STOCHASTISCHE OPTIMIERUNG

Vorlesungsskript, Sommersemester 2023

Christian Clason

Stand vom 25. Juni 2025

Institut für Mathematik und wissenschaftliches Rechnen
Universität Graz

Inhaltsverzeichnis

1 Optimierung von Erwartungswerten 1
- 1.1 Erwartungswertfunktionale 2
- 1.2 Optimalitätsbedingungen 7

2 Monte-Carlo-Approximation 10
- 2.1 Konsistenz 12
- 2.2 Konzentrationsungleichungen 13
- 2.3 Sampling-Verfahren 23

3 Stochastische Gradientenverfahren 26

4 Robuste Optimierung 32
- 4.1 Konvexe Risikomaße 33
- 4.2 Optimierung mit Risikomaßen 41

A Grundlagen der konvexen Analysis 48
- A.1 Direkte Methode der Variationsrechnung 48
- A.2 Konvexe Funktionen 50
- A.3 Das konvexe Subdifferential 51
- A.4 Fenchel-Dualität 53
- A.5 Epikonvergenz 56
- A.6 Trägerfunktionale 58

B Grundlagen der Wahrscheinlichkeitstheorie 60
- B.1 Wahrscheinlichkeitsmaße 60
- B.2 Integration 61
- B.3 Zufallsvariablen 64
- B.4 Bedingte Erwartung 65

1 OPTIMIERUNG VON ERWARTUNGSWERTEN

Die stochastische Optimierung beschäftigt sich mit der Optimierung von Funktionalen, die von zufälligen Variablen abhängen. Das prototypische Beispiel ist das *Trafikantenproblem*: Ein Trafikant muss im Voraus die Zeitungen ordern, die er am nächsten Tag verkaufen will. Angenommen, der Einkaufspreis für eine Zeitung ist $p_e > 0$, der Verkaufspreis ist $p_v > p_e$, und $x \geq 0$ ist die Anzahl der verkauften Zeitungen. Dann ist der Verlust (da wir minimieren möchten) offensichtlich $f(x) = (p_e - p_v)x < 0$. Allerdings wird diese Anzahl bestimmt von der tatsächlichen Nachfrage ξ, so dass der Verlust gegeben ist durch

$$f(x, \xi) = \begin{cases} p_e x - p_v x & \text{falls } \xi \geq x, \\ p_e x - p_v \xi & \text{sonst,} \end{cases}$$

und das Minimum wird in $x = \xi$ angenommen. Nun ist die Nachfrage im Voraus nicht bekannt; wie üblich wird diese Unsicherheit durch eine Zufallsvariable modelliert. Die Zielfunktion ist nun ebenfalls zufällig und wir können nur versuchen, den *erwarteten Verlust* zu minimieren. Nehmen wir an, dass die Verteilung der Zufallsvariable ξ eine Dichtefunktion $\rho : \mathbb{R} \to [0, \infty)$ besitzt, führt das auf die Funktion

$$\begin{aligned} F(x) = \mathbb{E}[f(x, \xi)] &= \int_0^x (p_e x - p_v \xi)\rho(\xi)\, d\xi + \int_x^\infty (p_e - p_v)x\rho(\xi)\, d\xi \\ &= (p_e - p_v)x + p_v \int_0^x (x - \xi)\rho(\xi)\, d\xi \\ &= (p_e - p_v)x + p_v \int_0^\infty \max\{0, x - \xi\}\rho(\xi)\, d\xi \\ &= (p_e - p_v)x + p_v\, \mathbb{E}[\max\{0, x - \xi\}]. \end{aligned}$$

Dies zeigt bereits die typische mehrstufige Struktur eines stochastischen Optimierungsproblems: In einer ersten Stufe muss eine deterministische Entscheidung getroffen werden, die mit deterministischen Kosten verbunden ist; in unserem Beispiel der (negative) Verlust pro bestellter Zeitung, beschrieben durch den ersten Term. In einer zweiten Stufe trifft dann ein unsicheres Ereignis ein, das verbunden ist mit unsicheren *Rekurs-Kosten*; in diesem Fall der Verkaufsausfall durch zu geringe Nachfrage, beschrieben durch den zweiten Term. (In einem komplizierteren Modell könnte der *Rekurs* darin bestehen, Zeitungen kurzfristig zu einem höheren Preis nachzuordern.) Entsprechend bezeichnet man die Funktion $z \mapsto \max\{0, z\}$ hier auch als *Rekursfunktion*.

Wir beschäftigen uns daher in diesem Kapitel mit Optimierungsproblemen der Form

$$\min_{x \in X} \mathbb{E}[f(x, \omega)] \tag{1.1}$$

für eine Teilmenge $X \subseteq \mathbb{R}^d$ und eine Funktion $f : \mathbb{R}^d \times \Omega \to \mathbb{R}$, abhängig von der Entscheidungsvariable $x \in \mathbb{R}^d$ und einem zufälligen Ereignis $\omega \in \Omega$. Wir nehmen in Folge an, dass $(\Omega, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum ist.

1.1 ERWARTUNGSWERTFUNKTIONALE

Wesentlich dabei ist, welche Eigenschaften das *Erwartungswertfunktional*

$$F : \mathbb{R}^d \to \overline{\mathbb{R}}, \qquad F(x) = \mathbb{E}[f(x, \omega)] = \int_\Omega f(x, \omega)\, d\mu(\omega),$$

von f erbt. Ein wichtiger – und der mit Abstand technischste und komplizierteste – Aspekt dabei ist, wann der Integrand auf der rechten Seite messbar und damit F wohldefiniert ist. Wir beschränken uns hier auf eine einfache Situation, lassen dafür aber (zunächst) zu, dass die Entscheidungsvariable ebenfalls zufällig sein kann. Wir betrachten daher für $1 \leq p \leq \infty$ und $\Omega \subset \mathbb{R}^d$ offen und beschränkt

$$F : L^p(\Omega; \mathbb{R}^d) \to \overline{\mathbb{R}}, \qquad F(u) = \int_\Omega f(u(\omega), \omega)\, d\mu(\omega).$$

Lemma 1.1. *Angenommen, für $f : \mathbb{R}^d \times \Omega \to \overline{\mathbb{R}}$ gilt*

(i) für fast alle $\omega \in \Omega$ ist $x \mapsto f(x, \omega)$ konvex und unterhalbstetig;

(ii) $\omega \to f(u(\omega), \omega))$ ist meßbar für alle $u \in L^1(\Omega)$;

(iii) es existiert ein $g \in L^1(\Omega)$ mit $f(x, \omega) \geq -g(\omega)$ für fast alle $\omega \in \Omega$ und alle $x \in \mathbb{R}^d$.

Dann ist für alle $1 \leq p \leq \infty$ auch

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

konvex und unterhalbstetig.

Beweis. Zunächst können wir durch äquivalente Betrachtung von $\tilde{f}(x, \omega) := f(x, \omega) + g(\omega) \geq 0$ nach Annahme (iii) annehmen, dass f nicht-negativ ist.

Für $u, v \in \operatorname{dom} F$ (sonst ist (A.1) trivialerweise erfüllt) und $\lambda \in [0, 1]$ folgt aus Annahme (i), dass für fast alle $\omega \in \Omega$ gilt

$$0 \leq f(\lambda u(\omega) + (1 - \lambda)v(\omega), \omega) \leq \lambda f(u(\omega), \omega) + (1 - \lambda) f(v(\omega), \omega).$$

Beachte, dass die rechte Seite nach Annahme (ii) meßbar ist. Da $L^1(\Omega;\mathbb{R}^d)$ ein Vektorraum ist, gilt wegen $u, v \in \operatorname{dom} F$ auch $\lambda f(u,\cdot) + (1-\lambda) f(v,\cdot) \in L^1(\Omega;\mathbb{R}^d)$; damit ist der mittlere Ausdruck ebenfalls in $L^1(\Omega)$, und durch Integration der Ungleichung über Ω folgt die Konvexität von F.

Für die Unterhalbstetigkeit verwenden wir Lemma A.5. Sei dafür $\{(u_n, t_n)\}_{n\in\mathbb{N}} \subset \operatorname{epi} F$ mit $u_n \to u$ in $L^p(\Omega;\mathbb{R}^d)$ und $t_n \to t$ in $\mathbb{R}$. Dann existiert eine Teilfolge $\{u_{n_k}\}_{k\in\mathbb{N}}$ mit $u_{n_k} \to u$ punktweise fast überall. Wegen der Unterhalbstetigkeit und Nichtnegativität von f folgt mit dem Lemma von Fatou

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

d. h. $(u,t) \in \operatorname{epi} F$. Also ist $\operatorname{epi} F$ abgeschlossen und damit F unterhalbstetig nach Lemma A.5 (iii). □

Die Annahme (ii) ist hier der kritische (und äußerst technische) Punkt; sie ist insbesondere für sogenannte *normale Integranden* erfüllt. Wir schränken uns hier auf eine typische Klasse ein.

Lemma 1.2. *Sei $f : \mathbb{R}^d \times \Omega \to \mathbb{R}$ eine* Carathéodory-Funktion, *d. h.*

(i) $x \mapsto f(x,\omega)$ ist stetig für fast alle $\omega \in \Omega$;

(ii) $\omega \mapsto f(x,\omega)$ ist meßbar für alle $x \in \mathbb{R}^d$.

Dann ist $\omega \to f(u(\omega),\omega)$ ist meßbar für alle $u \in L^1(\Omega)$.

Beweis. Jedes $u \in L^1(\Omega)$ ist insbesondere meßbar und damit der punktweise Grenzwert von einfachen Funktionen $\{u_n\}_{n\in\mathbb{N}}$ mit $u_n = \sum_{i\in I_n} u_{n,i} \mathbb{1}_{A_i}$ für A_i meßbar und disjunkt. Also ist nach Annahme auch

$$f(u_n(\omega),\omega) = \sum_{i\in I_n} f(u_{n,i},\omega)\mathbb{1}_{A_i}(\omega)$$

meßbar und konvergiert für $n \to \infty$ punktweise gegen $f(u(\omega),\omega)$ für fast alle $\omega \in \Omega$. Also ist auch der Grenzwert $f(u(\cdot),\cdot)$ meßbar. □

Ist $x \mapsto f(x,\omega)$ fast sicher stetig und $|f(x,\omega)| \le g(\omega)$ für alle $x \in \mathbb{R}^d$ und fast alle $\omega \in \Omega$, so zeigt man analog mit dem Satz von Lebesgue die Stetigkeit von F. Beachten Sie, dass wir noch keine Aussage gemacht haben, dass F eigentlich ist; dies erfordert stärkere Annahmen. Zum Beispiel genügt es, dass Ω beschränkt ist und ein $x_0 \in \mathbb{R}^d$ existiert, so dass $f(x_0,\cdot)$

fast sicher endlich ist; in diesem Fall gilt wegen $L^\infty(\Omega;\mathbb{R}^d) \subset L^p(\Omega;\mathbb{R}^d)$ für die konstante Funktion $u_0(\omega) = x_0$, dass $F(u_0) < \infty$ ist.

Die zentrale Frage ist nun, inwieweit wir Integration und Optimierung vertauschen können; insbesondere, ob der Erwartungswert des punktweisen Minimums von f auch das Minimum von F ergibt. Auch hier ist der wesentliche Punkt die Messbarkeit des punktweisen Minimums, die wir nur unter bestimmten Annahmen erhalten. Der folgende Satz ist prototypisch.

Satz 1.3. *Angenommen, für $f : \mathbb{R}^d \times \Omega \to \overline{\mathbb{R}}$ gilt*

(i) für fast alle $\omega \in \Omega$ ist $x \mapsto f(x,\omega)$ konvex und unterhalbstetig;

(ii) für fast alle $\omega \in \Omega$ hat $\operatorname{dom}(f(\cdot,\omega))$ nichtleeres Inneres.

Sei $1 \le p \le \infty$. Ist $F : L^p(\Omega;\mathbb{R}^d) \to \overline{\mathbb{R}}$ wie in Lemma 1.1 definiert eigentlich, dann gilt

$$\inf_{u\in L^p(\Omega;\mathbb{R}^d)} \int_\Omega f(u(\omega),\omega)\,d\mu(\omega) = \int_\Omega \inf_{v\in\mathbb{R}^d} f(v,\omega)\,d\mu(\omega).$$

Beweis. Wegen der Monotonie des Integrals gilt sicher

$$\inf_{u\in L^p(\Omega;\mathbb{R}^d)} \int_\Omega f(u(\omega),\omega)\,d\mu(\omega) \ge \int_\Omega \inf_{v\in\mathbb{R}^d} f(v,\omega)\,d\mu(\omega).$$

Für die umgekehrte Ungleichung genügt es, eine Folge $\{u_k\}k \in \mathbb{N}$ zu konstruieren mit

$$\lim_{k\to\infty} F(u_k) = \lim_{k\to\infty} \int_\Omega f(u_k(\omega),\omega) \le \int_\Omega \inf_{v\in\mathbb{R}^d} f(v,\omega)\,d\mu(\omega).$$

Sei dafür $\{a_k\}_{k\in\mathbb{N}}$ dicht in $\mathbb{R}^d$ (z. B. $\mathbb{Q}^N$ mit der Abzählung aus [Calkin & Wilf 2000]). Nach Annahme an F existiert zunächst ein $u_0 \in L^p(\Omega;\mathbb{R}^d)$ so, dass $f(u_0(\omega),\omega)$ integrierbar ist. Wir definieren nun rekursiv eine Folge $\{u_k\}_{k\in\mathbb{N}} \subset L^p(\Omega;\mathbb{R}^d)$ punktweise fast überall mit

$$u_k(\omega) := \begin{cases} u_{k-1}(\omega) & \text{falls } f(u_{k-1}(\omega),\omega) \le f(a_k,\omega), \\ a_k & \text{sonst.} \end{cases}$$

Dann ist u_k für alle $k \in \mathbb{N}$ messbar (da die Modifikation wegen der Annahme der Messbarkeit von $f(u(\cdot),\cdot)$ für alle $u \in L^1(\Omega;\mathbb{R}^d)$ nur auf einer messbaren Menge passiert) und in $L^p(\Omega;\mathbb{R}^d)$ (da die Modifikation eine Konstante und damit in $L^\infty(\Omega;\mathbb{R}^d)$ ist). Außerdem ist $\{f(u_k(\omega),\omega)\}_{k\in\mathbb{N}}$ nach Konstruktion monoton fallend. Damit ist $u_k \in \operatorname{dom} F$ für alle $k \in \mathbb{N}$.

Wir zeigen nun, dass $\lim_{k\to\infty} f(u_k(\omega),\omega) = \inf_{v\in\mathbb{R}^d} f(v,\omega)$ fast überall gilt. Sei $\omega \in \Omega$ fest. Nach Konstruktion ist dann $\lim_{k\to\infty} f(u_k(\omega),\omega) = \inf_{v\in E_\omega} f(v,\omega)$ für $E_\omega := \{a_k\}_{k\in\mathbb{N}} \cup$

$\{u_0(\omega)\}$. Angenommen, dieses Infimum ist strikt größer als das über alle $v \in \mathbb{R}^d$. Dann existiert ein $\bar{v} \in \mathbb{R}^d$ mit

$$\inf_{v\in\mathbb{R}^d} f(v,\omega) < f(\bar{v},\omega) < \inf_{v\in E_\omega} f(v,\omega).$$

Da $\operatorname{dom} f(\cdot,\omega)$ nichtleeres Inneres hat, ist $E_\omega \cap (\operatorname{dom} f(\cdot,\omega))^o$ dicht in $\operatorname{dom} f(\cdot,\omega)$. Wir finden also eine Folge $\{v_k\}_{k\in\mathbb{N}} \subset E_\omega \cap (\operatorname{dom} f(\cdot,\omega))^o$ mit $v_k \to \bar{v}$. Nun ist $x \mapsto f(x,\omega)$ nach Satz A.9 auf $(\operatorname{dom} f(\cdot,\omega))^o$ stetig, woraus folgt

$$f(\bar{v},\omega) = \lim_{k\to\infty} f(v_k,\omega) \geq \inf_{v\in E_\omega} f(v,\omega),$$

im Widerspruch zur Annahme.

Daraus folgt insbesondere, dass $\inf_{v\in\mathbb{R}^d} f(v,\omega)$ als punktweiser Grenzwert messbarer Funktionen messbar ist. Aus dem Satz von der monotonen Konvergenz folgt nun

$$\begin{aligned}\inf_{u\in L^p(\Omega;\mathbb{R}^d)} \int_\Omega f(u(\omega),\omega)\,d\mu(\omega) &\leq \lim_{k\to\infty} \int_\Omega f(u_k(\omega),\omega)\,d\mu(\omega)\\ &= \int_\Omega \lim_{k\to\infty} f(u_k(\omega),\omega)\,d\mu(\omega)\\ &= \int_\Omega \inf_{v\in\mathbb{R}^d} f(v,\omega)\,d\mu(\omega).\end{aligned}$$

und damit die behauptete Gleichheit. □

Die in dem Beweis konstruierte dichte Folge nennt man einen *Castaing-Repräsentanten* der mengenwertigen Abbildung $x \mapsto \operatorname{dom} f(x,\omega)$. Die Existenz einer solchen Folge (die man mit etwas mehr Aufwand unter noch etwas schwächeren Voraussetzungen zeigen kann, siehe z. B. [Pennanen & Perkkiö 2024, Chap. 1]) ist die wesentliche Voraussetzung für die Vertauschbarkeit von Infimum und Integral.

Aus diesem fundamentalen Resultat können wir nun weitere „Vertauschungsregeln“ ableiten.

Satz 1.4. *Angenommen, für $f : \mathbb{R}^d \times \Omega \to \overline{\mathbb{R}}$ gilt*

(i) für fast alle $\omega \in \Omega$ ist $x \mapsto f(x,\omega)$ konvex und unterhalbstetig;

(ii) für fast alle $\omega \in \Omega$ hat $\operatorname{dom}(f(\cdot,\omega))$ nichtleeres Inneres.

Sei $1 \leq p < \infty$. Ist $F : L^p(\Omega;\mathbb{R}^d) \to \overline{\mathbb{R}}$ wie in Lemma 1.1 definiert eigentlich, dann gilt für $u^* \in L^q(\Omega;\mathbb{R}^d)$, $\frac{1}{p} + \frac{1}{q} = 1$,

$$F^*(u^*) = \int_\Omega f^*(u^*(\omega),\omega)\,d\mu(\omega).$$

Beweis. Erfüllt f die Annahmen (i) und (ii), so auch

$$g(x,\omega) := f(x,\omega) - u^*(\omega)\cdot x$$

für alle $u^* \in L^q(\Omega;\mathbb{R}^d)$, wobei $\cdot$ das Skalarprodukt im $\mathbb{R}^d$ bezeichnet. Aus Satz 1.3 folgt daher sofort

$$\begin{aligned}
F^*(u^*) &= \sup_{u\in L^p(\Omega;\mathbb{R}^d)} \langle u^*, u\rangle_{L^p(\Omega;\mathbb{R}^d)} - F(u)\\
&= \sup_{u\in L^p(\Omega;\mathbb{R}^d)} \int_\Omega u^*(\omega)\cdot u(\omega) - f(u(\omega),\omega)\,d\mu(\omega)\\
&= -\inf_{u\in L^p(\Omega;\mathbb{R}^d)} \int_\Omega f(u(\omega),\omega) - u^*(\omega)\cdot u(\omega)\,d\mu(\omega)\\
&= \int_\Omega -\inf_{x\in\mathbb{R}^d} f(x,\omega) - u^*(\omega)\cdot x\,d\mu(\omega)\\
&= \int_\Omega \sup_{x\in\mathbb{R}^d} u^*(\omega)\cdot x - f(x,\omega)\,d\mu(\omega)\\
&= \int_\Omega f^*(u^*(\omega),\omega)\,d\mu(\omega).
\end{aligned}$$

□

Für das nächste Resultat verwenden wir die Schreibweise $\partial f(x;\omega)$ für das Subdifferential der Abbildung $x \mapsto f(x,\omega)$ für festes $\omega \in \Omega$.

Satz 1.5. *Angenommen, für $f : \mathbb{R}^d \times \Omega \to \overline{\mathbb{R}}$ gilt*

(i) für fast alle $\omega \in \Omega$ ist $x \mapsto f(x,\omega)$ konvex und unterhalbstetig;

(ii) für fast alle $\omega \in \Omega$ hat $\operatorname{dom}(f(\cdot,\omega))$ nichtleeres Inneres.

Sei $1 \le p < \infty$. Ist $F : L^p(\Omega;\mathbb{R}^d) \to \overline{\mathbb{R}}$ wie in Lemma 1.1 definiert eigentlich, dann gilt für $u \in \operatorname{dom} F$ und $\frac{1}{p} + \frac{1}{q} = 1$

$$\partial F(u) = \left\{ u^* \in L^q(\Omega;\mathbb{R}^d) \,\middle|\, u^*(\omega) \in \partial f(u(\omega);\omega) \quad \text{für fast alle } \omega \in \Omega \right\}.$$

Beweis. Aus der Fenchel–Young-Ungleichung folgt zusammen mit Satz 1.4 für alle $u^* \in L^q(\Omega;\mathbb{R}^d)$

$$\begin{aligned}
0 &\le F(u) + F^*(u^*) - \langle u^*, u\rangle_{L^p(\Omega;\mathbb{R}^d)}\\
&= \int_\Omega f(u(\omega),\omega) + f^*(u^*(\omega),\omega) - u^*(\omega)\cdot u(\omega)\,d\mu(\omega),
\end{aligned}$$

mit Gleichheit nach Satz A.21 genau dann, wenn $u^* \in \partial F(u)$ gilt. Da der Integrand – ebenfalls nach der Fenchel–Young-Ungleichung, diesmal für $f(\cdot,\omega)$ und $f^*(\cdot,\omega)$ – punktweise nichtnegativ ist, kann das Integral nur verschwinden, wenn der Integrand punktweise fast überall verschwindet. Das ist aber, wieder nach Satz 1.4, genau dann der Fall, wenn $u^*(\omega) \in \partial f(u(\omega);\omega)$ fast überall gilt. □

1.2 OPTIMALITÄTSBEDINGUNGEN

Wir kehren nun zurück zu dem Problem (1.1) mit deterministischer Entscheidungsvariable, für das eine punktweise fast sichere Minimierung natürlich nicht mehr funktioniert. (Wie man sich anhand des Beispiels $f(x,\xi) = \xi(x-1)^2 + (x+1)^2$ mit ξ gleichverteilt auf $(0,1)$ leicht vergewissert, ist der Erwartungswert der punktweisen Minimierer auch kein Minimierer des Erwartungswerts.) Die „Koppelung“ über alle Realisierungen (auch *Nichtantizipativitätsbedingung* genannt) ist also wesentlich.

Satz 1.6. *Angenommen, für $f : \mathbb{R}^d \times \Omega \to \overline{\mathbb{R}}$ gilt*

(i) für fast alle $\omega \in \Omega$ ist $x \mapsto f(x,\omega)$ konvex und unterhalbstetig;

(ii) es existiert ein $x_0 \in (\operatorname{dom} f(\cdot,\omega))^o$ für fast alle $\omega \in \Omega$;

(iii) es existiert ein $g \in L^1(\Omega)$ mit $f(x,\omega) \geq -g(\omega)$ für fast alle $\omega \in \Omega$ und alle $x \in \mathbb{R}^d$.

Sei

$$F : \mathbb{R}^d \to \overline{\mathbb{R}}, \qquad F(x) := \mathbb{E}[f(x,\omega)].$$

Dann gilt für alle $x \in \operatorname{dom} F$

$$\partial F(x) = \left\{ \mathbb{E}[u^*] \,\middle|\, u^* \in L^1(\Omega;\mathbb{R}^d), \quad u^*(\omega) \in \partial f(x;\omega) \text{ für fast alle } \omega \in \Omega \right\}.$$

Beweis. Wir kombinieren Satz 1.5 mit der Kettenregel Satz A.16. Sei dafür $1 \leq p < \infty$ beliebig und

$$A : \mathbb{R}^d \to L^p(\Omega;\mathbb{R}^d), \qquad [Ax](\omega) := x \quad \text{für fast alle } \omega \in \Omega.$$

Dann ist A offensichtlich linear und stetig (wegen $\mu(\Omega) = 1 < \infty$). Weiter ist nach Annahme Ax_0 ein innerer Punkt von $\operatorname{dom} F$. Da F nach Lemma 1.1 konvex und unterhalbstetig ist, können wir Satz A.16 anwenden und erhalten

$$\partial(F \circ A)(x) = A^* \partial F(Ax).$$

Es bleibt also nur noch, die Adjungierte A^* zu bestimmen. Dazu sei $x \in \mathbb{R}^d$ und $u^* \in L^q(\Omega;\mathbb{R}^d)$ mit $1 < q \leq \infty$ beliebig. Dann gilt

$$\langle u^*, Ax \rangle_{L^p} = \int_\Omega u^*(\omega) \cdot x \, d\mu(\omega) = x \cdot \int_\Omega u^*(\omega) \, d\mu(\omega) = x \cdot \mathbb{E}[u^*]$$

und damit $A^* u^* = \mathbb{E}[u^*]$. Da $p > 1$ beliebig war, folgt die Behauptung nun aus Satz 1.5. □

Wir wenden dieses Resultat auf das eingangs beschriebene Trafikantenproblem an.

