UNIVERSITÄT HEIDELBERG

Bachelorarbeit

Präsentiert zum Abschluss des

Bachelor of Science Mathematik

Marvin Koß

Immatrikulationsnummer: 3495081

Seestraße 35

69214 Eppelheim

marvinklg@outlook.de

Zusammenfassung

Partikelschwarmoptimierung ist ein gradientfreier Optimierungsalgorithmus, der in diversen Anwendungsgebieten zum Einsatz kommt (Bonyadi and Michalewicz, 2017). Er basiert auf der gemeinsamer Suche mehrerer Partikel nach einem Optimum in der Landschaft einer Kostenfunktion, die dabei ihre beste bisher gefundenen Positionen speichern. Die weitere Suche basiert auf diesen bisherigen Optima. Für die Analyse der Konvergenz des Optimierers müssen Annahmen auf sowohl die Beschaffenheit der Kostenfunktion als auch auf den Algorithmus selbst gemacht werden. Eine mögliche Methode dafür ist die Formulierung einer regularisierten, zeitlich kontinuierlichen Version des Algorithmus', welche die Anwendung von Werkzeugen aus dem Feld der stochastischen Differentialgleichungen sowie der Maßtheorie erlaubt. Diese Arbeit untersucht einen kürzlich in (Grassi et al., 2021; Huang and Qiu, 2021) veröffentlichten Beweis dafür, dass sich die Verteilungen der Partikel einer solchen Formulierung, die dem Algorithmus zusätzlich zum Verleihen zeitlicher Kontinuität die Speicherkomponenten nimmt, im sogenannten Mean-Field Limit gut verhalten. Das Mean-Field Limit bezeichnet dabei die Analyse makroskopischer Eigenschaften der Verteilung des Schwarms im Grenzwert der Partikelzahl gegen unendlich, wobei die Vielzahl der Partikel den Einfluss der einzelnen mikroskopischen Interaktionseffekte abschwächt. Der untersuchte Beweis zeigt unter Verwendung maßtheoretischer Argumente und bestimmten Annahmen auf die Kostenfunktion, dass die Folge der empirischen Maße des mikroskopischen Systems mit zunehmender Partikelzahl nicht *gegen unendlich auswandert*, sondern gegen die eindeutige schwache Lösung der sogenannten Fokker-Planck Gleichung des Mean-Field Systems konvergiert. Das Ergebnis des Beweises stellt einen Zwischenschritt in der Konvergenzanalyse dar, und kann zum Beispiel zum Erhalten globaler Konvergenz und einer exponentiellen Konvergenzrate des regularisierten Algorithmus' verwendet werden (Huang et al., 2022).

Inhaltsverzeichnis

1 Einleitung **2**

1.1 Verwandte Forschung 4

1.1.1 PSO 4

1.1.2 Mean-Field Analysen von Partikelsystemen 5

1.2 Motivation für eine SDGL-Perspektive 5

1.3 Gliederung der Arbeit 6

2 Mathematischer Hintergrund **8**

2.1 Konventionen 8

2.2 Stochastische Differenzialgleichungen 9

2.2.1 Stochastische Prozesse und Martingale 9

2.2.2 Kanonische Konstruktion 13

2.2.3 Brownsche Bewegung 14

2.2.4 Ito-Integral 16

2.2.5 Ito-Kalkül 19

2.2.6 Starke Lösungen von SDGLn 22

2.2.7 Diskretisierungen von Differentialgleichungen 24

2.3 Konvergenz von Maßen 25

2.3.1 Schwache Konvergenz 26

2.3.2 Straffheit 30

2.4 Kolmogorovgleichungen 32

3 SDE-Formulierung von PSO **35**

3.1 Kanonische Partikelschwarmoptimierung 35

3.2 Stochastische Differentielle Partikelschwarmoptimierung 37

3.2.1 Klasse der Kostenfunktionen 37

3.2.2 Reduziertes SD-PSO 38
3.2.3 Vollständiges SD-PSO 40
3.2.4 Entsprechung des *vollständigen SD-PSO* und *CanPSO* 42
3.3 Regularität des reduzierten SD-PSO 44

4 Mean-Field Limit **49**
4.1 Präliminarien 50
4.2 Ziel des Mean-Field Beweises 50
4.3 Momentenabschätzung 55
4.4 Straffheit der empirischen Maße 63
4.5 Konvergenz der empirischen Maße 65
4.5.1 Grenzwerte konvergenter Teilfolgen lösen die FPG schwach . . . 67
4.5.2 Eindeutigkeit des Grenzwerts 77
4.6 Abschluss des Beweises 81

5 Zusammenfassung **82**
5.1 Ausblick 83

6 Eigenständigkeitserklärung **84**

References **84**

Kapitel 1

Einleitung

Partikelschwarmoptimierung (PSO) (Kennedy and Eberhart, 1995) ist ein globaler, gradientfreier Optimierungsalgorithmus, und wird weitgehend angewandt (Bonyadi and Michalewicz, 2017), zum Beispiel zur Optimierung Neuronaler Netzwerke (Shi et al., 2001; Huang et al., 2022). Die in der Literatur typischerweise (Bonyadi and Michalewicz, 2017) behandelte Formulierung des Algorithmus von (Shi and Eberhart, 1998) mit einem Trägheitshyperparameter wird **Kanonisches PSO** (*CanPSO*) genannt und im Folgenden rekapituliert.

Zum Finden eines Minimums einer Kostenfunktion $F : \mathbb{R}^D \to \mathbb{R}$ wird ein Schwarm von N Partikeln eingesetzt, deren jeweiliger Zustand zum Zeitpunkt $t \in \mathbb{N}_0$ ein Quadrupel $(X_t^i, V_t^i, Y_t^i, y_t^i), 1 \leq i \leq N$ ist. Partikel i hat Position $X_t^i \in \mathbb{R}^D$, Geschwindigkeit $V_t^i \in \mathbb{R}^D$, die *personal best position (pbest)* $Y_t^i \in \mathbb{R}^D$ und ein assoziertes *personal best value* $y_t^i \in \mathbb{R}^D$ mit $y_t^i = F(Y_t^i)$. Zusätzlich verwaltet der Optimierer ein Tupel $(\overline{Y}_t, \overline{y}_t)$, die *global best position (gbest)* $\overline{Y}_t \in \mathbb{R}^D$ und ein assoziertes *global best value* $\overline{y}_t \in \mathbb{R}^D$ mit $\overline{y}_t = F(\overline{Y}_t)$. Der Optimierungsalgorithmus geht nun nach zufälliger Initialisierung der Partikelpositionen und -geschwindigkeiten (z.B. nach Normalverteilung oder Gleichverteilung auf einem Kompaktum) in jedem diskreten Zeitschritt $t \in \mathbb{N}_0$ in zwei Updateschritten vor:

1. Zunächst wird die Kostenfunktion in den Partikelpositionen X_t^i ausgewertet, woraufhin im Fall einer Verbesserung Y_t^i und $\overline{Y}_t$ und ihre assozierten Werte y_t^i und $\overline{y}_t$ aktualisiert werden; andernfalls behalten diese Variablen ihre Werte aus dem vorherigen Zeitschritt.

2. Im zweiten Schritt des Updates wird jeder Partikel um einen Geschwindigkeitsvektor V_t^i geupdatet, der eine Summe dreier je durch einen Hyperparameter gewichteter Geschwindigkeitskomponenten ist, die auf den aktuellen *pbest* und *gbest* basieren:

1. Eine Trägheitskomponente $V_{inertial} = \omega \cdot V_{t-1}^i$, die von einem in (Shi and Eberhart,

1998) eingeführten Trägheits-hyperparameter $\omega \in [0,1]$ skaliert wird. ω mediiert zwischen *Exploration* ($\omega \approx 1$) und *Exploitation*[1] ($\omega \approx 0$) (Bonyadi and Michalewicz, 2017, Sec. 4.1.2) und wird daher auch häufig als abnehmende Funktion von t gewählt (Bonyadi and Michalewicz, 2017; Shi and Eberhart, 1998).

2. Eine stochastische Komponente V_{pbest}, die von einem skalaren Hyperparemeter c_1 modifiziert wird. Für die Komponenten des zum *pbest* zeigenden Vektor $Y_t^i - X_{t\,i}$ werden D Zuvallsvariablen $r_d^1 \overset{\text{u.i.d}}{\sim} \mathcal{U}[0,1], d = 1,\ldots,n$ realisiert, woraufhin in jeder Dimension $d = 1,\ldots,D$ die V_{pbest} Komponente d die um $c_1 \cdot r_d^1$ skalierte Komponente $(Y_t^i - X_t^i)_d$ ist.

3. Eine zweite stochastische Komponente V_{gbest}, deren Komponente in jeder Dimension $d = 1,\ldots,D$ entsprechend die um $c_2 \cdot r_d^2$ skalierte Komponente $(\overline{Y}_t^i - X_t^i)_d$ ist, mit $r_d^2 \overset{\text{u.i.d}}{\sim} \mathcal{U}[0,1], d = 1,\ldots,n$.

Die Prozedur fährt so lange fort, bis z.B. $\bar{y}_t$ konvergiert ist oder eine Maximalzahl an Schritten erreicht wurde. Algorithmus 1 gibt einen Pseudocode für die beschriebene Prozedur an.

[1]Die Terme *Exploration* vs. *Exploitation* bezeichnen in der Optimierungsliteratur den Kompromiss zwischen der Suche nach besseren Lösungen durch
1. eine umfassendere Erkundschaftung des Suchraums (Exploration) vs.
2. durch Suchen in der Nähe eines lokales Optimums (Exploitation)

Algorithm 1 Kanonisches PSO $F : X \subset \mathbb{R}^D \to \mathbb{R}$

procedure PSO($F, N, D, \omega, c_1, c_2$)
  $t \leftarrow 0 \quad \mathrm{X}_t^i, V_t^i, Y_t^i, \overline{Y}_t \leftarrow$random_vector(D), $\forall i = 1, \ldots, n$
  $y_t^i, \overline{y}_t \leftarrow \infty$
  while Abbruchkriterium($\overline{y}_t$) nicht erfüllt **do**
    for $i = 1, \ldots, N$ **do**
      $y \leftarrow \mathrm{F}(\mathrm{X}_t^i)$
      if $y < y_t^i$ **then**
        $Y_t^i \leftarrow \mathrm{X}_t^i;\ \ y_t^i \leftarrow y$
      if $y < \overline{y}_t$ **then**
        $\overline{Y}_t \leftarrow \mathrm{X}_t^i;\ \ \overline{y}_t \leftarrow y$
    for $i = 1, \ldots, N$ **do**
      for $d = 1, \ldots, D$ **do**
        $r_1, r_2 \sim U[0,1]$
        $V_{inertial} \leftarrow \omega \cdot (V_{t-1}^i)_d$
        $V_{personal} \leftarrow \mathrm{c}_1 \cdot r_1 \cdot (Y_t^i - X_t^i)_d$
        $V_{global} \leftarrow \mathrm{c}_2 \cdot r_2 \cdot (\overline{Y}_t - X_t^i)_d$
        $(V_t^i)_d \leftarrow \mathrm{V}_{inertial} + V_{personal} + V_{global}$
        $(X_t^i)_d \leftarrow (\mathrm{X}_{t-1}^i)_d + (V_t^i)_d$
    $t \leftarrow$t+1
  return $\overline{Y}_t$

1.1 Verwandte Forschung

Da diese Arbeit PSO von einer Mean-Field Perspektive analysiert, werden verwandte Arbeiten aus diesen beiden Forschungsbereichen vorgestellt.

1.1.1 PSO

PSO ist dank einiger praktischer Eigenschaften, nämlich Gradientfreiheit und dass es ohne extensives Hyperparametertuning gute Performanz in unterschiedlichen Domänen hat, ein weitgehend angewandter Optimierungsalgorithmus (Shi et al., 2001, Sec. 3). Das Forschungsfeld zu PSO ist entsprechend umfangreich und daher kann ein kurzer Überblick über die Literatur dieser kaum gerecht werden. Wir verweisen auf (Poli et al., 2007) für eine Übersicht über Anwendungen von PSO und auf (Bonyadi and Michalewicz, 2017) für einen Überblick über Erweiterungen und theoretische Analysen des Algorithmus'. Beide Artikel betonen, dass es viele oft anwendungsorientierte Erweiterungen und Umformulierungen von PSO gibt, und dass theoretische Analysen von PSO zwecks Anwendbarkeit unterschiedlicher mathematischer Werkzeuge entsprechende vereinfachende Annahmen auf den Algorithmus machen; (Poli et al., 2007, Sec. 5) enthält eine Liste qualitativer Gründe für die Notwendigkeit dieser Simplifikationen.

1.1.2 Mean-Field Analysen von Partikelsystemen

(Jabin and Wang, 2017) gibt allgemeine Strategien für das Erhalten von Mean-Field Limits für stochastische Partikelsysteme, legt allerdings einen Fokus auf Systeme mit globalen Lipschitzkonstanten. In (Bolley et al., 2011b) wird ein iteratives Schema für Existenz und Eindeutigkeit der Lösung der Fokker-Planck Gleichung von Partikelsystemen mit konstantem Driftterm verwendet.

(Pinnau et al., 2017) führte **Consensus Based Optimization** (*CBO*) ein, eine ableitbare Formulierung von PSO mittels **Stochastischer Differentialgleichungen** (SDGLn); für eine einfache Behandlung unterscheidet sich *CBO* dadurch von *CanPSO*, dass es keinen Trägheitsparameter und kein *pbest* gibt, und das (regularisierte) *gbest* aus den aktuellen Partikelpositionen errechnet wird; es also keine Gedächtniseffekte gibt. Eine rigorose Perspektive auf das Mean-Field Limit von *CBO* folgte in (Carrillo et al., 2018); in dieser Arbeit wird das Mean-Field Limit von *CBO* jedoch lediglich postuliert, und nicht bewiesen. Ein Übergang von *CBO* zu auf SDGLn basierenden Formulierungen von *CanPSO* mit *pbest* und Trägheitsparameter, genannt *SD-PSO*, wurde von (Grassi and Pareschi, 2021) unternommen, zusammen mit numerischen Analysen des Mean-Field Limits. (Huang and Qiu, 2021) schließlich zeigten das Mean-Field Limit für *CBO* und *SD-PSO*; diese Arbeit ist zusammen mit (Grassi et al., 2021), worin die Mean-Field Ergebnisse genauer dargelegt und ein Überblick über die Literatur zu *SD-PSO* gegeben wird, Grundlage dieser Arbeit. (Huang et al., 2022) verwendeten das Mean-Field Ergebnis von *SD-PSO*, um für eine stärker eingeschränkte Klasse von Kostenfunktionen globale Konvergenz und eine exponentielle Konvergenzrate zu erhalten.

1.2 Motivation für eine SDGL-Perspektive

Es gibt zahlreiche Ansätze und mathematische Werkzeuge, um PSO bezüglich globaler Konvergenz und Konvergenzraten zu analysieren, wobei verschiedene vereinfachende

Annahmen gemacht werden. Die in dieser Arbeit genommene Perspektive von Stochastischen Differentialgleichungen (SDGLn) lässt mithilfe des für diese von Itô (1944) entwickelten Ito-Kalküls für zeitlich kontinuierliche Formulierungen des Algorithmus die Berechnung von hinreichend regulären Funktionen der resultierenden Prozesse zu. Insbesondere können *makroskopische* Aussagen getroffen werden, was passiert, wenn wir die Anzahl der Partikel N gegen unendlich gehen lassen, was intuitiv dazu führen sollte, dass zwei beliebige Partikel weniger Einfluss aufeinander haben sollten. (Huang, 2021, S.4) und insb. (Bolley et al., 2011b, Sec. 1.1) enthalten tiefergehende Überblicke über Mean-Field Analysen auf sozialer Dynamik basierender Systeme wie PSO.
Obwohl die Anwendung des Ito-Kalküls eine Neuformulierung des diskreten Algorithmus' 1 notwendig macht, weshalb sich erarbeitete theoretische Schlüsse nicht mehr ohne Weiteres auf den ursprünglichen Algorithmus übertragen lassen, hat die SDGL-Perspektive einige nützliche Eigenschaften:

1. Mit bestimmter Wahl von Zeitdiskretisierung sind mit *CanPSO* übereinstimmende (Grassi and Pareschi, 2021) und effiziente (Huang et al., 2022) Implementationen möglich. Ersteres werden wir in Abschnitt 3.2.4 sehen.

2. Dadurch ist eine numerische Validierung der theoretischen Ergebnisse möglich (Grassi and Pareschi, 2021).

3. Die Theoretische Analyse der regularisierten Formulierung hat Zugang zu einem breiteren Spektrum mathematischer Werkzeuge, insbesondere zum Ito-Kalkül und Partiellen Differentialgleichungen.

4. Es können starke Aussagen zur globalen Konvergenz und Konvergenzraten der regularisierten Formulierung gemacht werden (Huang et al., 2022).

1.3 Gliederung der Arbeit

Der Rest der Arbeit ist folgendermaßen gegliedert: Das folgende Kapitel 2 enthält notwendige mathematische Präliminarien. Hierbei folgen einer Einführung in die Theorie der stochastischen Differentialgleichungen, das Ito-Kalkül, eine Reihe maßtheoretischer Definitionen und Ergebnisse. Darauf folgend führt Kapitel 3 den ursprünglichen PSO-Algorithmus sowie die untersuchte regularisierte Version ein, und gibt einen Beweis für die eindeutige starke Lösbarkeit der regularisierten Formulierung. Der Hauptteil der Arbeit besteht in Kapitel 4, welches den Beweis für die Wohlbeschaffenheit des Mean-Field

Limits enthält. Zuletzt fasst Kapitel 5 die Ergebnisse der Arbeit zusammen und gibt einige Unterschiede zwischen dem hier gegebenen Beweis und dem originalen Beweis von (Grassi et al., 2021; Pinnau et al., 2017) an.

Kapitel 2

Mathematischer Hintergrund

Für die Analyse des Mean-Field Ergebnisses von (Grassi et al., 2021) gilt es, Stochastische Differentialgleichungen (SDGLn) und einige Konzepte aus der Maßtheorie rigoros einzuführen. Wir beginnen mit einer Einführung in SDGLn sowie die zentralen verwendeten Sätze des Ito-Kalküls in Abschnitt 2.2, gefolgt von den verwendeten Ergebnissen der Wahrscheinlichkeitstheorie in Abschnitt 2.3. Grundlegende topologische und wahrscheinlichkeitstheoretische Begriffe werden hierbei vorausgesetzt. Für Einführungen in diese Themen siehe zum Beispiel (Klenke, 2006; Elstrodt, 1996; Billingsley, 1999).

2.1 Konventionen

Im Verlauf dieses Kapitels und im Rest der Arbeit arbeiten wir in $\mathbb{R}^D$ stets mit dem Standardskalarprodukt $\langle\cdot,\cdot\rangle$ und der von ihr induzierten 2-Norm

$$\begin{aligned} |\cdot| : \mathbb{R}^D &\to \mathbb{R}_{\geq 0} \\ (x_1,\dots,x_D)^T &\mapsto \sqrt{\sum_{d=1}^{D} x_d^2}, \end{aligned}$$

die die nützliche Eigenschaft hat, dass

$$|[X^1,\dots,X^k]^T|^2 = \sum_{i=1}^{k} |X^i|^2.$$

Hier bezeichnet

$$
\begin{aligned}
[\cdot,\dots,\cdot] : \bigtimes_{i=1}^{k} \mathbb{R}^{n_i} &\to \mathbb{R}^{\sum_{i=1}^{k} n_i} \\
(x_1,\dots,x_k) &\mapsto [x_1,\dots,x_k]
\end{aligned}
$$

das Aneinanderfügen von Vektoren. Die im Verlauf der Arbeit hiermit erhaltenen Abschätzungen gelten trotz der Verwendung dieser spezifischen Norm wegen der Äquivalenz der Normen mit entsprechend angepassten Konstanten für beliebige Normen auf $\mathbb{R}^D$. Desweiteren bezeichnen Integrale ohne Subskript oder Grenzen stets Integrale über den Raum $\mathbb{R}^D$ oder Phasenraum $\mathbb{R}^{2D}$ (nicht aber über die Zeit); welches von diesen beiden gemeint ist, sollte aus dem verwendeten Maß hervorgehen.
Für zwei topologische Räume A und B bezeichnet $\mathcal{C}(A;B)$ den Raum der stetigen Funktionen von A nach B, $\mathcal{C}(A)$ bezeichnet $\mathcal{C}(A;\mathbb{R})$; $\mathcal{C}_b(A;B)$ denjenigen der Funktionen, die zusätzlich beschränkt sind; $\mathcal{C}_c(A,B)$ denjenigen der Funktionen, die stetig sind und einen kompakten Träger haben. Zuletzt bezeichnet $\mathcal{C}^k(A;B)$ den Raum der k-mal stetig differenzierbaren Funktionen.

2.2 Stochastische Differenzialgleichungen

Dieser Abschnitt rekapituliert anhand von (Øksendal, 2003, Kapitel 2, 3 und 4) die zentralen Definitionen und Ergebnisse des Ito-Kalküls. Dieses entwickelt einen Begriff von Integration bezüglich eines stochastischen Prozesses, der Brownschen Bewegung, die somit als Grundlage von Modellen in diversen Domänen der realen Welt auftretender Prozesse verwendet werden kann.

2.2.1 Stochastische Prozesse und Martingale

Für diesen Abschnitt sei $(\Omega,\mathcal{F},\mathbb{P})$ ein Wahrscheinlichkeitsraum und $\mathcal{T}$ eine (geordnete, möglicherweise überabzählbare) Indexmenge, beispielsweise $\mathcal{T}=\mathbb{N}$ oder $\mathcal{T}=[0,\infty)=\mathbb{R}_{\geq 0}$. Die grundlegenden Gegenstände unseres Interesses sind die sogenannten **stochastischen Prozesse**:

Definition 2.2.1: Stochastischer Prozess, Messbarkeit

Eine Familie von Zufallsvariablen $\{X_t\}_{t\in\mathcal{T}}$ auf $(\Omega,\mathcal{F},\mathbb{P})$ mit X_t mit Werten in einem Messraum $(E,\mathcal{E})$ heißt **stochastischer Prozess auf** $(\Omega,\mathcal{F},\mathbb{P})$ **mit Zeitbereich** $\mathcal{T}$ **und Zustandsraum** E.
Hierbei ist für alle $t\in\mathcal{T}$ X_t $\mathcal{F}-\mathcal{E}$**-messbar**, d.h. $\forall E\in\mathcal{E}: X^{-1}(E)\in\mathcal{F}$.
Falls $\mathcal{T}=[0,T]\subset\mathbb{R}_{\geq 0}$ für ein $T>0$, nennen wir $\{X_t\}_{t\in\mathcal{T}}$ einen stochastischen Prozess **in kontinuierlicher Zeit**; falls $\mathcal{T}\subset\mathbb{Z}$, ist der Prozess **in diskreter Zeit**.

Wir sehen somit, dass ein stochastischer Prozess nichts anderes als eine Menge von Zufallsvariablen auf einem gemeinsamen Wahrscheinlichkeitsraum ist, für die eine geordnete Indexmenge eine Reihenfolge angibt.
Eine interessante Frage ist, ob zwei stochastische Prozesse *gleich* sind. Es gibt diesbezüglich die folgenden beiden grundlegenden Begriffe:

Definition 2.2.2: Versionen und Ununterscheidbarkeit (Øksendal, 2003, 2.2.2)

Seien $\{X_t\}_{t\in\mathcal{T}}$ und $\{Y_t\}_{t\in\mathcal{T}}$ stochastische Prozesse auf $(\Omega,\mathcal{F},\mathbb{P})$. Dann sagen wir, dass $\{X_t\}_{t\in\mathcal{T}}$ eine **Version** oder **Modifikation** von $\{Y_t\}_{t\in\mathcal{T}}$ ist, wenn

$$X_t = Y_t \quad \mathbb{P}\text{-f.s.}, \quad \forall t\in\mathcal{T}.$$

Falls die stärkere Bedingung

$$X_t = Y_t \quad \forall t\in\mathcal{T}, \quad \mathbb{P}\text{-f.s.}.$$

gilt, heißen die beiden Prozesse **ununterscheidbar**.

In dem Fall, dass $\mathcal{T}$ abzählbar ist, fallen die beiden Begriffe zusammen (Klenke, 2006, Lemma 21.5); im Verlauf der Arbeit werden wir aber vorwiegend mit $\mathcal{T}\subset\mathbb{R}_{\geq 0}$ arbeiten; hier ist es wichtig, diese Unterscheidung zu machen. Auch dann gibt es aber eine Klasse von Prozessen, für die Modifikation identisch zu Ununterscheidbarkeit ist:

Lemma 2.2.1: Versionen fast sicher rechtsstetiger Prozesse (Klenke, 2006, Lemma 21.5)

Sei $\mathcal{T} \subset \mathbb{R}_{\geq 0}$ ein Intervall. Seien $X = \{X_t\}_{t\in\mathcal{T}}$ und $Y = \{Y_t\}_{t\in\mathcal{T}}$. Seien ferner X und Y $\mathbb{P}$-f.s. rechtsstetig, d.h. für $t \in \mathcal{T}$ und $t_n \downarrow t$ gilt jeweils

$$\lim_{n\to\infty} X_{t_n} = X_t,\ \mathbb{P}\text{-f.s.} \quad \text{und} \quad \lim_{n\to\infty} Y_{t_n} = Y_t,\ \mathbb{P}\text{-f.s.}.$$

Dann gilt:

$$X,Y \quad \text{Versionen} \iff X,Y \quad \text{ununterscheidbar}$$

Eine wichtige Klasse stochastischer Prozesse sind sogenannte **Markovprozesse**. Auch wenn wir im Verlauf der Arbeit Markovprozessen in stetiger Zeit begegnen werden, ist für uns nur der Fall $\mathcal{T} = \mathbb{N}$ interessant:

Definition 2.2.3: Diskreter Markovprozess

Sei auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ ein stochastischer Prozess $\{X_t\}_{t\in\mathcal{T}}$ mit diskretem Zeitbereich $\mathcal{T}$ und Werten in einem Messraum $(E, \mathcal{E})$ gegeben.
Dann heißt der Prozess **Markovprozess** (in diskreter Zeit), falls für alle $t \in \mathcal{T}$ und $B \in \mathcal{E}$ gilt:

$$\mathbb{P}(X_{t+1} \in B | X_t, X_{t-1}, X_{t-2} \dots) = \mathbb{P}(X_{t+1} \in B | X_t)$$

Für bestimmte Arten von Prozessen ist eine Charakterisierung der Information, die zum Zeitpunkt $t \in \mathcal{T}$ zu Verfügung steht, notwendig. Dies liefert der nachstehende Begriff:

Definition 2.2.4: Filtration

Eine **Filtration** von $(\Omega, \mathcal{F}, \mathbb{P})$ ist eine Familie von σ-Algebren $\{\mathcal{F}_t\}_{t\in\mathcal{T}}$ mit $\mathcal{F}_t \subset \mathcal{F}$ mit der Eigenschaft

$$\mathcal{F}_s \subset \mathcal{F}_t \quad \forall s \leq t \in \mathcal{T}.$$

$(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathcal{T}}, \mathbb{P})$ heißt dann **filtrierter Wahrscheinlichkeitsraum**.
Auf diesem Raum heißt ein stochastischer Prozess $\{X_t\}_{t\in\mathcal{T}}$, wobei $\forall t \in \mathcal{T}$ X_t Werte in einem gemeinsamen Messraum $(E, \mathcal{E})$ annehme, **adaptiert** an $\{\mathcal{F}_t\}_{t\in\mathcal{T}}$, falls für alle $t \in \mathcal{T}$ gilt, dass X_t $\mathcal{F}_t - \mathcal{E}$**-messbar** ist, d.h. $\forall A \in \mathcal{E} : X_t^{-1}(A) \in \mathcal{F}_t$.
Eine Filtration $\{\mathcal{F}_t\}_{t\in\mathcal{T}}$ heißt **erzeugt** von einem stochastischen Prozess $\{X_t\}_{t\in\mathcal{T}}$, falls $\forall t \in \mathcal{T}$ gilt: $\mathcal{F}_t = \sigma(X_t) := \sigma(X_s, s \leq t, s \in \mathcal{T}) := \{X_s^{-1}(A) | s \leq t, s \in \mathcal{T}, A \in \mathcal{E}\}$.

