seien $x \in U$ und $h \in X$ so dass $x + th \in U$ für alle $t \in [0, 1]$ gilt. Dann ist*

$$F(x+h) - F(x) = \int_0^1 F'(x+th)h\,dt. \tag{2.5}$$

*Beweis.* Betrachte für beliebiges $y^* \in Y^*$ die Funktion

$$f : [0,1] \to \mathbb{R}, \qquad t \mapsto \langle y^*, F(x+th)\rangle_Y. \tag{2.6}$$

Nach Satz 2.5 ist $f$ (als Verkettung von Abbildungen im Banachraum) differenzierbar mit

$$f'(t) = \langle y^*, F'(x+th)h\rangle_Y, \tag{2.7}$$

und der Hauptsatz der Differential- und Integralrechnung in $\mathbb{R}$ ergibt

$$\langle y^*, F(x+h) - F(x)\rangle_Y = f(1) - f(0) = \int_0^1 f'(t)\,dt = \left\langle y^*, \int_0^1 F'(x+th)h\,dt \right\rangle_Y, \tag{2.8}$$

wobei die letzte Gleichung aus (2.3) folgt. Da $y^* \in Y^*$ beliebig war, folgt daraus mit Folgerung 1.7 die gewünschte Gleichung. □

Wir betrachten nun die Charakterisierung von Minimierern eines differenzierbaren Funktionals $F : X \to \mathbb{R}$.[3] In diesem Fall ist die Gâteaux-Ableitung (falls sie existiert) $DF(x) \in L(X;\mathbb{R}) = X^*$; wir schreiben daher auch $F'(x;h) = \langle DF(x), h\rangle_X$ für die Richtungsableitung in $x$ in Richtung $h \in X$.

Satz 2.7 (Fermat-Prinzip). *Sei $F : X \to \mathbb{R}$ Gâteaux-differenzierbar und $\bar{x} \in X$ ein lokaler Minimierer von $F$. Dann gilt $DF(\bar{x}) = 0 \in X^*$, d. h.*

$$\langle DF(\bar{x}), h\rangle_X = F'(x;h) = 0 \qquad \textit{für alle } h \in X.$$

*Beweis.* Sei $h \in X$ beliebig. Da $\bar{x}$ ein lokaler Minimierer ist, existiert nach Lemma 1.2 ein $\varepsilon > 0$ mit $F(\bar{x}) \le F(\bar{x}+th)$ für alle $t \in (0,\varepsilon)$ und damit

$$0 \le \frac{F(\bar{x}+th) - F(\bar{x})}{t} \to F'(\bar{x};h) = \langle DF(\bar{x}), h\rangle_X \quad \text{für } t \to 0,$$

wobei wir die Gâteaux- und damit Richtungsdifferenzierbarkeit von $F$ verwendet haben. Da die rechte Seite linear in $h$ ist, können wir analog für $-h$ argumentieren und erhalten $\langle DF(\bar{x}), h\rangle_X \le 0$ und damit die Behauptung. □

Beachten Sie, dass Gâteaux-Ableitungen eines Funktionals $F : X \to \mathbb{R}$ Elemente des *Dualraums* $X^* = L(X,\mathbb{R})$ sind und daher nicht zu Vektoren in $X$ addiert werden können. In Hilberträumen (und insbesondere $\mathbb{R}^n$) kann man aber $DF(x) \in X^*$ mit Hilfe des Satz von Fréchet–Riesz kanonisch mit einem Element $\nabla F(x) \in X$, genannt *Gradient* von $F$, identifizieren über

$$\langle DF(x), h\rangle_X = (\nabla F(x), h)_X \qquad \text{für alle } h \in X.$$

[3] In der *indirekten Methode* der Variationsrechnung zeigt man darüber auch die Existenz von Minimierern, etwa als Lösung einer partiellen Differentialgleichung.

Als Beispiel betrachten wir für die durch das Skalarprodukt induzierte Norm in einem Hilbertraum das Funktional $F(x) = \frac{1}{2}\|x\|_X^2$. Dann gilt für alle $x, h \in X$

$$F'(x;h) = \lim_{t\to 0^+} \frac{\frac{1}{2}(x+th, x+th)_X - \frac{1}{2}(x,x)_X}{t} = (x,h)_X = \langle DF(x), h\rangle_X,$$

da das Skalarprodukt für festes $x$ linear in $h$ ist. Die quadrierte Norm ist also Gâteaux-differenzierbar in $x$ mit Ableitung $DF(x) = h \mapsto (x,h)_X \in X^*$ und Gradient $\nabla F(x) = x \in X$; wegen

$$\lim_{\|h\|_X\to 0} \frac{\left|\frac{1}{2}\|x+h\|_X^2 - \frac{1}{2}\|x\|_X^2 - (x,h)_X\right|}{\|h\|_X} = \lim_{\|h\|_X\to 0} \frac{1}{2}\|h\|_X = 0$$

ist sie sogar Fréchet-differenzierbar. Fasst man nun dieselbe Abbildungsvorschrift auf als definiert auf einem kleineren Hilbertraum $X' \hookrightarrow X$ (zum Beispiel $X = L^2(\Omega)$, $X' = H^1(\Omega)$), so ist $DF(x) \in (X')^*$ immer noch gegeben durch $\langle DF(x), h\rangle_{X'} = (x,h)_X$ (nur noch für alle $h \in X'$), aber $\nabla F(x) \in X'$ ist nun charakterisiert durch

$$\langle DF(x), h\rangle_X = (\nabla F(x), h)_{X'} \qquad \text{für alle } h \in X'.$$

Unterschiedliche Skalarprodukte führen daher zu unterschiedlichen Gradienten.

## 2.3 SUPERPOSITIONSOPERATOREN

Eine besondere Klasse von Operatoren auf Funktionenräumen sind solche, die durch punktweise Anwendung einer reellen Funktion definiert sind wie etwa $u(x) \mapsto \sin(u(x))$. Wir betrachten daher für $f : \Omega\times\mathbb{R} \to \mathbb{R}$ mit $\Omega \subset \mathbb{R}^n$ offen und beschränkt sowie $1 \le p, q \le \infty$ den zugehörigen *Superpositions-* oder *Nemytskii-Operator*

$$F : L^p(\Omega) \to L^q(\Omega), \qquad [F(u)](x) = f(x, u(x)) \quad \text{für fast alle } x \in \Omega. \tag{2.9}$$

Damit dieser Operator wohldefiniert ist, muss $f$ bestimmte Anforderungen erfüllen. Wir nennen $f$ *Carathéodory-Funktion*, falls

(i) für alle $z \in \mathbb{R}$ die Abbildung $x \mapsto f(x,z)$ messbar ist, und

(ii) für fast alle $x \in \Omega$ die Abbildung $z \mapsto f(x,z)$ stetig ist.

Wir fordern zusätzlich die folgende Wachstumsbedingung: Für gegebenes $1 \le p, q < \infty$ existieren $a \in L^q(\Omega)$ und $b \in L^\infty(\Omega)$ mit

$$|f(x,z)| \le a(x) + b(x)|z|^{p/q}. \tag{2.10}$$

Unter diesen Bedingungen ist $F$ wohldefiniert und sogar stetig.

**Satz 2.8.** *Erfüllt $f : \Omega\times\mathbb{R}\to\mathbb{R}$ die Wachstumsbedingung (2.10) für $1 \le p, q < \infty$, so ist der durch (2.9) definierte Superpositionsoperator $F : L^p(\Omega) \to L^q(\Omega)$ stetig.*

*Beweis.* Wir skizzieren die wesentlichen Schritte; einen ausführlichen Beweis findet man in [Růžička 2004, Lemma 1.19]. Zunächst zeigt man für gegebenes $u \in L^p(\Omega)$ mit Hilfe der Carathéodory-Eigenschaften die Messbarkeit von $F(u)$. Aus (2.10) und der Dreiecksungleichung folgt daher

$$\|F(u)\|_{L^q} \leq \|a\|_{L^q} + \|b\|_{L^\infty} \||u|^{p/q}\|_{L^q} = \|a\|_{L^q} + \|b\|_{L^\infty} \|u\|_{L^p}^{p/q} < \infty,$$

d. h. $F(u) \in L^q(\Omega)$.

Für die Stetigkeit betrachten wir eine Folge $\{u_n\}_{n\in\mathbb{N}} \subset L^p(\Omega)$ mit $u_n \to u \in L^p(\Omega)$. Dann existiert eine Teilfolge, die wir wieder mit $\{u_n\}_{n\in\mathbb{N}}$ bezeichnen, die punktweise fast überall in $\Omega$ konvergiert, sowie ein $v \in L^p(\Omega)$ mit $|u_n(x)| \leq |v(x)| + |u_1(x)| =: g(x)$ für alle $n \in \mathbb{N}$ und fast alle $x \in \Omega$ (siehe z. B. [Dobrowolski 2010, Satz 4.25 sowie den Beweis von Satz 4.17]). Aus der Stetigkeit von $z \mapsto f(x, z)$ folgt dann $F(u_n) \to F(u)$ punktweise fast überall sowie

$$|[F(u_n)](x)| \leq a(x) + b(x)|u_n(x)|^{p/q} \leq a(x) + b(x)|g(x)|^{p/q} \quad \text{für fast alle } x \in \Omega.$$

Da wegen $g \in L^p(\Omega)$ die rechte Seite in $L^q(\Omega)$ liegt, können wir den Satz von Lebesgue anwenden, und erhalten $F(u_n) \to F(u)$ in $L^q(\Omega)$. Da wir dieses Argument auf jede beliebige Teilfolge anwenden können, muss die gesamte Folge gegen $F(u)$ konvergieren, woraus die Stetigkeit folgt. □

Tatsächlich ist die Wachstumsbedingung (2.10) sogar notwendig für die Stetigkeit, siehe [Appell & Zabreiko 1990, Satz 3.2]. Außerdem zeigt man recht leicht, dass für $p = q = \infty$ aus (2.10) (in diesem Fall mit $p/q := 0$) folgt, dass $F$ sogar lokal Lipschitz-stetig ist.

Nun hätte man gerne, dass für differenzierbare $f$ auch $F$ differenzierbar ist, am besten mit punktweiser Ableitung $[F'(u)h](x) = f'(u(x))h(x)$. Dies ist aber nicht unbedingt der Fall; z. B. ist der durch $f(x, z) = \sin(z)$ definierte Superpositionsoperator für $1 \leq p = q < \infty$ in $u = 0$ *nicht* differenzierbar. Dies liegt daran, dass für einen Fréchet-differenzierbaren Superpositionsoperator $F : L^p(\Omega) \to L^q(\Omega)$ und eine Richtung $h \in L^p(\Omega)$ das punktweise(!) Produkt $F'(u)h \in L^q(\Omega)$ erfüllen muss. Dies stellt zusätzliche Anforderungen an den durch $f'$ definierten Superpositionsoperator $F'$, und ist als *zwei-Norm-Diskrepanz* bekannt.

Satz 2.9. *Sei $f : \Omega \times \mathbb{R} \to \mathbb{R}$ eine Carathéodory-Funktion, die die Wachstumsbedingung* (2.10) *für $1 \leq q < p < \infty$ erfüllt. Ist die partielle Ableitung $f'_z$ ebenfalls eine Carathéodory-Funktion und erfüllt* (2.10) *für $p' = p - q$, so ist der Superpositionsoperator $F : L^p(\Omega) \to L^q(\Omega)$ stetig Fréchet-differenzierbar, und die Ableitung in $u \in L^p(\Omega)$ in Richtung $h \in L^p(\Omega)$ ist gegeben durch*

$$[F'(u)h](x) = f'_z(x, u(x))h(x) \qquad \text{für fast alle } x \in \Omega.$$

*Beweis.* Nach Satz 2.8 ist für $r := \frac{pq}{p-q}$ (d. h. $\frac{r}{p} = \frac{p'}{q}$) der Superpositionsoperator

$$G : L^p(\Omega) \to L^r(\Omega), \qquad [G(u)](x) = f_z'(x, u(x)) \quad \text{für fast alle } x \in \Omega,$$

wohldefiniert und stetig. Aus der Hölderschen Ungleichung folgt weiter für beliebige $u \in L^p(\Omega)$

$$\|G(u)h\|_{L^q} \leq \|G(u)\|_{L^r} \|h\|_{L^p} \qquad \text{für alle } h \in L^p(\Omega),$$

d. h. durch $h \mapsto G(u)h$ wird ein linearer beschränkter Operator $DF(u) : L^p(\Omega) \to L^q(\Omega)$ definiert.

Sei nun $h \in L^p(\Omega)$ beliebig. Da $z \mapsto f(x, z)$ nach Annahme stetig differenzierbar ist, folgt aus dem klassischen Mittelwertsatz zusammen mit (2.4) und (2.3) nun

$$\begin{aligned}
&\|F(u+h) - F(u) - DF(u)h\|_{L^q} \\
&\qquad = \left(\int_\Omega |f(x, u(x) + h(x)) - f(x, u(x)) - f_z'(x, u(x))h(x)|^q \, dx\right)^{\frac{1}{q}} \\
&\qquad = \left(\int_\Omega \left|\int_0^1 f_z'(x, u(x) + th(x))h(x)\, dt - f_z'(x, u(x))h(x)\right|^q dx\right)^{\frac{1}{q}} \\
&\qquad = \left\|\int_0^1 G(u+th)h \, dt - G(u)h\right\|_{L^q} \\
&\qquad \leq \int_0^1 \|(G(u+th) - G(u))h\|_{L^q} \, dt \\
&\qquad \leq \int_0^1 \|G(u+th) - G(u)\|_{L^r} \, dt \, \|h\|_{L^p}.
\end{aligned}$$

Wegen der Stetigkeit von $G : L^p(\Omega) \to L^r(\Omega)$ geht das Integral für $\|h\|_{L^p} \to 0$ gegen Null, und damit ist $F$ nach Definition Fréchet-differenzierbar mit Ableitung $F'(u) = DF(u)$ (die wie bereits gezeigt stetig ist). $\square$

# Teil II

# KONVEXE ANALYSIS

# 3 KONVEXE FUNKTIONEN

Die klassischen Ableitungsbegriffe des letzten Kapitels sind für unsere Zwecke nicht ausreichend, denn viele interessante Funktionale sind in diesem Sinne nicht differenzierbar; auch Funktionale mit Werten in $\overline{\mathbb{R}}$ können damit nicht behandelt werden. Wir brauchen einen Ableitungsbegriff, der allgemeiner als Gâteaux- und Fréchet-Ableitungen ist, aber immer noch ein Fermatsches Prinzip und praktische Rechenregeln erlaubt.

Wir betrachten zuerst die Klasse der Funktionale, die solch eine verallgemeinerte Ableitung zulassen. Ein eigentliches Funktional $F : X \to \overline{\mathbb{R}}$ heißt *konvex*, wenn für alle $x, y \in X$ und $\lambda \in [0, 1]$ gilt

$$F(\lambda x + (1-\lambda)y) \leq \lambda F(x) + (1-\lambda)F(y) \tag{3.1}$$

(dabei ist der Funktionswert $\infty$ auf beiden Seiten zugelassen). Gilt für $x \neq y$ und $\lambda \in (0, 1)$ sogar

$$F(\lambda x + (1-\lambda)y) < \lambda F(x) + (1-\lambda)F(y),$$

so heißt $F$ *strikt konvex*.

Eine alternative Charakterisierung der Konvexität eines Funktionals $F : X \to \overline{\mathbb{R}}$ basiert auf ihrem *Epigraph*

$$\operatorname{epi} F := \{(x, t) \in X \times \mathbb{R} : F(x) \leq t\}.$$

Lemma 3.1. *Für $F : X \to \overline{\mathbb{R}}$ ist* $\operatorname{epi} F$

*(i) nichtleer genau dann, wenn $F$ eigentlich ist;*

*(ii) konvex genau dann, wenn $F$ konvex ist;*

*(iii) (schwach) abgeschlossen genau dann, wenn $F$ (schwach) unterhalbstetig ist.*

*Beweis.* Aussage (i) folgt direkt aus der Definition: $F$ ist eigentlich genau dann, wenn ein $x \in X$ und ein $t \in \mathbb{R}$ existiert mit $F(x) \leq t < \infty$, d. h. $(x, t) \in \operatorname{epi} F$.

Für (ii) sei $F$ konvex und seien $(x, r), (y, s) \in \operatorname{epi} F$ gegeben. Für beliebige $\lambda \in [0, 1]$ folgt dann aus (3.1), dass gilt

$$F(\lambda x + (1-\lambda)y) \leq \lambda F(x) + (1-\lambda)F(y) \leq \lambda r + (1-\lambda)s,$$

d. h. es ist

$$\lambda(x,r) + (1-\lambda)(y,s) = (\lambda x + (1-\lambda)y, \lambda r + (1-\lambda)s) \in \operatorname{epi} F$$

und damit $\operatorname{epi} F$ konvex. Sei umgekehrt $\operatorname{epi} F$ konvex und $x, y \in X$ beliebig, wobei wir $F(x) < \infty$ und $F(y) < \infty$ annehmen können (ansonsten ist (3.1) trivialerweise erfüllt). Offensichtlich sind $(x, F(x)), (y, F(y)) \in \operatorname{epi} F$. Aus der Konvexität von $\operatorname{epi} F$ folgt dann, dass für alle $\lambda \in [0,1]$ gilt

$$(\lambda x + (1-\lambda)y, \lambda F(x) + (1-\lambda)F(y)) = \lambda(x, F(x)) + (1-\lambda)(y, F(y)) \in \operatorname{epi} F$$

und damit nach Definition von $\operatorname{epi} F$ auch (3.1).

Nun zu (iii): Sei zuerst $F$ unterhalbstetig und $\{(x_n, t_n)\}_{n\in\mathbb{N}} \subset \operatorname{epi} F$ eine beliebige Folge mit $(x_n, t_n) \to (x, t) \in X \times \mathbb{R}$. Dann gilt

$$F(x) \leq \liminf_{n\to\infty} F(x_n) \leq \limsup_{n\to\infty} t_n = t,$$

d. h $(x, t) \in \operatorname{epi} F$. Sei umgekehrt $\operatorname{epi} F$ abgeschlossen und angenommen, $F$ ist eigentlich (sonst ist die Aussage trivial) und nicht unterhalbstetig. Dann existiert eine Folge $\{x_n\}_{n\in\mathbb{N}} \subset X$ mit $x_n \to x \in X$ und

$$F(x) > \liminf_{n\to\infty} F(x_n) =: M \in [-\infty, \infty).$$

Wir machen nun eine Fallunterscheidung:

a) $x \in \operatorname{dom} F$: Dann können wir eine Teilfolge auswählen, die wir wieder mit $\{x_n\}_{n\in\mathbb{N}}$ bezeichnen, so dass ein $\varepsilon > 0$ existiert mit $F(x_n) \leq F(x) - \varepsilon$ und damit $(x_n, F(x) - \varepsilon) \in \operatorname{epi} F$ für alle $n \in \mathbb{N}$. Wegen $x_n \to x$ folgt aus Abgeschlossenheit von $\operatorname{epi} F$ auch $(x, F(x) - \varepsilon) \in \operatorname{epi} F$ und damit $F(x) \leq F(x) - \varepsilon$, im Widerspruch zu $\varepsilon > 0$.

b) $x \notin \operatorname{dom} F$: In diesem Fall argumentiert man analog mit $F(x_n) \leq M + \varepsilon$ für $M > -\infty$ bzw. $F(x_n) \leq \varepsilon$ für $M = -\infty$, um einen Widerspruch zu $F(x) = \infty$ zu erhalten.

Genauso zeigt man die Äquivalenz von schwacher Unterhalbstetigkeit und schwacher Abgeschlossenheit. □

Zusammen mit Lemma 1.9 erhalten wir daraus sofort

Folgerung 3.2. *Sei $F : X \to \overline{\mathbb{R}}$ konvex. Dann ist $F$ schwach unterhalbstetig genau dann, wenn $F$ unterhalbstetig ist.*

Ebenfalls nützlich für die Betrachtung eines Funktionals $F : X \to \overline{\mathbb{R}}$ sind die zugehörigen *Subniveaumengen*

$$F_t := \{x \in X : F(x) \leq t\}, \qquad t \in \mathbb{R},$$

für die man analog zu Lemma 3.1 die folgenden Eigenschaften zeigt.

**Lemma 3.3.** *Für $F : X \to \overline{\mathbb{R}}$ gilt:*

*(i) Ist $F$ konvex, so ist $F_t$ konvex für alle $t \in \mathbb{R}$ (die Umkehrung gilt aber nicht);*

*(ii) $F$ ist (schwach) unterhalbstetig genau dann, wenn $F_t$ (schwach) abgeschlossen ist für alle $t \in \mathbb{R}$.*

Direkt aus der Definition folgt die Konvexität

(i) *stetig affiner Funktionale*, d. h. der Form $x \mapsto \langle x^*, x\rangle_X - \alpha$ für $x^* \in X^*$ und $\alpha \in \mathbb{R}$;

(ii) der Norm $\|\cdot\|_X$ in einem normierten Raum $X$;

(iii) der Indikatorfunktion $\delta_C$ für eine konvexe Menge $C$.

Ist $X$ ein Hilbertraum, so ist $F(x) = \|x\|_X^2$ sogar strikt konvex: Für $x, y \in X$ mit $x \neq y$ und $t \in (0,1)$ beliebig gilt

$$
\begin{aligned}
\|\lambda x + (1-\lambda)y\|_X^2 &= (\lambda x + (1-\lambda)y, \lambda x + (1-\lambda)y)_X \\
&= \lambda^2 (x,x)_X + 2\lambda(1-\lambda)\,(x,y)_X + (1-\lambda)^2\,(y,y)_X \\
&= \lambda\Big(\lambda\,(x,x)_X + (1-\lambda)\,(x-y,y)_X + (1-\lambda)\,(y,y)_X\Big) \\
&\quad + (1-\lambda)\Big(\lambda\,(x,x)_X + \lambda\,(x-y,y)_X + (1-\lambda)\,(y,y)_X\Big) \\
&= (\lambda + (1-\lambda))\Big(\lambda\,(x,x)_X + (1-\lambda)\,(y,y)_X\Big) - \lambda(1-\lambda)\,(x-y,x-y)_X \\
&= \lambda\|x\|_X^2 + (1-\lambda)\|y\|_X^2 - \lambda(1-\lambda)\|x-y\|_X^2 \\
&< \lambda\|x\|_X^2 + (1-\lambda)\|y\|_X^2.
\end{aligned}
$$

Eine in der Variationsrechnung besonders nützliche Klasse von konvexen Funktionen entsteht durch Integralfunktionale mit konvexen Integranden, die durch Superpositionsoperatoren definiert sind.

**Lemma 3.4.** *Sei $f : \mathbb{R} \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig. Sei $f$ zusätzlich nichtnegativ. Ist $\Omega \subset \mathbb{R}^n$ beschränkt und $1 \le p \le \infty$, dann ist auch*

$$
F : L^p(\Omega) \to \overline{\mathbb{R}}, \qquad u \mapsto \begin{cases} \int_\Omega f(u(x))\,dx & f \circ u \in L^1(\Omega), \\ \infty & \text{sonst}, \end{cases}
$$

*eigentlich, konvex, und unterhalbstetig.*

*Beweis.* Da $f$ eigentlich ist, existiert ein $t_0 \in \operatorname{dom} f$. Also ist die konstante Funktion $u \equiv t_0 \in \operatorname{dom} F$, da $f(u) \equiv f(t_0) \in L^\infty(\Omega) \subset L^1(\Omega)$ wegen der Beschränktheit von $\Omega$.

Für $u, v \in \operatorname{dom} F$ (sonst ist (3.1) trivialerweise erfüllt) und $\lambda \in [0,1]$ folgt aus der Konvexität von $f$, dass

$$0 \leq f(\lambda u(x) + (1-\lambda)v(x)) \leq \lambda f(u(x)) + (1-\lambda)f(v(x)) \quad \text{für fast alle } x \in \Omega.$$

Da $L^1(\Omega)$ ein Vektorraum ist, gilt wegen $u, v \in \operatorname{dom} F$ auch $\lambda f(u) + (1-\lambda)f(v) \in L^1(\Omega)$; damit ist der mittlere Ausdruck ebenfalls in $L^1(\Omega)$, und durch Integration der Ungleichung über $\Omega$ folgt die Konvexität von $F$.

Für die Unterhalbstetigkeit verwenden wir Lemma 3.1. Sei dafür $\{(u_n, t_n)\}_{n\in\mathbb{N}} \subset \operatorname{epi} F$ mit $u_n \to u$ in $L^p(\Omega)$ und $t_n \to t$ in $\mathbb{R}$. Dann existiert eine Teilfolge $\{u_{n_k}\}_{k\in\mathbb{N}}$ mit $u_{n_k}(x) \to u(x)$ fast überall. Wegen der Unterhalbstetigkeit und Nichtnegativität von $f$ folgt mit dem Lemma von Fatou

$$\begin{aligned} F(u) = \int_\Omega f(u(x))\,dx &\leq \int_\Omega \liminf_{k\to\infty} f(u_{n_k}(x))\,dx \leq \liminf_{k\to\infty} \int_\Omega f(u_{n_k}(x))\,dx \\ &= \liminf_{k\to\infty} F(u_{n_k}) \leq \limsup_{k\to\infty} t_{n_k} = t, \end{aligned}$$

d. h. $(u,t) \in \operatorname{epi} F$. Also ist $\operatorname{epi} F$ abgeschlossen und damit $F$ unterhalbstetig nach Lemma 3.1 (iii). □

Weitere Beispiele lassen sich analog zu Lemma 2.2 durch folgende Operationen erzeugen.

**Lemma 3.5.** *Seien $X, Y$ normierte Räume und sei $F : X \to \overline{\mathbb{R}}$ konvex. Dann sind konvex*

*(i) $\alpha F$ für alle $\alpha \geq 0$;*

*(ii) $F + G$ für $G : X \to \overline{\mathbb{R}}$ konvex (ist $F$* oder *$G$ strikt konvex, so auch $F + G$);*

*(iii) $\varphi \circ F$ für $\varphi : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ konvex und monoton steigend;*

*(iv) $F \circ A$ für $A : Y \to X$ linear;*

*(v) $x \mapsto \sup_{i\in I} F_i(x)$ mit $F_i : X \to \overline{\mathbb{R}}$ konvex für eine beliebige Menge $I$.*

Nach Lemma 3.5 (v) ist also das punktweise Supremum von affinen Funktionalen stets konvex. Tatsächlich lässt sich sogar jedes konvexe Funktional so darstellen. Dafür definieren wir für ein eigentliches Funktional $F : X \to \overline{\mathbb{R}}$ die *konvexe Hülle*

$$F^\Gamma : X \to \mathbb{R}, \qquad x \mapsto \sup\{a(x) : a \text{ stetig affin mit } a(\tilde{x}) \leq F(\tilde{x}) \text{ für alle } \tilde{x} \in X\}.$$

**Lemma 3.6.** *Sei $F : X \to \overline{\mathbb{R}}$ eigentlich. Dann ist $F$ genau dann konvex und unterhalbstetig, wenn gilt $F = F^\Gamma$.*

*Beweis.* Da stetig affine Funktionale konvex und stetig sind, ist $F = F^\Gamma$ nach Lemma 3.5 (v) und Lemma 2.2 (v) immer konvex und unterhalbstetig.

Für die andere Richtung sei $F : X \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig. Aus der Definition von $F^\Gamma$ als Supremum ist offensichtlich, dass stets $F^\Gamma \leq F$ punktweise gilt. Angenommen, $F^\Gamma < F$. Dann existiert ein Punkt $x_0 \in X$ und ein $\lambda \in \mathbb{R}$ mit

$$F^\Gamma(x_0) < \lambda < F(x_0).$$

Wir konstruieren nun mit Hilfe des Hahn–Banach-Trennungssatzes ein stetig affines Funktional $a \in X^*$ mit $a \leq F$ aber $a(x_0) > \lambda > F^\Gamma(x_0)$, was zusammen mit der Definition von $F^\Gamma$ zum Widerspruch führt. Da $F$ eigentlich, konvex und unterhalbstetig ist, ist $\operatorname{epi} F$ nach Lemma 3.1 nichtleer, konvex und abgeschlossen. Weiter ist $\{(x_0, \lambda)\}$ kompakt und wegen $\lambda < F(x_0)$ disjunkt zu $\operatorname{epi} F$. Satz 1.5 (ii) liefert also die Existenz eines $z^* \in (X \times \mathbb{R})^*$ und eines $\alpha \in \mathbb{R}$ mit

$$\langle z^*, (x,t)\rangle_{X\times\mathbb{R}} \leq \alpha < \langle z^*, (x_0, \lambda)\rangle_{X\times\mathbb{R}} \qquad \text{für alle } (x,t) \in \operatorname{epi} F.$$

Wir definieren nun ein $x^* \in X^*$ durch $\langle x^*, x\rangle_X = \langle z^*, (x,0)\rangle_{X\times\mathbb{R}}$ für alle $x \in X$ und setzen $s := \langle z^*, (0,1)\rangle_{X\times\mathbb{R}} \in \mathbb{R}$. Dann gilt $\langle z^*, (x,t)\rangle_{X\times\mathbb{R}} = \langle x^*, x\rangle_X + st$ und damit

$$\langle x^*, x\rangle_X + st \leq \alpha < \langle x^*, x_0\rangle_X + s\lambda \qquad \text{für alle } (x,t) \in \operatorname{epi} F. \tag{3.2}$$

Nun ist für $(x,t) \in \operatorname{epi} F$ auch $(x,t') \in \operatorname{epi} F$ für alle $t' > t$, und aus der ersten Ungleichung in (3.2) folgt für alle $t' > 0$ groß genug

$$s \leq \frac{\alpha - \langle x^*, x\rangle_X}{t'} \to 0 \qquad \text{für } t' \to \infty.$$

Also ist $s \leq 0$. Wir machen nun eine Fallunterscheidung.

(i) $s < 0$: Wir setzen

$$a : X \to \mathbb{R}, \qquad x \mapsto \frac{\alpha - \langle x^*, x\rangle_X}{s}.$$

Dann ist $a$ affin und stetig. Für $x \in \operatorname{dom} F$ ist außerdem $(x, F(x)) \in \operatorname{epi} F$, und aus der produktiven Null und der ersten Ungleichung von (3.2) folgt (beachte $s < 0$!)

$$a(x) = \tfrac{1}{s}\left(\alpha - \langle x^*, x\rangle_X - sF(x)\right) + F(x) \leq F(x).$$

(Für $x \notin \operatorname{dom} F$ ist die Aussage klar.) Aus der zweiten Ungleichung von (3.2) folgt aber

$$a(x_0) = \tfrac{1}{s}\left(\alpha - \langle x^*, x_0\rangle_X\right) > \lambda.$$

(ii) $s = 0$: Dann folgt $\langle x^*, x\rangle_X \leq \alpha < \langle x^*, x_0\rangle_X$ für alle $x \in \operatorname{dom} F$, weshalb $x_0 \notin \operatorname{dom} F$ gelten muss. Allerdings ist $F$ eigentlich, so dass ein $y_0 \in \operatorname{dom} F$ existiert, für das wir analog zu Fall (i) durch Trennung von $\operatorname{epi} F$ und $(y_0, \mu)$ für $\mu$ klein genug ein stetiges affines Funktional $a_0 : X \to \mathbb{R}$ mit $a_0 \le F$ punktweise konstruieren können. Wir setzen nun für $\rho > 0$

$$a_\rho : X \to \mathbb{R}, \qquad x \mapsto a_0(x) + \rho\left(\langle x^*, x\rangle_X - \alpha\right).$$

Dann ist auch $a_\rho$ affin und stetig, und wegen $\langle x^*, x\rangle_X \le \alpha$ gilt $a_\rho(x) \le a_0(x) \le F(x)$ für alle $x \in \operatorname{dom} F$ und $\rho > 0$ beliebig. Wegen $\langle x^*, x_0\rangle_X > \alpha$ existiert aber ein $\rho > 0$ mit $a_\rho(x_0) > \lambda$.

In beiden Fällen muss nach Definition von $F^\Gamma$ als Supremum also auch $F^\Gamma(x_0) > \lambda$ gelten, im Widerspruch zur Annahme $F^\Gamma(x_0) < \lambda$. □

Nach all der Vorarbeit können wir nun schnell das Hauptresultat über die Existenz von Lösungen konvexer Minimierungsaufgaben beweisen.

Satz 3.7. *Sei $X$ ein reflexiver Banachraum und seien*

*(i) $U \subset X$ nichtleer, abgeschlossen, und konvex,*

*(ii) $F : U \to \overline{\mathbb{R}}$ eigentlich, konvex, und unterhalbstetig mit* $\operatorname{dom} F \cap U \neq \emptyset$,

*(iii) $U$ beschränkt oder $F$ koerziv.*

*Dann hat das Problem*

$$\min_{x\in U} F(x)$$

*eine Lösung $\bar{x} \in U \cap$* $\operatorname{dom} F$*. Ist $F$ strikt konvex, so ist die Lösung eindeutig.*

*Beweis.* Wir betrachten das Funktional $\bar{F} = F + \delta_U$. Aus Voraussetzung (i) folgt mit Lemma 2.2, dass $\delta_U$ eigentlich, konvex und unterhalbstetig ist. Wegen (ii) existiert ein Punkt $x_0 \in U$ mit $\bar{F}(x_0) < \infty$, daher ist auch $\bar{F}$ eigentlich, konvex und unterhalbstetig und damit nach Folgerung 3.2 sogar schwach unterhalbstetig. Aus (iii) folgt schließlich die Koerzivität von $\bar{F}$ (da die Summe koerziv ist, sobald es einer dieser Summanden ist). Da $X$ reflexiv ist, können wir nun Satz 2.1 anwenden und erhalten die Existenz eines Minimierers $\bar{x} \in \operatorname{dom} \bar{F} = U \cap \operatorname{dom} F$ von $\bar{F}$ mit

$$F(\bar{x}) = \bar{F}(\bar{x}) \le \bar{F}(x) = F(x) \qquad \text{für alle } x \in U,$$

d. h. $\bar{x}$ die gesuchte Lösung.

Sei nun $F$ strikt konvex, und seien $\bar{x}, \bar{x}' \in U$ zwei verschiedene Minimierer, d. h. $F(\bar{x}) = F(\bar{x}') = \min_{x\in U} F(x)$. Dann ist für alle $\lambda \in (0,1)$ wegen der Konvexität von $U$ auch

$$x_\lambda := \lambda\bar{x} + (1-\lambda)\bar{x}' \in U,$$

aber wegen der strikten Konvexität von $F$ gilt

$$F(x_\lambda) < \lambda F(\bar{x}) + (1-\lambda)F(\bar{x}') = F(\bar{x}),$$

im Widerspruch zu $F(\bar{x}) \le F(x)$ für alle $x \in U$. □

Zum Abschluß dieses Kapitels zeigen wir, das folgende erstaunliche Resultat: *Jede (lokal) beschränkte konvexe Funktion ist (lokal) Lipschitzstetig.* Neben seinem Nutzen in späteren Kapiteln demonstriert dieses Resultat die Eleganz der konvexen Analysis: eine algebraische, globale Eigenschaft (Konvexität) verknüpft zwei topologische, lokale Eigenschaften (Umgebungen und Stetigkeit). Hier verwenden wir die starke Topologie in einem normierten Vektorraum.

Satz 3.8. *Sei $X$ ein normierter Vektorraum und $F : X \to \overline{\mathbb{R}}$ konvex. Existiert für $x \in X$ ein $\rho > 0$ so, dass $F$ auf $O_\rho(x)$ nach oben beschränkt ist, so ist $F$ in $x$ lokal Lipschitz-stetig.*

*Beweis.* Nach Voraussetzung existiert ein $M \in \mathbb{R}$ mit $F(y) \le M$ für alle $y \in O_\rho(x)$. Wir zeigen zunächst, dass $F$ dann ebenfalls lokal nach unten beschränkt ist. Sei dafür $y \in O_\rho(x)$ beliebig. Dann ist wegen $\|x - y\|_X < \rho$ auch $z := 2x - y = x - (y - x) \in O_\rho(x)$, und aus der Konvexität von $F$ folgt $F(x) = F\left(\frac{1}{2}y + \frac{1}{2}z\right) \le \frac{1}{2}F(y) + \frac{1}{2}F(z)$ und damit

$$-F(y) \le F(z) - 2F(x) \le M - 2F(x) =: m,$$

d. h. $-m \le F(y) \le M$ für alle $y \in O_\rho(x)$.