Beispiel 1.7. Betrachte für $p_v > p_e > 0$ das Problem

$$\min_{x \geq 0} (p_e - p_v)x + p_v \, \mathbb{E}[\max\{0, x - \xi\}].$$

Offensichtlich ist $f(x, \xi) := \max\{0, x - \xi\} \geq 0$ konvex und stetig bezüglich x. Nach Lemma 1.1 ist $F(x) = \mathbb{E}[f(x, \xi)]$ also konvex und unterhalbstetig. Weiter ist $\operatorname{dom} f(\cdot, \xi) = \mathbb{R}$ für alle $\xi \in \mathbb{R}$ und damit F eigentlich. Damit ist das gesamte Funktional eigentlich, konvex, und unterhalbstetig; aus Satz A.10, Satz A.15, Satz A.12, und Satz 1.6 erhalten wir daher für einen Minimierer $\bar{x} \geq 0$ die Optimalitätsbedingung

$$0 \in \partial \delta_{[0,\infty)}(\bar{x}) + \{p_e - p_v\} + p_v \, \mathbb{E}[\partial f(\bar{x}; \xi)].$$

Für den ersten Term erhalten wir direkt aus der Definition

$$\partial \delta_{[0,\infty)}(x) = \begin{cases} (-\infty, 0] & \text{falls } x = 0, \\ \{0\} & \text{sonst.} \end{cases}$$

Für den letzten Term gilt zunächst nach Satz A.17

$$\partial f(x; \xi) = \begin{cases} \{0\} & \text{falls } x < \xi, \\ \{1\} & \text{falls } x > \xi, \\ [0, 1] & \text{falls } x = \xi. \end{cases}$$

Um den Erwartungswert zu berechnen, nehmen wir an, dass $\xi \in \{0, 1, \ldots, M\}$ für $M \in \mathbb{N}$ uniform verteilt ist, so dass gilt

$$\mathbb{E}[\partial f(x, \xi)] = \frac{1}{M+1} \sum_{i=0}^{M} \partial f(x; i).$$

Man sieht nun leicht, dass gilt

$$\begin{aligned} \sum_{i=0}^{M} \partial f(0; i) &= [0, 1] + \{0\} + \{0\} + \{0\} + \cdots = [0, 1] \\ \sum_{i=0}^{M} \partial f(1; i) &= \{1\} + [0, 1] + \{0\} + \{0\} + \cdots = [1, 2], \\ &\vdots \\ \sum_{i=0}^{M} \partial f(n; i) &= [n, n+1], \qquad n \leq M, \end{aligned}$$

und analog $\sum_{i=0}^{M} \partial f(x; i) = \{\lceil x \rceil\}$ für $M > x \notin \mathbb{N}$ sowie $\sum_{i=0}^{M} \partial f(x; i) = \{M + 1\}$ für $x > M$.

Wir können die Optimalitätsbedingung daher umformen in

$$(M+1)\frac{p_v - p_e}{p_v} \in \begin{cases} (-\infty, 1] & \text{falls } \bar{x} = 0, \\ [\bar{x}, \bar{x}+1] & \text{falls } \bar{x} \in \{1, \dots, M\}, \\ \{M+1\} & \text{falls } \bar{x} > M, \\ \{\lceil \bar{x} \rceil\} & \text{sonst.} \end{cases}$$

Der vorletzte Fall kann wegen der Annahme $p_v - p_e > p_v$ ausgeschlossen werden, was zusammen mit der Zulässigkeit $0 \leq \bar{x} \leq M$ impliziert. Die restlichen Fälle implizieren, dass das optimale $\bar{x}$ so zu wählen ist, dass gilt

$$\bar{x} \leq (M+1)\frac{p_v - p_e}{p_v} \leq \bar{x} + 1$$

(z. B. $\bar{x} = 7$ für $p_e = 1$, $p_v = 3$, und $M = 10$). Insbesondere ist für $(M+1)(p_v - p_e) < p_v$ die einzige Lösung $\bar{x} = 0$. (Ist der Gewinn selbst bei $M+1$ verkauften Zeitschriften nicht größer als der Verlust einer einzigen nicht verkauften Zeitschrift, lohnt sich also das Trafikantengeschäft nicht.) Ist der mittlere Term zufällig ganzzahlig, so ist die Lösung nicht mehr eindeutig, und jedes $\bar{x} \in [(M+1)\frac{p_v - p_e}{p_v} - 1, (M+1)\frac{p_v - p_e}{p_v}]$ ist optimal.
Im Vergleich ist der Erwartungswert des punktweisen Mimimierers $\bar{x}(\xi) = \xi$

$$\mathbb{E}[\bar{x}] = \frac{1}{M+1}\sum_{i=0}^{M} i = \frac{1}{2}M$$

unabhängig von Einkaufs- und Verkaufspreis und daher suboptimal (z. B. $\mathbb{E}[\bar{x}] = 5 < 7$ für $M = 10$).

2 MONTE-CARLO-APPROXIMATION

Die Berechnung von Erwartungswerten ist im Allgemeinen schwierig, da nur in wenigen besonderen Fällen eine geschlossene Formel bekannt ist; oft ist auch die Verteilung μ der Zufallsvariable ξ nicht bekannt. Die Frage ist daher, inwiefern das stochastische Optimierungsproblem (1.1) mit Hilfe der *empirischen Verteilung* $\mu_N := \sum_{i=1}^N \delta_{\xi_i}$ für $N > 0$ *Stichproben* $\xi_1, \dots, \xi_N$ von ξ – z. B. aus historischen Daten oder über Monte-Carlo-Verfahren erzeugt – approximiert werden kann. Wir betrachten also statt $F(x) = \mathbb{E}[f(x, \xi)]$ die Monte-Carlo-Approximation

$$F_N(x) := \frac{1}{N} \sum_{i=1}^N f(x, \xi_i). \tag{2.1}$$

Wir nehmen hier und in Folge an, dass die ξ_i unabhängig und identisch verteilt sind nach der selben Verteilung wie ξ.

Beachte, dass die Stichproben zufällig sind und damit $F_N(x)$ selber wieder eine Zufallsvariable (auf dem Produkt-Wahrscheinlichkeitsraum $(\Omega^N, \bigotimes_{n=1}^N \mathcal{A}, \mu^n)$) ist. Diese hat den Erwartungswert

$$\mathbb{E}[F_N(x)] = \frac{1}{N} \sum_{i=1}^N \mathbb{E}[f(x, \xi_i)] = \frac{1}{N} \sum_{i=1}^N \mathbb{E}[f(x, \xi)] = \mathbb{E}[f(x, \xi)] = F(x)$$

wegen der Linearität des Erwartungswerts sowie der Annahme an die Stichproben. (In der Statistik nennt man F_N einen *verzerrungsfreien Standardschätzer* des Erwartungswerts.) Für die Varianz gilt analog für $x \in \operatorname{dom} F$

$$\begin{aligned}
\mathbb{V}[F_N(x)] &= \mathbb{E}[(F_N(x) - \mathbb{E}[F_N(x)])^2] = \mathbb{E}[(F_N(x) - F(x))^2] \\
&= \mathbb{E}\left[\left(\frac{1}{N} \sum_{i=1}^N f(x, \xi_i) - \frac{1}{N}\left(\sum_{i=1}^N F(x)\right)\right)^2\right] \\
&= \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}[(f(x, \xi_i)] = \frac{1}{N} \mathbb{V}[f(x, \xi)]
\end{aligned}$$

da die Stichproben unabhängig verteilt und damit die $f(x, \xi_i)$ unkorreliert mit dem selben Erwartungswert $F(x)$ sind. Insbesondere erhalten wir daraus eine Darstellung des

punktweisen L^2*-Approximationsfehlers*

$$\|F_N(x) - F(x)\|_{L^2} = \int_\Omega |[F_N(x)](\omega) - F(x)|^2 \, d\mu(\omega) = N^{-1/2}\sqrt{\mathbb{V}[f(x,\xi)]}. \tag{2.2}$$

Nun kann man sich fragen, ob unter diesen Annahmen das Infimum der Monte-Carlo-Approximation F_N ebenfalls ein unverzerrter Schätzer des Infimums von F ist. Das ist im Allgemeinen aber *nicht* der Fall.

Lemma 2.1. *Es bezeichne* $F^* := \inf_{x\in\mathbb{R}^d} F(x)$ *und* $F_N^* := \inf_{x\in\mathbb{R}^d} F_N(x)$. *Dann gilt für alle* $N \in \mathbb{N}$

$$\mathbb{E}[F_N^*] \leq \mathbb{E}[F_{N+1}^*] \leq F^*.$$

Beweis. Aus der Definition des Infimums und der Monotonie des Erwartungswerts folgt direkt

$$\mathbb{E}[F_N^*] = \mathbb{E}\left[\inf_{x\in\mathbb{R}^d} F_N(x)\right] \leq \mathbb{E}[F_N(\tilde{x})] = F(\tilde{x}) \quad \text{für alle } \tilde{x} \in \mathbb{R}^d,$$

und Infimum über alle $\tilde{x} \in \mathbb{R}^d$ ergibt $\mathbb{E}[F_N^*] \leq F^*$ für alle $N \in \mathbb{N}$.

Für die Monotonie zeigt man zunächst per Induktion, dass gilt

$$F_{N+1}(x) = \frac{1}{N+1}\sum_{i=1}^{N+1}\left(\frac{1}{N}\sum_{j\neq i}^{N+1} f(x,\xi_j)\right).$$

Daraus folgt

$$\begin{aligned}
\mathbb{E}[F_{N+1}^*] = \mathbb{E}\left[\inf_{x\in\mathbb{R}^d} F_{N+1}(x)\right] &= \mathbb{E}\left[\inf_{x\in\mathbb{R}^d} \frac{1}{N+1}\sum_{i=1}^{N+1}\left(\frac{1}{N}\sum_{j\neq i}^{N+1} f(x,\xi_j)\right)\right] \\
&\geq \mathbb{E}\left[\frac{1}{N+1}\sum_{i=1}^{N+1}\left(\inf_{x_i\in\mathbb{R}^d} \frac{1}{N}\sum_{j\neq i}^{N+1} f(x_i,\xi_j)\right)\right] \\
&= \frac{1}{N+1}\sum_{i=1}^{N+1}\mathbb{E}\left[\inf_{x\in\mathbb{R}^d} \frac{1}{N}\sum_{j\neq i}^{N+1} f(x,\xi_j)\right] \\
&= \frac{1}{N+1}\sum_{i=1}^{N+1}\mathbb{E}[F_N^*] = \mathbb{E}[F_N^*],
\end{aligned}$$

da nach Annahme an die Stichproben die innere Summe eine Realisierung des Schätzers $F_N(x)$ liefert. □

Der Schätzer F_N^* von F^* hat also stets eine negative Verzerrung, die monoton steigend ist.

2.1 KONSISTENZ

Wir untersuchen nun die Konvergenz $N \to \infty$. Dazu kombinieren wir das Gesetz der Großen Zahl mit der Epikonvergenz aus Anhang A.5. Da F_N selber wieder eine Zufallsvariable auf Ω ist, hängt die Konvergenz von $\omega \in \Omega$ ab. Wir sagen F_N epikonvergiert *fast sicher* gegen F, wenn gilt $[F_N](\omega) \to_e F$ für μ-fast alle $\omega \in \Omega$.

Lemma 2.2. *Angenommen, für $f : \mathbb{R}^d \times \Omega \to \overline{\mathbb{R}}$ gilt*

(i) für fast alle $\omega \in \Omega$ ist $x \mapsto f(x, \omega)$ konvex und unterhalbstetig;

(ii) es existiert ein $g \in L^1(\Omega)$ mit $f(x, \omega) \geq -g(\omega)$ für fast alle $\omega \in \Omega$ und alle $x \in \mathbb{R}^d$.

Hat dom F *nichtleeres Inneres, dann gilt $F_N \to_e F$ fast sicher.*

Beweis. Aus den Annahmen folgt nach Lemma 1.1, dass F eigentlich, konvex, und unterhalbstetig ist. Weiter ist F_N als endliche Summe konvexer und unterhalbstetiger Funktionen ebenfalls fast sicher konvex und unterhalbstetig. Außerdem ist $f(x, \xi_i) \in L^1(\Omega)$ für jede Stichprobe ξ_i. Nach dem Starken Gesetz der Großen Zahl (Satz B.7) gilt daher für *festes* $x \in \mathbb{R}^d$

$$\frac{1}{N} \sum_{i=1}^{N} f(x, \xi_i) \to \mathbb{E}[f(x, \xi)] \quad \text{mit Wahrscheinlichkeit 1,} \tag{2.3}$$

d. h. es existiert eine μ-Nullmenge $\mathcal{N}_x \subset \Omega$ mit $[F_N(x)](\omega) \to F(x)$ für alle $\omega \in \Omega \setminus \mathcal{N}_x$. Das genügt nach Satz A.25 schon fast für die Behauptung; allerdings darf für die fast sichere Epikonvergenz die Nullmenge nicht von x abhängen.

Sei nun $E \subset \mathbb{R}^d$ eine abzählbare dichte Teilmenge. Da E abzählbar ist, muss auch $\mathcal{N} := \bigcup_{x \in E} \mathcal{N}_x$ eine μ-Nullmenge sein. Für alle $\omega \in \Omega \setminus \mathcal{N}$ gilt nach (2.3) also $[F_N(x)](\omega) \to F(x)$ auf einer dichten Teilmenge. Aus Satz A.25 folgt daher $[F_N](\omega) \to_e F$ für $\omega \in \Omega \setminus \mathcal{N}$ und damit die fast sichere Epikonvergenz. □

Zusammen mit einem Kompaktheitsargument erhalten wir daraus die fast sichere Konvergenz von Minimierern der Monte-Carlo-Approximation. Da diese nicht eindeutig sein müssen, brauchen wir noch etwas Notation. Die Menge der Minimierer von F bzw. F_N bezeichnen wir mit

$$S := \left\{ x \in \mathbb{R}^d \,\middle|\, F(x) = \inf_{x \in \mathbb{R}^d} F(x) \right\},$$
$$S_N := \left\{ x \in \mathbb{R}^d \,\middle|\, F_N(x) = \inf_{x \in \mathbb{R}^d} F_N(x) \right\}.$$

Die *Abweichung* (englisch: „deviation“) dieser beiden Mengen ist dann definiert als

$$D(S_N, S) := \sup_{x \in S_N} \inf_{y \in S} \|x - y\|,$$

d. h. den maximalen Abstand von allen Minimierern von F_N zur Menge der Minimierer von F. Beachte, dass diese Definition nicht symmetrisch ist, im Allgemeinen also $D(S_N, S) \neq D(S, S_N)$ gilt. Da F_N eine Zufallsvariable ist, gilt dies natürlich auch für S_N und damit für $D(S_N, S)$.

Satz 2.3. *Angenommen, für $f : \mathbb{R}^d \times \Omega \to \overline{\mathbb{R}}$ gilt*

(i) für fast alle $\omega \in \Omega$ ist $x \mapsto f(x, \omega)$ konvex und unterhalbstetig;

(ii) es existiert ein $g \in L^1(\Omega)$ mit $f(x, \omega) \geq -g(\omega)$ für fast alle $\omega \in \Omega$ und alle $x \in \mathbb{R}^d$.

Hat $\operatorname{dom} F$ *nichtleeres Inneres und ist S nichtleer und beschränkt, dann gilt $D(S_N, S) \to 0$ und $F_N^* := \inf_{x \in \mathbb{R}^d} F_N \to \inf_{x \in \mathbb{R}^d} F(x) =: F^*$ fast sicher.*

Beweis. Unter diesen Annahmen gilt nach Lemma 2.2 fast sicher $F_N \to_e F$. Da $S \subset \mathbb{R}^d$ beschränkt ist, existiert eine konvexe und kompakte Menge $K \subset \mathbb{R}^d$ mit $S \subset K^o$. Betrachte nun die Menge $\hat{S}_N$ der Lösungen von $\inf_{x \in K} F_N(x)$. Da K nichtleer und kompakt ist und nach Satz B.7 $F_N(x) \to F(x) < \infty$ fast sicher für alle $x \in S$ konvergiert, ist auch $F_N(x) < \infty$ fast sicher für alle $x \in S$ und $N \in \mathbb{N}$ groß genug. Also ist $\hat{S}_N$ fast sicher nichtleer. Weiterhin erfüllt wegen $K^o \supset S \neq \emptyset$ auch $\hat{F} := F + \delta_K$ die Voraussetzungen von Lemma 2.2, es gilt also auch $F_N + \delta_K =: \hat{F}_N \to_e \hat{F}$ fast sicher. Außerdem hat nach Konstruktion $\hat{F}$ die selben Minimierer wie F.

Wir zeigen nun durch Widerspruch, dass $D(\hat{S}_N, S) \to 0$ fast sicher konvergiert. Sei $\omega \in \Omega$ beliebig mit $[F_N](\omega) \to_e F$. Angenommen, es existieren unendlich viele $\hat{x}_N \in \hat{S}_N$ mit $\inf_{x \in S} \|x - \hat{x}_N\| \geq \varepsilon$ für ein $\varepsilon > 0$. Da K kompakt ist, existiert daher eine Teilfolge von $\{\hat{x}_N\}_{N \in \mathbb{N}}$, die gegen ein $\hat{x} \in K$ konvergiert, für das aus Stetigkeitsgründen dann ebenfalls $\inf_{x \in S} \|x - \hat{x}\| \geq \varepsilon > 0$ gelten muss. Insbesondere ist $\hat{x} \notin S$. Andererseits folgt aus der Epikonvergenz von $\hat{F}_N \to_e \hat{F}$ nach Satz A.24 aber $\hat{x} \in S$ und damit der gesuchte Widerspruch.

Aus $D(\hat{S}_N, S) \to 0$ und $S \subset K^o$ folgt nun, dass für N groß genug jedes $\hat{x}_N \in \hat{S}_N$ auch $\hat{x}_N \in K^o$ erfüllt. Wegen der Konvexität von K und F_N ist daher $\hat{x}_N$ auch ein globaler Minimierer von F_N. Damit gilt $\hat{S}_N = S_N$ für N groß genug, woraus die erste Behauptung folgt. Aus der Kompaktheit von K und Satz A.24 folgt außerdem $F_N^* \to F^*$ (zunächst nur entlang einer Teilfolge; da jede konvergente Teilfolge aber den selben Grenzwert F^* haben muss, sogar für die gesamte Folge). □

2.2 KONZENTRATIONSUNGLEICHUNGEN

Wir betrachten nun Fehlerabschätzungen für *festes* $N < \infty$. Das fundamentale Werkzeug aus der Stochastik ist die folgende Ungleichung.

Lemma 2.4 (Chernov-Ungleichung). *Sei $\xi \in L^1(\Omega)$ mit $\mathbb{E}[\xi] = 0$. Dann gilt*

$$\mathbb{P}[\xi \geq \varepsilon] \leq e^{-\psi^*(\varepsilon)} \qquad \text{für alle } \varepsilon \geq 0 \tag{2.4}$$

mit der Fenchel-Konjugierten ψ^* der log-momentenerzeugenden Funktion

$$\psi(t) := \ln \mathbb{E}[e^{t\xi}].$$

Beweis. Sei $\varepsilon \geq 0$ und $t \geq 0$ beliebig. Dann gilt $x \geq \varepsilon$ genau dann, wenn $e^{tx} \geq e^{t\varepsilon}$ ist. Zusammen mit der Markov-Ungleichung (B.1) folgt daraus

$$\begin{aligned}\mathbb{P}[\xi \geq \varepsilon] &= \int_{-\infty}^{\infty} \mathbb{1}_{[\varepsilon,\infty)}(\xi(\omega))\, d\mu(\omega) = \int_{-\infty}^{\infty} \mathbb{1}_{[\exp(t\varepsilon),\infty)}\left(\exp(t\xi(\omega))\right)\, d\mu(\omega) \\ &= \mathbb{P}[e^{t\xi} \geq e^{t\varepsilon}] \leq e^{-t\varepsilon}\, \mathbb{E}[e^{t\xi}].\end{aligned}$$

Wir suchen nun $t \geq 0$ so, dass die rechte Seite minimal ist, bzw. äquivalent deren Logarithmus, d. h.

$$\inf_{t\geq 0} \ln\left(e^{-t\varepsilon}\, \mathbb{E}[e^{t\xi}]\right) = \inf_{t\geq 0} \psi(t) - t\varepsilon = -\sup_{t\geq 0} t\varepsilon - \psi(t).$$

Um daraus die Behauptung zu erhalten, müssen wir nur noch argumentieren, dass wir hier das Supremum über alle $t \in \mathbb{R}$ nehmen können. Zunächst gilt wegen $\psi(0) = 0$

$$\psi^*(s) = \sup_{t\in\mathbb{R}} ts - \psi(s) \geq 0s - \psi(0) = 0.$$

Andererseits folgt aus der Konvexität von $x \mapsto e^{tx}$ für $t \in \mathbb{R}$ zusammen mit der Jensen-Ungleichung (B.2)

$$\mathbb{E}[e^{t\xi}] \geq e^{t\,\mathbb{E}[\xi]} = 1$$

und damit wegen der Monotonie des Logarithmus $\psi(t) \geq 0$. Für $t < 0$ und $\varepsilon \geq 0$ folgt daher $\psi(t) \geq 0 \geq t\varepsilon$ und damit

$$t\varepsilon - \psi(t) \leq 0 \qquad \text{für alle } t < 0.$$

Also ist

$$\inf_{t\geq 0} \ln\left(e^{-t\varepsilon}\, \mathbb{E}[e^{t\xi}]\right) = -\sup_{t\in\mathbb{R}} t\varepsilon - \psi(t) = -\psi^*(\varepsilon),$$

und aus der Monotonie der Exponentialfunktion folgt die Behauptung. □

Beispiel 2.5. Wir betrachten zwei kanonische Beispiele.

(i) $\xi \sim \mathcal{N}_{0,\sigma^2}$. Dann ist $\mathbb{E}[\xi] = 0$ und

$$\mathbb{E}[e^{t\xi}] = \int_{-\infty}^{\infty} e^{tx} \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}\, dx = e^{\frac{1}{2}\sigma^2 t^2}$$

und damit

$$\psi(t) = \frac{\sigma^2}{2} t^2.$$

Aus Beispiel A.19 (iii) zusammen mit Lemma A.20 (i) folgt dann

$$\psi^*(s) = \frac{1}{2\sigma^2} s^2.$$

(ii) $\xi \sim \mathcal{U}_{(-a,a)}$. Dann ist ebenfalls $\mathbb{E}[\xi] = 0$, und eine elementare aber längliche Abschätzung mit Hilfe der Konvexität der Exponentialfunktion (siehe z. B. [Shapiro, Dentcheva & Ruszczynski 2021, Prop. 9.80]) zeigt, dass gilt

$$\mathbb{E}[e^{t\xi}] \leq e^{t^2 a^2/2}$$

und damit

$$\psi(t) \leq \frac{a^2}{2} t^2.$$

Analog zu oben folgt dann

$$\psi^*(s) = \sup_t ts - \psi(t) \geq \sup_t ts - \frac{a^2}{2} t^2 = \frac{1}{2a^2} s^2.$$

Eine Zufallsvariable ξ mit $\mathbb{E}[\xi] = 0$ und $\psi(t) \leq \frac{\sigma^2}{2} t^2$ und deshalb $\psi^*(s) \geq \frac{1}{2\sigma^2} s^2$ heißt *subnormal* (mit *Pseudo-Varianz* σ^2).

Dieses Resultat wenden wir nun auf den Approximationsfehler der Monte-Carlo-Approximation an.

Satz 2.6 (Konzentrationsungleichung). *Sei $\xi - \mathbb{E}[\xi]$ subnormal mit Pseudo-Varianz $\sigma^2 > 0$, und seien $\xi_1, \ldots, \xi_N$ unabhängig und identisch verteilte Stichproben. Dann gilt*

$$\mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N} \xi_i - \mathbb{E}[\xi] \geq \varepsilon\right] \leq \exp\left(-\frac{N\varepsilon^2}{2\sigma^2}\right) \qquad \text{für alle } \varepsilon \geq 0.$$

Beweis. Wegen $\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} \xi_i - \mathbb{E}[\xi]\right] = 0$ können wir Lemma 2.4 auf den Monte-Carlo-Approximationsfehler anwenden; es bleibt nur, dessen log-momentenerzeugende Funktion

ψ_N sowie die Fenchel-Konjugierte zu berechnen. Zunächst gilt offensichtlich für alle $t \in \mathbb{R}$

$$\begin{aligned}\mathbb{E}\left[\exp\left(t\left(\frac{1}{N}\sum_{i=1}^{N}\xi_i - \mathbb{E}[\xi]\right)\right)\right] &= \mathbb{E}\left[\exp\left(\sum_{i=1}^{N}\frac{t}{N}(\xi_i - \mathbb{E}[\xi])\right)\right] \\ &= \mathbb{E}\left[\prod_{i=1}^{N}\exp\left(\frac{t}{N}(\xi_i - \mathbb{E}[\xi])\right)\right] \\ &= \prod_{i=1}^{N}\mathbb{E}\left[\exp\left(\frac{t}{N}(\xi_i - \mathbb{E}[\xi])\right)\right]\end{aligned}$$

da die ξ_i unabhängig sind. Sei nun ψ die log-momentenerzeugende Funktion von $\xi - \mathbb{E}[\xi]$. Dann folgt

$$\begin{aligned}\psi_N(t) &= \ln\mathbb{E}\left[\exp\left(t\left(\frac{1}{N}\sum_{i=1}^{N}\xi_i - \mathbb{E}[\xi]\right)\right)\right] \\ &= \sum_{i=1}^{N}\ln\mathbb{E}\left[\exp\left(\frac{t}{N}(\xi_i - \mathbb{E}[\xi])\right)\right] \\ &= N\psi(N^{-1}t),\end{aligned}$$

wieder aus der identischen Verteilung der ξ_i. Aus Lemma A.20 (i) erhalten wir nun zusammen mit der Subnormalität

$$\psi_N^*(s) = N\psi^*(s) \geq \frac{N}{2\sigma^2}s^2$$

und damit die Behauptung. □

Wir kehren nun zurück zur Monte-Carlo-Approximation von Erwartungswertfunktionalen. Hier ist die Situation etwas diffiziler, da wir nicht nur ein einzige Approximation benötigen, sondern $F(x) = \mathbb{E}[f(x,\xi)]$ für verschiedene $x \in \mathbb{R}^d$ auswerten wollen. Wir müssen also untersuchen, mit welcher Wahrscheinlichkeit *mehrere* unabhängige Approximationen *alle* einen kleinen Fehler haben.