Eine Art von stochastischem Prozess, die eine solche Filtration benötigt, ist das **Martingal**:

Definition 2.2.5: Martingal (Liu and Röckner, 2015, Def. 2.2.4)

Sei auf einem filtrierten Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathcal{T}}, \mathbb{P})$ ein stochastischer Prozess $\{M_t\}_{t\in\mathcal{T}}$ mit Zeitbereich $\mathcal{T}$ und Werten in einem Messraum $(E, \mathcal{E})$ gegeben. Der Prozess heißt $\{\mathcal{F}_t\}_{t\in\mathcal{T}}$**-Martingal**, falls gilt:

1. $\{M_t\}_{t\in\mathcal{T}}$ ist an $\{\mathcal{F}_t\}_{t\in\mathcal{T}}$ adaptiert.
2. (**Martingaleigenschaft**) $\mathcal{E}[M_t | \mathcal{F}_s] = M_s$ $\mathbb{P}$-fast sicher $\forall s \langle t \in \mathcal{T}$.

Ein Martingal heißt **integrierbar**, falls $\mathbb{E}[|M_t|] < \infty$ für alle $t \in \mathcal{T}$.

Dieser Begriff lässt insofern eine Aussage darüber zu, ob ein Prozess eine Tendenz hat, sich bevorzugt in eine bestimmte Richtung zu bewegen; wenn das nicht so ist, ist er ein Martingal. Dies wird formalisiert über die Martingaleigenschaft, die von einem festen, *aktuellen* Zeitpunkt aus die erwartete Position in allen zukünftigen Zeitpunkten mit der aktuellen Position gleichsetzt, anhand der gegenwärtig verfügbaren Information, welche über die Filtration bestimmt ist.

2.2.2 Kanonische Konstruktion

Für das Arbeiten mit der gemeinsamen Verteilung $\mathbb{P}^X$ eines stochastischen Prozesses $X = \{X_t\}_{t\in\mathcal{T}}$ in kontinuierlicher Zeit $\mathcal{T} = \mathbb{R}_{\geq 0}$, wobei X_t Werte in einem Messraum $(E,\mathcal{E})$ hat, benötigen wir eine kanonische Konstruktion der gemeinsamen Verteilung $\mathbb{P}^X$ auf dem Produktraum. Dafür konstruieren wir zunächst eine ausreichend ausdrucksstarke σ-Algebra:

Definition 2.2.6: Kanonische Konstruktion (Johannes, 2021, Abschnitt 04.02)

Sei $\mathcal{T} \subset \mathbb{R}$, und $\{(E_t,\mathcal{B}(E_t))\}_{t_\mathcal{T}}$ eine Familie von Messräumen, die jeweils mit der Borel'schen σ-Algebra ausgestattet sind. Wir definieren für endliches $\mathcal{J}$ mit $\mathcal{J} \subset \mathcal{K} \subseteq \mathcal{T}$ und den **Produktraum** $E_\mathcal{J} := \bigtimes_{j\in\mathcal{J}} E_j$ die **Koordinatenabbildungen**:

$$\begin{aligned} \Pi^\mathcal{K}_\mathcal{J} &: E_\mathcal{K} \to E_\mathcal{J} \\ &(e_k)_{k\in\mathcal{K}} \mapsto (e_j)_{j\in\mathcal{J}}. \end{aligned}$$

Hiermit definieren wir die **endlichen Zylindermengen**:

$$\begin{aligned} \mathcal{Z}_\mathcal{J} &:= \{(\Pi^\mathcal{T}_\mathcal{J})^{-1}(A) | A \in \mathcal{B}(E)_\mathcal{J}\}, \\ \text{wobei} \quad \mathcal{B}(E)_\mathcal{J} &:= \bigotimes_{j\in\mathcal{J}} \mathcal{B}(E_j). \end{aligned}$$

Hiermit erhalten wir auf dem Produktraum $E_\mathcal{T} := \bigtimes_{t\in\mathcal{T}} E_t$ die **kanonische σ-Algebra**

$$\mathcal{B}_\mathcal{T} := \sigma\Big(\bigcup_{\substack{\mathcal{J}\subset\mathcal{T} \\ \text{endlich}}} \mathcal{Z}_\mathcal{J}\Big) \quad \Big(= \bigotimes_{t\in\mathcal{T}} \mathcal{B}(E_t) \text{ (Johannes, 2021, Bem. 04.26)}\Big). \tag{2.1}$$

Wir nennen $(E_\mathcal{T},\mathcal{B}_\mathcal{T})$ den (für E und $\mathcal{T}$) **kanonisch konstruierten Messraum**.

Mit dieser Konstruktion der σ-Algebra auf $\mathbb{R}^D_\mathcal{T}$ definieren wir die endlich-dimensionalen Verteilungen eines Prozesses:

Definition 2.2.7: Endlichdimensionale Verteilungen eines Prozesses (Johannes, 2021)

Sei $(\Omega, \mathscr{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X = \{X_t\}_{t \in \mathcal{T}}$ ein stochastischer Prozess auf diesem Raum mit Verteilung $\mathbb{P}^X$ auf einem kanonisch konstruierten Messraum $(E_\mathcal{T}, \mathcal{B}_\mathcal{T}))$. Für endliches $\mathcal{J} \subset \mathcal{T}$ sei $\mathbb{P}^X_\mathcal{J} := \mathbb{P}^X \circ (\Pi^\mathcal{T}_\mathcal{J})^{-1}$ die gemeinsame Verteilung von $\Pi_\mathcal{J} \circ X$. Die Familie von Wahrscheinlichkeitsmaßen$\{\mathbb{P}^X_\mathcal{J}\}_{\mathcal{J} \subset \mathcal{T}, \mathcal{J} \text{ endlich}}$ heißt dann die Familie der **endlichdimensionalen Verteilungen** von $\mathbb{P}^X$.

Die endlichdimensionalen Verteilungen der Form $\mathbb{P}^X_\mathcal{J}$ bestimmen die Verteilung $\mathbb{P}^X$ eines Prozesses $X = \{X_t\}_{t \geq 0}$ in kontinuierlicher Zeit **eindeutig bis auf Modifikation** durch ihren **projektiven Limes**:

Satz 2.2.1: Erweiterungssatz von Kolmogorov (Klenke, 2006, Satz 14.39)

Sei $\mathcal{T} \subset \mathbb{R}$, und $(E_\mathcal{T}, \mathcal{B}_\mathcal{T}))$ ein kanonisch konstruierter Messraum. Sei ferner $\{\mathbb{P}_\mathcal{J}\}_{\mathcal{J} \subset \mathcal{T}}$ eine Familie von Wahrscheinlichkeitsmaßen, sodass für endliche $\mathcal{J} \subset \mathcal{K} \subset \mathcal{T}$ gilt:

$$\mathbb{P}_\mathcal{J} = \mathbb{P}_\mathcal{K} \circ (\Pi^\mathcal{K}_\mathcal{J})^{-1}$$

Dann gibt es ein eindeutig bestimmtes Wahrscheinlichkeitsmaß $\mathbb{P}$, **projektiver Limes** der Familie genannt, auf $(E_\mathcal{T}, \mathcal{B}_\mathcal{T})$ mit $\mathbb{P}_\mathcal{J} = \mathbb{P} \circ (\Pi^\mathcal{T}_\mathcal{J})^{-1}$ für jedes endliche $\mathcal{J} \subset \mathcal{T}$.

Die endlich dimensionalen Verteilungen erlauben uns also, einen Prozess eindeutig bis auf Modifikation zu definieren. Dies wird im folgenden Abschnitt für die D-dimensionale Brownsche Bewegung getan.

2.2.3 Brownsche Bewegung

In diesem Abschnitt wird in kontinuierlicher Zeit $\mathcal{T} \subset \mathbb{R}_{\geq 0}$ und mit Werten in dem oben kanonisch konstruierten Messraum $(\mathbb{R}^D_\mathcal{T}, \mathcal{B}_\mathcal{T})$ gearbeitet.

Ein grundlegender stochastischer Prozess ist die sogenannte **Brownsche Bewegung**. Sie kann auf unterschiedliche Weisen definiert werden; hier wird sie über die oben definierten **endlichdimensionalen Verteilungen** gemäß Satz 2.2.1 eindeutig bis auf Modifikation konstruiert.

Satz 2.2.2: Existenz Brownscher Bewegung (Øksendal, 2003, 11)

Sei $x_0 \in \mathbb{R}^D$ ein festes Anfangswert. Wir definieren für $x \in \mathbb{R}^D$ und $t > 0$

$$p(t, x_0, x) = \left[\frac{1}{\sqrt{2\pi t}}\right]^D \cdot \exp - \frac{|x - x_0|^2}{2t}$$

Dann existiert ein stochastischer Prozess $B = \{B_t\}_{t \geq 0}$ auf einem W-Raum $(\Omega, \mathscr{A}, \mathbb{P})$ sodass für $k \in \mathbb{N}$ und endliches $\mathcal{J} := \{0 \leq t_1 \leq \ldots \leq t_k\} \subset \mathbb{R}_{\geq 0}$ die **endlichdimensionale Verteilung** von B gegeben ist durch

$$\begin{aligned} &\mathbb{P}^B_{\mathcal{J}}(B_{t_1} \in A_1, \cdot, B_{t_k} \in A_k) \\ &= \int\limits_{A_1 \times \ldots \times A_k} p(t_1, x_0, x_1) \cdots p(t_k - t_{k-1}, x_{k-1}, x_k) dx_1 \cdots dx_k \end{aligned} \tag{2.2}$$

Die Verteilung von B ist dann durch Satz 2.2.1 eindeutig bis auf Modifikation gegeben durch den projektiven Limes $\mathbb{P}^B$ der $\mathbb{P}^B_{\mathcal{J}}$ der obigen Form.

Definition 2.2.8: Definition Brownscher Bewegung

Ein Prozess $B = \{B_t\}_{t \geq 0}$ heißt *eine Version* **Brownscher Bewegung**, falls seine endlichdimensionalen Verteilungen $\mathbb{P}^B_{\mathcal{J}}$ (2.2) erfüllen.

Wir sehen, dass die Verteilung der Brownschen Bewegung zum Zeitpunkt t eine *isotrope* multivariate Normalverteilung $\mathcal{N}(x, t\mathbb{I})$ ist[1]. Weitere zentrale Eigenschaften der Brownschen Bewegung fasst der folgende Satz zusammen:

[1] Hier ist $\mathbb{I} \in \mathbb{R}^{D \times D}$ die Identität.

Satz 2.2.3: Eigenschaften Brownscher Bewegung (Øksendal, 2003, S. 12-14)

1. $\mathbb{P}(B_0 = x_0) = 1$
2. Die D-dimensionale Brownsche Bewegung mit $B_t = (B_t^1, \dots, B_t^D)$ hat Komponenten $B_d, 1 \leq d \leq D$, die D unabhängige 1-dimensionale Brownsche Bewegungen sind, also $\mathop{\perp\!\!\!\perp}_{d=1}^{D} B_d$.
3. (**Gauß-Prozess**) $\{B_t\}_{t \geq 0}$ ist ein Gauß-Prozess. Genauer folgt die endlich dimensionale Verteilung für $0 \leq t_1 \leq \ldots \leq t_k$

$$\mathbb{R}^{Dk} \ni (B_{t_1}, \dots, B_{t_k}) \sim \mathcal{N}((x, \dots, x), \Sigma),$$

   wobei $\Sigma \in \mathbb{R}^{Dk \times Dk}$ eine (pos. def., symm.) Blockmatrix ist, deren $n \times D$-Blöcke die Form $\Sigma_{ij} = \min(t_i, t_j)\mathbb{I}_D, 1 \leq i, j \leq k$ haben.
4. (**Unabhängigkeit der Inkremente**) $B_t - B_s$ ist unabhängig von $\mathscr{A}_s$ falls $t > s$.
5. (**Stationarität**) B_t ist stationär, d.h. die endlich dimensionale Verteilung $(B_{t_1+t}, \dots, B_{t_k+t})$ hängt nicht von t ab.
6. Die **Pfade** $t \to B_t$ sind nicht differenzierbar in t_0 für alle $t_0 \in \mathcal{T}$, $\mathbb{P}$-fast sicher.
7. (**Kolmogorov-Tschenzow**) Es existiert eine Version von Brownscher Bewegung, genannt **Wiener-Prozess** $\{W\}_{t \in \mathcal{T}}$, so dass die **Pfade** $t \to W_t$ $\mathbb{P}$-f.s. punktweise stetig sind.

Im Rest der Arbeit nehmen wir an, dass B_t punktweise stetige Pfade hat, dass wir also mit dem Wiener-Prozess arbeiten. Dies ist eine Voraussetzung für die Aussagen in den Abschnitten 2.2.6 und 2.4.

2.2.4 Ito-Integral

Die in 2.2.3 genannten Eigenschaften, insbesondere die fehlende Differenzierbarkeit, sind Grund dafür, dass bezüglich des Maßes $\mathbb{P}^B$ der Integrationsbegriff von Riemann-Stieltjes scheitert. Es wurden dafür neue Integralbegriffe eingeführt. Wir verwenden in dieser Arbeit das Ito-Integral Itô (1944).

Eindimensional

Im Folgenden sei $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathcal{T}}, \mathbb{P}^x)$ ein durch die von einer fixen Version $B = \{B_t\}_{t\in\mathcal{T}}$ der *eindimensionalen* Brownschen Bewegung erzeugten Filtration $\{\mathcal{F}_t\}_{t\in\mathcal{T}} = \{\sigma(B_t)\}_{t\in\mathcal{T}}$ filtrierter Wahrscheinlichkeitsraum. Als nächstes beschränken wir uns auf eine Klasse von Funktionen, die regulär genug sind, um als Integranden bezüglich $\mathbb{P}^B$ handhabbar sein zu können.

Definition 2.2.9: Menge der Integranden (Øksendal, 2003, 25)

Für $0\langle S \leq T$ sei dann $\mathcal{V} = \mathcal{V}(S,T)$ die Klasse von Funktionen

$$f(t,\omega) : [0,\infty) \times \Omega \to \mathbb{R}$$

sodass

1. $(t,\omega) \to f(t,\omega)$ eine $\mathcal{B}_[S,T] - \mathcal{F}$ messbar ist.
2. Der Prozess $\{f(t,\cdot)\}_{t\in S,\mathcal{T}}$ an $\{\mathcal{F}_t\}_{t\in\mathcal{T}}$ adaptiert ist.
3. $f \in \mathbb{L}^2(\mathbb{P}\otimes\lambda)$, das heißt $\mathbb{E}[\int_S^T f(t,\omega)^2 dt] < \infty$.

Das Ito-Integral $I[f](\omega)$ wird für allgemeine Funktionen aus dieser Menge ähnlich wie Lebesgue-Integrierbarkeit schrittweise eingeführt mittels Approximation durch **einfache** Funktionen $\phi_N \in \mathcal{V}$ der Form

$$\phi_N(t,\omega) = \sum_{j=0}^{N} e_j(\omega) \cdot \mathbb{1}_{[t_j,t_{j+1})}(t) \tag{2.3}$$

haben, wobei für alle $0 \leq j \leq N$ e_j $\mathcal{F}_{t_j}$-m.b. sein muss und $\bigcup_{j=1}^{N}[t_j,t_{j+1}) = [S,T)$ eine disjunkte Zerlegung des halboffenen Intervalls ist.
Das Integral einfacher Funktionen wird definiert durch

$$\begin{aligned} I[\phi_N](\omega) &:= \int_S^T \phi_N(t,\omega) dB_t(\omega) \\ &:= \sum_{j=0}^{N} e_j(\omega)[B_{t_{j+1}}(\omega) - B_{t_j}(\omega)] \end{aligned}$$

Nun lässt sich für $f \in \mathcal{V}$ stets eine passende Folge einfacher Funktionen $\{\phi_N\}_{N\in\mathbb{N}}$ wie in (2.3) finden mit $\phi_N \overset{N\to\infty}{\to} f$ in $\mathbb{L}^2(\mathbb{P}\otimes\lambda)$, sprich

$$\mathbb{E}\left[\int_S^T (f-\phi_N)^2 dt\right] \overset{N\to\infty}{\to} 0. \quad \text{(Øksendal, 2003, S. 27-28)} \tag{2.4}$$

Den Grenzwert der assoziierten Folge der Integrale solch einer Folge einfacher Funktionen wollen wir zur Definition des Integrals von f verwenden; dabei ist es allerdings wichtig, eine Konvention für die approximierenden ϕ_N zu wählen, da unterschiedliche Folgen existieren, die (2.4) erfüllen, deren Folgen von Integralen aber unterschiedliche Grenzwerte haben können. Für das Ito-Integral wird folgende Konvention gewählt:

Definition 2.2.10: Approximierende Folge einfacher Funktionen

Sei $f \in \mathcal{V}$. Wir wählen als approximierende Folge einfacher Funktionen $\{\phi_N^f\}_{N\in\mathbb{N}}$ mit

$$\phi_N^f(t,\omega) := \sum_{j=0}^{N} f(t_j,\omega)\cdot \mathbb{1}_{[t_j,t_{j+1})}(t)$$

Das lässt für allgemeine Funktionen $f \in \mathcal{V}$ die folgende Definition zu:

Definition 2.2.11: Ito-Integral (eindimensional)

Für $f \in \mathcal{V}(S,T)$ und eine Folge einfacher Funktionen ϕ_n^f wie in Definition 2.2.10 mit

$$\phi_N^f \overset{N\to\infty}{\to} f \quad \text{in} \quad \mathbb{L}^2(\mathbb{P}\otimes\lambda)$$

ist das Ito-Integral $I[f](\omega)$ von S bis T definiert durch den Grenzwert

$$\begin{aligned} I[f](\omega) &:= \int_S^T f(t,\omega)dB_t(\omega) \\ &:= \lim_{\substack{N\to\infty \\ \text{in } \mathbb{L}^2(\mathbb{P})}} \int_S^T \phi_N^f(t,\omega)dB_t(\omega) \end{aligned}$$

Es sei angemerkt, dass ein alternativer Integralbegriff, das Stratonovich-Integral, definiert werden kann, indem man in Definition 2.2.10 setzt:

$$\phi_N(t,\omega) := \sum_{j=0}^{N} f(\frac{t_j+t_{j+1}}{2},\omega)\cdot \mathbb{1}_{[t_j,t_{j+1})}(t) \quad \text{(Øksendal, 2003, S. 24)}$$

Es folgen einige wichtige Eigenschaften zum Rechnen mit Ito-Integralen.

Satz 2.2.4: Eigenschaften des Ito-Integrals (Øksendal, 2003, Thm. 3.2.1)

Seien $f, g \in \mathcal{V}(0,T)$, c konstant und $0 \leq S < U < T$. Dann gelten

1. **(Ito-Isometrie)**

$$\mathbb{E}\left[(\int_S^T f(t,\omega)dB_t)^2\right] = \mathbb{E}\left[\int_S^T f(t,\omega)^2 dt\right]$$

2. $\int_S^T f dB_t = \int_S^U f dB_t + \int_U^T f dB_t$ fast sicher.
3. **(Linearität)** $\int_S^T (cf+g)dB_t = c\int_S^T f dB_t + \int_S^T g dB_t$ fast sicher.
4. $\mathbb{E}[\int_S^T f dB_t] = 0.$[a]
5. $\int_S^T f dB_t$ ist $\mathcal{F}_T$-messbar.

[a] Hierfür ist die Voraussetzung notwendig, dass $\mathbb{E}[\int_S^T f(t,\omega)^2 dt] < \infty$.

D-dimensional

Im Folgenden bezeichnet $\{B_t\}_{t\in[0,T]}$ eine Version D-dimensionaler Brownscher Bewegung. Unser Ziel ist jetzt, für Matrizen in $v \in \mathbb{R}^{m\times D}$ das Ito-Integral zu definieren. Dafür gehen wir komponentenweise vor:

Definition 2.2.12: Ito-Integral (*D*-dimensional) (Øksendal, 2003, 35)

Für $0 < S \leq T$ sei $\mathcal{V}^{m\times D} = \mathcal{V}^{m\times D}(S,T)$ die Menge der Matrizen $v = v(t,\omega) \in \mathbb{R}^{m\times D}$ sodass $\forall 1 \leq i \leq m$ und $\forall 1 \leq d \leq D$ die Funktion $v_{ij}(t,\omega)$ die Bedingungen von $\mathcal{V}(S,T)$ erfüllt. Dann definieren wir

$$\int_S^T v dB := [\sum_{d=1}^D \int_S^T v_{1d}(s,\omega)dB_s^d(\omega), \ldots, \sum_{d=1}^D \int_S^T v_{mj}(s,\omega)dB_s^d(\omega)]^\top \in \mathbb{R}^{m\times D}.$$

2.2.5 Ito-Kalkül

eindimensional

Im Folgenden ist $\{B_t\}_{t\in[0,T]}$ wieder eine eindimensionale Brownsche Bewegung auf . Für die folgende Klasse von Prozessen ist es möglich, ein Analogon der bekannten Kettenregel aus der Analysis zu formulieren:

Definition 2.2.13: Ito-Prozess (eindimensional) (Øksendal, 2003, 4.1.1)

Ein eindimensionaler stochastischer Prozess $\{X_t\}_{t\in[0,T]}$ auf $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathbb{P})$ der Form

$$X_t = x_0 + \int_0^t u(s,\omega)ds + \int_0^t v(s,\omega)dB_s$$
$$\underset{\text{Notation}}{\Longleftrightarrow} dX_t = udt + vdB_t; \quad X_0 = x_0$$

wird Ito-Prozess genannt, falls der **Driftterm** $u : [0,T]\times\Omega \to \mathbb{R}$ und der **Diffusionsterm** $v : [0,T]\times\Omega\to\mathbb{R}$ beide in $\mathcal{V}$ sind.

Bezüglich solch eines Prozesses definieren wir nun das Ito-Integral:

Satz 2.2.5: Ito-Integral bezüglich eines Ito-Prozess

Falls $dX_t = udt + vdB_t$ ein Ito-Prozess ist, und $f, |fu|^{\frac{1}{2}}, |fv|^{\frac{1}{2}} \in \mathcal{V}$, dann definieren wir das **Ito-Integral** bezüglich X für $t\in[0,T]$ durch

$$\int_0^t f(s,\omega)dX_s := \int_0^t f(s,\omega)u(s,\omega)ds + \int_0^t f(\omega,s)b(\omega,s)dB_s$$
$$\underset{\text{Notation}}{\Longleftrightarrow} fdX_t := fudt + fvdB_t.$$

Die Kettenregel für Ito-Prozesse nimmt im eindimensionalen Fall die folgende Form an:

Satz 2.2.6: Ito-Formel (eindimensional) (Øksendal, 2003, 4.1.2)

Sei X_t ein Ito-Prozess der Form $dX_t = udt + vdB_t$. Sei ferner $g(t,x)\in\mathcal{C}^2([0,\infty)\times\mathbb{R})$. Dann ist $Y_t = g(t,X_t)$ wieder ein Ito-Prozess und es gilt

$$dY_t = \frac{\partial g}{\partial t}(t,X_t)dt + \frac{\partial g}{\partial x}(t,X_t)dX_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t,X_t)(dX_t)^2.$$

Im Vergleich zu gewöhnlichen Differenzialgleichungen bleibt hier also ein Term zweiter Ordnung.
Für die Berechnung von $(dX_t)^2$ gelten die Rechenregeln

$$dt\cdot dt = dt\cdot dB_t = dB_t\cdot dt = 0 \tag{2.5}$$
$$dB_t\cdot dB_t = dt.$$

D-dimensional

Sei $\{B_t\}_{t\in[S,T]}$ nun eine Version m-dimensionaler Brownscher Bewegung auf dem filtrierten Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \{\mathcal{F}\}_{t\in[0,T]}, \mathbb{P})$. Die D-dimensionale Verallgemeinerung des eindimensionalen Ito-Prozess hat die folgende Form:

Definition 2.2.14: Ito-Prozess (D-dimensional) (Øksendal, 2003, 48)

Falls $\forall 1 \leq i \leq D$ und $\forall 1 \leq j \leq m$ die Prozesse $u_i(t,\omega), v_{ij}(t,\omega) \in \mathcal{V}$, ist ein **$D$-dimensionaler Ito-Prozess** ein Prozess $X : [0,T] \times \Omega \to \mathbb{R}^D$ der Form[a]

$$\begin{aligned} dX_1 &= u_1 dt + v_{11} dB_1 + \ldots + v_{1m} dB_m \\ \vdots &= \vdots \qquad \vdots \qquad\qquad \vdots \\ dX_D &= u_D dt + v_{D1} dB_1 + \ldots + v_{Dm} dB_m, \end{aligned}$$

oder in Matrixnotation: $dX(t) = u(t)dt + v(t)dB_t$.
Hier sind $X(t) \in \mathbb{R}^D, \quad u(t) \in \mathbb{R}^{D\times 1}, \quad v(t) \in \mathbb{R}^{D\times m}, \quad B(t) \in \mathbb{R}^m$.

[a] Der Übersichtlichkeit halber schreiben wir $u_i = u_i(t,\omega)$ und $v_{ij} = v_{ij}(t,\omega)$.