Wir zeigen nun, dass daraus die Lipschitz-Stetigkeit auf $O_{\frac{\rho}{2}}(x)$ folgt. Seien $y_1, y_2 \in O_{\frac{\rho}{2}}(x)$ mit $y_1 \neq y_2$ und setze

$$z := y_1 + \frac{\rho}{2}\frac{y_1 - y_2}{\|y_1 - y_2\|_X} \in O_\rho(x)$$

wegen $\|z - x\|_X \le \|y_1 - x\|_X + \frac{\rho}{2} < \rho$. Nach Konstruktion ist nun

$$y_1 = \lambda z + (1 - \lambda)y_2 \quad \text{für} \quad \lambda := \frac{\|y_1 - y_2\|_X}{\|y_1 - y_2\|_X + \frac{\rho}{2}} \in (0, 1),$$

so dass wegen der Konvexität von $F$ gilt $F(y_1) \le \lambda F(z) + (1 - \lambda)F(y_2)$. Daraus folgt zusammen mit der Definition von $\lambda$ sowie $F(z) \le M$ und $-F(y_2) \le m = M - 2F(x)$ die Abschätzung

$$\begin{aligned} F(y_1) - F(y_2) \le \lambda(F(z) - F(y_2)) &\le \lambda(2M - 2F(x)) \\ &= \frac{2(M - F(x))}{\|y_1 - y_2\|_X + \frac{\rho}{2}}\|y_1 - y_2\|_X \\ &\le \frac{2(M - F(x))}{\rho/2}\|y_1 - y_2\|_X. \end{aligned}$$

Durch Vertauschung von $y_1$ und $y_2$ in der Konstruktion von $z$ erhalten wir damit

$$|F(y_1) - F(y_2)| \le \frac{2(M - F(x))}{\rho/2}\|y_1 - y_2\|_X \quad \text{für alle } y_1, y_2 \in O_{\frac{\rho}{2}}(x)$$

und damit die lokale Lipschitz-Stetigkeit mit Konstante $L(x, \rho/2) := 4(M - F(x))/\rho$. □

Wir folgern daraus die gewünschte Aussage zunächst für skalare Funktionen.

Folgerung 3.9. *Sei $f : \mathbb{R} \to \overline{\mathbb{R}}$ konvex. Dann ist $f$ lokal Lipschitz-stetig auf* $(\operatorname{dom} f)^o$.

*Beweis.* Sei $x \in (\operatorname{dom} f)^o$. Dann existieren $a, b \in \mathbb{R}$ mit $x \in (a, b) \subset \operatorname{dom} f$. Sei nun $z \in (a, b)$. Da Intervalle konvex sind, existiert ein $\lambda \in (0, 1)$ mit $z = \lambda a + (1 - \lambda)b$. Aus der Konvexität folgt daher

$$f(z) \leq \lambda f(a) + (1 - \lambda) f(b) \leq \max\{f(a), f(b)\} < \infty.$$

Also ist $f$ in $x$ lokal nach oben beschränkt, und die Behauptung folgt aus Satz 3.8. □

Mit etwas mehr Aufwand zeigt man, dass die Aussage auch für $F : \mathbb{R}^n \to \overline{\mathbb{R}}$, $n \in \mathbb{N}$ beliebig, gilt; siehe z. B. [Brokate 2014, Satz 2.18].

Um daraus den allgemeinen Fall zu erhalten, sind weitere Annahmen an $X$ und $F$ nötig.

Satz 3.10. *Sei $X$ ein Banachraum und $F : X \to \overline{\mathbb{R}}$ konvex und unterhalbstetig. Dann ist $F$ auf* $(\operatorname{dom} F)^o$ *lokal Lipschitz-stetig.*

*Beweis.* Wir zeigen die Aussage zuerst für den Fall $x = 0 \in (\operatorname{dom} F)^o$. Dann ist insbesondere $M := |F(0)| < \infty$. Betrachte nun für $h \in X$ beliebig die Abbildung

$$f : \mathbb{R} \to \overline{\mathbb{R}}, \qquad t \mapsto F(th).$$

Man vergewissert sich leicht, das $f$ ebenfalls konvex und unterhalbstetig ist mit $0 \in (\operatorname{dom} f)^o$. Nach Folgerung 3.9 ist $f$ daher lokal Lipschitz-stetig in 0, d. h. $|f(t) - f(0)| \leq Lt \leq 1$ für $t > 0$ klein genug. Aus der umgekehrten Dreiecksungleichung erhalten wir also ein $\delta > 0$ mit

$$F(0 + th) \leq |F(0 + th)| = |f(t)| \leq |f(0)| + 1 = M + 1 \qquad \text{für alle } t \in [0, \delta],$$

und damit liegt 0 im algebraischen Inneren der Subniveaumenge $F_{M+1}$, die nach Lemma 3.3 konvex und abgeschlossen ist. Aus Lemma 1.2 folgt nun $0 \in (F_{M+1})^o$, es gibt daher ein $\rho > 0$ mit $F(z) \leq M + 1$ für alle $z \in O_\rho(0)$. Also ist $F$ in 0 lokal nach oben beschränkt und daher nach Satz 3.8 lokal Lipschitz-stetig.

Für den allgemeinen Fall $x \in (\operatorname{dom} F)^o$ betrachte

$$\tilde{F} : X \to \overline{\mathbb{R}}, \qquad y \mapsto F(y - x).$$

Wieder vergewissert man sich leicht, dass $\tilde{F}$ ebenfalls konvex und unterhalbstetig ist mit $0 \in (\operatorname{dom} \tilde{F})^o$. Also ist wie eben bewiesen $\tilde{F}$ lokal Lipschitz-stetig in einer Umgebung $O_\rho(0)$, woraus folgt

$$|F(y_1) - F(y_2)| = |\tilde{F}(y_1 + x) - \tilde{F}(y_2 + x)| \leq L\|y_1 - y_2\|_X \quad \text{für alle } y_1, y_2 \in O_\rho(x)$$

und damit die behauptete lokale Lipschitz-Stetigkeit von $F$. □

# 4 DAS KONVEXE SUBDIFFERENTIAL

Wir wenden uns nun der Charakterisierung von Minimierern konvexer Funktionen durch ein Fermatsches Prinzip zu. Ein erster Kandidat für den nötigen Ableitungsbegriff ist die Richtungsableitung, denn diese existiert (zumindest in den erweiterten reellen Zahlen) für jede konvexe Funktion.

Lemma 4.1. *Sei $F : X \to \overline{\mathbb{R}}$ konvex und seien $x \in \operatorname{dom} F$ und $h \in X$ gegeben. Dann gilt:*

*(i) Die Funktion*

$$\varphi : (0, \infty) \to \overline{\mathbb{R}}, \qquad t \mapsto \frac{F(x + th) - F(x)}{t},$$

*ist monoton steigend.*

*(ii) Der Grenzwert $F'(x; h) = \lim_{t \to 0^+} \varphi(t) \in [-\infty, \infty]$ existiert und erfüllt*

$$F'(x; h) \leq F(x + h) - F(x). \tag{4.1}$$

*Beweis.* Zu (i): Durch Einsetzen und Umformen sieht man, dass für alle $0 < s < t$ die Bedingung $\varphi(s) \leq \varphi(t)$ äquivalent ist zu

$$F(x + sh) \leq \frac{s}{t} F(x + th) + \left(1 - \frac{s}{t}\right) F(x).$$

Dies folgt aber wegen $x + sh = (1 - \frac{s}{t})x + \frac{s}{t}(x + th)$ aus der Konvexität von $F$.

Aussage (ii) folgt sofort nun aus (i) wegen

$$F'(x; h) = \lim_{t \to 0^+} \varphi(t) = \inf_{t > 0} \varphi(t) \leq \varphi(1) = F(x + h) - F(x). \qquad \square$$

Leider liefert dieser Begriff noch nicht das Gewünschte, denn die konvexe Funktion $f : \mathbb{R} \to \mathbb{R}$, $f(t) = |t|$ hat ein Minimum in $t = 0$, dort aber nur die Richtungsableitung $f'(0; h) = |h| > 0$ für $h \in \mathbb{R} \setminus \{0\}$. Es gilt also nicht $f'(0; h) = 0$ für irgendein $h \neq 0$ – aber es gilt zumindest $0 \leq f'(0; h)$ für alle $h \in \mathbb{R}$. Diese Bedingung wollen wir auf allgemeine normierte Räume verallgemeinern. Dafür betrachten wir für $F : X \to \overline{\mathbb{R}}$ konvex und $x \in \operatorname{dom} F$ die Menge

$$\{x^* \in X^* : \langle x^*, h \rangle_X \leq F'(x; h) \quad \text{für alle } h \in X\}. \tag{4.2}$$

Diese Menge (die auch leer sein kann!) ist mit Hilfe von Lemma 4.1 auch ohne Richtungsableitung charakterisierbar.

Lemma 4.2. *Seien $F : X \to \overline{\mathbb{R}}$ konvex und $x \in \operatorname{dom} F$. Dann sind für $x^* \in X^*$ äquivalent*

*(i)* $\langle x^*, h\rangle_X \leq F'(x;h)$ *für alle $h \in X$;*

*(ii)* $\langle x^*, h\rangle_X \leq F(x+h) - F(x)$ *für alle $h \in X$.*

*Beweis.* Gilt (i), so folgt direkt aus Lemma 4.1 (ii), dass für alle $h \in X$ gilt

$$\langle x^*, h\rangle_X \leq F'(x;h) \leq F(x+h) - F(x).$$

Gilt (ii) für alle $h \in X$, so auch für $th$ für alle $h \in X$ und $t > 0$. Division durch $t$ und Grenzübergang liefert dann

$$\langle x^*, h\rangle_X \leq \lim_{t\to 0^+} \frac{F(x+th) - F(x)}{t} = F'(x;h). \qquad \square$$

Führen wir $\tilde{x} = x + h \in X$ ein, so führt die zweite Bedingung auf die folgende Definition. Für $F : X \to \overline{\mathbb{R}}$ und $x \in \operatorname{dom} F$ definieren wir das *(konvexe) Subdifferential* als

$$\partial F(x) := \{x^* \in X^* : \langle x^*, \tilde{x} - x\rangle_X \leq F(\tilde{x}) - F(x) \quad \text{für alle } \tilde{x} \in X\}. \tag{4.3}$$

(Beachten Sie, dass $\tilde{x} \notin \operatorname{dom} F$ zugelassen ist, da dann die Ungleichung trivialerweise erfüllt ist.) Für $x \notin \operatorname{dom} F$ setzen wir $\partial F(x) = \emptyset$. Direkt aus der Definition folgt, dass $\partial F(x)$ konvex und schwach-$*$ abgeschlossen ist. Ein Element $\xi \in \partial F(x)$ heißt *Subgradient*.

Diese Definition liefert nun das Gewünschte.

Satz 4.3. *Seien $F : X \to \overline{\mathbb{R}}$ und $\bar{x} \in \operatorname{dom} F$. Dann sind äquivalent:*

*(i)* $0 \in \partial F(\bar{x})$;

*(ii)* $F(\bar{x}) = \min_{x\in X} F(x)$.

*Beweis.* Dies folgt direkt aus den Definitionen: Es ist $0 \in \partial F(\bar{x})$ genau dann, wenn gilt

$$0 = \langle 0, \tilde{x} - \bar{x}\rangle_X \leq F(\tilde{x}) - F(\bar{x}) \qquad \text{für alle } \tilde{x} \in X,$$

d. h. $F(\bar{x}) \leq F(\tilde{x})$ für alle $\tilde{x} \in X$. $\square$

Dies entspricht auch der geometrischen Anschauung: Für $X = \mathbb{R} = X^*$ beschreibt $\tilde{y} := f(\tilde{x}) = f(x) + \xi(\tilde{x} - x)$ mit $\xi \in \partial f(x)$ eine Tangente an $y = f(x)$ mit Steigung $\xi$; die Bedingung $\xi = 0 \in \partial f(\tilde{x})$ bedeutet also, dass $f$ in $\bar{x}$ eine waagerechte Tangente hat.[1]

[1]Beachten Sie, dass in Satz 4.3 nirgendwo die Konvexität von $F$ eingeht! Tatsächlich charakterisiert $0 \in \partial F(\bar{x})$ die *globalen* Minimierer jeder Funktion. Nichtkonvexe Funktionen können aber auch lokale Minimierer haben, für die die Subdifferentialinklusion nicht erfüllt ist. Tatsächlich sind (konvexe) Subdifferentiale nichtkonvexer Funktionen in der Regel leer. Dies führt insbesondere für den Beweis einiger Rechenregeln zu Problemen, für die wir in der Tat Konvexität voraussetzen müssen.

Wir betrachten einige Beispiele. Zunächst ist aus der Konstruktion ersichtlich, dass das Subdifferential die Gâteaux-Ableitung verallgemeinert.

Satz 4.4. *Sei $F : X \to \overline{\mathbb{R}}$ konvex und Gâteaux-differenzierbar in $x$. Dann ist $\partial F(x) = \{DF(x)\}$.*

*Beweis.* Nach Definition der Gâteaux-Ableitung gilt

$$\langle DF(x), h\rangle_X = DF(x)h = F'(x;h) \quad \text{für alle } h \in X.$$

Aus Lemma 4.2 folgt nun mit $\tilde{x} := x + h$ sofort $DF(x) \in \partial F(x)$.

Umgekehrt folgt aus $\xi \in \partial F(x)$ wieder mit Lemma 4.2 für $h := \tilde{x} - x \in X$, dass gilt

$$\langle \xi, h\rangle_X \le F'(x;h) = \langle DF(x), h\rangle_X.$$

Da $\tilde{x} \in X$ beliebig war, gilt dies für alle $h \in X$. Supremum über alle $h$ mit $\|h\|_X \le 1$ liefert dann $\|\xi - DF(x)\|_{X^*} \le 0$, d. h. $\xi = DF(x)$. □

Natürlich möchten wir auch Subdifferentiale von Funktionen haben, die nicht differenzierbar sind. Das kanonische Beispiel ist die Norm $\|x\|_X$ in einem normierten Raum, die ja in $x = 0$ nicht differenzierbar ist.

Satz 4.5. *Für $x \in X$ ist*

$$\partial(\|\cdot\|_X)(x) = \begin{cases} \{x^* \in X^* : \langle x^*, x\rangle_X = \|x\|_X \text{ und } \|x^*\|_{X^*} = 1\} & \text{falls } x \neq 0, \\ B_{X^*} & \text{falls } x = 0. \end{cases} \tag{4.4}$$

*Beweis.* Für $x = 0$ ist nach Definition $\xi \in \partial(\|\cdot\|_X)(x)$ genau dann, wenn gilt

$$\langle \xi, \tilde{x}\rangle_X \le \|\tilde{x}\|_X \qquad \text{für alle } \tilde{x} \in X \setminus \{0\}$$

(für $\tilde{x} = 0$ ist die Ungleichung trivial). Dies ist aber wegen der Definition der Operatornorm äquivalent mit $\|\xi\|_{X^*} \le 1$.

Sei nun $x \neq 0$ und betrachte $\xi \in \partial(\|\cdot\|_X)(x)$. Indem wir nacheinander $\tilde{x} = 0$ und $\tilde{x} = 2x$ in die Definition (4.3) einsetzen, erhalten wir

$$\|x\|_X \le \langle \xi, x\rangle_X = \langle \xi, 2x - x\rangle \le \|2x\|_X - \|x\|_X = \|x\|_X,$$

d. h. $\langle \xi, x\rangle_X = \|x\|_X$. Analog haben wir für alle $\tilde{x} \in X$, dass gilt

$$\langle \xi, \tilde{x}\rangle_X = \langle \xi, (\tilde{x} + x) - x\rangle_X \le \|\tilde{x} + x\|_X - \|x\|_X \le \|\tilde{x}\|_X,$$

woraus wie im Fall $x = 0$ folgt $\|\xi\|_{X^*} \le 1$. Für $\tilde{x} = x/\|x\|_X$ gilt nun

$$\langle \xi, \tilde{x}\rangle_X = \|x\|_X^{-1}\langle \xi, x\rangle_X = \|x\|_X^{-1}\|x\|_X = 1.$$

Also ist tatsächlich $\|\xi\|_{X^*} = 1$.

Es sei umgekehrt $x^* \in X^*$ mit $\langle x^*, x\rangle_X = \|x\|_X$ und $\|x^*\|_{X^*} = 1$. Dann gilt mit (1.2) für alle $\tilde{x} \in X$ die Relation

$$\langle x^*, \tilde{x} - x\rangle_X = \langle x^*, \tilde{x}\rangle_X - \langle x^*, x\rangle_X \leq \|\tilde{x}\|_X - \|x\|_X,$$

und daher nach Definition $x^* \in \partial(\|\cdot\|_X)(x)$ □

Für den Fall $X = \mathbb{R}$ erhalten wir daraus das Subdifferential der Betragsfunktion als

$$\partial(|\cdot|)(t) = \operatorname{sign}(t) := \begin{cases} \{1\} & \text{falls } t > 0, \\ \{-1\} & \text{falls } t < 0, \\ [-1,1] & \text{falls } t = 0. \end{cases} \tag{4.5}$$

Schließlich haben wir auch eine einfache Darstellung für das Subdifferential der Indikatorfunktion einer konvexen Menge $C \subset X$. Für $x \in C = \operatorname{dom} \delta_C$ gilt nämlich

$$\begin{aligned} x^* \in \partial\delta_C(x) &\Leftrightarrow \langle x^*, \tilde{x} - x\rangle_X \leq \delta_C(\tilde{x}) \quad \text{für alle } \tilde{x} \in X \\ &\Leftrightarrow \langle x^*, \tilde{x} - x\rangle_X \leq 0 \quad \text{für alle } \tilde{x} \in C, \end{aligned}$$

da die Ungleichung für alle $\tilde{x} \notin C$ trivialerweise erfüllt ist. Die Menge $\partial\delta_C(x)$ nennt man auch *Normalenkegel* an $C$ in $x$. Wieder ist das Beispiel $X = \mathbb{R}$ erhellend: Betrachte $C = [-1,1]$ und $t \in C$. Dann ist $\xi \in \partial\delta_{[-1,1]}(t)$ genau dann, wenn $\xi(\tilde{t} - t) \leq 0$ für alle $\tilde{t} \in [-1,1]$ gilt. Wir machen nun eine Fallunterscheidung:

(i) $t = 1$. Dann ist $\tilde{t} - t \in [-2, 0]$ und damit gilt die Bedingung genau dann, wenn $\xi \geq 0$ ist.

(ii) $t = -1$. Dann ist $\tilde{t} - t \in [0, 2]$ und damit gilt die Bedingung genau dann, wenn $\xi \leq 0$ ist.

(iii) $t \in (-1, 1)$. Dann kann $\tilde{t} - t$ sowohl positive als auch negative Werte annehmen, und damit muss $\xi = 0$ sein.

Also ist

$$\partial\delta_{[-1,1]}(t) = \begin{cases} [0,\infty) & \text{falls } t = 1, \\ (-\infty, 0] & \text{falls } t = -1, \\ \{0\} & \text{falls } t \in (-1,1), \\ \emptyset & \text{falls } t \in \mathbb{R} \setminus [-1,1]. \end{cases}$$

Dies entspricht den aus der linearen Optimierung bekannten *Komplementaritätsbedingungen* für Lagrange-Multiplikatoren zu den Ungleichungen $-1 \leq t \leq 1$.

Für konvexe Integralfunktionale lässt sich das Subdifferential punktweise charakterisieren. Im Folgenden sei stets $\Omega \subset \mathbb{R}^n$ beschränkt und $p^{-1} + q^{-1} = 1$, d. h. $L^q(\Omega) \cong L^p(\Omega)^*$ für $1 \leq p < \infty$, angenommen.

Lemma 4.6. *Sei $f : \mathbb{R} \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig und $F : L^p(\Omega) \to \overline{\mathbb{R}}$ mit $1 \leq p < \infty$ wie in Lemma 3.4. Dann ist für alle $u \in \operatorname{dom} F$ für $q = \frac{p}{p-1}$*

$$\partial F(u) = \{u^* \in L^q(\Omega) : u^*(x) \in \partial f(u(x)) \quad \textit{für fast alle } x \in \Omega\}\,.$$

*Beweis.* Sei $u \in \operatorname{dom} F$, d. h. $f \circ u \in L^1(\Omega)$. Gilt für $u^* \in L^q(\Omega)$ nun $u^*(x) \in \partial f(u(x))$ fast überall, so folgt daraus sofort für alle $\tilde{u} \in L^p(\Omega)$ durch Integration

$$F(\tilde{u}) - F(u) = \int_\Omega f(\tilde{u}(x)) - f(u(x))\,dx \geq \int_\Omega u^*(x)(\tilde{u}(x) - u(x))\,dx = \langle u^*, \tilde{u} - u\rangle_{L^p}.$$

Sei umgekehrt $u^* \in \partial F(u)$. Dann gilt nach Definition

$$\int_\Omega u^*(x)(\tilde{u}(x) - u(x))\,dx \leq \int_\Omega f(\tilde{u}(x)) - f(u(x))\,dx \quad \text{für alle } \tilde{u} \in L^p(\Omega).$$

Sei nun $t \in \mathbb{R}$ beliebig und sei $A \subset \Omega$ eine beliebige messbare Menge. Für die Wahl

$$\tilde{u}(x) := \begin{cases} t & \text{falls } x \in A, \\ u(x) & \text{falls } x \notin A, \end{cases}$$

folgt wegen $\tilde{u} \in L^p(\Omega)$ aus der obigen Ungleichung

$$\int_A u^*(x)(t - u(x))\,dx \leq \int_A f(t) - f(u(x))\,dx.$$

Da $A$ beliebig war, muss gelten

$$u^*(x)(t - u(x)) \leq f(t) - f(u(x)) \qquad \text{für fast alle } x \in \Omega.$$

Da $t \in \mathbb{R}$ beliebig war, folgt daraus $u^*(x) \in \partial u(x)$ für fast alle $x \in \Omega$. □

Analog zeigt man für $F : \mathbb{R}^N \to \overline{\mathbb{R}}$ mit $F(x) = \sum_{i=1}^N f_i(x_i)$ und $f_i : \mathbb{R} \to \overline{\mathbb{R}}$ konvex, dass für $x \in \operatorname{dom} F$ gilt

$$\partial F(x) = \left\{\xi \in \mathbb{R}^N : \xi_i \in \partial f_i(x_i), \quad 1 \leq i \leq N\right\}.$$

Zusammen mit den obigen Beispielen erhält man daraus die Subdifferentiale der Norm $\|\cdot\|_1$ sowie der Indikatorfunktion der Einheitskugel bezüglich der Supremumsnorm in $\mathbb{R}^N$.

Subdifferentiale weiterer Funktionale erhält man durch Rechenregeln. Es ist naheliegend, dass diese umso aufwändiger zu beweisen sind, je schwächer der Differenzierbarkeitsbegriff ist (d. h. je mehr Funktionen in diesem Sinne differenzierbar sind). Die ersten beiden Regeln folgen noch direkt aus der Definition.

Lemma 4.7. *Für $F : X \to \overline{\mathbb{R}}$ konvex und $x \in \operatorname{dom} F$ gilt*

(i) $\partial(\lambda F)(x) = \lambda(\partial F(x)) := \{\lambda\xi : \xi \in \partial F(x)\}$ *für* $\lambda \geq 0$;

(ii) $\partial F(\cdot + x_0)(x) = \partial F(x + x_0)$ *für* $x_0 \in X$ *mit* $x + x_0 \in \operatorname{dom} F$.

Schon die Summenregel ist deutlich aufwändiger.

Satz 4.8 (Summenregel). *Seien $F, G : X \to \overline{\mathbb{R}}$ konvex. Dann gilt für alle $x \in \operatorname{dom} F \cap \operatorname{dom} G$*

$$\partial F(x) + \partial G(x) \subset \partial(F + G)(x).$$

*Existiert ein $x_0 \in (\operatorname{dom} F)^o \cap \operatorname{dom} G$, so gilt Gleichheit.*

*Beweis.* Die Inklusion folgt direkt aus der Definition des Subdifferentials durch Addition. Seien daher $x \in \operatorname{dom} F \cap \operatorname{dom} G$ und $\xi \in \partial(F + G)(x)$, erfüllen also

$$\langle \xi, \tilde{x} - x \rangle_X \leq (F(\tilde{x}) + G(\tilde{x})) - (F(x) + G(x)) \quad \text{für alle } \tilde{x} \in X. \tag{4.6}$$

Unser Ziel ist nun, ähnlich wie im Beweis von Lemma 3.6 mit Hilfe der Charakterisierung konvexer Funktionale durch ihren Epigraphen und des Trennungssatzes ein lineares Funktional $\zeta \in \partial G(x) \subset X^*$ zu finden mit $\xi - \zeta \in \partial F(x)$, d. h.

$$\begin{aligned} F(\tilde{x}) - F(x) - \langle \xi, \tilde{x} - x \rangle_X &\geq \langle \zeta, x - \tilde{x} \rangle_X \quad && \text{für alle } \tilde{x} \in \operatorname{dom} F, \\ G(x) - G(\tilde{x}) &\leq \langle \zeta, x - \tilde{x} \rangle_X && \text{für alle } \tilde{x} \in \operatorname{dom} G. \end{aligned}$$

Wir definieren dafür die Mengen

$$C_1 := \{(\tilde{x}, t - (F(x) - \langle \xi, x \rangle_X)) : F(\tilde{x}) - \langle \xi, \tilde{x} \rangle_X \leq t\}, \tag{4.7}$$

$$C_2 := \{(\tilde{x}, G(x) - t) : G(\tilde{x}) \leq t\}, \tag{4.8}$$

d. h.

$$C_1 = \operatorname{epi}(F - \xi) - (0, F(x) - \langle \xi, x \rangle_X), \qquad C_2 = -(\operatorname{epi} G - (0, G(x))).$$

Da es sich um verschobene (und, für $C_2$, gespiegelte) Epigraphen eigentlicher konvexer Funktionen (lineare Funktionale sind konvex) handelt, sind $C_1$ und $C_2$ nichtleer und konvex. Weiter ist $x_0$ ein innerer Punkt von $\operatorname{dom} F = \operatorname{dom}(F - \xi)$ und damit $(x_0, \alpha)$ für $\alpha$ groß genug ein innerer Punkt von $C_1$. Das Innere $(C_1)^o$ von $C_1$ ist also nichtleer. Bleibt zu zeigen, dass $(C_1)^o$ und $C_2$ disjunkt sind. Dafür sei $(\tilde{x}, \alpha) \in (C_1)^o \cap C_2$, für das nach Definition gilt

$$F(\tilde{x}) - F(x) - \langle \xi, \tilde{x} - x \rangle_X < \alpha \leq G(x) - G(\tilde{x}),$$

im Widerspruch zu (4.6). Folgerung 1.6 liefert also ein $(x^*, s) \in (X \times \mathbb{R})^* \setminus \{(0, 0)\}$ und ein $\lambda \in \mathbb{R}$ mit

$$\langle x^*, \tilde{x} \rangle_X + s(t - (F(x) - \langle \xi, x \rangle_X)) \leq \lambda, \quad \tilde{x} \in \operatorname{dom} F, t \geq F(\tilde{x}) - \langle \xi, \tilde{x} \rangle_X, \tag{4.9}$$

$$\langle x^*, \tilde{x} \rangle_X + s(G(x) - t) \geq \lambda, \quad \tilde{x} \in \operatorname{dom} G, t \geq G(\tilde{x}). \tag{4.10}$$

Wir zeigen nun, dass $s < 0$ ist. Für $s = 0$ folgt mit $\tilde{x} = x_0 \in \operatorname{dom} F \cap \operatorname{dom} G$ sofort der Widerspruch

$$\langle x^*, x_0 \rangle_X < \lambda \leq \langle x^*, x_0 \rangle_X,$$

da $(x_0, \alpha)$ für $\alpha$ groß genug innerer Punkt von $C_1$ ist und daher nach Satz 1.5 die Trennung sogar mit strikter Ungleichung erfüllt ist. Gilt $s > 0$, so ist für $t > F(x) - \langle \xi, x \rangle_X$ die Klammer in (4.9) positiv, und $t \to \infty$ mit $\tilde{x}$ fest führt zum Widerspruch zur Beschränktheit durch $\lambda$.

Also ist $s < 0$, und aus (4.9) mit $t = F(\tilde{x}) - \langle \xi, \tilde{x} \rangle_X$ und aus (4.10) mit $t = G(\tilde{x})$ folgt

$$F(\tilde{x}) - F(x) + \langle \xi, \tilde{x} - x \rangle_X \geq s^{-1}(\lambda - \langle x^*, \tilde{x} \rangle_X), \quad \text{für alle } \tilde{x} \in \operatorname{dom} F, \tag{4.11}$$

$$G(x) - G(\tilde{x}) \leq s^{-1}(\lambda - \langle x^*, \tilde{x} \rangle_X), \quad \text{für alle } \tilde{x} \in \operatorname{dom} G. \tag{4.12}$$

Setzen wir $\tilde{x} = x \in \operatorname{dom} F \cap \operatorname{dom} G$ in beiden Ungleichungen, so folgt sofort $\lambda = \langle x^*, x \rangle_X$. Damit ist $\zeta = s^{-1} x^*$ das gewünschte Funktional mit $(\xi - \zeta) \in \partial F(x)$ und $\zeta \in \partial G(x)$, d. h. $\xi \in \partial F(x) + \partial G(x)$. □

Daraus erhält man per Induktion Summenregeln für beliebig viele Summanden (wobei $x_0$ im Inneren aller bis auf eines effektiven Definitionsbereichs liegen muss). Analog beweist man eine Kettenregel für lineare Operatoren.

**Satz 4.9 (Kettenregel).** *Seien $X, Y$ normierte Räume, $A \in L(X, Y)$, und $F : Y \to \overline{\mathbb{R}}$ konvex und unterhalbstetig. Dann gilt für alle $x \in \operatorname{dom}(F \circ A)$*

$$\partial(F \circ A)(x) \supset A^* \partial F(Ax) := \{A^* y^* : y^* \in \partial F(Ax)\}.$$

*Existiert ein $x_0 \in X$ mit $Ax_0 \in (\operatorname{dom} F)^o$, so gilt Gleichheit.*

*Beweis.* Die Inklusion folgt wieder direkt aus der Definition: Für $\eta \in \partial F(Ax) \subset Y^*$ gilt insbesondere für alle $\tilde{y} = A\tilde{x} \in Y$ mit $\tilde{x} \in X$, dass gilt

$$F(A\tilde{x}) - F(Ax) \geq \langle \eta, A\tilde{x} - Ax \rangle_Y = \langle A^* \eta, \tilde{x} - x \rangle_X,$$

d. h. $\xi := A^* \eta \in \partial(F \circ A) \subset X^*$.

Sei nun $x \in \operatorname{dom}(F \circ A)$ und $\xi \in \partial(F \circ A)(x)$, d. h.

$$F(Ax) + \langle \xi, \tilde{x} - x \rangle_X \leq F(A\tilde{x}) \quad \text{für alle } \tilde{x} \in X. \tag{4.13}$$

Analog zur Summenregel könnten wir nun ein $\eta \in \partial F(Ax)$ mit $\xi = A^* \eta$ durch Trennung von $C_1 := \operatorname{epi} F$ und

$$C_2 := \{(A\tilde{x}, F(Ax) + \langle \xi, \tilde{x} - x \rangle_X) : \tilde{x} \in X\}$$

konstruieren. Der Abwechslung halber (und weil die verwendte Technik des „Liftens“ breiteren Nutzen hat) wenden wir stattdessen die Summenregel an auf

$$H : X \times Y \to \overline{\mathbb{R}}, \qquad H(x, y) := F(y) + \delta_{\operatorname{graph} A}(x, y).$$

Da $A$ linear ist, ist $\operatorname{graph} A$ und damit $\delta_{\operatorname{graph} A}$ konvex. Weiter ist nach Annahme $Ax \in \operatorname{dom} F$ und damit $(x, Ax) \in \operatorname{dom} H$.

Zuerst zeigen wir, dass $\xi \in \partial(F \circ A)(x)$ genau dann gilt, wenn $(\xi, 0) \in \partial H(x, Ax)$ ist. Sei dafür $(\xi, 0) \in \partial H(x, Ax)$. Dann gilt für alle $\tilde{x} \in X$, $\tilde{y} \in Y$

$$\langle \xi, \tilde{x} - x\rangle_X + \langle 0, \tilde{y} - Ax\rangle_Y \le F(\tilde{y}) - F(Ax) + \delta_{\operatorname{graph} A}(\tilde{x}, \tilde{y}) - \delta_{\operatorname{graph} A}(x, Ax).$$

Insbesondere gilt dies für alle $\tilde{y} \in \operatorname{ran}(A) = \{A\tilde{x} : \tilde{x} \in X\}$. Also ist wegen $\delta_{\operatorname{graph} A}(\tilde{x}, A\tilde{x}) = 0$

$$\langle \xi, \tilde{x} - x\rangle_X \le F(A\tilde{x}) - F(Ax) \quad \text{für alle } \tilde{x} \in X,$$

d. h. $\xi \in \partial(F \circ A)(x)$. Sei umgekehrt $\xi \in \partial(F \circ A)(x)$. Dann ist für alle $\tilde{x} \in X$ und $\tilde{y} \in Y$ wegen $\delta_{\operatorname{graph} A}(x, Ax) = 0$ und $\delta_{\operatorname{graph} A}(\tilde{x}, \tilde{y}) \ge 0$

$$\begin{aligned}
\langle \xi, \tilde{x} - x\rangle_X + \langle 0, \tilde{y} - Ax\rangle_Y &= \langle \xi, \tilde{x} - x\rangle_X \\
&\le F(A\tilde{x}) - F(Ax) + \delta_{\operatorname{graph} A}(\tilde{x}, \tilde{y}) - \delta_{\operatorname{graph} A}(x, Ax) \\
&= F(\tilde{y}) - F(Ax) + \delta_{\operatorname{graph} A}(\tilde{x}, \tilde{y}) - \delta_{\operatorname{graph} A}(x, Ax),
\end{aligned}$$

da für $\tilde{y} \ne A\tilde{x}$ beide Seiten der letzten Gleichung unendlich sind. Also ist $(\xi, 0) \in \partial H(x, Ax)$.

Da $Ax_0 \in (\operatorname{dom} F)^o$, ist $(x_0, Ax_0) \in (\operatorname{graph} A)^o = (\operatorname{dom} \delta_{\operatorname{graph} A})^o$. Wir können daher Satz 4.8 anwenden und erhalten

$$(\xi, 0) \in \partial H(x, Ax) = \partial F(Ax) + \partial \delta_{\operatorname{graph} A}(x, Ax).$$

Also ist $(\xi, 0) = (x^*, y^*) + (w^*, z^*)$ für ein $(x^*, y^*) \in \partial F(Ax)$ (mit $F$ wieder aufgefasst als $(x, y) \mapsto F(y)$) und ein $(w^*, z^*) \in \partial \delta_{\operatorname{graph} A}(x, Ax)$.

Nun ist $(x^*, y^*) \in \partial F(Ax)$ genau dann, wenn gilt

$$\langle x^*, \tilde{x} - x\rangle_X + \langle y^*, \tilde{y} - Ax\rangle_Y \le F(\tilde{y}) - F(Ax) \quad \text{für alle } \tilde{x} \in X, \tilde{y} \in Y.$$

Festhalten von $\tilde{x} = x$ bzw. $\tilde{y} = Ax$ liefert $y^* \in \partial F(Ax)$ (wobei nun wieder $F$ als Abbildung auf $Y$ aufgefasst wird) und $x^* = 0$. Weiter ist $(w^*, z^*) \in \partial \delta_{\operatorname{graph} A}(x, Ax)$ genau dann, wenn

$$\langle w^*, \tilde{x} - x\rangle_X + \langle z^*, \tilde{y} - Ax\rangle_Y \le 0 \quad \text{für alle } (\tilde{x}, \tilde{y}) \in \operatorname{graph} A,$$

d. h. für alle $\tilde{x} \in X$ und $\tilde{y} = A\tilde{x}$. Also ist

$$\langle w^* + A^* z^*, \tilde{x} - x\rangle_X \le 0 \quad \text{für alle } \tilde{x} \in X$$

und damit $w^* = -A^* z^*$. Zusammen erhalten wir

$$(\xi, 0) = (0, y^*) + (-A^* z^*, z^*),$$

woraus $y^* = -z^*$ und daher $\xi = -A^* z^* = A^* y^*$ mit $y^* \in \partial F(Ax)$ folgt, was zu zeigen war. □

Damit erhalten wir eine Charakterisierung von Minimierern konvexer Funktionen unter (konvexen) Nebenbedingungen.

Folgerung 4.10. *Sei $U \subset X$ nichtleer, konvex und abgeschlossen, und sei $F : X \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig. Existiert ein $x_0 \in U^o \cap \operatorname{dom} F$, so ist $\bar{x} \in U$ Lösung von*

$$\min_{x\in U} F(x)$$

*genau dann, wenn ein $\xi \in X^*$ existiert mit*

$$\begin{cases} -\xi \in \partial F(\bar{x}), \\ \langle \xi, \tilde{x} - x\rangle \leq 0 \quad \textit{für alle } \tilde{x} \in U. \end{cases} \tag{4.14}$$

*Beweis.* Aufgrund der Voraussetzungen an $F$ und $U$ können wir Satz 4.3 auf $J := F + \delta_U$ anwenden; und da $x_0 \in U^o = (\operatorname{dom} \delta_U)^o$ ist, können wir auch die Summenregel anwenden. Also hat $F$ in $\bar{x}$ ein Minimum genau dann, wenn gilt

$$0 \in \partial J(\bar{x}) = \partial F(\bar{x}) + \partial\delta_U(\bar{x}).$$

Zusammen mit der Charakterisierung des Subdifferentials der Indikatorfunktion als Normalenkegel erhält man damit (4.14). □

Für eine Gâteaux-differenzierbare (und damit stetige) Funktion $F : X \to \mathbb{R}$ ergibt (4.14) die klassischen *Karush–Kuhn–Tucker-Bedingungen*; die Existenz des inneren Punktes $x_0 \in U^o$ entspricht dabei genau der *Slater-Bedingung*.