Lemma 2.7. *Für $j = 1,\dots,M$ seien $\xi_j - \mathbb{E}[\xi_j]$ subnormal mit Pseudo-Varianz $\sigma_j^2 > 0$, und seien $\xi_{j,1},\dots,\xi_{j,N}$ unabhängig und identisch verteilte Stichproben von ξ_j. Dann gilt*

$$\mathbb{P}\left[\max_{j=1,\dots,M}\left|\frac{1}{N}\sum_{i=1}^{N}\xi_{j,i} - \mathbb{E}[\xi_j]\right| \geq \varepsilon\right] \leq 2M\exp\left(-\frac{N\varepsilon^2}{2\sigma^2}\right) \qquad \text{für alle } \varepsilon \geq 0$$

und $\sigma^2 := \max_{j=1,\dots,M}\sigma_j^2$.

Beweis. Sei $\varepsilon \geq 0$. Dann folgt für alle $j = 1,\dots,M$ zunächst aus der Konzentrationsungleichung in Satz 2.6

$$\mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N}\xi_{j,i} - \mathbb{E}[\xi_j] \geq \varepsilon\right] \leq \exp\left(-\frac{N\varepsilon^2}{2\sigma_j^2}\right) \leq \exp\left(-\frac{N\varepsilon^2}{2\sigma^2}\right).$$

Betrachte nun $-\xi_j + \mathbb{E}[\xi_j]$. Dann ist natürlich auch $\mathbb{E}[-\xi_j + \mathbb{E}[\xi_j]] = 0$, und für die log-momentenerzeugende Funktion gilt

$$\overline{\psi}_j(t) := \ln \mathbb{E}[\exp(t(-\xi_j + \mathbb{E}[\xi_j]))] = \ln \mathbb{E}[\exp(-t(\xi_j - \mathbb{E}[\xi_j]))] = \psi_j(-t) \leq \frac{\sigma_j^2}{2} t^2$$

wegen der Subnormalität von $\xi_j - \mathbb{E}[\xi_j]$. Also ist auch $-\xi_j + \mathbb{E}[\xi_j]$ subnormal mit Pseudo-Varianz $\sigma_j^2 > 0$, und wir erhalten analog

$$\mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N} \xi_{j,i} - \mathbb{E}[\xi_j] \leq -\varepsilon\right] = \mathbb{P}\left[-\frac{1}{N}\sum_{i=1}^{N} \xi_{j,i} + \mathbb{E}[\xi_j] \geq \varepsilon\right] \leq \exp\left(-\frac{N\varepsilon^2}{2\sigma^2}\right).$$

Aus der Subadditivität von Wahrscheinlichkeitsmaßen folgt dann

$$\begin{aligned}
&\mathbb{P}\left[\max_{j=1,\dots,M}\left|\frac{1}{N}\sum_{i=1}^{N} \xi_{j,i} - \mathbb{E}[\xi_j]\right| \geq \varepsilon\right] \\
&\qquad = \mathbb{P}\left[\bigcap_{j=1}^{M}\left(\left\{\frac{1}{N}\sum_{i=1}^{N} \xi_{j,i} - \mathbb{E}[\xi_j] \geq \varepsilon\right\} \cap \left\{\frac{1}{N}\sum_{i=1}^{N} \xi_{j,i} - \mathbb{E}[\xi_j] \leq -\varepsilon\right\}\right)\right] \\
&\qquad \leq \sum_{j=1}^{M}\left(\mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N} \xi_{j,i} - \mathbb{E}[\xi_j] \geq \varepsilon\right] + \mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N} \xi_{j,i} - \mathbb{E}[\xi_j] \leq -\varepsilon\right]\right) \\
&\qquad \leq 2M \exp\left(-\frac{N\varepsilon^2}{2\sigma^2}\right)
\end{aligned}$$

und damit die Behauptung. □

Die Anzahl der benötigten Stichproben wächst also linear in der Anzahl der Punkte, in denen wir das Erwartungsfunktional gleichmäßig approximieren möchten. Das ist insbesondere ein Problem, wenn wir das Maximum über überabzählbar viele $x \in \mathbb{R}^d$ bilden wollen. Wir benötigen daher zusätzliche Struktur (insbesondere Kompaktheit), um nur endlich viele Punkte berücksichtigen zu müssen.

Satz 2.8. *Angenommen, für $f : \mathbb{R}^d \times \Omega \to \overline{\mathbb{R}}$ und eine kompakte Menge $X \subset \mathbb{R}^d$ gilt*

(i) für alle $x \in X$ ist $\omega \mapsto f(x, \omega)$ integrierbar;

(ii) es existiert eine integrierbare Funktion $g : \Omega \to \mathbb{R}$ mit

$$|f(x_1, \omega) - f(x_2, \omega)| \leq g(\omega)\|x_1 - x_2\|_2 \qquad \text{für fast alle } \omega \in \Omega, \text{ und } x_1, x_2 \in X;$$

(iii) für alle $x \in X$ ist $f(x, \xi) - \mathbb{E}[f(x, \xi)]$ subnormal mit Pseudo-Varianz $\sigma(x)^2 > 0$;

(iv) $g(\xi) - \mathbb{E}[g(\xi)]$ ist subnormal mit Pseudo-Varianz $\sigma_0^2 > 0$.

Dann existieren für jedes $\varepsilon > 0$ Konstanten $C, c > 0$ mit

$$\mathbb{P}\left[\sup_{x\in X}\left|\frac{1}{N}\sum_{i=1}^{N} f(x,\xi_i) - \mathbb{E}[f(x,\xi)]\right| \geq \varepsilon\right] \leq Ce^{-cN}.$$

Beweis. Da $X \subset \mathbb{R}^d$ kompakt ist, existiert für beliebiges $\delta > 0$ eine Überdeckung mit endlich vielen Kugeln $B_\delta(x_j)$, $j = 1, \ldots, r$. Sei nun $x \in X$ beliebig mit $x \in B_\delta(x_j)$. Dann können wir abschätzen

$$\begin{aligned}\left|\frac{1}{N}\sum_{i=1}^{N} f(x,\xi_i) - \mathbb{E}[f(x,\xi)]\right| \leq{}& \left|\frac{1}{N}\sum_{i=1}^{N} f(x,\xi_i) - \frac{1}{N}\sum_{i=1}^{N} f(x_j,\xi_i)\right| \\ &+ \left|\frac{1}{N}\sum_{i=1}^{N} f(x_j,\xi_i) - \mathbb{E}[f(x_j,\xi)]\right| \\ &+ \left|\mathbb{E}[f(x_j,\xi)] - \mathbb{E}[f(x,\xi)]\right|.\end{aligned}$$

Die Terme auf der rechten Seite schätzen wir nun separat ab. Für den ersten Term verwenden wir die fast sichere Lipschitz-Stetigkeit (Annahme (ii)) und erhalten

$$\left|\frac{1}{N}\sum_{i=1}^{N} f(x,\xi_i) - \frac{1}{N}\sum_{i=1}^{N} f(x_j,\xi_i)\right| \leq \frac{1}{N}\sum_{i=1}^{N} g(\xi_i)\|x - x_j\|_2 \leq \delta\frac{1}{N}\sum_{i=1}^{N} g(\xi_i).$$

Nach Annahme (iv) können wir auf $g(\xi)$ die Konzentrationsungleichung aus Satz 2.6 anwenden; für beliebiges $\gamma > 0$ gilt daher

$$\mathbb{P}\left[\frac{1}{N}\sum_{i=1}^{N} g(\xi_i) - \mathbb{E}[g(\xi)] \geq \gamma\right] \leq \exp\left(-\frac{N\gamma^2}{2\sigma_0^2}\right) =: p_\gamma,$$

d. h. mit Wahrscheinlichkeit $1 - p_\gamma$ ist

$$\left|\frac{1}{N}\sum_{i=1}^{N} f(x,\xi_i) - \frac{1}{N}\sum_{i=1}^{N} f(x_j,\xi_i)\right| \leq \delta\frac{1}{N}\sum_{i=1}^{N} g(\xi_i) \leq \delta(\mathbb{E}[g(\xi)] + \gamma).$$

Für den zweiten Term verwenden wir Lemma 2.7 und erhalten für beliebiges $\rho > 0$

$$\mathbb{P}\left[\max_{j=1,\ldots,r}\left|\frac{1}{N}\sum_{i=1}^{N} f(x_j,\xi_i) - \mathbb{E}[f(x_j,\xi)]\right| \geq \rho\right] \leq 2r\exp\left(-\frac{N\rho^2}{2\sigma^2}\right) =: p_\rho$$

für $\sigma^2 := \max_{j=1,\ldots,n} \sigma(x_j)^2$, d. h. mit Wahrscheinlichkeit $1 - p_\rho$ ist

$$\left|\frac{1}{N}\sum_{i=1}^{N} f(x_j,\xi_i) - \mathbb{E}[f(x_j,\xi)]\right| \leq \rho \qquad \text{für alle } j = 1, \ldots, r.$$

Der dritte Term folgt ebenfalls aus der fast sicheren Lipschitzstetigkeit: Es gilt

$$\left|\mathbb{E}[f(x_j,\xi)] - \mathbb{E}[f(x,\xi)]\right| \le \mathbb{E}[|f(x_j,\xi) - f(x,\xi)|] \le \mathbb{E}[g(\xi)]\|x - x_j\|_2 \le \mathbb{E}[g(\xi)]\delta.$$

Zusammen können wir folgern, dass mit Wahrscheinlichkeit $1 - p_\gamma - p_\rho$ gilt

$$\left|\frac{1}{N}\sum_{i=1}^{N} f(x,\xi_i) - \mathbb{E}[f(x,\xi)]\right| \le (2\,\mathbb{E}[g(\xi)] + \gamma)\delta + \rho.$$

Es bleibt nun, $\delta, \gamma, \rho > 0$ (unabhängig von N) so zu wählen, dass die rechte Seite kleiner als ε ist. Wir versuchen hier nicht, die optimale Schranke zu finden, und wählen einfach $\gamma = 1$ und $\rho = \varepsilon - (2\,\mathbb{E}[g(\xi)] + 1)\delta$. Damit $\rho > 0$ ist, müssen wir also z. B. $\delta = \varepsilon(2\,\mathbb{E}[g(\xi)] + 1)^{-1}$ wählen; in diesem Fall ist $\rho = \frac{\varepsilon}{2}$. Wir können aus geometrischen Überlegungen auch eine Obergrenze für r angeben: Da X kompakt ist, existiert ein $M > 0$ mit $X \subset B_M(0)$, und eine Kugel mit Radius M in $\mathbb{R}^d$ kann mit $(1 + \frac{2M}{\delta})^d$ Kugeln vom Radius $\delta < M$ überdeckt werden.

Der Approximationsfehler ist also für alle $x \in X$ größer als ε mit Wahrscheinlichkeit

$$p := p_\gamma + p_\rho = \exp\left(-\frac{N}{2\sigma_0^2}\right) + 2r\exp\left(-\frac{N\varepsilon^2}{8\sigma^2}\right) \le Ce^{-cN}$$

für

$$C := 2\left(1 + \frac{4M(2\,\mathbb{E}[g(\xi)] + 1)}{\varepsilon}\right)^d + 1, \qquad c := \min\left\{\frac{1}{2\sigma_0^2}, \frac{\varepsilon^2}{8\sigma^2}\right\}. \qquad \square$$

Beachte, dass Annahme (ii) impliziert, dass $X \subset \operatorname{dom} f(\cdot,\omega)$ fast sicher gilt. Man sieht in dem Beweis auch deutlich die typische exponentielle Abhängigkeit der Konstanten C von der Dimension d der Entscheidungsvariablen sowie die bekannte Relation $\varepsilon \propto N^{-1/2}$ von Monte-Carlo-Verfahren, vergleiche (2.2).

Analog zu Abschnitt 2.1 leiten wir daraus Konvergenzraten für die Monte-Carlo-Approximation unter einer Kompaktheitsannahme her. Seien F^* bzw. F_N^* wieder der optimale Funktionswert des Erwartungswertfunktionals bzw. seiner Monte-Carlo-Approximation. Wir benötigen auch noch für $\varepsilon \ge 0$ die Menge der ε-Minimierer

$$S^\varepsilon := \left\{x \in \mathbb{R}^d \,\middle|\, F(x) \le F^* + \varepsilon\right\}, \qquad S_N^\varepsilon := \left\{x \in \mathbb{R}^d \,\middle|\, F_N(x) \le F_N^* + \varepsilon\right\}.$$

Wir betrachten nun für $\varepsilon \ge \delta \ge 0$ das Ereignis $\{S_N^\delta \subset S^\varepsilon\}$, d. h. dass alle δ-Minimierer von F_N auch ε-Minimierer von F sind, und zeigen, dass dessen Wahrscheinlichkeit für $N \to \infty$ exponentiell gegen 1 geht.

Wieder nehmen wir in einem ersten Schritt an, dass wir nur endlich viele Punkte $x \in \mathbb{R}^d$ betrachten müssen. Sei dafür $X \subset \mathbb{R}^d$ eine endliche Menge. Wir bezeichnen die optimalen Funktionswerte und Mengen der ε-Minimierer über X mit $\hat{F}^*$, $\hat{S}^\varepsilon$ bzw. $\hat{F}_N^*$, $\hat{S}_N^\varepsilon$.

Satz 2.9. *Angenommen, für $f : \mathbb{R}^d \times \Omega \to \overline{\mathbb{R}}$ und eine endliche Menge $X \subset \mathbb{R}^d$ gilt*

(i) für alle $x \in X$ ist $\omega \mapsto f(x, \omega)$ integrierbar;

(ii) für alle $x \in X$ ist $f(x, \xi) - \mathbb{E}[f(x, \xi)]$ subnormal mit Pseudo-Varianz $\sigma(x)^2 > 0$.

Dann gilt für jedes $\varepsilon \geq \delta \geq 0$

$$\mathbb{P}\left[\hat{S}_N^\delta \not\subset \hat{S}^\varepsilon\right] \leq |X| \exp\left(-\frac{N(\varepsilon - \delta)^2}{8\sigma^2}\right)$$

mit $\sigma^2 := \max_{x \in X} \sigma(x)^2$.

Beweis. Wir müssen die Wahrscheinlichkeit abschätzen, dass ein $x \in X$ existiert mit $x \notin \hat{S}^\varepsilon$ und $x \in \hat{S}_N^\delta$; in anderen Worten, die Wahrscheinlichkeit des Ereignisses

$$\{\hat{S}_N^\delta \not\subset \hat{S}^\varepsilon\} = \bigcup_{x \in X \setminus \hat{S}^\varepsilon} \bigcap_{y \in X} \{F_N(x) \leq F_N(y) + \delta\}.$$

Sei nun $\hat{x}^* \in \hat{S} \subset X$ beliebig. Dann gilt

$$\begin{aligned} \mathbb{P}\left[\hat{S}_N^\delta \not\subset \hat{S}^\varepsilon\right] &\leq \sum_{x \in X \setminus \hat{S}^\varepsilon} \mathbb{P}\left[\bigcap_{y \in X} \{F_N(x) \leq F_N(y) + \delta\}\right] \\ &\leq \sum_{x \in X \setminus \hat{S}^\varepsilon} \mathbb{P}\left[F_N(x) \leq F_N(\hat{x}^*) + \delta\right]. \end{aligned} \tag{2.5}$$

Angenommen, $X \setminus \hat{S}^\varepsilon$ ist nicht leer (ansonsten wäre jedes $x \in X$ ein ε-Minimierer von F über X, und es ist nichts zu zeigen). Dann gilt

$$\begin{aligned} F(\hat{x}^*) &= \hat{F}^* - \min_{x \in X \setminus \hat{S}^\varepsilon} F(x) + \min_{x \in X \setminus \hat{S}^\varepsilon} F(x) \\ &\leq \hat{F}^* - \min_{x \in X \setminus \hat{S}^\varepsilon} F(x) + F(\tilde{x}) \\ &=: F(\tilde{x}) - \varepsilon^* \qquad \text{für alle } \tilde{x} \in X \setminus \hat{S}^\varepsilon \end{aligned}$$

mit $\varepsilon^* := \min_{x \in X \setminus \hat{S}^\varepsilon} F(x) - \hat{F}^* > \varepsilon \geq 0$, da X endlich ist.

Definiere nun für $x \in X \setminus \hat{S}^\varepsilon$

$$g(x, \xi) := f(\hat{x}^*, \xi) - f(x, \xi).$$

Dann ist

$$\mathbb{E}[g(x, \xi)] = F(\hat{x}^*) - F(x) \leq -\varepsilon^* \qquad \text{für alle } x \in X \setminus \hat{S}^\varepsilon. \tag{2.6}$$

Definiere weiter

$$G_N(x) := \frac{1}{N}\sum_{i=1}^{N} g(x,\xi_j) = F_N(\hat{x}^*) - F_N(x),$$

wobei $\xi_1,\dots,\xi_N$ die unabhängig und identisch verteilten Stichproben für die Monte-Carlo-Approximation von F sind. Dann ist

$$\begin{aligned}\mathbb{P}\left[F_N(x) \le F_N(\hat{x}^*) + \delta\right] &= \mathbb{P}\left[G_N(x) \ge -\delta\right] \\ &= \mathbb{P}\left[G_N(x) - \mathbb{E}[g(x,\xi)] \ge -\delta - \mathbb{E}\left[g(x,\xi)\right]\right].\end{aligned}$$

Diese Wahrscheinlichkeit schätzen wir jetzt mit Satz 2.6 ab. Es gilt

$$\begin{aligned}g(x,\xi) - \mathbb{E}[g(x,\xi)] &:= f(\hat{x}^*,\xi) - f(x,\xi) - \mathbb{E}[f(\hat{x}^*,\xi) - f(x,\xi)] \\ &= (f(\hat{x}^*,\xi) - \mathbb{E}[f(\hat{x}^*,\xi)]) + (f(x,\xi) - \mathbb{E}[f(x,\xi)]) \\ &=: \hat{E}^* + E,\end{aligned}$$

wobei $\hat{E}^*$ und E nach Annahme subnormal mit Pseudo-Varianz $\sigma(\hat{x}^*)^2$ bzw. $\sigma(x)^2$ sind. Also ist auch $\mathbb{E}[\hat{E}^* + E] = 0$, und für die momentenerzeugende Funktion folgt aus der Hölderschen Ungleichung

$$\begin{aligned}\mathbb{E}\left[e^{t(\hat{E}^*+E)}\right] = \int_\Omega e^{t\hat{E}^*} e^{tE}\, d\mu &\le \left(\int_\Omega \exp(t\hat{E}^*)^2\, d\mu(\omega)\right)^{1/2}\left(\int_\Omega \exp(tE)^2\, d\mu(\omega)\right)^{1/2} \\ &= \left(\mathbb{E}\left[e^{(2t)\hat{E}^*}\right]\right)^{1/2}\left(\mathbb{E}\left[e^{(2t)E}\right]\right)^{1/2} \\ &\le \exp\left(\frac{\sigma(x^*)^2}{2}(2t)^2\right)^{1/2}\exp\left(\frac{\sigma(x)^2}{2}(2t)^2\right)^{1/2} \\ &= \exp\left(\frac{2\sigma(x^*)^2 + 2\sigma(x)^2}{2}t^2\right) \le \exp\left(\frac{4\sigma^2}{2}t^2\right).\end{aligned}$$

Also ist $g(x,\xi) - \mathbb{E}[g(x,\xi)]$ subnormal mit Pseudo-Varianz $4\sigma^2 > 0$.

Aus Satz 2.6 folgt daher

$$\begin{aligned}\mathbb{P}\left[G_N(x) - \mathbb{E}[g(x,\xi)] \ge -\delta - \mathbb{E}\left[g(x,\xi)\right]\right] &\le \exp\left(-\frac{N(-\delta - \mathbb{E}[g(x,\xi)])^2}{8\sigma^2}\right) \\ &\le \exp\left(-\frac{N(-\delta+\varepsilon^*))^2}{8\sigma^2}\right) \\ &\le \exp\left(-\frac{N(-\delta+\varepsilon))^2}{8\sigma^2}\right)\end{aligned}$$

wegen $-\delta - \mathbb{E}[g(x,\xi)] \ge -\delta + \varepsilon^* > -\delta + \varepsilon \ge 0$ nach (2.6) und Annahme.

Zusammen mit (2.5) folgt nun die Behauptung. □

Die Anzahl der benötigten Stichproben wächst also logarithmisch sowohl in der Größe der endlichen Menge X als auch der gewünschten Wahrscheinlichkeit $1 - p$:

Folgerung 2.10. *Unter den Annahmen von Satz 2.9 gilt für alle $\varepsilon > \delta \geq 0$ und $0 < p < 1$*

$$\mathbb{P}\left[\hat{S}_N^\delta \subset \hat{S}^\varepsilon\right] \geq 1 - p$$

für

$$N \geq \frac{8\sigma^2}{(\varepsilon - \delta)^2} \log\left(\frac{|X|}{p}\right).$$

Beachte, dass wir im Beweis die Abschätzung tatsächlich mit $\varepsilon^* > \varepsilon$ hergeleitet haben; wir bekommen also exponentielle Konvergenz auch für $\varepsilon = \delta = 0$, allerdings mit einer von X abhängigen Konstanten, die beliebig nahe an 0 liegen kann:

Folgerung 2.11. *Unter den Annahmen von Satz 2.9 existiert eine Konstante $C > 0$ mit*

$$\limsup_{N\to\infty} \frac{1}{N} \ln\left(1 - \mathbb{P}\left[\hat{S}_N \not\subset \hat{S}\right]\right) \leq -C.$$

Zusammen mit der bekannten Kompaktheitsannahme erhalten wir daraus das Resultat für den allgemeinen Fall.

Satz 2.12. *Angenommen, für $f : \mathbb{R}^d \times \Omega \to \overline{\mathbb{R}}$ gilt*

(i) für alle $x \in \mathbb{R}^d$ ist $\omega \mapsto f(x, \omega)$ integrierbar;

(ii) es existiert eine integrierbare Funktion $g : \Omega \to \mathbb{R}$ mit

$$|f(x_1, \omega) - f(x_2, \omega)| \leq g(\omega)\|x_1 - x_2\|_2 \qquad \text{für fast alle } \omega \in \Omega, \text{ und } x_1, x_2 \in \mathbb{R}^d;$$

(iii) für alle $x \in \mathbb{R}^d$ ist $f(x, \xi) - \mathbb{E}[f(x, \xi)]$ subnormal mit Pseudo-Varianz $\sigma(x)^2 > 0$;

(iv) $g(\xi) - \mathbb{E}[g(\xi)]$ ist subnormal mit Pseudo-Varianz $\sigma_0^2 > 0$.

Ist $S \subset \mathbb{R}^d$ nichtleer und beschränkt, dann existieren für jedes $\varepsilon > \delta > 0$ Konstanten $C, c > 0$ abhängig von $\varepsilon - \delta > 0$ mit

$$\mathbb{P}\left[S_N^\delta \not\subset S^\varepsilon\right] \leq Ce^{-cN}.$$

Beweis. Wähle $\gamma > \mathbb{E}[g(\xi)] \geq 0$ und $\rho := \frac{\varepsilon - \delta}{4\gamma} > 0$. Da S nichtleer und beschränkt ist, existiert eine konvexe und kompakte Menge $X \subset \mathbb{R}^d$ mit $S \subset K^o$. Diese besitzt für $\rho > 0$ eine Überdeckung mit endlich vielen Kugeln $B_\delta(x_j)$, $j = 1, \ldots, r$. Sei weiter $x^* \in S$ beliebig.

Wir betrachten nun die endliche Menge $\hat{X} := \{x_1, \dots, x_r, x^*\}$. Dann gilt für das entsprechend eingeschränkte Problem offensichtlich $\hat{S} \subset S$ und $\hat{F}^* = F^*$. Setze nun

$$\hat{\varepsilon} := \varepsilon - \gamma\rho = \frac{3}{4}\varepsilon + \frac{1}{4}\delta > 0, \qquad \hat{\delta} := \delta + \gamma\rho = \frac{1}{4}\varepsilon + \frac{3}{4}\delta > 0,$$

so dass gilt $\hat{\varepsilon} - \hat{\delta} = \frac{1}{2}(\varepsilon - \delta) > 0$.

Sei nun x_N^δ ein δ-Minimierer von F_N. Dann existiert für δ hinreichend klein und N hinreichend groß wegen $D(S_N, S) \to 0$ nach Satz 2.3 ein $\hat{x} \in \hat{X}$ mit $\|\hat{x} - x_N^\delta\| \leq \rho$. Aus Annahme (iii) und (iv) folgt dann zusammen mit Satz 2.6 für $\tau := \gamma - \mathbb{E}[g(\xi)] > 0$, dass mit Wahrscheinlichkeit

$$1 - p_\gamma := 1 - \exp\left(-\frac{N\tau^2}{2\sigma_0^2}\right)$$

gilt

$$F_N(\hat{x}) - F_N(x_N^\delta) \leq \frac{1}{N}\sum_{i=1}^{N} g(\xi_i)\|\hat{x} - x_N^\delta\| \leq (\mathbb{E}[g(\xi)] + \tau)\rho = \gamma\rho.$$

In diesem Fall ist

$$F_N(\hat{x}) \leq F_N(x_N^\delta) + \gamma\rho \leq F_N^* + \delta + \gamma\rho \leq \hat{F}_N^* + \hat{\delta},$$

da das Minimum von F_N über $\hat{X} \subset \mathbb{R}^d$ sicher nicht kleiner sein kann als das unbeschränkte Minimum. Also ist $\hat{x} \in \hat{S}_N^{\hat{\delta}}$, und aus Satz 2.9 folgt, dass mit Wahrscheinlichkeit

$$1 - p_\varepsilon := 1 - |\hat{X}| \exp\left(-\frac{N(\hat{\varepsilon} - \hat{\delta})^2}{8\sigma^2}\right) = 1 - (r+1)\exp\left(\frac{N(\varepsilon - \delta)^2}{32\sigma^2}\right)$$

gilt $\hat{x} \in \hat{S}^{\hat{\varepsilon}}$. Schließlich ist nach Annahme (ii) analog zu oben

$$F(x_N^\delta) - F(\hat{x}) \leq \mathbb{E}[g(\xi)]\|x_N^\delta - \hat{x}\| \leq \gamma\rho,$$

und daher gilt mit Wahrscheinlichkeit $1 - p_\gamma - p_\varepsilon$

$$F(x_N^\delta) \leq F(\hat{x}) + \gamma\rho \leq \hat{F}^* + \hat{\varepsilon} + \gamma\rho = F^* + \varepsilon$$

wegen $\hat{x} \in \hat{S}^{\hat{\varepsilon}}$ und der Definition von $\hat{\varepsilon}$. Also ist x_N^δ ein ε-Minimierer von F.