Nun kommen wir zur D-dimensionalen Version der Kettenregel im Ito-Kalkül:

Satz 2.2.7: Ito-Formel (D-dimensional) (Øksendal, 2003, 4.2.1)

Sei X_t ein D-dimensionaler Ito-Prozess der Form $dX_t = udt + vdB_t$. Sei ferner $g(t,x) = (g_1(t,x), \ldots, g_p(t,x)) \in C^2([0,\infty) \times \mathbb{R}^D; \mathbb{R}^p)$. Dann ist $Y_t = g(t,X_t)$ wieder ein Ito-Prozess, dessen Komponenten für $1 \leq k \leq p$ gegeben sind durch

$$dY_k = \frac{\partial g_k}{\partial t}(t,X_t)dt + \sum_{i=1}^{D} \frac{\partial g_k}{\partial x_i}(t,X_t)dX_i + \frac{1}{2}\sum_{i=1}^{D}\sum_{j=1}^{D} \frac{\partial^2 g_k}{\partial x_i \partial x_j}(t,X_t)dX_i dX_j. \tag{2.6}$$

Hier gelten für die Berechnung von $dX_i dX_j$ die Rechenregeln

$$\begin{aligned} dt \cdot dt = dt \cdot dB_t^j &= dB_t^i \cdot dt = 0 \\ dB_t^i \cdot dB_t^j &= dt \cdot \delta_{ij} \end{aligned} \tag{2.7}$$

Nun haben wir die notwendigen Werkzeuge, um **stochastische Differentialgleichungen (SDGLn)** einzuführen:

Definition 2.2.15: Stochastische Differentialgleichungen (Øksendal, 2003, Kap. 5)

Falls für alle $1 \leq i \leq D$ und $1 \leq j \leq m$ die Prozesse $u_i(t,\omega)$ und $v_{ij}(t,\omega)$ die Bedingungen in Definition 2.2.13 erfüllen, ist eine **D-dimensionale stochastische Differentialgleichung (SDGL)** eine Gleichung der Form

$$X_t = x_0 + \int_0^t u(s,X_s)ds + \int_0^t v(s,X_s)dB_s$$
$$\overset{\text{Notation}}{\iff} dX_t = u(t,X_t)dt + v(t,X_t)dB_t; \quad X_0 = x_0.$$

2.2.6 Starke Lösungen von SDGLn

Dieser Abschnitt hält sich hauptsächlich an (Khasminskii, 2011, Sec. 3.3,3.4). Wir übernehmen den Kontext aus dem letzten Abschnitt, also den filtrierten Wahrscheinlichkeitsraum und eine Version D-dimensionaler Brownscher Bewegung.

Definition 2.2.16: Starke Lösungen (Øksendal, 2003, Thm. 5.2.1 und S. 70)

Zu einer SDGLN wie in Definition 2.2.15 der Form

$$dX_t = u(t,X_t)dt + v(t,X_t)dB_t$$

mit einer *vorgegebenen* Version der Brownschen Bewegung B_t und einem *vorgegebenen* Anfangswert x heißt ein Prozess $X = \{X_t\}_{t\in[0,T]}$, der die obige Gleichung und $X_0 = x$ erfüllt, eine **starke Lösung** der SDGL.

Definition 2.2.17: Lyapunov-Operator
(Khasminskii, 2011, S. 72)

Sei $dX_t = udt + vdB_t$ ein D-**dimensionaler Ito-Prozess** wie in Definition 2.2.14. Dann definieren wir den **Generator des Ito-Prozesses** bzw. **Lyapunov-Operator** wie folgt:

$$\begin{aligned}
L : C^2([0,T],\mathbb{R}^D) &\to C^0([0,T],\mathbb{R}^D) \\
g(t,x) &\mapsto Lg(t,x) := \frac{\partial g}{\partial t}(t,x) + \sum_{i=1}^{D} u_i(t,x)\frac{\partial g}{\partial x_i}(t,x) \\
&\qquad + \frac{1}{2}\sum_{i=1}^{D}\sum_{j=1}^{D} v_{ij}(t,x)\frac{\partial^2 g}{\partial x_i \partial x_j}(t,x).
\end{aligned}$$

Der folgende Satz liefert ein hinreichendes Kriterium für die Existenz einer starken Lösung mittels Lipschitzstetigkeit und linear beschränktem Wachstum auf Zylindermengen.

Satz 2.2.8: Ein Kriterium für Existenz Starker Lösungen (Khasminskii, 2011, Theorem 3.5)

Sei $dX_t = udt + vdB_t$ ein D**-dimensionaler Ito-Prozess** (2.2.14) mit $t \in [0,T]$ und vorgegebener stetiger Version Brownscher Bewegung B_t, und davon erzeugter Filtration $\{\mathcal{F}_t\}_{t\in\mathcal{T}}$.

Falls für alle $R \in \mathbb{R}_{>0}, I = [a,b] \subset [0,T]$ eine Konstante B existiert, sodass für alle $t \in I$ und x, y mit $|x|, |y| < R$ lokale Lipschitzstetigkeit, also

$$|u(t,x) - u(t,y)| + \sum_{i=1}^{n} |v_i(t,x) - v_i(t,y)| \leq B|x-y|$$

gilt, und das Wachstum lokal linear beschränkt ist; genauer, dass

$$|u(t,x)| + \sum_{i=1}^{n} |v_i(t,x) - v_i(t,y)| \leq B(1+|x|)$$

gilt, und weiterhin ein nichtnegatives $g \in C^2([0,T],\mathbb{R}^D)$ existiert, sodass für eine Konstante $c \in \mathbb{R}_{>0}$

$$Lg \leq cg \quad \text{auf ganz } C^2([0,T],\mathbb{R}^D)$$
$$g_R := \inf_{|x|>R} g(t,x) \overset{R\to\infty}{\to} \infty, \forall t \in [0,T],$$

dann gilt:

1. Für jede Zufallsvariable Y auf $\mathbb{R}^D$, die unabhängig von $\{\mathcal{F}_t\}_{t\in[0,T]}$ ist, existiert eine **bis auf Ununterscheidbarkeit eindeutige starke Lösung** $\{X_t\}_{t\in[0,T]}$ mit $X_0 = Y$, die fast sicher stetig ist.
2. Diese Lösung ist ein Markov Prozess.
3. Falls u und v nicht von t abhängen, ist die Lösung zeitlich homogen.

2.2.7 Diskretisierungen von Differentialgleichungen

Für die numerische Validierung mittels des Ito-Kalküls erhaltener SDGLn, ist es wichtig, eine angemessene Methode für die Zeitdiskretisierung zu verwenden. Wir führen zunächst eine Methode für deterministische SDGLn ein, und werden dann eine Methode sehen, die für SDGLn mit Diffusionsterm verwendet werden kann.

Definition 2.2.18: Implizite-Euler Methode (Platen, 1999)

Sei eine D-dimensionale SDGLn mit Anfangswert $X_0 = x$ der Form

$$dX_t = u(t, X_t)dt$$

gegeben[a]. Dann ist die **Implizite Euler Methode** mit Schrittweite Δt definiert durch das Vorgehen:

1. Setze $Y_0 = x$.
2. Setze $Y_{n+1} = Y_n + u(n\Delta t, Y_n)\Delta t$.

[a] Die Gleichung ist also deterministisch; wir umgehen durch diese Notation das separate einführen Partieller Differentialgleichungen.

Die folgende Methode wird typischerweise zum Diskretisieren von SDGLn verwendet:

Definition 2.2.19: Euler-Maryuama Methode (Platen, 1999)

Sei auf $[0, T]$ eine D-dimensionale SDGLn mit Anfangswert $X_0 = x$ gegeben durch

$$dX_t = u(t, X_t)dt + v(t, X_t)dB_t.$$

Dann ist die **Euler-Maryuama Methode** mit Schrittweite Δt definiert durch das Vorgehen:

1. Setze $Y_0 = x$.
2. Setze $Y_{n+1} = Y_n + u(n\Delta t, Y_n)\Delta t + v(n\Delta t, Y_n)(B((n+1)\Delta t) - B((n)\Delta t))$.

2.3 Konvergenz von Maßen

Im Folgenden präsentieren wir einige maß- und wahrscheinlichkeitstheoretische Begriffe und Ergebnisse, die im Beweis des Mean Field Limits (Grassi et al., 2021) vorkommen. Der dort verwendete Konvergenzbegriff ist der der **schwachen Konvergenz** und wird im folgenden Abschnitt 2.3.1 diskutiert. Im Beweis wird die schwache Konvergenz einer Folge von Maßen erreicht durch den **Satz von Prokhorov**, der eine Verbindung zwischen schwacher Konvergenz und **Straffheit** von Maßen herstellt, welche im Abschnitt 2.3.2

diskutiert wird.
Im Verlauf des gesamten Abschnitts werden verschiedene Anforderungen an einen Raum E gestellt. Es ist wichtig im Hinterkopf zu behalten, dass für E in (Grassi et al., 2021, Thm. 3.3) stets der Raum

$$E = \mathcal{C}([0,T];\mathbb{R}^{2D})$$

eingesetzt und mit der kanonisch konstruierten σ-Algebra $\mathcal{B}_{[0,T]}$ aus (2.1) versehen wird. Hierbei ist $C([0,T];\mathbb{R}^{2D})$ der Raum der stetigen Funktionen von $[0,T] \subset \mathbb{R}_{\geq 0}$ nach $\mathbb{R}^D$.

2.3.1 Schwache Konvergenz

Wir sind interessiert an einer Art der Konvergenz von Maßen aus der Menge

$$\mathcal{M}_f := \mathcal{M}_f(\mathcal{B}(E)) := \{\mu | \mu \text{ ist Maß auf } (E, \mathcal{B}(E)) \text{ und } \mu(E) < \infty\}$$

der **endlichen (Borel-)Maße** auf einem metrischen Raum E (Elstrodt, 1996, 379). Entsprechend bezeichnet $\mathcal{P} \subset \mathcal{M}_f$ die Menge der **Wahrscheinlichkeitsmaße** (also Maße μ mit $\mu(E) = 1$), und für $p \in (0, \infty)$ bezeichnet

$$\mathcal{P}_p := \{\mu | \mu \in \mathcal{P} \text{ und } \int_E |x|^p d\mu < \infty\}$$

den Raum der Wahrscheinlichkeitsmaße mit endlichem p-ten Moment. Die hier verwendete Definition der **schwachen Konvergenz** charakterisiert Konvergenz von Maßen über stetige, beschränkte *Testfunktionen*:

Definition 2.3.1: Schwache Konvergenz von W-Maßen

Auf einem metrischen Raum E **konvergiert** eine Folge von endlichen Maßen $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_f$ **schwach** gegen μ, falls für jede **stetige, beschränkte** Funktion $\phi \in \mathcal{C}_b(E)$ gilt

$$\lim_{n\to\infty} \int_E \phi d\mu_n = \int_E \phi d\mu.$$

Wir notieren schwache Konvergenz schlicht mit $\mu_n \overset{n\to\infty}{\to} \mu$.
Für eine Folge von Zufallsvariablen $\{X_n\}_{n\in\mathbb{N}}$ mit Verteilungen $\{\mathbb{P}^{X_n}\}_{n\in\mathbb{N}}$ und eine Zufallsvariable X mit Verteilung $\mathbb{P}^X$ schreiben wir $X_n \overset{\mathcal{D}}{\to} X$ (X_n **konvergiert in Verteilung gegen** X), gdw. $\mathbb{P}^{X_n} \overset{n\to\infty}{\to} \mathbb{P}^X$.

Das *Portemanteau-Theorem* liefert eine Reihe äquivalenter Definitionen schwacher Konvergenz (Klenke, 2006, 13.16).

Der Begriff der schwachen Konvergenz induziert auf natürliche Weise eine Topologie auf dem Raum $\mathcal{M}_f$. Es ist hier notwendig, diese Topologie genauer zu betrachten, um die notwendigen Kriterien sowie die Aussage des in Kapitel 4 verwendeten Konvergenzbegriffes genau zu verstehen, da dieser die von der verwendeten Topologie abhängenden Konzepte des **topologischen Abschlusses** und einer **kompatiblen Metrik** involviert.

Definition 2.3.2: Schwache Topologie (Elstrodt, 1996, S. 382)

Auf $\mathcal{M}_f(\mathcal{B}(E))$ ist die **schwache Topologie** τ_w die *gröbste* Topologie, bezüglich der für alle $g \in C_b(E)$ die Abbildung

$$\begin{aligned} I_g &: \mathcal{M}_f \to \mathbb{R} \\ I_g(\mu) &= \int_E g d\mu \end{aligned}$$

stetig ist.

Die folgende Definition liefert bezüglich der schwachen Topologie einen Kompaktheitsbegriff für Familien von Maßen:

Definition 2.3.3: Schwach relativ folgenkompakt
(Klenke, 2006, 270, 287)

Eine Familie von Maßen $\mathscr{F} = \{f_i\}_{i \in \mathbb{I}} \subset \mathcal{M}_f$, wobei $(\mathcal{M}_f, \tau_w)$ mit der schwachen Topologie versehen ist, heißt **schwach relativ folgenkompakt**, falls für jede Folge $(\mu_n)_{n\in\mathbb{N}} \subset \mathscr{F}$ gilt, dass ein $\mu \in \overline{\mathscr{F}}$ im **topologischen Abschluss** der Familie von Maßen existiert mit $\mu_n \overset{n\to\infty}{\to} \mu$.

Die nächsten Definitionen werden mit dem Ziel eingeführt, einen Begriff von Äquivalenz von Maßen zu haben, der in bestimmtem Sinne die schwache Topologie respektiert. Dieses liefert der Begriff der **Kompatibilität**:

Definition 2.3.4: Kompatibilität einer Metrik mit einer Topologie (Carrillo et al., 2018, Abschnitt 3)

Sei (E,τ) ein topologischer Raum. Eine Metrik d auf E **induziert** eine Topologie τ_d auf E dadurch, dass man ihre Basis B durch ε-Bälle wie folgt konstruiert:

$$U_\varepsilon(x) := \{y | d(x,y) < \varepsilon\}$$
$$B := \{U_\varepsilon(x) | x \in E, \varepsilon > 0\}$$

d heißt **kompatibel** mit τ, falls $\tau_d = \tau$.

Der Begriff der Kompatibilität erlaubt es uns, polnische Räume einzuführen:

Definition 2.3.5: Polnischer Raum

Sei (E,τ) ein topologischer Raum. Dann heißt E

- **separabel**, wenn es eine abzählbare Teilmenge $Q \subset E$ gibt, die dicht in E ist.
- **vollständig metrisierbar**, wenn es eine Metrik d gibt, welche mit τ kompatibel und **vollständig** ist.
- **polnisch**, wenn er separabel und vollständig metrisierbar ist.

Im Verlauf der Arbeit werden wir für einen gegebenen polnischen Raum (E,d) mit dem topologischen Raum $(\mathcal{P}_p(E), \tau_w)$ arbeiten, für den wir dann wiederum gerne zeigen würden, dass er polnisch ist. Dies wird mit der Wassersteinmetrik getan; wir werden später sehen, dass diese mit der schwachen Topologie kompatibel ist. Diese Metrik benötigt zunächst den Begriff der **Kopplung** zweier Maße:

Definition 2.3.6: Kopplung (Klenke, 2006, Def. 17.54)

Seien $(\Omega_1, \mathscr{A}_1, \mu_1)$ und $(\Omega_2, \mathscr{A}_2, \mu_2)$ Wahrscheinlichkeitsräume. Wir nennen ein Wahrscheinlichkeitsmaß μ auf dem Messraum $(\Omega_1 \times \Omega_2, \mathscr{A}_1 \otimes \mathscr{A}_2)$ mit

$$\mu(\cdot \times \Omega_2) = \mu_1(\cdot) \quad \text{und} \quad \mu(\Omega_1 \times \cdot) = \mu_2(\cdot)$$

eine **Kopplung** von μ_1 und μ_2. Die Menge solcher Kopplungen bezeichnen wir mit $\Pi(\mu, \nu)$.

Mit dem Begriff der Kopplung lässt sich die Wassersteinmetrik definieren:

Definition 2.3.7: Wassersteinmetrik (Carrillo et al., 2018, Sec. 3)

Sei (E,d) ein polnischer Raum mit einer gegebenen Metrik d. Für $p \in (0,\infty)$ ist die p-**Wassersteinmetrik** W_p auf $\mathcal{P}_p(E)$ definiert durch

$$\begin{aligned} W_p(\mu,\nu) &:= \Big(\inf_{\pi\in\Pi(\mu,\nu)} \mathbb{E}[d(X,Y)^p]\Big)^{\frac{1}{p}} \\ &= \Big(\inf_{\pi\in\Pi(\mu,\nu)} \int_E\int_E d(x,y)^p \pi(dx,dy)\Big)^{\frac{1}{p}} \end{aligned}$$

Hierbei bezeichnet $\Pi(\mu,\nu)$ die Menge der Kopplungen von μ,ν wie in Definition 2.3.6.

Diese Metrik ist wohldefiniert, da beispielsweise das Produktmaß $\mu\otimes\nu \in \Pi(\mu,\nu)$ (Johannes, 2021, Satz 04.16).
Der folgende Satz lässt zu, die Eigenschaft, dass ein Raum polnisch ist, auf den darauf konstruierten Raum der Maße mit endlichem p-ten Moment übertragen:

Satz 2.3.1: Kompatibilität der Wassersteinmetrik mit der schwachen Topologie (Bolley et al., 2011a), (Villani, 2009, Thm. 6.9)

Sei (E,d) ein polnischer Raum. Dann wird für $p\in[1,\infty)$ auch der mit der schwachen Topologie ausgestattete Raum $(\mathcal{P}_p(E),\tau_w)$ polnisch mit der p-Wassersteinmetrik; d.h. $(\mathcal{P}_p(E),W_p)$ ist ein vollständiger metrischer Raum.

Dass $(\mathcal{P}_p(E),W_p)$ polnisch ist, hat auch den Vorteil, dass wir mittels des folgenden Satzes von einer in diesem Raum schwach konvergenten Folge übergehen zu einer $\mathbb{P}$-f.s. konvergenten **Representation** der Folge:

Lemma 2.3.1: Skorokhod Representation (Billingsley, 1999, 6.7)

Sei $\{\mu_n\}_{n\in\mathbb{N}}$ eine Folge von W-Maßen auf einem Messraum $(E,\mathcal{B}(E))$, die schwach gegen ein μ mit separablem Träger konvergiert. Dann existiert ein W-Raum $(\Omega,\mathscr{A},\mathbb{P})$ mit einem stochastischen Prozess $\{Y_n\}_{n\in\mathbb{N}}$ und eine Zufallsvariable Y, die beide in $(E,\mathcal{B}(E))$ Werte annehmen, sodass:

- $\mathbb{P}^Y = \mu$ sowie $\mathbb{P}^{Y_n} = \mu_n \quad$ für alle $n \in \mathbb{N}$ und
- $Y_n \overset{f.s.}{\to} Y$.

2.3.2 Straffheit

Umgangssprachlich ist der Begriff der Straffheit so zu verstehen, dass er sicher stellt, dass eine Folge von Maßen nicht "ins Unendliche auswandert" (Klenke, 2006, 287).

Definition 2.3.8: Straffheit von Maßen (Klenke, 2006, 13.26)

Eine Familie von endlichen Maßen $\mathscr{F} = \{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{M}_f(\mathcal{B}(E))$ heißt **straff auf** E (bzw. **in** $\mathcal{M}_f(\mathcal{B}(E))$), falls zu jedem $\varepsilon > 0$ eine kompakte Menge $K \subset E$ existiert mit

$$\sup_{\mu\in\mathscr{F}} \mu(E \setminus K) < \varepsilon.$$

Wir sagen, dass eine Folge von Zufallsvariablen $\{X^n\}_{n\in\mathbb{N}}$ **straff auf** E ist, falls die Folge der W-Maße $\{\mathbb{P}^{X^n}\}_{n\in\mathbb{N}}$ **straff** (**auf** E bzw. **in** $\mathcal{P}(E)$) ist (Billingsley, 1999, 57).

Für eine Folge von W-Maßen $\{\mathbb{P}_n\}_{n\in\mathbb{N}} \subset \mathcal{P}(E)$ bedeutet Straffheit also etwas anschaulicher, dass $\forall\varepsilon > 0$ ein Kompaktum $K_\varepsilon \subset E$ existiert, sodass

$$\mathbb{P}_n(K_\varepsilon) > 1-\varepsilon \quad \forall n \in \mathbb{N}. \quad \text{(Billingsley, 1999, 37)}$$

Das folgende Kriterium gibt für ein Partikelsystem, dessen Partikel anhand ihrer gemeinsamen Verteilung nicht voneinander unterscheidbar sind, an, dass der empirische Prozess straff ist, genau wenn für einen beliebigen Partikel die Folge seiner Randverteilungen straff ist:

Satz 2.3.2: Straffheit Empirischer Verteilungen unter Symmetrie (Sznitman, 1991, Proposition 2.2 ii))

Falls zu einer endlichen Menge von stochastischen Prozessen der Form

$$Z^N = (Z^{i,N} = \{Z_t^{i,N}\}_{t\in[0,T]})_{i\in[[N]]}$$

mit Werten in einem Raum $E_T := \mathcal{C}([0,T];E)$, die gemeinsame Verteilung $\mathbb{P}^{Z^N} \in \mathcal{P}(E_T^N)$ invariant unter Permutationen der Indexmenge $\sigma : [[N]] \tilde{\to} [[N]]$ und ferner E_T polnisch ist, sind die folgenden zwei Aussagen äquivalent:

1. Die Folge der empirischen Verteilungen $\{\frac{1}{N}\sum_{i=1}^N \delta_{Z^{i,N}}\}_{N\in\mathbb{N}}$ ist straff in $\mathcal{P}(E_T)$.
2. Die Folge $\{\mathbb{P}^{Z^{1,N}}\}_{N\in\mathbb{N}}$ ist straff in $\mathcal{P}(E_T)$.

Hier bezeichnet $\mathbb{P}^{Z^{1,N}} = \mathbb{P}^{Z_N} \circ \Pi_{\{1\}}^{-1}$ die Randverteilung von $Z^{1,N}$.

Es folgt ein hinreichendes Kriterium für die Straffheit einer Folge von Prozessen:

Lemma 2.3.2: Aldous' Kriterien für Straffheit (Billingsley, 1999, Thm. 16.10), (Grassi et al., 2021, Lemma 3.2)

Sei $\{X^n\}_{n\in\mathbb{N}}$ eine Folge stochastischer Prozesse auf $(\Omega,\mathcal{F},\mathbb{P})$ mit $X^n : \Omega \to C([0,T];\mathbb{R}^d)$. $\{X^n\}_{n\in\mathbb{N}}$ ist straff auf $C([0,T];\mathbb{R}^d)$ falls die folgenden zwei Bedingungen gelten:

1. $\{\mathbb{P}^{X^n(0)}\}_{n\in\mathbb{N}}$ ist straff auf $\mathbb{R}^d$.
2. $\forall \varepsilon > 0, \eta > 0$ und m existiert ein $\delta_0 \in (0,T)$ und $n_0 \in \mathbb{N}$ sodass, wenn $n \geq n_0$ und $\delta \leq \delta_0$ und β eine diskrete[a] $\sigma(X_s^n, s \in [0,T])$-Stoppzeit ist mit $\beta \leq m$, dann gilt:

$$\mathbb{P}(|X_{\beta+\delta}^n - X_\beta^n| \geq \varepsilon) \leq \eta$$

[a] D.h. mit endlichem Wertebereich.

Der **Satz von Prokhorov** verbindet die oben eingeführten Begriffe der schwachen Konvergenz und der Straffheit, und ist der Grund, warum der Konvergenzbeweis der Mean-Field Lösung den Weg über die Straffheit gehen kann.

Satz 2.3.3: Satz von Prokhorov(Klenke, 2006, 13.29)

Sei (E,d) ein metrischer Raum und $\mathscr{F} = (\mu_n)_{n\in\mathbb{N}}$ eine Folge von W-Maßen auf E. Dann gilt:
Falls $\mathscr{F}$ straff ist, ist $\mathscr{F}$ schwach relativ folgenkompakt.
(*Ist E außerdem polnisch, gilt auch die Umkehrung:*
Falls $\mathscr{F}$ schwach relativ folgenkompakt ist, ist $\mathscr{F}$ straff.)