# 5 FENCHEL-DUALITÄT

Ein Grund für die Nützlichkeit des konvexen Subdifferentials ist, wie wir sehen werden, seine Verbindung mit der Fenchel–Legendre-Transformation. Sei $X$ ein normierter Raum und $F : X \to \overline{\mathbb{R}}$ eigentlich. Dann ist die *Fenchel-Konjugierte* zu $F$ definiert als

$$F^* : X^* \to \overline{\mathbb{R}}, \qquad F^*(x^*) = \sup_{x \in X} \langle x^*, x \rangle_X - F(x).$$

(Da $\operatorname{dom} F \neq \emptyset$ angenommen ist, gilt $F^*(x^*) > -\infty$ für alle $x^* \in X^*$, also ist die Definition sinnvoll.) Aus Lemma 3.5 (v) und Lemma 2.2 (v) folgt sofort, dass $F^*$ für eigentliche $F$ stets konvex und schwach-$*$ unterhalbstetig ist. Ist $F$ nach unten durch ein affin-lineares Funktional beschränkt (was nach Lemma 3.6 für konvexe und unterhalbstetige Funktionen immer der Fall ist), so ist auch $F^*$ eigentlich. Die Definition liefert außerdem sofort die *Fenchel–Young-Ungleichung*

$$\langle x^*, x \rangle_X \leq F(x) + F^*(x^*) \qquad \text{für alle } x \in X, x^* \in X^*. \tag{5.1}$$

Anschaulich ist $F^*(x^*)$ der (negative) affine Anteil der Tangente an $F$ (im Punkt $x$, in dem das Supremum angenommen wird) mit der Steigung $x^*$. Analog definieren wir für $F : X^* \to \overline{\mathbb{R}}$ die *Fenchel-Präkonjugierte* als

$$F_* : X \to \overline{\mathbb{R}}, \qquad F_*(x) = \sup_{x^* \in X^*} \langle x^*, x \rangle_X - F(x^*).$$

Durch diese Konvention ist die *Bikonjugierte* $F^{**} := (F^*)_*$ selbst für nicht-reflexive Räume wieder auf $X$ definiert (anstatt auf $X^{**}$). Anschaulich ist $F^{**}$ die untere konvexe Hülle von $F$, die für konvexe Funktionen nach Lemma 3.6 ja mit $F$ übereinstimmt.

**Satz 5.1 (Fenchel–Moreau–Rockafellar).** *Sei $F : X \to \overline{\mathbb{R}}$ eigentlich. Dann gilt*

*(i)* $F^{**} \leq F$;

*(ii)* $F^{**} = F^\Gamma$;

*(iii)* $F^{**} = F$ *genau dann, wenn $F$ konvex und unterhalbstetig ist.*

*Beweis.* Für Aussage (i) nehmen wir in der Fenchel–Young-Ungleichung (5.1) das Supremum über alle $x^* \in X^*$ und erhalten

$$F(x) \geq \sup_{x^* \in X^*} \langle x^*, x \rangle_X - F^*(x^*) = F^{**}(x).$$

Für (ii) stellen wir zuerst fest, dass $F^{**}$ nach Definition der Fenchel-Konjugierten konvex und unterhalbstetig und wegen (i) auch eigentlich ist. Also gilt nach Lemma 3.6

$$F^{**}(x) = (F^{**})^\Gamma(x) = \sup \left\{ a(x) : a : X \to \mathbb{R} \text{ stetig affin mit } a \leq F^{**} \right\}.$$

Wir zeigen nun, dass wir auf der rechten Seite $F^{**}$ durch $F$ ersetzen können. Sei dafür $a(x) = \langle x^*, x \rangle_X - \alpha$ für $x^* \in X^*$ und $\alpha \in \mathbb{R}$ beliebig. Gilt $a \leq F^{**}$, so folgt aus (i) sofort $a \leq F$. Gilt umgekehrt $a \leq F$, so ist $\langle x^*, x \rangle_X - F(x) \leq \alpha$ für alle $x \in X$, und durch Supremum über alle $x \in X$ erhalten wir $\alpha \geq F^*(x^*)$. Daraus folgt nun nach Definition von $F^{**}$

$$a(x) = \langle x^*, x \rangle_X - \alpha \leq \langle x^*, x \rangle_X - F^*(x^*) \leq F^{**}(x) \quad \text{für alle } x \in X,$$

d. h. $a \leq F^{**}$.

Aussage (iii) folgt nun sofort aus (ii) und Lemma 3.6. □

Wir betrachten wieder relevante Beispiele.

**Beispiel 5.2.**

(i) Sei $X$ ein Hilbertraum und $F(x) = \frac{1}{2}\|x\|_X^2$. Wir identifizieren $X$ über den Satz von Fréchet–Riesz mit seinem Dualraum $X^*$, so dass die duale Paarung durch das Skalarprodukt ausgedrückt werden kann. Da $F$ Fréchet-differenzierbar ist mit Gradient $\nabla F(x) = x$, muss die Lösung $\bar{x}$ von

$$\sup_{x \in X} (x^*, x)_X - \frac{1}{2}(x, x)_X$$

das Fermat-Prinzip erfüllen, d. h. $x^* = \bar{x}$. Einsetzen und vereinfachen liefert die Konjugierte

$$F^* : X \to \mathbb{R}, \qquad F^*(x^*) = \frac{1}{2}\|x^*\|_X^2.$$

(ii) Sei $B_X$ die Einheitskugel im normierten Raum $X$ und setze $F = \delta_{B_X}$. Dann ist für $x^* \in X^*$

$$(\delta_{B_X})^*(x^*) = \sup_{x \in X} \langle x^*, x \rangle_X - \delta_{B_X}(x) = \sup_{\|x\|_X \leq 1} \langle x^*, x \rangle_X = \|x^*\|_{X^*}.$$

Analog zeigt man unter Verwendung der Definition der Präkonjugierten und Folgerung 1.7, dass $(\delta_{B_{X^*}})_*(x) = \|x\|_X$ ist.

(iii) Sei $X$ ein normierter Raum und setze $F(x) = \|x\|_X$. Für $x^* \in X^*$ unterscheiden wir nun zwei Fälle:

a) $\|x^*\|_{X^*} \leq 1$. Dann gilt mit (1.2) für alle $x \in X$ dass $\langle x^*, x\rangle_X - \|x\|_X \leq 0$ ist. Weiterhin ist $\langle x^*, 0\rangle = 0 = \|0\|_X$. Also gilt

$$F^*(x^*) = \sup_{x\in X} \langle x^*, x\rangle_X - \|x\|_X = 0.$$

b) $\|x^*\|_{X^*} > 1$. Nach Definition der Norm im Dualraum existiert dann ein $x_0 \in X$ mit $\langle x^*, x_0\rangle_X > \|x_0\|_X$. Lassen wir daher $t \to \infty$ gehen in

$$0 < t(\langle x^*, x_0\rangle_X - \|x_0\|_X) = \langle x^*, tx_0\rangle_X - \|tx_0\|_X \leq F^*(x^*),$$

so erhalten wir $F^*(x^*) = \infty$.

Zusammen ergibt dies $F^* = \delta_{B_{X^*}}$. Wie oben zeigt man analog, dass $(\|\cdot\|_{X^*})_* = \delta_{B_X}$ ist.

Analog zu Lemma 4.6 zeigt man, dass sich auch die Fenchel-Konjugierte punktweise berechnen lässt; siehe z. B. [Ekeland & Témam 1999, Proposition IV.1.2].

**Lemma 5.3.** *Sei $f : \mathbb{R} \to \overline{\mathbb{R}}$ eigentlich und unterhalbstetig, und sei $F : L^p(\Omega) \to \overline{\mathbb{R}}$ mit $1 \leq p < \infty$ wie in Lemma 3.4. Dann ist für $q = \frac{p}{p-1}$*

$$F^* : L^q(\Omega) \to \overline{\mathbb{R}}, \qquad F^*(u^*) = \int_\Omega f^*(u^*(x))\, dx.$$

Wir notieren noch einige nützliche Rechenregeln.

**Lemma 5.4.** *Sei $F : X \to \overline{\mathbb{R}}$ eigentlich. Dann ist*

*(i)* $(\alpha F)^* = \alpha F^* \circ (\alpha^{-1} \mathrm{Id})$ *für* $\alpha > 0$;

*(ii)* $(F(\cdot + x_0) + \langle x_0^*, \cdot\rangle_X)^* = F^*(\cdot - x_0^*) - \langle \cdot - x_0^*, x_0\rangle_X$ *für alle* $x_0 \in X$, $x_0^* \in X^*$;

*(iii)* $(F \circ A)^* = F^* \circ A^{-*}$ *für* $A \in L(Y, X)$ *stetig invertierbar und* $A^{-*} := (A^{-1})^*$.

*Beweis.* Die Regeln folgen direkt aus den Eigenschaften des Supremums.

Aussage (i) gilt wegen $\alpha > 0$ und

$$(\alpha F)^*(x^*) = \sup_{x\in X} \left(\alpha\langle \alpha^{-1}x^*, x\rangle_X - \alpha F(x)\right) = \alpha \sup_{x\in X} \left(\langle \alpha^{-1}x^*, x\rangle_X - F(x)\right) = \alpha F^*(\alpha^{-1}x^*).$$

Aussage (ii) gilt wegen $\{x + x_0 : x \in X\} = X$ und

$$\begin{aligned}
(F(\cdot + x_0) + \langle x_0^*, \cdot\rangle_X)^*(x^*) &= \sup_{x\in X} \langle x^*, x\rangle_X - F(x^* + x_0) - \langle x_0^*, x\rangle_X \\
&= \sup_{x\in X} \left(\langle x^* - x_0^*, x + x_0\rangle_X - F(x + x_0)\right) - \langle x^* - x_0^*, x_0\rangle_X \\
&= \sup_{\tilde{x}=x+x_0, x\in X} \left(\langle x^* - x_0^*, \tilde{x}\rangle_X - F(\tilde{x})\right) - \langle x^* - x_0^*, x_0\rangle_X \\
&= F^*(x^* - x_0^*) - \langle x^* - x_0^*, x_0\rangle_X.
\end{aligned}$$

Aussage (iii) gilt wegen $X = \operatorname{ran} A$ und

$$\begin{aligned}
(F \circ A)^*(y^*) &= \sup_{y\in Y} \langle y^*, A^{-1}Ay\rangle_Y - F(Ay) \\
&= \sup_{x=Ay, y\in Y} \langle A^{-*}y^*, x\rangle_X - F(x) = F^*(A^{-*}y^*).
\end{aligned}$$

□

Die Definition der Fenchel-Konjugierten ist auf besondere Weise verträglich mit der des Subdifferentials.

Lemma 5.5. *Sei $F : X \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig. Dann sind äquivalent für $x \in X$ und $x^* \in X^*$:*

*(i)* $\langle x^*, x\rangle_X = F(x) + F^*(x^*)$;

*(ii)* $x^* \in \partial F(x)$;

*(iii)* $x \in \partial F^*(x^*)$.

*Beweis.* Gilt (i), so folgt aus der Definition von $F^*$ als Supremum

$$\langle x^*, x\rangle_X - F(x) = F^*(x^*) \geq \langle x^*, \tilde{x}\rangle_X - F(\tilde{x}) \qquad \text{für alle } \tilde{x} \in X, \tag{5.2}$$

was nach Definition äquivalent ist zu $x^* \in \partial F(x)$. Umgekehrt ergibt Supremum über alle $\tilde{x} \in X$ auf beiden Seiten von (5.2)

$$\langle x^*, x\rangle_X \geq F(x) + F^*(x^*),$$

und zusammen mit der Fenchel–Young-Ungleichung (5.1) folgt (i).

Analog erhält man aus (i) zusammen mit Satz 5.1 für alle $\tilde{x}^* \in X^*$ die Ungleichung

$$\langle x^*, x\rangle_X - F^*(x^*) = F(x) = F^{**}(x) \geq \langle \tilde{x}^*, x\rangle - F^*(\tilde{x}^*),$$

woraus wie oben die Äquivalenz von (i) und (iii) folgt. □

Bemerkung. Ist $X$ nicht reflexiv, so ist dabei $x \in \partial F^*(x^*) \subset X^{**}$ über die kanonische Injektion aufzufassen, d. h. via

$$\langle J(x), \tilde{x}^* - x^* \rangle_{X^*} = \langle \tilde{x}^* - x^*, x \rangle_X \le F^*(\tilde{x}^*) - F^*(x^*) \quad \text{für alle } \tilde{x}^* \in X.$$

Gleichheit in der Fenchel–Young-Ungleichung (oder äquivalent die Subdifferentialinklusion (ii)) garantiert also in (iii) die Existenz eines Subgradienten in $X \subset X^{**}$ und nicht nur in $X^{**}$; umgekehrt ist in (iii) die Existenz eines Subgradienten in $X \subset X^{**}$ notwendig für die Gleichheit (und damit (ii)). (Analoges gilt für $F : X^* \to \overline{\mathbb{R}}$ und $F_* : X \to \overline{\mathbb{R}}$.)

Lemma 5.5 spielt die Rolle des „Satzes von der konvexen Umkehrfunktion“. Damit kann man insbesondere das Subdifferential einer komplizierten Norm durch das (einfachere) der konjugierten Indikatorfunktion ersetzen. Hat man zum Beispiel ein Problem der Form

$$\text{(P)} \qquad \inf_{x \in X} F(x) + G(Ax)$$

für $F : X \to \overline{\mathbb{R}}$ und $G : Y \to \overline{\mathbb{R}}$ eigentlich, konvex, und unterhalbstetig sowie $A \in L(X, Y)$, so können wir $G$ mit Hilfe von Satz 5.1 ersetzen durch die Definition von $G^{**}$ und erhalten

$$\text{(5.3)} \qquad \inf_{x \in X} \sup_{Y^* \in Y^*} F(x) + \langle y^*, Ax \rangle_Y - G^*(y^*).$$

Dürften wir nun inf und sup vertauschen, so könnten wir schreiben (mit $\inf F = -\sup(-F)$)

$$\begin{aligned} \inf_{x \in X} \sup_{y^* \in y^*} F(x) + \langle y^*, Ax \rangle_Y - G^*(y^*) &= \sup_{y^* \in Y^*} \inf_{x \in X} F(x) + \langle y^*, Ax \rangle_Y - G^*(y^*) \\ &= \sup_{y^* \in Y^*} - \left( \sup_{x \in X} -F(x) + \langle -A^* y^*, x \rangle_X \right) - G^*(y^*). \end{aligned}$$

Einsetzen der Definition von $F^*$ ergibt dann das *duale Problem*

$$\text{(D)} \qquad \sup_{y^* \in Y^*} -F^*(-A^* y^*) - G^*(y^*).$$

Als Nebeneffekt haben wir den Operator $A$ zwischen den Funktionalen verschoben.

Der folgende Satz nutzt auf elegante Weise das Fermat-Prinzip, Summen- und Kettenregel sowie die Fenchel–Young-Gleichung, um hinreichende Bedingungen für die Vertauschbarkeit zu geben.

Satz 5.6 (Fenchel–Rockafellar). *Seien $X, Y$ normierte Räume, $F : X \to \overline{\mathbb{R}}$ und $G : Y \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig, und $A \in L(X, Y)$. Gelte weiterhin:*

*(i) das* primale Problem (P) *hat eine Lösung $\bar{x} \in X$;*

*(ii) es existiert ein $x_0 \in \operatorname{dom} F \cap \operatorname{dom}(G \circ A)$ mit $Ax_0 \in (\operatorname{dom} G)^o$.*

*Dann hat das duale Problem (D) eine Lösung $\bar{y}^* \in Y^*$ und es gilt*

$$\min_{x\in X} F(x) + G(Ax) = \max_{y^*\in Y^*} -F^*(-A^*y^*) - G^*(y^*). \tag{5.4}$$

*Weiterhin sind $\bar{x}$ und $\bar{y}^*$ Lösungen von (P) bzw. (D) genau dann, wenn die* Fenchel-Extremalitätsbedingungen

$$\begin{cases} -A^*\bar{y}^* \in \partial F(\bar{x}), \\ \quad\ \ \bar{y}^* \in \partial G(A\bar{x}), \end{cases} \tag{E}$$

*erfüllt sind.*

*Beweis.* Sei zuerst $\bar{x} \in X$ eine Lösung von (P). Wegen Voraussetzung (ii) sind Satz 4.8 und Satz 4.9 anwendbar; aus Satz 4.3 folgt daher

$$0 \in \partial(F + G \circ A)(\bar{x}) = \partial F(\bar{x}) + A^*\partial G(A\bar{x})$$

und damit die Existenz eines $\bar{y}^* \in \partial G(A\bar{x})$ mit $-A^*\bar{y}^* \in \partial F(\bar{x})$, d. h. (E) sind erfüllt.

Es gelten umgekehrt (E) für $\bar{x} \in X$ und $\bar{y}^* \in Y^*$. Dann ist wiederum aufgrund der Sätze 4.3, 4.8 und 4.9 $\bar{x}$ Lösung von (P). Weiter folgt aus (E) mit Lemma 5.5 die Gleichheit in den Fenchel–Youngschen Ungleichungen für $F$ und $G$, d. h.

$$\begin{cases} \langle -A^*\bar{y}^*, \bar{x}\rangle_X = F(\bar{x}) + F^*(-A^*\bar{y}^*), \\ \quad\ \langle \bar{y}^*, A\bar{x}\rangle_Y = G(A\bar{x}) + G^*(\bar{y}^*). \end{cases} \tag{5.5}$$

Durch Summieren beider Gleichungen erhalten wir

$$F(\bar{x}) + G(A\bar{x}) = -F^*(-A^*\bar{y}^*) - G^*(\bar{y}^*). \tag{5.6}$$

Es bleibt zu zeigen, dass $\bar{y}^*$ Lösung von (D) ist. Setze dafür

$$L : X \times Y^* \to \overline{\mathbb{R}}, \qquad L(x, y^*) = F(x) + \langle y^*, Ax\rangle_Y - G^*(y^*).$$

Dann gilt für alle $\tilde{x} \in X$ und $\tilde{y}^* \in Y^*$ stets

$$\sup_{y^*\in Y^*} L(\tilde{x}, y^*) \geq L(\tilde{x}, \tilde{y}^*) \geq \inf_{x\in X} L(x, \tilde{y}^*),$$

und damit (Infimum über alle $\tilde{x}$ in der ersten und Supremum über alle $\tilde{y}^*$ in der zweiten Ungleichung)

$$\inf_{x\in X} \sup_{y^*\in Y^*} L(x, y^*) \geq \sup_{y^*\in Y^*} \inf_{x\in X} L(x, y^*).$$

Also ist

$$\begin{aligned} F(\bar{x}) + G(A\bar{x}) &= \inf_{x\in X} \sup_{Y^*\in Y^*} F(x) + \langle y^*, Ax\rangle_Y - G^*(y^*) \\ &\geq \sup_{Y^*\in Y^*} \inf_{x\in X} F(x) + \langle y^*, Ax\rangle_Y - G^*(y^*) \\ &= \sup_{y^*\in Y^*} -F^*(-A^*y^*) - G^*(y^*). \end{aligned} \tag{5.7}$$

Zusammen mit der Gleichung (5.6) erhalten wir damit

$$-F^*(-A^*\bar{y}^*) - G(\bar{y}^*) = F(\bar{x}) + G(A\bar{x}) \geq \sup_{y^* \in Y^*} -F^*(-A^* y^*) - G^*(y^*),$$

d. h. $\bar{y}^*$ ist Lösung von (D), woraus insbesondere die behauptete Existenz einer Lösung folgt.

Da alle Lösungen von (D) nach Definition den selben (maximalen) Funktionalwert haben, folgt aus (5.6) auch (5.4).

Sind schließlich $\bar{x} \in X$ und $\bar{y}^* \in Y^*$ Lösungen von (P) bzw. (D), so folgt aus der gerade gezeigten starken Dualität die Gleichung (5.6). Zusammen mit der produktiven Null erhalten wir daraus

$$0 = [F(\bar{x}) + F^*(-A^*\bar{y}^*) - \langle -A^*\bar{y}^*, \bar{x}\rangle_X] + [G(A\bar{x}) + G^*(\bar{y}^*) - \langle \bar{y}^*, A\bar{x}\rangle_Y] .$$

Da beide Klammern wegen der Fenchel–Young-Ungleichung nicht-negativ sind, müssen sie einzeln verschwinden. Also gilt (5.5), und aus Lemma 5.5 folgt (E). □

Mit Hilfe von Lemma 5.5 können wir aus den Fenchel-Extremalitätsbedingungen weitere äquivalente Optimalitätsbedingungen erzeugen, indem wir eine oder beide Subdifferentialinklusionen invertieren. Dies werden wir später nutzen, um praktisch durchführbare Algorithmen für die Lösung von Optimierungsproblemen dieser Form herzuleiten. Der Satz erlaubt auch, den Subgradienten $\bar{y}^*$ aus der Kettenregel als Minimierer eines konvexen Funktionals zu charakterisieren. Ist zum Beispiel $F^*$ oder $G^*$ strikt konvex, so ist $\bar{y}^*$ eindeutig; unter dieser Voraussetzung sind oft stärkere Aussagen über die Stabilität von Minimierern von (P) oder die Konvergenz von Verfahren zu deren Berechnung möglich.

# 6 MONOTONE OPERATOREN UND PROXIMALPUNKTE

Ist $\bar{x} \in X$ ein Minimierer der konvexen Funktion $F : X \to \overline{\mathbb{R}}$, so gilt nach Satz 4.3 das Fermat-Prinzip $0 \in \partial F(\bar{x})$. Um daraus Informationen über (und später konkrete Verfahren zur Berechnung von) $\bar{x}$ zu erhalten, müssen wir daher die (mengenwertige!) Abbildung $x \mapsto \partial F(x)$ untersuchen. Um technischen Schwierigkeiten aus dem Weg zu gehen – und da wir die folgenden Resultate primär für numerische Algorithmen, d. h. in $X = \mathbb{R}^N$, benötigen – beschränken wir uns in diesem und dem nächsten Kapitel auf den Fall, dass $X$ ein Hilbertraum ist. Dies erlaubt, $X^*$ mit $X$ identifizieren; insbesondere identifizieren wir $\partial F(x) \subset X^*$ mit der Menge der entsprechenden Riesz-Repräsentanten in $X$.

## 6.1 MONOTONE OPERATOREN

Für zwei normierte Räume $X, Y$ betrachten wir eine *mengenwertige Abbildung* $A : X \to \mathcal{P}(Y)$, geschrieben auch $A : X \rightrightarrows Y$, und definieren

- den *Definitionsbereich* $\operatorname{dom} A = \{x \in X : Ax \neq \emptyset\}$;
- das *Bild* $\operatorname{ran} A = \bigcup_{x \in X} Ax$;
- der *Graph* $\operatorname{graph} A = \{(x, y) \in X \times Y : y \in Ax\}$;
- die *Inverse* $A^{-1} : Y \rightrightarrows X$ durch $A^{-1}(y) = \{x \in X : y \in Ax\}$ für alle $y \in Y$. (Beachten Sie, dass $A^{-1}(y) = \emptyset$ möglich ist; für mengenwertige Abbildungen fallen also Inverse und Urbild – das ja stets existiert – zusammen.)

Für $A, B : X \rightrightarrows Y$, $C : Y \rightrightarrows Z$, und $\lambda \in \mathbb{R}$ definieren wir weiter

- $\lambda A : X \rightrightarrows Y$ durch $(\lambda A)(x) = \{\lambda y : y \in Ax\}$;
- $A + B : X \rightrightarrows Y$ durch $(A + B)(x) = \{y + z : y \in Ax, z \in Bx\}$;
- $C \circ A : X \rightrightarrows Z$ durch $(C \circ A)(x) = \{z : \text{es gibt } y \in Ax \text{ mit } z \in Cy\}$.

Eine mengenwertige Abbildung $A : X \rightrightarrows X$ heißt *monoton*, falls gilt $\operatorname{graph} A \neq \emptyset$ (um den trivialen Fall auszuschliessen) und

$$\left(x_1^* - x_2^*, x_1 - x_2\right)_X \geq 0 \quad \text{für alle } (x_1, x_1^*), (x_2, x_2^*) \in \operatorname{graph} A.$$

Klarerweise ist die Identität $\mathrm{Id} : X \rightrightarrows X, x \mapsto \{x\}$, monoton. Für eine konvexe Funktion $F : X \to \overline{\mathbb{R}}$ ist $\partial F : X \rightrightarrows X, x \mapsto \partial F(x)$ monoton: Sind $x_1, x_2 \in X$ und $x_1^* \in \partial F(x_1)$ und $x_2^* \in \partial F(x_2)$, so gilt nach Definition

$$\left(x_1^*, \tilde{x} - x_1\right) \leq F(\tilde{x}) - F(x_1) \qquad \text{für alle } \tilde{x} \in X, \tag{6.1}$$

$$\left(x_2^*, \tilde{x} - x_2\right) \leq F(\tilde{x}) - F(x_2) \qquad \text{für alle } \tilde{x} \in X, \tag{6.2}$$

woraus nach Addition der ersten Ungleichung mit $\tilde{x} = x_2$ und der zweiten Ungleichung mit $\tilde{x} = x_1$ durch Umformen die Monotonie folgt. Ebenso sind für $A, B$ monoton und $\lambda \geq 0$ auch $\lambda A$ und $A + B$ monoton. Wir werden aber noch eine stärkere Eigenschaft benötigen, die zusätzlich die Abgeschlossenheit von $A$ garantiert: Wir nennen einen monotonen Operator $A : X \rightrightarrows X$ *maximal monoton*, wenn für beliebige $x \in X$ und $x^* \in X$ die Bedingung

$$(x^* - \tilde{x}^*, x - \tilde{x})_X \geq 0 \qquad \text{für alle } (\tilde{x}, \tilde{x}^*) \in \operatorname{graph} A \tag{6.3}$$

impliziert, dass $x^* \in Ax$ ist. (In anderen Worten, (6.3) gilt dann *und nur dann*, wenn $(x, x^*) \in \operatorname{graph} A$.) Für feste $x \in X$ und $x^* \in X$ besagt die Bedingung gerade, dass mit $A$ auch die Erweiterung

$$\tilde{A} : X \rightrightarrows X, \qquad \tilde{x} \mapsto \begin{cases} Ax \cup \{x^*\} & \text{falls } \tilde{x} = x, \\ A\tilde{x} & \text{falls } \tilde{x} \neq x, \end{cases}$$

monoton ist. Die Behauptung ist dann genau, dass dies keine echte Erweiterung darstellt. Beispielsweise ist

$$A : \mathbb{R} \rightrightarrows \mathbb{R}, \qquad t \mapsto \begin{cases} \{1\} & \text{falls } t > 0, \\ \{0\} & \text{falls } t = 0, \\ \{-1\} & \text{falls } t < 0, \end{cases}$$

monoton aber nicht maximal monoton, da $A$ eine echte Teilmenge des durch $\tilde{A}t = \operatorname{sign}(t) = \partial(|\cdot|)(t)$ definierten monotonen Operators ist.

Aus der Definition folgen sofort einige nützliche Aussagen.

**Lemma 6.1.** *Ist $A : X \rightrightarrows X$ maximal monoton, dann ist es auch $\lambda A$ für alle $\lambda > 0$.*

*Beweis.* Seien $x, x^* \in X$ und gelte

$$0 \leq (x^* - \tilde{x}^*, x - \tilde{x})_X = \lambda \left(\lambda^{-1} x^* - \lambda^{-1} \tilde{x}^*, x - \tilde{x}\right)_X \quad \text{für alle } (\tilde{x}, \tilde{x}^*) \in \operatorname{graph} \lambda A.$$

Da für $\tilde{x}^* \in \lambda A x$ gilt $\lambda^{-1}\tilde{x}^* \in Ax$ und $A$ maximal monoton ist, folgt daraus $\lambda^{-1}\bar{x}^* \in A\bar{x}$, d. h. $\bar{x}^* \in (\lambda A)\bar{x}$. Damit ist $\lambda A$ maximal monoton. $\square$

**Lemma 6.2.** *Ist $A : X \rightrightarrows X$ maximal monoton, dann ist $A$ schwach–stark abgeschlossen, d. h. aus $x_n \rightharpoonup x$ und $Ax_n \ni x_n^* \to x^*$ folgt $x^* \in Ax$.*

*Beweis.* Für $\tilde{x} \in X$ und $\tilde{x}^* \in A\tilde{x}$ beliebig gilt wegen der Monotonie von $A$

$$0 \leq \left(x_n^* - \tilde{x}^*, x_n - \tilde{x}\right)_X \to (x^* - \tilde{x}^*, x - \tilde{x})_X\,,$$

da $x_n$ schwach und $x_n^*$ stark in $X$ konvergieren. Da $A$ maximal monoton ist, folgt daraus $x^* \in Ax$. □

Von zentraler Bedeutung für die Theorie der monotonen Operatoren ist der *Satz von Minty*, der besagt, dass ein monotoner Operator $A$ genau dann maximal ist, wenn $\operatorname{Id} + A$ surjektiv ist. Als Vorbereitung beweisen wir eine wichtige Teilaussage.

**Lemma 6.3.** *Für $F : X \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig ist $\operatorname{Id} + \partial F$ surjektiv.*

*Beweis.* Für gegebenes $z \in X$ betrachten wir das Funktional $J : X \to \overline{\mathbb{R}}$,

$$J(x) = \frac{1}{2}\|x - z\|_X^2 + F(x).$$

Mit $F$ ist auch $J$ eigentlich, (strikt) konvex und unterhalbstetig. Weiterhin ist $J$ koerziv, da $F$ als konvexe Funktion wegen Lemma 3.6 nach unten durch eine affine Funktion beschränkt ist, so dass für $\|x\|_X \to \infty$ die quadrierte Norm garantiert „gewinnt". (Man nennt sie daher auch *superkoerziv*.) Nach Satz 3.7 existiert also ein (eindeutiges) $\bar{x} \in X$ mit $J(\bar{x}) = \min_{x \in X} J(x)$, für welches nach Satz 4.3, Satz 4.8 und Satz 4.4 gilt

$$0 \in \partial J(\bar{x}) = \{\bar{x} - z\} + \partial F(\bar{x}),$$

d. h. $z \in \{\bar{x}\} + \partial F(\bar{x}) = (\operatorname{Id} + \partial F)(\bar{x})$. □

Nun zum allgemeinen Fall.

**Satz 6.4 (Minty).** *Sei $A : X \rightrightarrows X$ monoton. Dann ist $A$ maximal monoton genau dann, wenn $\operatorname{Id} + A$ surjektiv ist.*

*Beweis.* Sei zuerst $\operatorname{Id} + A$ surjektiv, und seien $x \in X$ und $x^* \in X$ mit

$$(x^* - \tilde{x}^*, x - \tilde{x})_X \geq 0 \qquad \text{für alle } (\tilde{x}, \tilde{x}^*) \in \operatorname{graph} A. \tag{6.4}$$

Nach Annahme existiert nun für $x + x^* \in X$ ein $z \in X$ und ein $z^* \in Az$ mit

$$x + x^* = z + z^* \in (\operatorname{Id} + A)z. \tag{6.5}$$

Einsetzen von $(\tilde{x}, \tilde{x}^*) = (z, z^*)$ in (6.4) ergibt dann

$$0 \le (x^* - z^*, x - z)_X = (z - x, x - z)_X = -\|x - z\|_X^2 \le 0,$$

d. h. $x = z$. Aus (6.5) folgt dann auch $x^* = z^* \in Az = Ax$, und damit ist $A$ maximal monoton.

Die umgekehrte Richtung ist deutlich aufwändiger. Den Spezialfall $A = \partial F$ für ein konvexes Funktional $F$ haben wir bereits in Lemma 6.3 gezeigt. Für den allgemeinen Fall gehen wir ähnlich vor, indem wir ein Funktional $F_A$ konstruieren, das für $A$ die selbe Rolle spielt wie $F$ für $\partial F$. Für $A : X \rightrightarrows X$ maximal monoton definieren wir dazu die *Fitzpatrick-Funktion*

$$F_A : X \times X \to [-\infty, \infty], \qquad (x, y) \mapsto \sup_{(z,w)\in\operatorname{graph} A} \left((z, y)_X + (x, w)_X - (z, w)_X\right). \tag{6.6}$$

Durch Umformen erhält man die äquivalente Darstellung

$$F_A(x, y) = (x, y)_X - \inf_{(z,w)\in\operatorname{graph} A} (x - z, y - w)_X\,. \tag{6.7}$$

Daraus folgen sofort einige nützliche Eigenschaften.

(i) Wegen der maximalen Monotonie von $A$ gilt nach Definition $(x - z, y - w)_X \ge 0$ für alle $(z, w) \in \operatorname{graph} A$ genau dann, wenn $(x, y) \in \operatorname{graph} A$; insbesondere ist für $(x, y) \notin \operatorname{graph} A$ stets $(x - z, y - w)_X < 0$. Aus (6.7) folgt also $F_A(x, y) \ge (x, y)_X$ mit Gleichheit genau dann, wenn $(x, y) \in \operatorname{graph} A$ (denn in diesem Fall wird das Infimum in $(z, w) = (x, y)$ angenommen). Insbesondere ist $F_A$ eigentlich.

(ii) Aus (6.6) erhält man

$$F_A = (G_A)^* \qquad \text{für } G_A(w, z) = (w, z)_X + \delta_{\operatorname{graph} A^{-1}}(w, z)$$

(denn mit $(z, w) \in \operatorname{graph} A$ ist $(w, z) \in \operatorname{graph} A^{-1}$). Wegen der Monotonie von $A$ ist ebenfalls nach Definition $\operatorname{graph} A \neq \emptyset$; also ist $F_A$ die Fenchel-Konjugierte eines eigentlichen Funktionals und damit konvex und unterhalbstetig.

Als ersten Schritt zeigen wir nun $0 \in \operatorname{ran}(\operatorname{Id} + A)$ und führen den allgemeinen Fall $z \neq 0$ darauf zurück. Wir setzen dafür in Folge $Z := X \times X$ sowie $\xi := (x, y)$ und betrachten

$$J_A : Z \to \overline{\mathbb{R}}, \qquad \xi \mapsto F_A(\xi) + \frac{1}{2}\|\xi\|_Z^2.$$

Zunächst halten wir fest, dass wegen (i) für alle $\xi \in Z$ gilt

$$\begin{aligned} J_A(\xi) = F_A(\xi) + \frac{1}{2}\|\xi\|_Z^2 &= F_A(x, y) + \frac{1}{2}\|x\|_X^2 + \frac{1}{2}\|y\|_X^2 \\ &\ge (x, y)_X + \frac{1}{2}\|x\|_X^2 + \frac{1}{2}\|y\|_X^2 \\ &\ge 0. \end{aligned} \tag{6.8}$$

Weiter ist $J_A$ eigentlich, (strikt) konvex und unterhalbstetig sowie koerziv, denn $F_A$ ist nach (6.6) nach unten durch eine affine Funktion beschränkt und die quadrierte Norm superkoerziv. Nach Satz 3.7 existiert also wieder ein (eindeutiges) $\bar\xi := (\bar x, \bar y) \in Z$ mit $J_A(\bar\xi) = \min_{\xi\in Z} J_A(\xi)$, für welches nach Satz 4.3, Satz 4.8 und Satz 4.4 gilt

$$0 \in \partial J_A(\bar\xi) = \{\bar\xi\} + \partial F_A(\bar\xi),$$

d. h. $-\bar\xi \in \partial F_A(\bar\xi)$. Nach Definition gilt also für alle $\xi \in Z$

$$\begin{aligned} F_A(\xi) \geq F_A(\bar\xi) + \left(-\bar\xi, \xi - \bar\xi\right)_Z &= J_A(\bar\xi) + \frac{1}{2}\|-\bar\xi\|_Z^2 + \left(-\bar\xi, \xi\right)_Z \\ &\geq \frac{1}{2}\|-\bar\xi\|_Z^2 + \left(-\bar\xi, \xi\right)_Z, \end{aligned}$$

wobei wir im letzten Schritt (6.8) verwendet haben. Der Übersichtlichkeit halber ersetzen wir in Folge $\bar\xi \mapsto -\bar\xi$; zusammen mit Eigenschaft (i) folgt dann für alle $(x, y) \in \operatorname{graph} A$

$$\begin{aligned} (x, y)_X = F_A(x, y) &\geq \frac{1}{2}\|\bar x\|_X^2 + (\bar x, x)_X + \frac{1}{2}\|\bar y\|_X^2 + (\bar y, y)_X \\ &\geq -(\bar x, \bar y)_X + (\bar x, x)_X + (\bar y, y)_X \end{aligned} \tag{6.9}$$

und damit $(y - \bar x, x - \bar y)_X \geq 0$. Wegen der maximalen Monotonie von $A$ ist daher $\bar x \in A\bar y$, d. h. $(\bar y, \bar x) \in \operatorname{graph} A$. Einsetzen in die erste Ungleichung von (6.9) ergibt dann

$$(\bar y, \bar x)_X \geq \frac{1}{2}\|\bar x\|_X^2 + (\bar x, \bar y)_X + \frac{1}{2}\|\bar y\|_X^2 + (\bar y, \bar x)_X = \frac{1}{2}\|\bar x + \bar y\|_X^2 + (\bar y, \bar x)_X \geq (\bar y, \bar x)_X$$

und damit $\|\bar x + \bar y\|_X = 0$, d. h. $0 = \bar y + \bar x \in (\mathrm{Id} + A)(\bar y)$.