Eine analoge Abschätzung von $p_\gamma + p_\varepsilon$ wie in Satz 2.8 ergibt nun die Behauptung. □

2.3 SAMPLING-VERFAHREN

Wir beschreiben zum Abschluss dieses Kapitels, wie man in der Praxis Stichproben $\xi_i \sim X$, $i = 1, \dots, N$, einer Zufallsvariable X für die Monte-Carlo-Approximation generiert.

Während es für uniform oder normalverteilte Zufallsvariablen einfache Algorithmen gibt, ist dies für allgemeine Verteilungen schwierig. Hier haben sich in den letzten Jahren *Langevin-Sampler* als Verfahren der Wahl durchgesetzt.

Ausgangspunkt dafür ist, die Zufallsvariable X als stationären Punkt eines stochastischen Prozesses zu beschreiben. Wir nehmen dafür an, dass die Verteilung μ_X von X eine stetig differenzierbare Dichte $\rho : \mathbb{R}^d \to (0, \infty)$ mit $\rho(x) > 0$ für alle $x \in \mathbb{R}^d$ besitzt. Dann heißt die stochastische Differentialgleichung

$$dX_t = \frac{1}{2}\nabla \log \rho(X_t)dt + dW_t, \tag{2.7}$$

wobei $\{W_t\}_{t\geq 0}$ ein Wiener-Prozess ist, *Langevin-Diffusion*. Unter weiteren Annahmen an die Dichte ρ kann man zeigen, dass eine Lösung $\{X_t\}_{t\geq 0}$ von (2.7) existiert und invariant ist unter ρ, siehe [Roberts & Tweedie 1996, Theorem 2.1]. Entsprechend sind diese Verfahren besonders populär für sogenannte *Gibbs-Verteilungen* mit Dichten der Form $\rho(x) = e^{-G(x)}$ für $G : \mathbb{R}^d \to \mathbb{R}$ konvex.

Um daraus ein Sampling-Verfahren zu gewinnen, wendet man auf (2.7) eine Zeitdiskretisierung, speziell das *Euler-Maruyama-Verfahren*, an: Für eine Schrittweite $\tau > 0$ wähle

$$\begin{cases} X_0 \sim \mathcal{N}(0, \tau\,\mathrm{Id}), \\ X_{n+1} = X_n + \dfrac{\tau}{2}\nabla \log \rho(X_n) + \sqrt{\tau}\eta, \qquad \eta \sim \mathcal{N}(0, \mathrm{Id}). \end{cases}$$

Dies ist der *unadjusted Langevin algorithm (ULA)*. Beachte, dass dann nach Konstruktion gilt

$$X_{n+1} \sim \mathcal{N}(X_n + \tfrac{\tau}{2}\nabla \log \rho(X_n), \tau\,\mathrm{Id}), \qquad n \geq 0. \tag{2.8}$$

Anschaulich schiebt der deterministische Term $\frac{\tau}{2}\nabla \log \rho(X_n)$ den Prozess in Richtung von lokalen Maxima der (logarithmischen) Dichte ρ, um diese Bereiche (die in der Monte-Carlo-Approximation mit hohem Gewicht eingehen) bevorzugt zu besuchen. Der stochastische Term $\sqrt{\tau}\xi \sim \mathcal{N}(0, \tau\,\mathrm{Id})$ sorgt dagegen dafür, dass der Prozess dort nicht stecken bleibt und (aymptotisch!) den ganzen $\mathbb{R}^d$ abtastet. Die Kopplung von Varianz und deterministischer Schrittweite ist hier kritisch und macht den wesentlichen Unterschied zu klassischen MCMC-Samplern aus.

Leider ist – genau wie das explizite Euler-Verfahren, dessen stochastische Variante das Euler-Maruyama-Verfahren ist – der ULA nicht absolut stabil und konvergiert für die meisten Schrittweiten nicht. Er wird daher in ein Metropolis–Hastings-Verfahren eingebettet: Für $n \geq 0$,

1. erzeuge einen *Kandidaten*

$$Y_{n+1} \sim \mathcal{N}(X_n + \tfrac{\tau}{2}\nabla \log \rho(X_n), \tau\,\mathrm{Id});$$

2. bestimme das *Akzeptanzverhältnis*

$$\alpha(Y_{n+1}, X_n) = \begin{cases} \min\left\{1, \frac{\rho(Y_{n+1})\eta(Y_{n+1};X_n)}{\rho(X_n)\eta(X_n;Y_{n+1})}\right\} & \text{falls } \rho(X_n)\eta(Y_{n+1};X_n) > 0, \\ 1 & \text{sonst;} \end{cases}$$

3. wähle $\tau \sim \mathcal{U}([0,1])$ und setze

$$X_{n+1} = \begin{cases} Y_{n+1} & \text{falls } \tau \le \alpha(Y_{n+1}, X_n), \\ X_n & \text{sonst.} \end{cases}$$

Hier ist $\eta(y;x)$ die Dichte der Übergangsverteilung von X zu Y; wegen (2.8) ist dies der Gauß-Kern

$$\eta(y;x) = \frac{1}{(2\pi\tau)^{d/2}} e^{-\frac{1}{2\tau}\|y-\left(x+\frac{\tau}{2}\nabla\log\rho(x)\right)\|_2^2}.$$

Schritte werden also nur dann sicher akzeptiert, wenn sie die Wahrscheinlichkeit (Dichte) erhöhen; ansonsten werden sie mit desto höherer Wahrscheinlichkeit verworfen, desto unwahrscheinlicher sie sind. (Beachte, dass Y_{n+1} wegen $\xi \sim \mathcal{N}(0, \mathrm{Id})$ zufällig ist und sich daher selbst bei gleichbleibendem X_n ändern wird!) Dies ist der *Metropolis-adjusted Langevin algorithm (MALA)*.

Unter weiteren technischen (und in der Praxis schwer nachprüfbaren) Annahmen an ρ kann man zeigen, dass X_n (sogar geometrisch) ergodisch konvergiert und daher asymptotisch nach der Dichte ρ verteilt ist, siehe [Roberts & Tweedie 1996, Theorem 4.1]. Da dies nur asymptotisch gilt, ignoriert man die ersten $M \gg 1$ Schritte („*burn-in*“) und wählt als Stichproben $\xi_i = X_{M+i}$, $i = 1, \dots, N$.

Schließlich ist noch anzumerken, dass die Konvergenz desto langsamer ist, je schneller $\nabla \log \pi(x) \to 0$ für $\|x\| \to \infty$ geht. Umgekehrt führt ein großer Wert von $\nabla \log \pi(x)$ in einem Bereich dazu, dass X_n diesen Bereich schnell verlässt und damit nicht ausreichend abtastet. Dies wird durch den *Metropolis-adjusted Langevin truncated algorithm (MALTA)* vermieden, der die Kandidaten

$$Y_{n+1} \sim \mathcal{N}\left(X_n + \frac{\tau\nabla\log\rho(X_n)}{\max\{\tau, \|\nabla\log\rho(X_n)\|_2\}}, \tau\,\mathrm{Id}\right).$$

(mit entsprechend angepasster Übergangsdichte η) erzeugt. Hier bleibt der Schritt gleichmäßig beschränkt, so dass auch Bereiche kleinerer Wahrscheinlichkeit hinreichend schnell abgetastet werden.

3 STOCHASTISCHE GRADIENTENVERFAHREN

Die Monte-Carlo-Approximation von Erwartungswertfunktionalen ist insbesondere motiviert durch die numerische Lösung: Ist $F(x) = \mathbb{E}[f(x,\xi)]$ und $F_N = \frac{1}{N}\sum_{i=1}^N f(x,\xi_i)$ für unabhängig und identisch verteilte Stichproben $\xi_1,\dots,\xi_N$, so gilt für $x \in \bigcap_{i=1}^N \operatorname{dom} f(\cdot;\xi_i)^o$ nach der Summenregel

$$\partial F_N(x) = \left\{ \frac{1}{N}\sum_{i=1}^N x_i^* \,\middle|\, x_i^* \in \partial f(x;\xi_i) \right\}.$$

Alternativ kann man $x_i^* \in \partial f(x,\xi_i)$ auch als Stichprobe von $\partial F(x)$ auffassen, die in jeder Iteration des Subgradientenverfahrens neu gezogen wird. Dies führt auf das *stochastische Subgradientenverfahren.*

Anders als in der Theorie ist es hier vorteilhafter, etwaige Nebenbedingungen separat zu behandeln; wir betrachten daher für $X \subset \mathbb{R}^d$ nichtleer, konvex, und abgeschlossen, und $f : \mathbb{R}^d \times \Omega \to \mathbb{R}$ fast sicher konvex und unterhalbstetig das Problem

$$\min_{x\in X} \mathbb{E}[f(x,\xi)].$$

Die Idee ist nun, in jeder Iteration eine Stichprobe aus dem Subgradienten zu wählen, diesen als Suchrichtung zu verwenden, und dann auf die zulässige Menge zu projizieren:

Algorithmus 3.1 : Stochastisches Subgradientenverfahren

1 Wähle einen Startpunkt $x^0 \in X$
2 **for** $k = 0,\dots$ **do**
3 Wähle Stichprobe $\xi_k \sim \xi$
4 Wähle Subgradient $d^k \in \partial f(x^k;\xi_k)$
5 Wähle Schrittweite $t_k > 0$
6 Setze $x^{k+1} = \operatorname{proj}_X(x^k - t_k d^k)$

Ist f zusätzlich fast sicher differenzierbar, so ist natürlich $d^k = \nabla f(x^k;\xi_k)$; man spricht dann vom *stochastischen Gradientenverfahren.*[1]

[1] Ein ähnlicher Ansatz unter dem selben Namen kann angewendet werden auf $F(x) = \sum_{i=1}^N f_i(x)$ mit $d^k = \nabla f_{i(k)}(x^k)$ für $i(k) \in \{1,\dots,N\}$ zufällig gewählt.

Wir untersuchen nun die Konvergenzeigenschaften dieses Verfahrens. Dazu zeigen wir zuerst, dass die Projektion auf eine konvexe Menge *nichtexpansiv* ist. Zur Erinnerung: Die Projektion von x auf eine nichtleere, konvexe, und abgeschlossene Menge ist definiert als

$$\operatorname{proj}_X(z) := \arg\min_{x \in X} \|x - z\|_2^2,$$

charakterisiert durch die Variationsungleichung

$$(\operatorname{proj}_X(z) - z)^T(\tilde{x} - \operatorname{proj}_X(z)) \geq 0 \qquad \text{für alle } \tilde{x} \in X. \tag{3.1}$$

Lemma 3.1. *Sei $X \subset \mathbb{R}^d$ nichtleer, konvex, und abgeschlossen. Dann gilt*

$$\|\operatorname{proj}_X(x) - \operatorname{proj}_X(y)\|_2 \leq \|x - y\|_2 \qquad \textit{für alle } x, y \in \mathbb{R}^d.$$

Beweis. Für $x, y \in \mathbb{R}^d$ können wir mit Hilfe der produktiven Null schreiben

$$x - y = \operatorname{proj}_X(x) - \operatorname{proj}_X(y) + (x - \operatorname{proj}_X(x)) + (\operatorname{proj}_X(y) - y).$$

Aus dem Satz von Pythagoras folgt dann

$$\begin{aligned}\|x - y\|_2^2 = \|\operatorname{proj}_X(x) - \operatorname{proj}_X(y)\|_2^2 &+ \|(x - \operatorname{proj}_X(x)) + (\operatorname{proj}_X(y) - y)\|_2^2 \\ &+ 2(x - \operatorname{proj}_X(x))^T(\operatorname{proj}_X(x) - \operatorname{proj}_X(y)) \\ &+ 2(\operatorname{proj}_X(y) - y)^T(\operatorname{proj}_X(x) - \operatorname{proj}_X(y)).\end{aligned}$$

Aus (3.1) folgt aber für $z = x$ und $\tilde{x} = \operatorname{proj}_X(y) \in X$ bzw. $z = y$ und $\tilde{x} = \operatorname{proj}_X(x) \in X$

$$\begin{aligned}(\operatorname{proj}_X(x) - x)^T(\operatorname{proj}_X(y) - \operatorname{proj}_X(x)) \geq 0, \\ (\operatorname{proj}_X(y) - y)^T(\operatorname{proj}_X(x) - \operatorname{proj}_X(y)) \geq 0,\end{aligned}$$

und damit

$$\|x - y\|_2^2 \geq \|\operatorname{proj}_X(x) - \operatorname{proj}_X(y)\|_2^2. \qquad \square$$

Der Preis, den wir für die Verwendung von Stichproben zahlen müssen, ist, dass wir Konvergenz nicht für die Iterierten selber, sondern nur für das laufende gewichtete Mittel

$$\bar{x}_\gamma^k := \sum_{j=0}^{k} \gamma_j x^j, \qquad \gamma_1, \ldots, \gamma_k > 0$$

(welches ebenfalls eine Zufallsvariable ist) zeigen können; man spricht dann von *ergodischer Konvergenz*. Wir brauchen weiter eine gleichmäßige Beschränkung der Subgradienten. Da die Suchrichtungen zufällig sind, ist auch die Folge $\{x^k\}_{k \in \mathbb{N}}$ der erzeugten Iterierten eine zufällige Folge, die von der Wahl aller vorhergehenden Stichproben $\xi_0, \ldots, \xi_{k-1}$ abhängt. Bezeichne wieder $x^* \in X$ einen Minimierer von F über X und F^* den minimalen Funktionswert.

Satz 3.2. *Angenommen, für $f : \mathbb{R}^d \times \Omega \to \mathbb{R}$ gilt*

(i) für fast alle $\omega \in \Omega$ ist $x \mapsto f(x, \omega)$ konvex und unterhalbstetig;

(ii) es existiert ein $g \in L^1(\Omega)$ mit $f(x, \omega) \geq -g(\omega)$ für fast alle $\omega \in \Omega$ und alle $x \in \mathbb{R}^d$.

Sei $X \subset \mathbb{R}^d$ nichtleer, konvex, und abgeschlossen, und es existiere ein $M > 0$ mit

$$\sup_{x \in X} \left\{ \mathbb{E}\left[\|x^*\|_2^2 \right] \,\middle|\, x^*(\omega) \in \partial f(x; \omega) \text{ für fast alle } \omega \in \Omega \right\} \leq M^2.$$

Dann gilt für alle $k \in \mathbb{N}$ und $\gamma_j := \frac{t_j}{\sum_{i=0}^k t_i}$, $j = 0, \ldots, k$,

$$\mathbb{E}\left[F(\bar{x}_\gamma^k) - F^* \right] \leq \frac{\|x^0 - x^*\|_2^2 + M^2 \sum_{j=0}^k t_j^2}{2 \sum_{j=0}^k t_j}.$$

Beweis. Wir halten zunächst fest, dass unter den Annahmen an f nach Satz 1.6 gilt $\mathbb{E}[z^*] \in \partial F(x)$ für $z^*(\omega) \in \partial f(x; \omega)$ fast überall.

Wir betrachten nun einen einzelnen Iterationsschritt. Aus Algorithmus 3.1 und Lemma 3.1 folgt wegen $x^* \in X$

$$\begin{aligned} \|x^{k+1} - x^*\|_2^2 &= \|\mathrm{proj}_X(x^k - t_k d^k) - \mathrm{proj}_X(x^*)\|_2^2 \\ &\leq \|x^k - x^* - t_k d^k\|_2^2 \\ &= \|x^k - x^*\|_2^2 - 2t_k (x^k - x^*)^T d^k + t_k^2 \|d^k\|_2^2. \end{aligned}$$

Auf diese Abschätzung wenden wir nun den Erwartungswert an. Für den letzten Term gilt nach Annahme und der Wahl $d^k \in \partial f(x^k; \xi_k)$ mit $\xi_k \in \Omega$

$$t_k^2 \, \mathbb{E}[\|d^k\|_2^2] \leq t_k^2 M^2.$$

Der zweite Term ist der kritische. Die Idee ist, dass die zufällige Suchrichtung d^k zwar von allen bisherigen Stichproben auf komplizierte Weise $\xi_0, \ldots, \xi_k$ abhängt; fixiert man jedoch eine konkrete Wahl von $\xi_0, \ldots, \xi_{k-1}$, verhält sich $d^k \in \partial f(x^k; \xi_k)$ wegen $\xi_k \sim \xi$ wie eine unabhängige Stichprobe eines vollen Subgradienten. Um dies zu formalisieren, verwenden wir bedingte Erwartungswerte. Sei dafür $\{\mathcal{F}_k\}_{k \in \mathbb{N}}$ die natürliche Filtration bezüglich der in Algorithmus 3.1 gewählten Folge $\{\xi_k\}_{k \in \mathbb{N}}$, so dass insbesondere ξ_k und damit auch $d^k \in \partial f(x^k; \xi_k)$ nach Satz 1.6 sowie x^{k+1} messbar sind bezüglich $\mathcal{F}_j$ für alle $j \leq k$. Da x^k unabhängig ist von ξ_k gilt dann

$$\begin{aligned} \mathbb{E}\left[(x^k - x^*)^T d^k \right] &= \mathbb{E}\left[\mathbb{E}\left[(x^k - x^*)^T d^k \mid \mathcal{F}_k \right] \right] \\ &= \mathbb{E}\left[(x^k - x^*)^T \, \mathbb{E}\left[d^k \mid \mathcal{F}_k \right] \right] \\ &= \mathbb{E}\left[(x^k - x^*)^T \, \mathbb{E}[d^k] \right]. \end{aligned}$$

Nach Satz 1.6 gilt nun $Z^* := \mathbb{E}[d^k] \in \partial F(x^k)$. Also folgt

$$\mathbb{E}[\|x^{k+1} - x^*\|_2^2] \leq \mathbb{E}[\|x^k - x^*\|_2^2] - 2t_k \,\mathbb{E}[(x^k - x^*)^T Z^*] + t_k^2 M^2.$$

Wir verwenden nun, dass F nach Lemma 1.1 konvex ist und daher gilt

$$\begin{aligned} t_k \,\mathbb{E}\left[F(x^k) - F(x^*)\right] &\leq t_k \,\mathbb{E}\left[(x^k - x^*)^T Z^*\right] \\ &\leq \frac{1}{2}\,\mathbb{E}\left[\|x^k - x^*\|_2^2\right] - \frac{1}{2}\,\mathbb{E}\left[\|x^{k+1} - x^*\|_2^2\right] + \frac{1}{2} t_k^2 M^2. \end{aligned}$$

Summieren dieser Ungleichung von $j = 0, \ldots, k$ und Verwenden der Teleskopsumme ergibt dann

$$\begin{aligned} \sum_{j=0}^{k} t_j \,\mathbb{E}\left[F(x^j) - F(x^*)\right] &\leq \frac{1}{2}\,\mathbb{E}\left[\|x^0 - x^*\|_2^2\right] - \frac{1}{2}\,\mathbb{E}\left[\|x^{k+1} - x^*\|_2^2\right] + \frac{1}{2}\sum_{j=0}^{k} t_j^2 M^2 \\ &\leq \frac{1}{2}\|x^0 - x^*\|_2^2 + \frac{1}{2}\sum_{j=0}^{k} t_j^2 M^2 \end{aligned}$$

da der Startwert x^0 nicht zufällig ist. Setze nun $\gamma_j := \frac{t_j}{\sum_{i=0}^{k} t_i}$ für $j = 0, \ldots, k$. Dann ist $\gamma_j \in (0,1)$ und $\sum_{j=0}^{k} \gamma_j = 1$. Wegen der Zulässigkeit der x^k durch die Projektion folgt aus der Konvexität von X nun $\bar{x}_\gamma^k := \sum_{j=0}^{k} \gamma_j x^j \in X$, und die Konvexität von F ergibt $F(\bar{x}_\gamma^k) \leq \sum_{j=0}^{k} \gamma_j F(x^j)$. Zusammen mit der Monotonie des Erwartungswerts erhalten wir

$$\mathbb{E}\left[F(\bar{x}_\gamma^k) - F(x^*)\right] \leq \sum_{j=0}^{k} \gamma_j \,\mathbb{E}\left[F(x^j) - F(x^*)\right] \leq \frac{\|x^0 - x^*\|_2^2 + M^2 \sum_{j=0}^{k} t_j^2}{2\sum_{j=0}^{k} t_j}$$

und damit die Behauptung. □

Beachte, dass wir die Beschränktheitsannahme nur für die tatsächlich verwendeten Abstiegsrichtungen $\{d^k\}_{k\in\mathbb{N}}$ verwendet haben.

Daraus folgt sofort die (ergodische) Konvergenz unter der üblichen Annahme, dass die Schrittweiten hinreichend langsam gegen Null gehen.

Folgerung 3.3. *Erfüllen die Schrittweiten $\sum_{k\in\mathbb{N}} t_k = \infty$ und $\sum_{k\in\mathbb{N}} t_k^2 < \infty$, so gilt*

$$F(\bar{x}_\gamma^k) \to F^* \qquad \text{für } k \to \infty.$$

Die Voraussetzungen sind z. B. erfüllt für die typische Wahl $t_k = \frac{1}{k+1}$. Diese Konvergenz kann aber wegen der schnell verschwindenden Schrittweiten beliebig langsam sein. Mit einer etwas langsameren Wahl erhalten wir auch eine Konvergenzrate.

Folgerung 3.4. *Für die Schrittweite* $t_k = \frac{\|x^0 - x^*\|_2 M^{-1}}{\sqrt{k+1}}$ *gilt*

$$\mathbb{E}\left[F(\bar{x}_\gamma^k) - F(x^*)\right] \leq M\|x^0 - x^*\|_2 \frac{2 + \log(k)}{2\sqrt{k}}.$$

Beweis. Mit dieser Wahl der Schrittweite vereinfacht sich die Abschätzung aus Satz 3.2 zu

$$\mathbb{E}\left[F(\bar{x}_\gamma^k) - F^*\right] \leq \|x^0 - x^*\|_2 M \frac{1 + \sum_{j=0}^k \frac{1}{j+1}}{2\sum_{j=0}^k \frac{1}{\sqrt{j+1}}}.$$

Wir schätzen nun den Nenner ab durch

$$\sum_{j=0}^k \frac{1}{\sqrt{j+1}} \geq \sum_{j=0}^k \frac{1}{\sqrt{k+1}} = \frac{k+1}{\sqrt{k+1}} = \sqrt{k+1} \geq \sqrt{k}.$$

Für den Zähler verwenden wir die Euler–Maclaurin-Formel: Da $x \mapsto \frac{1}{x}$ monoton fallend ist auf $[1, \infty)$, gilt

$$\sum_{j=0}^k \frac{1}{j+1} = 1 + \sum_{j=1}^k \frac{1}{j+1} \leq 1 + \int_1^k \frac{1}{x}\, dx = 1 + \log(k). \qquad \square$$

Dies ist zwar langsamer als $\frac{1}{\sqrt{k}}$ aber immer noch schneller als $\frac{1}{k^s}$ für alle $s < \frac{1}{2}$.

Wir können sogar konstante Schrittweiten wählen, wenn wir nur eine endliche Zahl von Schritten $k = 1, \ldots, K$ für $K < \infty$ (eine *Epoche*) betrachten; natürlich können wir danach das Verfahren mit $x^0 = \bar{x}_\gamma^K$ beliebig oft neu starten.

In diesem Fall ist für $t_k = t$ und $k = K$

$$\mathbb{E}\left[F(\bar{x}_\gamma^K) - F(x^*)\right] \leq \frac{\|x^0 - x^*\|_2^2 + M^2(K+1)t^2}{2(K+1)t},$$

und eine einfache Rechnung zeigt, dass die rechte Seite minimal ist für

$$t = \frac{\|x^0 - x^*\|_2}{M\sqrt{K+1}}$$

mit entsprechender Fehlerabschätzung

$$\mathbb{E}\left[F(\bar{x}_\gamma^K) - F(x^*)\right] \leq \frac{\|x^0 - x^*\|_2 M}{\sqrt{K+1}}.$$

Dann vereinfacht sich auch das ergodische Gewicht zu

$$\gamma_j = \frac{t}{(K+1)t} = \frac{1}{K+1}, \qquad j = 0, \ldots, K,$$

so dass gilt

$$\bar{x}^K := \bar{x}_\gamma^K = \frac{1}{K+1}\sum_{k=0}^{K} x^k.$$

Durch Induktion zeigt man dann leicht, dass sich $\bar{x}^K$ auch rekursiv berechnen lässt als

$$\bar{x}^{k+1} = \frac{1}{k+1}x^{k+1} + \left(1 - \frac{1}{k+1}\right)\bar{x}^k,$$

es ist also nicht nötig, die Iterierten zwischenzuspeichern.