2.4 Kolmogorovgleichungen

Der im Hauptteil der Arbeit analysierte Mean-Field Beweis von (Grassi et al., 2021) zeigt, dass im Fall des untersuchten Ito-Prozesses (3.3) eine mit diesem assoziierte **Partielle Differentialgleichung** (PDGL), die von den empirischen Verteilung der Lösungen $\{(X_t^i, V_t^i)\}_{i=1,\dots,N}$ erfüllt werden muss, überhaupt Lösungen (bestimmter Art) zulässt. Hier geben wir zunächst die allgemeine Form dieser PDGL an.
Die Anwendung der Ito-Formel (2.6) auf die Verteilung eines Ito-Prozess liefert die folgende Partielle Differentialgleichung:

Satz 2.4.1: Fokker-Planck-Gleichung (Pavliotis, 2015, Thm. 2.8)

Sei $X = \{X_t\}_{t\in[0,T]}$ ein D-dimensionaler Ito-Prozess der Form

$$X_t = x_0 + \int_0^t u(x,s)ds + \int_0^t v(x,s)dB_s$$

mit stetigen Pfaden, Anfangswert $x_0 \sim f_0$ wie in Def. 2.2.14 und über die laut Satz 2.2.1 eindeutig über ihre endlichdimensionalen Verteilungen gegebener Verteilung $\mathbb{P}^X$. Die mit X assoziierte Dichte

$$f(x,t) := \frac{d}{dx}\mathbb{P}^{X_t}(x|X_0 = x_0) \quad \text{(Pavliotis, 2015, Eq. 2.5)}$$

ist dann eine Lösung der folgenden **Partiellen Differentialgleichung** (PDGL):

$$\begin{aligned} \frac{\partial f}{\partial t}(x,t) &= \nabla_x \cdot (-[u(x,t)f(x,t)] + \frac{1}{2}\nabla_x \cdot [v^2(x,t)f(x,t)]) \\ &= -\sum_{i=1}^{D} \frac{\partial}{\partial x_i}[u_i(x,t)f(x,t)] + \frac{1}{2}\sum_{i=1}^{D}\sum_{j=1}^{D} \frac{\partial}{\partial x_j \partial x_i}[v_{ij}^2(x,t)f(x,t)] \end{aligned} \tag{2.8}$$

welche **Fokker-Planck-Gleichung** (FPG) oder auch
Kolmogorov-Vorwärtsgleichung heißt. Hier ist

$$v_{ij}^2(x,t) = \sum_{k=1}^{D} v_{ik}(x,t)v_{kj}(x,t)$$

Die folgende PDGL beschreibt das Verhalten der Erwartung eines ausreichend regulären Ito-Prozesses in einem aktuellen Zeitpunkt unter Variationen einer zurückliegenden Ausgangsbedingung in Zeit und Raum:

Satz 2.4.2: Kolmogorov-Rückwärtsgleichung (Pavliotis, 2015, Thm 2.7)

Sei eine SDGL in D Dimensionen der Form

$$dX_t = u(t,X_t)dt + v(t,X_t)dB_t \tag{2.9}$$

mit glattem Driftterm $u(t,\omega) \in \mathcal{C}^{1,2}([0,T] \times \mathbb{R}^D; \mathbb{R}^D)$, glattem Diffusionsterm $v(t,\omega)$: $\mathcal{C}^{1,2}([0,T] \times \mathbb{R}^D; \mathbb{R}^{D\times m})$ und einer stetigen Version der D-dimensionalen Brownschen Bewegung gegeben.
Seien ein fixes $t \in [0,T]$ und ein Anfangswert x_0 gegeben. Dann setzen wir für $\phi \in \mathcal{C}_b(\mathbb{R}^D)$ für alle $s \in [0,t]$:

$$f(s,x) := \mathbb{E}[\phi(X_t)|X_s = x_0]$$

Dann löst f die folgende Partielle Differentialgleichung, genannt **Kolmogorov-Rückwärtsgleichung**:

$$\begin{aligned}
-\frac{\partial f}{\partial s} &= u(s,x)\frac{\partial f}{\partial x} + \frac{1}{2}v(s,x)\frac{\partial^2 f}{\partial x^2}, \\
f(t,x) &= \phi(x_0).
\end{aligned}$$

Kapitel 3

SDE-Formulierung von PSO

Dieses Kapitel rekapituliert (Grassi et al., 2021, Abschnitt 2.2 und 2.3). Es wird im ersten Abschnitt 3.1 zunächst die Updateregel des Kanonischen PSO-Algorithmus *CanPSO* eingeführt. Danach werden in Abschnitt 3.2 zwei Formulierungen mittels SDGLn eingeführt, genannt *SD-PSO*. Abschnitt 3.2.4 enthält eine Diskussion zur Entsprechung von *SD-PSO* mit passender Zeitdiskretisierung zu *CanPSO*. Im letzten Abschnitt 3.3 wird die eindeutige Lösbarkeit von *SD-PSO* gezeigt.

3.1 Kanonische Partikelschwarmoptimierung

Zunächst geben wir für den in der Einleitung vorgestellten *CanPSO* eine Updateregel an, um diese mit den Updateregeln der im folgenden Abschnitt 3.2 vorgestellten regularisierten Formulierungen des Optimierers vergleichen zu können.

Definition 3.1.1: Kanonische Partikelschwarmoptimierung (Grassi et al., 2021, S.7,8)

Die N-Partikel **Kanonische Partikelschwarmoptimierung (*CanPSO*)** hat Hyperparameter $\lambda_j, \sigma_j > 0, j = 1,2$, wird zum Optimieren einer Kostenfunktion $F : \mathbb{R}^D \to \mathbb{R}$ eingesetzt und verwaltet je Zeitschritt $n \in \mathbb{N}_0$ die $3N+1$ Vektoren

$$X_i^n, V_i^n, Y_i^n, i = 1, \dots, N \quad \text{sowie } \overline{Y}^n.$$

Für $n = 0$ wird *CanPSO* initialisiert durch

$$\begin{aligned}
\{X_i^0, V_i^0\}_{i=1,\dots,N} &\overset{\text{u.i.v.}}{\sim} f_0 \in \mathcal{P}(\mathbb{R}^{2D}); \\
Y_i^0 &= \underset{Y \in \{X_1^0,\dots,X_N^0\}}{\operatorname{argmin}} F(Y) \\
\overline{Y}^0 &= \underset{Y \in \{Y_1^0,\dots,Y_N^0\}}{\operatorname{argmin}} F(Y).
\end{aligned}$$

In Schritt $n \in \mathbb{N}$ werden Geschwindigkeits-, Positions- und Gedächtnisupdate des i-ten Partikel, $i = 1, \dots, N$, wie folgt geschrieben:

$$X_i^{n+1} = X_i^n + V_i^n \tag{3.1}$$

$$\begin{aligned}
V_i^{n+1} &= \omega \cdot V_i^n + \frac{c_1}{2}(Y_i^n - X_i^n) + \frac{c_2}{2}(Y_i^n - X_i^n) \\
&\quad + \frac{c_1}{2}\tilde{R}_1^n(Y_i^n - X_i^n) + \frac{c_2}{2}\tilde{R}_2^n(\overline{Y}_i^n - X_i^n)
\end{aligned} \tag{3.2}$$

$$\begin{aligned}
Y_i^{n+1} &= \underset{Y \in \{Y_1^n, X_i^{n+1}\}}{\operatorname{argmin}} F(Y) \\
\overline{Y}^{n+1} &= \underset{Y \in \{\overline{Y}^n, Y_1^n,\dots,Y_N^n\}}{\operatorname{argmin}} F(Y),
\end{aligned}$$

wobei für $j = 1,2$ $\tilde{R}_t^j := 2R_t^j - \mathbb{I}$ gesetzt wurde, mit der Diagonalmatrix $R_j^n \in \mathbb{R}^{D \times D}$. R_1^n und R_2^n haben unabhängig und identisch $[0,1]$-gleichverteilte Zufallsvariablen

$$r_d^j \overset{\text{u.i.v.}}{\sim} \mathcal{U}[0,1], j = 1,2, d = 1,\dots,D$$

auf ihren Diagonalen.

Wie wir in Abschnitt 3.2.4 nachweisen werden, ist das Geschwindigkeitsupdate hierbei so umformuliert, dass es als mittels des Euler-Maryuama-Schema aus Definition 2.2.19 diskretisiert interpretiert werden kann (Grassi et al., 2021, Sec. 2.2 u. Remark 6.1), da diese Methode für das auf Brownscher Bewegung basierendende Geschwindigkeitsupdate der diskretisierten Version einer der folgenden auf SDGLn basierenden Formulierungen angewandt werden muss.

3.2 Stochastische Differentielle Partikelschwarmoptimierung

(Grassi and Pareschi, 2021) führten drei kontinuierliche Formulierungen des *PSO*-Algorithmus' ein: Eine ohne, eine mit Gedächtniseffekten, sowie eine mit Gedächtniseffekten und Hinzunahme des *pbest*. In dieser Arbeit werden nur die erste und letzte dieser mittels SDGLn formulierten Behandlungen diskutiert, und ab hier jeweils als *reduziertes SD-PSO* (Abschnitt 3.2.2) bzw. *vollständiges SD-PSO* (Abschnitt 3.2.3) bezeichnet. Die Formulierung zeitlich kontinuierlicher Versionen mit normalverteilten Zufallskomponenten von *CanPSO* erfolgt dabei mit dem Ziel, ein im Ito-Kalkül behandelbares Analogon von *CanPSO* zu haben, um Konvergenzergebnisse für die Verteilungen der erhaltenen System erarbeiten zu können. Dabei ist ein wichtiges Kriterium, dass mit entsprechender Zeitdiskretisierung der SDGL wieder *CanPSO* bzw. der jeweilig behandelte Algorithmus erhalten wird; dies wird in Abschnitt 3.2.4 diskutiert. Da in allen Behandlungen Regularitätsannahmen auf die Kostenfunktion gemacht werden, geben wir diese zunächst an.

3.2.1 Klasse der Kostenfunktionen

Die mittels SDGLn formulierten Versionen von *SD-PSO* machen folgende Annahmen auf die Kostenfunktion:

Definition 3.2.1: Kostenfunktionen

Wir definieren die Klasse der für die theoretische SD-PSO Behandlung ausreichend wohlbeschaffenen Kostenfunktionen

$$\mathfrak{F} := \{F : \mathbb{R}^D \to \mathbb{R} | \text{F erfüllt Bedingung 1., 2. und 3.}\}$$

wobei die drei Bedingungen wie folgt sind:

1. (**Lokal Lipschitz mit wachsender Konstante**) Es existiert eine Konstante $L \rangle 0$ s.d. für alle $x, y \in \mathbb{R}^D$

$$|F(x) - F(y)| \leq L \cdot (|x| + |y|)|x - y|.$$

2. F ist von unten beschränkt mit $-\infty \langle \underline{F} := \inf F$ und es existiert eine Konstante $C_u > 0$, sodass für alle Punkte $x \in \mathbb{R}^D$ gilt:

$$F(x) - \underline{F} \leq C_u(1 + |x|^2).$$

3. (**Quadratisches Wachstum**) Es existieren Konstanten $C_l, M > 0$ sodass für alle $x \in \mathbb{R}^D$ mit $|x| > M$:

$$F(x) - \underline{F} \geq C_l(1 + |x|^2)$$

In diesem Kapitel ist zunächst nur die Annahme wichtig, dass F lokal Lipschitz ist.

3.2.2 Reduziertes SD-PSO

Zunächst führen wir die Updateregeln für das *reduzierte SD-PSO*, in welchem der einfachen Analyse halber keine Gedächtniseffekte und kein *pbest* involviert sind.

Definition 3.2.2: ***Reduziertes SD-PSO***

Sei $D \in \mathbb{N}$, $F : \mathbb{R}^D \to \mathbb{R}$ eine Kostenfunktion, und $\lambda, m, \sigma > 0, \alpha >> 1$. Für gegebene $T > 0$, $N \in \mathbb{N}$, vorgegebene N Versionen D-dimensionaler Brownscher Bewegung

$$B^{i,N} := \{B^{i,N}\}_{t \in [0,T]}, \quad i = 1, \dots, N$$

und vorgegebene Anfangswerten

$$(X_0^{i,N}, V_0^{i,N}) \overset{\text{u.i.v.}}{\sim} f_0 \in \mathcal{P}_4(\mathbb{R}^{2D})$$

ist dann das N**-Partikel** ***reduzierte SD-PSO*** beschrieben durch folgendes System von $2DN$ SDGLn ($i = 1, \dots, N$):

$$\begin{aligned} m dV_t^{i,N} &= (1-m) V_t^{i,N} dt + \lambda (X^\alpha(\rho_t^N) - X_t^{i,N}) dt \qquad (3.3) \\ &\quad + \sigma \mathbb{D}(X^\alpha(\rho_t^N) - X_t^{i,N}) dB_t^{i,N} \\ dX_t^{i,N} &= V_t^{i,N} dt. \qquad (3.4) \end{aligned}$$

Hier bezeichnet

$$\rho^N(\cdot) := \frac{1}{N} \sum_{i=1}^{N} \delta_{X^{i,N}}(\cdot) = \int_{\mathbb{R}^D} f^N(\cdot, dv) \in \mathcal{P}(C([0,T]; \mathbb{R}^D))$$

die räumliche Randverteilung der **empirischen Partikelverteilung**

$$f_t^N(\cdot) := \frac{1}{N} \sum_{i=1}^{N} \delta_{(X_t^{i,N}, V_t^{i,N})}(\cdot) \in \mathcal{P}(C([0,T]; \mathbb{R}^{2D})). \qquad (3.5)$$

Die Partikel werden von einer regularisierten Version des aktuellen *gbest* angezogen:

$$\begin{aligned} X^\alpha(\rho_t^N) &:= \frac{\int_{\mathbb{R}^D} x \omega_\alpha^F(x) \rho_t^N(dx)}{\int_{\mathbb{R}^D} \omega_\alpha^F(x) \rho_t^N(dx)}, \\ \omega_\alpha^F(x) &:= \exp(-\alpha F(x)) \end{aligned}$$

Es sei angemerkt, dass Grassi et al. (2021) einen weiterer Hyperparameter $\gamma > 0$ zur Formulierung von *SD-PSO* verwenden; dieser wird in dieser Arbeit immer als $1 - m$ geschrieben, gemäß (Grassi et al., 2021, S. 10).

Die Nomenklatur von λ und σ stammt daher, dass $[X_t, V_t] \in \mathbb{R}^{2D}$ ein Ito-Prozess im Sinne von Definition 2.2.14 ist. Das heißt, λ skaliert den Driftterm und σ den Diffusionsterm des Ito-Prozesses. Diese in (Grassi and Pareschi, 2021, 3.1) vorgestellte Behandlung von *CanPSO* ohne Gedächtniskomponente oder *pbest* ist nicht die erste SDGL-basierte Behandlung von *CanPSO*. Tatsächlich basiert diese Formulierung auf dem in (Pinnau et al., 2017) vorgestellten Algorithmus **Consensus-based Optimization** *(CBO)*. In (Grassi et al., 2021, Sec. 4) wird gezeigt, dass im Limes $m \to 0$ der obige Algorithmus 3.2.2 mit *CBO* übereinstimmt.

3.2.3 Vollständiges SD-PSO

In (Grassi et al., 2021, Sec. 3.2) wird eine regularisierte Variante von *CanPSO* vorgestellt, die unter bestimmter Wahl der Hyperparameter sowie Zeitdiskretisierung den originalen Algorithmus entsprechen soll (siehe Abschnitt 3.2.4). Diese Formulierung modelliert daher die Gedächtniseffekte durch eine zusätzliche Variable Y, und involviert auch eine regularisierte Version des *pbest*. In dieser Arbeit wird diese Variante *vollständiges SD-PSO* genannt.

Definition 3.2.3: *Vollständiges SD-PSO* (Grassi et al., 2021, Sec. 3.2)

Sei $D \in \mathbb{N}$, $F : \mathbb{R}^D \to \mathbb{R}$ eine Kostenfunktion, und $\lambda_1, \lambda_2, \sigma_1, \sigma_2, \nu > 0, \alpha, \beta >> 1$. Für gegebene $T > 0$, $N \in \mathbb{N}$, vorgegebene $2N$ Versionen D-dimensionaler Brownscher Bewegung $B^{j,i} := \{B^{j,i}\}_{t\in[0,T]}$, $i = 1,\dots,N$; $j = 1,2$ und vorgegebene Anfangswerten $(X_0^{i,N}, V_0^{i,N}) \overset{\text{u.i.v.}}{\sim} f_0 \in \mathcal{P}_4(\mathbb{R}^{2D})$ sowie $Y_0^{i,N} := X_0^{i,N}$ ist das N-**Partikel *vollständige SD-PSO*** beschrieben durch folgendes System von $3DN$ SDGLn ($i = 1,\dots,N$):

$$\begin{aligned} dX_t^{i,N} &= V_t^{i,N} dt \\ dY_t^{i,N} &= \nu(X_t^{i,N} - Y_t^{i,N}) S^\beta(X^{i,N}, Y_t^{i,N}) dt \\ m dV_t^{i,N} &= (1-m) V_t^{i,N} dt + \lambda_1 (Y_t^{i,N} - X_t^{i,N}) dt \\ &\quad + \lambda_2 (Y^\alpha(\overline{\rho}_t^N) - X_t^{i,N}) dt \\ &\quad + \sigma_1 \mathbb{D}(Y_t^{i,N} - X_t^{i,N}) dB_t^{1,i} \\ &\quad + \sigma_2 \mathbb{D}(Y^\alpha(\overline{\rho}_t^N) - X_t^{i,N}) dB_t^{2,i}. \end{aligned} \tag{3.6}$$

Hier bezeichnet

$$\overline{\rho}^N(\cdot) := \frac{1}{N} \sum_{i=1}^N \delta_{Y^{i,N}}(\cdot) = \int_{\mathbb{R}^{2D}} f^N(x,\cdot,v) dv \in \mathcal{P}(C([0,T];\mathbb{R}^D))$$

die Gedächtnis-Randverteilung der **empirischen Verteilung** der Partikel

$$f_t^N(\cdot) := \frac{1}{N} \sum_{i=1}^N \delta_{(X_t^{i,N}, Y_t^{i,N}, V_t^{i,N})}(\cdot) \in \mathcal{P}(\mathcal{C}([0,T];\mathbb{R}^{3D}).$$

Die Partikel werden von einer regularisierten Version des historischen *gbest* angezogen:

$$\begin{aligned} Y^\alpha(\overline{\rho}_t^N) &:= \frac{\int_{\mathbb{R}^D} y \omega_\alpha^F(y) \rho_t^N(dx)}{\int_{\mathbb{R}^D} \omega_\alpha^F(y) \overline{\rho}_t^N(dy)}, \\ \omega_\alpha^F(y) &:= \exp(-\alpha F(y)). \end{aligned}$$

Das Gedächtnisupdate wird von einer stetigen Approximation der Vorzeichenfunktion angenähert:

$$S^\beta(x,y) := 1 + \tanh(\beta(F(y) - F(x))).$$

In der obigen Definition und im Rest der Arbeit bezeichnet für $x \in \mathbb{R}^D$ die Matrix

$\mathbb{D}(x) \in \mathbb{R}^{D\times D}$ die Diagonalmatrix mit x auf ihrer Diagonale (Entsprechend ist $\mathbb{D}(x)^2 := \mathbb{D}(x)\cdot\mathbb{D}(x)$.).

3.2.4 Entsprechung des *vollständigen SD-PSO* und *CanPSO*

In (Grassi et al., 2021, Remark 6.1) wird kommentiert, dass das *vollständige SD-PSO* bei bestimmter Wahl der Zeitdiskretisierungen für die beiden Updateregeln dem originalen *CanPSO* entspricht. Dies wird in den folgenden beiden Aussagen zusammengefasst.

Definition 3.2.4: Diskretisiertes *vollständiges SD-PSO*
(Grassi et al., 2021, Eq. 6.3)

Indem man die ersten beiden Gleichungen des Systems (3.6) aus Definition 3.2.3 mit der **Impliziten Euler-Methode** und die Geschwindigkeitsgleichung mit der **Euler-Maryuama-Methode** diskretisiert, erhält man das **diskretisierte vollständige SD-PSO**, das für $n \in \mathbb{N}_0$ und $\theta_{j,i}^n \overset{u.i.v.}{\sim} \mathcal{N}(0, \Delta t \mathbb{I}), j = 1,2$ gegeben ist durch

$$
\begin{aligned}
\tilde{X}_i^{n+1} &= \tilde{X}_i^n + \Delta t \tilde{Y}_i^{n+1}, && (3.7)\\
\tilde{V}_i^{n+1} &= \frac{m}{m+(1-m)\Delta t}\tilde{V}_i^n + \frac{\lambda_1 \Delta t}{m+(1-m)\Delta t}(\tilde{Y}_i^n - \tilde{X}_i^n) && (3.8)\\
&+ \frac{\lambda_2 \Delta t}{m+(1-m)\Delta t}(\tilde{\bar{Y}}_\alpha^n - X_i^n)\\
&+ \frac{\sigma_1\sqrt{\Delta t}}{m+(1-m)\Delta t}\mathbb{D}(\tilde{Y}_i^n - \tilde{X}_i^n)\theta_{1,i}^n\\
&+ \frac{\sigma_2\sqrt{\Delta t}}{m+(1-m)\Delta t}\mathbb{D}(\tilde{\bar{Y}}_\alpha^n - \tilde{X}_i^n)\theta_{2,i}^n,\\
\tilde{Y}_i^{n+1} &= \tilde{Y}_i^n + \nu\Delta t(\tilde{X}_i^{n+1} - \tilde{Y}_i^n)S^\beta(\tilde{X}_i^{n+1}, \tilde{X}_i^n)\\
\tilde{\bar{Y}}_\alpha^{n+1} &= (\sum_{i=1}^N \tilde{Y}_i^n \omega_\alpha^F(\tilde{Y}_i^n)) \;/\; (\sum_{j=1}^N \omega_\alpha^F(\tilde{Y}_j^n))
\end{aligned}
$$

Hierbei entspricht die Diskretisierung des Geschwindigkeitsupdates der Euler-Maryuama Methode aus Definition 2.2.19, da wegen Satz 2.2.3 gilt:

$$
B_{(n+1)\Delta t}^{j,i} - B_{n\Delta t}^{j,i} \overset{\text{stationär}}{=} B_{\Delta t}^{j,i} \overset{\mathcal{D}}{=} \theta_{j,i}^n
$$

Für diese Diskretisierung gilt es jetzt, die richtigen Hyperparameter für eine Entsprechung zu *CanPSO* zu identifizieren. Sie werden im Beweis des folgenden Satzes nachgewiesen:

Satz 3.2.1: Entsprechung des Diskretisierten *vollständigen SD-PSO* und *CanPSO* (Grassi et al., 2021, Remark 6.1)

Indem man im *diskretisierten vollständigen SD-PSO* in Definition 3.2.4 setzt:

$$\begin{aligned}
&\lambda_j = \frac{c_k}{2}, \quad \sigma_j = \frac{c_k}{2\sqrt{3}}, \quad j = 1,2; \\
&\Delta t = 1; \quad \nu = \frac{1}{2}; \quad m = \omega \\
&\alpha,\ \beta \to \infty,
\end{aligned}$$

entspricht dieses dem *CanPSO* aus Definition 3.1.1 in folgendem Sinn:
Falls die Kostenfunktion F und die Anfangsverteilungen $f_0 \in \mathcal{P}(\mathbb{R}^{2D})$ gleich sind, und wir jeweils den **Zustandsprozess** von *CanPSO* bzw. vom *diskretisierten vollständigen SD-PSO* definieren durch

$$\begin{aligned}
\{Z^n\}_{n\in\mathbb{N}} &= \{(X_i^n, V_i^n, Y_i^n, i = 1,\ldots,N, \overline{Y}^n)\}_{n\in\mathbb{N}}, \\
\{\tilde{Z}^n\}_{n\in\mathbb{N}} &= \{(\tilde{X}_i^n, \tilde{V}_i^n, \tilde{Y}_i^n, i = 1,\ldots,N, \tilde{\overline{Y}}^n_\alpha)\}_{n\in\mathbb{N}},
\end{aligned}$$

jeweils mit Werten in $\bigtimes_{k=1}^{3N+1} \mathbb{R}^D$, gilt für alle $n \in \mathbb{N}$, gegebenes Z^n für alle $A^{n+1} \in \{X_i^{n+1}, V_i^{n+1}, Y_i^{n+1}\}_{i=1,\ldots,N} \cup \{\overline{Y}^{n+1}\}$, dass:

$$\mathbb{E}\left[\tilde{A}^{n+1} | \tilde{Z}^n = Z^n\right] = \mathbb{E}\left[A^{n+1} | Z^n\right]$$

Beweis: Wir prüfen also die vier Updateregeln (3.7) mit den postulierten Hyperparametern.
Das Positionsupdate (3.7) ist deterministisch und wegen $\Delta t = 1$ trivialerweise in Erwartung gleich (3.1).
Für das Geschwindigkeitsupdate (3.8) gilt wegen $\mathbb{E}[\theta_{j,i}] = 0$:

$$\mathbb{E}\left[\tilde{V}^{n+1} | \tilde{Z}^n = Z^n\right] = mV_i^n + \frac{c_1}{2}(Y_i^n - X_i^n) + \frac{c_2}{2}(\overline{Y}^n - X_i^n)$$

Da in (3.2) für $j = 1,2, d = 1,\ldots,D$ $\mathbb{E}\left[\tilde{R}^j_{t\,d}\right] = 2\mathbb{E}\left[r^j_{t\,d}\right] - 1 = 0$ gilt, haben wir:

$$\mathbb{E}\left[\tilde{V}^{n+1} | \tilde{Z}^n = Z^n\right] = \mathbb{E}\left[V^{n+1} | Z^n\right].$$

Für das regularisierte *gbest* $\tilde{\bar{Y}}_\alpha^{\,n+1}$ ist die Gewichtungsfunktion $\omega_\alpha^F(x) := \exp(-\alpha F(x)) \in C_b(\mathbb{R}^D, \mathbb{R})$ (siehe 1. in Definition 3.2.1) so gewählt, dass (Pinnau et al., 2017):

$$\lim_{\alpha\to\infty} \mathbb{E}\left[\tilde{\bar{Y}}_\alpha^n | \tilde{Z}^n = Z^n\right] = \operatorname*{argmin}_{X_i^n, i=1,\dots,N} F(X_i^n)$$

Zuletzt gilt für das Update der persönlichen Speicherkomponenten

$$\begin{aligned}
\lim_{\beta\to\infty} \mathbb{E}\left[\tilde{Y}_i^{n+1} | \tilde{Z}^n = Z^n\right] &= \lim_{\beta\to\infty} Y_i^n + \frac{1}{2}(X_i^{n+1} - Y_i^n) S^\beta(X_i^{n+1}, X_i^n) \\
&= \lim_{\beta\to\infty} Y_i^n + \frac{1}{2}(X_i^{n+1} - Y_i^n)(1 + \tanh(\beta(F(X_i^n) - F(X_i^{n+1})))) \\
&= Y_i^n + (X_i^{n+1} - Y_i^n)\mathbb{1}_{F(X_i^n) < F(X_i^{n+1})} \\
&= \operatorname*{argmin}_{Y \in \{Y_1^n, X_i^{n+1}\}} F(Y)
\end{aligned}$$

Das schließt den Beweis ab. ■

3.3 Regularität des reduzierten SD-PSO

In diesem Abschnitt wollen wir zeigen, dass die *reduzierte* Formulierung starke und eindeutige Lösungen zulässt. Dafür verwenden wir das hinreichende Kriterium 2.2.8 und die Annahmen 3.2.1 auf unsere Kostenfunktion F.

Satz 3.3.1: Starke Lösung des *reduzierten SD-PSO*

Das *reduzierte SD-PSO* lässt für $F \in \mathfrak{F}$ (3.2.1), eine vorgegebene Version der Brownschen Bewegung und eine vorgegebene Anfangsbedingung X_0 eine bis auf Ununterscheidbarkeit eindeutige, fast sicher stetige starke Lösung X_t, die ein zeitlich homogener Markovprozess ist.