Sei nun $z \in X$ beliebig und definiere $B : X \rightrightarrows X$, $x \mapsto \{-z\} + Ax$. Man vergewissert sich leicht anhand der Definition, dass mit $A$ auch $B$ maximal monoton ist. Nach dem soeben Bewiesenen existiert also ein $\bar y \in X$ mit $0 \in (\mathrm{Id} + B)(\bar y) = \{\bar y\} + \{-z\} + A\bar y$, d. h. $z \in (\mathrm{Id} + A)(\bar y)$. Also ist $\mathrm{Id} + A$ surjektiv. □

Zusammen mit Lemma 6.3 erhalten wir daraus die maximale Monotonie von konvexen Subdifferentialen.

**Folgerung 6.5.** *Für $F : X \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig ist $\partial F : X \rightrightarrows X$ maximal monoton.*

## 6.2 RESOLVENTEN UND PROXIMALPUNKTE

Nach Lemma 6.3 ist $\mathrm{Id} + \partial F$ surjektiv; da die quadrierte Norm im Hilbertraum und damit $J$ strikt konvex ist, ist der Minimierer von $J$ und damit das Urbild sogar eindeutig. Also ist $(\mathrm{Id} + \partial F)^{-1}$ einwertig, auch wenn $\partial F$ dies nicht ist. Es besteht daher die Hoffnung, dieses

Objekt anstelle des Subdifferentials – oder allgemeiner, eines monotonen Operators – für Algorithmen nutzen zu können.

Wir definieren daher für $A : X \rightrightarrows X$ maximal monoton die *Resolvente*

$$\mathcal{R}_A : X \rightrightarrows X, \qquad \mathcal{R}_A(x) = (\mathrm{Id} + A)^{-1}x,$$

und für $F : X \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig die *Proximalpunktabbildung*

$$\mathrm{prox}_F : X \to X, \qquad \mathrm{prox}_F(x) = \arg\min_{z\in X} \frac{1}{2}\|z - x\|_X^2 + F(z).$$

Da $w \in \mathcal{R}_{\partial F}(x)$ die notwendige und hinreichende Bedingung dafür ist, dass $w$ der eindeutige Minimierer (man sagt auch *Proximalpunkt*) dieser strikt konvexen Funktion ist, gilt

$$\mathrm{prox}_F = (\mathrm{Id} + \partial F)^{-1} = \mathcal{R}_{\partial F}.$$

Wir müssen noch zeigen, dass auch die Resolvente von beliebigen maximal monotonen Operators einwertig ist.

**Lemma 6.6.** *Sei $A : X \rightrightarrows X$ maximal monoton. Dann ist $\mathcal{R}_A$ auf ganz $X$ definiert und einwertig.*

*Beweis.* Da $A$ maximal monoton ist, ist $\mathrm{Id} + A$ surjektiv, woraus $\mathrm{dom}\, \mathcal{R}_A = X$ folgt. Seien nun $x_1, x_2 \in X$ beliebig und $z_1 \in \mathcal{R}_A(x_1)$ und $z_2 \in \mathcal{R}_A(x_2)$, d. h. es gilt $x_1 \in \{z_1\} + Az_1$ und $x_2 \in \{z_2\} + Az_2$. Aus $x_1 - z_1 \in Az_1$ und $x_2 - z_2 \in Az_2$ folgt mit der Monotonie von $A$, dass gilt

$$\|z_1 - z_2\|_X^2 \le (x_1 - x_2, z_1 - z_2)_X . \tag{6.10}$$

Ist nun $x_1 = x_2$, so muss daher auch $z_1 = z_2$ sein. Also ist $\mathcal{R}_A$ einwertig. □

Aus dem Beweis folgen sogar noch stärkere Eigenschaften.

**Lemma 6.7.** *Sei $A : X \rightrightarrows X$ maximal monoton. Dann gilt:*

(i) *$\mathcal{R}$ ist* nichtexpansiv, *d. h. Lipschitzstetig mit Konstante* $1$;

(ii) *$\mathcal{R}$ ist* stark nichtexpansiv, *d. h.*

$$\|\mathcal{R}_A x_1 - \mathcal{R}_A x_2\|_X^2 \le (\mathcal{R}_A x_1 - \mathcal{R}_A x_2, x_1 - x_2)_X \qquad \text{für alle } x_1, x_2 \in X;$$

(iii) $\|\mathcal{R}_A x_1 - \mathcal{R}_A x_2\|_X^2 + \|(\mathrm{Id} - \mathcal{R}_A)x_1 - (\mathrm{Id} - \mathcal{R}_A)x_2\|_X^2 \le \|x_1 - x_2\|_X^2$ *für alle* $x_1, x_2 \in X$.

*Beweis.* Die Aussage (ii) haben wir bereits mit (6.10) gezeigt, und Aussage (i) folgt daraus sofort mit Hilfe der Cauchy–Schwarz-Ungleichung. Ebenso folgt (iii) aus (ii) zusammen mit dem Satz von Pythagoras, denn

$$\begin{aligned}\|(\operatorname{Id}-\mathcal{R}_A)x_1-(\operatorname{Id}-\mathcal{R}_A)x_2\|_X^2 &= \|x_1-x_2\|_X^2+\|\mathcal{R}_Ax_1-\mathcal{R}_Ax_2\|_X^2\\ &\quad -2\,(x_1-x_2,\mathcal{R}_Ax_1-\mathcal{R}_Ax_2)_X\\ &\le \|x_1-x_2\|_X^2-\|\mathcal{R}_Ax_1-\mathcal{R}_Ax_2\|_X^2.\end{aligned}$$

□

Folgerung 6.8. *Für $F: X \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig ist $\operatorname{prox}_F : X \to X$ Lipschitz-stetig mit Konstante $L = 1$.*

Tatsächlich lassen sich Minimierer konvexer Funktionen auch als Proximalpunkte charakterisieren, indem man den folgende nützlichen Zusammenhang verwendet.

Lemma 6.9. *Sei $F : X \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig, und seien $x, x^* \in X$. Dann ist für beliebige $\gamma > 0$*

$$x^* \in \partial F(x) \quad\Leftrightarrow\quad x = \operatorname{prox}_{\gamma F}(x+\gamma x^*).$$

*Beweis.* Aus den Rechenregeln für mengenwertige Abbildungen folgt dann

$$\begin{aligned}x^* \in \partial F(x) &\Leftrightarrow x+\gamma x^* \in (\operatorname{Id}+\gamma\partial F)(x)\\ &\Leftrightarrow x \in (\operatorname{Id}+\gamma\partial F)^{-1}(x+\gamma x^*)\\ &\Leftrightarrow x = \operatorname{prox}_{\gamma F}(x+\gamma x^*).\end{aligned}$$

Im letzten Schritt haben wir dabei $\gamma\partial F = \partial(\gamma F)$ nach Lemma 4.7 (ii) und damit $\operatorname{prox}_{\gamma F} = \mathcal{R}_{\gamma\partial F}$ verwendet. □

Wendet man diese Umformulierung auf das Fermat-Prinzip $0 \in \partial F(\bar{x})$ an, erhält man sofort das folgende Resultat.

Folgerung 6.10. *Sei $F : X \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig und $\gamma > 0$ beliebig. Dann ist $\bar{x} \in \operatorname{dom} F$ ein Minimierer von $F$ genau dann, wenn gilt*

$$\bar{x} = \operatorname{prox}_{\gamma F}(\bar{x}).$$

Damit liefert uns die Proximalpunkt-Abbildung eine Möglichkeit, Minimierer konvexer Funktionen statt durch (explizite) Mengeninklusion durch eine (implizite) Gleichung zu charakterisieren (und durch eine Fixpunktiteration zu berechnen).

Weiter erhält man daraus eine Verallgemeinerung der orthogonalen Zerlegung von Vektorräumen.

**Satz 6.11 (Moreau-Zerlegung).** *Sei $F : X \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig. Dann gilt für alle $x \in X$*

$$x = \operatorname{prox}_F(x) + \operatorname{prox}_{F^*}(x).$$

*Beweis.* Setze $w = \operatorname{prox}_F(x)$. Aus Lemma 6.9 und Lemma 5.5 folgt dann

$$\begin{aligned} w = \operatorname{prox}_F(x) = \operatorname{prox}_F(w + (x-w)) &\Leftrightarrow x - w \in \partial F(w) \\ &\Leftrightarrow w \in \partial F^*(x-w) \\ &\Leftrightarrow x - w = \operatorname{prox}_{F^*}((x-w)+w) = \operatorname{prox}_{F^*}(x). \qquad \square \end{aligned}$$

Folgende Rechenregeln werden nützlich sein.

**Lemma 6.12.** *Sei $F : X \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig.*

*(i) Für $\lambda \neq 0$ und $z \in X$ gilt mit $H(x) = F(\lambda x + z)$*

$$\operatorname{prox}_H(x) = \lambda^{-1}(\operatorname{prox}_{\lambda^2 F}(\lambda x + z) - z).$$

*(ii) Für $\gamma > 0$ gilt*

$$\operatorname{prox}_{\gamma F^*}(x) = x - \gamma \operatorname{prox}_{\gamma^{-1}F}(\gamma^{-1}x).$$

*(iii) Für $G : Y \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig gilt mit $H : X \times Y \to \overline{\mathbb{R}}$, $H(x,y) = F(x) + G(y)$,*

$$\operatorname{prox}_{\gamma H}(x,y) = \begin{pmatrix} \operatorname{prox}_{\gamma F}(x) \\ \operatorname{prox}_{\gamma G}(y) \end{pmatrix}.$$

*Beweis.* Zu (i): Nach Definition ist

$$\operatorname{prox}_H(x) = \operatorname*{arg\,min}_{w \in X} \frac{1}{2}\|w - x\|_X^2 + F(\lambda w + z) =: \bar w.$$

Mit der Substitution $v = \lambda w + z$ folgt wegen $\min_{w\in X} J(w) = \min_{v \in X} J(v)$ für alle $J$

$$\begin{aligned} \bar v &= \operatorname*{arg\,min}_{v\in X} \frac{1}{2}\|\lambda^{-1}(v-z) - x\|_X^2 + F(v) \\ &= \operatorname*{arg\,min}_{v\in X} \frac{1}{2\lambda^2}\|v - (\lambda x + z)\|_X^2 + F(v) \\ &= \operatorname*{arg\,min}_{v\in X} \frac{1}{2}\|v - (\lambda x + z)\|_X^2 + \lambda^2 F(v) \\ &= \operatorname{prox}_{\lambda^2 F}(\lambda x + z). \end{aligned}$$

Also ist $\bar w = \lambda^{-1}(\bar v - z)$ der gesuchte Minimierer.

Zu (ii): Aus der Moreau-Zerlegung, Lemma 5.4 (i), und (i) für $\lambda = \gamma^{-1}$ und $z = 0$, folgt

$$\begin{aligned}\operatorname{prox}_{\gamma F}(x) &= x - \operatorname{prox}_{(\gamma F)^*}(x)\\ &= x - \operatorname{prox}_{\gamma F^* \circ (\gamma^{-1}\operatorname{Id})}(x)\\ &= x - \gamma \operatorname{prox}_{\gamma(\gamma^{-2}F^*)}(\gamma^{-1}x).\end{aligned}$$

Anwenden auf $F^*$ unter Verwendung von $F^{**} = F$ liefert nun die Aussage.

Zu (iii): Nach Definition der Norm auf $X \times Y$ gilt

$$\begin{aligned}\operatorname{prox}_{\gamma H}(x, y) &= \operatorname*{arg\,min}_{(u,v)\in X\times Y} \frac{1}{2}\|(u, v) - (x, y)\|^2_{X\times Y} + \gamma H(u, v)\\ &= \operatorname*{arg\,min}_{u\in X, v\in Y} \left(\frac{1}{2}\|u - x\|^2_X + \gamma F(u)\right) + \left(\frac{1}{2}\|v - y\|^2_Y + \gamma G(v)\right).\end{aligned}$$

Da keine gemischten Terme in $u$ und $v$ auftreten, können die Klammern separat minimiert werden. Also ist $\operatorname{prox}_{\gamma H}(x, y) = (\bar u, \bar v)$ mit

$$\begin{aligned}\bar u &= \operatorname*{arg\,min}_{u\in X} \frac{1}{2}\|u - x\|^2_X + \gamma F(u) = \operatorname{prox}_{\gamma F(x)},\\ \bar v &= \operatorname*{arg\,min}_{v\in Y} \frac{1}{2}\|v - y\|^2_Y + \gamma G(v) = \operatorname{prox}_{\gamma G(x)}.\end{aligned}$$ □

Das Ausrechnen von Proximalpunkten ist in der Regel schwierig, denn die Auswertung von $\operatorname{prox}_F$ enthält ja bereits die Minimierung von $F$. In manchen Fällen kann man aber eine geschlossene Formel angeben.

Beispiel 6.13. Wir betrachten zuerst skalare $f : \mathbb{R} \to \overline{\mathbb{R}}$.

(i) $f(t) = \frac{1}{2}|t|^2$. Da $f$ differenzierbar ist, können wir die Ableitung von $\frac{1}{2}(s - t)^2 + \frac{\gamma}{2}s^2$ gleich Null setzen und nach $s$ auflösen und erhalten $\operatorname{prox}_{\gamma f}(t) = (1 + \gamma)^{-1}t$.

(ii) $f(t) = |t|$. Nach (4.5) gilt $\partial f(t) = \operatorname{sign}(t)$; also ist $s := \operatorname{prox}_{\gamma f}(t) = (\operatorname{Id} + \gamma \operatorname{sign})^{-1}(t)$ genau dann, wenn $t \in \{s\} + \gamma \operatorname{sign}(s)$ ist. Angenommen, für gegebenes $t$ gelte diese Inklusion für ein $\bar s$. Wir machen nun eine Fallunterscheidung.

a) $\bar s > 0$. Dies impliziert $t = \bar s + \gamma$, d. h. $\bar s = t - \gamma$ und damit $t > \gamma$.

b) $\bar s < 0$. Dies impliziert $t = \bar s - \gamma$, d. h. $\bar s = t + \gamma$ und damit $t < -\gamma$.

c) $\bar s = 0$. Dies impliziert $t \in \gamma[-1, 1] = [-\gamma, \gamma]$.

Da die so erhaltenen Bedingungen an $t$ disjunkt und vollständig sind, erhalten wir

$$\operatorname{prox}_{\gamma f}(t) = \begin{cases} t - \gamma & \text{falls } t > \gamma, \\ 0 & \text{falls } t \in [-\gamma, \gamma], \\ t + \gamma & \text{falls } t < -\gamma. \end{cases}$$

Diese Abbildung wird auch als *soft-shrinkage*-Operator bezeichnet.

(iii) $f(t) = \delta_{[-1,1]}(t)$. Nach Beispiel 5.2 (iii) gilt $f^*(t) = |t|$, und daraus erhalten wir mit Lemma 6.12 (ii)

$$\begin{aligned} \operatorname{prox}_{\gamma f}(t) &= t - \gamma \operatorname{prox}_{\gamma^{-1} f^*}(\gamma^{-1} t) \\ &= \begin{cases} t - \gamma(\gamma^{-1} t - \gamma^{-1}) & \text{falls } \gamma^{-1} t > \gamma^{-1}, \\ t - 0 & \text{falls } \gamma^{-1} t \in [-\gamma^{-1}, \gamma^{-1}], \\ t - \gamma(\gamma^{-1} t + \gamma^{-1}) & \text{falls } \gamma^{-1} t < -\gamma^{-1} \end{cases} \\ &= \begin{cases} 1 & \text{falls } t > 1, \\ t & \text{falls } t \in [-1, 1], \\ -1 & \text{falls } t < -1. \end{cases} \end{aligned}$$

Die Proximalpunktabbildung ist also – für alle $\gamma$ – die Projektion von $t$ auf $[-1, 1]$.

Beispiel 6.14. Beispiel 6.13 kann auf $X = \mathbb{R}^N$ (versehen mit dem Euklidischen Skalarprodukt) verallgemeinert werden, indem man $N$ mal Lemma 6.12 (iii) anwendet. Man erhält dann komponentenweise

(i) für $F(x) = \frac{1}{2}\|x\|_2^2 = \sum_{i=1}^N \frac{1}{2} x_i^2$:

$$[\operatorname{prox}_{\gamma F}(x)]_i = \left(\frac{1}{1+\gamma}\right) x_i, \quad 1 \le i \le N;$$

(ii) für $F(x) = \|x\|_1 = \sum_{i=1}^N |x_i|$:

$$[\operatorname{prox}_{\gamma F}(x)]_i = (|x_i| - \gamma)^+ \operatorname{sign}(x_i), \quad 1 \le i \le N;$$

(iii) für $F(x) = \delta_{B_\infty}(x) = \sum_{i=1}^N \delta_{[-1,1]}(x_i)$:

$$[\operatorname{prox}_{\gamma F}(x)]_i = x_i - (x_i - 1)^+ - (x_i + 1)^- = \frac{x_i}{\max\{1, |x_i|\}}, \qquad 1 \le i \le N;$$

mit der Schreibweise $(t)^+ := \max\{t, 0\}$ und $(t)^- := \min\{t, 0\}$.

Viele weitere Beispiele findet man in [Parikh & Boyd 2014, § 6.5].

Da für konvexe Integralfunktionale nach Lemma 4.6 das Subdifferential punktweise ausgewertet werden kann, gilt dasselbe auch für die Definition (6.2) der Proximalpunktabbildung.

Folgerung 6.15. *Sei $f : \mathbb{R} \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig und $F : L^2(\Omega) \to \overline{\mathbb{R}}$ wie in Lemma 3.4. Dann ist für alle $\gamma > 0$ und $u \in L^2(\Omega)$*

$$[\operatorname{prox}_{\gamma F}(u)](x) = \operatorname{prox}_{\gamma f}(u(x)) \qquad \text{für fast alle } x \in \Omega.$$

Beispiel 6.16. Durch Fallunterscheidung analog zu Beispiel 6.13 zeigt man für allgemeine Hilberträume $X$:

(i) Ist $F = \frac{1}{2}\|\cdot\|_X^2 = \frac{1}{2}(\cdot,\cdot)_X$, so ist

$$\operatorname{prox}_{\gamma F}(x) = \left(\frac{1}{1+\gamma}\right) x.$$

(ii) Ist $F = \|\cdot\|_X$, so zeigt man analog mit Fallunterscheidung wie in Satz 4.5

$$\operatorname{prox}_{\gamma F}(x) = \left(1 - \frac{\gamma}{\|x\|_X}\right)^+ x.$$

(iii) Ist $F = \delta_C$ für $C \subset X$ nichtleer, konvex und abgeschlossen, so ist nach Definition

$$\operatorname{prox}_{\gamma F}(x) = \operatorname{proj}_C(x) := \arg\min_{z \in C} \|z - x\|_X$$

die *Projektion* von $x$ auf $C$; die Proximalpunktabbildung verallgemeinert also die metrische Projektion auf (konvexe) Mengen. Projektionsformeln bzw. -verfahren für verschiedene Klassen von Mengen sind in [Cegielski 2012, Kapitel 4.1] aufgeführt.

## 6.3 MOREAU–YOSIDA-REGULARISIERUNG

Bevor wir zu Algorithmen für die Minimierung konvexer Funktionen kommen, betrachten wir noch eine weitere Möglichkeit, Optimalitätsbedingungen mit Hilfe von Proximalpunktabbildungen umzuformulieren. Dies ist zwar keine äquivalente Umformung mehr, bildet aber die Brücke zu den Newton-artigen Verfahren im letzten Kapitel.

Sei $A : X \rightrightarrows X$ ein maximal monotoner Operator und $\gamma > 0$. Dann ist die *Yosida-Approximation* von $A$ definiert durch

$$A_\gamma := \frac{1}{\gamma}\left(\operatorname{Id} - \mathcal{R}_{\gamma A}\right).$$

Insbesondere gilt für das Subdifferential einer eigentlichen, konvexen und unterhalbstetigen Funktion $F : X \to \overline{\mathbb{R}}$

$$(\partial F)_\gamma := \frac{1}{\gamma}\left(\mathrm{Id} - \mathrm{prox}_{\gamma F}\right).$$

Nach Folgerung 6.8 ist die Yosida-Approximation also stets Lipschitz-stetig mit Konstante $L = \gamma^{-1}$.

Eine andere Sichtweise erlaubt die folgende Definition. Für eine eigentliche, konvexe und unterhalbstetige Funktion $F : X \to \overline{\mathbb{R}}$ und $\gamma > 0$ ist die *Moreau-Hülle*[1] von $F$ definiert durch

$$F_\gamma : X \to \mathbb{R}, \qquad x \mapsto \inf_{z\in X} \frac{1}{2\gamma}\|z - x\|_X^2 + F(z).$$

Vergleich mit der Definition (6.2) der Proximalpunktabbildung liefert

$$F_\gamma(x) = \frac{1}{2\gamma}\|\mathrm{prox}_{\gamma F}(x) - x\|_X^2 + F(\mathrm{prox}_{\gamma F}(x)). \tag{6.11}$$

(Beachte, dass sich der Minimierer nicht ändert, wenn das Funktional mit $\gamma > 0$ multipliziert wird.) Also ist $F_\gamma$ in der Tat wohldefiniert und einwertig. Zudem ist mit $F$ auch $F_\gamma$ konvex.

**Lemma 6.17.** *Sei $F : X \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig. Dann ist für alle $\gamma > 0$ auch $F_\gamma$ konvex.*

*Beweis.* Wir zeigen zunächst, dass für $G : X \to \overline{\mathbb{R}}$ konvex die Abbildung

$$H : X \times X \to \overline{\mathbb{R}}, \qquad (x, z) \mapsto F(z) + G(z - x)$$

ebenfalls konvex ist. Seien dafür $(x_1, z_1), (x_2, z_2) \in X \times X$ und $\lambda \in [0, 1]$. Aus der Konvexität von $F$ und $G$ folgt dann

$$\begin{aligned}
H(\lambda(x_1, z_1) + (1-\lambda)(x_2, z_2)) &= F(\lambda z_1 + (1-\lambda)z_2) + G(\lambda(z_1 - x_1) + (1-\lambda)(z_2 - x_2)) \\
&\le \lambda (F(z_1) + G(z_1 - x_1)) + (1-\lambda)(F(z_2) + G(z_2 - x_2)) \\
&= \lambda H(x_1, z_1) + (1-\lambda) H(x_2, z_2).
\end{aligned}$$

Seien nun $x_1, x_2 \in X$ und $\lambda \in [0, 1]$. Wegen $F_\gamma(x) = \inf_{z\in X} H(x, z)$ für $G(y) := \frac{1}{2\gamma}\|y\|_X^2$ existieren dann zwei Minimalfolgen $\{z_n^1\}_{n\in\mathbb{N}}, \{z_n^2\}_{n\in\mathbb{N}} \subset X$ mit

$$H(x_1, z_n^1) \to F_\gamma(x_1), \qquad H(x_2, z_n^2) \to F_\gamma(x_2).$$

Aus der Definition des Infimums zusammen mit der Konvexität von $H$ folgt daher für alle $n \in \mathbb{N}$

$$\begin{aligned}
F_\gamma(\lambda x_1 + (1-\lambda)x_2) &\le H(\lambda(x_1, z_n^1) + (1-\lambda)(x_2, z_n^2)) \\
&\le \lambda H(x_1, z_n^1) + (1-\lambda) H(x_2, z_n^2),
\end{aligned}$$

und Grenzübergang $n \to \infty$ auf beiden Seiten ergibt die behauptete Konvexität. □

[1] nicht zu verwechseln mit der *konvexen* Hülle $F^\Gamma$!

Der nächste Satz stellt den Zusammenhang zwischen den beiden Definitionen her und rechtfertigt damit den Begriff *Moreau–Yosida-Regularisierung*.

Satz 6.18. *Sei $F : X \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig. Dann ist $F_\gamma$ für alle $\gamma > 0$ Fréchet-differenzierbar, und es gilt*

$$\nabla(F_\gamma) = (\partial F)_\gamma.$$

*Beweis.* Seien $x, y \in X$ beliebig und $x^* = \operatorname{prox}_{\gamma F}(x)$ sowie $y^* = \operatorname{prox}_{\gamma F}(y)$. Wir zeigen zunächst, dass gilt

$$\frac{1}{\gamma}\left(y^* - x^*, x - x^*\right)_X \leq F(y^*) - F(x^*). \tag{6.12}$$

(Beachte, dass für $F$ eigentlich aus der Definition von Proximalpunkten als Minimierer notwendigerweise $x^*, y^* \in \operatorname{dom} F$ folgt.) Dafür betrachten wir für $t \in (0,1)$ den Punkt $x_t^* := ty^* + (1-t)x^*$. Aus der Definition des Proximalpunktes $x^*$ als Minimierer und der Konvexität von $F$ folgt durch quadratisches Ergänzen

$$\begin{aligned} F(x^*) &\leq F(x_t^*) + \frac{1}{2\gamma}\|x_t^* - x\|_X^2 - \frac{1}{2\gamma}\|x^* - x\|_X^2 \\ &\leq tF(y^*) + (1-t)F(x^*) - \frac{t}{\gamma}\left(x - x^*, y^* - x^*\right)_X + \frac{t^2}{2\gamma}\|x^* - y^*\|_X^2. \end{aligned}$$

Umstellen, Division durch $t > 0$ und Grenzübergang $t \to 0$ ergibt dann die gewünschte Ungleichung (6.12). Zusammen mit (6.11) folgt daraus

$$\begin{aligned} F_\gamma(y) - F_\gamma(x) &= F(y^*) - F(x^*) + \frac{1}{2\gamma}\left(\|y - y^*\|_X^2 - \|x - x^*\|_X^2\right) \\ &\geq \frac{1}{2\gamma}\left(2\left(y^* - x^*, x - x^*\right)_X + \|y - y^*\|_X^2 - \|x - x^*\|_X^2\right) \\ &= \frac{1}{2\gamma}\left(2\left(y - x, x - x^*\right)_X + \|y - y^* - x + x^*\|_X^2\right) \\ &\geq \frac{1}{\gamma}\left(y - x, x - x^*\right)_X. \end{aligned}$$

Analog zeigt man durch Vertauschen der Rollen von $x^*$ und $y^*$ in (6.12), dass gilt

$$F_\gamma(y) - F_\gamma(x) \leq \frac{1}{\gamma}\left(y - x, y - y^*\right)_X.$$

Kombination dieser beiden Ungleichungen liefert dann

$$\begin{aligned}
0 &\leq F_\gamma(y) - F_\gamma(x) - \frac{1}{\gamma}\,(y - x, x - x^*)_X \\
&\leq \frac{1}{\gamma}\,(y - x, (y - y^*) - (x - x^*))_X \\
&\leq \frac{1}{\gamma}\left(\|y - x\|_X^2 - \|y^* - x^*\|_X^2\right) \\
&\leq \frac{1}{\gamma}\|y - x\|_X^2,
\end{aligned}$$

wobei wir im vorletzten Schritt die starke Nichtexpansivität der Proximalpunktabbildung (Lemma 6.7 (ii)) verwendet haben.

Setzen wir nun $y = x + h$ für $h \in X$ beliebig, so folgt daraus

$$0 \leq \frac{F_\gamma(x+h) - F_\gamma(x) - \left(\gamma^{-1}(x - x^*), h\right)_X}{\|h\|_X} \leq \frac{1}{\gamma}\|h\|_X \to 0 \qquad \text{für } h \to 0,$$

d. h. $F_\gamma$ ist Fréchet-differenzierbar, und der Gradient ist genau $\frac{1}{\gamma}(x - x^*) = (\partial F)_\gamma$. □

Da $F_\gamma$ nach Lemma 6.17 konvex ist, folgt daraus zusammen mit Satz 4.4 die eingängige Gleichung $\partial(F_\gamma) = (\partial F)_\gamma$.

Beispiel 6.19. Wir betrachten wieder $X = \mathbb{R}^N$.

(i) $F(x) = \|x\|_1$. Nach Beispiel 6.14 (ii) ist die Proximalpunktabbildung der komponentenweise soft-shrinkage-Operator, und damit ist komponentenweise

$$\left[(\partial\|\cdot\|_1)_\gamma(x)\right]_i = \begin{cases} \frac{1}{\gamma}(x_i - (x_i - \gamma)) = 1 & x_i > \gamma, \\ \frac{1}{\gamma}x_i & x_i \in [-\gamma, \gamma], \\ \frac{1}{\gamma}(x_i - (x_i + \gamma)) = -1 & x_i < \gamma. \end{cases}$$

Im Vergleich zum entsprechenden Subdifferential (4.5) ist also der mengenwertige Fall in $x_i = 0$ durch eine lineare Funktion auf einem kleinen Intervall ersetzt worden.

Analog erhält man durch Einsetzen in (6.11)

$$F_\gamma(x) = \sum_{i=1}^N f_\gamma(x_i) \quad \text{für} \quad f_\gamma(t) := \begin{cases} \frac{1}{2\gamma}|t - (t - \gamma)|^2 + |t - \gamma| = t - \frac{\gamma}{2} & t > \gamma, \\ \frac{1}{2\gamma}|t|^2 & t \in [-\gamma, \gamma], \\ \frac{1}{2\gamma}|t - (t + \gamma)|^2 + |t + \gamma| = -t + \frac{\gamma}{2} & t < -\gamma. \end{cases}$$

Der Betrag wird also für kleine Argumente durch eine quadratische Funktion ersetzt (wodurch die Nichtdifferenzierbarkeit im Ursprung verschwindet). Diese Modifikation ist seit langem als *Huber-Norm* bekannt.

(ii) $F(x) = \delta_{B_\infty}(x)$. Nach Beispiel 6.14 (iii) ist die Proximalpunktabbildung gegeben durch die komponentenweise Projektion auf $[-1,1]$, und damit ist

$$\left[(\partial\delta_{B_\infty})_\gamma(x)\right]_i = \frac{1}{\gamma}\Big(x_i - \big(x_i - (x_i-1)^+ - (x_i+1)^-\big)\Big) = \frac{1}{\gamma}(x_i-1)^+ + \frac{1}{\gamma}(x_i+1)^-.$$

Analog folgt durch Einsetzen wegen $\operatorname{prox}_{\gamma F}(x) \in B_\infty$ und $((x+1)^+, (x-1)^-)_X = 0$

$$(\delta_{B_\infty})_\gamma(x) = \frac{1}{2\gamma}\|(x-1)^+\|_2^2 + \frac{1}{2\gamma}\|(x+1)^-\|_2^2,$$

was genau der aus der nichtlinearen Optimierung bekannten Penalty-Funktion für die Ungleichungsnebenbedingungen $x-1 \leq 0$ und $x+1 \geq 0$ entspricht.

Ein weiterer Zusammenhang besteht zwischen der Moreau-Hülle $F_\gamma$ und der Fenchel-Konjugierten $F^*$.

Satz 6.20. *Sei $F : X \to \overline{\mathbb{R}}$ eigentlich, konvex und unterhalbstetig. Dann gilt für alle $\gamma > 0$ und $x^* \in X$*

$$(F_\gamma)^*(x^*) = F^*(x^*) + \frac{\gamma}{2}\|x^*\|_X^2.$$

*Beweis.* Die Definition der Fenchel-Konjugierten im Hilbertraum und der Moreau-Hülle ergibt direkt

$$\begin{aligned}
(F_\gamma)^*(x^*) &= \sup_{x\in X}\left((x^*,x)_X - \inf_{z\in X}\left(\tfrac{1}{2\gamma}\|x-z\|_X^2 + F(z)\right)\right)\\
&= \sup_{x\in X}\left((x^*,x)_X + \sup_{z\in X}\left(-\tfrac{1}{2\gamma}\|x-z\|_X^2 - F(z)\right)\right)\\
&= \sup_{z\in X}\left((x^*,z)_X - F(z) + \sup_{x\in X}\left((x^*,x-z)_X - \tfrac{1}{2\gamma}\|x-z\|_X^2\right)\right)\\
&= F^*(x^*) + \left(\tfrac{1}{2\gamma}\|\cdot\|_X^2\right)^*(x^*),
\end{aligned}$$

da unabhängig von der Wahl von $z \in X$ im inneren Supremum $x - z$ stets den ganzen Raum abdeckt. Aus Beispiel 5.2 (i) und Lemma 5.4 (i) folgt nun die Aussage. □

Wir skizzieren kurz die Relevanz für die Optimierung. Ist $F : X \to \overline{\mathbb{R}}$ konvex, so erfüllt jeder Minimierer das Fermat-Prinzip $0 \in \partial F(\bar{x})$, welches wir äquivalent schreiben können als $\bar{x} \in \partial F^*(0)$. Ersetzen wir $\partial F^*$ durch die Yosida-Approximation $(\partial F^*)_\gamma$, so erhalten wir eine regularisierte Bedingung

$$x_\gamma = (\partial F^*)_\gamma(0) = -\frac{1}{\gamma}\operatorname{prox}_{\gamma F^*}(0).$$

Dies ist nun eine *explizite* und sogar Lipschitz-stetige Relation. Zwar ist $x_\gamma$ selber kein Minimierer von $F$, wegen der Konvexität von $F_\gamma$ ist $x_\gamma \in (\partial F^*)_\gamma(0) = \partial(F_\gamma^*)(0)$ aber äquivalent mit

$$0 \in \partial(F_\gamma^*)^*(x_\gamma) = \partial\left(F^{**} + \tfrac{\gamma}{2}\|\cdot\|_X^2\right)(x_\gamma) = \partial\left(F + \tfrac{\gamma}{2}\|\cdot\|_X^2\right)(x_\gamma),$$

d. h. $x_\gamma$ ist der (wegen der strikten Konvexität der quadrierten Norm) einzige Minimierer von $F + \frac{\gamma}{2}\|\cdot\|_X^2$. Durch die Regularisierung von $\partial F^*$ wurde das zu minimierende Funktional also nicht geglättet, sondern lediglich stärker konvexifiziert. Aus der Äquivalenz folgt auch (ähnlich wie im Beweis von Satz 2.1) $x_\gamma \rightharpoonup \bar{x}$ für $\gamma \to 0$. In der Praxis scheitert dieses Vorgehen an der Berechnung von $F^*$, weshalb man es mit dem im nächsten Kapitel vorgestellten Splitting-Ansatz kombiniert.

# 7 PROXIMALPUNKT- UND SPLITTING-VERFAHREN

Wir beschäftigen uns jetzt mit Verfahren zur numerischen Berechnung von Minimierern von Funktionalen $J : X \to \overline{\mathbb{R}}$ der Form

$$J(x) := F(x) + G(x)$$

für $F, G$ konvex aber nicht differenzierbar. Die Schwierigkeit dabei ist, dass im Gegensatz zu differenzierbaren Funktionen das Äquivalent zum Gradientenabstieg, die Iteration

$$x^{k+1} \in x^k - \tau_k \partial J(x^k),$$

nicht praktikabel ist, da ein beliebiges Element aus dem Subdifferential keine Abstiegsrichtung sein muss; dies kann man nur für den Subgradienten minimaler Norm garantieren – und für $J$ erhält man diesen auch nicht aus den entsprechenden Subgradienten von $F$ und $G$. Wir ändern daher unseren Blickwinkel und suchen eine Nullstelle der mengenwertigen Abbildung $x \mapsto \partial J(x) \subset X^* \cong X$.

## 7.1 PROXIMALPUNKT-VERFAHREN

Wir haben in Folgerung 6.10 gesehen, dass eine Nullstelle $\bar{x}$ von $\partial F : X \rightrightarrows X$ charakterisiert werden kann als Fixpunkt von $\operatorname{prox}_{\gamma F}$ für beliebiges $\gamma > 0$. Dies legt nahe, diese durch eine Fixpunktiteration zu berechnen: Wähle $x^0 \in X$ und setze für eine geeignete Folge $\{\gamma_k\}_{k\in\mathbb{N}}$

$$x^{k+1} = \operatorname{prox}_{\gamma_k F}(x^k). \tag{7.1}$$

Wie bei der klassischen Fixpunktiteration müssen wir nun nachweisen, dass die Fixpunktabbildung in einem passenden Sinne kontrahierend wirkt. Es zeigt sich, dass die starke Nichtexpansivität ausreicht, was nach Folgerung 6.8 für Resolventen maximal monotoner Operatoren stets der Fall ist. Für den späteren Gebrauch betrachten wir gleich die allgemeine Version für beliebige maximal monotone Operatoren.