Diese optimale Schrittweite ist in der Praxis natürlich nicht anwendbar; allerdings ist die Fehlerabschätzung robust in dem Sinne, dass eine Multiplikation von t mit einer Konstanten nur den Vorfaktor aber nicht die Rate beeinflusst. Wir halten daher fest:

Folgerung 3.5. *Für festes $K \in \mathbb{N}$ und konstante Schrittweiten $t_k = \frac{\tau}{\sqrt{K+1}}$ mit $\tau > 0$ gilt für das empirische Mittel $\bar{x}^K := \frac{1}{K+1}\sum_{k=0}^{K} x^k$*

$$\mathbb{E}\left[F(\bar{x}^K) - F(x^*)\right] \leq \tau\frac{\|x^0 - x^*\|^2 + M^2}{2\sqrt{K+1}}.$$

Mit Hilfe der Chernov-Ungleichung erhält man daraus (für $\tau = 1$) eine Fehlerabschätzung der Form

$$\mathbb{P}[F(\bar{x}^K) - F^* \geq \varepsilon] \leq \frac{\mathbb{E}[F(\bar{x}^K) - F^*]}{\varepsilon} \leq \frac{\|x^0 - x^*\|_2^2 + M^2}{\varepsilon\sqrt{K+1}}.$$

Um also mit Wahrscheinlichkeit $1 - p$ ein ε-Minimum zu erreichen, muss die Epoche mindestens

$$K \geq \frac{(\|x^0 - x^*\|^2 + M^2)^2}{\varepsilon^2 p^2},$$

d. h. $O(p^{-2})$ Iterationen enthalten. Dies ist zwar deutlich mehr als die $O\left(\log(p^{-1})\right)$ für die Monte-Carlo-Approximation aus Folgerung 2.10; allerdings kann man mit dem entsprechenden Mehraufwand unter Verwendung von Konzentrationsungleichungen ähnliche Abschätzungen erhalten. In jedem Fall ist der Aufwand für festes $K = N$ im stochastischen Subgradientenverfahren erheblich geringer.

Beide Ideen können auch verbunden werden: Statt einer einzelnen Stichprobe bildet man das empirische Mittel von mehreren unabhängigen Stichproben, d. h. verwendet *Batch-Subgradienten*

$$d^k = \frac{1}{N(k)}\sum_{i=1}^{N(k)} d_i^k, \qquad d_i^k \in \partial f(x^k; \xi_i),$$

für $\xi_1, \ldots, \xi_{N(k)} \sim \xi$ unabhängig und identisch verteilt. Der Beweis von Satz 3.2 kann mit leichten Modifikationen auf diesen Fall verallgemeinert werden. Die Mittelung reduziert die Varianz des Schätzers von $Z^* \in \partial F(x^k)$ und kann dadurch die Konvergenzgeschwindigkeit (durch Verkleinerung der Konstanten) verbessern.

4 ROBUSTE OPTIMIERUNG

Die Minimierung von Erwartungswertfunktionalen hat aus Anwendungssicht den erheblichen Nachteil, dass extreme Ereignisse ω (mit sehr – und zu – hohem Funktionswert $f(x, \omega)$) bei der Optimierung hingenommen werden, wenn sie nur unwahrscheinlich genug sind. In der *robusten Optimierung* will man zusätzlich das Risiko, dass extreme Ereignisse auftauchen, minimieren.

Um den Zugang zu motivieren, nehmen wir der Einfachheit halber an, dass wir positive Funktionswerte verhindern wollen, d. h. garantieren wollen, dass die *fast sichere Nebenbedingung*

$$f(x, \xi) \leq 0 \qquad \text{für fast alle } \xi \in \Omega$$

erfüllt ist. Das ist in der Regel nicht möglich; wir können dann stattdessen versuchen für eine (sehr kleine) Wahrscheinlichkeit $\alpha > 0$ die *probabilistische Nebenbedingung*

$$\mathbb{P}[f(x, \xi) > 0] \leq \alpha$$

zu erfüllen. Wir ignorieren in Folge der Übersichtlichkeit die Abhängigkeit von der Entscheidungsvariablen $x \in \mathbb{R}^d$ und betrachten die Zufallsvariable $Z := f(x, \xi)$. Die Schwierigkeit hier ist, dass $\mathbb{P}[Z > t] = \mathbb{E}[\mathbb{1}_{(t,\infty]}(Z)]$ nicht konvex und unterhalbstetig ist, da die charakteristische Funktion nicht konvex ist; also definiert die probabilistische Nebenbedingung im allgemeinen keine konvexe zulässige Menge.

Wir suchen daher eine konvexe Reformulierung. Dafür definieren wir zunächst

$$\mathrm{VaR}_\alpha[Z] := \inf\{t > 0 \mid \mathbb{P}[Z \leq t] \geq 1 - \alpha\} = \inf\{t > 0 \mid \mathbb{P}[Z > t] \leq \alpha\}.$$

Anschaulich ist der *Value-at-Risk* $\mathrm{VaR}_\alpha(Z)$ der größte Verlust, der mit Wahrscheinlichkeit $1 - \alpha$ eintreffen kann; größere Verluste haben also Wahrscheinlichkeit kleiner als α. Es gilt dann $\mathbb{P}[Z > 0] \leq \alpha$ genau dann, wenn $\mathrm{VaR}_\alpha[Z] \leq 0$ ist, d. h. ein Verlust tritt mit Wahrscheinlichkeit höchstens α ein. Beachte, dass der Value-at-Risk im Gegensatz zur Wahrscheinlichkeit direkt im Wertebereich der Zufallsvariablen formuliert ist; dies ist für die Minimierung vorteilhafter.

Leider ist der Value-at-Risk immer noch nicht konvex. Die Idee ist nun, eine konvexe Majorante G zu konstruieren, so dass $G(Z) \leq 0$ auch $\mathrm{VaR}_\alpha(Z) \leq 0$ impliziert. Dafür ersetzen

wir die charakteristische Funktion durch eine geeignete stückweise lineare Funktion. Durch Fallunterscheidung sieht man leicht, dass für jedes $\gamma > 0$ gilt

$$\mathbb{P}[Z > 0] = \mathbb{E}\left[\mathbb{1}_{(0,\infty)}(Z)\right] \leq \mathbb{E}\left[\max\{0, 1 + \gamma Z\}\right] = \gamma\, \mathbb{E}\left[\max\{0, \gamma^{-1} + Z\}\right].$$

Gilt also

$$\inf_{\gamma>0} \gamma\, \mathbb{E}\left[\max\{0, \gamma^{-1} + Z\}\right] - \alpha \leq 0$$

so folgt daraus $\mathbb{P}[Z > 0] \leq \alpha$ und damit auch $\mathrm{VaR}_\alpha[Z] \leq 0$, da das Infimum dann nicht größer als 0 sein kann.

Wir können diese Majorante noch etwas umformulieren. Division durch $\alpha, \gamma > 0$ und Setzen von $t := -\gamma^{-1} < 0$ führt auf die äquivalente Bedingung

$$\inf_{t<0} t + \alpha^{-1}\, \mathbb{E}\left[\max\{0, Z - t\}\right] \leq 0.$$

Ist $t > 0$, dann ist auch $t + \alpha^{-1}\, \mathbb{E}\left[\max\{0, Z - t\}\right] \geq t > 0$. Damit ist unsere konvexe Majorante gegeben durch den *Conditional Value-at-Risk* (auch *Average Value-at-Risk*)

$$\mathrm{CVaR}_\alpha[Z] := \inf_{t\in\mathbb{R}} t + \alpha^{-1}\, \mathbb{E}\left[\max\{0, Z - t\}\right] \leq 0. \tag{4.1}$$

(Für $\alpha \to 0$ erhält man formal die fast sichere Nebenbedingung $Z(\omega) \leq 0$ wieder.)

Eine Abbildung, die einer Zufallsvariable solch eine Risikobewertung zuordnet, nennt man naheliegenderweise *Risikomaß*. Anstelle einer harten Nebenbedingung kann man ein Risikomaß G auch als Straffunktion verwenden und für $\gamma > 0$ ein *Mittelwert-Risiko-Modell* der Form

$$\min_{x\in\mathbb{R}^d} \mathbb{E}[f(x, \xi)] + \gamma G[f(x, \xi)]$$

betrachten. Man spricht dann auch von *risikoaverser Optimierung* (und entsprechend für $\gamma = 0$ von *risikoneutraler Optimierung*).

4.1 KONVEXE RISIKOMASSE

Wir betrachten zuerst eine allgemeine Klasse von Risikomaßen, die für die risikoaverse Optimierung vorteilhaft sind. Sei dafür $(\Omega, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum und $Z \in L^p(\Omega)$ für $1 \leq p < \infty$. Wir versehen $L^p(\Omega)$ mit der Halbordnung $Z \succeq Z'$, falls $Z(\omega) \geq Z'(\omega)$ für fast alle $\omega \in \Omega$ gilt. (Man nennt dies auch *stochastische Dominanz 1. Ordnung*.)

Sei nun $\mathcal{R} : L^p(\Omega) \to \overline{\mathbb{R}}$ eigentlich. Wir nennen dann $\mathcal{R}$ ein *konvexes Risikomaß*, falls gilt

(R1) $\mathcal{R}$ ist konvex;

(R2) $\mathcal{R}$ ist monoton, d. h. $\mathcal{R}(Z) \geq \mathcal{R}(Z')$ falls $Z \succeq Z'$;

(R3) $\mathcal{R}$ ist translationsinvariant, d. h. für alle $a \in \mathbb{R}$ und $Z \in L^p(\Omega)$ gilt $\mathcal{R}(Z+a) = \mathcal{R}(Z)+a$.

Gilt zusätzlich

(R4) $\mathcal{R}$ ist positiv homogen, d h. $\mathcal{R}(tZ) = t\mathcal{R}(Z)$ für alle $t > 0$ und $Z \in L^p(\Omega)$,

dann heißt $\mathcal{R}$ ein *kohärentes Risikomaß*. Offensichtlich ist $\mathcal{R}(Z) = \mathbb{E}[Z]$ ein kohärentes Riskomaß auf $L^1(\Omega)$.

Aus der Definition folgen einige intuitiv wünschenswerte Eigenschaften. Aus (R4) und (R1) folgt zunächst sofort

$$\frac{1}{2}\mathcal{R}(Z+Z') = \mathcal{R}\left(\tfrac{1}{2}Z + \tfrac{1}{2}Z'\right) \leq \frac{1}{2}\mathcal{R}(Z) + \frac{1}{2}\mathcal{R}(Z')$$

für alle $Z, Z' \in L^p(\Omega)$, d. h. ein kohärentes Risikomaß ist *subadditiv*. Aus (R4) folgt auch

$$\mathcal{R}(0) = \mathcal{R}(t0) = t\mathcal{R}(0) \qquad \text{für alle } t > 0,$$

was $\mathcal{R}(0) = 0$ impliziert. Daraus folgt zusammen mit (R3) weiter

$$\mathcal{R}(a) = \mathcal{R}(0) + a = a \qquad \text{für alle } a \in \mathbb{R},$$

d. h. das Risik einer im Voraus bekannten (nicht zufälligen) Größe ist genau durch die Größe gegeben. Gilt

(R5) $\mathcal{R}(Z) > \mathbb{E}[Z]$ für alle $Z \in L^p(\Omega)$ nicht konstant,

so nennt man $\mathcal{R}$ *risikoavers*. (Ein konvexes risikoaverses Risikomaß wird auch *reguläres Risikomaß* gennannt.)

Ein Vorteil von kohärenten Risikomaßen ist, dass sie eine besonders schöne duale Charakterisierung haben. Wir nehmen in Folge an, dass $\mathcal{R} : L^p(\Omega)$ ein eigentliches und unterhalbstetiges konvexes Riskomaß ist.[1] Aus Satz A.18 folgt dann für $1 < q \leq \infty$ mit $\frac{1}{q} + \frac{1}{p} = 1$

$$\begin{aligned} \mathcal{R}(Z) = \mathcal{R}^{**}(Z) &= \sup_{Z^* \in L^q(\Omega)} \langle Z, Z^* \rangle_{L^p(\Omega)} - \mathcal{R}^*(Z^*) \\ &= \sup_{Z^* \in \operatorname{dom} \mathcal{R}^*} \langle Z, Z^* \rangle_{L^p(\Omega)} - \mathcal{R}^*(Z^*). \end{aligned} \tag{4.2}$$

Die Menge $\Pi := \operatorname{dom} \mathcal{R}^*$ nennt man in diesem Zusammenhang auch *Risikohülle* (aus Gründen, die später klar werden). Die Eigenschaften eines Risikomaßes können dann durch die Risikohülle charakterisiert werden. Wir erinnern, dass gilt

$$\langle Z^*, Z \rangle_{L^p(\Omega)} = \int_\Omega Z^*(\omega) Z(\omega)\, d\mu(\omega) = \mathbb{E}[Z^* Z] \tag{4.3}$$

für alle $Z^* \in L^q(\Omega)$ und $Z \in L^p(\Omega)$.

[1] Die Voraussetzung der Unterhalbstetigkeit kann für einige der folgenden Aussagen fallengelassen werden, da sie zusammen mit der Konvexität aus der Monotonie folgt.

Satz 4.1. *Sei $\mathcal{R} : L^p(\Omega) \to \overline{\mathbb{R}}$ eigentlich, konvex, und unterhalbstetig und $\Pi := \operatorname{dom} \mathcal{R}^*$. Dann ist*

(i) $\mathcal{R}$ monoton genau dann, wenn $Z^* \succeq 0$ für alle $Z^* \in \Pi$ gilt;

(ii) $\mathcal{R}$ translationsinvariant genau dann, wenn $\mathbb{E}[Z^*] = 1$ für alle $Z^* \in \Pi$ gilt;

(iii) $\mathcal{R}$ positiv homogen genau dann, wenn $\mathcal{R}(Z) = \sup_{Z^* \in \Pi} \langle Z^*, Z \rangle_{L^p(\Omega)}$ für alle $Z \in L^p(\Omega)$ gilt (d. h. $\mathcal{R}$ das Trägerfunktional von Π ist).

Beweis. Zu (i): Angenommen, $\mathcal{R}$ ist monoton. Sei $Z^* \in L^q(\Omega)$ so, dass für eine Menge $A \in \mathcal{A}$ gilt $Z^*(\omega) < 0$ für fast alle $\omega \in A$. Dann ist für $\mathbb{1}_A \in L^p(\Omega)$

$$\langle Z^*, \mathbb{1}_A \rangle_{L^p(\Omega)} = \int_A Z^*(\omega)\, d\mu(\omega) < 0.$$

Wähle nun $Z \in \operatorname{dom} \mathcal{R}$ und betrachte $Z_t := Z - t\mathbb{1}_A \in L^p(\Omega)$ für $t \geq 0$. Dann gilt $Z \succeq Z_t$ und damit wegen der Monotonie von $\mathcal{R}$ auch $\mathcal{R}(Z) \geq \mathcal{R}(Z_t)$. Daraus folgt

$$\begin{aligned} \mathcal{R}^*(Z^*) = \sup_{Z \in L^p(\Omega)} \langle Z^*, Z \rangle_{L^p(\Omega)} - \mathcal{R}(Z) &\geq \sup_{t>0} \langle Z^*, Z_t \rangle_{L^p(\Omega)} - \mathcal{R}(Z_t) \\ &\geq \sup_{t>0} \langle Z^*, Z \rangle_{L^p(\Omega)} - t\langle Z^*, \mathbb{1}_A \rangle_{L^p(\Omega)} - \mathcal{R}(Z) = \infty, \end{aligned}$$

d. h. $Z^* \notin \operatorname{dom} \mathcal{R}^* = \Pi$.

Sei umgekehrt $Z^* \succeq 0$ für alle $Z^* \in \Pi$. Dann gilt für alle $Z^* \in \Pi$ und $Z, Z' \in L^p(\Omega)$ mit $Z \succeq Z'$ auch

$$\langle Z^*, Z \rangle_{L^p(\Omega)} = \int_\Omega Z^*(\omega) Z(\omega)\, d\mu(\omega) \geq \int_\Omega Z^*(\omega) Z'(\omega)\, d\mu(\omega) = \langle Z^*, Z' \rangle_{L^p(\Omega)}.$$

Aus (4.2) folgt dann

$$\mathcal{R}(Z) = \sup_{Z^* \in \Pi} \langle Z^*, Z \rangle_{L^p(\Omega)} - \mathcal{R}^*(Z^*) \geq \sup_{Z^* \in \Pi} \langle Z^*, Z' \rangle_{L^p(\Omega)} - \mathcal{R}^*(Z^*) = \mathcal{R}(Z').$$

Zu (ii): Angenommen, $\mathcal{R}$ ist translationsinvariant. Sei $Z^* \in \Pi$ so, dass $\mathbb{E}[Z^*] \neq 1$ gilt. Dann folgt für alle $Z \in \operatorname{dom} \mathcal{R}$ aus der Translationsinvarianz

$$\begin{aligned} \mathcal{R}^*(Z^*) &\geq \sup_{a \in \mathbb{R}} \langle Z^*, Z + a \rangle_{L^p(\Omega)} - \mathcal{R}(Z + a) \\ &= \sup_{a \in \mathbb{R}} \langle Z^*, Z \rangle_{L^p(\Omega)} + a\, \mathbb{E}[Z^*] - \mathcal{R}(Z) - a \\ &= \sup_{a \in \mathbb{R}} a\, (\mathbb{E}[Z^*] - 1) + \langle Z^*, Z \rangle_{L^p(\Omega)} - \mathcal{R}(Z) = \infty \end{aligned}$$

und damit $Z^* \notin \operatorname{dom} \mathcal{R}^* = \Pi$.

Sei umgekehrt $\mathbb{E}[Z^*] = 1$ für alle $Z^* \in \Pi$. Dann gilt für alle $Z^* \in \Pi$ und $Z \in L^p(\Omega)$

$$\begin{aligned}\langle Z^*, Z + a\rangle_{L^p(\Omega)} - \mathcal{R}^*(Z^*) &= \langle Z^*, Z\rangle_{L^p(\Omega)} + a\,\mathbb{E}[Z^*] - \mathcal{R}^*(Z^*) \\ &= \langle Z^*, Z\rangle_{L^p(\Omega)} - \mathcal{R}^*(Z^*) + a.\end{aligned}$$

Aus (4.2) folgt dann

$$\begin{aligned}\mathcal{R}(Z + a) &= \sup_{Z^* \in \Pi} \langle Z^*, Z + a\rangle_{L^p(\Omega)} - \mathcal{R}^*(Z^*) \\ &= \sup_{Z^* \in \Pi} \langle Z^*, Z\rangle_{L^p(\Omega)} - \mathcal{R}^*(Z^*) + a = \mathcal{R}(Z) + a.\end{aligned}$$

Zu (iii): Angenommen, $\mathcal{R}$ ist positiv homogen. Dann erfüllt $\mathcal{R}$ alle Voraussetzungen von Lemma A.26, woraus $\mathcal{R}^* = \delta_A$ und

$$\mathcal{R}(Z) = \sup_{Z^* \in A} \langle Z^*, Z\rangle_{L^p(\Omega)}$$

für

$$A := \left\{ Z^* \in \Delta_q \,\middle|\, \langle Z^*, Z\rangle_{L^p(\Omega)} \leq \mathcal{R}(Z) \text{ für alle } Z \in \operatorname{dom} \mathcal{R} \right\}$$

folgen. Wegen $\Pi := \operatorname{dom} \mathcal{R}^* = A$ folgt daraus die Behauptung.

Die umgekehrte Richtung folgt direkt aus der positiven Homogenität von Trägerfunktionalen. □

Für ein konvexes Risikomaß ist die Risikohülle also eine Teilmenge des *Wahrscheinlichkeitssimplex*

$$\Delta_q := \{Z^* \in L^q(\Omega) \mid \mathbb{E}[Z^*] = 1, \quad Z^* \succeq 0\}.$$

Insbesondere für kohärente Risikomaße erhalten wir aus dem Beweis von Satz 4.1 (iii) wegen $\Pi = \operatorname{dom} \mathcal{R}^*$ die folgende Charakterisierung.

Folgerung 4.2. *Sei $\mathcal{R} : L^p(\Omega) \to \overline{\mathbb{R}}$ ein eigentliches und unterhalbstetiges kohärentes Risikomaß. Dann gilt*

$$\Pi = \left\{ Z^* \in \Delta_q \,\middle|\, \langle Z^*, Z\rangle_{L^p(\Omega)} \leq \mathcal{R}(Z) \text{ *für alle* } Z \in \operatorname{dom} \mathcal{R} \right\}$$

sowie

$$\mathcal{R}^*(Z^*) = \delta_\Pi(Z^*) \qquad \text{*für alle* } Z^* \in L^q(\Omega).$$

Daraus folgt insbesondere, dass die Risikohülle eines kohärenten (und unterhalbstetigen) Risikomaßes stets konvex und schwach-$*$ abgeschlossen ist.

Umgekehrt definiert jede nichtleere, konvexe, und schwach-$*$ abgeschlossene Teilmenge $\Pi \subset \Delta_q$ ein unterhalbstetiges kohärentes Risikomaß $\mathcal{R}(Z) := \sup_{Z^* \in \Pi} \langle Z^*, Z\rangle_{L^p(\Omega)}$.

Beispiel 4.3. Wir betrachten folgende Teilmengen von $\Delta := \Delta_1 \subset L^\infty(\Omega)$.

(i) $\Pi := \{Z^* \in L^\infty(\Omega) \mid Z^*(\omega) = 1 \text{ fast überall}\} \subset \Delta$. Dann ist offensichtlich

$$\mathcal{R}(Z) = \sup_{Z^* \in \Pi} \langle Z^*, Z \rangle_{L^1(\Omega)} = \mathbb{E}[Z].$$

(ii) $\Pi = \Delta$. Dann ist

$$\mathcal{R}(Z) = \sup_{Z^* \in \Delta} \langle Z^*, Z \rangle_{L^1(\Omega)} \leq \operatorname*{ess\,sup}_{\omega \in \Omega} Z(\omega) \int_\Omega Z^*(\omega)\, d\mu(\omega) = \operatorname*{ess\,sup}_{\omega \in \Omega} Z(\omega),$$

mit der Konvention $\operatorname{ess\,sup}_{\omega\in\Omega} Z(\omega) = \infty$ für $Z \notin L^\infty(\Omega)$.

Umgekehrt ist $Z \mapsto \operatorname{ess\,sup}_{\omega\in\Omega} Z(\omega)$ eigentlich, unterhalbstetig, konvex, monoton, und translationsinvariant, und damit folgt aus Satz 4.1

$$\operatorname*{ess\,sup}_{\omega \in \Omega} Z(\omega) = \sup_{Z^* \in \Pi'} \langle Z^*, Z \rangle_{L^1(\Omega)}$$

für ein $\Pi' \subset \Delta$. Aus Lemma A.27 folgt dann

$$\operatorname*{ess\,sup}_{\omega \in \Omega} Z(\omega) = \sup_{Z^* \in \Pi'} \langle Z^*, Z \rangle_{L^1(\Omega)} \leq \sup_{Z^* \in \Delta} \langle Z^*, Z \rangle_{L^1(\Omega)} = \mathcal{R}(Z)$$

und damit $\mathcal{R}(Z) = \operatorname{ess\,sup}_{\omega\in\Omega} Z(\omega)$, genannt *worst-case-Risiko*.

Wir kehren nun zurück zum Conditional Value-at-Risk.

Lemma 4.4. *Für alle $\alpha \in (0,1)$ ist CVaR_α ein eigentliches, unterhalbstetiges, und kohärentes Risikomaß mit Risikohülle*

$$\Pi_\alpha = \left\{ Z^* \in \Delta_q \,\middle|\, Z^*(\omega) \leq \alpha^{-1} \text{ für fast alle } \omega \in \Omega \right\}.$$

Beweis. Wir zeigen zunächst, dass CVaR_α eigentlich, konvex, und unterhalbstetig ist. Zunächst gilt

(i) für alle $t \geq 0$, dass $t + \alpha^{-1} \max\{0, -t\} = t \geq 0$;

(ii) für alle $t < 0$, dass $t + \alpha^{-1} \max\{0, -t\} = (1 - \alpha^{-1})t > 0$;

und damit $\mathrm{CVaR}_\alpha(0) = \inf_{t\in\mathbb{R}} t + \alpha^{-1} \max\{0, -t\} = 0 < \infty$.

Seien $Z_1, Z_2 \in L^p(\Omega)$ und $\lambda \in (0,1)$. Dann existieren für $\varepsilon > 0$ beliebig $t_1, t_2 \in \mathbb{R}$ mit

$$t_1 + \frac{1}{\alpha}\mathbb{E}[\max\{0, Z_1 - t_1\}] \leq \mathrm{CVaR}_\alpha(Z_1) + \varepsilon,$$
$$t_2 + \frac{1}{\alpha}\mathbb{E}[\max\{0, Z_2 - t_2\}] \leq \mathrm{CVaR}_\alpha(Z_2) + \varepsilon.$$

Aus der Konvexität von $x \mapsto \max\{0, x\}$ und der Linearität des Erwartungswerts folgt dann

$$\begin{aligned}
\mathrm{CVaR}_\alpha(\lambda Z_1 + (1-\lambda)Z_2) &= \inf_{t\in\mathbb{R}} t + \mathbb{E}[\max\{0, \lambda Z_1 + (1-\lambda)Z_2 - t\}] \\
&\le (\lambda t_1 + (1-\lambda)t_2) + \mathbb{E}[\max\{0, \lambda(Z_1 - t_1) + (1-\lambda)(Z_2 - t_2)\}] \\
&\le \lambda\,(t_1 + \mathbb{E}[\max\{0, Z_1 - t_1\}]) + (1-\lambda)\,(t_2 + \mathbb{E}[\max\{0, Z_2 - t_2\}]) \\
&\le \lambda(\mathrm{CVaR}_\alpha(Z_1) + \varepsilon) + (1-\lambda)(\mathrm{CVaR}_\alpha(Z_2) + \varepsilon) \\
&= \lambda\,\mathrm{CVaR}_\alpha(Z_1) + (1-\lambda)\,\mathrm{CVaR}_\alpha(Z_2) + \varepsilon.
\end{aligned}$$

Da $\varepsilon > 0$ beliebig war, folgt die behauptete Konvexität.