Beweis: Wir prüfen zunächst die lokale Stetigkeit. Dafür betrachten wir den zusammengefügten Prozess

$$
\begin{aligned}
Z_t^{(N)} &= [X_t^{(N)}; V_t^{(N)}]^T \quad \in \mathbb{R}^{2ND}, \text{ wobei} \\
X_t^{(N)} &= [X^{1,N}, \dots, X^{N,N}]^T \quad \in \mathbb{R}^{ND} \text{ und} \\
V_t^{(N)} &= [V^{1,N}, \dots, V^{N,N}]^T \quad \in \mathbb{R}^{ND}
\end{aligned}
$$

und formulieren anlog zu (Carrillo et al., 2018, Sec. 2) das System (3.3) mittels des $2ND$-dimensionalen Ito-Prozesses $Z_t^{(N)}$, definiert durch

$$
\begin{aligned}
dV_t^{(N)} &= \frac{1-m}{m} V_t^{(N)} dt + \frac{\lambda}{m} F_N(X_t^{(N)}) dt + \frac{\sigma}{m} M_N(X_t^{(N)}) dB_t^{(N)} \\
dX_t^{(N)} &= V_t^{(N)} dt.
\end{aligned}
$$

Hier verwenden wir die Formulierung des Driftterms und des Weiteren

$$
\begin{aligned}
F^{i,N}(X) &:= \frac{\sum_{j \neq i} (X^i - X^j) \omega_F^\alpha (X^j)}{\sum_{k=1}^N \omega_F^\alpha (X^k)} = X^i - X^\alpha(\rho_t) \in \mathbb{R}^D \\
F_N(X) &= [F^{1,N}(X), \dots, F^{N,N}(X)]^T \in \mathbb{R}^{ND} \\
M_N(X) &= diag(|F_N^1(X)| \mathbb{I}_d, \dots, |F_N^N(X)| \mathbb{I}_d) \in \mathbb{R}^{ND \times ND}
\end{aligned}
$$

Hierbei bezeichnet für $i = 1, \dots, N$ und $X \in \mathbb{R}^{ND}$ der Vektor X^i den D-dimensionalen Teilvektor von X, der in den Komponenten $(i-1)D+1$ bis iD eingetragen ist. Die Drift- und Diffusionsterme u und v im Sinne von 2.2.8 sind also:

$$
\begin{aligned}
u_k(t, Z_t) &= \tilde{V}_t^k + \tilde{F}_N(X_t^{(N)})_k \\
\text{wobei } \tilde{V}_t^k &:= (\mathbb{1}_{k \leq ND} \cdot \frac{1-m}{m} + \mathbb{1}_{k > ND}) \cdot (V_t)_k \\
v_{kl}(t, Z_t) &= \frac{\sigma}{m} M_N(X_t^{(N)})
\end{aligned}
$$

Wir bemerken, dass die obigen Ausdrücke nicht von $t \in I$ abhängen, womit zeitliche Homogeneität folgt.
Also ist die erste Stetigkeitsbedingung, dass auf allen Zylindermengen $I \times U_R$ mit $I := [t_0, t_1] \subset [0, T]$, und $R > 0$ fest und $U_R := \{(x, v) | x, v \in \mathbb{R}^{ND}, |x| + |v| < R\}$, eine lokale Konstante $B > 0$ existiert, s.d. gilt:

$$
\begin{aligned}
&\frac{1}{m}|v-w|+\frac{\lambda}{m}|F_N(x)-F_N(y)|+\frac{\sigma}{m}\sum_{k=1}^{ND}|(M_N(x)-M_N(y))_k| \leq B(|x-y|+|v-w|)\\
\iff &\frac{1}{m}|v-w|+\frac{\lambda}{m}L_{F_N}|x-y|+\frac{D\sigma}{m}||F_N(x)|-|F_N(y)|| \leq B(|x-y|+|v-w|)\\
&\Leftarrow \frac{1}{m}|v-w|+\frac{\lambda+D\sigma}{m}L_{F_N}|x-y| \leq B(|x-y|+|v-w|)\\
&\Leftarrow (\frac{1}{m}+\frac{\lambda+D\sigma}{m}L_{F_N})(|x-y|+|v-w|) \leq B(|x-y|+|v-w|)\\
&\Leftarrow \frac{1}{m}+\frac{\lambda+D\sigma}{m}L_{F_N} \leq B
\end{aligned}
$$

Dabei haben wir verwendet, dass F_N *lokal Lipschitzstetig* (LL) mit Konstante L_{F_N} ist, weil die $F^{i,N}$ LL sind:

$$
\begin{aligned}
|F^{i,N}(X)-F^{i,N}(Y)| &= |\frac{\sum_{j\neq i}(X^i-X^j)\omega_F^\alpha(X^j)}{\sum_{k=1}^N \omega_F^\alpha(X^k)}-\frac{\sum_{j\neq i}(Y^i-Y^j)\omega_F^\alpha(Y^j)}{\sum_{k=1}^N \omega_F^\alpha(Y^k)}|\\
&\leq L_{\omega_F^\alpha}|X-Y|,
\end{aligned}
$$

wobei wir für die rechten Terme die LL-konstante $L_{\omega_F^\alpha}$ erhalten durch LL der Funktionen

1. F (gemäß Definition 3.2.1),
2. $\omega_F^\alpha(x)=\exp(-\alpha F(x))$
3. $h([x^1,\dots,x^n]) := \sum_{j=1}^N \frac{x^j g(x^j)}{\sum_{k=1}^N g(x^k)}$, für $g>0$ und lokal lipschitz,

da die Komposition von LL Funktionen wieder LL ist.
Die obigen drei Funktionen sind außerdem lokal Linear beschränkt, daher existiert eine Konstante $H_{\omega_F^\alpha}$, s.d. $F_N(X)\leq H_{\omega_F^\alpha}(1+|X|)$. Dadurch ist die zweite lokale Stetigkeitsbedingung erfüllt:

$$
\begin{aligned}
\frac{1}{m}|v|+\frac{\lambda}{m}|F_N(x)|+\frac{\sigma}{m}\sum_{k=1}^{ND}|(M_N(x)-M_N(y))_k| &\leq \frac{1}{m}|v|+\frac{\lambda}{m}H_{\omega_F^\alpha}(1+|x|)+\frac{D\sigma}{m}L_{F_N}|x|\\
&\leq (\frac{1}{m}+\frac{\lambda}{m}H_{\omega_F^\alpha}+\frac{D\sigma}{m}L_{F_N})(1+|x|+|v|)
\end{aligned}
$$

Also gelten die Stetigkeitsvoraussetzungen mit

$$\begin{aligned}
B &:= max(\frac{1}{m} + \frac{\lambda}{m}H_{\omega_F^\alpha} + \frac{D\sigma}{m}L_{F_N}, \frac{1}{m} + \frac{\lambda + D\sigma}{m}L_{F_N}) \\
&= \frac{1}{m} + \frac{\lambda}{m}H_{\omega_F^\alpha} + \frac{D\sigma}{m}L_{F_N}
\end{aligned}$$

Nun ist der zweite Teil der Voraussetzungen zu prüfen. Dazu setzen wir

$$g(t,z) = g(z) := |z|^2 + C.$$

Hierbei ist $C \geq 0$ ein konstanter Term, den wir im Verlauf des Beweises ermitteln. Er fällt offensichtlich durch die Anwendung des Lyapunov-Operators weg. Wir haben $g \geq 0, g \in C^2([0,T], \mathbb{R}^{2ND})$ und $g_R = R^2 + C \overset{R\to\infty}{\to} \infty$. Außerdem gilt

$$\begin{aligned}
Lg(z_t) &= \sum_{k=1}^{2ND} u_k(t,z_t)\frac{\partial |z_t|^2}{\partial z_k} + \frac{1}{2}\sum_{k,l=1}^{2ND} v_{kl}(t,z_t)\frac{\partial^2 |z_t|^2}{\partial z_k \partial z_l} \\
&= \langle \frac{1-m}{m}v_t + \frac{\lambda}{m}F_N(x_t), 2v_t\rangle + \langle v_t, 2v_t\rangle + \frac{\sigma^2 D}{m^2}|F_N(x_t)|^2 \\
&\leq 2\frac{1-m}{m}\langle v_t, v_t\rangle + 2\frac{\lambda}{m}\langle x_t - x^\alpha(\rho_t), v_t\rangle \\
&\qquad + 2\langle v_t, v_t\rangle + \frac{\sigma^2 D}{m^2}|x_t - x^\alpha(\rho_t)|^2 \\
&\leq \frac{2}{m}|v_t|^2 + \frac{2\lambda}{m}(\langle x_t - x^\alpha(\rho_t), v_t\rangle) + \frac{\sigma^2 D}{m^2}(|x_t|^2 + |x^\alpha(\rho_t)|^2) \\
&\overset{4.3.1}{\leq} \frac{2}{m}|v_t|^2 + \frac{2\lambda}{m}(\langle x_t - x^\alpha(\rho_t), v_t\rangle) \\
&\qquad + \frac{\sigma^2 DN}{m^2}(c_1 + (c_2+1)|x_t|^2)
\end{aligned}$$

Die Terme $|v_t|^2$ und $|x_t|^2$ können wir durch $|z_t|^2$ abschätzen.

Für den Term $\langle x_t - x^\alpha(\rho_t), v_t\rangle$ gilt folgende Abschätzung:

$$\begin{aligned}
\langle x_t - x^\alpha(\rho_t), v_t\rangle &= \langle x_t, v_t\rangle - \langle x^\alpha(\rho_t), v_t\rangle \\
&\leq |\langle x_t, v_t\rangle - \langle x^\alpha(\rho_t), v_t\rangle| \\
&\leq |\langle x_t, v_t\rangle| + |\langle x^\alpha(\rho_t), v_t\rangle| \\
&\overset{(4.12)}{\leq} 2(|x_t|^2 + 2|v_t|^2 + |x^\alpha(\rho_t)|^2) \\
&\overset{4.3.1}{\leq} 2(c_1 + (c_2+1)|x_t|^2 + 2|v_t|^2) \\
&\leq 2c_1 + 2(c_2+2)|z_t|^2
\end{aligned}$$

Also haben wir:

$$\begin{aligned}
Lg(z_t) &\leq (\frac{\sigma^2 DN}{m^2} + \frac{4\lambda}{m})c_1 + (\frac{2}{m} + \frac{2\lambda}{m} + \frac{(c_2+1)\sigma^2 DN}{m^2} + 2(c_2+2)\frac{\lambda}{m})|z_t|^2 \\
&=: c(|z_t|^2 + C) = cV(z_t)
\end{aligned}$$

Da alle Konstanten positiv sind, ist $|z_t|^2 + C \geq 0$.

Das schließt den Beweis der Aussage ab. $\blacksquare$

Kapitel 4

Mean-Field Limit

Dieses Kapitel führt den in (Grassi et al., 2021, Abschnitt 3.1) für Theorem 3.1 gegebenen Beweis des zuerst in (Grassi and Pareschi, 2021, Abschnitt 3.1) postulierten Mean-Field Limit des in Abschnitt 3.2.2 eingeführten *SD-PSO ohne Gedächtnis*.
Dieses Kapitel ist wie folgt aufgebaut:

1. Zunächst werden in Abschnitt 4.1 die Eigenschaften der Räume, mit denen gearbeitet wird, eingeführt.

2. Im folgenden Abschnitt 4.2 wird erarbeitet, was überhaupt zu prüfen ist, damit wir sagen, dass das Mean-Field Limit erfüllt ist. Der Mittelpunkt des Abschnitts ist Satz 4.2.2, dessen Beweis das Ziel des gesamten Kapitels ist.

3. Es folgt eine gleichmäßige Abschätzung der Momente 4.3.2 in ihrem eigenen Abschnitt 4.3, da sie von mehreren Teilen des Beweises verwendet wird.

4. Der nächste Abschnitt 4.4 befasst sich mit der Straffheit der Folge der empirischen Maße.

5. In Abschnitt 4.5 wird ausgehend von der Straffheit mithilfe eines Kompaktheitsarguments die Konvergenz der empirischen Maße gezeigt. Dieses gibt der Beweis von Satz 4.5.1, welcher die zentralen Argumente von (Grassi et al., 2021, Sec. 3) beinhaltet. Satz 4.5.1 ist somit ein guter Anhaltspunkt für einen Überblick über die Strategie, die verfolgt wird.

6. Abschnitt 4.6 gibt den Beweis von Satz 4.2.2.

4.1 Präliminarien

Zunächst prüfen wir die Eigenschaften des *Pfadraums* $C([0,T];\mathbb{R}^{2D})$, in dem die Pfade der Lösung X_t der SDGL (3.3) liegen, zu prüfen. Die Beschaffenheit dieses Raums ist wichtig, da wir die Konvergenz in Verteilung von Zufallsmaßen f^N innerhalb des darauf konstruierten Raumes der Maße $\mathcal{P}(C([0,T];\mathbb{R}^{2D}))$ betrachten.

Lemma 4.1.1: Polnischer Pfadraum

Der Pfadraum $C([0,T];\mathbb{R}^D)$ ist polnisch.

Beweis: Sei $E := C([0,T];\mathbb{R}^D)$. Dann ist E separabel (Billingsley, 1999, 57). E ist vollständig metrisierbar, zum Beispiel mittels der Skorokhod Metrik (Billingsley, 1999, S. 124). Also ist E polnisch. ■

Der Beweis der Konvergenz in Verteilung im Raum der Maße $\mathcal{P}(C([0,T];\mathbb{R}^{2D}))$ wird darüber laufen, dass wir zeigen, dass Teilfolgen der Folge der Verteilungen $\mathbb{P}^{f^N}$ in $\mathcal{P}(\mathcal{P}(C([0,T];\mathbb{R}^{2D})))$ einen eindeutigen schwachen Grenzwert haben[1]. Für die Eindeutigkeit dieses Grenzwerts ist es wichtig, eine Metrik auf $\mathcal{P}(\mathcal{P}(C([0,T];\mathbb{R}^{2D})))$ zu verwenden, die mit der Topologie der Schwachen Konvergenz aus Definition 2.3.2 kompatibel gemäß Definition 2.3.4 ist. So eine Metrik liefert mithilfe von Satz 2.3.1 das folgende Lemma:

Lemma 4.1.2: Polnischer Raum der Maße (Carrillo et al., 2018, Sec. 3)

Für $p,q \in [1,\infty)$ ist der (topologische) Raum der Wahrscheinlichkeitsmaße mit endlichem p-ten Moment $(\mathcal{P}_p(\mathcal{P}_q(C([0,T];\mathbb{R}^{2D}))),\tau_w)$, ausgestattet mit der schwachen Topologie, polnisch, da er durch W_p, die p-Wassersteinmetrik vollständig metrisiert wird, und W_p mit τ_w kompatibel ist.

Beweis: Gemäß Lemma 4.1.1 ist $C([0,T];\mathbb{R}^{2D})$ polnisch. Nun liefert zweifache Anwendung von Satz 2.3.1 die Aussage.

4.2 Ziel des Mean-Field Beweises

In diesem Abschnitt wird eine genaue Definition davon erarbeitet, was im Fall von *SD-PSO* heißt, dass das Mean-Field Limit gilt.

[1] Nach Definition 2.3.1 ist schwache Konvergenz der $\mathbb{P}^{f^N}$ gleichbedeutend mit Konvergenz in Verteilung von f^N.

Es ist zu prüfen, ob die Folge der empirischen Maße

$$\{f^N\}_{N\in\mathbb{N}} := \{\frac{1}{N}\sum_{i=1}^{N}\delta_{Z^{i,N}}\}_{N\in\mathbb{N}}$$

der laut Satz 3.3.1 existierenden zeitlich homogenen, eindeutigen starken Lösung $\{(Z_t^N)\}_{t\in[0,T]}$ des N-Partikel *SD-PSO*-Systems aus Definition 3.2.2, in gewissem Sinne konvergiert, und ob der Grenzwert eine schwache Lösung der nach Satz 2.4.1 geltenden Fokker-Planck Gleichung ist.

Zunächst betrachten wir, welche Form die FPG für *SD-PSO* überhaupt annimmt. Durch Anwendung von Satz 2.4.1 auf das SD-PSO System aus Satz 3.2.2 erhalten wir:

Satz 4.2.1: Fokker-Planck Gleichung für SD-PSO
(Grassi et al., 2021, Eq. (3.4))

Die **Fokker-Planck-Gleichung** nimmt im Fall des *SD-PSO ohne Gedächtnis* aus Satz die folgende Form an:

$$\begin{aligned}\frac{\partial f}{\partial t} = -v\cdot\nabla_x f + \nabla_v\cdot(\frac{1-m}{m}vf + \frac{\lambda}{m}(x - X^\alpha(\rho)) + \\ \frac{\sigma^2}{2m^2}\mathbb{D}(x - X^\alpha(\rho))^2\nabla_v f)).\end{aligned} \tag{4.1}$$

Hierbei ist $f \in \mathcal{P}_2(C([0,T],\mathbb{R}^D)^2)$, mit räumlichem Randmaß $\rho(\cdot) = \int_{\mathbb{R}^D} f(\cdot,dv) \in \mathcal{P}_2(C([0,T],\mathbb{R}^D))$.

Beweis: Wir identifizieren zunächst die Drift- und Diffusionsterme des Systems (3.3) als:

$$\begin{aligned} dZ_t &= u(Z_t,t)dt + v(Z_t,t)dB_t; \\ \text{wobei} \quad Z_t &:= [X_t,V_t] \in \mathbb{R}^{2D}, \\ u(z,t) &:= u(x,v,t) = [v, \frac{1-m}{m}v + \frac{\lambda}{m}(X^\alpha(\rho)-x)] \in \mathbb{R}^{2D}, \\ \text{und} \quad v(z,t) &:= v(x,v,t) = [0, \frac{\sigma}{m}(X^\alpha(\rho)-x)] \in \mathbb{R}^{2D} \end{aligned}$$

Somit erhalten wir

$$\nabla_z \cdot (u(z,t)f) = \nabla_x \cdot (vf) + \nabla_v \cdot (\frac{1-m}{m}vf + \frac{\lambda}{m}(X^\alpha(\rho) - x)$$
$$\nabla_z \cdot (v^2(z,t)f) = \nabla_v \cdot (\frac{\sigma^2}{2m^2}(X^\alpha(\rho) - x)^2 f)$$

Damit erhalten wir folgende FPG:

$$\begin{aligned}\frac{\partial f}{\partial t} &= \nabla_z \cdot (-[u(z,t)f(z,t)] + \frac{1}{2}\nabla_z \cdot [v^2(z,t)f(z,t)]) \\ &= \nabla_x(-vf) + \nabla_v \cdot (\frac{1-m}{m}vf + \frac{\lambda}{m}(x - X^\alpha(\rho)) + \\ &\qquad \nabla_v \cdot (\frac{\sigma^2}{2m^2}(x - X^\alpha(\rho))^2 f)) \\ &= -v \cdot \nabla_x f + \nabla_v \cdot (\frac{1-m}{m}vf + \frac{\lambda}{m}(x - X^\alpha(\rho)) + \\ &\qquad \frac{\sigma^2}{2m^2}\mathbb{D}(x - X^\alpha(\rho))^2 \nabla_v f))\end{aligned}$$

Dies schließt den Beweis ab. ■

(Grassi et al., 2021) konstruieren zunächst folgendes Funktional, anhand dessen sie schwache Lösungen von (4.1) charakterisieren:

Definition 4.2.1: Funktional

Für eine Testfunktion mit kompaktem Träger $\varphi \in C_c^2(\mathbb{R}^{2D})$ definieren wir ein Funktional wie folgt

$$\mathbb{F}_\varphi : \mathcal{P}_2(C([0,T];\mathbb{R}^{2D})) \to \mathbb{R} \tag{4.2}$$

$$\begin{aligned}
\mathbb{F}_\varphi(f) &:= \int \varphi d f_T - \int \varphi d f_0 + \int_0^T L\varphi ds \\
&= \int \varphi d f_T - \int \varphi f_0 \qquad (4.3) \\
&\quad + \int_0^T \int v_s \cdot \nabla_x \varphi d f_s ds \\
&\quad - \frac{1-m}{m} \int_0^T \int v_s \cdot \nabla_v \varphi d f_s ds \\
&\quad + \frac{\lambda}{m} \int_0^T \int (x_s - X^\alpha(\rho^{f_s^N})) \cdot \nabla_v \varphi d f_s ds \qquad (4.4) \\
&\quad - \frac{\sigma^2}{2m^2} \int_0^T \sum_{d=1}^D \int (x_s - X^\alpha(\rho^{f_s^N}))_d^2 \frac{\partial^2}{\partial {v_d}^2} \varphi d f_s ds \qquad (4.5)
\end{aligned}$$

Hier ist L der Lyapunov Operator aus Def. 2.2.17.

Die Folgende Definition gibt mithilfe dieses Funktionals hinreichende Bedingungen an, dass ein Maß eine Schwache Lösung der FPG (4.1) im Sinne der Theorie von PDGLn ist:

Definition 4.2.2: Schwache Lösung der Fokker-Planck-Gleichung (Huang and Qiu, 2021, Def. 3.1)

$f \in \mathcal{P}_2(C([0,T];\mathbb{R}^{2D}))$ ist eine **schwache Lösung** der PDGL (4.1), falls

1. Stetigkeit der zeitlichen Randmaße gilt: Für alle $t_n \to t$ gilt $f_{t_n} \to f_t$, d.h. für alle $\phi \in \mathcal{C}_b(\mathbb{R}^{2D})$ gilt

$$\int_{\mathbb{R}^{2D}} \phi(x,v) f_{t_n}(dx,dv) \to \int_{\mathbb{R}^{2D}} \phi(x,v) f_t(dx,dv). \tag{4.6}$$

2. Für alle $\phi \in C_c^2(\mathbb{R}^{2D})$ gilt:

$$\mathbb{F}_\varphi(f) = 0, \quad \mathbb{P}\text{-f.s.}$$

Da das Prüfen von Definition 4.2.2 unter bestimmten Voraussetzungen gemacht wird, die in einigen Sätzen und Lemmata reiteriert werden, gruppieren wir hier die Voraussetzungen in eine Definition:

Definition 4.2.3: Voraussetzungen für das MFL

Wir sagen, dass *die Annahmen $\mathbb{A}$ für das MFL gelten*, falls gilt:

1. Die Kostenfunktion F erfüllt die Annahmen 3.2.1, also $F \in \mathfrak{F}$.
2. Die Anfangswerte sind unabhängig und identisch verteilt, also $Z_0^{i,N} \overset{\text{u.i.v.}}{\sim} f_0, i = 1, \dots N$, mit $f_0 \in \mathcal{P}_4(\mathbb{R}^{2D})$.

Mit diesen Voraussetzungen sind immer für jedes $N \in \mathbb{N}$ die folgenden zwei Objekte assoziiert:

1. Die nach Satz 3.3.1 existierende und bis auf Ununterscheidbarkeit eindeutige Lösung des *SD-PSO*-Systems ohne Gedächtnis (3.3):

$$\{\{Z_t^{i,N} := [X_t^{i,N}, V_t^{i,N}]\}_{i \in [[N]]}\}_{t \in [0,T]} : [0,T] \to (\mathbb{R}^{2D})^N.$$

2. Das **empirische Maß**:

$$f^N := \frac{1}{N} \sum_{i=1}^{N} \delta_{Z^{i,N}} \in \mathcal{P}(C([0,T]; \mathbb{R}^{2D})) \tag{4.7}$$

Der folgende Satz fasst das Mean-Field Limit Ergebnis des gesamten Kapitels zusammen:

Satz 4.2.2: Mean-Field Limit des gedächtnislosen SD-PSO

Unter den Voraussetzungen $\mathbb{A}$ (4.2.3) gilt:
Mit $N \to \infty$ konvergieren die empirischen Maße f^N, die jeweils (3.5) für ein gegebenes N lösen, gegen eine schwache Lösung f der **FPG** im Sinne von 4.2.2.
Ferner hat mit diesem vorgegebenen f das folgende nichtlineare System eine eindeutige Lösung $\overline{Z}_t$:

$$\begin{aligned} md\overline{V}_t &= (1-m)\overline{V}_t dt + \lambda(X^\alpha(\rho_t) - \overline{X}_t)dt \\ &\quad + \sigma\mathbb{D}(X^\alpha(\rho_t) - \overline{X}_t)dB_t \\ d\overline{X}_t &= \overline{V}_t dt, \\ \text{wobei} \quad \rho_t &:= \int_{\mathbb{R}^D} f_t(x,y)dy = \mathbb{P}^{\overline{X}_t}, \\ f_t &= \mathbb{P}^{\overline{Z}_t} \\ \text{und} \quad \overline{Z}_t &:= [\overline{X}_t, \overline{V}_t] \end{aligned} \tag{4.8}$$

Beweis: Siehe Abschnitt 4.6.
Die Lösung $\{\overline{Z}_t\}_{t\in[0,T]}$ ist ein sogenannter **McKean**-Prozess.

4.3 Momentenabschätzung

Ziel dieses Abschnitts ist eine in der Zeit gleichmäßige Abschätzung gerader Momente der Positionen, der Geschwindigkeiten und des regularisierten *gbest*, die im Verlauf des Kapitels mehrfach Anwendung finden wird.
Wir beginnen damit, dass wir gerade Momente $|X^\alpha(\rho_t^N)|^{2p}, p \geq 1$ des *regularisierten gbest* durch das entsprechende Moment der gemittelten Partikelpositionen $\int |x|^{2p} d\rho_t^N$ abschätzen:

Lemma 4.3.1: Abschätzung 1 (Grassi et al., 2021, Lemma 3.5), (Carrillo et al., 2018, Lemma 3.3, 3.4)

Für $p \geq 1, \alpha \geq 1$ existieren Konstanten $a_1, a_2 \in \mathbb{R}_{\geq 0}$, sodass für beliebige $F \in \mathfrak{F}$, $\mu \in \mathcal{P}_{2p}(\mathbb{R}^D)$ und

$$\eta_\mu^\alpha := \omega_F^\alpha \mu / ||\omega_F^\alpha||_{L^1(\mu)} \in \mathcal{P}(\mathbb{R}^D)$$
$$X^\alpha(\mu) := \int x d\eta_\mu^\alpha \in \mathbb{R}^D$$

folgende Abschätzung gilt:

$$|X^\alpha(\mu)|^{2p} = |\int x d\eta_{\mu_t}{}^\alpha|^{2p} \leq a_1 + a_2 \int |x|^{2p} d\mu \tag{4.9}$$

Beweis: Wir beginnen mit der Abschätzung für $p = 1$.
Zunächst haben wir mit Jensen Ungleichung wegen der Konvexität von $|\cdot|^2$:

$$|X^\alpha(\mu)|^2 = |\int x d\eta_\mu^\alpha|^2 \overset{Jensen}{\leq} \int |x|^2 d\eta_\mu^\alpha$$

In (Carrillo et al., 2018, Lemma 3.4) wird die Aussage

$$\int |x|^2 d\eta_\mu^\alpha \leq c_1 + c_2 \int |x|^2 d\mu$$

gezeigt, mit $c_1, c_2 \in \mathbb{R}_{\geq 0}$, die nur von den Konstanten in $\mathfrak{F}$ (3.2.1), nicht aber von μ abhängen. Dadurch erhalten wir wegen $p \geq 1$ und der Konvexität von $g : \mathbb{R}_{\geq 0} \to \mathbb{R}, x \mapsto x^p$, dass

$$\begin{aligned} |X^\alpha(\mu)|^{2p} &\overset{(4.11)}{\leq} 2^{p-1}(c_1^p + c_2^p(\int |x|^2 d\mu)^p) \\ &\overset{Jensen}{\leq} 2^{p-1}(c_1^p + c_2^p \int |x|^{2p} d\mu) \end{aligned}$$

Somit haben wir die Aussage mit $a_1 := 2^{p-1}c_1^p$, $a_2 := 2^{p-1}c_2^p$. ■

Für die Verwendung der obigen Abschätzung ist uns besonders wichtig, dass a_1, a_2 nicht von μ abhängen, da wir für μ die räumliche Randverteilung des N-Partikelsystems verwenden werden, das von N abhängt. Somit haben wir sichergestellt, dass a_1, a_2 keine Funktionen von N sind.
Das folgende Lemma gibt eine zeitlich gleichmäßige Abschätzung der Momente an, die in verschiedenen Stellen des Mean-Field Beweises als obere Schranke verwendet wird:

Lemma 4.3.2: Abschätzung der Momente (Grassi et al., 2021, Lemma 3.1)

Seien $F, f_0, Z_t^{i,N}$ und f_t^N wie in $\mathbb{A}$ (4.2.3). Dann existiert eine Konstante $K > 0$, **die nicht von N abhängt**, sodass

$$\sup_{i=1,\dots,N} \{ \sup_{t\in[0,T]} \mathbb{E}[|X_t^{i,N}|^2 + |X_t^{i,N}|^4 + |V_t^{i,N}|^2 + |V_t^{i,N}|^4]\} + \sup_{t\in[0,T]} \mathbb{E}[|X^\alpha(\rho_t^N)|^2 + |X^\alpha(\rho_t^N)|^4]\} \leq K.$$

Beweis: In (Grassi et al., 2021) wird der Leser auf (Carrillo et al., 2018, Lemma 3.4) verwiesen, wo der Beweis für das ähnliche *CBO* umrissen wird, daher ist an dieser Stelle das Ziel, den Beweis rigoros zu führen.
Für das gedächtnislose *SD-PSO* verläuft der Beweis ähnlich zu (Carrillo et al., 2018, Lemma 3.4). Ein Unterschied ist, dass in seiner Formulierung Partikel i durch zwei Prozesse dargestellt wird; seine Position X_t^i und Geschwindigkeit V_t^i. Außerdem enthält *SD-PSO* einen Trägheitsterm m, welcher berücksichtigt werden muss.