**Satz 7.1.** *Sei $A : X \rightrightarrows X$ maximal monoton und habe eine Nullstelle $x^* \in X$, und sei $\{\gamma_k\}_{k\in\mathbb{N}} \subset (0,\infty)$ mit $\sum_{k=0}^\infty \gamma_k^2 = \infty$. Erfüllt $\{x^k\}_{k\in\mathbb{N}} \subset X$ die Rekursion*

$$x^{k+1} = \mathcal{R}_{\gamma_k A} x^k,$$

*dann konvergiert $x^k \rightharpoonup \bar{x}$ mit $0 \in A\bar{x}$.*

*Beweis.* Aus der Rekursion $x^{k+1} = \mathcal{R}_{\gamma_k A}x^k = (\mathrm{Id} + \gamma_k A)^{-1}x^k$ folgt

$$w^k := \gamma_k^{-1}(x^k - x^{k+1}) \in Ax^{k+1}$$

und damit auch $x^{k+1} - x^{k+2} = \gamma_{k+1}w^{k+1}$. (Der Vektor $w^k$ spielt die Rolle des Residuums in der verallgemeinerten Gleichung $0 \in Ax$.) Da $A$ monoton ist, gilt für $\gamma_{k+1} > 0$

$$\begin{aligned} 0 &\leq \gamma_{k+1}^{-1}\left(w^k - w^{k+1}, x^{k+1} - x^{k+2}\right)_X \\ &= \left(w^k - w^{k+1}, w^{k+1}\right)_X \\ &= \left(w^k, w^{k+1}\right)_X - \|w^{k+1}\|_X^2 \\ &\leq \|w^{k+1}\|_X \left(\|w^k\|_X - \|w^{k+1}\|_X\right). \end{aligned}$$

Also ist die nicht-negative Folge $\{\|w^k\|_X\}_{k\in\mathbb{N}} \subset \mathbb{R}$ monoton fallend und daher konvergent (zumindest solange $w^{k+1} \neq 0$ gilt, aber andernfalls ist wegen $w^{k+1} \in Ax^{k+2}$ bereits $x^{k+2}$ die gesuchte Nullstelle.)

Sei nun $x^* \in X$ eine Nullstelle von $A$, d. h. $0 \in Ax^*$, die nach Voraussetzung existiert. Dann folgt wie im Beweis von Folgerung 6.10 $x^* = \mathcal{R}_{\gamma A}x^*$ für alle $\gamma > 0$. Aus Lemma 6.7 (iii) folgt dann mit $(\mathrm{Id} - \mathcal{R}_{\gamma_k A})x^k = x^k - x^{k+1} = \gamma_k w^k$

$$\begin{aligned} \|x^{k+1} - x^*\|_X^2 &= \|\mathcal{R}_{\gamma_k A}x^k - \mathcal{R}_{\gamma_k A}x^*\|_X^2 \\ &\leq \|x^k - x^*\|_X^2 - \|(\mathrm{Id} - \mathcal{R}_{\gamma_k A})x^k - (\mathrm{Id} - \mathcal{R}_{\gamma_k A})x^*\|_X^2 \\ &= \|x^k - x^*\|_X^2 - \gamma_k^2\|w^k\|_X^2. \end{aligned} \tag{7.2}$$

Also ist $\{\|x^k - x^*\|_X\}_{k\in\mathbb{N}}$ für jede Nullstelle $x^*$ monoton fallend (man nennt solche Folgen *Féjer-monoton*) und damit beschränkt. Also ist auch $\{x^k\}_{k\in\mathbb{N}} \subset X$ beschränkt und hat daher eine schwach konvergente Teilfolge $x^{k_l} \rightharpoonup \bar{x}$.

Rekursives Anwenden von (7.2) ergibt außerdem

$$0 \leq \|x^{k+1} - x^*\|_X^2 \leq \|x^0 - x^*\|_X^2 - \sum_{j=0}^{k} \gamma_j^2\|w^j\|_X^2. \tag{7.3}$$

Die (monoton wachsende) Partialsummenfolge auf der rechten Seite ist also beschränkt, und deshalb konvergiert $\sum_{k=0}^{\infty} \gamma_k^2\|w^k\|_X^2$. Da die Folge $\{\gamma_k^2\}_{k\in\mathbb{N}}$ nach Annahme nicht summierbar ist, muss $\liminf_{k\to\infty} \|w^k\|_X^2 = 0$ gelten. Zusammen mit der Konvergenz von $\{\|w^k\|_X\}_{k\in\mathbb{N}}$ folgt daraus $w^k \to 0$. Insbesondere gilt $Ax^{k_l+1} \ni w^{k_l} \to 0$ für $x^{k_l+1} \rightharpoonup \bar{x}$, und aus der Abgeschlossenheit von maximal monotonen Operatoren (Lemma 6.2) folgt dann $0 \in A\bar{x}$, d. h. jeder schwache Häufungspunkt von $\{x^k\}_{k\in\mathbb{N}}$ ist eine Nullstelle von $A$.

Wir zeigen schließlich, dass die gesamte Folge $\{x^k\}_{k\in\mathbb{N}}$ schwach konvergiert. (Dieser Beweisschritt ist auch unabhängig als *Opial-Lemma* bekannt.) Seien $\bar{x}$ und $\hat{x}$ schwache Häufungspunkte und damit Nullstellen von $A$. Dann sind wegen der Féjer-Monotonie von

$\{x^k\}_{k\in\mathbb{N}}$ sowohl $\{\|x^k - \bar{x}\|_X\}_{k\in\mathbb{N}}$ als auch $\{\|x^k - \hat{x}\|_X\}_{k\in\mathbb{N}}$ monoton fallend und beschränkt und daher konvergent. Also konvergiert auch

$$\left(x^k, \bar{x} - \hat{x}\right)_X = \frac{1}{2}\left(\|x^k - \hat{x}\|_X^2 - \|x^k - \bar{x}\|_X^2 + \|\bar{x}\|_X^2 - \|\hat{x}\|_X^2\right) \to c \in \mathbb{R}.$$

Da $\bar{x}$ schwacher Häufungspunkt ist, existiert eine Teilfolge $\{x^{k_n}\}_{n\in\mathbb{N}}$ mit $x^{k_n} \rightharpoonup \bar{x}$; ebenso existiert eine Teilfolge $\{x^{k_m}\}_{m\in\mathbb{N}}$ mit $x^{k_m} \rightharpoonup \hat{x}$. Also ist

$$(\bar{x}, \bar{x} - \hat{x})_X = \lim_{n\to\infty} \left(x^{k_n}, \bar{x} - \hat{x}\right)_X = c = \lim_{m\to\infty} \left(x^{k_m}, \bar{x} - \hat{x}\right)_X = (\hat{x}, \bar{x} - \hat{x})_X,$$

und damit

$$0 = (\bar{x} - \hat{x}, \bar{x} - \hat{x})_X = \|\bar{x} - \hat{x}\|_X^2,$$

d. h. $\bar{x} = \hat{x}$. Jede Teilfolge hat daher den selben Grenzwert, und deshalb konvergiert die gesamte Folge $\{x^k\}_{k\in\mathbb{N}}$ gegen diesen. □

## 7.2 EXPLIZITES SPLITTING

Für Funktionale der Form $J(x) = F(x) + G(x)$ ist das Proximalpunkt-Verfahren in der Regel nicht praktikabel, da die Auswertung von $\mathrm{prox}_J$ nicht wesentlich einfacher ist als das ursprüngliche Problem – selbst wenn $\mathrm{prox}_F$ und $\mathrm{prox}_G$ eine geschlossene Darstellung haben. Anstatt die Proximalpunkt-Formulierung auf $0 \in \partial J(\bar{x})$ anzuwenden, verwendet man daher zuerst die Summenregel und erhält dadurch ein $\bar{p} \in X$ mit

$$\begin{cases} \bar{p} \in \partial F(\bar{x}), \\ -\bar{p} \in \partial G(\bar{x}). \end{cases} \tag{7.4}$$

Man kann nun eine oder beide der Subdifferentialinklusionen durch eine Proximalpunkt-abbildung ersetzen.

In expliziten Splitting-Verfahren – auch *forward–backward splitting* genannt – wendet man Lemma 6.9 nur auf die zweite Inklusion in (7.4) an und erhält

$$\begin{cases} \bar{p} \in \partial F(\bar{x}), \\ \bar{x} = \mathrm{prox}_{\gamma G}(\bar{x} - \gamma\bar{p}). \end{cases} \tag{7.5}$$

Die zugehörige Fixpunkt-Iteration ist dann

(i) Wähle $p^k \in \partial F(x^k)$ (mit minimaler Norm).

(ii) Setze $x^{k+1} = \mathrm{prox}_{\gamma_k G}(x^k - \gamma_k p^k)$.

Dies ist im Allgemeinen kein praktikables Verfahren, da die Bestimmung eines Subgradienten minimaler Norm aufwändig ist. Eine Ausnahme ist jedoch, wenn $F$ zusätzlich differenzierbar ist: Dann gilt (im Hilbertraum) $\partial F(x) = \{\nabla F(x)\}$. In diesem Fall erhalten wir das *Proximal-Gradientenverfahren*

$$x^{k+1} = \mathrm{prox}_{\gamma_k G}(x^k - \gamma_k \nabla F(x^k)) \tag{7.6}$$

(im Spezialfall $G = \delta_C$ – d. h. $\mathrm{prox}_{\gamma G}(x) = \mathrm{proj}_C(x)$ – auch *projiziertes Gradientenverfahren* genannt).

Um nun analog zum Proximalpunkt-Verfahren Konvergenz zu zeigen, müssen wir die Lipschitz-Stetigkeit des Gradienten voraussetzen (da wir für $F$ ja keine Resolvente verwenden, die automatisch stark nichtexpansiv und damit Lipschitz-stetig ist).

Lemma 7.2. *Sei $F : X \to \mathbb{R}$ Gâteaux-differenzierbar mit Lipschitz-stetigem Gradienten. Dann gilt*

$$F(y) \leq F(x) + (\nabla F(x), y - x)_X + \frac{L}{2}\|x - y\|_X^2 \quad \textit{für alle } x, y \in X.$$

*Beweis.* Aus der Gâteaux-Differenzierbarkeit von $F$ folgt

$$\frac{d}{dt}F(x + t(y - x)) = (\nabla F(x + t(y - x)), y - x)_X \quad \text{für alle } x, y \in X.$$

Integration über $t$ auf $[0, 1]$ liefert damit

$$\int_0^1 (\nabla F(x + t(y - x)), y - x)_X \, dt = F(y) - F(x).$$

Zusammen mit der produktiven Null erhalten wir dann unter Verwendung der Cauchy–Schwarz-Ungleichung und Lipschitz-Stetigkeit

$$\begin{aligned} F(y) &= F(x) + (\nabla F(x), y - x)_X + \int_0^1 (\nabla F(x + t(y - x)) - \nabla F(x), y - x)_X \, dt \\ &\leq F(x) + (\nabla F(x), y - x)_X + \int_0^1 \|\nabla F(x + t(y - x)) - \nabla F(x)\|_X \|x - y\|_X \, dt \\ &\leq F(x) + (\nabla F(x), y - x)_X + \int_0^1 Lt\|x - y\|_X^2 \, dt \\ &= F(x) + (\nabla F(x), y - x)_X + \frac{L}{2}\|x - y\|_X^2. \end{aligned}$$ □

Für genügend kleine Schrittweiten kann man damit die Konvergenz des Proximal-Gradientenverfahrens zeigen.

Satz 7.3. *Seien $F : X \to \mathbb{R}$ und $G : X \to \overline{\mathbb{R}}$ eigentlich, konvex, und unterhalbstetig. Sei zusätzlich $F$ Gâteaux-differenzierbar mit Lipschitz-stetigem Gradienten. Gilt $0 < \gamma_{\min} \leq \gamma_k \leq L^{-1}$, dann konvergiert die durch (7.6) erzeugte Folge $x^k \rightharpoonup \bar{x}$ mit $0 \in \partial J(\bar{x})$.*

*Beweis.* Wir argumentieren ähnlich wie im Beweis von Satz 7.1, wobei wir die Monotonie der Folge der Residuumsnormen $w^k \in Ax^{k+1}$ durch die der Funktionswerte $J(x^k)$ ersetzen. Wir definieren dafür

$$T_\gamma : X \to X, \qquad x \mapsto \gamma^{-1}(x - \operatorname{prox}_{\gamma G}(x - \gamma \nabla F(x))),$$

womit wir die Iteration (7.6) umschreiben können als

$$x^{k+1} = \operatorname{prox}_{\gamma_k G}(x^k - \gamma_k \nabla F(x^k)) = x^k - \gamma_k T_{\gamma_k}(x^k).$$

Nach Lemma 6.9 gilt daher

$$T_{\gamma_k}(x^k) - \nabla F(x^k) \in \partial G(x^k - \gamma_k T_{\gamma_k}(x^k)),$$

d. h. für alle $z \in X$ gilt

$$G(z) - G(x^k - \gamma_k T_{\gamma_k}(x^k)) \geq \left(T_{\gamma_k}(x^k) - \nabla F(x^k), z - x^k - \gamma_k T_{\gamma_k}(x^k)\right)_X. \tag{7.7}$$

Analog folgt aus $\nabla F(x^k) \in \partial F(x^k)$

$$F(z) - F(x^k) \geq \left(\nabla F(x^k), z - x^k\right)_X. \tag{7.8}$$

Aus Lemma 7.2 folgt weiter mit $x = x^k$, $y = x^{k+1} = x^k - \gamma_k T_{\gamma_k}(x^k)$, und $\gamma_k \leq L^{-1}$

$$\begin{aligned} F(x^k - \gamma_k T_{\gamma_k}(x^k)) &\leq F(x^k) - \gamma_k \left(\nabla F(x^k), T_{\gamma_k}(x^k)\right)_X + \frac{\gamma_k^2 L}{2} \|T_{\gamma_k}(x^k)\|_X^2 \\ &\leq F(x^k) - \gamma_k \left(\nabla F(x^k), T_{\gamma_k}(x^k)\right)_X + \frac{\gamma_k}{2} \|T_{\gamma_k}(x^k)\|_X^2. \end{aligned} \tag{7.9}$$

Zusammen erhalten wir aus (7.9), (7.7), und (7.8), dass für alle $z \in X$ gilt

$$\begin{aligned} J(x^{k+1}) &= F(x^k - \gamma_k T_{\gamma_k}(x^k)) + G(x^k - \gamma_k T_{\gamma_k}(x^k)) \\ &\leq F(x^k) - \gamma_k \left(\nabla F(x^k), T_{\gamma_k}(x^k)\right)_X + \frac{\gamma_k}{2} \|T_{\gamma_k}(x^k)\|_X^2 \\ &\quad + G(z) + \left(T_{\gamma_k}(x^k) - \nabla F(x^k), x^k - \gamma_k T_{\gamma_k}(x^k) - z\right)_X \\ &\leq F(z) + \left(\nabla F(x^k), x^k - z\right)_X - \gamma_k \left(\nabla F(x^k), T_{\gamma_k}(x^k)\right)_X + \frac{\gamma_k}{2} \|T_{\gamma_k}(x^k)\|_X^2 \\ &\quad + G(z) + \left(T_{\gamma_k}(x^k) - \nabla F(x^k), x^k - z - \gamma_k T_{\gamma_k}(x^k)\right)_X \\ &= J(z) + \left(T_{\gamma_k}(x^k), x^k - z\right)_X - \frac{\gamma_k}{2} \|T_{\gamma_k}(x^k)\|_X^2. \end{aligned} \tag{7.10}$$

Für $z = x^k$ folgt daraus

$$J(x^{k+1}) \leq J(x^k) - \frac{\gamma_k}{2}\|T_{\gamma_k}(x^k)\|_X^2, \tag{7.11}$$

d. h. $\{J(x^k)\}_{k\in\mathbb{N}}$ ist monoton fallend. (Das Proximal-Gradientenverfahren ist also ein *Abstiegsverfahren.*) Für $z = x^*$ mit $J(x^*) = \min_{x\in X} J(x)$ folgt weiter durch quadratisches Ergänzen

$$\begin{aligned} 0 \leq J(x^{k+1}) - J(x^*) &\leq \left(T_{\gamma_k}(x^k), x^k - x^*\right)_X - \frac{\gamma_k}{2}\|T_{\gamma_k}(x^k)\|_X^2 \\ &= \frac{1}{2\gamma_k}\left(\|x^k - x^*\|_X^2 - \|x^k - x^* - \gamma_k T_{\gamma_k}(x^k)\|_X^2\right) \\ &= \frac{1}{2\gamma_k}\left(\|x^k - x^*\|_X^2 - \|x^{k+1} - x^*\|_X^2\right). \end{aligned} \tag{7.12}$$

Insbesondere ist $\{\|x^k - x^*\|_X\}_{k\in\mathbb{N}}$ monoton fallend und damit $\{x^k\}_{k\in\mathbb{N}}$ Féjer-monoton und beschränkt. Es existiert also eine schwach konvergente Teilfolge $\{x^{k_l}\}_{l\in\mathbb{N}}$ mit $x^{k_l} \rightharpoonup \bar{x}$.

Summieren über $k = 0, \dots, n-1$ liefert nun

$$\begin{aligned} 0 \leq \sum_{k=1}^{n}(J(x^k) - J(x^*)) &\leq \frac{1}{2\gamma_{\min}}\sum_{k=1}^{n}\left(\|x^{k-1} - x^*\|_X^2 - \|x^k - x^*\|_X^2\right) \\ &= \frac{1}{2\gamma_{\min}}\left(\|x^0 - x^*\|_X^2 - \|x^n - x^*\|_X^2\right) \\ &\leq \frac{1}{2\gamma_{\min}}\|x^0 - x^*\|_X^2. \end{aligned}$$

Da $\{J(x^k)\}_{k\in\mathbb{N}}$ monoton fallend ist, folgt

$$0 \leq J(x^n) - J(x^*) \leq \frac{1}{n}\sum_{k=1}^{n}(J(x^k) - J(x^*)) \leq \frac{1}{2n\gamma_{\min}}\|x^0 - x^*\|_X^2 \tag{7.13}$$

und damit $J(x^n) \to J(x^*)$ für $n \to \infty$. Also gilt auch wegen der schwachen Unterhalbstetigkeit von $F$ und $G$, dass

$$J(\bar{x}) \leq \liminf_{l\to\infty} J(x^{k_l}) = J(x^*).$$

Wie im Beweis von Satz 7.1 folgt nun aus der Féjer-Monotonie von $\{x^k\}_{k\in\mathbb{N}}$, dass $x^k \rightharpoonup \bar{x}$ für die gesamte Folge gilt. □

Aus (7.13) folgt insbesondere $J(x^k) = J(x^*) + O(k^{-1})$. Um $J(x^k) \leq J(x^*) + \varepsilon$ zu erreichen, sind also $O(\varepsilon^{-1})$ Iterationen notwendig. Durch eine geschickte Extrapolation lässt sich dies auf $O(\varepsilon^{-1/2})$ reduzieren, was beweisbar optimal ist; siehe [Nesterov 1983], [Nesterov 2004, Theorem 2.1.7]. (Dafür ist die Folge der Funktionswerte allerdings nicht mehr monoton fallend.) Die Iterationsvorschrift dafür ist

$$\begin{cases} x^{k+1} = \operatorname{prox}_{\gamma_k G}(\bar x^k - \gamma_k \nabla F(\bar x^k)), \\ \bar x^{k+1} = \left(1 - \frac{1-\tau_k}{\tau_{k+1}}\right) x^{k+1} + \frac{1-\tau_k}{\tau_{k+1}} x^k, \end{cases} \tag{7.14}$$

für die (schwer zu motivierende) Folge

$$\tau_0 = 1, \qquad \tau_k = \frac{1+\sqrt{1+4\tau_{k-1}^2}}{2} \quad (\to \infty),$$

siehe [Beck & Teboulle 2009, § 4].

Ein Nachteil des expliziten Splittings ist die Notwendigkeit, für die Wahl einer zulässigen Schrittweite $\gamma_k$ die Lipschitzkonstante $L$ von $\nabla F$ zu kennen. Im Beweis von Satz 7.3 erkennt man, dass dies lediglich dazu dient, die Abschätzung (7.9) zu garantieren. Ist die Lipschitzkonstante nicht bekannt, kann man versuchen, die Gültigkeit von (7.9) durch Liniensuche zu erfüllen: Man startet mit $\gamma^0 > 0$ und verkleinert $\gamma_k$ (etwa durch Halbieren) so lange, bis

$$F(x^k - \gamma_k T_{\gamma_k}(x^k)) \le F(x^k) - \gamma_k \left(\nabla F(x^k), T_{\gamma_k}(x^k)\right)_X + \frac{\gamma_k}{2} \|T_{\gamma_k}(x^k)\|_X^2$$

erfüllt ist (was spätestens für $\gamma_k < L^{-1}$ der Fall ist). Dafür ist in jedem Schritt der Liniensuche jeweils eine Auswertung von $F$ und $\operatorname{prox}_{\gamma_k G}$ erforderlich (wobei Letzteres durch Vertauschen von Proximal- und Gradientenschritt – *backward–forward splitting* – vermieden werden kann).

## 7.3 PRIMAL-DUALES SPLITTING

Dieses Verfahren wurden speziell für Probleme der Form

$$\min_{x\in X} F(x) + G(Ax)$$

für $F, G$ eigentlich, konvex und unterhalbstetig und $A \in L(X, Y)$ entwickelt. Aus Satz 5.6 und Lemma 5.5 erhält man dafür die Fenchel-Extremalitätsbedingungen

$$\begin{cases} -A^*\bar y \in \partial F(\bar x), \\ \quad \bar y \in \partial G(A\bar x), \end{cases} \Leftrightarrow \begin{cases} -A^*\bar y \in \partial F(\bar x), \\ \quad A\bar x \in \partial G^*(\bar y). \end{cases} \tag{7.15}$$

Die zugehörige Proximalpunktformulierung nach Lemma 6.9 ist

$$\begin{cases} \bar x = \operatorname{prox}_{\tau F}(\bar x - \tau A^*\bar y), \\ \bar y = \operatorname{prox}_{\sigma G^*}(\bar y + \sigma A\bar x), \end{cases}$$

für beliebige $\sigma, \tau > 0$. Daraus lässt sich direkt eine Fixpunktiteration gewinnen:

$$\begin{cases} x^{k+1} = \mathrm{prox}_{\tau F}(x^k - \tau A^* y^k), \\ y^{k+1} = \mathrm{prox}_{\sigma G^*}(y^k + \sigma A x^{k+1}) \end{cases} \tag{7.16}$$

(der Einfachheit halber halten wir hier die Schrittweiten konstant). Um zu zeigen, dass diese Iteration konvergiert, versuchen wir sie in die Form eines Proximalpunktverfahrens zu bringen. Dazu formen wir (7.16) so um, dass $(x^k, y^k)$ und $(x^{k+1}, y^{k+1})$ auf jeweils einer Seite stehen. Verwenden von $\mathrm{prox}_{\tau F} = (\mathrm{Id} + \tau\partial F)^{-1}$ ergibt für die erste Gleichung:

$$\begin{aligned} x^{k+1} = \mathrm{prox}_{\tau F}(x^k - \tau A^* y^k) &\Leftrightarrow x^k - \tau A^* y^k \in \{x^{k+1}\} + \tau\partial F(x^{k+1}) \\ &\Leftrightarrow \tau^{-1}x^k - A^* y^k \in \{\tau^{-1}x^{k+1}\} + \partial F(x^{k+1}). \end{aligned}$$

Analog haben wir für die zweite Gleichung

$$y^{k+1} = \mathrm{prox}_{\sigma G^*}(y^k + \sigma A x^{k+1}) \Leftrightarrow \sigma^{-1}y^k \in \{\sigma^{-1}y^{k+1} - Ax^{k+1}\} + \partial G^*(y^{k+1}).$$

Setzen wir $Z = X \times Y$, $z = (x, y)$,

$$M = \begin{pmatrix} \tau^{-1}\,\mathrm{Id} & -A^* \\ 0 & \sigma^{-1}\,\mathrm{Id} \end{pmatrix}, \qquad T = \begin{pmatrix} \partial F & A^* \\ -A & \partial G^* \end{pmatrix},$$

so ist (7.16) äquivalent zu

$$Mz^k \in (M+T)z^{k+1} \qquad \Leftrightarrow \qquad z^{k+1} \in (M+T)^{-1}Mz^k. \tag{7.17}$$

Wäre $M$ invertierbar, so würde $M = (M^{-1})^{-1}$ und damit $(M+T)^{-1}Mz^k = (\mathrm{Id} + M^{-1}T)^{-1}z^k$ gelten; es handelt sich dann also in der Tat um ein Proximalpunkt-Verfahren für den Operator $M^{-1}T$. Ein Vergleich mit (7.15) zeigt außerdem $0 \in T(\bar z)$ für $\bar z = (\bar x, \bar y)$.

Leider ist die Invertierbarkeit von $M$ in dieser Form fraglich; wir modifizieren $M$ daher, um einen selbstadjungierten Operator zu erhalten, für den es dann ausreicht, positive Definitheit nachzuweisen. Wir betrachten also stattdessen

$$M = \begin{pmatrix} \tau^{-1}\,\mathrm{Id} & -A^* \\ -A & \sigma^{-1}\,\mathrm{Id} \end{pmatrix},$$

für das umgekehrt

$$\sigma^{-1}y^k - Ax^k \in \{\sigma^{-1}y^{k+1} - 2Ax^{k+1}\} + \partial G^*(y^{k+1}) \Leftrightarrow y^{k+1} = \mathrm{prox}_{\sigma G^*}(y^k + \sigma A(2x^{k+1} - x^k))$$

gilt und damit (7.17) geschrieben werden kann als (immer noch explizites) *primal-duales Extragradienten-Verfahren*[1]

$$\begin{cases} x^{k+1} = \mathrm{prox}_{\tau F}(x^k - \tau A^* y^k), \\ \bar x^{k+1} = 2x^{k+1} - x^k, \\ y^{k+1} = \mathrm{prox}_{\sigma G^*}(y^k + \sigma A \bar x^{k+1}). \end{cases} \tag{7.18}$$

[1]Eingeführt wurde das Verfahren in [Chambolle & Pock 2011], weshalb es in der Literatur auch oft als *Chambolle–Pock-Verfahren* bezeichnet wird; der Zusammenhang mit Proximalpunkt-Verfahren wurde in [He & Yuan 2012] beschrieben.

Wir zeigen nun, dass – unter Voraussetzungen an $\sigma, \tau$ – der modifizierte Operator $M$ selbstadjungiert und positiv definit ist bezüglich des Skalarproduktes

$$(z_1, z_2)_Z = (x_1, x_2)_X + (y_1, y_2)_Y \qquad \text{für alle } z_1 = (x_1, y_1) \in Z, z_2 = (x_2, y_2) \in Z.$$

Lemma 7.4. *Der lineare Operator $M : Z \to Z$ ist beschränkt und selbstadjungiert. Gilt $\sigma\tau\|A\|^2_{L(X,Y)} < 1$, so ist $M$ positiv definit.*

*Beweis.* Direkt aus der Definition von $M$ folgt die Beschränktheit (da $A \in L(X, Y)$ beschränkt ist) und Selbstadjungiertheit. Sei nun $z = (x, y) \in Z$ beliebig. Dann ist

$$\begin{aligned}
(Mz, z)_Z &= \left(\tau^{-1}x - A^*y, x\right)_X + \left(\sigma^{-1}y - Ax, y\right)_Y \\
&= \tau^{-1}\|x\|_X^2 - 2\,(x, A^*y)_X + \sigma^{-1}\|y\|_Y^2 \\
&\geq \tau^{-1}\|x\|_X^2 - 2\|A\|_{L(X,Y)}\|x\|_X\|y\|_Y + \sigma^{-1}\|y\|_Y^2 \\
&\geq \tau^{-1}\|x\|_X^2 - \|A\|_{L(X,Y)}\sqrt{\sigma\tau}(\tau^{-1}\|x\|_X^2 + \sigma^{-1}\|y\|_Y^2) + \sigma^{-1}\|y\|_Y^2 \\
&= (1 - \|A\|_{L(X,Y)}\sqrt{\sigma\tau})(\tau^{-1}\|x\|_X^2 + \sigma^{-1}\|y\|_Y^2) \\
&\geq C(\|x\|_X^2 + \|y\|_Y^2)
\end{aligned}$$

für $C := (1 - \|A\|_{L(X,Y)}\sqrt{\sigma\tau}) \min\{\tau^{-1}, \sigma^{-1}\} > 0$. Also ist $(Mz, z)_Z \geq C\|z\|_Z^2$ für alle $z \in Z$ und damit $M$ positiv definit. □

Der Operator $M$ induziert also ein Skalarprodukt $(z_1, z_2)_M := (Mz_1, z_2)_Z$ und dadurch eine Norm $\|z\|_M^2 = (z, z)_M$, für die gilt

$$c_1\|z\|_Z \leq \|z\|_M \leq c_2\|z\|_Z \qquad \text{für alle } z \in Z.$$

Folgerung 7.5. *Gilt $\sigma\tau\|A\|^2_{L(X,Y)} < 1$, so ist $M$ stetig invertierbar, d. h. es gilt $M^{-1} \in L(Z, Z)$.*

*Beweis.* Wegen (7.3) ist für beliebige $z \in Z$ die Abbildung $v \mapsto (z, v)_Z$ ein (bezüglich $\|\cdot\|_M$) beschränktes Funktional. Nach dem Satz von Fréchet–Riesz, angewendet auf den Hilbertraum $(Z, (\cdot,\cdot)_M)$, existiert also ein eindeutiges $z^* \in Z$ mit

$$(Mz^*, v)_Z = (z^*, v)_M = (z, v)_Z \qquad \text{für alle } v \in Z,$$

und die Abbildung $M^{-1} : z \mapsto z^*$ ist linear. Damit gilt

$$c_1^2\|z^*\|_Z^2 \leq \|z^*\|_M^2 = (Mz^*, z^*)_Z = (z, z^*)_Z \leq \|z\|_Z\|z^*\|_Z,$$

und nach Division durch $c_1^2\|z^*\|_Z$ folgt die behauptete Beschränktheit von $M^{-1}$. □

Also ist $M^{-1}T$ wohldefiniert, d. h. $\operatorname{graph} M^{-1}T \neq \emptyset$. Wir zeigen nun, dass $M^{-1}T$ maximal monoton ist bezüglich des Skalarproduktes $(\cdot,\cdot)_M$.

Lemma 7.6. *Gilt $\sigma\tau\|A\|^2_{L(X,Y)} < 1$, so ist $M^{-1}T$ maximal monoton in $(Z, (\cdot,\cdot)_M)$.*

*Beweis.* Wir zeigen zunächst die Monotonie. Sei $z \in Z$ und $z^* \in M^{-1}Tz$, d. h. $Mz^* \in Tz$. Nach Definition von $T$ existieren daher für $z = (x, y)$ ein $\xi \in \partial F(x)$ und ein $\eta \in \partial G^*(y)$ mit $Mz^* = (\xi + A^*y, \eta - Ax)$. Analog können wir für $\bar z = (\bar x, \bar y) \in Z$ und $\bar z^* \in M^{-1}T\bar z$ schreiben $M\bar z^* = (\bar\xi + A^*\bar y, \bar\eta - A\bar x)$ für $\bar\xi \in \partial F(\bar x)$ und $\bar\eta \in \partial G^*(\bar y)$. Dann gilt

$$\begin{aligned}(\bar z^* - z^*, \bar z - z)_M = (M\bar z^* - Mz^*, \bar z - z)_Z &= \big((\bar\xi + A^*\bar y) - (\xi + A^*y), \bar x - x\big)_X \\ &\quad + ((\bar\eta - A\bar x) - (\eta - Ax), \bar y - y)_Y \\ &= \big(\bar\xi - \xi, \bar x - x\big)_X + (A^*(\bar y - y), \bar x - x)_X \\ &\quad - (A(\bar x - x), \bar y - y)_Y + (\bar\eta - \eta, \bar y - y)_Y \\ &= \big(\bar\xi - \xi, \bar x - x\big)_X + (\bar\eta - \eta, \bar y - y)_Y \geq 0\end{aligned}$$

wegen der Monotonie der Subdifferentiale.

Für die maximale Monotonie seien nun $\bar z^*, \bar z \in Z$ und es gelte

$$(M\bar z^* - Mz^*, \bar z - z)_Z = (\bar z^* - z^*, \bar z - z)_M \geq 0 \quad \text{für alle } (z, z^*) \in \operatorname{graph} M^{-1}T, \tag{7.19}$$

d. h. für alle $z \in Z$ und $Mz^* \in Tz$. Wie oben ist nun $Mz^* = (\xi + A^*y, \eta - Ax)$ für ein $\xi \in \partial F(x)$ und ein $\eta \in \partial G^*(y)$. Setze nun $\bar\xi := \tilde x^* - A^*\bar y$ und $\bar\eta := \tilde y^* + A\bar x$ für $M\bar z^* = (\tilde x^*, \tilde y^*)$ und $\bar z = (\bar x, \bar y)$. Dann ist $M\bar z^* = (\bar\xi + A^*\bar y, \bar\eta - A\bar x)$, und aus (7.19) folgt daher für alle $(x, y) \in Z$, dass gilt

$$\begin{aligned}0 &\leq \big((\bar\xi + A^*\bar y) - (\xi + A^*y), \bar x - x\big)_X + ((\bar\eta - A\bar x) - (\eta - Ax), \bar y - y)_Y \\ &= \big(\bar\xi - \xi, \bar x - x\big)_X + (\bar\eta - \eta, \bar y - y)_Y\,.\end{aligned}$$

Setzen wir speziell Paare der Form $(x, \bar y)$ für $x \in X$ beliebig und $(\bar x, y)$ für $y \in Y$ beliebig ein, erhalten wir, dass beide Skalarprodukte auf der rechten Seiten nichtnegativ sind. Aus der maximalen Monotonie von Subdifferentialen folgt nun $\bar\xi \in \partial F(\bar x)$ und $\bar\eta \in \partial G^*(\bar y)$. Also ist

$$M\bar z^* = (\bar\xi + A^*\bar y, \bar\eta - A\bar x) \in T\bar z$$

und damit $\bar z^* \in M^{-1}T\bar z$. Dies zeigt die maximale Monotonie von $M^{-1}T$. □

Das primal-duale Extragradienten-Verfahren (7.18) ist also äquivalent mit der Proximalpunktiteration $z^{k+1} = R_{M^{-1}T}z^k$ für $M^{-1}T$ maximal monoton, so dass aus Satz 7.1 zusammen mit der Invertierbarkeit von $M$ die schwache Konvergenz folgt.

Satz 7.7. *Erfüllen $F, G, A$ die Voraussetzungen von Satz 5.6 und gilt $\sigma\tau\|A\|^2_{L(X,Y)} < 1$, so konvergiert $\{(x^k, y^k)\}_{k\in\mathbb{N}}$ schwach in $Z$ gegen eine Lösung $(\bar x, \bar y)$ von (7.15).*

*Beweis.* Aus Satz 5.6 folgt zunächst die Existenz eines Paares $\bar{z} := (\bar{x}, \bar{y})$, welches die Fenchel-Extremalitätsbedingungen (7.15) erfüllt. Nach Definition von $T$ ist dies äquivalent zu $0 \in T\bar{z}$, was wegen der Invertierbarkeit von $M$ aus Folgerung 7.5 genau dann gilt, falls $0 \in M^{-1}T\bar{z}$. Also hat $M^{-1}T$ eine Nullstelle. Nach Lemma 7.6 ist $M^{-1}T$ weiterhin maximal monoton (bezüglich $(\cdot,\cdot)_M$), so dass wir Satz 7.1 anwenden können. Daraus folgt

$$\left(z^k, Mw\right)_Z = (z_k, w)_M \to (\bar{z}, w)_M = (\bar{z}, Mw)_Z \quad \text{für alle } w \in Z$$

für ein $\bar{z} \in Z$ mit $0 \in M^{-1}T\bar{z}$ und damit auch $0 \in T\bar{z}$. Da $M$ invertierbar und damit insbesondere surjektiv ist, erhalten wir $(z_k, \tilde{w})_Z \to (\bar{z}, \tilde{w})_Z$ für alle $\tilde{w} := Mw \in Z$ und damit die behauptete schwache Konvergenz. □

Beachten Sie, dass das Verfahren durch die Proximalpunktformulierung der Fenchel-Extremalitätsbedingungen zwar implizit in $F$ und $G$, nicht jedoch in $A$ ist; es ist also nicht verwunderlich, dass dadurch eine Schrittweitenrestriktion an $\tau$ und $\sigma$ durch $A$ entsteht.[2]

## 7.4 STARKE KONVERGENZ UND RATEN

Der Kern der bisherigen Konvergenzresultate war, aus der Iterationsvorschrift und der Monotonie der Operatoren, deren Nullstelle $\bar{x}$ gesucht war, die Beschränktheit des Fehlers $\|x^k - \bar{x}\|_X$ und damit die schwache Konvergenz (zunächst einer Teilfolge) zu erhalten. Für starke Konvergenz brauchen wir zusätzliche Eigenschaften, die eine direktere Verknüpfung der Vorschrift mit dem Fehler liefern. Eine Möglichkeit ist, *starke Monotonie* zu fordern: Eine mengenwertige Abbildung $H : X \rightrightarrows X$ heißt *stark monoton*, falls ein $\gamma > 0$ existiert mit

$$(x_1^* - x_2^*, x_1 - x_2)_X \geq \gamma \|x_1 - x_2\|_X^2 \quad \text{für alle } (x_1, x_1^*), (x_2, x_2^*) \in \operatorname{graph} H. \tag{7.20}$$

Zum Beispiel ist $H = \partial F$ für $F(x) = \frac{1}{2}\|x\|_X^2$ offensichtlich stark monoton mit $\gamma = 1$; allgemeiner ist $\partial F$ stark monoton, wenn $F - \frac{\gamma}{2}\|x\|_X^2$ immer noch konvex ist.