Für die Unterhalbstetigkeit sei $\{Z_n, s_n\}_{n\in\mathbb{N}} \subset \operatorname{epi}\mathrm{CVaR}_\alpha$ mit $Z_n \to Z$ in $L^p(\Omega)$ und $s_n \to s$ in $\mathbb{R}$. Nach Definition existiert dann für alle $n \in \mathbb{N}$ und $\varepsilon > 0$ beliebig ein $t_n \in \mathbb{R}$ mit

$$t_n + \mathbb{E}[\max\{0, Z_n - t_n\}] \le \mathrm{CVaR}_\alpha(Z_n) + \varepsilon \le s_n + \varepsilon.$$

Da für eine Teilfolge von $\{Z_n\}_{n\in\mathbb{N}}$ gilt $Z_{n_k} \to Z$ punktweise und $x \mapsto \max\{0, x - t\}$ stetig ist, muss wegen der Konvergenz von $s_n \to s$ die Folge $\{t_n\}_{n\in\mathbb{N}}$ beschränkt sein. Also können wir – falls nötig durch Übergang zu einer weiteren Teilfolge – annehmen, dass auch $t_{n_k} \to t \in \mathbb{R}$ konvergiert. Damit gilt

$$\begin{aligned}
s + \varepsilon = \lim_{k\to\infty} s_{n_k} + \varepsilon &\ge \liminf_{k\to\infty} t_{n_k} + \mathbb{E}[\max\{0, Z_{n_k} - t_{n_k}\}] \\
&= t + \mathbb{E}[\max\{0, Z - t\}] \\
&\ge \mathrm{CVaR}_\alpha(Z).
\end{aligned}$$

Da $\varepsilon > 0$ beliebig war, folgt $s \ge \mathrm{CVaR}_\alpha(Z)$, d. h. $(Z, s) \in \operatorname{epi}\mathrm{CVaR}_\alpha$.

Für die Darstellung der Risikohülle verwenden wir Folgerung A.28. Einerseits gilt für alle $Z^* \in \Pi_\alpha$, alle $Z \in L^p(\Omega)$, und $t \in \mathbb{R}$ beliebig

$$\begin{aligned}
\langle Z^*, Z\rangle_{L^p(\Omega)} &= \mathbb{E}[Z^*(Z - t)] + t\,\mathbb{E}[Z^*] \\
&\le t + \mathbb{E}[Z^* \max\{0, Z - t\}] \\
&\le t + \frac{1}{\alpha}\,\mathbb{E}[\max\{0, Z - t\}].
\end{aligned}$$

Supremum über alle $Z^* \in \Pi_\alpha$ und Infimum über alle $t \in \mathbb{R}$ ergibt dann

$$\sup_{Z^*\in\Pi_\alpha} \langle Z^*, Z\rangle_{L^p(\Omega)} \le \mathrm{CVaR}_\alpha(Z).$$

Für die umgekehrte Richtung verwenden wir

$$\mathrm{VaR}_\alpha(Z) = \inf\{t > 0 \mid \mathbb{P}[Z > t] \le \alpha\},$$

für das gilt

$$\mathbb{P}[Z > \mathrm{VaR}_\alpha(Z)] \le \alpha \le \mathbb{P}[Z \ge \mathrm{VaR}_\alpha(Z)].$$

Also existiert ein $\lambda \in [0,1]$ mit

$$\alpha = \lambda\, \mathbb{P}[Z > \mathrm{VaR}_\alpha(Z)] + (1-\lambda)\, \mathbb{P}[Z \geq \mathrm{VaR}_\alpha(Z)].$$

Setze nun

$$Z_\alpha^* := \frac{1}{\alpha}\left(\lambda \mathbb{1}_{\{Z > \mathrm{VaR}_\alpha(Z)\}} + (1-\lambda)\mathbb{1}_{\{Z \geq \mathrm{VaR}_\alpha(Z)\}}\right) \in L^\infty(\Omega).$$

Dann ist $\frac{1}{\alpha} \succeq Z_\alpha^* \succeq 0$ sowie

$$\mathbb{E}[Z_\alpha^*] = \frac{1}{\alpha}\left(\lambda\, \mathbb{P}[Z > \mathrm{VaR}_\alpha(Z)] + (1-\lambda)\, \mathbb{P}[Z \geq \mathrm{VaR}_\alpha(Z)]\right) = 1$$

und damit $Z_\alpha^* \in \Pi_\alpha$. Daraus folgt mit der produktiven Null

$$\begin{aligned}
\sup_{Z^* \in \Pi_\alpha} \langle Z^*, Z\rangle_{L^p(\Omega)} \geq \langle Z_\alpha^*, Z\rangle_{L^p(\Omega)} &= \mathbb{E}[Z_\alpha^* Z] \\
&= \mathbb{E}[Z_\alpha^*(Z - \mathrm{VaR}_\alpha(Z))] + \mathrm{VaR}_\alpha(Z)\, \mathbb{E}[Z_\alpha^*] \\
&= \mathrm{VaR}_\alpha(Z) + \frac{1}{\alpha}\, \mathbb{E}[\max\{0, Z - \mathrm{VaR}_\alpha(Z)\}] \\
&\geq \inf_{t \in \mathbb{R}} t + \frac{1}{\alpha}\, \mathbb{E}[\max\{0, Z-t\}] = \mathrm{CVaR}_\alpha(Z),
\end{aligned}$$

wobei wir in der vorletzten Gleichung die Definition von Z_α^* verwendet haben zusammen mit

$$(Z - \mathrm{VaR}_\alpha(Z))\mathbb{1}_{\{Z > \mathrm{VaR}_\alpha(Z)\}} = \max\{0, Z - \mathrm{VaR}_\alpha(Z)\} = (Z - \mathrm{VaR}_\alpha(Z))\mathbb{1}_{\{Z \geq \mathrm{VaR}_\alpha(Z)\}}.$$

Zusammen folgt $\mathrm{CVaR}_\alpha(Z) = \sup_{Z^* \in \Pi_\alpha} \langle Z^*, Z\rangle_{L^p(\Omega)}$. Damit ist CVaR_α nach Satz 4.1 (iii) positiv homogen und insbesondere kohärent.

Umgekehrt folgt aus dem eben Bewiesenen zusammen mit (4.2), dass gilt

$$\sup_{Z^* \in \Pi_\alpha} \langle Z^*, Z\rangle_{L^p(\Omega)} = \mathrm{CVaR}_\alpha(Z) = \sup_{Z^* \in \mathrm{dom}\, \mathcal{R}^*} \langle Z^*, Z\rangle_{L^p(\Omega)}.$$

Nach Folgerung A.28 ist daher $\Pi_\alpha = \mathrm{dom}\, \mathcal{R}^*$ die behauptete Risikohülle. □

Der Beweis zeigt insbesondere, dass das Infimum in der Definition von $\mathrm{CVaR}_\alpha(Z)$ in $t = \mathrm{VaR}_\alpha(Z)$ angenommen wird.

Ohne Beweis geben wir noch das folgende Beispiel (siehe [Royset & Wets 2021, Example 8.31 (c)]).

Beispiel 4.5. Auf $L^2(\Omega)$ ist das *Markowitz-Risiko* für $\gamma > 0$ definiert als

$$\mathcal{R}(Z) = \mathbb{E}[Z] + \gamma\, \mathbb{V}[Z]^{1/2}.$$

Die zugehörige Risikohülle ist

$$\Pi = \left\{1 + \gamma Z^* \,\middle|\, \mathbb{E}[(Z^*)^2] \leq 1,\ \mathbb{E}[Z^*] = 0\right\}.$$

Nach Satz 4.1 (ii) kann das Markowitz-Risiko für $\gamma > 0$ groß genug daher nicht monoton und damit kein konvexes Risikomaß sein.

Zusammen mit dem Satz von Fenchel–Moreau–Rockafellar erhalten wir aus Satz A.22 auch eine Charakterisierung des Subdifferentials von Risikomaßen.

Satz 4.6. *Sei $\mathcal{R} : L^p(\Omega) \to \overline{\mathbb{R}}$ eigentlich und unterhalbstetig und $Z \in \operatorname{dom} \mathcal{R}$. Ist $\mathcal{R}$ ein konvexes Risikomaß, dann gilt*

$$\partial\mathcal{R}(Z) = \operatorname*{arg\,max}_{Z^* \in \Pi} \langle Z^*, Z\rangle_{L^p(\Omega)} - \mathcal{R}^*(Z^*).$$

Ist $\mathcal{R}$ ein kohärentes Risikomaß, so gilt

$$\partial\mathcal{R}(Z) = \operatorname*{arg\,max}_{Z^* \in \Pi} \langle Z^*, Z\rangle_{L^p(\Omega)}.$$

Beweis. Nach Satz A.18 ist $Z^* \in \partial\mathcal{R}(Z)$ genau dann, wenn gilt

$$\langle Z^*, Z\rangle_{L^p(\Omega)} = \mathcal{R}(Z) + \mathcal{R}^*(Z^*).$$

Für konvexe Risikomaße folgt daraus sofort, dass $Z^* \in \partial\mathcal{R}(Z)$ äquivalent ist zu

$$\begin{aligned}\langle Z^*, Z\rangle_{L^p(\Omega)} - \mathcal{R}^*(Z^*) = \mathcal{R}(Z) &= \sup_{Z^* \in L^q(\Omega)} \langle Z^*, Z\rangle_{L^p(\Omega)} - \mathcal{R}^*(Z^*) \\ &= \sup_{Z^* \in \operatorname{dom} \mathcal{R}^*} \langle Z^*, Z\rangle_{L^p(\Omega)} - \mathcal{R}^*(Z^*).\end{aligned}$$

und damit die Behauptung, da wegen Gleichheit das Maximum angenommen wird.

Für kohärente Risikomaße folgt analog mit Satz 4.1 (iii), dass $Z^* \in \partial\mathcal{R}(Z)$ äquivalent ist zu

$$\langle Z^*, Z\rangle_{L^p(\Omega)} = \langle Z^*, Z\rangle_{L^p(\Omega)} - \mathcal{R}^*(Z^*) = \mathcal{R}(Z) = \sup_{Z^* \in \Pi} \langle Z^*, Z\rangle_{L^p(\Omega)},$$

wegen $\mathcal{R}^*(Z^*) = \delta_\Pi(Z^*) = 0$ für $Z^* \in \operatorname{dom} \mathcal{R}^*$. □

Insbesondere folgt für kohärente Risikomaße $\partial\mathcal{R}(0) = \arg\max_{Z^* \in \Pi} 0 = \Pi$.

4.2 OPTIMIERUNG MIT RISIKOMASSEN

Wir betrachten nun wieder die Optimierung von Funktionalen unter Unsicherheiten, d. h. einer Funktion $f : \mathbb{R}^d \times \Omega \to \overline{\mathbb{R}}$, abhängig von der Entscheidungsvariable $x \in \mathbb{R}^d$ und einem zufälligen Ereignis $\omega \in \Omega$. Anders als in Kapitel 1 verwenden wir zunächst F für den *Superpositionsoperator*

$$F : L^p(\Omega; \mathbb{R}^d) \to L^1(\Omega), \qquad [F(u)](\omega) = f(u(\omega), \omega) \quad \text{für fast alle } \omega \in \Omega.$$

Wir untersuchen zuerst die Eigenschaften der Komposition $\mathcal{R} \circ F$, bevor wir dann wieder auf den Spezialfall $u(\omega) \equiv x \in \mathbb{R}^d$ zurückkommen.

Lemma 4.7. *Sei $\mathcal{R} : L^1(\Omega) \to \overline{\mathbb{R}}$ ein konvexes Risikomaß und*

(i) für fast alle $\omega \in \Omega$ ist $x \mapsto f(x, \omega)$ konvex;

(ii) es existiert ein $g \in L^1(\Omega)$ mit $f(x, \omega) \geq -g(\omega)$ für fast alle $\omega \in \Omega$ und alle $x \in \mathbb{R}^d$.

Dann ist auch $\mathcal{R} \circ F : L^p(\Omega; \mathbb{R}^d) \to \overline{\mathbb{R}}$ konvex.

Beweis. Seien $u_1, u_2 \in \operatorname{dom}(\mathcal{R} \circ F)$ (sonst ist nichts zu zeigen) und $\lambda \in (0, 1)$ beliebig. Dann folgt aus der fast sicheren Konvexität von f

$$f(\lambda u_1(\omega) + (1-\lambda)u_2(\omega), \omega) \leq \lambda f(u_1(\omega), \omega) + (1-\lambda) f(u_2(\omega), \omega) \quad \text{für fast alle } \omega \in \Omega.$$

Zusammen mit (ii) folgt daraus $F(\lambda u_1 + (1-\lambda)u_2) \in L^1(\Omega)$, und aus der Monotonie und der Konvexität von $\mathcal{R}$ folgt daher

$$\begin{aligned} \mathcal{R}\left(F(\lambda u_1 + (1-\lambda)u_2)\right) &\leq \mathcal{R}\left(\lambda F(u_1) + (1-\lambda)F(u_2)\right) \\ &\leq \lambda \mathcal{R}\left(F(u_1)\right) + (1-\lambda)\mathcal{R}\left(F(u_2)\right). \end{aligned}$$

□

Beachte, dass die Aussage ohne die Voraussetzung der Monotonie von $\mathcal{R}$ im Allgemeinen nicht gilt.

Lemma 4.8. *Sei $\mathcal{R} : L^1(\Omega) \to \overline{\mathbb{R}}$ ein unterhalbstetiges konvexes Risikomaß und*

(i) für fast alle $\omega \in \Omega$ ist $x \mapsto f(x, \omega)$ unterhalbstetig;

(ii) es existiert ein $g \in L^1(\Omega)$ mit $f(x, \omega) \geq -g(\omega)$ für fast alle $\omega \in \Omega$ und alle $x \in \mathbb{R}^d$.

Dann ist auch $\mathcal{R} \circ F : L^p(\Omega; \mathbb{R}^d) \to \overline{\mathbb{R}}$ unterhalbstetig.

Beweis. Sei $\{(u_n, t_n)\}_{n\in\mathbb{N}} \subset \operatorname{epi}(\mathcal{R} \circ F)$ mit $u_n \to u$ in $L^p(\Omega;\mathbb{R}^d)$ und $t_n \to t$. Durch Übergang zu einer Teilfolge können wir annehmen, dass $u_n \to u$ punktweise fast überall gilt. Aus der fast sicheren Unterhalbstetigkeit von f folgt dann

$$f(u(\omega),\omega) \le \liminf_{n\to\infty} f(u_n(\omega),\omega) \quad \text{für fast alle } \omega \in \Omega.$$

Sei nun $\{u_{n_k}\}_{k\in\mathbb{N}}$ die Teilfolge, die den lim inf realisiert, d. h.

$$f(u_{n_k}(\omega),\omega) \to \liminf_{n\to\infty} f(u_n(\omega),\omega) =: \tilde{F}(\omega)$$

punktweise fast überall. Aus der Monotonie und Unterhalbstetigkeit von $\mathcal{R}$ folgt dann

$$\mathcal{R}(F(u)) \le \mathcal{R}(\liminf_{n\to\infty} F(u_n)) = \mathcal{R}(\tilde{F}) \le \liminf_{k\to\infty} \mathcal{R}(F(u_{n_k})) \le \limsup_{k\to\infty} t_{n_k} = t,$$

und damit $(u,t) \in \operatorname{epi}\mathcal{R} \circ F$. □

Die Existenz von Lösungen zu (4.4) kann daher (unter üblichen Annahmen, die die Unbeschränktheit nach unten verhindern) mit Hilfe von Satz A.1 gezeigt werden.

Analog zu Satz 1.3 zeigen wir nun ein Resultat zur Vertauschbarkeit von Risikomaßen und Infimum.

Satz 4.9. *Angenommen, für $f : \mathbb{R}^d \times \Omega \to \overline{\mathbb{R}}$ gilt*

(i) für fast alle $\omega \in \Omega$ ist $x \mapsto f(x,\omega)$ konvex und unterhalbstetig;

(ii) für fast alle $\omega \in \Omega$ hat $\operatorname{dom}(f(\cdot,\omega))$ nichtleeres Inneres.

Sei $\mathcal{R} : L^1(\Omega) \to \overline{\mathbb{R}}$ ein eigentliches, unterhalbstetiges, und konvexes Risikomaß. Setze

$$[F^*](\omega) := \inf_{v\in\mathbb{R}^d} f(v,\omega) \qquad \text{für fast alle } \omega \in \Omega.$$

Ist $F^* \in (\operatorname{dom}\mathcal{R})^o$, dann gilt

$$\mathcal{R}(F^*) = \inf_{u\in L^p(\Omega;\mathbb{R}^d)} \mathcal{R}(F(u)),$$

und aus $f(\bar{u}(\omega),\omega) = \min_{u\in L^p(\Omega;\mathbb{R}^d)} f(u(\omega),\omega)$ für fast alle $\omega \in \Omega$ folgt

$$\mathcal{R}(F(\bar{u})) = \min_{u\in L^p(\Omega;\mathbb{R}^d)} \mathcal{R}(F(u)).$$

Ist $\mathcal{R}$ strikt monoton, d. h. aus $Z' \succeq Z$ und $Z' \neq Z$ folgt $\mathcal{R}(Z') > \mathcal{R}(Z)$, so gilt auch die umgekehrte Implikation.

Beweis. Zunächst folgt aus der Monotonie von $\mathcal{R}$ zusammen mit der Annahme $F^* \in \operatorname{dom}\mathcal{R}$ sofort

$$\mathcal{R}(F^*) \leq \mathcal{R}(F(u))) \qquad \text{für alle } u \in L^p(\Omega;\mathbb{R}^d)$$

und damit durch Infimum auf der rechten Seite $\mathcal{R}(F^*) \leq \inf_{u\in L^p(\Omega;\mathbb{R}^d)} \mathcal{R}(F(u))$.

Genau wie im Beweis von Satz 1.3 konstruieren wir nun eine Folge $\{u_n\}_{n\in\mathbb{N}} \subset L^p(\Omega;\mathbb{R}^d)$ mit

$$\lim_{n\to\infty} f(u_n(\omega),\omega) = \inf_{v\in\mathbb{R}^d} f(v,\omega) \quad \text{für fast alle } \omega \in \Omega.$$

Dann folgt insbesondere auch $F(u_n) \in L^1(\Omega)$ für alle $n \in \mathbb{N}$ und $F(u_n) \to F^*$ in $L^1(\Omega)$. Nach Satz A.9 ist nun das auf $L^1(\Omega)$ unterhalbstetige und konvexe Funktional $\mathcal{R}$ stetig in $F^* \in (\operatorname{dom}\mathcal{R})^o$, und wir erhalten

$$\mathcal{R}(F^*) = \lim_{n\to\infty} \mathcal{R}(F(u_n)) \geq \inf_{u\in L^p(\Omega;\mathbb{R}^d)} \mathcal{R}(F(u)).$$

Ist $\mathcal{R}$ strikt monoton, dann folgt aus $\mathcal{R}(F(u)) = \mathcal{R}(F^*)$ zusammen mit $F(u) \succeq F^*$ nach Definition von F^* zwingenderweise $F(u) = F^*$. □

Im Vergleich zu Satz 1.3 mussten wir hier zusätzlich annehmen, dass das punktweise Minimum F^* in $\operatorname{dom}\mathcal{R}$ liegt und $\mathcal{R}$ in F^* stetig ist, da wir nicht auf die Eigenschaften des Integrals (und insbesondere $\operatorname{dom}\mathbb{E} = L^1(\Omega)$) zurückgreifen können.

Wir betrachten nun wieder $u \equiv x \in \mathbb{R}^d$, d. h.

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

Für ein gegebenes Risikomaß $\mathcal{R} : L^p(\Omega) \to \overline{\mathbb{R}}$ ist das zugehörige *robuste Optimierungsproblem* dann

$$\min_{x\in\mathbb{R}^d} \mathcal{R}(F(x)). \tag{4.4}$$

Mit Hilfe von (4.2) und (4.3) können wir das Problem schreiben als *Sattelpunktproblem*

$$\min_{x\in\mathbb{R}^d} \sup_{Z^*\in\Pi} \mathbb{E}[f(x,\omega)Z^*(\omega)] - \mathcal{R}^*(Z^*). \tag{4.5}$$

Unter einer Regularitätsannahme können wir Existenz eines Sattelpunktes zeigen und Optimalitätsbedingungen herleiten; vergleiche Satz A.22.

Satz 4.10. *Angenommen, für $f : \mathbb{R}^d \times \Omega \to \overline{\mathbb{R}}$ gilt*

(i) für fast alle $\omega \in \Omega$ ist $x \mapsto f(x,\omega)$ konvex und unterhalbstetig;

(ii) es existiert ein $x_0 \in (\operatorname{dom} f(\cdot,\omega))^o$ für fast alle $\omega \in \Omega$;

(iii) es existiert ein $h \in L^p(\Omega)$ mit $f(x,\omega) \geq -h(\omega)$ für fast alle $\omega \in \Omega$ und alle $x \in \mathbb{R}^d$.

Sei $\mathcal{R} : L^p(\Omega) \to \overline{\mathbb{R}}$, $1 \leq p < \infty$, ein eigentliches, unterhalbstetiges, konvexes Risikomaß. Ist $\bar{x} \in \mathbb{R}^d$ eine Lösung von (4.4) mit $\bar{Z} := F(\bar{x}) \in (\operatorname{dom}\mathcal{R})^o$, dann existiert ein $\bar{Z}^* \in \Pi$ mit

$$\begin{cases} \bar{Z}^* \in \partial\mathcal{R}(\bar{Z}), \\ u^* \in \partial f(\bar{x};\cdot) \quad \text{fast überall}, \\ 0 = \mathbb{E}[u^*\bar{Z}^*]. \end{cases}$$

Weiter ist $(\bar{x}, \bar{Z}^*)$ Sattelpunkt von (4.5), d. h.

$$\begin{aligned} \sup_{Z^* \in \Pi} \mathbb{E}[f(\bar{x},\omega)Z^*(\omega)] - \mathcal{R}^*(Z^*) &\leq \mathbb{E}[f(\bar{x},\omega)\bar{Z}^*(\omega)] - \mathcal{R}^*(\bar{Z}^*) \\ &\leq \min_{x \in \mathbb{R}^d} \mathbb{E}[f(x,\omega)\bar{Z}^*(\omega)] - \mathcal{R}^*(\bar{Z}^*). \end{aligned} \tag{4.6}$$

Beweis. Die Aussage ist ähnlich dem Satz A.22 von Fenchel–Rockafellar; leider können wir nicht direkt die Kettenregel (Satz A.16) für das konvexe Subdifferential anwenden, da das innere Funktional F nicht linear ist. Wir müssen daher "zu Fuß" vorgehen und analog zum Beweis von Satz A.16 die Funktionale $\mathcal{R}$ und F separieren, indem wir den Funktionswert $Z := F(x)$ als unabhängige Variable einführen. Da eine Gleichungsnebenbedingung für nichtlineare Funktionen nicht konvex ist, gehen wir dafür wie folgt vor: Wegen der Monotonie von $\mathcal{R}$ ist $\bar{x} \in \mathbb{R}^d$ eine Lösung von (4.4) genau dann, wenn $(\bar{x}, \bar{Z})$ eine Lösung ist von

$$\min_{(x,Z) \in S} \mathcal{R}(Z), \qquad S := \left\{ (x,Z) \in \mathbb{R}^d \times L^p(\Omega) \,\middle|\, Z \succeq F(x) \right\}.$$

Wegen $\bar{Z} = F(\bar{x}) \in \operatorname{dom}\mathcal{R} \subset L^p(\Omega)$ ist $(\bar{x}, \bar{Z}) \in S$, also ist S nichtleer. Weiter ist S konvex, denn für $(x_1, Z_1), (x_2, Z_2) \in S$ und $\lambda \in (0,1)$ folgt aus der fast sicheren Konvexität von S

$$f(\lambda x_1 + (1-\lambda)x_2, \omega) \leq \lambda f(x_1,\omega) + (1-\lambda) f(x_2,\omega) \leq \lambda Z_1(\omega) + (1-\lambda) Z_2(\omega)$$

für fast alle $\omega \in \Omega$. Also ist $\lambda Z_1 + (1-\lambda)Z_2 \succeq F(\lambda x_1 + (1-\lambda)x_2)$ und damit $\lambda(x_1, Z_1) + (1-\lambda)(x_2, Z_2) \in S$, d. h. S ist konvex. Analog folgt aus der fast sicheren Unterhalbstetigkeit von f für $\{(x_n, Z_n)\}_{n\in\mathbb{N}}$ mit $x_n \to x$ und $Z_n \to Z$ in $L^p(\Omega)$, dass nach Übergang zu einer Teilfolge mit $Z_{n_k} \to Z$ punktweise fast überall gilt

$$f(x,\omega) \leq \liminf_{k\to\infty} f(x_{n_k},\omega) \leq \lim_{k\to\infty} Z_{n_k}(\omega) = Z(\omega)$$

für fast alle $\omega \in \Omega$. Also ist $Z \succeq F(x)$ und damit ist S abgeschlossen.

Außerdem ist $\mathcal{R}$ eigentlich, konvex, und unterhalbstetig sowie stetig in $\bar{Z}$. Nach Satz A.10 zusammen mit Satz A.15 (angewendet auf $\mathcal{R} + \delta_S$) gilt daher

$$0 \in \{0\} \times \partial\mathcal{R}(\bar{Z}) + \partial\delta_S(\bar{x}, \bar{Z}).$$

Es existiert daher ein $\bar{Z}^* \in \partial\mathcal{R}(\bar{Z})$ mit $(0, -\bar{Z}^*) \in \partial\delta_S(\bar{x}, \bar{Z})$. Aus der Darstellung von $\partial\delta_S$ als Normalenkegel folgt daraus

$$\langle \bar{Z}^*, Z - \bar{Z} \rangle_{L^p(\Omega)} \geq 0 \qquad \text{für alle } (x, Z) \in S.$$

Speziell für $Z = F(x)$ ist $(x, F(x)) \in S$ und daher

$$\langle \bar{Z}^*, F(x) - F(\bar{x}) \rangle_{L^p(\Omega)} \geq 0 \qquad \text{für alle } x \in \mathbb{R}^d.$$

Also ist $\bar{x}$ Minimierer von

$$G : \mathbb{R}^d \to \overline{\mathbb{R}}, \qquad G(x) = \int_\Omega f(x, \omega)\bar{Z}^*(\omega)\, d\mu(\omega) = \langle \bar{Z}^*, F(x) \rangle_{L^p(\Omega)}.$$

Wegen $\bar{Z}^* \in \Pi$ gilt $\bar{Z}^* \in L^q(\Omega)$ und $\bar{Z}^* \succeq 0$; also können wir Satz 1.6 anwenden auf die Funktion

$$g : \mathbb{R}^d \times \Omega \to \mathbb{R}, \qquad g(x, \omega) = f(x, \omega)\bar{Z}^*(\omega)$$

mit integrierbarer Minorante $h \cdot \bar{Z}^* \in L^1(\Omega)$ und erhalten

$$0 \in \partial G(\bar{x}) = \{\mathbb{E}[v^*] \mid v^*(\omega) \in \partial g(x; \omega) \text{ fast überall}\}.$$

Mit

$$\partial g(x; \omega) = \bar{Z}^*(\omega)\partial f(x; \omega) \quad \text{für fast alle } \omega \in \Omega$$

wegen Lemma A.14 (i) erhalten wir daraus die behaupteten Extremalitätsbedingungen.