Wir betrachten das $2p$-te Moment des zusammengefügten Prozesses Z_t^i der Position und Geschwindigkeit eines Partikels in der $2-$Norm $|\cdot|$ und mit dem Standardskalarprodukt $\langle\cdot,\cdot\rangle$:

$$Z_t^i := [X_t^i, V_t^i] \in \mathbb{R}^{2D}.$$

Wir bemerken, dass

$$|Z_t^i|^2 = |X_t^i|^2 + |V_t^i|^2. \tag{4.10}$$

Summieren wir das über $i = 1, \dots, N$ und teilen durch N, können wir es äquivalent schreiben als

$$\int_{\mathbb{R}^{2D}} |z|^2 df_t^N = \int_{\mathbb{R}^D} |x|^2 d\rho_t^N + \int_{\mathbb{R}^D} |v|^2 d\nu_t^N$$

wobei das Räumliche Randmaß von f_s^N definiert ist durch

$$\begin{aligned}\rho_s^N(\cdot) &:= \int_{\mathbb{R}^D} f_s^N(\cdot, dv) = \int_{\mathbb{R}^D} \Big(\frac{1}{N}\sum_{i=1}^N \delta_{(X_s^i, V_s^i)}(\cdot, dv)\Big) \\ &= \frac{1}{N}\sum_{i=1}^N \delta_{X_s^i}(\cdot)\end{aligned}$$

Um die Ito-Formel (2.6) anzuwenden, setzen wir jetzt

$$g(t, Z_t^i) := |Z_t^i|^{2p} = \Big(\sum_{j=1}^D (Z_{t_j}^i)^2\Big)^p = \langle Z_t^i, Z_t^i\rangle^p$$

und erhalten für $j = 1, \dots, D$

$$\begin{aligned}&\frac{\partial g}{\partial t}(t, Z_t^i) = 0 \\ &\frac{\partial g}{\partial z_j}(t, Z_t^i) \overset{\text{kettenregel}}{=} 2pZ_{t_j}^i |Z_t^i|^{2(p-1)} \\ &\frac{\partial^2 g}{\partial z_j^2}(t, Z_t^i) \overset{\text{produktregel}}{=} 2p|Z_t^i|^{2(p-1)} + 4p(p-1)|Z_t^i|^{2(p-2)}((Z_t^i)_j)^2.\end{aligned}$$

Durch Anwendung der D-dimensionalen Ito-Formel (2.6) erhalten wir somit[2]:

$$\begin{aligned}d|Z_t^i|^{2p} &= 2p|Z_t^i|^{2(p-1)}\langle Z_t^i, dZ_t^i\rangle \\ &\quad + p|Z_t^i|^{2(p-1)}\langle dZ_t^i, dZ_t^i\rangle + 2p(p-1)|Z_t^i|^{2(p-2)}\sum_{d=1}^{2D}(Z_{t\,d}^i)^2(dZ_{t\,d}^i)^2 \\ &= 2p|Z_t^i|^{2(p-1)}\Big(\langle X_t^i, V_t^i\rangle dt \\ &\quad + \langle V_t^i, \Big(\frac{m-1}{m}V_t^i + \frac{\lambda}{m}(X^\alpha(\rho_t^N) - X_t^i)\Big)dt + \frac{\sigma}{m}(X^\alpha(\rho_t^N) - X_t^i)dB_t^i\rangle\Big) \\ &\quad + p|Z_t^i|^{2(p-1)}\frac{\sigma^2}{m^2}|X^\alpha(\rho_t^N) - X_t^i|^2 dt \\ &\quad + 2p(p-1)|Z_t^i|^{2(p-2)}\frac{\sigma^2}{m^2}\sum_{d=1}^D (V_{t\,d}^i)^2(X^\alpha(\rho_t^N) - X_t^i)_d^2 dt.\end{aligned}$$

Der Diffusionsterm ist unproblematisch, da er nach Anwendung des Erwartungswertes wegfallen wird (cf. (4)). Daher ist unser Ziel im Folgenden lediglich, nach Integration die obigen Driftterme durch $|Z_t^i|^{2p}$ abzuschätzen, um die Gronwallsche Ungleichung an-

[2]Wir lassen Terme der Form $dZ_{t\,d}^i dZ_{t\,j}^i$ für $d \neq j$ aus, da diese nach den Rechenregeln (2.7) später wegfallen werden.

wenden zu können. Für diese Abschätzung sind folgende drei Ungleichungen hilfreich: Zuerst verwenden wir wie in (Carrillo et al., 2018, Lemma 3.4) die Ungleichung

$$(a+b)^q \leq 2^{q-1}(a^q+b^q), \quad \forall q \geq 1; a,b \in \mathbb{R} \tag{4.11}$$

Weiter überlegen wir uns anhand komponentenweiser Anwendung der binomischen Formel und (4.11) Ungleichungen der Form

$$\begin{aligned} \langle V_s^i, X_s^i \rangle &\leq |V_s^i + X_s^i|^2 \\ &\overset{(4.11)}{\leq} 2(|V_s^i|^2 + |X_s^i|^2) \\ &= 2|Z_s^i|^2 \end{aligned} \tag{4.12}$$

Zuletzt haben wir, da $a^q, q \geq 1, a \geq 0$ monoton in a ist, die Abschätzung

$$a^q b^r \leq \max(a,b)^{q+r} \leq a^{q+r} + b^{q+r}, \quad \forall a,b \in \mathbb{R}_{\geq 0}, q,r \in \mathbb{R}_{\geq 1}. \tag{4.13}$$

Hiermit erreichen wir für die Erwartung des $2p$-ten Momenten des Prozesses in Integraler Form folgende Abschätzung:[3]

$$\begin{aligned} \mathbb{E}|Z_t^i|^{2p} = \mathbb{E}|Z_0^i|^{2p} + \int_0^t &\mathbb{E}[2p|Z_s^i|^{2(p-1)}(\langle X_s^i, V_s^i \rangle \\ &+ \langle V_s^i, (\frac{m-1}{m} V_s^i + \frac{\lambda}{m}(X^\alpha(\rho_s^N) - X_s^i)) \rangle) \\ &+ p|Z_s^i|^{2(p-1)} \frac{\sigma^2}{m^2} |X^\alpha(\rho_s^N) - X_s^i|^2 \\ &+ 2p(p-1)|Z_s^i|^{2(p-2)} \frac{\sigma^2}{m^2} \sum_{d=1}^D (V_{s\,d}^i)^2 (X^\alpha(\rho_s^N) - X_s^i)_d^2] ds. \end{aligned}$$

[3]Wir lassen den Diffusionsterm direkt wegfallen, obwohl er technisch erst später bei Anwendung der Linearität von $\mathbb{E}$ wegfällt.

Indem wir die Koeffizienten durch ihre Summe abschätzen und herausziehen, erhalten wir

$$
\begin{aligned}
\mathbb{E}|Z_t^i|^{2p} \leq{}& \mathbb{E}|Z_0^i|^{2p} + (2p\frac{m-1+\lambda}{m} + (p+2p(p-1))\frac{\sigma^2}{m^2}) \\
&\cdot \int_0^t \mathbb{E}[|Z_s^i|^{2(p-1)}(\langle X_s^i, V_s^i\rangle \\
&+ \langle V_s^i, V_s^i\rangle + \langle V_s^i, X^\alpha(\rho_s^N)\rangle - \langle V_s^i, X_s^i\rangle) \\
&+ |Z_s^i|^{2(p-1)}|X^\alpha(\rho_s^N) - X_s^i|^2 \\
&+ |Z_s^i|^{2(p-2)} \sum_{d=1}^{D} (V_{s\,d}^i)^2 (X^\alpha(\rho_s^N) - X_s^i)_d^2] ds \\
\overset{(4.12)}{\leq}{}& \mathbb{E}|Z_0^i|^{2p} + (2p\frac{m-1+\lambda}{m} + (p+2p(p-1))\frac{\sigma^2}{m^2}) \\
&\cdot \int_0^t \mathbb{E}[|Z_s^i|^{2(p-1)}(2|Z_s^i|^2 \\
&+ |V_s^i|^2 + 2(|V_s^i| + |X^\alpha(\rho_s^N)|^2) + |Z_s^i|^2) \\
&+ |Z_s^i|^{2(p-1)}(|X^\alpha(\rho_s^N)|^2 + |X_s^i|^2) \\
&+ |Z_s^i|^{2(p-2)} \sum_{d=1}^{D} (V_{s\,d}^i)^2 (X^\alpha(\rho_s^N) - X_s^i)_d^2] ds. \qquad (4.14)
\end{aligned}
$$

Nun erhalten wir durch Summieren über $i = 1, \ldots, N$ und teilen durch N mit der Abschätzung 4.3.1 Folgendes[4]:

$$
\begin{aligned}
\mathbb{E}\int |z|^{2p} df_t^N \leq \mathbb{E}\int |z|^{2p} df_0^{\otimes N} &+ (2p\frac{m-1+\lambda}{m} + (p+2p(p-1))\frac{\sigma^2}{m^2}) \\
&\cdot \int_0^t \mathbb{E}[\int |z|^{2(p-1)}(3|z|^2 \\
&+ |v|^2 + 2(|v|^2 + a_1 + a_2|x|^2)) \\
&+ |z|^{2(p-1)}(a_1 + a_2|x|^2 + |x|^2) \\
&+ |z|^{2(p-2)} + (\sum_{d=1}^{D} v_d^2 x_d^2 + v_d^2 X^\alpha(\rho_s^N)_d^2) f_s^N(dx,dv)]ds \\
&\overset{(4.13)}{\leq} \mathbb{E}\int |z|^{2p} df_0^{\otimes N} + (2p\frac{m-1+\lambda}{m} + (p+2p(p-1))\frac{\sigma^2}{m^2}) \\
&\cdot \int_0^t \mathbb{E}[\int |z|^{2(p-1)}(3|z|^2 \\
&+ 3|v|^2 + 3(a_1 + a_2|x|^2) + |x|^2) \\
&+ |z|^{2(p-2)} + (|x|^4 + 2|v|^4 + a_1 + a_2|x|^4) f_s^N(dx,dv)]ds \\
&\leq \mathbb{E}\int |z|^{2p} df_0^{\otimes N} + (2p\frac{m-1+\lambda}{m} + (p+2p(p-1))\frac{\sigma^2}{m^2}) \\
&\cdot \int_0^t \mathbb{E}[\int (4(a_2+1)) + 2 + a_2)|z|^{2p} \\
&+ a_1(3|z|^{2(p-1)} + |z|^{2(p-2)}) f_s^N(dx,dv)]ds \\
&\overset{Jensen}{\leq} \mathbb{E}\int |z|^{2p} df_0^{\otimes N} + (2p\frac{m-1+\lambda}{m} + (p+2p(p-1))\frac{\sigma^2}{m^2}) \\
&\cdot (4(a_2+1) + 2 + a_2 + 3a_1 + 1) \cdot \int_0^t \mathbb{E}[\int |z|^{2p} f_s^N(dx,dv)]ds \\
&=: \mathbb{E}\int |z|^{2p} df_0^{\otimes N} + C(\lambda, m, \sigma, p, a_1, a_2) \\
&\cdot \int_0^t \mathbb{E}[\int |z|^{2p} f_s^N(dx,dv)]ds. \qquad (4.15)
\end{aligned}
$$

Auf die letzte obige Abschätzung können wir nun die Gronwallsche Ungleichung anwenden und erhalten für alle $t \in [0,T]$

$$
\begin{aligned}
\mathbb{E}\int |z|^{2p} df_t^N &\leq \mathbb{E}\int |z|^{2p} df_0^{\otimes N} \exp(C(\lambda, m, \sigma, p, a_1, a_2)T) \qquad (4.16) \\
&=: K(\lambda, m, \sigma, p, a_1, a_2, f_0, T)
\end{aligned}
$$

[4]Indem wir die Notation ausnutzen, schreiben wir für $q \in [0,\infty)$: $|z|^{2q} := (|x|^2 + |v|^2)^q$ (cf. (4.10)).

Die Schranke K hängt trotz des Terms $\mathbb{E}\int|z|^{2p}df_0^{\otimes N}$ nicht von N ab, weil für beliebiges $g:\mathbb{R}^D\to\mathbb{R}$ gilt

$$\begin{aligned}\mathbb{E}[\int_{\mathbb{R}^{2D}}g(z)f_0^N(dz)] &= \mathbb{E}[\frac{1}{N}\sum_{i=1}^{N}g(Z_0^i)]\\ &\overset{\mathbb{E}\,linear}{=} \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}_{Z_0^i\sim f_0}[(g(Z_0^i)]\\ &\overset{ident.\,vert.}{=} \mathbb{E}_{Z_0^1\sim f_0}[g(Z_0^1)],\end{aligned}$$

da die Partikel n.V. anfangs identisch verteilt sind ($Z_0^{i,N}\overset{\text{u.i.v.}}{\sim}f_0, i=1,\dots N$).
Aus (4.16) ergibt sich sofort 4.3.2 dadurch, dass man $p=1,2$ setzt, was wegen $f_0\in\mathcal{P}_4(\mathbb{R}^{2D})$ zu $K<\infty$ führt. Also hat man

$$\begin{aligned}\infty > K &\overset{(4.16)}{\geq} \sup_{t\in[0,T]}\mathbb{E}\int|z|^{2p}df_t^N\\ \Rightarrow 2^{1-2p}K &\overset{(4.11),\int linear}{\geq} \sup_{t\in[0,T]}\mathbb{E}\left[\int|x|^{2p}d\rho_t^N+\int|v|^{2p}d\nu_t^N\right]\\ \Rightarrow 2^{1-2p}\frac{K}{a_2} &\overset{(4.9)}{\geq} \sup_{t\in[0,T]}\mathbb{E}|X^\alpha(\rho_t^N)|^{2p}.\end{aligned}$$

Die übrige Abschätzung der Form

$$K' \geq \sup_{i=1,\dots,N}\left\{\sup_{t\in[0,T]}\mathbb{E}\left[|X_t^i|^{2p}+|V_t^i|^{2p}\right]\right\}$$

erhält man zum Beispiel, in dem man eine Abschätzung wie in (4.15) direkt auf (4.14) macht, ohne das Mittel über die Partikel zu bilden, und dann auf dieses Ergebnis die Gronwallsche Ungleichung anwendet. Zusammengenommen schließen diese Abschätzungen den Beweis von 4.3.2 ab. ■

4.4 Straffheit der empirischen Maße

Satz 4.4.1: Straffheit der Partikelverteilungen (Grassi et al., 2021, Thm 3.3)

Seien $F, f_0, Z_t^{i,N}$ und f^N wie in $\mathbb{A}$ (4.2.3). Dann ist die Folge $\{\mathbb{P}^{f^N}\}_{N\in\mathbb{N}}$ straff in $\mathcal{P}(\mathcal{P}(C([0,T];\mathbb{R}^{2D})))$.

Beweis: Wir können Satz 2.3.2 anwenden:

1. $C([0,T];\mathbb{R}^{2D})$ ist polnisch laut 4.1.1.
2. Unter Permutationen $\sigma \in S_n$ der N Partikel ist f_t^N invariant aufgrund der Kommutativität der Addition in (4.7).

Somit reicht es, die Straffheit der Folge $\{\mathbb{P}^{Z^{1,N}}\}_{N\in\mathbb{N}}$ in $\mathcal{P}(C([0,T];\mathbb{R}^{2D}))$ nachzuweisen. Dafür verwenden wir die Kriterien von Aldous aus Lemma 2.3.2.

Bedingung 1: Wir können aufgrund der Voraussetzungen die Momentenabschätzung 4.3.2 anwenden und erhalten ein K, unabhängig von N, sodass $\mathbb{E}[|Z_t^{1,N}|^2] \leq K$. Für $\varepsilon > 0$ ist $C_\varepsilon := \{z||z|^2 \leq \frac{K}{\varepsilon}\}$ offensichtlich kompakt. Mithilfe Markovs Ungleichung erhalten wir daher

$$\begin{aligned}\mathbb{P}^{(Z^{1,N})}(C_\varepsilon^c) &= \mathbb{P}(|Z_t^{1,N}|^2 > \frac{K}{\varepsilon})\\ &\overset{Markov}{\leq} \frac{\varepsilon}{K}\mathbb{E}[|Z_t^{1,N}|^2] \overset{4.3.2}{\leq} \varepsilon, \forall N \in \mathbb{N}.\end{aligned}$$

Also ist für $t \in [0,T]$ fix, z.B. $t = 0$, die Folge $\mathbb{P}^{Z^{1,N}}(0)$ straff auf $\mathbb{R}^{2D}$.
Bedingung 2: Für eine vorgegebene $\sigma(Z_s^{1,N}, s \in [0,T])$-Stoppzeit β mit diskreten Werten, s.d. $\beta + \delta_0 \leq T$, wollen wir zunächst eine Konstante $C < \infty$, unabhängig von N, finden, s.d. wir auf

$$\mathbb{E}\left[|Z_{\beta+\delta}^{1,N} - Z_\beta^{1,N}|^2\right] = \mathbb{E}\left[|X_{\beta+\delta}^{1,N} - X_\beta^{1,N}|^2\right] + \mathbb{E}\left[|V_{\beta+\delta}^{1,N} - V_\beta^{1,N}|^2\right] \leq C\delta \qquad (4.17)$$

ebenfalls Markovs Ungleichung anwenden können.

Es gilt für $\delta \in (\beta, \beta + \delta_0)$:

$$
\begin{aligned}
\mathbb{E}\left[|X^{1,N}_{\beta+\delta} - X^{1,N}_{\beta}|^2\right] &\overset{(3.4)}{=} \mathbb{E}\left[|\int_{\beta}^{\beta+\delta} V^{1,N}_s ds|^2\right] \\
&\overset{Jensen}{\leq} \mathbb{E}\left[\int_{\beta}^{\beta+\delta} |V^{1,N}_s|^2 ds\right] \\
&\overset{Fubini}{=} \int_{\beta}^{\beta+\delta} \mathbb{E}\left[|V^{1,N}_s|^2\right] ds \\
&\leq \int_0^T \mathbb{E}\left[|V^{1,N}_s|^2\right] ds \\
&\overset{4.3.2}{\leq} \delta K T,
\end{aligned}
\tag{4.18}
$$

wobei wir die Integrationsreihenfolge wegen der Momentenabschätzung tauschen durften.

Damit ist der Positions-Summand von (4.17) abgeschätzt. Für den Geschwindigkeitsteil helfen folgende Überlegungen:

$$
\begin{aligned}
\mathbb{E}\left[|V^{1,N}_{\beta+\delta} - V^{1,N}_{\beta}|^2\right] &\overset{(3.3)}{=} \mathbb{E}[| - \frac{1-m}{m}\int_{\beta}^{\beta+\delta} V^{1,N}_s ds + \frac{\lambda}{m}\int_{\beta}^{\beta+\delta} (X^{\alpha}(\rho^N_s) - X^{1,N}_s) ds \\
&\quad + \frac{\sigma}{m}\int_{\beta}^{\beta+\delta} \mathbb{D}(X^{\alpha}(\rho^N_s) - X^{1,N}_s) dB^1_s|^2] \\
&\overset{Jensen}{\leq} (\frac{1-m}{m})^2 \mathbb{E}\left[|\int_{\beta}^{\beta+\delta} V^{1,N}_s ds|^2\right] \\
&\quad + (\frac{\lambda}{m})^2 \mathbb{E}\left[|\int_{\beta}^{\beta+\delta} (X^{\alpha}(\rho^N_s) - X^{1,N}_s) ds|^2\right] \\
&\quad + (\frac{\sigma}{m})^2 \mathbb{E}\left[|\int_{\beta}^{\beta+\delta} \mathbb{D}(X^{\alpha}(\rho^N_s) - X^{1,N}_s) dB^1_s|^2\right]
\end{aligned}
$$

Auf die erste Erwartung können wir dieselbe Abschätzung wie für den Positionsteil verwenden. Für den zweiten Teil verfahren wir ebenfalls wie in den Abschätzungen (4.18):

$$
\begin{aligned}
\mathbb{E}\left[|\int_{\beta}^{\beta+\delta} X^{\alpha}(\rho^N_s) - X^{1,N}_s ds|^2\right] &\leq \delta \mathbb{E}\left[|\int_0^T X^{\alpha}(\rho^N_s) - X^{1,N}_s) ds|^2\right] \\
&\overset{(4.11)}{\leq} 2T\delta \sup_{t\in[0,T]} (\mathbb{E}\left[|X^{1,N}_t|^2\right] + \mathbb{E}\left[|X^{\alpha}(\rho^N_t)|^2\right]) \\
&\overset{4.3.2}{\leq} 2TK\delta
\end{aligned}
$$

Für den letzten Term verwenden wir die Ito-Isometrie:

$$\begin{aligned}
\mathbb{E}\left[|\int_{\beta}^{\beta+\delta}\mathbb{D}(X^{\alpha}(\rho_s^N)-X_s^{1,N})dB_s^1|^2\right] &\overset{Ito-Isom.}{=} \mathbb{E}\left[\int_{\beta}^{\beta+\delta}|X^{\alpha}(\rho_s^N)-X_s^{1,N}|^2 ds\right] \\
&\overset{binom.\ Form.}{\leq} \mathbb{E}\left[\int_{\beta}^{\beta+\delta}|X^{\alpha}(\rho_s^N)|^2+|X_s^{1,N}|^2 ds\right] \\
&\leq 2TK\delta.
\end{aligned}$$

Insgesamt haben wir also

$$\begin{aligned}
&\mathbb{E}[|X_{\beta+\delta}^{1,N}-X_{\beta}^{1,N}|^2+|V_{\beta+\delta}^{1,N}-V_{\beta}^{1,N}|^2] \\
&\leq TK((2(\frac{\lambda}{m})^2+(\frac{1-m}{m})^2)+2\frac{\sigma^2}{m^2})\delta \\
&=: C(T,K,\lambda,m,\sigma)\delta < \infty.
\end{aligned} \tag{4.19}$$

Um nun das zweite Aldous-Kriterium zu erfüllen, setzen wir für gegebene $\varepsilon,\eta > 0$

$$\delta_0 := \min\left(\frac{\varepsilon\eta}{C(T,K,\lambda,m,\sigma)}, T\right)$$

und wenden die Markov-Ungleichug an:

$$\sup_{\delta\in[0,\delta_0]}\mathbb{P}(|Z_{\beta+\delta}^{1,N}-Z_{\beta}^{1,N}|^2\rangle\eta) \overset{Markov}{\leq} \sup_{\delta\in[0,\delta_0]}\frac{1}{\eta}\mathbb{E}\left[|Z_{\beta+\delta}^{1,N}-Z_{\beta}^{1,N}|^2\right] \leq \varepsilon.$$

Die rechte Ungleichung gilt dabei für das Supremum über $\delta \in [0,\delta_0]$, da die oben erhaltene Abschätzung (4.19) monoton wachsend in δ ist. Das schließt den Beweis des Satzes ab. ∎

4.5 Konvergenz der empirischen Maße

Ziel dieses Abschnitts ist, die fast sichere Konvergenz der gesamten Folge der empirischen Maße f^N gegen ein f, das eine schwache Lösung der FPG ist, zu zeigen. Diese zentrale Aussage wird im folgenden Satz 4.5.1 gemacht und ihr Beweis, welcher in einem Kompaktheitsargument besteht, zunächst auf einem abstrakten Niveau gegeben. Dabei wird auf die jeweiligen Lemmata verwiesen, welche die einzelne Schritte des Beweises zeigen und in ihren eigenen Unterabschnitten folgen.

Satz 4.5.1: Fast sichere Konvergenz der empirischen Maße gegen eine schwache Lösung der FPG

Gelte $\mathbb{A}$ (4.2.3). Dann existiert ein $f \in \mathcal{P}(C([0,T];\mathbb{R}^{2D}))$, und ein Wahrscheinlichkeitsraum $(\Omega, \mathscr{A}, \tilde{\mathbb{P}})$, sodass die Folge der empirischen Maße f^N als Folge von Zufallsvariablen in Ω $\tilde{\mathbb{P}}$-fast sicher gegen f konvergiert:

$$f^N \overset{\tilde{\mathbb{P}}-\text{f.s.}}{\rightarrow} f.$$

Der Grenzwert f ist die eindeutige schwache Lösung der FPG 4.2.1.