Unter dieser Voraussetzung zeigen wir beispielhaft die starke Konvergenz des Proximalpunktverfahrens.

**Satz 7.8.** *Sei $H : X \rightrightarrows X$ stark monoton mit $\gamma > 0$ und $\bar{x} \in X$ mit $0 \in H(\bar{x})$. Sei weiterhin $x^0 \in X$ beliebig und $\{\tau_k\}_{k\in\mathbb{N}} \subset (0,\infty)$. Dann gilt für die durch das Proximalpunktverfahren $x^{k+1} = \mathcal{R}_{\tau_k H}(x^k)$ erzeugte Folge:*

[2] Die Proximalpunktabbildung zu $G \circ A$ würde zu einem voll impliziten Verfahren führen, dafür aber aufgrund der Rechenregeln die Anwendung von $A^{-1}$ benötigen. Es ist gerade Sinn und Zweck des primal-dualen Extragradientenverfahrens, die Invertierung von $A$ – was in der Praxis oft schlecht konditioniert oder sogar unmöglich ist – zu vermeiden.

(i) *Ist* $\tau_k \equiv \tau$ *konstant, so konvergiert* $x^k \to x$ *linear, d. h.* $\lim_{k\to\infty} \frac{\|x^{k+1}-\bar x\|_X}{\|x^k-\bar x\|_X} = \mu < 1$.

(ii) *Gilt* $\tau_k \to \infty$*, so konvergiert* $x^k \to x$ *superlinear, d. h.* $\lim_{k\to\infty} \frac{\|x^{k+1}-\bar x\|_X}{\|x^k-\bar x\|_X} = 0$.

*Beweis.* Nach Definition der Resolvente ist die Proximalpunktiteration äquivalent zu

$$-\frac{1}{\tau_k}(x^{k+1}-x^k) \in H(x^{k+1}).$$

Zusammen mit $0 \in H(\bar x)$ folgt daher aus (7.20)

$$-\left(x^{k+1}-x^k, x^{k+1}-\bar x\right)_X \geq \tau_k \gamma \|x^{k+1}-\bar x\|_X^2.$$

Wir wenden nun auf die linke Seite die leicht zu verifizierende *Pythagoras-Identität*

$$(x-y, x-z)_X = \frac{1}{2}\|x-y\|_X^2 - \frac{1}{2}\|y-z\|_X^2 + \frac{1}{2}\|x-z\|_X^2 \quad \text{für alle } x, y, z \in X$$

an und erhalten nach Umsortieren

$$\frac{1+2\tau_k\gamma}{2}\|x^{k+1}-\bar x\|_X^2 + \frac{1}{2}\|x^{k+1}-x^k\|_X^2 \leq \frac{1}{2}\|x^k-\bar x\|_X^2.$$

Insbesondere folgt daraus

$$\frac{\|x^{k+1}-\bar x\|_X^2}{\|x^k-\bar x\|_X^2} \leq \frac{1}{1+2\tau_k\gamma}.$$

Wir machen nun die Fallunterscheidung:

(i) Für $\tau_k \equiv \tau$ ist $\mu := (1+2\tau\gamma)^{-1/2} < 1$ und damit konvergiert $x^k$ linear.

(ii) Für $\tau_k \to \infty$ gilt $(1+2\tau_k\gamma)^{-1/2} \to 0$ und damit konvergiert $x^k$ superlinear. □

Analog zeigt man mit etwas Aufwand die lineare Konvergenz des expliziten Splittings für $F$ Lipschitz-stetig differenzierbar und $G$ stark konvex (aber nicht die superlineare Konvergenz, da $\tau_k \leq L^{-1}$ beschränkt bleiben muss). Für das primal-duale Splitting ist das (mit deutlich mehr Aufwand) ebenfalls möglich, wobei nun die Schrittweiten und die gewünschte Konvergenzrate mit Hilfe eines *Test-Operators* bereits in die Definition der starken Monotonie (und der Pythagoras-Identität) eingebracht werden müssen; siehe [Valkonen 2018].

# Teil III

# LIPSCHITZ-ANALYSIS

# 8 DAS CLARKE-SUBDIFFERENTIAL

Wir suchen nun nach einem verallgemeinerten Ableitungskonzept, das sowohl die Fréchet-Ableitung als auch das konvexe Subdifferential verallgemeinert. Dafür brauchen wir zuerst eine geeignete, größere, Funktionenklasse. Dabei ist klar, dass wir irgendeine Form der Stetigkeit voraussetzen müssen, denn ohne diese bestünde kein Zusammenhang zwischen Funktionalwerten an benachbarten Punkten und damit keine Hoffnung, aus punktweisen Informationen auf (lokale) Optimalität schließen zu können. In Teil II war dies die Unterhalbstetigkeit, die zusammen mit der Konvexität die nötigen Eigenschaften garantiert hat. Die letztere globale Forderung lassen wir nun fallen; dafür benötigen wir stärkere lokale Stetigkeitsforderungen. Wir betrachten daher in diesem Kapitel lokal Lipschitz-stetige Funktionen. Zur Erinnnerung: Eine Funktion $F : X \to \mathbb{R}$ ist lokal Lipschitz-stetig in $x \in X$, falls ein $\delta > 0$ und ein $L > 0$ (womit wir im Folgenden stets die Lipschitz-Konstante bezeichnen werden) existieren mit

$$|F(x_1) - F(x_2)| \le L\|x_1 - x_2\|_X \qquad \text{für alle } x_1, x_2 \in O_\delta(x).$$

Beachte, dass wir dafür $F$ als (lokal) endlich-wertig voraussetzen müssen.

Wir gehen analog zum konvexen Subdifferential vor und definieren zunächst die *verallgemeinerte Richtungsableitung* in $x \in X$ in Richtung $h \in X$ über

$$F^\circ(x; h) := \limsup_{\substack{y \to x\\ t \to 0^+}} \frac{F(y + th) - F(y)}{t}. \tag{8.1}$$

Beachte den Unterschied zur Richtungsableitung: Es wird nicht verlangt, dass irgendein Grenzwert existiert, nur Häufungspunkte. Wir benötigen die folgenden Eigenschaften.

**Lemma 8.1.** *Sei $F : X \to \mathbb{R}$ lokal Lipschitz-stetig in $x$. Dann ist die Abbildung $h \mapsto F^\circ(x; h)$*

*(i) Lipschitz-stetig mit Konstante $L$ und erfüllt $|F^\circ(x; h)| \le L\|h\|_X < \infty$;*

*(ii) subadditiv, d. h. $F^\circ(x; h + g) \le F^\circ(x; h) + F^\circ(x; g)$ für alle $h, g \in X$;*

*(iii) positiv homogen, d. h. für alle $\alpha > 0$ ist $F^\circ(x; \alpha h) = (\alpha F)^\circ(x; h)$ für alle $h \in X$;*

*(iv) reflektiv, d. h. $F^\circ(x; -h) = (-F)^\circ(x; h)$ für alle $h \in X$.*

*Beweis.* Zu (i): Seien $h, g \in X$ beliebig. Aus der lokalen Lipschitz-Stetigkeit von $F$ folgt dann

$$F(y+th) - F(y) \leq F(y+tg) - F(y) + tL\|h-g\|_X$$

für alle $y$ hinreichend nahe an $x$ und $t$ klein genug. Division durch $t$ und Bilden des Limes superior ergibt dann

$$F^\circ(x;h) \leq F^\circ(x;g) + L\|h-g\|_X.$$

Vertauschen von $h$ und $g$ liefert dann die Lipschitz-Stetigkeit. Weiter folgt direkt aus der Definition $F^\circ(x;g) = 0$ für $g = 0$ und damit auch die behauptete Beschränktheit.

Zu (ii): Aus der Definition des Limes superior und der produktiven Null folgt sofort

$$\begin{aligned} F^\circ(x;h+g) &= \limsup_{\substack{y\to x\\ t\to 0^+}} \frac{F(y+th+tg)-F(y)}{t} \\ &\leq \limsup_{\substack{y\to x\\ t\to 0^+}} \frac{F(y+th+tg)-F(y+tg)}{t} + \limsup_{\substack{y\to x\\ t\to 0^+}} \frac{F(y+tg)-F(y)}{t} \\ &= F^\circ(x;h) + F^\circ(x;g), \end{aligned}$$

da für $y \to x$ und $t \to 0$ auch $y + tg \to x$ konvergiert und umgekehrt.

Zu (iii): Direkt aus der Definition folgt wegen $\alpha > 0$

$$\begin{aligned} F^\circ(x;\alpha h) &= \limsup_{\substack{y\to x\\ t\to 0^+}} \frac{F(y-t(\alpha h))-F(y)}{t} \\ &= \limsup_{\substack{y\to x\\ \alpha t\to 0^+}} \alpha\frac{F(y+(\alpha t)h)-F(y)}{\alpha t} = (\alpha F)^\circ(x;h). \end{aligned}$$

Zu (iv): Ebenso folgt

$$\begin{aligned} F^\circ(x;-h) &= \limsup_{\substack{y\to x\\ t\to 0^+}} \frac{F(y-th)-F(y)}{t} \\ &= \limsup_{\substack{w\to x\\ t\to 0^+}} \frac{-F(w+th)-(-F(w))}{t} = (-F)^\circ(x;h), \end{aligned}$$

da für $y \to x$ und $t \to 0$ auch $w := y - th \to x$ konvergiert und umgekehrt. □

Aus Lemma 8.1 (i–iii) folgt insbesondere, dass $h \mapsto F^\circ(x;h)$ eigentlich, konvex und unterhalbstetig ist.

Wir definieren nun für $F : X \to \mathbb{R}$ lokal Lipschitz-stetig das *Clarke-Subdifferential* in $x \in X$ durch

$$\partial_C F(x) := \{x^* \in X^* : \langle x^*, h\rangle_X \leq F^\circ(x;h) \quad \text{für alle } h \in X\}. \tag{8.2}$$

Direkt aus der Definition und Lemma 8.1 (i) folgt, dass $\partial_C F(x)$ konvex, schwach-$*$ abgeschlossen und beschränkt (da $\|\xi\|_{X^*} \leq L$ für alle $\xi \in \partial_C F(x)$ nach Definition der Norm im Dualraum) ist. Auch hier gilt nach Konstruktion das Fermat-Prinzip.

Satz 8.2. *Hat $F : X \to \mathbb{R}$ ein lokales Minimum in $\bar{x}$, so gilt $0 \in \partial_C F(\bar{x})$.*

*Beweis.* Ist $\bar{x} \in X$ ein lokaler Minimierer von $F$, so gilt insbesondere $F(\bar{x}) \leq F(\bar{x} + th)$ für alle $h \in X$ und $t > 0$ hinreichend klein. Daraus folgt

$$\langle 0, h\rangle_X = 0 \leq \liminf_{t\to 0^+} \frac{F(\bar{x}+th) - F(\bar{x})}{t} \leq \limsup_{t\to 0^+} \frac{F(\bar{x}+th) - F(\bar{x})}{t} \leq F^\circ(x; h)$$

und damit $0 \in \partial_C F(\bar{x})$ nach Definition. □

Beachte, dass $F$ nicht konvex und damit die Bedingung nicht mehr hinreichend sein muss! (Betrachte z. B. $f(t) = -|t|$.)[1]

Wir werden später auch noch die folgende Abgeschlossenheitseigenschaft benötigen.

Lemma 8.3. *Sei $F : X \to \mathbb{R}$ und $\{x_n\}_{n\in\mathbb{N}}$ eine Folge mit $x_n \to x \in X$ sowie $x_n^* \in \partial_C F(x_n)$ für alle $n \in \mathbb{N}$ mit $x_n^* \rightharpoonup^* x^*$ in $X^*$. Dann ist $x^* \in \partial_C F(x)$.*

*Beweis.* Sei $h \in X$ beliebig. Nach Annahme gilt dann $\langle x_n^*, h\rangle_X \leq F^\circ(x_n; h)$ für alle $n \in \mathbb{N}$. Aus der schwach-$*$ Konvergenz folgt dann

$$\langle x^*, h\rangle_X = \lim_{n\to\infty} \langle x_n^*, h\rangle_X \leq \limsup_{n\to\infty} F^\circ(x_n; h).$$

Wir sind also fertig, wenn wir zeigen können, dass $\limsup_{n\to\infty} F^\circ(x_n; h) \leq F^\circ(x; h)$ gilt (denn dann ist nach Definition $x^* \in \partial_C F(x)$).

Nach Definition von $F^\circ(x_n; h)$ existieren Folgen $\{y_{n,m}\}_{m\in\mathbb{N}}$ und $\{t_{n,m}\}_{m\in\mathbb{N}}$ mit $y_{n,m} \to x_n$ und $t_{n,m} \to 0$ für $m \to \infty$, die den Limes superior realisieren. Wir finden also für alle $n \in \mathbb{N}$ ein $y_n := y_{n,m(n)}$ und $t_n := t_{n,m(n)}$ mit $\|y_n - x_n\|_X + t_n < n^{-1}$ (und damit insbesondere $y_n \to x$ und $t_n \to 0$) sowie

$$F^\circ(x_n; h) - \tfrac{1}{n} \leq \frac{F(y_n + t_n h) - F(y_n)}{t_n}$$

für $n$ hinreichend groß. Übergang zum Limes superior für $n \to \infty$ auf beiden Seiten liefert dann die gewünschte Ungleichung. □

[1] Ähnlich wie in Satz 4.3 musste auch hier die Lipschitz-Stetigkeit von $F$ nicht vorausgesetzt werden – das Fermat-Prinzip für das Clarke-Subdifferential charakterisiert (unter anderem) *jedes* lokale Minimum. Will man aber diese Charakterisierung verwenden, um nachzuprüfen ob ein gegebenes $\bar{x} \in X$ tatsächlich ein (Kandidat für einen) Minimierer ist, benötigt man eine vernünftige Darstellung des Subdifferentials – und das ist nur für (bestimmte) lokal Lipschitz-stetige Funktionale möglich.

Wir zeigen nun, dass das Clarke-Subdifferential in der Tat eine Verallgemeinerung der bereits bekannten Ableitungen darstellt.

Satz 8.4. *Sei $F : X \to \mathbb{R}$ stetig Fréchet-differenzierbar in einer Umgebung $U$ um $x \in X$. Dann ist $\partial_C F(x) = \{F'(x)\}$.*

*Beweis.* Zunächst folgt aus der Annahme die lokale Lipschitz-Stetigkeit: Da $F'$ in $U$ stetig ist, gibt es ein $\delta > 0$ mit $\|F'(z) - F'(x)\|_{X^*} \leq 1$ und damit $\|F'(z)\|_{X^*} \leq 1 + \|F'(x)\|_{X^*}$ für alle $z \in K_\delta(x) \subset U$. Aus der Konvexität von Kugeln folgt dann $x_1, x_2 \in K_\delta(x)$ beliebig auch $x_2 + t(x_1 - x_2) \in K_\delta(x)$ für alle $t \in [0,1]$ und damit aus Satz 2.6

$$\begin{aligned} |F(x_1) - F(x_2)| &\leq \int_0^1 \|F'(x_2 + t(x_1 - x_2))\|_{X^*} t \|x_1 - x_2\|_X \, dt \\ &\leq \frac{1 + \|F'(x)\|_{X^*}}{2} \|x_1 - x_2\|_X. \end{aligned}$$

Wir zeigen nun $F^\circ(x;h) = F'(x)h$ $(= F'(x;h))$ für alle $h \in X$. Seien dafür $\{y_n\}_{n\in\mathbb{N}}$ und $\{t_n\}_{n\in\mathbb{N}}$ Folgen mit $y_n \to x$ und $t_n \to 0^+$, die den Limes superior realisieren. Dann folgt wieder aus dem Mittelwertsatz und der Stetigkeit von $F'$ für $h \in X$ beliebig

$$\begin{aligned} F^\circ(x;h) &= \lim_{n\to\infty} \frac{F(y_n + t_n h) - F(y_n)}{t_n} \\ &= \lim_{n\to\infty} \int_0^1 \frac{1}{t_n} \langle F'(y_n + t(t_n h)), t_n h\rangle_X \, dt \\ &= \langle F'(x), h\rangle_X, \end{aligned}$$

da der Integrand gleichmäßig in $t \in [0,1]$ gegen $\langle F'(x), h\rangle_X$ konvergiert. Also ist nach Definition $x^* \in \partial_C F(x)$ genau dann, wenn $\langle x^*, h\rangle_X \leq \langle F'(x), h\rangle_X$ für alle $h \in X$ gilt, was nur für $x^* = F'(x)$ möglich ist. □

Satz 8.5. *Sei $F : X \to \mathbb{R}$ konvex und unterhalbstetig. Dann ist $\partial_C F(x) = \partial F(x)$ für alle $x \in X$.*

*Beweis.* Sei $x \in X$ beliebig. Da $F$ endlichwertig ist, gilt $(\operatorname{dom} F)^o = X$, und damit ist $F$ nach Satz 3.10 lokal Lipschitz-stetig in $x$. Wir zeigen nun $F^\circ(x;h) = F'(x;h)$ für alle $h \in X$, woraus mit der motivierenden Definition (4.2) des konvexen Subdifferentials (die nach Lemma 4.2 mit Definition (4.3) äquivalent ist) die Aussage folgt. Zunächst gilt stets

$$F'(x;h) = \lim_{t\to 0^+} \frac{F(x+th) - F(x)}{t} \leq \limsup_{\substack{y\to x \\ t\to 0^+}} \frac{F(y+th) - F(y)}{t} = F^\circ(x;h).$$

Für die umgekehrte Richtung sei $\delta > 0$ beliebig. Da nach Lemma 4.1 (i) für konvexe Funktionale der Differenzenquotient monoton wachsend ist, können wir abschätzen

$$\begin{aligned} F^\circ(x;h) &= \lim_{\varepsilon\to 0^+} \sup_{y\in K_{\delta\varepsilon}(x)} \sup_{0<t<\varepsilon} \frac{F(y+th)-F(y)}{t} \\ &\leq \lim_{\varepsilon\to 0^+} \sup_{y\in K_{\delta\varepsilon}(x)} \frac{F(y+\varepsilon h)-F(y)}{\varepsilon} \\ &\leq \lim_{\varepsilon\to 0^+} \frac{F(x+\varepsilon h)-F(x)}{\varepsilon} + 2L\delta \\ &= F'(x;h) + 2L\delta, \end{aligned}$$

wobei wir im vorletzten Schritt zweimal die produktive Null zusammen mit der lokalen Lipschitz-Stetigkeit in $x$ verwendet haben. Da $\delta > 0$ beliebig war, muss also $F^\circ(x;h) \leq F'(x;h)$ gelten. □

Ist $F : X \to \mathbb{R}$ lokal Lipschitz-stetig und gilt $F^\circ(x;h) = F'(x;h)$ für alle $h \in X$, so heißt $F$ *regulär* in $x$. Wie wir gerade gezeigt haben, ist jedes stetig differenzierbare sowie jedes konvexe und unterhalbstetige Funktional regulär. Anschaulich sind Funktionen also regulär in Punkten, in denen sie entweder differenzierbar sind oder einen "konvexen Knick" haben.

Wir kommen nun zu Rechenregeln. Die erste folgt wieder direkt aus der Definition.

Satz 8.6. *Sei $F : X \to \mathbb{R}$ lokal Lipschitz-stetig in $x \in X$. Dann gilt für alle $\alpha \in \mathbb{R}$*

$$\partial_C(\alpha F)(x) = \alpha\partial_C(F)(x).$$

*Beweis.* Zunächst ist für $\alpha \in \mathbb{R}$ offensichtlich $\alpha F$ lokal Lipschitz-stetig. Für $\alpha = 0$ steht nach Satz 8.4 auf beiden Seiten $\{0\}$. Für alle $\alpha > 0$ ist nach Definition stets $(\alpha F)^\circ(x;h) = \alpha F^\circ(x;h)$ für alle $h \in X$. Es gilt also

$$\begin{aligned} \alpha\partial_C F(x) &= \{\alpha x^* \in X^* : \langle x^*, h\rangle_X \leq F^\circ(x;h) \quad \text{für alle } h \in X\} \\ &= \{\alpha x^* \in X^* : \langle \alpha x^*, h\rangle_X \leq \alpha F^\circ(x;h) \quad \text{für alle } h \in X\} \\ &= \{y^* \in X^* : \langle y^*, h\rangle_X \leq (\alpha F)^\circ(x;h) \quad \text{für alle } h \in X\} \\ &= \partial_C(\alpha F)(x). \end{aligned}$$

Für den allgemeinen Fall muss daher nur noch die Aussage für $\alpha = -1$ gezeigt werden. Dazu schreiben wir analog mit Lemma 8.1 (iv)

$$\begin{aligned} \partial_C(-F)(x) &= \{x^* \in X^* : \langle x^*, h\rangle_X \leq (-F)^\circ(x;h) \quad \text{für alle } h \in X\} \\ &= \{x^* \in X^* : \langle -x^*, -h\rangle_X \leq F^\circ(x;-h) \quad \text{für alle } h \in X\} \\ &= \{-y^* \in X^* : \langle y^*, g\rangle_X \leq F^\circ(x;g) \quad \text{für alle } g \in X\} \\ &= -\partial_C(F)(x). \end{aligned}$$

□

Folgerung 8.7. *Sei $F : X \to \mathbb{R}$ lokal Lipschitz-stetig in $\bar{x}$. Hat $F$ ein lokales Maximum in $\bar{x}$, so gilt $0 \in \partial_C F(\bar{x})$.*

*Beweis.* Hat $F$ ein lokales Maximum in $\bar{x}$, so hat $-F$ ein lokales Minimum. Aus Satz 8.2 folgt dann

$$0 \in \partial_C(-F)(\bar{x}) = -\partial_C F(\bar{x}),$$

d. h. $0 = -0 \in \partial_C F(\bar{x})$. □

Die weiteren Regeln sind wieder aufwändiger. Dazu müssen wir Mengen der Form (8.2) vergleichen, wofür wir wiederholt die folgenden Argumente anwenden.

Lemma 8.8. *Sei $F : X \to \mathbb{R}$ positiv homogen, subadditiv und unterhalbstetig, und sei*

$$A = \{x^* \in X^* : \langle x^*, x\rangle_X \leq F(x) \quad \text{für alle } x \in X\}.$$

*Dann ist*

$$F(x) = \sup_{x^* \in A} \langle x^*, x\rangle_X \qquad \text{für alle } x \in X.$$

*Beweis.* Nach Definition gilt für alle $x^* \in A$ (und nur diese!) die Ungleichung $\langle x^*, x\rangle_X - F(x) \leq 0$ für alle $x \in X$. Durch Fallunterscheidung analog zu Beispiel 5.2 (iii) folgt daraus zusammen mit der positiven Homogenität von $F$ für $x^* \in X^*$ beliebig

$$F^*(x^*) = \sup_{x \in X} \langle x^*, x\rangle_X - F(x) = \begin{cases} 0 & x^* \in A, \\ \infty & x^* \notin A, \end{cases}$$

d. h. $F^* = \delta_A$. Nun ist $F$ nach Voraussetzung zusätzlich subadditiv und damit konvex sowie unterhalbstetig. Aus Satz 5.1 folgt daher für alle $x \in X$

$$F(x) = F^{**}(x) = (\delta_A)^*(x) = \sup_{x^* \in A} \langle x^*, x\rangle_X.$$ □

Man nennt die Abbildung $x \mapsto \sup_{x^* \in A} \langle x^*, x\rangle_X$ auch *Trägerfunktional* von $A$. Die nächsten Aussagen zeigen, dass man Mengen äquivalent durch ihre Trägerfunktionale charakterisieren kann.

Lemma 8.9. *Seien $A, B \subset X^*$ nichtleer, konvex, und schwach-$*$ abgeschlossen. Dann ist $A \subset B$ genau dann, wenn gilt*

$$\sup_{x^* \in A} \langle x^*, x\rangle_X \leq \sup_{x^* \in B} \langle x^*, x\rangle_X \qquad \text{für alle } x \in X. \tag{8.3}$$

*Beweis.* Ist $A \subset B$, so ist die rechte Seite von (8.3) offensichtlich mindestens so groß wie die linke. Für die andere Richtung sei angenommen, es existiert ein $x^* \in A$ mit $x^* \notin B$. Wegen der Voraussetzung an $A$ und $B$ existiert dann nach Satz 1.11 ein $x \in X$ und ein $\lambda \in \mathbb{R}$ mit

$$\langle z^*, x \rangle_X \leq \lambda < \langle x^*, x \rangle_X \qquad \text{für alle } z^* \in B.$$

Supremum über alle $z^*$ und Abschätzen von $x^*$ durch das Supremum ergibt dann

$$\sup_{z^* \in B} \langle z^*, x \rangle_X \leq \lambda < \sup_{x^* \in A} \langle x^*, x \rangle_X.$$

Damit ist (8.3) verletzt, und die Aussage folgt durch Kontraposition. □

Folgerung 8.10. *Seien $A, B \subset X^*$ konvex und schwach-∗ abgeschlossen. Dann ist $A = B$ genau dann, wenn gilt*

$$\sup_{x^* \in A} \langle x^*, x \rangle_X = \sup_{x^* \in B} \langle x^*, x \rangle_X \qquad \textit{für alle } x \in X. \tag{8.4}$$

*Beweis.* Die eine Richtung ist klar. Gilt umgekehrt (8.4), dann folgt daraus insbesondere (8.3) und damit nach Lemma 8.9 auch $A \subset B$. Vertauschen der Rollen von $A$ und $B$ ergibt die Aussage. □

Damit können wir zum Beispiel eine Summenregel beweisen.

Satz 8.11. *Seien $F, G : X \to \mathbb{R}$ lokal Lipschitz-stetig in $x \in X$. Dann gilt*

$$\partial_C(F + G)(x) \subset \partial_C F(x) + \partial_C G(x).$$

*Sind $F$ und $G$ regulär in $x$, so auch $F + G$ und es gilt Gleichheit.*

*Beweis.* Man sieht sofort die lokale Lipschitz-Stetigkeit von $F + G$ in $x$. Aus der Definition des Limes superior folgt weiter für beliebiges $h \in X$

$$(F + G)^\circ(x; h) \leq F^\circ(x; h) + G^\circ(x; h).$$

Sind $F$ und $G$ regulär in $x$, so erhalten wir aus den Rechenregeln für konvergente Folgen

$$F^\circ(x; h) + G^\circ(x; h) = F'(x; h) + G'(x; h) = (F + G)'(x; h) \leq (F + G)^\circ(x; h)$$

und damit $(F + G)^\circ(x; h) = (F + G)'(x; h)$ und damit die behauptete Regularität.

Nach Definition sind wir also fertig, wenn wir gezeigt haben, dass gilt

$$\partial_C F(x) + \partial_C G(x) = \{x^* \in X^* : \langle x^*, h \rangle_X \leq F^\circ(x; h) + G^\circ(x; h) \text{ für alle } h \in X\} =: A.$$

Dafür verwenden wir, dass beide Mengen konvex und schwach-∗ abgeschlossen sind, sowie dass verallgemeinerte Richtungsableitungen und damit auch ihre Summe nach Lemma 8.1 positiv homogen, subadditiv und unterhalbstetig sind. Nach Lemma 8.8 gilt daher

$$\begin{aligned}\sup_{x^*\in\partial_C F(x)+\partial_C G(x)} \langle x^*, h\rangle_X &= \sup_{x_1^*\in\partial_C F(x)} \langle x_1^*, h\rangle_X + \sup_{x_2^*\in\partial_C G(x)} \langle x_2^*, h\rangle_X \\ &= F^\circ(x;h) + G^\circ(x;h) = \sup_{x^*\in A} \langle x^*, h\rangle_X\end{aligned}$$

für alle $h \in X$, woraus mit Folgerung 8.10 die behauptete Inklusion bzw. Gleichheit der Subdifferentiale folgt. □

Beachten Sie den Unterschied zur konvexen Summenregel: Die Inklusion ist genau umgekehrt; auch müssen jetzt *beide* Summanden regulär sein, und zwar in genau dem Punkt, in dem die Summenregel angewendet wird. Wieder erhält man per Induktion Summenregeln für mehr Summanden (wobei diesmal alle regulär sein müssen).

Für die Kettenregel brauchen wir den folgenden „nichtglatten“ Mittelwertsatz.

Satz 8.12. *Sei $F : X \to \mathbb{R}$ lokal Lipschitz-stetig in $x \in X$. Dann existieren für alle $\tilde{x} \in X$ hinreichend nahe an $x$ ein $\lambda \in (0,1)$ und ein $x^* \in \partial_C F(x + \lambda(\tilde{x} - x))$ mit*

$$F(\tilde{x}) - F(x) = \langle x^*, \tilde{x} - x\rangle_X.$$

*Beweis.* Wir definieren $\psi, \varphi : [0,1] \to \mathbb{R}$ durch

$$\psi(\lambda) := F(x + \lambda(\tilde{x} - x)), \qquad \varphi(\lambda) := \psi(\lambda) + \lambda(F(x) - F(\tilde{x})).$$

Da $F$ lokal Lipschitz-stetig und $\tilde{x}$ nahe genug an $x$ ist, sind sowohl $\psi$ als auch $\varphi$ Lipschitz-stetig. Wegen $\varphi(0) = F(x) = \varphi(1)$ hat $\varphi$ daher ein lokales Minimum oder Maximum in einem Punkt $\bar{\lambda} \in (0,1)$. Nach Satz 8.2 oder Folgerung 8.7 zusammen mit der Summenregel aus Satz 8.11 sowie Satz 8.4 ist daher

$$0 \in \partial_C \varphi(\bar{\lambda}) \subset \partial_C \psi(\bar{\lambda}) + \{F(x) - F(\tilde{x})\}.$$

Wir sind also fertig, wenn wir zeigen können, dass für $x_{\bar{\lambda}} := x + \bar{\lambda}(\tilde{x} - x)$ gilt

$$\partial_C \psi(\bar{\lambda}) \subset \left\{\langle x^*, \tilde{x} - x\rangle_X : x^* \in \partial_C F(x_{\bar{\lambda}})\right\} =: A. \tag{8.5}$$

Dafür betrachten wir zunächst für $s \in \mathbb{R}$ beliebig die verallgemeinerte Richtungsableitung

$$\begin{aligned}\psi^\circ(\bar{\lambda}; s) &= \limsup_{\substack{\lambda\to\bar{\lambda}\\ t\to 0^+}} \frac{\psi(\bar{\lambda} + ts) - \psi(\bar{\lambda})}{t} \\ &= \limsup_{\substack{\lambda\to\bar{\lambda}\\ t\to 0^+}} \frac{F(x + (\lambda + ts)(\tilde{x} - x)) - F(x + \lambda(\tilde{x} - x))}{t} \\ &\le \limsup_{\substack{z\to x_{\bar{\lambda}}\\ t\to 0^+}} \frac{F(z + ts(\tilde{x} - x)) - F(z)}{t} = F^\circ(x_{\bar{\lambda}}; s(\tilde{x} - x)),\end{aligned}$$

da wir im letzten Limes superior *beliebige* Folgen $z \to x_{\bar\lambda}$ (anstelle von Folgen der speziellen Form $z_n = x + \lambda_n(\tilde x - x)$) betrachten. Nach Definition gilt daher

$$\partial_C\psi(\bar\lambda) \subset \left\{t^* \in \mathbb{R} : t^*s \le F^\circ(x_{\bar\lambda}; s(\tilde x - x)) \text{ für alle } s \in \mathbb{R}\right\} =: B. \tag{8.6}$$

Bleibt nur noch zu zeigen, dass die Mengen $A$ und $B$ in (8.5) bzw. (8.6) übereinstimmen. Dies folgt aber wieder aus Lemma 8.8 und Folgerung 8.10, denn für alle $s \in \mathbb{R}$ gilt

$$\sup_{t^*\in A} t^*s = \sup_{x^*\in\partial_C F(x_{\bar\lambda})} \langle x^*, s(\tilde x - x)\rangle_X = F^\circ(x_{\bar\lambda}; s(\tilde x - x)) = \sup_{t^*\in B} t^*s. \qquad \square$$

Wir kommen nun zur Kettenregel, für die im Gegensatz zum konvexen Fall die innere Abbildung nicht mehr linear sein muss; dies ist einer der wesentlichen Vorzüge des Clarke-Subdifferentials in unserem Kontext.

Satz 8.13. *Sei $Y$ separabel, sei $F : X \to Y$ stetig differenzierbar in $x \in X$, und sei $G : Y \to \mathbb{R}$ lokal Lipschitz-stetig in $F(x)$. Dann gilt*

$$\partial_C(G \circ F)(x) \subset F'(x)^*\partial_C G(F(x)) := \{F'(x)^*y^* : y^* \in \partial_C G(F(x))\}.$$

*Ist $G$ regulär in $F(x)$, so ist auch $G \circ F$ regulär in $x$, und es gilt Gleichheit.*

*Beweis.* Die Lipschitz-Stetigkeit von $G \circ F$ folgt sofort aus der von $G$ und $F$ (analog zu Satz 8.4. Für die Mengeninklusion bzw. Gleichheit argumentieren wir wie zuvor. Zuerst zeigen wir, dass für jedes $h \in X$ ein $y^* \in \partial_C G(F(x))$ existiert mit

$$(G \circ F)^\circ(x; h) = \langle y^*, F'(x)h\rangle_Y. \tag{8.7}$$

Dafür betrachte zwei Folgen $\{x_n\}_{n\in\mathbb{N}} \subset X$ und $\{t_n\}_{n\in\mathbb{N}} \subset (0,\infty)$ mit $x_n \to x$, $t_n \to 0$ und

$$(G \circ F)^\circ(x; h) = \lim_{n\to\infty} \frac{G(F(x_n + t_nh)) - G(F(x_n))}{t_n}.$$

Nun ist $F$ insbesondere stetig; wir können also ein $n_0 \in \mathbb{N}$ finden, so dass für alle $n \ge n_0$ sowohl $x_n$ nahe genug an $x$ als auch $t_n$ klein genug ist, dass $F(x_n)$ und $F(x_n + t_nh)$ nahe genug an $F(x)$ liegen, um Satz 8.12 anwenden zu können. Es existiert also für alle $n \ge n_0$ ein $\lambda_n \in (0,1)$ und ein $y_n^* \in \partial_C G(y_n)$ für $y_n := F(x_n) + \lambda_n(F(x_n + t_nh) - F(x_n))$ mit

$$\frac{G(F(x_n + t_nh)) - G(F(x_n))}{t_n} = \left\langle y_n^*, \frac{F(x_n + t_nh) - F(x_n)}{t_n}\right\rangle_Y. \tag{8.8}$$

Wegen $x_n \to x$, $t_n \to 0$, und $\lambda_n \in (0,1)$ sowie der Stetigkeit von $F$ gilt nun $y_n \to F(x)$; für $n \in \mathbb{N}$ groß genug ist $y_n$ also ebenfalls nahe genug an $F(x)$, so dass dort $y_n^* \in \partial_C G(y_n) \subset K_L(0)$ aufgrund der Lipschitz-Stetigkeit von $G$ gilt. Also ist $\{y_n^*\}_{n\in\mathbb{N}} \subset Y^*$ beschränkt und hat daher nach dem Satz 1.10 von Banach–Alaoglu eine schwach-$*$ konvergente Teilfolge, für dessen Grenzwert nach Lemma 8.3 gilt $y^* \in \partial_C G(F(x))$. Andererseits ist $F$ stetig differenzierbar in $x$, und deshalb konvergiert der Differenzenquotient auf der rechten Seite stark gegen $F'(x)h$ (was nicht ganz offensichtlich ist). Da das Produkt von schwach-$*$ und stark konvergenten Folgen konvergiert, erhalten wir durch Grenzübergang $n \to \infty$ in (8.8) die Gleichung (8.7) (zunächst nur entlang der oben gewählten Teilfolge; wegen der Konvergenz der linken Seite und Eindeutigkeit des Grenzwerts aber dann auch für die gesamte Folge). Für $y^* \in \partial_C G(F(x))$ gilt daher nach Definition des Clarke-Subdifferentials

$$(G \circ F)^\circ(x; h) = \langle y^*, F'(x)h\rangle_Y \leq G^\circ(F(x); F'(x)h). \tag{8.9}$$

Ist nun $G$ regulär, so gilt $G^\circ(F(x); F'(x)h) = G'(F(x); F'(x)h)$ und damit wegen der Lipschitz-Stetigkeit von $G$ und der Fréchet-Differenzierbarkeit von $F$

$$\begin{aligned}
G^\circ&(F(x); F'(x)h)\\
&= \lim_{t\to 0^+} \frac{G(F(x) + tF'(x)h) - G(F(x))}{t}\\
&= \lim_{t\to 0^+} \frac{G(F(x) + tF'(x)h) - G(F(x+th)) + G(F(x+th)) - G(F(x))}{t}\\
&\leq \lim_{t\to 0^+} \left(\frac{G(F(x+th)) - G(F(x))}{t} + L\|h\|_X \frac{\|F(x) + F'(x)th - F(x+th)\|_Y}{\|th\|_X}\right)\\
&= (G \circ F)'(x; h) \leq (G \circ F)^\circ(x; h).
\end{aligned}$$

(Da sowohl die Summe als auch der zweite Summand konvergieren, muss auch der erste Summand konvergieren.) Zusammen mit (8.9) folgt daraus $(G \circ F)'(x; h) = (G \circ F)^\circ(x; h)$ (d. h. $G \circ F$ ist regulär) sowie $(G \circ F)^\circ(x; h) = G^\circ(F(x); F'(x)h)$.