Aus Satz 4.6 folgt schließlich

$$\bar{Z}^* \in \partial\mathcal{R}(\bar{Z}) = \arg\max_{Z^* \in \Pi} \langle Z^*, F(\bar{x}) \rangle_{L^p(\Omega)} - \mathcal{R}^*(Z^*),$$

und damit ist $(\bar{x}, \bar{Z}^*)$ ein Sattelpunkt von (4.5). □

Da das Fermat-Prinzip für konvexe Funktionale notwendig und hinreichend ist, gilt umgekehrt, dass jedes Paar $(\bar{x}, \bar{Z}^*)$, das (4.6) erfüllt, auch ein Sattelpunkt von (4.5) ist, und insbesondere $\bar{x}$ eine Lösung von (4.4) liefert.

Satz 4.10 zusammen mit Satz 4.6 zeigt insbesondere, dass die Optimierung von konvexen Risikomaßen äquivalent ist zur Optimierung von *gewichteten* Erwartungswertfunktionalen für ein – von $\bar{x}$ abhängigen – Gewicht $\bar{Z}^* \in \Pi$, das genau die "worst case" Szenarien $f(\bar{x}, \omega)$ am stärksten gewichtet. Das konkrete Risikomaß $\mathcal{R}$ schränkt dabei die zulässigen Gewichte über die zugehörige Risikohülle Π ein.

Beispiel 4.11. Wir betrachten wieder das Trafikantenproblem aus Beispiel 1.7, diesmal in

der risiko-aversen Form

$$\min_{x\geq 0}(p_e - p_c)x + p_v \operatorname{CVaR}_\alpha(\max\{0, x-\xi\}) \\ = \min_x \operatorname{CVaR}_\alpha\left((p_e-p_c)x + p_v(\max\{0,x-\xi\}) + \delta_{(0,\infty)}(x)\right)$$

(wegen der Translationsinvarianz von konvexen Risikomaßen) für $\xi \in \{0,\dots,M\}$ uniform verteilt und $\alpha \in (0,1)$. Wieder ist

$$f(x,\xi) := (p_e - p_v)x + p_v \max\{0, x-\xi\} + \delta_{(0,\infty)}(x) \geq (p_e - p_v)M$$

konvex und stetig bezüglich x für alle ξ, so dass wir Satz 4.10 anwenden können und ein $\bar{Z}^* \in \mathbb{R}^{M+1}$ erhalten mit

$$\begin{cases} \bar{Z}^* \in \partial \operatorname{CVaR}_\alpha(f(\bar{x},\xi)), \\ u_i^* \in \partial f(\bar{x}; i) \quad i = 0,\dots,M, \\ 0 = \dfrac{1}{M+1}\displaystyle\sum_{i=0}^{M} u_i^* \bar{Z}_i^*. \end{cases}$$

Nach Satz 4.6 und Satz 4.1 gilt

$$\bar{Z}^* = \arg\max\left\{ \frac{1}{M+1}\sum_{i=0}^{M} Z_i f(\bar{x}, i) \,\middle|\, Z_i \in [0, \alpha^{-1}], \quad \frac{1}{M+1}\sum_{i=0}^{M} Z_i = 1 \right\}.$$

Diese Charakterisierung ist normalerweise nicht explizit auflösbar; in diesem Fall können wir aber ausnutzen, dass $i \mapsto f(x,i)$ monoton fallend ist für alle $x \geq 0$, so dass für $N := \alpha(M+1) < M \in \mathbb{N}$ die Lösung gegeben ist durch

$$\bar{Z}_i^* = \begin{cases} \alpha^{-1} & i < N, \\ 0 & \text{sonst.} \end{cases}$$

(Andernfalls ist noch ein $\bar{Z}_N^* \in (0, \alpha^{-1})$ zu berücksichtigen.)
Die letzte Gleichung ergibt daher die Optimalitätsbedingung

$$0 = \frac{1}{M+1}\sum_{i=0}^{N-1}\frac{1}{\alpha}u_i^* = \frac{1}{N}\sum_{i=0}^{N-1} u_i^*,$$

woraus wie in Beispiel 1.7 folgt, dass die optimale Lösung $\bar{x}$ die Bedingung

$$\bar{x} \leq N\frac{p_v - p_e}{p_v} \leq \bar{x} + 1$$

erfüllen muss.
Im Gegensatz zum risikoneutralen Trafikantenproblem werden hier also nur die N "worst case"-Szenarien berücksichtigt (die in diesem einfachen Beispiel unabhängig von x sind). Der risikoaverse Trafikant ist also ein Pessimist: Er geht davon aus, dass nur ein Bruchteil (je größer α, desto kleiner) der möglichen Käufer:innen auftauchen werden.

Für kohärente Risikomaße hat das Sattelpunktproblem die besonders einfache Form

$$\min_{x\in\mathbb{R}^d} \sup_{Z^*\in\Pi} \mathbb{E}[f(x,\omega)Z^*(\omega)].$$

Diese kann nun Ausgangspunkt für weitere Untersuchungen sein: Zum Beispiel kann wie in Kapitel 2 der Erwartungswert durch eine Monte-Carlo-Approximation ersetzt werden; zusammen mit einer passenden Approximation der Risikohülle Π kann man dann ähnlich wie in Satz 2.3 Konsistenz zeigen; siehe [Shapiro 2013].

Ebenfalls kann man daraus numerische Verfahren ableiten. Ist zum Beispiel f fast sicher stetig differenzierbar und die Projektion auf Π effizient realisierbar, kann man einen Sattelpunkt durch das folgende *primal-duale Splitting-Verfahren* berechnen:

$$\begin{aligned} Z^{k+1} &= \operatorname{proj}_\Pi(Z^k + \sigma_k F(x^k)),\\ \hat{Z}^{k+1} &= 2Z^{k+1} - Z^k,\\ x^{k+1} &= x^k - \tau_k F'(x^k)^* \hat{Z}^{k+1}, \end{aligned}$$

für geeignete Schrittweiten $\sigma_k, \tau_k > 0$; siehe [Angerhausen 2022] für den speziellen Fall $\mathcal{R} = \mathrm{CVaR}_\alpha$.

A GRUNDLAGEN DER KONVEXEN ANALYSIS

In diesem Kapitel sammeln wir die wesentlichen Definitionen und Resultate aus der konvexen Analysis, die in dieser Vorlesung benötigt werden. Für Details und Beweise wird auf [Clason & Valkonen 2020] verwiesen.

Sei X ein normierter Vektorraum und $F : X \to \overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty\}$, wobei $\overline{\mathbb{R}}$ mit der üblichen Arithmetik versehen wird, d. h. $t < \infty$ und $t + \infty = \infty$ für alle $t \in \mathbb{R}$; Subtraktion und Multiplikation von negativen Zahlen mit ∞ und insbesondere $F(x) = -\infty$ sind nicht zugelassen. Die Menge, auf der F endlich ist, bezeichnet man als (*effektiven*) *Definitionsbereich*

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

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

A.1 DIREKTE METHODE DER VARIATIONSRECHNUNG

Wir verallgemeinern 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 F. 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*.

Satz A.1 (Tonelli). *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 $\bar{x} \in \operatorname{dom} F$.

Ist X nicht reflexiv, aber Dualraum eines separablen Banachraums, so hat man analog die Existenz von Minimierern schwach-$*$ unterhalbstetiger Funktionale.

Unterhalbstetigkeit bleibt unter verschiedenen Operationen erhalten.

Lemma A.2. *Seien X, Y Banachräume und sei $F : X \to \overline{\mathbb{R}}$ schwach unterhalbstetig. Dann sind schwach unterhalbstetig*

(i) *αF 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.*

Da beschränkte lineare Funktionale stetig und damit unterhalbstetig sind, folgt sofort:

Folgerung A.3. *Sei X ein Banachraum. Dann ist $\| \cdot \|_X$ eigentlich, koerziv und schwach unterhalbstetig.*

Ein weiteres häufig auftretendes Funktional ist die *Indikator-Funktion*[1] einer Menge $U \subset X$, definiert als

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

Lemma A.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.*

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

A.2 KONVEXE FUNKTIONEN

Ein eigentliches Funktional $F : X \to \overline{\mathbb{R}}$ heißt *konvex*, wenn für alle $x, y \in X$ und $\lambda \in [0,1]$ gilt

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

(dabei ist der Funktionswert ∞ 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*

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

Lemma A.5. *Für $F : X \to \overline{\mathbb{R}}$ ist* 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.

Folgerung A.6. *Sei $F : X \to \overline{\mathbb{R}}$ konvex. Dann ist F schwach unterhalbstetig genau dann, wenn F unterhalbstetig ist.*

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 δ_C für eine konvexe Menge C.

Ist X ein Hilbertraum, so ist $F(x) = \|x\|_X^2$ sogar strikt konvex.

Weitere Beispiele lassen sich analog zu Lemma A.2 durch folgende Operationen erzeugen.

Lemma A.7. *Seien X, Y normierte Räume und sei $F : X \to \overline{\mathbb{R}}$ konvex. Dann sind konvex*

(i) αF 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 A.7 (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*

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

Lemma A.8. *Sei* $F : X \to \overline{\mathbb{R}}$ *eigentlich. Dann ist* F *genau dann konvex und unterhalbstetig, wenn gilt* $F = F^\Gamma$.

Schließlich benötigen wir noch die folgende überraschende und nützliche Eigenschaft.

Satz A.9. *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.*

A.3 DAS KONVEXE SUBDIFFERENTIAL

Konvexe Funktionen müssen nicht differenzierbar sein; geometrisch bedeutet das, dass anstelle einer eindeutigen Tangente in einem Punkt mehrere existieren können. Wir betrachten daher als Ersatz für die Ableitung die *Menge* aller Tangentensteigungen in einem Punkt. Für $F : X \to \overline{\mathbb{R}}$ und $x \in \operatorname{dom} F$ is das *Subdifferential* von F in x definiert als

$$\partial F(x) := \{x^* \in X^* \mid \langle x^*, \tilde{x} - x\rangle_X \le F(\tilde{x}) - F(x) \quad \text{für alle } \tilde{x} \in X\}. \tag{A.2}$$

(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; man zeigt weiter, dass $\partial F(x) \neq$ für alle $x \in (\operatorname{dom} F)^o$ ist. Ein Element $\xi \in \partial F(x)$ heißt *Subgradient*.

Direkt aus der Definition folgt eine notwendige und hinreichende Optimalitätsbedingung.

Satz A.10. *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) $F(\bar{x}) = \min_{x\in X} F(x)$.

Eine äquivalente Definition erhält man mit Hilfe der *Richtungsableitung* von F in x dom F in Richtung $h \in X$

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

die für konvexe Funktionen (zumindest als uneigentlicher Grenzwert) stets existiert.

Lemma A.11. *Sei $F : X \to \overline{\mathbb{R}}$ konvex und $x \in \operatorname{dom} F$. Dann ist*

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

Definiert $h \mapsto F'(x;h)$ einen stetigen linearen Operator $DF(x) : X \to \mathbb{R}$, so ist F *Gâteaux-differenzierbar* in x; aus Lemma A.11 folgt also, dass das Subdifferential eine Verallgemeinerung der klassischen Ableitung darstellt.

Satz A.12. *Sei $F : X \to \overline{\mathbb{R}}$ konvex und Gâteaux-differenzierbar in x. Dann ist $\partial F(x) = \{DF(x)\}$.*

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 A.13. *Für $x \in X$ ist*

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

Das Subdifferential der Indikatorfunktion einer konvexen Menge $C \subset X$ hat eine einfache Darstellung: 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.

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 A.14. *Für $F : X \to \overline{\mathbb{R}}$ konvex und $x \in \operatorname{dom} F$ gilt*

(i) $\partial(\lambda F)(x) = \lambda(\partial F(x)) := \{\lambda\xi \mid \xi \in \partial F(x)\}$ *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 und benötigt den Satz von Hahn–Banach.

Satz A.15 (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.

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 stetige lineare Operatoren.

Satz A.16 (Kettenregel). *Seien X, Y normierte Räume, $A \in L(X,Y)$ (d. h. ein stetiger linearer Operator von X nach Y), und $F : Y \to \overline{\mathbb{R}}$ konvex und unterhalbstetig. Dann gilt für alle $x \in \operatorname{dom}(F \circ A)$*

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

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

Nützlich ist auch das folgende Resultat.

Satz A.17 (Maximumsregel). *Sei $F : X \to \mathbb{R}$ konvex und gelte*

$$F(x) = \max_{i \in I} F_i(x)$$

für eine endliche Indexmenge I und $F_i : X \to \mathbb{R}$ Gâteaux-differenzierbar. Dann ist

$$\partial F(x) = \overline{\operatorname{co}}\{DF_i(x) \mid F(x) = F_i(x)\},$$

wobei $\overline{\operatorname{co}} A$ die schwach-$*$ abgeschlossene konvexe Hülle der Menge $A \subset X^*$ bezeichnet.

A.4 FENCHEL-DUALITÄT

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 A.7 (v) und Lemma A.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 A.8 für konvexe und unterhalbstetige Funktionen immer der Fall ist), so ist auch F^* eigentlich. Die Definition liefert außerdem sofort die *Fenchel–Young-Ungleichung*

$$\langle x^*, x \rangle_X \leq F(x) + F^*(x^*) \qquad \text{für alle } x \in X, x^* \in X^*. \tag{A.3}$$

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 A.8 ja mit F übereinstimmt.

Satz A.18 (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.*

Wir betrachten wieder relevante Beispiele.

Beispiel A.19.

(i) 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, dass $(\delta_{B_{X^*}})_*(x) = \|x\|_X$ ist.

(ii) Sei X ein normierter Raum und setze $F(x) = \|x\|_X$. Dann ist $F^* = \delta_{B_{X^*}}$. Wie oben zeigt man analog, dass $(\|\cdot\|_{X^*})_* = \delta_{B_X}$ ist.

(iii) Sei $F(x) = \frac{1}{2}\|x\|_X^2$ und X ein Hilbertraum. Dann können wir $X^* \simeq X$ identifizieren, so dass $F(x)$ differenzierbar ist mit Gradient $\nabla F(x) = x$, und die notwendigen und hinreichenden Optimalitätsbedingungen für $\sup_{x \in X} \langle x^*, x \rangle_X - \frac{1}{2}\langle x, x \rangle_X$ ergeben

$$F^*(x^*) = \frac{1}{2}\|x^*\|_X^2.$$

Wir notieren noch einige nützliche Rechenregeln.

Lemma A.20. *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* $A^{-*} := (A^{-1})^*$.

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

Satz A.21 (Fenchel–Moreau). *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) $x \in \partial F^*(x^*)$.

Mit Hilfe dieses Satzes erhalten wir ein Dualitätsresultat für Probleme der Form

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

für $F : X \to \overline{\mathbb{R}}$ und $G : Y \to \overline{\mathbb{R}}$ eigentlich, konvex, und unterhalbstetig sowie $A \in L(X, Y)$. Durch Einsetzen der Definition von $G = G^{**}$ und formales Vertauschen von inf und sup erhält man daraus das *duale Problem*

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

Dieses Vertauschen wird durch das folgenden Satz legitimiert.

Satz A.22 (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*

$$\min_{x\in X} F(x) + G(Ax) = \max_{y^*\in Y^*} -F^*(-A^*y^*) - G^*(y^*). \tag{A.4}$$

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.

A.5 EPIKONVERGENZ

Oft betrachtet man zu einem Optimierungsproblem eine Folge von approximierenden (z. B. diskretisierten oder regularisierten) Problemen, die einfacher zu lösen oder zu analysieren sind. Dabei ist natürlich die wesentliche Frage, ob Minimierer der approximierenden Funktionale gegen einen Minimierer des ursprünglichen Funktionals konvergieren. Dafür reicht punktweise Konvergenz der Funktionale nicht aus, wie das folgende Beispiel zeigt.

Beispiel A.23. Betrachte für $n \in \mathbb{N}$

$$f_n : \mathbb{R} \to \mathbb{R}, \qquad f_n(x) = \frac{1}{n}|x - n| - 1.$$

Dann gilt

$$\lim_{n\to\infty} f_n(x) = 0 \qquad \text{für alle } x \in \mathbb{R},$$

und damit $f_n \to 0 =: f$ punktweise, aber

$$\lim_{n\to\infty} \inf_{x\in\mathbb{R}} f_n(x) = \lim_{n\to\infty} f_n(n) = -1 \neq 0 = \inf_{x\in\mathbb{R}} f(x).$$

Die Konvergenz muss also in einem passenden Sinn gleichmäßig sein. Für erweitert-reellwertige Funktionale führt das auf den Begriff der *Epikonvergenz* (im Kontext der Variationsrechnung auch *Gamma-Konvergenz* genannt).

Sei $F_n : X \to \overline{\mathbb{R}}$, $n \in \mathbb{N}$. Wir sagen, F_n *epikonvergiert* gegen $F : X \to \overline{\mathbb{R}}$ und schreiben $F_n \to_e F$, falls für alle $x \in X$ gilt

(i) für alle Folgen $\{x_n\}_{n\in\mathbb{N}} \subset X$ mit $x_n \to x$ gilt

$$\liminf_{n\to\infty} F_n(x_n) \geq F(x);$$

(ii) es existiert eine *Recovery-Folge* $\{x_n\}_{n\in\mathbb{N}} \subset X$ mit $x_n \to x$ so, dass gilt

$$\lim_{n\to\infty} F_n(x_n) = F(x).$$

In Anbetracht von (i) genügt es anstelle von (ii) für eine Recovery-Folge nachzuweisen, dass gilt

(ii') $x_n \to x$ und

$$\limsup_{n\to\infty} F_n(x_n) \leq F(x).$$

(Eine äquivalente Charakterisierung von Epikonvergenz ist, dass epi $F_n \to$ epi F im Sinne der Painlevé–Kuratowski-Konvergenz von Mengen gilt, woher auch der Name stammt.)

Beachte, dass punktweise Konvergenz weder hinreichend noch notwendig für Epikonvergenz ist. Aus der Definition folgt direkt, dass der Grenzwert einer epikonvergenten Folge unterhalbstetig sein muss; sind alle F_n konvex, so ist auch F konvex.

Unter einer – separat nachzuweisenden – Kompaktheitsbedingung erhält man nun das gewünschte Ergebnis.

Satz A.24. *Sei $F_n : X \to \overline{\mathbb{R}}$, $n \in \mathbb{N}$, und $F : X \to \overline{\mathbb{R}}$ mit $F_n \to_e F$. Dann gilt*

$$\limsup_{n\to\infty} \inf_{x\in X} F_n(x) \leq \inf_{x\in X} F(x).$$

Existiert weiter eine Nullfolge $\{\varepsilon_n\}_{n\in\mathbb{N}}$ und eine konvergente Folge $\{x_n\}_{n\in\mathbb{N}} \subset X$ mit

$$F_n(x_n) \leq \inf_{x\in X} F_n(x) + \varepsilon_n \qquad \text{für alle } n \in \mathbb{N},$$

so ist der Grenzwert $\bar{x} \in X$ ein Minimierer von F und es gilt

$$\lim_{n\to\infty} \inf_{x\in X} F_n(x) = \inf_{x\in X} F(x).$$

Beweis. Sei $x \in X$ beliebig und $\{x_n\}_{n\in\mathbb{N}}$ eine zugehörige Recovery-Folge. Dann folgt sofort aus Bedingung (ii) für die Epikonvergenz

$$F(x) \geq \limsup_{n\to\infty} F_n(x_n) \geq \limsup_{n\to\infty} \inf_{x\in X} F_n(x),$$

und Infimum über alle $x \in X$ ergibt die erste Behauptung.

Sei nun $\{x_n\}_{n\in\mathbb{N}}$ eine konvergente Folge von ε_n-Minimierern mit Grenzwert $\bar{x} \in X$. Dann folgt aus der Definition zusammen mit Bedingung (i)

$$\liminf_{n\to\infty} \inf_{x\in X} F_n(x) = \liminf_{n\to\infty} (\inf_{x\in X} F_n(x) + \varepsilon_n) \geq \liminf_{n\to\infty} F_n(x_n) \geq F(\bar{x}) \geq \inf_{x\in X} F(x).$$

Zusammen mit der bereits bewiesenen ersten Ungleichung folgt daraus die behauptete Gleichung sowie $F(\bar{x}) = \inf_{x\in X} F(x)$. □

Das folgende (aufwändig zu beweisende) Resultat charakterisiert Epikonvergenz konvexer Funktionale in $\mathbb{R}^d$.

Satz A.25 ([Rockafellar & Wets 1998, Theorem 7.17]). *Seien $F_n : \mathbb{R}^d \to \overline{\mathbb{R}}$, $n \in \mathbb{N}$, konvex und unterhalbstetig, und sei $F : \mathbb{R}^d \to \overline{\mathbb{R}}$ konvex und unterhalbstetig mit $(\operatorname{dom} F)^o \neq \emptyset$. Dann sind äquivalent:*

(i) $F_n \to_e F$;

(ii) es existiert eine dichte Teilmenge $E \subset \mathbb{R}^d$ mit $F_n \to F$ punktweise auf E;

(iii) $F_n \to F$ gleichmäßig auf jeder kompakten Teilmenge $K \subset \mathbb{R}^d$, die keinen Randpunkt von $\operatorname{dom} F$ enthält.

A.6 TRÄGERFUNKTIONALE

Sei X ein Banachraum und $A \subset X^*$. Dann heißt die Abbildung

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

das *Trägerfunktional* von A. Wegen Lemma A.11 ist das folgende Resultat nützlich.

Lemma A.26. *Sei $F : X \to \overline{\mathbb{R}}$ positiv homogen, subadditiv und unterhalbstetig, und sei*

$$A = \{x^* \in X^* \mid \langle x^*, x \rangle_X \leq F(x) \quad \textit{für alle } x \in X\}.$$

Dann ist

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

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 $x \in X$. Wir machen dafür eine Fallunterscheidung:

(i) $\langle x^*, x \rangle_X \leq F(x)$ für alle $x \in X$. Dann folgt sofort

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

Aus der positiven Homogenität folgt weiter

$$F(0) = F(t0) = tF(0) \qquad \text{für alle } t > 0$$

und damit $F(0) = 0$. Also wird das Supremum für $x = 0$ in $F^*(x^*) = 0$ angenommen.

(ii) $\langle x^*, \bar{x} \rangle_X > F(\bar{x})$ für ein $\bar{x} \in X$. Dann folgt für $t > 0$ beliebig aus der positiven Homogenität

$$\begin{aligned} F^*(x^*) = \sup_{x \in X} \langle x^*, x \rangle_X - F(x) &\geq \langle x^*, t\bar{x} \rangle_X - F(t\bar{x}) \\ &= t\left(\langle x^*, \bar{x} \rangle_X - F(\bar{x})\right) > 0, \end{aligned}$$

und Grenzübergang $t \to \infty$ zeigt $F^*(x^*) = \infty$.

Dies zeigt $F^* = \delta_A$.

Nun ist F nach Voraussetzung zusätzlich subadditiv und damit konvex sowie unterhalbstetig. Aus Satz A.18 folgt daher für alle $x \in X$

$$F(x) = F^{**}(x) = (\delta_A)^*(x) = \sup_{x^* \in A} \langle x^*, x \rangle_X. \qquad \square$$

Die nächsten Aussagen zeigen, dass man Mengen äquivalent durch ihre Trägerfunktionale charakterisieren kann.

Lemma A.27. *Seien $A, B \subset X^*$ nichtleer, konvex, und schwach-$*$ abgeschlossen. Dann ist $A \subset B$ genau dann, wenn gilt*

$$\sup_{x^* \in A} \langle x^*, x \rangle_X \leq \sup_{x^* \in B} \langle x^*, x \rangle_X \qquad \textit{für alle } x \in X. \tag{A.5}$$

Beweis. Ist $A \subset B$, so ist die linke Seite von (A.5) offensichtlich nicht größer als die rechte. 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 dem Trennungssatz von Hahn–Banach (in der schwach-$*$-Topologie, bezüglich der die stetigen linearen Funktionale nach Definition der Form $x^* \mapsto \langle x^*, x \rangle_X$ für ein $x \in X$ sind) 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 (A.5) verletzt, und die Aussage folgt durch Kontraposition. □

Folgerung A.28. *Seien $A, B \subset X^*$ konvex und schwach-$*$ abgeschlossen. Dann ist $A = B$ genau dann, wenn gilt*

$$\sup_{x^* \in A} \langle x^*, x \rangle_X = \sup_{x^* \in B} \langle x^*, x \rangle_X \qquad \textit{für alle } x \in X. \tag{A.6}$$

Beweis. Die eine Richtung ist klar. Gilt umgekehrt (A.6), dann folgt daraus insbesondere (A.5) und damit nach Lemma A.27 auch $A \subset B$. Vertauschen der Rollen von A und B ergibt die Aussage. □

B GRUNDLAGEN DER WAHRSCHEINLICHKEITSTHEORIE

Wir sammeln hier die wesentlichsten Begriffe und Resultate aus der Wahrscheinlichkeitstheorie, die wir in dieser Vorlesung brauchen. Für Details und Beweise sei auf Standard-Lehrbücher wie etwa [Brokate & Kersting 2019; Klenke 2020] verwiesen.