Beweis: Die Existenz des Grenzwerts und seine Eigenschaften werden durch ein Kompaktheitsargument gezeigt, das folgende Schritte hat:

1. Die Familie $\mathscr{F} = \{\mathbb{P}^{f^N}\}_{N\in\mathbb{N}}$ ist straff in $\mathcal{P}(\mathcal{P}(\mathcal{C}([0,T];\mathbb{R}^{2D})))$ laut dem Satz 4.4.1 des letzten Abschnitts.

2. Aus Lemma 4.1.1 und dem Satz von Prohorov 2.3.3 folgt: $\mathscr{F}$ ist schwach relativ folgenkompakt (**srfk**).

3. Der folgende Abschnitt 4.5.1, insbesondere Lemma 4.5.5, zeigt, dass der Grenzwert einer beliebigen als *konvergent* in Verteilung angenommenenn Teilfolge f^{N_k} eine Schwache Lösung f der FPG 4.2.1 ist.[5]

4. Im darauf folgenden Abschnitt 4.5.2 zeigt Lemma 4.5.7 anhand eines Eindeutigkeitsbeweises, dass der Grenzwert aller *konvergenten* Teilfolgen gleich ist.

5. Da jede beliebige Teilfolge f^{N_j} wegen **srfk** wiederum eine konvergente Teilfolge hat, ist f^{N_j} ebenfalls konvergent, also kann es keine divergente Teilfolge geben, und somit konvergiert die gesamte Folge $f^N \overset{\mathcal{D}}{\rightarrow} f$.

Wir verwenden also die Eindeutigkeit der Lösung der FPG, um zu zeigen, dass die gesamte Folge $f^N \overset{\mathcal{D}}{\rightarrow} f$ in $\mathcal{P}(C([0,T];\mathbb{R}^{2D}))$ konvergiert.
Nach Def. 2.3.1 heißt das, dass

$$\mathbb{P}^{f^N} \overset{n\to\infty}{\rightarrow} \mathbb{P}^f \text{ (schwach) auf } \mathcal{P}(C([0,T];\mathbb{R}^{2D})).$$

Da dieser Raum laut Lemma 4.1.2 polnisch ist, gibt es somit dank dem Lemma von Skorokhod 2.3.1 Representationen Y^N, Y von f^N, f auf einem Raum $(\Omega, \mathscr{A}, \tilde{\mathbb{P}})$, in dem Sinne,

[5] Hier können wir bereits mit fast sicherer Konvergenz einer Representation von f^{N_k} arbeiten.

dass

$$\tilde{\mathbb{P}}^{Y^N} = \mathbb{P}^{f^N}; \quad \tilde{\mathbb{P}}^{Y} = \mathbb{P}^{f}$$
$$Y^N \overset{\tilde{\mathbb{P}}-f.s.}{\to} Y.$$

Dies schließt den Beweis des Satzes ab. ■

Im weiteren Verlauf des Kapitels (D.h.: Innerhalb des Beweises des obigen Satzes.) können wir mit diesen Representationen von f^N und f arbeiten, sprich z.B. im dritten Schritt oben bereits annehmen, dass die Teilfolge f^{N_k} $\tilde{\mathbb{P}}$-f.s. gegen ein f konvergiert. Wir schreiben im Folgenden trotzdem weiterhin $\mathbb{P} := \tilde{\mathbb{P}}$.

4.5.1 Grenzwerte konvergenter Teilfolgen lösen die FPG schwach

Die ersten zwei Schritte des Beweises von 4.5.1 sind bereits gegeben. Als nächstes wollen wir also Schritt 3. des Beweises führen, also für eine als in Verteilung gegen ein f konvergent angenommene Teilfolge $\{f^{N_k}\}_{k\in\mathbb{N}}$ der empirischen Maße zeigen, dass der Grenzwert f eine Schwache Lösung der Fokker-Planck Gleichung gemäß Definition 4.2.2 ist. Die Strategie dieses Abschnitts ist daher, wie in (Grassi et al., 2021, S. 16-20) einige sinnvolle Abschätzungen zu erhalten, die erlauben zu zeigen dass

1. $\mathbb{F}_\varphi(f^{N_k}) \overset{\mathscr{L}_1(\mathbb{P})}{\to} 0$ und
2. $\mathbb{F}_\varphi(f^{N_k}) \overset{\mathscr{L}_1(\mathbb{P})}{\to} \mathbb{F}_\varphi(f)$.

Der erste Teil ist relativ leicht zu erhalten, indem wir $\mathscr{L}_2(\mathbb{P})$-Konvergenz betrachten, und wird in folgendem Lemma gezeigt:

Lemma 4.5.1: Abschätzung des Funktionals
(Grassi et al., 2021, Lemma 3.4), (Huang and Qiu, 2021, Prop. 3.2, Eq. (4.13))

Sei f^N wie in $\mathbb{A}$ (4.2.3). Dann gilt für alle $\varphi \in C_c^2(\mathbb{R}^{2D})$:

$$\mathbb{F}_\varphi(f^N) \quad \overset{\mathscr{L}_2(\mathbb{P})}{\to} \quad 0.$$

Beweis: Wir zeigen genauer, dass eine Konstante

$$C(\sigma, K, T, ||\nabla\varphi||_\infty) > 0$$

existiert, die nicht von N abhängt, sodass

$$\mathbb{E}\left[|\mathbb{F}_\varphi(f^N)|^2\right] \leq \frac{C}{N},$$

woraus direkt die Behauptung folgt. Indem wir (4.7) in (4.2) einsetzen, können wir $\mathbb{F}_\varphi(f^N)$ schreiben als[6]

$$\begin{aligned}
\mathbb{F}_\varphi(f^N) &:= \int_{\mathbb{R}^{2D}} \phi(x)df_t^N - \int_{\mathbb{R}^{2D}} \phi(x)df_0^N + \int_0^t L\phi dt \\
&= \frac{1}{N}\sum_{i=1}^N \varphi(Z_t^{i,N}) - \varphi(Z_0^{i,N}) \\
&\quad + \int_0^t \frac{1}{N}\sum_{i=1}^N V_s^{i,N} \cdot \nabla_x \varphi(Z_s^{i,N})ds \\
&\quad - \frac{1-m}{m}\int_0^t \frac{1}{N}\sum_{i=1}^N V_s^{i,N}\nabla_v\varphi(Z_s^{i,N})ds \\
&\quad + \frac{\lambda}{m}\int_0^t \frac{1}{N}\sum_{i=1}^N (X_s^{i,N} - X^\alpha(\rho^{f_s^N}))\nabla_v\varphi(Z_s^{i,N})ds \\
&\quad - \frac{\sigma^2}{2m^2}\int_0^t \sum_{d=1}^D (X_s^{i,N} - X^\alpha(\rho^{f_s^N}))_d^2 \frac{\partial^2}{\partial {v_d}^2}\varphi(Z_s^{i,N})ds \\
&\overset{\int \text{ linear}}{=} \frac{1}{N}\sum_{i=1}^N ((\varphi(Z_t^{i,N}) - \varphi(Z_0^{i,N}) \\
&\quad + \int_0^t V_s^{i,N} \cdot (\nabla_x\varphi(Z_s^{i,N}) - \frac{1-m}{m}\nabla_v\varphi(Z_s^{i,N}))ds \\
&\quad + \frac{\lambda}{m}\int_0^t (X_s^{i,N} - X^\alpha(\rho^{f_s^N}))\nabla_v\varphi(Z_s^{i,N})ds \\
&\quad - \frac{\sigma^2}{2m^2}\int_0^t \sum_{d=1}^D (X_s^{i,N} - X^\alpha(\rho^{f_s^N}))_d^2 \frac{\partial^2}{\partial {v_d}^2}\varphi(Z_s^{i,N})ds.
\end{aligned} \tag{4.20}$$

Da φ zweimal stetig diffbar ist, können wir die Ito-Formel (2.6) für $2D$ Dimensionen auf $\varphi(Z_t^{i,N})$ anwenden und erhalten für $i = 1, \ldots, N$

[6](Grassi et al., 2021, S. 16)

$$
\begin{aligned}
d\varphi(Z_t^{i,N}) = V_t^{i,N} \cdot \Big(&\frac{1-m}{m}\nabla_v \varphi(Z_t^{i,N}) - \nabla_x \varphi(Z_t^{i,N}\Big) \\
&- \frac{\lambda}{m}\mathbb{D}(X_t^{i,N} - X^\alpha(\rho^{f_t^N}))\nabla_v \varphi(Z_t^{i,N}))dB_t^i \\
&+ \frac{\sigma^2}{2m^2}\sum_{d=1}^{D}(X_t^{i,N} - X^\alpha(\rho^{f_t^N}))_d^2 \frac{\partial^2}{\partial {v_d}^2}\varphi(Z_t^{i,N})dt,
\end{aligned}
$$

beziehungsweise in integraler Form

$$
\begin{aligned}
\varphi(Z_t^{i,N}) = \varphi(Z_0^{i,N}) + \int_0^t V_s^{i,N} \cdot \Big(&\frac{1-m}{m}\nabla_v \varphi(Z_s^{i,N})ds - \nabla_x \varphi(Z_s^{i,N})\Big) \\
&- \frac{\lambda}{m}\mathbb{D}(X_s^{i,N} - X^\alpha(\rho^{f_s^N}))\nabla_v \varphi(Z_s^{i,N})dB_s^i \\
&+ \frac{\sigma^2}{2m^2}\int_0^t \sum_{d=1}^{D}(X_s^{i,N} - X^\alpha(\rho^{f_s^N}))_d^2 \frac{\partial^2}{\partial {v_d}^2}\varphi(Z_s^{i,N})ds.
\end{aligned}
\tag{4.21}
$$

Setzen wir also (4.21) in (4.20) für alle $\varphi(Z_t^{i,N})$ ein, erhalten wir

$$
\mathbb{F}_\varphi(f^N) = \frac{\lambda}{m}\frac{1}{N}\sum_{i=1}^{N}\int_0^t \mathbb{D}(X_t^{i,N} - X^\alpha(\rho^{f_s^N}))\nabla_v \varphi(Z_s^{i,N})dB_s^i.
$$

Folglich gilt

$$
\begin{aligned}
\mathbb{E}\left[|\mathbb{F}_\varphi(f^N)|^2\right] &= \frac{\lambda^2}{m^2N^2}\mathbb{E}\left[|\sum_{i=1}^N \int_0^t \mathbb{D}(X_t^{i,N} - X^\alpha(\rho^{f_s^N}))\nabla_v\varphi(Z_s^{i,N})dB_s^i|^2\right] \\
&\overset{\text{dreiecksungl.}}{\leq} \frac{\lambda^2}{m^2N^2}\sum_{i=1}^N \mathbb{E}\left[|\int_0^t \mathbb{D}(X_t^{i,N} - X^\alpha(\rho^{f_s^N}))\nabla_v\varphi(Z_s^{i,N})dB_s^i|^2\right] \\
&\overset{\text{Ito Isom.}}{=} \frac{\lambda^2}{m^2N^2}\sum_{i=1}^N \mathbb{E}\left[\int_0^t \sum_{d=1}^D (X_t^{i,N} - X^\alpha(\rho^{f_s^N}))_d^2(\nabla_v\varphi(Z_s^{i,N}))_d^2 ds\right] \\
&= \frac{\lambda^2}{m^2N}\frac{1}{N}\sum_{i=1}^N \mathbb{E}\left[\int_0^t |(X_t^{i,N} - X^\alpha(\rho^{f_s^N}))(\nabla_v\varphi(Z_s^{i,N}))|^2 ds\right] \\
&\leq ||\nabla_v\varphi||_\infty^2 \frac{\lambda^2}{m^2N}\frac{1}{N}\sum_{i=1}^N \mathbb{E}\left[\int_0^t |X_s^{i,N} - X^\alpha(\rho^{f_s^N})|^2 ds\right] \\
&\leq ||\nabla_v\varphi||_\infty^2 \frac{\lambda^2}{m^2N}\frac{1}{N}\sum_{i=1}^N \mathbb{E}\left[\int_0^t |X_s^{i,N} - X^\alpha(\rho^{f_s^N})|^2 ds\right] \\
&\overset{(4.11)}{\leq} 2||\nabla_v\varphi||_\infty^2 \frac{\lambda^2}{m^2N}\frac{1}{N}\sum_{i=1}^N \mathbb{E}\left[\int_0^t |X_s^{i,N}|^2 + |X^\alpha(\rho^{f_s^N})|^2 ds\right] \\
&\overset{4.3.2}{\leq} \frac{1}{N}\cdot(2||\nabla_v\varphi||_\infty^2 \frac{\lambda^2}{m^2}\cdot K) \\
&=: \frac{1}{N}\cdot C(||\nabla_v\varphi||_\infty, \lambda, m, K).
\end{aligned}
$$

Dies schließt den Beweis des Lemmas ab. ■

Damit ist der erste Teil der Beweisstrategie dieses Abschnitts 4.5.1 erfüllt. Die Behauptung, dass $\mathbb{F}_\varphi(f^{N_k}) \overset{\mathscr{L}_1 \to \infty}{\to} \mathbb{F}_\varphi(f)$ ist schwerer zu erhalten, und hierfür müssen wir zur Skorokhod Representation 2.3.1 unserer als in Verteilung konvergent angenommenen Teilfolge übergehen, d.h. wir können hier bereits wie in Satz 4.5.1 beschrieben mit fast sicherer Konvergenz der f^{N_k} als Zufallsvariablen arbeiten (Grassi et al., 2021, S. 16).
Zunächst fassen wir zwei sinnvolle Abschätzungen in einem Lemma zusammen:

Lemma 4.5.2: (Grassi et al., 2021, S. 16,17), (Huang and Qiu, 2021, S. 6,7)

Sei $\{f^N\}_{N\in\mathbb{N}}$ wie in $\mathbb{A}$ (4.2.3). Sei ferner $\{f^{N_k}\}_{k\in\mathbb{N}}$ eine Teilfolge, die in $\mathbb{P}$-f.s. gegen ein f konvergiert. Definiere für $t \in [0,T], \phi \in \mathcal{C}_b(\mathbb{R}^{2D})$

$$M_t^k(\phi) := |\int_{\mathbb{R}^{2D}} \phi f_t^{N_k}(dz) - \int_{\mathbb{R}^{2D}} \phi f_t(dz)|$$
$$D_t^k := |X^\alpha(\rho_t^{N_k}) - X^\alpha(\rho_t)|$$

Dann gilt für alle $t \in [0,T], \phi \in \mathcal{C}_b(\mathbb{R}^{2D})$ (1.)

$$\lim_{k\to\infty} M_t^k(\phi) + D_t^k = 0,\ \mathbb{P}\text{-f.s.}$$

und außerdem (2.)

$$\lim_{k\to\infty} \mathbb{E}\left[M_t^k(\phi)^2 + (D_t^k)^2\right] = 0. \tag{4.22}$$

Beweis: (1.): Wegen der $\mathbb{P}$-f.s. Konvergenz gilt

$$\int_{\mathbb{R}^{2D}} \phi f_t^{N_k}(dz) - \int_{\mathbb{R}^{2D}} \phi f_t(dz) \overset{k\to\infty}{\to} 0$$

Indem man zur Skorokhod Representation übergeht, gilt diese Aussage $\mathbb{P}$-f.s.. Diese Argumentation funktioniert auch für das regularisierte *gbest* $X^\alpha(\cdot)$, da wegen Annahmen 2. und 3. in 3.2.1 $\omega_\alpha^F(x), x\omega_\alpha^F(x) \in \mathcal{C}_b(\mathbb{R}^()\mathbb{R}^D)$ sind, also:

$$\begin{aligned}
\lim_{k\to\infty} X^\alpha(\rho_t^{N_k}) &= \lim_{k\to\infty} \frac{\int_{\mathbb{R}^D} x\omega_\alpha^F(x)\rho_t^{N_k}(dx)}{\int_{\mathbb{R}^D} \omega_\alpha^F(x)\rho_t^{N_k}(dx)} \\
&= \lim_{k\to\infty} \frac{\int_{\mathbb{R}^D} xe^{-\alpha F(x)}\rho_t^{N_k}(dx)}{\int_{\mathbb{R}^D} e^{-\alpha F(x)}\rho_t^{N_k}(dx)} \\
&= \frac{\int_{\mathbb{R}^D} xe^{-\alpha F(x)}\rho_t(dx)}{\int_{\mathbb{R}^D} e^{-\alpha F(x)}\rho_t(dx)},\ \mathbb{P}\text{-f.s.} \\
&= X^\alpha(\rho_t),\ \mathbb{P}\text{-f.s.}
\end{aligned}$$

wobei $\rho_t(\cdot) := \int_{\mathbb{R}^D} f_t(\cdot, v)dv$.

Die Aussage (2.) gilt, da für $A > 0$ gilt:

$$\begin{aligned}
&\sup_{t\in[0,T]} \mathbb{E}\left[\int_{\mathbb{R}^{2D}} ((|x|^4+|v|^4)\wedge A) f_t(dx,dv)\right] \\
&\overset{\text{schwache Konv.}}{=} \sup_{t\in[0,T]} \mathbb{E}\left[\lim_{k\to\infty}\int_{\mathbb{R}^{2D}} ((|x|^4+|v|^4)\wedge A) f_t^{N_k}(dx,dv)\right] \\
&= \sup_{t\in[0,T]} \lim_{k\to\infty} \frac{1}{N_k}\sum_{i=1}^{N_k} (\mathbb{E}\left[(|X_t^{i,N_k}|^4+|V_t^{i,N_k}|^4\wedge A)\right]) \\
&\overset{4.3.2}{\leq} K.
\end{aligned} \tag{4.23}$$

Lassen wir nun $A \to \infty$ gehen, erhalten wir

$$\sup_{t\in[0,T]} \mathbb{E}\left[\int\int_{\mathbb{R}^{2D}} (|x|^4+|v|^4) f_t(dx,dv)\right] \leq K.$$

Zusammen mit 4.3.1 impliziert das

$$\mathbb{E}\left[|X^{\alpha}(\rho_t)|^4\right] < \infty. \tag{4.24}$$

Nun können wir dank der punktweisen Konvergenz in der ersten Aussage, der gleichmäßigen Momentenabschätzung 4.3.2 und (4.24) dominierte Konvergenz anwenden und erhalten

$$\lim_{k\to\infty} \mathbb{E}\left[M_t^k(\phi)^2 + (D_t^k)^2\right] = 0 \tag{4.25}$$

Dies schließt den Beweis des Lemmas ab. ■

Es folgt ein Lemma, welches die Konvergenz in Verteilung unserer Teilfolge auf ihre zeitlichen Randmaße überträgt. Dies erfolgt im Hinblick auf die Bedingung (4.6) zeitlicher Stetigkeit von Definition 4.2.2.

Lemma 4.5.3: Konvergenz der zeitlichen Randmaße (Huang and Qiu, 2021, Lemma 2.3)

Seien $\{f^N\}_{N\in\mathbb{N}}$ wie in $\mathbb{A}$ (4.2.3). Sei ferner $\{f^{N_k}\}_{k\in\mathbb{N}}$ eine in Verteilung gegen ein $f \in \mathcal{P}(C([0,T];\mathbb{R}^{2D}))$ konvergente Teilfolge, also

$$f^{N_k} \xrightarrow[k\to\infty]{\mathcal{D}} f \quad \text{in } \mathcal{P}(C([0,T];\mathbb{R}^{2D})).$$

Dann gilt für die zeitlichen Randmaße $\{f_t^{N_k}\}_{k\in\mathbb{N}}$:

$$f_t^{N_k} \xrightarrow[k\to\infty]{\mathcal{D}} f_t \quad \text{in } \mathcal{P}(\mathbb{R}^{2D})$$

Hierbei ist für $\nu \in \mathcal{P}(C([0,T];\mathbb{R}^{2D}))$ und $t \in [0,T]$ das zeitliche Randmaß $\nu_t \in \mathcal{P}(\mathbb{R}^{2D})$ gegeben durch

$$\nu_t(B) := \nu(t,B) = \nu \odot \Pi_t^{-1}(B) = \nu(\Pi_t^{-1}(B)), \quad \forall B \in \mathcal{B}(\mathbb{R}^{2D})$$

Beweis: Zum Prüfen von schwacher Konvergenz nehmen wir uns ein $\phi \in C_b(\mathbb{R}^{2D})$, und können[7] dank der Beschränktheit der Funktion den Satz von Fubini (Johannes, 2021, Satz 04.43) anwenden, um zu dem Raum $C([0,T];\mathbb{R}^{2D})$ überzugehen:

$$\begin{aligned}\int_{\mathbb{R}^{2D}} \phi(z) f_t^{N_k}(dz) &= \int_{\mathbb{R}^{2D}} \phi(z) f^{N_k} \odot \Pi_t^{-1}(dz) \\ &\overset{FT}{=} \int_{C([0,T];\mathbb{R}^{2D})} \phi(z_t) f^{N_k}(dz)\end{aligned}$$

Mit dieser Überlegung lässt sich die schwache Konvergenz in $\mathcal{P}(\mathbb{R}^{2D})$ auf die schwache Konvergenz in $\mathcal{P}(C([0,T],\mathbb{R}^{2D})$ zurückführen:

$$\begin{aligned}\lim_{k\to\infty} \int_{\mathbb{R}^{2D}} \phi(z) f_t^{N_k}(dz) &= \lim_{k\to\infty} \int_{C([0,T];\mathbb{R}^{2D})} \phi(z_t) f^{N_k}(dz) \\ &= \int_{C([0,T];\mathbb{R}^{2D})} \phi(z_t) f(dz) \\ &= \int_{\mathbb{R}^{2D}} \phi(z_t) f(dz)\end{aligned} \tag{4.26}$$

Dies schließt den Beweis des Lemmas ab. ■

[7] (Li et al., 2019, Lemma 2.8). Diese Aussage gilt offensichtlich auch für $f \odot \Pi_t^{-1}$.

Lemma 4.5.4: $\mathscr{L}_1$-Konvergenz der Funktionale (Grassi et al., 2021, 18-20)

Seien $F, f_0, Z_t^{i,N}$ und f_t^N wie in $\mathbb{A}$ (4.2.3). Sei ferner $\{f_t^{N_k}\}_{k\in\mathbb{N}}$ eine in Verteilung gegen ein f konvergente Teilfolge der empirischen Maße. Dann gilt für das in (4.2) definierte Funktional:

$$\mathbb{F}_\varphi(f^{N_k}) \quad \overset{\mathscr{L}_1(\mathbb{P})}{\to} \quad \mathbb{F}_\varphi(f).$$

Beweis: Es ist zu zeigen, dass $\lim\limits_{k\to\infty}\mathbb{E}\left[|\mathbb{F}_\varphi(f^{N_k})-\mathbb{F}_\varphi(f)|\right]=0$. Dafür untersuchen wir die fünf Summanden des Funktionals (4.3) - (4.5) je auf $\mathscr{L}_1(\mathbb{P})$-Konvergenz.
Ausgehend von (4.25) gilt für eine Testfunktion mit kompaktem Träger $\varphi \in \mathscr{C}_c^2(\mathbb{R}^{2D}) \subset C_b(\mathbb{R}^{2D})$:

$$\lim_{k\to\infty}\mathbb{E}\left[|(\int \varphi d f_t^{N_k} - \int \varphi d f_0^{N_k}) - (\int \varphi d f_t - \int \varphi d f_0)|\right] = 0$$

Die folgende Differenz teilen wir auf:

$$\begin{aligned}
&|\int_0^t \int (x - X^\alpha(\rho_s^{N_k}))\nabla_v \varphi d f_s^{N_k} ds - \int_0^t \int (x - X^\alpha(\rho_s))\nabla_v \varphi d f_s ds| \\
&\overset{\text{Dreiecksungl.}}{\leq} \int_0^t |\int (x - X^\alpha(\rho_s^{N_k}))\nabla_v \varphi d f_s^{N_k} - \int (x - X^\alpha(\rho_s^{N_k}))\nabla_v \varphi d f_s| ds \\
&+ \int_0^t |\int (X^\alpha(\rho_s) - X^\alpha(\rho_s^{N_k}))\nabla_v \varphi d f_s| ds \\
&=: \int_0^t |I_1^k(s)| ds + \int_0^t |I_2^k(s)| ds
\end{aligned}$$

Nun gilt

$$\begin{aligned}
\mathbb{E}\left[|I_1^k(s)|\right] &\leq \mathbb{E}\left[|\int x\cdot\nabla_v\varphi d f_s^{N_k} - \int x\cdot\nabla_v\varphi d f_s|\right] \\
&+ \mathbb{E}\left[|X^\alpha(\rho_s^{N_k})\cdot(\int \nabla_v\varphi d f_s^{N_k} - \int \nabla_v\varphi d f_s)|\right] \\
&\overset{Jensen;\, 4.3.2}{\leq} \mathbb{E}\left[|\int x\cdot\nabla_v\varphi d f_s^{N_k} - \int x\cdot\nabla_v\varphi d f_s|\right] \\
&+ \sqrt{K}\mathbb{E}\left[|\int \nabla_v\varphi d f_s^{N_k} - \int \nabla_v\varphi d f_s|^2\right]^{\frac{1}{2}} \\
&\overset{k\to\infty}{\to} 0,
\end{aligned}$$

weil φ einen Kompakten Träger hat, und somit $x \cdot \nabla_v \varphi, \nabla_v \varphi \in \mathcal{C}_b(\mathbb{R}^{2D})$. Die obigen Terme sind durch (4.23) und (4.24) und die Abschätzungen in 4.3.2 gleichmäßig dominiert, also können wir für die Erwartung des zeitlichen Integrals den Satz von Fubini anwenden:

$$\lim_{k\to\infty} \mathbb{E}\left[\int_0^t |I_1^k(s)| ds\right] = \lim_{k\to\infty} \int_0^t \mathbb{E}\left[|I_1^k(s)|\right] ds \overset{\text{dom. Konv.}}{=} 0.$$

Die Erwartung des Terms $|I_2^k(s)|$ verschwindet ebenfalls:

$$\begin{aligned}
\mathbb{E}\left[|I_2^k(s)| ds\right] &= \mathbb{E}\left[|\int (X^\alpha(\rho_s) - X^\alpha(\rho_s^{N_k})) \cdot \nabla_v \varphi d f_s|\right] \\
&\leq ||\nabla_v \varphi||_\infty \mathbb{E}\left[|X^\alpha(\rho_s) - X^\alpha(\rho_s^{N_k})|\right] \\
&\underset{k\to\infty}{\overset{(4.22)}{\to}} 0.
\end{aligned}$$

Jetzt können wir wieder dominierte Konvergenz anwenden, um zu erhalten:

$$\lim_{k\to\infty} \mathbb{E}\left[\int_0^t |I_2^k(s)| ds\right] = \lim_{k\to\infty} \int_0^t \mathbb{E}\left[|I_2^k(s)|\right] ds \overset{\text{dom. Konv.}}{=} 0.$$

Somit haben wir die Erwartete Differenz des Summanden (4.4) abgeschätzt.
Für den Diffusionsterm (4.5) funktioniert dasselbe Vorgehen wie oben, indem man die Differenz der d-ten Summanden aufteilt:

$$\begin{aligned}
&|\int_0^t \int (x - X^\alpha(\rho_s^{N_k}))_d \frac{\delta^2}{\delta v_d^2} \varphi d f_s^{N_k} ds - \int_0^t \int (x - X^\alpha(\rho_s))_d \frac{\delta^2}{\delta v_d^2} \varphi d f_s ds| \\
&\overset{\text{Dreiecksungl.}}{\leq} \int_0^t |\int (x - X^\alpha(\rho_s^{N_k}))_d \frac{\delta^2}{\delta v_d^2} \varphi d f_s^{N_k} - \int (x - X^\alpha(\rho_s^{N_k}))_d \frac{\delta^2}{\delta v_d^2} \varphi d f_s| ds \\
&+ \int_0^t |\int (X^\alpha(\rho_s) - X^\alpha(\rho_s^{N_k}))_d \frac{\delta^2}{\delta v_d^2} \varphi d f_s| ds \\
&=: \int_0^t |I_{d,3}^k(s)| ds + \int_0^t |I_{d,4}^k(s)| ds.
\end{aligned}$$

Wie zuvor können wir dominierte Konvergenz anwenden, und erhalten

$$\lim_{k\to\infty} \mathbb{E}\left[|\int_0^t \sum_{d=1}^D (x - X^\alpha(\rho_s^{N_k}))_d \frac{\delta^2}{\delta v_d^2} \varphi d f_s^{N_k} - \int_0^t \sum_{d=1}^D (x - X^\alpha(\rho_s))_d \frac{\delta^2}{\delta v_d^2} \varphi d f_s|\right] = 0$$

Die Konvergenz des zweiten Summanden erhalten wir direkt durch Anwenden von (4.23) wegen $v \cdot \nabla_v \varphi \in \mathcal{C}_b(\mathbb{R}^{\mathbb{R}^{2D}})$:

$$\mathbb{E}\left[\Big| \int_0^t \int v_s \nabla_x \varphi d f_s^{N_k} ds - \int_0^t \int v_s \nabla_x \varphi d f_s ds \Big| \right] \overset{k\to\infty}{\to} 0$$

Analog erhält man die Konvergenz des folgenden Summanden:

$$\mathbb{E}\left[\Big| \int_0^t \int v_s \nabla_v \varphi d f_s^{N_k} ds - \int_0^t \int v_s \nabla_v \varphi d f_s ds \Big| \right] \overset{k\to\infty}{\to} 0$$

Somit haben wir die $\mathscr{L}_1(\mathbb{P})$-Konvergenz der Differenz aller Summanden von (4.2) gezeigt und erhalten mit der Dreiecksungleichung:

$$\lim_{k\to\infty} \mathbb{E}\left[|\mathbb{F}_\varphi(f^{N_k}) - \mathbb{F}_\varphi(f)| \right] = 0$$

Dies schließt den Beweis des Lemmas ab. ■

Zuletzt führen wir die obigen Ergebnisse dieses Unterabschnitts zusammen, um Schritt 3. des Beweises von 4.5.1 zu zeigen.