Aus Lemma 8.8 folgt nun wieder für alle $h \in X$, dass gilt

$$\sup_{x^* \in F'(x)^* \partial_C G(F(x))} \langle x^*, h\rangle_X = \sup_{y^* \in \partial_C G(F(x))} \langle y^*, F'(x)h\rangle_Y = G^\circ(F(x); F'(x)h)$$

und damit nach Lemma 8.9

$$F'(x)^* \partial_C G(F(x)) = \{x^* \in X^* : \langle x^*, h\rangle_X \leq G^\circ(F(x); F'(x)h) \text{ für alle } h \in X\},$$

woraus die restlichen Behauptungen folgen. □

Wieder ist die Inklusion umgekehrt zur konvexen Kettenregel. Ist $G$ nicht regulär aber $F'(x)$ surjektiv, so kann man auf ähnliche Weise zeigen, dass die Kettenregel mit Gleichheit (aber nicht die Regularität von $G \circ F$) gilt; siehe [Clarke 2013, Theorem 10.19].

Ist $X$ endlich-dimensional, so ist eine explizitere Charakterisierung des Clarke-Subdifferentials möglich. Die Grundlage ist der *Satz von Rademacher*, der nur in $\mathbb{R}^N$ gilt; siehe z. B. [DiBenedetto 2002, Theorem 23.2] oder [Heinonen 2005, Theorem 3.1].

Satz 8.14 (Rademacher). *Sei $U \subset \mathbb{R}^N$ offen und $F : U \to \mathbb{R}$ Lipschitz-stetig. Dann ist $F$ fast überall Fréchet-differenzierbar.*

Dies erlaubt, den Limes superior in der Definition des Clarke-Subdifferentials (das wir nun als Teilmenge des $\mathbb{R}^N$ auffassen, d. h. den Dualraum von $\mathbb{R}^N$ mit $\mathbb{R}^N$ identifizieren) durch einen Grenzwert zu ersetzen.

Satz 8.15. *Sei $F : \mathbb{R}^N \to \mathbb{R}$ lokal Lipschitz-stetig in $x \in \mathbb{R}^N$. Dann ist $F$ Fréchet-differenzierbar auf $\mathbb{R}^N \setminus E_F$ für eine Lebesgue-Nullmenge $E_F \subset \mathbb{R}^N$ und es gilt*

$$\partial_C F(x) = \operatorname{co}\left\{\lim_{n\to\infty} \nabla F(x_n) : x_n \to x,\ x_n \notin E_F\right\}, \tag{8.10}$$

*wobei* $\operatorname{co} A$ *die konvexe Hülle der Menge $A \subset \mathbb{R}^N$ bezeichnet.*

*Beweis.* Zunächst folgt aus dem Satz von Rademacher, dass so ein $E_F$ existiert und – eingeschränkt auf die Umgebung, in der $F$ Lipschitz-stetig ist – eine Nullmenge ist. Also existieren überhaupt Folgen $\{x_n\}_{n\in\mathbb{N}} \in \mathbb{R}^N \setminus E_F$ mit $x_n \to x$. Weiter folgt aus der lokalen Lipschitz-Stetigkeit von $F$ für $x_n$ hinreichend nahe an $x$ und $h \in \mathbb{R}^N$ beliebig

$$|\left(\nabla F(x_n), h\right)| = \left|\lim_{t\to 0^+} \frac{F(x_n + th) - F(x_n)}{t}\right| \leq L\|h\|$$

und damit $\|\nabla F(x_n)\| \leq L$. Also ist $\{\nabla F(x_n)\}_{n\in\mathbb{N}}$ beschränkt und enthält daher eine konvergente Teilfolge. Damit ist die Menge auf der rechten Seite von (8.10) nichtleer.

Sei nun $x^* := \lim_{n\to\infty} \nabla F(x_n)$ für eine beliebige zulässige Folge $\{x_n\}_{n\in\mathbb{N}} \subset \mathbb{R}^N \setminus E_F$. Da $F$ nach Konstruktion in jedem $x_n \notin E_F$ differenzierbar ist, folgt

$$(\nabla F(x_n), h) = F'(x_n; h) \leq F^\circ(x_n; h) \quad \text{für alle } h \in X$$

und damit $\nabla F(x_n) \in \partial_C F(x_n)$. Nach Lemma 8.3 ist daher auch $x^* \in \partial_C F(x)$. Wegen der Konvexität von $\partial_C F(x)$ ist daher auch jede Konvexkombination von solchen $x^*$ in $\partial_C F(x)$ enthalten. Damit folgt die Inklusion “$\supset$” der beiden Mengen.

Für die andere Inklusion zeigen wir zunächst, dass für alle $h \in \mathbb{R}^N$ und $\varepsilon > 0$ gilt

$$F^\circ(x; h) - \varepsilon \leq \limsup_{E_F \not\ni y \to x} (\nabla F(y), h) =: M(h). \tag{8.11}$$

Nach Definition von $M(h)$ und des Limes superior existiert nun für jedes $\varepsilon > 0$ ein $\delta > 0$, so dass gilt

$$(\nabla F(y), h) \leq M(h) + \varepsilon \qquad \text{für alle } y \in O_\delta(x) \setminus E_F.$$

Dabei kann $\delta$ so klein gewählt werden, dass $F$ Lipschitz-stetig auf $O_\delta(x)$ ist. Insbesondere ist $E_F \cap O_\delta(x)$ eine Nullmenge. Für fast alle $y \in O_{\delta/2}(x)$ ist also $F$ in $y + th$ für fast alle

$t \in (0, \frac{\delta}{2\|h\|})$ differenzierbar (dies folgt aus dem Satz von Fubini). Für diese $y$ und $t$ gilt nach dem klassischen Mittelwertsatz

$$F(y + th) - F(y) = \int_0^t (\nabla F(y + sh), h)\; ds \leq t(M(h) + \varepsilon), \tag{8.12}$$

denn $y + sh \in O_\delta(x)$ für alle $s \in (0, t)$ nach Wahl von $t$. Da $F$ stetig ist, gilt die gesamte Ungleichung in (8.12) sogar für *alle* $y \in O_{\delta/2}(x)$ und *alle* $t \in (0, \frac{\delta}{2\|h\|})$. Division durch $t$ und Übergang zum Limes superior über alle $y \to x$ und $t \to 0^+$ ergibt nun die gewünschte Ungleichung (8.11). Da $\varepsilon > 0$ beliebig war, folgt $F^\circ(x; h) \leq M(h)$ für alle $h \in \mathbb{R}^N$.

Nun kann man wie in Lemma 8.1 zeigen, dass $h \mapsto M(h)$ positiv homogen, subadditiv und unterhalbstetig ist. Wir sind daher fertig, wenn wir zeigen können, dass wir die Menge auf der rechten Seite von (8.10) – in Folge als $\operatorname{co} A$ bezeichnet – schreiben können als

$$\operatorname{co} A = \left\{x^* \in \mathbb{R}^N : (x^*, h) \leq M(h) \quad \text{für alle } h \in \mathbb{R}^N\right\}.$$

Dafür verwenden wir wieder Folgerung 8.10 (denn beide Mengen sind nichtleer, konvex, und abgeschlossen). Zunächst gilt nach Definition der konvexen Hülle für alle $h \in \mathbb{R}^N$

$$\sup_{x^* \in \operatorname{co} A} (x^*, h) = \sup_{\substack{x_i^* \in A \\ \sum_i t_i = 1, t_i \geq 0}} \sum_i t_i \left(x_i^*, h\right) = \sup_{\sum_i t_i = 1, t_i \geq 0} \sum_i t_i \sup_{x_i^* \in A} \left(x_i^*, h\right) = \sup_{x^* \in A} (x^*, h),$$

denn die Summe ist maximal genau dann, wenn jeder Summand maximal ist. Nun gilt

$$M(h) = \limsup_{E_F \not\ni y \to x} (\nabla F(y), h) = \sup_{E_F \not\ni x_n \to x} (\lim\nolimits_{n\to\infty} \nabla F(x_n), h) = \sup_{x^* \in A} (x^*, h),$$

woraus mit Lemma 8.8 die Aussage folgt. □

Schließlich gilt, dass man ähnlich wie in Lemma 4.6 das Clarke-Subdifferential von Integralfunktionalen mit Lipschitz-stetigen Integranden punktweise darstellen kann; siehe z. B. [Clarke 1990, Theorem 2.7.3, 2.7.5].

**Satz 8.16.** *Sei $f : \mathbb{R} \to \mathbb{R}$ Lipschitz-stetig und $F : L^p(\Omega) \to \overline{\mathbb{R}}$ mit $1 \leq p < \infty$ wie in Lemma 3.4. Dann ist für alle $u \in L^p(\Omega)$ mit $q = \frac{p}{p-1}$ ($q = \infty$ für $p = 1$)*

$$\partial_C F(u) \subset \{u^* \in L^q(\Omega) : u^*(x) \in \partial_C f(u(x)) \quad \text{für fast alle } x \in \Omega\}.$$

*Ist $f$ regulär in $u(x)$ für fast alle $x \in \Omega$, so ist $F$ regulär in $u$ und es gilt Gleichheit.*

*Beweis.* Zunächst folgt aus der Monotonie des Lebesgue-Integrals und der Lipschitz-Stetigkeit von $f$, dass für alle $u, v \in L^p(\Omega)$ gilt

$$|F(u) - F(v)| \leq \int_\Omega |f(u(x)) - f(v(x))|\, dx \leq L \int_\Omega |u(x) - v(x)|\, dx \leq LC_p \|u - v\|_{L^p},$$

wobei $L$ die Lipschitz-Konstante von $f$ und die Konstante aus der stetigen Einbettung $L^p(\Omega) \hookrightarrow L^1(\Omega)$ für $1 \le p \le \infty$ ist. Also ist $F : L^p(\Omega) \to \mathbb{R}$ Lipschitz-stetig und daher insbesondere endlich-wertig.

Seien nun $\xi \in \partial_C F(u)$ fest und $h \in L^p(\Omega)$ beliebig. Nach Definition haben wir dann

$$
\begin{aligned}
\text{(8.13)}\qquad \langle \xi, h\rangle_{L^p} \le F^\circ(u;h) &= \limsup_{\substack{v\to u\\ t\to 0}} \frac{F(v+th)-F(v)}{t}\\
&\le \int_\Omega \limsup_{\substack{v\to u\\ t\to 0}} \frac{f(v(x)+th(x))-f(v(x))}{t}\,dx\\
&\le \int_\Omega \limsup_{\substack{v_x\to u(x)\\ t_x\to 0}} \frac{f(v_x+t_xh(x))-f(v_x)}{t_x}\,dx\\
&= \int_\Omega f^\circ(u(x);h(x))\,dx,
\end{aligned}
$$

wobei wir in der ersten Ungleichung den limes superior und das Integral mit Hilfe des umgekehrten Fatou-Lemmas vertauschen durften (denn der Integrand ist nach oben beschränkt durch die integrierbare Funktion $h \mapsto L|h|$ wegen Lemma 8.1 (i)); die zweite Ungleichung entsteht durch punktweises Abschätzen der Folgen, die den limes superior in $L^p(\Omega)$ realisiert, durch den punktweisen limes superior über alle zulässigen Folgen.

Um (8.13) punktweise zu interpretieren, definieren wir für $x \in \Omega$

$$
g_x : \mathbb{R} \to \mathbb{R}, \qquad g_x(t) := f^\circ(u(x);t).
$$

Aus Lemma 8.1 (ii)–(iii) folgt dann, dass $g_x$ konvex ist; Lemma 8.1 (i) impliziert weiterhin, dass die Funktion $x \mapsto g_x(h(x))$ messbar ist für alle $h \in L^p(\Omega)$. Wegen $g_x(0) = 0$ folgt aus (8.13), dass

$$
\langle \xi, h-0\rangle_{L^p} \le \int_\Omega g_x(h(x))\,dx - \int_\Omega g_x(0)\,dx,
$$

d. h. es gilt $\xi \in \partial G(0)$ für den Superpositionsoperator $G(h) := \int_\Omega g_x(h(x))\,dx$. Genau wie im Beweis von Lemma 4.6 zeigt man nun, dass dann $\xi = u^* \in L^q(\Omega)$ mit $u^*(x) \in \partial g_x(0)$ für fast alle $x \in \Omega$ gilt, d. h.

$$
u^*(x)h(x) = u^*(x)(h(x)-0) \le g_x(h(x)) - g_x(0) = f^\circ(u(x);h(x))
$$

für fast alle $x \in \Omega$. Da $h \in L^p(\Omega)$ beliebig war, folgt daraus $u^*(x) \in \partial_C f(u(x))$ fast überall.

Es bleiben noch die restlichen Aussagen für reguläres $f$ zu zeigen. In diesem Fall folgt aus (8.13), dass für alle $h \in L^p(\Omega)$ gilt

$$
\begin{aligned}
\text{(8.14)}\qquad F^\circ(u;h) &\le \int_\Omega f^\circ(u(x);h(x))\,dx = \int_\Omega f'(u(x);h(x))\,dx\\
&\le \lim_{t\to 0} \frac{F(u+th)-F(u)}{t} = F'(u;h) \le F^\circ(u;h),
\end{aligned}
$$

wobei die erste Ungleichung aus (8.13) folgt und die zweite Ungleichung aus dem Lemma von Fatou, diesmal mit Hilfe der integrierbaren unteren Schranke $-L|h(x)|$. Damit erhalten wir, dass $F'(u;h) = F^\circ(u;h)$ and daher $F$ regulär ist. Weiterhin folgt für alle $u^* \in L^q(\Omega)$ mit $u^*(x) \in \partial_C f(u(x))$ fast überall und alle $h \in L^p(\Omega)$, dass

$$\langle u^*, h\rangle_{L^p} = \int_\Omega u^*(x)h(x)\,dx \leq \int_\Omega f^\circ(u(x); h(x))\,dx \leq F^\circ(u,h),$$

wobei wir in der letzten Ungleichung (8.14) verwendet haben. Weil $h \in L^p(\Omega)$ beliebig war, folgt $u^* \in \partial_C F(u)$. □

# 9 SEMIGLATTE NEWTON-VERFAHREN

Die Proximalpunkt- und Splitting-Verfahren aus Kapitel 7 stellen Verallgemeinerungen von Gradientenverfahren dar und weisen im Allgemeinen auch nur deren lineare Konvergenz auf. In diesem Kapitel suchen wir daher eine Verallgemeinerung des Newton-Verfahrens, für das (lokal) superlineare Konvergenz gilt.

Als Motivation betrachten wir zuerst die allgemeine Form eines Newton-artigen Verfahrens. Sei $F : X \to Y$ für zwei Banachräume $X$ und $Y$, und gesucht sei ein $\bar{x} \in X$ mit $F(\bar{x}) = 0$. Ein Newton-artiges Verfahren hat dann die folgende Form:

1. wähle $M_k := M(x^k) \in L(X, Y)$ invertierbar;
2. löse $M_k s^k = -F(x^k)$;
3. setze $x^{k+1} = x^k + s^k$.

Wir können uns nun fragen, unter welchen Bedingungen diese Iteration konvergiert, und insbesondere, wann die Konvergenz superlinear ist, d. h. wann gilt

$$\lim_{k\to\infty} \frac{\|x^{k+1} - \bar{x}\|_X}{\|x^k - \bar{x}\|_X} = 0. \tag{9.1}$$

Dafür setzen wir $e^k := x^k - \bar{x}$ und verwenden den Newton-Schritt sowie $F(\bar{x}) = 0$ um zu schreiben

$$\begin{aligned}
\|x^{k+1} - \bar{x}\|_X &= \|x^k - M(x^k)^{-1}F(x^k) - \bar{x}\|_X \\
&= \|M(x^k)^{-1}\left[F(x^k) - F(\bar{x}) - M(x^k)(x^k - \bar{x})\right]\|_X \\
&= \|M(\bar{x} + e^k)^{-1}\left[F(\bar{x} + e^k) - F(\bar{x}) - M(\bar{x} + e^k)e^k\right]\|_X \\
&\leq \|M(\bar{x} + e^k)^{-1}\|_{L(Y,X)}\|F(\bar{x} + e^k) - F(\bar{x}) - M(\bar{x} + e^k)e^k\|_Y.
\end{aligned}$$

Also gilt (9.1), falls erfüllt sind

(i) eine *Regularitätsbedingung*: es existiert ein $C > 0$ mit

$$\|M(\bar{x} + e^k)^{-1}\|_{L(Y,X)} \leq C \qquad \text{für alle } k \in \mathbb{N},$$

(ii) eine *Approximationsbedingung*:

$$\lim_{k\to\infty} \frac{\|F(\bar x + e^k) - F(\bar x) - M(\bar x + e^k)e^k\|_Y}{\|e^k\|_X} = 0.$$

Dies motiviert die folgende Definition: Wir nennen $F : X \to Y$ *Newton-differenzierbar* in $x \in X$ mit *Newton-Ableitung* $D_N F(x) \in L(X, Y)$, falls gilt

$$\lim_{\|h\|_X\to 0} \frac{\|F(x+h) - F(x) - D_N F(x+h)h\|_Y}{\|h\|_X} = 0.$$

Beachte die Unterschiede zur Fréchet-Ableitung: Zum einen wird die Newton-Ableitung in $x + h$ anstelle von $x$ ausgewertet. Wichtiger ist aber, dass kein konkreter Zusammenhang von $D_N F$ mit $F$ gefordert wurde (im Gegensatz zur Fréchet-Ableitung, für die nur die Gâteaux-Ableitung als Kandidat in Frage kam); eine Funktion ist also nur Newton-differenzierbar (oder nicht) in Bezug auf eine konkrete Wahl von $D_N F$. Insbesondere sind Newton-Ableitungen nicht eindeutig![1]

Ist $F$ Newton-differenzierbar mit Newton-Ableitung $D_N F$, so erhält man das *semiglatte Newton-Verfahren*

$$x^{k+1} = x^k - D_N F(x^k)^{-1} F(x^k).$$

Direkt aus der Konstruktion folgt dann die lokal superlineare Konvergenz.

Satz 9.1. *Sei $F : X \to Y$ und $\bar x \in X$ mit $F(\bar x) = 0$. Ist $F$ Newton-differenzierbar in $\bar x$ mit Newton-Ableitung $D_N F(\bar x)$, und existieren $\delta > 0$ und $C > 0$ mit $\|D_N F(x)^{-1}\|_{L(Y,X)} \le C$ für alle $x \in O_\delta(\bar x)$, so konvergiert für alle $x^0$ hinreichend nahe an $\bar x$ das semiglatte Newton-Verfahren superlinear gegen $\bar x$.*

*Beweis.* Der Beweis ist völlig analog zum Konvergenzbeweis für das klassische Newton-Verfahren. Für $x^0 \in O_\delta(\bar x)$ gilt wie schon gezeigt

$$\|e^1\|_X \le C\|F(\bar x + e^0) - F(\bar x) - D_N F(\bar x + e^0)e^0\|_Y. \tag{9.2}$$

Sei nun $\varepsilon \in (0, 1)$ beliebig. Aufgrund der Newton-Differenzierbarkeit existiert dann ein $\rho > 0$ mit

$$\|F(\bar x + h) - F(\bar x) - D_N F(\bar x + h)h\|_Y \le \frac{\varepsilon}{C}\|h\|_X \qquad \text{für alle } \|h\|_X \le \rho.$$

[1] Wir folgen hier [Chen, Nashed & Qi 2000; Ito & Kunisch 2008; Schiela 2008] und betrachten nur einwertige Newton-Ableitungen (in ersterer Arbeit als *slanting function* eingeführt). Alternativ kann man für jedes $x \in X$ eine Menge $\partial_N F(x)$ festlegen, aus der $M(x)$ zu wählen ist. Gelten die Approximations- und eine Beschränktheitsbedingung *gleichmäßig* für alle $M \in \partial_N F(x)$, so nennt man $F$ *semiglatt*, was die Namensgebung in diesem Kapitel erklärt. Dieser Zugang wird z. B. in [Mifflin 1977; Kummer 1988; Ulbrich 2011] verfolgt.

Wählen wir also $x^0$ so, dass $\|\bar{x} - x^0\|_X \leq \min\{\delta, \rho\}$ gilt, so folgt aus (9.2) die Abschätzung $\|\bar{x} - x^1\|_X \leq \varepsilon\|\bar{x} - x^0\|_X$ und damit durch Induktion $\|\bar{x} - x^k\|_X \leq \varepsilon^k\|\bar{x} - x^0\|_X \to 0$. Da $\varepsilon$ beliebig war, kann man für jeden Schritt ein neues $\varepsilon_k \to 0$ wählen, und daher ist die Konvergenz superlinear. □

Der Rest des Kapitels ist nun der Konstruktion von Newton-Ableitungen gewidmet (wobei nicht verschwiegen werden soll, dass in der Praxis der Nachweis der Regularitätsbedingung die deutlich aufwändigere Aufgabe ist). Wir beginnen mit dem offensichtlichen Zusammenhang mit der Fréchet-Differenzierbarkeit.

Satz 9.2. *Ist $F : X \to Y$ stetig differenzierbar in $x \in X$, so ist $F$ Newton-differenzierbar in $x$ mit Newton-Ableitung $D_N F(x) = F'(x)$.*

*Beweis.* Für beliebige $h \in X$ gilt

$$\begin{aligned}\|F(x+h) - F(x) - F'(x+h)h\|_Y \leq{}& \|F(x+h) - F(x) - F'(x)h\|_Y \\ &+ \|F'(x) - F'(x+h)\|_{L(X,Y)}\|h\|_X,\end{aligned}$$

wobei beide Summanden $o(\|h\|_X)$ sind: der erste nach Definition der Fréchet-Ableitung und der zweite wegen der Stetigkeit von $F'$. □

Rechenregeln beweist man nun analog zu denen für Fréchet-Ableitungen. Die Summenregel ist offensichtlich; wir beweisen beispielhaft die Kettenregel.

Satz 9.3. *Seien $X, Y, Z$ Banachräume und $F : X \to Y$ Newton-differenzierbar in $x \in X$ mit Newton-Ableitung $D_N F(x)$ und $G : Y \to Z$ Newton-differenzierbar in $y := F(x) \in Y$ mit Newton-Ableitung $D_N G(y)$. Sind $D_N F$ und $D_N G$ gleichmäßig beschränkt in einer Umgebung von $x$ bzw. $y$, so ist $G \circ F$ Newton-differenzierbar in $x$ mit Newton-Ableitung*

$$D_N(G \circ F)(x) = D_N G(F(x)) \circ D_N F(x).$$

*Beweis.* Wir gehen wie im Beweis von Satz 2.5 vor. Für $h \in X$ und $g := F(x+h) - F(x)$ ist

$$(G \circ F)(x+h) - (G \circ F)(x) = G(y+g) - G(y).$$

Aus der Newton-Differenzierbarkeit von $G$ folgt nun

$$\|(G \circ F)(x+h) - (G \circ F)(x) - D_N G(y+g)g\|_Z = r_1(\|g\|_Y)$$

mit $r_1(t)/t \to 0$ für $t \to 0$. Aus der Newton-Differenzierbarkeit von $F$ folgt weiter

$$\|g - D_N F(x+h)h\|_Y = r_2(\|h\|_X)$$

mit $r_2(t)/t \to 0$ für $t \to 0$. Insbesondere ist

$$\|g\|_Y \leq \|D_N F(x+h)\|_{L(X,Y)}\|h\|_Y + r_2(\|h\|_X).$$

Wegen der gleichmäßigen Beschränktheit von $D_N F$ gilt also $\|g\|_Y \to 0$ für $\|h\|_X \to 0$. Daher gilt

$$\begin{aligned}
&\|(G\circ F)(x+h) - (G\circ F)(x) - D_N G(F(x+h))D_N F(x+h)h\|_Z \\
&\qquad\qquad \leq \|G(y+g) - G(g) - D_N G(y+g)g\|_Z \\
&\qquad\qquad\quad + \|D_N G(y+g)\,[g - D_N F(x+h)h]\,\|_Z \\
&\qquad\qquad \leq r_1(\|g\|_Y) + \|D_N G(y+g)\|_{L(Y,Z)} r_2(\|h\|_X),
\end{aligned}$$

woraus wegen der gleichmäßigen Beschränktheit von $D_N G$ die gewünschte Aussage folgt. □

Schließlich folgt direkt aus der Definition der Produktnorm und der Newton-Differenzierbarkeit, dass für vektorwertige Funktionen die Newton-Ableitung komponentenweise berechnet werden kann.

Satz 9.4. *Sei $m \in \mathbb{N}$ beliebig und seien $F_i : X \to Y_i$ Newton-differenzierbar mit Newton-Ableitung $D_N F_i$ für $1 \leq i \leq m$. Dann ist*

$$F : X \to (Y_1 \times \cdots \times Y_m), \qquad x \mapsto (F_1(x), \dots, F_m(x))^T$$

*Newton-differenzierbar mit Newton-Ableitung*

$$D_N F(x) = (D_N F_1(x), \dots, D_N F_m(x))^T.$$

Da die Definition keine konstruktive Vorschrift für die Newton-Ableitung enthält, bleibt die Frage, wie man dafür Kandidaten erhält, für die die Approximationsbedingung nachgeprüft werden kann. Für zwei Klassen von Funktionen ist solch eine explizite Konstruktion bekannt.

### LOKAL LIPSCHITZ-STETIGE FUNKTIONEN AUF $\mathbb{R}^N$

Ist $F : \mathbb{R}^N \to \mathbb{R}$ lokal Lipschitz-stetig, so liefert das Clarke-Subdifferential die gesuchten Kandidaten, die wegen Satz 8.15 eine explizite Darstellung haben. Unter zusätzlichen Annahmen sind diese Kandidaten auch tatsächlich Newton-Ableitungen.[2]

[2] Dies war die ursprüngliche Herleitung der semiglatten Newton-Verfahren.

Eine Funktion $F : \mathbb{R}^N \to \mathbb{R}$ heißt *stückweise (stetig) differenzierbar* oder *PC*$^1$*-Funktion*, falls $F$ stetig ist und für alle $x \in \mathbb{R}^N$ eine Umgebung $U_x \subset \mathbb{R}^N$ um $x$ sowie eine endliche Menge $\{F_i : U_x \to \mathbb{R}\}_{i\in I_x}$ von stetig differenzierbaren Funktionen existiert mit

$$F(y) \in \{F_i(y)\}_{i\in I_x} \qquad \text{für alle } y \in U_x.$$

Wir nennen $F$ dann eine *stetige Auswahl* der $F_i$ in $U_x$. Die Menge

$$I_a(x) := \{i \in I : F(x) = F_i(x)\}$$

heißt *aktive Index-Menge*. Aufgrund der Stetigkeit der $F_i$ gilt $F(y) \neq F_j(y)$ für alle $j \notin I_a(x)$ und $y$ hinreichend nahe an $x$. Indizes, die nur auf einer Nullmenge aktiv sind, müssen wir in Folge nicht berücksichtigen. Wir definieren daher die *essentiell aktive Index-Menge*

$$I_e(x) := \{i \in I_x : x \in \mathrm{cl}\left(\{y \in U_x : F(y) = F_i(y)\}^o\right)\} \subset I_a(x).$$

Zum Beispiel ist für $f(t) = \max\{0, t, \frac{1}{2}t\}$ die Funktion $f_3(t) = \frac{1}{2}t$ nur in $t = 0$ aktiv und daher nicht essentiell aktiv.

PC$^1$-Funktionen sind stets lokal Lipschitz-stetig; siehe [Scholtes 2012, Corollary 4.1.1].

Satz 9.5. *Sei $F : \mathbb{R}^N \to \mathbb{R}$ stückweise differenzierbar. Dann ist $F$ lokal Lipschitz-stetig mit Konstante $L = \max_{i\in I_a(x)} L_i$ für $L_i = \sup_{y\in U_x} |\nabla F_i(y)|$ bei $x$.*

Aus Satz 8.15 erhalten wir daher die folgende Darstellung des Clarke-Subdifferentials.

Satz 9.6. *Sei $F : \mathbb{R}^N \to \mathbb{R}$ stückweise differenzierbar. Dann gilt für alle $x \in \mathbb{R}^N$*

$$\partial_C F(x) = \mathrm{co}\left\{\nabla F_i(x) : i \in I_e(x)\right\}.$$

*Beweis.* Sei $x \in \mathbb{R}^N$ beliebig. Nach Satz 8.15 genügt zu zeigen, dass gilt

$$\left\{\lim_{n\to\infty} \nabla F(x_n) : x_n \to x,\ x_n \notin E_F\right\} = \left\{\nabla F_i(x) : i \in I_e(x)\right\}.$$

Sei dafür $\{x_n\}_{n\in\mathbb{N}}$ eine Folge in $\mathbb{R}^N$ mit $x_n \to x$, $F$ differenzierbar in $x_n$ für alle $n \in \mathbb{N}$, und $\nabla F(x_n) \to x^* \in \mathbb{R}^N$. Da $F$ differenzierbar ist in $x_n$, muss $F(y) = F_{i_n}(y)$ für ein $i_n$ und alle $y$ hinreichend nahe an $x_n$ gelten, woraus $\nabla F(x_n) = \nabla F_{i_n}(x_n)$ folgt. Für $n \in \mathbb{N}$ groß genug können wir darüberhinaus $i_n \in I_e(x)$ (eventuell durch Hinzunahme einer weiteren Nullmenge zu $E_F$) annehmen. Betrachten wir Teilfolgen $\{x_{n_k}\}_{k\in\mathbb{N}}$ mit konstanten Indizes $i_{n_k} = i \in I_e(x)$ (solche existieren, da $I_e(x)$ endlich ist), so erhalten wir wegen der Stetigkeit von $\nabla F_i$

$$x^* = \lim_{k\to\infty} \nabla F(x_{n_k}) = \lim_{k\to\infty} \nabla F_i(x_{n_k}) \in \left\{\nabla F_i(x) : i \in I_e(x)\right\}.$$

Umgekehrt existiert für $\nabla F_i(x)$ mit $i \in I_e(x)$ nach Definition der essentiell aktiven Indizes eine Folge $\{x_n\}_{n\in\mathbb{N}}$ mit $x_n \to x$ und $F = F_i$ in einer hinreichend kleinen Umgebung aller $x_n$

für $n$ groß genug. Wegen der stetigen Differenzierbarkeit der $F_i$ ist dann $\nabla F(x_n) = \nabla F_i(x_n)$ für alle $n \in \mathbb{N}$ groß genug und damit

$$\nabla F_i(x) = \lim_{n\to\infty} \nabla F_i(x_n) = \lim_{n\to\infty} \nabla F(x_n). \qquad \square$$

Daraus folgt die Newton-Differenzierbarkeit von $\mathrm{PC}^1$-Funktionen.

Satz 9.7. *Sei $F : \mathbb{R}^N \to \mathbb{R}$ stückweise differenzierbar. Dann ist $F$ Newton-differenzierbar für alle $x \in \mathbb{R}^N$, und jedes $D_N F(x) \in \partial_C F(x)$ ist eine Newton-Ableitung.*

*Beweis.* Sei $x \in \mathbb{R}^N$ beliebig und $h \in X$ mit $x+h \in U_x$. Nach Satz 9.6 hat jedes $D_N F(x+h) \in \partial_C F(x+h)$ die Form

$$D_N F(x+h) = \sum_{i\in I_e(x+h)} t_i \nabla F_i(x+h) \qquad \text{für} \sum_{i\in I_e(x+h)} t_i = 1, t_i \geq 0.$$

Da $F$ stetig ist, gilt für alle $h \in \mathbb{R}^N$ mit $\|h\|$ hinreichend klein $I_e(x+h) \subset I_a(x+h) \subset I_a(x)$. Also ist sowohl $F(x+h) = F_i(x+h)$ als auch $F(x) = F_i(x)$ für alle $i \in I_e(x+h)$. Damit folgt aus Satz 9.2

$$|F(x+h) - F(x) - D_N F(x+h)h| \leq \sum_{i\in I_e(x+h)} t_i |F_i(x+h) - F_i(x) - \nabla F_i(x+h)h| = o(\|h\|),$$

da alle $F_i$ nach Annahme stetig differenzierbar sind. $\square$

Als Anwendung haben wir natürlich die Proximalpunkt-Formulierung von Optimalitätsbedingungen für konvexe Probleme vor Augen.

Beispiel 9.8. Wir betrachten die Minimierung von $F + G$ für $F : \mathbb{R}^N \to \mathbb{R}$ zweimal stetig differenzierbar und $G = \|\cdot\|_1$. Analog zur Herleitung des expliziten Splitting-Verfahrens (7.6) schreiben wir die notwendige Optimalitätsbedingung $0 \in \partial_C(F+G)(\bar{x})$ mit Hilfe der Regularität von $F$ und $G$ äquivalent um als

$$\bar{x} - \mathrm{prox}_{\gamma G}(\bar{x} - \gamma \nabla F(\bar{x})) = 0$$

für $\gamma > 0$ beliebig. Nach Beispiel 6.14 (ii) ist die Proximalpunkt-Abbildung für $G$ komponentenweise gegeben durch

$$[\mathrm{prox}_{\gamma G}(x)]_i = \begin{cases} x_i - \gamma & \text{falls } x_i > \gamma, \\ 0 & \text{falls } x \in [-\gamma, \gamma], \\ x_i + \gamma & \text{falls } x_i < -\gamma, \end{cases}$$

was offensichtlich eine stückweise differenzierbare Funktion ist. Nach Satz 9.6 ist daher (ebenfalls komponentenweise)

$$\partial_C[\operatorname{prox}_{\gamma G}(x)]_i = \begin{cases} \{1\} & \text{falls } |x_i| > \gamma, \\ \{0\} & \text{falls } |x_i| < \gamma, \\ [0,1] & \text{falls } |x_i| = \gamma. \end{cases}$$

Nach Satz 9.7 zusammen mit Satz 9.4 ist daher eine mögliche Newton-Ableitung

$$[D_N\operatorname{prox}_{\gamma G}(x)h]_i = [\mathbb{1}_{\{|x|\geq\gamma\}}h]_i := \begin{cases} h_i & \text{falls } |x_i| \geq \gamma, \\ 0 & \text{falls } |x_i| < \gamma. \end{cases}$$

(Welchem Fall wir die Gleichheit zuschlagen, ist dabei willkürlich.) Nun sind $D_N\operatorname{prox}_{\gamma G}(x)$ und $D_N(\nabla F)(x) = \nabla^2 F(x)$ (wegen der stetigen Differenzierbarkeit) lokal gleichmäßig beschränkt, und mit Hilfe der Kettenregel (Satz 9.3) erhalten wir nach etwas Umformen den semiglatten Newton-Schritt

$$\left(\mathbb{1}_{\mathcal{A}_k} + \gamma \mathbb{1}_{\mathcal{I}_k}\nabla^2 F(x^k)\right) s^k = -x^k + \operatorname{prox}_{\gamma G}(x^k - \gamma\nabla F(x^k)),$$

wobei wir die *aktive* bzw. *inaktive Menge* definiert haben als

$$\mathcal{A}_k := \left\{i \in \{1,\dots,N\} : |x_i^k - \gamma[\nabla F(x^k)]_i| < \gamma\right\}, \qquad \mathcal{I}_k := \{1,\dots,N\} \setminus \mathcal{A}_k.$$

Partitionieren wir auch $s^k$ sowie die rechte Seite in aktive und inaktive Komponenten, setzen die Fallunterscheidung für $\operatorname{prox}_{\gamma G}$ ein, und sortieren das lineare Gleichungssystem so um, dass aktive und inaktive Komponenten zu Blöcken zusammengefasst sind, so erhält der Newton-Schritt die Form einer *aktiven-Mengen-Strategie*; siehe auch [Ito & Kunisch 2008, Kapitel 8.4].