B.1 WAHRSCHEINLICHKEITSMASSE

Sei Ω eine beliebige Menge. Eine Teilmenge $\mathcal{A} \subset \mathcal{P}(\Omega)$ der Potenzmenge von Ω heißt σ-Algebra, falls sie Ω enthält und abgeschlossen ist unter Komplementbildung und abzählbaren Vereinigungen. (Daraus folgt sofort, dass die leere Menge sowie abzählbare Schnitte ebenfalls in $\mathcal{A}$ liegen.) In diesem Fall nennen wir $(\Omega, \mathcal{A})$ einen *Maßraum* und $A \subset \mathcal{A}$ eine *messbare Menge*. Ein $\mathcal{S} \subset \mathcal{P}(\Omega)$, für das $\mathcal{A}$ die kleinste σ-Algebra ist, die $\mathcal{S}$ enthält, nennt man *Erzeuger von* $\mathcal{A}$. Ist $\Xi \subset \mathbb{R}^d$ und enthält $\mathcal{S}$ genau die offenen Teilmengen von Ξ, so wird dadurch die *Borel-Algebra* $\mathcal{B}$ erzeugt; $B \in \mathcal{B}$ heißt dann *Borel-Menge*.

Eine Funktion $\mu : \mathcal{A} \to [0, \infty]$ heißt *Maß* auf $(\Omega, \mathcal{A})$, wenn für alle $A_n \in \mathcal{A}$, $n \in \mathbb{N}$, mit $A_i \cap A_j = \emptyset$ für $i \neq j$ gilt

$$\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} \mu(A_n).$$

Gilt für alle $A \in \mathcal{A}$ – und damit insbesondere für Ω selber – dass $\mu(A) < \infty$, so heißt μ *endlich*; kann Ω zerlegt werden in die Vereinigung abzählbar vieler meßbarer Mengen $A_n \in \mathcal{A}$ mit $\mu(A_n) < \infty$ für alle $n \in \mathbb{N}$, so heißt μ *σ-endlich*; ein Beispiel ist das Lebesgue-Maß auf $\mathbb{R}^d$. Eine σ-Algebra $\mathcal{A}$ heißt *μ-vollständig*, wenn aus $A \in \mathcal{A}$ mit $\mu(A) = 0$ und $B \subset A$ auch $B \in \mathcal{A}$ und damit $\mu(B) = 0$ folgt. Gilt schließlich $\mu(\Omega) = 1$, so heißt μ ein *Wahrscheinlichkeitsmaß* und $(\Omega, \mathcal{A}, \mu)$ ein *Wahrscheinlichkeitsraum*. In diesem Fall wird das Wahrscheinlichkeitsmaß auch oft mit $\mathbb{P}$ bezeichnet.

Eine Menge $A \in \mathcal{A}$ mit $\mu(A) = 0$ heißt *(μ-)Nullmenge*. Gilt umgekehrt $\mu(\Omega \setminus A) = 0$, dann sagt man, das Ereignis A tritt *(μ-)fast sicher* (oder *fast überall*) ein; ist μ ein Wahrscheinlichkeitsmaß, so gilt in diesem Fall $\mu(A) = 1$, und wir sagen A tritt ein *mit Wahrscheinlichkeit* 1. Ein sehr nützliches Resultat dazu ist das folgende

Lemma B.1 (Borel–Cantelli). *Sei* $\{A_1, \dots, A_n\}_{n\in\mathbb{N}} \subset \mathcal{A}$ *und*

$$A^* := \{\omega \in \Omega \mid \omega \in A_n \text{ für unendlich viele } n \in \mathbb{N}\}.$$

Dann gilt

(i) Ist $\sum_{n\in\mathbb{N}} \mu(A_n) < \infty$, *so ist* $\mu(A^*) = 0$.

(ii) Ist $\sum_{n\in\mathbb{N}} \mu(A_n) = \infty$ *und* $\{A_n\}_{n\in\mathbb{N}}$ *unabhängig, so ist* $\mu(A^*) = 1$.

Hier nennt man eine Familie $\{A_i\}_{i\in I} \subset \mathcal{A}$ *unabhängig*, wenn für alle endlichen Teilmengen $J \subset I$ gilt $\mu(\bigcap_{j\in J} A_j) = \prod_{j\in J} \mu(A_j)$.

Ist $f : \Omega \to \mathbb{R}$ eine Funktion und $(\Omega, \mathcal{A})$ ein Maßraum, dann heißt f *messbar*, wenn für alle Borel-Mengen $B \in \mathcal{B}$ das Urbild $f^{-1}[B] := \{\omega \in \Omega \mid f(\omega) \in B\} \in \mathcal{A}$, d. h. messbar ist. Stetige Funktionen sind immer messbar; ist f messbar, so auch $f^+ := \max\{0, f\}$, $f^- := \max\{0, -f\}$, $|f|$, und $\operatorname{sign} f$. Weiter sind punktweise Summe, Differenz, Produkt, und Quotient von messbaren Funktionen messbar. Ist $\{f_n\}_{n\in\mathbb{N}}$ eine Folge von messbaren Funktionen, sind auch $\inf_n f_n$, $\sup_n f_n$, $\liminf_n f_n$, und $\limsup_n f_n$ messbar. Schließlich ist eine vektorwertige Funktion messbar genau dann, wenn die Koordinatenfunktionen messbar sind. Diese Aussagen gelten auch für erweitert-reellwertige Funktionen.

Jede nicht-negative meßbare Funktion $f : \Omega \to [0, \infty)$ ist Grenzwert einer fast überall monoton steigenden Folge von *Elementarfunktionen*, d. h. es gibt eine Folge $\{f_n\}_{n\in\mathbb{N}}$ mit

$$f_n(\omega) = \sum_{k=1}^{N} \mathbb{1}_{A_k} \alpha_k, \qquad A_1, \dots, A_N \in \mathcal{A}, \quad \alpha_1, \dots, \alpha_N \geq 0,$$

und $f_n \nearrow f$ punktweise fast überall. (Hier bezeichnet $\mathbb{1}_A(\omega) = 1$ für $\omega \in A$ und 0 sonst die *charakteristische Funktion* von $A \subset \Omega$). Es gilt daher

$$f(\omega) = \sum_{k=1}^{\infty} \mathbb{1}_{A_k} \alpha_k, \qquad A_k \in \mathcal{A}, \quad \alpha_k \geq 0 \text{ für alle } k \in \mathbb{N}.$$

B.2 INTEGRATION

Ist $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $f : \Omega \to \mathbb{R}$ eine messbare Funktion, ist das *Integral* bezüglich μ definiert als

$$\int_\Omega f(\omega)\, d\mu(\omega) := \sum_{k=1}^{\infty} \alpha_k \mu(A_k) - \sum_{k=1}^{\infty} \beta_k \mu(B_k)$$

für $f^+ = \sum_k \alpha_k \mathbb{1}_{A_k}$ und $f^- = \sum_k \beta_k \mathbb{1}_{B_k}$ mit $A_k, B_k \in \mathcal{A}$ und $\alpha_k, \beta_k \geq 0$, solange wenigstens eine der beiden Summen endlich ist. Wir schreiben auch oft kurz $\int_\Omega f \, d\mu$. Das Integral über eine Teilmenge $A \subset \mathcal{A}$ ist definiert als

$$\int_A f(\omega) \, d\mu(\omega) := \int_\Omega f(\omega) \mathbb{1}_A(\omega) \, d\mu(\omega).$$

Das Integral erfüllt die folgenden elementaren Eigenschaften:

(i) *Definitheit*: für $f \geq 0$ gilt $\int_\Omega f \, d\mu = 0$ genau dann, wenn $f = 0$ fast überall gilt;

(ii) *Monotonie*: $f \leq g$ fast überall impliziert $\int_\Omega f \, d\mu \leq \int_\Omega g \, d\mu$;

(iii) *Dreiecksungleichung*: $\left|\int_\Omega f \, d\mu\right| \leq \int_\Omega |f| \, d\mu$;

(iv) *Linearität*: $\int_\Omega \alpha f + g \, d\mu = \alpha \int_\Omega f \, d\mu + \int_\Omega g \, d\mu$.

Zwei wichtige Spezialfälle sind $(\Omega, \mathcal{B}, \lambda)$ mit $\Omega \subset \mathbb{R}^d$, $\mathcal{B}$ die Borel-Algebra, und λ das Lebesgue-Maß; in diesem Fall erhält man das *Lebesgue-Integral*. Ist $\Omega = \{\omega_1, \ldots, \omega_N\}$ eine endliche Menge und $\mu = \sum_{k=1}^N \alpha_k \delta_{\omega_k}$ für $\alpha_k \geq 0$ und $\delta_{\omega_k}(A) = 1$ falls $\omega_k \in A$ und 0 sonst, so gilt

$$\int_\Omega f(\omega) \, d\mu(\omega) = \sum_{k=1}^N \alpha_k f(\omega_k).$$

Wir definieren für eine messbare Funktion $f : \Omega \to \mathbb{R}$ nun

$$\|f\|_p := \left(\int_\Omega |f(\omega)|^p \, d\mu(\omega)\right)^{1/p}, \quad 1 \leq p < \infty,$$
$$\|f\|_\infty := \inf \{K \geq 0 \mid \mu(\{\omega \mid |f(\omega)| > K\}) = 0\}$$

und setzen für $1 \leq p \leq \infty$

$$L^p_\mu(\Omega) := \left\{f : \Omega \to \mathbb{R} \text{ messbar} \,\middle|\, \|f\|_p < \infty\right\}.$$

Ist $f \in L^1_\mu(\Omega)$, so sagt man, f ist *integrierbar*. Es gelten die üblichen Ungleichungen, etwa die *Höldersche Ungleichung* für $f \in L^p_\mu(\Omega)$ und $g \in L^p_\mu(\Omega)$ mit $\frac{1}{p} + 1q = 1$:

$$\int_\Omega f(\omega) g(\omega) \, d\mu(\omega) \leq \left(\int_\Omega f(\omega)^p \, d\mu(\omega)\right)^{1/p} \left(\int_\Omega g(\omega)^q \, d\mu(\omega)\right)^{1/q}.$$

Ist das Maß aus dem Kontext offensichtlich, schreibt man auch nur $f \in L^p(\Omega)$. Das Integral vektorwertiger Funktionen $f : \Omega \to \mathbb{R}^d$ und die entsprechenden Räume $L^p_\mu(\Omega; \mathbb{R}^d)$ integrierbarer vektorwertiger Funktionen werden analog komponentenweise definiert.

Die Hauptsätze über integrierbare Funktionen betreffen nun die Vertauschbarkeit von Grenzwert und Integral.

Satz B.2 (Monotone Konvergenz, Beppo Levi). *Sei $\{f_n\}_{n\in\mathbb{N}}$ eine Folge integrierbarer Funktionen mit $f_n \nearrow f$ fast überall. Dann ist*

$$\lim_{n\to\infty} \int_\Omega f_n(\omega)\, d\mu(\omega) = \int_\Omega f(\omega)\, d\mu(\omega).$$

Satz B.3 (Lemma von Fatou). *Sei $\{f_n\}_{n\in\mathbb{N}}$ eine Folge messbarer Funktionen. Existiert eine integrierbare Funktion g mit $f_n \geq g$ fast überall für alle $n \in \mathbb{N}$, dann gilt*

$$\liminf_{n\to\infty} \int_\Omega f_n(\omega)\, d\mu(\omega) \geq \int_\Omega \liminf_{n\to\infty} f_n(\omega)\, d\mu(\omega).$$

Satz B.4 (Majorisierte Konvergenz, Lebesgue). *Sei $\{f_n\}_{n\in\mathbb{N}}$ eine Folge integrierbarer Funktionen mit $f_n \to f$ fast überall. Existiert eine integrierbare Funktion g mit $|f_n| \leq g$ fast überall für alle $n \in \mathbb{N}$, dann ist auch f integrierbar und es gilt*

$$\lim_{n\to\infty} \int_\Omega f_n(\omega)\, d\mu(\omega) = \int_\Omega f(\omega)\, d\mu(\omega).$$

Sind μ, ν zwei Maße auf dem messbaren Raum $(\Omega, \mathcal{A})$ und gilt für alle $A \in \mathcal{A}$ mit $\mu(A) = 0$ auch $\nu(A) = 0$, so heißt ν *absolutstetig* bezüglich μ.

Satz B.5 (Radon–Nikodym). *Seien μ, ν zwei σ-endliche Maße auf $(\Omega, \mathcal{A})$. Dann ist ν absolutstetig bezüglich μ genau dann, wenn eine meßbare Funktion $f : \Omega \to [0, \infty)$ existiert mit*

$$\nu(A) = \int_A f(\omega)\, d\mu(\omega) \qquad \text{für alle } A \in \mathcal{A}.$$

Die Funktion f wird *Dichte* von ν (bezüglich μ) genannt; ist ν ein Wahrscheinlichkeitsmaß, spricht man auch von der *Wahrscheinlichkeitsdichte*.

Sind $(\Omega_1, \mathcal{A}_1, \mu_1)$ und $(\Omega_2, \mathcal{A}_2, \mu_2)$ zwei Wahrscheinlichkeitsräume, so können wir den *Produktraum* $(\Omega, \mathcal{A}, \mu)$ definieren via

$$\begin{aligned}
\Omega &:= \Omega_1 \times \Omega_2, \\
\mathcal{A} &:= \mathcal{A}_1 \otimes \mathcal{A}_2 := \{A_1 \times A_2 \mid A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}, \\
\mu &:= \mu_1 \otimes \mu_2, \quad \mu(A_1 \times A_2) := \mu(A_1)\mu(A_2).
\end{aligned}$$

Für das Integral auf dem Produktraum gilt der

Satz B.6 (Fubini). *Ist $f : \Omega \to \mathbb{R}$ meßbar bezüglich $\mathcal{A}$ und $f \in L^1_\mu(\Omega)$, dann gilt*

$$\begin{aligned}
\int_\Omega f(\omega)\, d\mu(\omega) &= \int_{\Omega_2} \left(\int_{\Omega_1} f(\omega_1, \omega_2)\, d\mu_1(\omega_1) \right) d\mu_2(\omega_2) \\
&= \int_{\Omega_1} \left(\int_{\Omega_2} f(\omega_1, \omega_2)\, d\mu_2(\omega_2) \right) d\mu_1(\omega_1).
\end{aligned}$$

B.3 ZUFALLSVARIABLEN

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Eine messbare Funktion $X : \Omega \to \mathbb{R}$ (versehen mit der Borel-Algebra) nennt man dann auch *Zufallsvariable*. Für $B \in \mathcal{B}$ setze

$$\mathbb{P}[X \in B] := \mu(X^{-1}[B]) = \int_{-\infty}^{\infty} \mathbb{1}_B(X(\omega))\, d\,\mathbb{P}(\omega).$$

Wir nennen das Wahrscheinlichkeitsmaß $\mathbb{P}_X := \mathbb{P} \circ X^{-1}$, $B \mapsto \mathbb{P}[X \in B]$, *Verteilung* von X. Wir schreiben dann auch $X \sim \mathbb{P}_X$ und sagen, X ist *verteilt nach* $\mathbb{P}_X$. Nimmt X höchstens abzählbar viele Werte an, spricht man von einer *diskreten Zufallsvariable* mit einer *diskreten Verteilung*. Ist $f : \mathbb{R} \to \mathbb{R}$ messbar, so ist auch $f(X)$ eine Zufallsvariable.

Da die Borel-Algebra $\mathcal{B}$ durch die halboffenen Mengen erzeugt wird, genügt es, Mengen der Form $B = \{X \le t\} := X^{-1}[\{x \in \mathbb{R} \mid x \le t\}]$ zu betrachten. Wir schreiben dann auch kurz $\mathbb{P}[X \le t]$, und analog für ähnliche Mengen. Die Abbildung $t \mapsto \mathbb{P}[X \le t]$ heißt *Verteilungsfunktion* von X; diese legt die Zufallsvariable eindeutig fest. Beispiele sind die *Normalverteilung* $\mathcal{N}_{\mu,\sigma^2}$ für $\mu, \sigma \in \mathbb{R}$ mit

$$\mathbb{P}[X \le t] = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{t} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx,$$

sowie die *uniforme Verteilung* $\mathcal{U}_{a,b}$ für $a < b$ mit

$$\mathbb{P}[X \le t] = \begin{cases} 0 & \text{falls } t \le a, \\ \frac{t-a}{b-a} & \text{falls } t \in (a,b), \\ 1 & \text{falls } t \ge b. \end{cases}$$

Eine Familie $\{X_i\}_{i \in I}$ heißt *identisch verteilt*, wenn $\mu_{X_i} = \mu_{X_j}$ für alle $i, j \in I$ gilt. Sie heißt *unabhängig verteilt*, falls für alle endlichen Teilmengen $J \subset I$ gilt

$$\mathbb{P}\left[\bigcap_{j \in J} \{X_j \in A_j\}\right] = \prod_{j \in J} \mathbb{P}[X_j \in A_j] \quad \text{für alle } A_j \in \mathcal{A},\ j \in J.$$

Für $X \in L^1_\mu(\Omega)$ ist der *Erwartungswert* definiert als

$$\mathbb{E}[X] := \int_\Omega X(\omega)\, d\mu(\omega).$$

Beispielsweise ist $\mathbb{E}[X] = \mu$ für $X \sim \mathcal{N}_{\mu,\sigma^2}$ und $\mathbb{E}[X] = \frac{a+b}{2}$ für $X \sim \mathcal{U}_{a,b}$.

Direkt aus den Eigenschaften des Integrals folgt die Linearität, Monotonie, und Dreiecksungleichung des Erwartungswerts. Haben zwei Zufallsvariablen die gleiche Verteilung, so haben sie auch den gleichen Erwartungswert. Sind $\{X_1, \ldots, X_n\}$ unabhängig, so gilt

$$\mathbb{E}\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} \mathbb{E}[X_i].$$

Weiter gilt die *Markov-Ungleichung*

$$\mathbb{P}[|X| \geq \varepsilon] \leq \frac{\mathbb{E}[|X|]}{\varepsilon} \qquad \text{für alle } \varepsilon > 0. \tag{B.1}$$

Ist $f : \mathbb{R} \to \mathbb{R}$ konvex, so gilt außerdem die *Jensen-Ungleichung*

$$\mathbb{E}[f(X)] \geq f(\mathbb{E}[X]). \tag{B.2}$$

Für $X \in L^1_\mu(\Omega)$ ist die *Varianz* definiert als

$$\mathbb{V}[X] := \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}[(X - \mathbb{E}[X])^2] \geq 0.$$

Beispielsweise ist $\mathbb{V}[X] = \sigma^2$ für $X \sim \mathcal{N}_{\mu,\sigma^2}$ und $\mathbb{V}[X] = \frac{(a+b)^2}{12}$ für $X \sim \mathcal{U}_{a,b}$.

Sind $\{X_1, \ldots, X_n\}$ unabhängig, so gilt (*Bienaymé-Gleichung*)

$$\mathbb{V}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{V}[X_i].$$

Anwenden der Markov-Ungleichung auf $(X - \mathbb{E}[X])^2$ liefert die *Chebyshev-Ungleichung*

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq \varepsilon] \leq \frac{\mathbb{V}[X]}{\varepsilon^2} \qquad \text{für alle } \varepsilon > 0. \tag{B.3}$$

Grenzwertsätze machen eine Aussage über das asymptotische Verhalten von Mittelwerten von Zufallsvariablen. Das folgende Resultat ist bekannt als das *starke Gesetz der großen Zahl.*

Satz B.7 (Etemadi). *Sei $(\Omega, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum und $\{X_n\}_{n\in\mathbb{N}} \subset L^1_\mu(\Omega)$ eine Folge von unabhängig und identisch verteilten Zufallsvariablen. Dann gilt*

$$\mathbb{P}\left[\limsup_{n\to\infty} \frac{1}{n}\left|\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right| = 0\right] = 1.$$

B.4 BEDINGTE ERWARTUNG

Sei $(\Omega, \mathcal{A}, \mu)$ ein Wahrscheinlichkeitsraum und $\mathcal{F} \subset \mathcal{A}$ eine Unter-σ-Algebra. Ist X eine integrierbare Zufallsvariable und ist Y eine Zufallsvariable, für die gilt

(i) Y ist messbar bezüglich $\mathcal{F}$;

(ii) $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]$ für alle $A \in \mathcal{F}$;

dann heißt Y *bedingter Erwartungswert* von X gegeben $\mathcal{F}$, und wir schreiben $\mathbb{E}[X \mid \mathcal{F}] := Y$. Der bedingte Erwartungswert existiert für alle $X \in L^1_\mu(\Omega)$ und alle $\mathcal{F} \subset \mathcal{A}$ und ist eindeutig bis auf μ-Nullmengen. (Alle Eigenschaften gelten daher immer nur fast sicher.) Direkt aus der Definition (wähle $A = \Omega \in \mathcal{F}$) folgt

$$\mathbb{E}\left[\mathbb{E}[X \mid \mathcal{F}]\right] = \mathbb{E}[X].$$

Ist X unabhängig von $\mathcal{F}$, so gilt sogar

$$\mathbb{E}[X \mid \mathcal{F}] = \mathbb{E}[X].$$

Der bedingte Erwartungswert erbt die Eigenschaften des Erwartungswerts wie Linearität, Monotonie, Dreiecksungleichung; ebenso gilt für $f : \mathbb{R} \to \mathbb{R}$ konvex die *bedingte Jensen-Ungleichung*

$$\mathbb{E}[f(X) \mid \mathcal{F}] \geq f(\mathbb{E}[X \mid \mathcal{F}]). \tag{B.4}$$

Ist $\mathcal{G} \subset \mathcal{F}$ eine weitere Unter-σ-Algebra, gilt die *Turmeigenschaft*

$$\mathbb{E}\left[\mathbb{E}[X \mid \mathcal{F}] \mid \mathcal{G}\right] = \mathbb{E}\left[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{F}\right] = \mathbb{E}[X \mid \mathcal{G}].$$

Für eine Zufallsvariable Z definieren wir den bedingten Erwartungswert von X gegeben Z als $\mathbb{E}[X \mid Z] := \mathbb{E}[X \mid \sigma(Z)]$, wobei $\sigma(Z) := \left\{X^{-1}[B] \,\middle|\, B \in \mathcal{B}\right\}$ die von Z erzeugte σ-Algebra ist (die kleinste σ-Algebra, bezüglich der Z messbar ist). Analog definiert man die von einer Menge $\{X_1, \ldots, X_n\}$ erzeugte σ-Algebra als die kleinste σ-Algebra, bezüglich der X_k messbar ist für alle $k = 1, \ldots, n$.

Eine Folge $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ von Unter-σ-Algebren mit $\mathcal{F}_1 \subset \mathcal{F}_2 \cdots \subset \mathcal{F}$ heißt *Filtration*. Eine Folge $\{X_n\}_{n\in\mathbb{N}}$ von Zufallsvariablen heißt *adaptiert* an die Filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$, wenn X_n messbar ist bezüglich $\mathcal{F}_n$ für alle $n \in \mathbb{N}$. Dies gilt insbesondere offensichtlich für die *natürliche Filtration* $\mathcal{F}_n = \sigma(\{X_1, \ldots, X_n\})$ für alle $n \in \mathbb{N}$.

Literatur

M. Ang, J. Sun & Q. Yao (2018), On the dual representation of coherent risk measures, *Ann. Oper. Res.* 262(1), 29–46, DOI: 10.1007/s10479-017-2441-3.

S. Angerhausen (2022), A Stochastic Primal-Dual Proximal Splitting Method for Risk-averse Optimal Control of PDEs, Diss., Universität Duisburg-Essen, URL: https://primo.uni-due.de/permalink/49HBZ_UDE/1uttatt/alma99208412927406446.

J. F. Bonnans (2019), *Convex and Stochastic Optimization*, English, Universitext, Springer, DOI: 10.1007/978-3-030-14977-2.

M. Brokate & G. Kersting (2019), *Maß und Integral*, 2. Aufl., Math. Kompakt, Cham: Birkhäuser, DOI: 10.1007/978-3-0348-0988-7.

N. Calkin & H. S. Wilf (2000), Recounting the rationals, *Am. Math. Mon.* 107(4), 360–363, DOI: 10.2307/2589182.

C. Clason & T. Valkonen (2020), Introduction to Nonsmooth Analysis and Optimization, ARXIV: 2001.00216.

A. Klenke (2020), *Wahrscheinlichkeitstheorie*, 4. Aufl., Masterclass, Berlin: Springer Spektrum, DOI: 10.1007/978-3-662-62089-2.

T. Pennanen & A.-P. Perkkiö (2024), *Convex Stochastic Optimization, Dynamic programming and duality in discrete time*, Bd. 107, Probab. Theory Stoch. Model. Cham: Springer, DOI: 10.1007/978-3-031-76432-5.

G. O. Roberts & R. L. Tweedie (1996), Exponential convergence of Langevin distributions and their discrete approximations, *Bernoulli* 2(4), 341–363, DOI: 10.2307/3318418.

R. T. Rockafellar & R. J.-B. Wets (1998), *Variational Analysis*, Bd. 317, Grundlehren der mathematischen Wissenschaften, Springer, DOI: 10.1007/978-3-642-02431-3.

J. O. Royset & R. J.-B. Wets (2021), *An Optimization Primer*, English, Springer Ser. Oper. Res. Financ. Eng. Springer, DOI: 10.1007/978-3-030-76275-9.

A. Shapiro (2013), Consistency of sample estimates of risk averse stochastic programs, *J. Appl. Probab.* 50(2), 533–541, DOI: 10.1239/jap/1371648959.

A. Shapiro, D. Dentcheva & A. Ruszczynski (2021), *Lectures on Stochastic Programming. Modeling and Theory*, 3. Aufl., Bd. 28, MPS/SIAM Ser. Optim. SIAM, DOI: 10.1137/1.9781611976595.