Lemma 4.5.5: Mean Field Limit konvergenter Teilfolgen der Partikelverteilungen (Grassi et al., 2021, Thm 1)

Seien $\mathbb{F}_\varphi, f_0, Z_t^{i,N}$ und f_t^N wie in $\mathbb{A}$ (4.2.3). Sei ferner $\{f_t^{N_k}\}_{k\in\mathbb{N}}$ eine in Verteilung gegen ein f konvergente Teilfolge der empirischen Maße.
Dann ist f eine schwache Lösung von (4.1) im Sinne der Bedingungen der Definition 4.2.2.

Beweis: Wir müssen Stetigkeit der zeitlichen Randmaße und $\mathbb{F}_\varphi(f) = 0$, $\mathbb{P}$-f.s. zeigen. Ersteres, also (4.6), folgt mit dominierter Konvergenz für beschränkte Testfunktionen $\phi \in \mathcal{C}_b(\mathbb{R}^{2D})$ dank Lemma 4.5.3:

$$\begin{aligned} &\int_{C([0,T];\mathbb{R}^{2D})} \phi(x_{t_n}, v_{t_n}) f(dx,dv) \to \int_{C([0,T];\mathbb{R}^{2D})} \phi(x,v) f(dx,dv) \\ \overset{(4.26)}{\iff} &\int_{\mathbb{R}^{2D}} \phi(x,v) f_{t_n}(dx,dv) \to \int_{\mathbb{R}^{2D}} \phi(x,v) f_t(dx,dv) \end{aligned}$$

Der zweite Teil der Bedingung, $\mathbb{F}_\varphi(f) = 0$ $\mathbb{P}$-f.s., folgt direkt aus der $\mathscr{L}_1(\mathbb{P})$-Konvergenz der Funktionale, gezeigt in Lemma 4.5.4, mit folgender Überlegung:

$$\begin{aligned}
\mathbb{E}\left[|\mathbb{F}_\varphi(f)|\right] &= \mathbb{E}\left[|\mathbb{F}_\varphi(f) - \mathbb{F}_\varphi(f^{N_k}) + \mathbb{F}_\varphi(f^{N_k})|\right] \\
&\overset{\text{dreiecksungl.}}{\leq} \mathbb{E}\left[|\mathbb{F}_\varphi(f) - \mathbb{F}_\varphi(f^{N_k})|\right] + \mathbb{E}\left[|\mathbb{F}_\varphi(f^{N_k})|\right] \\
&\overset{4.5.1}{\leq} \mathbb{E}\left[|\mathbb{F}_\varphi(f) - \mathbb{F}_\varphi(f^{N_k})|\right] + \frac{C}{\sqrt{N_k}} \\
&\underset{4.5.4}{\overset{k\to\infty}{\to}} 0.
\end{aligned}$$

Nach Übergehen zur Skorokhod-Repräsentation haben wir somit

$$\mathbb{F}_\varphi(f) = 0, \ \mathbb{P}\text{-f.s.}.$$

Das schließt den Beweis des Lemmas ab. ■

4.5.2 Eindeutigkeit des Grenzwerts

Im nächsten Schritt des Beweises von Satz 4.5.1 ist zu zeigen, dass alle schwach konvergenten Teilfolgen denselben Grenzwert haben. Hierbei ist es wichtig, den richtigen Begriff von Eindeutigkeit zu verwenden

Da wir mit schacher Konvergenz in $\mathcal{P}(\mathcal{P}(C([0,T];\mathbb{R}^{2D})))$ arbeiten, ist es notwendig, eine Metrik auf diesem Raum zu verwenden, die eine Topologie erzeugt, die identisch zu der von schwacher Konvergenz erzeugten (2.3.2) ist (Carrillo et al., 2018, Abschnitt 3). Eine solche Metrik ist die p-Wassersteinmetrik (2.3.7):

Lemma 4.5.6: Identifikation schwacher Grenzwerte (Villani, 2009, Thm 6.9)

Sei $q \in [1,\infty)$ und $g^N, g, h \in \mathcal{P}_q(C([0,T];\mathbb{R}^{2D}))$, sodass $g^N \underset{N\to\infty}{\overset{\mathcal{D}}{\to}} g$.
Für $p \in [1,\infty)$ impliziert dann $W_p(\mathbb{P}^g, \mathbb{P}^h) = 0$, dass $g^N \underset{N\to\infty}{\overset{\mathcal{D}}{\to}} h$.

Beweis: In Lemma 4.1.2 haben wir gesehen, dass $(\mathcal{P}_p(\mathcal{P}_q(C([0,T];\mathbb{R}^{2D}))), W_p)$ polnisch bezüglich der Topologie der schwachen Konvergenz τ_w ist, das heißt, es gilt (Villani, 2009, Thm. 6.9):

$$\mathbb{P}^{g^N} \overset{N\to\infty}{\to} \mathbb{P}^h \quad \Longleftrightarrow \quad W_p(\mathbb{P}^{g^N}, \mathbb{P}^h) \overset{N\to\infty}{\to} 0.$$

Also haben wir

$$W_p(\mathbb{P}^{g^N}, \mathbb{P}^h) \leq W_p(\mathbb{P}^{g^N}, \mathbb{P}^g) + W_p(\mathbb{P}^g, \mathbb{P}^h) \overset{N \to \infty}{\to} 0.$$

Das schließt den Beweis ab. ■

Das folgende Lemma liefert die Eindeutigkeit der Lösung der FPG entsprechend der obigen Charakterisierung:

Lemma 4.5.7: Eindeutigkeit der Schwachen Lösungen der FPG (Huang and Qiu, 2021, Lemma 3.2)

Seien die Maße $f^1, f^2 \in \mathcal{P}_2(C([0,T], \mathbb{R}^{2D}))$ zwei Schwache Lösungen im Sinne von 4.2.2 der FPG (4.1) mit derselben Anfangsverteilung $f_0 \in \mathcal{P}_4(\mathbb{R}^{2D})$.
Dann gilt

$$W_2(\mathbb{P}^{f^1}, \mathbb{P}^{f^2}) = 0,$$

wobei W_2 die 2-Wassersteinmetrik auf $\mathcal{P}_2(\mathcal{P}_2(C([0,T];\mathbb{R}^{2D})))$ ist.

Beweis: Zu den zwei vorgegebenen Maßen $f^i, i = 1,2$ konstruieren wir je einen Prozess als starke Lösung der SDGL

$$\begin{aligned} md\hat{V}_t^i &= (1-m)\hat{V}_t^i dt + \lambda(X^\alpha(\rho_t^i) - \hat{X}_t^i)dt + \sigma\mathbb{D}(X^\alpha(\rho_t^i) - \hat{X}_t^i)dB_t \\ d\hat{X}_t^i &= \hat{V}_t^i dt, \\ \text{mit} \quad \rho_t^i &:= \int_{\mathbb{R}^D} f_t^i dy \end{aligned}$$

mit identischem Anfangswert $\hat{Z}_0 \sim f_0$. Diese SDGL ist linear, da im Gegensatz zur Definition des McKean-Prozess' (4.8) die Maße f^i fest vorgegeben sind. Wir definieren $\hat{f}_t^i := \mathbb{P}^{\hat{Z}_t^i}$ und $\hat{Z}_t^i := [\hat{X}_t^i, \hat{V}_t^i]$. Mit der Fokker-Planck-Gleichung erhalten wir, dass die beiden Maße $\hat{f}_t^i, i = 1,2$, schwache Lösungen im Sinne von 4.2.1 der linearen PDGLn:

$$\begin{aligned} \frac{\partial \hat{f}_t^i}{\partial t} = -v \cdot \nabla_x \hat{f}_t^i + \nabla_v \cdot (\frac{1-m}{m} v \hat{f}_t^i + \frac{\lambda}{m}(x - b_t) + \\ \frac{\sigma^2}{2m^2}\mathbb{D}((x-b_t)^2 \nabla_v \hat{f}_t^i)) \end{aligned} \tag{4.27}$$

sind. Die schwache Lösung dieser beiden PDGLn ist jeweils laut dem folgenden Lemma 4.5.8 eindeutig. Da die obige PDGL (4.27) auch von f^i gelöst wird, gilt also $\hat{f}_t^i = f_t^i, i =$

$1,2$, für alle $t \in [0,T]$. Der mit f^i assoziierte McKean-Prozess $(\overline{Z}_t^i)_{t\in[0,T]}$ (die Lösung von (4.8)) ist allerdings bei gleichen Anfangswerten eindeutig bis auf Ununterscheidbarkeit, was aufgrund der lokalen Lipschitzstetigkeit auf Zylindermengen, die analog zu Satz 3.3.1 erhalten werden kann, durch Lemma 2.2.1 folgt.
Also sind $\overline{Z}^1, \overline{Z}^2, \hat{Z}^1, \hat{Z}^2$ ununterscheidbar; daher haben wir

$$\begin{aligned}
W_2(\mathbb{P}^{f^1}, \mathbb{P}^{f^2})^2 &:= \inf_{\pi\in\Pi(\mathbb{P}^{f^1},\mathbb{P}^{f^2})} \int\int_{\mathcal{P}_2(C([0,T];\mathbb{R}^{2D}))^2} W_2'(\mu,\nu)^2 \pi(d\mu, d\nu) \\
&\leq \mathbb{E}\left[W_2'(f^1,f^2)^2\right] \\
&\leq \mathbb{E}\left[\mathbb{E}\left[d(\overline{Z}^1,\overline{Z}^2)^2\right]\right] \\
&= 0.
\end{aligned}$$

Hierbei ist W_2' die 2-Wassersteinmetrik auf $\mathcal{P}(C([0,T];\mathbb{R}^{2D}))$ und $d(\cdot,\cdot)$ beispielsweise die Skorokhodmetrik (Billingsley, 1999, 124).

Lemma 4.5.8: Existenz und Eindeutigkeit der Schwachen Lösung der FPG mit vorgegebenem Global Best (Huang and Qiu, 2021, Theorem 4.3)

Sei $T > 0$, $b \in C([0,T];\mathbb{R}^{2D})$ und $f_0 \in \mathcal{P}_2(\mathbb{R}^{2D})$. Dann hat die folgende lineare[a] PDGL eine eindeutige schwache Lösung

$$\begin{aligned}
\frac{\partial f_t}{\partial t} &= -v \cdot \nabla_x f_t + \nabla_v \cdot \left(\frac{1-m}{m} v f_t + \frac{\lambda}{m}(x - b_t)\right. \\
&\quad \left. + \frac{\sigma^2}{2m^2} \mathbb{D}((x-b_t)^2 \nabla_v f_t)\right).
\end{aligned} \tag{4.28}$$

[a]Der Unterschied zu der FPG (4.1) ist, dass das regularisierte *gbest* hier vorgegeben ist, also nicht vom räumlichen Randmaß der Lösung abhängt.

Beweis (Skizze): Zunächst ist die Existenz zu zeigen; sie folgt direkt durch die Fokker-Planck-Gleichung (2.8) als Verteilung der Lösung $\tilde{Z}_t := [\tilde{X}_t; \tilde{V}_t]$ der folgenden linearen SDGL:

$$\begin{aligned}
md\tilde{V}_t &= (1-m)\tilde{V}_t dt + \lambda(b_t - \tilde{X}_t)dt + \sigma\mathbb{D}(b_t - \tilde{X}_t)dB_t \\
d\tilde{X}_t &= \tilde{V}_t dt.
\end{aligned} \tag{4.29}$$

Das Vorgehen ist dabei analog zu Satz 4.2.1.

Die Eindeutigkeitsaussage ist schwieriger zu erhalten; wir umreißen hier lediglich den Beweis anhand von (Huang and Qiu, 2021, Theorem 4.3), wo er für CBO geführt wird; das heißt dass sich die dort untersuchte PDGL von der obigen PDGL (4.28) nur durch das Fehlen des Trägheitsterms $\nabla_v \cdot \frac{1-m}{m} v \hat{f}_t^i$ und des hier durch das Verwenden einer Geschwindigkeitsvariable entstehenden Terms $-v \cdot \nabla_x \hat{f}_t^i$ unterscheidet.

Für $t_0 \in (0,T]$ und eine glatte Testfunktion $\psi \in \mathcal{C}_c^\infty(\mathbb{R}^{2D})$ konstruieren wir eine klassische Lösung $h \in \mathcal{C}^{1,2}([0,t_0], \mathbb{R}^{2D})$ der Kolmogorov Rückwärtsgleichung (2.9) von (4.29), nämlich

$$
\begin{aligned}
-\frac{\partial h_t}{\partial t} &= -v \cdot \nabla_x h_t + \nabla_v \cdot (\frac{1-m}{m} v h_t + \frac{\lambda}{m}(x - b_t) \\
&\quad + \frac{\sigma^2}{2m^2} \mathbb{D}((x-b_t)^2 \nabla_v h_t), \\
&\quad \text{mit} \quad (t,x,v) \in [0,t_0] \times \mathbb{R}^D \times \mathbb{R}^D; \quad h_{t_0} \stackrel{!}{=} \psi.
\end{aligned}
\tag{4.30}
$$

Diese kann konstruiert werden, indem man setzt

$$
h_t(z) := \mathbb{E}\left[\psi(\hat{Z}_{t_0}^{t,z})\right],
$$

wobei zu gegebenem Anfangswert $z = [x,v]$ hier $\{\hat{Z}_{t_0}^{t,z} := [\hat{X}_{t_0}^{t,x}; \hat{V}_{t_0}^{t,v}]\}_{s \in [t,t_0]}$ die starke Lösung der obigen SDGL (4.29) ist[8].

An dieser Stelle postulieren wir lediglich die ausreichende Regularität von h_t und dass es (4.29) löst; beides sollte ähnlich zu dem Vorgehen in (Huang and Qiu, 2021, Theorem 4.3) erhalten zu sein.

Um mittels h_t die Eindeutigkeit zu zeigen, nehmen wir schließlich an, dass f^1, f^2 zwei Lösungen der PDGL (4.28) mit der selben Anfangsverteilung $f_0 = f_0^1 = f_0^2$ sind. Ferner setzen wir $\delta f := f^1 - f^2$. Für unsere Testfunktion $\psi \in \mathcal{C}_c^\infty(\mathbb{R}^{2D})$ können wir jetzt schrei-

[8] Existenz der starken Lösung folgt analog zu Satz 3.3.1.

ben:

$$
\begin{aligned}
\int \psi(z) d\delta f_{t_0} &= \int h_{t_0}(z) d\delta f_{t_0} \\
&\overset{(4.30)}{=} \int_0^{t_0} \int \partial_s h_s(z) d\delta f_s ds + \int_0^{t_0} \int -v \cdot \nabla_x h_s - \\
&+ \nabla_v \cdot \left(\frac{1-m}{m} v h_s + \frac{\lambda}{m}(x - b_s)\right) \\
&+ \frac{\sigma^2}{2m^2} \mathbb{D}((x-b_s)^2 \nabla_v h_s) d\delta f_s ds \\
&= \int_0^{t_0} \int \partial_s h_s(z) d\delta f_s ds + \int_0^{t_0} \int -\partial_s h_s(z) d\delta f_s ds \\
&= 0.
\end{aligned}
$$

Da die Testfunktion und $t_0 \in (0, T]$ beliebig waren, gilt also $f^1 = f^2$. ■

Somit haben wir alle Schritte der Beweisstrategie von Satz 4.5.1 gezeigt und den Satz bewiesen.

4.6 Abschluss des Beweises

Nun haben wir mittels Satz 4.5.1 die Voraussetzungen von Satz 4.2.2 im Sinne von Definition 4.2.2 gezeigt. Indem wir auf die Verteilung des McKean-Prozess $\overline{Z}$, der Lösung von (4.8), die FPG aus Satz 2.4.1 anwenden, schließen wir den Beweis ab.

Kapitel 5

Zusammenfassung

In dieser Arbeit haben wir gesehen, wie Maßtheorie, Wahrscheinlichkeitstheorie und das Ito-Kalkül verwendet werden können, um nachzuweisen, wie sich die Verteilungen der Partikelpositionen und -geschwindigkeiten einer reduzierten, regularisierten Formulierung des Partikelschwarmoptimierers im Grenzwert der Partikelzahl gegen unendlich verhalten. Dabei waren neben dem Wechseln zu dieser Umformulierung des Algorithmus' einige einschränkende Annahmen auf die Beschaffenheit der Kostenfunktion nötig. Wir haben dabei stellenweise Ergebnisse erhalten, die in (Grassi et al., 2021; Grassi and Pareschi, 2021; Huang and Qiu, 2021) nicht exakt für das *reduzierte SD-PSO* gezeigt worden waren, sondern dort direkt aus Ergebnissen für *CBO* übernommen wurden.
Wir heben diese Unterschiede hier noch einmal hervor:

1. Die hier in Satz 3.3.1 erhaltene eindeutige starke Lösbarkeit des *reduzierten SD-PSO* wurde nach unserem Wissen noch nicht gezeigt, sondern bisher nur für *CBO*, z.B. in (Carrillo et al., 2018, Thm. 3.2).

2. Die gleichförmige Abschätzung der Momente wurde bisher nicht für das reduzierte SD-PSO, sondern lediglich in (Carrillo et al., 2018, Sec. 3.2) für *CBO* angegeben. Die dort erhaltene Schranke wächst mit in der Zeit mit $\exp(T^3)$; hier wurde eine Schranke erhalten, die lediglich proportional zu $\exp(T)$ wächst.

3. Anders als in (Huang and Qiu, 2021, 3.2) haben wir in 4.5.7 die Eindeutigkeit der Lösung der FPG unter expliziter Kompatibilität der verwendeten Metrik (Lemma 4.5.6) mit der von der schwachen Konvergenz in $\mathcal{P}(\mathcal{P}(C([0,T];\mathbb{R}^{2D})))$ erzeugten Topologie geprüft. Dabei haben wir angenommen, dass der McKean-Prozess fast

sicher rechtsstetig ist; es gibt auch andere Methoden, den fehlenden Schritt durchzuführen; beispielsweise mittels Integration bezüglich Testfunktionen.

5.1 Ausblick

Wie am Anfang der Arbeit erwähnt, ist das Erhalten des Mean-Field Limits des Systems lediglich ein Zwischenschritt in der Analyse von PSO mittels SDGLn.
Ein nächster Schritt wäre, für das hier *vollständig* genannte *SD-PSO* die Gültigkeit des Mean-Field Limits zu prüfen. Grassi et al. (2021) geben an, dass dies mit einer analogen Beweisstrategie zu dem *reduzierten SD-PSO* zu erhalten sein sollte. Die zentralen rechnerischen Schritte wären hierbei zu zeigen, dass das System (3.6) lokale Lipschitzkonstanten und lokal linear beschränktes Wachstum hat, und eine gleichförmige Momentenabschätzung gilt. Wie in (Grassi et al., 2021, Sec. 3.2) umrissen wird, sollte die Speicherkomponente 3.2.3 dabei leicht zu behandeln sein.

Kapitel 6

Eigenständigkeitserklärung

Hiermit erkläre ich, dass ich die Vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Alle sinngemäß und wörtlich übernommenen Textstellen aus fremden Quellen wurden kenntlich gemacht.
I hereby declare that I have written the present thesis independently and have not used any sources or tools other than those indicated. All text passages from external sources have been marked accordingly.

Heidelberg, den 28.Juli 2022,

Marvin Koß

Literaturverzeichnis

Patrick Billingsley. *Convergence of probability measures, 2nd Edition*. John Wiley & Sons, 1999.

François Bolley, José A Canizo, and José A Carrillo. Stochastic mean-field limit: non-lipschitz forces and swarming. *Mathematical Models and Methods in Applied Sciences*, 21(11):2179–2210, 2011a.

François Bolley, José A Canizo, and José A Carrillo. Stochastic mean-field limit: non-lipschitz forces and swarming. *Mathematical Models and Methods in Applied Sciences*, 21(11):2179–2210, 2011b.

Mohammad Reza Bonyadi and Zbigniew Michalewicz. Particle swarm optimization for single objective continuous space problems: a review. *Evolutionary computation*, 25 (1):1–54, 2017.

José A Carrillo, Young-Pil Choi, Claudia Totzeck, and Oliver Tse. An analytical framework for consensus-based global optimization method. *Mathematical Models and Methods in Applied Sciences*, 28(06):1037–1066, 2018.

Jürgen Elstrodt. *Maß-und Integrationstheorie*, volume 6. Springer, 1996.

Sara Grassi and Lorenzo Pareschi. From particle swarm optimization to consensus based optimization: stochastic modeling and mean-field limit. *Mathematical Models and Methods in Applied Sciences*, 31(08):1625–1657, 2021.

Sara Grassi, Hui Huang, Lorenzo Pareschi, and Jinniao Qiu. Mean-field particle swarm optimization. *arXiv preprint arXiv:2108.00393*, 2021.

Hui Huang. A note on the mean-field limit for the particle swarm optimization. *Applied Mathematics Letters*, 117:107133, 2021.

Hui Huang and Jinniao Qiu. On the mean-field limit for the consensus-based optimization. *arXiv preprint arXiv:2105.12919*, 2021.

Hui Huang, Jinniao Qiu, and Konstantin Riedl. On the global convergence of particle swarm optimization methods. *arXiv preprint arXiv:2201.12460*, 2022.

Kiyosi Itô. Stochastic integral. *Proceedings of the Imperial Academy*, 20(8):519–524, 1944.

Pierre-Emmanuel Jabin and Zhenfu Wang. Mean field limit for stochastic particle systems. In *Active Particles, Volume 1*, pages 379–402. Springer, 2017.

Jan Johannes. *Vorlesungsskript Wahrscheinlichkeitstheorie I*. 2021. URL `https://sip.math.uni-heidelberg.de/vl/wt1-ss21/src/Skript-WT1-SS21-%C2%A701-%C2%A719.pdf`.

James Kennedy and Russell Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95-international conference on neural networks*, volume 4, pages 1942–1948. IEEE, 1995.

Rafail Khasminskii. *Stochastic stability of differential equations*, volume 66. Springer Science & Business Media, 2011.

Achim Klenke. *Wahrscheinlichkeitstheorie*, volume 1. Springer, 2006.

Lei Li, Jian-Guo Liu, and Pu Yu. On the mean field limit for brownian particles with coulomb interaction in 3d. *Journal of Mathematical Physics*, 60(11):111501, 2019.

Wei Liu and Michael Röckner. *Stochastic partial differential equations: an introduction*. Springer, 2015.

Bernt Øksendal. Stochastic differential equations. Springer, 2003.

Grigorios A Pavliotis. Stochastic processes and applications. *Informe técnico*, 2015.

René Pinnau, Claudia Totzeck, Oliver Tse, and Stephan Martin. A consensus-based model for global optimization and its mean-field limit. *Mathematical Models and Methods in Applied Sciences*, 27(01):183–204, 2017.

Eckhard Platen. An introduction to numerical methods for stochastic differential equations. *Acta numerica*, 8:197–246, 1999.

Riccardo Poli, James Kennedy, and Tim Blackwell. Particle swarm optimization. *Swarm intelligence*, 1(1):33–57, 2007.

Yuhui Shi and Russell Eberhart. A modified particle swarm optimizer. In *IEEE 1998*, pages 69–73. IEEE, 1998.

Yuhui Shi et al. Particle swarm optimization: developments, applications and resources. In *Proceedings of the 2001 congress on evolutionary computation (IEEE Cat. No. 01TH8546)*, volume 1, pages 81–86. IEEE, 2001.

Alain-Sol Sznitman. Topics in propagation of chaos. In *Ecole d'été de probabilités de Saint-Flour XIX—1989*, pages 165–251. Springer, 1991.

Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2009.