### SUPERPOSITIONSOPERATOREN AUF $L^p(\Omega)$

In unendlichdimensionalen Funktionenräumen gilt der Satz von Rademacher nicht, so dass das Clarke-Subdifferential keinen algorithmisch nutzbaren Kandidaten für eine Newton-Ableitung mehr liefert. Eine Ausnahme bilden Superpositionsoperatoren, die durch Newton-differenzierbare skalare Funktionen definiert sind; für diese kann die Newton-Ableitung punktweise charakterisiert werden.

Wir betrachten wieder für ein offenes und beschränktes Gebiet $\Omega \subset \mathbb{R}^N$ eine Carathéodory-Funktion $f : \Omega \times \mathbb{R} \to \mathbb{R}$ (d. h. $f$ ist messbar in $x$ und stetig in $z$) sowie für $1 \leq p, q \leq \infty$ den zugehörigen Superpositionsoperator

$$F : L^p(\Omega) \to L^q(\Omega), \qquad [F(u)](x) = f(x, u(x)) \quad \text{für fast alle } x \in \Omega.$$

Wir möchten nun die Newton-Ableitung $D_N F$ von $F$ ebenfalls als Superpositionsoperator der Newton-Ableitung $D_N f(x,z)$ von $z \mapsto f(x,z)$ darstellen. Dabei ist die Forderung, $D_N f$ sei ebenfalls Carathéodory-Funktion, zu einschränkend, denn wir wollen auch unstetige Ableitungen zulassen. Für unsere Zwecke genügt eine schwächere Eigenschaft: Eine Funktion heißt *Baire–Carathéodory-Funktion*, wenn sie punktweiser Grenzwert von Carathéodory-Funktionen ist.

Unter bestimmten Wachstumsbedingungen[3] an $f$ und $D_N f$ können wir die Newton-Differenzierbarkeit von $f$ auf $F$ übertragen, wobei wir wieder eine zwei-Norm-Diskrepanz berücksichtigen müssen.

Satz 9.9. *Sei $f : \Omega \times \mathbb{R} \to \mathbb{R}$ eine Carathéodory-Funktion. Es gelte weiterhin*

*(i) $z \mapsto f(x,z)$ ist gleichmäßig Lipschitz-stetig für fast alle $x \in \Omega$ und $f(x,0)$ ist beschränkt;*

*(ii) $z \mapsto f(x,z)$ ist Newton-differenzierbar mit Newton-Ableitung $z \mapsto D_N f(x,z)$ für fast alle $x \in \Omega$;*

*(iii) $D_N f$ ist eine gleichmäßig beschränkte Baire–Carathéodory-Funktion.*

*Dann ist für alle $1 \le q < p < \infty$ der zugehörige Superpositionsoperator $F : L^p(\Omega) \to L^q(\Omega)$ Newton-differenzierbar mit Newton-Ableitung*

$$D_N F : L^p(\Omega) \to L(L^p(\Omega), L^q(\Omega)), \qquad [D_N F(u)h](x) = D_N f(x,u(x))h(x),$$

*für alle $h \in L^p(\Omega)$ und fast alle $x \in \Omega$.*

*Beweis.* Zunächst folgt aus der gleichmäßigen Lipschitz-Stetigkeit in $z$ zusammen mit der umgekehrten Dreiecksungleichung

$$|f(x,z)| \le |f(x,0)| + L|z| \le C + L|z|^{q/q} \qquad \text{für fast alle } x \in \Omega, z \in \mathbb{R},$$

und damit die Wachstumsbedingung (2.10) für alle $1 \le q \le \infty$. Wegen der stetigen Einbettung $L^p(\Omega) \hookrightarrow L^q(\Omega)$ für alle $1 \le q \le p \le \infty$ ist also $F : L^p(\Omega) \to L^q(\Omega)$ nach Satz 2.8 wohldefiniert und stetig.

Ist nun $u : \Omega \to \mathbb{R}$ messbar, so ist $x \mapsto D_N f(x,u(x))$ nach Annahme Grenzwert einer Folge von messbaren Funktionen und damit selber messbar. Aus der gleichmäßigen Beschränktheit folgt nun insbesondere die Wachstumsbedingung für $p' = p$ und $q' = p - q > 0$. Analog zum Beweis von Satz 2.9 erhält man daraus, dass der zugehörige Superpositionsoperator $D_N F : L^p(\Omega) \to L^s(\Omega)$ für $s := \frac{pq}{p-q}$ wohldefiniert und stetig ist, und deshalb für $u \in L^p(\Omega)$ durch $h \mapsto D_N F(u)h$ ein linearer und beschränkter Operator $D_N F(u) : L^p(\Omega) \to L^q(\Omega)$ definiert wird. (Wir unterscheiden diesmal nicht in der Notation zwischen linearem Operator und der Funktion, die durch punktweise Multiplikation diesen Operator definiert.)

[3] die deutlich abgeschwächt werden können; siehe [Schiela 2008, Proposition A.1]

Um zu zeigen, dass $D_N F(u)$ eine Newton-Ableitung von $F$ in $u \in L^p(\Omega)$ ist, betrachten wir das punktweise Residuum

$$r : \Omega \times \mathbb{R} \to \mathbb{R}, \qquad r(x,z) := \begin{cases} \frac{|f(x,z)-f(x,u(x))-D_N f(x,z)(z-u(x))|}{|z-u(x)|} & \text{falls } z \neq u(x), \\ 0 & \text{falls } z = u(x). \end{cases}$$

Da $f$ Carathéodory- und $D_N f$ Baire–Carathéodory-Funktion ist, ist für beliebiges $\tilde{u} : \Omega \to \mathbb{R}$ messbar auch $x \mapsto r(x, \tilde{u}(x)) =: [R(\tilde{u})](x)$ messbar (Summe, Produkt und Quotient von messbaren Funktionen ist messbar). Für $\tilde{u} \in L^p(\Omega)$ folgt aus der gleichmäßigen Lipschitz-Stetigkeit von $f$ und der gleichmäßigen Beschränktheit von $D_N f$ für fast alle $x \in \Omega$ mit $\tilde{u}(x) \neq u(x)$

$$|[R(\tilde{u})](x)| = \frac{|f(x,\tilde{u}(x)) - f(x,u(x)) - D_N f(x,\tilde{u}(x))(\tilde{u}(x) - u(x))|}{|\tilde{u}(x) - u(x)|} \leq L + C \tag{9.3}$$

und damit $R(\tilde{u}) \in L^\infty(\Omega)$. Also ist der Superpositionsoperator $R : L^p(\Omega) \to L^s(\Omega)$ wohldefiniert. Sei nun $\{u_n\}_{n\in\mathbb{N}} \subset L^p(\Omega)$ eine Folge mit $u_n \to u \in L^p(\Omega)$. Dann existiert eine Teilfolge, die wir wieder mit $\{u_n\}_{n\in\mathbb{N}}$ bezeichnen, mit $u_n(x) \to u(x)$ für fast alle $x \in \Omega$. Da $z \mapsto f(x,z)$ Newton-differenzierbar ist, gilt nach Definition $r(x, u_n(x)) \to 0$ punktweise fast überall. Zusammen mit der Beschränktheit (9.3) gilt daher nach dem Satz von Lebesgue auch $R(u_n) \to 0$ in $L^s(\Omega)$ (wegen Eindeutigkeit des Grenzwertes für die gesamte Folge).[4] Also folgt für beliebige $\tilde{u} \in L^p(\Omega)$ wegen $\frac{1}{p} + \frac{1}{s} = \frac{1}{q}$ aus der Hölderschen Ungleichung

$$\|F(\tilde{u}) - F(u) - D_N F(\tilde{u})(\tilde{u} - u)\|_{L^q} = \|R(\tilde{u})(\tilde{u} - u)\|_{L^q} \leq \|R(\tilde{u})\|_{L^s} \|\tilde{u} - u\|_{L^p}.$$

Setzen wir nun $\tilde{u} := u + h$ für $h \in L^p(\Omega)$ mit $\|h\|_{L^p} \to 0$, so ist dies wegen $\|R(u+h)\|_{L^s} \to 0$ genau die Definition der Newton-Differenzierbarkeit von $F$ in $u$. □

Für $p = q \in [1, \infty]$ ist die Aussage dagegen im Allgemeinen falsch, was durch Gegenbeispiele belegt werden kann. (Betrachte z. B. $f(z) = \max\{0, z\}$, $\Omega = (-1, 1)$, $u(x) = -|x|$, und $h_n(x) = 1/n$ für $|x| \leq 1/n$ und $0$ sonst.)

Aufgrund der zwei-Norm-Diskrepanz können wir das semiglatte Newton-Verfahren im Funktionenraum in der Regel nicht direkt auf Proximalpunkt-Formulierungen anwenden, weshalb wir auf die Moreau–Yosida-Regularisierung ausweichen müssen.

**Beispiel 9.10.** Wir betrachten analog zu Beispiel 9.8 die Minimierung von $F + G$ für $F : L^2(\Omega) \to \mathbb{R}$ zweimal stetig differenzierbar und $G = \|\cdot\|_{L^1}$. Die Proximalpunktformulierung der notwendigen Optimalitätsbedingung $0 \in \partial(F + G)(\bar{u})$,

$$\bar{u} - \operatorname{prox}_{\gamma G}(\bar{u} - \gamma \nabla F(\bar{u})) = 0,$$

[4] Dieser Schritt scheitert für $F : L^\infty(\Omega) \to L^\infty(\Omega)$, denn punktweise Konvergenz impliziert nicht gleichmäßige Konvergenz fast überall.

müssen wir nun als Gleichung in $L^2(\Omega)$ auffassen, aber $\operatorname{prox}_{\gamma G}$ ist *nicht* Newton-differenzierbar von $L^2(\Omega)$ nach $L^2(\Omega)$. Wir ersetzen daher in der äquivalenten Optimalitätsbedingung

$$\begin{cases} -\bar p = \nabla F(\bar u), \\ \quad \bar u \in \partial G^*(\bar p), \end{cases}$$

das Subdifferential von $G^*$ durch die Moreau–Yosida-Regularisierung $H_\gamma := (\partial G^*)_\gamma$, welche nach Folgerung 6.15 und Beispiel 6.19 punktweise gegeben ist durch $[H_\gamma(p)](x) = h_\gamma(p(x))$ für

$$h_\gamma : \mathbb{R} \to \mathbb{R}, \qquad t \mapsto \begin{cases} \frac{1}{\gamma}(t-1) & \text{falls } t > 1, \\ 0 & \text{falls } t \in [-1,1], \\ \frac{1}{\gamma}(t+1) & \text{falls } t < -1. \end{cases}$$

Diese Funktion ist offensichtlich stückweise stetig differenzierbar und damit Lipschitz-stetig. Nach Satz 9.6 ist daher

$$\partial_C h_\gamma(t) = \begin{cases} \left\{\frac{1}{\gamma}\right\} & \text{falls } |t| > 1, \\ \{0\} & \text{falls } |t| < 1, \\ \left[0, \frac{1}{\gamma}\right] & \text{falls } |t| = 1, \end{cases}$$

und nach Satz 9.7 zusammen mit Satz 9.4 ist eine mögliche Newton-Ableitung

$$D_N h_\gamma(t)h = \frac{1}{\gamma}\mathbb{1}_{\{|t|\geq 1\}} h = \begin{cases} \frac{1}{\gamma}h & \text{falls } |t| \geq 1, \\ 0 & \text{falls } |t| < 1. \end{cases}$$

Die Funktion $D_N h_\gamma$ ist nun gleichmäßig beschränkt und kann, wie man leicht zeigt, punktweise durch eine Folge von stetigen Funktionen approximiert werden. Also ist $H_\gamma : L^p(\Omega) \to L^2(\Omega)$ Newton-differenzierbar für alle $p > 2$, und eine Newton-Ableitung ist gegeben durch

$$[D_N H_\gamma(p)h](x) = \frac{1}{\gamma}\mathbb{1}_{\{|p|\geq 1\}}(x)h(x).$$

Angenommen, für $u \in L^2(\Omega)$ ist $p = -\nabla F(u) \in L^p(\Omega)$ für ein $p > 2$. Dann ist nach Satz 9.2 und der Kettenregel (Satz 9.3) die regularisierte Optimalitätsbedingung

$$u_\gamma - H_\gamma(-\nabla F(u_\gamma)) = 0$$

Newton-differenzierbar, und wir erhalten den semiglatten Newton-Schritt

$$\left(\mathrm{Id} + \frac{1}{\gamma}\mathbb{1}_{\{|\nabla F(u^k)|\geq 1\}} \nabla^2 F(u^k)\right) s^k = -u^k + H_\gamma(-\nabla F(u^k)).$$

In der Praxis hängt der Konvergenzbereich des Newton-Verfahrens von $\gamma$ ab. Man löst daher oft eine Folge von Problemen mit abnehmendem $\gamma$, wobei man die Lösung des vorhergehenden Problems als Startwert für das folgende nimmt.

# 10 NICHTKONVEXE SUBDIFFERENTIALE

Obwohl das Clarke-Subdifferential ein geeignetes Konzept sowohl für nichtglatte aber konvexe als auch für nichtkonvexe aber glatte Funktionale ist, hat es Schwachpunkte für nichtglatte *und* nichtkonvexe Funktionale: Wie in Folgerung 8.7 gezeigt wurde, kann sein Fermat-Prinzip nicht zwischen Minimierern und Maximierern unterscheiden. Der Grund dafür ist, dass das Clarke-Subdifferential immer konvex ist; dies ist eine direkte Konsequenz aus der Konstruktion (8.2) über Dualität mit der verallgemeinerten Richtungsableitung. Um schärfere Resultate zu erhalten, müssen wir *nichtkonvexe* Subdifferentiale direkt über einen *primalen* Grenzwertprozess definieren. Allerdings gibt es nichts umsonst: Die Rechenregeln für die bisherigen Subdifferentiale basierten in fundamentaler Weise auf dem Trennungssatz von Hahn–Banach für konvexe Mengen; analoge Regeln für nichtkonvexe Subdifferentiale sind daher deutlich schwieriger zu erhalten.

## 10.1 DAS BOULIGAND-SUBDIFFERENTIAL

Die erste Definition ist motiviert durch Satz 8.15: Wir *definieren* ein Subdifferential durch geeignete Grenzwerte von klassischen Ableitungen (ohne eine konvexe Hülle). Für $F : X \to \overline{\mathbb{R}}$ definierien wir zuerst die *Menge von Gâteaux-Punkten*

$$G_F := \{x \in X : F \text{ ist Gâteaux-differenzierbar in } x\} \subset (\operatorname{dom} F)^o$$

und dann das *Bouligand-Subdifferential* von $F$ in $x$ als

$$\partial_B F(x) := \{x^* \in X^* : DF(x_n) \rightharpoonup^* x^* \text{ für } G_F \ni x_n \to x\}\,. \tag{10.1}$$

Ist $F : \mathbb{R}^N \to \mathbb{R}$ lokal Lipschitz-stetig, dann folgt aus Satz 8.15 dass $\partial_C F(x) = \operatorname{co} \partial_B F(x)$ ist. Ist $X$ nicht endlich-dimensional, so ist jedoch a priori nicht klar, dass das Bouligand-Subdifferential nichtleer ist selbst für $x \in \operatorname{dom} F$. Außerdem erlaubt dieses Subdifferential wenige nützliche Rechenregeln; es gilt nicht mal ein Fermat-Prinzip.

Beispiel 10.1. Sei $F : \mathbb{R} \to \mathbb{R}$, $F(x) := |x|$. Dann ist $F$ differenzierbar in allen $x \neq 0$ mit $F'(x) = \operatorname{sign}(x)$, und daher ist

$$0 \notin \{-1, 1\} = \partial_B F(0).$$

Dieser Ansatz braucht daher einen subtileren Grenzwertprozess. Im Rest dieses Kapitels betrachten wir einen solchen Ansatz, wobei wir uns auf einen Überblick und Nennung relevanter Resultate aus [Mordukhovich 2006] beschränken. Ein alternativer, axiomatischerer, Zugang zu verallgemeinerten Ableitungen für nichtkonvexe Funktionale findet man in [Penot 2013; Ioffe 2017].

## 10.2 DAS FRÉCHET-SUBDIFFERENTIAL

Wir beginnen mit der folgenden Konstruktion, die die Charakterisierungen der Fréchet-Ableitung und des konvexen Subdifferentials verbindet. Sei $X$ ein Banachraum und $F : X \to \overline{\mathbb{R}}$. Das *Fréchet-Subdifferential* (auch *reguläres Subdifferential* oder *Präsubdifferential* genannt) von $F$ in $x$ ist definiert als

$$\partial_F F(x) := \left\{ x^* \in X^* : \liminf_{y\to x} \frac{F(y) - F(x) - \langle x^*, y - x\rangle_X}{\|y - x\|_X} \geq 0 \right\}. \tag{10.2}$$

Beachte, wie hier die Definition des konvexen Subdifferentials um $x$ „lokalisiert“ wird: Der Zähler muss nicht nicht-negativ für alle $y$ sein; es genügt, wenn dies für $y$ nahe genug an $x$ gilt. Ein ähnliches Argument wie in Satz 4.3 ergibt daher ein Fermat-Prinzip für *lokale* Minimierer.

**Satz 10.2.** *Sei $F : X \to \overline{\mathbb{R}}$ eigentlich und $\bar{x} \in \operatorname{dom} F$ ein lokaler Minimierer. Dann gilt $0 \in \partial_F F(\bar{x})$.*

*Beweis.* Sei $\bar{x} \in \operatorname{dom} F$ ein a lokaler Minimierer. Dann existiert ein $\varepsilon > 0$ so dass $F(\bar{x}) \leq F(y)$ für alle $y \in O_\varepsilon(\bar{x})$ gilt. Dies ist äquivalent zu

$$\frac{F(y) - F(\bar{x}) - \langle 0, y - \bar{x}\rangle_X}{\|y - \bar{x}\|_X} \geq 0 \quad \text{für alle } y \in O_\varepsilon(\bar{x}).$$

Gilt nun $y_n \to \bar{x}$, so ist $y_n \in O_\varepsilon(\bar{x})$ für $n$ groß genug. Übergang zum $\liminf$ in der Ungleichung ergibt daher $0 \in \partial_F(\bar{x})$. □

Für konvexe Funktionale ist der Zähler nach Definition immer nicht-negativ, und das Fréchet-Subdifferential stimmt mit dem konvexen Subdifferential überein.

**Satz 10.3.** *Sei $F : X \to \overline{\mathbb{R}}$ eigentlich, konvex, und unterhalbstetig, und sei $x \in \operatorname{dom} F$. Dann gilt $\partial_F F(x) = \partial F(x)$.*

*Beweis.* Nach Definition des konvexen Subdifferentials gilt für alle $x^* \in \partial F(x)$ dass

$$F(y) - F(x) - \langle x^*, y - x\rangle_X \geq 0 \quad \text{für alle } y \in X.$$

Dividieren durch $\|x - y\|_X > 0$ für $y \neq x$ und Übergang zum $\liminf$ für $y \to x$ ergibt daher $x^* \in \partial_F F(x)$.

Sei andererseits $x^* \in \partial_F F(x)$ und $h \in X \setminus \{0\}$ beliebig. Dann existiert ein $\varepsilon > 0$ mit

$$\frac{F(x + th) - F(x) - \langle x^*, th\rangle_X}{t\|h\|_X} \geq 0 \quad \text{für alle } t \in (0, \varepsilon).$$

Multiplizieren mit $\|h\|_X > 0$ und Grenzübergang $t \to 0$ ergibt dann nach Lemma 4.1

$$\langle x^*, h\rangle_X \leq \frac{F(x + th) - F(x)}{t} \to F'(x; h), \tag{10.3}$$

was nach Lemma 4.2 äquivalent zu $x^* \in \partial F(x)$ ist. □

Ähnliches gilt für Fréchet-differenzierbare Funktionale: Dort ist der Grenzwert in (10.2) gleich Null für alle Folgen.

Satz 10.4. *Sei $F : X \to \mathbb{R}$ Fréchet-differenzierbar in $x \in X$. Dann gilt $\partial_F F(x) = \{F'(x)\}$.*

*Beweis.* Aus der Definition der Fréchet-Ableitung folgt sofort

$$\lim_{y \to x} \frac{F(y) - F(x) - \langle F'(x), y - x\rangle_X}{\|x - y\|_X} = \lim_{\|h\|_X \to 0} \frac{F(x + h) - F(x) - F'(x)h}{\|h\|_X} = 0$$

und damit $F'(x) \in \partial_F F(x)$.

Sei andererseits $x^* \in \partial_F F(x)$ und $h \in X \setminus \{0\}$ beliebig. Wie im Beweis von Satz 10.3 erhalten wir dann, dass gilt

$$\langle x^*, h\rangle_X \leq F'(x; h) = \langle F'(x), h\rangle_X. \tag{10.4}$$

Das gleiche Argument für $-h$ ergibt dann $\langle x^*, h\rangle_X = \langle F'(x), h\rangle_X$ für alle $h \in X$, d. h. $x^* = F'(x)$. □

Für nichtglatte, nichtkonvexe Funktionale kann das Fréchet-Subdifferential aber strikt kleiner als das Clarke-Subdifferential sein.

Beispiel 10.5. Betrachte $F : \mathbb{R} \to \mathbb{R}$, $F(x) := -|x|$. Für alle $x \neq 0$ folgt dann aus Satz 10.4 dass $\partial_F F(x) = \{-\operatorname{sign} x\}$. Für $x = 0$ ist aber für beliebiges $x^* \in \mathbb{R}$

$$\liminf_{y \to 0} \frac{F(y) - F(0) - \langle x^*, y - 0\rangle}{|y - 0|} = \liminf_{y \to 0} (-1 - x^* \cdot \operatorname{sign}(y)) = -1 - |x^*| < 0,$$

und daher gilt

$$\partial_F F(0) = \emptyset \subsetneq [-1, 1] = \partial_C F(0).$$

Beachte, dass in diesem Beispiel $0 \in \operatorname{dom} F$ gilt. Obwohl das Fréchet-Subdifferential im Gegensatz zum Clarke-Subdifferential den Maximierer nicht anzeigt, kann die Tatsache, dass $\partial_F F(x)$ leer sein kann für $x \in \operatorname{dom} F$, zum Problem werden, Rechenregeln mit Gleichheit zu erhalten. Beispiel 10.5 zeigt weiterhin, dass das Fréchet-Subdifferential nicht abgeschlossen ist, was ebenfalls von Nachteil ist. Dies führt auf die folgende und letzte Definition.

## 10.3 DAS MORDUKHOVICH-SUBDIFFERENTIAL

Sei $X$ ein reflexiver Banachraum und $F : X \to \overline{\mathbb{R}}$. Das *Mordukhovich-Subdifferential* (im Englischen auch *basic subdifferential* oder *limiting subdifferential* genannt) von $F$ in $x \in \operatorname{dom} F$ ist definiert als der stark–schwach* Abschluss von $\partial_F F(x)$, d. h. als

$$\partial_M F(x) := \left\{x^* \in X^* : x_n^* \rightharpoonup^* x^* \text{ mit } x_n^* \in \partial_F F(x_n) \text{ für } x_n \to x\right\}. \tag{10.5}$$

Dies kann als Verallgemeinerung der Definition (10.1) des Bouligand-Subdifferentials interpretiert werden. Beachte, dass hier im Gegensatz zu (10.1) die konstante Folge $x_n^* \equiv x^*$ auch in nichtdifferenzierbaren Punkten zugelassen ist. Daraus folgt insbesondere $\partial_F F(x) \subset \partial_M F(x)$, und aus Satz 10.2 folgt daher sofort ein Fermat-Prinzip.

Folgerung 10.6. *Sei $F : X \to \overline{\mathbb{R}}$ eigentlich und $\bar{x} \in \operatorname{dom} F$ ein lokaler Minimierer. Dann gilt $0 \in \partial_M F(\bar{x})$.*

Analog zum Fréchet-Subdifferential erfüllen aber lokale Maximierer nicht das Fermat-Prinzip.

Beispiel 10.7. Betrachte wieder $F : \mathbb{R} \to \mathbb{R}$, $F(x) := -|x|$. Zusammen mit Beispiel 10.5 folgt aus (10.5) sofort, dass gilt $\partial_M F(0) = \{-1, 1\} = \partial_B F(0)$.

Auch das Mordukhovich-Subdifferential fällt mit dem konvexen Subdifferential zusammen.

Satz 10.8. *Sei $X$ ein reflexiver Banachraum, $F : X \to \overline{\mathbb{R}}$ eigentlich, konvex, und unterhalbstetig, und $x \in \operatorname{dom} F$. Dann gilt $\partial_M F(x) = \partial F(x)$.*

*Beweis.* Aus Satz 10.3 folgt $\partial F(x) = \partial_F F(x) \subset \partial_M F(x)$. Sei daher $x^* \in \partial_M F(x)$. Nach Definition existiert dann eine Folge $\{x_n^*\}_{n\in\mathbb{N}} \subset$ mit $x_n^* \rightharpoonup^* x^*$ und $x_n^* \in \partial_F F(x_n) = \partial F(x_n)$ für $x_n \to x$. Wie im Beweis von Folgerung 6.5 folgt daraus $x^* \in \partial F(x)$. □

Ein analoges Resultat gilt für *stetig* differenzierbare Funktionale.

Satz 10.9. *Sei $X$ ein reflexiver Banachraum und $F : X \to \overline{\mathbb{R}}$ stetig differenzierbar in $x \in X$. Dann gilt $\partial_M F(x) = \{F'(x)\}$.*

*Beweis.* Aus Satz 10.3 folgt $\{F'(x)\} = \partial_F F(x) \subset \partial_M F(x)$. Sei daher $x^* \in \partial_M F(x)$. Nach Definition existiert dann eine Folge $\{x_n^*\}_{n\in\mathbb{N}} \subset$ mit $x_n^* \rightharpoonup^* x^*$ und $x_n^* \in \partial_F F(x_n) = \{F'(x_n)\}$ für $x_n \to x$. Aus der Stetigkeit von $F'$ folgt sofort, dass $F'(x_n) \to F(x)$. Da starke Grenzwerte auch schwach-∗ Grenzwerte sind, erhalten wir $x^* = F'(x)$. □

Es gilt auch der folgende Zusammenhang mit dem Clarke-Subdifferential; vergleiche Satz 8.15.

Satz 10.10 ([Mordukhovich 2006, Theorem 3.57]). *Sei $X$ ein reflexiver Banachraum und $F : X \to \mathbb{R}$ lokal Lipschitz-stetig um $x \in X$. Dann gilt $\partial_C F(x) = \operatorname{cl}^* \operatorname{co} \partial_M F(x)$, wobei $\operatorname{cl}^* A$ den schwach-∗ Abschluss von $A \subset X^*$ bezeichnet.*[1]

Das folgende Beispiel illustriert, dass das Mordukhovich-Subdifferential nichtkonvex sein kann.

Beispiel 10.11. Betrachte $F : \mathbb{R}^2 \to \mathbb{R}$, $F(x_1, x_2) = |x_1| - |x_2|$. Da $F$ stetig differenzierbar ist für alle $(x_1, x_2)$, $x_1, x_2 \neq 0$, mit

$$\nabla F(x_1, x_2) \in \{(1,1), (-1,1), (1,-1), (-1,-1)\},$$

erhalten wir aus (10.2), dass gilt

$$\partial_F F(x_1, x_2) = \begin{cases} \{(1,-1)\} & \text{falls } x_1 > 0, x_2 > 0, \\ \{(-1,-1)\} & \text{falls } x_1 < 0, x_2 > 0, \\ \{(-1,1)\} & \text{falls } x_1 < 0, x_2 < 0, \\ \{(-1,-1)\} & \text{falls } x_1 > 0, x_2 < 0, \\ \{(t,-1) : t \in [-1,1]\} & \text{falls } x_1 = 0, x_2 > 0, \\ \{(t,1) : t \in [-1,1]\} & \text{falls } x_1 = 0, x_2 < 0, \\ \emptyset & \text{falls } x_2 = 0. \end{cases}$$

Insbesondere gilt $\partial_F F(0,0) = \emptyset$. Aus (10.5) folgt jedoch, dass gilt

$$\partial_M F(0,0) = \{(t,-1) : t \in [-1,1]\} \cup \{(t,1) : t \in [-1,1]\}.$$

Insbesondere gilt $0 \notin \partial_M F(0,0)$. Andererseits folgt aus Satz 10.10, dass gilt

$$\partial_C F(0,0) = \{(t,s) : t, s \in [-1,1]\} = [-1,1]^2 \tag{10.6}$$

und damit $0 \in \partial_C F(0,0)$. (Beachte, dass $F$ weder ein Minimum noch ein Maximum in $\mathbb{R}^2$ hat, während $(0,0)$ ein nichtglatter Sattelpunkt ist.)

[1] In reflexiven Banach-Räumen stimmt natürlich der schwach-∗ Abschluss mit dem schwachen Abschluss überein. Die Aussage gilt aber allgemeiner in sogenannten *Asplund-Räumen*, die auch nicht-reflexive Banach-Räume umfassen.

Im Gegensatz zum Bouligand-Subdifferential erfüllt das Mordukhovich-Subdifferential nützliche Rechenregeln, obwohl die Annahmen dafür nachvollziehbarerweise einschränkender sind als im konvexen Fall. Die erste Regel folgt wie immer direkt aus der Definition.

Satz 10.12. *Sei $X$ ein reflexiver Banachraum und $F : X \to \overline{\mathbb{R}}$. Dann gilt für alle $\lambda \geq 0$ und $x \in X$, dass*

$$\partial_M(\lambda F)(x) = \lambda \partial_M F(x).$$

Weitere Rechenregeln halten im Unendlichdimensionalen mit Gleichheit nur in eingeschränkten Situationen.

Satz 10.13 ([Mordukhovich 2006, Proposition 1.107]). *Sei $X$ ein reflexiver Banachraum, $F : X \to \mathbb{R}$ stetig differenzierbar, und $G : X \to \overline{\mathbb{R}}$. Dann gilt für alle $x \in \operatorname{dom} G$, dass*

$$\partial_M(F + G)(x) = \{F'(x)\} + \partial_M G(x).$$

Während die letzten beiden Sätze auch für das Fréchet-Subdifferential gelten, erfüllt nur das Mordukhovich-Subdifferential die folgende Kettenregel. Im Gegensatz zu Satz 8.13 darf hier das äußere Funktional sogar erweitert-reellwertig sein.

Satz 10.14 ([Mordukhovich 2006, Proposition 1.112]). *Sei $X$ ein reflexiver Banachraum, $F : X \to Y$ stetig differenzierbar, und $G : Y \to \overline{\mathbb{R}}$. Dann gilt für alle $x \in X$ mit $F(x) \in \operatorname{dom} G$ und $F'(x) : X \to Y$ surjektiv, dass*

$$\partial_M(G \circ F)(x) = F'(x)^* \partial_M G(F(x)).$$

Allgemeinere Rechenregeln erfordern zusätzliche nicht-triviale Annahmen an $F$ und $G$; siehe z. B. [Mordukhovich 2006, Theorem 3.36 und Theorem 3.41].

# LITERATUR


J. Appell & P. Zabreiko (1990), *Nonlinear Superposition Operators*, Cambridge University Press, New York.

H. H. Bauschke & P. L. Combettes (2017), *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2. Aufl., CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, Springer, New York, doi: 10.1007/978-3-319-48311-5.

A. Beck & M. Teboulle (2009), A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM J. Imaging Sci.* 2(1), 183–202, doi: 10.1137/080716542.

M. Brokate (2014), Konvexe Analysis und Evolutionsprobleme, Vorlesungsskript, Zentrum Mathematik, TU München, url: http://www-m6.ma.tum.de/~brokate/cev_ss14.pdf.

A. Cegielski (2012), *Iterative methods for fixed point problems in Hilbert spaces*, Bd. 2057, Lecture Notes in Mathematics, Springer, Heidelberg, doi: 10.1007/978-3-642-30901-4.

A. Chambolle & T. Pock (2011), A first-order primal-dual algorithm for convex problems with applications to imaging, *J Math Imaging Vis* 40(1), 120–145, doi: 10.1007/s10851-010-0251-1.

X. Chen, Z. Nashed & L. Qi (2000), Smoothing methods and semismooth methods for nondifferentiable operator equations, *SIAM J. Numer. Anal.* 38(4), 1200–1216, doi: 10.1137/s0036142999356719.

F. Clarke (2013), *Functional Analysis, Calculus of Variations and Optimal Control*, Springer, London, doi: 10.1007/978-1-4471-4820-3.

F. H. Clarke (1990), *Optimization and Nonsmooth Analysis*, Bd. 5, Classics Appl. Math. SIAM, Philadelphia, PA, doi: 10.1137/1.9781611971309.

C. Clason (2019), *Einführung in die Funktionalanalysis*, Mathematik Kompakt, Birkhäuser, Basel, doi: 10.1007/978-3-030-24876-5.

E. DiBenedetto (2002), *Real analysis*, Birkhäuser Boston, Inc., Boston, MA, doi: 10.1007/978-1-4612-0117-5.

M. Dobrowolski (2010), *Angewandte Funktionalanalysis*, Springer-Verlag, Berlin, doi: 10.1007/978-3-642-15269-6.

I. Ekeland & R. Témam (1999), *Convex Analysis and Variational Problems*, Bd. 28, Classics Appl. Math. SIAM, Philadelphia, PA, doi: 10.1137/1.9781611971088.

B. He & X. Yuan (2012), Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective, *SIAM J. Imag. Sci.* 5(1), 119–149, doi: 10.1137/100814494.

J. Heinonen (2005), *Lectures on Lipschitz analysis*, Bd. 100, Rep. Univ. Jyväskylä Dept. Math. Stat. University of Jyväskylä, URL: http://www.math.jyu.fi/research/reports/rep100.pdf.

A. Ioffe (2017), *Variational Analysis of Regular Mappings: Theory and Applications*, Springer Monographs in Mathematics, Springer International Publishing, DOI: 10.1007/978-3-319-64277-2.

K. Ito & K. Kunisch (2008), *Lagrange Multiplier Approach to Variational Problems and Applications*, Bd. 15, Advances in Design and Control, SIAM, Philadelphia, PA, DOI: 10.1137/1.9780898718614.

B. Kummer (1988), Newton's method for non-differentiable functions, *Mathematical Research* 45, 114–125.

R. Mifflin (1977), Semismooth and semiconvex functions in constrained optimization, *SIAM J. Control Optimization* 15(6), 959–972, DOI: 10.1137/0315061.

B. S. Mordukhovich (2006), *Variational Analysis and Generalized Differentiation I, Basic Theory*, Bd. 330, Grundlehren der mathematischen Wissenschaften, Springer, DOI: 10.1007/3-540-31247-1.

Y. E. Nesterov (1983), A method for solving the convex programming problem with convergence rate $O(1/k^2)$, *Soviet Math. Doklad.* 27(2), 372–376.

Y. Nesterov (2004), *Introductory Lectures on Convex Optimization*, Bd. 87, Applied Optimization, Kluwer Academic Publishers, Boston, MA, DOI: 10.1007/978-1-4419-8853-9.

N. Parikh & S. Boyd (2014), Proximal algorithms, *Foundations and Trends in Optimization* 1(3), 123–231, DOI: 10.1561/2400000003.

J.-P. Penot (2013), *Calculus Without Derivatives*, Bd. 266, Graduate Texts in Mathematics, Springer, New York, DOI: 10.1007/978-1-4614-4538-8.

M. Růžička (2004), *Nichtlineare Funktionalanalysis*, Springer, Berlin, DOI: 10.1007/3-540-35022-5.

A. Schiela (2008), A simplified approach to semismooth Newton methods in function space, *SIAM J. Opt.* 19(3), 1417–1432, DOI: 10.1137/060674375.

W. Schirotzek (2007), *Nonsmooth Analysis*, Universitext, Springer, Berlin, DOI: 10.1007/978-3-540-71333-3.

S. Scholtes (2012), *Introduction to piecewise differentiable equations*, Springer Briefs in Optimization, Springer, New York, DOI: 10.1007/978-1-4614-4340-7.

M. Ulbrich (2011), *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*, Bd. 11, MOS-SIAM Series on Optimization, SIAM, Philadelphia, PA, DOI: 10.1137/1.9781611970692.

T. Valkonen (2018), Testing and non-linear preconditioning of the proximal point method, *Applied Mathematics and Optimization*, DOI: 10.1007/s00245-018-9541-6, ARXIV: 1703.05705.

D. Werner (2011), *Funktionalanalysis*, 7. Aufl., Springer-Verlag, Berlin, DOI: 10.1007/978-3-642-21017-4.